

Chapter C

心理尺度・調査票の作成とデータ収集

心理尺度や質問紙調査の作成に関するお話を紹介しています。

本講義資料では、心理尺度を用いて測定されたデータを分析する代表的な方法を紹介しています。その背後には、「正しく測定できる」という前提があることは言うまでもありません。

そこでこの補足チャプターでは、主に心理尺度について、作成からデータ収集までに關するポイントや注意点などを紹介します。ただし一部の内容は、心理尺度に限らず、人間を対象に実施するアンケート調査全般に關係する話です。

C.1 心理尺度開発の道のり

心理尺度を作るのは、多分一般の人が思っているよりも大変な作業です。心理テストのように「なんかそれっぽいことを聞けばいいんでしょ」では、学術的には否定されてしまう可能性があります^{*1}。

そこでまずは、心理尺度というものがどのように作られているのか、そのプロセスを既存の尺度論文をもとに見ていきたいと思います。

C.1.1 質問紙ができるまで

心理尺度は、質問紙（紙以外にもウェブなど）による調査で用いられます。すなわち、心理尺度の利用においては、**実際に参加者に答えてもらう必要がある**わけです。そして、作成した尺度が「良いもの」になっているかを確認する段階でも、誰かに答えてもらう必要があります。したがって、心理尺度ができるまでのプロセスを知るためには、まず質問紙調査（アンケート調査）ができるまでのプロセスを確認するのが良いでしょう。

ひとつの質問紙ができるまでのプロセスは、細かく分けると以下のような手順になります（鈴木, 2016）。

1. 調査テーマと調査目的の再確認

^{*1} すべての心理テストがダメ、というわけではありませんが、「学術的に認められる心理測定」になるためには、いくつかのハードルを超える必要があります。

2. サブテーマごとの質問項目候補の選択
3. 質問項目の決定
4. デモグラフィック項目群の決定
5. 質問の種類と順序の決定
6. 回答形式の選択
7. 質問文・回答選択肢の作成
8. 教示文の作成
9. 最終項目群の作成
10. 挨拶状の作成
11. 表紙の作成
12. 体裁・書式・レイアウトの検討
13. 質問紙ドラフトの完成
14. 予備調査の実施
15. 質問紙ドラフトの修正
16. 質問紙の完成

以下では、質問紙作成の流れを心理尺度作成という目的に寄せる形で詳細に見ていきます。

1. 調査テーマと調査目的の再確認

最初に質問紙調査の目的を明確にしておきます。心理尺度を作るという目的であれば、**どのような心理的特性を測定したいか**を可能な限り厳密に言葉として定義しておく必要があります。心理学では一つ一つの心理的特性を**構成概念 (Construct)** と呼びます。例えば、UWES を作成した Schaufeli らは、ワークエンゲージメントを以下のように定義しました (島津, 2015)。

i ワークエンゲージメントの定義

ワーク・エンゲイジメントは、仕事に関連するポジティブで充実した心理状態であり、活力、熱意、没頭によって特徴づけられる。エンゲイジメントは、特定の対象、出来事、個人、行動などに向けられた一時的な状態ではなく、仕事に向けられた持続的かつ全般的な感情と認知である。

もちろんこの定義は独りよがりではいけません。Schaufeli らは、この尺度を開発する前に行った “theoretical analysis”^{*2} を通して、この定義を導き出していました。基本的には、構成概念の定義は先行研究で展開された理論や、実際にデータ分析を行った結果をもとに行われます。正直言うと、ここは「明確な根拠があって、みんなが納得できる定義」が示せたら OK という感じです。だって「構成概念」そのものを見たことがある人はいないんだから。

^{*2} ただ、この “theoretical analysis” を行った論文へのアクセスが大変＆オランダ語で書かれているっぽいので、具体的な分析の手順は不明です。ごめんなさい。

2. サブテーマごとの質問項目候補の選択

構成概念の定義ができると、次はこれを実際に観測可能な形に落とし込む作業が始まります。これは、構成概念の定義を細分化・具体化していく作業と言い換えられるかもしれません。健康診断でも、例えば「脂質異常症（高脂血症）」という**血液**の異常の有無を判断しようというときに、医者が血液を目視で確認するわけではありません。より具体的に「総コレステロール 220mg/dl 以上」や「中性脂肪 150mg/dl 以上」といった基準を設けて、血液中の成分量を調べています。このように、定義を**操作または観測可能な形で細分化・具体化することによって**、測定客観性を得ることができます。科学である以上、客観性は非常に重要なのです。

血液中の成分ならば直接量を測ることができるので幾分ラクなわけですが、心理学の構成概念は目に見えない上に曖昧なのでその具体化はより重要かつ厄介です。血液のように「検査する側が定義を理解できていれば良い」だけではなく、**回答する側が客観的に答えられる必要**があります。例えば、あなたは「今、ワーク（研究）エンゲージメントを感じていますか？」と聞かれたら、どう答えるでしょうか。構成概念は、通常そのまま聞いても回答者が答えづらいことが多く、また答えたとしても人によって異なる基準で回答してしまう可能性が非常に高くなります。例えばある人は「仕事はとても楽しいけど、趣味の方がそれ以上に楽しいから（相対的に）ワークエンゲージメントは高くない」と答える一方で、別の人は「生きるために必死で働いている（生活の中での優先度が高い）からエンゲージメントは高い」と答えるかもしれません。上述のワークエンゲージメントの定義に照らし合わせると、前者は仕事に対してポジティブなので高いとみなせる一方で、後者はあまり高くなさそうですが、実際に調査者が観測できる回答（値）では全く異なる結果が得られてしまうわけです。

ということで、構成概念を細分化した上で、誰が見てもほぼ同じ意味で解釈して回答できる形の質問文を考えていく必要があります。ワークエンゲージメントが「活力、熱意、没頭」の3つの要素によって特徴づけられるというならば、ワークエンゲージメントを直接聞くのではなく「活力」「熱意」「没頭」の程度をそれぞれ聞いてあげるようにします。ただ、「あなたは仕事に対して活力がありますか」というだけだとまだ曖昧かもしれません。というのも、「活力がある人」を想像すると何パターンかの人が浮かんできます。例えば「テキパキ動いている」とか「長時間働き続けている」とか「新しいことに挑戦している」とか「声が大きい」…とか？ こういった心理状態や行動のなかから、なるべく多様な解釈が生まれ得ないものを実際の項目として選択していくことになります。

とはいえ、データを取る前からどの項目が良い項目かを判断するのは難しいものです。そこで尺度作成の研究では通常、最終的に完成する尺度の項目数よりも多くの（大量の）項目候補を用意した上で、収集したデータから統計的な観点などによって「良い項目」だけを残していくことが多いです。ですが項目数を無尽蔵に増やすわけにもいきません。回答するのも人間なので、あまり項目数が多くなりすぎると疲れて適当に答えたりしてしまうかもしれません。そのため、あれもこれもと項目案を大量に作りまくるのは得策では無いと言えます。ある程度の客観性をもった項目案を作る方法としては、以下のような方法がある気がします。

- 自由記述アンケートを行う

- － 例えば「あなたの周りにいる、仕事に活力がある人はどんな人ですか」などとアバウ

トな質問を投げかけ、多く得られた回答は「多くの人がイメージする活力」ということで客観的に納得されやすい可能性が高いです。

ー ただ、そのために別でアンケートを立てる必要があるのでコストが大きいのがネックです。

- 先行研究に根拠を求める

ー 「ある論文で、活力が高い人の特徴はこう書いてあった」といった根拠があると納得しやすいですね。これは、構成概念の定義を明確化する手続きと近いものがあると思われます。

- 生成 AI に作らせる

ー 近年では、生成 AI に項目案を作らせて、その内容を人間が判断したり、実際に回答してもらったデータに基づいて評価したり、といった研究も見られるようになっていきます (e.g., 佐々木 & 豊田, 2024)。生成 AI が提案するすべての項目が「良い」ということは無いですが、大量に作成させると、その中には使える項目がある程度混ざっているはずで、そのような方法で項目案を増やすのもありでしょう。

3. 質問項目の決定

心理尺度は大きく分けると以下の 3 つのパーツで構成されています。

教示文 (ヘッダー) 項目全体に関する教示。例: 「以下の文章について、どの程度当てはまるかを選択肢の中から選んでください。」

項目 (本文) 具体的な一つ一つの文章。例: 「私は自分は良い人間だと思う」「私は朝起きるのが苦手だ」

選択肢 回答者が実際に選択するもの。例: 「まったく当てはまらない」「とても当てはまる」

心理尺度というと「項目」ばかりに目が行きがちですが、尺度を作成する立場からは**教示文と選択肢もかなり重要です**。心理尺度は巻き尺や体重計のように客観的に何かを測定するための「モノサシ」です。体重計になぞらえると、教示文は使い方 (e.g., つま先立ちしない, なるべく服は脱ぐ) に相当し、選択肢は目盛り (e.g., キログラムかポンドか, 小数点以下第何位まで表示するか) に影響するという感じです。つまり、とりわけ**その心理尺度で測定された結果を他の研究と比べる**, といった用途で正しく用いるためには、どのような聞き方をしているか・選択肢は何択かなども含めてコピーする必要があります^{*3}。

また、尺度作成のためのデータ収集の際には、手順 2 で作成した項目案はもちろんですが、それと合わせて「**関連がある・ないはずの構成概念を聞く尺度・項目**」を用意しておきます。詳しくは Section C.2 で説明しますが、心理尺度の「良さ」の一つである**妥当性 (Validity)** を検証する主要な方法の一つに「別の構成概念や行動指標などとの相関を見る」というものがあります。したがって、作成した尺度が、他のどの尺度や項目とどんな相関関係にあるかを予想した上でその尺度・項目を合わせて聞いてあげます。ワークエンゲージメントで言えば、「バーンアウトと

^{*3} さらに言えば、フォントやフォントサイズ、背景色やレイアウトなど、細かい見た目も含めて統制すべきだと個人的には思っていますが、なかなか難しい話です……

は強い負の相関（対極ではないにせよ）がある」と考えられるので、MBI-GS を一緒に聞いたりすると良いわけです。MBI-GS には消耗 (exhaustion)、シニシズム (cynicism)、低い職務効力感 (reduced professional efficacy) という 3 つの構成概念があり、UWES には活力 (vigor)、熱意 (dedication)、没頭 (absorption) という 3 つの構成概念がありました。これらの構成概念の間の関係について、Schaufeli らは

- 消耗と活力は対極にある
- シニシズムと熱意も対極にある
- 職務効力感と没頭は別の次元

と考えていたので、全体として構成概念間の相関の強さが

- MBI-GS と UWES の間には負の相関がある
- MBI-GS 内の 3 つ（消耗・シニシズム・低い職務効力感）の間には正の相関がある
- UWES 内の 3 つ（活力・熱意・没頭）の間にも正の相関がある
- 活力は、MBI-GS の 3 つの中では消耗と最も強い負の相関がある
- 熱意は、MBI-GS の 3 つの中ではシニシズムと最も強い負の相関がある
- （職務効力感と没頭の間はよくわからない）

となっていれば、理論的に想定していた構成概念が測定できていることの傍証の一つにはなりそう、というわけです^{*4}。

妥当性検証の相手は心理尺度である必要はありません。例えば「ワークエンゲージメントが高い人ほど早く出社しているはず」という仮説が立てられるならば、タイムカードの始業時間との相関を見る、という方法も考えられるかもしれません。むしろ自己報告式の心理尺度同士では、方法が共通していることに起因する共分散がバイアスとして含まれる可能性（common method bias: Podsakoff, 2003）も考えられます。

4. デモグラフィック項目群の決定

ここまでで心理尺度構築のための質問紙づくりのベースは完了です。これ以降は心理尺度の作成に限らない一般的な「質問紙づくり」と共通しているので、Section C.2 にて解説することになります。ただデモグラフィック項目群（性別・年齢・職業などその人の属性に関する項目）は心理尺度作成でも特別な意味を持ちます。

尺度の性質を検証して、良い項目を選定するために用いたデータがどのような属性・割合で構成されているかは重要な情報です。例えばワークエンゲージメントの測定では、ヒラ社員と管理職の間や、男性と女性の間ではそもそも業務の性質が異なっているなどの理由から、一貫しない結果が出るかもしれません。心理尺度は、構成概念を客観的に測定するためのツールなので、基本的にはいつ誰が答えても同じ構成概念を反映した答えが返ってくることを期待されています。そのため、構成概念に対して何らかの影響を及ぼしそうな属性が想定される場合には、デモグラ

^{*4} もちろん、これだけをもって「作った尺度が想定通りのワークエンゲージメントを測れている」とは言い切れないわけですが、現在の主流の考え方では、このような傍証を集めることで妥当性を浮かび上がらせようと考えている気がします。

フィック項目としてそうした内容を聞いておき、サブグループごとに分析を行って一貫した結果が得られるかを検証することがあります。このような分析は、心理尺度の結果がどこまで一般化できるかを評価する上で重要なのです。

とはいえ、デモグラフィック項目はその人のプライベートに関する内容だったりもするため、不用意に聞きまくるのも良くないです。特に近年ではコンプライアンスが倫理的にかなり重視される用になっているため、回答者が身構えてしまうようなセンシティブな内容は避けつつ、重要そうなデモグラフィック項目を選択する必要があります。

また、回答者の属性はデモグラフィック項目以外のところからも影響を受けているかもしれません。質問紙を配布するのは、ふつう自分のツテのある場所になります。例えば大学生を対象とした調査であれば、自分の大学の中でも同じサークルの学生や、特定の講義の受講者に絞られるでしょう。また会社員を対象とした場合でも、特定の企業に限定されるかもしれません。あるいは Web 調査を行う場合でも、調査会社のアンケートモニタの登録者に限定されたり、そもそもインターネットに容易にアクセスできる層に限定されていたりする事が考えられます。

C.1.2 データが取れたら

ここからは、質問紙を使って集めたデータをもとに、心理尺度が完成するまでの残りのプロセスを見ていきます。なお、具体的な内容や方法は、本編のほうで解説しているので、ここではその概略だけを説明するにとどめておきます。

回答の処理

紙で実施した場合は、得られた回答を打ち込んでいく必要があります。回答の中には、無効回答や適当な回答があるかもしれません。分析する上ではこうした回答は除外する必要があるわけですが、尺度作成という観点で見ると**特定の項目で無回答や無効回答が多い場合には、その項目に何らかの答えにくさなどが含まれている可能性があったりします**。こうした答えにくさを持った項目は尺度から除外、あるいは聞き方を改善したほうが良いかもしれません。

基礎集計

デモグラフィック項目や各項目の回答分布を見ることで、回答者の属性に偏りが無いか、答えが特定の選択肢に集中していることがないか、などを判断します。

因子構造の検討

(詳細は 第 6 章 および 第 7 章 でたっぷり解説します) 得られた回答のデータから、事前に想定していた仮説通りの因子が確認できるかを因子分析や構造方程式モデリング (SEM) を用いて評価します。ここでは因子分析のイメージだけを説明しておきましょう。一つの尺度 (例: UWES) は通常複数の構成概念に関する項目 (活力 6 項目・熱意 5 項目・没頭 6 項目) から構成されています。これらの構成概念同士は、基本的にはあまり相関が高くはないという想定がされます (相関が高いならばわざわざ別の構成概念として取り扱う必要がないため)。一方で、同じ構成概念に関する項目の間の相関は高いはずで、したがって、項目間の相関行列は表 C.1 のよ

表 C.1: 仮想的な因子構造において想定される相関構造

	構成概念 A		構成概念 B		構成概念 C	
	項目 1	項目 2	項目 3	項目 4	項目 5	項目 6
項目 1	1	高	低	低	低	低
項目 2	高	1	低	低	低	低
項目 3	低	低	1	高	低	低
項目 4	低	低	高	1	低	低
項目 5	低	低	低	低	1	高
項目 6	低	低	低	低	高	1

