確率統計 前回の解答

第１３回分

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

x と y のそれぞれの標本平均、標本分散を求めると

$$\bar{x} = 15.0, \quad s_x^2 = 10.0 \qquad \bar{y} = 39.2, \quad s_y^2 = 175.01$$

相関係数と回帰直線を求めるために $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$ の値を求めると。

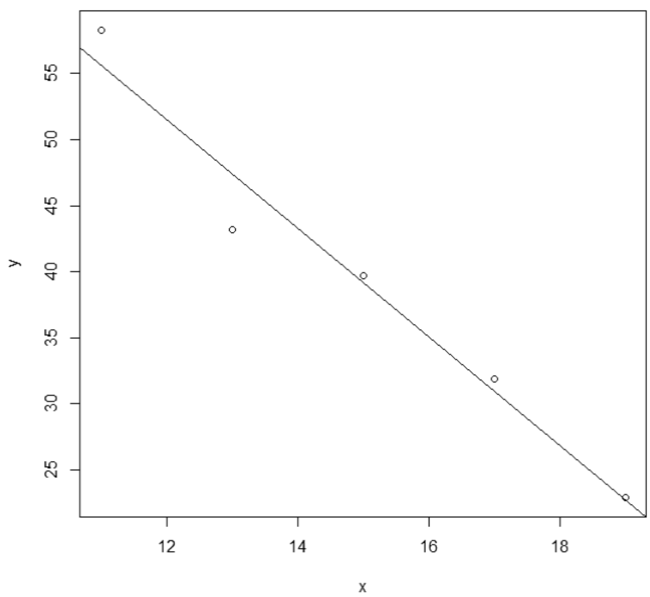
《$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ の場合》

番号	1	2	3	4	5	計
x	11	13	15	17	19	
y	58.3	43.2	39.7	31.9	22.9	
$x - \bar{x}$	-4	-2	0	2	4	0
$y - \bar{y}$	19.1	4	0.5	-7.3	-16.3	0
積	-76.4	-8	0	-14.6	-65.2	-164.2

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

番号	1	2	3	4	5	計
x	11	13	15	17	19	
y	58.3	43.2	39.7	31.9	22.9	
積	641.3	561.6	595.5	542.3	435.1	2775.8

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&\quad = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \\
&\quad = 2775.8 - 5 \times 15.0 \times 39.2 = -164.2
\end{aligned}$$



上記の結果を使って、相関係数を求めると

$$\begin{aligned}
r &= \frac{-164.2}{(5-1) \times \sqrt{10.0} \times \sqrt{175.0}} \\
&= -0.98128\cdots = -0.981
\end{aligned}$$

と求めることができる。

今日講義する回帰直線については

$$\hat{b} = \frac{-164.2}{(5-1) \times 10.0} = -4.105$$

$$\hat{a} = 39.2 - (-4.105) \times 15 = 100.775$$

なので、回帰直線の式は

$$y = -4.11x + 101$$

であることが分かる。

・回帰直線について

　回帰分析はデータの予測に用いられる手法で、いくつかの説明変数 $x_1, x_2, \cdots, x_p$ から目的変数 y を求める（予測する）関係式 $y = f(x_1, \cdots, x_p)$ を推定する。簡単な例として、おもりの重さとばねばかりの伸びは正比例の関係があることが知られている。つまりおもりの重さを x g、ばねばかりの伸びを y cm としたとき、

$$y = a + bx$$

の関係がある（この直線を回帰直線と呼ぶ）。ばねばかりに関する情報が何も無い場合、何回かの実験によって得られた (x_i, y_i) の組で与えられる標本から a, b の値を推定する必要がある。測定にはある程度の誤差が考えられるので、すべての標本が $y = a + bx$ 上に現れるわけではない。そこで次のような誤差 ε_i を考慮した関係式

$$y_i = a + bx_i + \varepsilon_i$$

を考えて、誤差を最小にする a, b の値を求める。誤差の単純和は正の値と負の値が打ち消しあってしまうため、実際には誤差の 2 乗和を最小にする a, b の値を計算する。詳しい説明は省くが、関数の最大値を求める問題なので、関数を a, b で偏微分し、0 となる a, b を求めればよい。実際に n 個の標本が与えられたとき、$\sum \varepsilon_i^2$ を最小にする a, b の推定値 $\hat{a}, \hat{b}$ は次のようになる。

$$\begin{aligned}
\hat{a} &= \bar{y} - \hat{b}\bar{x} \\
\hat{b} &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2} \\
&= \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x^2}
\end{aligned}$$

つまり平均および分散と x, y の積和があれば求めることが可能である。ちなみに相関係数はこの回帰直線にどれだけ近いかを表す統計量と考えることも可能である。

例．次のデータの相関係数と回帰直線を求めよ。

番号	1	2	3	4	5	6
x	3	4	5	6	7	8
y	11.8	19.4	18.5	16.1	20.5	30.1

　x と y のそれぞれの標本平均、標本分散を求めると

　　$\bar{x} = 5.5,\ \ s_x^2 = 3.50 \qquad \bar{y} = 19.4,\ \ s_y^2 = 37.032$

相関係数と回帰直線を求めるためには上記以外に $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$ を求めればよい。

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

番号	1	2	3	4	5	6	計
x	3	4	5	6	7	8	33
y	11.8	19.4	18.5	16.1	20.5	30.1	116.4
$x-\bar{x}$	-2.5	-1.5	-0.5	0.5	1.5	2.5	0
$y-\bar{y}$	-7.6	0	-0.9	-3.3	1.1	10.7	0
積	19	0	0.45	-1.65	1.65	26.75	46.2

$$\left(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=46.2\right)$$

相関係数：$r=\dfrac{46.2}{(6-1)\sqrt{3.50}\sqrt{37.032}}=0.81161\cdots$

回帰直線：$b=\dfrac{46.2}{(6-1)\times 3.50}=2.64,\quad a=19.4-2.64\times 5.5=4.88$

より回帰直線は $y=2.64x+4.88$

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

番号	1	2	3	4	5	6	計
x	3	4	5	6	7	8	33
y	11.8	19.4	18.5	16.1	20.5	30.1	116.4
積	35.4	77.6	92.5	96.6	143.5	240.8	686.4

$$\left(\sum_{i=1}^{n}x_iy_i=686.4\right)$$

相関係数：$r=\dfrac{686.4-6\times 5.5\times 19.4}{(6-1)\sqrt{3.50}\sqrt{37.032}}=\dfrac{46.2}{(6-1)\sqrt{3.50}\sqrt{37.032}}=0.81161\cdots$

回帰直線：$b=\dfrac{686.4-6\times 5.5\times 19.4}{(6-1)\times 3.50}=2.64,\quad a=19.4-2.64\times 5.5=4.88$

より回帰直線は $y=2.64x+4.88$

関数電卓の種類によってはデータを入力するだけで相関係数及び回帰直線を求めることが可能なものもある（２変数統計機能)。テストの際は使い方を自分で調べて途中計算無しに関数電卓の結果を使ってもかまわない。

（ただし途中計算が無い場合、部分点が付かない〔つまり０点か満点か〕）

資料置場

https://www.gen.kanagawa-it.ac.jp/takeda/class/

2019年度 確率統計　小テスト解答用紙　2020. 1. 7

・相関係数と回帰直線

次のデータが与えられたときの、回帰直線と相関係数を求めよ。

番号	1	2	3	4	5
x	5	6	7	8	9
y	50.8	41.1	38.3	36.7	27.1

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　8日（水）１７時頃まで