

第7講 データの作成から報告書の執筆まで

田中重人 (東北大学文学部准教授)

[テーマ] 報告書の役割と書きかたを理解する

1 Google Forms の設定について

通常の社会調査では、ログイン不要で、回答者のメールアドレスなどは収集せず、送信後は編集不可にしておくのがよい

- メールアドレスは収集しない
- ログインは不要 (ユーザーを限定しない)

「プレゼンテーション」関連の設定

- 「別の回答を送信するためのリンク」は表示しない
- 「進行状況バー」を表示すると、調査票全体のなかのどのあたりを回答しているかわかるので、便利かもしれない

調査票の URL は「送信」を押すと得られる

- いきなりメールで送るのではなく、リンクのアイコンをいったん押して、フォームのリンクを取得するとよい
- 「URL を短縮」をチェックしておく、短くなるので扱いやすい

2 調査結果の取得

Google Forms 画面上部のタブを「回答」に切り替えると、回答状況がみられる。

- 「回答を受付中」を OFF にすると、回答ができなくなる

回答の全データを取得するには、次のどちらか

- 右上のメニューから「回答をダウンロード」を選ぶ CSV ファイルを ZIP 圧縮したファイルがダウンロードできる (文字コードが UTF-8 であるため、Windows の Excel などでは文字化けすることがある)
- 右上の緑の「スプレッドシート」マークをクリックすると、Google スプレッドシートで結果を開くことができる

「印刷」で、フォーム全体を印刷できる。この機能を利用して、調査票を PDF ファイルに保存することができる。

3 データ回収後の作業

3.1 ID 番号の付与

データの最初には、回答日付時刻が入っている。これとは別に、通し番号をあたえて、これで回答者を識別するとよい。

3.2 エディティング (editing)

記入漏れ、不完全な回答、誤りなどを探して訂正する。可能であれば、再調査をおこなうこともある。訂正する際には、古いファイルを残して、いつどこを変えたのかをたどれるようにしておく。

論理的にありえない回答の組み合わせや、珍しい回答についても注意する。

記入漏れの非常に多いもの、内容の信頼性が低いものは無効とする。

この時点で有効回収率を計算する。

$$\text{有効回収率} = \text{有効票数} / \text{配布票数}$$

4 コード表

4.1 コード表の作成

コンピュータでの分析のためには、データはすべて数値で表すのが便利である。回答を数値に変換するための対応表を「コード表」という。

コード表は、調査票作成と同時に作っておくとよい。ただし、調査後に自由回答のアフターコードなどを追加する場合や、データの内容をみて変数の配列を変えざるをえない場合もありえる。そのときには、コード表のほうも改訂する必要がある。

Google スプレッドシートなどのデータには回答選択肢がそのまま入っているので、それらを適切な値に変更する。

4.2 自由回答のコード

文章での回答を求めた場合や、選択肢に「その他」が入っていて具体的な内容を書いてもらった場合は、適切な数値 (code) を与える必要がある (アフターコード)。

既存のコードがある場合は、それを使えばよい。たとえば、都道府県・市町村コード、学校名コード、職業分類など。各省庁のサイトや日本工業規格 (JIS) などを探してみるとよい。

具体的な内容を見ないとコード自体が決められない場合は、つぎのようにする

- すべての回答をみてから、コードのあたえかたを決めるのがのぞましい。回答の量が多い場合には、一部だけを見て方針を決めることもありうるが、その場合、実際にコーディング作業に入ってから、対応できないケースが出てくるおそれがある。
- 「7. その他」などの場合、もとのコード「7」はそのまま残しておき、それよりも大きい数値 (たとえば11以降) をアフターコード用に使うとよい
- 分類の基準が微妙である場合には、3人以上でコーディングをおこなって、2人以上が一致したコードを採用する、などの方法をとることがある。
- 文章をいったんコンピュータ・ファイルに入力してからコーディングをおこなう方法もある。また、そのような作業をある程度自動的におこなうプログラムも存在する。

- どのような基準でコードを付与したかは、コード表に書いておくこと。
- コードはできる限り細かくしておくのがよい。あとで分析する時には「変数値の再割り当て」(SPSSの場合)などで統合することができる。

4.3 欠損値のコード

通常の調査では、欠損値としては、無回答 (NA = No Answer) によるものと非該当 (対象外の質問) によるものがある。これ以外に、「わからない」(DK = Don't Know) やクリーニング段階で欠損値を与えたものなどについても別扱いにする場合がある。

