

■S3 群 (脳・知能・人間) - 10 編 (福祉情報)

2 章 心理計測と統計的分析手法並びに評価

(執筆者: 木村貴彦・篠原一光) [2011年3月 受領]

■概要■

製品の評価や利用者のニーズや判断といったことを組織的に得ようとしたとき、それらを出力として得る過程には人間の心理的な特性が多様に影響していることは直感的にも理解することができるだろう。しかしながら、そのような、いわゆる心理的な特性のほとんどは、機器やインタフェースの側の諸特性とは異なり、その特性をとらえることは容易ではない。その大きな理由の一つとして、対象となる人間には機械やシステムとは異なり、例え物理的に同様な状況にあったとしても、感情や加齢などの様々な背景に由来する心理特性における個人間、個人内での変動(差)が見られることがあげられる。例えば、ある機械を開発して商品として売り出そうとするときに、ある年齢層では大変好評な使用感がデータとして得られたにもかかわらず、別の年齢層ではそれほど高い評価が得られなかった、ということはよく見られることであるし、製品開発者やシステム評価者が研究を始める前に「このような結果が得られるのではないか」とあらかじめ準備していたこととは全く逆の結果が最終的に得られたということもあるかもしれない。近年になって、工学分野や情報分野において従来は主にハードウェアの側面から検討されてきた多くの問題が、人間がもつ様々な特性と直接的・間接的に結びついており、本来の目的である製品やシステムの開発や評価にとって欠かせないものであることが広く認知されるようになった。それにつれて、人間の心理特性をどのように扱うべきであるのか、また何らかの手法でもって計測した心理データをどのように分析していけば目的を達することができるのかという関心が様々な研究領域で高まっている。

心理学は人間や動物の行動に関する科学としての一領域である(例えば、高木, 1958)¹⁾。科学としての心理学において注意を払うべきことの一つは、その研究によって得られた心理データは妥当な手法で得られたものであるのかということと、そして何らかの結論を得られた場合に、その結論は適切な手法でもって分析されて導かれたものであるのかどうかという点である。もちろん、このような心理データを扱う実務者の多くは採用された研究手法の妥当性を検討する専門家でもなければ、得られたデータを統計的に分析する専門家でもない場合がほとんどであろう。また、専門家であっても、実際に心理計測を行う場合、または得られたデータを分析する際には心理データを扱う場合に特有の慎重さが求められる。すなわち、どのような手法であれ、心理計測を行う場合には、その対象はほとんどの場合が人間であるため、研究手法や分析手法の適用に倫理的問題がなく、かつ妥当性が認められることを十分に吟味する必要がある。

実証的なアプローチによる心理学は、その成立のときから現在に至るまで人間が行う様々な精神的な営みについての可視化を実現しようと試みてきた。換言すると、目に見えない「こころ」の働きを数値に置き換え、そのなかから一般的な法則を見いだしていくことを研究の主眼としてきたともいえる。そのため、このような心理的特性に関する研究の場においては、どのようなデータが扱われているのかという点や、実験・調査・観察を主な軸としたデータに基づいた研究手法とそこで得られたデータの分析手法について、ある程度の共通した理解

が求められることが一般的である。昨今では、心理学がかかわる研究領域も増大の一途をたどり、心理データを用いて製品やシステムなどにかかわる基礎的・実務的な研究をする機会も多くなっている。そこで、本章では人間がもつ様々な心理特性（感覚・知覚・認知・情動・判断・感性など）を調べる必要が生じた際に、どのような点に注意を払うべきであるのかという点を見ていくこととする。

【本章の構成】

本章は三つの節から構成されている。まず 2-1 節として、「心理学によるアプローチ (2-1)」を概観し、それらのアプローチを踏まえたうえで、2-2 節として「心理学における主な研究手法 (2-2)」について述べる。最後に、2-3 節として、それらの研究手法を用いて得られたデータについての分析手法、すなわち、「心理学におけるデータの分析 (2-3)」について、データの種類によって異なる分析手法が採用される点に触れることとする。

本章の大きな目的は、研究・分析手法についての詳細や具体的計算についての記述を目指したものではなく、データの扱いについての心理学的手法とその特徴の一部分を概観して説明することにある。そのため、研究手法や扱おうとするデータの特性によっても多岐に渡る分析手法や値の算出手法の詳細などについては表面的な記述に留めたものとなっている。必要に応じて末尾の文献を参照することで補って頂きたい。

■引用文献

- 1) 高木貞二，“心理学研究法，”岩波書店，1958.

S3 群 - 10 編 - 2 章

2-1 心理学によるアプローチ

(執筆者：木村貴彦・篠原一光) [2011 年 3 月 受領]

本節では心理データを扱う研究を行うために採用し得る複数のアプローチとそれらのアプローチにおける特徴について見ていくこととする。

2-1-1 探索的な研究と仮説検証的な研究

心理学に関係する研究が行われる場合に、研究の種類 (type) として、探索的な研究 (exploratory study) と、仮説検証的な研究 (hypothesis verification study) に区別されることがある。探索的な研究では、あらかじめ何らかの仮説を用意することをせずに研究が進められることが一般的であり、まさに「探索的」に心理的な現象、すなわち観察可能な種々の行動 (行為や発話、質問への回答、生理的反応などのこと) における様々な条件を明らかにしようとする試みとされる¹⁾。したがって、例えば研究の初期の段階で探索的な検討を行うような場合には、試行錯誤的に条件を変化させ、そのような条件の違いに対して行動がどのように変化するかを観察するということが行われる²⁾。また、問題を設定するために要因を整理しようとする場合や、予備的な位置づけとしての研究としても扱われることもある³⁾。

