

§2 データリテラシー (2) 演習問題

📎 問題の難易度の目安【易】☆☆☆ 【基礎】★★☆ 【標準】★★★

1 (★★☆)(統計的観点からの考察)

ある大学の学生の成績表には、S(秀), A(優), B(良), C(可)の4種類と不合格Fのみが記載されている。Sは90点から100点, Aは80点から89点, Bは70点から79点, Cは60点から69点を意味する。この成績表から素点換算する必要のあったXさんは、S, A, B, Cをそれぞれ90点, 80点, 70点, 60点とし、対応する成績科目の個数を掛け合わせて、科目個数の合計で割ったものをその学生の成績の換算点とした。統計的観点から、この計算の問題点を指摘し、適切な総合評価の方法を示せ。ここではGPAにおける計算方法との比較などは考えなくて良い。

2 (★★☆)(重み付き分散)

データ x_1, \dots, x_n における重み付き平均を

$$\bar{x}_w := \sum_{i=1}^n w_i x_i, \quad \forall w_i \geq 0, \quad \sum_{i=1}^n w_i = 1$$

と定義するとき、通常の本分散の定義において、通常の本平均の代わりに重み付き平均を用いた分散を

$$s_w^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_w)^2$$

とする。このとき、通常の本分散 s_x^2 と比べてどちらの方が大きいか。

3 (★★★)(重み付き相加相乗平均の不等式)

正値データ x_1, \dots, x_n に対する重み付き相加相乗平均の不等式

$$\bar{x}_w = \sum_{i=1}^n w_i x_i \geq x_1^{w_1} \cdots x_n^{w_n}$$

を示せ。ここに、重み $(w_i)_{i=1}^n$ は $w_i \geq 0, \sum_{i=1}^n w_i = 1$ を満たすとする。

4 (★★☆)(回帰直線)

以下のような2次元データ (x, y) を得た：

x	6	8	10	12	14	16
y	6	11	20	29	27	33

以下の問いに答えよ。

- (1) x の平均 \bar{x} と本分散 s_x^2 を求めよ。

- (2) y の標準偏差 s_y を求めよ.
- (3) (x, y) の共分散 s_{xy} を求めよ.
- (4) y の x への回帰直線の式を $y = ax + b$ の形で表せ.

5 (★★☆)(スクリーニング検査 (PCR 検査))

スクリーニング検査 (PCR 検査) において、各数値の求め方は次である：

	病気に罹患している人	病気に罹患していない人	計
陽性	x	z	$x + z$
陰性	y	w	$y + w$
計	$x + y$	$z + w$	$x + y + z + w$

- 有病率 = $\frac{\text{病人}}{\text{母集団}} = \frac{x+y}{x+y+z+w}$
- 感度 = $\frac{\text{真に陽性}}{\text{病人}} = \frac{x}{x+y}$
- 特異度 = $\frac{\text{真に陰性}}{\text{非病人}} = \frac{w}{z+w}$
- 偽陰性率 = $1 - \text{感度}$
- 偽陽性率 = $1 - \text{特異度}$
- 陽性適中度 = $\frac{x}{x+z} = \text{検査陽性者中における病人の割合}$
- 陰性適中度 = $\frac{w}{y+w} = \text{検査陰性者中における非病人の割合}$

この定義から、有病率が低いほど陽性適中度も低く、逆に有病率が低いほど陽性適中度も低いことがわかる。以下、ある感染症 A, B に対して、感度 95%、特異度 95% のスクリーニング検査を実施したところ、以下の表を得た。ただし、感染症 A は有病率 10%、感染症 B は有病率 1% であるとする：

【表 1】

	感染症 A に罹患している人	感染症 A に罹患していない人	計
陽性	950	450	1400
陰性	50	8550	8600
計	1000	9000	10000

【表 2】

	感染症 B に罹患している人	感染症 B に罹患していない人	計
陽性	95	495	590
陰性	5	9405	9410
計	100	9900	10000

このとき、以下の問いに答えよ。

- (1) 【表 1】【表 2】をもとに、感染症 A, B に対する偽陽性率、陰性適中度、陽性適中度を求めよ。
- (2) (1) の結果から分かることを論ぜよ。