

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

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

- 「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**

【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

【実習】

本人収入 (Q44_1) を

- (1) 万円単位の金額に変換
- (2) 5つのカテゴリに区切る

(2) については、わけかたを説明すること。

Word に出力とシンタックスを貼り付けて印刷して提出

1. 度数分布表
2. 累積%とパーセンタイル
3. グラフの利用

1

【度数分布表】

Frequencies コマンド

「分析」

→ 「記述統計」

→ 「度数分布表」

2

出力 :

- ★ 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)

3

【累積%とパーセンタイル】

- 順序に意味がある場合のみ有効
- 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)

4

【実習】

家族収入 (q44_3) について、度数分布表を出力し、中央値、25%点、75%点を求めよ

5

【グラフの利用】

- 表 (table)……正確な数値がわかるが、全体の傾向を読み取るには熟練が必要
- グラフ (graph/chart)……全体の傾向が簡単に読み取れるが、正確さは犠牲になる

初心のうち、表とグラフの両方を作成して読んでいくのがよい

6

【棒グラフとヒストグラム】

- 棒グラフ……棒同士の間空白をあける。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%

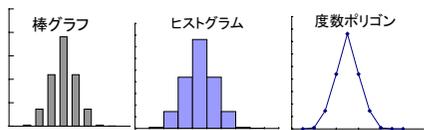
7

- ★ 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)

8



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

9

Excel を使う場合 :

- ★ recode で整形した上で度数分布表を出力
- ★ 表を Excel にコピーする
- ★ 必要なら変数値のラベルをつける (横軸に表示される)
- ★ ヒストグラムや度数多角形のときは両端に度数 0 の行をつくる
- ★ グラフを作成

10

棒グラフをヒストグラム風にするには

- ★ グラフの棒の上で右クリック
- 「データ系列の書式設定」
- 「オプション」
- 「棒の間隔」を 0 にする

※見た目がそれらしくなるだけなので、横軸のラベルや階級幅の調整はむずかしい。本当のヒストグラムを書くには、グラフ専用のソフトウェアを使う。

11

【実習】

年齢について 5 歳刻みのヒストグラムを作成する:
(21-25, 26-30, ... のようにラベルをつける)

12

【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

【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」

5

		β			
α		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

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・全体) を追加

→ これらはどのように計算されているか
どのようにつかいわけるか?

7

- ★ 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

8

【課題】

性別 × 適当な変数でクロス表作成、%からわかることをコメントする。

9

第6講「クロス表の性質」

1. 帯グラフの作成
2. 自由度 (degree of freedom)
3. クロス表分析のふたつの系列
4. 2×2 クロス表の性質

1

【帯グラフの作成】

- ・ SPSS 出力を Excel にコピー
- ・ %の部分だけ残す
- ・ 適切なラベルをつける
- ・ グラフ作成
「積み上げ棒」グラフの右端がそろったのを選ぶ
「系列」を「行」にするか「列」にするか

2

【自由度】

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

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

3

3×3 クロス表：セル度数が4つ決めれば...

α	β			合計
	1	2	3	
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)

$$d.f. = (k-1)(l-1)$$

4

【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

5

【2×2 クロス表の性質】

以下、つぎの記号法を使う

α	β		合計
	1	2	
1	a	c	g
2	b	d	h
合計	i	j	N

6

(1) 行%は1列について比較すればよい:

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

7

(2) 行%の差がゼロなら列%の差もゼロ

(3) 行%の差が100なら列%の差も100

(4) $g=i$ or $g=j$ なら行%の差と列%の差は同じ:

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

(5) これら以外の場合、行%の差と列%の差はちがう値になる

8

(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%

9

【キーワード】

φ係数 (phi coefficient)

連関 (association), 独立 (independence),

期待度数 (expected frequency),

1

【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)

2

【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

(→ VI セメスタ授業)

3. |φ| = 行%差と列%差の中間の値

(教科書 p. 103 表 4-1 について計算してみよう)

3

4. φ² = 標準残差の2乗の総計 / N

(→ 2×2 以上のクロス表に拡張できる)

4

【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:

$$\frac{gi/N}{hi/N} \quad \frac{gj/N}{hj/N}$$

5

各セルの期待度数は?

