

比較現代日本論研究演習 I

大学院生対象 : 2002 年度前期
<金 2> コンピュータ実習室 (文学部本館 7F 711-2)

『講義概要』 p. 396 記載内容

- ◆ 授業内容 : 小規模標本調査を念頭において、統計分析の基礎的な手法を学ぶ。講義・講読をとおした知識の習得と統計パッケージを用いた実習とを交互におこない、調査データの適切な加工ができる基礎的な分析力を身につけることを目指す。
- ◇ 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。
- ◇ 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 (4/12)
2. SPSS 入門 (4/19)
3. 統計分析の基礎 (4/26)
4. 記述統計(1): 度数分布とクロス表 (5/10~6/7)
5. 中間試験 (6/14)
6. 記述統計(2): 平均値の比較 (6/21~7/19)

※ () 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
- 他のソフトウェアについて (電卓, Excel, Word?)

2. SPSS 入門

- SPSS の概要
- 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
- 度数分布表

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
- サンプルングの概念と手順

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

- 4 点相関係数とクラマーの V
- 公表に際してなにを書くべきか

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. 平均値の比較

- 平均の差と差の平均
- 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/g020412.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 I 統計分析入門]

第 1 回 (2002-04-12)

この授業の概要・スケジュール・評価方法

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。

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

ディスプレイの電源を落とすのを忘れないこと。

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 My Document 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ディスクにコピーして、内蔵ディスクのほうのファイルは削除すること。

フロッピー (3.5 インチ) を各自購入しておくこと。

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、ISBN 4-563-00888-5。

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。

「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いじらなくてよい
- 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って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 My Document 内に保存してみる。

「エクスプローラ」で My Document 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。(フロッピーにコピーする必要はない。)

※ 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。

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 2002 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 (4 月)
- クロス表分析 (5-6 月)
- 平均値の比較 (6-7 月)

2

★ 推測統計は後期の
「研究演習 II」であつかう。

→連続履修することがのぞましい

3

【教科書】

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。

※生協文系書籍部に入荷済み

4

【コンピュータ実習室について】

- ★ 入室に**学生証**が必要
- ★ 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

5

【コンピュータの起動と終了】

6

【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 My Document
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

7

必要なデータは各自でフロッピー
にコピーして持ち帰る

→ フロッピーディスクを
各自で購入しておくこと。

8

【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

9

【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。

10

模擬データ入力実習

11

【その他のソフトウェア】

- 表計算 (Excel)
- 電卓 (「アクセサリ」のなか)
- ワードプロ (Word)

12

【印刷】

- ★ モノクロプリンタ 2 台
- ★ 出力先の切り替えかた
- ★ 印刷前にプレビューで確認

13

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

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. 『日本の階層システム』(全 6 巻)
東京大学出版会、2000 年。

2

- ★ 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

3

【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)

4

★ 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

5

【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」

6

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

7

【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。

8

- ★ 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。

9

- ★ 層化抽出法＝母集団を層別にわけ、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

10

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

11

【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。

12

【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 (.)

13

【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

14

【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 (▶)

15

【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

16

- 「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

17

【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

18

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

1

【データ収集から分析まで】

- データの収集
 - ・ **実験**……状況を人工的に統制
 - ・ **観察**……自然の状態そのまま (調査)

※この授業では観察を念頭にしている

2

- 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

【変数の種類】

- 名義尺度 (nominal scale) (質的変数とも)
 - 順序尺度 (ordinal —)
 - 間隔尺度 (interval —)
 - 比率尺度 (ratio —)
- (教科書 p. 8)

4

【尺度の変換】

- ★ 上位の尺度は下位の尺度の特徴を兼ね備えている
- ★ 上位の尺度のほうが使える演算が豊富

→ 分析手法の選択幅が広い

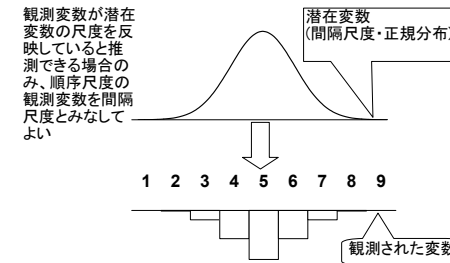
5

私たちが測定するものはいてい**順序尺度**

- ★ 上位の尺度への変換には一定の理論的根拠が必要

※ 2値の変数は間隔尺度とみなせるが、若干の注意が必要。

6



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

- ★ 最頻値 (mode)
- ★ 中央値 (median), 4分位 (quartile), %点 (percentile)

