

比較現代日本論研究演習 I

大学院生対象: 2010 年度前期
<木2> コンピュータ実習室 (文学部本館 7F 711-2)

『講義概要』 記載内容

- ◆ 講義題目: 統計分析入門
- ◆ 到達目標: (1) 統計分析の基礎を理解する; (2) 実際にデータ分析をできるようになる
- ◆ 授業内容: 意識調査・テスト・実験などのデータはどのように分析すればいいでしょうか。この授業では、データの特徴を要約する記述統計の手法を中心に、統計分析の基礎を学びます。統計解析パッケージを使ってデータ分析の実習を毎回おこないます。
- ◇ テキスト: 吉田寿夫、1998『本当にわかりやすいすぐ大切なことが書いてあるごく初歩の統計の本』北大路書房。
- ◇ 成績評価の方法: 各回の授業中の課題 (50%)、中間試験 (20%)、期末レポート (30%) を合計して評価する。
- ◇ その他: 実習室で使用できるコンピュータ台数が限られているため、受講人数を制限することがある。

授業の概要

目次

1. イントロダクション (4/15)
2. 統計分析の基礎 (4/22)
3. SPSS 入門・データ配布 (5/6)
4. 度数分布とクロス表 (5/13~6/3)
5. 中間試験 (6/10)
6. 平均値の比較 (6/17~7/15)
7. 予備日 (7/23)
8. 期末レポート (8 月上旬提出)

※ () 内の日付は、学期前のおおよその計画をあらわしているが、実際の授業の進行状況によって前後にずれることがある。

修士論文等で質問紙調査を予定している者は、2 学期開講の

- 比較現代日本論研究演習 II 「質問紙法調査の理論と実践」(木1) および
- 比較現代日本論研究演習 III 「実践的統計分析法」(木2: 学部と合同)

も受講することがのぞましい。

1. イントロダクション

- 授業の概要・スケジュール・評価方法
- 部屋とコンピュータの使いかた
- 模擬データ入力実習

2. 統計分析の基礎

- 記述統計と推測統計
- SPSS の起動
- データ行列 (データセット) とは
- その他のソフトウェアについて

3. SPSS 入門・データ配布

- データの配布と説明
- データの種類
- SPSS コマンド・シンタックス
- メニューによるシンタックス作成
- 変数値の再割り当て
- 標本調査とは

4. 記述統計 (1): 度数分布とクロス表

4.1. 度数分布表

- frequencies コマンド
- 相対度数 (パーセンテージ)
- 棒グラフ
- ヒストグラム・度数ポリゴン
- Excel で整形, グラフ作成

4.2. クロス表

- 度数分布表のグループ化
- クロス表表記
- 行と列の%
- 周辺度数 (marginal distribution)
- crosstabs コマンドとそのオプション

4.3. 無関連状態と期待度数

- Φ 係数
- 期待度数・残差・連関係数
- クロス表とグラフの書きかた

5. 中間試験

6. 記述統計 (2): 平均値の比較

6.1. 平均と分散

- データの種類: 復習
- 平均値
- 分散と標準偏差
- 分布と外れ値
- ノンパラメトリックな代表値 (中央値と四分位偏差)

6.2. 平均値の層別比較

- 平均の差と差の平均
- 層別平均
- エフェクト・サイズ
- 相関比から分散分析へ
- 公表に際してなにを書くべきか

カードをとって
適当なところに着席

電源はまだ入れない

0

比較現代日本論研究演習 I

統計分析入門

東北大学大学院文学研究科 2010 年度
田中 重人 (准教授)

1

【目的】

統計分析の基礎的な手法の習得

- SPSS の操作
- クロス表分析
- 平均値の比較

2

【教科書】

吉田 寿夫 (1998)
『本当にわかりやすいすぐく大切なことが
書いてあるごく初歩の統計の本』
北大路書房。

3

【成績評価】

- ・ 授業中の課題 (50%)
- ・ 中間試験 (20%)
- ・ 期末レポート (30%)

4

【関連する授業】

- 2 学期
- ・ 比較現代日本論研究演習 II
「質問紙調査の理論と実践」(木 1)
 - ・ 比較現代日本論研究演習 III
「実践的統計分析法」(木 2)

