心理統計学

吉村宰

長崎大学アドミッションセンター

目 次

第1章　はじめに

1.1　この文書について

この文書は吉村が自分が担当する授業用に作成したものであり公表を目的とするものではありません。作成途中であり不完全です。「授業中に補足説明すればいいや」と考えているので，記述には間違いも含まれますし引用関係も明確ではありません。利用に当たっては下記に留意し「自己責任」でお願いします。

- 内容に責任は持てません。自身でウラをとって下さい。
- 複製及び二次的配布は止めてください。
- 面識のある方からの質問・間違いの指摘については歓迎します。どうぞご遠慮なく。

1.2　なぜ「統計」なのか

理由は簡単,「必要だから」である[1]．心理学系ゼミへの所属を希望する者にとっては，より専門的な文献を読むために，卒業研究を進めるために，統計的手法についての十分な理解が欠かせない．教員になるのであれば，教育評価の場面で統計的手法が必要になる（残念ながら，十分なスキルを持った教員はごく少数というのが現状だろう）．

それ以外の者にとっても全くの無駄にはならない．現代は高度情報化社会である．誰でも簡単に大量の比較的質のよいデータを手に入れることができるし，そのようなデータに触れる機会も格段に増えている．そうしたデータから有用な情報を読みとりたい，引き出したい，というときには統計的手法が役に立つ．

この授業で学習するのは，統計「学」ではなく，統計の「使い方」である.「使い方」は場面・状況によって様々であるが，ここでは主として心理学研究の場面を想定して話を進める．

1.3　統計を勉強して統計から自由になろう

心理学研究における統計的手法の「使い方」には，心理学独自の部分（ローカルルール）がある，ということを知っておいて欲しい．心理学研究の文脈においては「正しいこと」「そうすべきこと」とされていても，統計学的には,「妙なこと」「意見の別れること」「よく分からないこと」「間違っていること」「どうでもよいこと」などが結構ある．率直に言うと，心理学研究における統計の使い方を見ると「細かいことにウルサくて窮屈で，その割には大事な部分には無頓着だなあ」と感じる．

心理学研究における統計の「誤用」はこれまで何度も指摘されてきた．しかし事態は改善されず，例えばアメリカ心理学会では「心理学研究の危機」という認識にまで至っている[2].「こういう場合にはこの分析をする」というようなHow to的な統計手法の利用には,「なぜその分析を行うのか」という「思考」がない．「よく分からない」で済まされる問題ではない．自分が何をやっているかよく分からない「研究」などあり得ないからである．

---

[1] 個人的には必ずしも必要であるとは考えない．統計を使わない心理学研究がもっとたくさんあってもよい

[2] できれば the American Psychological Association (APA) による「統計利用に関するガイドライン」参照のこと．これは WEB 上に公開されている（URL：http://www.apa.org/journals/amp/amp548594.html）

心理学研究のローカルルール（こういう場合にはこうしなければならない）が心理学の研究者自身の足枷となっている側面があることは否定できない．この不自由さは研究をツマラナイものにする．以下は「The Earth is Round ($p < .05$)」(Cohen, 1994) という論文からの引用である．

> 同僚が「統計について質問がある」といって私を訪ねてきた．彼は，ある非常にまれな病気が特定の母集団（例えば日本）には存在しないと考えている．つまり帰無仮説 $H_0: P_0 = 0$ である（P はその病気が母集団において存在する確率）．そしてその母集団から 30 程のランダムサンプルを抽出して調べたところ，1 名にその病気が確認された．つまり P_s（サンプルに基づく P_0 の推定値）$= 1/30 = .033$ である．彼は，帰無仮説 H_0 の検定を行うのに，Yates の修正を施した χ^2 検定と Fisher の exact test のどちらを使えばよいのか，また検定力は十分かどうか，がよく分からないという．さらに彼は，統計的仮説検定をしなければこの論文は査読を通らないだろう，ともいうのである．そんなバカな！

問：上の話のどこが「そんなバカな」なのだろうか

なぜ，こういうことが起こるのか．それは統計という道具をきちんと理解しないまま使うからである.「統計の使い方」と「統計ソフトの使い方」とを混同してはならない．ある統計的手法を適用する場合には，それが何を意味するのかを理解していなければならない．なぜその手法を使うかをきちんと説明できなければならない.「先行研究でそうしていたから」というのは理由にはならない．

「〜ならない」の連続でウンザリしたかも知れないが，逆に言えば，きちんと理解さえしておれば，先行研究の方法（形式）に縛られず，自分のアイデアを活かして自由に研究ができる，ということである．

1.4 学習を進める上でのアドバイス

統計的手法を学習する場合，複数の文献を参考にするとよい．同じ事柄についての説明も文献によって結構異なる場合があることに気付くだろう．統計学には,「これが唯一正解である」とは言えないような事柄が結構ある．統計学はたった 1 つのものではなく，いくつかの立場がある．立場が異なれば主張も異なる．A という立場から見れば「NG」でも，B という立場では「OK」だったりする．

数式を避けないことが重要．統計的手法の説明には必ず数式が登場する．これを嫌がっていてはきちんと理解できない．数式が「考え方」そのものだからである．統計の説明で登場する数式は，実はそれほど難しくない．ゆっくり読み進めれば必ず分かる．内容は，四則演算，簡単な因数分解（中学生レベル），指数・対数（高校生レベル）である．微積分も少し出てくるが直感的な理解で問題はない.「難しい」という人は，実は数式を見て嫌がっているだけで，内容を理解しようとしたことのない人だと思ってよい．以下に文献を紹介する．

- 「心理学論の誕生」，佐藤達哉他，北大路書房，￥2,800 （心理学研究のありかたについての議論）
- 「実践としての統計学」，佐伯胖他，東京大学出版会，￥2,600 （統計の使われ方についての議論）
- 「本当にわかりやすいすごく大切なことが書いてあるごく初歩の統計の本」，吉田寿夫，北大路書房，￥2,500 （タイトルの通り．統計的仮説検定の話が中心）
- 「Q & A 心理データ解析」，服部環他，福村出版，￥2,600 （心理分野で用いられる主な手法を網羅）
- 「違いを見ぬく統計学—実験計画と分散分析入門—」，豊田秀樹，講談社ブルーバックス，￥940
- 「複雑さに挑む科学—多変量解析入門—」，柳井晴夫他，講談社ブルーバックス，￥880

- 「原因をさぐる統計学—共分散構造分析入門—」，豊田秀樹他，講談社ブルーバックス，￥760
- 「心理・教育のための多変量解析法入門（基礎編）」，渡部洋，福村出版，￥2,600
  （数式があまり出てこない多変量解析の本）
- 「Q & A で知る統計データ解析 DOs and DON'Ts」，繁桝算男他，サイエンス社，￥2,400
  （「Q」が結構マニアックなので，初学者にはあまり使えないかも）
- 「統計的方法のしくみ」，永田靖，日科技連，￥2,500 （難しく感じるかもしれない）
- 「入門数理統計学」，ホーエル，培風館 （難しく感じるに違いない．しかし重要）

第2章　心理学研究の概要

2.1　心理現象を説明する

心理学研究は一般に，心理現象の成立の仕組みや法則を明らかにすることを目的とする．ある心理現象の仕組みや法則はどのようにでも説明できる．

1. 子どもに学習意欲が起こらないのは，精霊が学習意欲を食べてしまうからである
2. 子どもに学習意欲が起こらないのは，両親が先祖供養を怠っているからである
3. 子どもに学習意欲が起こらないのは，両親が無意識にそれを望んでいるからである
4. 子どもに学習意欲が起こらないのは，知的好奇心が不足しているからである

以上はどれも「子どもに学習意欲が起こらない」ことの説明である．説明の仕方は様々であるが，学習意欲が起こらないという「結果」に対する「原因」は何か，について言及していることは共通している．

心理現象の説明は，どのような事がどのようにその心理現象を引き起こしているか，という因果関係の説明という形をとることが一般的である．しかし，上に述べたような「原因」を指摘するだけの説明では説明とは言えない．何が原因となるかだけでなく，どのようにその現象を引き起こすかについての説得力ある説明も必要である．

2.1.1　因果関係と共変関係（相関関係）

ある心理現象に関する因果関係を明らかにしたいときに注意すべきことは，因果関係と共変関係（相関関係）を混同しない，ということである．

共変関係とは，２つ以上の事柄が伴って変化する，という関係である．伴って変化するからといって必ずしも因果関係が成立するとは限らない．

1. 「雨が降っている／いない」「傘をさしている／いない」
2. 「傘をさしている／いない」「レインシューズを履いている／いない」

1は共変関係である．雨が降ると傘をさしている人が多く，雨が降っていないと傘をさしている人が少ない．雨の有無と傘をさしている人の数とが伴って変化するからである．

１では「雨が降る→傘をさす」という因果関係が成立する．雨が降ることが傘をさすことの原因となっている（傘をさしたら雨が降るわけではない）．

２も共変関係である．レインシューズを履いている人の多くは傘をさしているだろう．つまりレインシューズを履いている人の数と傘をさしている人の数は伴って変化する．

しかしここには因果関係が成立しない．傘をさすからレインシューズを履くのでも，レインシューズを履いたから傘をさすのでもな．雨が降ったから，傘をさし，レインシューズを履くのである．

共変関係にある２つの事柄を観察したからといって，そのことから直ちに因果関係を論じることはできない．その２つの事柄に共通する別の原因がある可能性も考えてみる必要がある．また，２つの事柄が互いに互いの原因となっているような場合もある（循環的なシステムの場合）．

因果関係に言及するには，原因となる事柄を人為的に操作して，それによって結果が変化することを示す必要がある．上の例では,（可能かどうかは別として）雨を降らせ，そのことで傘をさす人が増えるかどうかを観察する．また，傘をさしてみて，雨が降るかどうかを観察する．このような操作を通して初めて因果関係を明らかにすることができる．

心理現象に関わる共変関係を示すことは容易である．しかし因果関係の成立—特にその現象を引き起こすメカニズムにおける因果関係—を示すことは容易ではない．

問：学習意欲と知的好奇心との例について考察してみよう

2.2 説明の妥当性をデータに基づいて主張する（実証研究）

ある説明（理論・仮説）が人々に受け入れられるためには，その正しさを立証するデータ（証拠）を示す必要がある．説明はデータを通し検証され，修正・改良されていく．

問：上の説明例 1〜4 を立証するには，それぞれどのようなデータが必要だろうか．またそのデータはどうやったら収集することができるだろうか

もちろん，明確な仮説がありそれを検証するためだけにデータが収集されるわけではない．これまで知られていなかった現象がデータよって示される場合もある（興味深い心理現象の発見）．

しかし，いずれにせよデータを収集する際には，必ず「こういう状況ではこのような反応が得られるはずである」のように確認したいこと，あるいは「こういう状況ではどのような反応が起こるのか」のように調べたいことが前もってあるはずである．データは無計画に収集されるわけではない．何らかの視点や切り口を必ず持っている．

知りたいこと，確かめたいことが明確になるようなデータを収集することが重要である．そのためにはまず，データを収集する目的を明確にする必要がある（要は「何を研究するのか」をはっきりさせるということ）．次に，その目的を達成するためにはどのようなデータをどういう手続きで収集し，どのように解析すればよいかをよく考えなければならない[1]．統計的データ解析手法は，無計画に収集されたデータに対しては無力である．データ収集計画段階での手落ちは解析手法をいくら駆使しても回復不能である．ゴミデータから宝が見つかることはまずない．

2.3 標本に基づく母集団の特徴や性質に関する統計的推論

2.3.1 主張のいろいろ

知りたいこと，確かめたいことが明確になるようなデータを収集することが重要である，と述べた．では，どのようのデータを収集すれば知りたいことや確かめたいことが明確になるのか．それは知りたいことや確かめたいこと（主張）がどのようなものかによって異なる．

2.3.1.1 ある特定の事実に関する主張

- 「私はあなたより背が高い」[2]
- 心理学実験の結果「実験に参加した男性は実験に参加した女性に比べ，平均的に攻撃性尺度得点が高かった」[3]

[1]データを取ってから「このデータをどうやって解析したらいいですかぁ」などと言うようではダメである
[2]両者の身長差が非常に小さい場合，統計的に推測することが考えられる
[3]攻撃性尺度の妥当性と信頼性が問題となるが，ここでは触れない

2.3.1.2　証拠（反証）が 1 つでもあればよい主張

- 「超能力は存在する」
- 「すべてのカラスは黒い」（白いカラスが 1 羽でもいると「論理的には」成り立たない）

2.3.1.3　集団の特徴や性質に関する主張

- 「男性は女性より身長が高い」：男性より背が高い女性はいくらでもいる．
- 「身長が高ければ体重も重い」：常に成り立つわけじゃないけど，我々はこれを受け入れる
- 「男性は女性に比べ攻撃性が強い」
- 「子どもに学習意欲が起こらないのは，知的好奇心が不足しているからである」

心理学研究における主張（仮説）は通常，このタイプである．実験や調査を行うのは，その対象となった人についてのみ何かを言うためではなく，例えば，「幼児は … 」「小学生は … 」「大学生は … 」「男性は … 」「人は … 」等の一般的な主張を行いたいからである．

一般的な主張の対象となる一般的な集団を母集団という．「幼児は … 」という場合，すべての幼児が母集団となる．すべての幼児を対象に調査や実験を行ったとすると，大手を振って「幼児は … 」という主張が可能となる．これは「事実」に関する主張であるからである．

2.3.2　一部から全体を推し量る

心理学研究では，より一般的な集団の特徴や性質に関心がある．しかし通常，集団のメンバー「すべて」を調査や実験の対象とすることは不可能である．そこで，すべてのメンバーの一部を取り出し，その一部を調査や実験の対象とする．このとき取り出された一部を「標本」という．標本の性質を調べ，より一般的な母集団の性質について何かを主張しようとするのが，我々のやり方である．

スープの味見をするとき，普通の人ならば，よくかき混ぜて小皿に少しだけとる．スープを全部飲む人はいない．スープ全体のほんの一部を調べることによって，スープ全体の味を知るのである．これは，標本（小皿）に基づく母集団（スープ全体）の性質に関する推論の一つであると言える．

スープの味見で重要なのは「よくかき混ぜる」ことである．塩こしょうのあと，かき混ぜもせずに小皿にとって味見をして「塩気が足りない」と塩を足したりすると大変なことになってしまう．よくかき混ぜることによって，どの部分をとっても味が「均一」であることが保証される．言いかえると，小皿のスープの味がスープ全体の味とよく似ることを目指してかき混ぜるのである．よくかき混ぜることによって小皿のスープがスープ全体の代わりとなるのである．この状態を小皿のスープはスープ全体を代表している，という．

標本から母集団を推論する場合，標本が母集団をよく代表していることが重要である．標本が母集団の中の特定の一部に偏っていては，標本に基づいた母集団全体の推論はできない．標本は母集団のよい模型でなければならない．

ところが我々の対象は，例えば「幼児」「小学生」「青年」「男性」「女性」「人」など人の集団であり，スープのようにかき混ぜることはできない．どうするか？母集団のどの部分も等しい確率で標本として抽出されるようなやり方で標本を構成するのである．この手続きを「標本の母集団からの無作為抽出」という．標本が母集団から無作為に取り出されたとき，標本に基づく結果を母集団にまで一般化することが可能になる．統計学はその理論的背景を与えるものである．このように一部から全体を推測するための統計を推測統計という．

2.3.2.1 無作為抽出実験

ここに1万人の集団がある．これを母集団とする．知りたいのはこの1万人の平均身長である．少数の標本から母集団の性質（平均身長）を推論してみる．表2.1は，1万人から無作為に10人，100人，500人を抽出し，その平均身長を測るという作業をそれぞれ10回ずつ繰り返した結果である．500人も抽出すれば，かなりの精度で1万人の平均身長の本当の値（真値）を知ることができることが分かる．

表2.1: 無作為抽出実験

	標本数		
	10人	100人	500人
1回目	161.6	165.0	165.4
2回目	166.9	166.2	165.9
3回目	160.3	164.0	164.9
4回目	168.8	166.6	165.9
5回目	165.9	165.8	165.6
6回目	165.5	166.9	165.5
7回目	166.4	163.7	164.5
8回目	164.5	164.6	165.9
9回目	166.4	166.1	165.1
10回目	160.6	163.5	164.1
標本平均の平均値	164.7	165.2	165.3
平均的なばらつき	2.67	1.23	0.62
1万人の平均身長	165.5		

2.3.2.2 実体的有限母集団と仮想的無限母集団

母集団を想定するとき，例えば，日本における有権者全体，というように母集団の実体が捉えられる場合，実体的有限母集団という．これに対し，特に実験の場合に当てはまるが，「ある薬が効くかどうか」を知りたい場合，実験的に薬を飲んだ人だけではなく，「将来薬を飲む人」あるいは「もし薬を飲んだら」という状況を想定している．このとき薬を飲んだ人全体というものは実体的には存在せず，仮想的な母集団である．また将来に渡って，ということを考えると有限でもない．こうした場合の母集団を仮想的無限母集団という．

どちらの母集団も，無作為標本に基づく統計的推論の対象である．

2.3.2.3 無作為性と不規則性

繰り返しになるが，無作為に標本を抽出するということは，どのメンバーも等しい確率で標本として抽出され得るようなやり方で抽出するということである．通常は乱数を用いる[4]．

無作為ということと，抽出の結果の見た目の規則性とはまったく別の話である．例えば，たくさんの白玉と黒玉が入った大きな袋の中から無作為に1つの玉を抽出する，という操作を5回繰り返したとき，その結果が「黒黒黒黒黒」となることがある．

[4]コンピュータによる擬似乱数を用いればよい

2.3.3 心理学研究における標本構成の実際と研究結果の一般化

少数の標本に基づいて母集団の性質を推論することが統計的に保証されるのは，標本が母集団から無作為抽出された場合に限る．しかし，ほとんど全ての心理学研究では，実験や調査の対象となる被験者（標本）は母集団からの無作為抽出によるものではない．したがって，その研究の結果は一般化できない（少なくとも統計的には）．

統計的には一般化できなくても，通常「常識的な判断」を根拠に研究結果はある程度一般化されるし一般化してもよいだろう．重要なのは，その一般化の根拠が統計ではないことを自覚しておくことである．

統計的に一般化できないからといって，研究が無意味となるわけではない．研究の方法がしっかりしていれば，そのときの対象については，何らかの心理学上の重要な知見が得られるわけである．もちろん，対象が変われば結果も変わるかも知れない．いずれにせよ，個々の研究結果の蓄積が全体の財産となる．

どんな統計的手法を用いても，標本が母集団から無作為に抽出されていなければ，研究結果を（統計理論に基づいて）一般化することはできない．その研究は，あくまでも対象となった標本についての事例研究である．

2.4 データ収集の方法

研究テーマが決まれば，その研究をどのように行うことが最も適切かをよく考える．研究方法の決定である．研究方法には，誰を対象に，どんなデータを，どのような方法で収集するか，といった事柄が含まれる．

データを得る方法にはおよそ次のようなものがある．どれか 1 つを選ぶということではない．研究の目的に応じてうまく組み合わせるとよい．

- 実験法：特定の心理現象の原因と考えられる要因を少数取り上げ，その要因の違いを人為的に作り出す（処理水準，条件と呼ぶ）．条件間での心理現象の違いを検討することで，因果関係を明らかにしようとする．
- 観察法：研究者が（比較的自然な文脈で）観察したものを記録する．
  - 自然観察：種々の現象を文字通り観察する．
  - 実験的観察：実験条件を設定し，観察される現象の条件間での比較を行う．
  - 参加観察：積極的に現象に関与しながら観察する．
- 質問紙法：特に人の「内面」に起こっている主観的な心理現象（感情や意見，態度，動機，期待，願望など）について質問紙を作成し本人に「尋ねる」．
- 面接法：面接における会話を通じ，目的に応じた情報・資料を得る．
- 検査法：すでに確立している心理検査等を用い個人の特性を記述する．

ただし，どの方法でデータを収集するにしても，どうすれば研究の目的が果たせるかをよく考えなければならない．この点に関しては，実験法と質問紙法を中心に後で詳しく述べる．

問：「子どもの学習意欲」の説明例 3，4 を検証するためには，どのようなデータ収集法が考えられるだろうか

2.5 研究の方法を考える視点

研究における皆さんの課題は，「方法を自分で考える」である．当たり前のことのはずなのだが，先行研究に盲従していると思われる研究がかなり多いように思われる．また，そうすべきである，というような「雰囲気」があるようにも感じられる．前例に必ずしも従う必要はない．主体的に，テーマに最も適切な方法を模索することにも力を注ぐことが重要だろう．

佐藤ら（2000）は，心理学研究における従来型研究スタイルとありうる研究スタイルの対立軸（視点）として以下を挙げている．もっと多様な研究スタイルがあるべきである，というのが彼らの主張である．

1. 問題意識の発見：生活実感的／先行知見重視的
2. 介入目的の有無：実践的／純粋学問的
3. 研究の計画性：臨機応変的／実験計画的
4. 論の立て方：仮説生成的／仮説検証的
5. 研究の目的：個性記述的／法則定立的
6. 統計の利用：定性的（質的）／定量的（量的）
7. 追試可能性：一回的／反復的
8. データを取る場：現場（フィールド）／実験室（教室・講義室）
9. 研究対象との関係性：関与的／権威的
10. データの産みだし方：生態的／人工的
11. データ発生の場の見方：文脈重視／要因重視
12. 解釈の対象：テクスト主義的／データ主義的
13. 利用する統計：記述統計／推測統計

問：「子どもの性格は親の育て方によって影響を受けるのか？」を研究したいとする．どのような方法が考えられるか．複数挙げよ．

2.6 研究の手順

1. テーマを決める：興味のあることならなんでも
2. テーマについて調べる：どこまで分かっているか／分かっていないか
3. どうしたらそれを研究できるかを考える
   - 必ずしもデータを取る必要はない（論考を中心とした理論研究もあり得る）
   - どのようなデータをどのように収集するか（必ずしも数量的なものである必要はない）
4. 何が分かったかを要領よく説明し，結果を考察する
   - データの整理，要約，説明
   - データの解析
   - 解析結果の解釈

ただし，お金，時間，人手，機会などの事情から，卒論・修論などでは研究法が限られる場合が多い．研究法に制限が加われば必然的に研究テーマも限られる．

第3章　データの整理・要約：基本統計量

データの収集後，最初にすべきことは，データを整理・要約しデータの持つ全体的な特徴を把握することである．○○解析とか××分析といった複雑な手法を適用しなくても，この段階でおよそのことは分かるものである（分からないなら，それはデータ収集計画がマズイからである）．ここでは，データの持つ全体的な特徴を数値的に記述するための次の基本的な数量（基本統計量）を中心に説明する．

- 平均値（代表値）
- 分散・標準偏差（データ変動の指標）
- 共分散・相関係数（共変関係の指標）

データの持つ全体的な特徴を基本統計量で表現することも大事だが，グラフを用いて視覚的に確認することはもっと大事である．よく用いられるグラフには以下がある．

- ヒストグラム：データの分布
- 箱ひげ図：データの分布および代表値（中央値）の群間比較
- 散布図：２変量の共変関係（量的データ）
- 棒グラフ：平均値等の群間比較
- 折れ線グラフ：数量の継時的変化，プロフィールの表現
- 帯グラフ，円グラフ：カテゴリ構成比の比較
- モザイク図：２変量の共変関係（質的データ）
- レーダーチャート：プロフィールの表現

3.1　尺度の種類

我々が扱うデータは何らかの形で数値化されていることが多い．数値であるからといって，すべて同列に扱うことはできない．その数値がどのようなモノサシ（尺度）によって得られたものであるかによって，処理の仕方は異なる．モノサシの種類は次の４つである．

名義尺度 男性=1，女性=2 のように特定のカテゴリに対し便宜的に数値を割り当てる場合をいう．
数値の大小関係には全く意味がない．足したり引いたりすることに意味がないことは言うまでもない．

順序尺度 質問に対して「よくあてはまる」から「全くあてはまらない」までの５段階で回答させ,「よくあてはまる=5」～「全くあてはまらない=1」のように得点化することがよく行われる．このように割り当てられた数量の順序に意味がある場合を順序尺度という．
ここでの５段階評定の場合，得られた数量の順序には意味を認めることができるが，５点と４点の間の差と２点と１点との差が等しいことは保証されていない．したがって足したり引いたりすることには「本来」意味がない．さらに言えば，いわゆる学力テストの得点も順序尺度である．99 点と 100 点の人の「学力」の差が 55 点と 56 点の人の「学力」の差と等しいとは言えないからである
（しかし通常，質問紙やテストで得られた数値は次の間隔尺度として扱われる）．

間隔尺度　数値間の差が等しい尺度である．例えば，温度．20 ℃と 25 ℃との 5 ℃の差は 10 ℃と 15 ℃との 5 ℃の差に等しい．ただし，原点 0 ℃は温度がないことを意味せず，任意である．

比例尺度との違いは数値間の比に意味が認められないという点にある．「10 ℃は 5 ℃の 2 倍である」は温度が 2 倍あることを意味しない．「10 ℃は-5 ℃の-2 倍である」を考えればよく分かるだろう．

比例尺度　数値間の比に意味がある尺度である．例えば，長さ，重さなどである．

原点 0 はその測定の対象が「ない」ことを意味する．0cm は長さがない．また，「10cm は 5cm の 2 倍である」という主張も意味をもつ．長さが 2 倍なのである．

また，名義尺度，順序尺度によるデータを質的データ，間隔尺度，比例尺度によるデータを量的データと呼ぶ．質的データと量的データとでは，データの整理・要約の仕方やその後の解析の手法が異なるので，自分の扱おうとするデータが，質的なのか量的なのかをしっかりと把握しておく必要がある．

さらに，心理学研究で扱う数量は，厳密に言えば，名義尺度，順序尺度がほとんどである．しかし，上でも述べたが，順序尺度データの多くは間隔尺度データとして扱われる．このような実状をよく認識した上で結果の解釈を行う必要がある．統計解析ソフトが出力する数値に振り回されないように注意しよう．

3.2 量的データ

3.2.1 データの要約：平均値，分散，標準偏差

我々の関心はある「集団」についての性質である．そして集団から得られたデータに基づいて何かを主張しようとする．このとき，その集団の性質や特徴をよく表す数量が必要となる．個々のデータはもちろん重要であるが，例えば 100 名のデータをそのまま見せられてもなんだかよく分からない．何らかの形で，それらを要約し有用な情報を取り出す必要がある．

集団の特徴を表現するものとして，代表値がある．代表値として用いられる数量には，平均値の他，中央値（median），最頻値（mode）などがある．これらは状況によって使い分ける必要がある．ところが，代表値だけでは集団の特徴は十分に記述しきれない．例えば，以下のデータを見てみよう．

表 3.1: 2 つのクラスの同一数学テストの得点

番号	A 組	B 組
1	50	80
2	60	75
3	40	75
4	45	90
5	55	30
6	65	25
7	70	15
8	75	85
9	50	60
10	55	30
平均値	56.5	56.5
標準偏差	10.5	27.0

A 組，B 組とも平均値は 56.5 点である．このことから，両クラスは数学の成績に関して性質が似ていると言えるだろうか．それぞれのデータをよく眺めてみると，決して似ているとは言えないことが分かるだ

ろう．A 組では，ほとんどの生徒がだいたい 60 点前後の得点であるのに対し，B 組では 80 点以上のものが 3 名いる一方で 15 点や 25 点といった者もいる．つまり個々のデータのちらばり具合が両クラスで異なるのである．こうしたことも性質や特徴の 1 つとして記述できれば，有用であろう．集団内のデータの散らばり具合を記述する数量に分散及び標準偏差がある．この 2 つは本質的に同じものであり，標準偏差は分散の正の平方根である．

表 3.1 にはそれぞれのクラスの標準偏差を示してある．これは，A 組では個々のデータは平均的に平均値から 10.5 点離れているのに対し，B 組では平均的に平均値から 27.0 点離れていることを表現している．なお，平均値，分散，標準偏差は以下に従って求める．

n 人から成るある集団から変数 x, y について（例えば身長と体重）データを得たとする．
すなわち，$(x,y) = (x_1,y_1),(x_2,y_2),\cdots,(x_i,y_i),\cdots,(x_n,y_n)$.

