経済統計：期末試験

村澤 康友

2017 年 7 月 31 日

注意：3 問とも解答すること．結果より思考過程を重視するので，途中計算等も必ず書くこと（部分点は大いに与えるが，結果のみの解答は 0 点とする）．

1. （20 点）以下の用語の定義を式または言葉で書きなさい（各 20 字程度）．
   （a）両側検定問題
   （b）F 統計量
   （c）回帰モデル
   （d）最良線形不偏推定量（BLUE）
2. （30 点）大きさ n の 1 変量データを $(x_1, \dots, x_n)$ とする．$\mu := \mathrm{E}(x_i)$ を OLS で推定したい．
   （a）OLS 問題を書きなさい．
   （b）μ の OLS 推定量を求めなさい．
   （c）OLS 残差の和が 0 となることを示しなさい．
3. （50 点）X 新聞と Y 放送が同日に行った内閣支持率の世論調査の結果は次の通りであった．

	有効回答数	内閣支持率
X 新聞	400	50%
Y 放送	400	40%

   両調査とも母集団は日本の有権者全体である．調査結果の差異について以下の問いに答えなさい．
   （a）$\mathrm{Bin}(1, p)$ の平均と分散を求めなさい．
   （b）$\mathrm{Bin}(1, p_X), \mathrm{Bin}(1, p_Y)$ から独立に抽出した無作為標本 $(X_1, \dots, X_m), (Y_1, \dots, Y_n)$ の標本比率（＝標本平均）を $\hat{p}_X, \hat{p}_Y$ とする．$\hat{p}_X - \hat{p}_Y$ の漸近分布を求めなさい．（ヒント：まず $\hat{p}_X, \hat{p}_Y$ の漸近分布を求める．）
   （c）次の検定を（近似的に）行うための検定統計量を式で与えなさい．

$$H_0 : p_X = p_Y \quad \text{vs.} \quad H_1 : p_X \neq p_Y$$

   （d）検定の有意水準を 5 %とする．標準正規分布表を利用して検定の棄却域を定めなさい．
   （e）上の世論調査の結果について検定を実行しなさい．

解答例

1. 統計学の基本用語
   (a) $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.
      - 検定問題として書かなければ 0 点.
   (b) H_0 の下で F 分布にしたがう検定統計量.
   (c) $\mathrm{E}(Y|X)$ を与える式.
      - $\mathrm{E}(Y|X)$ のみは 3 点.
   (d) 分散が最小となる線形不偏推定量.
      - 「誤差が最小」はダメ.
2. OLS
   (a)
$$\begin{aligned} \min_m \quad & \sum_{i=1}^n (x_i - m)^2 \\ \text{and} \quad & m \in \mathbb{R} \end{aligned}$$
   (b) 1 階の条件は
$$\sum_{i=1}^n (-2)(x_i - m^*) = 0$$
   正規方程式は
$$\sum_{i=1}^n x_i - nm^* = 0$$
   したがって
$$m^* = \frac{1}{n} \sum_{i=1}^n x_i$$
      - 正規方程式で 5 点.
   (c) OLS 残差は $e_i := x_i - m^*$. したがって
$$\sum_{i=1}^n e_i = \sum_{i=1}^n (x_i - m^*)$$
   正規方程式より右辺は 0.
      - OLS 残差で 5 点.
      - OLS 残差を示さなければ 0 点.
3. 母比率の差の検定
   (a) 平均は
$$1 \cdot p + 0 \cdot (1-p) = p$$
   分散は
$$\begin{aligned} (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) &= p(1-p)^2 + p^2(1-p) \\ &= p(1-p) \end{aligned}$$
      - 各 5 点.

（b）標本比率の漸近分布は

$$\hat{p}_X \overset{a}{\sim} \mathrm{N}\left(p_X, \frac{p_X(1-p_X)}{m}\right)$$
$$\hat{p}_Y \overset{a}{\sim} \mathrm{N}\left(p_Y, \frac{p_Y(1-p_Y)}{n}\right)$$

両者は独立なので

$$\hat{p}_X - \hat{p}_Y \overset{a}{\sim} \mathrm{N}\left(p_X - p_Y, \frac{p_X(1-p_X)}{m} + \frac{p_Y(1-p_Y)}{n}\right)$$

- 標本比率の漸近分布で 5 点.
- 母分散を単に σ_X^2, σ_Y^2 と書いたら 5 点減.
- この段階で $p_X = p_Y$ としたら 0 点.
- 母数 p_X, p_Y と推定量 $\hat{p}_X, \hat{p}_Y$ を混同したら 0 点.

（c）標準化すると

$$\frac{\hat{p}_X - \hat{p}_Y - (p_X - p_Y)}{\sqrt{p_X(1-p_X)/m + p_Y(1-p_Y)/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

検定統計量は

$$Z := \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_X(1-\hat{p}_X)/m + \hat{p}_Y(1-\hat{p}_Y)/n}}$$

- 分母を p_X, p_Y で書いたら 5 点減. $p_X = p_Y = p$ としても同様.
- 標本分散を単に s_X^2, s_Y^2 と書いたら 5 点減.

（d）棄却域は $(-\infty, -1.96] \cup [1.96, \infty]$.

- $[1.96, \infty]$ のみは 5 点減.
- 片側検定は 0 点.

（e）検定統計量の値は

$$\begin{aligned} Z &= \frac{.5 - .4}{\sqrt{.5(1-.5)/400 + .4(1-.4)/400}} \\ &= \frac{.1}{\sqrt{.49/400}} \\ &= \frac{.1}{.7/20} \\ &\approx 2.86 \end{aligned}$$

したがって H_0 は棄却される.

- 検定統計量の値が正しくなければ 0 点.
- 検定統計量の値が正しくても結論を間違えたら 5 点減.