

中級統計学／経済統計 II：復習テスト 26

学籍番号 _____ 氏名 _____

2021年7月20日（甲南）／7月26日（府大）

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 21～26 を（左上で）ホチキス止めし、定期試験実施日にまとめて提出すること。

1. 次の OLS 問題を考える。

$$\min_{a,b} \sum_{i=1}^n (y_i - a - bx_i)^2$$

and $a, b \in \mathbb{R}$

OLS 問題の解を (a^*, b^*) ，残差を $e_i := y_i - a^* - b^*x_i$ とする。また $((y_1, x_1), \dots, (y_n, x_n))$ の標本平均を (\bar{y}, \bar{x}) とする。以下の式を証明しなさい。

(a)

$$\sum_{i=1}^n e_i = 0$$
$$\sum_{i=1}^n x_i e_i = 0$$

(b)

$$\sum_{i=1}^n (x_i - \bar{x}) e_i = 0$$

(c)

$$\sum_{i=1}^n (y_i - \bar{y})^2 = b^{*2} \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n e_i^2$$

2. k 個の正規母集団 (群) $N(\mu_1, \sigma^2), \dots, N(\mu_k, \sigma^2)$ から独立に抽出した大きさ n_1, \dots, n_k の無作為標本を $(y_{1,1}, \dots, y_{1,n_1}), \dots, (y_{k,1}, \dots, y_{k,n_k})$ とする. $n := n_1 + \dots + n_k$ とする. 第 h 群の標本平均を \bar{y}_h , 全群の標本平均を \bar{y} とする.

(a) 全変動 S , 群間変動 S_b , 群内変動 S_w をそれぞれ定義しなさい.

(b) $S = S_b + S_w$ が成り立つことを示しなさい.

解答例

1. (a) 1 階の条件より

$$\begin{aligned}\sum_{i=1}^n (y_i - a^* - b^* x_i) &= 0 \\ \sum_{i=1}^n x_i (y_i - a^* - b^* x_i) &= 0\end{aligned}$$

すなわち

$$\begin{aligned}\sum_{i=1}^n e_i &= 0 \\ \sum_{i=1}^n x_i e_i &= 0\end{aligned}$$

(b)

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x}) e_i &= \sum_{i=1}^n x_i e_i - \bar{x} \sum_{i=1}^n e_i \\ &= 0\end{aligned}$$

(c) $y_i = a^* + b^* x_i + e_i$ より

$$\begin{aligned}\bar{y} &:= \frac{1}{n} \sum_{i=1}^n y_i \\ &= \frac{1}{n} \sum_{i=1}^n (a^* + b^* x_i + e_i) \\ &= a^* + b^* \bar{x}\end{aligned}$$

したがって

$$\begin{aligned}\sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [b^* (x_i - \bar{x}) + e_i]^2 \\ &= \sum_{i=1}^n [b^{*2} (x_i - \bar{x})^2 + 2b^* (x_i - \bar{x}) e_i + e_i^2] \\ &= b^{*2} \sum_{i=1}^n (x_i - \bar{x})^2 + 2b^* \sum_{i=1}^n (x_i - \bar{x}) e_i + \sum_{i=1}^n e_i^2\end{aligned}$$

第 2 項は 0.

2. (a)

$$\begin{aligned}S &:= \sum_{h=1}^k \sum_{i=1}^{n_h} (y_{h,i} - \bar{y})^2 \\ S_b &:= \sum_{h=1}^k \sum_{i=1}^{n_h} (\bar{y}_h - \bar{y})^2 \\ S_w &:= \sum_{h=1}^k \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)^2\end{aligned}$$

(b)

$$\begin{aligned} S &= \sum_{h=1}^k \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h + \bar{y}_h - \bar{y})^2 \\ &= \sum_{h=1}^k \sum_{i=1}^{n_h} [(y_{h,i} - \bar{y}_h)^2 + 2(y_{h,i} - \bar{y}_h)(\bar{y}_h - \bar{y}) + (\bar{y}_h - \bar{y})^2] \\ &= \sum_{h=1}^k \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)^2 + 2 \sum_{h=1}^k \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)(\bar{y}_h - \bar{y}) + \sum_{h=1}^k \sum_{i=1}^{n_h} (\bar{y}_h - \bar{y})^2 \\ &= S_w + 2 \sum_{h=1}^k (\bar{y}_h - \bar{y}) \sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h) + S_b \end{aligned}$$

第2項は0.