

中級統計学：復習テスト 2

学籍番号 _____ 氏名 _____

2023年9月29日

注意：すべての質問に解答しなければ提出とは認めない。正答に修正した上で、復習テスト1～8をまとめて左上でホチキス止めし、第1回中間試験実施日（10月23日の予定）に提出すること。

1. (a) 棒グラフとヒストグラム（柱状グラフ）の違いを説明しなさい。

(b) ヒストグラムと累積相対度数グラフの長所・短所を説明しなさい。

(c) (教科書 pp. 32-33 参照) データ (1, 1, 1, 1, 2, 3, 4, 5, 16, 20) の平均・中位数・最頻値を求めなさい。

2. (教科書 p. 38 参照) データを (x_1, \dots, x_n) とする.

(a) $y_i := a + bx_i$ と一次変換すると,

$$\mu_y = a + b\mu_x$$

$$\sigma_y^2 = b^2\sigma_x^2$$

となることを示しなさい. ただし μ_x, μ_y は平均, σ_x^2, σ_y^2 は分散を表す.

(b) 上の結果を利用して, $z_i := (x_i - \mu_x)/\sigma_x$ と標準化すると, 平均が 0, 分散が 1 となることを示しなさい. (ヒント: $z_i = -\mu_x/\sigma_x + (1/\sigma_x)x_i$ と書ける.)

解答例

1. (a) 棒グラフは横軸が分類を表し、柱の高さで（相対）度数を表す。ヒストグラムは横軸が数値を表し、柱の面積で（相対）度数を表す。
- (b) ヒストグラム
- 長所** 度数の大小が把握しやすい。
- 短所** 適切な階級の取り方が難しい。
- 累積相対度数グラフ
- 長所** 分位点を読み取るのに適しており、階級が細かいほど滑らかなグラフとなる。
- 短所** 度数の大小が把握しにくい。
- (c) 平均 5.4, 中位数 2.5, 最頻値 1
2. (a)

$$\begin{aligned}\mu_y &:= \frac{y_1 + \cdots + y_n}{n} \\ &= \frac{(a + bx_1) + \cdots + (a + bx_n)}{n} \\ &= \frac{(a + \cdots + a) + (bx_1 + \cdots + bx_n)}{n} \\ &= \frac{na + b(x_1 + \cdots + x_n)}{n} \\ &= a + b \frac{x_1 + \cdots + x_n}{n} \\ &= a + b\mu_x \\ \sigma_y^2 &:= \frac{(y_1 - \mu_y)^2 + \cdots + (y_n - \mu_y)^2}{n} \\ &= \frac{[(a + bx_1) - (a + b\mu_x)]^2 + \cdots + [(a + bx_n) - (a + b\mu_x)]^2}{n} \\ &= \frac{(bx_1 - b\mu_x)^2 + \cdots + (bx_n - b\mu_x)^2}{n} \\ &= \frac{[b(x_1 - \mu_x)]^2 + \cdots + [b(x_n - \mu_x)]^2}{n} \\ &= \frac{b^2(x_1 - \mu_x)^2 + \cdots + b^2(x_n - \mu_x)^2}{n} \\ &= b^2 \frac{(x_1 - \mu_x)^2 + \cdots + (x_n - \mu_x)^2}{n} \\ &= b^2 \sigma_x^2\end{aligned}$$

(b) $z_i := (x_i - \mu_x)/\sigma_x = -\mu_x/\sigma_x + (1/\sigma_x)x_i$ と書けるから, $a = -\mu_x/\sigma_x$, $b = 1/\sigma_x$ と置くと,

$$\begin{aligned}\mu_z &= a + b\mu_x \\ &= -\frac{\mu_x}{\sigma_x} + \frac{1}{\sigma_x}\mu_x \\ &= 0 \\ \sigma_y^2 &= b^2\sigma_x^2 \\ &= \left(\frac{1}{\sigma_x}\right)^2 \sigma_x^2 \\ &= 1\end{aligned}$$