URL: http://www.nik.sal.tohoku.ac.jp/~tsigeto/statg/
作成：田中重人（講師）〈tsigeto@nik.sal.tohoku.ac.jp〉

比較現代日本論研究演習

大学院生対象：2004 年度前期
＜木２＞コンピュータ実習室 (文学部本館 7F 711-2) 授業コード=LM14205

『講義概要』p. 398 記載内容

◆講義題目：統計分析入門
◆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。
◇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、受講人数の制限をおこなうことがある。
◇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

※（）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. SPSS 入門・データ配布

- 模擬データ入力実習
- データの配布
- 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

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
  - 名義・順序・間隔・比例
  - 順序尺度と間隔尺度の変換
  - 正規分布とは
- 標本抽出の 4 段階モデル
- サンプリングの概念と手順

- 新聞・雑誌・論文などにみられる調査の母集団・標本などについて各自報告

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

URL: http://www.nik.sal.tohoku.ac.jp/~tsigeto/statg/g040415.html
作成：田中重人（講師）〈tsigeto@nik.sal.tohoku.ac.jp〉

[比較現代日本論研究演習 統計分析入門]

第 1 回 (2004-04-15)

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)
- 「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

使い終えるときは……

- 「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「ハードウエ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 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。)

「どのような作業を行いますか?」ときかれたら「データに入力」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 a, b, ..., f 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 6 個なので、 3×6 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便な方法であるが、 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。

比較現代日本論研究演習
「統計分析入門」

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

2004.4.15

比較現代日本論研究演習

統計分析入門

東北大学文学部 2004 年度
田中 亘人（講師）

【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

【教科書】

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】

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

受講登録フォーム記入

【コンピュータ実習室について】

★ 入室に学生証が必要 (ない人は教務掛で)
★ 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

【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)
- 「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

【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

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。

【SPSS】

データ解析用ソフトウエア

★ Windows での開発に
特に力を入れている
★ 購入しやすい

【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。

模擬データ入力実習

2004.4.15

比較現代日本論研究演習 (田中重人) 受講登録フォーム

氏名：

学年：

学籍番号：

所属 (文学研究科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：

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

2004.4.22 比較現代日本論研究演習 (田中重人)

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. データの配布
3. 標本抽出

【データ収集から分析まで】

●データの収集 (実験／観察)

●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)

【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

(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

★ データが流出しないように

★ 変数ラベルは菅野剛 (日本大学) 氏による

★ 毎回の授業で使うので、忘れないこと

★ 期末レポート提出時に返却

【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

統計的な推測は、等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、標本抽出の 4 段階にそって紹介する。

人数分コピーを用意してきて、次回授業時に報告。

2004.5.13 比較現代日本論研究演習（田中重人）

第 4 回「SPSS 入門」目次

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

【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

【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（▶）

【度数分布表】

「分析」→
「記述統計」→「度数分布表」

→ 変数を選ぶ

【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
★ 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
★ タイル印刷 (2 面, 4 面, ...)

2004.5.13 比較現代日本論研究演習（田中重人）

第 4 回「変数の種類と度数分布」目次

1. 変数の種類
2. 尺度の変換
3. 度数分布表
4. 棒グラフとヒストグラム

【変数の種類】

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

(教科書 p. 8)

【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

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

【実習】

SSM 調査の調査票中で、比率尺度とみなせるものはどれか

【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)

【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)

【棒グラフとヒストグラム】

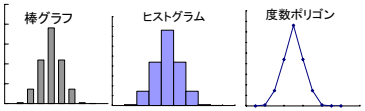
●棒グラフ……棒同士の間に空白をあける。**高さ（長さ）をよむ。**
●histogram (柱グラフ)……柱の間隔をあけない。**面積をよむ。**

※縦軸は度数または%

★ 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)



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

【実習】

(1) 本人年齢の度数分布表を出力し、中央値と 80%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

2004.6.3 比較現代日本論研究演習　(田中重人)

第 6 回「クロス表分析の基礎」

【キーワード】

行 (row)　　列 (column)　　セル (cell)

周辺度数 (marginal frequency)

行% (row percent)　列% (column percent)

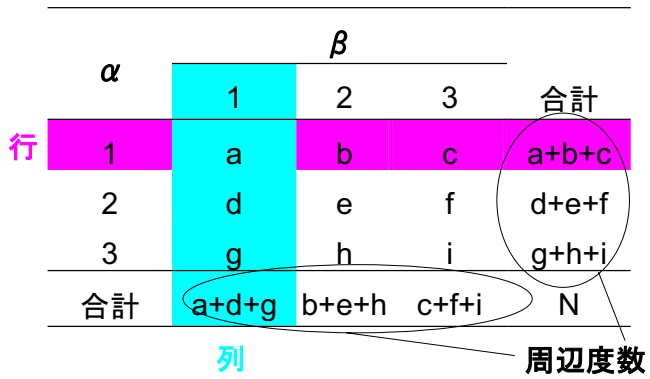
【度数分布表の比較】

- 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

- 「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

【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



【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」

【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**

【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**

【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

2004.6.10 比較現代日本論研究演習 （田中重人）

第 7 回「ϕ 係数」

1. 自由度 (degree of freedom)
2. クロス表分析のふたつの系列
3. 2×2 クロス表の性質
4. ϕ 係数 (phi coefficient)

