中級統計学：第 3 回中間試験

村澤 康友

2020 年 12 月 18 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）有限母集団
   （b）母分散
   （c）標本和
   （d）漸近（大標本）特性
2. （30 点）分布表を用いて以下の問いに答えなさい．
   （a）$X \sim \chi^2(8)$ とする．$\Pr[a \leq X \leq b] = .99$ となる a, b を求めなさい．
   （b）$Y \sim \mathrm{t}(36)$ とする．$\Pr[|Y| \leq c] = .99$ となる c を求めなさい．
   （c）$Z \sim \mathrm{F}(8, 30)$ とする．$\Pr[d \leq Z \leq e] = .99$ となる d, e を求めなさい．
   なお a〜e はすべて正の実数（$0, \infty$ は含まない）とする．
3. （50 点）Go To トラベル事業の 2020 年 8 月末までの利用者と非利用者で，9 月末までに発熱症状があった人の割合を比較したい．独立に抽出した大きさ n_X, n_Y の無作為標本で，発熱症状があった人の割合を $\hat{p}_X$, $\hat{p}_Y$ とする．
   （a）2 項母集団 $\mathrm{Bin}(1, p_X)$, $\mathrm{Bin}(1, p_Y)$ の平均と分散を求めなさい．
   （b）$\hat{p}_X$, $\hat{p}_Y$ の漸近分布を求めなさい．
   （c）$\hat{p}_X - \hat{p}_Y$ の漸近分布を求めなさい．
   （d）$p_X - p_Y$ の 95% 信頼区間を近似的に求めなさい．
   （e）$n_X = 2500$, $n_Y = 6400$, $\hat{p}_X = .05$, $\hat{p}_Y = .04$ として 95% 信頼区間を近似的に計算し，それが $p_X - p_Y = 0$ を含むかどうか調べなさい．
   ※数値例はフィクションです．この分析は Go To トラベル事業と発熱症状の相関関係の検証であり，結果を因果関係と解釈するのは誤りです．

解答例

1. 統計学の基本用語
   (a) 有限個の個体から成る母集団.
   (b) 母集団分布の分散.
      - 「母集団の分散」は意味が不明確なので 2 点.
   (c) 標本 $(X_1, \dots, X_n)$ の標本和は $T := X_1 + \cdots + X_n$.
      - 「標本の和」は意味が不明確なので 0 点.「標本の観測値の和」は OK.
   (d) 推定量の漸近分布に関する性質.
2. 分布表の読み方
   (a)

$$\begin{aligned}\Pr[a \le X \le b] &= \Pr[X \ge a] - \Pr[X > b] \\ &= .99\end{aligned}$$

   これを満たす例は

$$\begin{aligned}\Pr[X \ge a] &= .995 \\ \Pr[X > b] &= .005\end{aligned}$$

   χ^2 分布表より $X \sim \chi^2(8)$ なら $a = 1.34441$，$b = 21.9550$.
   - 各 5 点.

   (b) t 分布の対称性より

$$\begin{aligned}\Pr[|Y| \le c] &= \Pr[-c \le Y \le c] \\ &= 1 - 2\Pr[Y > c] \\ &= .99\end{aligned}$$

   すなわち

$$\Pr[Y > c] = .005$$

   t 分布表より $Y \sim \mathrm{t}(36)$ なら $c = 2.719$.

   (c)

$$\begin{aligned}\Pr[d \le Z \le e] &= 1 - \Pr[Z < d] - \Pr[Z > e] \\ &= .99\end{aligned}$$

   これを満たす例は

$$\begin{aligned}\Pr[Z < d] &= \Pr\left[\frac{1}{Z} > \frac{1}{d}\right] \\ &= .005 \\ \Pr[Z > e] &= .005\end{aligned}$$

   $Z \sim \mathrm{F}(8, 30)$ なら $1/Z \sim \mathrm{F}(30, 8)$ なので F 分布表より $1/d = 6.396$，すなわち $d = 1/6.396$．同じく F 分布表より $Z \sim \mathrm{F}(8, 30)$ なら $e = 3.580$.
   - 各 5 点.

3. 母比率の差の信頼区間

（a）$\mathrm{Bin}(1, p_X)$ の平均は

$$1 \cdot p_X + 0 \cdot (1 - p_X) = p_X$$

分散は

$$\begin{aligned}(1 - p_X)^2 \cdot p_X + (0 - p_X)^2 \cdot (1 - p_X) &= (1 - p_X)^2 p_X + p_X^2 (1 - p_X) \\ &= p_X(1 - p_X)\end{aligned}$$

$\mathrm{Bin}(1, p_Y)$ についても同様.

（b）

$$\begin{aligned}\hat{p}_X &\overset{a}{\sim} \mathrm{N}\left(p_X, \frac{p_X(1 - p_X)}{n_X}\right) \\ \hat{p}_Y &\overset{a}{\sim} \mathrm{N}\left(p_Y, \frac{p_Y(1 - p_Y)}{n_Y}\right)\end{aligned}$$

（c）

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

- 前問の解答と整合的なら OK.

（d）標準化すると

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

分母の p_X, p_Y を $\hat{p}_X, \hat{p}_Y$ に置き換えると

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

したがって

$$\Pr\left[-1.96 \le \frac{\hat{p}_X - \hat{p}_Y - (p_X - p_Y)}{\sqrt{\hat{p}_X(1 - \hat{p}_X)/n_X + \hat{p}_Y(1 - \hat{p}_Y)/n_Y}} \le 1.96\right] \approx .95$$

（近似的な）95% 信頼区間の上限・下限は

$$\hat{p}_X - \hat{p}_Y \pm 1.96\sqrt{\frac{\hat{p}_X(1 - \hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1 - \hat{p}_Y)}{n_Y}}$$

(e)

$$\begin{aligned}
\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} &= \frac{(1/20)(1-1/20)}{2500} \\
&= \frac{(1/20)(19/20)}{50^2} \\
&= \frac{19}{1000^2} \\
\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} &= \frac{(1/25)(1-1/25)}{6400} \\
&= \frac{(1/25)(24/25)}{80^2} \\
&= \frac{24}{25^2 80^2} \\
&= \frac{6}{25^2 40^2} \\
&= \frac{6}{1000^2}
\end{aligned}$$

したがって

$$\begin{aligned}
\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y} &= \frac{19}{1000^2} + \frac{6}{1000^2} \\
&= \frac{25}{1000^2}
\end{aligned}$$

すなわち

$$\begin{aligned}
\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}} &= \frac{5}{1000} \\
&= \frac{1}{200}
\end{aligned}$$

（近似的な）95% 信頼区間は

$$\left[.01 - \frac{1.96}{200}, .01 + \frac{1.96}{200}\right] = [.0002, .0198]$$

したがって 0 を含まない.