

# 心理学における実験計画の手引き

## 〔その1〕 実験計画の基本

～におい・かおりに関わる心理学研究で気をつけることは？～

このリーフレットは、におい・かおりに関わる研究をする中で、ヒトを対象として何らかの測定・評価を行う際に必要とされる基本的知識をまとめたものです。

心理学実験の初心者向けに易しく解説しています。皆さんの研究の一助となれば幸いです。

小林 剛史・山本 晃輔・小峯 裕己  
におい・かおり環境協会 技術課(重岡久美子・中辻 康・石井 進)

2025年1月10日

公益社団法人 おい・かおり環境協会

## はじめに

### ■ 心理学における行動とは

心理学では、ヒトの心的状態が外部に現れる動作や反応を総称して「**行動**」と呼びます。行動には、目に見えるような**動作はもちろん、感情の変化や発話、さらには脳や心臓血管系の活動**といった生理学的活動も含まれます。ヒトは学習や経験等により、多様な形の行動を示します。これらの行動をできる限り定量的に捉え、分析することが心理学の研究方法の基本です。

### ■ 主観的反応もヒトの行動

ヒトの言動やアンケートへの回答など、いわゆる「望ましき」や「**欺瞞**」といった要因が介入する主観的指標は客観性に乏しいと見なされることがあります。しかし、**主観的反応もまた、ヒトの特徴的「行動」の一部**であり、適切な方法で分析すれば重要な研究データとなります。一方、瞬目や脳活動、発汗など、意識的な要因が介入しにくい生理的指標は信頼できる、という考えがありますが、これらの指標だけで心理状態を正確に推測することはできません。主観的指標と生理的指標の特徴を理解し、組み合わせて活用することが、心理学研究において有効なアプローチとなります。

### ■ 生理的指標の変化は心理的指標の裏付けが必要

企業との共同研究では、しばしば、「ある条件下で脳活動を測定したい」という要望が挙がります。しかし、多くの場合、脳波の測定結果から解釈可能な結論には限界があります。

例えば、あるにおいを提示している状況で、脳波のある成分が上昇した場合を考えてみましょう。その上昇が脳波測定への緊張によるものか、それともそのにおいを嗅いで過去の楽しい記憶を思い出したことによるものか、経験ある生理心理学者でも判断することはとても難しいと言えます。心拍変動や発汗といった自律神経指標でも同様です。

最も重要な点は、**心理状態と表出した行動との対応（相関関係といってもいいでしょう）を繊細に検討**することです。しかし、その対応（相関関係）が研究者の恣意的推論によって解釈されることをしばしば目にします。

また、多くの生理指標を同時計測することは、実験参加者の心理的負担を増加させ、緊張を引き起こす可能性があります。その結果、本来取得したい生理データと異なったデータが得られる恐れがあるため、必要最低限の生理指標の測定に留めるべきでしょう。

**生理的指標における変化が心理状態と関連するものであるのかを明らかにするためには、以下の工夫が必要です：**

- ・ 他の生理的指標を同時に測定する。
- ・ 質問紙や直接聴取による主観的指標のデータも併せて収集する。
- ・ これらのデータを適切な分析に基づいて考察する。

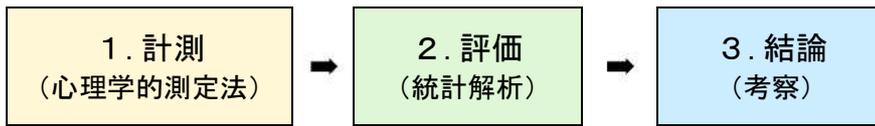
主観的指標のデータを丁寧に取得することは、心理学研究においてきわめて有効な研究手段です。

### ■ 実験計画の重要性と心理学の課題

心理学実験は、遵守すべきルールとスキルに基づいて行わない場合、結果の信頼性・妥当性が損なわれます。吟味された実験計画を立て、適切な分析を用いることで、実験結果の再現性を高めることが可能です。

特に、におい・かおりに対するヒトの反応は多様であり、これを定量的に測定することはきわめて難しいことです。巷には直感的に作成されたアンケートや、再現性に問題が生じそうな測定方法が散見されます。そこで、このリーフレットでは、心理学実験の経験が少ない方を主たる対象とし、ヒトの行動の測定・評価・分析において留意すべき事項の一部を、可能な限り平易な内容で解説します。

以下の流れに沿って、試験研究を計画する際に押さえておくべき留意点を説明します。



まず「その1. 実験計画の基本と計測」では、心理学の実験計画と計測に関する基礎的事項についてまとめました。続く「その2. 量的データと質的データの測定・評価」では、心理学における種々のデータの取り扱い方や、測定・評価について留意する点を説明しています。具体的な統計解析および結論（考察）については、リーフレットのその2で解説を行っています。

研究の目的等に応じて、適宜ご参考いただければ幸いです。

## 1. 実験計画の基本と計測

### ■ 研究における信頼性とは

心理学における**信頼性**とは、対象に対する測定の精度を指し、「**再現性**」、「**等価性**」、「**内的整合性**」の3つの指標によって評価されます。

信頼性に関する用語として「**妥当性**」がありますが、両者の違いは射的で例えることができます。

- ・ **信頼性**：「**的**の同じ場所に何度も命中できるか」の度合いを指します。
- ・ **妥当性**：「**的**の中心（狙い）に命中しているか」の度合いを指します。

例えば、何度も同じ場所に命中する（**信頼性が高い**）場合でも、その場所が**的の外側**であれば**妥当性は低い**と言えます。このように**信頼性**と**妥当性**は異なる概念であり、必ずしも一致するものではないことを覚えておきましょう。



