

母集団と標本抽出, 平均値の推測

樋口さぶろお <http://hig3.net>

龍谷大学工学部数理情報学科

生活の中の統計技術 L06(2018-10-29 Mon)

最終更新: Time-stamp: "2018-11-05 Mon 14:31 JST hig"

今日の目標

- 母集団の平均値の区間推定, 検定ができる



L05-Q1

Quiz 解答:回帰係数と回帰直線

$$y + 4 = \frac{-25\sqrt{36}}{\sqrt{49}\sqrt{36}\sqrt{49}} \times (x - 9).$$

L05-Q2

Quiz 解答:回帰係数と回帰直線

- 期末 = 0.813 中間 + 14.458
- $0.813 \cdot 50 + 14.458$

L05-Q3

Quiz 解答:回帰係数と回帰直線

- ① 期末 = $0.03 \times \text{レポート} + 0.81 \times \text{中間} + 13.26$
- ② 期末 = $0.03 \times 40 + 0.81 \times 50 + 13.26$

ここまで来たよ

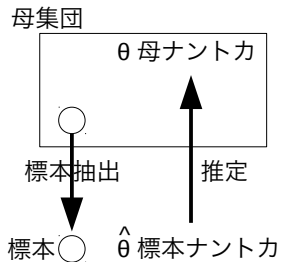
4 略解:重回帰分析

5 母集団と標本抽出, 平均値の推測

- 母集団と標本, 標本抽出と推定
- 平均値の区間推定
- 平均値の統計的仮説検定

母集団と標本, 標本抽出と推定

日本の大学生の〇〇の平均値, を調べるのはたいへん.
一部だけの情報から全体のことを言おう.



例

説明	母集団	サイズ	標本	N
全国の高校1年生の平均点を, いくつかの高校のデータから推定				
学校の生徒全員の平均点を, 何人かのデータから推定する				
候補Aに投票した人の比率を, 100人に1人ずつの出口調査から推定する				
工場で作られるドーナツの糖度を, 100個に1個の割合で壊して測定して推定する				

例 2

仮想的な母集団もある。

母集団=無限回の繰り返し. その平均値というのは, 「確率変数の平均値」

説明	母集団	サイズ	標本	N
ある選手のやり投げの記録は毎回異なる. 3回投げてデータを取り, 「本当の平均値」を推定する				
ある生徒の2桁の計算問題x10問の点数は毎回異なる. 3回受験してデータを取り, 「本当の平均点」を推定する				

例 3

両方の組み合わせ

説明	母集団	サイズ	標本	N
日本の小学3年生の2桁の計算問題 x10 問の平均点を知りたい。ランダムに選んだ小学校の児童に3回(1回)受験してもらってデータを取り、「本当の平均点」を推定する				

ここまで来たよ

4 略解:重回帰分析

5 母集団と標本抽出, 平均値の推測

- 母集団と標本, 標本抽出と推定
- 平均値の区間推定
- 平均値の統計的仮説検定

平均値の区間推定 I

(防ぎたいが) 抽出した標本は偏っているかも. どのくらいの範囲なの?

平均値の区間推定

母集団の平均値 μ の, 信頼係数 $1 - \alpha = 0.95, 0.99$ の信頼区間は,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{N}} < \mu < \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{N}}.$$

\bar{x} : 標本の平均値

s : 標本の標準偏差

N : 標本のサイズ

$$\text{係数 } z_{\alpha/2} = \begin{cases} 1.96 & (1 - \alpha = 0.95) \\ 2.58 & (1 - \alpha = 0.99) \end{cases}$$

信頼係数の意味: この方法で, 標本抽出と推定を何回も繰り返かえすとき, 確率 $1 - \alpha$ で区間内に正解がはいっている.

