

第8講 調査票回収からデータセット作成まで (7/6)

田中重人 (東北大学文学部准教授)

[テーマ] 調査票の回収からデータ作成までの手順を理解する

1 回収直後の作業

1.1 ID 番号の付与

調査票上部などに通し番号を振る。以降は、この ID 番号によって回答者を識別する。

1.2 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、回答とは区別できるように色ペンなどを使う。

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。

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

2 コード表

2.1 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。

2.2 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。

2.3 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。

いずれの場合も、「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。

2.4 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にしておくと、そのまま合計や平均を計算して使えるという利点がある。

3 コーディング

コード表にしたがって回答を数値化していく作業。つぎのどちらかの方式による。

- 調査票の回答選択肢の数値をそのまま使う。ただし、判断を要する場合や自由回答のコードなどは、調査票に直接書き込む（色ペンで回答と区別する）
- 専用の用紙に転記する

このコード結果をみながらコンピュータに打ち込むことになる。

4 コンピュータファイルの作成

変数名を決めておく（コード表をつくるときに決めておくとうい）

- 英字と数字のみ (A-Z 0-9 _)。ただし最初の1文字は数字を使えない
- 英字は大文字 / 小文字を区別しない
- 8文字以内
- 問番号にしたがって、Q10_4A などとつけると便利
- 転記用紙を使うときは、最初から変数名を書き込んでおくとうい

下線 (_) 以外の記号のほとんどは変数名には使えないので、注意すること。

コンピュータに打ち込むには、最初から SPSS 等で作成してもよいが、Excel 等の表計算ソフトが使いやすい。その場合、つぎのように進める。

- 最初の行に変数名を書き込んでおく
- 「ウインドウ枠を固定」しておくとうい
- 別の人に打ち込んでもらって、ファイルの相違をチェックするのがよい。Excel では、おなじファイルの複数のシートを使った計算ができる。対応するセル同士の差をとって、すべてゼロになっているかを確認する。

SPSS で分析するときは、ファイルの種類で「xls」を指定すると、Excel のファイルをそのまま読み込める。オプションで、1行目が変数名であることを指定すること。この方式では変数のラベルを読み込めないで、ラベルが必要ならあらためて SPSS の「変数」ビュー画面で入力する。SPSS で 9999 などを欠損値にするには、「変数ビュー」で、ひとつの変数について欠損値を指定しておいて、それをほかの変数のところにコピーすると、一気に処理できる。

5 クリーニング

データにみられるまちがいを摘発・修正する

- 度数分布表やクロス表を出力して、ありえない値をチェックする。コード表作成の段階で、変数の値の範囲や、変数間の依存関係をあらかじめリストアップして、コード表に書き込んでおくとよい。
- データをすべて調査票とつきあわせて、一致しているかを確認することもある。この方法の場合、データを打ち込んだのとはちがう人がおこなうのがよい。
- 調査票の回答そのものがおかしい場合にどのように修正するかについては、さまざまな立場がありうる。教科書を参照のこと。

あとで分析を進めていくうちにまちがいがみつかることもある。そのときは適宜データを修正して、データセットのバージョンを更新していく。一定の規則にしたがってファイル名をつけていくとよい。

6 課題

配布した調査票について、つぎの手順で作業し、コード表にしたがってデータセットを作成する

- 調査票とコード表の内容を確認
- IDは付与済み
- エディティング (色ペンを使う)
 - 全体をみて、おかしいところをなんとかする ある程度は原則をたてて考える
 - 問14-1: 自分が入っているか?
- コーディング
 - 住所コードの付与
 - 非該当の処理 コード表をみる
 - 無回答については、この段階で処理する必要はない
 - シートに転記

【以下は宿題】

- Excel でデータ入力
 - コード表にしたがって、1行目に変数名を入力
 - 調査票をみながらデータ入力
 - ファイルを次回までにメールで提出