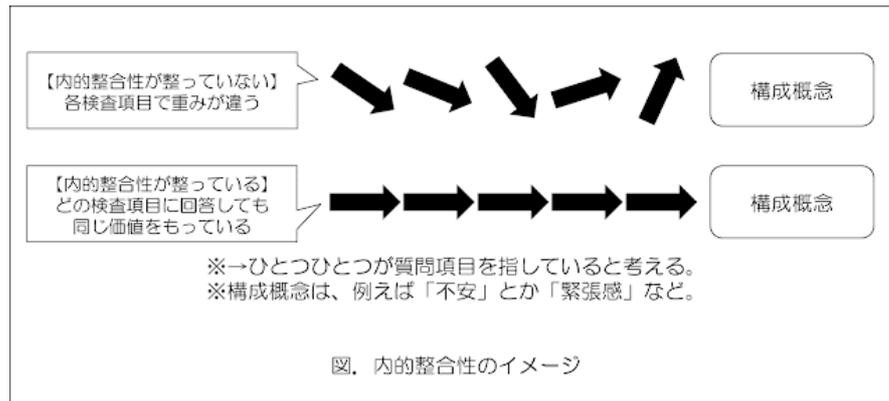
<https://healthpolicyhealthecon.com/2014/12/15/validity-and-reliability/>

再現性を平易に言えば、「**繰り返し測定しても同じ測定結果が得られること**」です。再現性が高ければ、**信頼性も高い**と言えます。

等価性とは、「**その検査と類似した構成概念の検査と一定の関係が認められること**」を指します。平易に言えば、検査の内容や難易度が元の検査と同程度である検査結果が、元の検査結果と相関しているかを確認することです。等価性が高ければ**信頼性も高い**と考えられます。

内的整合性とは、「**尺度内の各項目が同じ構成概念を測定しているかを確認すること**」です。内的整合性が高ければ、**信頼性も高くなります**。

信頼性に関する重要な指標の1つである**再現性の基準**は、学問領域によって大きく異なることがあります。物理学の世界では、何十万回もの**再現性が担保されなければ確固たる知見**として認めないとする考えもあります。



<https://public->

[psychologist.systems/16-%E5%BF%83%E7%90%86%E5%AD%A6%E3%81%AB%E3%81%8A%E3%81%91%E3%82%8B%E7%A0%94%E7%A9%B6%E3%80%81%E5%BF%83%E7%90%86%E5%AD%A6%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8B%E5%AE%9F%E9%A8%93/%E5%85%AC%E8%AA%8D%E5%BF%83%E7%90%86%E5%B8%AB%E3%80%802020-12/](https://public-psychologist.systems/16-%E5%BF%83%E7%90%86%E5%AD%A6%E3%81%AB%E3%81%8A%E3%81%91%E3%82%8B%E7%A0%94%E7%A9%B6%E3%80%81%E5%BF%83%E7%90%86%E5%AD%A6%E3%81%AB%E9%96%A2%E3%81%99%E3%82%8B%E5%AE%9F%E9%A8%93/%E5%85%AC%E8%AA%8D%E5%BF%83%E7%90%86%E5%B8%AB%E3%80%802020-12/)

心理学の分野では、個人差・個体差の大きいヒトや動物を対象とし、検査方法が物理学の検査とは大幅に異なるため、物理学のような高い再現性を求めることは現実的ではありません。それでも、可能なかぎり再現性を高める工夫や努力を行うことが重要です。

#### ■ 再現性の妨害要因は多種多様

心理学の教科書に出てくる「ある刺激に対する反応時間を計測する」という実験を授業の一環として実施した際、同様の結果が得られない、すなわち、再現性がないことがありました。その原因として、以下の点が挙げられます：

- ・「遅刻してきた学生を途中から実験に参加させない。」、「学生に私語をさせながら実験を行わせない。」など、基本的な注意事項を守らなかったこと。
- ・実験手順や注意事項を全ての参加者に等しく提示するための教示が不十分であったこと（通常は印刷物で行い、参加者によって偏りが生じないようにします）。

このように、実験過程には、些細な部分で再現性を妨げるノイズやアーチファクトを生み出す要素が多く存在します。これら要素一つ一つを除外するか否かの判断は、実験者の経験や感受性に異音する部分もありますが、初心者であれば可能な限り全ての要素を除外する方が良いと考えられます。

また、論文は再現性が担保されるよう詳細に記述する必要がありますが、実際には記述以上に留意すべき事項が多いことを念頭に置く必要があります。

さらに、恣意的に有意差が見られた部分のみを抽出して研究論文として公表するような行為では、再現性が低いことは明白です。研究者は常に再現性の高い研究を目指すべきです。

#### ■ 妥当性の確保は難しい

妥当性に関しては既に簡単に触れましたが、別の言い方で述べれば「研究対象とした現象を適切に研究できているか？」を問う概念です。しかし、妥当性に問題がある研究は依然として散見されます。におい・かおり環境学会の発表でさえ、妥当性に疑問が生じる事例が見られることがあります。

例えば、ある薬物にリラックス効果があることを検証したい時、どのような実験方法が考えられるでしょうか。その薬物Aを投与し血中濃度が高まる時間を見計らって、「あなたはどのくらいリラックスしていますか」と質問する方法が考えられるかもしれませんが。ヒトは言葉を用いる生き物ですので、言葉で直接ズバツと聞いてしまうのが一番手っ取り早いと思いがちです。しかし、この方法は短絡的であり、リラックス状態の多様な要素を無視しています。

妥当性を持ってリラックス状態であることを判断するには、以下の要素を総合的に考慮する必要があります。

#### ・生理学的要素

- ① 心臓の鼓動およびその変動
- ② 血圧
- ③ 発汗
- ④ 瞳孔の動き
- ⑤ 呼吸
- ⑥ 筋肉の緊張状態

#### ・情動・認知的要素

- ① おだやかな気持ちか
- ② 安心しているか
- ③ 自身が感じている問題についてどのくらい不安を感じるか
- ④ 将来に対する不安

こうした項目を質問項目に含めることで、総合的にリラックス状態を適切に評価することが可能です。もちろん、直接「リラックスしているか」を尋ねる方法に意味がないわけではありません。しかし、研究においては、**リラックス状態を検証するための信頼性および妥当性の高い質問紙**を活用し、精度の高い評価を行うことが求められます。

#### ■ 統計的検定手法は実験の計画段階から決まっている

心理学の実験では、統計的検定手法を実験の計画段階であらかじめ決定しておくことが重要です。このプロセスは、実験の信頼性や妥当性を担保するだけでなく、結果の解釈や科学的な透明性を高めるために欠かせません。適切な検定手法の選択は、実験計画に基づく独立変数と従属変数の設定、実験参加者の割り当て方法、そして仮説の性質に応じて決定されます。