5

質的研究法の授業？

6

受講登録フォーム記入

7

【コンピュータ実習室について】

- ★ 入室に**学生証**が必要
→ 研究生などは、オンラインで登録
(他学部の場合は文書で申請)
- ★ 土足・飲食・喫煙 **厳禁**
- ★ 退出時は必要事項を紙に書く
(書けるところを書いてみよう)
- ★ ドアの開けかた

8

【コンピュータの起動と終了】

- ・ 本体とディスプレイの電源を ON
- ・ 表示されるお知らせの内容をよく読む
- ・ 「NumLock」ランプ点灯を確認
- ・ 終了するときは、ディスプレイの電源を切
ることをわすれないように

9

【ファイルの保存場所】

授業でつかうファイルは、
授業開始時に マイドキュメント
フォルダにコピーして使う。
授業終了時に削除してかえること。

★ 内蔵 Disk にデータは置けない

10

必要なデータは各自で
USB スティックメモリなど
にコピーして持ち帰る

→ 各自で購入しておくこと。

11

【SPSS】

データ解析用ソフトウェア

- ★ Windows での開発に
特に力を入れている
- ★ 購入しやすい

12

【データ入力実習】

3 人 × 5 変数 (a~e)

- ・ どのように入力するか
- ・ 変数名
- ・ データ保存

13

【この授業で使用するデータ】

1995 年 SSM 調査 B 票の一部

cf. 『日本の階層システム』(全 6 巻)
東京大学出版会、2000 年。

SSM 調査については <http://www.sal.tohoku.ac.jp/~tsigetof/ssm/> 参照

14

2010.4.15

比較現代日本論研究演習 I (田中重人)

受講登録フォーム

氏名：

学年：

学生番号：

所属 (文学研究科日本語教育学専攻以外の場合)：

研究内容：

- ・ 自宅でパソコンを使えますか? **ある / ない**
- ・ SPSS を使った経験がありますか? **ある / ない**
- ・ コンピュータ・プログラムを作成したり、プログラミングの授業を受けたりしたことがありますか? **ある / ない**
 ある場合 → 言語名 ()

数学的予備知識の調査 (成績評価には関係ありません)

(1) 1次方程式 $y = 0.5x + 1.2$ をグラフに書いたとき、傾き (gradient) と切片 (intercept) はそれぞれいくつか。

傾き = _____ ; 切片 = _____

(2) 「必要十分条件」 (necessary and sufficient condition) とは何か。簡単に説明せよ。

(2) つぎの数式の値を求めよ。

$$\log_2 16 =$$

(4) つぎの数式の値を求めよ。計算のプロセスがわかるように解答すること

$$\sum_{k=1}^{10} k =$$

受講者の興味と数学的知識の調査

→別紙

コンピュータ実習室について

入室・退室

学生証が必要 (ない人は、教務係で臨時カードを借りること)。

土足・飲食・喫煙厳禁。

退出時には必要事項を紙に記入。

コンピュータの起動と終了

使いはじめるときは……

- コンピュータ本体の電源を入れる
- ディスプレイの電源を入れる (2-3 秒押しつづけないと入らないので注意)
- 表示されるお知らせをひととおりよむこと
- キーボード右上の「NumLock」ランプがついているか確認

使い終わるときは……

- 「マイドキュメント」などに保存してある自分のファイルを削除
- 画面左下の「スタートメニュー」から「終了オプション」→「電源を切る」を選択
- コンピュータ本体の電源が切れたことを確認
- ディスプレイの電源を切る
- フロッピーディスク、USB スティック・メモリなどをわすれないこと

ファイルの保存場所について

教室のコンピュータの内蔵ディスクには、個人のファイルを置いてはならない。授業中に必要なファイルは「マイドキュメント」フォルダに一時的に保存してよいが、授業が終わったら自分のフロッピーかスティック・メモリ等にコピーして、内蔵ディスクのほうのファイルは削除すること。

