

第2講 PSPP 入門

田中重人 (東北大学文学部教授)

[テーマ] データ配布 ; PSPP の基礎知識

1 前回課題について

予備知識の調査について解説

2 模擬データ入力実習

配布した架空の回答票 (別紙) をもとに、データを入力してみよう。

まず PSPP を起動する。通常、「データエディタ」ウインドウの「変数ビュー」タブが表示された状態になるはず。

ここで、まず「変数」を定義する。

- 変数名を必要なだけつくる。今回は a, b, ..., e とでもしておこう。変数名は自分がわかればどんなものでもよい。日本語も使える。なお、変数名以外のフィールドは入力しなくてよい
- 書き終わったら「データビュー」タブに切り替えて、いちばん上の行に変数名がなっていることを確認する。

つづいてデータを入力していく。今回は3人分のデータを用意してあって、変数は5個なので、 3×5 の行列型のデータができるはずである。

適当な名前でも保存してみる。

- PSPP データファイル (なんとか.sav) ができていることをたしかめる。
- このデータファイルは授業終了時に削除すること。(次回以降の授業ではつかわないので、おいておく必要はない。)

この方式は PSPP でデータを入力するときのいちばん簡便な方法であるが、大きなデータはあつかいにくい。実際の調査データの入力では、Excel ファイルやテキストファイルでデータを用意しておいて、PSPP に読み込むのがふつうである。

3 データ配布

この授業で使用するのは、1995 年 SSM 調査 B 票の一部。調査については、配布資料のほか、『日本の階層システム』(2000 年、全 6 巻、東京大学出版会) を参照。

- 全国から 70 歳以下の有権者を層化 2 段無作為抽出
- 訪問面接法

ただし、配布したのはこの調査データの一部に限定したものである。

- 意識項目と基本的属性に限定 (調査票の×印はデータセットにない項目)
- 250 ケースをランダムに抽出
- 菅野剛さん (日本大学) による変数ラベルが入っている

毎回の授業で使うので、忘れないこと (調査票も)。

このデータは、この授業でのみ使用を許可されているものである。データが流出しないように注意すること。また、期末レポート提出時に、データを削除すること。

なお、自分の研究用のデータがある人は、課題などではそれを使ってもよい。ただし事前に相談すること。

4 PSPP の基礎知識

4.1 データ・セット

PSPP のデータ (「データエディタ」ウインドウで見られる) は、ケース × 変数の行列型になっている。

- 「ケース」は、個々の調査回答者にあたる
- 変数には「変数名」がついている (歴史的事情により、英数字 8 文字以内)。これだけだとわかりにくいので、変数名以外に「ラベル」をつけるのがふつう
- 無回答などの欠損値はどうなっているか?

4.2 ウインドウ構成

- データ・エディタ (上記)
- 出力ビューア (分析結果やエラーメッセージなど)
- シンタックス・エディタ (プログラムを直接編集するときに使う)

4.3 分析の一般的な手続き

「データエディタ」のメニューの使いかた

- (1) 分析手法をえらぶ
- (2) 変数を指定
- (3) 必要なオプションを指定
- (4) 「OK」をクリック

結果は別ウインドウ (出力ビューア) に表示される

- 左側に目次、右側に出力内容
- エラー表示もここに出る
- PSPP のプログラム (シンタックス) も表示される

4.4 度数分布表を出してみる

- データエディタのメニュー 「分析」 「記述統計量」 「度数分布表」
- 左側の変数リストから、分析対象とするものを選択して、右側のパレットに移動させる
- 下側の「統計」のチェックをぜんぶ外す
- 「OK」

4.5 他のアプリケーションとの連携

PSPP の出力はあまりきれいでないので、レポートを作成するときなどは、Excel や Word に表を貼り付けて整形することになる。が、出力ビューアの表をそのままコピーすることができない。

そこで、出力結果をいったん HTML で保存するなどして変換する。

- 出力ビューアのメニューから「ファイル」 「書き出し」を選択
- ファイル名の最後を「.html」にして保存
- 該当ファイルをブラウザで開く
- 該当部分をコピーして、他のアプリケーションに貼り付ける

アプリケーションの種類によっては、「オープンドキュメント」「CSV」等での保存も使える可能性がある。

川口秀樹 (2016) 「[PSPP]インポート、エクスポート、マージ」<<https://note.com/xinzuzhai/n/n63b900f0bb86>>などを参照。

5 変数値の再割り当て

ウインドウ上部のメニューバーから操作する

- 「変換」 「他の変数への値の再割り当て」
- 変換先変数の名前をつけ、「変更」を押す。名前は英数字だけにしておくのが無難 (記号や日本語を使うと、問題がおきることがある)
- 「今までの値と新しい値」の組を順次指定する。「今までの値」は範囲で指定することも、単一の値を指定することもできる
- 値の組を指定したら「続行」を押す (元の画面に戻る)
- 「OK」ボタンを押して実行する
- 出力ビューアを右端までスクロールして、新変数ができていることを確認
- 度数分布を確認
- 問題がなければ、名前をつけてデータセットを保存 (どこに保存されるかを確認しておくこと)
- 再割り当ての手順を示したシンタックスが出力ビューアに出るので、それも保存しておくこと

6 宿題

配布したデータを使い、年齢についての度数分布表を出力する。ただし、適当な年齢幅に区切ること。結果の表を、年齢幅の設定などがわかるよう整形して、どの年齢層が多いかなどのコメントをつけて提出。また、課題の途中でどこでつまずいたかなどの経過について書いてもよい。木曜日正午までに Google Classroom に提出。

ほかの人と自由に相談してよい。

教科書のほか、つぎの資料を参考にしよう。

- 小木曾道夫「SPSSの使い方」<<http://www2.kokugakuin.ac.jp/~ogiso/spss/>>
- 浦上昌則「SPSSおたすけマニュアル」<http://www.ic.nanzan-u.ac.jp/~urakami/u-spss/SPSS_f.html>
- 保田時男「SPSS操作メモ 岩井・保田(2007)準拠版」<http://www2.itc.kansai-u.ac.jp/~tyasuda/files/2013/methoda/spss_memo_2.pdf>

ただ、これらはSPSSについての説明であるため、PSPP操作とは一部くいちがいがある。
これら以外の資料を使ったときは、課題中に書いておくこと。