実験計画段階で統計的検定手法の選択が必要な理由として、まず実験目的を明確にすることが不可欠である、ということが言えます。たとえば、「あるにおい・かおりが注意力に与える影響」を調べる実験を行う場合、「香りの種類」（**独立変数**）と「注意力のスコア」（**従属変数**）を設定します。この段階で、結果を検証するためにどの統計的手法が適切かを見極めておく必要があります。たとえば、比較する条件が複数ある場合は**分散分析**、2条件での比較なら **t 検定**を用いることが一般的です。

さらに、実験参加者の割り当て方によっても選択すべき検定手法が異なります。同じ参加者が複数の条件を経験する実験では、**対応のある t 検定**や**分散分析（参加者内計画）**を適用すべきです。一方、異なる参加者が各条件を経験する実験では、**対応のない t 検定**や**分散分析（参加者間計画）**を用いることが適切です。

また、次のような仮説がある場合、それぞれに応じた検定が必要になります。

- ・「香り A の提示が注意力を向上させるか」  
→ **片側検定**を用いて、特定の方向性のみを検証します。
- ・「香り A と香り B の注意力への影響に違いがあるか」  
→ **両側検定**を選び、どちらの条件が優れているかは問わずに差異を検証します。

以上の統計的分析の詳細については後述します。また、片側・両側検定については、**におい・かおりリーフレットその 2**で詳しく説明します。

なお、計画段階で決定した検定手法が十分でない場合、追加で事後分析を行うこともあります。しかし、**事後分析は計画段階の検定手法に従ったデータ収集を補完するものであるべき**です。新しい検定手法を適用する際は、その選択が研究の妥当性を損なわないよう注意が必要です。

#### ■ 実験計画段階で統計的検定手法を決定する意味

統計的検定手法を計画段階で決定する理由は、次の通りです。

- ① データ収集：必要なデータの種類や量を明確化し、不要なデータを収集する負担を回避します。
- ② 再現性：計画段階で手法を定めることで、他の研究者が同様の手順を追試できます。

③信頼性・妥当性：適切な検定手法を用いることで、実験の妥当性・信頼性を担保します。

#### ■ 実験参加者内計画と実験参加者間計画の違い

においAとにおいBに対するに感覚強度や発汗を評価・測定する実験を例として、説明します。

**実験参加者内計画**とは、同じ実験参加者がにおいAとにおいBの両方を嗅ぐ、という実験方法です。「**対応あり**」とも呼ばれます。一方、**参加者間計画**とは、異なる参加者が2種類のにおいのうち一方のにおいのみを嗅ぐ、という実験方法です。「**対応なし**」とも呼ばれます。

もうひとつ、事例を挙げます。A大学の教師Aによる「心理学概論」の教育効果を測定する際に、月に1回判定テストを実施するとします。履修者全員が4月から翌年の3月まで12回テストを受けるという手続きは、同じ実験参加者から12回データを取得することから、参加者内計画となります。一方、A大学、B大学、C大学、D大学において、教師Aによる単発の授業がどのような教育効果を呈するかをテストによって判定する場合、それぞれの大学における履修者のグループは異なるので、参加者間計画となります。

#### ■ 独立変数と従属変数は？

**独立変数**<sup>どくりつへんすう</sup>とは、平易に言えば、比較する条件に関わる変数のことです。たくさんの用語がでてきて混乱するかもしれませんが、1つの独立変数を「1要因」と呼ぶこともあります。例えば、かおりAを「高濃度」で呈示する場合と、「低濃度」で呈示する場合を比較するとき、独立変数は「濃度の高低」ということとなります。もちろん、より条件を多く設定して、濃度の条件を4種類、6種類と設定できます。この種類のことを「**水準**」と呼ぶこともあります。

一方、かおりAに加えて、かおりBを使用する場合、「かおりの種類」という独立変数に加わることとなります。この場合は、それぞれのかおりについて先述の濃度条件が4種類設定されるならば、全体の条件数は8つとなります。このような場合、独立変数の考え方として、濃度の要因（4種類）×かおりの種類の要因（2種類）の2要因の実験計画（要因計画ともいう）となります。

一方、何をデータとして比較するかが、「**従属変数**」<sup>じゅうぞくへんすう</sup>です。例えば、心拍数の変動を測定する場合には、この心拍数およびその変動が従属変数となります。

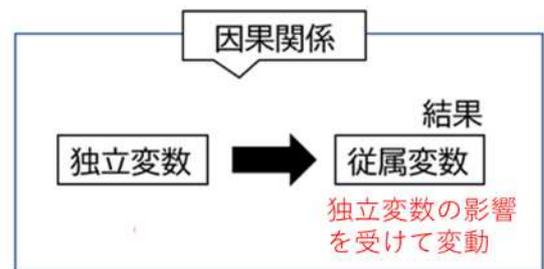
実験を計画する際、この独立変数と従属変数が混乱するケースがあるようですので、実験実施に際して**2変数の関係性をしっかり把握**しておくことが必要です。

心理学の実験では、基本的に実験を開始する前に、独立変数と従属変数に基づいて実験計画を綿密に立てることが重要です。この段階で、実験後に使用する可能性のある分析手法もある程度想定することが一般的です。たとえば、先に挙げた例では、2要因分散分析を用いることを計画段階で想定することが一般的です。なお、前述のように、実験後に新たな分析手法が必要だと気づき、事後的に追加で分析を行う場合もあります。それでも、独立変数、従属変数、実験計画、分析方法を一貫して関連付けて考えておくことで、実験をスムーズに進めることができるでしょう。

#### ■ 実験参加者のスクリーニング（抽出）は必要？

理想的には、**実験参加者は、母集団の中から無作為に選出**（心理学では多くの場合、「抽出」と呼びます。）されることが望ましいと言えます。しかしながら、実験目的に沿った適切な参加者が選出されなければ、当初の目的の実験を行うことができないということも事実です。そのため、実験参加者の募集に際して、ある程度のスクリーニングを行うことが一般的です。

