RとJASPで学ぶフリーで楽しい心理統計の世界

心理学データ解析 基礎

「見えないもの」をみよう・測ろう・考えよう

小杉考司

この本は Creative Commons BY-SA(CC BY-SA) ライセンス Version 4.0 に基づいて提供されています。著者に適切なクレジットを与える限り，この本を再利用，再編集，保持，改訂，再頒布 (商用利用を含む) をすることができます。もし再編集したり，このオープンなテキストを変更したい場合，すべてのバージョンにわたってこれと同じライセンス，CC BY-SA を適用しなければなりません。
https://creativecommons.org/licenses/by-sa/4.0/deed.ja

This book is published under a Creative Commons BY-SA license (CC BY-SA) version 4.0. This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license - CC BY-SA.

https://creativecommons.org/licenses/by-sa/4.0/

心理学データ解析基礎

小杉 考司

Last Compiled on 2025.1.13

はじめに

授業のテーマ

「見えないもの」をみよう・測ろう・考えようとするのが心理統計の世界である。

一年を通じて伝えたいポイント

母数を知るための推測　大学での統計，心理統計は標本の記述統計量を検証したいのではなく，母集団を表す数字を求めるための推測統計である。

確率を用いた推論　母数を知るための方法は 3 つある。代表値を用いたモーメント法，確率分布をつかった最尤法，ベイズ法である。

要因計画と線形モデル　心理学では，要因計画による研究条件をコントロールし，平均値の差をもちいた考察 (平均因果効果) をおこなうことが多い。そしてそれらは総じて線形モデルとして表現できる。

モデル比較と意思決定　モデルによる母数の推定だけではなく，そこから一定の「結論」あるいは「意思決定」を行うための方法として，モデル比較や NHST といった方法がある。

統計環境 R による実践　統計環境 R を利用して，理論的な理解だけでなく実際に計算ができることも身につけるべき技術である。

注)「科目の中でのこのコマの位置づけ」の中で示されているのは，上記のポイントに対応した，その時間の進度 (Level) のようなものです[*1]。たとえば，次のようにあった場合，

> 記述統計 [3]　確率 [2]　モデル [1]　推測・推定 [1]　検定 [0]　実践 [1]

記述統計は第 3 段階，確率についての話は第 2 段階，モデルについての話は第 1 段階，推測・推定についての話はまだ含まれておらず，実践は第 1 段階にきたことを表します。この進度は，次の表 1 に対応し，累積するものです。

なお，各コマにおける「標準シラバスにおける位置づけ」とは，公認心理師大学カリキュラム　標準シラバス (2018 年 8 月 22 日版)[*2] との対応を表しています。

その他

授業シラバスとこの講義資料を掲載したサイト (https://kosugitti.github.io/psychometrics_syllabus/) で，最新版のシラバスと授業資料，授業で用いるサンプルデータやコードの配布を行なっています。

[*1] 記述統計は「一年を通じて伝えたいポイント」の基礎および準備であり，本質ではありませんので「ポイント」には含まれませんが，「授業の進度」には含まれます。

[*2] https://psych.or.jp/wp-content/uploads/2018/04/standard_syllabus_2018-8-22.pdf

表 1　テーマと講義との対応

	記述統計	確率	モデル	推測・推定	検定	演習	page
心理学と統計法の関係							p.15
数値データの種類	1						p.21
統計環境 R の導入						1	p.37
統計環境 R の利用	2					2	p.59
記述統計量	3						p.81
測定の基礎			1				p.95
誤差と確率		1					p.107
相関と因果			2				p.129
単回帰分析			3				p.141
重回帰分析			4				p.153
R をつかった回帰分析			5			3	p.165
実験計画法 1			6				p.177
実験計画法 2			7				p.189
実験計画法 3			8				p.199
確率と統計的推論		2		1			p.213
推論と判断		3			1		p.223
帰無仮説検定		4			2		p.233
帰無仮説検定の注意点		5			3		p.243
二群の平均値差		6	9	2	4	4	p.253
要因計画と平均値差の検定		7	10		5		p.269
R による分散分析		8	11	3	6	5	p.285
確率分布とデータの関係		9	12	4			p.305
回帰分析の最尤推定		10	13	5			p.315
ベイズ推定の基礎		11		6			p.325
ベイズ推定の実際		12	14	7			p.335
ベイズ推定と判断		13	15	8			p.347
JASP によるベイズ推定		14		9		6	p.357
R によるベイズ推定		15	16	10		7	p.373

目次

第 I 部　心理学データ解析基礎 1A(心理学統計法) 13

第 1 章　心理統計を始める前に 15
1.1 はじめに . . . 15
1.2 心理学とはどういう学問か . . . 16
1.3 近代科学の特徴 . . . 17
1.4 心理学のいとなみ . . . 19
1.5 課題 . . . 20

第 2 章　数値データの種類 21
2.1 データと変数 . . . 21
2.2 変数の分類 . . . 24
2.3 尺度水準による分類 . . . 27
2.4 データの可視化 . . . 30
2.5 付録 . . . 34
2.6 課題 . . . 35

第 3 章　統計環境 R の導入 37
3.1 パソコンのファイル構造 . . . 37
3.2 統計環境 R/RStudio の準備 . . . 40
3.3 R/RStudio の基本操作 . . . 42
3.4 R の基本操作 . . . 49
3.5 再現性のある研究実践へ . . . 55
3.6 課題 . . . 57

第 4 章　統計環境 R の利用 59
4.1 Rmarkdown による文書作成 . . . 59
4.2 外部データの読み込み . . . 64
4.3 グラフの文法 . . . 70
4.4 課題 . . . 72
4.5 補遺 . . . 74

第 5 章　記述統計量 81
5.1 中心化傾向の指標 . . . 82

5.2 散らばりの指標 . . . 85
5.3 データの相対的位置 . . . 89
5.4 標準化された差 . . . 91
5.5 課題 . . . 93

第 6 章 測定の基礎 95
6.1 テストと心理学 . . . 95
6.2 測定のモデル . . . 98
6.3 誤差の基本的性質 . . . 98
6.4 テストの分散と信頼性･妥当性の関係 . . . 100
6.5 課題 . . . 105

第 7 章 誤差と確率 107
7.1 確率と呼ばれる数字 . . . 107
7.2 確率分布 . . . 108
7.3 ベルヌーイ分布と二項分布 . . . 113
7.4 正規分布の形 . . . 115
7.5 正規分布の関数 . . . 120
7.6 標準正規分布の利用 . . . 123
7.7 課題 . . . 124
7.8 補遺 . . . 125

第 8 章 相関と因果 129
8.1 変数間関係の指標 . . . 129
8.2 相関係数の落とし穴 . . . 132
8.3 因果関係 . . . 136
8.4 交絡という罠 . . . 138
8.5 課題 . . . 139

第 9 章 線形モデルの基礎 141
9.1 相関関係とモデル . . . 141
9.2 予測と線形モデル . . . 144
9.3 最小二乗法 . . . 145
9.4 回帰分析の性質 . . . 147
9.5 予測の評価 . . . 150
9.6 課題 . . . 151

第 10 章 重回帰分析 153
10.1 重回帰分析 . . . 153
10.2 部分相関･偏相関･偏回帰係数 . . . 156
10.3 標準化された係数 . . . 161
10.4 今日のまとめ . . . 162
10.5 課題 . . . 162

第 11 章　R をつかった回帰分析 165
11.1 R の環境再確認 . . . 165
11.2 線形回帰の実行 . . . 168
11.3 残差からわかること . . . 169
11.4 重回帰分析 . . . 171
11.5 課題 . . . 174

第 12 章　実験計画法 1；実験計画法とは 177
12.1 ランダム化 . . . 177
12.2 用語の整理 . . . 179
12.3 実験でなぜ効果がわかるのか . . . 182
12.4 効果はどのように評価するのか . . . 184
12.5 課題 . . . 186

第 13 章　実験計画法 2；Between 計画 189
13.1 二要因の場合 . . . 189
13.2 効果の大きさを算出する . . . 191
13.3 交互作用 . . . 192
13.4 課題 . . . 198

第 14 章　実験計画法 3；一要因モデル (Within) 199
14.1 Within 計画の場合 . . . 199
14.2 群間計画と群内計画の比較 . . . 201
14.3 線形モデルによる表現と自由度 . . . 203
14.4 分布を使って考える . . . 206
14.5 課題 . . . 208

第 15 章　前期のまとめ 209

第 II 部　心理学データ解析基礎 1B 211

第 16 章　確率と統計的推論 213
16.1 母集団と標本 . . . 213
16.2 母集団分布とモデル . . . 216
16.3 標本分布と標準誤差 . . . 217
16.4 まとめ . . . 221
16.5 課題 . . . 221

第 17 章　推論と判断 223
17.1 不偏推定量 . . . 223
17.2 標本平均の区間推定 . . . 225
17.3 標本だけから推測したい . . . 228
17.4 推定と検定 . . . 229

17.5 課題 . . . 232

第 18 章 帰無仮説検定 233
18.1 帰無仮説検定の手続き . . . 234
18.2 1 つの平均値の検定 . . . 236
18.3 相関係数の検定 . . . 238
18.4 課題 . . . 241

第 19 章 帰無仮説検定の注意点 243
19.1 結果の報告の仕方 . . . 243
19.2 統計的に有意とは . . . 244
19.3 検定における 2 種類の間違い . . . 246
19.4 効果量と例数設計 . . . 248
19.5 課題 . . . 250

第 20 章 二群の平均値差 253
20.1 一要因 Within デザイン，2 水準 . . . 253
20.2 一要因 Between デザイン，2 水準 . . . 256
20.3 R による実践;対応のある t 検定 . . . 260
20.4 R による実践;対応のない t 検定 . . . 262
20.5 課題 . . . 267

第 21 章 要因計画と平均値差の検定 269
21.1 平均値の比較とモデル比較 . . . 269
21.2 一要因 Between デザインの検定 . . . 273
21.3 二要因計画モデルの検定 . . . 279
21.4 課題 . . . 280

第 22 章 R による分散分析 285
22.1 R で分散分析を行う上でのポイント . . . 285
22.2 一要因 Between 計画の実際 . . . 288
22.3 二要因 Between 計画の実際 . . . 292
22.4 一要因 Within 計画の実際 . . . 297
22.5 課題 . . . 301

第 23 章 確率分布とデータの関係 305
23.1 確率モデルによる推論 . . . 305
23.2 確率関数と尤度関数 . . . 306
23.3 最尤推定 . . . 307
23.4 対数尤度関数 . . . 308
23.5 正規分布の尤度関数 . . . 309
23.6 課題 . . . 314

第 24 章 回帰分析の最尤推定 315

24.1 回帰分析と誤差分布 . . . 315
24.2 確率モデル . . . 316
24.3 確率モデルの最尤推定 . . . 318
24.4 課題 . . . 323

第 25 章 ベイズ推測の基礎 325
25.1 ベイズの定理 . . . 325
25.2 ベイズの定理の意味 . . . 328
25.3 ベイズの定理を使った推論の例 . . . 331
25.4 課題 . . . 333

第 26 章 ベイズ推測の実際 335
26.1 事前分布と事後分布 . . . 335
26.2 事後分布を代表する数字 . . . 339
26.3 3 つの推定法 . . . 341
26.4 事前分布と尤度の折衷 . . . 343
26.5 課題 . . . 345

第 27 章 ベイズ推定と判断 347
27.1 ベイズ法と帰無仮説検定 . . . 347
27.2 区間による判定 . . . 349
27.3 ベイズファクター . . . 352
27.4 ベイズ推定の利点 . . . 353

第 28 章 JASP によるベイズ推定の実践 357
28.1 ベイズ推定の方法 . . . 357
28.2 JASP の導入 . . . 361
28.3 JASP による相関･回帰分析 . . . 365
28.4 平均値の比較 . . . 368

第 29 章 R によるベイズ推定 373
29.1 環境の準備 . . . 373
29.2 回帰分析の推定 . . . 374
29.3 要因計画の推定 . . . 380
29.4 群内モデルの推定 . . . 382
29.5 課題 . . . 384

第 30 章 後期のまとめ 387

付録 A 電子計算機のイロハ 391
A.1 前置き . . . 391
A.2 コンピュータの基礎 . . . 391
A.3 コンピュータの歴史 . . . 392
A.4 情報の単位 . . . 395

A.5 ファイルの種類と拡張子 . . . 396
A.6 クラウドとは . . . 398
A.7 ファイルの位置の指定 . . . 399
A.8 ファイルのバージョン管理 . . . 401
A.9 おわりに . . . 402

付録 B よくある質問とミスの例 405
B.1 Frequently Miss and Comments . . . 405
B.2 Frequently Asked Questions;よくある質問と答え . . . 409

付録 C ギリシア文字一覧 417

付録 D 記号の入力とキーボードの場所 419

付録 E 本講義に対応する詳細シラバス 423
E.1 心理学と統計法 . . . 423
E.2 数値データの種類 . . . 424
E.3 統計環境 R の導入 . . . 426
E.4 統計環境 R の利用 . . . 428
E.5 記述統計量 . . . 429
E.6 測定の基礎 . . . 431
E.7 誤差と確率 . . . 433
E.8 相関と因果 . . . 435
E.9 単回帰分析 . . . 436
E.10 重回帰分析 . . . 438
E.11 R をつかった回帰分析 . . . 439
E.12 実験計画法 1;実験計画法とは . . . 441
E.13 実験計画法 2; Between 計画 . . . 442
E.14 実験計画法 3;一要因モデル (Within) . . . 444
E.15 前期のテスト . . . 446
E.16 確率と統計的推論 . . . 446
E.17 推論と判断 . . . 448
E.18 帰無仮説検定 . . . 450
E.19 帰無仮説検定の注意点 . . . 451
E.20 二群の平均値差 . . . 453
E.21 要因計画と平均値差の検定 . . . 454
E.22 R による分散分析 . . . 456
E.23 確率分布とデータの関係; 尤度 . . . 457
E.24 回帰分析の最尤推定 . . . 459
E.25 ベイズ統計の基礎 . . . 460
E.26 ベイズ推定の実際 . . . 462
E.27 ベイズ推定と判断 . . . 463
E.28 JASP によるベイズ推定の実践 . . . 465

E.29 R によるベイズ推定 466
E.30 後期のテスト 468

引用文献 468

索引 473

第 I 部

心理学データ解析基礎 1A(心理学統計法)

第 1 章

心理統計を始める前に

これから「心理統計学」を学んでいくわけですが，皆さんはそもそも心理学と統計学が深く関わっていることをイメージしていましたでしょうか。もし心理学なのに数学の話があるの？ 心理学と統計になんの関係があるの？ と思った人は，そもそも心理学というのはどういうものをイメージしてましたでしょうか。

第 1 回ではまず，心理学を学ぶにあたってなぜ統計学が必要なのかを考えます。そのためには，心理学が科学 (サイエンス) であること，科学であるとはどういうことかを知る必要があり，そのために数量的表現が有用であるから，ということをしっかりと理解してもらわなければなりません。

1.1 はじめに

みなさん，心理統計の世界，いいえ心理学の世界へようこそいらっしゃいました。

みなさんは心理学という言葉を，はじめて聞いたのはいつだったのか覚えてないのではないでしょうか。それぐらい一般的に使われていますが，さてそこではどのようなことを研究しているのか，皆さんはどんなイメージをお持ちでしょうか。皆さんが何をきっかけで心理学を志したのかによって，それは随分と異なってくると思います。

ある人はテレビドラマ「逃げるは恥だが役に立つ」をみて，主人公の女性が心理学の大学院をでたことから興味を持ったのかもしれません。あるいはスクールカウンセラーが学校にいたから，なんなら少し相談に乗ってもらったから，心理職というのがどのようなものなのか興味を持ったのかもしれません。

あるいはまた，テレビドラマ「科捜研の女」をみて，主人公が働く科学捜査研究所における心理職を知ったのかもしれません[*1]。実際の科捜研における心理職は虚偽検出を行っており，これには (生理) 心理学の知見が生かされています。

他にも高校までの科目として，倫理や現代社会でフロイド，エビングハウス，マズローなどの名前を聞いた人もいるかもしれません。知らなくても今から学ぶので問題ないのですが，それでもフロイドとポリグラフ検査はどう繋がるんだろう，と首を傾げるひともいるでしょう。心理学の知見は他にも，進路相談・職業指導や，街中に溢れるいろいろな工業製品のデザイン，トリックアートから神経生理学，果ては AI(人工知能) などにも関わってきます。

たとえば法学部では法律の勉強をするだろうということはわかりますが，「心理学を学んでいるよ」と誰かに言ってみても「それは何をするの？ 」と聞かれ，「心が読めるのか」とからかわれることもあるかもしれません。道又 (2009) には，次のような一節があります。

[*1] 科捜研には物理，化学，法医，文書，心理の 5 職種があり，文書鑑定 (筆跡鑑定) と心理 (ポリグラフ検査) を一人の技師が兼ねているところもあります。

心理学を「盛岡名物わんこそば」だとしよう。わんこそばはうまいし，おもしろいし，実に素晴らしいものだ。しかし，十分な知識のない人が，わんこそばについて勝手に空想的なイメージ (かわいいちいさなお椀にはいったおいしいおそばなど) を形成して，何の心構えもなく実際のわんこそばを注文したらどうなるだろうか。それはきっとかなり恐ろしい体験に違いない。

皆さんはまるでこの人のように，こんなものが心理学だったのかと驚かれるのではないかと思います。心理学は**名前は知られているが中身を知られていない学問**なのです。心理学が自分のイメージと違うなと思った人，とくにイメージはなかった人，色々いると思いますが，皆さんが今持っているイメージはこの 4 年間ですっかり変わってしまうと思います。少なくとも最初のメッセージとして，ともかく色々な領域に触れて自分のイメージが変わることを恐れないでください，ということを伝えておきたいと思います。(心理) 統計なんて嫌だよ，と思ったひとも，いったん真っ白な気持ちで受講してくれれば嬉しく思います。

1.2　心理学とはどういう学問か

ところで「しんりがく」という言葉はどこで区切るかご存知ですか？「し/んりがく」?「しん/りがく」?「しんり/がく」?

正解は「しん/りがく」です。心の理学，つまり「こころの理(ことわり)」についての学が心理学なのです。さらにいうと，「学」ですから，エンターテインメントや怪しげな術ではなく，知識の体系であるというのが大前提です。そして知識の体系とは，「積み重ねることができるもの」であり「誰にでも共有できるもの」でなければなりません。だからこそ大学で教育できるのです。逆に考えてみるとよくわかります。積み重ねられず，天啓を受けて，あるいは生まれ持った特殊な技能で達成できるものや，誰にも言えない特別な秘伝により限られた人や集団にしか教えられない情報は，学問ではなく宗教や奇跡，マジックや詐欺の領域に属するからです。心理学に神秘的で言葉にできない特殊な何かを求めている人には，残念ですがそれは心理学ではない，とお伝えしなければなりません[*2][*3]。

すでに述べたように，心理学は幅広い領域をカバーしています。それでも大きく分けると 2 つの分類ができるでしょう。下山 (2001) は次の表 1.1 のように分類しています。

表 1.1　下山 (2001) による心理学の分類

アプローチ	実証主義的アプローチ	了解的アプローチ
主な分野	基礎心理学	臨床心理学
手法	自然科学的手法	臨床的手法
対象となる側面	客観的事実	主観的事実
データ収集	実験・調査	臨床実践
データの特徴	量的データ (客観的データ)	質的データ (物語性)
モデルの有効性	理論的な検証可能性	現実状況を理解する際の有効性
対象者との関係	介入しない	積極的な関与 (関与観察)

[*2] 「なんだよー，さっきいろいろな心理学があるって言ったじゃないかよー。本当はこっそり秘伝の術があるんだろう，カウンセリングとか催眠術とかさあ。」と思っている人もいるかもしれませんが，残念ながらそれは心理学どころかそもそも学問じゃないんです。世の中には超心理学だとか，サイコなんとかだとかメンタルなんとかだといって，商売ネタに “心理” というワードを使う輩が後を立たないから，心理学の誤解は無くならないんですよねえ。

[*3] 余談ですが，大学での学問はすべてこうした「秘伝」を否定するものです。もし言語的に伝えられない合理的でないものがあるならば，それは大学で取り扱うものではありません。もし皆さんが学びたいと思っているのに，合理的な理由なく大学教員が教えてあげないというならば，アカデミックハラスメントに該当します。

ここでは心理学のアプローチを，実証主義的なアプローチと了解的なアプローチに二分しています。一般的には基礎と応用といった名称で呼ばれることもありますが，内容的にはこちらの区分の方が良いでしょう。主な分野としてそれぞれ基礎心理学，臨床心理学が挙げられていますが，基礎心理学のなかには知覚，生理，学習，認知といったさらに細かい学術領域があります。

2 つのアプローチが大きく異なるのは，続く研究の方法論的箇所です。実証主義的アプローチは自然科学的手法を模範とし，客観的事実に基づく実験や調査を行い，理論的な検証・予測を目的とします。一方で了解的アプローチの目的は第一に，クライアントの状態が良くなるところにあり，対象に積極的な関与をしますし，理論よりも現実状況を理解する有効性を目的とするのです。しかしいずれにも共通するのは，これが科学 (Science) であり，データに基づく知識の積み重ねであるという点です。

いずれにせよ，心理学者は心を読んだり，操ったりしているわけではありません。心というものがあるとしたら，どういう機序 (＝メカニズム，仕組み，因果関係) があるかを明らかにしようとしているのです。実は，心というものがあるのか，ないのか，それもわかりません。目にも見えず触ることもできないのですから。あなたの隣にいる人だってよくできたロボットかもしれないですよ？ [*4]そしてこの，対象が客観的・物理的に把握できない，社会的実在や主観的経験であることが心理学の難しさなのです。

心理学は「目に見えない対象を研究する」という難しい課題にチャレンジし続けている学問です。それでいて物理学 (物質の理の学) のように，規則性，法則性を探そうとしている学問です。気を付けないと，超魔術やスピリチュアル，宗教のようなものになってしまいます。心理学は，「知識の体系」＝「学」＝科学 (Science) であることを忘れないで欲しいのです。

1.3 近代科学の特徴

さて心理学が科学であるというならば，そもそもその科学とはなんでしょうか。西洋の歴史を考えてみると，そもそもギリシア哲学から学問が発生し，その後は中世に至るまでキリスト教が「世界を説明する教義」として支配的な考え方を提供していました。これが近代科学になるためには，ニュートンやガリレオなどによるルネサンスを待たねばなりません。そこでは一体どういう考え方の転換があったのでしょうか？

近代科学の特徴は，次の 4 つにまとめることができるでしょう。また，それぞれの考え方に関係した人の名前を挙げています。

- 実験主義：ティコ・ブラーエやケプラーによる天文記録，ガリレオのさまざまな実験
- 数量主義（**数学的現象主義**）：ガリレオやニュートンによる記述
- 機械論的発想：デカルトや，マクスウェルのデモンに代表される考え方
- 要素還元主義的：デカルトの「第二の格率」，「分割と統合」といった考え方

まず実験主義である，というのは「教科書に書いてあるから確かめなくてもいいでしょ」という考え方への反発です。「本当かどうか，確かめてみようじゃないか」というスタイルです。西洋では長らく聖書に書いてあることが世界の答えでした。聖書には重たいものが軽いものよりも先に落ちる，と書いてありました。だって神様がそうしてくれたのだから，というわけです。このことを疑うのは，神を疑うことであり，そんな人は地獄に落ちるぞ，と言われたりします。しかしこれを確かめてみよう，と実験したのがガリレオです。ピサの斜塔から重たいものと軽いものを同時に落とし，同時に着地することを確認する。実際にやって確かめなければならない，思弁的に論じているだけではダメだ，というわけです。他にも，ティコ・ブラーエやケプラーといった天文学者は天体の動きを何年にもわたって記録し，どういう法則性があるのかを考えていました。これも実際にデータ

[*4] 「スワンプマン」と呼ばれる思考実験の例があります。興味があれば「スワンプマン」で調べてみてください。

をとって考えてみるという意味では実験主義的な態度であり，これが科学の基本的なスタイルなのです。

数量主義あるいは数学的現象主義は，数学で現象を記述しようという考え方です。ニュートンは重力を発見した，と言われますが，正確には力がどのように働くのかを数式で表現したのであり，重力を見たわけではありませんよね。データをとってその関係性を記述することで，世界の理解を深めようと考えているのです[*5]。大事なポイントは，重力が何か (What) ではなく，どのように働くか (How)，という問いの大転換をしたことであり，これが近代科学の基礎を築いたのです。

機械論的発想というのは，「世界は機械＝メカだ。メカニズム (機序) があるんだ」という考え方です。西洋では長らく，世界は神様が作ったものだから，その仕組みを探ったりしたらバチが当たる，と考えられていました。デカルトは「神様が作ったのは「心」で，肉体とかの「体」はいくらでも分解して調べて，神様の考えを勉強させてもらったらええんやで」と考えました[*6]。このメカニズムを探ろう！というのが科学の目的であるということです。

最後に要素還元主義的であるというのは，「要素」に戻していこう，という考え方です。大きな問題を小さな要素に分解していき，確実にわかることを積み重ねていけば，いつかは大きな問題全体がわかる，という考え方です。これもデカルトの考え方で，「方法序説」という本に書いてあるので興味があれば読んでみてください (ルネ・デカルト, 1637 谷川訳 1997)。大問題を小問題に分割していこう，というのも近代科学的アプローチの基本です。

さて，心理学は科学ですから，ここにあげた 4 つの基本的な姿勢はそのまま継承しています。また，この 4 つの特徴の根本にあるのは**私秘性の排除**にあるということにも注意してください。私秘性とは聞き慣れない言葉かもしれませんが，要するに「私だけが知っている秘密」という性質であり，これを徹底的に排除しようとするのが科学的な姿勢なのです。私，というのは一個人はもちろん，教会といった特定の組織，ひいては全知全能の神様すら含まれるでしょう。説明できないところなくし，徹底的に言語化すること，誰にでもわかるように世界を説明することが，科学の究極的な目的なのです[*7]。

ですから心理学も当然，「実験的」で「数量的」で「機械論的」で「要素還元主義的」なアプローチをとります。問題を細分化しつつ，実際に実験や介入をへて，その機序を明らかにしていくわけです。でもここで，なんで数字を使うんだ！と思う人もいるかもしれませんね。誤解しないでいてほしいのですが，数字は「言葉」です。いわゆる日常的な言葉 (日常言語) ではありませんが，科学的な営みを進めていく上では，次の点で日常言語より優れているのです。

- 正確であること；5 点は誰が見ても 5 点。客観的で正確な記述に向いており，相対的に確実な事実の伝達ができる。日常言語で「すごい」「すごくすごい」と言いながら伝えることと比べてみるとすぐにわかる。
- 比較可能性；5 点は 3 点より大きい。確率的判断により予測が可能。日常言語で「やばい」「えぐい」といった副詞の程度を比較することを考えてみると，これも明らか。
- 要約可能性；代表値等で全体を表現できる。簡潔な表現でデータの特徴や傾向がわかる。日常言語で「要するに・・・」といった表現をすることがあるが，どのように要約しているのかも記述できる。
- モデルの検証が可能；モデルにデータを与えて仮説を検証したり，客観的で合理的な結論を導くことができる。

[*5] 神様の存在を疑ったり反抗したりするのではなく，逆に「こんなにも完璧に自然を設計している神様まじパネエ」という信仰心から研究が重ねられていたようです。詳しくは池内 (2014) や蛇蔵 (2014) などがおすすめです。

[*6] 関西弁は著者の意訳によるものです。

[*7] 科学というのはそういう意味で，本質的にコミュニケーションなのです。

1.4 心理学のいとなみ

上で述べたように，数字 (数学) の言葉を使って現象を表現することは，正確で比較検証可能な表現であるだけでなく，膨大な情報を要約することもできるなど，利点は明らかです。ここまで来てもなお，「いや，数字で表現できないことがある。それこそ心だ」と言いたい人がいるかもしれません。

いいでしょう，確かに心理学の研究対象は「心」ですが，それが目に見えず触ることもできないものだというのは認めます。さらに言えば，あるかどうか，その存在すら疑わしいですね。ばかな，心はあるに決まってるじゃないか，という人もいるかもしれませんが，近代科学の作法に則ってそのことを表現してみてください。すなわち，心があるかどうかは私にはわかっている，というのは不適切な論法である (私秘性の排除が必要) ということを踏まえて，考えてみて欲しいのです。

心理学者はこの問題について，次のように考えてきました。つまり，「あるかないかはわからないけど，仮にあるとしよう」，と仮定を置きます。その上で，もし心というものがあるのなら，外部から何らかの刺激を与えた場合，その心を経由して，何らかの反応があるだろう。たとえば人を拳で殴ります (刺激)。するとその人は泣き出すわけです (反応)。なぜそうなるのか，それはきっと「殴られて悲しい」からだ，それこそが心の働きだ，というわけです。これを図にしたものが図 1.1 です。

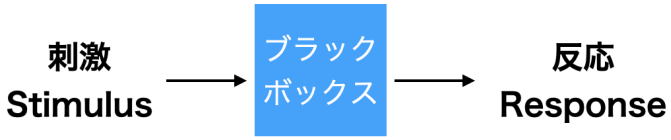


図 1.1 刺激と反応からブラックボックスを探る

心理学では科学的に考えるために，主体 (人間に限らず心があると思われるのであれば何でも) に対してさまざまな刺激を与え，その反応パターンでもって「心があると言えるかも」と考えるのです。これはニュートンが「りんご →?→ 落ちる」の関係から，そこに「重力」があるのかも？ と思ったのと同じ理屈です。彼はこの関係を数式で表現したのでしたね (数学的現象主義)。「こころ」も「こうすれば → 心のせいで → こうなる」という因果を説明する表現方法だといえます。

さきほどの図 1.1 もそうですが具体的な現象を抽象化して表現することで，理解が進んだり新しい解釈ができたりします。この抽象化された，理想的な状態の表現方法を**モデル (model)** と言います。心理学では主体に心というモデルを想定し，もしそのモデルが現実と対応しているのであれば，(理論から導かれた特定の) 刺激に対して，(理論から導かれる結果としての) 反応が現れるはずである，と考えることができます。そこでさまざまな刺激を考え，さまざまな反応を受け止め，心というモデルの正しさを検証しようとしているのです。

もちろん刺激にはいろいろなパターンが考えられます。それこそ無数に考えられてきました。そしてそれに対応する反応もさまざまなものがあるでしょう。1 つの刺激にだって，個人差がありますから，さまざまなバリエーションのある反応が出てくるはずです。こうしたたくさんの刺激 – 反応を正確に記録するために数値化が必要であり，数値化することで要約が可能になり，数式でその関係を表現することで一般化できますし，その数式に実際の数値を代入することで事実に合致しているかどうかを検証できます。このように考えると，心理学で数字を使わないはずがないのです。もちろんそのほかの学問でも，数値化や数字は必須です。物理学のように純粋に数学を扱う領域もあれば，工学のように測定されたデータを扱って物理理論との対応を考える，

いわゆる理系の学問では当然ですが，人文社会科学においても使わないわけはないのです。

統計は，単なる集計に関する学問ではありません。とくに「心理」統計は，心という目に見えないものを見るために，どのように測定するか，どのようにデータを取るか，と言ったところが決定的に重要です。これらは**心理測定法 (psychometrics)** といわれ，心理学のさまざまな領域で考えられています。しかし一度数値化されたデータが手に入れば，それからどのようにメカニズムを解釈するのか，どのようにモデルを立てていくのかについては領域を超えてさまざまな手法があります。それがこの授業，心理学データ解析でみなさんが学ぶべきことなのです。みなさんはこの後，さまざまな専門領域に進んでいきますが，得られたデータを分析するという点については共通した知識・知恵の蓄積がありますので，心理学のごく基礎的な段階から学ぶ必要があります。

数字で心がわからない，と思う人がいるかもしれませんが，これについてはどのように数字にしたのか，という心理測定の問題が理由となっているのではないでしょうか。すなわち，嬉しい気持ち，悲しい気持ちを 4 とか 23 とか言われても，そういうんじゃないんだよな，といいたくなるのはわかりますが，それは嬉しい $=4$ だと誰がどうやって決めたんだ，という問題なのです。仮に数字が正しく心の状態を表現していたとすると，その後その大量の数値から要約したり個人差を除外して，真なる状態を求めることができるかもしれません。近年はあらゆるところに人の行動が記録される機会があります。こうした記録は膨大な量に登るので，ビッグデータと呼ばれています。ビッグデータの中から，人の行動パターンを分析するのも心理統計の範疇です。犯罪プロファイリングやマーケティングなども心理統計の応用的実践例です。もちろん臨床的介入の効果を検証するとか，実験的操作の影響力を評価することも心理統計の守備範囲です。数値化することによって，目に見えなくて触れなくてあるのかないのかもわからないものが，どの程度わからないのか，どこまでならわかるのかといったことも表現できるようになるのです。

心理学と統計法の関係，心理統計の必要性についてご理解いただけましたでしょうか。納得をいただいた上で，心理統計の具体的な話に進んでいきたいと思います。

1.5　課題

■心理学とは何だろうか　みなさんの中で心理学とはどういうイメージを持っていますか。この講義を受けるまえはどのように考えていましたか。この講義を受けた後で何か変わりましたか。できれば今考えている，あなたなりの心理学像を，どこかに書き起こしておくと良いでしょう。前期の終わり，あるいは一年の終わり，そして四年後にそれがどう変わるかを楽しみに。

■心理学にどうして統計が必要なのか　第 1 回は心理統計の中身に入る前に，どうして心理学なのに数字を扱って統計を学ぶのかについてを解説しました。みなさんの身近な人に，みなさんなりの言葉でこのことを説明してみてください。そのときに，心理統計と心理測定の違い，科学としての心理学，というキーワードを入れながら説明できれば完璧です[*8]。

[*8] また後ほど触れることになりますが，ある人が何かを理解している，わかっているということは，そのことについてさまざまな表現を延々と生み出すことができる，ということでもあります。もし言葉に詰まったり，うまく説明できないことがあれば，そこの理解ができていないと考えられます。このことは第 6 講で触れる古典的テスト理論とも関係しますし，心理学の基本モデルである図 1.1 とも直結する観点です。

第 2 章

数値データの種類

さて前回は，科学としての心理学なので，数量的なデータを扱って研究が進められる，という話をしました。今回は，何をどのようにデータにするか，データの種類と言った点についてお話しします。

2.1 データと変数

心理学では目に見えないものを研究対象とするため，心を持つと考えられる主体にたいして刺激を与え，その反応を見て「こうすれば (刺激)→ 心のせいで → こうなる (反応)」という因果を説明するためのモデルとして研究するのだ，ということでした。ここで刺激と反応は，研究対象である「心」の外にあって，観測できるものです。これをデータにします。

図 2.1 は心理学研究の流れを説明したものです。ここで述べている「刺激と反応からデータを作る」というところは**心理測定法**と呼ばれます。何をどのようにデータにするか，ということを考えるところであり，それぞれの心理学領域によってさまざまなバリエーションがあります。神経生理のような領域での研究であれば，データは計測機から得られる電気反応の強さ，間隔などになるでしょうし，認知・知覚・学習心理学のような領域であれば，たとえば被験者のキー反応の早さなどを PC で計測してデータとするでしょう。社会・学習・臨床心理学のような応用的領域であれば，被験者の発話，調査票への回答パターンなどがデータになるでしょう。いずれの領域でデータを取るにしても，ここで重要なのはそのデータの**信頼性 (reliability)** と**妥当性 (validity)** です[*1]。

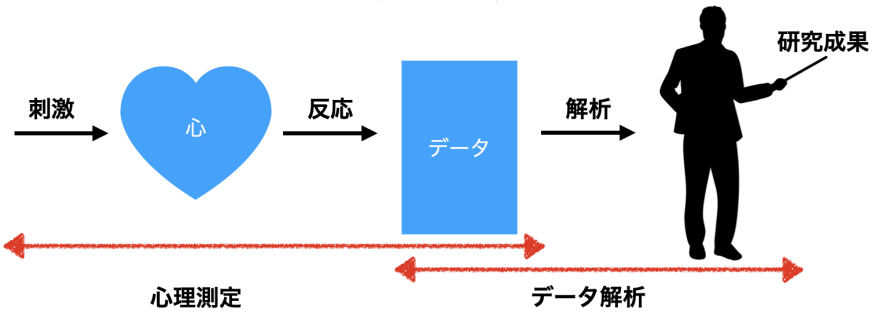
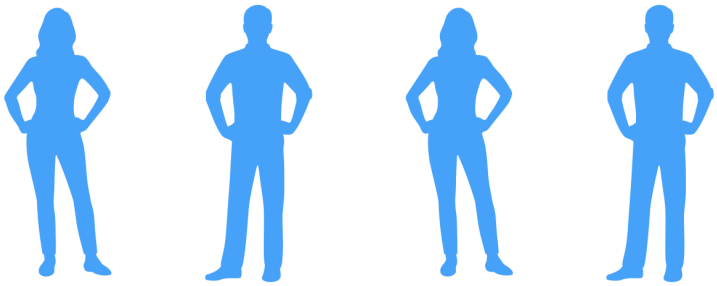


図 2.1　心理学研究のプロセス

それに対して，この講義でとくに重点的に議論するのは後半のデータ解析のところになります。つまりデータが得られた後で，それをどのように分析していくか，という点です。これについては次回も言及しますが[*2]，

[*1] 詳しくは第 6 講で触れます。
[*2] → セクション 3.5,Pp.55

客観性 (objectivity) や**再現性 (reproducibility)** が重要になってきます。同じデータであれば誰でも同じ結果に到達できる，ということが重要ですし，言い方を変えると数学的に正誤の判断がついてしまう領域でもありますので，正しく理解し利用できるように学んでいきましょう。



名前	Aさん	Bさん	Cさん	Dさん
性別	female	male	female	male
所持金	4200円	35,200円	7213円	15円

図 2.2　データって何?

さてこれからデータを分析していく話をするのですが，改めてデータとは何でしょうか。図 2.2 を見てください。ここには 4 人の研究対象者がいて，色々な情報を集めてみました。この中でデータとはいったい何を表すでしょうか。所持金は数字ですので，これだけがデータだ，と思ったら大間違いです。実は名前や性別もデータです。これらはすべて数値化し，データとして扱えるものです。

少し用語の説明をしておきますと，名前，性別，所持金などは人によって[*3]変わる数字です。人によって変わる数字のことを**変数 (variables)** と言います。名前・性別・所持金は変数名であり，それがケースごとに「A さん」「男性 male」「15 円」と変わるのです。ここで 15 円はともかく，名前や性別は数字じゃないぞ？ と思われるかもしれません。先にも述べたように，数字じゃないような実体であっても，これに数字を割り振るルールを決めて数字に置き換えていきます。このルールを決めることを**尺度化 (scaling)** と言います。

図 2.3 に数値化した例を挙げました。上の段にあるのが本来の情報であり，数値化されたのが下の段にあります。下の段はすべて数字になっていますね。どこがどう変わっているのか，よく見ておきましょう。実際の統計的な分析をするときは電子計算機を使いますが，電子計算機にデータを与えるときは下の段にあるように数字だけにするのだ，と理解してください[*4]。

こうしてみていても，数字には 2 種類あることがわかります。まず所持金を見てください。これは最初から数字でしたので，そのまま数字として扱っていきます[*5]。このように本来の数字としての意味を持っているものは**量的変数 (numeric variable)** と呼ばれます。一方で，性別は「女性 female」を 1 に，「男性 male」を 2 に置き換えています。名前に至っては，「A さん」を 1 に，「B さん」を 2 に，「C さん」を 3 に，「D さん」を 4 に置き換えています。これらの数字は人や性質を分類するためのものであって，数字に大小関係など量的な違いを表すものではありません。いわば質的な違いを表すための数字で，こうした変数は**質的変数 (categorical variable)** と呼ばれます。

[*3] 心理学でも研究対象が人でない場合がありますから，データの単位は他にも「個体」「ケース」などということがあります。英語の Observation を略して Obs. と書かれることもあります。

[*4] Excel や Numbers などの表計算ソフトでは，その枠の中に日本語を書き込むことができます。しかし統計ソフトで扱う場合は，機械が内部では下段のような数字に置き換えているのだと思っておいてください。電子計算機の中では日本語も 0/1 の数列にすぎませんが。

[*5] 単位の「円」という文字は取れています。表計算ソフトに入力するときはこのように，数字以外のものは削ぎ落とすようにしてください。文字が入っていると数字として計算できないからです。ましてや 2 バイト文字なんて!

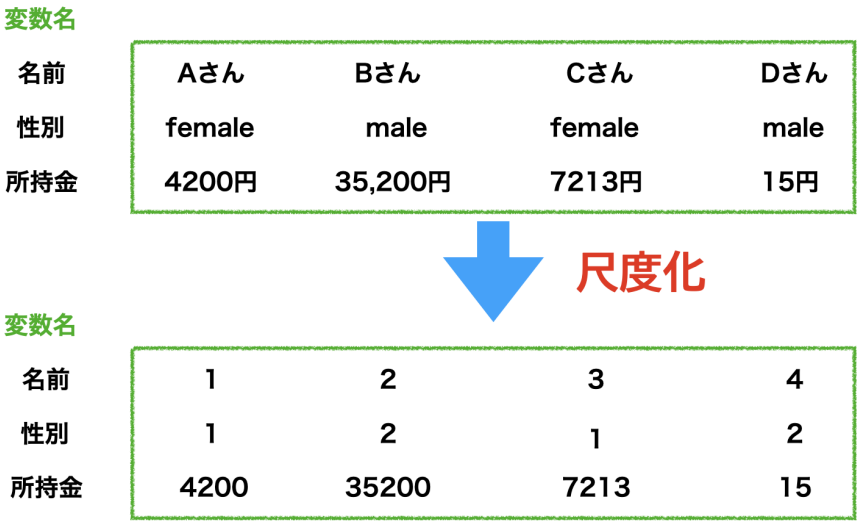


図 2.3 尺度化＝数値に置き換える

このことからわかるように,「なんでも数字にして扱うが, 数字には種類がある」ということに注意が必要です。なんでも数字にする, というと眉をひそめる人がいるかもしれません。心・気持ち・想いは数字にできない！と直観的に反論したくなる人もいるでしょう。しかしまったくできないわけではないのです。というのも, 言葉にできるのであればそれに適当な数字を割り振ることは可能だからです。数字が必ずしも大小関係を表すものばかりではないこともあるのですから。言葉にできるということは, 質的に異なって分類できるということであり, それに数字のラベルを貼ることは可能です。一方, 言葉にできない気持ちというのもあるでしょう。しかし, 適切に言葉にできないのであれば, 科学的な研究対象にはならないと言わざるを得ません。なぜなら「私だけが知りうる本当の気持ち」というのは, 私秘性を徹底的に排除するという科学の基本姿勢に反するからです[*6][*7]。

心理学の研究法の 1 つに, 質問紙法というものがあります。複数の項目が並んでいて, それぞれに対して「当てはまる」「どちらかと言えば当てはまる」「どちらとも言えない」「どちらかと言えば当てはまらない」「まったく当てはまらない」のどこかに丸をつけてください, という聞き方をするものが一般的です。たとえば「Q. この世の中では, 努力がいつか報われるようになっていると思う」という問いに対して「自分の考え（気持ち, 意見, 態度）に最も近いところに丸をしてください」と言われ, ある人は「当てはまる」に丸を, 別の人は「どちらかと言えば当てはまらない」に丸をつけたとしましょう。ここで「当てはまる」を 5 点,「どちらかと言えば当てはまらない」を 3 点, といった形で数値化すれば, 大小関係は比較できるので量的変数として扱うことができます。この時ここで測りたいのは,「世界は公平だと思うか」というその人の心理的態度 (信念) を測定していると考えられます。

このように回答のパターンを見ることで, 目に見えない心の状態を数値化することもできます。見たいものは目に見えない心の状態ですが, 測定して, 点数をつけたりします。他にも, 行動や振る舞いの記録など, 知

[*6] 言葉に言えないなんとも表現できない気持ちがない, と言ってるわけではないのです。科学的に研究できない, というだけで, 私個人も色々な心の動きを経験しています。ただそれは私一人の思い出であって, 研究対象にならないと言っているだけです。

[*7] 言葉にして数値化するというのは, 心の一側面を捉えているにすぎない, という指摘も当然ありましょう。しかしだからと言って数字を使わない, というのはもったいないのです。また数字で表すことによって, どこまで表現できてどこまで表現できないか, というのも表現できたりします (正確には誤差の大きさを見積もる, という話で本講義の第 6 講ごろから触れていきます)。表現できるところはし尽くし, できないところについては黙って味わえば良いのであって, 表現することを禁じるのは建設的な態度ではありません。

りたいことについて集められる情報はすべてデータであり，分析対象になります。注意してほしいのは，抽象的で目に見えないものだからなんでも好き勝手に「○○力」と名付ければいいというものではない，というところです。心理学者は空想世界に住んでいるわけではありません。現実の世界に起こる見て触れることのできるものがあり，それに基づく現象 (phenomenon) があります。この現象は抽象的なもので直接その対象が見えるわけではありませんが，現象として観測される何かがあるわけです。たとえば「流行現象」というのはモノやサービスが多くの人に利用される現象ですが，見えるのはそのモノやサービスの利用者でしかありません。ただしそれが多く観測されて，この状態を「流行」と呼ぼう，というわけです。このように抽象化されたものを**構成概念 (theoretical construct)** と言います。構成概念は現実や現象に立脚したものですが，構成概念に立脚した別の構成概念というのも考えられます。これは概念同士の関係から論理的に導出されるもので，その概念を設定する意義 (その概念がないと説明できない，あるいはその概念があることで他の概念や理論との共通点が見られるなど) がなければなりません。

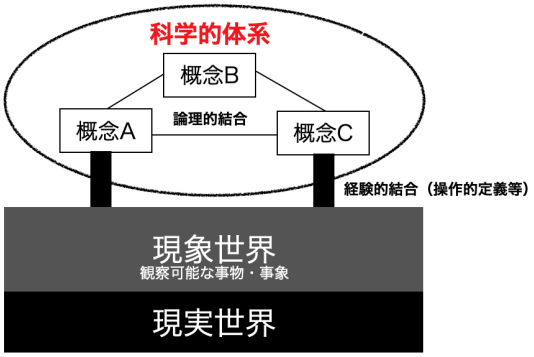


図 2.4　構成概念と現実との対応 (末永 (1987) を元に作図)

図 2.4 にあるように，概念は現実や現象に深く根ざした (経験的結合をした) ものであるか，概念同士の関係に強く結び付けられた論理的結合をしたものである必要があります。この抽象的で目に見えないものに対して，筋の通った論理的な説明体系を私たちは科学と呼ぶのです。

概念を操作して記録・数値化したものが変数となる，とも言えるでしょう。心理学的な研究対象について，何をどう変数化するか，その結びつきをどう考えるかを知るためには，先行研究を広く調べたり現実世界・現象世界についての知識と教養が必要なのはいうまでもありません。

2.2　変数の分類

変数は，文脈や用途によって色々な分類がなされます。先ほど量的変数，質的変数という分類をしましたが，これは数値データの特徴による分類でした。変数は，他にもその役割によって区別されることがあります。

2.2.1　原因変数と結果変数

たとえば次のような研究仮説を考えてみてください。

1. 家族療法を一定期間受けた拒食症の若年女性の，治療前と治療後の体重を記録した。家族療法が有効な治療法であったと言えるか?
2. ラットのある群には自由に食事を与えられ，もう一方には非常に低カロリーな生活をさせた。食事スタイルは寿命にどのように影響するか?

さてこれらのケースで，変数に当たるのはどこになるでしょうか。分析するときに必要な情報の単位が変数になりますので，注意深く読んで考えてみてください。

最初の例では，「治療前か治療後か」という違いが分析上必要ですので，この測定のタイミングが 1 つの変数になります。もう 1 つはもちろん体重ですね。この体重が変わったかどうかが知りたいのですから。2 つ目の例では，寿命の長さ (何日生きるか) が変数だというのはすぐに気づくと思いますが，もう 1 つ，「ある群」と「もう一方」，つまり「自由に食事ができるか，低カロリーなものしか食べられないか」という違いも変数なのです。これらの情報をデータとして記録しておかないと，体重や寿命の長さの数字だけ見せられても意味がないですね。

ここでの例はいずれも，複数の変数間関係を「原因」と「結果」に分けて検証するというものでした。

1. 介入の前と後 (原因)→ 体重 (結果)
2. 食事カロリーの違い (原因)→ 寿命 (結果)

「介入前のデータなのか，後のデータなのか」も “変化する” 指標なので，変数として扱う，というのがポイントです。この例にあるように研究仮説の観点から，原因となる変数と結果となる変数，言い換えれば説明する変数と説明される変数という，違いで分類することが少なくありません。

変数は数字で表されます。たとえば介入前を 0，介入後を 1 というように，です。介入の前後を表す変数を x，結果としての体重を y と表すと，この研究は $y = f(x)$，つまり関数関係を明らかにすることであるとも言えます[*8]。心理学が科学であること，科学であるということは数学的現象主義をとること，数値化によってモデルを検証できること，がこのように繋がっているのです。もちろん関数の形がどんなものなのかは，現象によってさまざまなものが考えられるでしょう。

ただし，データを使って一方の変数から他方の変数を説明・予測する，というのは，研究者の「仮説」に過ぎないことに注意してください。本当のことがわからないので，一方を原因，他方を結果とおいて考えようというだけで，本当の因果の向きは逆かもしれないのです。その仮説が正しいかどうかは，予測が当たるかどうかで評価することが基本です。もっとも，人間の行動は複雑なので (あるいは測定が非常に難しいので)，なかなかうまくいくものではありませんが。

また因果関係は「原因」になるのが先にあり，「結果」が後に来るはずです。研究の中には，同時に取ることで得られるデータもあります。同時に取られた複数の変数間関係からは，変数同士の関係の強さを数値化できます (第 8 講で詳しく説明しますが，相関係数と呼ばれます)。ただし，相関関係＝因果関係ではないことに注意が必要です。相関関係は相互に関係があることを示すだけで，どちらが原因でどちらが結果なのかは分かりません。因果関係は時間的な位置関係は明確です。それに加えて，原因変数の背後にさらなる原因変数があるかもしれませんし，原因変数と結果変数の間に共通する媒介変数があるかもしれません。確認されていないだけで，逆の因果関係があるかもしれません。このように，データや数字から表現できることは色々ありますが，それを解釈するのはあくまでも人間側の問題なのです。まったく中立なデータやモデルというのはなく，常になんらかの仮定や前提，主張を含んだ表現であることを忘れないでください。

さて，この区分に沿った変数の名前ですが，原因としたい変数のことは**予測子 (predictor)**(予測変数ともいう) とか，**説明変数 (explanatory variable)** と呼びます。やや古いテキストには，**独立変数 (independent variable)** という名称がついているかもしれません。これはその変数が何かに依存することなく，独立しているというところから来ています。しかし現実的には，(心理的な) 変数は常に何らかと関係があると考えた方が良いでしょうから[*9]，この表現は誤解を招くものかもしれません。

[*8] $y = f(x)$ の f は function の f で，function は関数とも機能とも訳されます。
[*9] 全く独立している！というのはとても強い仮定です。

また，結果としたい変数は**被予測変数 (predicted variable)**，**被説明変数 (explained variable)** などと言います。**応答変数 (response variable)** と**目的変数 (target variable)** と呼ぶこともあります。こちらの変数が研究のターゲットとなる変数であり，研究を計画するときはまずこちらを慎重に見定める必要があります。そうそう，古典的な名称としては**従属変数 (dependent variables)** というのもありました。これは独立変数に対応して，従っているということを強調した言い方ですね。用語が多く出てきたので，簡単な表 2.1 にまとめておきます。いずれの用語も「説明 (予測) する側」「説明 (予測) される側」という役割で分類していますので，この違いが研究計画から読み取れるようになれば OK です。

表 2.1 変数の分類

役割	名称	数学記号[*10]
説明する	予測子 予測変数 説明変数 独立変数	$X_1, X_2, \cdots$
説明される	被予測変数 被説明変数 目的変数 応答変数 従属変数	Y

2.2.2 潜在変数と観測変数

変数の分類ということで，もう一言だけ。

先ほど，個々人の考えや気持ちの強弱をその当てはまりの程度に置き換え，言葉の強弱から大小関係をもった数字で表す方法について説明しました (図 2.5)。言葉を点数に変える数字のルールを決めることを尺度化というのでしたね。どうして「どちらとも言えない」が 3 になるのか，ということについてはしっかりした理論が背後にありますがそれはおいおい説明していくことにして，話を先に進めます。

心理学の場合，ある心の状態 (たとえば抑うつ状態にあるとか，性格特性がどれぐらいあるとか) を数値化するのに，複数の質問項目の合計得点を利用することがあります。このように，目に見えているデータを加工して，目に見えない変数の代わりにすることがあります。たとえばテストの正答数から合計得点を計算し，成績をつけるということもあるでしょう。これは正答率が学力を反映しているという前提の下での尺度化です。学力も心と同じで，目に見えない想定される心的状態のひとつですね。他にも，「幸せの国」としてブータンという南アジアの国がありますが，この「幸せ」は 2 年ごとに 1 万人ほどの人を対象に行われる調査で，持続可能な社会経済開発 (Promotion of sustainable development)，環境保護 (Conservation of the natural

[*10] 説明する変数 X には $1, 2\ldots$ がついて説明される変数 Y にはつかないのか，という声が聞こえてきそうです。これはそういうものなのです。というのも，説明する変数は複数あることが容易に想像できますし，現にそうした状況に対応した分析モデルがありますので，説明する側には添え字をつけて区別できるのです。説明される変数が複数あることもありますが，これは研究のターゲットとなる変数なので，少数 (基本はひとつ) に絞ることが一般的です。どうしても複数の説明変数がある，という場合はそれらを統合した潜在変数をつくり，それを被予測変数とした関数を作ります。この方法は**多変量分散分析 (Multivariate Analysis Of VAriance:MANOVA)** とか，**構造方程式モデリング (Structural Equation Modeling:SEM)** と呼ばれる**多変量解析 (Multivariate Analysis)** の領域で，姉妹講義のデータ解析応用で取り扱います。もっと言えば，被説明変数も説明変数も複数あるという状況は，何がどうなってるか問題意識が明確になっていない状況かと思いますので，明確な説明<->被説明という関係を置かずに探索的に変数間関係を分析する方法をとることが多いでしょう。

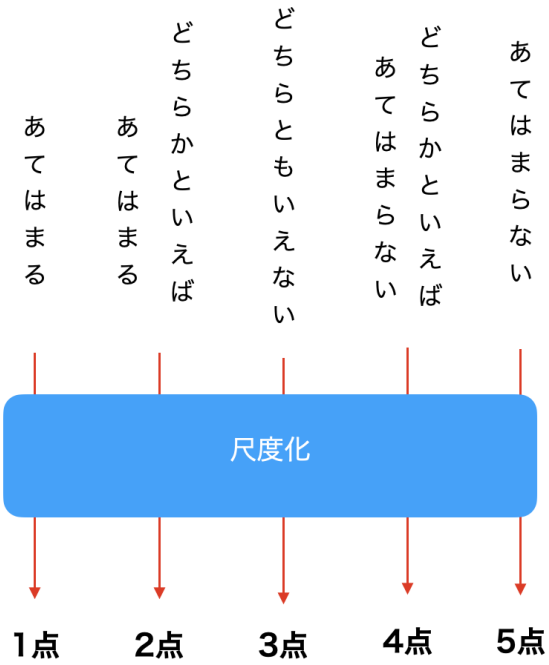


図 2.5 尺度化による数字への置き換え

environment), 伝統文化の振興 (Preservation and promotion of culturalvalues), 良好なガバナンス (Establishment of good governance) という 4 つの側面から聞き取られた数値を加工して算出される「幸福度」を指標にしています。この幸福度も目に見えないものを数値化しているので, 潜在変数です。

このように心理学では, 人間の振る舞いの背後に心というモデルを仮定し, それをデータという目に見える数値をつかって表現するものです。目に見えるように尺度化されたデータのことを**観測変数 (observed variables)** とか**アウトカム変数 (outcome variables)** と言います。これに対して, 観測変数から算出してモデルの状態を表現した変数のことを**潜在変数 (latent variables)** と言います。心理学では測定を数理的にモデル化し, モデル上想定される変数をモデルの中で加工したりすることがあります。潜在変数は具体的な数字に顕現化していませんが, 抽象的なモデルのままで計算を進めたりできるのです。具体的には測定モデルについての説明が必要になりますが, ここでは観測変数だけでないものも扱えるのだ, ということを知っておいていただければ十分です。

2.3 尺度水準による分類

変数にはさまざまな名称があることを見てきましたが, 今度は尺度化された数値に着目してみましょう。数字になってしまえば全部同じか, というと実はそうではありません。Stevens (1946) は数値が持つ情報とそれに伴う計算可能性の観点から, 尺度を 4 つの水準に分類しています。それぞれ次のような名前がついています。

1. **名義尺度水準 (nominal scale)**
2. **順序尺度水準 (ordinal scale)**
3. **間隔尺度水準 (interval scale)**
4. **比率尺度水準 (ratio scale)**

名義尺度水準が最も下のレベルで，比率尺度水準が最も上のレベルという階層性を持っています。上に上がるほど徐々に条件が強くなっていきます。逆に言えば，上の尺度水準に該当する変数をより下位の水準の変数として扱うことが可能です。このことについて詳しくは後ほど説明するとして，それぞれがどういう意味なのかを説明していきましょう。

■名義尺度水準　最も下位のレベルにある水準です。この水準にある数値は，数字ではありますが数字としての意味がないレベルです。単にラベルとして対象に数字を割り当てた，というだけのとき，その数字は**名義尺度水準**にある，と言います。この講のセクション 2.2,Pp.24 で，性別や名前に数字を割り当てたのを思い出してください。たとえば女性に 1 を，男性に 2 を割り当てたとします。この割り当てられた数字に，深い意味はないですね。別に男性が 2 でもいいし，女性に 43543 という数字を当てておいても良いわけです。数字だけ見ても，女性が男性の 2 倍あるとか，男性より少ない･小さいことを意味しているわけではありません。あるいはみなさんが持っている携帯電話の番号もこの尺度水準です。みなさんの数字はどれひとつとして同じではないですが，その数字に深い意味はないですね[*11]。他にも学籍番号や，名前に割り当てられた数字 (A さんが 1,B さんが 2 など) に数字としての特徴はなく，ただの名前を簡略化したものに過ぎません。こういう数字のことを名義尺度水準にある数字，というわけです。

この水準の数字は，数字が大小関係や順序関係を表さないので，計算ができません。たとえば女性を 1，男性を 2 と尺度化したデータを集めて平均値を出したら 1.7 でした，と言っても意味がないのです。1.7 に該当する対象がないのですから。この尺度水準では**数字と対象が一対一対応していること**が重要なのであって，数字を計算して変化させても意味がありません。ではこの尺度水準の数字はなんの役にも立たないのか，と言われるとそんなことはありません。違う数字が割り振られていることから違う対象であることは分かりますので，分類するという性質は持っていますし，何よりどんな複雑な対象であっても名義尺度水準の数値は割り当てられるのです。その意味では，この後に出てくるどんな尺度水準よりも豊富な情報を含んでいる数値である，とも言えます。また，言葉にできなければ研究対象にできないという話をしましたが，言い換えると言葉にできるのであれば名義尺度水準の数値が割り当てられるということでもあります。

■順序尺度水準　名義尺度水準は数字としての意味を持たない，と言いました。これに順序関係，すなわち数字の大小に意味があるという条件が加われば，それは**順序尺度水準**の数字だということができます。たとえば好きな夕飯のメニューとして 1 番がカレー，2 番がハンバーグだと答えた人がいたら，カレーの方がハンバーグよりも好まれている，価値が高い，優先されているという情報は得られるわけです。あるいは 100m 走での順位を表す数字だと，2 と 7 を比べれば 2 位の人の方が速いんだな，ということがわかります。この程度の情報が含まれている数字のことを，順序尺度水準であると言います。ただこの水準の数字は大小関係を表すだけで，それ以外のことはわかりません。たとえば着順の話でも，1，2，3 の順で優劣がついていることはわかるのですが，「1 位がぶっちぎりだった」「団子レースだった」といった違いは表現できていないわけです。図 2.6 を見てもらうとわかるように，データとしては 1,2,3 という数字だけですから，上図のような状態でも下図のような状態でもデータとしては 1,2,3,4 であり，上図の 2 と下図の 2 を区別できません。まだ少し不便ですね。このように大小しかわからないので，計算にもまだ使えません。1 位と 2 位を足して 3 位になるはずもないですし (そもそも逆転してますが)，2 位と 3 位を足して 1 位と勝負できるかということもわからないのです。もちろんかけたり割ったりもできませんので，この水準の数字は四則演算ができません。

とはいえ，この水準の数字は，心理測定において最もおもしろいところであると言えるかもしれません。みなさんご自身の心の状態を数字で表すとして，さてその数字はどの水準にあるのでしょうか。たとえば「カレーはハンバーグの 100 倍美味しい！」と思っていても，本当に 100 倍という計算ができるような精度の高い数字

[*11] 語呂合わせがしやすいかどうか，というのはありますが。

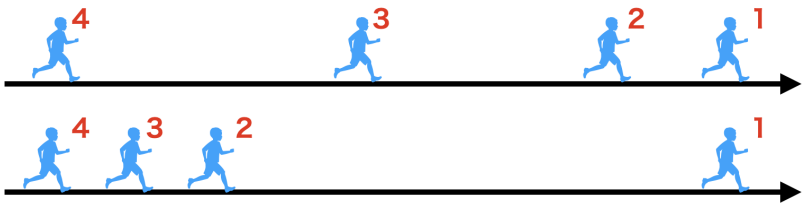


図 2.6 順序では程度がわからない

でしょうか？ せいぜい優先順位の違いぐらいで，カレーを 5 点，ハンバーグを 4 点，餃子を 3 点･･･と採点していっても，カレーとハンバーグの違いが，ハンバーグと餃子との差と同程度であるとまでは言えないのではないでしょうか。ことほど左様に，心理学的な評価水準はせいぜい順序尺度水準であって，差や比の計算をするには，さらに上の尺度水準でなければなりません[*12]。

■間隔尺度水準　間隔尺度水準は，順序尺度水準に加えて数値と数値の間隔が等しいという条件を追加したものです。順序尺度水準は 1 と 2 の差が 2 と 3 の差と違っても構わないのですが，間隔尺度水準ではこれがあっている必要があります。代表的な例として，摂氏 (温度) があげられます。これは水が氷になる氷点を 0，水が水蒸気になる沸点を 100 とし，その間隔を百等分した目盛りによって温度を表現するものです[*13]。このようにそもそも等間隔になるようにつけた目盛りですから，当然間隔尺度水準の数値であると言えます。この時，10 ℃と 20 ℃の差は 20 ℃と 30 ℃の差と等しいといえますから，これは $20-10=30-20$ と引き算の計算ができるレベルになっているわけです。

しかし，10 ℃は 20 ℃の 2 倍の熱量を持っている，とは言えません。というのも，この数字は 0 ℃に意味がないからです。言い換えると，0 ℃は温度が「ない」ことを意味するのではなく，0 ℃という温度が「ある」状態です。0 がスタート地点ではないのです。スタート地点は任意に決められますから，たとえば氷点を 100，沸点を 200 とした新しい温度体系 X を作ったとします。このとき 10 ℃は 110X で，20 ℃は 120X ですから，$20/10 \neq 120/110$ ですね。同じ状態を表している数字なのに計算が合わないのはおかしいわけです。このような割り算の計算をするためには，さらにもうひとつ条件を追加する必要があります。

■比率尺度水準　比率尺度水準は，間隔尺度水準に加えて絶対 0 点を定めたものです。絶対 0 点とは，0 が「ない」ことを意味する点のことです。たとえば 0cm は長さがないのです。0kg は重さがないし，0 秒は時間が経ってない，です。このように原点 0 が決まっている数値体系になってはじめて，加減乗除の四則演算すべてに対応できます。例に挙げたように，物理的な測定単位はほとんど比率尺度水準ですから，精緻に測定すればかなり細かい予測値まで計算できるのですね。これに対して心理学のデータで主観的につけられる評定値はせいぜいが順序尺度水準なので，物理学のようになかなかうまくいかないところもあるというのが実際です。心理学のデータでも客観的で比率尺度水準の数字を扱えるものもありますし (反応時間を指標とする，など)，主観的な評定値でも工夫して間隔尺度水準以上の数値にすることもあります。

みなさんが今後，どのようなデータをとって分析するようになるかは進む専門領域によって変わると思いますが，どの数字がどの尺度水準にあるのかについて，常に注意しておいてください。というのも，名義尺度水準や順序尺度水準では加減乗除の計算ができないので，分析の方法にも限界があるからです。これら尺度水準の違いは，統計ソフトウェアで分析するときもどの尺度水準にあるのか指定する必要があります。指定が

[*12] しかし実際には，評定値を間隔尺度水準の値を持っているとみなして分析をすることが多くあります。これには**心理測定法**の理論と歴史的な経緯があって，その両方を踏まえた上でことであればそのような「みなし」も正当化されます。残念なことに，それらを踏まえずして仮定が満たされていないのに勝手に間隔尺度水準だと思って使っている誤用例も数多くあるのですが。

[*13] セルシウスさんが考えた方法なのでセ氏といいます。ちなみに定義上このようになっているのであって，「水ってちょうど 0 度で氷になって，ちょうど 100 度で水蒸気になるの，奇跡じゃない?」というのは悪いジョークです。

なくすべて計算可能な数値であると機械が誤認すると，頓珍漢な分析をしてしまうことになります。表 2.2 に Stevens (1946) によるまとめを引用しましたので確認しておいてください[*14]。

表 2.2 尺度水準のまとめ

尺度水準	基本的な操作	数学的な群構造	可能な統計量
名義 Nominal	同質性の判定	置換群 $x' = f(x)$ $f(x)$ は一対一対応	ケース数, 最頻値, 一致係数
順序 Ordinal	大小関係の判定	同調群 $x' = f(x)$ $f(x)$ は単調増加関数	中央値, パーセンタイル
間隔 Interval	間隔や差分の等しさの判定	一般線形群 $x' = ax + b$	平均値 , 標準偏差, 順序相関, 積率相関
比率 Ratio	比率の等しさの判定	類似性群 $x' = ax$	変動係数

統計処理の観点からは，間隔尺度水準と比率尺度水準はいずれも計算できるので量的データ，名義尺度水準と順序尺度水準は (計算できないので) 質的なデータと 2 分割で済ませることもあります。他にも正の整数しか取らない離散的なデータをカウントデータと呼ぶこともあります。また，0 と 1 の 2 つの値しか取らないデータを二値 binary データと呼ぶこともあります。二値データは，0 と 1 がそれぞれ男性と女性，正と誤，真と偽といったようにさまざまな解釈ができますし，名義尺度水準でありながら数学的には量的データのように計算できることもあって，少し特別な位置付けにあるといえるでしょう。

2.4 データの可視化

さて，ここまで変数の分類や数値の分類について語ってきました。こうした違いに注意しながら，これから分析を進めていくことになります。分析の第一歩は**可視化 (visualization)** です。**データは図にする**のが基本中の基本です。みなさんもこれから演習や研究の一環でさまざまなデータを取ると思いますが，**データは図にする**のが基本です。これを忘れないでください。大事なことなので 2 度いいました。また，これから分析の話が 28 講分ありますが，**データは図にする**ということの重要性を強調するのは今回限りになります。しかし本当はこの重要性を常々言い続けていたいぐらいなのです。分析が進み，さまざまな知識や技術をみなさんに伝えていく中で，相対的に**データは図にする**ということの重要性が下がっていくように感じられるかもしれませんが，それは間違いです。取ったデータは最初に必ず図にしてください。さまざまな角度から図にしてください。なんならまだ取っていないデータであっても，このような図が得られるはずだとか，このような図を得るためのデータが欲しいと考えるようになってください。可視化することでわかりやすくなります。気づかなかったことが目で見てわかるようなヒントが得られます。一枚の図からさまざまな情報を読み取ることができるようになります。それほど可視化は重要なのです。

[*14] Stevens (1946) の仕事をより正確に言うなら「測定対象全体に数値を与える適正な規則とは何か」を考えたのではなく，何らかの測定がなされた尺度がいくつかあれば，それらを相互に変換する規則が定められて，それが表にあるような 4 種類になるだろう，と言う考え方です。詳しくは吉野 (2007) を参照してください。

逆に言えば，可視化しなければ数字だけの羅列になります。数字の羅列が好きだという人は少ないと思いますが，そういう人はかえって数字を過大評価し，嫌いな数字を見つめているのが嫌だからでしょうか？ 1 つの数字だけ見て大体わかったような気になってしまうのです。それではデータの本質がわかりません。データサイエンスの領域は今後ますますニーズが高まる領域ですが，優秀なデータサイエンティストは手に入ったデータを可視化して眺める時間をたっぷりとるものです[*15]。とりあえず (決まり切った) 分析をして指標を算出するというのは，時間の無駄になることも少なくないので，このことをしっかり意識しておいて欲しいと思います。

得られたデータは基本的に可視化しなければならず，そこから分析は始まっていると言えます。またそこにおいても再現性が重要です。ボタンをぽちぽち適当に押していたら出来上がった図，というのは再現性に乏しくチェックすることもできないので，図を書くためにも図を書くルールが必要です。表計算ソフトウェア[*16]は簡単に図を作ってくれますが，GUI 操作になりますので，統計環境の R やコンピュータ言語の Python などでは CUI，すなわちコマンド (スクリプト，プログラム) で図を書きます[*17]。図を書くのもプログラミングなのです。

この講義では統計環境 R を使っていきますが，R には図を書くためのパッケージとして `ggplot2`(Wickham, 2016) というものがあります[*18]。可視化の文法については，心理学業界ではまだ広く浸透した概念とは言い難いのですが，見た目の美しさのみならず，論理的な美しさや再現可能性の観点から考えて極めて重要ですのでここで概略だけ説明しておきます。，まずはどのようなグラフが描けるのか，いくつか例をあげてみましょう。

図 2.7 には複数のプロット例を挙げました[*19]。一番上の段，左側にあるのは散布図です。X 軸 Y 軸に変数があり，データがどのように散らばっているかを目で見ることができます。点を打つときは群ごとに色を変えてみました。最上段右図はボックスプロットと呼ばれる図です。4 つの群の平均値とデータの散らばりが図示されています。真ん中の段の左図はボックスプロットと同じ情報ですが，ボックスプロットとは違う表現をしたものです。すなわちデータの散らばりを左右対称のヒストグラムのように描いたもので，わかりやすくするためにデータ点もその上にプロットしてみました。真ん中の段の右図はバープロット，棒グラフと呼ばれるもので，4 つの群の平均値を高さで表しています。一番下の段にあるのはヒストグラムで，どの階級に何件のデータが入っているかを高さで表たものになっています。

図 2.8 に示したのは散布図ですが，群ごとに分けて表示してみました。またプロットするときの模様も，別のラベルごとに変えてみました。

図 2.9 に示したのは折れ線グラフです。これは私の体重変化の記録で，横軸が日付で縦軸が体重になっています。横軸に連続的な意味がある場合は各データ点がつながるように表現することで変化を表すことができます。

ここではさまざまな図の表現の仕方がある，ということをわかってもらうことが目的であり，詳しいプログラムの書き方は第 3 講でやりますので割愛します[*20]。強調したいのは，一枚の図の中にはラベルや色を変える

[*15] とある分析屋さんは，データの可視化に 1 週間ぐらいかけて，何百枚というグラフを眺めながらどうやって分析するかの筋道を見つけ出すと言ってました。

[*16] Microsoft の Excel, Apple の Numbers, フリーソフトウェアの LibreOffice にある Calcs などがそうです。

[*17] ひとつ 1 つの命令文のことをコマンド，複数のコマンドからなる一連の命令文をスクリプト，スクリプトがたくさん集まったものをプログラム，といいます。コマンドやスクリプトを描くことをプログラミングと呼ぶこともあります。

[*18] 統計環境 R については第 3 講で触れます。ひとまずは統計分析用のプログラミング言語だと理解しておいてください。またパッケージとは，特定の分析や機能に特化し追加で導入できる関数群のことです。

[*19] この図の元になったデータは，2020 年のプロ野球選手の身長や体重，年棒などをまとめたものです。ここでは身長，体重を守備位置やチームといった群の関係とともに表しています。

[*20] 気になる人は Healy (2018 瓜生他訳 2021) や松村他 (2021) を参照してください。またインターネット上にはさまざまな資料が公開されているので，`ggplot2` で検索してみてください。

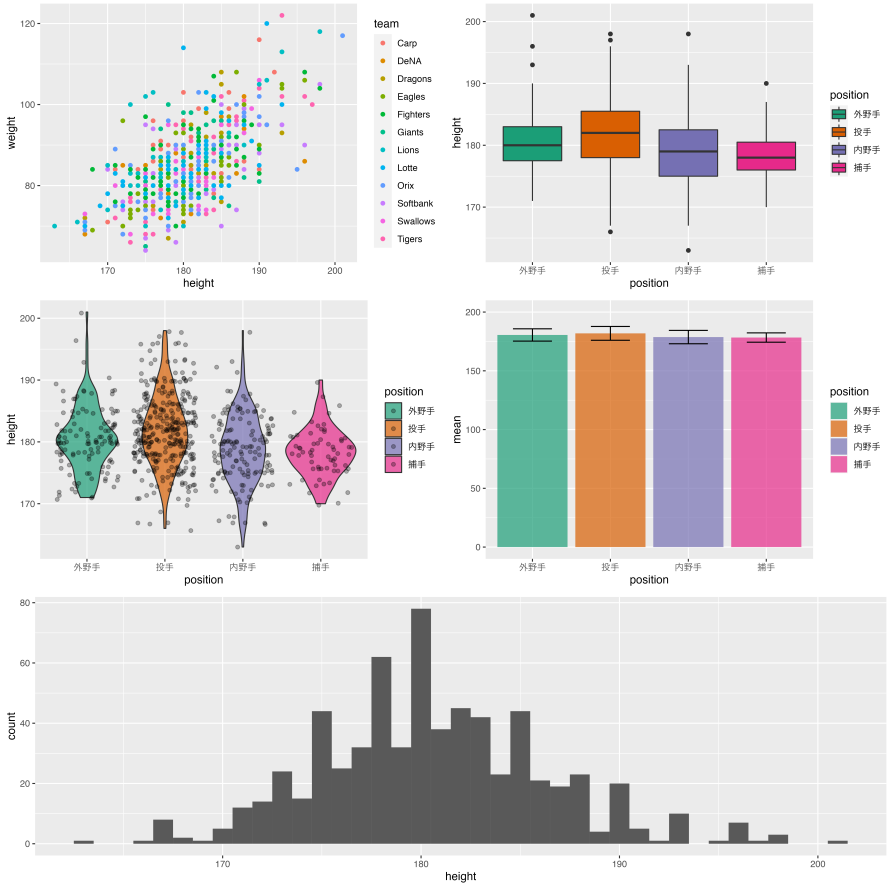


図 2.7　ggplot2 による出力例 1

ことでいろいろな情報を埋め込むことができるということと，これらの図はいずれも描画言語による命令文の積み重ね＝言葉で書かれている，ということです。コンピュータ言語という特殊な言葉ではありますが，どのように描くかということを指定するだけで，形状は機械がデータから作り上げるのです。このスクリプトを見ると誰でも同じ図を描くことができるので再現性が確保されていることを理解してください。

データ分析をするに当たってデータを図にするのですが，どのような図・色・形にすれば最もわかりやすく情報を表現できるか，ということを考えて図表のデザインをしなければなりません。図のデザイン (設計) は「図に書き込みたい情報は何かを考え，準備する」ところから始まり，縦軸・横軸で何を表すか，色・形・大きさで何を表現するか等を順に情報として追加していくことで行います。基本的には次のようなステップでデザインしていきます。

1. X 軸 Y 軸にどのような変数を置くかを考える。
2. X 軸 Y 軸から出来上がるキャンバスに，点や線，何をプロットするのかを指定する。
3. その上にモデルを加えて，元のデータとの対応を考えたりする。別のデータを重ねることもある。

ここでプロットと尺度水準のことを考えてみましょう。X 軸も Y 軸も量的な変数であれば散布図でその変数の関係を表現できます。X 軸が名義尺度水準で Y 軸が量的な変数であれば，散布図も X 軸の対応する水準の上に並ぶような形になります。図 2.10 には 2 つの散布図の例がありますが，図の左は量的変数と量的変数のプロット，右は質的変数と量的変数のプロットになっています。このように同じ散布図であっても，変数の種類が違うと見え方が異なります。とくに右の図は散布図にするより，各水準の平均値を比較するような

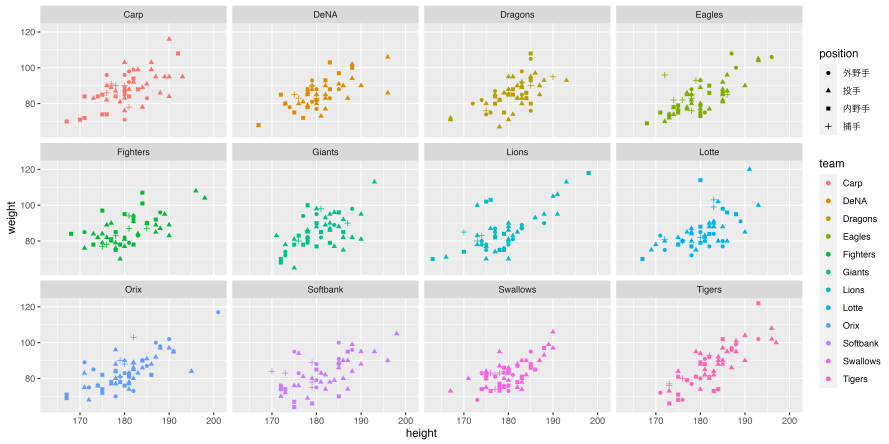


図 2.8 ggplot2 による出力例 2

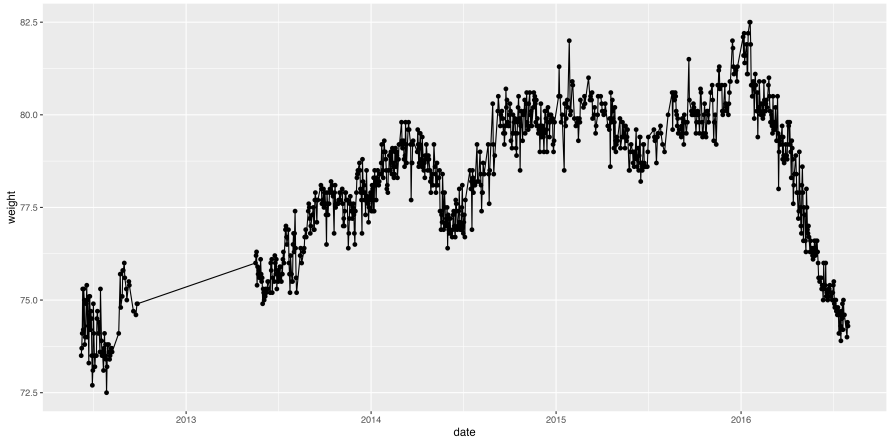


図 2.9 ggplot2 による出力例 3

バープロットのほうが情報を伝えやすいかもしれません。しかしいずれも散布図で，X 軸が質的変数であるときは特殊なならびになるだけだ，という点はこの後の話と大きく関わってきますので，頭の片隅のどこかに置いておくようにしてください。

ところで，たかがグラフを描くためだけに，設計手順を書き下すなんて面倒なことを，と思うひともいるかもしれません。しかし描画の文法化の逆を考えればその利点が見えてきます。作図をこのように文法に沿って言葉で表現するようにしたことで，客観性の担保，再現性の確保にくわえ，どのデータをどのように表現したかったのかが明らかになるわけです。これの逆とはつまり，意図を隠し，主観的で再現できない情報を提示していることになります。これは科学的な態度と正反対です。残念ながら，世の中には悪意を持った人間による人為的で不合理な印象操作を目的とした図表が溢れています。データは可視化することでその特徴を把握しやすくするのですが，逆に視覚的特徴に引っ張られることで異なった印象を持たされることがあります。意図的に悪用すれば，議論をミスリードできます。統計的なデータというのはとくに嘘をついたり誤った印象を与えることがやりやすい領域であるが故に，とくにこの点については注意が必要なのです。

メディアリテラシーという観点からは，Huff (1954 高木訳 1968) など多くの良書があります。一方，どうして図表をよみまちがえるのか，その知覚・認知的な観点からは Healy (2018 瓜生他訳 2021) に詳しく書かれています。Healy (2018 瓜生他訳 2021) は上でもふれた `ggplot` の入門書なのですが，その 1 章はほと

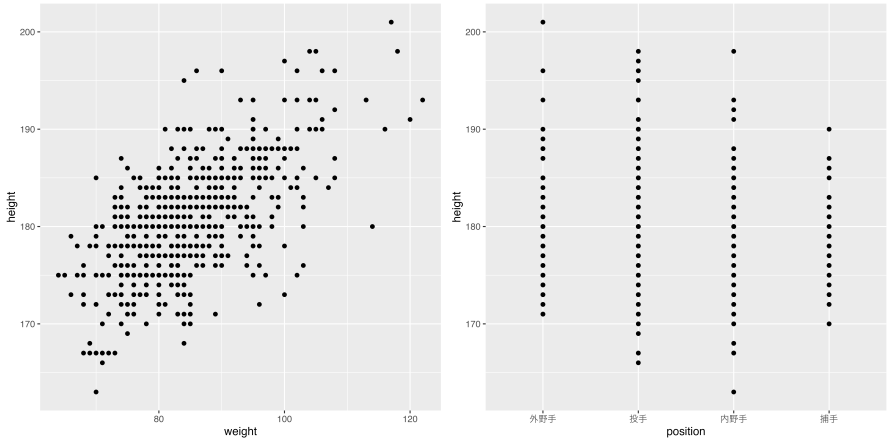


図 2.10 横軸と図の形

んど心理学です。視覚のゲシュタルト性などを例に挙げながらどういう見え方，見せ方をしなければならないかについて丁寧に論じられているので，是非目を通してみてください。

2.5 付録

参考までに，図 2.6,2.7,2.8 を描写したコードを掲載しておきます。R によるコードの書き方や，これらの意味については第 4 講で行います。

なお，描画に使われているデータは，野球選手の身長や体重のデータ (コード内オブジェクト名は `sample`) と，筆者の時系列的な体重変化のデータ (コード内オブジェクト名は `weight`) であり，これらを読み込むコードは記載されていません。

code : 2.1 ggplot2 による描画コード

```
1  # 散布図
2  g1 <- ggplot(sample, aes(x = height, y = weight,
3         colour = team, fill = team)) +
4           geom_point()
5
6  # ボックスプロット
7  g2 <- sample %>%
8    select(position, height) %>%
9    na.omit() %>%
10   ggplot(aes(x = position, y = height, fill = position)) +
11   geom_boxplot()
12
13 # バイオリンプロット
14 g3 <- sample %>%
15   select(position, height) %>%
16   na.omit() %>%
17   ggplot(aes(x = position, y = height, fill = position)) +
18   geom_violin(alpha = 0.7) +
19   geom_jitter(alpha = 0.3)
20
21 # バープロット
```

```r
g4 <- sample %>%
  select(height, position) %>%
  na.omit() %>%
  group_by(position) %>%
  summarise(mean = mean(height), sd = sd(height)) %>%
  ggplot(aes(x = position, y = mean, fill = position)) +
  geom_bar(stat = "identity", alpha = 0.7) +
  ylim(0, 200) +
  geom_errorbar(aes(ymin = mean - sd, ymax = mean + sd), width = .5)

# ヒストグラム
g5 <- sample %>% ggplot(aes(x = height)) +
  geom_histogram(binwidth = 1)

g6 <- ggplot(sample, aes(x = height, y = weight,
            colour = team, fill = team, shape = position)) +
  geom_point() +
  xlab("height") +
  ylab("weight") +
  facet_wrap(~team)

# 折れ線グラフ
g7 <- weight %>% ggplot(aes(x = date, y = weight)) +
  geom_point() +
  geom_line()
```

2.6 課題

■変数を見つける 次のような研究計画を立てた時に，変数として記録しておくべきなのは何でしょうか。

1. 人の身長と体重の情報から，性別を予測することに興味がある。
2. 人は寝ている間にたくさんの汗をかくという。一晩にかく平均的な寝汗の量に男女差があるのかどうかを調べたい。
3. 結婚している人としていない人とで，幸福度に差があるかどうかを知りたい。

■尺度水準を理解する 次のそれぞれのデータは，どの尺度水準に該当するでしょうか。

1. 個人を識別するために回答を求めた携帯電話番号の下 4 桁の数値
2. 納税証明書に記載されている，個人の昨年度の納税額
3. 年齢を直接質問するのがはばかられたため，「10 代」「20 代」「30 代」「40 代」「50 代以上」の 5 段階について回答を求め，それぞれ 1,2,3,4,5 とコード化した数値
4. 質問に対して「とてもそう思う」を 5，「まったくそう思わない」を 1 として，両者の間を均等に分割し 2,3,4 とする。また，「答えたくない」を 0 としてコード化した数値
5. セルシウスさんが開発した，沸点を 100，氷点を 0 として百分割した温度
6. 温度とは分子の平均運動量であり，一定量の物質の温度をどんどん下げていくとこれ以上低くはならないという点が出てくる。そこを原点として作られた絶対温度 K
7. テレビ番組「帰れま 10」では，ある外食チェーン店の豊富なメニューの中から，上位 10 品が何である

583 かを当てなければならない。このとき, それぞれのメニューに付される人気の順位を表した数字

第 3 章

統計環境 R の導入

ここまで，データを可視化したりデータの代表値を計算する方法について学んできましたが，実践上これらはすべてコンピュータを使って行います。今回は PC を使った実践的な内容になります。パソコンを実際に触りながら，意図したことを正しく機械に伝え，意図した結果を得られるようにスキルアップしていきましょう。

3.1 パソコンのファイル構造

3.1.1 パソコンの世界

皆さんはパソコンをどういう用途で使うことが多いでしょうか。インターネットで調べ物をしたり[*1]メールの読み書きをする，というのが多いかもしれません。Facebook や Twitter，Instagram など SNS を利用する，これもインターネットに関係するアプリケーションですね。これらはスマートフォンやタブレットなどの方が手軽で便利にアクセスできるので，それらを使うことの方が多いかもしれません。

インターネットにアクセスするのではなく，手元の PC で (ローカル環境で) 作業する例としては，文章を書くことが多いでしょう。大学生をはじめ研究者というのは基本的にコミュニケーションをする人ですから，半ば文筆業であると言っても過言ではありません。またこの文章を作成するために，必要な資料を作ることがあり，数値データを扱う場合は表計算ソフトを使うことになるでしょう。表計算ソフトの代表例として，Excel，Numbers，Calcs などがあります。表計算ソフトは記述統計量やグラフの作成など，これまで学んできたことが簡単にできるものがありますが，それ以上の複数の変数を一度に扱った分析をするのには，機能的に不十分です。

そこで，より専門的な統計的処理をするためには，統計専用のソフトウェアを使うことになります。統計専用のソフトウェアとしては，SPSS や SAS，Stata などの商用ソフトウェアもありますが，専修大学人間科学部心理学専攻では現在，R を学科共通のツールとして利用することにしています。パソコンでこうしたソフトウェアを使うに先立って，重要な「ファイルの場所」について説明します。すでに情報の授業などで聞いたことがある話でしょうし，使い慣れている人にとっては当たり前の話なのですが，携帯電話端末やタブレットではファイルの位置がユーザに見えにくくなっていることも確かです。パソコンに苦手意識があったり，PC のトラブルに慣れていない人は今からのお話と，できれば付録 (付録 A) も併せて目を通しておいてください。

[*1] 正確には，検索サイトを使って目的とするウェブサイトにアクセスする，という言い方になります。ウェブサイトを見る (ブラウズする) のはブラウザーというアプリケーションを使います。Chrome，Safari，Edge，Firefox などがブラウザーアプリケーションの名前で，Google，Yahoo など最初に開くウェブサイトのことをホームページと言います。ウェブサイトは個人や企業が提供しているものが多く，Google や Yahoo などは検索サイトと言われます。

3.1.2 ファイルの場所

パソコンを使うにあたって話をしておかなければならないのは，まず「ファイル」という考え方です。パソコンをはじめ，スマホやタブレットなどコンピュータ，あるいは万能計算機は，いろいろな機能を有しています。文字を書いたり数字を計算したり，音楽を鳴らしたり映像を加工したり，ですが，これらはすべて 0/1 というデジタルなデータであり，文字，数字，音，画像，動画などすべて 0/1 というデータに規格化したことが**情報化社会**の本質です。その中身が何であれ，データのまとまりのことをファイルと言います。

アプリケーションを実行するのもファイルですし，データを保存しておくのもファイルです。パソコンが機能するのに必要なファイルは非常にたくさんあるので，整理するためにディレクトリ (フォルダ) の中にまとめて，階層構造を持って保存されています。複数のファイルはフォルダにまとめますが，フォルダをまとめたフォルダ，というように「階層構造」をつくっています。図 3.1 や 3.2 に，Windows や Mac でファイルを見るための画面 (アプリケーション名がそれぞれエクスプローラ，ファインダーと異なります) を示しています。いずれもアイコンの色や形は違いますが，フォルダの中にフォルダやファイルが含まれる，という形になっているのがわかると思います。

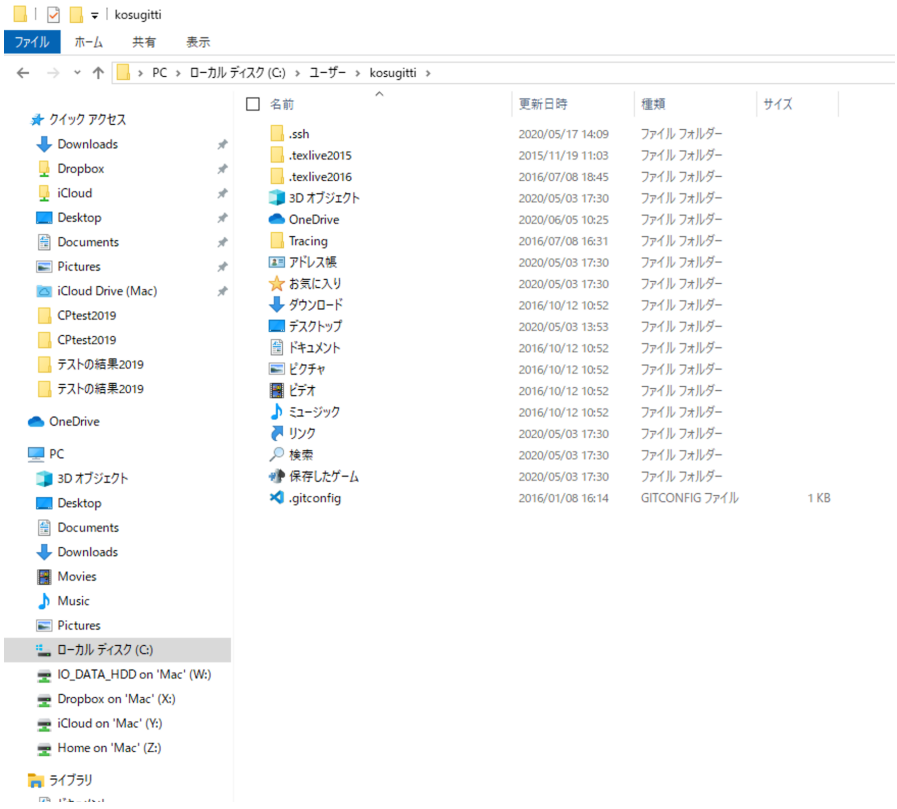


図 3.1 Windows のエクスプローラ画面

ここで Windows の画面にある，エクスプローラの上部を見てみましょう。この例では「PC>ローカルディスク (C:)>ユーザー> kosugitti>」とありますね。日本語の文字が入っているので 2 バイト文字なのかと思いきや，ここをクリックしてみると「C:¥Users¥kosugitti」と出てきます。実はこちらが本当の名前で，ローカルディスクとかユーザーと表示されていたのは使い手に優しく表現しているだけのイリュージョンなのです。こ

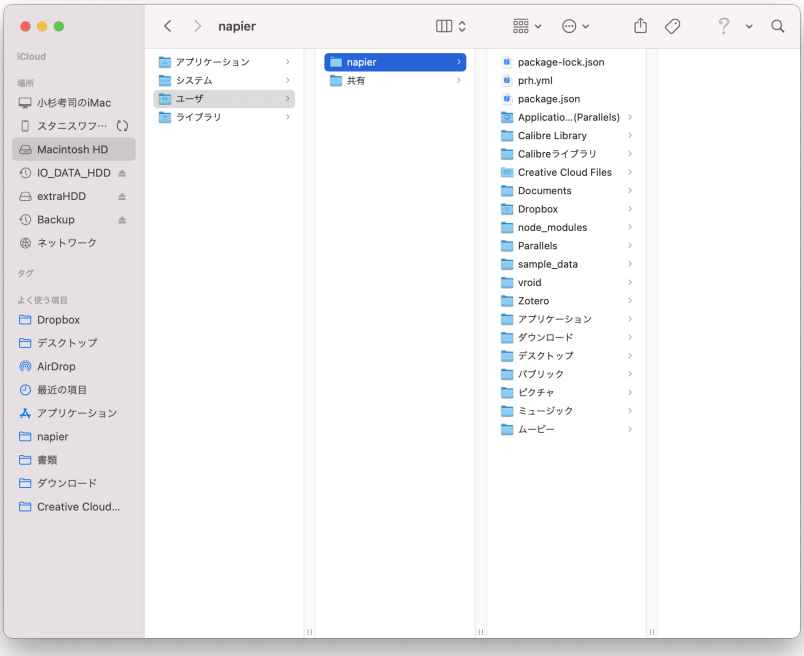


図 3.2 Mac OS のファインダー画面

こで C: とあるのはドライブと呼ばれ[*2]，記憶装置の実体を指します。この場合はパソコンに内蔵された記憶装置 (SDD や HDD) です[*3]。この C ドライブの中に，Users というフォルダがあり，その中に kosugitti というユーザが使う領域がある，という区分になっています[*4]。図 3.4 には Mac の画面を示していますが，ここでもドライブ，フォルダ，フォルダ，という構造になっていること，ユーザのなかに napier というユーザ名があることなどが確認できると思います。Ubuntu など別の OS でもこうした仕組みは同じです。これらの画面を見ながら自分で使うファイル，作ったファイルがどこにあるのかをはっきり把握できるようにしておいてください。

これからの大学生活で，皆さんはいろいろな授業でレポートを書いたり，統計計算をしたりします。それらを「書類」というフォルダの中にまとめて入れてしまうと，後で探す時に大変ですから，授業ごと，あるいは研究テーマごとなどにフォルダを作り管理するようにしてください。この授業でも，よく相談されるエラーが「ファイルがありません」というものです。作ったはずのものがない，保存したはずのものがない，というトラブルの原因はほぼ確実にユーザの責任で，置いた場所がわからなくなっているのです。これについては相談されても「自分で思い出してください」としか言えませんので，ファイルの整理整頓を常に心がけてくださいね。

[*2] なんで C から始まってて，A や B じゃないんだ，と思われるかもしれませんが，これにも歴史的経緯があります。かつてのパソコンは，フロッピーディスクドライブ (FDD) が 2 器ついているのが基本で，それらに A,B と割り振られていました。その後に出てきた内蔵ハードディスクドライブは C とするのが順当だったわけです。その後 FDD は時代とともに使われなくなりドライブも無くなったのですが，ドライブレター (ドライブに割り当てた文字) は残ったのです。

[*3] ハードディスクドライブ (HDD) とは，FDD の中に入っている磁気媒体の円盤が積み重なったもので，FDD よりも大容量のデータを保管できます。FDD が 1MB ぐらいであるのに対し，HDD は数 TB はあります。一方ソリッドステートドライブ (SDD) とは，USB メモリのようにメモリチップに記憶するもので，HDD のように回転することはないので静かですし，アクセス速度が HDD よりも速いため，最近ではこちらが内蔵ドライブになっていることが多いでしょう。いずれにせよ，内蔵ドライブの種類の違いで悩むことはありません。

[*4] 最初にパソコンを購入した時に，このパソコンで使うユーザの名前は何ですかと聞かれると思います。その時に入力したものがユーザ名になりますが，その時に 2 バイト文字でユーザ名を入力していると，その後のアプリケーションがうまく動かないエラーが出てくることになったりします。

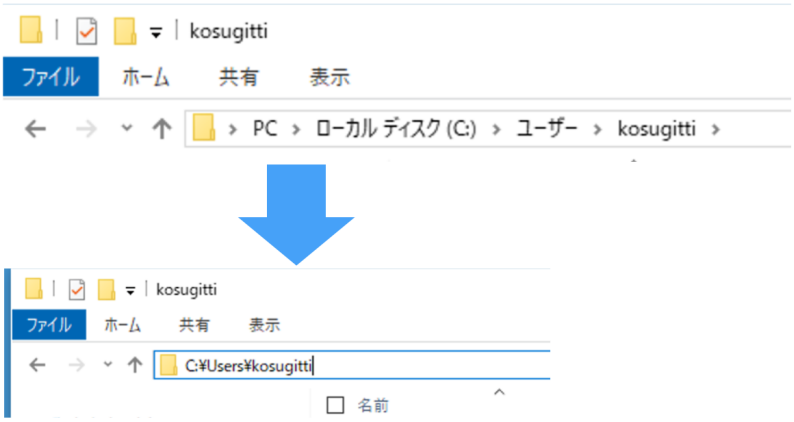


図 3.3 Windows のフォルダ階層

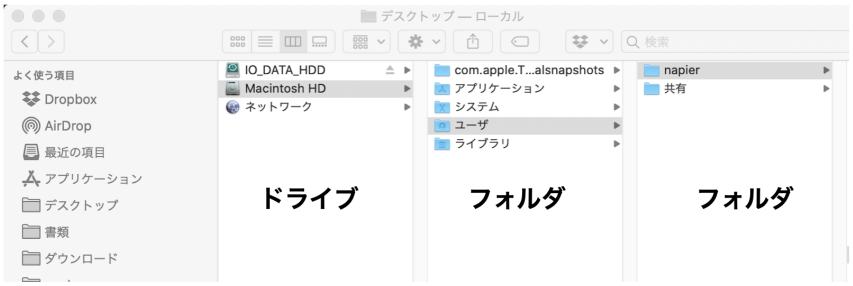


図 3.4 Mac OS のフォルダ階層

3.2 統計環境 R/RStudio の準備

それではやっと本質，パソコンを使っていくことにしましょう。とくにここでは統計環境 **R** について解説します。

R は統計環境，統計言語とも呼ばれることがあります。統計に関する分析を CUI で行うので統計 “言語” と呼んだり，数字だけでなく図を描くなど数値処理に関わることを全体的にカバーしているので “環境” と呼ばれたりします。R が優れている点はフリーソフトウェアであるという点で，誰でも無償で利用できることです。商用ソフトウェアと違って，ソースが公開されているのでその気になれば誰でもその内部構造を調べることができますから，間違い (エラー，バグ) がないかどうかをみんなの目でチェックできるというのが利点です。金銭的に補償があるわけではないのですが，そもそも学問の世界は市井の資本主義的発想とは異なる価値で評価される世界ですから，無償であることは問題になりません。

また R は多くの研究者が利用しており，その研究成果をフリーソフトウェアの精神で共有しています。こうした R の基本関数以外の機能は，パッケージという形で提供されます。R のパッケージを共有している場所は The Comprehensive R Archive Network(**CRAN**) と呼ばれ，ここに登録されているパッケージは誰でも利用できます。最先端の分析ツールがここを介して共有され，間違いや問題が生じたら随時修正されていきますので，成長する統計ソフトウェアということができるでしょう。

またこの授業では，R を使うにあたって **RStudio** という総合開発環境アプリを使います。R というアプリケーションを便利に使うためのアプリケーション，それが RStudio です。

R は優れた統計ツールではありますが，本体だけでは簡素な言語的やりとりができるに過ぎません。分析を進めていくとすぐにわかりますが，ファイルの操作や描画機能，過去ログの参照や管理など，統計言語を扱う際にそれを取り巻くさまざまな操作が必要になります。ここをカバーしてくれるのが RStudio というアプリケーションなのです。

RStudio はそれ単体では使うことができません。R が入っている PC 上で，R を取り込んで初めて意味のあるアプリケーションになります。R の計算機能を活用するだけでなく，文書作成機能やデータベースとの連携など非常に高機能な統合的サービスを提供してくれます。文章を書き，その文章に必要な計算を行い，その結果がすべて記録 (スクリプト，コマンド，プログラム) として残るわけですから，科学論文を書くためには RStudio だけでも十分です。

3.2.1 環境を構築しよう

専修大学では 2023 年度から，仮想デスクトップ環境 (VDI) の活用を推進しています。大学が仮想空間の中に個々人のデスクトップ環境を用意しているので，私たち利用者は自分自身のデバイスからアクセスするだけで，全員共通の環境が利用できるというのが利点です[*5]。ここへのアクセスの仕方，利用方法については専修大学情報科学センターからのガイダンスや，利用案内ページがあるのでそこを参照してください。

この，本学が提供する仮想デスクトップ環境や，心理学科が管理する 4 号館 4F の PC 室には R や RStudio がすでにセットアップされていますので，これを利用するのであればとくに準備は必要ありません[*6]。

しかし大学などの設備は大学が管理していますので，自分でパッケージを追加したい，とおもっても管理権限がなくうまく行かないこともあるかもしれません。幸い R や RStudio はフリーソフトウェアですので，手元の PC にインストールして環境を構築するにしても，元手は一切かかりません。これを機に，自分のパソコン環境に R/RStudio を準備しておくと良いでしょう。

R や RStudio の環境構築については，その方法を検索すればさまざまなものが出てきますので，そちらを参考にしてください[*7]。新しいものでは，高知工科大学矢内先生のサイト[*8]とか，Syunsuke Fukuda 先生の初心者向け R のインストールガイド[*9]などは網羅的かつ丁寧なので，よく読んで始めると良いでしょう。Quiita というエンジニアがよく参照するブログの記事[*10]なども良いと思います。書籍では，手前味噌で恐縮ですが小杉 (2019) などが良いでしょう[*11]。

以降は環境が準備できているものとして解説を続けます[*12]。

[*5] 自分のデバイスを持ってきて繋げる (Bring Your Own Device;BYOD) ので，スマートフォンやタブレットでも PC と同じような利用の仕方ができるのがメリットですが，教育的な側面では大学が準備すべき設備備品費用を個々人に押し付けているとも言えます。大学が PC を全員に配布できればいいのですが。

[*6] 他にも心理学科には，計算機サーバというのがあり，ここに RStudio Server が準備されています。RStudioServer は，インターネットを介するウェブブラウザから RStudio にアクセスすることができるものです。残念ながらまだ試験的導入段階で，心理学科の学生が一気にアクセスするとサーバがダウンしてしまいますので，希望者のみの利用にとどまっています。

[*7] R の唯一の弱点は検索可能性が低いこと。R という一文字で検索してもまともなものにヒットしないので，“CRAN” とか “R と RStudio インストール” など周りに言葉を足して検索してみてください。

[*8] http://yukiyanai.GitHub.io/jp/resources/

[*9] https://syunsuke.GitHub.io/r_install_guide_for_beginners/

[*10] https://bit.ly/3MfK3SD

[*11] なお，インストールに際しては，RStudio Desktop の Free edition を選んでください。RStudio Cloud や RStudio Server は別のものです。

[*12] トラブルがあって環境の準備がうまくできていない場合は，`classkosugitti@gmail.com` までご相談ください。

3.3 R/RStudio の基本操作

それでは今から，RStudio を使った統計処理の演習を始めましょう。少し不便かもしれませんが，この資料を見ながら，**必ず自分でコードを入力しながら**進めていってください。テキストと同じ結果が手元で再現できることを確認しながら進むことは，再現できることの重要性の確認になります。また，最初のうちは打ち間違えなどで思うように動かないことも頻発するでしょうが，その経験を繰り返すことで徐々に使い慣れていきます。失敗の経験はテキストの文中に書き込まれていませんから，読むだけでは身につきません。また，読み進んで気が向いたところだけ実行するときは，その前のデータや内容を参照している場合にエラーになるので，結局元に戻ってやり直すことになります。最初のうちは写経して作法を身につけることも重要なのです。最初の基礎的な訓練[*13]ができていると，徐々に大きなプロジェクト[*14]を動かすことができるようになりますので頑張って！

3.3.1 RStudio の機能と画面

最初に RStudio を起動すると，図 3.5 のような画面になっていると思います[*15]。ここでメニューから `File > NewFile > RScript` と進むと (図 3.6)，画面に白いページの領域が追加されます (図 3.7)。大きくみると画面が四分割されたようになります。この 4 つの領域を基本として話を進めていきます[*16]。

■コンソール・ペイン デフォルトでは画面の左下になりますが，**コンソール (R の出力)** ペインにあるのが R そのものです。RStudio は R を取り込んで使いやすくする環境を提供するもの，という話をしましたが，とりこまれた R だけの画面が左下にあります。そこには R のバージョン情報や，次のような注意書きが書かれています。

R の出力 3.1

```
R version 4.4.0 (2024-04-24) -- "Puppy Cup"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin23.4.0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
```

その下に>という記号が出ていると思います。これは入力を待っている状態でこの記号の左側に命令をどう

[*13] キーボードの入力に慣れていない人は，ブラインドタッチの練習をすることをお勧めします。キー入力は今後も活用できる社会人の基本スキルですので，なるべく早く習得しておいた方が人生が楽しくなります。ブラインドタッチの練習には，寿司打などのサイト http://typingx0.net/sushida/play.html で遊びながら学べていいですよ。ちなみに今やってみたら，私の成績は「高級 10,000 円コース【普通】で、★ 5,820 円分 お得でした!(速度: 6.1key/秒、ミス: 57key」でした。

[*14] 卒論は 4 年間の集大成になる個人プロジェクトです。

[*15] こちらは MacOS11.1，RStudio は 1.4.1103，R は 4.0.3 の環境におけるスクリーンショットです (撮影時 2021 年 2 月 2 日)。2024 年 4 月 25 日の段階では，MacOS14.1.1，RStudio 2023.12.1+402 "Ocean Storm" Release，R version 4.4.0 (2024-04-24) です。OS が違うとか，バージョンが違う，サーバ版を使っているとかで細部が異なっている可能性がありますが，ほとんど変わらないと思います。。

[*16] RStudio version1.4 から，列を追加することができるようになりましたので，必ずしも四分割で使わなければならないというわけではありません。どこに何があるのかを把握していれば OK です。ちなみに列を増やしたり配置を変えたりするのは，`Tool > Global Option > Pane Layout` から可能です。

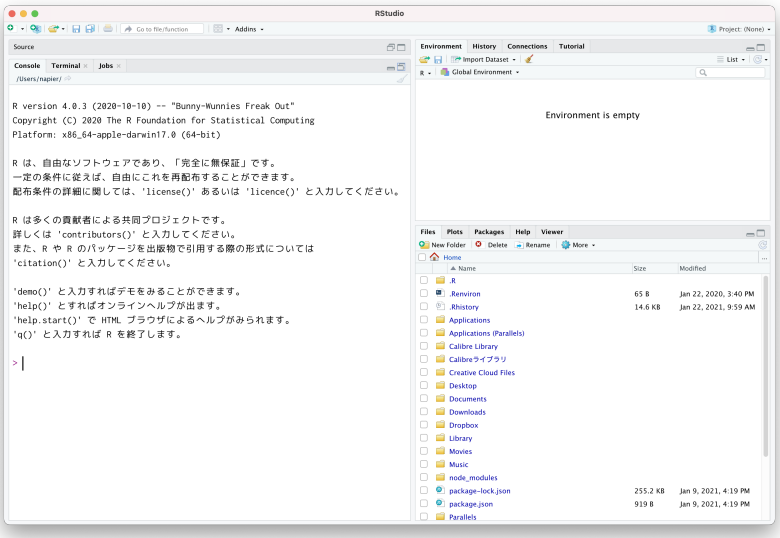


図 3.5 RStudio 起動初期の画面

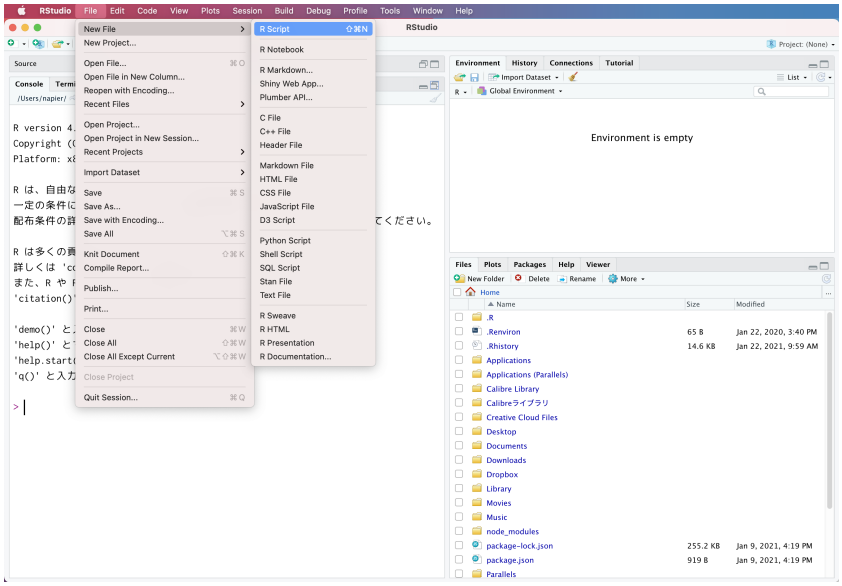


図 3.6 新しくスクリプトファイルを開く

ぞ，という意味です。この記号が出ているときは，R は次の動作準備 OK で指示を待っているという状態です。ここにコードを入力すれば計算結果が出力されるのですが，その前にスクリプトペインの説明をします。

■スクリプト・ペイン スクリプトペインは，R のコードを書くところです。R は一問一答型 (インタプリタ[*17]方式ともいいます) なのですが，実際の統計処理はあれをして，これをして... とたくさんの処理が連なる

[*17] プログラミング言語の実行方法として，コードを逐次翻訳していく方法と，プログラム全体を一気に機械語に翻訳して実行する方法のに種類があります。前者をインタプリタ方式，後者をコンパイラ方式と言います。インタプリタ方式の利点は書いたコードがすぐに実行されるので，どこでどういう処理がされているか分かり易いことです。エラーがあってもすぐにその場で帰ってきます。コンパイラ方式の利点は実行速度です。全部の処理を一気に機械語に翻訳してあるので，計算速度が非常に早くなります。コンパイラ方式の欠点は，どこでエラーが生じたかわかりにくい点です。

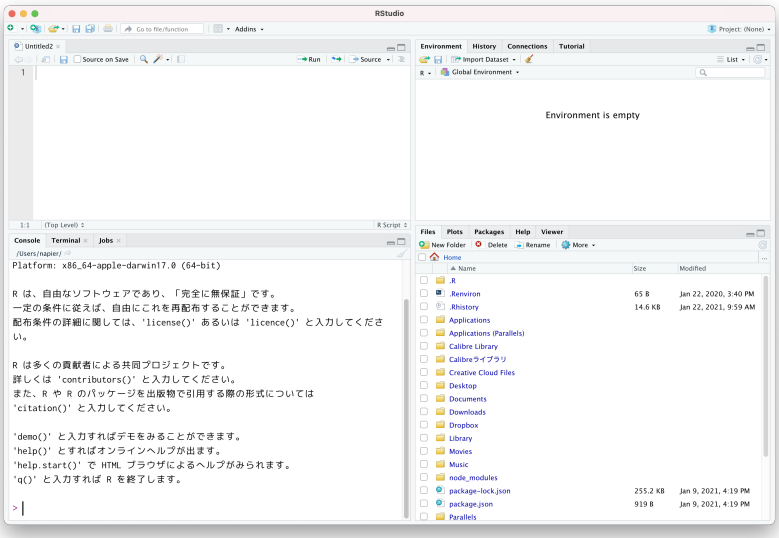


図 3.7 RStudio の画面

ことになります。この一連の手続きを書いておくのがこの場所です。

スクリプトとはお料理のレシピのようなものです。1 つ 1 つの計算は，たとえば「玉ねぎをみじん切りにする」といった単一の作業をさせることになりますが，全体としてハンバーグを作るのが目的であれば，1 つ 1 つの作業をすべて順番通りに実行しなければなりません。直接コンソールペインに書き込んで言ってもいいのですが，途中で「今どの工程だっけ」というのがわからなくならないとも限りませんので，いったんスクリプトのところに書いてから実行する，という手段を取るのが便利です。

実際にコマンドを 1 つ 1 つ書いていくときには，書き間違えたり違う方法を試したくなったりするので，書いたり消したりしながら進むのが普通です。ですが最終的には清書したものだけを残すようにしましょう。下書きの要素もすべて記録していると，あとで何がどうなったかわからなくなります。また，何度か繰り返しているうちに「何にどのような操作をしたか」がわかっているようでわからなくなることがあります。最終的には，1 つのスクリプトファイルを一行目から順に実行していけば，すべての作業が再現できるようにする，ということが重要です。途中で別のレシピを参照したり，記録されていない操作を挟むのはやめましょう。

さてでは，このスクリプトペインに code:3.1 を書いてみてください[*18]。

code : 3.1 四則演算のコード

```
1  # 四則演算
2  1+2
3  2-4
4  3*5
5  6/3
6  sqrt(16)
```

■コード解説

1 行目 コメントアウト。#記号で始まる行は，その後ろに何を書かれていても R は無視します。なので，この記号を冒頭に描いて，ここに何をしているかという説明を書き入れています。このコメントは，あとで見

[*18] この資料では指示がわかりやすくなるように行番号を含めて表示していますが，実際は行番号を入力する必要はありません。

直す時に何をしていたか判るためのものなので，必ず書かなければならないというものではないのですが，あればあるほど親切で丁寧なコードと言えるでしょう。

2-5 行目 順に和，差，積、商を求めるコードです。積*と商/の記号は PC ではこのように書きますので注意してください。

6 行目 これは関数を使って答えを出しています。`sqrt` 関数は正の平方根を返します。

書いたあと，スクリプトペインの右上にある `Run` を押すと一行ずつ実行されていきます。あるいは，CTRL+Enter でもスクリプトをコンソールに送り込むことができます。マウスで複数行を選んで `Run` ボタンを押すと囲んだ箇所を実行することもできます。このスクリプトペイン全体を実行したいときは `Source` ボタンを押します。

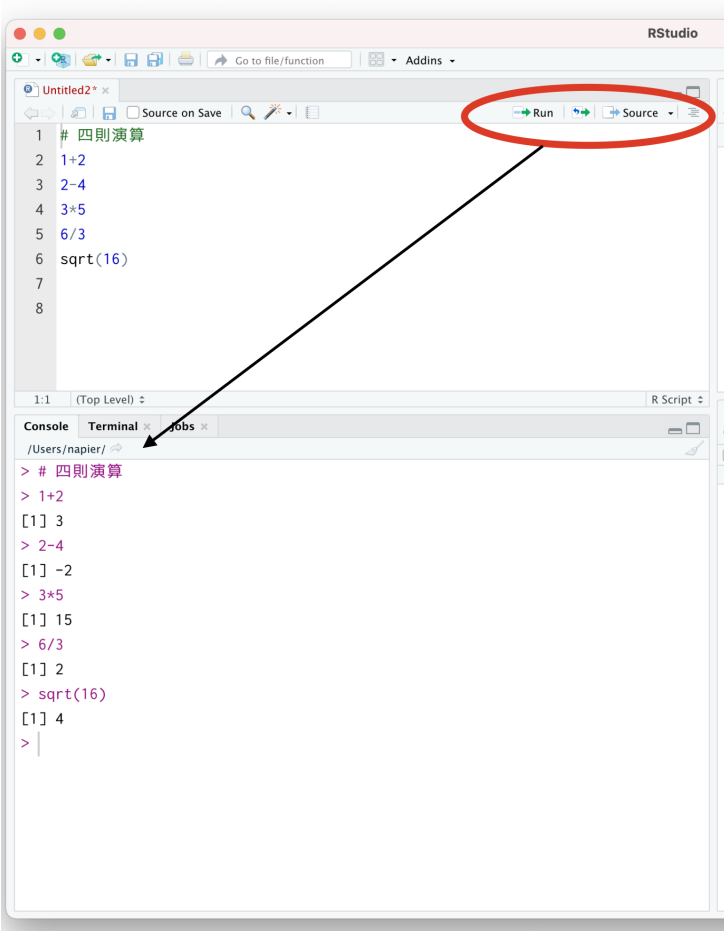


図 3.8 スクリプトを実行する

これらのボタンを押すと，コードがコンソールに送られて実行結果が返ってきます (図 3.8)。コンソールに示されているものは，この資料では次のような囲みで表現します (**R の出力 3.2**) が，さて，みなさんの環境で同じ結果が得られましだでしょうか。

R の出力 3.2: 四則演算

```
> # 四則演算
> 1+2
[1] 3
> 2-4
[1] -2
> 3*5
[1] 15
> 6/3
[1] 2
> sqrt(16)
[1] 4
```

出力をみると，1 つ 1 つの問い (たとえば 1+2) に対して，毎回答え (たとえば [1]　3) が返ってきています。一行ずつ実行した人はわかると思いますが，毎回>の記号がでて R が回答準備ができている時に，次の命令を受け入れることになります。もしおかしなコマンドが入っていれば，エラーが出ているはずですので，それをよく読んでどういうエラーなのかを考えましょう。我々が書くコードは，ミスやバグが多分に含まれます。私もいまだにスペルミスでエラーを発生させています。複数行を一気に実行してしまうと，どこにエラーが出ているかわかりにくくなりますので，一行ずつ実行することを強くお勧めします。

ところで，$1+2$ の答えは 3 ですが，出力の [1] はなんでしょう？ これは答えの要素がひとつ，という意味です。これからたくさんのデータを一度に処理することがありますが，そのときは答えがたくさん出てくることがあるので，その場合は何行にもわたって答えが返されることがあります。詳しくは後ほどの実践例で説明していきます。

さて，これが基本中の基本の使い方です。実践的な内容に入る前に，そのほかの RStudio ペインの説明をしていきましょう。

■Environment/History・**ペイン**　画面の右上にあるのは Environment や History のペインです。Environment は環境，すなわちこの R の頭の中にどういう変数、関数が入っているかを示してくれるところです。たとえばたくさんのデータセットを読み込んだ，というときは，それがどういうデータなのかを示してくれたりします。

History は履歴を表すタブです。ここをみると，今実行した過去のログが残っています。「さっき書いた命令をもう一度実行したい」というときはここで探すといいでしょう。このペインには，`to Source` や `to Console` といったボタンもあります。ある行を選んでこのボタンを実行すると何が起こるか，確かめてみてください。

■Files/Plots/Packages/Help/Viewer・**ペイン**　右下のペインはよく使う便利なところです。

1. Files ペインは，ファイル操作をする場所です。ファイルのコピー，移動，フォルダの作成などができます。エクスプローラや Finder と同じようなものだと思ってください。
2. Plots ペインは図の出力を確認するところです。ここで作られた図を適当なサイズやファイル形式に変換して出力することもできます。
3. Packages ペインは，R のパッケージを管理するところです。R はパッケージを追加することで機能がアップしますが，この追加やアップデートをこのペインで実行すると便利です。
4. Help ペインは，関数のヘルプが表示されるところです。表記はすべて英語なのが残念ですが。
5. Viewer ペインは HTML などの出力をみる画面です。

これらは必要に応じてみていくことになりますので，ここでは大体のイメージが掴めていれば OK です。また これらのペインは場所を変えることも可能ですので，自分の良いようにレイアウトしてみてください。

3.3.2 プロジェクト管理

さて環境の意味がわかったところで，いよいよ実践に入っていきます。先ほども少しコードを書いてみましたが，今度はより本格的に使っていこうと思います。

プロジェクトによる管理

RStudio を使う時の基本は，プロジェクトによる管理です。分析するデータ、テーマ、内容によってプロジェクトを切り替えながら使います。みなさんはこの授業の他にも，R を使う授業があるかもしれません。また，学年が上がっていくについれて，さまざまな心理学実験，研究に関わっていくことになります。その時、すべてを同じドキュメントフォルダに入れてしまうと、何が何だかわからなくなるので，フォルダごとに区分しますよね。それと同じように，R で作業する場所を RStudio でも区切りながら管理しようということです。

メニューバーから `File > New Project` と進んでください。すると「New Directory」「Existing Directory」「Version Control」の選択肢が出てきます。New Directory は新しいディレクトリ[*19]を作ってそこで関連ファイルをまとめますよ，という意味です。Existing Directory はすでに存在するディレクトリをまとめる場所にしますよ，という意味です。みなさんの環境に応じて使い分けてください。ここでは新しくフォルダを作る例を説明します[*20]。

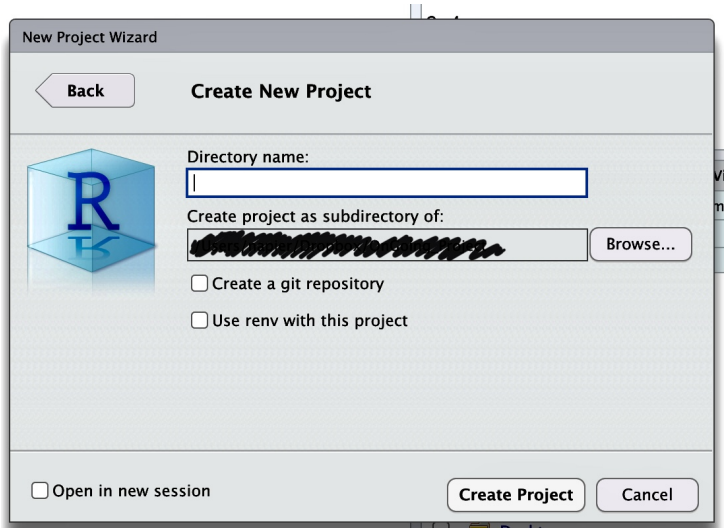


図 3.9 プロジェクトを作る場所を選ぶ

図 3.9 には `New Directory > New Project` と進んだ次のウィンドウを示しています。黒く塗りつぶしているところは私の個人的な PC 環境の内容が反映されるので誤魔化しているのですが，みなさんの環境にあった場所 (Document とか Desktop と言った場所) を選び，上の空白になっているところに新しいプロ

[*19] 計算機におけるディレクトリとフォルダは同じ意味です。

[*20] Version Control はパスしましたが、研究をする上ではこれが重要になってきます。つまり，何月何日にどのような作業をしたか，という記録をすべて残しておく必要があるからです。研究ノートやラボノートと言われる記録をつける時に必要です。この記録は不正をしていない証拠にもなります。ここを選ぶと Git や Subversion といったバージョン管理ソフトと連携できます。バージョン管理ソフトについては付録の A.8 も参照してください。

ジェクト名 (たとえば toukei など[*21]) を入力して OK とします。するとしばらくすると，画面が変わるはずです。一見同じ状態のようにも見えますが，実はところどころ違いがあります。気がつきましたか？ 図 3.10 には

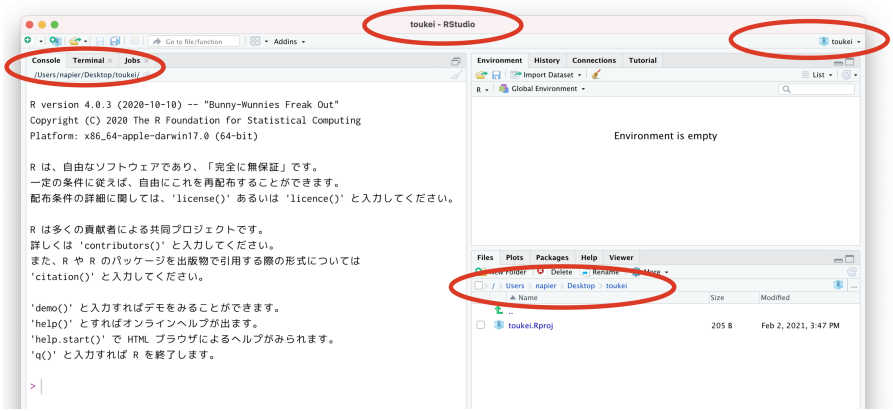


図 3.10 プロジェクトで開いている場合

プロジェクトを開いている時の RStudio 画面を示しています。赤く丸で囲っているところが，プロジェクトを開いていない時との違いがある箇所です[*22]。これをみると，コンソールペインや File ペインの場所に，今指定したフォルダの位置が示されていることがわかります[*23]。このプロジェクトが開いているフォルダに移動していることが重要で，このフォルダの中が今作業している場所として認識されるので，このフォルダの中にあるファイルにはファイル名だけでアクセスできるようになります。プロジェクトが違ったり，プロジェクトの起動をしていなかったりすると，作業している場所[*24]が変わりますので，注意が必要です[*25]。

3.3.3 エラーはヒントです。怖がらないで

以下ではみなさん自身で R コードを書いていくのですが，誰しも，間違いなく，エラーに遭遇します。エラーはムカつきますし，理由がわかりませんし，何書いているかわからん！ってなりますよね。

でも，エラーはヒントなのです。ここでこういうエラーが生じましたよ，というヒントをくれているし，それに対応しさえすればちゃんと動いてくれます。プログラムは思った通りに動くのではなく，書いた通りに動くのです。ですから，赤字で何か出てきた場合はどういうエラーなのかを読み取るようにしましょう。何行目の，どの部分で生じたエラーかがわかれば，対応の仕方がわかります。

エラーが出て困ったら，遠慮なく私や TA を頼っていただいて構いません。質問するときのコツは，エラーが出た箇所「だけ」を見せてはいけないということです。R は一行ずつ実行していきますから，ある箇所でエラーが出たとしても，その原因がその上流で生じている可能性があります。学生さんの中には，エラーが出た画面の写真をとって送ってくる人も少なくありませんが，大抵はその写真に写っていないところにエラーがあります。そんなときは「ごめん，もうちょっと上をみせて」という返信をするので，回答まで無駄なラリーが続くことがあります。質問するときは，実行しているスクリプトファイルすべてを添付する方が確実で，「これの XXX 行目でエラーと言われます」と報告しましょう。スクリプトのやり取りの利点は，私も R を持っていますから，私の環境で同じことをしてエラーの再現ができる点にあります。

[*21] 半角英数のみで作るべきです。
[*22] MacOS の例なので Windows 版では上のバーのところなど，違いもあります。
[*23] この例では `C:\Users\napier\Desktop\toukei` です。napier は私の PC の名前，toukei はプロジェクト名です。
[*24] ワーキングディレクトリといいます。詳しくは付録の A.7 を参照。
[*25] 実際に授業をしていて，ファイルへのアクセスエラーで相談を受ける原因第 2 位です。第 1 位はファイルエンコードの問題です。

自分の書いたコードが恥ずかしくて見せたくない，と言う人もいるかもしれません。あるいは「こんな簡単なことで質問してくるな」と思われるのでは，と思ってしまう人もいるかもしれません[*26]。そうした心理的障壁を乗り越えるのは難しいですが，そんな場合は ChatGPT をはじめとする生成 AI を利用しましょう。最近は (アカウントの登録は必要ですが) 無料で利用できます。そこにエラーメッセージやスクリプトを貼り付けて，「このエラーがこのスクリプトから出た。どこが問題か教えて」というとスラスラっと答えてくれます。生成 AI はシレッと間違えた答えをしてくるのが難点ですが，答えの通りにやってもエラーが出たら「まだエラーが出るんですけど」と言えば，さらにそれっぽい答えを教えてくれます[*27]。

繰り返しますが，エラーメッセージはヒントなのです。メッセージが自分にとっては意味不明でも，わかる人が見ればわかる情報が満載なので，エラーメッセージとスクリプトのセットで質問するようにしてください。

3.4 R の基本操作

ではこのプロジェクトの中で、演習を進めていきましょう。先ほどと同じように，`File > New File` から `R Script` と進んで新しいスクリプトファイルを開きます。

今から色々書いていきますが，これを記録しておくには，ファイルを保存する必要があります。`File > Save` と進むと保存できますので，ファイル名をつけて保存します。拡張子[*28]は `.R` にしておきましょう (小文字の `.r` でも結構です)。とくに思いつかなければ，`Lesson01.R` とでもしておいてください。

3.4.1 関数の利用

先ほど，`sqrt(16)` という関数の例をみました。この `sqrt` 関数はスクエアルート，つまり平方根を求めろというものです。何の平方根を求めるか，というと 16 という数字の平方根ですので，実行すると 4 という答えが返ってくるわけです。

このように，関数名で命令を指定し，その対象は小かっこ () で囲って渡すというやり方をします。$f(x)$ ということですね。括弧の中に入るものは**引数 (argument)** といい，今回のように目的となる数字を渡すこともありますが，そのほか関数を動かす上でのオプションを指定することもあります。また，1 つの引数だけではなく，複数の数字を渡したり，複数のオプションを指定することもあります。

たとえば関数がどのように動くか知りたいという時，ヘルプを開くのも関数で行います。`code:3.2` を実行すると，Help ペインにその説明が出てくるので確認してみてください。

code : 3.2 ヘルプのコード

```
1    # ヘルプも関数
2    help(sqrt)
```

3.4.2 パッケージの利用

R では関数を使ってさまざまな計算をこなします。R をインストールした段階でも，簡単な統計計算は十分できるようになっています。ですが，より専門的な分析やより高機能な情報処理をしようとするならば，**パッ**

[*26] 断言しますが，こちらはそのようには思っていません。教えるのは仕事ですし，簡単なことでも奥深いことがありますから，なんでも聞いてください。ちなみに，教えないのは教員から学生に対するアカデミックハラスメントです。

[*27] 生成 AI を利用するときは，こちらが望ましい答えや正解を知っている場合に限った方が良いでしょう。生成 AI の答えが合っているかどうかを自分でチェックできるのであれば，これはとても有用です。わからない課題を生成 AI にやらせてみると，合っているかどうかもわからないので，余計にひどいことになります！

[*28] 拡張子については，付録 A.5 を参照のこと。

ケージ (package) を利用することになります。たとえば，包丁一本で料理をするのは大変ですよね。包丁は確かに切ったり，皮を剥いたり，削ったり･･･と色々できることはできるのですが，皮を剥くのはピーラーで，細かく削るのはフードプロセッサーで，といった便利グッズを使えばもっと楽です。この「便利グッズ」が R でいうところのパッケージです。

R は豊富なパッケージが利用できることがその特徴で[*29]，パッケージは CRAN を通じて世界に公開されています。RStudio のパッケージペインを開くと `install` と `update` というボタンがありますが，`install` ボタンを押すと何のパッケージをとってくるか，入力できるようになります (図 3.11)[*30]。“Packages” とあるところに必要なパッケージ名を入力すると，自動的に CRAN にアクセスしてダウンロードが始まります。パッケージのインストールとダウンロードはその環境下では一度実行するだけで結構です。パッケージというのは関数群であり，関数を記載したファイルのダウンロードが終われば適切な場所 (Library) に保存されます。

ただ，ダウンロードしただけではすぐにその関数が使えるようになりません。ダウンロード後は，必要に応じて `library` 関数でパッケージを呼び出してやると，パッケージに含まれる関数が使えるようになります。購入したものを装備するかどうかは別，ということを頭の片隅に置いておいてください[*31]。また，パッケージの中には関係するパッケージを一まとめにしたパッケージというのもあります。インターネットへのアクセスは必要ですが，優れたパッケージを無償で入手できますので，便利なものをいくつか決めていれとくと良いでしょう。本講義で必要なパッケージは随時指定しますが，ひとまずは `tidyverse` パッケージ (Wickham et al., 2019) を入れておきましょう[*32]。

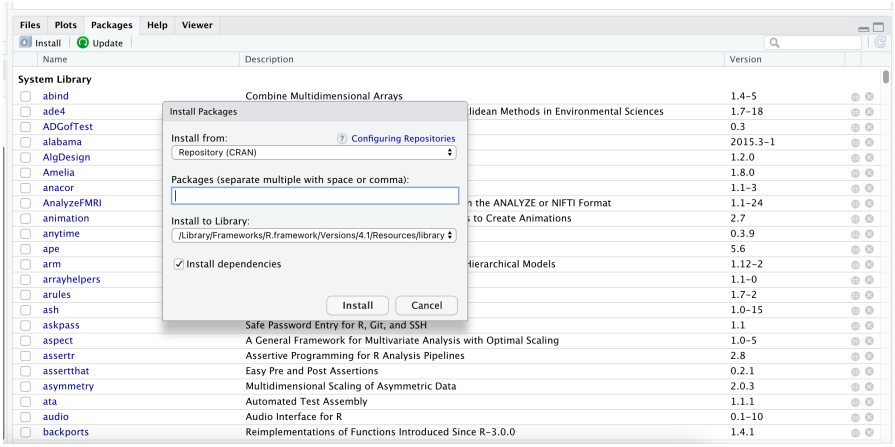


図 3.11 Package タブから Install をクリック

このインストールは，RStudio のパッケージペインからも実行できますが，`code:3.3` のようにコードを使って実行することも可能です。

code : 3.3 コードからインストールする

```
1  install.packages("tidyverse")
```

[*29] 本講執筆時点 (2021/02/05) で 289,680 件のパッケージが CRAN に登録されています。それ以外にも無数のパッケージが GitHub などで公開されています。

[*30] 出てくるボックスに示された，“Install From” が「どこからとってくるか」を示している欄になりますが，これは CRAN が自動的に選択されているので変更する必要はありません。

[*31] と同時に，同じパッケージを何度もダウンロードする必要はないということです。もっとも，R のバージョンが上がるとか，違う環境で実行する場合は再ダウンロードが必要なこともあります。

[*32] この `tidyverse` パッケージは，実はパッケージのパッケージですので，これをインストールしようとすると関連するその他のパッケージを次々と自動的に呼び出して環境を整えてくれます。初回は少し時間がかかりますが，気長に待ちましょう。大学の PC 室や計算機サーバなどには，すでにインストールが終わっているのでその作業は必要ありません。

3.4.3 オブジェクトの利用

R で扱う変数、関数、データセットなどにはすべて，名前がついています。なまえがついた「それ」のことをオブジェクトと呼びます[*33]。code:3.4 を書いてみてください。

code : 3.4 オブジェクトに代入

```
1    # オブジェクト
2    A <- 2
3    B <- 3
4    A + B
5    A
```

1 行目 コメント

2 行目 A と名付けたものに数字の 2 を代入する。代入の記号は小なり (<) とハイフン (-) の二文字でできていて，矢印記号 ← を表しています。RStudio の機能で，Mac の場合は option とマイナス記号，Windows の場合は Alt キーとマイナスを同時に押すと入力できます。。

3 行目 B と名付けたものに数字の 3 を代入しています。

4 行目 オブジェクト A とオブジェクト B を足し合わせています。

5 行目 オブジェクト A の中身を表示させろ (中身はなんだっけ)，という意味です。

実行結果は R **の出力** 3.3 のようになるでしょう。

R の出力 3.3: オブジェクトに代入する

```
> # オブジェクト
> A <- 2
> B <- 3
> A + B
[1] 5
> A
[1] 2
```

一行目は当然なんの回答もありませんが，2，3 行目を実行してもすぐに次のプロンプト (>) がでて指示待ちになります。これは「代入しろ」といわれたから代入したのであって，とくに反応を求めてないからです。次の 4 行目になって初めて足し算をせよ、と言われたのでその結果が 5 である，と返答しています。また，5 行目は A とだけ入力されました。このように指示待ちの時にそのオブジェクト名を入れると，その中身をすべて表示してくれます。今回は 2 がはいっているので，2 ですよ，と答えてくれています。

このようにオブジェクト名だけで計算を進めることが可能です。オブジェクト名はなんでもいいのですが，TRUE，FALSE，NA など R が特別な意味を持たせている言葉や，pi などの数学的意味がある言葉は予約語といって既に使われているので，そのままの方がいいでしょう。かといって，A, B,C といった無味乾燥な記号でもわかりにくいので，その時々に応じて任意の名前をつけてあげてください。

ちなみにこの「オブジェクトに代入」は基本的に上書きされるというルールです。さきほどのコードに続いて code:3.5 を書いてみてください。

[*33] 適当な日本語訳がありません。object はもの，対象という抽象的な意味なので，カタカナでオブジェクトとするしかないのです。

code : 3.5 オブジェクトは上書き

```
1  A <- 4
2  A + B
```

今度は結果として 7 が表示されたと思います。先ほどは A に 2 を代入せよといい，そのあとで A に 4 を代入せよと命じたので，いま A は 4 に上書きされています[*34]。この基本的に上書きされる，ということはよく注意しておいてください。というのも，これから長いコードを書くことがあったときに，部分的に実行を繰り返すと上書きに上書きを繰り返すこともあり得るからです。そうすると，ある時うまくいったけど，次に同じことをしようとしたらうまくいかなかった，ということにもなりかねません。くどいようですが記録は再現性の担保のためにやっているわけですから，一時的な記憶やたまたまうまくいった例に依存しては困ります。環境が空っぽの状態，クリアな状態で一行目から順に走らせて，何度やっても同じ答えが出るようにしましょう。

環境をクリアにするには，Environment タブのバーにある箒のアイコンをクリックします。これで R の頭の中が空っぽになります。あるいは `rm(list=ls())` というコードを実行するとクリアされます。スクリプトの一行目にこれを書いておくことをお勧めします。

また，R は終了すると，それまでの作業を保存しますから，同じプロジェクトをもう一度開くと以前の記憶が蘇ります。これが便利なのですが実は少し厄介で，学生さんが「自分の環境ではできた」といっても私の手元で再現できない (からレポートやり直し) ということがあります。きっと記録されているスクリプトの外で，なんらかの作業がなされているのです。清書できたスクリプトは，R の頭の中を空っぽにしてから再チェックすると良いでしょう。また，`Tools > Global Option` から，Workspace の “Restore .RData into workspace at a startup”(起動時に.RData からワークスペースの情報を復元する) のチェックを外し，“Save workspace to .RData on exit”(終了時に.RData にワークスペースの情報を保存する) を “never”(絶対そんなことしない) にしておくことで，いつでもクリーンな環境から始められます (図 3.12)。

オブジェクトの型 1・ベクトル

`code:3.6` を実行してみましょう。

code : 3.6 ベクトル処理

```
1    obj1 <- 1:10
2    obj1 + 2
3    obj2 <- c(2,3,4,1,2,3,4,5,6,7)
4    obj1 + obj2
```

1 行目 `obj1` オブジェクトに “1:10” を代入

2 行目 `obj1` オブジェクトに 2 を加える

3 行目 `obj2` オブジェクトに “c(2,3,4,1,2,3,4,5,6,7)” を代入

4 行目 `obj1` オブジェクトに `obj2` オブジェクトを加える

一行目で代入している “1:10” は R 独特の表記法で，コロンで繋いだ数字は連続した数字を意味します。ここでは 1 から 10 までの連続した数字，すなわち $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ です。このように数字のセットを一度に扱うことができます。このオブジェクトに 2 を加えるというのは，各要素に 2 を加えることと同じですので，そのような結果が出力されていますよね。また三行目で代入しているのは $2, 3, 4, 1, 2, 3, 4, 5, 6, 7$ という

[*34] オブジェクトの中に何が入っているかは，Environment タブをみるとわかることがあります。今見ると，A に 4，B に 3 が入っていることが示されているでしょう。

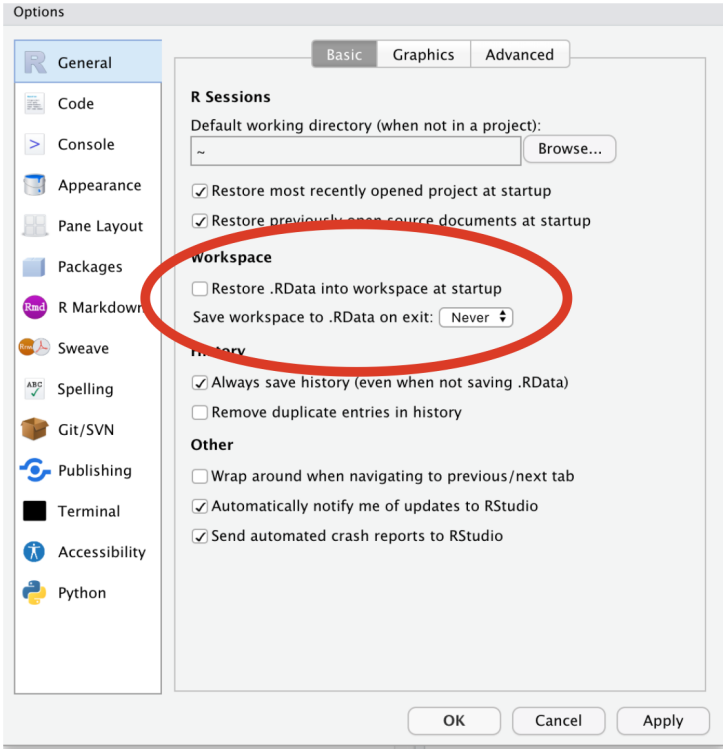


図 3.12　クリーンな環境をいつでも保つための設定

数字列です。これを c 関数で括っています。この c は一文字ですが関数で[*35]，数字をセットにするという働きをします。そして四行目で，対応する各要素を足し合わせるという計算を行っています。

　このように，数字をセットで扱うことができるのが電卓とは違うところです。統計ではたくさんのデータを扱うわけですから，この機能は基本中の基本になります。ちなみに，数字のセットのことは数学でベクトルというのでした。今回のオブジェクトは数字をベクトルとして保管していることになります。

オブジェクトの型 2・リスト

　データセットは数字だけではなく，文字列などを含むことがあります。そうした対象をまとめて扱うのが**リスト (list)** という形式です。つづいて code:3.7 を実行してみましょう。

code : 3.7　リスト型

```
1    obj3 <- list(
2      name = c("Taro", "Hanako", "Ken"),
3      gender = c("male", "female", "male"),
4      height = c(170,160),
5      weight = 55
6    )
7    obj3
8    obj3$name
```

1 行目　obj3 オブジェクトに代入するのは list であることを，list 関数を使って明示しています。この関数の中身は複数行にわたり，括弧が閉じられるのは 6 行目になりますので，そこまでで 1 つの命令文

[*35] combine や concatenate の略語と言われています。

だということになります。

2 行目　name という名前のついたベクトルを作ります。

3 行目　gender という名前のついたベクトルを作ります。

4 行目　height という名前のついたベクトルを作ります。

5 行目　weight という名前のついたベクトルを作ります。要素が 1 つでも，要素 1 のベクトルと R は解釈します。

6 行目　1 行目から続く `list` 関数の終わり。リストの中に含まれるベクトルは間まで区切られていること，対応する括弧がきちんと閉じられていること，が重要です。

7 行目　オブジェクトの中身はなあに，ということで中身を表示させています。

8 行目　`obj3` オブジェクトの name 要素だけにアクセスしてみました (要素を見せて，と命じている)。

出力された画面を見て，何が起きているかを確認してみてください。コードの 2 から 6 行目に当たるところでは，コンソールの左端が>ではなく + になっていると思います。これは前の行から続いて入力されている，つまり入力がまだ終わっていないことを意味する記号になっています。また 7,8 行目については，表示させる指示をしているので，その結果が返ってきているはずです。

このように，名前 (文字列) や性別 (名義尺度水準)，身長や体重 (比率尺度水準/量的変数) などをひとまとめにして `obj3` というオブジェクトに保管する，ということができています。このオブジェクトのある変数だけ参照したい，というときは，オブジェクト名の背後にドルマーク (`$`) をつけて繋げます。RStudio で入力してる時は，`obj3$`と書いただけで「中身はこれですが」という候補のリストが示されたと思います。このように入力補完をしてくれるのも RStudio のいいところです。

ところでこれがデータセットだとすると奇妙ですね。名前や性別は 3 つありますが，身長は 2 つ，体重は 1 つしかデータがありません。普通こんなことはなくて，1 つのケースについて複数の変数があるので，縦に個体 × 横に変数の長方形型をしているデータセットが一般的になるはずです。リスト形式は各要素の長さに関係なく，なんでもまとめて置けるので便利なのですが，データ分析においては長方形であるという形の制約があって良いはずです。この制約を持ったリスト形式のことを**データフレーム (data.frame)** 型と言います。

オブジェクトの型 3・データフレーム型

データフレーム型にするには，`data.frame` という関数を使います。`code:3.8` を実行してください。

code : 3.8　データフレーム型

```
1    df <- data.frame(
2      name = c("Taro", "Hanako", "Ken"),
3      gender = c("male", "female", "male"),
4      height = c(170,160,155),
5      weight = c(55,60,43)
6    )
7    df
8    df$name
```

1 行目　`df` オブジェクトに代入するのは `data.frame` であることを，`data.frame` 関数を使って明示しています。

2 行目　name という名前のついたベクトルを作ります。

3 行目　gender という名前のついたベクトルを作ります。

4 行目 height という名前のついたベクトルを作ります。要素数を合わせます。

5 行目 weight という名前のついたベクトルを作ります。要素数を合わせます。

6 行目 1 行目から続く `data.frame` 関数の終わり。ここに含まれるベクトルは間まで区切られていること，対応する括弧がきちんと閉じられていること，が重要です。

7 行目 オブジェクトの中身はなあに，ということで中身を表示させています。

8 行目 `df` オブジェクトの name 要素だけにアクセスしてみました (要素を見せて，と命じている)。

コンソールには R **の出力** 3.4 のように表されていると思います。

R の出力 3.4: オブジェクトに代入する

```
> df <- data.frame(
 +     name = c("Taro", "Hanako", "Ken"),
 +     gender = c("male", "female", "male"),
 +     height = c(170,160,155),
 +     weight = c(55,60,43)
 + )
> df
 name gender height weight
1   Taro   male    170     55
2 Hanako female    160     60
3    Ken   male    155     43
> df$name
[1] "Taro"   "Hanako" "Ken"
```

このオブジェクト，`df` の出力を見るとわかるように，行に個人，列に変数の長方形方になっています。変数にアクセスする時は，list 型の時と同じようにオブジェクト名の背後に`$`をつけて，`df$name` のようにします。

今回はひとまずデータを作るために，各ケースの値を入力して行ってもらいましたが，実際に統計解析をする時は表計算ソフトで素データを入力し，ファイルから R にデータを読み込むことが一般的です。これについては第 4 講で説明します。

3.5 再現性のある研究実践へ

さて，これで演習の第一段階は終了です。R のスクリプトを書いてみて，いかがだったでしょうか。思った通りに動かない，とイライラしたひともいるかもしれません。その通り，**コードは思った通りに動くのではなく，書いた通りに動く**のです。正しく指示し，正しい答えを得るために，実行記録が残るということは科学的営みとして重要なことなのです。

今後，みなさんはパソコンを使って文章 (レポートや論文) を作成するという機会が増えてくることは間違いありません。文字を入力するだけであればスマートフォンでもいいですが，レイアウトを整えるためにはパソコンが必要になってきます。文章だけでなく，図表も作成することになりますし，図表を作るのに必要な統計的データの整理，分析もパソコンを使って行います。とくに図表や統計分析は，プログラムを書いて行うことになります。機械は 0 と 1 の羅列で情報処理をしますが，その文字列を人間が考えるのは大変です。機械にやって欲しい操作，計算を人間がわかる言葉で書いたものをプログラムと言います。機械と人間を繋ぐ言語だと思ってください。この言語には色々な種類がありますが，基本的に統計環境 R という言語を使えるようになっ

てください[*36]。

計算だけでなく図表を書くのも，プログラミングすることになります。なぜそんな面倒なことをするのでしょう？ 表計算ソフトを使うと，誰でもマウスでカチカチっとすれば簡単にグラフが作れるのに，と思った人もいるかもしれません。その答えは端的にいうと，そのやり方だと記録が残らないからです。同じデータに同じプログラムをあてがえば，別の計算機上でも結果が同じになります。この誰がどこでやっても同じ結果になる，というのが科学の営みとして重要なことであるのは第 1 講で説明した通りです。また，教員と研究指導のやりとりをする上でも，プログラムに書き起こされていることは利点があります。マウスでの動きを口頭で説明するのは難しいですね。さらに自分がどういう設定・環境で実行したかを完全に意識していない場合，「なんだかわからないけど，カチカチやっているうちにこの結果が出ました」ということにもなりかねません。そういう報告を受けた教員の気持ちも想像してみてください。指導のしようがありません。

プログラムで実行するというのはすなわち，何をどのように操作するのかを書き下すことでもあります。すごく複雑な統計処理も，中身は 1 つ 1 つの「計算」の積み重ねです。何がどのように積み重なっているかをしっかり理解しなければ書けません。逆に言えば，書いてあるものを見ればどのように組み立て，どのような実行をしているのかがわかります。うまくいかないプログラム，つまりエラーがおきるプログラムであっても，プログラムされたコード (スクリプトとも言います) を見せてもらえれば，どこにエラーが生じているかを明らかにできます。研究の実践とその成果が「動作」ではなく「指示書」として記録に残ることは，研究の再現性を高めるためにも重要な営みなのです。

教育や学習の上でも利点はあります。プログラムを書くときに必要なのはこのアルゴリズム的思考，すなわち目標に向かうための必要な手続きを 1 つ 1 つ特定し，順序立てて実行していくという考え方です。お料理のように，「玉子焼き」というゴールに到達するために「卵を取る」「卵を割る」「卵をかき混ぜる」「フライパンを温める」... といった各手続きを順序立てて実行できなければなりません。これから先，統計の話が複雑になってくると，わからない！と音を上げたくなることが出てくるかもしれません。そんなときは (たっぷり休憩して心身ともに元気な状態で落ち着いて) 見直すことで，どこからどこまで理解していて，どこから理解できていないのか，確認できます。プログラムのエラーの場合，どこからどこまでが正常に動いていて，どこから計算がおかしくなるのかを見ることができるのです。諦めなければ，確実に積み重ね，前進できます。エラーが出ると言いましたが，エラーはたいていの場合エラーメッセージとともに明らかになります。このメッセージをよく読めば，解決のいとぐちが掴めるのです。わからなければエラーメッセージとともに，教員に相談してください。どこでどのようなエラーが出たかをじっくり確認すれば，何が間違いで何がわかっていなかったのかがわかるというものです。エラーを 1 つ 1 つ無くしていき，うまく動くプログラムが書けたときは嬉しいものですよ。

最後に，プログラムを含めて記録の取れる研究実践について少しだけ説明しておきます。科学は「自分だけがわかっている」という私秘性を排除した公共的な営みである，ということはすでにお話しした通りです。個々人の興味関心によって研究を積み重ねていくのですが，研究の結果明らかになったことはオープンな議論の場に出すべきです。なぜなら，自分が知りたいと思って始めた研究について，他の人から自分では気づかなかった問題点を指摘してもらえるかもしれないし，もっとおもしろいテーマに展開できるかもしれない，あるいは他の人と一緒に研究しようということになるかもしれないからです[*37]。自分の研究も他人の研究の成果を踏まえて，あるいはそこからインスパイアされて進めることになりますが，この研究のスタートとなるこれまでの成果に間違いや嘘が含まれていては困りますね。人間を対象にした研究の場合は，間違いや嘘がまったく含まれていなくても完全に同じ状況や同じ対象に研究するというのが難しいので，先行研究の結果が再現でき

[*36] 2024 年度も，専修大学心理学科で公式に使う統計環境になっています。

[*37] 間違いを指摘されたら恥ずかしい，などと思ってはいけません。間違いは正せばいいだけですし，間違いが直るとより良いものになっていくだけだからです。小中高の学校教育とは違って，正解がある問題を考えているわけではないので，教員はもちろん周りの人の目を気にする必要はまったくありません。そんなことでからかう人がいないのが大学という界隈の楽しいところです。

ないという問題があります[*38]。

そこで，何をどのように研究したか，どういうデータをとってどのように分析したか，ということを事細かく記録しておく必要があります。最近では，研究を始める前に事前にどのような研究計画で進めていくかを公開し[*39]，とったデータや分析プログラムも公開するということが行われています。こうした研究計画やデータ，プログラムの公開は Open Science Framework というウェブサイト[*40]が今最もよく利用されています。ここでファイルを共有すると，いつどのようなファイルが置かれたか，どのように更新されたかという記録が残るので，記録の改竄などを疑われるような懸念がなくなるという利点もあります。

皆さんも，ごく初心者とは言え心理学という科学の営みをするメンバーの一人になったわけですから，自分の記録やファイルを管理し，公開するという意識を持って取り組むようにしてください。

3.6 課題

■自分の電子機器のスペックを調べる 皆さんが使っているパソコンの，CPU のメーカー，型版はどういったものですか。一次記憶装置はどれぐらいの容量がありますか。二次記憶装置の容量はどうですか。いま二次記憶装置の空き容量はどれぐらいですか。調べて回答してください。

■自分のファイルの場所を調べる 同じくパソコンのユーザ名は何になっていますか。また何か適当なファイルを作り，どこかに保存し，「どこにどのようなファイル名で保存したか」を報告してください。拡張子をつけることを忘れずに。

■R で遊んでみよう 今日実行した R のコードはごく簡単な計算ばかりです。数字を色々書き換えて遊んでみましょう！

[*38] 実際，社会心理学の教科書に載っているような研究は再現率が 30% しかない，という批判もあります。詳しくは池田・平石 (2016) などを読んでください。

[*39] 研究をする前に計画だけの論文をチェックしてもらう方法もあります。プレレジストレーション，略してプレレジと言います。

[*40] https://osf.io/

第 4 章

統計環境 R の利用

4.1 Rmarkdown による文書作成

前回は R に計算させたりデータを入力したり，と統計的な処理をいろいろさせてきました。RStudio は総合開発環境として，こうした R の計算機能に加えて文書作成機能も持っています。ここでは RStudio を使った文書作成を解説したいと思います。

表計算ソフトと文書作成ソフトは別々のものですがセットで販売・利用されることが多いものです。これらをまとめてオフィスソフトということがあります。Microsoft が提供するのは Word, Excel, PowerPoint などの Office シリーズですし，Apple が提供するのは Pages, Numbers, Keynote などです。他にもフリーソフトウェアとして LibreOffice シリーズがあり，そこでは Writer, Calc, Impress Presentation という名称になっていますが，何も文書作成，表計算，プレゼンテーションのアプリです。こうしたものが一般的に提供されているので，表計算ソフトで数値計算し，文書ファイルで報告書を書いて，プレゼンテーションアプリで紙芝居形式の発表をする，というのがよく行われています。これから皆さんもたくさんこうした資料を作ることになると思いますが，作成に当たっては，表計算ソフトで図表を作って，それをコピー＆ペーストして文書ファイルやプレゼンテーションファイルに切り貼りする，という方法を取ることでしょう。これが短い資料であれば問題ないのですが，長い資料や多くのデータになってきたりすると，貼り付ける図を間違えるとか，数字の修正をする前の図版だった (更新されたグラフにアップデートし忘れた) といった問題が生じやすくなります。これを**コピペ汚染**と言ったりします。

文書作成とそこに含まれる図表や数字の計算が，同じ環境で行われればこの問題は無くなります。RStudio はそれを実現してくれる環境です。計算環境は R が持っていますから，それを文書に組み込む仕組みがあれば良いのです。RStudio では Rmarkdown という書式を使ってこれに対応します。Rmarkdown は Markdown といわれる文書形式に R を組み込んだ RStudio 独自のファイル形式です。メニューから `File > New File > R Markdown` とすすむと，最終的にどの出力形式のファイルが必要か聞いてくるところがあります (図 4.1)。ここにあるように，最終的に HTML，PDF，Word といったファイル形式で出力ができます[*1]。ここではひとまず HTML を選びましょう。Title には文書タイトルを，Author には作者名を入れますので，ここも適当なタイトルと皆さんの名前を入れて OK を押してください。

すると Rmarkdown 形式のファイルが開かれます。すでに何か書き込まれたファイルになっていますが，これは書き方の例が書いてあるのです。図 4.2 に最初の数行をキャプチャしたものを示しました。皆さんの手元の環境でも同じになっているか，確認しながら以下の解説を読んでください[*2]。

このファイルの 1〜6 行目は，文書全体の設定をしているところです。ここは基本的に変更する必要はあり

[*1] PDF 形式に出力するには，TeX という別のシステムがまた必要ですが，これも RStudio からインストールすることができます。

[*2] もちろん Title と Author はみなさんが書いたものに変わっているはずです。

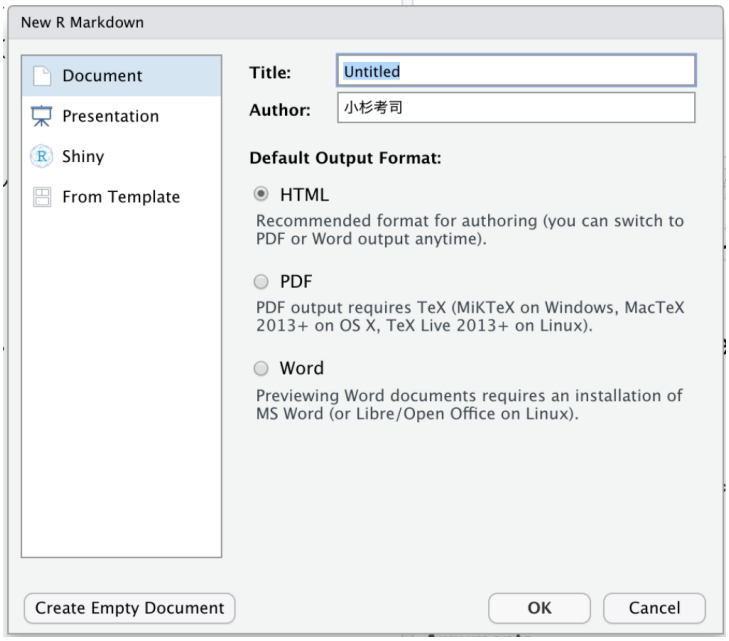


図 4.1　R Markdown からの最終出力を選択する

図 4.2　R Markdown からの最終出力を選択する

ません。1 行目，6 行目には---とありますが，この区切り記号が重要な意味を持ちますので，変更しないようにとくに注意してください。2 行目はタイトル，3 行目は作者，4 行目は作成日，5 行目は出力形式が書いてあります。これらはさきほどの図 4.1 で指定したものが反映されているはずです。これらは変更しても結構ですが，書式ルールから外れると後々うまくいかなくなるので注意してください[*3]。

7 行目からが本文です。よく見ると，背景が灰色のゾーンと白色のゾーンがあると思います。灰色のゾーン

[*3] chank の実行ボタンが表示されない，といったエラーは誤ってこの YAML 部分の最後の---を消してしまっており，Rmd ファイルとして認められないから，ということがあります。消しすぎにご注意！

(8〜10 行目，18〜20 行目など) は**チャンク (chunk)** と呼ばれる R のコードを書く領域です。白色のゾーン (11〜17 行目，21〜25 行目など) は文書，本文を書くところになります。7 行目以降はサンプルの文章が入っていますから，すべて削除してしまっても構いません。実際に皆さんがレポート等を書く時は削除しましょう[*4]。でも今いったんそのままで，ファイルの上にある `knit` というボタンを押してみましょう。これは**ニット (knit)** と言い，文章を編み上げるという意味です。このファイルが未保存の場合は，保存するファイル名を指定する画面が出ますが，任意の名前にして実行してみましょう[*5][*6]

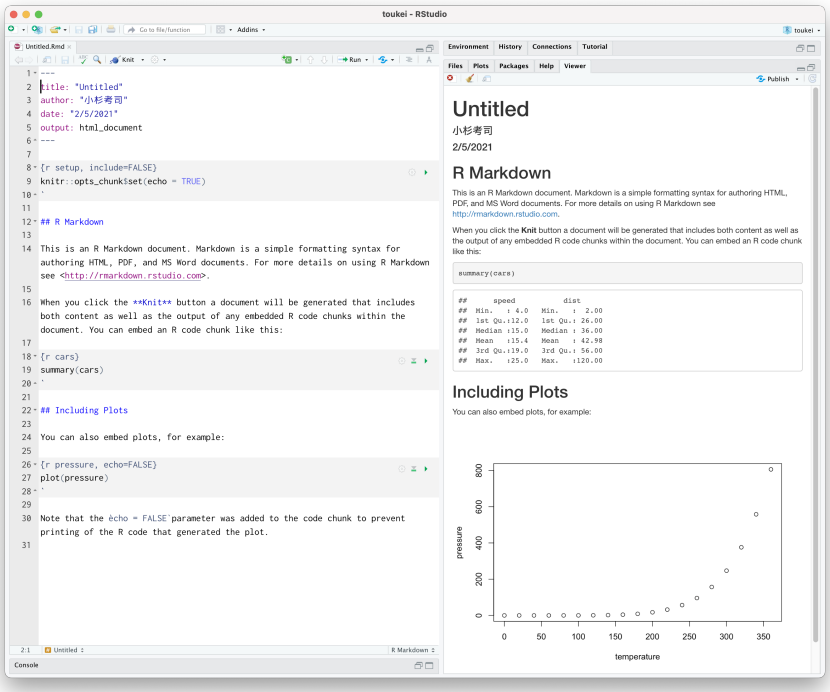


図 4.3　R Markdown ファイルと Knit されたものの比較

図 4.3 には，Rmarkdown のファイルと出力結果を並べてみました[*7]。対応関係がわかりますでしょうか。設定のところで入力したタイトルが大きく表示され，ついで作者名，日付が示されています。その後で本文が続きます。

出力の方では本文が R Markdown から始まっています。これはもとのファイルの方では##で示されたところに対応していますね。Rmarkdown ファイルでは，#記号はコメントアウトではなく，見出しの階層を示しています[*8]。##はレベル 2，###とするとレベル 3 と，#の数で階層の深さを表現しているのです。階層が低いほど文字が大きく表示される仕組みです[*9]。

[*4] 後ほど指示しますが，本講の課題は Rmarkdown 形式のファイルを提出してもらいます

[*5] 初めての場合，RStudio が `knit` に必要な関連パッケージをインターネットからダウンロードしようとします。ダウンロードしても良いか? と聞いてきますので OK を選択してください。

[*6] 保存先はプロジェクトフォルダの中にね！ Rmd は Knit の時にファイルの置かれた場所が Working Directory だと判断しますので，別のところにおくとファイルが見つからないエラーが出ます。

[*7] knit がうまく完了していれば，Viewer ペインかブラウザに表示されているはずです。

[*8] あれ，同じように入力しているのに見出しの所の色が変わらない (=強調表示されない)，という人へ。#マークの後ろに半角スペースを入れてください。区切りがないと，R は記号として認識できないんです！

[*9] レポートの書き方についてはこの授業で教える範囲を超えますので詳細は他の授業に譲りますが，アカデミックライティングの基本は構造化された文章であることは踏まえておいてください。科学論文や学術的な文書というものは，段落＞見出し＞小見出しといった階層構造を持っています。とくに心理学論文の場合は，段落が「問題」「方法」「結果」「考察」の 4 つであることが決まって

