心理教育測定法

「見えないもの」のはかりかた

小杉　考司

心理教育測定法

小杉　考司

Last Compiled on 2025.12.3

目次

第 1 章

心理尺度による測定の基礎

1.1 はじめに

心理学の研究法は調査，実験，観察の 3 つが代表的なものです。とくに調査法は，質問紙を用いて多くの人に回答を依頼し，それを統計的に分析することで研究仮説を確認しようとするものです。調査対象者をどのように集めるか，調査票をどのようにデザインするか，得られたデータをどのように分析するか，といったことでそれぞれ 90 分以上話せるぐらいに色々考えるべきことはありますが，ここではとくにその本質について考えてみましょう。すなわち，なぜ紙で用紙に丸をつけたものを集めただけで，心の何かがわかったと思えるのか，ということです。

アンケート調査は一見すると，なんのひねりもない研究法に思えます。すなわち，聞きたいことを当の本人に聞いてみる，これだけです。しかも答えやすいように (集計しやすいように？) 紙に問題が書いてあって，該当するところに丸をつけるだけでよかったりするのです。この手の研究は少なくとも 100 人，できれば数百人，大きい規模だと千以上の桁数の協力者に回答を求めます。それを PC に入力して集計するわけです[*1]。しかし「そう思う」を 5 点，「ややそう思う」を 4 点，以下 3，2，1 点と入力するのはなぜでしょう。これらの反応は**順序尺度水準**でしかない，と言われますが，それを間隔尺度水準であると「みなし」て，平均値を求めたりします。なぜそんなことが許されるのでしょう。

このような問題に答えるのが心理測定と呼ばれる学問領域です。この講義では，目に見えないものを測定するとはどういうことかを考え，測定の正当性や理論的裏付けを確認しつつ，測定された数値からどのように情報を取り出すことができるのか，あるいは測定されたものをうまく数値化したり可視化したりする方法について，解説していきたいと思います。

1.1.1 私たちは何を測ろうとしているのか

測定 (scaling) とは，一定のルールに沿って対象に数字を与えることを指します。

数値を与えなくても，ある特徴量についての比較判断は可能で，たとえば 2 つの物体について天秤の両方にそれを置き，傾けば「一方が他方よりも重い」と判断できます。これを多くの対象について行えば，重さという特徴量について序列をつけることができますし，なんらかの単位をもって表現することができるようになります。その単位の数を「測定値」とすること，これが測定の基本といえるでしょう。こうして重さ，長さといった単位が作られていき，それが異文化交流の中で単位の互換性を高めるために統一化・標準化されてきています。さらに「時間」という抽象的な概念についても測定され，単位が与えられるようになりました。このように考えると，われわれは何だって測定できるような気になってきます。

[*1] もっとも最近では，web で調査の回答を得ることで，入力や誤入力のチェックなどの手間が大幅に軽減されています。

心理学においては，物体と違って目に見えないものを研究対象にするわけですから，そもそも「人間の感覚も測定できるだろうか」というところから考えなければなりません。人間の感覚はあやふやなものですから，物理学のように安定して測定できるかどうかがまず疑問です。たとえば「ハンバーグは美味しい」という「おいしさ」についての評価も，空腹の時の評価と満腹の時の評価で変わってしまいますから，条件を統制するとか小さな違いに敏感な測定方法を考えなくてはなりせん。さらに，物理的な量のように大小の比較ができるかどうか，昨日のハンバーグと今日のハンバーグのどちらが美味しいか，と考えられるかどうかが問題です。昨日の経験と今日の経験は質的に違うものだ，というのであればこれはもう量的な数字を割り振る根本的な原理がないことになります。私たち何かを測定しているということは，何らかの側面において大小を比較できる，包含関係があるはずだという仮定を置いていることになります[*2]。

さらに問題がややこしくなる点が，心理学の研究対象が個々人の主観的な経験であるということです。わたしの「おいしい」とあなたの「おいしい」が同じ経験といえるのかどうか。物理的に同じ刺激であっても感じ方は人それぞれで，わたしの「かなりおいしい」とあなたの「ちょっとおいしい」が同じなのかどうかということになると，どこで単位を合わせれば良いのかわかりません。ここで徹底的行動主義者であれば，言語報告がどうであれ刺激に対する振る舞いが同じであれば同じ刺激の大きさとみなす，ということができるでしょう。あるいは社会心理学者であれば，「言語」が社会的に共通な単位として機能しているということがあるかもしれません。いずれにせよ，何らかの仮定，理論的なみなしを踏まえているはずなので，そこには自覚的であるべきです。

ここで確認しておきたい点は，心理尺度なるもので測定する対象について，次の 3 つの側面でどのように正当化されるのか，ということです。

1. それは時間的に安定した状態のものかどうか。どの程度持続しうる状態か。
2. それは大小関係を比較できるものかどうか。比較しうる同一の次元上にある状態か。
3. それは個人間で比較できるものかどうか。個人と個人を繋ぐ次元があるかどうか。

改めて考えると，これはなかなか難易度の高い問題であるように思います。とてもナイーブに考えれば，私たち自身の経験は私たち自身のものでしっかりと意識できるものですし，「自分は A よりも B が好きだ」と確信を持って判断することはできる，といいたいところです。しかし心理学を学べば学ぶほど，人間の認知的操作には限界があり，意図せず易きに流れる傾向があり，間違った判断をするし，他者との交わりにおいては嘘をついたり誇張したりすることがわかってきます。これを踏まえて慎重に，何をどのように測定するのかを考えていきましょう。

1.1.2 態度概念

心理学は何を研究しているのかについて，少しでも概論的な心理学をかじったことがある人は，一言では答えられないことに気づくと思います。心理学がカバーする領域は幅広く，右手のしていることを左手が知らない，ということがあり得る業界です。研究方法も大きく 2 つに分類され，たとえば下山 (2001) は心理学を次の表 1.1 のように分類しています。

実証主義的なアプローチとしては，生理心理学，知覚心理学，認知心理学，学習心理学などがあり，了解的なアプローチとしては臨床心理学，発達心理学などがあります。その間には，社会心理学や教育心理学など，実証的であったり了解的であったりするものも含まれています。

[*2] 数の大小というのは数直線のように，序列づけた時に大きい方が小さい方より後に出てくるということであり，その直線にそって小さい方を経由して大きい方に到達します。大きい方が小さい方を踏まえている，含んでいるとことで包含関係があるといえます。

表 1.1　下山 (2001) による心理学の分類

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

より基礎的な心理学である生理心理学や知覚心理学では，人間 (や動物) のハードウェアを解析するところから意識の物理的境界を探ろうとします。そこでは種としての人間が対象ですから，個体差はあるにしても基本的な組成は同じであるとみなして研究できます。気質や動機づけ (食欲・性欲・睡眠欲などの生き物としてのモチベーション) は人間の根源的なところに埋め込まれたものであり，それに基づく感情などもかなり共通のパターンを見てとることができます。

認知心理学は，ハードウェアに基づく OS(オペレーションシステム) や，それに伴う汎用アプリケーションを研究対象にしているといえるでしょう。考え方のパターン，アルゴリズム，効率の良い考え方や計算論理にカスタマイズされた思考ルートの癖，学習によって変化した辞書など，ある程度の個人差が見られるところでありますが，その中で比較的共通したパターンを探ろうとしています。

それよりも個人差や変化の度合いが大きいもの，時間的な安定性がより低いものは「行動」から研究対象を捉えようとします。もちろんその中でも比較的安定的な「行動パターン」を研究するのが，性格心理学や学習心理学です。これらの領域では個々人 (個体) の心の機微をみるというより，多くの個々人 (個体) を集めてはじめて見出せる全体的傾向であったり，特定の個体に確実に見られる刺激と反応の組み合わせを研究しようとします。個々人の主観的経験を丁寧に除去することで不安定さをなくし，安定的パターンを見出そうとするわけです。

了解的なアプローチに振り切ってしまうと，個々人の主観的な経験を直接的に扱うことが目的になります。その人の考え方，その人の悩みを受け止め，社会的な生きづらさや障害があるようであればそれを取り除く。そうした手法についてはマニュアル化できますが，扱っているケースはまさに個別の人生そのもの，といえるかもしれません。

さてこうした心理学のアプローチではちょっと物足りない，つまりもう少し個人の経験や感覚を扱いつつ，かといって「人生いろいろ」で終わってしまわない程度にパターン化できる何かが欲しい，というときに心理尺度は使われます。個別の行動を測定するのは一般化できないけれども，行動を生み出す心の働きの段階で共通する安定的な要素，これをいかにして取り出すかについての仕掛けが必要なわけです。

この微妙な問題に答えるために生み出されたのが**態度 (Attitude)** と呼ばれるものです。社会的態度は，次のように定義されています。

> An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.
> 態度とは心身の準備状態であって，経験を通して組織化され，それが関わっているあらゆる対象や状況への個人の反応の上に，指示的あるいは力動的な影響を与えているものである (Allport, 1967)。

なんだか非常に抽象的な表現で，わかったようなわからないような，というものです。藤原 (2001) は態度の性質を次のようにまとめています。

- 態度とは反応のための先有傾向準備状態である。したがって態度は，刺激と反応との媒介物であり，直接的には観察不可能な構成概念である。測定論的には，多数の一貫した反応群から推定される潜在的反応と考えられる。
- 態度は常に対象を持つ。
- 態度は動機的一情緒的性質を持つ。すなわち，態度はある一定の対象について「良い – 悪い」，「好き – 嫌い」といった評価を下す。その評価は中立を通り，ポジティブからネガティブにその方向と強度を変える。
- 態度は一時的な生物体の状態ではなく，一旦形成されると比較的安定しており，持続的である。
- 態度は先天的なものというよりむしろ，学習されたものである。
- ここの態度は cluster をなし，そうした cluster は constellation[*3] を形成する。

そして態度は価値，動機，動因，判断，セット，習慣，信念，意見といった諸概念とは異なる，という説明がなされることがあります。やや形而上学的で，「それ」が何なのか，というのを明示するのが難しいのですが，ポイントは，

1. 態度は対象を持つ
2. 態度は正と負の量をもつ
3. 態度は測定できる

というところです。

態度は一般に特定の対象に対する，正負・量的な評価が可能なものと考えられています。たとえば「自民党に対する態度」というのは，自民党という対象に対して，「好き」「嫌い」というポジティブ・ネガティブの評価ができ，さらに「とても好き」「やや嫌い」のように量的に表現できるものでもある，と仮定されます。多くの項目で調査することによって，項目同士の誤差が相殺しあって真のスコアにちかづくことができ (→ セクション 2,Pp.15)，学習によって後天的に何らかの要素の積み重ねでそれが形成されるとすれば，全体的な特徴として正規分布を仮定できます。社会的な態度は極端な値が少なく中庸な態度がもっとも多くなるように分布し，相対的に評価されます。

これらの仮定は心理尺度を構成する上で非常に扱いやすい性質といえます。態度概念が「あれでもない，これでもない」と他の類似概念との差別化を図りつつ，「社会心理学においてもっとも特徴的で欠くことのできない概念 (Allport, 1967)」と神格化することで広く使われるようになりましたが，今改めて見直してみると随分とご都合主義的な仮定であったようにも見えます。うがった見方をすれば，

1. 人の行動パターン，考え方を研究しよう
2. 客観的にデータを集めないと，心理学は科学じゃなくなっちゃうから測定できることにしよう
3. 行動だけでなく生物体の中にある組織化された概念の連合も研究したい
4. 行動が起こる前に，だいたいの予測はたてられないものであれば使い勝手が良い

という流れの中で生み出されてきた，非常に抽象的で便利なもの，それがあるものと想定して研究を積み重ねていこう，というツールが態度だといえるかもしれません。

そんなよくわからない想定のものを研究したいわけじゃない，私はもっとリアルな，生の人間の心が知りた

[*3] 類似したものの集まり。星座ではないよ。

いんだ，という反論が聞こえてきそうです。しかし調査法でよく使われているリッカート尺度というのは「態度が測定できる」という仮定のもと，態度の測定論的な仮定・性質に基づいてスコアリングされる尺度作成法があり，その簡易版として導入されたものである，という事実を忘れてはいけません。心理尺度を使って研究する以上，態度尺度の構成原理に従っているはずであり，もし態度でない何かの構成原理に従っていると言うのであれば，その理論的妥当性を訊さなければならないのです。

1.2 尺度の四水準

さて，これから尺度を作ったり作られたものについて考えていくわけですが，その前に測られた数字がどのような計算に耐えられるのかについて，今一度確認しておきたいと思います。心理統計のテキストは何を開いても，まず Stevens (1946) による**尺度水準 (level of measurement)** についての言及があります。何を測定した数値であるかとは別に，その数値にどのような算術処理を施すことができるかによって，数字を 4 つのレベルに分けるのでした。

■名義尺度水準　名義尺度水準 (nominal scale) は数字と対象が 1 対 1 で対応していることだけが重要です。男性を 1，女性を 2 とコード化するようなもので，この時「女性は男性の 2 倍である」といった数としての意味はありません。男性を 0，女性を 42 としても本質的に変わりがないからです。このような名前だけの数字は，計算ができませんので，せいぜい 1 が何件あったかという度数を数えて集計するにとどまります。

とはいえ，数字が直接対象を指し示しているわけですから，もっとも意味のある数字かもしれません。

■順序尺度水準　順序尺度水準 (ordinal scale) は，数字が大小関係の意味を持っているものです。レースで 1 位 2 位と順番がつくと，1 位のほうが 2 位より優れていることがわかります。人間の心理的な反応，とくに 5 段階や 7 段階で評定させる心理尺度は，この水準に相当します。選好の順序は明確でも，量的な違いがわからないからです。

■間隔尺度水準　間隔尺度水準 (interval scale) は，数字と数字の間隔が等しいことが制約として加わります。たとえば気温で 10 ℃と 20 ℃の差は，25 ℃と 35 ℃の差に等しいと言えます。これは摂氏が氷点を 0 ℃，沸点を 100 ℃としたうえで百等分したという定義から明らかなことです。間隔が整っているので加法・減法の計算は可能ですが，原点が定かでないので比を考えることはできません。たとえば 10 ℃は 20 ℃の倍の熱量を持っている，とは言えないのです。なぜでしょうか。たとえば，同じエネルギー状態を別の温度体系に置き換えてみたとしましょう。新しい温度体系は，氷点が 100 で沸点が 200 だったとします。そうすると 10 ℃は 110，20 ℃は 120 に該当しますが，120 は 110 の 2 倍にはなっていないからです。

■比率尺度水準　比率尺度水準 (ratio scale) は，さらに絶対 0 点の制約を付け加えたものです。これで原点からどれ位離れているか，を基準にして計算ができますので，乗法・除法もできることになりました。物理的な単位系はこの尺度水準にあるものがほとんどですから，緻密な計算モデルを作ることができるのですね。

さて早足で 4 つの水準について説明をしてきました。これが重要なのは，尺度水準によってできる計算が変わってくる点にあります。名義尺度水準は数え上げぐらいしかできません。順序尺度水準も同様で，名義や順序といった**質的変数 (categorical variables)** の場合，たとえば代表値を求める時も度数を数えて最頻値を報告する，というぐらいがせいぜいなのです。

これに対して，間隔尺度水準や比率尺度水準の**量的変数 (numeric variables)** では，加減乗除の計算ができますので，平均値を求めたり標準偏差を求めたり，ということができるようになります。

このように，データが得られた時にその数値がどのような処理に耐えうるのかを，しっかり意識しておく必要があります。より踏み込んだ言い方をすれば，私たちが「非常にそう思う」と強い程度で表現した時に，その

“心の状態” が量的なものか，順序的なものか，質的なものかについて考えることなしに，結果を解釈することはできないはずなのです。

こうした数値的な特徴を把握した上で，態度尺度がどのように作成され，どのように数値化しうるのかについて，改めて見ていくことにしましょう。

1.3 心理尺度の作り方

心理尺度の作り方には大きく分けて 3 つのスタイルがあります。

1 つ目はサーストン法による尺度で，**等現間隔法 (method of equal-appearing intervals)** と呼ばれるものです。2 つ目はリッカートの**シグマ法**と呼ばれるもので，5 件法，7 件法など数段階のカテゴリラベルのもっとも近いところに丸をつけるという方法です。態度尺度以外にも，測定に際して非常によく使われる方法で，一般に**リッカート法 (Likert Scale)** と呼ばれることもあります。3 つ目は SD(Semantic Differential 法，意味微分法と訳されることも) 法とよばれるものです。SD 法は態度測定というより，イメージの測定を目的としたもので，対象を提示しつつペアになった形容詞を列挙して提示します。たとえば「広島大学」という対象に対して，「激しい – 落ち着いた」「慎重な – 軽快な」といった形容詞対ではどちらの表現が近いかを評定してもらいます。形容詞の対が作る軸上でいうと平均的にどのあたりに対象がプロットされるのか，を見ることで対象ごとのイメージの違いを表現するのが基本的なアイデアです。SD 法は複数の対象に対するイメージの相対的比較ですから，スコアの点数化にはそれほど重きを置いていないのでここでは取り上げません。

サーストン法とリッカート法は，これを使って態度のスコアをつけることができます。小杉の自民党に対する態度は 4.8 点だ，といったように，です。こうした数値化がどのような理屈でなされるのかを，今からみていこうと思います。

1.3.1 サーストンの等現間隔法

サーストンの等現間隔法(とうげんかんかくほう)は，社会的な態度について絶対評価を与える方法です。この方法で作成された尺度は，各項目 (態度表明文と呼ばれます) に尺度値がついており，回答者は提示された項目に賛成であればその尺度値がその人の態度得点になります。この尺度値は事前に複数の評定者によって決めておく必要があります。すなわち，尺度を作る前の入念な準備が必要です。また，サーストンの尺度は 1 次元性を有していることが前提となります。

具体的な作成方法は次のような手順で行います。

1. 項目の収集
2. 評定者集団による評定
3. 尺度値の算出
4. 項目の選定

以下順に説明します。

■項目の収集　まずは測定したい社会的態度のテーマに沿って，項目を準備します。たとえば「自民党に対する態度」のように，誰でも思い描ける具体的な対象が良いでしょう。このようなテーマが決まれば，これに対する態度項目を色々考えます。「自民党のことを考えると夜も眠れない」とか「自民党に関係したニュースはなるべく見るようにしている」「近所の自民党員の事務所に行くことがある」というポジティブな態度もあるでしょうし，「自民党のニュースはなるべく聞きたくない」「自民党には投票しない」「自民党は不正まみれの悪い政党である」といったネガティブな態度もあるでしょう。こうした文言をなるべく多く，強い態度から弱い態度まで，

ポジティブなものからネガティブなものまで網羅的に準備します。ニュートラルな項目も考えておく必要があります。

■評定者集団による評定　尺度値を決めるための事前準備です。まず評定者を無作為に集めます。少なくとも十数人は必要でしょう。評定者には事前に準備した項目が好意的－非好意的 (または肯定的－否定的) の 1 次元にそって，7～11 段階ぐらいの多段階に分類してもらいます。「自民党のことを考えると夜も眠れない」というのは非常にポジティブなので 11 点，「自民党には投票しない」というのはかなりネガティブなので 2 点，といったようにです。

■尺度値の算出　このように各態度表明文を複数人で評価してもらいますから，その項目の平均値，中央値，分散などの記述統計量を計算できます。この中央値 (あるいは平均値) をその項目の尺度値とします。ただし，ここで分散が大きい項目は，評定者によって評定の仕方がバラバラだということを意味しますよね。値が人によって定まらないというのは，その項目が刺激としてあまり好ましくないと考えられるので，項目候補から削除します。誰がみても 10 点とか誰がみても 3 点，といった分散が少ない項目が望ましいでしょう。

■項目の選出　さてこうして尺度値が計算できたら，それを順に並べていきます。「夜も眠れない」は 10.7 点，「事務所に行くことがある」は 9.5 点，「ニュースをなるべく見る」は 7.9 点･･･というようにしていくことができますね[*4]。このとき，項目間の間隔が均等になるように項目を選別します。たとえば「夜も眠れない」と「事務所に行くことがある」の間隔は 1.2 点ですが，「事務所に行くことがある」と「ニュースをなるべく見る」の間隔は 1.6 点になっています。これでは等間隔と言えないので，1.2 点間隔すなわち 8.3 点ぐらいの項目を選出します。

このことからわかるように，サーストン法で尺度を作る場合は，事前に多くの項目を準備しておかないと「ちょうどいいところの表明文がない」となってしまう恐れがあります。ですから最終的にできる尺度に含まれる項目の，5 倍から 10 倍ぐらいの数の項目候補を事前に準備し，うまく等間隔に項目が選出できるようにしなければなりません。

なぜ等間隔に選ぶのかというと，もうお分かりですね，これで得られる尺度値を**間隔尺度水準**として扱いたいからです。間隔が等しくなければ順序尺度にしかなりませんが，間隔が等しいことがわかっていると，平均や分散などの計算をし，相対的な比較をできるからです。またこの尺度を使うときは，すでに評定者集団によって尺度値がわかっていますから，回答者にずらりと並べられた尺度を見てもっとも自分の意見に近い項目を選出してもらえば，その項目の尺度値がその人の態度得点だということができます。

評定者集団をなるべく偏りなく多く集めることで，事前に尺度の値を確定させておき，あとは本来研究対象にしたかった人にその尺度を当てれば尺度値 (尺度得点) が求められる方法ですから，準備が大変だけど使うときは確実で絶対的なスコアを与えることができるというのがこの方法の利点です。欠点はその準備コストの高さと，1 次元的な態度しか用いられないことでしょうか。また尺度構成の観点から重要なのは，選出プロセスによって項目間の尺度値が均等であることが保証されている点です。均等に選んだ後で，1，2，3，4，5 と数字を付け直しても構いません。大事なのは，こうしたプロセスのおかげで間隔尺度水準が維持され，以後の分析に耐えうるスコアになっているという点です。

[*4] 中央値なのになぜ小数点があるのだ，と思う人がいるかもしれません。平均値でもいいですし，評定者が偶数人の場合は中央値も両得点の平均や重みつき平均で小数点が出るからです。

1.3.2 リッカートのシグマ法

次に紹介するのはリッカートの**シグマ法 (sigma method)** です (Likert, 1932)。これはいわゆる 5 件法，7 件法と呼ばれる採点方法で，「私は自民党の政治のやり方が好きだ」といった項目に対して，「まったく当てはまる」「かなり当てはまる」「やや当てはまる」「どちらとも言えない」「やや当てはまらない」「あまり当てはまらない」「まったく当てはまらない」といった順序づけられたカテゴリーにたいしてもっとも自分の考え・態度と近いところに丸をする，という方法で反応が得られます。この時の反応カテゴリーが，今回は 7 つありますから 7 件法 (7-points scale) で回答を求めた，などと言います。5 段階なら 5 件法，4 段階なら 4 件法です。普通は「どちらとも言えない」というところを用意するために奇数 (3,5,7,9) 件法を使いますが，日本人は「どちらとも言えない」を選びやすいという中庸傾向があるとも言われていますので，意見をはっきりさせるために 4,6 件法も使われたりします。

項目はこれも複数あって，たくさん集められた項目を分析するために，もっとも当てはまるを 7，かなり当てはまるを 6，以下同様にしてまったく当てはまらないを 1，とコード化し分析するのが一般的です。ただし注意して欲しいのは，もっとも当てはまる＝ 7 としたのは**名義尺度水準**の数字の割り当て方と一緒で，このカテゴリーが 7 という尺度値を持っているわけではない点です。そもそもこの評定カテゴリーは，統計学的にはせいぜい順序尺度水準の性質しか持っていませんから (もっとも＞かなり＞やや)，カテゴリーに割り当てた数字からそのまま平均や分散の計算をするのはおかしいはずなのです。もしこれらのカテゴリーの間隔が等しかったとしても，3，4 段階しかないようであればやはり間隔尺度水準の計算ができるほどの精度は持っていません。数量的に分析するには (等間隔が担保された上で)9 から 11 段階は必要といわれています。

それではリッカート法ではどのようにして尺度値を決めるのでしょうか。リッカート法も測定しようとしているのは態度であって，表に出てくる反応カテゴリーの背後には連続的な心理的態度というのがある，と仮定しています。またこの (社会) 心理学的態度は，向きと大きさがあって正規分布を仮定できます。リッカート法も正規分布に従う潜在的な連続変数があると仮定するのです。

さて，ある項目について，多くの人からデータを集めて「もっとも当てはまる」「かなり当てはまる」といったカテゴリーごとの集計ができたとしましょう。多くの人の態度も集積すれば正規分布に従いますから，きっとこのヒストグラムも正規分布を反映したものになっているはずです。しかし我々が知りたいのはその背後にある連続体上の数字なわけです。

ここで図 1.1 を見てください。上段にあるのがある項目のヒストグラムの例です。しかし知りたいのは，下段にあるような正規分布の形をした連続体の変数のはずです。上段のヒストグラムは下段の状態を反映しているはずですから，上段のカテゴリの相対頻度を元に，下段の正規分布を分割します。具体的な数字との対応は表 1.2 を見てください。出現度数を相対頻度にし，正規分布の面積を順に分割していくことになります。カテゴリの下の方から順に分割するということで，表 1.2 の三段目には累積 (相対) 頻度を書いてあります。そしてこれを使って，標準正規分布の下から面積を考えます。統計環境 R では，`qnorm` 関数をつかうと累積確率の確率点が求められます (表 1.2 の 4 段目)。またその時の確率密度も求めてあります (表の 5 段目。R では `dnorm` 関数を使って求めます)。

さて，ではここからどのようにして尺度値を求めればいいでしょうか。一般に C 件法で，下から $1, 2, 3, \cdots c \cdots, C$ とカテゴリ順に数字を割り振ったとして，第 c カテゴリの尺度値 Z_c を考えるとします。このカテゴリ c は標準正規分布において上限 z_c，下限 z_{c-1} の確率点で挟まれる領域としていますから，この幅 ($[z_{c-1}, z_c]$) の平均を取ることを考えます。この点は，次の式で求められます (証明は付録 A,Pp.147 参照)。

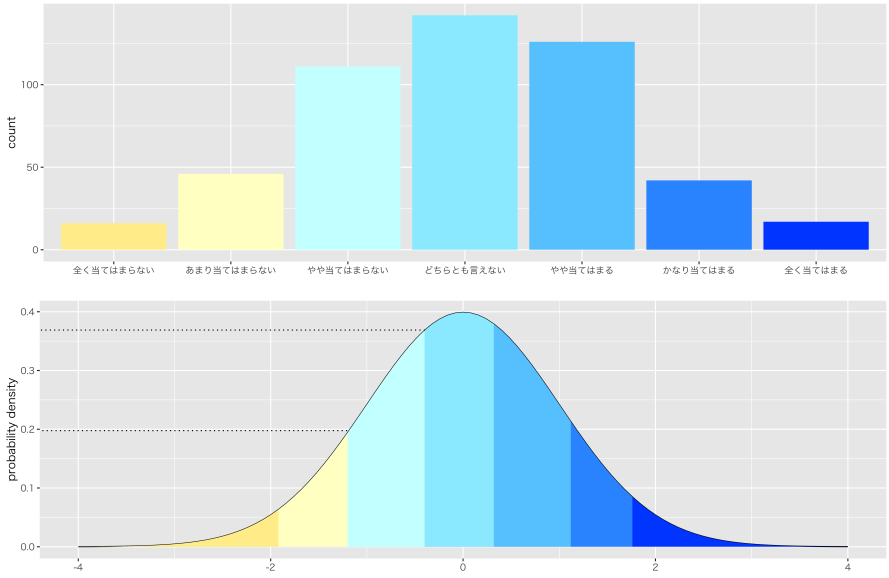

図 1.1　カテゴリ反応と背後の連続値

表 1.2　カテゴリと数値の対応

反応カテゴリ	まったく当てはまらない	あまり当てはまらない	やや当てはまらない	どちらとも言えない	やや当てはまる	かなり当てはまる	まったく当てはまる
出現度数	16	46	111	142	126	42	17
相対度数	0.03	0.09	0.22	0.28	0.25	0.08	0.03
累積相対度数	0.03	0.12	0.35	0.63	0.88	0.97	1.00
累積相対度数の確率点	-1.85	-1.16	-0.40	0.33	1.19	1.83	∞
累積相対度数の確率密度	0.07	0.20	0.37	0.38	0.20	0.08	0.00
付与される尺度値	-2.33	-1.44	-0.77	-0.04	0.72	1.50	2.67

$$Z_c = \frac{(y_{z_{c-1}} - y_{z_c})}{p_c}$$

ここで y_c は z_c, z_{c-1} における確率密度，p_c はカテゴリ c の相対頻度です。具体例でいきましょう。表 1.2 の数字を使うと，「やや当てはまらない」($c = 3$) の尺度値は，$\frac{0.20 - (0.37)}{0.22} = -0.7727273$ となります (分子は図 1.1 の点線部の差分，分母は該当箇所の面積になります)[*5]。このようにして計算された尺度値が，表 1.2 の一番下の段にある数字です。

このように，累積度数をつかって尺度値を決めるリッカートの方法を，**シグマ法 (σ method)** と呼びます。こうして作られた尺度値は連続体上の数字ですから間隔尺度水準になり，平均や分散をはじめとした数値計算

[*5] 表 1.2 は小数点下 2 桁までに丸めているので，正確な値ではありません。

に耐えうる値になっています。機械的に「まったく当てはまらない」から「まったく当てはまる」まで，1.0 刻みで数字を割り振っているのではないのです！

・・・と言いたいところですが，今回の尺度値を眺めてみるとそこそこ等間隔 (間隔は 0.8 ぐらいでしょうか) に並んでいますね。試しに各尺度値を 0.8 で割ってみると，$-2.91, -1.81, -0.97, -0.04, 0.90, 1.88, 3.33$ となります。四捨五入して小数点をなくしてみると，$-3, -2, -1, 0, 1, 2, 3$ となりますね。そう，つまり**非常にラフな近似値でよければ，機械的に 1,2,3... と数字を振っても同じことになります**。ですから，実際の研究ではとくに深く考えずに 1,2,3... と割り振った数字をそのまま使われたりするのです。大山鳴動して鼠一匹といいますか，苦労した割に得るものがないじゃないか，とお叱りを受けそうですが，少なくとも「なぜリッカート法は順序尺度ではなく間隔尺度のように扱って良いのか」という問いには答えられると思います。また，ここにくるまでに，態度の 1 次元性や正規分布の仮定などが含まれていたことを改めて思い出してください。今回は数値例ですので，綺麗に七段階に分かれるようなものを用意しましたが，実際の調査では正規分布しないものや，平均が低すぎるとか高すぎるものが結構みられます。それらに対して機械的につけた数字で分析するのは決して適切な方法ではなく，シグマ法を始めその他の手法で適切な尺度値を付与すべきなのですが，人間は易きに流れるもので**ほとんど考慮されていないのが現状です**[*6]。

[*6] この状況は決して良いものではなく，悪しき研究上の風習だと思われます。幸い，第 2 講で説明する**項目反応理論 (Item Response Theory)** の一種，**段階反応モデル (Greaded Response Model)** を用いると，この問題点をカヴァーしつつ有用な情報が得られますので，みなさんは一足飛びにその手法を身につけた方が良いと思います。

第 2 章

テスト理論

前回は態度尺度の作り方を見てきました。態度は個人の中にある比較的安定した傾向性であり，刺激ー反応を媒介するものと考えれています。これはいわゆる S-R 図式，つまり刺激と反応のセットから，ブラックボックスの性質を表現しようという**機能主義 (functionalism)** の考え方そのものです。*funciton* は関数という意味もありますから，**関数主義**と言い換えてもいいかもしれません。

直接は目に見えない，潜在的な状態を測定しようという発想は**テスト**と同じ発想です。皆さんが幾度となく受けてきたあのテスト，学力テストです。テストは「正答数が多ければ点数が高く，正答数が低ければ点数が低い」ものであり，幾つもの問題を被検者に投げかけて，「正答する」という一貫した傾向が見られれば，その人の中に「知識」「学力」「能力」といった性質があるだろうと考える，というものだからです。

心の状態 (態度) も一定の刺激に対する反応パターンから類推するものですので，テスト理論の考え方は態度測定の話と非常に密接な関係があります。そこでここでは，測るということについて理論的な裏付けを積み重ねてきたテスト理論を概観して，心の測定の参考にしてみましょう。

2.0.1 古典的テスト理論

心理学における測定は，物理学などハードサイエンスと違って誤差が大きいことを見越して考えなければなりません。知能検査，性格特性などを測定するために考え出された古典的テスト理論のモデルは，

$$X = t + e$$

で表されます。すなわち，テストの点数 X は真のスコア t と誤差 e に分割できるというものです。非常に単純なモデルですが，測定したものには誤差がついているという考え方，言い換えると目に見えるものだけが真実ではないという考え方がしめされています。この考え方は，ソフトサイエンスの領域においては重要なことです。

またこの古典的テスト理論から，いくつかの重要な考えを読み取ることができます。ひとつは誤差についての考え方です。このモデルを $X_j = t_j + e_j$ のように，ある個人の変化しない属性について，j 回測定したとします。この時，測定の平均は次のように計算できます[*1]。

[*1] この式は，ある測定を多くの個人 i に対して行ったものとして，$X_i = t_i + e_i$ と考えることもできますが，添字が異なるだけで式の展開に違いはありません。

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

さて古典的テスト理論では，誤差に関して次のことが仮定されます。

- 誤差の平均はゼロ。つまり誤差が出現するときは，「正に偏る」「負に偏る」といった一貫した傾向がないと考える。
- 真のスコアと誤差との相関はゼロ。つまり誤差は真のスコアに関係なく影響してくるもので，真のスコアと共変動するようであれば偶然によるものとはいえない。
- 異なる測定誤差間の相関はゼロ。誤差同士がなにか意味のある変動をしているのであれば，それはもう偶然によるものとはいえない。

これらはいずれも，誤差が「偶然によって現れる影響で，制御不可能なもの」という考え方からは自然な仮定だといえるでしょう。より詳しくいえば，誤差は測定に応じて毎回一定の傾向で生じる**系統誤差 (systematic error)** と，全く傾向のつかめない**偶然誤差 (random error)** に分けて考えられますが，系統誤差は測定に際して工夫して取り除くべき問題であり，ここでは全くの偶然による誤差についての議論だからです。

さて誤差の平均がゼロ，つまり $\bar{e} = 0$ ですから，$\bar{X} = \bar{t}$ となって，いつかは誤差がなくなって真のスコアを得ることができるようになる，ということが示されます。

また平均は 0 ですが分散はゼロではありません[*2]。このテストの分散を考えると，次のようなことがわかります。

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

[*2] ガウスの考えた誤差論から，誤差は確率**正規分布 (Gaussian Curve)** に従うと考えられます。

ここで Cov とは共分散を表しています。第二項の $2Cov(te)$ は真のスコアと誤差との共分散 (を 2 倍したもの) を表していることになりますが, 共分散 (相関) がそもそも線形関係を表す指標であったことを思い出してください。相関係数は共分散を標準化したものだったわけですが, そういう意味ではこの $Cov(te)$ というのは真のスコアと誤差との相関関係を表しているようなものです。さて, ここでも誤差の仮定から, 相関はゼロです。すなわち, 誤差とはどのような傾向もなく出現するもの, という考えられているのです。どのような傾向もないわけですから, 当然なにかと相関関係にあるはずがない, すなわち $Cov(te) = 0$ であるとなります。

これを踏まえると, $Var(X) = Var(t) + Var(e)$ のように, テストの分散が真のスコアの分散と誤差の分散に完全に分割されました。ここから, 信頼性の定義は

$$\frac{Var(t)}{Var(X)} = \frac{Var(t)}{Var(t) + Var(e)}$$

と表すことができます。言葉で言えば, **信頼性**の定義は「全分散中にしめる真のスコアの分散の割合」ということになります。

2.1 尺度を評価する

テストは目に見えないものを測定するためのツールです。できあがったテストがきちんと測定できているのかについては, 信頼性と妥当性という 2 つの側面から評価しなければなりません。

2.1.1 信頼性

信頼性は測定の安定性と言い換えてもいいかもしれません。すなわち, 同じものを 2 回測れば, 同じ数字がえられることが安定した測定です。測定するたびに数字が変わるようでは, その測定器 (ここでは尺度ですが) は信用ならない, というわけです。信頼性 Rel は, 次のように表現できるのでした。

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

言葉で表せば, 信頼性は全分散に占める真分散の割合であり, 全体から誤差分散の割合を引き算したものであるとも言えます。

信頼性のない尺度があれば, その後のお話は先に進みませんから, まずもって「尺度が信頼できるかどうか」を評価する必要があります。このことを信頼性は妥当性の上限, と表現したりします。

信頼性を評価する方法は, 測定値の安定の程度を評価できればいいのですから, 複数の測定を持ってその相関係数を計算することでひとまず達成できます。同じ検査を 2 回行って, 測定値の相関係数を見ることでその安定性を評価する方法を, **再検査信頼性 (Test-Retest reliability)** といいます。

しかし同じ尺度を何度も使うのは, 調査回答者に要らぬ構えを持たせてしまいますね。昨日と同じテストを今日もやる, ということになれば対策を練ることができますし, 同じ文言が繰り返されると飽きてきてわざと違う答えを書いてやろう, と思われるかもしれません。そこでまったく同じテストではなく, よく似た別のテストを使って一貫性の指標とすることが考えられます。これは**平行検査信頼性 (Parallel test reliability)** と呼ばれます。

平行検査信頼性は, テストの質が同じぐらいで内容が違うもの, ということになります。算数のテストであれば数字を少し変えることで, 同質の異なるものをつくることができたといえるかもしれませんが, 他の教科や心理学一般であれば, なかなか難しいことかもしれません。そこで 1 つのテスト項目を, たとえば偶数項目と奇数項目のように無作為に 2 分割し, それらの相関係数でもって安定性の測度とすることが考えられます。これを**折半法**といいます。

折半法では 1 つのテスト項目群を半分に分けるのでした。これはなにも半分ではなく，3 分割，4 分割と細かく割ってそれぞれのブロックの相関係数を計算し，その平均でもって全体的な安定の指標と考えることができます。いやそこまでいくのであれば，5,6,7 分割... といって究極的には M 項目の質問があれば M 分割してしまえ，ということが考えられますね。

個々の項目が他の項目とどれぐらい相関しているかを考えてみましょう。まず尺度全体から計算される回答者の値は，項目の尺度値の合計であるのが普通です (テストの点数も正答数に対応していますね)。ですから，ある項目 j の尺度値は，j を除いた残り $M-j$ 個の尺度値の和と高い相関をするはずです。このように，各項目が尺度全体の値とどの程度相関しているかを見る **IT 相関 (Item-Total correlations)** は，尺度の信頼性を見る 1 つの指標になります。ある項目が，尺度全体と相関していなければ，それはその項目が尺度の中で目的と違うものを測定している可能性があり，それは必然的に尺度の安定を損ねる結果になるからです。

この考え方を発展させ，各項目が他の項目とどの程度相関しあっているか，つまり尺度の中でどの程度整合性がとれたもの，すなわち一貫して同じものを測定しているのかを評価する指標として，α **係数 (alpha coefficient)** があります[*3]。これは，テストに含まれる項目数を M，テスト全体の分散を V_t，項目 j の分散を V_j と表すと，次の式で表されます。

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

この指標は，各項目が他の項目とどれぐらい相関するかを総合的に表した指標で，とくに**内的整合性信頼性**と呼ばれます。

このようにして，尺度の安定の程度である信頼性は数値化して評価されます。安定した尺度であることが確認できた上で，では安定して「何を測っていたのか」を考えることにしましょう。それが妥当性の概念です。

2.1.2 妥当性

妥当性 (validity) は，信頼性をその上限とした上で，それが何を測っているのかを改めて考える指標です。

たとえば身長を測ろうとして，体重計を使うとします。成長に応じて，身長が伸びますが，それは体重とも関係がありますので，身長の伸びに応じて体重も増えていくでしょう。体重計は安定した計測器で，信頼性は十分あると思いますが，体重計で身長が測れていると言えるでしょうか。相関する変数ですので，部分的に Yes といえそうですが，やはり身長は身長計で測ったほうが良いでしょう。身長計の方が，身長という特徴を的確に捉え，より本質に近い測定をしているからです。このように，作ったものがしっかりとその本質を捉えているかどうか，これが妥当性の基本的なポイントです。

妥当性はですから，そもそも概念としてその測定しようとしているものが適切かどうか (**構成概念妥当性 (construct validity)**) とか，その文言でちゃんと質問できているか (**内容的妥当性 (content validity)**)，理屈通りその測定値が結果と変動しているか (**基準関連妥当性 (criterion validity)**) といった面から検証されます。最後の基準関連妥当性については，基準値と尺度値をつかって数量的に検証できますが，構成概念妥当性や内容的妥当性は中身の問題であったり，言葉と概念の対応であったりするので，数理モデル的アプローチができるものではありません。数値化できないか大きな問題ではない，というのはもちろん逆で，数値化できないところであるからこそ，専門的な観点，幅広い視野，批判的思考でもって検証していかなければなりません。

[*3] **クロンバックのアルファ (Cronbach's alpha)** とも呼ばれます。

2.2 テスト理論の展開

ここまでのテスト理論は，古典的テスト理論という名称でした。古典的，というのであれば現代的なテスト理論は何がどうなっているのでしょうか。

現代的なテスト理論，新しいテスト理論ともいわれますが，それは**項目反応理論 (Item Response Theory)**，あるいは項目応答理論とよばれます。略して **IRT** と表現されます。これを考えるに当たって，テストの項目を分析するというのはどのようなことがなされているのかをみていきたいと思います。

2.2.1 通過率と累積正規分布

みなさんは大学入学共通テスト (旧センター試験，さらにその前は共通一次試験といいました) や，学内の定期テスト，模試など色々なシーンでテストを受けてきたことと思います。模試などでは偏差値が明らかになり，自分の実力が相対的にどのあたりにあるのかがわかるようになっていたかと思います。大学共通テストなどは 50 万人ぐらいが一度に受験しますから，さまざまな学力の人がそこには含まれるのですが，成績を図にするととても綺麗な正規分布になることが知られています[*4]。正規分布は誤差の分布でもありますが，多くの要因が考えられる際の集積的データも，自然とこの形になります。

さて，学力のような潜在変数が標準正規分布に従うと仮定しましょう。この分布の形はどこの確率点がどれぐらいの確率密度を持っているか，あるいはある確率点以上・以下の面積が全体の何 % かを表すものですが，縦軸をある点以下の累積確率に書き直してみましょう (図 2.1)。図 2.1 の上の図がいわゆる正規分布の分布の形，確率密度関数です。下の図はこれを累積確率に書き換えたものになっています。累積確率は 0% から始まって，最終的に 100% にまでどのように増えていくかを示した図になります。

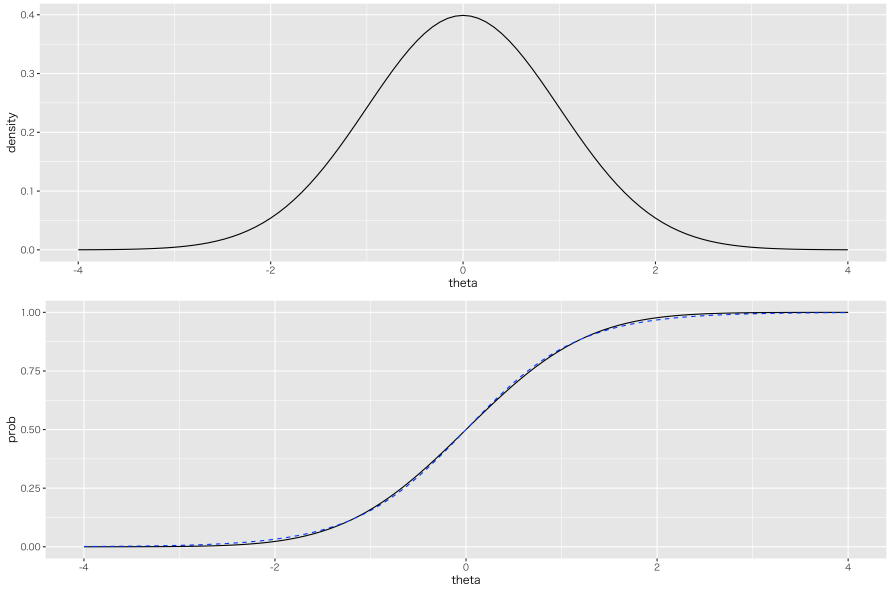

図 2.1 正規分布の確率密度関数 (上) と累積確率関数 (下)

累積正規確率は，テスト理論と密接な関係があります。というのも，学力が正規分布すると考えるなら，累

[*4] 山内 (2010) の見返し (表紙を開いた最初の内側のページ) に，センター試験の成績分布が載っており，綺麗な正規分布であることが示されています。

積正規分布の形はあるテスト項目の**通過率 (pass ratio)** と同じ形になると考えられるからです。

通過率とは，あるテスト項目に正答する人の割合のことです。ここで複数の項目からなる，あるテストをしたとしましょう。正答数を数えてその人の成績とすると，よくできたテストであれば成績は正規分布に従います。さらに，ある項目と成績との相関 (**IT 相関 (Item-Total correlations)**) は高いはずですね。つまりその項目に正答することが，テスト全体の成績と高く関係しているので，その項目はテスト全体が測ろうとしているものを反映していると考えられるからです。また，成績をもとに被験者集団を 5 群に分けたとしましょう。「成績上位群 (HH)」，「成績やや上位 (MH)」，「成績中程度 (M)」，「成績やや下位 (ML)」，「成績下位 (LL)」です。このとき，各群の平均通過率を考えると，図 2.2 左上図のようになるのが理想的です。つまり，成績が高い人たちの通過率は高く，成績が低い人達の通過率は低くなるはずですね。同じ図の右上は，LL 群でも半分ぐらいが通過し，その後の群は過半数，ほとんどが通過するようになっています。これは，この項目が簡単すぎたことを意味しています。簡単すぎる問題は，それはそれで被験者の特徴が弁別できないという意味で悪い項目です。逆に図の左下にあるのは，HH 群でも半分以下の通過率しかありません。つまり難しすぎる問題です。ほとんどの人が間違えてしまうわけですから，これも良い試験問題とは言えないでしょう。右下に至っては逆転していて，どうやったらこういう項目が作れるのか却ってわからないほどですが，学力の低い人だけが正答できて，学力の高い人は正答しない項目ということになります。もちろんこんな項目はよくないわけで，IT 相関で負の相関が出ているわけですから，テストの文脈でいうなら「そのテストで測っていない何か別の能力を測っている」と考えるしかありません。

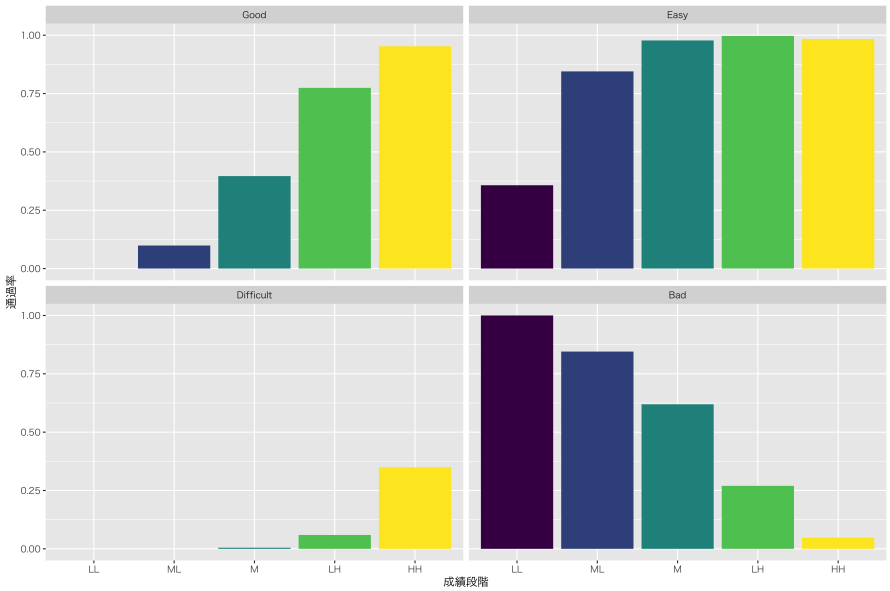

図 2.2　群ごとの平均通過率。左上が良いパターン。右上は簡単すぎる，左下は難しすぎる項目。右下は逆転していて良くない項目。

ともあれ，このようなやり方で項目の良し悪しを見ていくことができます。また，図 2.2 は 5 段階ですが，7 段階，9 段階とよりきめ細かくしていくと，理想的な形は累積正規分布により近づいていきます。新しいテスト理論による項目分析はこの累積正規分布の形を基本とし，これを拡張することで各項目の特徴を描いていくことになります。

ところで 1 つ前の図 2.1 の下の図には，実線と点線の 2 つの線が絡んでいることにお気づきでしょうか。実

線の方は確率分布関数から累積確率を出して描いたものですが[*5]，点線のほうは次の関数を使って描いています[*6]。

$$f(x) = \frac{1}{1 + exp(-1.7x)}$$

この関数，図から明らかなように累積正規分布とほとんど同じですよね。累積正規分布の関数を直接使うと，積分計算 ($\int$ を使うやつ) が入ってくるのでちょっと計算が面倒ですから，こちらの関数の方を近似関数として用います。この関数のことを**ロジスティック関数 (logistic function)** と言います。ロジスティック関数そのものは，先の式から 1.7 という係数を除いた $\frac{1}{1 + exp(-x)}$ で表されるもので，$-\infty$ から $+\infty$ までのどんな数字が与えられても，答えを 0 から 1 の範囲に変換してしまう関数です。この特徴はとても便利です。というのも，結果が 0 と 1 の間に入るということは，比率を表していると考えられるからです。0/1 のバイナリデータが従属変数のときに，独立変数をこのロジスティック関数で変換してやれば 0 か 1 のどちらに近いか，どれぐらいの比率で 1 の目が出るかがわかります。項目反応理論も結果が 0/1(誤答/正答) ですから，「学力」のような目に見えない能力をロジスティック変換してやれば，結果が正答率になるというのは大変便利なのですね。

それではこのロジスティック関数をつかった項目分析の話に進んでいきましょう。

2.2.2 項目母数の特徴

ロジスティック曲線が累積正規分布の近似関数になっていること，テスト項目の分析には通過率を使って考えることを見てきました。とくに通過率の分析 (図 2.2) では，その項目が難しい設問だったのか，簡単なものだったのかを見ることができました。ロジスティック曲線もこの「項目の難しさ」を表現できるように，次のように拡張できます。

$$p(\theta) = \frac{1}{1 + exp(-1.7(\theta - b))}$$

左辺の $p(\theta)$ に含まれる θ は，潜在特性であり，ここでは標準化された学力のことを想定しますから，偏差値のようなものだと思ってください[*7]。$p(\theta)$ は能力 θ の人がこの項目に正答する確率=(通過率) です。

ここで b という変数が入ってきました。これが**困難度 (difficulty)** を表す指標です。$b = 0$ のときは標準正規分布と同じ形になりますが，$b = 1$ ならばこの式は右に，$b = -1$ ならば左に動きます。つまり $b > 0$ であれば難しく，$b < 0$ であれば簡単であることを表現していることになります。図 2.3 に困難度が $b = -1, 0, +1$ の時の曲線を書いてみましたので確認してください。このように困難度を表現するパラメータを追加したモデルを **1 パラメータ・ロジスティックモデル (One Parameter Logistic model)** といいます。実際のテストの回答パターンにたいし，各項目にこの曲線を当てはめて困難度を推定することで，項目を評価できるようになります。このように項目の特徴を描く曲線のことを**項目特性曲線 (Item Characteristic Curve, ICC)** といいます。

