

# 現代日本論演習

3年生対象: 2004年度前期(5 Semester: 授業コード=L52505)  
<火5>コンピュータ実習室(文学部本館 7F 711-2)

## 『講義概要』 p. 166 記載内容

◆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、小規模の標本調査を念頭において、統計分析の基礎的な手法を学びます。これまで統計的な分析をおこなったことのない人を対象に、初歩から講義します。同時に、コンピュータを実際に使って、毎回データ分析の実習をおこないます。

◇テキスト: 吉田寿夫(1998)『本当にわかりやすいすご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。

◇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。

## 授業の概要(予定)

### 目次

1. イントロダクション (4/13)
2. SPSS 入門 (4/20)
3. 統計分析の基礎 (4/27)
4. 記述統計(1): 度数分布とクロス表 (5/11~5/18)
5. 中間試験 (5/25)
6. 記述統計(2): 平均値の比較 (6/1~6/15)
7. 推測統計 (6/29~7/20)
8. 期末レポート

※ () 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。

## 1. イントロダクション

- この授業の概要・スケジュール・評価方法
- 部屋とコンピュータの使いかた
- SPSS の起動
- データ行列(データセット)
- 模擬データ入力実習

## 2. データ配布・SPSS 入門

- データの配布
- SPSS の概要
- 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
- 他のソフトウェアについて(電卓, Excel, Word?)
- 印刷

## 3. 統計分析の基礎

- 実験と観察
- データの記述
- データの種類

## 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

- $\Phi$ 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
- 順序尺度と間隔尺度の変換
- 平均値
- 分散と標準偏差
- 分布と外れ値

### 6.2. 平均値の層別比較

- 層別平均
- エフェクト・サイズ
- 相関比から分散分析へ
- 表とグラフの書きかた

## 7. 推測統計

### 7.1. 誤差の評価

- データの記述と誤差の評価
- Case, Sample, Population, Universe
- 無作為抽出
- 非標本誤差
- 標本誤差の統計的推測

### 7.2. 平均値の推定

- 平均値の点推定
- 区間推定と  $t$  分布
- 平均値の差の区間推定
- エフェクトサイズ・相関比と区間推定

### 7.3. 統計的検定

- 区間推定の簡易表記としての有意水準
- 平均値の差の  $t$  検定
- 連関係数の  $\chi^2$  検定
- 分散分析と  $F$  検定
- 検定結果の表記

## 8. 期末レポート

## この授業の概要・スケジュール・評価方法

→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。

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)
- 「開始するユーザ名をクリックしてください」の画面にきたら「Guest」を選択
- 表示されるお知らせをひととおりによむこと
- 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 フラッシュ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フラッシュ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。

## 受講者の興味と数学的知識の調査

→別紙

## 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 10.0J」→「SPSS for Windows 10.0J」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」すること。

「どのような作業を行いますか?」ときかれたら「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
- 変数名を必要なだけつくる。今回は q35a, q35b, ..., q35f 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。(ほかのフォルダには保存できません。)

「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単な方法であるが、大きなデータはあつかいにくいので、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。

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

2004.4.13

現代日本論演習  
統計分析の基礎

東北大学文学部 2004 年度  
田中 重人 (講師)

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
- 推測統計の手法

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

受講登録フォーム記入

5

【コンピュータ実習室について】

- ★ 入室に**学生証**が必要
- ★ 入り口段差に注意
- ★ 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)
- ★ ドアが開かなくなったときは電話で連絡

6

【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
- ・ ユーザ名は「Guest」(パスワード不要)
- ・ 終了するときには、ディスプレイの電源を切ることをわすれないように

7

【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

8

必要なデータは各自で  
フロッピーかフラッシュメモリ  
にコピーして持ち帰る

→ 各自で購入しておくこと。

9

【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

10

【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。

11

模擬データ入力実習

12

2004.4.13

## 現代日本論演習 (田中重人) 受講登録フォーム

氏名 :

学年 :

学籍番号 :

所属 (文学部日本語教育以外の場合) :

興味のあること (非学術的な話題も可) :

