平均の区間推定

点推定と区間推定

推定

- 点推定
  - １点のみで推定
  - 例：母平均μを標本平均で推定
- 区間推定
  - 幅を持たせて推定
  - 「平均μは確率0.95でAとBの間にある」などと言う

例：区間推定

- 生後4日目の新生児の血色素量(ヘモグロビン)を推定
  - 生後4日目の新生児10人を無作為に抽出し，血色素量を測定(g%)
    - データ 1

      25.0, 22.0, 28.4, 23.2, 20.8, 20.0, 16.4, 17.2, 22.4, 18.6
      - 標本平均：21.4，<u>標本分散：13.3</u>
    - データ 2

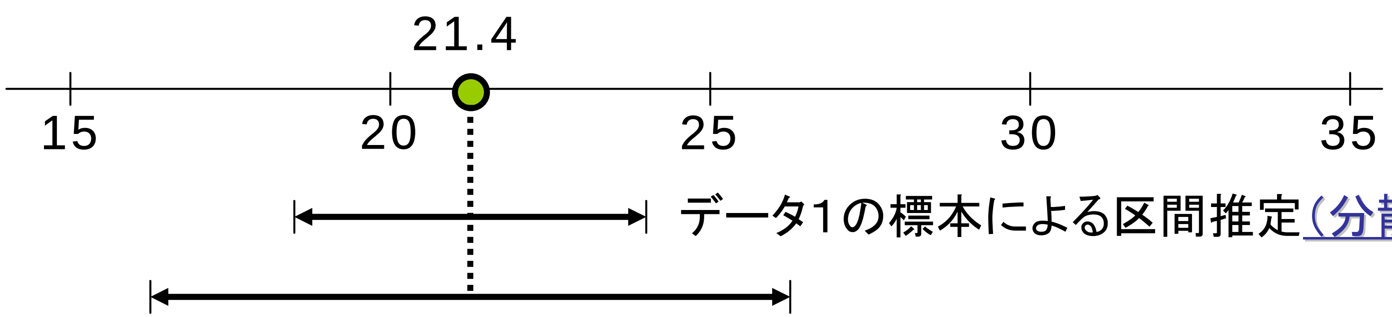
      29.0, 16.0, 33.4, 25.2, 16.8, 25.0, 13.4, 12.2, 26.4, 16.6
      - 標本平均：21.4，<u>標本分散：52.9</u>

データ1とデータ2では標本平均は等しいが，標本分散が異なる

区間推定のイメージ

区間推定のイメージ

どの程度の正確さで母数を推定しているかを
その区間が示している

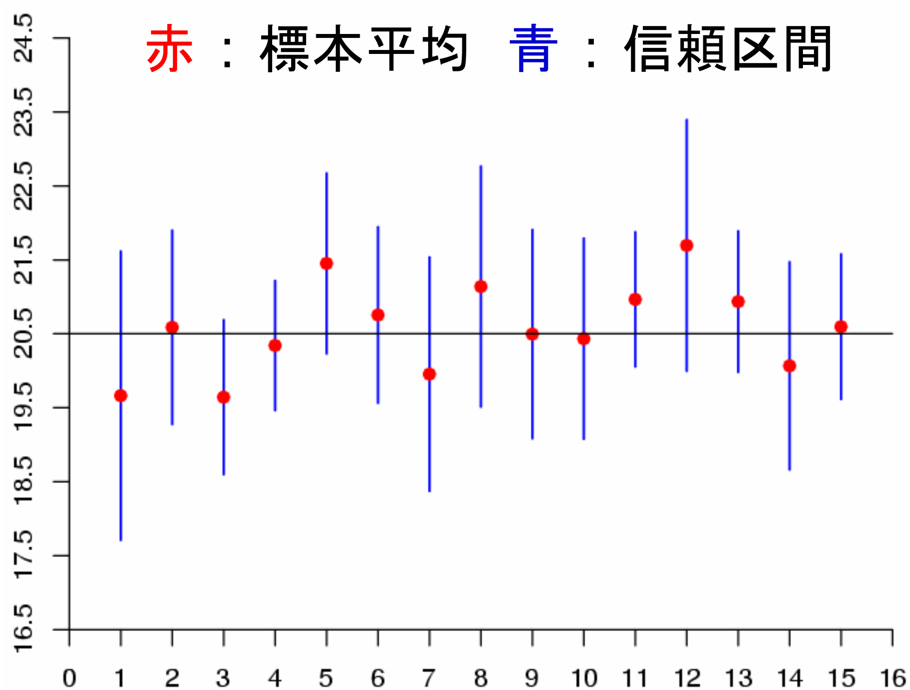


データ1の標本による区間推定(分散が小さい)