多くの場合、におい・かおりの心理実験では、**鼻、副鼻腔に関する疾患（風邪や花粉症を含む）の有無、喫煙習慣の有無、アロマや化粧品などの香料に関する職歴の有無、においへの関心の程度**などを事前に尋ね、参加者を選定します。また、参加者には予め実験当日に香水などのフレグラン



スを身につけないこと、香辛料などのおいしさを抑える強い飲食物の摂取を控えることを伝えておき、実験当日も確認するとよいでしょう。

なお、研究領域にもよりますが、近年の学術雑誌では参加者の詳細なデモグラフィックデータが求められることも増えてきました。性別、年齢、教育歴、年収、家族構成など、きわめて個人的な情報も含まれますので、スクリーニングの際には通常の選択肢に加えて、「回答したくない」という選択肢を設けることも必要でしょう。また、性別についての尋ね方として、「あなたが自認する性別は～」、「あなた自身が考える自分の性別は～」といった表現の使用も一般的になってきています。

実験参加者人数について、近年、心理学系の学会誌ではサンプルサイズの検定を求められることも増えてきています。調査研究では、数万人や数千人といったデータもみられ、ビッグデータ化していることは否めません。実験計画、使用する変数等で必要人数は変動しますが、**1群10名以上は必須**と考えていただき、過不足のない参加者募集を行うことが求められています。

### ■ 「課題要求特性」に対する配慮

人を対象とした研究に関する倫理指針を遵守することが求められて以降、実験に協力してくれる参加者には、予め実験の目的を認識して貰った上で、実験に臨む「**インフォームドコンセント**」を得ることがスタンダードになってきています。インフォームドコンセントを得る過程で、においに対する先入観を与えてしまう、においの有無による各条件の比較を行う研究であるなど、実験に関する情報を前もって与えてしまうと、参加者が条件の違いを読みとってしまうことなどから、研究者が得たいと考えているデータが取得できないケースが多々あると思われます。

例えば、におい物質Aの提示が認知的機能を改善させる効果があるかを検討する実験を行うとします。実験条件として、におい物質Aを呈示した状態で認知課題を実施します。統制条件(対照条件)として、無臭空気を提示した状態で同様の認知課題を実施します。このとき、実験前の説明により、においが課題遂行に及ぼす影響を検討する実験であると参加者が認識してしまうと、においが提示されている条件で、認知課題を「頑張っ<sup>て</sup>」行ってしま<sup>う</sup>ことがあります。このように、参加者が、研究者が期待する結果に対して何らかの反応を示してしまうことを、「**課題要求特性を刺激する**」と言います。心理学実験では、こうした影響を回避するために、**参加者間計画**を採用する、若しくは<sup>せんざいれんごうがくしゅうかだい</sup>**潜在連合学習課題** (**Implicit Association Test**)<sup>インパクト アソシエーション テスト</sup>を採用することがあります。

### ■ 順序効果を回避する実験手法

かおりA、かおりB、かおりC、無臭、の4つの条件における注意維持課題(内田クレペリン検査のような)の成績を比較するという実験を想定しましょう。ここでは、参加者内計画と参加者間計画の両方が設定可能となりますが、参加者内計画を採用する問題点を考えてみましょう。

例えば30名の参加者に4つのかおりを同じ順番で嗅いでもら<sup>う</sup>と、当然、「**順序効果**」(経験する条件の順番が結果に影響すること)が生じることとなります。これを回避するために、理想的には、30名が**ランダムな異なる順序**で4つの条件を経験する計画を立てる必要があります。しかし、それでは $4! = 24$ 通りの条件の組み合わせを設定する必要があり、参加者数や実際の実験実施を考えると非現実的です。

そこで、多くの場合には「**カウンターバランス**」(ある参加者は、A、B、C、無臭、の順序、別の参加者は無臭、C、B、A、の順序を設定するなど、**独立変数の提示順序等のバランスを取る**こと)という手法を用いますが、当然、4回目の課題を行う際の疲労が大きいと想定されますし、同じ課題を4回行うことの学習効果も想定されます。このように、参加者内計画は、参加者数の節約にはつながりますが、複数の要因が介入する恐れがあります。

これらのことから、参加者間計画を採用する方が妥当と判断できます。しかし、順序効果の影響は回避できるものの、それぞれの条件で30名ずつの参加者を確保しなければならず、4条件で120名の参加者が必要となります。その際に重要なことは、30名の参加者の年齢の分布など、**想定している母集団の属性を正しく反映した参加者**であることが求められます。

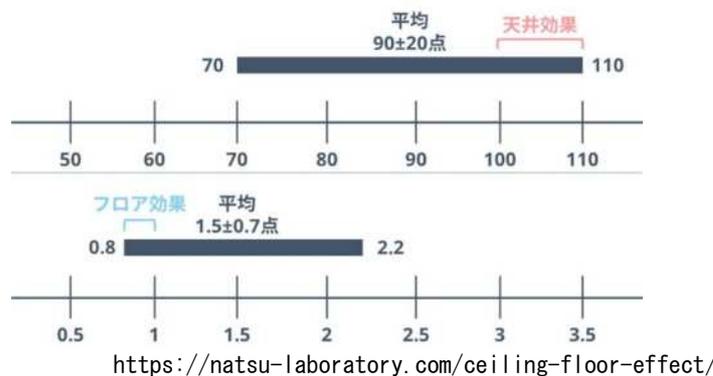
## ■ 従属変数の天井効果と床効果を回避する設定

独立変数によって変化すべき従属変数の天井効果と床効果についても、繊細な設定が必要です。心理学実験で最も難しい設定の1つが、**認知課題等の課題の難易度設定**です。

全ての参加者が100点をとってしまう、あるいは0点をとってしまうような課題を与えてしまうと、仮に実験計画における仮説が正しかったとしても、その正しさが統計的に検出されることはありません。

例を挙げましょう。教師Aと教師Bがある国家試験対策講座を担当し、1ヶ月に1回行われる8ヶ月間の模擬試験の受講生の成績の推移を調べます。教師Aの講座は分かりやすく知識の定着率が高いと想定し、模擬試験の成績がより上昇傾向を示すという仮説を立てます。

