現代日本学演習 V

実践的統計分析

田中重人 (東北大学文学部准教授)

3 年生対象：2022 年度 2 学期 (6 セメスタ) ＜金 4＞ Google Classroom クラスコード **yva6hj3**

1 授業の概要

(『講義概要』記載内容)

授業題目: 実践的統計分析法

学習目標: さまざまな統計分析手法を理解し、使いこなせるようになる

授業内容: 研究の現場で必要となる統計分析手法は、分析の目的とデータの特徴によってさまざまです。この授業の前半では、推測統計学の基本的な概念について解説し、統計的推定および検定の方法について学びます。後半では、さまざまな分析手法をとりあげて、それらの特徴と使い方を習得していきます。どのような分析手法をとりあげるかについては、受講者の関心と必要性を考慮します。統計解析パッケージを使ってデータ分析の実習をおこないます。

履修要件: 5 セメスタ開講の 現代日本学演習 II「統計分析の基礎」 を履修済みか、それと同等の知識を習得済みの者を対象とする。

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.

成績評価の方法: 授業中の課題と宿題 (70 %)、期末レポート (30%) を合計して評価する。

2 授業の予定

(1) 推測統計 (10/7～10/21)
(2) 相関係数 (11/4～11/25)
(3) 対応のあるデータの分析 (12/02～12/9)
(4) 多変量解析 (12/16～1/20) 〔回帰分析を予定しているが、受講者の希望を受け付ける〕
(5) 期末レポート (2/3 提出期限)

3 復習事項

3.1 SPSS の操作

- データエディタにおける「変数ビュー」の使いかた
- 「欠損値」とは何か
- シンタックスとは何か
- 度数分布における「パーセント」と「有効パーセント」のちがい
- 変数値の再割り当ての方法
- グループに分割して集計する方法
- 中央値と四分位の求め方

3.2　クロス表

- 「行％」と「列％」の使い分け
- 「独立」とはどういう意味か
- 期待度数と残差の計算方法
- V と $\chi 2$ の計算方法
- クロス表をグラフにするときは、どのような種類のグラフが適切か

3.3　平均値

- 平均値を計算してよいのはどのような場合か
- 標準偏差の計算方法
- エフェクト・サイズと相関比の計算方法

3.4　その他

- 尺度水準とは何か。それはなぜ重要か
- 分析結果を表にするときの一般的な書式
- 棒グラフ、帯グラフ、折れ線グラフなどの書きかた

4　コンピュータ実習室について

- 入室・退室に学生証が必要 (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) は登録しておくこと。
- 土足・飲食・喫煙厳禁。
- 退出時には必要事項を紙に記入。

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

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「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、個人のファイルを置いておくことはてできない。
- 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。

5　連絡先

田中重人 (東北大学文学部現代日本学研究室)

〒: 980-8576 仙台市青葉区川内 27-1 文学部棟 6F

Homepage: http://tsigeto.info/officej.html

オフィス・アワーは定めていない。質問等がある場合は、あらかじめ適当な時間に予約をとること。受講者への連絡は、Google Classroom または電子メールによる。

現代日本学演習 V「実践的統計分析」

第1講　推測統計の基礎

田中重人 (東北大学文学部教授)

[テーマ]　推測統計の基礎: 母比率の区間推定

1　復習

- 記述統計と推測統計 (教科書 pp. 3–5)
- 母集団と標本
- 無作為抽出
- 区間推定と統計的検定の考えかた

2　標本比率 m はわかっているが母比率 M が不明の場合の区間推定

つぎのような情報 (= 標本統計量) から、母集団における統計量 (= 母比率) を推測する → 母比率はたぶん ○○ から ×× の範囲にある (区間推定)

> 袋のなかに色つきの玉がたくさん入っている。ここから 8 個取り出したところ、すべて赤であった。→袋のなかの玉のうち、赤玉の占める比率はどれくらいか?

この例題では、 m=1 である (100%赤玉であった) ことがわかっているが、 M が不明である。このとき、95%信頼区間を求めるには、M を適当に仮定し、その仮定の下で m=1 になる確率を計算することを繰り返す：

- もし $M = 0.9$ なら……
- もし $M = 0.8$ なら……
- もし $M =$ 　 なら……

このようにして、m=1 になる確率が **2.5%以上** である M の範囲を求める。(母集団は無限大の規模であると考えてよい。)

課題 1:　解答を木曜正午までに Google Classroom に提出。プロセスがわかるように書くこと。

累乗 (0.9 の 8 乗など) を求めることが必要になる。Windows の「電卓」ではメニューから［表示］→［関数電卓］に切り替えるとよい。Excel では ^ という演算子が使える (掛け算を 8 回繰り返してもよい)。

3　もっと複雑な例

> 全世界から 400 人を無作為抽出してある意見を訊いたところ、「賛成」と答えた人が 240 人であった。このとき、母集団 (全世界の人々) における賛成の比率の 95%信頼区間を求めよ (欠損値はないものとする)。

原理的には上記とおなじやりかたで計算できるが、計算量が膨大になるので実際的でない。このような問いに答えるためには、「二項分布」(binomial distribution) の知識を利用する。

4　二項分布の簡単な例題

硬貨を 4 回投げて、そのうち表が出る回数 x を数える。表 = ○, 裏 = ▲ であらわすと

▲ ▲ ▲ ▲ (x=0)
▲ ▲ ▲ ○ (x=1)
▲ ▲ ○ ▲ (x=1)
▲ ▲ ○ ○ (x=2)
....
○ ○ ○ ○ (x=4)

どれも等しい確率 (1/16) で起こるとすると、つぎのそれぞれの場合の確率が求められる：

表が 1 回も出ない (x=0) 確率:
表が 1 回出る (x=1) 確率:
表が 2 回出る (x=2) 確率:
表が 3 回出る (x=3) 確率:
表が 4 回出る (x=4) 確率:

課題 2:　解答を木曜正午までに Google Classroom に提出。プロセスがわかるように書くこと。

参考資料

- Wikipedia の「二項分布」の項 <http://ja.wikipedia.org/wiki/二項分布 >
- 高校までの数学の教科書で、順列・組合せと確率・統計をあつかった部分