

現代日本論演習「統計分析の基礎」

3年生対象：2012年度5セメスタ

<火4>コンピュータ実習室(文学部本館 7F 711-2) 授業コード=LB52407

『講義概要』 記載内容 + α

- ◆ 講義題目：統計分析の基礎
- ◆ 到達目標：(1) 統計分析の基礎的な手法を理解する; (2) 実際に統計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、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。
- ◇ 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。
- ◇ 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。

卒業論文等で質問紙調査を予定している者は、5セメスタ開講の現代日本論演習「質問紙法の基礎」(金2)および6セメスタ開講の現代日本論演習「実践的統計分析法」(火4：大学院と合同)も受講することがのぞましい。

授業の概要

1. イントロダクション(4/10)
2. SPSS 入門(4/17)
3. 統計分析の基礎(4/24)
4. 記述統計(1): クロス表の分析(5/8~5/29)
5. 中間試験(6/5)
6. 記述統計(2): 平均値の比較(6/12~6/26)
7. 推測統計(7/3~7/24)
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(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)

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(月曜)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-11.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。

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
- 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
- frequencies コマンドと度数分布表
- Excel によるグラフ作成
- 印刷

3. 統計分析の基礎 [序章]

- 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 [1 章]

- 度数分布表
- パーセンタイル、中央値、四分値
- 度数分布表のグループ化

4.2. クロス表 [4 章]

- 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・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 [2 章]

- 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 [5 章]

- 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 [6 章]

- データの記述と誤差の評価
- 標本抽出の 4 段階モデル
- 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 [8 章]

- 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 [8 章]

- 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 χ^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

[] 内は、教科書の参照箇所

現代日本論演習 (田中重人)

受講登録フォーム

氏名 (よみがな):

学年:

学籍番号:

所属 (文学部日本語教育学専修以外の場合):

研究内容:

- ・ 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 ()
- ・ つぎのうち利用したことがあるものは? **ISTU / DCW / SRP**

以下は採点用

	4/17	4/24	5/8	5/15	5/22	5/29	6/5	6/12	6/19	6/26	7/3	7/10	7/17	7/24
宿題														
課題														
参加														

中間			
期末			

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 (gradient) と切片 (intercept) はそれぞれいくつか。

傾き = _____ ; 切片 = _____

(2) 「必要十分条件」 (necessary and sufficient condition) とは何か。簡単に説明せよ。

(3) 「偏差値」はどのような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

(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

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

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。

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。

- 1. データ配布
- 2. 標本抽出
- 3. SPSS の基礎知識

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

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

- ★ 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

4

【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)

5

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

6

【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)

7

【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 (.)

8

【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

9

【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)

10

【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
- ★ 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, ...)

11

【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

12

【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

13

- 「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、名前をつけて
データセットを保存

14

【実習】

満年齢 (Q1_2a) を適當年齢幅に区切って
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

15

1. 尺度水準
2. 度数分布表
3. 累積%とパーセンタイル
4. グラフの利用

1

【課題】

教科書 pp. 7-16 をもとに、
「データの種類」の分類について
まとめよ。

2

【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)

3

- ★ 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

4

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

5

【度数分布表】

Frequencies コマンド

「分析」

→「記述統計」

→「度数分布表」

6

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)

7

【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)

8

【宿題】

- (1) 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。
- (2) この変数は、どの尺度水準にあたるか。
- (3) この変数を比率尺度として扱えるよう変
数値の再割り当てをおこない、シンタックス
と度数分布表を出力せよ。

5/7 12:00 までに IST U で提出

9

第4講「グラフの利用」

- 1. 変数値の再割り当て (再)
- 2. 棒グラフ・ヒストグラム・度数ポリゴン

1

【変数値の再割り当て】

収入を比率尺度の値に割り当てなおすには
→ 変数値の再割り当て

※ SSM 調査データ作成時に、収入の値は
17: 2300 ~ 2800
18: 2800 ~ 3300
.....
のようにコードされている。

2

「学歴」の場合はどうすればよいか?

