

第4講 度数分布表とグラフの利用

田中重人 (東北大学文学部准教授)

[テーマ] グラフの種類と、それらの適切な使い分け

1 前回宿題について

(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) 累積パーセント (つまり中央値や四分位など) 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	

2 グラフの利用

分析結果は、通常、表またはグラフで示す。

表 (table): 正確な数値がわかるが、全体の傾向を読み取るには熟練が必要

グラフ (graph/chart): 全体の傾向が簡単に読み取れるが、正確さは犠牲になる

初心のうち、表とグラフの両方を作成して読んでいくのがよい

3 度数分布 (の比較) をあらわすグラフの種類

- パイチャート (円グラフ)半数を超えているかの判別に便利
- バーチャート (棒グラフ)離散量のそのままの分布を示す
- ヒストグラム連続量を階級に区切って示す
- 度数ポリゴン (度数多角形)複数の分布の比較に便利 (教科書 p. 34)
- 帯グラフ (積み上げ棒グラフ) (教科書 p. 106)

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

4 グラフを描く方法

4.1 PSPP でグラフを描く

- 度数分布表のオプションで「チャート」を選び、「パイチャート」「バーチャート」「ヒストグラム」のなかから選択。尺度を「度数分布」または「パーセンテージ」にする。
- データエディタのメニューから「グラフ」「バーチャート」を選び、「カテゴリ軸」に変数を指定。「バーの表示」を「ケース数」または「ケースのパーセント」に。

4.2 他のソフトを使う

いったん HTML ファイルに出力 (第 2 講資料など参照) し、ブラウザで開いてコピーする手もある (かなりめんどくさい)。

Google スプレッドシートでは、データ範囲を選択して「挿入」「グラフ」。「設定」でグラフ種類など選択し、「カスタマイズ」でこまかい設定を行う。

- 軸に入れる数値は「グリッドラインと目盛」で調整する
- 各カテゴリの人数など表示するには「系列」「データラベル」
- 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

5 宿題

つぎの3種類の度数分布について、PSPPで適切なグラフを描け。

- (1) 男女比
- (2) 本人年収の分布
- (3) 本人年収の分布の男女比較

教科書 32-37 ページを参照。

PSPPを使う場合、出力ビューアのメニュー「ファイル」 「書き出し」でファイル名の末尾を「.pdf」にするとPDF形式のファイルを作成できる。