

推定・検定の考え方

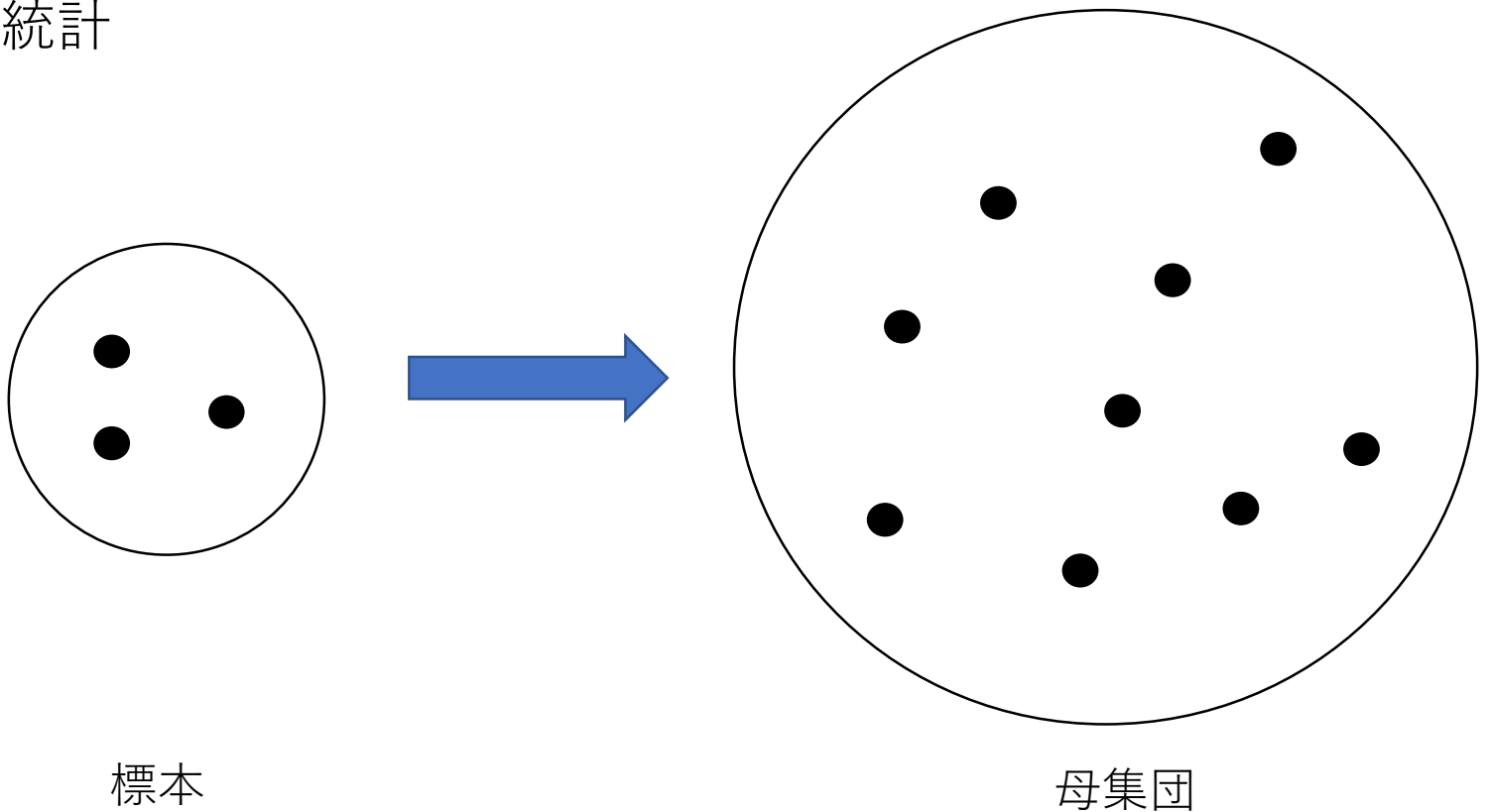
意味が分かる統計学 pp.88-125

60216064

加藤 寛瑛

推定

- 部分を見て, 全体の様子を知る方法
- 推測統計



- 使われる場面

→すべての要素からデータをとることが実質不可能である場合

(例)

選挙前の事前調査や当日の出口調査.