うな状態になっていることが想定されます。

因子分析を用いると、データの相関行列が上のような状態にどれだけ近いかを検証できます。また SEM を使うと、異なる構成概念間の相関を「0」に置き換えても大きな問題がないかを評価したりできます。もし異なる構成概念間の相関が高ければ、そこを無理やり 0 に固定してしまうことで、いろいろな悪影響がありそうですが、もともと相関が低い箇所であれば 0 に置き換えてもあまり大きな問題はなさそうですね。

このように、因子分析や SEM を用いることで「どの項目とどの項目が同じ構成概念を測定しているのか」を統計的に判断できるわけです。

信頼性および妥当性の検討

信頼性 (reliability) は一言で言えば「測定が安定しているか」の指標です。例えば体重計は信頼性が高い測定ツールですが、体重計の信頼性が低い状態では、乗るたびに体重の値が変わってしまったり、製品ごとに異なる体重を表示したりしてしまいます。心理尺度でも同じように、信頼性が低い尺度では、同じ人が回答しても毎回異なる値が得られたり、同じ構成概念を測定しているはずの項目間で回答がばらついてしまいます。

信頼性を検証するためには、「同じ人が回答しても毎回異なる値が得られない」ことや、「同じ構成概念を測定しているはずの項目間で回答がばらついていない」ことを確認したら良いわけです。特に後者の検証が心理尺度の作成では最もよく検証される信頼性であり^{*5}、その検証には α 係数や ω 係数というものが用いられます。

また妥当性は上で説明したように、基本的には「別の構成概念との相関を見る」という形である程度検証されます。

^{*5} ここまでの記述でなんとなく感じているかもしれませんが、「信頼性」と一口に言ってもその評価方法は一つではありません。これは主に信頼性を下げうる要因の多様性に起因するわけですが……このあたりは第 4 章で。

尺度得点の使用例など（使い方）

人間の身長や体重は客観的な指標が測定可能なため、自分の身長を外国の人や 100 年以上前の人と比べることも（理論的には）可能です。これと同じように、心理尺度が客観的な指標として受け入れられるようになると、同じ尺度を用いた異なるサンプル間での比較が可能になるというメリットがあります。特に心理学では、一つ一つの研究では結果の再現性があまりよろしくないという問題が近年注目されてしまっています。研究によって得られた知見を一般化するためには、複数の研究で得られた結果を統合することで「いろいろなところで同じような結果が出た」という事実を積み上げていく必要があります。いわゆるメタ分析というやつですが、これを可能にするためには「同じモノサシで測定している」ことがほぼ必須になるわけです。

近年の尺度作成論文では、妥当性の検討なども兼ねて「作った尺度を使ってみた」的な研究例があわせて載っていることが多いです。その研究例などの中で尺度得点がどのように算出されているかをチェックしておいてください。…と言っても、ほとんどの尺度では「合計点」か「平均点」か「因子得点」が使われていると思います*6。

C.1.3 尺度開発論文を見てみよう

ここまでで、尺度作成論文を見る上で必要そうな尺度作成プロセスを概観してきました。ここからは、実際に尺度を作成したという内容の論文で、これらの内容がどのように記述されているかを見てみましょう。

尺度を探す

心理尺度を使いたいということは、何かしらの構成概念を測定したいという目的があるはずです。ここでは、「学生のリーダーシップ行動」を測定したいとして話を進めていきます。

心理尺度はほぼ確実に「ナントカ尺度 (scale)」という名前がついているはずなので、Ciniiなどで「学生 リーダーシップ 尺度」などとキーワードを入力して使えそうな尺度が開発されていないかをチェックします。基本的に最初は調査対象者の国籍（以下、日本語とします）に合わせた言語で尺度が開発されていないかを探しましょう。もし日本語では使えそうな尺度がなければ、英語でも同様に、検索キーワードに「Scale」を加えて探してみます。

なかなか見つからない場合は、以下のような方策を取ってみるのも良いかもしれません。

- 先行研究を引用している論文リストをあさる
 - 「何かしらの構成概念を測定したいという目的」が固まるまでには先行研究をいくらか見ているはずです。特に測定したい構成概念の定義について言及している文献や、その構成概念に関するレビュー論文がある場合、その構成概念を測定しようとしている尺度を作成する論文でも引用している可能性が高いので、引用している文献リストの中から探すと見つかるかもしれません。
 - Cinii や J-STAGE では被引用文献を探すのが簡単ではないのですが、Web of Science

*6 実際の尺度作成論文では、因子分析ベースで信頼性の検討やモデルフィットの確認をしているにも関わらず、普通に合計点を使っているものが多かったりします。これが理論的に問題にならないのかは少し気になるのですが、現状ではあまり気にされていないようです。

や Google Scholar では容易にリストに当たることができます。下図は、“Leadership” という書籍を Google Scholar で調べた画面です。この「引用元」というところをクリックすると、その文献を引用しているあれこれが出てきます^{*7}。その先で「引用している記事内を検索」を行うと、被引用文献リストの中から特定のキーワードで検索できるので、ここで「尺度」と入れて検索すると、“Leadership” という書籍を引用している「尺度」の論文が色々見つかるわけです。

- 心理測定尺度集を見してみる
 - 今となってはもう古いかもしれませんが、サイエンス社から「心理測定尺度集」という本が出ています。それなりに色々な心理尺度が載っているので、パラパラと見てみると使えそうな尺度や項目のインスピレーションが湧くかもしれません。
 - 英語の尺度ならば、アメリカ心理学会 (APA) が提供している PsycTests というデータベースは、心理測定を含むいろいろなテストの情報が載っているのですが、大学によっては購読していないので、気になる方は自費で APA の会員になる必要があります……

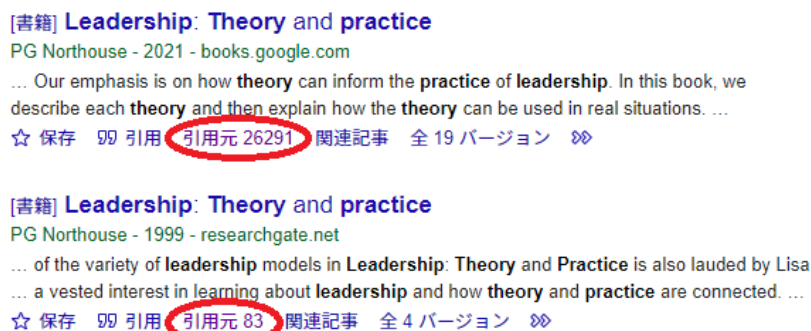


図 C.1: Google Scholar で検索

そんなこんなで、今回は以下の論文を見つけたとしましょう。

木村充, 舘野泰一, 松井彩子, 中原淳 (2019) 大学の経験学習型リーダーシップ教育における学生のリーダーシップ行動尺度の開発と信頼性および妥当性の検討. 日本教育工学会論文誌, 43(2): 105-115

中身を見ていく

【構成概念の定義】 この論文が作成しようとしている尺度は「大学の経験学習型リーダーシップ教育における学生のリーダーシップ行動」を測定するためのものだと思います。そこでこの論

^{*7} Google Scholar では学術論文以外のアヤシイものもヒットしたり、同じ文献が別のエントリーになったりすることがあるのでちょっと注意。

文では、「経験学習型リーダーシップ教育」「リーダーシップ行動」についての定義を、先行研究から導いています。詳しくは論文を見てもらうとわかりませんが、それぞれの定義は一言で抜き出すと

リーダーシップ行動 職場やチームの目標を達成するために他のメンバーに影響を及ぼす行動
経験学習型リーダーシップ教育 リーダーシップ行動を、集団活動を通して実際に行い、その行動を省察するというアプローチ

といった感じでしょうか。ここでは、リーダーシップ教育には「経験学習型」と「知識・スキル型」があり、そのうちの「経験学習型」における効果測定に用いるための尺度の作成を目指しているようです。

近年では心理尺度の測定したいものはどんどん細分化されています。研究知見が蓄積されることで事象の違いをより細かく区別できるようになっている一方で、むやみやたらに心理尺度が「乱立」している現状を嘆く人もいたりします（仲嶺& 上條, 2019）。そのため、尺度を作成するときには「既存の尺度ではだめ、ここが違うんだ!」ということをきちんと示さなければいけません。

今回の尺度は「大学の経験学習型リーダーシップ教育における学生のリーダーシップ行動」の尺度ですが、単純に「リーダーシップ」の測定だとすると、いくらでも先行事例がありそうなものです。この論文では、先行事例についてもきちんと整理した上で「他に使える尺度はない」ことを示しています。

- 経験学習型リーダーシップ教育の評価基準に関しては、データで信頼性・妥当性が検証された尺度はまだないこと
- 海外のリーダーシップ教育の尺度に関しては、翻訳の難しさや「リーダーシップ行動として何が重要か」に関する文化的差異があるため、ただ翻訳しても多分使いものにならないこと
- 企業におけるリーダーシップには上司部下の関係が入るため、参加者間が対等であるはずの大学の経験学習型リーダーシップ教育にはそのまま適用できないこと

そういうわけで、実際に「大学の経験学習型リーダーシップ教育における学生のリーダーシップ行動」尺度の作成が始まります。

【項目案の作成】 項目案は、以下の手順で作成されました。

1. 大量の先行研究から「リーダーシップ行動」の定義に当てはまりそうな項目を抽出する
2. 学生に対する自由記述アンケートで「効果的なリーダーシップを発揮するための行動はなにか」を尋ね、その回答をもとに著者が作成した項目を追加し、1. と合わせて 91 項目の素案を作成
3. 心理学・教育学を専門とするリーダーシップ教育に携わる大学教員 2 名で協議し、最終的に 54 項目にまとめた

教示文は

あなたのリーダーシップ行動についてお聞きします。あなたは、グループ活動において、以下の各行動をどのくらい行っていますか。最もあてはまるものを1つずつ選んでください

となっており、注釈としてグループ活動を『部活動、委員会活動、授業におけるプロジェクト型学習、ゼミナール、サークル、NPO、インターン等、ある程度の期間、同じメンバーで協力して、何かの目標に対して取り組むものを指す（数時間や1日だけのものを除く）』と教示しています。

また選択肢は「まったくしていない」から「いつもしている」の5件法としています。つまり選択肢はこの間にまだ3つあるわけですが、どうやらこの論文中には言及されていません。これは、本来尺度作成論文としてはあまりよろしくないことだと思います。

このように作成した尺度の情報が一部掲載されていない場合には、まず同じ著者や研究グループが行っている先行研究を見てみましょう。特に論文になる前には予備調査などで学会発表している可能性が結構あるのでそのあたりを探してみると見つかるかもしれません。あるいは著者に直接コンタクトするのも最終手段としてはあります。いずれにせよ、**元論文と同じ選択肢が用意できない場合には、元論文と同じ尺度を使っても結果の比較はできない可能性がある**という点は気をつけてください^{*8}。

【同時に尋ねる尺度】 先行研究から『効果的なチームワークがチームの目標達成を促し、リーダーシップ行動の発揮が効果的なチームワークを促すと予測される』ことから、妥当性検証のため「チームワーク状態」の尺度をあわせて尋ねています。ここでは『三沢・佐相・山口（2009）のチームワーク測定尺度のうち、因子負荷量が高い項目を各因子から3項目ずつ、18項目を抽出し、大学生のチームワークの測定に適した文言に変更して使用した』ということで、元論文とは異なる形で利用しているようです。全体としての質問紙の長さなどの理由があるのでしょうか。ただ理想的には、このように色々改変した場合にはこの改変によってもとの尺度の性能が損なわれていないかを検討する必要があります。一応この論文では、信頼性の確認として α 係数は算出されているようです。

【回答者の属性】 「方法」セクションの最初のところに、調査対象や調査手続きなどが記されています。

まず調査対象は『A 大学 B 学部の経験学習型リーダーシップ教育を経験したことがある学生』となっています。大学生を対象とする場合には、大学のレベルや規模などが構成概念に影響する可能性があります。例えば個人が主体的に学習する方法を習得できていないと思われるレベルの大学（あるいは教育困難校）では、学習方略の尺度やアクティブラーニングの尺度はかなり歪んだ結果になってしまいます。このように、尺度作成のための調査に用いられた集団が、世間一般を代表できているか、また自分の研究のターゲットの属性から大きく離れていないかはしっかりと確認しておく必要があります。今回の論文で言うと、この A 大学 B 学部では様々なリーダー

^{*8} 逆に言えば、他の研究と比較するつもりが将来にわたって起こり得ないのであれば、そこまで神経質になる必要はないかもしれません。

シップ科目が開講されているため、学生の間でも「リーダーシップとは何か」のようなものがある程度確立している可能性があります。そういったバックグラウンドを持った学生が回答しているデータをもとにこの先の分析が行われている、という点は、場合によっては気にする必要があるかもしれません。

調査手続きには以下の内容が書かれています。

調査時期 2017年11月と12月の2回。1ヶ月空けて2回実施した理由は再検査信頼性（同じ人であれば、何回答しても同じような回答が得られる、という意味での安定性）を確認するため。

調査方法 webによる質問紙調査。

有効回答数 全体数とともに、性別・学年別の回答者数が記載されている。

調査時期に関して言えば、基本的には新しい方が利用する側としては安心です。例えば「防災意識」に関する尺度が1980年に開発されていたとすると、直感的にこの尺度が問うている「防災意識」は阪神大震災や東日本大震災を期に変化している可能性が高そうです。このような感じで、古い尺度の場合にはそのまま利用可能かを注意して検討する必要があります。その尺度を利用している直近の研究例や、その尺度を引用しながら新しい尺度を提案している論文などを探してみましょう。

調査方法は回答者の属性に大きく影響します。この論文では、（たぶん著者自身の）大学の学生に対するweb調査を実施しているため回答者の層はある程度わかるのですが、通常のweb調査では調査会社などによって属性に違いがあるかもしれません（要出典）。また、同じweb調査でもクラウドソーシング系とアンケートモニタ系では属性が異なることでしょう。近年の尺度作成論文では、可能な限りこうした情報も載っているはずなのでしっかり確認しておきましょう。

有効回答数は、特定の属性の人に偏っていたりしないか・少なすぎないかなど注意してください。特に因子構造を求める場合には、回答数は多ければ多いほど良い^{*9}ものです。

【回答の処理・基礎集計】 すべての項目案について、はじめに度数分布・平均値・標準偏差を算出して、回答の分布を確認しています。いくつかの項目で若干の天井効果が見られたようですが、特に処理はしていません。

天井効果（床効果）は、回答が一番上（下）の選択肢に偏っている傾向のことです。多少の効果は気にする必要がないのですが、極端な天井・床効果が生じている項目は除外したほうが良いとされます。極端な天井・床効果が生じているということは、その項目ではほとんどの人が同じ選択肢を選んでいる、ということです。通常、心理尺度の項目に対しては図 C.2 の横軸のように、背後に連続量としての潜在特性を仮定し、各人の回答からその潜在特性の強さを推定します。つまり各回答者（A,B,Cさん）は自分の潜在特性の強さがどの程度かに応じて、適切な選択肢（1から4）を選択しているはず、と考えます。これに対して、例えば床効果が生じている状態は図 C.3 のような状態です。図 C.2 と比較して、各回答者の位置は変わっていないですが全員が1と回答しています。実際に私達が観測できるのは回答結果としての「1」のみなので、この項目に「1」

^{*9} もちろん、適当に回答する人が混じらなければ、という仮定のもとですが。

と回答したことはその回答者の潜在特性を推定する上でなんの情報も持っていないということになってしまいます。もちろんごく少数の「極端な人」を識別し「どの程度極端か」を評価する上では有用な項目なので、項目数の制約がなければ入れておいても良いのですが、項目数をなるべく少なくしたいなら、こういった項目は削除したほうが良さそう、と判断できます。また、項目反応理論などの方法では、こういった極端な項目が含まれていると他の項目の特性の推定にもあまり良くない影響を及ぼしたり、推定自体が不安定になってしまうケースがあります。そういった観点からも、極端な天井効果や床効果が見られる項目は除外したほうが良いのです。