さらにパラメータを追加して，次のようにすると **2 パラメータ・ロジスティックモデル (Two Parameter Logistic model)** になります。

$$p(\theta) = \frac{1}{1 + exp(-1.7a(\theta - b))}$$

[*5] R では pnorm 関数を使って描きます。数式でいうなら，$\int_{-\infty}^{p} \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$ となります。

[*6] exp というのは指数関数で，$exp(x) = e^x$ のことです。ここで e は数学定数で，$e = 2.718282...$ という数字です。

[*7] 偏差値は標準化スコア z_i を $10z_i + 50$ と変換したものを指します。ここはその変換前の z_i と同じです。

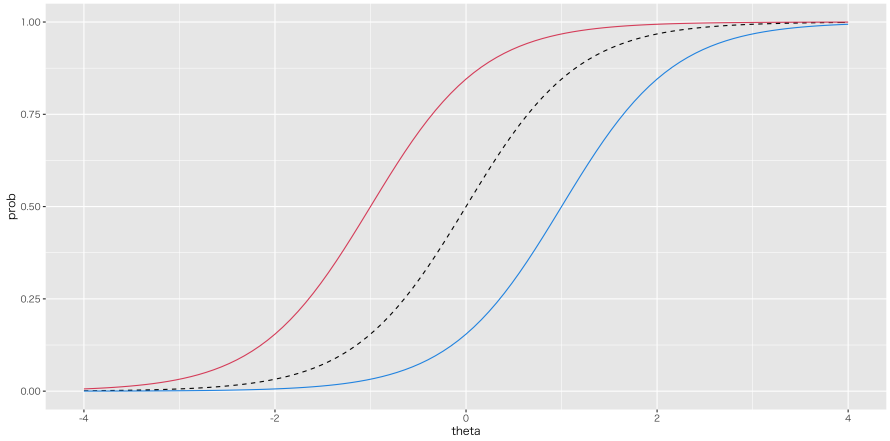

図 2.3　困難度母数の入ったロジスティック曲線。1PL ロジスティックモデルともいう。点線が $b = 0$ の標準的曲線。赤が $b = -1$, 青が $b = +1$ の例

ここで a という母数 (パラメータ) が入ってきました。これは $a = 1$ だともとのモデルのままなのですが, これが小さくなると曲線が傾き, 大きくなると曲線のカーブが強くなります (図 2.4)。曲線が緩やかになると (図 2.3 の赤線), θ の違いに対して通過率の変化が乏しくなります。言い換えると感度が悪くなるわけです。逆に曲線の立ち上がりが強くなると (図 2.4 の青線), θ がある一定のところを超えるかどうかで正答率がグッと変化することになります。つまりこのパラメータは, 回答者の能力 θ を分類する力の強さを表しているのです。このパラメータのことをとくに**識別力 (discriminant)** と呼びます。

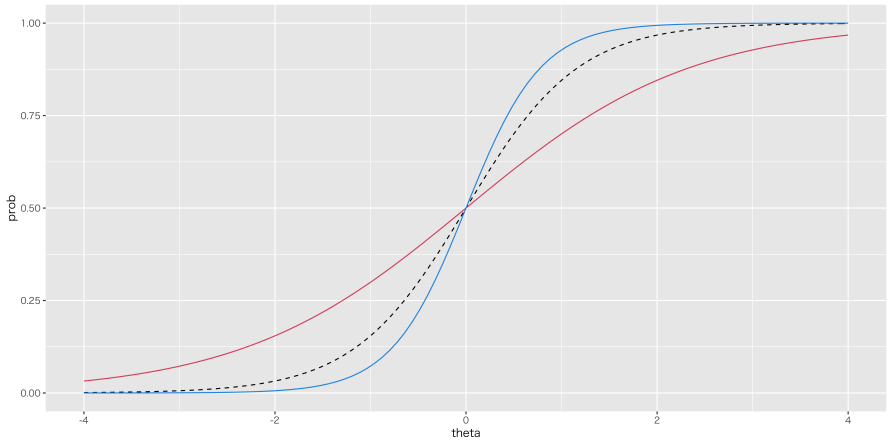

図 2.4　識別力母数の入ったロジスティック曲線。2PL ロジスティックモデルともいう。点線が $a = 1$ の標準的曲線。赤が $a = 0.5$, 青が $a = 1.5$ の例

一般にはここまで紹介した 2PL モデルがよく用いられます。他にも 3，4，5 つとパラメータが増えたモデルもありますが[*8], もとのロジスティック曲線に特徴を付け足していったもので, 曲線をずらしたり, 曲げたりしながらもとのデータにうまく当てはまるようにしつつ, その特徴を解釈できるように工夫しています。

いずれにせよ, テストの結果から各項目の特徴を記述します。少し例示したほうがわかりやすいでしょう。

[*8] 詳しくは http://antlers.rd.dnc.ac.jp/~shojima/exmk/jindex.htm を参照してください。3 つめのパラメータはあて推量母数，4 つ目は上方漸近母数，5 つ目は非対称母数と呼ばれています。

図 2.5 にはとある心理統計の授業で行った試験の結果から，2PL ロジスティックモデルを当てはめて項目分析をした例です[*9]。同じデータの項目母数を表 2.1 にも示しました。表 2.1 の数字と図 2.5 の曲線の対応をよく確認してください。

たとえば，項目 I0022 は困難度が-1.92，識別力が 1.30 です。困難度がマイナスですので，これはかなり簡単な問題だということになります。具体的には，偏差値 50 すなわち $\theta = 0$ の人であっても 98.58% 正解するわけですから，ほとんどの人にとって簡単な問題であることがわかります。ちなみにこの問題，具体的には帰無仮説検定に関する問いで，「「差がない」「偏りがない」といった仮説は何と呼ばれるか」というものでした[*10]。

一方，困難度が 0 近いところの例として項目 M0605 をあげますが，これが平均的な難易度の質問になっています。困難度母数 $b = 0.0$ であれば偏差値 50，すなわち $\theta = 0$ の人が正答する確率が 50% の質問ということになりますが，今回の M0605 はそれより少し難しいので，偏差値 50 の人で 20.70% の割合で正答できます。この問いについて偏差値が 70($\theta = 2$) であれば，92.96% の確率で正答できることになりますし，偏差値 30($\theta = -2$) であれば 0.51%，つまりほとんど正答は望めないということになります。ちなみにこの問題は「重回帰分析において，標準化されたデータを使って分析をすることで，モデルの適合度を上げることができる」を Yes か No かで判断させるという質問でした。

他にも項目 I0017 は困難度が 2.17，識別力は 0.69 です。困難度がもっとも高いグラフで，図の曲線は一番右にあるラインになっています。ただ識別力がやや低いので，曲線はよりなだらかになっていますね。困難度が高いので，$\theta = 0$ でも 7.24% しか正答できません。難しい！ $\theta = 2$ で 45.10% ですから，かなり能力の高い人でも半分は間違えるような問題です。ちなみにこれは標本分散の期待値が母分散からどれぐらいずれるのかを計算する問題でした。

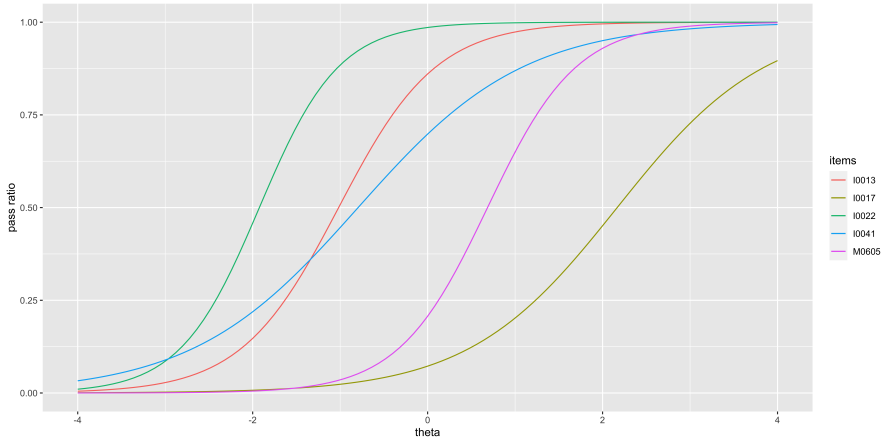

図 2.5 実際のテストに 2PL モデルを適用した例

2.2.3 被験者母数の推定

項目反応理論における潜在特性の推定は，項目の特徴を表す項目母数に対して**被験者母数**と呼ばれ，上述の項目特性に基づいて行われます。先ほどの図 2.5 をもとに説明します。ある回答者が項目 I0013 に正

[*9] 私が本務校で行っている統計の授業では，過去の受講生のデータとさまざまな質問をプールしたデータを貯めてあります。このデータに基づき項目の困難度，テストの平均値を事前に設計し，一貫した基準で評価できるように心がけています。

[*10] ちなみに答えは「帰無仮説」です。

表 2.1 各項目の困難度と識別力

	項目 ID	困難度	識別力
1	I0013	-1.02	1.05
2	I0017	2.17	0.69
3	I0022	-1.92	1.30
4	I0041	-0.79	0.62
5	M0605	0.68	1.15

答したとしましょう。その人の能力値 θ はどの辺りにあるかといえば，確率の曲線に沿った下の領域のどこかということになります (図 2.6)。この曲線，ICC は項目の特徴を表したもので，ある θ の値の能力があればどの程度の確率で正答できるかを表した項目の特徴でもありますが，逆にある人の θ がどのあたりにありそうかを示しているともいえます。たとえばこの ICC において，θ が 0.5 のときの通過率は 86.05% ですが，言い換えればこの項目に正解した人が $\theta = 0.5$ である可能性も高そうです。$\theta = -2$ の通過率は 14.71% ぐらいですから，能力がこんなに低いとは思えませんし，1 つの項目の話でしかないですが，希望的観測をするなら θ がもっと高い可能性もあるでしょう[*11]。

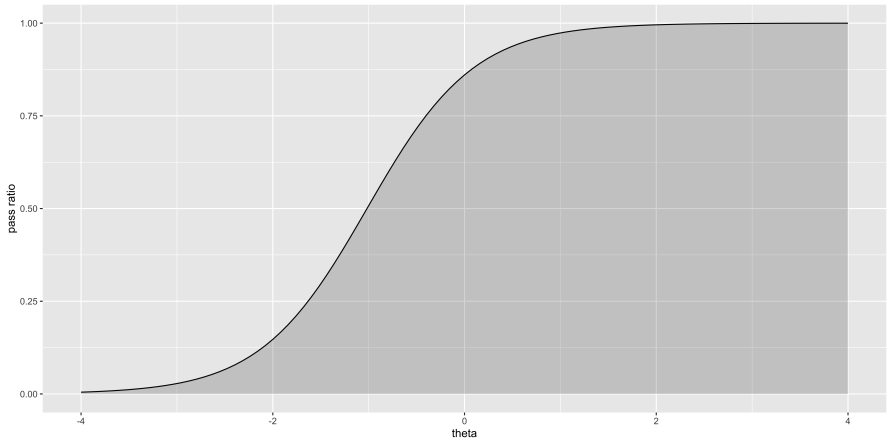

図 2.6 実際のテスト項目 I0013 に正答した人の能力がありそうな領域

次に，困難度のより高い項目である I0017 には**誤答**したとしましょう。その人は，I0017 の ICC の下の領域には**ない**はずです。項目 I0013 の ICC の下で，かつ，項目 I0017 の ICC の上にあるはずなので，図 2.7 のように塗りつぶされた領域の中に入ります。

この 2 つの曲線の間にある，濃く彩られた領域の幅が，回答者の能力値がありそうな程度を表しているのです。図 2.7 には $\theta = 0$ の可能性の大きさを矢印で示してありますが，θ の値はここだけに限らずこの曲線の幅のどこかです。ただ $\theta = 2$ や $\theta = -2$ よりは $\theta = 0$ のほうが，より「ありそう」な値だということがわかります。

ここでさらに同じ人に，項目 M0605 の質問をして，この人がそれにも正解したとしましょう。この人の能力 θ のありそうな範囲はさらに絞り込むことができます (図 2.8)。項目 I0013 よりは難しい質問に正解したわけですから，$\theta = 0$ の可能性はグッと小さくなり，それよりも $\theta = 2$ ぐらいの方がありそうだ，ということになっ

[*11] θ のありそうな「確率」とはいってないことに注意してください。すぐにわかることですが，この ICC の下の面積を積分しても 1.0 にはなりませんのでこれは確率ではなく，**尤度 (likelihood)** のほうなのです！

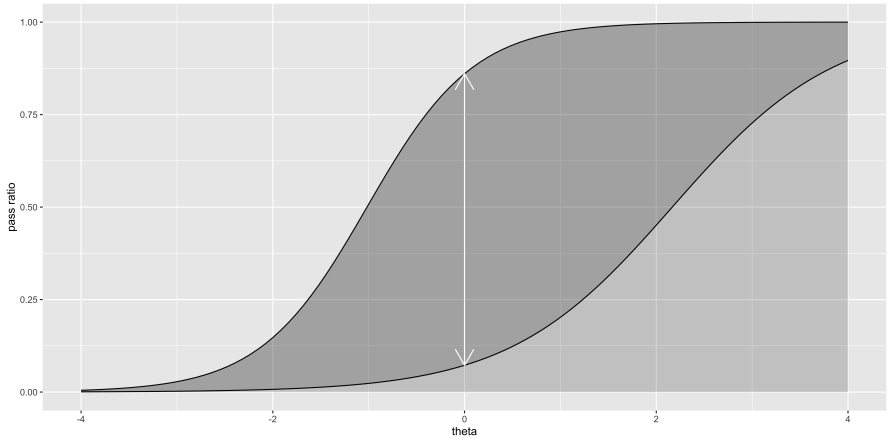

図 2.7　つぎのテスト項目 I0013 には誤答した人の能力がありそうな領域

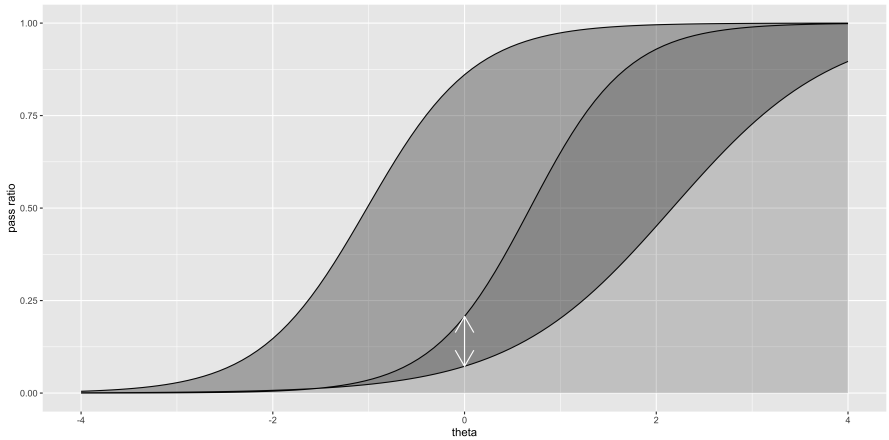

図 2.8　さらにテスト項目 M0605 には正答した人の能力がありそうな領域

てきます。このように，1 つ 1 つの項目からこの人の能力 θ がどのあたりにありそうか，というのを絞っていき，テストに含まれる項目全部を使えば，かなり狭い領域で「この辺りにあるはずだ」と推定できるでしょう。

この IRT を用いた推定方法は，このようにテストの項目ごとにその特徴を分析できるのが特徴です。テストの中でも良い項目，悪い項目というのはあるでしょうが，どのように悪いのかを困難度や識別力といった項目の特徴を使って表現できます。これらの数字は，テストに含まれる項目ごとの難易度を相対的に比較していく中で作られるものであり，回答者の能力に依存するものではありません。テスト理論が被験者の特徴と項目の特徴を分離するところから始まったことを，改めて思い出してください。

また，このような項目同士の比較から求められた項目の特徴をつかって，被験者の能力 (潜在特性) を推定する方法についても紹介しました。その過程の中で気づいたと思いますが，複数の回答を通じてある回答者の能力 θ のありそうな領域が狭められていく中で，明らかにその人にとって簡単すぎる問題，難しすぎる問題は意味を成しません。たとえば図 2.8 の段階まで絞り込まれているときに，項目 I0013 よりも簡単な I0022 を出題しても，おそらくほぼ確実に正答し，そのことは「領域を狭める」ことにはなんの貢献もしないでしょう。回答者の能力に相応しい質問を選んで出すことができれば，とても効率よくその領域を絞り込んでいくことができます。こうした考え方は，**コンピュータ適応型テスト (Computer Adapted Test)** として実装されます。

ここではテスト理論について紹介してきましたが，この考え方は後ほど性格テストなどで用いられているリッカート形式の尺度に応用できるように発展します。

2.3 現代テスト理論の特徴

前回は現代テスト理論として項目反応理論 (IRT) を紹介しました。ロジスティック曲線を応用して項目の特徴を描画し，それを使って被験者母数を推定する方法についても解説しました。この一連の手続きに基づき，現代テスト理論の利点を考えてみたいと思います。

2.3.1 現代テスト理論の利点 1: 項目母数と被験者母数の分離

古典的テスト理論からの発展として，現代テスト理論では被験者母数 θ_i と項目母数 a_j, b_j を区分して考えるようになりました。項目母数は通過率のアイデアを精緻にしたものですが，この通過率は項目群の総和を元に考えられていたことを思い出してください。すなわち，項目母数の計算には項目の相対的な困難度だけを用いています。イメージとしては鉱物の硬度検査のようなものです。2 つの異なる硬さの物をぶつけて崩れた方が負け＝より硬度が低いと考えるように，2 つの異なる項目を被験者に与えて，より正答者数が少ない方がより難しいと考えるのです。これはつまり，回答者の学力が高かろうが低かろうが，困難度が $b_x < b_y$ であるという関係に違いはないという考え方です。

これまでのテストや心理尺度の作り方に比べると，この点が大きく違います。学校などのテストは教員が作っていますから，教員が自分の感覚で「こちらの方がより発展的な内容だ」「こちらの方が難しいだろう」という問いに大きな配点がなされたりするでしょう。その後テストの平均点をみて「今回のテストは簡単にしすぎたか」という判断をしたりするでしょう。しかし IRT では項目それ自体に困難度を決めさせますから，そこに作成者や回答者の意図は含まれません。平均正答率が高いからといって簡単な問題なのではなく，項目の特徴として困難度が決まるのです。

たとえばサーストン尺度の作り方を思い出してください (セクション 1.3.1,Pp.10 参照)。サーストン尺度では尺度適用前に評定者集団によって尺度値を決めます。この評定者集団が偏った思想の持ち主だけで固められていた場合，尺度の点数は極端なものになり，普通の人がその尺度に回答すると極めて低い点数，高い点数になってしまうかもしれません。あるいはリッカート尺度の作り方を思い出してください。(セクション 1.3.2,Pp.12 参照)。リッカート尺度では回答者の累積度数から尺度値を算出します。先ほど同様，回答者集団の態度に偏りがあれば，尺度の点数は標準的な物ではなくなるでしょう。つまり「誰を対象に測定するか」によって目盛りが変わるようなものです。これでは測定結果の一般化は難しいでしょう。たとえばある大学で作られた尺度を，他大学でやってみると違った尺度値になるのですから，研究結果はせいぜい「その大学ではそうなんだろう」となり一般化できなくなります。従来の方法は，回答者と項目の特徴が関連しすぎていたのです。

これに対し，IRT を使って項目の特徴を計算する場合は，相対的な難易度に違いはありませんから，どの大学で作った尺度であっても統一的な解釈が可能です。尺度作成時に幅広くデータを集め，項目母数を確定させてしまえばどこででも統一的な評価ができます。テストなど学力を測定する際に大学間での違いが見られたとしても，その難易度を調整するのも簡単で，共通する項目を入れておけばそこを基準に相対的な困難度調整ができます[*12]。良問と悪問の評価と，回答者の評価を分けることは重要なポイントなのです。

[*12] こうしたテスト間の困難度調整のことをテストの**等価 (equation)** といいます。

2.3.2 現代テスト理論の利点 2: 被験者母数の推定の利点

被験者母数と項目母数の分離は，さらに別の利点も生み出します。それはデータが一部欠落した場合の補完に関係します。

リッカート尺度では回答者全体の相対頻度から，カテゴリの尺度値を決定するのでした。ここである人が特定の項目にだけ回答をし忘れたとします (調査研究ではよくあることで，同じような目盛りが並んでいると一行飛ばして丸をつけちゃうようなことはよくあります)。そうすると，その項目だけ合計人数が変わりますから，計算が面倒です。また相関係数を計算する時にも，その人のその項目については計算できなくなります。相関係数を考える時に，一箇所でも欠測値があるとその人のデータを抜いてしまうか[*13]，他の値を代入して補完するか[*14]，手間でも計算の時にそこだけ外して計算するか[*15]といった工夫が必要です。相関係数に基づいて指標化するときは，欠測値があればその人について算出できないことに変わりはありません。

それに対して，IRT の被験者母数の推定は，項目母数をつかった ICC をもとに一人ずつ絞り込んでいくというものでした。もしある人が回答していないことがあっても，計算ができなくなることはありません。その項目の情報が得られないので絞り込み精度は上がりませんが，ヒントが減っただけで回答できないわけではないのです。このように，得られた情報すべてを使ってその人の被験者母数を推定する方法のことを，**完全情報最尤推定 (Full Information Maximum Likelihood)** といいます。このように IRT を使うと，必ずしも全問に回答していなくてもスコアは計算できるということになります。欠測値があるからその人のデータは使えない，ということがないのでいいですね！

もっというと，IRT では全員が全員同じ問題に回答する必要はありません。たとえば能力値が $\theta_i = -0.3$ ぐらいにありそうだ，と絞り込めている段階で，次の問題の困難度母数が $b_j = 2.5$ であれば，おそらくほぼ確実にその人は回答できないでしょう[*16]。その人にいくら困難度母数の高い質問を繰り返しても，ほとんど誤答がつづくだけで，とくにその人の θ がありそうな領域を狭めるヒントにはなりません。むしろ $b_j = -1$ とか $b_j = -0.5$ のような簡単な問題を出して[*17]，それらに正答できるかどうかを見極め，絞り込んで行った方が効率的です。紙に印刷されているテスト (Paper Based Test) であれば，印刷された問題は変えようがありませんから，困難度順に問題を並べると，あるところから先はずーっと不正解が続く人が続出します。ずーっと不正解なところの問題はいくら良問でも，その人の能力を測るのには役立ちません。ヒントが増えないからです。であれば，回答者の学力に合った問題を，その都度その都度出題した方がいいですよね。コンピュータを使って回答者に相応しい質問をダイナミックに組み替える，コンピュータに基づいたテスト (Computer Based Test)，別名**コンピュータ適応型テスト (Computer Adaptive Test)** というのがそれです。CAT になると，回答者ごとに出題が変わりますからカンニング対策の必要も無くなって，とても便利になること間違いなしです。

2.3.3 現代テスト理論の利点 3: 信頼性についての考え方

IRT の考え方からいうと，どんな項目でも何らかの情報を提供してくれるはずです。たとえば $b_j = 3$ のようなとても難しい項目が合ったとします。これがどれぐらい難しいかというと，偏差値 70 の人でも 15% しか

[*13] リストワイズ削除といいます。

[*14] 欠測値補完については，平均値や中央値を代入したり，同じようなパターンで回答している人の値を使い回したり，回帰分析でその項目の値の推定値を入れたり，とさまざまな方法が考えられてきました。欠測値発生のメカニズムにもよりますが，いずれもある程度バイアスのかかった値になってしまいます。統計的によりバイアスの少ない適切な代入法が考えられてはいますが，そもそも欠測値がないのが最も望ましいことに変わりはありません。

[*15] ペアワイズ削除といいます。

[*16] 2PL モデルで $a = 1, b = 2.5$ のとき，$\theta = -0.3$ が正解する確率は 0.8492% です。

[*17] 2PL モデルで $a = 1, b = -1$ のとき，$\theta = -0.3$ が正解する確率は 76.674%，$b = -0.5$ のであれば 58.419% です。

正解できないレベルです。ほとんどの人にとっては誤答にしかならない難しすぎる悪問だ，といいたくなるかもしれませんが，学力が高い人がどれほど高いレベルでやれるのかを検証するためには必要な問題です。偏差値が 70 なのか，75 なのか，80 までいけるのか，といったことを見極めるためにはこの問題でないと情報が得られないことになります。逆に簡単すぎる問題であっても，より低いレベルを精緻に検証するためには必要なものなのです。

つまり，どの項目にもその項目が得意とする領域があるはずです。この項目はこの領域の回答者を絞り込む時に，最も有用な情報をもたらしてくれるはず，という θ の場所があるはずなのですね。これを表現するのが**項目情報曲線 (Item Information Curve)** といい，次の式で表される項目情報関数で描くことができます。

$$I(\theta) = a_j^2 p_j(\theta) q_j(\theta)$$

ここで $p_j(\theta)$ はその項目 j の θ における正答率 (通過率)，$q_j(\theta)$ は誤答率を表しています。能力が平均的，すなわち $\theta = 0$ のときは，0.5×0.5 になる平均的な困難度 ($b = 0$) の設問が最も大きな値になる，というわけですね。図 2.9 にいくつかの ICC とそれに対応する IIC を描きましたので，確認してください。IIC の

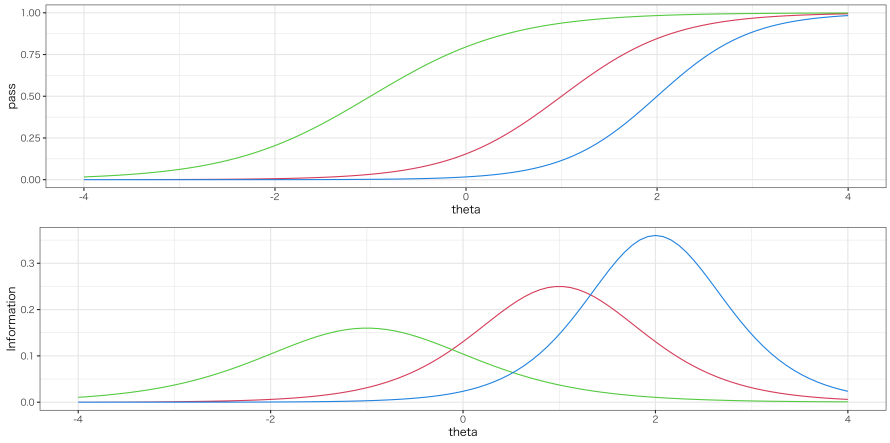

図 2.9　項目特徴曲線 (ICC, 上図) と項目情報曲線 (IIC, 下図) の例。左から順に $a_1 = 0.8, b_1 = -1$ の識別力が弱く簡単な項目，$a_2 = 1, b = 1$ のやや困難な項目，$a_3 = 1.2, b_3 = 2$ の識別力が高く困難な項目。

ピークは，対応する ICC の $\theta = 0.5$ のところにあること，識別力はピークの尖り具合に関わっていることを確認してください。

さて，IIC はその項目がどこで情報をもたらしてくれるか，ということを表しているのでした。言い方を変えると，IIC のピークはその項目のもっとも信頼できるところであるともいえます。つまり IRT において**信頼性は潜在特性の関数**になっているのです。古典的テスト理論ではテスト全体の分散に占める真分散の割合のことを信頼性というのでした。IRT では項目ごとにこれを考え，さらにその項目のもっとも感度の良いところを探る関数として，その信頼性を評価できるようになったといえるでしょう。

また IIC はある項目から得られる情報のことを意味しますが，テストに含まれているすべての情報関数を足し合わせることで，そのテストから得られる情報の大きさを関数として評価できます。すなわちテスト全体の情報量 $I_T(\theta)$ は，$I_T(\theta) = \sum_{j=1}^{M} I_j(\theta)$ であり，この関数のことを**テスト情報関数 (Test Infromation Curve)** といいます。図 2.9 の 3 つの項目からなる TIC を示したのが図 2.10 です。これを見ると，この 3 つの項目か

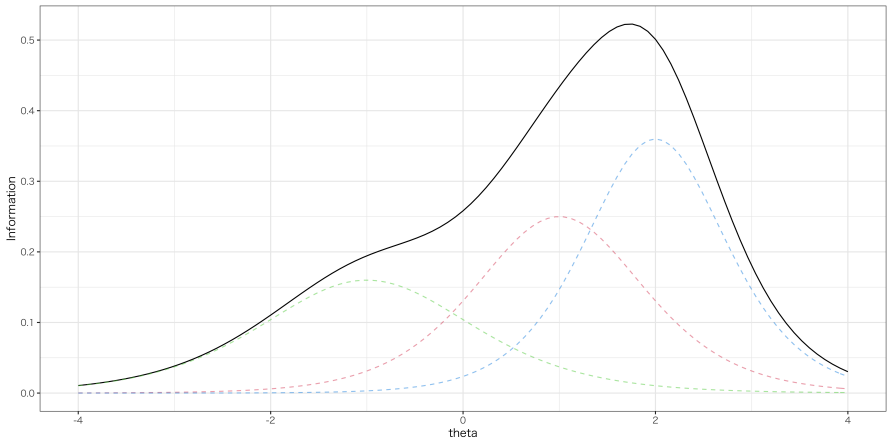

図 2.10　テスト情報関数 (黒の実線部)。理解を進めるために各 IIC を薄い点線で表現した。

らなるテストは $\theta = 2$ より少し低いレベルを測定するときに最も鋭敏に働くということがわかります。このように，項目母数がわかっていれば事前にテストの特徴をどのあたりに持ってくるかを決定でき，自由自在にテストをデザインできるようになるわけです。

テストの例で話をしていますが，心理学的な領域でももちろん便利な手法です。たとえば高い認知能力レベルの人をとくに選出したいとか，精神的な健康度がごく低い人をしっかりと検出したいといった目的があれば，そのあたりにピークが来るような項目からなる質問項目からなる調査票を構成すれば良いのです[*18]。

2.3.4　現代テスト理論の問題点

ここまで見てきたように，IRT はさまざまな利点があります。しかし欠点がないわけではありません。

ここまでの話はすべて，項目母数が既にわかっていれば，という前提付きで進めてきました。では項目母数はどのようにして定めるのでしょう？ これはもちろん得られたデータから算出できるのですが，そのためには事前に多くの被験者から回答を集め，項目母数の値はほぼ間違いなくこれぐらいだろう，といえるほど安定したものである必要があります。テストの場合，回答が 0/1 というバイナリデータで得られますから，そもそもそれほど情報がある反応ではありません。バイナリデータからその項目の特徴を安定して推定するためには，かなり多くの被験者 (数千から数万単位) を集めて項目に回答させておく必要があります。テスト項目は一度使ってみないと，項目母数がどうなるかわからないというのもポイントです。

また，CAT など項目をダイナミックに組み合わせるためには，選べるぐらいさまざまな項目を準備しておかなければなりません。項目を集めたものを**項目プール (Pool of Items)** といいますが，これも数千から数万の単位で用意しておく必要があります。なぜなら，テスト項目は事前に 1 回は使ってあるわけですから，数えるほどしか項目がなければ受験生が正答を事前に丸暗記できてしまうからです。もちろん項目プールが数千から数万あっても一度どこかで使われていますから，過去問をすべて完全に丸暗記すればその人は満点が取れてしまいます。もっともそれだけのものを覚えられるのは，ある意味学力が高いといっても差し支えないと思いますが。

日本でおこなわれる大学入学共通試験をはじめ，試験問題というのは極めて厳重な管理下に置かれ，事前にその情報が漏れることは公平性の観点から言って不適切であるとされています。しかし IRT で分析す

[*18] 具体例として小杉 (2014) をあげておきます。学校適応感を測定するため，テストのピークがやや低いところに来るようになっています。

るためには，事前に項目の特徴を知っていなければならないのです。CAT をつかって入学試験などができれば，カンニング対策にもなりますし，受験生は何度でもチャレンジできるので利点も多いのですが，「公平性のために新しいテストでなければならない」となるとなかなか実用化できないところがあるというのも事実です[*19]。

ところで，心理学の場合は学力テストのように正答・誤答があるものではなく，「当てはまる」から「当てはまらない」といった軸上で多段階の反応を求めることが一般的です。それでは多段階反応に拡張されたモデルを見ていきましょう。

2.4 段階反応モデル

リッカート法などで作られる尺度は，一般に 5，7 段階のものが多くあります。少ないものでは 3 件法[*20]であったり，ものによっては 4，6 件法であることもあります[*21]。しかしこれらの段階はいずれも順序尺度水準の情報しか持っておらず，そのままでは尺度値として使うことができません。シグマ法などで数値化すれば良いのですが，その手間を省いて分析する悪い習慣もあることは既に指摘した通りです。

IRT の多段階版はこうした問題に対応できる方法です。IRT の多段階モデルは大きくわけて 2 つあり，1 つは**段階反応モデル (Graded Response Model; GRM)**(Samejima, 1997)，もうひとつは**多段採点モデル (Partial Credit Model)**(Muraki, 1992) と呼ばれています。どちらも発想は似たようなところがあり，ここでは GRM について解説します。興味がある人は，豊田 (2012) や加藤 et al. (2014) など専門書を参考にしてください。

GRM の考え方の基本は，段階反応の背後には正規分布する連続的な潜在特性 θ がある，と仮定するところです。心理的な能力，学力，性質などは連続的なのですが，それが表に出てくる時は離散的 (順序的) だというわけです。図 2.11 に図示されているように，正規分布がある**閾値 (threshold)**(これを b_k と表しますが) を超えると出現する時は次のカテゴリになる，ということを考えます。図は三段階の例ですが，図から明らかなように k 段階であれば閾値の数は $k-1$ 個あることになります。横軸 θ は心理的な態度や性質の強さだと思ってください。さてそうすると，$\theta = 2.0$ ぐらいであれば，ほぼ間違いなく「当てはまる」に回答することになりますし，$\theta = -2$ であれば「当てはまらない」に回答するようになるはずです。このように θ が大きくなればなるほど最後のカテゴリに反応する確率は上がっていきますから，ここは 2PL モデルの時のようにロジスティック曲線で「当てはまる」に回答する確率を表現できるでしょう。問題は，それより下の段階に反応する確率をどのように表現するか，です。

ここである段階に反応する確率を考えるために，少し表現を改めます。すなわち，個人 i の項目 j に対する反応が，カテゴリ k **以上**になる確率として，$P^{+}_{jk}(\theta) = P(x_{ij} \geq k|\theta)$ をまず考えます。ここで $k = 0, 1, 2, ..., K$ とします。先ほど示したように，$k = K$，すなわち一番上のカテゴリ (ここでは「当てはまる」) に回答する確率は，2PL ロジスティック関数と同じ形をしていますから，次のように表現できます。

$$P_{jK}(\theta) = P^{+}_{jK}(\theta) = \frac{1}{1 + exp(-a_j(\theta - b_{jK}))}$$

ここで右辺の a_j は**識別力**，b_{jK} はカテゴリ K の**困難度**を表しています。左辺の $P_{jK}(\theta)$ は θ の人が項

[*19] 令和 2 年度に大学入試センター試験から大学入学共通試験に変わりましたが，改革前の計画では CAT 化することが盛り込まれていました。しかし実際には，受験生のためのコンピュータやタブレットを準備したり，安定した通信網が必要であったり，というハード的な問題もあって見送られてしまいました。

[*20] たとえば YG 性格検査は 3 段階です。

[*21] 偶数の段階にすることで，必ずどちらかの極に寄るようにして集計できます。日本人は「どちらでもない」に回答しがちな中点集中傾向があるとも言われているので，わざと肯定・否定のどちらかに寄せようという考え方です。

目 j のカテゴリ K に反応する確率で，それは k 以上に反応する確率 $P^{+}_{jk}(\theta)$ と一致していることを表しています。

次に，もっとも低い段階に回答する確率を考えましょう。これは θ が大きくなればなるほど減っていくはずで，いわばロジスティック曲線の逆のような形をするはずです。反応確率は最大でも 1.0 ですから，逆ということは 1.0 から引いてやればよいでしょう。

$$P^{+}_{j0}(\theta) = 1.0 - \frac{1}{1 + exp(-a_j(\theta - b_{j0}))}$$

問題は「どちらでもない」に反応する確率です。これは引き算で考えることができます。すなわち「どちらでもない」以上に反応する確率から，「当てはまる」以上に反応する確率を引いてやれば良いのです。

$$P_{jk}(\theta) = P^{+}_{jk}(\theta) - P^{+}_{j,k+1}(\theta)$$

ここにあるように，段階数が増えたとしても k 番目のカテゴリ以上に反応する確率から，$k+1$ 番目に反応する確率を引いてやれば，k 番目のカテゴリに反応する確率が得られます。この計算をして描かれる曲線のことを**項目反応カテゴリ特性曲線 (Item Response Category Charactaristic Curve; IRCCC)**，あるいは単に**カテゴリ確率曲線 (Category Probability Curve)** と呼ばれます。IRCCC は図 2.11 の下段に示されています。

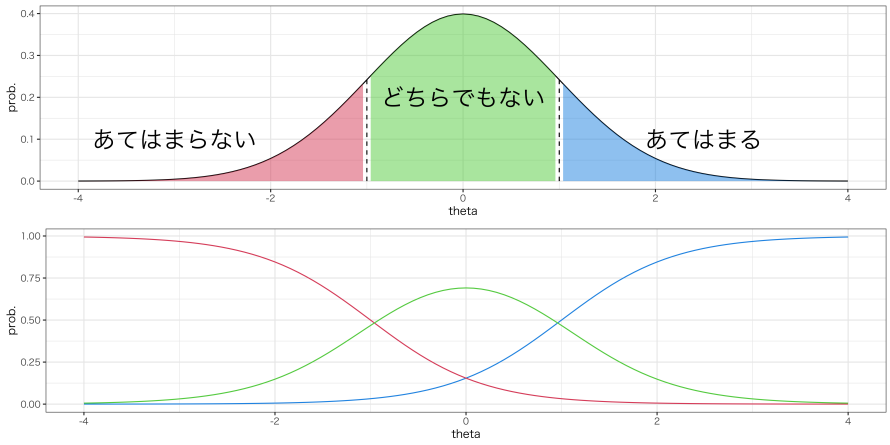

図 2.11 正規分布と閾値 (上図) と IRCCC(下図)

IRCCC を，θ の値がマイナスからプラスの方向に動かしながら見ていってください。最初は当然「当てはまらない」に反応する確率が一番高いのですが，それが徐々に下がっていきます。「どちらでもない」の反応確率は徐々に増えていき，閾値 b_{j1} で「当てはまらない」と逆転しピークを迎えることになります。その頃「当てはまる」も徐々増えていき，閾値 b_{j2} でピークが逆転する，というようになります。ピークは逆転されても，他の確率が 0 になっているわけではなく，可能性は残っています。また IRCCC も ICC 同様に変換して，情報曲線に帰ることができます。すなわちどのあたりで鋭敏に情報を検出できるかを表現する項目反応カテゴリ情報曲線を描くこともできます。このようにして，段階反応でもその項目の特徴をデザインできるのです。

2.4.1 適切な反応段階を考える

実際の調査研究をすると，図 2.12 の上の段のような IRCCC が描かれることがあります。何かおかしいところがありませんか？これは 5 段階の反応モデルですが，4 番目の反応カテゴリがずいぶん低く，そのピー

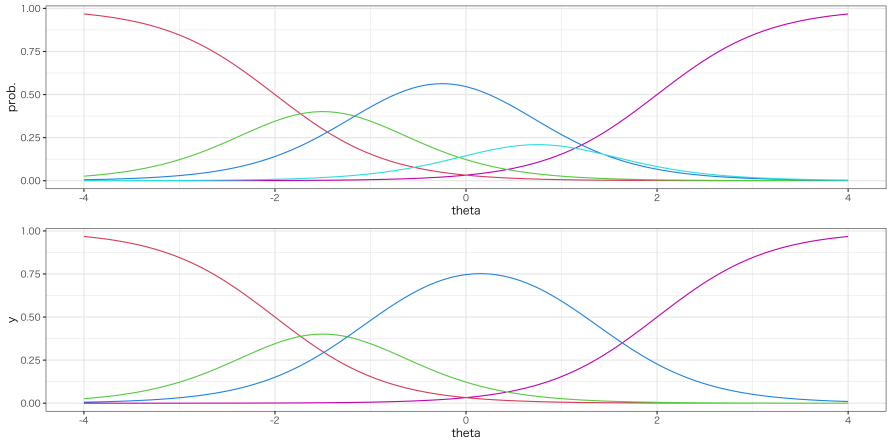

図 2.12　適切な反応段階をデザインする

クが 3 番目と 5 番目の反応カテゴリに潰されてしまっていますね。つまり，4 番目の反応がもっとも際立つシーンがないということです。言い換えるならば，これは尺度作成側が 5 段階だと思っていたにもかかわらず，回答者はどういう時に「やや当てはまる」と答えるのかがはっきりせず，実質 4 段階でしか反応していないことを表しています。

このような IRCCC が描かれてしまう場合は，$k = 3$ の反応と $k = 4$ の反応を合わせてひとつにしてしまうなど，段階の修正を考えると良いでしょう。具体的にはデータで 4 と入力していたものを，3 に置き換えたりします[*22]。修正したのが下の図になります。このように修正しても，情報関数は変わりません。同じデータから得られる情報は同じだからです。

このように 5 段階，7 段階を設定して回答者に無理やり回答を求めても，分析するとカテゴリのピークが潰れていることがあります。回答者の反応しやすいカテゴリ数を丁寧に設計してやることが重要です。もちろん無分別に尺度値をつけて，そのまま分析するのはもっとも不適切な方法です。

[*22] 4 を 5 に書き換えても構いません。ヒストグラムを見て，より正規分布っぽくなるようにすると良いでしょう

第 3 章

因子分析法

3.1 因子分析モデル

3.1.1 単因子モデル

今からお話しするのは，**因子分析 (Factor Analysis)** というモデルです。因子分析モデルは古典的テスト理論の拡張であり，もっとも簡単な 1 因子モデルは次のように表されます。

$$z_{ij} = a_j f_i + e_{ij}$$

ここで Z_{ij} は個人 i の項目 j に対する反応を標準得点で表したもの[*1]，a_j は項目 j の**因子負荷量 (factor loading)**，f_i は個人 i の**因子得点 (factor score)**，e_{ij} は個人 i と項目 j の組み合わさった時に生じた誤差と呼ばれます。

因子負荷量 (factor loading) とは，因子というこのテストで測定したい特性と，項目との関係の強さを表しているものです。**因子得点 (factor score)** とは，因子というこのテストで測定したい特性と，個人との関係の強さを表しているもので，その人のスコアだということができます。

記号についている添字に注目してください。添字 i は個人を，添字 j は項目を表していますが，因子負荷量は a_j と表されています。つまり項目によって変わる変量です。因子得点は f_i と表されています。つまり人によって変わる変量です。古典的テスト理論をこの添字を使って表現するならば，$X_{ij} = t_i + e_{ij}$ となりますが，これと比べてみると t_i が $a_j f_i$ に変わったのが因子分析モデルだということになります。t_i は個人についての真のスコアなのですが，古典的テスト理論の場合，テストの点数は個人のこの能力だけを反映していると考えられていたことになります。もしテストの問題が難しすぎて，まったく答えることができなければ，その人の能力はゼロということになるわけです。しかし中には悪い問題というのもあって，たとえば小学生に高校で習う知識が必要な問題を解かせるような問題があれば，誰だって解けないかもしれません。解けない問題を出して「学力が低い」と結論づけるのはやや暴力的ですらありますね。このように，古典的テスト理論は項目の良し悪しといったものが評価できないモデルだったのです。

因子分析モデルはこれを改良し，$a_j f_i$ としました。すなわち，ある項目に対する反応 z_{ij} は，その項目が測定したい特徴を十分に反映しているかどうか (a_j) と，その人がその特徴を有しているかどうか (f_i) の両方が成立している必要があるわけです。掛け算ですから，一方がゼロであれば結果もゼロになります。すなわち測定したい特徴を反映していない項目 $(a_j = 0)$ であれば，どれほどそれについての能力 (f_i) が高くても反応

[*1] **標準得点 (Standard Score)** とは，素点 X_j を $Z_j = \dfrac{X_j - \bar{X}}{\sigma_j}$ と変換したものであることを思い出してください。標準化されたスコアは平均が 0，分散が 1 になりますので，単位の異なるもの同士であっても標準得点を使うと相互に比較可能になるのでした。

できないのです。たとえば美的センスに非常に秀でた人がいても，数学のテストでその能力を反映させることはできませんよね。これは数学のテストというのが数学力を測定するものであって，美的センスを測定するものではないからです。

因子分析は知能検査や性格検査の文脈から生まれてきたものです。心理学において「知能」とは，何にでも応用可能な一般的な知能というのがあるのか，あるいは語彙力や計算力といった複数の個別の能力があるのか，という議論がありました。知能検査としていろいろなものが考えられますが，それらがきちんと当該能力を測定する検査法だったかどうかもわからないわけで，因子分析モデルはそこを評価できるようにした，ともいえます。性格検査についても同様で，特性論的に考えるならば人間の性質というのは複数のもの，たとえば外向性，神経症傾向，開放性，協調性，勤勉性[*2]などがあり，ある性格検査の項目は協調性を測定するのには向いているけれども，神経症傾向を測定するには向いていないということがあるわけです。このように，心理学と因子分析モデルは深い関係があります。

3.1.2 多因子モデル

さて先ほどは一因子，あるいは単因子ともいいますが，測定したい特徴が 1 つだけのモデルでした。学力テストなどは一因子で問題ありません。国語のテストは国語の能力を，数学のテストは数学の能力を測定していれば良いのであって，数学のテストを解くのに美的センス (真美的能力) が必要というのは，むしろ困った状況ですらありますね。しかし，知能検査や性格検査の場合はそうではありません。ある行動傾向，ある形容詞，ある検査がたった 1 つの能力・性質・心理的要因だけを反映しているとは限りません。たとえば人に優しく振る舞うといっても，その背後に外向性があるのか，あるいはそうすると自分がよく見られるからという利己的な性格があるのか等々が考えられます。1 つの項目に複数の要素 (因子) が複合的に影響していることを考えるべきです。そこで因子の数が 1 つではなく，複数ある多因子モデルを考えることにします。多因子モデルは次のように表現されます。

$$z_{ij} = a_{j1}f_{i1} + a_{j2}f_{i2} + a_{j3}f_{i3} + \cdots + a_{jm}f_{im} + d_j u_{ij} \tag{3.1}$$

記号の意味は単因子モデルと基本的には同じです。z_{ij} は個人 i の項目 j に対する反応を標準得点で表したものであり，a_{jm} は第 m 因子の**因子負荷量**，f_{im} は第 m 因子の**因子得点**を表しています。因子の数が複数あるモデルですから，$a_{j1}, a_{j2}, \cdots, a_{jm}$ とか $f_{i1}, f_{i2}, \cdots, f_{im}$ のように因子の番号と項目・個人の添字の組み合わせになっていることを確認してください。最後の e_{ij} が $d_j u_{ij}$ となっていますが，これは誤差についても他の因子と形を同じくし，項目に依存するものとそれ以外に区別しているだけです。

さて，ここでは因子を m 個あるとしています。この因子はどの項目にも共通して働くので，**共通因子 (common factor)** と呼ばれます。共通因子がいくつあるかは事前にわかりませんが，一般的にその数は数個〜十数個になります。性格心理学は長い研究の中で，性格を表す言葉に共通する因子はおおよそ 5 つぐらいであろう，という答えを得るに至りましたが，それ以外の領域では領域ごとの見解があるでしょう。知能が何種類の因子に分かれるのか，あるいはとある心の状態がどういう構造をしているのか，というのは心理学的にみても十分興味のある考え方です。もちろん因子分析によって得られる因子が，人間の潜在的な知能や概念に直接対応しているとはいえないのですが[*3]，それでも因子がどのような構造 (しくみ) をしているのかについての一定の情報を与えてくれます。多因子モデルが心理学一般で広まったのは，こうした心の「構造」に注目する学問との相性が良かったということでしょう。

[*2] 小塩 (2020) のビッグファイブについての解説に基づいています。

[*3] むしろテスト項目や調査票などに対する反応パターンが因子として出てくるだけで，性格や知能が数次元あるというより，我々は性格や知能を数次元で捉えることしかできない，という方が正しいでしょう。詳しくは第 6 章で議論します。

3.2 因子分析の定理

3.2.1 因子分析モデルの展開

因子分析モデルも古典的テスト理論のように，式の展開から何が見えてくるか考えてみましょう[*4]。

左辺の z_{ij} は観測されたデータから算出できるものですが，右辺の因子負荷量，因子得点はいずれも未知数です。データに対して未知数が多すぎるようで，これではどのようにして答えを見つけ出せば良いのかわからないかもしれません。たとえばある人のある項目に対する標準得点が 0.12 であるとして，それが 0.4×0.3 で得られるのか，0.2×0.6 で得られるのか，はたまた他の数値の組み合わせで得られるのか，を解く数学的技術は存在しません。この方程式はこのままでは解けないのです。

そこで，この未知数だらけの方程式を解くために，因子について以下のような条件を置きます。

- 共通因子の因子得点，独自因子の因子得点は，標準化されている。すなわち，いずれの因子得点も平均点は 0 であり，分散は 1 である。
- 独自因子は共通因子，他の独自因子と相関しない。

この他に，状況に応じて因子同士の間に相関を仮定します。

- 共通因子同士の相関を認めないのを「直交因子モデル」，認めるのを「斜交因子モデル」と呼ぶ。

このような仮定を置いたら問題が解けるようになるのでしょうか？ 実はこの問題を解く鍵は，多変量データであればなんとかなるのです！

2 つの変数，j と k の標準得点から，

$$r_{jk} = \frac{1}{N}\sum_{i=1}^{N} z_{ij}z_{ik} \tag{3.2}$$

のように，相関係数が算出されることを思い出してください。先ほどの因子分析の基本代数式 (式 3.1) をこの式に代入してみましょう。

$$\begin{aligned} r_{jk} &= \frac{1}{N}\sum_{i=1}^{N} z_{ij}z_{ik} \\ &= \frac{1}{N}\sum_{i=1}^{N}(a_{j1}f_{i1} + a_{j2}f_{i2} + \ldots + a_{jm}f_{im} + d_j u_{ij})(a_{k1}f_{i1} + a_{k2}f_{i2} + \ldots + a_{km}f_{im} + d_k u_{ik}) \end{aligned} \tag{3.3}$$

これは代数の計算としてやっていくと，非常に煩雑で間違いが起きやすそうです。そこで，少し視覚化してわかりやすくしてみましょう。多項式の掛け算は，各項目の総当たり戦ですので，列方向に z_{ij}，行方向に z_{ik} の各要素を置いて，要素同士の組み合わせ表を作ります (図 3.1)。

図 3.1 に示されたのは個人 i についてのものであり，これが人数分ある，すなわち $\sum_{i=1}^{N}$ をつけないといけないことに注意してください。さて図を軽く色分けしてあるのですが，ここにあるように計算すべき領域を 4 つに分けて考えていきましょう。