データ2の標本による区間推定(分散が大きい)

区間推定に関するシミュレーション

- 平均20.5，分散4の正規分布にしたがう乱数を15個生成し，その標本平均と95%信頼区間を求めた
- 繰り返し回数：15回



平均の区間推定

平均の区間推定

正規分布をしている母集団の平均 μ の区間推定

- 手順1：標本平均を求める

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \qquad \boxed{\bar{x} \sim N(\mu, \sigma^2/n)}$$

※標本平均 $\bar{x}$ が平均 μ, 分散 σ^2/n の正規分布にしたがう

- 手順2：変数の基準化を行う

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}, \qquad z \sim N(0, 1)$$

平均の区間推定

- 手順3： 標準正規分布の 95% 点から z の信頼区間を構成

$$-1.96 < \frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < 1.96$$

- 手順4：

手順３で得た信頼区間を μ について書き換える

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

μ の95%信頼区間

標本数と信頼区間

- 標本数が大きくなるにつれ、信頼区間の幅が狭くなる。すなわち、標本数が大きくなると、高い精度で μ を推定できる.

標準偏差を5とした場合の標本数と信頼区間の関係

標本数	5	25	100	500
信頼幅	8.77	3.92	2.00	0.88

$$※\ \text{信頼幅} = 2 \times 1.96 \boxed{\frac{\sigma}{\sqrt{n}}}$$

標本数が大きくなるにつれ，小さくなる

分散未知の場合の区間推定

これまでは、母分散が既知として信頼区間を求めた

- 一般には母分散は未知

母分散 σ^2 を s^2 で推定し，次の統計量を構成

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

統計量 t は自由度 $n-1$ の t 分布にしたがう

※正規分布ではない

分散未知の場合の区間推定

自由度 m の t 分布の両側 5%点を $t_{0.05}(m)$ とすると,
t 統計量の 95%信頼区間は

$$-t_{0.05}(m) < \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)} < t_{0.05}(m)$$

※ $m = n - 1$

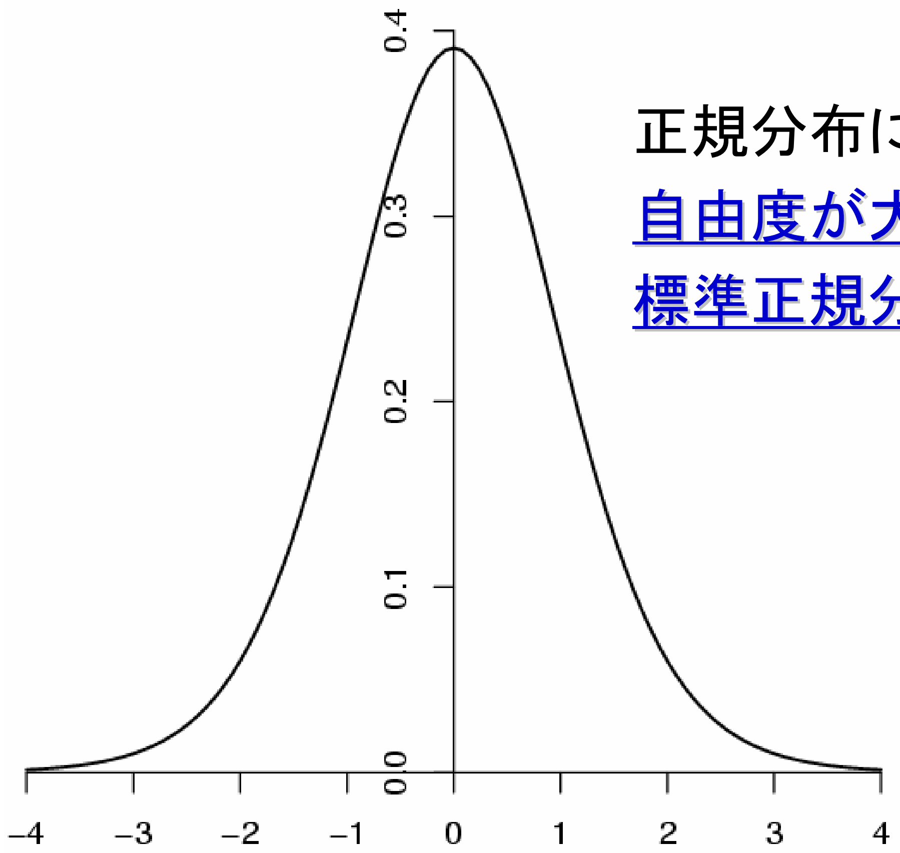
これを μ について解いて

$$\bar{x} - t_{0.05}(m)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}(m)\frac{s}{\sqrt{n}}$$

