URL: http://www.nik.sal.tohoku.ac.jp/~tsigeto/statg/2.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/現代日本論演習

学部３年生以上・大学院生対象：2004 年度後期
＜木２＞コンピュータ実習室 (文学部本館 7F 711-2)
授業コード=LM24205,L64207

授業の概要

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 SPSS を使ってデータ分析の実習をおこないます。

履修要件

前期開講の 現代日本論演習 「統計分析の基礎」 / 比較現代日本論研究演習「統計分析入門」を履修済みであるか、それと同等の知識を習得済みであること。

テキスト

なし

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。

備考

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 (卒業論文／修士論文等で統計分析をおこなう予定の者を優先する)。

授業の予定

目次

※（）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

- 確率密度と理論分布
- 誤差の対策
- 標本誤差の推定
- 平均値の点推定・区間推定
- 平均値の差の区間推定と t 検定
- 連関係数の区間推定と $\chi 2$ 検定
- サンプル・サイズと検定力

2. 相関係数

- 尺度水準についての復習
- 散布図
- 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

5. 多変量解析

未定 (多元配置の分散分析?)

6. 期末レポート

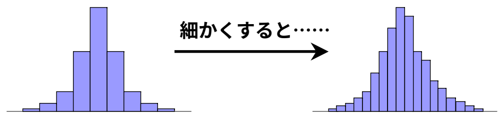
比較現代日本論研究演習/現代日本論演習（田中重人）

第 1 回「確率密度と理論分布」(2004.10.7)

1. ヒストグラムと確率密度
2. 確率の理論分布
3. 正規分布とそのファミリー
4. 確率密度関数の使いかた

【ヒストグラム】

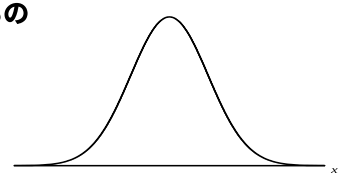
連続量を階級分けして度数分布を示したもの



【確率密度のグラフ】

Probability density

連続量に対応して、連続的に変化する確率を表したもの



【確率の理論分布】

特定の仮定から
理論的に導出された確率の分布

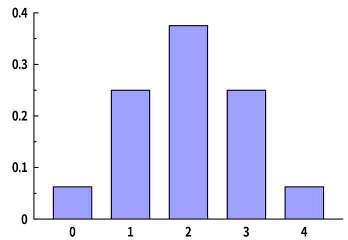
例：硬貨を投げるとき
　　表が出る →
　　裏が出る →

【2 項分布】

硬貨を n 回投げる。
表が出る回数を x とする。

$n = 4$ のとき、 x はどのような値を
どのような確率でとるか?

$n = 4$, 確率= 0.5 の 2 項分布



【期待値】

Expected value
値 (x) に確率 (p) を掛けたものの総和：

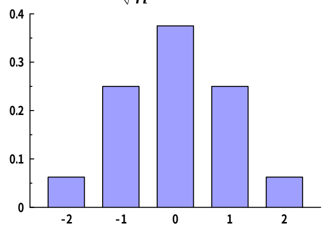
$$E = \Sigma\,(x \times p)$$

※「平均値」と呼ばれることもある

$n = 4$ の 2 項分布の期待値は?

【標準化】

$Z = (x - E)\dfrac{2}{\sqrt{n}}$ に変換すると



【標準正規分布】

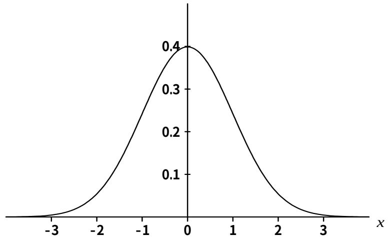
Standard normal distribution

n が大きければ、Z は
標準正規分布の確率密度関数

$$\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]$$

で近似できる

標準正規分布の確率密度のグラフ：



※ 標準正規分布に定数による加減乗除を加えたものを総称して「正規分布」(normal distribution) という

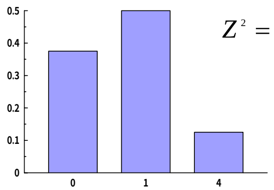
※ 0.5 以外の確率による 2 項分布でも、適当な標準化を行って n を増加させると正規分布に近づく

