確率統計 前回の解答

2019.12.17 分

２標本の平均の差の検定

消費電力が低くなるように電球に改良を加えた結果、改良前後の消費電力は次の通りであった。

改良前 (x)	:	7.3	7.8	7.5	7.7	8.2	8.0	
改良後 (y)	:	7.5	6.9	6.4	7.2	7.1		（単位 W）

消費電力の分布は同じ母分散を持つ正規分布であるとするとき、次の問いに答えよ。

1) 標本平均 $(\bar{x}, \bar{y})$・標本分散 (s_x^2, s_y^2) を求めよ。

$\bar{x} = 7.75,\ s_x^2 = 0.107,\ \bar{y} = 7.02.\ s_y^2 = 0.167$

2) 共通の母分散の推定量 s^2 を求めよ。

$$s^2 = \frac{1}{5+4}(5 \times 0.107 + 4 \times 0.167) = \frac{1.203}{9} = 0.13366\cdots$$

3) 消費電力が低くなるといえるかどうか有意水準１％で検定せよ。
消費電力が低くなるかどうかなので、帰無仮説と対立仮説は下記のようになる。

帰無仮説：消費電力が高くなるか同じである　H_0: $\mu_x \geqq \mu_y$
対立仮説：消費電力が低くなる　H_1: $\mu_x < \mu_y$

実際の検定のときは H_0: $\mu_x - \mu_y = 0$ v.s. H_1: $\mu_x - \mu_y < 0$ として検定する。

帰無仮説（$\mu_x - \mu_y = 0$）が真のときの統計量の値を求めると下記のとおりである。

$$t = \frac{(7.75 - 7.02) - (0)}{\sqrt{0.1337} \times \sqrt{\frac{1}{6} + \frac{1}{5}}} = \frac{0.73}{0.2214} = 3.2971\cdots$$

（精密に計算した場合 $3.2974\cdots$）

t は自由度 $(6-1)+(5-1)=9$ の t 分布に従うことを使って判定を行う。

この問題は対立仮説が「＜」の片側検定なので、棄却域はグラフの片側に作る。t が自由度９の t 分布に従うことから、片側１％（＝両側２％）の点を表から求めると 2.8214 である。$|t| = 3.297 > 2.8214$ なので、帰無仮説を棄却する。

つまり、改良後の方が消費電力が下がっていると言える。

注）今回の問題は**「消費電力が低くなる」**なので、対立仮説は $\mu_x < \mu_y$ の片側検定である。
「消費電力に違いがあるかどうか」の検定の場合、両側検定になる。
何を調べているのか問題文から読み取って正しい検定を行うこと。

区間推定の場合の結果は両側１％の点を用いて

$$-3.2498 < \frac{(7.75 - 7.02) - (\mu_x - \mu_y)}{\sqrt{0.1337} \times \sqrt{\frac{1}{6} + \frac{1}{5}}} < 3.2498 \quad \Rightarrow \quad 0.0105 < \mu_x - \mu_y < 1.45$$

確率統計　相関係数と回帰直線　　第１３回

・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。

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

μ_i：i 番目の変数の母平均（$E[X]$）
$\sigma_{ii} = \sigma_i^2$：i 番目の変数の母分散（$E[(X-\mu)^2]$）
σ_{ij}：i 番目の変数と j 番目の変数と共分散（$E[(X_i - \mu_{x_i})(X_j - \mu_{x_j})]$）

・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

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

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。

多変量の場合にも同様に定義することが可能で、i 番目の変数と j 番目の変数の相関係数 ρ_{ij} と分散 σ_i^2, σ_j^2、共分散 σ_{ij} の関係は次のとおりになる。

$$\sigma_{ij} = \rho_{ij}\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}} = \rho_{ij}\,\sigma_i \sigma_j$$

相関係数を並べた行列（形式上 $\rho_{ii} = 1$）を相関行列と呼ぶ。

・相関係数の求め方

次のデータが与えられたときの、相関係数を求める。

番号	1	2	3	4	5
x	7	8	9	10	11
y	64.5	64.8	48.3	45.2	40.7

x と y のそれぞれの標本平均、標本分散を求めると

$\bar{x} = 9.00,\ \ s_x^2 = 2.50 \qquad \bar{y} = 52.7,\ \ s_y^2 = 126.315 = 126.3$

相関係数を求めるために $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$ の値を求めると。

《$\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ の場合》

番号	1	2	3	4	5	計
x	7	8	9	10	11	
y	64.5	64.8	48.3	45.2	40.7	
$x - \bar{x}$	-2	-1	0	1	2	0
$y - \bar{y}$	11.8	12.1	-4.4	-7.5	-12.0	0
積	-23.6	-12.1	0.0	-7.5	-24.0	-67.2

《$\displaystyle\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$ の場合》

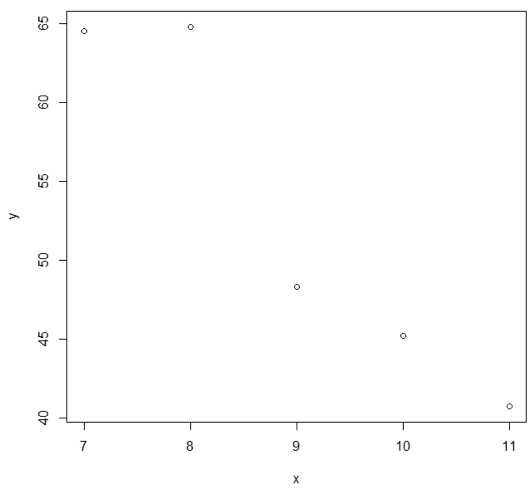
番号	1	2	3	4	5	計
x	7	8	9	10	11	
y	64.5	64.8	48.3	45.2	40.7	
積	451.5	518.4	434.7	452.0	447.7	2304.3

$$
\begin{aligned}
&\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&\quad = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \\
&\quad = 2304.3 - 5 \times 9.00 \times 52.7 \\
&\quad = -67.2
\end{aligned}
$$

上記の結果を使って、相関係数を求めると

$$
\begin{aligned}
r &= \frac{-67.2}{(5-1)\sqrt{2.50}\sqrt{126.3}} \\
&= -0.94544\cdots = -0.945
\end{aligned}
$$

と求めることができる（四捨五入の場所によっては $-0.9455\cdots$ となるので -0.946 も正解）。散布図を書くと右図の通り強い負の相関があることが分かる。



注）途中計算では有効数字４桁以上使用してください。

2019年度 確率統計　小テスト解答用紙　2019.12.23

・相関係数

次のデータが与えられたときの、相関係数を求めよ。

番号	1	2	3	4	5
x	11	13	15	17	19
y	58.3	43.2	39.7	31.9	22.9

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

提出先：K3-3309号室前　17番のボックス　提出期限：1月6日（月）１７時まで【時間厳守】