・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、傾きと切片はそれぞれいくつか。

傾き = \_\_\_\_\_ ; 切片 = \_\_\_\_\_

(2) 「偏差値」はどういう目的のために使われるか。またどうやって求めるか。

簡単に説明せよ

(3) 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

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

## 数学的予備知識の調査：解答のポイント

(1) 1次方程式  $y = 0.5x + 1.2$  をグラフに書いたとき…

↓            ↓  
傾き        切片

(2) 「偏差値」は

平均と分散が違う複数の得点分布のなかでの相対的位置を示す

$$50 + 10 \frac{\text{生の得点} - \text{平均}}{\text{標準偏差}}$$

(3)  $16 = 2^x$  となる  $x$  をさがせばよい： $x = 4$

(4) つぎの数式の値：

$$\sum_{k=1}^{10} k = 1+2+3+4+5+6+7+8+9+10 =$$

1. データの配布
2. SPSS のウインドウ構成
3. メニューとシンタックス
4. 変数値の再割り当て
5. 出力の読みかた・印刷

1

## 【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。

2

- ★ 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

3

## 【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 (.)

4

## 【SPSS のウインドウ構成】

- データ・エディタ
- シンタックス・エディタ
- 出力ビューア

5

## 【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 (▶)

6

## 【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

7

- 「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

8

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

9

1. データ収集から分析まで
2. 変数の種類
3. 度数分布表とヒストグラム

1

### 【データ収集の方法】

- 実験とは
- 観察とは

※「調査」は、通常、「観察」にふくめる  
※ この授業では、実験の方法はあつかわない

2

### 【データ分析のふたつの水準】

- データの特徴を少数の数値に要約して記述 = **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

- データ収集の際は、誤差をできるだけ少なくする方法を考へること
- 分析の際は、まず手元のデータの特徴をしっかりとつかむ  
→ 記述統計 (教科書 第5章まで)
- そのうえで、データがふくむ「誤差」を推測し、「真の値」について考へる  
→ 推測統計 (教科書 第6章以降) +  $\alpha$

4

### 【変数の種類】

- 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)

5

### 【尺度の変換】

- ★ 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

6

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

7

### 【度数分布表】

Frequencies コマンドを使う

- ★ 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)

8

### 【累積%とパーセンタイル】

- 順序尺度以上の場合のみ意味を持つ
- Percentile (= %点)
- 中央値 (median) = 50%点
- 「割り切れてしまう」場合は中点をとる  
(教科書 p. 43)
- 同じ値が並ぶ場合は多少の操作が必要  
(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, p. 15)

9

### 【棒グラフとヒストグラム】

- 棒グラフ……棒同士に空白をあける。**高さ(長さ)をよむ。**
- histogram (柱グラフ)……柱の間隔をあけない。**面積をよむ。**

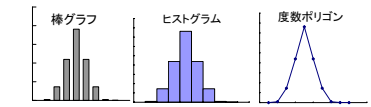
※縦軸は度数または%

10

- ★ 連続量を階級分けした場合 → ヒストグラム
  - ★ それ以外の場合 (離散量 / 名義尺度) → 棒グラフ
- ※度数多角形 (polygon) は複数の変数の分布を比較するときに便利。

(教科書 p. 32-36)

11



SPSS では histogram が書きにくい。  
★ recode で整形した上で度数分布表のメニューで「図表...」指定。棒グラフを書く  
★ グラフ → インタラクティブ → ヒストグラム  
では等間隔の区間に分割してくれる

12

### 【実習】

- (1) 本人年齢の度数分布表を出力し、中央値と上側 20% 点に印をつけよ
- (2) 適当な変数について棒グラフまたはヒストグラムを作成

13

### 【キーワード】

行 (row) 列 (column) セル (cell)  
周辺度数 (marginal frequency)  
行% (row percent) 列% (column percent)

1

### 【度数分布表の比較】

- データエディタのメニューで  
「データ」→「ファイルの分割」  
→「グループの比較」
- 度数分布表を出力

2

- 「データ」→「ファイルの分割」  
→「すべてのケースを分析」  
でもとにもどしておく

3

### 【クロス表の基本型】

質的変数 (名義尺度) 同士の関連  
についての基本的な分析法

4

	$\beta$			
$\alpha$	1	2	3	合計
行 1	a	b	c	a+b+c
2	d	e	f	d+e+f
3	g	h	i	g+h+i
合計	a+d+g	b+e+h	c+f+i	N

列 周辺度数

5

### 【Crosstabs コマンド】

性別×「性別による不公平」  
のクロス表を書いてみよう

「分析」→「記述統計」→「クロス集計表」

6

### 【行%と列%】

「クロス集計表」メニューで「セル」にパー  
センテージ (行・列) を追加

- ★ 行%, 列%のつかいわけは  
説明→被説明の関係に対応  
行→列の説明をすることが多い
- ★ 周辺度数の%とも比較する

7

### 【グラフを書いてみる】

- ★ クロス表は帯 (積み上げ棒)  
グラフで表現することが多い  
SPSS ではうまくかけない。コピーして  
Excel に貼付けてグラフを書くのがよい
- ★ 度数にも注意

8

### 【課題】

性別×適当な変数でクロス表作成、  
グラフも書いて印刷して提出

9