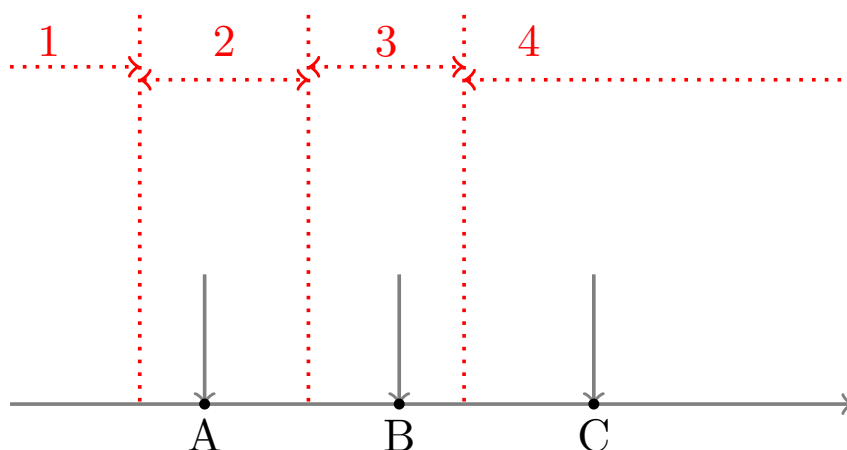


図 C.2: ふつうの項目

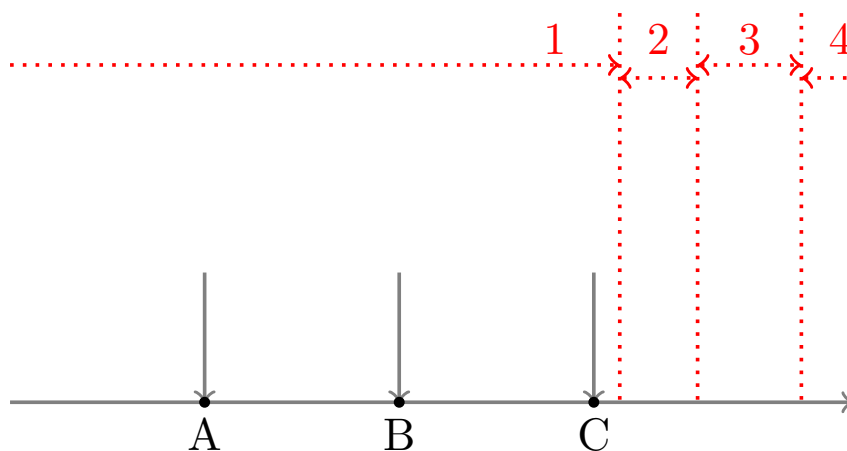


図 C.3: 床効果が生じている項目

【因子構造の検討】 細かい話は第6章 および 第7章で取り上げますが、因子構造の検討の方法は大きく分けると2種類あります。

1つ目はこの論文のように、事前の仮定はなく、データから因子構造を抽出する方法（探索的因子分析）です。表 C.1 のように、相関が高い項目同士が同じ構成概念（因子）を反映している

と考えられるため、データから得られた相関行列をもとに、どの項目同士が同じ因子を反映していると言えそうかを特定します。この場合、因子の名前は当然決まっていないので、実際に「同じ因子を反映していると言えそう」な項目を眺めて、共通点などを見つけてボトムアップに名前を決めてあげます。

2 つ目は先行研究などをもとに、どの項目がどの因子を反映しているかを決めた上で、その決定に無理がないかを検証する方法（確認的因子分析・SEM）です。この場合、因子の名前はトップダウンで決まっているはずです。

この論文では以下の手順によって、54 項目から 6 つの因子を抽出しています。

1. 探索的因子分析によって 6 つの因子を抽出した。
2. 因子の名前がつけられない（項目の内容的にキッパリと分けられない？）ので、探索的因子分析の結果から演繹的に項目を分類・因子構造を仮定した上で確認的因子分析を行った。
3. 因子負荷量が小さい項目を削除して、38 項目からなる 6 因子にまとまった。
4. まだ多いので、いい感じの項目を残して 30 項目まで減らした。

因子負荷量とは、各項目の背後にある潜在特性が回答に与える影響の強さを表したものです。因子負荷量が高い項目の場合「潜在特性が高い人ほど上の選択肢を選びやすく、潜在特性が低い人ほど下の選択肢を選びやすい」というようになる一方で、因子負荷量が小さい項目では「潜在特性の高低と選択肢の間の関係があまりない」となります。したがって、**因子負荷量が小さい項目は「その項目でどの選択肢を選んだかが潜在特性の予測にあまり意味をなさない」ということ**になるため、項目数を減らしたいならば優先的に削除される候補となります^{*10}。ただ、この論文で行っているように因子負荷量がどの程度より低かったら項目を削除するかについての明確なルールはないため、削除には慎重になったほうが良いです。

最後に『適合度』を確認し、『許容できる値が得られた』としています^{*11}。適合度とは、確認的因子分析で算出される指標です。確認的因子分析では、分散共分散行列の特定の箇所をゼロにして推定を行い、その仮定に無理がないかを検証する方法でした。この「無理がない」程度を表しているのが適合度です。適合度が許容できるということは、つまり各項目が特定の因子を反映している、という仮定には（少なくとも今回のデータでは）無理がなかった、ということになります。

【尺度得点の使用】『各因子を構成する項目の合計点を項目数で割った平均値を尺度得点とし』ています。

【信頼性の検討】 信頼性の検討には、内的整合性（すべての項目が同じような内容を問うているか）の指標として α 係数を使用しているようです。加えて、再検査信頼性を確認して、問題ない

^{*10} 項目数の制限がないならば、あったほうが良いです。というかあっても特に問題にはなりません。

^{*11} ただし一般的には、同じデータに対して探索的因子分析を行ったあとで確認的因子分析を行うことはタブーとされています。確認的因子分析は「想定していた因子構造にデータがフィットするか」を確認しているわけなので、その「想定していた因子構造」自体を同じデータから作り出せば、フィットするに決まっている、というわけです。

ということが確認されています^{*12}。

また、リーダーシップ・プログラムの受講経験の有無によって異なる因子構造が得られる可能性、つまりリーダーシップ教育を受けていない人ではリーダーシップ観が異なる可能性を考慮して、『学年を多母集団とする確認的因子分析』を行っています。このように、確認的因子分析・SEMでは特定の属性間で因子構造が異なるかどうかを検証することもできます。詳細は第7章で。

【妥当性の検討】 妥当性の中でも『基準関連妥当性』を検討するために、用意しておいた「チームワーク」との相関係数を算出し、想定通り正の相関が見られたことから、『本尺度で測定されるリーダーシップ行動は、良好なチームワークの形成を予測するものであり、内容的妥当性は支持された』としています。ただし、これだけで作成した尺度が測定しているものが「リーダーシップ行動」であるとは言い切れません。この相関は（イヤな言い方をするならば）今回作成した尺度が「良好なチームワークの形成を予測する何か」を測定している、ということの証拠の一つではないのです。

【今後の課題】 尺度作成論文に限らないですが、「今後の課題」はその研究の限界を知る上で非常に重要です。特に尺度作成論文では、**その尺度がどこまで適用可能か**に関してヒントを与えてくれるため、その尺度を別の研究に利用しようとするときには必ず目を通しておきましょう。

今回の尺度に関しては、まず第一に学生の自己評定である点が挙げられています。つまり客観的な「評価」ではないということです。例えば授業の成績をつけるためには使いづらかったりするでしょう。また、「自分はリーダーシップ行動が取れていると思っていても、実際には取れていない」といったケースも考えられます。割と質問紙全体に挙げられる課題ですが、**質問紙はあくまでも自己評価である**という点は常に気をつけましょう。

次に、一つの大学のみを対象とした研究であることも挙げられています。特に経験学習型リーダーシップ教育を取り入れている大学はまだ少ないため、教育方法も十分に確立されていなかったりするかもしれません。すると、他の大学で同じように調査を行ったときに、同じ因子構造が得られるとは限らないのです。

三点目の「項目数が多い」は一長一短なので割愛します。

まとめ

ここまで、尺度作成論文の実例を詳細に確認してみました。尺度作成論文はいわばそのモノサシの説明書的な役割を果たします。既存の心理尺度を利用する際には、ただ項目の文章をコピーしてくるだけでなく、できるだけ詳細に目を通しておくと良いでしょう。仮に先行研究と異なる結果が得られた場合でも、「ここが先行研究と違うからだ」といったことにも気づけるようになるはずです。

^{*12} 実はい内的整合性の指標としても ω という係数があるのですが、ここでは ω を再検査信頼性を表す記号として使用しているようです。ややこしい…

C.2 心理尺度・項目の作成

前節では、既存の心理尺度を利用する場合をイメージして、尺度作成論文を眺めながら心理尺度作成の大まかなフローを確認しました。ここからは、心理尺度を自分でイチから作成する場合を想定して、心理尺度の項目を作る際の注意点などを説明していきます。

C.2.1 そもそも本当に尺度を作る必要がある？

これまでに見てきたように、新しい心理尺度を作るというのはかなり大変です。少なくとも1回は尺度を作成するためだけにデータを収集する必要がありますし、その結果を踏まえて項目を修正する、または質問紙のデザインを変更するという場合には複数回データを収集する必要があります。ただでさえ心理尺度が乱立している現状において、**その構成概念を測定する・その心理尺度を作成する必要があるかはよく考える必要があります**。

仲嶺 & 上條 (2019) では、新しい構成概念を定義して心理尺度を作成する必要性は以下の3点に分類できるとしています。

関連的必要性 ある構成概念を予測する、あるいはそれと関連するために新しい構成概念が重要で、それを測定するための新しい心理尺度が必要である

弁別的必要性 既存の類似の構成概念とは異なる側面を新しい構成概念が記述しているために、その測定のための新しい心理尺度が必要である

利点的必要性 新しい構成概念のほうが既存の構成概念よりも利点があるために、その測定のための新しい心理尺度が必要である

前節で見たリーダーシップ行動の尺度の例でいうと

- **関連的必要性**：大学の経験学習型リーダーシップ教育が広がりを見せている一方で、その評価方法については十分に整備されていないため必要
- **弁別的必要性**：既存の尺度（社会人向け・海外の尺度）ではダメな理由を先行研究をもとにしっかりと説明している
- **利点的必要性**（書かれていない）

ということで、関連的必要性と弁別的必要性について明確に言及しています。尺度作成時には、まず関連的必要性については考えていることが多いかと思います。しかし、関連的妥当性は「こんな尺度を作る必要があります」ということを説明する一方で、「**既存の構成概念ではダメ・既存の構成概念より良いから新しいものを作るべきです**」ということは説明できていません。このために、新しい尺度を作成する際には関連的必要性だけでは不十分なのです。ですが近年の心理尺度作成論文では、関連的必要性のみに言及して弁別的・利点的必要性を無視しているものが多く、結果的に同じような構成概念を測定するための心理尺度が生まれている、という事態になっているようです。

前節では尺度作成論文の読み方をざっくりと説明しました。何かしらの構成概念を測定したいと考えたときには、まず第一に既存の心理尺度をこれでもかというほどにレビューし、構成概念

の定義を押さえましょう。新しい尺度作成に手を出すのは、既存の尺度ではどうしても不十分だと説明できるだけの十分な論拠が溜まってからの話です。ストップ、車輪の再発明。

C.2.2 心理尺度の項目を考える

前節では、心理尺度の項目の素案の作り方について軽く触れました。基本的には「先行研究から使えそうな文章・項目を探してくる」「自由記述アンケートの回答から抽出する」「専門家による会議によってひねり出す」のいずれかになるのではないかと思います。特に「先行研究から使えそうな文章・項目を探してくる」に関しては、他の構成概念用に作成された項目であっても、単一の項目としてみると別の構成概念にもそのまま当てはまる文章はあるはずです。とりわけ「いま作成しようとしている構成概念」と関連のある構成概念の尺度は、尺度作成の必要性の面から事前にリサーチしているはず（べき）なので、そういった情報を駆使して項目を作成してみるのが良いでしょう。

ただ項目作成について、内容面に関してはこれ以上の体系的な方法論の説明はありません。いわば職人技の世界になってきます。というのも、はじめから完璧な尺度を一発で作成するのはほぼ不可能な話です。だからこそ、項目の案を多めに作っておいて、あとから「統計的に良さそうな」項目を抽出するというやり方が一般的だと考えられます。項目案作成の段階で「専門家による会議」あるいは今では「生成 AI に作らせてみた」などという再現性の高くないやり方が事実上認められている以上、

- 構成概念の定義が先行研究に則っていて、受け入れられるものであり
- 一つ一つの項目が、構成概念の定義に照らして考えても問題なく
- 統計的な手続きによって因子構造や信頼性・妥当性がきちんと確認されている

ならば「心理尺度」は完成と見て良いのでしょう。

ということでここからは、項目作成における体系的な（あるいは小手先の）テクニックに関して紹介します。ただ、ここで説明する内容は、心理尺度に限らず社会調査などで行われる質問紙調査の項目作成全般で役に立つ可能性のある話です。

質問・教示文のタイプ

心理尺度は「項目」「教示文」「選択肢」の3つの要素で構成されています。このうち「項目」についてはこれまでも説明してきました。ここでは、目的に応じた質問あるいは教示文のタイプ（宮本& 宇井, 2014）について見ていきたいと思います。

心理尺度では複数の項目を並べて尋ねるため、共通の状況設定や条件などは「教示文」に置かれます。ここで適切に状況や場面を指定してあげることで、より目的に沿った回答が得られるようになるかもしれません。当然ですが、どのような状況を設定すべきかは構成概念の定義によって決まります。項目や教示文に関して、細かい言葉遣いの違いによっていくつかのタイプに分かれるということを知識として知っておくことで、適切な尺度づくりに役立つのではないかと思います。以下では、UWES 日本語版の「仕事に熱心である」という項目を例に、質問のタイプを見ていきます。聞き方によって得られる回答がどのように変化しそうか想像してみましょう。

【個人的質問と一般的質問】 個人的質問では回答者自身の自己評価を尋ねているのに対して、一般的質問では世間一般に対する回答者から見た評価を尋ねています。つまり聞かれている対象が異なっているということです。ワークエンゲージメントの例で言えば、

- 【個人的質問】：あなたは仕事に熱心である
- 【一般的質問】：あなたの会社の人は仕事に熱心である

という感じの違いになります。すべての項目で「あなたの会社の人は」とするのは冗長なので、教示文で「あなたの会社の人についてお聞きします。」などと書くほうが良いかもしれません。

【意見と行動】

- 【意見】：仕事に熱心であるべきだと思う
- 【行動】：私は熱心に仕事をしている

「そう思っている」と「実際にそうである」ことの間には大きな違いがあります。一般的には、行動よりも意見を尋ねたときのほうが社会的に望ましい方向に回答が歪められる傾向があるようです。

【常態的質問と実態的質問】

- 【常態的質問】：ふだん仕事に熱心であると感じている
- 【実態的質問】：ここ1週間のうちに、仕事に熱心であると感じることがあった

常態的質問では漠然とした印象を回答させていますが、人によっては適当な期間内においてそのようなことがあったかを参考にして回答しているかもしれません。また実態的質問では期間の設定が重要になります。期間が長いほど常態的質問に近づく（印象による回答になる）可能性がある一方で、期間が短すぎると「たまたまそういうことがあった／なかった」の影響を受けてしまうため、特定の心理的特性をあまり反映していない回答が得られてしまうかもしれません。心理尺度の場合は、通常すべての項目に共通で期間を設定するため、教示文で「ここ1週間のうちに、以下のようなことがありましたか」といった聞き方のほうが良さそうです。

【場面想定法】 教示文で特定の場面を設定し、その場面における感情や行動などを尋ねる方法です。例としては

あなたは入社して5年目の中堅社員です。あなたは上司から大きなプロジェクトリーダーを任されました。このプロジェクトが成功すると一気に昇進できそうですが、もし失敗すると会社が倒産してしまうかもしれません。このような状況で、あなたは以下のようなことをどの程度感じるでしょうか。