- 1 → 6
- 2 → 8
- 3, 4, 5 → 11
- 6 → 14
- 7 → 17
- 12 → 9
- 13 → 12
- 14 → 14
- 15 → 16
- 16 → 18

3

【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

4

【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%

5

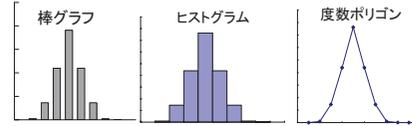
★ 連続量を階級分けした場合
→ ヒストグラム

★ 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)

6



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

7

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ときは両端に度数 0 の行をつくる

8

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。

9

【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)

10

【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 5/14 12:00 まで)

11

【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

(教科書 第4章)

4

		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、
%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 5/21 12:00 まで)

次回 (5/22) は観察室で授業

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】

- φ = 交差積の差 / √(周辺度数の積)
- φ = 相関係数の特殊ケース
(→ 2学期授業)
- |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)

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:

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

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/√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

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

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

9

【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

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

10

【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

11

【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

12

【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

13

【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

14

【宿題】

つぎのひとつについてクロス表と V を出力:
・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?

15

【次回予告】

6/5 は中間試験をおこないます。
持ち込み可。範囲は、今日の授業内容まで。
試験後は、通常の授業をおこないます。

16

第9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

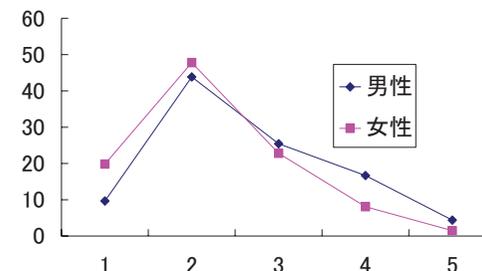
	平均	SD	(人数)
男性	2.62	1.02	(114)
女性	2.24	0.91	(136)
合計	2.41	0.98	(250)

平均の差 =
併合 SD \approx
ES =

※ 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
3. 推測統計の基礎
4. 区間推定

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解析法を「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

【宿題】

適当な変数の平均値の年齢層によるちがいについて、分散分析を行い、結果とそれについての解釈を提出。

年齢層は 3 つ以上に分割すること。

(ISTU で来週月曜 12:00 まで)

10

【記述統計と推測統計】

記述統計 (descriptive statistics)

= データ (ケース) の特徴を
数値や図表にまとめる

推測統計 (inferential statistics)

= 確率的な誤差を考慮して、
母集団の特徴を推測する

(教科書 pp. 3-5)

11

【無作為抽出】

random sampling

母集団から計画標本を選ぶ際に、

すべての個体の抽出確率が等しくなる

ように抽出する

→ 「等確率標本」(probability sample)

12

袋のなかに色つきの玉が 60 万個入っている:

赤玉: 30 万個

青玉: 30 万個

玉を n 個取り出したとき、その色は……?

→ 全世界から n 人を無作為抽出したとき、そのなかに ○○ の人は何%ふくまれるか?

13

【区間推定】

interval estimation

「答えは たぶん この範囲内にある」

↓

信頼率 (confidence level) を適当に設定して

信頼区間 (confidence interval) を求める

14

赤・青の玉が多数入った袋から 400 個抽出:

赤玉: 240 個

青玉: 160 個

袋の中の赤玉の比率は?

$$0.6 \pm 1.96 \times \sqrt{(0.6 \times 0.4 / 400)}$$

答: _____ ~ _____ % (95%信頼区間)

15

【比率の区間推定】

標本の規模がじゅうぶん大きく ($n > 30$)、

比率があまり偏っていない ($0.1 < m < 0.9$) とき、

95%信頼区間は

$$m \pm 1.96 \times \sqrt{\frac{m(1-m)}{n}}$$

標準誤差
(standard error)

16

第 11 講「区間推定」(2012.7.3)