他方、仮説検証的な研究では、すでに得られている知見や事実を基にして仮説を提出し、その仮説について実験や調査などの研究によって検証 (verification) していくことで研究を進めていく。したがって、ある仮説の真偽を確かめようとする場合には、検証可能なかどうかという検証可能性 (verifiability) や、仮説が誤っているという帰結を得ることが可能かどうかという反証可能性 (falsifiability) について考慮することも必要となる。もしも反証ができないことが明らかである場合、仮説の誤りを明らかにできないということと同時に、正しいことも明らかにできないことを示していることとなる。また、研究を進めていくうえで提出される仮説のことを作業仮説 (working hypothesis) と呼ぶこともある。そして、現代の心理学において重要視されている点として、行おうとする研究が探索的であり、仮説検証的であり、何らかの考察を進めていくうえでデータを用いることがあげられる⁴⁾。

2-1-2 質的データと量的データ

心理状態を計測する際に重要となるものにデータがあることは前項で述べたとおりであるが、データには質的データ (qualitative data) と量的データ (quantitative data) の二つがある。質的データの場合に扱われるデータとは、文字などが中心となり、量的データの場合に扱われるデータは数値によって表される。また、詳細については次節で説明するが、質的データには名義尺度と順序尺度という水準があり、量的データには間隔尺度と比率尺度という水準がある (表 2・1)。

心理学的な事象についての何らかの関心を検討するうえで心理データを得るために用いられる手法として、例えば個人的な経験を主観的に記述していく場合のような、主として質的データを用いる場合を質的方法 (qualitative method) とし、他方、現象や経験について、数量的に説明するための量的データが主に用いられる場合を量的研究 (quantitative method) としている⁵⁾。このとき、前者を定性的研究、後者を定量的研究と呼ぶ場合もある。

表 2・1 質的データと量的データ

データの種類	質的データ	量的データ
尺度	名義尺度 順序尺度	間隔尺度 比率尺度
研究方法	質的研究	量的研究
	面接 観察 事例	実験
例	日記など生活の中での記録 面接などでの言語の記述 エピソード	長さ パーセンテージ 大きさ

質的方法（定性的方法）が中心となる手法は、面接・観察・事例研究といった人間の営みや行動のなかでも数値として表現することが困難な側面を検討しようとする際に有効なものといえる。また、量的方法（定量的方法）が中心となる手法として実験法があるが、実験法では人間の行動を定量的に表現し、何らかの法則や規則性を見いだそうとする試みが多くみられる。どちらの研究手法を用いる場合であっても、心理計測を実際に行おうとする際には、得られるデータのもつ意味について留意すべきなのはいうまでもない。

2-1-3 横断的研究と縦断的研究

横断的研究（cross-sectional study）と縦断的研究（longitudinal study）の最も大きな相違点は、収集されるデータの出所ともいえる標本（sample）を抽出する際の扱いについて、どのように標本を選択するかという点である。すなわち、例えば発達に関する研究を行おうとする場合（表 2・2）、横断的研究においては、ある年齢の範囲内で複数の集団に対して調査や観察などを実施してデータを収集していく。すなわち、ある小学校に在籍する 1 年生から 6 年生までの児童に対して同一の調査を実施してデータを得るといったものである。このような手法であれば、比較的短期間に研究を行うことが可能であり、なおかつ多数のデータを収集して異なる集団の比較を行うこともできる。ただし、横断的研究の場合には、学年の違いによる結果の相違についての言及は可能となるが、それぞれの個人がどのように発達の変化を見せるのかという点を扱うことはできない。

表 2・2 横断的研究と縦断的研究の例（高木，1958 をまとめたもの）

	横断的研究	縦断的研究
研究内容	小学生の読字力の発達についての研究	
研究機関	比較的短期	長期
集団	ある年の小学 1 年生から 6 年生	ある年に入学した児童
データ	学年別に収集	個人に反復して収集

それに対して、縦断的研究では同一の集団において、長期間に渡って成長の変化を追跡してデータを収集していく方法である。この場合、どちらかといえば比較的小規模な集団が研究対象となることが多い。先ほどの小学校の例でみた場合、その小学校にある年に入学した小学生について、卒業までの 6 年間を通じて反復してデータを収集することがそれに当たる。

縦断的研究の手法の場合には、個人の年齢を契機として反復した測定を実施するため、それによって得られたデータはそれぞれの個人の発達による変化をとらえることが可能となる。そのため、個人内でどのような発達がなされるのかについても言及可能となり非常に貴重なデータとなり得るが、その一方で研究を完全に実施していくためには大きな労力を伴う。

2-1-4 心理学研究における制約

心理学は、目に見えない心の働きをシステマティックに検討するための様々な手法を工夫して編み出してきた歴史を有するが、その一方で、ある意味で人間の極めて個人的な部分である心を扱うことに伴う制約も確かに存在している。

まず、個人内・個人間でのおぼろつきについては、科学としての心理学が成立してから今日に至るまで多くの試みがなされてきた。心の働きは置かれた環境やその日の体調といった、様々な細かな要因によって変動することもあるし、同じ内容の実験や調査を繰り返し行った場合であってもその研究での参加者が異なるなどすると、前回とは異なる結果が得られる場合もみられる。人間が営む様々な活動でみられるこのような変動 (variation) については、後述する剰余変数 (extraneous variable) を除く努力をしたうえそれが偶然によるものなのか、あるいは独立変数 (independent variable) によるものなのかを統計的に判断するなどの手続きが行われる。また、そのような配慮のもとに得られた知見であったとしても、あらゆる状況で普遍的に採用できるものであるかどうかという点では問題が残る場合も多く、Eysenck は『心理学におけるあらゆる理論と研究は、幾つかの点で限定されているものであることを忘れるな (p15)』⁹⁾と述べている。これには例えば実験室内で得られた知見をそのまま実際の場面における行動の原理として扱って良いかどうかという問題や、大学生を対象として調査を行った結果が、大学生以外の集団に適用可能かどうかといった例があげられ、心理計測の際に得られたデータをどの範囲まで適用可能であるのか見極める必要性を示唆している。