<u>分散が未知の場合の μ の95%信頼区間</u>

t 分布

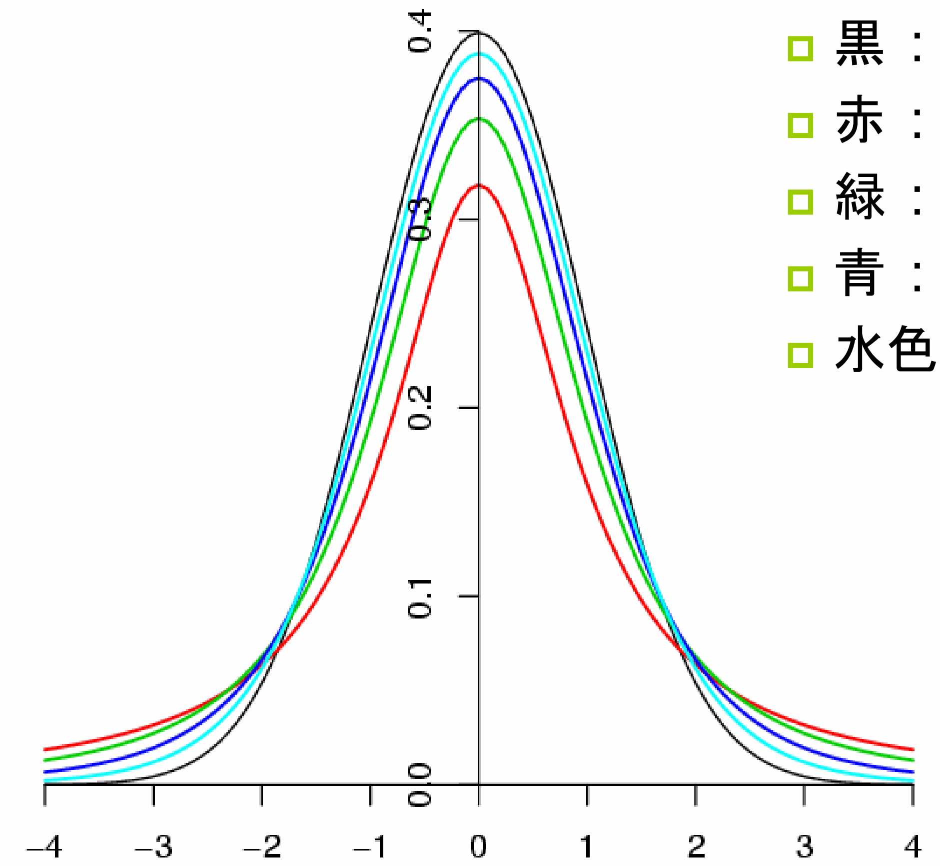
- 下の図は，自由度12のt 分布を描いたものである



正規分布によく似ており、
自由度が大きくなるにつれて
標準正規分布に近づいていく

t 分布

- 自由度1，2，4，9のt 分布と標準正規分布



- 黒 ： 標準正規分布
- 赤 ： 自由度 1
- 緑 ： 自由度 2
- 青 ： 自由度 4
- 水色 ： 自由度 9

例：分散未知の場合の区間推定

- 目的
  - 花びらの母平均μの95%信頼区間を求める
- データ
  - ひあふぎあやめを20本採取し，花びらの長さを測定

    1.4, 1.5, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5

    1.5, 1.6, 1.4, 1.1, 1.2, 1.5, 1.3, 1.4, 1.7, 1.5 (cm)
- μの95%信頼区間

$$\bar{x} - t_{0.05}(m)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.05}(m)\frac{s}{\sqrt{n}}$$