ここで、模擬試験の内容が難しすぎて、何ヶ月やり続けても2人の教師の受講生の試験成績は0点でずっと推移し続けると、2つの受講生群の成績の有意差も、受講期間による成績の上昇を示す有意差も測定できません。当然、仮説は検証不可能です。このような現象を**床効果**と呼びます。



一方、模擬試験の内容が簡単すぎると、受講生全員の成績が100点になってしまいます。**天井効果**と呼びますが、床効果同様、有意差が測定できないので仮説は検証不可能です。

**試験の成績が適度に分散する**ように、実験者は模擬試験に関して適切な難易度を設定することが求められます。なお、学習効果をできるだけ回避する上で、模擬試験の回数により、成績が大幅に向上することがないよう、模擬試験問題の難易度を適切に設定する必要があります。

さらに、成績が集中しないように配慮する必要があります。難易度の設定が適切でも、問題が4問しかない場合、参加者のデータのほとんどが50点と75点に集約しがちです。この場合、従属変数がある値に集中してしまい、分散分析の検定に不適となってしまいます。

## ■ 内的妥当性と外的妥当性を確保した研究を

良い研究には、研究成果の信頼性・妥当性が高いことが求められます。研究の信頼性や妥当性を評価するために、「内的妥当性」や「外的妥当性」という概念が用いられます。

**「内的妥当性」**とは、研究や実験で導き出された結果の正当性、および因果推論の適切さを示す概念です。因果推論とは、独立変数のデータと従属変数のデータから、両者の因果関係（原因とそれによって生じる結果との関係）を統計的に推定していく考え方です。

**「外的妥当性」**とは、研究結果が現実の多様な場面に適用できるかを示す指標です。別の集団に当てはめた場合でも、同様な研究結果を得られるか、一般化できるかの程度を示す概念です。

### ・例1：遠隔操作可能な手術装置を使った手術とビデオゲーム

小さい頃からビデオゲームをたくさん行ってきた医学生と、全く行ってこなかった医学生で、遠隔操作可能な手術装置の操作テストを比較したところ、ビデオゲーム経験が豊富な医学生の方が操作テストの成績が有意に高い結果が得られました。

ビデオゲームで培ったコントローラーの微妙な操作方法が役に立ったことが、その理由だと推測できます。この場合、内的妥当性は高いと考えられます。

しかし、この結果を受けて、ビデオゲームをたくさん行ってきた医学生だけが、実際の患者に対する手術に適すると判断して良いでしょうか。実際の手術では、統制された実験計画に含まれないような多様な要因が存在します。外科手術は他の医師や看護師との協働的コミュニケーションによって成立するため、こうした連携のスキルも要求されます。手技のみに集中するだけではなく、こうした要因が手術結果に大きく影響を与えるでしょう。このようなことを考慮すると、この結果は、外的妥当性は低いと言えます。

#### ・例2：購買欲求と実際の購買行動

ある限られた実験状況で、商品Aに対する購買欲求が商品B、C、D、Eよりも著しく高いという結果が得られたとします。しかし、実際にコンビニエンスストアで販売した場合、商品Eが圧倒的によく売れるケースが起こることがあります。これは、実験状況で得られる購買欲求と、実際の購買行動の乖離があることを示しており、外的妥当性がない実験でした。研究の難しさを反映する典型例です。

### ■ 内的妥当性と外的妥当性のバランス

実験計画とそれに伴う仮説が妥当であれば、内的妥当性の高い結果が得られる可能性は高まります。しかし、研究成果が本来目指すべき研究目的から逸脱してしまうことがあることも心得ておく必要があります。研究者は、内的妥当性のみにこだわることで、現実の場面で有益とはならない成果を得てしまうリスクについて再考する必要があります。研究目的を明確にし、内的妥当性と外的妥当性のバランスを考慮した実験計画を立てることが重要です。

## 2. 量的データと質的データの測定・評価

### ■ 心理学実験における心理尺度使用の留意点

「あるにおいを使用した際の快適さを測定したい」という場合、一般的には、実験参加者に「あるにおい」を呈示し、「そのにおいはどれくらい快適ですか」と尋ね、「まったく快適ではない」から「非常に快適である」の5段階で評定させます。この方法で、そのにおいの快適さは適切に測定されたといえるでしょうか。もちろん、このように直接的な表現を用いた方法にも意義はあります。しかし、「快適さ」といった言葉の理解は、人によって異なる可能性があります。ある人は「楽しい気分」を「快適」と感じるかもしれませんが、また別の人は「落ち着いた気分」を快適と感じるかもしれません。

このように1つの表現においても多様な感じ方があり、それをできるだけ多角的にかつ詳細に測定するために、心理学領域ではさまざまな表現を用いた質問項目リストを使用します。

質問項目リストは、通常、質問紙と呼ばれ、項目リストをどのように得点化するかも含めてまとめたものは**心理尺度（あるいは単に尺度）**と呼ばれます。心理学領では、新しい心理尺度の開発や海外で開発されたものの日本語化は、複数の調査や実験を経た上でその信頼性や妥当性を確立させる必要があるため、それ自体が1つの重要な独立した研究とされています。これは、複数の調査や実験を経て信頼性や妥当性を確立するためです。

これらの尺度における教示や評定段階、選択肢などの要素は、それぞれの開発者が熟考し、実証的な研究結果に基づいて採用したものです。したがって、それらを使用する際には、原則として表現を安易に変更しないよう注意しなければなりません。同じ方法で測定することで、先行研究との比較が可能となるといった利点があります。

心理学的評価を行うのであれば、**できるだけ既存の心理尺度を用いる**ことが望ましいといえます。におい・かおりに関する心理尺度もいくつか開発されていますので、研究目的に応じて使い分けましょう。