更に、心理学的手法を用いた研究を行うにあたっては、研究者が望むいかなる内容の研究を自由に実施して良いわけではない。研究にあたっての制約として研究開始時や論文執筆時に問題となる点の一つに、研究についての倫理的配慮 (ethical considerations) がある。ここで言及する倫理的配慮とは、剽窃などの類ではなく (それらも重要な問題であることはもちろんである)、主に人間や動物を研究対象とすることに由来するものである。すなわち、恐怖や痛みが伴う研究や、人格への影響や後遺症が考えられるような研究など、例え過去に行うことが可能であった内容であっても今日では検討することができないという場合もみられる。そのような許容できない内容を判断するうえでは、研究者の所属機関が中心となつての倫理審査を受けることや、人間の場合にはインフォームド・コンセント (informed consent) を得ること、動物の場合には不必要な痛みを伴わないなどケアに注意することなどの点が求められており、American Psychological Association (APA : アメリカ心理学会) が発行するマニュアルにも詳細が記載されている⁷⁾。これらの点については、研究における制限といったネガティブなものとして扱うのではなく、心理データを取得する上の基本的なルールとしてとらえ、それぞれの研究機関においても関係の議論を深めていく必要がある。

■引用文献

- 1) Christensen, L. B. “Experimental methodology,” Allyn and Bacon Needham Heights, MA, 1994.
- 2) 高木貞二, “心理学研究法,” 岩波書店, 1958.
- 3) 加藤 司, “心理学の研究法 実験法・測定法・統計法,” 北樹出版, 2008.
- 4) 南風原朝和・市川伸一・下山晴彦, “心理学研究法入門,” 東京大学出版会, 2001.
- 5) Ray, W.J., 岡田圭二 (訳), “エンサイクロペディア 心理学研究方法論,” 北大路書房, 2003.
- 6) Eysenk, M.W., 山内光哉 (監修), “アイゼンク教授の心理学ハンドブック,” ナカニシヤ出版, 2008.
- 7) American Psychological Association, “Publication Manual of the American Psychological Association Sixth edition,” American Psychological Association, 2009.

S3 群 - 10 編 - 2 章

2-2 心理学における主な研究手法

(執筆者：木村貴彦・篠原一光)

心理計測のため研究手法は、近年になって、機器やインタフェースの使いやすさや快適性、ユーザからの受け入れやすさといった様々な評価を行うときに有効な手段として用いられるようになった¹⁾。これは、人間の心理計測について多くの手法を工夫してきた心理学が、実際の生活やモノづくりにとっても大きな貢献が可能であることを意味しており、前掲したほかにも製品の印象や好き嫌い、購買意欲の高低といったように、通常の物理計測の範ちゆうだけではとらえきることが難しい内容を計測しようとする動機で行われることも多い。このようにときに用いられるのが尺度構成法 (scaling method) である。尺度構成はすなわち、目に見えない心の働きを何らかの規則に従って並べていく作業であり、その結果様々な用途を目的とした心理データを収集することができる。このような尺度構成のための方法としては、精神物理学的方法 (psychophysical method) や評定尺度法 (rating scale method)、心理検査法 (psychological testing method) といった手法がある。

このような心理データを得るために用いられる代表的な研究手法としてあげられるものが、(1) 実験 (experiment)、(2) 調査 (survey)、(3) 観察 (observation) である。これらの研究手法を研究の目的に応じて、単独で、あるいは組み合わせて心理計測を行うことが多い。それぞれの手法に長所と短所があり、いずれの方法が適切であるのかを吟味したうえで用いる必要があることはいうまでもない。

2-2-1 実験による手法

心理計測における実験的手法は、一般的な心理学に対するイメージとは異なり、科学としての心理学の幕開けと共に用いられてきたものである。実験を実施する際の根幹となる考え方の一つとして、実験計画 (experimental design) があげられる。これは、実験において実験者が調べようとする内容について、様々な実験変数の影響を排除し、検討したい要因のみを導入してその影響を検討できるよう計画・実施されるもので、実験計画に不備がある場合、当然結果に影響することになる。したがって、心理計測のために実験的手法を用いる場合には、実験的手法についての訓練を受けた人間が実施することが望ましい。通常、実験的手法においては厳密な実験計画のもとに、原因と結果、すなわち因果関係を明確にしていけることが可能となり、この点はほかの研究手法と比較したときの大きな特徴といえる。このとき、原因となる変数のことを独立変数 (independent variable)、結果となる変数のことを従属変数 (dependent variable) と呼ぶ。実験者 (experimenter) は独立変数を操作し、実験参加者 (participant) が示した反応 (response) による従属変数の変化を組織的に検討していく。例えば、明るさの違いが作業速度に影響するかどうかを検討する場合を例に考えると、実験者が設定するのはおそらく「明るい・暗い」という違いであろう。このような違いのことを条件 (condition)、あるいは水準 (level) と呼び、「明るい・暗い」という二つの条件をまとめたものを要因 (factor) とする。設定された二つの条件の下で、実験参加者の作業速度を計測し、明るい場合と暗い場合における違いを計測していくことになる。ところで、このような

ときに問題となるのが剰余変数 (extraneous variable) と呼ばれる変数である。すなわち、実験を行う場合に、実験者が検討したいのは、本来は独立変数として設定した明るさの要因であるにもかかわらず、それ以外の要因 (剰余変数) が結果に影響を及ぼす可能性である。例えば、明るさを独立変数として用いているときに、周囲の音 (静かな場合と騒々しい場合) に違いが見られたというときに、それらが剰余変数として作業速度に影響を及ぼしている可能性を否定できないことになる。このような剰余変数による影響をできるだけ排除するために行われるのが統制 (control) である。もし、例にあげたような、音による剰余変数を排除したい場合には、例えば防音室を利用するなど、音に関する影響が一定となる環境のなかで実験を行うなどの対応がなされ、これを恒常化という²⁾。