1. 区間推定の方法と論理

2. 平均値の区間推定

1

【区間推定の方法】

確率不明のくじを 10 回ひいたところ、すべて
当たりであった。このとき、当たりくじの確率
はどれくらいだと考えるのが合理的か？



可能性の少ないケースを排除

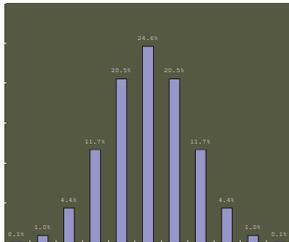
具体的には…… さまざまな仮説を立て、
確率 < 0.025 のものを否定する。

2

- ・ 「信頼率」を適当に設定 (通常 95%)
- ・ (1 - 信頼率) の確率を両極の事象に設定 (通常 2.5% ずつ)
- ・ 母集団における値がいくつであればこの両極端を除いた区間に測定値が入るかを、総当りで計算する
- ・ こうして求めた、母集団においてありうる値の集合が「信頼区間」である。

3

当たる確率 0.5 のくじを 10 回引いた場合



4

標本について計算できる統計量については、
すべて統計的推測が可能である

ただし、計算方法はさまざま

通常、適当な仮定を置いて計算を簡略化する

5

【平均値の区間推定】

母集団における平均値の 95% 信頼区間は次の式で求める。
ただし、正規分布を仮定：

$$\underbrace{m}_{\text{標本平均}} \pm \underbrace{1.96}_{t \text{ 臨界値}} \times \underbrace{\frac{SD}{\sqrt{n}}}_{\text{標準誤差}}$$

※ 「t 臨界値」は n によって変化するが、n > 200 で 1.96 に収束 (教科書 p. 281)。

6

【SPSS コマンド】

「分析」 → 「記述統計」 → 「探索的」

- ◎ 「従属変数」を指定
- ◎ パネル左下の「統計」だけをチェック

※ 信頼率を変更するには「統計」を選択
 ※ 「因子」を指定すると層別に分析できる

7

【宿題】

適当な変数の平均値について、男女別に区間
推定を行う。結果とそれについての解釈を
書いて提出。

(ISTU で来週月曜 12:00 まで)

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。
 推測統計の結果を含めること。
 図表は読みやすく整形し、論文としての体裁を整える。

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 12 講「統計的検定」(2012.7.10)

1. 平均値の差の区間推定
2. 統計的検定とは
3. 平均値の差の検定
4. 有意確率

1

【平均値の差の区間推定】

「平均値の比較」→「独立したサンプルの T 検定」

- ◎ 「グループ化変数」は、数値を指定しないといけない。
連続量を一定の値で切ることもできる

出力は「独立サンプルの検定」の 1 行目
「等分散を仮定する」を見る

2

【統計的検定】

Statistical test

統計的検定 = 特定の値を設定して、その値が信頼区間に含まれているかどうかを判定する
0 に設定するのがふつう

3

【統計的検定用語】

帰無仮説 (null hypothesis):

母集団における統計量が
この「特定の値」に等しい、という仮説

有意 (significant): 「特定の値」が信頼区間に
入っていないことをあらわす

危険率 (critical level): 1 - 信頼率

4

平均値の差の検定の場合:

「5%水準で有意」とは……

- 95%信頼区間が 0 をふくまない
- = すくなくとも 95%の確率で、
母集団において平均値の差がある
といえる

5

「5%水準で非有意」とは……

- 95%信頼区間が 0 をふくむ
= 母集団においては平均値の差がない
という可能性を無視できない
- 平均値の差があるとはいえない

6

【有意確率とは】

信頼区間をひろげていくと、
どこかでゼロをふくむようになる

→このときの危険率のことを「有意確率」または「p 値」という。

7

分析の際は、

- ・ 前もって危険率を設定しておく
(通常は 5%または 1%)
- ・ 有意確率とその値を
下回っているかどうか判別する

例:

- 有意確率が 0.007 →
- 有意確率が 0.023 →
- 有意確率が 0.088 →

8

【宿題】

男女別に平均値の差の検定をしたときに有意な差がある変数とない変数をひとつずつ探す。
結果とそれについての解釈を書いて提出。

(ISTU で来週月曜 12:00 まで)

9