### ■ 質問紙で類似した質問項目が沢山あるのはなぜ？

心理学の研究では、質問紙を用いて対象の特性や反応を測定することが一般的に行われます。

この際、似たような内容の質問が多く含まれていることに気付くかもしれません。これは測定・評価の信頼性と妥当性を高めるための重要な工夫です。信頼性の観点では、**内的整合性**を確保することが挙げられます。同じ構成概念を異なる表現で繰り返し測定することで、項目ごとの偏りや曖昧さを補完し、一貫性のある評価が可能になります。また、妥当性の観点では、測定対象を多角的に評価することで、測ろうとする概念を適切に捉える精度が高まります。さらに、類似した質問項目を含めることで、回答の一貫性や信憑性を確認することもでき、矛盾した回答が多いデータを特定して除外するなどの処理が可能です。また、統計解析では複数の質問項目を因子ごとの合計得点として扱うことでデータの分布が安定し、解析結果の信頼性が向上します。

このように、類似した質問項目は単なる重複ではなく、データの質を確保し、研究の信頼性や妥当性を向上させるために重要な役割を果たしており、心理学における質問紙設計において不可欠な要素です。

## ■ 量的データ

心理尺度を用いて得られるデータは基本的に**量的データ**です。量的データは統計的検定を行う際に重要な役割を果たします。量的データには、**間隔尺度**と**比率尺度**の2種類があります。

間隔尺度は、気温のように絶対的なゼロ点が存在せず、数値で測定可能なデータを指します。この尺度では、数値間の間隔が等しく保たれているため、データの差を比較することが可能です。

一方、比率尺度は、身長や体重のように絶対的なゼロ点が存在し、データの間隔だけでなく比率にも意味を持つデータを指します。このため、たとえば「ある値が別の値の2倍である」といった比較が可能です。

心理学の実験では、これらの尺度に基づいたデータを収集し、それをもとに記述統計や推測統計を用いてデータの分析を行います。これにより、得られた結果を基に仮説の検証や心理現象の理解が進められます。

## ■ 主観的指標の尺度はVASがよい

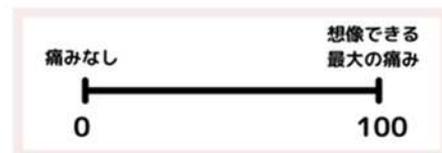
定量的な変数の尺度である間隔尺度、比率尺度は、t検定や分散分析等の統計解析を行うことが可能です。ところで、においの快不快度や臭気強度という主観的指標は、間隔尺度や比率尺度として扱うことが可能でしょうか。例えば、6段階臭気強度では0を「無臭」、5を「強烈」とした尺度において、「0、1、2、3、4、5の感覚強度は等間隔である。」と言えるでしょうか。

心理学では、こうした数値間の距離が心理的に等間隔であるか検討を重ねてきました。その結果、この間隔が絶対的に等間隔である保証はないことが明らかになりました。今までは、統計解析を行うための便宜的な方法として、数値に○をつける、あるいはチェックをいれるような回答方法を採用し、「便宜的に」数値間の心理的距離が等間隔と仮定して、t検定や分散分析を行ってきました。現在でも、こうした手続きをとることはある程度容認されているものの、本来、1は0よりも大きく、3は2よりも大きい、そして、その数値間の間隔は等間隔であることが保証されない**順序尺度**です。順序尺度では、加減乗除を行うような検定は通常できません。

そこで、数値や言語ラベルの間を数直線でつなぎ、数直線のどの位置に斜線やチェックを記入できる方法として、**ビジュアルアナログスケール (VAS)** が用いられるようになりました。VASは主観的指標を連続変数として扱うための有効な手段として用いられるケースが増えています。連続変数として扱えるため、特に同一の実験参加者における経時的な変化の比較や刺激提示前後の変化の比較等に有効です。このように、VASは主観的指標を細かく定量化するのに適した方法です。

### VAS (visual analog scale)

「痛みなし」と「想像できる最大の痛み」を両端とする直線上で、痛みの程度を示す方法。視覚的アナログスケールともいいます。



痛みの強さに対するVASによる尺度

<https://www.https://yakumedical.com/pain-scale/539>より引用

## ■ SD 法

主観的指標を評価する方法として **SD 法（意味微分法）** もよく使用されます。SD 法は、評価対象に対する印象や感情を「快-不快」「強い-弱い」など、対になる形容詞を両端に設定した尺度を用いて測定します。評価者はこの尺度上の点を選ぶことで、自分の印象を定量的に表現します。例えば、あるにおいに対する感覚が「心地よい」か「不快」かを 7 段階で評価する場合、各段階の間隔を統計解析上の間隔尺度として扱うことが可能です。SD 法は、評価者の直感的な反応を詳細に捉えることができるため、におい・かおりの心理学研究で広く活用されています。また、SD 法に VAS に準ずる評定方法を取り入れ、連続尺度として使用することで、より信頼性の高い評価手法として用いることもできますでしょう。

## ■ 質的データ

前述のように、心理学では基本的に量的データをもとに分析が行われますが、量的データだけでは収集できない側面があるため、質的データを採用することもあります。**質的データは**、数量としての意味をもたないデータであり、例えば実験参加者の内実などを自由記述やインタビューで集めたデータのことを指します。

実践研究や事例研究では、質的データのみを取り扱うことが多いと言えます。質的データを取り上げる際には、研究目的に合致する例や記述などを恣意的に選択してしまうといったある種のバイアスが生じがちですが、データとして扱う以上、また研究の遡上にのせる以上、**客観性を担保**する必要があります。

**KJ 法**や**グランデッドセオリー**は、得られた質的データがどのようなカテゴリーに分類可能かを検討する際に使用されます。詳細な手続きは個々の専門書に委ねますが、例えば KJ 法の初期では、研究目的を伝えていない数名の研究者が、それぞれ独立して質的なデータをいくつかのグループに分類します。最終的な調整は研究者が行うものの、このように分類作業の一部を第三者に依頼することで客観性を担保しています。