また、実験計画を行うに当たっては、用いる計画に対してどのように実験参加者を割り当てていくのかについての考慮も求められる。それが実験参加者間計画 (between-participant design) と実験参加者内計画 (within-participant design) である。これは、データを得る際に、個人内の変動と個人間の変動という剰余変数となり得るものについて配慮することを意味しているといえる。例えば 10 名の実験参加者が得られた場合に、その実験において設定されたすべての条件について実験参加者が経験するような実験計画が実験参加者内計画とされる一方、例えば、明るい条件で 5 名、暗い条件で 5 名といった具合に条件ごとに実験参加者を割り当てるといった実験計画が実験参加者間計画とされる。どのような場合にどちらの実験計画を用いるのかを判断することは、行おうとする実験の内容によって選択する必要がある。記憶や運動が関連するような課題の場合、実験参加者内計画によってすべての条件を実験参加者が行うことになると、後に行う条件が前に行う条件の影響を受ける練習効果がみられるかもしれない。なお、このように、条件の順序が結果に影響を及ぼすことを順序効果 (order effect) というが、順序効果を取り除くために一般的に用いられる手法がカウンターバランス (counterbalance) である。これは、例えば A 条件と B 条件の 2 条件があったときに、A 条件から始める実験参加者と B 条件から始める実験参加者に区別することで、結果に対する条件の順序の影響を相殺しようと試みるものである。また、全く異なる属性をもつような実験参加者をいくつかのグループに分け、実験参加者間計画を用いる場合には、その属性の違いが剰余変数として結果に影響を及ぼす可能性もある。更に、実験によっては複数の要因を用いる場合もあるが (例えば、明るさと音の 2 要因など)、そのようなときに、例えば性差を扱う場合のように、実験参加者内計画と実験参加者間計画の両方が用いられる実験計画もあり得る。

次に、生態学的妥当性 (ecological validity) の高い実験をどのようにして実現するのかを実験者は考える必要がある。実験実施の際には厳密に条件を統制することが求められるため、実験室内での実験が選択されることが多いが、実験室内という特殊な環境での人間の行動が、実際の人間の行動と同じものとみなしてよいかには常に疑問が付きまとう。したがって、実際の人間の行動についての、どのような側面を検討しようとしているのか考えてみることも必要である。フィールド実験には常に剰余変数の影響が存在する可能性が高くなり、目的によっては必ずしも妥当とはいえないかもしれないが、製品評価やユーザビリティについて検討する場合には、例え実験室内の実験を行おうとするときであっても実際の行動との関連に目を向けて条件を設定することを忘れるべきではないだろう。

2-2-2 調査による手法

心理学における研究では、ある学年や特殊な集団に対して実施されるような小規模な調査から、ある地域の住民全体といった大規模な集団に対して実施するような調査まで、極めて多種多様な形態によって調査が行われてきた。このような調査では、質問紙 (questionnaire) が用いられるのが一般的である。実験法による検討をしていた場合であっても、大抵の場合は実験参加者による内観 (introspection) やデモグラフィックな属性、更には、その実験に関する簡単なアンケート結果 (例えば、課題の困難度や、その実験に特化した内容: 運転が関係する実験の場合には免許の有無や事故経験など) が記録されることも多く、その意味では質問紙を用いた調査的な手法を併用しているともいえる。

本来、調査による手法を用いてデータを収集しようとする場合、対象となる母集団におけるすべての調査対象者 (participant) から目的に応じたデータの収集することが可能であれば望ましい場合もあり、このような手法のことを悉皆調査と呼んでいる。しかしながら、実際にはそのような方法を取るだけの時間的・金銭的・人的コストをかけることができず、そのために母集団のなかから何らかの方法でもって標本を抽出することになることが多く、これを標本調査という。質問紙によるデータ収集は、全く同じ内容のものを多数の調査対象者に対して同時に実施することが可能であり、ある種の目的のためには効率的ともいえる。しかしながら、質問紙の作成上の問題や、調査対象の偏りなど、データ収集上の問題がある場合、得られたデータは、事実をゆがめたものや過大・過小評価したものとなる危険性もあるために注意しなければならない。すなわち、質問紙によって測定されたデータが正確なものであるためには、『精密な統計分析による適切な質問項目の選定と項目間の関連性を明らかにした上での尺度構成の明確化 (p166)』³⁾が必要となる。このことは、質問紙を用いた調査による手法によって心理計測を行う場合、質問紙に含まれる項目をどのように設定し、それぞれを組み合わせた結果としてどのような分析を行っているのかをあらかじめ検討していく必要性を示唆している。例えば、質問項目内での言葉の使い方 (ワーディング) には細心の注意を払う必要がある。なぜなら、仮に、質問紙を作成して新製品の使いやすさに関する評価に用いようという場合、質問紙の項目の作成の仕方によっては、使いやすい、使いにくいという判断について、調査実施者の意図にかかわらず調査対象者からの回答をある方向に誘導してしまう可能性があるためである。また、一つの質問項目の文中に二重の意味を含むようなものは、調査対象者の回答がどちらの意味によって導かれたものであるのかが判断できないため避けなければならない。更には、質問項目の回答方式をどのように設定するのかという問題も重要である。例えば、用意された項目 (例えば、バナナ・リンゴ・ブドウ・イチゴ・ミカン) のなかから調査対象者が一つだけを選択するような場合は単一回答法と呼ばれるが、該当するものをすべて選択するような場合には複数回答法と呼ばれる。このほかに、用意された項目に順序づけをするような順位法や、それぞれを組み合わせて比較していく一対比較法 (method of paired comparisons) などの手法があり、調査対象者が質問紙に回答する際の負担や、調査者がどのように分析を進めていこうと考えているのかといった点から決定していく必要がある。

また、調査による手法は一度だけの実施で終える場合だけではない。一度だけの実施の場合には、あるタイミングにおける対象とした集団の特徴を記述することになり、横断的な性

質をもつ調査といえる。それに対して、時間において同じ質問紙を用いた調査が複数回行われることがあり、これは縦断的な性質をもつ調査といえる。このとき、同じ集団に対して同じ質問紙を用いて調査を実施する形式のパネル調査と、一つの集団内から抽出された異なる人を対象として調査が行われる形式の繰り返し調査がある。このような手法を用いれば、複数の時点におけるデータを収集することができるため、研究したい内容についての心理的な変動を検討することが可能となる。