		100	
		100.0%	
		50	
		100.0%	
78	72	150	
52.0%	48.0%	100.0%	

6

★ 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. $A = \frac{a - gi/N}{\sqrt{gi/N}}$

7

観測度数が下記の場合、残差と標準残差は?

40	60	100	
		100.0%	
38	12	50	
		100.0%	
78	72	150	
52.0%	48.0%	100.0%	

【宿題】教科書の表 4-1 について
期待度数・残差・標準残差・χ² を求める

8

χ² (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

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

χ² を人数で割った値が φ の 2 乗 に等しい

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

9

1. 連関係数
2. SPSS で実習
3. 表の書きかた

1

【クラメールの連関係数 V 】

$k \times l$ 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

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

2

【 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

3

【SPSS で実習】

クロス表のオプションを指定：

「統計」で

「カイ 2 乗」「ファイと Cramer の V 」

※「セル」で「度数」(観測/期待)と

「残差」(標準化なし/標準化)を指定することもできる

4

つぎの変数についてクロス表を解釈：

- ・ 性別 (q1_1) × 性別役割意識 (q35a)
- ・ 年齢 10 歳階級 × 性別役割意識 (q35a)
- ・ 生活水準の変化 (q36) × 満足度 (q37)

V がどれくらいか

→ どこに%の差があるか?

5

【注意事項】

期待度数の小さいセルがある場合、
連関係数は適切な指標にならない

→ 期待度数 < 5 のセルがないか、
カイ 2 乗値の表の下の警告で確認

6

【表と図の約束ごと】

★ 「表 1」「図 1」のように
それぞれ通し番号をつけて参照

★ 表のタイトルは上、
図のタイトルは下

「それだけでわかる」ように

7

【他人に見せる表】

● 資料としての表…データを詳細に再現したものがよい

● プレゼンテーション用の表…
わかりやすく情報を圧縮する
→ どう圧縮するかがセンスの見せどころ

8

【予告】

来週は中間試験

- ・ 出題範囲は、「クロス表」まで
- ・ コンピュータで答案を作成、印刷して提出
- ・ 何でも持ち込み可
ただし通信と相談は禁止

9

授業資料

表の書きかた

表1 性別と性別による不公平感との関連

性別	性別による不公平			合計	(人)
	「大いにある」	「少しはある」	「ない」		
男性	36.0	50.5	13.5	100.0	(111)
女性	27.3	56.8	15.9	100.0	(132)
合計	31.3	53.9	14.8	100.0	(243)

Cramer's $V=0.094$. 無回答=7.

表2 県や市町村の部課長以上の役人に知り合いがいる比率の男女差

性別	%	(人)
男性	46.0	(113)
女性	27.6	(134)
合計	36.0	(247)

$\phi=0.191$. 無回答=3.

プレゼンテーションのための表

- ・ カテゴリーの並べ順や行列のくみあわせをわかりやすく
- ・ 変数とカテゴリーの命名
- ・ 表のタイトル

書くべき要素

- ・ 各セルの行(列)%
- ・ 行(列)合計の度数と「100.0%」
- ・ 列(行)合計の%
- ・ 全体の度数
- ・ Cramer の V (または ϕ)
- ・ 欠損数とその原因

行→列の因果を想定するのがふつうだが、列→行でもよい。方向は、合計の「100.0」で区別する。

- ・ 全度数が 1000 人以下であれば、%は小数第 1 位まで
- ・ V や ϕ などの係数は小数第 3 位まで

2 列表の場合は 1 列の%だけ示してもよい

- ・ 縦罫線はなるべく引かない
- ・ 文字列は左揃え、数字は小数点揃えが基本

タイトル、表本体、注釈を読めばそれだけでわかるように書くこと

→ タイトルと行・列頭の見出し (heading) を工夫する

第9講「平均値と標準偏差」

1. 尺度水準と分析法
2. 代表値と散布度
3. 平均値と標準偏差

1

【尺度水準と分析法】

名義×名義 → クロス表

名義×間隔 → 平均値の比較

2

【代表値と散布度】

★ 中央値 (median) — 四分位偏差 (Q)
(順序尺度以上)

★ 平均値 (mean) — 標準偏差 (SD)
(間隔尺度以上)

(教科書 p.42-51)

3

【平均値】

総和をデータ数で割ったもの

【標準偏差】

平均値からの偏差の2乗値の平均が「分散」
分散の平方根が「標準偏差」

★ 平均値と標準偏差はセットで使う

4

★ 次のデータの平均と SD は?

