

次のように作成した18個のデータに対して、母分散が未知の場合の母平均の99%信頼区間を求めよ。

```
> set.seed(学籍番号)
> x <- round(rnorm(18,35,4.3),digits=1)
```

解答例：学籍番号が1825200の場合

データから標本平均・標本分散を求めると

```
> set.seed(1825200)
> x <- round(rnorm(18,35,4.3),digits=1)
> mean(x)
[1] 35.98889
> var(x)
[1] 32.32575
> sd(x)
[1] 5.685574      ← 標本標準偏差  $s$ 
> length(x)
[1] 18           ← データの数  $n$ 
```

母分散 σ^2 が未知なので、

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

で変換すると、 T は自由度 $17 (= 18 - 1)$ の t 分布に従うことがわかる。ここで、

$$\Pr\{-t_\alpha \leq T \leq t_\alpha\} = 0.99 = 1 - 0.01$$

となる t_α を R を使って求めると

```
> qt(1-0.01/2,18-1)
[1] 2.898231
```

であることが求められる。以上のことから、母平均 μ の95%信頼区間は

$$\begin{aligned} -2.898 &\leq \frac{35.99 - \mu}{\frac{5.686}{\sqrt{18}}} \leq 2.898 \\ 35.99 - 2.898 \times \frac{5.686}{\sqrt{18}} &\leq \mu \leq 35.99 + 2.898 \times \frac{5.686}{\sqrt{18}} \\ 32.106 \cdots &\leq \mu \leq 39.873 \cdots \\ 32.1 &\leq \mu \leq 39.9 \end{aligned}$$

であるから、 $[32.1, 39.9]$ である。R で計算する場合は

```
> mean(x) - qt(1-0.01/2,18-1)*sd(x)/sqrt(18)
```

のように求められる（上記は下限を求める式で、上限は $\text{mean}(x) + \dots$ とすればよい）。

ちなみに95%信頼区間は $t_\alpha = 2.109816$ を使って $[33.2, 38.8]$ である。

注) レポートもそうであるが、途中計算がある程度分かるように解答を書くこと!! (標本平均, 標本分散, ε_α の値と最初の不等式など)

学籍番号による解答一覧

学籍番号	標本平均	標本分散	下限	上限	下限(3桁)	上限(3桁)
1822007	33.86	15.29	31.190	36.532	31.2	36.5
1822008	36.63	25.30	33.191	40.064	33.2	40.1
1822015	36.52	24.05	33.166	39.867	33.2	39.9
1822024	34.95	27.74	31.352	38.548	31.4	38.5
1822031	36.18	11.94	33.818	38.538	33.8	38.5
1822050	34.18	18.53	31.237	37.119	31.2	37.1
1822065	35.58	12.84	33.130	38.025	33.1	38.0
1822070	35.23	21.00	32.097	38.358	32.1	38.4
1822074	35.14	17.58	32.281	38.008	32.3	38.0
1822078	35.26	18.34	32.335	38.187	32.3	38.2
1822094	38.29	13.88	35.744	40.834	35.7	40.8
1822096	34.14	23.97	30.800	37.489	30.8	37.5
1822097	34.65	11.86	32.297	37.003	32.3	37.0
1822100	34.56	13.65	32.037	37.085	32.0	37.1

・平均の区間推定（二標本の差の区間推定）

いままでは、一つの母集団に対して標本平均や標本分散の推定を行ってきたが、男女の平均の差などのように、二つの母集団の平均の差を調べたいこともある。ここでは、二つの母集団がそれぞれ母分散が等しい正規分布 $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ に従うとき、母平均の差 $\mu_1 - \mu_2$ の区間推定を行う。

まず、2つの母集団からそれぞれ

$$(x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_m)$$

と1つ目の母集団から n 個、2つ目の母集団から m 個の標本が無作為に得られたとする。このとき母平均の差 $\mu_1 - \mu_2$ の点推定は

$$\bar{x} - \bar{y} = \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{m} \sum_{j=1}^m y_j$$

のように、それぞれの標本平均のを計算した差で求めることができる。次にこの点推定の値を使って区間推定を求める。正規分布の和や差が正規分布に従うことから、 $\bar{x} - \bar{y}$ も正規分布に従うことになる。実際に $\bar{x} - \bar{y}$ は母平均 $\mu_1 - \mu_2$ 、母分散 $\frac{\sigma^2}{n} + \frac{\sigma^2}{m}$ の正規分布 $N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ に従うことが理論的にわかっている。よって、母分散が既知の場合と未知の場合で次のように区間推定を行うことができる。

・母分散が既知のとき

$\bar{x} - \bar{y}$ が正規分布 $N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ に従うので、

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

と変換すれば Z が標準正規分布に従うので、前回同様に次のような z_α を決め、

$$\Pr\{-z_\alpha \leq Z \leq z_\alpha\} = 1 - \alpha$$

括弧の中の不等式に実際の値 \bar{x}, \bar{y} を代入し、 $\mu_1 - \mu_2$ について解くことで

$$-z_\alpha \leq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq z_\alpha$$

$$-z_\alpha \times \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \leq (\bar{x} - \bar{y}) - (\mu_x - \mu_y) \leq z_\alpha \times \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$(\bar{x} - \bar{y}) - z_\alpha \times \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + z_\alpha \times \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