という感じで場面を設定し、項目はいつもどおり「仕事に熱心である」などを尋ねます。この方法はうまく使えど、特定の場面（e.g., 商品を選んでる時、株で大損した時）における感情や行動を知ることができる一方で、すべての人が同じような場面を想定できるのかなどの面で、

表 C.2: リッカート法の例

	1	2	3	4	5
	全くあてはまらない	あまりあてはまらない	どちらでもない	ややあてはまる	非常にあてはまる
朝起きるのがつらい	○	○	●	○	○
昼寝をよくする	○	○	○	○	●
夜、なかなか寝付けない	●	○	○	○	○
起きてすぐに朝食を食べることができる	○	○	○	○	●

難易度はかなり高いと思います。回答者からしても、そこそこ長い教示文を読まされ、自分ではない何者かの感情を押し量って答える必要があるため、結構負荷が大きい方法です。

選択肢のタイプ

「選択肢」も回答の方向づけに大きく寄与します。例えば「私は朝起きるのが苦手だ」という項目に対して「まったくあてはまらない」から「とてもあてはまる」のように感覚的な程度をグラデーションで尋ねた場合と、「まったくない」から「ほぼ毎日」のように事実としての頻度をグラデーションで尋ねた場合には、同じ人であっても異なる回答が得られるはずです。前者の尋ね方では、回答者が日頃どう思っているかによって回答が決まる一方で、後者の場合は特定の期間内（e.g., 直近1週間, 1ヶ月）において実際にそのようなこと（e.g., 寝坊した, 二度寝した）がどの程度あったかの認識によって回答が決まります。結局のところ、選択肢の作り方も構成概念の定義によって決まるわけですが、心理尺度にはどのようなタイプがあるかを知っておくと、より適切な尋ね方が見つかるかもしれません。

【リッカート法】 心理尺度で最も多いのが、リッカート法（リッカート尺度）です。皆さんも目にしたことがあるかもしれませんが、表 C.2 のような形式のものです^{*13}。この例では選択肢が5個用意されているので、「5件法（5-point scale）のリッカート尺度」などと呼ばれます。リッカート法では、回答者は各項目について一つを選択肢を選び回答します。

リッカート法では、基本的にすべての項目が同じ方向を向くように項目を並べます。ただし、表 C.2 の項目4のように、意図的に反対の意味の項目を置くことがあります。これを逆転項目と呼ぶのですが、逆転項目には「慣れ」の影響を抑制する効果があると言われています。通常、すべての項目は右に行くほど肯定的（否定的）で統一されているため、ずっと同じ方向で回答し

^{*13} 私の L^AT_EX 力不足で選択肢のラベルが横書きになっていますが通常は縦書きです。

ていると、**内容に関わらず**「この質問群なら自分はだいたい『ややあてはまる』くらいだなあ」などと考えてしまう可能性があります。人間はラクをしようとする生き物なのです。そこで、たまに逆の方向の項目を混ぜておくことで緩急がつき、一つ一つの項目に向き合ってくれるだろうということが期待されます。また、逆転項目に気づかずに『ややあてはまる』を選び続けた人がいた場合、適当に回答しているかもと判断する材料にもなりえます。

ただ、実際のところ逆転項目が全く無くてもあまり大きな問題ではないかもしれません。また、近年では逆転項目それ自体が回答に影響を及ぼす、といった研究 (e.g., [Weijters, 2013](#)) も見られるようになっており、不用意に逆転項目を無理やり入れるなどは避けたほうが良いかもしれません。ということで、**逆転項目はマストというほどのものでもない**ということ覚えておきましょう。

リッカート法のメリットは、その分析のしやすさにあります。リッカート法では、上の例のように複数の項目を尋ねた合計点などを、その人の「得点」として扱います。表 C.2 の例ならば、 $3 + 5 + 1 + 1 = 10$ 点^{*14}となります。この得点は cm や kg のような明確な単位こそ設定されていないですが、一般的には**間隔尺度として扱うことができる**ため、普通に平均値を求めたり差を取ったりできるものと「見なす」ことができるのです。ここで「見なす」と言っているのは、本質的には間隔尺度ではない可能性が否定出来ないためです。本当にこの尺度得点が間隔尺度であるためには、各項目の選択肢も全て等間隔である必要があります。つまり 図 C.4 のように各選択肢が等間隔に並んでいるという仮定が置かれます。ただし実際には、すべての回答者が選択肢を等間隔で捉えているとは言えないはずで、fig-ChapA3-likertbad のように、特に「やや」や「あまり」といったラベルから受ける印象は人によってずれている可能性があるわけです。したがって、項目単位で見ると間隔尺度ではなく順序尺度だと考えたほうが妥当な気がしてきます。それでも、複数の項目を合算した場合にはこのような選択肢の位置のズレは概ね相殺されて（誤差の中に押し込まれて）、尺度得点は間隔尺度と見なせるだろうと考えるのです^{*15}。

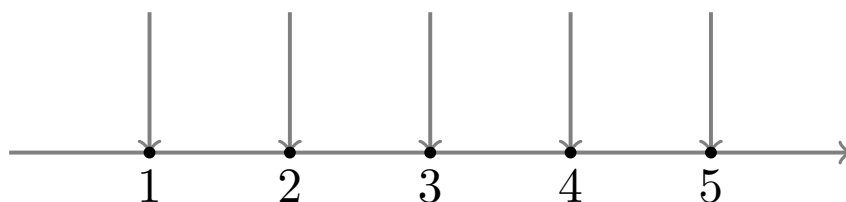


図 C.4: 正しい間隔尺度の項目

ほとんどの場合、リッカート尺度の選択肢には意味的に連続したラベル（「あてはまらない」などの言葉）がついています。意味的に連続していれば良いので、例えば UWES（日本語版）では様々な感情（項目）をどの程度の頻度で感じるかを尋ね、選択肢のラベルは

1. 全くない
2. 1年に数回以下

^{*14} 項目 4 は逆転項目なので、1 が 5 点、5 が 1 点となります。

^{*15} 現実的な問題として、順序尺度だと平均を取るなど基本的なレベルの演算すらできなくなってしまうので「多少のズレは見逃してよ」的に考えている面もあると思います。

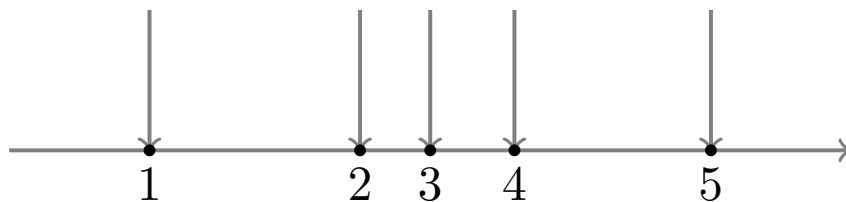


図 C.5: ある人の捉え方

3. 1 ヶ月に 1 回以下
4. 1 ヶ月に数回
5. 1 週間に 1 回
6. 1 週間に数回
7. 毎日

となっています。このラベルを見ると、@表 C.2 の例よりも一層「間隔尺度じゃない感」がありますね。

選択肢の数も回答行動に影響を及ぼしうる要因です。「あてはまらない」～「あてはまる」系のラベルを置く場合、ちょうど真ん中には「どちらでもない」が置かれるのが一般的ですが、特に中庸を好む傾向のある（ことなかれ主義的）日本人は「どちらでもない」に吸い寄せられる傾向があることが知られています（田崎& 申, 2017）。「どちらでもない」を選択することで態度の表明を放棄されてしまうことを避けるために、あえて選択肢を偶数個にして、「どちらでもない」を削除する設定もよく行われます。これにより、回答者に強制的に意思表示をさせることができるわけです。ただ、本当にどちらでもない（ちょうど中間の）人や、情報不足で回答できないような人に対してもしいずれかの態度を表明させるということが常に良いかはわかりません。あるいは回答に困った人が今度は「無回答」に流れてしまう可能性もあります。山田（2010）は、

「ほとんどの回答者が問題について考えたことがあり、かつ、答えを出せるはずである」という仮定が妥当である場合には、「どちらともいえない」という選択肢をおかないほうがよいと考えられます。（中略）これに対し、複雑な問題についての意見をたずねるような場合には、「どちらともいえない」を入れるか、あるいは（中略）両義的な選択肢を加え、さらに詳しく調査するのがよいと考えられます。

としています。ここでいう「両義的な選択肢」とは、表 C.3 のように「どちらでもない」を空白にして、そこを選択する理由を具体的に尋ねる方法のようです。

選択肢の数という点では、奇数が偶数かだけでなく、どこまで細かく尋ねるかも重要です。体重計も、1kg 刻みでしか表示してくれないものよりも 0.1kg 刻みで表示してくれるもののほうが性能が良いといえます。これと同じように、選択肢の数が増えると態度などについての程度をより細かく取ることができるため、より「良い」測定のためには選択肢は多いほうが良いと思うかもしれません。体重計と心理尺度の最大の違いは、心理尺度では回答者が自己評価をするという点です。リッカート法では通常、数字に対してラベルを付けるため、選択肢の数を多くするとそ

表 C.3: 両義的な選択肢の例

	1	2	3	4	5
	全くあてはまらない	あまりあてはまらない		ややあてはまる	非常にあてはまる
朝起きるのがつらい	○	○	(L)	○	○
昼寝をよくする	○	○	()	○	●
夜、なかなか寝付けない	●	○	()	○	○

「左右どちらともいえない場合」には () のなかに L と、「左と右の間である場合」には () のなかに M と、「考えたことがない場合」には () のなかに N と、「考えてもわからない場合」には () のなかに O と記入してください。

れだけ異なる程度表現の言葉を用意する必要があります。例えば「ややあてはまる」と「どちらかといえばあてはまる」と「すこしあてはまる」は、どの順番で並べるのが正しいのでしょうか。織田（1970）の研究では、このような程度表現について一対比較判断をしてもらい、高低の判断が人によらず概ね同じとなったもので分類を行いました。その結果、リッカート法で使われるような程度表現では、一致した区別ができるのはせいぜい5段階くらいだということがわかりました。したがって、7件法や9件法ですべての選択肢にラベルを付与した場合、人によっては選択肢のラベルの順序に違和感を覚えてしまう可能性が高くなる、ということです。

結局のところ、何件法が最適なのかについての明確な答えは出ていません。ですが、特に「どちらでもない」の有無は回答行動に大きな影響を与えそうなので、じっくり考えておくのが良さそうです。

【SD 法】 SD 法 (Semantic Differential 法) は、特定の物事に対する印象を評価するために用いられる方法です。表 C.4 のように、ヘッダーに特定の物事を呈示して、項目には**意味的に対極にある形容詞**を配置します。

基本的には「つまらない」－「楽しい」のようにそのまま解釈できる項目を用いて顧客満足度の調査などには使いやすいはずですが。その一方で、「赤い」－「青い」というようになかなか抽象的なイメージを尋ねる項目を設定することも可能ではあります。それが解釈できるかは保証しませんが……

SD 法によって得られたデータは、基本的にはリッカート法のときと同じように分析できるでしょう。データの形式がリッカート法と同じということは、SD 法も選択肢数の影響を受けます。

表 C.4: SD 法の例

	1	2	3	4	5	6	7	
【問】この授業に対して、どのような印象を持っていますか。最もあてはまるものをえらんでください。								
つまらない	○	○	●	○	○	○	○	楽しい
弱い	○	○	○	○	○	○	●	強い
小さい	○	●	○	○	○	○	○	大きい

表 C.5: VAS の例

まったく痛くない	—————	死ぬほど痛い
----------	-------	--------

一般的に SD 法では選択肢にラベルを付けることはしませんが、それでも選択肢数が奇数であれば、回答者自身の過去の（特にリッカート法の）経験から、ちょうど真ん中の選択肢を「どちらでもない」として考え、判断保留などの意味で真ん中を選ぶ人が出てもおかしくないでしょう。

【VAS】 Visual Analogue Scale (VAS) は、患者の痛みの程度を尋ねる際などによく使われる方法ですが、通常の心理尺度に対しても使おうと思えば使える方法です。得点化などにそれなりのテクニックが要するため現状ではあまり使用されていないですが……

表 C.5 のように、左端を「ゼロ」、右端を「最大」とした直線を用意し、回答者は自分の主観で程度をマークします。紙での実施ならば、定規で左端からの長さを測って得点化するため結構面倒ですが、コンピュータでの実施ならば自動的に長さを記録できるため、昔よりは VAS の使い勝手は良くなっています。また因子分析や項目反応理論の枠組みでは、VAS で得られた回答を分析可能なモデルはすでにある（e.g., Samejima, 1973; 澁谷, 2020）ので、そういった意味でも今後 VAS を使った尺度がそれなりに登場する可能性があるかもしれません。

リッカート法などと比べたときの VAS のメリットは、ラベルを設定する必要が無いという点です。つまり程度表現の受け取り方の違いを気にする必要がなくなったり、回答者の言語レベルの制限を下げて、外国人や小さい子どもに対しても実施しやすい可能性が高まります。

ワーディング

心理尺度では、項目が言葉によって呈示され、回答者がそれをどう解釈したかによって実際の回答行動が決まります。したがって、テキトウな言葉を選んでテキトウな項目を作ってしまうと、想定していたような回答が得られない可能性もあります。割と重要なことなのですが、**回答者は思っているよりも質問紙を丁寧に読み込まないものです。**というよりは、丁寧に読んでくれるだろうという思い込み自体がすでに調査者の傲慢なのです。つまり理想としては、わざわざ時間を割いて回答してくれる人に極限まで寄り添うように、努力せずとも内容がスルッと頭に入ってくるくらいのレベルを目指すべきです。構成概念的に難しいこともあると思いますが、可能な

範囲で読みやすい項目を目指していきましょう。

そこでここからは、項目作成時に気をつけるべき言葉遣い（ワーディング）について説明していききたいと思います。

【「良い」ワーディングとは】 鈴木（2016）では、ワーディングの曖昧さをなくして質問の意味を明確にするための条件として、以下の5つを挙げ、推敲のためのガイドラインを示しています*16。つまりこれに当てはまらない項目は「避けるべき項目」といえます。

1. 簡潔性

- **文末表現を統一する**：一つの尺度の中で、文章の項目（「仕事に熱心である」）と単語・体言止めの項目（「仕事に前向き」）が混ざったりすると気持ち悪いです。
- **語句の意味と用語には一貫性を持たせる**：同じものを指す言葉をコロコロ変えないようにしましょう（e.g., 「仕事」「労働」）。意図的に変える（使い分ける）場合には明確な違いを持たせて、それが回答者にも分かるようにすべきです。
- **自然で平易な表現にする**：理想は一回読めば分かる状態です。「もし俺が謝って来られてきてたとしたら、絶対に認められてたと思うか？（◎ かまいたち）」的な項目は絶対にやめましょう。

2. 客観性・中立性

- **客観的あるいは中立的な表現を使用する**：質問の中で偏見をのぞかせると、回答者がそれに引っ張られる可能性があります。
- **主観的かつ断定的な表現は使用を避ける**：「…すべきである」「…しなければならない」といった表現は客観性に欠け、不快感を招いたり回答を誘導したりする可能性があります。

3. 具体性

- **具体的に表現する**：人によって解釈が異なりうる表現は使わないようにしましょう。例えば
 - －（副詞）「ときどき」ってどれくらい？
 - －（期間）「最近」ってどこから？
 - －（きょうだい）姉妹はどうする？ 義理の関係は？
 - －（住んでいるところ）行動範囲内？ それとも市や区のレベル？
 - －（収入）年収か月収か、手取りか税込みか などに注意が必要です。
- **指示のあいまいな指示代名詞は使用しない**：いわゆる「こそあど言葉」に注意（特に教示文や表紙など）