2-2-3 観察による手法

研究を行うにあたって、観察 (observation) は極めて重要なものである。心理学の目的の一つには、人間の理解があるが、人間行動を観察することによってその目的に接近していくことは有効な手段といえる。Wundt などの古典的な心理学において用いられていた内観 (introspection) も自分自身の内面における心の働きを訓練された実験参加者が言語化する試みとも言え、その点では観察的な意味合いも含まれている。観察法がもつ魅力的で大きな特徴として、研究の初期の段階において、観察対象がどのような行動をとるのかといった現象的な側面を記述して探索的に検討していくことが可能な点があげられる。そのようにして得られた現象についての記述に基づいて、研究を次のステップに発展させていくことは、研究の進め方における一つの方法である。しかしながら、科学的見地からの観察とは、目的や訓練なく行われるものを指さないのが一般的である。

観察による手法には、統制が設けられず、観察対象が日常的な行動の場としている場面で実施される自然観察法 (natural observation) と、観察者が研究の目的に沿って条件についての操作を行ったような状況における観察対象の行動を記録するような実験的観察法 (experimental observation) に区別することが可能である⁴⁾。自然観察法であれば、観察対象の全く通常の行動を記録していくことが可能となる点は長所といえるし、実験のような手法で扱うことが不可能な対象を研究することが可能なこともある⁵⁾。しかしながら、その一方で、研究の目的において、例えば生起する頻度が低いものを調べようとしているときなどにおいては長時間観察を継続する必要性が生ずるなどの問題が生じることもある。また、実験的観察法の場合の大きな問題点としては、研究者が設定する場面が実際の状況とはそぐわないような事態で観察を行う危険性があげられる。このことは、生態学的妥当性が低い状況で観察を実施しようとするものであり、実験室での制限的な環境で実施される実験を日常生活における知見としてどこまで拡大可能かという、実験的手法による研究と共通した問題点ともいえる。

具体的な観察のための手法としては、時間見本法 (time sampling method)、事象見本法 (event sampling method) などがあげられる。時間見本法の場合、あらかじめ設定された時間間隔ごとに観察の対象となる行動が生じたのかどうかを記録していくものである。一方、事象見本法の場合には、時間ではなく行動そのものが直接的に記録の対象となり、行動が生じたときの一連の流れを記録していき、最終的にどのように帰結していくのかを観察していくものである。

さて、実際に観察を行ううえで大きな問題として、観察を行うことそのものが観察対象の行動に影響を及ぼす可能性が考えられる。観察場面において、観察者の存在は基本的には異物であるわけで、乳幼児や動物などが観察の対象である場合には観察者がいることで普段と

は異なる行動を示すこともあり得る。そのような事態を回避するためには、例えば観察に入る前に観察者の存在に観察対象が慣れるまで時間を置くなどの対処や、集団の一員ではない第三者として観察を行う非参加観察法ではなく、集団の一員として参加したうえで観察を実施する参加観察法の採用などが考えられる。ただし、参加観察法を採用した場合には、観察者自身も対象となる集団の一員であることに関連し、観察者がもつ先入観や研究上の期待などで観察したデータ（事実）が変容する可能性に注意することが必要となる。

適切な場面を選択したうえで観察を行ったときに求められることは、観察された現象について、どのように記述していくのかという点であり⁹⁾、換言すると、どのようなサンプリングを行うのかを決定することといえる。質的な研究方法ととらえられがちな観察法であるが、定量的なデータの収集と分析も可能である。例えば、行動目録法や評定尺度法を用いれば、あらかじめ作成されたチェックリストに行動の生起をチェックしていくことで対象となる行動が生起する頻度を数値的にとらえることも可能であるし、評定尺度に記録していくことで観察された行動の背景にある要因について考えることも可能である。

■引用文献

- 1) 熊田孝恒・遠藤信貴，“心理量の計測,” 計測と制御, vol.42, pp.1039-1043, 2003.
- 2) 森 敏昭・吉田寿夫，“心理学のためのデータ解析テクニカルブック,” 北大路書房, 1990.
- 3) 心理学実験指導会，“実験とテスト=心理学の基礎 (解説編),” 培風館, 1985.
- 4) 澤田英三・南 博文，“質的調査 観察・面接・フィールドワーク,” 南風原朝和・市川伸一・下山晴彦 (編) “心理学研究法入門 東京大学出版会” pp.19-62, 2001.
- 5) Christensen, L.B., “*Experimental methodology*,” Allyn and Bacon Needham Heights, MA, 1994.
- 6) 高野陽太郎・岡 隆，“心理学研究法 心を見つめる科学のまなざし,” 有斐閣アルマ, 2004.

S3 群 - 10 章 - 2 章

2-3 心理学におけるデータの分析

(執筆著者：木村貴彦・篠原一光)

2-3-1 心理学で用いられるデータ

心理計測を行う場合、様々なかたちでデータを収集するが、それらのデータには表 2・1 のとおり、大別して質的データと量的データの二つがある。データは無目的に集められるものではなく、何らかの意図をもって系統的に集められるものであり、集められたデータについては分析 (analysis) が行われる。更に、数字によって表すことができる量的データには、人数やサイコロの目のように間に取り得る値がない離散変量 (discrete variable) といわれるものと、時間のように基本的にどの位置でも数値を示すことが可能な連続変量 (continuous variable) といわれるものがあり、その場合、値は無限に小さくしていくことが可能になる。ただし、心理計測を実施する際に、仮に連続変量を用いる場合であっても、どの単位までの値を用いるのかは研究の目的に応じて検討すべきである。