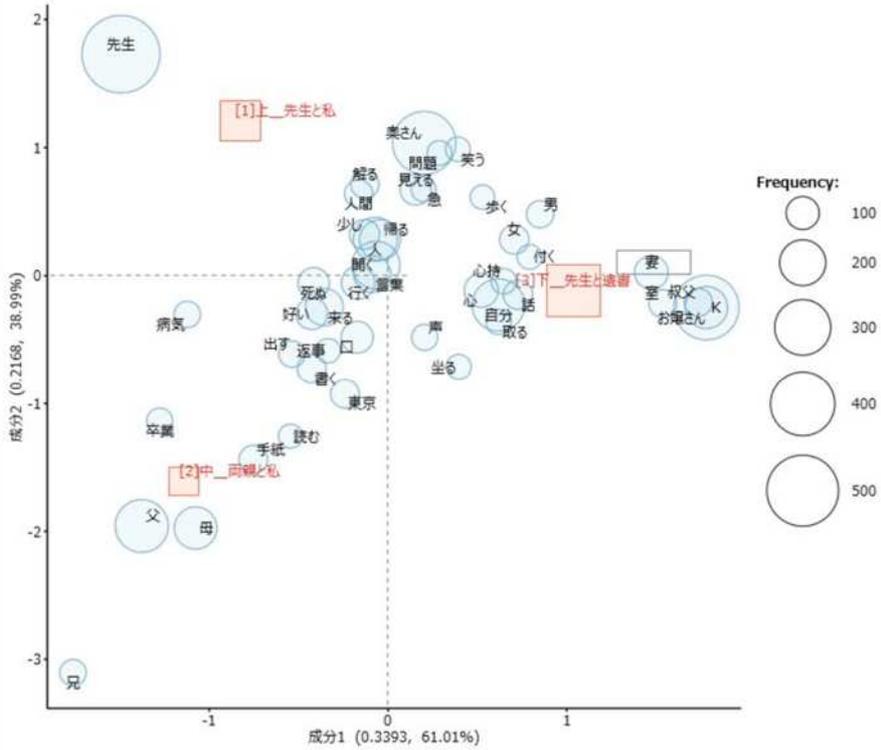
また近年では、自由記述などのテキスト（文章）型データを数量的、客観的に分析する手法の 1 つとして、**テキストマイニング**が使用されることが増えてきました。テキストマイニングでは、テキストデータのなかの特定の語の頻度や出現パターンの似通った、すなわち共起の程度を視覚化することが可能であり、それらのネットワークから内容をカテゴライズすることができます。専用のソフトが多数開発されており、比較的簡単に行えることから利用機会が高まっています。

テキストマイニングの解析手法はいくつかありますが、その一つに**コレスポネンス分析（対応分析）**があります。この分析では、テキスト内における語の出現度をベースに、各条件や群ごとに特徴的に出現している語を視覚的に把握するために、図内にその語をプロットします。円の大きさは出現語の多さを示しています。

事例の図は、夏目漱石の「こころ」の全文について、上、中、下の三巻を要因として設定し、コレスポネンス分析を行ったものです。原点付近に集まっているのは、どの巻にも登場する特徴のない語です。一方で、「先生」は「上」巻、「父」、「母」は「中」巻、「妻」は「下」巻の近くに付置されており、これらは各巻に特徴的に出現する語であることが視覚的にみてとれます。

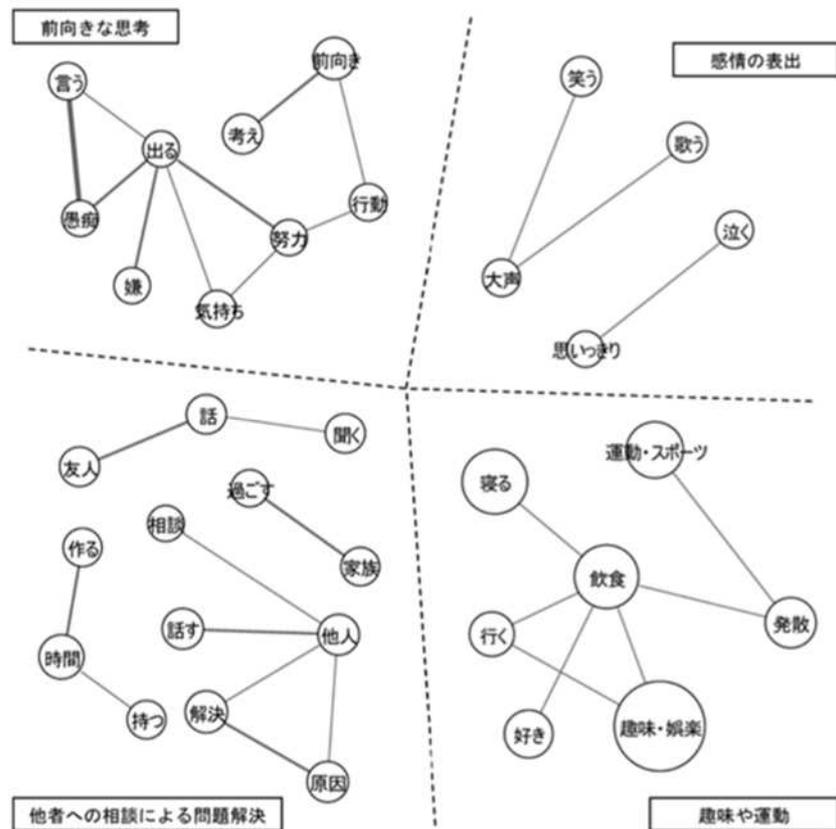
この他にも多用されている解析手法として、**共起ネットワーク**というものがあります。この解析では、単語がどれくらいの頻度でテキスト内に出現していて、どのような単語同士と一緒に出現がしているかを線で結び、ネットワークとして描くことで、語と語が互いにどのように結びついているかを図で提示する手法です。

事例の図は、ストレスをためずに生活する技能に関する自由記述の結果をまとめたものです。図の左上では、「前向き」「気持ち」「行動」「考え」などの語に関連がみられます。これらの語群は「考え方を前向きに変える」ことを示すため、「前向きな思考」として類型されます。



コレスポネンズ分析の一例（夏目漱石のころを対象として）

([https://photos.google.com/share/AF1QipMsNie80sVfSIBowz2VtushGagZ8KndQb6rMUwoXOAG0I3Z2sUZK1x8gTLae2h\\_MA?k ey=b1JEQzFtcIIJd25CY0xnWExtUHR5UWMOemhTRzVn](https://photos.google.com/share/AF1QipMsNie80sVfSIBowz2VtushGagZ8KndQb6rMUwoXOAG0I3Z2sUZK1x8gTLae2h_MA?k ey=b1JEQzFtcIIJd25CY0xnWExtUHR5UWMOemhTRzVn))