見出しのついてない文章に関しては，そのまま本文として出力されています。注目すべきは，灰色の背景を持った R の領域の箇所です。たとえば元ファイルの 19 行目には，`summary(cars)` と書かれていますね。これは R のコマンドで，`cars` というオブジェクトの要約を，`summary` 関数を使って表示せよ，というものになっています。**本文にはこの命令文しか書いてありませんが**，右の結果を見ると灰色の背景でどのような命令がなされたかを示すと同時に，結果が表示されています (speed 変数と dist 変数の記述統計量が表示されています)。また，元のファイルの 27 行目では `plot(pressure)` というコードが書かれています。これは `plot` 関数によって図を描けという命令文なのですが，これもファイルの方に反映されているはずです。

このように，計算結果をコピー＆ペーストで文書に書き写すのではなく，明示するのは計算手順だけでよいのです。このようにすることで，この出力された文書を受け取る側は，どのような計算がなされ，どのような結果になるかを確認できます。そしてそこに**コピペ汚染**は生じるはずがないのです！これは科学論文を作成するときに，方法を明示することと間違いがないことを担保しているという意味で，大変便利で画期的な方法であることがおわかりいただけたかと思います。

■自分で書いてみる　それでは元のファイルの 7 行目以下をすべて削除し，自分なりに Rmarkdown を書いてみましょう。`code:4.1` のように書くのを目標とします。このとき，R のコードを書くチャンク領域を挿入するには，図 4.4 に示した上のボタンを押して，R を選択します。`code:4.1` に書いてあるように，バッククォーテーションを 3 つ繋げて，中括弧で r と書き (```` ```{r} ````)，バッククォーテーションで閉じるという方法でも結構です。

チャンクには，設定や一時的な実行を行うためのボタンが表示されます (図 4.5)。実行ボタンを押すと knit ではなくその下に実行結果が表示されるので，自分の意図した結果が示されているか，確認できます。

code : 4.1　Rmarkdown を書いてみる

````
1  ---
2  title: "れんしゅう"
3  author: "小杉考司"
4  date: "2/5/2021"
5  output: html_document
6  ---
7
8  # 計算をする
9
10 ## 四則演算
11
12 ### 足し算
13
14 足し算をします。
15
16 ```{r}
17 2+3
18 ```
19
20 ### 引き算
21
22 引き算をします。
23
24 ```{r}
````

いて，独自のスタイルに勝手に変えることはできません。

```
25  5-2
26  ```
```

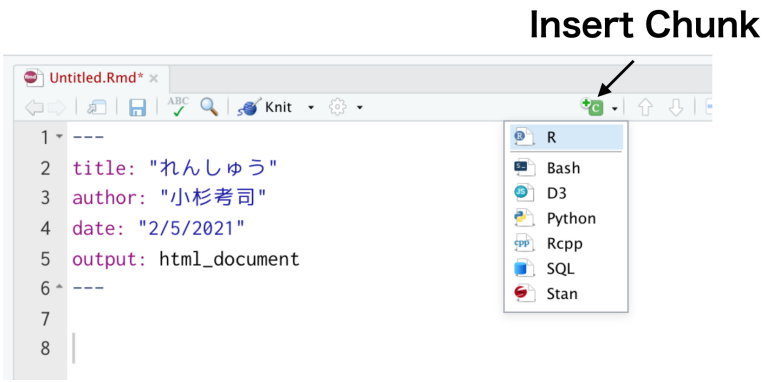


図 4.4　R チャンクを挿入するボタン

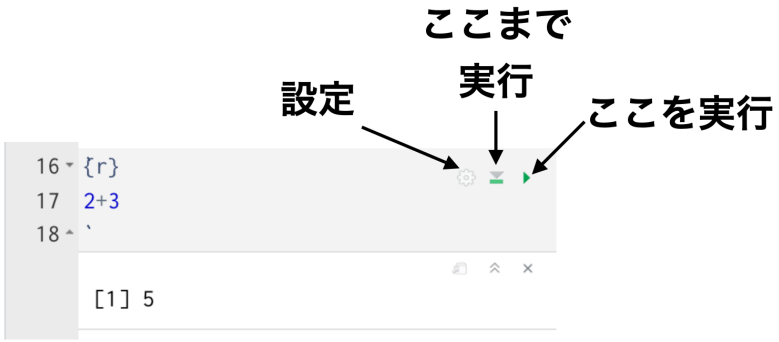


図 4.5　チャンクの設定や実行

うまくできましたでしょうか。コードや設定に間違いがあれば knit は中断してしまいますが，エラーメッセージを読んで対応しましょう。RMarkdown について，詳しくは高橋 (2018) が一冊を使って解説していますし，松村他 (2021) や小杉 (2019) なども参考になります[*10]。

Rmarkdown を使うと，R スクリプト (計算の指示書) とレポートの文章を 1 つにまとめ，knit することで欲しい形式に出力できることがわかりました。ここで，スクリプトに描かれているのは計算の指示書だけであり，データをどのように加工するかの手続きが記載されています。言い換えると，データを変更しても計算手続きだけは決まった形で進めることができるのです。

もし定期的に同じようなデータが上がってきて，何度も同じようなレポートをしなければならない，というような仕事があったとします。これは仮定の話ではありますが，実際は社会に出て，毎月，毎年の集計をしなければならないようなことはよくあることです。この時いつも同じ計算をマウスでカチカチと繰り返しているのは，馬鹿らしいことですね。同じ手続き，同じ文章で数字だけ違うのであれば，データだけ変えて同じ指示書に流し込んでやればいいのです。Rmarkdown はまさにそういう時に向いています。

[*10] 専修大学の国里ゼミと小杉ゼミでは，卒論を Rmarkdown で書くことを推奨しています。再現性の観点から便利であることはもちろん，文書の規定に合わせたり引用文献リストを自動生成させたりする，専修大学専用パッケージ `senshuRmd` を共同開発しており，これを利用します。

もちろんそれだけでなく，計算手続の記録をつけるという意味で科学的にも重要だという話でしたね。研究実践の例として，データが手に入ったらデータファイルは加工せず，加工の部分はすべて R スクリプト/Rmarkdown でやることで計算結果の捏造が起きていないかどうかをチェックできるのです。研究で得られたデータは公開することで，いろいろな人にチェックしてもらおうというオープンサイエンスの動きも広がっています。

4.2 外部データの読み込み

実際の研究のシーンでは，データ入力は表計算ソフトを使うことが一般的です[*11]。データの入力についてもお作法がありますが，それについてはセクション 4.5.1,Pp.74 を参照してください。入力は商用ソフトを使ってもいいですが，それを分析するのは R のお仕事です。

データファイルが外部にある場合，それを R の中に取り込む必要があります。外部データの取り込みは，分析を始める前の手続きではありますが，初心者がトラブルに巻き込まれやすいシーンでもあります。手順を説明していきますので，注意深く読んでおいてください。

4.2.1 データファイルを読み込む

さて，では練習用のデータセットを使って，これを読み込むところから分析を始めてみましょう。授業シラバスのサイト[*12]から「野球選手のデータ 10 年分」をダウンロードし，**プロジェクトフォルダに保存してください**。

このデータセットはプロ野球データ Freak[*13]さんからスクレイピングしてきたものであり，2011 年から 10 年間のプロ野球選手のデータが入っているものです。データの中身については表 4.1 を参照してください。

さてこのデータファイルを読み込むには，読み込むための関数を使います。R が持っている基本関数である `read.csv` 関数を使っても読み込めるのですが，ここではより便利なパッケージの関数を使ってみましょう。

ここでは Rmarkdown ではなく，プレーンな R スクリプトファイルをつくって話を進めます。作業を始める前に，データファイルがプロジェクトフォルダの中にあること，そもそもプロジェクトを開いていることを確認してくださいね[*14]。その上で，`FIle > New File > R Script` から新しくファイルを開くか，先ほどの R スクリプトファイルの下に `code:4.2` のコードを書いて実行してください。

code：4.2 CSV ファイルを読み込む

```
1  library(tidyverse)
2  dat <- read_csv("BaseballDecade.csv",na = "NA")
```

1 行目 `tidyverse` パッケージを読み込む。次の `read_csv` 関数がこのパッケージに含まれているからです。

2 行目 `dat` オブジェクトに，`read_csv` 関数で読み込むファイルを代入します。関数の第一引数はファイル名，`na` は欠損値の指定で，`NA` という文字列を欠損値として扱うことを宣言しています。

[*11] 流石に数字の入力は表計算ソフトの方が有利で，統計ソフトの R をつかってデータを入力するのは難しいです。

[*12] https://kosugitti.github.io/psychometrics_syllabus/

[*13] https://baseball-data.com/

[*14] プロジェクトを開くことで，R のカレントディレクトリ (現在作業中のディレクトリ。ワーキングディレクトリ，ワーキングフォルダとも言います) が当該フォルダの中にセットされます。これがセットされているので，ファイルの読み込みにあたってファイル名だけの指定で済むようになります。プロジェクトを開いていないとか，違うフォルダのファイルを参照したい場合は，ファイル名のところに相対パスを指定する必要があります。ファイルの場所の指定については，付録 A の Pp.399 を参照してください。

このようにすることで，外部ファイルを読み込むことができます。ここで示した tidyverse パッケージは，R におけるデータ加工のためのパッケージです[*15]。作ったデータを分析しやすいように加工することを**データハンドリング (Data Handling)** といいます。次のセクションでは簡単にデータハンドリングの方法と，どのように加工すると良いのかについて説明します。

4.2.2 データハンドリング

ここではデータハンドリングの話をします。データハンドリングとは，データを加工するという意味です。進んだ統計的な分析などをする場合も，まずは元のデータを分析しやすい形に整える作業が必要なのです。面白いもので，このデータハンドリングだけについて書かれた本もあります。松村他 (2021) がそうで，どんなデータをどのように分析するか，について触れることなく，どのようにデータを加工するか，加工しやすくするための準備をするか，ということに特化しています。手元に一冊置いておきたいおすすめの一冊です。

ここではよく使われる基本的な関数を抜粋して，ご紹介します。

■tidyverse パッケージ　ここでは tidyverse パッケージと呼ばれる，データハンドリング専門パッケージの使い方を説明していきます。このパッケージは，パッケージをひとまとめにしたパッケージです。つまり便利グッズのパッケージ，「初心者お料理セット」みたいなものですね。

インストールは前回 (セクション 3.4.2,Pp.49) で既に終わっているでしょうから，これを読み込んでハンドリングしてみましょう。ダウンロードとインストールは同じ環境で 1 回やれば終わりますので，起動するたび，実行するたびに毎回行う必要はありませんが，実行に際してはパッケージを**実装**しなければなりません。ダウン

表 4.1　サンプルデータファイルの構造

変数名	型	内容
Year	文字列	何年度のデータか
Name	文字列	選手名
team	文字列	チーム名
salary	数字	年報 (推定)
bloodType	文字列	血液型。A,B,O,AB 型の 4 分類
height	数字	選手の身長
weight	数字	選手の体重
UniformNum	数字	背番号
position	文字列	守備位置。投手，捕手，内野手，外野手の 4 分類
Games	数字	出場試合数
AtBats	数字	打席に立った回数。打数
Hit	数字	ヒットを打った回数
HR	数字	ホームランを打った回数
Win	数字	勝ち投手になった回数。投手のみ
Lose	数字	負け投手になった回数。投手のみ
Save	数字	セーブ数。投手のみ
Hold	数字	ホールド数。投手のみ

[*15] ファイルが読み込めない，文字化けするなどのエラーについては，付録 B.2,Pp.409 を参照のこと。

ロードとインストールはグッズの購入であって，実装とはそのグッズを身につけるということですね[*16]。実装する関数は `library` で，実行すると図 4.1 のように表示されます。

R の出力 4.1: tidyverse パッケージの読み込み

```
> library(tidyverse)
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.3     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package to force all conflicts to become errors
```

ここに示されているのは，`dplyr`，`readr`，`forcats`，`stringr`，`ggplot2`，`tibble`，`lubridate`，`tidyr`，`purrr` というパッケージを読み込みましたよ，ということと，読み込んだ `dplyr` パッケージの `filter` 関数と `lag` 関数が他のパッケージと干渉 (conflict) しているので，上書きしましたよということです。基本的に気にしなくてよい情報です[*17]。

ところで，先ほどの `read_csv` 関数を実行すると図??のような出力が出てくるかと思います。赤い文字なのでエラーだ！と思って驚くかもしれませんが，よくみてください。6546 行 (Rows)，17 列 (Columns) とあるように，データのサイズが表示されています。続く表示も，区切り文字がカンマであること (`Delimiter: ","`)，文字列の変数と数値 (dbl) の変数[*18]であることなどがでています。これはエラーではなくて，正しく読み込めているかどうか確認するためのメッセージなのです。親切！

[*16] ドラゴンクエストというゲームでは，武器や防具を武器屋で買いますが，買って持っているだけでは身につけたことになりませんし，当然その効果も発動しません。購入した武器や防具を「身に着ける」というアクションが必要です。武器屋で購入すると，店主が「ここで身につけていくかい」というセリフで注意喚起してくれます。

[*17] とは言え気になる人のために。パッケージは数多くありますから，あるパッケージ A で使う関数と，あるパッケージで使う関数が同じ名前だった，ということがあり得るわけです。これが「名前空間の干渉」というやつで，これを解消するためには，A パッケージの関数 f だよ，B パッケージの関数 f だよ，ということを明示してやる必要があります。R ではコロンを 2 つ繋げて，`A::f()` とすることで A パッケージの関数 f を表しています。ここでは `dplyr::filter()` が `stats::filter()` をマスクしましたよ，ということを意味しており，パッケージの指定なしで `filter()` と書くとこの後は `dplyr()` パッケージの `filter()` を使いますよ，ということです。このあとも `stats` パッケージの `filter()` を使いたい場合は，`stats::filter()` とすれば OK です。ちなみに `stats` パッケージとは，R を起動した段階で読み込まれている基本中の基本パッケージです。

[*18] `dbl` はダブル double，つまり倍精度の長さを持つ実数変数だということです。要するに数字です。

R の出力 4.2: 外部データの読み込み時メッセージ

```
> dat <- read_csv("BaseballDecade.csv",na = "NA")
Rows: 6546 Columns: 17
── Column specification ────────────────────────────────────────────────────────
—
Delimiter: ","
chr  (6): Year, Name, team, bloodType, UniformNum, position
dbl (11): salary, height, weight, Games, AtBats, Hit, HR, Win, Lose, Save, Hold

☒ Use `spec()` to retrieve the full column specification for this data.
☒ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

さあでは tidyverse パッケージをつかってデータハンドリングしてみましょう。今からいくつかの関数を紹介します。それらはレゴブロックの 1 ピースのようなもので,「これだけ渡されてもどうしろと」と思うかもしれませんが, これらを組み合わせると大きなことができる・・・！と思ってお付き合いください。

■列選択：select　まずは列選択の関数，select について。先ほど読み込んだ野球データは表 4.1 にあるように多くの変数があります。ここから, 身長 height と体重 weight だけの変数を取り出したい, というときにこの関数を使います。コード 4.3 の結果が出力 4.3 です。

code : 4.3　select 関数

```
1  dat %>% select(height,weight)
```

R の出力 4.3: select 関数の出力結果

```
> dat %>% select(height,weight)
# A tibble: 6,546 × 2
   height weight
    <dbl>  <dbl>
 1    188     97
 2    182     73
 3    183     95
 4    171     73
 5    201    100
 6    183     90
 7    177     82
 8    173     73
 9    176     80
10    188     97
# … with 6,536 more rows
```

変数が 2 つに絞られていることが確認できますね。

ここで重要なポイントがあります。それはパイプ演算子 %>% です。これは前のオブジェクトを後の関数の第一引数に渡す, という意味があります。たとえば，$y = f(x)$ という変換をし, さらに $z = g(y)$ という変換をしたい場合, 頭の中では $x \to y \to z$ と考えていると思いますが, コードで書くと $y \leftarrow x$ で $z \leftarrow y$ という逆の向きで考えなければなりません。あるいは $z = g(f(x))$ と内側にくり入れていくので, 見た目にわかりにくく

なります。この思考の流れに沿った変換にするようにつなげるのが %>% です[*19][*20]。

■行選択：filter　今度は行選択。filter という関数です。先ほど読み込んだデータは 6546 行もあるのですが，たとえば 2020 年度のデータだけが欲しい，というときは行を選択することになります。ここで使うのが filter です。コード 4.4 の結果が出力 4.4 です。

code：4.4　filter 関数

```
1  dat %>% filter(Year == "2020年度")
```

R の出力 4.4: filter 関数の出力結果

```
> dat %>% filter(Year == "2020年度")
# A tibble: 664 × 17
   Year     Name     team  salary bloodType height weight UniformNum position
   <chr>    <chr>    <chr>  <dbl> <chr>       <dbl>  <dbl> <chr>      <chr>
 1 2020年度 菊池 … Carp   30000 A型           171     72 33         内野手
 2 2020年度 鈴木 … Carp   28000 A型           181     98 1          外野手
 3 2020年度 Ｋ．ジ… Carp   27250 不明          193     95 42         投手
 4 2020年度 會澤 … Carp   18000 B型           177     91 27         捕手
 5 2020年度 大瀬良… Carp   17500 AB型          187     95 14         投手
 6 2020年度 長野 … Carp   17000 O型           180     85 5          外野手
 7 2020年度 田中 … Carp   15000 A型           171     84 2          内野手
 8 2020年度 中﨑 … Carp   14500 B型           186     99 21         投手
 9 2020年度 野村 … Carp   12000 AB型          177     89 19         投手
10 2020年度 松山 … Carp    8500 B型           176     96 55         外野手
# … with 654 more rows, and 8 more variables: Games <dbl>, AtBats <dbl>,
#   Hit <dbl>, HR <dbl>, Win <dbl>, Lose <dbl>, Save <dbl>, Hold <dbl>
```

これで出来上がったデータセットは 664 行 ×17 変数です。10 年分で 6500 行ほどだったので，一年分だけ切り出すとその 1/10 ほどのデータサイズになるわけですね。

ここで，filter 関数の中では，Year == "2020 年度"と書いてあります。これは Year 変数の値が「2020 年度」だった行だけ抜き出しましょう，という意味です。イコールを 2 つ重ねている==という記号は，「左が右と等しいかどうか」を判断するための関数のような動きをしています。イコールが 1 つしかなければ，R は代入せよという意味，すなわち<-と同じ意味に解釈してしまうのです[*21]。

この条件判断式は色々な使い方があります。コード 4.5 にいくつかの例を示しますので，実行してどのようになるか確かめてみてください。

code：4.5　filter 関数

```
1  # 身長が180cm以上の人だけ取り出す
2  dat %>% filter(height >= 180)
3  # 体重が100kg以下の人だけ取り出す
```

[*19] これを読むときは，頭の中で「経て」と読みましょう。dat を経て select へ。なんのこっちゃと思った人は，ダウンタウンのごっつええ感じという昔の番組で「経て」というコントがあったので，ググってみてみてください。

[*20] パーセント，大なり，パーセントと 3 回も入力するのは面倒ですね！ RStudio ではショートカットキーとして，Ctrl+Shift+M があります。Ctrl キーと Shift キーと M を同時押し，です。こうした入力ショートカットについては他にもいろいろあり，RStudio の Help メニューから CheetSheet を選択してください。チートコマンドを書き込んださまざまな情報が 1-2 枚にまとまっています。

[*21] 実は普段から，dat=read_csv(***) のように書いても良かったのですが，いわゆる等式を表す記号と混乱するのでここでは<-にしていました。

```
4  dat %>% filter(weight <=100)
5  # 投手だけ取り出す
6  dat %>% filter(position == "投手")
7  # 野手だけ取り出す，野手とは投手でない人のことです
8  dat %>% filter(position != "投手")
9  # どちらかの条件を満たす，という場合
10 dat %>% filter(position == "捕手" | position == "内野手" |
11                    position =="外野手")
```

野手だけ取り出すときに!=という記号が出てきましたが，これは $\neq$ の意味です。$\neq$ が入力できないので，!で否定を表すようになっているのですね。このデータは投手のデータと野手のデータが含まれており，position は守備位置を表しますから，野手のデータは捕手か内野手か外野手ということになります。A か B か C, という選択を表す表現は|という表現を使いますが，今回は長くなるので「投手でない」という条件で抜き出しました。

ここまで列選択 select と行選択 filter について説明しました。両者の違いをイメージで理解しておいてください (図 4.6)。

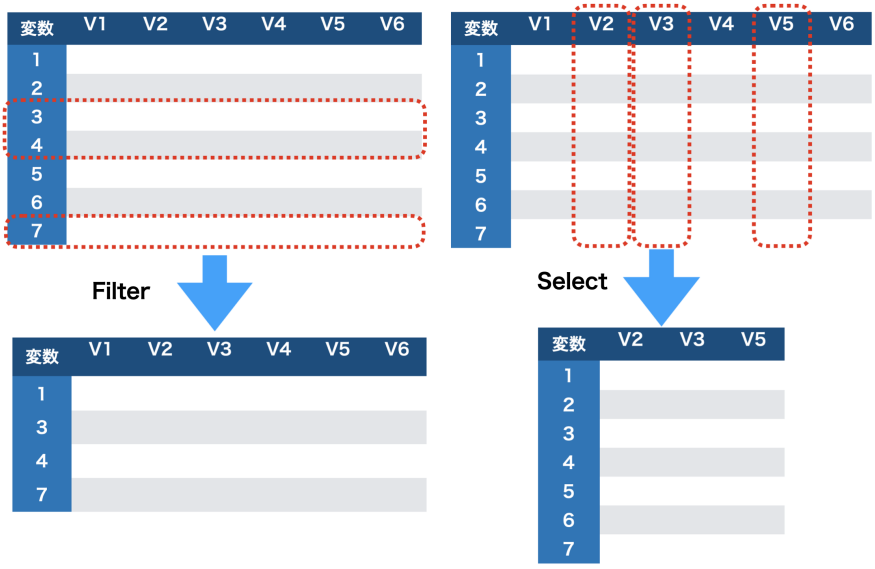


図 4.6 filter(左) と select(右) の違い

■変数生成；mutate 今度は変数を作る，mutate という関数です。たとえば打率は安打数を打席数で割ったもの，です。今回のデータセットには安打数 Hit と打席数 AtBats はありますが，打率のデータがありません。これを作ろう，というときはコード 4.6 のようにします。

code：4.6 mutate 関数

```
1  dat %>% select(AtBats,Hit) %>% mutate(BattingAverage = Hit / AtBats)
```

今回はいったん変数を選択してから，mutate 関数に入れました。もちろん直接入れてもらっても構いません。いずれにせよ，これで変数が作れることがわかったと思います。

この「新しく変数を作る」というのは結構重要で，たとえばデータの中に欠損値があったとき，それを別の数字で置き換えるということがしたくなるかもしれません。mutate 関数で作る変数名を既存の変数名と同じにしておくと**上書き**されますから，これで書き換えることができます。

またデータサイエンスの実践上，変数同士の組み合わせからデータの特徴を表す別の変数を作り出したいということはよくあります。こうした加工のことを**特徴量エンジニアリング**といいます。

このように変数を加工することで，既存の変数から新しい変数をどんどん作っていくことができます。言い換えれば，データの加工はこれらの関数を使って，R 上で記録を取りながら加工していきましょう，ということです。データを表計算ソフトで作っていて，そこで加工してしまうとデータの捏造が疑われる可能性があります。入力ミスや欠損値など不完全なデータであっても，それをどのように修正したかをすべて R で記録しながら進めていくことが，オープンサイエンスの取り組みとして重要なことなのです。入手したデータファイルは手を加えず，加工のプロセスをすべてオープンに明示しながら分析を進めていくことが重要です。

4.3　グラフの文法

それでは最後に，描画するグラフの文法について紹介します。

4.3.1　グラフの描画

セクション 2.4，Pp.30 でも触れましたが，**データは図にする**のが基本中の基本ですし，この図をプログラミングすることが客観性，再現性の観点からも重要です。適切な図からは適切な情報の読み取りができますし，人為的でなくても不適切な図から間違った印象を受け取ることもありますから，なぜ・どういうつもりで図にするかを明文化できるようにしておきましょう。このことについての専門書，Healy (2018 瓜生他訳 2021) の第 1 章は，知覚心理学や認知心理学の観点からグラフの書き方について論じられています。是非一度読んでみてください。

ここでは**可視化の文法 (Grammar of Graphics)** にしたがって作図をする，`ggplot2` パッケージ (Wickham, 2016) を使いながら説明します。

■可視化の文法　可視化の文法に沿った作図のコードは，次のように考えていきます。

1. X 軸 Y 軸にどのような変数を置くかを考える。
2. X 軸 Y 軸から出来上がるキャンバスに，点や線，何をプロットするのかを指定する。
3. その上にモデルを加えて，元のデータとの対応を考えたりする。別のデータを重ねることもある。

これを踏まえて，さきほどのコードに続けて[*22]，`code:4.7` を入力してください。

code：4.7　まずはキャンパスを設定する

```
1  g1 <- ggplot(data = dat,  mapping = aes(x = height, y = weight))
2  g1
```

1 行目　`g1` オブジェクトに ggplot で作った図の情報を渡します。

2 行目　`g1` って何？ つまりこれを表示しろ，という命令です。`plot(g1)` と書いても結構です。

この 1 行目で，データは `dat` というオブジェクトを使い，x 軸に `dat` オブジェクトの身長変数 `height`，y 軸に体重変数 `weight` を対応させたマッピングをしなさい，ということを指定しています。`aes` は aesthetic の略です。エステティック，審美的・美学的思想という意味です。データの美的対応を考える箇所ですので，ちょうどいい名称ですね。この `aes` 関数の中にはこのほかにも，色を何に対応させるか，サイズを何に対応させる

[*22] 続きにかくというのは，パッケージやデータの読み込みが終わっていることが必要だからです。描画に必要な `ggplot2` パッケージは，`tidyverse` パッケージを読み込むと同時に読み込まれています。

か，群わけはどの変数に対応しているか，と言った対応関係を表す関数情報を設定するところだと思っておいてください。

さてこのように，データの縦横だけを指定したオブジェクトを作り，それを表示させると，プロットのところには灰色の背景を持ったキャンパスだけが表示されたと思います (図 4.7 の左上)[*23]。

これがベースになって，次々と要素を付け加えていきます。次のコード code:4.8，code:4.9 を書いて実行してみてください。

code：4.8　キャンパスにデータ点をプロット

```
1  g2 <- g1 + geom_point()
2  g2
```

1 行目　g1 オブジェクトに点 (point) を打つ関数を追加したものを，g2 オブジェクトとします。

2 行目　g2 って何? つまりこれを表示しろ，という命令です。plot(g2) と書いても結構です。

code：4.9　線形モデルを当てはめる

```
1  g3 <- g2 + geom_smooth(method="lm")
2  g3
```

1 行目　g2 オブジェクトに滑らかな線を引く関数を追加し，g3 オブジェクトとします。線の関数は lm，Linear Model の略で直線を引く時に使います。

2 行目　g3 って何? つまりこれを表示しろ，という命令です。plot(g3) と書いても結構です。

図がどのように変わっていくか，確認できたでしょうか。g1 オブジェクトはただの白いキャンパスでしたが，それに点を打ったり (散布図)，線を引いたり (モデルの当てはめ) することでグラフがどんどん出来上がっていきます。オブジェクトに次々追加していくことでこれができています。今回は理解のために逐一オブジェクトを作りましたが，慣れてくると一文で表現できます。

また今回使った geom_という関数は，geometry(幾何学) の頭文字ですが，これを使ってさまざまなマッピングをします。よく使われるものに，次のようなものがあります。

- geom_point 点を打つ。散布図を書くことができる。
- geom_boxplot ボックスプロットを書く。
- geom_bar 棒グラフを書く。
- geom_line 線を引く。折れ線グラフなどに使う。
- geom_smooth 滑らかな線を引く。どのような線を引くかは色々な関数が考えられる。

これらの組み合わせで，さまざまなグラフが表現できるのです。少し書き換えるだけですぐにできますので，色々変えて遊んでみましょう。

最後に，次のコード code:4.10 を実行してみましょう。

code：4.10　チームごとに分割

```
1  g4 <- g3 + facet_wrap(~team)
2  g4
```

[*23] Mac ユーザであれば，x 軸，y 軸のラベル名が表示されず四角い豆腐のようなものしか出ないかもしれません。これは文字コードの問題で，次のコードを実行すれば直ります。old <- theme_set(theme_gray(base_family = "HiraKakuProN-W3"))。あるいは，「Mac ggplot2 豆腐 文字化け」などで検索してみてください。

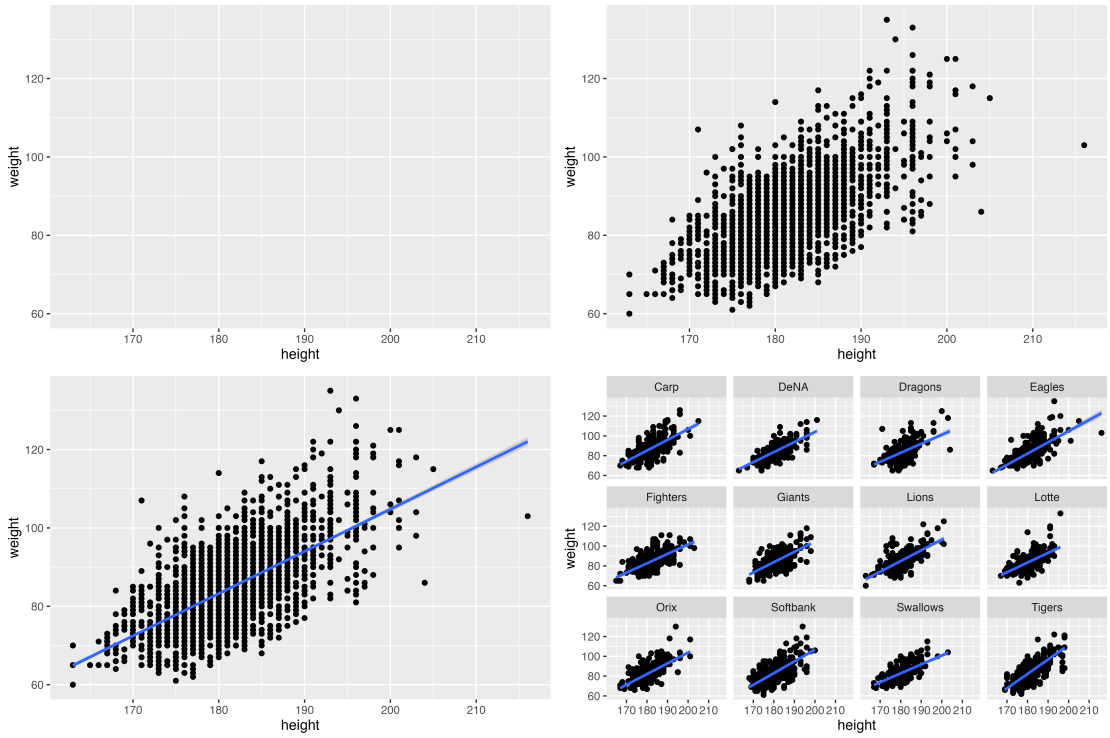


図 4.7 ggplot による描画。左上は code:4.7, 右上は code:4.8, 左下は code:4.9, 右下は code:4.10 の結果

1 行目 g3 オブジェクトをチーム変数 team で区切った図にせよ, という対応を指示。

2 行目 g4 を表示しろ, という命令です。

今度は facet_wrap という関数が出てきました。facet_というのは分割, 区切りを考える関数です。データにはさまざまな情報を書き込むことができますが, 適当なグループで書き分けることでわかることも出てきます。今回は team という変数で分割せよ, という表現をしており, その結果が図 4.7 の右下のようになっています。

可視化の文法については, 習うより慣れろといった側面があります。さまざまな出力例をみながら, どのようにこのグラフが書けたのか, 文法を読み解く練習もしてみましょう。ggplot による出力例は, https://www.r-graph-gallery.com/ggplot2-package.html などが参考になります。

4.4 課題

今回の課題は, Rmarkdown ファイルで作成してください。タイトルは「課題 5」とし, 作者名 (Author) のところに学籍番号, 氏名を記入してください。出力形式は HTML を選んでおいていただければ結構です。**提出するファイルはエラーのない完備な.Rmd ファイルです**。ファイル形式の間違い, 出力にエラーが含まれている (未完成な) ものは返却しますので注意してください。

■記述統計量を計算する 次の手順を実行し, データの記述統計量を計算してください。

1. サンプルデータ BaseballDecade.csv をダウンロードし, プロジェクトフォルダに保存します。
2. code:4.11 を実行し, データを整えます。
3. データ全体の要約をする関数 summary を使って, データオブジェクトの記述統計量を表示してくだ

さい。

4. 身長変数 height の平均値を算出してください。平均値の関数は mean です。
5. 体重変数 weight の標準偏差を算出してください。標準偏差の関数は sd です。
6. 年収変数 salary の中央値を算出してください。中央値の関数は median です。
7. 身長と体重の相関係数を算出してください。相関係数の関数は cor で，2 つの引数をとります。

code：4.11 データの読み込みと整え

```
1 library(tidyverse)
2 dat <- read_csv("BaseballDecade.csv")
3 dat$team <- as.factor(dat$team)
4 dat$position <- as.factor(dat$position)
5 dat$bloodType <- as.factor(dat$bloodType)
```

1 行目 tidyverse パッケージを読み込みます。

2 行目 データファイルを読み込みます。

3-5 行目 team,position,bloodType などの変数は，**名義尺度水準**です。これが名義尺度水準であると設定し，同じ変数名に上書きしています。名義尺度水準にする関数は as.factor です。

■図を書いてみる 1 同じデータセットを使って，code:4.11 を実行してください。

code：4.12 さまざまなグラフを書いてみる

```
1 g1 <- ggplot(data = dat,
2              mapping = aes(x = height, y = weight, color= team))
3 g1 + geom_point()
4 g1 + geom_smooth()
5 g2 <- ggplot(data = dat,
6              mapping = aes(x = team, y = salary, fill= team))
7 g2 + geom_boxplot()
```

1 行目 データは dat オブジェクト，x 軸は身長変数，y 軸は体重変数で，チームごとに色分けをします。

2 行目 散布図を書いてみます。

3 行目 データに合うような線を引いてみます。

4 行目 データは dat オブジェクト，x 軸にチームを，y 軸に年収変数をおいて，チームごとに色で塗りつぶします。

5 行目 箱ひげ図を書いてみます。

■図を書いてみる 2 dat オブジェクトを使って図 4.8 と同じ図を書いてください。データは 2020 年度のものだけ使っています。

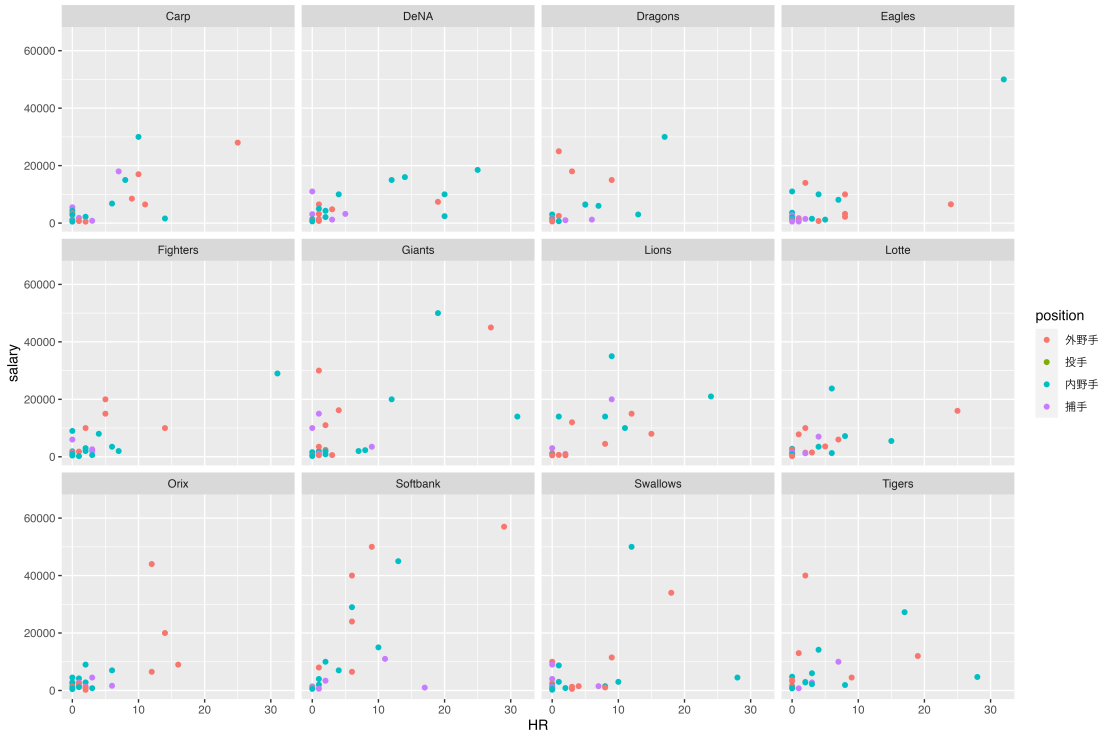


図 4.8 課題の図

4.5 補遺

4.5.1 データの作成

少し横に話がそれますが, どのようにデータファイルを作るかについても説明しておきましょう。

データが数件しかないのであれば, 上で解説した方法で `data.frame` 型に 1 つ 1 つ入力していけば良いのですが, データを手入力してファイルを作る場合は, 表計算ソフトを使う方が便利です。このときに注意すべきことは, **1 つのセルには 1 つのデータしか入れない**ということです。これからデータを色々加工していきますが, 加工する前のデータはなるべくプレーンな状態であったほうが良いのです (すでに加工された食材を違う料理に利用するのは大変ですから！)。

例えば図 4.9 は, 悪いデータファイル作成の例です。これをまずはじっくり眺めて, どこがまずいのか考えてみてください。少なくとも 7 つの点で問題があります。

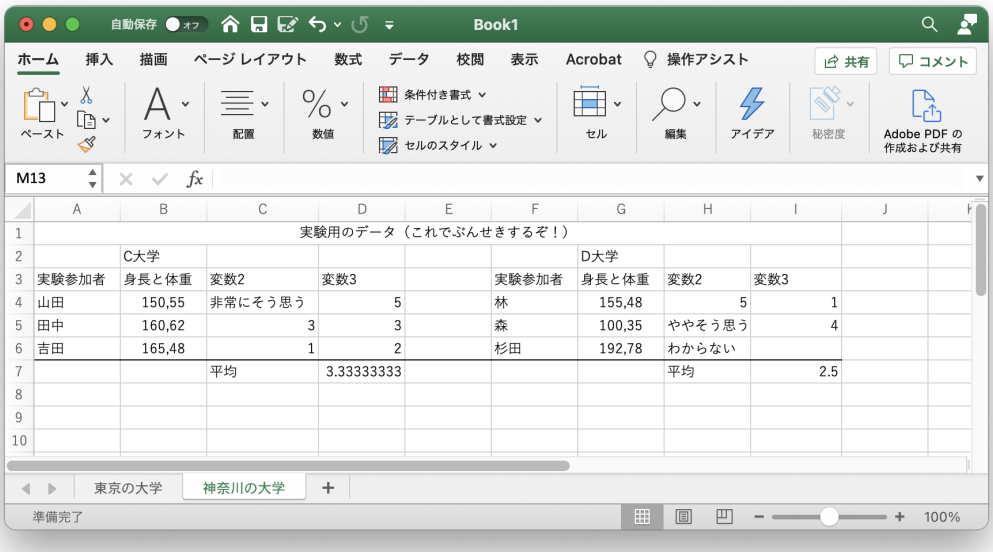


図 4.9 分析に不向きなデータファイル

分析しにくいよくないデータになっているのは次の箇所です。

1. 1 行目にこの表のタイトルのようなものが入っていますが，これは分析には使わない情報なので不要です。
2. 1 行目はセルを結合しています。データセットは `data.frame` 型のように矩形であるべきなので，データのサイズを狂わせる行の存在は不適切です。
3. 2 行目には大学名が入っているようです。A から D 列は C 大学，F から I は D 大学のデータのようですが，1 つのデータセットに複数のデータが混在しているのはよくありません。
4. 身長と体重，という変数名のところには 1 つのセルに複数の数字が入っています。カンマで区切られており，見た目にはわかりますが，機械的には (カンマのせいで) 数字ではなく文字列として扱われます。データがデータになっていないのです。
5. 変数 2 の列は「非常にそう思う」という言葉の回答と，数字が入っています。1 つの列の中に複数の種類のデータが含まれるので分析できません (文字と数字の集計ができない。ちなみにこの場合は文字変数として扱われます)。
6. 7 行目に平均とありますが，データの平均はデータから計算できる量なので，R で算出するものです。データの中に計算結果を入れておくべきではありません。
7. I 列 6 行目，D 大学の杉田さんに関する変数 3 の箇所が空欄です。これはデータが入手できなかったところかもしれませんが，これは欠損値であることを明示すべきです。
8. 「東京の大学」と「神奈川の大学」という複数のシートがあります。データファイルとしては 1 つのシートにすべての情報が入っていなければなりません。同時に分析できるデータなのであれば，一枚のシートにまとめるべきです。

皆さんはいくつ気づいたでしょうか。これらの問題点を通じて言えることは「1 つのデータセット，1 つのセルには過不足ない情報を入れる」ということです。表を作って見せるのであればこのような入力でも良いのですが，ここでは数値データを分析することが目的なのですから，レイアウトや数字でない情報，計算してここから

作れる情報は，分析に必要のない過剰な情報なのです。ちなみにアンダーラインや文字色，セルの色などは分析の時に自動的に受け落ちます (データではない情報なので)。表計算ソフトは数字を表にして計算するソフトウェアであり，方眼用紙やレイアウトする文書とは異なることを意識しましょう。

加えて分析の時に不便なのは，文字と数字の関係です。1 つのセルに複数の数字が入る場合，区切り記号などをつけることが多いかと思いますが (スペースも機械にとっては文字です！)，こうした文字種・変数の種類・水準が混在しているのも情報過多ということになります。

今回の問題点を修正したのが図 4.10 のようなデータセットです。単一のデータセットで完結しています。また 1 つの列に一種類の変数，不要な情報はなく，過剰な情報もないデータセットになっているのがわかるかと思います。なお，変数 3 のところに `NA` とあるのは `No Answer` の略で**欠損値 (missing value)** を意味しています。これは読み込む時に，欠損値をどのような文字にしているか，という指定ができますので問題になりません。

	A	B	C	D	E	F	G
1	実験参加者	地域	所属大学	身長	体重	変数2	変数3
2	山田	神奈川	C	158.1	62.3	非常にそう思う	1
3	田中	神奈川	C	149.7	55.3	非常にそう思う	5
4	吉田	神奈川	C	158.2	53.9	非常にそう思う	3
5	林	神奈川	D	149.5	53.8	ややそう思う	2
6	森	神奈川	D	158.7	54	非常にそう思う	4
7	杉田	神奈川	D	161.8	57.6	非常にそう思う	5
8	久保田	東京	A	168.5	55.2	ややそう思う	6
9	悲鳴嶼	東京	A	155.7	57.9	非常にそう思う	2
10	荒木	東京	A	151.1	60.8	非常にそう思う	3
11	荒尾	東京	B	156.9	63.3	ややそう思う	2
12	田邊	東京	B	156	56.5	非常にそう思う	NA
13	末次	東京	B	163.3	59.6	わからない	5
14	青山	東京	B	158.6	59.7	非常にそう思う	3
15	諫山	東京	B	154.7	55.9	非常にそう思う	NA
16							

図 4.10　分析に適したデータファイル

このように，データの入力の基本は，**1 行に 1 ケースの情報が入っている，過不足のない 1 つのデータセットを作ること**である，と言えます。

このようにして作られたデータファイルは，UTF-8 の文字コードによる csv 形式で保存するのが基本です (→ 付録 A，A.5 節，Pp.396 参照)。ファイルを保存する時に，ファイル形式をしっかり指定し，拡張子も `.csv` にしましょう。csv ファイルは ASCII ファイルですし，文字コードも世界基準の UTF-8 にしておけばどの PC からでもみることができます。ある一企業が開発した特定の OS，特定のアプリケーションを持っていなければデータを見ることができない，というのは科学的研究の観点からは不適切なのです。

ちなみに 2020 年 12 月，総務省により機械判読可能なデータの表記方法の統一ルールが策定されました

(総務省, 2020)。それには次のようなチェック項目が含まれています。

チェック項目 1-1　ファイル形式は Excel か CSV となっているか
チェック項目 1-2　1 セル 1 データとなっているか
チェック項目 1-3　数値データは数値属性とし, 文字列を含まないこと
チェック項目 1-4　セルの結合をしていないか
チェック項目 1-5　スペースや改行等で体裁を整えていないか
チェック項目 1-6　項目名を省略していないか
チェック項目 1-7　数式を使用している場合は, 数値データに修正しているか
チェック項目 1-8　オブジェクトを使用していないか
チェック項目 1-9　データの単位を記載しているか
チェック項目 1-10　機種依存文字を使用していないか
チェック項目 2-1　データが分断されていないか
チェック項目 2-2　1 シートに複数の表が掲載されていないか

みなさんも自分でデータを作るときは, こうした基準に沿うように心がけてください。また表計算ソフトの中で, 代表値を計算したりグラフを書いたりしたくなりますが, みなさんは既に統計環境を手に入れているのですから, そうした加工・可視化などは R のほうでやり, データは「誰が触っても良いプレーンな素材」として保管しておくように心がけましょう。

4.5.2 整然データ

データ入力の時に見たように, データファイルは 1 セルに 1 情報, 重複や集計, オブジェクトや文字列のないデータセットであることが基本です。また図 4.10 にあるように, 一行ごとに 1 人 (1 ケース, 1 オブザベーション, 1 レコードとも言います) のデータが入っており, 各列が変数を表していることが一般的です。変数が増えると横に伸びていきますので, こうしたデータの入力形式を**ワイド型**のデータということがあります。

ここで表 4.2 を見てください。このデータを見ると我々は, 「大阪の午後は曇りなんだな」ということがすぐにわかります。

表 4.2　ワイド型データ

	午前	午後	夕方	深夜
東京	晴	晴	雨	雨
大阪	晴	曇	晴	晴
福岡	晴	曇	曇	雨

しかしその判断は, 「大阪」行の「午後」列を見て, すなわち行と列に素早く目を走らせてそれのクロスするところで「曇」という情報を得ています。夕方や深夜の情報が知りたければ, 列を変えますね。列ごとに違う変数になっているから当然です。行と列の名前を参照するというのは, 目線を左右と上下に動かす 2 アクション必要なことです。

しかしここで参照している「列名 (変数名)」も変数, データの一部ではないでしょうか。この観点に立ち, 1 行見るだけですべての情報がわかるようにしてみましょう。表 4.3 には表 4.2 と同じ情報しか含まれていませんが, 一行見るだけですべての情報がわかるようになっています。

表 4.3　ロング型データ

地域	時間帯	天候
東京	午前	晴
東京	午後	晴
東京	夕方	雨
東京	深夜	雨
大阪	午前	晴
大阪	午後	曇
大阪	夕方	晴
大阪	深夜	晴
福岡	午前	晴
福岡	午後	曇
福岡	夕方	曇
福岡	深夜	雨

表 4.3 のような形のデータのことを，**ロング型**のデータと言います。このような形にしておくと，情報の欠落 (欠損値) があった時の対応が楽に行えたり，レコードの選択がやりやすくなるなどの利点があります。

このような分析しやすいデータの形について，Hadley (2014) が提唱したのが**整然データ (Tidy Data)** という考え方です。整然データとは，次の 4 つの特徴を持ったデータ形式のことです。

- 個々の変数 (variable) が 1 つの列 (column) をなす。
- 個々の観測 (observation) が 1 つの行 (row) をなす。
- 個々の観測の構成単位の類型 (type of observational unit) が 1 つの表 (table) をなす。
- 個々の値 (value) が 1 つのセル (cell) をなす。

さきほどの天気の例では，時間，場所，天候という 3 つの情報が表に含まれていますが，これが一行の中にまとまっているのが整然データです。人間にとってはワイド型のデータの方がわかりやすいこともありますが，分析を進めていく上ではロング型・整然データの方が便利なことも多くあります。たとえばこの後で行う作図においては，整然データの方が便利な形です。X 軸に割り当てる列，Y 軸に割り当てる列，色や形を指定する列，など列ごとに持つ特徴を図の設定に割り当てることができるからです。

ここで見たようにロング/ワイドどちらの形式でも持っている情報は同じですが，用途に応じて使い分けます。これを実行するのもまた，データハンドリング技術の応用です。少し見てみましょう。

■ロング型とワイド型の変換　データをロング型の整然データにすることを考えましょう。次のコード 4.13 を実行してください。

code : 4.13　ロング型にする関数

```
1  dat.long <- dat %>%
2    select(Year, Name, position) %>%
3    rownames_to_column("ID") %>%
4    pivot_longer(-ID) %>%
5    print()
```

1 行目　いまから dat オブジェクトを加工していきます。加工した結果を dat.long オブジェクトに代入します。

2 行目　dat オブジェクトから年度，名前，ポジション変数だけ取り出しています。

3 行目 行変数名，あるいは何行目かという情報をもった変数を一列作ります。変数名は ID とします。

4 行目 pivot_longer 関数でロング型にします。このとき，ID 列がキーになっています。

5 行目 結果を出力させています。

このコードを 3 行目まで実行した場合，出力 4.5 のようになります。ID という列ができていますね。この列変数をキーに，Year,Name,Position とあった変数を変数名とその値というペアに置き換えるのが，pivot_longer 関数です。

R の出力 4.5: 行番号をつけた段階

```
# A tibble: 664 × 4
   ID    Year      Name            position
   <chr> <chr>     <chr>           <fct>
 1 1     2020 年度 菊池　涼介      内野手
 2 2     2020 年度 鈴木　誠也      外野手
 3 3     2020 年度 K．ジョンソン   投手
 4 4     2020 年度 會澤　翼        捕手
 5 5     2020 年度 大瀬良　大地    投手
 6 6     2020 年度 長野　久義      外野手
 7 7     2020 年度 田中　広輔      内野手
 8 8     2020 年度 中﨑　翔太      投手
 9 9     2020 年度 野村　祐輔      投手
10 10    2020 年度 松山　竜平      外野手
# … with 654 more rows
```

最後までいくと，結果は出力 4.6 のようになっているはずです。

R の出力 4.6: 行番号をつけた段階

```
# A tibble: 1,992 × 3
   ID    name      value
   <chr> <chr>     <chr>
 1 1     Year      2020 年度
 2 1     Name      菊池　涼介
 3 1     position  内野手
 4 2     Year      2020 年度
 5 2     Name      鈴木　誠也
 6 2     position  外野手
 7 3     Year      2020 年度
 8 3     Name      K．ジョンソン
 9 3     position  投手
10 4     Year      2020 年度
# … with 1,982 more rows
```

これで整然データの形になっていますね。すなわち，name のところに変数名が，value のところにその変数の値が入っているわけです。一行目は，ID1 番の，変数 Year の値が 2020 **年度**であることを示し，2 行目は ID1 番の，変数 Name の値が**菊池　涼介**であることをしているわけです。このようにして，ロング型にすることができました。

ちなみに，新しく作られる変数の変数名を指定することもできます。コード 4.14 の例では，変数名を

Variables, 値が入る変数名を Value としてみました。

code : 4.14　出力する変数名の指定

```
dat %>%
    select(Year, Name, position) %>%
    rownames_to_column("ID") %>%
    pivot_longer(-ID, names_to = "Variables", values_to = "Value")
```

さて, このようにロング型のデータができたら, その逆もみておきましょう。ロング型をワイド型に戻したいこともあるわけです。次のコード 4.15 を実行してください。

code : 4.15　ワイド型にする関数

```
dat.long %>%
    pivot_wider(id_cols = ID, names_from = name, values_from = value)
```

これは一行だけですが, 引数が色々ありますね。まずどの変数をキーにして横に並べるかを id_cols で指定する必要があります。この ID が同じものを横に並べていくからですね。次は変数名はどの列からくるのか, 値はどこからくるのかを指定する必要があります。

pivot_longer や pivot_wider 関数の基本的な使い方はこれだけですが, 変数を作る時に変数名を加工したりとか, 空白セルができてしまう場合はどのように処理するかといった設定もできます。詳しくは松村他 (2021) の 3 章などを参照してください。

第 5 章

記述統計量

さて，ここまでで心理学においてなぜ統計学が必要か，計算機とはどういうものか，計算機に与えるデータはどういうものか，という話をしてきました。下準備はそこまでです。いよいよ，今回から心理統計の具体的な話に入っていきます。

初めに，心理統計の 3 つの枠組みを紹介しておきます。

記述統計 descriptive statistics　データの特徴を記述する統計法。可視化，代表値による要約など。データを取ったらまずやるべき作業がこちら。手元のデータでわかることを，隅々まで書き表すのが狙いです。

推測統計 inferential statistics　部分的なデータから全体像を推測する統計学。心理統計の目的は一部のサンプルから全体を予測することにあるのです。この授業の後半 (後期) は主にこの話になります。

多変量解析 multivariate analysis　大量のデータから意味のある情報を引き出す，あるいは無意味な情報を削ぎ落とす技術です。調査研究をする場合はもちろん，センサーなどの機器から大量に取れるデータを扱う場合にも使われる技術です。データサイエンス，機械学習 (いわゆる AI) などにも通じる領域です。本学では 2 年のデータ解析応用で扱います。

ということで，ここにあるように記述統計学が最初のステップになります。手元にあるデータの特徴を記述することが目的になります。記述という意味では可視化もそうなのですが，今回は数字によってデータの特徴を記述することを考えます。複数のデータの代表的な値ですから，代表値と言われたりします。

研究対象を数値化したとしましょう。それを眺めて「いろいろな数字があるなあ」と感心していたのでは，数値化した意味がありません。得られた**データは図にする**のが基本です。すぐに図にします。いろいろな図にします。前回も話しましたが，とにかく図にします。しなければなりません。さて，図を見て考えをまとめるのが重要なのですが，具体的にどういう特徴があるのか，あるいはどういう違いがあるのかといったことを，要約して表現できればなお良いですね。そこで代表値を考えることになります。

いきなり身も蓋もないことをいうようですが，たくさんの数値を 1 つの代表値を使って記述する，というのは基本的に無理があります。十人十色，いろいろな違いがあるからこそ，そのデータセットは意味があるのに，たった 1 つの数字に変えて表現するのですから。代表値は，何らかの側面で評価することで 1 つに定まるものであって，あくまでもその評価次元での代表値に過ぎません。つまり，1 つの代表値だけで全体が分かるわけではないのです。あるいは，1 つの代表値にするということは，少し無理をして情報を削ぎ落とすということでもあります。代表値は何であれ，必ず何らかの情報が欠け落ちているのです。ですから，統計データを扱うときは，複数の代表値を使って多角的に判断するのが基本です。そのことをまずしっかり理解しておいてください。

その上で，代表値を 2 つの種類に分けて説明します。ひとつは中心化傾向の指標，もうひとつは散らばりの

指標です。

5.1 中心化傾向の指標

5.1.1 平均値 mean

平均値 (mean)[*1], より正確には算術平均[*2]は最もよく知られた代表値でしょう。計算方法は, すべてのデータをたしあわせて, 足したデータの個数で割る, というものです。

図 5.1 に 5 人の所持金を示しています。この例では, 平均値は次のように計算できますね。

$$(15000 + 30000 + 20000 + 10000 + 20000)/5 = 19000$$

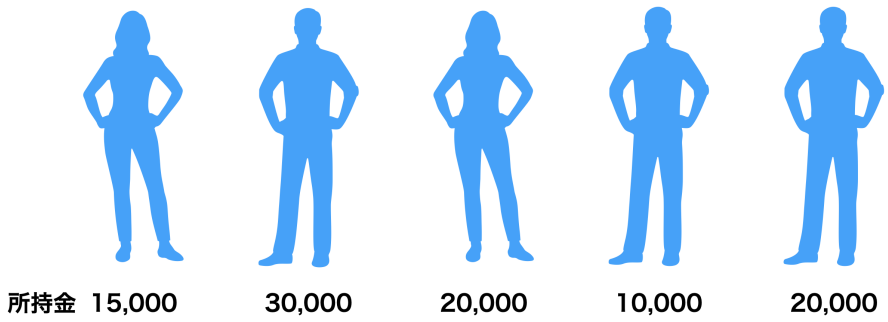


図 5.1 5 人の所持金データ

■文字と記号の導入 さてこれを, 今後のために記号で一般的に表記する方法について知っておきましょう。変数は一般に x や y で表されます。今回は 5 人分のデータがありますが, 1 人目のデータ, 2 人目のデータ, といった違いを表すときは $x_1, x_2, \cdots$ のようにアルファベットの右斜め下に小さく数字を添えます。添字と言います。さて今回は所持金という 1 つの変数でしたが, 2 つ目の変数, 3 つ目の変数・・・とたくさんのデータが手に入ることもあります。そのようなとき, 変数は x_{ij} と添字を 2 つ書いて表現します。ここで i がケース (個体, 観測番号) を表し, j が何番目の変数であるかを表す, というルールにします[*3]。

次に足し算の記号を紹介します。$\sum$ で総和を表すことにします。この記号はギリシア文字の大文字の S で, 足し合わせる summation の頭文字をとって $\sum$, サムと読みます。この記号には上と下に文字を追加して, 次のような意味で使います。

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

ここにあるように, $\sum$ の下に i という文字があります。この文字 i が 1 から始まって 5 まで変化し総和する, という数学記号です。$\sum_{i=1}^{5} x_i$ の x_i という変数について, i が 1 から 5 まで変わるわけですから, 右辺にある

[*1] 英語では average というのも平均を意味します。mean の方がより意味が狭く, average はより広い意味で代表的であるという違いがあるようです。

[*2] 他にも調和平均, 幾何平均などがあります。

[*3] この講義ではこのルールで通します。必ずしも個体が前で変数が後, という決まりがあるわけではありませんが, このような記号を導入するときはどの文字が何に対応しているか明記することが一般的です。この講義でとくに明記していなければ, 前の添字が個体で後ろの添字が変数を表すのだと思ってください。

ように添字が変わりながらすべて足し合わせる，という意味の記号です。この $\sum$ も，文脈から考えてどこからどこまで足すのかが明らかな場合，上下の添字は省略されることがあります。

さて，先ほどの 5 人の平均ですが，足し算の場所は $x_1 + x_2 + \cdots + x_5 = \sum_{i=1}^{5} x_i$ と書けます。これを 5 で割れば平均になりますから，$\frac{1}{5}\sum x_i = 19000$ というわけです。このように記号で表現しておくと，平均値は一般に次のように表すことができます。

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

今回は 5 人でしたが，一般に N 個のデータを扱うときにまで拡張できました。左辺は平均値 $\bar{x}$ で，エックスバーで表すことが一般的です。

5.1.2 中央値 median

次の代表値は**中央値 (median)** です。データを大きさの順に並べて行って，ちょうど真ん中のケースの値をその代表値として使う，というものです。さきほどの図 5.1 の場合は，並べ替えると $10000 < 15000 < 20000 = 20000 < 30000$ となります。上から 3 番目 (下からでもいいです) が 20000 円ですから，中央値は 20000 円ということになります。

今回はたまたま，5 人のケースですから真ん中は 3 番目となりましたが，データの数が偶数しかなかったらどうするか，という問題があります。その場合は，真ん中に来た 2 つの値の算術平均でもって中央値とすることが一般的です。

5.1.3 最頻値 mode

最後に**最頻値 (mode)** についてです。漢字の成り立ちからわかるように，頻度が最も多いものを代表値とするのが最頻値です。さきほどの図 5.1 の場合は，10000 円が 1 人，15000 円が 1 人，30000 円が 1 人ですが 20000 円持っている人が 2 人いましたので，最頻値は 20000 円になります。

このように，データの中心をどのように考えるかによっても 3 つも考え方があり，それぞれが一致するわけではないのです。それぞれの意義がありますので，見比べながら考える必要があります。

では練習です。次の表 5.1 のようなデータがあった場合，平均値，中央値，最頻値はそれぞれどうなるでしょう？

表 5.1 5 人の所持金データその 2

A	B	C	D	E
10000	10000	20000	25000	100000

この場合，平均は 33000 円，中央値は 20000 円，最頻値は 10000 円となりました。平均値が他の 2 つより大きくなりましたね。これは**外れ値 (outlier)** があるからです。E さんが，平均値をぐっと押し上げているのです。平均値は量的な中心をとりますので，1 人でも遠く離れたところに位置すると全体的に中心が引っ張られます。中央値は順番の情報しか使いませんので，大金持ちが 1 人いても 1 位だ，と判断するだけです。最頻値は度数の問題ですので，大金持ちが 1 人いても度数 1 にしかなりません。平均値は外れ値に弱い，あるいは中央値や最頻値がそれに強いというだけです。

では平均値は使い物にならないのでしょうか？ もちろんそうではなくて，左右対称に散らばっているような

データであれば，平均値は良い代表値になります。データの散らばりが左右対称ではなく歪んでいるようであれば，平均値は適切な代表値になりません。

図 5.2 は，厚生労働省による「2021 年・国民生活基礎調査」で示されたデータです[*4]。これを見ると所得分布は左側に歪んでおり，平均値は 564 万 3 千円ですが，中央値は 440 万円，最頻値は 300-400 万円のクラスということになります。最頻値は一番データ点が多いところですので，「国民の平均所得は 564 万円です」と言われると多くの人が (この例では少なくとも 55.7% の人が)，「えっ，私の年収低すぎ・・・?」という印象を抱きます。もちろん所得分布が低い方に歪んでいる状況も悪いのですが，代表値として平均値を使うことがそもそも適切でない例ともいえます[*5]。ニュース報道などでは平均値しか報道されませんが，それでは不当な印象を与えることになります。**データは図にする**ことが大事なのです。

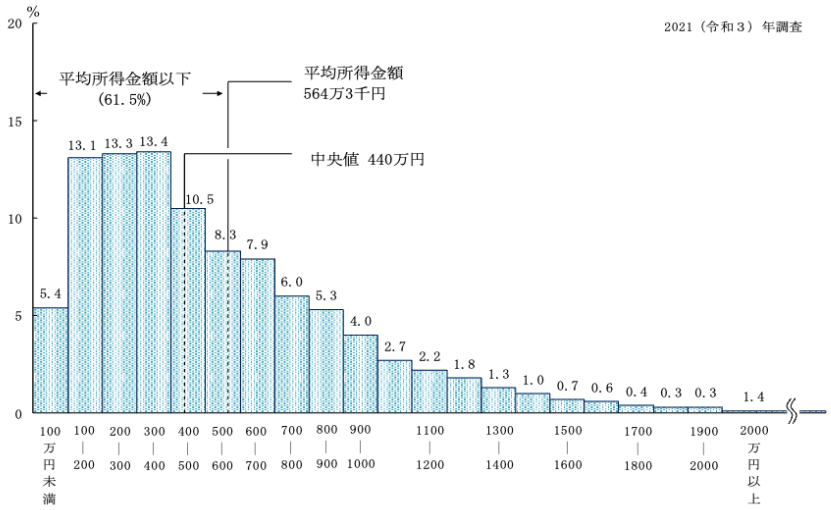


図 5.2　所得金額階級別世帯数の相対度数分布

ところで，最頻値は頻度を数えるものですから，間隔尺度水準以上の連続的な数字の場合はいくつかの区間に区切って数えることになります。図 5.2 の例もそうですね。ただしこの時，最頻値は区間幅の区切り方が違うと結果が変わってしまうことに注意しましょう。図 5.3 の 2 つのヒストグラムは，実は同じデータなのです。違いは区間幅 binwidth のとり方で，左図は 15 点間隔，右図は 5 点間隔でグラフを書きました。最頻値はそれぞれ赤で囲ったところになりますが，左図の場合の最頻値は 40-55 点クラス，右図の場合は 47-52 点のクラスになります。最頻値の場合は，どのような区間幅でカウントするかということにも注意が必要ですね。

では，ここで出てきた 3 つの中心化傾向の指標をいったんまとめておきましょう。

平均値　データを全部使って計算しており，外れ値がなければ全データが反映される良い指標。実際，統計分析で最も活躍している指標です。左右対称の分布に適しています。

中央値　外れ値の影響を受けにくい指標で，ちょうど真ん中，という意味も分かり易いです。ただ，分布の真ん中の数字だけしか使ってません。データが偶数個のときの中央値の計算はいくつかあり得ます。

最頻値　外れ値の影響を受けにくく，左右対称でない分布の時も意味が分かり易いです。データの値ではなく度数分布にのみ注目しており，度数分布の区間幅が変わると変わってしまうという欠点があります。

[*4] https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa21/dl/03.pdf

[*5] もちろん為政者の目から見れば，平均は 564 万円もあるんだから日本は豊かだ，とアピールできるので良い指標なのかもしれません。数字に悪意はなく，何の数字をどのように使うのかという人の側に問題があるだけです。

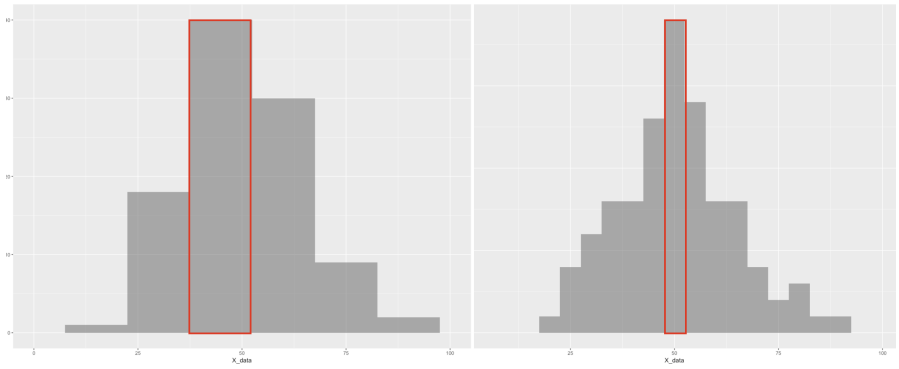


図 5.3 区間幅の区切り方で結果が変わる

5.2 散らばりの指標

5.2.1 分散と標準偏差

つぎは散らばりの指標について考えていきましょう。データの中心がどこにあるか，というだけではわからないこともあるのです。表 5.2 のような 2 つのデータがあったとしましょう。これを図示したのが図 5.4 です。

表 5.2 2 つのデータ

A	B
4	1
4	2
5	3
5	5
5	5
5	7
6	8
6	9

このデータ，A/B グループについてそれぞれ平均値，中央値，最頻値はどうなるでしょう。何と，平均値は A 群も B 群も 5，中央値も A/B 群ともに 5，最頻値も A/B 群ともに 5 となります。中心化傾向の指標だけでは，A 群と B 群の違いが表現できません。こんなにも違う見た目をしているのに！

この 2 つの群は何が違うかといって，図 5.5 を見れば明らかなように中心から左右にどれほど散らばっているかが違うのです。

■分散 これをうまく表現する代表値は，**分散 (variance)** があります。指標化のポイントは，「中心から個々のデータ点が平均的にどの程度離れているか」を考慮することにあります。数式では次のように表されます。

$$s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

この式にある $(x_i - \bar{x})$ は，各データ点 x_i が平均 $\bar{x}$ から離れている程度を表しています。これをとくに**平**

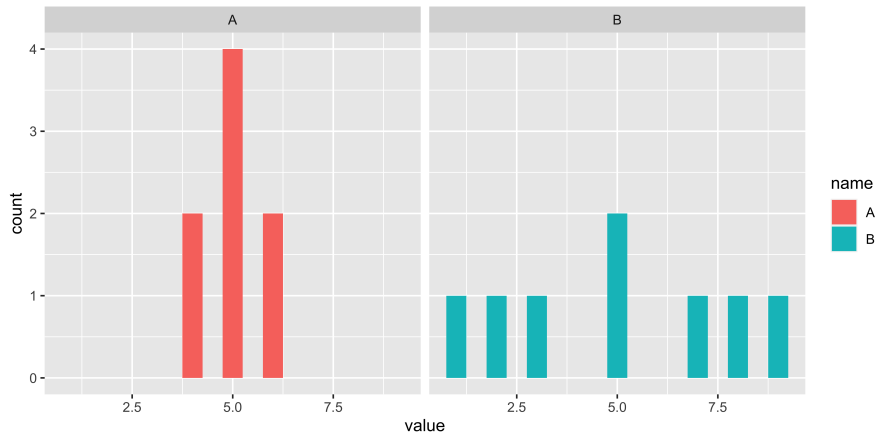


図 5.4 2 つのデータを見比べてみよう

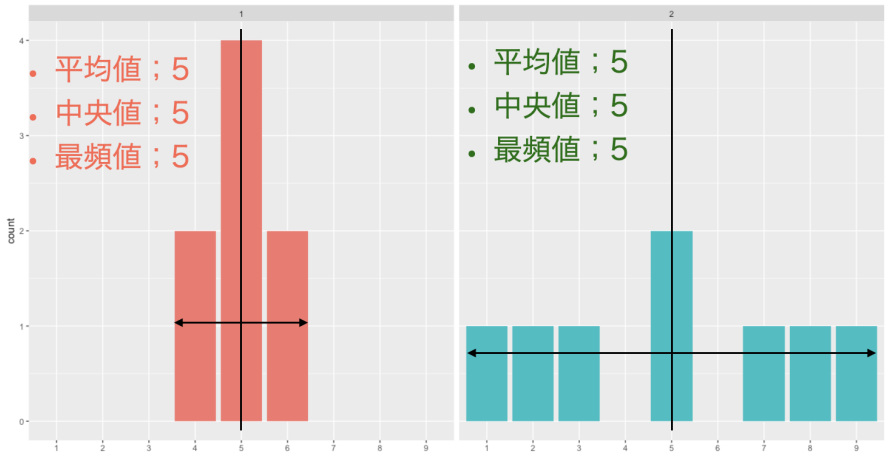


図 5.5 中心から各データ点が平均的にどの程度離れているか

均偏差 (mean deviation) と言います。それを二乗していますね。これがポイントで，二乗しておかないと平均偏差の総和はゼロになってしまうからです[*6]。プラスとマイナスが打ち消し合うからゼロになってしまうわけで，符号の影響をなくすべく二乗しているのだと思ってください。符号をなくすのには絶対値でもいいのですが，絶対値は数式で書けないから使いにくいのです[*7]

実際に計算してみましょうか。A 群の分散は，次のようになります。

$$\frac{1}{8}((4-5)^2+(4-5)^2+(5-5)^2+(5-5)^2+(5-5)^2+(5-5)^2+(6-5)^2+(6-5)^2)$$

$$=\frac{1}{8}(1^2+1^2+0^2+0^2+0^2+0^2+1^2+1^2)=\frac{4}{8}=0.5$$

B 群の分散は，次のようになります。

$$\frac{1}{8}((1-5)^2+(2-5)^2+(3-5)^2+(5-5)^2+(5-5)^2+(7-5)^2+(8-5)^2+(9-5)^2)$$

[*6] 検算してみよう!

[*7] 絶対値は $|x|$ と書くじゃないか，と思った人もいるかもしれません。おしい，もう一歩。この記号の中身，計算式を考えてみてください。$0 \leq x$ のとき，$|x| = x$ です。$x < 0$ のとき，$|x| = -x$ です。これ，「･･･の時」という条件で分岐してて，この箇所が数式で書けてないですよね。

$$= \frac{1}{8}(4^2 + 3^2 + 2^2 + 0^2 + 0^2 + 2^2 + 3^2 + 4^2) = \frac{58}{8} = 7.25$$

このように，A 群の分散は 0.5，B 群の分散は 7.25 と，B 群の方が散らばりが大きいことがわかります。

分散はデータの散らばりを表現しています。分散が大きい方がデータの散らばりが大きい，というわけですね。逆に分散がゼロになると，そのデータには散らばりがない，ということです。実際取ったデータの分散がゼロだと，そこから得られる情報がない，ということになります。なぜなら，たとえばクラスの全員がテストで 100 点を取ると，分散は 0 です。この時は「どういう勉強をすれば成績が上がるか」とか「成績が低い人をどうサポートすれば良いか」という知見を得ることができません。分散が 0 のデータセットには，個々の違いがないので情報が得られないのです。つまり，分散はそのデータから取り出せる「意味」「情報」の上限という点で非常に重要な指標なのです！

ちなみに分散の式は，次のように展開できます。

$$\begin{aligned}
s_x^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \qquad \textstyle\sum \text{記号は分配できるから} \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\frac{1}{N}\sum_{i=1}^{N}x_i\bar{x} + \frac{1}{N}\sum_{i=1}^{N}\bar{x}^2) \qquad \text{定数を N 回足すのは N 倍と同じなので} \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\frac{1}{N}N\bar{x}\sum_{i=1}^{N}x_i + \frac{1}{N}N\bar{x}^2) \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\bar{x}\frac{1}{N}\sum_{i=1}^{N}x_i + \bar{x}^2) \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\bar{x}\bar{x} + \bar{x}^2) \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\bar{x}^2 + \bar{x}^2) \\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \bar{x}^2
\end{aligned}$$

て計算するときなどは，先に全ての変数を二乗したり，変数の平均の計算をしたりしますので，こちらの方が計算が早くなるかもしれません。もっとも電子計算機の発展している現代において，手計算で分散を求めることなんか少なくなりましたけどね。

■標準偏差 ところで，分散は元のデータを二乗した値から計算しているのでした。ですから，その単位が元の尺度と異なることに注意が必要です。つまり，元が cm 位のデータであれば，分散は cm^2 の単位になっているのです。ですから数字が必然的に大きくなりがちですし，分散の大きさが実際どれぐらいなのか，という実感が持ちにくいという欠点があります。

そこで，単位を元に戻すことを考えます。二乗して大きくなっているのですから，平方根をとってやれば元に

戻りますね。

$$s_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

このようにして計算された s_x のことを**標準偏差 (standard deviation)**(SD) と言います。これはルートを取るだけですので，すぐに計算できますね。さきほどの例ですと，A 群の SD は 0.7071，B 群の SD は 2.6925 です。

■四分位数 (しぶんい) と IQR　他にも範囲を表現する指標はあります。分散や標準偏差が平均値に関する指標であったのに対し，中央値に関係した指標として**四分位 (Quantile)** というのがあります。文字通り，データを四分割した時の切れ目にあたる数字のことです。中央値はデータを大きい順に並べ替えて，ちょうど真ん中になる数字という話でしたが，四分割した場合は図 5.6 にあるように，下の方が第 1 四分位，上の方が第 3 四分位ということになります。中央値は第 2 四分位と同じことになるのですね。

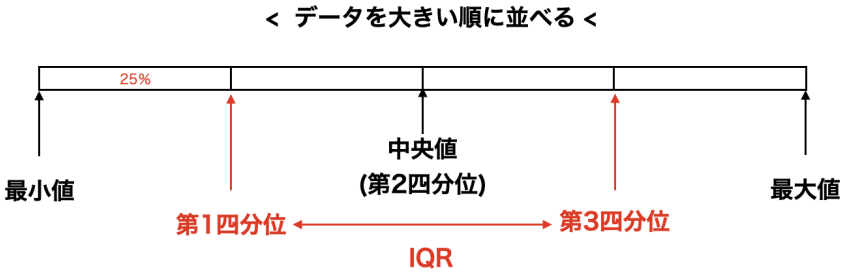


図 5.6　四分位と IQR

この第 1 と第 3 をつかった幅，つまり第 3 四分位数から第 1 四分位数を引いた数字のことを**四分位範囲 (Inter-quantile Rangel IQR)** と呼びます。中央値と同じように外れ値の影響を受けにくく，また標準偏差のように元データと同じ単位ですので，理解しやすい指標です。さてここで，四分割じゃなくて五分割がいい，と思った人もいるかもしれません。いや三分割でも何分割でもいいんですが，要するになんらかの割合で区切った値というのを考えることは可能ですよね。この「割合に応じた値」のことを，**パーセンタイル (Percentile)** と言います。たとえば第 1 四分位は 25 パーセンタイル，第 3 四分位は 75 パーセンタイルと同じ意味です。この表現が使えるのなら，任意の数字で分割できるのでいいですね。分布の形状によっては，95 パーセンタイルとか 5 パーセンタイルといった値を表示することもあります[*8]。

ところで第 4 章で，ボックスプロットについて言及しました。みなさんはもしかすると**箱ひげ図 (boxplot)** という名称で，高校までの数学で学んだことがあるかもしれません。この図示の方法は，データがどのぐらいの幅を持っているかを箱とひげで表現しています。箱の真ん中にあるのが中央値，箱の下のラインが第 1 四分位，上のラインが第 3 四分位を表すのが一般的です。ですが，髭に当たる部分については 2 通りの書き方があります。図 5.7 にあるのは同じデータですが，ひげの長さを「最大値から最小値まで」とする方法 (上段) もありますし，「IQR の 1.5 倍まで」とする方法 (下段) もあります。どちらの方法で描画されているのか，注意するようにしてください[*9]。

[*8] テスト理論業界では，**スタナイン (Stanine)** という区切り型もあります。全体を均等に 9 区分するのではなく，上から順に 4%,7%,12%,17%,20%,17%,12%,7%,4% で区切ります。一桁で表現できる均等な区切りであり，この不均一な数字は学力の分布が正規分布に従うことから来ています。詳しくは，Shojima (2022) を参照してください。

[*9] 高校数学，引いてはセンター試験などでは最大値から最小値まで表す，という記法が使われているようですが，データ分析の文脈では IQR の 1.5 倍までをひげにして，それ以上の数字は外れ値として点で表す方法が一般的です。どちらが正しいとも言え

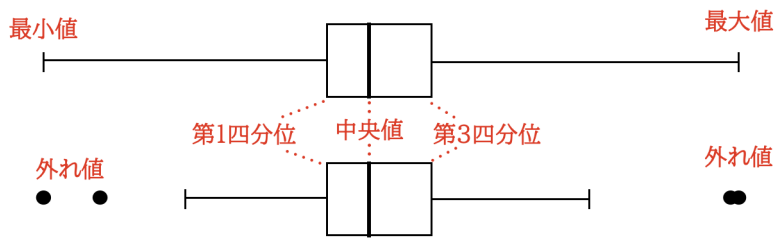


図 5.7 ボックスプロット (箱ひげ図) の 2 通りの書き方

■最大値，最小値，範囲 次に最も簡単な散らばりを示す数値を紹介します。**最大値 (Max)** はデータのなかで一番大きな値を指します。**最小値 (Min)** は逆に，データの中で一番小さな値です。**範囲 (Range)** は最大値から最小値を引いたものです。これは単純ですね。

さてここまでで，中心化傾向の指標と散らばりの指標について説明してきました。冒頭でお伝えしたように，これらの代表値は 1 つだけで表現できることに限界があります。論文などでは，取得したデータについての記述統計量をまず記載することが基本です。データの平均値，中央値，標準偏差，最大値と最小値などを記載することで，どういうデータだったかが (データ点すべてを書かなくても) 見えてきます。逆にいうと，平均値だけではデータの真ん中を示すだけですし，歪んだデータになっているかもしれないことが読み取れません。ちなみにデータが左右対称でない場合，よく使われる統計的分析手法が適用できないこともあります。それを無視してデータを分析すると，間違った結論に辿り着くかもしれないのです。データをさまざまな角度でみる (描画し，複数の代表値を確認する) ことを厭わないでください。幸い，多くの統計パッケージはすぐにこれらの指標を算出してくれますので，計算の苦労はありません。

5.3 データの相対的位置

さてデータ全体を代表値で見ることができるようになりました。ここで平均値と標準偏差を見ることができるようになりましたが，これを使って各データ点の相対的な位置を表現することを考えましょう。

心理学の研究対象の多くは，絶対ゼロ点を持たないものです。たとえば性格というのを考えると，とある性格特性が「ない」ということはありません。もちろん他の人とくらべてより明るい人だとか，より勤勉な人だとか，より情緒が不安定な人だ，といったことはあります。しかしこれらはいずれも，多くの人の中での相対的な位置であって，明るい人でない＝暗い人ではあっても，明るいか暗いといった次元に存在しない人，というのはありません。すなわちデータの特徴として間隔尺度水準であることがせいぜいなのです。そうした他者との比較において考える時，相対的な位置を表現する方法を考えなければなりません。

ゼロ点が存在しないので，相対的な位置を決める中心点を平均値におくとしましょう。平均値に比べてその人がどの程度離れているか，というのは平均偏差ですから $x_i - \bar{x}$ で表現できます。表 5.3 にとある学生の成績のデータを載せてみました。これをみると，この人の素点は英語が最も高く，この人は英語が得意なんだな，と思うかもしれません。でもクラス平均と比べてみると，あらびっくり，むしろ平均より低いので，相対的に英語は苦手なんじゃないか，ということがわかります。素点だけでは困りますよね[*10]。さて，偏差を見ると国語

ないのですが，今後は後者の書き方に出会うことの方が多いと思いますので，注意を促すために書いています。

[*10] 私の娘が通っていた中学校では，定期試験の結果を表にまとめたプリントを親に見せてくれるのですが，本人の素点と本人の平均点しか表示されておらず，まったく無意味な数値の羅列でした。学校に抗議するよう娘には伝えたのですが，卒業まで決して直ることはありませんでした。

と数学のどちらもその平均偏差は +14 です。なるほど，国語と数学は同じぐらい出来がいいのだな，と考えたいところですが，果たしてこの考え方もあっているでしょうか？ 実はこれでもまだ足りません。というのも，

表 5.3 成績の比較

科目	成績 X_i	クラス平均 $\bar{X}$	偏差 $(x_i - \bar{x})$	SD(s_x)
英語	86	90	-4	8
国語	67	53	14	10
数学	44	30	14	5

このクラスの標準偏差をみると，国語と数学では倍ほど違いがあります。標準偏差の違いは，データの散らばり方の違いです。図 5.8 にプロットしてみました。このテストの点数が，左右対称の分布をしているとした時の例です。

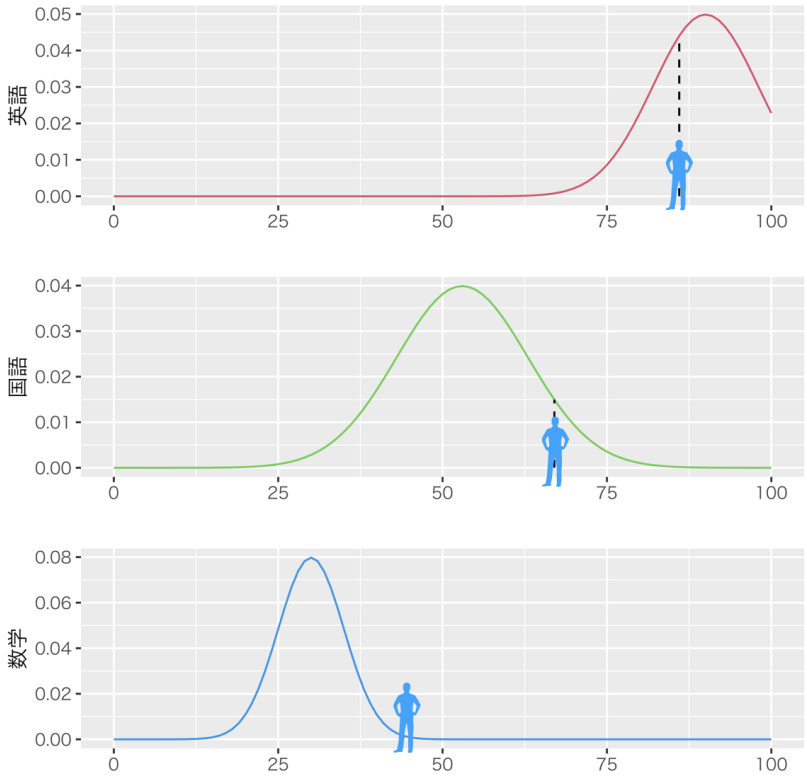


図 5.8 データの散らばりと相対的な位置

図 5.8 の最上段には英語の分布とこの人の位置が書いてあります。分布のピーク＝最も人が多いところは 90 点ぐらいのところにあり，この人はそれの左側ですから相対的にやや劣ることがわかります。二段目・三段目に国語と数学があり，相対的に優れていることがわかります。同時に，国語の分布はやや幅広く，数学は平均点が低く散らばりの幅も少ないことがわかります。国語は幅広ですから，この人よりも高い点数を取る人もいるようです。一方数学は幅が狭いため，この人よりも高い点数を取れる人がほとんどいません。ということは，かなり上位の数学力を持っていることになります。このように，相対的な位置を考えるときはこのデータがどれぐらい広がっているかを考えなければなりません。そこで平均からの偏差を標準偏差で割り，標準偏差いくつ分はずれているのかを考えましょう。この点数を z_i とし，次のように計算します。

$$z_i = \frac{(x_i - \bar{x})}{s_x}$$

このスコア z_i を**標準化スコア (standardized score)** と言います。もっと簡単に z スコアということもあります。このように標準化したスコアでこの人の成績をみてみますと，英語は -0.5，国語は $+1.4$，数学は $+2.8$ となります。国語の優秀さよりも数学の優秀さのほうが際立っていることが，数字で表されましたね。

この標準化されたスコアは平均が 0 ですし，標準化されたスコアの分散は 1 になっています (標準偏差も 1 です)。平均を取り除いて標準偏差いくつ分なのか，という表現にしたのですから当然ですね。言い方を変えると，この数字からは平均や分散を取り除いたことで，単位の大きさに左右されずに複数の変数で比較できるようになったということです。先ほど，国語と数学の点数を比較しました。素点だと平均や分散によって変わってくるので，それを取り除いたことで比較できたのですが，これは国語が 200 点満点，数学が 1000 点満点だったりしても可能です。相対的な位置だけの問題になっているからです。他にも，身長と体重と言った単位がまったく違うものを比較できます。心理学のデータは単位がよくわからないのですが，このスコアにすると相対的にどれぐらい「明るい人」「真面目な人」なのか，ということがわかり，比較できるようになります。標準化スコアは心理学研究の中でなくてはならないものなのです。

またみなさんは**偏差値 (deviation score)** というのを知っていると思います。学業成績に関係してくるアレですね。偏差値は，実はこの z スコアと同じです。ただ z スコアだとマイナスになったり，単位が小さくなったりするので，平均 50，標準偏差 10 になるように変換して，

$$\text{偏差値} = 10z_i + 50$$

という計算で算出します。今回の例でいうと，この人の偏差値は英語で 45，国語は 64，数学は 78 ということになります。

5.4 標準化された差

標準化されたスコアの特徴は，単位がないことです。たとえば身長は cm，体重は kg で表現できますが[*11]，当然のことながら 170cm，60kg というデータを見て「わあ 170 のほうが 60 より大きいね」と思う人はいないわけです。単位が違うものを比較してはいけません，というのは算数の初歩で習うことですね。

しかしこの数字も標準化して，身長の z スコアが 0.3，体重の z スコアが 1.8 だとすると，これは比較できます。標準化すると単位がなくなり，平均から標準偏差 0.3 分，1.8 分大きいわけですから，「体重の方が平均からの偏差が大きい」すなわち「身長の割に体重が重い＝おでぶちゃん」ということがわかります。0.3 と 1.8 を比べて考えられるのです。また標準化されたスコアがマイナスであれば，平均より下回っているわけですから，たとえば身長と体重の z スコアが-0.3 と $+1.8$ であれば「身長は平均より低く，体重は平均より高い。しかも体重の平均からの偏差のほうが大きい」すなわち「かなりおでぶちゃん」ということもできるのです。

さて心理学では目に見えないものを測ります[*12]。抑うつの程度，自尊心の高さ，攻撃性の強さなどは，しっかりと計算されて開発された**心理尺度 (Psychological Scales)** で数値化されるのです。しかしこれらには単位がありません。抑うつの程度がセンチメートルやキログラム，デシベルやルクスやジュールで表現できるはずがないのです[*13]。スコアリングするときは，非常に多くの人からデータを集め，尺度点の分布を描いて

[*11] 日本にはポンド・ヤード法を使う人はいないと思います。世界的にもメートル法，CGS 単位系で表記することがスタンダードで，心理学の論文もそれに倣っています。かつて Pockemon Go!が米国で流行した時に「Pockemon Go の一番の功績はアメリカ人にメートル法を使わせたことだ」と言われたことがありました。

[*12] どうやって測るの?!ということについては次回あるいは 2 年次の「データ解析応用」でお話しします。

[*13] 痛みの単位を 1Hanage とし，鼻毛を一本抜いた時の痛みと定義する，というジョークがありますが，痛みのような日常的なもの

偏差値のような形で表現し，相対的に比較することになります。また標準化することで異なる測度の比較もできますから，「抑うつの程度が高いと攻撃性が強くなる」といった仮説を検証することもできるようになります。心理学にとって，標準化はなくてはならないプロセスなのです。

また群間の比較に際しても標準化プロセスは有用です。

たとえば，男性と女性の平均身長を考えてみたとします。男性の平均身長が 170cm，女性の平均身長が 165cm だということが調査の結果わかったとしましょう。その差は 5cm です。これが大きいと見るか小さいと見るかは，もちろん状況によりけりだとは思いますが，事実としてそうなっていることが示されたのであれば，一考の価値がある量だということになります。しかしこのような話をすると「え，でも私の周りには背の低い男性も多いよ」とか「俺の姉ちゃんは俺よりでかいよ」というような声が聞こえてきそうです。この手の問題は，所謂「主語が大きい」話というやつで，「男性が」「女性が」という主語で話し始めると「そうでない男性もいる」「該当しない女性もいる」という意見・感想を持ちやすいわけです。しかしデータも感想も間違っているわけではありません。集団全体で見た時に差があるかどうかという話と，個別の話をした時に差があるかどうかという話はレベルが違うので一概に比較することができるものではないからです。

また，こうした「実感の差」が出るのは群の分散が大きい時です。図 5.9 にはいくつかのパターンを描いてみました。左上の図は群間の差が大きく，群の分散が小さい場合です。このようなときは差がはっきりとみて取れますし，群の重なりも小さいので「差がある」といわれたらなるほど，と思います。しかし同じ差があっても，群の分散が大きいとどうなるでしょうか。左の下には群間の差があるものの，群の分散も大きい場合を示しました。このときは重複も結構ありますから，「群間に差がある」という話をされても「でも違うこともあるよね」と感じやすいでしょう。右列は差が小さい時で，上段のように差が小さくても分散が小さければ「例外もあるけど差があるね」となりがちですが，右下のように差が小さく分散が大きい場合は例外も多く観測されますから実感が持ちにくいでしょう。

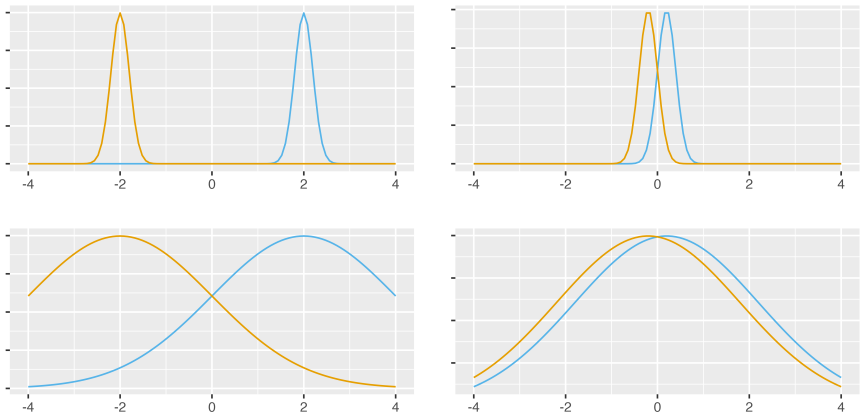


図 5.9 差のパターン。左列は右列に比べて群間差が大きく，下段は上段に比べて分散が大きい。

男女差の話などは典型的な右下のケースで，集約すると差があるのですが個人差が大きく，「男性は力持ち」だとか「女性は少食」などというとすぐ炎上するでしょう。事実として平均値差はあるのですが，それだけでは十分に表現できていないことが多すぎるのです。

ではどうすれば良いのでしょうか。ここで標準化がまた役に立ちます。つまり群間の差は標準偏差に比べてどれぐらい大きいのか，ということを考えればいいのです。この標準化された差のことをとくに**効果量 (effect**

であっても単位はないのです。ちなみにこうしたジョークだけを集めた「新しい単位」(世界単位認定協会, 2017) という本がありましたが絶版になってしまいました。残念。

size) と言います。

$$ef = \frac{\bar{X_1} - \bar{X_2}}{s_x}$$

図の例では差が大きいときその差を 4, 小さいときは 0.4 としています。標準偏差は 2 と 0.2 です。効果量で考えると, 左上は $4/0.2 = 20$, 左下は $4/2 = 2$, 右上は $0.4/0.2 = 2$, 右下は $0.4/2 = 0.2$ となります。左上と右下では意味が全然違うというのが数値化できましたね。

後期になると群の差を統計的にどのように検出するか, という話がでてきます (→ 第 17 講)。統計的に「差を検出する」というときは, 差があるかないかがわかるのですが, 図 5.9 のどの場合でも差はゼロではないですよね。問題は効果量, つまり標準化した時の差を考えることが重要です。もちろん単位がある状態で, 実質的な大きさを検証できればそれに越したことはありません。

実際にデータ分析するときは, 可視化したり単位・分布に気をつけながら考えること, 表面的な指標だけで考えないようにすること, が重要なポイントです。

5.5 課題

■代表値を算出する データ $X = \{17, 15, 30, 30, 43\}$ があったとき, 平均値, 中央値, 最頻値はそれぞれ何になりますか。

■標準化の計算 次の表 5.4 の表の空欄をすべて埋めてください。

表 5.4 二変数 X と Y

ID	X	Y	X の平均偏差	Y の平均偏差	z_x	z_y
1	149	37	-2.2	-12.4		
2	137	52				
3	164	67				
4	159	43				
5	147	48				
平均						
分散						
標準偏差						

第 6 章

測定の基礎

この講義の目的である心理統計は，**「見えないもの」をみよう・測ろう・考えようとする**学問です。見えないものを見るためには，測定に対する考え方，方法論の整備が必要です。ここでは測定に対する考え方と基本的なモデルについて解説します。

6.1　テストと心理学

心理学の研究対象である心は，見えないし触れることもできません。なんなら存在するかどうかもわかりません。我々は自分も含めて，他者，人間の中には「心がある」と思っていますが，物質的なものとは違って本当に存在するのかどうかを確認する方法がありませんから[*1]，いわば「心がある」と考えることにしているにすぎません。逆に言えば，それで存在することで合意するなら，ペットにも心があると言っていいでしょう[*2]。ともかくここでのポイントは，心の存在は「みなし」あるいは「仮定」にすぎない，ということです。心とは，機能的な面を表現した**モデル (model)** なのでしょう。

モデルとは，模型とか理想形という意味でもありますが，ここでは抽象的に考えられた機構 (メカニズム) のことを指します。人の心は，人を理解するために我々が想定したモデルだというわけです。モデルはそれがどのように振る舞うかを記述したものですが，記述の方法が数学的であれば**数理モデル**と言いますし，確率を取り入れたモデルであれば**確率モデル**，統計的な指標同士の関係を記述したものであれば**統計モデル**などと言います。記述は自然言語でも構いませんが，第 1 講でもお話しした通り，関係を数式で表現する**数学的現象主義**は近代科学の基本的なスタイルです。数学的に表現することで，厳密さや抽象化，一般化が進みます。

さて心理学の研究パラダイム，あるいは心理学の基本的な心のモデルとは何かといえば，第 1 章の図 1.1 に示した，**SR 図式**です。刺激と反応の組み合わせから，ブラックボックスである人の中に「心」というモデルがあることを考えるわけです。この刺激と反応のペアについて，心理学はさまざまな方法を試してきた学問であるといえるでしょう。一方で，同じような枠組みで見えないものを見ようとする試みがあります。みなさんにも馴染み深い，「テスト」と呼ばれるものです。

みなさんはこれまで，さまざまなシーンでテストを受けてきたと思います。そのほとんどは，学校関係での学

[*1] 目に見えている物体だってそこにあるかどうかわかりません。カントはその著書「純粋理性批判」(カント, 1787 篠田訳 1961) の中でものが見えて，存在するとはどういうことかを考え続けました。興味がある人は入門書として Kurosaki (2000) などを読んでみるといいでしょう。

[*2] ペットは家族の一員である，などと言いますが，それはその動物にも自我があると家族全員がみなしてふるまっているのですから，少なくとも社会的実在としてその家庭の中ではペットに自我があることになります。もっとも，動物嫌いの人にとっては他人の愛玩動物であっても獣の一種にしか見えないでしょうし，法律的にも愛玩動物を傷つけることは傷害罪ではなく器物破損罪として扱われる事実があります。

力テストでしょう。ところで改めて考えると,「学力」とはなんでしょうか。文部科学省は学習指導要領という教育の指針のなかで, さまざまな力を育てることを謳っていますが,「学ぶ力」とか「生きる力」といったものは明確に定義できているようで, よくわかりません。ただ少なくとも, 学力テストでは学力を測っているということだけは間違いありません。学力は目に見えない, 個人のうちに秘めたる力であり, これを測定するためのテストは心理学とも深い関係を持っています。

学力テストは, 紙とペンを使って行うのがほとんどです。テスト問題用紙には, さまざまな設問が用意されており, 受験生はそれに回答していきます。回答は正答と誤答に区分され, 正答数が多いと学力が高く, 正答数が低いと学力が低いと判断されます[*3]つまり, 同じ人に対するさまざまな設問という刺激を与え, 正答という反応がたくさん帰ってくると, その人の中に設問を正しく捉えて判断するための「知識」「学力」「能力」という一貫した性質が保持されていると考えているのです。反応のパターンが同じであるということは, ブラックボックスにその性質が宿っていると考えるのは, 心理学のモデルそのものですね。

6.1.1 人の性格についての考え方

たとえば心理学では, 人の性格 (人格) というものを「個人の持つ安定的な行動傾向」と捉えたりします。この「安定的ななにか」がある, と考えるところが重要で, 安定した性質であれば, 複数の刺激に対して同じパターンの回答を返すだろう, ということが予想できます。性格検査はそのような目的で作られています。

性格心理学は, 古くは外見などの見た目で人を判断するところから始まりました。筋肉質な体つきだから真面目だろうとか, 太っているから朗らかだろう, という思いつきのようなところから始まっています。ガルというひとが「骨相学」を考えましたが, これは頭蓋骨の形を見て性格を判断しよう, というものです。これは偏見や差別にもつながる話であり, 今では完全に否定された考え方ですが, 歴史的にはそうした時代があったのです[*4]。ところで, こうした分類で考える方法を類型論といいます。類型論は話が単純になっていいのですが, このカテゴリーに収まらない例外もすぐに思い付いてしまいます。そこで性格心理学では, 類型論から特性論へとシフトしてきました。特性論とは, 性格のような人を特徴付ける性質は人類全体で共通していて, その程度が個々人によって違うだけだ, と考えるのです。たとえば明るい人, 暗い人というのがいますが, これは「外向性」という性質が高い人なのか低い人なのかという違いであって, 暗い人は外向性スコアの低い人, 明るい人は外向性スコアの高い人, という相対的な関係で理解するのです (小塩, 2020)。

このような特性を測定するのが, 性格検査と呼ばれるものです。これはテスト理論の性格版ともいえます。学力テストと違うのは, 測りたい目に見えないものが一種類ではないということです。たとえば国語のテストでは国語の能力を測ります。算数のテストでは算数の能力を測ります。つまり測りたいものが 1 次元なのです。これに対して性格は, さまざまな行動に反映される個人の特性ですから, それが何次元あるのかが問題になってきます。性格検査はその歴史と統計的な分析の結果から, 性格は 5 次元で測るのがよかろう, とひとまず結論づけられています。ここでは学力テストと性格テストが同じ目的を持ったツールであること, ただし主眼点がことなることを踏まえておいてください。

6.1.2 良いテストと悪いテスト

ところで, 学校で行われる試験の中には, 学力の程度を明確に測定することを目的にしたものもありますが, 授業内容や基礎的な知識を有しているかどうかといった理解度・達成度を目的にしているものもあります。小学校の低学年などはえてして後者の目的で作られており, 満点が出ることも少なくありません。ここで

[*3] 実際には, 正答数を数えるだけではなく, 問題の重要度などに応じて配点を変えて合計することがあります。

[*4] 詳しくは渡辺 (2019) を参照。

は前者の，学力の高さ低さを精緻に測定することを目的にしたテストについて考えていきます[*5]。

さて，目に見えない能力を正確に測定することを目的にしたテストを考えた場合，理想はどういうテストになるでしょうか。考えやすくするために，理想の逆，最悪のテストとは何か，ということを考えてみましょう。みなさんが今まで受けてきた試験で，これは良くないテストだな，と思ったことはありませんか？ それはどのようなものでしたか？回答例はいくつか挙げられると思います。1 つは「不公平なテスト」でしょう。同じ回答をしていても，ある人は正答と判断され，ある人は誤答と判断される，というのでは困ります。これは問題と採点者の両方が原因です。良いテストはまず問題に対して，正答が厳格に定義されていることが必要です。そして採点者は，回答者が誰であってもその「正答の定義」に則って正しく判断する必要があります。記述・論述試験は，正答の定義や採点者の裁量に影響されることがありますから，本来あまり望ましいテストの形式ではありません。

他にも，「問題が間違っていて答えがない」ものも良いテストとは言えないですね。こうしたテストは全員誤答になるか全員正答になるしかありません。その場合，テストの点数は受験者全員で一律になり，**分散 (variance)** が 0 になります。分散が 0 というのは，そこから情報を得ることができないので，よくないテストだということになります。

逆に言えば，良いテストとは評価基準が一律で，誰がどこで誰に対して採点しても正答・誤答の判断が一意に定まり，かつ，そのテストから得られる分散がなるべく大きいものであること，と言えます。もしテストが一問しか作れないのであれば，受験生の半数が正答し，残る半数が誤答するのが良いテストです。それが分散を最大にするからです。みなさんは Akinator[*6]というのをご存知でしょうか。古くからある 20Quesutions(20 の扉) というゲームのオンライン版なのですが，これはいくつかの設問に対する反応から答えを当てるというゲームです。Akinator は，我々が考えたことに次々と質問してきます。それに答えていくと，20 問と行かないうちに当てられることがほとんどです。この Akinator の出題方法をみていればわかりますが，その問いは「出題者の考えそうな問題空間」を二分の一に分割していくようなものを選んでいます。たとえば第一問は「女性ですか？」と聞かれたりします。これに Yes と答えるか No と答えるかで，少なくとも半分に絞り込めますね[*7]。次に「名前に漢字が入っている？」など，さらに回答を絞り込むようにしています。この絞り込み方が 1/2 のペースでいけば，一番効率よく世界を狭められるのです[*8]。人間がこの手のゲームをやると，下手な質問としていきなり答えを特定するような質問をしたりします。たとえば 2,3 問目に「あなたの想定した答えは “綾瀬はるか” ですか？」と聞いて，Yes ならいいのですが No であれば何も分かったことになりません。綾瀬はるかでない何かのほうが，選択肢が多いからです。良いテストとはこのように，世界を半分にわけ，それをさらに半分にわけ･･･と段階的に区分できるような問題が次々あるもの，ということになります。もし 1 回の試験でさまざまな区分を作りたいのであれば，事前に分かっている問題の難易度順に 1,2,3... と並べ，正答の程度が回答者に応じて分布するようにすれば大きな分散＝情報を得ることができます。

このように，テストは目に見えないものを測定するためのモデルであり，それを数理的に記述して特徴を明

[*5] いわゆる心理テストと銘打って，「うさぎさんを選んだからあなたは心の底では助平なのです」といったお遊びとは違うことに注意してください。そうしたお遊びは，箔をつけるために心理学という学問の名前を利用しているにすぎず，その分類や予言?にいかなるデータも根拠もありませんし，ましてや何かを測定することを目的にしたものではありません。そのような遊びをする人はよく，これは統計だから，と言ったりしますが，科学的な手続きが取られたものでもなければ，データとして蓄積されたものが数量的に評価できず，統計的に分析できるわけでもありません。いちいちこのような断りを入れなければならないのは，個人的には甚だ迷惑で，こちらの営業妨害だと言いたくなります。やれやれ。

[*6] https://jp.akinator.com/

[*7] もちろん性別は 2 種類に限りませんし，Akinator への回答選択肢は「はい」「いいえ」「わからない」「多分そう/部分的にそう」「多分違う/そうでもない」になっていますが，ここでは話を簡単にするために回答を Yes/No の二択に限定しています。

[*8] Akinator を実際にやってみると，二問目に「Youtuber？」とか「闘う？」とか聞かれたりします。これは Akinator の持っている回答データベースがこれに偏っているか，あるいはエンターテイメント性を出すためにあえて 1/2 に拘っていないかといった可能性が考えられます。

らかにしようという営みがあります。これを**テスト理論 (Test Theory)** といいます。次に**古典的テスト理論 (Classical Test Theory)** について解説していきます。

6.2 測定のモデル

古典的テスト理論は，テスト理論のスタートとなった (今とは古典である) 理論です。そのモデルは非常に単純な形で表現できます。古典的テスト理論の測定モデルは次のようになります。

$$X = t + e$$

なんとシンプルなのでしょう。ここで X とはテストの点数を表します。また t は**真のスコア (true score)**，e は**誤差 (error)** と言われます。t は本当に測りたいもの，たとえば本当の学力，真の実力といったもの，と思ってください。古典的テスト理論のモデルが言っていることは，テストの点数は本当の能力を反映してはいるが，それに誤差がついたものになっているよ，ということです。この**誤差と真のスコアの分離**が重要なポイントです。たとえばハードサイエンス (物理学など) では，測定は心を持たないものを対象にし，また測定精度の高いセンサー等で測定しますので，測定した値は大体その通りで間違い無いと考えます。体重計が 83kg だったら体重は 83kg であって，60kg や 90kg ということはありえそうにありません。ですが，学力や性格のようなものは，心を持った (と想定される) ものの反応です。人間は嘘をつくし，間違えるし，易きに流れる生き物ですから，外向的なふりをして答えようと振る舞うこともできますし，うっかり計算ミスをしてしまうこともありますし，「あなたはだらしないところがあるのではないか」といわれたら「そうかもなあ」とおもって反応してしまうかもしれません。そういう意味で，表に出てきた回答（反応やテストの得点）はそのまま信用できないのです。本当は t なんだけど，なんらかの事情である e が加わったものが，目に見える X なんだよ，と考えるのが，このモデルの意味するところです。

測定されたものをそのまま信用しない，といえば聞こえが悪いですが，測定の精度はそれほど高くないのだということを自省的に取り込んでいるという意味で，ソフトサイエンスの良心があらわれているとも言えます。みなさんは人を対象に研究をするのですから，この点を強く自覚しておいてください[*9]。

6.3 誤差の基本的性質

ここで誤差のことを考えてみましょう。誤差は本当に測定したいものからずれている分，を表します。たとえば体重 80kg の人がいたとして，その人が体重計に乗ったら 81kg だったとします。このずれた 1kg がこの測定の誤差になります。ところで，この体重計に誰も乗っていない時に，その目盛りが 0.5kg を表示していたらどうでしょうか。つまり，何もなくてもプラス 0.5kg になるようにずれているわけですから，誰が乗ってもプラス 0.5kg のずれが常に生じることになります。これは体重計が壊れているので，ゼロにリセットできるように調整が必要です。このような，常に生じる誤差のことをとくに**系統誤差 (systematic error)** といいます。このような誤差が紛れ込むのはある程度仕方がないものかもしれませんが，常にずれているのであれば，その分を差し引くなどして補正できる誤差です。

ではこの人の，残りの 0.5kg のずれはなんでしょうか。たとえば測定の前にたくさん食べたとか，たまたま厚手の下着を身につけていたとかいった，その場限りの一貫性のない誤差というのもあり得ましょう。こう言った

[*9] 個人的な経験ですが，他分野の人，とくに理系領域から来られた人と話をする時に，測定されたものを盲信的に使おうとする態度に驚かされることがあります。心理学者としては当然「回答者が答えたからと言ってその通りとは限らない」と考えるのですが，回答がそのまま正しいと思う人もいます。それらの数字は間違っているかもしれず，回答者の心理的な要因によって数字がハックされることもある，ということは常に注意した方がいいでしょう。数値目標を建てると，裏技的な攻略方法が編み出されるため，測定が意味をなさなくなることがある，ということを Muller・Muller (2018 松本訳 2019) は詳しく論じています。

誤差のことを**偶然誤差 (random error)** と言います。この誤差は，どうしてそのようになったのか，その理由がわかりません。「たまたま」ですから，たまたまプラスになる要因もあれば，たまたまマイナスになる要因もあるでしょう。この偶然誤差こそ，測定の時に入り込む困った問題です。系統誤差は調整して直すことができますが，偶然誤差は傾向がないので対応のしようがないからです。

テストの時に考えなければならない誤差 e は，基本的にこの偶然誤差です。たまたま前日ヤマを張ったところが出たとか，たまたま欠落した単語帳のページに書いてあった項目が問題に入ったとか，体調がたまたま良かった・悪かった，と言った要因など，真の実力が発揮できない要素は必ず入ってくると考えなければなりません。

さてでは，テスト理論のモデルに戻ってこのことを考えてみましょう。もとのモデルを少し拡張して，個人 i についてのテストであるということを明確にするために，X_i という添字をつけて書くとします。

$$X_i = t_i + e_i$$

さて，このテストを n 人のクラスで実施したとします。そこからクラスの平均点が算出できますね。

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

これを展開すると次のようになります。

$$\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(t_i + e_i) && \text{定義より} \\
&= \frac{1}{n}\sum_{i=1}^{n} t_i + \frac{1}{n}\sum_{i=1}^{n} e_i && \text{分配して} \\
&= \bar{t} + \bar{e}
\end{aligned}$$

さて，ここでポイントです。$\bar{e}$ は偶然誤差の平均を表しています。偶然誤差は一貫性のない誤差なので，プラスに出るかマイナスに出るかわかりませんが，**理屈上，その平均は 0 であると**考えられます。これがなんらかの定数 c であれば，それは系統誤差であり補正すべき大きさですが，そうではなくまったく何の傾向もない**偶然にすぎない**のですから。この特徴，$\bar{e} = 0$ を代入すると上の式は，

$$\bar{X} = \bar{t}$$

となります。やったね！すごい！誤差が消えて無くなりました。

このように，測定の平均値は真の値の情報だけになることがわかります。測定すれば真の値にたどり着く，というのが古典的テスト理論のポイントなのです。

ちょっと補足をしておきます。最終的に得られた $\bar{X} = \bar{t}$ というのは，テストの平均が真の値の平均になるということですが，それはクラス全体の平均が本当の能力だというだけで，個人の能力が本当かどうかはわからないじゃないか，と思われるかもしれません。それはそうなのですが，この話を個人と学級という対比ではなく，テストの複数の項目と，合計点としてのその人のスコアだと考えてみてください。すなわち，X_{ij} として，個人 i の項目 j に対する正答，誤答が採点され，最終的に $X_i = \sum_{j=1}^{M} X_{ij}$ と点数が与えられるとします (項目数は M 個)。

ここでも真の能力 t_i と，誤差が考えられますが，誤差は e_{ij} と項目ごとに逐一入り込むと考えられます。すると上の式展開は次のようになります。

$$\begin{aligned}\bar{X_i} &= \frac{1}{M}\sum_{j=1}^{M} X_{ij} \\ &= \frac{1}{M}\sum_{j=1}^{M}(t_i + e_{ij}) \\ &= \frac{1}{M}\sum_{j=1}^{M} t_i + \frac{1}{M}\sum_{j=1}^{M} e_{ij} \\ &= t_i + \bar{e_i}\end{aligned}$$

ここで $\bar{e_i}$ は個人に生じる誤差の平均ですが，これもとくに傾向が考えられないので 0 と考えられますから，$\bar{X_i} = t_i$ と個人の真の能力を測っている，と考えることができます。

6.4 テストの分散と信頼性・妥当性の関係

6.4.1 信頼性

では，今度はテストの分散を考えてみましょう。テスト X の分散を Var(X) という添字をつけて書くとします。

$$\begin{aligned}Var(X) &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 && \text{定義より} \\ &= \frac{1}{n}\sum_{i=1}^{n}((t_i + e_i) - (\bar{t} + \bar{e}))^2 && \text{X をテスト理論のモデルに} \\ &= \frac{1}{n}\sum_{i=1}^{n}((t_i - \bar{t}) + (e_i - \bar{e}))^2 && \text{同じ記号でまとめる} \\ &= \frac{1}{n}\sum_{i=1}^{n}((t_i - \bar{t})^2 + 2(t_i - \bar{t})(e_i - \bar{e}) + (e_i - \bar{e})^2) && \text{展開する} \\ &= \frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})^2 + \frac{1}{n}\sum_{i=1}^{n}2(t_i - \bar{t})(e_i - \bar{e}) + \frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2 && \textstyle\sum \text{を分配}\end{aligned}$$

ここで，第一項 $(\frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})^2)$ と第三項 $(\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2)$ の式が分散の定義式と同じ形になっていますね。これらはそれぞれ，真のスコアの分散と，誤差の分散を表しているわけです。では第二項，$\frac{1}{n}\sum_{i=1}^{n}2(t_i - \bar{t})(e_i - \bar{e})$ は何かというと，真のスコアと誤差との**共分散 (covariance)** の式であることがわかります。詳しくは第 8 講で扱うのですが，ここでは共分散とは 2 つの変数の関係の強さを表す数字だ，と思っておいてください[*10]。さて，誤差は定義上，一貫した傾向のないものです。傾向があると予測できます

[*10] 既に高校までの数学で**相関係数**を習っている人もいると思います。共分散は，標準化されていない相関係数のことでもあります。相関係数のことだと思って読み進めてもらってもいいでしょう。

が，まったく予測できない部分を誤差として表現したのでした。ですから**真のスコアと誤差は関係し合わない**，つまり $\frac{1}{n}\sum_{i=1}^{n} 2(t_i - \bar{t})(e_i - \bar{e}) = 0$ です。これを踏まえて考えると，次の関係が得られます。

$$Var(X) = Var(t) + Var(e)$$

言葉で言えば，**テストの分散は真の分散と誤差の分散に分割される**ということです。テストの分散は「真のスコア」と「誤差」の 2 つの要素で完全に説明され，他の要素が入り込まないという意味です。

さて，分散とは，そのテストから得られる情報の大きさそのものだという話をしました。同じ分散のテストがあれば得られる情報が同じですが，本当に知りたいのは $Var(t)$ の部分なのですから，こちらの量が相対的に大きい方がありがたいわけです。つまり，全体の分散のうち，真のスコアの分散 (真分散) が占める割合が大きい方が良いテストだということができます。この全分散中に占める真分散の割合のことを，そのテストの**信頼性 (reliability)** と言います。信頼性は，テストが測りたいものを測れているかを考える上での 1 つの指標になります。信頼性 Rel は次の式で表されます。

$$Rel = \frac{Var(t)}{Var(X)} = 1 - \frac{Var(e)}{Var(X)}$$

6.4.2 信頼性を評価する

テストの信頼性はどのように評価すれば良いでしょうか。これを考える上で，また誤差の性質がヒントになります。(偶然) 誤差は平均が 0 ですが，誤差の分散はゼロではありません。分散とは散らばりの幅であり，誤差の幅とはすなわちスコアがどれぐらいブレるかの幅を示しています。ここでたとえば体重 80kg の人が体重計に乗って，80.1kg だったとしましょう。このとき誤差が 100g 出ているわけですが，体重計からおりて，すぐもう一度乗ってみたら今度は 79.9kg だった，としましょう。今度は誤差が-100g 出たわけです。自宅の体重計でしたら，100 グラムぐらいの違いがちょくちょく出るのは仕方ないかもしれません。しかし，これが ±1kg だったら，ちょっと誤差が大きいなと思うでしょう。±3kg とか ±5kg だったら，これはこの体重計が「信頼できない」わけです。このように，信頼性とは測定値の安定性とも言えます。というのも，誤差は他の何ものとも相関せず，一環した傾向を持たないはずですから，1 回目の測定の時に出た誤差と 2 回目の測定の時に出た誤差との間にも，相関関係がないと考えられるからです。複数回の測定において，誤差同士が相関しないわけですから，何度か測定するとその誤差の要素が打ち消し合って，真のスコアが得られるのでしたね。

テストの信頼性は，このように，同じものに対する複数の測定の相関係数を計算することで評価できます。あるテストを 1 回行って，少し期間を空けて同じテストをもう 1 回行い，その 2 回のテストの相関係数を算出して十分に高ければ，そのテストは信頼性が高いということができます。このような評価の仕方を**再検査信頼性 (Test-Retest reliability)** と言います。

ところで，気をつけなければいけないことは，私たちが測定しようとしている対象が体重など物理的な特性であればいいのですが，心理学の場合は人間の心理的な特徴であるということです。たとえば学力であれば，テストを 1 回行って，3 日ほど開けて同じテストをしたとすると，受験者は前の問題を覚えていて対策を練っているかもしれません。そうすると，1 回目に測定した対象と 2 回目に測定した対象が厳密には異なることになり，どちらが本当の学力だったのかがわからなくなります。あるいは性格検査の場合，1 回目の回答パターンを回答者が覚えていて，「前と同じ答えにしておこう」と思うかもしれませんし，逆に「前とは違う答えにしちゃえ」と思うかもしれません。このように，測定対象が意図的あるいは無意図的にでも行動を変えることがあり得ます。測定は同じ状態に同じスコアが与えられることが重要ですが，測定対象の方が時事刻々と変わる可能性があることに注意が必要です。もちろん，記憶が薄れるのをまったり，嘘がつけないような仕組みを

考えるなどの対策も必要ですが，毎回それをするというのも大変です。

そこで，同じテストを 2 回するのではなくて，同程度の内容のテストを 2 種類作って実施することを考えます。たとえば計算問題を 50 問だして数学力を測ることとし，それを 2 種類作ってその相関を見れば，その形式のテストの信頼性だと言えます。このようにして信頼性を評価する方法を**並行検査信頼性 (parallel test reliability)** と言います。

ここで難しいのは同程度の内容のテストを 2 種類作ることです。たとえば計算問題のテストで難易度を同じぐらいにするのは，数字を少し書き換えればいいのですからそれほど苦労はしませんが，性格検査などの場合は同じような側面を測っている言葉を探さなければならないので，必ずしも同程度のテストにはならないかもしれません。そこで，同じ 1 つのテストを半分に分割する**折半法**の出番です。これは 1 つのテストだけを作って実施し，そのテストの前半と後半とか，偶数番号の問いと奇数番号の問いとかで任意の半分に分け，それの相関係数を求めることにするのです。こうすると 1 回のテストで信頼性を計算できます。

半分にして検査できるのであれば，さらにその半分にしてみたらどうでしょう。4 分の 1 にしてすべての組み合わせについて相関係数をとって平均化すると，分割の違いによるブレを抑制できて良さそうです。ではその半分なら？ さらにその半分なら？ ･･･ということを考えていくと，結局のところテストに含まれるある設問 (項目) が，他のすべての項目とどの程度相関するかを計算し，その平均を取れば良いのではないか，ということがわかります。この考え方に基づいて計算される信頼性の指標が α **係数 (alpha coefficient)** です。開発者の名前を取って**クロンバックの** α**(Cronbach's alpha)** ともいいます。これは，テストに含まれる項目数を M，テスト全体の分散を V_t，項目 j の分散を V_j と表すと，次の式で表されます。

$$\alpha = \frac{M}{M-1} \times \left(1 - \frac{\sum V_j}{V_t}\right)$$

この方法は項目ごとの相関係数のペアをテスト全体について評価していますから，言い換えるとテスト全体の一貫性を評価していることでもあります。ですからこの信頼性のことを**内的整合性信頼性 (Internal consistency reliability)** とも言います。この数値は 0 から 1 の間をとり，0.8 程度が必要であるという目安もありますが，式を見るとわかる通り項目ごとの分散を小さくし，テスト全体の分散を大きくすると大きくなります。また項目数が多い方が大きくなりやすくなります。

いずれにせよこのようにして，テスト全体の信頼性は数値的に評価できることがわかります。しかしこれで測りたいものがちゃんと測れているかどうかのチェックになっているかというと，残念ながらその答えは No です。

6.4.3 妥当性

測定が安定している，つまり信頼性がある測定ができれば良いことは間違いないのですが，信頼性があっても**妥当性 (validity)** がなければその測定は意味がありません。妥当性とは，測定したい対象をきちんと捉えている程度のことです。たとえば身長を測ろうとして，体重計を使うとします。成長に応じて，身長が伸びますが，それは体重とも関係がありますので，身長の伸びに応じて体重も増えていくでしょう。体重計は安定した計測器で，信頼性は十分あると思いますが，体重計で身長が測れていると言えるでしょうか。相関する変数ですので，部分的に Yes といえそうですが，やはり身長は身長計で測ったほうが良いでしょう。身長計の方が，身長という特徴を的確に捉え，より本質に近い測定をしているからです。

身長の例では，「身長とは背の高さのことである」という定義が明確で，誰しも身長とは何か，という問いに疑問を持たない程度に共通了解のとれた概念であり，誰しもが信じる身長の本質＝背の高さは，身長計で測定できるというのが自明だと思われます。ですが，心理学が測定しようとしている概念は，それほどはっきりした定義，共通了解の取れた概念ではなく，何がその概念の本質かということについてもいろいろ異論があり

そうです。たとえば「自分らしさ」というものが，「自己実現の程度だ」と定義されたとして，さて自己実現とはなんでしょうか。自分のやりたいことを自分でできている程度のこと？ 私は爪を切ろうと思ったら切れるし，お風呂に入ろうと思ったら入れますが，だからといって自己実現ができている，といったら誰しもそれはちょっと・・・って思うのではないでしょうか。もっと下世話に，たとえばコミュニケーション力というのを考えてみましょうか。コミュニケーション力とは，字義通りにはコミュニケーションをする力ですが，コミュニケーションは話し手と受け手が相互にやりとりをすることであり，話し手が一方的に話すだけでもだめ，受け手が一方的に話を引き出すだけでもだめなのではないでしょうか。コミュニケーション力が高いひとでも冠婚葬祭のような厳格な式典の時には黙るでしょう。沈黙を続けていたら，コミュニケーション力が無いと思われても仕方ないのではないでしょうか？ これが屁理屈だと思われる人は，誰しもが納得するコミュニケーション力の定義を考えてみてください。すぐに例外や定義から漏れるシーンを思いつき，明確にならないことがわかるでしょう。

心理学者が測ろうとしているものの多くは，状況や個人差によって大きく変わりうるものであり，その中で安定した (信頼性のある) 測定をするのは難しいことですが，さらにその測定されたものが本当に欲しかった数値であるかどうかについては，信頼性だけでは補償されないのです。信頼性があるのは当然とした上で，本当にそれで測れているのかどうかは，きちんと定義され，共通了解の取れる一般的な概念であって，かつ測定がその定義された状態と合致していることが必要になります。この程度のことを妥当性と言います。信頼性は相関係数などで数値化できますが，妥当性は本質的に数値化しにくく，論証や考察を通じて担保されていると納得するしかありません。**信頼性は妥当性の上限**といいますが，信頼性に加えて妥当性の高い測定をしなければなりません。

6.4.4 妥当性を評価する

妥当性はいくつかの観点から評価されます。ここでは代表的な 3 つの評価軸を説明します。

■構成概念妥当性 妥当性は「それって本当に本質を突いてるの？ 」ということの度合いですから，「本質が何か」というところから考えねばなりません。心理学者が測定したい本質とは心の性質ですから，それをきちんと言葉で定義した上で，その測定が定義に合致しているかどうかを考えることになります。その前に，そもそもそういう心の性質を考えることが適切かどうか，ということを考えるのがこの**構成概念妥当性 (construct validity)** です。たとえばさきほど悪い例としてあげた，コミュニケーション力は，そもそもそれがどのように定義できるか，そういったものを想定する意味があるのかというところから疑念があります。他にも性差を強調するだけの概念とか (たとえば女子力など)，ゲーム脳のようなエビデンスの弱いものなども，構成概念として妥当かどうかが怪しいわけです。構成概念妥当性を高めるためには，先行研究を調べることでその概念が他の概念とどのように違うか，その概念を提示しなければならない理由は何か，といったことが説明できるようにならなければなりません。さもないと，訳のわからない用語が乱立するだけになってしまうので，注意が必要です。

■基準関連妥当性 心理学的な構成概念を考えて，もしそれが本当に有用な概念なのであれば，なんらかの実際的反応や行動に反映されるはずです。そのとき，その概念に関連する他の指標も一緒に変化するはずです。たとえば抑うつ状態というのは「甘えだ」などといって認めない人もいるかもしれませんが，実際に抑うつ症状のあるクライアントは食事ができなくなったり，なにもないのに涙が出たり，という事実があります。それらを総合して，医者が診断を下すわけです。さて心理学の質問紙で抑うつの程度を測定するものを作ったとします。いくつかの項目に回答してもらい，採点することで，点数が高ければ抑うつ状態が高いと判断できるとしましょう。であれば，すでに医学的に抑鬱状態であると診断されたクライアントと，そうでないクライアントにこの質問紙に回答を求めると，点数に差があるはずです。このように外的な基準に照らし合わせた時に，測定

したものがその基準と対応しているかどうかで判断するのが**基準関連妥当性 (criterion validity)** です。

■内容的妥当性 測定項目をたくさん作って，それで測定したものが，測定したい内容を過不足なく含んでいるといえるか。こうした観点から判断するのが**内容的妥当性 (content validity)** です。パーソナリティ (性格) を測定するとしましょう。パーソナリティには色々な側面がありますから，色々な項目で多角的にアプローチしなければなりません。真面目か，気性が荒いか，人と話すのは好きか･･･とさまざまな角度からの項目を作って，取りこぼしの内容にちゃんと捉えることができているかを，チェックする必要があります。内容的妥当性には，統計的な分析をした結果が仮説モデルと合致しているかどうかをチェックする方法[*11]があります。

妥当性はこのように色々な側面がありますし，「たくさんある妥当性概念の妥当性はあるのか」といった議論になることもあります。妥当性についてより詳しくは Grimm・Yarnold (2001 小杉他訳 2016) の第 4 章などにまとめられていますので，一度目を通してみてください。

さて，これらの観点から考えて，きちんと測りたいものが測れているかを検証していきます。妥当性と信頼性の関係は，図 6.1 に示したように射的の例で考えるとわかりやすいでしょう。1 つ 1 つの質問項目が，射的の弾ひとつ 1 つだと思ってください。テスト･項目による測定は，心という目に見えない的に向かって問いかけて当たるかどうかを確かめているようなものなのです。図 6.1 の左端のように，当たっているものもあれば外れているものもある，というのは信頼性も妥当性もないテストです。信頼性は測定の安定度ですから，それがあればいいのかというとそうではなく，左から 2 つ目の状態のように一箇所に集中しているけど中心は射抜いていない，ということがあるわけです。この中心を射抜いているかどうかの程度は妥当性ですが，妥当性があっても信頼性がない (図の左から 3 つ目) 状態でも困ります。やはり，それぞれの問いかけが的をしっかり射抜いている，安定した測定が良い測定であると言えるでしょう。

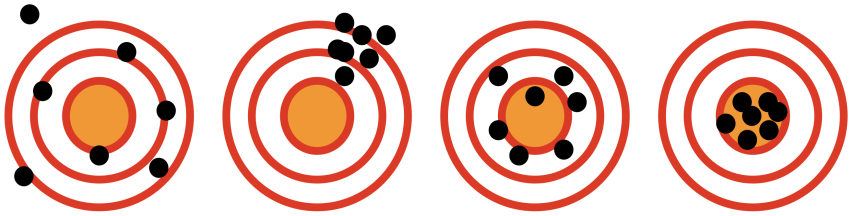


図 6.1 信頼性と妥当性のイメージ。左から 1. 信頼性も妥当性もない，2. 信頼性はあるが妥当性はない，3. 妥当性はあるが信頼性がない，4. 信頼性も妥当性もある，という状態

ただし，心理学の怖いところは，質問を投げかけた相手が人間だということです。人間は間違えるし，嘘をつくし，易きに流れるのです。悪気はなくとも，いや悪気がないからこそ，なにを聞かれているかわからなくても「とりあえず適当に答えておこうか」とか，「1 から 5 のどこかに丸をつけておけばいいだろう」と反応してしまうことがあります。白紙回答が返ってくることは滅多にないですが，本当にきちんと理解し，正しく回答しているかどうかはわかりません。調査用紙に誤字脱字があったというだけで，ヘソを曲げてやる気を無くす回答者もいますし，構成概念が練り上げられていなければなにを聞かれているかわからない (から適当に答えておこう) という反応を誘発することもあります。また，真面目な研究者が自分の仮説にこだわるあまり，ついつい自分に都合の良い結果を見出してしまうこともあり得ます[*12]。

[*11] 因子分析という分析方法によって，次元数がおなじであるか，得られた次元と項目の対応が適切か，といったところから判断することができます。詳しくは 2 年次の応用的授業で扱います。他にも因子分析を使うと，因子的妥当性，弁別的妥当性，収束的妥当性などを検証することができます。

[*12] このことを専門的に**てっちゃんの手品 (Teppei's Magic)** といいます。詳しくは小杉 (2018) を参照。

心理学はまじめに学問をやっており，場を盛り上げるための “心理テスト” などとは違って，綿密に考えられ丁寧に分析されたものだけが概念として後世に積み重ねられていくものなのです。それでなくても，目に見えない深淵なるものが研究対象なのですから，研究法はより精緻に，分析はより客観的かつ正確であらねばなりません。

6.5 課題

■古典的テスト理論 古典的テスト理論のモデル式を書き，その平均と分散がなにを意味するか説明しなさい。

■信頼性と妥当性 テストの信頼性と妥当性を自分の言葉で説明してみてください。

■古典的テスト理論その 2 つぎのテストについての記述のうち，正しいものすべてに丸をつけなさい。

1. 良いテスト項目とは，誰にでも正答できる項目である。
2. 悪いテスト項目とは，受験者の分散を最大にする項目である。
3. 能力や性格といった心理変数は実体が明らかではないので，一見なんの関係もない質問であっても測定に役立つ。
4. テスト得点の分散が大きければ大きいほど，その測定から得られる情報量は大きいと考えてよい。
5. 妥当性とは，測定に先立って，測定しようとする対象概念を想定できることの程度を意味する。
6. 古典的テスト理論は，測定しても必ず誤差の影響が入るため，真の能力は決して測定できないという不可能性を証明したものである
7. 古典的テスト理論からは，テスト全体の分散を真の分散と誤差の分散に分割できることが導かれる。
8. テストによって測定したい真の値と誤差の共分散は一般に正の相関をすると考えられる。
9. 信頼性は，全分散中に占める真の分散の割合で表される。
10. 信頼性は，1-(誤差分散/全分散)で表される。
11. 誤差の平均と分散は 0 である

第 7 章

誤差と確率

今回は確率のお話です。

前回は心理学的な測定の話をしました。そこから急になぜ数学の話になるの，と思った人もいるかもしれませんが，確率と統計は切っても切れない関係にあるのです。心理学における測定には，誤差が生じるという話をしました。この誤差は，目に見えなければどういう振る舞いをするかもわからない，規則性のないものです。ただこの規則性のなさを，どうコントロールするかというときに，確率の概念が出てきます。

確率と呼ばれる数字は，偶然性や毎回の不確実さを表しつつ，その奥にある大局的な性質を記述できます。測定の誤差が積もり重なると平均がゼロになるように，毎回の事実はわからなくても全体的な傾向を記述することはできるのです。

これは大規模調査を行う時にも応用できます。すなわち，個々人の違いはわからなくても，集計して統計を取った時に，その全体の平均的な傾向を掴むことは可能なのです。

偶然性をハンドリングし，全体的な傾向としてそれをコントロールするために，確率の考え方は避けては通れないのです。

7.1　確率と呼ばれる数字

確率と統計については高校数学でもある程度習うでしょう。「場合の数」を数え上げることから始まって，順列，組み合わせ･･･ああ面倒くさい！という印象を持っている人もいるかもしれません。

でもそれと同時に，皆さんはもっと日常的に確率という言葉を使っています。たとえば天気予報で降水確率は 30% だとか，100% 勝てる！といった表現をすることがありますね。

これのどちらも**確率 (probability)** です。確率という言葉で表される数字が従うべき条件は，実は次の 3 つに集約されます。

1. 確率の値は非負（正または 0 の数字）
2. 全標本空間におけるすべての事象の確率の合計は 1.0
3. 任意の 2 つの相互に排他的な事象について，どちらか一方が発生する確率は個々の確率の和である

まず，確率は正の数字ですから (たとえば 10%=0.1 という正の実数です)，1 つ目の条件は問題ないでしょう。

2 つ目はややこしい言葉が並んでいます。実現する出来事のことを事象といい，ここの出来事のことを標本，標本が含まれる世界全体のことを全標本空間と呼んでいますから，ここでは「あり得るすべてのこと」ぐらいに理解しておいてください。つまりこの 2 つ目の条件は「あり得るすべての確率を足し合わせたら 1.0 になる」ということを指しています。だから高校数学での確率は，あり得るすべての事象を数え上げさせて，その中

の相対的な頻度でもって確率としていたのです。でもたとえば，天気予報が「晴れ」「雨」「曇り」の 3 つの状態 (事象) に限定されていて，明日の天気はその中でも「雨」になりそうな度合いが 0.3 ぐらいだというのであれば，それでもいいのです[*1]。皆さんも決意や意志の強さを確率で表現したいことがあるかもしれませんが，その時は心の中のあり得る状態をすべて考えたうち，どの程度の強い思いなのかを表現すれば良いのです。ちなみに天気予報は未来の予測であり，起こりうるすべての未来を書き出して天気の割合を算出しているわけではありません。過去の同じような天気図になった時に，実際に生じた事象を集計したものを参考に，未来に託す想いの強さを確率の言葉で表現しているのです。

さて，第 3 の条件は少しややこしいようです。でも言葉をよく読んでみましょう。「任意の 2 つの相互に排他的な事象」は「なんでもいい 2 つの事象，ただし排他的」ということです。排他的というのは一方が成立すると他方は成立しないというように，他を排除する，両立しない，という意味です。確率の例でいうなら，サイコロの出目で 1 が出るという事象と，2 が出るという事象は相互に排他的です。1 と 2 の面が同時に上を向くなんてことはあり得ないわけですから。これを踏まえて後段を考えると，「どちらか一方が発生する」確率です。先の例で言えばサイコロで「1 か 2 が出る」状態ですね。その場合は個々の確率の和，すなわち $1/6 + 1/6 = 1/3$ ということを説明しているだけです。

この 3 つの条件を満たしていると，その数字は確率と言って構いません。高校までの数学で習った確率は，たとえば「サイコロの目が出る確率は，サイコロを無限回転がしていくとその相対頻度は最終的に $1/6$ に収束するはず」として定義しています。この無限回試行を行う，というのは実際にはできませんが，それぞれの出目の確率が等しいならいつかは $1/6$ にたどり着くはずで，出目の割合がそれぞれ $\{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$ だとすると，非負で，全部足すと 1 で，第 3 の条件も成立しているので，これは確率だと言えます。また「降水確率」などとして使われる「確率」の場合，「今日が無限回繰り返されれば･･･」という考え方は成立しないのですが，「雨が降る確率が 60 %，曇りが 30 %，晴れが 10 %の確率」と表現し，「雨かつ曇り，という状態はない」と考えれば，確率の定義には当てはまっていますから，これも確率です。**この条件を満たす数字が確率であり，確率に関する数学的法則はいずれも同じように作用する**，というところがポイントです。

持って回ったような言い方をしていますが，ここで伝えたいのは要するに「無限回の数え上げだけが確率じゃない」ということです。雨が降る確率とか，お昼ご飯にカレーを食べる確率とか，そういう言い方でも問題ないよということです[*2]。むしろそうした確率の方が馴染みやすい数字ではないでしょうか。ちなみに無限回試行の相対頻度から確率を算出するやりかたで定義する方法を**頻度主義 (frequensist)** と呼ぶことがあります。それに対して，確信できる度合いの強さを表現するものとして定義する方法を**主観確率 (subjective probability)** ということがあります。どちらが正しいという話ではないので，定義に則して考えるとさまざまな解釈が可能だ，ということをここでは抑えておいてください。

7.2 確率分布

ある事象が確率的に振る舞う，あるいは個々の出来事を正確に予測はできなくとも全体的にある傾向に従うことが示せる場合，その事象は確率的な動きをする変数という意味で**確率変数 (random variable)** と呼ばれます。確率変数がどのような振る舞いをするのか，その起こりうるすべての事象と対応した確率を

[*1] 日常用語で使う確率表現の場合は，間違った使い方があり得ます。たとえば元大阪府知事の橋下徹さんは，2007 年の選挙戦前のインタビューで「出馬は 20000% ない」と発言しましたが (しかもそれを数日後に覆して出馬し，当選しましたが，それはおいといて)，これは確率の数字ではありません。同様に，Dr.Stone の主人公，千空の口癖で「100 億 % ない」などという表現がありますが，これも確率の数字とは言えません。博学の主人公にしては残念なセリフです。

[*2] もちろん条件を満たしている必要があるので，橋下さんの表現は数学的に正しくはありません。

網羅的に書き出した一覧のことを**確率分布 (probability distribution)** といいます。たとえばサイコロの出目に関して，次のように表現したとしましょう。

- 1 が出る確率 $\cdots 1/6$
- 2 が出る確率 $\cdots 1/6$
- 3 が出る確率 $\cdots 1/6$
- 4 が出る確率 $\cdots 1/6$
- 5 が出る確率 $\cdots 1/6$
- 6 が出る確率 $\cdots 1/6$

これはすべての出目と確率を対応させた一覧ですから，サイコロの目の確率分布を表していると言ってもいいでしょう。一般的な六面ダイスで説明しましたらからこれでいいですが，20 面サイコロなどになるとちょっと表記が面倒ですね。また，サイコロの出目は 1.5 とか 2.3 のようになることはありません。これは出目が離散的，あるいはカテゴリカルなものと言ってもいいでしょう。尺度水準で言うところの，**名義尺度水準**や**順序尺度水準**の数字であれば，離散的，カテゴリカルなものになるわけです。このように離散的な変数についての確率分布のことを**離散分布**と言います。これに対して，連続的な変数の確率分布は**連続分布**といいます。間隔尺度水準，比率尺度水準の数字は連続分布ですね[*3]。連続分布の場合はとくに，すべての数字を網羅的に書き下すことはできません。そこで分布を数式・関数で表現することが一般的です。関数で表された確率分布のことを**確率分布関数 (probability distribution function)** といいます。

7.2.1 離散分布と確率質量

離散的な確率変数の例をさらにみていきましょう。サイコロを例にとります。サイコロの出目は離散的ですからね。サイコロを 100 回投げてその出目を記録し，グラフにしたのが図 7.1 です。

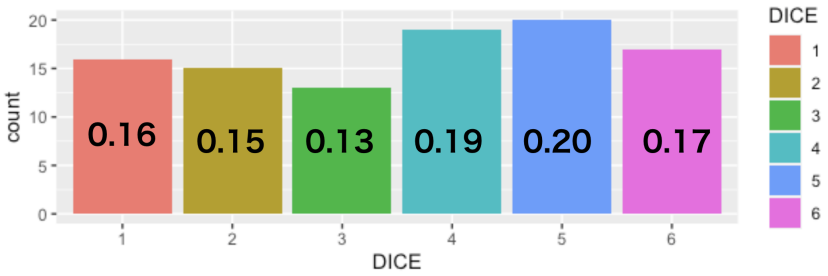


図 7.1 サイコロの確率分布

ここにはすべての出目を数えてその相対頻度を書き表してあります。このような離散的な確率変数の大きさを表す数字のことをとくに，**確率質量 (Probability Mass)** といいます。また相対頻度は先に述べた確率の条件に合致しているので，確率と考えることができます。ですから確率質量の総和は 1.0 になります。

[*3] 連続とは何か，というのは数学的には結構込み入ったお話ですが，直感的には「どのような値と値の間にも中点があるなら連続」と思っておけばいいでしょう。たとえば身長などの連続変数は，160 と 170 の間に 165，160 と 165 の間に 162.5，160 と 162.5 の間に 161.75，160 と 161.75 の間に... とどこまでも行けるわけです (測定精度が出るかどうかはともかく，理論的に)。

7.2.2　連続分布と確率密度数

次に連続変数の場合を考えてみましょう。先ほどと同じく，今度は 1000 人の身長を測定したデータを例にします。身長は連続量だと考えることができます。図 7.2 にそのヒストグラムを書いてみました。このヒストグラムは 1 階級が 5cm 幅で，各階級に何人いるかの数字も同時に書いています。このように階級ごとに区切ったものは，先のサイコロの例と同じように相対頻度にできますから，たとえば 160-165cm の階級の確率質量は 0.165，という言い方をしてもいいかもしれません。しかし「5cm で区切る」ということに深い意味は

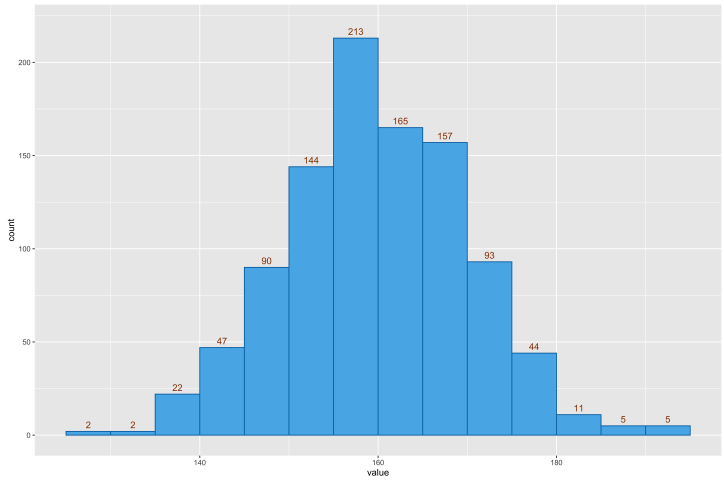


図 7.2　身長の確率分布。bin 幅は 5cm

なく，これが別に 2.5cm でも 1cm でもいいわけです。たとえば同じデータを 2.5cm 間隔で区切りで書き直したのが図 7.3 になります。このように，度数は階級の区間に依存しますので，どれぐらいの幅の中にどれ

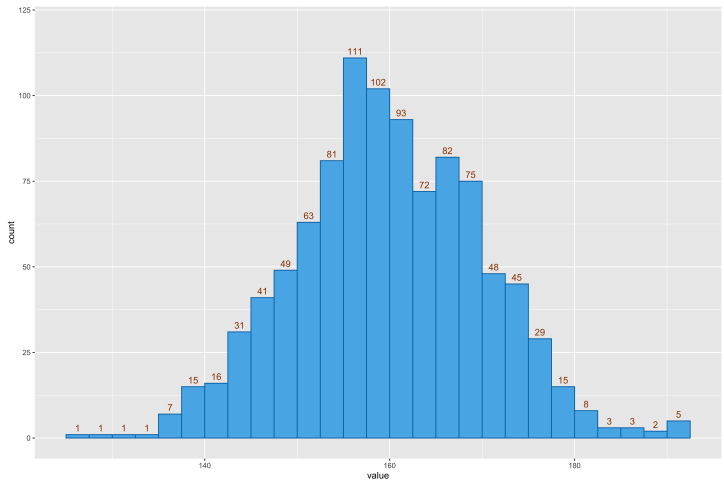


図 7.3　身長の確率分布。bin 幅は 2.5cm

ぐらいの量が詰まっているか，で考える必要があります。確率変数の確率質量に関する密度ですから，これを**確率密度 (probability density)** といい，確率密度＝確率質量/区間幅で計算します。たとえば図 7.2 の 160-165cm の階級は 0.165 の確率質量でしたから，密度は $0.165/5 = 0.033$，というように考えます。図 7.3 の場合は 160-162.5cm の階級に 0.093 ですから，$0.093/2.5 = 0.0372$ というようになります。

しかしこれでは，連続変数を離散変数のようにカテゴリカルに区切っているだけですから，連続変数は連続変数らしくもっと細かく，連続的になるようにこの階級をどんどん短くしていくことを考えましょう。数学的には，間隔をどんどん短くしていった極限を考えることを微分というのでした。この考え方を使って，間隔を極限まで短くして確率密度を表現したものが，連続分布になるわけです。

ここで注意することがいくつかあります。まず，離散変数の時は全事象の確率質量を足し合わせると 1.0 になるのでした。連続変数の場合は微分してその密度を考えてきましたが，全体のことを考えるには微分の逆である積分の計算が必要です。つまり連続変数の場合は，全区間の確率密度を積分すると 1.0 になる，ということになります。次に，確率質量と違って確率密度の値は 1.0 を超えることもあるということです。これは密度が確率質量 (1.0 未満) を階級で除したものであることを思い出すとすぐにわかります。分母がどこまでも小さくなっていくのですから，ある点の密度はどんどん大きくなっていくしかないのです。分母が極限まで小さくなると，密度は無限大になります。それと関係して，連続変数の場合ある一点の確率はゼロであるということになっています。つまり，**確率密度は常に「区間」を伴って考えますので，ある一点の確率，というのは連続変数の場合存在しません (0 と定義します)**。これはそうしないと色々都合が悪いので，このようなルールにしたということです。面倒な言い方になるのですが，たとえば身長が 165cm である確率は，と言われるとこの定義からゼロになるのです。そうではなく，身長が 165cm 以上である確率は，とか，身長が 165cm 以上 170cm 未満の間にくる確率は，という言い方をしなければならないということです。

7.2.3 確率分布関数

さて，離散変数であれ連続変数であれ，確率的な振る舞いのすべてを書き出したものが分布として用いられるということでした。離散変数ならまだしも，連続変数の場合にすべての実現値の値を書くことは不可能ですから，これらは数学関数の形を使って表現されます。こうした関数は，離散変数の場合は**確率質量関数 (probability mass function)**，連続変数の場合は**確率密度関数 (probability density function)** と言います。「質量」と「密度」の違いに注意しておいてください。確率密度関数の場合は，細かい話をすると微分や積分の話が出てくることになりますが，この授業では積分計算をいちいちやったりしませんので安心してください。性質として，全部足し合わせると 1.0 (連続変数の場合は全部の面積が 1.0) であること，確率は非負の数字であること，を把握していれば問題ありません。

また表記法についてですが，ある変数 X がある確率分布に従うことを「$X \sim$ 確率分布 (パラメータ)」の形で表現します。この $\sim$ はチルダという記号ですが，このほにゃっとしたカーブで左側の変数が右側の分布に従うことを意味するわけです。右側にくる確率分布関数には，かっこ付きで**パラメータ (parameter)** を記載します。パラメータとは分布の見た目を決める数字のことで，この数字を調整すると分布の形がかわります。形といっても Form(数式) が変わるのではなく，Shape/Figure(見た目) の形が違うということです。たとえば先ほどの図 7.1 は横一線ですし，図 7.3 は単峰のなだらかな山のような見た目をしていますが，これを関数で表す際にはパラメータに数字をある数字を与えて図を書かせる必要がありますね。分布の形を決めるのは，パラメータの値を決めることと同じであるということを，頭の片隅においといていただけたらと思います。

7.2.4 確率関数の期待値

ここで確率分布関数の特徴を記述することについて触れておきます。確率分布とは，確率的に変動する数字の実現値と，出現確率を一覧にしたもので，確率質量関数や確率密度関数で表現されるのでした。離散変数であれ，連続変数であれ，いろいろな数字が出てくる可能性があるので，その特徴を記述する方法をみて

おきましょう。ここでは確率分布の平均と分散を考えます。確率分布の平均はとくに**期待値 (expectation)** と呼ばれます。

確率的にしか得られない数字に対して，平均的にどれぐらいの数字が得られるか，は「トータルどの程度の数字が期待できるか」という意味で期待値と呼ぶのです。サイコロの例で説明しましょう。サイコロの出目は 1，2，3，4，5，6 ですが，サイコロを振って出る目の平均はどれぐらいの値になるのか？ を考えます。ここではもちろん正当なサイコロ，出目がいずれも 1/6 の確率であることとします。確率変数 X の期待値は $E(X)$ で表し，次のように計算します。

$$E(X) = \sum x_i \times p_i$$

ここで x_i は確率変数がとりうる値，p_i はその値が出てくる確率です。そしてすべての可能性を足し合わせ ($\sum$) ます。サイコロの期待値は次の通りです。

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

実際に 3.5 という出目があるわけではありませんが，平均はこうなるということです。このように，確率分布の平均値は実現値 × その確率をすべて足し合わせたもの，と思っておいてください。たとえば，これがわかれば宝くじの期待値がわかります。2004 年のサマージャンボ宝くじ，1 ユニット，1000 万通り当たりの当選金は表 7.1 のようになります。

表 7.1　2004 年のサマージャンボ宝くじ当選金一覧

等級	当選金	本数	当選確率
1 等	2 億円	1	0.00001%
前後賞	5 千万円	2	0.00002%
組違い	10 万円	99	0.00099%
2 等	1 億円	1	0.00001%
3 等	100 万円	10	0.0001%
4 等	10 万円	100	0.001%
5 等	3000 円	100000	1%
6 等	300 円	1000000	10%
夏ラッキー賞	ラッキー賞	40000	0.4%

これを元に期待値を計算すると次のようになります。

$$200000000 \times 0.0000001 + 50000000 \times 0.0000002 + \cdots + 10000 \times 0.4 = 143.0$$

宝くじは一枚 300 円で買えますから，300 円買っても平均 143 円しか当たらないことになります。言い換えると 157 円=52% は胴元の儲けになるので，ギャンブルとしては効率が悪い部類に入ることがよく知られています[*4][*5]。

話を戻します。このようにして，(離散) 確率分布関数の平均値が計算できます。平均値だけでなく，分散も計算できます。式は次のようになります。

[*4] 外国では，「宝くじは貧乏人の税金だ」などと言われることもあるようです。一攫千金を求めてつい･･･ね。

[*5] ちなみにこれは「バラ」で買った場合です。連番で買うとちょっと計算がややこしいです。

$$Var(X) = \sum (x_i - E(X))^2 \times p_i$$

実際の出目から期待値を引いたものの二乗に，その確率をかけたものを総和するのですね。サイコロの場合は次のようになります。

$$(1-3.5)^2\frac{1}{6} + (2-3.5)^2\frac{1}{6} + \cdots + (5-3.5)^2\frac{1}{6} + (6-3.5)^2\frac{1}{6} = 2.9166...$$

連続分布の場合は足し算でなく積分をすることになりますが，基本的な発想は同じです。

7.3 ベルヌーイ分布と二項分布

7.3.1 ベルヌーイ分布

さてここまでで，確率という考え方を導入しました。離散的イベントが確率的に生じる場合，離散確率分布に従う，といいます。連続的な量が確率的に得られる場合，連続確率分布に従うといいます。確率の大きさをすべて書き表したものを確率分布といい，確率分布は関数で表現されます。離散変数は確率質量関数，連続変量は確率密度関数でそれぞれの事象・量の生じやすさが表されます。

少しややこしくなってきたので，具体的で最も簡単な例をあげながら話を進めることにします。コイントスを考えましょう。コインを弾いて表が出るか，裏が出るかは確率的な事象ですね。**結果変数 (outcome variable)** は，「表」か「裏」かのどちらかになります[*6]。これしかないので，離散的な確率変数です。

このような，結果が 2 種類だけであり，毎回の試行が独立（前に出た結果が続く結果に影響することはない）な試行のことを，ベルヌーイ試行といいます。17 世紀の数学者，ヤコブ・ベルヌーイに因んだ名前です。コイン X の表が出ることを，$X = 1$，裏が出ることを $X = 0$ と書くとします。表が出る確率が θ であれば，裏が出る確率は $1 - \theta$ になります。すべての確率を足し合わせると 1.0 になるし，事象は表か裏しかないからです。これは次のように表現できます。

$$X \sim Bernoulli(\theta)$$

べるぬーい，のスペルが面倒ですが，先ほど説明した表記法が具体例として示されていることを確認してください。ここで $\theta = 0.5$ であればこのコインは表が出るか，裏が出るかが半々だということなので，公平なコインだと言えます。マジシャンが細工をして表が出やすくなったコインというのがあるとすれば，たとえば $X \sim Bernoulli(0.8)$ のように表現できるでしょう。このように，確率分布関数の名前 (ここでは $Bernoulli$) の後ろのカッコ内にパラメータ θ を，そしてそのパラメータの値が幾つになるかで確率分布の性質を記述したことになります。このようにパラメータが 1 つだけの分布関数もありますし，複数あるものもあります。

ベルヌーイ分布関数の中身はどうなのでしょうか。$X = 1$ になる確率が θ で，$X = 0$ になる確率が $1 - \theta$ である，というのを数式で表現するには，次のようにします。

$$P(X = k) = \theta^k(1-\theta)^{1-k} \text{ただし } k = \{0, 1\}$$

ここで $P(X = k)$ は X が k になる確率を表します。k は 0 か 1 です。表が出る確率 $P(X = 1)$ は $\theta^1(1-\theta)^0 = \theta \times 1 = \theta$，となっていますね ($x^0 = 1$ という指数の定義がポイント)。

[*6] コインが地面に突き刺さって立ったらどうするんだ！というツッコミがあるかもしれませんが，「その場合は表とする」とか，ルールを決めておけばいいですよね。ここでは例なので単純な 2 種類のアウトカムだけで話をさせてください。

7.3.2 二項分布

さて，コイントスを 2 回行ったとします。同じコインでピン，ピンと指ではねあげるのです。結果としてあり得るのは，表･表，表･裏，裏･表，裏･裏の 4 通りです。確率はそれぞれ θ^2,$\theta(1-\theta)$,$\theta(1-\theta)$,$(1-\theta)^2$ となります。ここで表･裏と裏･表は同じ $\theta(1-\theta) = \theta(1-\theta)$ ですから，$2\theta(1-\theta)$ ですね。$\theta = 0.5$ だとすると，表が 2 回でる確率が 0.25，1 回出るのが 0.5，0 回出るのが 0.25 ということになります。この 3 つの数字を足しても 1.0 になりますから，n 回中 k 回表が出るという組み合わせもまた確率分布になります。

n 回中 k 回成功する[*7]確率を表す確率分布は，**二項分布 (binomial distribution)** といい，次の式で表します。

$$P(X = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

ここで $\binom{n}{k}$ は組み合わせの記号で，${}_nC_k$ と同じ意味です。こちらの表記法で習った人が多いかもしれません。これは二項係数とも呼ばれるもので，詳しくは高校数学の教科書に戻って確認して欲しいのですが，ここでは要するに複数回試行の成功率を考えるときには二項分布になるという仕組みを理解してください[*8]。

さて，二項分布に従う確率変数は，「何回試行したか」という試行回数 n にも依ることなので，パラメータに n を表記に追加しなければなりません。分布の形を決めるパラメータが成功率 θ と試行回数 n の 2 つになり，次のように表記します。

$$X \sim Binom(\theta, n)$$

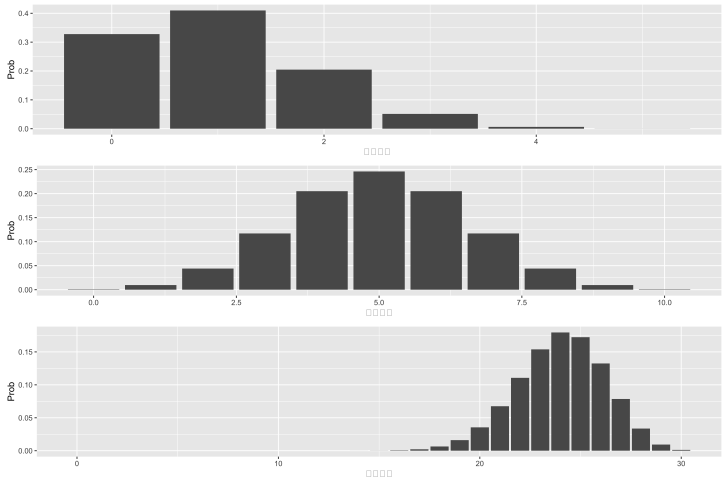


図 7.4 二項分布の例。最上段は $n = 5, \theta = 0.2$, 中段は $n = 10, \theta = 0.5$, 下段は $n = 30, \theta = 0.8$

図 7.4 にはいくつかの二項分布の例を示しました。パラメータの設定はそれぞれ，最上段は $n = 5, \theta =$

[*7] 表が出る，でもいいのですが 0/1 の二択は裏表以外にも，男女，正誤，合否，生死など色々なもののメタファになりますので，ここは表が出ることを成立する，成功する，という言い方にしたいと思います。

[*8] 一応少し補足しておきます。n 個の中から順番を区別して取り出す方法が順列，順番の区別をしない (裏･表＝表･裏) のが組み合わせというのでした。順列は n 個から k 個取り出すとき ${}_nP_k = n \times (n-1) \times (n-2) \times \cdots (n-k+1)$ という数だけあります。この中で重複している組み合わせは ${}_kP_k = k \times (k-1) \times \cdots 1 = k!$ ありますので，組み合わせは $\binom{n}{r} = \frac{{}_nP_k}{k!}$ 個，ということになります。

0.2, 中段は $n = 10, \theta = 0.5$, 下段は $n = 30, \theta = 0.8$ となっています。同じ関数ですが, パラメータ n, θ の違いで見た目が変わっていることに注意してください。

7.4 正規分布の形

今度は連続的な確率変数の代表格である**正規分布 (Normal Distribution)** について説明します。心理学を始め, 統計的アプローチでこの正規分布は中心的な役割を担います。統計モデルのほとんどは正規分布帝国に含まれる, といわれるほどです (三中, 2018)。数式的には小難しいのですが, その意味的なところを理解できるように解説します。

7.4.1 正規分布ができるまで

正規分布の始まりは, ドイツの物理学者カール・フレードリヒ・ガウスの研究からです。彼の貢献に敬意を評して, 正規分布のことを**ガウス分布 (Gaussian Distribution)** とも呼びます。彼が考えていたのは**誤差 (Error)** に関することでした。天文学ではケプラーが何十年にもわたる観測データを集めたりしていましたが, 人間が観測するデータには誤差が含まれています。この誤差をなんとかハンドリングしたいというのがそもそものモチベーションです。

誤差は**系統誤差 (systematic error)** と**偶然誤差 (random error)** に分かれます。系統誤差は機械の精度や測定者の癖などで, 必ず一方向にずれるもの。これはその機械の特徴や測定者の癖がわかれば, 修正することが可能です。一方, 偶然誤差の方は「偶然生じるもの」なので, どのように出てくるのかは全然予測できません。他の何ものとも相関しない (関係が見出せない) というのが特徴です。さらに, 誤差は, 「いろいろな原因が積み重なって生まれる」と考えられます。たとえば天体観測の場合, レンズの歪み, 気温や湿度などの条件, 観測者のちょっとしたミス・・・これらの積み重ねで「正しい値からずれる」と考えるのです。

天体観測の例で考えましょう。ある星の位置を測定していたとします。星がちょうどスコープの中に入るように, 天体望遠鏡の位置を左右に調整していくとします。話を簡単にするために, 右側にずらすことをプラス, 左側にずらすことをマイナスと考えます。さて, ではまず何かの原因で, 測定値が本当の値からちょっとプラスの方向にずれる, ということがあったとしましょう。次に第二の「誤差の原因」があり, それが今度はちょっとマイナスに働いたとします。さらに第 3 の「誤差の原因」があり, それが今度はちょっとプラスに働いたとします。誤差は何段にも何段にも重なっていくとすると, 途中には「同じぐらいの大きさで向きが違う」誤差, というのも含まれてくるでしょう。これは測定上はプラス・マイナスの大きさが相殺し合って, 表に出てこないことになります (図 7.5)。

誤差の大きさはさまざまですが, たとえば大きなプラスの誤差は ($\rightarrow\rightarrow$), 標準的なサイズの誤差 ($\rightarrow$) が二度連続して出た, と考えることもできます。積み重なる原因はたくさんたくさんあるのですから, この際ちょっとぐらい増えても変わりません。また, どうせ足し合わせるんですから, 順番を入れ替えても構いません。そうすると, 相殺のペアはあちこちで見つけることができるようになります。結局プラスに何回, マイナスに何回, という組み合わせがどれだけ出るかという話になります。

さて, 誤差がプラス側に出るのを 1, マイナス側に出るのを 0 と表したとして, その試行を何度も何度も繰り返し, 結果的に「1 が 1 つだけ」「0 が 1 つだけ」「1 が 2 つ」「0 が 2 つ」「1 が 3 つ」・・・「全部 1」「全部 0」というパターンを数え上げていくとすると, これはどういう形になるでしょうか。

そう, **二項分布**ですね！

復習です。n 回試行したうち, k 回表が出る (成功する) 確率を表す確率分布のことを, 二項分布というのでした。今回の例では $n \rightarrow \infty$, つまり何度も何度もコイントスをする, 無限回する, というイメージで考えて

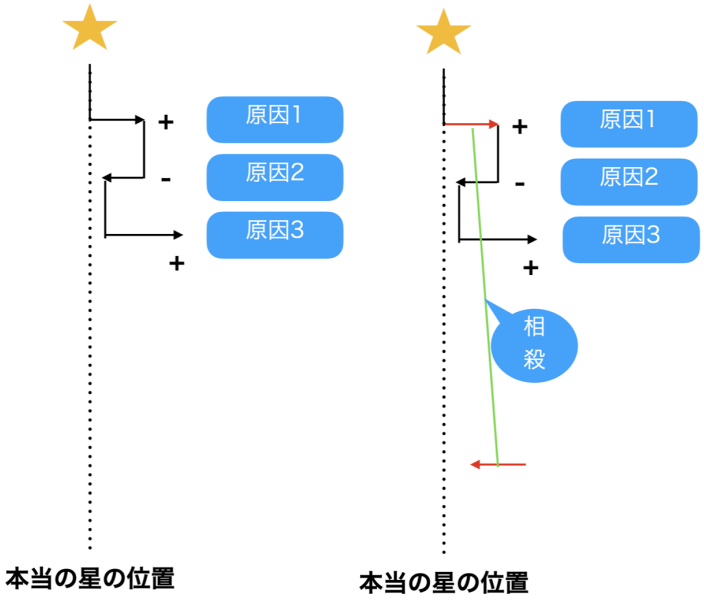


図 7.5 色々な大きさの誤差が重なり合う。時には同じ大きさで相殺し合うことも。

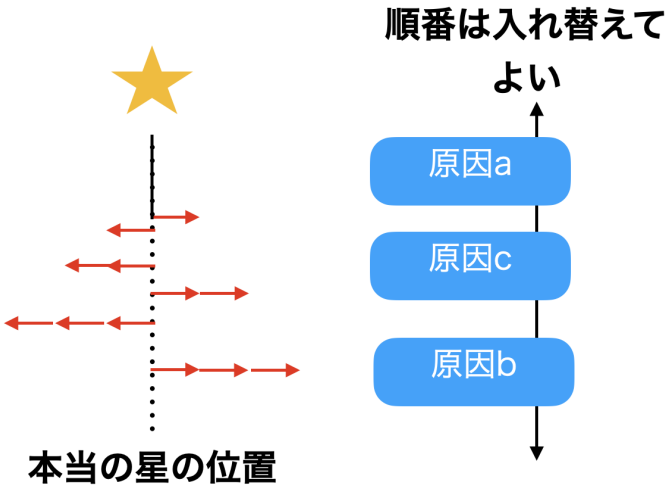


図 7.6 積み重ねて考えるので順番を入れ替えても良い

ください。そのうち，たまたま 1 回表がでた（ちょっとプラスになった）のは何回か，2 回表が出た，3 回出たのはどうか，・・・・全部表だった場合はどうか，という時の組み合わせの形を考えていきます。その行き着く先が，今回の**正規分布**なのです。

違った例から説明してみましょう。みなさんは**ゴルトンボード (Galton Board)** というおもちゃをご存知でしょうか (図 7.7)。釘が等間隔に打ち付けられている板で，その上からパチンコ玉を落とすのです。パチンコ玉は釘にあたります。その時，右側に落ちるか左側に落ちるかの確率は 1/2 ですね。たとえば左側に落ちたとしましょう。落ちた先にもまた釘があり，そこに当たって右にいくか左に行くかは，また 1/2 の確率になります。右側に落ちていたとしても同様で，次の釘で左右のどちらに弾かれるかは半々です。2 段目までで考えると，「左，左」に来たケースと「左，右」あるいは「右，左」に来たケース，「右，右」に来たケースがありますが，それぞれのケースになる確率は $0.25 : 0.5 : 0.25$ ですね。これは二項分布の $\theta = 0.5, n = 2$ のケースですが，

この n がどこまでも連なっている，つまり釘の段がどこまでも続いているのが正規分布の世界であり，それを模したのがゴルトンボードです。

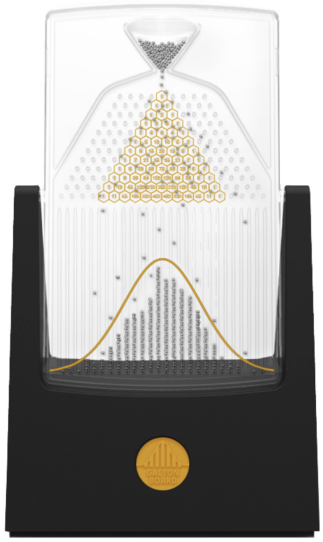


図 7.7　二項分布と正規分布の関係を表すゴルトンボード。図は Wikipedia の「ゴルトンボード」の項より。

このボードの下には左右対称のなだらかな山の形が描かれていますが，これが正規分布の形です。美しいですね。

さて，「誤差は偶然生じる」「誤差は多段階で積み重なる」ということから，二項分布の極限を考えました。その結果得られる形のことを正規分布といいます。正規分布は連続値の確率分布の代表格で，さまざまなところにその形を見出すことができます。この確率分布のポイントは左右対称になっていることと，より極端な値はより出てきにくい，ということです，つまり誤差を生み出すさまざまな要素がすべてプラスに働くこと，すべてマイナスに働くことというのはめっっっっっったにない，ということです。

たとえば心理学で考えたい性質，性格とか社会的態度といった心理学的傾向は，何がその形成要因になっているかわかりません。親の学歴，親の年収，育った地域の教育水準，幼稚園に通ったのか保育園に通ったのか，食事は肉と魚どっちが多かったかとか，それはもう無数の原因があると思います。すべてがその心理学的傾向に影響したとして，それぞれが「高かった」「低かった」「よかった」「幼稚園だった」「魚だった」というように分岐していくと，無数のパターンが考えられます。その結果，人の心理学的傾向はどのように分布するかというと，とある平均的な程度を中心にして，「やや外向性が高い人」「やや不安定性が高い人」といった相対的な比較がされるようになり，かつ「極端に外向性が高い人」というのは滅多にいない，ということになるでしょう。つまりそうした性質は正規分布に従うと考えられるのです。学力なども同様です[*9]。

正規分布はこのように，さまざまな影響源による累積的影響の帰結として捉えることができます。意味的には，誤差のように確率的な散らばりの形を表すものでもあり，心理学的傾向の例のように集団レベルでの性質が従う分布としてとらえられることもあります。この形状をとくに Normal Curve ということもありますが，これの中心から外れることが Abnormal だというわけではありませんので，最近は Normal というより Gaussian という形容詞で呼ぶことが多いのかもしれません[*10]。

[*9] 収入などは偶然性による影響ではなく，人為的な傾向が入ってくるので正規分布になりません。

[*10] 平均的な人間が正常 normal で，平均から逸脱した人間は異常 abnormal である，という考え方は人類がその評価基準を社会的なものに求めるようになったからであり，このことをニーチェは「神は死んだ」と表現したわけです。正義や善悪を神の視点で考えるのではなく，人間同士の相対的な態度で考える態度です。しかし集団的傾向の中心にいることが正しいことを意味するも

7.4.2 正規分布を特徴づける

正規分布の形で大事なのは,「位置」と「幅」です。系統誤差がどれぐらい出てくるのか, に該当するのが位置です。偶然誤差は相殺し合うので平均 0 になります。偶然誤差がどれぐらい出てくるのか (散らばるのか) については, 幅で特徴付けることができます (図 7.8)。

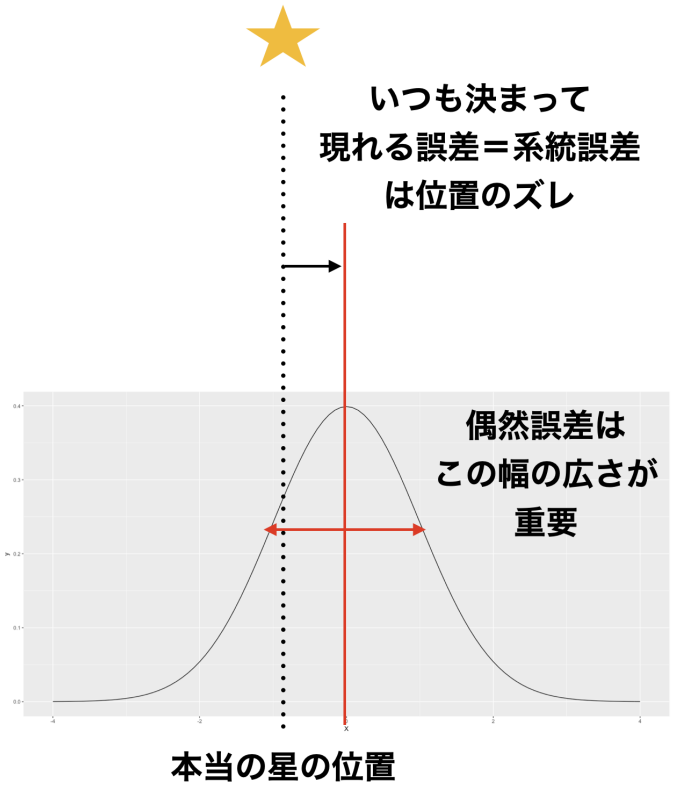


図 7.8 正規分布で重要なのは位置と幅

ところで, 正規分布の幅というのはどこで測れば良いのでしょう？ 分布の形はなだらかな山形ですから, 標準偏差がどこを表しているのかイメージしにくいのではないでしょうか。もしこの誤差の分布が三角形であれば, その幅を定義するには底辺の長さを使えば良いでしょう。しかし正規分布は緩やかなカーブをしていますし, 理論的にその裾野は $-\infty$ から $+\infty$ まで広がっているのです。

これについては, 長沼 (2016) がとてもわかりやすい解説をしてくれています。すなわち, この正規分布のカーブを飛行機が飛んでいると考えれば特徴が見つけやすいというのです (図 7.8)。飛行機は操縦桿を引くと機首が上がり, 上に上がります。操縦桿を押すと機首が下がり, 下に下がります。正規分布カーブのように飛ぶ時は, 二箇所で操縦桿が水平になるシーンが出てくるはず。そこを結ぶとカーブの幅が定義できます。この幅の長さが標準偏差 2 つ分に該当します。平均を挟んで, 左右に標準偏差 1 つ分ずつでてくるわけですね[*11]。

ということで, 正規分布は位置と幅が重要だということがわかりました。記述統計でいえば, 位置は平均で

のではないことは, みなさんもお分かりいただけると思います。

[*11] 数学的には, この点は変曲点と呼ばれます。関数を微分すると変化量がわかります。距離 (と時間) から速度がわかるようなものですね。変化量をさらに微分すると加速度がわかります。この加速度に対応するのが操縦桿の位置です。この加速度がゼロになるところは, 関数を 2 階微分したものがゼロになる点として算出できます。

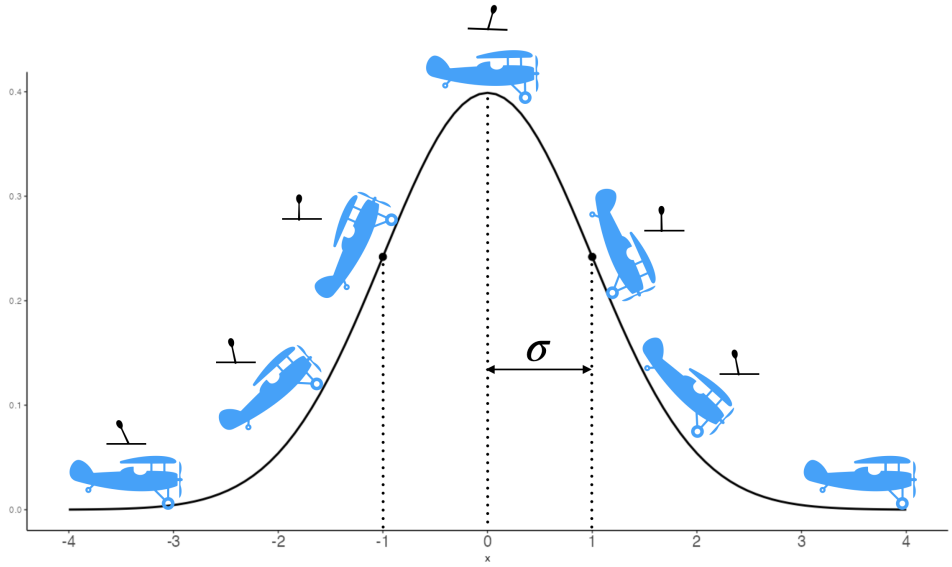


図 7.9 正規分布の幅を測る位置。長沼 (2016) を参考に作図しました。

す。左右対称な分布の中心にあたりますから，これが適切な代表値でしょう。同様に幅は標準偏差で考えることができます。数学的に正規分布を記述する場合，ある変数 X が正規分布に従うことを，$X \sim N(\mu, \sigma)$ と記述します。ここで N は Normal の N です。μ は平均を表しており，ギリシア文字 μ は「ミュー」と読みます。σ は標準偏差を表しており，ギリシア文字 σ は「シグマ」と読みます。総和記号 $\sum$ はギリシア文字 S の大文字 Σ からきていますが，これは小文字の s です。

正規分布の表記法は，テキストによっては $N(\mu, \sigma^2)$ と書かれることも少なくありません。σ^2 は標準偏差 σ の二乗ですから分散を表しています。二乗するかしないかの違いはありますが，幅を表している数字ですし，二乗する/平方根をとる，という操作で相互に変換可能ですから，どちらで表現しても間違いではありません。本講義では σ で表記を統一したいと思います[*12]。

図 7.10 には，さまざまな平均値と標準偏差を持った正規分布の形を描いてみました。ここに描かれている曲線は**すべて正規分布です**。全然違うようにも見えますが，平均と SD の値の取り方によっては，このように見え方としての形 (Shape,Figure) は変わります。が，これを描いている数式としての形 (Form) は同じであることに注意してください。

図の中心にある黒い線が**標準正規分布**と呼ばれる，そのなの通り標準的な形の正規分布です。これは平均が $\mu = 0$，SD が $\sigma = 1$ の分布です。平均値が変わることは，その位置がずれることです。標準正規分布と同じ形をしていて位置が右にずれている赤い線がありますが，これは $\mu = 1, \sigma = 1$ のときの形状です。

標準偏差が変わることは，その幅が変わることになります。標準偏差を小さくすると分布の山は尖っていき，標準偏差を大きくするとなだらかな分布に変わっていきます。図の最も左にピークを持っている分布は，$\mu = -2, \sigma = 0.5$ にしたときの正規分布の形です。位置が左にずれ，山が尖っていることがわかりますね。最もなだらかなラインは $\mu = 1, \sigma = 3$ にして描きました。ほとんど横ばいのような形ではありますが，これでも左右対称で山形の正規分布の一種です。

一口に正規分布と言っても，パラメータの設定によって見え方が大きく変わることがお分かりいただけたの

[*12] 心理統計のテキストでは，一般的に σ^2 で表すことの方が多いようです。数学や統計ソフトウェアでは二乗せずに計算することが多く，たとえば**確率的プログラミング言語 (stocastic programming language)** の Stan がもっている正規分布関数は σ として扱っています。

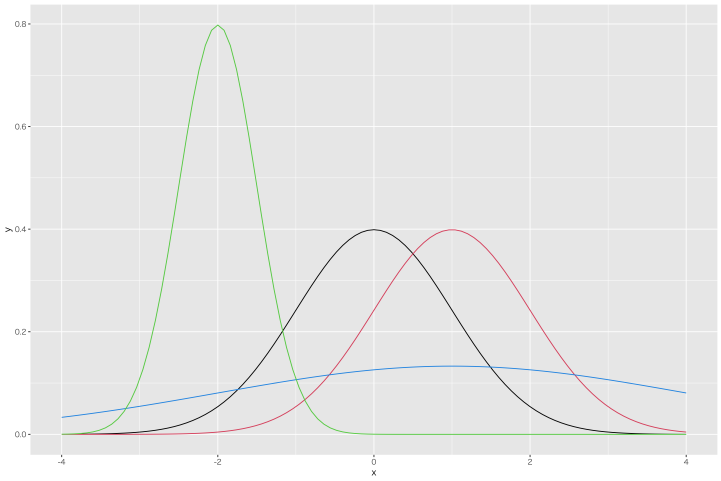


図 7.10 さまざまな幅・位置の正規分布

ではないでしょうか。言い換えると，パラメータさえ調節すればさまざまなデータにも応用できそうです。これが正規分布帝国が幅広く用いられる 1 つの理由であると言えるでしょう。

7.5 正規分布の関数

正規分布の形状は，ガウスが発見した確率密度関数で描かれます。その関数は次のような形をしています。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{7.1}$$

ここで $\mu = 0, \sigma = 1$ という標準化されたスコアであれば，この関数の形は一気に簡単になります。標準化された正規分布のことを**標準正規分布 (standard normal distribution)** といいますが，次の式になります。

$$f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$$

さて，数式が出てきてびっくりされた方もいるかも知れません。しかし実際は，この複雑にみえる数式も計算機を使って値を出したり確率を計算したりしますから，式を丸暗記してどこかで使わないと，というようなシーンはほぼありません[*13]。それよりも正規分布の形や，ソフトウェアはなんでもいいですから計算機で必要な値を算出できるようになることの方が重要です[*14]。

ここでは統計環境 R をつかった確率分布の数字を算出する方法を説明します。R は基本的な確率分布関数を最初から持っており，これを呼び出すだけで使えます。正規分布は `norm`，二項分布は `binom` など，英語名を省略した形が関数になっています。そのほかにも色々な分布があって，たとえば t 分布であれば `t`，χ^2 分布であれば `chisq`，F 分布であれば `F` などとします。しかし，その関数だけでは使えません[*15]。関数

[*13] もちろん腕に自信がある人は，積分計算で答えを計算して出してもらっても構いません。

[*14] 計算機が今ほど普及していなかった頃，心理統計のテキストにはたいてい付録として数表がついているものでした。代表的な関数については，ある程度計算したものの一覧表があると，計算機を使う必要がないから便利だったのですね。数表の例はセクション 7.8.2,Pp.126 を参照してください。

[*15] とくに `F` は R の中では `FALSE` の予約語になっており，それだけで偽/0/スイッチオフ，のような特別な意味を持っています。

の略称の前に，文字を付け加えることで確率密度関数をさまざまな角度から捉えた数字を提示します。そのルールを表 7.2 にまとめました。

表 7.2 R の確率に関するルール

用途	コード名 XXX の関数名	説明
確率密度	dxxx(q)	q は確率点を表す。コード名が norm なら dnorm(q) となる。
累積確率	pxxx(q)	q は確率点を表す。コード名が norm なら pnorm(q) となる。
確率点	qxxx(p)	p は確率を表す。コード名が norm なら qnorm(p) となる。
乱数	rxxx(n)	n は乱数の数を表す。コード名が norm なら rnorm(n) となる。

XXX のあたりが少しわかりにくいですが，ここには確率分布を表す関数名が入ると思ってください。正規分布でしたら norm であり，これに d, p, q, r をつけることで正規分布の確率密度，累積確率，確率点などの値を返してくることになります。

図 7.11 に dXXX，pXXX，qXXX が何を表すかを図示しました。確率密度 d は関数の高さを意味します。累積確率 p は下限から指定した確率点 (x 座標，q で表す) までの面積を意味します。また，下限からの面積が p になる時の x 軸上の座標 q を求めることもできます。

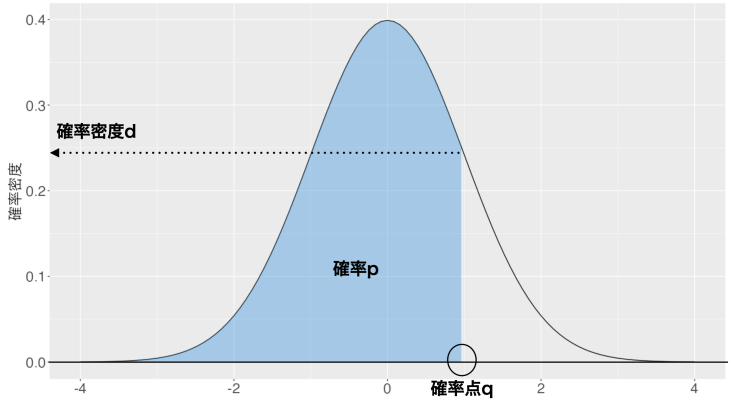


図 7.11 dXXX，pXXX，qXXX の示すところを確認しよう

具体例を見てみましょう。R で次のように入力し，実行した例を示します。

code：7.1 確率に関する関数使用例