【正規分布の応用上の意義】

偶然による現象の生起確率や、
その組み合わせで決まる物事は、
正規分布（またはそのファミリー）
で近似できることが多い

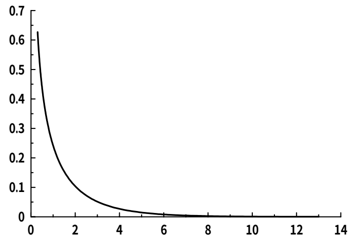
【自由度 1 の χ^2 分布】

先に標準化した変数の 2 乗を考える：



$$Z^2 = \left((x-E)\frac{2}{\sqrt{n}}\right)^2$$

n が増加すると、Z^2 の確率分布は
自由度 1 の χ^2 分布に近づく



【χ^2 分布の一般形】

硬貨を n 回投げる作業を c 回繰り返す。

それぞれについて表が出る回数 x_i を数え、それを標準化して 2 乗して総和を求める：

$$\chi_c^2 = \sum_{i=1}^{c}\left((x_i - E)\frac{2}{\sqrt{n}}\right)^2$$

n が大きければ、自由度 c の χ^2 分布に近似

【χ^2 分布の応用上の意義】

期待値や平均値からのずれを予測するときに使う

【t 分布】

(1) 硬貨を n 回投げる作業を 1 回おこない、表が出た回数を x とする

(2) 硬貨を n 回投げる作業を d 回繰り返し、それぞれについて表が出る回数 y_j (j=1...d)を数える

このとき

$$t_d = (x-E)\sqrt{\frac{d}{\sum_{j=1}^{d}(y_j - E)^2}}$$

の確率分布は、n が大きければ、
自由度 d の t 分布で近似できる。

d が大きければ、標準正規分布で近似できる。

【t 分布の応用上の意義】

平均値とそこからのずれの両方を
同時に予測するときに使う

【F 分布】

(1) 硬貨を n 回投げる作業を c 回おこない、それぞれについて表が出る回数 x_i (i=1 ... c)を数える

(2) 硬貨を n 回投げる作業を d 回繰り返し、それぞれについて表が出る回数 y_j (j=1 ... d)を数える

このとき

$$F_{(c,d)} = \frac{\sum_{i=1}^{c}(x_i - E)^2}{\sum_{j=1}^{d}(y_j - E)^2} \times \frac{d}{c}$$

の確率分布は、n が大きければ、
自由度(c, d) の F 分布で近似できる。

※$\sqrt{F_{(1,d)}} = t_d$である。
※また、d が大きければ、$F_{(c,d)}$ は χ^2_d に近似する。

【F 分布の応用上の意義】

平均値からのずれの大きさを
比較するときに使う

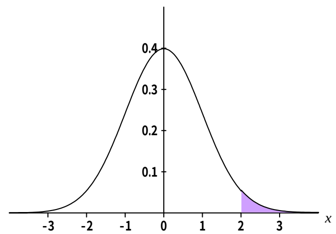
【確率密度関数の使いかた】

積分して、特定の区間内の値をとる確率

を求める。

例：標準正規分布において、2 以上の値をとる確率

$$\int_2^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$



実際には、

$$0.05 = \int_v^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$

のような方程式を（v について）解いて

臨界値 (critical value) を求めることが多い

【宿題】

別紙の乱数表から、1 桁の数字を 5 個と 50 個抜き出し、それぞれ平均値を求めよ。

乱数表の使いかた：
(1) 適当なところに目をつける
(2) 適当な方向（上・下・右・左・斜めなど）
　　へ移動しながらひとつずつ数字を拾う

【文献】

宮川公男（1999）『基本統計学［第 3 版］』有斐閣。

比較現代日本論研究演習/現代日本論演習　(田中重人)

2004.10.07 課題

学年：
氏名：

別紙の乱数表から、1 桁の数字を 5 個と 50 個抜き出し、それぞれ平均値を求めよ

5 個：

合計＝　　　　　　　　　　平均＝

50 個：

合計＝　　　　　　　　　　平均＝

比較現代日本論研究演習/現代日本論演習　(田中重人)

2004.10.07 課題

学年：
氏名：

別紙の乱数表から、1 桁の数字を 5 個と 50 個抜き出し、それぞれ平均値を求めよ

5 個：

合計＝　　　　　　　　　　平均＝

50 個：

合計＝　　　　　　　　　　平均＝