4. 正確さ

- **句読点は正確に使用する**

*16 誤字脱字が無いのは、「もってのほか」だからなのでしょう。

- ・ 助詞や係り結びの使い方を正確にする
- ・ 修飾語と被修飾語の位置に注意する：多くの場合、修飾語と被修飾語は近いほうが読みやすいと考えられます。
- ・ 専門用語、業界用語、抽象的なことば、聞き慣れないことば、多義的なことばは使用しない
- ・ 日本語は、文章の前半より後半のほうが意味が強く伝わることを理解しておく：「この仕事は大変だが楽しい」では「楽しさ」基準で回答されますが、「この仕事は楽しいが大変だ」だと「大変さ」基準で回答されそうです。

5. 丁寧さ

- ・ 適切な敬語を使用する：項目まで敬語にする必要は無いですが、教示文や表紙、デモグラフィック項目などでは敬語にしましょう。
- ・ 話しことば、略語、流行語、俗語は使用しない：世代間で理解度の異なる言葉は特に危険です。また「公衆電話」などのように時代とともに消えゆくものを出すと、せっかく作った尺度がすぐに古くなってしまう可能性もあります。

大前提として、質問紙調査に協力してくれている方々に対する敬意を忘れないようにしてください。わざわざ時間をとって答えてくれているわけなので、少しでもストレスを無くすのが肝心です。一瞬でも項目の意味について悩ませたりしたら負けです。最悪の場合回答を放棄されてしまうかもしれません。

また、多義的な解釈の可能性については、尺度・項目作成者以外に見てもらうのが良いでしょう。新しく項目を作成したならば、少なくとも1回は（できれば繰り返し）他人に見てもらべきです。初見の人に、なるべく「ひねくれ者」の目線で見てもらう事によって、項目作成時には思いもよらなかった解釈が飛び出す可能性があります。「そんな解釈するやついるかねえ？」と思わずに、「どうあがいてもこれ以外には解釈のしようが無い」というレベルをぜひ目指してください。

【避けるべき項目の例】 避けるべき項目は他にもいくつかあります。

【誘導尋問】 聞き方次第では、回答を望ましい方向に誘導することができます。もちろん正しい回答が得られなくなってしまうのでダメなのですが、回答者が「誘導されてるな」と感じてしまうと、調査全体に不信感を抱いてしまうかもしれません。

- ・ 「熱心に仕事をするのは当然だと思いませんか。」
- ・ 「近年若者の労働意欲低下が叫ばれて久しいですが、あなたは普段熱心に仕事をしていますか」
- ・ 「ハーバード大学の教授が『意欲的に労働したほうが良い』と言っていますが、あなたはこの意見に賛成ですか」
- ・ 「あなたは夫婦別姓に賛成ですか」「…に反対ですか」→「どう思いますか」と聞いたほうが良いです
- ・ 「あなたは毎日何時間働いていますか」→働いていることを前提としているので、働いて

いない人が困ります

【虚偽回答・タテマエの回答を招きやすいもの】 質問紙調査は自己回答です。嘘をつくのも自由なので、調査者としてはなるべく「嘘をつきたくなる項目」を避けたほうが良いでしょう。

- ・道徳・倫理に関する質問では「こうあるべき」に引っ張られやすくなります（例：「信号はきちんと守っている」）
- ・プライバシーに関する質問では、逃げ道を作っておかないと嘘をつかれてしまう可能性があります。学歴や年齢など（デモグラフィック項目）を尋ねる際には「答えたくない」という選択肢を入れておきましょう。
- ・回答次第で自分に有利に働きそうな項目は危険です。（例：「今の手取り月収は少ない」という質問では、「そう思う」と答えておいたほうが何らかの動きに繋がりそう）
- ・調査の意図がバレないようにしましょう。優しい人は意図を汲んであえて嘘をついてくれるかもしれません。

【ダブルバーレル質問】 一つの項目に2つの内容が含まれている質問は、どのように答えたら良いかが一貫しなくなってしまうので避けるべきです。例えば「仕事や趣味に熱心である」という項目では「仕事には熱心ではないけど趣味には熱心」といった人が困ってしまいます。また、「あてはまる」という回答も、それが仕事についてなのか趣味についてなのか、どのくらいの割合で熱心なのかなど、解釈が難しくなってしまいます。こうした項目は、できるだけ一つずつに分解してあげましょう（例：「仕事に熱心である」「趣味に熱心である」）

また、「趣味を楽しむために仕事に熱心に取り組んでいる」といった質問もダブルバーレルです。この場合「仕事には熱心だが、別に趣味のためではない」といった人が困ってしまいます。この場合は、まず「仕事に熱心である」を聞いた上で、「あてはまる」と回答した人には続けてその理由（「趣味を楽しむためである」）を尋ねるようにするなどの方法があります。

【回答者が答えにくい質問】 その一つは「回答者に該当しない質問」です。上で挙げた「あなたは毎日何時間働いていますか」という質問には、働いていない人は答えに窮してしまいます。それだけでなく、この回答者から得られるデータは使い物にならないにも関わらず、謝礼はきちんと支払う必要が生じてしまいます。これを避けるためには、まず「あなたは働いていますか」という質問を用意するのが良いでしょう。Web 調査のアンケートモニターでは、その調査に適した属性を持っているかを尋ねる事前調査を行うことで、効率的な回答の収集を可能にしています。

もう一つは「デリケートな質問」です。答えたくない項目が登場したときには、回答者は「嘘をつく」か「無回答」のどちらかを選ぶでしょう。できればこういった項目は無い方が良いでしょうが、場合によってはどうしても必要なこともあるかもしれません。そのような場合は、事前にそのような質問があることを念押ししておき、あわせて虚偽回答よりは無回答のほうがマシなので、「回答しない」という選択が回答者の不利益にならないことを説明しておくとう良いでしょう。そのうえでどのくらいの回答が得られるかは、どれだけ丁寧に質問紙を作成できているかによって多少変わってくるかもしれません。できるだけ誠実にいきましょう。

【否定の入った質問】 例えば「働くのは嫌いだ」という項目よりも「働くのは好きではない」のほうが、文字数が多いだけでなく認知的な負荷もわずかに高くなります。できるだけ、文章の中に否定形を入れるのではなく、反対の意味の形容詞を用いるなどしましょう。

ただ、厳密に言うと「嫌い」と「好きではない」は異なるニュアンスを持っています。項目として「好きではない（＝嫌い or なんとも思っていない？）」という表現を使いたければ、無理に反対の意味の形容詞を持つてくる必要はありません。

それでも、二重否定となるとほぼ確実に止めたほうが良いといえます。例えば「熱心に働かないことは適切ではないことだと思わないですか」といった項目は、いよいよ何が言いたいのか一目ではわからなくなってしまいます。

C.3 質問紙の作成

前節では、心理尺度の項目作成に関する話をしました。ここからは、心理尺度を含んだ質問紙全体に関するお話です。基本的には回答者ファーストで答えやすいようにデザインしましょう、という話です。

C.3.1 質問紙の構成要素

まずは、質問紙全体の構成に関する話をしていきます。質問紙を構成する際の大前提として、回答者は思っているよりも質問紙を丁寧に読み込まないということを再確認しておきましょう。

教示文の役割

質問紙全体で見ると、教示文は「ガイド」の役割を持ちます。一つの心理尺度の前には「ここからはこういう状況について答えてね」という説明を行い、またある項目の前では「この項目は前の項目で●●を選んだ人だけ答えてね」と指示したりします。ただ無機質に項目を並べ立てるのではなく、それぞれの尺度やデモグラフィック項目の前などで適切に教示文を入れておくことで、回答者の認知的な負荷を軽減できると良いですね。

ただし、回答者は思っているよりも質問紙を丁寧に読み込まないものなので、可能であれば図 C.6 のように、教示文を読み飛ばしても問題ないくらいに懇切丁寧な配置にするとなお良いです。

項目の並び順

一つの尺度内の項目の並び順や、尺度の並び順は、他の項目や尺度の回答に影響を与える可能性があります。特定の質問項目の存在が、それ以降の項目への回答に影響を与える効果のことをキャリアオーバー効果と呼びますが、可能な限りキャリアオーバー効果を抑制する質問紙づくりが求められます。

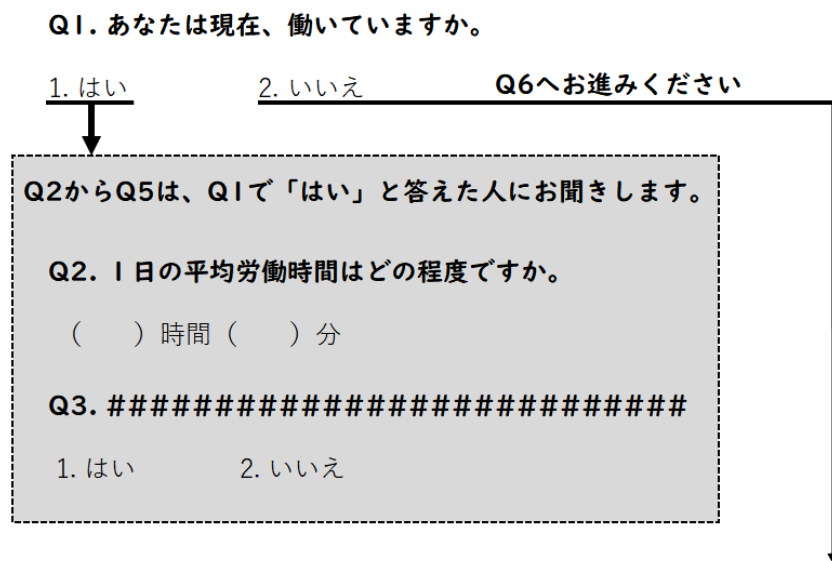


図 C.6: 矢印とか使っても良いですね

i 怪しげなアンケート? の例

Tip

2030年の冬季オリンピック（五輪）・パラリンピック開催を目指す札幌市が16日、今月実施した招致の是非を問う市民1万人アンケートの郵送調査で「賛成」と「どちらかといえば賛成」の回答が計52%だったと公表した結果に関し、先月まで行われた北京五輪で日本選手団の団長を務めた伊東秀仁氏（60）が歓迎した。

調査は3種類あり、インターネットの57%、街頭の65%が賛成。市は合計を公表していないが、6割弱になる見通しとなった。この受け止めとして「我々が戦った北京の（良い）印象もあったのかなと、うれしく思っている」と語った。

日刊スポーツ、2022/03/16

この調査で使用した調査票が2025年9月時点では閲覧可能なのですが（PDFはこちら）、中身を見るとどうにもキナ臭い感じがします。

具体的には、問3から問7で

- 札幌市では2030年大会について、「単なる一過性のスポーツイベントではなく、北海道・札幌が将来に渡って輝き続けるためのまちづくりに関するプロジェクト」としています。
- 冬季パラリンピックは北海道・札幌市で初めての開催となります。札幌市は、パラリンピックの開催が、障がいの有無に関わらず誰もが生き生きと暮らせる「共生社会」の実現に貢献すると考えています。
- オリンピック・パラリンピックを開催する際の懸念として、大会後の施設の後利用

や維持・管理についての負担があげられます。札幌市では施設整備の考え方として、「すでにある施設を最大限活用し、大会開催のための新たな施設は設けない」としています。

- 札幌市では財政面の考え方として、「大会の運営のための費用は、IOC（国際オリンピック委員会）の負担金やスポンサー収入などで賄い、原則、税金は投入しない計画」としています。
- オリンピック・パラリンピックを開催する際の懸念として、災害や感染症などのリスクがあげられます。札幌市では「新型コロナウイルスへの対応等、不測の事態への対応のため、予算の10%程度を予備費として確保する」としています。

というそれぞれの内容について、理解できたかを尋ねたあとで、問8で『北海道・札幌市で冬季オリンピック・パラリンピックを開催することを、あなたはごどう思いますか。』を「賛成」～「反対」の5件法で尋ねています。先程の『「賛成」と「どちらかといえば賛成」の回答が計52%だった』はこの問8の結果なのですが、問3から問7でオリパラの良い面ばかりを説明したあとでの問8は、本当に市民の素直な気持ちだけを反映した結果なのでしょうか…？

このように、ある質問への回答がその前の質問の内容や回答の影響を受けてしまうことは、**文脈効果** (context effect) と呼ばれたりします。心理尺度の場合には、そもそも同じ構成概念を反映した項目は当然似た内容になるため、実は文脈効果の影響が結構無視できない気もするのですが、あまり気にされていないような気がします……

鈴木 (2016) や 宮本 & 宇井 (2014) では、項目の並び順についての指針が示されていますが、ざっくりまとめると以下のような感じです。

序盤は答えやすいものを置く いきなり答えにくい質問が来ると回答者が逃げてしまいます。まずは答えやすい項目で流れをつけてから核心に迫っていくのが良さそうです。例えば意見や態度よりは、客観的な事実についてのほうが回答は容易と考えられます。

回答者の心理的な流れを重視する 例えば過去に関する質問の後で未来についての質問をおいたほうがしっくりきます。また、関連する内容の尺度は並べておくと認知的な負荷が減りそうです。ただしキャリーオーバー効果が出ないか要注意。

重要な質問は後ろに置きすぎない 後ろに置かれた質問は、前の質問の影響を受けてしまったり、回答者が疲れてしまって適当に回答する可能性が高まります。最も問いたい項目・尺度は中間あたりに置いておくと良さそうです。

プライバシーに関わる項目は最後の方がよい デモグラフィック項目を含む、プライバシーにかかわる項目は回答拒否の原因となりやすい。最後においておけば、回答拒否されてもそれまでの項目への回答はデータとして使えるので良さそう。

その他の要素

実際に回答してもらう項目の部分以外にも、気をつけておいたほうが良いことはいろいろあります。基本的に、質問紙の回答者はほぼボランティアで協力してくれています。あるいは授業で突然配られたりすると、いやいや協力している可能性すらあります。もしかしたら協力をやめる口実を探しているかもしれません。なので、とにかく**失礼のないようにしなければいけません**。なるべくそっぽを向かれないように気をつけていきましょう。

【表紙（カバーページ）】 表紙は、回答者が最初に見るものです。つまり**第一印象を決めるのはここ**です。世の中には「第一印象で n 割決まる」みたいな言葉があったりするので、ここで「答えてやってもいいか」と思わせられたらひとまずクリアでしょう。

表紙には、概ね以下の内容が書かれていることが多いようです（鈴木, 2016）。具体的な例は、ゼミの先輩が使ったものを見せてもらったりすると良いと思います。また、Google で「質問紙 表紙 テンプレ」などと調べるといくつか出てきます。絶対的な正解はないので、自分でいろいろ見てみてマネしたり反面教師にしたりすると良いでしょう。一応、図 C.7 にもテンプレを載せてみました。以下に示す内容を全て記載しようとすると、1 ページには収まらない気がするので、いろいろな表紙の例を参考にしながら、必ず書くべきこと・より重要なものを取捨選択して載せるようにしましょう。

【調査のタイトル】 タイトルはだいたい「（調査テーマ）に関する調査へのご協力をお願い」のような感じだと思います。この段階で、**回答者にある程度質問の内容を推測させてあげると良い**と思います。例えば「車の購入に関する調査」というタイトルであれば、車に興味ない人は初めから協力しないという選択ができるでしょう。一度「協力する」と決断してもらってから項目を見て「やっぱり分からん」と思わせるよりは誠実です。ただ一方で、論文のタイトルのように厳密・詳細にしすぎるとそれはそれで回答者の興味を惹かない可能性があったり、調査者の意図を読まれすぎて回答を歪めてしまう可能性があります。これを避けたり、テーマを伝えること自体が調査内容・回答行動に何らかの影響を及ぼしかねない場合には、かなりぼやかして「心理学に関する調査へのご協力をお願い」くらいにすることもあります^{*17}。

【調査実施年月日、調査対象者への挨拶】 手紙などと同じように、丁寧に書いておいて損はありません。

【調査に対する説明】

調査目的と意義 調査の目的（の要約）をタイトルよりも少し詳細に書いてあげましょう。可能であれば、その結果がどのような形で活かされるかなどがあると協力者のモチベーションが上がるかもしれません。