のような区間推定を得る。

・母分散が未知の場合

母分散 σ^2 の代わりに標本分散 s^2 を使うことを考える。しかし標本分散の計算は母集団ごとに s_x^2, s_y^2 と2つ計算できるので、等分散の条件をもとに

$$s^2 = \frac{1}{(n-1) + (m-1)} \left\{ (n-1)s_x^2 + (m-1)s_y^2 \right\}$$

と計算し、これを共通の母分散 σ^2 の推定量とする。これで、母分散 σ^2 のかわりに s^2 を使うことによって、

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

と変換すると t は自由度 $(n-1) + (m-1)$ の t 分布に従うことがわかっている。後は前回同様に次のような t_α を求め、

$$\Pr\{-t_\alpha \leq T \leq t_\alpha\} = 1 - \alpha$$

確率の括弧の中身に \bar{x}, \bar{y}, s^2 を代入し $\mu_1 - \mu_2$ について解くことで

$$(\bar{x} - \bar{y}) - t_\alpha \times s \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + t_\alpha \times s \sqrt{\frac{1}{n} + \frac{1}{m}}$$

のような区間推定を得る。

例. 二標本の差の区間推定

ある工場で作られた電球7個の寿命を調べたところ、下記のような結果を得た。

76.5, 82.4, 93.0, 78.7, 86.6, 94.2, 82.9 (単位: 日)

また、これとは別の工場で作られた電球6個の寿命を調べたところ下記のようにであった。

77.7, 69.5, 79.6, 68.2, 71.4, 67.4 (単位: 日)

このとき、母平均の差 $\mu_x - \mu_y$ の95%信頼区間を求めよ。

解答例)

標本平均と標本分散を計算すると、

```
> x <- c(76.5, 82.4, 93.0, 78.7, 86.6, 94.2, 82.9)
> y <- c(77.7, 69.5, 79.6, 68.2, 71.4, 67.4)
> cbind( mean(x), mean(y) )
      [,1] [,2]
[1,] 84.9 72.3
> cbind( var(x), var(y) )
      [,1] [,2]
[1,] 45.70667 26.384
> cbind( sd(x), sd(y) )
      [,1] [,2]
[1,] 6.76067 5.136536
```

となるので、標本平均や標本分散は次のようになる。

標本平均 $\bar{x} = 84.9$ 、標本分散 $s_x^2 = 45.7$ 、標本標準偏差 $s_x = 6.76$

標本平均 $\bar{y} = 72.3$ 、標本分散 $s_y^2 = 26.4$ 、標本標準偏差 $s_y = 5.14$

母分散が未知なので、共通の母分散の推定量 s^2 を求めると

$$\begin{aligned}s^2 &= \frac{1}{(7-1) + (6-1)} \{(7-1) \times 45.70667 + (6-1) \times 26.384\} \\ &= \frac{1}{11} (274.24 + 131.92) = 36.92364 \quad (s = 6.076483)\end{aligned}$$

となる。また、

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

と変換すると、 T は自由度 $(7-1) - (6-1) = 11$ の t 分布に従うことになる。

したがって、 $\Pr\{-t_\alpha \leq T \leq t_\alpha\} = 0.95 = 1 - 0.05$ となる t_α は

$$\text{qt}(1-0.05/2, 11) = 2.200985$$

となる。よって母平均の 95%信頼区間は

$$-2.200985 \leq \frac{(84.9 - 72.3) - (\mu_x - \mu_y)}{(6.076483) \times \sqrt{\frac{1}{7} + \frac{1}{6}}} \leq 2.200985$$

の不等式を解いて、

$$-2.200985 \times 6.076483 \times \sqrt{\frac{13}{42}} \leq 12.6 - (\mu_x - \mu_y) \leq 2.200985 \times 6.076483 \times \sqrt{\frac{13}{42}}$$

$$12.6 - 7.440745 \leq \mu_x - \mu_y \leq 12.6 + 7.440745$$

$$5.159255 \leq \mu_x - \mu_y \leq 20.04075$$

$$5.16 \leq \mu_x - \mu_y \leq 20.0$$

と求められる。

ちなみに下記のようにして求めることも可能

```
> s <- sqrt((6*var(x)+5*var(y))/(6+5))
> s
[1] 6.076482
> s^2
[1] 36.92364
> (84.9-72.3)-2.200985*6.076483*sqrt(1/7+1/6)
[1] 5.159256
> (84.9-72.3)+2.200985*6.076483*sqrt(1/7+1/6)
[1] 20.04074
```

もしくは

```
> (mean(x)-mean(y))-qt(1-0.05/2,6+5)*s*sqrt(1/7+1/6)
[1] 5.159256
> (mean(x)-mean(y))+qt(1-0.05/2,6+5)*s*sqrt(1/7+1/6)
[1] 20.04074
```

次のように作成した x が 9 個, y が 10 個のデータに対して、母分散の値が共通で未知と仮定したとき、母平均の差の 95% 信頼区間を求めよ。

```
> set.seed(学籍番号)
> x <- round(rnorm(9,15,3.5),digits=1)
> y <- round(rnorm(10,17,3.5),digits=1)
```

注) x の 1 番目のデータと y の 10 番目のデータは裏面で確認すること

2019年度神奈川工科大学 数理統計学 演習問題	学科	学年	組	学 籍 番 号	氏 名	

提出先：K3-3309号室前 19番のボックス 提出期限：11月21日（木）17時頃まで

学籍番号別：データの確認

学籍番号	$x[1]$	$y[10]$
1825200	16.0	16.2
1822007	7.1	10.9
1822008	18.4	13.2
1822015	16.4	15.8
1822024	15.6	12.5
1822031	21.3	16.6
1822050	9.5	17.0
1822065	14.1	14.9
1822070	14.5	19.0
1822074	14.0	26.0
1822078	15.6	17.5
1822094	19.6	21.3
1822096	17.0	15.4
1822097	17.0	15.8
1822100	12.2	14.6