

経済統計学 - 宿題 No.1

(解答例)

蛭川雅之

2025 年 10 月 24 日

第 1 部

問題 1-1

空白部分を埋めた度数分布表は次のようになる。

階 級	階級値	度 数	相対度数	累積度数
2000 ~ 2350	2175	2	0.0333	2
2350 ~ 2700	2525	8	0.1333	10
2700 ~ 3050	2875	16	0.2667	26
3050 ~ 3400	3225	18	0.3000	44
3400 ~ 3750	3575	7	0.1167	51
3750 ~ 4100	3925	4	0.0667	55
4100 ~ 4450	4275	3	0.0500	58
4450 ~ 4800	4625	2	0.0333	60

- (1) 階級値は階級の中央値であるから、 $(2350 + 2700) \div 2 = 2525$ である。
- (2) (1) と同様、 $(3400 + 3750) \div 2 = 3575$ である。
- (3) 2 番目の階級“2350 ~ 2700”の相対度数が 8、同階級までの累積度数が 10 であるから、求める相対度数は $10 - 8 = 2$ である。
- (4) この階級までの累積度数が 26、直前の階級“2350 ~ 2700”までの累積度数が 10 であるから、求める相対度数は $26 - 10 = 16$ である。
- (5) $18 \div 60 = 0.3000$ である。

(6) $4 \div 60 = 0.06666 \dots \approx 0.0667$ である。

(7) 最小の階級までの累積度数は同階級の度数に等しいため 2 である。

(8) 最大の階級までの累積度数はデータ数に等しいため 60 である。■

問題 1-2

降順に並べた n 個の順位データを $(X_1^b, Y_1^b), \dots, (X_n^b, Y_n^b)$ 、一方、昇順に並べた n 個の順位データを $(X_1^\sharp, Y_1^\sharp), \dots, (X_n^\sharp, Y_n^\sharp)$ と表記する。なお、 (X_i^b, Y_i^b) および $(X_i^\sharp, Y_i^\sharp)$ は生データ (X_i, Y_i) に付された順位である。ここで、降順に並べた順位データから計算したスピアマン順位相関係数は、標本平均 $\bar{X}^b = \frac{1}{n} \sum_{i=1}^n X_i^b$ および $\bar{Y}^b = \frac{1}{n} \sum_{i=1}^n Y_i^b$ を使って、

$$\rho_{XY}^b = \frac{\frac{1}{n} \sum_{i=1}^n (X_i^b - \bar{X}^b) (Y_i^b - \bar{Y}^b)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i^b - \bar{X}^b)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i^b - \bar{Y}^b)^2}}$$

と表現できる。同様に、昇順に並べた順位データから計算したスピアマン順位相関係数は、標本平均 $\bar{X}^\sharp = \frac{1}{n} \sum_{i=1}^n X_i^\sharp$ および $\bar{Y}^\sharp = \frac{1}{n} \sum_{i=1}^n Y_i^\sharp$ を使って、

$$\rho_{XY}^\sharp = \frac{\frac{1}{n} \sum_{i=1}^n (X_i^\sharp - \bar{X}^\sharp) (Y_i^\sharp - \bar{Y}^\sharp)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i^\sharp - \bar{X}^\sharp)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i^\sharp - \bar{Y}^\sharp)^2}}$$

となる。

ここで、同順位がないことから、 (X_i^b, Y_i^b) と $(X_i^\sharp, Y_i^\sharp)$ との間に

$$\begin{cases} X_i^\sharp = (n+1) - X_i^b \\ Y_i^\sharp = (n+1) - Y_i^b \end{cases}$$

が成り立つ。従って、

$$\bar{X}^\sharp = \frac{1}{n} \sum_{i=1}^n \{(n+1) - X_i^b\} = \frac{1}{n} \sum_{i=1}^n (n+1) - \frac{1}{n} \sum_{i=1}^n X_i^b = (n+1) - \bar{X}^b$$

であり、同様に、 $\bar{Y}^\sharp = (n+1) - \bar{Y}^b$ も得られる。従って、 $(X_i^\sharp, Y_i^\sharp)$ の標本平均周りの偏差は $X_i^\sharp - \bar{X}^\sharp = \{(n+1) - X_i^b\} - \{(n+1) - \bar{X}^b\} = -(X_i^b - \bar{X}^b)$ および $Y_i^\sharp - \bar{Y}^\sharp = -(Y_i^b - \bar{Y}^b)$

となり、最終的に

$$\begin{aligned}\rho_{XY}^{\#} &= \frac{\frac{1}{n} \sum_{i=1}^n \left\{ - \left(X_i^b - \overline{X^b} \right) \right\} \left\{ - \left(Y_i^b - \overline{Y^b} \right) \right\}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left\{ - \left(X_i^b - \overline{X^b} \right) \right\}^2} \sqrt{\frac{1}{n} \sum_{i=1}^n \left\{ - \left(Y_i^b - \overline{Y^b} \right) \right\}^2}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n \left(X_i^b - \overline{X^b} \right) \left(Y_i^b - \overline{Y^b} \right)}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(X_i^b - \overline{X^b} \right)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n \left(Y_i^b - \overline{Y^b} \right)^2}} \\ &= \rho_{XY}^b\end{aligned}$$