①**の領域** 因子 p と p の積和部分 (同じ因子同士の掛け合わせ)
②**の領域** 因子 p と q の積和部分 (異なる因子同士の掛け合わせ)

[*4] 以下このセクションは小杉 (2018) の原稿を再構成したものです。

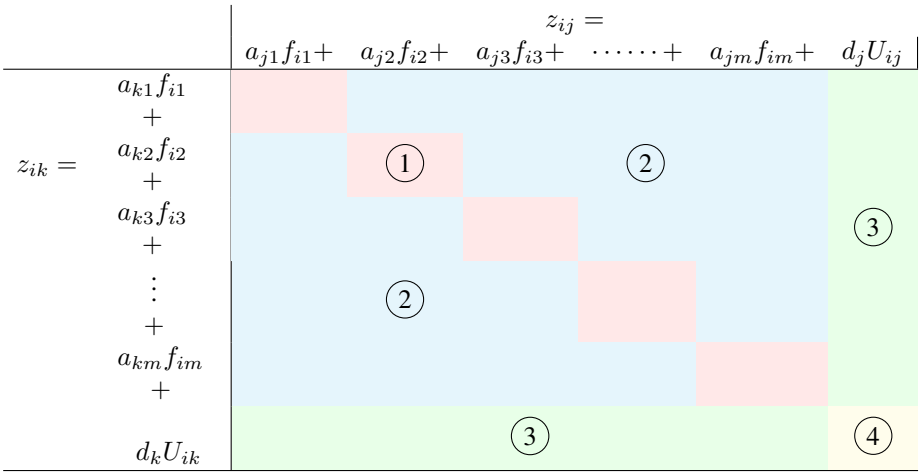

図 3.1　項目同士の総当たりを考える

③**の領域**　因子 $p(q)$ と独自因子の積和部分
④**の領域**　独自因子同士の積和部分

この各パートを順に計算していきましょう。まず①ですが，たとえば第一因子については

$$\frac{1}{N}\sum_{i=1}^{N} a_{j1}a_{k1}F_{i1}^2 \tag{3.4}$$

となることがわかります。ここで，a_{j1} と a_{k1} は N には関係がない ($\sum$ は i が 1 から N まで変化することを意味しているが，係数 i はどちらにも入っていない) ので，総和して割る意味がないことに気づきます。そうなると，必然的にこの式は，

$$\frac{1}{N}\sum_{i=1}^{N} a_{j1}a_{k1}F_{i1}^2 = a_{j1}a_{k1}\frac{1}{N}\sum F_{i1}^2 \tag{3.5}$$

となります。この F_{i1} は因子得点であり，上の仮定より標準化されたものだということになります。さて，標準得点と標準得点の積和平均は相関係数になることをもう一度思い出してください！そうすると，これは自分自身との相関係数を表していることになりますから，当然 $F_{i1}^2 = 1.0$ であることがわかります。

結局，①のエリアは

$$\frac{1}{N}\sum_{i=1}^{N} a_{j1}a_{k1}F_{i1}^2 = a_{j1}a_{k1}\frac{1}{N}\sum F_{i1}^2 = a_{j1}a_{k1} \tag{3.6}$$

と，とてもあっさり書き下すことができるのです。

つづいて②を見てみましょう。ここは異なる因子がかけ合わさった部分ですね。落ち着いて，第一因子と第二因子を例にして考えてみましょう。この箇所で得られるのは，

$$\frac{1}{N}\sum a_{j1}F_{i1}a_{k2}F_{i2} = a_{j1}a_{k2}\frac{1}{N}\sum F_{i1}F_{i2} \tag{3.7}$$

ということになります。ここで，F_{i1} および F_{i2} はそれぞれ第一，第二因子における個人 i の因子得点を意味しています。因子得点は標準化されていることをもう一度思い出すと，これは第一因子と第二因子の相関係数になります。ここで，この因子分析が直交因子モデルだと考えますと，因子同士に相関がないわけですから，数字としては 0.0 で消えてしまいます。するとこの部分は，

$$\frac{1}{N}\sum a_{j1}F_{i1}a_{k2}F_{i2} = a_{j1}a_{k2}\frac{1}{N}\sum F_{i1}F_{i2} = 0 \tag{3.8}$$

となることがわかりました。つまり，この領域(2)は，すべて 0 になってしまうのです。

続いて(3)の部分について考えてみましょう。これはある共通因子と独自因子の積和部分です。例によって標準得点同士の関係から，相関係数を算出することになりますが，独自因子は共通因子と無相関であることを考えると，

$$\frac{1}{N}\sum a_{i1}F_{i1}d_jU_{ij} = a_{ik}d_j\frac{1}{N}\sum U_{ij}F_{i1} = 0 \tag{3.9}$$

とこのように，この箇所もすべて 0 になってしまいます。

最後の(4)に至っては，独自因子と独自因子の積和ですから，これも

$$\frac{1}{N}\sum d_jd_kU_{ij}U_{ik} = d_jd_k\frac{1}{N}\sum U_{ij}U_{ik} = 0 \tag{3.10}$$

のように 0 になります。結局，消えて無くなるのがほとんどで，残るのは①の部分だけであり，r_{jk} を考えるときはそこだけ考慮すればよいことになります。

整理すると，

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jm}a_{km} \tag{3.11}$$

ということがわかります。つまり，項目 j と項目 k の相関係数は，項目 j の因子負荷量と項目 k の因子負荷量を，すべての因子について総和したものであるということです。因子分析の基本モデルから導出されるこの定理を，とくに**因子分析の第二定理**と呼びます。

ここで同じ項目同士の相関を考えてみましょう。項目 j と項目 j の相関係数は，もちろん 1.0 になりますね。これを因子分析の基本式で表すと，次のように表現できます。

$$r_{jj} = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + d_j^2 = 1.0 \tag{3.12}$$

さて，この式が意味するのはなんでしょうか。意味を考えてみると，ある項目それ自身との相関係数は，因子負荷の二乗和からなっている，ということがわかります。これこそ**因子分析の第一定理**と呼ばれるものであり，解けるはずのなかった方程式を解くための鍵となる式なのです。

3.3 因子分析の定理

数式の展開はいったんここまでにして，第一定理は次のような形をしているのでした。

$$a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + d_j^2 = 1.0$$

ここで共通因子部分を，$a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 = h_j^2$ のようにすると，この式は単に $h_j^2 + d_j^2 = 1.0$ となります。この h_j^2 のことをとくに**共通性 (communality)** といいますが，この式は共通性と独自因子の二乗和が 1.0 になることを意味しています。言い換えると，全体を 100% とした比率で共通性と誤差を比較できるということです。共通性は因子負荷量の二乗和で，共通因子はそのテストの背後にある共通の要因，すなわちテストで測定したいものだったわけです。古典的テスト理論では，モデル式の展開から信頼性を全分散中に示る真のスコアの割合と定義しましたが，因子分析モデルはこのように 1 つの項目 j における共通因子の割合を算出し，項目の信頼性を考えることができます。因子分析モデルにおける信頼性は，1 項目の中に含まれる共通因子の大きさだともいえるわけです。逆に $d_j^2 = 1 - h_j^2$ は**独自性 (uniqueness)** は，当該項目がそのテストで測っていないものの大きさを表しており，この割合があまりにも大きいと「この項目は全然関係ない

ものを測っちゃってるんじゃないか？」と疑われることになります。多因子モデルにおいては，多角的に対象を切り分けるために多くの質問を投げかけるわけですが，独自性の高い項目は回答者に負担をかけるだけの邪魔なものですから，実践上はこうした項目を除外することが少なくありません。因子分析には**単純構造の原則 (principle of simple structure)** と呼ばれるものがあり，項目は該当する因子を適切に反映し，かつ，他の因子と関係ないことが美しいとされます。尺度構成段階では，共通性 (独自性) をみて項目の良し悪しが判断されるのです。

次に第二定理を見てみましょう。第二定理は次のような形をしているのでした。

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jm}a_{km}$$

2 つの項目 j と k の相関係数は，それぞれの因子負荷量の積和の形で表される，というものです。ここに誤差の話は入ってこず，共通因子だけで話ができています。

左辺の相関係数は，2 つの項目がどれほど同じものを測定しているかの指標です。相関係数 (の絶対値) が高ければ高いほど，2 つの項目は同じものを指し示しているわけです。逆に相関係数が低いということは，2 つの項目に関係がないことを表します。ここで右辺に目をやりますと，右辺の各項目は因子負荷量の積の形になっています。左辺の値が小さくなる 1 つの理由は，ある共通因子 m が項目 j, k に対して，異なる方向で寄与しているからだと考えられるでしょう。そしてそのパターンが一貫していないという状況です。そもそも相関係数が小さいところからは因子を見つけ出すのは難しいのですが，そうした状況があるのはある項目ペアについて因子同士の向きがバラバラに影響しているからだといえます。そのような状況は，測定がきちんとできているかどうか怪しいですね。**測定の一義性**ともいわれますが，そのような尺度は妥当性が低いといえるでしょう。

このように，因子分析モデルは第一定理で信頼性を，第二定理で妥当性をあらわすものになっているのです。

3.4 因子分析の歴史と展開

ところで，因子分析モデルもテスト理論も，潜在変数モデルとしては同じでいずれも古典的テスト理論からの発展系なのでした。因子分析モデルは多段階反応が一般的で，多因子モデルで「潜在的な (心理学的な) 構造はどうか」ということを問題にします。ここでの目的は全体に共通する要素やその構造であり，何種類に別れて要素間の関係はどうなっているのか，というところが中心的関心事になります。一方，項目反応理論はバイナリ反応が一般的で，因子数はひとつです。学力テストはその学力が測定できていることが重要で，因子の構造よりも因子得点をより精緻に推定できることの方が重要だからです。因子分析の言葉で表すなら，因子得点をより精緻に表現しようとしているわけです。

さて，GRM は，項目反応理論の多段階モデルでした。実は GRM は因子分析モデルの発展系でもあるのです。因子分析は相関係数のモデルであったことを思い出してください。因子分析モデル自体は z_{ij} から始まっていましたが，変数同士の共分散 r_{ij} を考えるといくつかの仮定から**因子負荷量**だけのモデルに簡略化され，推定できるようになるのでした[*5]。この標準化された共分散，すなわち相関係数はピアソンの積率相関係数ともいわれ，間隔尺度水準以上の数値を使って計算されます。多段階の反応は順序尺度水準ですから，相関係数を計算するのは不適切で，そのまま因子分析することはできません[*6]。では**順序尺度水準**の相関係数は，というとスピアマンの順序相関などを思い出す人がいるかもしれませんが，より統計的なモデルで表現

[*5] 具体的にどうやって因子負荷量を算出するかは次回以降のお楽しみです。

[*6] できないのですが，尺度値に変換することもなく機械的に分析してしまう悪い習慣が蔓延っているのは何度も指摘している通りです。くどいと思われるかもしれませんが，私は憤っているのです。

しやすいものがあります。

順序尺度水準の変数 × 順序尺度水準の変数の相関は**ポリコリック相関係数 (polychoric correlation)** といいます。順序尺度水準の変数 × 間隔尺度水準の変数の相関は**ポリシリアル相関係数 (polyserial correlation)** といいます。ついでにバイナリ変数 × バイナリ変数の相関係数は**テトラコリック相関係数 (tetracholic correlation)** といいます。

これらの相関係数はいずれも，順序 (あるいはバイナリ) 変数の背後に連続体があると考え，潜在的な連続体 × 潜在的な連続体の相関係数を連続体のカテゴリが変わる閾値とともに推定するのです (図 3.2)。

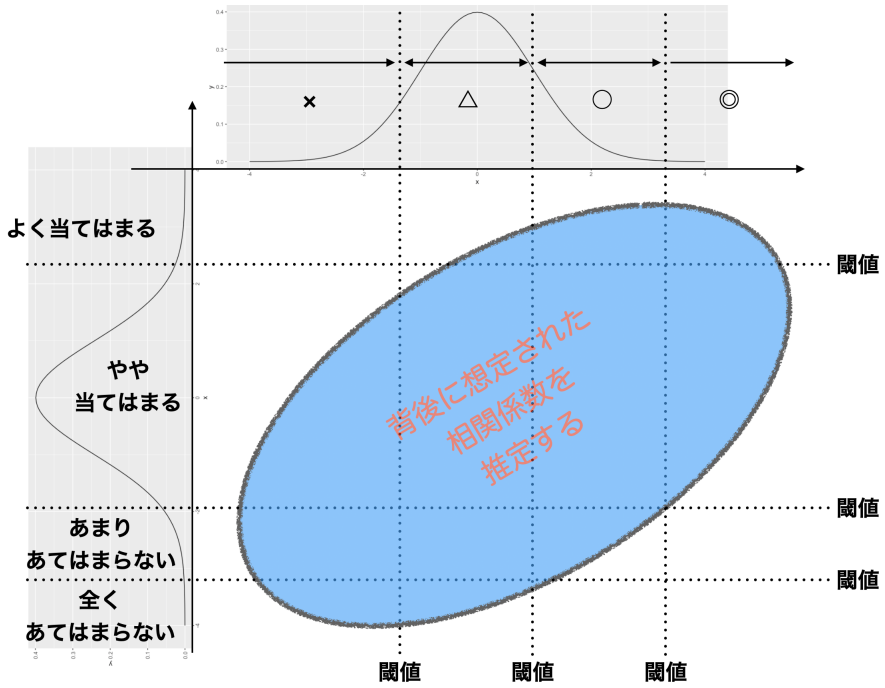

図 3.2 ポリコリック相関係数のイメージ

こうして推定された相関係数を使って因子分析を行うと，順序尺度水準に適した因子分析を行うことができます。この方法をとくに**カテゴリカル因子分析 (categorical factor analysis)** といいます。もっともこの名前を覚えておく必要はありません。**カテゴリカル因子分析は段階反応モデルと数学的に等価である**ことがわかっています。GRM をすることはカテゴリカル因子分析をしていることと同じ，なのです。

もっとも GRM は項目反応理論の系列ですから，単因子構造を仮定しています。複数の因子を想定するカテゴリカル因子分析に対応するのは，正確には**多次元項目反応理論 (Multidimensional Item Response Theory)** といいます。しかし数学的・技術的には同じであり，カテゴリカル因子分析をした結果から IRCCC を描くこともできますし，実際に分析するソフト上では使用する変数がカテゴリカル (順序尺度水準) であることを指定するだけです。われわれユーザはもはや悩む必要はなく，ただただデータに適した分析をするだけで良いのです。

3.4.1 系譜の違いはどこに関係するか

因子分析モデルと項目反応理論が，カテゴリカル因子分析として概念的に統合されることを話してきました。本質的にはこのように違いがないのですが，系譜の違い，出自の違いはそれぞれの利用される文脈で，何を強調するかに影響してくることがあります。

たとえば因子分析の文脈では，共通性が低い項目は削除し，綺麗な因子構造を目指そうという考え方があります。尺度作成の中で 1 つの項目は 1 つの因子に対応しているべきであるという考え方があり (**単純構造の原理 (Principle of Simple Structure)** といいます)，もし 1 つの項目が複数の因子の影響を受けているようであれば，「美しくないので」削除されることがあります。因子分析は知能，性格，態度の研究で展開されてきたため，美しい「構造」を見つけ出すことに狙いがありますから，この美しさを損なうもの (項目) は取り除く，という方向に行きがちです。

一方で，項目反応理論はテスト理論の生まれです。もちろん回答者の能力や特性を測定するのに優れた項目とそうでない項目，という峻別はしますが，中でも「この項目は測定能力の偏差値 30 程度の回答者を測定するのに適している」とか，「偏差値 70 程度の回答者を測定するのに適している」と判断できます。偏差値 30 や 70 を測定するのに適した項目とは，非常に簡単 (ほとんどの人が正答する) だったり，非常に難易度が高い (ほとんどの人が誤答する) 項目です。心理尺度でいうと床効果，天井効果がみられる項目とされる，どちらかに偏った分布をもつ項目です。しかし，それを捨てるということにはならず，どのような項目でも回答者の能力を推定するための情報量がゼロではない，という考えから，さまざまな項目をどんどんためていく方向にいきがちです。項目反応理論の文脈では，あらゆる人に対する測定を準備しておく必要があり，むしろ回答者の特性にあわせて設問の方を選んだり，事前の項目特性から，前もってテストの項目構成をデザインする，ということを目的とするのです。

このように使われるシーンによって，「構造」か「機能 (=得点)」のどちらに注目するかが変わり，結果的に実践の方針がちがってくることもあるのです。図 3.3 にこの心理学的系譜 (左ルート) とテスト理論的系譜 (右ルート) の流れを描いてみました。

ポイントは「最終的には同じところに辿り着く」という点ですから，歴史的流れや個々のモデルの細かい数式を完全に理解していなくてもいいかもしれません。それよりは，心理学者として，あるいはテストをする側として，回答者に無理のない，それでいて必要な程度に精緻な情報が得られる適切な分析方法を選択できるようになることが重要です。

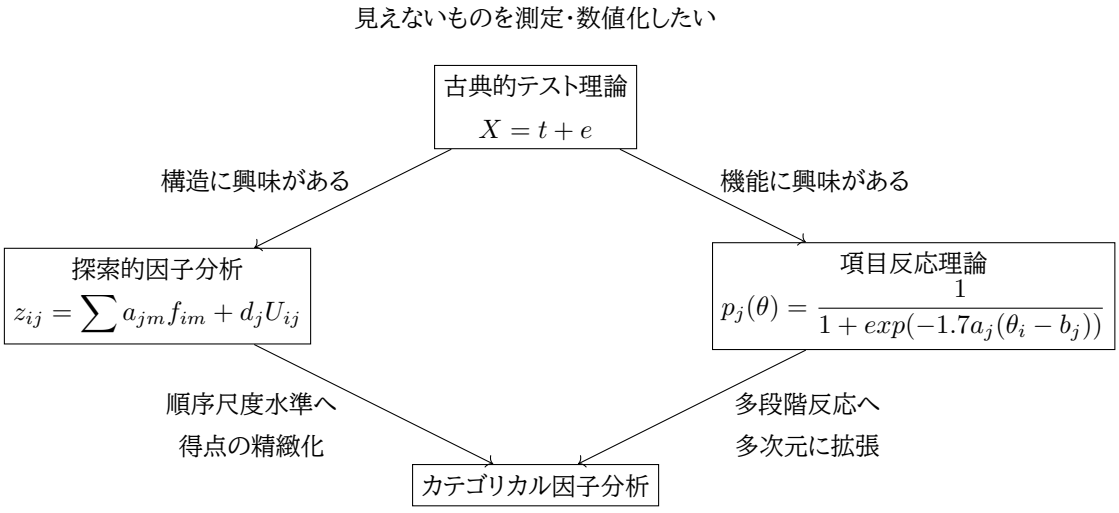

図 3.3　理論・モデルの流れ。左側のルートが心理学的系譜，右側のルートがテスト理論的系譜

3.5　調査研究の手順

最後に，一般的におこなわれている心理尺度作成の手順を大まかに理解しておきましょう。

■構成概念の設定　心理尺度を作り始めるにあたって，まずは何を測りたいのかを明確にする必要があります。**構成概念妥当性 (Construct Varidity)** についてのそもそも論ですね。ある概念を測定したいとして，そういう概念が本当に存在するのか，どこの誰がどのように持つ概念なのかを明確にする必要があります。たとえば文部科学省がスローガンのように掲げる**生きる力**というのを測ってみたいと思っても，それは何なのかわかりません。死んでなければ生きているのですから，生きる力はありそうに思えます。でもそういうことじゃないらしい。じゃあどういうことなのか，ということを考えて行かなければなりません。あるいは，みなさんは**いちびり**という関西弁をご存知ですか。いたずらっ子，わざと変なことをして注目を引きたい子，のような意味なのですが，このいちびり感は関西の人間にしかわからない感覚かもしれません。であれば調査対象者が限定的な，一般的ではない概念ですから，因子分析の行う多くの人の反応から共通成分を抜き出すという考え方には適さない概念です。より一般的なものに考え直す必要があるでしょう。

ほかにも，心理学的な態度なのか，パーソナリティのような行動の傾向なのか，抑うつ傾向のように心身両方の問題なのか，そしてそれらが紙とペンで測定可能なものなのかどうか，改めて考えましょう。何を当たり前のことを，と思うかもしれませんが，意外とおかしな問題設定に入り込んでないとも限らないのです。また，心理尺度や心理学的な概念はすでにさまざまなものが存在します。これらを見比べて，自分の測定したい概念が他の概念とどのように同じで，どのように違うのかを論理的に考えられなければなりません。どこかにある尺度とほとんど同じであればオリジナリティが存在しないだけでなく，車輪の再発明というムダな努力をすることになります。完全にオリジナルなものを考えるのは難しいかもしれませんが，新しい概念を使うことによってこれまでの概念で説明できたことを含み，さらにこれまでの概念で説明できなかったことも説明できるようになる，という利点が必要です。これらは**弁別的妥当性 (distincive validity)** にも関わる問題です。

■項目の選出　測定したい概念が明確になれば，これに関わる項目を作り出す必要があります。ある態度を強く持っている人はどのような振る舞いをし，どのような振る舞いをしないのか。どのような意見に賛成し，どのような意見には反対するのか，といったことをさまざまな角度から検証します。「目に見えないものを測定する」のが心理測定であり，質問紙調査というのはこの見えない的に向かって項目という矢を射掛けるようなものです。矢の数はなるべく多く，また人は嘘をついたり間違えたりする生き物ですから多角的に聞くことで，捉えるべき本質に迫っていく必要があります。言葉を使って聞くわけですから，古すぎる・新しすぎる表現は避け，専門用語は使わずなるべく平易な言葉で聞く必要があります。ワーディング (言い回し) についても細部まで注意しましょう[*7]。

■予備調査の実施　ある程度項目が集まれば，予備調査を行います。用意した項目の 5 倍から 10 倍の人を集めるのが良い，と一般的にいわれています (Grimm & Yarnold, 1994)。ここでの予備調査項目は，十分考えられたものではあると思いますが，分析してみると意外と不適切な項目というのが出てくるものですので，多めに用意されていることでしょう。必然的に，予備調査の対象サイズも百人以上の大きなものになるのが一般的です。

■探索的因子分析　ここが尺度作成に関わる分析のメインです。ここで因子数を決める (あるいは確認する) ことになり，因子的な妥当性みたり，不適切な項目は除外したりします。別の聞き方の方が良いような項目があれば，項目の入れ替えを行なって再調査することもあります。この探索的な因子分析と予備調査を繰り返して，何度行っても同じ因子構造になり，同じ項目が同じ因子に含まれるというのが確認できれば，項目セットは完成するといえるでしょう。

[*7] たとえば「テレビやラジををよく見聞きしますか」という質問は悪い質問です。テレビしか見ない人，ラジオしか聞かない人がどう答えていいか悩むからです。これはダブルバーレルと呼ばれる悪手ですが，ほかにも色々ありますので調査法の専門書を参考にしてみてください。

■項目反応理論　ある因子に関連する項目とそのデータセットを抜き出して，項目反応理論 (段階反応モデル) を実行しましょう。なぜ一部を抜き出すかというと，項目反応理論モデルは 1 次元性を仮定したモデルであることが基本だからです。これを多次元に展開した多次元段階反応モデルもありますので，直接そちらを使っても構いません。ともかくこれを行うことで，反応段階数を確認でき，また項目情報曲線やテスト上表曲線を書くことで，この尺度がどのような領域を得意とする項目群からなるのかを記述できることになります。

■本調査へ　最終的に項目群が確定したら，大規模な調査を行ってテストの**標準化**を目指します。ここでいう標準化とは，標準得点にする数値的演算ではなく，大規模調査に基づいてこの尺度がどのような得点分布をするかに基づき，ロウデータを採点結果としての点数に換算する手続きを明確化する，という意味です。因子構造などはほぼ確定しているはずですから，このテストを使うとどのようなデータの散らばりが得られるのかを確定することができるはずです。この最後のステップの目的はこのテストで測定される人の平均や散らばり，分布の形状を確認し，確定することにあります。使うときに逐一分析しなくてもいいように，項目回答パターンがどれぐらいであれば，全体のどのあたりに位置するかといったプロフィールをつくるのです。膨大なデータに裏付けられた採点なので，尺度の値が何らかの目安になり得るわけです。

以上が大体の流れになります。ここにあるように，予備調査を何度も繰り返して因子構造を確定し，そこから本調査を行うことで，ほぼこの尺度はどれぐらいの点数で分布するか，といったことがわかるようになります。かつては因子分析を 1 回実行するだけでも多大な時間がかかりましたから，心理尺度を作るというのはそれだけでライフワークになるような作業でした。幸い最近は分析のスピードが飛躍的に向上しましたので，簡単に分析してやり直すということができます。しかしだからといって，構成概念妥当性を疎かにしてはいけませんし，「やったらこうなった」というようなやり逃げ研究になるのではなく，使うための尺度を作るのだということは忘れてはいけないでしょう。

第 4 章

行列計算の基礎

これまで古典的テスト理論，因子分析論，現代テスト理論を通じて，目に見えない潜在変数を数値化する方法について学んできました。潜在変数という心のモデルは，心理学の中心的関心事であり，実際多くの調査研究で潜在変数をモデルに組み込んで検証されています。その割には，どういったメカニズムで潜在変数が見出されているのかについての理解は十分行き渡っていないようです。たとえば因子分析モデルは，統計パッケージを使うと瞬時に「3 因子構造で因子負荷量はこれこれ，因子得点はこのようになっています」と答えを出してくれます。しかし，なぜそのような数字になったのか，どのようにそれが算出されたのかを知らなければ，何もわかっていないのと同じではないでしょうか？ 因子は「機械がやってくれるもの」と思考停止してしまうと，結局のところ私たちの知りたいことには辿り着けませんし，誤用の元になってしまいます。

なぜその肝心の箇所が放置されているかというと，数学的には**線形代数 (linear algebra)** と呼ばれる計算が必要であり，そこについての文系数学的解説がないからです。線形代数はベクトルや行列の計算，文字と式の便利な表現形式です。これを知ることの利点は，**多くの数字のセットを簡単な記号で一般的に表現できるようになること**です。変数や回答者数が数十，数百，時には数万のサイズで得られた時，1 つ 1 つのデータにアルファベットを割り振っていたのでは間に合いませんので，線形代数はデータ解析には必須の知識です。

本講義では，線形代数の基礎を導入した上で，最終的には潜在変数，共通因子や**因子負荷量**と呼ばれるものがどのように算出されるのかを理解することを目的としています。事前の知識は必要なく，また目的に必要な最小限の知識だけで進めていきますので，一歩ずつ確実にフォローしてください[*1]。

4.1　行列とベクトル

行列やベクトルは，複数の数字をひとまとめにして扱うためのものです。まずはその基本的な形からみていきます。

■ベクトル　複数の数字を一行，あるいは一列にまとめて表現したものを，**行ベクトル (row vector)**，**列ベクトル (column vector)** といいます。

行ベクトルは次のように表します。

$$\boldsymbol{a} = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix}$$

[*1] ここからの話は小杉 (2018) の pp.148–179 に同内容のものがあります。もちろん線形代数のテキストとしては他にもいろいろあり，数学的な入門としては，基礎的には村上 et al. (2016) が，発展的なところでは永田 (2005) が参考になるでしょう。より文系のデータ解析的解説が多いのは，絶版になってしまいましたが岡太 (2008) が最高です。

列ベクトルは次のように表します。

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

具体的には，a_1 とか b_2 のところには数字が入っています。つまり次のような形です。

$$\boldsymbol{a} = \begin{pmatrix} 1 & 3 & 5 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 12 \\ 8 \end{pmatrix}$$

ここで今回の $\boldsymbol{a}$ は 3 つの要素が入っていますので，サイズは 3，同じく $\boldsymbol{b}$ はサイズが 5 のベクトルです。行列の言い方に合わせて 1×3 の (行) ベクトル，5×1 の (列) ベクトル，という言い方をすることもあります。このベクトルの中の数字は，とくに関係があるわけではありません。前に入っている数字がえらいとか，横にある方が重要だ，といったことはなく，ただただ数字をまとめて扱っているだけです。数字のセットを記号ひとつで表せるので，ずいぶん楽ですよね。

さて，行数も列数も 1 であるものつまり行列でない数字は，とくに**スカラー (scalar)** と呼びます。今までは $1 + 2 = 3$ といった計算をしていましたが，この $1, 2, 3$ はすべてスカラーだといえるわけです。

■行列　行列 (matrix) とは数を長方形に並べたものです。行列として並べられた数を成分といい，成分の横の並びを行，縦の並びを列と呼びます。

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

この例は，n 行 m 列の行列を表しています。お気づきかと思いますが，行列やベクトルを表す場合は，アルファベットを太字にするのが慣例です。たとえば，A とか x は 1 つの数字を表していますが，$\boldsymbol{A}$ や $\boldsymbol{x}$ であれば行列やベクトルを表していることになります。成分を表す文字は，一般に a_{ij} のように，はじめの添え字で行番号，次の添え字で列番号を表します。行列の大きさは行数と列数とによって，$n \times m$ のように表現します。n と m が同じ，つまり行数と列数が同じであれば，これをとくに正方行列といいます。正方行列の例を次にあげておきます。

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

正方行列の中でも，i 行 j 列目の値が j 行 i 列目の値と同じである行列 ($a_{ij} = a_{ji}$) のことを**対称行列 (Symmetric Matrix)** といいます。

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

この (正方) 対称行列の形は，データ解析の中ではよくでてきます。たとえば 3 つの変数 x_1, x_2, x_3 について，その相関係数を考えたいとしましょう。相関係数は 2 つの数字の組み合わせですから，x_1 と x_2，x_1 と x_3，x_2 と x_3 について計算でき，それぞれ r_{12}, r_{13}, r_{23} と表したとします。i と j の相関係数 r_{ij} は，j と i

の相関係数と同じ ($r_{ij} = r_{ji}$) であり，また $r_{jj} = 1.0$ なのは定義から明らかです。これを行列で表すと次のようになります。

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}$$

このように対称行列になっています。この行列をとくに**相関行列 (Correlation Matrix)** といいます。また相関係数は標準化された共分散でもありました。標準化するまえの相関行列は，**分散共分散行列 (Covariance Matrix)** と言います。その名前の通り，自分自身との共分散が分散になるわけですから，右上から右下にかけての対角線上にある要素 (これをとくに**対角 (diagonal)** 要素といいます) が分散であり，それ以外が共分散になっている行列です。

$$\boldsymbol{V} = \begin{pmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{pmatrix}$$

また，正方行列の中でもとくに対角要素にのみ値があって，それ以外はすべて 0 になっている行列のことを**対角行列 (diagonal matrix)**，対角行列の中でもとくに，対角項が 1 になっているものは**単位行列 (identity matrix)** と呼びます。単位行列は $\boldsymbol{I}$ とか $\boldsymbol{E}$ で表されます。

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

これは後ほど，掛け算をするときに「かけても変わらない状態」を表すために用いられます。

4.2 行列の四則演算と操作

行列の四則演算は，通常のスカラーのそれとは異なります。改めて，行列としての加減乗除を定義するのだと思ってください。

■加法・減法 まずは行列の足し算 (加法)，引き算 (減法) から説明します[*2] 。これはそれぞれ対応する位置にある成分を加え合わせる (減じる) ことで表されます。

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}$$

数値例をみておきましょう。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

これからわかるように，行列の加法，減法は大きさの等しい行列でないと成り立ちません。サイズが違うものを足そうとすると，演算できない箇所が出てしまうのです。このように行列では，「計算できない」という状態になることが少なからずあります。行列のサイズに注意が必要，ということがお分かりいただけるかと思います。

[*2] ベクトルは行列の中でも，行数あるいは列数が 1 のものですので，これで一般的に表現します。

■乗法 続いて掛け算です。まずスカラーと行列の積を見てみましょう[*3]。

$$\lambda \boldsymbol{A} = \boldsymbol{A}\lambda = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1m} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{n1} & \lambda a_{n2} & \cdots & \lambda a_{nm} \end{pmatrix}$$

実際の計算は，各成分をスカラー倍すればよいだけですので，比較的簡単ですね。

$$2 \times \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 \times 1 & 2 \times 2 \\ 2 \times 3 & 2 \times 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

次はベクトルとベクトルの掛け算です。これは形が変わってしまうので，注意が必要です。まずは行ベクトルに列ベクトルをかける例からみていきましょう。

$$\boldsymbol{ab} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{j=1}^{n} a_j b_j$$

掛け算なのですが，足し合わせるという計算プロセスが入り込んでいるので，結果はスカラーになります。掛け算なのにどうして足し算の要素が入るんだ，というクレームは，今はなしです。このように計算することに決めたことで，あとあと便利なことが出て来ますから，作法にまず慣れてからにしましょう。数値例も確認しておきます。

$$\begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} = 1 \times 3 + 2 \times 4 + 1 \times 2 = 13$$

ここで注意して欲しいのは，両方のベクトルのサイズが同じ ($1 \times n$ ベクトルと，$n \times 1$ ベクトル，いずれもサイズは n) ということです。サイズが違うと，演算が対応しない要素が出てくるので，計算できない，が答えになります。

今度は向きを変えて，列ベクトルに右から行ベクトルをかけてみましょう。

$$\boldsymbol{ab} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_n b_1 & a_2 b_2 & \cdots & a_n b_n \end{pmatrix}$$

今度は行列になりました。かける順番が変わるとサイズが変わる (ここでは，上の例では 1×1 のサイズ，下の例では $n \times n$ のサイズ) ことに注意してください。スカラーの計算では順番を入れ替えても，たとえば $2 \times 3 = 3 \times 2$ のように同じ答えになりましたが，行列の場合は必ずしもそうはならない，ということです。

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 \times 3 & 1 \times 4 & 1 \times 2 \\ 2 \times 3 & 2 \times 4 & 2 \times 2 \\ 1 \times 3 & 1 \times 4 & 1 \times 2 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 2 \\ 6 & 8 & 4 \\ 3 & 4 & 2 \end{pmatrix}$$

行列とベクトルの積や，行列と行列の積はこの応用になってきます。まず行列に列ベクトルを右からかける例を見てみましょう。結果は列ベクトルになります。

[*3] 式中にでてくる λ はギリシア文字でラムダといいます。小文字が λ，大文字では Λ と書きます。

$$\boldsymbol{Ab} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} a_{1j}b_j \\ \sum_{j=1}^{m} a_{2j}b_j \\ \vdots \\ \sum_{j=1}^{m} a_{nj}b_j \end{pmatrix}$$

ここでも掛け算なのに足し算のプロセスが入ってきています。注意深く記号を読んでみてください。数値例でも確認しておきます。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \times 2 + 2 \times 1 \\ 3 \times 2 + 4 \times 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$$

今度は行列に行ベクトルを左からかけましょう。結果は行ベクトルになります。

$$\boldsymbol{cA} = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{j1}c_j & \sum_{j=1}^{n} a_{j2}c_j & \cdots & \sum_{j=1}^{n} a_{jm}c_j \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 2 \times 3 & 1 \times 0 + 3 \times 3 \end{pmatrix} = \begin{pmatrix} 7 & 9 \end{pmatrix}$$

さて，最後に行列と行列の積を考えます。行列 $\boldsymbol{A}$ と $\boldsymbol{B}$ の積が成立するのは，前者の列数と後者の行数とが等しいときに限られます。行列 $\boldsymbol{A}$ のサイズが $n \times m$，行列 $\boldsymbol{B}$ のサイズが $m \times l$ とすると，その積は $n \times l$ の行列になります。計算手続きは，次のようになります，

$$\boldsymbol{AB} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{ml} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} a_{1j}b_{j1} & \cdots & \sum_{j=1}^{m} a_{1j}b_{jl} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} a_{nj}b_{j1} & \cdots & \sum_{j=1}^{m} a_{nj}b_{jl} \end{pmatrix}$$

どうにもこれはややこしいかもしれません。足し算や掛け算が入り乱れるし，計算途中でどの要素を計算しているかわからなくなるからです。ベクトルと行列の積の時のように，前の行列の要素は左に進み，後ろの行列の要素は縦に進みますから，左手と右手で違う図形を描く認知課題のように，そもそも混乱しやすい作業なのです。

しかし 2 つほど注意をしておくと，間違いにくくなります。1 つは積によって得られる**結果の行列サイズを意識すること**です。先ほど，前の行列の列数と，後ろの行列の行数が同じでないと計算ができないといいました。つまり，$n \times m$ 行列と $m \times l$ 行列でないと計算できない (m が同じ) ということです。また，結果は $n \times l$ 行列になります。前の行列の行数，後ろの行列の列数が結果のサイズです。ここに注目しておくと，計算を始める前に，計算が可能かどうかと結果の行列サイズは想像がつくのです (図 4.1)。

また，実際に計算する際は，**前の行列に横の，後ろの行列に縦の補助線を入れる**とわかりやすいかもしれません。こうすることで，間違えて計算を進めることがないようになるからです。

$$\left(\begin{array}{cc} 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & 6 \end{array}\right) \left(\begin{array}{c|c|c} 0 & 1 & 1 \\ 1 & 0 & 1 \end{array}\right) =$$

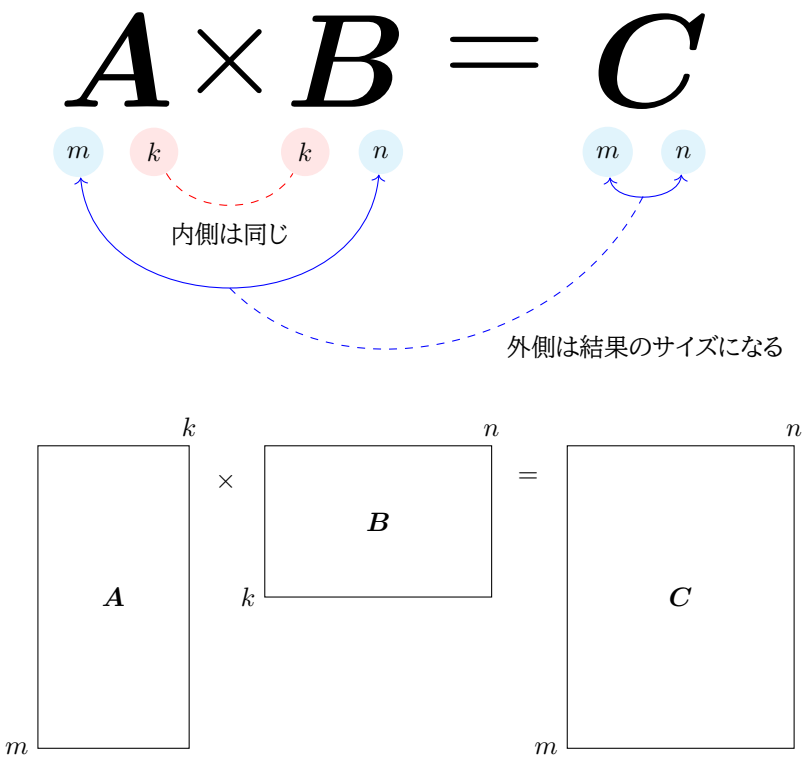

図 4.1 行列のサイズを意識する

$$\begin{pmatrix} 1\times 0+2\times 1 & 1\times 1+2\times 0 & 1\times 1+2\times 1 \\ 3\times 0+4\times 1 & 3\times 1+4\times 0 & 3\times 1+4\times 1 \\ 5\times 0+6\times 1 & 5\times 1+6\times 0 & 5\times 1+6\times 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 7 \\ 6 & 5 & 11 \end{pmatrix}$$

■転置 次に**転置 (transpose)** と呼ばれる操作を説明します。これは計算の便宜上，よく使われる行列操作のひとつです。

大きさ $n \times m$ の行列 $\boldsymbol{A}$ における i 行 j 列成分を j 行 i 列成分とする $m \times n$ 行列のことを，元の行列 $\boldsymbol{A}$ の転置とよび，$\boldsymbol{A}'$ や $\boldsymbol{A}^T$ と表します。行列を転ばせたようなイメージです。

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \text{のとき，} \boldsymbol{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

ベクトルも転置でき，行ベクトルを転置すると列ベクトルに，列ベクトルを転置すると行ベクトルになります。

$$\boldsymbol{a} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \text{のとき，} \boldsymbol{a}' = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

また，転置には以下のような性質があります。これは知識として知っておくだけでよいでしょう。

1. $(\boldsymbol{A}')' = \boldsymbol{A}$
2. $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$

3. $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$
4. $(c\boldsymbol{A})' = c\boldsymbol{A}'$

■逆行列　最後に**逆行列**のお話をします。逆行列は割り算のイメージです。ある行列にその逆行列をかけると単位行列になる，つまり割ると 1 になるような行列のことです。

正確に表現すると，ある正方行列 $\boldsymbol{A}$ に対し，$\boldsymbol{AX} = \boldsymbol{I}$ となるような行列 $\boldsymbol{X}$ が存在するとき，これを $\boldsymbol{A}$ の**逆行列 (inverse)** と呼び，$\boldsymbol{A}^{-1}$ で表します。正方行列でない場合に逆行列はありませんし，正方行列であっても逆行列が存在しない場合もあります。逆行列の例をみてみましょう。$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$ とすると，次の計算が成り立ちます。

$$\boldsymbol{AB} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

このとき $\boldsymbol{B}$ は $\boldsymbol{A}$ の逆行列，すなわち $\boldsymbol{A}^{-1} = \boldsymbol{B}$ といえます。

とくに対角行列の逆行列は，対角成分の逆数をそれぞれ対角成分とする行列になります。

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} \text{のとき，} \boldsymbol{D}^{-1} = \begin{pmatrix} \dfrac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{d_n} \end{pmatrix}$$

逆行列は，行列の世界の割り算のようなものです。これで一通り四則演算の定義ができました。

4.3 行列を使うと便利なこと

さて，ここまでで行列の計算の話をして来ましたが，どこが良いのかいまいちピンとこない，という人もいるかもしれません。そこで最後にどうしてこのような計算をするのか，何が良いのかを説明してみたいと思います。

4.3.1 行列と方程式

線形代数は「便利な書き方」の学問です。便利な書き方をするためにルールが作られていますから，ルールから学ぶと「なんでそんな変な操作をするんだ」という気持ちになるのもわかります。

では何が便利になるのでしょうか。これは方程式を解くことと関係があります。たとえば，以下のような連立方程式があったとしましょう。

$$\begin{cases} x - 2y - 5z &= 3 \\ 5x + 4y + 3z &= 1 \\ 3x + y - 3z &= 6 \end{cases}$$

これは行列で表現すると，次のようになります。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

この左辺を行列とベクトルの式の計算ルールにのっとって展開してみてください。ちゃんと最初の連立方程

式の左辺になることがわかると思います。かけて足して，という面倒な計算ルールは，連立方程式を簡単に表記するためのものだったのですね。

最終的にはこの方程式を解いて，次のように答えを求めます。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x=-1 \\ y=3 \\ z=-2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

皆さんも学校で習ったように，このような連立方程式を解く方法として，加減法や代入法というのがあります。ですがここはひとつ，行列を使った解法を考えて見ましょう。

そのような解法のひとつ，消去法は，ひとつの方程式を何倍かして，他の方程式に加えることにより，方程式をどんどん簡単にしていくというものです。まず，第一の式を 5 倍，あるいは 3 倍して，第二，第三の式から x の項を消去します。

$$\begin{cases} x-2y-5z &= 3 \\ -14y-28z &= 14 \\ -7y-12z &= 3 \end{cases}$$

第二の式の係数を簡単にしておきましょう。

$$\begin{cases} x-2y-5z &= 3 \\ y+2z &= -1 \\ -7y-12z &= 3 \end{cases}$$

第二の式を 7 倍して，第三の式から y を消去します。

$$\begin{cases} x-2y-5z &= 3 \\ y+2z &= -1 \\ 2z &= -4 \end{cases}$$

あとはこれの 3 行目から $z=-2$ が得られ，芋づる式に $x=-1$，$y=3$ が得られました。

この操作は，式を一本ずつ，あるいは 2 つの式を組み合わせて文字を消していく消去法を係数全体に行う操作になっています。実際，ここで操作される係数だけ見ていくと，次のようになります。

■第一段階

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & -7 & -12 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix}$$

■第二段階

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -4 \end{pmatrix}$$

さらにこの方法を改良した，ガウスージョルダンの消去法というものがあります。この手法による係数の変化を，行列表記で見ていくことにします。

まず第一段目は同じです。

$$\begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 2 \\ 0 & -7 & -12 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix}$$

次に，第二の方程式を用いて第一と第三の式から y の係数を消してしまいます。

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$$

最後に，第三の式の z の係数を 1 にして，第一，第二式の z の係数を消してしまいましょう。

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$$

最後の形を見ると，左辺は単位行列になっていますから，

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$$

と解が求められたことがわかります。ここで注目すべきは，連立方程式の解を求めるプロセスは係数行列を単位行列に変えていくプロセスだった，ということです。係数行列が単位行列になれば，それはもう答えを出したことになるのです。

さて，係数行列を $\boldsymbol{A}$ とすると，その逆行列 $\boldsymbol{A}^{-1}$ があれば $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ となるのでした。であれば，連立方程式の右辺にあったベクトルに $\boldsymbol{A}^{-1}$ をかけてやれば，一気に答えが求まるではないですか。

実際に見て見ましょう。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

この連立方程式に対して，次のような操作をします。

$$\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix}$$

とします[*4]。すると左辺は単位行列になりますから，次のように計算すれば一気に答えが求まることになるのです。

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$$

つまり，連立方程式を解くという問題が，係数行列の逆行列を求める問題になります。また，逆行列は存在しないこともある，ということでしたが，その場合その連立方程式は解けない，ということになります。

[*4] 数値的には $\begin{pmatrix} 1 & -2 & -5 \\ 5 & 4 & 3 \\ 3 & 1 & -3 \end{pmatrix}^{-1} = \begin{pmatrix} 15/28 & 11/28 & -1/2 \\ -6/7 & -3/7 & 1 \\ 1/4 & 1/4 & -1/2 \end{pmatrix}$ という行列です

第 5 章

因子分析の行列表現

5.1　データの行列表現

ここまで行列の形ばかり見て来ましたが，狙いはあくまでも調査研究など，多変量データを扱う場面での利用です。なぜ多変量データ分析をする際にこのような知識が必要なのか，思うかもしれません。ですが，得られるデータは行列として扱うと表現が大変便利なのです。たとえば質問項目が m 個あって，調査対象者 n 人から回答を得たとすると，データは次のように表現できます。

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

データ全体をこうして，ひとつの記号で表現できたら便利ですよね。これらを使ったデータの表記に慣れておきましょう。

各反応の平均点は以下のように表現されます。まず，要素がすべて 1 からなるベクトルを次のように表します。

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

わかりにくいかもしれませんが，この $\mathbf{1}$ は太字でベクトルを表しており，スカラーの 1 とは違うことに注意してください。

さて，各項目の和はベクトルの掛け算の定義によって次のように表現できます。

$$\boldsymbol{X}'\mathbf{1} = \begin{pmatrix} \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i2} \\ \vdots \\ \sum_{i=1}^{n} x_{im} \end{pmatrix}$$

これを使って平均値 (列) ベクトル $\boldsymbol{m}$ を次のように表すことができます。

$$\boldsymbol{m} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{1} = \begin{pmatrix} 1/n\sum_{i=1}^{n} x_{i1} \\ 1/n\sum_{i=1}^{n} x_{i2} \\ \vdots \\ 1/n\sum_{i=1}^{n} x_{im} \end{pmatrix} = \begin{pmatrix} \bar{x}_{.1} \\ \bar{x}_{.2} \\ \vdots \\ \bar{x}_{.m} \end{pmatrix}$$

ここで $\bar{x}_{.1}$ は 1 つめの添字 i を足し合わせて割ることでなくしていますから，$\bar{x}_1$ のように省略して書くことがあります。このときの 1 は「第一番目の変数」という意味であり，個人の情報がなくなっている変数を意味していることに注意してください。

さて，平均からの偏差を要素に持つ行列 $\boldsymbol{V}$ を考えたとします。

$$\begin{aligned} \boldsymbol{V} &= \boldsymbol{X} - \boldsymbol{1}\boldsymbol{m}' \\ &= \boldsymbol{X} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \end{pmatrix} \\ &= \boldsymbol{X} - \begin{pmatrix} \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{.1} & \bar{x}_{.2} & \cdots & \bar{x}_{.m} \end{pmatrix} \end{aligned}$$

この行列 $\boldsymbol{V}$ のサイズは $n \times m$ であることに注意してください。これはまた，次のように表すこともできます。

$$\boldsymbol{V} = (\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}')\boldsymbol{X}$$

ここで $\boldsymbol{I}$ は適切なサイズの単位行列です[*1]。

これを使うと，たとえば分散共分散行列 $\boldsymbol{S}$ は次のようになります。

$$\boldsymbol{S} = \frac{1}{n}\boldsymbol{V}'\boldsymbol{V} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1m} \\ s_{21} & s_2^2 & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_m^2 \end{pmatrix}$$

ここで s_j とあるのは第 j 変数の標準偏差を，s_{jk} とあるのは第 j 変数と第 k 変数の共分散です。添え字は変数番号になっています。また，ここでもサイズに注目してください。$\boldsymbol{V}'\boldsymbol{V}$ は，サイズで言うと $m \times n$ と $n \times m$ の積ですから，$m \times m$ になります。この行列は正方対称行列です。

また，対角項に各変数の標準偏差 s_j が入った行列 $\boldsymbol{Q}$ を以下のように定めるとしましょう。次のような行列

[*1] 適切なサイズってなんだよ，と思いますよね。これは計算に合うようなサイズ，という意味です。具体的に考えてみますと，$\boldsymbol{I}$ の後ろは $\boldsymbol{1}\boldsymbol{1}'$ です。$\boldsymbol{1}$ は $n \times 1$ の列ベクトルで，転置したものと掛け合わせますから，$\boldsymbol{1}\boldsymbol{1}'$ のサイズは $n \times n$ です。行列の引き算は同じサイズでないと成立しませんから，ここでの $\boldsymbol{I}$ も $n \times n$ でなければなりません。カッコの中身が $n \times n$ で，それにサイズ $n \times m$ である $\boldsymbol{X}$ をかけますから，計算結果や右辺のサイズは $n \times m$ になります。

です。

$$\boldsymbol{Q} = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m \end{pmatrix}$$

そうすると，これの**逆行列**をつかって標準得点行列 $\boldsymbol{Z}$ を次のように表すことができます。

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{Q}^{-1}$$

さらに，これを用いて相関行列 $\boldsymbol{R}$ を次のように表すことができます。

$$\boldsymbol{R} = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{pmatrix}$$

データのサイズにかかわらず，一般的にこのように表現できるのはとてもわかりやすいですね。

5.2 固有値と固有ベクトル

つづいて正方行列にみられるおもしろい特徴である，**固有値 (eigenvalue)** と**固有ベクトル (eigenvector)** についての話を見てみましょう。

ある正方行列 $\boldsymbol{A}$，列ベクトル $\boldsymbol{x}$，スカラー λ が次のような関係にあった時，λ を固有値，$\boldsymbol{x}$ を固有ベクトルと言います。

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

一見すると，$\boldsymbol{x}$ が両辺に入っていますから，$\boldsymbol{A}$ が λ に置き換わった等式に見えます。しかし一方は行列で，他方はスカラーです。こんな奇妙なことが本当にあるのでしょうか？具体的な数値例をみてみましょう。

$$\boldsymbol{A} = \begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

を例にします。この時次の関係が成り立ちます。

$$\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \end{pmatrix} = 7\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 7\boldsymbol{x}$$

