

第5講 クロス表分析の基礎

田中重人 (東北大学文学部准教授)

[テーマ] クロス表の書きかたと読みかた

1 前回課題について

グラフは大きく2種類に分かれる：

- (1) 一定の面積を分割して割合を示す: 円グラフ、帯グラフ、ヒストグラム、度数ポリゴンなど
- (2) 位置または長さで量を示す: 棒グラフ、折れ線グラフ、散布図など

構成比 (全部足すと100%になる) を示すには (1) のグラフを使うのが原則……であるが、実際にはそうでないことも多い。

- 円グラフは「半数」を基準としてみるときに使えるが、それ以外の目的には不適當
- 連続量の度数分布は適當な階級幅に分けてヒストグラムを書くのが本来であるが、Excel などでは描きにくい (棒グラフの距離をゼロにして見た目をヒストグラム風にする)
- 複数の分布を比較するには度数ポリゴンがよい (実際には折れ線グラフとして書く)。二つの分布の比較では棒グラフを並べてもよいが、棒の色をはっきり違えないと、識別しにくい
- 棒グラフ・折れ線グラフでは、縦軸の数値、目盛り、範囲に注意すること
- 3次元 (3D) グラフは正確な数値がつかみにくい
- Excel では、凡例や軸数値やタイトルのほか、各カテゴリの人数なども表示できる
- カラーで作成すると、白黒印刷では読みにくくなることが多い

2 データセットを分割する方法

SPSS には、特定の変数の値によってデータセットを分割するコマンドがある

- メニューから「データ」→「ファイルの分割」を選ぶ
- 適當な変数を選び、「グループの比較」を選び、OK

いったんこの操作をすると、それ以降は、すべての分析が、その変数の値ごとに別々におこなわれる。元に戻すときは、「データ」→「ファイルの分割」→「すべてのケースを分析」

3 今回の課題

「性別」と「性別による不公平」について、次の手順で「クロス表」(cross table)を作成する：

- (1) メニューから「分析」→「記述統計」→「クロス集計表」
- (2) 適当な変数を「行」「列」にひとつずつ指定
- (3) 「セル」にパーセンテージ(行・列の両方)を追加

出力を元に、次のことを考える(参照：教科書第4章)

- この表から何がわかるか
- 「行」の%と「列」の%は何を表しているか。またこのクロス表を解釈するときにはどちらを見るのが適切か
- このクロス表をわかりやすく表示するにはどのようなグラフが適切か考え、実際に作成してみる(Excelを使用)

提出は、ISTU で水曜日正午まで。