Google Forms では、答えがなかった項目は、そのセルに何も入っていない状態になる。通常は、このままでも問題ない。

欠損値の種類を区別したい場合は、「9999」「9998」などの通常は出てこない値を入れておいて、それを SPSS で処理する際に「ユーザー欠損値」として指定するとよい。

4.4 複数回答のあつかい

「いくつでも をつけてください」のような質問の回答は、ひとつの回答についてひとつ変数をつくり、つぎのような数値をあたえるのが通例である。

- 0: 丸がついていない
- 1: 丸がついている

0と1にしておく、そのまま合計 (= その項目を選択した人数) や平均 (= その項目を選択した人の比率) を使えるという利点がある。

5 分析

ダウンロードしたデータを統計ソフトに読み込んで分析

- 適当なソフトをインストールする (PSPP, R など)
- 文学部 7F のコンピュータ実習室が使える?

統計分析の方法はこの授業の対象外

6 報告書

6.1 調査結果を公表する媒体

- 報告書
- 雑誌論文
- 書籍
- プレス発表

学位論文や通常の授業レポートは、報告書と雑誌論文の間くらいか。

6.2 報告書の役割

報告書はつぎのような役割を持つ(用途別にちがう版をつくることもある)。印刷した物を関係者に配布するほか、図書館に寄贈したり、インターネットで公開したりする。

- 協力者への結果報告
- 他の研究者との情報の共有
- 分析結果を利用して論文を書く場合の参照先
- データ・チェックの基礎資料

調査の企画段階から、報告書作成のための予算・時間を見越しておくこと

7 報告書の様式

7.1 調査の概要

- 背景
- 目的
- 調査の内容
- 調査主体(場合によっては費用負担)
- 企画から調査票作成、予備調査、実施、分析までのスケジュールと内容
- 調査方法(配布・記入・回収の過程をくわしく)
- 調査母集団と計画標本の抽出法
- 回収数、回収率、非回収の原因、基本的属性の分布
- 調査協力者

文章で詳しく書いてもよいし、表の形式で簡潔にまとめてもよい。

7.2 分析結果

- 分析結果の提示(表または図)
- 分析結果の解釈(文章)
- 提言など(報告書の目的による)

7.3 資料

- 調査票
- 単純集計(度数分布表)
- あいさつ状など(あれば)
- 調査票をつくるうえで参考にした既存調査等の一覧
- コーディングやエディティングのルール
- コード表(データを公開・共有する場合; データ分析用マニュアルとして、別に用意する場合もある)

「資料」には、通常、すべての変数（ただしIDなどのケース別に固有の値のものは除く）について度数分布を表示する（巻末に「資料」などとして載せる）。カテゴリ数の多いものについては、適当な階級に分けてよい（たとえば、年齢を5歳刻みにするなど）。

度数分布表に載せる相対度数（パーセンテージ）には、欠損値を含めて計算したものと、除いて計算したものがある。どちらを使ってもよい。ただし、どのくらいの欠損値が出ているかという情報は重要なので、欠損値の種類と数は必ず表に載せること。

8 報告書本文の分析

報告書の中心部分には、データ分析をおこなった結果を載せる。調査をおこなうにあたっては、問題関心（または仮説）があるはずなので、それにしたがって分析し、結論を出すことが望ましい。

分析結果をどのようなかたちで報告書に記述するかは、想定読者によって変わる

- (1) 分析結果だけを載せる（解釈しない）
- (2) 分析結果 + 解釈
- (3) 実証研究論文としての体裁をとったもの
- (4) 政策提言や企画立案をおこなうもの

また、分析結果そのものの表示の仕方もいろいろである。

- 一般の読者相手であれば、表よりもグラフを多用して視覚的に訴えるのが望ましいことが多い。また、細かい技術的な情報は割愛して、結論がストレートにわかるようにする。
- 学術的などところで発表する場合には、表を利用して細かい数値をきちんと書く。技術的な情報を十分に書いておくこと（たとえばカテゴリーへの数値の与え方、欠損値の処理、統計的検定の方法など）

報告書は繰り返しコピーされることを念頭において作成する。網掛けやカラー印刷はなるべく使わないのがよい。グラフを使用する場合、細部まではっきり見えるかどうかに注意すること。3次元（3-D）グラフはわかりにくくなるので、使わないほうがよい。

9 この授業のレポート

締切: 8/11 (火) 17:00

提出先: Google Classroom

内容: 調査報告書（上記の(2)または(3)のかたちで分析結果を含めたもの。調査票や度数分布表などの資料もつける）

10 発表会

- 持ち時間は、それぞれ10分（説明3分 + 質疑7分）。
- 配布資料を Google Classroom で配布
- 上記の「調査の概要」にあたる情報と、調査票を必ず含めること