平均 $\bar{x}, \bar{y}$：集団の代表値として用いられることが多い．

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \end{aligned}$$

分散 s_x^2, s_y^2：データの散らばり度の指標として用いられる．

$$\begin{aligned} s_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \end{aligned}$$

標準偏差 s_x, s_y：データの散らばり度の指標として用いられる．分散の正の平方根．

$$\begin{aligned} s_x &= \sqrt{s_x^2} \\ s_y &= \sqrt{s_y^2} \end{aligned}$$

3.2.2 データの標準化

表 3.2 は，国語と数学の 10 名分の試験成績である．8 番の学生は国語が 75 点，数学が 85 点である．この点数だけを比較すると数学の方が成績がよかったと言える．平均値と比較しても，国語では 18.5 点，数学では 28.5 点平均値よりも高い．やはり数学の方が成績がよい．ところが順位をみると，国語は 10 名中 1 位，数学は 2 位である．だったら国語の方が成績はよかったと言えるのではないか．

このような困ったことが起こるのは，国語と数学とでは，データの平均値は同じでも散らばり度が異なるからである．国語では個々のデータは平均的に平均値から 10.5 点散らばっているのに対し，数学では 27.0 点である．平均的に 10.5 点散らばる中で平均値より 18.5 点高いことと，平均的に 27.0 点散らばる中で平均値より 28.5 点高いことでは，どちらに価値があるだろうか．

平均値からの距離（偏差）はそのままでは比較できない．集団内の散らばりの程度が異なるからである．そこでこれを比較可能にするために，平均的な平均値からの散らばり＝標準偏差を単位として平均値からどれだけ高いか低いかを求めることにする．すると，国語の場合の 18.5 点は標準偏差の 1.76 倍となり，一方，数学の場合の 28.5 点は標準偏差の 1.06 倍となる．どうやら国語の方がよくできていたと言えそうである．

注：ここで成績の良し悪しの観点は，集団の中での相対的な位置でしかない．国語の能力と数学の能力のように性質の異なるものはデータをどのように変換しても，そもそも比較できない（非常に重要）

表 3.2: 2 つの科目のテスト得点

番号	国語	数学
1	50	80
2	60	75
3	40	75
4	45	90
5	55	30
6	65	25
7	70	15
8	75	85
9	50	60
10	55	30
平均値	56.5	56.5
標準偏差	10.5	27.0

この例のように，2 つの科目間の平均値が同じ場合でも，データの散らばり度が異なれば，なかなか単純には比較できない．そのときには，平均的なデータの散らばり＝標準偏差を単位として，その何倍平均値から離れているかに基づいて比較するのである．平均値が異なれば，全体から平均値を引いて，ともに平均値を 0 とすれば平均値を揃えることができる．この作業をデータの標準化という．少し形式的に言うと次の通りである．

データ $x_i : i = 1, \cdots, n$ について，以下の変換を標準化と言い，標準化された得点（無名数）を標準得点という（z で表される）．

$$z_i = \frac{x_i - \bar{x}}{s}$$

ただし

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdots \text{平均値}$$

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdots \text{標準偏差}$$

言い換えると，標準得点はもとの変量からその平均値を引き，標準偏差で割った数量である．標準得点の平均値は 0，標準偏差（及び分散）は 1 である．

問：「標準得点の平均値は 0，標準偏差（及び分散）は 1 である」ことを確認しよう

※データを標準化すると「平均と標準偏差に関する情報が失われる」ということも知っておこう．

3.2.3 連関：共分散，ピアソンの積率相関係数

集団からデータを得る場合，それは必ずしも 1 種類とは限らない．複数の変量間の関係も重要な情報であり，その集団を特徴づけるものである．例えば，身長が高ければ体重も重い，という関係はおそらく一般的な集団の特徴として指摘することができるだろう．心理学研究においても，種々のデータ間の連関関係を通して，新しい事実が発見されたり，あるいは理論が検証されたりする．

データ間の連関を記述する指標として，共分散，ピアソンの積率相関係数（略して相関係数）がある．特に相関係数は非常によく用いられる．

表 3.3: 3 科目の試験成績間の関係

	得点			平均値からの偏差			偏差の積	
受験者	数学	物理	英語	数学	物理	英語	数-物	数-英
1	53	53	28	-3.2	-13.4	-30.6	43.5	99.5
2	51	65	63	-6.0	-1.4	4.0	8.3	-24.1
3	46	68	67	-10.3	1.6	7.6	-16.4	-78.5
4	65	76	69	8.1	9.6	10.0	78.0	81.5
5	64	75	68	7.1	8.6	9.0	61.4	64.6
6	74	78	64	17.6	11.6	5.2	204.2	91.2
7	39	53	57	-17.6	-13.4	-2.1	236.0	36.6
8	66	62	43	9.7	-4.4	-16.4	-42.8	-159.1
9	56	69	66	-1.1	2.6	7.3	-2.8	-8.0
10	52	65	65	-4.4	-1.4	5.8	6.2	-25.6
平均	56.7	66.4	59.0				575.6	78.3

表 3.2 には数学，物理，英語の試験得点 10 人分が示してある.「平均値からの偏差」には，それぞれの科目の 10 人の平均値から，各受験者の得点がどれだけ離れているかを表したものである．例えば，1 番の受験者の数学は 53 点，数学の平均点は 56.7 点であり，平均値からの偏差は，53-56.7=-3.2 である（-3.3 とならないのは丸め誤差）.

数学と物理について，この平均値からの偏差を「正負に注意して」眺めてみよう．数学がマイナスであれば物理もマイナス，数学がプラスであれば物理もプラスとなっていることが分かるだろう．このことは，数学が平均値よりも高ければ（低ければ），物理も平均以上（以下）であることを意味する．このように，一方が変化すれば他方もそれに伴って変化する量のことを共変量という．

次に偏差の積を見てみよう．積なので，プラスとプラスでプラス，マイナスとマイナスでプラスとなる.「数-物」の列をみると全体的にプラスの値が多く大きい．このことは，数学と物理の得点間に，一方が高ければ他方も高く，一方が低ければ他方も低いという共変関係があることを意味する．なお，偏差の積の列の一番下の数値 575.6 は，偏差の積の総和である．これに対し,「数-英」を見ると，プラスマイナス半分ずつであり，偏差の積の総和も 78.3 と小さい．つまり数学と英語の得点には数学と物理に比べ共変関係は小さいと言える．

共分散 s_{xy}：2 つの変数の共変量.

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

さて，共分散は例えば，数学-物理の 575.6 と数学-英語の 78.3 を比較すると，数学-物理間の共変関係の方が強いことが分かるが，数値そのものからは，それが大きいとか小さいとかは分からない．まったく同じ集団を対象にした場合でも，例えば身長 cm-体重 kg の共分散と身長 m-体重 kg の共分散とでは身長 cm-体重 kg の方が大きくなる．このように，共分散の大きさには単位の違い（これは平均値を標準偏差の違いと言い換えることができる）が影響する．これでは，様々なデータについて共変関係の強さを互いに比較することができない．

単位の違いの影響を受けない共変関係の指標にピアソンの積率相関係数がある（次式）．相関係数は-1～1 の値を取り，絶対値が大きいほど 2 つの変量間の直線的な共変関係は強いといえる．我々の扱う類のデー

タでは $r = 0.6$ もあればかなり強い共変関係があると言える．ただし，共変関係と因果関係とを混同してはならない．

ピアソンの積率相関係数 r_{xy}：2つの変数の共変関係の指標．

$$
\begin{aligned}
r_{xy} &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \\
&= \frac{s_{xy}}{s_x s_y}
\end{aligned}
$$

上式を注意深く見れば分かるが，これは共分散をそれぞれ2つの変数の標準偏差で割ったものである（つまり標準化したもの）．もとのデータを標準化してから共分散を求めれば，それはピアソンの積率相関係数と一致する．

3.2.3.1　相関係数の利用上の注意

相関係数は便利であるが，相関係数だけをみて変数間の関係を論じるのは危険である．散布図などによりデータがどのようになっているかを確認する必要がある．

(1) ハズレ値の影響：相関係数はハズレ値の影響を強く受ける．図 3.1 において，全データの相関係数は 0.408 となる．しかし，右上の 2 点を除けば-0.206 となる．

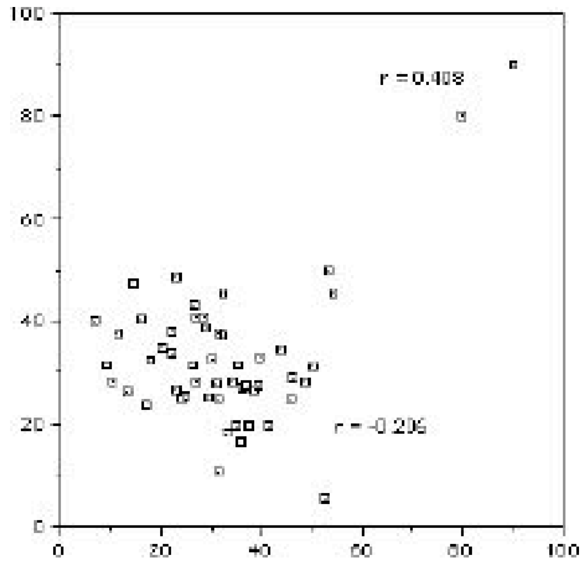

図 3.1: ハズレ値の影響

(2) 切断効果：集団を分割することによって本来ある共変関係が消える場合がある．例えば，入試の成績と入学後の成績との間には本来共変関係があると考えられる．しかし現実には入学者についてしかデータは得られない．入試成績と入学後の成績との相関係数は一般に小さいと言われているが，図 3.2 のようなことが起こっていることが考えられる．

(3) 群の融合効果：異なる性質をもつ2つ以上の群を合わせて相関係数を求めると，本来ない共変関係が表れたり，逆に本来ある共変関係が消えたりので注意を要する（図 3.3）．

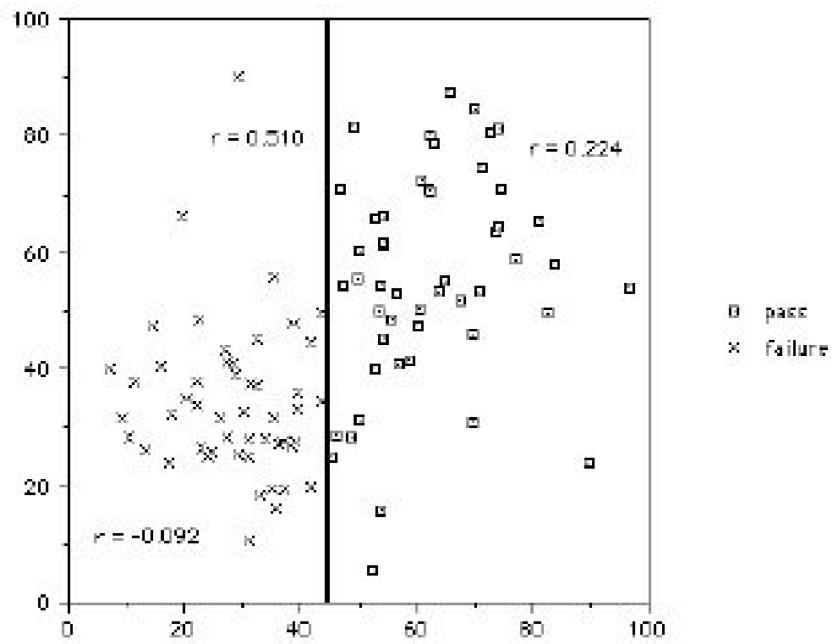

図 3.2: 切断効果

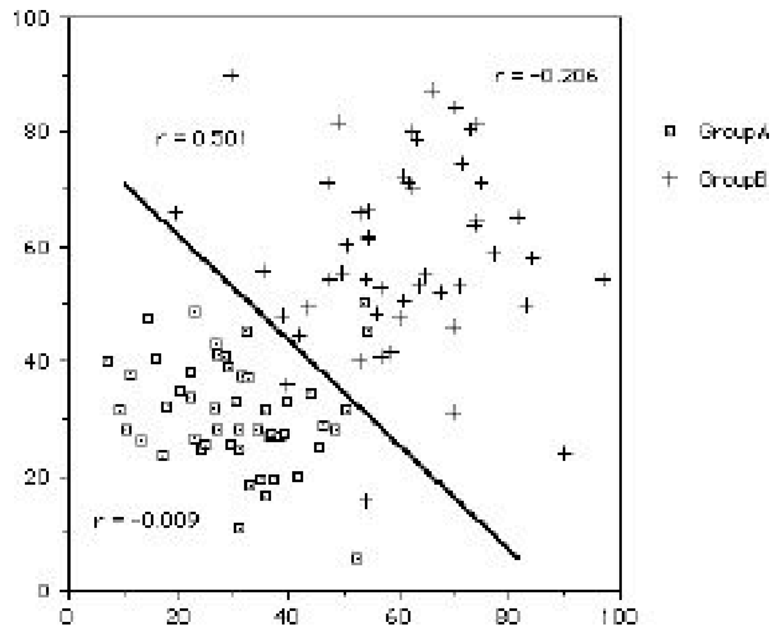

図 3.3: 群の融合効果

3.3 質的データ

3.3.1 データの要約：集計

質的データは足したり引いたりできない．基本は「数える」である．カテゴリへの反応数を数え上げることを「集計」という．集計結果は表（例えば表 3.4）やグラフで示される．

表 3.4: 集計例（教育学部の男女構成）

	男性	女性	計
N	126	154	280
%	45.0	55.0	100.0

3.3.2 連関：分割表，連関係数

質的データの場合でも，変数間の連関は重要な情報となるが，やはり基本は数える＝集計である．2 つ以上の変数間のカテゴリの組み合わせ毎に集計し表示するのが分割表（クロス表）である．例えば表 3.5 は，学部（4 カテゴリ）×性別（2 カテゴリ）のクロス表である．

表 3.5: 学部別男女構成

	N	男%	女%
教育	280	45.0	55.0
文学	90	50.0	50.0
農学	120	65.0	35.0
工学	250	75.0	25.0

ここから，例えば，教育学部は工学部や農学部に比べ女性が多い，などということが分かる．つまり「学部の違いによって男女構成が異なる」のである．このことは，学部と男女構成との間に何らかの関連があることを意味する．

質的データ間の連関の様子はクロス表を眺めれば大体分かるが，量的データにおける相関係数のように，関連の程度の指標となる数量がいくつかあるので以下に紹介する．

- 相関比：名義尺度×間隔尺度（学部×テスト得点など）：名義尺度の水準数が 2 の時は点双列相関係数と一致する

$$
\begin{aligned}
\eta^2 &= \frac{\sum_i^m n_i(\bar{x}_{i.} - \bar{x})^2}{\sum_i^m \sum_j^{n_i}(x_{ij} - \bar{x})^2} \\
m &: \text{名義尺度の水準数} \\
n_i &: \text{名義尺度水準 i におけるデータ数} \\
\bar{x}_{i.} &: \text{水準 i における } n_i \text{ 個のデータの平均} \\
\bar{x} &: \text{全データ平均} \\
x_{ij} &: \text{水準 i における j 番目のデータ}
\end{aligned}
$$

- 点双列相関係数：相関比において名義尺度が 2 水準の場合（男女×テスト得点など）

$$
\begin{aligned}
r &= \frac{\bar{x}_2 - \bar{x}_1}{s_x} \frac{\sqrt{n_1 n_2}}{n_1 + n_2} \\
\bar{x}_1 &: \text{名義尺度水準=1 の x の平均} \\
\bar{x}_2 &: \text{名義尺度水準=2 の x の平均} \\
n_1 &: \text{名義尺度水準=1 のデータ数} \\
n_2 &: \text{名義尺度水準=2 のデータ数} \\
s_x &: \text{x の標本標準偏差}
\end{aligned}
$$

- 四分点相関係数（ファイ係数）：2×2 の分割表におけるピアソンの積率相関係数に一致．２つの変数 x, y がそれぞれ１か０のどちらかの値を取るとする．値の組み合わせは $(x,y)=(1,1),(0,1),(1,0),(0,0)$ の４通りであり，それぞれのデータ数を a,b,c,d とすると，x,y の四分点相関係数は以下で定義される．

$$\phi=\frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

- 双列相関係数：２変数 x,y が二変量正規分布に従うと仮定され，一方の変数（例えば y）が適当に二分され２値変数 $y'(y'=1 \text{ or } 2)$ によって表されている場合に用いられる x と y' 間の相関関係を表す指標．

$$\begin{aligned}
r_{bis} &= \frac{\bar{x}_2-\bar{x}_1}{s_x}\frac{p(1-p)}{k}\\
\bar{x}_1 &: \text{y'=1 となる群の x の平均}\\
\bar{x}_2 &: \text{y'=2 となる群の x の平均}\\
p &: \text{y'=1 となるデータの比率}\\
s_x &: \text{x の標本標準偏差}\\
k &: \text{標準正規分布を p:(1-p) に分割する点における確率密度（縦座標の値）}
\end{aligned}$$

- スピアマンの順位相関係数：順序尺度×順序尺度．対応する n 個の２変数の値 (x_i,y_i) が１から n までの順位で与えられているとき，その順位の差 $d_i=x_i-y_i$ をもとに以下のように定義される．

$$r_s = 1-\frac{6\sum d_i^2}{n(n^2-1)}$$

同順位が含まれる場合は以下の通り

$$r_s = \frac{\{\sum x_i^2+\sum y_i^2-\sum d_i^2\}\times n/2-T^2}{\sqrt{n\sum x_i^2-T^2}\sqrt{n\sum y_i^2-T^2}}$$

ただし，　$T=\frac{1}{2}n(n+1)$

- 連関係数：名義尺度×名義尺度．クラメールの連関係数の定義は次の通り．

$k\times l$ の分割表において

$$\begin{aligned}
V &= \sqrt{\frac{\chi^2}{N[min(k,l)-1]}}\cdots 0\le V\le 1\\
\chi^2 &: \text{分割表から計算されるカイ二乗}\\
\min(k,l) &: \text{行数 k，列数 l の小さい方}\\
N &: \text{観測総数}
\end{aligned}$$

第4章　実験法と質問紙法

ここでは，心理学研究における代表的な（よく目にする）研究方法，実験法と質問紙法について理解しておくべきことをやや詳しく述べる．

4.1　実験法

心理学の実験では，心理現象の成立に関与していると思われる要因を組織的に変化させ，その上で被験者の反応を観測し，変化させた要因の反応への影響（効果）を吟味する（原因を操作することによって結果がどう変化するかを見るのである）．このことによって，その心理現象と取り上げた要因との因果関係を明らかにしようとする．

データの分析には分散分析が用いられる．

4.1.1　要因・処理・処理水準

要因が研究者によって操作可能な場合，その要因は実験要因，あるいは処理と呼ばれる．また処理の取りうる値をその要因の水準という．例えば，

> 「数学の授業内容の理解度が３つの異なる教授法でどのように異なるか」を調べるために，A 組の生徒には教授法 a，B 組の生徒には教授法 b，C 組の生徒には教授法 c で，それぞれ担任に授業をしてもらい，授業終了後に共通の理解度テストを実施しその成績を比較した．

この場合，教授法が実験要因であり，a,b,c の３つの（処理）水準を持つ．

問：理解度テストの成績に影響を与えうる他の要因をできるだけ多く挙げよ

4.1.2　独立変数と従属変数

関数 $y=f(x)$ では，x の値によって y の値が変化する．このような関係にある場合，x を独立変数，y を従属変数と呼ぶ．例えば，年齢の違いによって現在の生活への満足度が異なる，という場合，年齢が独立変数，満足度が従属変数となる．上の教授法の例では，教授法が独立変数，理解度テストの成績が従属変数となる．なお，独立変数が実験者によって操作可能な場合，実験変数という．これに対し従属変数は観測変数という．

4.1.3　剰余変数の影響の統制

上述の実験の結果，A 組の理解度テストの成績が他のクラスに比べて非常に良かったとする．しかし，このことから，教授法 a が数学の授業内容の理解度を最も高めるものであると結論づけるわけにはいかない．A 組はもともとよくできる生徒が多いクラスだったかも知れないからである．

理解度テストの成績（従属変数）には，教授法以外にも，もともとの数学の能力などが影響を及ぼす．このように実験要因以外で従属変数に影響を及ぼす変数を剰余変数（あるいは外的変数）という．つまり，実験に取り上げられなかった要因のことである．上記の実験の場合，もともとの数学の能力は剰余変数であると言える．剰余変数によるデータの変動を系統誤差という．

実験的手法においては，剰余変数がいかにきちんと統制されているかがポイントとなる．剰余変数が統制されていなければ，従属変数（結果）の違いが，実験要因（原因）の違いによるものか，剰余変数（その他の原因）の違いによるものかの区別がつかない（この状態を交絡という）．

従属変数に及ぼす剰余変数の影響を統制し，より高い精度で処理の効果を検証するための一連の技術・手続きが実験計画法である．実験によるデータ収集後でも，統制できていない剰余変数が見つかれば（指摘されれば）その実験は失敗である．

4.1.4 剰余変数の影響の統制の方法

実験的手法において最も大事なことは，従属変数に影響を与える要因を特定できることである．そのためには剰余変数を統制しなければならないことは上に述べた通りである．剰余変数の影響を統制する方法には，次のようなものがある．

- 無作為化：実験要因の各処理水準に被験者を無作為に割り当てる方法．

  実験に参加する生徒をすべて集め，個々の生徒に教授法 a,b,c を無作為に割り当てる．注意しなければならないのは，教授法を無作為に割り当てるというのは，適当にということではなく，生徒が a,b,c それぞれの教授法に割り当てられる確率が等しく 1/3 であるということである．こうした操作を行うことで，理解度テストの成績から，もともとの数学能力による系統誤差を除去し，これを確率（偶然）誤差として扱うことが可能になる．

  検討したい要因が 1 つで，要因の各処理水準に被験者を無作為に割り当てる実験を完全無作為配置（あるいは単純無作為配置，一元配置）という．また，このように水準間で被験者が異なる場合，要因は被験者間変数と呼ばれる．

- ブロック化・層別化：無作為化よりも積極的に剰余変数の影響を統制しようとするもの．

  もともとの数学の能力が等しいとみなせる 3 人の生徒の組を作る（この組をブロックという）．それぞれのブロックの中では被験者に各教授法を無作為に割り当てる．こうして作られた各教授法を受けるグループ（各処理水準）内ではもともとの数学の能力は同じような構成となるはずである．したがって，理解度テストの成績に処理水準間で違いがあったとしても，それはともとの数学の能力の影響によるものではないと言える．

