

第3講 統計分析の基礎

田中重人 (東北大学文学部准教授)

[テーマ] 度数の利用と統計分析の基礎

1 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)

2 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)

3 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）

4 課題

（前回資料と同じ）

配布したデータを使い、年齢についての度数分布表を出力する。ただし、適当な年齢幅に区切ること。結果の表を、年齢幅の設定などがわかるよう整形して、どの年齢層が多いかなどのコメントをつけて提出。また、課題の途中でどこでつまずいたかなどの経過について書いてもよい。木曜日正午までに Google Classroom に提出。

ほかの人と自由に相談してよい。

教科書のほか、前回資料を参考にすること。