(教科書 p. 43-50)

9

【棒グラフとヒストグラム】

- 棒グラフ……棒同士の間を空白をあける。**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% (状況次第)

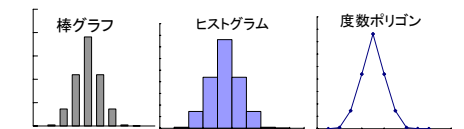
10

- ★ 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 (順序尺度/離散量) → **棒グラフ**

11

※**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)



12

SPSS では histogram は書きにくい

- ★ 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

13

- ★ 出力ビューアの内容は copy& paste できる:
→ Excel などでグラフを描く

14

【実習】

- (1) 本人年齢の度数分布表を出力し、中央値と上側 10% 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

15

【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」
- そのあとで度数分布表を書いてみる

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

		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

4

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

比較現代日本論研究演習I (田中 重人)
2002.5.24 課題

氏名:

学年:

所属:

学籍番号:

つぎのクロス表に周辺度数, 行%, 列%を書き入れよ

x	y			合計
	1	2	3	
1	40	56	15	
2	36	75	21	
合計				

(N=243)

第 6 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)

1

【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		合計
	1	2	
1	a	g-a	g
2	i-a	h-i+a	h
合計	i	j	N

2

3×3 クロス表：セル度数が 4 つ決まれば…

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

k×l クロス表の自由度 (degree of freedom)

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

3

【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

4

【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

5

(1) 行%は 1 列についてだけ比較すればよい:

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

(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:

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

6

(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%	

(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%	
78	72	150
52.0%	48.0%	100.0%

7

【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)

8

【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

9

比較現代日本論研究演習 I (田中 重人)

2002.5.31 課題

氏名：
 学年：
 所属：
 学生番号：

周辺度数、%、 ϕ を計算して下の表に書き入れよ。

α	β		合計
	1	2	
1	52	61	
2	37	97	
合計			

行%の差=

列%の差=

ϕ =

比較現代日本論研究演習 I (田中 重人)

2002.5.31 課題 解答例

周辺度数、%、 ϕ を計算して下の表に書き入れよ。

α	β		合計
	1	2	
1	52 46.0 58.4	61 54.0 38.6	113 100.0
2	37 27.6 41.6	97 72.4 61.4	134 100.0
合計	89 36.0	158 64.0	247 100.0

行%の差=18.4

列%の差=19.8

ϕ = 0.191

【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)

1

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%

4

【クラメールの連関係数 V 】

$k \times l$ 表への ϕ 係数の拡張

- ★ 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

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

7

【 ϕ 係数の性質】

1. $\phi = \text{交差積の差} / \sqrt{\text{周辺度数の積}}$
2. $\phi = \text{相関係数の特殊ケース}$
3. $|\phi| = \text{行\%差と列\%差の中間の値}$
4. $\phi^2 = \text{標準残差の総計} / N$
($\rightarrow 2 \times 2$ 以上のクロス表に拡張できる)

2

- ★ 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 (standardised ---) 残差/ $\sqrt{\text{期待度数}}$

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

5

【 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

8

【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:

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

3

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

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

6

【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
ファイとクラメールの V

9

1. 他人に見せる表
2. 表と図のあつかい
3. 表の書きかた
4. グラフの書きかた
5. クロス表を説明する文章

1

【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

2

【他人に見せられない表】

- ★ セル数が多すぎて周辺度数が偏っているもの
期待度数が5未満のセルがあると、
V係数は無意味
→適切なカテゴリー統合を行う必要

※資料としての意味はまた別である

3

- ★ 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

4

【表と図】

表 (table) …活字と罫線で行列型に組む。

図 (figure) …活字・罫線以外の要素を含む。グラフのほか、概念図や写真を使うことも

5

【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

6

【表に書くべき要素】

- 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

7

- ★ 行→列の因果を想定するのがふつうだが、列→行でもよい。(％の「100.0」で区別)
- ★ 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
- ★ 統計的検定をした場合は、その結果も

