

比較現代日本論研究演習 I 「統計分析入門」

大学院生対象: 2012 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=LM14206

『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。
- ◇ 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 (4/12)
2. SPSS 入門 (4/19)
3. 統計分析の基礎 (4/26, 5/10)
4. 度数分布とクロス表 (5/17~6/7)
5. 中間試験 (6/14)
6. 平均値の比較 (6/21~7/19)
7. 予備日 (7/26)
8. 期末レポート (8/14 提出) → 9/11 以降に返却

※ () 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2) および
- 比較現代日本論研究演習 III 「実践的統計分析法」(火4: 学部と合同)

も受講することがのぞましい。

受講者との連絡とフィードバック

- 毎回の課題・宿題は、コメントをつけて返却します (内容によっては再提出を求めることもあります)。

- 中間試験、期末レポートは、採点後に返却します。
- 課題・宿題は、特に指示のあるものをのぞき、ISTU（東北大学インターネットスクール：<http://www.istu.jp>）のレポート機能による提出とします。提出期限は、原則として授業前日（ 昇水曜）正午（12:00）です。
- ISTU には、この授業の「受講申請」をしておいてください（受講者情報の自動的登録は、履修登録完了以降）
- 研究生などで東北大 ID のないかたは、所属学部の教務係に相談してみてください。
- 教員からの連絡は、ISTU「お知らせ」「掲示板」のほか、個人ブログ <http://tsigeto.blog.fc2.com/blog-category-14.html>（RSS フィード利用可）に出る場合があります
- オフィス・アワーは定めていません。教員への相談は、適当な時間に予約をとってください。

内容の詳細

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
- SPSS の起動

2. SPSS 入門

- 模擬データ入力実習
- データの配布と説明
- データ行列（データセット）とは
- メニューによるシンタックス作成
- 変数値の再割り当て
- その他のソフトウェアについて

3. 統計分析の基礎 [序章]

- 記述統計と推測統計
- 標本調査とは
- データの種類（尺度水準）

4. 度数分布とクロス表

4.1. 度数分布表 [1 章]

- 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 によるグラフ作成

4.2. クロス表 [4 章]

- 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 章]

- ϕ 係数
- 期待度数・残差と
- Cramer の連関係数 V
- 表とグラフの書きかた

5. 中間試験

6. 平均値の比較

6.1. 平均と分散 [2 章]

- 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 [5 章]

- 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

[] 内は、教科書の参照箇所

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 (正規の学生以外は、登録申し込みが必要。ない人は、教務係で臨時カードを借りること)。文学部正規学生以外 (研究生や他学部の学生など) は登録が必要。

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
- 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「シャットダウン」を選択
- コンピュータ本体の電源が切れたことを確認
- 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。

- 1. 模擬データ入力実習
- 2. データ配布
- 3. SPSS の基礎知識
- 4. 変数値の再割り当て

1

【模擬データ入力実習】

別紙

2

【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 (次回説明)
- ★ 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。

調査票は <http://srdq.hus.osaka-u.ac.jp/PDF/SSM95BJ.pdf> にもある

3

- ★ 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

4

- ★ 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削除

自分の研究用のデータがある人は、
課題などではそれを使ってもよい。
ただし事前に相談すること。

5

【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 (.)

6

【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

7

【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
- ★ 「OK」をクリック

※ 古いバージョンでは「貼り付け」でシン
タックスを作成して分析の記録を残したほ
うがよい (必要部分を選択して実行できる)

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
- ★ Ver. 19 ではシンタックスも表示される

【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

【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

10

【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」
→ 「他の変数への値の再割り当て」
- 変換先変数の名前をつける (「変更」)
- 「今までの値と新しい値」
- 値の組を指定したら「続行」
- 「OK」ボタンを押して実行

11

- 新変数ができていることを確認
- 度数分布を確認
- 問題がなければ、名前をつけてデータセッ
トを保存
(どこに保存されるかを確認しておくこと)

12

【実習】

本人年収 (Q44_1) を適当なカテゴリに区切
って度数分布表を出力する。
出力ビューアの内容を保存して .spv ファイル
を提出 (シンタックス部分を含めること)

13

模擬データ入力実習

SPSS の起動

スタートメニューから「プログラム」→「IBM SPSS Statistics」→「IBM SPSS Statistics 19」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 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。

第3講「統計分析の基礎」

- 1. データ収集から分析まで
- 2. 標本抽出
- 3. 度数分布表

1

【カテゴリ統合の方針】

- 何かの意味があるところで区切る
- 人数が均等になるように区切る

2

【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

3

- 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)