ストレス対処の共起ネットワーク（嘉瀬他，2016）<sup>3)</sup>

## ■ 嗅覚能力を測るには？

参考文献 1)

実験参加者のスクリーニングにおいて、参加者に自身の嗅覚能力を主観的に報告させることがあります。しかし、それが実際の嗅覚能力を正確に反映しているかどうかはわかりません。自信過剰、あるいは自信過小により、実際の嗅覚能力と一致しない可能性は十分に考えられます。そのため、**客観的な指標となる嗅覚測定検査法を実施**することがあります。

何を明らかにするかにより、いくつかの嗅覚検査法があります。

例えば、嗅覚測定に適した嗅覚の有無を明らかにするためには、「臭気指数等の算定の方法」(平成7年環境庁告示 63 号)に定められているパネルの選定方法があります。その他にも、<sup>べんべつ</sup>においの弁別能力(あるにおいが他のにおいと違うことがわかる能力)や<sup>どうてい</sup>同定能力(呈示されたにおいが何のにおいかわかる能力)を測定する検査方法があり、日本人向けに改良された検査法が開発されています(例えば、UPSIT-JやOSIT-J)。

特にOSIT-Jをさらに簡便化した**嗅覚同定能力研究用カードキット Open Essence**は実施も容易であり、高齢者にも適用可能です。



Open Essence (嗅覚同定能力研究用カードキット)

<https://staff.aist.go.jp/kobayakawa-tatsu/smell-db/new.html>

## ■ 適切な嗅覚刺激の呈示方法は？

参考文献 2)

嗅覚刺激は視覚や聴覚刺激と異なり、呈示のオン・オフを簡単に制御できるものではありません。オフにしたとしても、実験室内や鼻腔内にかおりが残ってしまうため、実験刺激としてにおい・かおりを呈示することは非常に難しいといえます。

このような嗅覚刺激による特徴に留意しつつ、**心理学的実験では必要最低限に制御**された方法を用いて、これまで研究や実験が行われてきました。

におい・かおり刺激の呈示には、嗅覚物質から揮発した気体を用いますが、その方法には専門的な装置を用いたものから、一般的な市販の既製品を用いるものまで様々です。

専門的な装置の1つとして、**オルファクトメーター**があります。これは、においガスの濃度や単位時間当たりの呈示量を機械的に制御、調整してにおいガスを呈示する装置です。

さらに、におい物質を充填したマイクロカプセルを無香の粘性のある練り物に混ぜて成型した**においスティック**があります。スティックを塗布した薬包紙を擦ってカプセルをつぶすことでカプセル内から揮発するにおいガスを呈示する簡便な方法です。既述の Open Essence も類似した方法です。

オルファクトメーターやマイクロカプセルによる呈示は精度の高い方法ですが、ある程度の専門的な知識や技術、環境が必要であるため、個人が使用することは難しいかもしれません。

比較的簡便な方法の1つとして、**においボトル**があります。においボトルは、ボトルの中に一定容量のにおい物質を入れ、ボトルの上層部にたまったにおいガスを呼吸によって吸い込ませる方法です。ボトルから鼻先までの距離や嗅ぎ方および呈示時間を一定にすることで、呈示量をある程度制御することが可能です。ボトルにもさまざまなものがありますが、ボトル自体が無臭であり、においの吸着がしにくいものを選ぶ必要があります。製品によっては予め蓋を開けて空気に晒し、材質のにおいを除去することもあります。材質としては、テフロン素材のものが望ましいのですが、

高価でありまた硬いため容器を押しづらく実際には使いにくいかもしれません。その点、ポリプロピレン製のボトルは安価でかつ入手も容易であり、材質も柔らかく使いやすいといえます。

におい物質が固体の場合には、そのまま、あるいは適度なサイズに切断をしたものをボトルに入れます。香料などの液体の場合には、綿球を予めボトルにいれておくことで転倒による液体の漏出を防ぐことができます。

におい物質の量については、実験目的を伝えていない人などを対象に事前に予備実験を行い、**強度を一定に調整する必要があります**。秤や測量機能のあるピペットなどで調整しましょう。また、香料の強度が強い場合には、プロピレングリコールなどの無臭の溶液で希釈するという方法もあります。

その他、におい・かおりの評価実験では、におい紙が使用されることも多いでしょう。こちらも香料を付与する紙の長さを調整することで強度の調整が可能です。

また、一室すべてににおいを拡散、充満させるような実験では、市販されているアロマディフューザーも効果的です。



オルファクトメーター

[https://www.tokyokankyo.jp/kankyoken\\_contents/research-meeting/h17-01/1703-pp.pdf](https://www.tokyokankyo.jp/kankyoken_contents/research-meeting/h17-01/1703-pp.pdf)



においスティック (OSIT-J)

<https://j-ichiyaku.com/stick/>

「その2. 量的データと質的データの測定・評価」では、心理学における種々のデータの取り扱い方や、測定・評価について留意する点を説明します。

#### 参考文献：

- 1) 綾部早穂・中野詩織 (2018) 2.4.2 嗅覚に関する検査および質問紙 日本基礎心理学会監修 坂上貴之・河原 純一郎・木村 英司・三浦 佳世・行場 次朗・石金 浩史(編)基礎心理学実験法ハンドブック 朝倉書店 pp.112-113.
- 2) 綾部早穂・中野詩織 (2018) 2.4.1 嗅覚刺激提示法 日本基礎心理学会監修 坂上 貴之・河原 純一郎・木村 英司・三浦 佳世・行場 次朗・石金 浩史(編)基礎心理学実験法ハンドブック 朝倉書店 pp.108-111.
- 3) 嘉瀬貴祥・坂内くらら・大石和男 (2016) 日本人成人のライフスキルを構成する行動および思考：計量テキスト分析による探索的検討 社会心理学研究, 32, 60-67.
- 4) 山田剛史・村井潤一郎 (2004) よくわかる心理統計 ミネルヴァ書房