確かに成立する組み合わせがありますね。この行列 $\boldsymbol{A}$ に対して，7 が固有値，$(1, 1)$ が固有ベクトルになっています。また，実はこの行列 $\boldsymbol{A}$ については，-1 も固有値であり，そのときの固有ベクトルは $(-3, 1)$ も固有ベクトルです。

この固有値分解こそ，因子分析を元とする多変量解析の中心的な数学原理なのです。多変量解析の世界においては，分散共分散行列やそれを標準化した相関行列など，変数同士の関係を分析のスタートにおくのでした。これらの行列は正方行列ですから，その固有値や固有ベクトルを計算することで正方行列の特徴を別の視点から分解して考えられるようになります。

5.2.1 固有値の特徴

この固有値の数学的特徴は色々あるのですが，データ分析をする上で重要な点を押さえておきましょう。

固有値の特徴として，**固有値の総和が正方行列の対角要素の総和に合致する**，というのがあります。数式で表現すると，次のようになります。

$$\sum_{i=1}^{N} \lambda_i = trace(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$$

ここで a_{ij} は行列 $\boldsymbol{A}$ の要素であり，a_{ii} は i 行 i 列目，つまり対角要素です。この正方行列 $\boldsymbol{A}$ のサイズは N で，対角要素の総和をとくに**トレース (trace)** といい $trace(\boldsymbol{A})$ と表します。それが固有値の総和とイコールになる，ということを表しています。サイズ N の正方行列からは固有値が N 個算出できることがわかっており，それをすべて足し合わせたものがトレースと同じになっているのですね。先ほどの例で言えば，$\boldsymbol{A}$ のトレースは $1 + 5 = 6$ で，固有値の総和は $7 - 1 = 6$ であり，確かにこの関係が成立していることがわかります。

分散共分散行列のトレースは，分散の総和を意味します。項目同士の関係を表した行列であれば，分散はその項目から得られる情報の大きさであり，それを総和するということは，その調査研究・項目群から得られる情報の総和であると言ってもいいでしょう。相関行列のトレースは，対角項に入っているのが $r_{ii} = 1.0$ ですから，項目の数と一致します。1 つの項目の情報量を 1.0 に基準化して N 項目分の情報がある，ということを表しています。固有値と行列の関係は冒頭で示したように，$\boldsymbol{Ax} = \lambda \boldsymbol{x}$ であり，正方行列の特徴をスカラーにしてしまうというものです。得られる N 個の固有値は，元の正方行列のエッセンスをスカラーにして表現しているわけです。

ところで $\boldsymbol{A}$ を n 次正方対称行列，つまり $n \times n$ サイズの対称行列だとすると，n 個の固有値が求められます。これを $\lambda_1, \lambda_2, \cdots, \lambda_n$ として，対応する固有ベクトルを $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ とします。ここで各固有ベクトルのノルムが 1 であるとしましょう。行列と固有値・固有ベクトルの関係から，

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$$

となりますが，このベクトルを並べた行列 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{pmatrix}$ を考えると，

$$\boldsymbol{AX} = \boldsymbol{X\Lambda}$$

と書くことができます。ここで $\boldsymbol{\Lambda}$ は

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

のような行列です。

この両辺に $\boldsymbol{X}'$ をかけると

$$\boldsymbol{AXX}' = \boldsymbol{X\Lambda X}'$$

となりますが，固有ベクトルの性質とノルムを整えていることから $\boldsymbol{XX}' = \boldsymbol{I}$ であり，そこから

$$\boldsymbol{A} = \boldsymbol{X\Lambda X}'$$

と書くことができます。

ここであらためて要素に注目すると，行列 $\boldsymbol{A}$ が次のように分解されていることがわかります。

$$\boldsymbol{A} = \lambda_1 \boldsymbol{x}_1 \boldsymbol{x}_1' + \lambda_2 \boldsymbol{x}_2 \boldsymbol{x}_2' + \cdots \lambda_m \boldsymbol{x}_m \boldsymbol{x}_m' = \sum \lambda_i \boldsymbol{x}_i \boldsymbol{x}_i'$$

となります。

この分解例は 2×2 の簡単な例で確認しておきましょう。たとえば $\boldsymbol{A} = \left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ という行列と，対角行列 $\boldsymbol{\Psi} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$ があったとして，$\boldsymbol{A\Psi A'}$ の計算をしてみたいと思います。

$$
\begin{aligned}
\boldsymbol{A\Psi A'} &= \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
&= \begin{pmatrix} \alpha a & \beta c \\ \alpha b & \beta d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
&= \begin{pmatrix} \alpha aa + \beta cc & \alpha ab + \beta cd \\ \alpha ab + \beta cd & \alpha bb + \beta dd \end{pmatrix} \\
&= \alpha \begin{pmatrix} aa & ab \\ ab & bb \end{pmatrix} + \beta \begin{pmatrix} cc & cd \\ cd & dd \end{pmatrix} \\
&= \alpha \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} a & b \end{pmatrix} + \beta \begin{pmatrix} c \\ d \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix}
\end{aligned}
$$

と，このようにスカラーとベクトルの積和の形に書き換えられるのですね[*2]。

さて，これらをまとめて考えると，

1. 固有値分解は行列を列ベクトルとその転置ベクトルの積の形に分解する。
2. 固有値の総和は元の行列の対角要素の総和である。
3. 元の行列の対角要素は各項目の分散を表している。

ということですから，固有ベクトルは全体の情報量をそのままに重要度の大きさに並べ替えたもの，固有値分解は行列をその要素の重要度ごとに分解していくことである，といえます。これこそ**因子分析**で取り出そうとしている因子であり，固有値の大きさはその因子の重要度として，共通次元の判別 (どこまで共通次元とみなすか) に使われるのです。

5.3 固有値と固有ベクトルを求める

ここで少し数学の方に話を戻して，固有値と固有ベクトルの計算方法を考えましょう。元の式を書き換えて次のような方程式を考えます。

$$
(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}
$$

固有ベクトルは $\boldsymbol{x} = \boldsymbol{0}$ すなわち全部ゼロであれば当然成り立ちますから (自明な解といいます)，これは除外することにします ($\boldsymbol{x} \neq \boldsymbol{0}$)。行列の表現は連立方程式の解を求めることと同じなのでした。$\boldsymbol{A} - \lambda \boldsymbol{I}$ を連立方程式の係数行列だと考えれば，それが**逆行列**を持つと左辺にそれをかけてしまえば全部ゼロの答えになってしまいますから，そうでない答えを求めるには，この係数行列が逆行列を持たないことが重要です。

[*2] ただし，この計算が可能なのは分解する元の行列が実対称行列だからです。実対称行列は固有値と固有ベクトルで対角化可能であることが証明できます。実対称行列の固有値は全て実数ですし，固有値が全て実数であれば適当な直交行列をつかって対角化でき，実対称行列の固有ベクトルは互いに独立するのでこれらを使って直交行列を作ることができるからです。これらの性質は線形代数のテキストなどの証明を参照してください。また，計算プロセスからもわかるように，同じ要素を持つ列ベクトルと行ベクトルの積ですから，結果は対称になってしまうからです。

さて，この授業の中では説明してきませんでしたが，方程式が解を持つかどうかを決定する計算方法があります。これを**行列式 (determinant)** といい[*3]，この値がゼロでなければその方程式は解を持つ，ということがわかっています。説明しなかったのは，この値を求める計算がとても面倒だからで，詳しくは線形代数のテキストにお任せするとして[*4]，ここでは簡便のために 2×2 方程式の行列について紹介します。

2×2 の係数行列，$\boldsymbol{P} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ の逆行列は次の式で求められることがわかっています。

$$\boldsymbol{P}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

この式から考えると，$ad - bc$ のところが 0 になるとこの計算はできませんから，逆行列が存在しないことになります。この $ad - bc$ にあたるところが行列式であり，$|\boldsymbol{P}|$ とか $det(\boldsymbol{P})$ のように表します。$ad - bc$ が 0 でなければ方程式は解けるのですが，今回の場合は解けると自明になってしまうので困ります。今回は $ad - bc = 0$ でなければならないのです。つまり一般的に書くと次のようになります。

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$

$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ とすると，この式は次のようになります。

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = 0$$

$$(a - \lambda)(d - \lambda) - bc = 0$$

この方程式をとくに**固有方程式**といいますが，これを解いてやれば良いことになりますね。具体的に $\begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix}$ の例で計算してみましょう。

$$\begin{vmatrix} 1 - \lambda & 6 \\ 2 & 5 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)(5 - \lambda) - 12 = 0$$

$$\lambda^2 - 6\lambda - 7 = 0$$

$$(\lambda - 7)(\lambda + 1) = 0$$

ここから $\lambda = 7, -1$ が得られますね。

では固有ベクトルはどうなるでしょうか。固有値 7 の例で計算してみます。

$$\begin{pmatrix} 1 & 6 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 7 \begin{pmatrix} x \\ y \end{pmatrix}$$

$$x + 6y = 7x \rightarrow 6x = 6y \rightarrow x = y$$

$$2x + 5y = 7y \rightarrow 2x = xy \rightarrow x = y$$

あれっ？ なんじゃこりゃ？ と思った人もいるかもしれません。でもこれ，間違いではありません。実は固有ベクトルは大きさが定まっておらず，今回の例で言えば $x = y$ つまり $x : y = 1 : 1$ の関係であればあらゆる

[*3] 行列式は数値であり，解が求まるかどうか決定 determinant する，という意味なのに，日本語訳はなぜか「式」といいます。変なの。

[*4] たとえば村上 et al. (2016) の第 3 章をみてください。

値が成立してしまうのです。固有ベクトルが $(1,1)$ でも $(2,2)$ でも $(100,100)$ でも，$\boldsymbol{Ax} = \lambda\boldsymbol{x}$ の関係に影響しませんから，普通はベクトルの長さ (**ノルム (norm)**) を 1.0 に規格化するという方策が取られます[*5]。先ほど「固有ベクトルを適当な大きさに選んでやれば」というような表現をしましたが，それはベクトルの大きさはいくらでもいいからできることなのですね。

ここでみたように，2×2 の方程式であれば固有方程式を解くことはできるのですが，行列のサイズがどんどん大きくなると一般的に解けなくなって行くことは想像にかたくないと思います。実際我々は正方行列として，項目の情報が詰まった分散共分散行列とか相関行列を使いますから，それが 2 項目しかないなんてことはなくて，もっともっと大きなサイズになります。そうすると計算機を使って近似的に答えを求めて行くことになります。

5.4 固有値と固有ベクトルの幾何学的意味

固有値，固有ベクトルについて，今度は違う側面から見直してみましょう。

ある正方行列から固有値 λ と固有ベクトル $\boldsymbol{a}$ が得られたとします。このベクトル $\boldsymbol{a}$ のすべての要素を定数 c 倍したベクトル $\boldsymbol{b} = c\boldsymbol{a}$ を考えると，これもやはり同じ関係が成り立ちます。

$$\boldsymbol{Ab} = \lambda c\boldsymbol{a} = \lambda\boldsymbol{b} \tag{5.1}$$

先ほどの計算でも明らかになりましたが，固有ベクトルの値は絶対的なものではなく，要素間の相対的大きさを反映しているに過ぎないのでしたね。

さてこれを幾何学的に，図形として考えてみましょう。要素が 2 つのベクトルは，2 次元座標に表現できます。ベクトル $\boldsymbol{x} = (x, y)$ という座標を表しているというわけです。固有ベクトルも要素が 2 つであれば，座標で表現できます。先ほどの，要素を c 倍しても固有ベクトルとしての性質は変わらない，という話は，「固有ベクトルは大きさに意味はなく，方向を表したもの」ということになります。では何の方向を指し示しているのでしょうか。

固有値と固有ベクトルの話の最初にあった，$\boldsymbol{Ax} = \lambda\boldsymbol{x}$ というのを見直してみましょう。$\boldsymbol{x}$ がなんらかの座標を表していると考えると，それに正方行列をかけるとはどういう意味でしょうか。次の計算式を見てください。

$$\boldsymbol{Ax} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix} \tag{5.2}$$

これをみると，座標 $\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ に $\boldsymbol{A}$ をかけたことで，座標が $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ に変わった，と見ることもできますね。このように，ある座標が別の座標に移ることをとくに「変換」と呼びます[*6]。つまり**正方行列はなんらかの変換を施すもの**だ，と考えることができます。

今回の例では，行列 $\boldsymbol{A}$ には次のような性質があります。

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{5.3}$$

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.4}$$

[*5] ノルムとは，要素の二乗和の平方根，$\sqrt{x_1^2 + x_2^2 \cdots x_n^2}$ のことです。

[*6] より一般的にいうと，以下のようになります。: 集合 X の各元 x に集合 Y の元 $f(x)$ を対応させる対応 f のことを，集合 X から集合 Y への写像 (mapping)，関数 (function)，あるいは変換 (transformation) という。

そう，お気づきのように，これは固有値・固有ベクトルです。この行列の固有値・固有ベクトルはそれぞれ $\lambda_1 = 2, \boldsymbol{x}_1 = (1,0)$，$\lambda_2 = 3, \boldsymbol{x}_2 = (0,1)$ であることがわかります。この固有値，固有ベクトルの組み合わせをじっとみていると，おもしろい特徴がわかって来ます。

今回の行列から得られた 2 つの固有ベクトル，$(1,0)$ と $(0,1)$ は，2 次元平面の単位ベクトルと呼ばれるものです。**2 次元座標の任意の点は，これら 2 つのベクトルの任意の線型結合で表現できます**。座標 (a,b) は $a \times (1,0)$ と $b \times (0,1)$ からなるベクトルですから，$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$ というように，です。

このように，単位ベクトルは 2 次元世界の基礎となる単位ともいうべきもので，ここで $(1,0)$ は x 座標の，$(0,1)$ は y 座標の基盤となるベクトルであるということができます (これをとくに**基底**といいます)。つまり，正方行列 $\boldsymbol{A}$ から得られる固有ベクトルは，その正方行列が作る空間の基盤を明らかにするものであったのです。

では固有値はどうでしょうか？ 今回は x 座標を 2 倍，y 座標を 3 倍に引き延ばす変換をしたわけですが，この座標の歪み (重み) が固有値に対応していますね。つまり，固有値と固有ベクトルは新しい座標に変換する，その変換先の空間的性質を表していることになります。元の座標空間は $(1,0), (0,1)$ で作られる空間ですが，変換先の空間は $(2,0), (0,3)$ で作られている空間，ということになります。

つまり，正方行列は空間を変換するもの，あるいは正方行列の中に固有ベクトルを基底とした空間があるもの，ということです。

すべての正方行列に，こういった「変換」という解釈ができるのであれば，相関行列にも同様のことがいえるでしょう。相関行列は正方行列ですので，固有値分解できるのです。相関行列を固有値分解することは，相関行列の中に潜む次元 (dimension) を抽出してくることです。固有ベクトル (因子負荷行列) は，正方行列によって変換される，変換先の単位ベクトルのことだったのです。そして，固有値はその次元のゆがみ (重み，重要性) という意味があったのです。

5.5 因子分析モデルの行列表現

さて，行列とベクトルを学んできた皆さんは，このようなエレメントワイズの計算はうんざりで，行列でまるっと表現できるはずなのに，とうずうずしていることと思います。ということで，因子分析モデルの代数的表現ですが，これも行列を使って表現すると非常にシンプルに表現できるということを説明していきましょう。

内容はまったく同じですが，確認しておきましょう。標準得点行列 $\boldsymbol{Z}$ を因子負荷行列 $\boldsymbol{A}$ と因子得点行列 $\boldsymbol{F}$ をつかって，次のように表します。

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D} \tag{5.5}$$

ここで，各行列の要素のサイズ感をつかんでおきましょう。まず $\boldsymbol{R}$ というのは相関行列ですから，m 個項目があるのでサイズは $m \times m$ の正方行列になります。次に $\boldsymbol{F}$ ですが，これは因子得点の行列です。得点は人数分ありますから行は n，因子の数が列になるのでこれを p とすると $n \times p$ です。$\boldsymbol{A}$ は因子負荷行列。因子負荷行列は因子の数と項目の数の組み合わせだけあるわけですから，$m \times p$ になりますね。$\boldsymbol{U}$ は独自因子得点です。得点ですから人数分，独自性分は各項目にありますから，サイズとしては $n \times m$ になります。最後に $\boldsymbol{D}$ ですが，これは独自因子の負荷量です。項目の数だけあるのですが，列ベクトルや行ベクトルで表現すると計算の時にサイズが変わって不便なことになります。ですから，対角項に d_j をもつ正方行列 $m \times m$ として表現しています。

サイズを確認したところで，実際に行列計算をしてみましょう。

$$
\begin{aligned}
\boldsymbol{R} &= \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z} && \boldsymbol{Z} \text{ を因子分析のモデル式にして} \\
&= \frac{1}{n}(\boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D})'(\boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D}) && \text{前の項の転置を中に入れます} \\
&= \frac{1}{n}\left\{(\boldsymbol{F}\boldsymbol{A}')' + (\boldsymbol{U}\boldsymbol{D})'\right\}(\boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D}) && \text{転置のルールより} \\
&= \frac{1}{n}(\boldsymbol{A}\boldsymbol{F}' + \boldsymbol{D}'\boldsymbol{U}')(\boldsymbol{F}\boldsymbol{A}' + \boldsymbol{U}\boldsymbol{D}) \\
&= \frac{1}{n}\boldsymbol{A}\boldsymbol{F}'\boldsymbol{F}\boldsymbol{A}' + \frac{1}{n}\boldsymbol{A}\boldsymbol{F}'\boldsymbol{U}\boldsymbol{D} + \frac{1}{n}\boldsymbol{D}'\boldsymbol{U}'\boldsymbol{F}\boldsymbol{A}' + \frac{1}{n}\boldsymbol{D}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{D}
\end{aligned}
\tag{5.6}
$$

と，このように展開できました。記号を見ているとわかりにくいので，サイズ感を確認しましょう。最終的には，次のようになっています。

$$
\underset{m\times m}{\boldsymbol{R}} = \frac{1}{n}\underset{m\times p}{\boldsymbol{A}}\underset{p\times n}{\boldsymbol{F}'}\underset{n\times p}{\boldsymbol{F}}\underset{p\times m}{\boldsymbol{A}'} + \frac{1}{n}\underset{m\times p}{\boldsymbol{A}}\underset{p\times n}{\boldsymbol{F}'}\underset{n\times m}{\boldsymbol{U}}\underset{m\times m}{\boldsymbol{D}} + \frac{1}{n}\underset{m\times m}{\boldsymbol{D}'}\underset{m\times n}{\boldsymbol{U}'}\underset{n\times p}{\boldsymbol{F}}\underset{p\times m}{\boldsymbol{A}'} + \frac{1}{n}\underset{m\times m}{\boldsymbol{D}'}\underset{m\times n}{\boldsymbol{U}'}\underset{n\times m}{\boldsymbol{U}}\underset{m\times m}{\boldsymbol{D}}
$$

ここで，要素ごとに計算していた時のことを思い出してください。第二項 $\frac{1}{n}\boldsymbol{A}\boldsymbol{F}'\boldsymbol{U}\boldsymbol{D}'$ と第三項 $\boldsymbol{D}'\boldsymbol{U}'\boldsymbol{F}\boldsymbol{A}'$ の中にある，$\boldsymbol{F}'\boldsymbol{U}$ と $\boldsymbol{U}'\boldsymbol{F}$ のところは，共通因子得点と独自因子得点の積ですし，いずれも標準化されていますから，$\frac{1}{n}$ と合わせて考えると，これは相関係数を表していることになります。また，共通因子と独自因子は相関しませんので，これはイコール $\boldsymbol{0}$ となり，この 2 つの項が消えてしまうのでした。

また，第一項の $\frac{1}{n}\boldsymbol{F}'\boldsymbol{F}$ は，共通因子同士の相関を表しています。$\boldsymbol{F}'\boldsymbol{F} = \boldsymbol{C}$ とすると，これは因子得点間相関 $\boldsymbol{C}$ を表すことになります。これが直交であると仮定する，つまり他の因子と相関しないと考えると，$\boldsymbol{C} = \boldsymbol{I}$，つまり単位行列です。単位行列は計算に影響を与えませんから，$\boldsymbol{A}\boldsymbol{F}'\boldsymbol{F}\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{C}\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{A}'$ となり，この式は簡単に次のように変形できます。

$$
\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{D}^2 \tag{5.7}
$$

先ほどの代数的展開を，そのまま行列で表現しただけですが，この方がシンプルに表現できていますね。この表現は，因子分析の第一定理と第二定理の両方を含んで一度に表せているのです。

5.6 因子分析の数学的理解

さあ因子分析に戻って考えてみましょう。因子分析のモデルは次のようなものでした。

$$
\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{D}^2 \tag{5.8}
$$

ここで，左辺の正方行列を相関行列 $\boldsymbol{R}$ とし，$\lambda\boldsymbol{x}\boldsymbol{x}' = \boldsymbol{a}\boldsymbol{a}'$ となるようにベクトルの大きさを整えてみましょう。これは λ を分解してベクトルの中に溶け込ませるようなものですから，$\boldsymbol{a} = \sqrt{\lambda}\boldsymbol{x}$ とすればよいでしょう。すると相関行列は次のように分解できます。

$$
\boldsymbol{R} = \boldsymbol{a}_1\boldsymbol{a}_1' + \boldsymbol{a}_2\boldsymbol{a}_2' + \cdots + \boldsymbol{a}_m\boldsymbol{a}_m' + \boldsymbol{d}\boldsymbol{d}' \tag{5.9}
$$

ここでは共通因子の数が m 個だとわかっている体で分解していますが，基本的にはサイズ N の行列からは N 個の固有値がずらーっと並ぶわけです。それをどこかで「共通しているのはここまで」と判断し，残りは誤差であるとしてまとめて $\boldsymbol{d}\boldsymbol{d}'$ にしているだけですね。このように数学的にはここからが共通因子，ここから

が独自因子といった区別をすることなく，最後のひとかけらまで固有値分解を行なっているのですが，その次元の重要度でもって共通因子と誤差因子に (研究者が恣意的に) 分割しているのが因子分析のやっていることなのです。

一般に，N よりも m のほうがグッと少なくなります。たとえば YG 性格検査では $N = 120$ であり，m はせいぜい 5 から 10 数個です。120 項目のつくる 120 次元空間の中で，そこに働きかけても方向の変わらない基礎的な少数の次元にのみ注目すれば，効率よく情報圧縮ができるというものです。

相関係数を固有値分解すると，その固有値はすべて足し合わせるとサイズ N になるのでした。元のデータから計算される相関行列は，1 つの項目が一単位分の情報を持っていると考えますが，固有値分解はそれを次元の重要性順に並べ替えます。固有値は項目いくつ分の重要度があるかということを表す指標だと考えることができます。どこから共通因子でどこからが誤差か，ということを考えるときに，たとえば固有値が 1.0 よりも小さくなるようであれば，項目 1 つ分の情報もないのだからということで誤差因子だと判断することがあります。

ところで因子分析モデルの $\boldsymbol{A}$，因子負荷行列ですが，これは共通因子の固有ベクトルをセットで扱ったものです。つまり，$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m \end{pmatrix}$ と縦ベクトルを並べたものになっています。エレメントワイズで表現すると次のようになります。

$$\boldsymbol{A} = \left(\begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1N} \end{pmatrix}, \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2N} \end{pmatrix}, \cdots, \begin{pmatrix} a_{m1} \\ a_{m2} \\ \vdots \\ a_{mN} \end{pmatrix} \right)$$

ここで $\boldsymbol{AA'}$ の間に単位行列 $\boldsymbol{I}$ を挟んでも，別に結果は変わりませんよね。単位行列はかけても変わらないのが特徴ですから，$\boldsymbol{AA'} = \boldsymbol{AIA}$ です。ここでの $\boldsymbol{I}$ のサイズは $m \times m$ であることに注意しつつ聞いてください。

$\boldsymbol{I} = \boldsymbol{TT'} = \boldsymbol{T'T}$ になるような m 次の行列を使うと，$\boldsymbol{AA'} = \boldsymbol{ATT'A'}$ の関係は保たれたままです。この $\boldsymbol{T}$ はこれまた行列の座標を変えてしまう変換行列であり，これを挟むことができるということは**因子負荷量の値はなんでもあり**だということになってしまいます。この変換行列 $\boldsymbol{T}$ はとくに**回転行列 (rotation matrix)** とも呼ばれますが，因子分析はこのようにどんな回転でもできる，**回転に関する不定性**があるのです。あらゆる因子負荷量の組み合わせがあり得る，というのは困りものなので，なんらかの形で因子負荷行列 $\boldsymbol{A}$ に制約をかける必要があります。言い方を変えれば，色々な制約の中ではあっても好きな値を取ることができますから，ユーザにとって便利な基準を考えてやれば良いでしょう。因子分析では一般に，**因子軸の回転**を行いますが，それはこうした理由からです。

ちなみに $\boldsymbol{TT'}$ に $diag(\boldsymbol{T\Phi T'}) = \boldsymbol{I}_m$ という制約のある行列 $\boldsymbol{\Phi}$ を挟んでやっても，元の計算モデルに影響はありません。この $\boldsymbol{\Phi}$ は**因子間相関 (factor correlations)** とよばれ，相関を持った因子軸の回転，すなわち**斜交回転 (oblique rotation)** も色々なものが考えられています[*7]。

[*7] これに対して $\boldsymbol{TT'} = \boldsymbol{I}$ のような因子間相関がない (単位行列) な回転を**直交回転 (orthogonal rotation)** といいます。

第 6 章

心理尺度で測定しているもの

6.1 Elephant in the room

ここまで，態度尺度の原理やテスト理論，因子分析法など理論的・数理的側面での心理尺度というのを見てきました。少し理屈っぽい話が続いたので辟易している向きもあるかもしれませんが，ここまでで改めて測定の原理，数理を見直してきたのは，「紙とペンで質問紙に回答することで，何がどうわかったといえるのか」という問題をしっかりと確認しておく必要があったからです。

公認心理師の標準カリキュラムの中には，心理統計という授業があります。この授業では，得られたデータを分析し，報告するための技術について修めることがもとめられています。というのも，心理学は実証科学ですから，データを分析するというプロセス・手続きはすべて書き出すことができるはずだからです。客観性，再現性を担保するのが科学という営みの根本的な発想であり，誤った分析結果からは誤った結果が得られ，しっかりと正すことができることが重要です。

しかしそれ以前に，分析するデータは何をどのように数値化したのか，何をデータと見做したのかについてが間違っていると，その後の心理統計的処理が適切であっても，意味のない数字遊びをしていることになってしまいます。根本的な土台が腐っていると，その上にどのような家屋も建築できないのです。

質問紙による測定は一見，簡単なものに見えます。「わからないなら，聞けばいいじゃない」という軽い気持ちでデータを集めてもいいじゃないか，と思われるかもしれません。しかし，聞いても本当の答えが返ってくるかどうかわからないので，心理学では何をデータにするかについて 100 年以上の労力をかけてきたのです。人間は，嘘をつくし，間違えるし，易きに流れる生き物です。あるいは何かの指標を目標にすると，それは必ずハックされます。たとえば感染症対策の不備を指摘されるのに困るようであれば，検査回数を減らしてしまえばいいのです。大学進学率，就職率が高く評価されるのであれば，進学や就職ができそうにない人を「希望者ではない」と分母から除外してやることで改善できます。学生が授業にどれぐらい積極的に参加しているかを評価したいのであれば，椅子に対する前傾の角度や手を上げた回数を数えてもいいですし，椅子を取り除いて演習にしてしまえばみんな立ち上がって議論していることになります。これは半分冗談のようですが，半分は笑えない側面があります。繰り返しますが，安易な指標化はハックされ，実質的な意味がないことになってしまうのです[*1]。

心理尺度が心理学のデータになっているのであれば，その妥当性について目をつぶることはできないはずです。*Elephant in the room* という慣用句にあるように，問題の存在に気づいていながら見て見ぬ振りをするのは，やはり適切な態度とはいえません。改めて，心理尺度がどのように使われ，何を測定しているのかを考えてみたいと思います。

[*1] このことについては Muller & Muller (2018) という本に詳しく書かれているので，ぜひ手に取ってみてください。

6.2 心理尺度の使われ方

ここで取り上げるのは，2021 年度の心理学研究という論文誌です。日本心理学会が年に 6 冊だしている機関紙で，専門家によるピアレビューを経て掲載された，日本語心理学系論文の最高峰といっていいでしょう[*2]。

2021 年の心理学研究 92 巻から，「尺度」や「質問項目」など言葉で測定している研究をピックアップしてみると，33 本の論文 (資料，特集含む) が該当します。そこでは 81 件の言葉による測定がなされており，測定されているのは 184 の概念や因子です。実験による反応だけ，あるいは逐語録のような「語り」だけをデータとしている研究は少なく，また人口動態のような集計されたデータだけで議論される研究もほとんどありません[*3]。つまり，心理学の研究はほとんど個人の言語刺激に対する反応，それを集計したものから構成されているのです。またこうした言語反応は，自記式であるものがほとんどで，他記式の研究は 1 件永谷他 (2022) だけでした。すなわち，現代心理学の研究とは「本人に言語で解答してもらう」ものになっています。

表 6.2 には，それぞれの論文中で利用されている尺度名をリストアップしました。少し時間をかけてみていただければ，実にさまざまな概念が測定されていることがわかります。個々の研究について問題点を指摘し，“論破” することが目的ではありませんが，批判的に考えていきたいと思います[*4]。

6.2.1 更新されない研究実践

これを見るとまず，あまりにも多くの側面について測られていることに驚かれると思います。1 つの学問領域としてまとまっているのが不思議なほどで，同じ研究テーマを持って集まった同人誌 (機関誌) でとは思えないかもしれません。また，それぞれの論文を参照すればわかりますが，反応カテゴリについては「1. あてはまらないから 5. あてはまるの 5 件法」などと表記されるのがほとんどで，2,3,4 など中間点にどのようなカテゴリが付与されているか，細かく言及してある研究もありません[*5]。各カテゴリについての言及がないということは，カテゴリに対する反応ではなく 1,2,3,4,5 という連続体を想定していることと考えられます。リッカート式のやり方である以上，態度についての連続体に正規分布を仮定した数値化をすることが基本原則ですが，分布の正規性，尺度値の等間隔性という仮定に言及するまでもなく前提として利用しているように見えます。心理尺度による測定値が，そこまでの精度をもたないものとして考えられているのかもしれません。

尺度を作るときは，何度も構成概念の確らしさ，尺度の信頼性・妥当性の高さを検証し，かなりの確らしさで当該心理概念を測定している，ということを明らかにする必要があります。その上で，作成に使用したデータに基づいて尺度をどのようにスコアリングするか，仕様の手続を標準化することが一般的です。物差しとしての精度や目盛りを確定してから公開するのは当然であり，原基が変化しないからこそ尺度間・研究間での比較，知見の累積につながるはずです。

さて，各研究を見直してみると，その 3 割弱 (27.0%) では改めて探索的因子分析をして因子を確認しています。そのほかの多くでも，検証的因子分析によってどの因子がどの項目で測定されているかを確認してい

[*2] もちろんより優れた論文は英文で書かれて国際誌に載っているものではないかという批判はあるでしょうが，ここで論じる尺度の使われ方という意味では国内外でそれほど大きく違いがないようです。

[*3] 92 巻 5 号は「新型コロナウイルス感染症と心理学」を表題とした特集号であり，この号は語りや集計データが利用される傾向がありました。

[*4] 「批判的 critical」であることとは，問題点を指摘して悪い印象をあたえることではありません。長所も短所も言語化して明確にし，問題点を乗り越えてより良いものを作るためのステップです。疑ったり批判したりするということは，対象をより深く知ろうとすることです。また「論破」はコミュニケーションを終了させるためのものであり，批判的なコミュニケーションの持続的な生産を目的とする科学的営みとは真逆の精神であることについて理解してください。

[*5] これは論文の紙幅制限によるところも大きいかもしれません。昨今は電子的な付録を Web などで公開する取り組みもあります。

ます[*6]。改めて使用した尺度の概念的妥当性を確認しておく，という考え方によるものかもしれません。またそうしたときは，妥当性の上限たる信頼性の推定値として，α 係数が報告されることが少なくありませんが，今回の研究プールのなかで報告された α 係数の最小値は 0.48，中央値は 0.83，最大値は 0.97 でした。α 係数は信頼性係数の精度としては高くなく，因子負荷量で表される項目の因子に対する貢献度を考慮しないスコアになっているのですが[*7]，いまだにこうした伝統的な手法から脱却できていないのが現状です。あるいはこれも，心理尺度の測定精度がそれほど高くないことを，著者や読者が自覚しているからかもしれません。

また，α 係数は目安として 0.8 以上であれば内的整合性信頼性が取れている，といわれています。言い換えるならば，それ以下であれば尺度として一貫した何かを測っているとはいいにくいことになります。今回 α 係数を報告している 217 の概念について，この基準をクリアしたのは 69.1% であり，30% 近くはこの基準に満たないものでした。小岩他 (2021) では $\alpha = 0.48$ の尺度について報告されており，この基準を厳格に適用するなら考慮に値しない概念ということになります。ちなみにこの尺度は新型コロナウイルス感染症に対する対処行動を分類するものであり，「不安からの逃避」因子と名付けられています。項目の内容を考慮した上で有用であると判断されたこと，また「いずれの項目を除外しても信頼性係数の値が上がらなかったことから((小岩他, 2021),P.447,L.11)」，以降の分析から除外せず，そのまま考察の対象になっています。繰り返し指摘しておきますが，ここではこうした個々の事例をあげつらうことが目的なのではありません。新型コロナウイルス感染症のような，突発的で先例のない現象に対して素早く対処するため，急遽用意された心理尺度なので十分な信頼性，妥当性の検証がおいつかず，精度を犠牲にしてでも概略を掴もうという目的があったため仕方がない，という考え方もあろうかと思います。しかしことこの研究に限らず，精度の低い概算的な基準を用いて，かつその基準に満たなくても問題ないと許容される傾向が学会全体にあることは，指摘しておきたいと思います。

また尺度得点については，尺度を作成したときの標準化された手続によるものではなく，因子を構成する項目の平均をもちいた**簡便的因子得点**が用いられることがほとんどです。その傍証として，多くの研究では尺度得点の平均値と標準偏差のような記述統計量が報告されていることが挙げられます。尺度とはある規則に則って対象に数値を割り振ることとされますが，必ずしも先行研究で標準化されたものではなく，研究ごとに標準化された相対的な関係を見るために使われることが一般的な慣例となっているようです。既に因子分析を学んだみなさんは，因子分析の第一定理から，ある項目 j についての標準化された分散 r_{jj} には誤差分散 d_j^2 が含まれていることをご存知ですね (→ セクション 3.3,Pp.37 参照)。ある研究において探索的因子分析をし，因子に寄与する項目群の平均値で因子得点の推定値とするというのは，数理モデルの観点から見ると誤差を分離した後でまた誤差を戻し入れて分析していることと同じです。たとえるなら，昆布や鰹節で丁寧に出汁をとりきれいなスープを作った上で，出汁が出尽くした昆布と鰹節を戻し入れるようなものです。このことが正当化されるのは，推定された因子得点ではなく項目丸つけ行動という具体的な行動を研究対象にしているという考え方か，尺度の平均値に意味があるため標準化されたスコアを用いないという考え方に依拠しているからです。しかし後者の点については，当該研究で得られた平均点を標準的手続きでスコアリングされた尺度得点と比較していない時点で，その正当性には疑義があります。またたとえば，反応連続体の中点に付与されたカテゴリが「3. どちらでもない」であり，それより小さければポジティブ，大きければネガティブといった意味的な違いを考慮するのであれば，意味を持つといって良いかもしれません。しかし尺度を使った分析においてあくまでも相対的な比較しかしない場合や，反応カテゴリについての報告がない場合は，こうした点が十分に検証できなくなっているといえるでしょう。

[*6] 検証的因子分析とは，尺度の因子構造，測定モデルをあらかじめ分析者が準備しており，そのモデルとデータが合致したかどうかで妥当性を検証する方法です。これは**構造方程式モデリング**をつかうことで分析できます。

[*7] 清水 (2007) は α 係数ではなく，因子負荷量の情報を含んだ ω 係数を使うことを提唱していますが，今回のデータセットの中で ω 係数を報告しているのは永井 (2021) の 1 件のみでした。

表 6.1 測っているものの分類

自他の区分	基準の所在	分類名	項目例
自分	内的	自分の意図，信念，自己評価，願望	今の自分が客観的にどう見えるか知りたい
自分	内的	自分の今の感情	怒りを感じる
自分	内的	自身の行為の理由	できないことができるようになるとうれしいから
自分	外的	自分の普段の振る舞いについての自己報告	厳しく命令したり注意したりする
自分	外的	近況など経験に基づく自身の状態	風邪やインフルエンザなどにとても感染しやすい
他者	内的	他者に対する評価	黒い衛生マスクを着用する人物についてどう感じますか
他者	内的	推測に基づく自他の行為の予測	(あなたが詐欺電話について相談した場合，配偶者は) 電話を詐欺だと見抜くことができると思いますか
他者	内的	事象・現象に対する評価や理由づけ	音楽は友人たちとの話題にできるものだから
他者	外的	事実の言語報告	両親はよく喧嘩をする

信頼性係数の使い方や，尺度値の与え方，報告の仕方についてはこのように，数理モデル的な正当性が保持されているものというよりも，研究実践上の慣習に依存しているところが少なくありません。このこと自体について批判し，改めることも重要ですが，それよりもこうした慣習で測定されているものについて，さらに詳しく見ていきたいと思います。

6.2.2 測っているものは何か

実際の使われ方を見た上で，どのようなものが測定されたと考えられているのか，改めて確認してみたいと思います。内容をいくつかに分類したのが表 6.1 です。

これらを見るだけでも，既にサーストン法やリッカート法が誕生した経緯である，「態度対象に対する意見文についての本人の位置付けの評定」という枠組みは大きく飛び越えていることは明らかです。これを 2 つの次元で分類してみましょう。

- 個人の主観的経験世界，自己の内部についての言及と，他者や事象・現象についての言及
- 判断基準の所在が自分の内部におくもの (内部状態の参照。「そう思う」「あてはまる」など) と，客観的なもの (頻度や程度の報告，社会的価値基準で評価できるもの。「毎日ある」，「全く問題にならない」，「経済的に困っている」など)

その上で，各領域について何を測定しているか，測定上の問題があるとすればどのようなものかについて考えてみたいと思います。

■内部についての言及を，内的基準で評価する場合 自分の意図や信念，感情を評定させるケースがここに当たります。当然のことながら，本人の主観的な経験，感覚を，本人の主観的な語感にもとづいて報告させるわけですから，その判断の正当性に疑問を挟むことができます。その判断は本当に正しいのでしょうか。自分のことは自分が一番よくわかっている，などと言いますが，本当にそうなのでしょうか。

仮に自分の意識では疑いないことだ，という表明がされたとしても，その人が嘘をついてないといえるでしょうか。嘘とは言わなくても少し表現を誇張したり，社会的望ましさを考えて表現を変えてないといえるでしょうか。さらにいえば，真実でないことを自覚しているかどうかという問題もあります。社会的状況にプライミングされている場合は気づきませんし，お酒を飲んでいる時に「自分は酔っていない」と主張しても誰も信じてくれないように，本人の意識的な報告が誤っているということは当然あり得るわけです。本人は至って真剣で (もちろん素面で)，真実であるという自覚があってもです。

もちろんそうした間違いが生じないように，方法論上の工夫はいろいろ考えられています。類似の文言で多角的にアプローチしたり，虚偽報告が含まれないよう引っかけ問題を用意することもあります。三浦 & 小林 (2015) ではとくに Web 調査での手抜き回答者，Satisficer を検出する研究なども行っています。そのほか Lie 項目，フィラー項目などを入れるといった調査上のテクニックなどもかんがえられるでしょう。

しかし虚偽報告を除外したとしても，さらに大きな哲学的問題が残ります。すなわち，その人が自らの感覚を報告する場合，その目盛が正しいかどうかをどのように判断すれば良いのでしょうか。ある人にとって「非常に嬉しい」という感覚と，別の人にとって「やや嬉しい」といった感覚，この「非常に」と「やや」の大小を比較することはいかにして可能なのでしょうか。病院臨床の現場では，「最も痛かった時を 10，痛みがない時を 0 として，今の痛みはどの程度ですか」と言語報告を求めることがあります。これは Numerical Rating Scale(NRS) と呼ばれ，実際に筆者も骨折で入院した時に毎朝の健診で質問されました。ところが，入院初日に「今の痛みはどの程度ですか」と聞かれ，「5 点ぐらいでしょうか?」と回答したところ，看護婦さんに「そんなに?!」と驚かれたため，慌てて「あ，3 ぐらいです，たぶん」と言い直した経験があります。確かに上限，下限があればある程度の相対比較は可能で，他者とスケーリングを合わせることもできますが，これについても多分に**言語という共通の意味空間**に依存したスコアであることは間違いありません。言語の意味が互いに共有できているからこそ，質問と回答や合算ができることに辛うじて根拠があたえられるのであり，「集計した累積度数から算出した相対的位置」ほどの精度を期待することはほぼ不可能でしょう。

■内部についての言及を，外的基準で評価する場合 自分が普段ある振る舞いを「どの程度頻繁に」，「どの程度の強度で」行っているかを自己報告させて測度としているのがこの場合です。

もちろん自己報告ですから，嘘をついていないか，社会的望ましさの影響はどうか，嘘をついていることの自覚はあるかといった問題が含まれるのは先ほどと同様です。また，今現在その行為・行動をおこなっているのでなければ記憶に頼ることになりますが，記憶が正しいかどうかについても考えなければなりません。過去を振り返って評価する方法には回顧法などと名称はついていますが，過去の経験を思い出させることにどれほどの信憑性があるのかについては，記憶の研究を引用するまでもないでしょう。思い出したくない，思い出したところで人に言いたくない，ということも少なくないので，その数値の信憑性はかなり低いと見積もっておいた方が良いのではないでしょうか。

またこうした頻度，程度の多くは順序尺度水準であると考えられます。大小関係は維持されているといえるでしょうが，加算したり平均を出したりすることはできませんので，研究報告をする上では注意が必要です。

■他者や事象についての言及を，内的基準で評価する場合 他者に対する評価，事象・現象に対する評価や理由づけは，社会的態度を測定している領域といえそうです。社会的に実在性が共有されているものに対して意見を表明すること，その意見は心理的状態を反映している，少なくとも「言質が取れた」ものとして

考えることができます。そうした意見の分布は，様々な要因からの影響が想定されるので，正規分布していると仮定することもできるでしょう。こうした分布が確認できるのであれば，評定者集団をもって事前に数値化する基準を与えた尺度を作成し，尺度値として数値化することもできます。また正規分布が仮定できるのであれば，集計して累積的な分布を見るとある平均値をもって左右対称に分布していることが確認できるはずであり，正規分布の確率密度を相対頻度で分割することで各意見項目の核反応カテゴリに数値を付与することもでき，個々人の態度得点として数値化できます。前者はサーストン法，後者はリッカート法という尺度作成法であり，リッカート法での尺度はサーストン法での尺度と高い相関を示すことがわかっているのでした (Likert, 1932)。

ですから，リッカート法をつかって測定することができる，といっても良いでしょう。ただし，その数値化が非常に簡便なものになっていることに注意が必要です。本当にその集計値は正規分布の形をしているでしょうか。態度理論の仮定を満たしたものになっているでしょうか。3 件法，5 件法，7 件法など，さまざまな運用形態がありますが，そのこと自体は問題ではありません。問題はその幅の中で正しく分布できているかどうかであり，平均点が高すぎる (**天井効果**) とか，低すぎる (**床効果**) といった問題があれば，下から順に 1,2,3... と数値化することの正当性が担保されません。図 6.1 には左右対称等間隔なスコアが正当化される事例 (左上) も示しましたが，天井効果 (左上)，床効果 (右下) が見られる場合，あるいは正規分布でなく二峰性 (右下) があるばあいの，シグマ法による尺度化例を示しました。こうした分布の状況を顧みず，機械的に数値化することに問題がないといえるはずもありません。

天井効果，床効果，あるいは正規分布とは思えないような分布をしている場合，統計モデルを駆使して尺度値を補正することが可能です。切断正規分布や複数の確率分布の組み合わせなど，モデリングによって補正するアイデアが考えられますが，少なくとも今回参照した文献の中でそうした補正は行われていません。測定された数値がそれでいいのかどうか，あらためて計算手順を考え直す必要がありそうです。

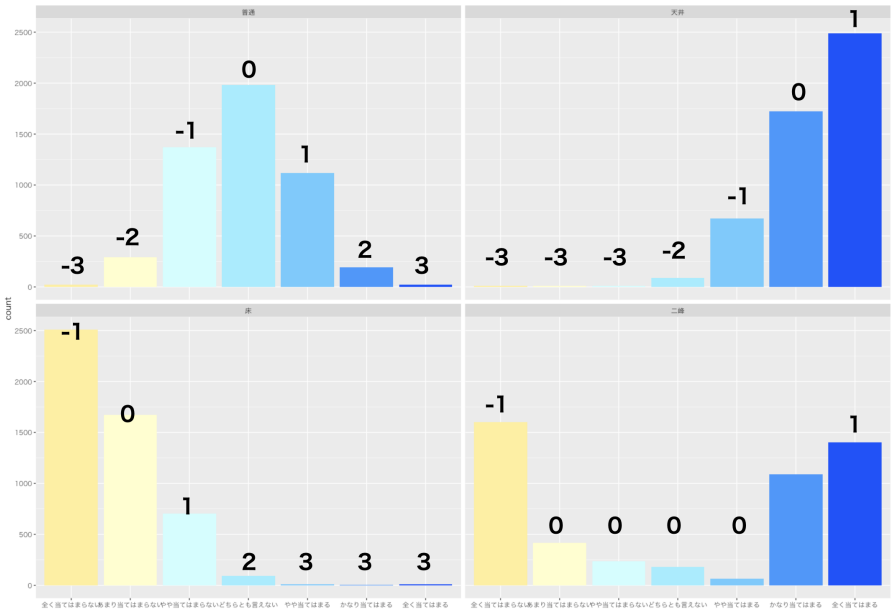

図 6.1 天井効果・床効果の見られた尺度

また言葉で指示した他者・事象・現象が同一のものであるかどうかについて，妥当性あるいは言語という共通基盤の頑健性について考えておく必要があります。社会的態度の研究の場合，例えば政治的態度を研究するのことを考えたとしましょう。「自民党についての態度」を測定する場合，「自民党について」の意見が自分

にどの程度当てはまるかを回答者個々人が判断することになりますが，このとき全員の頭の中に想定された「自民党」は同じ対象でなければなりません。日本における一般的な社会人であれば，「自民党」といえばあの自由民主党のこと，と誰でも同じものを思い浮かべられるでしょう。ここでアメリカの民主党を考えたり，イギリスの労働党を考えながら回答しているような人がいれば，回答パターンは無茶苦茶なものになってしまうでしょう。

しかし評価してもらう対象が「配偶者」「教師」「音楽」といったものであればどうでしょうか。誰 1 人として同じ配偶者を持っている人はいませんから，ある人のパートナーに対する態度と，別の人のパートナーに対する態度が同じもの (並べて，集計して，正規分布するもの) といえるでしょうか。「世話になった教師」といえば，ある時代のある学校においては「あの先生！」と特定できるかもしれませんが，それでも「自分は別の先生に世話になった」という人もいるでしょうし，時代と場所が変われば当然同じ人ではなくなります。「好きな音楽について」であっても，ロック，テクノ，K-Pop などなど，同じ「音楽」という括りで表現して良いものでしょうか。雅楽を想定している人と，パンクロックを想定している人が，「音楽とは穏やかな気持ちにさせてくれる」という意見に同じ次元で反応しているといえるでしょうか。

この批判はややもすると，言いがかりに聞こえるかもしれません。「配偶者」「教師」という言葉で表される典型的なパターンを研究したい，つまり言葉の使われ方の研究であれば問題ないかもしれません。あるいは社会通念上共通する架空の典型例，すなわち「あるある」ネタ研究であるなら，文化的な研究として価値があることかもしれません。しかし一人ひとりの主観的経験や心情を測定するというのとは，少し次元が違ってきているようです。

社会的態度は，個々人に数値を付与して個人間で相対的に比較することを可能にしてくれます。しかしそのことと，「その人が社会的態度を有している」かどうかについては，別の観点から考える必要もありそうです。このことについてはまた後で触れることになります。

■他者や事象についての言及を，外的基準で評価する場合 この領域は，事実の言語報告ですから，社会調査や実態調査として使われるところです。本人の言語報告に依存することなく，調べて裏取りをすることも可能ではありますが，「聞いた方が早い」から聞いているわけです。例えば家計収入など，本人の前年度の納税証明書を提示してもらえれば，円の単位まで正確に測定することができますが，そこまでの精度や手間が必要ではないので「100 万円未満，100～200 万円，200～500 万円･･･」等々の範囲で聞いていったりします。

プライベイトな情報であれば社会的望ましさの影響が，過去の記録を回顧して回答してもらう場合は記憶の歪曲といった問題が考えられますから，正しく事実を記録したいのであれば裏取りをするべきでしょう。言語による解答はあくまでも近似的，簡便的なスコアでしかないことに留意することが重要です。またこうした尺度で報告されるものは，調査前後の文脈によって大きく影響される可能性がありますから，「このような報告があるから」ということをエビデンスとするにはやや弱いかもしれない，と考えておいた方がいいかもしれません。

さて各領域にわたって批判的に眺めてきました。そもそも紙とペン，言語による報告にそこまでの精度や正確性がないことは，改めてしっかりと理解しておかなければなりません。測定されたデータに基づいて研究を構築するのであれば，その土台の頑健さを考え，可能であれば改良することを考えなければなりません。以下では改めて，どのような改善の可能性があるのかを見ていきたいと思います。

6.3 尺度構成の原理を考え直す

6.3.1 因子分析で取り出すことができるもの

心理学研究で，測定され議論されているのは，一体何なのでしょうか。リッカート法は態度測定法ですが，本来個々人の外部にある社会的実在に対する評価を測定する方法が，個人の内部にある心理的状態の主観的評価にすり替わっている，という事例も数多くみられます。

テスト理論のように，正誤の判断が個人の主観的経験を超えたところ，客観的基準として外部に存在する場合，その刺激群 (項目群) に対する反応パターンを測定することから得られるものについては，理論的にも正当化されやすいところです。被検者の能力が正規分布に従うという仮定も，生理生態学的指標の多くが正規分布しているという事実からも，受け入れやすいアイデアではないでしょうか。テストのように 0/1 という反応パターンについての限界も，多くのデータを取ったり数理モデルを工夫することで乗り越えていくことができますし，理論的正当化がされやすい領域ですので数理モデルもさまざまなかたちで発展しています。

