

# 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（部分点は大きいに与えるが、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）。

- (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）。
  - 標本の大きさ
  - (母数の) 尤度
  - 信頼区間
  - 母比率
- (30点) 分布表を用いて以下の問いに答えなさい。
  - $X \sim \chi^2(20)$  とする。  $\Pr[a \leq X \leq b] = .95$  となる  $a, b$  を求めなさい。
  - $Y \sim t(21)$  とする。  $\Pr[|Y| \leq c] = .98$  となる  $c$  を求めなさい。
  - $Z \sim F(7, 2)$  とする。  $\Pr[d \leq Z \leq e] = .99$  となる  $d, e$  を求めなさい。なお  $a \sim e$  はすべて正の実数 ( $0, \infty$  は含まない) とする。
- (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  $(X_1, \dots, X_n)$  とする。また標本比率 (= 標本平均) を  $\hat{p}$  とする。
  - $E(X_i)$  と  $\text{var}(X_i)$  を求めなさい。
  - $E(\hat{p})$  と  $\text{var}(\hat{p})$  を求めなさい。
  - $\hat{p}$  の漸近分布を求めなさい。
  - $p$  の95% 信頼区間を近似的に求めなさい。
  - $n = 10000$ ,  $\hat{p} = .1$  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)

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

これを満たす例は

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

$\chi^2$  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

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

すなわち

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

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

(c)

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

これを満たす例は

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

$Z \sim F(7, 2)$  なら  $1/Z \sim F(2, 7)$  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  $Z \sim F(7, 2)$  なら  $e = 199.357$ .

- 各 5 点.

### 3. 母比率の信頼区間

(a)

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

- $\mu, \sigma^2$  で表すのはダメ. 以下同様.

(b) 期待値の線形性より

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

$X_1, \dots, X_n$  は独立なので

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

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

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

(d) 標準化すると

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

または

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

標準正規分布表より

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

ここで

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

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  $\Pr[.09422 \leq p \leq .10588] \approx .95$  と書くのは重大な誤り.