8

- ★ 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
→タイトルと行・列頭の見出し (heading) を工夫する

9

【Excel による作表】

★SPSS の出力をコピーして
必要なところを残す

★「書式」→「セル」メニューの
「表示形式」「配置」「罫線」
で整形

10

【グラフの書きかた】

クロス表は積み上げ棒グラフで
表現するのがふつう

(2 列表の場合は棒グラフでも)
(度数多角形を重ねる場合もある)

11

★グラフは細かい数字がのせにくいし、紙幅
を食う。基本的には表を使用して、特に視覚
的インパクトを狙う場合に限ってグラフを使
うのがよい

12

【Excel によるグラフ作成】

★注意点：

- ・ データ系列の順序
- ・ 凡例の表示
- ・ 区分線の表示
- ・ 棒同士の間隔

13

【クロス表を説明する文章】

★まず周辺度数分布を説明

★ 線形の関連か？

大小のある変数で、行によって大きいほう
または小さいほうに偏っている関連

14

★どのセルに注目するか？

特徴的なセルをみつけて、それ
をほかの適当なセル (周辺度
数) と比較する。カテゴリ一統
合のセンスがものをいう

15

★ %とポイント差：

「Y は X より 10%大きい」とは

$Y=1.1X$ それとも $Y=0.1+X$?

後者の場合は「10 ポイント大きい」と書く

★表に示される**事実**と、
自分が加える**解釈**との峻別

16

【解釈のツボ】

★実感に照らして納得できるか

★測定・分析のミスではないか

★標本の偏りで説明できないか
(→統計的検定)

17

【先行研究と照合】

★命題を立てる

★既存のデータと整合的か

★ (後知恵ではないかたちで)
理論的に説明できるか

18

授業資料

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。

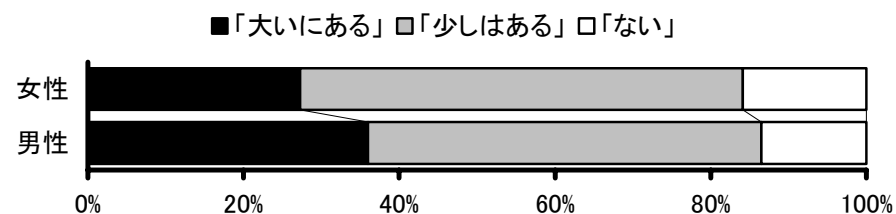


図1 性別と性別による不公平感との関連

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。

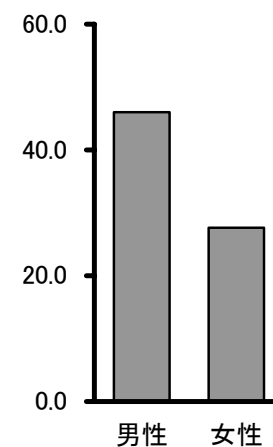


図2 県や市町村の部課長以上の役人に知り合いがいる%の男女差

比較現代日本論研究演習 I (田中重人)

中間試験 (2002.6.28)

配布済みの論文「意識の男女差と結婚の遅れ」(田中重人+秋山哲也 1999) を読み、下記の問題に答えよ。

【回答上の注意】

- ① 他の回答者の画面が見えないよう、互いに離れて座ること
- ② コンピュータで回答を書き、印刷して提出
- ③ 小数の回答については、小数第3位まで書くこと
- ④ 何を持ち込んで参照してもよいが、人に相談してはならない

1. 図表1の45-49才女性回答者の「不詳」セルについて、次の値を求めよ。

観測度数:

期待度数:

2. 図表3の女性回答者についてグラフを作成せよ (Excel を使用)

