

2001.10.17 比較現代日本論研究演習 II
第2回「データ配布・SPSS入門」（田中 重人）

1. データの配布
2. 標本抽出について
3. SPSS 入門
4. データの変換

【データの配布】

1995年SSM調査B票の一部

★ 全国から70歳以下の有権者を
層化2段無作為抽出

★ 訪問面接法

cf. 『日本の階層システム』（全6巻）

東京大学出版会、2000年。

- ★ 意識項目と基本的属性に限定
- ★ 250 ケースをランダムに抽出
- ★ 未公開のデータなので
流出しないように
- ★ 変数ラベルは菅野剛
(大阪大学) 氏による

【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)

★ 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あるので、
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、
標本抽出の4段階にそって紹介する。

【データ・セット】

- ★ ケース × 変数
- ★ 変数は変数名で管理
- ★ 変数名以外に「ラベル」
- ★ 無回答などの欠損値 (.)

【SPSS のウィンドウ構成】

- データ・エディタ
- シンタックス・エディタ
- 出力ビューア

【メニューとシンタックス】

- ★ 分析手法をえらぶ
- ★ 必要なオプションを指定
- ★ 「貼り付け」をクリック
- ★ シンタックスの必要部分を選択して実行（▶）

【変数値の再割り当て】

データエディタのメニューバーで

- 「変換」 → 「値の再割り当て」
→ 「他の変数へ」
- 変換先変数の名前をつける

- 「今までの値と新しい値」
- 値の組を指定したら「続行」
- シンタックスを貼付けて実行
- 新変数の度数分布を確認
- 問題がなければデータセットを保存する

【出力ビューア】

- ★ 左側に目次、右側に出力内容
- ★ エラー表示もここに出る

【印刷】

- ★ 左側の目次で選択
- ★ 印刷前にプレビューで確認

【実習】

本人年収 (Q44_1) を 5~7 程度の
適当な間隔に区切って度数分布表
を出力し、印刷して提出