中級統計学：復習テスト 26

学籍番号____________________氏名____________________

2024 年 1 月 22 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 21〜26 を順に重ねて左上でホチキス止めし，定期試験実施日（1 月 26 日の予定）に提出すること．

1. k 個の正規母集団（群）$\mathrm{N}\left(\mu_1, \sigma^2\right), \ldots, \mathrm{N}\left(\mu_k, \sigma^2\right)$ から独立に抽出した大きさ $n_1, \ldots, n_k$ の無作為標本を $(y_{1,1}, \ldots, y_{1,n_1}), \ldots, (y_{k,1}, \ldots, y_{k,n_k})$ とする．$n := n_1 + \cdots + n_k$ 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．

2. 次の OLS 問題を考える．

$$\min_{a,b} \quad \sum_{i=1}^{n}(y_i - a - bx_i)^2$$
$$\text{and} \quad a, b \in \mathbb{R}$$

OLS 問題の解を (a^*, b^*)，回帰残差を $e_i := y_i - a^* - b^* x_i$ とする．また $((y_1, x_1), \ldots, (y_n, x_n))$ 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)

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

解答例

1.（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）

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

ここで

$$
\begin{aligned}
\sum_{i=1}^{n_h}(y_{h,i}-\bar{y}_h) &= \sum_{i=1}^{n_h}y_{h,i}-n_h\bar{y}_h \\
&= \sum_{i=1}^{n_h}y_{h,i}-\sum_{i=1}^{n_h}y_{h,i} \\
&= 0
\end{aligned}
$$

したがって第 2 項は 0.

2.（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

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

(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

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

前問より第 2 項は 0.