調査対象者を選んだ理由と手続きの説明 統計的推測を行う目的であれば、母集団からのランダムサンプリングによって、協力者を選ぶのが最も望ましいです。そのため、ここではどう

^{*17} 実際のところ、タイトルをどう設定したら回収率がどうなるか、みたいな研究は見たことがありません。（あるのかもしれませんが）

対人関係に関する調査へのご協力をお願い

本調査は、神戸大学経営学部から無作為に選ばれた 200 名を対象に、高校までの友人関係と大学での人間関係のかかわりを明らかにすることを目的として実施されるものです。本調査の結果を通して、大学での人間関係構築に関する様々な問題に対して有益な情報を提供できればと考えています。

以下の説明をお読みいただき、同意いただける場合は回答にお進みください。

- 冊子は 6 ページ、全体での所要時間はおよそ 10 分です。
- 本調査では、過去の人間関係などプライバシーに関する内容をお尋ねします。回答は強制ではないため、回答したくないと感じた場合には、回答の途中でであっても拒否することが出来ます。回答を拒否・中断することによってあなたに不利益が生じることはありません。
- 回答は完全に匿名であり、個人を特定できる情報（氏名・住所など）は収集いたしません。
- 集まった回答は統計的に処理を行い、学会発表や学術論文などの形で公表する可能性があります。
- 回答者のプライバシーの保護に配慮し、無記名によるご回答は統計的に処理して研究期間終了後にデータを破棄しますので、ご迷惑をおかけすることはありません。
- ご回答頂いた冊子は、5 月 31 日までに調査実施者にお渡しください。期限を過ぎた場合も、そのままお返しただけであれば幸いです。
- 最後までご回答いただいた方には、調査実施者に本冊子をお渡しいただく際に、50 円分の Amazon ギフト券のコードをお渡しいたします。
- ご希望の方には調査結果の概略をメールでお送りいたしますので、お手数ですが最終ページにてメールアドレスのご記入をお願いいたします。記入いただいたメールアドレスは、調査結果の報告以外の用途には使用いたしません。

お忙しいところ誠に申し訳ございませんが、ご協力をよろしくお願いいたします。

ご記入にあたってのお願い

1. この質問紙には、調査協力をお願いした本人が回答してください。
2. 質問紙は回収後すべて整理番号に従って匿名化した上で取り扱いますので、質問紙にあなたのお名前など個人が特定可能な情報を記入していただく必要はありません。
3. 質問への回答は特に指示のない限り、当てはまる回答選択肢の番号にひとつだけ○をつけてください。
4. ご記入は、鉛筆あるいは黒か青のボールペンなどをお願いいたします。

2022 年 5 月 10 日

神戸大学大学院経営学研究科

調査実施者：大学院生・分寺杏介

調査責任者：教授・神戸太郎

問い合わせ先：(メールアドレス) bunji@bear.kobe-u.ac.jp

本研究は JSPS 科研費 JP22K00000 の助成を受けたものです。

図 C.7: 表紙の例（体裁は全く整えていません）

いった属性の人を、どんな手続きで選んできたかを書いておきます。授業で配るような場合にはここは省略しても良いかもしれませんが、郵送やウェブ調査の場合には書いておくのが良いと思います。

特に対象となる人の割合が少ない、珍しい属性の母集団を想定しているためにランダムサンプリングのしょうが無い場合や、質問内容がセンシティブな場合には、選ばれた理由をしっかりと書いておくことで理解を得る必要があります。

手続きに関しては、ランダムサンプリングができていれば「無作為に選んだ」で問題ありません。基本的には、何が無作為で何が作為（対象の属性）かが明記されていれば OK です。

回答者への倫理的配慮 近年とくに厳しくチェックされるのが、倫理的な配慮です。心理学系の論文誌では、どのような倫理的配慮を行ったかを本文中などで記すことが実質的に必ず必要になっているものもあり、アンケート調査を行う場合には倫理的配慮はもう必須と言っても過言ではありません^{*18}。

倫理的配慮には、以下のような要素が書かれることが多い気がします。

- ・ 調査への協力は自発的なものであり、答えたくない場合には回答を拒否したり、途中で回答を中止してもよいこと。また、それによって協力者が不利益を被ることはないということ。
- ・ 集めた回答は匿名化処理を行い、統計的に処理されること。個人が特定される情報は収集しないこと。
- ・ 論文や学会発表で利用する場合にはその旨も示した上で、調査目的以外には使用しないこと。もし論文文化の際にオープンデータとする可能性がある場合にはその旨も示しておく。
- ・ テーマによっては、この調査への回答内容が回答者の利益・不利益にはならないことを明記しておく（授業評価アンケートなどをイメージしてください）
- ・ 調査に協力することによるリスクがあれば記しておく。リスクとは、例えば「過去の辛い出来事を思い出させる事による心理的ストレス」や「項目の内容による不快感」など。
- ・ 協力者には「結果を知る権利」があること。合わせて結果通知の方法も記しておく（調査者のウェブサイトに掲載するのか、デモグラフィック項目でメールアドレスを収集するのか）。

もちろん、調査の目的によってはあえてこれを見捨てる必要があるかもしれません（無回答を認めないなど：ウェブ調査ではよくあります）。そういった場合には、その旨を示した上で、同意書にサインをもらうようにしましょう。

質問紙のページ数・質問数・およその回答所要時間 テーマパークでは待ち時間の表示があるお陰で、並ぶか別のアトラクションに行くかを決めたり、色々と覚悟を決めることができます。これと同じで、質問紙でも**事前に全体のサイズ感を示しておく**ことで、「やってみたら思ったより時間がかかってイライラした」的な不幸を防ぐことができます。また所要時

^{*18} さらにいえば、調査計画について倫理審査委員会のチェックを受けることが必須になっている論文誌も増えています。

間が短ければ、思ったよりも多くの人が協力してくれるかもしれません。

質問数のカウント方法は人によって異なるイメージがあります。アンケート調査会社では一般的に、ブロック単位で質問数を数えます。なので心理尺度では、一つの尺度の中に 20 問あっても「1 問」としてカウントしていたり、イメージ調査のどでかいマトリックスでも「1 問」としてカウントしています^{*19}。だからといって全部の項目を「1 問」としてカウントしてしまうと、数分で終わる調査でも 100 問とかあることになってしまい、それはそれで印象が良くないでしょう。いずれにせよ、個人的には質問数を示すよりはおよその所要時間を示しておくほうが良い気がします。また、某テーマパークと同じように「最長の所要時間」を見せておけば、不満は減少すると思います。

回収方法と締め切り日等 授業で配るような場合やウェブ調査では書かなくても良いかもしれませんが、質問紙を配って各自回答してもらうような場合には必ず書いておきましょう。合わせて、締切を過ぎてしまった質問紙をどう処理するか（各自で処分してもらうか、無回答でも回収するか）も、余裕があれば書いておくとも良いかもしれません。

謝礼の内容 謝礼の内容（いくら分）および形態（Amazon ギフト券とか図書カードとか）を記します。倫理的配慮として「途中でやめても良い」とは伝えていますが、ここでは「**謝礼は最後まで回答した人のみ**」であることを明記しておきましょう。後でトラブルになります。

調査結果の使い道 倫理的配慮のところに含めても良いと思います。「論文などで公表する予定だが、その際には個人情報に配慮して…」という感じで、**具体的に決まっていなくても、可能性があるものは書いておいたほうが良い**でしょう。

助成機関名と助成内容 科研費などから調査の費用を出している場合には、その旨を書いておく必要があります。書いておいたほうがしっかりした研究感が出るので良いと思います。

【回答記入にあたっての注意事項】 質問紙調査では、回答者は基本的に最初から最後まで自分ひとりで回答する必要があるため、「これってどうするんだろう？」が発生しないように事前にフォローしておく必要があります。

- ・ 回答するのは必ず本人であること。
- ・ 空いてるスペースに名前を書かないこと。
- ・ 記入に使用する文具の指定（あれば）。特に「こすると消えるペン」は夏には勝手に消える可能性がある^{*20}ので、やめるように書いておいても良いと思います。また、色も指定しておけば、後から調査者が記入するものと区別できて良いかもしれません。
- ・ 回答方法について。リッカート式の項目であれば、一つの項目では一つの数字にマルを付けるようにする。数字の間にマルをつけたり、複数個マルを付けないように注意する。余

^{*19} これは、調査会社の料金が項目数 × 人数で決まっており、「1 問」の単位も所定のルールがあるためです。このルールが回答者の感覚とかけ離れているのは問題だと個人的には思っていますが、調査会社がどう思っているかは知りません。

^{*20} そんなに簡単に消えるものではないですが、可能性はゼロではありません。ちなみに『フリクションインキは60度以上になると無色になり、マイナス10度以下になるともとの色が復元し始め、マイナス20度前後になると完全に色が戻るという特性をもっています。』（パイロット社ホームページより）ということで、例えば炎天下の自動車の中に放置したり、直射日光が当たる窓際においていたりすると消える可能性があるので気をつけましょう。

裕があれば図 C.8 のように例を載せると良い。

- － 単答と多答が混ざっていたり、いろいろな回答方式が混ざっている場合は、各設問のところで説明するようにしたほうがよいかも。

他にも色々あるかもしれませんが、必要に応じて書いておきましょう。

各設問では、下の「正しい例」のように数字を一つだけマルで囲んでください。
複数の数字を囲んだり、数字の間にマルを書かないようにしてください。

【正しい例】

1 2 3 ④ 5

【正しくない例】

1	②	3	4	⑤
1	2	3	○	5
1	2	3	4	5

図 C.8: 回答方法の指示の例：リッカーート

【調査協力の依頼と回答前の謝辞】 「3. 調査対象者への挨拶」が「調査に興味を持ってくれてありがとう」だとすると、こちらは「調査に協力してくれて（してくれるならば）ありがとう」といった感じでしょうか。丁寧に越したことは無いですが、スペース的にきつければ「3. 調査対象者への挨拶」とどちらか一つでも良い気がします。

【調査主体の自己紹介・連絡先・所属先】 倫理的配慮の一環として「後からでも回答を取り下げできる」というようにしておくのも望ましいため、調査に何か疑問やクレームがあった際に、どこに連絡したら良いかを明記しておきます。同時に、身元を明らかにすることで、調査の信憑性が高まります。基本的には

- ・ 調査実施者（自分）と調査責任者（大学ならば指導教員）の氏名
- ・ 調査実施者のメールアドレス and/or 電話番号
- ・ （大学・会社の）住所
- ・ （大学・会社の）名前と役職

などがあれば良いかと思います。メールアドレスはプライベートなもの（Gmail など）よりも、所属機関のもの（.ac.jp など）が良いです。また（学生の場合）、調査責任者として指導教員の名前を入れておくのは、検索したときに身元がハッキリするためでもあります。もちろん指導教員に内緒で勝手に名前を入れるのはダメですよ。

【整理番号記入欄】 紙で実施した場合、集めた回答はパソコンに入力していくことになります。データ入力後に、おかしな点に気づいた場合に、元の回答が記入された質問紙を確認するために整理番号を記入する欄を作っておきましょう。整理番号には、群を表す桁などを用意しておくこととあとあとラクになります。例えば男女で比較をしたい場合には「男性は 101 から 150、女性は 201 から 250」という感じです。こうしておくことで、「整理番号の 1 桁目」を引っ張り出すだけで自動的に群分けができるようになるのです。

この段階では、整理番号はまだ記入しない方がよい場合があるため「欄」という言い方をしています。例えば授業で配布する場合に、左から配り始めたとなると、整理番号 1 番の人は多分いちばん左前に座っていた学生になるでしょう。このように、一部でも整理番号によって個人が特定できてしまう可能性が生じてしまう限りは、整理番号の記入は全ての回答が集まってからのほうがお互いに安心です。

【フォローアップ】 項目の本体およびデモグラフィック項目まですべて回答が終わったら、最後のページにはフォローアップの内容を記しておきます。図 C.9 に例を載せています。

謝辞 ここでは「回答してくれてありがとう」の謝辞です。

調査の意見や感想など 自由記述にて、調査の意見や感想などを記入できる欄を用意しておいてください。誤字や回答に困った人がいれば教えてくれるかもしれないので、調査の改善に役立つ意見が得られる可能性があります。また、調査を楽しんでもらえたりという感想をいただけることもあります。ありがたいです。

調査の真意について 例えば以下に示すような例では、表紙の時点では調査の意図がわからないように、あえて別の目的を伝えることがあります。

💡 本当の目的を伝えないケース

リッカート尺度での回答に関して、「わからない」という選択肢の位置を真ん中に置くか右端に置くかによって回答行動がどう変化するかを検証したいとする。この場合、本当の目的である『「わからない」という選択肢の位置の効果の検証』をそのまま伝えてしまうと、その事自体が回答者の「わからない」という選択肢を選ぶプロセスに何らかの影響を及ぼしかねない。

この調査では、尺度として「友人関係に関する尺度」を使用することに決めていたので、調査のタイトルを「友人関係に関する調査」として、目的では「高校までの人間関係が大学での友人関係にどのように影響するかを検証する」と書くことにした。

このような場合、そのウソがどんなウソであれ相手を騙していることには変わらない^{*21}ので、フォローアップのところで本当の目的を伝えます。

調査の意図に関して 上に関連してたまにあるのが、「この調査の意図は何だと思いましたか」

^{*21} この例のようなウソで傷つく人はいないような気もしますが、倫理審査委員会的には「騙されたと感じる人がいるかもしれない」と考えます。リスクは限りなくゼロに近づければいけないのです。

以上で質問は終わりです。ご協力ありがとうございました。

- 本調査の結果報告をご希望の方は、送付先のメールアドレスのご記入をお願い致します。間違いやすい文字にはフリガナをつけていただけると幸いです
(「エル (l)」と「いち (1)」、「オー (O)」と「ゼロ (0)」など)

[]

本調査に関してご意見・ご感想がある方は以下の欄にご記入をお願い致します。

[]

調査協力の同意後（調査実施者に本冊子をお渡しいただいた後）においても、お渡しいただいた回答は随時撤回することができます。

また、本調査および本研究の内容に関してご意見・ご質問などが生じた場合には、お気軽に調査実施者までメールでお問い合わせください。

調査へのご協力ありがとうございました。

最後に、回答の記入漏れが無いかをお確かめのうえ、5月31日までに調査実施者までお渡しください。

神戸大学大学院経営学研究科
調査実施者：大学院生・分寺杏介
調査責任者：教授・神戸太郎
問い合わせ先：(メールアドレス) bunji@bear.kobe-u.ac.jp

図 C.9: フォローアップの例（体裁は全く整えていません）

「調査の意図に気づきましたか」といった類の質問です。上の例で尋ねるならば、本当の目的を伝える前（デモグラフィック項目の前とか？）に配置するほうが良いかもしれません。これは、調査の意図に気づいた人がいた場合、それによって回答が歪んでいる可能性を考慮してのことなのかもしれません。私も見たことがあるだけで自分ではこの項目を使ったことがないのでこれ以上わかりませんが……

C.3.2 もっと細かいデザインの話

どれだけ良い質問紙ができて、回答者が正しく理解して回答してくれるかは分かりません。何度でも言いますが回答者は思っているよりも質問紙を丁寧に読み込まないので、「**読まなくても正しく回答できる**」あるいは「**何がなんでも読ませる工夫をする**」必要があります。そのためには、デザイン（や UI）の知識が参考になります。

ここからは「**伝わるデザイン**」という Web ページの情報をもとに、質問紙の「見た目」の話をしていきます。「伝わるデザイン」では基本的にプレゼン資料やポスターの作成時のポイントについて解説していますが、質問紙にも流用できるテクニックが色々あります。