```
1       dnorm(x = 0, mean = 0, sd = 1)
2       pnorm(q = 1.0, mean = 0, sd = 1)
3       qnorm(p = 0.618, mean = 0, sd = 1)
4       rnorm(n = 10, mean = 0, sd = 1)
```

R の出力 7.1: 確率に関する関数の出力例

```
> dnorm(x = 0, mean = 0, sd = 1)
[1] 0.3989423
> pnorm(q = 1.0, mean = 0, sd = 1)
[1] 0.8413447
> qnorm(p = 0.841, mean = 0, sd = 1)
[1] 0.9985763
> rnorm(n = 10, mean = 0, sd = 1)
[1] -1.00458071 -1.16956265  0.59980728 -0.07192335  0.66660905 -0.10629118 -
0.94452719 -0.04108743  2.39762300 -0.10580504
```

1 つずつ確認しましょう。`dnorm(x = 0, mean = 0, sd = 1)` は $\mu = 0, \sigma = 1$ の正規分布における，確率点 0 の確率密度，すなわち正規分布曲線の高さを示しています。

`pnorm(q = 1.0, mean = 0, sd = 1)` は $\mu = 0, \sigma = 1$ の正規分布において，確率点 1.0 までの面積を求めています。0.841 ですから 84.1% ですね。これは数式で言うところの $\int_{-\infty}^{1.0} f(x)dx$ を計算した結果，数表でいうところの $P(0 \leq Z \leq 1.0) = 0.341$ に対応しています[*16]。

`qnorm(p = 0.618, mean = 0, sd = 1)` は $\mu = 0, \sigma = 1$ の正規分布において，$-\infty$ からの面積が $p = 0.841$ になるときの確率点を答える関数です。先ほどの関数とは逆の操作ですね。ちょうど 1.0 になるべきなのですが，0.998563 なのでほぼ 1.0 ということでヨシとしてください。

こうした関数を使うことで，たとえば 0.3 から 1.0 までの面積は，`pnorm(q = 1.0, mean = 0, sd = 1) - pnorm(q = 0.3, mean = 0, sd = 1)` とすることで算出できます。実際にはこうした計算をすることは稀ですが，確率分布の関数をどのように使うかは想像できるようにしておいてください。

また最後の `rnorm` については説明がまだでした。関数 `rnorm(n = 10, mean = 0, sd = 1)` を実行すると，10 個の数字が出てきています。10 個というのは関数の中で $n = 10$ と設定したからです。幾つに設定しても構いませんが，さて出力される数字はなんでしょうか。

これは乱数といって，なんの規則性もない数字の連なりのことです。1 つ 1 つの数字の現れ方には規則性がないのですが，長い目で見ると正規分布に従っているといった，確率変数の実現値を擬似的に作り出すのがこの関数なのです。このようにコンピュータには擬似乱数発生装置が用意されていて，関数の形がわかっている代表的な確率分布であれば，そこからどういう数字が出てくるのかシミュレーションしてみることができるようになります。たとえば先ほど図 7.1 で，サイコロを 100 回投げた例を説明しましたが，アレは実際に 100 回投げたわけではありません。すみません。この乱数発生機能を使って，擬似的に一様分布する乱数を作ったのでした。皆さんも RPG などをしていて敵に出くわす，敵が攻撃してくる，というシーンを経験したことがあると思いますが，あれも機械の中で乱数を振っていて出くわしたかどうか，攻撃するか逃げるか，攻撃するとしたらどの程度 HP が減るかということを擬似的に計算しています。ソーシャルゲームなどでは「ガチャを回す」ということがありますが，あれもプログラムによってどのような数字が出るか計算済みです。偶然を，確率分布の力と計算機の能力で制御しているのです。乱数の使い方は，統計や心理学の研究上でも役に立ちます。具体的には第 28 講までお待ちいただくことになりますが，乱数によって確率的な振る舞いを近似し，データがどのような特徴を持つかを考えることができるようになります。この話は長くなりますのでひとまず横に置いておいて，先を急ぎましょう。

[*16] 数表は確率点 0 からの差分で表現しますので，$-\infty$ から 0 までのちょうど半分の確率を足して，$0.5 + 0.341 = 0.841$ になります。詳しくはセクション 7.8.2，Pp.126 参照。

7.6 標準正規分布の利用

最後に (標準) 正規分布をどのように利用するかについて, 少しだけ触れておきます。

正規分布は誤差だけでなく, 集合的な性質を見るときにもよく現れるという話をしました。たとえば, 大学入試共通試験の点数の分布はきれいな正規分布になっていたりします。心理学で扱う心理変数にも正規分布が仮定されます。というのも, 人の行動を決めるのには, 生まれてから今まで無数の選択を繰り返し, その積み重ねで出てきていると考えられるからで, そうしたデータをたくさん集めると (ゴルトンボードのように) 正規分布になると考えられるからです。

ここで次のように考えてみましょう。手元に何人かぶんのデータがあったとします。理論的にいって, このデータは正規分布に従うと無理なく仮定できるとします。手元のデータは数十, 数百ぐらいしかなかったとしても, さっきの乱数の話のように, 正規乱数に従う数十, 数百の実現値があると考えると, この標本分布も正規分布を近似していると言えるかもしれません。であれば, このデータをつかってある種の推測ができるようになります。

皆さん, 模擬試験で「あなたの偏差値は XXXX です」とか「もしこの試験を 10000 人が受けていたとしたら, あなたは XXX 番目です」というフィードバックを受け取ったこと, ありませんか？繰り返しになりますが, 学力テストのようないろいろな要因が絡むもの (ex. 本人の努力, 使ってるテキスト, 学校の雰囲気, テスト会場の気温と湿度, 家庭環境, 子供の頃の経験‥‥) は, 正規分布に従うと考えられ, 全体が正規分布するのなら, その相対的な位置は正規分布の式から推定できるので, このような推測が成り立つのです！

この授業のテーマは「見えないものを見るための心理統計」でした。その一端がここから伺えます。本当に 1 万人もテストを受けているとは言えない模擬試験でも, 分布の仮定から「実際には見えない位置」まで推測できちゃうのです。

正規分布は平均と標準偏差で特徴付けられ, それがどういう値になっているかはわかりませんが, このデータを**標準化**すればそのことを気にする必要もなくなります[*17]。標準正規分布はその特徴がよくわかっています。数表や計算機を使うと特定の範囲の面積も明らかになります。たとえば正規分布は平均 ± 1 標準偏差の中に全体の 68.2% が, 平均 ± 2 標準偏差の中に全体の 95.4% が, 平均 ± 3 標準偏差の中に全体の 99.7% が含まれるのです。であれば, 標準化したスコアが 1 であれば全体の上位 15.8% に, 標準化したスコアが 2 であれば全体の上位 2.2% にいると言えます[*18]。この確率を 1 万人換算すれば,「偏差値 60 のあなたは上位 15.8%」とか「偏差値 70 のあなたより上には 227 人しかいません！」といったことが言えるようになります。

この話を数式で表現しておきましょう。

1. データ x が正規分布に従っている, というのを $x \sim N(\mu, \sigma)$ と書く。
2. データを標準化する。$z_i = \dfrac{x_i - \mu}{\sigma}$
3. $z_i \sim N(0, 1)$ より, 標準正規分布の特徴を使って推測する。

ここで第二のステップ, 標準化そのものは, データが正規分布に従っていると言えなくても行える計算作業であることに注意してください。集合的なデータが正規分布しているという仮定と組み合わさることではじめて, 推測につかえるようになります。推測に用いる統計学, **推測統計学 (inferential statistics)** は後期に入るとじっくり取り掛かることになります。ここでは正規分布の特徴の活用例として示すにとどめます。

[*17] 標準化したスコアについてはセクション 5.3, Pp.89 を参照。

[*18] R を使って計算をしました。関数を利用して `1-pnorm(q=1,mean=0,sd=1)=0.15865`, あるいは `1-pnorm(q=2,mean=0,sd=1)=0.02275013` とします。

7.7 課題

■確率と呼ばれる数字 次の確率についての記述において，正しいものすべてに丸をしてください。

1. 連続変数の場合，確率密度が負の値をとることもある。
2. サイコロを振った時，1 か 2 が出る確率は 1/36 である。
3. 離散変数の場合，すべての事象を分母にとった相対頻度で確率を表現できる。
4. 降水確率のような「確率」という言葉の使い方は，主観的な評価を含んでいるのであり正しい確率の定義ではない。
5. 「わからない」ことを確率で表現する 1 つの方法は，すべての事象の生起確率を 0 と見積もることである。
6. サイコロの目やコイントスの結果などは，相互に排他的な事象である。
7. 確率質量は 1 を超える場合もある。
8. 確率密度は 1 を超える場合もある。
9. 連続変数の確率分布を関数で表したものは，確率密度関数である。
10. 離散変数の確率分布を関数で表したものは，確率質量関数である。

■正規分布の面積を求める 標準正規分布における，1. 確率点 $p = -1$ の確率密度，2. 確率点 $q = 0$ までの累積確率，および 3. 確率点 -1.96〜$+1.96$ の範囲の面積を求めてみましょう。

■正規分布を推測に利用する あるクラスの試験の結果は，平均点が 75，標準偏差が 15 点の正規分布に従っていると考えられる。このとき，40 点以上 50 点未満の人は何 % になるか求めてみましょう。ここで「50 点未満」は 50 点未満であり，49 点以下ではないことに注意してください。

7.8 補遺

7.8.1 標準正規分布の関数形

すでに述べたように，正規分布の関数形，式 7.1 を必ずしも覚えておく必要はありません。しかし，どうしてこのような形になっているかを理解すると，この式も少しはわかりやすくなりますので，少し解説してみましょう。

まず今から求めたいのは，誤差の確率分布ですから，そこからいくつか満たすべき条件というのが出てきます。確率という数字は非負です。また，ここでは連続変数ですから，確率密度関数の面積を積分すると 1.0 になる必要があります。つぎに誤差の分布ですから，中心から外れるほどその出現確率が少なくなるようにならなければなりません。しかも傾向がないので左右対称です。これらの条件を満たすような関数を作ることを考えます。

左右対称の分布からいきましょう。左右対称の関数として，真っ先に思い浮かぶのは $y = x^2$ のような二次式ではないでしょうか (図 7.12 の 1 段目左)。ただ $y = x^2$ だと原点から離れるほど大きくなってしまいます。誤差は逆で，一箇所をピークにどんどん下がっていって欲しいので，$y = -x^2$ にした方が良さそうです (図 7.12 の 1 段目右)。さてそうすると，今度は値 y がすべてマイナスになってしまいます。確率関数を作ろうとしているのにこれはまずい。確率はすべて正の数字でなければならないからです。ここで**指数関数 (exponential function)** を導入します。指数関数とは $y = e^x$ という形をしています。これは $y = exp(x)$ とも書き，自然対数 e を何乗するかという関数です[*19]。関数の形をグラフにすると，図 7.12 の 2 段目左のようになります。指数の定理から，指数 x がマイナスであるとは逆数にするという意味ですから，マイナスになることはありません[*20]。この関数を先ほどの $y = -x^2$ と組み合わせて，$y = exp(-x^2)$ とします (図 7.12 の 2 段目右)。この段階で，すでに正規分布っぽい形にはなっているのです。すなわち，左右対称で両端に行くほど小さくなる関数ですね。足りないのは，パラメータ μ, σ によって位置と幅を調節する機能をつけたいこと，面積が 1 になるように整えたいこと，この 2 点です。これを追加します。

そしてパラメータを追加します。位置をコントロールできるようにするには，x に数字を足したり引いたりしてやればいいことになります。平均を μ として，$y = exp(-(x-\mu)^2)$ とすることでこの問題は解決します (図 7.12 の 3 段目左)。次に幅をコントロールしたいですね。これは x の倍率を調節してやれば良く，x を分散 σ^2 で割ってやることで対応できます。すなわち，$y = exp(-\frac{x^2}{\sigma^2})$ です (図 7.12 の 3 段目右)。位置と幅の両方をコントロールするには，$y = exp(-\frac{(x-\mu)^2}{\sigma^2})$ とすれば良いでしょう。

あとはこの関数が，積分して面積 1 になるように整えてやる必要があります。このようにしてやると，関数が確率を表す数字として扱うことができるようになるからです。この関数の面積を計算して，その面積で割ってやればいいことになりますね。ただこの複雑な関数，面積を求めるのが大変難しい。ただの積分ではうまくいきそうにないのです。その数学的な問題を突破したのがガウスで，凄すぎるのでガウス積分と名前がついているほどですが，これについては説明すると長くなるので解説は他所に譲り[*21]，答えだけ言っておくとその面積は $\sqrt{2\pi\sigma^2}$ になります[*22]。ということで，その面積で割ってやるように関数を規格化しましょう。そうすると

例の式になります。

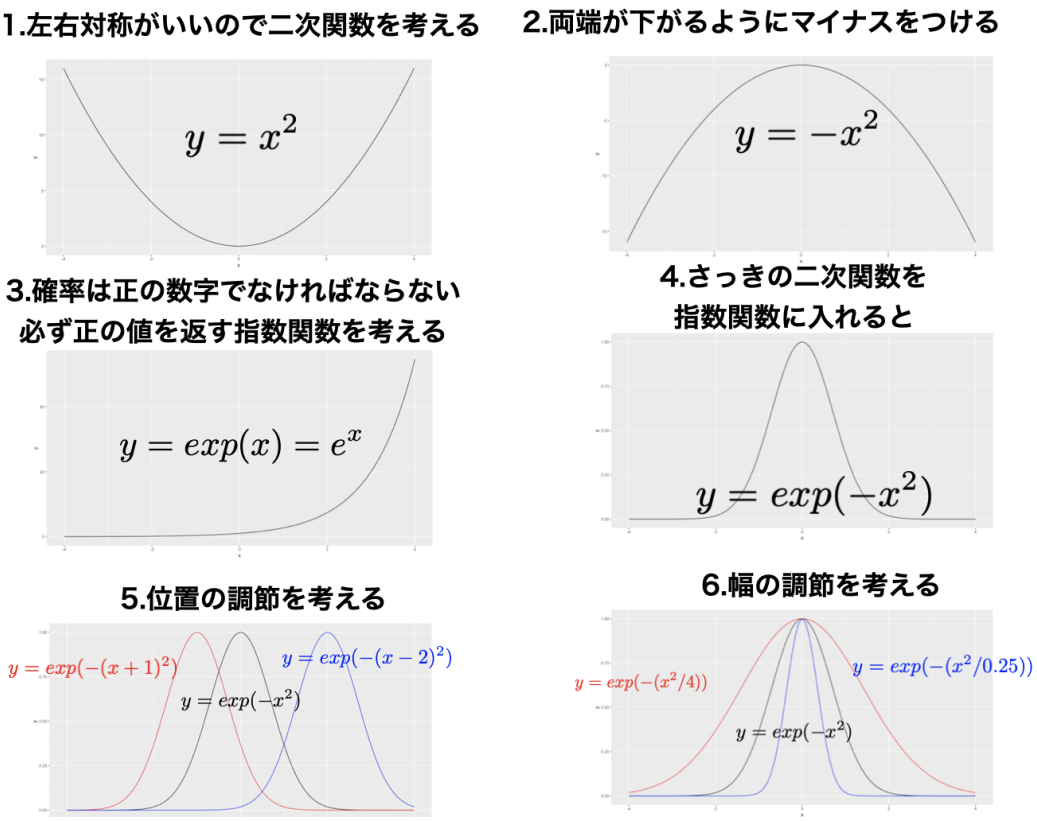


図 7.12 正規分布の確率密度関数導出の経緯

7.8.2 数表とその見方

表 7.3 が正規分布に関する値の一覧表です。ここには，z と lower , upper とした列があります。lower は $P(0 \leq Z \leq z)$ の範囲を，upper は $P(Z \geq z)$ の範囲を表しています (図 7.13)。

ここで z 列は今知りたい確率点，今回でいえば $z = 0.3$ とか $z = 1.0$ のことになります。次の lower= $P(0 \leq Z \leq z)$ 列には，$z = 0$ からその求めたい確率点までの面積が入っています。たとえば $z = 0.3$ のところを見ると，$P(0 \leq Z \leq 0.3) = 0.118$ となっていますから，この確率が 11.8% であることがわかります。今回は 0.3 から 1.0 までの面積が知りたかった訳ですが，それは直接この表には載っていません。ただ，$z = 1.0$ の数字を見ることができますから，それを参照して $P(0 \leq Z \leq 1.0) = 0.341$ がわかります。0 から 0.3 までが 0.118 で，0 から 1.0 までが 0.341 ですから，0.3 から 1.0 までは $0.341 - 0.118 = 0.223$，22.3% だと読み取ることができます。

この数表には $P(Z \geq z)$ の確率も載っています。たとえば upper= $P(Z \geq 0.3) = 0.382$ ですから，0.3

*19 e は何かというと 2.718... という実数です。微分しても変わらない数字ということで，数学の世界では重宝されます。e については Maor (1994 伊理訳 1999) が読み物として楽しく理解できます。

*20 指数は $a \neq 0$ のとき，$a^0 = 1$，また，$a^{-n} = \dfrac{1}{a^n}$ と定義されています。

*21 こういうの見ると，「あ，ずるい。著者め逃げたな」と思いませんか。私なら思います。そう思われると悔しいので，参考図書として西内 (2017) をあげておきます。また長沼 (2016) の P.62–69 も参考になります。

*22 π は 3.141592... という実数です。円周率ですね。面積になぜ円周率が出てくるのか，それがガウスの凄いところです。

以上になる確率は 38.2% ということがわかります。こうした数表を駆使することで，ある範囲からある範囲までの面積がいくらか，という計算をできます。

またたとえば，確率点がマイナスの場合はどうするか，というのを疑問に思われる方もいるかもしれません。その時は，正規分布が左右対称であることを思い出してください。$z = -0.3$ 以上の面積が知りたい，という時は，まず $z = 0.3$ のときの数字から $P(Z \geq 0.3) = 0.382$ を見つけます。左右がひっくり返っても同じですから，$P(Z \leq -0.3) = 0.382$ になります。-0.3 までの面積がわかっても，知りたいのは -0.3 以上です。しかし確率はすべての面積が 1.0 ですので，$P(Z \geq -0.3) = 1 - 0.382 = 0.618$ と計算できます。あるいは，$P(0 \leq Z \leq 0.3) = 0.118$ で，平均 $z = 0$ までの面積が半分，すなわち 0.5 ですから，$0.5 + 0.118 = 0.618$ と求めても構いません。ともかく，半分しかない数表であっても，これを使えばさまざまな面積計算ができるということです。

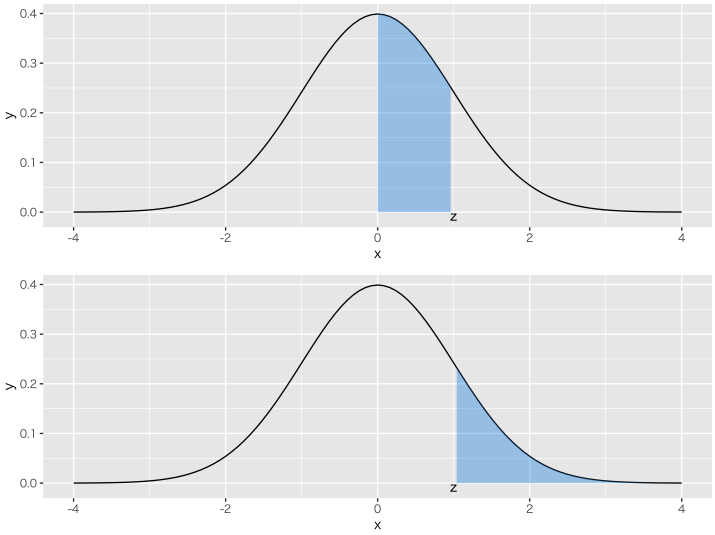


図 7.13 数表の見方。上の段が $P(0 \leq Z \leq z)$，下の段は $P(Z \geq z)$ のイメージ図

表 7.3 標準正規分布の表

z	lower	upper	z	lower	upper	z	lower	upper	z	lower	upper	z	lower	upper
0.00	0.000	0.500	0.60	0.226	0.274	1.20	0.385	0.115	1.80	0.464	0.036	2.40	0.492	0.008
0.01	0.004	0.496	0.61	0.229	0.271	1.21	0.387	0.113	1.81	0.465	0.035	2.41	0.492	0.008
0.02	0.008	0.492	0.62	0.232	0.268	1.22	0.389	0.111	1.82	0.466	0.034	2.42	0.492	0.008
0.03	0.012	0.488	0.63	0.236	0.264	1.23	0.391	0.109	1.83	0.466	0.034	2.43	0.492	0.008
0.04	0.016	0.484	0.64	0.239	0.261	1.24	0.393	0.107	1.84	0.467	0.033	2.44	0.493	0.007
0.05	0.020	0.480	0.65	0.242	0.258	1.25	0.394	0.106	1.85	0.468	0.032	2.45	0.493	0.007
0.06	0.024	0.476	0.66	0.245	0.255	1.26	0.396	0.104	1.86	0.469	0.031	2.46	0.493	0.007
0.07	0.028	0.472	0.67	0.249	0.251	1.27	0.398	0.102	1.87	0.469	0.031	2.47	0.493	0.007
0.08	0.032	0.468	0.68	0.252	0.248	1.28	0.400	0.100	1.88	0.470	0.030	2.48	0.493	0.007
0.09	0.036	0.464	0.69	0.255	0.245	1.29	0.401	0.099	1.89	0.471	0.029	2.49	0.494	0.006
0.10	0.040	0.460	0.70	0.258	0.242	1.30	0.403	0.097	1.90	0.471	0.029	2.50	0.494	0.006
0.11	0.044	0.456	0.71	0.261	0.239	1.31	0.405	0.095	1.91	0.472	0.028	2.51	0.494	0.006
0.12	0.048	0.452	0.72	0.264	0.236	1.32	0.407	0.093	1.92	0.473	0.027	2.52	0.494	0.006
0.13	0.052	0.448	0.73	0.267	0.233	1.33	0.408	0.092	1.93	0.473	0.027	2.53	0.494	0.006
0.14	0.056	0.444	0.74	0.270	0.230	1.34	0.410	0.090	1.94	0.474	0.026	2.54	0.494	0.006
0.15	0.060	0.440	0.75	0.273	0.227	1.35	0.411	0.089	1.95	0.474	0.026	2.55	0.495	0.005
0.16	0.064	0.436	0.76	0.276	0.224	1.36	0.413	0.087	1.96	0.475	0.025	2.56	0.495	0.005
0.17	0.067	0.433	0.77	0.279	0.221	1.37	0.415	0.085	1.97	0.476	0.024	2.57	0.495	0.005
0.18	0.071	0.429	0.78	0.282	0.218	1.38	0.416	0.084	1.98	0.476	0.024	2.58	0.495	0.005
0.19	0.075	0.425	0.79	0.285	0.215	1.39	0.418	0.082	1.99	0.477	0.023	2.59	0.495	0.005
0.20	0.079	0.421	0.80	0.288	0.212	1.40	0.419	0.081	2.00	0.477	0.023	2.60	0.495	0.005
0.21	0.083	0.417	0.81	0.291	0.209	1.41	0.421	0.079	2.01	0.478	0.022	2.61	0.495	0.005
0.22	0.087	0.413	0.82	0.294	0.206	1.42	0.422	0.078	2.02	0.478	0.022	2.62	0.496	0.004
0.23	0.091	0.409	0.83	0.297	0.203	1.43	0.424	0.076	2.03	0.479	0.021	2.63	0.496	0.004
0.24	0.095	0.405	0.84	0.300	0.200	1.44	0.425	0.075	2.04	0.479	0.021	2.64	0.496	0.004
0.25	0.099	0.401	0.85	0.302	0.198	1.45	0.426	0.074	2.05	0.480	0.020	2.65	0.496	0.004
0.26	0.103	0.397	0.86	0.305	0.195	1.46	0.428	0.072	2.06	0.480	0.020	2.66	0.496	0.004
0.27	0.106	0.394	0.87	0.308	0.192	1.47	0.429	0.071	2.07	0.481	0.019	2.67	0.496	0.004
0.28	0.110	0.390	0.88	0.311	0.189	1.48	0.431	0.069	2.08	0.481	0.019	2.68	0.496	0.004
0.29	0.114	0.386	0.89	0.313	0.187	1.49	0.432	0.068	2.09	0.482	0.018	2.69	0.496	0.004
0.30	0.118	0.382	0.90	0.316	0.184	1.50	0.433	0.067	2.10	0.482	0.018	2.70	0.497	0.003
0.31	0.122	0.378	0.91	0.319	0.181	1.51	0.434	0.066	2.11	0.483	0.017	2.71	0.497	0.003
0.32	0.126	0.374	0.92	0.321	0.179	1.52	0.436	0.064	2.12	0.483	0.017	2.72	0.497	0.003
0.33	0.129	0.371	0.93	0.324	0.176	1.53	0.437	0.063	2.13	0.483	0.017	2.73	0.497	0.003
0.34	0.133	0.367	0.94	0.326	0.174	1.54	0.438	0.062	2.14	0.484	0.016	2.74	0.497	0.003
0.35	0.137	0.363	0.95	0.329	0.171	1.55	0.439	0.061	2.15	0.484	0.016	2.75	0.497	0.003
0.36	0.141	0.359	0.96	0.331	0.169	1.56	0.441	0.059	2.16	0.485	0.015	2.76	0.497	0.003
0.37	0.144	0.356	0.97	0.334	0.166	1.57	0.442	0.058	2.17	0.485	0.015	2.77	0.497	0.003
0.38	0.148	0.352	0.98	0.336	0.164	1.58	0.443	0.057	2.18	0.485	0.015	2.78	0.497	0.003
0.39	0.152	0.348	0.99	0.339	0.161	1.59	0.444	0.056	2.19	0.486	0.014	2.79	0.497	0.003
0.40	0.155	0.345	1.00	0.341	0.159	1.60	0.445	0.055	2.20	0.486	0.014	2.80	0.497	0.003
0.41	0.159	0.341	1.01	0.344	0.156	1.61	0.446	0.054	2.21	0.486	0.014	2.81	0.498	0.002
0.42	0.163	0.337	1.02	0.346	0.154	1.62	0.447	0.053	2.22	0.487	0.013	2.82	0.498	0.002
0.43	0.166	0.334	1.03	0.348	0.152	1.63	0.448	0.052	2.23	0.487	0.013	2.83	0.498	0.002
0.44	0.170	0.330	1.04	0.351	0.149	1.64	0.449	0.051	2.24	0.487	0.013	2.84	0.498	0.002
0.45	0.174	0.326	1.05	0.353	0.147	1.65	0.451	0.049	2.25	0.488	0.012	2.85	0.498	0.002
0.46	0.177	0.323	1.06	0.355	0.145	1.66	0.452	0.048	2.26	0.488	0.012	2.86	0.498	0.002
0.47	0.181	0.319	1.07	0.358	0.142	1.67	0.453	0.047	2.27	0.488	0.012	2.87	0.498	0.002
0.48	0.184	0.316	1.08	0.360	0.140	1.68	0.454	0.046	2.28	0.489	0.011	2.88	0.498	0.002
0.49	0.188	0.312	1.09	0.362	0.138	1.69	0.454	0.046	2.29	0.489	0.011	2.89	0.498	0.002
0.50	0.191	0.309	1.10	0.364	0.136	1.70	0.455	0.045	2.30	0.489	0.011	2.90	0.498	0.002
0.51	0.195	0.305	1.11	0.367	0.133	1.71	0.456	0.044	2.31	0.490	0.010	2.91	0.498	0.002
0.52	0.198	0.302	1.12	0.369	0.131	1.72	0.457	0.043	2.32	0.490	0.010	2.92	0.498	0.002
0.53	0.202	0.298	1.13	0.371	0.129	1.73	0.458	0.042	2.33	0.490	0.010	2.93	0.498	0.002
0.54	0.205	0.295	1.14	0.373	0.127	1.74	0.459	0.041	2.34	0.490	0.010	2.94	0.498	0.002
0.55	0.209	0.291	1.15	0.375	0.125	1.75	0.460	0.040	2.35	0.491	0.009	2.95	0.498	0.002
0.56	0.212	0.288	1.16	0.377	0.123	1.76	0.461	0.039	2.36	0.491	0.009	2.96	0.498	0.002
0.57	0.216	0.284	1.17	0.379	0.121	1.77	0.462	0.038	2.37	0.491	0.009	2.97	0.499	0.001
0.58	0.219	0.281	1.18	0.381	0.119	1.78	0.462	0.038	2.38	0.491	0.009	2.98	0.499	0.001
0.59	0.222	0.278	1.19	0.383	0.117	1.79	0.463	0.037	2.39	0.492	0.008	2.99	0.499	0.001

第 8 章

相関と因果

さて今回は，複数の変数の関係について考えていきたいと思います。ここまでの記述統計量 (平均値，中央値，最頻値，分散，標準偏差，**四分位**，パーセンタイル，最小値，最大値，範囲) はいずれも 1 つの変数に対して計算される量でした。今度は身長と体重とか，国語の点数と算数の点数といった，2 つの変数がどういう関係にあるかを示す指標を考えていきたいとおもいます。

8.1　変数間関係の指標

指標の話をする前に，可視化を考えましょう。複数の変数を図にするとき，連続的な変数であれば**散布図**を描くのが直感的にもわかりやすいでしょう。

図 8.1　散布図の例

図 8.1 に，身長と体重のデータをプロットしたものを示しました[*1]。データの一部は表 8.1 に示した通りです[*2]。X 軸のラベルに height(身長)，Y 軸のラベルに weight(体重) とありますように，2 つの変数を同時にプロットしています。左下に目立つように 1 点色とサイズを変えた点を打ちましたが，これは表 8.1 の ID=1 のケース，つまり曽根選手を表しています。彼は身長が 175cm，体重が 68kg であり，対応する X 軸，Y 軸をみるとちょうどその値になっていますね。このように，各点が 1 つのケースに対応しているのが散布図です。

これを見ると次のことがいえそうです。すなわち，身長が高い人は，体重も大きく，背が低い人は体重も小

[*1] このデータは第 4 章で使ったものと同じ，2020 年のプロ野球選手のデータセットです。

[*2] データは全体で 924 人分あります。

表 8.1 データの一部

ID	Name	height	weight
1	曽根 海成	175	68
2	上本 崇司	170	71
3	鈴木 誠也	181	96
4	田中 広輔	171	82
5	小窪 哲也	175	83
6	長野 久義	180	85

さいが多い傾向がある，と。というのも，身長が高くて体重が軽い領域 (グラフの右下) や，身長が低くて体重が重い領域 (グラフの左上) に点が打たれていないので，そういう人はいない，と思えるからです。この図のように，散布図に打たれた点が右上がりの方向にあると，2 つの変数の間になんらかの関係があるということが推察できますね。その関係がどのようなものであれ，このような関係にあると私たちは「何があるんだろう」と考えたり，「背が高いということは体重も重いだろうな」と予測したりできます。来年チームにやってくる新しい選手が，どんなポジションでどんな特技を持っているかはわからなくても，身長が 200cm だということだけ分かれば，まあ体重が 60kg ってことはないよな，と予測できますものね。

さて，前期後期を通じてこの授業のポイントは「見えないものを見ようとする」でした。皆さんはここで変数間の相関を推察したわけです。関係がありそうだと予測したと言ってもいいかもしれません。直接は見えない関係をデータから見いだすには，変数間関係が重要であることがお分かりいただけますでしょうか。

図 8.1 に示した関係は，とくに「正の相関関係」と言います。方向が同じ，つまり一方が大きければ他方も大きい，一方が小さければ他方も小さい，という関係にあるからです。逆に負の相関関係というのもあります。一方が大きければ他方は小さく，その逆も成り立つような状況がそうで，散布図でいうと点が右下がりになるような散らばりをするときです。負の相関関係でも関係は関係ですから，予測できますね。

どちらの傾向もないデータは相関がない，無相関である，といいます。散布図で書くと，各点が円のように広がる形になります。無相関であれば，関係性・規則性がないので，予測はできません。性格心理学では血液型と性格は無相関である，ということがわかっていますが，この話をすると「でも僕は血液型は A 型で，多くの人に几帳面な性格だ，って言われるんですよ！」と反論する人がいます。実際に授業しているとよくある質問です。何がおかしいのかわかりますか？ 心理学者の主張は「無相関である」，です。関係がない，一般的な傾向がないということです。無関係なので，A 型で几帳面な人もいるし，几帳面でない人もいます。B，O，AB 型で几帳面な人もいますし，そうでない人もいるというだけなのです。無相関であれば，そこから考察することはできません。

このように，変数間関係は色々な面において重要な指標です。分散がなければ情報が取り出せない，というのと同じで，相関関係がなければ関連する情報が取り出せないのです。このことを念頭に置いた上で，変数間関係を指標化することを考えましょう。

■共分散 二変数間関係の強さを表す指標は**共分散 (covariance)** といいます。さきほどの図 8.1 を例にとって考えてみましょう。身長変数を x，体重変数を y とします。ID=1 の曽根選手のデータは，$x_1 = 175, y_1 = 68$ のように表します。データ全体を使って身長の平均，体重の平均を計算できますが，これをそれぞれ $\bar{x}, \bar{y}$ と表しましょう。共分散は次の式で表されます。

$$s_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

これはよく見ると，分散の式に似ていると思いませんか。分散の式は次のように書くのでした。

$$s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

この二乗になっているところを展開して見ると次のようになります。

$$s_{xx} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})$$

こうして s_{xy} と s_{xx} を見比べるとわかりやすいと思いますが，分散は変数 x と x 自身との共分散であるとも言えるわけです。

もう少し違う見方をしてみましょう。共分散はそれぞれの変数についての平均偏差の積の平均である，とも言えます。$(x_i - \bar{x})$ は変数 x について，i さんの値が平均からどれほどずれているかを表していて，平均より上であればプラス，下であればマイナスになります。$(y_i - \bar{y})$ は同様に変数 y についての話ですが同じ個人 i さんについてのデータでもあります。なので，ある個人 i が，変数 x でも平均より上，変数 y でも平均より上であれば，この 2 つを掛け合わせた $(x_i - \bar{x})(y_i - \bar{y})$ はプラス $\times$ プラスで，プラスの数字になります。同様に，ある人がどちらの変数でも平均より下であれば，マイナス $\times$ マイナスでプラスですから，この数字は大きくなります。この数字が小さくなるのは，ある人が一方の変数で平均より上，他方の変数で平均より下，というときです。このとき掛け合わせた数字はマイナスになるので，総和を減らすことになります。このような数字を人数ぶん足し合わせる $(\sum_{i=1}^{N})$ という操作ですから，最終的にプラスであれば正の相関関係マイナスであれば負の相関関係がある，ということになるわけです。

注意すべきことは，共分散の大きさは変数の単位に依存するということです。たとえば身長と体重の共分散の場合ですと，身長の単位は cm，3 桁の数字で，体重は単位が kg，2 桁の数字が多いでしょうから，かけ合わせたもの平均は 2〜3 桁の数字になるでしょう。この時の単位は cm×kg でよくわかりません。ただ，身長を mm に換算し，体重をインチに換算するといったことをすると，それに応じて共分散の大きさは変わってきます。関係が同じなのに，単位を変えただけで数字が変わってしまっては困りますね。

■相関係数　そこで**相関係数 (correlation coefficients)** の出番です。相関係数[*3]は，単位に影響されずに変数間関係をみるものです。共分散の大きさは，元データの単位に依存するので，変数間関係を表す指標の大きさが単位に依存するということは，異なる変数間関係を比較できないことになって不便だからです。「身長と体重の関係の強さ」と「国語の成績と算数の成績の関係の強さ」はどっちが強い結びつきをしているのでしょう？ との問いには共分散を見ても答えられません。さてでは，それぞれの変数の単位を揃えるにはどうしたらよいのでしょう？

そう，標準化ですね。それぞれのデータを z スコアに計算しなおせばよいのです。変数 x を z スコアにした

[*3] ここではピアソンの積率相関係数についてのみ説明しています。相関係数は他にも順序相関や，ポリコリック相関係数，ポリシリアル相関係数，テトラコリック相関係数など，色々なものがあります。

ものを z_x と表すとして，これの共分散を考えましょう。変数 x と y の相関係数は r_{xy} と表します。

$$
\begin{aligned}
r_{xy} &= \frac{1}{N}\sum_{i=1}^{N}(z_{x_i}-\bar{z_x})(z_{y_i}-\bar{z_y}) && \text{z スコアの平均は 0}\\
&= \frac{1}{N}\sum_{i=1}^{N} z_{x_i} z_{y_i} && \text{z スコアの式に戻そう}\\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y} && i \text{ は分子にしか関係ないから}\\
&= \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y} && \text{分子が共分散だ}\\
&= \frac{s_{xy}}{s_x s_y}
\end{aligned}
$$

このように展開できました。相関係数は共分散をそれぞれの標準偏差で割ったもの，と考えれば良いわけですね。この数字は単位が違う 2 つの変数でも計算できますし，単位が整えられているので $-1 \leq r_{xy} \leq 1$ の範囲に収まります。

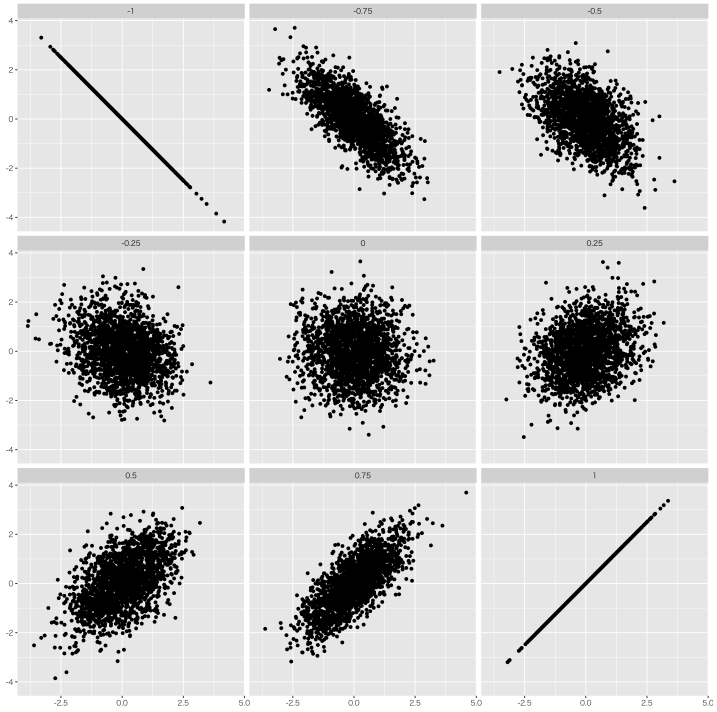


図 8.2　さまざまな相関係数

図 8.2 にはさまざまな相関係数のプロットを示しました。ちなみにさきほどの野球選手の身長と体重のデータは，相関係数 $r = 0.601$ になります。大体の散布図のイメージを持っておいてもらうことも重要ですし，何かデータを取ったら図にする，ということを忘れないでください。

8.2　相関係数の落とし穴

さてここで，相関係数を見る時の注意すべき点をいくつかあげていきます。

相関係数を見て考える話に入る前に，注意喚起から。表 8.2 のようなデータセットがあったとしましょう[*4]

表 8.2 さまざまなデータセット

dataset	Mean.X	Mean.Y	SD.x	SD.y	Min.x	Min.y	Max.x	Max.y	Corr
circle	54.267	47.838	16.760	26.930	21.864	16.327	85.665	85.578	-0.068
dino	54.263	47.832	16.765	26.935	22.308	2.949	98.205	99.487	-0.064
star	54.267	47.840	16.769	26.930	27.025	14.366	86.436	92.215	-0.063

データセット，`circle` も `dino` も `star` も，2 つの変数 X,Y を持っているのですが，いずれも平均，標準偏差にほとんど違いがありません。相関係数も-0.06 となっており，ほぼ無相関，関係がなさそうです。

ところがどっこい，これらの変数をプロットしてみたのが図 8.3 です。ご覧の通り，無関係ではなくプロットすると図が見えてきましたね。

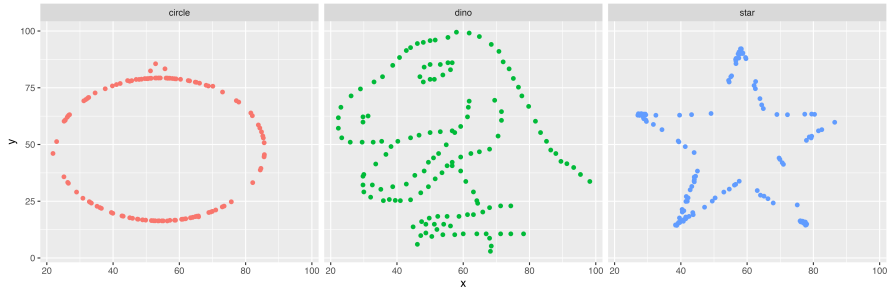


図 8.3 プロットしてびっくり

実際のデータではこんなことはないと思いますが，これは「数字だけ見ていてもわからないよ」という警告です。たくさんのデータを手に入れると，ついつい計算機で一気に数字を出してみたくなります。実際のデータ分析をするときは (あまりにも簡単に相関係数が計算できてしまうがために)，見落とされるところでもあります。数値化は楽ですが，その分なにかを見落としているかも知れない，ということについて常に注意するようにしましょう。

以下では相関関係があったとしても，注意深く考えないといけない例を挙げておきます。

■落とし穴 1；直線じゃないかも 図 8.2 の左上には $r = -1$ の例，右下には $r = +1$ の例があります。このように，相関係数はその極限である ± 1 のときに，直線になる関係であることを理解しておいてください。言い換えると，相関係数はあくまでも変数間の直線的関係を指標化したものであって，曲線的な関係の場合はうまく表現できないのです。図 8.4 には曲がった散布図の例を示しました。このデータの相関係数は $r = 0.005$ です。ほとんど関係ないですね！じゃあこの X 軸の変数と Y 軸の変数は何の関係もない，といっていいでしょうか。たとえばこのデータ，X 軸がコーヒーの温度，Y 軸が美味しさだったらどうでしょう。熱すぎても冷たすぎても，コーヒーは美味しくありません。適温というのがあるのです。あるいは X 軸が血圧，Y 軸が健康度だったらどうでしょう。血圧も高ければ高いほど健康，低ければ低いほど不健康というわけではないですよね。これらの例にあるように，心理的なデータや整理的なデータというのは，中庸であることが重要であることがあります。その時に相関係数を算出して，それがとても低いので「血圧と健康には関係がない」と結論づけるのは間違えています。直線性を表す指標がよくない，というだけなのですから。

[*4] このデータセットは，R の `datasauRus` パッケージ (Matejka & Fitzmaurice, 2017) の一部です。

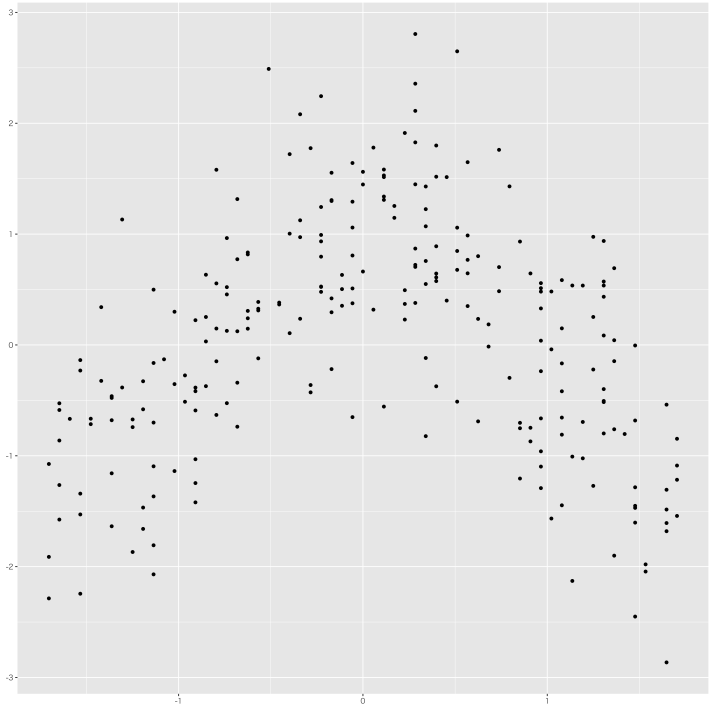


図 8.4 曲がった関係には向かない

■落とし穴 2；全体の傾向じゃないかも 逆に，ある 2 変数の相関係数が $r = 0.922$ でした。この 2 変数には強い相関があるといって良いでしょうか。待って待って，いったん落ち着いて散布図を書いてみましょう。たとえば図 8.5 のようなデータなのかもしれないのですから。図 8.5 は，確かに相関係数は高いのですが，明らかにおかしなポジションにいるデータが 1 件ありますよね。平均値が外れ値に引っ張られたように，平均値をその計算新規の中に取り込んでいる相関係数も，外れ値に弱いところがあります。今回この極端な値を取る 1 ケースを取り除くと，相関係数は $r = 0.3$ にまで下がります。左の小さな群だけで見れば，むしろ関係ないと考えるべきなのです。これもデータを図にしなければ気づかない落とし穴です。

■落とし穴 3；データの一部が切り取られている? とある大学の入学試験のデータを取ってみたとします。入試の成績と，入学後の成績のデータを集め，入試の成績がどれぐらい入学後の成績と関係があるのかを考えたい，というわけです。良い成績で入学したら，入学後も成績が高いというのが普通に考えられる結果ですが，実はその相関係数は $r = 0.4$ ぐらいだった，ということもあります。なぁんだ，入試の成績はその後の頑張りと関係ないんじゃないか，入試なんか意味がないんじゃないか，と思うのは大間違い。

さて図 8.6 には，X 軸に入試の成績を，Y 軸には入学後の成績をプロットしてみました。相関係数は全体で 0.7 あるのです。ところが，大学入試というのは世知辛いもので，ある点数を取らない人はそもそもその大学に入れませんよね。図でいうと，ある線 (60 点) より右側の黒いデータしか分析する時に手に入らないのです。この黒い点だけみて相関係数が低いから，関係ないんだと思ってしまうのは間違いです。このように，データがそもそも打ち切られている可能性にも注意が必要です。このようなデータは切断データといわれます。これは相関係数の問題というより，データの問題なのですが，データの性質をよく考えて考察しなければならないということはしっかり意識しておいていただきたいところです。

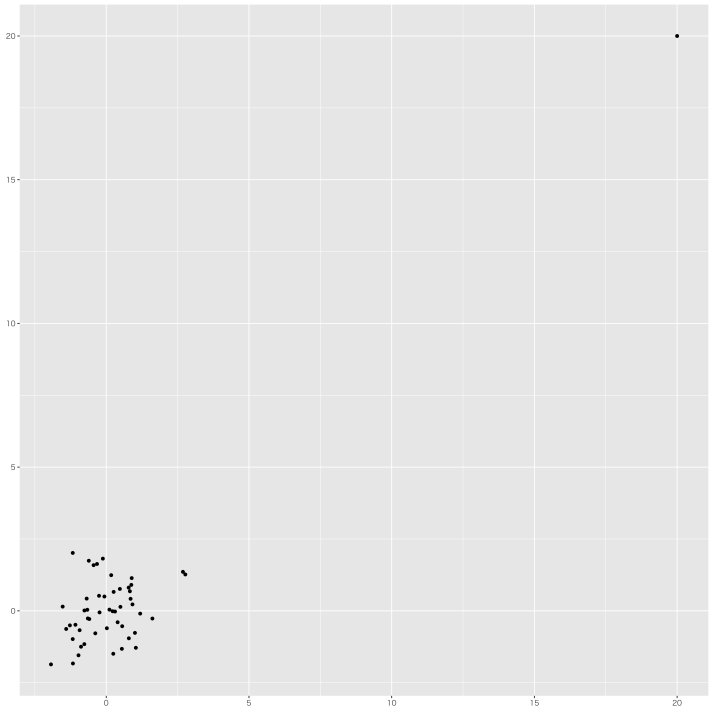


図 8.5　外れ値がひっぱる

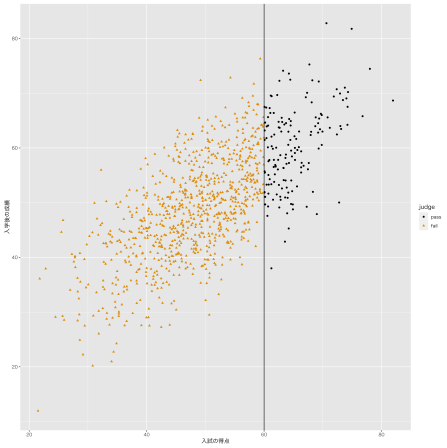


図 8.6　打ち切りデータの例

■落とし穴 4；データは色々な情報が混在している？　図 8.7 の左を見てください。これは右上がりの散布図ですから，なるほど変数間に関係はありそうです。実際に相関係数を計算すると $r = 0.67$ になります。今度は散布図もみているからバッチリ！と思うかもしれませんが，さにあらず。実はこのデータ，3 つのグループのデータが混ざった全体的な傾向です。図の右側には，グループごとに色分けをしてみました。おや，これを見るとグループの中ではむしろ右下がりの分布をしていませんか？ 実はその通りで，グループ A は $r_A = -0.29$，グループ B は $r_B = -0.42$，グループ C は $r_C = -0.35$ と，いずれも負の相関をしているのです。にもかかわらず，全体で見ると正の相関をしている，ということもあるわけですね。

　このデータ，たとえば X 軸がゲームで遊ぶ時間，Y 軸が学校の成績だったとしますと，全体で見るとゲームをすれば成績が上がる，という結論になってしまいますね。でも学校別 (グループ別) でみるとむしろ逆，こ

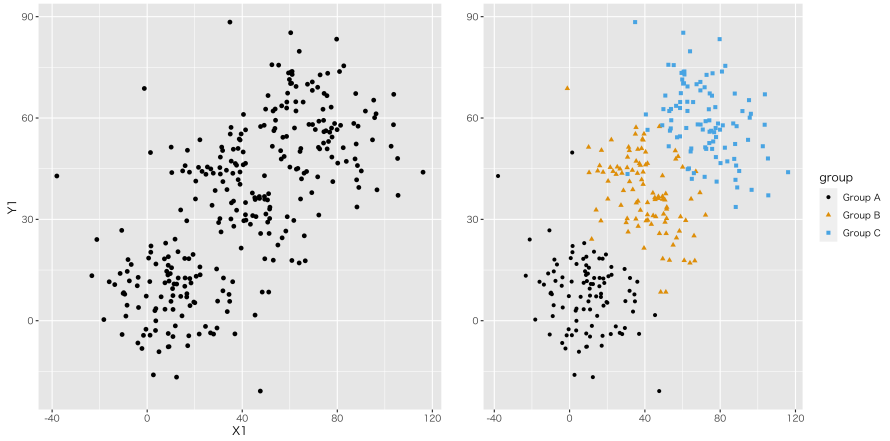


図 8.7 階層的なデータ

こにあるのは 3 つの学校のそもそもの成績の違いが反映されているだけでした，ということもあるわけです。このように，外的な指標で群が変わることがわかっているなら，プロットする時にその情報を加えて，このような情報の混在が生じていないかをチェックする必要があります。プロットするのは当然ですが，プロットのやり方にも注意が必要なのですね。また，群ごとの違いがわかっていなくても，潜在的に何か違うグループに分かれているのかもしれません。回答者の特性を考慮して，いくつかのクラスターに分けて分析するという手法も統計的に考えられています。データを色々な角度から見る，という姿勢は忘れてはいけません[*5]。

■疑似相関 さて，いろいろ「気をつけないといけない」相関係数の話をしてきましたが，最後にたまたま相関があるだけ，という可能性も指摘しておきます。

たとえば次の変数間には高い相関があることがわかっています[*6]。

- ニコラス・ケイジ映画出演数とプール溺死数
- マーガリン消費率とメイン州の離婚率
- 自分のベッドシーツに絡まって死んだ人の数とチーズ消費量

このように，相関があるからと言って必ず意味があるわけではありません。おそらくニコラス・ケイジ (ハリウッド俳優) が頑張って映画に出たからといって，米国でのプール溺死者数が増えるわけではないし，マーガリンを塗らなければ離婚率が下がるというものでもないでしょう。ただデータを取ると，見かけ上，相関係数が高くなってしまっただけなのです。

こうした関係のことを**疑似相関**と言います。我々は目に見えない心理学的な現象を，相関関係を手掛かりに分析していくのですが，たまたまみられる表面的な関係性を「ほらやっぱり！」と思い込んでしまわないように，慎重に考えなければなりません。

8.3 因果関係

色々問題点を指摘しましたが，変数間に相関関係があると，我々は色々なことを推察できます。そういう意味で，相関関係は研究のスタートにもなる重要な指標です。

[*5] マーケティングの領域では，セグメンテーションと呼ばれることもあります。購買層ですね。クラスターというのは COVID-19 の関係で一般的な用語になりましたが，この言葉そのものに悪い意味はなく，塊という意味です。

[*6] 参考；http://tylervigen.com/spurious-correlations

先ほども例に挙げましたが，たとえば身長と体重といった生理的な指標は相関することが一般的ですから，「新しいクラスメイトは身長が 185cm だって」という情報を得たら，まさか体重が 50kg を切っているだろうなんてことは想像しませんね (実際あり得るかも知れないに)。このように，変数間関係があるということは，一方が決まれば他方が決まる，いわゆる関数関係をイメージしがちなわけです。

ただし，**相関関係 (Correlational Relationships)** があるということと，**因果関係 (Causal Relationships)** があるということは別です。たとえば，デスクの上が散らかっている人ほど出世しない，というような話がありますが[*7]，これを聞いて「デスクをきれいにすれば出世する」と言えるでしょうか?実証的なデータに基づいた仮説だったとしましょう。元になるデータは，デスクがきれいな人ときれいではない人をたくさん集めてきて，出世の程度と比べるわけです。出世の程度というのは分かりにくいので，年収でもいいですし，会長 – 社長 – 専務 – 部長 – 課長のように順序尺度水準で並べた数字でもいいとしましょう。そしてこの両者に相関関係が確かにあったとします。つまり，年収が高い人，役職が高い人ほどデスクのきれい度が高く，逆もまた真なり，というわけです。ではデスクの上をきれいにすると出世すると言えるでしょうか？答えは No です。

A と B，2 つの変数があって，相関関係にあるのは事実としましょう。すなわち 2 つの変数には何らかの関係があるとします。しかし因果関係は一方が原因，他方が結果という意味です。$A \to B$ なのか $B \to A$ なのか，相関関係だけでは分かりませんね。デスクがきれいから出世したのか，出世したから (秘書がつくなどして) デスクがきれいになったのか，わからないからです。このように，相関関係と因果関係では前者の方が縛りが緩く (関係があると言ってるだけ)，後者の方はそれに加えてさらに色々な条件を検証しなければなりません。

マスコミなどの報道には，浅慮なのか故意なのかわかりませんが，この手のミスリーディング事例には事欠きません。「お人好しほど収入が少ない[*8]」とか「テレビの暴力シーンを見ると子供が凶暴化する[*9]」といった，相関を因果にすり替えて議論してしまっている例が見られます。他にもたとえば，マイナスイオンを謳った商品などがありますが，あれも「滝など自然の環境にはマイナスイオンが多い。そこで人は緊張がほぐれ，リラックスしている」ことから「マイナスイオンを撒き散らせばリラックスする」という因果の向きをねじ込んで根拠としています。たとえば，町の消防車の数と火災発生件数に相関があったとして，消防車が増えると火災が増える，消防署は火をつけて回っているのだと考えるのはおかしな話だ，というのは誰でもわかると思うのですが。

相関関係以上に深い関係である，因果関係をいうためには，他にも条件が必要です。それらをまとめると次のようになります。原因変数を A, 結果変数を B とすると，

- A が先，B が後に生じていなければならない (時間的先行性)。
- 原因とされる変数が他の多くの変数の原因になっていない。すなわち，$A \to B, A \to C, A \to D \cdots$ のようになっていない (ex. 朝ごはんを食べたから犯罪を犯した，とは言えない)。
- A,B の両方に共通する別の原因 Z，すなわち $Z \to A, Z \to B$ が存在しない (ex. 親が暴力映像が好きで，わんぱくな子供に育てたかった，という共通要因の存在)。

といったことも考えなければなりません[*10]。図 8.8 に「暴力映像を見ると子供が暴力的になる」という因果関

[*7] 金田博之 (2020) デスクが汚い人は出世できない！ 3 つの視点,https://ameblo.jp/dolphin55/entry-12589304622.html.2022 年 2 月 18 日アクセス

[*8] ライブドアニュース (2011) 研究結果：お人好しほど収入は少なくなる https://news.livedoor.com/article/detail/6103564/,.2022 年 2 月 18 日アクセス.

[*9] 朝日新聞 (1999) 暴力 TV，子供に影響/なぐる･けると相関/小中学生を総務庁調査,1999.10.31.

[*10] 事象から因果関係を主張するにはどうすればよいか，というのは哲学的な問題で，J.S.Mill や D.Hume などの思索が有名ですので，色々考えてみてください。また統計的に因果を推論する方法について，近年はさまざまな技術が開発されており，統計的

係を主張するときに，考えるべき影響力の例を挙げてみました。背後の変数 (教育変数)，媒介変数 (子供の性格)，逆向きの因果などの関係がすべて否定されて，始めて暴力映像を見たことが暴力行為の原因である，と言えるのです。

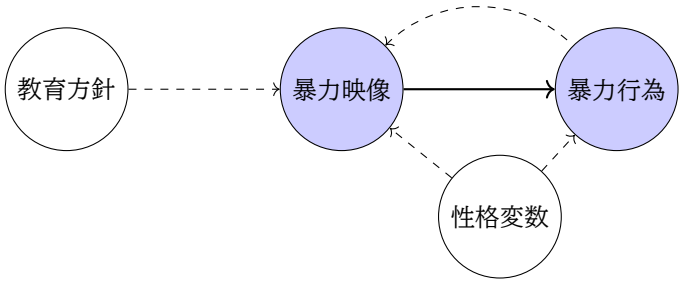


図 8.8 因果を特定するのは難しい。点線のパスがないことを確認しなければならない。

心理学で扱う多くのことは，相関関係にあって因果関係まで特定できることが少ないでしょう。ですから因果関係を考えるときには，他の原因変数がないか，因果の向きが逆ではないかということは常に意識しておかなければなりません。ある結果 R にたどり着くルートが $A \to R, B \to R, C \to R \ldots$ とあるのに，A だけが原因だと言ってしまうのも問題ですし，もっと強く影響を与えている変数を見逃していることにもなりかねません。また「暴力映像が好きだから暴力的になった」のか，「暴力的だから暴力映画も好きだ」となったのかはわからない，あるいはどちらともあり得る話なのです。また，北野武が記者に『暴力映画の影響で実際に暴力行為に出る人間も居ると思いますが、その点についてどう思われますか？』と聞かれた時に『じゃあ、何でおなみだ頂戴の良い話だらけなのに、一向に世の中は良くならねえんだよ』と聞き返した，という逸話があります。このように，ほかの可能性，逆の可能性を考えることが本質に迫る重要なポイントになるのです。研究を深めるためにも，**批判的思考 (Critical Thinking)** の癖をつけておきましょう[*11]。

8.4 交絡という罠

すでに述べたように，心理学は目に見えない心 (というもの) を研究対象にしているため，直接それを操作することができず，変数の相関関係からその存在を推察するしかありません。そして相関関係はここで論じてきたように，因果関係を意味するものではなく，単に関係がありそうだというヒントにすぎないわけです。

それでも私たちは，そのヒントをかき集めてなんとかして研究対象を探ろうとします。事象の原因を心に求め過ぎてしまうことこそ，心の働きの 1 つだということも明らかにされていますが[*12]，我々はつい心を原因とした結果としての現象と考えてしまう，つまり因果関係があると思い込んでしまいがちです。

因果関係を明らかにするのが難しいのは，その背後にあるかもしれない変数が常に見つけやすいわけではないからです。図 8.8 で示したように，原因変数と結果変数の両方に関係している変数がどこかにあるかもしれないわけで，こうした原因・結果変数の両方に関係する要因のことをとくに**交絡要因 (confounding factor)** と言います。交絡という言葉は，字義通り「交わり，からまっている」という意味で，本当に見たい因果関係を切り出せないところが難しい問題です。

にも因果関係はホットなトピックスです。

[*11] 批判というと否定したり，貶めたり，論破したりすることだと勘違いしている人が少なくありませんが，疑ってよく考えようという意味です。そうすると人を疑うなんて！と嫌がる人もいるかもしれませんが，疑うという行為は「よく知ろう」という行為であって，無条件に信じるというのは「考えたくない，知りたくない」ということでもあるのです。漫画「ライアー・ゲーム」にもこのようなセリフが出てきますよ。

[*12] 社会心理学の中ではこれを基本的帰属錯誤 Fundamental Attribution Error と呼んでいます。

たとえば次のような例を考えてみましょう。本屋さんの販促として文庫本フェアを始めたとしましょう。かわいいポップを書いて,「文庫本を読もう！」というコーナーを作ったとします。するとその月の文庫本の売り上げが上がったとします。そうすると誰しも,フェアの効果だ,ポップ広告が効いたね,などと考えるでしょう。販売促進の努力をした → 売上が上がった,というのは誰の目にも明らかな因果関係に思えます。しかし,本当にそうでしょうか。本当にそれだけの効果があったのでしょうか。たとえば同時期にテレビで文庫本の良さを宣伝していたかもしれません。単にある季節になれば文庫本の売り上げが増えるという傾向があるのかもしれません。「販売促進の努力をしたから」という原因以外に,交絡している要因はいろいろ考えられそうです。

本来,因果効果を考えるためには「販促をしなかった時」と「販促をした時」の 2 つの場合で,他の要因をまったく同じにして比較しなければなりません。同じ日,同じ場所,同じ時間帯に販促をした時としなかった時の両方の条件を作らなければ,正しく因果推論できないのです。「販促をしなかったはずの世界線[*13]」というのがあって確認できれば良いのですが,そんなことは土台無理なことです (因果推論の根本問題 (安井, 2019) と呼ばれています)。

もし因果関係を探りたいのであれば,実験的に他の要因を統制する必要があります。みなさんが小中学校で学んできたように,他の条件をすべて同じにした対照実験で,はじめてその効果がわかるのです。もちろん,人間や社会を対象にした研究の場合,個人差もふくめてすべて同じ条件にするというのは無理ですが,さまざまな工夫でできるだけこの無理難題を解決しようとするのです。詳しくは第 12 講で説明しますが,一番良いのは**ランダム化比較実験 (Randomized Controled Ttials; RCT)** という方法です。また,次回以降説明する**回帰分析 (Regression Analysis)** という方法でも,統計的に変数を統制する一助にできます。より専門的には他書 (安井 (2019) や高橋 (2022)) に譲りますが,技巧を駆使して因果推論の根本問題を少しでもマシな形で解決しようと考えられています。

ここでは「一時点の調査では相関関係しかわからないので,そもそも因果関係はわからない」ことや,「完全に統制できない因果関係の根本問題がある」こと,それでも事前にあらゆる交絡要因を探し,できるだけの手を打とうと科学者は努力していること,をしっかり理解しておいていただければ十分です。

8.5 課題

■相関係数の算出 表 8.3 は表 5.4 を再掲したものですが,一列 $z_x z_y$ を追加しています。追加された列および,元データ X, Y の共分散,相関係数を計算してください。また共分散や相関係数が表中のどこに出てくるか確認してください。。なお,——部は計算しなくて結構です。

[*13] 世界線というのは,時間が過去から未来へ一方向的に進む次元とし,時間を線で表した時に,その線上に成立している世界という意味で使っています。専門用語ではなく一般的によく用いられている表現ですが,タイムマシンもの SF で出てきた概念のようです。出典としては, Back to the Future という映画の第二作目でドクが説明する時に線を引いたものや,アニメ STAINS;GATE などが挙げられます。一方,量子物理学でも複数の世界が並行して存在する可能性としての多元世界論というのがあります。える・ぷさい・こんぐるぅ。

表 8.3 二変数 X と Y

ID	X	Y	X の平均偏差	Y の平均偏差	z_x	z_y	$z_x z_y$
1	149	37	-2.2	-12.4			
2	137	52					
3	164	67					
4	159	43					
5	147	48					
平均							
分散			———	———			———-
標準偏差			———	———			———-

第 9 章

線形モデルの基礎

ここまで記述統計による記述と可視化，テスト理論による「見えないもの」の測定，相関関係・因果関係の推論について学んできました。

では，ここからの話はどうなるかというと，いよいよ本題に入っていきます。データに誤差がつくのは分かったとして，知りたいのは真の値，誤差を除いた全体の平均値です。科学の目的の 1 つは予測と制御です[*1]。個々のデータには誤差 (個人差) が含まれますが，全体的にはなんらかの傾向があるだろうと考え，そのなんらかの傾向を**モデル**で表現します。このモデルを数式で書くことで予測ができ，予測ができるのであれば対応を考えることができる，という流れです。

本講義のテーマである「見えないものを見るための心理統計」は，ここからいよいよ，**データを通じて全体的な傾向を見つける**という意味で展開していくことになります。

9.1 相関関係とモデル

モデルとか傾向を予測するというと，なんだか複雑な話だと思われるかもしれませんが，全然そんなことはありません。

たとえばみなさんよく気になる「成績」を例に考えてみましょう。たとえば表 9.1 のように，高校での内申点と大学入試の成績がデータとして手に入ったとします。

高校のデータは内申点なので 5 点満点にしてみました。大学のデータは 1000 点満点にしてみました。仮想データですのでなんでもよかったのですが，ここでは単位が違う例を出したかったのでこのようにしました。

さて，データは可視化するのが基本です。可視化してみましょう。ここでは 2 つの変数の関係を見たいので，X 軸に高校の成績を，Y 軸に大学の成績をとった**散布図 (scatter plot)** が適しています。図 9.1 に散布図を書いてみました。

散布図を見ると，データからなんらかの傾向が見えてきますよね。左下から右上にかけて，点が散らばっているのがわかります。そう，このデータには**正の相関関係**がありそうだということがわかります。

もちろん，**相関関係**があることが，すぐに**因果関係**があることを意味するわけではないのでしたね。が，今回の場合は，高校の成績は大学の成績に時間的に先行しますし，どちらも人の能力を反映しているものとして理解ができそうです。ということで，この相関関係から予測することは (理論的，あるいは社会通念的にも) 妥当ですから，私たちは「高校の内申点がよかった」ときくと「その人はきっと頭もいいんだろうね」「大学の成績も高いだろうな」と自然と考える＝予測するわけです。

予測は因果を仮定しなくても可能です。たとえば町の消防署の数と火災発生件数は高い相関係数を持ち

[*1] もう 1 つは「説明する」です。

表 9.1 高校と大学の成績例

学生 ID	高校の内申点	大学入試の成績
A	2.13	460
B	2.42	500
C	2.26	473
D	3.87	620
E	3.90	690
F	2.43	512
G	3.44	582
H	2.15	550
I	2.18	485
J	3.00	650
K	3.42	593
L	2.55	528
M	3.19	585
N	3.05	569
O	2.52	518
平均	2.834	554.33

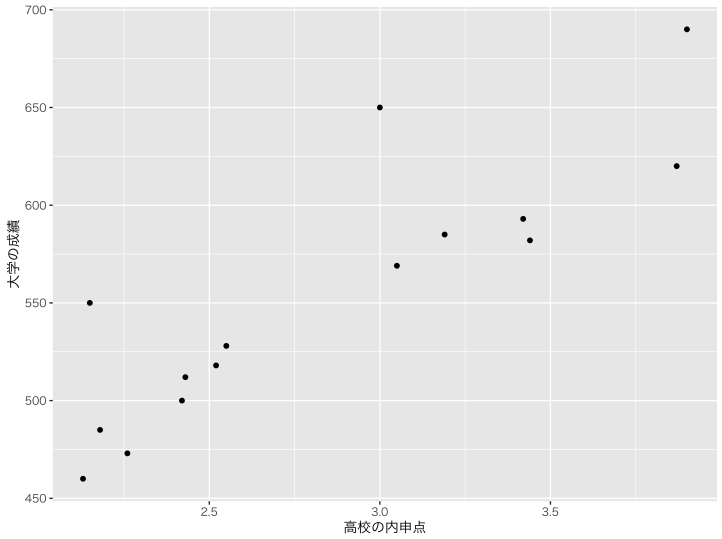


図 9.1 散布図を見てみよう

ますが，ここから消防署が火災発生の原因だ，と考える人はいないでしょう。でもデータで相関関係が示されていることがわかれば，ある町の消防署の数が多いと聞くと，火災の多い町なのかな，と予測することはできるのです。また，相関関係だけが予測のツールではありません。相関関係は変数の直線的な関係を表しているに過ぎませんから，直線的でない現象に対しては相関係数は低くなります。たとえば，血圧が高過ぎたり低過ぎたりすると，その人は不健康だなという予測はできるわけですから (前回の議論を参照してください)。

こうした注意は必要ですが，今回のデータは散布図 (図 9.1) からわかるように，相関関係も高く直線的な

関係がありそうだ，ということになります。この直線的な関係を，**関数 (function)** で表すことを考えたいのです。

データが数字で得られていると，変数間関係を関数で表現できます。関数は一般的に次のように書きます。

$$y = f(x)$$

これは，x になんらかの数字が入ると，関数 f でなんらかの加工がなされて，結果としての数字が y になる，という意味ですね。心理学が科学であるとはすなわち，数学的現象主義でアプローチすること，変数間関係を関数で表現するということです。この関数のことを**モデル (model)** とも言います。モデルとは抽象化された理想的なものですが，科学においては数学的な関数のことをモデルというのですね。もちろんこのモデル，つまり関数の形がどんなものなのかは，現象によってさまざまなものが考えられます。

ところで今回のデータは左下から右上にかけて，直線的な関係があるように思えます。直線的な関数は，みなさんすでに中学校の数学で学んだことがあると思います。直線的な関数は次のように表現できますよね。

$$y = ax + b$$

これこそ心理学で最もよく使われている，**線形関数 (linear function)** です。なぁんだ！と思いませんか。みなさんは多項式，二次式，三角関数など色々な関数を学んできたと思いますが，ここで使うのはただの $y = ax + b$ なんです。

ただし，中学校で学んだような一次式とちょっと違うところもあります。中学校の一次間数は，たとえば「ある 2 点 P,Q を通る一次関数の切片 b と傾き a を求めよ」，なんて話だったかと思いますが，今回のこの散布図にはご覧の通り，点は複数あってどこか二点を選んで線を引くわけではありません。しかも，全部の点を通るような直線はないのです。

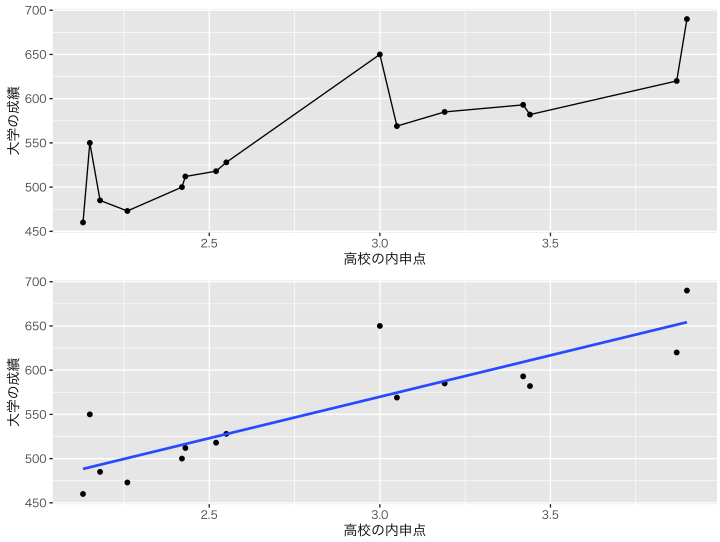


図 9.2 線を引くのも難しい

図 9.2 の上の図は，すべての点を通る線を引いてみたのですが，当然直線にはなりません。下の図は直線ではありますが，すべての点を通っているわけではありません。しかし，目指すのは下のような直線です。すべてを説明することはできませんが，だいたい当てはまっている，そういう関係を探そうというわけです。データには誤差や個人差がありますから，そのせいで直線関係が当てはまらないところがありますが，それらを除外して理想的な関係を考えれば直線なんじゃないか，というのがここでの課題になります。

線形関数は傾き a と切片 b が変われば，色々な直線を引くことができます。ここでの狙いは**データに最も当てはまりがマシな直線関係を引く**ということになります。

少し余談ですが，データが直線関係にない場合や，データになるべくぴったり当てはめることを目的にすると，直線であるという縛りを捨てる必要があります。それは**非線形関数 (Non-Linear function)** と呼ばれ，AI(人工知能) や機械学習と呼ばれているのはこの非線形関数を使ってデータの予測精度を上げていくものです。ですが，非線形関数は n 次多項式になりますから，何が何やら意味がわかりません。意味がわからなくてもデータに当てはまればいいんだ，と言われたらそういうこともあるか，と思いますが，心理学のような学術的アプローチは人間一般に通じる理論が欲しいのですから，説明できない関数は不要なのです。なので，データにぴったり当てはまるとはいかないけれども，線形で大体説明できるところを探そうじゃないか，というわけです。

9.2 予測と線形モデル

それでは具体的に，目に見えない理想的な線形関係 $y = ax + b$ を探していきましょう。

ここで改めて，記号の表記やデータとの対応を丁寧に考えてみたいと思います。確かに中学数学でやるような $y = ax + b$ の関数は見つかりませんが，今回はたくさんのデータ点があります。図 9.1 の各点は，表 9.1 の各行に対応しています。このグラフの中の点が持つ座標は，データの値なのですから，たとえば一番右上にある点は E さんの $(3.90, 690)$ というデータの組み合わせになります。このように x, y の組み合わせがたくさんあるので，座標点は (x_i, y_i) のように書くべきですね。この y_i は説明される側の変数ですから，**被説明変数 (predicted variable)**，あるいは**従属変数 (dependent variable)** といい，x_i は説明する側の変数ですから**説明変数 (predict variables)**，あるいは**独立変数 (independent variables)** というのでした。さてこれを踏まえて，求める式を次のように表します。

$$\hat{y}_i = b_0 + b_1 x_i$$

ここで $\hat{y}_i$ の y の上についている記号は「ハット」といいます。関数がぴったりと当てはまるわけではないのですから，y_i と書くわけにはいかなかったのです。あくまでも理論的な**予測値 (predicted value)** ですよ，ということを表すための記号がハットです。また，a の代わりに b_1，b のかわりに b_0 としました。a とか b はただの記号ですので，他のどんな文字でもいいのですが，「1 つ目の説明変数についてくる係数」という意味で b_1 にしました[*2]。この 1 は 1 つ目の変数，という意味です。切片は説明変数を持ちませんから，0 番目にしたということです。今はまだ b_0, b_1 の値がわかりませんが，これがわかれば，x_i のところに適当な数字を入れることで，変数 y の値はこれぐらいなんじゃないか，という予測ができるともいえます。

このようにして引かれた直線 $\hat{y}$ は，線型結合で表された関数 $b_0 + b_1 x_i$ の解の集合です。b_0, b_1 の値はわからないので今からそれを探しに行くわけですが，x_i, y_i の組み合わせはデータとしてわかっているのでした。たとえば J さんのデータ，記号でいうなら x_{10}, y_{10} ですが，それはそれぞれ $x_{10} = 3.0, y_{10} = 650$ となります。しかし $\hat{y_{10}}$ は実際のデータのずいぶん下の方にあります (図 9.3)。これは J さんについてはこの関数の予測が外れていることを意味しますね。この外れている大きさを**誤差 (error)** あるいは**残差 (residuals)** といい，記号 e_i で表します。残差の記号 e に個人の違いを表す添字 i がついてあることからわかるように，残差はすべてのデータについて生じます。残差は次のように定義されます。

$$e_i = y_i - \hat{y}_i$$

[*2] なんで b なんだ a でもいいじゃないか，と言われたら返す言葉もないのですが，慣例的に b で係数を表しています。ちなみに今はアルファベットの b ですが，後期になると b のギリシア文字 β に変わります。ギリシア文字で表現するのは推測統計学的な推定に関わるようになってからです。

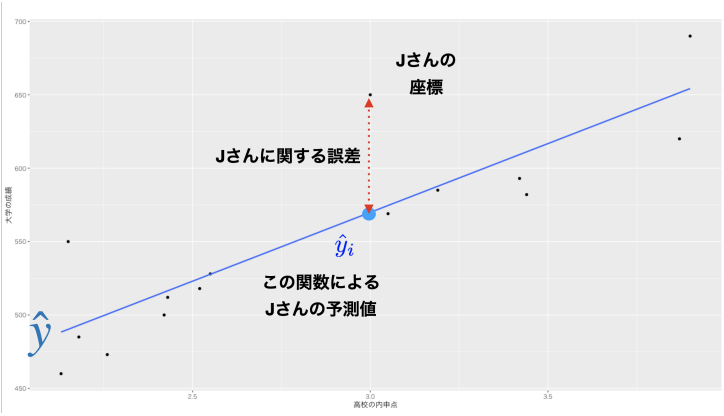


図 9.3 データと予測値と誤差の関係

予測値 $\hat{y_i}$ と実測値 y_i の差分が誤差 e_i だ，というわけですね。この式は次のように展開できます。

$$y_i = \hat{y}_i + e_i = b_0 + b_1 x_i + e_i$$

ここで左辺がデータ (実測値) になりました。モデルからデータを予測する式ができたわけです。

これで表記上の準備ができました。あとは b_0, b_1 の値を求めるだけです。b_0 は直線がどこから始まるかという基準となる高さを表しますし，b_1 は直線の角度，傾きを表すものです。この 2 つが決まっていないということは，どのような直線でも引くことができてしまうということでもあります。全部のデータ点にピッタリ合うような線を見つけることができないのですから，全体的に一番いい感じの線にしたいのですが，この「全体的に一番いい感じ」を数学的にどう言い表せば良いでしょうか？

9.3 最小二乗法

図 9.4 の上に，直線 A,B という 2 つの線を試しに引いてみました。直線 A のようなモデルだと，ある点は通っている＝あるデータ点については予測が実際の値に一致しているのですが，他のデータ点はすべて線 (予測値の集まり) の下方向，つまり予測値が全部実際より大きい方向にズレていることになります。逆に B のモデルだと実際のデータ点はひとつを除いてすべてモデル直線 (予測値の集まり) の上方向にあります。つまり予測値は全部低すぎるので，これも全体としてのズレは大きいと言わざるを得ないでしょう。

いいモデル式というのは，すべてのデータ点にフィットすることはない代わりに，プラスにもマイナスにも，均等にズレているものだということです (図 9.4)。個別には誤差がゼロにならないけど，全体としてはちょうど中庸のラインを通るように，切片と傾きを定めるのが良いようです。これを数式的に考えてみましょう。個別の誤差は次のように表現できるのでした。

$$e_i = y_i - \hat{y}_i$$

これが全体的に真ん中を通るようにしたいわけです。全体的に，というのは言い換えると誤差を総和した時であり，真ん中を通るというのはプラスにもマイナスにも同じだけ，つまり合わせてゼロになるのが目標です。

$$\sum_{i=1}^{N} e_i = 0$$

これが目標なのですが，これは平均するとゼロになるということと同じで，誤差がどれぐらいの大きさだったのかがわかりません。誤差をなるべく小さくするような直線を求めているのですが，誤差の大きさが評価で

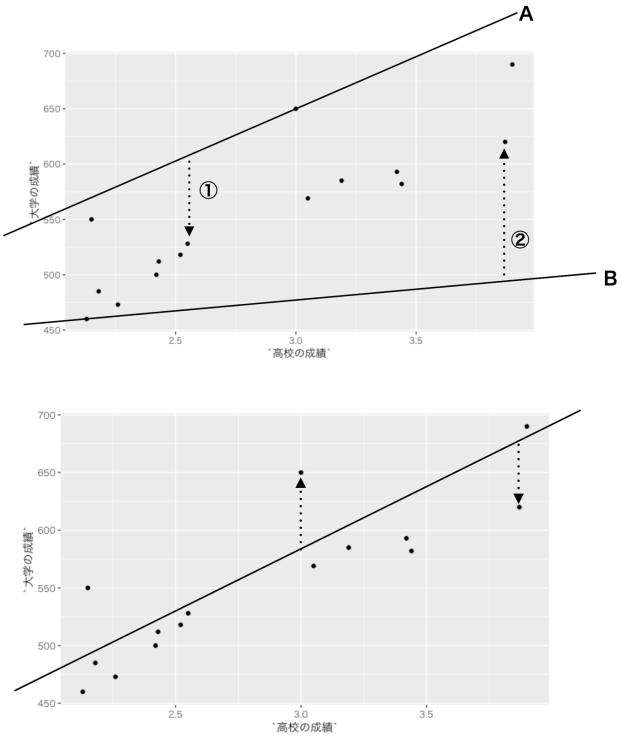


図 9.4 考えうる色々な予測の線

きないのでは困ります。ここで思い出して欲しいのですが, 正規分布の時にあったように, 誤差は位置よりも (中心はゼロですから), 幅が重要です。位置ではなく幅でその大きさを評価するのです。幅, すなわち散らばりを表す指標を考えるには, 分散の時のようにこれを二乗した値を考えます。

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - \hat{y_i})^2 = \sum_{i=1}^{N} (y_i - (b_0 + b_1 x_i))^2 = Q(b_0, b_1)$$

この式の最後にあるように, 最終的には未知数 b_0 と b_1 の関数が出てきます。これを Q とあらわし, この Q が最小値を取るように b_0, b_1 を求める, という条件で方程式を解いていけば良いことになります。

最後に出てきたこの関数, 仮に Q としてありますが, これは切片 b_0 や傾き b_1 の取り方で変わる量ですが, 二乗の関数ですので極値を求めることができます。関数の極値を求めるためには微分が必要で, しかも今回は 2 変数の微分ですから偏微分という手法が必要になります。この手法について詳しくは小杉 (2018) や西内 (2017) などを参考にしていただければと思いますが, この授業の範疇を超えるのでここでは細かい計算については割愛します。

このように誤差を最小にするようにデータにモデルをフィットさせる方法, あるいは誤差を最小にするという基準で未知数を推定する方法とも言えますが, この方法のことを**最小二乗法 (Least Squares Method)** と呼ぶということを知っておいてください。この方法を考え出したのが, 正規分布を導出したガウスさんなんですね。すごいですねえ。

最小二乗法で係数を求めるのは, 具体的には次のような計算式になります。

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = r_{xy} \frac{s_y}{s_x}$$

ここで r_{xy} は変数 x, y の相関係数，s_x, s_y はそれぞれ x と y の標準偏差，$\bar{x}, \bar{y}$ はそれぞれ x と y の平均値です。いかがですか！すべてデータから計算できる記述統計量で，切片と傾きを求めることができました。もちろん具体的に数字を入れて計算していってもいいのですが，実際には統計環境 R など統計パッケージを使えば一瞬で答えを出してくれます。計算過程を理解しておくことも重要なのですが，まずはデータと条件から解が求められることを理解し，どのような条件，どういう基準で何を求めたのかをしっかり確認しておいてください。

数値例でみてみましょう。表 9.1 のデータから計算される記述統計量は，まず $r_{xy} = 0.868357, s_x = 0.6185097, s_y = 66.75292$ を計算して，傾きを求めます。

$$b_1 = 0.868357 \times \frac{66.75292}{0.6185097} = 93.71779$$

次にこの b_1 と $\bar{x} = 2.834, \bar{y} = 554.333$ を使って，切片を求めます。

$$b_0 = 554.3333 - 93.71779 \times 2.834 = 288.7371$$

つまり，予測式は次のようにすれば誤差を最小にした，言い換えればすべてのデータ点から満遍なく隔たった直線が引けるということです。

$$\hat{y}_i = 288.7371 + 93.71779 x_i$$

この分析方法のことをとくに**回帰分析 (regression analysis)** といいます。今回は変数 X から変数 Y を予測する関数，$y = f(x)$ を作りましたが，同じデータを使って変数 Y から変数 X を予測する関数，$x = g(y)$ を考えて係数を求めることができます。この関数を交互に使って，$y^1 = f(x) \to x^1 = g(y^1) \to y^2 = f(x^1) \to x^2 = g(y^2) \to \cdots$ と続けて行くと，最終的に x^n は変数 X の平均値に，y^n は変数 Y の平均値に近づいて行きます。このように回って回って平均に帰って行くので，**回帰**という名前になっています[*3]。回帰分析によって得られた係数のことを**回帰係数 (regression coefficients)** ともいいます。

ここで回帰係数の切片項，b_0 は何かというと，$x = 0$ の時の値になります。今回の例でいうと，内申点が 0 点であったら大学でどれぐらいの成績が取れるか，ということですね。つまりこれはベースラインだといえます。では傾きを表す b_1 は何でしょうか。これは x の変数が 1 上がった時の，$\hat{y}$ の上昇率を表しています。今回の例では，内申点が 1 点上がったら大学での成績がどれぐらい上がると予想されるか，ということですね。

9.4 回帰分析の性質

ここで，回帰分析の数理的特徴を確認しておきましょう。

[*3] 親の身長から子供の身長を予測することを考えてみてください。次にその子が親になって，その子がまた親になって･･･という時に，「背の高い親からは背の高い子しかうまれない」というルールであれば，世界はものすごく高身長なひとと低身長な人に二分されてしまいます。そうではないということは，背の高い親から背の低い子が生まれることもあるし，その逆もあるということです。また野球選手などでルーキーのときは成績がいいのに，2 年目になったら成績が下がる，というのは 2 年目のジンクスなどといわれますが，その人の平均的な能力以上のスコアが出た次の年に，平均的な値に帰るために低くなりがちなことは，当然ありえることでもあるのです。こうした効果のことを回帰効果といいます。

9.4.1 説明変数と残差の共分散

説明変数 x_i と残差 e_i の共分散を計算してみましょう。

$$s_{xe} = \frac{1}{N}\sum(x_i - \bar{x})(e_i - \bar{e}) \tag{9.1}$$

$\bar{e} = 0$ ですから, (9.2)

$$= \frac{1}{N}\sum(x_i - \bar{x})e_i \tag{9.3}$$

$$= \frac{1}{N}\sum(x_i e_i - \bar{x}e_i) \tag{9.4}$$

$$= \frac{1}{N}\sum x_i e_i - \frac{1}{N}\bar{x}e_i \tag{9.5}$$

$$= \frac{1}{N}\sum x_i e_i - \bar{x}\frac{1}{N}e_i \tag{9.6}$$

$$= \frac{1}{N}\sum x_i e_i - \bar{x}\bar{e} \tag{9.7}$$

$\bar{e} = 0$ ですから, (9.8)

$$= \frac{1}{N}\sum x_i e_i \tag{9.9}$$

ここで，$e_i = y_i - \hat{y}_i$ であること，$\hat{y}_i = b_0 + b_1 x_i$ であることから，$e_i = y_i - (b_0 + b_1 x_i)$ であることを思い出して，式の中に入れてみます。

$$\frac{1}{N}\sum x_i e_i = \frac{1}{N}\sum x_i(y_i - (\bar{y} - b_1\bar{x} + b_1 x_i)) \tag{9.10}$$

$$= \frac{1}{N}\sum x_i y_i - x_i\bar{y} + x_i b_1\bar{x} - x_i^2 b_1 \tag{9.11}$$

$$= \frac{1}{N}\sum x_i y_i - \frac{1}{N}x_i\bar{y} + \frac{1}{N}x_i b_1\bar{x} - \frac{1}{N}x_i^2 b_1 \tag{9.12}$$

$$= \frac{1}{N}\sum x_i y_i - \bar{y}\frac{1}{N}\sum x_i + b_1\bar{x}\frac{1}{N}\sum x_i - b_1\frac{1}{N}\sum x_i^2 \tag{9.13}$$

$$= \frac{1}{N}\sum x_i y_i - \bar{x}\bar{y} + b_1\bar{x}^2 - b_1\frac{1}{N}\sum x_i^2 \tag{9.14}$$

$$= \frac{1}{N}\sum x_i y_i - \bar{x}\bar{y} - b_1(\frac{1}{N}\sum x_i^2 - \bar{x}^2) \tag{9.15}$$

(9.16)

さらにここで，分散 (→Pp.87) と共分散の式 (→Pp.130) を思い出してみましょう。

$$s_x^2 = \frac{1}{N}\sum x_i^2 - \bar{x}^2$$

$$s_{xy} = \frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum x_i y_i - \bar{x}\bar{y}$$

このように展開できるのですから，この式は，

$$= s_{xy} - b_1 s_x^2 \tag{9.17}$$

となります。

さらに，$b_1 = r_{xy}\frac{s_y}{s_x} = \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x}$ ですから，

$$= s_{xy} - \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x}s_x^2 \tag{9.18}$$
$$= s_{xy} - s_{xy} \tag{9.19}$$
$$= 0 \tag{9.20}$$

となります。

この式が意味するのは，説明変数 x_i と残差 e_i は相関しない，共変動しないということです。まずはこのポイントを押さえておいて，次に被説明変数 y_i と残差 e_i の関係を見ていきます。

9.4.2 説明変数の分散

話に先立って，変数 $x_i = a_i + b_i$ としたときの分散 s_x^2 を考えてみます。

$$s_x^2 = \frac{1}{N}(x_i - \bar{x})^2 \tag{9.21}$$
$$= \frac{1}{N}((a_i + b_i) - (\bar{a} + \bar{b}))^2 \tag{9.22}$$
$$= \frac{1}{N}((a_i - \bar{x}) + (b_i - \bar{b}))^2 \tag{9.23}$$
$$= \frac{1}{N}((a_i - \bar{x})^2 + 2(a_i - \bar{x})(b_i - \bar{b}) + (b_i - \bar{b})^2) \tag{9.24}$$
$$= \frac{1}{N}(a_i - \bar{x})^2 + 2\frac{1}{N}(a_i - \bar{x})(b_i - \bar{b}) + \frac{1}{N}(b_i - \bar{b})^2 \tag{9.25}$$
$$= s_a^2 + 2s_{ab} + s_b^2 \tag{9.26}$$

このように，合成された変数の分散はそれぞれの分散と共分散の 2 倍との和になります。これを踏まえて，$y_i = \hat{y}_i + e_i$ であることから，

$$s_y^2 = s_{\hat{y}+e} \tag{9.27}$$
$$= s_{\hat{y}}^2 + 2s_{\hat{y}e} + s_e^2 \tag{9.28}$$

となりますが，$\hat{y}$ と e の共分散は 0 ですから[*4]，

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

となります。つまり，被説明変数 y の分散は，予測値 $\hat{y}$ の分散と残差 e の分散とにキッパリ分離されるということです。

残差は説明変数 x と相関しないわけですから，回帰分析の式は予測変数 x で y を説明し尽くしていて，どうにも説明できない y のあまりの部分が全部残差になっている，ということがわかります。この「回帰分析は，説明変数で余すことなく被説明変数の分散を取り尽くしている」という点が重要なポイントです。式から明らかなように，説明変数は精一杯のことはしているんですが，それでも説明できない部分はどうしても残ります。それが残差になっているわけで，そこには説明変数の要素がカケラも入っていない，というところが重要です。

では，説明変数の分散 s_y^2 のうち，説明できた $s_{\hat{y}}^2$ の部分と s_e^2 の部分は，どれぐらいの比率なのでしょうか。それが次の評価のところに関わるわけです。

[*4] 確認してみよう！詳しくは小杉 (2018) にあるよ。

9.5 予測の評価

さて，ともかくこうして回帰係数を求めることができました。これはとても嬉しいことですね！嬉しいことですよ！だってデータから関数関係がわかったのですから。関数関係がわかるということは，$y = f(x)$ の x に数字を入れれば y の値がわかるということです。実際のデータ x_i をいれると $\hat{y}_i$ がわかります。これを見ることで，そのデータ点がどの程度モデルから外れているかということがわかります。誤差の大きさを量的に見積もれるわけです。さらに，もちろん，未来が予測できますね！実際のデータではなく，新しい数字を入れれば予測値が出てくるわけです。たとえば今回の例で言えば「高校の内申点が 2 点だった」とすれば，大学での成績は $288.7371 + 2 \times 93.71779 = 476.1727$ と計算できます。わざわざ大学に進学しなくても，この程度の内申点ならこれぐらいの成績をあげるだろう，ということが事前にわかるということです。素敵！

実は，BIG データだとか，AI だとかで騒いでいる話の基本は，これのことなのです。数式に入れるデータが大量であること，複数の予測変数を使うことなど，このモデルを拡張していく必要はありますが，基本はこのような「モデルをデータに当てはめて予測する」というところにあるわけです。今回はその最も単純な形，すなわち一次関数・線形モデルでしたが，人間の行動，市場の動向，あるいは次のレースで勝つ馬はどれか，なんてこともモデルを複雑にしつつ予測して行くことができるのです。

ただここで注意しておいてもらいたいのは，このように計算ができることと予測が当たることは別だ，ということです。本当に内申点 2 点のひとが，大学で 476 点になるかどうかはわかりません。なぜなら，今回のこの係数は得られたデータに最もフィットした形で算出しましたが，それでもすべての点に合致しているわけでなく，一番当てはまりの程度がマシだ，というだけであり，実際には必ず予測できない残りの部分，すなわち**残差 (residuals)** があるからです。

このように，モデルをデータに当てはめて分析する時は，その当てはまりの程度を別途評価しなければなりません。この評価の指標のことを**適合度 (fit index)** といいます。回帰分析の場合，予測値は $\hat{y}_i$ であり，実際のデータは y_i です。両者の関係は $\hat{y}_i - y_i = e_i$ でしたね。予測値と実際のデータがどれぐらい近いかを表すために，この 2 つの相関係数 $r_{\hat{y}y}$ を考えることにします。相関係数ですから，この数字は $-1 \leq r_{\hat{y}y} \leq 1$ の範囲内に収まるのですが，負の相関であっても相関が高ければある意味合致はしているのですから，ここでの符号にはそれほど興味がありません。そこでこの相関係数を二乗してしまいましょう。この $r^2_{\hat{y}y}$ のことを，R^2 **値 (R-squared)** と言い，適合の指標と考えます。R^2 **値**は別名**決定係数 (coefficient of determination)** ともいいます。

この R^2 値は，計算すると次の関係にあることがわかります[*5]。

$$R^2 = \frac{s^2_{\hat{y}}}{s^2_y}$$

言葉でいうと，**被予測変数の分散中に占める予測値の分散の割合**ですね。つまり，この予測モデルが予測される変数の情報のうち，何 % を説明したかがわかるわけです。

今回のデータについていうと，$R^2 = 0.754$ でした。予測モデルで 75% ぐらいは説明するので，まあまあ良いモデルだったと言えるかもしれません。論文などで回帰分析について書かれている場合は，この R^2 あるいは類する適合度指標をよく見るようにしましょう。結果は出ても，説明率がうんと低ければ，実質的になんの説明もできていない，予測にはまったく役立たないということでもあるのですから。

[*5] 細かい計算式については，小杉 (2018) の Pp.133 をみてください。

9.6 課題

■回帰係数を求めてみよう 表 9.1 のデータを使って，大学入試の成績から高校の内申点を予想する回帰分析を行なった時，その切片と傾きはいくらになるか計算してみてください。なお，大学の成績から高校の成績を予測するのですから，因果の向き (時間の向き) は逆転しています。これで因果関係がわかることはないのですが，データから相関関係があれば回帰はできるのであり，それをどういう意味で使うか，伝えるかは使う側の人の問題だということを忘れないでください。

第 10 章

重回帰分析

前回は**回帰分析**というタイトルで，データの相関関係を元に予測モデルを作るということをしたのでした。予測と制御は科学の基本ですから，それができるとなったら素敵なことです。ただし，我々が使うデータは人間を対象にしているので，誤差がたくさん含まれていますから理論モデルのようにぴったりフィットすることはありません。それでも大まかに線形関数が成り立つのであれば (直線の分布をしてそうならば)，有用なことは多いのだという話をしました。

今回は前回のモデルを少し複雑にしていきます。前回は $y = b_0 + b_1 x$ で，説明される変数 y に対して説明する変数 x がひとつだけ，という簡単なモデルでした。今回は説明する変数が 2 つ以上のモデルを考えます。

10.1 重回帰分析

説明変数が複数に増えた場合の回帰分析は，**重回帰分析 (Multiple Regression Analysis)** と言います[*1]。説明変数が $x_1, x_2 \cdots x_m$ と増えますので，係数もそれに応じて $b_1, b_2, \cdots, b_m$ とつきます。切片は b_0 ひとつで変わりません。これを式で表すと次のようになります。

$$\hat{y_i} = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_m x_{mi}$$

説明変数に x_{mi} とありますが，これは m 番目の変数についての i さんのデータということです。今示した式は予測値 $\hat{y}$ についての式ですが，実際にはこれに誤差 e_i がついて実測値 y_i になります。

$$y_i = \hat{y_i} + e_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_m x_{mi} + e_i$$

一般的に書くために x_m まで用意しましたが，たとえば 2 変数であれば $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_i$ となります。どこに i がついているか，どこについていないかといった細かいところまでしっかり確認してくださいね。

さて $y = b_0 + b_1 x$(あるいは懐かしい $y = ax + b$ という形でもいいですが) のときは直線になるので線形モデルという話をしました。説明変数が 2 つあると，x_1 の軸，x_2 の軸と従属変数 y の 3 つの軸からなる 3D 平面で図示することになります (図 10.1)。説明変数が 3 つ以上になると図示することはできません。しかし，抽象的・数学的なモデルとしては何の問題もありません。また，平面になってもある側面から見れば直線になっていますから，説明変数が増えても一般的に**線形回帰**あるいは**線形モデル**という名称はかわりません。

説明変数の数が増えると，回帰係数を求める計算はより複雑になります。複雑になりますが，結局のところ**最小二乗法**と同じ考え方，すなわち誤差 (の分散) が最も小さくなるように係数を定めるという基準で係数を

[*1] これに対して，説明変数が 1 つしかないモデルのことを**単回帰分析**と呼んで区別することもあります。

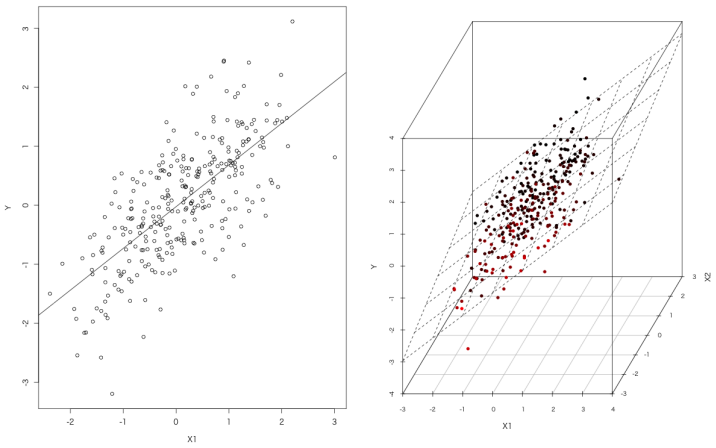


図 10.1 立体でも線形モデル

求めることができます[*2]。具体的な計算式はここで示しませんが，R を使えば単回帰分析と同じやり方で答えが出せる，ということだけ知っておいていただきたいと思います。

具体的な数値例でみてみましょう。表 10.1 には表 9.1 に一列追加したデータを用意しました。今度は大学の GPA つまり大学成績の平均値が入っています。高校の内申点，大学の入試の点数が，大学の評定平均に影響している，という考え方は十分あり得そうですね。

表 10.1 高校と大学の成績例 2

学生 ID	高校の内申点	大学入試の成績	大学の GPA
A	2.13	460	5.34
B	2.42	500	7.97
C	2.26	473	6.30
D	3.87	620	7.82
E	3.90	690	8.40
F	2.43	512	6.60
G	3.44	582	7.80
H	2.15	550	7.10
I	2.18	485	7.30
J	3.00	650	7.20
K	3.42	593	7.80
L	2.55	528	6.50
M	3.19	585	8.20
N	3.05	569	7.50
O	2.52	518	7.60

[*2] 後期に入ると，推測統計学や確率モデルの考え方を導入した，**最尤法 (Maximum Likelihood Method)** や**ベイズ法 (Bayesian Estimation Method)** といった他の推定方法・推定基準が導入されますが，いずれも未知数を求める方法の一種です。

このようなデータが得られたら，次にすることは，そう，可視化です。**データは図にする**が分析の基本だからです。

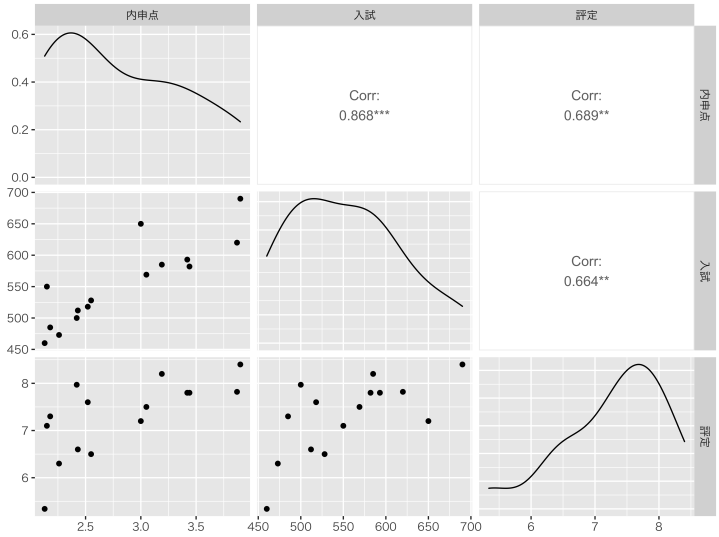


図 10.2　変数間関係の可視化

図 10.2 は 3×3 の行列にならべてデータを可視化する方法です。行に内申点，成績，GPA が，列にも内申点，成績，GPA が入っています。左上から右下にかけての対角 (セルの番地でいうと n 行 n 列目) には，データの分布を表す図が入っています。左下のほうにあるのは，2 つの変数ペアの散布図です。右上にあるのはその相関係数です。当然ながら，左下と右上には同じ情報が入っているのですが，これを可視化したものと数値化したものになっているのです。

このデータを見ると，3 つの変数の間には十分相関がありそうです。とくに，大学の評定平均値を従属変数と考えると，内申点と評定平均の相関が 0.689，入試と評定平均の相関が 0.664 とあり，散布図を見てもどちらも左下から右上にかけての直線関係があるように見えます。そこで，評定平均を y，内申点を x_1，入試の成績を x_2 と考えて重回帰分析を実行してみたとします。

R でチョチョイと計算できるのですが，詳しくは次回に回すとして，結果として次のような回帰式が得られました。

$$\hat{y} = 3.761 + 0.604x_1 + 0.003x_2$$

つまり，内申点を 0.604 倍し，入試の成績を 0.003 倍したものに 3.761 を足せば大学の平均評定 GPA の予測値になるのです。具体的に，たとえば内申点が 2 点，入試の成績が 500 点の人であれば，この人は $3.761 + 0.604 \times 2 + 0.003 \times 500 = 6.469$ ぐらいの成績を得るんじゃないの，ということがわかります。

また，仮想の値ではなく今回のデータを入れると，それぞれの予測値 $\hat{y}$ と，誤差の大きさが計算できます。表 10.2 には予測値を `Yhat` 列に，誤差の大きさ (残差) を `Residuals` 列に入れてみました。

予測値と実際の従属変数との相関係数が，モデルの**適合度**を表す指標になるのでしたね。その相関係数を二乗した R^2 **値**を計算すると，今回は $R^2 = 0.9418$ になります。入学後の成績の 9.5 割ぐらいは，入試と内申点で予測できるデータだったということができます。

重回帰分析は単回帰分析と変わらず，便利な使い方ができそうだ，ということがわかってきました。ただし，結果の解釈やそれを言葉にする時は，ちょっと注意が必要です。まず言葉上の注意ですが，重回帰分析の回帰係数はとくに**偏回帰係数**といいます。「偏」という字がつくのですが，この一文字が曲者です。

単回帰の時，傾きを表す係数 b_1 は説明変数 x_1 が一単位上昇した時の，従属変数 y の伸び率を表してい

表 10.2 予測値と残差

ID	内申点	入試	評定	Yhat	Residuals
A	2.13	460	5.34	6.560388	-1.2203875
B	2.42	500	7.97	6.866879	1.1031209
C	2.26	473	6.30	6.681601	-0.3816011
D	3.87	620	7.82	8.136849	-0.3168493
E	3.90	690	8.40	8.384655	0.0153447
F	2.43	512	6.60	6.912295	-0.3122954
G	3.44	582	7.80	7.752318	0.0476824
H	2.15	550	7.10	6.867773	0.2322275
I	2.18	485	7.30	6.672630	0.6273699
J	3.00	650	7.20	7.709539	-0.5095393
K	3.42	593	7.80	7.776324	0.0236764
L	2.55	528	6.50	7.037309	-0.5373092
M	3.19	585	8.20	7.611085	0.5889147
N	3.05	569	7.50	7.473985	0.0260146
O	2.52	518	7.60	6.986369	0.6136308

るのでした。傾きってそういうことですよね。たとえば前回の例で言えば，$\hat{y} = 288.74 + 93.72x$ でしたから，内申点が 1 点上がれば入試の成績は 93 上がるぞ，というようなことが言えたわけです。

今回は $\hat{y} = 3.764 + 0.604x_1 + 0.003x_2$ ですから，「内申点が 1 点上がれば GPA は 0.604 上がるし，入試の点が 1 点上がれば GPA は 0.003 上がるぞ」と言いたいところです。がちょっと注意して欲しいのは，「内申点が 1 点あがれば」「入試の点が 1 点上がれば」という言い方です。この 2 つはまったく独立に効果が出ているようですが，図 10.1 を思い出せばわかるように，重回帰分析の場合の説明関数は面になっているということです。面の場合，x_1 が 1 ポイント上がるというのは，x_1 軸の方に水平移動するイメージですが，その時 x_2 は変化していません。x_2 が 1 ポイントあがるというのは，x_2 軸の方に水平移動するイメージですが，その時 x_1 は変化していないのです。

ところで，元のデータを図示した図 10.2 をみてください。内申点と入試の成績，つまり x_1 と x_2 の間には相関がありますね ($r = 0.868$)。ということは，x_1 が動くと x_2 もつられて動きます。データの空間上は斜めに動いているようなものです。であれば一概に，「内申点が 1 点上がれば･･･」ということはできません。重回帰分析の場合の回帰係数が表しているのは，「入試の点が同じであれば，内申点が 1 点上がれば 0.604 点上がる」という条件付きの係数だったのです。この点，注意が必要ですのでもう少し詳しく読みといていきましょう。

10.2 部分相関・偏相関・偏回帰係数

先ほどの表 10.2 から散布図を書いた図 10.3 をみてください。

お気づきでしょうか。これを見ると，相関係数がきれいに $r = 0.0$ になっている箇所がありますね。そうです，残差 e_i と 2 つの説明変数 x_1, x_2，および残差 e と予測値 $\hat{y}$ のところがゼロです。前回の回帰分析のところ (→Pp.147) でもやりましたが，これがどういう意味を持つのか考えていきます。

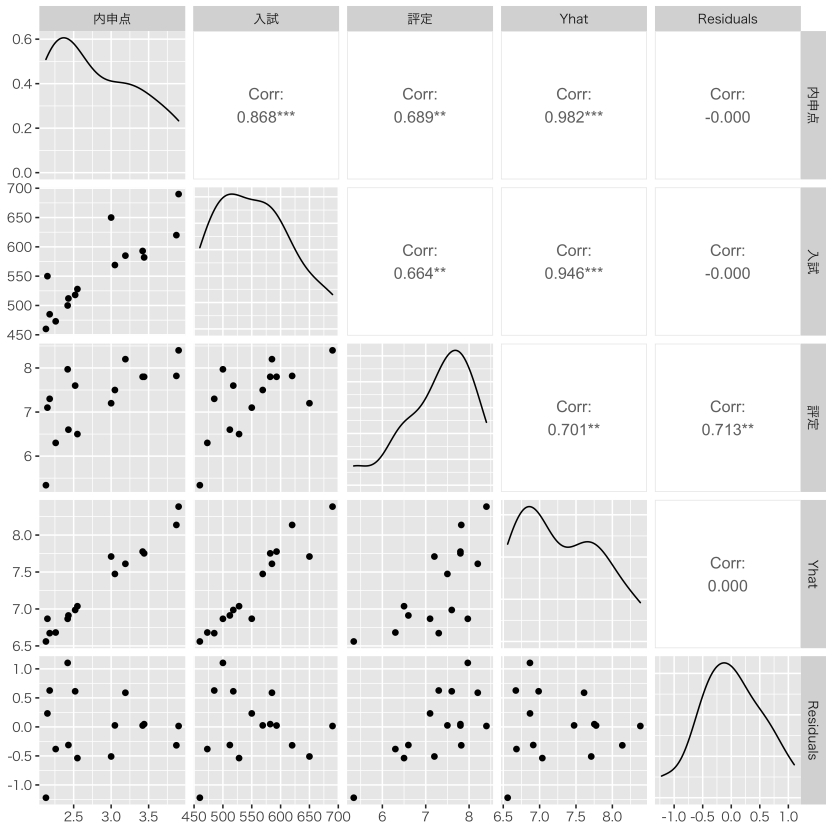


図 10.3 予測値や残差を含んだ散布図

まず残差と説明変数の相関がないことについて。被説明変数 y を説明変数 x_1 で回帰すると，この変数で説明できるパートとできないパートに分割できます (図 10.4)。説明できないパートが残差 e となって取り出されているわけです。ここで x_1 と e に相関関係が残っているようであれば，回帰分析の改良の余地があるということになってしまいます。回帰係数は今回のデータにベストフィットするように選ばれているはずなので，原理的にここには相関があるはずないのです。図 10.4 にそのイメージを描いてみました。

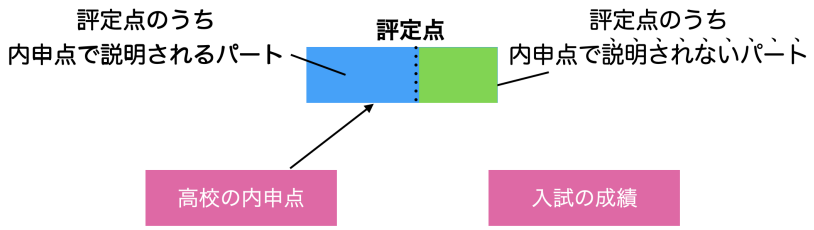


図 10.4 回帰分析は説明変数を分割する

さて，この説明変数 x_1 で説明できなかった残りの部分と，もう 1 つの説明変数 x_2 について相関係数を計算できます。この相関係数のことをとくに**部分相関係数 (semi-partial correlation coefficient)** といいます。図 10.5 のイメージ図をみてもらったらわかりやすいと思います。

これは被説明変数 y である「大学での評定平均点」を，第一の説明変数 x_1 である「高校での内申点」で**統制した**，「大学入試得点」(x_2) との部分相関係数，というように言います。統制したというのは，その変数でコ

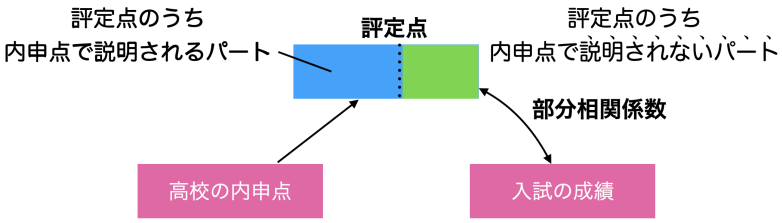


図 10.5 部分相関係数

ントロールした，あるいはその変数の影響を除外した，という意味だと思ってください。なぜ影響を除外したと言えるかというと，内申点 (x_1) で説明できるところは説明し尽くした残りとの相関を見ているわけですから，「説明し尽くした」=「その変数の影響を除外した」と解釈できるからです。

さらに，第一の説明変数 x_1 をつかって，第二の説明変数 x_2 を従属変数とした回帰分析をしたとします。そうすると，その残差は x_2 の中からも x_1 の影響を取り除いたものだ，ということができます。このように，内申点 x_1 をつかって，入試の点 x_2 と評定点 y 両方を回帰分析し，それぞれの残差の相関を求めたものを**偏相関係数 (partial correlation coefficient)** と言います。両方からある変数の影響力を取り除いた，偏りのない相関係数という意味です。図 10.6 にそのイメージ図を描いてみました。

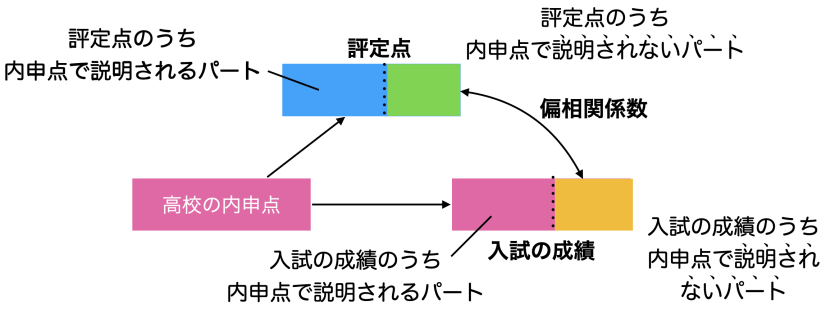


図 10.6 偏相関係数

入試の成績にも，大学での評定点にも共通の「高校時代の成績」といった要因がある，と考えることができますね。つまり「入試の成績と大学での成績を純粋に比較したい」というときであっても，背後の変数の影響がどうしても残ってしまっています。でもこの方法を使うと，背後の変数の影響を取り除いた純粋な「入試の成績と大学での成績」の関係を見ることができるようになります。これはすごいことなのです。たとえば，大学での評定平均値は他に何で予測できるだろうか，ということを考えたとします。それに影響してくる要素の候補として，たとえば通学にかかる時間だとか，その学生がアルバイトしているかどうか，サークルに入っているかどうか，あるいはこれまで塾に通ってきたかどうか，ひいては親の収入が高かったかどうか，兄弟姉妹の数が何人か・・・などなど，考えなければいけないことは山のように出てくるでしょう。しかもこういった情報は，調査の段階で統制できるものではありません。アルバイトしているかどうかで勉強に費やせる時間が違うな，と思っても，アルバイトしている人だけ回答者を募集したり，していない人だけ募集したり，というのではコストも大変です。実験的には統制できることでも，調査の実際を考えると統制すべき変数が多すぎて制御しきれないのです。

しかしここで，考えうる諸般の事情をデータ化して回帰分析を行い，その評定平均の残差を使うことにすると，それは諸般の事情が取り除かれたものとして扱うことができます。残差を研究することで，統制できないけどいろいろ考えなければならない効果からの影響を，綺麗に取り除ことができるともいえるのです！表 10.2

の Residuals の列には，2 つの説明変数の影響を取り除いた大学での評定平均点情報が入っています。内申点や入試の点は人それぞれ違いますが，それらの影響を取り除いて，評定平均点はどういう違いがあるのかな，ということを考えることができるのです。

さらに，ここでは x_1 で統制された y と x_2 との相関関係を求めましたが，x_1 で x_2 を統制した残差をつかって，y を回帰することもできます (図 10.7)。実はこれが重回帰分析の時に計算されている，回帰係数になっているのです。

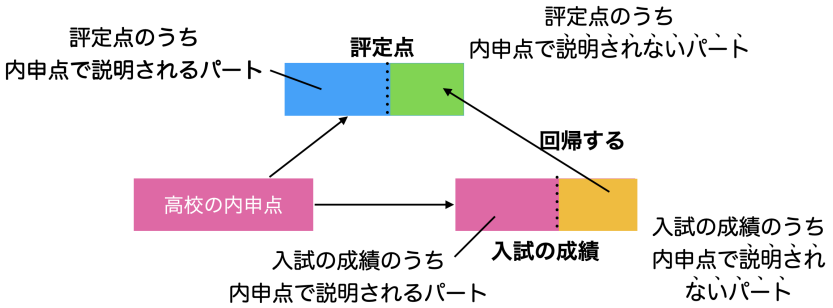


図 10.7 偏回帰係数

重回帰分析の回帰係数はとくに，**偏回帰係数 (partial regression coefficient)** といいます。偏という字がついているのは，相互に影響を取り除いたという意味です。前の節の最後に，回帰係数の意味は x_1, x_2 それぞれの軸に沿った水平移動をした時の上昇率だ，ということを伝えました (図 10.8)。ここで x_1 の係数を解釈するときには，x_2 が統制されている (x_2 は変化していない) ことに注意が必要です。言葉で表現するのは難しいのですが，丁寧にいうなら「入試の成績 (x_2) が同じだとすると，内申点 (x_1) が 1 点上がるごとに入学後の成績 (y) は 0.604 点あがる」が正しい表現ということになります。

「なんかややこしいなあ」「何がダメなのかわかりにくい」という人もいるかと思いますので，別の例で考えてみます。大学の成績 y を予測するのに，今度はサークルにかける時間 x_1 とバイトにかける時間 x_2 をというデータを集めたとします[*3]。まず，それぞれの変数で単回帰分析をしてみました。

$$\hat{y} = 2.8743 + 0.1407x_1$$

$$\hat{y} = 15.4221 - 0.5443x_2$$

なるほど，サークルをすると成績が上がる (+0.1407) のですが，バイトをすると成績が下がる (-0.5443) ようだ，という解釈ができるかと思います。

しかしここで重回帰分析をしてみます。2 つの変数を同時に予測方程式の中に入れるのです。

$$\hat{y} = 21.9910 - 0.1209x_1 - 0.5956x_2$$

おや，どちらの係数もマイナスになってしまいました。サークルをすると成績が下がるし，バイトをしても成績が下がるという結果です。サークル活動はいいんじゃなかったっけ？ なぜこのような逆転現象が起きたのでしょうか。

実は，このデータは x_1 と x_2 の間に相関関係があったのです ($r_{12} = -0.4360$)。サークル活動をしているとバイトしている暇がない，あるいはその逆でしょうか。いずれにせよ負の相関があることをこのように解釈できます。また，被説明変数との相関は，$r_{y1} = 0.1421, r_{y2} = -0.5529$ のようになっています。これは単回帰分析の結果と同じ傾向をしていますね。サークルに費やす時間が多い人は成績が高い，バイトの時間が長い

[*3] 説明をするための架空のデータです。

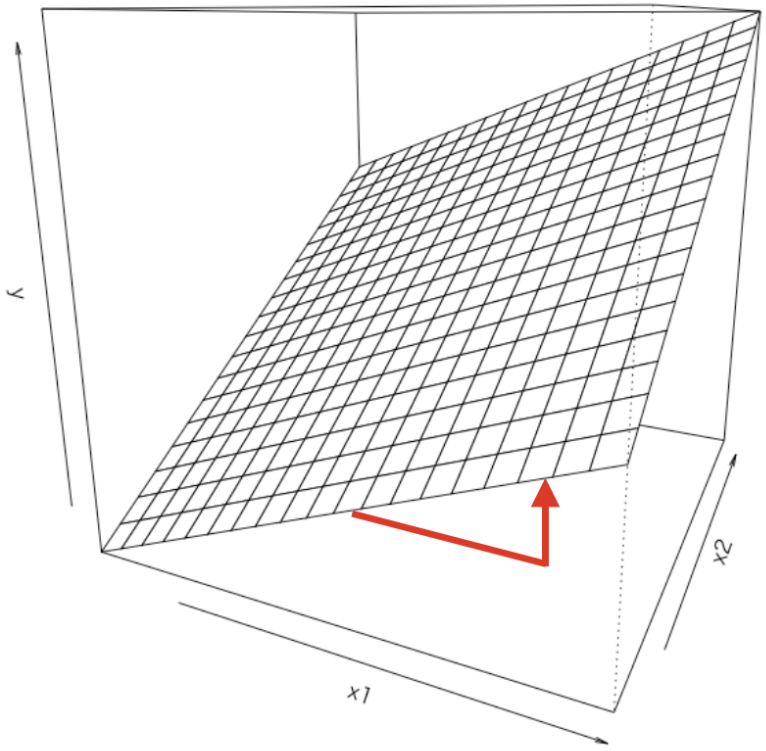


図 10.8 偏回帰係数が表す値の意味

人は成績が低い，ということがあるようです。しかし注意が必要なのは，このように $x_1 \to y$，$x_2 \to y$ という直接的な関係には，実は見えない影響が混ざっていることです。今回は x_1 と x_2 に関係がありますから，$x_1 \to x_2 \to y$ という影響力のルートがあるのです。これを**間接効果 (indirect effect)** といいます。

重回帰分析は間接効果を統制した影響を示してくれています。「サークル活動時間が多ければ成績が下がり，バイトにかける時間が多くても成績が下がる」というのではなく，「バイトにかける時間が同じであれば，サークル活動時間が多ければ成績が下がる」「サークル活動時間が同じであれば，バイトにかける時間が多いと成績は下がる」というのが正しいのです。両方が独立しているような，並列的な結論を書くのではなく，条件付きで解釈する必要があります。また，もちろん，バイト時間のことを調べていなければ，「サークルはいいよ，成績も上がることがデータで示されているよ」という結論を出してしまうかもしれません。分析する時は，被説明変数に影響しそうな変数は網羅的に考えておかないと，間違った結論を導いてしまうのです。

先ほど見たように，単純な相関係数 r_{y1} は，統制された直接的な影響 (偏相関係数) と，間接効果が入り混じったものになっているので，しっかり統制して評価しなければなりません。重回帰分析の結果を解釈するときは，その表現も注意しなければならないことがおわかりいただけましたでしょうか。

また，説明変数間に強い相関関係があるときは，重回帰分析の結果そのものが不正確なものになることがあります。この問題をとくに**多重共線性 (multi-collinearity)** の問題といいます。仮に x_1 と x_2 が $r_{12} = 0.9$ とか 1.0 に近いような，かなり相関が強い関係にあったとしましょう。$x_1 \simeq x_2$ といってもいいような感じです。仮に x_2 を x_1 に置き換えてみたとすると，回帰式は次のようになります。

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

ここで $x_1 \simeq x_2$ と，ほとんど同じなのですから，$x_1 = x_2$ として置き換えてしまいましょう。

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_1$$

係数をまとめて次のようになります。

$$\hat{y} = b_0 + (b_1 + b_2) x_1$$

最後の式は，単回帰分析と同じ形になっていますね。単回帰分析の係数は 1 つだけですからある係数 $b_s = b_1 + b_2$ を満たす b_1, b_2 のペアは無数に考えられます。解が定まらないのです。実際は $x_1 = x_2$ と置き換えられるほどのことはないでしょうが，かなり相関が高いと同じようなことが起こります。多重共線性が生じると，結果的に回帰係数の符号が真の関係と逆になって出てくるなど，正しく答えが出てこない問題が生じます。

このように，説明変数間に相関関係がある場合はとくにいろいろ注意しなければなりません。説明変数同士の相関がないときに用いると重回帰分析は大変シンプルなのですが，実際はそういうことばかりではありません。解釈や多重共線性に注意するためにも，得られたデータはまず可視化して，説明変数同士に相関関係があるかどうかをチェックするようにするといいでしょう。

10.3 標準化された係数

さて，話は戻って大学の評定平均値 y を，高校の内申点 x_1 と入試成績 x_2 で予測する回帰式をもう一度確認したいと思います。

$$\hat{y} = 3.761 + 0.604 x_1 + 0.003 x_2$$

これを見ると，高校の内申点の回帰係数が 0.604，入試の成績の回帰係数が 0.003 です。この 2 つの数字を見ると，入試の成績がずいぶん小さいですね。高校の内申点は入試の成績の 200 倍も重要なのでしょうか？ いえ，そうではありません。今回の場合，従属変数は最小値が 5.340，最大値が 8.385 です。説明変数はこの従属変数に合うように伸ばしたり (係数をかける) ずらしたり (切片を足す) するわけです。内申点は最小値が 2.13，最大値が 3.90 ですから 5–8 ぐらいの幅に合わせるためには等倍ぐらいの伸ばし方でいいのですが，入試の成績は最小値が 460，最大値は 690 にも及びます。これを 5–8 ぐらいの，一桁の数字にするためにはまず 1/100 ぐらいにスケールダウンしないとどうしようもないですね。ということで，x_2 の係数は小数点以下の小さな数字になっているのです。

このように，ロウデータ[*4]を使った表現では，単位が違うということになりますから，そのまま偏回帰係数を比較して，こちらの影響力が大きかった，という判断をできません。ではどうすればよいのでしょうか。単位が異なるから相対比較できない，ということだったので，単位がなくなれば相対比較できるようになるわけです。そう，このような操作を**標準化**というのでしたね。

説明変数も従属変数も単位がついていますから，これらを整えるためにそれぞれの変数を標準化します。標準化したデータセットは表 10.3 のようになります。

この標準化されたデータセットは，各変数平均が 0 分散が 1 となっており，このままで相対比較が可能です。これを使って回帰分析を行うと，結果は次のようになります。

$$\hat{y.z_i} = 0.4564 x.z_{1i} + 0.2674 x.z_{2i}$$

[*4] raw data，生のデータという意味です。ここでは取ったデータそのものという意味で使っています。

表 10.3 標準化したデータ (一部)

ID	内申点 z	入試 z	評定 z
A	-1.1382198	-1.4131716	-2.3874746
B	-0.6693508	-0.8139469	0.8237724
C	-0.9280371	-1.2184236	-1.2153084
D	1.6749939	0.9837272	0.6406214
N	0.3492265	0.2197157	0.2498993
⋮	⋮	⋮	⋮
O	-0.5076719	-0.5442958	0.3720000

標準化された内申点の回帰係数は 0.4564, 標準化された入試点数の回帰係数は 0.2674 でした。これを比較して, 内申点のほうが入試よりも影響力が大きい, と適切に判断することが可能です。このように各変数を標準化して算出した偏回帰係数のことを**標準化偏回帰係数 (standardized partial regression coefficient)** といいます。長い名称ですが, 意味を考えると理解できると思います。標準化偏回帰係数は, x_2 の得点が同じとすると x_1 が 1SD 上昇するときに従属変数がどれぐらい変化するか, を表している係数になります。解釈はちょっとストレートではないですね。標準化しない回帰係数は単位がついていたので理解はしやすいが比較が難しい, 標準化した回帰係数は単位がないので理解はしにくいけど比較がしやすい, ということです。

さて標準化したスコアで回帰分析した時の特徴をあと 2 つほど指摘しておきましょう。まずは切片 b_0 の値がないことです。気づきましたでしょうか。切片項は変数全体の高さ調整に使われるものであり, 各変数の平均値を加味して計算されるものだったわけですが, 標準化したデータを使った場合すべての変数の平均が 0 になっていますから, この調整が必要なくなったわけです。ということで標準化した回帰分析のばあいは切片がないのも当然なのです。もうひとつ, モデル適合について考えてみます。実は, 標準化しない時の R^2 値と標準化した時の R^2 値は同じです。データが標準化されていようがされていまいが, 線形モデルの当てはめをする上で変数間関係 (相関関係) が変わるわけではありません。位置とか幅が違うだけであり, 散布図の見た目が変わるわけではないのです。考えてみれば当たり前のことなのですが, 改めて確認しておくといいでしょう。

10.4 今日のまとめ

- 説明変数が複数になった回帰分析を重回帰分析という。そのときの回帰係数は偏回帰係数という。
- 偏回帰係数は「他の変数を統制したときの」回帰係数である, ということに注意。部分相関係数, 偏相関係数, 偏回帰係数の順に理解していこう。
- 説明変数が複数ある場合は単位がそれぞれ異なるので, 影響力同士を比較するには標準化偏回帰係数をつかう。

10.5 課題

大学周辺のワンルームマンションの相場を調べるため, 8 件ほどの物件のデータを集めてみました (表 10.4, 図 10.9)。これをもとに, 家賃を従属変数とし, 広さ, 築年数, 最寄りの交通機関からの移動時間を独立

変数とした重回帰分析を行いました。

表 10.4　家賃のデータ

平米	築年数	駅から徒歩	家賃
41.40	14	10	8.60
20.28	16	9	6.10
18.20	28	10	4.20
19.87	16	13	5.70
20.28	19	7	6.30
23.18	20	14	5.40
19.87	17	12	6.25
19.87	20	17	5.80

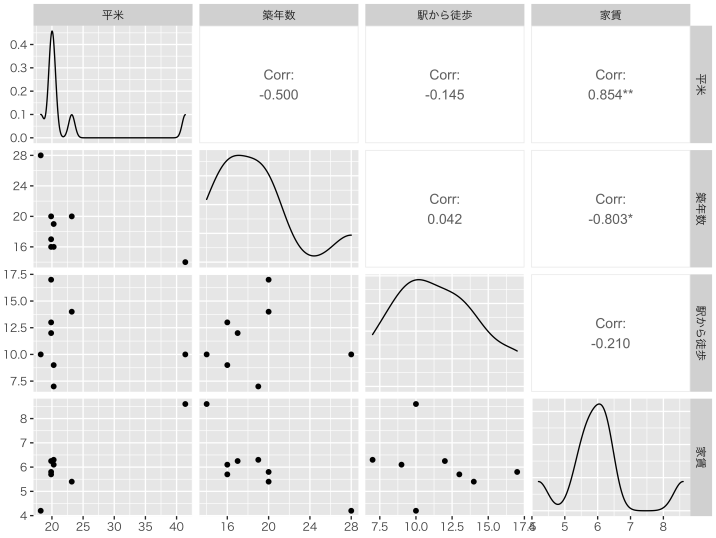


図 10.9　サンプルデータのプロット

ロウデータの重回帰式は，

$$\text{予測値} \quad = \quad 7.05 \quad + \quad 0.09\text{平米} \quad - \quad 0.14\text{築年数} \quad - \quad 0.04\text{移動時間}$$

となり，すべてのデータを変数ごとに標準化したデータの重回帰式は

$$\text{予測値} \quad = \quad 0.58\text{平米} \quad - \quad 0.51\text{築年数} \quad - \quad 0.10\text{移動時間}$$

となりました。また，適合度指標は $R^2 = 0.93$ になりました。

このデータや結果を踏まえて，次の設問に答えてください。

■重回帰分析の結果を読み取る 1　ロウデータの回帰分析とその結果について，正しい表現をしているものすべてにチェックを入れてください。

- 切片は 7.05 である
- 移動時間が同じ物件であれば，築年数が 1 年長いと家賃が 1400 円低い

- 家賃を決めるのに最も重要な変数は，回帰分析からはわからない
- 家賃を決めるのに最も重要な変数は，係数の絶対値が一番大きい築年数である。
- 家賃を決めるのに最も重要な変数は，係数が正の平米である。
- 移動時間と築年数が同じ物件であれば，1 平米広くなると家賃が 900 円高い
- 築年数が 1 年増えると家賃が 1400 円下がる
- 従属変数である家賃と平米，あるいは家賃と築年数の相関係数が高いため，多重共線性が生じている可能性を疑うべきである。

■重回帰分析の結果を読み取る 2　標準化されたデータの回帰分析とその結果について，正しい表現をしているものすべてにチェックを入れてください。

- 移動時間と築年数が同じ物件であれば，1 平米広くなると家賃が 5800 円高い。
- 家賃を決めるのに最も重要な変数は，回帰分析からはわからない。
- 家賃を決めるのに最も重要な変数は，係数の絶対値が一番大きい平米である。
- 切片は記載されていないので，分析結果が解釈できない。
- 家賃を決めるに最も重要な変数は，係数が正の平米である。
- 移動時間が同じ物件であれば，築年数が 1 年長いと家賃が 5600 円低い。
- 標準化されたデータの散布図が描かれていないので，分析に進むのは良くない。
- 説明変数が平米だけの回帰分析より，平米と築年数の 2 変数を使った回帰分析の方が，また平米と築年数の回帰分析より，平米と築年数と駅からの時間の 3 変数を使った回帰分析の方が，適合度 R 二乗は大きくなる

第 11 章

R をつかった回帰分析

さてここまでは，回帰分析や**重回帰分析 (Multiple Regression Analysis)** の仕組みを言葉で説明してきましたが，自分で計算できなければ絵に描いた餅です。重回帰分析の回帰係数を求める方法についてはとくに説明してきませんでしたが，R で計算する時は単回帰・重回帰の区別なく表現できるので，ここで一気に習得してしまいましょう。

11.1 R の環境再確認

第 3,4 章で R の使い方について解説してきました。思い出しながら進めていきましょう。

11.1.1 環境の準備 (確認)

まずは環境の準備です。既に R や RStudio のインストールは終わっているものとして話を進めます[*1]。
次の 3 つのステップをたどって，実行の準備をしてください。

■RStudio の起動 RStudio を起動してください。パッケージのインストールをするときなど，管理人権限が必要になるケースがありますので，Windows ユーザは管理人として実行するようにしたほうが良いかもしれません。

■プロジェクトを開く プロジェクトは前回作成したもので結構です。前回と同じフォルダの中で，新しくスクリプトファイルを作っていくことにしましょう。メニューからプロジェクトを開くを選び (File→Open Project)，目的のフォルダのにある `.Rproj` ファイルを選択してください。それがプロジェクト情報を持っているファイルで，それを選択することでワーキングディレクトリが当該フォルダにセットされます。プロジェクトが開いたかどうかは RStudio の右上やメニューバーなどで，確認できます。

■R スクリプトを開く 今日のコードを書くための R スクリプトファイルを準備します。File > NewFile > R Script と進み，何も書いてない R スクリプトの画面を表示させてください。真っ白いファイルが開いたら問題ありません。Rmd ファイルで書き進めてもらっても結構です。
これでスクリプトのところにコードを書いていく準備ができました。
統計に関するデータは，手入力するのが嫌になる程度にはたくさんのデータがあるはずです。今回もデータをファイルから読み込んで使うことにします。ファイルから読み込むときに注意が必要なのが，テキストエン

[*1] まだだよー！という人は第 3 章に戻って追いついてきてください。ネットで RStudio インストールといったキーワードで検索すると，親切にも案内してくれているサイトに色々出会えるでしょう。

コードです。ASCII データを読み込みますが，エンコードが違うと文字化けが生じてしまいます[*2]。OS が Windows で世界非標準のエンコード Shift-JIS(CP-932) を使っている人は，世界標準の UTF-8 を指定して読み込む必要があるかもしれません。

今回，第 4 章の課題と同じ野球選手のデータを読み込んでみましょう。次の 3 行を実行してください。

code : 11.1 環境の初期化とパッケージ，ファイルの読み込み

```
1  rm(list=ls())
2  library(tidyverse)
3  dat <- read_csv("baseballDecade.csv", na="NA",
4          locale=locale(encoding="utf8")) %>%
5          filter(Year=="2020年度")
6  summary(dat)
```

■コード解説

1 行目 環境 Environment を初期化するコードです。Rmd 形式で書いている人はこの行は書かないようにしてください[*3][*4]。

2 行目 tidyverse パッケージを読み込んでいます。このパッケージはファイルの読み込みや並べ替え操作など，便利な機能がまとめられたものです。

3-4 行目 baseballDecade.csv というファイルを読み込んでいます。このファイルがプロジェクトフォルダ内にあることを確認してから実行してください。

5 行目 ファイルの一部，2020 年度のデータだけに絞るようフィルターをかけています。

6 行目 読み込んだデータの記述統計量など，基本的な要約を行っています。

11.1.2 散布図の描画

データが手に入ったとき，最初にすることは可視化です。**データは図にする**という約束は忘れないでくださいね。今回はさまざまな変数があるのですが，ここで打者の身長と体重のデータだけ抜き出して図にしてみたいと思います。

code : 11.2 環境の初期化とパッケージ，ファイルの読み込み

```
1  batter <- dat %>%
2    dplyr::filter(position != "投手") %>%
3    dplyr::select(height, weight)
4  g <- ggplot(batter, aes(x = height, y = weight))
5  g <-  g + geom_point()
6  g
```

■コード解説

1-3 行目 batter というオブジェクトを作っています。これは dat オブジェクトを加工して作るのですが，詳しくはこの後で説明します。

[*2] なんのことかわからない，という人は付録 A を参照してください。

[*3] Rmd ファイルはそのまま使うのではなく，出力ファイル形式に knit して使います。その際必ず環境情報は初期化されるので書く必要がなく，むしろ書いてあると knit するファイルの情報なども消してしまうので knit できなくなるのです。

[*4] Rmd で保存するときの拡張子は .Rmd になります。ファイル名にピリオドが入っていると拡張子が変わっちゃうことがあるので，注意してください。拡張子って何，というひとは A.5,Pp.396 を参照。

4 行目 batter オブジェクトを使って，x 軸を height 変数，y 軸を weight 変数をするキャンパスを指定します。

5 行目 キャンパスに点を打つように geom_point します。

6 行目 描画オブジェクト g を表示します。

結果的に図 11.1 のような図ができていれば OK です。

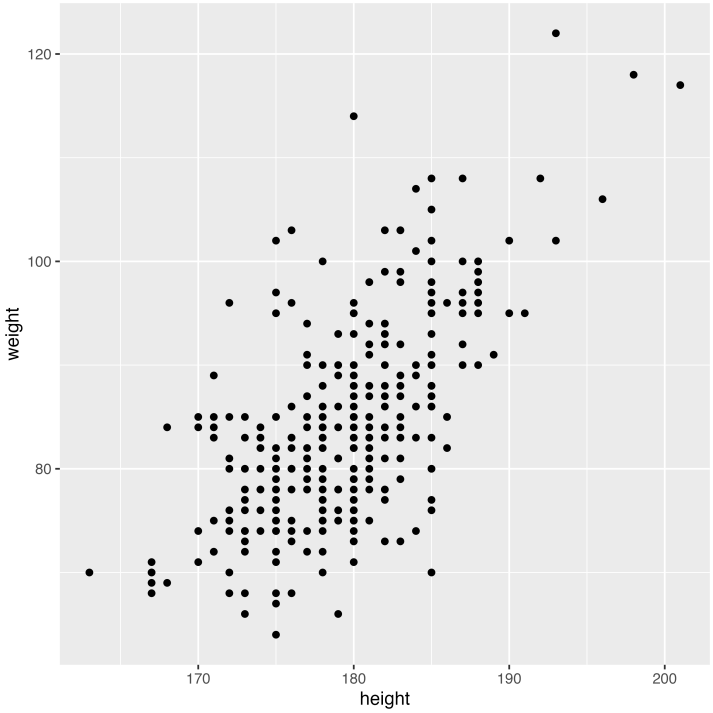


図 11.1 身長と体重のデータ

さて 1-3 行目の操作ですが，ここでやりたかったことは dat オブジェクトの中から打者のデータに限って，身長と体重の変数を抜き出すことでした。これらの操作は tidyverse パッケージに含まれる dplyr パッケージの filter 関数と select 関数で実装されます。この両者の違いについては，セクション 4.2.2,Pp.65 を参照してください。ともかくこれで欲しいデータの散布図が書け，どうやら正の相関がありそうだということがわかりました。直線関係を仮定しても良さそうです。ちなみに次のコードを書くと ggplot で直線を書くことができますよ。

code : 11.3 回帰線を引く

```
1  g <- g + geom_smooth(method = "lm", se = FALSE)
2  g
```

■コード解説

1 行目 散布図を書いているオブジェクト g にスムージング関数 geom_smooth でデータにスムーズな線を引く。線の引き方は lm すなわち**線形モデル**。

2 行目 オブジェクト g の描画

11.2 線形回帰の実行

それではここから回帰分析を行います。回帰分析は線形関数を当てはめるので**線形モデル (linerar model)** と呼ばれますが, これの頭文字をとった lm 関数が線形回帰モデルです。次の code:11.4 を実行してみましょう。

code : 11.4 回帰分析の実行

```
1       result <- lm(weight ~ height, data= batter)
2       summary(result)
```

■コード解説

1 行目 lm 関数に回帰式とデータを入れて分析を実行し, その結果を result オブジェクトに代入する。ちなみに~はチルダといいます[*5]。

2 行目 分析結果の要約を出力させる。

ここで重要なのは, 1 行目の回帰式の表現です。R では $y \sim x$ という書き方で, $y = f(x)$ を表現します。これは回帰式に限らず, 一般的に従属変数 y と独立変数 x の関係を $\sim$ でつなげる, という**式 (formula)** の書き方なのです。今回は lm という関数を使って表現しましたが, 線形モデルではない関数で独立変数, 従属変数の関係を書く時も y ~ x で表現します。この表記に慣れると, あの分析もこの分析も, 形としては $y = f(x)$ なんだな, ということがわかりやすくていいですね。今回は, 体重を身長で予測したということになっています。

この分析の出力結果が出力 11.1 の通りです。

R の出力 11.1: 回帰分析の出力結果

```
Call:
lm(formula = weight ~ height, data = batter)

Residuals:
     Min       1Q   Median       3Q      Max
-20.9217  -4.9213  -0.4214   4.0783  28.9118

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -124.91755   13.89317  -8.991   <2e-16 ***
height         1.16670    0.07746  15.063   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.435 on 327 degrees of freedom
Multiple R-squared:  0.4096, Adjusted R-squared:  0.4078
F-statistic: 226.9 on 1 and 327 DF,  p-value: < 2.2e-16
```

[*5] 日本語キー配列の場合,「ほ」の右, アットマークの上にある「へ」のキーを Shift を押しながら押すと出ます。US キー配列の場合は tab キーの上にあります。

何やら色々出ていますが，順に確認していきましょう。まず Residuals とありますが，ここは残差 e_i の記述統計情報を表しています。どれぐらい残差が大きく出るのか，確認すると良いでしょう。次の Coefficientes は係数という意味です。ここに回帰係数が出ています。少しみにくいかもしれませんが，表形式になっていて，Estimate と書かれたところが推定値，今回知りたい回帰係数の値になっています。回帰式は次のようになります。

$$\hat{y} = -124.91755 + 1.16670 \times height$$

どの数字が結果の何に対応しているか確認しておいてください。また，この式が得られたということは身長のところに新しい数字を入れると体重の予測ができるということもわかってもらえると思います。

その他にもいろいろな情報が出ていますが，その内容は後期でお話しする推測統計に関するところですので，今のところはまだ知らなくて OK です[*6]。それよりも大事なのは下から二行目にある，Multiple R-squared というところです。これがいわゆる R^2 値であり，モデルの**適合度**を表しています。今回は $R^2 = 0.4096$ ですから，予測値 $\hat{y}$ と従属変数 y の相関係数が 0.62 ぐらい，分散説明率が 40.9% ぐらいだったということがわかります。悪い当てはまとしては，やや悪いといったところでしょうか。

11.3 残差からわかること

今回の lm 関数から返される結果を格納したオブジェクトの中には，他にも予測値 $\hat{y}_i$ や残差 e_i といった情報が含まれています。それらが架空の値ではなく，算出されたここの値であるということを確認するために，表示させてみましょう。

code : 11.5 回帰分析のその他の情報

```
1 batter <- batter %>%
2     dplyr::mutate(residuals = result$residuals,
3                   yhat = result$fitted.values)
4 batter
5 summary(batter)
6 plot(batter)
7 cor(batter)
```

■コード解説

1-3 行目 batter オブジェクトに batter オブジェクトを加工したものを上書きしていきます。加工内容は，dplyr パッケージの mutate 関数を使って変数を追加することです。追加する変数名は residuals と yhat で，residuals には result オブジェクトの residuals 変数を，yhat には result オブジェクトの fitted.values 変数をあてがっています。

[*6] それでも少し解説しておきます。まず推定値というのは今回のデータから得られる値ではなく，このデータが大きな母集団から撮られた標本だと考えたときに，標本のデータから母集団の関係を推測するならこの値になる，という意味があるからです。今回のデータについての最小二乗法での答えが欲しいのですから，標本から推定するのではなく，標本の特徴を考えてればいいじゃない，と思うかもしれません。いやほんとその通りなので，今はまだ説明しないのです。幸い，最小二乗法で求めた回帰係数の値と，推定法 (ここでは最尤法という方法を使っていますが) による答えが一致しますから，悩まなくてもいいのですが。また，今回の標本ではなく別の標本だった場合，この推定値がどれぐらい違う値になる可能性があったかを計算して出しています。それが Std.Error のところです。それを踏まえて考えると，回帰係数が 0，すなわち説明変数がまったく予測しないという状態になる可能性がどれぐらいあるか，というのを検証しているのが t value や Pr(>|t|) のところで，これをみながら今回の分析から何が言えるかを結論づけたりするわけです。

4 行目 上書きされた batter オブジェクトを表示させています。長くなりますので，上の方の一部分だけが表示されます。

5 行目 上書きされた batter オブジェクトの要約統計量を表示させています。

6 行目 batter オブジェクトのプロットを描いています。

7 行目 batter オブジェクトの相関係数を計算しています。

回帰分析の結果を代入したオブジェクト result には，residuals 変数として残差が含まれています。同じく fitted.values 変数として予測値が含まれています。4 行目の出力結果の一部を見てみましょう (出力 11.2)。

R の出力 11.2: 回帰分析の出力結果