日本人全体の平均年収について統計したい. etc.

用語

- 母集団…特徴を調べたいデータ全体
- 標本…特徴を調べるために母集団からランダムに取り出したもの
- 母平均, 母分散, 母標準偏差
…母集団の平均, 分散, 標準偏差
- 標本平均, 標本分散, 標本標準偏差
…標本の平均, 分散, 標準偏差

- 点推定…

「母平均は50である」「母分散は47である」とピンポイントで予測すること

- 区間推定…

「母平均は45~55までの間にある」と幅を持たせて予測すること

点推定(最尤法による推定)

例題

500人がテストを受けました. この母集団の中から標本8人をとって点数を調べると次のようになりました. この時, 母平均, 母分散, を予想してください.

34 35 47 48 51 58 62 81

$$\bar{x} = \frac{34+35+47+48+51+58+62+81}{8} = 52$$

$$s^2 = \frac{(34-52)^2 + (35-52)^2 + \cdots + (81-52)^2}{8}$$

$$s^2 = 204$$

$$s \doteq 14.28$$

点推定(不偏推定量による推定)

不偏分散を U^2 とする

$$\bar{x} = 52$$

$$U^2 = \frac{(34 - 52)^2 + (35 - 52)^2 + \cdots + (81 - 52)^2}{8 - 1}$$

$$U^2 = 233.14$$

$$U = 15.27$$

区間推定

「過去のデータから、未来のデータを予測し、確率を割り出す」
(例) 打率, 降水確率

大数の法則により...

サンプルの変量 X の相対度数分布グラフは、母集団の相対度数分布グラフと等しくなる.