3. 図表4について、男女それぞれのクラメールの連関係数Vをもとめよ。(計算のプロセスも書くこと)

女性:

男性:

比較現代日本論研究演習 I (田中重人) 中間試験 (2002.6.28)

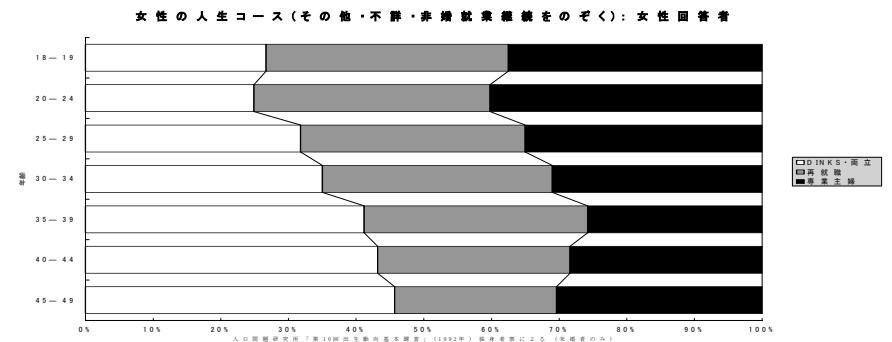
解答

1. 図表1の45-49才女性回答者の「不詳」セルについて、次の値を求めよ。

観測度数: 23

期待度数: 6.873

2. 図表3の女性回答者についてグラフを作成せよ (Excel を使用)



3. 図表4について、男女それぞれのクラメールの連関係数をもとめよ。

女性: $V = \sqrt{21.533 / (3160 \times 2)} = 0.058$

男性: $V = \sqrt{43.243 / (4089 \times 2)} = 0.073$

第 10 回「平均値と標準偏差」

1. 代表値と散布度
2. 平均値
3. 標準偏差
4. SPSS のコマンド
5. 平均値を使うときの注意事項

1

【代表値と散布度】

平均値 (mean) - 標準偏差 (SD)
(間隔尺度以上)

中央値 (median) - 四分位偏差 (Q)
(順序尺度以上)

(教科書 p. 42-51)

2

【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

3

【SPSS のコマンド】

「記述統計」 「記述統計」

変数とオプションを指定

4

【平均値を使うときの注意事項】

- 平均値ははず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%を取りのぞいたデータセットで計算する (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

5

平均値・標準偏差は**間隔尺度**以上のデータ
に対してしか意味をもたない。
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

(4/26 資料参照)

6

- ➡ 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

7

- ➡ 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

8

間隔尺度のデータでも、
左右対称でないものについては
平均値よりも中央値のほうが
適当であることが多い

典型例: 収入・人口など

9

次のデータの平均とSDは?

{0, 1, 4, 5, 7}

10

【課題】

適当な変数について、ヒストグラムの軸上に
平均値と標準偏差を（手書きで）書き入れた
ものを作って提出

11

1. 平均値の層別比較
2. SPSS のコマンド
3. エフェクト・サイズ
4. 分散分析と相関比

1

【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; 相関比)

2

【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)

3

【エフェクト・サイズ】

ES = 平均値の差 / 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)

4

【例】

性別による不公平

	平均	SD	(人数)
男性	1.77	0.67	(111)
女性	1.89	0.65	(132)
合計	1.84	0.66	(243)

平均の差=0.11 併合 SD=0.66
ES=0.17

※ ES は SPSS では計算してくれない

5

【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？

6

【相関比】

- ★ 各層の個体が全員その層の平均値を持つ
ような状況を仮定して 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) であらわす
- ★ 相関比の 2 乗 η^2 を
「決定係数」「分散説明率」などという

※ η^2 を「相関比」ということもある

7

- ★ SPSS では、「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
影響力が大きいほど大きくなる
- ★ 同じ大きさの 2 層で平均値を比べる場合、

$$ES^2 = \frac{4\eta^2}{1-\eta^2} \text{ という関係がある。}$$

層の大きさがちがえば、ES はこの式よりも大きくなる

※ ES は最小値 0、最大値 ∞

8

- ★ 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。

9