

§3 データ分析で注意すべき点 演習問題

📖 問題の難易度の目安【易】☆☆☆ 【基礎】★★☆ 【標準】★★★

1 (☆☆☆)(はずれ値)

同じ実験を6回繰り返して、次の測定データが得られたとする：

実験データ	2.3	2.5	1.8	3.0	2.4	0.6
-------	-----	-----	-----	-----	-----	-----

- (1) 平均と分散を求めよ.
- (2) 0.6だけが異常に低い値と思われる. このデータだけ取り除いた, 残り5個のデータの平均と分散を求めよ.
- (3) データ0.6の扱いについて考察せよ.

2 (★★☆)(1次式によるデータの変換①)

$\alpha, \beta, \gamma, \delta$ を定数とし, $\beta \neq 0, \delta \neq 0$ とする. 2次元データ $(x_1, y_1), \dots, (x_N, y_N)$ を

$$u_k := \frac{x_k - \alpha}{\beta}, \quad v_k := \frac{y_k - \gamma}{\delta}$$

により, データ $(u_1, v_1), \dots, (u_N, v_N)$ に変換する. 変換前後の共分散 s_{xy} と s_{uv} に対して, 等式

$$s_{xy} = \beta\delta s_{uv}$$

が成り立つことを示せ. また, $\beta, \delta > 0$ ならば, 変換前後の相関係数 r_{xy} と r_{uv} に関して, 等式

$$r_{xy} = r_{uv}$$

が成り立つことを示せ.

3 (★★☆)(1次式によるデータの変換②)

ある駅の不動産屋で8件の賃貸物件(1LDK)の駅からの徒歩時間(分)と1ヶ月の賃貸料(万)を調べたところ次の表のようになった：

徒歩時間	1	3	3	4	6	7	7	9
賃貸料	8	6	5	7	6	5	6	5

徒歩時間を変数 x , 賃貸料を変数 y とし, 次の問いに答えよ.

- (1) 散布図を描け.
- (2) 相関係数 r_{xy} を求めよ. また, 徒歩時間と賃貸料にはどの程度の相関があると言えるか.
- (3) 回帰直線を求め, 散布図に描け.

4 (★★☆)(第3ファクターの影響を無視した場合の相関係数)

3つ組データ $(x_1, y_1, z_1), \dots, (x_N, y_N, z_N)$ を考える. 以下 $(y_1, z_1), \dots, (y_N, z_N)$ の関係を調べる際に, 第3ファクター x_i の影響がなくなるようにするために,

$$\begin{cases} \hat{y}_i := \frac{s_{xy}}{s_x^2} x_i + \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} \\ \hat{z}_i := \frac{s_{xz}}{s_x^2} x_i + \bar{z} - \frac{s_{xz}}{s_x^2} \bar{x} \end{cases}$$

とにおいて, $(y_1 - \hat{y}_1, z_1 - \hat{z}_1), \dots, (y_N - \hat{y}_N, z_N - \hat{z}_N)$ の相関係数を調べよう. 以下簡単のため

$$y'_i := y_i - \hat{y}_i, \quad z'_i := z_i - \hat{z}_i, \quad i = 1, \dots, N$$

とおく.

- (1) $\bar{y}' = 0, \bar{z}' = 0$ を確かめよ. また, 共分散 $s_{y'z'}$ が

$$s_{y'z'} = s_{yz} - \frac{s_{xy}s_{xz}}{s_x^2}$$

で与えられることを示せ.

- (2) $s_{y'}^2, s_{z'}^2$ について

$$s_{y'}^2 = s_y^2 \left\{ 1 - \left(\frac{s_{xy}}{s_x s_y} \right)^2 \right\}, \quad s_{z'}^2 = s_z^2 \left\{ 1 - \left(\frac{s_{xz}}{s_x s_z} \right)^2 \right\}$$

で与えられることを示せ.

- (3) (1), (2) を用いて, 第3ファクター x を除いた新しいデータ (y', z') に関する相関係数 $r_{y'z' \setminus x} := \frac{s_{y'z'}}{s_{y'} s_{z'}}$ は

$$r_{y'z' \setminus x} = \frac{r_{yz} - r_{xy} r_{xz}}{\sqrt{1 - r_{xy}^2} \sqrt{1 - r_{xz}^2}}$$

で与えられることを示せ. $r_{y'z' \setminus x}$ を **偏相関係数** という. ここに, r_{yz} はデータ (y, z) に関する相関係数であり, r_{xy}, r_{xz} についても同様である.

5 (★★☆)(偏相関係数)

ある会社の社員の体重と年収について、データを取ってみたところ、それらの間には相関関係があった。3次元データ (x, y, z) を

$$(x, y, z) = (\text{年齢}, \text{体重}, \text{年収})$$

とする。ここで、体重 y と年収 z の相関係数 r_{yz} は $r_{yz} = 0.90$ で高い相関を持っていた。ところがAさんはこの相関に疑問を感じたため、年収・体重の両方に影響を及ぼしている第3のファクターとして年齢 x があるのではないかと思い、実際に調べてみたところ、

- ・年齢 x 体重 y の相関係数 $r_{xy} = 0.75$

- ・年齢 x 年収 z の相関係数 $r_{xz} = 0.80$

であった。このとき年齢 x の影響を取り除いた年収 y 、体重 z の偏相関係数を求めよ。