中級統計学：定期試験

村澤 康友

2024 年 1 月 26 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）以下で定義される統計学の専門用語をそれぞれ書きなさい．
   (a) 帰無仮説を棄却するとき代わりに採択する仮説
   (b) 帰無仮説が真なのに帰無仮説を棄却する誤り
   (c) $\mathrm{E}(Y|X)$ を与える式
   (d) X の 1 単位の増加に対する Y の変化
2. （30 点）次表は自作のサイコロを 60 回振った結果である．このサイコロが公正かどうかを調べたい．

出目	1	2	3	4	5	6	計
度数	14	11	11	7	9	8	60

   (a) このサイコロで $1,\dots,6$ が出る確率を $p_1,\dots,p_6$ とする．検定問題を定式化しなさい．
   (b) 適合度検定統計量は H_0 の下でどのような分布に近似的に従うか？また有意水準 5% の検定の棄却域を定めなさい．
   (c) 適合度検定統計量の値を求め，有意水準 5% の検定の結果を述べなさい．
3. （50 点）サッカーの PK 戦における先攻有利説を検証したい．先攻の勝ちを 1，負けを 0 で表すと，先攻の勝敗は $\mathrm{Bin}(1,p)$ にしたがう．無作為に選んだ n 回の PK 戦の結果を $(X_1,\dots,X_n)$ とし，先攻の勝率（標本比率）を $\hat{p}_n$ 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) $\hat{p}_n$ の漸近分布を求めなさい（要証明）．
   (c) 検定統計量を定義し，その H_0 の下での分布を求め，有意水準 5 %の検定の棄却域を定めなさい．
   (d) $n=100$，$\hat{p}_n=.58$ として検定統計量の値を求め，検定を実行しなさい．
   (e) 検定統計量の値の（漸近）p 値を求めなさい．

解答例

1. 統計学の基本用語
   (a) 対立仮説
   (b) 第 1 種の誤り
   (c) 回帰モデル（回帰式，回帰関数）
      - 「回帰」のみは 1 点減.
   (d) 限界効果
2. 適合度検定
   (a)

$$H_0 : \begin{pmatrix} p_1 \\ \vdots \\ p_6 \end{pmatrix} = \begin{pmatrix} 1/6 \\ \vdots \\ 1/6 \end{pmatrix} \quad \text{vs} \quad H_1 : \begin{pmatrix} p_1 \\ \vdots \\ p_6 \end{pmatrix} \neq \begin{pmatrix} 1/6 \\ \vdots \\ 1/6 \end{pmatrix}$$

   (b) 適合度検定統計量は H_0 の下で $\chi^2 \overset{a}{\sim} \chi^2(5)$. 棄却域は $[11.0705, \infty)$.
   (c) 適合度検定統計量の値は

$$\begin{aligned}\chi^2 &:= \frac{(14-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(8-10)^2}{10} \\ &= \frac{16+1+1+9+1+4}{10} \\ &= \frac{32}{10}\end{aligned}$$

   χ^2 値が棄却域に入らないので H_0 は棄却されない. すなわちサイコロは不公正とは言えない.
3. 母比率の片側検定
   (a)

$$H_0 : p = .5 \quad \text{vs} \quad H_1 : p > .5$$

   (b) $X \sim \mathrm{Bin}(1, p)$ とすると

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

   期待値の線形性より

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

$X_1, \ldots, X_n$ は独立なので

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

$X_1, \ldots, X_n$ は iid なので，中心極限定理より

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

- 平均 4 点，分散 4 点，分布 2 点.
- 結果のみは 0 点.

（c）標準化すると

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

$H_0 : p = .5$ を代入すると，検定統計量は

$$Z := \frac{\hat{p} - .5}{\sqrt{.5(1-.5)/n}}$$

H_0 の下で

$$Z \overset{a}{\sim} \mathrm{N}(0,1)$$

標準正規分布表より H_0 の下で

$$\Pr[Z \geq 1.65] \approx .05$$

したがって棄却域は $[1.65, \infty)$.

- 検定統計量 5 点，棄却域 5 点.

（d）$n = 100$，$\hat{p}_n = .58$ なら

$$\begin{aligned}
Z &:= \frac{.58 - .5}{\sqrt{.5(1-.5)/100}} \\
&= \frac{.08}{\sqrt{1/400}} \\
&= .08 \cdot 20 \\
&= 1.6
\end{aligned}$$

これは棄却域に入らないので H_0 は採択.

（e）標準正規分布表より $Z = 1.6$ なら p 値は.0547993.