

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 前回宿題 1 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%) を示すには (1) のグラフを使うのが原則.....だが、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
- 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 度数にするかパーセントにするか
- 複数の分布の比較で棒グラフを並べる場合、棒の色をはっきり違えないと、識別しにくい
- 複数の分布 (特に 3 つ以上) 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)
- 人数がゼロのところがあるので注意
- 欠損値のあつかい
- 連続量の度数分布は適当な階級幅に分けて「ヒストグラム」(柱状グラフ histogram) を書くのが本来である

いったん HTML ファイルに出力し、ブラウザで開いて Excel にコピーする手もある (かなりめんどくさい)

- 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
- 3次元 (3D) グラフは正確な数値がつかみにくい
- ヒストグラムは Excel 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)

2 前回宿題 2 について

- (1) 「データの種類」の分類について

- 尺度水準によって使える計算方法が違う (= 使える分析法が違う) ことを理解しておくこと
- 測定している対象そのものの性質ではなく、データにおいてどのような数値が割り当てられているかが問題である
- 上位の尺度は下位の尺度の性質を兼ね備える (たとえば間隔尺度のデータは順序尺度としても分析できる)

- (2) SSM 調査の質問項目のうち、比率尺度に当たるものはどれか 人数、年数など

- (3) 「中央値」「四分位」などに意味があるのはどの尺度水準か 順序尺度以上

(4) 「収入」や「学歴」を比率尺度として分析するにはどのようにすればよいか 「変数値の再割り当て」で適当な値に変換：

- 「収入」については、各カテゴリを適当な金額に変換すればよい (たとえば各階級の真ん中の値をとる)
- 「学歴」については、その学歴を取得するのに必要な標準的年限で置き換えることが行なわれている (単に「教育年数」とよぶことが多い)

1	6	
2	8	
3, 4, 5		11
6	14	
7	17	
12	9	
13	12	
14	14	
15	16	
16	18	

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) 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 等を使用)