

第10講 データの入力と点検

田中重人 (東北大学文学部准教授)

[テーマ] データ作成と基本的クリーニング

1 コンピュータファイルの作成

変数名を決めておく (コード表をつくる時に決めておくとよい)

- 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よい

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。

※ 最近の SPSS では、変数名に日本語文字が使えて、長くても OK ではあるが、トラブルが起こりやすい

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」しておくと便利
- 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」などを指定すると、Excel のファイルをそのまま読み込める。オプションで、1行目に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。

2 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。

3 課題

- (1) 前回作成した転記シートデータを、Excel に入力
- (2) 1 ファイルにまとめ、シート間の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
- (3) SPSS に読み込む
- (4) 9999, 9998 を「ユーザー欠損値」に指定