```
# A tibble: 329 × 4
   height weight residuals  yhat
    <dbl>  <dbl>     <dbl> <dbl>
 1    171     72   -2.59    74.6
 2    181     98   11.7     86.3
 3    177     91    9.41    81.6
 4    180     85   -0.0882  85.1
 5    171     84    9.41    74.6
```

これを見ると，たとえば 1 行目のデータは身長 171cm 体重 72kg の選手であり，回帰式 $-124.91755 + 1.16670 \times height$ を使った計算結果は 74.6 ということがわかります。74.6 という予測値 $\hat{y_1}$ は，実際の値 y_1 と比べて 2.59 ずれていますので，$\hat{y_1} - y_1$ で-2.59 となっています。以下同様で，329 人分のデータそれぞれについて予測値と残差が計算できました。

またプロットを見てみましょう。これは ggplot を使っての出力ではなく R オリジナルの出力例ですが (そのせいか少し色気のない出力ですが)，これを見て幾つか気づく点があると思います。1 つは height と yhat の組み合わせが直線になっているところですね。次の cor 関数でも，対応するところが 1.0 になっています。これは当然で，身長データを伸ばしてずらしたのが予測値であり，この両者はいわば一対一対応，100% の相関関係にあることがわかります。また，説明変数 height と残差 residuals の組み合わせは丸に近く，相関係数はほぼ 0.0 です[*7]。説明変数と残差の間に相関関係があるというのは，原理的におかしなことですよね。説明変数は従属変数をできるだけ説明しようとしていて，その残りが残差なのですから，残差のところに説明変数と関係する要素があるとすれば，それは残差ではなく従属変数のまだ説明できる箇所だということになってしまうからです。

同様に，残差 residuals と予測値 yhat の相関関係もほぼ 0.0 になっています。これも原理的には 0.0 になるのです。回帰分析は説明変数によって，説明できる部分 (予測値) と説明できない部分 (残差) にきっちり分離するのですから，分離されたもの同士に関係が残っているはずがないのです。数理的な導出は小杉 (2018) の P.127–135 を見て貰えばわかりますが，データを見ることで一目瞭然だと思います。

[*7] いやいや，何か数字があるよとツッコミが入りそうですね。計算結果のところには $5.451636e-16$ という表示があるかと思います。この表現は，桁数があまりにも大きいので普通の十進法で表現できなかった場合に出てくるもので，$e+$ のところが 10 の累乗を表しています。今回は $e-16$ ですから，$10^{-16} = 0.0000000000000001$ というごく小さな値を，5.451636 にかけた値が答えとなっています。数学的にはちょうど 0.0 になることが証明できるのですが，機械の計算は丸め誤差などがあるので少しずれが生じます。が実質的にゼロだということです。

11.4 重回帰分析

では続いて**重回帰分析**に行きましょう。R において重回帰分析を行う時は，説明項を 1 つ追加するだけでいいので，作業としては簡単なものです。ただし結果の解釈には注意が必要であることを忘れずに。

code：11.6 重回帰分析のデータを作って可視化

```
1  batter2 <- dat %>%
2      dplyr::filter(position != "投手") %>%
3      dplyr::select(HR, Hit, salary) %>%
4      na.omit()
5  g1 <- ggplot(batter2, aes(x = HR, y = salary)) +
6      geom_point()
7  g1
8  g2 <- ggplot(batter2, aes(x = Hit, y = salary)) +
9      geom_point()
10 g2
11 cor(batter2)
```

■コード解説

1-4 行目 batter2 というオブジェクトを作っています。これは dat オブジェクトを加工して作っており，2 行目で野手 (投手でない人) を選択し，3 行目で変数として本塁打，安打，salary を選択しています。4 行目は欠損値を一括で取り除く処理をしています。ホームランを打っていない人などはデータに含められないからです。

5-7 行目 x 軸に本塁打の数，y 軸に年俸 (salary) をとった散布図を書いています。

8-10 行目 x 軸に安打の数，y 軸に年俸 (salary) をとった散布図を書いています。

11 行目 データの相関関係を数字でみています。

プロットを見ると，ホームランの数や安打の数が増えるほど，年俸は上がっているようです (図 11.2)。さすがプロ野球は実力勝負の世界，グランドには銭が埋まっているというやつです。少しデータの左端の方，つまりホームランや安打を打たない人が数多くいるため，よく打つ人に引っ張られて相関係数が大きく出ている可能性はありますし，安打と本塁打の間にも相関があるので少し気になりますが，ひとまずはこれで分析することにしましょう。

回帰分析のコード code:11.7 と出力 11.3) は次の通りです。

code：11.7 重回帰分析の実行

```
1  result2 <- lm(salary ~ HR + Hit, data = batter2)
2  summary(result2)
```

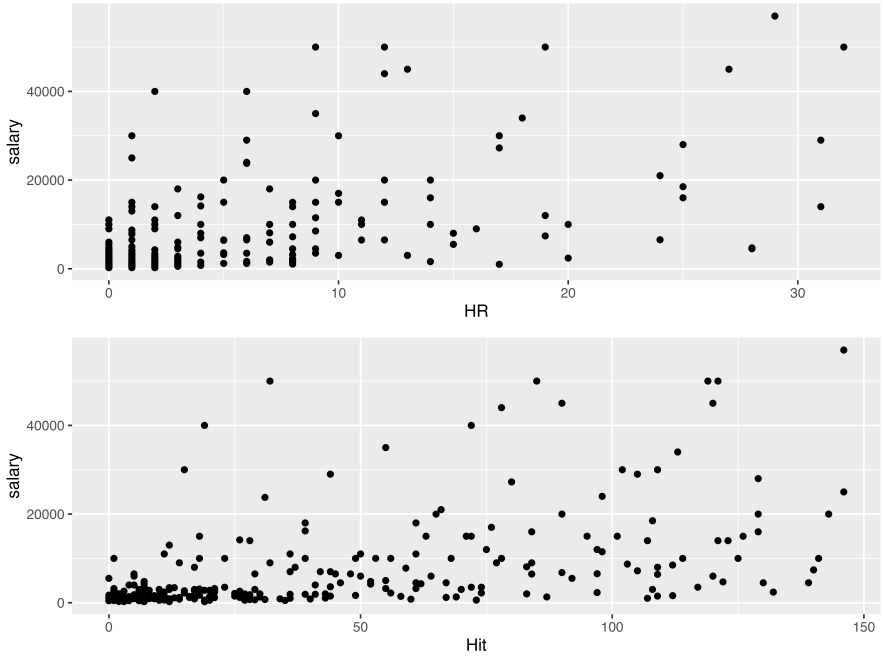


図 11.2 本塁打, 安打と年俸の関係

R の出力 11.3: 回帰分析の出力結果

```
Call:
lm(formula = salary ~ HR + Hit, data = batter2)

Residuals:
   Min     1Q Median     3Q    Max
-22581  -2394  -1025    264  40723

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1620.46     575.40   2.816  0.00516 **
HR            659.15     104.09   6.333 7.98e-10 ***
Hit            53.88      16.61   3.245  0.00130 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7761 on 326 degrees of freedom
Multiple R-squared:  0.378, Adjusted R-squared:  0.3742
F-statistic: 99.06 on 2 and 326 DF,  p-value: < 2.2e-16
```

結果の読み取りはよろしいでしょうか。重回帰式は次のようになります。係数をどこから読み取るか, 確認してくださいね。

$$\hat{y} = 1620.46 + 659.15 \times HR + 53.88 \times Hit$$

これをみると, 安打の数が同じくらいであれば, ホームラン一本につき 660 万円年俸が上昇するようです。また本塁打の数が同じぐらいであれば, ヒット一本で 53 万円年俸が上昇します。またこのモデルの R^2 は

0.378 ですから，予測モデルは全体の 37% ぐらいしか説明できておらず，あまり良いモデルとは言えないということがわかりました。

ヒットも本塁打も単位は「本」ですので，解釈も同じようにできて簡単かもしれません。あるいはヒットと本塁打は意味が違う，というのであれば直接の比較はおかしいことになります。ここでは後者の可能性を考えて，変数同士の単位が違うから標準化しよう，という方向で分析を進めたいと思います。

データセットの各変数を標準化する必要がありますが，それには scale 関数を使います。

code : 11.8　標準化したデータで重回帰分析を実行

```
1  batter2z <- scale(batter2) %>% as.data.frame()
2  summary(batter2z)
3  result2z <- lm(salary ~ HR + Hit, data = batter2z)
4  summary(result2z)
```

■コード解説

1 行目　batter2z というオブジェクトを作っています。これは batter2 オブジェクトを scale 関数で標準化することで作れるのですが，scale 関数は出来上がったものがデータを表す data.frame 型になっていませんので，%>% でつないで as.data.frame という関数でデータの形にしています。

2 行目　出来上がった batter2z というオブジェクトの記述統計量をみています。平均値が 0.0 になっていることなどから，標準化されていることが確認できます。

3 行目　batter2z オブジェクトで同じように重回帰分析をしています。

11 行目　結果を出力させています。

R の出力 11.4: 回帰分析の出力結果

```
Call:
lm(formula = salary ~ HR + Hit, data = batter2z)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3016 -0.2440 -0.1045  0.0269  4.1507

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.006e-16  4.361e-02   0.000   1.0000
HR           4.298e-01  6.786e-02   6.333 7.98e-10 ***
Hit          2.202e-01  6.786e-02   3.245   0.0013 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7911 on 326 degrees of freedom
Multiple R-squared:  0.378, Adjusted R-squared:  0.3742
F-statistic: 99.06 on 2 and 326 DF,  p-value: < 2.2e-16
```

結果を見てみましょう (出力 11.4))。推定値の読み方，わかるでしょうか。ここにも e-01 なんて表示があるのですが，これは $10^{-1} = \dfrac{1}{10} = 0.1$ と同じ意味です。標準化した回帰分析の結果は次の通りです。

$$\hat{y} = 0.4298HR + 0.2202Hit$$

これを見比べると，ホームランよりもヒットのほうが年俸の上昇には強い影響を与えているようですね。ここでの回帰式では切片を記入しませんでした。推定値は $1.006e-16$ であり，これは $e-16=10^{-16}$ がかけられていますからほぼゼロになっています。原理的にも 0 であるはずなので，ここでは書きませんでした。また R^2 が標準化しなかったときと変わっていないことも確認できますね。相対的な関係には影響しないので，このようになります。

計算自体は一瞬でできることがわかりましたね。でも計算ができた時に，何をやっていたのかわからないとか，何のためにやっていたのかわからない，ということでは意味がありません。またもちろん，結果の読み取り方が間違っていてもいけませんし，この結果から何が言えるだろうか，と考えることで初めて研究が始まるのです。「やった，できた」だけではなく，「なぜ，どうやって，どういう意味か」と考えを深めるためのツールとして，統計技術を使い倒してみてください。

11.5 課題

課題は Rmd ファイルで提出していただきます。提出するファイル名には学籍番号を含めてください。

R チャンクの中に，必要な計算式を記入してください。R チャンクの外に，結果の読み取りや解釈を，言葉で記述してください。

11.5.1 課題

次のデータ (表 11.1) は，向ヶ丘駅周辺の家賃について調べたもので，部屋の広さ (`heibei`，単位 m^2) と，築年数 (`chiku`)，駅から徒歩何分かかるか (`ekitoho`)，家賃 (`yachin`) からなるものです[*8]。なおこのデータは `yachin.csv` として配布しています。

表 11.1 家賃相場

heibei	chiku	ekitoho	yachin
41.40	14	10	8.60
20.28	16	9	6.10
18.20	28	10	4.20
19.87	16	13	5.70
20.28	19	7	6.30
23.18	20	14	5.40
19.87	17	12	6.25
19.87	20	17	5.80

■課題 家賃を従属変数，平米を説明変数とした回帰分析を行い，回帰式を求めてください。

■課題 15 平米で家賃 6 万円の物件を新たに見つけました。この物件はこのデータに照らし合わせて，お得な物件だと言えるでしょうか。データに基づいて判断してください。

■課題 家賃を従属変数，平米と築年数を説明変数とした重回帰分析を行い，回帰式を求めてください。

[*8] 2019 年 5 月に実際に賃貸住宅サイトを利用して調べたものです。

■課題 家賃を従属変数，平米と築年数と駅からの距離を説明変数とした重回帰分析を行い，回帰式を求めてください。

■課題 ある従属変数に対して，説明変数 A だけの単回帰分析をした時 (これを model1 とします)，説明変数 A と B を用いた重回帰分析を行なった時 (これを model2 とします)，説明変数 A,B,C を用いた重回帰分析を行なった時 (model3) をそれぞれ比べると，モデルの適合度が一番良いのはどのモデルになるでしょうか。またそれはなぜそうなるのか，考察してください。

第 12 章

実験計画法 1；実験計画法とは

さてここまで線形モデルによって変数間の関係を記述し，予測できるというお話をしてきました。相関係数や回帰分析による影響力で判断する方法でした。心理学の研究においてこうした相関関係をもとに考察を進める事は少なくありませんが，もう 1 つ，実験的なアプローチによってより因果的な効果を検証する方法がとられることも少なくありません。

ここからは心理学における実験と，その分析方法についての話に進んでいきます。

12.1 ランダム化

心理学は人間を対象にした学問であり，研究の方法はさまざまなものがあります。一般的には，調査，観察，実験，面接などの方法がとられます。ここでいう実験とは，皆さんが小中学校の理科の時間の時に習ったような実験と，同じようなものだと思っていただいて構いません。ただしその対象が人間になっているというところに，注意が必要なだけです。その注意というのは，もちろん倫理的な問題も多く含みますが，それに加えて，物理学の実験と違って研究の対象が意思を持っているというところです。そうした研究対象特有の問題に考慮しつつ，実験とはどのように進めていくべきかを考えてみましょう。

簡単なところから，小学校の理科でやるような実験として，次のような例を考えてみましょう。植物の種をまいて，植物の成長にはどういった要因が必要なのかを考える授業をするとします。ここでいろいろな条件を考えて実験を計画しますね。たとえば日向と日陰を比べればどちらが成長が早いのかとか，水をあげた場合とあげない場合とではどちらが成長が早いのかとか，どういった肥料をどれぐらいあげれば成長が速くなるのかといったことを検証したいとします。

たとえば日光の効果を検証したいのであれば，半分は日向に残りの半分は日陰に種をまいて，どちらの方が成長が良いかを記録するという方法で実験することになると思います。当然のことながらこの時重要になってくるのは，日向か日陰かという条件以外の要素を変更しないことです。たとえば日向にまいた種だけに水を与えるとか肥料与えるといったことをすると，最終的に得られた成長の違いが日光によるものなのか，それ以外の要因によるものなのか区別できません。これでも研究の目的が果たせないわけです。実験をする場合は，検証したい条件の違い以外は均一にしておくというのは絶対条件です[*1]。

何を当たり前のことをと思われるかもしれませんが，人間の場合はさまざまな要因が絡み合っているので，何か 1 つの効果を見るために他のものを統制しようとしても，統制すべき要因が無数に考えられてしまうので，おのずと限界があるわけです。

人間を研究の対象にする場合の問題は他にもあります。まずこの授業でも冒頭のほうにお話ししたと思い

[*1] このように介入が入り混じっていて効果が特定できなくなることを，要因が**交絡 (confounding)** するというのでした。これについては，セクション 8.4,Pp.138 参照。

ますが，何を結果変数にするかというところも明確ではありません。人の気持ちのようなものが従属変数になっていたとして，その気持ちをどのように測定するのかという測定方法にもさまざまな注意すべき点があるでしょう。測定するべき対象が明確になっていたとしても，被験者に負担をかけすぎないかどうかは倫理的に検証されるべきです。また，研究対象が人間ですから，人間の回答は嘘をつかれることもあれば，間違うこともあれば，実験者の意図を汲んで自然な状態でない反応することもあります。こうした問題に対しては，**二重盲検法 (double blind experiment)** を始めさまざまな方法で問題を回避しなければならないといった事は，心理学研究法の方で話を聞いていることと思います。

先に述べた測定の問題と，後半で述べた倫理的な問題を始めとする研究法の問題は横に置いたとしても，実験にはコントロールしてもコントロールしきれないものが含まれる事はお分かりいただけると思います。このコントロールしきれない部分のことを**偶然誤差**と言いますが，偶然誤差に振り回されるだけではなくこれを積極的に利用することによって，実験の質を高めるために利用できます。偶然をハンドリングするのです！

人間を対象にした実験の例を考えてみましょう。たとえば新しく商品開発された 3 つのミネラルウォーターがあったとして，いろんな人に試し飲みしてもらって採点してもらったとします。このドリンク A,B,C について評価してもらうとして，その場合に気にしないといけないことがあります。1 つはどの順番で飲んでもらうかということです。たとえばドリンク A に甘い味がついていたとして，その後で B,C を飲むと味が薄いと評価されるかもしれません。あるいは 1 種類目 2 種類目 3 種類目と順番に飲んでいたとして，最後のほうはもうお腹がタポタポで全然味の評価ができないということがあるかもしれません。このようにどういう順番で飲むかとか，どの並びで飲むかといったことも評価に影響してきます。こうなってくるとこれはもう統制できませんので，諦めるのではなくて，逆に偶然の力を借りるのです。つまり，サイコロを振って 1 か 2 が出れば A，3 か 4 が出れば B をまず飲む，といったようにし，人によって飲む順番も並びもバラバラになってしまえば，飲む順番や並びの効果があったとしても，相互に影響を殺しあって，平均をとればそれぞれのミネラルウォーターの評価として使うことができると考えるわけです。

ここでこのようにバラバラにするということを，言い方を変えると**ランダム化 (randomize)** するといいます。英語でランダム random は確率とも訳されます[*2]。確率化すると言い換えても構いません。偶然をハンドリングするために，確率の力を借りようというわけです。

第 7 講で**正規分布**の話をしました。正規分布はさまざまな要因の積み重ねとして，つまり二項分布の極限の形で出てくる分布だということでした。この話をここで思い出してみましょう。人にはさまざまな個性があり，それを統制した上で実験をするなんて事は不可能です。しかし 1 人ではなくて多くの人間を集めてくると，ある変数については正規分布の形に近づいてくるでしょう。たとえば身長や体重といった生理的な指標や，学力や心理的な変数なども正規分布にしたがうと考えられるわけです。このときの正規分布の平均は，集団の代表的な値，真ん中あたりの人ということになるでしょう。偶然誤差の部分は平均がゼロになっていますから[*3]，系統的な特徴が平均のところに現れていると考えられるわけです。

このように，被験者一人一人の特徴はあるとしても，平均を取れば何らかの傾向があるはずです。誰が飲んでも美味しい飲み物とか，その逆ということはあるわけです。平均を取ることで，個性をキャンセルアウトしようということですね。その上で，振り分けにも確率の力をかりてランダム化します。サイコロの目が偶数なら群 A，奇数なら群 B，というように無作為に群分けするのです。無作為にグループを分けていますから，どちらのグループにもさまざまな人が入ることになります。味覚に敏感な人やそうでない人など，今回の実験に関わる個性があったとしても，2 つのグループの平均は，個性がキャンセルアウトされていますから，平均的には同程度の群が 2 つできたことになります。

[*2] 確率的に決まる変数のことを確率変数と言いますが，確率変数は英語で random variables です。

[*3] プラスの要因とマイナスの要因，大きさもさまざまなので，それらが相殺し合う＝キャンセルアウトされる，といいます。

心理学における実験で因果的な効果を考える場合は，無作為[*4]に群分けをし，一方を**統制群 (control group)**，他方を**実験群 (experimental group)** と呼びます。実験群には何らかの操作が加えられます。これに対して統制群には何もしない，あるいは何の効果もない同程度の介入をします[*5]。無作為に群分けしているので実験群のグループに含まれるさまざまな原因の効果は，統制群にも同程度に含まれているはずです。それが課題に影響与えるものであったとしても，これは両群に同程度の効果を持っているはずです。介入の後で両群に何らかの心理学的な課題をさせるとか，反応を見るとかします。その結果のところで平均的な成績の差を比較します。何らかの要因があって，それが課題の成績にプラスに出るかマイナスに出るかといったことがあるかも知れませんが，群全体としてはそれは同程度のはずです。それを超えて群の間に差があるとしたら，その原因は操作をしたかしなかったかにあると言えるではありませんか。このように平均化することで，因果的な効果を考えることができるようなるのです (図 12.1)。この効果のことをとくに**平均因果効果**と言います。

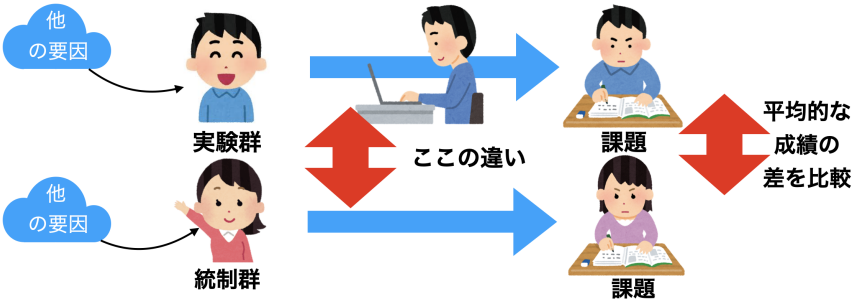


図 12.1 RCT の基本的枠組み

これが実験の基本的な枠組みです。集団の平均的な違いは，無作為に群分けすることで個体差を相互に打ち消し合うための仕掛けです。誤差をコントロールする方法の 1 つと言えるでしょう。この方法を**ランダム化比較実験 (Randomized Controled Ttials; RCT)** といいます。偶然誤差，あるいは個体差が相互に打ち消し合うのは，それらが正規分布に従うという特徴を生かした方法だとも言えるでしょう。

12.2 用語の整理

12.2.1 要因と水準

このセクションでは今後のために実験計画に関わる用語を整理していきます。

実験の例として，心理学的な話ではありませんが，小学校の理科でやったような植物の成長に関する実験をとりあげます。植物の成長には何が必要かを学ぶための実験で，植木鉢を日向に置いたり日陰に置いたり，水分をあげたりあげなかったり，肥料をあげたりあげなかったりします。そのような条件のもとで，1 ヵ月間でその植物が何センチ成長したかを記録することにしました。

この実験の例でいうと，**独立変数** (**説明変数**ともいう) と**従属変数** (**被説明変数**ともいう) は明らかですね。従属変数は，1 ヵ月間で成長した植物の長さということになります。独立変数は，日光があるかどうか (日陰か日向か)，水分があるかどうか (水をあげない，時々あげる，毎日あげる)，肥料をあげるかあげないか，といった 3 つになります。従属変数は**比率尺度水準**で，独立変数は**名義尺度水準**になります。

[*4] 無作為の反対，作為とは意図を持ったという意味です。無作為はその否定で，何の意図も持たずに，という意味です。
[*5] 何もしないことが逆に何らかの意味や効果を持ってしまう可能性があるからです。人間を対象にした研究はややこしい！

このように，実験や研究で検証したいもののことを，**要因 (factor)** といいます。ここでは日光，水分，肥料の 3 つの要因がありますので，「この実験は 3 要因計画である」と言ったりします。

3 つの要因にはそれぞれ比較したい**水準 (level)** が複数あります。日光という要因は，日向/日陰という 2 つの水準がありますので，日光という要因は 2 水準ある，などと表現します。今回の例の場合，水分は 3 水準，肥料は 2 水準です。

要因と**水準**を使ってこの実験を表現するときは，「3 要因の 2× 3 × 2 デザイン」と言ったりします。実験がどのようにデザイン (計画) されているか，ということを説明する用語です。

12.2.2 群内計画と群間計画

さて要因と水準という言葉の違いがわかったところで，実験計画に関わるもう 1 つの区別について説明しましょう。さきほどと同じように植物を測定する実験の例を考えます。ただし今回は要因として 1 つは「日光」，つまり日向なのか日陰なのかの比較をすることを考えたとします。この条件のもとで 1 ヵ月間記録し，日向と日陰で成長がどれぐらい違うのかを，0 日目，15 日目，30 日目に記録して比較することを考えたとします。得られるデータは図 12.2 のようになることがお分かりいただけますでしょうか。

	日光の要因	0日目	15日目	30日目
植木鉢A	ひなた	5	16	27
植木鉢B	ひなた	12	23	34
植木鉢C	ひかげ	1	6.5	12
植木鉢D	ひかげ	10	21	31

図 12.2 得られるデータ

この時に考えられる要因の 1 つはもちろん日光ですが，これは 1 行目 +2 行目と 3 行目 +4 行目の条件を比較すればわかることになります。ただしこの実験はそれ以外にもたくさんの情報を含んでいます。それは 1 つの行が 1 つの植木鉢についての繰り返し獲得できた情報だということに起因します。データを見ると，植木鉢 A とか植木鉢 C はそもそも初日でほとんど芽が出ていませんでした。それに対して植木鉢 B と植木鉢 D は，初日ですでに 10 センチ以上の芽が出ていたことになります。こうした実験開始時の個体差というのは，今回の研究でどのように考えるべきでしょうか。今回の研究では日光がどれぐらい成長を促進するかということに興味があるのであって，そもそもの個体差は興味がなく，何なら実験にとって邪魔な情報ですらあります。

成長の促進の度合いは 0 日目と 15 日目との差分，15 日目と 30 日目との差分でもって表されることであって，スタート地点の情報は必要ありません。同じ答えから反復してデータを得ているわけですから，行方向に数字を見ていくことによって，個体差を取り除いた変化量を考えることができます (図 12.3)。

このように，日向と日陰といったグループ間の違いを見る要因と，個体差を考慮して 1 つ 1 つの個体がどのように変化していったかを見るグループ内の違いを見る要因があります。用語としてそれぞれ**群間要因 (Between Factor)** と**群内要因 (Within Factor)** といいます。今回の例では，日光を群間要因，測定日を群内要因と呼ぶことになります (図 12.4)。

この実験は群内要因と群間要因が組み合わさっていますので，とくに**混合要因計画 (Mixed Design)**

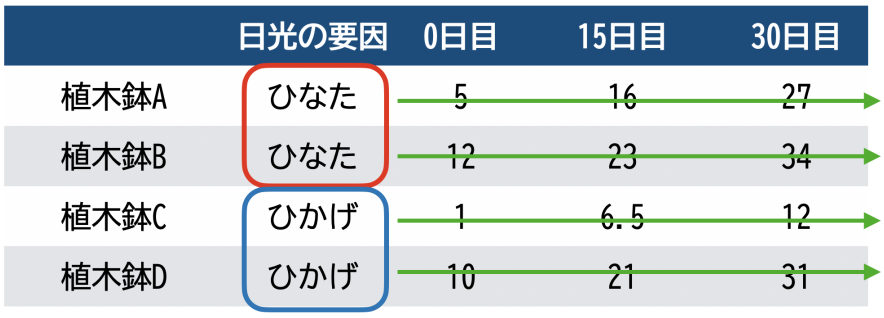


図 12.3 比較できる方向

と呼ばれることもあります。水準数にも配慮して表現すれば，この研究は 2 要因の 2× 3 デザインあるいは「間 2× 内 3 の混合計画」という言い方をします。

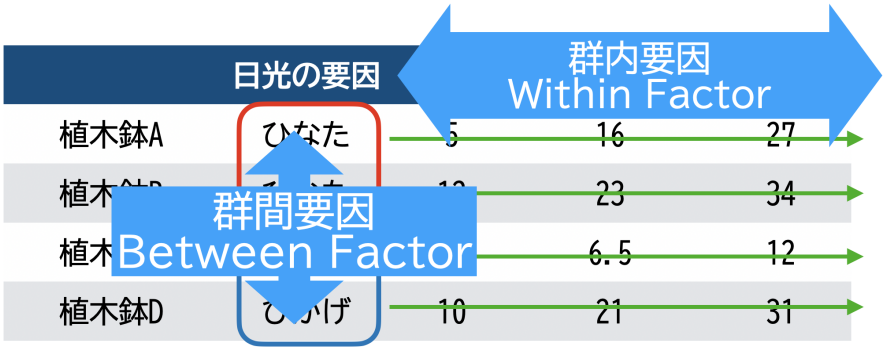


図 12.4 群内要因と群間要因

皆さんが実験を計画するときには，これらの言葉を使ってどういった要因を考え，どういった計画を立てているのかを表現する必要があります。ちなみに群間要因は別名**対応のない**要因とも言います。対して群内要因は，**対応のある**要因，あるいはそのデータの性質から**反復測定 (Repeated Measure)** 要因とも呼ばれます。

後ほどの講義で詳しく説明されますが，群間要因計画と群内要因計画を比べると，群内要因計画の方が個人差を算出できることから，より精度の高い分析になります。群間要因計画は**個人差**と**偶然誤差**を区別できないため，群内要因計画に比べると分析の精度は下がります。じゃあ全部群内要因で実験を計画すればいいじゃないかと思うかもしれませんが，群内要因の別名，「反復測定」という言葉にあるように，人を対象にする実験計画はそのまま実験参加者に対する負担になります。そういったことも考えると，内容によっては群間要因計画にせざるをえないということも少なくありません。

また，人間を研究対象にする場合は，考えるべき要因がたくさんあることも少なくありません。あの要因もこの要因も関係してくるに違いないと考えると，実験計画の要因数はどんどん増えていきます。ただしこの後見ていくように，要因をどんどん増やすと組み合わせを考慮しなければならない数も多くなり，解釈も難しくなりますから，せいぜい 3 要因で，できれば 2 要因までで研究計画をすることをお勧めします。

12.3 実験でなぜ効果がわかるのか

ここまで実験計画の用語の整理をしてきました。因果関係をわかるためには実験をしなければならない。自分で実験の計画を立てて，要因の効果があるかどうかを確かめようというのは心理学でも長らく行われてきた研究のやり方です。ここでそもそも実験でなぜ効果がわかるのかということを考えてみたいと思います。その大枠については，ランダム化のセクションのところでお話ししましたが，ここではもう少し掘り下げて，数字を使ってどこにどのような形で違いが現れ，何を評価すれば良いのかについて考えていきたいと思います。

12.3.1 効果と偶然誤差の切り分け

表 12.1 をみてください。これは一要因 3 水準，Between デザインの実験データです。カバーストーリーとして，植物の成長を見る実験で水を毎日あげる，時々あげる，あげない，という違いを作って比較したものだと思ってください。要因は「水」で，水準数は 3，実験期間終了後に成長した長さを測って従属変数にしたとします。各水準に 4 つずつ植木鉢を用意し，無作為に種を巻いて記録します。複数の植木鉢を用意するのは，その他の条件を相殺させるための手続きですね。

3 水準 × 各群 4 鉢ですので，合計 12 点のデータがあります。

表 12.1 一要因 3 水準 Between 計画のデータ

ID	水準	記号表記	value
1	毎日あげる	y_1	6
2	毎日あげる	y_2	6
3	毎日あげる	y_3	5
4	毎日あげる	y_4	5
5	時々あげる	y_5	7
6	時々あげる	y_6	4
7	時々あげる	y_7	7
8	時々あげる	y_8	6
9	あげない	y_9	4
10	あげない	y_{10}	3
11	あげない	y_{11}	4
12	あげない	y_{12}	6

これを見ると結果としての数字はバラバラです。ですが，群ごとに平均を求めると，個体差等今回見たくない他の要因は消しあって効果の大きさだけが見られるということでしたよね。そこでそれぞれの群の平均を計算してみましょう。それぞれの群平均を計算してみましょう。水をあげる群は，

$$\bar{y}_{\text{あげる}} = \frac{1}{4}\sum_{i=1}^{4} y_i = 5.5$$

時々あげる群は，

$$\bar{y}_{\text{ときどき}} = \frac{1}{4}\sum_{i=5}^{8} y_i = 6.0$$

あげない群は，

$$\bar{y}_{\text{あげない}} = \frac{1}{4}\sum_{i=9}^{12} y_i = 4.25$$

であり，全体平均は

$$\bar{y}_{gm} = \frac{1}{12}\sum_{i=1}^{12} y_i = 5.25$$

となります[*6]。

ここで結果から逆に，どうしてこういった数字になったのかを考えてみたいと思います。もし仮に，水に成長を変えるような効果がまったくなかったとすれば結果はどうなったでしょうか。ランダム化したことによって，見たい要因以外の効果はすべてキャンセルアウトされていたはずです。個体差も今回みたい要因ではなかったわけですから，そのために各群に 4 つの鉢を用意しておいたのです。複数用意することで平均が算出でき，この平均は個体差をキャンセルアウトしあった代表的な値のはずです。

であれば，12 鉢全部の平均は，個性を無くした全体の傾向のはずですね。つまり，群わけのような人為的な操作がなければ，原理的には全部同じ数字すなわち全体平均である 5.25cm になるはずなのです。

しかしもちろん実際には，個体差があります。さらに水も何らかの効果があったと思われます。その大きさを考えてみましょう。水の効果はどのぐらいあったのでしょうか。水をあげた群，すなわち y_{i1} のグループは，平均で 5.5cm の伸びになっています。何の効果も考えなければ 5.25cm だったはずなのに，水をあげた群は 5.5cm になっていた。この事実から考えるに，差分すなわち $5.5 - 5.25 = 0.25$cm が水をあげたことによる効果だったと考えることができます。水を時々あげたときの効果はどうでしょうか。y_{i2} のグループは，平均で 6.0cm の成長しています。まったく何の効果もなければ，5.25cm しか伸びなかったのに，この群は 6.0cm 伸びていたわけですから，$6.0 - 5.25 = 0.75$cm 普通より大きいことになります。これが水を時々あげたことによる効果です。水をまったくあげなかったときの効果はどうでしょうか。y_{i3} のグループの平均は，4.25cm です。まったく何の効果もなければ，5.25cm の成長ができたはずなのに，今回は 4.25cm しか成長できませんでした。これは，水をあげなかったことによる効果が，$4.25 - 5.25 = -1.0$cm だったということができます。もちろん平易な日本語で言えば，水をあげなかったことによって 1 センチメートル成長が抑えられた，成長の効果はなかったというところですが，ここは表現を整えるためにもマイナスの効果があった，という言い方をしていきましょう。

このように，**群平均－全体平均＝効果**と計算できます。これが実験計画の基本的なアイディアです。ポイントは，すでにさまざまな影響が入り込んでいる研究対象を，ランダムに配置することによって，平均点を揃えることができる。その上で，介入によって平均値を変えることができる。その変わった平均値の大きさが，効果の大きさだったと考えるということです。ここで注意して欲しいのは，それぞれの群に見られた効果をすべて足し合わせると，ゼロになってしまうということです。すなわち，$+0.25 + 0.75 - 1.0 = 0$ です。このようにあくまでも相対的な比較で検証していることに注意しておいてください。

実はこれで話が終わりではありません。もう 1 歩進んで，個体差のことも考えます。さきほどの考え方の例で言えば，全体の平均が 5.25，水をあげる群には +2.5 の効果が出てきますから，群 1 に含まれる植木鉢の数字はすべて 5.5 であるべきなのです。もちろんそのようにはなっていません。$y_{11} = 6$ で $y_{41} = 5$ です。この，理想通りにいかなかった理由は何でしょうか？ そうこれが，今回の実験では考えられなかった効果です。

[*6] 総和記号の添字に注意して読みましょう。もう一つ別の表記法として，水準も記号にしてしまう方法があります。すなわち 3 つの水準を記号 j で表し，データを y_{ij} のようにします。ここで i は第 j 群の何番目の鉢か，を表す数字であり，1 から 4 までしか値をとりません。これを使って，$\bar{y}_j = \frac{1}{4}\sum_{i=1}^{4} y_{ij}$ と書くとスマートですね。数式は文字を含まない方が美しいと思いませんか。

植物の持っている個体差なのかもしれません。植木鉢に埋められた種と土の接触面積が違ったのかもしれません。あるいはたまたま風が強く当たる場所だったとかそうでなかったとか。ともかく今回の実験では考えられなかった効果です。これらをひっくるめて，**偶然誤差**と呼ぶしかありません。この偶然誤差がどれぐらいそれぞれの答えに出たかを算出してみましょう。

y_{11} は 5.5 であるべきところが 6 だったのですから，$6 - 5.5 = 0.5$ の誤差が出た，ということができます。同様にすべての個体に対して，**結果－群平均＝誤差**という計算式を当てはめることができます。

記号だと分かりにくいかもしれませんが，データで考えると表 12.2 のようになります。

表 12.2　一要因 3 水準 Between 計画のデータを分解する

ID	condition	notation	value(y_i)	全体平均 (b_0)	群の効果 (b_j)	誤差 (e_{ij})
1	毎日あげる	y_1	6	5.25	+0.25	+0.50
2	毎日あげる	y_2	6	5.25	+0.25	+0.50
3	毎日あげる	y_3	5	5.25	+0.25	−0.50
4	毎日あげる	y_4	5	5.25	+0.25	−0.50
5	時々あげる	y_5	7	5.25	+0.75	+1.00
6	時々あげる	y_6	4	5.25	+0.75	−2.00
7	時々あげる	y_7	7	5.25	+0.75	+1.00
8	時々あげる	y_8	6	5.25	+0.75	+0.00
9	あげない	y_9	4	5.25	−1.00	−0.25
10	あげない	y_{10}	3	5.25	−1.00	−1.25
11	あげない	y_{11}	4	5.25	−1.00	−0.25
12	あげない	y_{12}	6	5.25	−1.00	+1.75
合計					0.00	0.00

ここでの計算式を言葉で表現すると，個々の値＝全体平均＋効果＋誤差，という形になります。

Oh La La!どこかで見たことありませんか，この式の形？

そう，**線形モデル**ですね！ $y_i = b_0 + b_1 x_i + e_i$ と同じ，すなわち個々の値 (y_i) が，全体平均 (b_0) と効果 ($b_1 x_i$) と誤差 (e_i) に分解されています。**実験計画は線形モデルだったのです。**違いは x_i が説明変数の大きさではなくて，どの群に含まれているかを表す離散的な変数になっているということだけです。回帰分析では独立変数も従属変数も連続変数だったのですが，説明変数が離散的な回帰分析のことを我々は実験 (計画) と呼んでいるのです。

12.4　効果はどのように評価するのか

このようにデータを効果と誤差に分解できました。あとは，効果がどれぐらいあったかをどう評価するかを考えます。たとえば「水を毎日上げた群」はどれほど効果があったのでしょうか。4 つの植木鉢いずれについても +0.25 の効果があるのですが，誤差は +0.5 から-0.5 までさまざまです。誤差の大きさを評価するために，足し合わせちゃいましょう！となると当然ながら，どの群においても誤差の総和は 0 になります。効果が全体平均からの相対的な違いであったことと同じで，誤差というのも群平均からの相対的な違いでしかないからです。

これは「大きさ」を足し合わせることで探そうとしたことに問題があります。誤差はすべて足し合わせるとゼロになりますが，それは偶然誤差の位置の話であって，大事なのはどの程度のばらつきをもって現れるか，に

あります。誤差は分散で大きさを考えなければなりません。つまり，真値から離れている距離の平均を考えることになるのです。

効果にしてもそうです。効果も大きさで考えるときに総和で考えるのではなく，「水をあげたりあげなかったり，時々あげたりする」という変化に応じて，結果がどの程度ばらついたのかを考えなければなりません。

誤差にしても効果にしても，その大きさを見積もるためには表 12.2 の各要素を二乗して足し合わせたものを使います。これをとくに**平方和 (Sum of Squares:SS)** と言います。平方和を足し合わせた数で割って平均化したものが分散ですから，平方和と分散は同じ情報を持っているものと考えることができます。どうして平均化せずに総和のままにしておくのかについては，後期の第 12 講あたりで解説しますので，今は言葉として知っておいてください。

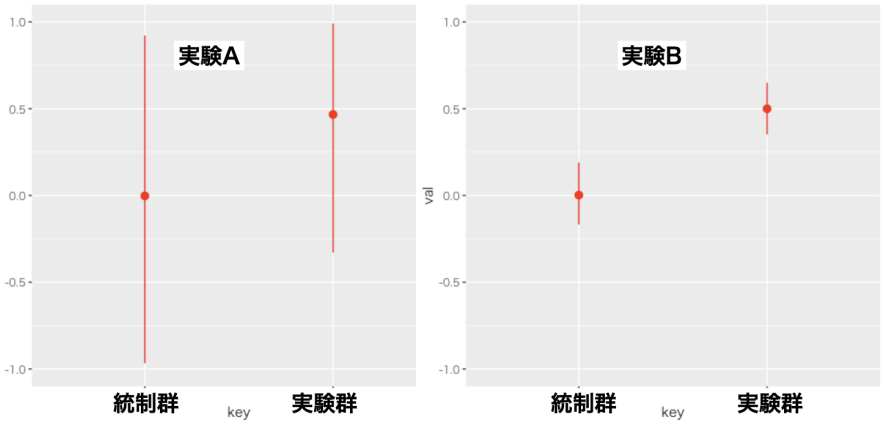


図 12.5 平均値の差 (群の効果) が同じ 2 つの実験の例。どちらの実験が効果があると言いやすいか。

ここで図 12.5 を見てください。左に実験 A，右に実験 B とふたつの実験結果を示しています。これは箱髭図の箱をとったような図で，赤い点が群の平均値，そこから上下に伸びる線は群の最大値と最小値を結んでいます。実験 A は実験 B に比べて，分散が大きいのがすぐにわかると思います。実は実験 A と実験 B は，平均値間の差は同じです。この時，実験の効果があった，とはっきり言い切れるのはどちらでしょうか？

実験 A の方は，確かに実験群のほうが平均的には平均値が高くなっていますが，統制群の中にいる最大値の人は実験群の中の数字と遜色ないですし，実験群の最小値の人は統制群に入ってもおかしくない数字です。これに対して実験 B の方は，統制群の最大値でも実験群の最小値に及びません。いくら頑張っても実験群ほどのスコアを叩き出せないのですから，これは実験群による介入が大きかったと判断しやすくなりますね。

つまり実験によって効果があった，なかったという評価をする時は，**群内の分散に比べて群間の分散が十分大きいかどうか**で考えることになります (図 12.6)。

この考え方は，**古典的テスト理論**の**信頼性**の話とも繋がります。信頼性の定義は，全分散に占める真のスコアの分散の比率，すなわち次の式になります。

$$Rel = \frac{Var(t)}{Var(X)} = \frac{Var(t)}{Var(t) + Var(e)}$$

実験の効果を評価するのも同じで，全体の分散 SS_t を群間の分散 (あるいは効果の分散)SS_f と誤差の分散 SS_e との関係から次のように表せます。

$$\frac{SS_f}{SS_t} = \frac{SS_f}{SS_f + SS_e}$$

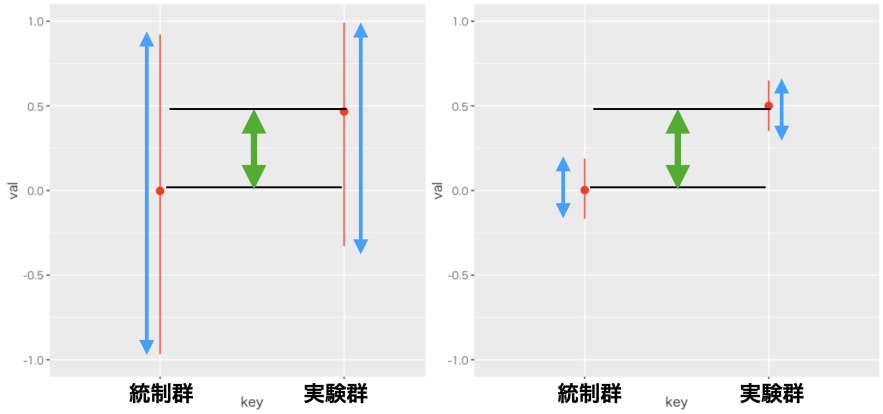


図 12.6 平均値の差 (群の効果) が同じ 2 つの実験の例。左側は群間の分散 (緑の矢印) に比べて群内の分散 (青の矢印) が大きいため, 効果があるとは言いにくい。右側の例では, 統制群の最大値が実験群の最小値にも及ばないほどなので, 効果があると言いやすい。

分散を分解し, その比率を評価するこの考え方は**分散分析 (Analysis of Variance)** とも呼ばれ, 心理学における古典的かつ基礎的な研究手法になっています。

12.5 課題

■要因と水準 次の心理学実験は, どのような要因計画でできているでしょうか。要因, 水準という言葉の違いに注意しながら, 正しく表現してみましょう。

例)「3 つの学校で共通試験をした。成績の平均値に差があるかどうかを検証したい。」→ 一要因 3 水準, 群間 (Between) デザイン

- アサガオの生長を観察するため、水をあげる群とあげない群、肥料をあげる群とあげない群をつくり、成長の程度を比較した。
- ラットの種類 A/B それぞれに対して, トランキライザーの効果を見るため, 0.4mg/0.7mg/1.5mg 投与する群を作り比較した。
- 4 つの学校で, 1 年間で共通の試験を実施した。一学期, 二学期, 三学期の各校における成績変化を比較したい。
- 二つの社会的強化の方法 (a1,a2) があり, この条件に無作為に大学生を選び割り当てた。被験者は 4 つのセッション (b1,b2,b3,b4) を体験し, 各セッションの作業量を調べ, 比較する。

■効果の算出 先の例である 3 つの学校で行われた共通試験の結果は, 表 12.3 のようになりました (児童は各校に 4 名しかいません!)。全体平均, 群平均から学校ごとの差分 (効果の大きさ) を計算し, 個々人の誤差の大きさも算出しましょう。

表 12.3　一要因 3 水準 Between 計画のデータを分解する

School	ID	Score	群平均	効果	誤差
A	1	58			
A	2	58			
A	3	64			
A	4	56			
B	1	53			
B	2	53			
B	3	54			
B	4	53			
C	1	39			
C	2	38			
C	3	35			
C	4	38			

第 13 章

実験計画法 2；Between 計画

実験計画について，用語の整理や効果の評価の方法について説明してきました。前回は一要因の実験計画でしたが，今回はこれを二要因に増やした場合について考えていきます。

13.1　二要因の場合

今度は，二要因の群間計画に進んでいきたいと思います。二要因計画とはたとえば，次のような実験計画です。

> 植物の生長を観察するため，水をあげる群とあげない群，肥料をあげる群とあげない群をつくり，成長の程度を比較したい

この例では要因は「水」と「肥料」，水準はそれぞれ「あげる・あげない」なので 2 × 2 の実験計画というのでした。

アサガオの例のまま進めても良いのですが，もう少し心理学っぽいほうがイメージしやすいかもしれませんので，以下では表 13.1 のようなコーヒーテイスティング事例を使うことにします。

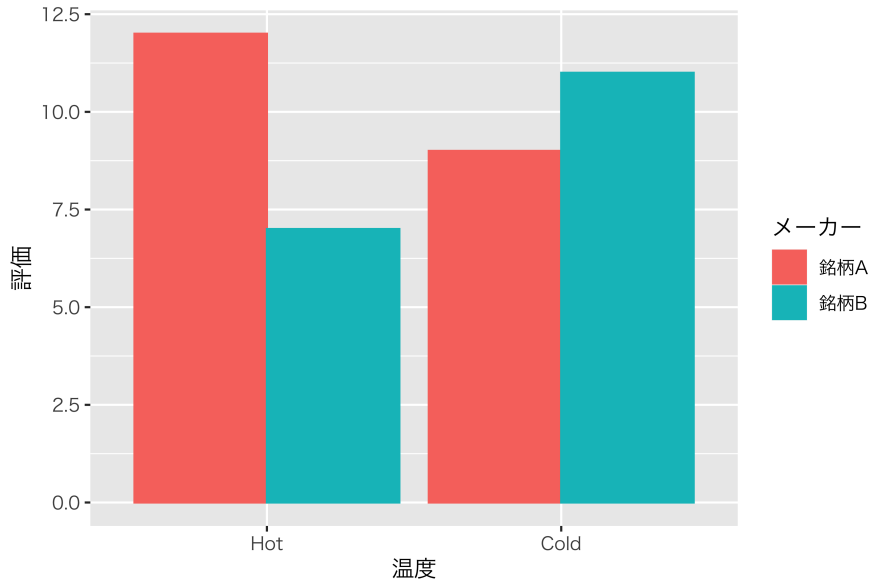


図 13.1　2 種類のコーヒーと 2 通りの飲み方による評価

2 つのメーカーがそれぞれ缶コーヒーを作っていて，ホットコーヒーにした場合，アイスコーヒーにした場合の美味しさを評価してもらいます。各条件に 3 名ずつ試飲をお願いして，20 点満点で評価してもらったら，表 13.1 のようになりました[*1]。これを図示したのが図 13.1 にあります。これを見るとわかるように，ホットコーヒーにすると銘柄 A は高い評価，アイスコーヒーだと両者の違いはなくなることがわかります。どうも銘柄 B はホットコーヒーにして売らない方がいいように思いますが，たまたまホットの銘柄 B を飲んだ 3 人がコーヒーが苦手だったのかもしれないですね。このような誤差に比べて，温度と銘柄の別の効果はどの程度あるのかを考えていきたいと思います。

表 13.1　二要因のデータ例

ID	温度	メーカー	notation	評価 (y_i)
1	Hot	銘柄 A	$X_{11}=+1, X_{12}=+1$	13
2	Hot	銘柄 A	$X_{21}=+1, X_{22}=+1$	11
3	Hot	銘柄 A	$X_{31}=+1, X_{32}=+1$	12
4	Hot	銘柄 B	$X_{41}=+1, X_{42}=-1$	7
5	Hot	銘柄 B	$X_{51}=+1, X_{52}=-1$	6
6	Hot	銘柄 B	$X_{61}=+1, X_{62}=-1$	8
7	Cold	銘柄 A	$X_{71}=-1, X_{72}=+1$	9
8	Cold	銘柄 A	$X_{81}=-1, X_{82}=+1$	9
9	Cold	銘柄 A	$X_{91}=-1, X_{92}=+1$	9
10	Cold	銘柄 B	$X_{10,1}=-1, X_{10,2}=-1$	13
11	Cold	銘柄 B	$X_{11,1}=-1, X_{11,2}=-1$	11
12	Cold	銘柄 B	$X_{12,1}=-1, X_{12,2}=-1$	9

さて，一要因の計画は，線形モデルと見ることができるという話をしていました。二要因の計画も当然，同じように線形モデルとして考えることができます。今回は二要因計画です。要因は「温度」と「銘柄」の二種類ありますから，効果として b_1 と b_2 の 2 つが必要です。これを式に書き入れてみますと，次のようになります。

$$y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + e_i \tag{13.1}$$

ここで左辺の y_i は ID が i の人の評価で，たとえば $y_1 = 13, y_2 = 11, \cdots, y_{12} = 9$ ということになります。X_{ij} は i さんが要因 j のときにどちらに割り振られたかを表しているのだと思ってください。$j = 1$ は温度の要因，$j = 2$ は銘柄の要因を表しています。ですから，X_{i1} は i さんが温度要因で Hot/Cold のどちらを飲んだかを表します。具体的には，X_{11} は Hot 条件です。X_{71} は Cold 条件です。これは**名義尺度水準**であり，Hot/Cold になんらかの数字を割り振らなければなりません。数字は何でもいいといえば何でもいいのですが，ここでは数式的にわかりやすくするために Hot に $+1$ を，Cold に-1 を割り振ることにします。こうすることで，b_1 が温度効果の大きさを表すことになりますね。Hot/Cold で符号が違って大きさが同じですから，相対的な大きさを表しているだけになるからです[*2]。$X_{11} = +1, X_{71} = -1$ のような数字が入るものだと思って読み進めてください。同様に，X_{i2} は銘柄水準の割当を表していて，銘柄 A は $+1$, 銘柄 B は-1 を表しています。具体的に書き起こしてみましょう。

[*1] 添字にカンマが入っているのが美しくないですね。しかし他の良い方法が思いつきませんでした。

[*2] 他にも Hot を 0,Cold を 1 とすることもできます。すると Hot 条件を基準にした Cold 条件の変化の大きさを表したことになります。詳しくは次講で扱います。

$$
\left\{
\begin{array}{rcccl}
y_1 & = & b_0 + b_1 X_{11} + b_2 X_{12} + e_1 & = & b_0 + b_1 + b_2 + e_1 \quad \text{Hot で銘柄 A} \\
\vdots & & \vdots & & \vdots \\
y_4 & = & b_0 + b_1 X_{41} + b_2 X_{42} + e_4 & = & b_0 + b_1 - b_2 + e_4 \quad \text{Hot で銘柄 B} \\
\vdots & & \vdots & & \vdots \\
y_7 & = & b_0 + b_1 X_{71} + b_2 X_{72} + e_7 & = & b_0 - b_1 + b_2 + e_7 \quad \text{Cold で銘柄 A} \\
\vdots & & \vdots & & \vdots \\
y_{10} & = & b_0 + b_1 X_{10,1} + b_2 X_{10,2} + e_{10} & = & b_0 - b_1 - b_2 + e_{10} \quad \text{Cold で銘柄 B} \\
\vdots & & \vdots & & \vdots
\end{array}
\right.
$$

ここで X_{ij} は符号だけを表している名義尺度水準であることを確認しておいてください。

改めて式 13.1 や式 13.1 を見ると，これも線形モデル，とくに説明変数が複数ある**重回帰分析**のモデルと同じ形をしているのが分かりますね。重回帰分析は第 10 講でやりましたが，そのモデル式と同じ形をしているのが確認できると思います。要因計画は線形モデルなのです。

13.2 効果の大きさを算出する

それでは，効果の大きさ，誤差の大きさを見積もってみましょう。二要因の場合でも考え方は同じです。無作為に割り当てたのですから，人間の手を加えなかったらデータの散らばりのすべては誤差 (個体差) です。誤差はキャンセルし合うはずなので，理論的にはすべて平均値に一致するはずです。

しかし，やれ温度の操作だとか，やれメーカーの違いだとかで人によって違う介入があり，しかも人間のやることですので個人差も加わってしまいました。だからデータがばらついてしまったのだ，と考えるのです。

理想的には全部同じはずと考えると，実験もしなければ個人差もない，理想的なデータの値は全体平均 $\bar{X}_{gm} = 9.75$ にぴったり合致するはずなのでした。しかしそこに温度の効果 b_1 が関係してきます。図 13.2 の左にあるように，温度で群わけして Hot 群だけの平均を計算してみましょう。

$$\bar{y}_H = \frac{1}{6}\sum_{i=1}^{6} y_i = \frac{1}{6}(13 + 11 + 12 + 7 + 6 + 8) = 9.5$$

同じく，Cold 群だけの平均は次のようになります。

$$\bar{y}_C = \frac{1}{6}\sum_{i=7}^{12} y_i = \frac{1}{6}(9 + 9 + 9 + 13 + 11 + 9) = 10.0$$

ここで，群平均-全体平均=効果だと考えられますから，Hot 群に割り当てられると -0.25 の効果が，Cold 群に割り当てられると $+0.25$ の効果があったということになります。

同様に，メーカーの効果も考えましょう。温度の効果を潰して (温度の効果で分けずに) 銘柄 A のデータだけ取り出して，平均値を計算してみます[*3]。

$$\bar{y}_A = \frac{1}{6}\left(\sum_{i=1}^{3} y_i + \sum_{i=7}^{9} y_i\right) = \frac{1}{6}(13 + 11 + 12 + 9 + 9 + 9) = 10.5$$

同様に，銘柄 B の平均値を計算します。

$$\bar{y}_B = \frac{1}{6}\left(\sum_{i=4}^{6} y_i + \sum_{i=10}^{12} y_i\right) = \frac{1}{6}(7 + 6 + 8 + 13 + 11 + 9) = 9.0$$

[*3] この数式の美しくないこと！ $\sum$ の中身が 2 つに分裂してしまっています。しかしこれは i を通し番号でつけたから仕方のないこと。ここでは i が ID をあらわしていますから，その何番目から何番目までが加算されているかを確認してください。

群平均-全体平均=効果ですから，銘柄 A は全体平均より $10.5 - 9.75 = 0.75$ 高くなる効果が，相対的に銘柄 B は全体平均より 0.75 下がる効果がある (-0.75 あげる効果がある)，ということがわかります。

図 13.2 に，群ごとに平均値をプロットし，全体平均 (9.75) のラインを横一線で引いてみました。これを見ると，各群の効果が全体平均 (b_0) を挟んで，相対的に同じ大きさで影響しているのがわかると思います。

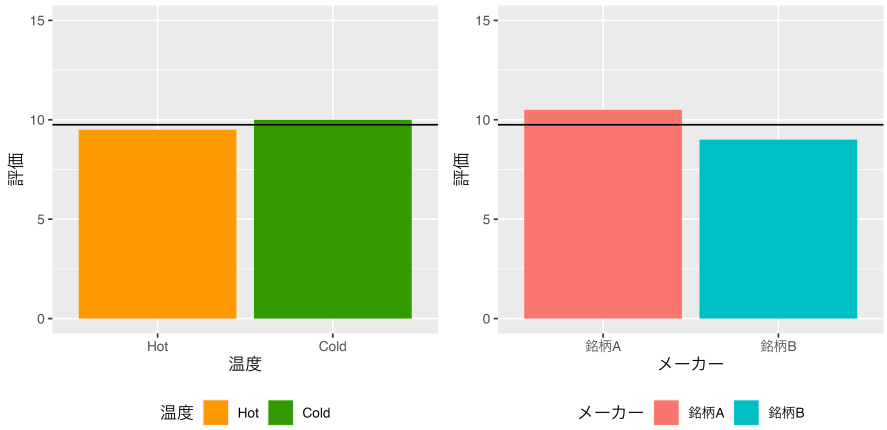


図 13.2 群ごとにプロットしたもの。全体平均から見た相対評価になっていることが確認できる。

さて，これらの効果がわかったところで，足し算を組み上げていきましょう。全体平均にそれぞれの効果を加えていきます。銘柄 A をホットコーヒーでいただくとすると，全体平均 9.75 に Hot の効果があって (-0.25) の銘柄 A の効果 ($+0.75$) も入りますから，$9.75 - 0.25 + 0.75 = 10.25$ となり，これが銘柄 A をホットで飲んだ時の平均値になるはずですが・・・

$$\bar{y}_{H,A} = \frac{1}{3}\sum_{i=1}^{3} y_i = \frac{1}{3}(13 + 11 + 12) = 12$$

あれ？ なりませんでした。

13.3 交互作用

13.3.1 交互作用の算出

実は，同様のことは他の組み合わせのところでも生じます。群平均からの差分が誤差なのだ，と考えれば良いはずですが，群平均が理論通りにならないのです。具体的に，表 13.2 に計算してみました。効果の総和と群平均，理論的には合致するはずです。これはどうしたことでしょう。

銘柄 A($+0.75$) のホットコーヒー (-0.25) が 10.25 で，群平均が 12.0 ということは，$12.0 - 10.25 = +1.75$ の差分があります。これも実は，考慮すべき効果のひとつなのです。これは「A をホットで」という組み合わせの時に**のみ**生じる組み合わせの効果で，とくに**交互作用 (interaction)** といいます[*4]。

交互作用とはこのように，水準同士の組み合わせによって生じる特別な効果のことを指します。交互作用に対して，これまでみてきた群の平均は**主効果 (main effect)** と呼んで区別します。式 13.1 にあった線形モデルは，次の式 13.1 のように交互作用項も含めたものにアップデートしなければなりません。

$$y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i \tag{13.2}$$

[*4] 余談ですが，統計的な文脈では interaction を「交互作用」と訳します。社会心理学などの文脈では，「相互作用」と訳されます。

表 13.2 二要因のデータ例

全体平均	温度の効果	メーカーの効果	効果を足し合わせると	群平均は	評価
9.75	-0.25	+0.75			13
9.75	-0.25	+0.75	10.25	12.0	11
9.75	-0.25	+0.75			12
9.75	-0.25	-0.75			7
9.75	-0.25	-0.75	8.75	7.0	6
9.75	-0.25	-0.75			8
9.75	+0.25	+0.75			9
9.75	+0.25	+0.75	10.75	9.0	9
9.75	+0.25	+0.75			9
9.75	+0.25	-0.75			13
9.75	+0.25	-0.75	9.25	11.0	11
9.75	+0.25	-0.75			9

この交互作用は，式 13.2 にもあるように，主効果とは別の独立した影響源です。ですから，主効果がなくても，交互作用だけある，ということがあり得ます。今回の例で言えば，温度の違いや銘柄の違いははっきりしないが，銘柄 B は温めるととんでもなくうまくなる，といった場合ですね。

つまり，交互作用はいわば「ただしイケメンに限る」という条件のようなものです。かっこいいセリフや仕草は，それぞれに評価を上げる効果がなくとも (あってもいいんですが)，イケメンがするからとくに良いのであって，イケメンでない人がするとむしろ逆効果，ということもあるのでそれを考慮しておこう，ということです[*5]。

さて，この組み合わせの効果も相対的なもので，全部足し合わせると 0 になるようにしなければなりません。相対的に評価する，というのは効果をただ足し合わせると 0 になるということだからです。他の交互作用の大きさを計算して確かめてみましょう。

銘柄 A(+0.75) の Cold 条件 (−0.25) の効果が全体平均に加わると 8.75 で，群平均が 7.0 ですから，ここには −1.75 の交互作用効果が出ています。同様に銘柄 B の Hot 条件には −1.75，銘柄 B の Cold 条件には +1.75 ですから，4 つの要素を足すとキャンセルしあって合計 0 になっていますね。これを踏まえて，式 13.1 をアップデートしましょう。

[*5] このたとえはポリコレ的に正しくないというか，恋愛至上主義的で容姿に基づく人物評価を助長するような笑いの形式になっているので，個人的には好ましく思っていないのです。が，みなさんが交互作用を身近に感じてもらえるのでは，と思って書いています。個人的には熱燗で飲む日本酒と冷やで飲む日本酒って違うよね，という例のが一番ピンとくるんですが，お酒を飲まない人，あるいは未成年の方にとってはこれも不適切な例かなと思ったり。本文はコーヒーにしていますが，わかりやすい例になっているでしょうか。何か良いアイデア，たとえがありましたら，ご一報ください。

$$
\left\{
\begin{array}{rcccl}
y_1 & = & b_0 + b_1 X_{11} + b_2 X_{12} + b_3 X_{13} + e_1 & = & b_0 + b_1 + b_2 + b_3 + e_1 \quad \text{Hot で銘柄 A} \\
\vdots & & \vdots & & \vdots \\
y_4 & = & b_0 + b_1 X_{41} + b_2 X_{42} + b_3 X_{43} + e_4 & = & b_0 + b_1 - b_2 - b_3 + e_4 \quad \text{Hot で銘柄 B} \\
\vdots & & \vdots & & \vdots \\
y_7 & = & b_0 + b_1 X_{71} + b_2 X_{72} + b_3 X_{73} + e_7 & = & b_0 - b_1 + b_2 - b_3 + e_7 \quad \text{Cold で銘柄 A} \\
\vdots & \vdots & \vdots & & \\
y_{10} & = & b_0 + b_1 X_{10,1} + b_2 X_{10,2} + b_3 X_{10,3} + e_{10} & = & b_0 - b_1 - b_2 + b_3 + e_{10} \quad \text{Cold で銘柄 B} \\
\vdots & & \vdots & & \vdots
\end{array}
\right.
$$

交互作用項 (b_3) のもつ符号は，プラスとマイナスが上から順に $+, -, -, +$ となっています。どうしてこういうことになっているのかについて，説明を追加しておきましょう。

今回は 2×2 の計画ですから，図 13.3 にあるように組み合わせの表を書いてみると，主効果はその周辺の平均値で考えていたことがわかります。たとえば 1 つ目の要因である温度 (図中では要因 A とします) の，「Hot(水準 a1)」か「Cold(水準 b1)」かの話は，もう 1 つの要因であるメーカーを潰して，列方向に 1 つのグループにした平均からその効果を計算していました (b_1)。この水準同士の比較は相対的なので，水準 1 と 2 は符号違いで $+b_1, -b_1$ になっています。同様に，2 つ目の要因である銘柄 (図中では要因 B とします) についても，要因 A の水準を潰して行方向に 1 つのグループと考え，平均を出して効果を算出し，相対比較にしていました $(+b_2, -b_2)$。

これに対して，交互作用はセル内，各組み合わせについて生じている効果です。行方向，列方向についてはそれぞれの要因の主効果として計算されていますから，交互作用は行方向に足し合わせても ± 0，列方向に足しあわせても ± 0 になる必要があります。そうでないと，周辺の計算が合わなくなってしまうからです。図 13.4 の左上，a1b1 水準に交互作用効果 b_3 が生じたとすると，列方向，つまり同じ水準 a1 のなかで，総和がゼロにならないといけませんから，a1b2 水準には $-b_3$ が入ることになります。これはまた，行方向にも総和がゼロでなければなりませんから，a2b1 も $-b_3$ である必要があります。残る右下の a2b2 は，行方向にも列方向にも総和ゼロにするためには $+b_3$ でなければなりません。このようにして，交互作用の符号は周りとの関係で定まっていたのです。交互作用の大きさは，どの行方向，列方向をみても総和ゼロになっている，というのは水準数が増えても同じです。

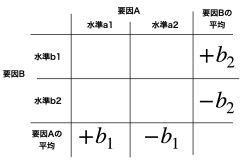


図 13.3 主効果の計算は周辺の平均値から計算する

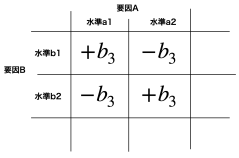


図 13.4 交互作用はセルの内部で相殺するように

このように交互作用を含めて計算すると，組み合わせ水準のところまで含めて，群平均の計算が合うようになります。それでも一人ひとりのレベルまで考えてみると，やはりズレが出てしまいますが，これはどうしようもない個人差＝誤差ですので，これを計算することで誤差を算出できます。

表 13.3 には，それぞれの効果の大きさと，それを二乗して足したものを書きました。二乗して足し合わせたものはとくに**平方和 (Sum of Squares; SS)** といいます。これはもちろん，大きさを比較するために幅の計算にするためです。評価の列にも平方和がありますが，これは各評価と全体平均の差の平方和になっています。評価の平方和はデータ全体のばらつきを表していますが，この全体平方和＝

表 13.3 平方和の計算

全体平均	温度の効果	銘柄の効果	交互作用	誤差	評価
9.75	-0.25	+0.75	+1.75	+1.00	13
9.75	-0.25	+0.75	+1.75	-1.00	11
9.75	-0.25	+0.75	+1.75	0.00	12
9.75	-0.25	-0.75	-1.75	0.00	7
9.75	-0.25	-0.75	-1.75	-1.00	6
9.75	-0.25	-0.75	-1.75	+1.00	8
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	-0.75	+1.75	+2.00	13
9.75	+0.25	-0.75	+1.75	0.00	11
9.75	+0.25	-0.75	+1.75	-2.00	9
平方和	0.75	6.75	36.75	12.00	56.25

温度効果の平方和＋銘柄の平方和＋交互作用の平方和＋誤差の平方和となっているのがわかります ($56.25 = 0.75 + 6.75 + 36.75 + 12.00$ です)。2 要因計画の場合も，交互作用を考慮して，データの散らばりを足し算の形に分解できました。

13.3.2 交互作用のパターン

もう一度冒頭の図 13.1 にもどって，2 水準 ×2 水準のスコアをそれぞれ見てみましょう。横軸は温度の 2 水準で，名義尺度水準の数字すなわち離散的な変数ですから適切ではないのですが，同じ条件のところを線で結んでみますとクロスしているのがみて取れると思います (図 13.5)。交互作用が見られる典型的なパターンの 1 つが，このように平均点を結ぶとクロスする形になります。

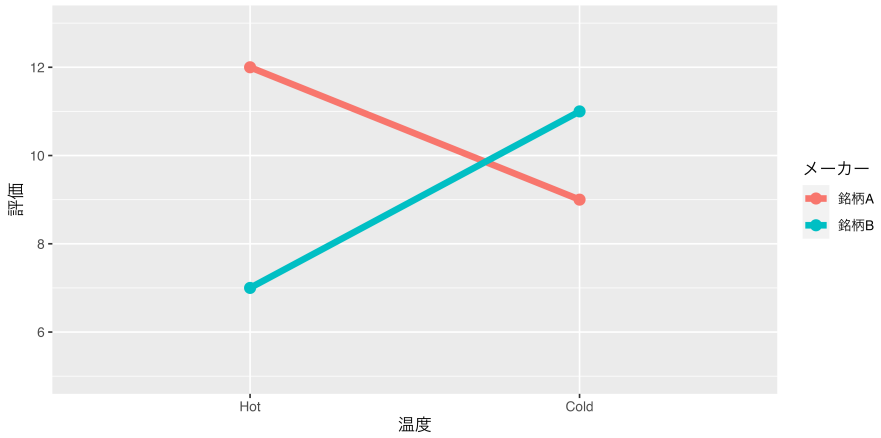


図 13.5 平均点を結ぶとクロスする

ここで再確認ですが，主効果というのは各群の平均における差を見るものです。温度の効果を見るときはメーカーの水準に関係なく，温度の Hot/Cold の別だけに注目して平均を見ます。逆にメーカーの効果を見

るときは，温度の水準に関係なく銘柄 A/B の違いを見ます。図 13.6 には先ほどのクロスしたラインに主効果の位置を書き加えました。図 13.6 の左にあるのが，温度の主効果を比べる時の点で，右にあるのがメーカーの主効果を見るときの場所です。主効果はこのように，それ以外の群の水準を潰して平均を見ているわけですね。図 13.3 のように，表で理解した方がわかりやすい人もいるかもしれませんが，表でピンとこなかった人は図でイメージしてみてください。

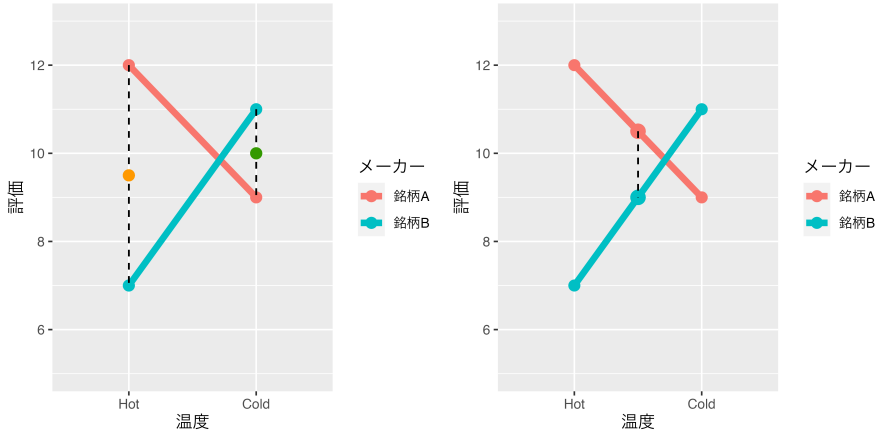


図 13.6 主効果はどこを比べているのか確認しておこう

ところで，交互作用は必ずしもこのクロスしたような状況だけを指すものではありません。基本的に主効果ではなく，組み合わせた時に初めてみられる効果のことを言いますから，交互作用の現れ方は他にもあり得ます。順にいろいろなパターンを見てみましょう。

まず要因 A,B それぞれの主効果があり，交互作用がない例が図 13.7 です。要因 A の第一，第二水準の間に差があるので要因 A の主効果がありますし，要因 B のみ，つまり左右で見比べてもその平均に違いがあります。

一方の要因しか効果がない場合が図 13.8,13.9 です。図 13.8 は上下の違いだけで左右に違いがありませんから，要因 A の効果だけが出ていることになります。図 13.9 は左側と右側の違いがありますが，要因 A の水準に違いがありません。またこの図には交互作用が入っておらず，組み合わせの効果がないのでパターンが比較的みやすくなっています。

図 13.10 には，主効果も交互作用も含まれている例を示しました。上側・下側の比較でも差がありますし，左側・右側だけでも差があります。さらに左右を結んだ線が平行になっていないということは，特別な組み合わせの効果がどこかに出ていることを意味します。

図 13.11 は，コーヒーテイスティングの例の極端にしたようなパターンです。これは要因 A の第一水準の平均と第二水準の平均を取っても同じ，要因 B のそれも同じなのですが，明らかに効果の現れ方が一方向てきではありませんね。それぞれの主効果がないのに，交互作用だけあるというパターンです。

今回のコーヒーテイスティング事例では，主効果の大きさよりも大き差の方が強かったので，図 13.11 のような形になったのでしょう。実際に「どれぐらい大きければ意味があるのか」といったことについては，後期の推測統計学に入ってから学ぶことになります。そこでは効果の大きさを評価し，意味があるのかどうかを判断するということについて学びますが，今のところはひとまず算出方法と，出現パターンのあり方について把握しておいてください。何にせよ，**データは図にする**ことの重要性が，改めて分かったのではないかと思います。

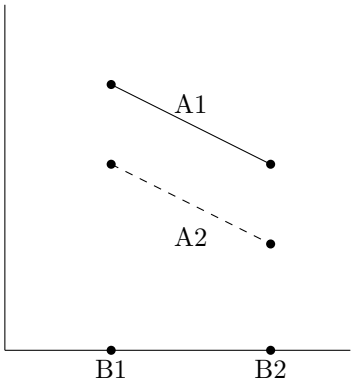


図 13.7　両要因の主効果だけがある

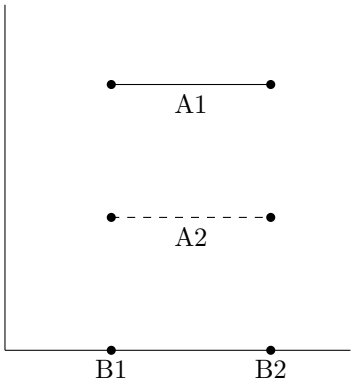


図 13.8　要因 A の主効果だけがある

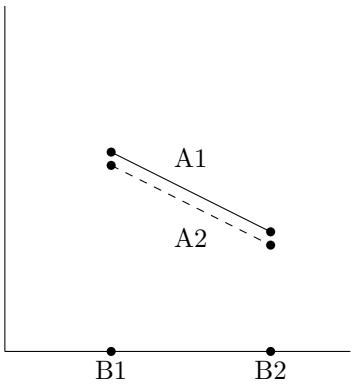


図 13.9　要因 B の主効果だけがある

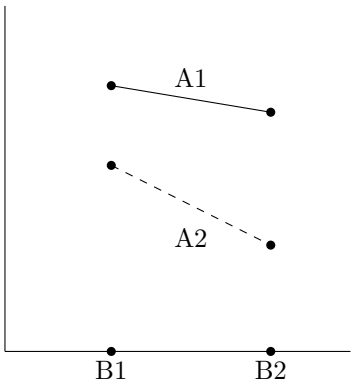


図 13.10　主効果も交互作用もある

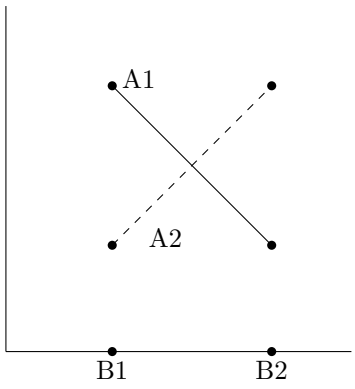


図 13.11　主効果はないが交互作用だけがある

13.3.3　組み合わせが増えると

ところで今回，要因の組み合わせによる影響，つまり交互作用を考えようという話でした。今回は 2 要因の例でしたが，3 要因になるとさらにその組み合わせは増えます。下に要因計画ごとの要素の組み合わせを書

いてみました。

一要因群間計画	全体変動	=	要因 A + 誤差
一要因群内計画	全体変動	=	要因 A + 個人差 + 誤差
二要因群間計画	全体変動	=	要因 A + 要因 B + 交互作用 AB + 誤差
三要因群間計画	全体変動	=	要因 A + 要因 B + 要因 C + 交互作用 AB + 交互作用 AC + 交互作用 BC +交互作用 ABC + 誤差
四要因群間計画	全体変動	=	要因 A + 要因 B + 要因 C + 要因 D +交互作用 AB + 交互作用 AC + 交互作用 AD + 交互作用 BC +交互作用 BD + 交互作用 CD + 交互作用 ABC + 交互作用 ABD +交互作用 BCD + 交互作用 ABCD + 誤差

3 要因になると，3 つの主効果 (効果 A,B,C) に加え，組み合わせの効果が AB,AC,BC の一次の交互作用だけでなく，ABC3 つ揃った時の二次の交互作用も考える必要があります。4 要因についても書いてみましたが，もう訳が分からないレベルです。

このことからわかるように，要因計画はいくらでも複雑にしても良い，というものではありません。要因が増えるごとに考えるべき要素が増えるだけであり，結局統一的な解釈が難しくなります。お勧めとしては，交互作用のないような現象について，二要因までの研究計画にしておいた方が良いように思います。

13.4 課題

間 (2)× 間 (2) の実験計画で，各セル n 人になるように配置してとある課題をやってもらったとします。2 つの要因をそれぞれ A,B とし，A の第一水準を a_1 のように表記することにします。

a_1b_1 の条件での平均点は 65 点，a_1b_2 の条件での平均点は 45 点，a_2b_1 の条件での平均点は 41 点，a_2b_2 の条件での平均点は 49 点でした。

この時，次の大きさを算出してください。効果の大きさは群平均と全体平均の差で表現できます。

1. 全体平均 (Grand Mean, GM)
2. 要因 A の効果の大きさ ($|b_1|$)
3. 要因 B の効果の大きさ ($|b_2|$)
4. 交互作用 AB の効果の大きさ ($|b_3|$)

なお，水準の組み合わせごとの平均を表にしたものが表 13.4 です。参考にしてください。

表 13.4 実験計画と効果

	a_1	a_2	
b_1	65	41	$+b_2$
b_2	45	49	$-b_2$
	$+b_1$	$-b_1$	GM

第 14 章

実験計画法 3；一要因モデル (Within)

ここまで，線形モデルとしての実験計画法を見てきました。一要因，二要因がそれぞれ単回帰分析，**重回帰分析**に対応すること，二要因以上になると組み合わせの効果が出現することなどを見てきました。しかしここまでの話はいずれも，**群間要因**計画の話でした。そこで**群内要因**計画の場合はどのようになるか，というところに目を向けていきたいと思います。

14.1 Within 計画の場合

群内要因での例を考えていくにあたって，今回も数値例を基に話を始めてから，線形モデルでの表現へ，と進めていきたいと思います。

今回は一要因 3 水準の，群内計画の要因計画の話をしましょう。データに対応がある，あるいは反復測定であることがポイントです。同じ人に複数回データを提供してもらう，次のような状況をイメージしてください。

臨床心理学者がある介入法をつかってメンタルケアに取り組んでいます。4 人のクライアントにこの手法を適用し，抑うつ度が改善されていくかどうかチェックします。表 14.1 のようなスコアが得られたとき，この介入法に効果あるかどうかを考えたい，とします。このデータを図にしてみましょう。ポイントは，3 つの調査時点

表 14.1 Within Data Example

ID	Time1	Time2	Time3
1	10	5	9
2	9	4	5
3	4	2	3
4	7	3	5

(Time1,2,3) それぞれが独立ではなく，クライアント番号 (ID) は共通しているというところです。横軸は時間だとすると，その軸に意味がありますから，同じ人のデータ点を線で結ぶことができるのです。

図 14.1 をみていると，個人差こそあれ，第一期 (Time1) から二期 (Time2) にかけてはいったん下降傾向があり，第二期から三期 (Time3) にかけては上昇傾向があるのかなあ，と思ったりします。この直感をデータで検証してくことが大事ですね。

ここで考えたいのは，「個人差こそあれ」のところです。**個人差**があるというのは，最初 10 点から始まるのか (ID=1)，4 点から始まるのか (ID=3)，というところですでに 7 点差ついてるのですが，今回これは個人差であって見たい効果の部分ではない，と考えるということです。見たい効果は「時期」要因による，3 つの水準それぞれへの影響，b_j です。数式的に表現すると，個人 i の第 j 期におけるデータ y_{ij} は，次のように表

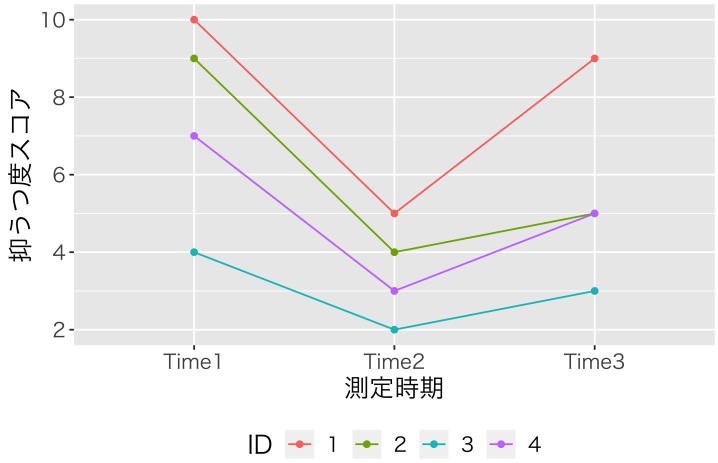


図 14.1 介入効果によるクライアントの状態変化

すことができます。

$$y_{ij} = b_0 + d_i + b_j X_{ij} + e_{ij}$$

ここで X_{ij} はまたしても，第 j 期に b_j の効果がある (1) かない (0) かを表している記号だと思ってください。具体的には，i さんの 3 つの時期 $j = 1, 2, 3$ の効果を隠さずに書くとしたら，次のようになっていることに注意です。

$$\begin{aligned}
y_{i1} &= b_0 + d_i + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_{i1} \\
&= b_0 + d_i + b_1 \times 1 + b_2 \times 0 + b_3 \times 0 + e_{i1} \\
&= b_0 + d_i + b_1 + e_{i1} \\
y_{i2} &= b_0 + d_i + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_{i2} \\
&= b_0 + d_i + b_1 \times 0 + b_2 \times 1 + b_3 \times 0 + e_{i2} \\
&= b_0 + d_i + b_2 + e_{i2} \\
y_{i3} &= b_0 + d_i + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_{i3} \\
&= b_0 + d_i + b_1 \times 0 + b_2 \times 0 + b_3 \times 1 + e_{i3} \\
&= b_0 + d_i + b_3 + e_{i3}
\end{aligned} \tag{14.1}$$

また，ポイントとして，新たに d_i というのが出てきていることに注意してください。これは i に伴って変わる数字で，時期 j の添字がついていませんから，「時期を通して一貫した個人の平均値」ということになります。b_0 は i も j もついていませんから，個人にも時期にも関係なく変化しない全体の平均値を表しています。

これを図で表したのが図 14.2 です。全体平均 b_0 とは別に，個別の平均 d_i の水平線が引かれています。個人平均は個々人のベースラインですので，時期によって変わりません。時期の効果 b_j は 4 人に等しく影響します。添字 i がないことでこれを表現しています。効果は同じであったとしても，個人の平均が違うので，データとして得られる数字はさまざまになっている，というわけです。もちろん個人差，効果を考慮しても析出できない誤差 e_{ij} がありますし，これまで同様 b_j は相対的，つまりすべての水準を足し合わせると 0 になってしまうものです。また，個人の平均値 d_i も相対的なものとして捉えられます。全体平均 b_0 が別にあるからですね。これまで同様に無作為に割り付けているので，何の効果も個人差もなかったら，全体平均の横一線に並ぶはずだ，というのが基本的なアイデアだからです。

具体的な数字を使って考えてみましょう。表 14.1 の行平均，列平均，全体平均を計算してみました。行方向の平均は個人ごとの平均です。最初の人 (ID=1) は平均が 8 ですから，全体平均から考えて +2.5 の個人差がある人なのです。また，列方向の平均は水準ごとの平均で，これが今回見たい効果ということになりま

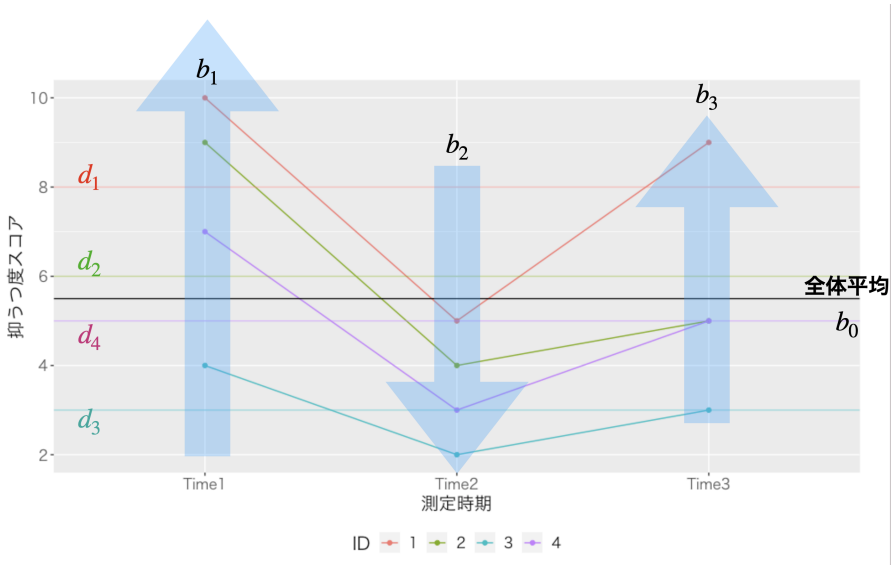


図 14.2 線形モデルで表現

す。第 1 期 (Time1) は平均が 7.5 ですから全体平均から考えて +2.0 の効果があった時期だ，ということです。

個別のデータ，たとえば ID=1 の第 2 期のデータは，理論通りに行けば，全体平均 5.5 に個人差 +2.5 と時期の効果 −2.0 が加わって出てくるはずです。5.5 + 2.5 − 2.0=6.0 になるはずですが，実際は 5 ですね。この違いは，理論では考えられないところなので，誤差になります。こうして，効果，個人差，誤差を計算

表 14.2 水準ごとの平均から効果を考える

ID	Time1	Time2	Time3	平均
1	10	5	9	8
2	9	4	5	6
3	4	2	3	3
4	7	3	5	5
平均	7.5	3.5	5.5	5.5

して表にしたのが表 14.3 です。例によって総和すると 0 ですので，平方和を計算することで相対的な大きさを比較できるようになります。

ここにあるように，個人 i の時期 j における誤差，と各セルに細かく誤差が計算できるようになりました。これが群内要因計画の強みなのです。

14.2 群間計画と群内計画の比較

ここで**群間要因 (Between Design)** 計画と**群内要因 (Within Design)** 計画の構造的な違いに触れておきたいと思います。

群内計画では，個人が誰であるかという情報があったので個人差を算出でき，それに基づいて分析しました。が，群内計画の時のデータを群間計画のように分析してみるとどうなるでしょうか。表 14.2 にある数字が同じであれば，当然全体の平方和 (分散) は同じです。ただ個人情報が抜けているのであれば，効果と誤差

表 14.3 効果, 個人差, 誤差を析出する

ID	時期	スコア	全体平均	個人差	効果	誤差
1	Time1	10	5.5	+2.5	+2.0	0.0
1	Time2	5	5.5	+2.5	-2.0	-1.0
1	Time3	9	5.5	+2.5	0.0	+1.0
2	Time1	9	5.5	+0.5	+2.0	+1.0
2	Time2	4	5.5	+0.5	-2.0	0.0
2	Time3	5	5.5	+0.5	0.0	-1.0
3	Time1	4	5.5	-2.5	+2.0	-1.0
3	Time2	2	5.5	-2.5	-2.0	+1.0
3	Time3	3	5.5	-2.5	0.0	0.0
4	Time1	7	5.5	-0.5	+2.0	0.0
4	Time2	3	5.5	-0.5	-2.0	0.0
4	Time3	5	5.5	-0.5	0.0	0.0
平方和				39.0	32.0	6.0

にしか分割できません。言い換えると, 群間計画というのは誤差の中に個人差が含まれており, 「これは誤差なのかな, 個人差なのかな？ 」という区別がつかない状態である, ということです (図 14.3)。

最終的には, 誤差の大きさに対して効果の大きさはどうか, という観点から評価します[*1]。効果の大きさ＝分子が同じデータであれば, 分母が小さい方が相対的に大きいと判断されることになりますね。群間計画の時の誤差は偶然誤差と個人差を含んでおり, 群内計画の時は個人差を含みませんから, 群内計画の方が群間計画よりも, 分析の精度が高いということになります。

であれば, すべての研究が群内計画であればよいのに, と思うかもしれません。ただし現実的な問題として, 同じ人に何度も何度もデータを提供してもらうのは, 非常に負担をかけることになります (群内計画は反復測定計画とも呼ばれるのでした)。研究倫理的な観点, あるいは計画上の手間を考えると, 群間計画にせざるを得ないということが少なくありません。研究者, 研究に協力してくれる人, のコストを考えて実験計画はデザインする必要があるということです。

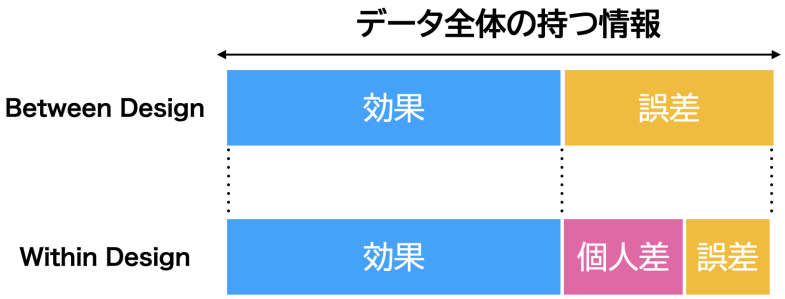


図 14.3 群内計画と群間計画の違い

[*1] 統計的な評価, 価値判断の方法については後期の推測統計学で扱うので, ここでは比較して考える要素が何であるかを理解していただければ十分です。

14.3 線形モデルによる表現と自由度

さてこれまで，実験計画は説明変数が離散変数になっただけで，線形モデルの一種であるという説明をしてきました。数学的な表現が苦手な人もいるかもしれませんが，$y = ax + b$，あるいは回帰分析の表現に則って $y = b_0 + b_1 x$ でもいいですが，要因計画も一次関数の特殊なパターンなのだという点を強調してきました。

なぜそのような流れで話をするかというと，回帰分析と要因計画は実は同じ**モデル**なんだよ，ということがわかると今度はそのモデルをこう変えたらどうなる？こっちのこれをいじったらどうなる？とどんどん発展させていくことができるからです。実際，心理学の研究で使われている分析は回帰分析系の子孫たちであり，今後これらを活用するために「要因計画で考えたら実質的にどういう違いを表しているのか」「線形モデルではどう表せるのか」と多角的なものの見方ができるようになってほしいからです。

その第一歩である，要因計画と線形モデルで一般的に[*2]説明できるということを表して，**一般線形モデル (General Linear model)** という言い方をします。これを少し詳しく見てみましょう。

説明変数が連続変数である，いわゆる回帰分析モデルは，次のような式で表現されるのでした。

$$y_i = b_0 + b_1 x_i + e_i$$

回帰分析の下になるのは相関関係ですから，変数 x_i, y_i のいずれも x 軸，y 軸上のあらゆる値をとり得るものと考えられます。**間隔尺度水準**以上の数字で，説明変数が一単位増加したら，従属変数がどれぐらい増加するかを b_1 で表しているのでした。

要因計画の場合は，説明変数が離散的です。離散的である，というのは尺度水準でいうと**名義尺度水準**にあるということです。たとえば統制群に 1，実験群に 2 という数字を割り当てた，といった数値化をした時ですね。ただし 1,2 とコーディングしてしまうと，数式上は 2 の群が 1 の群の 2 倍という意味になるので，このままでは一般的に書くことができません。数学的にも成り立つし，実質的な意味としても通じるようなコーディングの仕方を考えましょう。

一番シンプルなのは，統制群を 0，実験群を 1 とする方法です。たとえば ID=1 の人を統制群としたら $x_1 = 0$ ですし，ID=4 の人が実験群であれば $x_4 = 1$ である，と決めるのです。こうすると，統制群の回帰モデルは次のようになります。

$$y_i = b_0 + b_1 \times 0 + e_i = b_0 + e_i$$

同様に，実験群の回帰式は次のようになります。

$$y_i = b_0 + b_1 \times 1 + e_i = b_0 + b_1 + e_i$$

どちらの群も b_0 がありますが，これは何も処置をしない群の平均値を意味します。統制群のデータの散らばりは，すべて誤差 e_i で説明するモデルということになります。

他方，実験群は，何もしない状態 b_0 に**加えて** b_1 があり，それ以外の変化は誤差 e_i で説明するということになります。ここで効果の大きさ b_1 は添字 i がついていないことを確認しましょう。効果の大きさは，個人ごとに変わるものではないからです。この b_1 の総量が効果の大きさになりますが，比較すべき誤差 e_i の総和はゼロになりますから，そのままでは比較せずに二乗して比較するようになります。

[*2] 一般といわれると「ありふれていること」「広く行き渡っていること」を意味しているように聞こえますが，元々の意味は「同一である」というものです。他のものと同一，他のものを通じて変わらないもの，と思ってください。

この表現の仕方は，統制群に比べて効果の大きさ b_1 を見るというもので，比較的わかりやすい表現かと思います。でも水準数が 3 つになったらどう表したらいいのでしょうか。

大丈夫です，このときも基本的に考え方は変わりません。3 つの群として前回の例，「水を毎日あげる」「水をときどきあげる」「水をあげない」とあったとしましょう。その上で，第一の群である「毎日水をあげる」群を基本に考えて式を立ててみましょう。

$$\begin{cases} y_{i1} = b_0 + e_{i1} & \text{毎日水をあげる群} \\ y_{i2} = b_0 + b_1 + e_{i2} & \text{ときどきあげる群} \\ y_{i3} = b_0 + b_2 + e_{i3} & \text{あげない群} \end{cases}$$

このようにすると，まず「水をあげる群」は $b_0 + e_i$ となっています。e_i は誤差・個体差を表しているので，この式は水をあげる群の各個体は，群平均と誤差からなるということを意味しています。続いて第二の群ですが，式は $b_0 + b_1 + e_i$ になりました。これは b_0 すなわち水をあげる群の平均ですから，それに比べて「ときどき」群の平均が b_1 だけ増え，さらに群内の個体差があるという順になります。あげない群も同様に，b_0 に比べて b_2 だけ平均が違う，ということになります。

しかしなぜ「水をあげる群」からスタートしたんでしょうね。あげない群と比べた方が，意味的にはわかりやすかったですかね。でも他のシーンで，A,B,C....X 群とたくさんの水準がある時に，なぜ特定の群を基準にしたのだと言われても合理的な説明ができないかもしれません。

また要因計画では前提として，「群わけしなければ同じ状態だったもの」に「群わけ」という操作を加えて効果を算出しているのですから，原理的には分ける前の状態，すなわちすべての個体を 1 つの群としてみたときの状態が基本だったわけです。ですから，その基本状態を b_0(全体平均) として，毎日水をあげる効果を b_1，時々あげる効果を b_2，あげない効果を b_3 として，同様に考えてみましょう。

そうすると，3 つの群のデータはそれぞれ次のように表現できることになります。

$$\begin{cases} y_i = b_0 + b_1 + e_i & \text{毎日水をあげる群} \\ y_i = b_0 + b_2 + e_i & \text{ときどきあげる群} \\ y_i = b_0 + b_3 + e_i & \text{あげない群} \end{cases}$$

ここで b_0 とあるのは，全体平均です。水のあげかたの違いになんの効果もなければ，すべて全体平均＋個体差で説明できてしまうはずだからです。よろしいでしょうか。

このようにすると，一次関数の式のように，X を使った表現をしやすくなりました。ここで X は $X = \{0, 1\}$ でその項について「該当する/当てはまる/割り当てられている (1)」か「該当しない/当てはまらない/割り当てられていない (0)」かだけを意味するのだ，というルールにします。たとえば i 番目のデータが「毎日水をあげる群」だったとすると，$X_{i1} = 1, X_{i2} = 0, X_{i3} = 0$ のように表します。すなわち，「水をあげる群」であり (1)，「ときどきあげる群」「あげない群」ではない (0) とするのです。これを踏まえて式を書き改めると，次のようになります。

$$\begin{cases} y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i & \text{毎日水をあげる群} \\ y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i & \text{ときどきあげる群} \\ y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i & \text{あげない群} \end{cases}$$

すべて同じ式になってしまいました！どの群も $y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i$ という 1 つの式で表現できていますから，非常に簡略化できましたね。またこうすると，群 j のところが繰り返されていますから，$\sum$ の記号を使って次のように書いてもいいかもしれません。

$$y_i = b_0 + \sum_{j=1}^{3} b_j X_{ij} + e_i$$

さらにさらに，b_0 だけ表に出ているのが美しくないですね！ここで $X_{i0} = 1$ と決めてしまうと，さらにスマートに表現できます。

$$y_i = \sum_{j=0}^{3} b_j X_{ij} + e_i$$

うん，スッキリ。

ところでこれはちょっと余談ですが，数字をまとめて表現する，**行列 (matrix form)** による表現を使うと，もっと表現が楽になります。i 番目のデータがどの群に割り当てられているかは，$\boldsymbol{X}_i = (1, 0, 0)$ のように数字のセット，**ベクトル**で表現されます。

ベクトルをセットで扱うのが**行列**と呼ばれるもので，これを使うとデータセットは次のように表現できます。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_0 \\ b_0 \\ \vdots \\ b_0 \end{pmatrix} + \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & b_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

ここで右辺の第二項は $b_{ij}X_{ij}$ の計算結果を表したもので，これを掛け算の式のまま表現すると，次のように書くことができます[*3]。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

こうした上で行列やベクトルを記号で置き換えると，$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$ となり，このとき $\boldsymbol{X}$ のことを (どの要因が割り当てられているかを表している行列という意味で) **デザイン行列 (design matrix)** といいます。

さてここで式をよくみてください。2 水準のときに比べると少し記号が増え過ぎな気がしませんか？ 統制群と実験群だけの 2 群のときは，統制群が b_0 で実験群が b_1，計 2 つだけだったのですが，3 水準になると b_0, b_1, b_2, b_3 と 4 つの係数を考えることになっています。なぜ係数が 3 つで済まなかったのでしょう。

実はこの式には，ひとつ大事な**制約**が欠けているのです。全体平均からの相対的な大きさで効果を考える場合，「相対的」ですから全体平均を中心にしなければなりません。言い換えるならば，効果だけを足し合わせると総和が 0 になるような制約が必要なのです。これを式で表現すると次のようになります。

$$b_1 + b_2 + b_3 = 0$$

この式を展開すると次のようになります。

$$b_3 = 0 - (b_1 + b_2) = -b_1 - b_2$$

これを踏まえて式 14.3 を書き直すと，次のようになります。

$$\begin{cases} y_i = b_0 + b_1 + e_i & \text{毎日水をあげる群} \\ y_i = b_0 + b_2 + e_i & \text{ときどきあげる群} \\ y_i = b_0 - b_1 - b_2 + e_i & \text{あげない群} \end{cases}$$

[*3] 行列とベクトルを習っていない人は，この掛け算どうなってるの？と思われるかもしれません。こうした行列の表現は線形代数と呼ばれるもので，心理統計のように変数が多くなったときの簡便的な表記法についての体系です。この授業の発展形である，データ解析応用の授業では行列を扱いますが，足し算掛け算から丁寧にやっていきますのでご心配なく。むしろ美しく表現できることにときめいてください。

こうすると，モデル式の係数は b_0, b_1, b_2 の 3 つになり，3 水準で係数が 3 つなので計算が合いますね。ちなみに 2 水準のときも全体平均を用いで表現するならば，次のようになります。

$$\begin{cases} y_i = b_0 + b_1 + e_i & \text{統制群} \\ y_i = b_0 - b_1 + e_i & \text{実験群} \end{cases}$$

このように表記すると，効果 b_1 の総和も 0，誤差 e_i の総和も 0 なので，二乗しないと計算できないなあということが理解しやすいかと思います。

ここで伝えたかったポイントは，「効果や誤差の大きさは相対的で，正負が同じ大きさだけ出るから足し合わせるとゼロになるから，大きさを評価するには二乗することで符号を消さなければならない」ということと，「効果の大きさを表す係数は，その要因の水準数 -1 個しか求められない」ということです。2 つ目のポイントはとくに，効果を評価するときに確率分布を応用する推測統計学の話に入ったとき，**自由度 (degrees of freedom)** というパラメータに繋がる話なので，先走って解説しておきました。また第 21 講あたりで出てくる話になりますので，それまで少し頭の片隅にでも引っ掛けておいていただければと思います。

14.4 分布を使って考える

さて，ここで前期の最後に少し発展的な要素も含んだお話をしておきましょう。

要因計画がどのようなものであれ，データの散らばりを生んでいるのは効果，個人差，偶然誤差によるものだと考えるのでした。逆に言えば，これらがなければすべての値は同じになるはずだ，という観点から逆算し，効果，個人差，偶然誤差の大きさを見積もっていったのでした。

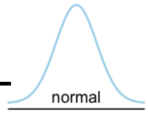
ここで，誤差を考える時には確率の考え方を利用してそれをコントロールする，という話を思い出してください。ガウスが考えた誤差論は，平均が 0，分散 σ^2 で散らばる正規分布に従うのでした。すなわち誤差を e_i とすると，次のように表現できます。

$$e_i \sim Normal(0, \sigma)$$

群間計画の場合は，各項目の値 y_i を切片 b_0 と効果 b_1 と誤差に分けます。

$$y_i = b_0 + b_1 + e_i$$
$$e_i \sim Normal(0, \sigma)$$

数式の中に図を入れるのは御法度かもしれませんが，イメージでいうなら次のような形です。

$$y_i = b_0 + b_1 +$$ 

e_i が正規分布に従うのですが，$b_0 + b_1$ のところは確率的に変わるものではないことに注意してください。ここが確率的に変わらないということをよく考えれば，数式に沿ってある値が足されることと同じです。確率分布にある数字を加えると，その分布の位置がずれますね。正規分布は位置と幅でその特徴が作られているのですが，右辺全体で見ると，$b_0 + b_1$ のところは，正規分布の位置をずらすことでもあります。

ずれた分布は，正規分布の平均値が $b_0 + b_1$ ですから，式全体を整えると次のようになります。

$$y_i \sim Normal(b_0 + b_1, \sigma)$$

この式のポイントは，$y =$ ではなく $y \sim$ になっているところです。データが正規分布に従う，という形に変わったのです。線形モデル (要因計画) はこのように，**確率分布の位置パラメータに数式 (モデル) を当てはめる**こととも言えます。

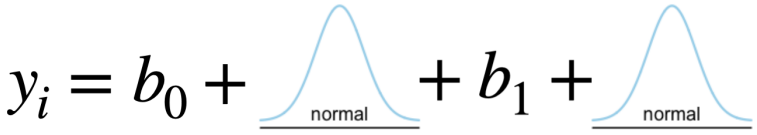
さらにさらに。群内計画の場合は，個人差 d_i も計算できるのでした。この個人差というのも，たくさん集めると確率的な傾向を表してくるはずで，とくに個々人の生得的な性質に関係するのであれば，それはまた正規分布に従うと言えるでしょう。今回の個人差も相対的に捉えるので，平均をひとまず ψ とでも置いて[*4]，これを踏まえて表現すると次のようになります。

$$d_i \sim Normal(\psi, \delta)$$

さてそうすると，群内計画の場合は 2 つの異なる正規分布が混ぜ合わさった形になっていることになります。式で書くと次のようになりますね。

$$y_{ij} = b_0 + d_i + b_{ij} + e_{ij}$$
$$d_i \sim N(\psi, \delta), e_{ij} \sim N(0, \sigma)$$

この式をイメージで表すと，次のような形になります。



このように確率をふまえたモデルで考えると，群内計画のデータには 2 つの分布が混ざり込んでいるので，**混合効果モデル (Mixed Effect Model)** と呼ぶことがあります。先ほどの群間計画では，$b_0 + b_1$ で平均値をずらした正規分布にデータが従うことになりましたが，群内計画では $b_0 + b_1 + d_i$ がすでに確率的に分布し，それが平均値となった確率分布がデータの従う分布になる，つまり (誤差の) 正規分布の中に (個人差の) 正規分布が入れ子になった形でモデルが表現されたことになります。分布が混ざったので**混合**というのですね。またより複雑なモデルの場合，この個人差を説明する他の変数や要因をつかって，個人差の分布もモデル化することになります。そのような場合，個人差に関係なく全体に及ぼす効果のことを**固定効果 (Fixed Effect)**，個人差とともに変わりうる効果のことを**変量効果 (Random Effect)** と呼んで区別することがあります。名前だけでも覚えておいてください。

さて本講義はこの後，線形モデルに含まれる分布のことを考えながら，回帰係数や効果の大きさ，誤差の大きさを推定・評価していく**推測統計学 (inferential statistics)** へと展開していきます。そのためにはデータがどのような分布をしているかを考え，理論的な分布の形に合わせて比較検討する，という手続きになります。具体的には「大体当てはまっている線形モデルの係数を考える」とか「線形モデルで計画された効果の大きさを推定する」「誤差の分布 (散らばり) の大きさと効果の分布の大きさを比べる」といった手続きです。基本的にはこうした線形モデルであることを知っておいていただければ十分です。

また，統計モデルを使った推論は，どんどん発展させていくことができます。要因が 1 つ，2 つと増えたように，混ぜ合わせる分布が複数あっても構わないのです。想像力を膨らませて，モデルにいろいろなパーツを組み立てることができるようになれば，心理学の研究や実験ももっとクリエイティブなものに感じられるでしょう。心理学の研究は，統計的な分析は，楽しいことなのですよ。

[*4] ψ はギリシア文字でプサイと読みます。大文字は Ψ であり，心理学 Psychology をギリシア語で表現した最初の文字になります。

14.5 課題

■Within 計画の分析例 本講で扱ったのと同じような, 一要因 3 水準 Within 計画のデータを表 14.4 のように用意しました。要因の効果, 個人差, 誤差の大きさ (平方和) を計算する練習をしてみましょう。

表 14.4 効果, 個人差, 誤差の計算練習

ID	Time1	Time2	Time3
A	48	48	40
B	49	51	45
C	58	59	46
D	46	44	37
E	52	52	46

第 15 章

前期のまとめ

これで前期のお話はいったん終わりです。これまでの話を振り返ってみましょう。

まず前期の授業では，「記述」「因果と相関」がメインの内容でした。

そもそも心理学ではなぜ数学が必要なのか，ということを冒頭でお話ししました。数学が受験科目にないので文系を選んだ，なかでも心という非常にウェットな，文学的な対象なので数学は出てこないだろうと思っていた，と学生さんにはよく言われます。しかし今や，文学でも法学でも歴史学でも，数学的なアプローチは避けては通れません。それはあらゆる科学がデータに基づく客観的な思考の積み重ねであるからです。数学的な考え方は無味乾燥で面白くないと思われるかもしれませんが，無味乾燥，すなわち主観的な色合いのないプレーンで非感情的な言語だからこそ，科学的に物事を考えるのに必要なことなのです。最近はデータサイエンティストの需要が高まっていますが，私たちが扱うデータは人間の反応であり，その奥にいる「ひと」「こころ」を知りたいという目的は微塵もぶれていませんし，データをただの数字としてしかみない (見れない) アプローチよりも，よっぽど社会に必要とされています。心理学をやりながらデータを扱う，数学的発想を持っていることは，今後の社会にとても魅力的な人材として映ると思いますので，自信をもってください。

みなさんは最近のあたらしい学校指導要領の方針で，記述統計の基本的なことについては既に学んできたことと思います。ヒストグラムや折れ線グラフ，箱髭図や散布図の描き方を学ぶというのは面白みを感じない科目だったかもしれませんが，そこに現れてくるのが人間の反応であり，心の現れであるとかんがえるなら，数字がただの意味のない羅列には見えてこないはずです。しかも，個々のデータではわからなかったことが，たくさん集めると何らかの傾向を持っている (正規分布や相関関係など) ということがわかって来れば，日々目の前の事例だけで振り回されるようなことがない，俯瞰的な視点を持つこともできるようになるのです。ワクワクしてきませんか。

そうした記述統計や可視化の方法も，いまは計算機を使って簡単に表現することができます。R 言語は簡単じゃないよ！と思うかもしれませんが，これまた無味乾燥な機械の言葉で指示通りの反応を機械的にしか返してこないことが，客観性や手続的正当性を担保するために必要なものなのです。大学では科学をやっていますので，こうしたお作法にもなれていただく必要があります。最初は苦労することも多いと思いますが，計算機という文房具をしっかり使いこなせるようになってください。

さて，途中で測定 (第 6 講) と確率 (第 7 講) の話を挟みましたが，その後は相関，因果，要因計画とより心理学的な研究実践に肉薄する内容で進んできました。そもそも目に見えないものを研究するというだけでも，難易度が高い学問なのです。なのに/だからこそ，データを取って統計や確率の力を借りながら，考えを進めていかなければなりません。みなさんはご自身の心を持っている (と思っている) でしょうから，考えたらわかるだろうと思うかもしれませんが，逆です。私たちは心があると思いがちですし，心が原因であると思いがちですし，実験して得られた結果もそれっぽかったら納得しがちなのです。私たちはそうした心の「自らを欺く作用」を掻い潜り，データを使って客観的に真実を見つけなければなりません。また個人差や偶然誤差など，物

にはない意志を持った研究対象特有のゆらぎがありますが，これも確率や要因計画法などを通じて，統制する方法を学ばなければなりません。

心理学というのが思った以上に厄介な学問である，というのが実感としてわかってきたのではないでしょうか。逆に言えば，ちょろっと考えたぐらいでわかる程度に，人間や心というものが簡単なはずがないんです。知識や技術をしっかり積み重ねて，「自分はこんなもんだ」「社会はこの程度なんだ」と自分を騙そうとする自分から自由になってください。

さて後期に入ると，この話がさらに複雑度を増していきます。ここまでは，目の前のデータが意味すること，データに含まれている情報を探り出すことを目指してやってきました。後期に入ると，目の前のデータを飛び越えて一般化することを考えます。手元のデータはこうなっているが，次のデータ，未来のデータはどのようになっているか？手元のデータはせいぜい身近な人から得た情報だが，社会は，世界はどうなっているか？とさらに発想を展開していく必要があります。その時に必要になってくるのが，またしても確率と統計なのです。

どうぞお楽しみに。

第 II 部

心理学データ解析基礎 1B

第 16 章

確率と統計的推論

さて、授業は後半のほうに入って行きます。確率と統計的推論というところからお話が始まります。

この授業は全体を通じてのテーマとして,「見えないものを見る」があります。たとえば潜在変数ですとか誤差のようなものも見えないものの代表格ですが、ここから始まる話は一部のデータから全体を推測するという統計的な推測の考え方についてのお話です。

このようにそもそも統計学というものが、不完全な情報からできるだけ多くの知識を得ようとする営みであることを知っておいてください。そして, 一見不可能に見えるようなこの営みを成立させるためには, 確率の力を借り, いくつかの仮定をおいて議論を進める必要があります。この「仮定」の上に成り立った “みなし” の学問であることは, ずっと忘れないでいてください。

16.1 母集団と標本

さてでは始めていきましょう。例として, ある小さな学校で身体測定をしたと考えて下さい。とても小さな学校で, 生徒の数はたったの 3 人です。3 人の身長はそれぞれ, 165cm,173cm,182cm でした。さて, この 3 人の平均身長は？

そう,

$$(165 + 173 + 182)/3 = 173.33\cdots$$

ですね。大体 173cm と結論づけましょう。

こうした代表値の計算 ($\rightarrow$ 第 5 講) は, 統計学の中でも**記述統計 (descriptive statistics)** と呼ばれる領域に属します。そこでは, 手元のデータの状態を記述することが目的なのでした。

ところで私たちがこうして何らかのデータを得たときに、手元のデータの状態だけではなくこれを使って何らかの推論をしたいということがあります。このままでは、ある地方の、ある学校の特徴がわかっただけだからです。しかもここは生徒数がたった 3 人ですから、特殊な例だと言われても仕方がありません。手元のデータを特殊な事例だといわれて困るのは, 逆に言えば一般的な特徴を知りたかったからですよね。一般性を確保するという事は、他の学校ではどうか、他の地方ではどうか、日本全体ではどうか、世界ではどうか、と言及できる対象を広げたいと考えているということです。

もちろん日本全体や世界全体の児童生徒の身長を測って回るというのはとても大変なことですから一朝一夕にできることではありません。だからといって完全に諦めてしまうのではなく、手元のデータをヒントにして、この全体像を推測しよう。これが**推測統計学 (inferential statistics)** の基本的な狙いです。

推測統計学の話を始めるにあたって、いくつかの言葉を整理しておきましょう。図 16.1 を見てください。

推測統計学では、私たちが知りたい世界全体のことを**母集団 (population)** といいます。たとえばあなたが日本の小学生の平均身長が知りたい、というのであれば母集団は日本の小学生全体です。また母集団の

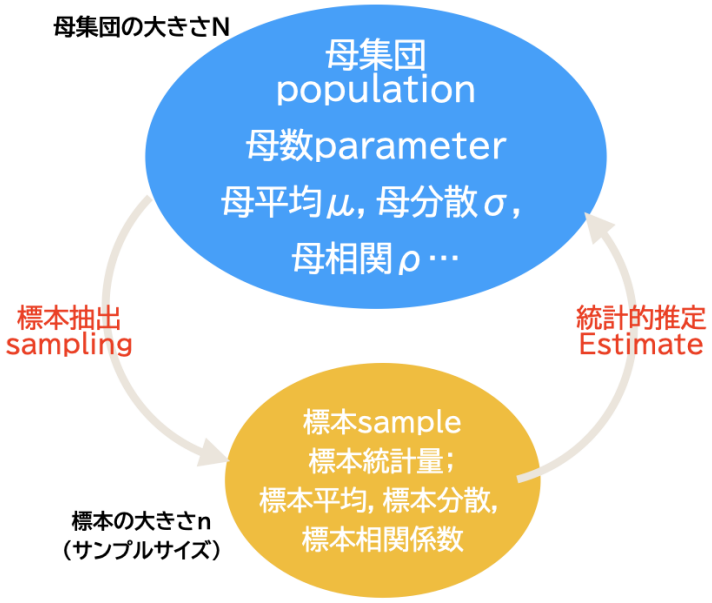


図 16.1 母集団と標本

統計的な特徴のことを，**母数 (parameter)** といいます。[*1]

母数として、たとえば母集団の平均のことであれば母平均、分散のことであれば母分散、相関係数のことであれば母相関、などと，「母」をつけて区別します。

母集団は非常に大きな集団を仮定することが一般的ですので、すべてを調査し尽くす事は非常に困難です。日本では 5 年に 1 度国勢調査が行われますが、これは日本という母集団全体のことを調べようとする調査です (悉皆調査といいます)。5 年に 1 度しか調査ができませんし、その結果をまとめるのにも数年かかります。私たち心理学者が、何らかの心理学的な特徴を知りたいと考える場合、母集団は日本を飛び越えて、世界全体あるいは人類全体を指すことになります。この全体の特徴を測ろうとすると，おそらく一生かかっても研究は終わりません。そこで全体の特徴を反映しているような、一部分を取り出してそれを詳しく調べることで、全体を考えるヒントにしようとします。この母集団から取り出されたもののことを、**標本 (sample)** と呼びます。また，取り出し方は**標本抽出 (sampling)** と呼び、一般に**無作為標本抽出 (random sampling)** をすることが必要です。無作為という言葉は聞きなれないかもしれませんが、要するに、作為がない、つまり意図がないという意味です。何らかの意図を持ったサンプリングというのはたとえば、背の高い人だけを連れてきてその平均身長を測り、全体を調べるヒントにしようというようなものです。そして計算された平均は、おそらく一般よりも高いものになるでしょう。だって背の高い人を “意図的に選んで” 連れてきたのですから。そうではなく，無作為に選び出すことが，全体の特徴を知るために重要なポイントになります。[*2]サンプリングは、お料理でいうところの味見のようなものです。中身をよくかき混ぜれば，ごく一部だけ取り出すことでも、十分全体の味を推測できるわけです。

[*1] 最近巷では、母数という言葉を，調査した全体を指す意味で用いられている誤用が散見されます。たとえば「データが少ないから，もう少し母数を増やさなければ」といった用法です。本来の母数という言葉を知っていれば，これはまったく意味の通じない表現です。おそらく，「分母」という言葉と混同しているのだと思われます。

[*2] 社会調査などでは，この標本抽出の方法について，どの程度の母集団に対し，ある精度で平均値を推定するためには，何人のサンプルをとってくる必要があるか，といった設計や技法が駆使されます。社会調査の方法については心理統計学の領域を超えますので，他書に譲ります。

さて母集団におけるパラメーターと区別するために、標本における統計量の事は、標本平均、標本分散、標本相関係数、などと標本という言葉を前につけておきます。これらの数値のことを**標本統計量 (sample statistics)** といいます。平均や分散, 標準偏差, 相関係数など、これまで標本に対して計算してきた記述統計量は、すべて標本における統計量なので, このように呼んで区別しておきます。この標本の数値や統計をヒントにして、母数がどうなっているかに言及することを, **統計的推定 (statistical inferenece)** と呼びます。

統計学では一般に, 母数は未知であり, ギリシア文字で表します。母平均は μ (ミューと読む), 母分散は σ^2 (シグマの二乗, と読む。総和を表すシグマ $\sum$ の小文字です), 母相関は ρ(ローと読む) などがよく使われます。これに対して, 標本統計量は計算できる数字であり, 一般にアルファベットを用いて表現します。標本平均は $\bar{X}$, 標本相関係数は r_{xy} などです。

ここで誤解しやすい 2 つの言葉について解説を追加します。1 つは標本の大きさと数についてのお話です。図 16.2 を見てください。さきほどのとある学校での標本についての話ですが、そこでは 3 年の児童生徒を取り出したのでした。この時サンプルの大きさ、つまり標本に含まれる人数 (観測度数) は**サンプルサイズ (sample size)** と呼ばれます。このサイズの調査を他でも行ったとしましょう。つまり A 校, B 校, C 校... といろいろな学校で 3 人ずつ調査を繰り返していくのです。この時, 得られた標本の数は**サンプル数 (Number of samples)** といいます。誤用がおおいのは, このサンプルサイズとサンプル数の取り違えです。「人数 n のサンプルを取りました」というべきところを, 「サンプル数 n です」というと, 「で, 何人取ったの？ 」とか「そんなに繰り返し調査したの？ 」という返事がかえってきて, 混乱することにもなりますので注意してください。

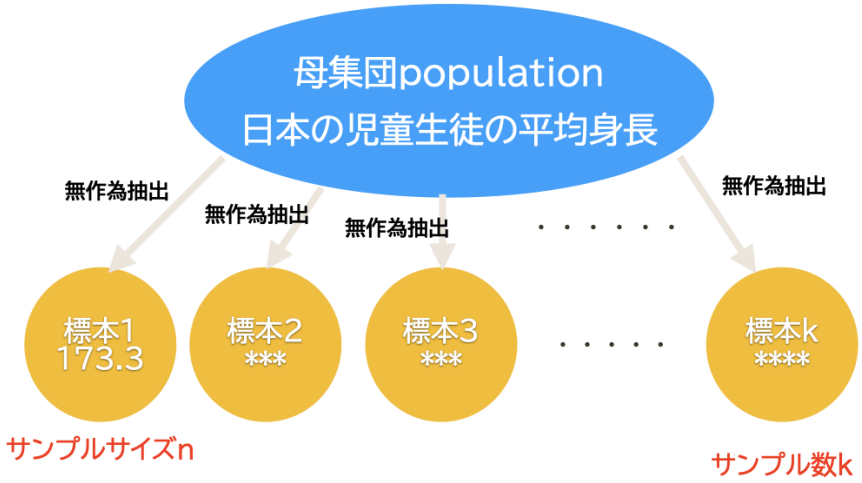


図 16.2 サンプルサイズとサンプル数

もう 1 つ。冒頭の例では, 平均身長が 173cm だったわけです。これはもちろん, 違う標本を取れば (i.e. 違う学校の 3 人を無作為に取ってくる), 違う数字になるでしょう。標本統計量はこのように, 標本によって変わる変数なのです。メカニズムはわからないけどゆらぎがある, そう, 確率的に変化するものと考えられます。確率的に変わる変数は, **確率変数 (random variable)** と呼ばれますが, 標本統計量は確率変数なのです。このとき, 173 という数字は, 確率変数がたまたま今回定まった値をとったことになりますね。そこでこの数字

は，確率変数の**実現値 (observed value)**，あるいは**標本統計量の実現値**と呼ぶことにします。標本統計量そのものは確率的に変動するものであり，それが具現化したものとは区別しておきましょう。[*3]

16.2 母集団分布とモデル

さて用語の準備が整ったところで，実際にどのように推定していこうか，考えていきたいと思います。さきほどの標本では，平均身長 (標本平均の実現値) が 173 だということがわかりました。であれば母平均も 173 だろう，と考えて良いでしょうか。少し性急すぎるような気がしませんか。違う標本を取れば違う数字になるでしょうから，1 つの標本から答えを出すのは少し無理があるようです。

そこで，母集団の数字 (**母集団分布 (population distribution)**) がどのような分布をしているかを考えましょう。これはもう，仮定の話です。だって実際に母集団の数字を全部見ることは叶わないわけですから。たとえば身長に影響してきそうな変数は無数に考えられます。たとえばご両親の身長，生育環境，日々の睡眠時間，食事の量，運動の量，etc... 何もかもが，影響しているように思えます。さて，これら無数の条件がいろいろ積み重なって，1 つの数字となったのであれば，そうした数字を集めると正規分布に従うのではないか，と考えることができそうです。このように，仮定ではありますが，合理的な論拠があれば母集団の分布の形は正規分布である，と考えて話を進めても良さそうです。実際，心理学の多くの研究は，データの母集団分布に正規分布を仮定しています。心という目に見えないものも，無数の影響に晒されて出来上がるものと考えているからです。

重要なのは，母集団分布がどのような分布か，あるいは心的反応の結果として得られる数字の分布が，どのようなものであるかはすべて「モデル」であり[*4]，「仮定」に過ぎない，ということです。反応の結果として得られる数字 (アウトカム変数ともいいます) が，「はい」「いいえ」「どちらでもない」のような三択であれば，これの母集団分布は正規分布のような連続分布ではなく，離散的な分布を仮定しなければならないでしょう。

心理学は見えないものを見ようとする学問であり，統計は一部の不完全なデータから全体を見ようとする学問です。この一見不可能とも思えるような推論を進めるにあたって，仮定の力を借りて (モデルを置いて) いるのだということは，忘れてはいけません。

正規分布の意味

ここで母集団に正規分布を仮定する，という話をしました。この後も，正規分布はあちこちで顔を出します。心理学において (それ以外の統計モデルにおいても)，正規分布はとても重要な意味を持つ分布です。

この授業で前期の間は，正規分布 1 つは誤差の分布として導入されました。偶然誤差はどの程度散らばるかということがわかれば，それを見込んで評価できます。データは真の値 + 誤差 ($X = t + e$) という**古典的テスト理論**はこの特徴を利用しているのでしたね。

そしてここで，集団全体の傾向を表す分布として正規分布を使うことになりました。推測統計ではとくに平均がどのあたりにあるのかに注目し，その傾向を掴むといった統計的な使い方をします。平均身長や平均寿命を考えるのはこういったところにあるわけですね。この時は個人の特徴を平均からの逸脱の程度だという考えているといえます。この考え方は心理学全域に関わるモデルです。平均的な傾向を探ったり比較したりすることで，人間全体の傾向を考えようというわけです。もちろん臨床心理学のように，全体の傾向ではなく個人内の変動を考えるという研究アプローチもありますが。

このように正規分布には大きく分けると 2 つの意味があるのです。第一の意味は偶然誤差の分布。平均が

[*3] サイコロの目は確率変数ですが，⚀や⚄は確率変数の実現値です。

[*4] モデルとは，雛形や模型という意味もありますが，ここでは些末な事例を省略して作られた，抽象的な理想型という意味です。

0 で分散が σ^2 の正規分布に誤差が従うと考えます。第二の意味は集団全体の傾向です。個々人に影響するさまざまな要因を特定することはできませんが，集団全体としては 1 つの傾向を持つ，つまりある平均値があり，そこから離れるほどその出現度数が少なくなるというパターン全体を表しています。前者の立場からは測定値=真値 + 誤差と考えていることになりますし，後者の立場では測定値=期待値 + 個人差と考えていることになります (期待値とは確率分布の平均のことです。)。

元々はガウス流の，誤差分布の流れから見つけ出された正規分布ではありますが，社会調査が始まった頃に社会全体の特徴分布として正規分布が改めて見直されたといえるでしょう。椎名 (2013) は他にも正規分布が見つけられた経緯として，7 種類のルートがあると論じています。

16.3 標本分布と標準誤差

母集団に分布を仮定すると標本統計量がどのような分布をするかが，数理的に導出できます。この計算はとても専門的ですので，興味がある方は「数理統計学」に関する他書をあたっていただくとして，そのエッセンスだけいただくことにしましょう。

そのエッセンスとはつまり，母集団に分布を仮定すると，標本統計量の分布がわかるということです。正確に言えば，標本統計量がどのような確率分布に従うのかがわかる，という点です。標本統計量がどのような確率分布に従うのかがわかれば，標本ごとにどの程度数字がブレるのか (分散・標準偏差)，標本分布の平均はどのあたりにあるのか (期待値) がわかります。であれば，1 回の調査＝標本が 1 つであっても，この標本の数字がどの程度ブレていくのかがわかる，つまり実現値の精度を評価できるということになります。ありがたいことに，母集団からいくつもの標本を取り出してそれぞれの標本平均を計算すると，その分布が正規分布になる，ということがわかっています[*5]。具体例で見ていきましょう。

表 16.1 100 人村のテスト分布

69	63	58	54	55	44	63	55	50	53
58	38	43	59	52	58	34	52	55	61
37	47	41	41	61	53	43	39	53	74
62	45	39	44	65	39	47	63	44	64
50	55	52	66	47	74	49	70	57	63
50	56	40	44	43	52	38	31	58	51
22	72	60	52	61	45	48	57	59	54
62	52	45	56	49	50	46	63	40	43
57	65	41	37	35	30	50	63	42	51
51	70	62	36	45	59	45	37	43	48

ここに表 16.1 を用意しました。10 行 10 列なので 100 の数字が入っています。100 人のテストの成績だ，とでも思っておいてください。さてこの 100 人が母集団だとします (世界が 100 人の村だったら，みたいに)。母平均はこの 100 個の平均ですから，計算してみますと $\mu = 51.24$ でした。

さて，これは仮想データなので，母集団分布は正規分布であることがわかっているとします[*6]。100 点しかデータポイントがありませんから，ヒストグラムもあまり正規分布っぽくはなりませんが，一応図にすると次のようになります (図 16.3)。

[*5] この性質は**中心極限定理 (central limit theorem)** として知られています。

[*6] これは R を使って `rnorm(100,50,10)` として発生させた乱数です。

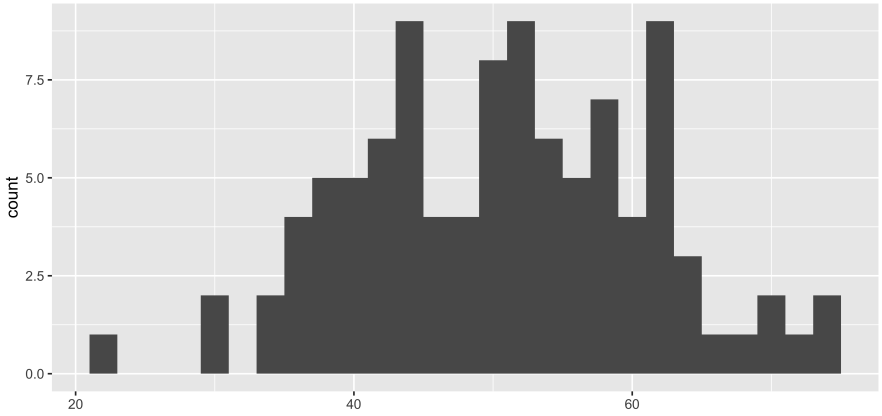


図 16.3 100 人村のデータ

ここから，無作為にサンプルサイズ 5 のデータをとってみたところ，次のような数字が取れたとしましょう[*7]。

$$X_1 = \{57, 43, 47, 52, 63\}$$

これの平均は (標本平均は)$\bar{X}_1 = 52.4$ です。母平均が $\mu = 51.24$ ですから，結構近いですね。
次に，もう一度同じ母集団から，同じサンプルサイズのデータを取りましょう。

$$X_2 = \{63, 61, 52, 31, 62\}$$

これの平均は $\bar{X}_2 = 53.8$ でした。母平均からみると，少し大きいほうにずれてしまいました。
次にもう一度‥‥というのを繰り返していくと，表 16.2 のようになりました。

表 16.2 サンプルサイズ 5 のサンプリングを繰り返す

サンプル番号	標本平均
1	52.40
2	53.80
3	49.40
4	55.00
5	50.20
6	44.40
7	55.00
8	50.80
9	54.00
10	52.80

これを見るとわかるように，標本平均は標本を取るごとに変わります。分布するのです。標本を取るごとに

[*7] 無作為にサンプリングするのは非常に難しいです。「ぱっとみて」「適当に」やっても，人間がやると何らかのバイアスがかかって完全に無作為にすることはできません。ここでは 10 面サイコロを 2 つ振って行と列の番号を選び，数字をピックアップした‥‥とでも思っておいてください。実際は，R にある無作為抽出関数 `sample` を使っています。

数値が変わりますので，母平均に近い数字を取ることもあれば，母平均から離れた数字になることもあります。ところが，この標本平均**の平均**を取ると，$\bar{\bar{X}}_i = 51.78$ になります。母平均は $\mu = 51.24$ ですから，個々のどの標本平均よりもぐっと母平均，つまり推定したい正解に近づいています。

今度は，同じ母集団からサンプルサイズ 10 でサンプリングするとしましょう。

$$X_1 = \{50, 39, 65, 66, 44, 50, 36, 54, 70, 74\}$$

こちらの平均は $\bar{X}_1 = 54.8$ です。これは母平均には程遠い数字ですが，頑張って次々サンプリングを繰り返したとします。10 回分の結果が表 16.3 です。

表 16.3 サンプルサイズ 10 のサンプリングを繰り返す

サンプル番号	標本平均
1	54.80
2	49.40
3	51.90
4	49.80
5	47.40
6	54.10
7	50.50
8	52.80
9	54.60
10	47.70

この時の標本平均の平均は，$\bar{\bar{X}} = 51.3$ で，先ほどのサンプルサイズ 5 の繰り返しよりも母平均 $\mu = 51.24$ に近くなりました。ふむぅ。

さて，これらの実験から何がわかるでしょうか。まずは標本平均は分布するということです。取るたびに数字が変わるわけです。また，数理統計学の知見から，標本の平均は正規分布に従うことがわかっているのでした。つまり表 16.2 にある標本平均 (実現値 51.60,49.00.50.60,.... など) は平均を中心として左右対称の単峰分布だということです。また，平均から離れるほど，その出現確率はどんどん小さくなるはずです。

標本分布が正規分布になるわけですが，標本分布の期待値＝平均値は，究極的には (サンプル数を無限に増やしていけば)，母平均に一致することがわかっています。標本平均の平均は母平均に一致することをとくに**大数の法則**といいますが，これが推測統計学の基本です。つまり，標本をたくさん取ることで母平均を推定できるようになるということです！

このことをふまえて，表 16.2 と表 16.3 にある 2 つのサンプルサイズ ($n = 5$ と $n = 10$) を比較すると，次のことがわかります。まず，サンプルサイズが大きい方が，母平均に近い数字が出やすそうだ，ということです。これはデータをたくさんとっているからで，推測したい母集団のヒントがたくさん手に入るからです。直感的にもわかりやすいでしょう。今回も，サンプルサイズ 5 の時よりも 10 の時の方が，標本平均の平均が母平均に近くなりました。

また，サンプルサイズを大きくすると，標本平均のばらつきは小さくなります。多くのデータをとっているのですから，極端な値に振り回される可能性が減るからです。

この標本平均のばらつきのことを表す言葉があります。標本平均は正規分布に従うのですから，この幅，つまり標準偏差がそのばらつきを表すことになります。標本平均**の標準偏差**はとくに**標準誤差 (standard**

error) と呼ばれ，推定値としての標本平均の精度を表す数字になります。今回，表 16.2 にあるサンプルサイズ 5 の場合の標準誤差は 2.92，同じく表 16.3 にあるサンプルサイズ 10 では 2.48 です。サンプルサイズを大きくすると，標準誤差が小さくなる，つまり測定の精度が上がるわけですね。

具体的には，標本分布の分散は母分散 σ^2 をサンプルサイズ n で割ったものに一致することがわかっています。標本分布の標準偏差，つまり標準誤差は，$\frac{\sigma}{\sqrt{n}}$ に一致することになります。分母にサンプルサイズが入っているので，これが大きくなると標準誤差が小さくなることがわかります (図 16.4)。

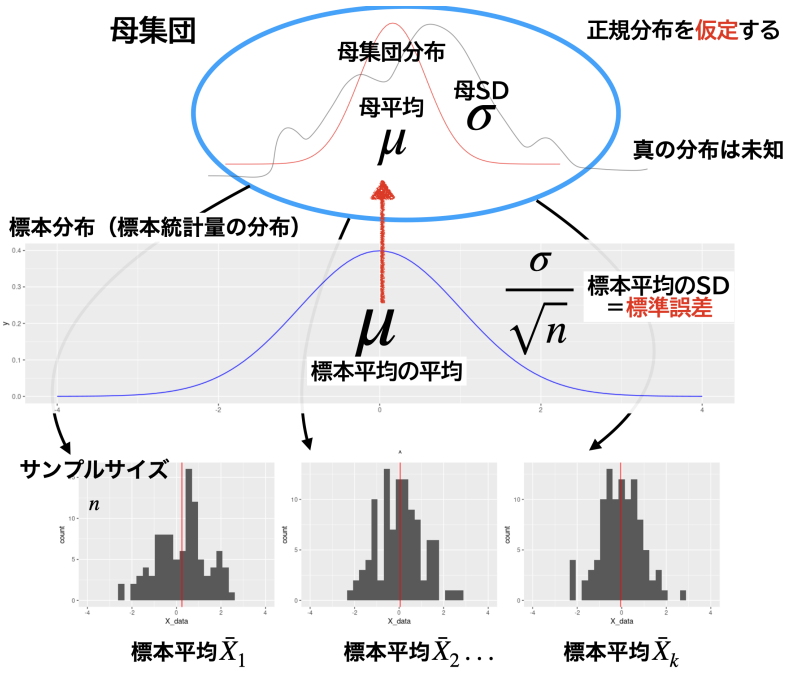


図 16.4　標本，標本統計量，標本分布と統計的推測。母集団からサイズ n の標本を得る。個々のデータ (標本 1 の中にいる人物フィギュア) は，母集団と標本の枠組みの中で捉える。標本統計量は標本をたくさん取ると $(1, 2, \cdots k$, このときサンプル数 $k)$, 毎回異なる値を出すが，標本分布に従っている。標本分布の期待値は母平均に一致するし，ある標本平均の精度は標準誤差で評価できる。

このことを数値シミュレーションでみてみましょう。先ほどと同じように，仮想的な数字を使って検討します。今度は一気に 100 万点のデータを作り，そこからランダムサンプリングをするようにしました。こうすることで，複数回無作為抽出しても，たまたま同じものがかぶる (再サンプリングされてしまう) ような危険性を回避できます。さて，図 16.5 をみてください。まず，100 万点のデータセットから無作為にサンプルサイズ 3,5,10,50 の標本を 100 回 (サンプル数 100) とりました。それらの標本平均も 100 あるわけですが，標本 1 の標本平均，標本 1 と 2 の平均，標本 1,2,3 の平均･･･と標本平均を 1 つずつ増やしていったときに，母集団の平均 μ(青の点線) に本当に近づいていくかどうかを検証したものです。図 16.5 からわかるように，サンプルサイズが小さくても 100 回もとれば母平均に一致していくようです。ただ，この折れ線の最初の頃をみてみると，サンプルサイズが小さい(3 や 5) ときは，大きく変動しています。サンプルサイズが大きくなると(n=50)，最初から母平均近くの値が得られているわけです。データをたくさん取ることの利点は，標準誤差が小さくなることにあります。

ここまでのことをまとめておきます。ポイントは，次の 2 つです。

- 標本平均の標本分布の平均 $\bar{\bar{X}} = \mu$
- 標本分布の標準偏差は $S.E. = \frac{\sigma}{\sqrt{n}}$

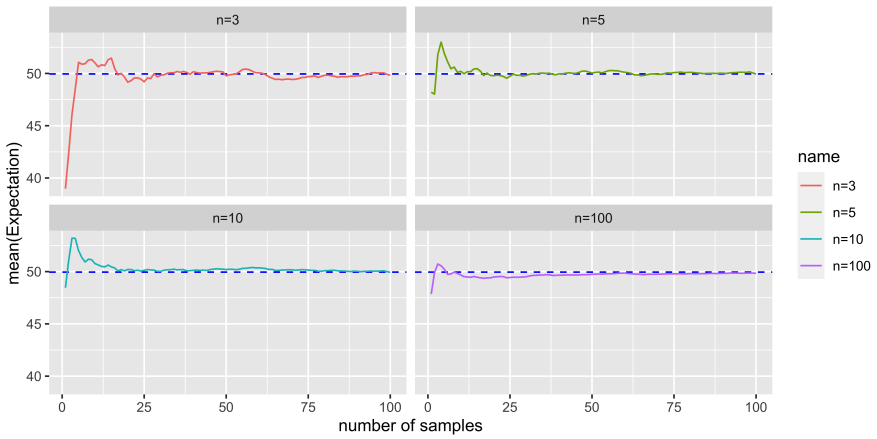


図 16.5 サンプリングを繰り返し，標本平均の平均を求めると，母平均に近づく

16.4 まとめ

今回は推測統計学の入り口でした。まずは用語として，母集団，標本，サンプルサイズ，サンプル数，母数，サンプリング，推定，標本分布といった言葉の整理を行いました。ちょっとした言葉の違い，ちょっとした数字の違いでも意味や結果が変わってくるので，慎重に読みこなしてください。

次に，母集団に確率分布を仮定することで，一部の標本から母集団のことを推測できる，という基本的なロジックをご紹介しました。推測統計学は，目に見えないものを想定して，仮定に仮定を重ねて推論していくことになります。仮定が多くてなにかいやだな，と思う人もいるかもしれませんが，そもそも全体像がわからないものを少数のヒントで推定するわけですから，無理な問題設定なのです。それをなんとか解決するためには，いくつかの仮定をおかざるを得ないのです。これらのお話は，あくまでも正規分布のような仮定・モデルの上での話であることは，いつまでも肝に銘じておいてください (それをいうと心理学そのものも，モデル上の話なんですけど！)。

ところで，最後に紹介した標本統計量をつかった母数の推定は，**モーメント法 (moment method)** とよばれるもので，実は他にも推定する方法があります。ただ，いずれにせよ確率分布は活用することになりますので，確率的な表現にも慣れておく必要があるでしょう。

16.5 課題

■前期と後期の学ぶ内容の違い 後期に入ってから推測統計学の領域に入ります。なぜ推測統計学を学ぶのか，前期で学んだ内容とどう変わってくるのかを，自分の言葉で説明してみてください (400-800 字)

第 17 章

推論と判断

前回は「モーメント法」という，標本の記述統計量をつかって母数を推定する方法をお話ししました。さて今回は，この話の続きをしていきましょう。どうやって母数を推定していくかについて，心理学でよく使われる正規分布を例に話を進めていきます。また，心理学の目的としては「母数がどういう数字だったか」ということよりも，相対的な比較によって考えることがあります。この考え方は，ひいてはモデル比較という話になっていきます。

17.1 不偏推定量

ここまでの話で，標本平均の平均が母平均に一致し，そのばらつきもわかるのですから，標本から母集団を推測できそうだ，ということが見えてきましたかと思います。標本平均は多少のズレを含んでいるかもしれませんが，母平均の推定値として使うことができそうです。3 人学校の身長の例で言えば，母平均も 173cm ぐらいかもしれない，という予測ができそうだということですね。もちろん，サンプルサイズが小さいので，173± 何センチかのズレはありそうですが。

ところで，「母平均の推定値は XXX だ！」と一点張りで推定することを**点推定 (point estimation)** といいます。ただ，これはなかなかのギャンブルで，正解 (母平均) 近くにはあるかもしれませんが，正解そのものにピタリと当たることはなさそうです。とくに連続変数の場合は，小数点下何桁までぴったりと当たる，というのはほぼなさそうですから。そこで，「何センチかのズレ」を予測に盛り込むことが考えられます。これが**区間推定 (interval estimation)** です。母平均は 173cm だ，と言い切ってしまうのではなくて，160cm から 180cm の間に入ってるんじゃないの，というような予測の仕方をするのです。こうすると，外れる可能性は低くなります。ただ，0cm 以上 500cm 未満だったら 100% 当てられるぜ，という推定は何も推定していないことと実質同じですから，どれぐらいの幅で，どれぐらいの確率で･･･ということを考える必要があります。たとえば，173cm±5cm の区間に入ってる確率が 95%，といった具合に。

このように婉曲的な，少し保険をかけた推定の仕方はうまくいきそうですが，まだ問題が残っています。この幅を表現するためには，標本平均がどれぐらいブレるのか，その幅を推定しなければなりません。そう，標準誤差が必要なのです。ただこの標本平均のばらつき (標準誤差) は母分散によるので，母分散がどれぐらいになるのかを考えなければなりません。

母分散の推定値として，平均の時と同じように，標本分散をあてがうというのは 1 つのアイデアです。ですが，残念ながらこの方法だとうまく推定できません。図 17.1 をみてください。これは標本平均の時と同様に，サンプルサイズ 10 のサンプリングを 200 回繰り返し，標本分散の平均を順次とって行ったものです。これをみると･･･あれあれ，真値 (母分散 σ^2) の青いラインの，下側に収束して行ってるようですね。このズレのことをバイアスと言います。標本平均値は偏りのない推定量 (不偏推定量) なのですが，分散は偏り (バイアス)

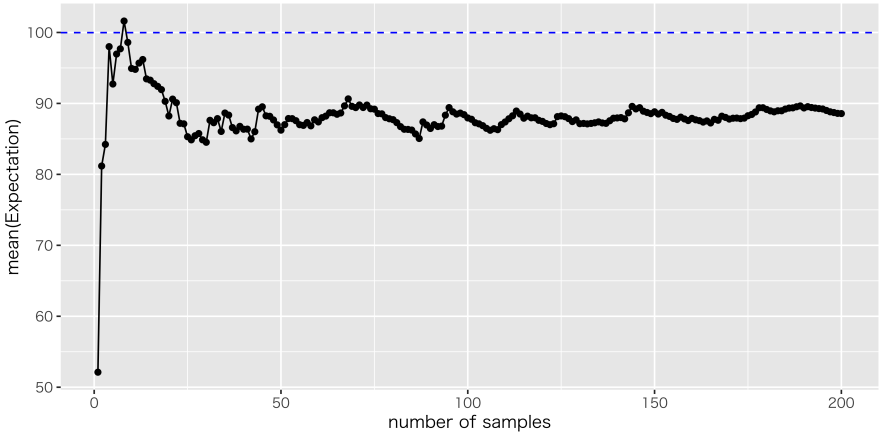


図 17.1 サンプリングを繰り返し，標本分散の平均を求めると，母分散から少しずれる

があるのです。

実は，標本分散の期待値 (平均を取ったもの) は，母分散に一致しないことがわかっています。ただ，このズレの加減は一定しているようですね。実際，これは計算できて (数理統計学の中身に入り込むのは避けて，答だけもらってきますが)，ズレの程度は $\frac{\sigma^2}{n}$ であることがわかっています。

ズレの程度がわかっているのなら，計算が合うように最初から計算を工夫しておけばいいでしょう。標本分散の計算式を覚えているでしょうか？ こんな式でしたね。

$$s_X^2 = \frac{1}{n}\sum(X_i - \bar{X})^2 \tag{17.1}$$

これが標本分布の期待値[*1]になると，母分散から $\frac{\sigma^2}{n}$ だけずれるのですから，母分散 σ^2 の推定値にするためには，事前にこのズレのことを考慮したいと思います。式を展開しましょう。

$$\begin{aligned}
\sigma^2 - E(s_X^2) &= \frac{\sigma^2}{n} \\
\sigma^2 - \frac{\sigma^2}{n} &= E(s_X^2) \\
n\sigma^2 - \sigma^2 &= nE(s_X^2) \\
\sigma^2(n-1) &= nE(s_X^2) \\
\sigma^2 &= \frac{n}{n-1}E(s_X^2)
\end{aligned}$$

こうなりますので，先に推定値としてこの計算を先にすることにします。

$$\hat{\sigma}^2 = \frac{n}{n-1}\frac{1}{n}\sum(X_i - \bar{X})^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 \tag{17.2}$$

ということで，事前にこのずれが起きないように補正できました。

記述統計としての分散 (標本分散) はサンプルサイズ n で割り，推測統計量としての分散推定値 $\hat{\sigma^2}$ はサンプルサイズより少し小さな，$n-1$ で割るのだ，と考えれば良いでしょう。後者のことはバイアスのない標本統計量ということで，**不偏分散 (unbiased variance)** と呼びます。

先ほどのシミュレーションを，不偏分散の式を使って行ったのが図 17.2 です。補正してある不偏分散だと，母分散に近づいているのがわかりますね。

[*1] 確率変数の平均値のことを期待値というのでした。忘れてた人はセクション 7.2.4,Pp.111 を参照してください。

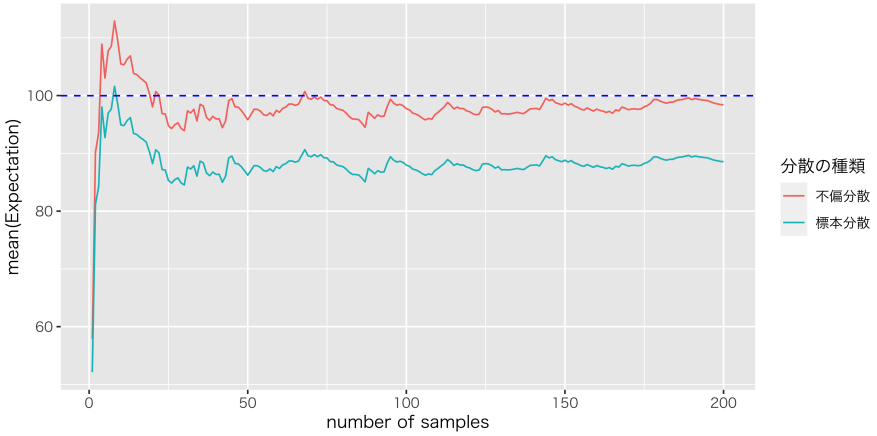


図 17.2 サンプリングを繰り返し，標本分散と不偏分散の期待値を求めたもの

17.2 標本平均の区間推定

以前，正規分布を利用した推定のお話しをしました (セクション 7.6,Pp.123 を参照してください)。あの話をもう一度思い出してみましょう。あるスコア x が正規分布に従っているとします。これは次のように表記されます[*2]。

$$x \sim N(\mu, \sigma)$$

この変数 x を標準化しておくと，標準正規分布に従うことになります。つまり，

$$z = \frac{x - \mu}{\sigma} \tag{17.3}$$

$$z \sim N(0, 1)$$

ですから，標準正規分布の表などを参考にすると相対的な位置が推測できる，という話でした。

これを母集団と標本という枠組みの中で考え直してみたいと思います (図 17.3)。先ほどの話は，あるスコアが平均 μ, 標準偏差 σ の正規分布に従っている，というところから始まりました。これ，ギリシア文字で書いてありますから，母数ですよね。母数がわかっていれば苦労しないのですが。

そこで，μ の代わりに標本平均 $\bar{X}$ を使って – 標本平均は母平均の推定値として悪くないはずですから – 考えてみたいと思います。この標本平均 $\bar{X}$ は標本を取る度に変わります。どの程度変わるか，つまりどの程度標本平均が散らばるかを表すのが，標本平均の分散であり，標本平均の標準偏差でした。標本平均の分散は，$\frac{\sigma^2}{n}$ であり，標準偏差 (標準誤差) はその平方根ですから $\frac{\sigma}{\sqrt{n}}$ で表されます。つまり，

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

より，これを標準化して，

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

[*2] 正規分布は Normal Curve なので，Normal, あるいは単に略してここにあるように N, とだけかかれます。正規分布は平均と標準偏差という 2 つのパラメータを持ちますが，統計の本によっては，平均と分散で書いてある場合があります。分散は標準偏差の正の平方根なので (→ 第 5 講)，持っている情報としては同じですから，単に表記の問題です。本書では統計環境 R と表記を合わせるために，標準偏差で記載しています。

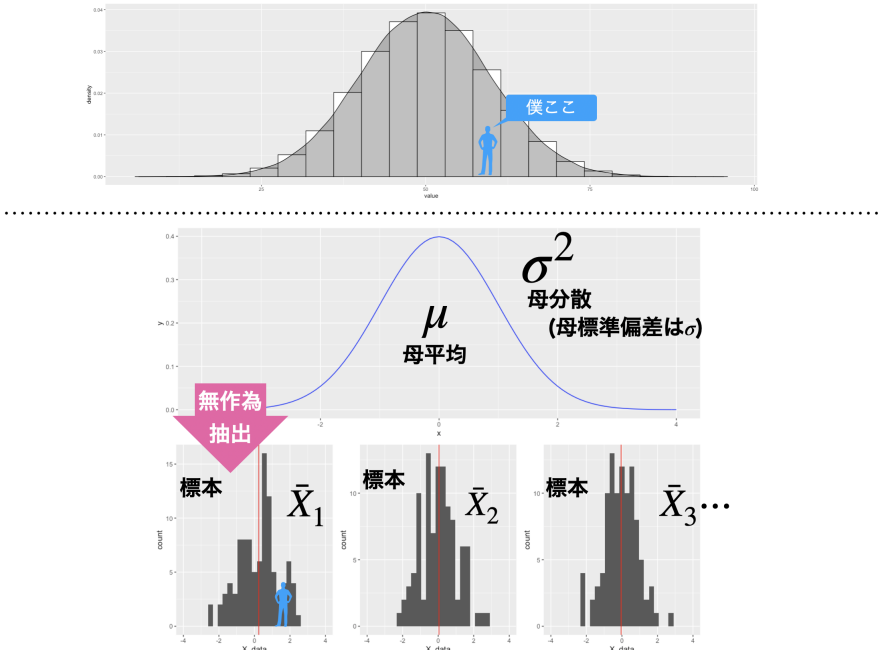


図 17.3　推測統計学の枠組みで考えることに注意。以前 (セクション 7.6) は上の図のように，推測するのではなくて母集団そのものの中に一個人のスコアを位置付けて，相対的な位置を考えよう，という話でした。今回は下のように，母集団が正規分布だと仮定して，そこから得られた標本のことについて考えています。個々のデータは標本の中にあります。また今回考えるのは標本統計量で，それが標本分布に従うのでした。

$$Z \sim N(0, 1)$$

となります。この Z は標準正規分布に従いますから，母平均がどのあたりにあるかを正規分布を使って推測できます (図 17.4)。

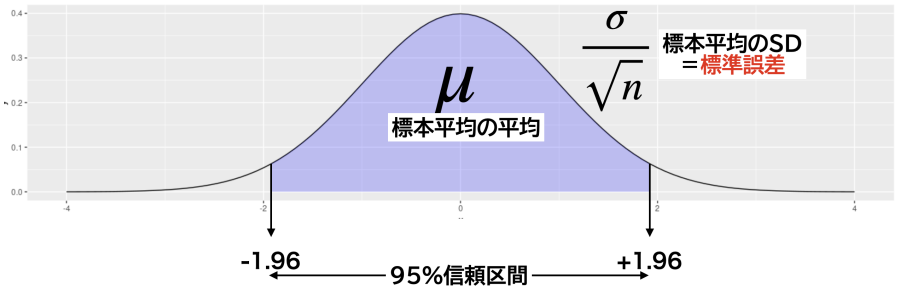


図 17.4　標本統計量の従う分布も正規分布。幅は母分散とサンプルサイズに依存します。

たとえば 95% の幅を持って推測するとしましょう[*3]。標準正規分布で 95% の幅を持つ区間はおおよそ

[*3] 心理学を始め，推測統計の業界ではよくこの 95%，あるいはその逆の 5% という数字が使われます。これにはなんの根拠もありません。50% や 30% の幅で考えたって，全然問題ありません。

±1.96 ですから[*4]，

$$-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

となります。この式を変形していきます。

$$-1.96\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}$$

真ん中を μ だけにしますと，

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

となります。

具体的な数字をあてがって考えてみましょうか。母平均が 50，母標準偏差が 10 の母集団から，サンプルサイズ $n = 9$ のデータをとってきたところ，次のようになったとしましょう。

$$X = \{55.22, 50.10, 45.59, 61.99, 48.83, 50.38, 61.95, 53.44, 46.71\}$$

これの平均を考えると，$\bar{X} = 52.69$ となります。ただ，違う標本を取れば違う平均値になるはずなので，母平均 μ の推定値を 52.69 と一点張り (点推定) するのは外れてしまうかもしれないですね。実際，母平均 50 の設定だったので，点推定による推測は外れているわけです。ただ，標本統計量の散らばりが

$$\frac{10}{\sqrt{9}} = 3.33$$

ですから，

$$52.69 - 1.96 \times 3.33 < \mu < 52.69 + 1.96 \times 3.33$$

すなわち

$$46.1632 < \mu < 59.2168$$

となります。これを平易な言葉でいうと，「母平均は 46.1632 から 59.2168 の間に入っている確率が 95%」ということになります。この区間の中には設定した母平均 $\mu = 50$ が含まれていますから，区間推定をすることによって推測は当たった，と言えるわけです。

これが**区間推定 (interval estimation)** の実際です。ある値がはいる範囲を確率で表現するというのは，随分もってまわった言い回しのように思えますが，そもそも少ない標本から全体を推定しようとしている，という難しい問題をやっているのでこういう言い回しになります。少しずつ慣れて行ってください。ちなみにこの区間のことを**信頼区間 (Confidential Intervals)** といい，CI[46.1632,59.2168] と書くことがあります。

また，これには母数が定数であるという隠れた考え方があります。母集団が推測したいデータの全体であれば，母数を知ることは絶望的に難しいですが，それでもなんらかの 1 つの値のはずです。であれば，この 95% の区間の中に「入っているか，入っていないか」の**どちらか**であり，入っている確率が 95%，という言い方であることに注意してください[*5]。

[*4] 以前，R で計算した課題があったことを思い出してください。確率なので全部の面積は 1.0，95% 区間は両端 2.5%=0.025 を除いたことになりますから，`qnorm(0.975,0,1)` より，1.959964 です。左右対象なので，下側は符号を反転させれば OK です。

[*5] 最近の統計学の動向の中には，標本が定数で母数が確率変数であり，パラメータは定数ではないという考え方や (ベイズ理論的立場)，標本は確率変数で母数も確率変数であるという考えかた (情報統計学的ベイズ理論の立場) などがあります。テキストによっては立場によって，母数の区間推定についての説明が微妙に違ったりしますが，あまり気にしなくてよろしい。

17.3 標本だけから推測したい

さて，式 17.3 を使って区間推定できることがわかりましたが，よく見るとまだギリシア文字＝未知数が入っていることに気づきます。母平均 μ は推測したいものだからいいとして，母分散 σ^2 がわかっていないのです。母数がわかっていれば苦労しないのですが，わからないので推測するわけで，ここは母分散の代わりに推定量 $\hat{\sigma^2}$ を使うことにしましょう。標準化もこのスコアを使って行うことになります。

数値例でみてみます。先ほどのこのデータから，

$$X = \{55.22, 50.10, 45.59, 61.99, 48.83, 50.38, 61.95, 53.44, 46.71\}$$

標本平均は 52.69 でした。不偏分散は，

$$\frac{1}{9-1}\left\{(55.22-52.69)^2+(50.10-52.69)^2+(45.59-52.69)^2+\cdots+(46.71-52.69)^2\right\} = 36.5394$$

となります。標準偏差の不偏推定量は，不偏分散の正の平方根で代用して[*6]，

$$\hat{\sigma} = \sqrt{\hat{\sigma^2}} = \sqrt{36.5394} = 6.044783$$

となります。

これを使って標準化する，つまり $(\dfrac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}})$ として区間推定にしたいのですが，残念ながらこれは標準正規分布に従いません。その代わりと言ってはなんですが，母分散がわからないときに不偏分散で代用した場合，t **分布**というのに従うことがわかっています。ここでは，(また) 数理統計学による t 分布の導出を回避して結果だけ使わせてもらうことにして，この t 分布を使って推定することを考えます。

この t 分布，標準正規分布と違ってパラメータとして自由度 ν(ギリシア文字で「ニュー」と読みます。[*7]) というものを使います[*8]。t 分布の形状は標準正規分布に近く，自由度が大きくなればほぼ標準正規分布と同じ形になるのがわかると思います (図 17.5)。この自由度とはなんでしょう？ t 分布の使い所は，母分散がわからない時に標本から計算した不偏分散で代用するというシーンだったことを思い出してください。この分散の推定量は，当然サンプルサイズによってその精度が変わります。このサンプルサイズによって推定の精度を調節しているのが，自由度というパラメータで表されているのです。しかも，この自由度は「サンプルサイズ -1」で計算できるという優れものです[*9]。

さて，これを使って母平均の信頼区間を標本だけから推定していきましょう。不偏分散で代用した標準化スコアは t 分布に従う，つまり次の数字を計算することになります。

$$t = \frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$$

[*6] 標準偏差は分散の正の平方根なので，分散を不偏分散にかえたものの正の平方根にすればそれが不偏標準偏差じゃないのかと言いたいところですが，実はそうではありません。不偏標準偏差を求めるためにはガンマ関数を利用した別のものにしなければなりませんで，大変複雑ですのでここでは「代用する」にとどめています。

[*7] ニューガンダムのニューです。ニューガンダムについては「逆襲のシャア」を観てください。ジークジオン。

[*8] 正規分布のパラメータは平均 μ と標準偏差 σ でした。t 分布の平均は 0 で，自由度 ν は標準偏差に当たるような散らばりの指標だと思っていただければ結構です。

[*9] またマイナス 1 がでてきたー，なんだこれ。と思う人もいるかもしれません。イメージとして，標本における分散の代わりになる指標ですから，サンプルサイズ 1 の時は分散の計算ができず，サンプルサイズ 2 から計算し始めることができる，つまりサンプルサイズ 2 が最初の (1 番目の)t 分布の形なので，サンプルサイズ -1 が t 分布のかたちなのだ，と考えてみてはいかがでしょうか。

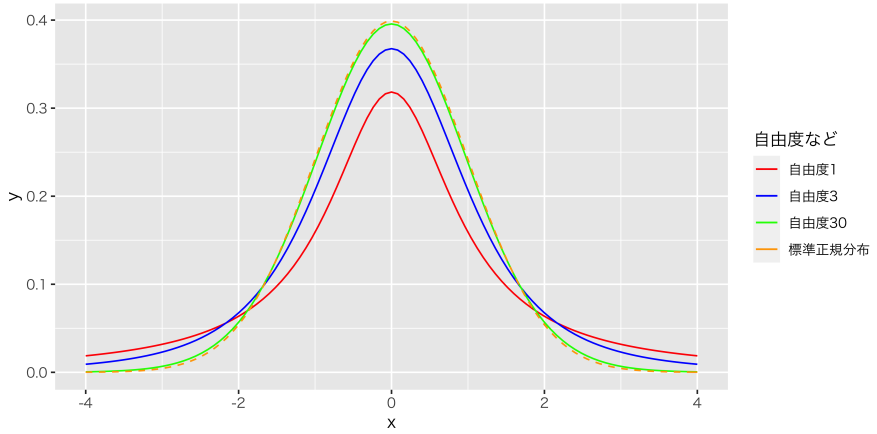


図 17.5 標本統計量の従う分布も正規分布。幅は母分散とサンプルサイズに依存します。

ここで，自由度 8 の t 分布について，95% 区間を R で求めると，両端は ±2.306004 ですから[*10]

$$-2.306 < \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} < 2.306$$

分母を両端にかけて，

$$-2.306 \times \hat{\sigma}/\sqrt{n} < \bar{X} - \mu < 2.306 \times \hat{\sigma}/\sqrt{n}$$

$\bar{X}$ を移項して，

$$\bar{X} - 2.306 \times \hat{\sigma}/\sqrt{n} < \mu < \bar{X} + 2.306 \times \hat{\sigma}/\sqrt{n}$$

これで標本統計量だけで計算できるようになりました。代入しましょう。

$$52.69 - 2.306 \times \frac{6.045}{\sqrt{9}} < \mu < 52.69 + 2.306 \times \frac{6.045}{\sqrt{9}}$$

ですから，

$$48.044 < \mu < 57.336$$

となります。この範囲に 95% の確率で母平均がある，と推定できました。レポートなどで報告するときには，CI[48.044,57.336] と書くことができます。

17.4 推定と検定

さて，ここまでで「標本から母平均を推定する方法」を説明してきましたが，これが実際の研究シーンではどのように使われるのでしょうか。

ここで，第 12 講を思い出して欲しいのですが，心理学ではランダム化比較実験とよばれる「群わけによる操作の効果」を検証するという研究スタイルについてお話ししました。人間の心の状態に影響する要因は無数に考えられ，すべてを統制することはできませんが，ランダムに群わけすることで，この目に見えぬ無数の影響はどちらの群にも同じ程度に入っているだろう，と考えるのです。であれば，それらの影響は相互に相殺しあって，結局平均値としては同じになるでしょう。その後，操作によって集団の平均値が変わることがあれば，それは操作の影響と言っていいでしょう，というロジックですね (図 17.6)。

[*10] 標準正規分布と同じで，t 分布も R が関数を最初から持っています。d,p,q,r などを分布の前につけて (`qt(0.975,df=8)` で計算します。また，t 分布も左右対象ですので，片方が求まればもう片方はその符号反転です。

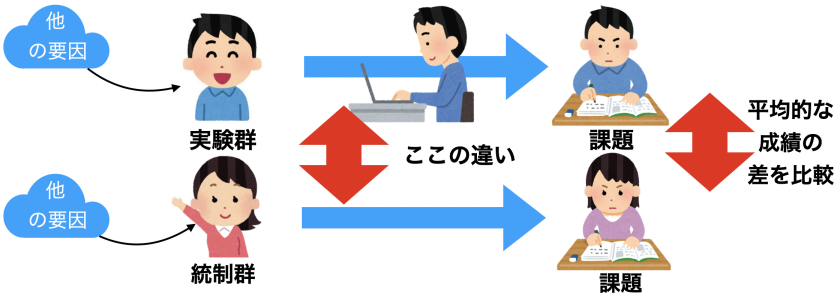


図 17.6 RCT による心理学研究のアプローチ

4595 このように，集団の平均的な因果効果をみたい，というのがそもそものアイデアでした。これに推測統計学
4596 の話を合わせて考えてみましょう。そうです，心理学的な実験のシーンで集められた集団も，母集団からの標
4597 本だと考えるわけです。母集団の標本である以上，興味があるのはデータの平均ではなくて (標本平均では
4598 なくて)，母集団の平均であり，標本平均 (と不偏分散) から推定した母平均の比較がしたいのだ，ということ
4599 になります。

4600 このように推測統計学を使うことで，手元の標本だけではなく，結果を一般化して論じることができるよう
4601 になります。推測統計学を使わずに記述統計だけで考えると，「あるデータをとったらこうなった，だからこう
4602 考えた」というだけで終わりです。これでは「あなたのデータがたまたまそういうデータだったのでしょう」と言
4603 われるかもしれません。小中学校の夏休みの自由研究であればこれでもいいのですが，専門的に科学的推
4604 論をするためにはこの謗りを排除し，手元のデータを超えて，目に見えぬ母集団の平均値で語りたいわけで
4605 す (図 17.7)。

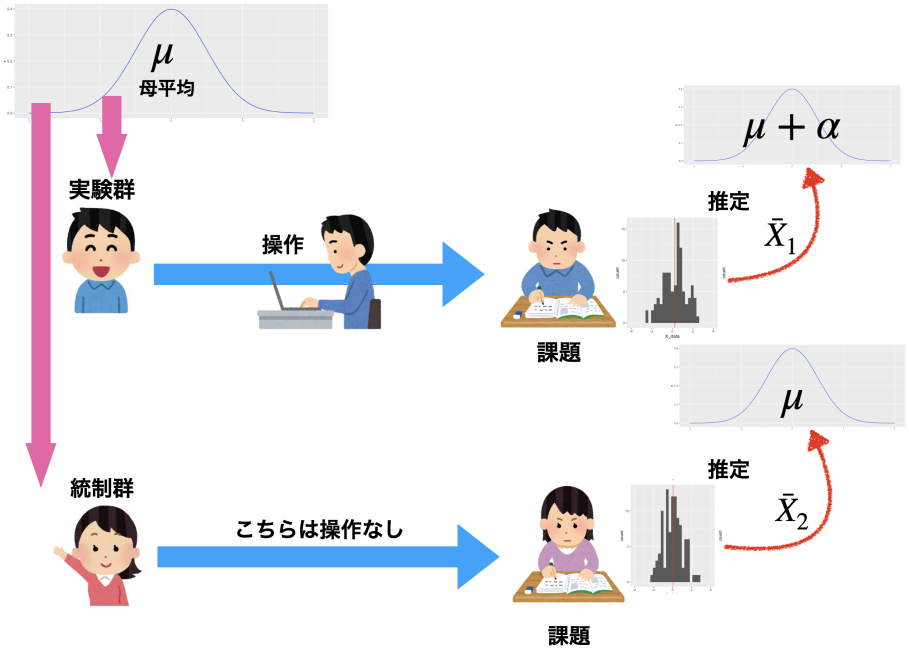


図 17.7 RCT と推測統計を合わせて考える。標本平均に差があるのは当たり前で，問題は母平均に差があると言って良いかどうか，である。

4606 さらに，セクション 12.4,Pp.184 を思い出して欲しいのですが，平均の間に差があった，というためには平

均だけではなく，その散らばりも考えなければなりません。今回の区間推定はここと関係してくるのです。つまり，点推定によって母平均に違いがある，というのでは不十分で，信頼区間も考慮しておかなければなりません (図 17.8)。というのも，母平均がどのあたりにあるのか，標本平均に基づく点推定では，ほぼ確実に差がある[*11]のですが，母平均を当てているとも思えないわけです。なので，測定精度を利用して「この範囲であれば入ってそう」と考えるのですが，その幅が非常に広かったり，他方の群の領域と重複しているようであれば，断言はできないですね。

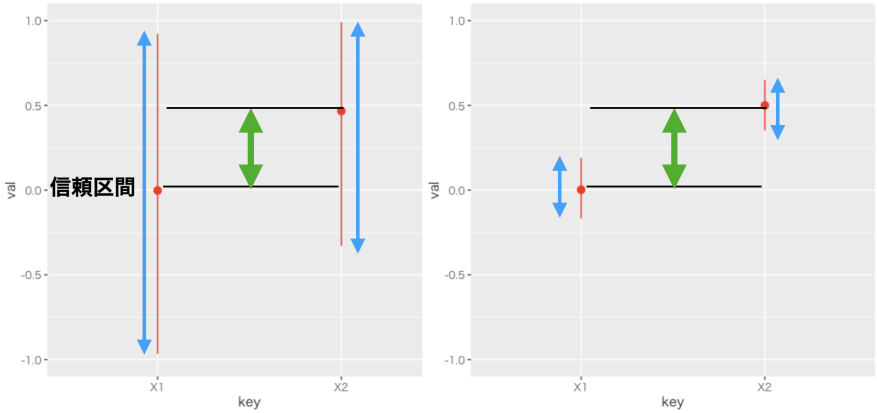


図 17.8 信頼区間 (青い双方向矢印) と合わせて考える。左右 2 つの実験で，点推定値の差 (緑の双方向矢印) は同じであるが，左の実験では信頼区間が広く，右の実験では信頼区間が狭い。この場合，右の実験は差があると言っても良さそうだが，左の実験は差があるとは言いにくい。

点推定だけ考えて，「差があるのかないのか」「違うのか，同じなのか」といった結論を急ぐと，誤った判断をしてしまいかねません。ここで**効果量**の説明を思い出して欲しいのですが，たとえば性差を考えてみます (効果量についてはセクション 5.4, Pp.91)。性差はあちこちで見受けられ，男女は見るからに違うし，得意なこともいろいろ違いそうです。皆さんもこれまでの経験で，男性と女性は違うな，というシーンをいろいろみてきたことと思います。しかし，人間は多様です。その分散は非常に大きく，力強い女性もいれば，少食の男性もいます。外向的な男性も内向的な女性もいますし，几帳面な男性もそうでない女性も。とにかくさまざまです。母分散が大きいと標本平均のゆらぎも大きくなります (標準誤差は $\frac{\sigma}{\sqrt{n}}$ であることを思い出してください)。標準誤差の分子が大きいので，分母も相当たくさんとらないと，その差があるかどうか断言するのは難しいですね。

また，信頼区間は確率を含んだ表現ですから，解釈に注意が必要です。95% 信頼区間は，この区間の中に母平均が入っているはずだ，という幅です。これは区間で表しているものの，分布の情報を持った表現ではありません。確率変数のように，この区間の中の面積が 95% というのではなく，「95% の確率で当たるくじ」のようなもので，当たるか外れるかは，常にどちらか 1 つです[*12]。この確率の言葉を使った表現は，平易な表現からはかけ離れていますし，「確率で母平均に差があるということは 95% の確率で間違っている」と言われてもピンとこないかもしれません。

そこで区間推定に加えて「差があるのかないのか」という判断をする方法が求められます。推測統計学の中で，これは**検定 (test)** と呼ばれます。次回はこの検定の進め方について，具体的に解説していきたいと思

[*11] 差がないということは等しいわけで，有効桁数にもよりますが，連続確率分布の実現値が等しい状況というのはほぼ考えられないでしょう。

[*12] 95% の確率で成功する手術があるとしましょう。それでも手術が失敗して死んでしまうということはあり得ます。たくさん手術をこなす医師にとっては 95% は成功率かもしれませんが，手術を受ける方にとっては生か死か，二択の事実があるだけです。

4630 います。

4631 17.5 課題

4632 **■復習；標準正規分布の区間を求める** R を使って, 次の値を求めましょう。

4633 1. 標準正規分布において, 平均を挟んで 95% が含まれる領域はどこからどこまでですか。
4634 2. 標準正規分布において, 平均を挟んで 90% が含まれる領域はどこからどこまでですか。
4635 3. 標準正規分布において, 平均を挟んで 50% が含まれる領域はどこからどこまでですか。
4636 4. 平均 $\mu = 50$, 標準偏差 $\sigma = 10$ の正規分布において, 平均を挟んで 95% が含まれる領域はどこか
4637 らどこまでですか。
4638 5. 自由度 2 の t 分布において, 平均を挟んで 95% が含まれる領域はどこからどこまでですか。なお, t
4639 分布は pnorm や qnorm の代わりに, pt や qt といった関数を使い, 引数に自由度 df を指定してく
4640 ださい。

4641 **■標準正規分布を使った区間推定** 母 SD が 10 の母集団分布から, サンプルサイズ $n = 16$ の標本を
4642 取ったときの標本平均値が $\bar{X} = 30$ でした。このとき, 次の値を求めましょう。

4643 1. 点推定値
4644 2. 95 % 信頼区間
4645 3. 90 % 信頼区間

4646 **■t 分布を使った区間推定** サンプルサイズ $n = 16$ の標本を取ったときの標本平均値が $\bar{X} = 30$, 標本の
4647 不偏 SD が $\hat{\sigma} = 9$ でした。このとき, 次の値を求めましょう。

4648 1. 点推定値
4649 2. 95 % 信頼区間
4650 3. 90 % 信頼区間

第 18 章

帰無仮説検定

前回は標本から母平均を区間推定する方法について説明しました。今回は，これを使って「差があるかないか」という判断をする方法について考えます。

母平均を区間推定することで，「このあたりにある，と予測したら XX% の確率で当たる」という範囲を定めることができました。この推定を実験群・統制群それぞれについて行うとすると，実験群の平均値はこの範囲，統制群の平均値はこの範囲にあると予測する，という言い方をすることになります。その範囲が重複していなければ，実験群と統制群の平均値は異なると言い切れるでしょうが，信頼区間が重複することは当然起こり得ます。そのときどうやって「差がある」「差がない」という判断をするのか，ということが問題になるのです。この判断は**検定 (test)** と呼ばれます。とくにその形式から，**帰無仮説検定 (Null Hypothesis Significane Test)**，略して **NHST** と呼ばれます。

どこに注意が必要なのでしょうか？ いろいろな問題があるのですが，ここでは説明しながら代表的なものを 1 つ挙げてみましょう。検定はモデル比較の考え方の一種です。統計的なモデル，というのは本講義の中では一貫して述べてきたように，データに対する理想形 + 誤差の考え方で，回帰分析に代表される線形的なモデルが代表的なものです。モデル比較とは，複数のモデルを用意した時にどちらのモデルがより良いか，という優劣を決めることです。統計的検定は，とくに**帰無仮説検定 (Null Hypothesis Significance Test)** と呼ばれ，帰無仮説 (H_0 と表記) と対立仮説 (H_1 と表記) という 2 つのモデルを戦わせ，どちらが良いかを決定することにあります。

つまり，検定は「推定」に「判断」が加わったものです。判断は，何らかの価値に基づいて，あるいは実践的な必要性に応えるためになされることです。たとえば感染症対策として新しい薬を開発し，この効果があるかどうというときに，「効果あり」とか「効果なし」という判断をするのですね。この判断が必要なのは，効果がないのに投薬をすると時間とお金の無駄，ひいては何らかの副作用で人命に影響が出るかもしれない，というシーンだからです。幸い，心理学において人命が関わるような実験というのはほとんどあり得ませんが，実験や介入を行った効果があるのかどうかを判断するのにわかりやすいので，これまでよく使われてきました。実際とてもよく使われているのですが，同時に多くの誤用・悪用を生んでしまいました。あまりにも間違った使われ方，悪い使われ方が広まっていることから，アメリカ統計学会や権威ある科学雑誌である Nature が「間違えた使い方はダメだぞ」という声明を出しており (Amrhein et al., 2019; Wasserstein & Lazar, 2016)，「統計学は瀕死だ」という人もいるほどです (豊田, 2020)。念のためにいっておきますが，帰無仮説検定は間違った理論だ，というわけではありません。理論的に筋は通っているのですが，理論的仮定を犯さずに実践し，正しく解釈するのに細心の注意が必要だ，ということです[*1]。

[*1] これらの声明のタイトルもよく見ると，“p-valuie”(p 値) や “statistical significance”(統計的有意) について注意が必要だ，ということであって，帰無仮説検定 “significance test” そのものが間違っているというわけではないのです。

これら誤用・悪用を生み出す根本的な原因は，帰無仮説検定が「効果があった，なかった」という 1bit 判断[*2]を提供することにあります。判断に迷うような推定結果を，とある基準で Yes/No に振り分けますが，そうするとそのインパクトが大きくなりすぎるのです。たとえばやり方によっては，二群にわけた群の平均の差が 0.1 点でも「差がある！」と判断することになったりします。これが二群の身長の差であれば，0.1cm は気にならないどころか実質的に意味のない差なのですが，「結果としてこの介入 (投薬) をする群と，しない群を比較すると**成長に差が見られた！**」なんて伝わってしまうかもしれません。またこの 0.1 点の差が男女の平均獲得賃金 (単位・100 万円) などであれば大きく意味が違います。この場合は男女差は 1 円だってあるべきではないのです。つまり，本来，差があるかどうかという程度の問題はドメイン知識[*3]に照らし合わせて考えなければならないものなのです。心理学のような目に見えない対象を研究する場合は，この判断が非常に難しく，どの程度の量であれば検出することに意味があるのか，どの程度なら無視して良いのか，ということについて，統一的な見解が作りにくいのです。心理学の場合は，判断の前の推定こそ重要であって，1bit に落とし込まれた派手な結果だけに目移りすることがないよう，注意する必要があります。

さて，いきなり警句から始まってしまってすみません。上に書いたことには充分注意しつつ，具体的に説明を始めていきたいと思います。

18.1 帰無仮説検定の手続き

統計的帰無仮説検定の手続きは一般に，次のようなステップですすみます。

1. 仮説を設定する。
2. 統計的仮説検定に用いられる検定統計量を選択する。
3. 仮説選定判断の基準になる確率を設定する。
4. 実際のデータから検定統計量の実現値を計算する。
5. 最初に定めた仮説を支持するかどうかを判断する。

このステップについて，順に見ていきましょう。

18.1.1 仮説を設定する

まず最初のステップは仮説を設定するところです。帰無仮説検定では，**帰無仮説 (Null Hypothesis)** と**対立仮説 (Alternative Hypothesis)** という 2 つの仮説を用意します。検定は差があるかないかの判断をすることですから，差があるとするモデル，差がないとするモデルを勝負させて，どちらに軍配が上がるかを見ようとすることです。

帰無仮説は Null という言葉があるように，空っぽ，空虚な仮説です[*4]。普通，比較するときには何らかの違いがあるとか，効果がある，というようなことをいいたいわけですが，これに対して Null，つまり違いがない，効果がないというような仮説を一方の側に用意します。帰無仮説というのは，無に帰してほしい，なくなって欲しい仮説ということで名付けられています。

これに対して，対立仮説は Alternative(別の可能性，代わりになるもの，という意味) であり，「差がある」

[*2] 1bit は付録 A にあるように，2 進数のデジタルデータの最小単位で，0/1 の 2 つの状態しかありません。

[*3] ドメインとはその領域のことです。生物学的な身長の話をしているのか，産業社会学的な賃金体系の話をしているのか，といった話題にまつわる領域のことだと思ってください

[*4] ゼロと Null は違います。ゼロはゼロという数字があるといえますが，Null はその数字すらない，虚無です。余談ですが，ネットスラングの「ぬるぽ」とは Null pointer Exception の意味で，ポインタが行き着かない例外（参照されたメモリ番地が存在しないエラー) という意味です。

「効果がある」という仮説になります。研究者としてはこちら側を応援したいところです。応援したい側があるのに，帰無仮説のようなものを置くのは，NHST の判断の仕方が背理法的なものになっているからです。つまり，差がないという帰無仮説を立てて考えると，おかしな事実が生じるのであれば，帰無仮説を棄却できる。**だから**対立仮説を採択できる。これが基本的な筋道です。

18.1.2 統計的仮説検定に用いられる検定統計量を選択する

たとえば平均値が違うといいたいのか，比率が違うといいたいのか，相関係数に意味があるのか，度数のばらつきが違うといいたいのか，などなど，研究のシーンに応じて考えたい標本統計量は変わります。

また，標本統計量をそのまま比較できない場合は，そこから計算して検定用の検定統計量を計算することになります。これまでの例では，たとえば平均の推定をするときに t 分布を使うという話がありました。平均の比較をする時は，この t 分布に従う統計量，すなわち t 値が検定統計量になります。しかし，相関係数などは推定に使う分布が簡単なものではないものがあります。その場合でも，母数の推定や信頼区間を考えなくても，検定の判断ができる簡単に計算できる統計量に変換するという迂回路を取ることができます。

相関係数の検定はこの迂回ルートをとって，t 分布を使った検定を行います。他にも F とか χ^2(カイ二乗，といいます) といった統計量を計算することがあります。

どのシーンでどの統計量が必要なのか，については数理統計学の力を借りてますので，天下り式に覚えるしかないところですが，逆に言えばいくつかの典型例を知っていれば数学的な展開を必要とせずに，結果を手に入れられるのだと思って奮い立ってください。

18.1.3 仮説選定判断の基準になる確率を設定する

検定はモデル比較の結果，優劣を定める勝負です。推定の結果を考察するにあたって，帰無仮説と対立仮説のどちらに軍配をあげるかという，「実践上での意思決定ツール」でもあるわけです。ですので，事前に「こうなったら決着したとしましょう」というルールを決めておかなければなりません。分析の前に判断ルールを決めておかないと，あとでルールの方が変わってしまうと困りますので，必ず事前に決めておく必要があります。

この確率のことを，**有意水準 (significance level)** といいます。有意 significant とは，統計的に意味が有る，ということであり[*5]，これ以上の開きが出たら勝負あったと判断します，という基準です。これは確率の言葉で表現しますが，心理学では一般に 5% が用いられます。なぜ 5 なのか。とくに意味はありません。6 でも 4 でも良かったのですが，最初に統計的検定をやった人が「根拠はないけどとりあえず 5 で」と始めたのが，いまだに使われています。ということで，こだわる必要はないのですが，一般的に 5%，ときどき 1% や 0.1%，10% という数字が使われることがあります。どの数字でも，別に深い理由はありません。

ちなみにこの数字は，あくまでも優劣の判定基準であって，データそのものに言及する数字ではありませんし，仮説の正しさ，仮説モデルとデータの適合度を表す数字ではないことに注意が必要です。

18.1.4 実際のデータから検定統計量の実現値を計算する

実際のデータを当てはめて，検定統計量を計算します。

[*5] 優位と誤変換しないように注意してください。普通の変換では優位のほうが先に候補になりますが，心理学畑出身の人は有意のほうが先に変換候補に出るようになるでしょう。

18.1.5 最初に定めた仮説を支持するかどうか判断する

ここまではルールの設定でした。このあとは，ルールに従って 2 つのモデル (仮説) を戦わせていきます。

実際のデータから，検定統計量がどれぐらいの数字になるのかを計算し，それが出現する確率を計算します。この確率をとくに p **値 (p-value)** と言います。この p 値が有意水準を上 (下) 回ったかどうかが，勝敗の結果になります。ちなみに勝った方の仮説を**採択 (accept)** し，他方を**棄却 (reject)** する，という言い方をします。

これが一連の流れです。このように手続きは形式的に定まっていますので，途中の計算は統計ソフトウェアが自動的にやってくれます。ですから計算の手間はほとんどないのですが，それだけに安易に使って安直に結果だけに飛びつくということがないように，注意が必要です。

18.2 1 つの平均値の検定

具体的な例で考えていきましょう。まずは平均値の推定の時と同じく，母集団の SD がわかっている場合に標準正規分布を使っての検定を，次に母集団の SD がわからない時の，t 分布を使っての検定を見ていくことにします。

18.2.1 母 SD がわかっている場合

さて，話を進めるために，具体的な例をあげてみましょう。ある学校で，学力テストしたとします。その学校で無作為に 49 人分の記録を取り出したところ，平均が 52 点でした。このテスト，全国平均は 50 点で，SD が 9 の正規分布であることがわかっていたとします。これとそのまま比べると，その学校の成績が良い，ということになりますが，たった 2 点差でそのように結論づけて良いかどうか，というのがここでの問題です。

帰無仮説検定のステップを思い出しましょう。まずは仮説を設定します。ここでやりたいことは，「この学校の成績が良いかどうかの判断」です。言い換えると，全国平均の分布が母集団でそれの平均や SD がわかっているのですから，この標本の平均は母平均と同じなのか，異なっているのかが知りたいのです。標本ですからたまたま高い点数が出ること (そしてその逆) もありますが，この標本平均から推測される母平均と，このテストの理論的な平均値 (全国平均) の違いが**統計的に有意な差なのかどうか**が知りたいわけです。数式で表現すると，標本平均は $\bar{X}$ で，$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ ですから，この時の $\mu = 50$ が帰無仮説，$\mu \neq 50$ が対立仮説ということになります。

次に検定統計量ですが，平均値の推定，しかも理論的に平均 50，SD9 の正規分布との照合ですから，推定したい対象の特徴がもうわかっている特殊なケースです。これは変換すれば標準正規分布に従うのですから，標準化された Z 得点が検定統計量になります。

第 3 のステップは基準になる確率を設定する，でした。勝敗を決めるルールの決定ですね。これはとくに意味もなく，慣例に従って $0.05(5\%)$ としておきましょう。

これで準備が終わりました。実際の検定統計量の計算です。平均 μ，分散 σ^2 の正規分布する母集団からの標本の場合，標本平均の標本分布も正規分布に従い，

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

になります。今回の場合は，

$$52 \sim N(50, 9/\sqrt{49})$$

だと考えられます。標準化することで標準正規分布を使うことができますから，

$$(52-50)/(9/7) \sim N(0,1)$$

として，$(52-50)/(9/7) = 2/1.285714 = 1.555...$ を得ます。この 1.556(下三桁までで四捨五入しました) という数字はどのあたりに位置するのでしょう？

そこで確率の計算をしてみましょう。これが次のステップの，実際のデータから検定統計量の実現値を計算する，というやつです。標準正規分布において，1.556 より大きな数字が得られる確率は[*6]，$p = 0.05985405$ でした[*7]。5.9% なのですね。

最後に判定です。5% が判定ラインだったわけですが，今回得られた「確率変数の実現値」は 5% より大きな確率です。つまり大きい＝出現する可能性が大きい，よくある話だったわけです。そんなにレアな数字じゃない。つまり，たまたま今回の標本が，平均から 2 点大きくなるようなことは十分あり得る話だということです。この学校の生徒は優秀だー！と言いたいところなんですが，偶然それぐらいの差がでることがあるのであれば，そこまで胸を張っていうことができません。ということで，軍配としては帰無仮説の勝ち，つまり，「その学校の成績が特段良いわけではない」「統計的には差がない」ということになります。

図で確認しておきましょう。図 18.1 には標準正規分布と今回の判定に係る数字を記載しました。今回計算した統計量は，標準正規分布にしたがいます。標準正規分布で 5% という基準を考えると，± 1.96 を超えた領域になります。今回のデータがこの領域に入るほどのものであれば，帰無仮説は棄却することになりますが，今回はその領域に入らない，つまり「このサンプルからこの平均値が出てくることも，あるよ」ということで棄却できないのでした。帰無仮説を棄却する領域のことを一般に**棄却域**といい，この数字を越えれば棄却域に入るというギリギリの数字のことを**臨界値**といいます。今回の臨界値は ± 1.96 で，両端が棄却域，今回の 1.55 という実現値は棄却域に入らなかったので帰無仮説の勝ち，ということです。

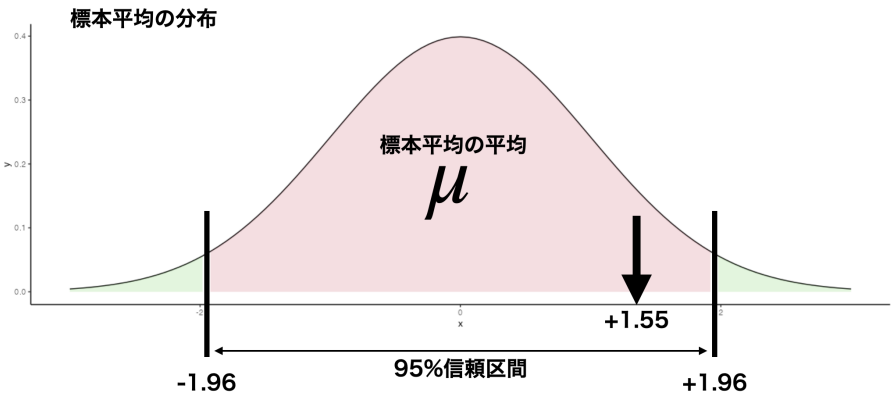


図 18.1 棄却域と臨界値

18.2.2 母 SD がわかっていない場合

先ほどの例は，母 SD がわかっているという話でした。第 17 講でも同じことを言いましたが，まあ母数がわかっていれば苦労はないわけで，それが分からないから推測するのですよね。

[*6] これまた確率の，とくに連続変数のややこしいところで，1.556 という数字が出てくる確率はゼロです。ユークリッド幾何学でいうところの，点は位置があって大きさがないというのと同じで，点の確率は大きさがないのです。確率は面積ですから，1.556より大きな数字が得られる確率，というのが本当は正しい表現です。

[*7] これは R で次のように計算しています：`1-pnorm(1.556,mean=0,sd=1)`

ということで，母 SD の代わりに標本 SD(**不偏分散**の正の平方根) を使って推定することとします。そのさいは (標準) 正規分布が使えないので，サンプルサイズに応じた t 分布を使うのでした。同じデータ例でやってみましょう。今度は標本の標準偏差が必要ですので，これが $\hat{\sigma} = 10$ だったとします。

帰無仮説・対立仮説はさきほどと同じく，$\mu = 50$ か $\mu \neq 50$ です。ただし推測に使う統計量が t になるわけです。判断基準も同じく慣例にならって，なんとなく 5% にしておきましょう。検定統計量 t は $t = \dfrac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}$ で求めることができますから，

$$t = (52 - 50)/(10/7) = 1.4$$

となります。これが自由度 $n - 1 = 48$ の t 分布に従うので，自由度 48 の t 分布において 1.4 以上の数字が出る確率を求めますと，$p = 0.08397194$ となりました[*8]。8.3% なのですね。

これで判定です。5% の判定ラインより大きな確率が出ましたので，これは滅多にないことではない。なくはないのだから，よくある話。珍しいことじゃない。つまり，母平均から外れて大きい数字，とまでは言えないねということになります。今回の判定に関するイメージを図 18.2 に示しました。今回は自由度 48 の t 分布を描き，検定統計量の実現値を示しています。実現値が棄却域に入るか入らないか，という判断をしてもいいですし，今回のように直接確率を算出して 5% になるかどうかの判定をしてもらっても結構です[*9]。

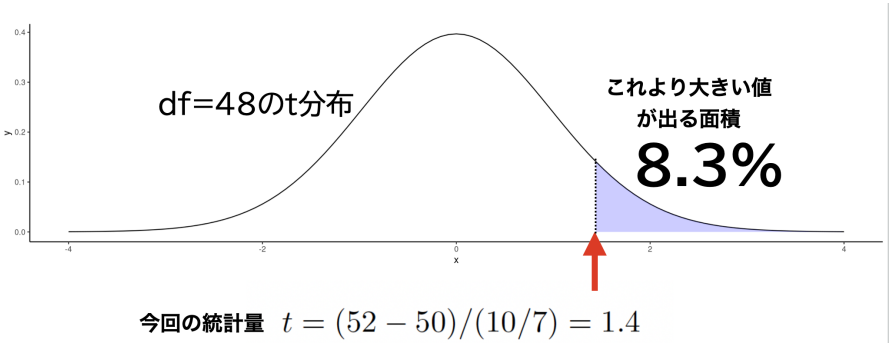


図 18.2 t 分布をつかったときの棄却域と臨界値

いかがでしたでしょうか。今回のデータでは (母 SD がわかっていても，わかっていなくても) 帰無仮説を棄却することはできませんでした。何だか持って回った言い方をしているように感じられるかもしれませんし，いまいちイメージが掴みにくいかもしれません。そこで，今度はまた別のケース，相関係数の例で考えてみたいと思います。徐々に手続きと結論の出し方に慣れていくと思います。

18.3 相関係数の検定

今度は相関係数の検定の話をしましょう。心理学では，平均値を推測するだけでなく，相関係数をテーマに研究することも少なくありません。実験など，介入の前後や群間の比較であれば平均値の話になりますが，こうした因果関係が明確でなく，何らかの関係があるかもしれない，という見込みで探索的に現象を解明していくこともよくあります。たとえばうつ病と起床時間に関係があるのかもしれない。自尊心の高さは学業成績と関係があるのかもしれない。恋愛に対する態度には性格が関係しているのかもしれない etc...。何だか心

[*8] これは R で次のように計算しています: `1-pt(1.4,df=48)`

[*9] 直接確率を計算する方が楽なのに，と思う方もいるかもしれません。これは歴史的な経緯があって，今でこそ R などを使って簡単に確率点が計算できますが，そういう機械がなかった頃は数字の表を見て判断していたのです。第 7 講義のセクション 7.8.2,Pp.126 で数表のことをご紹介しましたが，あれを見て臨界値の大小で棄却域に入ったかどうかを判断する，というのが一般だった時の名残だと思ってください。

理学っぽい話になってきましたね！

さて最後の例，恋愛に対する態度と性格の関係，を研究したいとしましょう。先行研究から，「恋愛についての積極性」を測定する尺度と，外向的な性格かどうかを測定する尺度があったとします[*10]。この尺度を使って，協力者を 20 人募って回答を求め，標本相関係数を算出したところ，$r = 0.5$ という数字が得られたとします (相関係数の算出については第 8 講を思い出してください)[*11]。

さて，標本相関係数は $r = 0.5$ ですが，母相関係数はどうでしょうか。母集団でもこの 2 変数は関係がある，と言って良いでしょうか。標本の話を母集団に広めるには推定，そして結論づけるには検定をするのでした。検定のステップを踏んで話を進めましょう。

■仮説の設定：相関係数の場合　まずは仮説の設定です。帰無仮説と対立仮説が必要なのでした。帰無仮説は Null な仮説ですから，今回は「相関関係がない」というのがそれになります。母相関はギリシア文字 ρ(ローとよむ) で表されますから，数学的な表現をすると帰無仮説は $H_0 : \rho = 0$ です。母相関が 0.0 と一致する，という強い主張ですね。これに対する対立仮説は，これの否定ですから $H_1 : \rho \neq 0$ です。既にこの 2 つのモデル対決，帰無仮説はピンポイント予測ですから，かなり狭い可能性のような気もしますが，帰無仮説検定の話の上ではこうならざるを得ないのです。

というのも，帰無仮説検定は背理法のロジックを使う，というところにその話の味噌があります。まずは極端かもしれませんが，帰無仮説の世界を想定しておき，それに従っていくとおかしなことが起こるなら，帰無仮説を棄却する，という話の進め方です。母相関が 0 な世界を想定してみましょう – 幸い，最近はコンピュータを使って簡単にシミュレーションをできます。相関係数が 0 の世界があり，そこから 20 人のサンプルをとっていくとしましょう。

図 18.3 に示したのは，母相関 $\rho = 0$ の世界から乱数を 10 万点発生させ，そこからランダムに 20 個の点を取り出した例です。黒い点の塊はほぼ円で，二変数に相関がないというデータを具現化したものが示されています。小さな赤い点がそこから 20 個取り出した場合で，今回この 20 個の点による標本相関係数は $r = 0.143$ でした。理論的には $\rho = 0$ の世界からのサンプルなのですから，r も 0 であって欲しいかもしれませんが，一部の標本だけ取り出すとぴったり 0 になることはありません。同様のサンプルを何度も繰り返すと，標本相関係数は $r_2 = 0.32, r_3 = -0.46, r_4 = -0.26$.... といろいろな数字を取ります。この分布が標本分布になるのです。

このように，いったん「母相関は $\rho = 0$」，としてから標本を取り出したとして，その標本分布がどうなるかを考え，たまたま今回得られた標本相関係数 (今回の例では 0.5 でしたが) が出てくる確率はどれぐらいか，を考えます。その確率が非常に小さければ，滅多にないことが目の前のデータで生じたわけですから，これは前

[*10] ここでは軽〜く書いていますが，心理的な状態を測定する物差しのことを**尺度 (scale)** といい，この尺度を作る・使い方を考えるのも心理学の重要な研究領域なのです。この領域をとくに**心理測定法**といいますが，テスト理論の発展系である因子分析モデルや構造方程式モデリングによって，調査項目に対する反応を細やかに検証する世界があります。それでなくても目に見えないものを測定するのですから，きちんと測定できているかどうかは大問題です。また世の中には論文化され，公開されている心理尺度はたくさんありますが，自分の研究に借用するときは「本当に自分が測りたいものと，この尺度が測ろうとするものが同じなのか」「今の，私の研究対象に同じ測定方法で測って同じ結果が得られるのか」，など常に批判的な態度を持っておく必要があります。

[*11] ちなみに，標本相関係数の期待値 (標本を何度もとったときの標本相関係数の平均値) は，母相関に**一致しません**。つまり不偏分散を計算したように，不偏相関係数を計算しなければならないのですが，簡単な式で表現できないので標本相関係数のままで推定することが慣例になっています (南風原, 2002; 平川, 2021)。また区間推定をする場合は，標本相関係数の標準誤差を考えなければなりません。標本相関係数の標準誤差の推定値 $\hat{\sigma}_r$ は，次の式で与えられます。

$$\hat{\sigma}_r = \frac{1 - r^2}{\sqrt{n}}$$

今回の場合ですと，$\hat{\sigma}_r = \frac{1 - 0.5^2}{\sqrt{20}} = 0.1677$ となります。ただし標本相関係数の標本分布は正規分布ではありませんので，区間推定をするには Fisher の Z 変換という手続きが必要です。

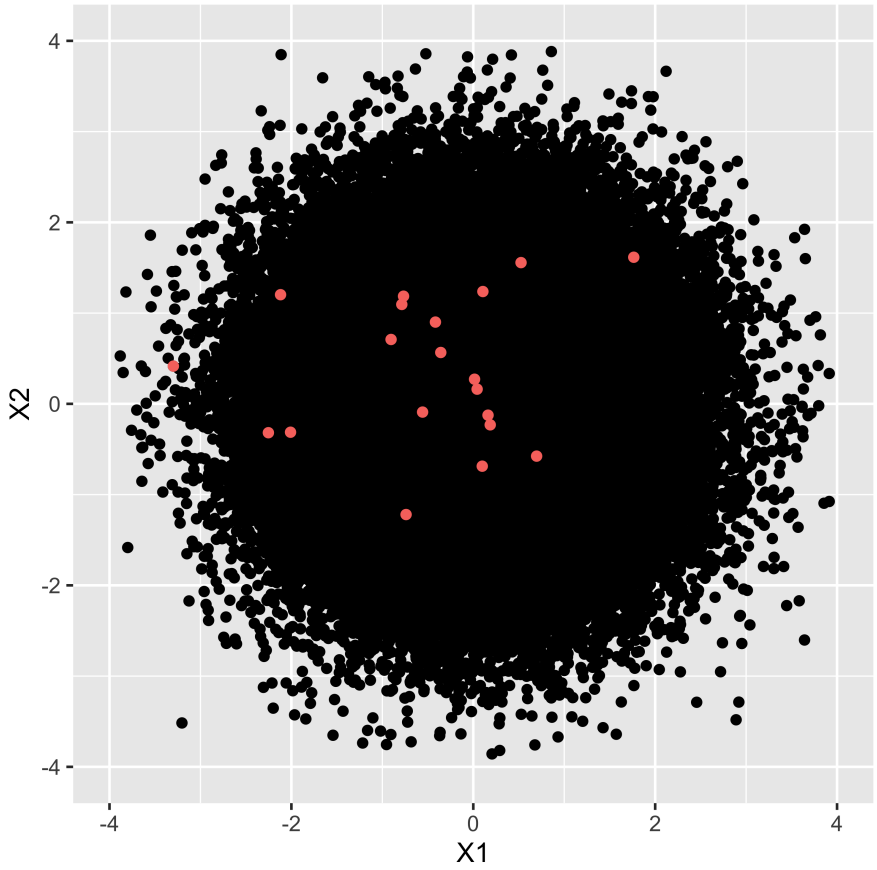


図 18.3 母相関 $\rho = 0$ の世界とそこから $n = 20$ のサンプルを取った例

提である「母相関は $\rho = 0$」という仮定の方がおかしい，ということで帰無仮説を捨てるわけです。ここが背理法になっているわけですね。また，帰無仮説として「母相関はありまぁす！」という仮定を置くのも難しいところなのがわかっていただけるかと思います。相関が「ある」といって，どの程度あるのかは無数に考えられるのに対し，相関が「ない」の場合は一意に定まるからです。

■検定統計量を定める ここで，標本相関係数の標本分布は，自由度 $n-2$ の t 分布に従うことがわかっています[*12]。ただ，そのまま相関係数を使うのではなくて，標本相関係数を次の式で変換する必要があります。

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

■判断基準の設定 今回も慣例に倣って $0.05(5\%)$ にしましょう。

■検定統計量の算出 今回の例では，

$$t = \frac{0.5 \times \sqrt{18}}{\sqrt{1-0.25}} = 2.44949$$

となりました。

[*12] ここでも数理統計学の結果だけお借りしましょう！

■では判定を 今回得られた数字，$t = 2.45$ よりも大きい数字が自由度 18 の t 分布から出てくる確率はどれぐらいでしょう？ これを計算すると[*13]その確率は $p = 0.01237173$ です。またここでは，$\rho = 0.0$ の元で標本相関係数が 0.5 より大きくなる確率を考えているのですが，この偏った相関係数の出方は $|r| > 0.5$ についての話ですから，左右対称の分布であることから $p = 0.0124 \times 2 = 0.0248$ と考えるべきです。いずれにせよ，絶対値 0.5 以上の標本相関係数が得られる確率は 5% よりも小さい数字で，勝負の判定としては「滅多にないことが起こった」と言って良いわけですから，これは帰無仮説がおかしい。そもそも母相関が $\rho = 0$ であれば，たまたま標本から相関係数 $r = 0.5$ が出てくるなんて，ありえへん，というわけです。そこで結論は，帰無仮説の棄却，対立仮説の採択となります。図 18.4 に t 分布，臨界値，実現値の関係を図示しましたので，確認しておいてください。

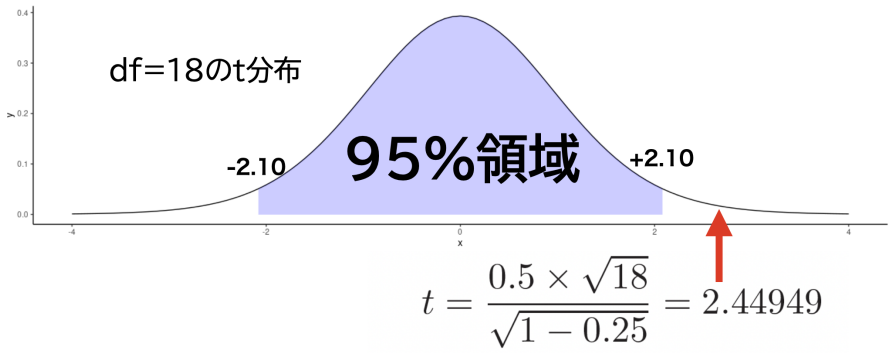


図 18.4 相関係数の検定結果を t 分布で判断する

いかがだったでしょうか。区間推定とは少し違ったロジックですが，判定基準を設けて 2 つの仮説 (モデル) を戦わせることで，判定はできるようになりました。次回はこの判定結果の報告の仕方や，帰無仮説検定のもたらした問題点などについて考えていきたいと思います。

18.4 課題

■母 SD がわかっている場合の平均値の検定をしてみる 母平均が 40，母 SD が 10 の母集団分布から，サンプルサイズ $n = 16$ の標本を取ったときの標本平均値が $\bar{X} = 45$ でした。この標本から推測される平均値が，母集団と比べて有意に異なっているかどうかを判断したいと思います。このとき，次の問いに答えてください。

1. 標本平均から推測した母平均の 95 % 信頼区間はどのような値になりますか。標本分布に照らし合わせて考えてください。
2. 母集団に照らし合わせた時，今回の標本平均以上の数字が出てくる確率はどれぐらいですか。母集団分布に照らし合わせて考えてください。
3. 有意水準を 5% として帰無仮説検定をします。この時の帰無仮説と対立仮説はそれぞれどのようになりますか。
4. 有意水準を 5% として帰無仮説検定をします。結論として，どちらの仮説が採択されますか。

[*13] R で `1-pt(2.45,df=18)` とします。

■t 分布を使った区間推定 母平均が 40, 母 SD はわからないのですが, サンプルサイズ $n = 16$ の標本を取ったときの標本平均値が $\bar{X} = 45$, 標本の不偏 SD が $\hat{\sigma} = 9$ でした。この標本から推測される平均値が, 母集団と比べて有意に異なっているかどうかを判断したいと思います。このとき, 次の問いに答えてください。

1. 標本平均から推測した母平均の 95 % 信頼区間はどのような値になりますか。
2. 母集団に照らし合わせた時, 今回の標本平均以上の数字がこの母集団から得られる確率はどれぐらいでしょうか。
3. 有意水準を 5% として帰無仮説検定をします。この時の帰無仮説と対立仮説はそれぞれどのようになりますか。
4. 有意水準を 5% として帰無仮説検定をします。結論として, どちらの仮説が採択されますか。

■t 分布を使った相関係数の検定 サンプルサイズ $N = 49$ で標本相関係数 $r = 0.3$ という結果が得られたとします。この時この標本相関係数が統計的に意味があると言っていいかどうかを判断したいと思います。このとき, 次の問いに答えてください。

1. 有意水準を 5% として帰無仮説検定をします。この時の帰無仮説と対立仮説はそれぞれどのようになりますか。
2. 有意水準を 5% として帰無仮説検定をします。結論として, どちらの仮説が採択されますか。

第 19 章

帰無仮説検定の注意点

さて前回は帰無仮説検定について説明しましたが，いかがだったでしょうか。まず「関係がない」「差がない」といった**帰無仮説**をおき，本来主張したいことを**対立仮説**とし，帰無仮説の否定をもって対立仮説の採択とする，という**背理法**のロジックで結論を導き出しますから，少しひねくれた考え方になっていることに注意が必要です。手元に得られたデータが，帰無仮説の下では滅多に生じ得ないことだということでもって，帰無仮説を棄却するということも理解していただいておりますでしょうか。

今回は，この帰無仮説検定を行った結果を，どのように報告し，どのように解釈するのかについての解説を行います。とくに解釈の仕方については注意が必要で，少し触れたように帰無仮説検定は**誤用・誤解が多い**ことも指摘されています。筆者の考えでは，背理法というひねくれた考え方と，基準となる確率の考え方が不自然であることに根本的な原因があり，初学者には帰無仮説検定は向いてないのではないかとさえ思います。それでもその手軽さから広く浸透した手法ですから，皆さんが先行研究などで触れる過去の論文の中にも間違えた使い方，間違った解釈をしているものが含まれているかもしれません[*1]。皆さんが同じ過ちを犯さないように，しっかりと基礎知識を身につけ，安易で誤った答えに飛びつかないようにしてください。

まずは帰無仮説検定の結果報告の仕方について，次のセクションで少し解説していきます。

19.1 結果の報告の仕方

帰無仮説検定は帰無仮説と対立仮説のモデル対決なので，結果は軍配をはっきり示す必要があります。帰無仮説が勝ったのか，対立仮説が勝ったのかを明記するのです。ちなみに仮説が勝った，というのを**採択 (accept)**，負けたというのを**棄却 (reject)** と言うのでした。ですから「帰無仮説を棄却して対立仮説を採択」，あるいは「帰無仮説が棄却できずに採択，対立仮説を棄却」のどちらかになります。

また，結果の判定ですから，判定に使った数字＝検定統計量も報告する必要があります。たとえば相関係数の検定の例では，t 統計量が使われましたので，これを報告します。また，t 統計量は t 分布に従い，これはサンプルサイズに依存する自由度によって調整されるものですから，自由度も報告する必要があります。前回の例では，相関係数 $r = 0.50$ がサンプルサイズ 20 のデータから得られ，統計的に有意であると結論づけられたのでした。この結果は，次のように報告しましょう。

[*1] プロの研究者 (大学教員など) が書いて，プロの研究者が査読して刊行されている学会論文においてさえ，間違った手法や解釈が含まれている可能性があります。一つは時代的なもので，かつて計算機が十分に発達していなかった頃には正しい分析法が適用できず，やむなく代替案・次善の策として別の方法を使わざるを得なかった，という事情です。当時は正しいとされていた分析手法が，のちの研究で問題が指摘され，より改良された正しい方法ができることもあります。いずれにせよ研究をする上では，常により正しい手法を使うように心がけなければなりません。また，正しくない手法と知りながら実践するような態度は厳に慎まなければなりません。

標本から得られた標本相関係数 $r = 0.50$ について仮説検定を行ったところ，$t(18) = 2.450, p = 0.0248$ で統計的に有意であった。

書き方のポイントとしては，統計量を書くこと，そしてその時の統計量はイタリック (斜体) であること[*2]，その統計量の実現値が得られる確率 (p 値) を記載すること，です。

p 値は上の例のように確率そのものを描くのが基本ですが，昔は判定基準を使って $p < 0.05$ とか $p < 0.01$ と書いていました。これは手計算で検定をしていた頃の名残で，統計環境 R などがなかったものですから確率計算が直接できませんでした。ではどうしていたか，というと代表的な確率分布の数字が載っている数表[*3]を使ってチェックするしかなかったので，細かい数字まで計算できなかったのです[*4]。今ではそういう問題がないので，そのまま p 値を書きましょう。

19.2 統計的に有意とは

さてここまで，(ひとつの) 平均値の検定や，相関係数の検定を行なってきました。想定された平均値と大きく違っている/いない，とか，母集団での相関関係がある/ない，という事例でしたが，最終的には帰無仮説を棄却する/採択する，という話になるのでした。

この時，帰無仮説が棄却できると「統計的に有意 statistically significant」というのでした。統計的に意味が有る，というわけです。ただし気をつけて欲しいのは，統計的に有意とはどういうことか，です。相関係数の検定の時のことを思い出しましょう。帰無仮説は $\rho = 0$ というものでした。これが棄却されたので，「統計的に有意な相関係数だ」と結論づけたわけですが，より正しくは「母相関が 0 であるとはいえない」ということですよね。関係がないとは言えない $\rightarrow$ 関係がある，というこの $\rightarrow$ の部分は，厳密に言えば少し飛躍があることになります[*5]。

この帰無仮説検定を考えた人は，R.A. フィッシャーという人なのですが，この人は帰無仮説が棄却されたら「なにかおかしなことが現実に起こっているのだから，実験手続きを見直したほうがいい」というチェックのために使うことが想定していました。これを E. ピアソンと J. ネイマンという人が「帰無仮説の棄却＝対立仮説の採択，とすれば良い」という判断基準と合わせて使う方法への応用したのです。心理学ではその流れを受けて帰無仮説検定を利用してきました。そういう意味で，本来の創始者の意図とは違った，より実践的な意味のある (言い換えれば実用上の責任を負わなければならない) 使い方をしているのです。

また既に述べたように，実際の相関係数の大きさや，平均値の差についての情報がかき消された表現になっていることにも注意が必要です。最終的な判断は p 値だけを見るのですが，この数字だけみても相関係数の大きさや，平均値の差がどれぐらいあったのかはわかりません。有意になったという結果だけをみて，関係があったのか，そうかー，と考察を巡らせても，実は標本相関係数が $r = 0.1$ だったということもありえるのです[*6]。これで「統計的に有意だ！無相関とは言えない！」と言っても，標本相関係数が 0.1 なのですから，「(相関) 関係がある」という結果をもとに議論するとおかしなことになってしまいます。

[*2] 字体を整えるのは，心理学だけでなく論文の書き方として一般的です。統計量や本のタイトルなどはイタリックであることが決まっています。誰が決めているかって? America Psychological Association の Publication Manual にそう定められていますし，日本の心理学会も基本的にこれに準じた出版作法を定めているのです。

[*3] 昔は確率分布の数字の一覧が載っている本がありました。確率のテキストには付録として，数表の一部が載っているものもあります。

[*4] たとえば t 分布は自由度によって変わりますが，自由度 $1, 2, 3 \cdots$ と 1 つずつ書いてあったのでは，付録のページ数がいくつあっても足りません。そこで $1, 2, 3 \ldots 10, 15, 30, 50$ のように，後半の方は「だいたいこれぐらい」という大雑把な幅でしか載っていませんでした。こんな時，じゃあ自由度 18 ならどうしたらいいんだ，と思うかもしれませんが，その時は手元のデータに最も近く，より厳しい自由度 15 のところを見て代替する，というやりかたでした。ですから具体的な p 値は書けなかったのです。

[*5] 「あなたのことが嫌いではない」といわれても好きだとは言われてないわけで。

[*6] 実際に，サンプルサイズが 271 人を超えると，標本相関係数が 0.1 でも 5% 水準で有意になります。

ところが実際，このような表面上の使われ方に引っ張られてしまうことはよくあります。とくに研究をしていると，努力したのに関係がなかったとは言いたくないという気持ちが働いてしまいます。そこで，有意になったんだから意味はあるよね，と考えて無理やり二変数間の関係を考察してしまう，ということになってしまいます。これではせっかくデータを取ったのに，正しく考えられたことになりません。

一般にサンプルサイズが大きいと，p 値は小さくできますから，帰無仮説は棄却しやすくなります。図 19.1 には，横軸にサンプルサイズ，縦軸に p 値をとり，標本相関係数が $r = 0.5, 0.3, 0.1$ のときにどうなるかをプロットしました。5% を有意水準 (点線) としていますが，サンプルサイズが大きくなるにつれて p 値は小さくなって行きます。

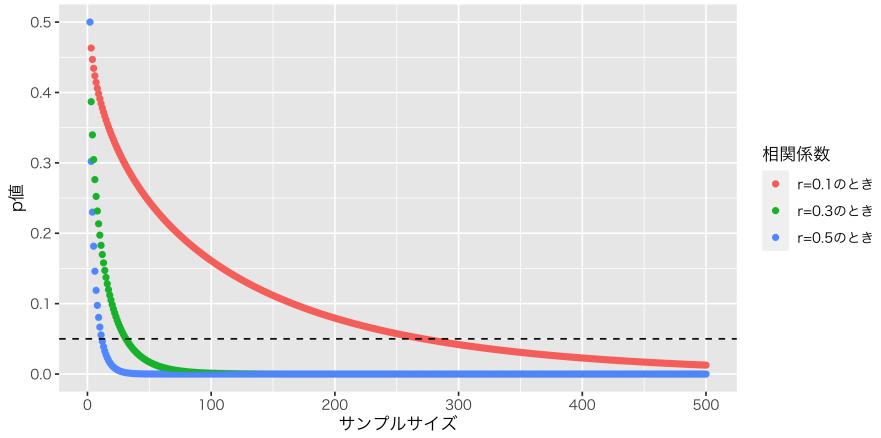


図 19.1　サンプルサイズと p 値の関係

この下がり方はどこからスタートするかにかかわらず同じ形ですから，判定はサンプルサイズが増えさえすればいずれ有意になるのです。これがまた，研究の実践ではおかしな風潮を呼んでしまいます。つまり，有意にしたければサンプルサイズをたくさん取る＝頑張れば良いのです。

しかしながら，科学的に正しいかどうかを考えるべきシーンで，「頑張れば有意になる」といった “努力” に依存するのはおかしいことだと思いませんか。頑張った → 有意になった → 著者の主張が正しい，というのは，論理的ではありません。結果を自分の都合の良いように変えることは，正しい努力ではないはずです。

帰無仮説検定は，標本から母集団のことを「判定」できる優れた技術ですが，表現とその解釈に気をつけないと，科学を間違った方向に導きかねないことに注意が必要です。残念ながら，これまでの心理学の歴史の中でもこのような間違った努力がなされてきた事実があります。これらは**問題のある研究実践 (Questionable Research Practices)**，すなわち **QRPs** と言われています。たとえば「有意な結果にするために」サンプルサイズを増やすとか，「有意な結果になるまで」検定を繰り返す，といったことが QRPs に当たります。そうした研究実践が実際に行われている現場も少なくないようです。しかしそんなことをしていると，必然的に科学の議論が「正しくない結果 (の解釈)」に基づきますから，後続の研究が土台から崩れ，長年の努力が無駄になってしまいます。データの捏造や，他人の書いた論文を盗む剽窃は明らかに悪意のある行為ですが，「頑張ってデータを増やす」というのは一見真面目で誠実な研究態度に思えてしまうので，ちょっとタチが悪い問題です。努力が人を誤らせるのですから。皆さんも騙されたり間違った道に進んでしまわないよう，十分に注意して研究を進めてください。

19.3 検定における 2 種類の間違い

19.3.1 p 値とは何か

ところで，帰無仮説検定は「推定」の応用である「判断」だということでした。判断が確率的に行われるのですが，人間は確率的な考え方が得意ではありませんから，うっかり間違って理解してしまうことも少なくありません。正しい判断，間違った判断というのはどういうときに起こるのかを考えてみましょう。

まず先ほどから p 値と呼ばれる数字を使って説明してきましたが，改めてこの p 値は何を表している数字でしょうか？ 次の選択肢の中から考えてみてください。

- 帰無仮説の成立する確率
- 帰無仮説の正しさを表している数字
- 有意の程度の強さを表している数字 (p 値が小さければ小さいほど効果が大きい)
- 検定の強さを表している数字

実は正解は，上のどれでもない，です。意地の悪い選択肢でごめん。

上の 2 つの選択肢，「帰無仮説の生じる確率」「帰無仮説の正しさを表している数字」はよくある勘違いです。仮説検定は「帰無仮説の下で･･･」という前提から始まります。無相関検定のときに，母相関が 0 の世界から標本をピックアップする例をお話ししましたね。母相関の世界を前提として･･･という話なので，帰無仮説の生じる確率は 100%，帰無仮説の正しさも 100% です (だって仮定したんだから！)。

有意の程度の大きさを表す数字でもありません。確かに差が大きいとか，帰無仮説の世界から大きく離れているというようなことがあれば，検定統計量の実現値は大きくなり，p 値は小さくなります。が，p 値が示しているのは(帰無仮説の下での) 検定統計量の出現確率に過ぎません。いわば帰無仮説という仮想空間の中での数字であって，現実の統計量について何か示す数字ではないのです。

検定の強さを表す数字，でもありません。p 値が小さくなればなるほど，確かにレア度は増しているのですが，それは帰無仮説という仮想空間の中でのレア度であって，帰無仮説が対立仮説をどれほど打ち負かしたか，という情報を持っているわけではありません。ですから，p 値が小さければ小さいほどすごい，などと思ってしまうと本質を見誤ります。

では検定の正しさや間違いについて，どのように考えれば良いのでしょうか。実はこれも，確率的表現の難しさがあるのです。帰無仮説検定最後のステップでは「最初に定めた仮説を支持するかどうか判断する」というのがありました。この判断の間違い方にも，2 種類あるのです。つまり，「帰無仮説が正しいのに棄却してしまう」という間違い方と，「帰無仮説が正しくないのに採択してしまう」という間違い方です。

図 19.2 に，間違い方の組み合わせを書いてみました。帰無仮説が正しいときに帰無仮説を採択する，帰無仮説が正しくないので帰無仮説を棄却する。これができれば正しい決定ですが，「帰無仮説が正しいのに帰無仮説を棄却する」のも「帰無仮説が正しくないのに帰無仮説を採択する」のも間違いです。

帰無仮説は差がないとか効果がないといったからっぽ (Null) な結論です。ここでは仮に，開発された新薬に治療効果があるかどうかを考えている，というシーンを想像してみてください。帰無仮説はこの新薬に効果がないときです。ロクな効果がないときに，帰無仮説を棄却する＝対立仮説を採択する＝効果があると言ってしまうのは，怖いことですね (患者さんに副作用など悪影響があるかも！)。このような功を焦って誤った判断をしてしまう危険な間違いを，**タイプ 1 エラー (Type I Error)** とよび，その確率を α で表します。実は，5% 水準で判断するというのは，この α の確率を 0.05 にしたことになります。慣例では $\alpha = 0.05$ となっていますが，別に 1% でも 10% でも，自由に決定して構いません。どの程度の水準にするかは，この間違い

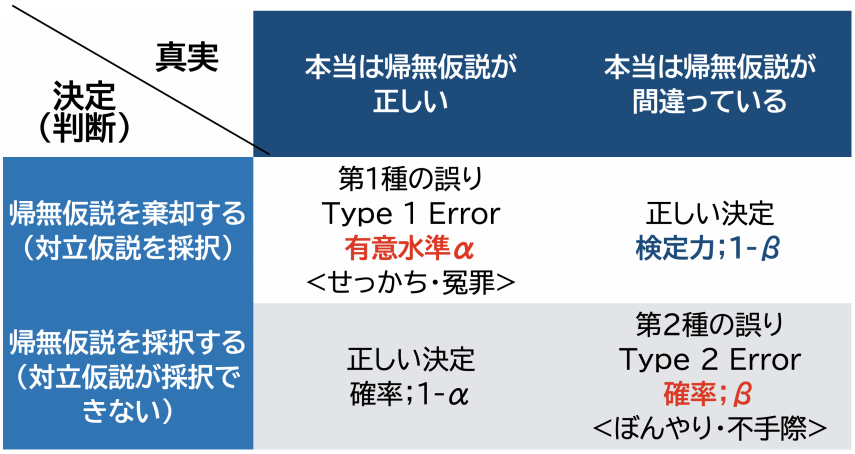


図 19.2 二種類の間違い方

を犯した時の重要さを参照しながら決めるべきことです。

一方，効果があるのに（帰無仮説が間違ってるのに），帰無仮説を採択するのは，せっかくの効果を見抜けなかった残念なケースです。これも困った結果ではありますが，少なくとも患者に迷惑がかかるようなことはありませんよね。なので，こちらについてはひとまず間違っても実際的な問題は生じないでしょう。こちらのケースは**タイプ 2 エラー (Type II Error)** と呼び，確率は β で表すのが一般的です。ただ，この β は帰無仮説が間違っている，つまり Null「でない」状態が，どう「である」のかわからないので，大きさを見積もることはできません。せめて，効果があるときはちゃんとあるって知りたいよね，と思いを馳せるだけです。この図 19.2 にある表は，列ごとに世界線が分かれていますから，右列の「帰無仮説が間違っている世界」で「帰無仮説を棄却してしまう確率」が β であれば，ちゃんと効果を見つけ出せる確率は $1-\beta$ です。この $1-\beta$ を**検定力 (power)** といい，検定力のある分析ができているかどうかをちゃんとチェックしておくのは，研究の手続き上重要なことです。

19.3.2 α のインフレ

先ほど，5% 水準という**有意水準**は α を設定したことと同義だという話をしました。つまり効果がないのに効果があると言ってしまう，うっかりミスを防いでおこうという方針で判断基準を設計しておきましょうということです。それを踏まえた上で，次のような話を考えてみてください[*7]。

> あるレストランの地下にはたくさんのワインが貯蔵されているが，20 本に 1 本は酸っぱくなった不良品が混じっている。客がやってきて 1 本のワインを注文する時，ソムリエは長い目で見て 20 回に 1 回冷や汗をかくことになる。このレストランで開かれるパーティには 10 本単位でワインを出すことになっているが，1 本でも酸っぱいワインが入っていたらパーティはダメになってしまう。パーティーのときは，ソムリエはどれほど謝らなければならないか？

何だか怪しいレストラン，いかがわしいソムリエですね。こんな店が実在していたら行きたくないですが，もちろんこれは架空の話で，20 本に 1 本のダメなワインがあるというのは，帰無仮説検定における 5% の有意水準のメタファーになっています。つまり，5% の判断ミスが生じる危険性がある，ということですね。うっかり

[*7] この例え話は永田・吉田 (1997) からの引用です。

ミスの水準を 5% にしているのは，言い換えると 95% の高確率で判断があっているのですから，ほとんどの場合は問題ないとも言えます。このレストランでは 20 回に 1 回しかミスはない，というわけです。

ところがパーティーが開かれると，ワインがまとめてどっさり提供されるのでした。1 回のパーティで 10 本のワインが出るんでしたね。ひとまず 2 本のワインを持ってきたとします。危険率は 5%，安全率は 95% ですが，これは一本あたりの話です。2 本連続で安全である確率は，$0.95 \times 0.95 = 0.9025$ になります。言い換えると，どちらか一本がアウトである確率は $1 - 0.9025 = 0.0975$ で，なんとほとんど 10% ではありませんか。

同じように考えて，10 本まとめて持ってくると，その中のどれか一本でも間違っている確率は，$1 - (0.95)^{10} = 0.4013$ ということになります。40% ですね。つまり 10 本まとめて持ってくると，一本ずつだと 5% だった危険率が 40% に跳ね上がっていることがわかります。

話を心理学の世界に戻しましょう。心理学では実験の結果を検証するときに，5% の危険率で判断するのでした。さて，とある熱心な研究者がある現象についてさまざまな角度から研究し，何度も統計的な検定を重ねて結論を導き出そうとしているとします。その研究者はとても熱心で慎重なので，1 つの研究のなかに 10 種類の実験を考え，10 回それぞれ 5% 水準で帰無仮説検定を行い，丁寧に判断を繰り返してすべて有意な効果が得られたとします。これは頑健な結果だ，一連の研究はうまくいったぞ･･･といえるでしょうか。

実はそうならないのです。さきほどのワインの話のように。つまり 1 つの研究で 10 回検定を行うと，たまたまどこかで有意になる (差がないのに差があるように判断してしまう) 可能性が 40% に膨れ上がっています。つまり，一連の研究で 5% の危険率を維持しようとするなら，たとえば 10 回の検定が含まれるのであれば，それぞれの有意水準はせめて 0.5% にしておかなければなりません。さもないと，うっかりミスの確率がインフレしてしまうのです。

これも悲しいかな，努力が裏目に出てしまう可能性ですね。一見，1 つの論文で複数回の実験や検定を行なっているのは，慎重で真面目であり，有意な結果が出ればそれは十分強い証拠だ，と考えてしまいます。しかしこれはあくまでも 0/1 の判断であり，しかも確率的な判断ですから，間違いの評価の仕方も確率の理論に沿って丁寧に考えねばなりません！

実際ひとつの論文の中に複数回の検定が行われていることは，少なくありません。しかしその中で，α 水準の全体的な調整がなされていることはあまり見かけません。そう，多くの研究者は気づいていながら間違った研究実践をしているのです。「みんな間違っていてみんないい」とはならないのでして，こうした状況は一刻も早く修正する必要がありますが，なぜかなかなか浸透しないところでもあります。本来は，1 つの論文全体で調整をかけるべきなのですが･･･。こうした誤りに振り回されないようにするには，「差がある/ない」といった 1bit 判断にするのではなく，「どの程度の」差が見られたのか，その効果の大きさをしっかり見極める目が必要になってきます。

19.4 効果量と例数設計

検定力のチェックは，分析が終わった後に事後的にすることもできますし，事前にすることもできます。事前に，というのは，みたい効果がどの程度あるのかを考え，その効果をきちんと検出するにはサンプルサイズをどれぐらいにすれば良いのかを考えること，でもあります。このサンプルサイズを考えることを，**例数設計 (sample size design)** といいます。

帰無仮説検定において，例数設計はとても大事なことです。既に述べたように，サンプルサイズが大きくなると有意だという判断はしやすくなります。これは p 値で判断する以上仕方のないことですが，だからと言ってそれを逆手にとって，事実よりも主張をするためにサンプルをたくさんとるのは，帰無仮説検定のハッキン

グのようなものです[*8]。ですので，検定をするときはまず「どの程度の効果が見込まれ，これを検証するためには何人ぐらいのサンプルをとらなければならない」と宣言してから，検定して判断するという手続きが必要です。この宣言をせずに検定をすると，「有意にするためにサンプルサイズを変えたのではないか」と批判されてしまうかもしれません。学部の授業や卒論で行う実験で，例数設計をしたりサンプルサイズの宣言をするようなことはないかもしれませんが，学術業界では事前に学会にこの宣言を伝えてから研究に取り掛かる，事前登録制 pre-registration が導入されはじめています。

ここで，検定の判定結果に関係してくる要素を改めてみて思い出してみましょう。1 つはもちろん**有意水準**ですね。これは判定を決める基準ですから，これが緩いと対立仮説の勝ち！という結果が出やすいことになります。2 つ目は**サンプルサイズ**ですね。サイズが大きければ区間推定の幅は狭くなりますから，たくさんデータをとれば取るほど対立仮説の勝ち！という結果が出やすいことになります。ここまではお分かりいただけていると思います。続いて 3 つ目は，**効果量 (Effect Size)** です。平均値差の検定の場合を例に簡単に言えば，群間の差の大きさのことです。母集団において本当に平均値の差が大きければ大きいほど，対立仮説の勝ち！という結果が出やすいことになります。これは当然ですよね。ただ，効果量を一般的にいうと**標準化された差**という言い方をします (セクション 5.4，Pp.91 に出てきた話がここにも！)。100 点満点のテストで 1 点の差がある場合と，1000 点満点のテストで 1 点の差がある場合は同列に考えることができないですから，標準偏差で割ることで「標準偏差何個分離れているか」という表現にするのです。標本から推定する効果量としては，**Cohen の d(Cohen's d)** と **Hedges の g(Hedges;s g)** が有名です。Cohens の d は次のようにして計算します。

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}}$$

ここで分母の標準偏差は，2 群で同じであれば問題ないですが，異なる場合は $\hat{\sigma} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$ として合わせて計算します。いずれにせよ，「本当に差が大きければ対立仮説が採択される」というのは当然そうあって欲しい性質ですから，これもお認めいただきやすいと思います。

そして最後にもう 1 つ。先ほどの二種類の間違いのうち，本当に差がある場合に正しく検出できる程度，つまり**検定力 (power)** というのがありました。うっかりミスである**タイプ 2 エラー**を犯す確率が β ですから，検定力は $1 - \beta$ です。これも大きければ大きいほど，正しく対立仮説を採択できます。このように，帰無仮説検定において判定に関係してくるのは有意水準，サンプルサイズ，効果量，検定力の 4 つであり，これらのうち 3 つが決まると残りの 1 つの要素は計算で決まります (水本・竹内, 2011)。また有意水準 α は一般に 5% を使いますし，検定力 $1 - \beta$ も 0.8 ぐらいにするのがよい，というのが通例になっています。であれば，効果量に応じてサンプルサイズを決めることができるわけです。小さな効果しかなさそうなものであればサンプルサイズはたくさん必要ですし，逆もまた真です。ただ，「サンプルサイズは大いに越したことないだろう」というのは間違いで，そうすると実質的にはごくわずかな違いであっても，検定で検出してしまうことになるからです。もっとも，研究を始める前に効果量がどれぐらいなのかを事前に知るのは難しいところです。効果がどの程度あるかがわからないので，研究しようというシーンが多いでしょう。それでも先行研究があれば，そこにはすでにサンプルサイズも，実際のデータによる効果量も計算できるわけですから，効果量がどれくらいある研究なのかを計算することは可能です。

研究を行う前に，事前にこれらの計算をすることを**検定力分析 (power analysis)** と言います。検定力分析は，**例数設計**をするためにも行われる場合と，研究後にどれぐらい効果量があったのかを調べる場合の二種類があります。またこうした計算をするための統計パッケージや R の関数などがありますので，実際はこ

[*8] 実際これを **p-hacking** と呼ばれるよくない研究実践の典型例です。

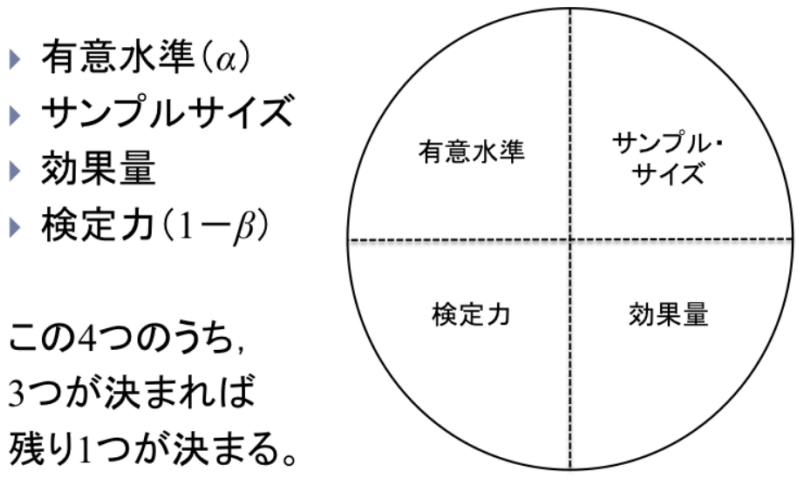


図 19.3 統計的検定における四つの要素の関係

れらを使って計算することになります。

19.5 課題

■検定結果の解釈 帰無仮説を $\rho = 0.0$ とし，標本相関係数 $r = 0.45$ が有意であるかどうかを 5% 水準で検定したところ，対立仮説が採択されました。結果の解釈として，次の文が正しいかどうか判断しましょう。

- 母集団において $\rho \neq 0$ であることが証明された。
- 母集団において $\rho = 0$ である確率は 5% 以下である。
- 母集団において ρ は 0 より大きな値である。
- 母集団において $\rho = 0.45$ である。
- 母集団において $\rho = 0$ であるとするならば，$|r| > 0.45$ という値が得られる確率は 5% 以下である。

■有意水準とサンプルサイズ 次の文章それぞれについて，正しいか間違っているかを判断しましょう。

- 他の条件が一定なら，有意水準が小さくなると有意になりにくくなる。
- 他の条件が一定なら，有意水準が小さくなると第二種の誤りを犯す確率が大きくなる。
- 他の条件が一定なら，サンプルサイズが大きくなると有意になりやすくなる。
- 他の条件が一定なら，サンプルサイズが大きくなると帰無仮説を棄却するときの理論的な統計量の絶対値が大きくなる。

■p 値の性質と意味 次の文章それぞれについて，正しいか間違っているかを判断しましょう。

- p 値が小さいとき，母集団における効果は大きい。
- p 値が小さいとき，帰無仮説が正しい確率は小さい。
- p 値が小さいとき，有意水準は小さい。
- p 値はデータから計算される統計量である。

■検定力分析 次の文章それぞれについて，正しいか間違っているかを判断しましょう。

- 他の条件が一定なら，サンプルサイズが大きくなると検定力も大きくなる。
- 検定力が 0.6 であるとすると，効果を見逃す確率が 0.4 あることを意味する。
- 検定力を求めるには，検出したい効果の大きさを仮定する必要がある。
- 検定力とは標本統計量が，未知母数の値と一致する確率である。

第 20 章

二群の平均値差

心理学の実践で最もよく用いられるのが，平均値差の検定です。実験計画のところで学んだように，RCT と組み合わせて使うことで，結果としてみられる群平均の差が効果の大きさと考えられるからです。本講ではまず，二群の平均値差の検定を考えます。対応がある二群の場合は実はひとつの「差の分布」についての判断になるため，これまでと同様の議論ができるのです。対応がない二群の場合は，線形モデルの最も単純な場合です。要因計画 (線形モデル) と統計的検定の話を接続することで，複数の要因計画へと展開していくことができます。

20.1 一要因 Within デザイン，2 水準

前回までは，帰無仮説検定の考え方とその形式化された手順を説明し，また平均値の検定や相関係数の検定を例に出して具体的な報告の仕方，結果の考え方について論じてきました。

今回は，心理学的な研究シーンで最もよく目にすることのある，2 つの群の平均値差の検定についての話をします。実験の基本的な枠組みは「介入した群としなかった群の違いを，群平均の差でみる」というものです。実験群と統制群という，2 つの群があった時の平均値の「差」が有意であったかどうかを検討するモデルですね。

これを実験計画 (design) の用語を使っていうと，一要因，間要因，2 水準の計画ということができます。要因，群間/群内，水準という言葉に不安がある人は，第 12 講の資料に戻って確認しておいてください。

さてこのセクションでは，まず一要因・群内 (Within) の実験デザイン，中でも 2 水準の計画についてとりあげます。これは群内，言い換えれば反復測定で，その水準が 2 つしかないということは，たとえば介入の前後で効果があったかなかったか，という実験デザインだということになります。細かいことを言えば，群内のデザインは反復測定だけでなく，データに対応がある場合も含まれます。双子の研究や，親密な関係の二人[*1]について研究する場合は，ペアに対応関係があり，相互に影響しあっている可能性が高いので，2 つの変数の間に相関関係があると思われます。こうした状況であれば，とくにその共有している背後の要因にも配慮した分析をする必要があります。群内計画は「対応のある分析」と呼ばれることもあります。

今回はプレ・ポスト (事前・事後) のデータを取った研究だとしましょう。たとえば心理的な健康状態について，事前にアンケート調査をしておき，その後なんらかの臨床的介入 (リラックス法のトレーニングをする，など) を行って，一定期間後に同じ人に事後のアンケートを取る，というような研究デザインです (図 20.1)。

事前に集められた被験者は母集団から無作為に集められているとし，事前調査による健康状態のスコアが正規分布にしたがっていたとします。これと同じテストを事後に行ったとすると，個人差はあるにせよ全体的に

[*1] 恋人同士や夫婦関係のことを，社会心理学では一般的に「親密な関係」と表現します。持って回った言い方ですが，性別に限定されない良い表現だと思います。

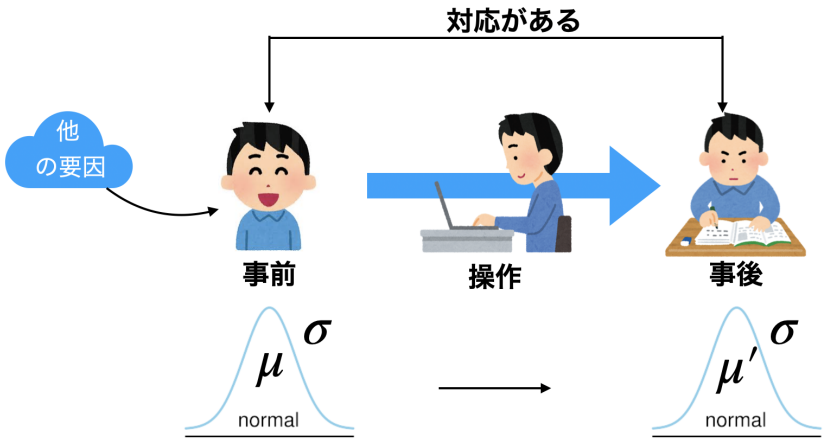


図 20.1 一要因群内 2 水準の実験モデル

効果があれば，この臨床的介入法は優れている，と言えそうです。個々のクライアントの変化も重要ですが，より効果を一般的に解釈するためには，推測統計学の力を借りて，母集団での変化，母平均の差を検証したいと考えるのは当然でしょう。そこで，標本平均から母平均を推定し，また母集団分布が正規分布であるという仮定をおくことで標本平均のブレ幅も想定できますから，これを使って母平均のありそうな場所を推定，それが事前・事後でどう違うかということを考えていきたいと思います。またここで，母平均はもちろん母 SD もわからないので，標本から計算した不偏分散を推定値として考えます。標本統計量から推定する場合は，正規分布ではなく t 分布に従いますので，ここでも t 分布を使った推定をすることになります。この推定値を使った検定を考えるのですが，とくにこの検定のことを **t 検定 (t-test)** といいます[*2]。

さて，事前・事後で変化があったかどうかを知りたいのですが，ありがたいことに正規分布に従う確率変数 X と，正規分布に従う確率変数 Y の差，$X - Y$ も，正規分布に従うことがわかっています。今回はデータに対応があり，事前と事後で変化があったかどうかが知りたいわけですから，この特徴を使うことができます。

数値例として，次のような 2 群を考えてみましょう (表 20.1)。ここには 10 人の実験参加者 (participants) がいます。表 20.1 には，それぞれの人の事前・事後のスコアと，その差分について示されています。数式で表しておきますと，まず事前のスコアを X_1，事後のスコアを X_2 とします。すると差分 D は

表 20.1 Pre-Post Design Data

Participants	1	2	3	4	5	6	7	8	9	10
pre	34	25	13	21	13	26	19	12	19	19
post	35	30	16	27	20	28	16	14	18	21
diff	1	5	3	6	7	2	-3	2	-1	2

$$D = X_2 - X_1$$

ですが，その平均についても

$$\bar{D} = \bar{X}_2 - \bar{X}_1$$

[*2] t 分布を使った検定なので，t 検定です。実は平均の差の検定意外にも t 分布を使った推定・検定は多くあるのですが，非常に応用例の多いパターンですので，2 群の平均値差の検定のことを t 検定ということがあるのです。ちなみにこの時の t は小文字であることにご注意を。大文字の T は別の検定の時に使ったりします。ややこしい。

という関係が成り立ちます。そしてこの D が，平均 μ_D，標準偏差 σ_D の正規分布に従いますから，

$$D \sim N(\mu_D, \sigma_D)$$

より

$$\bar{D} \sim N(\mu_D, \frac{\sigma_D}{\sqrt{n}})$$

ということになります。ここで σ_D がわかればいいのですが，わからないので D の不偏分散をその推定値としますと，

$$t = \frac{\bar{D} - \mu_D}{\hat{\sigma_D}/\sqrt{n}}$$

と自由度 $n-1$ の t 分布に従うことになるわけです。この一連の流れ，数式で表されるとピンとこないという人がいれば，第 17 講に立ち返って，確認してください。

さて，検定の前に区間推定してみましょう。自由度 $n-1 = 10-1 = 9$ の t 分布の 95% 区間は ± 2.262157 ですから[*3]，今回の数字を使って考えると，

$$-2.262 < \frac{\bar{D} - \mu_D}{\hat{\sigma_D}/\sqrt{n}} < 2.262$$

から，両辺に分母をかけて

$$-2.262 \times \hat{\sigma_D}/\sqrt{n} < \bar{D} - \mu_D < 2.262 \times \hat{\sigma_D}/\sqrt{n}$$

$\bar{D}$ を移項目して

$$\bar{D} - 2.262 \times \hat{\sigma_D}/\sqrt{n} < \mu_D < \bar{D} + 2.262 \times \hat{\sigma_D}/\sqrt{n}$$

ここで $\bar{D} = 2.400, \hat{\sigma_D} = 3.062$ ですから，

$$0.2097 < \mu_D < 4.590$$

ということがわかります。標本の差分は 2.4 点だったのですが，母平均の差は点推定で 2.4，95% の区間推定で [0.21,4.59] ぐらいだということでした。つまり，95% の幅で低く見積もっても，0.21 以上ではあるわけですから，差がないわけではない，ということがわかります。

これを検定のロジックに乗せて考えてみましょう。

■仮説の設定 帰無仮説と対立仮説が必要ですが，帰無仮説はないないづくし，前後の差がない，効果がないという仮説になります。記号を使っていうなら，$\mu_D = 0$ です。対立仮説は臨床的な効果があった，ということになりますから，従属変数の差は 0 プラスであるはず。つまり $\mu_D > 0$ です。

■検定統計量を選択 平均値の推定ですので，検定統計量は t ですね。

■判断の基準になる確率を設定 有意水準 α，今回も 5% で勝負を決めるとしましょう。

■検定統計量の実現値を計算 先ほどは区間推定をしたので $\frac{\bar{D} - \mu_D}{\hat{\sigma_D}/\sqrt{n}}$ の計算はしていませんでした。しておきましょう。帰無仮説に従うなら，$\mu_D = 0$ なのですから，

$$t = \frac{\bar{D} - \mu_D}{\hat{\sigma_D}/\sqrt{n}} = \frac{2.4 - 0}{3.062/\sqrt{10}} = 2.479$$

この $t = 2.479$ は自由度 9 の t 分布に従います。この実現値以上の t 値が出てくる確率は，$p = 0.0175$ です。

[*3] 念のために書いておくと，R で `qt(0.975,df=9)` とします。

■最初に定めた仮説を支持するかどうかを判断 判定基準 5% より小さい数字ですので，帰無仮説の下で計算したこの t の値は，珍しいと言えます。こんな珍しいことが生じたのだから，仮定が何か間違っている。つまり，差がないというのがおかしい。帰無仮説を棄却して，対立仮説を採択する，と判断します。

結果の報告の仕方は，

> 対応のある t 検定を行ったところ，$t(9) = 2.479, p = 0.0175$ で統計的に有意である。

ということになります。

ところで，先ほどの区間推定の結果を覚えているでしょうか？ 95% の幅で考えると，少なく見積もっても 0.21，大きく見積もると 4.59 も差があったかもしれません。この「少なく見積もっても」の方がゼロ (=帰無仮説) ではありませんので，この区間をみるだけでも帰無仮説を棄却することがわかります。この区間推定がゼロを挟む，つまり少なく見積もるとマイナス (逆効果！)，大きく見積もるとプラスだ，というときは検定をしても有意にはなりません。この区間推定の情報は，推定された平均値がどれほど変化しうるのか，の情報でもありますから，この大きさを一緒に報告しておくと「この実験がどれほど信用できるか」という目安にもなって良いですね。そこで先ほどの報告例を少し修正して，

> 対応のある t 検定を行ったところ，$t(9) = 2.479, p = 0.0175, 95\%C.I.[0.21, 4.59]$ で統計的に有意である。

とすると尚良いでしょう[*4]。

20.2 一要因 Between デザイン，2 水準

20.2.1 区間推定でまず考えてみよう

それでは次に，対応のないバージョン，つまり群間 (Between) 計画の例を考えてみましょう。これは両群に対応がない，それぞれの平均値の推定をして，その上でその推定値の間に差があるかどうか，という話になります。

具体的な数値例でみてみましょう。今回は $n = 5$ としました。5 人の実験参加者を実験群 (experimental) と統制群 (control) に無作為に割り付け，実験群にはある認知課題を課した後で，記憶力を測定するテストをした，と思ってください。テストのスコアが value 列に入っている数字です。さて，認知課題をした群とそうでない群の平均値に違いはあるでしょうか？

これまた母 SD がわからない中で，標本平均から母平均を推定するので t 分布を使うことになります。実験群，統制群ともに $n = 5$ ですから，自由度 $n - 1 = 5 - 1 = 4$ の t 分布を使って推定すると，統制群の母平均は $\bar{X}_{ctl} = 20.60$,$\hat{\sigma_{ctl}} = 4.393$ より，

$$15.145 < \mu_{ctl} < 26.055$$

となります。実験群は $\bar{X}_{exp} = 25.20$, $\hat{\sigma_{exp}} = 6.221$ より

$$17.476 < \mu_{exp} < 32.924$$

となります。

[*4] 5% 水準でカットするのも，95% 水準の幅を報告するのも，同じ情報で冗長だ，と思われるかもしれません。実にその通りで，かつ，帰無仮説検定の「有意差あり」という報告は信頼区間をみればわかる一方，その逆は成立しない (p 値は分布に関する情報ではない) のですから，有意差あり/なしという言い方は必要ないのです。今でもこの表現が使われるのは，旧世紀の名残でしかありません。

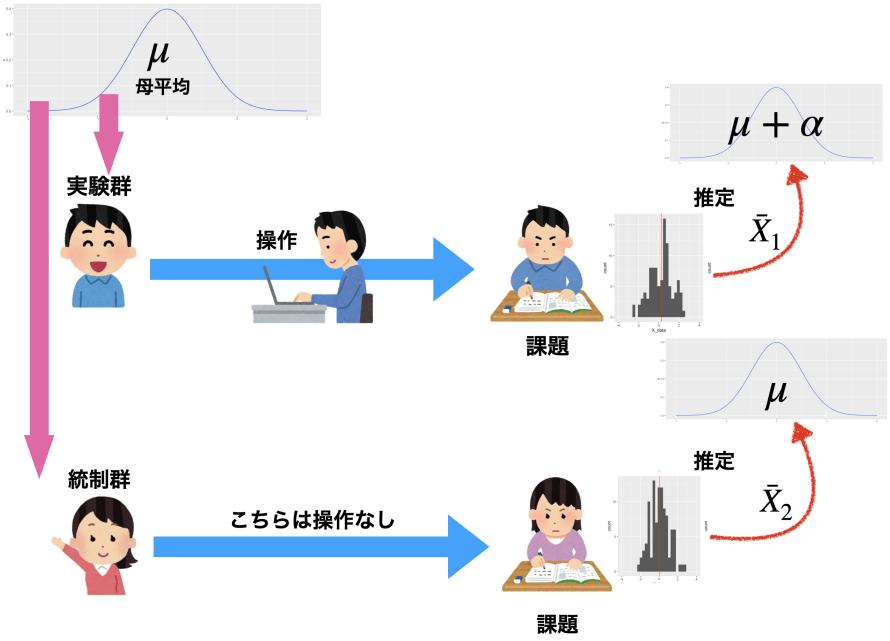


図 20.2 図 17.7 の再掲。一要因 Between デザイン，2 水準の場合。

表 20.2 Between Design のデータ例

ID	condition	value
1	control	25
2	control	15
3	control	24
4	control	17
5	control	22
6	experimental	31
7	experimental	28
8	experimental	30
9	experimental	19
10	experimental	18

差がありますか？ 微妙ですね！統制群は大きく見積もるなら 26.05 ぐらいの標本平均が出てきたかもしれなかったし，実験群は小さく見積もるなら 17.48 ぐらいの標本平均だったかもしれないわけです。区間推定だけだと判断に迷うところです。

20.2.2 検定で勝負を決めようじゃないか

そこで検定の出番。先ほどの例と同じく，「正規分布と正規分布の引き算はまた，正規分布」という特徴を利用して検定を行います。正規分布である母集団 $N(\mu, \sigma)$ から無作為標本を取ってきたとすると，標本平均 $\bar{X}$ は $N(\mu, \frac{\sigma}{\sqrt{n}})$ に従うのでした。ここで，2 つの群の標本平均の差，$\bar{X}_1 - \bar{X}_2$ を考えると，その分布は平均 $\mu_1 - \mu_2$，標準偏差 $\sigma / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ に従うことがわかっています。ただ，この σ がわかりませんので，不偏

推定量で置き換える必要があります。今回は 2 つの群があって，実験群の不偏分散を使っても，統制群の不偏分散を使ってもおかしいので，両群をまとめて

$$\hat{\sigma^2}_{pooled} = \frac{(n_1-1)\hat{\sigma_1^2} + (n_2-1)\hat{\sigma_2^2}}{n_1+n_2-2}$$

を使います。pooled とはプールした，まとめた，という意味です。これは分散の推定量ですから，差の分布の標準偏差の推定値としてはこれの正の平方根，

$$\hat{\sigma}_{pooled} = \sqrt{\frac{(n_1-1)\hat{\sigma_1^2} + (n_2-1)\hat{\sigma_2^2}}{n_1+n_2-2}}$$

を使えば良いことになります。

ややこしくなってきましたが，これを使って考えると統計量は次のように計算できます。

$$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)\hat{\sigma_1^2}+(n_2-1)\hat{\sigma_2^2}}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

あっ，心閉ざさないで！数式は話の筋を通すために書いているだけです。文字がいっぱい出てきて混乱したかもしれませんが，この数式，覚えておく必要はありません。ここでは，教材なので一応説明していきますが，実際には統計ソフトウェアが組み込まれた関数で一気に答えを出してくれます。大事なのは，筋道を理解することですので，初めのうちは数式のところは「なんか書いてるわ」ぐらいの理解でも構いません。それよりも，この t の式には標本統計量しか入っていないことに感動しましょう。

「うそつけ，μ_1, μ_2 があるじゃないか」と思った人は，目の付け所がいいですね。確かにこれはわからないところ (だからギリシア文字で書いてあります) なのですが，これを使って今から検定を行うのです。検定では帰無仮説をおきます。帰無仮説はないないづくし，ですから，「2 群の母平均の間に差はない」と仮定するのです。これを言葉でいうと，$\mu_1 = \mu_2$，あるいは $\mu_1 - \mu_2 = 0$ です。上手くできてますよね！帰無仮説の下では，分母の $\mu_1 - \mu_2$ を $= 0$ とおくことができますので，

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)\hat{\sigma_1^2}+(n_2-1)\hat{\sigma_2^2}}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

となり，完全に標本統計量から計算できることになりました。それでは検定を進めて参りましょう。

■仮説の設定　帰無仮説と対立仮説が必要です。先ほども述べたように，帰無仮説は $\mu_1 = \mu_2$，対立仮説はこれの否定ですので $\mu_1 \neq \mu_2$ になります。

■検定統計量を選択　平均値の推定ですので，検定統計量は t ですね。

■判断の基準になる確率を設定　有意水準 α，今回も 5% で勝負を決めるとしましょう。

■検定統計量の実現値を計算　これはちょっと厄介ですが，数式にしたがって，統制群を X_1，実験群を X_2 とおくと，$\bar{X}_1 = 20.60, \hat{\sigma}_1 = 4.393, \bar{X}_2 = 25.20, \hat{\sigma}_2 = 6.221$ です。

$$t = \frac{20.60 - 25.20}{\sqrt{(4.393^2 + 6.221^2)/5}} = -1.3506$$

となりましたので，これが自由度 $(n_1-1)+(n_2-1) = n_1+n_2-2 = 10-2 = 8$ の t 分布に従うことになります。これより極端な t 値が得られる確率は，$p = 0.1068$ になりました。

20.2.3 検定と方向性

ここで注意が必要です。今回の帰無仮説は，2 群の母平均が等しくないというものでした。等しくないというのは，$\mu_1 > \mu_2$ でもいいし，$\mu_1 < \mu_2$ でもいいはずです。今回は検定統計量の実現値を計算するときに，分子を $\bar{X}_1 - \bar{X}_2$ で計算しましたが，$\bar{X}_2 - \bar{X}_1$ でもいいはず。そうすると，後者の場合 t の値は 1.3506 になることになります。

t 分布を見ながら，先ほどの t 値の考え方を見てみましょう。t の実現値が今回は-1.3506 だったわけですが，それ以上に極端な数字が出る確率を考えるときは，符号を考えずに `1-qt(1.3506,df=8)` で計算していました。この R での計算は，図でいうなら図 20.3 の上にあるように，1.3506 までの面積を 100% から引き算することによって求めた，青い色付けられた領域になっています。反転の反転，で説明がややこしくなっていますが，`qt(-1.3506,df=8)` と同じことになっています。

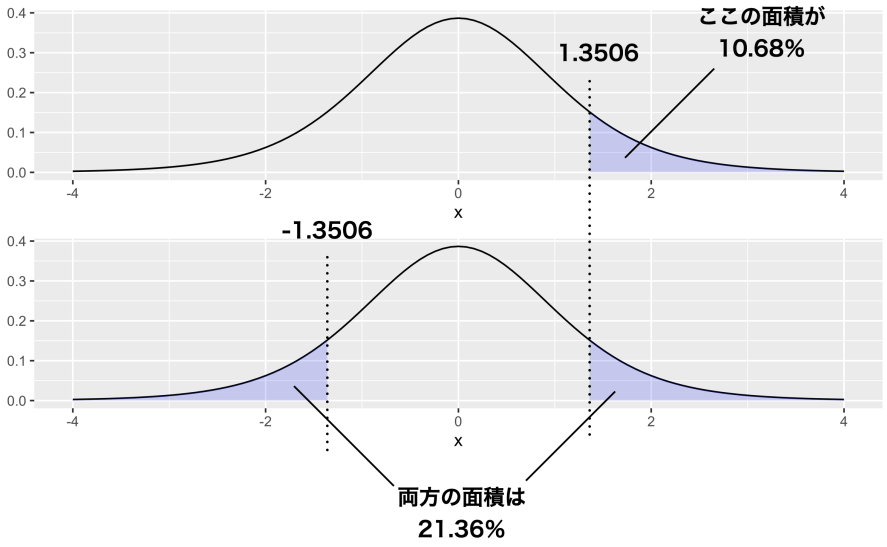


図 20.3 片側検定 (上) と両側検定 (下)

そしてその t 値は，統制群の平均値から実験群のそれを引くか，実験群から統制群を引くかによって符号が変わるものであり，今回の仮説はそのどちらにするかを考えていなかったので，どちらでもあり得る，つまり面積的にはこの倍の出現の可能性がある，と考えるべきなのです (図 20.3 の下)。

これは検定の「方向性」，といわれ，「一方が他方より大きいはず」という意味があるのなら，片方の面積を考えるだけで構いません。これは**片側検定 (one sided test)** と言われ，前のセクションで対応のある t 検定をした時は，臨床的介入の「正の」効果があった，と考えた片側検定をしていたのです。もし臨床的介入が悪化させることもある，という可能性があれば，面積を倍にして考えなければなりませんでした[*5]。

今回は，実験群と統制群の「違い」だけに興味があり，実験的な操作が成績を上げるのか下げるのかわからない，とにかく違うに違いない，と考えるので**両側検定 (two sided test)** であり，t の出現確率は $p = 0.2136$ ということになります。

まとめましょう。報告する時は次のようになります。

[*5] その場合でも，$0.0175 \times 2 = 0.035$ で 5% 水準を下回り，有意であったという結論は変わりませんが。

実験群と統制群の平均値の差を検定したところ，$t(8) = 1.356, p = 0.214$ で統計的に有意であるとは言えなかった。

2 群の平均値の差の検定については，このような手続きで行います。繰り返しになりますが，実際の計算は機械がやってくれるので，どのような数字がどのような意味を持っているのか，得られた結果からどのように判定するのか，という手続き的な知識をしっかりと踏まえてください。

また，ここでこのように検定の具体例をみてきたわけですから，是非この技術を身の回りでも活用してみてください。何か 2 つのグループがあって，その平均値が違うのであれば，これは偶然によるものなのかそうでないのか。一般化して良いのかどうか。この推測統計学的発想は，身の回りの至る所に応用できるものです。知識と技術を手に入れたら，是非具体例を自分で考えて，試してみてください[*6]

20.3 R による実践；対応のある t 検定

では具体的に，R を使って計算してみましょう。まずは対応のある **t 検定 (t-test)**，つまり一要因群内 (Within)2 水準の平均値差の検定です。面倒な計算式がありましたが，今回その辺の計算は機械がやってくれますので，例によって出力をしっかり確認するようにしましょう。

例として，表 20.1 の 10 人の参加者データを使います。元になるデータをオブジェクトに格納する code:20.1 を実行しましょう。

code：20.1 t 検定を行う