量的データを扱うことで比較や統計の手法の適用などが容易になり、現象の記述を行うときの利便性は向上する。その一方で、データをまとめて代表値として表現する傾向も強まるため、ある現象が有する特殊な特徴や特異な例については適用の範囲外となることもあり、その意味では質的データに優位性がみられる場合もある。また、研究によって得られたデータは、場合によっては変換 (transformation) したうえで用いられることもある。テストにおける点数が全体のどの程度の位置にあるのかを表現するための偏差値なども代表的な変換といえるが、そのほかのデータの変換として、対数変換 (logarithmic transformation) や逆数変換 (reciprocal transformation)、逆正弦変換 (arcsine transformation) などがある。これらの変換を用いる目的として、得られたデータと心理量との関係づけのためと、正規分布に従うなど統計的な処理を行うときの前提を満たすための 2 点をあげており、このとき、前者が重視されるべきであることが指摘されている¹⁾。

心理学で用いられる数値によるデータを区別すると、質的データとして名義尺度 (nominal scale) と順序尺度 (ordinal scale)、量的データとして間隔尺度 (interval scale) と比率尺度 (ratio scale) として区別される。ここでの尺度 (scale) とは、心理計測におけるある種の規則を表しているものであり¹⁾、高い水準をもつものほどデータを扱う際の範囲が広くなる。すなわち、尺度に水準が規定されていることによって、それぞれの尺度水準において得られたデータに特有の特性が見られ、その結果、適用できる分析手法にも違いが出てくることになる。

名義尺度とは、数を割り当てられてはいるが、分類や記号としての意味しか有していない。例えば、郵便番号や学籍番号などがこれに当たるものである。基本的にどのような数値を当てはめることも可能であるし、そもそも数値ではなく、何らかの記号を当てはめることも用足りる。例えば、車のナンバープレートは日本では数字であるが、海外では英字を用いられていることも多い。このような特性があることから、名義尺度を用いて和や差を求めても意味がない。順序尺度では、何らかの順序を表すために用いられるものであり、そのため、大小の関係を示すことが可能である。成績を並べて評価するときなどがこれに該当する。しかしながら、順序には意味があるものの、和や差といった違いについては示すことができない。仮に、成績の順番についての差分を取ったとしても、順序の違いを示しているだけで、

差そのものを扱うものではない。間隔尺度の場合、順序のみならず、その差や和にも意味をもたせることが可能である。したがって、値と値の間隔を取ったときに、それらの単位は同一のものでなければならない。間隔尺度としては、摂氏や華氏の温度が該当する。なお、心理学において用いられるデータでは間隔尺度に該当するものも多く、算術平均をはじめとして多くの分析に用いることが可能となる。比率尺度の大きな特徴は、原点である 0 が定まっていることである。長さや重さを計測したときのデータでは、0 が決められており、比率尺度の例の一つといえる。

2-3-2 記述統計

何らかの手法によって心理データが収集されたのであれば、それらを分析していかねばならない。分析をしないままのデータは、それだけでは傾向も特徴も判然としない場合が多く、“生のまま”であるという意味でローデータ (raw data) と呼ばれる。ローデータそのものはデータのまとまりであるにすぎず、それをを用いて議論をすることもままならず、研究の俎上にのせ、方向を見極めるためにはどのような分析を行うのが大きなポイントとなる。このように、収集されたデータを要約して整理していく作業のことを記述統計 (descriptive statistics) と呼ぶ。

記述統計の基本の一つとして、収集されたデータがどのように散らばっているのか知ることがあげられる。心理データの場合には個人間、個人内を問わず、計測ごとの違い (差) が見られることも多く、データの散らばり、すなわち分布 (distribution) をみることは重要な点である。データの分布を明らかにするために用いられる方法としては、度数分布 (frequency distribution) があげられる。このとき、何らかの値や、範囲についてデータがいくつ該当しているのかを数え、それを度数 (frequency) として用いる。また、度数分布をグラフ化する一つの表現方法として、ヒストグラム (histogram) と呼ばれるものがあり、横軸に値や範囲をとり、縦軸に度数を示すことでデータの分布を表現することができる。なお、度数には計測された度数を直接用いる場合と、収集された全体のデータとの比として表される相対度数 (relative frequency) が用いられる場合がある。

次に、収集された心理データを整理し、何らかのかたちで表現したい場合に用いられることが多いものが代表値 (average) である。代表値を用いる目的は、それによって集められたデータの特徴や傾向を容易にとらえることである。心理学において、最もよく用いられる代表値としては平均値 (mean) があげられる。平均値には算術平均 (arithmetic mean) や幾何平均 (geometric mean)、調和平均 (harmonic mean) などがあるが、間隔尺度にも適用される算術平均が一般的であり、 \bar{X} や M して表される。このように、代表値としてすぐにあげられるものが平均値であり、間隔尺度や比率尺度での適用にも耐えるために分析手法も広がるが、データ分析の際にはそれ以外の代表値の利用が適切となる場合もある。その例が最頻値 (mode: Mo) や中央値 (median: Me) である。いずれも代表値の一つであるが、最頻値はその名のとおり、最も度数が多く見られるデータの値であり、名義尺度の場合にも用いることが可能である。また、収集されたデータを順番に並べたときに、その中央に位置する値が代表値として用いられるのが中央値である。いずれの代表値を用いれば適切であるかを決定するために、先述した度数分布の形状を目安とすることが多い。例えば、絶対的な基準はないとしたうえで、歪度 (skewness) と呼ばれる、左右対称の分布であるか、あるいは左や右に