  この場合，もともとの数学の能力という新しい要因を導入したことになる．その処理水準数は作成されたブロック数となる．このように実験の目的ではない要因をブロック要因（層別要因）という．ブロック化を行うと，実験目的の要因（教授法）以外にブロック要因が加わり 2 つの要因となる．このような実験を乱塊法（乱塊配置）という．

  ところで，もともとの数学の能力によってブロック化したとしても，そのメンバーを見ている内にまた別の剰余変数，例えば他の教科の成績も影響するかもしれない．考え出すときりがなく，多くの剰余変数をブロック化で統制することは事実上不可能である．こうした剰余変数の総体であるとも言える「個人差」を統制する方法がある．被験者 1 人を 1 つのブロックとみなし，すべての処理水準に参加するという方法である．これは被験者内計画あるいは反復測定計画と呼ばれる．またこのときの実験変数は被験者内変数と呼ばれる．ただし，例の場合で言うと，1 人の生徒に対し a,b,c 3 つのすべての教授法の元で理解度テストを行うことになるので，この計画は不適切であることは言うまでもない．

問：「１人の生徒に対し a,b,c ３つのすべての教授法の元で理解度テストを行うこと」はなぜダメなのだろうか

被験者内計画では，１人の被験者がすべての処理水準に参加するので被験者数が少なくて済む．しかし一方で，その処理の順序や疲労などが従属変数に影響することが考えられる．これらは新たな剰余変数であり実験変数と交絡しないようにうまく統制する必要がある．

- 恒常化：剰余変数の値を一定に保つ．

  例の場合，３つのグループが異なる教授法で授業を受け，その後理解度テストを行うが，特定のグループが授業を受けた教室が極端に暑かったり騒音が激しかったりすると，そのことが理解度テストに影響を与える剰余変数となる可能性がある．こうした場合，実験変数以外の条件を可能な限り常に一定に保つ（恒常化する）ことで，剰余変数の影響を統制することができる．しかし，そうした統制の元で得られた実験結果は，一方で，特定の条件の元でのものなので一般化の段階において問題が生じる．剰余変数の値がそれ（実験状況）以外の場合に同じ結果が得られる保証はない．

次善の方法ではあるが，剰余変数の影響もあるかも知れないがそれはとても小さい，ということを他の研究で示すことで説得力を増すことも可能ではある．しかしまずはできる限り剰余変数を統制に配慮すべきである．

4.1.4.1 準実験的方法

無作為化の手続きを行う実験的方法は，やや理想的である．心理学の研究では性別など無作為化できない要因に関心がある場合も多く，常に被験者に各水準（性別の場合は，男／女の２水準となる）を無作為に割り当てることができるとは限らない（できないことの方が多い）．例えば，ある学校のあるクラスが全員同じ実験条件の下でデータが収集され，他の実験条件を割り当てられたクラスから得たデータと比較するような場合もそうである．

4.1.4.2 無作為化の手続きは統計的仮説検定の前提条件である

主に実験要因の観測値に対する効果の有無を確認するために行われる統計的仮説検定（典型的には分散分析，後述）は，観測値が独立に同一の分布に従うことが仮定されている．この仮定は実験要因の処理の無作為配置の手続きによって保証される．

無作為配置が行われない準実験的方法では，「観測値が独立に同一の分布に従うこと」が単に「想定されている，みなされている」だけであって，実際の無作為化の手続きによる裏付けがない．したがって，「観測値が独立に同一の分布に従わないんじゃないの？」と言われれば反論のしようがない．つまり得られた観測値の違いが，実験要因によるものか剰余変数によるものかの区別のしようがないのである．無作為配置が行われていれば，剰余変数の影響があったとしてもそれは偶然誤差として扱うことができ，例えば，「その可能性もあるかも知れないが，せいぜい１％未満の確率だよ」と言うことができる．

実際，上述の被験者内計画（反復測定計画）では，ほとんどの場合「観測値が独立に同一の分布に従うこと」は成り立たない．また，その場合には分散分析による仮説検定の拠り所となる検定統計量 F が帰無仮説を棄却しやすい方向に歪むことが知られている．こうしたことに配慮せずに単純に分散分析を適用し「有意水準×％で帰無仮説が棄却された」「××の主効果が有意であった」などと言っても，信頼性に欠ける根拠に基づく主張となり，説得力は乏しいものとなる．

問：なぜ前提が満たされないような状況でも「統計的仮説検定」が行われているのだろうか

4.1.4.3　標本の無作為抽出と実験要因の無作為配置

どちらも統計的な推論を行うための前提条件である．

- 無作為抽出：標本から母集団へ一般化するための前提条件
- 無作為配置：実験要因の効果の有無を確率的に判断するための前提条件

4.1.4.4　交絡と交互作用

交絡：実験変数に伴い変化する剰余変数が存在し，実験変数の影響なのか剰余変数の影響なのかが判別できない状態．

例の場合，各教授法を「それぞれの担任」が行ったが，この場合，授業者の違いが剰余変数となっており，無作為化を行ったとしても，理解度テストの成績の違いが教授法の違いによるものなのか授業者の違いによるものなのかは判別できない（この場合，同一人物が３つの教授法を行うことで対処できる）．

交互作用：

理解度テストの結果に影響を及ぼす要因として教授法（a,b,c の３水準）を考えたが，これ以外にテスト不安という要因も考慮に入れるとする（高，低の２水準とする）．そして，テスト不安が高い者では教授法 c が有効であり，テスト不安が低い者では教授法 a が有効である，というような実験結果が得られたとする．このとき，教授法とテスト不安との間に交互作用がある，という．２つ（以上）の要因があり，一方の要因の水準によって他方の要因の効果が異なるような場合，この２つの要因には交互作用があるという．交絡と混同しないように．

4.1.5　様々な実験計画

- 実験要因が１つ：完全無作為法，乱塊法（ブロック要因あり）
- 実験要因が２つ：２要因の完全無作為化法，分割プロット法（完全無作為法+乱塊法），２要因の乱塊法

実験計画の呼び方はさほど重要ではないが，ある要因が被験者間要因か被験者内要因かという点は把握しておきたい．また，交絡を起こすような剰余変数があるかどうかにも注意を払う必要がある．データを取った後に統制できていない剰余変数に気づいても遅い（データの取り直しとなる）．また欲張っていくつもの要因を１つの実験に盛り込もうとするのはやめた方がよい．理由は次の通り．

- 被験者間要因計画ならば多くの被験者を必要とする（２水準の要因が３つのとき８つの群が必要となる）．
- 被験者内要因計画ならば１人の被験者への負担が大きくなる（２水準の要因が３つのとき８つの水準での実験への参加が要求される．順序効果，疲労効果への対処が必要となる）
- 要因間に交互作用がある場合，その解釈が困難になる
- ラテン方格法によって実験数を減らすことも出来るが，交互作用に関する情報が得られない

4.2 質問紙法

上で述べた実験法は，種々の条件を人為的に設定し，そのことによる行動の違いを観察することで，人間のある種の側面を理解しようとするものである．しかし，感情や意見，態度，動機，期待，願望など，いわゆる「内面」に起こっている主観的な心理現象は，観察して分かるものではない．そこで「尋ねる」のである．これが質問紙法である．

質問紙を使って行われることは，大きく２つに分けられる．１つは多くの人々の意見や態度の実態を知ろうとする「社会調査」，もう１つは目に見えない人のこころの内面を量的に測定しようとする「心理測定」である．我々が質問紙と関わるのは，この「心理測定」の場合であると考えてよい．

4.2.1 相関的方法

質問紙法では，実験法とは異なり，無作為化の操作がない（できない），因果関係について特定の仮説を前提としないことが多い（普通仮説はあるが，実験的手法になじまないという場合が多い）．こうした場合，変数間の因果関係を直接確認することはできず，相関関係の分析が中心となる．この観点から，心理学研究を実験的方法と相関的方法とに分類できる．表 4.1 に種々の相関的方法における分析手法を示す（表中には分散分析もある．誤解しないで欲しいのは，分散分析が因果関係を明らかにするのではなく，適切な実験計画が因果関係を明らかにするのである）．

表 4.1: 相関分析法の分類

		基準変数	
説明変数		量的	カテゴリー
観測変数	量的	回帰分析	判別分析
	カテゴリー	分散分析・数量化１類	数量化２類
潜在変数	量的	因子分析・共分散構造分析	潜在プロフィール分析
	カテゴリー	理想点判別分析	潜在クラス分析

潜在変数：心理学研究の対象は，直接観測できない説明のための概念（構成概念）であることが非常に多い（例えば，知能，性格）．観測はされないが存在するものとして想定される変数が潜在変数である．潜在変数を想定し，観測変数（実際に観測される変数）との関係を記述する分析モデルが心理学では発展している．因子分析，共分散構造分析がその代表である．

相関的方法で因果関係を立証することはできない．相関関係（共変関係）と因果関係とが同じではないことを十分に認識して欲しい．しかし，相関的方法で，因果関係に関する仮説（因果モデル）を生成することや，もっともらしい仮説をデータを根拠に示唆することはできる．ただし，やみくもに因果モデルを探索することは勧められない．変数の数が増えれば，可能な因果モデルの数は天文学的数になるからである．

真のモデル（正解）に近い意味のあるモデル（説明）を見いだすには，理論的考察によって，あらかじめ因果モデルの数をごく少数に絞っておき，その上でどのモデルがデータから最も支持されるかを検討するのがよい．

4.2.2 質問紙は何かを測るためのものである

人の内面を測るためにはそのためのモノサシが必要である．これを心理尺度という．質問紙を用いて，いくつかの項目について回答してもらい，そのことで，例えば，攻撃性の高さや自尊感情の高さや内的動機付

けや自我同一性の達成の程度などを測ったとするのである．そして，こころの内側の種々の特性間の関係や特定の行動との関連を調べたりするのである．

質問紙法は実験に比べ手軽である．100 名の実験データを得るには相当の労力と日数を要するが，質問紙を使えば100 名程度のデータは非常に簡単に手に入る．このことを研究そのものの手軽さと混同してはならない．人の内面を測ることはそうたやすいものではなく，質問紙は慎重に丁寧に作成しなければならない．

4.2.3　直接測定と間接測定と構成概念

長さ，重さ，時間などはそれぞれの測定器具で直接測定できる．これに対し，教育・心理の領域で測定の対象となるのは，感情や意見，態度，動機，期待，願望，意欲，関心，判断力，思考力，自主性，主体性... などモノサシをあてて直接測ることができないものである．これらを構成概念（仮説構成体，理論概念）と言う．構成概念は思考や理論の簡便のための仮説なので，実体と混同しないことが重要である．

我々が質問紙や種々の心理検査（学力テストを含む）などを通して得ることができるのは，想定される○○性や××度の一つの「指標」としての数値である．注意しなければならないのは，指標はあくまでも指標であり,「測りたいもの」そのものではないということである．測りたいものと指標とを同一視してはならない．自分が手にした指標が「測りたいもの」の指標となり得ているのかを常に問い，吟味する姿勢が重要である．

4.2.4　尺度構成法

繰り返しになるが，心理学では，本来の研究上の関心の対象となる「何か」を直接測定できない場合が多く，集められたデータから，関心の対象となる「何か」のモノサシを作り（＝尺度構成），この尺度を使って「何か」を測定した結果を，理論やモデルの構築・検証に用いる．研究の信頼性は尺度の妥当性，信頼性にかかっているといっても過言ではない．

4.2.4.1　測定のモデル

測定値は本当の値（真値）と誤差の２つの成分からなるとする．

$$\text{測定値}\, X = \text{真値}\, T + \text{誤差}\, E$$

ところで，誤差には

- 系統誤差：測定器具や測定者の偏りに起因するもの．修正するしかない
- 偶然誤差：確率的に生じる誤差．測定回数を増やし，平均することで真値に近づけることができる

があるが，上記モデルでの誤差は偶然誤差（確率的に起こる誤差）である．

4.2.5　測定の信頼性と項目分析

信頼性のある測度（心理尺度，あるいはテスト）とは，測定の度に測定値がでたらめに変動することがなく，異なる２つの測定間で結果が一致するようなものである．そのためには,（偶然）誤差成分が小さくなければならない．cm 単位の測定値を得たとしても，m 単位の（偶然）誤差成分があるなら，この測定値は一貫性・再現性を持ち得ず，何の意味も持たなくなる．

また，測りたいものだけが測られている，ということも重要である．つまり，内部一貫性が保たれているということである．身長を測るとき，身長計が体重によって歪むような材質のものであれば，得られる測定値には，体重の影響も加わってしまう．このときこの身長計は身長を測ってるとは言えない．

心理尺度には，一つの単一な構成概念を測定するものであること，その概念を最小の誤差で測定することが求められることに留意しなければならない．

一般に，心理尺度による測定は精度が悪い．20cm 単位もしかすると m 単位で身長を測るようなものである．だからこそ，よりよい尺度を作成しなければならない．何を測っているのかがあいまいであれば，いくらデータを集めても得られる情報はぼんやりしたものとなる．

尺度の信頼性とは，

- 測定値の分散に占める真値の分散：尺度の信頼性係数（$\rho_X = \sigma_T^2/\sigma_X^2$）
- 測定値と真値との相関係数：尺度の信頼性指数（$\rho_{XT} = \sqrt{\rho_X}$）

である．いずれも，真値を知る必要がある．しかし，それはできないので（真値を知る方法があれば初めからそれを採用すればよい），信頼性係数は「推定」するしかない．

4.2.6 信頼性係数の推定

4.2.6.1 再テスト法

あるテストを一定の期間あけて２度行い，それらのテスト得点間の相関係数を推定値とする．テスト得点の安定性に関する誤差の評価に適している．しかし，現実には問題が多い（測定したいものが時間を経て変化する可能性がある場合，被験者が内容を覚えており意識的に同じ（異なる）回答を行う場合）．

4.2.6.2 平行テスト法

同質と見なせるテスト（平行テストといい，$T_1 = T = 2, \sigma_{E1} = \sigma_{E2}$ が想定される）を行い，それらのテスト得点間の相関係数を推定値とする．テストの特殊性による誤差の評価に適している．

4.2.6.3 クロンバックの α 係数

n 個の項目からなるテストの合計点を X，項目 j の得点を X_j とするとき，クロンバックの α 係数は次のように定義される．

$$\alpha \equiv \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum_j \sigma^2(X_j)}{\sigma^2(X)}\right)$$

α 係数は，すべての可能な折半法（下記）による信頼係数の平均に相当し，信頼性係数の推定値の下限を与える．

テストを構成する項目の等質性に関する誤差の評価に適している．

4.2.6.4 その他

他に次のようなものもある（興味のある人は調べてみて下さい）．

- スピアマン・ブラウンの公式
- 折半法
  複数の項目からなるテストを考える．項目を２つのグループに分割し，それらの相関係数を求め，それを r' とすると，信頼性係数 r は次のように求められる．

$$r = \frac{2r'}{1+r'}$$

- キューダー・リチャードソンの公式

4.2.7 項目分析：信頼性を高めるために

信頼性とは結局，あるテストなり尺度の内部一貫性の高さ（測りたい 1 つのものを測っているかどうか，項目の等質性）と同じであると言える．よりよい心理尺度の作成にあたっては，こうした信頼性指標を用い，尺度を構成する項目の吟味を十分に行うことが重要である．なお，下に挙げた信頼性指標を用いることも大事だが，何よりも重要なのは生のデータをよく見るということである．

4.2.7.1 項目ー総得点間相関

テストの総得点と個々の項目との相関を計算し，それらの相関係数の平均値を検討することでテストの内部一貫性を確かめることができる．また，総得点との相関が低い項目は，他の項目とは異なる何かを測っている可能性が高いので，そうした項目は除外されるべきである．

4.2.7.2 項目間相関

個々の項目間の得点の相関を検討することで，内部一貫性を確かめることができる．面倒くさいかもしれないが，まずはここから始めることを強く勧める．

4.2.8 構成概念妥当性

妥当性とは，測定しようとする属性・特性・能力などを測っているかどうかに関する概念である．たとえ信頼性の高い（安定した，誤差成分の少ない）測定値が得られたとしても，それが目的とする測りたい「何か」でなければ，妥当性が高いとは言えない．また，測定の結果は（心理学的）理論から導かれるさまざまな心理学的事実と整合的でなければならない．心理測定においては，信頼性の高さももちろんだが，妥当性が高いことが非常に重要である．

なお構成概念妥当性はその確認の方法によって次のように分類される（いくつもの妥当性があるわけではない

- 内容的妥当性：測りたい内容項目がむらなく抽出された項目で尺度が構成されているかどうか．尺度を構成する項目内容の検討を通して確認．
- 基準関連妥当性：他の外的基準との関連の検討を通して確認．
  - 併存的妥当性：外的基準による測定値がと尺度得点とほぼ同時期に得られる場合
  - 予測的妥当性：外的基準による測定値が尺度得点よりも後に得られる場合（例えば，模試と入試の関係）

4.2.9 付録：信頼性指数，信頼性係数の導出

真値 T，観測値 X とする．各受験者ごとの各回の測定値は X の T への回帰直線 $\hat{X} = T$ の周りにばらついていると考える．このとき X と T の差が誤差 E であり，これはこの回帰直線の周りにランダムに分布しているとする．このとき X の T への回帰係数 $b(X,T) = 1$ である．ところで，

$$b(X,T) = \rho(X,T)\frac{\sigma(X)}{\sigma(T)} = 1$$

より，

$$\rho(X,T)=\frac{\sigma(T)}{\sigma(X)}$$

この測定値と真値との相関係数をテストの信頼性指数という．

いま，T による X の回帰推定値を $\hat{X}$ とする．すなわち，$\hat{X}=b(X,T)T$．このとき，

$$\begin{aligned}
\sigma^2(\hat{X}) &= b(X,T)^2\sigma^2(T) \\
&= \rho^2(X,T)\frac{\sigma^2(X)}{\sigma^2(T)}\sigma^2(T) \\
&= \rho^2(X,T)\sigma^2(X) \\
&= \frac{\sigma^2(T)}{\sigma^2(X)}\sigma^2(X) \\
&= \sigma^2(T)
\end{aligned}$$

これより，

$$\rho^2(X,T)=\frac{\sigma^2(T)}{\sigma^2(X)}\equiv\rho(X)$$

が得られる．これをこのテストの信頼性係数といい，これは T による X の予測の際の決定係数である．

第5章　確率・確率変数・確率分布

5.1　確率

きっと〜だろう，おそらく〜だろう，など不確さを伴う推量は日常的に行われる．確率は，こうした不確かさを表現し，不確かな世界を理解するための一つの用具である．確率を用いた不確かさの数値的な表現は，日常生活ではあまりないが，例えば，降水確率を用いた天気予報もあるし，コインを投げあげて表が出る確率は大抵の場合，1/2 と評価される．

[確率実験（実験）]
現実世界において自然に起こりうる結果，あるいは人為的な状況における結果が，あらかじめ限定されており（可能な起こりうる結果が限定されている），そのうちの 1 つが実際に起こるという過程を言う．

[事象 (event)]
（確率）実験で起こりうる結果の任意の集合をさす．サイコロを振るという 1 つの実験では，結果は 1 から 6 までのサイコロの目が出ることに限定される．このとき，1 から 6 までのサイコロの目がでることのそれぞれが事象である．また，個々の結果の集合（例えば出た目が偶数である）も事象である．特に，これ以上分割できない事象を要素事象という．

5.1.1　等確率性による確率の定義

実験の結果が，同等に起こりやすい n 個の要素事象に分けられるとき，r 個の要素事象から成る事象が起こる確率は

$$\frac{r}{n}$$

である．例えばサイコロを振る実験の結果 1 の目が出る確率は，6 個の要素事象のうち，1 つの要素事象が起こるということなので，1/6 となる．また，出た目が偶数であるという確率は，6 個の要素事象のうち，3 個の要素事象（2，4，6）が起こるということなので $3/6 = 1/2$ となる．

5.1.2　頻度による確率の定義

n 回の実験のうち，ある事象が起こった回数が r 回であったとする．この比 r/n は，たまたま起こった数値であり，これをこの事象の確率とは言えない．しかし n の数を非常に大きくとると，r/n は安定した値を示すだろう．この安定した値を確率と呼ぶことはできそうである．

事象の確率を π として，

$$\pi = \frac{r}{n}, (n \to \infty)$$

とする．n の値を無限大に取ることは現実的には不可能であるが，このような実験を仮想することによって確率を定義する．

5.1.3 主観確率

上述の２つの確率の定義は，等確率の事象が存在するとは言えない状況での確率評価，あるいは無限回の実験を想定し得ないような１回的な状況での確率評価には適用できない．より広い適用性をもつ確率に主観確率がある．標準確率実験という理論上の実験を通して定義される．

5.2 確率変数

確率実験において，事象に数値を与えることにする．すなわち，事象にある１つの実数 x を対応させる．ある事象が起こるかどうかは，観測するまで確定できない不確実性を持つので，当然のこととして数値 x も同じ不確実性を持つ．ところで，確率実験においては，事象の生起に対し確率が定義されている（確率と事象との対応を決めるものを確率関数と呼ぶ）．従って，事象に対応した数値 x についてもその生起確率は定義される．

このような x の全体を X とおき，確率変数（**random variable**）と呼ぶ．また，特定の値 x が生起することを，$X=x$ と表す．また，実験の結果生じる X のすべての可能な値によって構成される集合を標本空間と呼ぶ．また，標本空間上に定義される確率を確率分布と呼ぶ．

- サイコロを振る（実験）→どの目がでるか，事象：１〜６の目のどれかが出る
- コイントス（実験）→コインのどちら側がでるか，事象：表か裏のどちらか
- テストを受ける（実験）→何点か，事象：0〜100 点
- 身長を測る（実験）→何 cm か，事象：0〜無限大 cm

上に挙げた例それぞれにおいて，例えばサイコロの場合，事象は「１の目が出る」「２の目 … 」である．そしてこれらの事象にそれぞれ $1/6$ の確率が与えられる．この確率を与えるという操作そのものが「確率関数」である．特に数式をイメージする必要はない．

サイコロ実験の各事象にそれぞれ 1〜6 の数値を対応させる．標本空間は $x=1,2,3,4,5,6$ である．x がどんな値をとるかは実験の前には分からないが，$x=1$ となる確率 $P(X=1)$ は，対応する事象に与えられた確率に等しく $1/6$ である．

次に，コイントスを３回繰り返すという実験を考える．起こりうる事象は,「表表表」「表表裏」「表裏表」「表裏裏」「裏表表」「裏表裏」「裏裏表」「裏裏裏」の８つである．これらの事象にそれぞれ $1/8$ の確率を与える．ここで，実験の結果「表」が出た回数に数値を与えるということにする．すなわち，確率変数 X は，3，2，1，0 のいずれかの値をとる．このとき標本空間は $(x=3,2,1,0)$ である．例えば，$X=2$ となる確率は，表が２回出るということなので，それには「表表裏」「表裏表」「裏表表」の事象が対応し，$P(X=2)=3/8$ となる．このように，何を X（確率変数）とするかは自由である．

確率変数には大きく分けて次の２つがある．

- 離散型：サイコロの目のように 1〜6 などの数値を取るがその中間はない．
- 連続型：身長などのような連続値

5.3 確率分布とデータ発生モデル

ある変数の実現値が特定の確率に従って起こることを想定することは，ある事象のデータ発生モデルを考えるということである．確率に基づくデータ発生モデルを確率モデルという．確率モデルの中で，確率変

数に与えられた確率の集合を確率分布という．確率変数が離散型の場合には，確率変数の実現値に対応して確率が定義され，それらの関係は確率密度関数で定義される．確率変数が連続型の場合には，確率が確率変数の区間に対して定義されるため，確率変数の実現値に対しては確率密度が定義される．

5.3.1 離散型確率分布の例

確率変数 X の可能な値が，例えばサイコロの目の場合のように不連続である場合，X の可能な各値ごとに確率が与えられる．

- サイコロを振って出る目を x（$x = 1,2,3,4,5,6$）とする．このとき，例えば $X = 1$ となる確率 $P(X=1)$ は次式で与えられる．
$$p(x) = \frac{1}{6},$$
- コイントスで上になる方を x（$x=1,2$）とする．このとき，
$$p(x) = \frac{1}{2},$$
- コイントスを 10 回行ったとき表がでる回数を確率変数 X（$x=0,1,\cdots,10$）とする．このとき，
$$p(x) = \frac{10!}{x!(10-x)!}(1/2)^x(1/2)^{10-x}\text{：2 項分布}$$
となり,「x は 2 項分布に従う」と言う（〜分布に従う，をよく理解することが重要である）．

これらのように確率変数 X がある特定の値 x をとる確率 $P(X=x)$ が関数 p(x) によって与えられるとき，$p(x)$ を確率密度関数という（確率関数ということもある）．

5.3.2 連続型確率分布の例

確率変数 X の可能な値が連続である場合，X の値のある区間に対して確率は定義される．すなわち，
$$P(a \le X \le b) = \int_a^b f(x)dx.$$
このとき $f(x)$ も確率密度関数という．注意すべきは，離散型の場合と異なり，$f(x)$ は $P(X=x)$ を表すものではない．X が連続であるとき，常に $P(X=x)=0$ である．なお，以降は $p(x), f(x)$ の区別をせず両者とも $p(x)$ と表すことにする．