正誤の客観的基準がないものの例として，パーソナリティ尺度をあげることができます。これは個々人の心理的な感覚を言語報告していますから，上で指摘した心理尺度の問題点のいくつかが該当するところかもしれません。しかし中身をよく見ると，性格とは安定的な人の行動パターンのことであり，言葉によるパターン分類の共通性の高さが担保されていることがほとんどです。「真面目な人」「明るい人」といった言葉の意味は社会的に広く共有されています。またパーソナリティ研究の多くは非常に多くのデータを集めることが一般的です。さらに性格という概念の特徴上，極端な人が少なく中庸的な人が多いということから，正規分布するという仮定を置くことの妥当性も認め得るでしょう。であれば，相対比較によってスコアリングすることの意味もまた妥当であるといえるでしょう。

パーソナリティ尺度の研究は，因子分析によって共通要因 (因子) を抽出し，その次元性すなわち構造について議論がなされます。ここで注意しておきたいのは，明確な意味 (広く社会に共有された意味) をもつ言語空間の，静的な構造を明らかにすることが研究目的であるという点です。因子分析は，その基本モデルとその展開から明らかなように，項目同士の相関関係を考えて，その相関構造の持っている次元的基盤 (基底) を重要度の高い順に確認することです。**言葉の意味空間における共通次元**を抜き出して，そこに個々人の反応パターンを与えると，特徴が際立った反応パターンを見ることができる。そうした反応パターン変換装置の仕組みが何かを考える，というのがパーソナリティを考えることです。その意味で，言語空間が持っている共通空間は，個人の内部にある構造ではなく，個人の外側，社会空間の共通基盤なのです。性格は個人が有する特性ではなく，多くの個人を分類するための社会的な基準であり，われわれがその基準を内在化して自他を表現しているに過ぎません。

あらためて因子分析で因子，概念を抽出するときに何をやっているか，数理モデルに基づいて考え直すと「変数内部の静的構造を抜き出している」ことはあきらかです。静的 static というのは動的 dynamic ではなく，一時的に止まっているもの，安定的で変化しない状態であるもののことです。構造 structure とはしくみのことであり，機能 function すなわちはたらきの背後にある共通した枠組みのことです。静的構造を取り出す技術ですから，いいかえると，不安定であったり共通していない個別性の高いものは，取り出すことができません。

時間的・概念的に不安定であったり，共通していない個別性の高いものを取り出すことができない，というわけですから，そこから分析ツールとしての特徴が見えてきます。たとえば過去の出来事や社会的共通認識，多くの人に共有できるものについての測定は可能です。事実，社会的態度は政治的態度や差別意識など，当時の人間にとって誰しも「それ」とわかる対象についての考え方のパターンをみることから研究が始まってい

るのは，それを測定すること，社会的態度の存在そのものが自明だったからです[*8]。逆に，時事刻々と変化する事象や個別性の高いものは態度調査で測定することはできません。個々人の感情や生理反応の言語報告や，恋愛関係・親子関係のようなプライベイトという個別性の高い話題，時事刻々と様相の変わるニュースに対する評価などは，尺度として安定した反応が返ってこないと考えるべきです。もちろん「怒っていますか」という質問にペンで回答することはできますが，「非常に怒っている」に静かに丸をつけた人は果たして怒っているのでしょうか？怒っているとしたいだけではないですか？

ところで昨今の研究の中に，場面想定法というものがあります。「あなたがもしこのような状況であれば，どのような反応をするでしょうか，その時の気持ちになってお答えください」という教示に従って，尺度に反応させていくわけです。仮想の世界，未来の状態などの反応を求めることは，事実に基づく科学としての心理学として認められるのか，という批判もあるかとおもいます。しかしこれなどは逆に「言葉」で制限された仮想世界を書き込むことで対象を限定的にし，同じ状況において取るべきパターンとして考えられている共通見解を取り出す，スクリプト処理のパターン分析という意味で，因子分析を適用できる適切な事例といえるかもしれません。ただし，「卒業を控えた学生の気持ちになって考えてみましょう」というような曖昧な状況設定では個別性が高くなってしまいますから，誰しもが没入できる共通空間，たとえば夏目漱石の「こゝろ」の第三部で展開される，友人 K が自殺したと聞かされた時の主人公の気持ち，などとすれば，その反応パターンから共通成分を抜き出すことができるでしょう。それを心理学と呼ぶかどうかはわかりませんが。

因子分析で取り出すことができるのは，社会的に安定的で共通した次元であり，個人の内部に存在するものではないということを議論してきましたが，実践上はさらに注意するべきことがあります。それは，因子分析をする元になる，質問紙調査の中からしか答えが出てこないということです。因子分析は項目感関係のなかの，次元的基盤を抜き出すものです。用意された項目にない次元については，当然のことながら抽出することができません。たとえば学校に対する「不適応の程度」を測定したいとして，「学校に友達がいない」「相談できる相手がいない」「相談できる先生がいない」といった質問項目群を児童生徒に与えて，その結果を因子分析したらどうでしょうか。当然のことながら「対人不安傾向」のような因子がみつかるでしょう。これをして，昨今の児童生徒は対人不安を抱えている，と結論づけるのは，おかしな振る舞いだとは思いませんか。「学校が楽しい」「勉強でいろいろなことを身につけたい」「図書館には本がいっぱいあってずっとこもっていたい」など，いろいろな楽しみ方や魅力についての評価を求めず，対人的でネガティブな項目ばかりを与えたとすると，いずれの項目にも「あてはまらない」「そう思わない」と答えていても相関係数は高くなり，因子として見出されてくることになります。このように，質問項目がもつ次元とは，研究者が先に項目として用意した項目群の中に埋め込まれたものに過ぎないのです。このことを専門用語で**てっちゃんの手品** (小杉, 2018) といいます。手品師が自ら仕込んだネタを，自らのタイミングで取り出して，あたかも発見したかのように表現するようなことがあってはなりません。

6.3.2 心の中を測るとは

心理尺度は，社会の中にある静的な構造を取り出したものです。個人の内部状態を自ら判断し，量的に報告することの妥当性はどこにあるのでしょうか。正直なところ，そのような妥当性，理論的な正当性を与えてくれる議論はまだまだ不十分であるとしかいえません。

Bergson (1889) は，意識に直接経験されるものは質的なものであり，言語というフィルターで量化されたもので程度の判断を報告することはできるが，それは心の状態そのものではないといった議論をしています。

[*8] 日本では欧米ほど明確な人種差別の歴史がありませんでした。黒人のように肌の色が違う，外見的特徴ですぐに「違う」とわかるものに対する話がどれほど受け入れやすかったかは，当時の当事者が持つ心象を想像するしかありませんが，かなり常識的，当たり前なことだったのではないでしょうか。

これに基づくと，心理学者は言葉の研究者に過ぎないのかもしれません。あるいは言葉を超えた，生きた経験を測定する方法を新たに考えるなど，何らかの哲学的決断を迫られることになりそうです。少なくとも，個人の内部状態を自覚的に判断する場合，間隔尺度水準以上の判断ができているとは考えにくく，せいぜい順序尺度，おそらく質的違いの言語報告である名義尺度水準として，各カテゴリをそのまま扱う必要があるでしょう。個人の感覚を数値化したものについて，いかにして他者と比較可能，合算可能なものにするかについては，哲学的にも数理モデル的にも根拠が必要であり，それ抜きには尺度で個人の何かを正しく測定したとはいえないことに間違いはありません。

6.4 尺度の正しい使い方

それでは心理学研究の，言葉での評定や数値化をしているものは全て研究として無意味である，ということになるのでしょうか。一足飛びにそこまで行くのではなく，たとえば心理学では「この時の主人公の気持ちを答えよ」というような典型パターンの分類をしているのだと割り切ってしまえば，それはそれで正当化しうるでしょう。問題は「その人の主観的経験，心の内部状態を数値化できたもの」とみなすには，もう少し丁寧なアプローチが必要だということです。

自己の経験についての，内部基準による判断でない領域については，しっかりと尺度化し，数値化の根拠を明らかにするところから始めなければなりません。心理学研究はそれでなくても，尺度が乱立すると批判されている現状がありますが，尺度作成時にはその標準化した採点方法も共有し，毎回のサンプルごとの探索的因子分析をやめて，「標準化された尺度原基に基づくスコア」を正しく適用していけばよいのです。当然のことながら，その正しく尺度を適用・運用するという実践のなかには，構成概念の妥当性を緻密に検討するというプロセスが含まれます。言葉による揺らぎがないように，測定する概念の固定化を明確にし，裏が取れるようなものがあればそれを使って尺度の外的妥当性も確保する必要があります[*9]。重要なのは，あくまでも静的な構造を見出しているだけであり，個人の内部状態に言及できるものではない，ということをしっかりと踏まえておくことです。

では個人の内部状態を，自らの判断基準で量的に評価する場合はいかなる正当化もできないのでしょうか。これについては，ひとつは NRS のように，個人の評定における最小値・最大値を教示することで尺度値の標準化をする，という解決策が考えられます。個々人の判断基準が異なる，すなわち個人ごとに尺度の分散が異なることが，個人間比較を許さない原因になっているわけですから，個人ごとに分散を整えた尺度をつくればよいのです。もちろん言葉による上限・下限の設定，程度の評定ですからどれほどの精度がみこめる測定なのか，疑問は残りますが多少は改善されるはずです。あるいは得られた尺度を個人ごとに標準化するといった処理を踏まえるだけでもいいかもしれません。あるいは，尺度に付与する値 X の上限を一定にするような調査法 (ex.100 点を重要度ごとに項目に分配するなど) が考えられます。こうしたケースごとの合計値が同じデータのことを**イプサティブデータ (ipsative data)** といい，統計的な処理をするにあたっては特別な配慮が必要[*10]ですが，こうすることで個人間の共通化を行うわけです。ハードウェアとしての人間の限界は個体が違っても大体同じぐらいだと見込めるので，その範囲での言語表現であれば相対的に比較可能だろう，という強い仮定の下で個人間のスコアを比較することが可能になります。

あるいは比較対象を，尺度項目の外部に置くこともできるかもしれません。**係留ビネット法 (Anchoring Vignettes)** は，尺度に回答させる前に回答者の回答傾向を別の項目で測定し，そこでみられた個人のバイア

[*9] 小杉 (in press) は身長と体重を心理尺度で測定し，言語報告の身長・体重と因子得点との相関が 0.8 程度あったことを報告しています。少なくともこの程度，あるいはせいぜいこの程度の外的基準があると考えることができるかもしれません。

[*10] ケースごとの合計値が同じということは，M 項目のうち自由度が M-1 しかないことであり，これを行列で処理する場合は正則でない，すなわち逆行列が求まらないため，さまざまな統計処理に不都合が生じます。

スを調整することで，個人間の回答幅が均等になるような調整をかけることを目的としています[*11]。このほかにも，INDSCAL や Tucker3 など，個人の違いをデータの特徴として分析の際に考慮する **3 相データ**の分析[*12]，多次元展開法など個人差をモデル化することを考えていけば，少なくとも個人間を比較するということについての分析的正当化はできそうです。これらは測定のモデリングとして，今後ますます展開の可能性が見込める領域です。

もちろん，そもそも論として，個人の判断はせいぜい順序尺度水準にすぎないという観点から，順序尺度に対応した分析モデルを積極的に活用することを考えるべきです。間隔，比率尺度水準で心理的判断・言語化ができているというのはあまりにも強い仮定ではないでしょうか。もっといえば，感覚の主体的な判断による言語報告は，名義尺度水準でしかないというのであれば，因子分析をやめて名義尺度水準に対応した多変量解析モデルを適用することを検討するべきです。いずれにせよ安易な数値化，反応カテゴリの省略はすぐにでも止めるべき悪習であるといえるでしょう。本書の以下の章ではこの観点から，順序，名義尺度水準の分析方法について考えていきたいと思います。

この章を閉じる前に，更なる根本的問題として，「心とは何であるか」を改めて考えてみましょう。因子分析は静的な構造を取り出すという話でした。それは項目が，言葉がある状態を一時点に切り取る作用をするからです。本当はさまざまな生理的反応，神経伝達物質の分泌が時々刻々と体内で変化し，胃のむかつき，発汗，まとまらない考え，不快感を次々と経験しているのに，それを「怒り」の一言で表現することができるでしょうか。筆舌に尽くし難い，言葉にできない，割り切れない気持ちというのが心の本性ではないでしょうか。言葉は時間を止めてしまうということを意識したとき，心理尺度で測定しているのはもはや，心ではなくて言葉・意味の世界のものであって，生きているということとは別の事象なのかもしれません。心理尺度の中には，言葉遣いの微妙な変化から状態と特性を区別することがありますが，それはあくまでも言葉遣いによる違いであって，現象の違いではないかもしれません。また TEM やネットワーク分析のように，時間を表現する萌芽的なアプローチも最近みられるようになってきましたが，言葉に依存している以上，根本的な解決にはなりません。

心理学，とくに社会心理学のように「言葉で切り取られた一貫性のある意味空間」を研究対象にするのだという割り切りがあれば，それはそれでひとつの哲学的判断ですが，もし「生きた心」を研究したいのであれば，音楽や動画のように動きそのものを表現する技法を手に入れなければならないのかもしれません。

[*11] 詳しくは北條 & 岡田 (2018) などを参照してください。
[*12] データの相と元については，セクション 8.1,Pp.107 で詳しく触れます。

表 6.2: 心理学研究 92 巻で用いられた尺度項目など

出典	尺度名など	項目数	項目例	因子名	α 係数
鈴 木 ・ 荒 俣 (2021)	動機づけ尺度	52	できないことができるようになるとうれしいから	内的調整	0.820
		52	集団活動や礼儀について学べるから	同一化調整	0.800
		52	サボっていると思われたくないから	取り入れ調整	0.810
		52	部活に参加しないと罰を受けるから	外的調整	0.810
		52	活動に参加したくない	無調整	0.840
	部活動への傾倒尺度	4	部活には自主的に参加している	傾倒	0.840
	基本的心理欲求充足尺度	12	私は, 能力のある人間だと感じている	有能感	0.730
		12	私は, 周りの人と有効な関係を築いていると思う	関係性	0.780
		12	私は自分の意見や考えを自分から自由に言えていると思う	自律性	0.740
	動機づけ尺度	52	できないことができるようになるとうれしいから	内的調整	0.870
		52	集団活動や礼儀について学べるから	同一化調整	0.840
		52	サボっていると思われたくないから	取り入れ調整	0.820
		52	部活に参加しないと罰を受けるから	外的調整	0.760
		52	活動に参加したくない	無調整	0.840
	リーダーシップ尺度	17	気まずい雰囲気があると解きほぐす	関係調整・統率	0.920
		17	活動内容に関する知識が豊富である	技術指導	0.850

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		17	厳しく命令したり注意したりする	規範的指導	0.840
神原 (2021)	バイアスについての説明文	7	自分の信念にあう情報は受け入れるが，自分の信念に矛盾する情報は否定する。		0.710
	驚いた程度	1	カテゴリ直接選択		
川崎・小塩 (2021)	病理的自己愛目録日本語版	52	私が人の手伝いをするのは，自分が良い人間であることをわかってもらうためだ	誇大性	
		52	私はよく，自分の偉業が認められるという空想をする	誇大妄想	
		52	誰に対しても，私の思うように物事を信じさせることができる	搾取性	
		52	私が言ったこと･したことに，興味を持ってもらえないとイライラする	脆弱性	
		52	心の中で感じている自分の弱さを他人に見せるのは耐えがたいことだ	自己隠蔽	
		52	人が自分の存在に気がついてくれないと，がっかりしてしまう	随伴的自尊感情	
		52	私がしてあげたことが感謝されないのではないかと思って，人を避けてしまうことがある	脱価値化	
生田目他 (2021)	日本語版幸せへのおそれ尺度	5	喜びの後にはたいてい悲しみがやってくるので，私はあまり喜びすぎないようにしている	幸せへのおそれ	0.880
	日本語版幸せの壊れやすさ	4	どんな時でも何かが起こってしまって，幸せをあっけなく失うかもしれない	幸せの壊れやすさ	0.900

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
小野他 (2021)	感情反応を測定する項目	17	やる気がわく	ポジティブ感情	0.920
		17	警戒する	不信感	0.890
		17	怒りを感じる	苛立ち	0.770
	魅力を測定する項目	1	メッセージカードの送り手は魅力的だと思う	魅力評定	
	感情反応を測定する項目	17	やる気がわく	ポジティブ感情	0.810
		17	警戒する	不信感	0.790
		17	怒りを感じる	苛立ち	0.870
新国他 (2021)	主体感を測定する尺度	17		精神的活動における主体の誤帰属	
		17	自分の身体を思うように動かせないと感じることがある	身体的活動における自己身体の制御不能性	
		17	周りに協調するよりも自分の主張を通すことがある	社会的活動における自己の主張性	
豊田他 (2021)	文章刺激に対する評価	1	先ほどの場面についての想像は, どのぐらい鮮明でしたか	想像の鮮明さ	
		1	○さんはどのような気持ちだと思いますか	相手の感情	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		1	○さんについての文章を読んで，あなたはどのような気持ちになりましたか	自分の感情	
		1	○さんに対して，あなたはどのぐらい助けようと思いましたか	援助意図	
		1	あなたがもし，○さんと同じ状況だとしたら，他者に助けを求めますか	援助希求度	
		1	あなたは過去に，○さんと同じような状況の人物に遭遇したことがありますか	遭遇頻度	
		1	あなたは過去に，○さんと同じような状況になったことがありますか	経験頻度	
		1	先ほどの人物はどのような気持ちだと思いますか	相手の感情	
		1	先ほどの人物についての文章を読んで，あなたはどのような気持ちになりましたか	自分の感情	
		1	先ほどの人物に対して，あなたはどのぐらい助けようと思いましたか	援助意図	
		1	あなたがもし，先ほどの人物と同じ状況だとしたら，他者に助けを求めますか	援助希求度	
		1	あなたは過去に，先ほどの人物と同じような状況の人物に遭遇したことがありますか	遭遇頻度	
		1	あなたは過去に，先ほどの人物と同じような状況になったことがありますか	経験頻度	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
廣 瀬 ・ 濱 口 (2021)	日本語版 SIS	32	私は悲しくなる	情緒覚醒	0.770
		32	私の一日は台無しになる	情緒統制不全	0.810
		32	私は自分の気持ちを隠そうとする	葛藤からの回避	0.820
		32	私はどちらか，または両方に対してかわいそうに感じる	葛藤への関与	0.750
		32	家族はまだうまくやっていける	建設的な家族表象	0.860
		32	私は家族の将来が心配になる	破壊的な家族表象	0.790
		32	私は板挟みになっているように感じる	巻き込まれ表象	0.800
	破壊的 IC	20	両親はよく喧嘩をする	葛藤の激しさ	0.810
		20	両親はけんかをしても，すぐに仲直りする	葛藤の持続性	0.800
		20	両親はけんかをしても，大きな声を出さずに話し合う	葛藤の解決	0.740
	IC のおそれ・身体反応	5	両親がけんかをしていると，不安になる	IC のおそれ・身体反応	0.880
	適応問題	33		不安・抑うつ	0.850
		33		攻撃的行動	0.860

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
天井 (2021)	中学生を対象とした情緒的援助期待尺度	64	とにかくただ話を聞いてほしい	受容期待	0.760
		64	今の自分が客観的にどう見えるか知りたい	再解釈期待	0.610
		64	自分はいい人間だと思わせてほしい	正当化期待	0.740
		64	がんばれと言ってほしい	楽観的期待	0.730
		64	一緒に何か夢中になることをして辛い状況を忘れたい	気晴らし期待	0.780
	悩みの認知評価	14	解決するための方法がわかっている	統制可能性	
		14	自分を傷つけることだと思う	影響性	
	援助要請行動	8	誰かから同情や理解を得る	情緒的サポートの利用	
		8	何をすべきか誰かからアドバイスを得ようとする	道具的サポートの利用	
	中学生を対象とした情緒的援助期待尺度	23	きちんと話を聞いてほしい	受容期待	0.730
		23	別な見方を教えてほしい	再解釈期待	0.750
		23	間違ってないよと言ってほしい	正当化期待	0.820
		23	そんなこと気にするなと笑い飛ばしてほしい	楽観的期待	0.640
		23	全く別の話をして話題から意識をそらしたい	気晴らし期待	0.750

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
	援助要請行動	8	誰かから同情や理解を得る	情緒的サポートの利用	
		8	何をすべきか誰かからアドバイスを得ようとする	道具的サポートの利用	
	中学生を対象とした情緒的援助期待尺度	23	きちんと話を聞いてほしい	受容期待	0.720
		23	別な見方を教えてほしい	再解釈期待	0.830
		23	間違ってないよと言ってほしい	正当化期待	0.840
		23	そんなこと気にするなと笑い飛ばしてほしい	楽観的期待	0.720
		23	全く別の話をして話題から意識をそらしたい	気晴らし期待	0.790
	援助評価	23	どうすればいいかがはっきりした	問題状況の改善	
		23	どうすればいいか余計にわからなくなった	退所の混乱	
		23	自分は一人じゃないんだと思った	他者からの支えの知覚	
		23	自分が他の人に頼りすぎていると思った	他者への依存	
	愛着	30	人に対して自分のイメージを悪くしないか恐れている	見捨てられ不安	
		30	他の人との間に壁を作っている	親密性の回避	
金政他 (2021)	愛着スタイル尺度	15		愛着不安	0.930
		15		愛着不安	0.940

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
	パートナーへの間接的暴力加害	6	相手の行動を制限することが	DaV 加害	0.900
		6	相手の行動を制限することが	DaV 加害	0.910
		6	相手の行動を制限することが	DaV 被害	0.900
	一般的他者版 ECR	18		愛着不安	0.940
	パートナーへの間接的暴力加害	6	相手の行動を制限することが	DaV 加害	0.940
		6	相手の行動を制限することが	DaV 加害	0.900
		6	相手の行動を制限することが	DaV 被害	0.900
讃井他 (2021)	詐欺電話に関する測定尺度	6	詐欺電話がかかってきたことを誰かに相談すると思いますか	相談行動意図	
		3	(あなたが詐欺電話について相談した場合，配偶者は) 電話を詐欺だと見抜くことができると思いますか」	相談相手としての信頼 (能力の予期)・妻	
		3	(あなたが詐欺電話について相談した場合，配偶者は) 電話を詐欺だと見抜くことができると思いますか」	相談相手としての信頼 (能力の予期)・夫	0.780

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		3	(あなたが詐欺電話について相談した場合，配偶者は) 誠実に対応すると思いますか」	相談相手としての信頼 (誠実さの予期)・妻	
		3	(あなたが詐欺電話について相談した場合，配偶者は) 誠実に対応すると思いますか」	相談相手としての信頼 (誠実さの予期)・夫	
		3	(あなたが詐欺電話について相談した場合，配偶者は) あなたの気持ちを分かってくれると思いますか」	相談相手としての信頼 (共感の予期)・妻	
		3	(あなたが詐欺電話について相談した場合，配偶者は) あなたの気持ちを分かってくれると思いますか」	相談相手としての信頼 (共感の予期)・夫	
		4	あなたは詐欺電話がかかって来た時に，自分がお金を騙し取られる被害にあう可能性がどの程度あると思いますか	被害リスク認知・妻	
		4	あなたは詐欺電話がかかって来た時に，自分がお金を騙し取られる被害にあう可能性がどの程度あると思いますか	被害リスク認知・夫	
	夫婦関係に関する測定尺度	19	うれしかったこと，たのしかったことについて	コミュニケーション・妻	0.910

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		19	うれしかったこと, たのしかったことについて	コミュニケーション・夫	0.910
		19	あなたは, 配偶者との関係についてどの程度満足していますか	関係満足度・妻	
		19	あなたは, 配偶者との関係についてどの程度満足していますか	関係満足度・夫	
		19	配偶者と話し合おうとする	問題対処法略の夫婦関係内アプローチ・妻	0.860
		19	配偶者と話し合おうとする	問題対処法略の夫婦関係内アプローチ・夫	0.850
		19	配偶者とは別に普段より社会的な催しに参加する	問題対処法略の夫婦関係外アプローチ・妻	0.710
		19	配偶者とは別に普段より社会的な催しに参加する	問題対処法略の夫婦関係外アプローチ・夫	0.760
水野他 (2021)	セルフコンパッション	26		セルフ・コンパッション	0.760
	新職業性ストレス簡易調査票短縮版	6		上司からのサポート	0.750

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		6		同僚からのサポート	0.830
		17		職務ストレッサー	0.630
	バーンアウト傾向	17		情緒的消耗感	0.820
		17		脱人格化	0.800
		17		個人的達成感の後退	0.810
太幡・佐藤 (2021)	インターネット上での情報プライバシー	20	最近 1 ヶ月のうちに，以下の情報を含めたツイート，リツイートをしたことがどのくらいありましたか	情報公開回数	0.910
	他者懸念	7	他人にプライベートな質問をされたくない	自己意識・維持行動	0.760
		4		他者意識	0.690
		4		他者維持行動	0.740
		5	総合的に考えて，他人についての情報を他の利用者と共有することは，深刻なプライバシーの問題を引き起こす	Twitter 上での他者懸念	0.890
石川他 (2021)	気分状態	24	いらいらしている	緊張と興奮	
		24	気持ちが滅入っている	抑うつ感	
		24	何か物足りない	不安感	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
池上他 (2021)	音楽聴取の心理的機能	133	音楽は自分自身の道を見つけてくれるのを助けてくれるから	自己認識	
		133	音楽は私の機敏を明るくすることができるから	感情調節	
		133	音楽は友人たちとの話題にできるものだから	コミュニケーション	
		133	私は周りの雑音を音楽でかき消したいから	道具的活用	
		133	大音量で聴くのが好きだから	身体性	
		133	音楽は私の信仰心を支えてくれるから	社会的距離調節	
		133	音楽は私の悲しみに寄り添ってくれるから	慰め	
赤澤他 (2021)	アサーション	16	好きな人には素直に愛情や好意を示す	関係形成	0.785
		16	買った商品に欠陥があったら交換してもらう	説得交渉	0.757
		16	好きな人には素直に愛情や好意を示す	関係形成	0.767
		16	買った商品に欠陥があったら交換してもらう	説得交渉	0.786
		16	好きな人には素直に愛情や好意を示す	関係形成	0.642
		16	買った商品に欠陥があったら交換してもらう	説得交渉	0.610
	視点取得	5		視点取得	0.912
		5		視点取得	0.944
		5		視点取得	0.929
	暴力観	5	けがをしない強さで叩く	暴力観	0.809
		5	けがをしない強さで叩く	暴力観	0.890
		5	けがをしない強さで叩く	暴力観	0.771

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
外山・長峯他 (2021)	制御焦点	16	どうやったら自分の目標や希望を叶えられるか, よく想像することがある	利得接近志向尺度	0.730
		16	どうやったら失敗を防げるかについて, よく考える	損失回避志向尺度	0.810
	エンゲージメント	14	課題に一生懸命に取り組んだ	行動的エンゲージメント	0.890
		14	課題は楽しかった	感情的エンゲージメント	0.930
		14	課題に没頭していた	状態的エンゲージメント	0.800
藤後他 (2021)	子の運動に対する養育態度尺度	10	子どもと一緒に運動する	共同活動	0.810
		10	試合 (あるいは進級試験等) について反省させる	支配的対応	0.780
		10	運動したり試合をしたりすることは, 子どもにとって良いことだよとつたえる	運動の価値伝授	0.830
	子供への対応	17	次は, どんな目的をもって練習する?	自律的対応	0.870
		17	これだから負けるのよ	支配・過干渉	0.810
		17	今日もよく頑張ったね	肯定的評価	0.720
		17	昨日のテレビの話など試合に関係ない話題をする	間接的支援	0.820
		17	次は, どんな目的をもって練習する?	自律的対応	0.880

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		17	これだから負けるのよ	支配･過干渉	0.850
		17	今日もよく頑張ったね	肯定的評価	0.690
		17	昨日のテレビの話など試合に関係ない話題をする	間接的支援	0.820
榊原・大薗 (2021)	マスク着用の理由		あなたがもし新型コロナウイルスに感染したら，症状は深刻なものになると思いますか マスクは，自分が周りからウイルスを移されないために着用している		
宮崎他 (2021)	他者に見られる不安	20	人前で文字を書かなければならない時，不安になる		0.940
	対人交流不安	20	あまり知らない人に会った時，あいさつするかどうか迷う		0.940
	他者に見られる不安	20	人前で文字を書かなければならない時，不安になる		0.950
	対人交流不安	20	あまり知らない人に会った時，あいさつするかどうか迷う		0.940
	他者に見られる不安	20	人前で文字を書かなければならない時，不安になる		0.930
	対人交流不安	20	あまり知らない人に会った時，あいさつするかどうか迷う		0.940
	他者に見られる不安	20	人前で文字を書かなければならない時，不安になる		0.930

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
	対人交流不安	20	あまり知らない人に会った時, あいさつするかどうか迷う		0.950
	特性不安	20	たのしい		0.930
	感染脆弱意識尺度	15	風邪やインフルエンザなどにとても感染しやすい	易感染性	0.810
		15	誰かと握手した後は手を洗いたくなる	感染嫌悪	0.750
	特性不安	20	たのしい		0.920
	感染脆弱意識尺度	15	風邪やインフルエンザなどにとても感染しやすい	易感染性	0.840
		15	誰かと握手した後は手を洗いたくなる	感染嫌悪	0.740
	特性不安	20	たのしい		0.920
	感染脆弱意識尺度	15	風邪やインフルエンザなどにとても感染しやすい	易感染性	0.790
		15	誰かと握手した後は手を洗いたくなる	感染嫌悪	0.720
	特性不安	20	たのしい		0.920
	感染脆弱意識尺度	15	風邪やインフルエンザなどにとても感染しやすい	易感染性	0.800
		15	誰かと握手した後は手を洗いたくなる	感染嫌悪	0.690
	特性不安	20	たのしい		0.910
		15	風邪やインフルエンザなどにとても感染しやすい	易感染性	0.820
	感染脆弱意識尺度	15	誰かと握手した後は手を洗いたくなる	感染嫌悪	0.740

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
鎌谷他 (2021)	黒色マスク着用者に対する信念	1	男性の顔の魅力は, 黒い衛生マスクを着用することで高まると思いますか	魅力	
		1	黒い衛生マスクを着用する人物についてどう感じますか	健康さ	
		1	黒い衛生マスクを着用する人物についてどう感じますか	ファッション性	
		1	黒い衛生マスクを着用する人物についてどう感じますか	イメージの良さ	
山本・岡 (2021)	感染者の特性に対するステレオタイプ的認知	20	責任感のある一責任感のない	社会性	0.902
		20	無気力な一意欲的な	活動性	0.755
	感染予防行動に対する認知	10	外出の際はマスクを着用しない	感染予防行動の欠如	0.972
	感染脆弱意識尺度	15	他の人よりも, 病気にかかりやすい方だ	易感染性	0.737
		15	前に使った人から何かうつりそうなので, 公衆電話は使いたくない	感染嫌悪	0.610
	感染予防行動に対する認知	10	外出の際はマスクを着用する	感染予防行動	0.760

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
飯田他 (2021)	経済的負担感	1	新型コロナウイルスが発生する前の，あなた自身の経済状態について，あなたはどのように感じていますか		
	孤独感	3			
	精神的健康	6		抑うつ	
		7		不安	
永井 (2021)	主体的な学修態度尺度	9			
		9			
中尾 (2021)	アタッチメント	9		不安	0.780
		9		回避	0.770
	孤独感	20		孤独感	0.890
	精神的健康	5		精神的健康	0.790
	アタッチメント	9		不安	0.800
		9		回避	0.740
	孤独感	20		孤独感	0.910
	精神的健康	5		精神的健康	0.850
杉山他 (2021)	孤独感	20		孤独感	
	不安・抑うつ	14		不安・抑うつ	
	睡眠の質	9		睡眠の質	
	クロノタイプ	5		クロノタイプ	
	孤独感	20		孤独感	
	不安・抑うつ	14		不安・抑うつ	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
	睡眠の質	9		睡眠の質	
	クロノタイプ	5		クロノタイプ	
	孤独感	20		孤独感	
	不安・抑うつ	14		不安・抑うつ	
	睡眠の質	9		睡眠の質	
	クロノタイプ	5		クロノタイプ	
髙坂 (2021)	臨時休業期間中の生活週間の変化	7	就寝時間が不規則になった	不規則な睡眠	0.810
		7	食事の量が変わった (増えた/減った)	食習慣の乱れ	0.760
		7	いつもより勉強する量が少なかった	学習時間の減少	0.810
		7	外出する機会が減った	外出の減少	0.710
		7	テレビやインターネットを見ている時間が増えた	テレビ・ネット視聴の増加	0.850
		7	これまでよりも運動しなくなった	運動の減少	0.670
		7	ゲームやスマートフォンをいつもよりも長く利用していた	ゲーム・スマホ利用の増加	0.740
平井・渡邊 (2021)	家庭と仕事における調整	10	職場でお互いの事情について説明しあう	仕事調整	0.822
		10	夫婦でお互いの役割分担について話し合う	家庭調整	0.806
	「家族する」概念	8	家族が自分にどうして欲しいかを考えて行動する	家族する	0.830

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		8	家族が自分にどうして欲しいかを考えて行動する	家族する	0.880
	仕事へのコミットメント	8	仕事においては人並み以上に努力をしている	仕事へのコミットメント	0.920
		8	仕事においては人並み以上に努力をしている	仕事へのコミットメント	0.930
	家庭, 仕事, および生活・人生の満足度	8	現在の夫婦の関係	家庭	0.900
		8	現在の仕事の内容	仕事	0.830
		8	現在の自分の生活	人生	0.840
小岩他 (2021)	新型コロナウイルス感染症に対する対処行動	10	換気が悪い場所には行かないようにした	三密の回避	0.850
		10	手洗い・うがいやアルコールによる手や指を消毒した	日々の感染予防	0.630
		10	気を紛らわすようなことをした	不安からの逃避	0.480
					0.870
清水他 (2022)	象徴的障害者偏見尺度	13	障害者は一生懸命努力しても, 目標をたいてい達成できない	個人主義	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		13	日本において，障害者に対する差別はもはや問題事ではない	現状の理解の無さ	
	構成世界信念	12	不公平に苦しむ全ての人々が報われる日がいつか来るに違いない	究極的構成世界信念	0.880
		12	どんな人であっても自分の働いた悪事の報いはいつか受けるものである	内在的構成世界信念	0.920
		12	世の中の大抵のことは不公平だ	不公正世界信念	0.830
	能力の認知	3	障害者は有能である		0.850
	保護の対象としての認知	1	障害者は守られるべきである		
	第三者公正感受性	10	他の人より不当に苦しい生活を送っている人がいると，落ち着かない気分になる		0.850
	弱者救済規範意識	5	虐げられている人を，まず敬うべきだ		0.790
	傲慢さの認知	1	障害者はわがままである		
外山・長峯 (2022)	困難な目標への対処法略尺度	25	その目標のことは考えないようにする	目標断念	0.870
		25	新しい他の目標にやりがいを見出す	目標の内容の調整	0.880
		25	最初に設定した目標よりも少しだけレベルを下げて取り組む	目標の水準の調整	0.900

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		25	その目標を達成数 r 為に他に良い方法はないか考える	目標達成宝暦の調整	0.890
		25	達成することができないとしても，その目標を追求し続ける	目標継続	0.880
		25	その目標のことは考えないようにする	目標断念	0.800
		25	新しい他の目標にやりがいを見出す	目標の内容の調整	0.810
		25	最初に設定した目標よりも少しだけレベルを下げて取り組む	目標の水準の調整	0.820
		25	その目標を達成数 r 為に他に良い方法はないか考える	目標達成宝暦の調整	0.810
		25	達成することができないとしても，その目標を追求し続ける	目標継続	0.810
	ホープ尺度	8	窮地 (困難な場面) から逃れる多くの方法を考えつくことができる	ホープ尺度	0.740
	努力の粘り強さ	6	私は頑張り屋だ	努力の粘り強さ	0.770
	認知の柔軟性尺度	10	困難な状況について対応する前に，複数の選択肢を考慮する	代替	0.850
	適応的諦観尺度	10	失敗してもなんとかやっていける気がする		0.890
	主観的 well-being	9		抑うつ	0.850

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		9		不安	0.830
		9		人生に対する満足	0.890
	心理的 well-being	12		人格的成長	0.840
		12		人生における目的	0.870
		12		自己受容	0.900
		12		環境制御力	0.890
	困難な目標への対処法略尺度	25	その目標のことは考えないようにする	目標断念	0.820
		25	新しい他の目標にやりがいを見出す	目標の内容の調整	0.830
		25	最初に設定した目標よりも少しだけレベルを下げて取り組む	目標の水準の調整	0.830
		25	その目標を達成する為に他に良い方法はないか考える	目標達成宝暦の調整	0.810
		25	達成することができないとしても，その目標を追求し続ける	目標継続	0.840
永谷他 (2022)	日本語版 BRIEF-T	86		抑制	0.930
		86		シフト	0.860

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
		86		情緒コントロール	0.640
		86		開始	0.900
		86		ワーキングメモリ	0.890
		86		計画/組織化	0.920
		86		整理	0.910
		86		モニタ	0.920
		86		抑制	0.920
		86		シフト	0.880
		86		情緒コントロール	0.900
		86		開始	0.810
		86		ワーキングメモリ	0.880
		86		計画/組織化	0.880
		86		整理	0.920
		86		モニタ	0.890
	日本語版 ADHD-RS	18		不注意尺度	
		18		多動・衝動性尺度	

表 6.2: 心理学研究 92 巻で用いられた尺度項目など (続き)

出典	尺度名など	項目数	項目例	因子名	α 係数
湯他 (2022)	Promotion/Prevention Focus Scale	16		利得接近志向尺度	0.838
		16		損失回避志向尺度	0.857

第 7 章

距離を分析する

前章では心理尺度の理論と実践を見てきました。理論的仮定から遠く離れて，実践的に「聞けばわかる」といわんばかりにさまざまな “心” の状態が測定されていることが明らかになってきました。

心がどのような状態なのか，言葉にすることが可能なのか，言葉にしたものは心とは違う何かなのか，といった哲学的な問題については，これから心理学業界全体をあげて考え直さなければならない問題です。あるいはこれまでのように，理論的正当性は欠いたままかもしれないが，測られたものによって言論空間が成立し，何かしら妥当性を感じさせる議論ができればそれでよい，というプラグマティックな考え方を選択することになるかもしれません。

こうした問題について，この講義の中で答えが出せるわけではありませんが，ここからは方法論的な改善という観点からアプローチしてみたいと思います。心理尺度の使われ方の問題は，等間隔性を仮定して数値化されたデータに対し，因子分析によって潜在変数を抽出し，見出された構成概念が実在するものと考えて論じるという一連の研究実践で醸造されてくるものでした。等間隔性の仮定や因子分析という手法については，測定モデルを工夫することで修正できる可能性が残されています。

以下では測定されたものが順序尺度水準や名義尺度水準といった，より低い尺度水準の値であることを想定したモデルの応用を考えていきたいと思います。実は心理学において，因子分析による尺度作成という慣例ができあがる前は，さまざまな分析手法が提案，適用されてきました。今一度これらの分析モデルに目を向けることで，新しい可能性について考えてみたいと思います。

7.1 距離と心理学のデータ

私たちが単位もない不確かなものを対象にしながらもそれを測定するモノサシをつくることができるのは，2 つの対象にたいしてその比較をして一方が他方よりも大である，と比較できると考えるからです。記号を使って言えば，x と y を比べて $x \succeq y$ である，ということから，尺度上の値 $p \geq q$ を対応させるということが，尺度を作るということです[*1]。このとき比較する 2 つが重さや広さのような物理的なものであればわかりやすいですし，「痛み」や「喜び」といった心理的な要素であっても構いません。あるいは「より賛成」といった態度表明のようなものでも良いかもしれません。ただしそれらがちゃんとした判断基準に沿って行われている，いいかえるなら**合理的な意思決定をしている**というためには，次の条件を満たす必要があります。

完備性　集合の任意の 2 つの元を取り出したとき，同じかどちらか一方が大である，と判断できること (わか

[*1] 経済学では学問の最初の段階で，財を源とした集合に対する二項関係 $x \circ y$ を定義し，それを効用関数 u で実数領域に写像して $u(x) \geq u(y)$ を考える，といった基本的な原理を抑えます。心理学ではなぜか，「集合」「元」「二項関係」という基本的な比較の要素について考えることなく，その分析ツールだけがどんどんと発展してきています。

らない, はナシ)。

推移性 $x \succeq y$ かつ $y \succeq z$ ならば $x \succeq z$, が成り立つこと。

この 2 つの条件が満たされているとき, **弱順序 (weak order)** が成立している, といいますが, 心理学のデータの場合ははてさて, このような合理性の前提すら怪しいところがあるというのが難しいところです。

ともかく, 人間の心理的な状態の判断を考えれば大小関係の評価ができるかどうかというところから怪しいわけですが, 何らかの大小関係, 選好の程度比較, 類似の程度などが量的に判断できるとするなら, 刺激間の関係を**距離 (distance)** という考え方で表現できます。

あらためて**距離 (distance)** とは何かを考えてみましょう。2 点 x, y の距離を $d(x, y)$ とすると, 距離とよばれる数字の条件は次のようになります。

非負性 $d(x, y) \geq 0$, 距離が負になることはない。

非退化性 $d(x, y) = 0 \Leftrightarrow x = y$, 距離がゼロということは同じ場所にいるということ

対称性 $d(x, y) = d(y, x)$, 一方から他方への距離は, その逆とおなじであること

三角不等式 $d(x, z) + d(z, y) \geq d(x, y)$, 第 3 の点 z を経由した距離は, 直接の距離よりも必ず大きくなること

もっとも一般的に使われるのはユークリッド距離で, 2 次元座標 $(x_1, y_1), (x_2, y_2)$ があった時のユークリッド距離は次のように計算します。

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

3 次元座標 $(x_1, y_1, z_1), (x_2, y_2, z_2)$ の場合は項を増やすだけです。

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

このようにして計算される距離ですが, 上の条件を考えると何も二乗してルートを取らなくても絶対値を足し合わせるような方法でも構いません。たとえば 2 次元座標の場合は, 次のような計算でもいいのです。

$$|x_1 - x_2| + |y_1 - y_2|$$

現にこのような距離のことを**マンハッタン距離 (Manhattan distance)** といいます。

また, ここまでは 2 乗して平方根をとってきましたが, これを一般化して p 乗して p 乗根をとる, と考えるとこれは**ミンコフスキー距離 (Minkowski distance)** とよばれます。

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - x_j|^p \right)^{\frac{1}{p}}$$

他にもマキシマム距離, バイナリ距離, チェビシェフの距離, キャンベラ距離など, 上記の条件を満たすさまざまな距離が考えられています[*2]。また, 相関係数 r_{ij} は $-1 \leq r_{ij} \leq +1$ の範囲にありますが, この絶対値を使って $1 - |r_{ij}|$ とするとこれも距離の条件に当てはまります。どの程度類似しているかというのも, 距離と考えることができるのです。

距離は (尺度のように間主観的な抽象空間にあるのではなく), 個々人の中にある評価基準として考えることができますし, 可能ならこれを集約して社会的なレベルで考えても良いかもしれません。

また距離とは類似度でもありますから, 何を距離と見なすかによって, さまざまな心理学的刺激がデータを形作ることになります。たとえば評定尺度で, n 個の項目で何らかの評価をしてもらったとします。x_{ij} を i 番

[*2] これらの距離はいずれも R の `dist` 関数のオプションで選ぶことができますので, ヘルプを参照してみるとよいでしょう。

目の項目における対象 j の評定値だとすると，$d(j,k) = \sqrt{\sum_{i=1}^{N}(x_{ij} - x_{ik})^2}$ とすれば対象の類似度が計算できます。尺度値になんらかの根拠が必要ではありますが，もし人が同じカテゴリには同じパタンで反応するというのであれば，個人内レベルでの一貫した尺度として用いることができるかもしれません。

あるいは，尺度によって繰り返し類似した項目を重ねるのではなく，回答者の自然な総合的判断で「意見 A と意見 B の類似度はどうか」と直接両者の関係を評価してもらうこともできます。項目群を作るというのは研究者があらかじめどのような評価軸があるかを決め打ちしているようなものですから，個々人の判断基準を不用意に深掘りするのではなく，総合的な判断を求めるほうが良いかもしれません。総合的な判断をいきなり求めることで，社会的望ましさなどのバイアスの影響を受けない形でデータを得ることができますし，評価が安定して得られているかどうかはデータ収集時にその一貫性をチェックすることもできるでしょう。

本書ではここまで散々，心理尺度の安易な使い方を批判してきましたから，それに共感して尺度評定から類似度を算出する，あるいは言語報告によって類似度を直接量的に評価させることはやはり許せぬ，というひともいるかもしれません。しかし心理学の測定法は尺度だけではありません。他にもより客観的な，あるいは実験的に明確な，かつ安定的な指標として取り出すことができる二項関係として，以下のようなものがあります (高根, 1980)。

■刺激の混同率 たとえば複数の刺激があったときに，刺激 A に対する反応と刺激 B に対する反応を取り間違えてしまう，といったことがあるかもしれません。こうした刺激の混同率を両者の類似度と考えれば，それを距離と見なすことができます。よく引き合いに出される例ですが，モールス信号による送受信のエラーなどがこのケースです。送信者が A(.-) を送っているのを，B(-...) と間違えることは少ないかもしれませんが，L(.-..) と間違えてしまうことはあるかもしれません。この場合，A-B の距離は A-L の距離よりも大きいといえるでしょう。他にも，「アボカドに醤油をかけたらトロみたいだ」とか，「プリンに醤油をかけたらウニみたいだ」という組み合わせの妙がありますが，これは「アボカドと醤油」と「トロ」の距離が近い，というように判定することができるでしょう。これも刺激の類似性・混同率に基づく距離の測定といえそうです。

■連想価 記憶の実験ではしばしば無意味綴りを利用しますが，気をつけないといけないのはランダムに言葉を繋いでも，それが何かの意味を持ってしまう可能性があることです。梅本 et al. (1955) は清音 2 音節の有意味度の研究をしていますが，「イカ」とか「ムシ」は漢字や連想したものを思い出しやすく，「ヌヨ」とか「リニ」はほとんど何も連想することがないですね。この研究のように，同じ程度の意味を持つ程度を実験的に集めることができるように，ある刺激から別の刺激へのアクセスのしやすさを距離としてデータとみなすことができます。

■刺激の汎化勾配 ある刺激 S に対して反応 R を誘発するように条件づけ実験を考えるとします。このとき，刺激 S_i から反応 R_i がみられる程度を確率で表現し，これを S_i と R_i の近さ，類似度と見ることができます。

■反応潜時 2 つの刺激が「同じ」か「違う」かを判断する課題を用意し，反応潜時を類似性の指標と考えることもできるでしょう。

■ソシオメトリックデータ 学級内の友人選択を調査や観察で記録し，小集団における人間関係の選好やネットワークの有無を両者の類似度とみなすこともできます[*3]。

[*3] ソシオメトリーの創始者 J.L.Moreno は，集団療法の一環として社会関係を様々なレベルで測定することを考え，そのもっとも直接的なレベルとして学級の児童生徒に対して選好関係を聞くという方法を提案しています。ただし調査して分析することが目的なのではなく，集団関係のカルテとして使うことが目的であり，ソシオメトリックテストはその一側面的補助資料に過ぎないという考

このようにいろいろなものが「類似しているかどうか」の指標として使えます尺度評定よりも，具体的な 2 つの刺激が似ているかどうかの反応の方がやりやすい，というのは誰しも実感としてわかることではないかと思います。さてこのように距離とみなすことができるデータが手に入れば，これをもとにデータの特徴なり個々人の特性なりを描き出すことができるかもしれません。

7.2 クラスター分析

刺激間，変数間，あるいは対人的な関係の近い・遠いなどの関係を，距離行列で表すことができたら，ここからどのような分析が可能でしょうか。

分散共分散行列も変数間の関係を表した行列でしたが，因子分析は変数間関係の背後に測定モデルを考えて生まれています。しかしそうしたメカニズムを抜きにして，データだけから読み取れることを考えてみたいと思います。ひとつは「類似性」を表現しているのですから，これをもとに分類することができるでしょう。分類は科学的アプローチの最初のステップで，表面的な特徴をもとに分類したうえでそのグループが何によってまとめられているのか考察することができます。これを機械的にやる方法が，**クラスター分析 (Cluster Analysis)** というものです。クラスターとはまとまり，塊という意味で，データの中から似ているものは同じクラスター，似ていないものは別のクラスターとして分類していきます。クラスター分析の方法はさまざまなものが考えられており，大きく分けて階層的なモデルと非階層的なモデルに分類できます。

7.2.1 階層的クラスター分析

階層的クラスター分析とは，クラスターが階層をなすように分類していく手法です。

たとえば 5 つの対象 A,B,C,D,E があって，その距離行列が得られているとします。

1. ここでもっとも距離の小さいペアを 1 つのクラスターとします。例えば A-B 間の値がもっとも小さければ，この 2 つをまとめて F という塊にしてしまいます。
2. C,D,E,F の距離行列を考え，そこでもっとも距離の小さいペアを次のクラスターとします。たとえば C-D 間の値がもっとも小さければ，この 2 つをまとめて G という塊にします。
3. E,F,G の距離行列を考え，そこでもっとも距離の小さいペアを次のクラスターとします。たとえば F-G 間の値がもっとも小さければ，この 2 つをまとめて H という塊にします。
4. さいごに E-H をまとめて塊にしてしまいます。

このようにクラスター化，クラスター化を次々重ねていくので，階層的なクラスターとよばれるわけです。

ここでのポイントは，まとめて作られた新しい点と他の対象との距離，今回の例でいえば第 2 段階の F と C,D,E との距離をどう考えるかという問題がでてきます。C,D,E は既にデータとして距離が与えられていますが，新しい F は A と B が合体したものなので，何かルールを決めないとこの段階では距離が定まっていないのです。ここでたとえば F と C の距離を考えるときに，C-A と C-B の短い方を採用する場合 (**最短距離法**)，長い方を採用する場合 (**最長距離法**) だけでなく，重心を考える場合や，クラスター内分散とクラスター間分散の差が最小となるように取る**ウォード法 (Ward method)** など，色々なパタンが考えられています。またこうした手法の違いによって，結果も変わってくるので注意が必要です。

え方でした。社会心理学領域で，ソシオメトリックテストが導入されたときは，治療や介入ではなく測定だけに注目しましたが，これは「冷たいソシオメトリー」であって，治療目的などで実践的に活用する「温かいソシオメトリー」」が目指すところではない，という批判があります (田中, 1960)。いずれにせよ，当事者の言葉による表出をそのままスコアリングすることの危険性は，心理尺度と同様，深く注意しなければならない問題です。

7.2.2 非階層的クラスター分析

クラスターのクラスター，というような階層性を考えない方法を非階層的クラスター分析といいます。有名な **k-means 法**とは，次のようなアルゴリズムで分類します。

1. 分けるべきクラスター数 k を考え，変数間関係の距離空間のなかに k 個の重心を適当に取る。
2. クラスターの重心と対象との距離を計算し，対象はそのもっとも近い重心点をもつクラスターに分類されるものとする。
3. 構成されたクラスターの重心を算出し，新しい重心とする。
4. あらためて新しく計算された k 個の重心と対象との距離を計算し，クラスターに再分類する。
5. 重心の計算，クラスターの分類，というプロセスを反復し，変化が生じなくなれば (収束すれば) 分類が完了したものとする。

この方法ですと，反復計算のなかでクラスターのクラスターという考え方は出てきませんから，最初に設定した k 個のクラスターが最終的に得られることになります。問題は k 個に分類するという，分類の数が事前に与えられることで，なぜ k 個なのか，他の数ではダメなのかという問題に答えることができません。