読みやすさの3条件

文字や文章の読みやすさは、以下の3つの条件からなります。

- **可読性**：読みやすさ
- **視認性**：パッと見た瞬間での判断のしやすさ
- **判読性**：誤読が少ないか

基本的にいずれも重要なのですが、時と場合によってより重要なポイントが変わってきます。例えば調査の概要などを比較的長めの文章で説明するであろう表紙では可読性が重要になり、一つ一つの質問項目単位では判読性が割と重要といえます。

読みやすさを高めるために、具体的にどのようなことができるかを見ていきましょう。

書体 日本語での書体は大きく分けると「明朝体」と「ゴシック体」の2つです。

長い文章では、太い書体が続くと黒が強く疲れてしまいやすいそうです。そのため長い文章では、ゴシック体と比べると、可読性の高い（細めの）明朝体のほうが好まれる傾向にあります。

一方、ゴシック体は明朝体と比べて視認性が高く、パッと見るのに適しています。そのため質問項目や箇条書きの場面、あるいはタイトルのところにはゴシック体のほうが良いと言われています。

ただし最近では、ユニバーサルデザインに対する意識が高まっています。特に視覚過敏の人や識字障害を持つ人では、明朝体特有の「ウロコ」などのせいで不快感を覚えたりするそうです。そのため、後述する UD フォントを使用したり、長文でも細めのゴシック体を使うなどの配慮をしてあげると良さそうです。

強調 文章全体が明朝体であっても、**強調は太めのゴシック体で行ったほうが良い**とされています。明朝体は元々細めの字体なので、多少太くしたところで遠目にはあまりわからないも

のなのです。

フォント 当然ですが、きれいなフォントを選びましょう。ただし「きれい」かどうかは状況によって変わります。質問紙をどのような紙に印刷するのか（紙質・サイズなど）、あるいは Web 調査で行う＝ディスプレイに表示するのかによって最適なフォントは変わるので、必ず本番と同じレイアウトで表示して確認してください。

また、太字を使用する場合は、Word などの太字機能は使用しないほうが良いです。というのも、Word などの太字機能は元々あるフォントの周りにフチを追加して擬似的に太く見せていることがあり、この場合明らかに見栄えが悪くなります。最近は自動的に「太くするために作られた別フォント（ウェイト）」に置き換えてくれたりしますが、確実なのは**わざわざ太いウェイトのフォントを直接指定**することです。

日本語と英語を併用する場合には、日本語と英語で異なるフォントを選択することになると思いますが、違和感の無い組み合わせを選ぶようにしましょう。

判読性の観点からは、読み間違いの少ないフォントが好ましいです。英語フォントの場合、おしゃれなフォントでは“a”と“o”の区別がつきにくかったりします。また日本語フォントでも濁点・半濁点が小さいフォントがあります。基本的には自分よりも目が悪い人を想定して、画面表示を小さくしたりしながらフォントを選ぶようにしましょう。

「読みやすさ」を追求したフォントとして、**ユニバーサルデザイン (UD) フォント**が近年注目を集めています。windows にもデフォルトで「UD デジタル教科書体」や「BIZ UD」フォントが入るようになりました。読みやすさという点ではかなり良いフォントなので、選択肢の一つとして考えてみてください。

行間 狭すぎても広すぎてもだめです。ちょうどよい行間は、**フォントサイズに対して 0.7 文字分くらい**だと言われています。頑張って設定してみてください。

字間 狭すぎても広すぎてもだめです。ちょうどよい字間は、**フォントサイズに対して 5-10% くらい**だと言われています。頑張って設定してみてください。

また、タイトルなどの大きな文字で記号を使うときには細かい注意が必要になります。可能であれば、図 C.10 のように細かく調整するとベターです（が、そこまでこだわりだと時間が溶けていきます）。

あるいは、フォントをプロポーショナルフォントにしておくという手もあります。フォントには「等幅（モノスペース）フォント」と「可変幅（プロポーショナル）フォント」の 2 種類があり、プロポーショナルフォントでは文字や記号の形に合わせて、見栄えが良くなるように自動的に前後の余白を調節してくれます。ということで、読ませる文章では基本的にプロポーショナルフォントにしておくと思います。一方で、等幅フォントはプログラムコードを記述する際や、縦で文字の位置を揃えたい場合に有効です。

上で紹介した「BIZ UD」のゴシック体フォントには、「BIZ UD ゴシック」と「BIZ UDP ゴシック」の 2 種類が用意されており、後者がプロポーショナルフォントです。このように、同じ名前で「P」がついているフォントがある場合、等幅とプロポーショナルの 2 種類だと思われるので、必要に応じて使い分けましょう。

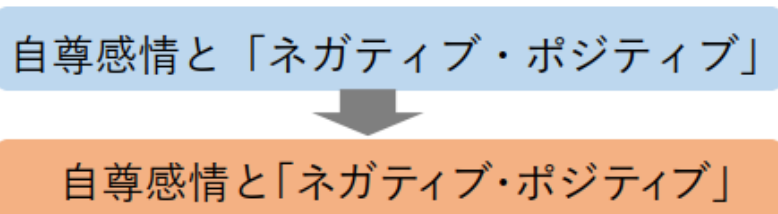


図 C.10: 上：調整前、下：カギカッコや中黒の前後のスペースを調整後

箇条書き・段落 番号付きの箇条書きの場合、図 C.11 のように文章の起点を揃えてあげると見えが良くなります。Word ならば「箇条書き」や「段落番号」の機能があるので、それを使ってあげましょう。

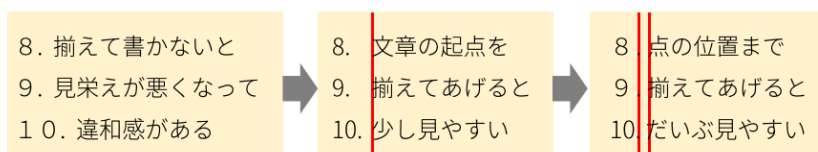


図 C.11: 箇条書きの調整

改行 できれば、単語の途中で改行しないように文章の長さを調整するとベターです。同様に、最後の行に1文字2文字だけはみ出すのもかっこ悪いとされています。

表 リッカート尺度のように、複数の項目に対して選択肢が共通の場合には、表形式にすると視認性が高まると考えられます。その表について、まず罫線は少ないほうが良いとされています。また、行の幅は地の文のときよりも少し広めのほうが視認性が高まります。心理尺度では項目を見たあとで横に平行移動して選択肢を選ぶため、横並びがわかりやすくなるように一行おきに薄い色をつけたりすると回答者に優しいデザインになりそうです。

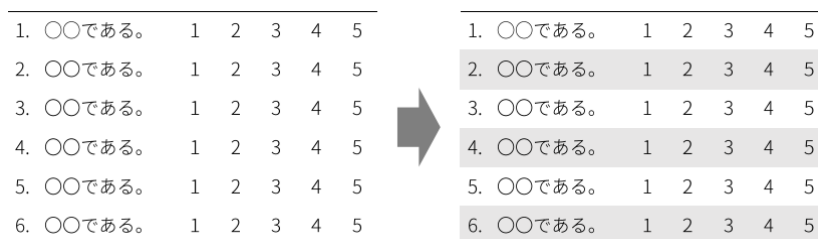


図 C.12: 見やすい表にしよう

全体の統一感 基本方針は、「統一感」です。フォントは質問紙全体で同じものを使用するし、枠や文頭、箇条書きの位置は縦でズレないようにするべきです。全体が統一感を持つと、そ

こからあえてずらすことで違和感が生まれます。意図的に違和感を生み出すことで、その部分を強調させることができるわけです。一箇所だけ太いゴシック体にしてみたり、あえて文頭を上げることでセクションの切れ目をはっきりさせたりすることができます。

その他

ここまでは文章の読みやすさにフォーカスしてきましたが、質問紙への回答しやすさという点では、他にもいくつか注意できそうなポイントがあります。

項目番号の付け方・質問紙の階層構造 階層構造の付け方は、文章の見出しの付け方と似ているところがある気がします。適切に構造化された質問紙であれば、見出し番号を見るだけで「ここまでが一つのまとまりだな」ということがハッキリします。特に複数の全く異なる内容に関して回答する必要がある質問紙であれば、見出し番号が変わるタイミングによって、「ここからは別のことを聞かれるんだな」という心の準備をさせてあげられるわけです。

また、質問番号が適切に構造化されていると、後で行うデータ分析のときにもいろいろとラクになります（正規表現が使えるようになるので）。

ページまたぎ 基本的には一つの尺度など、項目のまとまりの途中でページが切り替わらないように気をつけたいところです。といっても長い尺度ではどう頑張っても1ページには収まらない可能性があります。その場合は、2ページ目の頭にも「前ページの続きです」的な文言を入れたり、いっそ教示文を再掲しても良いと思います。また、ページ区切りが発生する場合でも、可能であれば見開きに並ぶようにするとまだマシかもしれません。

ラベルの頻度 表3.1のように、リッカート尺度では解答欄の頭の部分で、数字にラベル文を付けることが一般的です。ですが尺度が長くなると、「ラベル文はなんて書いてあったっけ？」となりやすいでしょう。A4の紙くらいならまだ良いのですが、ウェブ調査では何十問も並んでいるのに最初のところにしかラベルが書かれていなかったりして発狂するケースが実際に結構あります。ページまたぎがある場合には必ず改めてラベル文を記すべきだと思いますが、他にもページの一番下にもラベル文を記したり、ウェブ調査であれば5-10問おきにラベル文を挟むようにすると非常に親切です。

余白 上に示したページ区切りや質問紙全体の階層構造によっては、少し余白を調整してでも無理やり一つのページに同じ内容をねじ込みたくなるかもしれません。ただし、あまりに余白が少なくなりすぎるとそれだけで文字文字しくなってしまう、圧迫感が生まれてしまいます。少しだけ注意しましょう。

C.4 データの収集

前節までは、質問紙の作成までのお話でした。質問紙が完成したら、あとは配ってデータを集めて分析するだけです。ここからは質問紙調査の実施方法および、実際に行うときに気をつけるべきポイントについて紹介します。どんなに良い質問紙が完成したとしても、データの取り方を間違えるとゴミデータが集まってしまいます。

C.4.1 調査の方法

質問紙法と一口に言っても、「相手に質問に答えてもらう」方法としてはいくつかのパターンが考えられます。実施方法によって長所・短所がそれぞれあり、また注意すべきポイントも変わってきます。

以下では、心理尺度を用いた近年の質問紙調査でありえそうな4つの方法について紹介します。この4つ以外にも「電話調査」や「訪問調査」なども選択肢としてはありますが、正直いって学術研究の枠組み（特に心理尺度を使うような場合）ではほぼ行われていないと思って良いでしょう^{*22}。

集合調査

一箇所に集めた（集まった）人たちに対して質問紙を配布して実施する方法です。大学の授業などで実施するような場合がこれにあたります。

【長所】

- 一気にデータを集めることができるので、実施期間は短くて済む。
- 一度に全員に説明できるということは、教示の仕方が統一できる。
- 回収率がかかなり高い。授業で配るならばほぼ100%も狙える。
- 質問が出たときに対応できる。
- 配り方・集め方に気をつけておけば、高い匿名性を担保できる。

【短所】

- 調査を行う場所を用意する必要がある。授業で配布する場合は、担当の教員と打ち合わせしておくことになる。
- 集まった人たちは属性が偏っている可能性が結構高い。
- 回答スピードが人によって異なるので、早く終わった人はヒマになるかもしれない。

留置調査

質問紙を直接手渡しして、しばらく日にちが経ってから回収する方法です。大学の授業で配布する場合でも、次の授業のときに回収するとしたらこちらになります。

【長所】

- 直接お願いすることになるため、回収率は比較的高い。
- 自分のペースで回答できるので、回答者にやさしい。

【短所】

^{*22} 「電話調査」や「訪問調査」では調査員が直接回答者に質問して口頭で回答してもらうのですが、心理尺度のようなものを一問一問口頭で聞いていくのもいかなものか…ということでそもそもあまり相性が良くないのです。

- 自宅に赴く場合かなり大変。
- 本人が回答したという保証がない。
- 匿名性は高くない（直接のやり取りがある）ので、センシティブな質問をしにくい。
- 調査実施者がその場にいないため、わからないことがあったときにどう対応するかが読めない。

郵送調査

質問紙を郵送して、回答が完了したら返送してもらう方法です。調査協力者の住所を知っている必要があるため、特定の施設や調査会社に協力を依頼することになると思います。

【長所】

- 幅広いエリア・属性の人に質問紙を配りやすい。
- 自分のペースで回答できるので、回答者にやさしい。
- （調査会社等を介するならば）匿名性が高い。

【短所】

- 回収率はあまり高くない。
- 回収できるデータは「回答してくれるような人」という意味で属性が偏ってしまう（e.g., 学歴、調査内容への関心）可能性がある。
- 本人が回答したという保証がない。
- 発送→回答→返送を行う必要があるため、回収までに時間がかかる。
- 調査実施者がその場にいないため、わからないことがあったときにどう対応するかが読めない。

Web 調査

インターネット上に回答フォームを作成し、アクセスしてもらって回答する方法です。大きく分けると、リサーチパネルにて実施するクローズド調査と、クラウドソーシングや自身の SNS などを通じて実施するオープン調査の 2 種類があります。

【長所】

- 回収スピードが段違い。
- 幅広いエリア・属性の人からデータを集めやすい。
- やり方次第ではコストを相当抑えられる。
- データ起こしの手間が省ける。
- 匿名性は高い。
- 動画や音声を使ったり、やろうと思えばいろいろできる。
- 回答時間のデータや、操作ログなどの情報も同時に集められる。
- 間違った回答形式に対して、その場で訂正させることもできる。

- 回答の順序を厳密にコントロールできる。
- 項目や選択肢の順序をランダム化することで、回答の歪みの原因を抑制できる

【短所】

- 標本の偏りには注意が必要。高齢者ほど回答数は減少するだろうし、モニター登録している人はもともとそういうことに関心がある人だったりする。
- 適当な回答が発生しやすい。一部のプロは最低限の労力で報酬を手に入れようとしている。こういった人たちは適当に見えない（システムに弾かれない）ように適当な回答をするので、なかなか排除するのが難しい。
- 本人が回答したという保証がない。
- 調査実施者がその場にいないため、わからないことがあったときにどう対応するかが読めない。
- ネット環境や OS の違い、ディスプレイのサイズによって、項目の表示（改行の位置など）が変わってしまう可能性があり、それが回答にも影響する。

短所も色々ありますが、近年ではやはり「数こそパワー」的な風潮や、スピード主義的な風潮によって、短時間で大量のデータが収集できるウェブ調査の流れが強まっています。

C.4.2 調査対象の選び方

質問紙調査を行う目的は、基本的には標本抽出によって得られたデータをもとに母集団の性質を推定することにあります。ということは、大前提としては**なるべく母集団から無作為に抽出する**ように調査対象を選ばなければいけません。究極的には、母集団の全構成員から等確率で対象者を選定できれば良いわけですが、そんなことはほぼ不可能です。

ここからは、実際の質問紙調査で行われるであろういくつかのサンプリング方法を紹介します。いずれの方法も、完全な無作為抽出と比べると良くない点があるので、**何が良くないのかを正しく理解した上で、調査の目的に対してその問題点がクリティカルにならないようにサンプリング方法を意識できるように**なりましょう。

層化抽出法

特定の属性（年齢・性別など）に関して、母集団での比率と同じになるようにサンプリングする方法です。表 C.6 は母集団を「日本人」と設定して層化抽出を行う場合の抽出する数を表しています。日本の全人口のうち、20 代男性は 5.63% なので、抽出するサンプルが 1000 人だとすると、そのうち 56 人が 20 代男性となるように選べば良い、ということです。理想的なランダムサンプリングになるためには、さらに「日本国内の全 20 代男性からランダムに 56 人を選ぶ」必要があり、結局打つ手は無いわけですが、層化抽出法では**少なくとも層化抽出の対象の属性（表 C.6 の例で言えば年代と性別）に関しては偏りのないサンプリングが出来ます。**