信頼区間を求めるための統計量

- 標本平均

$$\bar{x} = \frac{1.4 + 1.5 + \cdots + 1.5}{20} = 1.44$$

- 標本分散・標準偏差

$$s^2 = \frac{(1.4 - 1.44)^2 + (1.5 - 1.44)^2 + \cdots + (1.5 - 1.44)^2}{20 - 1}$$

$$= 0.0215$$

- *t* 統計量の両側5%点（付表や統計パッケージから得られる）

$$\boldsymbol{t_{0.05}(19) = 2.09}$$

区間推定

- 信頼区間を求める.

$$\bar{x} - \boldsymbol{t}_{0.05}(\mathbf{19})\frac{s}{\sqrt{n}} < \mu < \bar{x} + \boldsymbol{t}_{0.05}(\mathbf{19})\frac{s}{\sqrt{n}}$$

$$\bar{x} = 1.44, \quad s = 0.147, \quad \boldsymbol{t}_{0.05}(\mathbf{19}) = 2.09$$

より

$$1.44 - 2.09\frac{0.147}{\sqrt{20}} < \mu < 1.44 + 2.09\frac{0.147}{\sqrt{20}}$$

よって、μ の95%信頼区間は

$$\boxed{1.37 < \mu < 1.51}$$

二標本の平均の差の区間推定

二標本問題

- $\mu_1 - \mu_2$ の信頼区間について考える

母集団１
正規分布
平均：μ_1
分散：σ_1^2

標本
$x_1, x_2, \ldots, x_n$

母集団２
正規分布
平均：μ_2
分散：σ_2^2

標本
$y_1, y_2, \ldots, y_m$

二標本問題

問題： $\mu_1 - \mu_2$ の区間推定を行う

- ここでは、2群の分散は等しいとする
- 次の統計量が自由度 $f = n + m - 2$ の t 分布にしたがう

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- 共通の分散 s^2 は次式で推定する

$$s^2 = \frac{\{(x_1 - \bar{x})^2 + \cdots (x_n - \bar{x})^2\} + \{(y_1 - \bar{y})^2 + \cdots (y_m - \bar{y})^2\}}{(n-1) + (m-1)}$$

信頼区間の構成

- 自由度 $f = n + m - 2$ の t 分布の両側 5 %点：$t_{0.05}(f)$
- $\mu_1 - \mu_2$ の信頼区間

$$-t_{0.05}(f) < \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} < t_{0.05}(f)$$

これを $\mu_1 - \mu_2$ について解いて，

$$(\bar{x} - \bar{y}) - t_{0.05}(f)\left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right) < \mu_1 - \mu_2 < (\bar{x} - \bar{y}) + t_{0.05}(f)\left(s\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

$\boxed{\mu_1 - \mu_2 \text{ の } 95\% \text{信頼区間}}$

例：二標本の平均の差の区間推定

目的： 男子と女子の血色素量の平均の差 $\mu_1 - \mu_2$ の95%信頼区間を求める

- 男子10名，女子10名の血色素量の測定値
  - 男子 x
    - 16.4, 15.5, 15.8, 16.1, 15.1, 16.9, 15.6, 14.8, 16.5, 16.3
  - 女子 y
    - 16.4, 15.5, 15.8, 16.1, 15.1, 16.9, 15.6, 14.8, 16.5, 16.3
- 標本平均

$$\bar{x} = 15.9, \quad \bar{y} = 14.2$$

信頼区間の計算

- 共通の分散・標準偏差

$$s^2 = \frac{1}{(10-1)+(10-1)}[\{(16.4-15.9)^2+\cdots+(16.3-15.9)^2\}$$
$$+\{(14.9-14.2)^2+\cdots+(14.0-14.2)^2\}]$$
$$= 0.414$$

- 信頼区間

$$(\bar{x}-\bar{y}) - \boldsymbol{t}_{0.05}(\boldsymbol{f})\left(s\sqrt{\frac{1}{n}+\frac{1}{m}+}\right) < \mu_1-\mu_2 < (\bar{x}-\bar{y}) + \boldsymbol{t}_{0.05}(\boldsymbol{f})\left(s\sqrt{\frac{1}{n}+\frac{1}{m}+}\right)$$

$$n = 10,\quad m = 10,\quad \bar{x} = 15.9,\quad \bar{y} = 14.2,\quad s = 0.644$$

より、男子と女子の血色素量の平均の差の95％信頼区間は

$$\boxed{1.1 < \mu_1-\mu_2 < 2.3}$$