が得られる。■

問題 1-3

$(\mu_X, \mu_Y) = (E(X), E(Y))$ と表記する。このとき、

$$\begin{aligned}Cov(X, Y) &= E\{(X - \mu_X)(Y - \mu_Y)\} : [\text{共分散の定義を用いる}] \\ &= E(XY - \mu_X Y - X \mu_Y + \mu_X \mu_Y) : [\{\} \text{内を展開する}] \\ &= E(XY) - E(\mu_X Y) - E(X \mu_Y) + E(\mu_X \mu_Y) : [\text{和の期待値を期待値の和に書き換える}] \\ &= E(XY) - \mu_X E(Y) - E(X) \mu_Y + \mu_X \mu_Y : [(\mu_X, \mu_Y) \text{ は定数である}] \\ &= E(XY) - E(X) E(Y) - E(X) E(Y) + E(X) E(Y) : [\text{本来の表記に戻す}] \\ &= E(XY) - E(X) E(Y)\end{aligned}$$

となる。次に、 X と Y が無相関、即ち

$$Cov(X, Y) = E(XY) - E(X) E(Y) = 0$$

であることは

$$E(XY) = E(X) E(Y)$$

と同値である。ここで、 X と Y が独立であれば、すべての i, j に対し

$$p_{XY}(x_i, y_j) = p_X(x_i) p_Y(y_j)$$

が成り立ち、これを

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{XY}(x_i, y_j) = \sum_{i=1}^n \left\{ \sum_{j=1}^m x_i y_j p_{XY}(x_i, y_j) \right\}$$

に代入することにより、

$$\begin{aligned}
E(XY) &= \sum_{i=1}^n \left\{ \sum_{j=1}^m x_i y_j p_X(x_i) p_Y(y_j) \right\} \\
&= \sum_{i=1}^n x_i p_X(x_i) \underbrace{\left\{ \sum_{j=1}^m y_j p_Y(y_j) \right\}}_{=E(Y)} : [j \text{ に依存しない部分を } \{\} \text{ の外に出す}] \\
&= \underbrace{\left\{ \sum_{i=1}^n x_i p_X(x_i) \right\}}_{=E(X)} E(Y) : [E(Y) \text{ は定数である}] \\
&= E(X) E(Y)
\end{aligned}$$

が得られる。従って、題意は証明された。

第 2 部

Excel ファイル“hw1_econstat_2025.xlsx”も併せて確認すること。

問題 2-1

シート“Q2-1”を参照せよ。■

問題 2-2

シート“Q2-2”を参照せよ。■

問題 2-3

シート“Q2-3”を参照せよ。なお、

$$p_{X,Y}(1,0) = \frac{3}{20} \neq \frac{3}{25} = \frac{3}{10} \times \frac{2}{5} = p_X(1) p_Y(0)$$

などから、 X と Y は独立ではないことが確認できる。一方、 $Cov(X, Y) = Corr(X, Y) = 0$ であり、この数値例は X と Y は無相関であるが独立ではない一例であることがわかる。■

問題 2-4

保険金請求を行う企業数を X と表記する。仮定より、保険を掛けている企業が詐欺被害に遭う確率は $p = 1/1000$ 、また、詐欺被害に遭う企業数の平均は $\lambda = np = 2500 \times 1/1000 = 2.5$

である。これらを利用すると、保険金請求を行う企業数がちょうど x 社である確率は

$$\Pr(X = x) = \begin{cases} {}^{2500}C_x \left(\frac{1}{1000}\right)^x \left(\frac{999}{1000}\right)^{2500-x} & : [\text{二項分布}] \\ \frac{2.5^x e^{-2.5}}{x!} & : [\text{ポアソン分布}] \end{cases}$$

となる。計算結果はシート“Q2-4”を参照せよ。今回の二項分布のポアソン近似は講義で紹介した例以上にいいものがあることが確認できる。■

問題 2-5

K 氏の調理時間を X と表記し、 $X \sim N(22, 5^2)$ を利用する。また、 X を標準化して得られる確率変数を Z と表記する。具体的な計算については、シート“Q2-5”を参照せよ。

(1) 求める確率は

$$\begin{aligned} \Pr(20 \leq X \leq 25) &= \Pr\left(\frac{20-22}{5} \leq \frac{X-22}{5} \leq \frac{25-22}{5}\right) \\ &= \Pr(-0.4 \leq Z \leq 0.6) \\ &= \Pr(Z \leq 0.6) - \Pr(Z \leq -0.4) \\ &= \text{normsdist}(0.60) - \text{normsdist}(-0.40) \\ &\approx 0.726 - 0.345 \\ &= 0.381 \end{aligned}$$

である。

(2) これは $\Pr(X > x) = 0.1$ を満たす分位点 x を求める問題である。左辺を $1 - \Pr(X \leq x)$ と書き直すことができることに注意すると、

$$1 - \Pr(X \leq x) = 0.1 \Rightarrow \Pr(X \leq x) = 0.9$$

となり、この問題は $N(22, 5^2)$ の 90% 分位点を求める問題に帰着する。さらに、

$$0.9 = \Pr\left(\frac{X-22}{5} \leq \frac{x-22}{5}\right) = \Phi\left(\frac{x-22}{5}\right)$$

であり、 $\text{normsinv}(0.90) \approx 1.282$ を利用すると、

$$\frac{x-22}{5} = 1.282$$

となる。この方程式を解くことにより、 $x = 22 + 5 \times 1.282 \approx 28.4$ (分) が得られる。

(3) 求める確率は

$$\begin{aligned}\Pr(X > 30) &= \Pr\left(\frac{X - 22}{5} > \frac{30 - 22}{5}\right) \\ &= \Pr(Z > 1.6) \\ &= 1 - \Pr(Z \leq 1.6) \\ &= 1 - \text{normsdist}(1.60) \\ &\approx 1 - 0.945 \\ &= 0.055\end{aligned}$$

である。■