こうした問題に答えるためには，データが何らかの特徴に基づいて得られているというモデルを想定し，そのモデルの適合度でもってクラスターの数や分類を考えることになります。こうしたクラスタリング手法を，**モデルベース・クラスタリング**と呼ぶことがあります。

たとえばデータがいくつかの正規分布によって生成されていると考えるものが，**正規混合モデル (normal mixture model)** です。このモデルを使うと，統計的な観点からデータの背後に隠れている正規分布の数をいくつにするのがもっとも適合度が高いか，という基準で判断することができますので，クラスター数の恣意性問題に答えることができます。

また確率を使ったモデルになりますから，ある対象 A がクラスター 1 に分類される確率，クラスター 2 に分類される確率･･･というように分類の分かれ方が少し曖昧になったりします。明確な分類をするものをハードクラスタリング，確率などで曖昧な分類を許すものをソフトクラスタリングなどと呼んで区別することもあります。

ここでみたクラスタリングの手法は，距離データの背後に構造や生成メカニズムを考えていません。あくまでも表面的な類似性だけから分類するため，さまざまな現象に応用可能であるという長所を持っています。類似性，あるいは距離と見做せるだけで良いのですから，心理学的な実験データはもちろん，社会調査など集計されたデータ (ex. 国家間の貿易額，国内の人口流動など) でも分析できますし，購買行動のようなデータだけからでも購買層を分類するなど，応用面での利用はさまざまなものが考えられます。心理尺度を使った分析でも，因子分析のようなメカニズムを想定することなく，表面的な変数同士の距離関係から分類することもできますし，回答者間の回答パターンの類似性から回答者のクラスターを見つけるということも可能です。モデルが複雑なメカニズムを要求しないからこそ，見えてくるところがあるかもしれません。

7.3 多次元尺度構成法

距離データの分析について，もう少しその構造的特徴を考えたいというのであれば，**多次元尺度構成法 (Multidimensional Scaling)** がいいでしょう。

多次元尺度法をもっとも簡単に説明するならば，「距離から地図を作る方法」ということができるかと思います。具体例で見てみましょう。R にはサンプルデータとして `eurodist` というヨーロッパ各地の都市間距離

のデータが用意されています。表 7.1 にその一部を示します。

表 7.1　ヨーロッパ都市間距離データの一部

	Athens	Barcelona	Brussels	Calais	Cherbourg	Cologne
Athens	0					
Barcelona	3313	0				
Brussels	2963	1318	0			
Calais	3175	1326	204	0		
Cherbourg	3339	1294	583	460	0	
Cologne	2762	1498	206	409	785	0

元のデータは 21 × 21 のサイズです。表から，たとえばアテネ (Athens) とバルセロナ (Barcelona) の距離が 3313km，と言うことが読み取れます。表の右上が空白になっていますが，これは**距離行列**が**対称行列**なので，ムダな情報を表示しないようにしているからです。アテネとバルセロナの距離は，バルセロナとアテネの距離に等しいですからね。

さて距離行列はこのように，**正方行列**ですから，**固有値分解**をすることで**基底**を求めることができます。基底は座標を形成する基本単位ですから，それを使って各変数の位置にあたる座標を計算できます[*4]。このように距離行列から座標を求めて，変数をプロットすることで変数間関係を可視化する手法のことを**多次元尺度構成法 (Multi-Dimensional Scaling: MDS)** と言います。

試しに `eurodist` データを MDS で 2 次元プロットしてみましょう。これを実行するコードは簡単で，`eurodist` データのようにデフォルトで組み込まれている `cmdscale` 関数を使います。

code : 7.1　計量 MDS の実践と描画

```
1  library(tidyverse)
2  library(ggrepel)
3  # MDSを実行。データや関数はRがデフォルトで持っているもの
4  result.MDS1 <- cmdscale(eurodist,k=3)
5  # y軸反転させつつ描画
6  g <- result.MDS1 %>%
7    as.data.frame() %>%
8    dplyr::mutate(label = rownames(.)) %>%
9    ggplot(aes(x = V1, y = V2, label = label)) +
10   geom_point() +
11   geom_text_repel() +
12   xlim(-2500, 2500) +
13   ylim(2500, -2500) +
14   xlab("dim␣1") +
15   ylab("dim2")
```

■コード解説

1-2 行目　データ整形や描画に必要なパッケージを読み込みます。

4 行目　`cmdscale` 関数に `eurodist` データを与えています。`k=3` は 3 次元解を出すように指定しています。地球は球体ですから，地球上の地理は 3 次元で表現できますよね。

[*4] 厳密には距離行列 $\boldsymbol{D}$ そのものではなく，それを二重中心化した行列の固有値分解になりますが，本質的には変わりません。詳しくは岡太 & 今泉 (1994) などを参照のこと。

5-14 **行目** ggplot2 による描画です。結果オブジェクトである result.MDS をデータフレームにし，行の名前になっていた都市名を変数として格納したのち，散布図のようにプロットしています。

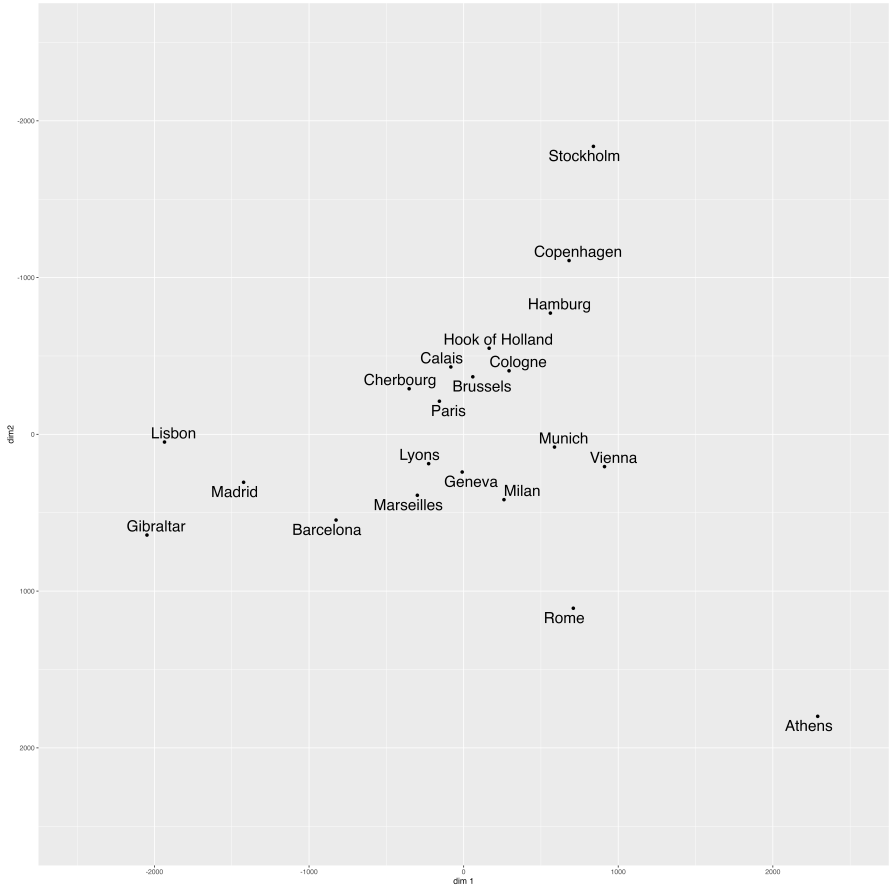

図 7.1 MDS のプロット

図 7.1 をみると，上の方にストックホルム (Stockholm) があって，右下にアテネが，中央にパリ (Paris) が･･･といった配置になっています。地球上の位置と完全に一致しているとは言えませんが，それでも概ねうまくプロットできていますね[*5]。今回はデモストレーションですから，地理 → 距離 → 地図という面倒なことをしていますが，最初の地理情報がなくとも距離関係だけから地図を作ることができるというのが，MDS という手法の利点です。

多次元尺度構成法はその名の通り，「尺度」を構成するための方法です。心理尺度 (態度尺度) のように項目に対する点数化についての仮定を考えなくとも，対象間の類似度から地図＝多次元空間を作ることができるというのは，もしかすると哲学的根拠が薄弱な心理尺度を救う一つの切り口になるかもしれません。人を対象にしたデータから刺激，変数，対象の近い･遠いを考え，MDS で分析し，心の地図を作るのですから，たとえば刺激の混同率のデータなどから「間違いやすいパターンとそうでないパターンはどういう特徴があるか」を，地図にプロットされた対象の**布置 (configurations)** から考察することができるようになるわけです。

ただし今回示した例は，比率尺度水準という最高レベルの測定方法で構成されたデータから地図を作っていたのでした。元のデータを固有値分解するということは，そのデータが加減乗除の計算に耐えうる水準の情

[*5] 完全に一致しない理由は，地球が球面であるのに対し平面にプロットしたから，という側面もあります。

報を持っていなければならないわけですが, 心理学のデータの場合はもしかするとその水準を持っていないかもしれない (というかおそらく持っていない) でしょう。

では MDS は絵に描いた餅か, というとそうではなく, むしろこの問題を解決し, さらにさまざまな拡張・応用が考えられています。次節ではこの点について, いろいろみていきたいと思います。

第 8 章

多次元尺度法とその応用

多次元尺度構成法 (MDS) は距離データから地図を作る方法であり，データに複雑なモデルを仮定しないためさまざまな適用例を考えることができます。ここではその応用例をいくつかみていくことにします。

その前に，分析するデータの種類について新しい概念を導入します。

8.1 データの相と元

数値として得られたデータがどの程度の算術操作に耐えうるかについては，尺度水準によって分類できるのでした。これとは別に，データがどのような要素の組み合わせでできているか，という観点から分類することがあります。それがデータの**相 (mode)** と**元 (way)** という区分です。

データの相とは，データを構成する要素の種類の数を指します。たとえば MDS で例に挙げた「都市間の距離」をみてみると，距離行列は都市 × 都市のデータセットです。つまり要素の種類としては「都市」の一種類しかありません。このようなデータは一相のデータといいます。あるいはまた，一般的な調査データを考えてみましょう。スプレッドシートに数値情報が並びますが，一般に一行に 1 ケース (個体，オブザベーション)，一列に 1 変数の矩形データです。これは個体 × 変数なので，データを構成する要素は 2 つありますから，二相のデータといいます。今度は追跡調査，つまり同じ人に複数回調査を行う例を考えてみましょう。スプレッドシートには「調査時期」という変数列が 1 つ増えただけと考えることもできますが，その中身は個体 × 変数 × 時点，という 3 つの要素から成り立っている 3 相データと考えることができます。このように含まれる要素の (種類の) 組み合わせであり，相は「そのまとまりから何らかの共通点がある」と考えて情報を探るべき側面という意味合いがあります。いろいろ複雑にしていくことはできますが，多くても 3，4 相ぐらいのデータセットまでになることが一般的です[*1]。

相は要素の種類の数でしたが，これに対して，データの元とはその組み合わせ回数のことを意味します。最初の都市間距離のデータは一相ですが，都市と都市の組み合わせで距離が作られます。都市を二回使っているので 2 元データです。両者を合わせて一相二元データといいます。一般的な調査データの場合は，個体 × 変数なので二相二元データ，追跡調査の場合は三相三元データということになります。特殊な例ですが，家庭内での人間関係を把握するための調査を考えてみましょう。父，母，子からなる家族を対象に，「お父さんはお母さんのことをどのように考えていますか」，「お父さんはお子さんがお母さんのことをどう考えているとお思いですか」といった調査です。家族構成員ひとりにつき，「父・母・子が」「父・母・子に対して」どのように考えているか，ということを尋ねますので，構成員 × 父母子 × 父母子の二相三元データということができます。ここで，評価する主体と評価される客体は同一人物であっても異なる側面である，と考えることもできま

[*1] 筆者が今まで見たことのある多相データのなかでもっとも複雑なのはテレビの視聴調査に関するもので，個体 × 変数 × 放送時間帯 × ジャンルという 4 相のものでした。

しょう。その場合は，構成員 × 主体 × 客体と考えて三相三元データになります。

つまり，相とは「その側面の特殊性・特徴を取り出して分析したい」と研究者が注目したい側面であって，同じデータであってもどのような情報を取り出すか，どこに情報が潜んでいるとみなすかによって扱いが変わるということです。一般的な調査で，因子分析をするプロセスの最初に $\boldsymbol{R} = 1/n\boldsymbol{Z}'\boldsymbol{Z}$ という計算がありました。これはつまり回答者の相を要約して潰し，変数の相だけの一相二元データ (変数 × 変数) に作り替えたことを意味します。もし個人の情報が重要であると考えるなら，個人の相から直接情報を取り出すことを考えるのも良いことかもしれません。**構造方程式モデリング (Structural Equation Modeling)** などを使って，注目したい相に含まれる潜在変数を仮定することで考察の光を当てることができます[*2]。

皆さんもデータの分析にあたって，どの側面のどの情報を重要とみなすか，あるいはどの情報であれば要約可能であると考えられるかに，少し配慮してみてください。

8.2 非計量多次元尺度法

さて前章では，一相二元データの代表例である距離データをつかって，クラスター分析や MDS によって分類したり可視化したりできるということを見てきました。ただし，この場合の距離行列とは，間隔尺度水準以上であることが必要なのでした。しかし心理学的な刺激に対する反応をデータ化する時は，同時に被験者はそこまで鋭敏に反応しているのか，いいかえれば本当に間隔尺度水準以上の精度で判断できているのかな，という人間側の問題が気になりますね。そこまで人間は鋭敏な反応をしていないかもしれません。

数値的な厳密さを持つデータのことを**計量 (metric)** データと呼ぶことがありますが，人文社会科学的なデータはせいぜい順序尺度程度の違いしかもたない**非計量 (non-metric)** データです。この非計量データを加工して，数値的特性を取り出そうとしても，元の数字に信憑性がないのですから結果も推して知るべしです。

でも大丈夫。MDS には非計量データに対応したものがあるのです。**非計量的多次元尺度構成法 (Non-Metric Multi-Dimensional Scaling)** と呼ばれる手法がそれです。非計量 MDS では，対象 j と k の類似度を δ_{jk} とし，分析によって埋め込む多次元空間での距離を d_{jk} としたときに，次の関係が保持されることだけを考えます[*3]。

$$\delta_{jk} > \delta_{lm} \text{ならば} d_{jk} \leq d_{lm}$$

これは評定値のような非計量データの世界を参考に，距離関係が保持されている仮想的・理論的な空間を作ってそこでの布置を考えようというモデルです。評定の世界と構成される世界が (対応関係はあるものの) 別立てになっているところに注目してください。

多次元空間をどのように構成するかについては，埋め込まれる空間での距離 d_{jk} を配置する仮想空間上での，推定される距離 $\hat{d}_{jk}$ の差の二乗和，特に

$$stress = \sqrt{\frac{\sum\sum_{j\neq k}(\hat{d}_{jk} - d_{jk})^2}{\sum\sum d_{jk}^2}}$$

が最小になるように，反復計算で座標を求めていくのが古典的な Kruscal の方法とよばれているものです (Kruskal, 1964b)。ここで定義した不適合度は特に**ストレス (Stress)** と呼ばれ，Kuruscal の提案の他にもい

[*2] 手前ミソながら具体例として小杉, 清水, et al. (2011) を挙げておきます。

[*3] 一般的な統計の文脈では，推定値をギリシア文字で，実測値をアルファベットで表記する習わしですが，MDS の文脈では逆に仮想空間の値・座標をアルファベットで，生データによる測定値をギリシア文字で表現することが一般的です。なんでだろう。

くつかの方法が考えられています。このストレス値が小さければ小さいほど，データと埋め込まれた空間の差が小さい，つまり当てはまりが良いと考えるのです。

非計量 MDS は R の MASS パッケージや smacof パッケージに実装されており，簡単に試すことができます。具体例で見てみましょう。

8.2.1 非計量多次元尺度法の例

ここでは MASS パッケージに含まれる isoMDS 関数を使って実践してみます。

使うデータは，2021 年末の M-1 グランプリの評定をデータ化したものです[*4]。2021 年度のスコアは表 8.1 にあるようなものでした

表 8.1 M-1 グランプリ 2021 の採点結果

演者	巨人	富澤	塙	志らく	礼二	松本	上沼
モグライダー	91	93	92	89	90	89	93
ランジャタイ	87	91	90	96	89	87	88
ゆにばーす	89	92	91	91	93	88	94
ハライチ	88	90	89	90	89	92	98
真空ジェシカ	90	89	92	94	94	90	89
オズワルド	94	95	95	96	96	96	93
ロングコートダディ	89	90	93	95	95	91	96
錦鯉	92	94	94	90	96	94	95
インディアンス	92	91	93	94	94	93	98
もも	91	90	91	96	95	92	90

M-1 の採点は審査員各自の主観に基づいて行われ，得点の絶対値はそれほど重要ではないかもしれません。すなわち，松本人志の 80 点が上沼恵美子の 80 点と同じぐらいの面白さを評価しているか，ということについては真偽判断ができないでしょう。それでも各審査員の中での相対的評価には，一貫性がありそうです。すなわちある審査員が漫才師 A に 80 点，漫才師 B に 85 点をつけたのなら B の方が面白かったということでしょうし，漫才師 C が 83 点なら $A < C < B$ という順序はあると思われます。つまり**順序尺度水準**程度の質はあると仮定することに無理はなさそうです。

そこでこのデータをもとに，10 組の漫才師の類似度を計算します。類似度はユークリッド距離を用いることにします。ユークリッド距離は既に述べたように差分の二乗を総和して平方根を取ったものですが，具体的な数字で見た方がわかりやすいかもしれませんので，表 8.2 を用意しました。表 8.2 にモグライダーとランジャタイの二組だけ取り出し，これで計算例を見てみます。それぞれの得点の差分，その二乗を計算し，それを総和したところ 103 という値になっています。これの平方根を取ったもの，すなわち $\sqrt{103} = 10.14889$ がこの二組の距離，すなわち非類似度ということになります。この計算を全ての組み合わせについて計算してくれるのが，dist 関数なのです。

[*4] 念のために解説しておきますが，M-1 グランプリとは 2001 年から始まった漫才の賞レースの 1 つで，年末に年間チャンピオンが決定します。開催年ごとにルールが少し変わることもありますが，基本的には予選を勝ち抜いた 10 組の漫才師が 4 分間のネタを披露し，6-7 名の審査員が 100 点満点で採点します。点数の上位 3 組が決勝戦を行い 2 本目のネタを披露，投票によりチャンピオンが選出されるという流れです。2021 年はオール巨人，富澤たけし，塙宣之，立川志らく，中川礼二，松本人志，上沼恵美子が審査員で，最終的には錦鯉がチャンピオンになりました。このデータセットは 1 本目のネタについての採点を 2001 年から集めたものになります。

表 8.2 距離の計算

	巨人	富澤	塙	志らく	礼二	松本	上沼	総和
モグライダー	91	93	92	89	90	89	93	637
ランジャタイ	87	91	90	96	89	87	88	628
差分	4	2	2	-7	1	2	5	9
差分の二乗	16	4	4	49	1	4	25	103

code : 8.1 距離行列の計算

```
1  dat <- read_csv("M1score2021.csv")
2  dat.mat <- dat %>%
3    dplyr::filter(年代 == 21) %>%
4    arrange(ネタ順) %>%
5    dplyr::select(-年代, -ネタ順) %>%
6    pivot_longer(-演者) %>%
7    na.omit() %>%
8    pivot_wider(id_cols = 演者,
9             names_from = name,
10          values_from = value) %>%
11   as.matrix()
12 rownames(dat.mat) <- dat.mat[, 1]
13 dat.mat <- dat.mat[, -1] %>%
14   dist()
```

■コード解説

1 行目 データファイルを読み込み，dat オブジェクトに格納します[*5]。

2-9 行目 必要なデータだけに絞り込む操作です。流れを解説しますが，他のやり方でもいいですしできあがったものが何かだけわかれば結構です。

3 行目 2021 年のデータだけに絞り込みます。

4 行目 ネタ順に並び替えています。

5 行目 年代変数とネタ順変数はもういらないので削除してしまっています。

6 行目 ロング型に変換しています。これで演者-審査員-採点のデータセットができます。

7 行目 欠測値を除外しています。実はこれがこの操作の目的で，というのも過去の審査データも入っているものですから，過去の審査員も大量に変数として含まれていて，それらが欠測値になってしまっていたのです。

8-10 行目 元のワイド型に戻しています。

11 行目 以下の行列処理のため，data.frame 型から matrix 型に変換しています。

12 行目 変数として一列目に演者名が入っていますが，これを行列の行名に入れています。matrix 型は行名・列名をデータの外に持つのです。

13-14 行目 変数としての演者名を除いて，パイプで dist 関数に入れ，距離行列を作っています。

できた距離行列は，表 8.3 のようになっています。モグライダーとランジャタイの距離が，先ほどの例で計算した値と一致していることを確認してください。

[*5] このデータセットは伴走サイト，https://kosugitti.github.io/psychometrics_syllabus/ にサンプルデータとし置いてあります。

表 8.3 演者の非類似度行列 (演者名は略記)

	モグ	ラン	ゆに	ハラ	真空	オズ	ロン	錦鯉	インデ	もも
モグ	0.000									
ラン	10.149	0.000								
ゆに	4.583	9.165	0.000							
ハラ	7.937	12.806	7.616	0.000						
真空	8.660	7.483	7.071	11.832	0.000					
オズ	12.490	15.652	12.207	14.933	11.000	0.000				
ロン	9.381	11.446	6.403	9.110	7.416	9.487	0.000			
錦鯉	8.485	15.264	8.307	10.909	10.247	7.071	7.874	0.000		
インデ	9.381	14.107	8.062	8.660	9.950	8.124	4.472	6.325	0.000	
もも	10.100	9.110	8.307	12.207	3.606	8.718	6.782	9.592	8.718	0.000

あとはこの距離行列を `isoMDS` 関数に渡すだけです。`isoMDS` 関数は引数として，何次元の解を求めるかを設定できます。地理データであれば 2,3 次元から作られていることは明らかですが，この評価が何次元かは事前にわかりません。そこで次元数を 1 から 7 まで変化させながら，Stress 値がどうなるかを表したのが図 8.1 になります[*6]。

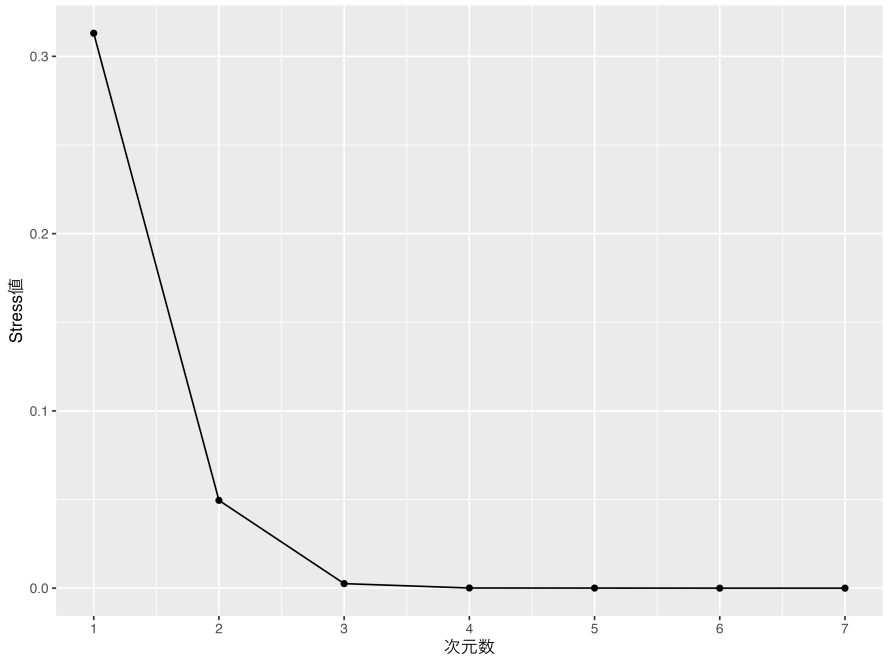

図 8.1 Stress 値の減衰

まるで因子分析の**スクリープロット**のようですね。横軸に次元数，縦軸に Stress 値を置いた折れ線グラフですが，当然のことながら反映させる MDS 空間の次元数が増えるとデータとの距離が縮んでいきます。

[*6] どうして 7 までか，というと評定者が 7 名だからです。7 名それぞれの評価次元があると考えると，この距離空間の中には最大 7 次元あるはずですから。あるいは 10 組の演者がいますから，組み合わせの自由度から考えて 9 次元まで試してもいいと思います。

Kruskal (1964a) の基準によれば，Stress 値は表 8.4 のように評価できます。この基準でいくと，今回は 2 次元で 4.9%(0.0049534) ですから，2 次元解で OK としましょう。

表 8.4　Stress 値の評価

Stress	Goodness of Fit
20%	poor
10%	fair
5%	good
$2\frac{1}{2}$%	excellent
0%	“perfect”

演者をプロットしたのが図 8.2 です。東西南北というか，上下左右の軸に特に意味はありませんから，因子分析のように因子軸の解釈や命名をすることはありません。近くにプロットされた対象は評価が似ていたんだな，ということがわかりますし，相対的な位置関係から，「ハライチとももは芸風が真逆だな」とか「ロングコートダディは中心に近いから中庸的な笑い，悪く言えばキャラが立ってないんだな」といったことが読み取れます。

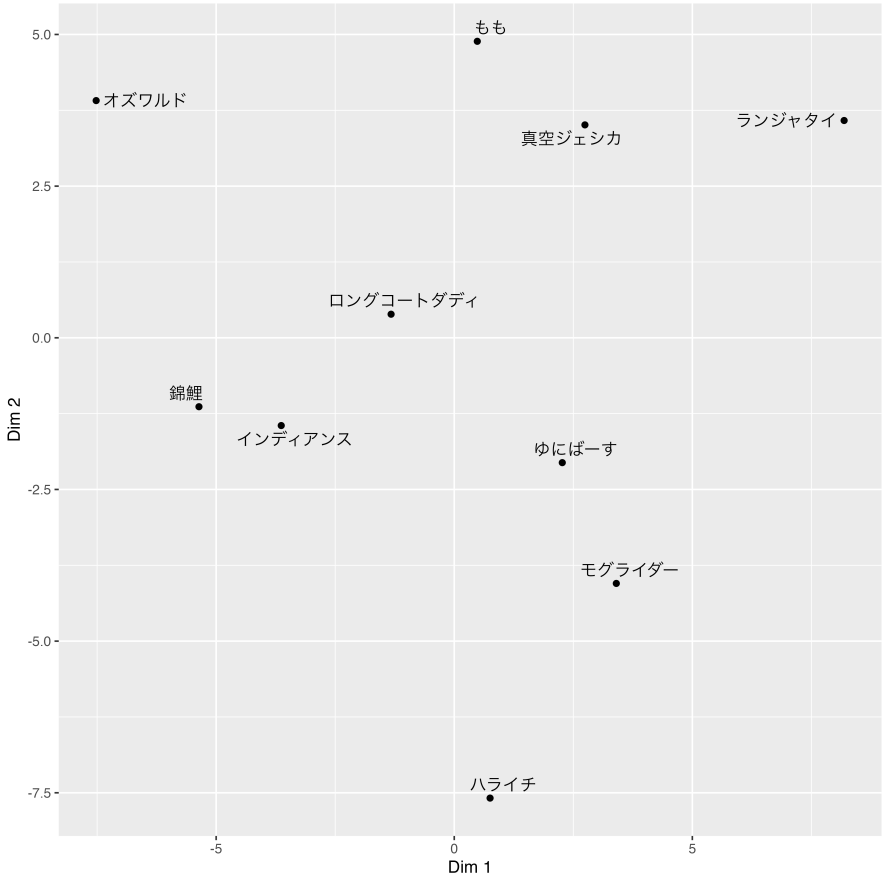

図 8.2　非計量 MDS のプロット

また MDS では，地図を書くことはできても東西南北を固定することはできませんから，因子分析のように軸の解釈をすることはあまりありません。しかし座標を説明変数として，外的な被説明変数に回帰分析をすることで，軸が何を反映しているのかを考えることができます。これについては Grimm & Yarnold (1994) の

MDS の章を参照してください。あるいは，MDS の結果と同じデータのクラスター分析の結果を組み合わせて考察することもあります。

このように，非計量 MDS は優劣・大小関係という順序尺度水準の評定だけからでもその空間的な特徴を描くことができますから，心理尺度のもつ仮定や測定モデルを考える必要がないため，今後その重要性が再発見されていくのではないかと思います。しかし MDS の面白さはこれだけにとどまりません。測定論的にも方法論的にも，大きな可能性を秘めた応用モデルが考えられています。

8.3 MDS の応用モデル

8.3.1 個人差多次元尺度構成法

MDS はある個人が主観的に判定した距離行列からでも，地図を書くことができます。ある人に，「ハライチとランジャタイは (芸風が) 似てると思う?」などと聞いて数字で答えさせ，距離行列を作って非計量 MDS で分析すれば，その人の心の中でお笑い芸人をどのように分類しているかを見ることができるわけです。心理尺度をわざわざ構成しなくても，ある人の感覚を数量的に表現してその大小関係が妥当なものであるならば，その人の心の中での配置がわかるのです。個人の内部状態を個人の基準で判断させるような研究においては，既に指摘したような心理尺度の理論的根拠の弱さは該当しませんから，目的によっては MDS の方がよほど適しているはずです。

しかし気になることがあるとすれば，個々人の心の状態がわかるのは良いにしても，それではケースバイケース，つまり「この人の中ではこうなっているようだ」という限定的な言及しかできないのが苦しいところです。心理学は曲がりなりにも，人間一般の行動なり感覚なりを研究したいわけですから，人生いろいろ，の一言で終わらせてしまうわけにはいかないのです。

でも大丈夫。個人差を踏まえて MDS を考える，つまり二相三元データを分析する MDS の方法も，古くから研究されています。もっとも有名なモデルのひとつが Carroll & Chang (1970) による **INDSCAL**(INdividual Differences SCALing, 個人差多次元尺度構成法)[*7]です。

個人 i が対象 j と k に対して与えた類似度を δ^i_{jk} と表すとしましょう。個人 i にかかわらず，全体的には p 次元空間において

$$d_{jk} = \sqrt{\sum_{t=1}^{p}(x_{jt} - x_{kt})^2}$$

という関係で表現されているとし，個人差を

$$d^i_{jk} = \sqrt{\sum_{t=1}^{p} w_{it}(x_{jt} - x_{kt})^2}$$

として表現します。つまり第 t 番目の次元に個々人の重み w_{it} があるとモデル化し，「人によって重視する次元がある」と考えるのです。モデルの推定にあたっては，$w_{it}0$ と $\sum_t w^2_{it} = 1$ などの制約を課します。

さて図 8.3 には共通次元のプロットと個人 A,B ごとの強調パターンを示しました。図中の X1,X2,X3,X4 が共通次元のプロットを表しています。個人 A は第一次元に重みを持つ人の例で，X1,2,3,4 が A1,2,3,4 のように左右に広がって布置されています。個人 B は第一次元に 1.0 未満の，第二次元に 1.0 以上の重みを持つ人の例で，B1,2,3,4 のように第一次元を短く，第二次元を大きく拡張しています。一般的な X1-X3 の

[*7] これの発音は「インスカル」です。“D” は読まないことに注意。

距離に比べて，A はその両者の違いが大きいと判断し，B はその違いが小さいと判断しているわけです。つまり A は第一次元の方向に敏感で，B は鈍感である，といった解釈をすることができます。

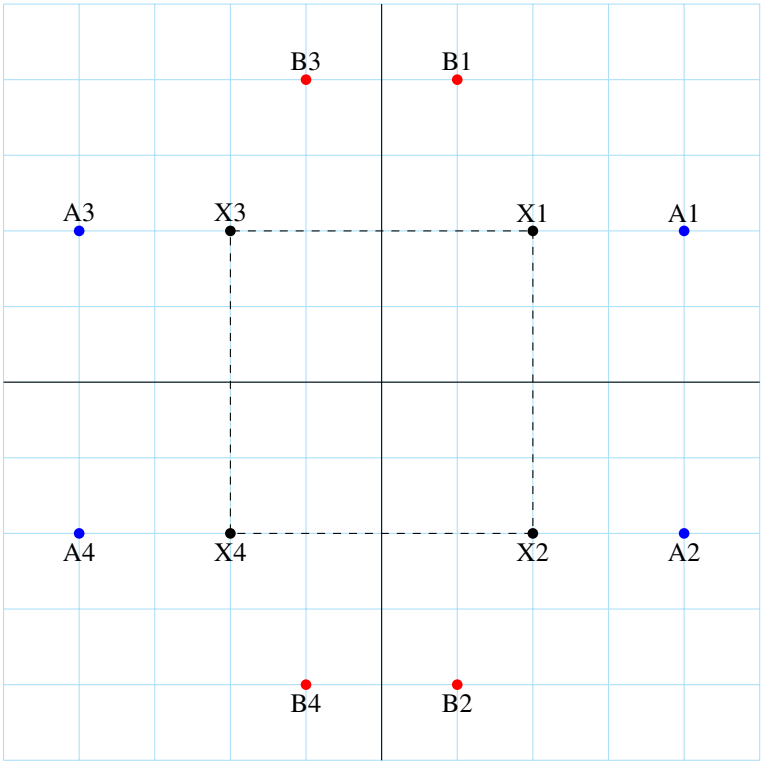

図 8.3 INDSCAL による軸の協調イメージ

このように，INDSCAL を使えば個人の判断次元の重要度をその重みとして表現することができます。ただの MDS と違って INDSCAL は軸の回転に対して不定性をもっていますから，それぞれの軸がどういう意味であるかを解釈することができます。何より，全体の特徴だけでなく個人差をモデルの中に表現することができることは，心理学の研究用途にとって非常に有用な特徴ではないでしょうか。

より慎重に考えるならば，この個人差の表し方は「個々人で原点が同じ」であり，「次元の重みが個人差である」という仮定を置いていることになります。第 6 章で心理尺度の問題として，個々人の評定をどのように合算するのかについて理論的根拠がない，という批判をしました。これについて INDSCAL ではこのような仮定でモデル化する，としているだけで，理論的根拠を示すものではありません。しかし，モデルの形で個人間の取り扱いを明示していることは，しばしばその存在にすら気づかれない水面下の仮定をとるよりも，建設的な議論を可能にします。もし個人差の表し方，個人間の表現に疑義があるのであれば，心理学者も積極的に三相データのモデル化に取り組むべきでしょう。

8.3.2 展開法

ここまでは距離データとして，直接的に変数間の距離が測定されたものとして考えられてきました。つまり一相二元のデータから分析が始まっており，一般的な個人 × 変数の二相二元のデータセットがあった場合は，いずれかの相を潰して分析しなければなりません。これに対して Coombs (1950) は，刺激に対する個人の選好順位から一次元尺度を作る**展開法 (Unfolding method)** を提案しました。この方法では，個人 × 変数の組み合わせで得られるデータを，両者の距離と考えるところがポイントあり，態度尺度作成法の一つとし

ても考えられてきました (藤原, 2001)。ある態度対象 A_j, たとえば政治的態度であれば「自民党」「共産党」などですが, これに個人 i が選好順序をつけたとします。自分は自民党のほうが共産党より好きだ, といった感じです。これは個人 i と態度対象 A_j の距離を表していると考えられます。この距離に応じて, 態度対象が一次元上に並んでいると考えるのです。態度対象が並んでいる連続体を J-Scale といいます。もちろん人によって選好度は異なります。別の人にとっては共産党のほうが自民党より好きだ, となることもあるのです。こうした個人の違いは, J-Scale 上の個人の立ち位置が違うのであって, 個人の尺度は態度と垂直する方向に伸びていると考えます。個々人の尺度を I-Scale といいますが, この尺度は J-Scale を個人の位置で折り畳んだものになっていると考えるのです。

図 8.4 には, 4 つのカテゴリをもつ J 尺度と, 個人 i, k の I 尺度を示しました。個人の J 尺度上の理想点 IP_i, IP_k の上に各カテゴリがたたみ込まれており, 表に出てくるときはこれが展開すると考えるわけです。

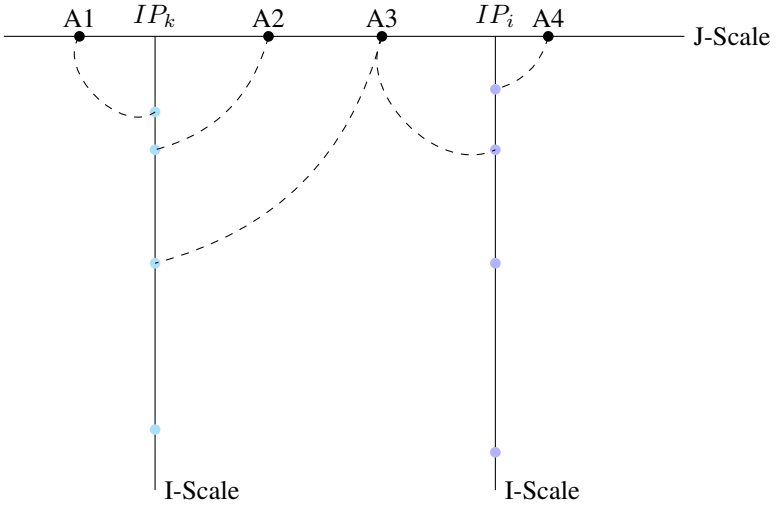

図 8.4 展開法による一次元尺度の構成

これを多次元に拡張したモデルもあります。多次元展開法は, p 次元空間における個人の理想点 P_i と, 態度対象の布置 A_j との距離 $d(P_i, A_j)$ を次のように考えます。

$$d(P_i, A_j) = \sqrt{\sum_{t=1}^{p}(P_{it} - A_{jt})^2}$$

さらに, 個人 i が態度対象 j に対して「4. かなり当てはまる」とか「1. 全く当てはまらない」などと評価するわけですが, この評価 R_{ij} が構成される多次元空間上の距離の関数になっている, すなわち

$$R_{ij} = \alpha - \beta d(P_i, A_j)$$

と考えます。ここで一次関数的に表現していることから, モデル上の距離と評定値との残差を最小にするような定数 (回帰係数) として α, β を求めるということです。

展開法は R の `smacof` パッケージなどでも実行できるようになっています。

この方法のポイントは, 個人と態度対象を同じ空間に埋め込み, その座標を考えることにあります (図 8.5)。評定値が尺度によるスコアと考えるのではなく, 態度対象との距離関係を表しているものとみなして分析するわけです。リッカート的なスコアリングは, 1,2,3... と振られる数字が大きくなるにつれて, その態度の強度が累積的に大きくなっていくことを含意しています。態度とはそのような強度を持つものである, という仮定があ

り，測定しているのが態度であれば問題ないのですが，リッカート法が必ずしも社会的態度だけでなくそれ以外のものも測定するのに用いられている現状を顧みるに，その数字が何を表しているのかについて別の角度から考えるのも，これからの心理測定のヒントになるかもしれません。

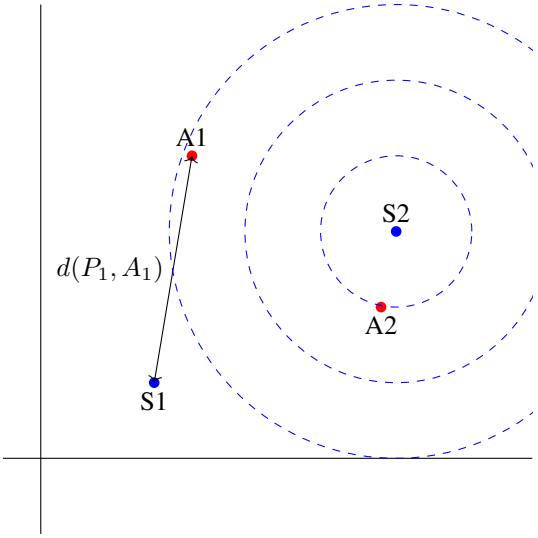

図 8.5　多次元展開法による態度対象と個人の同時プロット。個人 S2 は態度対象 A1 より A2 のほうに近い

8.3.3　非対称多次元尺度構成法

MDS の応用例として最後にご紹介するのは，非対称多次元尺度構成法です。距離とは対称性，すなわち $d(x, y) = x(y, x)$ という性質を満たすものですが，この条件を取り払ったものを考えます。

たとえば親近性のように，A さんが B さんのことを好きだと思っていても，B さんは A さんのことが嫌いである，といった非対称性は人間関係ではザラにあることです。このほかにも，日本から米国への輸出量と，米国から日本への輸出量 (日本の米国からの輸入量) が異なるとか，地方から東京への人口移動とその逆といったことも，対象間の非対称性の例です。あるいはビールから発泡酒に変える人はいるけど，逆の人は少ない，といった購買行動におけるブランドスイッチングなども非対称関係が見られます。こうした非対称な情報が測定の誤差でないとするならば，何らかの意味が読み取れるように分析しなければなりません。それを考えるのが非対称多次元尺度構成法です。

地図に加筆するモデル

MDS は他の多変量解析と違って，非対称関係を考えることができるだけでなく[*8]，地図を描いているわけですからその地図に情報を書き足すなど，表現上の工夫で対応することもできます。

非対称な関係を地図上に表された対象に書き加えるモデルは，基本的に非対称データを対称部と歪対称部に分割します。すなわち非対称行列 $\boldsymbol{S}$ に対して，次のような変換をします。

$$\boldsymbol{S} = s_{jk} = \frac{1}{2}(\boldsymbol{S} + \boldsymbol{S}') + \frac{1}{2}(\boldsymbol{S} - \boldsymbol{S}') = \boldsymbol{S}_s + \boldsymbol{S}_{sk}$$

これは要するに平均をとったところと，そこからのずれのところに分割するようなもので，数値例を見るとすぐ

[*8] たとえば分散共分散行列や相関行列は，その行列計算の手続きの中でどうしても正方対称行列になってしまいますから，非対称関係を表現することができません。

にお分かりいただけると思います。

$$\begin{pmatrix} 0 & 4 & 1 & 3 \\ 1 & 5 & 1 & 2 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 7 & 5 \end{pmatrix} = \begin{pmatrix} 0.0 & 2.5 & 1.5 & 3.0 \\ 2.5 & 5.0 & 1.5 & 2.5 \\ 1.5 & 1.5 & 5.0 & 5.5 \\ 3.0 & 2.5 & 5.5 & 5.0 \end{pmatrix} + \begin{pmatrix} 0.0 & +1.5 & -0.5 & 0.0 \\ -1.5 & 0.0 & -0.5 & -0.5 \\ +0.5 & +0.5 & 0.0 & -1.5 \\ 0.0 & +0.5 & +1.5 & 0.0 \end{pmatrix}$$

このように考えると，前半の対称部 $\boldsymbol{S}_s$ については一般的な MDS を施すことができます。これで布置が得られたら，その周りに非対称情報を書き加えれば良いのです。たとえば Okada-Imaizumi モデル (Okada & Imaizumi, 1987, 1997) では，歪対称な情報を円または楕円で表現します。2 点間の距離 r_{jk} は，対称な距離 d_{jk} に加えて，点 j,k がまとう楕円の半径 r_j, r_k を使って，

$$r_{jk} = d_{jk} - r_j + r_k$$

とします。両者の点間距離 d_{jk} から出発点の半径を引く，つまり円の外縁からスタートし，到達点の半径を足す，つまり円の外縁までを距離として表現するのです。各点がまとう円のサイズは違いますから，$j \to k$ と

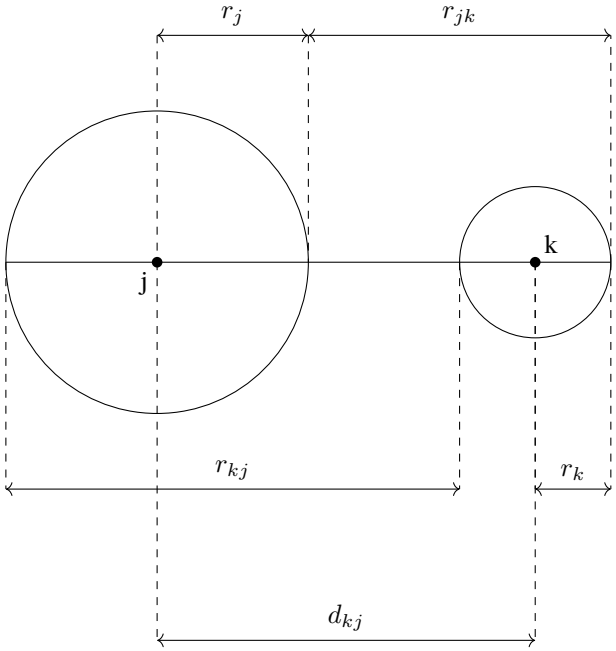

図 8.6 Okada-Imaizumi モデルによる非対称関係の表現

$k \to j$ の距離がその半径分プラスマイナスされることで非対称な関係として表すことができます。また図 8.6 は円を書き加えましたが，次元ごとに重みが違う楕円で表現することもできます[*9]。

円や楕円を纏うだけでなく，全体的に非対称性を持っている方向に矢印や確率分布を纏わせるモデル[*10]，空間に風向きや勾配などで非対称関係を表現するモデルなど，色々なものが提案されています。

これらはいずれも，地図の上に別の情報を追加するというアイデアからきています。MDS で描画されたものが地図ですから，その上に別の情報を載せることでほかにもさまざまな表現ができるかもしれませんね。

[*9] 藤澤 (1997) の付録にはソシオメトリックデータにこのモデルを応用した例が掲載されています。

[*10] 荘島 (2011) は，空間統計学で用いられるフォン・ミーゼス分布を点の周りに纏わせる，非対称フォン・ミーゼス尺度法を提案しています。

数学的に展開したモデル

千野 et al. (2012) は非対称 MDS を網羅的にまとめていますが，彼の提案する**エルミート形式モデル (Hermitian Form Model)** はさまざまな非対称 MDS モデルの中でも，数学的な拡張でもって非対称情報の描画を考えた野心的なモデルであると言えます。

非対称行列 $\boldsymbol{S}$ を対称部 $\boldsymbol{S}_s$ と歪対称部 $\boldsymbol{S}_{sk}$ に分解するところまでは同じなのですが，歪対称部を虚数で表す，つまり

$$\boldsymbol{H} = \boldsymbol{S} + i\boldsymbol{S}_{sk}$$

とすることで，非対称行列を複素数で表現することを考えます。複素数の世界[*11]において，虚部の符号が違うだけのもの，すなわち $a + bi$ に対する $a - bi$ は特に共役複素数と呼びます。非対角要素に共役複素数が入っている行列は，複素数の世界からみれば対称行列のような性質を持っており，特に**エルミート行列**と呼ばれてその性質が研究されています。

非対称行列は固有値分解すると，固有値も固有ベクトルも複素数で得られるのですが，エルミート行列になっていると固有値は実数になります。固有ベクトルは複素数になりますが，計量 MDS の時にみたように固有ベクトルが複素空間における座標として，また要素同士の内積が両者の距離として考えることができます。そもそも対称行列を実数の空間にプロットすることが難しいのが非対称行列を扱う上での問題だったわけですが，プロットする空間の方を複素空間に拡張することで，数学的に自然な形で MDS を展開することができるのです。小杉, 藤澤, et al. (2004) ではこのモデルを社会心理学におけるバランス理論に応用しています。

HFM は分析の元になるデータが計量的，すなわち間隔尺度水準以上であることが求められます。心理学の応用的側面としては，比率尺度水準以上のデータを入手することが困難かもしれませんが，複雑な人間関係などを線形関係や実数の中だけで理解するのはそもそも難しいかもしれません。複素数を扱った心理学理論など，考えてみるのも面白いかもしれませんね。

[*11] 念のために書いておきますが，複素数とは $i = \sqrt{-1}$ となるような虚数単位を導入し，実部と虚部の和，すなわち $a + bi$ のかたちで一つの数字を表現する方法です。

第 9 章

名義尺度水準の尺度モデル

多次元尺度構成法は距離行列から始め，順序尺度水準しか持たない情報であっても可視化など次元解析をすることが可能な技術でした。では順序尺度よりも低い水準である，名義尺度水準の次元を見るような分析方法はあるのでしょうか。ここではそれをみていくことにしましょう。

9.1 カテゴリに数値を与える

名義尺度水準は，数字の付け方が名義的です。たとえば男性に 1，女性に 2 という数字を割り振って集計する，というような数値の与え方が名義尺度水準です。このとき，男性に 2，女性に 1 であっても構いませんし，男性に 33465，女性に 8768432 を割り当てても問題ありません。ここでの数字は数字としての意味があるのではなく，<33465>という文字列が男性というカテゴリに対応している，という一対一対応だけがわかっていればそれで良いのです。数字にラベルとしての意味しかありませんから，これを使って計算することはありません。

間隔尺度水準の量的なデータであれば，線形モデルが色々開発されて利用されてきました。因子分析も回帰分析も主成分分析も，あるいはそれらを統合した構造方程式モデリングも，全て線形モデルの世界です。分散共分散行列を標準化した行列は相関行列といいますが，相関係数が 1.0(あるいは -1.0) の状態を考えれば明らかなように，分散共分散行列 (や相関行列) で表現されるのは変数の直線的な関係性に限った話なのです。

しかし心理学の場合は中庸が良いようなシーンも少なくありません。たとえば血圧と健康度の関係でいえば，低血圧でも高血圧でも不健康であり，ちょうどいい血圧が一番健康的だというのはすぐにわかります。このように中庸が良い場合は，散布図が U 字型に現れてきます。散布図が U 字型であれば，相関係数としては 0 近くになりますが，血圧と健康が無関係だとは誰もいえないでしょう。

これについて，ごくシンプルかつ具体的な例をあげてみましょう。血圧と頭痛の頻度について調査したとします。血圧は高い，普通，低いの 3 段階。頭痛の頻度は「ない」，「たまに」，「ときどき」，「いつも」の 4 段階です。表 9.1 のような結果を見ると，この集計表に直線的な関係は確かになさそうですね。でも関係ないわけではない，と思いませんか。

このような場合は，「血圧が高い人か低い人は，頭痛がある」という傾向が見て取れるはずです。しかしこのデータ，機械的に血圧が高いを 1，普通を 2，低いを 3 とし，同様に頭痛もない〜いつもを 1,2,3,4 として相関係数を計算すると，$r = 0.165$ になります。相関関係では傾向をうまく読み取れていないのです。

その根本的な原因は，もちろんスコアリングの方法にあります。共分散を計算するには間隔尺度水準以上の情報が必要で，今回のような 3,4 件法では相関係数を計算して良い数字ではありません。ではポリコリック相関係数のように順序尺度水準の相関係数なら良いか，といいたいところですが，それでもまだ不十分で

表 9.1 血圧と頭痛

血圧	ない	たまに	ときどき	いつも
高い	0	0	3	2
普通	5	0	0	0
低い	0	2	3	1

す。なぜなら，血圧が高い方から低い方まで，順番が保存されてしまっているからです。

関係を導き出すには抜本的な改革が必要です。たとえば表 9.2 のようにすればどうでしょうか。こうすると左下から右上にかけての直線的な関係がまだ見えてきます。これは血圧の順番を「高い」「低い」「普通」に並べ替えたものです。また，頭痛の順番も「いつも」と「ときどき」をひっくり返しました。順番を並び替えてしまっ