[正規分布]
X は $-\infty$ から ∞ の値を取りうる．正規分布の確率密度関数は次の通りである．
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$
ここで,

π ： 円周率（定数）

μ ： 母数：（母）平均

σ^2 ： 母数：（母）分散

知能，学力，身長など多くの現象のモデルとして用いられる．

[母数（パラメータ）]
確率分布を特徴づけるものであり，通常は未知である．これを観測されたデータから推測するというのが統計的推論の主な仕事である．

5.4　確率分布と統計的推論・推測

日本人の平均身長を知りたいとする．しかし日本人全員の身長を測定することは現実的に困難である．そこで日本人全体を母集団とし，その無作為標本（n=1000 とする）を調べることによって母集団の平均値を推測することにする．しかし，これだけでは，道具が足りない．1000 人の身長を調べることで日本人全体の平均身長にどれだけ迫れるのか．ここで次のような手続きが必要となる．

- 日本人全体から無作為に 1 人選び出したときのその身長を確率変数 X と考える
- 確率を与える確率分布に平均 μ，分散 σ^2 の正規分布を仮定する

つまり，現象（データを得る過程）を確率実験としてとらえ，何らかの確率分布を仮定する必要があるのである．こうした仮定をおいた後は，それぞれの確率分布の性質によって，演繹的に推論を進めることができる．例えば，

n 個の標本 $x_1, x_2, \cdots, x_n$ がそれぞれ独立に平均 μ，分散 σ^2 の正規分布に従うとき，μ 及び σ^2 の（不偏）推定値，$\hat{\mu}, \hat{\sigma^2}$ は以下で与えられる．

$$\hat{\mu} = \frac{1}{n}\sum_{i}^{n} x_i = \bar{x}\text{：標本平均}$$

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2\text{：不偏分散}$$

5.4.1　母数の点推定

上を利用して，1000 個の標本から日本人全体の平均身長 μ やその分散 σ^2 を推定することができる．このように母数についてその値を一点で推定しようする推定を点推定という．

5.4.2　母数の区間推定

上の場合，標本平均 $\bar{x}$ もまた確率変数であり，平均 μ，分散 σ^2/n の正規分布に従う，ということが分かっている．これを利用すると，例えば $\hat{\mu} = \bar{x} = 165cm$，$\hat{\sigma^2} = 225$（すなわち母集団における標準偏差の推定値が 15cm）のとき，「日本人全体の平均身長 μ の 95 %信頼区間は 164.07cm〜165.93cm である」というような幅をもたせた推定も行うことができる．

なお，95 %の信頼区間とは，同じ母集団から標本抽出を何回も繰り返しその度に信頼区間を求めたとき，そのうちの 95 %は母平均 mu を含むということである．95 %の確率でこの区間に母平均 mu が存在するということではない．いずれにせよ，この区間が狭いほど，点推定値の信頼性は高いと言えよう．
心理学研究で多用される統計的仮説検定も，統計的推論の一つであり，同様の仕組みで行われている．現象に確率モデルを仮定し，そこから演繹的に導き出される種々の統計量の確率分布に基づき，平均値の差の検定や，無相関の検定，独立性や適合度の検定などが行われる．

第6章　統計的仮説検定の基礎

6.1　導入：超能力は鍛えられたのか

TV 番組の話である．1人の男A氏が部屋に軟禁され，超能力を鍛えさせられている．食事をとるためには，並べられた5枚のカードを順に裏返し，1枚だけ用意されているアタリのカードを最後に残さなければならない．つまり食事をとるためには，4回連続してハズレのカードを選ばなければならないのである．超能力のうちの1つ，透視能力を鍛えるための訓練だということである．

視聴者にとってはドキドキする演出である．カードが1枚めくられる度に，それがハズレであれば「おーっ，やったー」となり，次のカードがめくられる．うまくいけばそれが4回も続くわけで，最後にアタリカードが残ることがとてもすごいことのように思えてくるのが不思議だ．

さて，まったくの偶然で食事にありつける確率を計算してみることにする．

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

「さいころを振って1の目が出たら食事を取れる」とした方が難しいことが分かる．TV の演出はすごい．

表 6.1: A氏の1週間の戦績：21戦8勝13敗

月	朝	○	金	朝	○
	昼	×		昼	×
	夜	×		夜	×
火	朝	×	土	朝	○
	昼	○		昼	×
	夜	×		夜	○
水	朝	×	日	朝	○
	昼	○		昼	×
	夜	×		夜	×
木	朝	×			
	昼	×			
	夜	○			

さて，上の表はA氏の初めの1週間の戦績である．全部で21回挑戦し，8回食事を取ることに成功している．彼に透視能力は養われているのだろうか．

6.1.1　偶然におこる可能性を考える：帰無仮説と対立仮説

A氏が透視能力をもつかどうかはどのように判断すればよいだろうか．説得力のある根拠が欲しいところである．そこで「偶然におこる可能性（確率）」を根拠として利用することにする．A氏のもたらした結果がまったくの偶然によって起こる確率を計算し，それに基づいて，次の2つの仮説のうちのどちらを選択するかを決定するのである．なお，仮説0は「帰無仮説」，仮説1は「対立仮説」と呼ばれる．

仮説 **0** ：A 氏は単に偶然に食事にありつけただけである

仮説 **1** ：偶然を超える何らかの「力」が作用している

さて上では「A 氏のもたらした結果がまったくの偶然によって起こる確率を計算し,..」と述べたが，この「まったくの偶然である確率」は帰無仮説に基づいてしか計算できない，という点が非常に重要である．少し考えれば分かることだが，透視能力があれば 1/2 の確率で食事にありつける，などとは言えない．あるかないかが分からない透視能力を根拠に確率を計算することなどはできないのである．これに対し，全くの偶然であるとする帰無仮説の元では，1 回の挑戦あたりの成功確率（食事にありつける確率）を 1/5 とすることができるし，21 回の挑戦で 8 回成功する確率も容易に計算できる．

このことをふまえると，上の「2 つの仮説のうちのどちらを選択するかを決定する」という表現もおかしい．仮説 1 が正しい確率はそもそも計算できないのだから，確率に基づいて仮説 1 を選択することもできない．実は確率に基づいて検討されるのは仮説 0，すなわち帰無仮説だけなのである．正確にいうとここでの意思決定は，

「帰無仮説を棄却するか，棄却しないか」

なのである．これが我々がお世話になる「統計的仮説検定」である．仮説検定とは帰無仮説の正しさのテスト（検定）であり検定されるのは帰無仮説である，ということをことを理解しておくことは非常に大切である．

[注]
上述の説明では，あたかも「仮説 0（H_0）=帰無仮説」のような表現をしているが，これは正確ではない．より正確に言うと，H_0 は「検定仮説」と呼ばれ，ある命題の真偽を検証するために設けられた統計的に表現される仮説である（統計的仮説）．統計的仮説検定とは，まさに検定仮説の検定である．例えば上の例では「1 回当たりの成功確率は 1/2 である」を H_0 とし，これを検定することもできるのである．この場合 H_0 を帰無仮説とは言わない．また対立仮説も様々に設定できる．心理学研究において我々が検定する仮説 H_0 はほとんどの場合，帰無仮説であると考えてよい．

6.1.2 決定のルール：Type1 エラーの危険率の設定

さて上の 2 つの仮説であるが，A 氏の残した戦績が，偶然ではほとんどあり得ないようなものであるならば，仮説 0 をあきらめてもよいだろう（これは仮説 1 を採択することを意味する）．

では，偶然では考えられないような事態をどのように設定すればよいか．例えば，同じことを 10 人同時に行ったとき，そのうちの 1 人に偶然起きるような事柄が起これば，それを「偶然では考えられないような事態」とするのはどうだろうか．10 人に 1 人程度であれば，40 人学級ならばクラスに 4 人程度は透視能力をもつことになってしまうので，この提案は却下されるだろう．では 1 万人に 1 人ではどうだろう．あるいは 100 万人に 1 人ぐらいを考えなければ「超」能力とは言えないかもしれない．

ここで注意しなければならないのは，帰無仮説を棄却するか否かは，あくまでも選択者本人による「意思決定」であるということである．確率は意思決定を行うにあたっての一つの指標として用いるだけであって，確率の値によって，自動的に帰無仮説が棄却されるわけではない．

このことをふまえ，とりあえずここでは偶然であれば 100 万人に 1 人しか成功しないようなことが起こっていたら，これはもう偶然に起こったできごとではない，と認めることにする．しかしこの場合でさえ，単なる偶然である可能性は 100 万分の 1 はある．つまり，「本当は偶然に起こったことなのに，うっかり間違えて仮説 **0** を捨ててしまう危険性が **100** 万分の **1** ある」ということである．

このように本当は帰無仮説が正しいのにも関わらず，間違って棄却してしまう検定の誤りのことを **Type1** のエラー（第一種の誤り）という．Type1 のエラーが生じる確率の上限値を危険率あるいは有意水準など

と呼び，通常 α で表す．例えば，「有意水準５％で検定を行った」は，Type1 のエラーが生じる可能性が５％（$\alpha = 0.05$）以下であれば，思い切って帰無仮説を棄却するのだ，と宣言していることに等しい．

以上を用いて表現し直すとここで行うのは，危険率 $\alpha = 0.000001$ での帰無仮説の検定である．

6.1.3 帰無仮説を正しいと仮定して確率を計算する：統計ソフトが行うこと

ある出来事が起こる確率を計算するためには，出来事に対して確率を割り当てることができなければならない．いま関心ある出来事は「A 氏の 21 回に及ぶ透視能力トレーニングにおいて 8 回うまく食事にありつけた」である．

全部で 21 回挑戦したのだから，起こりうる結果は 21 戦 0 勝から 21 戦 21 勝までの 22 通りである．仮説 0 が正しいとすると，1 回の挑戦あたりの成功確率は 1/5 とすることができる．このとき，21 回の挑戦の中で成功する回数を X とすると（21 戦 X 勝），X は成功確率 1/5，試行回数 21 回の二項分布（次式）に従うことになる．

$$p(x) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$$

ただし　π1 回の試行当たりの成功確率：このケースでは $\pi = 1/5$

n 試行回数：このケースでは $n = 21$

すべての可能な x についての確率及び累積確率を表にまとめると次のようになる．

表 6.2: 成功確率 1/5，試行回数 21 回の二項分布

成功回数 X	確率	累積確率	成功回数 X	確率	累積確率
0	0.009	0.00922337	11	0.001	0.99980593
1	0.048	0.05764608	12	0.000	0.99996752
2	0.121	0.17870283	13	0.000	0.99999549
3	0.192	0.37037603	14	0.000	0.99999949
4	0.216	0.58600838	15	0.000	0.99999995
5	0.183	0.76929588	16	0.000	1.00000000
6	0.122	0.89148755	17	0.000	1.00000000
7	0.065	0.95694737	18	0.000	1.00000000
8	0.029	0.98558604	19	0.000	1.00000000
9	0.010	0.99592778	20	0.000	1.00000000
10	0.003	0.99903030	21	0.000	1.00000000

6.1.4 決定

我々が設定した危険率は $\alpha = 0.000001$ であった．これは上に示した二項分布の表から判断すると，成功回数が 14 回以上の出来事が起こる確率に相当する．つまり A 氏が食事にありつけた回数が 14 回以上であれば帰無仮説は棄却される．帰無仮説が棄却されるような x の（一般にはデータから計算される何らかの統計量）値の領域を棄却域という．

しかし残念ながら A 氏が食事にありつけた回数は 8 回であり，これは棄却域に達しないものである．したがって，帰無仮説は棄却しない，つまり A 氏に起こった出来事は十分に偶然の範囲で起こりうることであると判断することにする．

6.1.5 本当は透視能力があるんじゃないか：Type2 のエラー，検定力

もしかするとA氏は過酷なトレーニングの中で本当は透視能力を養っていたのかもしれない．具体的に言えば，全くの偶然ならば，1回当たりの成功確率は 1/5 である．しかし，透視能力が養われたことによって，A氏の場合，1回当たりの成功確率が 1/3 となっていた可能性もあるわけである．

我々は誤って帰無仮説を捨ててしまう可能性を，危険率 $\alpha = 0.000001$ という具合に厳しく設定した．したがって本当は偶然なのに超能力があると判断してしまうような誤りを犯してしまう可能性は非常に小さい．ところが，もしかすると，我々は帰無仮説が実は間違っているにも関わらず，これを採択してしまった，つまり対立仮説を正しく採用しなかった，という誤りを犯しているかもしれないのである．この誤りを **Type2** のエラー（第二種の誤り）という．なお，Type2 のエラーを犯す確率は β で表される．

下の表は，A氏の1回当たりの成功確率が，実は 1/3 に上昇していた場合の二項分布である．

表 6.3: 成功確率 1/3，試行回数 21 回の二項分布

成功回数 X	確率	累積確率	成功回数 X	確率	累積確率
0	0.0002	0.0002	11	0.0345	0.9788
1	0.0021	0.0023	12	0.0144	0.9932
2	0.0105	0.0128	13	0.0050	0.9982
3	0.0333	0.0462	14	0.0014	0.9996
4	0.0750	0.1212	15	0.0003	0.9999
5	0.1275	0.2486	16	0.0001	1.0000
6	0.1700	0.4186	17	0.0000	1.0000
7	0.1821	0.6008	18	0.0000	1.0000
8	0.1594	0.7601	19	0.0000	1.0000
9	0.1151	0.8752	20	0.0000	1.0000
10	0.0691	0.9443	21	0.0000	1.0000

我々が設定した危険率から導かれた棄却域は $x \geq 14$ であった．したがって，$x \leq 13$ であれば帰無仮説は採択される．このことによって，帰無仮説が正しいにも関わらずこれを誤って棄却してしまう確率は 0.000001 と非常に小さいものとなった．しかし，もし本当は，A氏には透視能力が育っており，1回当たりの成功確率は 1/3 まで上昇させていたならば，我々の判断が間違っていた確率は，上の表から，$P(x \leq 13) = 0.9982 = \beta$ もあることが分かる．

このように，統計的仮説検定では，H_0 が真であるにも関わらずこれを棄却する誤り（Type1 のエラー）と H_1 が真であるにも関わらずこれを棄却する誤り（正しくは「H_0 が偽であるにも関わらずこれを棄却しない誤り」と言うべき：Type2 のエラー）がある．ただし，通常 Type2 のエラー確率 β は計算が困難であり，あまり考慮されない．

なお，H_0 が正しくないとき，正しく H_0 を棄却する，すなわち H_1 が正しいとき，正しく H_1 を採択する確率を検定力といい，この確率は $1 - \beta$ である．我々の行った検定は，実は $1 - \beta = 1 - 0.998 = 0.002$ という非常に検定力の小さいものだったのである．

以上は二項分布を利用した仮説検定についてであるが，次に，標準正規分布及び t 分布を利用した統計的仮説検定について述べる．

6.2 母集団の平均，分散が分かっている場合：z 検定

X を確率変数とし，母集団における確率分布が平均 μ，分散 σ^2 の正規分布（Normal distribution）であるとする．これを次のように表現することにする．

$$x \sim N(\mu, \sigma^2)$$

ここでは，μ 及び σ^2 は既知であるとする（このようなことはほとんどあり得ないが）.

次に，大きさ N の標本を母集団から無作為に抽出する．このとき，標本平均 $\bar{X}$ は，平均 μ，分散 σ^2/N の正規分布に従う．すなわち，

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{N})$$

さらに，$\bar{X}$ をこの平均及び標準偏差を用いて標準化する．すなわち，

$$z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/N}}$$

このとき，標準得点 z は平均 0，分散 1 の正規分布（これを標準正規分布という）に従う．すなわち，

$$z \sim N(0, 1)$$

以上を用いて統計的仮説検定を行うことができる．

例 **1**：**z** 検定（ 両側検定）

WISC-R で測定した小学校 3 年生全体の IQ は平均が 100，分散が 225（標準偏差が 15）であることが分かっているものとする．このとき，無作為に抽出した 5 月生まれの 3 年生 100 人に WISC-R を実施したところ，平均値は 103 となった．5 月生まれの 3 年生の IQ の平均は全国平均と異なるだろうか．

- 小学校 3 年生の IQ の全国平均を μ_0 とする（$\mu_0 = 100$）
- 5 月生まれの 3 年生の IQ の全国平均を μ_1 とする（μ_1 は不明）
- 帰無仮説 H_0：$\mu_0 = \mu_1$
  (5 月生まれの 3 年生の IQ の全国平均は 3 年生全体の全国平均に等しい)
- 対立仮説 H_1：$\mu_0 \neq \mu_1$
  (5 月生まれの 3 年生の IQ の全国平均は 3 年生全体の全国平均とは異なる)
- 危険率 5 %で帰無仮説の両側検定を行うことにする
  (帰無仮説が誤って棄却されてしまう [これを Type1 のエラーという] 危険性が 5 %であるということ)
- 帰無仮説が正しいと考えて推論を進める．
- 帰無仮説が正しい，すなわち大きさ 100 の標本（5 月生まれの 3 年生）は平均 100，分散 225 の正規分布からの無作為標本である，とすると，
- 大きさ 100 の標本平均は平均 100，分散 2.25（標準偏差 1.5）の正規分布に従う．
- 5 月生まれの 3 年生 100 人の IQ の平均値は 103 であった．これを標準化すると，

  $$z = \frac{103 - 100}{1.5} = 2.0$$

  となる．z は標準正規分布に従うことから $P(|z| \geq 2.0)$ を求める．

- 求められた確率が.05よりも小さければ帰無仮説を棄却する
  （対立仮説を採択し「5月生まれの3年生のIQの平均は全国平均と異なる」と結論づける）
- 求められた確率が.05よりも大きければ帰無仮説を採択する
  （「5月生まれの3年生のIQの平均は全国平均と異ならない」と結論づける）

例 **2**：**z** 検定（片側検定）
WISC-R で測定した小学校3年生全体のIQは平均が100，分散が225（標準偏差が15）であることが分かっているものとする．このとき，無作為に抽出した5月生まれの3年生100人にWISC-Rを実施したところ，平均値は103となった．5月生まれの3年生のIQの平均は全国平均よりも高いだろうか．

- 小学校3年生のIQの全国平均を μ_0 とする（$\mu_0 = 100$）
- 5月生まれの3年生のIQの全国平均を μ_1 とする（μ_1 は不明）
- 帰無仮説 H_0：$\mu_0 = \mu_1$
  (5月生まれの3年生のIQの全国平均は3年生全体の全国平均に等しい)
- 対立仮説 H_1：$\mu_0 < \mu_1$
  (5月生まれの3年生のIQの全国平均は3年生全体の全国平均より高い)
- 危険率5％で帰無仮説の片側検定を行うことにする

- 帰無仮説が正しいと考えて推論を進める．
- 帰無仮説が正しい，すなわち大きさ100の標本（5月生まれの3年生）は平均100，分散225の正規分布からの無作為標本である，とすると，
- 大きさ100の標本平均は平均100，分散2.25（標準偏差1.5）の正規分布に従う．
- 5月生まれの3年生100人のIQの平均値は103であった．これを標準化すると，

$$z = \frac{103 - 100}{1.5} = 2.0$$

  となる．z は標準正規分布に従うことから $P(z \geq 2.0)$ を求める．
- 求められた確率が.05よりも小さければ帰無仮説を棄却する
  （対立仮説を採択し「5月生まれの3年生のIQの平均は全国平均より高い」と結論づける）
- 求められた確率が.05よりも大きければ帰無仮説を採択する
  （「5月生まれの3年生のIQの平均は全国平均と異ならない」と結論づける）

6.2.1 検定される仮説は帰無仮説だけである

統計的仮説検定での検定仮説は帰無仮説（例えば，$H_0 : \mu_0 = \mu_1$）である．これを棄却するかしないかを確率に基づいて判断するのが統計的仮説検定である．以下に注意しなければならない．

1. 「帰無仮説が棄却されないこと＝帰無仮説が正しい」ではない
2. 「帰無仮説が棄却される＝対立仮説が正しい」ではない

あくまでも確率に基づく意思決定である．統計が正しさを「証明」してくれるわけではない．

6.2.2 両側検定と片側検定

帰無仮説を $H_0 : \mu_0 = \mu_1$ とすると，

- 両側検定：対立仮説が $\mu_0 \neq \mu_1$ のとき
- 片側検定：対立仮説が $\mu_0 < \mu_1$ あるいは $\mu_0 > \mu_1$ のとき

である．基本的には両側検定であると考えた方がよい．片側検定は，何らかの明確な根拠があって $\mu_0 < \mu_1$（あるいは $\mu_0 > \mu_1$）でしかあり得ないという特殊な場合であると考えて欲しい．分析者の主観によって選べる類のものではない（したがってここで挙げた例は不適切である）．

通常，分析者（研究者）の仮説は「男性よりも女性の方が○○が大きい（小さい）」である．しかし帰無仮説の棄却は「男性＝女性ではない」までしか意味しない．ここから「男性¿女性」という結論を導くことには「飛躍」があるということを知っておいて欲しい．

6.3 母集団の分散が分からない場合：t 検定

上の例ではいずれも，母集団に仮定した正規分布の平均も分散も分かっているものとした．ここでは，平均は分かっているけれども，分散は分からないものとする．すると，
大きさ N の標本を母集団から無作為に抽出すると，標本平均 $\bar{X}$ は，平均 μ，分散 σ^2/N の正規分布に従う．すなわち，

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{N})$$

ここまでは上の場合と同じである．上ではこのあと $\bar{X}$ を母集団の平均及び標準偏差を用いて標準化したのだが，困ったことにここでは $\sqrt{\sigma^2/N}$ が分からない．そこで，σ^2 の代わりにその推定値 $\hat{\sigma}^2$ で代用することにする．すでに述べたように $\hat{\sigma}^2$ はデータから求めることができ，

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

である．この推定値を使って標準化を行い，この標準得点を t で表す．すなわち，

$$t = \frac{\bar{X} - \mu}{\sqrt{\hat{\sigma}^2/N}}$$

このとき，t は自由度 $N-1$ の t 分布に従う．仮説の検定は，自由度 $N-1$ の t 分布を用いて行うことになる．

例 **3**：**t** 検定（片側検定）
WISC-R で測定した小学校 3 年生全体の IQ は平均が 100 であることが分かっているものとする．このとき，無作為に抽出した 5 月生まれの 3 年生 100 人に WISC-R を実施したところ，平均値は 102,（不偏）分散は 169 となった．5 月生まれの 3 年生の IQ の平均は全国平均よりも高いだろうか．

- 小学校 3 年生の IQ の全国平均を μ_0 とする（$\mu_0 = 100$）
- 5 月生まれの 3 年生の IQ の全国平均を μ_1 とする（μ_1 は不明）
- 帰無仮説 $H_0 : \mu_0 = \mu_1$
  (5 月生まれの 3 年生の IQ の全国平均は 3 年生全体の全国平均に等しい)

- 対立仮説 H_1：$\mu_0 < \mu_1$
  (5月生まれの3年生のIQの全国平均は3年生全体の全国平均より高い)
- 危険率5％で帰無仮説の片側検定を行うことにする

- 帰無仮説が正しいと考えて推論を進める.
- 帰無仮説が正しい，すなわち大きさ100の標本（5月生まれの3年生）は平均100，分散169の正規分布からの無作為標本である，とすると，以下の統計量 t は自由度99の t 分布に従う.

$$t = \frac{102 - 100}{1.3} = 1.538$$

- t 分布の表から対応する t の値（臨界値という，この場合 $t(.05, 99)$ と表現する）を求め，データから計算された t 値が $t(.05, 99)$ を超えていれば帰無仮説を棄却する
  （対立仮説を採択し「5月生まれの3年生のIQの平均は全国平均より高い」と結論づける）
- データから計算された t 値が $t(.05, 99)$ を超えなければ帰無仮説を採択する
  (「5月生まれの3年生のIQの平均は全国平均と異ならない」と結論づける)

なお両側検定の場合は（危険率を α とすると）t の臨界値は $t(\alpha/2, df)$ となる．上の例では $t(.25, 99)$ と計算された t とを比較することになる.

6.4 2群の平均値の差の検定：t検定

実際のところ，仮定される母集団分布については，平均も分散も未知である．このような場合，２群の（母集団における）平均値の差の検定を行うために非常によく用いられるのが t 検定である.

t 検定を行うにあたっては以下の条件が満たされていなければならない.「２群の平均値の差の検定だから t 検定だ」というような利用法は慎まなければならない.

1. 母集団分布に正規分布を仮定できる（ただし頑健性あり）
2. ２条件間の分散が等しいと見なせる（分散の等質性）
3. 標本が母集団から無作為に抽出されている（互いに独立に同一分布に従う）

こうした条件が明らかに満たされていない場合，ノン・パラメトリックな方法（いずれ述べる）を用いる方がよい．扱うデータが質的データの場合は特に注意を要する．なお，分散が２群で等質とはいえない場合には，ウェルチ検定（後述）などがある.

6.4.1 対応がない測定値の場合（被験者間要因）

標本A：標本数$=n_A$ … 正規分布 $N(\mu_A, \sigma_A^2)$ からの標本とする
標本B：標本数$=n_B$ … 正規分布 $N(\mu_B, \sigma_B^2)$ からの標本とする
ただし、$\sigma_A^2 = \sigma_B^2 = \sigma^2$ とする (分散の等質性の仮定).

