

第3講 統計分析の基礎

田中重人 (東北大学文学部教授)

[テーマ] 度数の利用と統計分析の基礎

1 前回課題について

操作については 前回資料 後半部分を参照

- 用紙上部に番号・名前を記載
- 70代はなぜ少ないのか
- 年齢層別の人口の変動 1995年の人口ピラミッド <<http://www.ipss.go.jp/site-ad/TopPageData/1995.png>> と出生力の変動 <http://www.ipss.go.jp/syoushika/tohkei/Popular/P_Detail2020.asp?fname=G04-01.gif>
- 変数ラベルの利用：値の再割り当ての変数を指定する際に、変数名とは別に「ラベル」をつけることができる。あとから「データビュー」の「変数ビュー」でも
- 値ラベルの利用：「データビュー」の「変数ビュー」タブで、変数の値に「ラベル」をつける 分析結果出力に表示される
- シンタックス (syntax) の利用

2 度数分布表の読みかた

- 度数
- 相対度数 (%)
- 累積度数・累積相対度数
- 欠損値のあつかい

(教科書 p. 27-31)

3 データセットを分割する方法

PSPP 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
- 適当な変数を選び、「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」

4 データ収集から分析まで

- (1) データの収集 (実験 / 観察)
- (2) 分析可能な形に加工
- (3) データ・セット作成
- (4) クリーニング
- (5) データの特徴を少数の数値に要約 = 記述統計
- (6) 誤差の評価 (この手続きの一部が推測統計)

(教科書 p. 1-6)

5 標本抽出

標本抽出の4段階モデル

- 理論母集団 = 興味の対象となる人や事物の全体
- 調査母集団 = 調査の対象とする具体的な範囲
- 計画標本 = 母集団から抽出した対象者のこと
- 有効標本 = 調査の結果あつまった有効なデータ

「無作為抽出」(random sampling) とは：

- 母集団から計画標本を選ぶ際に、母集団にふくまれるすべての個体の抽出確率が等しくなるように抽出する
- この結果として、「確率標本」(probability sample) がえられる

統計的な推測のための理屈は、確率標本を前提として組み立てられている。母集団の人口がわかっていて、全個体を網羅した台帳がないと、無作為抽出はできない。実際にはそういうことはないので、いろいろ工夫して無作為抽出に近づける。

「層化2段無作為抽出」はその方法のひとつ：

- まず「地点」を抽出 (第1次抽出)
- その際、地域・都市規模等で地点抽出数を割り当てておく (層化)
- その地点の台帳から個人を抽出 (第2次抽出)

6 宿題

- (1) 教科書 pp. 7-16 を元に、「データの種類」の分類 (名義尺度、順序尺度、間隔尺度、比率尺度) についてまとめよ
- (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
- (3) 累積パーセントに意味があるのはどの種類のデータか
- (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

木曜正午までに Google Classroom に提出。