4

【標本抽出の4段階モデル】

- 目標母集団 (universe)
- 調査母集団 (population)
- 計画標本 (designed sample)
- 有効標本 (valid sample / case)

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」が完備していない状況でも、工夫次第で無作為抽出に近づけることができる

7

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

8

【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)

9

【宿題】

自分の関心に沿って論文を探し、その論文中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。

(1) その論文の書誌情報と、目標母集団・調査母集団・計画標本・有効標本について簡単にまとめたもの

(2) その論文のなかで、データについての説明の部分

人数分コピーを用意してきて、次々回授業時に報告。

10

【度数分布表】

Frequencies コマンド
「分析」
→ 「記述統計」
→ 「度数分布表」

11

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)

12

【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)

13

第4講「グラフの利用」

- 1. 棒グラフ・ヒストグラム・度数ポリゴン
- 2. SPSS でグラフ作成
- 3. Excel でグラフ作成

1

【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

2

【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%

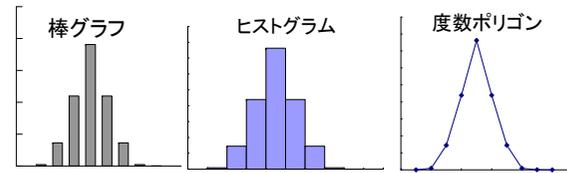
3

- ★ 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便利。

(教科書 p. 32-36)

4



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→レガシーダイアログの「ヒストグラム」では等間隔の区間に分割してくれる

5

Excel を使う場合：

- ★ 度数分布表を出力 (必要なら値を再割り当て)
- ★ 表を Excel にコピーする
- ★ 必要なら変数値のラベルをつける (横軸用)
- ★ ヒストグラムや度数多角形の場合は両端に度数 0 の行をつくる

6

通常は、縦の棒グラフ。
 度数ポリゴンは、折れ線グラフで描く。
 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」
- 「要素の間隔」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。

7

【実習】

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

8

【宿題】

問 27 の項目からふたつをえらび、
 度数ポリゴンを描いて比較せよ。
 グラフにコメントをつけて提出
 (ISTU で 5/16(水) 12:00 まで)

9

【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

【度数ポリゴンの比較】

- ・ 「有効パーセント」の数値を並べてコピー
- ・ 両端に度数0のセルをつくるのが正式

2

【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

3

- 「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

4

【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

(教科書 第4章)

SPSS では

「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) 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、%からわかることをコメントする。
表・グラフにコメントをつけて提出
(ISTU で 5/23 12:00 まで)

9

第 6 講 「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) 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：

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

6

(5) これら以外の場合、行%の差と列%の差はちがう値になる

(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%

8

【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)

9

【キーワード】

連関 (association), 独立 (independence),
期待度数 (expected frequency)

1

【φ係数の性質】

1. $\phi = \text{交差積の差} / \sqrt{\text{周辺度数の積}}$
2. $\phi = \text{相関係数の特殊ケース}$
(→ 2学期授業)
3. $|\phi| = \text{行\%差と列\%差の中間の値}$
(教科書 p. 103 表 4-1 について計算してみよう)

2

4. $\phi^2 = \text{標準残差の2乗の総計} / N$
(→ 2×2 以上のクロス表に拡張できる)

3

【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 :

$$\begin{array}{c|c} \frac{gi/N}{hi/N} & \frac{gj/N}{hj/N} \end{array}$$

4

各セルの期待度数は?

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

5

- ★ 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/ $\sqrt{\text{期待度数}}$

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

6

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%

7

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

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

8

【課題】

教科書の表 4-1 について
期待度数・残差・標準残差・ χ^2 を求める

ISTU で提出 (6/6 12:00 〆切)

9

1. 連関係数とその性質
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

【モデルとデータの乖離】

連関係数は、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) 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

4

【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

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

【宿題】

つぎのクロス表を整形したうえで解釈：

- ・ 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)

必要に応じてカテゴリーを統合すること

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.

人に見せる表

- ・ カテゴリーの並べ順や行列の組み合わせをわかりやすく
- ・ 変数とカテゴリーの命名
- ・ 表のタイトルとして適切なものをつける

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

書くべき要素

- ・ 各セルの行 (または列) %
- ・ 行 (または列) 合計の度数と 「100.0%」
- ・ 列 (または行) 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

第9講「平均値と標準偏差」

1. 平均値と標準偏差の計算
2. 尺度水準と代表値
3. 平均値を使うときの注意事項
4. SPSS コマンド

1

【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

2

★ 次のデータの平均と SD は?