{1, 2, 4, 6, 7}

(教科書 p. 42, 48)

5

値	偏差	偏差 ²
1		
2		
4		
6		
7		

平方和=
分散=
SD =

6

【SPSSのコマンド】

「記述統計」 → 「度数分布表」

「統計」オプションで
「平均値」と「標準偏差」をチェック

「記述統計」 → 「記述統計」でもよい

7

【宿題】

10歳刻み年齢と適当な変数とのクロス表について
・整形した表を書く
・連関係数および%をみて解釈

【予習】

教科書 pp. 44-46 を読んでおくこと。

8

【期末レポート】

期限：8/12 (木) 17:00

提出先：日本語教育学研究室 (文法合同棟 2F)
205 室の田中のレターケース (郵送・電子メール提出も可)

内容：クロス表と平均値の比較について適当な分析をして結果を解釈する。図表は読みやすく整形し、論文としての体裁を整えること。

備考：後期の授業を受講しない者は、SSM データのディスクをレポートと一緒に提出。データのコピーをすべて消去すること。

9

第10講「平均値の層別比較」

1. SPSS での平均値と標準偏差の計算
2. 層別 (group 別) 比較
3. 平均値を使うときの注意事項
4. Effect Size
5. 相関比

1

【平均値を使うときの注意事項】

- ★ 平均値は、外れ値の影響を受けやすい。
あまりにかけはなれたケースがあるときは
 - ・ 上下数%を取りのぞく (調整平均: 教科書 p. 46)
 - ・ 順位に変換したり中央値を使って分析
- ★ 左右対称でないデータでは平均値より中央値の方が適切な代表値であることが多い

2

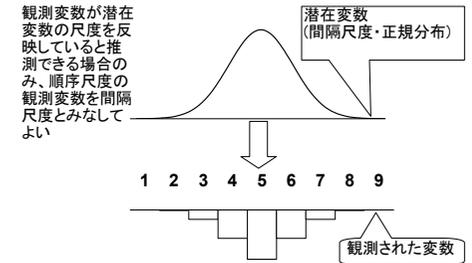
★ 順序尺度の平均値をとっていいのは

- ・ 潜在的には間隔尺度のはず
- ・ 測定のポイントが一定間隔

という2条件をともに満たす場合

※ 2 値の変数は間隔尺度とみなせるが、若干の注意が必要。

3



4

具体的には

- 4 点以上の尺度
 - 正規分布に近似 (教科書 p. 53-59):
 - ・ 単峰性
 - ・ 左右対称性 (歪度)
 - ・ 中央への集中度 (尖度)
- ヒストグラムを描いて検討するとよい。
- 正規分布との乖離度を統計的に検討する手法もある

5

歪度・尖度は「度数分布表」の「統計」オプションで指定できる

正規分布のとき0、絶対値が大きくなるほど、正規分布から外れる

これらの条件を満たさない場合は

- 非線形変換 (教科書 p.142-144)
- 順位に変換したり中央値を使って分析

6

【平均値の層別比較】

ふたつの層の間の平均値の比較

- ★ 平均値の差をもとめる (層別平均)
- ★ 標準偏差を基準にして差を評価 (effect size)

7

【SPSS のコマンド】

「平均の比較」→「グループの平均」

従属変数=平均値を求める変数 (間隔尺度)
独立変数=層を指定する変数 (名義尺度)

8

【エフェクト・サイズ】

$$ES = \frac{\text{平均値の差}}{\text{標準偏差}}$$

★ 正式には層別 SD の重みつき平均のような数値 (併合 SD) をつかう (教科書 p. 137)

9

【例】

性別による生活全般満足度の違い

	平均	SD	(人数)
男性	2.62	1.02	(114)
女性	2.24	0.91	(136)
合計	2.41	0.98	(250)

$$\begin{aligned} \text{平均の差} &= \\ \text{併合 SD} &\approx \\ \text{ES} &= \end{aligned}$$

※ ES は SPSS では計算してくれない

10

【ES の特徴と問題点】

- ★ 各層の人数を考慮せず平均値だけ比較
 - ➔ 大きさがちがう場合は?
- ★ 2 層間の比較だけ
 - ➔ 3 つ以上の層を比較したい場合は?