```
1  X <- c(34, 25, 13, 21, 13, 26, 19, 12, 19, 19)
2  Y <- c(35, 30, 16, 27, 20, 28, 16, 14, 18, 21)
3  t.test(X, Y, paired = T, alternative = "less")
```

■コード解説

1 行目 一方のデータをオブジェクト X に代入しています。c は結合 combine という関数です。

2 行目 他方のデータをオブジェクト Y に代入しています。

3 行目 t 検定を行っています。引数として，データ X,Y を指定し，次に paired オプションを T に[*7]，対立仮説 alternative を「X が Y よりも小さい」という片側検定にしてあります。

さて，このスクリプトは直感的にわかりやすく書きましたが，同じことを別の表現をすることもできます (code:20.2)。

code：20.2 t 検定その 2

```
1  X <- c(34, 25, 13, 21, 13, 26, 19, 12, 19, 19)
2  Y <- c(35, 30, 16, 27, 20, 28, 16, 14, 18, 21)
3  dat <- data.frame(condition = c(rep(1, 10), rep(2, 10)), value = c(X, Y))
```

[*6] たとえば「お兄ちゃんのフライドポテトの方が，いつも長いのばっかり入っていてずるい！」というような兄弟喧嘩の種も，両者のフライドポテトの長さを測り，その平均値を出して t 検定すれば「たまたま今回のポテトの問題だった」のか，いつもそうなる現象なのか検定で決着をつけることができます。実際に筆者が居酒屋のフライドポテトでこれの検証をした時，テーブルごとのポテトの長さに平均的な差はありませんでした。くわえて，フライドポテトの長さが非常に正規分布に近かったのを知って，感心したことがあります。

[*7] 大文字の T は $TRUE$，真であり，この逆は F とかいて $FALSE$，偽を意味する**特別な文字**です。真か偽か，本当か嘘かという 2 つの状態を表す論理値と呼ばれるもので，ここではスイッチオン/オフの意味です。対応スイッチ，オン！という意味ですね。大文字の T, F をオブジェクト名にして利用することも可能ですが，R が専門的な意味で使う予約語なので，一般的にはそのような使用は避けた方が望ましいです。

```
4  dat$condition <- factor(dat$condition, labels = c("pre", "post"))
5  t.test(value ~ condition, data = dat, paired = T, alternative = "less")
```

これは 1，2 行目は先ほどと同じですが，

3 行目　オブジェクトからデータフレームを作成。データの値は変数 value に，これを条件ごとに区別する条件変数 condition と共に作ります。condition の中にある rep 関数は，ある文字・数字を繰り返すもので，ここでは rep(1,10) で 1 を 10 個作っています。ここだけ実行するとその動作がよくわかりますよ (R **の出力** 20.1)。

R の出力 20.1: rep の出力

```
> rep(1,10)
[1] 1 1 1 1 1 1 1 1 1 1
```

4 行目　先ほど作った condition 変数をファクター型 (要因型，ラベル付き名義尺度水準) に置き換えています。

5 行目　t 検定ですが，書式が変わっています。チルダ ~ を使って，従属変数~独立変数という書き方に変えています。

こちらはデータフレームを作った書き方ですから，データフレームがどういう形なのかみておくと良いでしょう (R **の出力** 20.2)。

R の出力 20.2: dat の中身を確認

```
> dat
   condition value
1        pre    34
2        pre    25
3        pre    13
(中略)
11      post    35
12      post    30
13      post    16
(後略)
```

実際のデータ分析の時は、外部からデータファイルを読み込んだり，既にデータフレームに整形されたものを分析することが多いかと思います。このチルダを使って独立変数，従属変数の形に表すのが R では一般的な関数関係の表現形ですから，これを機にどちらの使い方でもできるように解説しました。

さて，こうして得られる結果は R **の出力** 20.3 の通りです。

R の出力 20.3: t.test の出力

```
Paired t-test
data:  value by condition
t = -2.4783, df = 9, p-value = 0.01754
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
        -Inf -0.6248331
sample estimates:
mean of the differences
                  -2.4
```

上から三行目に t 値が出ていますね。帰無仮説の下での t 統計量の実現値は $t(9) = -2.4783$ であり，これが出現する確率は $p = 0.01754$ です。対立仮説は，平均値の差が 0 より小さい (true difference in means is less than 0) という対立仮説になっています[*8]。これが片側検定になっているのは，`t.test` 関数の引数に `alternative="less"`と書いたからで，これを `alternative="greater"`とすれば逆の片側検定 (事前データのスコアの方が事後データのスコアより大きい) を検定することになり，とくに何も書かないか `alternative="two.sided"`とすれば両側検定になります。

前の章 (→ 第 19 講) を参考に，このような結果をどう解釈し，どうレポートに記載するかを確認しておいてください。

20.4 R による実践；対応のない t 検定

つづいて対応のない **t 検定 (t-test)** のほうも，R でできるようになりましょう。

さきほどの表 20.2 のデータを分析することを考えます。難しい計算は抜きにして，`code:20.3` を入力し，分析を実行してください。

code : 20.3 formula で指定

```
1  X <- c(25, 15, 24, 17, 22)
2  Y <- c(31, 28, 30, 19, 18)
3  dat <- data.frame(condition = c(rep(1, 5), rep(2, 5)), value = c(X, Y))
4  dat$condition <- factor(dat$condition,
5                          labels =  c("control","experimental"))
6  t.test(value ~ condition, data = dat)
```

スクリプトの内容については、先ほどとほぼ同じです。大きな違いは，`t.test` 関数の引数がなくなったところで，`paired` オプションもなければ，`alternative` オプションもありません。ないのは、デフォルトの状態を使っているからで，R の `t.test` 関数はデフォルトで対応のない，両側検定の t 検定をやってくれるのです。

結果は次のように表示されるはずです (R の出力 20.4)。

[*8] R では前の変数が後ろの変数に比べて大か，小かを表現します。

R の出力 20.4: Welch の補正

```
Welch Two Sample t-test

data:  value by condition
t = -1.3506, df = 7.195, p-value = 0.2178
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.609577   3.409577
sample estimates:
     mean in group control mean in group experimental
                      20.6                       25.2
```

前回, 手計算した結果と同じになっているかどうか確認してください。

んん？ 自由度が 7.195? 自由度って 8 になるんじゃなかったでしたっけ (実験群の自由度が $5-1=4$, 統制群の自由度も $5-1=4$, 足して 2 つで $4+4=8$)。

20.4.1 Welch の補正

そうです，t 検定の自由度は 8 のはずなのですが, 変な数字で出ていますね。これは, 検定に関わるまだ説明していない別の手続きがあったからです。

検定にはどのような仮定があったか, 思い出してください。母平均の差をゼロとする, という帰無仮説をおくのも仮定ではありますが, さらにその前にあった仮定です。それは，2 つの群が正規分布から得られていること, とくに平均値に差はあるけれども, 母標準偏差は同じところから取ってきていた, という仮定があったのです。

もしこの仮定が崩れていたら？ 母集団に正規分布が仮定できなければ，t 検定はできません。たとえば年収の差とか, データが比率や角度についての数字であったりすると, 正規分布を仮定できないので t 検定はできないのです。正規分布は正規分布だったとしても, 母標準偏差がちがっているなら, 推定の仕方も考え直さないといけないかもしれません。

この問題は，t 値の自由度を調整することで解決されています。前回までのいわゆる t 検定は, 等分散の仮定が満たされているケース (分散の均一性などということもあります) の計算方法であり, この仮定が満たされていない場合は **Welch の補正 (Welch's correlation)**(ウェルチの補正) とよばれる調整が必要です。

R の `t.test` 関数は, より仮定の緩やかな，Welch の補正をかけた結果をデフォルトで返してくれます。まず分散が等しいかどうかチェックして云々, という手続きが面倒なので, 最初から補正しておこう, ということですね。結果の一行目にも Welch Two Sample t-test と書いてあります。これを使って, 結果は

$$t(7.195) = 1.351, p = 0.2178 \text{ で 2 群の間に統計的に有意な差は認められなかった。}$$

と報告します[*9]。差があるかどうかの判断だけでなく, 差の 95% 信用区間にも目を向けておきましょう。結果から, 差の 95% 信頼区間は CI[-12.609577,3.409577] でした。差は-12 から +3 までの幅があり, 帰無仮説である差 $=0$ を含んでいるので, 有意になりません。

ところで, 分散が同じだ, と仮定できる場合は `t.test` 関数のオプションを使います (code:20.4)。

[*9] t の値は, 両側検定の場合符号に意味がないのでマイナスを取って記載します。

code : 20.4 分散のオプション指定

```
1 t.test(value ~ condition, data = dat,var.equal = T)
```

これで，R の出力 20.5 のような結果を得ることができます。

R の出力 20.5: 分散が等しいという仮定を置く

```
 Two Sample t-test

data:  value by condition
t = -1.3506, df = 8, p-value = 0.2138
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.453967   3.253967
sample estimates:
     mean in group control mean in group experimental
                      20.6                      25.2
```

このやり方ですと，自由度も整数になっていますね (判定結果は変わりませんが)。

20.4.2 事後的な効果量の検証

検定が終わった後ではありますが，このデータの実験は「勝ち目がある」ものだったのでしょうか。つまり，母集団における母平均の差が十分大きなものなのかどうかが気になります。母数は分かりませんが，これまでも分からないものは標本統計量で置き換えて推定してきたわけですから，データを取った後になって，これはどれぐらい効果量の大きなデータだったのか，を考えてみましょう。

効果量の計算についても，R を使って簡単に行うことができます。R ではパッケージを追加することでさまざまな統計関数が増えますが，効果量についての effsize パッケージ (Torchiano, 2020) を導入すれば，この計算も簡単にできます。効果量については第 19 講で触れましたので確認しておいてください。ここではより推奨される，Hedges の g の算出の仕方を紹介します (code:20.5)。

code : 20.5 効果量を求める

```
1 # パッケージの読み込み
2 library(effsize)
3 cohen.d(value ~ condition, data = dat, hedges.correction = T)
```

ここではまず 2 行目でパッケージを読み込み[*10]，次の行で cohen の d を Hedges の補正をかけて算出せよ，としています。

結果は R の出力 20.6 のように示されます。

[*10] このパッケージを持っていない場合は，当然インターネットからダウンロードしてローカル (手元の PC) に保存しておく必要があります。

R の出力 20.6: 効果量の出力

```
Hedges's g

g estimate: -0.7715342 (medium)
95 percent confidence interval:
       lower      upper
-2.1369693  0.5939009
```

g estimate: -0.7715342 (medium) とあるように, 推定値は 0.77 で, これは中ぐらい (medium) の大きさだよ, ということです。中程度の違いがありそうなのに, 今回有意にならなかったのは、サンプルサイズが小さかったからかもしれません。事前にきちんと設計してやっておけば, 実験手続きとしてより良くなったでしょう。

効果量の計算はこのように, 事後的に行って今回の標本から標準化された平均値の差がどれぐらいありそうか, と考えることもできます。もちろん事前に平均値の差を想定して, そこから例数設計 (どれぐらいのサンプルを取れば良いか考える) するのにも使います。今回ここで求めた標本効果量も, レポートに書くようにしましょう。

> $t(7.195) = 1.351, p = 0.2178$ で 2 群の間に統計的に有意な差は認められなかった。なお, Hedges の g で効果量を求めたところ, 0.77 であった。

20.4.3 補遺

付録的に 2 点付け加えます。時間的に余裕がある・興味がある場合はご一読ください。

■対応のある t 検定の効果量を計算する 対応のある t 検定をして, その効果量を求める場合は, 次のように paired オプションをオンにしておきます (code:20.6)。

code : 20.6 cohen の d

```
1  cohen.d(X,Y,paired = TRUE)
```

ただし, チルダを使った書き方をしている場合, code:20.7 のようにすると, 出力に次のような警告が付されます (R の出力 20.7)。

code : 20.7 cohen の d その 2

```
1  cohen.d(value ~ condition, data = dat, paired = TRUE)
```

R の出力 20.7: rnorm の出力

```
Cohen's d

d estimate: -0.344241 (small)
95 percent confidence interval:
      lower       upper
-0.64457930 -0.04390272

    警告メッセージ:
cohen.d.formula(value ~ condition, data = dat, paired = TRUE) で:
Trying to compute paired samples Cohen's d using formula input. Results may be incorrect
 if cases do not appear in the same order for both levels of the grouping factor. Use the
format 'value ~ treatment | Subject(id)' to specify a subject id variable.
```

この警告は,「対応のある個々人を識別する変数がデータセットに入ってないので, 間違った結果になっているかもしれませんよ。次の書式を使ってください」と言ってます。これに対応するためには、code:20.8 のようにします。

code : 20.8 対応のある場合の cohen.d

```
1  X <- c(34, 25, 13, 21, 13, 26, 19, 12, 19, 19)
2  Y <- c(35, 30, 16, 27, 20, 28, 16, 14, 18, 21)
3  dat <- data.frame(condition = c(rep(1, 10), rep(2, 10)), value = c(X, Y))
4  dat$condition <- factor(dat$condition, labels = c("pre", "post"))
5  dat$ID <- rep(1:10,2)
6  cohen.d(value ~ condition | Subject(ID), data = dat, paired = TRUE)
```

ここでのポイントは 5 行目に ID を識別する変数を作り[*11], 効果量を計算する式を警告の提案通りのフォーマットに書き換えています。

■例数設計のしかた 対応のない t 検定で, 効果量が 0.7 あるのに有意になりませんでした。これはサンプルサイズが小かったからかも, ということでしたが, では何人分のデータがあればよかったのでしょうか。code:20.9 ように R に入力することで, そのヒントが得られます。

code : 20.9 power.t.test 関数

```
1  power.t.test(power = 0.8, delta = 0.7, sig.level = 0.05)
```

ここで, delta=0.7 としたのは, 標準化された平均値差が 0.7 あるとして, という考え方です。これは先ほどの事後的検定力分析の結果, それぐらいあったんじゃないの？ と思われたのでその値を使いました。もっと小さく見積もったり, 逆にもう少し大きく見積もったりしても構いません。また, power=0.8 としましたが, これは**検定力 (power)** で, セクション 19.3.1(Pp.246) で出てきたものです。Type II Error, つまり差があるのにあると言えない間違い (β) を 20% ぐらいに抑え, 80% の確率で当てたいとしました ($1-\beta=0.8$)。また, 有意水準は 5% です ($\alpha=0.05$, スクリプトの中では sig.level=0.05 としています)。

さて, これの出力結果を見ると, 次のように示されました (**R の出力 20.8**)。

[*11] 関数 rep で 1:10 を 2 回繰り返しています。R ではコロンで繋げた数字は, 連番を意味します。

R の出力 20.8: rnorm の出力

```
Two-sample t test power calculation

n = 33.02467
delta = 0.7
   sd = 1
sig.level = 0.05
power = 0.8
alternative = two.sided

NOTE: n is number in *each* group
```

ここの n にあるように，効果量 0.7，検定力 0.8 で研究するのであれば，33 か 34 人は集めたいところ，だったわけです。逆に今回は 5 人しか集めていませんでしたが，これを使って，code:20.10 とすると，**R の出力 20.9** のように示されます。

code：20.10　power.t.test で n を固定

```
1  power.t.test(n = 5, delta = 0.7, sig.level = 0.05)
```

R の出力 20.9: 検定力の出力

```
Two-sample t test power calculation

n = 5
delta = 0.7
sd = 1
sig.level = 0.05
power = 0.16318
alternative = two.sided

NOTE: n is number in *each* group
```

ここにあるように，検定力は 0.16，つまりサンプルサイズ $n = 5$ であれば，この研究は 100 回やっても「有意差あり」と正しく判断できるのは 16 回ほどしかなかった，ということになります。分の悪い勝負に出たというところでしょうか。

20.5 課題

次の計算をする R コードが含まれる Rmd ファイルを提出してください。なお提出されたコード単体でバグがなく動くことが確認できないものは，未提出扱いになります。コードの書き方などわからないところがあれば，曜日別 TA か小杉までメールで連絡し，指導を受けてください。

■課題 1 対応のある t 検定　表 20.3 にあるように，ある介入の事前・事後の心理検査の成績が得られたとして，次の各指示に従い，結果について回答してください。

1. 事前スコアの方が事後スコアよりも低いかどうか，5% 水準で検定してください。

表 20.3 課題 2

pre	19	38	24	25	12	28	11	17
post	31	23	28	35	14	4	4	38

■課題 2 対応のない t 検定 表 20.4 にあるように，統制群と実験群の心理検査の成績が得られたとして，次の各指示に従い，結果について回答してください。

表 20.4 課題 3

Control	20	21	21	20	21	22	22	22	20	20
Experimental	20	18	22	20	21	18	19	21	21	21

1. 母分散が等しいと仮定した t 検定を実行してください。
2. 母分散が等しいという仮定をおかない t 検定を実行してください。
3. この検定に関して，効果量として Hedges の g を算出してください。

■課題 3 基礎実験 1 のデータを使って 基礎実験 1 の授業の中で，社会的影響についてのデータを取りましたね。そこでは，「他者の存在の有無」により，「課題遂行量」に違いが出るのかどうか，を検証したとおもいます (他にも変数をとったと思いますが，ここでは課題遂行量にのみ注目します)。そのデータを使って，次の問いに答えてください。なお，計算に使用したデータ (あるいはデータファイル) も合わせて提出してください。

1. 他者がいる群といない群，それぞれの標本平均値，標本分散，不偏分散をそれぞれ算出してください。
2. このデータセットが正規分布に従う母集団から得られたと考えて，各群の母平均について 95% の区間推定をしてください
3. このデータセットが正規分布に従う母集団から得られたと考えて，二群の平均値の差の検定をしましょう。その際，帰無仮説と対立仮説はそれぞれどのようなものになりますか。
4. 二群の平均値の差の検定をし，検定統計量とともに判断結果を報告してください。
5. その検定に関して，効果量として Hedges の g を算出してください。

第 21 章

要因計画と平均値差の検定

ここまでは主に，平均値の比較に確率的判断の手続きを加えた仮説検定の話をしてきました。ここでは前期の最後に学んだ，線形モデルから要因計画へと進んだ話に，検定の話を統合します。線形モデルに確率的判断の手続きを加えたものが，**分散分析 (ANalysis Of Variance;ANOVA)** と呼ばれる独自の体系になります。

今回の章のあらすじを先にまとめておきます。

1. 回帰分析における独立変数が離散的であれば，要因計画となる。回帰分析も要因計画もいずれも線形モデルで表現できるので，まとめて**一般線形モデル (General Linear Model)** と呼ぶ。
   - 全体平均からの群平均の差分を効果と呼ぶ (→ 第 12 講, セクション 12.3.1, Pp.182)。
   - 効果の全体を検証するためには**平方和 (Sum of Squares:SS)** を計算するのでした (セクション 12.4, Pp.184)。
   - 要因の数が増えるのは**重回帰分析**になるようなもの。ただし交互作用には注意して (→ 第 13 講, セクション 13.3, Pp.192)。
   - 群内要因であれば，個人差の分散 (平方和) も算出できる (→ 第 14 講)。
2. 平方和の比率が従う分布として F 分布というのが知られており，これを使うと標本の**誤差に対する効果の大きさ**がどれほど珍しいものであるのかを検証できる。
3. 帰無仮説が棄却されたことが，どの群間差で生じたのかを明らかにするために，**多重比較 (multiple comparison)** を行うことでさらに精緻に分析する。

ここにあるように，まずは前期でやった線形モデルのことをよく思い出していただく必要があります。その上で，効果や誤差の「大きさ」として計算された平方和の比が従う分布を利用して，検定と接合することになります。検定は帰無仮説と対立仮説の勝負による判定にすぎないことに注意しつつ，また何が帰無仮説として置かれていたのかに注意して読み進めてください。

21.1 平均値の比較とモデル比較

後期に入ってから，確率の使い方が母集団を対象にしたものに変わりました。これまでは誤差が正規分布するから，という理由で確率が導入されていましたが，さまざまな誤差に影響された誤差まみれの数字は，正規分布するという正規分布の別の側面からアプローチしてきたのです。また，母集団と標本という区別をおくことで，母集団の分布についての仮定から，標本統計量の分布が考えられ，それを使って母数を推測するという方法を取りました。母数を推測するだけでなく，実践的な判断をする帰無仮説検定という手続きについても，解説してきました。

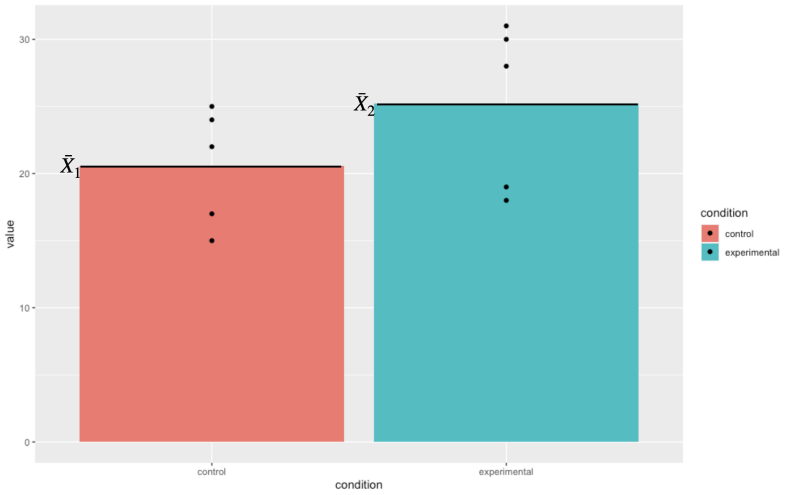


図 21.1　2 つの群の平均値のプロット。黒い点は各群のそれぞれの観測値

さて，最初に帰無仮説検定の話をしたときに，これはモデル比較だと言いました (→ セクション 18, Pp.233, およびセクション 18.1.3,Pp.235)。帰無仮説というモデル，対立仮説というモデルを比較して，どちらに軍配を上げるかという勝負判定をする，これが帰無仮説検定だと。

ここでの「モデル」という言葉は，たとえば帰無仮説は $\mu_1 - \mu_2 = 0$ とおく，といった表現がなされますから，前期の (線形) モデルという表現と異なるイメージを持ったかもしれません。しかしこれもモデル，しかも線形のモデルなのです。今からこの「帰無仮説検定」と「線形モデル」を概念的に統合することを説明してきます。

そのための例として，次のようなデータを用意しました (図 21.1)。これは対応のない，2 つの群 (仮に実験群と統制群とします) の平均値をプロットしたものです。ここで示されている平均値は標本平均ですから，そこに差があるのは見た通りです。ちなみに黒い点は各群における個体 1 つ 1 つの観測値です。

21.1.1 神様の目から見た実験計画

ここで，同じデータを別の見方で捉えてみましょう。

今回の例において，従属変数である心理変数はどういう特徴なのか，目に見えないものなのでわかりません。でも，心の問題は無数の要因が積み重なった結果であると考えると，正規分布を仮定するのが自然です。つまり，実験群のひとも統制群のひとも，大きな視点から考えれば同じ人間という正規母集団からのサンプルなのです。

さて，実験参加者たちは実験者によって無作為に実験群と統制群に割り当てられました (RCT)。個人差がなければ，各群 1 名ずつのスコアを比較して考えれば良いのですが，人間を対象にしている以上そうはいきません。個人差がどうしても生じてしまうので，複数人を集めて個人差を相殺することを考えます。各群の平均値の違いを見ることで，操作の効果を見ることにするのです。

もしこの実験に効果がなければ，2 群の平均値は同じになるでしょう。この時の平均値は，実験群と統制群に分けた効果がないのですから，両群のスコアを足しあわせた全体平均になるはずです。効果がない=全体平均，なのです。しかし，群ごとに見てみると全体平均から違いが生じています。図 21.2 では，統制群は相対的に低く，実験群は相対的に高い平均値を示しています。この差分こそ (図 21.2 では短い緑の矢印で示しています)，実験の効果なのです。

さて，群ごとの平均は次のように書くことができます。

$$\bar{X}_1 = \frac{1}{5}\sum_{i=1}^{5} X_{i1}$$

$$\bar{X}_2 = \frac{1}{5}\sum_{i=1}^{5} X_{i2}$$

また，全体平均次のように書くことができます[*1]。

$$\bar{X}_{gm} = \frac{1}{10}\sum_{i}\sum_{j} X_{ij}$$

これは標本を使った平均値ですが，本来は推定したい (群間の差がない) 母平均について議論したいところです。$\bar{X}_{gm}$ は母平均の推定値として使い道はありますが，理想化・抽象化して語れるのがモデルの良いところですから，ここでは全体平均を未知数を表すギリシア文字 β をつかって β_0 としておきましょう。

さて次に，群わけの効果を β_1 としましょう。すると各群の平均のモデルは

$$\mu_1 = \beta_0 + \beta_1$$

$$\mu_2 = \beta_0 - \beta_1$$

と表すことができます。左辺も各群の母平均なので，未知数を表すギリシア文字になっています。

つづいて，ここでの個人差はこの実験で制御できない誤差であり，群平均を中心にどういう向きで出るかわからないものですから，このことを表現するなら，

$$X_{ij} = \mu_j + \varepsilon_{ij}$$

ということになります。これらを組み合わせると，

$$X_{i1} = \beta_0 + \beta_1 + \varepsilon_{i1}$$

$$X_{i2} = \beta_0 - \beta_1 + \varepsilon_{i2}$$

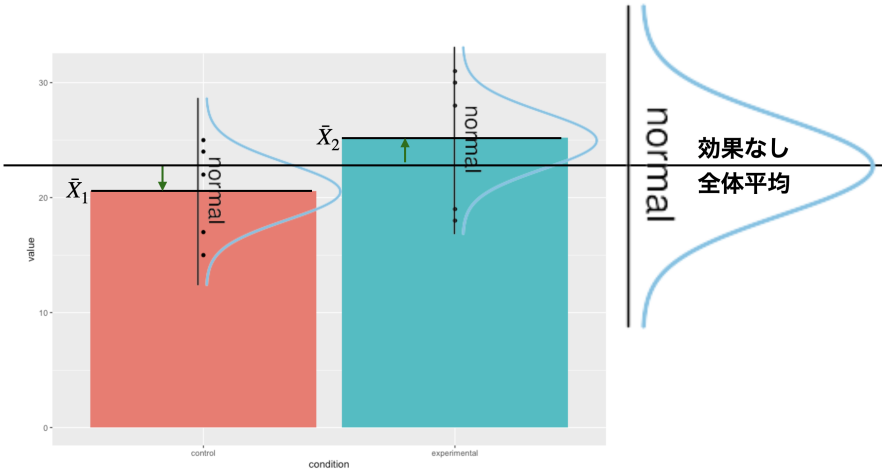


図 21.2 効果がなければ全体平均に同じ。群の効果は相対的に平均値をずらす。

[*1] gm は grand mean の略です。

と書くことができます。ここで一工夫。β_1 にプラス，マイナスの符号がついているのは，$+1, -1$ を掛け算していると考えてみましょう。この符号を決めているものを $X_1 = +1, X_2 = -1$ と表したとすると，まとめて次のように書くことができます。

$$X_{ij} = \beta_0 + \beta_1 X_j + \varepsilon_{ij} \tag{21.1}$$

これ，どこかで見覚えありませんか。そうです，前期の間にやった回帰分析と同じ形をしているのです (→ 第 9 講参照)。

さて回帰分析を今ここで大雑把に復習しておきますと，2 つの変数 x_i, y_i があったときに，両者の間に何らかの関数関係がある，と考えましょうという話でした。関数関係とは，一方を定めれば他方も定まるという関係ですから，この関係を数式で表現してやれば予測もできるぞ，嬉しいぞ。ということで，最も簡単な関数関係って何だ？ そう，中学校の時にやった一次関数 $y = ax + b$ だよね，ということでした。ただ中学校の時の話とは違って，データがたくさんあり，かつ，全部にぴったり当てはまる線は引けない。なので，ちょっと誤差はあるけどだいたいあってる線を引こうね，ということを考えたのでした。この関数 (一次関数，線形関数，線形モデル) で予測される値を $\hat{y}_i$ と表したとすると，

$$\hat{y}_i = b_0 + b_1 x_i$$

データ y_i は予測値 $\hat{y}_i$ に誤差 e_i がついて現れるから，

$$y_i = \hat{y}_i + e_i$$

なので，これを統合して

$$y_i = b_0 + b_1 x_i + e_i$$

としたのでした。

さてさて。記号で y_i, x_i とか b_0, b_1 だったのが β_0 とか β_1 などと変わってはいますが，基本的には $y = ax + b$ の一次線形関数の形です (図 21.3)。

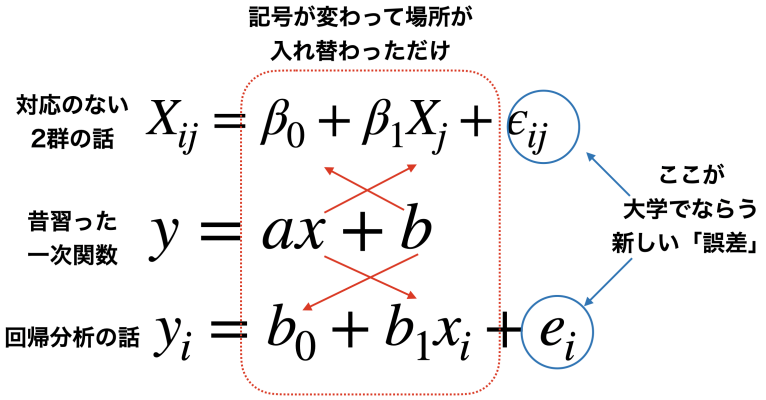


図 21.3 数学的には記号が変わっても使われ方が同じであれば同型と考えます

先ほど，効果がない＝全体平均，という言い方をしました。効果のことを，記号 β_1 で表しているのですから，効果がないというのは $\beta_1 = 0$ ということです。そうすると $\beta_1 X_j$ の項が消えてしまいます。つまり，全体平均 β_0 に誤差 ε_{ij} がついただけで，データは 1 つの正規分布から出てきたまま，ということになります。

1 つの共通の正規分布。2 つの群に分かれない，同じ正規分布。そろそろ帰無仮説検定と繋がってきたのではないでしょうか。効果がない，というのはまさに，帰無仮説のことですよね。つまり帰無仮説は，$\beta_1 = 0$ と

いう傾きゼロの直線，図 21.2 では群平均を横切っている，横一線の線形モデルだったのです。平均値の差の検定というのは，群平均が横一直線＝傾きゼロではないよ，という結論を出すためのロジックであり，比較される対立仮説は「傾きゼロでない何か」というモデルです。

- 帰無仮説 H0: $\mu_1 = \mu_2$ → 線形モデルにおける傾きの係数 $\beta_1 = 0$
- 対立仮説 H1: $\mu_1 \neq \mu_2$ → 線形モデルにおける傾きの係数 $\beta_1 \neq 0$

帰無仮説が傾きゼロという一点ばりに対し，対立仮説はゼロ以外であれば何でも良いというラフなモデルです。ですから帰無仮説が棄却されたとしても「差がなくはない」という程度のもので，「じゃあゼロでない，何なのか」ということまでは強く主張できません。帰無仮説検定は補助的に使いつつ，実質はどのように傾いているのか，差の信頼区間はどの程度あるのか，効果量 (標準化された平均値の差) は大きいのかといった，実質的な大きさを考えることが重要です。

また，前期の回帰分析の時は最小二乗法を使って回帰係数を求めたのでした (→ 第 9 講，セクション 9.3，Pp.145)。ただ，その時は推測統計学的な発想を導入していませんでした。データに最もうまく適合するために計算すると，とにかく求められるよ，という説明の仕方だったかと思います。手元のデータの話，つまり標本の話しかしていませんでしたので，係数として b_0, b_1 と書いていたのですが，推測統計学の枠組みでこれを考えると，我々が知りたいのは母集団における傾き，母数だということになりますので，今回ギリシア文字 β_1 とか ε_{ij} とします。

平均値差の検定と回帰分析は同じものなのだ，という話をしてきました。いずれも正規分布を仮定しており，その意味が「心理現象に仮定される全体傾向をあらわす分布」なのか「誤差の分布」なのか，という微妙な出自の違いを持っていると言えるかもしれません。しかし数学的には同じことなので，どちらの筋道からでも考えて，同じように利用できます。

また，回帰分析の係数を最小二乗法で求めたように，推測統計の枠組みで平均値を推定するのに最小二乗法を使うのだろうか，と考えている人もいるかもしれません。実はこれが大変興味深いところなのですが，推測統計の枠組みでは正規分布という確率分布を利用しますから，誤差を最小にしようという解析的な[*2]方法とは異なる考え方を使う必要があります。確率分布を使った母数の推定方法として，**最尤法 (Maximum Likelihood metod)** と**ベイズ法 (Bayesian metod)** という二種類の方法を今後説明していきますが，結論だけ先にいっておくと，幸いにして最小二乗法による答えと最尤法による答えは**ぴったり一致します**。

ちなみにこれまでやってきた，標本平均や標本分散から母数を推定する方法は，「**モーメント法 (Moment method)**」というやり方に分類されます。何であれ，同じデータであれば推定値はだいたい同じになります。同じものを推定しようとしてるのですから，違ってきたらむしろ困りますわね。

21.2 一要因 Between デザインの検定

さて，これまでは 2 水準の平均値の比較について話をしてきましたが，3 水準以上になるとどうなるのでしょうか？これも，基本的には線形モデルの延長で考えることができます。たとえば表 21.1 のようなデータがあったとしましょう。要因は 1 つですが水準は 3 つで，群間デザインの例です。それぞれの群平均は，$\bar{X}_1 = 5.5, \bar{X}_2 = 6.0, \bar{X}_3 = 4.25$ であり，全体平均は $\bar{X}_{gm} = 5.25$ になります。帰無仮説検定の文脈で考えれば，帰無仮説は「母平均に差がない」という仮説になりますから，記号で言えば $\mu_1 = \mu_2 = \mu_3$ になります。線形モデルで考えれば，図 21.4 にあるように，傾きのない線形モデル，つまり $X_{ij} = \bar{X}_{gm} + \varepsilon_{ij}$ の横一線ということになります。

[*2] 解析的なという言葉は，ここでは確率的でない，という意味で使っています。

表 21.1 一要因 3 水準 Between 計画のデータ

id	condition	notation	value
1	1	X_{11}	6
2	1	X_{21}	6
3	1	X_{31}	5
4	1	X_{41}	5
1	2	X_{12}	7
2	2	X_{22}	4
3	2	X_{32}	7
4	2	X_{42}	6
1	3	X_{13}	4
2	3	X_{23}	3
3	3	X_{33}	4
4	3	X_{43}	6

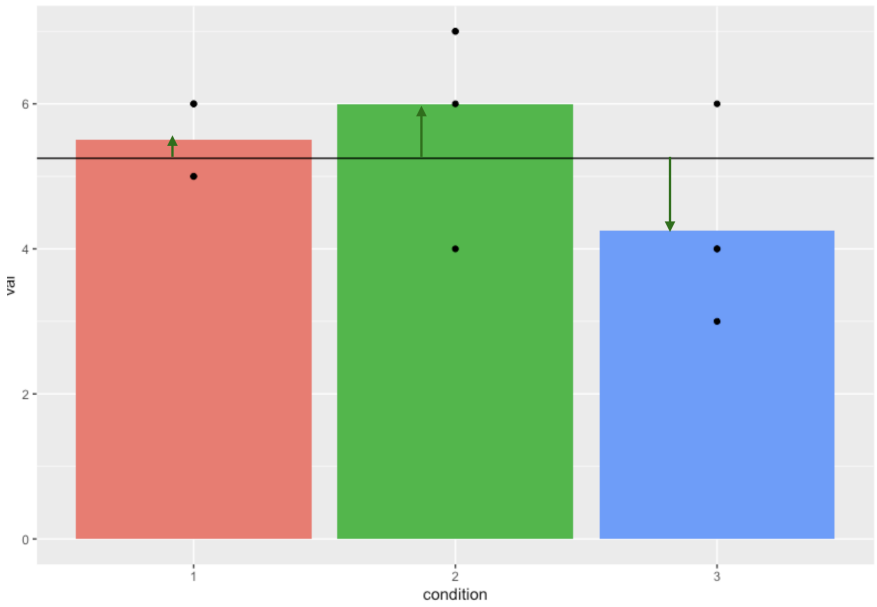


図 21.4 表 21.1 を可視化したもの。全体平均からの矢印は群効果の大きさ

対立仮説は帰無仮説の否定ですから，横一線の直線ではない (傾きがある)，あるいは母平均のどこかに差がある，というものになります。横一線ではない，という話ではありますが，右上がりの直線なのか，左下りの直線なのか，という話は一概にできないことに注意してください。というのも，横軸の水準 1/2/3 は名義尺度水準に過ぎないからです。水準 1/2/3 が「統制群」「操作 A をした群」「操作 B をした群」であったとして，統制群が 1 である理由はとくになく，2 でも 3 でも構いません。**名義尺度水準**は，対象と数字が一対一対応していることがその要件であって，数字の並びに意味がないからです。図 21.4 を見ると，水準 1 から 2 に向けては右上がり，水準 2 から 3 に向けては右下がりになって，一本の直線では表現できないですね。ただこれの並びを変えて，水準 3/1/2 の順番にすると右上がりの直線をあてがうことが可能です (図 21.5)。もちろん水準 2/1/3 の順にして右下がりの直線を当て嵌めても構いません。いずれもデータの持つ情報は同じです。です

から，対立仮説としては「横一線ではない」「母平均が同じではない」とするに留めましょう。

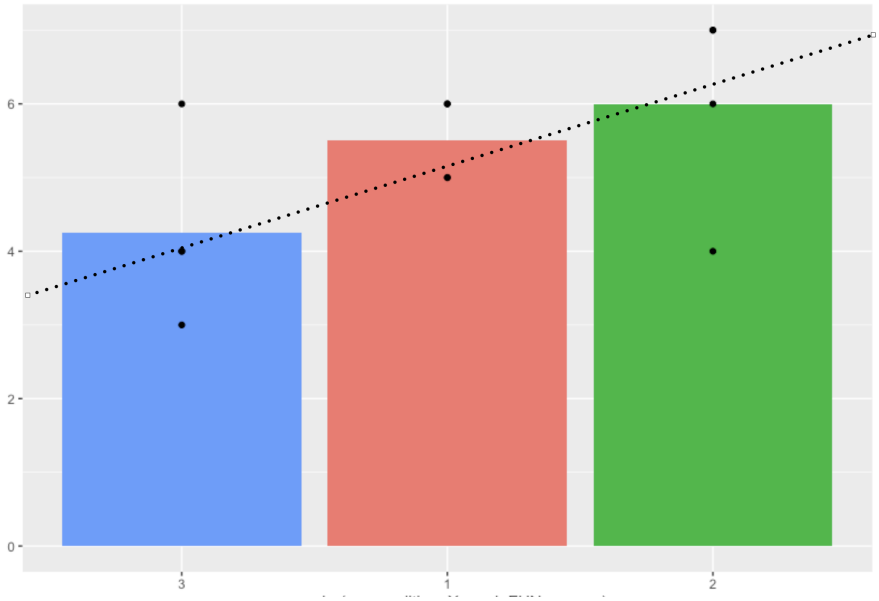


図 21.5 横軸は並べ替えても構わない

- 帰無仮説 H0: $\mu_1 = \mu_2 = \mu_3 \rightarrow$ 線形モデルにおける傾きの係数 $\beta_1 = 0$
- 対立仮説 H1: H0 ではない。$\rightarrow$ 線形モデルにおける傾きの係数 $\beta_1 \neq 0$

要因計画に帰無仮説検定の要素を加えた分析方法をとくに，**分散分析 (ANalysis Of VAriance;ANOVA)** と呼びます。ANOVA（アノーバ）と略して言われることもあります。「分散」分析といっていますが，やっていることは「平均値差の検定」であることに注意してください。どうして分散と呼んでいるのかは，直線が右上がりなのか，右下がりなのか一概に決められないことと関係があります。プラスにもマイナスにも変動するので，符号の効果をなくすべく二乗することで計算を進めるからです。

さて，前期の間は二乗して効果の大きさを算出するところまではやったのですが，効果と誤差をどうやって比較するか，ということについては伏せたままでした。というのも，推測統計学的な枠組みがなければ比較方法に言及できず，また帰無仮説検定の枠組みがなければ判断基準に言及できなかったからです。今ではこれらを使って説明できます。二乗してたし合わせたもの，すなわち**平方和**がどのような分布に従うのかを考え，理論分布から統計量を計算すれば良いわけです。その方法についてみていくことにしましょう。

21.2.1 帰無仮説検定の手続き

分散分析は，多水準の平均値差を検定する考え方です。帰無仮説検定の一種ですから，例の形式的手続きは踏襲していきます。

■仮説の設定 帰無仮説と対立仮説が必要です。既に述べたように，帰無仮説は母平均に差がないというものですから，$H_0 : \mu_1 = \mu_2 = \mu_3$ になり，対立仮説はこの否定になります。勘の良い人はここで「おや，まてよ？」と思うかもしれません。そうです，この帰無仮説，いろんな否定の仕方が考えられるのです。つまり，

- $\mu_1 \neq \mu_2$ だけど，μ_3 については $\mu_1 = \mu_3$ なのか $\mu_2 = \mu_3$ なのかわからない。
- $\mu_1 \neq \mu_3$ だけど，μ_2 については $\mu_1 = \mu_2$ なのか $\mu_3 = \mu_2$ なのかわからない。

- $\mu_2 \neq \mu_3$ だけど，μ_1 については $\mu_2 = \mu_1$ なのか $\mu_3 = \mu_1$ なのかわからない。
- $\mu_1 \neq \mu_2 \neq \mu_3$, つまりとにかく全部違う

の 4 パターン考えられます。実はこれ, すべて対立仮説の一員なのです。線形モデルとしては横一線でなければよい, ということなので, どこかに (統計的に有意なレベルでの) 傾きがあれば良いということになります。

■検定統計量を選択　平均値の推定ですが, 分散を検定対象にするので F 分布というものを使います。

■判断の基準になる確率を設定　有意水準 α, 今回も 5% で勝負を決めるとしましょう。

■検定統計量の実現値を計算　検定統計量 F の実現値の計算は, 先ほどの表 12.2 から行います。群の効果の列, 誤差の効果の列をすべて二乗して足し合わせることで, 平方和の計算ができます。

つまり, 次のように考えます[*3]。

$$\sum_i \sum_j (X_{ij} - \bar{X}_{gm})^2 = n \sum_j \beta_j + \sum_i \sum_j \varepsilon_{ij}$$

表 21.2　差分の計算

value	全体平均	群平均	平均偏差	群の効果	誤差	平均偏差の二乗	効果の二乗	誤差の二乗
6.000	5.250	5.500	0.750	0.250	0.500	0.562	0.062	0.250
6.000	5.250	5.500	0.750	0.250	0.500	0.562	0.062	0.250
5.000	5.250	5.500	-0.250	0.250	-0.500	0.062	0.062	0.250
5.000	5.250	5.500	-0.250	0.250	-0.500	0.062	0.062	0.250
7.000	5.250	6.000	1.750	0.750	1.000	3.062	0.562	1.000
4.000	5.250	6.000	-1.250	0.750	-2.000	1.562	0.562	4.000
7.000	5.250	6.000	1.750	0.750	1.000	3.062	0.562	1.000
6.000	5.250	6.000	0.750	0.750	0.000	0.562	0.562	0.000
4.000	5.250	4.250	-1.250	-1.000	-0.250	1.562	1.000	0.062
3.000	5.250	4.250	-2.250	-1.000	-1.250	5.062	1.000	1.562
4.000	5.250	4.250	-1.250	-1.000	-0.250	1.562	1.000	0.062
6.000	5.250	4.250	0.750	-1.000	1.750	0.562	1.000	3.062
合計			0.00	0.00	0.00	18.25	6.50	11.75

表 21.2 の最後の行を見てくださいね。平均偏差の二乗和 ($\sum_i \sum_j (X_{ij} - \bar{X}_{gm})^2$) が 18.25 で, これが効果の二乗 ($n \sum_j \beta_j$) の 6.50 と, 誤差の二乗 ($\sum_i \sum_j$) の 11.75 に分解できています。

この，11.75 と 6.50 を比べてどちらが大きいかという勝負になるのですが, この平方和を比べてみればもちろん誤差の方が大きいですよね。でも大事なのは, この分散が一単位あたりどれぐらい大きかったか, という点にあります。誤差の方，11.75 はデータ全体から生み出されていますが, 効果の方，6.50 は 3 つの水準から生み出されたわけです。この違いを加味して統計量を計算します。

検定統計量 F は, 分散の比の確率分布に従います。今はまだ**平方和**であって分散ではありません。二乗してたしただけで, 足した数で割らなければならないのです。ここで割る数が何かと考えると, 全体は (不偏分

[*3] 群の効果 β_j は同じ大きさのものが群の人数分, つまり $n = 4$ だけあるので，$n \sum_j \beta_j$ としています。

散で考えるので)$N-1$ ですが，要因や誤差は要因を生み出した要素の数 (マイナス 1)，誤差を生み出した要素の数 (マイナス 1) で考える必要があるのです。こうした一要素あたりの平均的な大きさのことを**平均平方 (Mean Squares)** と言います。

群の効果は，3 つの水準でこの平方和を作り出したわけですから，一単位は $3-1=2$ で考えて，MS は

$$6.50/(3-1)=3.250$$

となります。また，誤差の一単位あたりの大きさは，各群の平均と各値の差分で作られており，各群に 4 人配置されていますから，各群の誤差の自由度 $(4-1)$ が 3 群あると考えて

$$11.75/3(4-1)=11.75/9=1.306$$

です。

F 分布に従う検定統計量は，この MS の比率です。ここでは「誤差に比べてどれぐらい効果が大きかったか」がみたいのですから，効果の MS を誤差の MS で割って F 値を算出します。具体的には，

$$F=3.250/1.306=2.489$$

ということになります。

F 値は F 分布に従います。比の分布なので，分布の形状を決める自由度は分子と分母の 2 つが必要です。自由度は先ほどの一単位を計算した値であり，分子の自由度 $df_1=2$，分母の自由度 $df_2=9$ の F 分布からその出現確率を計算し，これが 0.1379077 とでます[*4]。5% 水準で有意，とは言えませんでしたね。

■分散分析表 なんだこの面倒な計算！と思った方も多いかと思います。少なくとも私は初めて習った時は，わけがわからなくなりました。まずは「分散 (平方和) の分割をする」ことと「分散の比で比較する」ということをしっかりと抑えてくれれば OK です。というのも，この手の面倒な計算はコンピュータがやってくれるからです。

その演習は次回に譲りますが，機械で計算した結果は次のような**分散分析表 (ANOVA Table)** にまとめて報告することが一般的です (表 21.3)。

表 21.3 ANOVA Table

	Df	SS	MS	F value	p
condition	2	6.500	3.250	2.489	0.138
Residuals	9	11.750	1.306		

この表にあるように，平方和 (SS)→ 自由度 (df) をつかって平均平方 (MS)→ 比率の F 比 →p 値の計算，という流れになっていることを確認してください。

結果をレポートなどで報告するときは，この表とともに，

> 一要因 3 水準の分散分析を行ったところ，$F(2,9)=2.489, p=0.138$ で，水準間に統計的に有意な差はみられなかった。

などとします。

[*4] R で F 分布の確率関数は `f` ですから，`1-pf(2.489,df1=2,df2=9)` として算出します。

■効果量は? 帰無仮説検定で，差があったかなかったか，という判断だけで終わらせるのではなく，どの程度の差があったのかについて報告しないといけないのでした。分散分析における効果量は，全体の分散のうち効果の分散がどれほどの量であったかという比率，つまり効果量=効果の分散/全体の分散=1 - (誤差の分散/全体の分散), で算出できます。これを ε^2(イプシロンの二乗) といいます[*5]。

とはいえ，母集団における効果の分散，母集団における誤差の分散は分かりませんので，標本の不偏分散をつかって標本効果量として推定するしかありません。今回，全体の標本 (不偏) 分散は，表 21.2 の「平均偏差の二乗」列の一番下にある 18.25，これを $12-1=11$ で割った $18.25/11=1.659091$ ですから，

$$\varepsilon^2 = 1 - \frac{1.306}{1.659} = 0.2127788$$

となりました。合わせて報告した方が良いでしょう。

21.2.2 事後の検定

さて，分散分析の結果として「有意である」という判断がなされた，つまり帰無仮説が棄却されて対立仮説が採択されたとき，複数考えられた対立仮説のどれが成立しているんでしょうか。実は，これはわかりません。帰無仮説じゃないことが明らかになっただけですから。

じゃあどうするか。簡単に思いつくのは，上のいくつかの対立仮説サブモデルを，虱潰しに検証していくことです。すなわち，1 群と 2 群，1 群と 3 群，2 群と 3 群というペアにして，対応のない t 検定を 3 回やれば良いのでは，と考える方もいるでしょう。水準数が 4 つになると 1-2,1-3,1-4,2-3,2-4,3-4 の 6 ペア，5 つになると 1-2,1-3,1-4,1-5,2-3,2-4,2-5,3-4,3-5,4-5 の 10 ペア･･･と水準数が増えれば手間は増えますが，なぁに機械が計算してくれるんだから努力で勝負！というところですね。

ですが，この考え方はこのままではうまく機能しません。既に第 19 講 (セクション 19.3.2，Pp.247) で学んだように，検定の繰り返しはタイプ 1 エラーの可能性を広げてしまいます。

ではどうすれば良いのか，ということになりますよね。この分散分析のあとの検定のことを**下位検定**とか**事後検定 (post-hoc analysis)** といったりしますが，この事後検定についてはさまざまな方法が提案されており，どれが良い･悪いと一概には言えないところがあります (それだけで専門書一冊出ているぐらいで (永田･吉田, 1997))。とは言え，それでは困るでしょうから，2 つほど方法を紹介しておきます。

■Bonferroni の方法 さてここで問題になっているのは，基本的に α レベルのインフレなのですから，これを低く調節してやれば良いのではないか，というのが 1 つの解決策です。たとえば今回は 3 つの対立仮説があって，全部で 5% 水準を保ちたいわけですから，$0.05/3=0.017$ として判断するようにしよう，というものです。これは直感的でわかりやすい基準ですね。この方法はさらにいろいろ改良版 (Holm 法，Shaffer 法，Holland-Copenhaver 法) があります。

■Tukey の HSD 法 これが最も一般的で，よく使われている方法です[*6]。ここでは各対について別の統計量 q を計算し，これが基準となるラインをクリアするかどうか，で判断します。

この他にも，統計パッケージによってさまざまな方法･オプションがあります。ですので，大事なのは「どの分析方法を使って，どのように判断したか」をしっかりと記録し，報告することだと思っておいてください。

[*5] 他にも同じような意味を持つ η^2 という指標もよく使われます。

[*6] ただし「よく使われている」から「正しい」とか「優れている」というわけではありません。それぞれの方法に，それぞれの仮定･前提，注目するところ･強調するところなどが変わって来ますし，「よく使われる」というのは実質的に「よく使う統計ソフトパッケージがこの統計量を計算してくれる」ぐらいの意味しかないこともあるからです。

21.3 二要因計画モデルの検定

では引き続き，二要因の実験デザインを検定することを考えましょう。

ここでは第 13 講 (セクション 13.1,Pp.189) でやった，二種類の銘柄のコーヒーをホットで飲むかアイスで飲むか，というデータを例にしたいと思います。要因は銘柄と温度，二要因ですから**交互作用**も計算に入れないといけない，という話でしたね。ここに表 13.3 を再掲しておきます。それぞれの効果や誤差の大きさ，平方和が算出されています。

表 21.4 再掲；平方和の計算

全体平均	温度の効果	銘柄の効果	交互作用	誤差	評価
9.75	-0.25	+0.75	+1.75	+1.00	13
9.75	-0.25	+0.75	+1.75	-1.00	11
9.75	-0.25	+0.75	+1.75	0.00	12
9.75	-0.25	-0.75	-1.75	0.00	7
9.75	-0.25	-0.75	-1.75	-1.00	6
9.75	-0.25	-0.75	-1.75	+1.00	8
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	+0.75	-1.75	0.00	9
9.75	+0.25	-0.75	+1.75	+2.00	13
9.75	+0.25	-0.75	+1.75	0.00	11
9.75	+0.25	-0.75	+1.75	-2.00	9
平方和	0.75	6.75	36.75	12.00	56.25

この平方和の比率が，F 分布に従うという性質を利用して，検定的判断を下します。形式的な手続きは次の通りです。

■仮説の設定 帰無仮説と対立仮説が必要です。既に述べたように，帰無仮説は母平均に差がないというものです。ここではどこをどう切り取っても母平均に差がない，ということになります。1 つ目の要因の第一水準 (温度要因のホット) で，2 つ目の要因の第二水準 (メーカー要因の銘柄 B) の母平均を μ_{12} とすると，帰無仮説は $H_0 : \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$ ということになります。対立仮説はこの否定です。

■検定統計量を選択 平均値の推定ですが，分散を検定対象にするので F 分布というものを使います。

■判断の基準になる確率を設定 有意水準 α，今回も 5% で勝負を決めるとしましょう。

■検定統計量の実現値を計算 検定統計量 F の実現値の計算は，先ほどの表 21.4 にある平方和を基に計算します。自由度を加味して 1 単位あたりの効果の大きさにしてから比べますので，分散分析表にまとめて表示しちゃいましょう。

ここでも，平方和 (SS)→ 自由度 (df) をつかって平均平方 (MS)→ 比率の F 比 →p 値の計算，という流れになっていることを確認しておきましょう。

さて，これをみると，温度の主効果はなく ($F(1,8) = 0.50, p = 0.500$)，メーカーの主効果もなく ($F(1,8) = 4.50, p = 0.067$)，交互作用がある ($F(1,8) = 24.50, p = 0.001$) ことがわかります。今回各 2

表 21.5 ANOVA Table

	Df	SS	MS	F value	p
温度	1.000	0.750	0.750	0.500	0.500
メーカー	1.000	6.750	6.750	4.500	0.067
交互作用	1.000	36.750	36.750	24.500	0.001
誤差	8.000	12.000	1.500		

水準ですので，主効果については事後検定をする必要はありません。交互作用がみられた場合は，どこに違いがあったのかさらに分析を進める必要があります。各水準ごとにデータを分割し (要因 A の第 1 水準，第 2 水準，要因 B の第 1 水準，第 2 水準)，その中でどこに差があるのかを探索します。今回の例で言えば，1. ホットの時に銘柄 A/B に差はあるか，2. アイスの時に銘柄 A/B に差があるか，3. 銘柄 A においてホット/アイスで差があるか，4. 銘柄 B においてホット/アイスで差があるか，の 4 パターンを検証することになります。もちろん検定の多重性の問題がありますから，Bonferroni や Tukey の方法などで補正する必要があります。

また，効果量も計算して報告するのが良いでしょう。これらの話は統計パッケージが自動的にしてくれることが多いので，R の演習の時まで，細かい話はお預けにしておきます。

21.4 課題

■Between 計画の ANOVA Table を書いてみよう　あるスナック菓子を三人のこどもそれぞれに買い与えたが，お菓子の長さに差があると言って喧嘩を始めてしまいました。彼らの主張が妥当かどうか検証するために，それぞれの菓子袋から棒状のお菓子 4 本ずつをサンプリングし，長さを測定したのが次の表 21.6 です。それぞれの効果などは計算してあります。これを使って分散分析表を書いてみましょう。

表 21.6 お菓子の長さデータ

id	袋	長さ	GM	mean	diff	eff	err	SS_{diff}	SS_{eff}	SS_{err}
1	A	6	10.333	8.25	-4.333	-2.083	-2.25	18.778	4.340	5.062
2	A	8	10.333	8.25	-2.333	-2.083	-0.25	5.444	4.340	0.062
3	A	10	10.333	8.25	-0.333	-2.083	1.75	0.111	4.340	3.062
4	A	9	10.333	8.25	-1.333	-2.083	0.75	1.778	4.340	0.562
5	B	17	10.333	13.75	6.667	3.417	3.25	44.444	11.674	10.562
6	B	11	10.333	13.75	0.667	3.417	-2.75	0.444	11.674	7.562
7	B	14	10.333	13.75	3.667	3.417	0.25	13.444	11.674	0.062
8	B	13	10.333	13.75	2.667	3.417	-0.75	7.111	11.674	0.562
9	C	7	10.333	9.00	-3.333	-1.333	-2.00	11.111	1.778	4.000
10	C	7	10.333	9.00	-3.333	-1.333	-2.00	11.111	1.778	4.000
11	C	12	10.333	9.00	1.667	-1.333	3.00	2.778	1.778	9.000
12	C	10	10.333	9.00	-0.333	-1.333	1.00	0.111	1.778	1.000
合計		124			0.000	0.000	0.00	116.667	71.167	45.500

表 21.7 の (あ)〜(お) に当てはまる数字を答えてください。

表 21.7　お菓子データの ANOVA Table

	SS	Df	MS	F value	p
お菓子の袋	(あ)	(う)	35.5833	(お)	0.0144
残差	(い)	9	(え)	5.0565	

■二要因計画の ANOVA Table を書いてみよう　この街には，ファーストフードのチェーン店が 2 つ展開しています (A/B)。いずれもハンバーガー (H) とチーズバーガー (C) がメニューにあります。各店舗でハンバーガーとチーズバーガーを 3 つずつ買ってきて，それぞれの重さを測った結果が表 21.8 です。ショップの違いによる効果やメニューの違い，交互作用の効果，誤差，それぞれの平方を取ったものなどは計算してあります。これを使って分散分析表を書いてみましょう。

表 21.8　ハンバーガーの重さ比較

id	Shop	Menu	Weight	GM	diff	E_{shop}	E_{menu}	$E_{int.}$	error	SS_{diff}	SS_{shop}	SS_{menu}	$SS_{int.}$	SS_e
1	A	H	55	50	5	2.333	2	1.667	-1.000	25	5.444	4	2.778	1.000
2	A	H	56	50	6	2.333	2	1.667	0.000	36	5.444	4	2.778	0.000
3	A	H	57	50	7	2.333	2	1.667	1.000	49	5.444	4	2.778	1.000
4	A	C	49	50	-1	2.333	-2	-1.667	0.333	1	5.444	4	2.778	0.111
5	A	C	49	50	-1	2.333	-2	-1.667	0.333	1	5.444	4	2.778	0.111
6	A	C	48	50	-2	2.333	-2	-1.667	-0.667	4	5.444	4	2.778	0.444
7	B	H	48	50	-2	-2.333	2	-1.667	0.000	4	5.444	4	2.778	0.000
8	B	H	49	50	-1	-2.333	2	-1.667	1.000	1	5.444	4	2.778	1.000
9	B	H	47	50	-3	-2.333	2	-1.667	-1.000	9	5.444	4	2.778	1.000
10	B	C	47	50	-3	-2.333	-2	1.667	-0.333	9	5.444	4	2.778	0.111
11	B	C	47	50	-3	-2.333	-2	1.667	-0.333	9	5.444	4	2.778	0.111
12	B	C	48	50	-2	-2.333	-2	1.667	0.667	4	5.444	4	2.778	0.444
合計						0.000	0	0.000	0.000	152	65.333	48.000	33.333	5.333

5778 表 21.9 の (か)〜(こ) に当てはまる数字を答えてください。

表 21.9 ANOVA Table

	SS	Df	MS	F value	p
Shop	(か)	1	65.333	(け)	0.001
Menu	(き)	1	48.000	(こ)	0.001
Interaction	(く)	1	33.333	50.000	0.001
Residual	5.333	8	0.6667		

第 22 章

R による分散分析

ここまでで，要因計画を線形モデルとして，あるいは帰無仮説検定の枠組みで考えるやり方を説明してきました。細かい計算は R で行う，ということで先送りしてきた話をここで回収したいと思います。R を使った演習でどのように計算が進むのか，どのように結果が表示されるのか，どのようにその結果を解釈するのかについて学んでいきましょう。

22.1 R で分散分析を行う上でのポイント

それでは R を起動するところから始めましょう。復習がてら，作業を始める前の最初のステップについて改めて書いておきます。

■RStudio の起動 RStudio を起動してください。パッケージのインストールをするときなど，管理人権限が必要になるケースがあります。

■プロジェクトを開く プロジェクトは前回作成した物で良いと思います。同じフォルダの中で，スクリプトファイルだけ変えれば良いでしょう。メニューからプロジェクトを開き (FIle→Open Project)，RStudio の右上などで，プロジェクトが開いていることを確認しましょう (→ 図 22.1)。

■R スクリプトを開く 今日のコードを書くための R スクリプトファイルを準備します。`File > NewFile > R Script` と進み，何も書いてない R スクリプトの画面を表示させてください。真っ白いファイルが開いたら問題ありません。

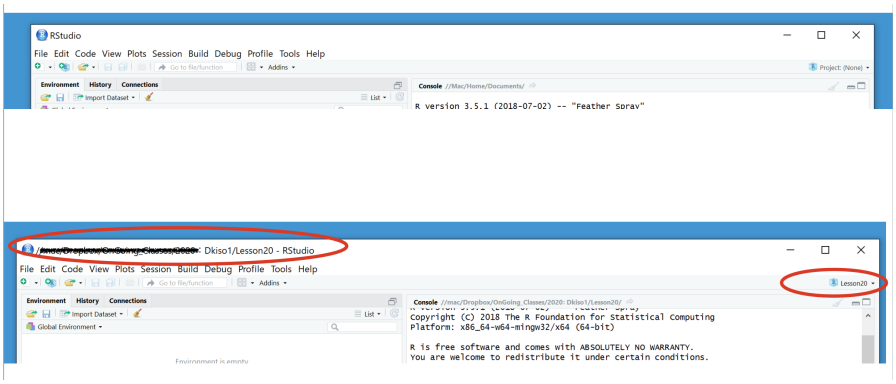


図 22.1 プロジェクトが開いていることを確認する。上が開いていない状態。下が開いている状態。開いていると，メニューバーにフォルダが表示され，右上にプロジェクト名が表示されている。

さて，準備 OK ですね。

22.1.1 数のタイプ

今回は要因計画を R で計算させるということですが，ここでポイントになるのが変数の型です。繰り返しお話ししていますように，前期でやった回帰分析と今回扱う要因計画とは，線形モデルという同じ形をしています。違うところがあるとすれば，回帰分析の説明変数が連続的であるのに対し，要因計画の説明変数は離散的 (カテゴリカル) であるということです。ですから，R で分析する時も関数の形は同じであっても，変数の種類が違うということを明らかにしておかなければなりません。

R にはいろいろな数字のタイプがあります。オブジェクトにいろいろな変数を格納できますが，その格納された変数のタイプには，次のような種類があります。

numeric　数字。加減乗除などの対象になるもの。尺度水準で言えば**間隔尺度水準**か**比率尺度水準**にあたるもの。

factor　ファクター型。訳して要因型ということもある。要因計画の「要因」です。この変数は数字とそれにつけるラベルとの組み合わせからなります。尺度水準で言えば**名義尺度水準**です。

orderd factor　順序付きファクター型。ファクター型の特殊系で，中についている数字に順序情報がある物。尺度水準で言えば当然**順序尺度水準**に該当します。あまり使われることはないけど。

chr　文字型 (Character)。一見ファクター型と見分けがつかないこともあるが，こちらは内部に対応する数字があるわけでもなく，ただの文字。シングルクォーテーション，ダブルクォーテーションで括ったものが変数に渡されるとこの型になる。

logical　論理型。*TRUE* か *FALSE* の 2 つの値しか持たない。*TRUE* は 1 に，*FALSE* は 0 に対応している。名義尺度水準の特殊な型で，オプションのスイッチオン･オフなどをこれで表したりする。大文字で書く TRUE や FALSE，あるいはそれらの一文字目 T や F はこの型のラベルとして予約されているので[*1]，オブジェクト名などにしない方が良い。

これらの型を確認するには，オブジェクトの仕組みがどうなっているかを調べる必要があります。調べるのは str 関数で，これを使うとオブジェクトの中身がわかるようになっています。

第 4 講のことを思い出しつつ，code:22.1 を入力してみてください。

code：22.1　オブジェクトに代入

```
1  obj <- data.frame(
2      list(name=c("kosugi","yamada","ueda"),
3           gender = c(1,2,1),
4           height = c(167,170,180),
5           weight = c(60,70,80))
6  )
7  str(obj)
```

■コード解説

1-6 行目　データフレームを作って obj オブジェクトに代入している。

7 行目　obj オブジェクトの構造を確認

[*1] 予約語といい，他にも何もないことを表す NULL や，無限大を表す Inf，欠損値を表す NA などがある。数学定数 pi も予約語の 1 つ。

さて，最後の 7 行目を実行すると，コンソールに次のように出力されていると思います (R の出力 22.1)。

R の出力 22.1: データフレームの中身

```
> str(obj)
'data.frame': 3 obs. of  4 variables:
 $ name  : chr  "kosugi" "yamada" "ueda"
 $ gender: num  1 2 1
 $ height: num  167 170 180
 $ weight: num  60 70 80
```

ここにあるように，obj オブジェクトの name 変数は chr 型で，他の 3 つの変数は数値型 (numeric) である，というのが示されています。ここで性別 gender を因子型にしてみましょう (code:22.2)。

code : 22.2 因子型にする

```
1  obj$gender <- factor(obj$gender,label=c('male','female'))
```

これの出力は R の出力 22.2 のようになります。

R の出力 22.2: 変数を因子型にした

```
> str(obj)
'data.frame': 3 obs. of  4 variables:
$ name  : chr  "kosugi" "yamada" "ueda"
$ gender: Factor w/ 2 levels "male","female": 1 2 1
$ height: num  167 170 180
$ weight: num  60 70 80
```

ここに示されているように，性別がファクター型になり，それは 2 つの水準を持っている。それぞれに「male」「female」というラベルがついている，ということがわかります。

この factor 型にしておくというのは，R が名義尺度水準の変数なんだなと理解したことになりますので，これで回帰分析をすると OK ということになります。逆に factor 型にしておかないと，量的変数として普通の回帰分析をしてしまうので，正しくない結果になります[*2]。

22.1.2 分散分析用の関数を読み込もう

factor 型にして分析するだけで，効果の大きさは見積もれるのですが，それを「判定」するには帰無仮説検定の枠組みで論じる必要があるでしょう。

また思い出して欲しいのですが，分散分析で群間・群内平方和の比較をした結果，有意であると判断されても「どこが有意なのか」までは分からないので，事後検定が必要なのでした。R はデフォルトで分散分析や Tukey の多重比較をやってくれる関数などを持っていますが，別々の関数になっていて，自分で 1 つ 1 つやっていくのは面倒なところもあります。この主な分析 → 事後の検定という一連の流れを，自動的にやってくれる便利な関数があります。

井関龍太さんの anovakun というのがそれで，ネットで一般に公開されています[*3]。CRAN からダ

[*2] 具体的には，3 水準に 1,2,3 という量的な数字を割り振って分析するということは，$1 < 2 < 3$ という順序性と 2-1=3-2 という等間隔性を持たせたことになります。要因計画の可視化の時に，横軸は入れ替え可能 (順序性がない) だったことを思い出してください。

[*3] https://riseki.cloudfree.jp/?ANOVA%E5%90%9B

ウンロードするパッケージのようにはなっていませんので，これを使うにはローカルに anovakun 関数を書いたファイル (執筆時点，2024 年 11 月現在で anovakun_489.txt) をダウンロードし読み込むか，R から直接インターネット経由でダウンロードして使うか，の 2 通りの方法のいずれかを使う必要があります。どちらにしても，R の source 関数をつかうもので，ダウンロードしたファイルを使う場合はプロジェクトフォルダに当該ファイルをダウンロードした上で，読み込むコードを実行します (code:22.3)。

code：22.3 ソースを読み込む

```
1 source("anovakun_489.txt")
```

ネットから直接ダウンロードする場合は，インターネットにつながっている状態で code:22.4 を実行します[*4]。

code：22.4 ANOVA 君のソースをネットからとってくる

```
1 source("https://shorturl.at/tsBhl")
```

これらのコードが実行されると，R の環境にたくさんの関数が入っているはずです (図 22.2)。これらが入っていれば OK です。それでは実際の使い方を見ていきましょう。

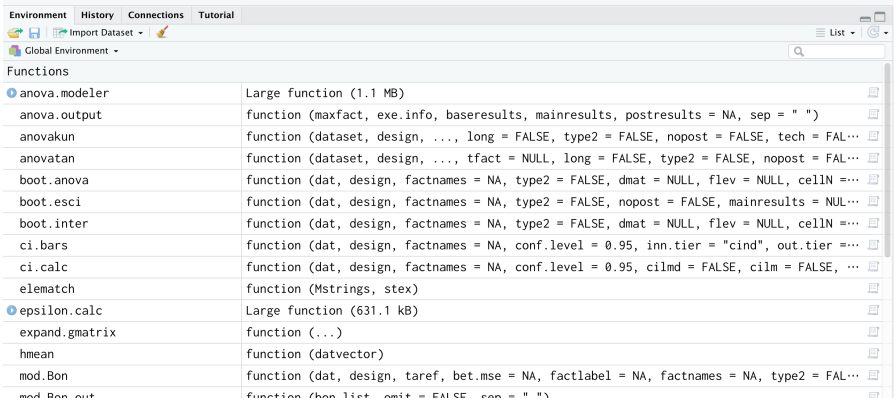


図 22.2 anovaku を読み込んだあとの環境タブ

22.2 一要因 Between 計画の実際

ではまず一要因 Between 計画の話から。例として第 21 講のデータを使います。一応再掲しておきましょう (表 22.1)。まずはこれを線形モデルとして分析するところからです。code:22.5 のようにデータフレームを組み上げます。

code：22.5 データの準備

```
1 # データフレームを組み上げる
2 Example1 <- data.frame(id = rep(1:4,3),
3                        condition = rep(1:3,each=4),
4                        value = c(6,6,5,5,7,4,7,6,4,3,4,6))
5 # 要因型にする
```

[*4] リンク先を転記するのは大変です。Google で「ANOVA 君」を検索し，井関先生のサイトに飛んでから，「下のアイコンをクリックしてファイルを保存してください。」とある ANOVA 君ソースファイルのリンク先を右クリック，リンク先をコピーすれば確実です。ちなみにこのアドレスはリンクの長さを縮める Link Shortner サービスを使っています。

表 22.1 一要因 3 水準 Between 計画のデータ

id	condition	notation	value
1	1	X_{11}	6
2	1	X_{12}	6
3	1	X_{13}	5
4	1	X_{14}	5
1	2	X_{21}	7
2	2	X_{22}	4
3	2	X_{23}	7
4	2	X_{24}	6
1	3	X_{31}	4
2	3	X_{32}	3
3	3	X_{33}	4
4	3	X_{34}	6

```
6  Example1$condition <- factor(Example1$condition,
7                               labels = c("control","exp1","exp2"))
```

■コード解説

1-3 行目 データフレームを作っています。ID は被験者番号ですが, 群内での番号なので 1 から 4 の繰り返しを 3 回 (`rep(1:4,3)`) です。条件 condition は 3 種類, 各 4 人分 (`rep(1:3,each=4)`), 最後の従属変数 (value) は手入力するしかありません。

5-7 行目 条件変数を要因型にしています。ラベルは統制群 (control) と実験群 1(exp1), 実験群 2(exp2) としています。

22.2.1 線形モデルとして分析

線形モデルとしての分析は, 前期でやったように `lm` 関数を使います (code:22.6)。

code : 22.6 lm 関数で分析

```
1  result.lm1 <- lm(value~condition,data=Example1)
2  summary(result.lm1)
```

1 行目が `lm` 関数を使っていて, 結果を `result.lm1` オブジェクトに入れているところです。従属変数と独立変数をチルダ~で繋げているところ, 思い出してください。2 行目はこの結果の出力になります。結果は **R の出力** 22.3 のように示されているはずです。

R の出力 22.3: lm 関数の出力

```
Call:
lm(formula = value ~ condition, data = Example1)

Residuals:
   Min     1Q Median     3Q    Max
-2.000 -0.500 -0.125  0.625  1.750

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.5000     0.5713   9.627 4.91e-06 ***
conditionexp1   0.5000     0.8079   0.619    0.551
conditionexp2  -1.2500     0.8079  -1.547    0.156
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.143 on 9 degrees of freedom
Multiple R-squared:  0.3562, Adjusted R-squared:  0.2131
F-statistic: 2.489 on 2 and 9 DF,  p-value: 0.1379
```

これを見ると，切片（Intercept）の推定値が 5.5，次に 2 つの説明変数の係数が書いてあります。文字が連なっているのでわかりにくいかもしれませんが，それぞれ condition と exp1 の関係，condition と exp2 の関係を表しています。図 22.3 を見ると結果の意味は分かり易いと思います。

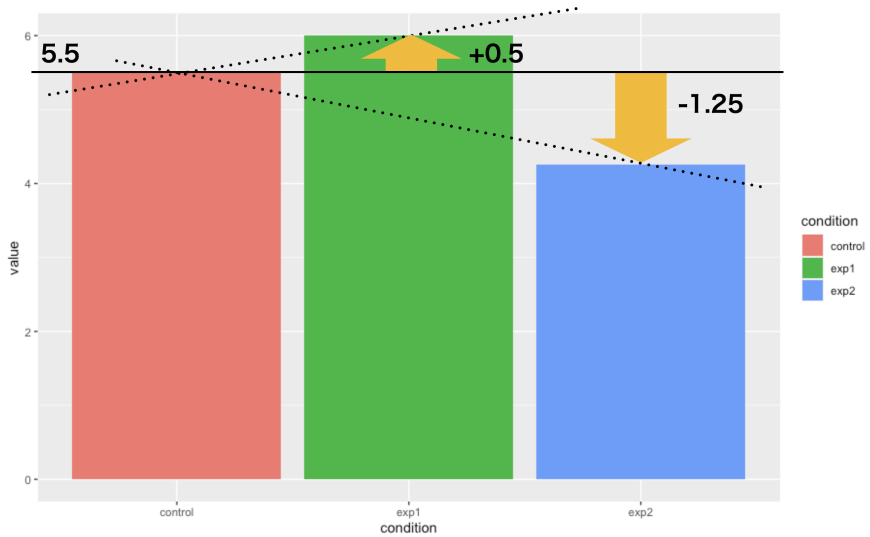


図 22.3 線形モデルとして分析した結果の意味するもの

第 21 講では，線形モデルを

$$X_{ij} = \beta_0 + \beta_j + \varepsilon_{ij}$$

としていました。ここで β_0 は全体平均でした。全体平均を基本として，各水準に対する効果 β_j がどうなったか，というモデル化でしたが，ここでは

$$X_{ij} = \bar{X}_1 + \beta_j + \varepsilon_{ij}$$

という形になっています (図 22.3)。本質的に違うものではありません。統制群に比べて実験群 1 の平均値は +0.5, 実験群 2 は −1.25 の平均値の違いがあるのだな, と思っていただければ結構です。その後ろに Std.Error や t 値が出ていますが, これらは推定された回帰係数の標準誤差および, 係数が 0 であるという帰無仮説を棄却できるかどうかの検定結果が出ています[*5]。

線形モデルですから, モデルの適合度 (`Multiple R-squared, Adjusted R-squared`) なども出ていますが, それよりも最後の一行, この線形モデル全体の検定結果が出ています。これは要因計画のもう 1 つの切り口, 分散分析による結果と同じものです。では分散分析の視点からこの分析結果を見てみましょう。

22.2.2 帰無仮説検定として分析

要因計画の帰無仮説検定版, つまり分散分析の結果ですが, 線形モデルと同じなので今回得られた出力をすこし変えるだけで対応できます。最も単純には, 先ほどの線形モデルの結果を `anova` 関数に入れると, 分散分析表が次のように示されます (code:22.7,R の出力 22.4)。

code : 22.7 分散分析としてみる

```
1  anova(result.lm1)
```

R の出力 22.4: 分散分表の出力

```
Analysis of Variance Table

Response: value
          Df Sum Sq Mean Sq F value Pr(>F)
condition  2   6.50  3.2500  2.4894 0.1379
Residuals  9  11.75  1.3056
```

ここから, $F(2, 9) = 2.4894, p = 0.1379$ で統計的に有意な差はみられなかった, と判断できます。これらの数字, どこをどう読み取ったのか確認してください。

これだけですから, `lm` 関数だけでいいじゃないかと考える人もいるかもしれません。基本的には問題ないのですが, 分散分析は「群のサンプルサイズが違う場合」や, 「分散の計算」においていくつかの計算バリエーションがあり, R デフォルトの `lm` 関数では十分対応できなくなるシーンが出てきます。

そこで先ほど導入した, 分散分析専用関数 `anovakun` を使います。実際の使い方は次のようになります (code:22.8)。

code : 22.8 anovakun で実行

```
1  anovakun(Example1[,2:3],"As",3,eps=T)
```

まず, `anovakun` には分散分析に必要なデータだけを与える必要があります。今回のデータフレームにある id 列は入りませんので, `Example1[,2:3]` としてデータの 2 列目と 3 列目だけ (2 列目から 3 列目まで) に限定しています[*6]。次に書いてある `As` は, `anovakun` 独特の記法です。ここでの `s` は被験者 subject を意味しており, これの左側に群間要因を, 右側に**群内要因**を書きます。今回は一要因の群間デザインですので, 左側に 1 つ要因 A をおく, という書き方になっています。これを `sA` のようにすると一要因群内デザイ

[*5] 各群の母平均の推定値でもありますが, 各群について一変数の t 検定をしているのではなく, これら線形モデル全体として (他の説明変数に条件づけられた分布の中で) 標準誤差を計算します。詳しくは南風原 (2002) の p.118 を参照してください。

[*6] 大括弧 [,] は, データフレームの行および列を指定するのに使います。一行目だけを指定したい場合は `Example1[1,]`, 一列目だけ指定したい場合は `Example1[,1]`, 1 行 1 列目の要素だけアクセスしたい場合は `Example1[1,1]` のようにします。

ン，AsB とすると，要因 A は群間で要因 B は群内の混合計画を表すことになります。最後に，今回の一要因 A の水準数を入力すれば，anovakun 関数はすべての分析を実行してくれます。オプションとして，効果量 ε^2 を表示するスイッチをオン (eps=T) とすると，効果量も出してくれます (R の出力 22.5)。

R の出力 22.5: anovakun の出力

```
[ As-Type Design ]

This output was generated by anovakun 4.8.5 under R version 4.0.2.
It was executed on Tue Sep 15 13:40:30 2020.


