

比較現代日本論研究演習 III / 現代日本論演習

大学院生対象：2013 年度後期

<水 2> コンピュータ実習室（文学部本館 7F 711-2）

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 / 応用統計分析

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

1 学期 / 5 セメスタ開講の 比較現代日本論研究演習 I / 現代日本論演習「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

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。

授業の予定

目次

1. 推測統計 (10/2～10/23)
2. 相関係数 (10/30～11/6)
3. 中間試験 (11/13)
4. 変数をキーにした分析 (11/20～12/4)
5. 多変量解析 (12/11～1/22)
6. 期末レポート (2/3 提出期限)

※ () 内の日付は、学期前のおおよその計画をあらわしていますが、実際の授業の進行状況によって前後にずれることがあります。

1. 推測統計

- 記述統計と推測統計 (復習)
- 母集団、標本、無作為抽出 (復習)
- 母比率の区間推定
- 平均値の区間推定
- 平均値の差の区間推定と t 検定
- 相関係数の区間推定と χ^2 検定
- サンプル・サイズと検定力
- 検定結果の解釈と表の書きかた

2. 相関係数

- 尺度水準 (復習)
- 相関図
- Kendall の順位相関係数
- Spearman の順位相関係数
- Pearson の積率相関係数
- 相関係数行列
- 欠損値の処理 (pairwise/listwise)

3. 中間試験

4. 変数をキーにした分析

- 個体間変動と変数間変動
- 対応のある分析
- 2項検定と符号検定
- ハッセ図の利用

5. 多変量解析

重回帰分析? (受講者の興味と必要性によります)

ISTU への登録

ISTU <http://www.istu.jp> の「比較現代日本論研究演習 III」に受講申請しておくこと。

確認事項

各自で復習しておくこと

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
- 中央値とパーセンタイルの求め方

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
- ϕ , V , χ^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

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

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
- Excel による棒グラフ、度数ポリゴン、帯グラフ、誤差範囲つき折れ線グラフの書きかた
- ISTU を利用したレポート提出

連絡先

田中重人（東北大学文学部日本語教育学研究室）

〒：980-8576 仙台市青葉区川内 27-1 文学部・法学部合同研究棟 2F

Homepage: <http://www.sal.thoku.ac.jp/~tsigeto/welcomej.html>

Blog: <http://b.tsigeto.info/school>

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、基本的に、授業においてまたは文学部 2F 教務係前の掲示板においておこなう。ただし、休講などで緊急を要する連絡は、田中の個人ブログ (School カテゴリの記事) に掲載することがある。 <http://www.sal.tohoku.ac.jp/~tsigeto/newsj.html>

第1講 推測統計の基礎 (10/2)

田中重人 (東北大学文学部准教授)

[テーマ] 推測統計の基礎

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

2 統計的推測のふたつの方法

- 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。
- 全世界から8人を無作為抽出して麺類の好みをきいたところ、全員が「うどんが好き」と答えた。
- 2010年 FIFA ワールドカップの際、ある水族館で、国旗を貼った餌箱のどちらからタコが餌を食べるかで勝敗を占った。全8回について、先に食べたほうのチームが勝利 <[http://ja.wikipedia.org/wiki/パウロ_\(タコ\)](http://ja.wikipedia.org/wiki/パウロ_(タコ))>

このような情報 (= 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

区間推定: 母比率はたぶん ○○ から ×× の範囲にある

統計的検定: 母比率が0.5だと考えてよいか?

統計的検定のほうが計算が簡単であるため、よくつかわれている。区間推定を論文等で目にする機会はあまりないが、きちんと理解するにはまず区間推定の考え方をおさえるのがよい。

3 母比率の区間推定

3.1 区間推定の原理

- (1) 「信頼率」を決めておく (たとえば95%)
- (2) データから統計量を計算する
- (3) 母集団分布についていろいろなケースを想定する。その想定のもとでの標本統計量の確率分布を計算し、95%の確率で出現する範囲を確定する。
- (4) この範囲のなかに、データから求めた統計量の値がふくまれるかを調べる
- (5) (4) の条件を満たす想定ケースのすべてについて統計量を求める
- (6) (5) で求めた値の集合が「95%信頼区間」である

3.2 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

上記の例題では、 $m=1$ であることがわかっているが、 M が不明である ($n=8$)。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。

課題 1: 解答を火曜正午までに ISTU 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 \wedge という演算子が使える (掛け算を 8 回繰り返してもよい)。

3.3 もっと複雑な例

全世界から 400 人を無作為抽出して麺類の好みを訊いたところ、「うどんが好き」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) におけるうどん好きの比率の 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。

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=1, 裏=0 であらわすと
0 0 0 0 ($x=0$)
0 0 0 1 ($x=1$)
0 0 1 0 ($x=1$)
0 0 1 1 ($x=2$)
.....
1 1 1 1 ($x=4$)

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: 解答を火曜正午までに ISTU 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