数理統計学　前回の解答

１）身長の分布が $N(170.5, 7.6^2)$ に従うと仮定できる場合、自分の身長《○○○ cm（小数第１位を四捨五入)》以上になる確率 $P(○○○ \leqq X)$ を求めよ

```
> 1-pnorm(○○○,170.5,7.6)
```
で確率が出てくる。《> 1-pnorm((○○○-170.5)/7.6) でも可》

主な確率の値

身長	150	155	160	165	170	175	180	185	190
確率	0.997	0.979	0.916	0.765	0.526	0.277	0.106	0.0282	5.15×10^{-3}

２）母集団が母分散 $(2.95)^2$ の正規分布に従うとき、９つのデータ

32.4, 39.2, 41.3, 43.7, 39.7, 37.2, 40.9, 39.9, 39.4

が与えられたときの、標本平均を計算し、母平均 μ の 95%信頼区間を求めよ。

```
> x <- c(32.4, 39.2, 41.3, 43.7, 39.7, 37.2, 40.9, 39.9, 39.4)
> mean(x)
 [1] 39.3
> var(x)
 [1] 9.81
```

ここまでで標本平均 $\bar{x} = 39.3$、標本分散 $s^2 = 9.81$ $(s = 3.1320\cdots)$ がわかる。
95%信頼区間を求めるための ε は

```
> qnorm(1-0.05/2)        《> qnorm((1+0.95)/2) でも可》
 [1] 1.959964
```

によって求められるので、信頼区間は

```
> 39.3-1.959964*2.95/sqrt(9)
 [1] 37.3727
> 39.3+1.959964*2.95/sqrt(9)
 [1] 41.2273
```

のように $[37.4, 41.2]$ であることがわかる。

もし、99%信頼区間を求めるのであれば、

```
> mean(x)-qnorm(1-0.01/2)*2.95/sqrt(9)
 [1] 36.7671
> mean(x)+qnorm(1-0.01/2)*2.95/sqrt(9)
 [1] 41.8329
```

によって、$[36.8, 41.8]$ である。

注）１．解答は有効数字３桁以上で答えること。
　　２．データの個数が９なのに例題のまま sqrt(10) とする人がいた。
　　３．区間推定の答えは区間 $[a, b]$ または $a \leqq \mu \leqq b$ で答えること。
　　　　≪ (a, b) や $a < \mu < b$ でも可≫
　　４．母分散 $\sigma^2 = (3.25)^2$ なのに、例題のまま３を使う人が多かった。
　　５．母分散既知なのに、s の値を使う人がいた。
　　６．qnorm とまちがえて pnorm を使い $\varepsilon = 0.8352$ としないように注意 !!

演習で使うデータ

学籍番号	1 番目のデータ	18 番目のデータ	学籍番号	1 番目のデータ	18 番目のデータ
1523051	37.0	39.1	1823072	45.2	35.1
1723016	29.1	27.5	1823076	35.8	32.8
1823019	38.8	35.6	1823091	35.8	33.5
1823038	40.3	32.0	1823100	31.8	32.6
1823058	29.7	35.5	1823102	32.5	39.7
1823067	38.1	27.9	1823109	43.2	37.3
1823068	38.4	43.3	1823131	45.1	36.2
1823070	25.6	32.6			

・平均の区間推定（母分散未知）

基本的には母分散既知と同じ考え方であるが、母分散 σ^2 の値がわからない（未知）のため

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

での変換ができない。そこで Z の変換において、母分散 σ^2 の代わりにその推定量である標本分散 s^2 を使うことを考える。つまり

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

と変換する。このとき、T は標準正規分布ではなく、自由度 $n-1$ の t 分布に従うことがわかっている。よって $\Pr\{-z_\alpha \leqq Z \leqq z_\alpha\} = 1-\alpha$ ではなく、