歪んだ分布であるのかといった、得られたデータによる分布についての歪みを示す指標でもって検討することが指摘されている¹⁾。

更に、例えば異なる集団の代表値を計測したときに、その代表値が同じ値を示していればそれらの集団の特徴が同じものとして良いのかどうかをとらえようとする場合に検討すべきものに散布度 (dispersion) がある。散布度が意味するものは、データがどのように散らばっているのかということであり、例えば異なる二つのクラスで身長を計測した時に、170 cm の平均が得られた場合、一つのクラスでは平均値の近くの範囲でデータが散らばっており、もう一つのクラスでは平均値から遠くの範囲でデータが散らばっているのであれば、それらの二つのクラスは平均値が同じであっても散らばりが異なる特徴をもつ集団としてとらえることが適切ともいえる。このような、データの散らばりを表現するための代表的な指標が分散 (variance: s^2) や標準偏差 (standard deviation: SD)、標準誤差 (standard error: SE) と呼ばれるものである。算出された分散についての正の平方根が標準偏差とされ、母集団から抽出された標本による分布から求められる標準偏差が標準誤差である。いずれもデータがどのように散らばっているのかという、収集されたデータの分布を表現するための指標として用いられる。なお、心理計測を行ったときに得られるデータに基づいた分布は確率分布 (probability distribution) の一つである正規分布 (normal distribution) に従うと仮定されていることが多い。標準偏差や標準誤差の場合、それぞれのデータが平均値からどの程度離れているのかを示し、それらの値が大きければ、その集団に属するデータの散らばりが大きいことを意味する。

心理学的手法を用いて収集されたデータについての散らばりを表す散布図 (scatter diagram) を示したときに、多くの場合、縦軸と横軸にそれぞれ独立変数をあてはめることでそれぞれの変数の関連性をみようとする。このような関連性を相関 (correlation) という。相関にはその形状から、一方の変数が大きくなるに従って他方も大きくなる正の相関 (positive correlation)、逆に一方の変数が小さくなるに従って他方は大きくなる負の相関 (negative correlation)、二つの変数によるデータの散らばりに関連性が見られない無相関 (uncorrelation) があり、変数間における関連性の程度によって相関の強さが表される。このような二つの変数の間における相関の強さについて、どの程度直線的であるかを明らかにするために求められるのが相関係数 (correlation coefficient) である。変数が間隔尺度や比率尺度である場合、ピアソンの積率相関係数 (Pearson's product-moment correlation coefficient) が用いられることが多い。

2-3-3 推測統計

推測統計 (inferential statistics) が記述統計と大きく異なる点は、得られた標本をもとにして、その標本が抽出された母集団 (population) についての推測を行うことを主目的としていることである。これは、得られた標本についての特徴を表現しようとする記述統計と比べて、より一般化した知見を得ようとするものともいえる。すなわち、推測統計においては母集団からある方法によって抽出した標本がもつ平均値や分散などの値を用いることによって、母集団がもつそれらの値を推測する作業を行われることとなる。ここで、推測という用語が持つ直接的な意味としては、標本が持つデータから母集団のデータを確率的に議論できるように様々な手法を用いるということである。このときの様々に用いられる手法のことを統計的検定 (statistical test) という。そして、統計的検定のために用いられる最も基本的な事項が、

帰無仮説 (null hypothesis) と対立仮説 (alternative hypothesis) である。帰無仮説とは、 H_0 と表記されるもので、例えば、A 条件と B 条件の間に差があるかどうか、ということを検討しようとする場合に、「A 条件と B 条件の間に差がない」とされる仮説のことを指す。そして、この帰無仮説が棄却される場合に用いられるとき、つまり「A 条件と B 条件の間に差がある」とされる仮説を H_1 として、対立仮説と呼ぶ。つまり、統計的検定を行うときには、まず帰無仮説を立て、その帰無仮説を棄却することで条件間に差があることを示すことになる。なお、ここで混同すべきではないのは、実際に研究を行ううえで用いる作業仮説 (working hypothesis) と統計的検定で用いられる仮説とは必ずしも同一ではないことである。これは、作業仮説が実験や調査などの研究を行おうとする際に立てられるものであるのに対して、帰無仮説はあくまでも統計的な分析を行っていく過程において便宜上用いられるためである。

次に、帰無仮説の採択を決めようとする際には何らかの基準が必要となるが、そのような基準のことを確率的な表現として有意水準 (level of significance) あるいは危険率として示され、心理学では 5% (0.05) をその値として用いることが一般的である。つまり、有意水準として設定された 5% の基準に照らして帰無仮説を棄却するかどうかの判断がなされる、棄却できるという判断が行われたときに条件の間に有意 (significant) な差が見られたことになる。他方、棄却ができない場合には有意な差がない (nonsignificant: *n.s.*) とされる。しかしながら、得られた結果はあくまでも確率的な範囲に留まるため、実際には仮に帰無仮説が棄却された場合であっても、依然としてその棄却の判断には誤りである確率が残されており、これを間違いの種類によって、第 1 種の過誤 (type I error)、第 2 種の過誤 (type II error) という。第 1 種の過誤の場合、実際には帰無仮説が正しいにもかかわらずそれを棄却し、対立仮説を採択してしまうタイプの誤りであり、第 2 種の過誤の場合、実際には帰無仮説が誤っているにもかかわらずそれを採択し、対立仮説を棄却してしまうタイプの誤りである。

このように、心理計測を行って得られた標本においてみられた違いや関連が、標本が属している母集団においても同様にみられるのかどうかを検討する手段が統計的検定である。標本から求められた違いや関連には、偶然によるものと偶然ではない現実のものが混在しているために、それらを区別しようとするものといえる。例えば、標本の代表値として得られた平均値に条件間の違いがみられるのかどうかという検定を行おうとするときに用いられるものの一つに、t 検定 (t-test) がある。この検定では、無作為抽出 (random sampling) されたデータにおける二つの条件の平均値の間に差があるかどうかということを調べることができるが、実験参加者内計画で行われた対応のあるデータであるか、実験参加者間計画で行われた対応のないデータであるかによって、t 値の算出方法や自由度 (degrees of freedom) が異なるため注意が必要である。また、二つの条件がもつ母分散が等質であるかどうかについて調べたうえで、もしも等質ではない場合にはウェルチの検定 (Welch test) などの異なる手法が用いられる。算出された t 値については、それをを用い、t 分布表からあらかじめ設定した有意水準と自由度の組み合わせから臨界値 (critical value) を求め、t 値が臨界値を超えているかどうかを有意な差が見られたかどうかの判断とする。なお、ここでの臨界値とは、分析を行う際に用いられる分布 (この場合は t 分布である) において、帰無仮説を採択するか棄却するかという区別を行うための値を指している。