○このとき、
・帰無仮説：$\mu_A - \mu_B = 0$（$\mu_A = \mu_B$）→　母平均は同じ
・対立仮説：$\mu_A - \mu_B \neq 0$（$\mu_A \neq \mu_B$）→　母平均が異なる

○母分散が分かっている場合、標本平均の差　$\bar{X}_A - \bar{X}_B$　の分布は
・平均＝0
・分散 $= \dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B} = \sigma^2\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)$ の正規分布に従う。

○これを標準化すると、

$$z = \frac{\bar{X}_A - \bar{X}_b}{\sqrt{\sigma^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

となり、あとはこの z の値を標準正規分布に照らし合わせればよい

○しかし、母分散 σ^2 は分からない　→　推定値 $\hat{\sigma}^2$ を用いる

$$\hat{\sigma}^2 = \frac{(n_A - 1)\hat{\sigma}_A^2 + (n_B - 1)\hat{\sigma}_B^2}{(n_A - 1) + (n_B - 1)}$$

○この推定値を用い、上と同様の標準化を行うと、

$$t = \frac{\bar{X}_A - \bar{X}_b}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

これは自由度　$n_A + n_B - 2$ の t 分布に従う．これから分かるように t は，２群の平均値の差の標準得点の一種である．

6.4.2　２群の分散が等質ではない場合：Welch test

上述した t の代わりに以下の t' を求める．

$$t' = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}}$$

これは次式によって求められた自由度の t 分布に従うことが分かっている．

$$\text{自由度}\nu = \frac{\left(\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}\right)^2}{\frac{\hat{\sigma}_A^4}{n_A^2(n_A-1)} + \frac{\hat{\sigma}_B^4}{n_B^2(n_B-1)}}$$

6.4.3　対応がある測定値の場合（被験者内要因）

例えば表 4.4 のようなデータが得られたとする．

このように１人の被験者から２つ以上の測定値を得る場合，対応がある測定値（**repeated measurement**, 反復測定デザイン）という．検定の手続きは次の通りである．

- 両条件下での得点の母平均をそれぞれ μ_L, μ_H とする．
- 帰無仮説は $\mu_L = \mu_H$，すなわち $\mu_L - \mu_H = 0 = \mu_D$ と表現される．
- 表中 D は各被験者についての条件間での得点の差であるが，帰無仮説が正しいとき，母集団におけるその平均値は $\mu_D = 0$ である．
- そこで，この各 D を平均 0，分散未知の正規分布からの標本であるとする．

表 6.4: 騒音の大きさが計算作業に及ぼす効果に関する実験データ（森・吉田,1990,p.65）

被験者	低騒音 (X_L)	高騒音 (X_H)	差 $D(X_L - X_H)$
A	78	76	2
B	70	72	-2
C	66	60	6
D	76	72	4
E	78	70	8
F	76	72	4
G	88	84	4
H	76	70	6
$\bar{X}$	76.0	72.0	$\bar{D} = 4.0$
s	6.00	6.24	$\hat{\sigma}_D^2 = 9.14$

- すると，先に述べた平均が既知，分散が未知の場合の t 検定となる．t 統計量は以下の通り．

$$
\begin{aligned}
t &= \frac{\bar{D} - \mu_D}{\sqrt{\hat{\sigma}_D^2/n}} \\
&= \frac{4 - 0}{\sqrt{9.14/8}}
\end{aligned}
$$

注
心理学研究では，例えば，反復デザイン計画は個人差を分離できるので優れている，というような説明がなされたりするが，これを鵜呑みにしてはいけない．反復測定デザイン下でのデータは実は扱いが困難である．反復測定デザインデータの解析については第？章で詳しく述べることにする．

6.5 仮説検定のまとめ

心理学研究におけるデータ収集計画から統計的仮説検定に至るまでのプロセスはおよそ次の通りである（例えば実験的研究の場合）．

1. 特定の心理学的現象とそのメカニズム等に関心をもつ．
   ：例えば,「ある行動に A という要因が影響を及ぼす」というような仮説をもつ．
2. ある行動の量や質を測定するモノサシを準備する．
3. 要因 A の水準を設定する（要因 A がある／ないなど．3 水準以上でもよい）．
4. 要因 A の水準毎に行動の量や質を比較できるように実験を計画する（剰余変数をうまく統制できるように）．
5. データ発生の確率モデル（母集団における確率分布，通常は正規分布）を仮定する．
6. 実験等で得られるデータを確率変数 X とみなす．
7. 検証したい仮説に対応する帰無仮説を用意する（母集団において要因 A の各水準間の平均値に差がない等）．

8. 仮説検定のルールを設定する
：Type1 のエラー（誤って帰無仮説を棄却してしまう危険性）の危険率 α の設定．ここでは $\alpha = .05$ とする．
：両側検定か片側検定かを決める．

9. 自分の計画した実験から得られるデータを解析することによって，言いたいことが言えるのか，調べたいことが分かるのかを予め確認する（どんな解析手法を用いればよいか）．

10. 母集団からの無作為標本を準備する（結果を一般化できる）

11. 被験者を実験要因の各水準に無作為に割り付ける（実験要因の効果の有無についての統計的推論ができる）

12. 確率変数 X が独立に同一の母集団分布に従うという仮定から，帰無仮説の元で，種々の統計量（平均値，分散，相関係数など）がどのような分布に従うかが導き出される（自分で導出する必要はない）．

13. 実験の結果得られたデータから計算されるそうした統計量が，当該の帰無仮説の元での分布に照らし合わせて，どの程度で起こりうることかを確認する．

14. 計算の結果得られた検定統計量（t, F, χ^2 など）が，帰無仮説の元では得られにくいものである場合（$p < .05$），帰無仮説を棄却する（=対立仮説の採択．ただしこの判断が誤っている可能性は５％ある）．

15. 計算の結果得られた検定統計量が，帰無仮説の元でも充分に得られるようなものである場合（$p > .05$），帰無仮説を採択する（=対立仮説の棄却）．

注意

- 用いようとする検定手法には種々の仮定が置かれている．自分の計画がそれを満たしているかどうかを確認する必要がある．特に反復測定計画にはいろいろと問題があることを知っておいた方がよい．

- 仮説検定では，Type1 のエラーだけではなく，Type2 のエラーも考慮すべき．

- Type2 のエラーに関して，その危険率を β とすると，$1 - \beta$ を検定力（検出力）という．これは，対立仮説 H_1 が正しいときに正しくそれを採択する力のことであり，検定力が低い検定（実験）のもとでの仮説検定はあまり意味がない．

- 検定の結果得られた確率値，いわゆる p 値の大小に一喜一憂する傾向が見られるが，これはナンセンスである．p 値の大小を効果や関係の大小・強弱と混同してはいけない．例えば，変量間の共変関係の強弱の指標としてよく用いられるものにピアソンの積率相関係数があるが，標本数が多ければ $r = 0.1$ でも $p < .01$ となる．$p < .01$ だけを見て「有意な相関が見られた」などと決して言わないこと．

- 効果の大きさの指標には効果量（effect size, ES）という指標がある．ES を確認することも必要である．

$$ES = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sum_i^{n1}(y_{1i} - \bar{y}_1)^2 + \sum_i^{n2}(y_{2i} - \bar{y}_2)^2}{n1 + n2 - 2}}}$$

  つまり，標準化された平均値の差（絶対値）であると考えればよい．サンプル数が大きければ ES が小さくても帰無仮説が棄却されやすくなる．実質的な差の意味も考えるべきである．

第7章　種々の解析手法とそのモデル

心理学研究におけるデータ解析手法は様々である．特によく用いられるものには，分散分析，重回帰分析，因子分析などがあるが，最近では，共分散構造分析のように複雑な解析手法が用いられることも増えてきた．しかし，十分に理解した上でそれらの手法を用いているとは思えないようなケースもよく見かける．不適切な統計解析手法によって得られた結果は信頼できない．研究報告を信頼性の高いものにするためには，目的に応じた方法の選択，目的及び方法に適したデータ解析手法の選択が欠かせない．そのためには，種々の解析手法について十分に理解しておく必要がある．以下に種々の解析手法の基本的な考え方と仕組みを説明する．

7.1　重回帰分析 (1)：基本モデル

1つの基準変数をp個の他の変数で説明（あるいは予測）しようとする方法である．モデルは次の通り．

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \\
&= \beta_0 + \sum_j^p \beta_j x_{ji} \cdots \text{観測値} \\
\hat{y}_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \\
&= \beta_0 + \sum_j^p \beta_j x_{ji} \cdots \text{説明変数による } y \text{ の予測値} \\
y_i &: \text{個人 i の基準変数 } y \text{ の観測値} \\
x_{ji} &: \text{個人 i の説明変数 j における観測値} \\
\beta_j &: \text{偏回帰係数：説明変数の重み} \\
\varepsilon_i &: \text{残差：観測値と予測値の差}
\end{aligned}
$$

回帰分析では，観測値yと説明変数による予測値$\hat{y}$とのズレを最小化するように偏回帰係数βが求められる．つまり，以下の量Qを最小化するのである．

$$
\begin{aligned}
Q &= \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2 \\
&= \frac{1}{n}\sum_i^n (y_i - \beta_0 - \sum_j^p \beta_j x_{ji})^2
\end{aligned}
$$

偏回帰係数は，上の式からも明らかなように，基準変数の予測式における説明変数の重み係数である．この値が大きいほど基準変数の説明（予測）への寄与が大きいと言える．

7.1.1　偏回帰係数・標準偏回帰係数

用いるデータが全て標準化されているとき（平均0，分散1）の偏回帰係数は標準偏回帰係数と呼ばれる（β^* と表す）．このとき$\beta_0 = 0$である．なお，偏回帰係数と標準偏回帰係数との間には次の関係がある．

$$
\beta_j = \frac{s_y}{s_j}\beta_j^*
$$

$$\begin{aligned}\beta_j^* &= \frac{s_j}{s_y}\beta_j \\ s_y &: \text{基準変数 } y \text{ の標準偏差} \\ s_j &: \text{説明変数 } x_j \text{ の標準偏差}\end{aligned}$$

y と x_j から残りの $p-1$ 個の x の影響を取り除いたあとの残りをそれぞれ $\tilde{y}$, $\tilde{x}_j$ （残差変数という）とおくと，x_j の偏回帰係数 β_j は，$\tilde{y}$ と $\tilde{x}_j$ との単回帰係数と一致する（なお「y と x_j から残りの $p-1$ 個の x の影響を取り除く」とは，それぞれ y と x_j から残りの $p-1$ 個の x による回帰予測値を引くことを意味する）．

また，y と x_j が標準化されている場合，これは y と x_j の偏相関係数となる．したがって，標準化偏回帰係数は y と x_j の偏相関係数である．

ところで，r_{yj} を y と x_j との相関係数，r_{jk} を x_j と x_k との相関係数とすると，

$$\begin{aligned}r_{yj} &= \beta_1^* r_{j1} + \beta_2^* r_{j2} + \cdots + \beta_j^* + \cdots + \beta_p^* r_{jp} \\ &= \sum_k^p \beta_k^* r_{jk}\end{aligned}$$

が成り立つ．なぜならば，y と x_j の共分散を s_{yj} とすると，

$$s_{yj} = \frac{1}{n}\sum_i^n (y_i - \bar{y})(x_{ji} - \bar{x}_j) \cdots \text{共分散の定義}$$

ここで，y, x をそれぞれ標準化すると，

$$\begin{aligned}s_{yj} &= r_{yj} = \frac{1}{n}\sum_i^n y_i x_{ji} \\ &= \frac{1}{n}\sum_i^n \{(\sum_k^p \beta_k^* x_{ki}) x_{ji}\} \\ &= \sum_k^p \{\beta_k^* \frac{1}{n}\sum_i^n x_{ki} x_{ji}\} \\ &= \sum_k^p \beta_k^* r_{jk}\end{aligned}$$

7.1.2 重相関係数

偏回帰係数 β は観測値と予測値のズレが最小となるように決められると述べたが，このことは，観測値 y と予測値 $\hat{y}$ との相関係数を最大化することと同じである．この観測値 y と予測値 $\hat{y}$ との相関係数 $r_{y\hat{y}}$ を重相関係数と言う．

$$\begin{aligned}r_{y\hat{y}} &= \frac{s_{y\hat{y}}}{s_y s_{\hat{y}}} \\ s_{y\hat{y}} &: \text{観測値 } y \text{ と予測値 } \hat{y} \text{ との共分散} \\ s_y &: \text{観測値 } y \text{ の標準偏差} \\ s_{\hat{y}} &: \text{予測値 } \hat{y} \text{ の標準偏差}\end{aligned}$$

7.1.3 決定係数

一般に変数の値の変動は平均値からの平方和で表現される．観測値 y の全変動（SS_T）は，

$$SS_T = \sum_i (y_i - \bar{y})^2$$

と表される．ところで，観測値は予測値と残差から構成され，

$$\begin{aligned} \bar{y} &= \bar{\hat{y}} \\ Cov(e, \hat{y}) &= 0 \cdots \text{残差と予測値の共分散は} 0 \end{aligned}$$

なので，

$$\begin{aligned} e_i &= y_i - \hat{y}_i \text{より} \\ y_i - \bar{y} &= (\hat{y}_i - \bar{\hat{y}}) + (y_i - \hat{y}_i) \end{aligned}$$

これを用いると，

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{\hat{y}})^2 + \sum_i (y_i - \hat{y}_i)^2$$

これは，観測値 y の全変動（SS_T）が回帰による変動（SS_R）と残差変動（SS_E）に分解できることを意味する．さらに上式を n で割ると

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

が成り立つことが分かる．これは観測値 y の分散が予測値 $\hat{y}$ の分散と残差の分散に分解できることを意味する．このとき以下の数量を考える．

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

これは，観測値の全変動（＝分散）が説明変数を用いた回帰による予測値によってどれだけ説明されるかの割合を示している．これを決定係数という．

ところで重相関係数の 2 乗は

$$\begin{aligned} r_{y\hat{y}}^2 &= \frac{s_{y\hat{y}} s_{y\hat{y}}}{s_y^2 s_{\hat{y}}^2} \\ &= \frac{(s_y^2 - s_e^2)^2}{s_y^2 s_{\hat{y}}^2} \\ &= \frac{s_{\hat{y}}^2 s_{\hat{y}}^2}{s_y^2 s_{\hat{y}}^2} \\ &= \frac{s_{\hat{y}}^2}{s_y^2} \\ &= R^2 \end{aligned}$$

となり決定係数と一致する．

さらに決定係数については，以下が成り立つ．

$$R^2 = \sum_j \beta_j^* r_{yj}$$

なぜならば，

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

であり，全ての変数が標準化されているならば，

$$
\begin{aligned}
R^2 &= \frac{s_{\hat{y}}^2}{s_y^2} = s_{\hat{y}}^2 = \frac{1}{n}\sum_i^n \hat{y}_i^2 \\
&= \frac{1}{n}\sum_i^n (\beta_1^* x_{1i} + \cdots + \beta_j^* x_{ji} + \cdots + \beta_p^* x_{pi})(\beta_1^* x_{1i} + \cdots + \beta_k^* x_{ki} + \cdots + \beta_p^* x_{pi}) \\
&= \frac{1}{n}\sum_i^n \sum_j^p \beta_j^* (\beta_1^* x_{1i} x_{ji} + \cdots + \beta_k^* x_{ki} x_{ji} + \cdots + \beta_p^* x_{pi} x_{ji}) \\
&= \sum_j^p \beta_j^* (\beta_1^* \frac{1}{n}\sum_i^n x_{1i} x_{ji} + \cdots + \beta_k^* \frac{1}{n}\sum_i^n x_{ki} x_{ji} + \cdots + \beta_p^* \frac{1}{n}\sum_i^n x_{pi} x_{ji}) \\
&= \sum_j^p \beta_j^* (\beta_1^* r_{j1} + \cdots + \beta_k^* r_{jk} + \cdots + \beta_p^* r_{jp}) \\
&= \sum_j^p \beta_j^* \sum_k^p \beta_k^* r_{jk} \\
&= \sum_j^p \beta_j^* r_{yj}
\end{aligned}
$$

つまり，基準変数 y と説明変数 x_j との相関係数 r_{yj} と標準偏回帰係数 β_j^* との積の全ての j についての和は，重相関係数と一致する．このことから，r_{yj} と β_j^* との積は当該説明変数 j の予測に対する寄与度として利用されることがある．

なお，重相関係数は，サンプル数 N に比べて説明変数の数 p が多くなるとともに 1 に近づいていき，$N=p+1$ のとき 1 に等しくなる．説明変数が増えるということは，それだけ説明のためのパラメータが増えることを意味する．$N=p+1$ のときにはすべてのサンプルを完全に説明することができるが，これは説明の仕方がデータの数だけあるに過ぎない（意味がない）．したがって，モデルの良さを判定するためには，説明変数の数も考慮に入れた方がよい．

説明変数の数を考慮に入れた適合度（説明度）の指標となるのが，自由度調整済み決定係数である．これは，説明変数の数が増えると自由度も増えることを考慮した決定係数 R^2 の変形である（次式）．

$$R^{*2} = 1 - \frac{s_e^2/df_e}{s_y^2/df_y} = 1 - \frac{N-1}{N-p-1}(1-R^2)$$

予測式の適否を判定する場合には，この自由度調整済みの決定係数を用いるようにすること．

7.1.4 モデル全体の検定

モデル全体の有効性に関しては以下の事実を用い F 検定を行うことができる．

$$
\begin{aligned}
\text{モデルが適合しているとき} &: V_E = \frac{SS_E}{n-p-1} \sim \chi^2_{n-p-1} \\
\text{全ての回帰係数が 0 であるとき} &: V_R = \frac{SS_R}{p} \sim \chi^2_p
\end{aligned}
$$

このとき帰無仮説 H_0 を次のように設定すると，

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$F_0 = \frac{V_R}{V_E}$$

は帰無仮説の元で自由度（$p, n-p-1$）の F 分布に従う．

7.1.5　偏回帰係数の検定

検定仮説は，各係数について

$$H_0 : \beta_j = 0$$

である．回帰係数の推定値を β_j 及びその標準誤差を $SE(\hat{\beta}_j)$ とすると，

$$F = \frac{\hat{\beta}_j^2}{SE(\hat{\beta}_j)^2}$$

が帰無仮説の元で，自由度（$1, n-p-1$）の F 分布に従うことを利用する．

7.1.6　説明変数の選択

p 個の説明変数による重回帰モデルを考え，偏回帰件数等を求めたとする．その結果，いくつかの説明変数において $\beta_j = 0$ の帰無仮説が棄却されなかったとする（つまり「有意でなかった」）．このとき，そのまま解析を終了するのはよくない．係数が 0 である変数を含んだモデルを採択したことになるからである．基準変数の説明あるいは予測に関係のないような無駄な説明変数はモデルからは取り除くべきである．実際には，大抵の統計ソフトには変数選択の機能があるので，それを利用すればよい．

7.2　重回帰分析 (2)：属性によって回帰係数が異なるモデル

以下の表のデータは，ある病院で生まれた男女 12 人ずつの乳児の出生時体重 (g) と推定妊娠期間 (週) を示したものである．また，これを図示したものが図 5.1 である．

表 7.1: 男児および女児の出生時体重と妊娠期間

	男児		女児	
	妊娠期間	出生時体重	妊娠期間	出生時体重
	40	2968	40	3317
	38	2795	36	2729
	40	3163	40	2935
	35	2925	38	2754
	36	2625	42	3210
	37	2847	39	2817
	41	3292	40	3126
	40	3473	37	2539
	37	2628	36	2412
	38	3176	38	2991
	40	3421	39	2875
	38	2975	40	3231
平均	38.3	3024.0	38.8	2811.3

図 5.1 から，妊娠期間が長ければ出生時体重が大きいという関係を読みとることができる．このようなケースでは，妊娠期間と出生時体重との関連（増加率）が男女で同じなのかどうかということが興味の対象

となることがある．このとき，増加率が男女で異なるというモデルは次のように表現される．

$$y_{ji} = \beta_{0j} + \beta_j x_{ji} + e_{ji}$$

ただし，

y_{ji} ： 性別 j における個人 i の出生時体重（男児：j=1，女児：j=2）

β_{0j} ： 性別 j の切片

β_j ： 性別 j の傾き（増加率）

x_{ji} ： 性別 j における個人 i の妊娠期間

e_{ji} ： 誤差項：それぞれ独立に同じ正規分布 $N(0, \sigma^2)$ に従う

帰無仮説および対立仮説は次の通りである．

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 (= \beta) \\ H_1 &: \beta_1 \neq \beta_2 \end{aligned}$$

また，これに対応する予測モデルは次の通りである．

$$\begin{aligned} y_{ji} &= \beta_{0j} + \beta x_{ji} + e_{ji} \cdots \text{モデル } 0 \\ y_{ji} &= \beta_{0j} + \beta_j x_{ji} + e_{ji} \cdots \text{モデル } 1 \end{aligned}$$

この２つのモデルの違いは回帰係数を男女で等しいとするか，異なるとするかである．検定は，回帰係数を異なるとすることによってどれだけモデルの適合度が増加するか，を基準に行われる．詳しくは次の通り．

H_1 が正しいとき，

$$\begin{aligned} SS_1 &= \sum_j \sum_i (y_{ji} - \beta_{0j} - \beta_j x_{ji})^2 \\ SS_1/\sigma^2 &\sim \chi^2_{24\ 4} \end{aligned}$$

H_0 が正しいとき，

$$\begin{aligned} SS_0 &= \sum_j \sum_i (y_{ji} - \beta_{0j} - \beta x_{ji})^2 \\ SS_0/\sigma^2 &\sim \chi^2_{24-3} \end{aligned}$$

となる．モデル 0 と比較するとモデル 1 はパラメータが 1 つ多い．このことによる当てはまりの改善度は

$$(SS_0 - SS_1)/\sigma^2 \sim \chi^2_1$$

である．帰無仮説が正しいとき，以下の比 F は自由度 $1, 24-4$ の F 分布に従うので，これを利用して検定を行う．

$$F = \frac{(SS_0 - SS_1)/1}{SS_1/(24-4)} \sim F_{1,24-4}$$

重回帰分析の仲間

- 数量化 I 類：基準変数，説明変数ともに質的データである場合の重回帰分析
- ロジスティック回帰分析：基準変数が 2 値変数である場合の回帰分析
- 分散分析・共分散分析：説明変数がダミー変数（処理水準）である場合の回帰分析であると言える

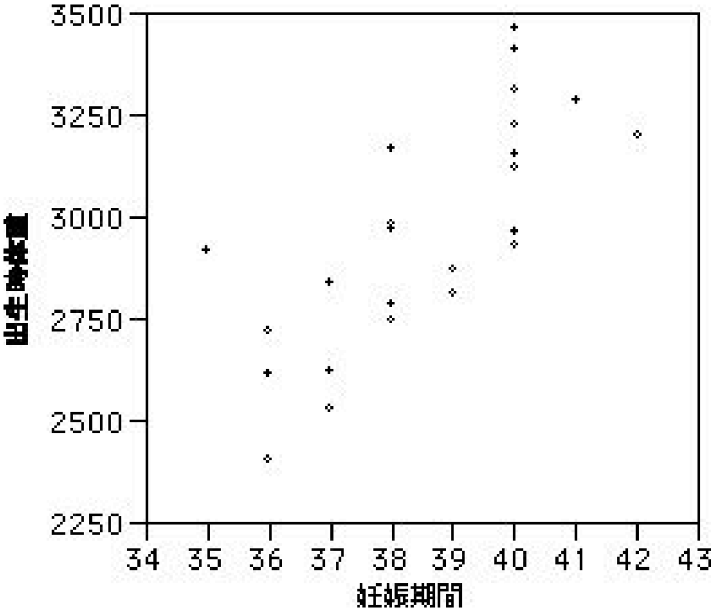

図 7.1: 男児および女児の出生時体重と妊娠期間

7.3 一要因分散分析（被験者間要因）

ある心理現象に関心があるとする．その心理現象にはどのような要因が影響しているのだろうか．予想される要因があり，その要因を操作できる場合には実験が計画される．

単語の記憶再生成績に対する記銘方略の影響を調べたい（森・吉田，1990）．互いに無関連な 20 個の単語を覚える際の記銘方略として，

- 自由方略：方略指示なし
- 反復方略：リハーサルを行う
- 文章方略：文章を作る
- イメージ方略：視覚的イメージを構成する

を設定する．各条件にそれぞれ 8 名ずつ無作為に被験者を割り当てる．再生できた単語数を記憶成績とする．この実験の特徴は，

- 要因：記銘方略
- 水準：自由，反復，文章，イメージの 4 水準
- 各水準に別々の被験者が無作為に割り当てられる＝被験者間計画

である．結果は表 5.2 の通りである．

表中「全体平均との差」に着目する．全体平均からの差は何に由来するのだろうか．考えられる他の要因がどの方略の元でも同じように働くと考えられるとき，この全体平均からの差は，記銘方略に由来すると考えることができる．この全体平均からの差をそれぞれの記銘方略の効果という．誤差による変動を考慮した上で，この効果がすべて 0 であるという仮説を検討するのが分散分析である．

表 7.2: 記銘方略別の平均単語再生数

記銘方略	平均	全体平均との差
自由方略	9.50	-0.88
反復方略	6.50	-3.88
文章方略	12.50	+2.12
イメージ方略	13.00	+2.62
全体	10.38	0.00