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

になるように t_α を求めてやればよい。確率の括弧の中身は母分散既知と同様に

$$-t_\alpha \leqq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leqq t_\alpha \quad \Rightarrow \quad \bar{X} - t_\alpha \times \frac{S}{\sqrt{n}} \leqq \mu \leqq \bar{X} + t_\alpha \times \frac{S}{\sqrt{n}}$$

と変形できるので、$[\bar{x} - t_\alpha \times \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \times \frac{s}{\sqrt{n}}]$ の間に真の平均 μ が入っている確率は、事前に決めた確率にすることができる。

・$\Pr\{-t_\alpha \leqq T \leqq t_\alpha\} = 1-\alpha$ となる t_α の求め方

t 分布は教科書 P.54 にあるように、ほとんど標準正規分布と変わらない形をしている。母分散を推定しているために、誤差が大きくなる分だけ正規分布に比べて少しつぶれた形をしている。自由度 $(= n-1)$ が大きいほど正規分布に近づき、$n \to \infty$ のとき t 分布は正規分布となる。コンピュータが無い場合は、t 分布の両側 $100\alpha\%$ 点の表を使って t_α を求めることになるが、R では正規分布と同様に T が t 分布に従うとき、$\Pr\{T \leqq t_\beta\} = \beta$ となる t_β は `qt(`β`, 自由度)` で与えられる。データの個数が７の場合、自由度が $6(= 7-1)$ となるので９５％信頼区間を求めるには、$\Pr\{T \leqq t_\alpha\} = 1 - 0.05/2 = 0.975$ となる t_α を

```
> qt(1-0.05/2,6)
[1]  2.446912
```

のように求める。データの数が少ないので、標準正規分布の 1.96 に比べるとかなり大きな数字であることがわかる。データの数を増やしていくと

```
> qt(1-0.05/2,50)
[1]  2.008559
> qt(1-0.05/2,100)
[1]  1.983972
```

と正規分布の場合の値である 1.96 に近づいていく。

例．母平均が未知の場合の区間推定

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)

このとき、母平均 μ の 99%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)
> mean(x)
 [1] 84.9
> var(x)
 [1] 45.70667
> sd(x)
 [1] 6.76067
```

となるので、標本平均 $\bar{x} = 84.9$、標本分散 $s^2 = 45.70667$、標本標準偏差 $s = 6.76067$ であることがわかる。母分散が未知なので、

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

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

よって、$\Pr\{-t_\alpha \leqq T \leqq t_\alpha\} = 0.99 = 1 - 0.01$ となる t_α は

$$\Pr\{T \leqq t_\alpha\} = 1 - \frac{0.01}{2} = 0.995 \quad \Rightarrow \texttt{qt(1-0.01/2,6)} = 3.707428$$

となる。よって母平均の 99%信頼区間は

$$-3.707428 \leqq \frac{84.9 - \mu}{\frac{6.76067}{\sqrt{7}}} \leqq 3.707428$$

の不等式を解いて、

$$84.9 - 3.707428 \times \frac{6.76067}{\sqrt{7}} \leqq \mu \leqq 84.9 + 3.707428 \times \frac{6.76067}{\sqrt{7}}$$

$$75.42643 \leqq \mu \leqq 94.37357$$

$$75.4 \leqq \mu \leqq 94.4$$

と求められる。

中間レポートについて

次回１１月２２日（金）に問題配布、１２月　６（金）締め切りで中間レポートを行います。

2019年度 数理統計学 D科　小テスト解答用紙　2019.11.15

・母分散が未知の場合の母平均の区間推定

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(18,34,5.5),digits=1)
```

注）1番目のデータと18番目のデータは別紙に一覧表で載せておくので確認すること

<table>
<tr><td rowspan="2">2019年度神奈川工科大学
数理統計学
演習問題</td><td>学科</td><td>学年</td><td>組</td><td>学　籍　番　号</td><td>氏　　　名</td><td rowspan="2"></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

提出先：K3-3309号室前　20番のボックス　提出期限：１１月１８日（月）１７時頃まで