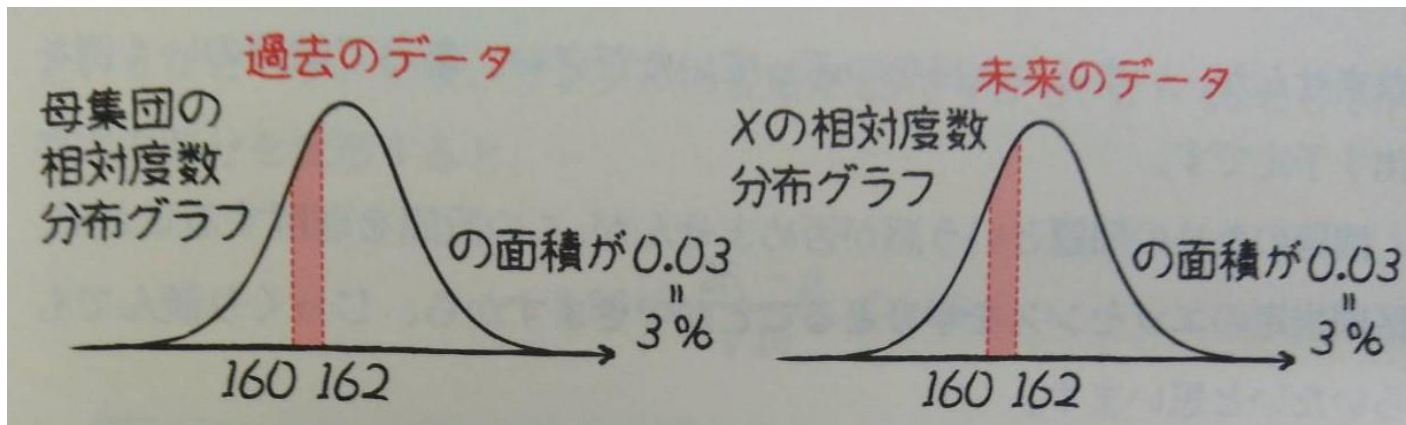


図1

区間推定

例題

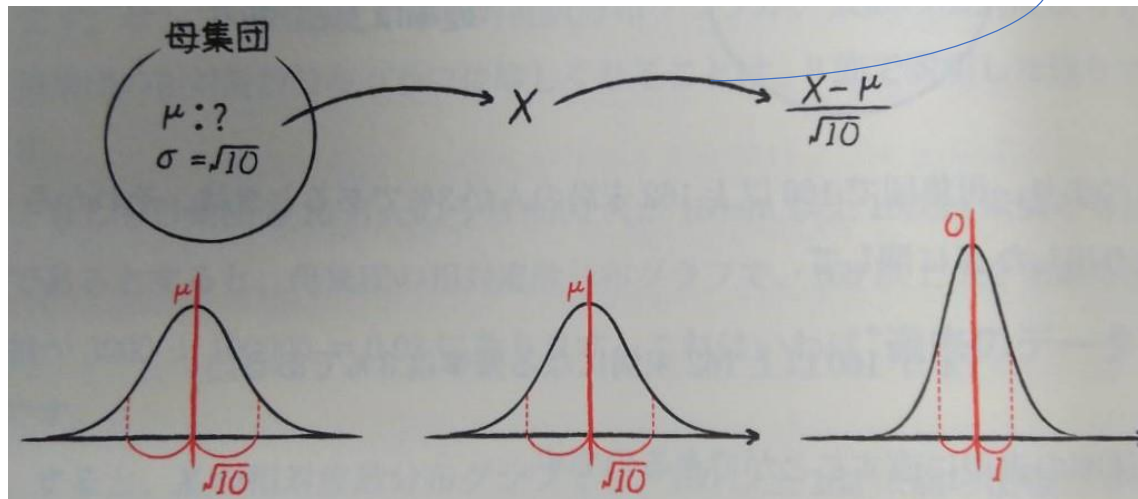
正規分布に従う母集団があり、平均 μ の値はわからないが、分散 σ^2 の値は10とわかっているものとする。この母集団から1個取り出した時、変量が25であったとする。このとき、信頼計数95%で μ の信頼区間を求めよ。

①母集団から取り出した標本の変量を $X_1, X_2, \dots, X_n$ として、 $\mu, \sigma, X_1, X_2, \dots, X_n$ から統計量 T を作る。

今回は X_1 のみなので $T = X_1$ となる。

区間推定

②統計量Tの従う分布を特定する.