表 C.6: 左：日本の人口比率、右：これに合わせて 1000 人を抽出

	男性	女性	計		男性	女性	計
70 歳以上	0.1056	0.1462	0.2518	70 歳以上	106	146	252
60 代	0.0693	0.0727	0.1420	60 代	69	73	142
50 代	0.0752	0.0754	0.1506	50 代	75	75	150
40 代	0.0836	0.0821	0.1657	40 代	84	82	166
30 代	0.0646	0.0628	0.1274	30 代	65	63	128
20 代	0.0563	0.0542	0.1105	20 代	56	54	110
15-19 歳	0.0266	0.0253	0.0519	15-19 歳	27	25	52
計	0.4812	0.5188	1.0000	計	482	518	1000

クラスター抽出法

クラスター抽出法は、母集団を属性によっていくつかのクラスターに分けた後、ランダムに選ばれたクラスターにおいて全数調査を行う、というものです^{*23}。例えば中学校では、各クラスには生徒がある程度均質になるように振り分けられている可能性が高いですね（各クラスにピアノができる子が必ず一人入るようにしていたり……）。そのため、特定のクラスの生徒全員に対して調査を行えば、学年全体に対して代表性を持っている可能性が高いかもしれません。そして一つのクラスで調査を実施するのであれば、ホームルームの時間などを利用するとなんとかなりそうです。

クラスター抽出法は、**母集団がある程度均質なクラスターに分けられる場合にはうまくいく方法**ですが、例えば上の例でも、母集団が「全国の中学生」であった場合には、偏りが生じてしまうことになります。

一つの授業で質問紙を配布するケースは、「その授業がランダムに選ばれた」と考えるとクラスター抽出法の一種と言えるかもしれません^{*24}。ただ履修する授業は個人の興味関心によって決まるため、各授業の履修者が均質になっているとは到底思えません。

応募法

特に Web 調査では、募集をかけると興味を持った人が回答をしてくれます。回答が一定数に到達すると締め切られるわけですが、こういった方法を応募法と呼ぶこともあります。

応募法では、**その調査内容に強い関心を持っている人や、報酬目当ての人に偏りやすい**ため、場合によっては結果に大きなバイアスが生じてしまうかもしれません。ただ、Web 調査の場合、どんなに上手にランダムサンプリングできたとしても、回答をお願いした人が絶対に回答してく

^{*23} ランダムに選ばれたクラスターの中から、さらにランダムに調査対象者を選び出す場合は「多段抽出法」と呼ばれたりします。

^{*24} 「たまたまその授業に出ただけ」なので、偶然法とも呼ばれたりします。「法」って名乗るほどのものでもないですが…

れるとは限らない以上、とりあえず答えてくれる人で数を集める、と言うのは仕方ない側面もあります。

ちなみに調査会社に依頼するような場合には、層化抽出法を組み合わせたデータ収集が行われていると考えられます。かなり多くのモニターに一斉にアンケートを配布した後、各層ごとに決められた必要数が満たされるまで早いもの勝ちで回答を受け付けている気がします。こうすることで、少なくとも年代と性別は母集団に近い比率にすることが出来ますが、調査対象に対する興味関心などについては事前に確認しようがないため、やはり注意が必要でしょう。

スノーボールサンプリング❶

「雪だるま」はゴロゴロ転がしていくと、新しい雪がくっついていくことで段々と大きくなっていきます。スノーボールサンプリングは、これと同じように知人の紹介などによって数珠つなぎ形式に次の回答者をリクルートしていく方法です。単純に次の回答者を紹介してもらうだけでなく、例えばリツイートによって回答フォームを拡散してもらうような場合も、回答者または調査実施者との知り合いばかりが集まっていくため、スノーボールサンプリングに該当します。

少し考えると分かることですが、**この方法では標本の代表性は基本的には満たされません**。ですが、**母集団が少ないような場合や、対象者がそう簡単には見つからないような場合（e.g., 依存症患者）**には、既に構築されているネットワークを利用することで、効率的なリクルーティングが可能となるでしょう。

機縁法

調査実施者の知り合いに調査を依頼する方法です。自身のサークルなどで質問紙を配布するような場合が該当します。スノーボールサンプリングと同様に、**この方法でも標本の代表性は基本的には満たされません**。ただ、よく知った仲の知り合いであれば属性に関してはかなり厳密なコントロールも可能かもしれません。また、知り合いだからこそ虚偽回答や適当回答をされてしまう可能性も低いでしょう。学術研究として形にする場合には機縁法はかなり厳しいですが、予備調査の段階であれば大アリな方法だと思います。

まとめ

究極的には、私たち平民には完璧な無作為抽出は不可能です。したがって、少なくとも結果に大きな問題が生じないように、妥協したり妥協しなかったりする必要があります。サンプリングにおける「大きな問題」は、ほぼ**標本の偏り**に起因します。標本の中で、特定の属性の人が必要以上に多く集まってしまうことで問題が顕在化してしまうわけですが、実際に分析内容に対して「大きな問題」とはならない偏り方というの也有ります。では、どのような場合には標本の偏りが問題視されるのでしょうか。

それは、**分析対象と強い関係があると想定される場合**です。ここでは、母集団を「大学生」と設定して、とある工学部の授業に出席している学生を対象に質問紙調査を行うケースを考えてみましょう。一般的に工学部の学生は、きっと平均的な大学生よりは新しいテクノロジーに関する関心が強いと考えられます。したがって「新しいテクノロジーの受容度」に関する心理尺度を尋ねた場合、ランダムサンプリングのときとはかなり異なる結果（平均値や因子構造）が得られる

可能性が高いと想像できます。

一方で、もし「親子関係の良好さ」に関する心理尺度を尋ねたならば、これは工学部に特異的な何かがあるとは考えにくいでしょう。もしかしたら「親と馬が合わない人ほど理系に傾きがち」といったことがある可能性はゼロではないですが、かなり無理筋な話に聞こえます。この場合、少なくとも学部による偏りは無視しても問題ないと判断される可能性が高いです。ただ、例えばこの大学は親元を離れて暮らす人の割合が高いとか、あるいは寮生活のような特殊な環境だとすると、この結果を「大学生」全体に対して一般化するのは難しいかもしれません^{*25}。

ということで、これはあくまでも個人的な考えになりますが、標本の偏りが問題になるかどうかを判断するためには、

1. まず想定しているサンプリング方法がどのような属性の偏りを生むかを想像し、
2. その属性の偏りと関連がありそうな要因を思い浮かべて、
3. その要因が調査の主目的とも関連があるかを考える

と良いのかな、と思います。このうち、1. についてはこれまでに紹介した各種サンプリング方法に関して理解をしていれば良いと思います。2-3. に関しては、ドメイン知識も要求されるでしょう。まずはヘリクツでも考えられそうな関係性を想像するところから始めると良いのかもしれませんが^{*26}。

上の工学部の例のように、場合によっては結局母集団全体に対して一般化するのは難しいかも、という話になる可能性もあります。そもそも完全なランダムサンプリングなんてものは一個人には不可能なので、どんなに頑張っても一般化の限界は必ず訪れます。そういう時には、「課題と今後の展望」的なセクションでその限界をきちんと説明できるように、事前に準備しておきましょう。

C.4.3 回答が歪むケース

ここからは、サンプリング以外のところで回答が歪んでしまういくつかのケースを取り上げてみたいと思います。（もちろん他にもいろいろな落とし穴があると思いますが）

疲れちゃった？

20 項目の 5 件法の心理尺度を 5 個載せた質問紙を配布して、回収したデータをチェックしたところ、1 個目と比べて 5 個目の心理尺度ではやけに無回答率が高いことに気づきました。

あまりに長いと途中で回答を投げ出してしまふ人が出るかもしれません。あるいは授業の最後

^{*25} 例えば神戸大生で考えると、例えば「ウォーキング・登山に対する気持ち」的なものは平均的な大学生と少し違う、かもしれませんね。好きになるか嫌いになるかはわかりませんが。

^{*26} テクニカルな対策としては、統計的因果推論を勉強すると良いと思います。統計的因果推論では、属性に偏りがある場合にどうやって補正して真の因果効果を推定するかといった方法を学ぶことが出来ます。特に既存のデータを別の目的に利用する二次分析を行う場合には、統計的因果推論の知識はかなり役立つはずです。

の10分で配布するときなど、時間の区切りがある場合には、時間が足りなくて回答できていない可能性があります。

また、無回答でなくても適当な回答が増える可能性もあったりするので、基本的な対策としては**長すぎる質問紙を作らない、十分な時間を確保する**といったあたりでしょうか。

Web調査の場合には、項目の順序自体をランダムにすることによって、特定の尺度・項目への回答が歪まないように対策することも可能です。ただしこの場合、項目の提示順がどのようになってもキャリーオーバー効果が生じないか慎重に検討する必要があります。

見られている？

「統計学の学習意欲」に関する心理尺度を作成するための調査を行うことにしました。項目作成時に協力してくれた先生が、自身が受け持つ経営統計の授業の受講生に対して調査を実施してくれると言ってくれたので、質問紙の配布と回収をお願いしました。回収したデータをチェックしたところ、どうやらみんな「統計学の学習意欲」は高いんだなあ、ということがわかりました。

いわゆる**社会的望ましさ**によるバイアスです。経営統計の授業で「統計学の学習意欲」に関する質問をすると、なんだか「意欲あるよ！」という回答をしたほうが先生の印象も良さそうですね。しかもその調査自体を先生が配って回収しているとなると、仮に無記名でもあまりネガティブな回答はしにくい感じがします。

無記名であったり、表紙などで「回答が成績に影響しない」ことを説明したとしても、やはり無意識のうちに回答は良い方向に歪んでしまうかもしれません。もちろん表紙などで十分な説明はした上で、例えば**配布から回収まで自分で行う、先生が教室を出てから行う**などの方法によって心理的安全性を確保してあげるようにするのが良さそうです^{*27}。

無回答率が高い？

「ダイエット意向」に関する調査を行うことにしました。「ダイエット意向尺度」に加えて、現状を把握したかったので「現在の体重」を回答してもらいました。回収できたデータをチェックしたところ、「現在の体重」が無回答になっている人が2割もいました。また「ダイエット意向」に関する調査をしたかったのに、メインターゲットである「現在肥満の人」が想定していたよりも少ないことが判明しました。

体重のようにセンシティブな項目は、回答拒否を招く危険性があるということは以前お話したとおりです。この例では体重を尋ねた結果、特に「現在肥満の人」が多めに無回答になってしまったようです。このような状況で、例えば「現在肥満な人ほど●●である」といった分析を行うことは、結果にバイアスを招く可能性があります。

無回答は、分析の際には**欠測値 (missing value)** として扱われます。テクニカルな話をする

^{*27} 会社のストレスチェックとかガッツリ名前が分かるようになっていたりしますが、あれに素直に回答する人ってどれくらいいるんですかね？

と欠測値には以下の 3 つのタイプがあり、そのタイプに応じた対処が求められます。

MCAR (Missing Completely At Random) とは、欠測値が完全にランダムに発生している状態です。簡単にいえばたまたま体重を記入し忘れた、という場合には MCAR として扱うことが出来ます。

MAR (Missing At Random) とは、欠測値がその変数**以外**に関連して発生している状態です。体重の例で言えば、女性の方が体重を答えるのをためらう傾向が強いと思われるので、「女性の方が体重の欠測が多い」のは MAR として扱うことができます。

MNAR (Missing Not At Random) とは、欠測値がその変数**自体**に関連して発生している状態です。体重の例で言えば、肥満傾向の人のほうが体重を答えるのをためらう傾向が強いと思われるので、「肥満の人の方が体重の欠測が多い」のは MNAR として扱われます。

具体的な対処法まで話すと長くなってしまうので省略しますが、簡単に言うと **MNAR でなければ統計的に欠測値を埋める方法があります**^{*28}。したがって、特定の属性について欠測値が多くなってもバイアスを補正した推定が可能となるわけです。ということで、質問紙調査を行う際には、MNAR が発生しそうなセンシティブな項目には特段の注意が必要となります。

いつ公開する？

「ワークライフバランスに関する調査」を、会社員を対象にクラウドソーシングで実施することにしました。頑張って回答フォームを作成し、午後 6 時過ぎに公開しました。回収したデータを確認したところ、想定していたよりもワークライフバランスが整っている人が多いということが判明しました。

クラウドソーシングを含む Web 調査では、回答は基本的に「早いもの勝ち」です。午後 6 時過ぎに公開した場合、すぐに回答できるのは午後 6 時の時点で仕事が終わっている人になるでしょう。そんなほぼ定時で仕事を上がっている人は、たぶんワークライフバランスがうまく行っている人が多いでしょう。

早いもの勝ち形式の Web 調査では、**調査を公開するタイミングも結果に影響する可能性があります**。特に仕事や家事・育児など、ライフスタイルに関する尺度・項目は注意が必要そうです。公開する時間および曜日（平日か土日か）に関して、偏りを抑制するためには、例えば 500 人集める場合でも 50 人 × 10 回に分けて様々な時間に公開する、などの対策が考えられるかもしれません。（この影響は個人的に気になっているところなので、共同研究したい人がいたらご連絡ください！）

^{*28} MCAR の場合には、最悪欠測値がある人をリストワイズ削除しても問題ありません（サンプルサイズが小さくなるだけでバイアスにはならない）。MAR の場合には、多重代入法 (MI) や完全情報最尤推定法 (FIML) を用いると、うまく欠測値を補完できるとされています。

C.4.4 まとめ

前の節では、標本の偏りが問題になるかどうかを判断する考え方として「1. まず想定しているサンプリング方法がどのような属性の偏りを生むかを想像し」と説明しました。実際には、**偏りを生じうる要因はサンプリング方法に限らず質問紙調査のあらゆるところに隠れている**のです。したがって、正しくは「1. まず想定している**調査方法**がどのような属性の偏りを生むかを想像し」となります。データが集まってから「うまくいきませんでした」ではあまりにも悲しいので、事前に考えられるあらゆるシナリオをシミュレートしておきましょう^{*29}。

そして、可能であれば予備調査を行うことで、想像していなかったトラブルのタネを炙り出してから本番に臨めると良いですね。

^{*29} 考え抜いたつもりでも、実際に調査を行うと思わぬトラブルが発生するものです。調査って難しいですね。

参考文献

- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, 38(2), 203–219. <https://doi.org/10.1007/BF02291114>
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, 18(3), 320–334. <https://doi.org/10.1037/a0032121>
- 仲嶺真., & 上條菜美子. (2019). 「心理学研究」の新心理尺度作成論文に記載された尺度作成の必要性. *心理学研究*, 90(2), 147–155. <https://doi.org/10.4992/jjpsy.90.17233>
- 佐々木研一., & 豊田秀樹. (2024). ChatGPT により生成された心理尺度項目の信頼性・妥当性の評価. *日本テスト学会誌*, 20(1), 111–133. https://doi.org/10.24690/jart.20.1_111
- 宮本聡介., & 宇井美代子. (2014). 質問紙調査と心理測定尺度—計画から実施・解析まで. サイエンス社.
- 山田一成. (2010). 聞き方の技術：リサーチのための調査票作成ガイド. 日本経済新聞出版社.
- 島津明人. (2015). ワーク・エンゲイジメントに注目した個人と組織の活性化. *日本職業・災害医学会会誌*, 63, 205–209.
- 澁谷拓巳. (2020). 連続反応を対象とするベータ反応モデルの拡張と EM アルゴリズムによる母数推定. *教育心理学研究*, 68(4), 373–387. <https://doi.org/10.5926/jjep.68.373>
- 田崎勝也., & 申知元. (2017). 日本人の回答バイアス—レスポンス・スタイルの種別間・文化間比較— . *心理学研究*, 88(1), 32–42. <https://doi.org/10.4992/jjpsy.88.15065>
- 織田揮準. (1970). 日本語の程度量表現用語に関する研究. *教育心理学研究*, 18(3), 166–176. https://doi.org/10.5926/jjep1953.18.3_166
- 鈴木淳子. (2016). 質問紙デザインの技法. ナカニシヤ出版.