

2003.1.7 現代日本論演習 II (田中重人)

第9回 「測定値と誤差」

1. 記述統計と推測統計
2. 「真の値」と測定値
3. 誤差の種類と対策
4. 標本抽出のプロセス

【記述統計と推測統計】

記述統計＝データ（**ケース**）の特徴を
数値や図表にまとめる

推測統計＝確率的な**誤差**を考慮して、
母集団の特徴を推測する

(教科書 pp. 3–5)

【「真の値」と測定値】

$$\text{測定値} = \text{真の値} + \text{誤差}$$

記述

推測

(教科書 pp. 17–20)

【誤差 (error) の種類】

● 測定上の誤差

計器の故障・測定精度の問題

回答者の間違い・虚偽の回答

調査員の間違い・不正

調査票の不備

入力ミス

● 対象者の選択に起因する誤差

【誤差への対策】

誤差の発生メカニズムを想定して対処する

- ★ 特定の方向へのかたより (bias)
 - できるだけ起こらないようにするか、かたよりの方向を想定して補正
- ★ 方向性を持たない (狭義の error)
 - できるだけ小さくする。
誤差の範囲を考慮してデータ解釈

【統計学があつかえる誤差】

- 発生メカニズムが既知
- 誤差の範囲が確率的に決まる

無作為標本抽出にともなう

「**標本誤差**」がその典型である

【標本抽出の4段階モデル】

ユニバース (universe)*

母集団 (population)

計画標本 (designed sample)

有効標本 (valid sample / case)

*: 一般的な用語ではないので注意

★ 伝統的な統計学では
4 段階にわけずに
2 段階で考えるのがふつう：
母集団=Universe + population
標本 = (designed/valid) sample

【無作為抽出】

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**」

つぎの条件が必要：

★ 母集団の人口が既知

★ 個体を網羅した「台帳」

※ 個体によって抽出確率が違う場合も、事後的に調整して
等確率標本と同様の統計処理をおこなうことは可能

※ 「台帳」が完備してない状況でも、工夫次第で
無作為抽出に近づけることができる

【無作為抽出の実際】

★ 2 段抽出 = 2 段階の抽出単位を設定

例：市町村→住民、学校→生徒

- 確率比例抽出法：その抽出単位が含む個体数に抽出確率を比例させる。
- 等確率抽出法：上位抽出単位の抽出確率は一定にしておき、個体の抽出数のほうを調整。

★ 系統抽出 = 「台帳」から等間隔に抽出。

- スタート番号は乱数で決める
 - 抽出間隔は次のことを考えてきめる
 - (1) 台帳のもつ周期性と同調しない
 - (2) 台帳全体をカバーできる
- 具体的には 台帳人数 / 計画標本数
に近い素数をえらぶのがよい。

- ★ 層化抽出法＝母集団を層別に分け、各層の人数に比例して標本数を割り当てる
- 結果に影響を与えそうな重要な属性についておこなう：性別・年齢・地域など
 - 抽出単位や個体がどの層に属しているかを台帳から判断できないと使えない

※「層別抽出法」「比例割当抽出法」ともいう

実際の調査で理想的な標本抽出ができることはまずない。
また計画標本のなかから無効回答がでるので、
無作為ではない誤差がかならず発生する。
この誤差は統計的には処理できないので、個別に推測する

- ・ どの層を過剰に代表しているかを把握する
- ・ おなじ母集団を対象にした調査と比較する

【宿題】

論文や新聞・雑誌記事で使われている調査データについて、

- ・ その記事等の標本抽出がわかる部分のコピー
- ・ 標本抽出の4段階にそった解説

を提出（次回授業時）

次回までに、教科書の

pp. 56–59, 147–150, 259–260

を読んできておいてください