表 9.2 並び替えられた「血圧と頭痛」

血圧	ない	たまに	いつも	ときどき
高い	0	0	2	3
低い	0	2	1	3
普通	5	0	0	0

たというのは，尺度水準的には数字を無視して「高い」「低い」「普通」というカテゴリーとして扱ったということになります。つまり**名義尺度水準**レベルにまで落として考えたのです。

このように並べ替えて，血圧のスコアを高い $\to 3$，低い $\to 2$，普通 $\to 1$ とし，また頭痛の頻度をいつも $\to 3$，ときどき $\to 4$ と付け替えて相関係数を計算すると，今度は $r = 0.810$ になりました。こちらの方が線形性は高いことが数字でも確認できました。

ここで行ったのは，すべての反応カテゴリをただの言葉だと考えて，付与された数字を無視して並べ替えたことになります。そうすることで，左下から右上にかけて線形性を高めることができました。しかし「高い」と「低い」，「低い」と「普通」の配置も等間隔である必要はなく，もっと直線性がはっきりするように配置してやってもよいのでは，というアイデアが浮かびます。

ここで考え方が反転していることに注意してください。一般的なリッカート尺度では，反応カテゴリに (シグマ法などで) 数字をつけて，変数間の線形性を算出して意味を考えるのでした。ここでは逆に，変数間の線形性を最大にするように反応カテゴリに数字をつけてやろう，という考え方です。データの直線性というのはデータの特徴を最も強調し解釈しやすい形です。そのようにデータを整えるためには，反応カテゴリにどういう数字を付与すれば良いか，と考えるのです。

分析者がデータを解釈しやすくする目的で，カテゴリに数値を与えること。これが名義尺度水準の分析をするときの発想であり，林の「数量化理論」と呼ばれる一連の分析手法群です。少し寄り道になりますが，林の数量化理論についてみておきましょう。

9.1.1 林の数量化理論

数量化の研究をしたのは，日本の偉大な統計学者，林知己夫 (はやし ちきお)[*1]という人です。その名を冠して林の数量化理論と呼ばれることがあります。

林はさまざまな研究成果を挙げており，論文ごとにデータに必要な分析方法を開発するというようなスタイルでした。その膨大な研究業績を，弟子である飽戸弘[*2]が分類して，同じような分析方法ごとに I 類，II 類，III 類，IV 類，と呼んでいきました。

表 9.3 林の数量化

手法	外的基準	データ	目的	関連する手法
数量化 I 類	量的変数	質的変数	外的基準の予測	重回帰分析
数量化 II 類	質的変数	質的変数	外的基準の判別	判別分析
数量化 III 類	なし	質的変数	変数間の関係の要約と記述	主成分分析・正準相関分析
数量化 IV 類	なし	対象間の類似度	対象間の関係の要約と記述	多次元尺度法

表 9.3 にあるように，基本的には質的変数，すなわち名義尺度水準や**順序尺度水準**のデータが得られたときに，それを解釈するためにはどのような数値を割り振ってやれば良いか，という発想から生まれたものになっています。

さきほどの表 9.1 のようなデータの並べ替えについては，数量化でいうところの III 類に該当します。外的基準を持たずに，変数間関係の要約と記述をするのですから，因子分析や主成分分析の名義尺度版だと考えてもいいかもしれません。ポイントは，データセットがカテゴリカル変数なので，数字の直線性を考えるのではなく，カテゴリの並べ替えも認めた上で尺度化するという，「分析者が主体的に数字を割り振る」という考え方が前面に打ち出されているところです。

カテゴリカル変数を対象にしたモデル群なので，分析の応用範囲は多岐にわたります。どのような変数でも名義尺度水準に落とすことができるからです。表 9.1.1 には一般的なデータセットの形を示しました。ID があって，Q1,Q2... と項目が列方向に並ぶ形です。一行が一人の反応を表しています。Q1 がたとえば性別などの名義尺度水準の変数であっても，男性 $\rightarrow$ 1，女性 $\rightarrow$ 2 のようにコード化するルールを決めて入力します。Q2 はたとえばリッカート法で当てはまる，やや当てはまる・・・などの 5 件法だったとしましょう。その場合はシグマ法によるスコアを入れる，あるいは簡便的に当てはまるを 5，やや当てはまるを 4，といったように数字を割り振って入力しますね。

これをカテゴリカル変数と見なしてデータセットにした例が表 9.1.1 です。ID は分析対象ではありませんから横に置くとして，Q1 が 2 列に増えています。ID=1 の人は男性なのですが，この人は Q1:Male=1 かつ Q1:Female=0 というようにコード化されています。同様に，Q2 には「どちらとも言えない」を選択しているのですが，これを 3 とするのではなく 00100 と 5 列にわたってコード化しているのです。このようにすることで，すべての変数をカテゴリーとして扱うことができるようになります。

また表 9.1 を並べ替えた時のように，こうした名義尺度のデータの線形性が最大になるように，行だけでな

[*1] 林 知己夫 (はやし ちきお，1918 年 6 月 7 日 - 2002 年 8 月 6 日) は，日本の統計学者。正四位勲二等，理学博士。統計数理研究所第 7 代所長。社会調査・世論調査におけるサンプリング方法の確立を始め，数量化理論 (Hayashi's Quantification Methods) の開発とその応用で知られる。1990 年代以降，データの科学 (Data Science) を提唱し，その研究・思想は現在へと引き継がれている。Wikipedia より。

[*2] 飽戸 弘 (あくと ひろし，1935 年 3 月 14 日 -) は，日本の社会学者，東京大学名誉教授。専門は，社会心理学，コミュニケーション論。Wikipedia より。

表 9.4 一般的なデータセット

ID	Q1	Q2	···
1	1	3	···
2	2	4	···
⋮	⋮	⋮	

表 9.5 カテゴリカル化したデータセット

ID	Q1:M	Q1:F	Q2:5	Q2:4	Q2:3	···
1	1	0	0	0	1	···
2	0	1	0	1	0	···
⋮	⋮	⋮				

く，列も並べ替えます。データの中で線形性が最大になるように，反応カテゴリと回答者に数字を割り振るのです。ちなみに表 9.1 は集計されたデータであり，今回の表 9.1.1 は集計前のデータです。カテゴリカルな変数の分析の場合は，どちらでも良いのです。行と列の関係が表されている数字であれば，「ある/ない」のような二値反応でも，集計された度数でも構いません。さらにいえば，行と列の関係が記述されていればなんでもいいのです。

こうしてカテゴリカルな変数として表されたデータセットに対し，並べ替えも許して適切な数字を割り振るアルゴリズムのひとつに，交互平均法と呼ばれるものがあります。

9.1.2 交互平均法

具体的に，どのようなアルゴリズムで行・列に対する重みをつけるか見てみましょう。例えば，以下のような分割表が得られたとします (表 9.6)。ここで各カテゴリに値を割り振る手続きは，次のような手順で行います。

表 9.6 データ例

血圧	頭痛なし	たまに頭痛	頭痛あり	計
低い	9	2	3	14
普通	3	1	4	8
高い	6	2	5	13
計	18	5	12	35

1. 行方向の「なし (X_1)」を 1 点，「たまに (X_2)」を 0 点，「あり (X_3)」を-1 点と仮におく。
2. 列方向の「低い (Y_1)」，「普通 (Y_2)」，「高い (Y_3)」を算出する。すなわち，

$$Y_1 = \frac{1 \times 9 + 0 \times 2 + (-1) \times 3}{14} = 0.4285$$

$$Y_2 = \frac{1 \times 3 + 0 \times 1 + (-1) \times 4}{8} = -0.125$$

$$Y_3 = \frac{1 \times 6 + 0 \times 2 + (-1) \times 5}{13} = 0.0769$$

3. Y の平均値を 0 にする。すなわち，

$$\bar{Y} = \frac{14 \times Y_1 + 8 \times Y_2 + 13 \times Y_3}{35} = 0.1714$$

$$Y_1 = Y_1 - \bar{Y} = 0.4285 - 0.1714 = 0.2571$$

$$Y_2 = Y_2 - \bar{Y} = -0.125 - 0.1714 = -0.2964$$

$$Y_3 = Y_3 - \bar{Y} = 0.0769 - 0.1714 = -0.0945$$

4. Y の大きさを整える。すなわち, 絶対値最大の要素で割って, 全体が 1 を超えないようにする。

$$Y_1 = Y_1/maxY = 0.2571/|-0.2964| = 0.8674$$

$$Y_2 = Y_2/maxY = -0.2964/|-0.2964| = -1.000$$

$$Y_3 = Y_3/maxY = -0.0945/|-0.2964| = -0.3188$$

5. これらの重みを使って, 今度は行の重みを算出する。

$$X_1 = \frac{Y_1 \times 9 + Y_2 \times 3 + Y_3 \times 6}{18} = 0.1608$$

$$X_2 = \frac{Y_1 \times 2 + Y_2 \times 1 + Y_3 \times 2}{5} = 0.0194$$

$$X_3 = \frac{Y_1 \times 3 + Y_2 \times 4 + Y_3 \times 5}{12} = -0.2493$$

6. X の平均値を 0 にする。

$$\bar{X} = \frac{18 \times X_1 + 5 \times X_2 + 12 \times X_3}{35} = -3.806e-17$$

$$X_1 = X_1 - \bar{X} = 0.1608$$

$$X_2 = X_2 - \bar{X} = -0.0194$$

$$X_3 = X_3 - \bar{X} = -0.2493$$

7. X の大きさを整える。

$$X_1 = X_1/maxX = 0.6450$$

$$X_2 = X_2/maxX = 0.0780$$

$$X_3 = X_3/maxX = -1.000$$

8. 2〜7 を収束するまで繰り返す。収束判定は,

$$\delta_Y = \sum Y_i^{t+1} - Y_i^t$$

$$\delta_X = \sum X_i^{t+1} - X_i^t$$

として, $\delta_X < \epsilon$ 且つ $\delta_Y < \epsilon$

このように, 適当な数字から始めても交互に平均 (項目の中心) を合わせていくことによって, 最終的に次のような最適な重みが得られます。

$$X = \begin{pmatrix} 0.644 \\ 0.081 \\ -1.00 \end{pmatrix}, Y = \begin{pmatrix} 0.851 \\ -1.00 \\ -0.301 \end{pmatrix}$$

これが交互平均法による重み付けアルゴリズムです。

この方法をコツコツとあらゆるデータに応用していくこともちろんできますが, 数学的にもう少し洗練された方法を使うことが一般的です。すなわち, 行列演算を用いるアプローチです。

これまでの回帰分析, 因子分析から SEM に至るまでの流れは, 変数間関係だけを考えてきました。N 行 M 列のデータ行列の計算の途中で, 行に関する情報は平均化して潰されてしまい, 最終的には $M \times M$ サイズの正方行列だけを扱うことになったのでした。そして M 個の変数に**因子負荷量**などの重みをつけて考察

してきました。これを数学的な面から説明すると，正方行列の場合は**固有値分解**をして，**固有値**と**固有ベクトル**を求め，固有ベクトルが変数の重みになるのでした。固有値分解とは

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$

の関係を見つけることでした。一般に $n \times n$ の行列からは n 個の固有値・固有ベクトルのセットが得られ，実対称行列であれば固有ベクトルからなる行列と，固有値を対角に持つ行列を使って，

$$\boldsymbol{A} = \boldsymbol{X\Lambda X}'$$

と書くことができます。ここで固有ベクトルからなる行列 $\boldsymbol{X}$ を縦・横に並べて掛け合わせたものが $\boldsymbol{A}$ になるのは，$\boldsymbol{A}$ が縦・横同じサイズの正方行列だったからです。

しかし今回のカテゴリカルなデータは，$N \times M$ 行の矩形行列になります。矩形行列の場合は縦と横の長さが違いますから，同じベクトルを縦・横にしてかけたところでサイズが整いません。そこで

$$\boldsymbol{B} = \boldsymbol{YMX}'$$

のように分解します。ここで $\boldsymbol{B}$ は $N \times M$ 行列とすると，$\boldsymbol{Y}$ は N 行，$\boldsymbol{X}$ は M 列の行列です。N 行 M 列のデータセットですから，行のベクトルと列のベクトルそれぞれが計算されることになります。また $\boldsymbol{M}$ は対角に ρ_i を含んだ正方行列であり，この値は**特異値 (Singular balue)** とよばれます。矩形行列版の固有値のようなものですね。

このように分解する方法は**特異値分解 (Singular value decomposition)** と呼ばれ，行および列に対応する**特異ベクトル**をその重み，座標と考えることになります。ちなみに固有値分解は特異値分解の特殊な例 (正方行列の場合) だと考えることができます。

9.2 双対尺度法

さて，この数量化 Ⅲ 類はおもしろいもので，同時期に独立にこの手法がフランス，カナダでも発展しており，2 つの別名を持っています。フランス学派が開発した手法は**対応分析 (correspondence analysis)** といい，カナダ在住の日本人，西里静彦[*3]が開発した手法は**双対尺度法 (Dual Scaling)** といいます。

では数量化 Ⅲ 類とコレスポンデンス分析，および双対尺度法とよばれる一群の手法が，それぞれどのように異なっているのかについてみていきましょう。

9.2.1 数量化 Ⅲ 類とコレスポンデンス分析

数量化 Ⅲ 類とコレスポンデンス分析は，分析のもとになる分割表の考え方の違いに求めることができます。例えば，「はい」と「いいえ」で回答が得られる項目が 3 つあったとしましょう。このとき，数量化 Ⅲ 類が分析対象にする行列は，「はい」を 1，「いいえ」を 0 とした分割表であり，表 9.7 のような表し方になります。

これに対して，コレスポンデンス分析と双対尺度法が分析対象とする行列は，表 9.8 のように考えます。

情報量としては同じですが，Yes でなければ No であるとするか，Yes がある/No があるという形で示すかの違いでしかありません。この分析するもととなる行列の違いがあるだけで，それを特異値分解して行・列の両方に重みを与える点では同じ分析手法であるといえます。

[*3] 西里 静彦 (にしさと しずひこ，1935 年 6 月 9 日 -) は，カナダの行動計量学者 (計量心理学)。学位は Ph.D.(ノースカロライナ大学・ 1966 年)。トロント大学名誉教授，アメリカ統計学会フェロー，日本行動計量学会名誉会員。北海道十勝郡浦幌町出身 (札幌市生まれ)。Wikipedia より。

表 9.7 数量化が分析する行列

被験者	項目 1	項目 2	項目 3	合計
S1	1	0	1	2
S2	0	0	1	1
S3	1	1	0	2
S4	1	0	1	2
S5	1	1	1	3
	4	2	4	

表 9.8 コレスポンデンス分析, および双対尺度法が分析する行列

被験者	項目 1(Yes)	項目 1(No)	項目 2(Yes)	項目 2(No)	項目 3(Yes)	項目 3(No)	合計
S1	1	0	0	1	1	0	3
S2	0	1	0	1	1	0	3
S3	1	0	1	0	0	1	3
S4	1	0	0	1	1	0	3
S5	1	0	1	0	1	0	3
	4	1	2	3	4	1	

9.2.2 コレスポンデンス分析と双対尺度法

コレスポンデンス分析と双対尺度法の違いは, もう少し複雑な説明になります。これらの違いについて考える前に, **双対性**について説明しなければなりません。

まず, 双対尺度法は名義尺度水準同士の変数間関係を分解するものですから, 表 9.6 のような分割表を分解することになります。分割表で表される行と列がどれぐらい関係しあっているか, あるいは独立であるかを見るための指標として, χ^2 値があります。双対尺度法は, この χ^2 値を次のように分解していくモデルでもあります。

$$\chi^2 = \chi_1^2 + \chi_2^2 + \chi_3^2 + \dots \tag{9.1}$$

主成分分析が固有値分解したように, 双対尺度法は特異値分解を使って分解していくのですが, これを要素レベルで見ていくと次のようになります。まず分割表の i 行 j 列目の度数を f_{ij} とし, 行 i の総和 (周辺度数) を $f_{i.}$, 列の総和を $f_{.j}$ と表すことにします。また, 総度数を f_t と書くとしますと, 各行・各列が周辺度数の比率通りの独立な関係であるならば,

$$f_{ij}^* = \frac{f_{i.} f_{.j}}{f_t}$$

となります。これは各セルの期待値を計算していることになります。χ^2 値は, これを各セルについて計算して, 期待値と観測値のずれの総和, すなわち各セルの分散の大きさのようなものとして表していることになります。

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}$$

双対尺度法は得位置分解を使って, 行と列に最適な重みを計算するのですが, 要素レベルで見ると次のよう

な分解になります。

$$f_{ij} = \underbrace{\underbrace{\underbrace{\frac{f_{i.}f_{.j}}{f_t}}_{\text{第0次近似}}(1 + \rho_1 y_{1i}x_{1j}}_{\text{第1次近似}} + \rho_2 y_{2i}x_{2j} + \ldots + \rho_k y_{ki}x_{kj})}_{\text{第2次近似}}$$

第 0 次近似は期待値の成分であり，これは行と列が独立な場合の大きさですから分析に際しては意味がなく，無意味解といわれたりします。分析をして意味があるのは次の第 1 次近似からであり，このとき y_{1i} は y 列の第 1 ウェイト，x_{1j} は j 行の第 1 ウェイトといいます。また ρ を特異値といい，この二乗 ρ^2 が固有値とよばれます。特異値や固有値はその次元の重要度を表すというのは，因子分析の時と同じですね。

特にどの項までで近似するかを示すのに，第 n 次近似という表現を用いるあたりは因子分析とも同じです。またデータが $m \times n$ の場合，第 $k = min(m, n) - 1$ 次近似でデータを完全に説明することができます。

さて双対の関係とは，行 Y と列 X の重みとして分解した式 (9.2) にみられる，次の関係です。

$$\begin{cases} Y_i &= \dfrac{1}{\rho}\dfrac{\sum f_{ij}X_j}{f_{i.}} \\ X_j &= \dfrac{1}{\rho}\dfrac{\sum f_{ij}Y_i}{f_{.j}} \end{cases} \tag{9.2}$$

この双対の関係が意味するのは，Y のベクトルが張る空間と X のベクトルが張る空間は等しくなく，一方が他方に射影されるためには特異値をもちいて変換しなければならない，ということです。たとえばある回答者 Y1 が，X1,X2,X3 という3つの項目を選択した場合，X 空間に射影される Y の重みづけられた座標は，X が作る三角形の重心に来ることになります。

双対尺度法をつかうと，行の空間を見出すことも，列の空間を見出すこともできますが，その両者は別の座標を表現しています。行変数と列変数を同時に表現できれば両者の関係がわかって良いのですが，行空間のなかに列変数をプロットするとか，列空間のなかに行変数をプロットすることを考えるときは，一方から他方への座標を変換しなければならないのです。双対尺度法の文脈では，行空間にプロットするのか列空間にプロットするのかを選択し，空間内での距離関係に注意して考えることが基本です。

ところがコレスポンデンス分析は，これをひとつの空間に射影します。ρY_i と ρX_j は，大きさ (ノルム) を整えてあるので分散が同じですから，ρY_i と ρ_j をひとつの空間に射影しようじゃないか，というのがコレスポンデンス分析のやり方なのです。厳密には数学的に正しくないやり方かもしれませんが，両者共通の空間に同時にプロットできますから，行変数と列変数の対応が見やすいという利点があります。あくまでも，同じ空間にプロットできているわけではない点には，注意が必要です。

9.3 双対尺度法の発展

双対尺度法をつかうことで，名義尺度水準のデータを次元解析できるようになりました。リッカート法でなんとなく 1,2,3,4,5 という数字をつけて，潜在的な次元を頂戴したとありがたがるのではなく，分析者が積極的に得られた反応パターンに数字を与える。その目的は得られたデータを最も説明しやすくする (線形性を最大にする) ことである，というのが「尺度化」「数量化」のいわんとすることです。

双対尺度法はデータや分析方法に工夫をすることで，他にも面白い分析をすることができます。いくつか双対尺度法の発展モデルを見てみましょう。

9.3.1 順序データの場合

例えば好きな政党を順に番号をつけてもらうといった，順序データがあったとしましょう (表 9.9)。この順序データを双対尺度法で分析するには，少し工夫するだけで OK です。

表 9.9 順序データの場合

被験者	自民党	民主党	公明党	自由党	社民党	合計
S1	1	2	4	5	3	15
S2	2	1	3	4	5	15
⋮	⋮	⋮	⋮	⋮	⋮	⋮

双対尺度法で分析するには，この順序関係が入った表を，**ドミナンス表 (dominance table)** に変換してから計算を行うことになります。ドミナンス表とは，R_{ij} を被験者 i の項目 j に対する順位とすると

$$d_{ij} = n + 1 - 2R_{ij} \tag{9.3}$$

で表される数値が入った表のことです。これは，ある項目が他の残りの項目それぞれと比べて順位が上であれば +1，下であれば-1 である数値で，いわば他の項目と何勝何敗であったかを表す数値になっています。双対尺度法では変数同士の相対的な結びつきの強さがロウデータになりますから，このように変換することで相対的な関係の強さに置き換えるわけです。今回の例では，ドミナンス表は以下のようになります (表 9.10)。

表 9.10 ドミナンス表

被験者	自民党	民主党	公明党	自由党	社民党	合計
S1	4	2	-2	-4	0	0
S2	2	4	0	-2	-4	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

あとはこれを特異値分解して，行・列の両方に重みを与えることになります。このようなドミナンス数の考え方は，例えば一対比較法のデータにも応用できます。一対比較法とは，例えば A,B,C の3つのカテゴリがあった場合に，A と B を比較してどちらが優位であったか，A と C は，B と C は，というように一対一での優劣 (似ている，似ていないとか，好ましい，好ましくないなど比較の次元は問いません) を競いあったデータのことです。このようなデータであっても，双対尺度法を使えば変数同士の尺度化，回答者の同時プロットも可能にします。

くどいようですが，尺度構成の基本は一対比較です。何らかの次元で，一方が他方より大か小かを判断する次元があるとするならば，それを取り出そうとする考え方です。得られたデータの中から最大の説明をするように数値を割り振る双対尺度法は，新しい心理尺度の可能性を開いてくれるかもしれません。

9.3.2 強制分類法

少しトリッキーな技術ですが，**強制分類法 (Forced Classification)** という考え方も心理尺度の中では有用かもしれません。

例えば 2 つの項目，F1,F2 があって，それがそれぞれ「はい」・「いいえ」で答えるデータであったとしましょう。データは行列 $[\boldsymbol{F1}, \boldsymbol{F2}]$ のようになりますが，ここで F2 の列を意図的に水増しすることを考えます。例えば F2 を 5 つ並べて，

$$[\boldsymbol{F1}, \boldsymbol{F2}, \boldsymbol{F2}, \boldsymbol{F2}, \boldsymbol{F2}, \boldsymbol{F2}] = [\boldsymbol{F1}, 5\boldsymbol{F2}]$$

のようにすることもできます。これは，項目 2 を 5 回繰り返したもので，データとしては表 9.11 になっていることとおなじです。

表 9.11 繰り返された表

x1	x2	x3	x4
1	0	5	0
1	0	0	5
0	1	0	5
0	1	5	0

一般に，ある項目 j を k 回繰り返したとすると，それは項目 j に重み k を掛け合わせたことと同じになります。では繰り返しを 5 回ではなく，10 回，20 回‥‥とどんどん増やしていくとしましょう。そして増やした行列それぞれに対して双対尺度法を繰り返すとどうなるでしょうか。当然のことながら，今回の例では F2 についての反応，カテゴリでいう x3,x4 の重要度が増していくことになりますから，構成される空間は F2 が第一次元に強く強く関係していきます。繰り返しが ∞ に近づくにつれ，F2 が双対尺度法で得られる第一次元の解に与える寄与率が 1.00 に近くなっていくわけです。これを言い換えれば，F2 の軸を基準にして，他の項目がそれに合うように強制的に分類して行くことでもあります (図 9.1)。実際には無限の重みをつける必要はなく，十数倍の重みづけで十分です

強制分類法は，何か明確な外的基準があるときに，そこを基準にして周りを見てみようという分析方法です。因子分析のように，目に見えない軸を取り出して「この軸は何だろう」と悩むのではなく，分析者の分析視点を明確にして「この観点から考えれば，対象はこのようにプロットされます」と論じる方が責任の所在がはっきりして良いかもしれません。応用例として，学術的な分類法や法的な基準があるような領域で，専門家と非専門家の評定がどの程度ずれているかといったことを考えるときに，専門家の判断パターンを基準に重みづけをし，非専門家がどのようなずれを生じさせるのかを考えるといった可能性が考えられます。

9.3.3 全情報分析

式 9.2 に挙げた双対の関係にあるように，列の最適な重みづけによる平均値は行の重みに特異値をかけたもの，行の最適な重みづけによる平均値は列の重みに特異値をかけたもの，になります。つまり双対尺度法によるプロットは，「列空間」に行・列の各カテゴリをプロットするか，「列空間」にそれらをプロットするかのどちらかを選ばなければなりません。行の空間と列の空間は異なる縮尺を持つ空間なのです。

ここで行変数と列変数を同一の空間にプロットすることを考えたときに，行変数と列変数の空間的隔たりがどの程度あるかを次の式で算出することができます。

$$d_{ij} = \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik}^2 + x_{jk}^2 - 2\rho_k y_{ik} x_{jk})}$$

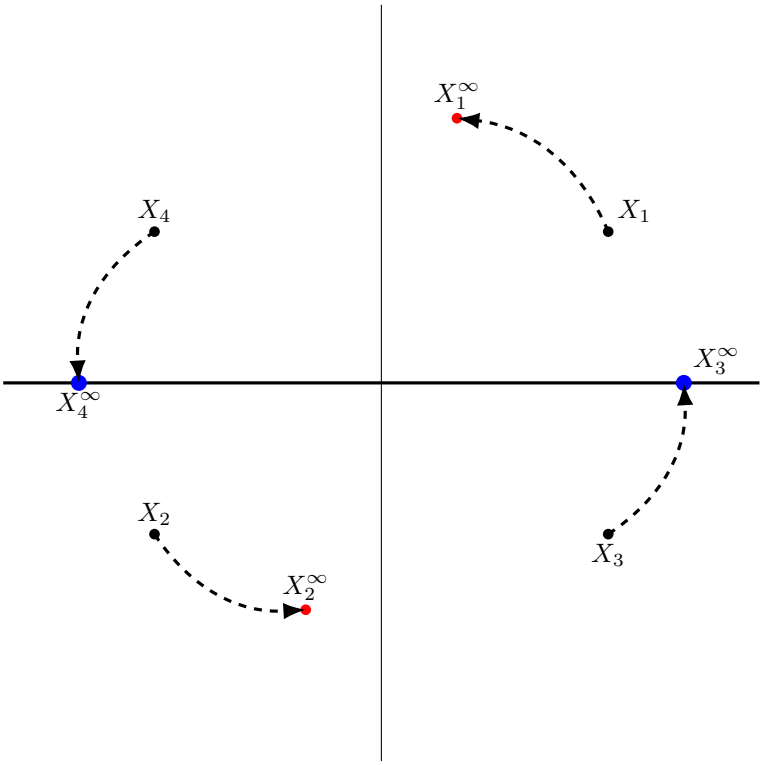

図 9.1 強制分類法による布置

また, 行変数同士 (y_i, y_{i*}) の距離や, 列変数同士の距離 (x_j, x_{j*}) は,

$$d_{ii*} = \sqrt{\sum_{k=1}^{K} \rho_k^2 (y_{ik} - y_{i*k})^2}$$

$$d_{jj*} = \sqrt{\sum_{k=1}^{K} \rho_k^2 (x_{jk} - x_{j*k})^2}$$

で求めることができます。これらを距離行列 $\boldsymbol{D}_{yy} = \{d_{ii*}\}, \boldsymbol{D}_{xx} = \{d_{jj*}\}, \boldsymbol{D}_{xy} = d_{ij}, \boldsymbol{D}_{yx} = d_{ji}$ からなる行列としてまとめれば, 次のようになります。

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{yy} & \boldsymbol{D}_{yx} \\ \boldsymbol{D}_{xy} & \boldsymbol{D}_{xx} \end{bmatrix} \tag{9.4}$$

これは行の空間, 列の空間, 行と列の空間 (と列と行の空間) の全てが含まれた, 大きなデータセットです。一方から他方の関係を考えるだけでなく, 全ての組み合わせを含んだ行列です。この行列を分析することを**全情報分析 (Total Infromation Analysis)** といいます。この行列は距離行列ですから, クラスター分析などで類似性をみつける分類などが考えられます。

双対尺度法は, 行にも列にも, データのあらゆる相にたいして重みをつけて分析することができます。もちろんカテゴリの数だけ, 人の数だけ情報がでてくるので単純な話ではなくなります。各項目の各反応カテゴリに対して, 最適な値が割り振られることになりはするのですが, その数が膨大になってしまうのです。例えば 5 件法で 3 変数分のデータを取って 1 因子が出た, というのであれば話は単純ですが, 双対尺度法は 15 カテゴリに重みがつけられ, しかもそれらが必ずしも直線上に並んでいるわけでもなく, 全体を眺めて意味を考え

なければならないのです。しかも同時に，回答者の数だけそのプロットもされています。先ほどの 3 変数でも，調査データの場合は 100 人，200 人とデータを集めることがあるでしょうが，変数の空間の中に投影された 100〜200 の回答者の点を考えると，もう何が何やらわからない，ということになるかもしれません。

しかし大事なのは，リッカート法で何を測っているか，何が見たかったのか，です。便利だからといって理論的根拠もなく中途半端な方法で，お決まりの分析手法しか使えないようになってしまっては，本質を欠いているのではないでしょうか。双対尺度法は，カテゴリに対して分析者が主体的に「意味づけする」ような数字を与えます。順序データでも一対比較砲のデータであっても構いません。また強制分類法のように判断基準も能動的に与えることもできます。

紙とペンで得られた反応の分析も，様々な可能性が考えられるはずなのです。

第 10 章

確率モデルとベイズ推測

ここまで「測定する」「数値を与える」ということについて考えてきました。心理統計は得られたデータを分析することについてデータを正しく処理することが求められますが，そもそも得られたデータが正しいのか，それでいいのかということについて考えてきたわけです。

これは「まずデータありき」の考え方だった，ともいえます。データが手に入ってから，それにどういう意味づけができるのか，尺度構成の理屈からみればこうなるはず，というような議論でした。データがあって初めて駆動する，という意味で**データドリブン**と呼びます。心理学者にとって，使えるツールが因子分析や回帰分析[*1]であれば，そうしたツールに当てはまるようにデータの取り方を工夫しよう，という考え方になっていったのもやむなしといったところです。心理学は方法論の科学だといわれることもありますが，それは (被験者の) 主観的な経験を扱うときにいかに (研究者の) 主観的な影響を取り除くか，といった工夫を手続きの中に埋め込むことを指しているのです。あるいはまた，最終的に因子の構造や平均値の差に落とし込めるように，刺激や実験状況を整えることこそが重要，と考えられてきたことを指しているともいえます。そういう意味では，統計的手法は添え物にすぎませんから，「このやり方でやっていれば問題ないだろう」といわんばかりに分析方法のパターン化，ルーチンワークになってしまっていたところがあるかも知れません。統計的ツールの背景に測定論としての仮定，理論的背景があることを改めて学ぶこともないままに。

しかしここまでで，測定の原理をしっかり踏まえた上で使うべきツールがある，ということがわかってきたかと思います。分析方法のアラカルトを学ぶよりも，測定の工夫や測定モデルの考え方を詳しく知る方が，心理学の研究としては重要なことのように思いませんか。少なくとも，何をどのように測っているかについて，自分のしていることを自覚しておく必要はあります。この測定のモデルをいろいろ考えて工夫することこそ，心理学の本質に近いと筆者は考えています。因子分析 (第 3 章) は測定の数理モデルです。項目反応理論 (セクション 2.2,Pp.19) は回答者の反応ひとつひとつに対する数理モデルです。このような，測定や反応についてのモデルを直接考えることが，心理学におけるブレイクスルーをもたらしてくれるかもしれません。

ここではこうした測定モデルを考えること，**モデリング (modeling)** について考えていきたいと思います。

10.1　データ生成モデリング

「心理学統計法」と名のつく講義のほとんどは，記述統計や推測統計，とくに帰無仮説検定を扱うことが中心であるのが現状です[*2]。心理学において帰無仮説検定を行う理由は，心理学実験の効果を検証すること，それも手元の標本についての記述ではなく母集団においてもその効果があるといえるかどうかを検証するた

[*1] 分散分析や t 検定など，平均の差の検定は回帰分析のグループに入りますのでここで改めてそれらを取り上げていません。

[*2] とくに公認心理士に必要な授業として，科目名「心理学統計法」を名乗る必要があり，そこで教える基本的な内容としてこれらが含まれています。なお測定論やそれに関する多変量解析法はそれほど中心的話題ではありません。みんな使うのになあ。

めです。手元のデータに基づいて，目に見えない母集団全体のことに推論するというのは，基本的に不良設定問題です。つまり，確実な正解を導き出すにはヒントが少なすぎる，圧倒的に不利な状況で知恵を絞る必要があるのです。こうした不利な状況下ですから，いくつかの仮定をおいて，確率的に推論するのでした。その推論の方法として，**モーメント法 (moment method)**，**最尤法 (Maximum Likelohood estimation)**，**ベイズ法 (Bayesian estimation)** があります。

モーメント法は，標本の平均や分散といった統計量を母数の推定値とする方法です。これは特に**実験計画 (Experimerntal Design)**(要因計画ともいう) と組み合わせで用いられ，帰無仮説検定という意思決定ツールとして使われています。心理学における実験的なアプローチは，実験群と統制群に無作為に割り当てた集団に対し，介入・処置のあとでの群平均を見ることでその効果を検証します。人間を相手にする研究ですから，当然誤差や個人差がデータには含まれますが，無作為割り当てと平均化によってそれらはキャンセルアウトされ，平均の比較をすることで効果を見ることができると考えるのでした。また標本の平均ではなく，それを用いて母平均を推測することで，結果の一般化を考えます。母平均の推定には点推定と区間推定とがあり，確率的表現を用いた区間推定をつかって慎重に結論を導き出すのです。ただし区間推定は判断基準が明確ではありませんから，帰無仮説と対立仮説という 2 つのモデルを戦わせて決着を見る，というやり方をして一応の決着を見るのでした。こうした「実験計画」+「推測統計学」=「帰無仮説検定」は，長らく心理学のスタンダードとして君臨しています。

こうしたアプローチについて，昨今批判があることについては，今は触れないでおきましょう[*3]。それよりもこうした問いの立て方や解決策に注目してみたいと思います。心理学は物理学を範として，科学 science の仲間入りをしたい，というモチベーションが強く根付いている世界であり，また人間というのは嘘をついたり間違えたりするものだ，ということが骨身に染みてわかっていますから[*4]，データを分析する際にも**客観性**を大事にすることが殊更重視されます。客観性の反対は主観性，つまり本人の思い込みや考え方の癖がもっとも邪魔になるのです。そこでデータに対しては真っ白な気持ちで向き合うものだ，という姿勢をとることになります。

言い方を変えると，分析をする前はデータについて何も考えてないよ，という態度をとるわけです。もちろんデータは要因計画の結果として得られるものですから，計画を立てる際は主観的な誤謬が微塵も入り込むことないように緻密に練り上げるのですが，出てきた結果は結果，数字の羅列に過ぎないと考えるのです。その結果を統計的に分析するときは，それが記憶実験の数字であろうが臨床実験の結果であろうが気にすることはなく，淡々と帰無仮説検定の俎上に乗せていくことになります。こうしたアプローチができるからこそ，そしてそれを許す統計ソフトウェアがあるからこそ，心理学の研究を積み重ねていくことができたのだという一面があります。

これに対して，データがどのように生まれてきたのか，そのメカニズムを考え，その仕組みを明らかにしていこうというのが**データ生成モデリング**です。松浦 (2016) は統計モデリングを「確率モデルをデータに当てはめて現象の理解と予測を促す営みのこと」と定義しています。すなわちこのアプローチは，まずデータがどのようなメカニズムで生まれてきたのかを，簡潔な数式を使って表現します。そのモデルをデータに当てはめ，このモデルからデータが出てきたといえるかどうか，他のモデルの方が今のデータをうまく説明するのではないか，といった比較検証をしていくことになります。データドリブン型分析では，こうしたメカニズムが実験計画の中に暗黙理に埋め込まれていたといえるかもしれません。データ生成モデル駆動型の分析は，メカニズムを明示して検証する，というところが違います。

[*3] Amrhein et al. (2019) などが指摘するように，誤った使い方をされることが多く心理学研究の意義そのものが疑われるほどになっている現状があります。また豊田 (2020) ではさらに辛辣に批判がなされています。

[*4] というか心理学の研究というのは，いかに人間がダメで偏った考え方をする生き物であるかを，滔々と明らかにしていくという側面もあります。

データが作られるメカニズムを考えるアプローチの利点は，予測にも向いています。メカニズムが正しい，あるいはうまく現状のデータを再現できるのであれば，おそらく今後も同じようにデータが作られていくでしょう。であれば将来の予測ができる，という考え方です。たとえば市場の動向の予測，消費者の傾向の予測ができればそのメリットは想像に難くありません。昨今はデータサイエンスという言葉が流行していますが，そこにはこうした研究アプローチの隆盛があるのです。もちろん，心理学においてもモデリングのアプローチは有用ですし，従来通りの平均値の比較をする上でも，さまざまなメリットがあります。

これを考える上で重要なのが**確率モデル (stochastic model)** です。言葉の通り，データ生成過程を確率を使って表現します。なぜ確率を？と思うかもしれませんが，手元のデータは誤差や個人差を含んで微小に変わる，偶然の成分が必ず入っているからです。こうした偶然をハンドリングし，偶然の中にも理論的な予測をするという点で確率が必要になるのです。最尤法とベイズ法は，母数の推測に確率モデルを使う方法であり，最尤法は特にデータにぴったり合うような確率のパラメータを探し当てるというものでした。ベイズ法についてはまだあまり一般的に知られていないところもあるかも知れませんが，これも確率のパラメータを確率で表現する推定方法です。モーメント法と最尤法の答えは原理的に一致するシーンが多く，「なにか名称が変わったな」とおもうだけでユーザにとっては対して考えなくても良い，とされてきたところがありますが，ベイズ法については結果も確率分布として得られるなど，少し考え方の修正が必要です。

進んだ確率モデルで考えるときには，最尤法よりもベイズ法の方がよく使われます。確率モデルが複雑になっていくにつれ，最尤法では計算コストが非常に高く，実質的に解が得られないということも少なくないからです。ベイズ法によるアプローチはこの問題をクリアしてくれるからです。

10.2 ベイズ推定の基礎

データ生成モデルが確率の言葉で記述される，というのはすでに述べたところです。推測統計というのがそもそも不良設定問題，すなわち少なすぎるヒントから未知なるものを当てるという状況に置かれた学問であり，この「わからないこと」を確率で表現するからです。この確率についての考え方として，ベイズの公式というのがあるのでした。ベイズの定理は次のように書き表されます。

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

ここで $P(X)$ は X についての確率，という表現です。$P(A|B)$ は条件付き確率というもので，B が与えられた時の A の確率，という意味です。

ベイズの定理の用語は次の通りです。まず右辺の分子に注目しましょう。$P(D|\theta)$ とありますが，ここで D はデータを表しています。θ はデータを生む確率のパラメータです。$P(D|\theta)$ はパラメータ θ のもとでのデータ D が得られる程度を表すもの，という意味になります。これは**尤度 (likelihood)** と呼ばれるもので，パラメータの関数になっているのでした。たとえば正規分布からデータが生成されると考えるのであれば，手元のデータがパラメータ θ から出てくる可能性がどれぐらいあるのか，を表現しているといえるわけです。この尤度関数は，データを生み出すメカニズムの表現そのものです。

右辺の分子の第二項，$P(\theta)$ はパラメータの確率です。正規分布からデータが生成される例でいえば，尤度はあるパラメータの値からデータが得られる程度を表現しているのですが，そのパラメータが出てくる確率はそもそもどの程度であるか，を考えているのです。これは別名**事前分布 (prior distribution)** とも呼ばれます。実際にデータが出てくる前の段階で，そもそもそのパラメータがどれほど出やすいかという確率を表現しており，これは今回のデータを取る前までの事前のデータや経験に基づいている，ということができます。

右辺の分母，$P(D)$ はデータ全体が得られる確率であり，**周辺尤度 (marginal likelihood)** とか**エビデン**

ス (evidence) と呼ばれます。**正規化定数 (normalized constant)** ということもあります。これについては理解しにくいところもあるかもしれませんが, 後に述べる理由によって, ひとまず深く考える必要はありません。

左辺はこの計算の結果得られる**事後分布 (posterior distribution)** を表しています。$P(\theta|D)$ はデータ D で条件づけられたパラメータ θ の確率です。我々は確率モデルを作るわけですが, そのパラメータがどうなっているかは事前にはわかりません。が, データを取ることで「データが与えられたのであれば, パラメータはこうだよ」というのがわかるわけです。あるいは事前にパラメータはこのあたりにあるのではないか, と経験上考えていたとしても, それがデータを取ることによって更新される (確信が強くなったり, 違うかもしれないと思い直したり), ということを意味しています。

この式を言葉で表現し直すと, 次のようになります。

$$\text{事後分布} = \frac{\text{尤度} \times \text{事前分布}}{\text{周辺尤度}} = \text{尤度} \times \frac{\text{事前分布}}{\text{周辺尤度}}$$

尤度がメカニズムそのものだ, ということはすでに書きました。たとえば**回帰分析**においても, 尤度を計算できます。普通の回帰分析は, 誤差が確率的に生じると考え, 誤差以外のところは線形モデルで表現します。すなわち, 従属変数 Y_i に対して, 独立変数 X_i があったとすると,

$$\begin{aligned} Y_i &= \hat{Y}_i + e_i = \beta_0 + \beta_1 X_i + e_i \\ e_i &\sim N(0, \sigma) \end{aligned}$$

より,

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma)$$

と表されます。ここで $N(\mu, \sigma)$ は平均 μ, 標準偏差 σ の正規分布を表し, データ Y_i が正規分布から生成されているというモデルになっています。回帰分析の場合は**最尤法**で未知数 β_0, β_1, σ を求めたりしましたが[*5], その名の通りこのモデルが示す尤度を最大にする未知数の求め方を最尤法というのでした。

ですが, 尤度は確率を表すものではありません。確率とは非負の実数で, すべての確率を足し合わせる (あるいは積分する) と 1.0 にならなければなりませんが, 尤度は足し合わせても 1.0 にならない数字なのです。ですから,「手元のデータがパラメータからでてくる**可能性**」という変な言い回しをしていました。「出てくる**確率**」とはいえないのですね。そこで, 尤度を確率に置き換えたい, と考えたときに便利なのがベイズの定理なのでした。ベイズの定理は尤度に事前分布をかけ, 周辺尤度で割るという操作によって, 事後分布が得られる式, と読むこともできます。事後分布は**確率分布**です。尤度にある数字をかけて (事後の) 確率分布にしていると考えるといいでしょう。ちなみに事前分布も確率分布ですが, **周辺尤度**は確率ではありません。たとえば度数を総度数で割った相対頻度は確率と考えることができますが, それと同じように, 分子をとある数字で割って, 全体を 1.0 に整えるための定数なのです。**正規化定数**というのはそういう意味ですね。定数倍 (正確には定数の逆数) をかけて大きさの調整をしているだけですので, ベイズの式はさらに次のように書き直すことができます。

$$\text{事後分布} \propto \text{尤度} \times \text{事前分布}$$

ここで $\propto$ は比例する, という関係を表しています。最終的には事後分布の形が知りたいのですが, その形を決めているのは分子の尤度と事前分布だけだ, ということになります。

[*5] あるいは**最小二乗法 (Least Square method)** で求めるのですが, これは幸い最尤法の結果と一致します。正規分布を使った線形回帰モデルの場合, データの記述統計的値と確率モデルによる推定値が一致するので, 気づかないまま推測統計学に足を踏み入れてしまえるのでした

ここで事前分布が**一様分布 (uniform distribution)** であれば，事後分布の形状には影響せず，事後分布は尤度そのものの形を反映することになります。従来のデータドリブンな分析では，データに対して事前の仮定を一切置かないということでしたが，ベイズ流の分析でもそのことは同様に表現できるわけです。

以上がベイズの定理についての簡単な説明でした。しかしベイズの定理自体は，1740 年代には明らかになっていたことであり，それがなぜ 21 世紀の今になって見直されてきたのでしょうか。それには色々な理由がありますが，その 1 つは最近まで「ベイズ流の分析は絵に描いた餅」だったからです。すなわち理屈ではこのような形になることは明らかだったのですが，実際に計算するのは難しいことが多々ありました。その問題を解決したのが**マルコフ連鎖モンテカルロ法 (Markov Chain Monte-Carlo method; MCMC)** と呼ばれる方法です。

10.3 マルコフ連鎖モンテカルロ法

ベイズの定理から，事前分布と尤度が分かれば，計算式を解いて事後分布の形を算出できます。しかし**確率分布**の式が複雑になればなるほど，その計算はとても難しくなります。確率関数を複数組み合わせたり，求めるべきパラメータの数が増えて行ったりすると，とてもじゃないけど計算して答えを出すことができない，ということになります。そのせいもあって，ベイズ統計学は長らく実践には向かない手法だったのですが，**マルコフ連鎖モンテカルロ法 (Markov Chain Monte-Carlo method)**，略して **MCMC** が出てきてからその状況が一変しました。

MCMC 法は近似的な答えを探す 1 つの方法です。MCMC という名前は「マルコフ連鎖」と「モンテカルロ法」の 2 つのパートからできあがっています。マルコフ連鎖は，目標となる確率分布状態になるような推移規則を作る方法であり，モンテカルロ法は乱数を発生させるアルゴリズムです。この 2 つが合体することで，確率分布の形がわからなくてもサンプルが得られるような，乱数発生器を作ることができるようになったのです。

ベイズ統計のゴールは事後分布を作ることです。事前分布を一様分布にして，尤度はベルヌーイ分布にする，といった簡単な場合ですととくに問題ないのですが，ある確率分布のパラメータがあって，さらにそれを生成する確率分布があるといった，階層的に入れ子になっているようなモデルが出てくると，「最終的な事後分布の形がわからない」という状況になります。正規分布とかポアソン分布といった，名前がついている確率分布であればその特徴がはっきりわかるのですが，それらを組み合わせて作られるものは名もなき合成関数になり，どんな形状をしているのか，まったく想像つかないことがあります。マルコフ連鎖を使うと，その名もなき合成関数の形状をとにかく作り上げることができる，そんな数学的技術です。

モンテカルロ法は乱数発生技術です。乱数というのは規則性のない数ですが，これを作るのはなかなか難しいものです。人間が 0-9 の数字を適当に書いていくと，知らず知らずのうちに均等でないパターンができてしまいます[*6]。コンピュータに (たとえば R に) 乱数があるじゃないか，と思われるかもしれませんが，コンピュータはあくまでも計算機でどこまでも合理的です。乱数によって動いているように見えますが，乱数に見えるような数字を生成するアルゴリズムがその中には埋め込まれていますから，正確には擬似乱数でしかありません。コンピュータの作る乱数は，もちろん人間が適当に思いつく乱数よりもより規則性が少ないですが，それでも基本的に

- 何らかの関数 g によって，内部状態 S_t をアップデートする；$S_{t+1} = g(S_t)$
- 内部状態 S_t から何らかの関数 h によって，実現値が生成される；$x = h(S_t)$

[*6] 嘘の 538(ゴサンパチ) という標語があって，人間がふと思いつきで数字を作ろうとすると 5，3，8 が多くできてしまうといわれています。エビデンス (出典) がわからないので本当かどうかわかりませんが･･･。

というステップを反復 ($t = 1, 2, 3 \ldots$) することで生成するのです。こうした乱数列ができれば，それを正規分布の形だとか二項分布の形に当てはめて出力することは簡単なのです。このステップ・バイ・ステップの計算法としてある確率分布に従う乱数発生技術があり，これがマルコフ過程とひっついてできたのが MCMC 法ということになります[*7]。

MCMC 法は，ですから，どんな形の確率分布であっても乱数のサンプリングは生成できる，という技術なわけです。事後分布の関数の形，形状がわからなくてもそこからの乱数は作れるという技術であり，コンピュータ技術の性能が発展した今では大量の乱数を生成することも瞬時に行われます。

乱数を用いるアプローチはいくつかの利点があります。たとえば**確率分布**の平均である期待値など，分布の特徴を記述するための計算は積分を含むので解析的に解くのは知識も技術も必要です。しかしその確率分布から乱数を発生させると，その平均値を求めることでその近似値を得ることができます。確率分布の計算が記述統計の計算に置き換えられるのであり，また統計ソフトウェアにとって大量のデータの記述統計量を描くのは造作もないことなのです。乱数による近似値は，あくまでも近似値，近くて似ている値に過ぎませんが，精度を上げるためにはその乱数の数を増やしてやるだけで良いことになります。この点もいいですね。

また求めたいパラメータが複数ある場合，たとえば正規分布だと平均と標準偏差が未知数ですし，回帰分析では平均の中に切片と傾きといった未知数が入っているわけですが，このような場合の事後分布は同時確率空間ということになります。すなわち平均と標準偏差という 2 つの変数の場合でも，可視化するなら 3 次元空間が必要です (x 軸に平均，y 軸に標準偏差，z 軸に確率密度)。こうした多次元空間において，あるパラメータだけについて期待値を計算したい，という場合はそれ以外のパラメータについては**周辺化**といって積分して全部の可能性をつぶしてしまう必要があるのですが，この計算も解析的にやるには実に大変なものです。しかし多次元の事後分布空間から生成された乱数があるなら，注目したい変数だけについての記述統計をすれば，他の変数を周辺化したこと同じになります。なんと便利なのでしょう。

このように，乱数を使ったアプローチは計算上非常に有利な特徴を揃えています。乱数を大量に発生させられるような計算機のパワーは，最近の PC でしたら十分持っていますから，最近になってベイズ統計学やモデリングアプローチが生きてきたわけですね。さらにありがたいことに，事後分布から乱数を発生させるためのプログラミング言語ツールが登場したのも大きいでしょう。古典的には BUGS というソフトウェアが，最近では JAGS や Stan といったのがそれで，**確率的プログラミング言語 (stochastic programming language)** と呼ばれたりします。これらの言語では，尤度と事前分布をモデルとして表記し，それにデータを与えてやることで，事後分布からの乱数を生成します。いわば万能乱数発生器なのです。こうした環境が整ったことで，簡単に分析できるようになりました。

10.4 モデリングの実践例

それではモデリングでどのようなことがわかるのか，ベイズ推定による分析結果はどのようなもので，どのような解釈の仕方があるのかを，いくつかの例とともに見ていくことにしましょう。

[*7] 余談ですが，スマホアプリなどで「ガチャを引く」というのも内部では乱数が生成されていて擬似的に「偶然あたりが出た」のを装っているに過ぎません。私はゲームなどをする時，擬似乱数に思いを馳せ「どこかの誰かに遊ばされている」と思うとやる気がなくなるので，ガチャを引くようなシーンは興醒めしてしまいます。やはりマルチプレイヤーゲームのように，人間が背後にいたほうがよっぽど意外な行動が見られますし，ゲームよりもリアルの世界の方が擬似的でない本物の偶然を楽しむことができておもしろいと思います。皆さんはどうお考えですか。