11

【相関比】

- ★ 各層の個体が全員その層の平均値を持つ状況を仮定して SD を求める
- ★ この仮想 SD を実際の SD で割った数値が「相関比」。η (イータ) であらわず
- ★ 相関比の 2 乗 η² を「決定係数」「分散説明率」などという
 - ※ η² を「相関比」ということもある

12

【SPSS コマンド】

「平均の比較」→「グループの平均」

「オプション」の「第1層の統計」で「分散分析表とイータ」をチェック

- ★ η は 0~1 の範囲の値をとり、独立変数の影響力をあらわす

※ ES は最小値 0、最大値 ∞

13

- ★ 3 層以上で平均値を比べる場合にも相関比が使える。

このように、層別平均値をあてはめて仮想分散を求める分析法を「分散分析」(ANOVA: ANalysis Of VAriance) という。

14

【宿題】

相関比は、なぜ層別の平均値の比較に使えるのか。説明せよ。

15

1. 相関比の意味

2. エフェクト・サイズと相関比

3. 表と図の書きかた

1

【相関比の意味】

ある個体の値を x 、全体平均を M 、層別平均を m とすると、全体平均との差（偏差）は

$$x - M = (x - m) + (m - M)$$

2

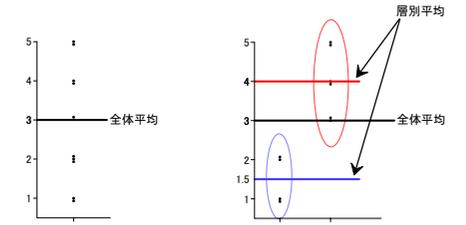
次のデータの平均値と SD は？

{1, 1, 2, 2, 3, 5, 4, 5, 4, 3}

これをふたつの層に分割すると：

{1, 1, 2, 2} {3, 5, 4, 5, 4, 3}

3



4

【モデルとデータの乖離】

連関係数も相関比も、モデルとデータの乖離を表した値と解釈できる

- 特定の仮定（モデル）の下で予測される値（期待度数・仮想 SD）を求める
 - 実際のデータの値と比較する
 - 0～1 の範囲の係数になるように調整する
- 多くの統計手法がこのタイプに属する

5

【ES と η の関係】

$$ES^2 = \frac{\eta^2}{1 - \eta^2} \times \frac{N^2}{n_1 n_2}$$

特に、2層の大きさが同じ ($n_1 = n_2$) なら、

$$ES^2 = \frac{4\eta^2}{1 - \eta^2}$$

層の大きさがちがえば、ES はこれより大きくなる

6

※ このように ES と η は互いに変換できる。

→ 両方示すのは冗長

7

【注意事項】

層別の平均値を分析する場合、各層の人数は一定以上必要

(最低 20 人?)

→ カテゴリ統合が必要になることがある

8

【表に書くべき要素】

- 各層と全体の平均値と標準偏差（測定水準の 2 桁下まで）
- 各層と全体の人数
- 相関比またはエフェクトサイズ（小数第 3 位まで）
- 欠損数とその原因

9

表 1 保守的意識の男女差

	平均	標準偏差	(人)
男性	4.15	1.01	(109)
女性	3.57	1.26	(130)
合計	3.83	1.18	(239)

「以前からなされていたやり方を守ることで、最上の結果を生む」に対する回答: 「1. そう思う」～「5. そう思わない」
相関比 $\eta = 0.244$ 、無回答 = 11.

10

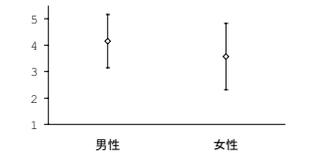
【グラフの書きかた】

平均値をプロットし、上下に SD を表示する。

Excel では

- ・折れ線グラフを描く
- ・マーカーの上でダブルクリック
- ・「Y 誤差範囲」をえらぶ
- ・「指定」で SD の値が入っているセル範囲を指定

11



「以前からなされていたやり方を守ることで、最上の結果を生む」に対する回答: 「1. そう思う」～「5. そう思わない」
相関比 $\eta = 0.244$ 、N = 239、無回答 = 11.

図 1 保守的意識の男女差 (平均±標準偏差)

12