値	偏差	偏差 ²
1		
2		
4		
6		
7		

平均 =

平方和 =

分散 =

SD =

(教科書 p. 42, 48)

3

【宿題】

教科書 p. 52 の練習問題 2-3 について、
同様の表をつくり、
平均値と標準偏差を計算せよ。

(ISTU で、水曜正午まで)

4

【尺度水準】

- 私は **20** 歳です
- 今年は **2012** 年です
- 今日は **9** 回目の授業です
- 郵便番号は **980** です

これらの違いについて教科書 pp. 7-16 をもとに考えてみよう

5

- 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)

6

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

7

上位の尺度への変換には
一定の理論的根拠が必要

8

【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)

9

第10講「平均値に関する注意事項」

1. SPSS コマンド
2. 平均値を使うときの注意事項

1

【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

2

【平均値を使うときの注意事項】

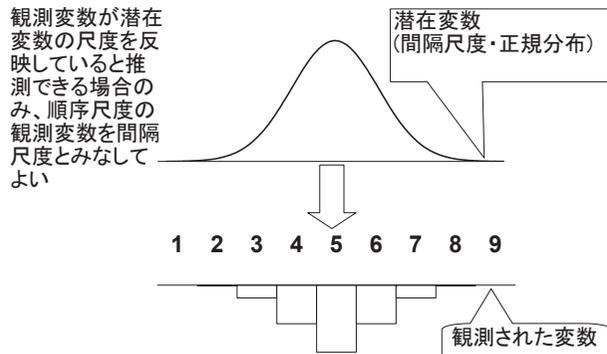
★順序尺度の平均値をとっていいのは

- ・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)

SPSS で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。

7

【宿題】

「生活全般満足度」(Q37) について
男女別に度数ポリゴンを作成し、そこに
平均値と標準偏差を書き入れたものを作成

ISTU で来週月曜の正午までにファイル提出

8

【期末レポート】

期限: 8/14 (火)

提出先: ISTU 「期末レポート」にファイルを提出

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、データのコピーをすべて消去すること。

9

第 11 講「平均値の層別比較」

1. 層別 (group 別) 比較
2. Effect Size
3. 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 または 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】

$$ES = \text{平均値の差} / \text{標準偏差}$$

★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)

4

【例】

性別による生活全般満足度の違い

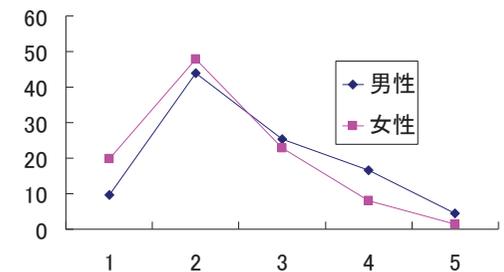
	平均	SD	(人数)
男性	2.62	1.02	(114)
女性	2.24	0.91	(136)
合計	2.41	0.98	(250)

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

※ ES は SPSS では計算してくれない

5

性別による生活満足度の違い



6

【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？

7

【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

8

【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 ∞

9

1. 相関比の意味

2. エフェクト・サイズと相関比

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

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

$\eta = 1$: すべての x について $x = m$

$\eta = 0$: すべての m について $m = M$

2

次のデータの平均値と SD は?

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

2 層に分割すると:

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

層別平均値をあてはめると:

{1.5, 1.5, 1.5, 1.5} {4, 4, 4, 4, 4, 4}

3

次の場合はどうか?

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

↓

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

↓

{3, 3, 3, 3, 3} {3, 3, 3, 3, 3}

4

$\eta = 1$: 層内のばらつきがない
(全員同じ値)

$\eta = 0$: 層別の平均値が同じ

5

【分散分析】

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。

3 層以上で平均値を比べる場合にも使える。

6

【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

7

※ 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

8

【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

9

1. モデルとデータの乖離

2. 表と図の書きかた

1

【モデルとデータの乖離】

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

- 「モデル」は何か？
- データとの乖離はどうやって計算しているか？
- 係数の取りうる値の範囲は？

2

【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

3

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が、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.

4

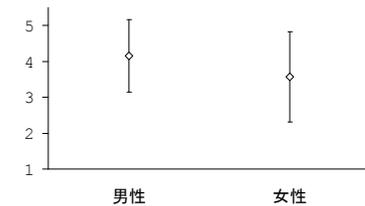
【グラフの書きかた】

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

5



「以前からなされていたやり方を守ることが、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)

6

【今後の予定】

- ・ 授業内容は今日で終了
- ・ 来週は補習 (出席任意)
- ・ 期末レポートは 8/14 締切 (ISTU)
- ・ レポート返却: 9/4-6 (火-木) 11:00-13:20

7