中級統計学：第 3 回中間試験

村澤 康友

2021 年 7 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) 標本の大きさ
   (b)（母数の）尤度
   (c) 信頼区間
   (d) 母比率
2. （30 点）分布表を用いて以下の問いに答えなさい．
   (a) $X \sim \chi^2(20)$ とする．$\Pr[a \leq X \leq b] = .95$ となる a, b を求めなさい．
   (b) $Y \sim \mathrm{t}(21)$ とする．$\Pr[|Y| \leq c] = .98$ となる c を求めなさい．
   (c) $Z \sim \mathrm{F}(7, 2)$ 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）ある母集団で新型コロナワクチンの接種を拒否する人の割合 p を区間推定したい．拒否を 1，その他を 0 で表し，無作為に選んだ n 人の回答を $(X_1, \ldots, X_n)$ とする．また標本比率（＝標本平均）を $\hat{p}$ とする．
   (a) $\mathrm{E}(X_i)$ と $\mathrm{var}(X_i)$ を求めなさい．
   (b) $\mathrm{E}(\hat{p})$ と $\mathrm{var}(\hat{p})$ を求めなさい．
   (c) $\hat{p}$ の漸近分布を求めなさい．
   (d) p の 95% 信頼区間を近似的に求めなさい．
   (e) $n = 10000$，$\hat{p} = .1$ として p の 95% 信頼区間を近似的に計算しなさい．

解答例

1. 統計学の基本用語
   (a) 標本に含まれる個体の数.
      - 「大きさ」の定義が必要.
      - 「標本 $(X_1,\dots,X_n)$ における n」は n の説明がないのでダメ.
      - 1つの標本の大きさの定義なので「標本の数」は間違い.
   (b) その母数の下で標本の実現値を観測する確率（密度）.
      - 「尤もらしさ」は定義でないのでダメ．その定義が必要.
   (c) ある確率で母数を含む確率的な区間.
      - 「ある確率で母数を含む区間」は母数と区間のどちらが確率的か不明確.「確率的な区間」と明示しなければダメ．※ベイズ統計学では母数を確率的とする.
      - 「1次元の信頼域」は可とする.
   (d) ベルヌーイ母集団における成功（= 1）の比率.
      - 「母集団の比率」は意味が不明なのでダメ.
2. 分布表の読み方
   (a)

$$\begin{aligned}\Pr[a \le X \le b] &= \Pr[X \ge a] - \Pr[X > b]\\ &= .95\end{aligned}$$

これを満たす例は

$$\begin{aligned}\Pr[X \ge a] &= .975\\ \Pr[X > b] &= .025\end{aligned}$$

χ^2 分布表より $X \sim \chi^2(20)$ なら $a = 9.59078$, $b = 34.1696$.

   - 各5点.

   (b) t 分布の対称性より

$$\begin{aligned}\Pr[|Y| \le c] &= \Pr[-c \le Y \le c]\\ &= 1 - 2\Pr[Y > c]\\ &= .98\end{aligned}$$

すなわち

$$\Pr[Y > c] = .01$$

t 分布表より $Y \sim \mathrm{t}(21)$ なら $c = 2.518$.

   (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

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

$Z \sim \mathrm{F}(7,2)$ なら $1/Z \sim \mathrm{F}(2,7)$ なので F 分布表より $1/d = 12.404$，すなわち $d = 1/12.404$．同じく F 分布表より $Z \sim \mathrm{F}(7,2)$ なら $e = 199.357$.

- 各 5 点.

3. 母比率の信頼区間

（a）

$$\begin{aligned}
\mathrm{E}(X_i) &:= 1 \cdot p + 0 \cdot (1-p) \\
&= p \\
\mathrm{var}(X_i) &= \mathrm{E}\left(X_i^2\right) - \mathrm{E}(X_i)^2 \\
&= \mathrm{E}(X_i) - \mathrm{E}(X_i)^2 \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}$$

- μ, σ^2 で表すのはダメ．以下同様.

（b）期待値の線形性より

$$\begin{aligned}
\mathrm{E}(\hat{p}) &= \mathrm{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{\mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)}{n} \\
&= \frac{np}{n} \\
&= p
\end{aligned}$$

$X_1, \ldots, X_n$ は独立なので

$$\begin{aligned}
\mathrm{var}(\hat{p}) &= \mathrm{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} \\
&= \frac{np(1-p)}{n^2} \\
&= \frac{p(1-p)}{n}
\end{aligned}$$

（c）$X_1, \ldots, X_n$ は iid なので，中心極限定理より

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

（d）標準化すると

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

または

$$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

標準正規分布表より

$$\Pr\left[-1.96 \le \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \le 1.96\right] \approx .95$$

ここで

$$\begin{aligned}
-1.96 \le \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \le 1.96 &\Longleftrightarrow -1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le \hat{p}-p \le 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&\Longleftrightarrow \hat{p}-1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p}+1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\end{aligned}$$

したがって p の 95 %信頼区間は近似的に

$$\left[\hat{p}-1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p}+1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

(e) $n=10000$, $\hat{p}=.1$ を代入すると

$$\begin{aligned}
\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= \sqrt{\frac{.1(1-.1)}{10000}} \\
&= \frac{.3}{100} \\
&= \frac{3}{1000}
\end{aligned}$$

したがって 95% 信頼区間は

$$\left[.1-\frac{1.96\cdot 3}{1000}, .1+\frac{1.96\cdot 3}{1000}\right] \approx [.09422, .10588]$$

- $\hat{p}=.1$ を代入した後で $\Pr[.09422 \le p \le .10588] \approx .95$ と書くのは重大な誤り.