<< DESCRIPTIVE STATISTICS >>

--------------------------
 A   n    Mean    S.D.
--------------------------
a1   4  5.5000  0.5774
a2   4  6.0000  1.4142
a3   4  4.2500  1.2583
--------------------------


<< ANOVA TABLE >>

-------------------------------------------------------------------
Source      SS  df     MS  F-ratio  p-value      epsilon^2
 ----------------------------------------------------------------
      A  6.5000   2 3.2500   2.4894   0.1379 ns      0.2131
  Error 11.7500   9 1.3056
-------------------------------------------------------------------
  Total 18.2500  11 1.6591
                    +p < .10, *p < .05, **p < .01, ***p < .001

output is over --------------------///
```

各群の平均値，標準偏差および，分散分析表が表示されていますね。これを見て，どことどこの数字を参照すれば良いか等，確認しておいてください。

これだけだと，とくに anovakun が便利そうに思えないかもしれません。が，以下さまざまな要因計画に対応しているので，徐々にその良さがわかってくると思います。

22.3 二要因 Between 計画の実際

それでは続いて，二要因の計画にいきましょう。データを再掲しておきましょう (表 22.2)。これをデータフレーム型にします。温度，メーカーは名義尺度水準なので要因型にします (code:22.9)。

code : 22.9 データフレームにする

```
1  Example2 <- data.frame(id = 1:12,
2                         num = rep(1:3,4),
```

表 22.2 二要因のデータ例

通し番号	群での番号	温度	メーカー	評価
1	1	Hot	銘柄 A	13
2	2	Hot	銘柄 A	11
3	3	Hot	銘柄 A	12
4	1	Hot	銘柄 B	7
5	2	Hot	銘柄 B	6
6	3	Hot	銘柄 B	8
7	1	Cold	銘柄 A	9
8	2	Cold	銘柄 A	9
9	3	Cold	銘柄 A	9
10	1	Cold	銘柄 B	13
11	2	Cold	銘柄 B	11
12	3	Cold	銘柄 B	9

```
3                          temp = rep(1:2,each=6),
4                          maker = rep(rep(1:2,each=3),2),
5                          value = c(13,11,12,7,6,8,9,9,9,13,11,9))
6  Example2$temp <- factor(Example2$temp,
7                           labels=c("Hot","Cold"))
8  Example2$maker <- factor(Example2$maker,
9                            labels=c("A","B"))
```

22.3.1 線形モデルとして分析

これでデータはできたので，分析にかけるだけですね。二要因以上に計画が複雑になる場合は，交互作用を考えなければならないことに注意が必要です。線形モデルで表現するには，次のようにします (code:22.10)。

code：22.10 線形モデルとして実行

```
1  result.lm2 <- lm(value ~ temp*maker,data=Example2)
2  summary(result.lm2)
```

ここで，説明変数が * でつなげられています。これで主効果と交互作用の両方をモデルの中に入れたことになります。結果は R の出力 22.6 のように表示されるはずです。

R の出力 22.6: 交互作用を入れたモデルの出力

```
 > summary(result.lm2)

Call:
lm(formula = value ~ temp * maker, data = Example2)

Residuals:
    Min      1Q  Median      3Q     Max
-2.00  -0.25   0.00   0.25   2.00

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      12.0000     0.7071   16.97 1.48e-07 ***
tempCold         -3.0000     1.0000   -3.00  0.01707 *
makerB           -5.0000     1.0000   -5.00  0.00105 **
tempCold:makerB   7.0000     1.4142    4.95  0.00112 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.225 on 8 degrees of freedom
Multiple R-squared:  0.7867, Adjusted R-squared:  0.7067
F-statistic: 9.833 on 3 and 8 DF,  p-value: 0.004641
```

ここに示されているのは，温度 temp 変数の第一水準 (Hot) かつメーカー maker 変数の第一水準 (銘柄 A) を基準にした時に，tempCold つまりメーカーは同じ (銘柄 A) で温度だけ違う群は平均値が -3.0，次にメーカーを B で makerB 温度は Hot の群と比べるとその平均値は -5.0，両方変えると $+7.0$ になる，ということです。それぞれ標準誤差や t 値が出ていますが，これは傾きが 0 である，という帰無仮説を棄却しているだけであって，効果の大きさについてはわかりません。

また，最後に F 統計量が出ています。自由度が $(3, 8)$ の時に $p = 0.004641$ ですが，これは線形モデル全体としてみた時に横一線の帰無仮説モデルとは違うね，というだけであって，具体的にどこがどう違うのか，というのがわかりにくくなっています。

そこでそれぞれの水準について検証できる，分散分析モデルで分析してみます。

22.3.2 帰無仮説検定として分析

要因計画の帰無仮説検定版，でやってみましょう。先ほどと同じように，R がデフォルトで持っている関数 anova に線形モデルの結果を入れてみます (code:22.11)。

code：22.11 分散分析としてみる

```
1  anova(result.lm2)
```

結果は R の出力 22.7 のようになります。

R の出力 22.7: 分散分析表の出力

```
> anova(result.lm2)
Analysis of Variance Table

Response: value
           Df Sum Sq Mean Sq F value   Pr(>F)
temp        1   0.75    0.75     0.5 0.499576
maker       1   6.75    6.75     4.5 0.066688 .
temp:maker  1  36.75   36.75    24.5 0.001121 **
Residuals   8  12.00    1.50
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

分散分析表だけでなく，ご丁寧に有意かどうかわかりやすいような記号 (. や*など) をつけてくれています。最後の一行には Signif. codes とあって，米印が 1 つなら 5%，2 つなら 1%，3 つなら 0.1% 水準で有意である，という読み方が書いてあります[*7]。これを見ると，maker 水準は 5% 水準で有意にはならず[*8]，交互作用項 (temp:maker) は有意になっていることがわかります。

ただ，このままだと「交互作用は有意だけど，どこがどう違うのかわからない」という状況ですね。4 つの母平均が同じである ($\mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$) という帰無仮説が棄却されただけで，どこがどう違うのかについては事後の検定を追加しなければならないからです。

ではこれを anovakun でやってみましょう。彼 (?) なら一気にすべての答えを出してくれるのです (code:22.12)。

code : 22.12 anovakun で実行

```
anovakun(Example2[,3:5],"ABs",2,2,eps = T)
```

先ほどと同様，anovakun には id 列や num 列は必要ありませんので，3 行目から 5 行目まで，要因 A,B と従属変数だけを与えます。要因計画は間 × 間ですので，被験者 s の左側に A,B としています。2 つの要因はそれぞれ 2 水準ですので，水準数を 2 つ書いています。最後の eps=T は，効果量である ε^2 を表示するスイッチをオン (TRUE)，ということです。

結果は R の出力 22.8 のように表示されます。

[*7] おそらくこれが諸悪の根源で，なんとなく星の数が多い方が「良い結果」のような印象を持ちますし，分析したらとりあえず星を探す，という仕草が生まれるようになっていますよね。帰無仮説検定はあくまでもモデル比較の判定であって，判定基準が厳しいほど「良い」とか，基準をクリアできなかったから「惜しい (残念)」という評価を伴うものではないはずなのに，です。有意差がなかったら，そういう結果が得られた，というだけですし，星の数が小さくてもえらいとかすごい，ということにはならない，ということを再確認してください。「心理学者は星を探して喜んでるんだ」というようなジョークの的にならないように，適切な分析的素養を身に付けましょう。

[*8] 世の中には「有意傾向」という変な言葉があります。今回の p 値は 6% でしたが，これは 5% 水準ではアウトですよね。こういった 5% をちょっとだけ上回っちゃった，惜しい！という時に，「有意になりそうな傾向があった」略して「有意傾向があった」という表現がされることがあるのです。あるいは「傾向差があった」という言葉もまことしやかに使われていたりします。が，これらは心理学独自の間違った表現であり，決して使うべきではありません。なぜなら，5% だろうが 6% だろうが，任意に決めた判定基準であって，p 値の大小関係は現実の効果の大きさとは関係がないからです。

R の出力 22.8: ABs タイプの出力

```
[ ABs-Type Design ]

This output was generated by anovakun 4.8.5 under R version 4.0.2.
It was executed on Tue Sep 15 11:50:09 2020.

<< DESCRIPTIVE STATISTICS >>

------------------------------
  A   B   n     Mean    S.D.
------------------------------
 a1  b1   3  12.0000  1.0000
 a1  b2   3   7.0000  1.0000
 a2  b1   3   9.0000  0.0000
 a2  b2   3  11.0000  2.0000
------------------------------

<< ANOVA TABLE >>

-------------------------------------------------------------
 Source      SS  df      MS  F-ratio  p-value      epsilon^2
-------------------------------------------------------------
      A  0.7500   1  0.7500   0.5000   0.4996 ns     -0.0133
      B  6.7500   1  6.7500   4.5000   0.0667 +       0.0933
  A x B 36.7500   1 36.7500  24.5000   0.0011 **      0.6267
  Error 12.0000   8  1.5000
-------------------------------------------------------------
  Total 56.2500  11  5.1136
                  +p < .10, *p < .05, **p < .01, ***p < .001

<< POST ANALYSES >>

< SIMPLE EFFECTS for "A x B" INTERACTION >

-------------------------------------------------------------
Source      SS  df      MS  F-ratio  p-value      epsilon^2
-------------------------------------------------------------
A at b1 13.5000   1 13.5000   9.0000   0.0171 *       0.2133
A at b2 24.0000   1 24.0000  16.0000   0.0039 **      0.4000
B at a1 37.5000   1 37.5000  25.0000   0.0011 **      0.6400
B at a2  6.0000   1  6.0000   4.0000   0.0805 +       0.0800
  Error 12.0000   8  1.5000
-------------------------------------------------------------
                    +p < .10, *p < .05, **p < .01, ***p < .001

output is over --------------------///
```

最初に，水準ごとの平均値と標準偏差が表示され，続いて分散分析表が示されています。これを見ると，要

因 A(温度) の主効果はなく ($F(1,8) = 0.500, p = 0.4996$)[*9], ついで要因 B の効果も 5% 水準では認められない (10% 水準で有意, その記号として + が使われています), 交互作用は 5% 水準で有意という結果になりました。ここまでは当然, 先ほどと同じですね。効果量 ε^2 の列を見ると, 交互作用は全体の 62% の分散を説明したことになりますから, これはかなり大きな効果だったようだ, ということがわかります。

続いて POST ANALYSIS として事後の検定が表示されています。SIMPLE EFFECTS for "A x B" INTERACTION とあるのは,「A と B の交互作用についての単純効果」という意味で (直訳！), 各水準に限定した時に他の水準の効果があったかどうかをみています。たとえば最初の行, A at b1 とありますが, これは要因 B(銘柄) の第一水準における要因 A(温度) の効果をみていることになります。以下同様ですが, これらは分散分析の表をより細かくしてチェックしているようなものです。これを見ると, 冷たい時の銘柄には差がない (B at a2) ようですが, それ以外の差は統計的に有意, と言えそうです。

これらの結果を見やすくするために, 図 22.4 に対応するところに線を引いてみました。結果と見比べて, どこに差があったか確認しておいてください。

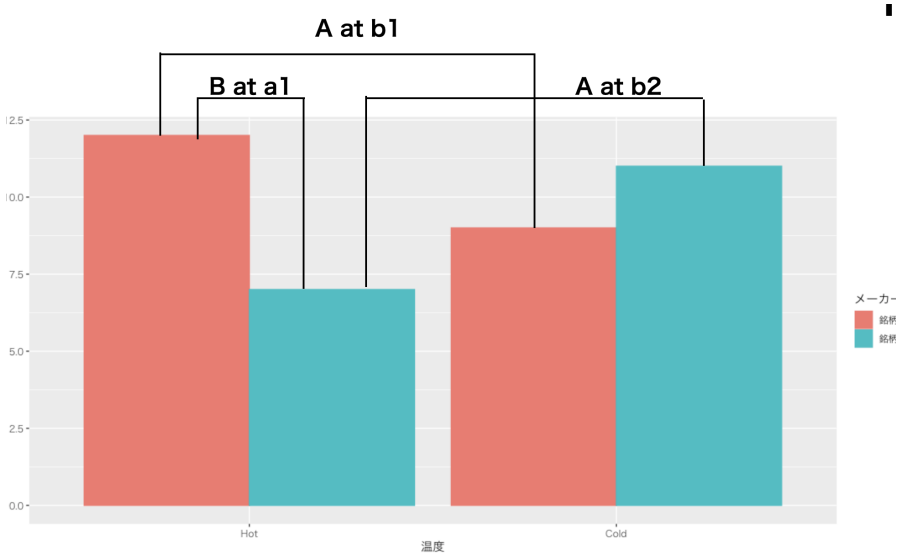


図 22.4 POST ANALYSIS の検定結果の確認

また, これらの結果を報告する時は, 次のように書くと良いでしょう。

> 温度と銘柄の効果を独立変数とした二要因分散分析 (2×2 の Between デザイン) をおこなったところ, 交互作用が有意 ($F(1,8) = 24.5, p = 0.0011, \varepsilon^2 = 0.6267$) になり, 事後の検定の結果, 銘柄 A における温度の効果 ($F(1,8) = 9.0, p = 0.0171, \varepsilon^2 = 0.2133$), 銘柄 B における温度の効果 ($F(1,8) = 16.0, p = 0.0039, \varepsilon^2 = 0.4000$), および温かい時の銘柄の効果 ($F(1,8) = 25.0, p = 0.0011, \varepsilon^2 = 0.6400$) が統計的に有意であった。

22.4 一要因 Within 計画の実際

最後に Within 計画の例をやってみましょう。仮想データとして次のような表 22.3 を使います。これをデータフレーム型にします (`code:22.13`)。

[*9] 米印があるべきところに ns とありますが, これは not significant, つまり有意ではないという意味です。

表 22.3 Within Data Example

ID	Time1	Time2	Time3
1	10	5	9
2	9	4	5
3	4	2	3
4	7	3	5

code : 22.13 データフレームにする

```
1  Example3 <- data.frame(ID=1:4,
2  Time1 = c(10,9,4,7),
3  Time2 = c(5,4,2,3),
4  Time3 = c(9,5,3,5))
```

線形モデルとして表現するには, 少し複雑な操作が必要ですので[*10], 今回は分散分析的アプローチのみ紹介したいと思います。

22.4.1 帰無仮説検定として分析

群内モデルの anovakun での入力は次のようになります (code:22.14)。

code : 22.14 anovakun で実行

```
1  anovakun(Example3[, 2:4], "sA", 3, eps = T)
```

データセットや変数の指定, 効果量オプションの指定はとくに追記することはありません。デザインの指定が"sA"と s の右側に来ているところがポイントです。これを実行すると, R **の出力** 22.9 のように結果が示されます。

まず群ごとの記述統計量が計算されています。その次に<< SPHERICITY INDICES >>とありますね。これは**球面性の検定 (sphericity test)** と訳されます。t 検定の時に Welch の補正をしたのを思い出して欲しいのですが, 分散分析の時も基本的に母分散が同じである, という仮定があります。この仮定が満たされない場合は補正をしなければなりません。とくに Within 計画の場合は, 全部の分散 (共分散) が等しいという仮定を置くのではなく, 少なくとも全方向に等しく広がっている空間であるはずだ, という仮定をおいています[*11]。

[*10] 具体的には, 一人ひとりに回帰分析をしつつ, 個々人の切片＝個人差がより上位の確率分布にしたがっている, という階層的線形モデルという形になり, 次回以降で取り上げる確率モデルの導入をしないといけないのです。R で実行する場合も別途パッケージを導入する必要があり, とても今回の残り時間で扱い切れる分量ではないので, 割愛させていただきました。興味がある人は二年次のデータ解析応用でお会いしましょう。

[*11] Within 計画の時にややこしいのは, モデルでいうと個々人に線形モデルを仮定して, 個人差＝切片が異なるような階層的なモデルになっているということです。個々人の分散が同じ, というのはなかなか厳しい仮定だと思われますが, 水準同士の相関関係 (共分散) も考慮に入れて, 分散共分散行列 (対角要素に分散, 非対角要素に共分散が入っている正方行列) が均質である, という仮定ぐらいはせめて保持したいわけです。水準数が増えるとこの相関関係を表現するする世界は多次元空間に広がっていくので, 究極的には丸くなる, 球面性の検定といわれる所以です。

R の出力 22.9: sA タイプの anovakun 出力

```
[ sA-Type Design ]

This output was generated by anovakun 4.8.5 under R version 4.0.2.
It was executed on Tue Sep 15 13:50:06 2020.


<< DESCRIPTIVE STATISTICS >>

--------------------------
  A   n    Mean    S.D.
--------------------------
 a1   4  7.5000  2.6458
 a2   4  3.5000  1.2910
 a3   4  5.5000  2.5166
--------------------------


<< SPHERICITY INDICES >>

== Mendoza's Multisample Sphericity Test and Epsilons ==

------------------------------------------------------------------------
 Effect  Lambda  approx.Chi  df      p         LB     GG     HF     CM
------------------------------------------------------------------------
      A  1.0000     -0.0000   2 1.0000 ns  0.5000 1.0000 3.0000 0.5000
------------------------------------------------------------------------
                             LB = lower.bound, GG = Greenhouse-Geisser
                            HF = Huynh-Feldt-Lecoutre, CM = Chi-Muller


<< ANOVA TABLE >>

-------------------------------------------------------------
 Source      SS  df      MS  F-ratio  p-value      epsilon^2
-------------------------------------------------------------
      s 39.0000   3 13.0000
-------------------------------------------------------------
      A 32.0000   2 16.0000  16.0000   0.0039 **      0.3896
  s x A  6.0000   6  1.0000
-------------------------------------------------------------
  Total 77.0000  11  7.0000
                  +p < .10, *p < .05, **p < .01, ***p < .001
```

これが満たされない場合は，Greenhouse-Geisser の補正や Huynh-Feldt-Lecoutre などの補正をしなければならないが，どうでしょうね？ というのを確認するのがこの箇所です (R **の出力** 22.10)。今回は p が 1.000 で ns，つまり not significant=有意に歪んでいるとは言えない，ということなので補正は不要です。もしここが有意である，つまり丸いという状況から異なっている現状があると思われるのであれば，自由度などの補正が必要で，その時 anovakun はオプションで補正を指定できます。ともかく今回は，このまま先に進め

そうです[*12]。

R の出力 22.10: 下位検定の出力

```
<< POST ANALYSES >>

< MULTIPLE COMPARISON for "A" >

== Shaffer's Modified Sequentially Rejective Bonferroni Procedure ==
== The factor < A > is analysed as dependent means. ==
== Alpha level is 0.05. ==

--------------------------
  A   n    Mean    S.D.
--------------------------
 a1   4  7.5000  2.6458
 a2   4  3.5000  1.2910
 a3   4  5.5000  2.5166
--------------------------

------------------------------------------------------------
  Pair     Diff  t-value  df       p   adj.p
------------------------------------------------------------
 a1-a2   4.0000   5.6569   3  0.0109  0.0328  a1 > a2 *
 a1-a3   2.0000   2.8284   3  0.0663  0.0663  a1 = a3
 a2-a3  -2.0000   2.8284   3  0.0663  0.0663  a2 = a3
------------------------------------------------------------


output is over --------------------///
```

その先にはやっと分散分析表が出てきました。これを見ると Source 列の s とあるところ，これが個人差です。ここについては検定の対象外なので，F 値などの計算は行われていません。その下の A からが要因の効果で，s x A とあるところ，つまり A に伴って生じる個人差，のところが比較対象となる誤差のところです。この F 値が 16.00，$p = 0.0039$ で統計的に有意 (ちなみに効果量 $\varepsilon^2 = 0.3896$) なので，水準間のどこかには差があるということがわかります。

続く<< POST ANALYSES >>のところに，Sheffer の方法を使った事後の検定があります。これによると，a1-a2 のペアは $a1 > a2$ の関係にあって有意で，そのほかのペアは有意ではない，ということがわかります。

結果としては，たとえば次のように報告するといいでしょう。

> 一要因 3 水準の分散分析を行ったところ，$F(2, 6) = 16.00, p = 0.0039, \varepsilon^2 = 0.386$ で統計的に有意であり，事後検定の結果 a1 水準が a2 水準よりも高いことが示された (Shaffer の方法による多重比較)。

[*12] まったく！検定の多重性が問題だとかなんとか言いながら，そもそも前提が満たされているかどうかの検定は挟まないといけないし，ここでは有意差がでると補正，出ないと OK ってなんかいつもとは違う判断をしないといけない。分散分析とか帰無仮説検定というのは，とても厳密な手続きが必要なのですが，こんな感じであちこちに補正，修正などが入ってきて，正直嫌になりますよね。私は嫌なのでそういう悩みのないベイズ統計的アプローチの方がいいと思うんですが，みんなよくこんなのまじめにやってきたなあと思いますよ。

以上で解説を終わります。分散分析を面倒で煩雑な作業だ，と思った人も少なくないと思います。事実そうなのですが，だからと言って適当に実行してしまうと，誤用・誤解など問題のある研究実践になる可能性がありますので，細部まで細かく配慮しながら，どこにどういう意味があって，どのような読み取り方をするべきなのかをしっかりと理解しておいてください。

このほか，たとえば三要因計画になるとか，混合計画になるなどすると，それぞれ分散分析表や事後の分析の読み取り方などでアウトプットがどんどん長くなります。が，そこで注意を切らすのではなく，落ち着いて眺めてみてください。基本的には同じような前提，作業工程を，繰り返し繰り返し適用していくだけなのです。複雑で難解なのではなく，単純で煩雑なので嫌気が差してしまうだけです。ここでしっかりと腹を据えて読み取れるようになれば，分散分析は意外と素直なモデルに見えるかもしれません。

それでも帰無仮説検定という実践的判断基準を実践するので，ある種の断定 (有意差があった，なかった) が以後の解釈を歪めてしまいかねないという可能性は否定できませんし，確率的判定をしている以上は検定の多重性の問題からは逃れることができません。これを乗り越えるためには，もう少し確率的なモデルのことを詳しく理解する必要がありますし，モーメント法 (平均値や不偏分散で母数を推定する) ではないほかの方法で考える必要もあります。続く講義ではこれらについて解説していきます[*13]。

22.5 課題

■課題 1 線形モデルとしての分析 ある学校の国語の授業で，ある教師が 4 つのクラスそれぞれに対して個別のテキストを使って授業を行い，同一のテストを行うことでテキストの効果があるかどうかを検証したとします。各クラスには 10 人の生徒がいて，それぞれのテスト結果が次の通りです (表 22.4)。次の各指示に従い，その結果について回答してください。なお，このデータセットは `chapter22exe1.csv` ファイルに入っています。

表 22.4 課題 1

textA	textB	textC	textD
56	49	40	48
55	53	41	53
62	49	47	54
53	52	48	51
53	52	43	51
56	46	40	48
59	49	44	45
51	51	43	49
57	54	43	51
59	50	49	49

1. 回帰分析を行ってください。切片の推定値はいくらになりましたか。
2. 回帰分析を行った結果，テキスト A にくらべてテキスト B の平均値はどれぐらい違うと思われますか。
3. 回帰分析を行った結果，テキスト A を基準とした時にテキスト C に向けて引いた回帰線について，正

[*13] 分散分析はオワコンだ，と思われて久しいのですが，今でも学術論文には帰無仮説検定が実践されているので，みなさんには古典を読むためのリテラシーとして教えています。

しい記述を選びなさい。

4. 分散分析を実行し，その時の F 値を答えてください。
5. 分散分析を実行し，ボンフェローニの下位検定の結果として平均値の差が認められなかったペアを答えてください。

■課題 2 Between デザインの分散分析 別の学校では，テキストを A,B,C の 3 種類とし，教員 ℵ, ∇ の 2 人がそれぞれのテキストを使って教えることにしました。1 クラス 12 人，一学年 6 クラスを使っての実験です。授業後のテストの点数が表 22.5 のようになっています。次の各指示に従い，その結果について回答してください。なお，このデータセットは `chapter22exe2.csv` ファイルに入っています。

表 22.5 課題 2

textA-ℵ	textB-ℵ	textC-ℵ	textA-∇	textB-∇	textC-∇
25	32	30	28	23	34
39	40	29	20	27	42
36	39	25	31	27	42
31	29	23	19	26	37
41	42	19	35	16	35
34	21	24	24	27	35
33	43	14	25	19	33
31	26	19	34	21	41
28	44	31	23	35	34
34	31	26	22	30	38
31	26	31	19	17	28
32	29	30	23	27	33

1. 分散分析を実行し，テキストの効果について正しく述べてある文章を選んでください。
2. 分散分析を実行し，教員の効果について正しく述べてある文章を選んでください。
3. 分散分析を実行し，交互作用の効果について正しく述べてある文章を選んでください。
4. 下位検定の結果，テキスト A における教員の効果は合ったと言って良いでしょうか。正しく述べている文章を選んでください。
5. 下位検定の結果，教員 B が受け持ったクラスにおいて，どういう効果があったと言えるでしょうか。正しく述べている文章を選んでください。

■課題 3 Within デザインの分散分析 さらに別の学校では，テキスト・教員は一種類・一人なのですが，4 つの時期において成績が変化するかどうかを追跡調査しました。各時期のテストの点数が表 22.6 のようになっています。次の各指示に従い，その結果について回答してください。なお，このデータセットは `chapter22exe3.csv` ファイルに入っています。

1. 分散分析を実行し，時期の主効果について正しく述べている文章を選んでください。
2. 下位検定を見て，効果があった箇所を選んでください。

表 22.6　課題 3

id	period1	period2	period3	period4
1	39	38	42	52
2	59	66	62	56
3	27	33	44	44
4	41	38	41	42
5	46	46	44	51
6	22	21	37	27
7	54	45	52	51
8	59	51	76	65
9	49	55	59	66
10	48	37	41	47
11	34	36	39	45
12	65	54	62	69
13	51	42	43	43
14	44	37	53	53
15	42	47	50	54
16	49	48	61	47
17	48	39	62	49
18	46	50	44	46
19	45	40	48	55
20	44	48	46	57

第 23 章

確率分布とデータの関係

検定では平均因果効果を検証するため，平均値や (不偏) 分散などの標本統計量がどのような確率分布に従うかを参照し，結果を判定するという使い方を行なってきました。検定はこのように，母数を推定する方法に，結果を判定するロジックが組み合わさったものなのでした。この時の母数の推定方法は，標本平均や不偏分散を推定値とする方法で，モーメント法とよばれるものです。これに対し，モデルである確率分布をデータに最も近しい形に調節することで，そのパラメータを推定値とする方法があります。ここではこの方法，最尤推定法とその根幹にある尤度という考え方を解説します。

23.1 確率モデルによる推論

もう一度，後期のテーマである統計的推論の基本を振り返ってみましょう。

私たちは，手元にあるデータの状態を記述するだけでなく，これを使って「推論」をしたいわけです。覚えていますか？ 手元のデータ (標本) において，群間に差があるかどうか！なんて考えなくても「そりゃあるでしょ」となります。微細な差でも差は差ですから，それは違いありません。問題は，手元の標本の特徴を使って，「だから人間は･･･」などと大きな主語のことを言おうとすると，ちょっと待て待て，となるわけです[*1]。たまたま標本の数字がそうだっただけじゃないか，というわけですね。ここでいう「大きな主語のことを言おうとする」，というのが統計的推論という話なので，そもそも雲を掴むようなというか，無理難題に挑戦しているというのが大前提，何も確定的に言えないんじゃないかとさえ思えてきます。

そこをぐっと堪えて，仮に，仮にですよ？ 仮に母集団の数字が正規分布に従うと仮定できるのであれば，そこから原理的に標本統計量の特徴が導かれますから，それを使って推論しようよ，という話をしてきたのでした。

さて，今回までの話は母集団が正規分布するなら，標本統計量がある程度の幅を持った正規分布に従うので，標本平均や標本分散を使って推定しましょう，というアプローチをとってきました。平均値や分散など，標本の特徴そのものから母数を推定してきましたが，今回からお話しするのは少し異なるアプローチです。それは，**確率モデル (stochastic model)** の考え方の入り口であり，これを知ることで線形モデルの表現がより豊かになっていきます。

[*1] でも人間は，自分の経験を使っていろいろ言いたい生き物です。政治家だって，タレントだって，それでいろいろ失敗してますよね。でもたとえば「どうして男性は浮気するのか！」と怒っている人に「サンプルサイズが少なすぎる。それでは妥当な推論ができないから，せめて 20 人以上と交際してから言え」なんていうと逆ギレされるかしらけられるかのどちらかです。統計家は辛いよ。

23.2 確率関数と尤度関数

確率とデータの関係について，改めて考えてみましょう。

確率分布の形がわかっていて，パラメータもわかっていたとします。たとえば $\theta = 0.5$ のベルヌーイ分布だとか，平均値 $\mu = 5$，標準偏差 $\sigma = 1$ の正規分布，といった具合に。そうすると，そこからデータがどう出てくるかを考えることができます。ベルヌーイ分布の例がわかりやすいので，それで考えてみましょう[*2]。とあるコインの表が出るか裏が出るかが，$\theta = 0.5$ のベルヌーイ分布に従う，というのがわかっていたら，きっと表 (1) と裏 (0) が半々で出てくるだろうな，と思えるわけです。$\theta = 0.7$ であれば，こりゃ表がよく出てくるだろうな，といったことがわかります。

ただ私たちが，確率的なことを考える時というのはもっぱら逆です。表，表，裏，表，表･･･とどうも表がたくさん出てくるという現実があり，「これのパラメータは $\theta = 0.5$ ではないのではないか (半々じゃない，イカサマコインだ)」といったように，実現値が既知でパラメータが未知な状況です。研究をしている時も，手元にデータがあって，パラメータがわからないわけです。これまで推論してきた母平均，母分散というのもパラメータですが，そこから得られたと思しきデータを使って推論してきたのでした。

もう少し考えを深めてみましょう。ベルヌーイ分布は次のような確率分布です。

$$p(k) = \theta^k (1-\theta)^{1-k}$$

ここで k は 1 か 0 のどちらかの数字をとります。表が出る確率 $p(1)$ は，$\theta^1(1-\theta)^0 = \theta$ になり[*3]，裏が出る確率は $p(0) = \theta^0(1-\theta)^k = (1-\theta)$ になる，ということを数式で書くとこうなるのです。試しに $\theta = 0.7$ とすると，表が出る確率が 0.7 で裏が出る確率が 0.3 になっていますね。わかりやすい。

ここで $\theta = 0.7$，のようにパラメータがわかっている状況は普通とは逆なのですから，パラメータがわからない未知のまま (θ の記号のまま) 考えを進めたいと思います。もしコインが表，表と 2 回続けて表が出るとすると，その確率は $\theta \times \theta = \theta^2$ です。3 回続けて出たら θ^3，表，表，裏，裏，表なら $\theta^3 \times (1-\theta)^2$，と文字を使って表現できます。ここで θ は 0 から 1 の間を取る数字ですから，横軸に θ をとり，，縦軸に $y = \theta^3 \times (1-\theta)^2$ として関数のグラフを書くことができますね。これを書いたのが図 23.1 です。これはパラメータがどのあたり

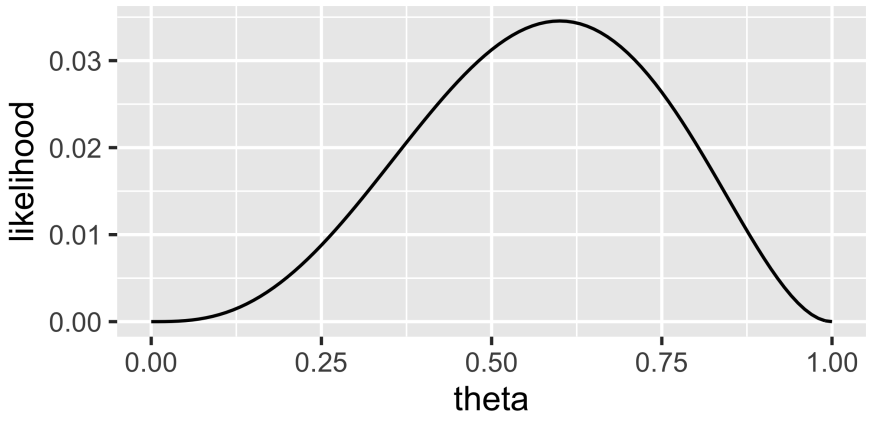


図 23.1 「表，表，裏，裏，表」と出たコインのパラメータの関数

[*2] ベルヌーイ分布については第 7 講，セクション 7.3,Pp.113 を参照。

[*3] ある数字の 0 乗は 1 になる，というのが累乗の計算のルールです。

にありそうか，という情報を持ったグラフになっています。このように，データが既知でパラメータが未知のとき，パラメータの関数のことを**尤度関数 (Likelihood function)** といい，このグラフの縦軸 y のことを**尤度 (likelihood)** と言います[*4]

尤度関数は確率関数と形 Form は同じですが，考える向きを逆にしたものだと言えます。つまり，

- 確率関数；パラメータがわかっている → 実現値がどのように出てくるかを表現
- 尤度関数；実現値がわかっている → パラメータがどのあたりならこのデータが得られそうかを表現

という関係にある，というわけです。

23.3 最尤推定

さてこの尤度関数，横軸が θ ですから，いろいろな θ の時にこのデータが得られる可能性はどれぐらいありそうか，というグラフだといえます。そうするとこの山のピーク，ここでは $\theta = 0.6$ なのですが，これが「このデータから考えられる最もありえそうなパラメータの値」ということになります。尤度が最大になるところをそのパラメータの推定値とする，という意味で，この推定値を**最尤推定値 (Maximum Likelihood Estimation)** といい，この方法でパラメータのことを推定する方法を**最尤法 (Maximum Likelihood method)** といいます (さいゆう推定，と読みます。略して ML 法とかかれることも)。

5 回中 3 回表が出たから，パラメータは 0.6 なのは当たり前じゃないか，と思ったかもしれません。それは標本平均を使って推定した，ということですが，同じ現象・データなので推定値が同じになるのはまあ当然のことです。答えは同じなのですが，その答えを出した筋道が違う，ということに注目してください。

最尤法は，関数の最大値を求めることによって最尤推定値を算出します。「関数の最大値を求める」という方法は，具体的には微分をつかって極限を求めることになります。微分について，高校数学で学んできたことを思い出して欲しいのですが，要するに「関数の傾きがフラットになる (傾き 0) ところがその関数の最大値，あるいは最小値になっている」という性質から，傾きの関数を求めてそれがどこなのか，を突き止める方法であり，まさに今回のようにピークを求めるにはもってこいのツールだったわけです[*5][*6]。

細かい数式はともかく，要するに関数のピークは最も尤もらしい，つまりこのデータが出てくる確率を最大にする値としてパラメータを推定できるということです。

ここで一点，注意を促しておきます。尤度関数は，確率分布関数の未知数と既知数を取り換えただけのものです。確率分布関数は，確率の出目 (実現値) に対応する確率の一覧を数式で表現したものだったことを思い出してください。また，確率とは非負で，その総和が 1.0 になるような数字なのでした。ここで図 24.1 を

[*4] ゆうど，と読みます。犬度 (いぬど) ではありません。尤度の尤は，「もっとも」と読み「尤もらしい」といったりします。

[*5] 高校までの数学は，こうした目的を先に教えておいてくれたらイメージしやすいのですが，先に概念や計算方法だけ教わるので苦手意識を持っている人もいるかもしれません。完成図を見ずにパズルのピースだけ渡されても，興味は持ちにくいかもしれませんが，幸い目的を持って学び始めると，いつ・どこから始めても必ず道がつながっているので，いつでも学び直すと良いでしょう。

[*6] 今回，この関数を微分すると，

$$\begin{aligned} y &= \theta^3(1-\theta)^2 \\ &= \theta^3(1-2\theta+\theta^2) \\ &= \theta^3 - 2\theta^4 + \theta^5 \text{より} \\ y' &= 5\theta^4 - 8\theta^3 + 3\theta^2 \\ &= \theta^2(5\theta^2 - 8\theta + 3) \end{aligned}$$

ここで $y' = 0$ とおいて極値を求めると

$$(5\theta^2 - 8\theta + 3) = (\theta - 1)(5\theta - 3)$$

より，$0 < \theta < 1$ の範囲では $\theta = 3/5 = 0.6$ が極大ということになります。

見ていると，これもなんだか確率分布のように見えたりもしますが，実はこれは確率分布ではありません。というのも，この関数の面積が 1.0 にならないからです。

簡単な例で解説しましょう。コイントスを 2 回連続でやることを考えてみたいと思います。この時何回表が出てくるかは，0 回 (裏・裏)，1 回 (裏・表，表・裏),2 回 (表・表) のいずれかです。

表 23.1 コインを二回トスします。

θ	0 回	1 回	2 回	合計
0.00	1.00	0.00	0.00	1.00
0.25	0.5625	0.375	0.0625	1.00
0.50	0.25	0.50	0.25	1.00
0.75	0.0625	0.375	0.5625	1.00
1.00	0.00	0.00	1.00	1.00
合計	1.9525	1.175	1.9525	

表 23.1 をみてください。これはコインを 2 回トスして，何回表が出たかについて，θ をいろいろに変えて計算してみた表です。たとえば $\theta = 0.0$ であれば，完全に裏しか出ないコインなのですから，0 回の確率が 100% で，1 回，2 回出る可能性は 0% です。この行方向の総和は 1.00 になっています。どの行についても，行の総和は $1.00 = 100\%$ で，θ が決まっているとしたらデータがどうなるかの確率は計算できるわけです。

一方，2 回コイントスして，1 回も表が出なかったという事実が先にあったとします。表 23.1 でいうところの 0 列，縦方向にみるのですね。θ は 0 から 1 のどこかにあるのですが，$0.00, 0.25, 0.50, 0.75, 1.00$ の 4 パターンだけピックアップしてみてみても，その総和は 1.00 を超えています。これを見ると，1 回も表が出ないパターンというのは，$\theta = 1.0$ であるはずがなく，$\theta = 0.75$ のような高い確率でもなさそう，$0.5, 0.25$ とパラメータが小さくなるに従って「そうなりそうな度合い」は高くなっていき，$\theta = 0.00$ であるというのが一番あり得そうではありますが，$\theta = 0.00$ の確率が高い，とは言えないのです。列の総和が 1.00 にならないから，これ (尤度) は確率じゃないからです。尤度は確率ではないのです[*7]。

尤度が最も高い，というのはそうなる可能性が最も高いわけですが，この数字は確率じゃない – ちょっと面倒な感じですが，言葉を正確にいうとこうなりますので，ご容赦ください。あくまでも，もっともらしさ，それっぽさが最大になるように推定値を定めるというのが最尤推定の正しい説明になります。

23.4 対数尤度関数

ところで，尤度の最大値を求める計算は，微分を使うのでちょっと面倒だな，と思うかもしれませんが，具体的な数値にしてしまえばそれほど大変ではありません。たとえば先ほどの例，ベルヌーイ試行 (コイントス) で表が 3 回，裏が 2 回でた，という場合ですが，$\theta = 0.3$ であれば，

$$0.3^3 \times (1 - 0.3)^2 = 0.3 \times 0.3 \times 0.3 \times 0.7 \times 0.7 = 0.01323$$

と計算できます。最尤推定値である $\theta = 0.6$ は，その尤度が

$$0.6^3 \times (1 - 0.6)^2 = 0.6 \times 0.6 \times 0.6 \times 0.4 \times 0.4 = 0.03456$$

である，と計算できるわけです。もちろん R などコンピュータに計算させれば良いわけですが，問題はこれ，数字が小さいってことなんですよね。

[*7] この表を使った説明の仕方は，私がみてきた「尤度は確率じゃない」の説明の中で最もわかりやすいもので，Lambert (2018b) にあった説明の仕方です。ここではその数字を少し改変して紹介しました。

そもそも確率は 1 を超える数字になることがないので，小さい数字に小さい数字を次々掛け合わせていくと，どんどんと桁数が小さくなります。コンピュータは計算機ですから，「掛け算せよ」と言われたらどこまでも掛け算はしますが，その内部では数字とは言え 0/1 のビット情報の連なりで，物理的上限があるので無限に小さな数字を計算できるわけではありません。無限どころか，たとえば今の PC は 64 ビット OS が基本ですが，どんなコンピュータも数字に割り当てれる上限を持っています[*8]。64 ビット OS の場合は，1 つの数字に 64 ビットの情報を持たせますが，これは桁数でいうと 2^{64}，およそ 20 桁の計算が限界です。何が言いたいかというと，コイントスが全部で 5 回ぐらいならコンピュータで計算できますが，10 回，20 回となるとだんだん限界に近づいて，そのうち「これ以上は小さすぎて計算できません！」というエラーが出てくるということです。

そこで実際の計算には少し工夫が必要です。尤度関数をそのまま使うのではなく，尤度関数の対数をとって計算するというのがそれにあたります。

対数 (logarithm) というのは，その数字を何乗したかというのを表すための表記法で，log という記号を使います。$a^P = M$ のときに，$p = \log_a M$ と表記し，この時「p を a を底とした M の対数」という言い方をするのです。これが便利なところとして，対数にしても数字の大小関係は変わらないこと，掛け算が足し算に，割り算が引き算に変わるので計算がしやすいことなどが挙げられます[*9]。記号を使ってより正確に書いておきますと，次の通りです。

- $M > N$ のとき，$\log_a M > \log_a N$；たとえば $100 > 10$ だが，$\log_{10} 100 = 2, \log_{10} 10 = 1$ より $\log_{10} 100 > \log_{10} 10$ です。
- $\log_a MN = \log_a M + \log_a N$；たとえば $\log_2 8 = 3, \log_2 16 = 4$ ですが，対数の定義より $2^3 = 8, 2^4 = 16$ という関係を表していることになります。ここで $\log_2 8 \times 16$ は，$2^3 \times 2^4$ の $\log_2$ をとったもの，ということになりますが，$2^3 \times 2^4 = 2^{(3+4)} = 2^7$ ですから，$\log_2 8 \times 16 = \log_2 8 + \log_2 16$ と，足し算が掛け算になります。
- $\log_a \dfrac{M}{N} = \log_a M - \log_a N$；これも同様に，$\log_2 \dfrac{16}{8} = \log_2 2 = 1$ ですが，$\log_2 16 - \log_2 8 = 3 - 2 = 1$ で計算できます。

たとえば先ほどのコイントスの計算も，$\theta = 0.3$ の例について対数を取って計算すると[*10]，

$$\log(y) = 3\log 0.3 + 2\log 0.7 = 3 \times -1.203973 + 2 \times -0.3566749 = -4.325268$$

となり，「桁数が小さすぎて計算できない」という問題が生じなくなります。また，対数をとったグラフは図 24.3 のようになりますが，ピークの箇所は変わらないことが確認できます[*11]。

23.5 正規分布の尤度関数

対数の話は計算上の問題なので知識として頭の片隅に置いておいてもらったら OK です。それよりも，関数の形をそのまま使ってパラメータを推定できる，ということをしっかり踏まえていただければと思います。この方法だと，不偏推定量を考えるとか，標準誤差を考えるといったことがなくなるので，もしかするとこちらの

[*8] 1995 年に Windows95 という OS が出ましたが，それは 32bit が単位の OS でした。Nintendo のファミリーコンピュータは 8bit で，Nintendo64 というハードは 64bit 単位で計算できるようになった，ということを示すために「64」が入っているのです。

[*9] 対数はそもそも天文学など，桁数がとても大きな数字の計算に使われるものでした。対数を取るとどういう数字になるか，という対応表を考えたネイピアさんのおかげで，天文学者の寿命は 2 倍に伸びたと言われるほど画期的に便利な発明でした。

[*10] この時の底は何？ と思うかもしれません。とくに明示されていない場合，底は自然対数 e を取ります。e はネイピア数とも呼ばれ，2.7 ぐらいの数字です。π のように，数学でよく出てくる定数の一種です。あるいは 10 を底にすることもあり，こちらは常用対数と呼ばれます。

[*11] 正直わかりにくいよね，でもあってますからご容赦。

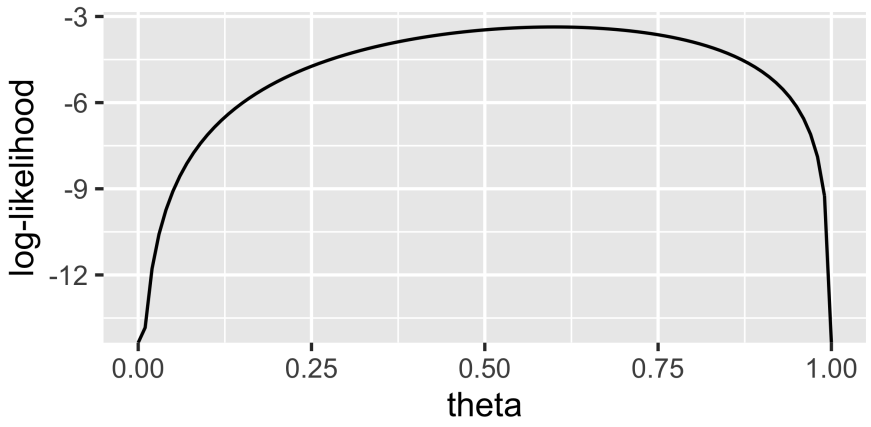


図 23.2 「表, 表, 裏, 裏, 表」と出たコインのパラメータの関数について対数をとったもの。縦軸は対数尤度

方が楽だな, と思われるかもしれません。注意しておきたいのは, これは「ある推定値が得られる確率」を考えているわけではないことです (尤度は確率ではないので)。また, データに最もぴったり合うようにパラメータを定めた, という方法なだけで, そもそもデータに当て嵌める確率分布関数 (=確率モデル) が間違っていたら, 何の意味もないことに注意してください。コイントスの結果にはベルヌーイ分布 (あるいは二項分布) を当て嵌めるのが正しく, 他の分布 – たとえば正規分布 – を考えると「答えは出るけど, そもそも全部間違ってる」というような話になります[*12]。とは言えここまで見てきたように, 心理学の場合は, 多くのデータに正規分布を仮定できます。私たちは正規分布という「仮定」, 「モデル」をおいて現象を理解しようとしているんだ, ということを忘れないでください。

そう言えば, 正規分布を使った最尤推定はどうなるんでしょうね。正規分布は平均 μ と標準偏差 σ という 2 つのパラメータで調整されますから, この 2 つのパラメータが両方未知であるという連立方程式を解くことになります。

数値例を使って考えてみましょう。第 16 講で最初に考えた例, 「ある小さな学校で身体測定をしたところ, 3 人の身長はそれぞれ, 165cm,173cm,182cm だった」という話で考えてみたいと思います。話を単純にするために, 母標準偏差は $\sigma = 10$ だとわかっていたとします。こうすることで, 先ほどのベルヌーイ試行と同じで推定すべきパラメータはひとつ, μ だけになります。ここで知りたいのは, どんな μ をもった $N(\mu, 10)$ から $\{165, 173, 182\}$ というデータが得られたか, ということです。

仮に $\mu = 160$ として考えてみます。データの 165 という数字が $N(160, 10)$ から出てくる尤度 (尤もらしさ) を計算したいのですが, まず正規分布の式を再確認しましょう。次のような式でしたね。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

ここで, データ $x = 165$ というのが, $\mu = 160, \sigma = 10$ というパラメータのもとで出てきたとすると, 数字を代入しますから,

$$\frac{1}{\sqrt{2\pi \times 10^2}} \exp\left(-\frac{(165-160)^2}{2 \times 10^2}\right) = 0.03520653$$

[*12] 適切な確率分布関数を選択できていないことなどを総称して, モデル誤設定問題と言います。

となります[*13]。同じように，この正規分布 $N(160, 10)$ から 173 というデータが出てくる尤度は，

$$\frac{1}{\sqrt{2\pi \times 10^2}} \exp\left(-\frac{(173-160)^2}{2 \times 10^2}\right) = 0.01713686$$

となります。また，182 というデータが出てくる尤度は，

$$\frac{1}{\sqrt{2\pi \times 10^2}} \exp\left(-\frac{(182-160)^2}{2 \times 10^2}\right) = 0.003547459$$

となります。ちなみにこれらの数字は，R で `dnorm(165,mean=160,sd=10)` とすれば得られます。

イメージで考えてみるとわかりやすいかもですね。図 23.3 に $N(160, 10)$ から 3 つのデータ点 165，173，182 が出てきたシーンを図示しました。

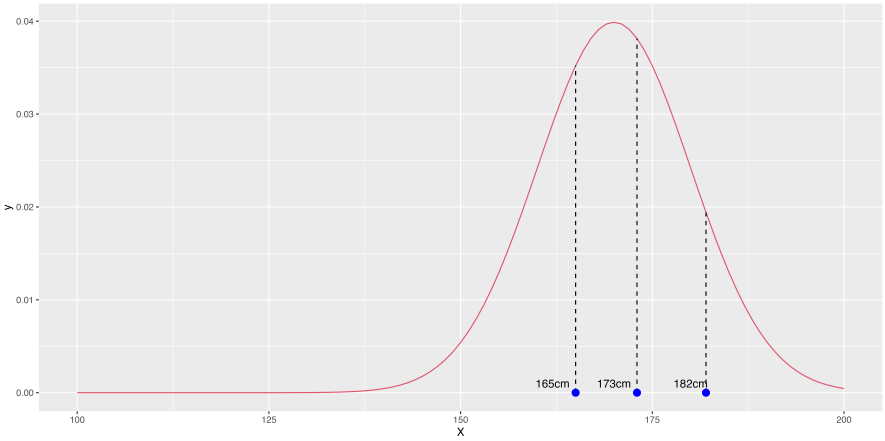


図 23.3 $N(160, 10)$ から 165, 173, 182 というデータが出てくる

これと同様の計算は，さまざまな μ について実行することができます。試しに $\mu = 150$ の時の図と，$\mu = 180$ の時の図を書いたものを図 23.4 に示します。

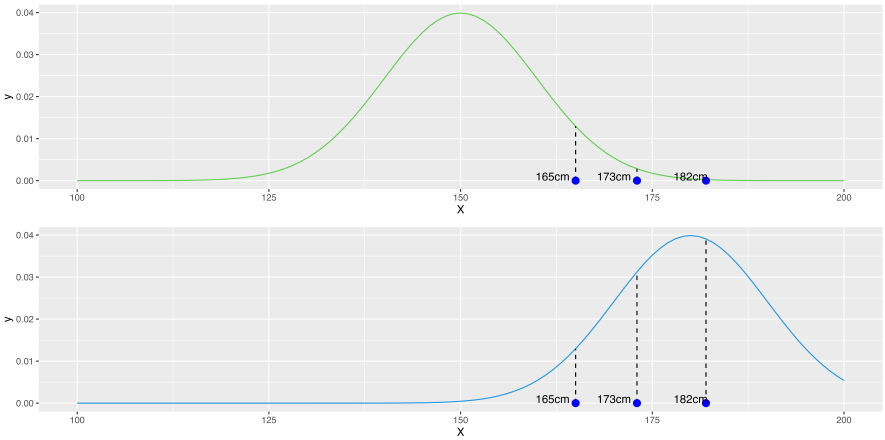


図 23.4 上の図は $N(150, 10)$，下の図は $N(180, 10)$ の例

[*13] 確認ですが，$\pi = 3.1415...$ という数字，$\exp(x)$ は $\exp(x) = e^x$ という指数関数で，ここで $e = 2.71828...$ という数字ですので，この式に未知数はないのです。

この図を見ると明らかなように，今回はいい μ を探したいのですが，正規分布の幅が決まっているので結局は山の平行移動をしていることになりますね。

さて，これら 3 つのデータは同じ母集団から独立に得られているわけです。独立に，というのはデータ間にとくに関連がないことですので，この 3 つが同時に成立する確率は掛け算で表すことができます。

考えてみてください。コイントスの場合もこの同じ分布から独立して得られている，という仮定がありました。同じコインですから同じ確率分布を仮定できます。**独立**というのは前の試行が次の試行に影響を与えない，という意味です。コイントスした時にコインが歪んでしまった (違う θ になった)，というようなことは考えていませんし，コインが「前は表になったから今度は裏を出してやろう」と考えるわけではありませんから，毎回確率 θ で表か裏になるはずです。ですから，繰り返した時に，掛け合わせて考えたのでした。ちなみにこの仮定のことを**独立同分布 (independently and identically distributed)** といい，i.i.d と略記して表現することがあります。

さて今回の 3 人，165cm,173cm,182cm という身長の持ち主も，同じ母集団からのサンプルですし，ある人の身長が他の人の身長に影響するとは思えませんから[*14]独立，つまりこのデータは i.i.d で，出現確率は同じ正規分布の関数を掛け合わせることで計算できますね。

これを計算すると，

$$L_{160} = 0.035 \times 0.017 \times 0.003 = 0.0000002140286$$

になります。このようにとても小さい数字になりますので，対数をとって

$$LL_{160} = \log 0.035 + \log 0.017 + \log 0.003 = -13.05457$$

とした方がいいかもしれません。

もちろんほかのケース，たとえば $\mu = 150$ ならどうか，$\mu = 180$ ならどうか，といろいろ試すことができますね。実際 `dnorm(173,mean=150,sd=10)` とか `dnorm(182,mean=180,sd=10)` と言った形で計算できます。尤度は次のようになります。

$$L_{150} = 0.0130 \times 0.0028 \times 0.0002 = 0.0000000008746$$

$$L_{180} = 0.0130 \times 0.031 \times 0.039 = 0.0000015814$$

数字が小さいので対数尤度でみましょう。

$$LL_{150} = \log 0.0130 + \log 0.028 + \log 0.0002 = -4.346524 - 5.866524 - 8.341524 = -18.55457$$

$$LL_{180} = \log 0.0130 + \log 0.031 + \log 0.039 = -4.346524 - 3.466524 - 3.241524 = -11.05457$$

となります。これを見ると，$LL_{150} < LL_{180} < LL160$ ですから，$\mu = 150$ と仮定するより $\mu = 180$ のほうが，そしてそれより $\mu = 170$ の方が**尤もらしい**わけです。もちろん μ は他にもあり得ますから，横軸を μ にした関数とみて，そのグラフを描いてみましょう (図 23.5)。

このようにしてみると，μ が 170 から 175 の間あたりにピークがありそうですね。このピークを計算するには微分[*15]の計算が入りますが，実際に計算すると 173.3 になります。ピーク，すなわち最も尤度が高いわけですから，これが今回の**最尤推定値**になります。

[*14] 双子だとか血縁関係ではないとします！

[*15] 微分はある点における関数の接線の傾きを表す，というのは聞いたことがあると思います。この傾きがゼロになっているということは，つまり関数がピークを迎えたということです。つまり，微分は極値を求める操作だったのです。忘れていても習っていなくても大丈夫，テキストはあちこちにありますから，よくわからんという人は少し調べてみましょう。

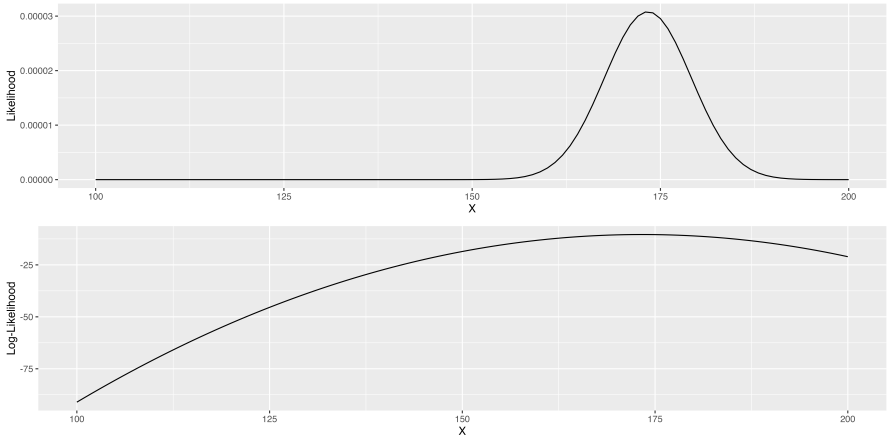


図 23.5 さまざまな μ における尤度 (上図) と対数尤度 (下図)

あれ, でもこれ標本平均と同じじゃないか, と思うかもしれません。その通り, 同じなのです。実は推定方法が違うだけで同じ母集団の推定方法なので, むしろ答えが一致してよかったわけですね。推定方法が違うと答えが違う, というのでは困りますから。

ともかく, このように正規分布でも最尤推定ができます。もっとも, 今回は母標準偏差がわかっている ($\sigma = 10$) という仮定の下で計算したのでした。本来はこちらも推定しなければなりません。つまり, 平均 μ と標準偏差 σ という 2 つの変数の関数になるのです。この 2 つの関数のグラフを書いたのが図 23.6 です。2 つの変数で 1 つの対数尤度が決まるので, 3 次元プロットになりますし, 2 変数の微分方程式を解くことになるので数学的には少し面倒になりますが, 同様に推定できることがわかります。

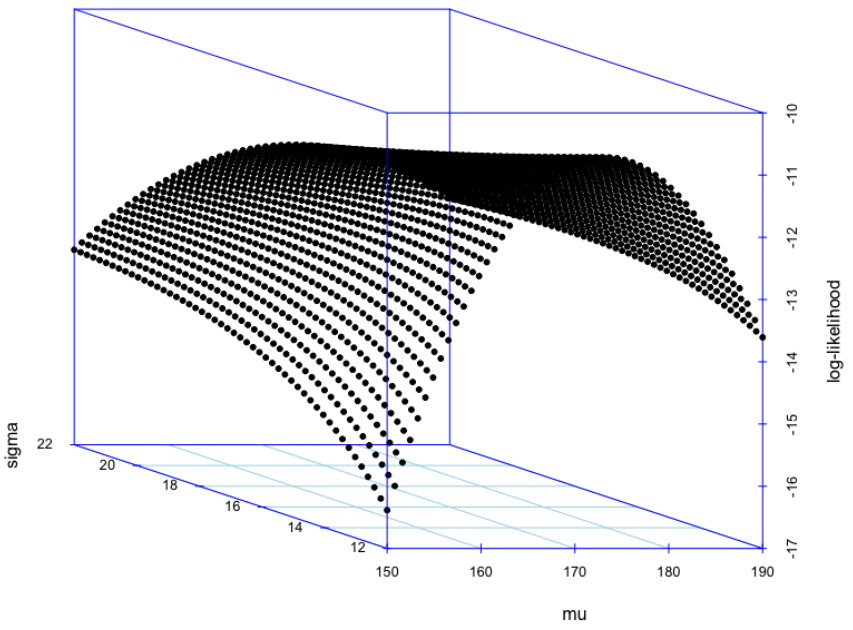


図 23.6 平均と標準偏差による 2 変数関数としての正規分布の対数尤度

もちろんこの 2 変数の計算は機械がやってくれるので，皆さんがそこで苦労する必要はありません。それよりも，確率分布に基づいて推定するという方法があることを理解していただきたいと思います。最後に今日のポイントをまとめておきます。

- **確率密度関数 (probability density function)** と同じ形で，パラメータの方が未知数の場合これを尤度関数という。
- 尤度が最大になる点をパラメータの推定値とする，という推定方法を最尤推定という。
- 尤度の計算は数字が小さくなりがちなので，対数をとった対数尤度で計算することが一般的。対数をとってもピークの場所は変わらない。

今回は統計モデルを確率モデルに拡張するというお話でした。確率モデルに一般化することで，さまざまなデータ形式に対応したモデルを考えることができるようになりますが，それはまたどこかの機会で。

23.6 課題

ある小学校で身体測定をしたところ，3 人の身長はそれぞれ,155cm,143cm,132cm だったとします。このデータに基づいて，次の問いに答えなさい。

■正規分布の尤度を求める 1 このデータが，平均 170，標準偏差 10の正規分布から独立に得られていたとすると，その時の**尤度**はどれぐらいになるでしょう。

■正規分布の尤度を求める 2 このデータが，平均 140，標準偏差 10の正規分布から独立に得られていたとすると，その時の**対数尤度**はどれぐらいになるでしょう。

■正規分布の尤度を求める 3 このデータが，平均 140，標準偏差 10の正規分布から得られていたと考えるモデルと，平均 140，標準偏差 8の正規分布から得られていたと考えるモデルがあったとします。データはどちらをより支持しますか。

■正規分布の尤度を求める 4 このデータが正規分布から得られると考えて，そのときの平均値の最尤推定値を求めましょう。

第 24 章

回帰分析の最尤推定

前回は確率分布関数に含まれる変数の中で，パラメータの方が未知であると考えた尤度関数として捉え，尤度が最大になるようにパラメータの値を推定する**最尤推定法**について学びました。

今回はこれを，線形モデルに適用するお話です。線形モデルについては前期から今まで，折に触れては説明してきました。まず第 9 講で回帰分析に触れ，続いて重回帰分析 (第 10 講)，それから要因計画へと展開してきたわけです。後期に入って推測統計学の話になり，要因計画と仮説検定の合体技，**分散分析**の話になりましたが，そこでの推定方法は**モーメント法**，すなわち**標本平均**や**不偏分散**を使ってのものでした。**平方和**を**自由度**で割って，**平均平方**を出し，その比が **F 分布**に従う，という考え方は，平均や分散から母数を考えようという発想と同じものだと言えます。

それに対して，前回は尤度から母数を考えようという発想でした。推定値は同じになるのですが，考え方が違っていることに注意して，線形モデルの最尤推定について考えていきます。

24.1 回帰分析と誤差分布

改めて線形モデル，とくに**単回帰分析**のモデルを復習しておきましょう。

データの組，x_i, y_i があった時に，変数 x をつかって y を予測する，線形モデルを当てはめるということを考えるのでした。線形モデルは最も単純な一次式で，

$$\hat{y}_i = b_0 + b_1 x_i$$

とし，予測値 $\hat{y}$ に誤差 e がついて実現値 y が得られると考えて，

$$y_i = b_0 + b_1 x_i + e_i$$

とするのでした。第 8 講でやった**最小二乗法**は，この式をデータに当てはめ，データの中で誤差が一番小さくなるように切片を求める，という方法だったことを思い出してください。あくまでも手元のデータの話しかしていませんでしたが，これに推測統計学のエッセンス，つまり手元のデータは標本に過ぎず，母集団の中でどうか，ということを考えたいと思います。

ところで，この時も仮定として誤差の分布に**正規分布**を仮定していましたね。誤差はさまざまな未知なる要因の積み重ねですから，数字を増やす方向に加わるのか，減らす方向に加わるのかは分かりません。それでも極端に大きな誤差というのは出にくく，平均周りの誤差が出やすいと考えるなら，単峰で左右対称のベルカーブのような形になるはずです。また，今回ここで考えているのは**偶然誤差**ですから[*1]，その平均はゼロだ

[*1] これに対して**系統誤差**というのがありました。システマチックな誤差，という意味で，常に定数が加わるようなイメージです。体重計などの秤で何も測っていない時にメモリがずれているようなものですから，今回のケースではその補正は触れません。測定する際に気をつけるべきことで，気をつけてもなくならないのが偶然誤差です。

と考えられます。ですから，$e_i = y_i - \hat{y}_i$ をデータ全体で最小にするようにモデルをフィッティングするのでした。ここまでは前期の復習になります。

24.2 確率モデル

この線形モデルを確率モデルと考えて，最尤推定したい，というのがここでの狙いです。前期の最後，第 14 講，セクション 14.4，Pp.206 のあたりを思い出していただきたいところですが，大事なことなのでもう一度説明しておきましょう。

最尤推定をするためには，この回帰分析の式を確率モデルで表現しなければなりません。先ほどのモデル式の中に出てくる記号のなかで，確率的に変わるのは何かと言われれば，誤差 e です。そしてこれは，正規分布に従うだろうと考えるのが自然なのでした。式で書くと，

$$e_i \sim N(0, \sigma)$$

ということになります。ここで誤差の平均が 0 である，ということを再確認してください。また，平均は 0 ですが，誤差は一定の幅で散らばって出てきますので，これを標準偏差 σ で表しています。

さて，誤差が確率的に求まることがわかりましたが，x と $\hat{y}$ の関係は確率的ではありません。切片や傾きは未知数ですが (母集団での切片と傾きのことになるので β_0, β_1 と表記しますが)，これがわかりさえすれば $\hat{y}_i$ と x_i の関係は確定的です。つまり，「確率分布に従う」を表す $\sim$ という記号ではなく，「左辺はこのようになっ**ている**」ということを記述するために，$=$ を使って書くことができます。

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

これが線形モデルです。誤差に邪魔されることがない，理想的な予測値 $\hat{y}$ は，x_i の一次関数の形で記述できる，ということです。実際のデータ y_i は $\hat{y}$ に誤差がついて出てくるものですから，

$$y_i = \hat{y}_i + e_i$$

と書くことができます。

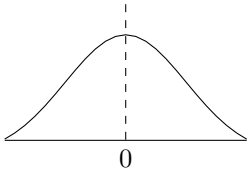
さて，これらを組み合わせていきましょう。最後の式から $\hat{y}$ を消去する (線形モデルで代入する) と，

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

ということになります。ここで $\beta_0 + \beta_1 x_i$ のところは確率的に変わるものではない，ということを思い出してください。

$$y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\text{確率的でないところ}} + \underbrace{e_i}_{\text{確率的}}$$

誤差 e_i は平均 0 を中心に確率的に散らばりますが，その前のところは確率的ではありませんので，これを組み合わせた式全体としては，最後に確率的な散らばりがひっついているような形になります。数式ではありませんが，イメージでいうとこんな感じ。

$y_i = \beta_0 + \beta_1 x_i +$ 

さて，この「確率的でないところ」が確率分布の中に入り込むとしたら，誤差 e_i の平均値を一定の量，$\beta_0 + \beta_1 x_i$ 分だけ動かすことになります。回帰直線が表しているのは，誤差のない予測の線でもあります。実際のデータはそれに誤差が伴って生じます。確率モデルになった線形モデルは，図 24.1 のようなイメージで考えるといいかもしれません。

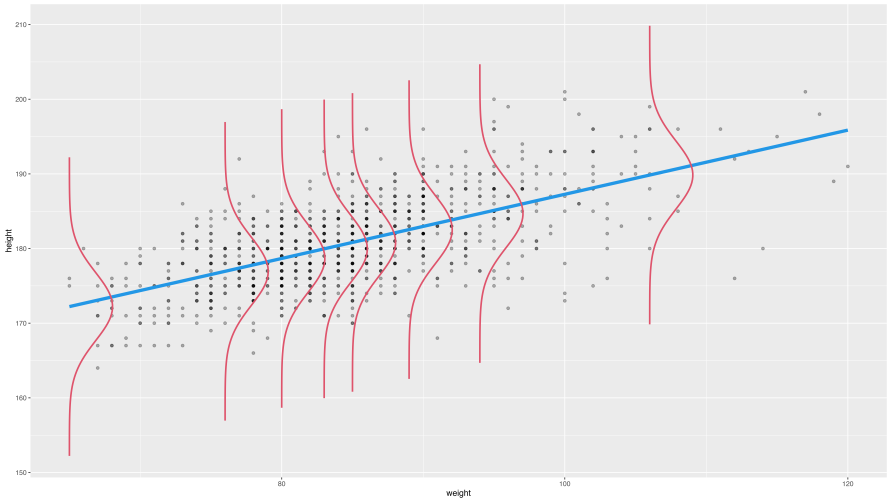


図 24.1 予測直線にまとわりつく誤差

ここで正規分布がもつ 2 つのパラメータ，μ と σ は，それぞれ**位置**と**幅**を表しているわけですが，確率的でないところはこの位置をズラすような作用をすることになります。

大事なところなので，もう一度式で確認しながら説明します。まず今回のモデルに含まれる情報は次のとおりです。

$$\hat{y}_i = \beta_0 + \beta_1 x_i \tag{24.1}$$
$$y_i = \hat{y}_i + e_i \tag{24.2}$$
$$e_i \sim N(0, \sigma) \tag{24.3}$$

ここで式 24.1 と 24.2 を合わせて，

$$\begin{array}{rcl} y_i & = & \beta_0 + \beta_1 x_i + e_i \\ e_i & \sim & N(0, \sigma) \end{array} \tag{24.4}$$

とします。ここで e_i の中心は 0 で，それに $\beta_0 + \beta_1 x_i$ が加わるのですから，

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma)$$

ということになります。

あるいは，式 24.2 と 24.3 から先に，

$$\hat{y}_i = \beta_0 + \beta_1 x_i \tag{24.5}$$
$$y_i \sim N(\hat{y}_i, \sigma) \tag{24.6}$$

と考えて，この式 24.6 の $\hat{y}_i$ に式 24.5 を代入しても，

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma)$$

が得られます。どちらのルートでも結構ですから，自分のわかりやすい方で理解してください。一覧の流れを図 24.2 にまとめてみました。

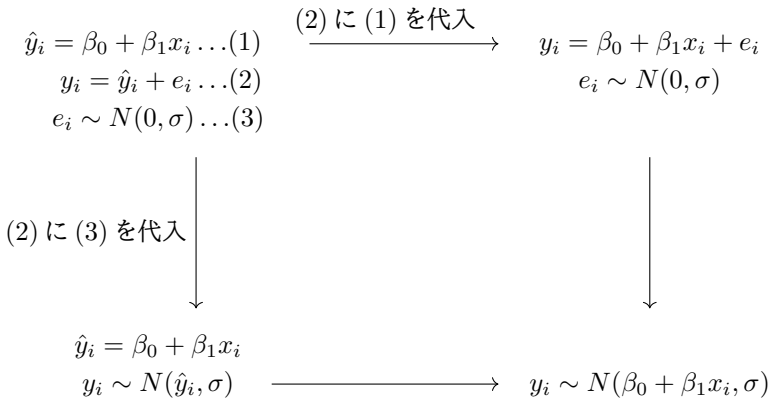


図 24.2 確率モデルと予測モデルの合体

これが回帰分析の**確率モデル**になります。ポイントは，確率分布の位置を表すパラメータに，数式で表現された**構造**を書き込んだところにあります。狭義の**モデリング**とは，確率分布のパラメータに構造を与えるもの，ということができるでしょう。発展的な話題になりますが，正規分布に限らず**ベルヌーイ分布**や**二項分布**，**ポアソン分布**などのパラメータを数式で表現したモデルを考えることもできます。正規分布以外の確率分布でパラメータに線形モデルを与えたものを**一般化線形モデル (Generalized Linear Model)** と呼びます[*2]。一般化線形モデルになると，比率や度数などより広い範囲のデータを扱う分析ができるようになります[*3]。

24.3 確率モデルの最尤推定

このようにして確率モデルを書くことができれば，あとはこれにしたがって「データがこの確率モデルから出てくる尤もらしさが最も高くなるように，未知パラメータを推定すれば良いということになります。改めて確認ですが，今回のモデル $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma)$ において未知な数字は β_0, β_1, σ の 3 つになります。

正規分布の式をちゃんと書くと，x という値が出てくる確率密度は

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

となります。この平均 μ のところに $\beta_0 + \beta_1 x_i$ という数式を入れることになります。

すでに入り組んだ形をしているので，イメージしづらいところがあるかもしれませんが，あるデータ y_i が得られる確率密度は

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)$$

ですから，$\mu = \hat{y}_i = \beta_0 + \beta_1 x_i$ を代入して

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}\right)$$

となります。

データは $i = 1, 2, \cdots, N$ とあり，それらが同じ母集団から独立に得られている，と考えます (i.i.d.)。つまり手元に同時にこれらのデータが出揃うときの尤度 L は，上記の確率密度 $(p(y_i))$ を全部掛け合

[*2] **一般線形モデル (General Linear Model)**(→ セクション 14.3,Pp.203) と混同しないように注意してください。
[*3] 本学では二年次選択科目のデータ解析応用で扱います。興味がある人，履修をお待ちしています。

わせたもの[*4], と考えることができますね。i を 1 から n まで足し合わせることを総和する, といって $x_1 + x_2 + \cdots + x_N = \sum_{i=1}^{N} x_i$ と表しますが, 同様にすべて掛け合わせる総積は,

$$x_1 \times x_2 \times \cdots x_N = \prod_{i=1}^{N} x_i$$

と書きます。ここでの記号 $\prod$ は総積(プロダクト)と言いますが, これを使って尤度 $L(Y_i)$ を表現すると, 次のようになります。

$$L(y_i|\beta_0, \beta_1, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

ここで $L(y_i|\beta_0, \beta_1, \sigma)$ という書き方の, $|$ は何かというと, この縦線の右側にある β_0, β_1, σ が与えられた場合の, という条件付き確率の書き方を借りた**尤度 (Likelihood)** を表しています。

さてこれは確率 (密度) の掛け算になりますから, すごく小さな数字になって実際の計算は大変です。そこで前回習ったように, 尤度そのものではなく**対数尤度 (log-likelihood)** をとって, その最大値を求めるということにしたいと思います。対数をとると掛け算は足し算に変わりますから, $\prod \to \sum$ に記号が変わりまして, 対数尤度の関数は次のようになります。

非常に面倒な形をしていますね！しかもこれは β_0, β_1, σ という 3 つの変数が同時に関わりますから, 連立微分方程式を解くことになります。

ここからは少しテクい話になりますが, この式を展開していきましょう。$\frac{1}{\sqrt{2\pi\sigma^2}} = (2\pi\sigma^2)^{-1/2}$ であることに注意して対数尤度を書き直すと,

$$\begin{aligned}
\log L &= \log \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}\right) \\
&= \sum_{i=1}^{n} \log\left[(2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}\right)\right] \\
&= \sum_{i=1}^{n} \left[\log(2\pi\sigma^2)^{-1/2} + \log\exp\left(-\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}\right)\right] \\
&= \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}\right] \\
&= \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2}\right] \\
&= \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2
\end{aligned}$$

こうしてみると, $\frac{n}{2}\log(2\pi)$ は定数ですし, $\frac{n}{2}\log(\sigma^2)$ は推定すべきパラメータ σ^2 で, データによって変わるところではありません。第 3 項の $\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$ がデータによって変わるところで, これを最小化することを考える必要がありますが, これは最小二乗法で考えるべきポイントと同じですね。ですから, この最尤推定の結果得られる β_0, β_1 の値は, 最小二乗法の結果と一致します。

[*4] コインが表, 表, 裏, と出たら $\theta \times \theta \times (1 - \theta)$ と書けたことを思い出して！

もちろんこの面倒な計算は，R がやってくれますから私たちはなんの心配もする必要はありません。ですから，実質的にはただの回帰分析をすれば良いのです。

24.3.1 実践例

数字としてはなんら変わるところはないのですが，復習も兼ねて第 9 講でやった回帰分析をもう一度見てみましょう。回帰分析の関数は `lm` という関数名で，R がデフォルトで持っている関数なのでした。実はこの関数，最尤推定値を返す関数だったのです。そのときは**最小二乗法**の説明しかしていなかったのですが，関数の使い勝手のこともあって，最尤推定する関数でご紹介しておりました。本当の意味で，最小二乗法のアルゴリズムで回帰分析をするときは `lsfit` 関数というものを使います。

使用例をコード 24.1 に示しました。ここでは野球選手のデータを使って，身長から体重を予測するというモデルを実行しています[*5]。

code：24.1 二つの推定法の比較

```
1 # 最小二乗法の推定
2 result.LS <- lsfit(x = batter$height, y = batter$weight)
3 result.LS$coefficients
4 # 最尤推定
5 result.LM <- lm(weight ~ height, data = batter)
6 result.LM
```

結果は出力 24.1 のようになります。

[*5] 第 11 講と同じ分析結果です。

R の出力 24.1: 比較した結果

```
> # 最小二乗法の推定
> result.LS <- lsfit(x = batter$height, y = batter$weight)
> result.LS$coefficients
  Intercept           X
-124.917552    1.166698
> # 最尤推定
> result.LM <- lm(weight ~ height, data = batter)
> summary(result.LM)

Call:
lm(formula = weight ~ height, data = batter)

Residuals:
     Min       1Q   Median       3Q      Max
-20.9217  -4.9213  -0.4214   4.0783  28.9118

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -124.91755   13.89317  -8.991   <2e-16 ***
height         1.16670    0.07746  15.063   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.435 on 327 degrees of freedom
Multiple R-squared:  0.4096, Adjusted R-squared:  0.4078
F-statistic: 226.9 on 1 and 327 DF,  p-value: < 2.2e-16
```

これを見ると明らかなように，どちらの方法でやっても同じ推定値 (切片 $\beta_0 = -124.918$, 傾き $\beta_1 = 1.167$) になっていますね。また少し余談ではありますが，最尤推定した結果の最終行を見ると，F 統計量が自由度 $F(1, 327) = 226.9$ とあり，この p 値も計算されています。ここで線形モデルと要因計画・帰無仮説検定も関連があることがわかります[*6]。

方法が違っても推定値が同じ，というのは前回もお話ししたように，同じ母数を推定しているわけですからむしろ当然の結果です。1 つの山の頂上に達するのに，複数のルートがあってもゴールは同じ，ということです。違う点があるとすれば，最尤推定の場合は誤差の標準偏差 σ も推定されているというところでしょうか。結果に $\sigma = ...$ という記載はありませんが，よく見ると `Residuals` のところに残差 e の最小値，最大値，四分位偏差，中央値などが示されており，誤差が分布するものと考えられていることがわかります。

今回は説明変数が 1 つの単回帰分析でしたが，説明変数の数が増えた**重回帰分析 (Multiple Regression Analysis)** でも同じで，`lm` 関数はデータが正規分布から得られたものとした上で，尤度を最大にする推定値を返してくれます。

R で回帰分析をしたついでに，この結果を plot 関数にいれてみましょう。

code : 24.2 回帰分析の結果オブジェクトをプロットに入れる

```
1    par(mfrow = c(2, 2)) # 2x2の配置にする
```

[*6] ここでは切片以外のすべての説明変数 (今回の例では単回帰なので説明変数は一つですが) の傾きが 0 である，という帰無仮説にたいして検定が行われています。

```
2    plot(result.LM)
```

4 つの図が表示されますが，ここでは 2 × 2 の配置にしています。出力されるのは図 24.3 のようになります。

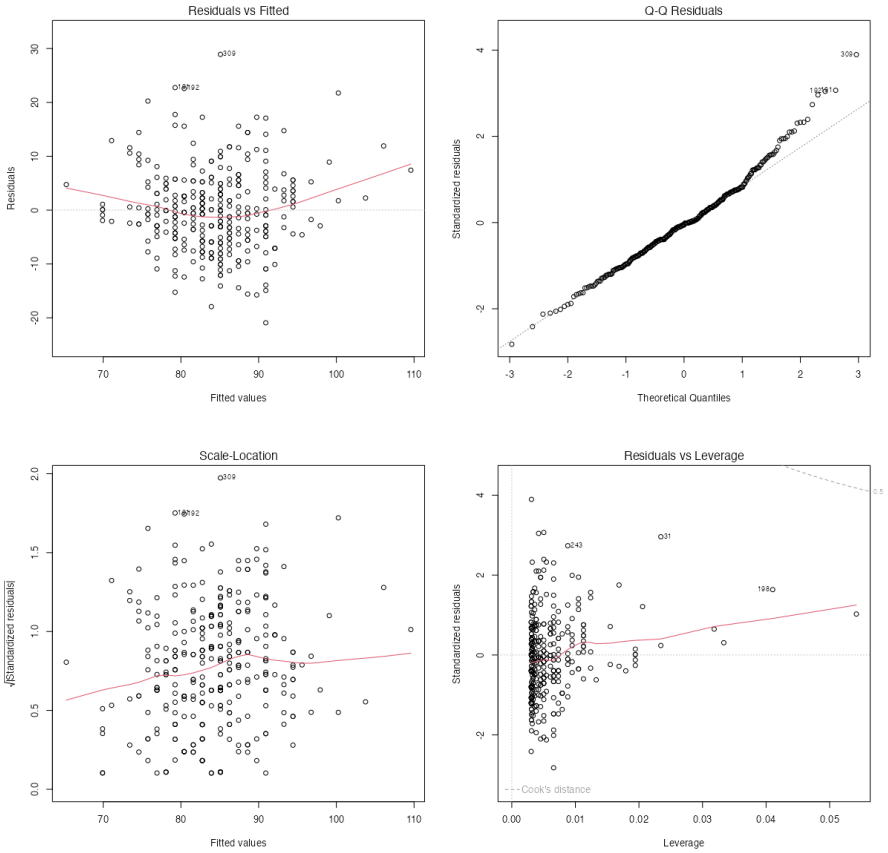


図 24.3 回帰分析の結果をプロットすると 4 枚の図が出てくる

最初の図 (図 24.3 左上) は，残差と予測値の散布図です。回帰分析において，残差 e_i と予測値 $\hat{y}_i$ の相関は理論的にゼロになります。実践上そのようになっているかを確認するための図が表示されています。今回の散布図をみると，一定の関係はなさそう (緩やかな U 字で，特定の関係があるように思えない) ですね。残差は規則性を持たず，均一に散らばっているはずです。この図を見てそれを確認しましょう。

第二の図 (図 24.3 右上) は **Q-Q プロット (Q-Q plot)** と呼ばれるものです。Q-Q プロットとは一般的に，データが仮定された確率分布に従っているかどうかを表示するためのもので，今回は横軸に正規分布の理論点を，縦軸に残差をプロットしています。理論通りの分布をしていると，左下から右上への 45 度線上にプロットされていくはずで，このラインから外れるのは理論分布から外れていることになります。今回は右の方，大きな値のところで少し歪みが生じていますが，ほとんどの点がライン上に載っていますので，まあ正規分布していたと考えられるでしょう。

第三の図 (図 24.3 左下) は残差標準偏差と予測値のプロット，第四の図 (図 24.3 右下) は残差とレバレッジの散布図です。レバレッジとは，各データ点がモデルに与える影響の大きさの指標です。これらも残差の分布が均一であるか，外れ値がないか，というところを検証するためのものです。

これらの図をみて残差が正規分布という仮定に従っているか，ということを確認することが重要です。こう

した**残差診断 (Residual Diagnostics)** は，モデルの適合性や家庭の妥当性を評価しておくことが重要です。最尤法は確率モデルなのですから。

さて，こうしてみてくると，最小二乗法と最尤法，結局答えが同じになるのに，なぜこんな回りくどい話をするのかと思うかもしれません。大事なのは線形モデルが確率モデルになったというところで，確率モデルになることで適用できる範囲が広がるから，というのが 1 つの答えです。ところがもう 1 つ大事なことがあるのです。それは次回以降，第三の母数推定法である**ベイズ法 (Bayesian estimation method)** に関係します。実はこの第三の方法も確率モデルを使ったものなのですが，詳しくは次回の講釈で。

24.4 課題

■回帰分析の結果を読み取る 出力 24.1 の読み取り方について，正しい記述をすべて選んでください。

- 回帰係数が 1.166 である確率は $p < 0.05$ なので，滅多に得られる数字ではないから統計的に意味がない。
- 回帰係数の有意性検定が示しているのは，傾きが 0 であるという帰無仮説のもとで今回の係数 1.166 が得られる確率が小さい ($p < 0.05$) ということだから，統計的に意味がある。
- 今回のデータから考えると，回帰係数を 1.166 としたときに尤度が最も高くなる。
- 今回のデータから考えると，回帰係数を 1.166 としたときに残差の総量 (二乗和) が最も小さくなる。
- 今回のデータから考えると，回帰係数が 1.166 となる確率は 95% 以上である。

■最尤推定の目的 回帰分析で使う「最尤推定」の目的として、最も適切なものを一つ選んでください。

- データの中から最大値を見つけること
- データがモデルから得られる確率が最も高くなるようにパラメータを推定すること
- 誤差が最も小さくなる点を見つけること
- データの平均値を求めること

■最尤推定の特徴 回帰分析における最尤推定について、正しい記述はどれですか？（複数選択可）

- 正規分布を仮定した場合、最尤推定による推定値は最小二乗法による推定値と一致する
- 誤差項の分布が正規分布でない場合でも、常に最小二乗法と同じ結果になる
- 対数尤度関数を最大化することで、パラメータの推定値を得る
- 最尤推定では必ず誤差の標準偏差 σ も推定される
- 最尤推定は常に不偏推定量を与える

■回帰分析の確率モデル 以下の式は何を表していますか？

$$y_i \ N(\beta_0 + \beta_1 x_i, \sigma)$$

- 誤差項の確率分布
- 回帰係数の確率分布
- 応答変数 yi の条件付き確率分布
- 説明変数 xi の確率分布
- 残差の確率分布

第 25 章

ベイズ推測の基礎

ここまで尤度を使った推定方法について説明してきました。モーメント法と最尤法，いずれも同じ答えが出るのですがその考え方が違うというところを確認してきましたね。さてここからは，もう 1 つの推定法である**ベイズ法**について解説します。そのためには，ベイズ法の根本的な法則について知っておいた方が良いので，少し回りくどいかもしれませんが，まずはそこから説明していきましょう。目的を先に教えてくれないとやだよ，という人のために一言で目的だけ言っておきます。ベイズ法は最尤法と同じく確率モデルを使うのですが，尤度が「確率じゃない」という妙な性質があったのに対し，ベイズ法は答えを確率で出します。尤度を確率にするための仕掛けがこれから始まるのです。

25.1　ベイズの定理

ベイズというのは人の名前で，トーマス・ベイズ（Thomas Bayes、1702 年 - 1761 年 4 月 17 日）はイギリスの長老派の牧師であり数学者でもある人です。この人の考えた「逆確率の法則」から端を発した推定法が，21 世紀の現代においてとても重要なものになってきています。ベイズ牧師の考えついた法則は非常に単純なもので，簡単に書くと次のような式で表現できます。

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

この式の意味するところを，確率の用語をいろいろ紹介しながら説明していきたいと思います。

話をわかりやすくするために，具体的な数値例とともに話を進めましょう。たとえば，ある学校の生徒数は男子が 40 人，女子が 60 人だったとします。そこからランダムに一人を選び出した時，それが男子である確率はどれぐらいでしょうか？ この確率を $P(\text{男子})$ とすると，

$$P(\text{男子}) = \frac{40}{(100)} = 0.4$$

であることはすぐにわかりますね。

では次に**同時確率 (joint probability)** という考え方を導入しましょう。これは先ほどの学校について，表 25.1 のように学年情報を追加した例を使って説明します。

記号で表現するために，性別の変数を A とし，A_1 = 男子，A_2 = 女子，学年の変数を B として B_1 =1 年生，B_2 =2 年生，B_3 =3 年生とします。

ここでランダムに一人を選び出した時，それが「3 年男子」である確率はどれぐらいでしょうか？ これは「学年は 3」かつ「性別は男」なので，学年と性別の両方の条件が入った同時確率といい，$P(A_i, B_j)$ と書きます。これは表から，次のように計算できますね。

$$P(A_3, B_1) = \frac{13}{100} = 0.13$$

表 25.1 ある学校の学年と性別のデータ

	1 年	2 年	3 年	合計
男子	15	12	13	40
女子	22	20	18	60
合計	37	32	31	100

次に用語として**周辺確率 (marginal probability)** というのを考えます。これはある分割に関してすべて足し上げたもの, というものですが, 要するに表 25.1 の合計欄のことです。たとえば「男子生徒についての周辺確率」は, 男子生徒が学年によって 3 つに分割されていますが, これをすべて足し上げるので, 1 年男子 +2 年男子 +3 年男子で求まるものです。式で言うと, 次のような意味になります。

$$\sum_{i=1}^{3} p(A_i, B_j) = p(B_j)$$

$$\sum_{j=1}^{2} p(A_i, B_j) = p(A_i)$$

全部の数字を足し合わせるので, その変数の確率の話ではなくなる, と思ってください。このように一方の変数のすべての確率を足し合わせることを**周辺化**する, といいます。

では次のステップ。今度は**条件つき確率 (conditional probability)** というのを導入します。無作為に学生を選び出す時, 「選ばれたのが女子である」ということが先にわかったとします[*1]。その時, その生徒が 2 年生である確率はどれくらいでしょうか？ この問いを「女子という条件がついた上での確率」という意味で, 条件つき確率といい, 表現としては $P(2\text{年生}|\text{女子})$ と書きます。2 つの分割を縦棒で区切り, 条件の方が後ろに書きます。「女子という条件が与えられた上での二年生の確率は」というように読んだりします。

ちょっと状況はややこしくなりましたが, 確率の計算そのものはそれほど難しくありませんね。女子である, というのが決まっているので, 分母が変わるだけなのです。つまり,

$$P(A_2|B_2) = \frac{20}{60} = 0.333...$$

となります。

条件つき確率の計算は分母が変わったと思えばいいのですね。この条件つき確率は, 同時確率を周辺確率で割っても計算できます。すなわち,

$$P(A_2|B_2) = \frac{P(A_2, B_2)}{P(B_2)} = \frac{0.2}{0.6} = 0.333...$$

となるわけです。

今回は性別という条件が与えられた時に, 学年はどうなるかという条件をつけた確率を考えました。性別が一年生の時, 二年生の時, 三年生の時, それぞれで分母が変わりますし, 条件付き確率も変化します。このように, 性別が変わると学年の確率が変わる, という関係にあるので, この性別と学年という 2 つの変数は関係しあっていることがわかります。逆に, 一方の変数が変わっても他方の変数に影響しない場合, 両者は**独立 (independent)** だ, といます。確率において独立であるとは, すべての i, j, k において次の関係が成り立つことを言います。

[*1] たとえば名前とか容易に類推できた, といったようなシーンです。最近では名前だけから性別を判断するのは難しくなりましたが, かつての名残で明らかにわかるような名前だったと思ってください。

$$p(A_i|B_j) = p(A_i|B_k)$$

これまでも (確率的に) 独立である例の話はいろいろしてきました。たとえばコイントスを続けて行うときは，それぞれの試行は独立だよね，という話をしました (→ セクション 23.5，Pp.312)。1 回のコイントスの結果が次のコイントスの結果に影響しない，という意味です[*2]。

ここで独立しているのならば，ここまでの式展開から，

$$\begin{aligned} p(A_i|B_j) &= p(A_i|B_k) \\ \frac{p(A_i,B_j)}{p(B_j)} &= \frac{p(A_i,B_k)}{p(B_k)} \\ p(A_i,B_j)p(B_k) &= p(A_i,B_k)p(B_j) \end{aligned} \tag{25.1}$$

ここですべての k の状態について B_k を数え上げる，つまり B を周辺化します。

$$\begin{aligned} p(A_i,B_j)\sum_{k=1}^{K} p(B_k) &= p(B_j)\sum_{k=1}^{K} p(A_i,B_k) \\ p(A_i,B_j) &= p(B_j)p(A_i) \end{aligned} \tag{25.2}$$

ここからわかるように，A と B が独立である場合は，同時確率と確率の積がおなじになります。独立している事象が起こる確率は，それぞれの確率を掛け合わせましょうということです。これまでも，コイントスやサイコロの例で，2 回続けて計算する場合はその積で表現してきましたよね。確率 θ で表が出るコインで，2 回続けて表が出る場合は $\theta \times \theta = \theta^2$，表・裏と出る場合は $\theta \times (1-\theta)$ といった表現をしてきました。独立した事象は掛け算で表せる，というのはこういう理屈からなんですね。

さて，これで用語の準備は終了です。ここから式の展開が面白くなってきます。

あらためて，条件付き確率は同時確率を周辺確率で割ったもの，でしたね。つまり，

$$P(B_j|A_i) = \frac{P(B_j,A_i)}{P(A_i)}$$

ここで右辺の分母を両辺にかけて，

$$P(A_i)P(B_j|A_i) = P(B_j,A_i)$$

とします。これは A,B が何であるかにかかわらず，一般的に成立します。たとえば $P(A_i|B_j)$ と記号をひっくり返しても，

$$P(A_i|B_j) = \frac{P(A_i,B_j)}{P(B_j)}$$

より

$$P(B_j)P(A_i|B_j) = P(A_i,B_j)$$

としても構いませんね。

ところで，「2 年の男子」と「男子の 2 年」というのは同じ意味です。つまり，

$$P(B_j,A_i) = P(A_i,B_j)$$

[*2] 逆に一回のコイントスが次のコイントスに影響するとはどういうシーンでしょうか？ 無理矢理考えるなら，コインが意思を持って「さっき表になったから，次は裏にしようとする」とか，トスしたコインを受け取り損ねて地面に落下し，片方の面が欠けて他方の面が出やすくなった，というようなシーンです。コインは意思を持ってないし，硬いものですから，これらはやはり考えにくい話ですよね。

ですから，上の式をつなぎ合わせて

$$P(A_i)P(B_j|A_i) = P(B_j)P(A_i|B_j)$$

とできます。ここで両辺を $P(B_j)$ で割って清書すると，

$$P(A_i|B_j) = \frac{P(B_j|A_i)P(A_i)}{P(B_j)}$$

が得られます。やったね！ベイズの定理が導出できました。

さて，これの何がすごいのでしょう？ 数式を「記号の羅列だ！」と思って拒否するのではなく，その意味をしっかり味わってみましょう。

25.2 ベイズの定理の意味

25.2.1 因果の向きについて

さてベイズの定理のすごいところは，左辺と右辺の中にある条件つき確率の中身が**ひっくり返っているところ**にあります。これがひっくり返ると何がうれしいのでしょうか。ベイズの定理の左辺には $P(A|B)$ があり，右辺は分子に $P(B|A)$ を含んでいます。$P(A)$(分子の 2 つ目の項目) や $P(B)$(分母) のことはいったん忘れてくれてもいいですし，

$$P(A|B) = P(B|A) \times \frac{P(A)}{P(B)}$$

と考えて，$P(A|B)$ が $P(B|A)$ の関数である，と読み取ってくれても構いません。

ここで A を「心のモデル」，B を「データ」と考えてみてください。私たちは心の仕組みを考える時に，モデルをたてます。「あの人が A という行動をとったのは，心の状態が X だったからではないか」というのは心というモデルを立てて考えるということなのでした。心の状態が X であれば，人は A という行動をとるであろう，というのが研究者の仮説，モデルなのです。つまり，P(データ|モデル) が心理学でやっていることなのです。

ところが実際には，手に入れることができるのはデータだけです。心がどうなっているかはわかりませんが，人が ◇ ◇ ◇◇ な行動をとった，という事実だけがそこにあるのです。この事実を使って，我々のモデルがどれぐらい正しいのか。あるいはモデルのパラメータはどうなっているのか，を検証したいわけです。つまり，研究の営みとは P(モデル|データ) であり，データを所与 (条件) としたときの，モデルの確率を考えることなのです。データをとってモデルを検証するという営みが成り立つ根拠を，ベイズの定理が示してくれていると言えるかもしれません。大事なのは，不確かな状況においては我々は確率的にしか判断をできず，ベイズの定理では条件の向きをひっくり返して計算できる，という点です。原因が結果を引き起こす確率が知りたい時に，結果を手にすることで，原因がどれほどあり得たかを考えることができる。それがベイズの定理のポイントなのです。ベイズの定理は因果の向きを逆にできる，ということを示しているので，**逆確率の法則**と呼ばれることもあります[*3]。

25.2.2 確率という数字，再考

ここまでベイズの定理の説明は，性別や学年など離散的な数字をもとに説明してきましたが，連続的な確率変数でも成り立ちます。連続的な確率変数の場合，すべての確率を足し合わせるという操作が積分することに変わりますが，それ以外は本質的に変わりません。一般的に書くと，

[*3] ベイズ自身がこの法則を発見した時に，「なんかおかしなことになってるかも」と思って結果を公表しなかったと言われています。神様が決めた世界を裏返しにすることに，背徳感のようなものを感じたのかもしれません。

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{x})}$$

となります。ここで $\boldsymbol{\theta}$ はパラメータで，複数ある場合があるのでベクトルの表記 (太字) にしてあります。$\boldsymbol{x}$ はデータのことで，これも 1 つのデータではないので太字にしています。f は連続的な確率関数だと思ってください。先ほど説明したように，この関数は

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta}) \times \frac{f(\boldsymbol{\theta})}{f(\boldsymbol{x})}$$

と考えても構いません。ここで $f(\boldsymbol{x}|\boldsymbol{\theta})$ は，パラメータが所与の元でのデータの出てくる確率ですから，**尤度**だ，ということがわかります。式の左辺は $f(\boldsymbol{\theta}|\boldsymbol{x})$ ですから，データが所与の時のパラメータの確率ですね。これをとくに**事後確率 (posterior probability)** と言い，それに対して左辺の $f(\boldsymbol{\theta})$ は**事前確率 (prior probability)** といいます。ベイズの式は，確率を確率に変換する公式なのです。

ところでこの**確率**って何なのでしょうね。パラメータについての事前の確率とは，一体どういう意味があるのでしょうか。

さて，確率の公理を思い出してほしいのですが，確率とは次の条件を満たす数字だということでした。

1. 確率の値は非負（正または 0 の数字）
2. 全標本空間におけるすべての事象の確率の合計は 1.0
3. 任意の 2 つの相互に排他的な事象について，どちらか一方が発生する確率は個々の確率の和である

そして，確率には**頻度主義**的な考え方と，**主観確率**という考え方があって，どちらでもこの公理を満たしている限り確率として計算できます，という話をしてきました。

高校数学までは頻度主義的な意味で確率が説明されてきたと思います。すなわち，コインやサイコロを**無限に**投げ続けると，パラメータの値は $1/2$ や $1/6$ に近づいていくだろう，ということです。この無限に繰り返すと 1 つの値に収束するというのは，確率的収束という数学的な証明がされる考え方ですが，現実には無限に行うというのはあり得ません。心理学の実験も，同じ実験を無限に繰り返すと，という設定のもとで考えて帰無仮説検定を行ったりしますが，それがそもそもあり得ない前提じゃないか，と言われたらどうしようもないのです[*4]。無限の回数を考えて，その相対頻度，割合で確率を考えるので**頻度主義**と呼ばれています。

これに対して天気予報などは，信念の確からしさを確率で表現しています。明日晴れる確率が 30% というのは，明日という日が無限に繰り返されればそのうちの 3 割は晴れている，と言うのはやはりおかしいわけです。どんなループに巻き込まれてしまったのでしょうか，明日が無限に繰り返されるなんて！そういう意味では，普段私たちが慣れ親しんでいるのは，自分達の確信の強さ，どの程度その事象を信じているかと言う信念の強さを表している数字と言えます。明日 100% 雨が降る，というのは完全にその状態を信じているということですし，50% の降水確率は雨が降る可能性が半分，降らない可能性が半分だと考えていると言うことですね。また，相対頻度・割合で考えるときは，全体で割った平均的な数字を出すことになります。この頻度主義的な考え方ですと，たとえば野球のシーズン最初の試合，勝ったチームは勝率 100% になります。一試合しかしていないので，$1/1$ だからです。でも誰も「このチームは今シーズン一度も負けないだろう」と本気で信じているわけではないですよね。あるいは最初の打席でヒットが出なかったからといって，打率 0 割，永遠に打てない打者だとは誰も考えません。本当はもう少し低い勝率，高い打率だと思っているのに，そう表現できていないのです。その意味では，「確からしさを表す数字」として確率を使うほうが自然かもしれません。

[*4] さらに帰無仮説というあり得ない仮定を置いてますから，一体どこの世界の話をしているのやら，という気持ちになります。

つまり，確率という数字の見方には大別して二種類あるということです。どちらが正しいとか，どちらが優れているというわけではありません。どちらも上に述べた 3 つの公理を満たすものですから，確率は確率なのです[*5][*6]。

頻度「主義」とか「主観」確率，という言葉の使い方をすると，個人の考え方で数学のルールや数字の意味が変わるのはおかしい，と思う人がいるかもしれません。数学のルールは人類が存在しなくても永遠普遍の真実だ！という考え方があるのもわかります。主義や信念の対立で優劣を決めたいわけではないのですが，頻度主義的な発想とベイズ主義的な発想とでは，母数についての考え方が違っているのも事実です。頻度主義的な発想では，無限に試行を繰り返せばいつかは答えにたどり着くはずで，言い換えれば「真実はいつも 1 つ」なわけです。ただそれが直接わかり得ないから，徐々に近づいていくしかない。図 25.1 の上の段にあるように毎回の試行で信用区間をとって表現できますが，これは本当の値 (図では 0.5 が真の値としています) を跨いでいるかどうか 0/1 の判定が続くわけです。

これに対して主観確率で考えるベイズ法による推定は，そもそも母数の値がわかりませんので，わからないことを確率で表現する。最初は幅の広い，全域に渡るような確率でぼんやり考えるのです。これに毎回の試行＝データを積み重ねることによって，この幅を狭めていき，「この辺にあるんじゃないかなあ」という確信度を高めていくという考え方です。

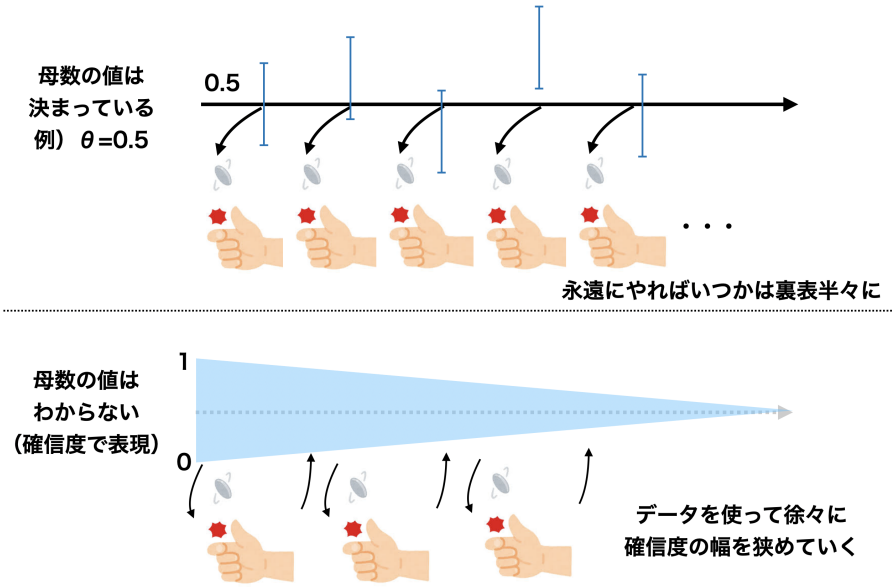


図 25.1 確率についての 2 つの考え方

この後，ベイズ法による推定を説明していきますが，このとき真の値が定数としてただ 1 つ定まっている，と考えられるかどうかについては，とくに定まりがありません。母数は定数で定まっている，ただそれが確率的に

[*5] というか，この公理はさまざまな確率の考え方の共通基盤・最低限のルールをまとめただけなので，必然的にどちらにも合うようになっているというだけです。

[*6] ちなみに，主観確率が成り立つためにはどのような条件が必要かということについては，Gilboa (2009 川越訳 2014) などに詳しく論じられています。また繁桝 (2019) は主観確率という確率が成立する条件として，1. すべての不確かな事象の蓋然性は相互に比較できる，2. 蓋然性に関する判断は相互に矛盾しない，3. まったく可能性がない状態を 0，確実に起こることを 1 として，0 から 1 までのすべての数値を取りうる不確定な状態を作る実験が存在する，の 3 つを挙げています。人間が必ずこのルールに則って判断している，という話ではなく合理的な判断があるとするならば，この規則に則っていると考えるべきであるという規範理論であり，「そんな状況はあり得ないからだめだ」というのは反論として的外れなものです。経済学では合理的に価値判断をする人間を仮定して理論を組み立てており，この不確実状況下での意思決定論もそれと同じ系譜，かつ，ベイズ統計学の応用を許す理論的基礎になっているといえるでしょう。

しかわからないだけだという考え方をするのか，母数がそもそも確率的に分布していると考えるのか，いずれの考えであってもいいでしょう[*7][*8]。

現代的に，あるいは統計のユーザとしてはプラグマティックに「使い物になるほうが正しい考え方だ」という割り切り方をしても良いでしょう。いずれにせよ，これらの手法によって表現されるものの意味の違いについて，少し注意しておいていただければと思います。

25.3 ベイズの定理を使った推論の例

さて，ベイズの定理を使うとどういうことがわかるのでしょうか。事前の確率が事後の確率に変換される，つまり推論情報のアップデートができることの数学的な意味はわかったとして，実際どのようなシーンで活用できるのでしょうか。ここでは 2 つの例を見てみましょう。

25.3.1 事前情報を活用しよう

事前確率があることで何がわかるかについて，疫病の診断を例に考えてみたいと思います。

コロナウイルスに感染しているかどうかは，PCR 検査で明らかになりますが，これらの検査の結果というのも確率的な要素があるのはご存知でしょうか。本当は感染していないのに，検査では陽性になってしまう，ということがゼロではありません。こういうのを**偽陽性 (False Positive)** といいます。逆に感染しているのに，検査で検出できずに陰性になる，**偽陰性 (False Negative)** ということもあります。いずれも確率的な話です[*9]。

コロナに感染している → 検査で陽性になる，というのが因果関係ですが，我々は実際には検査で陽性である → コロナに感染しているかもしれない，というように推論します。これを今日学んだ確率の言葉で書いてみましょう。

感染している状態で，陽性反応が出るというのは条件つき確率として表現できますね。$P(\text{陽性}|\text{感染})$，というわけです。さてここに，そもそもの感染率というのが別途見積もれたとします。$P(\text{感染})$ ですね。さらにもうひとつ。感染しているかどうかに関係なく，陽性であるという結果だけ取り上げて $P(\text{陽性})$ を考えることもできます。これが揃うと，先ほどの定理を使って

$$\frac{P(\text{陽性}|\text{感染})P(\text{感染})}{P(\text{陽性})} = P(\text{感染}|\text{陽性})$$

を計算できます。つまり陽性になったという結果から，本当にかかっているかどうかの確率を計算できるの

[*7] このような考え方の違いは，量子力学のコペンハーゲン解釈，ボーアとアインシュタインの哲学の違いに似たところがあります。アインシュタインは「神はサイコロをふらない」といい，物理現象は確たるものがあると信じていました。これに対してボーア率いるコペンハーゲン学派の解釈は，物理現象は究極的に確率的に存在していると言い，シュレディンガーの猫も生と死の重ね合わせた状態である (観測するまでわからない) と言い切るわけです。現代物理学としては量子力学が優勢ですが，アインシュタインの気持ちもわかりますね。

[*8] ベイズ統計の中にもさまざまな流派があり，Twitter では時折喧嘩になっています。争点はこの「主観確率」を認めるかどうかというところで，「人間はすべての情報を手に入れることはできないので，自分の知り得た情報の範囲の中で最適な判断をするしかない。」と考える主観ベイズ派と，「手に入れた情報はなんらかの確率分布から生成されたサンプルに過ぎず，確率的に変動するサンプルから真の分布の状態を推論するのだ」という情報統計学的立場があります。後者は日本の数理統計学者 (赤池弘次，渡辺澄夫) が展開している議論で，AIC や WAIC など有用な指標を提案していることもあって，日本ではよく知られています。しかし主観ベイズ派は，古くからいかにして主観的確率判断が成り立つかについて緻密な理論を打ち立てており，世界的にはこちらの方が多数派かもしれません。

[*9] 勘のいい人はお気づきかもしれませんが，これはセクション 19.3.1 で学んだ検定における 2 種類の間違い (→ セクション 19.3.1，Pp.246) と同じ話であり，偽陽性は**タイプ 1 エラー**，偽陰性は**タイプ 2 エラー**のことです。有意水準は偽陽性率 (本当は差がないのに差があると言ってしまう，本当は感染していないのに感染していると言ってしまう，せっかちなミス) として導入されていたのでした。

です。

多くの場合，検査で陽性だったら「もうだめだー」と思ってしまいがちです。これは陽性であると確実に検出できる，偽陽性がゼロの検査であれば正しい判断です。が，実際には確率的に判断すべき問題です。具体的に数字を例にして考えてみましょう。ここでは乳がん検査 (マンモグラフィー) の数値例でお話しします[*10]。

- 40 歳の女性が乳がんにかかる確率は 1% です。
- 乳がんである人が乳房 X 線検査で陽性と出る確率は 90% です。
- 本当は乳がんでなくても，検査で陽性と出る確率は 8% にすぎません。
- あなたが 40 歳の女性で検査結果が陽性だったとしたら，あなたが乳がんである確率はどれぐらいでしょうか。

これは先ほどの公式に当てはめて考えると，

$$\frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.08 \times 0.99} = 0.102$$

となり，10% ちょっとだということになります[*11]。ずいぶんと印象が変わりますよね。PCR 検査を複数回行うのは，これと同じような背景があるからです。

このように，事前確率があると，そこから遡って「診断された人が病気である確率」の計算をできるのです。逆確率，という言葉の意味が少しおわかりいただけたでしょうか。

25.3.2 モンティ・ホール問題

次は**モンティ・ホール問題**という，有名な事例を取り上げます。モンティ・ホールというのは人名で，この人がテレビショー「Let's make a deal」の司会をしていたのですが，その中で行われたゲームに由来するお話しです。問題はこんな内容です。

> プレーヤーの前に閉じた 3 つのドアがあって、1 つのドアの後ろには景品の新車が、2 つのドアの後ろには、はずれを意味するヤギがいる。プレーヤーは新車のドアを当てると新車がもらえる。プレーヤーが 1 つのドアを選択した後、司会のモンティが残りのドアのうちヤギがいるドアを開けてヤギを見せる。
>
> ここでプレーヤーは、最初に選んだドアを、残っている開けられていないドアに変更してもよいと言われる。ここでプレーヤーはドアを変更すべきだろうか？

直感的には，選択したドアを変えても変えなくても，新車が得られる確率は 1/3 で変わらない，と思いますよね。でも実際は，変えた方が当たる確率は 2 倍になるのです。

なんでそんなことになるのか，と思うかもしれません。変えても変えなくても，確率は変わらないじゃないか，と思うのではないでしょうか。

実は明文化されていませんが，この話のポイントは，司会者のモンティは答えを知っていて，テレビ番組を盛り上げるために，必ずヤギ (ハズレ) のドアを開けるということです。司会者が「じゃあこっちを開けてみましょ

[*10] 乳がんについては古くからの疫学的知見で，大体の罹患率が事前情報としてあるからです。コロナについては 2021 年現在でも日々状況が変化するので，事前確率も確定していませんので･･･。

[*11] おっと，計算についていけないぞ！と思った人。大丈夫です，いい方法があります。確率の計算は小数点を含みますし相対的な度数なのでわかりにくいですから，絶対的な度数に戻して考えればいいのです。たとえば 10000 人の世界だとして考えると，乳がんの人は 1%=100 人です。乳がんでない人は 9900 人です。陽性になる人は乳がんで陽性反応が出るひと ($100 \times 0.9 = 90$) と乳がんじゃないのに陽性反応が出る人 ($9900 \times 0.08 = 792$) の和で，882 人です。陽性反応が出た人の中で本当に乳がんであるひとは，$90/882 = 0.102$ となります。

う，わー新車だ。ザンネーン」といったらシラけてしまうからですね。司会者がハズレのドアを開けるということは，新しい情報が追加されているということでもあります。であれば私たちは，情報をアップデートできるはずであり，確信度は変わるはずです。

3 つのドアを A,B,C として，どこかに当たりが入っているのですが，確信度はどれも同じぐらいなので，それぞれ 1/3 の確率だと考えます。あなたが A のドアを選んだとします。当たっている確率は 1/3 です。B か C が当たっている確率は，A ではないので，2/3 になります。モンティはどのどあをあけるでしょうか？ A のドアは番組がしらけてしまうので開けません。B か C を開けるしかないのです。ここでもし，ドア B の後ろに新車があるとモンティはこれを開けられないので，ドア C が開けられることになります。そうすると当然，ヤギが出てくることになりますね。さて，これまで B か C があたりの確率は 2/3 でした。でも C があたりということはあり得ないので，この 2/3 の確率は B=2/3,C=0 にアップデートされるはずです。あなたが最初に選んだ A のドアは，1/3 の確率だったわけですから，「A から B に変えますか？」と言われると，変えた方が 2 倍得だ (A=1/3,B=2/3,C=0) ということになります。

これはたとえば，3 つのドアでなくて 100 枚のドアがある場合で考えてみると，わかりやすいかもしれません。100 枚のドアがあって，あなたが 1 番目のドアを選び，モンティが 100，99，98，...，5，4，3 番目のドア，と順番にドアを開け続けて，すべてにヤギが入っていたとします。それでも 2 番目のドアに変えないか？といわれると･･･モンティはどうして残り 99 枚のどれを開けてもよかったのに，隣のドアを開けなかったのか？と思いますよね。

モンティ・ホール問題は他にもいろいろな説明の仕方があります。納得いかなかった人は，Wikipedia などで調べてみると良いでしょう。

くどいようですが，ポイントは「情報がもたらされたときに，確率はアップデートできる」というところにあります。事前確率ともたらされる情報をつかって，確信度をどんどんアップデートしていく，これがベイズ法の考え方なのです。次回は具体的に，ベイズ法をつかった推論について考えていきましょう。

25.4 課題

■事後確率の計算 ある疾病は人口の 0.05% が罹患していることが分かっています。この病気の検査をしたとき，罹患している人が陽性になる確率は 70% です。また，罹患していない人が陽性となる確率は，0.1% です。このとき，この検査で陽性と判定された人が，この病気に罹患している確率を求めなさい。

第 26 章

ベイズ推測の実際

前回，ベイズの定理を導入しました。今回はこの定理を，これまで行ってきた推測統計学の枠組みでどのように利用するかを考えていきます。

26.1 事前分布と事後分布

さっそく**ベイズ法**で，実際にパラメータを推定する例を考えてみましょう。

今回は例として，学力テストを取り上げます。難易度が同じぐらいの 10 問のテストを受けたとします。推定したいのは回答者の学力です。この能力は，問題に正しく答える比率 θ である，と定義します。学力 θ は直接観測できず，観測できるのは「正答数」だけです。今回ある回答者が 4 問正解したとして，学力はどうやって見積もれば良いでしょうか？ というのがここでの課題になります。

ベイズ法による考え方は，基本的にたった 2 つのステップからなります[*1]。

1. 不確実性，信念の強さを確率によって数量化する (**事前分布 (prior distribution)** の設定)
2. 観測データを使って事前の情報/信念を更新し，事後の情報/信念 (**事後分布 (posterior distribution)**) にアップデートする

これだけです。

ではまずステップ 1, **事前分布 (prior distibution)** について考えてみましょう。

26.1.1 事前の信念の表し方

事前分布は，確率のパラメータについて事前に与える分布です。パラメータがどれぐらいになるのか知りたいのに，事前に何かわかっていることがあるっておかしくない？ と思う人もいるかもしれません。ここで，確率のベイズ的考え方，すなわち確率とは確信の程度を表現しているのだと思って思ってください。不確実で，わからないものを，どの程度わからないのか確率で表現するのです。推論とは確率の再分配であり，「他に可能性がない」のであれば 100% 確からしい，と言えるのでした。逆に「事前にパラメータがどのあたりにあるのかなんて，まったくわからない」という場合も，確率で表現できます。この表現方法として，「どこ」にあるのかわからないのであれば「全部あり得る」と広く広く守備範囲を広げておけば良いことになりますね。確率なので面積は 1.0 でなければなりませんが，その分うすく伸ばすように考えれば良いのです。

図 26.1 に，4 つの確率分布を示しました。これらはベルヌーイ分布のパラメータ θ についての，個人の考え方を示していると思ってください。ベルヌーイ分布はコイントスのような，二値の結果を生じさせる試

[*1] 帰無仮説検定より少ない！やったね！

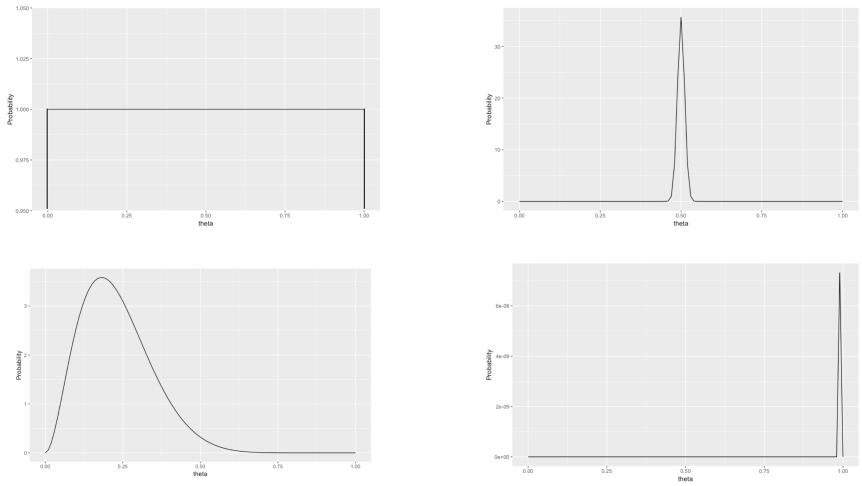


図 26.1 コイントスの出目についての四つの信念

行についての確率分布ですから，確率 θ は $0 \leq \theta \leq 1$ の範囲にあるはずです。図 26.1 の左上は範囲すべてにわたってフラットです。つまり，「どこ」にあるとも断言できない，どこにも同じ程度にあるはずだという信念で，言い方を変えると何もはっきりとは言えない，つまりわからない状態を表しています。データを分析するにあたって，事前に何もわからないのであれば，このようにフラットな事前分布 (とくに**一様分布 (uniform distribution)** といいます) を置けばよく，こういう事前分布の置き方をとくに**無情報事前分布 (uninformative prior distribution)** と呼びます。図 26.1 の右上は逆に，$\theta = 0.5$ のあたりにあるはずだ，と思っている人の考え方です。少なくとも 0.45，大きくとも 0.55 ぐらいでしょう，ということなので，かなり確信度の高い (確率分布の幅が狭い) 状態にあると言えます。そこまではっきりわからないけど，でもちょっと裏が出やすそうだなあ，というふんわりと低めの予想をしている人は，図 26.1 の左下のような信念をもっている，と表現できますし，「絶対に表が出る。面しか出ない」と強く信じている人は図 26.1 の右下のようになります。

このように，まったくわからないことや，どこかに特定の強い思い入れがあること，あるいはなんとなくこの辺じゃないかと思っていること，など事前の情報を確率分布として表現することは可能なのです。研究の実践場面において，事前に何かわかっていることが「まったくない」のであれば，一様分布のような無情報事前分布，あるいはすごく幅の広い弱情報事前分布を置けば良いでしょう。身長や体重などのデータを分析する時に，正規分布を仮定しますが，このような場合は少なくとも負の値をとることはありませんし，いくら背が高い，体重が重いとしても 10m や 5t になったりするはずはないので，$0 \leq \mu \leq 100000$ のような幅で事前分布をおいても構いません。あるいは先行研究などから，ある程度のアタリがつけられるのであれば，その値を中心にふんわりした事前分布を置いたっていいのです。逆にいうと，このように事前分布をおくことで，先行研究との一貫性を持たせたり，先行研究の情報を活用して知見を積み重ねていくことができる，とも言えます[*2]。

さて今回の例に戻りましょう。この受験生の学力がどれぐらいなのか，事前にはまったくわかっていません。しかし正答率を考えるのですから，推定値は $0 \leq \theta \leq 1$ の範囲にあるはず。まったくわかっていないのですから，この範囲の一様分布を設定しましょう。つまり，とくになんの偏見も持たずにデータに立ち向かうという

[*2] 分散分析や最尤法を考え出した R.A.Fisher 先生は，このベイズの考え方がお気に召さなかったようです。研究を始める前になんらかの思い込みを持って取り組むというのはおかしい，データに対してはなんら心構えを持たずに立ち向かうべきだ，という考え方です。Fisher 先生のベイズ統計学についての当たりの強さは，McGrayne (2011 冨永訳 2018) などにも記されています。

精神を，無情報事前分布で表現するわけです。

26.1.2 事後分布の計算

次にステップ 2 ですね。アップデートです。アップデートにはベイズの定理を使います。ベイズの定理を改めて見直しておきましょう。次のような式なのでしたね。

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

ここで $P()$ で表されているのは確率，とだけ言っていましたが，これは離散でも連続でも同じことですので，確率分布と呼びましょう。とくに左辺は「結果的にこうなる」という意味ですから，**事後分布 (posterior distribution)** と呼ばれます。

この確率的関係 (右辺も左辺も全部 P で囲われています！) を，実際に統計的推論の場面で用いることを考えます。左辺にある事後分布は，データをとったうえで，パラメータがどのあたりにあるのかを推論するということを条件付き確率で表しています。つまり，B がデータで A がパラメータ，事後分布 $P(A|B)$ が「データをとった上でのパラメータの確率」ということです。であれば，右辺にある $P(B|A)$ は「パラメータがあった時にデータが出てくる確率」を意味することになります。これってなんでしたっけ？ そう，**尤度 (likelihood)** ですね！

さて記号の意味を考えていきますと，$P(A)$ はデータを取る前のパラメータの確率，ということになります。これが先ほど説明した**事前分布**ですね。

また $P(B)$ はデータの確率，ということになります。データの確率？ データが得られるかどうかの確率？というのはわかるはずがないような気もしますが，前回の乳がんの罹患率の例を思い出すと「あり得るパラメータから考えられるすべてのデータが出てくる確率をかき集めたもの」，というようなイメージができるかもしれません。用語としては**周辺尤度 (marginal likelihood)** といいますが，実践上はこれについて深く考えなくてもよいものであることを後で説明します。のでこれについても話はここまで[*3]。

今，言葉で表現したものでベイズの定理を書き変えてみます。

$$\text{事後分布} = \text{尤度} \times \frac{\text{事前分布}}{\text{周辺尤度}}$$

ここにあるように，尤度に何かをかけた結果，左辺が分布になっている，ということに注目してください。前回，尤度は確率ではないという話をしました。尤度関数の面積 (積分) が 1.0 にならないので，確率にならないのです。ですがこの式は，その尤度に事前分布をかけて，周辺尤度で割って，事後分布を得ています。事後分布は確率分布です。つまり，尤度を確率にしてくれるのがベイズの定理だ，とも言えるのです。

さて，ベイズの定理は事後分布$=\frac{\text{尤度} \times \text{事前分布}}{\text{周辺尤度}}$ で，今回の尤度は 4 問正解した，つまり (順番は分かりませんが) 正答, 正答, 正答, 正答, 誤答, 誤答, 誤答, 誤答, 誤答, 誤答，というような状況です。ベルヌーイ分布ですから，尤度は $\theta^4(1-\theta)^6$ ということになります。

分子が確率と確率の掛け算になっているので，イメージしにくいかもしれません。私個人のイメージでは，波の衝突のような感じだと思っています。X 軸に沿って細かく数字を切り出して，逐一掛け合わせていくような方法のイメージでも良いかもしれません (図 26.3)。ちなみに事前分布が今回，無情報な一様分布ですので，尤度と組み合わせたところ，その高さは変わりますが全体的な形状は変化しません。このように，事前分布を一様分布にしておくと，データと確率分布の関係，尤度の形をそのまま結果に反映させることができるこ

[*3] 本当の本当は，この周辺尤度というのがデータの適合度などを考える時にとっても重要なものになってくるのですが，心理統計のライトユーザにとっては t 分布の導出のように「今ここでやらなくてもいい」ものです。興味がある人は浜田他 (2019) を読んでください。

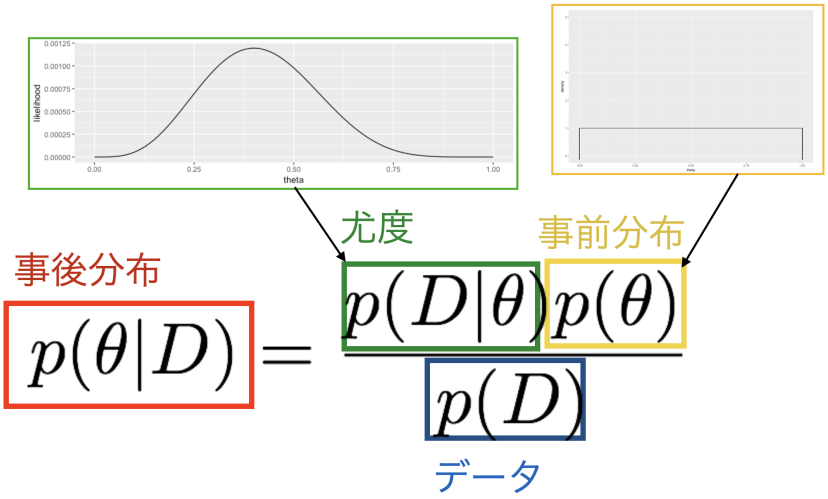


図 26.2 ベイズ推定の実際

とになります。さて，この掛け合わせた結果を $P(D)$，つまりデータの確率で割らないといけないのですが，こ

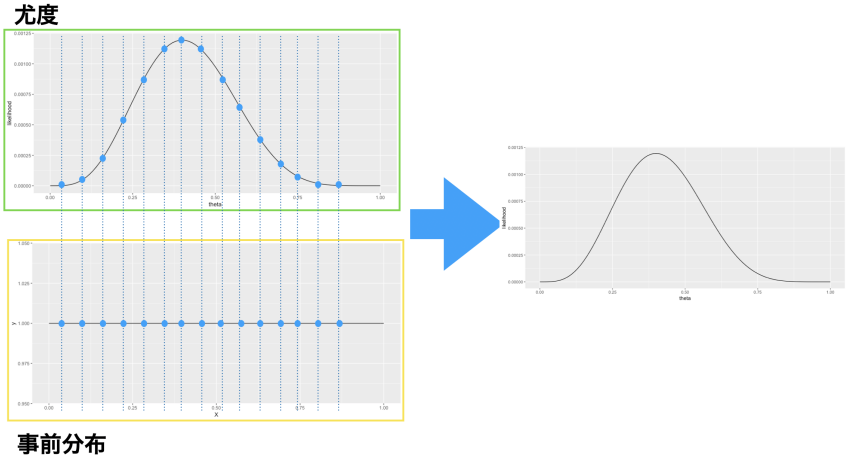


図 26.3 確率同士の掛け合わせ

こは無視してしまいましょう。というのも，この分母が何をやっているかを考えれば，パラメータのような変わる数 (変数) が含まれていないのでここは「定数」になっていることがわかります。分子のところは尤度と事前分布を掛け合わせたものですから，数字はすごく小さいのものになっているのでしょうが，最終的に左辺は事後分布という名の確率分布です。つまり，左辺は面積が 1.0 に規格化されており，分母は面積を 1.0 にするための，割り算用の定数に過ぎないということがわかります。我々はパラメータがどのあたりにあるのかが知りたいのですが，「ここにありそう」というのはそこの確率が高そうだ，ということであって，形さえわかれば縦軸の単位は別に詳しくわからなくても，横軸の位置を推測するのに苦労はない訳です。そういうわけで，分母は

無視しちゃいましょう[*4]。この「形がわかればいい」というのを数学的に表すと，

$$P(\text{事後分布}) \propto \text{尤度} \times \text{事前分布}$$

となります。$\propto$ というのは，相似形である，という意味です。

さて，図 26.4 にあるように，事後分布が計算できました。これが θ の推定結果ということになります。

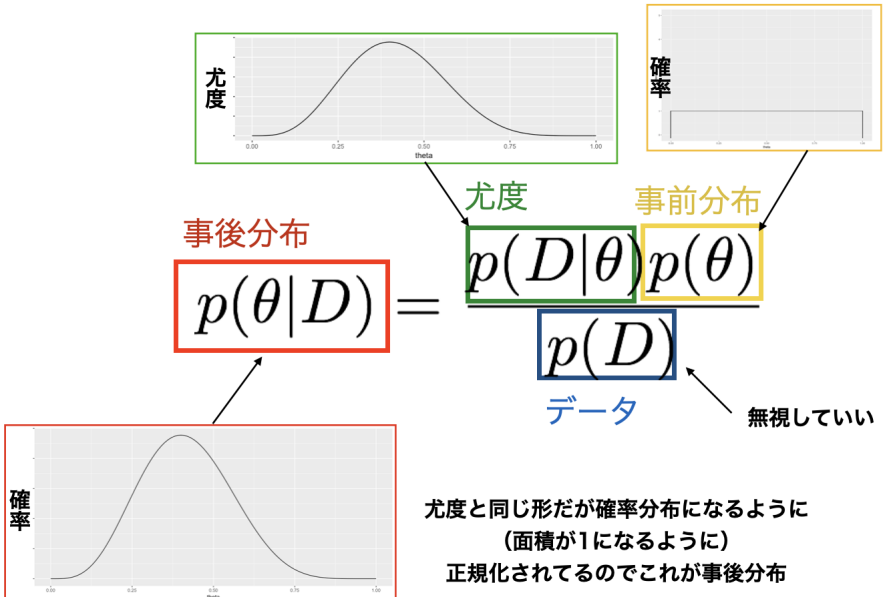


図 26.4 事後分布の計算

これで事後分布が計算できました。

26.2 事後分布を代表する数字

このように計算するのがベイズ法による推定方法です。推定結果である事後分布は図 26.4 の右下のような形になります。これが結果です，と言われても，ピンと来ないかもしれません。ベイズ法は推定結果も事後分布＝確率分布として得られます。モーメント法や最尤法は，θ の値が XXX です！と答えを出してはくれるのですが，ベイズ推定の場合は答えが「分布」で出ていますから，1 つの値にして回答を出すことができないのです。この分布全部が答えであり，事前分布の解釈のように「この辺りにありそうだ」という信念の強さを表しているとも言えます。

分布全体を見ながら「ふむふむ」でもいいのですが，分布の代表値を結果の報告に利用できます。たとえばこの分布関数のピーク，記述統計でいうところの最頻値を考えると，これは最も確率が高いところなので，最も尤もらしい，つまり最尤推定値と考えることができます。ちなみにこの関数の最尤推定値は**最大事後確率 (Maximum A Posterior)** ということで **MAP 推定値 (MAP estimation)** と言われます。また，確率関数の期待値を代表値とすることもできて，それは Expectation A Posterior ということで **EAP 推定値 (Expectation A Posterior estimation)**，と呼ばれます。またこの関数は左に歪んでいますから，平均値である EAP はあまり代表値として相応しくないため，中央値をつかって MEDian a posterior とい

[*4] 繰り返しになりますが，実は深くベイズ推論を進めていくとこの分母というのが結構おもしろいもので，このモデルの中でデータがどれぐらい出てきそうなのか，というモデルとデータの適合度を測る目安に使えちゃうのです。がそれはまたの講釈で。

うことで **MED 推定値 (MEDian a posterior estimation)** を使ったほうがいいかもしれません。ちなみに今回のこの関数のピーク，MAP 推定値は $\theta = 0.4$ です。0.4 になる確率が最も高いのですね。

図 26.6 に分布を要約する数字を描いてみました。

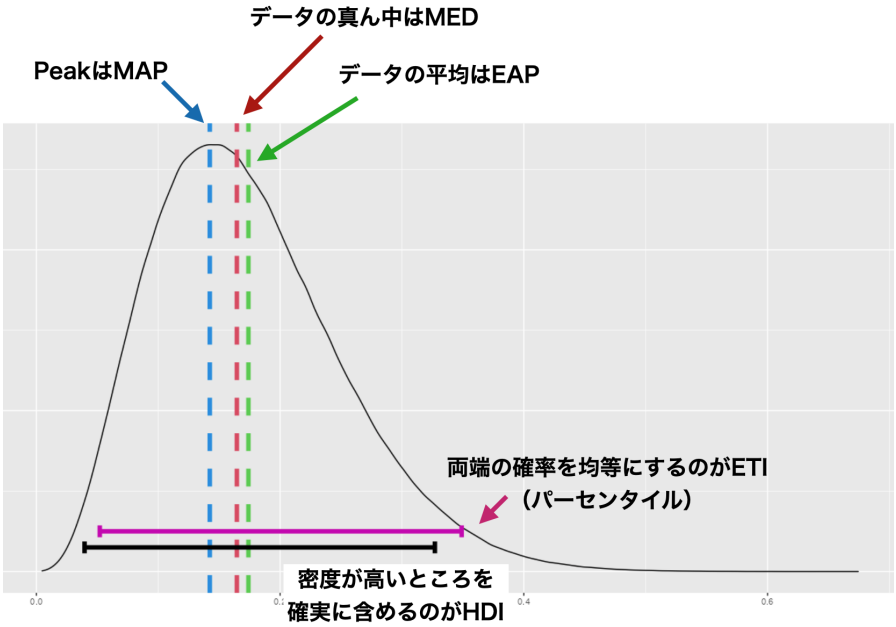


図 26.5　得られた事後分布を代表する数字

また，分布の要約ですから，幅を表す指標を使って表現することもできます。幅を表す統計量としては，**四分位 (quantile)** や**パーセンタイル (percentile)** などがあります。四分位はデータを上から 25%, 50%, 75% のところで区切って四分割し，25% や 75% に当たる数値で報告するものでした。なお第 2 クォンタイル，すなわち 50% 点は中央値と同じです。分割の場所はどこにおいても構わないので，任意のパーセント点で区切って報告するのがパーセンタイルです。33% だろうが 68% だろうが，なんでもいいのですが，下から 2.5%，97.5% の点を報告することで全体の 95% が含まれるので，この 95% 区間を報告することも少なくありません。こうした事後分布の区間をとくに，**確信区間 (Credible Interval,C.I.)** といいます。

ちなみに，今回の事後分布のように，分布の形は正規分布のような左右対称であるとは限りません。左右それぞれに 2.5% ずつの裾を取る確信区間は，**等裾区間 (Equal-tailed interval;ETI)** と呼びます。これに対して，分布の最も信頼できる部分，かつ分布の大部分をカバーし，区間内部のすべての点が，区間外の任意の点よりも高い密度を持っているように推定するものを**最高密度区間 (Highest-Density Interval;HDI)** と呼びます。HDI の方が計算が面倒ですので，一般的には ETI の方が使われることが多いようです。

図 26.6 に ETI と HDI の違いを説明する図を作りました。やや左に偏っている分布を作ったのですが，この分布の 75%ETI は $[0.0886, 0.266]$ で，HDI は $[0.075, 0.242]$ です。ETI は左右の残りの面積を均等にした幅ですから，左側の面積が少し多めに取られています。一方 HDI は「その区間に入っている密度は他よりも高い」区間ですので，比較的均等な幅が得られているようです。数値的には微妙な違いですが，考え方の違いがあることに注意してください[*5]。

これらの数字を使って，事後分布を要約しながら考えていくことになります。今回の MAP 推定値は 0.4 で

[*5] この区間を描くために，R で `bayestestR` パッケージ (Makowski et al., 2019) の `eti`,`hdi` 関数を使いました。

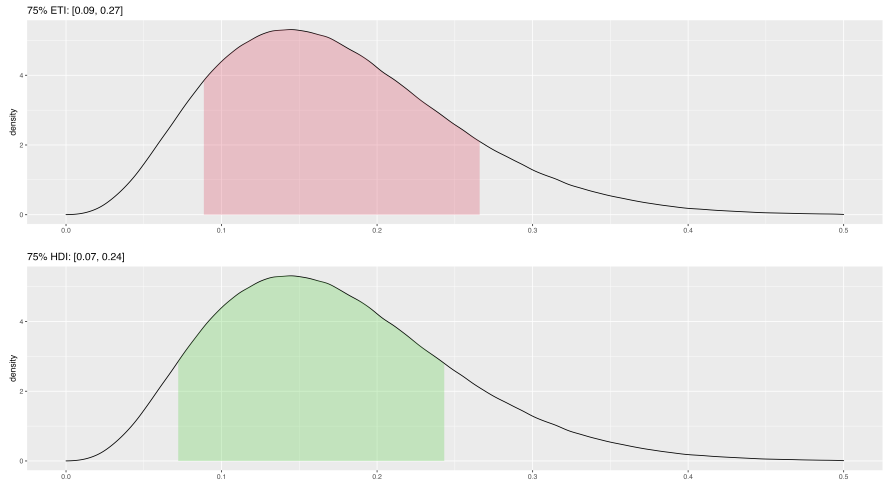


図 26.6 ETI と HDI の違い

したから，事後確率 0.4 である確率が最も高い，つまりこの人の推定値は 0.4 だ，とすれば良いでしょう。

・・・なんだそれ，と思ったかもしれません。10 問中 4 問正解だから 0.4 なのは当たり前じゃないか，ということだと思うのですが，繰り返しになりますがそれは標本平均値を使ったモーメント法の答え。尤度関数と事後分布の形が同じなので，最尤推定の結果も 0.4 になります。が最尤推定の場合は 0.4 になる「確率が最も高い」とは言えません。尤度関数は確率ではないからで，0.4 になるのが最も尤もらしい，というに止まります。ベイズ推定の場合は，(事後分布が)0.4 になる確率が最も高いといっても構いません。大事なのは，どの方法をとっても同じような答えが出るということです。同じデータで，同じ母数を推定しているのですから，筋道は違えど答えは同じということです。それよりも，各推定法の道のりの違い，表現の違い，どういうことが言えるかといった違いをしっかり踏まえておいてください。

ベイズ法の場合は確率での答えになっていますから，表現しやすい利点があります。たとえば MAP 推定値は 0.4 ですが，図 26.5 にあるように，「θ の値が 0.306 から 0.499 の間に入っている確率が 50%」だとか，「95% の確率で θ が 0.159 から 0.682 の間に分布している」という表現ができます。この区間推定の方法は，モーメント法による信用区間のそれとは違うことに注意してください。モーメント法による推定は，標準誤差を使って推定し，「この区間の中にある」と断言して，この宣言が的中する確率が 95% という意味です。実際には，あたっているかハズレているかの二択であり，基本的に点推定なのであるのに対し，こちらは基本的に区間推定で，逆に言えば「この区間の中にある」としか言えない，あるいは「この区間の中にある確信度が 95%」というような意味になります。頻度主義的な**信用区間 (confidence interval)** に対し，ベイズ法によるこの区間は**確信区間 (credible interval)** と呼んで区別します[*6]。

26.3 3 つの推定法

これで**モーメント法**，**最尤法**，**ベイズ法**という 3 つの推定方法が出そろったことになります。それぞれの特徴について，今一度確認しておきましょう。

最初に説明したのは，標本平均をつかって母平均の推定値とする，不偏分散をつかって母分散の推定値と

[*6] といっても，英語でいうといずれも C.I. と略されるのです。ややこしいですが，これはおそらく後発のベイズ推定が表現を真似した，いたずら心による洒落ではないかと思います。微妙に違いはしますが，CI95 といっておけばどっちとも取れるので便利っちゃあ便利・・・なのかな。

する，という，記述統計量に密着した推定法，モーメント法でした。母集団に正規分布を仮定し，標本統計量の分布を使って推定精度を表現するといった方法です (図 26.7)。これはそもそも，標本も繰り返せばいずれは 1 つの値に収束していくという，無限試行の極限として確率を定義する方法の延長線上にある考え方です。古くから積み重ねられてきた手法であり，また手続きのパターンが決まっているので統計パッケージに採

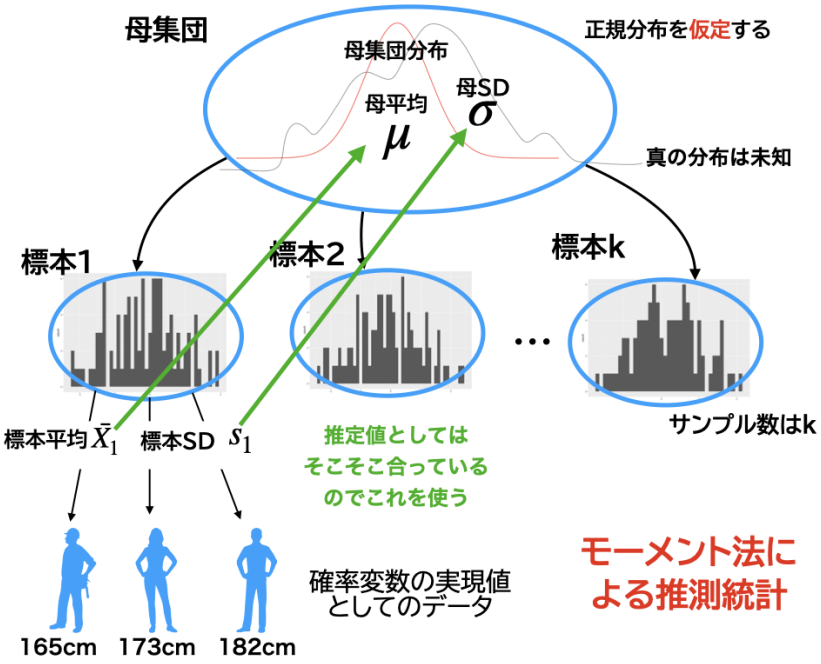


図 26.7 モーメント法による母数の推定

用されやすく，広く普及した方法です。この手法と相性の良かった帰無仮説検定は，心理学においても他の分野においても，長らく使われてきています。ただし手続きがパターン化，悪く言えばマニュアル化し，「これさえやっておけば良いだろう」という使われ方が，マニュアルの誤読や誤解，ルール違反を生み出すことにもなりました[*7]。マニュアル自体もアップデートされていっており，今ではよっぽどのことがない限り検定を使わないでおこう，という主張もあるぐらいです[*8]。皆さんも十分注意深く扱って欲しいと思います。

さてこれに対して，確率分布関数をつかってそれに合うパラメータを探すという，最尤法という考え方もあります (図 26.8)。これは分散分析の生みの親，R.A. フィッシャー先生が考えた方法で，データが正規分布する母集団から生成されている，という情報をヒントに，データが与えられた時のパラメータがありそうな場所を探す方法，手元のデータが最も出てきやすいパラメータは何かを考える，という方法でした。これは母集団 – 標本の考え方からいっても非常に素直な発想です。いわば，確率分布をデータ生成機と考え，このデータを生み出す機械の設定 (=パラメータ) はどうなっているんだろう，と考えるというわけです。手元のデータが最も再現されやすいところがその設定値だろうと考える，というわけですね。

この最尤法による推定も長く広く使われています。計算にもそれほど時間がかかるわけでもないですから，多くのデータを扱った推定法としてはスタンダードな手法です。この方法が確率モデルの基本中の基本，王道といってもいいかもしれません。ただ，尤度という考え方の使い方をちょっと注意しないといけないことと[*9]，

[*7] これはこの統計手法が悪いというのではなく，何でもルール化してしまうと同時にハックする方法もできてしまうということですので，研究に対する態度の問題です。

[*8] 帰無仮説検定が悪いとか間違っているというのではなくて，厳密な手続きを持って行わないと正しく判断できないことが，探索的な研究領域には向いていないという意味です。

[*9] 尤度は確率ではありません。また「データに最も合うように確率分布のパラメータを合わせた」のであって，そもそもデータにあて

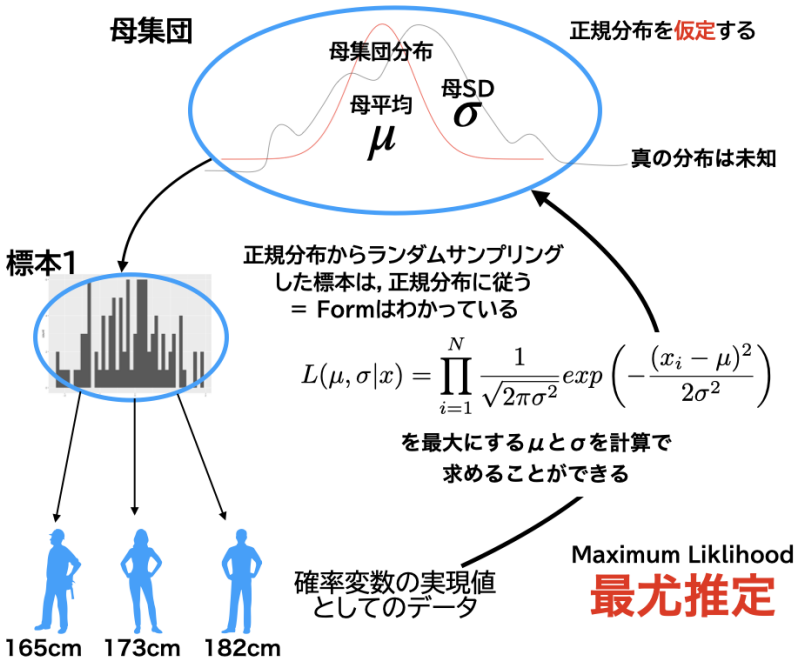


図 26.8 最尤法による母数の推定

モデルが複雑に入り組んでくると，昨今の計算機でも答えが出せなくなるような計算機上の限界が見えてきた，というところではあります。

第 3 の手法はベイズ法です (図 26.9)。このベイズ法の原理そのものは 18 世紀に考えられていたものですが，計算機科学の発展によって実用可能なものになり，一気にその利用が広がっています。この方法を使うと，最尤法でうまく答えが出せなかったような複雑なモデルであっても，問題なく答えが出せます。最尤法の場合はたくさんのデータがないと安定した答えが出ない，というようなことがあるのですが，ベイズ法の場合は少ないデータであっても答えを出してくれます。ベイジアン[*10]にとっての限界は，その人の想像力だけだ，という人もいるぐらいです (リー・ワゲンメーカーズ, 2013 井関訳 2017)。

ではどうして少ないデータであっても答えが出せたり，複雑なモデルであっても答えが出せたりするのでしょうか？ 技術的には乱数発生による近似法 (MCMC 法) が発達したこと，というのが答えになりますが，もう 1 つの答えは最尤法にはない「答えを出すためのヒント」，事前分布があるからでしょう。

それでは最後に，事前分布が事後分布にどのように影響を与えるのか見ておきましょう。

26.4 事前分布と尤度の折衷

ベイズ推定の (たった 2 つの) ステップを思い出してください。

1. 不確実性，信念の強さを確率によって数量化する (事前分布の設定)
2. 観測データを使って事前の情報/信念を更新し，事後の情報/信念 (事後分布) にアップデートする

まずは事前分布，それを事後分布に変える。これだけですね。データを得て事後分布にアップデートできれ

がう確率分布が間違っている可能性もあります。

[*10] ベイズ統計学をする人，ベイズ流の，というのをベイジアンといいます。ちなみにコンピュータ言語 Python を使う人はパイソニアン，版組みソフト LaTeX を使う人はテフニシャンなどといった呼称があります。なぜか R を使う人は「R おじさん」と言われています。

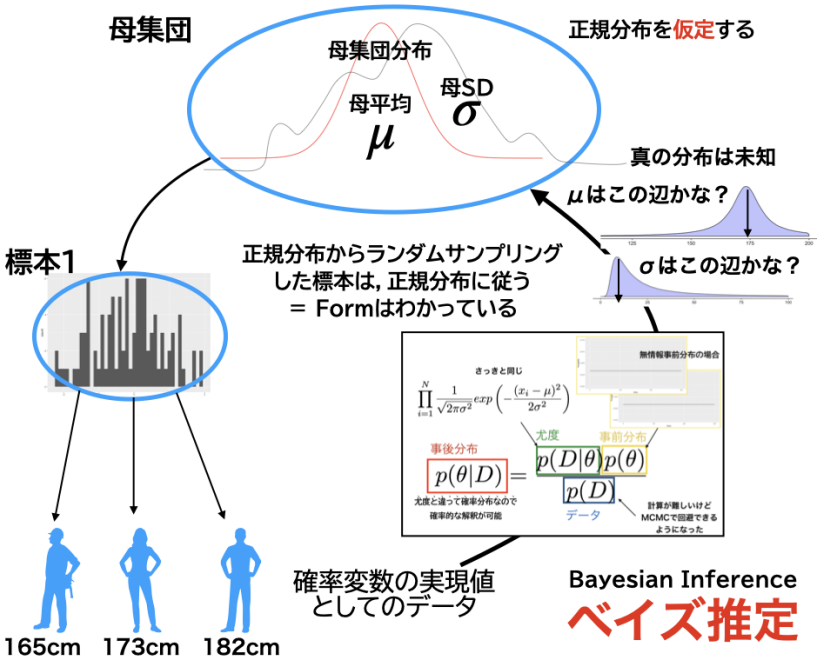


図 26.9 ベイズ法による母数の推定

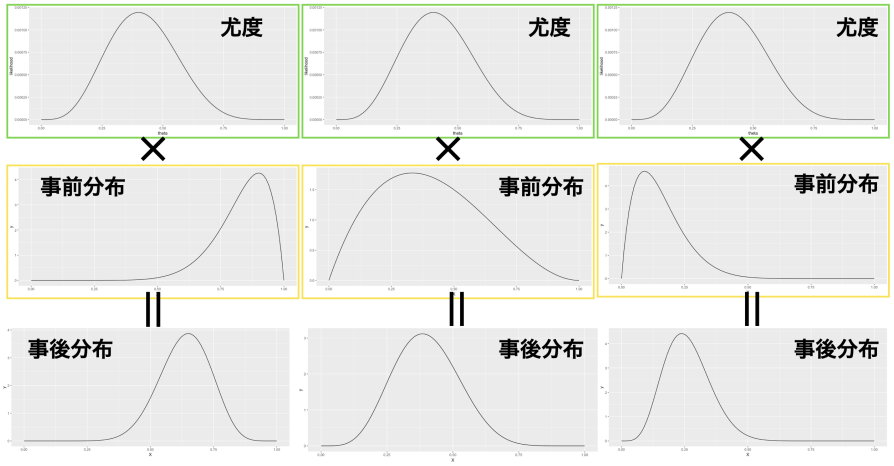


図 26.10 事前分布と事後分布

ば，その事後分布を事前分布として次の分析に使うことができます。このようにデータ＝情報が得られるたびに，どんどんそれを活用して次に分析を進めていくことができるのが，ベイズ推定の大きな特徴です。

事後分布は，事前分布と尤度の積によって形作られます。図 26.10 に 3 つの分析例を示しました。図 26.10 は縦に見ていって欲しいのですが，最上段の尤度に二段目の事前分布を掛け合わせた結果として，事後分布の形が三段目に得られる，というステップを描いています。ここでいずれもデータと確率分布の関係＝尤度は同じで，前回の例である 10 問中 4 回正答がでたというベルヌーイ尤度関数の形になっています。二段目の事前分布ですが，一番左は「この人の能力は高そうだ (正答率が高い)」という分析前の見込みを示しています。この時，事後分布は一番左の一番下で，尤度のピーク (0.4) と事前分布のピーク (0.9 ぐらい？) の間ぐらいにピークを持っています。一番右は逆に「この人の能力は低いに違いない」という分析前の見込み，を示

6938 しています。この時の事後分布は事前分布に引っ張られて，ずいぶんと低い方向によっています。真ん中の
6939 列はその中間で，「ちょっと出来が悪いかもな」，ぐらいの信念を表しています。これも事後分布は，尤度と事
6940 前分布の中間ぐらいの形になっています。
6941 私たちは事後分布を見て結果を解釈しますから，事前分布の持ち方 1 つでこんなにも結果が変わるので
6942 は，何かよくないのではないか，と思ってしまうかもしれません。事前分布が「分析者の恣意的な思い込み」に
6943 なってしまうもしれないということです。勝手な事前分布を置いて答えが出たとしても，違う事前分布を置い
6944 てまったく違う答えが得られるというのでは，科学が恣意的な営みになってしまい，客観的で公平な性質を疑
6945 われるかもしれません。そこで事前分布は客観的に決めるとか，無情報事前分布をおくといった方法が良い，
6946 ということになります。
6947 しかし逆に言えば，事前分布を持っているというのはこれまでの蓄積を活用できるということでもあります。
6948 また，事前分布と尤度が掛け合わされて判断されるのですから，尤度がものすごく尖っている＝たくさんの
6949 データによってある事実が支持されているのであれば，ほんわかした事前分布ぐらいでは事後分布の形が変
6950 わることはありません。逆に，事前分布がものすごく尖っている＝とても強烈な事前の信念がある，あるいは
6951 これまでのたくさんの知識の積み重ねがある，という場合は，少しのエビデンスで結果が大きく変わることも
6952 ありません。その意味で，ベイズの性質はフェアな精神であるとも言えるかもしれません。

6953 26.5 課題

6954 **■事前分布の解釈** 図 26.1 のように，確率分布で表現された確信の程度を自分の言葉で表現してみま
6955 しょう。

6956 **■分布の統計量の説明** EAP 推定値，MED 推定値，MAP 推定値，クォンタイルやパーセンタイル，EDI
6957 区間，HDI 区間についての理解を確認しておきましょう。

6958 **■CI って？** 信用区間と確信区間の違いを自分の言葉で説明できるでしょうか。

第 27 章

ベイズ推定と判断

さて，ベイズ法についてここまでお話ししてきましたが，いかがでしょうか。従来の頻度主義的方法と比べて，確率モデルを使って考えること，結果も確率分布で得られること，さらにその確率の考え方も従来とは違うことなどを考えると，どちらが良いかはなかなか言い切れないところがあるかもしれません[*1]

ところで従来のやり方で，統計的帰無仮説検定という方法で効果の有無を判断する，という話がありました。あれはベイズ法ではどうなるんだろう？というのが今回のテーマです。

27.1　ベイズ法と帰無仮説検定

帰無仮説検定の考え方をもう一度振り返っておきましょうか。

まずさまざまな個人差や誤差が含まれるであろうデータに対して 2 つのグループを作って (多群でもいいですが) 平均化し，そのことで個体差を相殺，一方の群にのみ影響があるような操作を行なって，結果の群平均を見ることで，平均的な**因果効果 (causual effect)** を見ることができるという**要因計画 (Factor Design)** を使って研究が行われます。この要因計画は，影響源をさまざまな要因と交互作用に分解していく**線形モデル**であり[*2]，これを使って効果の大きさを見ることができるのでした。

