

現代日本論演習 / 比較現代日本論研究演習 I 「統計分析入門」(2013)

## 第3講 統計分析の基礎 (4/24)

田中重人 (東北大学文学部准教授)

[テーマ] 累積度数の利用と統計分析の基礎

### 1 課題

#### 1.1 前回課題について

カテゴリ統合の方針：

- 何かの意味があるところで区切る
- 人数が均等になるように区切る

#### 1.2 今回の課題

年齢 (カテゴリ統合していない元の変数) の度数分布から、中央値と四分位を求めよ (提出は不要)。

参考資料：

- 教科書 p. 43
- 船津好明「統計計算の方法」 <<http://www.waq.jp/stacal.htm>>
- 総務省統計局「How to 統計」 <<http://www.stat.go.jp/howto/lecture4/01.htm>>

また、任意のパーセンタイル (percentile) を求める方法を考えること。

#### 1.3 発展問題 (余裕のある人のみ)

次の情報を参考にして、カテゴリ統合した後の年齢の度数分布表から中央値を求める方法を考える (森・吉田 (1990, p.15) も参照)

- Yahoo! 知恵袋 <[http://detail.chiebukuro.yahoo.co.jp/qa/question\\_detail/q1214126522](http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q1214126522)> の回答
- 青木繁伸「中央値 (M.e)」 <<http://aoki2.si.gunma-u.ac.jp/lecture/Univariate/median.html>> 「One more step!」以降

## 2 度数分布表の読みかた (復習)

- 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)

## 4 標本抽出

標本抽出の4段階モデル

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

無作為抽出とは：

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 (random sampling) 等確率標本

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いけない。

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)

## 5 宿題

調査データを使っている論文や新聞記事をひとつとりあげ、上記の「標本抽出の4段階」にそって紹介する。ISTUのこの授業の掲示板に、下記の内容を投稿すること。期限は5/14(火) 12:00

- (1) その論文などの書誌情報と、理論母集団・調査母集団・計画標本・有効標本について簡単にまとめたもの
- (2) その論文などがオンラインで読める場合は、URLを上記の「書誌情報」に含めておくこと。そうでない場合は、コピーを田中のレターケースまで提出

## 6 文献

- 森敏昭・吉田寿夫(編)(1990)『心理学のためのデータ解析テクニカルブック』北大路書房.

次回は5/15です(5/1, 5/8は休講)