経済統計 II：定期試験

村澤 康友

2021 年 8 月 2 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）．教科書のみ参照してよい（他の講義資料・ノートは持込不可）．

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）対立仮説
   （b）t 統計量
   （c）正規方程式
   （d）古典的線形回帰モデル
2. （30 点）家計の飲食費支出の所得弾力性を推定したい．そこで無作為標本を用いて両変数の対数値について単回帰分析を行い，次の結果を得た．

$$\widehat{\text{l\_foodexp}} = \underset{(0.13820)}{0.545142} + \underset{(0.020323)}{0.855897}\,\text{l\_income}$$
$$T = 235 \quad \bar{R}^2 = 0.8834 \quad F(1,233) = 1773.7 \quad \hat{\sigma} = 0.13679$$

(丸括弧内は標準誤差)

   （a）所得弾力性を β とする．β の t 値を求めなさい．
   （b）古典的正規線形回帰モデルを仮定する．β の t 値は H_0 の下でどのような分布に従うか？
   （c）β の 95 %信頼区間を求めなさい．
3. （50 点）X 社と Y 社の新型コロナワクチンについて，日本の 20 代の若者を対象に 2 回目接種後の副反応発生率の差異の有無を検証したい．両社のワクチン接種後に発熱する人の割合を p_X, p_Y，独立に抽出した大きさ n_X, n_Y の無作為標本で実際に発熱した人の割合を $\hat{p}_X, \hat{p}_Y$ とする．
   （a）検定問題を定式化しなさい（問題意識を踏まえること）．
   （b）$\hat{p}_X - \hat{p}_Y$ の漸近分布を求めなさい．
   （c）H_0 の下で $\mathrm{N}(0,1)$ に従う検定統計量を与えなさい．
   （d）H_0 の下で $\chi^2(1)$ に従う検定統計量を与えなさい．
   （e）$n_X = 250, n_Y = 400$, $\hat{p}_X = .5, \hat{p}_Y = .4$ として χ^2 検定統計量の値と漸近 p 値を求めなさい．
   ※数値例はフィクションです．

解答例

1. 統計学の基本用語
   (a) 帰無仮説を棄却するとき代わりに採択する仮説.
      - 「帰無仮説と対立する仮説」は「対立」の意味が不明（片側？両側？）なのでダメ.
   (b) H_0 の下で t 分布にしたがう検定統計量.
      - 「H_0 の下で」がなければ 0 点.
   (c) OLS 問題の 1 階の条件を整理した式.
   (d) 誤差項が無相関で分散が均一な線形回帰モデル.
      - 「古典的正規線形回帰モデル」の定義は 1 点.
2. 単回帰分析
   (a)
$$\begin{aligned} t &= \frac{0.855897}{0.020323} \\ &\approx 42.11 \end{aligned}$$
   (b) H_0 の下で $t \sim \mathrm{t}(n-k)$. $n = 235$, $k = 2$ より $t \sim \mathrm{t}(233)$.
   (c) β の OLS 推定量を b, その標準誤差を s とすると
$$\frac{b-\beta}{s} \sim \mathrm{t}(233)$$
   したがって
$$\Pr\left[-1.970 \le \frac{b-\beta}{s} \le 1.970\right] \approx .95$$
   または
$$\Pr[-1.970s \le b-\beta \le 1.970s] \approx .95$$
   または
$$\Pr[b-1.970s \le \beta \le b+1.970s] \approx .95$$
   $b = 0.855897$, $s = 0.020323$ より β の 95 %信頼区間は $[0.816, 0.896]$.
      - $\mathrm{t}(240)$ で近似できるので $\mathrm{N}(0,1)$ による近似は 1 点減.
3. 母比率の差の両側検定
   (a)
$$H_0 : p_X = p_Y \quad \text{vs} \quad H_1 : p_X \ne p_Y$$
   (b)
$$\begin{aligned} \hat{p}_X &\overset{a}{\sim} \mathrm{N}\left(p_X, \frac{p_X(1-p_X)}{n_X}\right) \\ \hat{p}_Y &\overset{a}{\sim} \mathrm{N}\left(p_Y, \frac{p_Y(1-p_Y)}{n_Y}\right) \end{aligned}$$
   したがって
$$\hat{p}_X - \hat{p}_Y \overset{a}{\sim} \mathrm{N}\left(p_X - p_Y, \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right)$$

（c）標準化すると

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

検定統計量は

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

（d）χ^2 検定統計量は

$$\begin{aligned}\chi^2 &:= Z^2 \\ &= \frac{(\hat{p}_X - \hat{p}_Y)^2}{\hat{p}_X(1-\hat{p}_X)/n_X + \hat{p}_Y(1-\hat{p}_Y)/n_Y}\end{aligned}$$

（e）

$$\begin{aligned}\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} &= \frac{(1/2)(1-1/2)}{250} \\ &= \frac{(1/2)(1/2)}{250} \\ &= \frac{1}{1000} \\ \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} &= \frac{(2/5)(1-2/5)}{400} \\ &= \frac{(2/5)(3/5)}{400} \\ &= \frac{6}{10000}\end{aligned}$$

したがって

$$\begin{aligned}\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} &= \frac{1}{1000} + \frac{6}{10000} \\ &= \frac{16}{10000} \\ &= \frac{4^2}{100^2}\end{aligned}$$

すなわち

$$\begin{aligned}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}} &= \frac{4}{100} \\ &= \frac{1}{25}\end{aligned}$$

χ^2 検定統計量は

$$\begin{aligned}\chi^2 &:= Z^2 \\ &= \left(\frac{.5 - .4}{1/25}\right)^2 \\ &= 2.5^2 \\ &= 6.25\end{aligned}$$

標準正規分布表より漸近 p 値は

$$
\begin{aligned}
\Pr\left[\chi^2 \geq 6.25|H_0\right] &= \Pr\left[Z^2 \geq 6.25|H_0\right] \\
&= \Pr[|Z| \geq 2.5|H_0] \\
&= 2\Pr[Z \geq 2.5|H_0] \\
&= 2 \cdot 0.00620967 \\
&= 0.0124193
\end{aligned}
$$

- χ^2 検定統計量 5 点，漸近 p 値 5 点.
- 漸近 p 値は χ^2 分布表からは読み取れないが，標準正規分布表から計算できる.