数理統計学　前回の解答

第１２回分

　Ｘ社，Ｙ社の同容量のバッテリーにモーターを繋ぎ、動作時間を調べた。下記の結果から両社のバッテリーに違いがあるかどうか有意水準１％で検定せよ（母分散は共通)。

Ｘ社	13.5	15.2	13.4	14.7	13.4	13.2
Ｙ社						

（単位：日）

なお、Ｙ社のデータは下記のプログラムで作成すること。

```
> set.seed(学籍番号)
> y <- round(rnorm(5,11.9,0.876),digits=1)
```

≪解答例（ 1625250 の場合）≫

（１）帰無仮説と対立仮説を立てる。

バッテリーに違いがあるかどうかを証明したいので

帰無仮説：バッテリーに違いがない　　H_0: $\mu_X = \mu_Y$

対立仮説：バッテリーに違いがある　　H_1: $\mu_X \neq \mu_Y$

（２）標本の抽出

```
> x <- c(13.5,15.2,13.4,14.7,13.4,13.2)
> set.seed(1825200)
> y <- round(rnorm(5,11.9,0.876),digits=1)
> x
 [1] 13.5 15.2 13.4 14.7 13.4 13.2
> y
 [1] 12.1 10.9 12.8 12.3 12.0
```

（３）帰無仮説が真の場合の（２）で得られた標本の出現確率を求める。

まず、標本平均と標本分散を計算すると

```
> c(mean(x),var(x),mean(y),var(y))
 [1] 13.900  0.696 12.020  0.487
```

であるから、$\bar{x} = 13.9$, $s_x^2 = 0.696$, $\bar{y} = 12.02$, $s_y^2 = 0.487$ である。この結果から、共通の母分散 σ^2 の推定値 s^2 および s の値を求めると次の通り

$$s^2 = \frac{1}{(6-1)+(5-1)}\{(6-1)\times 0.696 + (5-1)\times 0.487\} = 0.60311\cdots = 0.6031$$

$$s = \sqrt{0.6031} = 0.77659\cdots = 0.7766$$

さらに、帰無仮説 $\mu_X - \mu_Y = 0$ を用いて t の値を求めると、

$$t = \frac{(13.9-12.02)-(0)}{0.7766\sqrt{\frac{1}{6}+\frac{1}{5}}} = \frac{1.88}{0.4703} = 3.9974\cdots$$　（s, t の計算間違いが多数いたので注意）

である。t は自由度 $(6-1)+(5-1) = 9$ の t 分布に従うので、$\Pr\{X < 3.997\}$ の値は `pt(3.997,9)` より 0.9984377 である。

（４）判定

　この問題は有意水準が１％で対立仮説が「≠」の両側検定なので、（３）で求めた値が０．５％以下もしくは９９．５％以上のとき帰無仮説を棄却する。実際に（３）で求めた確率は９９．８％と棄却できる範囲にある。よって、帰無仮説を棄却し、対立仮説である「バッテリの動作時間が異なる」と言える。

ちなみに t.test を使う場合

```
> t.test(x,y,var.equal=T,alternative="two.side")

        Two Sample t-test

data:  x and y
t = 3.9978, df = 9, p-value = 0.003121
alternative hypothesis: true difference in means is not equal to 0
                    ⋮
```

となり、p 値 (p-value) が 0.003121 と有意水準の 1% 以下であるから帰無仮説を棄却する。

結果一覧

学籍番号	$\bar{y}$	s_y^2	s^2	t	確率	判定
1825200	12.02	0.487	0.6031	3.9978	0.99844	棄却する
1822007	11.52	0.922	0.7964	4.4042	0.99915	棄却する
1822008	12.20	0.905	0.7889	3.1609	0.99423	棄却できない
1822015	12.86	0.413	0.5702	2.2744	0.97550	棄却できない
1822024	11.32	0.947	0.8076	4.7413	0.99947	棄却する
1822031	12.54	0.603	0.6547	2.7758	0.98923	棄却できない
1822050	11.28	0.487	0.6031	5.5714	0.99983	棄却する
1822065	11.56	0.148	0.4524	5.7451	0.99986	棄却する
1822070	11.98	0.287	0.5142	4.4217	0.99917	棄却する
1822074	11.56	0.723	0.7080	4.5927	0.99935	棄却する
1822078	11.58	0.0920	0.4276	5.8594	0.99988	棄却する
1822094	12.50	0.925	0.7978	2.5885	0.98536	棄却できない
1822096	11.24	1.133	0.8902	4.6558	0.99940	棄却する
1822097	11.98	0.652	0.6764	3.8552	0.99806	棄却する
1822100	11.28	0.487	0.6031	5.5714	0.99983	棄却する

数理統計学

2020. 1. 8

・１変量から２変量（多変量）への拡張

今までは、身長なら身長だけ、体重なら体重だけのように、１つの変数に対して統計解析を行ってきた。しかしながら、身長と体重の間の関係を調べる場合、２つの変数を同時に扱う必要がある。そこで身長を x、体重を y で表し、(x, y) の組を１つの変量として扱うことを考える。n 人分のデータを得た場合、$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ となる。このとき、x_i と y_i は i 番目の人の（身長, 体重）の組であり、他人の身長・体重が組になることはない。

さらに（身長, 体重, 胸囲, 座高,⋯）のような多変数を組として１つの変量として扱うことも可能である。一般には p 変量を取り扱う場合、下記のようにベクトルと行列で表記するのが一般的であるが、ここでは２変量しか扱わないのでベクトル表記はしない。

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \text{平均ベクトル} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad \text{分散共分散行列} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

・散布図について

x, y の関係を簡単に見る方法として、散布図を描くことがある。単純に xy 平面上に (x_i, y_i) の点を書くだけであるが、どのような関連性があるか簡単に見ることができる。R では次のようにして書くことが可能（各自で確認すること）。

```
> x <- c(100,118,110,114,106,106,116,94,98,102,124)
> y <- c(110,150,144,130,140,108,124,114,122,120,146)
> plot(x,y)
```

・相関係数とは

(x, y) の間に次のような関係があるとき

・x が増えると y も増える　　　　・x が増えると y が減る

その関係の強さをはかるものさしとして、相関係数 ρ（ギリシャ文字：ロー）が用いられる。

一般に n 個の標本 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ が得られたときの標本相関係数 r は

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)}{s_x \times s_y}$$

のように、x, y の標本平均・標本分散と積和で表される。$-1 \leqq r \leqq 1$ の値を取り、絶対値が１に近いほど直線的な関係が見えるほか、

・r が正のとき：x が増えると y も増える。　散布図が右肩上がり
・r が負のとき：x が増えると y が減る。　　散布図が右肩下がり

のような関係性がわかる。注意する点は相関係数は直線的な関係しかはかることができないので、２次関数上や円の上にある点のように、曲線的な関連性があっても相関係数の値は低くなる。R で計算する場合、`cor(x,y)` で求められる（散布図の例のとき $r = 0.7060575$）。

2019年度 数理統計学 N科　小テスト解答用紙　2020. 1. 8

・相関係数

次のデータの散布図を描き、相関係数を求めよ。

番号	1	2	3	4	5	6	7	8
x	3	4	5	6	7	8	9	10
y	13.7	22.2	19.5	27.3	29.5	35.1	47.5	55.6

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：　1月　9日（木）１７時頃まで