7.3.1 モデル

a 個の処理水準を持つ要因 A を考える．処理の各水準を $A_1, A_2, \cdots, A_a$ とし，それぞれの水準にそれぞれ $n_1, n_2, \cdots, n_a$ 人の被験者を無作為に割り当て測定値を得る．そして，測定値の平均値を比較することによって，処理の効果を吟味する．

ここで，処理 A_j における i 番目の被験者（$i = 1, 2, \cdots n_j$）の測定値を X_ij とする．このとき測定値のモデルは，

$$X_{ij} = \mu + \alpha_j + e_{ij}$$

$$\text{ただし，} \quad \sum_{j=1}^{a} \alpha_j = 0 \text{ とする}$$

と書ける．ここで，α_j は処理 A_j の効果を示し，μ は全体平均（あるいは一般平均）である．各処理における平均を μ_j とすると，

$$\mu_j = \mu + \alpha_j$$

となる．すなわち，測定値は，全体的な平均値（本当の値）と水準 j による効果と誤差から成り立っていると考えるのである．また，誤差項 e_{ij} については，互いに独立に同一の正規分布 $N(0, \sigma_e^2)$ に従うことが仮定される（σ_e^2 は誤差分散で，一定とする=分散の等質性の仮定）．

分散分析では，処理効果がないという以下の帰無仮説をテストする．

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

7.3.2 分散分析における仮説検定の原理

[記号]

- X_{ij}：処理 j における被験者 i の測定値
- $\bar{X}_{.j}$：処理 j における測定値の平均値
- $\bar{X}_{..}$：測定値の全体平均

次の恒等式を利用する．

$$X_{ij} - \bar{X}_{..} = (\bar{X}_{.j} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{.j})$$

つまり，個々の測定値と全体平均との偏差は,「処理水準内の平均と全体平均との偏差」と「個々の平均値と処理水順内の平均との偏差」との和である．

これの両辺を平方すると，以下が得られる．

$$\sum_{j=1}^{a}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{..})^2=\sum_{j=1}^{a}n_j(\bar{X}_{.j}-\bar{X}_{..})^2+\sum_{j=1}^{a}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{.j})^2$$

つまり,（平均からの偏差の）全体平方和（SS_T）=処理間平方和（SS_B）+処理内平方和（SS_W）である．ここで次の量を考える．

$$\begin{aligned} MS_B &= \frac{SS_B}{a-1} \\ MS_W &= \frac{SS_W}{n-a} \\ F &= \frac{MS_B}{MS_W} \end{aligned}$$

このとき F は自由度 (a-1, n-a) の F 分布に従うことが分かっている．帰無仮説のもとでは，$MS_B=MS_W$ であり，分子分母ともに誤差成分のみとなる（処理の効果は 0）．処理の効果があれば，MS_B は大きくなり，従って F 値も大きくなる．

データから計算された F が自由度 (a-1, n-a) の F 分布における臨界値（設定した危険率における F の値）を超えれば，帰無仮説を棄却し，処理の効果があると結論づける．

7.3.2.1 χ^2 分布

さて，仮説のテストは，観測値から得られる何らかの統計量が，帰無仮説が正しいもとで特定の値以上（以下）になる「確率」に基づいて行われる.「帰無仮説が正しいときに統計量が○○以上である確率は××である．よって，帰無仮説を棄却（あるいは採択）する」という論理である．したがって，その「確率」を与える「確率分布」が必要となる．

$$X \sim N(\mu,\sigma^2)\cdots \text{X は母平均 } \mu\text{，母分散 } \sigma^2 \text{ の正規分布に従う}$$

のとき，

$$\chi^2_{(1)}=\frac{(X-\mu)^2}{\sigma^2}=z^2\cdots\text{標準得点の 2 乗}$$

は自由度 1 のカイ二乗（χ^2）分布に従う．

また，それぞれ独立に同一の分布に従う $X_1, X_2, \cdots, X_i, \cdots, X_n$ について，

$$\sum_{i}^{n}\frac{(X_i-\mu)^2}{\sigma^2}=\chi^2_{(n)}$$

は，自由度 n の χ^2 分布に従う．

しかし通常は母平均 μ は未知であり，これを標本平均 $\bar{X}$ で代用する．このとき，

$$\sum_{i}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2}=\chi^2_{(n-1)}$$

は，自由度 n-1 の χ^2 分布に従う．さらにこれは次のように変形できる．

$$\chi^2_{(n-1)}=\sum_{i}^{n}\frac{(n-1)(X_i-\bar{X})^2}{(n-1)\sigma^2}=\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

$\hat{\sigma}^2$ は母分散 σ^2 の標本に基づく（不偏）推定値である．

7.3.2.2 F 分布

２つの異なる母集団から得られた２つの χ^2 値，$\chi^2_{\nu_1}, \chi^2_{\nu_2}$ を考える．ν_1, ν_2 はそれぞれの自由度である．このとき次の量，

$$F = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$$

は自由度（ν_1, ν_2）の F 分布に従う．F 分布は２つの χ^2 分布に従う変数をそれぞれの自由度で割ったものの比の分布である．

さて，上述の χ^2 分布に従う２つの変数が共通の母集団から抽出されたものとすると，F は以下のようになる．

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

すなわち，同じ分散をもつ母集団から２つの標本を独立に抽出し，それぞれの標本における分散の不偏推定値を求めその比をとると F 分布に従う．これが分散分析における F 検定の原理である．反復測定計画ではこの仮定が成り立たないことをよく認識しておくべきである．

7.3.3 主効果と多重比較

分散分析における帰無仮説は，先にも述べたが以下の通りである．

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

この帰無仮説が棄却されたとき，要因 A の主効果がある，という．これは，要因 A の各処理水準における平均値の少なくともどれか 1 つは有意に異なっている，ということを意味する．水準数が２つであれば，どの水準間で平均値が異なっているかは明らかであるが，水準数が３つ以上の場合にそれを特定するには，水準間の多重比較が必要となる．

平均値間の比較であるから，水準を２つずつ取り出してそれぞれについて t 検定をすればよいではないか，と思うかも知れないがそれはマズイ．検定をやみくもに繰り返すと Type1 のエラーを犯す確率が大きくなってしまうからである．以下に示す多重比較の方法は，検定の全体での Type1 のエラーを犯す確率を統制するための方法である．

7.3.3.1 テューキーの HSD 検定：Tukey's honestly significant difference test

すべての処理水準における平均値を互いに比較したい場合に適している．第 1 種の誤りに対して保守的である．

Studentized range（スチューデント化された範囲）と呼ばれる以下の統計量がある．

$$q = \frac{\bar{X}_{max} - \bar{X}_{min}}{\sqrt{MS_e/n}}$$

ここで，　MS_e：当該の要因の主効果の検定に用いた誤差項の平均平方（不偏分散）

n：各処理水準のデータ数

$\bar{X}_{max}$：比較したい平均値の大きい方

$\bar{X}_{min}$：比較したい平均値の小さい方

この q の値は，有意水準 α，処理水準数 m，誤差項の自由度 df_e によって決まる（q の表は省略）．

テューキーの HSD 法では統計量 q を用い，以下の要領で HSD という量を求め，比較したい一対の平均値の差が，求められた HSD よりも大きい場合，帰無仮説（平均値に差がない）を棄却する．

$$HSD = q_{\alpha,m,df}\sqrt{\frac{MS_e}{n}}$$

ここで，　α：有意水準（主効果で用いたものと一致させること）

m：処理水準数

df：誤差項の自由度

n：平均値の計算に用いるデータ数：処理水準間で異なる場合はそれらの調和平均

7.3.3.2　ダネットの方法

ある特定の水準（例えば対照群．この水準の母平均を μ_1 とする）と他の水準 j との比較に関心がある場合に用いる．帰無仮説は次の通り．

$$H_{j0}: \mu_1 = \mu_j$$

仮説検定には次の t 統計量を用いる．

$$t_{j,1} = \sqrt{n}\frac{\bar{X}_j - \bar{X}_1}{\sqrt{2MS_e}}$$

この $t_{j,1}$ について，

$$t_{j,1} > d_{\alpha,m-1,df_e} \rightarrow \mu_j - \mu_1 > 0\text{：片側検定}$$

$$|t_{j,1}| > d_{\alpha,m-1,df_e} \rightarrow \mu_j \neq \mu_1\text{：両側検定}$$

ここで，　α：有意水準（主効果で用いたものと一致させること）

m：処理水準数

df_e：誤差項の自由度

n：セルのデータ数：処理水準間で異なる場合はそれらの調和平均

とする（d の表は省略）．

7.3.4　分散分析と多重比較

通常，統計のテキストでは分散分析の説明の中に多重比較の説明も含まれる．しかし，多重比較は分散分析の一部ではない．別のものである．「分散分析→主効果や交互作用が有意→多重比較」という手順が一般的な理由は，それが効率的であるからに過ぎない．

7.4　一要因分散分析（被験者内要因）

a 個の処理水準を持つ要因 A を考える．処理の各水準を $A_1, A_2, \cdots, A_a$ とし，n 人の被験者がおり，1 人の被験者がすべての水準に参加し測定値を得る．つまり，1 人の被験者について a 個の測定値が得られ，これが n 人分あるということである．

このとき，

- X_{ij}：処理 j における被験者 i の測定値
- $\bar{X}_{.j}$：処理 j における測定値の平均値

- $\bar{X}_{i.}$：被験者 i のすべての処理を通しての測定値の平均
- $\bar{X}_{..}$：測定値の全体平均

とすると，平方和について以下が成り立つ．

$$\sum_{j=1}^{a}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{..})^2=\sum_{j=1}^{a}n(\bar{X}_{.j}-\bar{X}_{..})^2=\sum_{i=1}^{n}a(\bar{X}_{i.}-\bar{X}_{..})^2+\sum_{j=1}^{a}\sum_{i=1}^{n}(X_{ij}-\bar{X}_{.j}-\bar{X}_{i.}+\bar{X}_{..})^2$$

つまり，全体平方和（SS_T）=処理間平方和（SS_B）+ 個人間平方和（SS_S）+ 誤差（SS_{RES}）である．ただし，計算上では，SS_{RES} は個人差と要因との交互作用である．しかしこれは分離できないので，こうした交互作用はないものと仮定し誤差として扱う（無茶な仮定である）．

平均平方 MS 及び自由度は次の通りであり，

- 要因 A：MS_A，自由度 a-1
- 個人差：MS_S，自由度 n-1
- 誤差：MS_{RES}　自由度 (a-1)(n-1)

要因 A の効果の検定は

$$F=\frac{MS_A}{MS_{RES}}$$

に基づいて行われる．個人差の要因が分離されている分，被験者間要因の場合と比較して，分母 MS_{RES} が小さくなる（のが普通である）．したがって，帰無仮説は棄却されやすくなる（検定力が高まる）．ただし，条件間の測定値の相関が高くなければ検定力は高まらない（正の相関が高いほど個人差要因が大きく関与していることになり，MS_{RES} は小さくなる）．

7.5 二要因分散分析（被験者間要因）

二要因分散分析では，関心のある要因が 2 つとなる．また，要因間の相互作用も検討の対象となる．

不安傾向の高低が課題 A の成績に影響するか，またそれは，実際に課題遂行の場に他者が存在することと関連があるのだろうか（森・吉田，1990）．あらかじめ測定された不安傾向尺度の成績によって，被験者を不安傾向が高い群と低い群に分けた．それぞれの群の被験者を無作為に他者あり条件と他者なし条件に割り当て課題 A の成績を比較した．結果は表 5.3 の通りである．

- 要因 A：不安傾向（高・低の 2 水準）
- 要因 B：他者存在（有・無の 2 水準）

表 7.3: 各要因の各水準別の課題の平均成績

	高不安	低不安	他者存在平均
他者あり	3.0	7.2	5.1
他者なし	4.0	3.6	3.8
不安傾向平均	3.5	5.4	全体平均=4.45

表 5.3 から全体平均を取り除くと次のようになる．

表 5.4 中，他者存在平均の 0.65，-0.65 が他者存在の効果を表している．同様に，不安傾向平均の-0.95, 0.95 が不安傾向の効果を表している．なお，全体平均を引いたので表 5.4 の全体平均は 0 となっている．次に不安傾向の効果を取り除く．

表 7.4: 各要因の各水準別の課題の平均成績（全体平均からの偏差を取り除く）

	高不安	低不安	他者存在平均
他者あり	-1.45	2.75	0.65
他者なし	-0.45	-0.85	-0.65
不安傾向平均	-0.95	0.95	全体平均=0.0

表 7.5: 各要因の各水準別の課題の平均成績（不安傾向平均からの偏差を取り除く）

	高不安	低不安	他者存在平均
他者あり	-0.5	1.8	0.65
他者なし	0.5	-1.8	-0.65
不安傾向平均	0.0	0.0	全体平均=0.0

表 7.6: 各要因の各水準別の課題の平均成績（他者存在平均からの偏差を取り除く）

	高不安	低不安	他者存在平均
他者あり	-1.15	1.15	0.0
他者なし	1.15	-1.15	0.0
不安傾向平均	0.0	0.0	全体平均=0.0

当然の事だが，不安傾向平均は0となる．さらに，ここから他者存在の効果を取り除くと，表5.6となる．

これは結局，データのセル毎の平均値から，2つの要因の効果を取り除いたものである．2つの要因の効果を取り除いたのにも関わらず，各セルは0とはなっていないことが分かる．これは何か．

以上の操作から分かるように，取り除かれた各要因の効果（0.65,-0.65や0.95,-0.95）は，それぞれの要因の「平均的な」効果であって，他方の要因の水準の違いは考慮されていない．このように各セルの平均値から，2つの要因の「平均的な」効果を取り除いた残りを交互作用という．この例（表5.6）でいうと，高不安群における他者ありの効果は-1.15，高不安群における他者なしの効果は1.15ということになる．

二要因分散分析では，各要因の平均的な効果，及び要因間の水準の組み合わせによる交互作用が0であるかどうかが調べられる．

7.5.1 モデル

2つの要因A,Bを考える．水準数はそれぞれa,bであるとする．このとき，処理jkにおけるi番目の被験者の測定値をy_{ijk}に以下を仮定するのがに要因分散分析モデルである．なお処理jkの被験者数をそれぞれn_{jk}ただし$\sum_{jk} n_{jk} = N$とする．

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk} \\
y_{ijk} &= \text{処理jkにおけるi番目の被験者の測定値} \\
\mu &: \text{全体平均（一般平均）} \\
\alpha_j &: \text{要因A水準jの効果} \\
\beta_k &: \text{要因B水準kの効果} \\
\gamma_{jk} &: \text{要因A水準jと要因B水準kの交互作用}
\end{aligned}
$$

ε_{ijk}　:　個人 i の要因 A 水準 j，要因 B 水準 k における誤差

ただし，すべての主効果及び交互作用の和は 0，すなわち，

$$\sum_j^a \alpha_j = \sum_k^b \beta_k = \sum_j^a \gamma_{jk} = \sum_k^b \gamma_{jk} = 0$$

であり，また，ε_{ijk} はそれぞれ独立に同一の正規分布 $N(0, \sigma_e^2)$ に従うことが仮定される（「独立に同一の」が非常に重要）．

分散分析モデルの考え方の中心は，処理 jk における平均を μ_{jk} を次のように構造化することにある．

$$\mu_{jk} = \mu + \alpha_j + \beta_k + \gamma_{jk}$$

つまり，処理 jk における平均は，全体平均，要因 A の水準 j の効果，要因 B の水準 k の効果，要因 A の水準 j と要因 B の水準 k との交互作用（要因 A,B の水準の組み合わせによる効果）から構成されているとする．そして，得られたデータ及び誤差 ε_{ijk} への仮定に基づき，以下の帰無仮説を検定する．

$$\begin{aligned}
\alpha_1 &= \alpha_2 = \cdots = \alpha_j = \cdots = \alpha_a = 0 \\
\beta_1 &= \beta_2 = \cdots = \beta_k = \cdots = \beta_b = 0 \\
\gamma_{11} &= \gamma_{12} = \cdots = \gamma_{jk} = \cdots = \gamma_{ab} = 0
\end{aligned}$$

以下に，2 つの要因の水準数がともに 2 の場合の主効果及び交互作用を表に表現する．

表 7.7: 二要因分散分析モデル（水準数=2）

要因		A 1	A 2	
B	1	γ_{11}	γ_{21}	β_1
B	2	γ_{12}	γ_{22}	β_2
		α_1	α_2	

表から分かるように，要因 A,B の主効果 α, β は，互いの水準の違いが考慮されていない（当たり前だが）．したがって，交互作用がない場合には考察の対象となるが，交互作用がある場合にはこれを取り上げることにあまり意味がない．

このモデルにおける帰無仮説は，

$$\begin{aligned}
\alpha_1 &= \alpha_2 = 0 \\
\beta_1 &= \beta_2 = 0 \\
\gamma_{11} &= \gamma_{12} = \gamma_{21} = \gamma_{22} = 0
\end{aligned}$$

である．言い換えれば，

$$\mu_{jk} = \mu$$

である．

7.5.2　観測値の変動の分解と F 検定

重回帰分析において,「観測値の変動＝予測値の変動+誤差変動」となることを述べた．分散分析でも同様のことが成り立つ．

$$
\begin{aligned}
SS_T &= \sum_j \sum_k \sum_i (y_{ijk} - \bar{y}_{...})^2 \\
SS_A &= \sum_j n_{j.} (\bar{y}_{j..} - \bar{y}_{...})^2 \\
SS_B &= \sum_k n_{.k} (\bar{y}_{.k.} - \bar{y}_{...})^2 \\
SS_{AB} &= \sum_j \sum_k (\bar{y}_{jk.} - \bar{y}_{j..} - \bar{y}_{.k.} + \bar{y}_{...})^2 \\
SS_W &= \sum_j \sum_k \sum_i (y_{ijk} - \bar{y}_{jk.})^2
\end{aligned}
$$

言葉で表すと

SS_T ： 個々の観測値と全体平均との差の平方和：全体変動

SS_A ： 要因 A の各水準の平均と全体平均との差の平方和：要因 A による変動

SS_B ： 要因 B の各水準の平均と全体平均との差の平方和：要因 B による変動

SS_{AB} ： 要因 A,B の水準の組み合わせによる変動

SS_W ： 要因 A,B の水準の組み合わせセル内の変動

である．これらの変動には以下が成り立つ．

$$SS_T = SS_A + SS_B + SS_{AB} + SS_W$$

さらに,

$$
\begin{aligned}
V_A &= \frac{SS_A}{a-1} \\
V_B &= \frac{SS_B}{b-1} \\
V_{AB} &= \frac{SS_{AB}}{(a-1)(b-1)} \\
V_W &= \frac{SS_W}{N-ab}
\end{aligned}
$$

とすると，帰無仮説の元では V_A, V_B, V_{AB}, V_W は全て等しく誤差分散の不偏推定値であり，それぞれ自由度 $a-1, b-1, (a-1)(b-1), N-ab$ の χ^2 分布に従う．このとき,

$$
\begin{aligned}
F_A &= \frac{V_A}{V_W} \sim F(a-1, N-ab) \\
F_B &= \frac{V_B}{V_W} \sim F(b-1, N-ab) \\
F_{AB} &= \frac{V_{AB}}{V_W} \sim F((a-1)(b-1), N-ab)
\end{aligned}
$$

が成り立つ．処理の効果や交互作用がある場合には，V_A, V_B, V_{AB} は大きくなり，したがって上記 F 値も大きな値となる．この F 値が設定された臨界値を超えれば，帰無仮説は棄却される．

7.5.3 変数ごとに分散分析を何回も行うことは好ましくない

例えば質問紙で収集したいくつかの変数の１つ１つについて，性などを要因とした分散分析を何回も行う，というようなことをよく見かける．これは好ましくない．できれば得られた変数全体に対して１つの検定を行う「多変量分散分析」を行って欲しい．理由は次の通りである．

1. 多変量検定では設定された危険率 α で検定できるが，p 個の単変量検定を行えば α が大きくなってしまう．例えば，10 個の変量についてそれぞれ $\alpha = 0.05$ の検定を行うとする．このとき，少なくとも１つの検定で Type 1 のエラーを犯してしまう確率は 0.05 よりも大きくなる．もし 10 個すべての変数が互いに独立ならば（こんなことはほとんどありえないが），帰無仮説の元で，少なくとも１つの仮説が誤って棄却される確率は次のようになる．

$$\begin{aligned} P(\text{少なくとも１つの仮説が誤って棄却される}) &= 1 - P(\text{すべての帰無仮説が正しく保留される}) \\ &= 1 - (0.95)^2 \\ &= 0.40. \end{aligned}$$

   つまり，少なくとも１つの検定で Type 1 のエラーを犯してしまう確率は 0.40 となるのである．

2. 単変量仮説検定は変数間の相関を無視するが，多変量検定では共分散行列を用いるので，変数間の関係に関する情報も考慮した検定を行う．

3. 多変量検定の方が検定力がある場合も多い．検定力とは，帰無仮説が間違っているときにそれを正しく棄却する確率である．単変量検定ではすべての p 個の変量で帰無仮説を棄却できない場合でも，多変量検定では正しく棄却することがある．各変数における小さな効果を総合的に判断するからである．

4. 多変量検定では，副産物として変数の線形結合が得られる．これにより，どのように変数が帰無仮説を棄却したのかについての情報も得られる．

7.6 共分散分析

共分散分析は，上で説明した「重回帰分析（2)」と非常によく似ている．「重回帰分析（2)」で挙げた例では２群間で回帰式の「傾き」が同じかどうかに関心があった．共分散分析では，群間の傾きは等しいとみなし「切片」が群間で同じかどうかを検討する．

表 5.3 は３種類の異なる訓練法による学業成績データである．学業成績は訓練開始時の適性検査スコアと共変関係にある．したがって，学業成績だけを群間で比較することはできない．適性検査スコア（補助変数という）を考慮した上での学業成績の訓練法間の違いを検討したい．言い換えれば，群間の適性検査スコアの違いを補正した上で学業成績の平均を比較するのである．

図 5.2 から学業成績は適性検査成績と共変関係にあること，全般に訓練 2,3 の学業成績が訓練１の群よりも高いことが読みとれる．

分析
以下のダミー変数を導入する．

$$\begin{aligned} d_1 &: \text{訓練法２であれば１，それ以外０} \\ d_2 &: \text{訓練法３であれば１，それ以外０} \end{aligned}$$

上のダミー変数を含んだ重回帰モデルを作る．

$$y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + e$$

表 7.8: 3 種の訓練法と学業成績

	訓練法 1		訓練法 2		訓練法 3	
	y	x	y	x	y	x
	6	3	8	4	6	3
	4	1	9	5	7	2
	5	3	7	5	7	2
	3	1	9	4	7	3
	4	2	8	3	8	4
	3	1	5	1	5	1
	6	4	7	2	7	4
平均	4.4	2.1	7.6	3.4	6.7	2.7

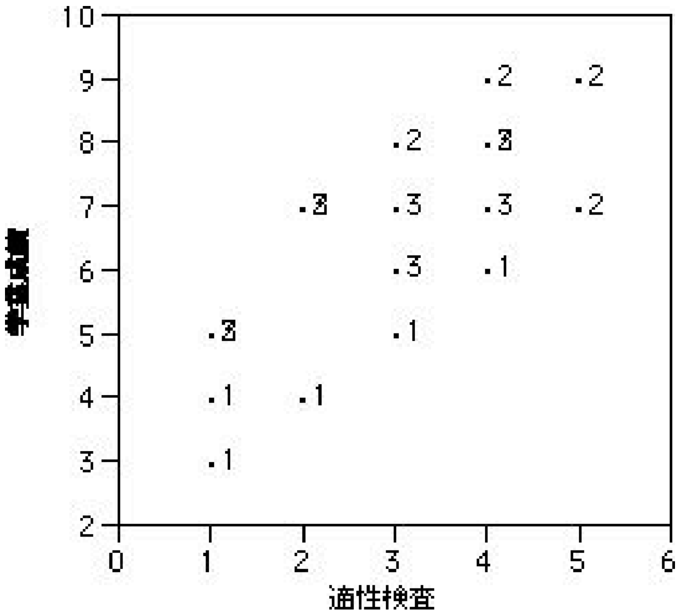

図 7.2: 3 種の訓練法と学業成績

これを訓練法別に表すと次のようになる．

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x + e \\
y_2 &= (\beta_0 + \beta_2) + \beta_1 x + e \\
y_3 &= (\beta_0 + \beta_3) + \beta_1 x + e
\end{aligned}
$$

したがって，上記のダミー変数を組み込んだ重回帰分析を行うと，3 種の訓練間で傾きが同じ，切片が異なる回帰式が得られることになる．通常，統計ソフトでは帰無仮説（$\beta = 0$）の検定が行われるので，例えば，$\beta_2 > 0$ かつ帰無仮説が棄却されれば，訓練法 2 は訓練法 1 に比べ学業成績を高める効果があると結論づけられる．

修正平均値

補助変数の値が補助変数の全体平均 $\bar{x}$ に等しい場合の基準変数の予測値である．補助変数の違いが補正された基準変数の平均値であると理解してよい．基準変数の修正平均値は上で求めた回帰係数 β_1 を使い以下