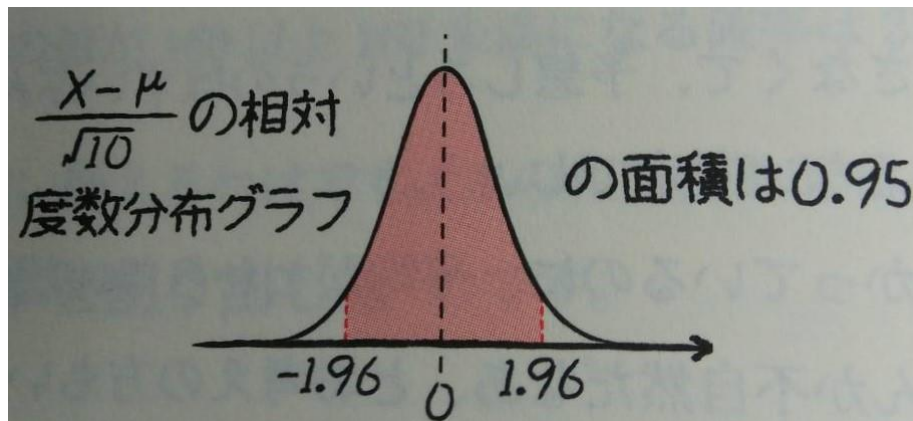
Xを標準化

標準正規分布
 $N(0, 1^2)$ に従う

-1.96以上1.96以下の区間の面積が0.95

$$-1.96 \leq \frac{X - \mu}{\sqrt{10}} \leq 1.96$$

となる確率が95%



区間推定

③面積が0.95になる区間を求めTについての不等式をたてる

$$-1.96 \leq \frac{X-\mu}{\sqrt{10}} \leq 1.96$$

$\sqrt{10}$ をかけて変量Xに25を
代入

$$-1.96 * \sqrt{10} \leq 25 - \mu \leq 1.96 * \sqrt{10}$$

$$25 - 1.96 * \sqrt{10} \leq \mu \leq 25 + 1.96 * \sqrt{10}$$

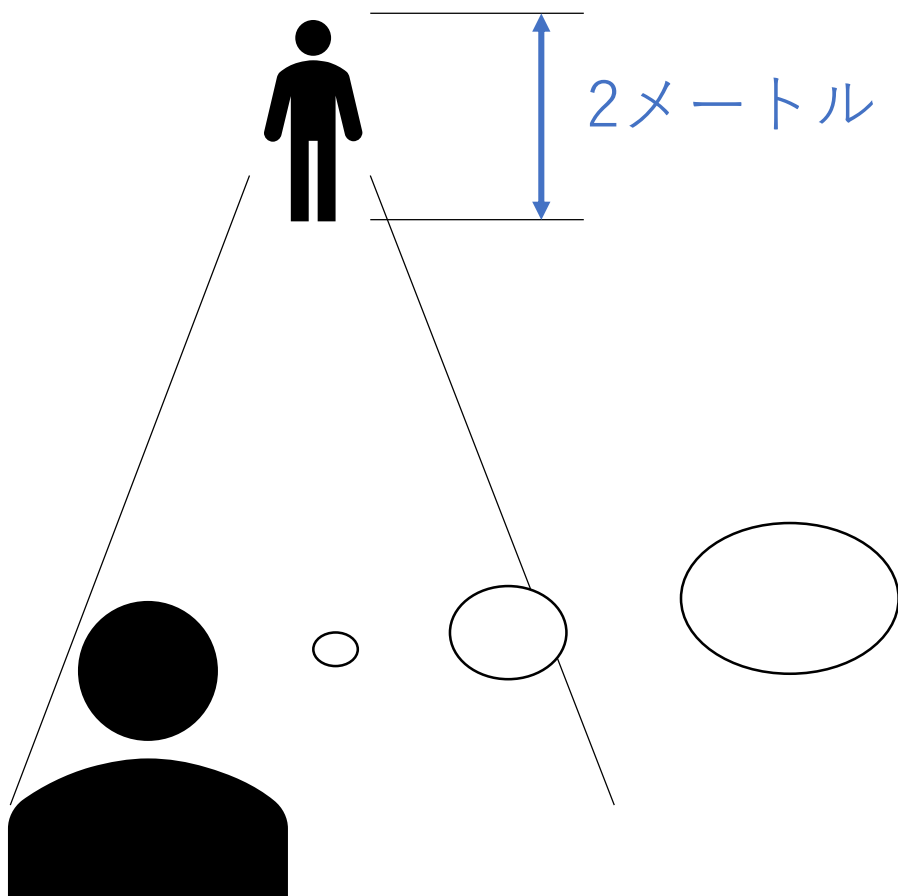
よって、 μ の信頼計数95%の信頼区間は、

$$25 - 1.96 * \sqrt{10} \leq \mu \leq 25 + 1.96 * \sqrt{10}$$

である

検定

- 部分から全体の特徴を把握する



日本人かな？

それにしても大きいな

外国人か

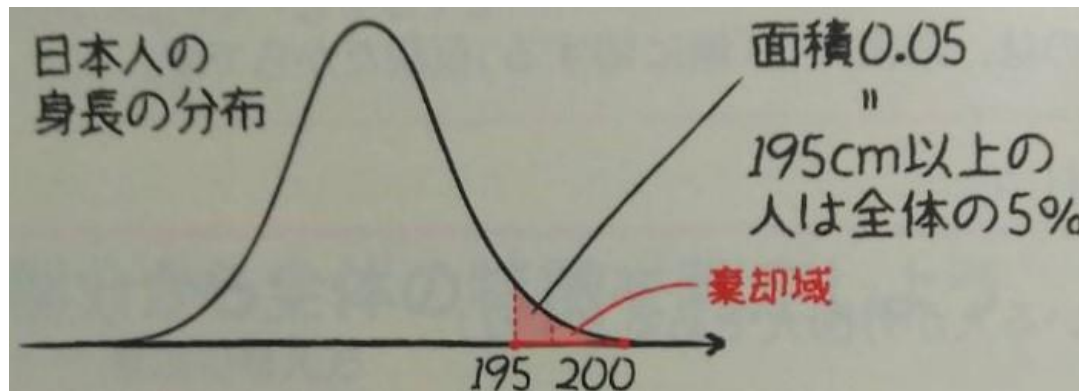
用語1

- 帰無仮説 H_0 …(最終的に)「無に帰す」仮説
歩いている人を日本人とする
- 対立仮説 H_1 …帰無仮説を否定した内容を持つ仮説
歩いている人を外国人とする

用語2

- ・ 有意水準…帰無仮説と対立仮説の判断基準

「身長が195cm以上の日本人は、5%しかいない」ならば、歩いてくる人が日本人である確率は5%以下である。この5%を有意水準という。



- ・ 棄却域…帰無仮説 H_0 を棄却する範囲のこと

用語3

- 第1種の誤り…帰無仮説 H_0 を棄却して，対立仮説 H_1 を採択したときに， H_0 が起きた.