10.4.1 7 人の科学者

ここでは Lee & Wagenmakers (2013) より「7 人の科学者」の例を紹介します。そこでのカバーストーリーは次のようなものです。

> 実験スキルが大きく異なる 7 人の科学者が，全員同じ量について測定を行う。彼らが得た数値結果は次の通り。
>
> $$Y = -27.020, 3.570, 8.191, 9.808, 9.603, 9.945, 10.056$$
>
> 直感的には，最初の二人の科学者はひどく適性を欠いた測定者であり，この量の真の値はおそらく 10 をわずかに下回るくらいであるように思えるのだが・・・？？

さあ，これをみて「あれ，どこが統計的な問題なんだ？」と思った人もいるかもしれません。なんらかの測定をして，データのばらつきがあるんだけど，それが測定者によって違うらしい，というのはわかったかと思いますが，これをどうやって統計的な問いにするのでしょうか。まず，ストーリーの背後に「正確な測定値 (真の値) はわからないけど，定数のはず」という前提があることを確認しましょう。測定の基本として測定誤差があるので真の値を特定出来はしませんが，測定を繰り返すと真の値に近づいていくはずではあります。また，今回の測定結果は 7 人それぞれ別の人による測定であることから，測定誤差の出方が測定毎すなわち測定者毎に異なるだろう，という仮説もあります。

これらを踏まえて謎解きをしていきましょう。わからない量を確率の言葉で表現し，データを使って少しでも正解に近づこうとするのが統計のやり方です。まずここでわからない量は，「正確な測定値」と「個人毎の誤差の大きさ」です。誤差は正規分布に従うと思われますから，データ Y は $Y \sim N(\mu, \sigma)$ と表すことができます。データ生成メカニズムという意味では，「正規分布に従ってデータが作られる」といってもいいかもしれません。また，正確な測定値はここでは μ ということになります[*8]。また，データは 7 点あって，それぞれ $Y_1, Y_2, \cdots, Y_7$，あるいは一般に Y_i と書くことにしましょう。ここで添字の i は第 i 回目の測定でもありますし，測定者 i のことでもあります。この測定者 i 毎に誤差の出方の大きさが変わります。誤差の大小，言い換えれば測定の精度は誤差の SD で表現されていますから，σ が個人毎に変わる，すなわち σ_i と書くことができます。これが今回のデータ生成モデルです！

しかしまだ，μ と σ_i は結局なんなんだ，どういう数字なんだということがわからないままです。**わからないことは確率の言葉で表現する**のがベイジアンの生き様。ここは未知のパラメータがどうやって出てくるかなんてわからないところなので，「わからない」ことを確率で表現する必要があります。パラメータに対して「この辺りにあるだろう」と事前に想定する確率分布のことを**事前分布 (prior distribution)** というのでした。

ここでは μ は平均 0，SD100 の正規分布からきていることにしましょう。正規分布というのは左右対称で平均値・中央値・最頻値が一致する単峰の分布です。これは正確な測定値を表しているのでしたから，なんらかの値を取るとしても 1 つの値だろう，というのは無理のない仮定でしょう。複数の値を取る可能性があるのなら多峰性の分布を仮定したらいいと思いますが，今回はそうではないだろう，と仮定するのです。平均 0 で SD100 ということは，-300 から +300 の範囲にある可能性が 99.7% だということでもあります[*9]。実際のデータが -27 から 10 ぐらいの範囲に入っているので，-300 から $+300$ の間というのも十分過ぎるぐらい広めの範囲にしてある数字といえるでしょう。このように，うっすら情報を持っているけどほとんど意味がない

[*8] 正確な測定値 μ に誤差 $N(0, \sigma)$ がついてデータになる，すなわち $Y = \mu + N(0, \sigma)$ ということですが，誤差の平均は 0 で μ の分だけ正規分布がズレる，そして「分布に従う」を表現する $\sim$ を使いたいので，$Y \sim N(\mu, \sigma)$ となるのです。

[*9] 正規分布は，平均 $\pm$1SD の範囲に全体の 68% が，$\pm$2SD の範囲に 95% が，$\pm$3SD の範囲に 99.7 が入ることが理論的にわかっています。わからない人は R で計算して確認してみよう。

程度に縛りをかけるのを，**弱情報事前分布 (weakly informative prior distribution)** といいます。たとえばこれが身長のデータであれば，$\pm$3m の範囲は十分すぎ，なんなら負の数は取るはずがないので過剰な安全策を置いているともいえます。それでも「なんらかの事前分布を持つのは，主観的な思い込みだ」というのであれば，$\pm\infty$ の一様分布を考えるか，ベイズ法をやめるかということになります。

次に σ_i です。これは標準偏差パラメータなので負の数を取ることはあり得ません。ここでは**半コーシー分布 (half-cauchy distribution)** を事前分布におくことにします。コーシー分布とは正規分布によくにた形ですが，正規分布より裾が重く[*10]，標準偏差 (分散) の事前分布に適しているといわれている分布です (Gelman et al., 2006)[*11]。このコーシー分布は正規分布のように左右対称ですが，0 を中心に半分に折りたたんで正の値しか取らないように考えたのが半コーシー分布です。今回はスケールパラメータに 5 を置いていますが，とくにこの数字に深い意味はありません。気になるようでしたら 10 でも 100 でも変えていただいて結構です[*12]。

これらをおくことで，モデルの設計図ができあがりました (図 10.1)。あとはこれをもとに，**確率的プログラミング言語 (stocastic programming language)** に書き起こしていけば良いでしょう。

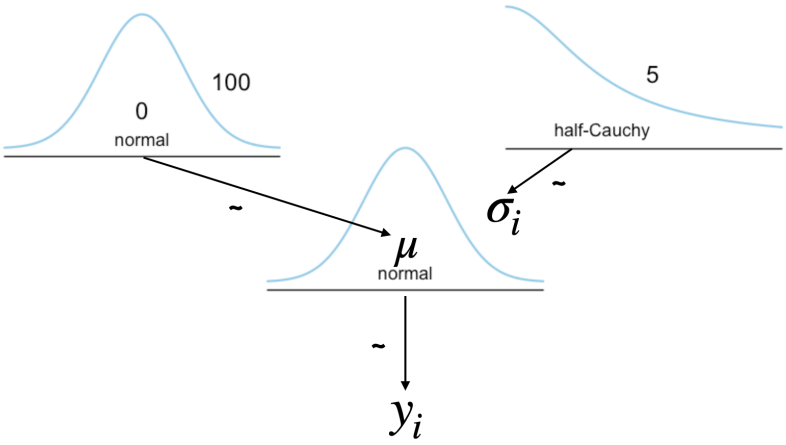

図 10.1　モデルの全体像

この授業では，確率プログラミング言語の演習をすることは目的としていませんから，書けたものとして結果だけ示すことにします[*13]。

ここでは**事後確率最大値 (Maximum A Posterior)** を見ることにしましょう。これによると，μ すなわち真の測定値は 9.596 ぐらいじゃないか，という推測がなされています。10 より少し小さいくらいではないか，と思っていたようですが，こうして計算してみると確かにそれぐらいのようですね。そして 7 人それぞれの測定精度が σ_i として算出されています。これをみると，最初の測定者は 23.101，2 番目の人は 6.849 になっています。これが大きいということは，誤差の幅広さを表しているのですから，測定精度が悪いということを意味します。確かに最初の二人は非常に精度が悪く，他のベテラン測定者が 0.3 とか 0.9 ぐらいの誤差であることに比べれば，圧倒的に悪いということがいえますね。

[*10] より極端な値が出る可能性が多いという意味です
[*11] この論文では他にも，student の t 分布などを提案していますが，パラメータ数がより少ないコーシー分布を選びました。
[*12] 半コーシー分布がどんな数字になるのか気になる，という人は R で `rcauchy` 関数を使って乱数を発生させ，その挙動を見てみればいいでしょう。乱数発生によって大体の分布感 (分布の感覚?) を得ることができる，ということを思い出してください。
[*13] 詳しく知りたい人は，テキスト伴走サイトにある (専修大学の授業講義資料の)「心理学データ解析応用」テキストを見てください。

表 10.1　MCMC の結果

パラメータ	EAP	MED	MAP	U95	L95
μ	9.570	9.622	9.596	10.547	7.753
σ_1	42.197	26.955	23.101	151.573	14.259
σ_2	8.776	6.853	6.849	30.356	2.590
σ_3	4.310	3.232	3.626	17.280	0.440
σ_4	2.756	1.102	0.993	15.750	0.085
σ_5	2.507	1.052	0.358	11.823	0.022
σ_6	2.503	1.676	2.021	12.339	0.076
σ_7	2.598	1.020	0.674	13.548	0.100

もちろんこれは点推定値で，真の測定値が 9.596 に違いないとか，最初の人の測定精度が 23.101 だと断言すると，おそらくほぼ確実に外れた予測ということになるでしょう。そこで幅を持った予測，すなわち**区間推定 (interval estimation)** を行えば良いのです。区間を考えれば，真の測定値は 10.547 から 7.753 の範囲にあるだろうとか，最初の測定者の精度も最悪 151.573，最善なら 14.259 ぐらいである可能性があるわけです。

ベイズ推定のポイントとして，これらは事後分布，すなわち**確率分布による予測**であるということが挙げられます。真の測定値 μ のありそうな範囲が 10.547 から 7.753 にある確率が 95% だ，といういい方ができるところです。**モーメント法**による**信頼区間 (confidential intervals)** とは違い，ベイズ法の予測は**確信区間 (Credible Intervals)** といういい方になります[*14]。

いかがでしょうか。この「7 人の科学者」がおもしろい点は，少ないサンプル (たった 7 件のデータ！) で推測ができること，簡単なコードで検証できること，というのもありますが，なにより標準偏差 (分散) を検証対象とすることにあると私は思います。心理学の実験のほとんどが，平均の比較，操作の効果の話になっているのですが，データの散らばりというのも測定精度のように心理学的に意味のあるものとして考えられ，検証の対象にしたっていいのです。もちろん正規分布でなくてもいいですし，さまざまな分布のさまざまなパラメータに意味があるなら，それをわからないものとして検証しようといえる，というのはとても可能性が広がる話ではないでしょうか。

10.4.2　帰無仮説検定の代わりに

心理学の一般的な研究は，データドリブン，すなわちデータが得られてからその意味を考えるというものでしたが，モデリングはデータ生成メカニズムを考える，すなわち数値の出自とその意味を考えるというアプローチです。このアプローチを逆に従来のデータドリブン型研究，たとえば帰無仮説検定の文脈で扱うとどのようなことがわかるでしょうか。

たとえば t 検定は，二群の平均値の差を比較し，結論を下すという最も典型的な帰無仮説検定の例の 1 つです。とくに独立した二群の検定は，最も単純な要因計画 (一要因二水準 Between デザイン) ということができるでしょう。ちょっとよくわからない，という人は，山田 & 村井 (2004)，清水 (2021) など優れたテキストがいろいろありますので，そちらで改めて分析の流れや仮定をもう一度確認しておいて欲しいと思います。

非常に駆け足ながら概略を説明してみますと，次のようになります。

[*14] 信頼区間の場合は，ここにあると予測したとして，その予測が当たる確率を表しています。すなわち当たるか外れるか，0/1 の話をしているのであって，この「範囲に存在する」という大きさ・幅の話をしているわけではないことに注意です。

1. 標本の母集団が正規分布していると考える。母集団から無作為に選ばれた標本が二群あるとする。
2. 二群の一方に何らかの処置を加え，他方には何もしない (あるいは検討したい処置と同等の効果のない処置を施す[*15])。前者を実験群，後者を統制群という。データ (標本) は正規分布するはずだから，その平均値を見ることで誤差や個人差は相殺される。つまり群間の平均値の差が効果の大きさであるということができる。
3. ところで標本平均値などの**標本統計量**も正規分布に従うことがわかっている。母集団が $N(\mu, \sigma)$ であれば，標本平均は $N(\mu, \frac{\sigma}{\sqrt{n}})$ に従う (ここで n はサンプルサイズ)。
4. 標本平均が従う分布 (散らばりの位置と幅) が分かれば，母平均がどのあたりにあるのかは区間推定することが可能。実験群・統制群ともに母平均の値を推定する。これが異なっていれば標本を超えて母集団で効果があった，と結果を一般化できる。もちろん点推定値は母平均の値とピッタリ同じとは思えないし，標本平均もピッタリ同じになるはずがないから，点推定値ではなく区間推定で考えたい。
5. ところで標本平均の従う分布の中には，σ つまり母 SD がパラメータとして入っている。母数がわからないという前提のもとでは，どの程度の幅で散らばるのかがわからないことと同じ。そこで σ の代わりに $\hat{\sigma}$ を用いて推定することを考える。この場合，標本平均は正規分布ではなく t 分布に従うことがわかっている。
6. t 分布を使った区間推定をしても，差があるのかないのか判断するために何らかの基準が必要である。そこで「差がない」という主張と「差がないとはいえない」という主張を戦わせて，判定するという形式を取る。前者の主張を**帰無仮説**，後者の主張を**対立仮説**という。また判断も確率的になるので，勝敗を決める基準を 5% とする。この確率は**有意水準**とか**危険率**と呼ばれる。
7. データから t 分布に従う統計量を計算し，その値が出てくる確率を算出する。この確率は p **値**と呼ばれる。これが有意水準よりも小さいようであれば，帰無仮説の仮定の下で算出した数字が滅多に生じ得ないことを意味するから，帰無仮説が間違っていたのだと判断してこれを**棄却**し，対立仮説を**採択**する。

さて，この二群の平均値を検定することの流れですが，とくに「帰無仮説のもとでの t 値を算出して判定する」というあたりがややこしかったかもしれません。数式で書くと嫌われそうですが，帰無仮説というのは二群の平均値に差がない，という仮定だったので，実験群から考えられる母平均 μ_A と統制群から考えられる母平均 μ_B に差がない，つまり $\mu_A = \mu_B$，からの $\mu_A - \mu_B = 0$ という位置母数が 0 の理論分布のを考えている，というところがポイントです。検定統計量である t 値は次のように計算できるのでした。

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{(n_1-1)\hat{\sigma_A}^2+(n_2-1)\hat{\sigma_B}^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}}$$

とっても複雑な式に見えますが，複雑そうなのは分母で，分子は標本平均の差を表しているに過ぎません。参照する t 分布は，標準正規分布が $N(0,1)$ だったように，位置と幅のパラメータを持ち，$t(0, df)$ で考えられる分布にこの統計量を照らし合わせると，差 0 で自由度に応じて変わる幅 df の分布から「どの程度極端な値が出てきたのか」の確率を計算できるわけです。ちなみに分母は，母分散ではなく標本から計算される分散 s^2 からバイアスを除いた不偏推定量 $\hat{\sigma}^2$ を使っています。この計算はサンプルサイズ n ではなく $n-1$ で割ることによって計算されるのですが，2 つの群それぞれについて $n_j - 1$ で割ったものを足し合わせるために，いったん $n_j - 1$ 倍して二群のサンプルサイズ -2 で割り直す，という作業をしているため，分母がとくに

[*15] たとえば「単語リストは声に出して覚えたほうが記憶の定着度が高い」ということを検証したい場合，声に出す群と出さない群で比較することになりますが，声に出さないことが (頭の中で反芻するなど) 従属変数に与える別の効果を持っていることがあるので，音のない映像を見せるなどして効果のない同等の処置をした群を比較対象にする，ということを考えたりするわけです。

複雑に見えているだけです。

ともあれ，こうして検定するんだったという流れを思い出したところで，データ生成モデリングの観点からこれを考え直してみましょう。仮定としておいてあるのは，両群ともに同一の正規分布から得られた標本であり，操作によってその平均値が異なっているはず，ということだけです。つまり実験群のデータは $X_{i,A} \sim N(\mu_A, \sigma)$，統制群のデータは $X_{i,B} \sim N(\mu_B, \sigma)$ という分布が前提とされているのです。

モデルの設計図を図 10.2 に挙げておきます。

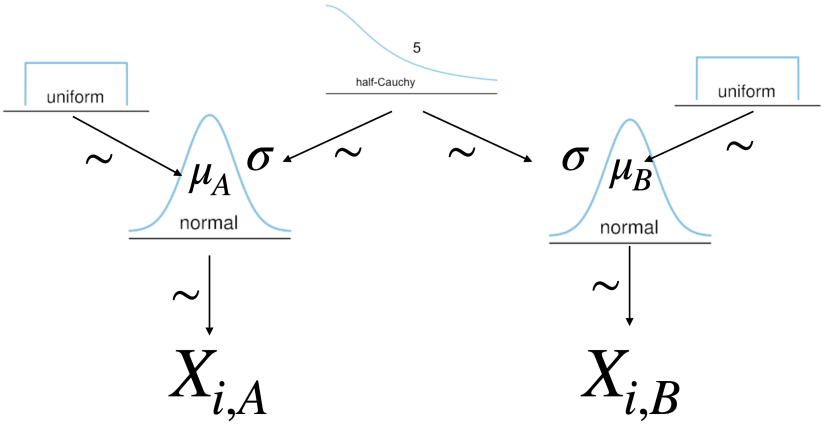

図 10.2　平均値の差を比べるときの設計図

では少し具体例をつかって，これがどういう結果をもたらしているのかを確認しましょう。

t 検定の具体例

> 統計学の新しい指導法の教育効果を見るため，全国の大学の心理学科から学生を無作為に 16 名選び，従来の指導法グループ (統制群) と，新指導法のグループ (実験群) にランダムに 8 名ずつわりつけました。プログラム終了後，心理統計のテストを行いました。統制群のスコアは $20, 40, 60, 40, 40, 50, 40, 30$，実験群のスコアは $30, 50, 70, 90, 60, 50, 70, 60$ となりました。新たしい指導法は，心理学科の学生に効果があるといえるでしょうか？ 5% 水準で検定してください。

これを t 検定するのは簡単ですね。計算式は複雑でも，機械がやってくれるから楽なのが NHST[*16]のいいところです。R でこれを分析するには，次のようなコードを書きます。

code : 10.1　帰無仮説検定のコード

```
1  groupA <- c(30, 50, 70, 90, 60, 50, 70, 60)
2  groupB <- c(20, 40, 60, 40, 40, 50, 40, 30)
3  ## t検定
4  t.test(groupA, groupB, var.equal = TRUE)
```

結果は出力 10.1 のようになります。検定統計量である t 値，検証のための自由度 df，帰無仮説のもとでの出現確率 p 値，が表示されています。5% より小さいので有意差あり，という判断ができます。もっとも，t 値とか自由度とかは，データに関係のない数字なのでピンと来ないというところもあるかと思います。

[*16] **帰無仮説検定 (Null Hypothesis Significance Test)** の略です

R の出力 10.1: 検定の結果

```
> t.test(groupA, groupB, var.equal = TRUE)

 Two Sample t-test

data:  groupA and groupB
t = 2.6458, df = 14, p-value = 0.01919
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.786937 36.213063
sample estimates:
mean of x mean of y
       60        40
```

モデルで推定してみる

つづいて，先ほどのモデルに基づくモデリングを行い，ベイズ推定をした結果を示します。モデリングのコードや分析方法については，本書の伴走サイト[*17]にある「データ解析応用」のテキスト等を参照してください。

cmdstanr の出力 1: MCMC の結果出力

```
variable   mean median   sd  mad     q5    q95 rhat ess_bulk ess_tail
    lp__ -51.56 -51.21 1.33 1.08 -54.10 -50.12 1.00     8313    10618
    mu1   60.07  60.03 5.57 5.30  51.01  69.19 1.00    15288    12028
    mu2   40.04  39.98 5.54 5.34  31.07  49.13 1.00    17592    12673
    sig   15.47  15.01 3.07 2.77  11.39  21.14 1.00    12956    11789
```

これを見ると，未知のパラメータ μ_1, μ_2, σ の値がそれぞれ推定されていますが，「判定」のような結果は出てきていません。というのも，勝負するような形で考えているのではなく，パラメータがどこにありそうか，ということを示す**確率分布**を求めているからです。それでも，**EAP** をみると 60 と 40，95%CI をみると実験群は `[51.01, 69.19]`，統制群は `[31.07, 49.13]` とあり，実験群の平均値が 51 より小さい値になる確率はほとんどなく (2.5% 以下)，統制群の平均値が 49 以上より大きくなることもまたほとんどないわけですから，確実に差があると言っても過言ではないでしょう。

帰無仮説検定の注意点とベイズ法の利点

帰無仮説検定の利点は分析手順の標準化が終わっているところです。統計パッケージの中に分析ツールが入っていますから，たとえば R では `t.test` と書けば答えを出してくれるのです。

しかし帰無仮説検定は，初学者を惑わすわかりにくい点があることにも注意が必要です。まず判定に使う p 値についての理解がわかりにくい点です。とりあえずこれが 5% より小さければいいや，星マーク (正確にはアスタリスク) がたくさんついてればいいや，と思っていませんか? p 値の定義は「帰無仮説が正しいとしたとき，手元のデータから得られる検定統計量以上の極端な値が出る確率」です。これはつまり，

- p 値は帰無仮説の正しさ/正しくなさの指標ではありません;「帰無仮説が正しい」という仮定のもとでの世界であり，この世界での帰無仮説の正しさは 100 %です。

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

- p 値はデータが偶然発生した確率ではありません;帰無仮説の世界といういわば空想上の世界の数字であって, 実際の世界と対応する数字ではありません。
- p 値は効果の大きさや結果の重要性を示唆するものではありません;帰無仮説は空想上の世界の数字であって (以下略)。

ということです。自分達が想定した「帰無仮説の世界」という仮想世界の中で, じつデータを当てはめた判定をしているのであって, p 値はその仮想世界の数字です。現実世界と直接の関係がない指標なので, これの大小を議論することは不毛なわけです。

p 値は「仮説が正しい時のデータ出現率, 統計の記号で書くと $P(D|H0)$ であって,「データから考えられる仮説の確らしさ, つまり $P(H|D)$ ではないのです。仮想世界のもとで算出される数字であり, 仮想世界を操作すれば p 値は容易に操作でき, これが**疑わしい研究実践 (Questionable Reserch Practices,QRPs)** を生むことにもなっているわけです。昨今では研究方法の事前登録や例数設計など, 研究マナーの向上が必要です。

しかしなにより, 重要な差は「有意差」ではなく,「実質的な差」ではないでしょうか。もし実質的な差が単位に依存し, 研究間の比較ができないことが問題になるなら「標準化された差」, つまり**効果量 (effect size)** を考えるべきなのです。このことは古くから言われ続けてきましたが, (リッカート法が無批判に使われてきたように!) 実践上はほとんど無視され続けてきました。

しかしモデリングの技法を使うと, これらの問題点が改善される見込みがあります。結果を見てわかるように, 知りたかったことは「平均値がどの程度異なるか」「母平均の位置はどこにありそうか」ということであり, それが幅をもって表現されているわけですから, 最初から実質的な差の程度を結果示しています。この程度の幅であれば差があるといっていいだろういえないだろう, という判断はドメイン知識に基づいて考えることができます。従来の「パラメータが同じだとした仮想空間のなかではあり得ないことなんだけど」といったもって回った考え方は不要で, 値がどれぐらいか, どれぐらいの差があるか, その結果の確からしさ (確信度) は高いといえるか, といったことが考えられます。またこれらは母平均というパラメータに対する仮説ですが, データ生成メカニズムを考えていますから, そのメカニズムに照らし合わせて, このモデルが作るデータは手元のデータと合致しているか, データ上での違いはどの程度見込まれるか, パラメータが変わればデータはどのように変わるか, といったことをシミュレーションすることもできます。

仮説の立て方も, より自由になるでしょう。パラメータがある量 c より大きくなる可能性はどれぐらいあるかとか, データ上での差がある量 d を上回る確率はどれぐらいあるか, といったことです。また帰無仮説検定ではできなかった,「差がないのではないか」という仮説を検証することもできます。

ベイズ統計が全ての問題を払拭してくれる万能な考え方だ, とはいいません。そうした持ち上げ方はいずれまた,「ベイズさえ使っていれば他のことを考えなくていい」という人間特有の弱さに負けてしまうでしょう。帰無仮説検定も正しく使えば問題のあるツールではありませんし, NHST やベイズといったツールをつかって研究をより良いものにしようという流れは結局,「ちゃんと勉強しよう」にしか行きつかないのです。

自身の立場, 仮定, 背景について深く理解し, 言語化できるようにしておこうという姿勢が必要だということだけは, 間違いのない人生訓のようです。

第 11 章

おわりに

この講義では，心理学における測定，データの扱いについて論じてきました。以下に各章の流れをまとめてみたいと思います。

第 1 章では心理学が何を研究対象にしているのかを考え，態度尺度と呼ばれる方法で作られる数値はなぜそれで間隔尺度水準の数値になるといえるのか，その理論的裏付けについて考えました。第 2 章では心理学と少し違った角度，テスト理論という文脈で，目に見えないものを測定するモデルとしてどのようなものが考えられてきたか，古典的テスト理論と現代テスト理論に言及しながら考えてきました。第 3 章では一転して心理学に戻り，心理学でもっともよく使われている測定モデルである因子分析法について，そのモデルと実践例について考えてきました。第 4,5 章は線形代数について学びました。数学的側面なので苦手意識を克服できなかった人もいるかも知れませんが，ここで正方行列のスペクトル分解を理解しておくと，その後のデータの扱い一般についての理解が格段に広がるため，どこかで触れなければならなかったのです。上手く理解できなければ，こちらの伝え方の方に問題があります。自分を責めずに，わかるまでトライしてもらえれば幸いです。第 6 章ではここまでの知識を踏まえた上で，実際に心理学の研究において心理尺度がどのように使われているか，言語による測定がどのようなものかを見てみました。態度尺度のような理論的裏付けが及ばない領域での利用が多くみられ，翻ってそれぞれの領域で考えるべき点，守られるべき点や，どのような改善方法が考えられるかについて検討しました。第 7 章では，心理尺度に至る前のプリミティブな心の判断，類似性をデータにした場合にどのような分析が可能かについて，クラスター分析や多次元尺度構成法について学びました。第 8 章では多次元尺度構成法の中でも，非計量的多次元尺度法や個人差モデル，展開法，非対称多次元尺度法などの応用例を見てきました。これらの方法は心理尺度による研究よりもデータが頑健で，個人差についての明示的なモデルがあり，尺度に対する反応も異なる観点から捉えてモデル化できること，対人関係のような非対称性が重要な側面でも応用可能なモデルがあることなど，その広がりについて論じました。第 9 章では双対尺度法を中心に，名義尺度水準のデータを時限解析する方法についてみていきました。ここに至って「カテゴリに数値を付与する」という明確な方向性を改めて確認し，分析者にとって意味のある数値化についてのいくつかの応用モデルもみていきました。第 10 章ではモデリングアプローチから，そもそもデータがどのような機序で生成されているか，データ生成メカニズムの観点から分析する方法を見てきました。数理モデルをデータと照合するための便利な方法として，ベイズ統計の枠組みである確率的プログラミング言語の助けを借りながら，研究関心を直接表現することでデータドリブンの研究スタイルを転換させる手段を導入しました。またこの手法から従来法を眺め直すと，いろいろ不便なところも見えてきましたが，両者の良いところを見ながら使うことで今後の分析が楽しくなれば良いなと思っています。

ここでお話しした内容は，「心理学統計法」という枠組みというより「心理測定法」の枠組みに入る内容です。測定法といえば，心理学研究法の下位分野と考えている人もいるかもしれません。

心理学界隈では昨今，公認心理師という国家資格がつくられたことにより，資格対応を考える大学ではカ

リキュラムが標準化される動きがありました。心理学は非常に幅広い領域をカバーする学問ですし，これに医療や法律の観点を加えて 4 年間で履修するプログラムを考えるとなると，各領域をどのようにまとめるか，その境界を引いたり複合領域をどう扱うかという問題に直面することになります。当然カリキュラムを考えるところでは激しい討論が重ねられたでしょうし，それによって提供された現状のものが完璧であるとはいえないと思います。しかしそれは資格取得を目指す学生側には関係のないことで，学ぶ人たちはそれぞれ自らの思いを持って，誠実に決められたことを学ぶしかないでしょう。

残念ながら，心理学の各領域の専門家と言われる人たちも，その領域のあらゆるジャンルをマスターしているわけではありません。教える側にも単元ごとに得手不得手があり，よく知っているところはうまく教えられるでしょうが，よく知らないところは教科書の表面的なところをなぞるだけ，ということになるかもしれません。心理統計の領域はそもそも専門家の数が少ないのですが，それに比べてユーザの数，手練れの経験者の数は非常に多いので，ややもするとノウハウ寄りの教え方をする人が多いかも知れません。また教わる側も，心理学と数学を同時に愛する人は少なく，そもそも数学嫌いだからこそ文系たる心理学に行く，という人もいるでしょうから，ややもするとノウハウのみを身につけて，試験をクリアすることを目的にする人が出てくるでしょう。測定の最初に学んだように，資格やその試験など数値目標を設定すると，それを獲得するためのルートは必ずハックされ，チートを探す人が出てきます。公認心理師の試験は膨大な領域に及びますから，合格するための良い戦略は「心理統計の問題が出たら捨てる」かもしれません。

これは「資格」というもの全体の問題で，資格になるからこそ専門性が下がるというパラドキシカルな現象はどこでもあり得る話です。とくに心理統計に限って声高にいうべきことではないのかもしれません。しかしここであえていいたいのは，**心理測定は心理統計や心理学研究法の下位領域ではなく，心理学そのものである**と筆者は考えるからです。ノウハウに振り切って教えられることで，原理原則が無視されたり軽んじられ，そもそもの発想からは合理的な裏付けがえられないほどに奇妙な研究実践が蔓延してしまっている様を，皆さんも垣間見たのではないでしょうか。資格や科目の一部に矮小化されたり，優先順位が下がることは仕方がないこととはいえ，**教えられなかったものは存在しなくなる**可能性があります。古い心理学の全集などをみると，かならず数理心理学，計量心理学で一冊は書かれるほどリッチな研究と考察が重ねられていました。今や実験実習の 1 つの単元で尺度作りが行われるほど，そのノウハウは一般化しましたが，その数字が意味するものが何なのか，誰も知らなくなる日がそこまできているのかもしれません。

心理学が心の学問である，行動の学問であるというのはジョークの一種であり，実は空想の世界の SF 小説で，架空の世界で架空の数値を使って，さらに架空の統計量を使って面白おかしくストーリーを物語っているだけだ，というのであれば，それはそれで構わないと筆者は思っています。心理学以外の領域，とくにハードサイエンスの人から見たら，私たちがやっていることは総じてコントなのかもしれません。いや違う，心を，人間を，行動を測って理解しようとしているのだというのであれば，馬鹿馬鹿しくも蔓延したルーチンワークのような尺度作成は捨て去って，より建設的な，人の心の深淵に触れうるための学問を積み重ねていかなければならないと思いませんか。

大事なものがこぼれ落ちていく世の中で，私のささやかな足掻きが読者のみなさんにちょっとした引っかかりを生むことができれば，望外の喜びです。

付録 A

標準正規分布から尺度値を求める計算方法

Likert 法では，態度が標準正規分布すると仮定するのでした。標準正規分布をカテゴリの相対度数で分割し，あるカテゴリ c の上限の確率点 z_c，下限の確率点 z_{c-1} の確率密度の差分を，相対度数 p_c で割ることで，尺度値 Z_c が下の式で得られます。

$$Z_c = \frac{(y_{z_{c-1}} - y_{z_c})}{p_c}$$

ここで y_{z_c} は標準正規分布をカテゴリで区分したときの，当該カテゴリ c までの累積確率点 z_c における確率密度，p_c はカテゴリ c の相対頻度です。図 A に記号の対応関係を示しましたので，確認してください。

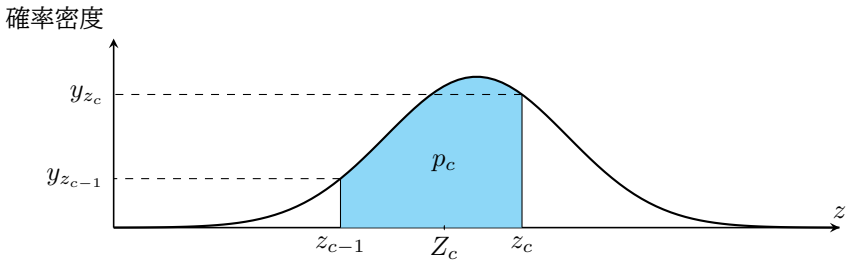

図 A.1　標準正規分布と対応する記号の確認

尺度値 Z_c を求める計算が確率密度 $y_{z_{c-1}}, y_{z_c}$ と，相対度数 p_c で算出されるというのは一見奇妙です。どうしてこのようになるのかを見ていきましょう。

まず標準正規分布の確率密度の式を確認しておきます。確率点 z における確率密度 y は次の式で算出できます[*1]。

$$y_z = f(z) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{z^2}{2})}$$

この関数は確率密度の曲線を表しており，確率はその面積です。z_{c-1} から z_c までの面積 (確率) は，積分を使って

$$p_c = \int_{z_{c-1}}^{z_c} f(z)dz = \int_{z_{c-1}}^{z_c} \frac{1}{\sqrt{2\pi}} e^{(-\frac{z^2}{2})} dz$$

[*1] e^x を $\exp(x)$ と書くことも少なくありませんし，この e は自然対数の底で $e = 2.718...$ という実数です。この数字は微分しても変わらない，$(e^x)' = e^x$ という便利な特徴を持っています。

で表されます。

さて，ある確率点 z における密度の高さを $f(z)$ としたとき，z の微小な増分 Δz を考えると，区間の面積 $S = f(z)\Delta z$ を考えることができますから，逆にある点 z が知りたい時は

$$z = \frac{\sum Sz}{\sum S}$$

とすれば良いことになります。積分はこの微増分 Δz の極限を

$$\lim_{\Delta z \to 0} \Delta z = dz$$

と考えることですから，ここで考えたいのは

$$Z_c = \frac{\int_{z_{c-1}}^{z_c} f(z)zdz}{\int_{z_{c-1}}^{z_c} f(z)dz}$$

になります。分母は確率密度関数の積分ですから面積すなわち確率で，ここでは p_c であり，その面積を実データの相対度数で代えてもいいでしょう。

分子については少し式の変形が必要です。記号を見やすくするために，$a = z_{c-1}, b = z_c$ と書き換えておきましょう。

$$\begin{aligned}\int_a^b yzdz &= \int_a^b \frac{1}{\sqrt{2\pi}} e^{(-\frac{z^2}{2})} zdz \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{(-\frac{z^2}{2})} zdz\end{aligned}$$

ここで変数変換をして，$u = -\dfrac{z^2}{2}$ とすると

$$\begin{aligned}\frac{du}{dz} &= -z \\ du &= -zdz\end{aligned}$$

となり，積分の下限は $-a^2/2$, 上限は $-b^2/2$ になりますから，以下のように展開できます。

$$\begin{aligned}(\text{与式}) &= -\frac{1}{\sqrt{2\pi}} \int_{-a^2/2}^{-b^2/2} e^u du \\ &\qquad\qquad \int e^u du \text{ は } e^u \text{ なので} \\ &= -\frac{1}{\sqrt{2\pi}} \left[e^u\right]_{-a^2/2}^{-b^2/2} \\ &= -\frac{1}{\sqrt{2\pi}} e^{(-\frac{b^2}{2})} - \left(-\frac{1}{\sqrt{2\pi}} e^{(-\frac{a^2}{2})}\right) \\ &= \frac{1}{\sqrt{2\pi}} e^{(-\frac{a^2}{2})} - \frac{1}{\sqrt{2\pi}} e^{(-\frac{b^2}{2})} \\ &\qquad\qquad y = \frac{1}{\sqrt{2\pi}} e^{(-\frac{z^2}{2})} \text{ であることから，} \\ &= y_a - y_b \\ &= y_{z_{c-1}} - y_{z_c}\end{aligned}$$

以上のことから，リッカート法において標準正規分布をもとに尺度得点を決めるには，

$$Z_c = \frac{(y_{z_{c-1}} - y_{z_c})}{\int_{z_{c-1}}^{z_c} f(z)dz} = \frac{(y_{z_{c-1}} - y_{z_c})}{p_c}$$

とすれば良いことになります。

この計算に至る理論的背景は，より専門的には**系列範疇法 (Method of Successive Categories)** と呼ばれ，順序カテゴリに数値を付与する心理測定論，精神物理学理論からきています。詳しくは Guilford (1954) や西村 (1977) も参照してください。

付録 B

ギリシア文字一覧

ギリシア文字ってかっこいいけど，読み方わからない･･･という人のための一覧。ついでに TEX 表記も。

表 B.1　ギリシア文字とその読み方

読み方	大文字	小文字	英語表記	TEX 表記
アルファ	A	α	alpha	`\alpha`
ベータ	B	β	beta	`\beta`
ガンマ	Γ	γ	gamma	`\gamma`
デルタ	Δ	δ	delta	`\delta`
エプシロン	E	ε	epsilon	`\varepsilon`
ゼータ	Z	ζ	zeta	`\zeta`
イータ	H	η	eta	`\eta`
シータ	Θ	θ	theta	`\theta`
イオタ	I	ι	iota	`\iota`
カッパ	K	κ	kappa	`\kappa`
ラムダ	Λ	λ	lambda	`\lambda`
ミュー	M	μ	mu	`\mu`
ニュー	N	ν	nu	`\nu`
クサイ	Ξ	ξ	xi	`\xi`
オミクロン	O	o	omicron	`\mathrm{o}`
パイ	Π	π	pi	`\pi`
ロー	R	ρ	rho	`\rho`
シグマ	Σ	σ	sigma	`\sigma`
タウ	T	τ	tau	`\tau`
ウプシロン	U	υ	upsilon	`\upsilon`
ファイ	Φ	ϕ	phi	`\phi`
カイ	X	χ	chi	`\chi`
プサイ	Ψ	ψ	psi	`\psi`
オメガ	Ω	ω	omega	`\omega`

引用文献

Allport, G. W. (1967). *Attitudes* (M. Fishbein, Series Ed.). John Wiley & Sons Inc.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance.

Bergson, H. (1889). *Essai sur les données immédiates de la conscience* (正. 合田 & 靖. 平井, Trans.). 筑摩書房. (Original work published)

Carroll, J. D. & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, *35* (3), 283–319.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological review*, *57* (3), 145.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, *1* (3), 515–534.

Grimm, L. G. & Yarnold, P. R. (1994). *Reading and undestanding multivariate statistics*. American Psychological Association.

Guilford, J. (1954). *Psychometric methods*. McGraw-Hill Book Company.

Kruskal Joseph, B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29* (1), 1–27.

Kruskal Joseph, B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, *29* (2), 115–129.

Lee, M. D. & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling:a practical course*. Cambridge University Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, *22*, 5–55.

Muller, J. & Muller, J. Z. (2018). *The tyranny of metrics* (裕. 松本, Trans.). Princeton University Press. (Original work published)

Muraki, E. (1992). A generalized partial credit model: application of an em algorithm. *ETS Research Report Series*, *1992* (1), i–30.

Okada, A. & Imaizumi, T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, *14* (21), 81–96.

Okada, A. & Imaizumi, T. (1997). Asymmetric multidimensional scaling of two-mode three-way proximities. *Journal of Classification*, *14* (2), 195–224.

Samejima, F. (1997). *Graded response model* (, Series Ed.). Springer.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103* (2684), 677–680.

三浦, 麻. & 小林, 哲. (2015). オンライン調査モニタの satisfice に関する実験的研究. *社会心理学研究*, *31* (1), 1–12. https://doi.org/10.14966/jssp.31.1_1

下山, 晴. (2001). *臨床心理学研究の多様性と可能性* (晴. 下山 & 義. 丹野, Series Eds.). *Vol.2*. 東京大学出版会.

中尾 達馬 (2021). コロナ禍での大学生におけるアタッチメントと孤独感や精神的健康との経時的な相互関係. 心理学研究, *92* (5), 390–396. https://doi.org/10.4992/jjpsy.92.20320

加藤, 健., 山田, 剛., & 川端, 一. (2014). *R による項目反応理論* (Kindle 版). オーム社.

北條, 大. & 岡田, 謙. (2018). 係留ビネット法による反応スタイルの分類 ヨーロッパの大規模健康調査を例に. *行動計量学*, *45*, 13–25.

千野, 直., 岡田, 謙., & 佐部利, 真. (2012). *非対称 mds の理論と応用* (単行本). 現代数学社.

外山 美樹・長峯 聖人 (2022). 人は困難な目標にどう対処すべきか？. 心理学研究, *92* (6), 543–553. https://doi.org/10.4992/jjpsy.92.20220

外山 美樹・長峯 聖人・海沼 亮・湯 立・三和 秀平・相川 充 (2021). 制御適合した欲求支援行動がエンゲージメントに及ぼす効果. 心理学研究, *92* (4), 257–266. https://doi.org/10.4992/jjpsy.92.20015

天井 響子 (2021). 青年期前期の援助評価に対する情緒的援助期待の影響. 心理学研究, *92* (2), 140–150. https://doi.org/10.4992/jjpsy.92.19233

太幡 直也・佐藤 広英 (2021). 他者のプライバシー意識と twitter 上での他者情報公開との関連. 心理学研究, *92* (3), 211–216. https://doi.org/10.4992/jjpsy.92.20327

宮崎 由樹・鎌谷 美希・河原 純一郎 (2021). 社交不安・特性不安・感染脆弱意識が衛生マスク着用頻度に及ぼす影響. 心理学研究, *92* (5), 339–349. https://doi.org/10.4992/jjpsy.92.20063

小塩, 真. (2020). *性格とは何か　より良く生きるための心理学 (中公新書)* (Kindle 版). 中央公論新社.

小岩 広平・若島 孔文・浅井 継悟・高木 源・吉井 初美 (2021). 我が国における看護師の新型コロナウイルス感染症への感染恐怖の規定要因. 心理学研究, *92* (5), 442–451. https://doi.org/10.4992/jjpsy.92.20048

小杉, 考. (2014). 学校適応感尺度 fit の開発. *研究論叢. 第 3 部, 芸術・体育・教育・心理*, *64*, 69–82.

小杉, 考. (2018). *言葉と数式で理解する多変量解析入門*. 北大路書房.

小杉, 考., 清水, 裕., 藤澤, 隆., 石盛, 真., 渡邊, 太., & 藤澤, 等. (2011). 家族関係尺度の構成とその階層的因子構造について. *行動計量学*, *38* (1), 93–99. https://doi.org/10.2333/jbhmk.38.93

小杉, 考., 藤澤, 隆., & 藤原, 武. (2004). バランス理論と固有値分解. *理論と方法*, *19* (1), 87–100. https://doi.org/10.11218/ojjams.19.87

小野 由莉花・及川 昌典・及川 晴 (2021). 撤回：性的過大知覚バイアス. 心理学研究, *92* (2), 79–88. https://doi.org/10.4992/jjpsy.92.19042

山内, 光. (2010). *心理・教育のための統計法* (第 3). サイエンス社.

山本 真菜・岡 隆 (2021). 新型コロナウイルス感染者に対するステレオタイプと行動免疫システム活性化の個人差との関連. 心理学研究, *92* (5), 360–366. https://doi.org/10.4992/jjpsy.92.20334

山田, 剛. & 村井, 潤. (2004). *よくわかる心理統計*. ミネルヴァ書房.

岡太, 彬. (2008). *データ分析のための線形代数*. 共立出版.

岡太, 彬. & 今泉, 忠. (1994). *パソコン多次元尺度構成法* (単行本). 共立出版.

川崎 直樹・小塩 真司 (2021). 病理的自己愛目録日本語版(pni-j)の作成. 心理学研究, *92* (1), 21–30. https://doi.org/10.4992/jjpsy.92.19217

平井 美佳・渡邊 寛 (2021). 乳幼児の父親におけるパンデミックによる働き方の変化と家庭と仕事への影響. 心理学研究, *92* (5), 417–427. https://doi.org/10.4992/jjpsy.92.20061

廣瀬 愛希子・濱口 佳和 (2021). 両親関係の情緒的安定性が青年の適応に与える影響. 心理学研究, *92* (2), 129–139. https://doi.org/10.4992/jjpsy.92.19229

新国 佳祐・里 麻奈美・邑本 俊亮 (2021). 自己主体感の個人差が主語省略文理解時の視点取得に及ぼす影響. 心理学研究, *92* (2), 89–99. https://doi.org/10.4992/jjpsy.92.19045

杉山 翔吾・廣康 衣里紗 まり・野村 圭史・林 正道・四本 裕子 (2021). 外出規制が孤独感・不安・抑うつに及ぼす影響. 心理学研究, *92* (5), 397–407. https://doi.org/10.4992/jjpsy.92.20081

村上, 正., 佐藤, 恒., 野澤, 宗., & 稲葉, 尚. (2016). *教養の線形代数* (単行本). 培風館.

松浦, 健. (2016). *Stan と r でベイズ統計モデリング (wonderful r)*. 共立出版.

梅本, 堯., 森川, 弥., & 伊吹, 昌. (1955). 清音 2 字音節の無連想価及び有意味度. *心理学研究*, *26* (3), 148–155. https://doi.org/10.4992/jjpsy.26.148

榊原 良太・大薗 博記 (2021). 人々がマスクを着用する理由とは. 心理学研究, *92* (5), 332–338. https://doi.org/10.4992/jjpsy.92.20323

水野 雅之・菅原 大地・谷 秀次郎・吹谷 和代・佐藤 純 (2021). 若手の理学療法士・作業療法士のバーンアウト傾向とセルフ・コンパッションの関連. 心理学研究, *92* (3), 197–203. https://doi.org/10.4992/jjpsy.92.20317

永井 暁行 (2021). コロナ禍の非対面授業における学生の主体的な学修態度. 心理学研究, *92* (5), 384–389. https://doi.org/10.4992/jjpsy.92.20322

永田, 靖. (2005). *統計学のための数学入門 30 講* (単行本). 朝倉書店.

永谷 文代・松嵜 順子・諏訪 絵里子・上西 裕之・谷池 雅子・毛利 育子 (2022). 教師記入式実行機能行動評定尺度の小学生に対する信頼性及び妥当性の検証. 心理学研究, *92* (6), 554–563. https://doi.org/10.4992/jjpsy.92.20226

池上 真平・佐藤 典子・羽藤 律・生駒 忍・宮澤 史穂・小西 潤子・星野 悦子 (2021). 日本人における音楽聴取の心理的機能と個人差. 心理学研究, *92* (4), 237–247. https://doi.org/10.4992/jjpsy.92.20005

清水 佑輔・ターン 有加里ジェシカ・橋本 剛明・唐沢 かおり (2022). 象徴的障害者偏見尺度日本語版(sas-j)の作成. 心理学研究, *92* (6), 532–542. https://doi.org/10.4992/jjpsy.92.20208

清水, 和. (2007). α はやめて ω にしよう. *日本心理学会大会発表論文集*, *71*, 2PM049–2PM049. https://doi.org/10.4992/pacjpa.71.0_2PM049

清水, 裕. (2021). *心理学統計法 (放送大学教材 1638)*. 放送大学教育振興会.

湯 立・外山 美樹・長峯 聖人・海沼 亮・三和 秀平・相川 充 (2022). 学習者のエンゲージメントにおける制御適合の効果. 心理学研究, *92* (6), 564–570. https://doi.org/10.4992/jjpsy.92.20326

生田目 光・猪原 あゆみ・浅野 良輔・五十嵐 祐・塚本 早織・沢宮 容子 (2021). 日本語版幸せへの恐れ尺度と日本語版幸せの壊れやすさ尺度の信頼性・妥当性の検討. 心理学研究, *92* (1), 31–39. https://doi.org/10.4992/jjpsy.92.20206

田中, 熊. (1960). *ソシオメトリーの理論と方法*. 明治図書出版.

石川 遥至・浮川 祐希・野田 萌加・越川 房子 (2021). 注意の分割を伴う気晴らしが気分とネガティブな思考に及ぼす影響. 心理学研究, *92* (4), 227–236. https://doi.org/10.4992/jjpsy.92.19037

神原 歩 (2021). 態度が相反する他者への過度なバイアス認知を錯視経験が緩和する効果. 心理学研究, *92* (1), 12–20. https://doi.org/10.4992/jjpsy.92.20014

荘島, 宏. (2011). Asymmetric von mises scaling : asymmetric multidimensional scaling using directional distribution. *日本行動計量学会大会発表論文抄録集*, *39*, 261–262.

藤原, 武. (2001). *社会的態度の理論・測定・応用*. 関西学院大学出版会.

藤後 悦子・大橋 恵・井梅 由美子 (2021). 子どもの運動に対する養育態度尺度作成の試み. 心理学研究, *92* (4), 267–277. https://doi.org/10.4992/jjpsy.92.20205

藤澤, 等. (1997). *ソシオン理論のコア*. 北大路書房.

西村, 武. (1977). 主観評価の理論と実際. *テレビジョン*, *31* (5), 369–377. https://doi.org/10.3169/itej1954.31.5_369

讃井 知・島田 貴仁・雨宮 護 (2021). 詐欺電話接触時の夫婦間における相談行動意図の規定因. 心理学研究, *92* (3), 167–177. https://doi.org/10.4992/jjpsy.92.20024

豊田, 秀. (2012). *項目反応理論 [入門編](第 2 版) (統計ライブラリー)* (単行本). 朝倉書店.

豊田, 秀. (2020). *瀕死の統計学を救え！*. 朝倉書店.

豊田 雪乃・小林 正法・大竹 恵子 (2021). 援助想像が援助意図に及ぼす影響. 心理学研究, *92* (2), 111–121. https://doi.org/10.4992/jjpsy.92.20002

赤澤 淳子・井ノ崎 敦子・上野 淳子・下村 淳子・松並 知子 (2021). デート dv 第 1 次予防プログラムの開発と効果検証. 心理学研究, *92* (4), 248–256. https://doi.org/10.4992/jjpsy.92.20008

金政 祐司・古村 健太郎・浅野 良輔・荒井 崇史 (2021). 愛着不安は親密な関係内の暴力の先行要因となり得るのか？. 心理学研究, *92* (3), 157–166. https://doi.org/10.4992/jjpsy.92.20013

鈴木 雅之・荒俣 祐介 (2021). 部活動における生徒の動機づけと指導者のリーダーシップとの関係. 心理学研究, *92* (1), 1–11. https://doi.org/10.4992/jjpsy.92.19051

鎌谷 美希・伊藤 資浩・宮崎 由樹・河原 純一郎 (2021). Covid-19 流行が黒色衛生マスク着用者への顕在的・潜在的態度に及ぼす影響. 心理学研究, *92* (5), 350–359. https://doi.org/10.4992/jjpsy.92.20046

飯田 昭人・水野 君平・入江 智也・川崎 直樹・斉藤 美香・西村 貴之 (2021). 新型コロナウイルス感染拡大状況における遠隔授業環境や経済的負担感と大学生の精神的健康の関連. 心理学研究, *92* (5), 367–373. https://doi.org/10.4992/jjpsy.92.20339

高根, 芳. (1980). *多次元尺度法* (－). 東京大学出版会.

髙坂 康雅 (2021). 親の認知した臨時休業中の小学生の生活習慣の変化とストレス反応との関連. 心理学研究, *92* (5), 408–416. https://doi.org/10.4992/jjpsy.92.20040

索引

た

な

は

ま

や

ら