で求められる．

$$
\begin{aligned}
y_1^* &= \bar{y}_1 + \beta_1(\bar{x} - \bar{x}_1) \\
y_2^* &= \bar{y}_2 + \beta_1(\bar{x} - \bar{x}_2) \\
y_3^* &= \bar{y}_3 + \beta_1(\bar{x} - \bar{x}_3)
\end{aligned}
$$

7.7 主成分分析

質問紙を使うと 1 人の被験者から多くの変数に関するデータが得られる．多くの変数に関するデータが得られることはありがたいのだが，例えば 20 個の変数を得た場合には，その 20 個をそれぞれ検討しなければならないのだろうか．それはそれで意義のあることだが，20 個の変数に含まれる情報をぎゅっと集約して表現することも重要である．そのことによって，個々の変数ごとにデータを眺めるだけでは分からないような情報を手にすることもできる．主成分分析は多数の変数から少数の合成変数を作り出すための手法である．

7.7.1 合成変数

合成変数とは何だろうか．大学入試では，センター試験の得点と二次試験の得点とをそのまま足して合否が決まる場合もあるが，センター試験の数学や英語の得点を 2 倍したり社会や理科を 0.75 倍したりした上で二次試験の得点に加え個人の成績とする場合もある．

$$
\text{入試得点} = 2 \times \text{数学} + 2 \times \text{英語} + \text{国語} + 0.75 \times \text{社会} + 0.75 \times \text{理科}
$$

このとき入試得点は数学や英語など各科目（変数）の「線形結合 **(linear combination)**」による合成変数であり，各科目の得点に掛けられている数値（2,0.75 など）を重み係数という．重み係数が大きい科目は入試得点に大きな影響を与える．あるいは，入試得点は重み係数が大きい科目の成績をより多く反映しているとも言える．

7.7.2 主成分

主成分とは，複数の変数の線形結合である．

「A 君は身長○○ cm，体重××kg」というように，個人の特徴が身長と体重という 2 つの変数で表現されるとする．健康診断などではもっと多くの変数で個人が特徴づけられる．20 項目からなる質問紙を実施した場合では，個々の回答者はその 20 個の変数によって特徴づけられる．

さて，図 7.3 は身長・体重データの散布図である．図からは，身長が高くなると体重も重くなるという直線的な関係が読みとれる．このことを考慮に入れると，例えば A 君の身長が 185cm であれば，体重は 60kg 以上あるだろう，ということぐらいは想像できる．身長が 185cm で体重が 25kg の人もいれば 80kg の人もいるのであれば，個人を特徴づけるものとして体重データも必要となるだろうが，図 5.3 のような状況では体重に関する情報は少々「冗長」であるとも言える．

そこで，身長の軸と体重の軸という 2 つの軸上の数値を用いる代わりに，それらの軸の関係を反映した図中斜めの線を用いることにするわけである．この新しい軸では，身長も体重も大きい個人が右上に，身長も体重も小さい人は左下に位置づけられる．このことから，この新しい軸は「体の大きさ」に関する軸であると言うことができるだろう．A 君が「身長 185cm，体重 85kg」であれば，この斜めの軸上の右上の方に位置づけられ，「体の大きさ得点」が高い人として A 君は特徴づけられることになる．「体の大きさ」は身長と体重の 2 つの変数から作り出された合成変数である．そしてこのように，もとの変数軸間に見いだされ

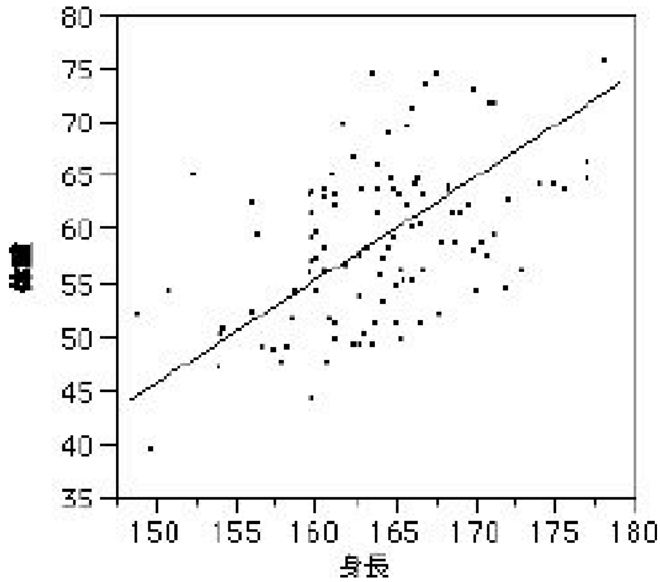

図 7.3: 身長と体重データの散布図

る新しい軸を主成分（**principal component**）という．

注：この説明は少々（かなり？）乱暴である．

7.7.3 モデル

n 人の個人から p 個の変数に関するデータが得られたとする．個人 i の j 番目の変数に関するデータを x_{ij} とすると，主成分分析のモデルは次の通りである．

$$\begin{aligned}
z_{1i} &= w_{11}x_{i1} + w_{12}x_{i2} + \cdots + w_{1j}x_{ij} + \cdots + w_{1p}x_{ip} \\
z_{2i} &= w_{21}x_{i1} + w_{22}x_{i2} + \cdots + w_{2j}x_{ij} + \cdots + w_{2p}x_{ip} \\
&\vdots \\
z_{ki} &= w_{k1}x_{i1} + w_{k2}x_{i2} + \cdots + w_{kj}x_{ij} + \cdots + w_{kp}x_{ip} \\
&\vdots \\
z_{pi} &= w_{p1}x_{i1} + w_{p2}x_{i2} + \cdots + w_{pj}x_{ij} + \cdots + w_{pp}x_{ip}
\end{aligned}$$

ただし，

z_{ki} ： 個人 i の第 k 主成分得点

w_{kj} ： 第 k 主成分を構成する変数 j の重み係数

また，

$$\sum_j w_j^2 = 1 \cdots \text{各主成分の重み係数の 2 乗和は 1}$$

p 個の変数があるとき，主成分は最大 p 個まで求めることができる（p 個の主成分が得られない場合もある）．しかし，p 個の主成分をすべて求めることはバカげている．p 個の変数をより少ない主成分で説明すること（多くの変数に含まれている情報を集約して表現すること）が主成分分析の目的だからである．

先程の身長・体重データに胸囲データを加え主成分分析を適用し，第 2 主成分まで求めると次の通りとなる．

$$z_1 = .425(w_{11}) \times \text{身長}\,(x_1) + .668(w_{12}) \times \text{体重}\,(x_2) + .610(w_{13}) \times \text{胸囲}\,(x_3) \cdots \text{寄与率 } 68.8\%$$

$$z_2 = .873(w_{21}) \times \text{身長}\,(x_1) - .124(w_{22}) \times \text{体重}\,(x_2) - .473(w_{23}) \times \text{胸囲}\,(x_3) \cdots \text{寄与率 } 27.1\%\text{（累積寄与率 } 95.9\%\text{）}$$

ここから以下を読みとる．

- 第 1 主成分はデータ全体の 68.8%を説明する
- 第 1 主成分では身長，体重，胸囲ともに重み係数が正であり，また体重と胸囲の重み係数が身長に比べて大きいことから，第 1 主成分は「体の大きさ」と解釈できる．
- 各変数の第 1 主成分への寄与を求めると，身長 18.1%，体重 44.6%，胸囲 37.2%である．
- 第 2 主成分はデータ全体の 27.1%を説明する．
- 第 2 主成分では身長の重み係数が正であり.873 と大きい．一方，体重，胸囲の係数が負であり，特に胸囲の係数が大きい．このことから第 2 主成分は「体の細長さ」と解釈できる．
- 各変数の第 2 主成分への寄与を求めると，身長 76.2%，体重 1.5%，胸囲 22.4%である．
- 第 2 主成分までの累積寄与率は 95.9%であり，このデータは以上の 2 つの主成分でほとんど全て説明できる．

7.7.4 主成分の求め方

主成分分析の理解のために主成分の求め方について簡単に説明しておく．結果の適切な解釈のためには必要なことだろう．

図 5.3 を見るとデータは散らばっている．その散らばりの大きい方向に斜めの線が引かれている．これが主成分分析の原理である．2 変量ではデータは平面上に散らばる．3 変量では立体的に散らばる．4 変量以上になると視覚的には表現できないが，p 変量では p 次元の空間にデータが散らばる．

与えられた空間内でのデータの散らばりの最も大きい方向に線を引く．これが第 1 主成分である．次に散らばりの大きい方向に第 2 主成分が求められる．ただし第 2 主成分は第 1 主成分と直角に交わるように決められる．したがって主成分間の相関を論じることはできない（主成分得点間の相関は常に 0 であるので論じようがない）．さらに散らばりの大きい方向を探して第 3 主成分が決まる．以下，同様にして p 変量の場合だと第 p 主成分まで得られる．

7.7.5 データに関する注意

主成分分析の解（重み係数）は，変数間の相関行列（共分散行列）の固有値分解によって得られる．つまり，変数間の関係が実際にはデータとなるのである．したがって，主成分分析の結果は，変数間の関係がうまくとらえられているかどうかにかかっていると言える．このことは，因子分析（次節）でも同じである．

ところで，心理学の領域で得られるデータ，特に主成分分析や因子分析の対象となるデータには，質問に対する 3〜7 段階での回答が多い．こうした目の粗いデータでは，変量間の相関係数がうまく求められない（薄められる）という問題が起こる．

表 5.4 には回答カテゴリ数を減らしたときに相関係数が小さくなる様子が示されている．まず，カテゴリ数 11 を真の値とする．このとき，2 つの変量 x と y の相関係数は r=.854 である．そして，これを 5 カテゴリにまとめると r=.707，3 カテゴリでは r=.553 となる．実際には，我々が想定する構成概念の量は 11 カテゴリどころか連続値であろうから，このような相関係数の希薄化は必ず起こっていることが予想される．

表 7.9: カテゴリ数を減らすと相関係数は小さくなる

11 カテゴリ		5 カテゴリ		3 カテゴリ	
x	y	x	y	x	y
11	10	5	5	3	3
10	9	5	4	3	3
9	7	4	3	3	2
8	7	4	3	2	2
7	9	3	4	2	3
6	4	3	2	2	2
5	6	3	3	2	2
4	3	2	2	2	1
3	4	2	2	1	2
2	5	1	3	1	2
1	3	1	2	1	1
r=.854		r=.707		r=.553	

ここで示したように，データ解析の結果は，どのようにデータがとられたかに大きく依存する．データ解析の結果が，ある心理現象のほんとうのところを示しているかどうかについては，十分な検討が必要であるし，解析結果の解釈も慎重に行わなければならない．思うような結果が得られない場合は，より精度の高い方法でデータを取り直してみることも一つの手である．

主成分分析の仲間

以下ははいずれも，質的データの主成分分析である．主成分分析と同様，多変量を少数の主成分で説明しようとするものである．また，クラスター分析などと併用し，サンプルを類型化することも多い．

- 数量化 III 類
- 対応分析 (corespondense analysis)
- 双対尺度法 (dual scaling)

7.8 因子分析

身長・体重・胸囲データについての主成分分析を振り返る．主成分分析の結果，これら 3 つの変量は「体の大きさ」「体の細長さ」という以下の 2 つの主成分にまとめあげることができた．

$$\text{体の大きさ} = .425\,\text{身長} + .668\,\text{体重} + .610\,\text{胸囲}$$
$$\text{体の細長さ} = .873\,\text{身長} - .124\,\text{体重} - .473\,\text{胸囲}$$

3 つの変量をうまく足し合わせることによってより少数の 2 つの合成変数を作ったのである．主成分は変数の合成の結果である．

これに対し因子分析では，身長，体重，胸囲のデータをもたらす原因を考えるのである．個々のサンプルは内在的に「体の大きさ」「体の細長さ」という因子をもち，その因子の影響によって身長や体重，胸囲のデータが決まると考えるのである．これを式で表すと次のようになる．因子分析の理解のポイントは観測さ

れる変量の式中の位置である．主成分分析では観測値は式の右辺にあり，主成分を説明する形となっているのに対し，因子分析では観測値が式の左辺にあり，因子によって説明される形となっている点に着目しよう．

$$
\begin{aligned}
\text{身長} &= .425\,\text{体の大きさ} + .873\,\text{体の細長さ} \\
\text{体重} &= .668\,\text{体の大きさ} - .124\,\text{体の細長さ} \\
\text{胸囲} &= .610\,\text{体の大きさ} - .473\,\text{体の細長さ}
\end{aligned}
$$

この式の意味するところは，

- 身長には「体の大きさ」因子と「体の細長さ」因子が正の方向に影響する
- 体重は「体の大きさ」因子から正の方向の影響を受けるが,「体の細長さ」因子からの影響はあまりない
- 胸囲は「体の大きさ」因子から正の方向の影響を受け,「体の細長さ」因子からは負の方向の影響を比較的強く受ける

「体の大きさ」因子を「体の大きさ」を規定する遺伝的要因と考えれば理解しやすいだろう．

上の式からは「体の大きさ」因子が身長にも体重にも胸囲にも正の方向で影響していることが見てとれる．このように１つの因子が複数の変量に影響を与える形になっていると，その変量間には共変関係が生じる.「体の大きさ」因子が大きい人は身長も体重も胸囲も大きくなり，逆にこの因子が小さい人は身長も体重も胸囲も小さくなる．したがって，身長が高ければ体重も重く胸囲も大きい，身長が低ければ体重も軽く胸囲も小さい，というようなことが起こるのである．なお，個々のサンプルにおける因子の量を因子得点という．この例で言うと「体の大きさ」因子得点が大きいサンプルは身長，体重，胸囲が大きくなる．

また,「体の大きさ」因子も「体の細長さ」因子も３つの変量すべてに共通して影響を与えるものとなっている．こうした因子を共通因子という．これに対し，身長，体重，胸囲のそれぞれに独自に影響を与える要因というものも考えられる．そのような各変量に独自に関係する要因を独自因子とよぶ．さらに，共通因子の直前の数値は，その共通因子がどの程度変量に影響を与えているかの指標である．これを因子負荷量という．重回帰分析の偏回帰係数と同じと考えて良い．

7.8.1 因子：観測されることのない変数（潜在変数）

因子分析の式の右辺には観測されるデータがない．観測される変量はすべて式の左辺，すなわち説明される側にある.「体の大きさ」「体の細長さ」という因子は直接観測されるものではない（このような変数を「潜在変数」という．共通因子も独自因子も潜在変数である）．何もないところからどうやって身長や体重や胸囲を説明する因子というものを求めるのか．

先に，複数の変量に共通して影響する因子は観測値間の共変関係をもたらすと述べた．これを利用するのである．共通因子が各変量間の共変関係を作り出すのであるから，逆に各変量間の共変関係に基づいて，それをもたらす共通因子及び因子負荷量を求めるのである．主成分分析と同様に，因子分析でも変量間の相関係数（分散共分散）が直接のデータとなる．

心理学では，知能や態度，動機，意欲，関心，感情など直接観測できない構成概念を興味・関心の対象とする場合がほとんどである．そうした対象であっても，研究・考察の対象とするからには何とか測定しなければならない．そこで，具体的に得られる観測値の背後にその観測値をもたらす構成概念を想定することで，目的の構成概念を扱おうとする．そうして生まれた手法が因子分析である．心理学研究とは非常に密接な関係をもつ解析手法である．

7.8.2 モデル

8個の項目があるとする．これら8個の項目への個人iの反応が，その個人iの3種の潜在的な特性の数量（f_{1i}, f_{2i}, f_{3i}）の重み（a, b, c）付け和で決まるとき，これは次のように表現される．

$$\begin{aligned}
x_{1i} &= a_1 f_{1i} + b_1 f_{2i} + c_1 f_{3i} + u_1 v_{1i} \\
x_{2i} &= a_2 f_{1i} + b_2 f_{2i} + c_2 f_{3i} + u_2 v_{2i} \\
x_{3i} &= a_3 f_{1i} + b_3 f_{2i} + c_3 f_{3i} + u_3 v_{3i} \\
x_{4i} &= a_4 f_{1i} + b_4 f_{2i} + c_4 f_{3i} + u_4 v_{4i} \\
x_{5i} &= a_5 f_{1i} + b_5 f_{2i} + c_5 f_{3i} + u_5 v_{5i} \\
x_{6i} &= a_6 f_{1i} + b_6 f_{2i} + c_6 f_{3i} + u_6 v_{6i} \\
x_{7i} &= a_7 f_{1i} + b_7 f_{2i} + c_7 f_{3i} + u_7 v_{7i} \\
x_{8i} &= a_8 f_{1i} + b_8 f_{2i} + c_8 f_{3i} + u_8 v_{8i}
\end{aligned}$$

ここで，

- x_{1i} ： 個人iの項目1への反応
- f_{1i} ： 個人iの潜在特性 f_1 の数量＝因子得点
- a_1 ： 項目1への反応に対し潜在特性（因子）f_1 がどの程度効いているか＝因子負荷量
- v_{1i} ： 個人iの項目1独自の因子（潜在特性 f_1, f_2, f_3 では説明できない部分）＝独自因子
- u_1 ： 項目1への反応に対し独自因子がどの程度効いているか

通常，個人は区別されず，モデルはもっと一般的に，項目j（$j = 1, 2, \cdots, m$）はq個の因子（潜在的な特性：$f_1, f_2, \cdots, f_k, \cdots, f_q$）の各項目ごとの重み（$a_{jk}$）付け和で以下のように表現される．

$$x_j = a_{j1} f_1 + a_{j2} f_2 + \cdots + a_{jk} f_k + \cdots + a_{jq} f_q + u_j v_j$$

一般的な仮定このモデルはこのままではどうしようもなく，一般的にはまず以下の仮定がおかれる．

- 全ての変数 x, f, v の平均（期待値）を0とする（標準化されているということ）
- 共通因子 f 及び独自因子 v の分散を1とする
- 共通因子 $f_1, \cdots, f_q$ と独自因子 v は直交する（相関がない）
- 異なる独自因子 $v_1, \cdots, v_j, \cdots, v_m$ は互いに直交する（相関がない）

7.8.3 直交モデルと斜交モデル

因子分析モデルは直交モデルと斜交モデル分けられる．直交モデルでは以下の仮定が加わる．また，この仮定をおかないのが斜交モデルである．

- 異なる共通因子 $f_1, \cdots, f_q$ は互いに直交する（相関がない）

この制約のもとで得られた解を直交解という．直交解を求めた後，因子間の相関を論じることは，この仮定と矛盾するので気をつけなければならない．一方，この制約を置かずに得られた解を斜交解という．

7.8.4 因子パターンと因子構造

因子パターンとは因子負荷量行列を差し，因子構造は変数と因子との相関係数行列を指す．直交モデルでは両者が一致するが，斜交解及び斜交回転（すなわち因子間相関を想定した場合）には両者が一致しないので注意を要する．なお，結果の解釈には因子パターンを用いるのが一般的である．

7.8.5 単純構造

単純構造とは，ある項目が特定の 1 つの因子によってのみ専ら説明されるような因子負荷量のパターンが得られた状態を指す．例えば，上記 8 項目の例で言えば，次のような状態である．

$$
\begin{aligned}
x_1 &= a_1 f_1 + 0 f_2 + 0 f_3 + u_1 v_1 \\
x_2 &= a_2 f_1 + 0 f_2 + 0 f_3 + u_2 v_2 \\
x_3 &= a_3 f_1 + 0 f_2 + 0 f_3 + u_3 v_3 \\
x_4 &= 0 f_1 + b_4 f_2 + 0 f_3 + u_4 v_4 \\
x_5 &= 0 f_1 + b_5 f_2 + 0 f_3 + u_5 v_5 \\
x_6 &= 0 f_1 + b_6 f_2 + 0 f_3 + u_6 v_6 \\
x_7 &= 0 f_1 + 0 f_2 + c_7 f_3 + u_7 v_7 \\
x_8 &= 0 f_1 + 0 f_2 + c_8 f_3 + u_8 v_8
\end{aligned}
$$

これは，項目 1〜3 への反応は因子 1，項目 4〜6 では因子 2，項目 7，8 は因子 3 の数量によってほぼ決まることを意味している（共通性は大きいものとする）．そして，通常，各項目の内容から f_1 は○○性因子，f_2 は××性因子... と名付けられる．

サーストンによる単純構造の基準

- どの変数も共通因子の総数よりも少ない因子で説明される
- 因子パターン行列の各列は少なくとも共通因子数分個の 0 の要素をもつ
- 2 つの列を比較するとき，いくつかの変数はいずれか一方の列に負荷を持ち，片方は 0 である
- 共通因子数が 4 以上のとき，2 つの列の因子負荷量が 0 である変数がかなり多い

7.8.6 尺度構成と尺度の一次元性と α 係数

○○性や××度といった心理学的構成概念の量を測定したい場合，複数の項目からなる質問紙を用いることが多い．

先の例で言うと，項目 1〜3 は潜在的な特性 f_1（○○性）を（間接的に）測定する尺度として用いることができる．直接観測できる項目への反応を用い，○○性という特性を測定したとみなすのである．同様に，項目 4〜6 は因子 2 に関わる尺度，項目 7，8 は因子 3 に関わる尺度として利用できる．したがって，因子分析の結果は単純構造をもつことが非常に望ましい．

ところで，上の例における項目 1〜3 における因子負荷量が次のようなものだとする．

$$
\begin{aligned}
x_1 &= 0.5 f_1 + 0.5 f_2 + 0.5 f_3 + u_1 v_1 \\
x_2 &= 0.9 f_1 + 0.0 f_2 + 0.0 f_3 + u_2 v_2 \\
x_3 &= 0.9 f_1 + 0.0 f_2 + 0.0 f_3 + u_3 v_3
\end{aligned}
$$

このとき，項目１は，何を測っているのか，という問題が生じる．

尺度構成の場面では，測定したい何かのみを測る項目群を用意しなければならない．このとき用意された項目群に求められるのが一次元性（内部一貫性）である．用意された複数の項目は，興味の対象である「一つの」何かを測定しなければならないということである．

複数の項目からなるある尺度の一次元性の指標としてよく用いられるものにα係数がある．α係数の定義は次の通りであり，この値が１に近ければその項目群（尺度）は一次元性を保っている，つまり測りたい１つの何かを測っている，と言える．

n 個の項目からなる尺度（テスト）があるとする．この尺度の合計点をX，項目 j の得点をX_jとすると，

$$\alpha \equiv \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum_j \sigma^2(X_j)}{\sigma^2(X)}\right)$$

7.8.7 因子の回転

論文などを読むと「バリマックス回転」という言葉に出会うことが多いだろう.「回転とは何だろう」と思っている人も多いのではないだろうか．因子分析で「解を求める」とは因子負荷量を求めることを指すのが一般的である（あるいは因子を抽出するなどともいう）．ところが因子分析には，解は一意に定まらない「解の不定性」という性質がある．この性質を逆に利用するのが因子の回転である．

因子分析の解を求めると，その結果は単純構造からはほど遠い場合も少なくない．変量がいくつもの因子に負荷しているような因子負荷パターンが得られると，因子の解釈が非常に困難になる．そこで単純構造や特定の因子パターンを目指して，得られた解（因子負荷量）を変換するのである．これが因子の回転である．

回転の方法は，

- 直交回転（因子間の直交性の制約あり）か斜交回転（直交性の制約を外す）か
- ターゲットがあるかないか

によって分類できる．

オーソマックス回転：直交回転，ターゲットなし

- コーティマックス回転
- バリマックス回転（基準バリマックス法）

直交なので因子間相関は論じないこと．

オブリミン回転：斜交回転，ターゲットなし
因子間相関を論じることができる

プロクラステス回転：直交回転／社交界点，ターゲットあり

- プロクラステス直交回転
- 斜交プロクラステス回転
- 斜交プロマックス回転

単純構造を目指すのではなく，特定の因子負荷パターンを設定し，そのパターンにできるだけ近づけるという基準で回転を行う．

7.8.8 変数間の相関と因子負荷量

直交解では，因子負荷量は当該因子と変数との相関係数となる．このことを知っておくと解釈のためにも役立つ．また，求められた因子負荷量から，変数間の相関係数を再現することもできる．変数 x_j と x_k の相関係数 r_{jk} は

$$r_{jk} = \sum_{l=1}^{q} a_{jl} a_{kl}$$

で得られる．これは因子分析の基本式と呼ばれ，因子分析の考え方がここに凝縮されている．

7.8.9 共通性

変数が基準化（平均 0，分散 1）されているとき，以下が成り立つ．

$$x_j \text{の分散} = a_{j1}^2 + \cdots + a_{jk}^2 + \cdots + a_{jq}^2 + u_j^2 = 1$$

このとき，

$$h_j^2 = a_{j1}^2 + \cdots + a_{jk}^2 + \cdots + a_{jq}^2$$

を項目 j の共通性という．これはその変数の分散を共通因子でどの程度説明できるかを表す．共通性が小さいということは，その変数は共通因子以外の要因（＝独自因子）によって説明される割合が大きいことを意味する．つまり，データを「因子」で説明することに意味がないということである．
なお，変数の分散（=1）から共通性を引いた残り（$1 - h^2$）を特殊性という．