また、三つ以上の条件については、それぞれの条件を組み合わせ (A 条件と B 条件、A 条件と C 条件、B 条件と C 条件のように) t 検定を繰り返して適用することはせず、ほかの

方法が選択されるが、その一つが分散分析 (*analysis of variance: ANOVA*) である。分散分析では *t* 検定とは異なり、要因についても複数を扱うことが可能であり、更にはそれらの要因が対応ありと対応なしの組合せとなっても適用可能である。しかしながら、あまりに多くの要因を扱うことは計算の複雑化に留まらず、結果の解釈に困難が伴うことも多いことから、研究を実施する際の計画を慎重に行う必要がある。

分散分析は、要因に関係する分散と誤差による分散のそれぞれを区別して求め、それらを用いて平均値における差を明らかにしていく手法である。分散分析を実施する場合には、どのような研究計画でもってデータを収集したのかを前もって十分に理解しておく必要がある。すなわち、要因数と条件数である。要因が一つの場合には 1 要因の分散分析、二つであれば 2 要因の分散分析と呼び、場合によって 2×2 などと要因のなかの条件数が表記されることもある。分散分析では、分散分析表と呼ばれる平方和や誤差が記載された表に基づいて、要因と誤差の変動の比として *F* 値を求め、*F* 分布に従って検定を行うことになる。また、分散分析による検定を行った結果を解釈するうえで理解しておくべき重要なものが主効果 (*main effect*) と交互作用 (*interaction*) である。主効果とは、分析に用いられた要因を構成する条件の間におけるデータの変動であり、交互作用とは、その名のとおり、複数の要因がどのように影響して変動としてみられるかを明らかにしていくものである。

分散分析を行った結果として主効果や交互作用に有意差がみられた場合、次に問題となることは、それでは複数の条件や要因の間でどのように有意な差がみられたのか、という点である。このような分析を行うことが下位検定である。例えば、二つ以上の条件をもつ要因についての主効果が見られたときには、多重比較を行うことでそれらの三つの条件からなる関係において、どのような差が見られたのかを詳細に明らかにすることができるし、二つの要因以上を用いて分析が行われたのであれば、その単純主効果 (*simple main effect*) を検討することによって、それぞれの要因がどのように影響し合っているのかを明らかにする必要がある。なお、文献 1) では、3 要因までの分散分析における計算手法が詳細に説明されているため、必要に応じて参照頂きたい。また、帰無仮説が棄却されるかどうかに関する記述に加えて、検討したい独立変数がどの程度の効果をもつのかを明らかにするためには効果量 (*effect size*) が用いられる。例えば、データの数が十分に大きくなると有意差は得られやすくなるが、このようにして得られた有意差には実際には重要な意味をもたせた主張が難しい。そのため、検討したい独立変数がデータに及ぼす効果の大きさを、効果量を併記することによって明らかにするのである。これについては文献 2) に詳しい。

t 検定や分散分析は、*t* 分布や *F* 分布といった母集団の分布形状を利用して検定を行っていくもので、このような手法をパラメトリック検定 (*parametric test*) というが、心理測定においては必ずしもそのような分布を利用せずともデータを分析することも可能であり、そのような手法をノンパラメトリック検定 (*nonparametric test*) という。

ノンパラメトリック検定は、名義尺度や順序尺度の水準によるデータにも適用することができ、例えば名義尺度の場合であれば χ^2 検定がこれに該当する。 χ^2 検定は、データがあるカテゴリにおいて生起する頻度 (人数や度数) に違いが見られるのかどうかを検定するための手法であり、 χ^2 分布を利用して行われ、実際に得られたデータである観測値と理論上得られるとされる期待値の間における差を明らかにしていく。したがって、帰無仮説として、それぞれのカテゴリでの生起頻度は等しいと設定し、それについて棄却できるかどうかを判断さ

れる。 χ^2 検定は、調査法によって収集されたデータなどをまとめた際の、クロス表におけるデータの頻度に偏りがみられるかどうかを調べるときなどに用いることが可能である。ノンパラメトリック検定は母集団の分布形が分からなくても使用でき適用が容易であるという長所をもつが、一方で検定効率が低いことや、交互作用の処理が容易でないという短所がある。

■引用文献

- 1) 森 敏昭・吉田寿夫, “心理学のためのデータ解析テクニカルブック,” 北大路書房, 1990.
- 2) 大久保街亜・岡田謙介, “伝えるための心理統計 効果量・信頼区間・検定力,” 勁草書房, 2012

■2章における参考文献

- 3) 南風原朝和, “心理統計学の基礎 統合的理解のために,” 有斐閣アルマ, 2002.
- 4) 岩淵千明 (編著), “あなたもできるデータの処理と解析,” 福村出版, 1997.
- 5) 海保博之・楠 見孝, “心理学総合事典,” 朝倉書店, 2006.
- 6) 大山 正・岩脇三良・宮埜壽夫, “心理学研究法 データ収集・分析から論文作成まで,” サイエンス社, 2005.
- 7) Pedhazur, E.J., Schmelkin, L.P., “*Measurement, Design, and Analysis: An integrated approach*,” Lawrence Erlbaum Associates, Inc. Hillsdale, NJ., 1991.
- 8) Searle, A., 宮本聡介・渡邊真由美 (訳), “心理学研究法入門,” 新曜社, 2005.
- 9) 杉山高一・藤越康祝・杉浦成昭・国友直人, “統計データ科学事典,” 朝倉書店, 2007.
- 10) 田中 敏・山際勇一郎, “ユーザのための教育・心理統計と実験計画法 方法の理解から論文の書き方まで,” 教育出版, 1989.
- 11) 吉田寿夫 編著, “心理学研究法の新しいかたち,” 誠信書房, 2006.