コンピュータ実習室で使えるリムーバブルメディアはつぎのふたつ。各自どちらかを購入しておくこと。

- フロッピーディスク (3.5 インチ) …… 「Windows フォーマット」のものが便利。安いがよく故障する。容量が小さい。
- フラッシュメモリ …… 「USB2.0 対応」のもの。値段は高いが容量が大きい。とりはずすときは画面右下の「ハードウェアの安全な取り外し」アイコンをクリックして、「USB 大容量記憶装置」を停止させてから、メモリ本体を引き抜く。

模擬データ入力実習

SPSS について

参考書: 宮脇典彦・和田悟・阪井和男 (2000) 『SPSS によるデータ解析の基礎』培風館。

SPSS の起動

スタートメニューから「プログラム」→「SPSS for Windows」→「SPSS for Windows 12.0J」で起動する。(※ここで何かエラーメッセージが出るかもしれないが、気にせず「続行」または「OK」する。)

「どのような作業を行いますか?」ときかかれたら「データを入力」をチェックして「OK」。

データ入力

配布した架空の回答票をもとに、データを入力してみよう。

まず変数を定義

- 「データエディタ」ウインドウのいちばん下の「変数ビュー」タブに切り替える
- 変数名を必要だけつくる。今回は a, b, ..., e とでもしておこう。変数名は自分がわかればどんなものでもよい。日本語も使える。なお、変数名以外のフィールドは入力しなくてよい
- 書き終わったら「データビュー」タブに切り替えて、いちばん上の行に変数名がならんでいることを確認する。

つづいてデータを入力していく。今回は3人分のデータを用意してあって、変数は5個なので、3×5の行列型のデータができるはずである。

適当な名前で「マイドキュメント」内に保存してみる。(ほかのフォルダに保存してはならない。)

「マイドキュメント」を開いて、SPSS データファイル (なんとか.sav) ができていることをたしかめる。

このデータファイルは授業終了時に削除すること。(次回以降の授業ではつかわないので、コピーしておく必要はない。)

※ この方式は SPSS でデータを入力するときのいちばん簡便な方法であるが、大きなデータはあつかにくいので、テキストファイルでデータを用意しておくのがふつうである。

第2回「統計分析の基礎」目次

- 1. データ収集から分析まで
- 2. データ配布
- 3. 標本抽出
- 4. SPSS の基礎知識
- 5. 尺度水準

1

【データ収集から分析まで】

- データの収集 (実験／観察)
- 分析可能な形に加工
 - ・ 分析の単位
 - ・ 変数の同定
 - ・ 変数値の付与 (coding)
- データ・セット作成
 - ・ クリーニング

2

- データの特徴を少数の数値に要約
= **記述統計**
- 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

【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. (2000)『日本の階層システム』(全6巻) 東京大学出版会。

4

- ★ 意識項目と基本的属性に限定
(調査票の×印はデータセットにない項目)
- ★ 250 ケースをランダムに抽出
- ★ 流出しないように
- ★ 変数ラベルは菅野剛
(日本大学) 氏による

5

- ★ 毎回の授業で使うので、忘れないこと (調査票も)
- ★ 期末レポート提出時に返却

6

【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)

→ 「**確率標本**」 (probability sample)

7

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

8

統計的な推測は、**確率標本を前提とする**

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

9

【層化 2 段無作為抽出】

- ・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)

10

【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 (.)

11

【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

12

【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 (▶)

13

【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
- ★ 出力先の切り替え
- ★ 印刷前にプレビュー
- ★ 電源の入れかた
- ★ ジョブの確認・取り消し
- ★ タイル印刷 (2 面, 4 面, ...)

14

【その他のアプリケーション】

- 文書作成 (Word)
- 表計算 (Excel)
- 電卓 (アクセサリ)

SPSS の出力ビューアから表を Excel や Word に貼り付けられる

15

【尺度水準】

- 比率尺度 (ratio scale)
- 間隔尺度 (interval —)
- 順序尺度 (ordinal —)
- 名義尺度 (nominal —)
(質的変数とも)

(教科書 p. 8)

16

- ★ 上位の尺度のほうがあつかえる演算が豊富
- ★ 上位の尺度は下位の尺度の特徴を兼ね備えている
→分析手法の選択幅がひろい

17

私たちが測定するものはたいてい
順序尺度以下である

上位の尺度への変換には
一定の理論的根拠が必要

18