ただ標本の効果の大きさだけでは限界があるので，これを**母集団**の世界での話に広げたい。そこで**推測統計学**の出番です。平均値の位置や効果の大きさを，標本から推定していくわけです。帰無仮説検定ではさらに，効果があったかなかったかという判断を行うための基準を用意してくれたのでした。効果がないというモデル vs 効果がなくはないという 2 つの仮説＝モデルを立てて，どちらのモデルが良いものかを検討します。その際，いったん効果がないというモデル (=**帰無仮説**) の立場に立って，効果がないと仮定した上でのモデル上の計算を進め，実現値がそのモデルからどれぐらい逸脱しているのかを考えるのでした。ここで効果が濃さに比べてどれぐらい大きいかという，分散の比の分布 (F 分布) を用いたり，二群の平均の差の分布 (t 分布) を用いたりして，実現値がどれぐらいあり得ないかを考えます。これが事前に決めた判定基準，すなわち**有意水準**よりも小さければ，滅多にないことが現実に生じているわけですから，帰無仮説が間違っていたのだ，と考えるのです。ここで，帰無仮説が間違っていたのだとなると，「差がないとは言えない」という結論になります。その結論を使って「もう少し検討しないとな」と考えるもよし，「差がないとは言えないというのは，差があるということだな」と判断するもよし，です。後者の場合は「本当は差がないのに差があると言ってしまう間違い (**タイプ 1 エラー**)」とか，「本当は差があるのに差がないと言ってしまう間違い (**タイプ 2 エラー**)」が

[*1] どっちも嫌いです，というのは勘弁してください。心理学に統計が必要なのはなぜか，ということを考えると，難しいかもしれませんが立ち向かわないわけにはいかないのです。

[*2] 分解していくというのは，全体の変動を要素の足し算に分解していくという意味です。単純な足し算にしているので線形 (一次関数) モデルです。

生じる可能性があるので，注意しなければなりません。差があるかないか，という判断に飛びつくのではなく，どれほど効果があるのか量的に判断すべきで，この効果の大きさについては**効果量**という指標を使って報告する，というフォーマットがあるのでした。また，判断を繰り返しすぎると，タイプ 1 エラーがコントロールできなくなりますから，分析は慎重に，まず**主効果**をみてから事後の検定へ，といった丁寧な進め方をしなければなりません。

ふう！何コマもかけて説明してきたことを，1 つのパラグラフにまとめるとこんな感じになります。十分面倒ですね！とくに「差がないとは言えない」とか，「差があるのにないと言ってしまう確率」のあたりで頭が混乱してくるのではないでしょうか。少なくとも私はいつも，教えていながら間違ってやしないか，ヒヤヒヤしております。どうしてこんなややこしい方法が有り難がられるのでしょう？それは実は，この考え方には**事前分布**のような発想がなく，出てきたデータがなんであれ，正規分布から得られていそうだということさえ仮定できれば，機械的にこの検定手続きを進めていくことができるからです。

帰無仮説検定の話は，農学からでてきました。農業試験場で作物の成長についての実験をしていたところから，区画を分割して効果を比較するといったアイデアが形になっていったわけです。しかし一度，この分析のフォーマットが出来上がると，心理学でも利用できますよね。人間は農業試験場で取れる生産物ではありませんが，さまざまな特徴は正規分布してそうなので，利用すればいいのです。別に心理学に限りません。生物学でも，経済学でも，物理学でも，社会学でも，なんでも結構です。データが分布するとか，誤差が正規分布すると言ったアイデアがあれば，応用は可能です。しかもこの分析手続に含まれる，面倒な計算！とくに t 分布がどうだとか，F 分布がどうだとかいうことを数理的に導出しようとするととても大変ですが，計算のフォーマットは定まっているわけですから，R の関数にポンといれれば結果を出してくれる。難しいことを考えなくても，結果は出てくるのです。これが帰無仮説検定がありがたい理由で，フォーマットが定まっているからさまざまな領域に応用できること，その計算フォーマットを実装した統計パッケージが利用できることで，必ずしも数学が得意ではないわたしたちでも統計の力を借りることができるようになりました[*3]。

私たちは心理学が専門ですから，統計のことはよくわからない，ツールをつかうユーザの立場でしかない。というのはその通りだと思います。しかし同時に，心理学が専門だというのであれば，その結果の解釈や意味の理解にこそ，もっと責任を持たなければなりません。先ほど帰無仮説が棄却された場合について言及しましたが，差がないと言えるかどうか，判断基準を機械的に決めるのは確かに 1 つの方法です。しかしその前に，どの程度の差が実質的な意味を持つのか，何をもって差があると言えるのかについて，丁寧に考える必要があるでしょう。帰無仮説検定はあくまでも，区間推定の結果を判定するやり方の 1 つであり，そもそも推定された平均値の差，その区間，効果量や効果量の信頼区間について，考える必要があるのです。

ここで注意が必要なのは，従来の頻度主義的な考え方では，推定される平均値は定数として定まったものであるということです。**信頼区間**は，その区間の中に推定された平均値が**含まれる割合**を示す数字であり，区間と言いながら幅の情報を持っていません。これに対して，ベイズ法での推定値，たとえば群ごとの平均値の事後分布から考えられる**確信区間**は，確率的に分布する平均値の**とりうる範囲**を示す数字であり，幅の情報そのものです。ベイズ法で推定される平均値の区間推定は，下限から上限の範囲にある，あるいはあると思われる信念の領域を表しており，当たり外れの問題ではないのです。

[*3] しかしこのことが問題にもなってきています。すなわち，よくわかってないのに使って，結果だけ見て考えるというようなことが起こってしまうわけです。免許も持っていないのに，とりあえずアクセルを踏んだら車が走るから踏んでみた，というような話ですね。そんな怖いことを現実社会でやったらとんでもないことなのですが，でも統計の話って難しいし，専門的すぎてどこかは機械や先人に任せるしかない。車だって，MT 車じゃなくて AT 車のほうが普及していますし，素人が AT 車のコンピュータ制御部分を修理したりできるほど専門知識を持ってるはずないですもんね。全部わからないと運転してはいけない，というのはちょっとやりすぎです。つまり私たちは，免許を持てる程度に基本的な知識を身につけ，悪用・誤用をしてしまわないように気をつけなければいけないのです。また，免許には更新があるように，統計の勉強を普段から怠っていてはいけません。新しい技術が使いこなせないと思った場合は，免許返納・車に乗らないという判断もするべきです。

また，ベイズの考え方は事前分布を必要としますから，「分析の前に差があるだろうとかないだろうとか，主観的な思い込みを持ってデータを弄るのはケシカラン」という批判がなされることもあります。しかし逆に，何の前提も置かないことが，果たして良いことなのかどうか，という話もあるのです (クルシュケ, 2014 前田・小杉監訳 2017)。

たとえばコイントスをして「24 回投げて 7 回は表が出た」話があったとしましょう。これは公平なコインか，というのを帰無仮説検定したいと思います。帰無仮説はコインが表になるか裏になるかに偏りはない，というもの。対立仮説は何らかの偏りがある，というものです。データは半分 (12 回) を少し下回っていますから，ちょっと怪しいところではありますが，実際に帰無仮説検定をするとこれは帰無仮説を棄却できません[*4]。つまり，7/24 という結果は 5% よりも生じやすい結果であり，驚くほどのことではないということになります。

しかし，この例がコインではなくて釘だったらどうでしょう？ 釘をトスして，頭の部分が底になって地面の上にピンと立ち上がったら表，そうでなかったら裏だと考えるのです。それで，24 施行中 7 回ピンと立ち上がったとします。なんで？ おかしくない？ 珍しくない？ すごくない？

こんな時でも，帰無仮説検定は棄却できません。24 回中 7 回表が出た，という事実は事実ですから，「おかしくない。こういうことだってある」という結論になります。私たちが「そんな馬鹿な，地面に磁石でもあるんだろ」などと思うのは，放り投げた釘が立ち上がることなんてあり得るはずがない，と強い信念を持っているからです。その信念は物理的な法則に裏付けられたものであり，十分に納得のいくものでしょう。しかしこうした信念すら「事前の思い込みはいけない」といって無視し，「事前に何も思い込みを持ってはいけない，出てきたデータがすべてだ」という態度を取るのであれば，偏りがあるとは言えない，という答えを支持するしかありません。これが科学者の定めでしょうか？

帰無仮説検定は，事前の思い込みだけでなく，データがどのように出てきたかというその生成過程やメカニズムにまで言及しません。一方，ベイズ推定をする時は，事前分布と尤度を明記しなければいけません。事前分布は研究者の事前知識を，尤度はデータが出てくるメカニズムを表していますから，これを明確に記録するベイズ統計のほうがより開かれた態度だということもできるでしょう。

27.2 区間による判定

帰無仮説検定が面倒になる一番の理由は，判断が確率的になることです。判断して間違える確率，というところまで考えないといけないのでややこしいですね。これに対して**ベイズ法**は，「正しい世界」「間違えた世界」を考えるわけではなく，「私が信じられる世界」を考えるので，こうした面倒が生じません。

例として，一要因 3 水準群間デザインをとりあげ，これをベイズ推定したとしましょう。このデザインは第 21 講でやったもので，3 つの群から推定される母平均が同じかどうかということが問題になるのでした。何にせよ正規分布を仮定していますから，3 つの群 A,B,C についての確率モデルは，

$$X_{i1} \sim N(\mu_A, \sigma), X_{i2} \sim N(\mu_B, \sigma), X_{i3} \sim N(\mu_C, \sigma)$$

となっている，と書くことができます。ここで X_{ij} は，j 群に属する i さんのデータ，という意味であり，対応関係がないのでそれぞれ異なる 3 つの平均値 μ_A, μ_B, μ_C の正規分布から出てきているということをこの式が表しています。分散分析の時は，母分散が均等であるという仮定をおきますから，σ には群わけの添字がありません。

これをベイズ推定して，それぞれの群の平均値が得られたとしましょう。結果を図 27.1 に，各群の平均値の推定結果とその 95% 区間とともに示しました。この結果から「差があるかどうか」という意思決定をどのよう

[*4] 表 7，裏 14 という度数分布に偏りがあるかどうかは，χ^2 検定 (カイにじょうけんてい) を使います。R で `chisq.test(c(7,14))` と書くと，統計量 $\chi^2(1) = 2.33, p = 0.1266$ という答えが得られます。

にすれば良いでしょうか。

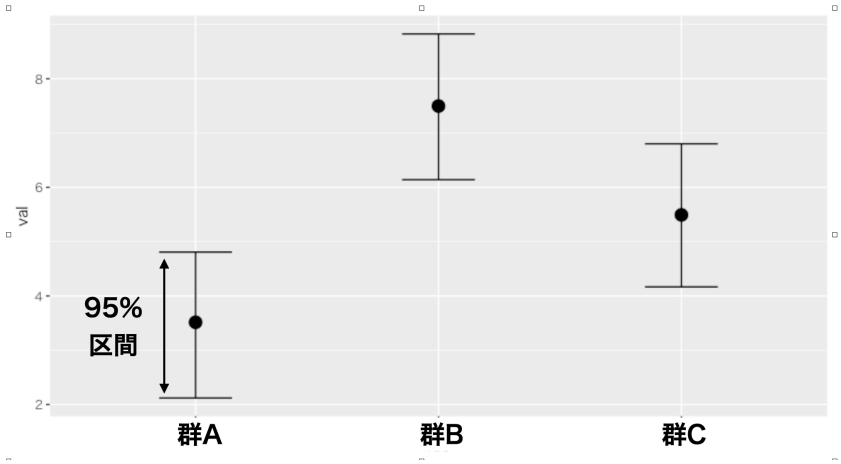


図 27.1 ベイズ推定による推定結果

意思決定に際して，帰無仮説検定は「帰無仮説と対立仮説というモデル対決」に持ち込んでいき，5% の判定ミスを許容した上で決断するのでした。どうしてこのような形になっているかというと，母平均はどこか一点にあって，われわれはそれを探っているから，究極的には推定が「あっているか間違っているか」の 2 択にしかならないからです。

ところがベイズ主義的には，母平均はどこにあるのかわかりません。わからないので「この辺りにありそうだ」というのを確率で表現しています。それは知り得ない 1 点があるのか，それともそもそも確率分布としてあるのかは，哲学的観点が分かれるところです。何にせよあるかないか，という点の勝負ではなく，そもそも分布の勝負になっているので，この 95% 区間のどこかにあるとだけしておきましょう。図には 95% 区間を描いていますが，この幅のどこかに存在するとしたら，群 A の上側 95% 点よりも群 B の下側 95% 点の方が上にあって，この両群の確信区間は重複する領域がありませんから，この 2 群ははっきりと差がある，ということができます。一方，群 A と群 C，群 B と群 C は区間に重複がありますから，これらの組み合わせは差があるとは言えない，ということになります。

That's all!これで終わりです。帰無仮説検定の場合は差があったとしても「差がないとは言えないし，そういってしまって間違える可能性が 5%」というのが正しい表現でした。まだるっこしいですね。これに対してベイズ推定した結果は，「95% の確信度で差がある」ということができます。ここでいう確信度とは言葉の通りこの主張の強さを表していますから，50% でも 99% でも構いませんし，95% よりも 99% の方が確信度が高い，つまりより強く主張していることになります。帰無仮説検定の時の p 値についてはこのような判断ができなかったのに比べて，非常に直感的でわかりやすい結果になっています。p 値については，5% と 1% を比べて 1% のほうが強い，などと言えないのでした。それは p 値がデータや母数についての数値ではなく，1. 帰無仮説という仮定を置いた上での，2. 現実値の珍しさの度合いを表す，3. 判定基準にすぎない，というものだったのでした。これに比べるとベイズ推定して得られた結果は，1. 母数の推定された区間，これだけですので，推定値で物事を考えるのにもってこいです。こうなってくると，5% にこだわった考え方が無意味であることが一層はっきりしてきますね。

この話は，ベイズ法でなくても別にいい話ではあります。というのも，頻度主義的なモーメント法による区間推定で，信頼区間をみて「差があるかなあ，ないかなあ」と考えることはできたわけです。どの群の信頼区間も大きくて，群間のそれが重複しているようであれば，平均値がどのあたりにありそうなのかわかりにくいわけ

ですから，判断するには時期尚早。信頼区間が小さければ，判断があたるか外れるかのチェックをしてもいいけど，という感じで利用することもできます。何が言いたいかというと，帰無仮説検定は「判断」を定めることに無理があったのであり，科学的な研究アプローチとしては「外れたから研究は失敗」という短絡的なものではなく，徐々に何が正しいのかに近づいていけさえすれば良いのです。豊田 (2009) が指摘するように，重要なのは**有意差**ではありません。むしろ重要度はそれが一番小さくて，

有意差 < 標準化された差 (効果量) < 実質的な差

の順に意味があるのです。データが実質的にどれぐらい違うのか，標準偏差いくつ分ぐらい違うのか，ということが大事なのであって，判断結果で有意差があるかどうかは大したことではありません。

とはいえ，確信区間や信用区間を見て「差があるかなあ，ないかなあ」と思いを馳せるだけでは，レポートには書きにくいですね。そこで事前に，どれぐらいの差があれば良いと判断するかという基準を決めておく，というのがいいかもしれません。Kruschke (2018) は，**実質的に等価な範囲 (Region Of Practical Equivalence,ROPE)** という幅を事前に定め，そこに事後分布の**最高密度区間**がどの程度含まれるかで判断すべきである，という主張をしています (図 27.2)。この区間をどうやって決めるんだ，ということについてまた議論があるかもしれませんが，それは分析を応用する各領域において，その領域専門の知識 (ドメイン知識) に依存する話なので一般論として語ることはできません。たとえば数字の上で 1 ポイント差があるというのが，男女の平均身長差の話であれば (そして単位が cm であれば)，なんだ 1cm の違いなんか実質的に意味ないね，という話になるでしょう。しかし 1 人の女性が一生のあいだに産む子どもの数 (合計特殊出生率) の国際比較であれば，子供が 1 人多いか少ないかというのは結構重要な差になります。錯視量が 1cm なのは大きいでしょうか？ 回転角が 1 度ちがうのは？ 反応時間が 1ms 違うのは？ こうした話はいずれも，単位とドメイン知識に依存することなので，一律に言えないのです。

帰無仮説検定は一律に 5% と言ってくれました。しかしそのことが，大事な何かを見失わせたのかもしれません。

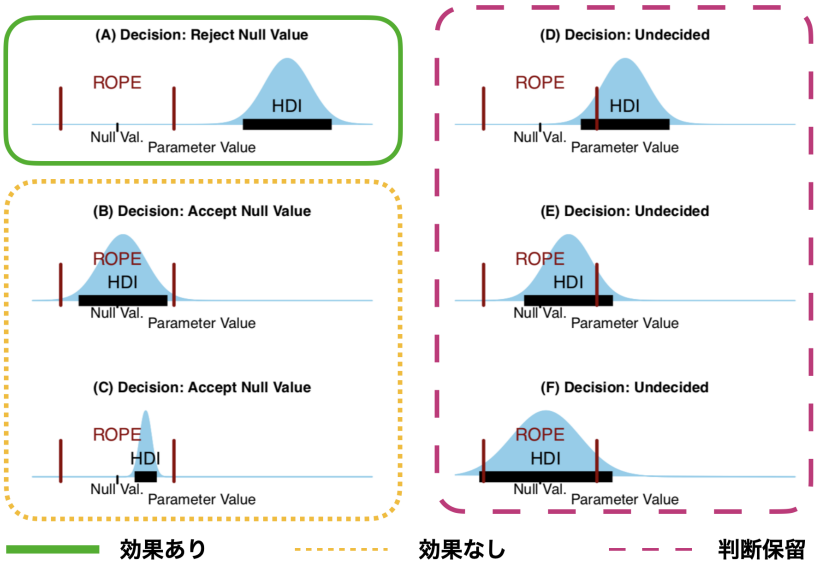


図 27.2 実質的に等価な範囲による判断 (Kruschke, 2018)

27.3 ベイズファクター

ベイズを使った判断基準について，もう 1 つ別の角度から紹介しておきましょう。それが**ベイズファクター (Bayes Factor)** です。

もう一度ベイズの公式を見てみましょう。ベイズの公式は次のようなものでした。

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

言葉で言うと，事後確率は事前確率と尤度の積を**周辺尤度**で割ったもの，となります。ここまで事後分布の形に注目してきたため，分子に当たる事前確率と尤度の方に話が集中していましたが，ここでは分母に目を向けます。分母は $p(D)$，つまりデータが出てくる確率で，定数なので考えなくていいよと言う話をしてきましたが，モデル比較ということを考えると，重要な意味を持ってきます。

さて帰無仮説検定は，帰無仮説というモデルと対立仮説というモデルの対決だ，という話をしました。あるモデル $\mathcal{M}$ をつかって検証する，という条件をこのベイズの公式に追加してみましょう。

$$p(\theta \mid D, \mathcal{M}) = \frac{p(D \mid \theta, \mathcal{M})p(\theta \mid \mathcal{M})}{p(D \mid \mathcal{M})}$$

すべての要素に「あるモデルという条件のもとで」という条件が追加されています。今度はここで分母に注目します。分母の $p(D \mid \mathcal{M})$ を言葉にするならば，モデル $\mathcal{M}$ が与えられた時のデータの確率，ということになりますね。ここでモデル $\mathcal{M}$ はパラメータ θ を持つモデルなのですが，このパラメータがかりに 2 つしか状態を持たない，たとえば θ_1 と θ_2 しかないのであれば，この 2 つを使って次の計算ができます。

$$p(D \mid \mathcal{M}) = p(D \mid \theta_1, \mathcal{M})p(\theta_1 \mid \mathcal{M}) + p(D \mid \theta_2, \mathcal{M})p(\theta_2 \mid \mathcal{M})$$

これはつまり，θ で**周辺化**したことになります。確率のすべてのパターンを足し合わせると，そのパラメータが消えてしまうわけですね。K 個の状態を保つ離散的なパラメータの場合に一般化して書くと，次のようになります。

$$p(D \mid \mathcal{M}) = \sum_{k=1}^{K} p(D \mid \theta_k, \mathcal{M})p(\theta_k \mid \mathcal{M})$$

連続的なパラメータであっても，総和 $\sim$ の記号が積分記号 $\int$ に変わるだけで，式としてはほぼ同じです。

$$p(D \mid \mathcal{M}) = \int p(D \mid \theta, \mathcal{M})p(\theta \mid \mathcal{M})d\theta$$

さてこのように考えると，$p(D \mid \mathcal{M})$ はモデルを考えた上でのデータが得られる確率，ということであり，求め方は「モデルが想定するすべての確率を網羅的に数え上げたもの」ということになります。検定のロジックは帰無仮説と対立仮説という 2 つのモデルでしたから，$p(D \mid \mathcal{M}_{null})$ と $p(D \mid \mathcal{M}_{alt})$ というモデルを比較するという話だったことになります。そしてこの確率は，想定したモデルからデータが得られる確率，という意味ですから，言い換えると想定したモデルがどれほどデータに支持されているかということでもあります。モデルの想定通りのデータが出てきたのであれば，この数字は大きな値になりますし，モデルの想定していなかったデータが出てきたのであれば，この数字は小さくなります。もしモデルが「どんなデータでも出てくるよ」というような漠然とした広ーい予測しかしていなかった場合，これはパラメータの範囲がごく広いということですか

ら，そのごく広い範囲すべての場合にわたってデータを計算すると，相対的に予測していないところからデータが出てきていることが多くなり，結局この値は小さくなります。つまり「なんでも説明できるモデル」は何も説明しないことになるのです。この値が，モデルがデータに支持されている程度を表している，ということの意味が伝わりましたでしょうか。

その上で，次のような数字を考えます。

$$BF_{12} = \frac{p(D \mid \mathcal{M}_1)}{p(D \mid \mathcal{M}_2)}$$

この数字は**ベイズファクター (Bayes Factor)** と呼ばれるもので，モデル 2 に比べモデル 1 がどれほどデータに支持されているかを示す数字です。この数字が 1.0 より大きければ，モデル 1 はモデル 2 に比べて相対的に強くデータに支持されているということになります。逆に 1.0 を下回ると，モデル 2 のほうがモデル 1 よりもいいね，ということです。モデル同士の比による表現ですので，BF_{21} としても構いません。ここでモデル 1 が対立仮説，モデル 2 が帰無仮説だとすると，どちらがよりデータをうまく表しているか，どちらがよりデータに支持されているかがわかるわけですが，このやりかただと「差がない」という帰無仮説を積極的に支持することもできます。このように指標化されると，私たちはつい「じゃあ BF がどれぐらいの大きさだったらいいんだ」と目安を求めてしまいがちですが (そしてそれは一般に 3 以上と言われていますが)，p 値の 5% を目指した研究が陥ったのと同じ無意味な目標数値になってしまうかもしれませんので，そういうことは考えない方が良いかもしれません。

ともかく，これがベイズ流のモデル比較の方法のひとつです。他にも情報量統計学の観点からモデルの予測力などで評価する方法[*5]などもありますが，いずれもこの周辺尤度を使うやり方です。問題は，この周辺尤度の計算がとても難しいということにあります。パラメータが複数になると，分母のところにあらゆるパラメータのあらゆる組み合わせについて，重積分をしないといけないことになります。これは数値計算上とても難しいことで，実用面ではなかなか厳しいというのが本当のところです。

しかし，帰無仮説と対立仮説という関係なら，簡便的にベイズファクターを算出する方法があります。たとえば二群の平均値の差の検定に関して言えば，帰無仮説は $\mu_1 = \mu_2$ であり，対立仮説は $\mu_1 \neq \mu_2$ というものでした。対立仮説は 2 つの母平均がイコールでなければすべて OK，という仮説であり，言い換えれば $\mu_1 - \mu_2 \neq 0$ ということです。つまり差である $\mu_1 - \mu_2$ を X 軸に取ったとすると，帰無仮説は $X = 0$ というピンポイントな，ごく部分的な仮説になっています。対立仮説の広い範囲の中で，帰無仮説はその一部・特殊ケースということであり，このようにネストされた仮説の場合は**サヴェージ・ディッキー法 (Savage-Dickey method)**(リー・ワゲンメーカーズ, 2013 井関訳 2017) という計算法でベイズファクターが計算できることがわかっています。

27.4 ベイズ推定の利点

いろいろ説明してきましたが，**ベイズ法**による推定でも，効果があるかどうか，平均に差があるかどうかを考える方法はあるということです。区間推定を使って考える方法，ベイズファクターという指標を使う方法などが考えられており，これらは帰無仮説検定にはなかったメリットがあります。以下，3 点挙げて説明しましょう。

■そのまま大きさの評価ができる まず 1 つは，上でも述べたように 1bit 判断に陥りにくいことがあります。差，つまりどれほどの効果があったのかを直接議論しますから，最初から効果量を，効果の大きさを考えていることになります。帰無仮説検定の問題点として，差が大きかったのか小さかったのかを無視して「とにか

[*5] WAIC という指標を使います。詳しくは浜田他 (2019) などを参照してください。

く有意である」の一点張りで議論を組み立ててしまう危うさがあったわけですが[*6]，ベイズ推定をもちいた考え方であれば，このような誤解に陥る危険性が未然に防ぐことができているとも言えます[*7]。とくに未知な状態が多い現象について探索的な研究をする場合は，帰無仮説検定が性急に判定結果を出すことで，誤った方向へ仮説や理論が誘導される可能性があるので，この問題は心理学全体にとって重要なポイントだと思われます。

■補正や多重性の問題から解き放たれる t 検定や分散分析において，分散が均等であるかどうかを検定し，均等であるという仮説が保たれていないのであれば Welch の補正をするだとか，球面性の仮定が保たれていないのであれば Greenhouse-Geisser の補正をする，といった説明を重ねてきました。これらは最初に考えられた ANOVA について，理論の展開とともに修正パッチをどんどん当てていくようなもので，非常に煩わしい問題でした。また分散分析の結果は，事後の検定が必要です。差がないとは言えない，じゃあどこにあったんだ，ということで改めて探しにいく必要がありましたし，この時に α 水準のインフレに怯えてさまざまな補正，修正を考えていかなければなりませんでした。これも非常に煩わしい問題です。

ベイズ推定になれば，この問題が一気に解決します。後者の α のインフレについては，帰無仮説を設定するわけでも確率を含んだ判断をするわけでもないのですから，そもそも問題になり得ません。分散が等しいかどうか，球面性の仮定が保たれているかどうかといった検定を積み重ねる必要もありません。分散や共分散も未知なるパラメータとして推定するだけで，確率的判断はそこにはいってこないからです。

また，帰無仮説検定が有している問題の 1 つに，**停止規則 (Stopping Rule)** に関する問題があります。たとえばデータを 5 人分集めて，有意にならなかったので，頑張って 10 人にデータを増やして分析し，それでもダメだったので (?)20 人に増やしたら有意になったので，差がある結論としてレポートした，という研究方法は，**帰無仮説検定的にはご法度**です。帰無仮説検定は，データを取る前に勝負のルールを決めて判定します。勝負の前になんら事前情報がない，フェアな形で進めなければなりません。サンプルサイズが大きくなると有意な判定になりがちである，という性質がありますから，「徐々に増やしていって有意になったところで止める」というのは勝つまで延長戦を続けるという不正なやり方なのです。これに対して，ベイズの流儀は基本的にデータが追加されると情報がアップデートされますので，徐々に確信度が変わっていくだけですから，最初の 5 人で得られた事後分布を事前分布として，追加の 5 人の情報を加えた事後分布を計算するというのはごく自然なやり方です[*8]。このように，帰無仮説検定が持っていた問題，とくに初心者が無意図的に陥りがちな問題を自然に回避できるのも，ベイズ法の利点と言えるかもしれません[*9]。

■帰無仮説を積極的に支持できる 研究の実践上は，関係がないとか効果がないということを示したいというシーンもあるでしょう。しかし帰無仮説検定の枠組みの中では，それはできませんでした。なぜなら帰無仮説を棄却することで決断するのであり，「差がない・効果がない」という仮定の下で計算を進める以上，その仮定が正しいかどうかはその技術で検証することはできないのです。

[*6] それにしたって「差がないとはいえない」「相関がないとは言えない」という程度の主張でしかないのが歯痒いところです。「君が僕と付き合う気がないとしても，今日こうやってデートできたのなら，君が僕のことを好きではないとは言えない」という帰無仮説検定のやり方って，あんまりかっこよくないですよね。「今日こうやってデートできたのだから，君が僕のことを好きだという確信がさらに強まった」というベイジアンの方がモテると思いませんか?

[*7] もちろんこれは私の「ベイズ贔屓」が入っている表現で，頻度主義的発想であっても帰無仮説検定のような判定をせず，推定値とその信頼区間で考察するに留めておけば良いのです。線形モデルでも最尤推定してその信頼区間を考えればよく，無理に判定までする必要はないからです。

[*8] この「最初の 5 人の事後分布を事前分布として 5 人追加する」というやり方と，「10 人分まとめて事後分布を計算する」というやり方，いずれも事後分布の形は同じになります。

[*9] 停止規則に関する問題については，「頑張れば有意になる」という努力と科学的判断の交わるところですので，研究実践や教育的配慮の観点から知らず知らずのうちに罪を犯してしまうという，危うい性質のあるものなのです。研究も人間の営みの一つですから，わかっていてもなかなか区別しにくいところがあるのでしょうね。もちろんわかっていて，有意にするためにデータを加工するのは捏造に近い科学的犯罪です。

ベイズ推定の流れでは，無理な帰無仮説をおくのではなく，自然な確率モデルをおくだけですので，傾きが 0 であるかどうかは確信区間の中に含まれているかどうかで表現できます。確信区間が狭くなれば狭くなるほど，その傾きについての確信度が高まっていくわけですから，はっきりと差がない，というような言い方をできます。

また帰無仮説検定はモデル比較だという話をしてきました。ベイズ推定でも，モデル同士を比較してどちらのモデルがよりデータに支持されているかを，ベイズファクターとして表現できます。この時，帰無仮説的な傾き 0 のモデルが，そのほかのモデルに比べてどれぐらいデータに支持されているか，ということができますから，「差がないぞ」ということを積極的に主張できます。

27.4.1 ベイズの世界の広がり

こうして見てくると，帰無仮説検定が少し窮屈に思えてくるでしょう。帰無仮説検定は正しく結論づけるために，非常に細かくルール・マニュアルを設計していきますが，そのマニュアルが窮屈すぎるきらいがあることは否めません。ここまでくると，そもそも線形モデルであるというのも気に入らなくなってきませんか。これまでは帰無仮説検定しかツールがなかったので，無理やり正規分布かつ線形の枠組みに研究を押し込めてきたという歴史的経緯がありました。

たとえば比率のデータは正規分布しませんから，分散分析の対象にはなりません。どうしても分散分析したい場合はデータを角変換して補正するとか，所得や反応時間のように正規分布しないものを対数変換して補正する，といったように，データの変換というのが行われてきました。これは分散分析に落とし込むしか検証の方法がなかった，全時代的な風習です。

線形モデルも発展していき，正規分布の仮定を取り払った**一般化線形モデル (Generalized Linear Model)**(GLM) では，二項分布やポアソン分布といったデータの生成過程にあった分布を選択できるようになりました[*10]。いまではデータの変換ではなく，適切な確率関数・尤度関数を選ぶのが基本的なやり方です。

それでも線形の性質は保持されたままです。もっと自由に変数間関係を表現することはできないのか，という考えが湧いてきた人もいるかもしれません。もちろんそれは可能で，線形ではないモデルを自分から考える，**モデリング (Modeling)** の考え方は現代心理学にもどんどん広まっています。そのモデルのパラメータ推定には，ベイズ統計の力が非常に強力で役立つので，とくに**ベイジアンモデリング (Bayesian Modeling)** と呼ばれています。

これからの心理学には，ベイズの力を借りたモデルや分析方法がどんどん広まっていくものと思われます。詳しくは姉妹講義，心理学データ解析応用でお話しすることになります。

[*10] 平均値の差の検定は線形モデルと同じ (一般) だぞ，ということを一言でいうと一般線形モデル General Linear Model といいます。日本語でいうと「化」，英語でいうと –zed があるかないかで意味が随分変わってしまうので要注意です。

第 28 章

JASP によるベイズ推定の実践

さて前回は，ベイズ推定を使った判断の方法について，**HDI** と **ROPE** をつかって考える方法と，**ベイズファクター**をつかって考える方法などを紹介してきました。

これらの方法も，実際に計算できなければ絵に描いた餅，ということでベイズ推定を実際に行う方法について具体的な方法を学んでいきましょう。

28.1 ベイズ推定の方法

28.1.1 解析的な方法

ベイズ推定を実際に行うには，もちろんベイズの定理をつかって計算すればよい，ということは誰にでもわかることですね。ベイズの定理は次のようなものでした。

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

なるほど，確かにこれを計算すればよいのですが，たとえば正規分布の平均値 μ と σ をこの計算方法で求めようという場合，$P(B|A)$ は尤度ですから，

$$P(X|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)$$

というように表せます。続いて事前分布，$P(A)$ はパラメータについての事前の信念ですから，$P(\mu,\sigma)$ を何らかの確率分布として表す必要があります。多次元確率分布ですから，面倒なことになりそうな予感がビシビシ伝わってきますね。分母の $P(B)$ は，パラメータ全区間についてのモデルの周辺尤度ですから，$\int_{\mu}\int_{\sigma} P(X|\mu,\sigma)P(\mu,\sigma)d\mu d\sigma$ という計算になります。$\int\int$ って！これは積分記号ですが，積分が複数重なるのでまぁ計算が面倒です。今回はパラメータが 2 つでしたが，これ以上になると実質的に計算できないレベルで難しいものです。ベイズの定理そのものは古くから発見されていましたが，なかなか実用化されなかった理由のひとつは，この積分計算の複雑さにあったのです。

それでもベイズ推定したいという場合はどうするかというと，「事前分布と事後分布が同じ関数の形になるもの」をうまく探し出す，というやり方でした。たとえば尤度が**ベルヌーイ分布**や**二項分布**であったときに，事前分布に**ベータ分布**というのを想定すると，事後分布もなんとベータ分布の形をしてくれるのです。これは計算式の展開から明らかで，こうした特徴を持っている分布のことを**共役事前分布 (conjugate prior distribution)** といいます。時折こうした特別な組み合わせが見つかりますから，うまくいけば計算で答えが出せるのですが，パラメータが多くなると事後分布は何だかよくわからない形，式はかけるけど計算できない形，一般化できない何かになってしまい，お手上げです。

28.1.2 グリッドサーチ

次にご紹介する方法は，グリッドサーチという方法です。第 26 講で，次のような表現をしました。

$$P(\text{事後分布}) \propto \text{尤度} \times \text{事前分布}$$

これはつまり，事後分布の計算をするには分母の計算は無視して比例の関係 $\propto$ に注目しましょう，ということです。というのも，分母は面積を 1，すなわち尤度と事前分布の積を確率にするための数字でしかなく，我々が興味があるのは事後分布の MAP 推定値など，事後分布のピークがどこにあるかということだったりするので，山の形状・勾配さえわかれば，標高はとくに気にならないわけです。

ですから，事前分布と事後分布の積を数式で表現して (これは面倒ですが頑張ればできます)，パラメータに適当な数字を入れていくとします。たとえばベルヌーイ尤度関数とベータ分布の組み合わせでしたら，パラメータは θ だけですから，$\theta = 0.0$，$\theta = 0.3$，$\theta = 0.6$ ･･･などパラメータを適当なサイズに区切り，式に代入して確率が高くなるところを探します。もし $\theta = 0.3$ より $\theta = 0.6$ のほうが事後分布のピークが高くなりそうであれば，$\theta = 0.55, 0.575, 0.600, 0.625...$ のようにさらに細かく区切り，さらにさらに･･･と繰り返していけば「ピークはこの辺りでしょう」という近似的な推定ができるでしょう。計算機はこの手の反復が得意ですから，時間を掛ければ何とかできるはずなのです。ちなみにもちろんこれは話を簡単にするために，解析的にも解ける例をわざわざ出しているだけで，本当はもっと複雑なモデルに対して適用する話になります。正規分布モデルであれば，μ, σ の 2 つのパラメータそれぞれについて細かく区切るわけです。区切られた区画のことをグリッドと言いますので，グリッドに分けて探す，すなわちグリッドサーチです。

この方法はうまくピークを探せるかもしれませんが，ピークが複数ある山脈のような事後分布になると，ひょっとしたらもっと高いところがあるのに見つけられないということがあるかもしれません。またパラメータの数がもっともっと多くなると，計算ももっともっと時間がかかるということで，大変計算量の多い方法になってしまい，これも現実的な解法ではないと言えるでしょう。

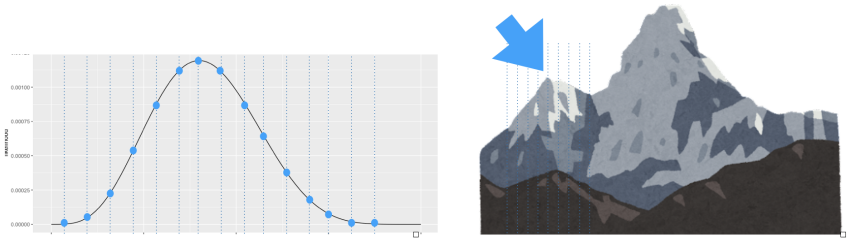


図 28.1 スライスしてピークを探すグリッドサーチ法

28.1.3 MCMC 法

さて最後に，現代ベイズ統計の盛り上げに一役も二役も買っている，**マルコフチェインモンテカルロ法 (Marcov Chain Monte-Carlo)** の登場です。通称 **MCMC** と呼ばれるこの手法は，マルコフチェインとモンテカルロの 2 つの技術が組み合わさったものでです。マルコフチェイン (**マルコフ連鎖**ともいいます) は，目標となる確率分布状態になるような推移規則を作る方法であり，モンテカルロ法は乱数を発生させるアルゴリズムです。この 2 つが合体することで，確率分布の形がわからなくてもサンプルが得られるような，乱数発生器を作る方法が生まれたのです。

第 7 講 (セクション 7.5, Pp.122) で少し触れたように, 乱数発生は確率分布を近似するための 1 つのアプローチです。R には正規分布や二項分布, t 分布や F 分布など, 有名な確率分布関数に関しては自前で準備しており, `dnorm` のようにすれば正規分布 `norm` の密度 density を算出するとか, `qt` のようにすれば `t` 分布におけるある確率を取る範囲の確率点 `q` を返すといったことができるのでした。さらに `rnorm(N,mean,sd)` という関数は, 任意の平均と標準偏差をもつ正規乱数を N 個生成できます。大量に乱数を発生させると, そのヒストグラムは正規分布の形を再生するのです (図 28.2)。

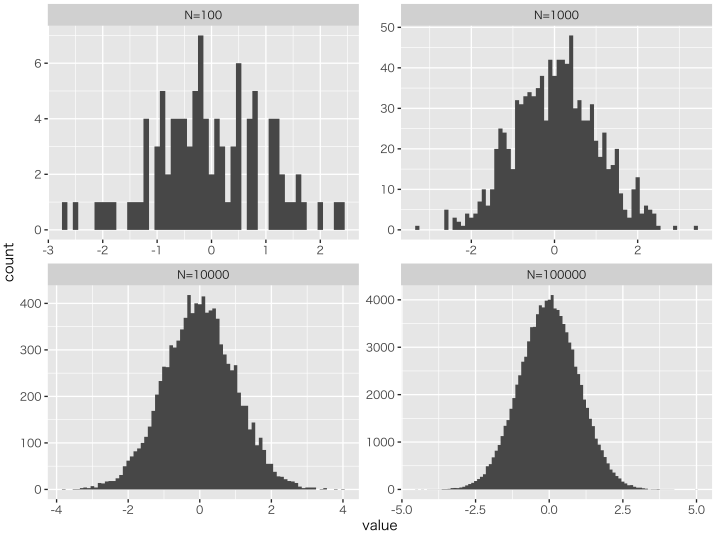


図 28.2 乱数が増えると正規分布の形に近づく。左上から順に 100,1000,10000,100000 点の乱数を発生させて描いたヒストグラム

コンピュータの中で乱数を発生させるのはあくまでも**擬似的**です。コンピュータの世界は計算され尽くした世界だからです。「規則正しく」「規則のない数列を生成する」というのは矛盾していますね。計算機科学の中では, この矛盾を解決し, 計算規則には則っているのだけれども規則のないような数字を発生させる方法について, いろいろな研究がなされてきました。そうした乱数発生技術の 1 つとして**モンテカルロ法**ができてきたのです[*1]。

ただし, ここで我々が求めたいのは事後分布です。事後分布も乱数で近似できればありがたいのですが, 事後分布の形は数式で書けても一般化できない名も無き確率分布なのでした。この事後分布の状態を乱数発生技術とつなげる方法として**マルコフ連鎖**というのがあります。マルコフ連鎖とは確率の推移に関する考え方で, ある時点 t の状態はその 1 ステップ前, $t-1$ 時点によって確定するという状況の中で, 推移のパターンを安定させることを研究する, 確率推移状態のモデルです。この安定したパターンが事後分布になるように上手く調節し, 推移ステップごとに乱数を生成すると, うまく事後分布を代表する数字の列が出来上がるという仕組みです。

非常に大雑把な説明ですが, 要するにどんな確率分布であってもそこからの乱数を発生させる方法が開発された, ということだと考えてくれれば OK です。事後分布の関数は難しいかもしれませんが, そこから大量の乱数を発生させることができれば, それのヒストグラムを描いたらその稜線が事後分布の形を表していることになります。大量のデータを計算したりそこから描画したりするのは, 最近の統計環境であれば簡単で, 数万・数十万のデータ (事後分布からの代表値=MCMC サンプル) でも一瞬でその手のデータ処理を

[*1] ちなみにモンテカルロとは, モナコの地区の一つでカジノが有名な都市です。カジノのルーレットのように乱数を発生させるという意味でこの名前がついています。

してくれます。

またこの乱数発生法のメリットとして，複数のパラメータがあっても困らないということがあります。求めたいパラメータが μ と σ の 2 つある場合，パラメータは 2 次元空間に広がっていますから，たとえば μ だけの分布が知りたい場合，解析的に考えると σ を全領域にわたって積分し，消してしまわないといけないことになります[*2]。

しかし乱数発生法の場合，この積分や周辺化消去についてはとくに意識する必要はありません。図 28.3 には 2 次元空間から乱数を発生させた散布図と，周辺化した密度プロットを描きましたが，この密度プロット，要するに変数ごとのヒストグラムを書いただけです。積分のような面倒なことを考えなくても，変数ごとに考えることがすなわち，他の変数の影響を排除して考えたことになるからです。

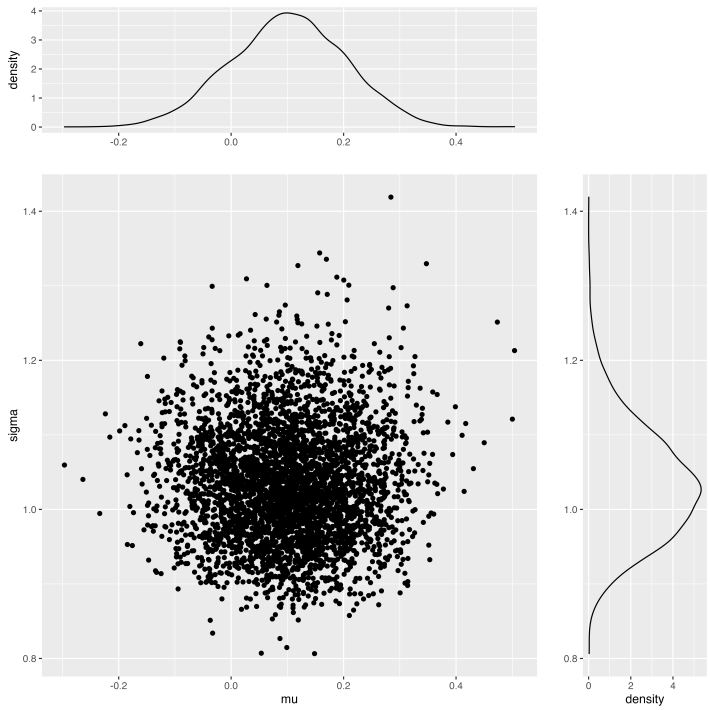


図 28.3 周辺化は変数ごとのヒストグラムでよいことになる

乱数を使うとこのように，積分の計算が簡単になります。多次元の場合でも注目する変数だけ集計すればよいことになりますし，ある変数 (パラメータ) の部分的な集計がその範囲の積分になるのですから。こんなに簡単で確実な方法はありません。

こうした MCMC 法をつかった推定ができたおかげで，簡単にベイズ統計を実践できるようになりました。この MCMC 法を実現するには，乱数発生法を持ったプログラミング環境が必要です。これら**確率的プログラミング言語 (stocastic programming language)** と呼ばれる代表的なものとして，**JAGS**(Wood, 2016) や **Stan**(Carpenter et al., 2017) と呼ばれるものがあります[*3]。これらはいずれも，データと確率モデル，事前分布を設定すれば，尤度と事前分布の積を内部で計算し，そこからの乱数を発生させてくれま

[*2] 前回やった**周辺化**というやつです。

[*3] 以前は BUGS や WinBUGS という言語がありました。BUGS は 1989 年から開発が始まり，後継の WinBUGS は 2007 年に最新版が出て終わりました。古いテキストを探すと，これらのソフトを使う例が載っています (Lee・Wagenmakers (2013 井関訳 2017) など)。その後発として JAGS が，さらに乱数発生アルゴリズムに改良を加えた Stan は 2021 年現在も最新版が開発され，リリースされ続けています。

す[*4]。これらの言語は単体で使うのではなく，R など他の統計環境から呼び出して使います。R は rstan パッケージや cmdstanr パッケージをつかって Stan と連携をとり，乱数発生は Stan に，集計や描画は R でできるようにしています (図 28.4)。

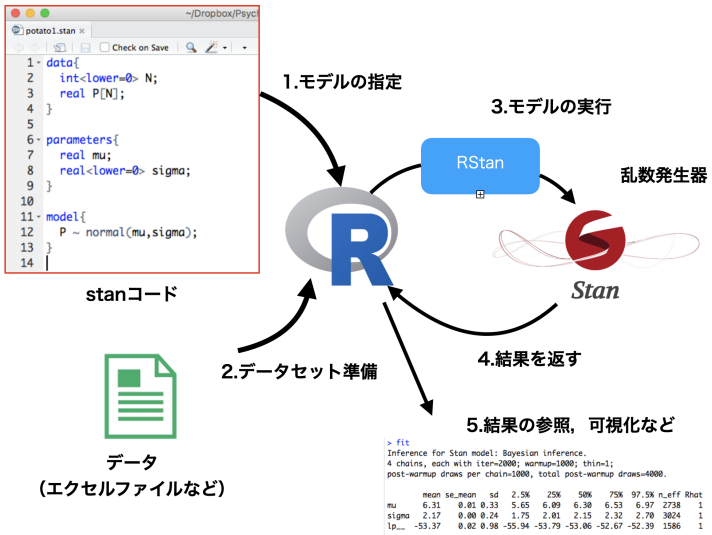


図 28.4　R から Stan を使うときのイメージ

Stan 用のプログラムコード，すなわち乱数発生のためのレシピも RStudio 上で編集できますので，これを使えばベイズ分析ができます。

それでも難しそう，敷居が高いと感じる人がいるかもしれません。大丈夫，Stan のコードを書かなくても Stan を使って分析するための R パッケージというのもあります (brms(Bürkner, 2017, 2018) や blavaaan(Merkle, Fitzsimmons, et al., 2021; Merkle & Rosseel, 2018) など)。そしてさらに，もっと簡単にベイズ分析ができるような統計パッケージがあります。さっそくそのチョー簡単な統計解析ソフト，**JASP** をみてみましょう。

28.2　JASP の導入

JASP もまた R と同じくフリーソフトウェアです。こちらはアムステル大学から支援を受けたオープンソースプロジェクトだとされています。JASP と検索すると，メインのサイト https://jasp-stats.org/ にアクセスできるようになるでしょう (図 28.5)。本稿執筆時点[*5]で最新版は ver0.16 となっています。ここに，ダウンロードのボタン (Download JASP) がありますのでクリックすると，各種 OS に対応した JASP が用意されています。自分の環境に合ったものをダウンロードし，アプリをインストールしてみましょう[*6]。インストールが無事終了したら，起動してみましょう。JASP の起動画面がすでに日本語化されていると思います。そこには JASP の特徴である，**フリー，フレンドリー，フレキシブル**の説明があります。とくに「フレキシブル」のところにあるように，このソフトは他のどのパッケージにもない特徴として，伝統的な分析方法とベイズ的な分析方法をシームレスにつなげているという特徴が明記されています。

[*4] これらの詳しい使い方については，姉妹講義であるデータ解析応用の方で取り扱いますので，興味があれば是非履修してください。

[*5] 2021 年 12 月 17 日 14 時 48 分

[*6] macOS ユーザで Homebrew によるソフトのパッケージ管理をしているひとは，`brew install JASP --cask` でインストールすることができます。

図 28.5 JASP の Web サイト

図 28.6 自分の環境に合ったものをダウンロードしよう

また下の方には「JASP は猛烈なスピードで開発されています！」とあります。JASP のバージョンはまだ 1.0 にも満たないものですが，すでに十分使えるレベルになっているのに加え，開発も進んでいます。ちなみに日本語化チームには著者も加わっており[*7]，もし不自然な訳語や誤った表記，誤訳などがあるようでしたらご連絡ください[*8]。

少し触ってみましょう。画面の左上にある 3 本線をクリックするとメインメニューが開きます。ここで「開く」→「コンピュータ」と進むと，自分の PC にあるデータファイルを読み込むことができます。例として，野球選手のデータファイル (csv) を読み込んでみます[*9]。読み込んだらデータファイルが表示されます。ここで変数の横に，ベン図のような記号がついていたり，物差しのような記号がついていたりするのがわかると思いま

[*7] 詳しくは https://jasp-stats.org/2021/10/28/日本語でjaspが使えるようになりました/
[*8] あるいは JASP ユーザの会 https://github.com/jasp-user-jp/jasp-user-jp に連絡いただいても結構です。
[*9] このサンプルデータは本稿の伴走サイトにて提供されています。

図 28.7　起動画面のご挨拶も是非一読してください

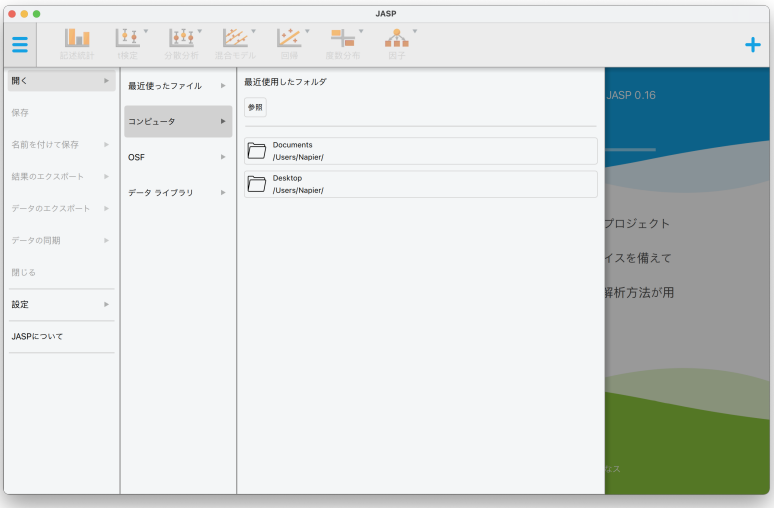


図 28.8　ファイルを読み込む

す。JASP ではデータを**名義尺度水準**, **順序尺度水準**, **スケール**の 3 つに分類します。スケールは**間隔尺度水準**と**比率尺度水準**を合わせたものです。スケールになると加減乗除など算術的処理ができるので, それ以外の 2 つと区別するのですね。

変数名のところをクリックすると, これら 3 つの水準が表示されますので, もし間違った分類があるようでしたら修正しておきましょう。

メニューにある「記述統計」をクリックすると, その名の通りさまざまな変数の情報を簡単に要約してくれます。たとえば身長変数 `height` と体重変数 `weight` をチームごとにみてみたい, というのであれば,「分割」のところに `team` 変数を投入し (右に向いた三角形ボタンを押すと, 選ばれた変数がボックス内に移動します), 分析対象の変数を上のボックスに投入すればよく, 右の画面に自動的に計算結果が出てきます (図 28.9)。簡単ですね！

記述統計メニューは, 下の方に「プロット」や「統計量」についてのオプション指定ができる場所があります。ここをクリックすると, さまざまなオプションが表示されます。たとえばプロットでは, 色を指定するカラーパレッ

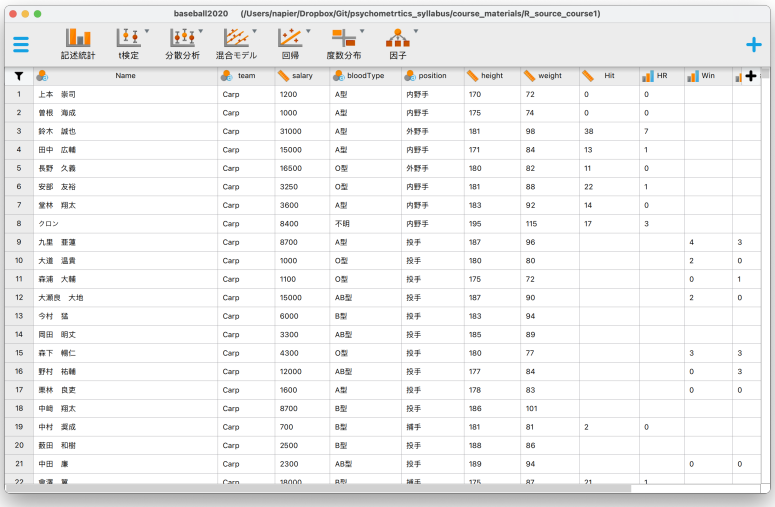


図 28.9 ファイルを読み込みました

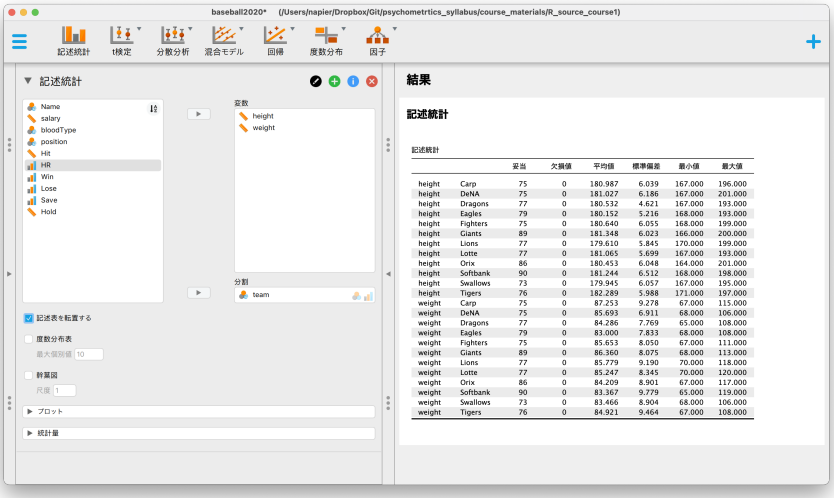


図 28.10 記述統計の出力

トが選べますし，箱ヒゲ図，散布図，回帰直線の追加などが選べます (図 28.10)。

もちろん細かな図版の調整には **ggplot2** の方が向いているのですが，これでも概略を掴むには十分ですね[*10]。

これらのプロットは，図の横にある下三角をクリックすると，コピーしたり画像ファイルとして保存することもできます (図 28.13)。JASP だけで画像の調整がうまくいくのであれば，プレゼン資料の作成などに利用できるでしょう。

[*10] JASP に準拠したテキストがオンラインで公開されていて，そちらも日本語訳が進められているのですが，そこでも「細かい調整は ggplot2 でないとね」といった内容の記述があります。詳しくは JASP で統計を学ぼう!のサイト https://jasp-user-jp.github.io/LSwithJASP/ をご覧ください。

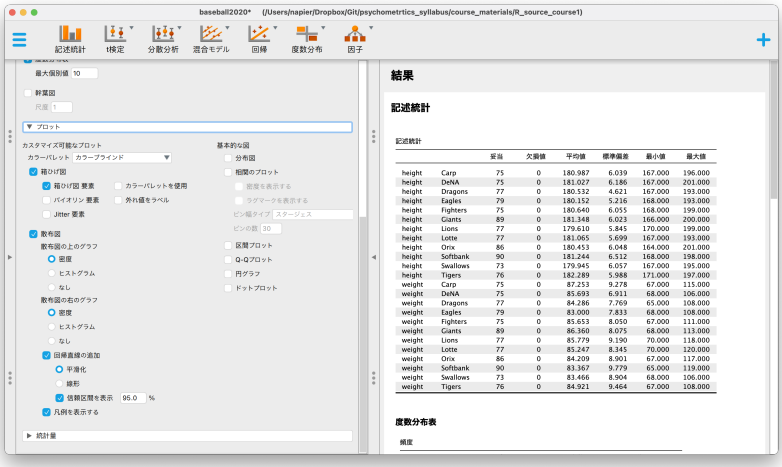


図 28.11 プロットオプション

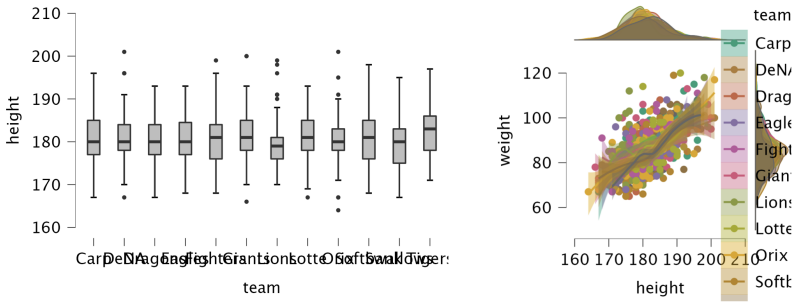


図 28.12 ボックスプロットや散布図

28.3 JASP による相関・回帰分析

JASP の特徴はすでに述べたように，1 つの分析法を従来型 (頻度主義的確率に基づくもの) とベイズ型 (朱 k ん確率に基づくもの) の両方で並列的に行えることです。メニューをクリックすると，伝統的な方法とベイジアン (ベイズ的な) が選べるようになっています (図 28.14)。また，もう 1 つの特徴として，サンプルデータが豊富なことが挙げられます。JASP が持っているサンプルデータだけでなく，OSF[*11]とよばれる Web 上で公開されている研究データにアクセスし，そこからデータを取ってくることもできるようになっているのです (図 28.15)。

ここで適当なサンプルデータを選んで，相関係数や回帰分析を見てみることにしましょう。`Big5` というデータを選びました。データライブラリ →4.Regression →Big Five Personality Traits と進めばデータが表示されます。これはパーソナリティ心理学の領域ではよく知られ散る，性格テストの結果から明らかになる性格の 5 次元で，解放性 Openness，誠実性 Conscientiousness，外向性 Extraversion，協調性 Agreebleness，神経症傾向 Neuroticism のスコアからなります。性格に関するスコアですから，当然相関がありそうな変数

[*11] Open Science Framework の略で，心理学の再現性問題などのとも相まって，研究で使ったロウデータそのものを公開したらいいじゃないか，という活動です。詳しくは https://osf.io/gceur/ を参照してください。

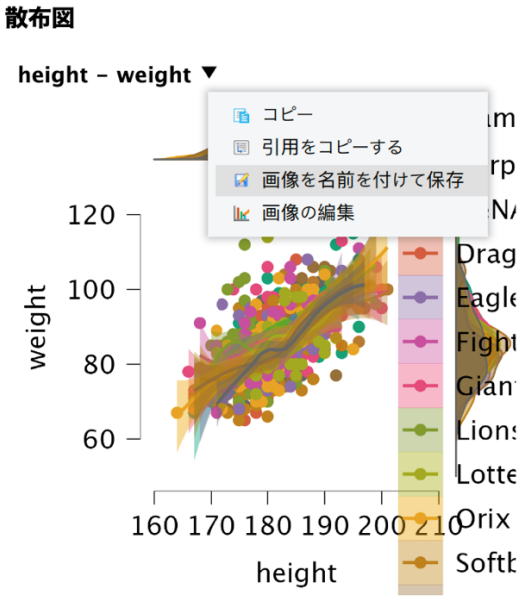


図 28.13　JASP の図オブジェクトは保存できる

図 28.14　一つのメニューで二つの分析

ですね。

普通の相関係数が欲しいだけでしたら，メニューの回帰 → 伝統的･相関を選びます。すべての変数を右側のボックスに移動するだけで，結果のところに相関行列が表示されます。オプションのところにはデフォルトで，「ピアソンの r」と「有意性の報告」にチェックが入っています。いわゆる相関係数は**ピアソンの積率相関係数**というのが正式名称ですが，JASP では順序尺度水準の数字に適用できるスピアマンの順位相関係数や，ケンドールのタウ b と呼ばれる相関係数にすることもできます。また，有意性の報告にチェックが入っていて，表には p 値がでています。思い出して欲しいのですが，帰無仮説検定として「母相関 ρ はゼロである」という無相関を仮定したモデルと，「母相関はゼロではない」という対立仮説を比べています。たとえば神経症傾向と外向性の相関係数は $r = -0.3650$ であり，これが母相関 $\rho = 0$ のところから得られたデータだと考えるにはちょっと珍しすぎるよね $(p < .001)$ ということが示されています。オプションで「有意な相関のフラグを立てる」にチェックを入れると，有意なところにはアスタリスクマークがつき，p 値が小さいほどアスタリスク

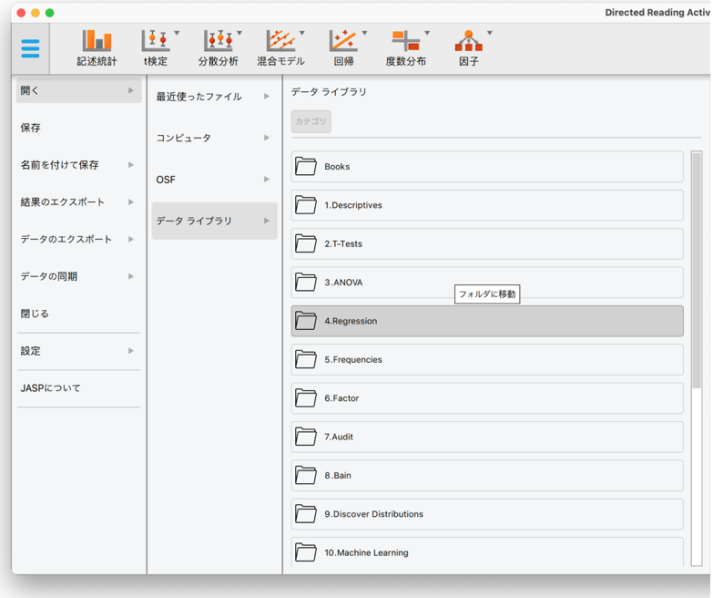


図 28.15　サンプルデータも豊富にある

図 28.16　伝統的な相関分析

の数は増えることでそれを示しています[*12]

同じ分析をベイズ的にやってみましょう。メニューの中で回帰 → ベイジアン・相関と進むだけです。変数の選び方も同じ操作です。分析結果の追加オプションには，p 値ではなく**ベイズファクター**が入っています。ここで比較するモデルは非ベイズ的分析と同じで，母相関がゼロというモデル 0 と，ゼロではないというモデル 1 を比べて BF_{10}，すなわちモデル 0 にくらべてモデル 1 がどれほどデータに支持されるかという比率を表してくれます。先ほど有意で合った，神経症傾向と外向性の相関については 6.765e+12，すなわち

[*12] これが良くない風習ですね！ p 値が小さくても，相関係数が大きいことを意味しませんし，仮説の正しさを支持するものでもありません。

6765000000000 倍もモデル 1 を支持しています。ベイズファクターはこのように，時折とても大きな数字になることがありますし，事前分布を変えれば変わることも大きいので，とりあえず 0 ではないと思うぐらいでいいかもしれません。

それよりも，信用区間のところをチェックしてみましょう。95%確信区間が [-0.270,-0.424] ですから，相関係数が-0.270 よりも大きくなる (0 に近づく) ことはほぼない，2.5% ぐらいしかあり得ないだろう，ということがわかります。**ROPE** をどの程度に置いておくかによりますが，「ほぼ無相関と同じレベル」とは言えないと判断すべきでしょう。ちなみに従来法での信用区間は，[-0.271,-0.425] です。ベイズ確信区間とほとんど同じですね。ただしこちらの場合は，この区間の中に真の相関係数が含まれる確率が 95%，言い換えるなら「この範囲に相関係数があると言って間違える確率が 5%」という意味です。

28.4 平均値の比較

それでは最後に平均値の比較をしてみましょう。メインメニューから，開く → データライブラリ →2.T-Tests→ と進んで，一番上にある Directed Reading Activities を選びます。実験群と統制群に分かれて，さを検証する実験計画のようです。メニューから t 検定 → 伝統的·独立したサンプルの t 検定を選びます。そこで，従属変数に唯一のスケール変数 `drp` をいれ，グループ化変数に `group` を投入します。検定は等分散性を仮定したスチューデント法と，その過程をしないウェルチの方法，両方を選んでみましょう。対立仮説は二群に差がある (グループ 1≠ グループ 2) とし，仮定のチェックや効果量など，適当に選んでみましょう。結果として，表 28.1 が示されます。ここで，分散が等しいという仮定の場合は Student の行をみて，

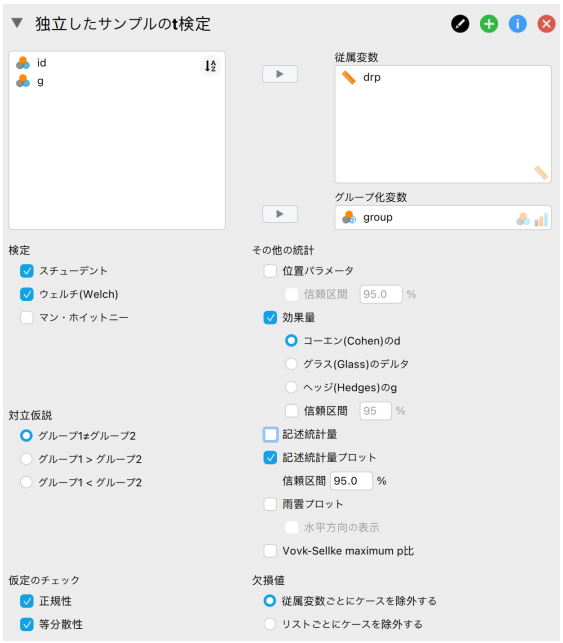


図 28.17 伝統的な t 検定

$t(42) = -2.267, p = 0.029, d = -0.684$ であることが読み取れます[*13]

補正が必要な場合は，Welch の方を見ましょう。このとき $t(37.855) = -2.311, p = 0.026, d = -0.691$

[*13] t 分布はペンネーム Student さんが書いた論文が元になっているので，異分散のための補正のない t 検定は Student の t 検定，と言います。ちなみに著者は William Sealy Gosset であることが明らかになっています。彼を雇用するギネスビール社が本名仕様を許さなかったため，ペンネームで論文を投稿したと言われています。

であるということがわかります。

表 28.1 独立したサンプルの t 検定

	検定	統計量	df	p	コーエン (Cohen) の d
drp	Student	−2.267	42.000	0.029	−0.684
	Welch	−2.311	37.855	0.026	−0.691

ではどういう時に Student 行をみて，どういう時は Welch 行を見ればいいのでしょうか。基本的に仮定が緩い Welch のほうをみればよいのですが，分散が同じと見做せるかどうかは「仮定のチェック」を見れば良いでしょう。仮定のチェックには，そもそもの前提である「データは正規分布に従っているのか」を検定する，シャピロ・ウィルク検定があります (表 28.2)。つづいて分散が同じと見做せるかどうかを検定するレヴィーン検定 (表 28.3) があります。

ここで注意して欲しいのは，これらの検定は帰無仮説が「正規分布に従っている」「分散が等しいといえる」であり，対立仮説は「従っていない」「等分散ではない」になります。つまり，ここでは分析結果が有意になると，「従っていない」「分散が等しくない」ということになります。シャピロ・ウィルク検定で有意になると，そもそも t 検定をしてはいけません。レヴィーンの検定で有意になると，分散が等しいという仮定が破られているわけですから，Welch の方を見るべきです。

表 28.2 正規性の検定 (シャピロ・ウィルク)

		W	p
drp	Control	0.972	0.732
	Treat	0.966	0.652

今回は幸い，レヴィーンの検定で $p = 0.132$ でしたから有意ではなく，Student の t 検定でいいということになります。それにしても少し面倒ですね。仮定のチェックのところでは帰無仮説を応援し，本番の検定で

表 28.3 Levene's の分散の等質性の検定

	F	df	p
drp	2.362	1	0.132

は対立仮説の方を応援する，という形になってしまいます。帰無仮説を積極的に支持できないじゃなかったのかよ，と言いたくなります。なにより検定を繰り返して行っていますから，α 水準が 5% ではなくなっているかもしれません。

そんな不安も，ベイズ統計の場合は必要ありません。早速ベイジアン分析をしてみましょう。今度はメニューから t 検定 → ベイジアン・独立したサンプルの t 検定を選びます。従属変数やグループ化変数の選び方は同じです。

結果は表 28.4 のように表されました。$BF_{10} = 2.217$ ですから，差がないという帰無仮説 (モデル 0) に比べて，差があるという対立仮説 (モデル 1) は 2.2 倍ぐらい支持されています。どちらかというと対立仮説の方が優勢，という感じですが，BF の 1 つの目安は 3 を超えるぐらいというのがありますので，「差があった」と断言するにはちょっと気が早いかもしれません。

表 28.4 ベイジアン独立サンプルの t 検定

	BF_{10}	エラー %
drp	2.217	0.002

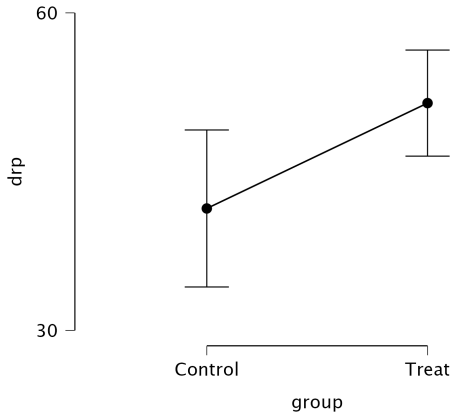


図 28.18 二群の平均値と確信区間

このことは記述統計的なプロットを見るとはっきりします (図 28.18)。これをみると，実験群と統制群の 95% 確信区間は重複しているところが結構ありそうです。重複している場合は，ROPE がどれぐらいであるにせよ，性急な判断はやめて保留し，データ (エビデンス) をもっと追加してから考えた方がいいでしょう。

ここでもう 1 つ，「事前分布と事後分布」の出力を見てみましょう (図 28.19)。

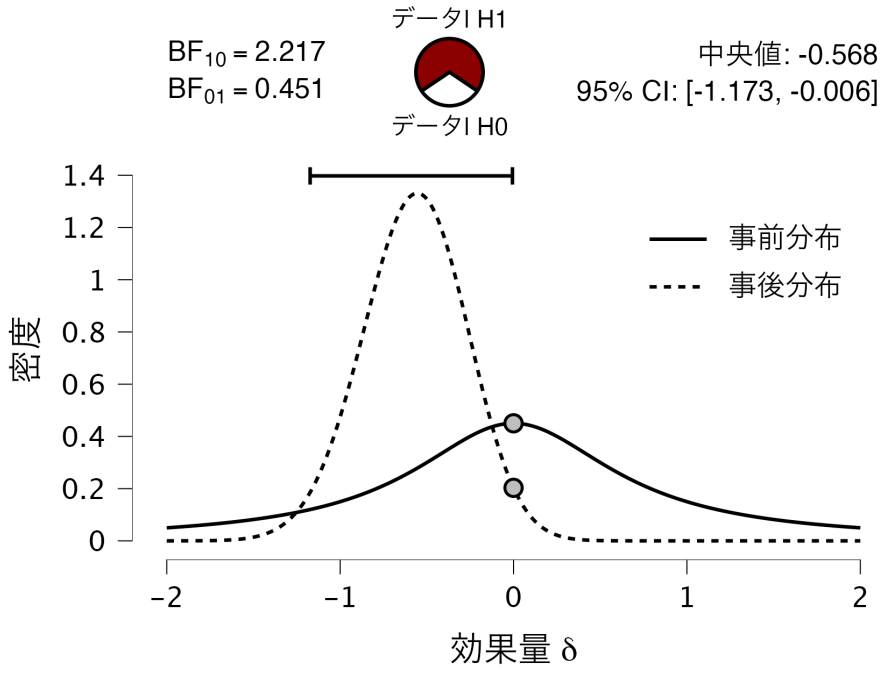


図 28.19 事前分布と事後分布

これは効果量 d の推定値 δ の事前分布と事後分布を示しています。効果量は標準化された平均値の差を表しており，事前分布では 0 を中心としたコーシー分布が置かれています。つまり，検定前の主観確率として

は「差はないんじゃないかな」とうっすら思っている，という状態です。これにデータ (尤度と事前確率) から計算された事後分布を重ねると，$\delta = 0$ の密度が事前分布よりも低い位置にきていますね。この事前分布と事後分布の，特定の点についての密度比はベイズファクターに一致することが知られており，この方法でベイズファクターを計算するのは**サヴェージ・ディッキー法**と呼ばれています (リー・ワゲンメーカーズ, 2013 井関訳 2017)。ベイズファクターの計算には，周辺尤度の積分が必要でなかなか難しい，というところだったのですが，帰無仮説のような点の仮説と，それを包含する対立仮説の関係にある場合，この事前分布と事後分布の密度比が利用できるのです。

ともかくこれを見ると，ちょこっと下がっただけ。また δ の確信区間は [-1.173,-0.006] です。上限がほとんどゼロですね。ゼロというのは差がないということですから，今回のデータを持って差があるよ，というのはちょっと厳しいかもしれません。

ここで注意を 1 つ。従来法と**ベイズ法**の両方の結果を見て，あれ？従来法なら有意差ありで，ベイズ法ならそこまで強く言えないんなら，従来法だけレポートに書いたいいんじゃない，と思った人。そういう「ちょっと自分にまずいことは隠しておこう」という態度は科学的誠実さには欠けますよ。今回はいずれも対立仮説の方が支持されていますが，もっと微妙な差であればもしかすると，従来法とベイズ法の結果が異なることがあるかもしれません。そのような場合でも今回の場合でも，両方の結果を記載して「このデータが本当に伝えていることは何か」をしっかり考えることの方が重要です。自分に都合の良い結果だけ利用するのではなく，差があるかないかという短絡的な判断で論じるのではなく，信頼区間や確信区間，p 値も効果量もベイズファクターもすべて記載して，現象の正しい理解をすることの方が重要なのですから。

ともかく，JASP というソフトを使うと，簡単に両方の手法が比較できることがわかっていただけたかと思います。今後は，t 検定と BF の併記のような形で，分析結果を報告する研究例が増えてくるでしょう。皆さんも JASP を通じて，さまざまなデータの分析を手軽に試してみてください。

第 29 章

R によるベイズ推定

JASP による分析は比較的簡単で，わかりやすかったのではないでしょうか。とはいえ，それはユーザーインターフェイスのおかげかもしれませんね。すなわちクリック，タップ，ドロップなどで視覚的に操作できる GUI のおかげで，簡単便利とおもえるわけです。もちろんそれを否定するわけではありませんが，GUI の欠点として，やったことを他人と共有する難しさがあります。「これをこうして･･･」と動きを言葉で説明するのは難しいことです。共有できないことは再現できないことにつながりますし，ひどい場合は「機械が勝手にやったので，自分では何をやったかわかりません」ということにもなりかねません[*1]。その点，R など CUI で実行指令書を残しておくやり方は，一見面倒なように思えるのですが，記録が残ることと再現可能性の高さから，研究教育の面では向いています。

さて，R でベイズ推定できるのかと言えば，当然それは可能です。今回はその方法について，演習的にやっていきましょう。と言っても，ベイズ対応のパッケージを使うだけなのですが。簡単なモデルであれば確率分布の計算から解析的に答えを出すことはできますが，今後のことも考えて MCMC 法による方法を説明します。これは R の `brms` パッケージ (Bürkner, 2017, 2018) で提供されています。R パッケージの `brms` を使うと，一般線形モデルが簡単に実装できます。関数名が `lm` から `brm` に変わるだけで，MCMC をつかった線形モデルのベイズ推定を行うことができ，またその結果の可視化も簡単に行えます。`lm` 関数と同じく，説明変数を離散型 (factor 型) にしておくと，要因の効果の大きさが推定でき，`brms` パッケージの `conditional_effects` 関数を使うと要因ごとに周辺化された結果も図示されます。

29.1 環境の準備

まずは環境の準備です。いつもの通り，次の 3 つのステップを経てください。

■RStudio の起動 RStudio を起動してください。パッケージのインストールをするときなど，管理人権限が必要になるケースがあります。

■プロジェクトを開く プロジェクトは前回作成した物で良いと思います。同じフォルダの中で，スクリプトファイルだけ変えれば良いでしょう。メニューからプロジェクトを開き (FIle→Open Project)，RStudio の右上などで，プロジェクトが開いていることを確認しましょう (→ 図 22.1)。

[*1] 実際，GUI の統計ソフトウェアで指導していると困るのがここで，自分では気づかない，デフォルトで設定されている条件でぶんせきしているとか，自分がどういう動きをしたかがわからないので，統計の原理や仮定を違えておかしな分析を無自覚にやっている，ということにもなります。

■R スクリプトを開く 今日のコードを書くための R スクリプトファイルを準備します。File > NewFile > R Script と進み，何も書いてない R スクリプトの画面を表示させてください。真っ白いファイルが開いたら問題ありません。

続いて，ベイズ推定用のパッケージを準備します。brms パッケージ[*2]をインストールするのですが，このパッケージをインストールする**前に**，このパッケージが必要とするコンパイラを用意する必要があります。詳しくは https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started-(Japanese) のサイトに説明がありますが[*3][*4]，Windows ユーザの場合はまず，code:29.1 を実行してください。

code：29.1 必要なツールをインストールする

```
1 pkgbuild::has_build_tools(debug = TRUE)
```

コンパイラである Rtools が環境にない場合には，ポップアップウィンドウが出てきて Rtools をインストールするか聞いてくるでしょう。Yes をクリックして 1 分ぐらいインストールが終わるのを待って下さい。

Mac ユーザの場合は，https://github.com/rmacoslib/r-macos-rtools にいき，Release page から macos-rtools-4.0.0.pkg を選択してインストールします[*5]。

これで準備 OK で，これではじめて brms パッケージをインストールできるようになりました[*6]。関連する複数のパッケージを次々取り込んで行きますので，少し時間がかかるかもしれません。RStudio のパッケージのところから brms を入力してインストール，あるいは R のスクリプトとして code:29.2 のようにすればインストールが始まります。

code：29.2 パッケージを入れる

```
1 install.packages("brms")
```

この作業は環境整備なので一度だけでよく，無事にインストールできれば今後はパッケージを呼び出すだけでよくなります。

29.2 回帰分析の推定

それでは回帰分析の推定をしていきましょう。使うデータは第 9 講で使った高校・大学の成績についてのものです。復習がてら，code:29.3 実行しましょう。

code：29.3 回帰分析をやってみる

```
1 # データを入力
2 koukou <- c(2.13,2.42,2.26,3.87,3.90,2.43,3.44,
3 2.15,2.18,3.00,3.42,2.55,3.19,3.05,2.52)
4 daigaku <- c(460,500,473,620,690,512,582,
5 550,485,650,593,528,585,569,518)
6 # データフレーム型に組み上げる
7 seiseki <- data.frame(koukou,daigaku)
```

[*2] ベイズで (Bayesian)，回帰モデルを (Regression Mode) 実行する，Stan を使ってね (using Stan)，という意味です。Stan とは**確率的プログラミング言語 (stocastic programming language)** の一種で，要するに乱数発生器を作ってくれるソフトウェアのことです。

[*3] 「RStan Getting Started (Japanese)」でググると出てきます。

[*4] 馬場 (2019) にも導入についての詳しい説明があります。

[*5] 2020 年 11 月 28 日段階での状況です。

[*6] この環境構築は初心者にはとっても荷が重いところですね。困ったらいつでも連絡してきてください。どうしても難しい場合は，サーバが利用できますのでこれまた連絡してください。

```
 8
 9  # データは図にする
10  ggplot(seiseki,aes(x = koukou, y = daigaku)) +
11      geom_point() +
12      geom_smooth(method = "lm", se = F)
13
14  # 線形モデルはlm関数。従属変数~説明変数の形で書く
15  fit <- lm(daigaku ~ koukou, data = seiseki)
16  # 結果をわかりやすく表示
17  summary(fit)
```

■コード解説

2-3 行目 高校と大学の時のデータを入力している

5 行目 それぞれのデータをデータフレーム型に組み上げている

8-10 行目 ggplot をつかって描画しています

13 行目 回帰分析を実行しています

15 行目 結果を出力しています

これの結果は次のようなものでした (R **の出力** 29.1)。

R の出力 29.1: lm の出力

```
> summary(fit)

Call:
lm(formula = daigaku ~ koukou, data = seiseki)

Residuals:
    Min      1Q  Median      3Q     Max
-31.425 -21.896  -6.906  -1.207  80.110

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   288.74      43.00   6.715 1.44e-05 ***
koukou         93.72      14.85   6.313 2.69e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 34.36 on 13 degrees of freedom
Multiple R-squared:  0.754, Adjusted R-squared:  0.7351
F-statistic: 39.85 on 1 and 13 DF,  p-value: 2.688e-05
```

回帰係数は切片が 288.74, 傾きが 93.72 と推定され, それらが 0 ではないと帰無仮説に対する検定結果が示されています。またモデル全体の適合について, $F(1,13) = 39.85, p = 0.000002688$ という結果や, $R^2 = 0.754$ などが示されています[*7]

[*7] 少し補足。p 値は 2.688×10^{-5} という意味です。桁数の表示が長くなりすぎるので, 機械はこのような省略記号を書きます。また適合度を表す R^2 の値ですが, ここでは Multiple R-squared の値をそのまま書いています。この値はサンプルサイズによって過大評価されることがあるので, 論文などではそれを補正した「自由度調節済み R^2」が使われることもあります。

早速これをベイズ推定していきましょう。何がどう変わるのでしょうか。まずはパッケージ brms を読み込み，回帰分析の関数を lm から brm に変えるだけでオッケーです (code:29.4)。

code：29.4 ベイズ推定してみる

```
1 library(brms)
2 fit.brm <- brm(daigaku ~ koukou, data = seiseki)
3 print(fit.brm)
```

画面にさまざまなメッセージが出力されたかと思いますがそれは後で説明するとして，とりあえず最後の 3 行目を実行して示された回帰分析の結果を確認してみましょう。

一行目に Family: gaussian とあります。これは正規分布 (gaussian curve) を使ったモデルであることが示されています。2，3 行目にモデルの説明，4 行目にデータの説明があった後，6 行目には Samples: とあります。これはデータ，標本のことを指しているのではなく，MCMC 法によって得られた乱数のサンプルのことを意味しています。MCMC の表記等については後でまとめて説明します。

また，MCMC による結果は毎回乱数を発生させて計算させています。ここに示した結果は私の環境での結果であって，皆さんがやってみると少し違った数字が出てくるかもしれません。というのも，パソコンがかわると生成される擬似乱数も変わってしまうからです。MCMC 法はあくまでも近似計算ですから，厳密に一致するかどうかを考えるのではなく，だいたいあっているかどうかで判断してください[*8]。

係数などは Population-Level Effects というところに書かれているもので，切片 (Intercept) の推定値 (Estimate) が 288.37，説明変数である高校の成績にかかる傾きが 93.86 であることが示されています。続いて標準誤差，95% 確信区間の下限 (lower)(l-95%CI)，上限 (upper)(u-95% CI) が示されています。検定ではなく，95% 区間を見てそれが 0 を跨いでいるかどうかで，検定と同じような判断ができるということですね。

最後にもう 1 つの推定値，σ が Family Specific Parameters のところに現れています[*9]。

R の出力 29.2 が基本的な結果の情報です。lm 関数を使った最尤推定の結果とほとんど同じ数字になっていることがわかります。結果はそれほど変わらないのですが，推定値が分布として得られていることが大きく異なる点です。

[*8] そんな適当でいいのか！と思うかもしれません。でも乱数を大量に発生させると，ほとんど同じ結果になります。また生成する乱数の数を 10 倍にすると，一桁精度がアップします。それでも完璧に一緒，というわけにはいきませんが，そんな場合は擬似乱数の開始点 seed を固定することで，再現可能な乱数を得ることができます。

[*9] 線形回帰は $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma)$ で，未知数は β_0, β_1, σ の 3 つであったことを思い出してください。

R の出力 29.2: brms の出力

```
Family: gaussian
Links: mu = identity; sigma = identity
Formula: daigaku ~ koukou
 Data: seiseki (Number of observations: 15)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup samples = 4000

Population-Level Effects:
        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   288.37     46.68   196.11   378.93 1.00     3212     2568
koukou       93.86     16.14    62.44   126.07 1.00     3167     2482

Family Specific Parameters:
    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    37.23      7.86    25.26    55.84 1.00     2746     2839

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

このことをわかりやすく表示するために，`plot(fit.brm)` としてみましょう。図 29.1 のように，分析結果がプロットされます。この図の左側にあるのが，事後分布の形です。推定値である 288.37 や 93.86 は分布

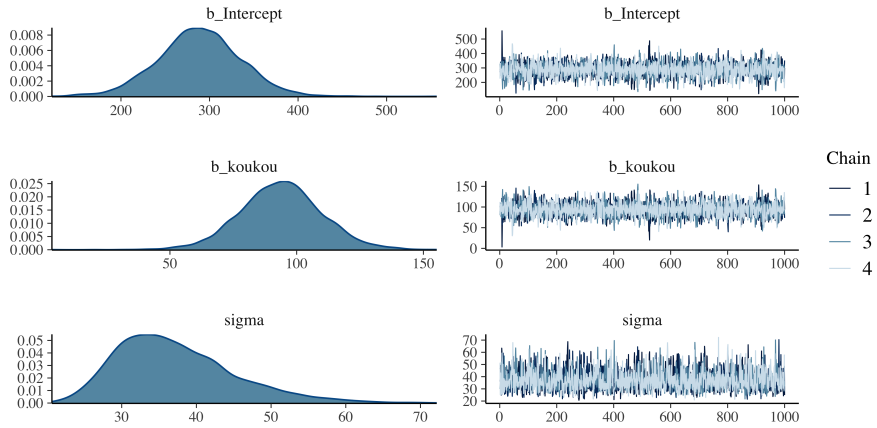


図 29.1 ベイズ法による回帰係数の推定

の平均値 (**EAP 推定値**) で，この分布のピーク (MAP 推定値) ではないことに注意してください。MAP 推定値が欲しい場合は，別途 `bayestestR` パッケージ (Makowski et al., 2019) を導入し，`map_estimate` 関数を実行する必要があります (`code:29.5`)。

code：29.5 結果の表示

```
1  install.packages(bayestestR)
2  library(bayestestR)
3  map_estimate(fit.brm)
```

■コード解説

1 行目 bayestestR パッケージがない人は，この 1 行でインストールできます

2 行目 パッケージを読み込み

3 行目 map_estimate 関数に分析結果を投入

最小二乗法や最尤法と違って，結果が分布で表されるという感覚が掴めてきたでしょうか。続いてはこの分布が乱数によって生成されていることに注目してみたいと思います。

29.2.1 MCMC 法の評価

既に述べたように，MCMC は乱数発生法であり，事後分布からのサンプリングでもって結果のヒストグラムを描くのでした。ヒストグラムが確率分布の稜線を描いてくれるからです。ただし，乱数ですから毎回違う結果が出ます。乱数がちゃんと事後分布からのものであるかどうかも，チェックしなければなりません。それらの情報について書いてあるのが，結果の 6 行目のあたりにあるのです (**R の出力** 29.3)。これを読み解いていきましょう。

R の出力 29.3: MCMC の評価

```
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
    total post-warmup samples = 4000
```

MCMC 法はステップ・バイ・ステップで，一歩ずつ次の乱数を生み出していってます。この MCMC サンプルは，推定したいパラメータの代表候補たちです。内部に保存されているデータを抜き出したのが図 29.2 です。今回は切片 (b_intercept) と傾き (b) と誤差の散らばり度 (sigma) について，毎回のステップで得られた乱数の値が書かれています。たとえば傾きは $77.55104 \to 100.42599 \to 133.46551 \to 77.90979 \to \cdots$ というように，前回の値を軸にしながら次々と乱数を生み出していってるのです。正確には，ここには 3 つの変数があり，赤枠で囲んでいるようにこの 3 つの変数からなる同時確率空間からサンプリングしてきます。つまり $(336.0793, 77.55104, 32.49790) \to (283.5139, 100.42599, 41.52797) \to \cdots$ というように，3 つずつセットでステップを踏んでいることになります。これがどんどん積み重なっていく，ステップの様子を示したのが図の中頃にあるギザギザの折れ線であり，これをまとめ上げて積み重ねたヒストグラムが最後に事後分布として得られる，というのが基本原理です。

これを踏まえて，出力されたものを見てみましょう。まず 4 chains とあります。これは MCMC のステップを 4 つ並行して実行したことを意味しています。MCMC はステップ・バイ・ステップで，一歩ずつ次の乱数を生み出していってるわけですが，どこからスタートするかによって全然違う答えにたどり着いてしまう可能性がゼロではありません。ステップの途中でも，もしかすると変なところにフラフラ外れていってしまっているかもしれません。こうした懸念をチェックするべく，1 つのマルコフチェインではなくて複数のチェインを同時に実行して，「どのチェインも同じところに行き着くよ」ということを持って，正しくサンプリングがなされたと判断するのです。

次に iter とあります。これは総ステップ数で，今回は各チェイン 2000 ステップしたことを意味します。次に warmup は，ウォームアップ期間とかバーンイン期間と呼ばれるものです。チェインの最初の方のステップは，初期値に依存したステップになっており，事後分布の代表値としては適切でないかもしれません。事後分布をうまく代表するように，最初の期間は採用しないという方法が取られており，それが最初の 1000 ステップであること記録されています。各チェインについて 2000 ステップ，うち 1000 ステップを捨てるので 1000 ステッ

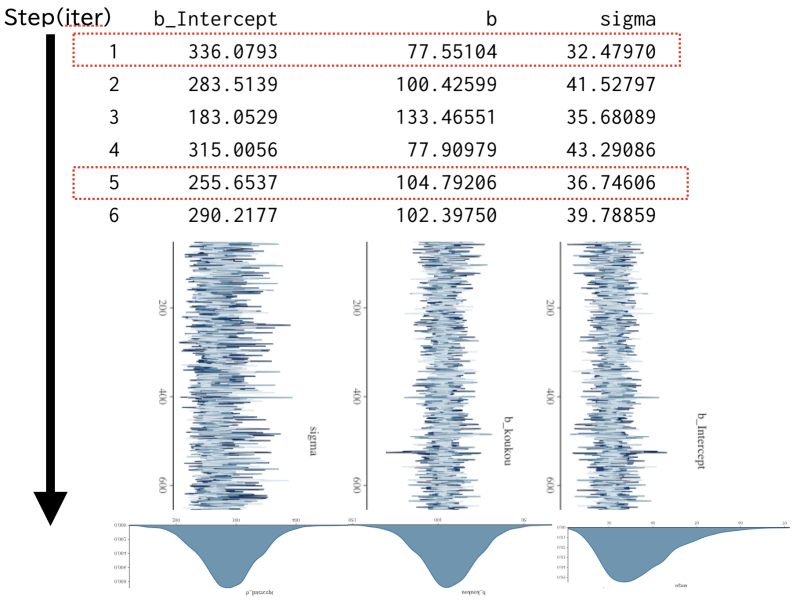


図 29.2 MCMC のやっていること

プだけ採用されますが，それが 4 チェインあるので，合計で 4000 サンプル。4000 点のヒストグラムで事後分布を描画したことになります。 `total post-warmup samples = 4000` というのはそういう意味ですね。最後に `thin=1` とありますが，これは間引きといわれ，サンプルとして採用するステップの間隔を意味します。今回は 1 ですので毎回のサンプルを (1step ずつ) 採用していきますが，これが 2 とか 3 になると，2 ステップ，3 ステップごとにしか結果として保存しないということになります。

図 29.3 をみてください。この図はトレースプロットといい，先ほどの図 29.2 の中頃にあったもの，あるいは `plot` で出力させた図 27.2 の右側にあったものです。ステップ数を横軸にとりながら鎖のステップを線で結んで，鎖がどのあたりをウロウロしているかを可視化しています。図の上には，複数のスタート地点から始まっても，途中で鎖が絡まりあって同じような領域をウロウロしている (サンプリングしている) ことが示されています。一方，下の方の図では鎖がほぐれています。下のような例は，サンプリングがうまくいっていないことの証左で，モデルが悪かったか，パラメータを特定する条件が不十分であったか，といった何らかの問題があったことを示しています。このように，結果が出力された後で MCMC の挙動をチェックすることが必要です。出力画面には最後にチェインの絡み具合をし評価した $\hat{R}$(Rhat) が 1 であれば収束している，ということを書いていて[*10]，今回の結果の Rhat の列にはいずれも 1.00 とありますから，MCMC は問題なく事後分布からの代表値を取り出している，と診断できます (R の出力 29.4)。

R の出力 29.4: MCMC に関するメッセージ

```
and Rhat is the potential scale reduction factor on split
chains (at convergence, Rhat = 1).
```

[*10] ひとところに集まってきているという意味の converge を収束，といいます。コロナウイルスの蔓延が終わる，という意味でのしゅうそくは「終息」と書きます。

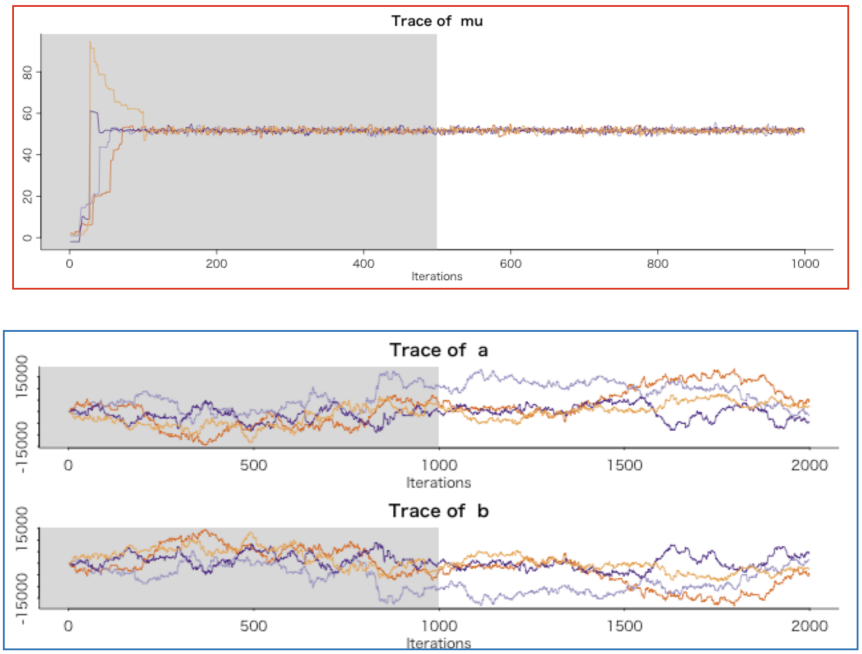


図 29.3 推定がうまくいってない時の例

29.3 要因計画の推定

続いて要因計画の例にいきましょう。既に線形モデルで推定する方法については説明してありますから，これの関数を lm から brm に変えるだけでオッケーです。二要因の例を使って説明してみましょう (code:29.6)。

code : 29.6 データを作って brm でベイズ推定

```
1  Example2 <- data.frame(
2      id = 1:12,
3      num = rep(1:3, 4),
4      temp = rep(1:2, each = 6),
5      maker = rep(rep(1:2, each = 3), 2),
6      value = c(13, 11, 12, 7, 6, 8, 9, 9, 9, 13, 11, 9)
7  )
8  Example2$temp <- factor(Example2$temp,
9                          labels = c("Hot", "Cold")
10 )
11 Example2$maker <- factor(Example2$maker,
12                           label = c("A", "B")
13 )
14
15 result.brm2 <- brm(value ~ temp * maker, data = Example2)
16 result.brm2
```

■コード解説

1-7 行目 データフレーム形にデータを組み上げています

8-13 **行目** 説明変数の中で名義尺度水準のものを要因型であると設定しています。

15 **行目** 線形モデルを実行しています。brm としてあることに注意して下さい。

16 **行目** 結果を出力させています。

これを実行した結果が次のようになります (R **の出力** 29.5)。

R の出力 29.5: 交互作用モデルの brms 結果

```
Family: gaussian
Links: mu = identity; sigma = identity
Formula: value ~ temp * maker
 Data: Example2 (Number of observations: 12)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup samples = 4000

Population-Level Effects:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept          12.01      0.84    10.34    13.65 1.00     1742     2016
tempCold           -3.04      1.21    -5.55    -0.68 1.00     1554     1583
makerB             -5.02      1.19    -7.40    -2.63 1.00     1622     1915
tempCold:makerB     7.05      1.65     3.69    10.42 1.01     1475     1563

Family Specific Parameters:
    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.42      0.42     0.85     2.49 1.00     1290     1796

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

データの生成に正規分布を仮定していること (Family: gaussian), モデル式や MCMC サンプルの特徴を描いた後に, 推定値が出ています。推定値の見方は lm の時と同じですが, 95% 区間が示されていること, MCMC の収束を判定する Rhat などがあるところが違うところですね。この 95% 区間が狭いと自信を持って結果を言える, 広いと確信度が低くまだまだデータが足りない, というところでしょうか。また今回は主効果にあたるところの 95% 区間が 0 を跨いでいるので, 傾きが 0 であるという帰無仮説を考えるのであれば, これは棄却できないということになります。交互作用は少なくとも 3.69, 大きければ 10.42 ポイントも動かすのですから, これは大きな効果があると言っていいでしょう。

このように直接的に大きさを評価することもできますが, 次のようなプロットを描くととくにわかりやすいでしょう (code:29.7)。

code : 29.7 周辺化して効果を見る

```
1  conditional_effects(result.brm2,effects = "temp")
2  conditional_effects(result.brm2,effects = "maker")
3  conditional_effects(result.brm2,effects = "temp:maker")
```

表示される 3 つの図 (29.4) はそれぞれ, 温度の主効果, 銘柄の主効果, そして交互作用であり, その EAP 推定値 (点) と 95% 確信区間がついています。これを見ると, 温度の主効果は確信区間に重複があり, 銘柄にそれはないこと, 交互作用も暖かい時の銘柄の違いが大きく, 冷たい時の銘柄の効果は薄いとか, 銘柄 B は温度ではっきりと違いがあるというようなことが目に見えてわかります。これらは帰無仮説検定ではありま

せんから，検定の多重性の問題がないため，新しい統計量を使って計算し直す必要もありません。非常に直感的に推論を進めていくことができるでしょう。

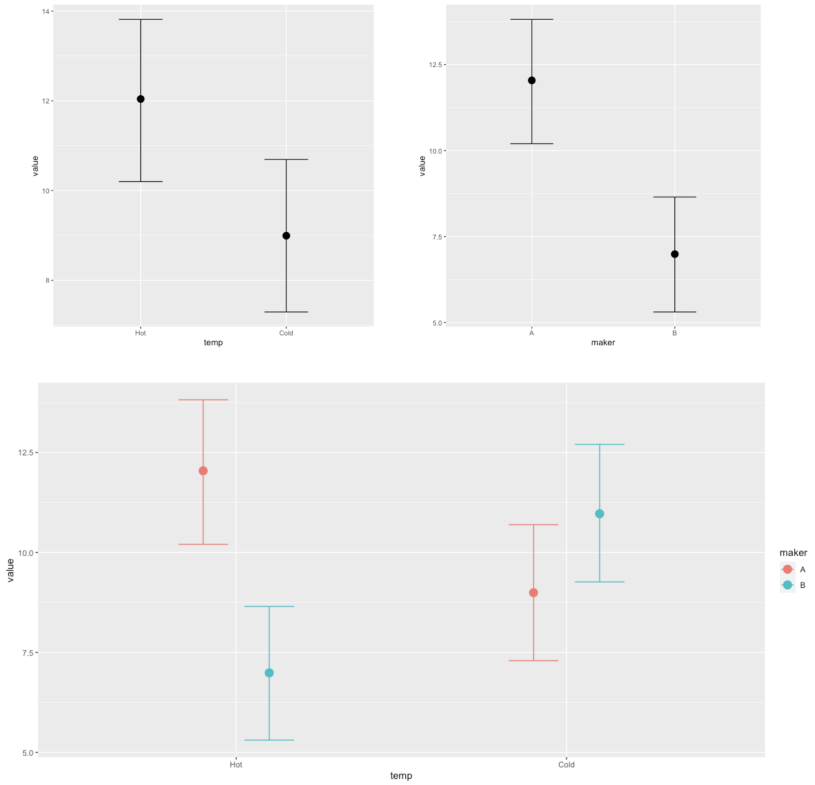


図 29.4 周辺平均値の推定結果

29.4 群内モデルの推定

最後に，群内のモデルについて解説します。第 22 講の，介入効果のモデルを思い出してください。群内要因がある線形モデルは，次のように表記できるのでした。

$$X_{ij} = \mu_i + \beta_0 + \beta_{ij} + \varepsilon_{ijk}$$

群間要因のモデルと違って，μ_i すなわち個人 i ごとの平均値が想定されていることがポイントです。これを確率モデルとして書くと，

$$X_{ij} \sim Normal(\mu_i + \beta_0 + \beta_{ij}, \sigma)$$

となります。平均にあたるところが線形モデルで表現され，誤差 ε が散らばることが正規分布の標準偏差 σ で表現されていることを確認してください。それ以外に違いはありません。ところで，個人差 i は介入前のスタート地点の状態を表しており，さまざまな値を取ると考えられますが，この個性の部分についても正規分布を仮定できます。つまり，

$$\mu_i \sim Normal(\mu_G, \tau)$$

のように，です (個人差の散らばりを表す標準偏差として，新たな記号 τ(タウ) を用意しました)。線形モデルで表現するには，このように正規分布の中に正規分布が入っている，という階層的な構造を考えることと同じなのです。これを `brms` で推定するには，次のようにコードを書きます (`code:29.8`)。

code : 29.8 データを作って brm でベイズ推定

```
1  Example3 <- data.frame(
2    ID = rep(1:4,3),
3    time = rep(1:3,each=4),
4    value = c(10, 9, 4, 7,5, 4, 2, 3,9, 5, 3, 5)
5  )
6  Example3$time <- factor(Example3$time,labels = c("time1","time2","time3"))
7
8  result.brm3 <- brm(value ~ time + (1 | ID), data = Example3)
9  result.brm3
10 conditional_effects(result.brm3)
```

■コード解説

1-5 行目 データフレーム形にデータを組み上げています。前回と同じデータなのですが，組み方が少し変わっている点に注意してください。確認のために，Example3 がどのような形になっているか一度表示してみるといいでしょう。

6 行目 説明変数の中で名義尺度水準のものを要因型であると設定しています。

8 行目 線形モデルを実行しています。brm としてあること，(1|ID) という項が増えていることに注意してください。

9-10 行目 結果を出力させています。

線形モデルの書き方が変わったことに注意してください。従属変数と独立変数をチルダで繋ぐのはこれまでと同じなのですが，これに加えて (1|ID) という表記があります。このカッコの中身は 1 が切片を表しており，それが変数 ID ごとに変わるよ，ということを意味しています。

これを実行した結果が次のようになります (**R の出力** 29.6)。

R の出力 29.6: ランダム切片モデル

```
Family: gaussian
Links: mu = identity; sigma = identity
Formula: value ~ time + (1 | ID)
 Data: Example3 (Number of observations: 12)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup samples = 4000

Group-Level Effects:
~ID (Number of levels: 4)
            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     2.33      1.15     0.69     5.20 1.00      970     1001

Population-Level Effects:
        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     7.36      1.30     4.79     9.83 1.00     1201     1112
timetime2    -3.99      1.00    -5.99    -1.86 1.00     2267     2017
timetime3    -2.00      0.98    -3.95    -0.02 1.00     2439     1736

Family Specific Parameters:
    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.31      0.51     0.70     2.58 1.01      957     1723

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

結果のところに大きく分けて，Group-Level Effects と Population-Level Effects の 2 つがあることみてください。群間要因の場合は Population-Level Effects の要素しかありませんでした。これは**固定効果 (fixed effect)** と呼ばれ，どの個人にも共通の効果があることを表しています。これに対して Group-Level Effects は**変量効果 (random effect)** と呼ばれ，人ごと，ケースごとに確率的に変わりうるところを意味しています。今回は最初の個人差 (スタート地点，切片) が人ごとに変わるので sd(Intercept) とあります。スタート地点が人によって 2 点ぐらいは左右するよ，ということがここでみて取れるのです。

固定効果の見方は，群間要因の時と同じです。第一期 (Time1) に比べて，第二期は平均点が -3.99，第三期は -2.00 となっています。この区間幅をみると，第一期よりも下がっているというところですが，第一期と第三期の下り幅はギリギリ 0 でない，というぐらいであってもう少し慎重にデータを追加して検証した方が良いかもしれませんね。

29.5 課題

第 22 講でやった 3 つの課題それぞれを，brms パッケージを使って実行してください。念の為，以下に簡単な問題設定を再掲いたします。

■課題 1 線形モデルとしての分析 ある学校の国語の授業で，ある教師が 4 つのクラスそれぞれに対して個別のテキストを使って授業を行い，同一のテストを行うことでテキストの効果があるかどうかを検証したとし

ます。各クラスには 10 人の生徒がいて，それぞれのテスト結果は表 22.4 の通りです。このデータについて，回帰分析のベイズ推定を行った上で，次の各問いに答えてください。

1. 切片の推定値はいくらになりましたか。おおよそで結構ですので，EAP 推定値を答えてください。
2. 回帰分析を行った結果，テキスト A にくらべてテキスト B の平均値はどれぐらい違うと思われますか。おおよそで結構ですので，95% 区間を答えてください。
3. 回帰分析を行った結果，テキスト A にくらべてテキスト C の平均値はどれぐらい違うと思われますか。おおよそで結構ですので，95% 区間を答えてください。
4. 回帰分析を行った結果，テキスト A にくらべてテキスト D の平均値に違いがあると断言しても良いでしょうか。conditional_effects の図をみて判断してください。

■課題 2 Between デザインの分散分析 別の学校では，テキストを A,B,C の 3 種類とし，教員が二人がそれぞれのテキストを使って教えることにしました。1 クラス 12 人，一学年 6 クラスを使っての実験です。授業後のテストの点数が表 22.5 の通りです。このデータに対して回帰分析のベイズ推定を行った上で，その結果について回答してください。

1. 切片の推定値はいくらになりましたか。おおよそで結構ですので，EAP 推定値を答えてください。
2. テキスト A,B,C の効果はどのように判断すれば良いでしょうか。conditional_effects の図をみて考察してください。
3. 教員 A,B の効果ははどのように判断すれば良いでしょうか。conditional_effects の図をみて考察してください。
4. 交互作用として，どのような特徴があるでしょうか。conditional_effects の図をみて考察してください。

■課題 3 Within デザインの分散分析 さらに別の学校では，テキスト･教員は一種類･一人なのですが，4 つの時期において成績が変化するかどうかを追跡調査しました。各時期のテストの点数は表 22.6 の通りです。次の各指示に従い，その結果について回答してください。

なお，このデータセットは Lesson8exe3.csv ファイルに入っていますが，分析に際しては次のコードを実行して，整理しなおしたデータセットを使ってください (code:29.9)。

code：29.9 データを外部ファイルから読み込んで実行する

```
1  library(tidyverse)
2  dat3 <- read_csv("Lesson8exe3.csv") %>%
3      pivot_longer(-id,names_to="period") %>%
4      print
```

■コード解説

1 行目 データファイル Lesson8exe3.csv を読み込んでいます

2 行目 id 列をキーにデータを縦型にしています

3 行目 データを表示させています

1. 個人差について正しく述べている文章を選んでください。
2. 時期の効果について正しく述べている文章を選んでください。

第 30 章

後期のまとめ

後期のテーマを振り返ってみます。

前期は記述統計，線形モデル (要因計画)，誤差としての確率が導入されました。後期は**確率**が中心的なテーマになっています。というのも**推測統計学**の領域に入ったからですね。前期の要因計画で，群平均の差を見ることで，介入の効果などが表現できることが論じられました。科学的実験はもちろん，カウンセリングなどの臨床的介入においても，その効果があったのかなかったのかを評価することは重要です。それが目の前の被験者，クライアントだけでなく，どれほど一般化できるかということが重要なのです。そのためには，目の前のデータに対して**母集団**と**標本**という区別をつけ，標本から母集団を推測する必要があるわけです。

標本から母集団を考えてものをいうためには，**推定**と**検定**という考え方を使います。

推定は，標本から母集団の状態を記述することを考えるものです。母平均，母相関，母比率など**母数**がどうなっているのかについて，手元の標本平均，標本相関係数，標本分布などからその数字を見積もるわけですね。方法としては，モーメント法，最尤法，ベイズ法の 3 つが考えられているのですが，モーメント法を使ったものが最も広く利用されています。

モーメント法の「モーメント」とは，積率という訳語もあるのですが，要するに平均，分散，歪度，尖度などのことです。標本平均や標本分散から，母平均や母分散を推定するというだけでも，少数のデータから全体像を推測するというなかなか難しいチャレンジをしているんだ，ということをまずは踏まえて欲しいところです。推測統計学というのはことほど左様に，そもそも問題を解決するためのヒントが少なすぎるという難易度の高いものでもあるのです。

標本分散は母分散に一致せず，ずれてしまいますので**不偏分散**という計算方法をしなければなりませんでした。標本平均がどの程度の散らばりをするかは，標本分の標準偏差，すなわち**標準誤差**を使って表現できますが，母 SD がわからないのであればそのままでは使えない。ということで不偏 SD と t 分布をつかって推定する必要があるのでした。**区間推定**はベイズ推定のそれとは確率の考え方が違っているので注意が必要ですが，要因計画で介入の効果を考えるのが，群平均に基づくものだったわけですから，これを使って効果の大きさを見積もるのだ，という方針をしっかり思い出してください。

もう 1 つの推測統計学の用法，統計的な**検定**というのは，こうした推定と判断がセットになった考え方です。差があるのかないのか，効果があるのかないのか，という判断をする必要があったときに用いられます。手続きは機械的に進められるカッチリしたもので，**帰無仮説**と**対立仮説**という 2 つの仮説による対決で勝敗を決するというものです。判断基準である**有意水準**や，**検定力**，確率的判断の正しさなど細々したルールがあるので，しっかりとその手続きについても理解しておいてください。また，相関係数の判断については t 分布，要因計画の判断については F 分布など，対応する統計量が何になるのかとか，それぞれの帰無仮説はどのように設定するのか，効果量はどのように見積もるのかといったことも，実践上は重要なポイントになってきます。

検定は誤用や悪用もあるのですが，そもそも母数を推定したい，科学的に判断したいというニーズに基づいていたことをしっかり思い出せば，こうした問題は幾らか軽減されるかもしれません。推定が目的であることを思い出せば，差があったのかなかったのか，モデルの勝敗はどうだったのかという一点にこだわりすぎることもなくなるでしょう。また「有意になりさえすればよい」「有意な検定結果が良い論文」という価値判断に陥ることもなくなるでしょう。科学論文は正しいかどうかが重要で，嘘をついてまで有意差を見出すことが目的ではないからです。こうした考え方は，モーメント法による有意差検定だけが母集団の推定方法だと思っていることが原因かもしれません。

最尤法やベイズ法では，確率モデルを元に母集団を推測する方法です。確率関数を直接用いることになるので，数式や計算方法がややこしいように思えてきますが，そもそも確率は「個々の事例はわからないけれども全体的な傾向はわかる」という偶然をハンドリングするための秘策でもあるのです。理論的にはこうなるはず，でも完全に制御できない要素がある，それがどの程度かを表現する，そういう方法が確率です。個々のデータは確率分布から生み出されているんだと考えると，得られたデータがどういう確率分布から出てきているんだろうか，ということが問題になってきます。得られたデータに合うように確率分布のパラメータを推定する，それが**最尤法**のやりかたなのでした。

尤度はデータが得られる確率から計算するもので，尤度関数自体は確率分布のように解釈することはできません。形は似ているのですが。これを確率のように解釈するには，ベイズのやり方を使う必要があります。注意すべきは，ベイズ法では確からしさ，確信の強さを確率分布で表現するという解釈に立っているということです。確率分布は信念の強さ，確率は**主観確率**というべきものとして理解した方がいいでしょう。であれば，母数がどのあたりにあると思われるかを表現するのがベイズ推定の結果であり，結果を表す**事後分布**はその名の通り，分布として区間の情報を持っているものになります。区間を使って推定する，判断するという枠組みは同じですが，帰無仮説検定のような架空の確率空間における p 値を見るのではなく，母数の分布そのものを見ることができるので，より直感的な理解が可能です。実践上は JASP や Stan など特殊な統計環境を必要としますから，まだ一般に広まった考え方とは言えませんが，今後ますます活用されてくることになるでしょう。

心理学における統計の使い方は，「心」はもちろん「実験・介入の効果」といった目に見えないものを，なんとか測定し，推定し，心理学的考察の補助とするためのものです。これから皆さんは，さまざまな専門的心理学の道に進んでいくことになりますが，そのどこにおいても方法論的基礎は必ず利用します。心理学を学べば学ぶほど，人間の直感的な理解や判断は間違えていることが多く，数学的あるいは理論的に正しい判断をするための補助的なツールが必要だということが身に染みてくると思います。そのための方法論ですから，決して蔑ろにすることなく，正しく身につけてもらいたいと思います。

どうしても数学的な表現をすることがありますので，その表面的な難しさに嫌気が差して目を逸らしてしまいたくなる気持ちもわかります。幸い，わからなければ立ち返って，わからなくなったところからやり直すことができるのが，数学のよいところです。私自身，間違えた理解や判断をしたことがないはずはなく，その度に学び直すということを繰り返しています。怖がらなくて結構ですから，「何のために統計分析をしているのか」を忘れずにこれからも学習し続けていってください。

姉妹講義であるデータ解析応用 (二年次配当，選択) の授業では，さらに進んだ理解，表現ができるようなことを学びます。基礎の授業では要因計画や帰無仮説検定など，決まり切った方法をなぞることを求めてきましたが，これから先は統計を駆使して，よりクリエイティブに表現することを考えます。応用まで履修していただくと，**心理測定法**の理論や心を表現するための新しいモデル，心のメカニズムを数式やプログラミングで表現できるようになります。これからの社会はデータサイエンティストが求められていますが，人とデータ，データとその表現について心理統計は一日の長がありますので，ぜひ履修してみてください。もちろんデータサイエンティストが流行らなくなった時代になったとしても，人間とは何か，心とは何かという問いについて，その

分析方法がもつ思想的背景からの観点を身につけておけば，人生をより豊かなものとして楽しむことができるでしょう。

付録 A

電子計算機のイロハ

A.1　前置き

このセクションは心理統計ではなく，コンピュータについての四方山話をダラダラと書いています。そんなの聞かなくてもわかってるよ，という人もいるかもしれませんし，知らなくてもみなさんはきっとスマートフォンやタブレットを使えていることと思います。しかし知らずに使うことと，知ってて使うことには大きな違いがありますし，今後大学でレポートや論文を書いたり，それに必要な統計処理をするためにも，計算機の基本的な特徴を知っておくと，トラブルに出会した時に「あぁこれってひょっとして」というヒントが得られたり，納得できるようになるかもしれません。知らなければ「何だかわからないけどパソコンが壊れた」というか，「パソコン運が悪い」「自分はパソコンが苦手なのだ」と間違えた帰属をしてしまうことになります。

コンピュータ関係の授業で聞いたことがある話，これから聞く話もあると思いますが，もし苦手意識を持っている人がいたらこれを機に再入門するつもりで読んでください。

A.2　コンピュータの基礎

21 世紀に生きる私たちは身の回りをコンピュータに囲まれて生きています[*1]。それは携帯電話の形をしていたり，ノートパソコン，デスクトップパソコンの形をしていたりします。また時々テレビなどで報道されますが，気象予報や飛沫がどのように飛び散るかをシミュレーションする大型計算機「富嶽」などもコンピュータですね。これらは形は違いますが，いずれも電子計算機であり，電子計算機には次の 5 つの装置があります。

入力装置　キーボードやマウス，タッチパネルなどを使って情報を取り込むデバイス (装置)

出力装置　モニタやプリンタ，タッチパネルなどを経由して情報を出力するデバイス

演算装置　プログラムの命令に従って計算処理 (四則演算や論理演算) をする装置。一般に電子計算機の中央で一括して処理するので，Central Processing Unit(CPU) と呼ばれるものです。

制御装置　演算結果に従って他の装置に指示を出す装置のこと。演算装置とまとめて CPU に実装されています。

記憶装置　計算結果などの情報をいったん保持しておく装置のこと。コンピュータの内部にあって一時的な

[*1] 私は 1976 年生まれですが，生まれた頃は周りにコンピュータなんかありませんでした。小学生の頃にマイコン (マイクロコンピュータ，小さなコンピュータという意味でもありますが，My Computer，私のコンピュータという意味でもあります。つまり個人単位でコンピュータが使えるようになった，というだけでも大きな出来事だったのです。) という言葉が出てきて，なんかかっこいいなと思った記憶があります。私が 10 歳になったころ，ビデオゲーム (テレビゲームでやるゲームが家庭でできるようになり，それをこのように呼びました。) が身の回りに出てきました。ファミリーコンピュータ，とくに「スーパーマリオブラザーズ」によって日本中の子供たちが熱狂したのが 11 歳の頃，「ドラゴンクエスト」によって社会問題になったのが 12 歳の頃になります。ともかくこの頃は，コンピュータといってもゲーム機のような扱いでした。

7873 計算に使われる一次記憶装置，ハードディスクドライブ (Hard Disk Drive,HDD) やソリッドステート
7874 ドライブ (Solid State Drive,SSD) などの二次記憶装置など。

7875 最後の記憶装置については，一次記憶装置，二次記憶装置と種類が分かれていますがこの区別は簡単
7876 で，電源を落とした時に記憶が消えてしまうのが一次記憶装置，電源を落としても記憶が消えないのが二次
7877 記憶装置です。たとえば $12 + 38 =$ という計算をするとき，頭の中で「えーっと一の位が 2 と 8 だから 10 に
7878 なって繰り上がるから･･･」と考えてから，ノートに $= 50$ という答えを書くと思います。次の問題に進むと，先
7879 ほどの「1 繰り上がるから･･･」という情報は忘れてますよね。でもノートに書いた $12 + 38 = 50$ というのは
7880 残っています。このノートがいわば二次記憶装置であり，頭の中で一時的に保持していた情報が一次記憶装
7881 置ということになります。一次記憶装置は RAM(Random Access Memory) とも呼ばれます[*2]。
7882 ところで，コンピュータがやっているのは計算だけです。私はこの資料を PC に向かって書いており，キー
7883 ボード (入力装置) を叩きながら，画面 (出力装置) を見て文字を連ねています。これも「キーが押されたら文
7884 字を表示させ，その文字列を記録する」という処理を機械が淡々とこなしているに過ぎません。あるいはマウ
7885 スやトラックパッドで，アイコンを指し示し，カチリと押す[*3]ことで選択し，押し込んだまま移動させ (ドラッグ)，
7886 離すことで別の場所に置いたりします (ドロップ)。トラックパッドの場合は 2 本指で同時に押したりしますし，
7887 タッチパネルの場合は二本指を広げたり (ピンチアウト)，逆に二本指を狭めたり (ピンチイン)，3 本以上の指
7888 でファサーっと触って場所を広げたりします。たとえば「ファイルを掴んでゴミ箱に捨てる (削除する)」というの
7889 が我々にとっての操作ですが，コンピュータの内部では実はこんなことをしていません。ファイルは (二次) 記
7890 憶装置に書き込まれた情報です。記憶装置は原稿用紙のように小さなマス目がたくさんあって，ファイルとは
7891 そのマス目の XXX 番目から YYY 番目までの情報，ということです。このファイルを削除するというのは，
7892 記憶装置のある場所 (アドレス) に「削除されたものなので画面に表示しない」という情報を書き込む，という
7893 操作をしているだけです。じゃあなぜ私たちは「ゴミ箱にドラッグ＆ドロップ」なんてするのでしょう？ それは
7894 そのほうがわかりやすいからですよね。「ファイルを削除する」というのは，「メモリアドレスの XXX 番地に別
7895 の情報を書き込む」という操作だと言われてもピンとこないので，コンピュータが人間にとってわかりやすい表
7896 現をして見せてくれているのです。このユーザにとってわかりやすい幻を見せてその気にさせてくれるという
7897 デザインのことを，ユーザーイリュージョンと言います。ともかくこういう「画面で見ながら操作する」ことをグ
7898 ラフィカル･ユーザ･インターフェイス (Graphical User Interface,GUI) といいますが，これのおかげでコン
7899 ピュータの操作は随分楽になっています[*4]。

7900 A.3 コンピュータの歴史

7901 コンピュータの装置，すなわちハードウェアについての解説につづいて，ソフトの側面についても解説を加
7902 えようと思うのですが，そのためには少し歴史的な流れを説明したほうがわかりやすいかもしれません。
7903 コンピュータの発展の歴史は，小型化の歴史でもあります。最初にできたコンピュータは ENIAC といい
7904 ます。1946 年の話です。この ENIAC は 27 トン，広さにして倉庫 1 つ分 ($167m^2$，90 畳以上の広さ) が

[*2] 実はこのように人間を 1 つのコンピュータに喩えて，そこではどのように計算がされているのか，人間の記憶装置や演算装置，入出力装置の特徴はどうなっているのか，というのを研究するのも心理学の仕事です。記憶や演算，制御については認知心理学や学習心理学，入出力については知覚心理学や生理心理学が専門的に扱っています。そういう意味でもコンピュータの登場は心理学に大きな影響を与えているのですが，それはまた別の講釈で。

[*3] 押すときの音から，この操作をクリック click と言います。2 回続けて押すことをダブルクリックと言います。

[*4] 実は人間の意識もこのユーザーイリュージョンのようなもので，実際の体の動かし方や感覚情報の受け止め方，処理の仕方は別ですよね。すべての情報を意識に上げるのではなく，「私は XXX をしている」と身体がそれっぽい幻想を投影して見せてくれているのが意識の正体ではないか，という議論があります。興味のある人は Norretranders (1999 柴田訳 2002) を読んでみてください。

必要なもので，真空管を使った計算機でした。軍事的な計画のために開発されたオーダーメイドのものですから，一般人が触れるはずがないものです。時代が下がって真空管が半導体，IC チップになった頃，やっとサイズが小さくなって，家庭用・個人用のコンピュータというのができるかもという時代が来ました。Apple コンピュータの創始者，スティーブ・ジョブズとスティーブ・ウォズニアックが最初のパソコン (Personal Computer,PC)，Apple I を発売したのが 1976 年。爆発的に売れた Apple II が発売されたのが 1977 年です。AppleII はブラウン管表示装置とキーボードを持っていましたので，今の PC の原型とも言えるかもしれません[*5]。PC を作っている会社は Apple だけでなく，IBM や DEC などがありましたが，まだこの頃はパーソナルなレベルのものよりも大型計算機の開発が進んでいました。IBM が PC を作ったのは 1980 年で，この頃から小型化が進められていきます。

私事で恐縮ですが，1976 年に生まれた私が初めて PC を手にしたのは 1991 年，高校入学のお祝いで買ってもらった Fujitsu の FM-Towns という機体でした。この頃は「マルチメディア」という名前もなく，「ハイパーメディア」と読んで売り出していました。この機体は他の PC(NEC の PC-9801 や Sharp の X68000 シリーズが有名でした) とは違って，CD-ROM ドライブをつけていたことが画期的だった時代です。その後 1994 年に大学生になりましたが，この頃は連絡を取り合うツールはポケベルが主流であり，携帯電話 (や PHS) のような個人端末は高級品という時代でした。大学に入るとコンピュータを学ぶ授業があり，アカウントをもらったりするのですが，それは大学が持っている大型計算機に端末からアクセスするためのものでした。いろんな部屋にあるのは「端末」で，それほど機能の優れた PC ではなく，複雑な計算 (統計的な計算など) は大型計算機に仕事を依頼しその返信を待つ，というスタイルでした。関西私立のマンモス校でしたので学生数は非常に多かったのですが，多くの学生が一度にアクセスしても，大型計算機はものすごくものすごーく計算が早かったので，瞬時に回答をもたらしてくれるものでした[*6]。つまり，まだ「専門的な計算は大型計算機」という時代であり，パーソナルなコンピュータになるにはもう少し時間が必要でした。

どれぐらいの時間が必要だったかというと，実はその次の年なのです。1995 年，Windows 95 という OS が発売されました。これを機に日本でも PC がどんどん浸透していくことになります。Windows95 は物凄いんだぞ，と発売前からテレビでも散々とりあげられ，発売日には行列ができて真夜中のカウントダウンと同時に大フィーバー，という売れ行きでした。当時のそれは何が凄かったのでしょう？ コンピュータにはそれ動かす基本ソフトが必要です。HDD にデータを書き込み，キーボードからの入力をディスプレイに表示する，といったごく基本的な装置を統括し，メモリ番地をファイルという単位で扱うと言った基本的な操作は OS いうソフトウェアが担当します (図 A.1)[*7]。この OS，大型計算機は Unix と呼ばれるものを使っていましたが，各企業が個人向けにコンピュータを売り始めるときにも当然必要で，各社で開発もしていましたが，PC の共通規格をつくることで OS 部分は共有できるようになりました。そこを提供したのが Microsoft 社のビル・ゲイツです。どんなパーツで作られた PC であっても Windows という OS が共通のフィールドを用意してくれるので，ユーザは Windows で動くアプリケーションを選ぶだけで良い，ということになったのです。

そして Windows95 は，GUI，つまり「ファイルを掴んでポイ」といった直感的な操作で使えることも大きな特徴でした。大型計算機で使われている Linux は基本的に Command User Interface,CUI で，黒い画面にプログラムを書いて実行するといった手法で，初心者には人気がなかったのです。GUI については，その頃 Apple を追放されていたスティーブ・ジョブズが，NeXT という会社で GUI を備えた OS を開発していました。この NeXT はその後 Apple に買収され，ジョブズは Apple に戻って活躍することになります。その頃から巷では，Windows の GUI は Apple の OS を真似したものだと非難されていたのですが，商業的に

[*5] Mac の歴史については Isaacson (1995 井口訳 2011) が読み物としておもしろいですよ。

[*6] Time Sharing System,TSS，時分割システムとよばれる機構です。命令を小さな単位に分割し，それを順次捌いていくという方法でした。

[*7] 物理的な機構と OS との間に Basic Input/Output System,BIOS というのが入りますが。

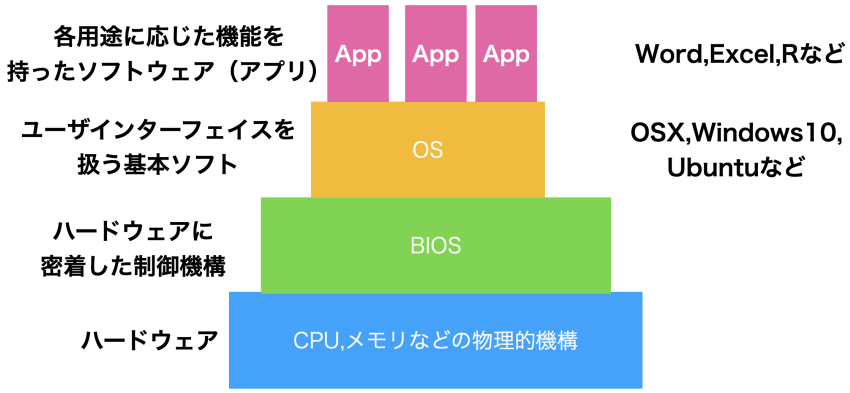


図 A.1 コンピュータのハードウェアとソフトウェアの関係

は Windows が大勝利，というわけです。ともかくこれを機に企業などはもちろん一般家庭でも PC を使うような時代になりました。

ちなみにインターネットが広まったのもこの頃です。私は大学 3 年生の時，大学の授業で初めてインターネットを介して世界の情報を得る，という経験をしました[*8]。その頃から徐々に，大型 PC のアカウントではなくインターネットで使えるメールアドレスというのを個々人が持つようになりました。携帯電話も廉価な PHS が広まり，メッセージだけでなく徐々に音声，写真，短い動画が遅れるようになっていきます。当初は当然画質・音質も今とは比べものにならないのですが，それでもインターネットを経由してつながるという経験は想像を超えたものでした。

皆さんはすでに，携帯電話やインターネットがある時代に生まれた世代だと思います。こんな話はすでに過去のものであり，不便な頃の話を聞かされても困る，と思うかもしれません。私の世代は，幸いにもこのようにちょうどコンピュータが使われはじめ，広がり，高性能になっていくにつれて育ってきていますので，そのぶん機械の根本的な理解に直結しやすい社会環境にあったのです。みなさんは，便利な時代ではありますが，言い換えると「初心者が苦労しないように補助輪をつけておく」「何もしなくてもできているような感覚が得られるようなイリュージョンを見せておく」という状態におかれているので，補助輪が対応できない道に進んだり，多くの人とは違う使い方をするときに，急にサポートが外れどこで困っているかわからなくなる，ということになりかねません。非常に大雑把な Historical Review ではありますが，理解の一助になればと思います。

ところで，先ほどサポートという言葉を使いましたが，このサポートを商売にしているのが Microsoft や Apple という IT 企業です。これらの会社は，ユーザが使いやすいようにソフトウェアを提供してくれますが，有償ですし想定外の使用をするユーザに対してはサポートをしてくれません。逆にいうと，「決められた路線を走らなければ面倒を見ない」ということでもあり，これはユーザに不自由を強いているとも言えます。また，どのような仕組みで動いているのかを尋ねても，企業秘密と言って答えてくれません。コンピュータは誰でも自由に使えるべきものであり，勝手にユーザの情報を盗んだりしていないか，と言ったチェックをするためにもオープンであるべきではないか。そういう考え方に基づく，フリーソフトウェアというソフトウェアのあり方があります。Unix という OS を PC 用にした Linux がそうであり，オフィスソフトの LibreOffice，統計環境の R などもこの精神に賛同するものです。フリーソフトウェアは自由であり，無償です。お金と秘密を払ってサポートを受けるのではなく，ユーザが相互に助け合ってオープンで自由な世界を広げていこうという活動です。急

[*8] 教職関係の科目で，教育技術として今後使われるだろうということで担当教員が実演してくれました。隣の部屋から電話線を延長し，モデムというパソコンの信号を音情報に帰る機器を繋げて，NASA の Web ページを Netscape というブラウザでみたのが初めての体験でした。

に何の話なんだ，と思うかもしれませんが，心理学ひいては科学的活動すべてにおいて重要な問題であることをご理解いただきたいと思います。

A.4 情報の単位

コンピュータを取り巻く世界の話はこれまでにして，ソフトの側面についての解説に入りましょう。

コンピュータは文字，音，絵，動画ファイルいずれについても，すべて 0 か 1 のデータとして管理します。0/1 の 1 つの単位を 1bit(ビット) と言います。1bit であれば Yes か No か，という二択の情報しか提供できませんが，これが 7 つあれば $2^7 = 128$ ですから，これで 128 種類の状態を表現できます。コンピュータは一般に，8bit で 1 つの単位として計算します。8bit のことを 1byte(バイト) と言います。この 1byte が 1024 集まったものを 1kb(キロバイト) といいます[*9]。1kb の次は，1024kb=1Mb(メガバイト) です。さらに 1024MB=1GB(ギガバイト) で，1024GB=1TB(テラバイト)，1024TB=1PB(ペタバイト)，1024PB=1EB(エクサバイト) と続きます。K,M,G,T,P,E といった名称は 1000 倍ごとに変わる大きさの桁をあらわしているのであって，「ギガが減る」というのは本来意味をなさない表現です[*10]。

このビット・バイトは情報に関する基本単位なのであちこちに使われます。記憶装置について使われるとき，一次記憶装置も二次記憶装置も同じ単位なので混同するかもしれませんが，2021 年現在では二次記憶装置の単位は GB から TB が使われます。USB フラッシュメモリーや外付け HDD などは数 TB の容量が一般的でしょう。これに対して，一次記憶装置は 4〜64GB ぐらいが相場かと思います。一次記憶装置は暗算の途中経過のように一時的に記憶する場所に過ぎないので十数 G でも問題ありませんが，二次記憶装置は結果の記録なので大きければ大きいほど余裕が持てますね。たとえば数 TB でもテレビドラマや映画を何本も記憶できるのですから，PB や EB なんて使うのかな，と侮ってはいけません[*11]。すぐにそれぐらいのサイズが必要な時代が来ることでしょう。

実際，それぞれ単位ではどれぐらいの情報が記録できるのでしょうか。1byte は 128 文字表現できるので，英語のアルファベット 26 文字に加え，数字や簡単な記号であれば 1byte で表現できます。たとえば A という文字は 01000001，B という文字は 01000010，小文字の a は 01100001，と言ったように 0/1 の文字列 8 個と一対一対応させて考えるのです。日本語は 1byte では足りませんので，一文字あたり 2byte が割り当てられます。また 1KB(=1024byte) は 500 文字ですから，原稿用紙一枚ぐらいになります。1MB(=1024KB) は文字だけだと新聞一紙 (朝刊の 40 ページ分) ぐらいで，昔の記録媒体であるフロッピーディスク一枚に保存できるのがちょうど 1MB でした[*12]。ちなみに「カメラ映像 + 音声」のオンラインビデオ会議を 1 時間やると 200〜300MB ぐらいの容量をやりとりしていることになります。ビデオ会議では文字 (チャット) だけでなく，画像 (動画) や音声も送り合いますね。実は画像や音声も，0/1 に置き換えています。音声の場合，1 秒間を短い間隔にくぎります。この区切りのことをサンプリング周波数といい，たとえば 44.1 kHz という単位は 1 秒間に $44.1 \times 1000 = 44100$ 点のデータの採取をします。この 6 データ点において，音の振動幅を区切ります。ビットレートと言いますが，たとえば 16bit で区切る場合は $2^16 = 65,536$ 段階で区

[*9] キロメートルやキログラムのように，キロは 1000 の単位を表す言葉ですが，コンピュータは 2 進数なので 1000 ちょうどではなく 1024 で 1 つ上の位に上がることになります。

[*10] 同様に「USB を紛失する」というのもよく聞くおかしな表現です。USB は Universal Serial Bus の略で，データ転送規格のことを指します。なくすことができるのはそれにつなげるメモリースティックなどです。

[*11] 私事ですが，私が 1991 年に生まれて初めて手した PC，FM-Towns はハードディスクが 40MB，RAM は 2MB で，CD-ROM がついていましたが，それは 360MB の容量でした。購入するときに，「ハードディスクが 40MB もあって何を記録するんです？ CD-ROM なんか情報が詰まり過ぎてますよ！」と店員さんに笑われたのを今でも覚えています。数年後，PC の動きが遅くなったので，2 万円で 2M の RAM を追加したのも良い思い出です。今でこそ，2MB なんて USB フラッシュメモリーでも売ってないほど微小なサイズですが。

[*12] 正確には 1.2MB 入る規格 (2DD) と 1.44MB 入る規格 (2HD) とがあります。

切り，その高さの音がある (1) かない (0) かで表すわけです。このように時間と音階を細かく区切り，その目に情報があるかないかを積み重ねてデータとするわけです。テキストよりも圧倒的に情報が多くなるのが分かりますね。画像も同様に，図面を細かい単位 (ピクセルなど) で分けて，そこの色合いを色々な段階で区切ります。色は R(赤)G(緑)B(青) の組み合わせで表現でき，それぞれを 8bit= 2^8 = 256 段階で表現したりします。一点一点にその情報がありますから，図の情報も非常に多くなるのがわかると思います。動画はその画像が時系列的に細かく分割されたものと思ってください。このように分解しますので，テキスト，音声，図，動画の順にデータサイズが大きくなります。通信機器が最初ポケベル＝数 byte の情報しか送れなかったものから，徐々に絵文字，ショートメッセージ (音声)，写真[*13]，動画が送れるように発展していきました。今では町中のあちこちで，誰もが手軽に動画をモバイル端末で見られるようになっています。あらためて，すごい進歩ですね[*14][*15]。

ところで，1byte は 256 種の情報が記録できるので，英語のアルファベットや数字は 1byte あれば十分だが，日本語や中国語など，英語以外の言語は文字種が多いので，2byte で一文字を表すという話をしました。この 2 バイト文字も，たとえば「あ」とか「亜」という文字に 1110001110000001100000010 とか 111001001011101010011100 という文字列を割り当てるのですが，言語ごとによってどの数字をどの文字に割り当てるかという対応表が変わってきます。これを文字コードと言います。日本語はかつて Shift-JIS というコードで変換していましたが，今は世界のあらゆる言語に対応している共通企画である，UTF-8 という文字コードで変換することが一般的です。ところがなぜか，日本の Windows OS だけ Shift-JIS をいまだに使い続けており，他の PC とファイルをやり取りするときに文字コードの変換エラー問題が起きます。ファイルを開いて文字化けをしたりとか，プログラムが実行される際に「ファイルにアクセスできない」というエラーが生じたりするのは，この文字コードの問題が大きいのです[*16]。受身的な対策法になってしまいますが，PC でつかうユーザ名やファイル名などは半角英数文字を使い，短くした方がこうしたエラーに出くわしにくくなります。逆に，全角文字やスペース (空白) などを含んだファイル名，やたらと長いファイル名を使っていると，こうした問題に出くわしやすくなるということです。

A.5 ファイルの種類と拡張子

ここまで述べてきたように，計算機というのは基本的に物理的実体 (記録装置，記憶装置) の上で 0/1 のデータをやり取りしているだけです。記録 (記憶) 装置上に置かれている情報のセットは「ファイル」という形で記録されています。スマートフォンやタブレットは，ユーザの利便性のためにファイルの存在を意識しなくても良いようになってはいますが，バックエンドでは実行されるアプリケーションもファイルですし，開かれる音声や動画もファイルです。とくにパソコンでは，どの媒体，どのアプリで使うどういうファイルかを識別するために，拡張子 (かくちょうし) と呼ばれる識別記号をファイルの後ろにつけています。拡張子はファイル名の背後にピリオドで区切って追記されています。ついてないように見えても，OS がそれを表示させない設定にして

[*13] できた当時は写真を撮ってメールができることをとくに「写メールする」と言い表したほどです。

[*14] 実は音でも画像でも，分割してそのままデータにしてしまうと膨大になりすぎるので，人間が気付きにくい周波数や色合いなどは削除して作ります。これを非可逆圧縮処理と言います。一度落としてしまった情報は戻らない，という意味です。ライブや生きている人間が処理している情報は，携帯の画面から得られるものの何億倍もの情報量なんですよ!

[*15] デジタル化のすごいところは，こうした文字，音，図版，動画といったものを bit という共通の単位に落とし込んだことです。こうすることですべて一元的 (bit という共通次元) で処理することができるようになったのです。メディアの違いが問題にならなくなり，0/1 の情報であれば複写も簡単にできてしまいます。情報化社会においては情報に特別性はなく，情報があるかないか，それを生み出せるかどうかこそが重要なのです。

[*16] Windows だけ世界標準から外れているので，早く修正して欲しいのですが，歴史的な経緯からユーザ数が多くて切り替えられないでいることと，こうした違いがあることをユーザに説明しない (素人は知らなくていい，と馬鹿にされているようなものです) ので，問題が解決される日はまだ先になりそうです。

あるだけであることに注意してください。代表的な拡張子と，それに対応づけられているアプリケーションは次のようなものがあります。

.docx マイクロソフト社の文書作成アプリケーション，Word で使うファイル

.xlsx マイクロソフト社の表計算アプリケーション，Excel で使うファイル

.pdf Adobe 社が開発した Portable Document Format 形式。OS が違っても同じレイアウトで文書を表示できるのが利点で，PDF 形式を読むことができるアプリケーションは多数。

.txt シンプルな文字だけのテキストファイル。文字の飾りやレイアウトなどの情報がない最もプレーンな形式なので，OS が違っても文字コードさえ合っていれば読むことができる。

.mp3 音楽，音声のファイル。音声データには他の種類もあります。

.jpg 画像のファイル形式の一種。

.png 画像のファイル形式の一種。

.csv comma/character separated variables ファイル。変数をカンマ (,)，あるいはタブ，半角スペースなど文字コードで区切ったファイルという意味で，中身は.txt と同じく装飾のない文字/数字だけであり，文字コードさえ間違えなければ OS を問わずに読み書きできる。データのやり取りはこの形式で行われることが多い。

.zip 圧縮ファイルの一種。1 つまたは複数のファイルをパックして圧縮してあるもの。ファイルの冗長な部分をうまくまとめてコンパクトにまとめ上げるため，ファイルサイズが小さくなるし，複数のファイルもひとまとめにできる。また圧縮の際にパスワードをかけることもできるため，メールなどに添付する場合はこの形式にまとめられることが多い[*17]。可逆圧縮であり，圧縮されたファイルは展開する (解凍するともいう) ことでパッキングを開封できる。zip ファイルの圧縮/展開は各種 OS が標準的に対応している。

`.txt` や `.csv` といった形式は，「装飾のない，文字だけの」ファイルです。こうした種類のことを ASCII ファイルと言います。メモ帳などのエディタと呼ばれるプログラムで読み書きできます。逆に `.docx` など特定の会社が提供するアプリケーションに対応しているファイルは，アプリの中でのさまざまな操作・装飾を暗号化して保存しており，メモ帳などで読んでも意味がわかりません。対応しないアプリでは開くこともできません。こうした形式は ASCII ファイルに対してバイナリファイルと言います。

OS は拡張子を見てファイルの種類を判別し，そのファイルを開くのに適したアプリケーションを自動的に起動し，開いてくれます[*18]。`.csv` ファイルは Excel などのアプリケーションで開くことも当然できますが，その際文字コードのエラーが生じたり，保存するときに文字コードを変えたりして，形式・内容が気づかずに変わっていることがあります。Windows も良かれと思ってやっていることなのですが，処理が徹底してないのか，かえって不便になってしまっています[*19]。

[*17] **PPAP** というとピコ太郎の楽曲を思い出す人もいるかもしれませんが，圧縮ファイルの文脈では“「Password つき zip ファイルを送ります。Password は次のメールで送ります」Angoka(暗号化) Protocol の略です。つまりメールでパスワード付きのファイルを送り，そのファイルを開くためのパスワードをまたメールで送るという，日本でよく見られるおかしな風習です。おかしな，というのは，メールがハッキングされていたらパスワードもどうせバレるわけで，同じメールに書いてあるのはもちろん馬鹿馬鹿しいですが，すぐ次のメールに書いてあるのも同じぐらいに馬鹿馬鹿しいことです。情報セキュリティ対策手法のつもりで行われる慣習が広まっていますが，PPAP の標語のもと，馬鹿馬鹿しいのでやめましょうという風潮になってきました。

[*18] 見たことのない拡張子の場合は，どのアプリケーションで開くべきか尋ねてくるでしょう。

[*19] 実際，この授業での課題データを UTF-8 形式の csv ファイルで提供しても，Excel で開いたばっかりに文字化けして分析できなくなる，という相談がこれまで多く寄せられています。根本的な解決策として，Windows を使うのをやめることをお勧めします。

A.6 クラウドとは

すでに述べたように，計算機は基本的に 0/1 データのやりとりであり，それを保存してあるのがファイルとよばれるものです。ファイルは HDD や USB メモリ，SSD に保存することができます。

ところで，最近はこうした手元の物理的実体にファイルを置くことに加えて，クラウドに保存することも少なくありません。クラウドとは雲という意味で，インターネットの向こう側のどこか，ということを意味します。ですが，基本的にはインターネットで繋いだその先にも電子計算機があるのです。たとえばパソコン A とパソコン B をケーブルで繋ぐと，パソコン A からパソコン B のファイルにアクセスできます[*20]。このケーブルをどんどん伸ばすと，遠く離れていてもこの操作ができます。このケーブル網を世界レベルに広げているのがインターネットです[*21]。

ここで覚えておいて欲しいのは，当たり前のようですが，ネットといっても基本的には電子計算機と電子計算機を繋げている実体がどこかにあって，ファイルのやりとりをしているだけだということです。ブラウザはウェブサイトの情報を書いたファイルを取り込んでホームページを見せてくれていますし[*22]，Youtube は動画のファイル，Instagram は画像のファイルへのアクセスをして見せてくれているのです。最近ではクラウドサー

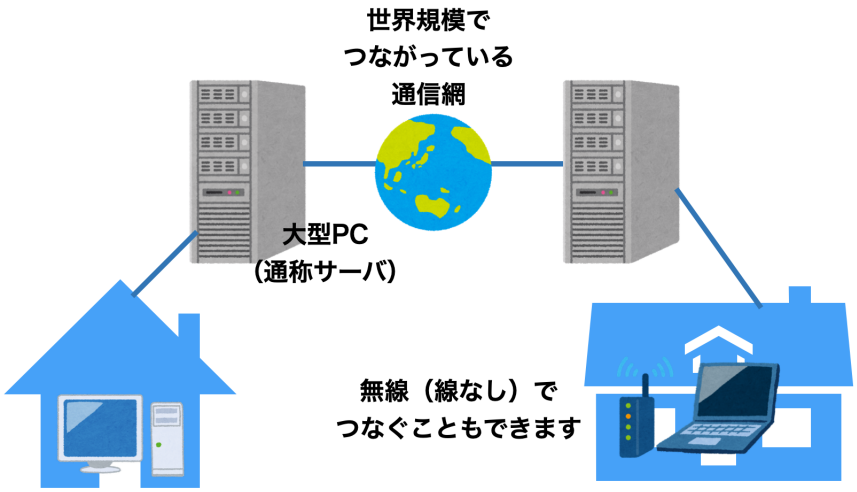


図 A.2 インターネットの世界

ビス，というのがよくあります。中でも Dropbox とか OneDrive，iCloud などが有名です。これらは，ファイルをインターネットを介した別の大型電子計算機に保存 (アップロード) し，インターネットを介して読み込み (ダウンロード) して使う，というものです。手元の電子計算機の中に記録されているファイルのことを「ローカル」，サーバなど遠隔地にある大型機に記録されているファイルのことを「サーバ」「クラウド」と呼んで区別しますが，このローカルとサーバのファイルを常に同じものに同期しておく便利なシステムです。こうしたクラウドサービスを使うと，たとえば大学で作業して保存したファイルを，USB メモリに保存して自宅のパソコンにコ

[*20] もちろんファイルの情報をどのように信号に変えて送受信するか，アクセス権限はどうするかといった細々したことを調整しなければなりませんが，そのあたりの仕事をやってくれるのが OS のありがたいところです。

[*21] インターネットは軍事通信網として始まりましたが，それを電話回線や企業内通信網，大学間通信網などと接続しあって世界中に広げています。網と網が相互に (inter) 繋がっているので inter net であり，大学内・企業内のネットワークのことはイントラ (intra) ネットと言います。ちなみに一般用語としてのインターネットは Internet と大文字で書き始めます。

[*22] ホームページとは本来ブラウザが最初に開くページのことを指し，企業や個人が情報発信しているページのことはウェブサイトというのが適切です。

ピーする，という操作をしなくても，大学でクラウドサービス上のフォルダ内に保存しただけで，自宅のパソコンにそのファイルがコピーされているため[*23]，意識せずにそのまま作業を続けられるということです。自動的にバックアップをとってくれているとも言えるので便利ですね。

もちろん注意すべきこともあります。「他の人に見られたら困るファイル」，「アプリケーションの実行ファイル」などは通信網の向こうではなく，手元 (ローカル) に置いておきましょう。個人情報・機密情報などを，クラウドドライブに保存すると，悪意を持った人が大型計算機に攻撃を仕掛けて情報を盗んでいく可能性があるからです。情報化の怖いところは，取られても気づかない (コピーすればいいだけで元ファイルになんの影響もない) ところにあります。加えて，取られたものをばら撒かれる＝インターネットを介して誰でもアクセスし保存できるようにされると，すべてを回収できなくなるのも問題です。失言が記録されて拡散されると大変なことになるのは，みなさんもこの時代に生きる人間のマナーとして色々見聞するところと思います。また，クラウドサービスで自動的に同期されるといっても，アプリケーションの実行ファイルなどはローカルに保存するべきです。アプリケーションは実行に際してさまざまな関連ファイルにアクセスしますので，1 つでも場所が違うところがあるとエラーになって動かなくなります。同じ OS でも，です。インターネットからとってきたソフトウェアをうっかり OneDrive に保存してしまうと，アクセスできないエラーで起動しない，ということもありますので注意してください[*24]。

余談ですが，R などプログラミングを推している時に最も頻発する質問が「ファイルがありません/読み込めません」です。その理由は大きく 2 つあって，1 つはスペルミスです。ファイル名の大文字・小文字の違いということもありますし，拡張子のまえに半角スペースが入っている，というよくみてみないと気づかないようなものもあります[*25]。2 つめの問題は，パスが通らないことです。デスクトップに置いたのに見つからない，ということもよくあります。教員や TA を悩ませるタイプの問題として，同じファイル名・フォルダ名が別の場所にあることがあります。フォルダ A の中にファイルを置いたのに見つかりません，という質問に対応していて，確かにフォルダ A を開いているのにな，と思ったら別の場所のフォルダ A だった，ということがあるのです。パスは隅々まで同じでなければ意味がないので，自分がどこに何を置いたのかをしっかり把握しておきましょう[*26]

A.7 ファイルの位置の指定

ここでファイルとそのパスについての話をしておきたいと思います。

計算機が情報を 0/1 で管理し，それらがファイルとなってどこかに保存されている，ということでした。私たちは Finder や Exploer などファイルブラウザをつかって，実行したいファイル，参照したいファイルを探していきますね。ファイルはまとめてフォルダの中に含まれていますし，フォルダの中にフォルダがあるといった，階層状態になっていることも少なくありません。ちなみに

[*23] これはユーザが特段の指示をしなくても，アプリケーションがユーザの見えないところで (バックグラウンドで) アップロード，ダウンロードの作業を進めているからです。パソコンはシャットダウンしていなければ，裏でこうした作業を淡々とこなし続けてくれています。

[*24] マイクロソフト社は，これまたユーザのためを思ってやっているのかもしれませんが，デフォルトで OneDrive に保存させようとします。それでうまくインストールできなかったという相談も多々寄せられています。抜本的な解決策として，Windows を使わないことをお勧めします。

[*25] k これはたとえば，`sample.csv` というファイルをダウンロードしてね，と指示した時に，複数回同じ場所に保存してしまって，機械が，`sample (1).csv`，`sample (2).csv` などと名前をつける時に起こります。この `(1)` を削除してもその前に半角スペースが入ってたりするのです。まったく Windows は！

[*26] Windows は One-Drive にもデスクトップがあります。同じ環境をローカルと同時にクラウドにもあって常に同期していればいいのですが，必ずしもそうではないので (OneDrive の設定によります)，`C:\Users\Username\Desktop` と `C:\Users\Username\OneDrive\Desktop` のどちらに保存したのかが分かりにくくなっています。デスクトップに保存したのに，と言われてもどっちのデスクトップなのか，という問題が生じるわけです。さらに専修大学では 2023 年から仮想デスクトップ環境 (VDI) を使うようになりました。これで 3 つ目のデスクトップができたことになるわけです。なんてこった！まったく Windows は！

フォルダと同じ意味で**ディレクトリ**という言葉が使われることもあります。

A.7.1 相対パスと絶対パス

普段 PC を使っているときは気にすることがありませんが，R や RStudio などプログラミング言語をつかっているときは，「今どこで作業しているか」という現在地が重要になってきます。たとえば，RStudio で `C:\User\kosugitti\Document\kiso1\`というところでプロジェクトを開いているとします。スラッシュ (\) はフォルダ，コロン (:) はドライブを表す記号です。プロジェクトフォルダは，C ドライブの User フォルダの下にある，kosugitti フォルダの下にある，Document フォルダの下にある，kiso1 というフォルダということになります (プロジェクト名が `kiso1` だとそうなります。)。この kiso1 フォルダが現在地です。

このフォルダの中で，Rmd ファイルや R スクリプトファイルを使って，他のファイルを参照するようなコードを書くとしましょう。たとえば `script1.R` というファイルに `read_csv` 関数を書いたとします。読み込みたいファイルは，同じフォルダの中にある，`sample.csv` とだとします。このとき，`read_csv` の書き方は次のようになるでしょう。

code：A.1 相対パスで読み込む

```
1 dat <- read_csv("sample.csv")
```

しかし別の書き方もあります。たとえば code:A.2 のような書き方でも問題ありません。

code：A.2 絶対パスで読み込む

```
1 dat <- read_csv("C:\User\kosugitti\Document\kiso1\sample.csv")
```

後者 code:A.2 の書き方は，ファイルの場所を全部書いてありますから，確実にその場所が特定できます。それに比べて前者 code:A.1 の書き方は，なぜファイルを書いただけでいいのでしょうか。これは，このコードを実行している現在地と同じフォルダの中に `sample.csv` ファイルがあるからです。プログラムは，命令を受けるとファイルを探しにいきますが，現在地と同じフォルダの中を探すことになっているのです。この現在地，すなわち現在作業しているフォルダのことを**作業フォルダ (working directory)** と言います。

では作業フォルダと別のフォルダの中にファイルがあれば，アクセスできないのでしょうか。そんなことはありません。code:A.2 の書き方を使えば，作業フォルダがどこにあっても位置を特定できますから，作業フォルダを問わずに書くことができます。ちょっと長くて面倒ですが，確実にある場所を指定しているからです。この書き方のことを**絶対パス**による指定といいます。一方，code:A.1 の書き方は，今の作業フォルダから見た場所，という相対的な書き方になっています。この書き方のことを**相対パス**による指定，といいます。相対パス指定で，違うフォルダにアクセスする場合には，次のようにします (code:A.3)。ここでは，`Document` フォルダの中に，kiso1,kiso2 フォルダがあり，kiso1 フォルダの中で作業している時に kiso2 フォルダの `sample2.csv` ファイルを読み込む例を書いています。

code：A.3 絶対パスで読み込む