より一般的な共通性の定義は以下の通り．

$$t_j = a_{j1} f_1 + a_{j2} f_2 + \cdots + a_{jk} f_k + \cdots + a_{jq} f_q$$

とすると，

$$x_j \text{の分散} = t_j \text{の分散} + u_j v_j \text{の分散}$$

上式中，t_jの分散 は，変数の分散のうち共通因子で説明できる部分であり，

$$h_j^2 = \frac{t_j \text{の分散}}{x_j \text{の分散}} = \text{共通性}$$

である．

7.8.10 寄与率

m 個の各変数それぞれにおける共通性を $h_1^2, \cdots, h_j^2 \cdots, h_m^2$ とする．
また，m 個の各変数それぞれにおける特殊性を $u_1^2, \cdots, u_j^2 \cdots, u_m^2$ とすると，

$$\text{変数全体の分散} = \sum_{j=1}^{m} h_j^2 + \sum_{j=1}^{m} u_j^2$$

である．つまり，変数全体の分散は，共通性の和と特殊性の和から成るということである．ところで，共通性は各変数の分散のうち共通因子で説明できる部分であった．このことから，共通性の和は，変数全体の分散のうち，共通因子で説明できる部分，すなわち因子分析モデルで説明できる部分を表すことが分かる．

因子寄与率とは，それぞれの因子が，共通性全体のどの程度を占めているかを表す指標である．変数 j の第 k 因子の因子負荷量を a_{jk} とすると，寄与率は次のように定義される．

$$\begin{aligned}\text{第 k 因子の寄与率} &= \frac{\text{第 k 因子に関する因子負荷量の平方の総和}}{\text{共通性の総和}} \\ &= \frac{\sum_{j=1}^{m} a_{jk}^2}{\text{共通性の総和}}\end{aligned}$$

斜交因子の場合の寄与率
一意に定義されない．いくつかの考え方がある．その 1 つに，重回帰分析の場合と同様に考える考え方がある．因子分析ではデータ x_j が因子によって説明される，というモデルである．これは重回帰分析と同じモデルである．したがって，特定の因子の寄与率を知りたいときは，まず，全ての因子を用いた場合の重相関係数の平方（決定係数）に相当する量を求め，その後，任意の因子（説明変数）を取り除いた場合の重相関係数の平方を計算する．そして，その差を見れば，当該因子の基準変数（データ）の説明に対する寄与の程度が分かる．

7.8.11 因子の抽出法：推定方法のいろいろとその特徴

主因子法

- 反復なし主因子法
  - 共通性の推定値=1 とした場合，主成分解と一致する
  - 初期値の影響を受けやすい
- 反復主因子法
  - 共通性の推定→固有方程式を解く→共通性の推定，の反復
  - 初期値の影響を受けやすい
  - 不適解がみつかりにくい
  - ちゃんと推定できていれば最小 2 乗解と一致する

最小 2 乗法

- 因子負荷量の 2 乗和の最大化という原理に基づき計算する．
- 不適解を見いだしやすい
- 因子数の決定時にモデルとデータとの適合度指標を使える

最尤法

- 観測変数の母集団分布に多変量正規分布を仮定し，尤度を最大化させる
- 不適解を見いだしやすい
- 因子数の決定時にモデルとデータとの適合度指標を使える

7.8.12 標本数

1. 解を求めるために用いる共通性の初期値（SMC と呼ばれる数量を用いることが多い）を得るには，標本数は変数の数よりも大きい必要がある
2. 不適解を避けるためには,（予想される）各因子ごとに少なくとも 3 つ以上の項目を用意し，さらに大きな標本数を用意する必要がある
3. 尤度比検定やその他適合度基準を用いたい場合，推定する母数の数の 5〜10 倍の標本数を用意する必要がある（1 つの目安として）

7.8.13 不適解

因子分析モデルでは，解を求めるのは実は非常に難しく．時に（結構）不適解が生じる．独自因子の分散が負になる（＝共通性が 1 を超える）Heywood case と呼ばれる不適解はしばしば生じる．不適解に関して気をつけなければならないのは，統計パッケージでは，何らかの形でその対処をしている，という点である．また主因子法では，本来不適解が生じるケースにも関わらず，解の推定が途中でうち切られる場合がある．おそらく多くの場合，不適解が生じていても，ユーザーはそのことに気がつかず（知らされず），結果を用いて何らかの主張を行うことになる．

不適解は，十分な標本数，事前の項目分析等で避けることができる．また，因子数を増やすと不適解が生じやすくなる．なお，ベイズ推定という推定法では不適解が生じないという利点があるが，一般的ではなく，おそらく自分で推定プログラムを作らなければならない．

7.8.14 因子数の決定

- 標本相関係数行列の固有値のうち 1 より大きいものの数
- スクリープロットの結果，急激に減少する 1 つ手前までを因子の数とする
- 尤度比検定（観測変数に多変量正規分布を仮定する）：m 因子モデルを帰無仮説とし，m を 1 から順に増やしていき，帰無仮説が棄却されない最小の m を因子数として採択する．
- 情報量基準による決定（観測変数に多変量正規分布を仮定する）：尤度を用いたある数量を指標とする
- ワオクライテリオン：因子数を変えながら解を求め，その度に残差行列を確認する．すると突然残差の絶対値が小さくなる．その直前の因子数を採用すればよい．

7.8.15 付録 1：主因子法

標本相関行列の対角要素に共通性を入れた行列を固有値分解したもの．しかし，共通性の本当の値は知ることはできないので，その推定値を用いる．

$$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D} \cdots \text{因子分析モデルの行列表現}$$

ここで,

$\boldsymbol{X}$: データ行列（N × n）

$\boldsymbol{F}$: 因子得点行列（N × m）

$\boldsymbol{A}$: 因子負荷量行列（n × m）

$\boldsymbol{U}$: 独自因子得点行列（N × n）

$$\begin{array}{rcl} \boldsymbol{D} & : & \text{独自因子負荷を対角要素に持つ行列（n × n）} \\ N & : & \text{データ数} \\ n & : & \text{変数の個数} \\ m & : & \text{共通因子の個数} \end{array}$$

さらに，因子分析の基本式から，

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{AA'} + \boldsymbol{D}^2 \cdots \text{直交解（理解のポイント）} \\ \boldsymbol{R} &= \boldsymbol{ALA'} + \boldsymbol{D}^2 \cdots \text{斜交解（理解のポイント）} \end{aligned}$$

ここで，

$$\begin{array}{rcl} \boldsymbol{R} & : & \text{標本相関行列} \\ \boldsymbol{L} & : & \text{因子間相関行列} \end{array}$$

直交解について，標本相関行列の対角要素に共通性（1-独自負荷の２乗）を入れた行列 $\boldsymbol{R}^+$ を考えると，$\boldsymbol{A}$ は $\boldsymbol{R}^+$ を固有値分解したときの固有ベクトル行列として得られる（互いに直交する複数のベクトルで表現される，ということ）．すなわち，

$$\begin{aligned} \boldsymbol{R}^+ &= \boldsymbol{R} - \boldsymbol{D}^2 = \boldsymbol{AA'} \\ \boldsymbol{R}^+\boldsymbol{A} &= \boldsymbol{A}\Lambda \end{aligned}$$

ここで

$$\Lambda \quad : \quad \boldsymbol{R}^+ \text{の固有値を成分とする対角行列}$$

7.8.16 付録２：因子の回転

因子分析モデルには,「解の不定性」という性質がある（以下は直交モデルの場合）.

$$\begin{aligned} \boldsymbol{x} &= \Lambda\boldsymbol{f} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{f}) &= 0 : \text{因子の期待値は} 0 \\ E(\varepsilon) &= 0 : \text{誤差の期待値は} 0 \\ V(\boldsymbol{f}) &= \boldsymbol{I} : \text{因子間の相関はなし} \end{aligned}$$

において，因子数を m とし m 次の直交行列 $\boldsymbol{T}$ を用いて

$$\boldsymbol{g} = \boldsymbol{T'f}$$

と変換する（因子 $\boldsymbol{f}$ の直交回転）．このとき $\boldsymbol{g}$ は

$$\begin{aligned} E(\boldsymbol{g}) &= 0 : \text{因子の期待値は} 0 \\ V(\boldsymbol{g}) &= \boldsymbol{I}_m : \text{因子間の相関はなし} \end{aligned}$$

を満たす．また，

$$\Lambda_g = \Lambda\boldsymbol{T}$$

という因子負荷量の変換を行うと，

$$\begin{aligned} \boldsymbol{x} &= \Lambda_g\boldsymbol{g} + \varepsilon \\ E(\boldsymbol{g}) &= 0 : \text{因子の期待値は} 0 \\ V(\boldsymbol{g}) &= \boldsymbol{I}_m : \text{因子間の相関はなし} \end{aligned}$$

となる．

これは直感的に言えば，どのように軸をとっても相対的に位置関係が同じであればよい，ということである．したがって，結果の解釈（軸の解釈）を行いやすいように，因子を回転する．因子負荷量行列が先に述べた「単純構造」を持つ場合，各因子の解釈は容易となる．

7.9 共分散構造分析

共分散構造分析は因子分析と密接な関連がある．通常の因子分析では，データから探索的に因子を抽出する．これを探索的因子分析という．これに対し，構成概念上の理論から先に因子に関するモデルを構成し，データによってそれを確かめるという因子分析のやり方もある．これを確認的因子分析という．確認的因子分析は，共分散構造分析の一種である．

7.9.1 観測変数と潜在変数を含む因果モデル

観測値が得られる変数を観測変数，直接観測されない変数（因子分析での因子がそうである）を潜在変数という．共分散構造分析では，複数の観測変数や潜在変数間の因果関係をモデル化する．そして，

- データに基づき変数間の因果関係の強さ（程度）を求める
- データに最もフィットするモデルを探す
- データを通して既存の理論上のモデルを検証する

などを行う．我々が目にする共分散構造分析モデルは，通常パス図によって表現される．また，共分散構造分析の解析結果の表現はパス解析の結果と非常によく似ている．パス図内に bf 潜在変数が含まれるということがパス解析とは異なる点である．

7.9.2 測定方程式と構造方程式

共分散構造分析における数式上の因果モデルは「測定方程式」と「構造方程式」によって表現される．観測変数及び潜在変数間の因果関係には次の 4 通りが考えられるが，

- 潜在変数が観測変数の原因となる
- 潜在変数が他の潜在変数の原因になる
- 観測変数が他の観測変数の原因になる
- 観測変数が潜在変数の原因になる

この内,「潜在変数が観測変数の原因となる」を表現する方程式が測定方程式である．これは原因としての潜在変数が複数個の観測変数に影響を与えている様子を記述するための方程式であり，その潜在変数が観測変数を用い，どのように測定されているかを記述する方程式であるとも言える．数式的に表現すると，

$$\text{観測変数} = \text{因果係数}\,1 \times \text{潜在変数（因子）}\,1 + \cdots + \text{因果係数}\,\mathrm{m} \times \text{潜在変数（因子）}\,\mathrm{m} + \text{誤差}$$

となる．上で述べた因子分析モデルにおける方程式は測定方程式である．

7.9.3 外生変数と内生変数

モデル内で取り上げるどの変数からも影響を受けない変数を外生変数という．モデルの外で生まれる変数という意味である．これに対して，モデル内で他の変数によって説明される変数を内生変数という．モデル内で生まれる変数という意味である．

7.9.4 解析例：進路指導に関する調査データ（豊田他,1992）

調査項目：4 段階評定

x_1 「進学率の高い高校が良い高校」という考え方に賛成か

x_2 人物評価基準として学歴・学校歴の効用を支持するか

x_3 「偏差値の高い学校への進学が将来有利」という考え方に賛成か

x_4 校外模擬試験の偏差値結果をどの程度重視するか

x_5 模擬試験結果の合格可能性をどの程度重視するか

x_6 偏差値中心の進学指導の必要性をどの程度感じるか

x_7 模擬試験の結果を積極的に利用しているか

x_8 模擬試験の結果をどの程度信頼しているか

x_9 生徒は進路決定で受験産業情報にどの程度依存しているか

x_{10} 教師は進路指導で受験産業情報をどの程度利用しているか

データ
平成元年度における共通 1 次の志願者が 1 名以上あったすべての高等学校 3691 校中，回答のあった 2616 校のデータ（各校，1 名の進路指導担当教員が回答）.

因子分析：プロマックス解（斜交解）

- 因子 1：偏差値に対する信頼（x_4, x_5, x_6, x_7, x_8）
- 因子 2：受験産業への依存（x_9, x_{10}）
- 因子 3：学歴志向（x_1, x_2, x_3）

この結果から,「因子 3 →因子 1 →因子 2」という因子間の因果モデルを設定することにする.

モデル及び結果の図示（パス図）
図 5.4 は進路指導調査データに関する共分散構造分析の結果である．このように結果はパス図で示される．図中の楕円及び四角はそれぞれ潜在変数，観測変数を表す．変数間に引かれた→は変数間の因果関係を表す．また→に添えられた数値はその関係の程度を表すものであり，因果係数と呼ばれる．

ここで示されているように，因子分析では潜在変数を特定するだけであったが，共分散構造分析ではそれら潜在変数間の因果関係を検討できる．これから多く利用される解析手法となるだろう．

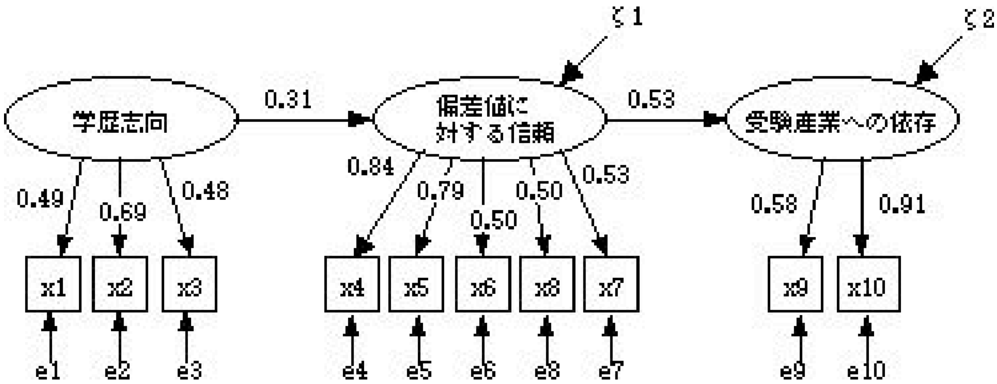

図 7.4: 進路指導に関する因果モデル

第8章　分割表の解析

8.1　χ^2 検定

χ^2 検定とは，χ^2 分布を用いた検定の総称である．非常に大雑把にいうと，χ^2 とは，理論上の値（通常，帰無仮説の元での期待値）からのズレの程度の分布である．

質的データの場合，そのデータの集約の仕方は「集計」である．各カテゴリへの反応の分布は集計表あるいはクロス集計表として表現される．観測された集計値は，必ずしも理論上の集計値とピッタリとは一致しない（一致することの方がまれである）．理論上の度数分布が正しくても，偶然誤差の範囲で観測値はばらつく．つまり理論上の集計値と観測値との間にはズレが生じる．このズレの大きさの分布が χ^2 分布なのである．

したがって，理論上の分布（帰無仮説の元での分布）が正しいならば $p < .01$ の程度でしか起こり得ないような大きさのズレが生じた場合，そのズレは偶然誤差によるものではなく,「有意」なズレであるとみなすのである．

8.1.1　一様性の検定・適合度の検定：理論値との一致を調べる

一様性の検定では，複数の事象の起こり方が一様であるか，それとも偏っているかの検定を行う．帰無仮説は「一様性が成り立っている」である．一方，適合度の検定では，観測値が理論値と一致しているかどうかの検定を行う．帰無仮説は「理論値が正しい」である．一様性の検定も一種の理論値との適合度の検定であると言え，両者に違いはない．以下に検定例を述べる．

ここに 1 つのサイコロがある．これをつかって賭をしようと，金に困っている友人が持ってきたものである．賭をしても構わないが，サイコロに何か細工でもしてあったら圧倒的に不利である．そこでサイコロを確認させてもらうことにした．一定の目が出やすくなっているかどうかを確かめるためである．全部で 60 回，サイコロを振ってみた．結果は表 7.1 の通りである．

表 8.1: 60 回サイコロを振った実際の結果と理想的な結果

目	1	2	3	4	5	6	計
観測値	13	11	8	8	9	11	60
理論値	10	10	10	10	10	10	60

何となく 1 の目がたくさん出たような気もするが,「たまたま」ということも十分ありえそうである．根拠もなく友人を疑うのもよくない．さて，何の細工もない理想的なサイコロの場合，1〜6 のどの目が出る確率も一様に 1/6 と考えることができる．すると，どの目も出る確率が 1/6 であるとするならば，理想的には表 7.1 中の「理論値」のようになるはずである．

では，観測値は理論値からどの程度ズレているのか．そのズレの程度を以下のように数量化する．

$$\text{ズレの大きさ} = \frac{(13-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10}$$

カテゴリ数を m，j 番目のカテゴリへの反応数を y_j，反応数の総和を Y とし，これを一般的に表現すると，

$$\text{ズレの大きさ} = \sum_{j=1}^{m} \frac{(y_j - E_j)^2}{E_j} = \chi^2$$

ただし，

E_j ： j 番目のカテゴリへの反応数の理論値（期待値）

となる．この χ^2 は，自由度 $m-1$ の χ^2 分布に従うことが分かっている．

なおこの例の場合，60 回の実験から得られた $\chi^2 = 2$ は，自由度 5 の χ^2 分布における $p = .01$ の臨界値 15.09 と比べ小さいので，ばらつきは偶然の範囲であると結論づけることになる（一様性の帰無仮説は棄却されない）．

8.1.2 独立性の検定：2 変数間の連関の有無の検討

やはりこれも理論値との適合度の検定である．ただし，独立性の検定では，変数が 2 つ（以上）となる．独立性の検定における理論値は，2 つ（以上）の変数が独立（関連がない）であるという仮説から得られる「期待値」である．

2 変数が独立であるということ

2 変数が独立であるということは，一方の変数の値が他方の変数の値に影響を与えないということである．例えば 2 つのサイコロを振るという事態を考える．2 つのサイコロをそれぞれ A,B とし，まず A を転がし，その後 B を転がすとする．何の細工も仕込みもなければ，A のサイコロの目と B のサイコロの目の出方は無関係である．A で 1 の目が出たからといって，B では 1 が出にくくなるということはない．A のサイコロで出る目を X，B のサイコロで出た目を Y とそれぞれ表すと，X の値は Y の値には影響しないということである．

さて，この場合，X の取りうる値は $1, 2, 3, 4, 5, 6$ のうちのどれかである．同様に Y の取りうる値も $1, 2, 3, 4, 5, 6$ のうちのどれかである．A のサイコロを振って x の目が出たことを $X = x$，その確率を $P(X = x)$ と表すことにする．同様に，B のサイコロを振って y の目が出たことを $Y = y$，その確率を $P(Y = y)$ と表すことにする．そして 2 つのサイコロを振った結果を $(X = x, Y = y)$，その確率を $P(X = x, Y = y)$ と表すことにする．このとき以下が成り立つ．

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

具体的に言うと，

$$\text{A,B ともに 1 が出る確率} = \text{A に 1 の目が出る確率} \times \text{B に 1 の目が出る確率} = \frac{1}{36}$$

ということである．このことは 1 の目に限らず，1〜6 のどの目についても成り立つ．このサイコロの例に限らず，$P(X = x, Y = y) = P(X = x)P(Y = y)$ が成り立つとき，2 つの変数 X, Y は独立であるという．逆に言えば，2 つの変数 X, Y が独立ならば，$P(X = x, Y = y) = P(X = x)P(Y = y)$ である．

期待値

2 つのサイコロを振るということを 360 回繰り返したとする．このとき，A,B ともに出た目が 1 となることは何回ぐらい起こるだろうか．

A の出る目と B の出る目は独立であると仮定できる．したがって，以下が成り立つ．

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

A,B ともに 1 の目が出るという出来事は確率 1/36 の出来事である．こうした出来事は，360 回のうち 10 回起こることが期待できる．この 10 回を期待値という．

$$10\,回 = 360\,回 \times \frac{1}{36}$$

独立の仮定に基づく分割表における期待値

さて別の例を考える．母集団から計 150 名を無作為に抽出したところ身体に何らかの異常を訴える者が 60 名，全く健康であるものが 90 名であった．またこの 150 名についてストレス尺度を実施したところ，ストレス値が高いと判定された者は 100 名，ストレス値に特に問題はないとされた者が 50 名であった．

表 8.2: 身体異常及びストレス値についてのそれぞれの集計値

身体異常	あり	なし	計
	60	90	150
ストレス値	高い	普通	計
	100	50	150

このことから，母集団における「身体異常あり」及び「ストレス値高い」の確率はそれぞれ $60/150 = 0.40$, $100/150 = 0.67$ とすることができる（最も合理的な判断である）．このとき，もし身体異常の有無とストレス値との間に何の関係もなければ，すなわち独立であれば，これらの組み合わせによる以下の出来事の確率は次のようになる．

$$\begin{aligned}
P(「身体異常あり」かつ「ストレス値高い」) &= P(身体異常あり)P(ストレス値高い) \\
P(「身体異常あり」かつ「ストレス値普通」) &= P(身体異常あり)P(ストレス値普通) \\
P(「身体異常なし」かつ「ストレス値高い」) &= P(身体異常なし)P(ストレス値高い) \\
P(「身体異常なし」かつ「ストレス値普通」) &= P(身体異常なし)P(ストレス値普通)
\end{aligned}$$

これを分割表で表現すれば次のようになる．

表 8.3: ストレス値と身体異常の有無に関する確率

	身体異常あり	身体異常なし	行和（周辺確率）
ストレス値高い	$60/150 \times 100/150$	$90/150 \times 100/150$	$100/150$
ストレス値普通	$60/150 \times 50/150$	$90/150 \times 50/150$	$50/150$
列和（周辺確率）	$60/150$	$90/150$	1

こうして，身体異常の有無とストレス値の高低の組み合わせの出来事への独立の仮定のもとの確率を用いると，150 名についての各出来事への人数の割り当てには以下が期待できる．

表 8.4: ストレス値の高低による身体異常の有無についての期待度数分布

	身体異常あり	身体異常なし	周辺度数
ストレス値高い	40	60	100
ストレス値普通	20	30	50
周辺度数	60	90	150

これを期待度数分布（度数分布の期待値）という．もし,「ストレス値」と「身体異常」という 2 つの変数が独立（関連がない）ならば，上の表のような人数の分布が期待されるのである．この期待度数分布が一様性・適合度の検定における「理論値」に相当する．

独立性の検定
さて，150 名の度数分布の観測値は表6.5 に示した通りであった．

表 8.5: ストレス値の高低による身体異常の有無についての観測値

	身体異常あり	身体異常なし	周辺度数
ストレス値高い	52	48	100
ストレス値普通	8	42	50
周辺度数	60	90	150

独立性の検定は，この観測値と表 6.4 における期待値とのズレに基づいて行われる．独立性の検定における帰無仮説は「**2 変数は独立である**」である．つまり，表6.4 の度数分布を理論上の分布であると考えるのである．あとは，適合度の検定と同様の手続きである．ズレの総和 χ^2 は次の通りである．

$$\chi^2 = \sum_i \sum_j \frac{(y_{ij} - E_{ij})^2}{E_{ij}} = 18$$

この $\chi^2 = 18$ は自由度 1 の χ^2 分布に従うことから，危険率を 1%とすると，$\chi^2_{1,\alpha=.01} = 6.63$ より，帰無仮説は棄却される．すなわち，ストレス値の高低と身体異常の有無は独立とはみなせず，両変数間には何らかの連関があると結論づけることになる．

注：自由度 1 の場合，計算される χ^2 値に対し「連続のための修正」がほどこされるが，ここでは行っていない（なお，統計ソフトでは自動的に修正する）．

8.1.2.1　下位検定：**Haberman** の方法

セルの観測度数 y_{ij} を帰無仮説（独立）のもとでの条件付き分布の期待値，標準偏差で規準化すると以下の規準化残差 e_{ij} が得られる．

$$e_{ij} = \frac{y_{ij} - y_{i.}y_{.j}/y_{..}}{\sqrt{\frac{y_{i.}y_{.j}}{y_{..}}(1 - \frac{y_{i.}}{y_{..}})(1 - \frac{y_{.j}}{y_{..}})}}$$

この e_{ij} は帰無仮説のもとで漸近的に標準正規分布に従うので，これを利用する．

8.2　対数線形モデル

変量の対数に線形構造を仮定するモデル．m を任意の自然数とし，$Y, X_i (i = 1, \cdots, m)$ を任意の変量とするとき，モデルは一般的に

$$\ln(Y) = \sum_{i=1}^{m} a_i X_i$$

と表現される．

8.2.1 $I \times J$ 分割表における対数線形モデル

変数 A,B があり，それぞれの水準数（カテゴリー数）は，I, J とする．このとき $I \times J$ の分割表が構成される．(i,j) セルの頻度の期待値を m_{ij} とすると，飽和モデルは，

$$\ln(m_{ij}) = u + u_i^A + u_j^B + u_{ij}^{AB}$$

となる．ここで，

$$\begin{aligned} u^A &: \text{要因 A の主効果} \\ u^B &: \text{要因 B の主効果} \\ u^{AB} &: \text{要因 AB の交互作用} \\ u &: \text{全基準…基準セルあるいは全平均の期待値の対数} \end{aligned}$$

である．ただし，

$$\begin{aligned} \sum_i^I u_i^A &= \sum_j^J u_j^B = 0 \\ \sum_i^I u_{ij}^{AB} &= \sum_j^J u_{ij}^{AB} = 0 \end{aligned}$$