言い訳

- 本当は, $z_{\alpha/2}$ は他の α に対してもわかってるが, $1 - \alpha = 0.95, 0.99$ がよく使われる.
- 本当は, 母集団が正規分布というよくある分布にしたがっていないと正しくない.
- 本当は, $N - 1$ で割った不偏標本標準偏差というのを使うが, N が大きいときはこれでもほぼ同じ).
- 本当は, t -分布表の $t_{\alpha/2}(N - 1)$ というのを使うが, N が大きいときはほぼ同じ.

区間推定の公式はもっとも

区間の中心は

で, 両側に同じ幅だけ広がってる.

幅は

●

ほどせまい

●

ほどせまい

●

ほどせまい

幅がせまい ⇔

例題 I

L06-Q1

Quiz(平均値の区間推定)

ある学生が, (ほぼ同じ問題からなる) 計算問題のテストを毎週受験したところ, 次のような点数をとった,

80 79 81 83 77

この学生の「本来の」点数の平均値はどのくらい? 信頼係数 0.95 で範囲を示そう.

例題 I

L06-Q2

Quiz(区間推定)

瀬田学舎の龍大生の通学時間の分布を知るために, 無作為に 5 人を選んで質問したところ次のようだった. 通学時間の母平均値を, 信頼係数 0.99 の信頼区間で区間推定しよう.

10 分, 20 分, 30 分, 30 分, 110 分.

ここまで来たよ

4 略解:重回帰分析

5 母集団と標本抽出, 平均値の推測

- 母集団と標本, 標本抽出と推定
- 平均値の区間推定
- 平均値の統計的仮説検定

平均値の統計的仮説検定

- あるテストは, 授業を受ける前 (事前 pre-) は平均点が 100 点満点で 50 点であることがわかっているとしよう (まったく未知の知識で, ○×問題だから).
- ある方式の授業を受けた後 (事後 post-), 成績があがることは確か, と主張したい.
- 授業後に標本抽出, 推定して, 信頼区間が

ならば主張できる.

「平均値の片側の統計的仮説検定で, 平均値が 50 であるという帰無仮説が, 有意水準 $\alpha/2$ で棄却できる」という

- そうならないとき, 事後の平均値と 50 との大小は主張できない.

「平均値の片側検定で, 平均値が 50 であるという帰無仮説が, 有意水準 $\alpha/2$ で棄却できなかった」という.

例題 I

L06-Q3

Quiz(平均値の片側検定)

日本全国の高校 2 年生の平均点が正確に 70 点であることが (なぜか) わかっているテストがある.

ある高校の平均点がこれより高いことを主張すべく, 高校の生徒 5 人をランダムに選んで受験させたところ, 次の点数だった.

有意水準 0.025 で判定しよう.

74 78 80 82 86

例題 I

L06-Q4

Quiz(片側 t 検定)

ドーナツ製造マシン 1 号の作るドーナツに含まれる不純物 X の質量は正規分布にしたがう. その母平均値は 100mg 以下であることが求められている. 1 号で製造したドーナツの不純物の量を実際に測定したところ, 不純物は次の量だけ含まれていた.

104mg, 106mg, 106mg, 104mg.

不純物 X の質量の母平均値 μ が $\mu_0 = 100\text{mg}$ を越えていないかどうか, 有意水準 $\alpha = 0.05$ で, 片側 t 検定で判定しよう.

お知らせ

● 中間試験計画

- ▶ 30 ピーナッツ/科目 100 ピーナッツ
- ▶ 60 分?
- ▶ 2018-11-12 月 かどうか?
- ▶ 出題計画

60% 計算問題. データが与えられたときに, 平均値, q -分位数, 中間値, 四分位数, 分散, 標準偏差, 共分散, 相関係数, 単回帰の回帰直線, データ中の 1 個の数値の偏差値が求められる.

30% これらの量の性質や意味についての正誤判定問題

10% 上記にあてはまらないかもしれない問題 (ワイルドカード)

★ Excel の操作方法については出題しない

- ▶ 持込 紙はコピーを含め何でも. 電子機器は単機能電卓 (平方根まで) のみ.