```
1 # 絶対パス指定
2 dat <- read_csv("C:\User\kosugitti\Document\kiso2\sample2.csv")
3 # 相対パス指定
4 dat <- read_csv("..\kiso2\sample2.csv")
```

絶対パスはそのままなのですが，相対パスは..\という記号になっていますね。このピリオドを 2 つ打つ方法で，「ひとつ上のレベルの」という意味になります。このように，現在地からの相対的な位置関係で，ファイルを指定することもできます。

絶対パスと相対パスのどちらが良いのか，というのは一概には決められません。絶対パスは，PC が変わったりフォルダの構造が変わったりすると役に立ちませんから，使い勝手が悪いと言えなくもないですが，確実に指定できる方法です。相対パスは，PC が変わったりしても「現在地から相対的に見てどこか」という話ですから，たとえばこの例で `kosugitti` フォルダごと別の場所に移しても (たとえば `D:\Univ\Classes\kosugitti\kiso1` のように)，コードはエラーなく動きます。kiso1 フォルダ，kiso2 フォルダの相対的な位置関係が変わらなければいいのですから。バックアップを取ったり，複数の環境で同期しながら作業する場合などは相対パスの方がいいでしょうね。

いずれにせよ，現在どこで作業しているかということ，すなわち作業フォルダの場所を，意外と意識しておかなければならないということには注意が必要です。ファイルをどこに置いたか，どんなファイルを置いたか，自分はどこにいるのか，これが変わってくると「ファイルが見つかりません」というエラーになるのです。言い方を変えると，**ファイルが見つかりませんエラーの原因は，この 3 つのどれかである**ことがほとんどです。

A.8 ファイルのバージョン管理

これからみなさんは大学生活の中で，たくさんのファイルを生み出していくことになるでしょう。たとえば 4 年生の時に卒業論文を書くことになりますが，データファイル，分析ファイル，図を書いたファイル，引用文献リストを書いたファイル，卒論本文などなど，1 つの研究でも複数のファイルが作られることはよくあります。さらに，これらのファイルは日々加工されますから，その度に上書き保存することになります。いわば，ファイルがバージョンアップしていくのです。

卒論などの場合はとくに，「途中で保存しておく」ことが重要です。途中まで書いていた時に，横に置いていたマグカップが倒れて PC にコーヒーがかかり，変な音を立てて PC が壊れてしまった，ということがあるかもしれませんからね。紙に書いていた時代は，その手のハプニングがあってもせいぜい原稿用紙数枚がダメになっただけで，「ちくしょう，やりなおしかぁ」で済んだのですが，電子データの場合は電子の藻屑になると復元させることができません。ですから**バックアップは非常に大事**なのです[*27]。

バックアップの基本は，「別の場所に」「別のファイル名で」というものです。同じ名前で上書きすると元に戻ることができませんから，面倒でもコツコツと違う名前をつけましょう。そうするとよくあるのが，`soturon1.docx`,`soturon2.docx`,`soturon3.docx`,`soturon3`(**修正**)`.docx`,`soturon3`(**最新**)`.docx`,`soturon3`(**最新**)(**修正**)`.docx`,`soturon3`(**最新**)(**修正**)(**提出版**)`.docx`,`soturon3`(**最新**)(**修正**)(**提出版** 2)`.docx`.... というようになっていくやつで，「どれが最新版だっけ...」と書いてる本人でも探すのに苦労することになります。

この問題の解決策として，`soturon1001.docx`，`soturon1005.docx` のように日付を入れるというものがあります。10 月 1 日分，10 月 5 日分，としていけば「いつまで戻れば良いか」もわかるのでいいやり方ですね。日付の数字をファイルの先頭につけておくと，並べ替えも簡単です。この日付をつけて保存するというのを習慣化し，1. 昨日までのファイルを開いて今日の日付で別名保存する，2. 作業を進めて，時々上書き保存，最後にも上書き保存，3.PC に USB メモリをさして，バックアップ保存して作業終了，というルーティンを作っておくと，確実に記録が残って良いでしょう。

ただし，この方法の問題は，ファイルサイズが大きくなりすぎることです。図表などを含めたファイルが数百 MB になることは少なくありません。それを次々複製していくわけですから，大容量の USB メモリでも限界が来るかもしれません。これは「丸ごとコピー」していることが原因で，たとえば昨日は 10 行目まで書いた，今

[*27] ちなみに私の経験上，レポートが電子の藻屑になったので助けてください！と言われることがよくあり，4 年間の学生生活の間では平均 10–15 名に 1 人の割合で発生することのようです。

日は 11 行目から 14 行目まで書いた，というのであれば，この増えた 4 行分 (差分) だけを追加保存すればいいのに･･･と思いませんか。

こうしたバックアップやバージョン管理をやってくれる仕組みとして，Git というものがあります。Git は作業フォルダの中身の差分だけを記録し，必要であれば過去のバージョンに戻すこともできるシステムです。毎回上書き保存 (commit する，といいます) のたびに「どこを変更したか」というメモを付けて保存しておけば，そのメモを見ながら「ここの時点まで戻ろう」という使い方をすることができます。ファイル名は変更する必要なく，同じファイル名で進めていけますから，同じようなファイルがたくさんあって訳がわからなくなるということもありません。さらに保存先をクラウドにした GitHub というものもあり，これを使うとクラウド上に追記していくことができます。この GitHub は IT 企業などでプログラムをチームで進めていく時にも使われている技術で，全体のプログラムに個別の機能を複数人が追加，管理者が必要なものだけ取り入れる，というように使われています。国里ゼミや小杉ゼミでは，卒論を GitHub で管理し，学生が書いた分を commit し，それを hub 上にアップロード (push といいます) する，というようにします。教員の方は学生の進捗が管理できますし，どこがどう変わったかが分かりやすく，バージョン管理と同時にバックアップもできるという便利な仕組みです。GitHub は無料でアカウントを作ることができますから，興味があれば皆さんもチャレンジしてみてください[*28]。

さてここで，ひとつ注意をしておきます。卒論の原稿やプログラムは日々変化するものですからバージョン管理が必要ですが，データファイルはアップデートする必要がありません。いや，アップデートしてはおかしいのです。たとえば 100 人分のデータを取って分析をしていて，後で「やっぱりこのデータを削ろう」というのは，研究不正が疑われかねません。自分に都合の良いデータだけで議論し，都合の悪いデータは削除して統計的に有意な結果が出るように細工しよう，なんてことがあれば困るのです。データファイルはバージョンアップせず，それを加工，計算するプログラムがバージョンアップしていく。そしてその加工プロセスは誰でも見ることができるように公開されている。少なくとも，科学的な営みをする上では，そうしたやり方が必要なのです。自分だけのデータで自分だけの分析方法で，良い結果だけ示すというのは適切な方法ではありません。

Open Science Framework(https://osf.io/) はこうした「オープンな科学」にむけた取り組みの一種です。このアカウントは誰でも無料で作ることができ，ここにファイルをアップロードしたり，分析計画を事前に記録しておくことができます。何も難しいことではなくて，クラウド上のファイル置き場だ，というぐらいに思っていただければ結構です。ここにおかれたファイルは自動的にバージョン管理され，同じファイル名のものがアップデートされてもその記録 (ログ) が残ります。最近はこの OSF をつかって，論文化されたデータやプログラムを公開するという取り組みも進んでいます。

クラウド，バックアップ，オープンサイエンスといった新しい研究方法は日々生まれてきています。皆さんも便利な機能はどんどんキャッチアップしていきましょう！

A.9 おわりに

古臭い話をしてしまうのは，私が歳をとった証拠でしょうか。皆さんはこんな話を知らなくても，スマホやタブレットを使いこなしていることと思います。細かいことを知らなくても，ユーザとして利用するだけなら知らなくて良い話なのかもしれません。私はここにも書いたように，高校生の頃からコンピュータの発展と一緒に大人になってきましたから，学ぶともなく学んできたところがあります。皆さんは生まれた頃からコンピュータやがあったネイティブ・デジタル・カウボーイですから，苦労なんかする必要なかったわけです。

[*28] このテキストやシラバスも GitHub で管理していますし，公開されているサイトも GitHub 上のものです。これからは教科書も日々成長していくものになるかもしれません。

しかし細かい仕組みを知らないということは，問題が生じた時に「何か・誰かが，どこかでどうにかなって，今私が困っている」という状況になります。問題を特定できないと，解決することもできません。コンピュータは文房具に過ぎませんから，それを使いこなせないほうが格好悪いのです。しかも今後ますますコンピュータに囲まれた世界になっていくのは自明ですから，ここに学習コストをかけない方が勿体無い。幸い，わからないことに対して，自ら調べて学んだ利する時間と環境が用意されているのが大学という世界なのですから，今のうちにしっかり基礎固めをしておきましょう。

ここのくだらない懐古的エッセイのような文章が，何かの足しになれば幸いです。

付録 B

よくある質問とミスの例

B.1 Frequently Miss and Comments

Rmd でレポートを提出したのに，なんだか中身の問題じゃないのに突き返された，中身を見てくれよ！と思う人もいるかもしれません。テキストでは R や Rmd での課題を提出するよう求めているところがありますが，その際よく見られる学生さんのミスとその対応についてのコメントをここにまとめておきます。

B.1.1 FMC1；そもそもファイルの書式が違う

Rmd で提出してください，R で提出してください，という指示に対して，違うものが提出されてくることがあります。書式があっていない，というのは些細なことのように思えるかもしれませんが，学術論文は書式に捉われず内容に集中するためにも，書式は整えられたものである必要があります。学会誌に掲載されている論文も，みなさんが書く卒業論文も，レポートに至るまで，書式や指示に沿ったものを準備する必要があります。書式があっていない場合は，門前払いになっても文句が言えないのがアカデミックの世界です。

ということですので，R や Rmd で提出してください，という指示があれば，R や Rmd で提出してください。ファイルの種類は拡張子で分類され[*1]，R ファイルは `.R`，Rmd ファイルは `.Rmd` という拡張子になっています。たまに `.Rproj` というファイルを提出してくる人がいますが，これは R プロジェクトのファイルで，これには R スクリプトも文書も含まれておらず，みなさんの計算環境情報が少し書いてあるだけです。また，`.Rmd` だけ入っていれば良いかというとそうではありません。まれに `Filename.Rmd.R` というようなファイルを送ってくる人がいます。これはファイル名にピリオドが含まれているだけで，ファイルの種類を識別する拡張子は `.R` ですから Rmd ファイルではありません[*2]。そもそもファイル名には 2 バイト文字はもちろん，!や?などの記号を含めるべきではありません。ピリオドも当然記号の一種ですから，ファイル名にするのは不適切です[*3]。

ではどうやって適切なファイル形式にするか，ということですが，最も素直な方法としては RStudio で新しくファイルを開くときに，R markdown 形式 (略して Rmd) を選ぶようにしましょう。もし間違って，R script 形式 (`.R` 形式) で開いてしまった，というときは，そのファイルを破棄して新しく作り直すのが一番ですが，エディタペインの右下にあるファイル種類表示 (図 B.1) をクリックして修正することもできます。

[*1] 拡張子についてはセクション A.5,Pp.396 参照

[*2] 最近の OS は拡張子を表示しない設定になっているものも少なくないので，このようなミスが生じます。

[*3] 長いファイル名などの場合，空白を入れるのも適切ではありません。できなくはないのですが，やるべきではないのです。どうしても空白を入れたければ，アンダースコアなどで区切ると良いでしょう。

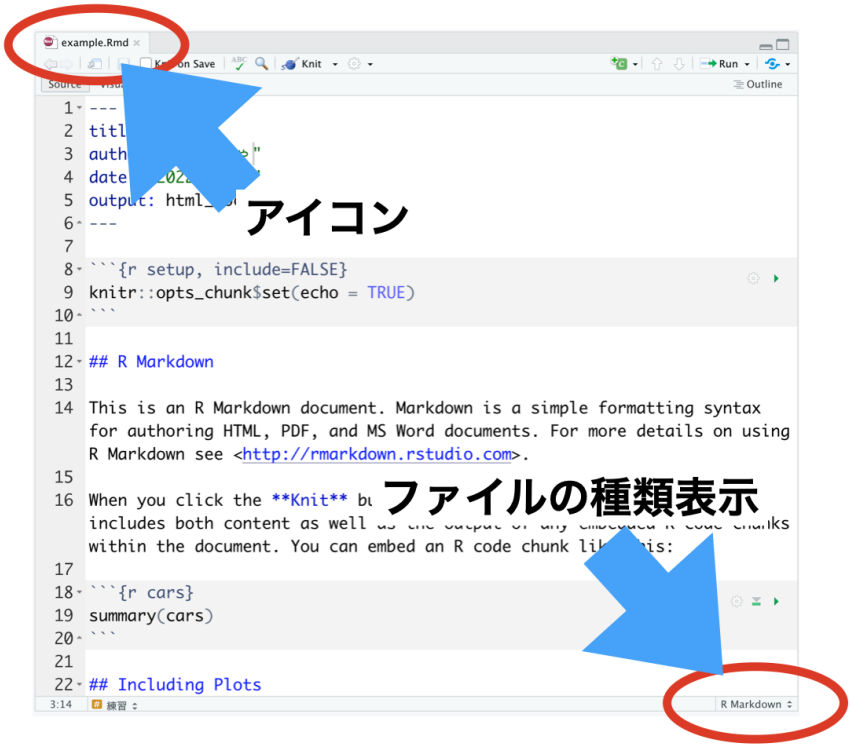


図 B.1 ファイルの種類を判別する方法

B.1.2 FMC2；やっぱりファイルの形式が違う

Rmd 形式のファイルは，拡張子だけで決まるものではありません。図 B.2 にあるように，Rmd ファイルは冒頭の 6 行 (正確には---で囲まれた領域) が YMAL と呼ばれるところで，文書全体の設定をしています。その下に，R のコードを実行する部分 (**チャンク (chunk)**) や文章の領域があります。文章のところは#記号で見出しを作ったりできます。

この YMAL 部分が壊れている，あるいはチャンクが正しく記述されていない場合，拡張子が.Rmd であっても適切な Rmd ファイルにはなっていません。YMAL 部分の書き方はよくわからない，という人も多いと思いますので，RStudio で Rmd ファイルを作ったときの状態をなるべく変更しないように注意すると良いでしょう[*4]。

チャンクは R のコードを書くところで，バッククォーテーション 3 つでくくるのが決まりです。チャンク領域が始まるところに{r}とかいて「ここが R で計算するところですよ」というのを指定するわけです[*5]。このバッククォーテーション 3 つ (```) が全角だったり (｀)，ダブルクォーテーションだったり (") すると，機械は正しくチャンクであると認識しません。フォントなどの見せかけ上は，微妙な違いのように見えますが，機械にとって違う文字列は違う意味を持ちますので注意してください[*6]。RStudio で編集する場合，チャンクの領域はや

[*4] Rmd ファイルを新しく開くときに，文書タイトルや著者名，日付，出力ファイル形式などを設定することができるウィンドウが開きますので，そこに必要な情報を書くと自動的にそれを使った YMAL が生成されます。また，本文として幾つかのサンプルコードが最初から含まれていますが，これらに関してはサンプルをそのまま使うことはないので，全て削除してしまって結構です。

[*5] ほかにも Python や Julia など他の計算言語を混ぜることもできます。

[*6] 記号の名称や入力については，セクション D,Pp.419 を参照してください。

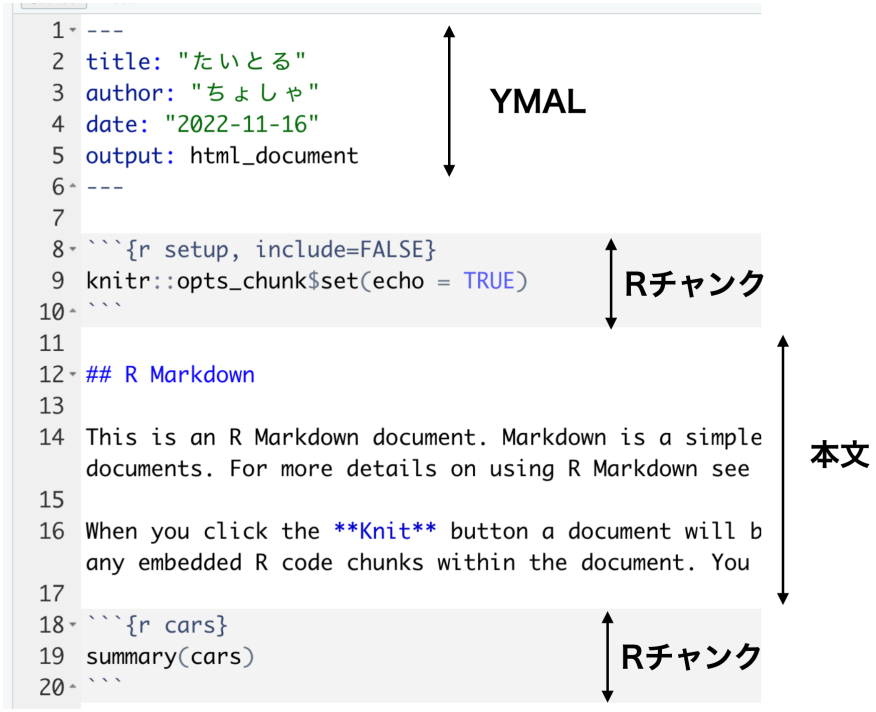


図 B.2　Rmd ファイルの中身における書式

や灰色がかった強調表示がされますので，どこにチャンクがあるかわかると思います。もし強調表示されていないようであれば，チャンクとして認識されていない可能性を疑った方が良いでしょう。

チャンクはバッククォーテーション 3 つで開き，おわったら同じくバッククォーテーション 3 つで領域を閉じます。閉じなければずっと R の計算領域が続くと解釈されますし，最後までチャンクが閉じられないと Rmd ファイルとして正しくない書式ということになります。チャンクが正しく入力できているかについて，常に注意を払っておく必要があります。また，自分でバッククォーテーション 3 つを使ってチャンク領域を開いたり閉じたりするのが面倒だ，という人は RStudio のチャンク挿入ボタンから挿入すると間違いないでしょう。

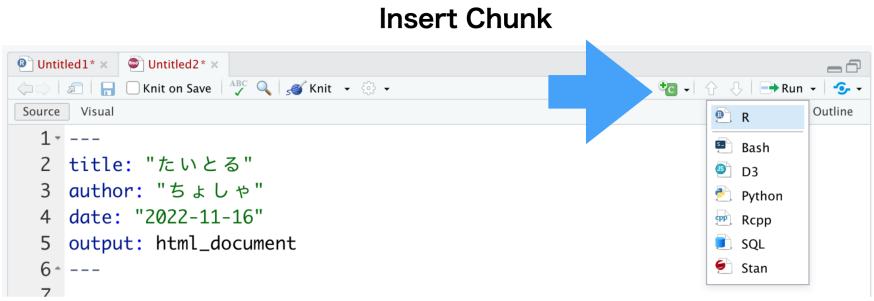


図 B.3　RStudio のチャンク挿入ボタン

B.1.3　FMC3；Rmd が knit できない

Rmd は文書作成に加えて，R での計算がセットになったファイル形式であり，文中の数字や分析結果は「書き写す」のではなく「その場で生成する」ものです。生成する，というのは Rmd ファイルを変換して，

HTML や DOCX，PDF 形式のファイルにすることを指します[*7]。このファイル変換を**ニット (knit)** といいます。編み物を編むようなイメージですね。このときに，タイトルをつけ，見出しのサイズを変え，R の計算をして結果を埋め込む作業をするわけです (図 B.4)。こうすることで，結果のコピペを避けること，再現性を担保することができるようになるわけです。すでに説明したように，チャンクを使って計算に必要な指示を Rmd

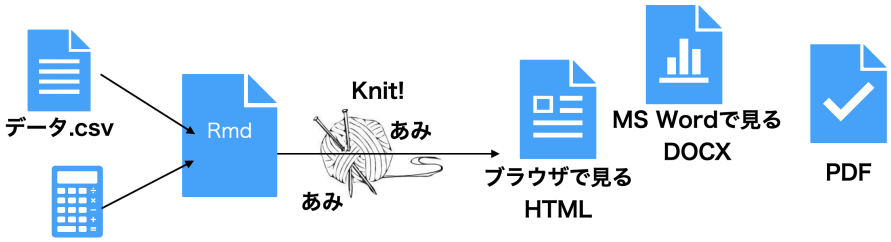


図 B.4 knit してレポートが完成する

ファイルに書きます。この指示はエラーのないコードである必要があります。当然ですが，スペルミスや間違った R コードは「実行できない」というエラーになるわけです。その場合，knit は中断され結果のファイルが出力されません。提出されたレポートは knit して出来上がったもののことを指しますから，knit できない Rmd ファイルを提出されてもレポートが提出されたことにはならないわけです。

knit できない Rmd ファイルになっていないか，というのはご自身の RStudio 環境で knit して確認できることです。提出前に一度，きちんと機能する Rmd ファイルになっているかどうか確認するようにしてください。以下で述べるように，自分の環境で knit できても，提出先の (私の) 環境で knit できないということもあり得ます。しかし，自分の環境で knit できないのに提出先でできる，ということはあり得ませんので，「knit できませんよ」というコメントをつけて返される前に自分で確認するようにしてください。

B.1.4 FMC4；外部環境を参照してしまう

さて，R で行う作業の中には，csv ファイルを読み込むといった「外部のファイルを使う」指示もありえます。例えば次のようなコードです。

code : B.1 Rmd 中の R コードが外部環境を参照してしまう

```
1  dat <- read.csv("social_effects.csv")
2  dat <- read.table("clipboard")
```

このコードでは，上の行は `social_effects.csv` というファイルを，下の行はクリップボードの内容を読み込むようになっています。ファイルを読み込む指示は，ファイルがなければ当然エラーになりますから注意が必要です。レポート等で Rmd ファイルを提出するとき，Rmd ファイルの他に必要な読み込むべきファイルがあれば一緒に提出するようにしてください。また，クリップボードの内容は再現できません。クリップボードとは，コピーアンドペーストのコピーを行ったとき，PC の内部で一時的にコピー内容を覚えておく場所のことです。つまり，みなさんがみなさんの環境で，クリップボード上にデータを一旦保持している場合は，このコードでエラーがしょうじることはありません。しかし，レポート提出先の (私の) 環境で，事前にデータのコピー作業を行なっているわけではないのですから，提出先でエラーが発生します。

[*7] HTML は Hyper Text Markup Language の略で，ブラウザで開くファイル形式です。DOCX は Microsoft Word のファイル形式です。PDF は Portable Document Format の略で，データを紙に印刷した状態のようにサイズを固定して出力したファイル形式であり，OS がもつビューワーや Adobe Acrobat Reader などで見ることができます。

そもそも Rmd ファイルは，作業の再現性を担保するためのファイルになっているわけですから，クリップボードの利用のような「自分の環境だけで可能な記録されない作業」をそのファイルに含めるのは適切ではありません。Rmd ファイルは Rmd ファイルだけで分析作業が完結するように，必要な記録は全て記載されている必要があります。同様に，分析作業に関係のない冗長な指示や無駄な指示は Rmd ファイルに含めるべきではありません。

B.1.5 FMC5；提出先の環境を変更してしまう

R はさまざまなパッケージを使って分析環境を拡張することができます。パッケージは **CRAN** を通じてインターネット経由で配布されますから，ネット環境があれば誰でも最新のパッケージをとってくることができます。みなさんが Rmd ファイルの中で使う R の関数の中には，パッケージの関数も含まれているでしょう。パッケージがなければ関数が動きませんから，パッケージをインストールする作業も Rmd に書いておきたいと思うかもしれません。しかし，これは推奨できません。

パッケージのインストールは，実行環境の準備にあたる作業です。Rmd ファイルを使ってコードのやり取りをするとき，提出先の環境で分析することになりますが，提出先の環境にどのようなパッケージをいつ入れるかは，提出先の環境の管理人が判断すべき問題です。提出する Rmd ファイルにパッケージをインストールする関数，すなわち `install.packages()` 関数が含まれているというのは，相手の環境を勝手に操作してしまうことと同じであり，セキュリティ的にも適切な発想ではありません。

環境の準備は提出先の管理人が管理すべき問題であり，またすでにパッケージが入っている場合は無駄な上書き作業をさせることにもなります。また `install.package` 関数は CRAN のサーバを参照したりしますから，適切な設定がなければ R チャンク実行時にエラーが発生します。いずれにせよ，R チャンクの中に `install.package` 関数を含めないようにしましょう。

以上が Rmd ファイルや R ファイルでレポートを提出するときの留意点です。その他にも R や RStudio を使うときによくある質問がありますので，それらも一問一答型で紹介しておきましょう。

B.2 Frequently Asked Questions；よくある質問と答え

Q. B.2.1: テキストを参考にパッケージをインストールしようとしたところ，エラーが発生しました。

A. 質問ではなくて報告ですね。お返事としては，「わかりました。エラーが発生して大変ですね。」としか言いようがありません。どの環境で，何をしようとして，どのようなエラーが出たのか，明示してください。メールのやり取りで指示が明確になるように，テキストを準備しています。何章，何ページの，どの文章を参考にしたのかも教えてください。

Q. B.2.2: 添付のようなエラーが発生しました (図 B.5)。対応策を教えてください。

A. エラーの発生画面を送ってこられていますが, この画面に写っていない上の方でエラーが発生していますから, これではどのエラーなのかわかりません。スクリーンショットを送ってこられることはよくありますが, ほとんどの場合, 適切な箇所が写っていません。複数枚添付してこられる人もいますが, 画面を拡大しながら読むのも難しいので, **R コードそのものを送ってきてください**。そうすると, 何行目のどこにエラーがあるかが明確になり, 対応に関係ない無駄なやり取り (ex.「もう少し上を写してください」「違う, もっと上」) が減ります。

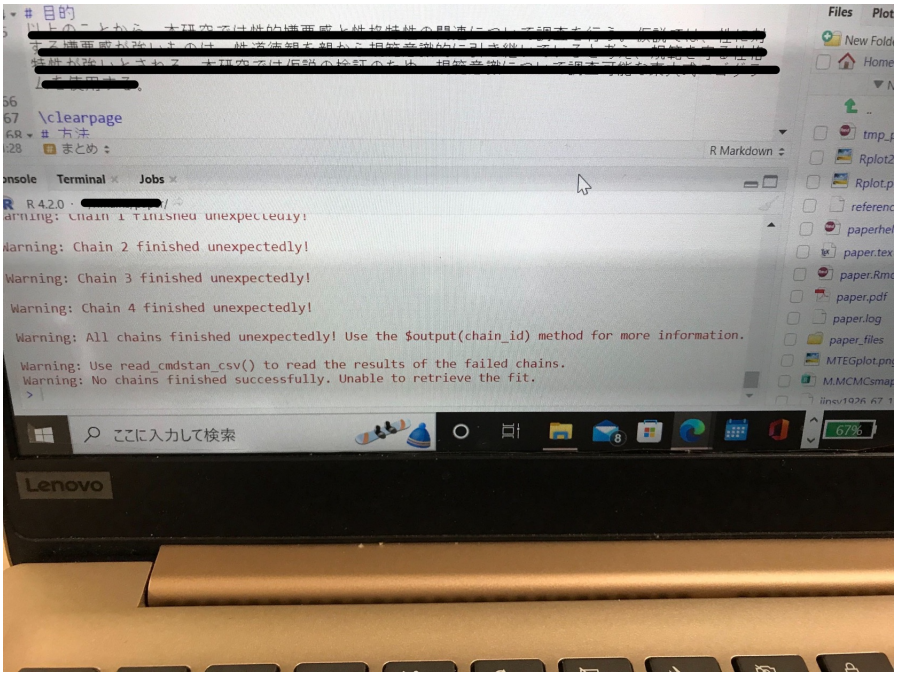


図 B.5 添付されてくるスクリーンショットの一例

Q. B.2.3: ファイルを読み込もうとすると次のようなエラーが出ます。どうすれば良いですか。

`'xxxxx'` に不正なマルチバイト文字があります

A. 文字化けの原因は, たとえば UTF-8 形式で提供されているファイルを, Windows 標準の CP932 形式で読み込もうとする, という文字コードの不一致です。ですからそれをオプションで指定してあげれば問題解決です。`read.csv` 関数を使っている場合, オプションの指定は `read.csv("foo.csv",fileEncoding="UTF-8")` のようにします。`tidyverse` の `read_csv` 関数を使っている場合, locale オプションで, locale 関数の encoding を明示的に UTF-8 とします。`read_csv("foo.csv", locale=locale(encoding = "UTF-8"))` のようにします。

Q. B.2.4: ファイルを読み込もうとすると次のようなエラーが出ます。どうすれば良いですか。

`'xxxxx'` に不正なマルチバイト文字があります

A. 文字化けの原因は，たとえば UTF-8 形式で提供されているファイルを，Windows 標準の CP932 形式で読み込もうとする，という文字コードの不一致です。ですからそれをオプションで指定してあげれば問題解決です。`read.csv` 関数を使っている場合，オプションの指定は `read.csv("foo.csv",fileEncoding="UTF-8")` のようにします。`tidyverse` の `read_csv` 関数を使っている場合，locale オプションで，locale 関数の encoding を明示的に UTF-8 とします。`read_csv("foo.csv", locale=locale(encoding = "UTF-8"))` のようにします。

Q. B.2.5: パッケージを読み込もうとすると次のようなエラーが出ます。どうすれば良いですか。

`library(tidyverse)` でエラー；`'tidyverse'` というパッケージはありません。

A. パッケージがないので，インストールしてください。

Q. B.2.6: パッケージをインストールしようとすると次のようなエラーが出ます。どうすれば良いですか。

```
Warning in install.packages:
    'lib = "C:/Program Files/R/R-4.0.5/library"' is not writable
```

A. ユーザがファイルに書き込みをする権限を持っていないので，(パッケージファイルをドライブに) 書き込みできないというエラーです。RStudio を実行する時に，「管理者として実行」を選びましょう。

Q. B.2.7: パッケージをインストールしようとすると次のようなエラーが出ます。どうすれば良いですか。

```
Warning in install.packages:
    ディレクトリ 'C:〜(任意の文字列)〜\????' を作成できません。理由は
```
```'Invalid argument' です。```

A. ユーザ名が全角文字を含んでいるため，文字化けして R 側から操作ができません。解決する方法は二つあります。

ユーザを作り直す方法 ユーザ名に全角文字を含まない，新しいユーザを作成します。設定＞アカウントから新しいアカウントを作りましょう。

インストールフォルダを指定する R のコンソールで `.libPaths()` と実行すると，パッケージをインストールするフォルダが出てきますが，ここを変更する方法です。次の 2 ステップで対応できます。

- まず C ドライブのすぐ下にインストール先のフォルダを作ります。`myLib` としましょう。
- 次に R のコンソールで `.libPaths("C:/myLib")` 書いて実行します。

これでインストール先が変わりますので，書き込み・インストールができるようになります。老婆心ながら付け加えますと，フォルダ名はなんでも構いませんが，全角文字ではいけません。また場所も C ドライブのすぐ下である必要はありませんが，全角文字や空白を含むフォルダの下に入れてしまってはいけません。OndDrive のようなクラウドを指定するのも良くありません (探しに行った時にオフラインだとまたエラーになります)。
```

Q. B.2.8: ファイルが読み込めません

```
file(filename, "r", encoding = encoding) でエラー:
  コネクションを開くことができません
 追加情報:  警告メッセージ:
 file(filename, "r", encoding = encoding) で:
  ファイル 'foo.csv' を開くことができません: No such file or directory
```

A. 指定された場所にファイルがないので，読み込むことができないエラーです。確認すべきは，「プロジェクトを開いているか」，「プロジェクトフォルダの中に当該ファイルはあるか」，「ファイル名のミススペルはないか」です。プロジェクトってなんだという人は，RStudio の基本に立ち戻り，プロジェクトでファイルやフォルダを管理するようにしてください。プロジェクト管理については，Pp.47 にもその説明があります。

プロジェクトによる管理とは要するに，R が今見ているフォルダの場所を固定する方法です。プロジェクトを開くと，プロジェクトフォルダが「今見ているフォルダ (ワーキングディレクトリ)」になりますので，その中のファイルを参照することになります。プロジェクトフォルダの中に当該ファイルがないと読み込むことができませんので[a]，当該ファイルをプロジェクトフォルダ内に移してきてください。

Rmarkdown や Quarto ファイルの場合も，必ずプロジェクトフォルダの中に置くようにしてください。これらは Knit するときはファイルの置かれた場所を WorkingDirectory にしますので，プロジェクトフォルダの中にないとパスが通らない問題が頻発します。

[a] フォルダの位置を相対的・絶対的に指定してやれば，どこにおいていても読み込むことはできますが，この問題で悩んでいる人は相対・絶対パスの指定というところでさらに疑問が深まることになると思いますので，気にしないでくれて結構です。相対・絶対パスが知りたい人は，付録 A,399 でファイル場所とは何かを再確認してください)。

Q. B.2.9: ANOVA 君が読み込めません

```
file(filename, "r", encoding = encoding) でエラー:
  コネクションを開くことができません
 追加情報:  警告メッセージ:
 file(filename, "r", encoding = encoding) で:
  ファイル 'http:/riseki.php.xdomain.jp/index.php?plugin=attach&refer=
  ANOVA  君&openfile=anovakun_486.txt' を開くことができません: Invalid argument
```

A. ファイルを読み込みにいく先が URL，すなわちインターネット上になっています。ANOVA 君のファイルを一度手元の PC のフォルダにダウンロードし，そのローカルのファイルの位置を指定して読み込むようにしてください。

Q. B.2.10: 効果量として Hedges の g を算出してください。という指示について実行すると g ではなく d が出てしまうようなのですが、どうすればよいでしょうか。コードは次の通りです。

```
cohen.d(value ~ condition, data = dat, hedges.crrection = T)
```

A. Hedges の補正 (correction) のオプションが通っていません。スペルミスです。オプションのスペルが間違っているので無視されたので, 関数名通り Cohen の d が算出されます。

Q. B.2.11: 描画の際に次のような注意が出てきます。これはどういう意味ですか。

```
Removed 671 rows containing missing values (geom_point).
```

A.「671 件の行で欠損値が含まれています」ということです。つまりデータセットの中に欠損値 (観測されていない, 数値が入っていない行) が 671 件あったので, それは表示できませんでしたよ, という意味です。警告が嫌だということであれば, データセットの中で欠けているものを除外する必要があります。R の関数では, `na.omit` で欠損値を除外することができます。

Q. B.2.12: lm 関数を実行するコードでオブジェクトがないと言われます。

```
result <- lm( weight - height, data = dat)
```

A. 従属変数と独立変数とをつなげるのはチルダ (~) という記号です。そこがハイフン (-) になっています。スペルミスの一種です。

Q. B.2.13: lm 関数を実行するコードでオブジェクトがないと言われます。

```
result <- lm( weight - height )
```

A. データを与えていないので, `weight` というオブジェクトを探しに行って, 見つからないというエラーです。`data` オプションでデータを与えてあげてください。

Q. B.2.14: 因子分析の Robust 法での RMSEA の p 値って超えてたらまずいですか。Standard(DWLS) の方だけで報告はだいじょぶでしょうか。

A. 非常に専門的な質問をしておられますが，まず，「因子分析」「Robust 法」「RMSEA」など，個々の用語の意味はわかっているでしょうか。キーワードによる検索で，「ロバスト法とは」「Robust の意味」などは答えが出てくると思いますが，これらを分析法の体系的な文脈の中に位置付けないと，答えられない種類の問題です。この例をもとに説明すると，Robust を辞書で引くと「壮健」「たくましいこと」などと出てきますが，もちろんそういう意味ではありません。ロバスト法を検索すると，「統計学の分野でロバスト推定法というやり方がある」「観測値に外れ値が含まれている可能性を考え，その影響を抑えることを目的とした手法」などの説明が出てきます。ロバスト推定法は因子分析に限らず，回帰分析など他の手法でも使われる考え方なので，このような一般的な解説になります。ですがここで知りたいのは「因子分析におけるロバスト推定」ですから，因子分析でロバスト推定するとはどういうことか，因子分析の観測値における外れ値とは何か，因子分析における推定とは何を推定するのか，そもそも因子分析とは何を目的としているのか，なぜ自分は因子分析方を使うのか，さらに何故因子分析の中のロバスト推定を使いたいと思っているのか，といったことがわかっていないと，この質問に正しく回答することができないのです。このように，複合的な要素について一回で質問しても，適切な答えに辿り着けないことがあります。聞くことは恥ずかしいことではありません。知らないことを知っているふりをすることが恥ずかしいことであり，知らないまま「みんなそうやっているから」「テキスト/ネットに書いてあったから」と看過してしまうことこそ恥ずべきことなのです。ちなみに，質問に対する回答を間違えることも恥ずかしいことではありませんから，「正しく答えられなければ恥をかく (から質問しない)」というのも同様に恥ずべきことです。こうした恥ずかしさ (保身) から，「専門用語を使ってそれっぽく質問してわかった感じになろう」というのは，かえって遠回りになります。実は回答よりも，質問の仕方でその人の理解度が明らかになってしまっています。このように書くと，なんでも反射的に「わかりません，教えてください」という人もいますが，それも適切な質問方法ではありません。教員がアドバイスできるのは，「答え」ではなく「理解」が欲しい人に対してだけなのです。質問する場合は，自分は何がわかっていて何がわかっていないのか，何が知りたいのかを明確に言語化して質問するようにしてください。

付録 C

ギリシア文字一覧

ギリシア文字ってかっこいいけど，読み方わからない･･･という人のための一覧。ついでに TEX 表記も。

表 C.1　ギリシア文字とその読み方

読み方	大文字	小文字	英語表記	TEX 表記
アルファ	A	α	alpha	`\alpha`
ベータ	B	β	beta	`\beta`
ガンマ	Γ	γ	gamma	`\gamma`
デルタ	Δ	δ	delta	`\delta`
エプシロン	E	ε	epsilon	`\varepsilon`
ゼータ	Z	ζ	zeta	`\zeta`
イータ	H	η	eta	`\eta`
シータ	Θ	θ	theta	`\theta`
イオタ	I	ι	iota	`\iota`
カッパ	K	κ	kappa	`\kappa`
ラムダ	Λ	λ	lambda	`\lambda`
ミュー	M	μ	mu	`\mu`
ニュー	N	ν	nu	`\nu`
クサイ	Ξ	ξ	xi	`\xi`
オミクロン	O	o	omicron	`\mathrm{o}`
パイ	Π	π	pi	`\pi`
ロー	R	ρ	rho	`\rho`
シグマ	Σ	σ	sigma	`\sigma`
タウ	T	τ	tau	`\tau`
ウプシロン	U	υ	upsilon	`\upsilon`
ファイ	Φ	ϕ	phi	`\phi`
カイ	X	χ	chi	`\chi`
プサイ	Ψ	ψ	psi	`\psi`
オメガ	Ω	ω	omega	`\omega`

付録 D

記号の入力とキーボードの場所

プログラミングのミスでよくあるのが打ち間違い，スペルミスです。X と x，S と s など大文字と小文字で形が同じものや，l(エルの小文字) と 1(数字のイチ) の違いなどは，プログラミング用フォントにして違いがわかるようにするとか，文字列の意味から類推する (Normal とあればノーマルであって，ノーマ・イチではないと察する) など工夫が必要かもしれません。

また，理由はよくわからないのですが頻発するスペルミスは，データ (data) をデート (date) と書いてしまうことです。データはラテン語の datum(与えられたもの) の複数形なのですが，最近のデータサイエンスの文脈では data も単数形と考えるようです。ともかく，日付を表す date とは由来も意味もスペルも全て違うので，気をつけましょう。

さて，これらはまだ序の口。プログラミングのコードを読んでも，日本語の五十音に入っていない記号の違いがわからない，どこでそれが入力できるかわからない，質問しようにも読み方がわからないといったものもあります。ここではこれらをまとめて解説します。記号の上では微妙な違いですが，当然形が違うのでプログラミング上は違う文字として扱われますので，形の細部までよくみてください。なお，プログラミングでつ変わる時は言語に依存することもありますので，ご注意ください[*1]。

特に目立つのはハイフンとチルダ，アンダースコアの入力ミスです。それぞれ文字を書く領域における位置が違ったり，形が違ったりするのでよくみてください。

ハイフンは真ん中	A-B
アンダースコアは下	A_B
オーバーバーは上	A¯B
チルダはニョロ	A~B

これを踏まえて，そのほかの記号の名称や意味を表 D.1 で確認しましょう。

一般的な日本語キー配列の場合，1 つのキーに 4 つの文字・記号が割り当てられていますが，英語入力モードの場合はキーの左側，日本語入力モードはキーの右側を見ることになります。キーを押すと下の段の文字が入力されますが，シフトキーを押しながらキーを押すことで上の段の文字が入力されることになります (図 D.1)。

これを踏まえて，日本語キーについては D.2，US キーについては図 D.3 に代表的な記号とキー配列の位置を示しました。入力に困った場合は一度図を見て確認してください[*2]。

[*1] たとえばコメントアウトは C 言語では\\，R では#，TeX では % など，それぞれ異なります。

[*2] US キー配列はキートップに一種類の文字しかなく，美しい配置なのでおすすめです。

表 D.1 記号と読み方

記号	読み方	解説
:	コロン	英文中では「すなわち」などの意味。セミコロンと間違えないように
;	セミコロン	英文中では文章の区切り，接続詞のようにつかう
.	ピリオド	英文の終わりを意味する。日本語で言う句点
,	カンマ	英文の区切りを意味する。日本語で言う読点
@	アットマーク	メールアドレスに用いられることで有名
$	ドルマーク	米国の通貨単位。R では変数名指定のときにもちいる
/	スラッシュ	割り算の記号
*	アスタリスク	掛け算の記号
+	プラス	足し算の記号
-	マイナス	引き算の記号
^	ハット	累乗の計算の記号
=	イコール	等号。プログラミングでは==で一致しているかどうかの判定にも
!	エクスクラメーション	強調。プログラミングでは否定 (NOT) の意味になることも
_	アンダースコア，アンダーバー	位置に注意。ハイフンではなく文字領域の下の線 変数名をつなげる時 (ex. snake_case) に使ったりする
~	チルダ	R では独立変数と従属変数をつなぐときに使う ハイフンやオーバーライン，アンダースコアと間違えられる率高め
‾	オーバーライン，オーバーバー	アンダースコアの逆。滅多に使わないが。文字化けを直したときにみられる。
%	パーセント	プログラミングではコメントアウトの時などに使われたりする
&	アンパサンド	プログラミングにおける AND(論理積) の記号など
|	縦棒	プログラミングにおける OR(論理和)，条件付き確率の記号にも
\	バックスラッシュ	プログラミングではコメントアウトの時などに使われたりする
"	ダブルクォーテーション	文字列の開始･終了を表す。同じ記号で閉じる
'	シングルクォーテーション	文字列の開始･終了を表す。同じ記号で閉じる
`	バッククォーテーション	文字列の開始･終了を表す。同じ記号で閉じる
[]	大括弧	プログラミングでは配列を意味することがある
{}	中括弧	プログラミングではブロックの開始と終了を意味することがある
()	小括弧	数式のまとまりを表す

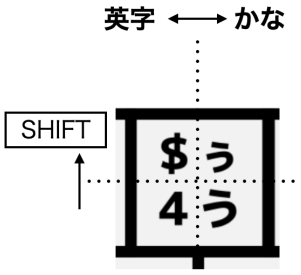


図 D.1 日本語キーで入力する場合

図 D.2　代表的な記号と日本語キー配列

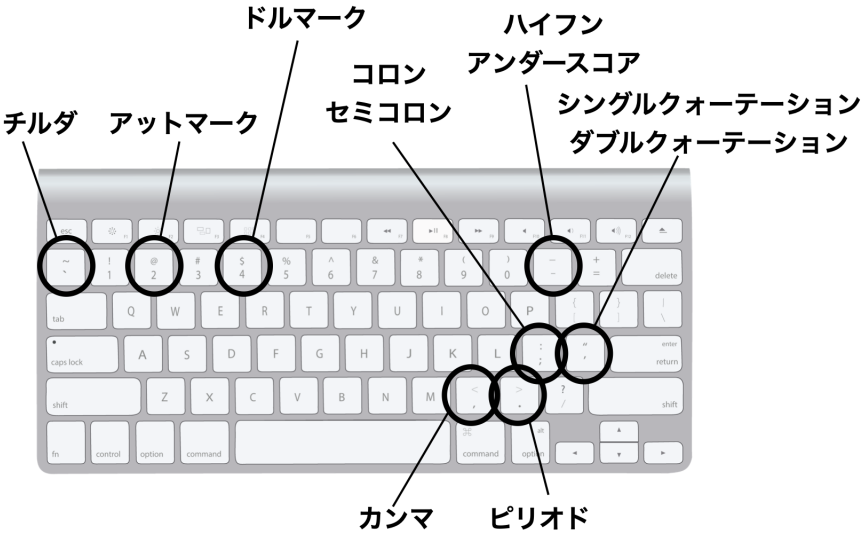


図 D.3　代表的な記号と US キー配列

付録 E

本講義に対応する詳細シラバス

E.1　心理学と統計法

E.1.1　授業内容

科目の中でのこのコマの位置づけ

記述統計 [0]　確率 [0]　モデル [0]　推測・推定 [0]　検定 [0]　実践 [0]

概要

心理学を学ぶに当たって，なぜ統計学を学ばなければならないのかを考える。そのためには，心理学がサイエンスであること，サイエンスであるとは客観的であること，積み重ねが必要であることを知る必要があり，そのために数量的表現が有用であることを理解する。

コマ主題細目

心理学とはどういう学問か　心理学は初等教育の中には取り入れられていないにもかかわらず，心や気持ちといった用語は日常的に頻繁に用いられ，かつ重要視されているものでもある。「心理学の過去は長いが歴史は短い」といわれるが，どのようにして近代科学の仲間入りをしたのかを理解するため，サイエンスとしての枠組みを改めて問い直す。

→ 三浦 (2017) の P.1–7.

心理学の 2 つの流派　心理学には基礎と応用，という区別がされることがあるが，内容やその方法論的な区分から行けば実証主義的アプローチと了解主義的アプローチの 2 つに分類する方が良い。また基礎と応用という区分では，基礎が先に，下にあって応用が後に，上にくるような印象を持つが，それらは理学と工学の違いのように，併存して然るべきものである。

→ 下山 (2001) による研究方法の分類

近代科学の特徴　心理学は科学である，というならばその科学とはそもそもなんだろうか。ここでは近代化学の特徴として実験，数学的現象主義，機械論，要素還元主義をあげる。ニュートンやガリレオが行なったのは，関数による数学的現象主義であり，心理学そのものは基本的にその枠組みの中に位置する。すなわちアウトカム変数に対して心というモデルを立てて，そのモデルの正しさを検証することが 1 つの目的である。アウトカム変数は数量化されていなくてもよいが，数量化されていた方が正確性，客観性，比較可能性，要約可能性に優れており，かつ統計モデルの活用により表面的に現れる数字か

ら見えない情報を引き出すことが可能になっている。

→ 三浦 (2017) の P.79–81. → 数値化の意味については川端・荘島 (2014) の P.1-6

キーワード

- 心理学における 2 つの検証方法
- 近代科学としての心理学
- 数学的現象主義

E.1.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 4;心理学研究法; 1. 心理学における実証的研究法(量的研究及び質的研究); A 科学と実証

予習・復習課題

■予習 心理学入門の入門書として, 道又 (2009) と中西・今田 (2020) などを一読しておくことで, これから 4 年間かけて学ぶ心理学についての事前のイメージを再構築すると良い。ほかにも, 日本心理学会が一般向けに出版している「心理学ワールド」が読みやすい入門雑誌であろう。あるいは心理学についてのイメージ調査については菊池 (2018) が詳しく, 心理学の歴史についての専門書としては大芦 (2016) なども参考にすると良い。

■復習 本学における心理学の学習が始まる前に持っていた, 心理学に対する事前のイメージをいったんわすれ, まずは広く心理学の分野について眺めた上で, 入学を志望したときに持っていた心理学についてのイメージを改めて位置づけてもらいたい。心理学のより細かい専門領域に対する位置付けを把握した上で, 心理学統計法のような研究方法が統一され共有されていること, とくに初年時の必須科目として位置付けられていることの意義を自分なりに見出せるように, ノートなどに授業の感想と現段階での今後の展望を書き留めるなどすると良い。

E.2 数値データの種類

E.2.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [1] 確率 [0] モデル [0] 推測・推定 [0] 検定 [0] 実践 [0]

概要

目に見えない心という対象を研究するために, 言語や行動によって得られた反応を記録しなければならず, 記録されたものは集計, 分析のために数値化される。数字やその統計的分析結果が, 対象を必ずしも反映していないという直観に反するかもしれないが, 数字の限界は言語の限界と同じであり, 言い換えれば論述・論考・論証の限界に等しいすべての科学的領域をカバーする。ここでは数値化することで何が得られるか, その長所を理解する。

コマ主題細目

データと変数 個人の特徴を記述するにあたって，心理統計ではすべての情報を数値化する。これらの数字は人によって変わりうるので，一般に変数と呼ぶ。また，目に見える変数を観測変数やアウトカムとよぶが，目に見えない心の状態をデータから取り出したりする場合は，それを潜在変数と呼んで区別する。

→ データの分類については，山田・村井 (2004) の P.18–21。

変数の分類 実際に心理学ではどのようなものを研究対象としているかの事例をいくつか見ることで，何が変数として使われているか，それがどういう枠組みで使われているかを理解する。心理学研究のほとんどが，予測変数/被予測変数の関係で議論されていることから明らかなように，心理学での研究関心の多くは因果的効果の検証にある。ただし，相関関係と因果関係は異なることに注意が必要である。

→ 使い方による変数の分類は，山内 (2010) の P.4-5，因果関係と相関関係の区別については三浦 (2017) の P.30–34，より詳細には豊田他 (1992) の P.24-66 に詳しい。

尺度水準による分類 すべての数字が四則演算が可能かと言われると，そうではない。尺度水準とは，数値が持つ情報や性質の段階を表す言葉であり，Steavens,S. が提唱した 4 段階の区別がとくに重要である。この尺度の四水準について理解し，得られた数字データがどの水準に該当するかを判別することができるようにならなければ，誤った統計処理をすることになるため注意が必要である。

→ 川端・荘島 (2014) の P.9–16，あるいは山田・村井 (2004) の P.22-25.

データの可視化 数量化されたデータは可視化することでその全体像がわかりやすくなる。他に，数字を見ていただけでは気づかない変数間関係が可視化されることで理解されることも少なくない。得られたデータは基本的に可視化しなければならず，その重要性および可視化に当たっても再現性を担保するため，グラフ描画の文法があることを理解する。ただし，ここでは統計環境 R のパッケージ，ggplot2 の用法に限定的に言及している。可視化の文法については，心理学業界ではまだ広く浸透した概念とは言い難いが，見た目の美しさのみならず，論理的な美しさや再現可能性の観点から考えて重要である。

→ 松村他 (2021) の 4 章に詳しい。近年では Healy (2018 瓜生他訳 2021) が一冊を通じて可視化について論じており，とくに第 1 章は心理学的観点から論じられているため必読。

キーワード

- 観測変数と潜在変数
- 尺度の四水準
- 相関関係と因果関係
- データの可視化

E.2.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 4;心理学研究法; 1. 心理学における実証的研究法(量的研究及び質的研究); C 実証の手続き

予習・復習課題

■予習 一般に, データに基づいて (Ebidence Based) 論証・考察することが主観的な思い込みから脱却する第一歩である。また常識的な知識であっても, その情報をアップデートしていかなければならない。このことについて, Rosling 他 (2018 上杉・関訳 2019) を呼んでみると良い。尺度水準については心理学や心理統計の入門的テキストでは必ず触れられる項目であり, 川端・荘島 (2014), 小杉 (2018, 2019), 三浦 (2017), および山田・村井 (2004) などを手に取って確認しておくと良い。

■復習 身の回りにあるさまざまな数字に注意を払い, それらがどの尺度水準の情報を持っているかを考えるだけで十分な復習になる。また逆に, 四つの水準それぞれの代表例を身近な例から考えて見ると良い。可視化についても, 身の回りのグラフが適切に情報を反映しているかどうか, 検証するための材料には事欠かない。とくにこれはおかしなグラフになっている, というのを発見した場合はすぐに担当教員まで連絡して欲しい (小杉は「間違ったグラフ」フォルダのなかにさまざまな悪用例をコレクションするのが趣味であり, 教材として用いることも少なくないため, 是非協力してもらいたい)[*1]。

E.3 統計環境 R の導入

E.3.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [0] 確率 [0] モデル [0] 推測・推定 [0] 検定 [0] 実践 [1]

概要

携帯電話やタブレットが誰でも身近に利用できるようになってきてはいるが, それらはあくまでも受動的にサービスを利用する媒体であり, 執筆や計算などユーザが能動的に電子機器を利用するにあたっては, パソコンの利用は不可欠である。本講義だけでなく, 心理学の研究をする上でも当然パソコンの利用は必須であり, とくに統計的データ処理をするにあたってはより深く知る必要がある。また, Excel や Numbers などの表計算でも記述統計量やグラフの作図はできるが, それ以上の「複数の変数を扱った多変量解析」や「群間の平均値の差を細かく検証する」といった心理統計的な作業には不十分なのである。そこで, より専門的なツールとして統計環境 R を導入する。統計環境 R は専修大学人間科学部心理学科が公式に採用している統計環境であり, これを活用する総合開発環境である RStudio の基本的な使い方, 結果の再現性を担保するための工夫などについても言及する。それに先立って, コンピュータの基礎知識についても解説する。

コマ主題細目

パソコンのファイル構造 あらゆる媒体が情報という一元化された単位で扱われるのは便利なことではあるが, 一見してその違いがわからないのは不便である。そこでファイル名には拡張子と呼ばれるファイルの種類を区別する情報を付与する。また, ファイルがたくさん出来上がって管理できないのも困るので, いくつかのファイルをまとめるフォルダ/ディレクトリという単位でまとめて置くことを理解する。また

[*1] mailto:kosugi@psy.senshu-u.ac.jp

これらのファイルがどこにあるかをしっかり理解する必要がある。とくにクラウド/ローカルの違いについて注意しておく。

統計環境 R/RStudio の準備　心理学科では基本的に統計環境 R を共通のツールとして用いる。統計環境 R はフリーソフトウェアであり，誰でも無償で利用できる環境である。また，R 単体では簡素な言語的やりとりができるに過ぎない。ファイルの操作や描画機能，記録，参照，管理など統計言語を扱うに当たって必要な総合的環境として RStudio を併用することで，その利便性は飛躍的に向上する。まずは R/RStudio を導入し，基本的な画面構成やプロジェクトによる管理ができることを確認する。

→ R についての入門書は多くあるがここでは小杉 (2019) の P.1–7 を挙げておく。同じく RStudio を使った入門書は多くあるがここでは小杉 (2019) の P.7–24 を挙げておく。

R の基本操作　環境が整ったところで，R の基本的な操作を見ていく。まずは R スクリプトをつかって，四則演算や関数計算，データの形式やパッケージの導入と利用について理解する。スクリプトペインとコンソールペインの違いを理解し，またスクリプトとして清書しておくことで計算手続きの記録がつけられることの利点を理解する。

再現性のある研究実践へ　今後，パソコンを使って文章のみならず，図表も作成することになる。統計的分析結果の出力も，プログラムによって行うことになる。コンピュータにおいてプログラムが実行されるとはどういうことか (適切なファイルへのアクセス) を理解し，またプログラムを書いて実行するときの注意点について理解する。とくにエラーとその報告の仕方については，恐怖心に駆られることなく対応するべきである。そして研究の実践とその成果が「動作」ではなく「指示書」として記録に残ることは，研究の再現性を高めるためにも重要な営みであることを理解する。Git による文書のバージョン管理や，OSF など情報の共有など，最新の研究実践方法についても言及する。

キーワード

- ファイル，フォルダ，拡張子
- R と RStudio
- プロジェクト
- 再現性のある研究実践

E.3.2 授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ　科目番号 5;心理学統計法; 2. 統計に関する基礎的な知識; C エクセル，R，SPSS 入門 (1) 統計分析のための言語の手ほどき　プログラムの基本的操作

予習・復習課題

■予習　大学生活では PC を使うことがほぼ確実に必要であり，今後の社会においてもその能力は間違いなく生きてくるので，なるべく早い段階で PC を所持し，ブラインドタッチなど基本的な技術を身につけておく。

■復習　自分の PC のユーザ名，ファイル名，拡張子の付け方などを確認し，半角英数になっていないか確認しよう。今後ファイルを作ることがあれば，半角英数文字で名前をつけるよう癖づけておいたほうが良い。これらは未然にエラーを防ぐことにもなる。

E.4 統計環境 R の利用

E.4.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [2] 確率 [0] モデル [0] 推測・推定 [0] 検定 [0] 実践 [2]

概要

前回に続き，R を使った演習を行う。今回はの分析結果と文章化を 1 つのファイルでまとめて利用できる，Rmarkdown というファイル形式について学ぶ。Rmarkdown を利用することで，分析手続の記録と結果の出力を一元管理することができる。具体的なデータ分析に先立って，外部ファイルの読み込みについての演習を行う。ファイルの場所，種類 (拡張子)，エンコードなどの基本的な理解ができているかを，ここで改めて確認する。今後，データを表計算ソフトなどで作成する機会も増えてくるが，それに向けてデータの入力，分析しやすいデータ形式である整然データの解説を行う。最後に，グラフなども表計算ソフトではなく，描画の文法に乗って行うことを `ggplot2` パッケージを利用しながら確認する。

コマ主題細目

Rmarkdown による動的な文書作成 RStudio を R の実行環境として捉えるだけでなく，R のもつ豊かなパッケージ群と RStudio のもつ統合的な環境とを合わせて活用することで，文書作成を 1 つの環境の中で統合的に扱うことができるようになる。Rmarkdown という R の書式の 1 つについて理解し，データの処理と報告書の作成を統一された環境で行うことで，複数の環境を跨ぐことによって生じる問題 (ex. コピペ汚染) から脱却し，再現性のある実行環境を維持する取り組みについて理解する。

→ 小杉 (2019) の P.171–188。再現可能性も視野に入れたより専門的な参考図書は高橋 (2018)。

外部データの読み込み 初学者が最もつまづきやすい点のひとつが，外部データファイルの読み込みである。つまづきやすい点として，ファイルが指定した場所にないこと，ファイルの形式が適切でないこと，ファイルのエンコードが異なることなどが挙げられる。プロジェクトで管理されたフォルダにおいて，CSV 形式ファイルを UTF-8 形式で読み込むことを徹底する。

整然データ データを分析する際には，データの持っている尺度水準と特定の値に紐づけられた情報が，行と列の指定で一意に特定できることが重要である。このようなデータのことをとくに整然データという。整然データを目標としたデータの入力の必要性を理解する。また不適切に入力されたものであっても整然データの形に整形するための技術が必要であり，こうしたデータ加工の基本的な操作について説明する。

→ 整然データについては https://id.fnshr.info/2017/01/09/tidy-data-intro/ がわかりやすい。

ggplot2 によるグラフの文法 大量のデータが整然とした形に整えられ，そこから情報を汲み取る時にまず行うのが可視化であり，データの要約統計量による理解である。ここでは先のデータの整然化を経て，理解を進めるグラフを描画するための文法について理解する。R の ggplot2 パッケージをもちいて，ヒストグラム，ボックスプロット，散布図の描画，それらの群ごとの色分けや掻き分けなどがスクリプトによって実現できることを理解し，一度描画言語で記述することで再現性を高めることができることを理解する。

→ 松村他 (2021) を参照

キーワード

- プロジェクト管理
- CSV ファイル
- UTF-8 形式
- 整然 (tidy)data
- ggplot2

E.4.2 授業情報

■コマの展開方法 演習; 4 号館 4F の PC ルーム 1/2 に別れて, 実際の PC を使う実習を伴った授業を行う

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法; 2 統計に関する基礎的な知識; C/D/E エクセル, R, SPSS 入門

予習・復習課題

■予習 受講生の中にはまだ, Excel などの表計算ソフトを実際に利用・活用したことがないものがいるかもしれない。住所録や家計簿といった身の回りの情報であっても, 数値化して入力されたものを一覧にし, 必要な関連キーワードによってまとめ, 特徴を描写するなどの方法は, 高度に情報化された現代社会では基本的な素養の 1 つである。身の回りでのデータの活用例について事前に情報収集し, 本時で学ぶ内容を即座に活用できるような側面についていろいろ考えておくことも予習になる。また有償であるソフトウェアは, 品質保証と引き換えにその内部構造が企業秘密＝ブラックボックスであり, そのことを問題視するフリーソフトウェアという考え方がある。フリーソフトウェアについてはストールマン, 2002 長尾訳 2003 に詳しく, 事前知識として一読すると良いだろう。

■復習 R および RStudio を使った実習については, 大学の PC 環境で利用できるのはもちろんであるが, フリーソフトウェアなので個人所有の PC にインストールすることができる。小杉 (2019) などを参考に, 手元の環境で実行環境を構築することで, 今後の分析やレポート作成に向けた準備にもなる。授業時間内に課される課題を Rmd 形式で提出することが求められる。時間内に学んだ技術を用いて, レポートを提出すること。なおこのレポートの提出は前期の単位認定にあたっての必要条件であり, 基準に満たないレポートは再提出を求める。前期試験の前にパスすることが, 前期末試験を受験するための前提条件であることを承知しておくこと。

E.5 記述統計量

E.5.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [0] モデル [0] 推測・推定 [0] 検定 [0] 実践 [2]

概要

統計の基本は集めたデータを集計し，要約統計量で表現することである。手元にあるデータの詳細を記述するため，これは記述統計 (descriptive statistics) と呼ばれる。いかなる進んだ分析を行うに当たっても，必ず可視化および記述統計による特徴の把握を欠かすことはできない。また要約統計量はその性質上，必ず有意味な情報を捨て去って代表的な性質を表すものであり，単一の指標だけで考えることは危険であることを理解し，活用できるようにする。

コマ主題細目 (3-5 項目箇条書き)

中心化傾向の指標 散らばりを持つデータの中心についての要約統計量として，平均値 (mean)，中央値 (median) , 最頻値 (mode) の 3 つがあげられる。それぞれの特徴および算出方法，その際の注意点についても理解する。また，計算や表現の利便性のため，$\sum$ という総和記号を用いる。

→ 山内 (2010) の P.27–35, 山田・村井 (2004) の P.30–33, 川端・荘島 (2014) の P.23–26.

散らばりの指標 データにとって重要なのは中心化傾向の指標だけではなく，むしろその散らばりについての指標である分散 (variance), 標準偏差 (standard deviation) こそ重要である。加えて最大値 (max)，最小値 (min) , 四分位偏差 (quantile)，パーセンタイル (percentile)，範囲 (range) などが散らばりを表す指標として用いられる。ここではそれぞれの特徴および算出法について理解する。

→ 山内 (2010) の P.37–48, 山田・村井 (2004) の P.34–37, 川端・荘島 (2014) の P.26-33.

標準化 平均値と標準偏差を導入したことで，データ点の相対的位置を表現することが可能になった。心理学の研究対象の多くは，絶対ゼロ点をもたない間隔尺度水準であり，個々のデータ点の特徴を記述する際には相対的な位置である標準得点で表されることが多い。ここでは標準化の方法と標準得点の特徴，標準得点の応用である偏差値の特徴および算出方法について理解する。

→ 山内 (2010) の P.54–56, 山田・村井 (2004) の P.38–41, 川端・荘島 (2014) の P.36–40.

標準化された差 心理学の多くは単位を持たない測度で考えるため，標準化されたスコアすなわち単位のない測度での比較検討を行うことが多い。これが差の効果量の基本的なアイデアなので，ここで一瞥しておく。

→ 効果量については大久保・岡田 (2012) を参照

キーワード

- 平均値，中央値，最頻値
- 分散，標準偏差，最大値，最小値，四分位偏差，パーセンタイル，範囲
- 標準得点，偏差値
- 標準化された差，効果量

E.5.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法; 1. 心理学で用いられる統計手法; A 記述統計: 代表値と散布度/B 記述統計:相関

予習・復習課題

■予習 統計分野は 2012 年 (平成 24 年) より, 高校数学の必修である数学 I で新たに扱われることとなっているため, 受講生諸君はすでに算術平均や相関係数について一度は習っていることと思われる。内容について自信がない場合は, 高校数学のテキストを読み返し, キーワードにあげられている要約統計量について予習をしておくこと。また心理統計は統計学のユーザでもあるので, PC を使ってこれらの指標を計算できるように準備, 練習しておくことが望ましい。

■復習 今回の内容については, 心理統計の各種テキストの中に必ず含まれている内容であるから, 類書の当該箇所について自分なりの復習をしてもらいたい。山田・村井 (2004) はユーザー目線から丁寧な記述がされている良書であり, とくに以降の推測統計については慎重かつ緻密な言葉遣いで記述されていることから, 日本の多くの心理統計業界でのテキストとして愛用されている。川端・荘島 (2014) は山田・村井 (2004) 以上に丁寧かつ細部にまで言及され, 言葉遣いも柔らかいので大変読みやすく感じる。とくにこの本はシリーズ化されていて, 研究領域ごとに特化した統計分析がテーマになっている[*2]。山内 (2010) は数学的記号が多用されており, 敷居が高く感じる読者もいるかもしれないが, 一冊で網羅的に記述されており相関・回帰を先に導入するなどモデルベースの観点から記述されている点が特徴的えだる。

数学的表現は難解に思えるかもしれないが, 必要十分な情報がすべて盛り込まれているし, 何より積み重ねの学問であるので, 不明な点が出た場合は定義に戻って確認しておくことを習慣付けよう。また記号だけでの理解ではなく, 実際に数点のデータをとって, 各統計量を手計算, あるいは PC を使って算出する練習をしておくと良い。記号 $\sum$ に関する例題としては, 山内 (2010) の P.10–13 が適当である。

E.6 測定の基礎

E.6.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [0] モデル [1] 推測・推定 [0] 検定 [0] 実践 [2]

概要

ここでは心理学がいかにしてデータを得るかについて, 心理学的測定論の話に触れる。心理手学的測定は古典的テスト理論に始まる。受講生はもちろん多くの人間が「テスト」で学力その他を測定された経験があると思うが, そもそも良いテストとはどのようなものか, 良いテストの項目とはどのようなものか, なぜテストをすることで学力等が正しく測定できると考えるのか, といったことについて体系的に習うことは少ない。テストは目に見えない学力を測定するためのものであり, 質問という項目に対する反応からそれを測定しようとするという営みであると考えれば, 同じく目に見えない心を測定し, 刺激に対する反応からその実態の存在を類推しようとする心理学との類似性・親近性は自ずと理解されるところである。

[*2] なお, このシリーズの各巻末にある著者紹介は大変ユニークであるので是非順番に読んでみよう。

コマ主題細目 (3-5 項目箇条書き)

テストと心理学 心理学の研究の基本は S→R 図式であり，関数主義による実態の把握であった。心理学者はもちろん，一般に我々が想定する「心」も，一定の反応傾向を見せる個人という実体の内部モデルとしてである。人間の性格についての歴史的変遷 (類型論から特性論へ) を踏まえ，質問紙調査などの研究法によってアプローチしようとしている研究の基本的枠組みを，テスト (測定) という観点から改めて捉える機会とする。

→ Grimm・Yarnold (2001 小杉他訳 2016) の P.20–25.

測定のモデル 古典的テスト理論は $X = t + e$ というごく単純な数式で表現されるが，ここから導出される含意は非常に大きいものがある。観測値をそのまま用いず，潜在変数としての真値 t と誤差 e の統合されたものとしてみるという視点は，ハードサイエンスのように測定の精度が高い科学領域や，経済学や商学などのように 1 円のズレも認められない社会科学領域とは違い，「測定はされたものの正しいかどうかは即断できない」という非常にあやふやな事実に立脚したソフトサイエンスたる心理学の，良心であるとも言える。

→ 熊谷･荘島 (2015) の P.10-12.

テストの平均値と真値の関係 古典的テスト理論における測定に含まれる誤差は系統誤差と偶然誤差に分けることができる。系統誤差は取り除くべき問題であり，これが適切に対応されたと仮定すると問題は偶然誤差の振る舞いである。偶然誤差の平均は 0 であると考えられるから，複数の項目を測定することで真値が得られると考えられる理論的根拠が導出される。ここではこの過程を確認するとともに，類似した項目を幾度となく繰り返して測定する尺度がなぜそのような形式を取ることになっているのか，その原理について理解する。

テストの分散と信頼性・妥当性の関係 古典的テスト理論のモデルを分散について同様の式展開を行う。ここでは誤差の新たな性質である，他の何者とも相関関係にないことから，テストで得られる分散が真の分散と誤差の分散「だけ」に分割されることを理解し，テストにおける信頼性の定義を行う。また目に見えないものを測定しようという営みである限り，信頼性だけでなく妥当性についても考えなければならないが，数値化できる信頼性とは違って妥当性は論証するしかないこと，信頼性は妥当性の上限であることを理解し，心理学における測定の重要性を再確認する。

→ テストの信頼性については上述の Grimm・Yarnold (2001 小杉他訳 2016) の P.20–25。

キーワード

- 古典的テスト理論
- 偶然誤差と系統誤差
- 信頼性
- 妥当性

E.6.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 4;心理学研究法; 1. 心理学における実証的研究法(量的研究及び質的研究); C 実証の手続き

予習・復習課題

■予習 これまで受けてきた試験のなかで, 良問, 悪問はどのようなものであったか考えてくると良い。翻って, 一番良いテスト項目とはどのようなものであるかについて検討できれば, 一歩進んだ予習になる。また, アキネータ (https://jp.akinator.com/) と呼ばれるインターネット上の遊び, 古くは 20 の扉 (20 Questions) などに触れ,「質問によって正統を導き出す」ための, 効率の良い問いの立て方について考えるだけでも予習になる。

■復習 $\sum$ や文字を使った平方の展開など, 基本的には中学数学程度の知識でフォローできる知識範囲であるが, 自分自身で式展開をフォローできるかどうか再確認しておくこと。とくに古典的テスト理論における平均や分散から導出される意味については, (非常に容易なため) テキストなどには逐一記載がないが, わかったふりをするのではなくしっかりと自分で再生できるか確認すると良い。

心理学をはじめ人文社会科学においては, 純粋な数学力というよりは, 数式のもつ意味を理解できるかどうかが肝要である。単純な式展開の中に多大な意味が含まれていることを理解すると, 数学が 1 つの言語であるということが実感できると同時に, 今後の心理統計における数式に対しても嫌悪感を低減することができる。

E.7 誤差と確率

E.7.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [1] 推測・推定 [0] 検定 [0] 実践 [2]

概要

前回は偶然誤差の平均について注目したが, どのように振る舞うのかを確定できない変数について対応するために, 我々は確率分布を利用する。統計学と確率は切っても切れない関係にある。記述統計ではデータの中心的な位置, ばらつきの幅を記述することができたが, それらがどのような規則性を持つかを記述するのが確率分布の考え方である。確率の基本的な性質と, 生じうる結果の一覧としての確率分布を理解する。最も単純な二項分布から始め, その施行回数を極限まで増やしたものとして正規分布を位置付ける。

コマ主題細目

確率分布 確率という言葉は不確実な事象を表現するための表現であることを確認し, コルモゴロフの公理に従う数字という事実のみまず確認する。その上で確率的な事象と見立てて確率の言葉で記述するとき, 起こりうるすべての可能性を網羅的に表したものを確率分布ということなど, 確率変数や確率分布の表現について理解する。離散変数の場合は確率質量関数とよばれる, すべての起こりうるカテゴリに付与された確率の一覧がその確率分布となる。連続的な分布の場合は確率密度関数と呼ばれる関数に変わることに注意が必要である。連続変数の場合, 一点の確率は定義できず, 区間の面積が確率の数字になる。

→ Kruschke (2014 前田・小杉監訳 2017) の P.83–88

二項分布 確率分布の中でも最も単純なのがベルヌーイ分布であり，それを拡張したのが二項分布である。離散変数の代表的な例としてベルヌーイ分布と二項分布について，数式や表記法とともに理解し，また確率分布の平均値である期待値，確率分布の分散を表す方法を学ぶ。

→ 野島他 (2019) の P.66-68

正規分布 測定に誤差が含まれている可能性については，ティコ・ブラーエ，ケプラーなどによって，天文学における膨大なデータと，それに対する天体運動の数式の当てはめを考えていた頃から懸念されていた問題であった。ガウスは誤差のありうるべき形を考え，そこから正規分布を導出しており，これがひいては総合的・社会的な分布として利用されるようになる。ただしここではあくまでも，誤差分布としての正規分布について解説する。

→ 長沼 (2016) の P.16–40，あるいは少し数学的な説明が長くなるが，西内 (2017) の P.408–430

標準正規分布 さまざまな現象が正規分布に従うとしても，平均値や標準偏差が異なるのであれば相対的な比較はしにくい。しかし素点を標準化しておくと，標準正規分布に照らし合わせることで推測をすることができる。確率密度，確率点，累積確率，などについて数字の表や統計環境 R をもちいた例を示しつつ，標準正規分布の活用の仕方について理解する。

→ 川端・荘島 (2014) の P.36–44，山田・村井 (2004) の P.86–89。

キーワード

- 確率
- 確率分布
- ベルヌーイ試行，ベルヌーイ分布，二項分布
- 正規分布

E.7.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法；1. 心理学で用いられる統計手法；D 離散分布：二項分布を中心として
科目番号 5 心理学統計法；1. 心理学で用いられる統計手法；E 連続分布：正規分布を中心として

予習・復習課題

■予習 これまで習ってきた高校数学の，確率・統計に関する記述を改めて読み直してこの授業への準備としよう。確率についての数学は比較的歴史が浅く，また専門家でも間違えることの多い学問領域であるが，同時にギャンブルや保険の利率計算など生活に密着した数字でもある。副読本として読み物浜田 (2018) や Kruschke (2014 前田・小杉監訳 2017) を挙げておくので，最初の数章だけでも目を通しておくと心構えができるだろう。

■復習 正規分布の数式は大変複雑な形をしているように見えるが，その本質は「左右対称」かつ「単峰」であることである。この点については長沼 (2016) が言葉による解説を行なっており，数式に苦手意識を持っていてもわかりやすいので一読するとよい。連続変数における確率は確率密度として表現されることについて，

Kruschke (2014 前田・小杉監訳 2017) の P.84–88 を読みながら復習すると良い。最後に身近なデータをつかって，変数の標準化と標準正規分布における相対的な位置の推測の仕方について，確認しておこう。計算に際しては Excel/Numbers/Calcs などの表計算ソフトが持っている確率密度関数や，統計環境 R のそれを使うとよい。

E.8 相関と因果

E.8.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3]　確率 [1]　モデル [2]　推測・推定 [0]　検定 [0]　実践 [2]

概要

これまでは 1 変数の記述や，潜在変数の説明をしてきたが，ここからは複数の変数について考えていくことになる。変数同士の関係は，共分散やその標準化した相関係数で記述することができるが，相関係数は直線的関係についての指数であることや，表面上の相関関係が必ずしも本質を反映していない可能性があることに注意が必要である。相関関係と因果関係の違いを注意深く理解し，交絡変数に影響された擬似変数になっていないか，といったところを注意する必要がある。複数の具体的な例を提示しつつ，因果関係と相関関係の違いや，スナップショット的研究では相関関係しかわからないこと，実験/介入の効果を見るためにどのような点に注意しなければならないかといった研究法の一部に言及しつつ理解を深める。

コマ主題細目

変数間関係の指標　ここまでは 1 つの変数についての特徴を記述するものであったが，次に複数の変数による変数同士の関係を考える。変数間関係の指標としては，順序相関や連関係数なども含まれるし，クロス集計表も変数間関係を表すものである。中でも連続的な変数に対する指標としての共分散，それを標準化したスコアにしたものがピアソンの積率相関係数であることを理解する。また相関係数はあくまでも変数間の直線的関係についての記述であり，相関係数だけを見て判断すると対象の理解を誤る場合があることに注意する。

→ 山内 (2010) の P.62–75, 山田・村井 (2004) の P.44–59, 川端・荘島 (2014) の P.48–58.

相関の落とし穴　見かけ上は相関があるように見える例として擬似相関がある。他にも，打ち切りデータや層別の問題など，相関係数の大小だけでは気づかない点も多い。具体的な例を参照しつつこれらの問題点を指摘し，またデータを入手した時に散布図を複数書くことで直感的な判断ミスを避けることができる。

→ `datasauRus` パッケージには，記述統計が同じだがプロットがまったく異なる 12 のケースが含まれていて参考になる。可視化の重要性については Healy (2018 瓜生他訳 2021) を参照。

因果関係　相関関係があるところに，因果的な説明をしてしまいがちであるが，因果的な説明をするためには相関以上に注意が必要である。因果関係を見るために必要な，時間的先行性，他の原因がないこと，他の結果に至らないことなどの諸条件を確認する。マスコミなど周りの例から，相関と因果を取り違えて判断している悪例を示す。

→ 相関と因果の違いについては，豊田他 (1992) に簡潔にまとめられている。

交絡 第三の隠れた変数によって，見かけ上の相関が生じている可能性もある。こうした変数のことを交絡変数と呼ぶ。調査研究などの場合は，関係するすべての変数を集めて統計的に統制することを考えないと，交絡変数の影響で結果が正しく評価できなくなることに注意が必要である。また現実問題として，実験的に統制できないことが多いことにも注意が必要である。

→ 因果効果を考えるための導入として安井 (2019) を参照。

キーワード

- 相関係数
- 相関関係と因果関係
- 擬似変数
- 交絡要因

E.8.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法; 1. 心理学で用いられる統計手法; B 記述統計:相関

予習・復習課題

■予習 分散，標準偏差，標準化の手続きについて再確認しておくこと。分散は自分自身との共分散という形で理解することができるし，共分散を標準化したものが相関係数になるからである。

■復習 我々がアクセスできる情報の中にも，相関関係を因果関係のように捉えて類推し，ミスリーディングな表現になっているものも少なくない。身の回りのこうした事例をひとつ挙げ，どのように間違いがあるか，正しくはどのように考えるべきかについて，論じてみよう。これはデータリテラシーの問題でもあるし，統計法の初歩的問題でもある。

E.9 単回帰分析

E.9.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [3] 推測・推定 [0] 検定 [0] 実践 [2]

概要

古典的テスト理論から測定に含まれる誤差を知り，誤差が正規分布に従っていると考えることで，平均の位置を推定することができるようになった。ここでは他の変数との関係から，平均の位置が関数で表現される最も単純な場合，線形モデルについて考える。

コマ主題細目 (3-5 項目箇条書き)

相関係数とモデル　記述統計の中で学んだ複数の変数間関係を記述する共分散，相関係数について再確認する。とくに相関係数が線形関係の指標になっていたことに留意する。しかし逆に考えれば，線形関係が見出せるような散布図であれば，我々が「見出している」関数がどのような形であるかを考えることができる。

→ 小杉 (2018) の P.51–52.

予測と線形モデル　変数間関係を記述する最も単純な形として，一次関数を用いたものを考える。現象に対して関数で記述する数学的現象主義，変数を説明変数と被説明変数として捉えること，予測に誤差が加わって現象となることなど，これまでの議論をデータと数式に置き換えただけであることをしっかりと理解する。さらに単回帰分析の数式的記述について理解する。とくにデータ (X_i, Y_i)，予測値 $(\hat{Y})$，誤差 (e_i)，回帰係数 (b_0, b_1) として表される記号の細部にまで注意を払い，その意味を理解する。

→ 豊田 (2017) の P.27–28.

最小二乗法　最小二乗法によって推定値が算出できることを理解する。数式の細かい展開はここでは行わず，最小化する目的関数の直観的理解に止める。また算出された係数 (傾きと切片) がどのような意味を持つのか改めて理解する。

→　豊田 (2017) の P.30

残差と決定係数　回帰係数が計算されれば，実際のデータに対して予測値を計算することができ，予測がどの程度適合していたかを検証することができる。予測値と被説明変数との差分を残差 residuals と呼ぶが，この両者の相関係数を重相関係数，その二乗したものを決定係数と呼んでモデルの適合度を測る指標にすることを理解する。

→ 野島他 (2019) の P.58-60.

キーワード

- 一次関数
- 予測値，誤差，回帰係数
- 最小二乗法

E.9.2　授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ　科目番号 5 心理学統計法；1. 心理学で用いられる統計手法；C 記述統計：回帰

予習・復習課題

■予習　共分散および相関係数の算出方法について，第 3 回の講義を参考に復習しておくこと。

■復習 山内 (2010) の P.62–81 は相関係数と回帰分析を 1 つの章にまとめて解説しているので，予習と復習の両方に適した参考書である。この本は続く章に母集団と平均の言及がある，すなわち回帰分析の最小二乗法による解の段階では，記述統計の範囲で議論できることではあることにも気づかされる。最小二乗法による回帰係数の算出については，偏微分連立方程式を用いれば簡単に計算できるが，偏微分連立方程式とは何かに遡って解説している本当して小杉 (2018) の P.104–112 がある。また最小二乗法に頼らずに算出することも，手数はかかるが不可能ではない。その場合については西内 (2017) を参考のこと。概念的理解が困難に思える場合は，小杉 (2018) の P.54–57 を参考に実際に R で計算して確かめてみると良い。

E.10 重回帰分析

E.10.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [4] 推測・推定 [0] 検定 [0] 実践 [2]

概要

回帰分析によって，データにモデルを当てはめて係数の推定値を得るということを行った。データにもっとも適した形を求めて計算を行ったが，そのこととモデルが適していたかどうかは別である。この点を理解するためには，モデルの適合度について考える必要がある。モデルが適合しているということを踏まえた上で，説明する変数が増える重回帰分析へと議論は展開する。ここでは説明変数の読み取り方への注意を促すべく，偏相関係数の考え方を解説し，偏回帰係数，標準化偏回帰係数などの用語について理解する。最後に，多重共線性の問題についても指摘する。

コマ主題細目 (3-5 項目箇条書き)

重回帰分析 説明変数が複数に増えた場合の回帰分析は，重回帰分析 Multiple Regression Analysis と呼ぶ。数式的な表現としては項が増えるだけであるが，データのプロットを考えると三次元以上の回帰平面に拡張されていること，それでも面は湾曲していない，すなわち線形関係であることを理解する。

→ 小杉 (2019) の P.168, 図 10.7 の回帰平面図を参照

部分相関と偏相関 続く偏回帰係数の理解に先立って，部分相関の考え方を理解しておく必要がある。回帰分析によって，被説明変数の分散が説明される部分と残差の部分に分離させられること，残差と説明変数は相関しないことを確認し，残差は説明変数の影響を除外したものと考えることができる。これは条件を必ずしも統制できない事象において，統計モデル的に統制していることでもある。これを踏まえて，変数間関係を改めて理解する。

偏回帰係数 部分相関の説明において，残差が分離されることを見てきたが，残差同士の相関関係はとくに偏相関ということができる。これは共通する変数からの影響を除いた変数間関係であり，これを見るだけでも変数間関係を見ることにつながる。とくに残差変数を使っての回帰分析を行うとき，得られる係数が偏回帰係数である。重回帰分析における回帰係数はこのように条件付きで考えなければならない。また変数を増やすことで重相関係数の増加することを理解する。

→ 小杉 (2018) の P.60–63.

標準化された係数 回帰係数については，素点による回帰係数が理解しやすいが，単位に左右されるもの

でもあるので相対比較をする場合はすべての変数を標準化した標準化解を用いた方がわかりやすい。標準化解を用いても適合度は変化しないことなど留意点とともに有用性を理解する。

→ 小杉 (2018) の P.66–67.

キーワード

- 残差
- 決定係数
- 重回帰分析
- 偏相関係数, 偏回帰係数
- 標準化解

E.10.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法; 1. 心理学で用いられる統計手法; C 記述統計:回帰

予習・復習課題

■予習 説明変数が 1 つの回帰分析について, 数式的表現や最適化関数 (最小二乗法) の意味について概念的な理解で十分なので再確認しておくこと。後者について今回も同じ基準が用いられるが, 説明変数が複数になる場合は解析的算出には行列計算の知識が必要である。行列計算については, 二年次の統計に関する講義で扱う予定であるが, 興味があるものは「線形代数」をキーワードに予習すると良い。初学者には結城 (2018) が, あるいは岡太 (2008) がよい。

■復習 豊田 (2017) の P.89–98 は本時の内容がわかりやすくまとめられているので, 通読しておくと良い復習になる。また重回帰分析は広く用いられる手法なので, Grimm・Yarnold (1994 小杉他訳 2016) の第二章などを読んで, どのような使い方をされているのか, また前提となっている条件や利用上の問題点などについて, 講義時間内で触れられなかった諸問題についての知識を補完しておいてもらいたい。

E.11 R をつかった回帰分析

E.11.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [5] 推測・推定 [0] 検定 [1] 実践 [3]

概要

回帰分析や重回帰分析について, 基本的な概念や用語・推定値の意味は前回までに習得済みであるが, それが統計環境では実際どのようにして算出されるか, 算出された数字はどのような意味があるかについて具体的に理解することを目的とする。座学の中では記号としてしか意味を持たなかったものが, 実際の数字として得られるとより理解が深まることにもなる。統計環境は R を用いるが, SPSS そのほかの統計パッケー

ジを用いても同様の出力が得られる。統計環境が変わる可能性もあるので，出力画面の位置で内容を覚えるのではなく，意味的な理解をしておく必要がある。

コマ主題細目

散布図の描画 線形回帰はその名の通り線形関係をモデルで表現するものであり，非線形な関係や外れ値の存在などがあれば適切な推定値が得られたことにはならない。分析をする前にまずデータの様子を確認するために，散布図の描画は必ず行わなければならない。第 E.4 講の復習を兼ねて，ggplot による散布図の描画を行う。

→ 松村他 (2021) の P.126–138.

線形回帰の実行 データが読み込まれれば，lm 関数によって最小二乗法による回帰分析を行うことができる。R による関数関係の記述 (formula) の方法を理解し，出力結果から適切な読み取りができるよう，どのような用語でどのような出力がなされているかをしっかり確認する。

→ 小杉 (2018) の P.54–60.

残差のプロット lm 関数から返される結果オブジェクトの中には，予測値 $\hat{Y}$ が fitted.values，残差が residuals という変数名で保存されている。これらを取り出したり図示したりすることで，回帰分析の諸特徴を理解する一助となる。数理的な導出は小杉 (2018) の P.127–135 にあるが，具体的な数字としてこれを確認し，理解する。

→ 小杉 (2018) の P.54–60.

重回帰分析の実行 統計環境において重回帰分析を行うのは，説明項を 1 つ追加するだけで良いので作業としては容易い。ただし解釈に注意が必要であることを再確認し，また標準化偏回帰係数の算出方法についても理解を進める。また変数を増やす前と増やした後で，重相関係数が増加することを確認する。

→ 小杉 (2018) の P.63–67.

キーワード

- ggplot2
- lm 関数
- formula 表記
- fitted.value と residuals
- 標準化偏回帰係数

E.11.2 授業情報

■コマの展開方法 演習；4 号館 4F の PC ルーム 1/2 に別れて，実際の PC を使う実習を伴った授業を行う

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法；2 統計に関する基礎的な知識；C/D/E エクセル，R，SPSS 入門

予習・復習課題

■予習 第 E.4 講と同じ教室・形式で行われる。R や RStudio, プロジェクトによる管理など基本的なことを改めて教示しないので, 前回の復習をかねて空いている時間に PC ルームで基本的な R/RStudio の挙動について理解しておくと良い。また R の内部でどのように数字やデータが扱われるかについて, 小杉 (2019) の P.24–36 などを参考にみておくと良い。

■復習 残差をデータの中で確認し, 部分相関係数や偏相関係数, 偏回帰係数がどのように算出されるか自分自身で確かめてみると良い。また, 授業時間内に課される課題を Rmd 形式で提出することが求められる。時間内に学んだ技術を用いて, レポートを提出すること。なおこのレポートの提出は前期の単位認定にあたっての必要条件であり, 基準に満たないレポートは再提出を求める。前期試験の前にパスすることが, 前期末試験を受験するための前提条件であることを承知しておくこと。

E.12 実験計画法 1；実験計画法とは

E.12.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [6] 推測・推定 [0] 検定 [1] 実践 [3]

概要

心理学の研究の多くは, 従属変数に影響を及ぼす要因をコントロールし, 統制群と比較することで平均因果効果を検証するところにある。心理学研究の具体例を基盤に, 実験計画の種類, 用語を確認し, 平均差の効果は誤差に対する相対的な大きさで検討される原理を理解する。心理統計のテキストの多くは「分散分析」の文脈でこのことを論じるものがほとんどであるが,「線形モデルであること」「誤差と効果の対比であること」という 2 つの観点から読み解き, 古典的テスト理論と同じ形の, 関数で表現し誤差を分離するという心理統計の基本方針を見出さなければならない。

コマ主題細目

ランダム化 平均因果効果を検証するためには, ランダム化による比較, いわゆるランダム化比較実験 (RCT,randomized controlled trial) が必要である。RCT によって誤差をコントロールすることができるようになることを理解する。

→ 西内 (2013) の P.101-107, 大村 (2013) の P.17–40.

「要因」と「水準」 実験計画法で用いられる用語,「要因」「水準」「群内計画」「群間計画」といった用語の使い方を整理する。とくに心理学系の論文からどのような計画で行われた実験であったかを読み取ったり, 自分自身が今後立案する場合はどのように考えれば良いか, 具体的な例に即して理解する。

→ 山田・村井 (2004) の P.174-177, 大村 (2013) の P.48–52.

誤差と効果 RCT デザインをとることで, もし, 見ようとする効果がないとするならば, すべての結果変数が全体平均に一致すると考えられる。群に分割することの効果があるなら, それは群内のどの要素に対しても生じているはずであり, 群の平均と全体平均の差で算出できるに違いない。こうした理論的な計算に沿って, なお実測値と異なる値が得られるとするならば, 測定や実験状況, 個人差に伴って生じ

8997 8998 る誤差になるはずである。このようにして効果と誤差を算出できる。具体的に考えやすくするため，山田・村井 (2004) の P.162 の表 7.1.1 にある数値例を用いて議論を進める。

8999 9000 → 大村 (2013) の P.64–72 が詳しい。山田・村井 (2004) の P.162–165 の数値例と計算式も参照。

9001 9002 9003 **平均差の効果** 実験計画を一般線形モデルとして考えた場合，残差がどのように現れるか，また効果が大きい・小さいというのはどのように現れるかを理解する。平方和を分解 ($SS_t = SS_f + SS_e$) すること，誤差に対する比率で考えるという原理を理解する。

9004 → 大村 (2013) の P.84-88.

9005 **キーワード**

9006 • 実験計画
9007 • ランダム化比較実験
9008 • 平均因果効果
9009 • 一般線形モデル

9010 E.12.2 授業情報

9011 **■コマの展開方法** 講義

9012 9013 **■標準シラバスにおける位置づけ** 科目番号 4;心理学研究法; 2. データを用いた実証的な思考方法; C データの統計的記述

9014 科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; G 推測統計: 代表値と散布度をめぐって (1)

9015 **予習・復習課題**

9016 9017 9018 9019 9020 **■予習** また心理統計のテキストから，「分散分析」として説明されているもの (たとえば山田・村井 (2004) の P.162–165 など) を一読しておくと良い。ただし，分散分析の文脈での議論は，最終的な効果の検証に当たって「帰無仮説検定」の枠組みを用いている。先んじてテキストを参照するものにとっては，F分布 や自由度などの耳慣れない用語が出てくるが，この講義において帰無仮説検定について扱うのはまだ先になる。効果検証の判断基準については現段階で理解する必要はない。

9021 9022 9023 9024 **■復習** 既に習った R の回帰分析関数をつかって，説明変数が離散的であるデータを使って分析を試み，どのような結果が出るかを確認すると良い。その際，説明変数が data.frame の中で factor 型になっている必要が亜あることに注意する。水準は少ない方がわかりやすいため，2 水準，3 水準と徐々に増やして結果がどのように変化するかを確認しよう。分析の前に ggplot で散布図を描くのを忘れないように。

9025 E.13 実験計画法 2；Between 計画

9026 E.13.1 授業内容

9027 **科目の中でのこのコマの位置づけ**

9028 記述統計 [3] 確率 [1] モデル [7] 推測・推定 [0] 検定 [1] 実践 [3]

概要

実験計画に基づいて，データをどのように扱っていくかを具体的に確認する。まずは一要因一要因 3 水準，Between デザインの数値例を用いる。ここで改めて，標本平均と母平均の違いや平均因果効果の推論をしようとしていることを確認し，統計的に (すなわち母集団において) 差があるかないかをどのように考えるかを理解する。回帰分析の時と同じように線形モデルで考え，数式の表現を行い，回帰分析と同じ形式 Form であることを確認する。

コマ主題細目

二要因のモデル　二要因の実験計画法モデルを考えるために，数値例を導入する。まず数値を可視化し，次に要因ごとに周辺化することで要因の効果をどのように書き出すことができるかを考える。回帰式に書き起こすことで，重回帰分析と同じ形式 Form になっていることを確認する。

→ データは山田・村井 (2004) の P.184, 表 7.6.1 を用いる。計算式も P.184–189 を参照すると良い。

効果の算出と平方和の分解　回帰式を参照しながら，平方和を分解する計算を進める。2 つの要因それぞれの大きさを各要因の水準ごとに考えることで計算する。要因ごとに区切るのは周辺化と呼ばれる。可視化したグラフを参考にしながら，計算箇所を辿りながら考えることが重要である。最後に 2 つの効果を全体平均に加え，水準の平均と合致「しない」ことを確認する。

→ 山田・村井 (2004) の P.184–189 を参照すると良い。

交互作用　合計が合致しなかったのは，組み合わせによる効果が発生していることによる。このような効果のことを交互作用 interaction という。交互作用の大きさを見積もることで平方和を完全に分解できたことになる。交互作用はその出現の仕方にさまざまなパターンがあることを理解する。また組み合わせ爆発という言葉があるように，三要因以上の実験計画を立てると，考慮すべき交互作用項がどんどん増えていくので，実質的に研究として使える実験計画は 2，3 要因の計画にとどめておいた方が良いことを理解する。

→ 山田・村井 (2004) の P.190–193. 小杉 (2019) の P.139–142

キーワード

- 誤差と効果
- 一般線形モデル
- 主効果
- 交互作用
- 多重共線性

E.13.2　授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ 科目番号 4;心理学研究法; 2. データを用いた実証的な思考方法; C データの統計的記述
科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; G 推測統計: 代表値と散布度をめぐって (1)

予習・復習課題

■予習 統計環境 R を使った線形モデルを実際に行い, 切片と傾きが算出されることを確認し, また説明変数が離散的であった場合の切片や傾きがどういう意味を持っているのかを改めて考えておくと理解が進む。また, 実験計画と回帰分析が同じである, という原理を十分に理解し授業に臨む必要がある。またその観点から, 複数の変数で行う回帰分析, すなわち重回帰分析についても復習しておく必要がある。

■復習 全体平均, 群平均を計算に用いて効果と誤差を算出する過程を, もう一度確認しておく。この時, 自分で収集した適当なデータがあると理解が進む。身の回りにあるデータ (ex. お菓子や食べ物の重さなど) や, 公開されているデータ (ex. スポーツに関するデータなどは公開されていることが多い) を使って計算するとよい。計算には Excel/Numbers/Calcs などの表計算ソフトを用いても良いし, 統計環境 R で行列のデータを対象に分析してももちろん良い。

E.14 実験計画法 3; 一要因モデル (Within)

E.14.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [1] モデル [8] 推測・推定 [0] 検定 [1] 実践 [3]

概要

再び一要因に戻るが, 今度は群内 (Within) 計画の場合で考える。群内計画の場合は, 事前-事後や第 I 期, II 期, III 期など反復測定をしている場合になり, 個体の識別ができていることになる。そのことで個人差を考慮することができるようになり, より精緻に誤差を取り除くことができるとも言える。効果の比較対象とするべき誤差の計算方法, 数式表現など Between デザインと異なるところに注意しながら理解を深める。

コマ主題細目 (箇条書き)

Within 計画のデータ 一要因 Within の実験計画法モデルを考えるために, 数値例を導入する。ここでの数値例は山田・村井 (2004) の P.178 の表 7.5.1 である。可視化に当たっては, 個人が識別できていることから散布図ではなく折れ線グラフで描画するのが適切である。可視化されたグラフを見ながら, どのような数式表現が可能かを考える。

モデルと誤差 個人差を個人ごとの平均と考え, 数式で表現する。次に各データから個人平均を取り除いて数値で考えてみる。こうすることで, 誤差がどこに現れているかを直観的に考えることができる。これらのプロセスを通じて数式を完成させ, また個々のセルにあらわれる誤差を算出する。

→ 計算の手順は山田・村井 (2004) の P.178–183.

Between と Within の違い 同じようなデータであっても, 個人が識別されていることで個人差を算出することができた。これは Between デザインでは個人差も要因計画では考慮できない誤差として捉えていたことを意味する。逆に, Within デザインは Between デザインよりも誤差を精緻に見積もることができるとも言える。心理学における実験計画は一般に Within デザインにする方が有利であるとも

言えるが，Within デザインは同じ個体に反復測定を行うことでもあり，実験上運用上のコストは大きいことに留意する。

→ 計算の手順は山田･村井 (2004) の P.179，図 7.5.1 の平方和の分解図式

デザイン行列と自由度　効果と誤差をそれぞれ未知母数 $\delta_j, \varepsilon_{ij}$ とおき，実測値 y_{ij} がどのように表現できるかを考える。この時，効果の有無を 0 と 1 で表した説明変数 X を用いることになる。このように数式で表現することで，その形が回帰分析と同じであること，説明変数のデータ種が違うだけであることを再確認できる。誤差の算出の時とは異なり，回帰分析的表現では切片を第一水準の平均値におき，そこからの相対比較で表現していた。これを全体平均からの相対比較に書き換えることで，両者の数学的表現を同一にすることが可能である。ただしこの場合は，推定できる効果の未知母数が「増える」のではなく，数学的性質から「制約をかける」ことで，実質的に推定すべき母数の数は同じであることが確認できる。

→ 小杉 (2018) の P.136–138

誤差の分布，個人差の分布　ガウスの考えた誤差論は，誤差は真値の周りで正規分布することを表していた。実験計画法における誤差も，理論値の周りで正規分布すると考えられる。一方で，さまざまな複合的な原因から影響されるものは自然と正規分布するため，個々人の生理･生態学的特徴も正規分布に従う。Within デザインは複数の正規分布が含まれる混合モデルである。

キーワード

- Within デザイン
- 個人差
- 階層線形モデル
- 変量効果と固定効果

E.14.2　授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ　科目番号 4;心理学研究法; 2. データを用いた実証的な思考方法; C データの統計的記述
科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; G 推測統計: 代表値と散布度をめぐって (1)

予習・復習課題

■予習　実験計画の用語を復習し，どういった実験計画が群内 Within デザインになるのかを考える。心理学系論文の中から，群内デザインの実験を行なっている例を探して読んでみるのも予習になる。

■復習　Within デザインにおいて個人差を算出するための手続きを，もう一度自分で確認しておくと良い。計算にあたっては身近な数値例を探すと良いが，個人が特定できるようなデータは見つけにくいかもしれない。そのような場合は，Between デザインの数値例を，借りに Within デザインで得られた数字であるとみなして，計算上の練習を行うのも効果的である。また今回，最後に後期の確率モデルにつながる「混合モデル」の考え方が示された。混合モデルについては，階層線形モデル，マルチレベルモデルなど複数の呼称が存在

するので，参考資料を探す場合はそれぞれのキーワードで探すと良いだろう。

E.15 前期のテスト

E.15.1 授業内容

科目の中でのこのコマの位置づけ

前期で学んできたことについて，基本的な知識，用語の理解，心理統計の基本的な考え方の筋道などを等テストをする。

E.15.2 授業情報

■コマの展開方法 マークシート式の試験を行う。問題数は 60-70 問程度である。

■諸注意

- 授業を行なっている教室で試験を実施する。着席場所は指定しないが，机の両端に座り，他の受験生と隣接しないようにする。
- 試験に際して資料等の持ち込みは認めない。
- 携帯電話，タブレット，PC など通信機能を有する機器の使用も認めない。
- (関数)電卓がなくても計算ができるよう工夫された出題をしているため，電卓などの使用も認めない。
- 受験者数を確定するため，当日は出席管理システムへの登録を行なってもらう。
- 不正行為を見つけた場合は即時退出を求め，単位を与えない。
- 試験会場に 20 分以上遅刻したものは受験を認めない。
- 試験開始から 30 分以上経過すると，希望者には退出を許可する。
- 退出に当たっては，質問用紙と解答用紙の両方を提出すること。
- マークシートへの回答であるため，鉛筆の利用が望ましい。
- マークをすべて塗りつぶしていないなど，読み取り上のエラーには対応しない。
- エクストラクレジットについての説明は別途行う。

E.16 確率と統計的推論

E.16.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [2] モデル [8] 推測・推定 [1] 検定 [0] 実践 [3]

概要

母集団に確率分布を仮定し，そこから得られる標本分布の特徴を利用して母数を推定する方法について学ぶ。標本の平均値が母平均に一致することを確認し，また標本の分散は母分散に一致しないことを確認する。ここでの推定方法はモーメントを用いたものであるが，確率分布を用いる方法もあり，その場合の “確率モデル” の考え方が必要なことを学ぶ。

コマ主題細目 (箇条書き)

母集団と標本 推測統計学の基本的な概念である母集団と標本の関係を理解し，統計学的な推定 (点推定，区間推定) の位置付けについて学ぶ。[*3]用語としての母数 (母平均，母分散，母比率など) と標本統計量 (標本平均，標本分散，標本比率など)，および標本統計量の実現値といった用語の意味するところも慎重に理解しなければならない。

→ 山田･村井 (2004) の P.68–73

母集団分布とモデル ここでは母集団に正規分布を仮定する “モデル” について考える。統計的な推測は基本的に不良設定問題であり，何らかの前提をおかないと解くことができない問題がほとんどである。心理学の場合は母集団分布を正規分布とする仮定が選ばれることが多く，この場合は標本統計量の分布も計算しやすくなる。

→ 山田･村井 (2004) の P.74–79.

正規分布の別の意味 正規分布はここまで，誤差の分布として解説してきたが，複数の要因が積み重なった時の全体的傾向として理解することもできる。平均値が集団の代表値としての意味もあること，また中心極限定理によって分布によらずその標本平均値が正規分布に従うことが示される。

中心極限定理については → 皆本 (2015) の P.180–181.

標本分布と標準誤差 母集団が正規分布に従うとすると，標本統計量も正規分布に従うことが導出できる。ここでは標本統計量としての平均値を例にあげ，「標本統計量の分布」のことを標本分布と呼ぶこと，また標本分布の標準偏差をとくに標準誤差と呼ぶことを理解する。たとえばサンプルサイズが大きくなったら，小さくなったらどのような値が得られるのかについて，具体的な例をみながら考えると理解が進む。

→ 標本分布については山田･村井 (2004) の P.90–97.

キーワード

- 母集団と標本
- 標本分布
- 中心極限定理
- 標本平均，標本分散

E.16.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5；心理学統計法；(2) 統計に関する基礎的な知識；A 統計分析の基礎：母集団と標本

[*3] 本書では検定をモデル比較の一種として扱うので，ここでは言及しない。

予習・復習課題

■予習 ここでは山田・村井 (2004) が丁寧に解説されているので，事前に一読しておくことを強く勧める。

■復習 いよいよ記述統計をこえ，直接知り得ないものについて推論する段階に進む。基本的に解けない問題を解こうとするようなものであり，仮定やモデル，みなしている箇所について自覚的でなければならない。今回の講義の中で，何が仮定され，何が理論的に導出されたものなのかを区別できるかどうか，ノートを整理しておくと良いだろう。また推測統計学の基本的概念である，母集団と標本の関係と推定の手続きについては，山田・村井 (2004) のテキストが最も丁寧で適切な記述をしているので，是非該当箇所を通読し，概念や用語を間違えないように注意されたし。

E.17 推論と判断

E.17.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [3] モデル [8] 推測・推定 [1] 検定 [1] 実践 [3]

概要

標本統計量をつかって母数を推定する方法として，幅を持たせた推論である区間推定という方法がある。この区間の中に母数があるとはどういうことか，注意して理解する必要がある。また標本統計量を標準化しておくことで，標準正規分布を使って一般的な推定ができる利便性があることを理解する。これらは推定の問題であったが，心理学における多くの場合には推定量そのものよりも，平均因果効果があったのかなかったのか，という判断を下したいという目的があり，そのために検定と呼ばれるモデル比較の考え方が導入される。正規分布など母集団分布におけるモデルはそのままに，パラメータがどういう値であるかについての仮説を戦わせるその作法を学ぶ。

コマ主題細目 (箇条書き)

不偏推定量 母集団に正規分布を仮定し，標本平均の平均 (期待値) が母平均に一致することを確認する。このことから，標本統計量が推定値として利用できることを確認する。ただし，分散の場合は一致しない (不偏性がない) ことをみる。分散については不偏分散を考えることでこの不一致を解決することができる。

→ 不偏推定量については山田・村井 (2004) の P.98–103 に詳しい。

点推定と区間推定 標本平均は母平均の不偏推定量ではあるが，ある値をそのまま母平均であると考えるのは (標準誤差の考え方があるとは言え) やや断定的である。これに対して一定の幅を使って予測する区間推定という方法がある。推定に確率の言葉が入ってくるので不慣れな表現が出てくるが，注意深く解釈する必要がある。

→ 川端・荘島 (2014) の P.111–116.

標準正規分布の利点 標本統計量を標準化しておくことで標準正規分布に対応させ，たとえば 95% 区間に対応する ± 1.96 という数字を使うことができれば便利である。ここでは標準化の計算についておさらいするとともに，標準正規分布の表の利用の仕方について改めて理解する。

→ 山田･村井 (2004) の P.86–89.

標本だけを使った推定 標準正規分布を使うと母数の推定ができることが分かったが，母分散が明らかでない場合は推定ができない。そこで母分散が未知の場合であっても推定できるような仕掛けを考える必要がある。ここでは t 分布を導入し，標準正規分布の時と同様の計算方法で標本統計量がけから区間推定ができることを理解する。

→ 山田･村井 (2004) の P.128–131.

推定と検定 ここまでは平均値などを推定してきたが，推定量を意思決定に応用する方法がある。特定の数値についての複数のモデルを比較し優劣を決めるモデル比較の考え方がこれで，心理学においてはとくに，帰無仮説と対立仮説という 2 つのモデルを用いる帰無仮説検定という考え方である。

→ 川端･荘島 (2014) の P.81–85.

キーワード

- 点推定と 区間推定
- 不偏推定量
- 標準正規分布
- t 分布
- 統計的検定

E.17.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; F 推測統計: その考え方

予習・復習課題

■予習 母集団と標本の関係，標本から母数を推定できることについて，前回の話の延長線上で進むので十分に復習してから授業に挑むこと。とくに限られた情報から未知なるものを推定するという手続きに当たっては，幾重にも過程が重ねられていくので，どのような仮定を置いて進めてきているかを再確認しておくと良いだろう。

■復習 帰無仮説検定については，触れていない心理統計のテキストはないと言っても過言ではないので，どのようなものでも良いから手に取り，その手続きや用語について復習しておくことが望ましい。テキストによっては，分布 (ex. t 検定，F 検定) や要因計画 (ex. 対応のある t 検定，分散分析) ごとに分けて書かれていたり，フローチャートのように説明されているものもあるが，本質は母数の推定とモデル比較であることに注意して読み解くと良い。

E.18 帰無仮説検定

E.18.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [4] モデル [8] 推測・推定 [1] 検定 [2] 実践 [3]

概要

心理学の研究において最もよく使われる意思決定法である，帰無仮説検定の具体例について学ぶ。帰無仮説検定は背理法をもちいて効果の有無を判断する方法であり，結果の報告に関する表記法も含めてその作法を学ぶ必要がある。注意するべきは，帰無仮説検定の技術は誤用されたりや悪用されたりすることが多い方法であり，前提や仮定をしっかり守って正しく運用することを常に心がけねばならない点である。ここでは最も単純な相関係数の帰無仮説検定に基づいて考える。また，検定は確率的判断の手続きであるが故に，判断の誤り方が 2 通り考えられる。ここで検定力の考え方が導かれ，どのような状況のもとで帰無仮説検定が行われているかを理解する必要がある。

コマ主題細目

統計的帰無仮説検定とその手順 統計的帰無仮説検定 (Null Hypothesis Significance Test;NHST) は帰無仮説に基づいてモデル比較をする一連の手続きである。この手続きはさまざまな統計量に対して一般化されているため，統計パッケージの中にも含まれており，統計法の専門家でなくても簡単に利用することができる。ただし安易な使い方は誤用，悪用を招くことになるので，慎重にその手続きとロジックを理解しておく必要があるだろう。

→ 手続きの一般化としては山田・村井 (2004) の P.108–109

有意水準 モデル比較という観点からは，どうしても「判定基準」を考える必要がある。NHST の文脈では，それは希少性の指標である有意水準になる。有意水準の定め方は領域ごとに定める任意であり，心理学では一般に 5% が用いられている。このことは，複数のモデルの優劣を判定する数字であって，モデルの強さ，正しさを表現する数字ではないことに注意する。

→ 山田・村井 (2004) の P.112–115.

1 つの平均値の検定 ここでは母分散がわかっている場合の，標本平均の検定を考える。標準化の手続きと，帰無仮説検定の考え方を一通り復習した上で，実際どのように検定が行われるかの実践例である。

→ 川端・荘島 (2014) の P.80–89, 山田・村井 (2004) の P.126–127.

1 つの平均値の検定 (母分散未知) 先ほどの例では，検定すると言いながら母分散がわかっているという特殊な状況であった。普通は母分散がわかっていることはなく，これも不偏分散を使って推定しなければならないが，このような場合は標準正規分布ではなく，t 分布と呼ばれる統計量が用いられる。t 分布の導出は行わないが，不偏分散を使った分布であること，サンプルサイズによる大きさの違いを調整する，自由度というパラメータがあることに注意する。

→ 川端・荘島 (2014) の P.93–96.

相関係数の検定　具体的な数字を使って，相関係数の検定を行う。Rde 相関係数を行う際は，cor.test 関数を用いる。ここでは無相関検定である，すなわち統計的に有意であることが相関に意味があることを指すわけではないことを理解し，また統計量に応じた分布を利用して推定する必要があることを理解する。

→ 清水 (2021) の P.150–162.

キーワード

- 有意水準
- 1 つの平均値の検定
- t 分布
- 無相関検定

E.18.2　授業情報

■コマの展開方法　適宜投影資料を用いる

■標準シラバスにおける位置づけ　科目番号 4;心理学研究法; 2. データを用いた実証的な思考方法; C データの統計的記述
科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; F 推測統計: その考え方

予習・復習課題

■予習　具体的な帰無仮説検定の手続きが進むので，帰無仮説，対立仮説，有意水準といった一般的な用語について復習しておく。

■復習　相関係数の大きさ，標本の大きさを任意の値にし，無相関検定を実際にやってみると良い。さまざまな数値を入れることで，検定結果がどのように変動するかを理解することは帰無仮説検定の本質を理解するのに役立つ。仮想的な数字では実感が得られにくい場合，身の回りの相関がありそうなデータを具体例として検定し，何が言えるのかを考えてみると良い。

E.19　帰無仮説検定の注意点

E.19.1　授業内容

科目の中でのこのコマの位置づけ

記述統計 [3]　確率 [5]　モデル [8]　推測・推定 [1]　検定 [3]　実践 [3]

概要

帰無仮説検定の流れを抑えたところで，実際に計算結果をどのように報告するか，結果からどのように考察するかについて解説する。昨今の批判にもあるように，結果の 1bit 判断だけに基づいて考察を進めることは，効果の大きさを無視した過剰な宣伝になってしまう可能性があり，ひいては再現性の低下，科学の信頼性の問題まで引き起こしたという事実を学ばなければならない。ここでは検定における二種類の間違い，および効果量，検定の多重性の問題などをしっかり確認し，正しく利用するための観点についてしっかり身につけて

おく必要がある。

コマ主題細目 (箇条書き)

結果の報告 相関係数の有意性検定について解説する。別名「無相関検定」ともよばれているように，帰無仮説が想定する世界は母相関が 0 の状態である。ここから得られた標本が作る標本相関の分布は t 分布に従うことがわかっており，これを使って検定を行うことができる。計算のプロセスを簡単にフォローするとともに，結果の報告における表記法について理解する。

→ 山田・村井 (2004) の P.132–133, 山内 (2010) の P.216–220.

統計的に有意とは 相関係数の帰無仮説検定においては，顕著な違いが示されること＝相関係数が大きいこと，を意味するわけではない。たとえばサンプルサイズが大きくなれば，非常に小さな相関係数であっても統計的に有意であると判断される。有意かどうかにこだわるのではなく，その大きさに基づいて判断することの重要性を学び，標準化された効果の大きさを表現する効果量について理解する。

→ 山田・村井 (2004) の P.142–143. 効果量については大久保・岡田 (2012) の P.43–46.

検定における 2 種類の間違い ここで検定における 2 種類の間違い方，すなわち Type I Error と Type II Error について解説する。モデルの定量的な比較でなく，仮定に基づく質的な確率的な判断であり，その組み合わせから 4 つの結果が起こり得る。ここで NHST では何をコントロールし，何を目的としているかを改めて確認すること，また検定を繰り返すことで Type I Error がインフレーションを起こしうることを理解することが必要である。

→ 山田・村井 (2004) の P.120–121，検定の多重性については同書 P.158–161. 川端・荘島 (2014) の P.90–93. も

キーワード

- 信頼区間
- Type I/II Error
- 効果量
- 検定力
- 例数設計

E.19.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; H 推測統計: 代表値と散布度をめぐって (2)

予習・復習課題

■予習 統計的帰無仮説検定の手続きを一通り復習し，とくに帰無仮説と対立仮説がどのように用意されているか，意思決定の判断基準はどういう意味であったかをしっかりと復習して挑む。

■復習 検定の話に入ると，要因計画や線形モデルとどのようにこれらが接合するのかが分かりにくい。回帰モデルの説明変数を離散化したのが要因計画 → 要因計画の効果の大きさ (傾きの大きさ) を質的に判断する手続きが，平均値差の検定である，という関係であることを整理しておこう。推測統計学は，基本的に母数の推定 (回帰係数のパラメータも母数である) をしているのであり，推定量を何らかの基準で判断するのが検定，という繋がりである。もちろん検定を行わずに，大きさをそのまま報告しても良い。むしろ昨今ではそちらの方が推奨されており，些細な違いや「差がないこと」を「差があること」の根拠とすることの問題点が数多く指摘されている。帰無仮説検定の手続きは，過去の論文に示された結果を読むための知識として重要であり，本質である効果の大きさを推定していることを忘れないように。

E.20 二群の平均値差

E.20.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [6] モデル [9] 推測・推定 [2] 検定 [4] 実践 [4]

概要

心理学の実践で最もよく用いられるのが，平均値差の検定である。帰無仮説検定としては，t 検定を行うことになる。対応のある t 検定，対応のない t 検定の順に，R での演習も交えて伝える。その際，結果から適切な読み取りができるようになること，記述することなどが求められる。統計環境は R を用いるが，SPSS そのほかの統計パッケージを用いても同様の出力が得られる。統計環境が変わる可能性もあるので，出力画面の位置で内容を覚えるのではなく，意味的な理解をしておく必要がある。

コマ主題細目 (箇条書き)

一要因 Within デザイン，2 水準 一要因 Wtihin デザイン，2 水準モデルの検定統計量もまた，t 分布を用いる。これは「対応がある」，すなわち個人が特定されているため，正規分布する変数の差の分布についての検定であり，既に習った 1 つの平均の差についての検定と同じである。

→ 橋本・荘島 (2016) の P.1–13.

一要因 Between デザイン，2 水準 一要因 Between デザイン，2 水準モデルの検定統計量は，t 分布を用いるため t 検定と呼ばれることもある。帰無仮説は二群の平均値差がないとするものであり，5% 水準による検定結果と差の信用区間が示されることが多い。また効果量 d を別途計算して報告されることにも注意する。

→ 橋本・荘島 (2016) の P.14–25.

R による実践；一要因 Within デザイン，2 水準 このデザインは R では t.test 関数を用いる。Within デザインの場合は paired オプションをつける必要がある。効果量の算出には，別途 `effsize` パッケージを導入すると良い。

→ 橋本・荘島 (2016) の P.1–25.

R による実践；一要因 Between デザイン，2 水準 このデザインは R では t.test 関数を用いる。Between デザインの場合は paired オプションをつける必要はないが，Welch の補正がかけられてい

る。仮説検定の前提となった仮定を思い出し，これが満たされない場合はこうした補正がかかることを学ぶ。

→ 橋本・荘島 (2016) の P.1–25.

キーワード

- 対応のない t 検定，対応のある t 検定
- 線形モデルと検定
- t 検定
- 効果量

E.20.2 授業情報

■コマの展開方法 講義および演習

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; H 推測統計: 代表値と散布度をめぐって (2) 科目番号 5 心理学統計法; 2 統計に関する基礎的な知識; C/D/E エクセル，R，SPSS 入門

予習・復習課題

■予習 R や RStudio，プロジェクトによる管理など基本的なことを改めて教示しないので，これまでの復習をかねて基本的な R/RStudio の挙動について再確認しておくこと。データファイルは事前に CoursePower などを通じて配布されるが，読み込みなどについては逐一立ち止まって解説しないので，できることを確認して授業に挑むこと。

■復習 時間内に収まりきらなかった課題については，復習用の課題となり，Rmd ファイルで提出することが求められる。適切な分析ができていない場合は，課題を不受理として扱う。これまでと同様，課題が提出されていることが単位取得の要件になることに注意し，しっかり対応して欲しい。また，今回の内容について橋本・荘島 (2016) が丁寧な解説をしているため，十分理解できなかった場合は同書を一読して理解しておく。内容的には，山田・村井 (2004) の P.162–207 も該当する。最も，これらを実践的に R で行うことが目的であり，R の使い方に言及している小杉 (2019) なども参考にしてほしい。効果量に関しては，大久保・岡田 (2012) の P.52–68 を参照しておく。

E.21 要因計画と平均値差の検定

E.21.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [7] モデル [10] 推測・推定 [2] 検定 [5] 実践 [4]

概要

線形モデルに確率的判断の手続きを加えたものが，平均値差の検定であった。要因計画にこれを応用したとき，分散分析と呼ばれた独自の体系になる。ここでは線形モデルの一環であることを意識しつつ，質的説明

変数によって水準同士の対比較になることによる分析の手順について理解する。

コマ主題細目 (箇条書き)

要因計画と検定　要因計画の考え方と平均値差の検定モデルを統合的に理解する。データの散布図を書けば明らかなことであるが，二群の平均値差の検定は，二群の平均値を通る回帰直線の傾きを求めていることと同じである。要因計画もその本質は線形モデルであり，帰無仮説検定はモデル比較である。帰無仮説の線形モデルがどのようなものであり，検定結果が何を報告するものであるかを知ることで，両者の類似点と相違点が明らかになる。

一要因 Between デザイン，多水準の場合　水準数が 3 以上になると，一般に分散分析と呼ばれる検定方法を用いることになる。多水準の場合も帰無仮説を線形モデルで表現することができるが，説明変数が名義尺度水準であるため，順不動に並べ替えることができるし，効果を検証するためには符号を無視した分散を考える必要がある。分散分析の基本的アルゴリズムは，分散の分解と，比による検定である。分解の手順を追うことで，線形モデルとの対応が分かりやすくなると同時に，変化量 (効果の大きさ) そのものを検証することのほうがより重要であることに注意する。帰無仮説検定としては分散の比をもちいるが，線形モデルとして効果の大きさそのものを考えるのが良い。

事後の検定　多水準モデルの際に置かれた帰無仮説は水準間のどこかに差があるというものであり，どこに差があるのかまでは明示してくれない。そこで事後的な検定を行い，水準間のどこに差があるのかまで明らかにしなければならない。このような込み入った手続きを取るのは，Type I Error のコントロールが必要な帰無仮説検定だからであることもまた，合わせて理解しておく必要がある。

二要因デザイン　要因が複数になると，交互作用について考える必要があるのであった (→ 第 E.13 講)。交互作用項も含めた分散分析表の読み取り方や，主効果・交互作用がそれぞれどこの平均値を比較するためのものであるかを理解しながら読む。また，二要因デザインでも効果があると判断された場合，多重比較によってその後の検定を続ける必要がある。統計環境によっては自動的に以下の検定を続けてくれることもある。

→ 橋本・荘島 (2016) の P.57–79, 山田・村井 (2004) の P.184–193.

キーワード

- 分散分析
- 一般線形モデル
- 分散分析表
- 多重比較

E.21.2 授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ　科目番号 5;心理学統計法; 1. 心理学で用いられる統計手法; H 推測統計: 代表値と散布度をめぐって (2)

予習・復習課題

■**予習** 前期に学んだ要因計画法は，F 分布と合わせて分散分析としてよく知られている。計算方法については，おそらく日本で最も容易な入門書は大村 (2013) である。同書の 3 章を読み，前期の振り返りとともに効果検証のステップを思い出しておこう。

■**復習** ここでの手続きは一見「難しい」ようである。しかしその難しさをよく考えてみよう。関係性が複雑で全体像がわからない「困難さ」ではなく，関係は単純であるが幾重にも反復されている「煩雑さ」がその原因であることに気づくだろう。つまり，単純 (simple) ではあるが容易 (easy) ではない，と思ってしまうのである。言い換えれば，落ち着いて結果を読み進めれば，同じことを繰り返しているだけなので，そのパターンが見えれば理解はしやすい。

くどいようであるが，要因計画に質的判断を持ち込んだのが分散分析であり，判断の分岐によって注意すべき点や検討すべき箇所がそれぞれ異なる。効果の大きさだけを考え流のであれば，ここまで複雑な考え方は必要ないように思えるが，確率を伴った判断と科学的な手続きである以上，避けて通れないことでもある。出力例をみながらどこでどのようなモデル比較が行われているのか，しっかり復習しておくことが大事である。

E.22 R による分散分析

E.22.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [8] モデル [11] 推測・推定 [3] 検定 [6] 実践 [5]

概要

今回は統計環境 R で分散分析を行う方法について演習的に学び，結果から適切な読み取りができるようになることを目的とする。統計環境は R を用いるが，SPSS そのほかの統計パッケージを用いても同様の出力が得られる。統計環境が変わる可能性もあるので，出力画面の位置で内容を覚えるのではなく，意味的な理解をしておく必要がある。

コマ主題細目 (箇条書き)

一要因 Between デザイン 3 水準 2 つ以上の水準の場合は，分散分析 (Analysis of variance) と呼ばれる検定グループに入る。分散分析については，独自の関数群 anovakun を導入すると大変便利である。anovakun の導入とデータ形式について簡単に説明した後，一要因 3 水準のモデルを実行し，どのような結果が得られるかを確認する。

→ 橋本・荘島 (2016) の P.26–55. 分散分析の効果量については大久保・岡田 (2012) の P.76–85.

二要因 Between デザイン 複数の要因が組み合わさった場合でも，anovakun で対応することができる。ここでは二要因計画，とくに間 × 間，内 × 内のようなものから，間 × 内のような混合計画と呼ばれるものまで，一通り実行する。同じ出力結果が得られるところ，異なるところなど，アウトプットを注意深く読み取ることが必要である。

→ 橋本・荘島 (2016) の P.57–120. 要因計画の効果量については大久保・岡田 (2012) の P.86–88.

一要因 Within デザイン WIthin デザインの場合は個人差を特定することができるのであった。分散分析表にも個人差の分散が算出されるが，そこは比較の対象ではなく，誤差との比較になることを確認する。また，分散分析の前提である分散の等質性についての検定が事前に必要になることにもまた，注

意が必要である。

キーワード

- 分散分析
- 要因計画
- 効果量

E.22.2 授業情報

■コマの展開方法 演習; 4 号館 4F の PC ルーム 1/2 に別れて, 実際の PC を使う実習を伴った授業を行う

■標準シラバスにおける位置づけ 科目番号 5 心理学統計法; 2 統計に関する基礎的な知識; C/D/E エクセル, R, SPSS 入門

予習・復習課題

■予習 R や RStudio, プロジェクトによる管理など基本的なことを改めて教示しないので, 前回の復習をかねて空いている時間に PC ルームで基本的な R/RStudio の挙動について再確認しておくこと。データファイルは事前に CoursePower などを通じて配布されるが, 読み込みなどについては逐一立ち止まって解説しないので, できることを確認して授業に挑むこと。

■復習 時間内に収まりきらなかった課題については, 復習用の課題となり, Rmd ファイルで提出することが求められる。適切な分析ができていない場合は, 課題を不受理として扱う。これ真同様, 課題が提出されていることが単位取得の要件になることに注意し, しっかり対応して欲しい。また, 今回の内容について橋本・荘島 (2016) が丁寧な解説をしているため, 十分理解できなかった場合は同書を一読して理解しておく。内容的には, 山田・村井 (2004) の P.162–207 も該当する。これらを実践的に R で行うことが目的であり, R の使い方に言及している小杉 (2019) なども参考にされたし。

E.23 確率分布とデータの関係; 尤度

E.23.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [9] モデル [12] 推測・推定 [4] 検定 [6] 実践 [5]

概要

検定では平均因果効果を検証するため, 平均値や (不偏) 分散などの標本統計量がどのような確率分布に従うかを参照し, 結果を判定するという使い方を行なってきた。検定はこのように, 母数を推定する方法に, 結果を判定するロジックが組み合わさったものである。この時の母数の推定方法は, 標本平均や不偏分散を推定値とする方法で, モーメント法とよばれるものである。これに対し, モデルである確率分布をデータに最も近しい形に調節することで, そのパラメータを推定値とする方法が最尤法である。ここでは最尤法の原理について理解する。

コマ主題細目 (箇条書き)

確率関数と尤度関数 最も単純な例として，ベルヌーイ分布を取り上げる。ベルヌーイ分布は確率 θ によって 0 または 1 が得られる分布であるが，推定に当たっては θ が未知でデータが既知である。この関係から，関数の読み取り方を逆にしたものがとくに尤度関数と呼ばれる。

最尤推定 データに基づいて尤度を計算できることから，このデータが得られるのに最も適したパラメータの位置を推定する方法が最尤推定である。最尤推定では尤度関数の総積を計算し，そのピークの値を推定値とする方法である。尤度関数を実際に描きながら解説する。ここで注意するべき点は，尤度関数もまたなんらかの分布関数のように見えるかもしれないが，公理よりこれが確率分布関数ではないことをしっかり意識することである。

→ 野島他 (2019) の P.74 にベルヌーイ分布の例。あるいは小杉 (2018) の P.117–119.

対数尤度関数 確率密度の総積を計算すると，その値は必然的にどんどん小さなものになっていき，計算が非常に困難である。そこで実際の計算では，尤度関数の対数を取ることが一般的である。対数を取っても値の大小関係は変わらないため，ピークを推定するのには問題なく利用できる。記号が $\prod$ から $\sum$ に変わることは対数の特徴である。

正規分布の尤度関数 心理学では主に正規分布をモデルに用いるが，正規分布のパラメータは 2 つある。話を簡単にするため，まずは分散パラメータを固定し，平均パラメータを変化させればどのように尤度関数が，対数尤度関数が変化するかを確認する。つづいて，分散パラメータも未知なものとして，両者が同時にピークを迎える点を最尤推定とする。

キーワード

- 確率密度関数，確率質量関数
- 尤度関数
- 最尤推定値
- 対数尤度

E.23.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5；心理学統計法；(2) 統計に関する基礎的な知識；B 確率と確率分布

予習・復習課題

■予習 改めて確率分布を用いたモデルに展開していくため，これまでに習った確率分布についての基礎知識 (確率質量，確率密度，確率変数やパラメータを使った記述の仕方，等) について，第 E.6，E.7 講を復習しておくことが望ましい。また，総和の記号 $\sum$ や対数についての考え方など，これまでに習っている教材を利用して復習しておくとよい。

■復習 非常に小さな数字の計算になるため，尤度関数を最大化する問題はコンピュータにとっても難しい。とはいえ，表計算ソフトや統計環境をもちいて，簡単な数値例を自ら考え，計算したり図示したりすることで直

観的に理解が進むこともある。また対数を取っても大小関係が変わらないことなども，手元の計算機 (電卓，表計算ソフト，統計環境 R) で簡単に確認できるので，是非数値例を用いた復習を行っておいて欲しい。

E.24 回帰分析の最尤推定

E.24.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [10] モデル [13] 推測・推定 [5] 検定 [6] 実践 [5]

概要

最尤法によって母数を推定する新しい方法を理解したところで，この方法を使ってパラメータに構造を入れる確率モデルの考え方に進むことができる。まずは回帰分析と誤差分布の関係を改めて確認し，誤差分布が正規分布に従うと仮定した場合，回帰モデルがどのような確率モデルとして表現できるかを理解する。これらの形が求まると，解答用紙の最尤推定値が得られることを理解する。

コマ主題細目 (箇条書き)

回帰分析と誤差分布 あらためて単回帰分析のモデルを考える。実測値，予測値，誤差がそれぞれどのように計算され，誤差を最小にする形で回帰係数を算出していたことを復習する。(→ 第 E.8 講)

→ 豊田 (2017) の P.27–30.

確率モデル 今度は回帰分析のモデルにおいて，誤差が正規分布するという確率的仮定を導入する。確率分布に従うことを意味する記号 ∼ を用いて回帰分析のモデルを書き改めることを目的とする。式の展開，そのプロセスについてしっかりと理解することが重要である。

→ 小杉 (2018) の P.120–122.

確率モデルと最尤推定 確率モデルで記述することができれば，データを既知，パラメータを未知と捉えることで尤度関数に書き換えることができる。尤度関数をデータごとに掛け合わせる (実際には対数を取って足し合わせる) ことで，データに確率分布を仮定した時に最も尤もらしい値としての推定値を算出することができる。これらの展開は，重回帰分析モデルになっても同様である。

→ 小杉 (2018) の P.120–122. あるいは浜田他 (2019) の P.26–32.

キーワード

- 誤差分布
- 正規分布
- 最尤推定
- 回帰分析

E.24.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; (2) 統計に関する基礎的な知識; B 確率と確率分布

予習・復習課題

■予習 再び回帰分析に戻ってくるので, 第 E.8, E.9 講を復習しておくことが望ましい。最小二乗法による推定方法の考え方も合わせて復習しておく。また 2 つのパラメータをもつ正規分布の尤度関数について, 理解が不十分なようであれば前回やその前に戻って確認しておくと良い。

■復習 最小二乗法を使っても最尤法をつかっても, 推定値に違いはないのだが, 背後にある基準の意味合いが大きく異なる。母数を推定するモデルとして, どこが変わったのかをしっかり理解し, 要因計画法など複雑なモデルになっても記述できることを確認しておく。

E.25 ベイズ統計の基礎

E.25.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [11] モデル [13] 推測・推定 [6] 検定 [6] 実践 [5]

概要

母数を推定する第三の方法として, ベイズ法による推定を考える。ベイズの定理の歴史は古く, 逆確率の法則とも呼ばれるものである。ベイズの定理を理解するにあたっては, 条件つき確率について理解する必要があり, これらを通じてベイズの定理を導く。ベイズ法の利点は, 尤度関数を確率分布関数に変えてくれることであり, 推定値の確からしさを確率で表現することができる点である。ここでは病気の罹患率の例を用いて説明し, 確率を信念の更新として理解することもできることを学ぶ。

コマ主題細目 (箇条書き)

二次元の確率変数 これまで扱ってきた確率分布は単一のものであったが, 二変数による (二次元的な) 確率分布を導入する。複数の事象が同時に生じることを表す同時確率, 同時確率のうち一方の変数だけに周辺化して考える周辺確率, 一方の情報が分かっている上での確率である条件つき確率, の 3 つの考え方が導かれる。また, 両者が独立であるとはどういうことかについて確率的な定義を与える。簡単な数値例を用いて解説する。

→ Kruschke (2014 前田・小杉監訳 2017) の P.93–109. 豊田 (2015) の P.3–11.

ベイズの定理の意味 ベイズの定理について, 数式をさまざまな角度から眺めてその意味を理解する。同時確率, 周辺確率の定義などを用いて, 条件付き確率の式を変形していくことができ, 最終的に右辺と左辺で条件が逆になる場合の確率が計算できることがわかる。これがベイズの定理と呼ばれるものであり, 逆確率の定理, 法則と呼ばれることもある。ベイズの定理は, 事前確率に尤度と周辺尤度の比をかけたものとして理解することもできるし, 尤度関数を確率として捉えるために変形させる項として事前確率と周辺尤度の比をかけたもの, と考えることもできる。

→ Kruschke (2014 前田・小杉監訳 2017) の P.93–109. 尤度を確率分布に変えることの分かりやすい説明としては, Lambert (2018a) の Table4.1.X が良い。

ベイズの定理を使った推論の例　ベイズの定理について，意味的な解釈を膨らませておこう。ベイズの定理は確率のアップデートとして考えることができる。我々が推論するというのは，確率をアップデートすることだと考えることもできる，ということについてはすでに触れており，確率という数字が従うべきルールさえ守られていれば，確率は人間の自然な推論モデルとして考えることができるだろう。ただし頻度主義的な確立の理解と，主観確率による理解の違いには改めて注意が必要である。またベイズの定理を使っての推論の例として，病気の罹患率の例とモンティ・ホール問題を取り上げる。これらのポイントは，「事前分布がわかっているということ」と「情報がアップデートされる」ということを意識するかどうかにある。

→ Kruschke (2014 前田・小杉監訳 2017) の P.15–23. あるいは豊田 (2015) の P.8–12.

キーワード

- 同時確率
- 周辺確率
- 条件付き確率
- ベイズの定理
- 事前確率，事後確率

E.25.2　授業情報

■コマの展開方法　講義

■標準シラバスにおける位置づけ　科目番号 5；心理学統計法；(1) 心理学で用いられる統計手法；J より高度な記述や推定を目指して

予習・復習課題

■予習　ここからは，より進んだ統計的内容に入る。ベイズ統計学を心理学領域で活用する歴史はまだ浅いが，現代社会では実践的に多用されている技術であり，心理学にとっても今後ますます重要性が高まることは間違いない。たとえば野島他 (2019) は公認心理師向けのテキストであるが，その最終章をベイズにあてている。内容的には，確率についてまた違った視点からの考え方が導入される。とはいえ，確率が守るべきルールについては変わりがない。改めて第 E.6 講の確率について復習しておくとよい。「ベイズ統計学」をキーワードとしたいくつかの入門書があるので，目を通しておくと良い。あるいは「モンティ・ホール問題」で検索すると，興味を覚える人もいるかもしれない。

■復習　さまざまな確率の書き方から，ベイズの定理の導出まで，筋道をもう一度自らフォローアップして導出できるかどうか確認しておくこと。罹患率のような具体例については，読み物として Giegernzer (2002 吉田訳 2003) があるので目を通すと良い。確率についての説明の仕方については，平岡・堀 (2009) の P.1–69 の説明も異なる切り口からでおもしろいし，奥村他 (2018) の P.1–12 もわかりやすい。ベイズ統計学についてのごく簡単な入門書としては，ラノベ風の石田・石田 (2019) なども良いだろう。

E.26 ベイズ推定の実際

E.26.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [12] モデル [14] 推測・推定 [7] 検定 [6] 実践 [5]

概要

ベイズの定理が導入され，確率に対するもう一つの考え方を導入した。ここでは具体的な例を使って，事前分布で確信度を表現すること，事後分布も確率分布であり，頻度主義的な解釈とは異なることを確認する。事後分布は確率分布なので，推定にあたっては分布を要約する必要があり，EAP,MAP,MED 推定値や確信区間，HDI など分布を要約して表現する方法について整理する。

コマ主題細目 (箇条書き)

事前分布と事後分布 ここまで事前・事後確率と呼んでいたものは，連続的な分布関数出会っても同様に機能する。ここで確率のベイズ的解釈により，事前分布はデータが得られる前の確信度を表現していること，またとくになんの情報もないとき，わからないことを一様分布のように表現できることを理解する。また，ベイズの定理は尤度を (事後) 分布に変換するものであることとして捉え，結果が分布として得られることを確認する。具体例としてベルヌーイ分布をもちいた学力の推定をとりあげ，最尤推定との違いや，分布による結果であることの意味を理解する。

→ Kruschke (2014 前田・小杉監訳 2017) の Pp.18-23. や，豊田 (2016) の P.18–21

事後分布を代表する数字 ベイズによる推定の結果は，事後分布という確率分布の形で得られる。結果が分布として得られるため，報告する場合はたいていその代表値を使って報告することになる。とくに MCMC によるベイズ推定の場合は，多くの MCMC サンプルの標本記述統計量がその代表値として考えられるという利点がある。中央値 (MED)，平均値 (EAP)，最頻値 (MAP) の名称と算出方法について理解し，そのほかにも最高密度区間 (HDI) といった区間で報告することもある。

→ 豊田 (2015) の P.51–55.,HDI については Kruschke (2014 前田・小杉監訳 2017) の P.91–93.

事前分布の影響 事前分布を一様分布においた場合，事後分布の形は尤度関数そのままの形で現れ，かつ確率分布として考えられる利点がある。しかし事前分布が異なれば，事後分布の形はそれに影響を受け，結果が変わってしまう。事後分布は事前分布と尤度関数との折衷であり，多くのデータをもつ鋭い尤度関数であればその影響は小さいが，緩やかな尤度関数であれば大きく影響を受けてしまう。研究に際して適切な事前分布を置くことの重要性を確認する。一方，帰無仮説検定においては不可解な結果になるケースもあり，有効に活用することで研究に与することもあることを理解する。

→Kruschke (2014 前田・小杉監訳 2017) の P.136–138

キーワード

- 事前分布，事後分布
- 事前分布の影響

- MED,EAP,MAP,HDI

E.26.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; (1) 心理学で用いられる統計手法; J より高度な記述や推定を目指して

予習・復習課題

■予習 確率の考え方やベイズの定理についての数式的理解, 意味的理解を踏まえ, これを心理統計に応用する段階に進むところであるから, これらのについて不明な点がないように, 事前に十分理解しておく必要がある。

■復習 帰無仮説検定をはじめとする, 頻度主義的な考え方とに違いに留意しつつ, 両者の長短所を自分なりに言葉にできるようにしておくことが重要であるだろう。

E.27 ベイズ推定と判断

E.27.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [13] モデル [15] 推測・推定 [8] 検定 [6] 実践 [5]

概要

ベイズ推定による結果の解釈は, 分布の区間として直接的な解釈ができるため, 帰無仮説検定で用いていたような「差があると言って間違っている確率」という持って回った言い回しを必要としない。ここではまず区間推定を行い, 判断基準としての帰無仮説・対立仮説の導入という従来法による群間差の判断原理を再確認する。つづいてベイズ法による区間推定と, 区間を使ってどのような判断が可能なのか (HDI 区間, ROPE による判断) を考える。また, 帰無仮説検定が二つのモデルの比較であったことを思い出し, ベイズ流のモデル比較であるベイズファクターについて考える。ベイズファクターの計算には周辺尤度が必要であるが, サヴェージ・ディッキー法をつかうとその簡便的な計算が可能であることにも触れる。

コマ主題細目 (箇条書き)

ベイズ法と帰無仮説検定 ベイズ法による事後分布を用いた区間推定は, 信用区間ではなく確信区間であり, 区間が分布の情報を含んだものであること, 母数に対する確信度であることを確認する。一方で, モーメント法での推定は母数を含んでいると判断すると, 当たっている確率としての幅であり, 判断そのものは当たっているか外れているかの 0/1 判断である。同様のことが帰無仮説検定における判断にも言える。改めて帰無仮説検定の判断ロジックを確認しておく。

→ 豊田 (2016) の P.42–43.

区間による判定 ベイズ法を用いて群間差を評価する方法として, 区間推定の結果を用いる ROPE を使う方法について解説する。区間が母数の分布を含んでいるために判断がしやすい反面, 明確な基準

がなく，また ROPE をどのように設定するのかについてはドメイン知識が必要になることが難点であるが，これはそもそも従来法でも考えるべき問題であった。またベイズ法による事後分布は一つに限定されるので，多重比較の問題が生じないことにも触れる。

→ Kruschke (2014 前田・小杉監訳 2017) の P.303-334

ベイズファクター ベイズ法によるもう一つの判断方法は，帰無仮説検定と同じくモデル比較の観点から考えることである。ベイズファクターは周辺尤度の比からいずれのモデルがデータに支持されているかを表す指標であり，帰無仮説モデルと対立仮説モデルをこれで比較することによって，どちらがよりデータに適合したモデルであるかを評価することができる。特徴として，帰無仮説のような Null 仮説を積極的に支持することができる点がある。そもそも帰無仮説モデルが対立仮説モデルに比べて非常に限定的であること，ネストされたモデルであることを指摘し，このような場合はサヴェージ・ディッキー法で比較できることを解説する。

→ Lee・Wagenmakers (2013 井関訳 2017) の P.88–103.

キーワード

- 標本分布と事後分布
- 区間推定と点推定
- ベイズファクター
- サヴェージ・ディッキー法

E.27.2 授業情報

■コマの展開方法 講義

■標準シラバスにおける位置づけ 科目番号 5；心理学統計法；(1) 心理学で用いられる統計手法；J より高度な記述や推定を目指して

予習・復習課題

■予習 帰無仮説検定の前段階である，母数の区間推定について改めて理解しておく。また帰無仮説検定が推定結果からある種の判断を行うためのツールであったこと，科学的探究のためには必ずしも 0/1 の判断結果に陥る必要がないことを確認しておく。

■復習 帰無仮説検定の時と違って，どのような結論の出し方や解釈の仕方になるのか，しっかり理解しておこう。統計ツールとして R のほかに JASP が導入される。JASP は 2021 年 12 月現在でもバージョン 1.0 に満たないソフトウェアではあるが，ベイズ法と従来法を並列的に活用できるツールであるので，興味があれば自分の PC にダウンロードして試してみるのも良いだろう。

E.28 JASP によるベイズ推定の実践

E.28.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [14] モデル [15] 推測・推定 [9] 検定 [6] 実践 [6]

概要

ベイズ統計を用いた考え方は，従来の推定やモデル比較と異なる観点からのアプローチができるようになる。もっとも原理的に優れたものであっても，実践できないのであれば絵に描いた餅にすぎない。JASP の登場により，GUI で比較的簡単にベイズ法による分析，実践を行うことが可能になった。ここでは JASP の導入と，JASP による回帰分析，帰無仮説検定を行い，従来法とベイズ統計とでどのような違いがあるのかを演習的に学ぶ。なお，JASP のバージョンは本シラバス執筆時点[*4]で ver.0.16 であり，未だ発展途上で急速に開発が進んでいる段階であるため，表示や操作の方法が今後変更になる可能性が高いことに留意する。

コマ主題細目 (箇条書き)

ベイズ推定の方法 ベイズ推定を実際に計算するためには，解析的な方法やグリッドサーチ，そして MCMC による推定方法が必要になる。とくにマルコフチェイン・モンテカルロ法は，困難であったベイズ推測を可能にした技術的ブレイクスルーであった。基本的なアイデアは，確率分布の形状を求める際に乱数を使って近似するということである。単純な乱数を用いた例から，これによる確率分布の近似と積分計算の簡略化ができることを理解し，事後分布の相対的な形状近似ができるようになったことを学ぶ。その上で，便利なアプリケーションである JASP の紹介を行う。

→ Kruschke (2014 前田・小杉監訳 2017) の 7 章，馬場 (2019)P.62–76

JASP の導入 JASP のインストールと起動，その基本的な特徴の解説など導入する演習を行う。各自の PC に JASP 環境を導入し，サンプルコードを開いて記述統計量やプロットを描く。現段階[*5]では JASP を用いたテキストや解説書は少ないが，標準的なアプリケーションの導入で対応できるため，初学者にもその障壁は低いと考えられる。

JASP による回帰分析 JASP の特徴の一つは，従来法とベイズ法が同じメニューから並列的に利用できることである。ここでは回帰分析を例に，最尤法とベイズ法で結果の表示の違いや解釈の方法について見ていく。

JASP による平均の比較 一要因二水準，一要因三水準の平均の比較について，従来法とベイズ法の両方を JASP で実践する。帰無仮説検定における p 値の解釈と，ベイズ法におけるベイズファクターでの評価それぞれについて，演習的に学ぶ。

キーワード

- MCMC 法
- JASP

[*4] 2021 年 12 月 11 日
[*5] 2021.12.01 現在

- BF

E.28.2 授業情報

■コマの展開方法 演集

■標準シラバスにおける位置づけ 科目番号 5;心理学統計法; (1) 心理学で用いられる統計手法; J より高度な記述や推定を目指して

予習・復習課題

■予習 本講では新しい統計ソフト，JASP が導入される。授業開始前に JASP を自らの PC にインストールし，起動するかどうかを確認しておこう。

■復習 これまでのデータ分析は R で行ってきたが，同じデータを JASP で行えばどうなるのか，進んで実行してもらいたい。また JASP にはサンプルコードや OSF との連携機能も含まれているので，これら既存のデータを使って JASP がどのような振る舞いをするのか，さまざまに試してみると良いだろう。

E.29 R によるベイズ推定

E.29.1 授業内容

科目の中でのこのコマの位置づけ

記述統計 [3] 確率 [15] モデル [16] 推測･推定 [10] 検定 [6] 実践 [7]

概要

JASP によるベイズ推測は，GUI で非常に簡便的に行うことができた。R でももちろんベイズ統計を扱うことができるが，初学者にはその準備に障壁がある。まず実践的なベイズ統計の場合は，MCMC 法を利用することになるが，この MCMC 法の概要について理解する。その上で，この MCMC 法を可能にした確率的プログラミング言語，JAGS や Stan を導入する。これらの言語を利用するには，インストールの際に困難を伴う可能性があり，またプログラミング言語で記述するのも初学者には敷居が高く感じられるところである。そこで Stan をバックエンドで利用するパッケージ，brms を導入する。R パッケージの brms を使うと，一般線形モデルが簡単に実装できる。関数名が lm から brm に変わるだけで，MCMC をつかった線形モデルのベイズ推定を行うことができ，またその結果の可視化も簡単に行える。lm 関数と同じく，説明変数を離散型 (factor 型) にしておくと，要因の効果の大きさが推定でき，brms パッケージの `marginal_effect` 関数を使うと要因ごとに周辺化された結果も図示され，分かり易い。ベイズ統計学的アプローチを，絵に描いた餅にせずに技術として習得することが狙いである。

コマ主題細目 (箇条書き)

回帰分析のベイズ推定 最も単純な例として，単回帰分析のベイズ推測を brms パッケージを使って行う。brms パッケージの導入にあたっては，Rtools というコンパイラを含んだ追加ツールの導入が必要であるため，これをインストールして環境を準備するところから始める。準備ができれば，データを読み込み，可視化し，brms パッケージによって回帰分析を実行，その結果を確認する。コンパイルや MCMC サンプリングに少し時間がかかるが，計算機のアルゴリズムを思い出して優しい気持ちで

9810 待つ。

9811 → 馬場 (2019) の P.186–195.

9812 **モデルを用いた予測** 回帰分析の結果をプロットし，確信区間をはじめ事後分布の可視化を行う。ベイズ推
9813 定の場合は結果が分布として得られるため，いかなる代表値も全情報の一部にすぎないことを確認す
9814 る。続いて同じモデルを使った予測も図示する。確信区間と予測区間の幅がどうして異なるのか，モデ
9815 ルに基づいて理解する。

9816 → 馬場 (2019) の P.196–199.

9817 **MCMC を使った分析のレポート** brms は MCMC 法によって近似的な解を得るものであり，その結果
9818 の報告には事後分布や確信区間だけでなく，事前分布や MCMC に必要な情報 (バーンイン期間，サ
9819 ンプル数，チェイン数，thining の数) も記載する必要がある。ここではそれぞれについて解説し，実
9820 際にレポートする練習を行う。

9821 **キーワード**

9822 - MCMC
9823 - brms パッケージ
9824 - ベイズ推定
9825 - 事後予測分布＼

9826 ## E.29.2 授業情報

9827 **■コマの展開方法** 講義

9828 **■標準シラバスにおける位置づけ** 科目番号 5;心理学統計法; (1) 心理学で用いられる統計手法; J よ
9829 り高度な記述や推定を目指して

9830 **予習・復習課題**

9831 **■予習** ベイズ法による推定の実際は，パッケージを使って行われる。とくに brms パッケージは一般線形モ
9832 デルのみならず，一般化線形モデルのベイズ的実践に用いられるパッケージである。これを使った推定の良
9833 書として馬場 (2019) がある。この本は R や ggplot2 パッケージの使い方など，一冊で統計分析ができるよ
9834 うになっているので，今回の講義内容に合わせて事前に読んでおくとよい。

9835 **■復習** ベイズ法を使った心理学論文はまだ数が少ないが，今後増えていくことが予見されている。レポー
9836 ティングの方法についてもしっかり理解しておく必要がある。そのため，授業時間内に課される課題を Rmd
9837 形式で提出することが求められる。時間内に学んだ技術を用いて，レポートを提出すること。なおこのレポート
9838 の提出は前期の単位認定にあたっての必要条件であり，基準に満たないレポートは再提出を求める。前期試
9839 験の前にパスすることが，前期末試験を受験するための前提条件であることを承知しておくこと。

E.30 後期のテスト

E.30.1 授業内容

科目の中でのこのコマの位置づけ

後期で学んできたことについて，基本的な知識，用語の理解，心理統計の基本的な考え方の筋道などを等テストをする。

E.30.2 授業情報

■コマの展開方法 マークシート式の試験を行う。問題数は 60-70 問程度である。

■諸注意

- 授業を行なっている教室で試験を実施する。着席場所は指定しないが，机の両端に座り，他の受験生と隣接しないようにする。
- 試験に際して資料等の持ち込みは認めない。
- 携帯電話，タブレット，PC など通信機能を有する機器の使用も認めない。
- (関数)電卓がなくても計算ができるよう工夫された出題をしているため，電卓などの使用も認めない。
- 受験者数を確定するため，当日は出席管理システムへの登録を行なってもらう。
- 不正行為を見つけた場合は即時退出を求め，単位を与えない。
- 試験会場に 20 分以上遅刻したものは受験を認めない。
- 試験開始から 30 分以上経過すると，希望者には退出を許可する。
- 退出に当たっては，質問用紙と解答用紙の両方を提出すること。
- マークシートへの回答であるため，鉛筆の利用が望ましい。
- マークをすべて塗りつぶしていないなど，読み取り上のエラーには対応しない。
- エクストラクレジットについての説明は別途行う。

引用文献

Amrhein, Valentin, Greenland, Sander, & McShane, Blake (2019). Scientists rise up against statistical significance.

馬場 真哉 (2019). R と Stan ではじめる ベイズ統計モデリングによるデータ分析入門 講談社

Bürkner, Paul-Christian (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software , 80 (1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, Paul-Christian (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal , 10 (1), 395–411. https://doi.org/10.32614/RJ-2018-017

Carpenter, Bob, Gelman, Andrew, Hoffman, Matthew D, Lee, Daniel, Goodrich, Ben, Betancourt, Michael, Brubaker, Marcus, Guo, Jiqiang, Li, Peter, & Riddell, Allen (2017). Stan: A probabilistic programming language. Journal of statistical software , 76 (1).

Giegernzer, G. (2002). Calculated Risks.
(ギーゲルンツァー, G. 吉田 利子 (訳) (2003). 数字に弱いあなたの驚くほど危険な生活——病院や裁判で統計にだまされないために—— 早川書房)

Gilboa, Itzhak. (2009). Theory of Decision under Uncertainty.
(ギルボア, I. 川越 敏司 (訳) (2014). 不確実性下の意思決定理論 勁草書房)

Grimm, Laurence・Yarnold, Paul. (1994). Reading and Understanding Multivariate Statistics. American Psychological Association.
(グリム, L.G. & ヤーノルド, P.R. 小杉 考司・高田 菜美・山根 嵩史 (訳) (2016). 研究論文を読み解くための多変量解析入門 基礎篇: 重回帰分析からメタ分析まで 北大路書房)

Grimm, Laurence・Yarnold, Paul. (2001). Reading and Understanding More Multivariate Statistics. American Psychological Association.
(グリム, L. & ヤーノルド, P. 小杉 考司・高田 菜美・山根 嵩史 (訳) (2016). 研究論文を読み解くための多変量解析入門 応用篇: SEM から生存分析まで 北大路書房)

Hadley, Wickham (2014). Tidy Data. Journal of Statistical Software , 59, 1–23. https://doi.org/10.18637/jss.v059.i10

南風原 朝和 (2002). 心理統計学の基礎——統合的理解のために—— 有斐閣

浜田 宏 (2018). その問題、数理モデルが解決します ベレ出版

浜田 宏・石田 淳・清水 裕士 (2019). 社会科学のためのベイズ統計モデリング 朝倉書店

橋本 貴充・荘島 宏二郎 (2016). 実験心理学のための統計学——[心理学のための統計学 2]: t 検定と分散分析—— 誠信書房

Healy, Kieran. (2018). Data Visualization: A Practical Introduction. Princeton Univ Pr.
(キーラン・ヒーリー 瓜生 真也・江口 哲史・三村 喬生 (訳) (2021). データ分析のためのデータ可視化入門 講談社)

蛇蔵 (2014). 決してマネしないでください。(1) 講談社コミックス

平川 真 (2021). 2 変数の関係を表す統計量の推定 清水 裕士 (編) 心理学統計法 (pp. 150–162) 放送大学教育振興会

平岡 和幸・堀 玄 (2009). プログラミングのための確率統計 オーム社

Huff, Darrell. (1954). How to Lie with Statistics. W. W. Norton & Company.
(ダレル・ハフ 高木 秀玄 (訳) (1968). 統計でウソをつく法——数式を使わない統計学入門—— 講談社)

池田 功毅・平石 界 (2016). 心理学における再現可能性危機:問題の構造と解決策 心理学評論 , 59 (1), 3–14. https://doi.org/10.24602/sjpr.59.1_3

池内 了 (2014). 宇宙論と神 (集英社新書) 集英社

Isaacson, Walter. (1995). Steve Jobs. JSTOR.
(アイザクソン, W. 井口 耕二 (訳) (2011). スティーブ・ジョブズ 講談社)

石田 基広・石田 和枝 (2019). 女子高生乱子によるベイズ統計学入門講座——とある弁当屋の統計技師—— 共立出版

Kant, Immanuel. (1787). Critique of Pure Reason.
(カント 篠田 英雄 (訳) (1961). 純粋理性批判 岩波書店)

川端 一光・荘島 宏二郎 (2014). 心理学のための統計学入門——[心理学のための統計学 1]: ココロのデータ分析—— 誠信書房

菊池 聡 (2018). 心理学者は誰の心も見透かせるの?– 学問と偽科学の違い 楠見 孝・日本心理学会 (編) 誠信書房

椎名 乾平 (2013). 七つの正規分布 心理学評論 , 56 (1), 7–34. https://doi.org/10.24602/sjpr.56.1_7

小杉 考司 (2018). 言葉と数式で理解する多変量解析入門 北大路書房

小杉 考司 (2019). R でらくらく心理統計 : RStudio 徹底活用 講談社

Kruschke, John K. (2014). Doing Bayesian Data Analysis. Elsevier.
(クルシュケ, J.K. 前田 和寛・小杉 考司 (監訳) 前田 和寛・小杉 考司・井関 龍太・井上 和哉・鬼田 崇作・紀ノ定 保礼・国里 愛彦・坂本 次郎・杣取 恵太・高田 菜美・竹林 由武・徳岡 大・難波 修史・西田 若葉・平川 真・福屋 いずみ・武藤 杏里・山根 嵩史・横山 仁史 (訳) (2017). ベイズ統計モデリング: R, JAGS, Stan によるチュートリアル 原著第 2 版 共立出版)

Kruschke, John K (2018). Rejecting or accepting parameter values in Bayesian estimation. Advances in Methods and Practices in Psychological Science , 1 (2), 270–280.

熊谷 龍一・荘島 宏二郎 (2015). 教育心理学のための統計学——[心理学のための統計学 4]: テストでココロをはかる—— 誠信書房

黒崎 政男 (2000). カント『純粋理性批判』入門 (講談社選書メチエ) 講談社

Lambert, Ben (2018a). A Student's Guide to Bayesian Statistics (1st ed.). SAGE Publications Ltd.

Lambert, Ben (2018b). A student's guide to Bayesian statistics. Sage.

Lee, M. D. · Wagenmakers, Eric-Jan. (2013). Bayesian Cognitive Modeling: A Practical Course. Cambridge University Press.
(リー, M.D & ワゲンメーカーズ, E-J. 井関 龍太 (訳) (2017). ベイズ統計で実践モデリング: 認知モデルのトレーニング)

Makowski, Dominique, Ben-Shachar, Mattan S., & Lüdecke, Daniel (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software , 4 (40), 1541. https://doi.org/10.21105/joss.01541

Maor, Eli. (1994). e:the story of a number. Princeton University Press.
(マオール, E. 伊理 由美 (訳) (1999). 不思議な数eの物語 筑摩書房)

Matejka, Justin & Fitzmaurice, George (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing Proceedings of the 2017 CHI conference on human factors in computing systems, 1290–1294.

松村 優哉・湯谷 啓明・紀ノ定 保礼・前田 和寛 (2021). 改訂 2 版 R ユーザのための RStudio[実践] 入門 ——tidyverse によるモダンな分析フローの世界—— 技術評論社

McGrayne, Sharon Bertsch. (2011). The Theory That Would Not Die. Yale University Press.
(シャロン・バーチュ・マグレイン 冨永 星 (訳) (2018). 異端の統計学ベイズ 草思社)

Merkle, Edgar C., Fitzsimmons, Ellen, Uanhoro, James, & Goodrich, Ben (2021). Efficient Bayesian Structural Equation Modeling in Stan. Journal of Statistical Software , 100 (6), 1–22. https://doi.org/10.18637/jss.v100.i06

Merkle, Edgar C. & Rosseel, Yves (2018). blavaan: Bayesian Structural Equation Models via Parameter Expansion. Journal of Statistical Software , 85 (4), 1–30. https://doi.org/10.18637/jss.v085.i04

道又 爾 (2009). 心理学入門一歩手前——「心の科学」のパラドックス—— 勁草書房

三中 信宏 (2018). 統計思考の世界——曼荼羅で読み解くデータ解析の基礎—— 技術評論社

皆本 晃弥 (2015). スッキリわかる確率統計——定理のくわしい証明つき—— 近代科学社

三浦 麻子 (2017). なるほど! 心理学研究法 (心理学ベーシック第 1 巻) 北大路書房

水本 篤・竹内 理 (2011). 効果量と検定力分析入門 : 統計的検定を正しく使うために 2010 年度部会報告論集「より良い外国語教育のための方法」

Muller, Jerry・Muller, Jerry Z. (2018). The tyranny of metrics. Princeton University Press.
(ミュラー, ジェリー & ミュラー, ジェリー・Z 松本 裕 (訳) (2019). 測りすぎ——なぜパフォーマンス評価は失敗するのか?—— みすず書房)

長沼 伸一郎 (2016). 経済数学の直観的方法 確率・統計編 講談社

永田 靖・吉田 道弘 (1997). 統計的多重比較法の基礎 サイエンティスト社

中西 大輔・今田 純雄 (編) (2020). あなたの知らない心理学——大学で学ぶ心理学入門—— ナカニシヤ出版

西内 啓 (2013). 統計学が最強の学問である ダイヤモンド社

西内 啓 (2017). 統計学が最強の学問である [数学編]——データ分析と機械学習のための新しい教科書—— ダイヤモンド社

Norretranders, Tor. (1999). The user illusiton. Penguin.
(ノーレットランダーシュ, T. 柴田 裕之 (訳) (2002). ユーザーイリュージョン——意識という幻想——紀伊國屋書店)

大芦 治 (2016). 心理学史 ナカニシヤ出版

大久保 街亜・岡田 謙介 (2012). 伝えるための心理統計: 効果量・信頼区間・検定力 勁草書房

大村 平 (2013). 実験計画と分散分析のはなし——効率よい計画とデータ解析のコツ—— 日科技連出版社

岡太 彬訓 (2008). データ分析のための線形代数 共立出版

奥村 晴彦・瓜生 真也・牧山 幸史 (2018). R で楽しむベイズ統計入門 [しくみから理解するベイズ推定の基礎] (Data Science Library) 技術評論社

小塩 真司 (2020). 性格とは何か——より良く生きるための心理学—— 中央公論新社

Rosling, Hans・Rosling, Ola・Rönnlund, Anna Rosling. (2018). Factfulness. Brockman Inc. (H. ロスリング・O. ロスリング, & A.R ・ロンランド 上杉 周作・関 美和 (訳) (2019). FACTFULNESS : 10 の思い込みを乗り越え、データを基に世界を正しく見る習慣 日経 BP 社)

繁桝 算男 (2019). ベイズ統計学 楠見 孝・日本心理学会 (編) 公認心理師の基礎と実践 5 心理学統計法 (pp. 275–296) 遠見書房

清水 裕士 (2021). 心理学統計法 放送大学教育振興会

下山 晴彦 (2001). 臨床心理学研究の多様性と可能性 下山 晴彦・丹野 義彦 (編) 講座臨床心理学 (pp. 3–24) 東京大学出版会

Shojima, Kojiro (2022). Test Data Engineering: Latent Rank Analysis, Biclustering, and Bayesian Network. Springer.

安井 翔太 (2019). 効果検証入門——正しい比較のための因果推論／計量経済学の基礎—— 技術評論社

Stallman, Richard M. (2002). Free software,free society. Lulu. com. (ストールマン, R.M. 長尾 高弘 (訳) (2003). フリーソフトウェアと自由な社会 アスキー)

Stevens, Stanley Smith (1946). On the theory of scales of measurement. Science , 103 (2684), 677–680.

末永 俊郎 (編) (1987). 社会心理学研究入門 東京大学出版会

高橋 康介 石田 基広 (編) (2018). 再現可能性のすゝめ 共立出版

高橋 将宜 石田 基広・市川 太祐・高橋 康介・高柳 慎一・福島 真太朗・松浦 健太郎 (編) (2022). 統計的因果推論の理論と実装 共立出版

Torchiano, Marco (2020). effsize: Efficient Effect Size Computation. R package version 0.8.1. https://doi.org/10.5281/zenodo.1480624. from https://CRAN.R-project.org/package=effsize

豊田 秀樹・前田 忠彦・柳井 晴夫 (1992). 原因をさぐる統計学——共分散構造分析入門—— 講談社

豊田 秀樹 (2009). 検定力分析入門——R で学ぶ最新データ解析ー—— 東京図書

豊田 秀樹 (2015). 基礎からのベイズ統計学——ハミルトニアンモンテカルロ法による実践的入門—— 朝倉書店

豊田 秀樹 (2016). はじめての 統計データ分析——ベイズ的〈ポスト p 値時代〉の統計学—— 朝倉書店

豊田 秀樹 (2017). もうひとつの重回帰分析 東京図書

豊田 秀樹 (2020). 瀕死の統計学を救え！ 朝倉書店

Wasserstein, Ronald L & Lazar, Nicole A (2016). The ASA statement on p-values: context, process, and purpose.

渡辺 茂 (2019). 動物に「心」は必要か: 擬人主義に立ち向かう 東京大学出版会

Wickham, Hadley (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, McGowan, Lucy D'Agostino, François, Romain, Grolemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas Lin, Miller, Evan, Bache, Stephan Milton, Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana Paige, Spinu, Vitalie, … Yutani, Hiroaki (2019). Welcome to

the tidyverse. Journal of Open Source Software , 4 (43), 1686. https://doi.org/10.21105/joss.01686

Wood, Simon N. (2016). Just Another Gibbs Additive Modeler: Interfacing JAGS and mgcv. Journal of Statistical Software , 75 (7). https://doi.org/10.18637/jss.v075.i07

山田 剛史・村井 潤一郎 (2004). よくわかる心理統計 ミネルヴァ書房

山内 光哉 (2010). 心理・教育のための統計法 サイエンス社

野島 一彦・繁桝 算男・山田 剛史 (編) (2019). 心理学統計法 遠見書房

吉野 諒三 (2007). 公理的測定論の基礎――心理表現の論理と実際―― 吉野 諒三・山岸 侯彦・千野 直仁 (編) 数理心理学 (pp. 49–108) 培風館

結城 浩 (2018). 数学ガールの秘密ノート/行列が描くもの (数学ガールの秘密ノートシリーズ) SB クリエイティブ

Descartes, René. (1637). Discours de la méthode.
(ルネ・デカルト 谷川 多佳子 (訳) (1997). 方法序説 岩波書店)

世界単位認定協会 (2017). 新しい単位 ポプラ社

総務省 (2020). 統計表における機械判別可能なデータ作成に関する表記方法.

索引

C

Cohen の d 249
CRAN 40, 409

E

EAP 推定値 339, 377

F

F 分布 315

G

ggplot2 364

H

HDI 357
Hedges の g 249

J

JAGS 360
JASP 361

K

ニット 61, 408

M

MAP 推定値 339
MCMC 358
MED 推定値 340

P

p-hacking 249
p 値 236

Q

Q-Q プロット 322
QRPs 245

R

R 40
ROPE 357, 368
RStudio 41
R^2 値 150, 155

S

SR 図式 95
Stan 360

T

t 検定 254, 260, 262
t 分布 228

W

Welch の補正 263

あ

アウトカム変数 27
α 係数 102
一様分布 336
一般化線形モデル 318, 355
一般線形モデル 203, 269, 318
因果関係 137, 141
因果効果 347
応答変数 26

か

回帰係数 147
回帰分析 139, 147, 153
下位検定 278
ガウス分布 115
確信区間 340, 341, 348
確率 107, 329, 387
確率質量 109
確率質量関数 111
確率的プログラミング言語 119, 360, 374
確率分布 109
確率分布関数 109
確率変数 108, 215
確率密度 110
確率密度関数 111, 314
確率モデル 95, 305, 318
可視化 30
可視化の文法 70
片側検定 259
間隔尺度水準 27, 29, 203, 286, 363
関数 143
間接効果 160
観測変数 27
偽陰性 331
棄却 236, 243
棄却域 237
記述統計 213
基準関連妥当性 104
期待値 112
帰無仮説 234, 243, 347, 387
帰無仮説検定 233, 347
逆確率の法則 328
客観性 22
球面性の検定 298
偽陽性 331
共分散 100, 130
共役事前分布 357
行列 205
偶然誤差 99, 115, 178, 181, 184, 315
区間推定 223, 227, 387
クロンバックの α 102
群間要因 180, 199, 201
群内要因 180, 199, 201, 291
系統誤差 98, 115, 315
結果変数 113
欠損値 76
決定係数 150
検定 231, 233, 387

検定力 247, 249, 266, 387
検定力分析 249
効果量 93, 231, 249, 348
交互作用 192, 279
構成概念 24
構成概念妥当性 103
構造方程式モデリング 26
交絡 177
交絡要因 138
誤差 98, 115, 144
個人差 181, 199
固定効果 207, 384
古典的テスト理論 98, 185, 216
コピペ汚染 59, 62
ゴルトンボード 116
混合効果モデル 207
混合要因計画 180
コンソール 42

さ

再検査信頼性 101
再現性 22
最高密度区間 340, 351
最小値 89
最小二乗法 146, 153, 315, 320
最大事後確率 339
最大値 89
採択 236, 243
最頻値 83
最尤推定 316
最尤推定値 307
最尤法 154, 273, 307, 341, 388
サヴェージ・ディッキー法 353, 371
作業フォルダ 400
残差 144, 150
残差診断 323
散布図 141
サンプルサイズ 215, 249
サンプル数 215
式 168
事後確率 329
事後検定 278
事後分布 335, 337, 388
指数関数 125
事前確率 329
事前分布 335, 337, 348
実験群 179
実現値 216
実質的に等価な範囲 351
質的変数 22
私秘性の排除 18
四分位 88, 129, 340
四分位範囲 88
尺度 239
尺度化 22
重回帰分析 153, 165, 171, 191, 199, 269, 321
従属変数 26, 144, 179
自由度 206, 315
周辺化 352, 360
周辺確率 326
周辺尤度 337, 352
主観確率 108, 329, 388
主効果 192, 348
順序尺度水準 27, 28, 109, 286, 363
条件つき確率 326
真のスコア 98
信用区間 341
信頼区間 227, 348
信頼性 21, 101, 185
心理尺度 91
心理測定法 20, 21, 29, 239, 388
水準 180
推測統計学 123, 207, 213, 387
推定 387
数学的現象主義 17, 95
数理モデル 95
スタナイン 88
正規分布 115, 116, 178, 315
整然データ 78
絶対パス 400
折半法 102
説明変数 25, 144, 179
線形回帰 153
線形関数 143
線形モデル 153, 167, 168, 184, 347
潜在変数 27
相関関係 137, 141
相関係数 100, 131
相対パス 400

た

対応のある 181
対応のない 181
対数 309
大数の法則 219
対数尤度 319
タイプ 1 エラー 246, 331, 347
タイプ 2 エラー 247, 249, 331, 347
対立仮説 234, 243, 387
多重共線性 160
多重比較 269
妥当性 21, 102
多変量解析 26
多変量分散分析 26
単回帰分析 153, 315
チャンク 61, 406
中央値 83
中心極限定理 217
データハンドリング 65
データフレーム 54
停止規則 354
適合度 150, 155, 169
デザイン行列 205
テスト理論 98
てっちゃんの手品 104
点推定 223
統計的推定 215
統計モデル 95
同時確率 325
等裾区間 340
統制群 179
特徴量エンジニアリング 70
独立 326
独立同分布 312
独立変数 25, 144, 179

な

内的整合性信頼性 102
内容的妥当性 104
二項分布 114, 115, 318, 357
二重盲検法 178

は

パーセンタイル 88, 340
箱ひげ図 88
外れ値 83
パッケージ 50

パラメータ 111
範囲 89
反復測定 181
引数 49
被説明変数 26, 144, 179
非線形関数 144
批判的思考 138
標準化 123, 161
標準化スコア 91
標準化偏回帰係数 162
標準誤差 220, 387
標準正規分布 119, 120
標準偏差 88
標本 214, 387
標本抽出 214
標本統計量 215
標本平均 315
被予測変数 26
比率尺度水準 27, 29, 179, 286, 363
頻度主義 108, 329
部分相関係数 157
不偏分散 224, 238, 315, 387
分散 85, 97
分散分析 186, 269, 275, 315
分散分析表 277
ベータ分布 357
平均因果効果 179
平均値 82
平均平方 277, 315
平均偏差 86
並行検査信頼性 102
ベイジアンモデリング 355
ベイズファクター 352, 353, 357, 367
ベイズ法 154, 273, 323, 325, 335, 341, 349, 353, 371
平方和 185, 194, 269, 275, 276, 315
ベクトル 205
ベルヌーイ分布 318, 357
偏回帰係数 155, 159
偏差値 91
変数 22
偏相関係数 158
変量効果 207, 384
ポアソン分布 318
母集団 213, 347, 387
母集団分布 216
母数 214, 387

ま

マルコフチェインモンテカルロ法 358
マルコフ連鎖 358
無作為標本抽出 214
無情報事前分布 336
名義尺度水準 27, 28, 73, 109, 179, 190, 203, 274, 286, 363
モーメント法 221, 273, 315, 341, 387
目的変数 26
モデリング 355
モデル 19, 95, 141, 143, 203
問題のある研究実践 245
モンティ・ホール問題 332

や

有意水準 235, 247, 249, 347, 387
尤度 307, 319, 329, 337
尤度関数 307
要因 180
要因計画 347
予測子 25
予測値 144

ら

ランダム化 178
ランダム化比較実験 139, 179
離散分布 109
リスト 53
両側検定 259
量的変数 22
臨界値 237
例数設計 248, 249
連続分布 109
ロング型 78

わ

ワイド型 77