- 第2種の誤り…帰無仮説 H_0 を採択して，対立仮説 H_1 を棄却したときに， H_1 が起きた.

事実 \ 判断結果	判断結果	
	帰無仮説 H_0 を採択	帰無仮説 H_0 を棄却
帰無仮説 H_0 が本当	正しい判断	第 1 種の誤り
帰無仮説 H_0 が嘘	第 2 種の誤り	正しい判断

例題

1975年の成人男子の平均身長は164cm, 標準偏差は6cmでした. 現在, 当時成人男子だった9人に協力してもらって身長の平均を計算すると, 170cmでした. 1975年に比べて, 男子の身長は伸びたと言えるでしょうか.

※1975年当時の標準偏差と現在の標準偏差は等しいものとする

(解)

1975年の平均身長を $\mu_0 = 164$, 標準偏差を $\sigma_0 = 6$ と置く

現在の平均身長を μ と置く

帰無仮説 $H_0 : \mu = \mu_0$

対立仮説 $H_1 : \mu > \mu_0$

有意水準は5%を用いる

例題

帰無仮説 H_0 のもとで、与えられた9人のデータの平均が170cmとなることが5%以下でしか起こらないようなことであるのか否かを分布から調べてみる

$\mu = \mu_0$ であるとすれば現在の成人男子の身長相対度数分布は、正規分布 $N(\mu_0, \sigma_0^2)$ に従う

標本となる9人の身長を $X_0, X_1, \dots, X_9$ とすると平均身長 $\bar{X}$ は、

$$\bar{X} = \frac{X_0 + X_1 + \dots + X_9}{9}$$

と表される.

この平均は μ_0 、標準偏差は $\frac{\sigma_0}{3}$ となる

つまり $\bar{X}$ は正規分布 $N(\mu_0, \frac{\sigma_0^2}{9})$ に従う.

これを標準化して、 $T = \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{3}}$ とおく.