【自由度】

2×2 クロス表では、周辺度数が所与なら、
1 つのセル度数が決まればほかも決まる

α	β 1	β 2	合計
1	a	$g-a$	g
2	$i-a$	$h-i+a$	h
合計	i	j	N

3×3 クロス表：セル度数が 4 つ決まれば…

α	β 1	β 2	β 3	合計
1				f
2				g
3				h
合計	I	j	m	N

$k \times l$ クロス表の自由度 (degree of freedom)

$$\text{d.f.} = (k-1)(l-1)$$

【クロス表分析の 2 つの系列】

- 「%の差」系 (期待度数との差)
  ＝連関係数
- オッズ比系 (乗法モデル)
  ＝対数線形分析、ロジット分析

この授業で取り上げるのは前者だけ

【2×2 クロス表の性質】

以下、つぎの記号法を使う

α	β 1	β 2	合計
1	a	c	g
2	b	d	h
合計	i	j	N

(1) 行%は 1 列について比較すればよい：

$$\frac{a}{g}-\frac{b}{h}=\frac{d}{h}-\frac{c}{g}$$

(2) 行%の差がゼロなら列%の差もゼロ

(3) $g=i$ なら行%の差と列%の差は同じ：

$$\frac{a}{g}-\frac{b}{h}=\frac{a}{i}-\frac{c}{j}$$

(例 1) 行%の差＝８％

60%	40%	100%
52%	48%	100%

(例 2) 行・列とも%に差なし

52	48	100
52.0%	48.0%	100.0%
66.7%	66.7%	
26	24	50
52.0%	48.0%	100.0%
33.3%	33.3%	
78	72	150
52.0%	48.0%	100.0%

(例 3) 行・列とも 10%の差

70	30	100
70.0%	30.0%	100.0%
70.0%	60.0%	
30	20	50
60.0%	40.0%	100.0%
30.0%	40.0%	
100	50	150
52.0%	48.0%	100.0%

【ϕ 係数】

2×2 クロス表の「連関」の尺度

$$\phi = \frac{ad-bc}{\sqrt{ghij}}$$

この係数の意味は?
(分子だけ取り出して考えてみよう)

【SPSS での ϕ 係数の計算】

「クロス集計表」の
「統計」で
「ファイとクラマーの V」をチェック

2004.6.17 比較現代日本論研究演習 （田中重人）

第 8 回「連関係数」

【キーワード】

連関 (association), **独立** (independence),

期待度数 (expected frequency),

クラメール の連関係数 (Cramer's *V*)

【φ係数の性質】

1. φ＝交差積の差 / √(周辺度数の積)
2. φ＝ 相関係数の特殊ケース
3. |φ| ＝ 行%差と列%差の中間の値
4. **ϕ^2 ＝ 標準残差の総計 / *N***
   (→ 2×2 以上のクロス表に拡張できる)

【期待度数とφ係数】

※記号法は前回と同じ

独立 (無関連)：a/b = c/d

期待度数 (expected frequency)
周辺度数を固定しておいて独立なクロス表を作ったとき、各セルに入る度数：

gi/N	gj/N
hi/N	hj/N

独立なクロス表の例

52	48	100
52.0%	48.0%	100.0%
66.7%	66.7%	
26	24	50
52.0%	48.0%	100.0%
33.3%	33.3%	
78	72	150
52.0%	48.0%	100.0%

★ 期待度数はたいてい小数になる
★ 期待度数について行%と列%を計算すると、周辺度数の%とおなじになる

観測度数　　各セルに入る実際の度数
残差 (residual) 観測度数と期待度数の差
標準残差 (standardized ---) 残差/√期待度数

ex. $A = \dfrac{a - gi/N}{\sqrt{gi/N}}$

χ^2 (chi-square)　　標準残差の平方和

各セルに入る標準残差を *A, B, C, D* とする

$$\chi^2 = A^2 + B^2 + C^2 + D^2 = N\left(\frac{a^2}{gi} + \frac{b^2}{hi} + \frac{c^2}{gj} + \frac{d^2}{hj} - 1\right)$$

χ^2 を人数で割った値が φの２乗 に等しい

$$\phi^2 = \frac{\chi^2}{N} \quad \text{すなわち} \quad |\phi| = \sqrt{\frac{\chi^2}{N}}$$

【クラメールの連関係数 *V*】

***k*×*l* 表へのφ係数の拡張**　　(教科書 p. 114–117)

★ k と l のうち小さいほうを m とする
★ 2×2 表と同様に期待度数・残差を求める
★ χ^2 を求める
★ χ^2 を N と$(m-1)$ で割って平方根をとる

$$V = \sqrt{\frac{\chi^2}{N(m-1)}}$$

【*V* の性質】

★ **行･列変数が独立のとき $V = 0$**
★ **関連が強くなると大きくなる**
★ **最大値は 1**

【SPSS で実習】

クロス表のオプションを指定：

●「セル」… 度数(観測／期待)
　　残差(標準化なし／標準化)
●「統計」…カイ２乗
　　ファイと Cramer の V