

## 第7講 調査票の回収とデータの作成

田中重人 (東北大学文学部教授)

[テーマ] 調査票の仕上げ ; 予備調査 ; 調査実施と回収

### 1 予備調査 (pre-test) について

本調査の前に、かならず予備調査をおこない、調査票を修正する。できるかぎり最終版の調査票と同一のものを使う。

- 質問文が適切に理解されるか
- 質問の流れに支障はないか
- 回答にかかる時間
- 回答者の負担感

対象者は少数でよい (数人から 20 人程度) が、本調査と同様の対象者を選ぶこと。回答に迷ったところや答えにくいところ、意味のつかみにくい質問文などについて、回答者の意見を集めることが目的である。このため、実際に回答してもらったあとで、かかった時間、気になったところ、疲労感などについて聞いておくこと。

### 2 調査票の校正

調査票は、調査開始前に入念にチェックする。小さなミスがのこっていると、それだけで調査が台無しになることがある。細心の注意を払って校正すること。

- ざっとながめる、画面上の配置や美しさを確認する
- 小さい単位に区切って、文字列を確認する
- 数字、記号、固有名詞などのチェック
- 印刷して、チェック済みのところに鉛筆や蛍光ペンなどでマークをつけながら校正するなど

校正作業に入る前に、しばらく調査票から離れて、内容についての記憶をできるだけうすれさせておくこと効果的。

### 3 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

## 4 データ回収後の作業

### 4.1 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。

### 4.2 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。

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

## 5 コード表

### 5.1 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 Forms の回答データには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。

### 5.2 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。

### 5.3 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、それを統計解析ソフトウェア等で処理する際に「ユーザー欠損値」として指定するとよい。

### 5.4 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。