Tを検定統計量といい、標準正規分布 $N(0, 1^2)$ に従う.

$$0.5 - 0.05 = 0.45$$

から，標準正規分布表を見て， p の値を求める．

表より， $p = 1.64$ となる．

ここで，

$$T = \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{3}}$$

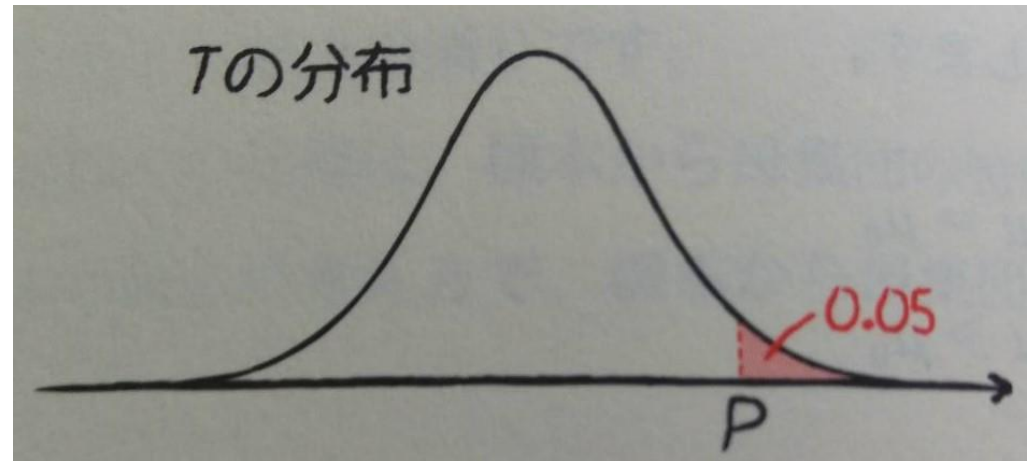
$$= \frac{170 - 164}{\frac{6}{3}}$$

$$= 3 > 1.64$$

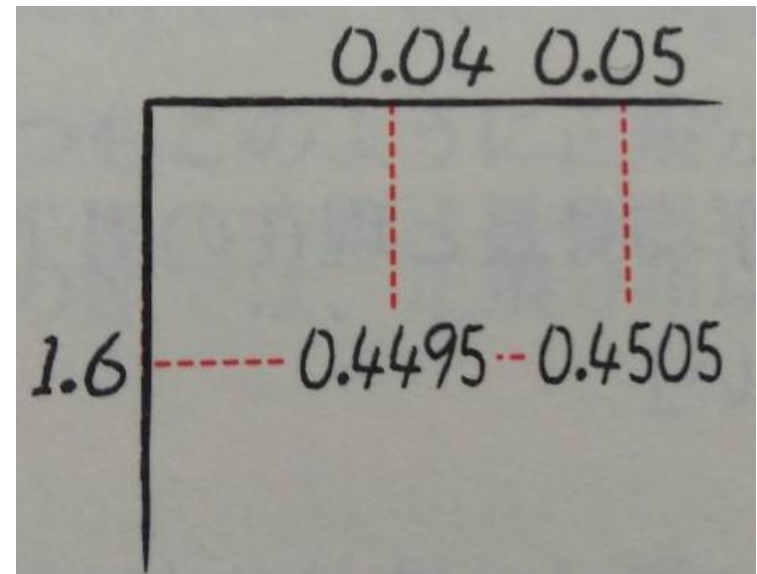
p の値より大きい



170cmは図の5%の棄却域のところに位置する



Tの相対度数分布グラフ



標準正規分布表

つまり，標準平均 $\bar{X}$ が $\bar{X} = 170$ となるのは，上側5%の棄却域にあるので，5%以下でしか起こりえないということ．



有意水準5%で

帰無仮説 $H_0 : \mu = \mu_0$ が棄却され

対立仮説 $H_1 : \mu > \mu_0$ が採択される



「1975年の男子の平均身長と現在の男子の平均身長は等しいとは言えない」ということが主張できる．

おしまい