データ解析

サンプルサイズ設計2

創域理工学部
Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年12月21日

Contents

- 母比率の差に関する2標本検定問題
  - Unconditional approach
  - Conditional approach

- サンプルサイズ設計法
  - Unconditional approach
  - Conditional approach

諸前提

- 試験治療Tと対照治療Cの応答変数の母比率の差に関する両側検定を考える
  - 試験治療Tの応答変数 $X_i \underset{\text{i.i.d}}{\sim} B(\pi_1, 1)\ (i = 1,2,\dots,n)$
  - 対照治療Cの応答変数 $Y_j \underset{\text{i.i.d}}{\sim} B(\pi_2, 1)\ (j = 1,2,\dots,m)$

- 母比率が大きいことが臨床的に望ましい状態とする
  - $\varepsilon = \pi_1 - \pi_2 > 0$が臨床的に望ましい

母比率の差に関する2標本検定

- 試験治療の対照治療に対する優越性を検証
- 仮説
  - 帰無仮説 $H_0: \varepsilon = 0$
  - 対立仮説 $H_1: \varepsilon \neq 0$
- 応答変数の標本平均

$$\bar{\pi}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{\pi}_2 = \frac{1}{m}\sum_{j=1}^{m} Y_j$$

- 割付比
  - 各群の参加者数の比を$r = m/n$とし，$\kappa = r/(r+1)$とする

相似検定

複合仮説　　　　　　　　　　　　　複合仮説

帰無仮説 $H_0: \theta \in \Theta_0$　　対立仮説 $H_1: \theta \in \Theta_1 = \Theta \cup \Theta_0^c$

$\Theta = \Theta_0 \cup \Theta_0^c$ かつ $\Theta_0 \cap \Theta_0^c = \phi$

すべての$\theta_0 \in \Theta_0$に対して,

$$\beta_W(\theta_0) = P_{\theta_0}\left((X_1, X_2, \dots, X_n) \in W\right) = \alpha$$

を満たす棄却域Wを用いた検定を相似検定 (similar test) という

母比率の差に関する2標本相似検定

- 母集団1
  - パラメータπ_1のベルヌーイ母集団
  - 無作為標本$X_1, X_2, \ldots, X_n$
- 母集団2
  - パラメータπ_2のベルヌーイ母集団
  - 無作為標本$Y_1, Y_2, \ldots, Y_m$
- 仮説
  - $H_0: \pi_1 = \pi_2$ vs $H_1: \pi_1 \neq \pi_2$

この統計的仮説検定に対する相似検定を構成する

Unconditional approach (1)

$$X_i \underset{\text{i.i.d}}{\sim} B(\pi_1, 1) \ \ (i = 1, \dots, n) \qquad \sum_{i=1}^{n} X_i \sim B(\pi_1, n)$$

$\perp$

$$Y_j \underset{\text{i.i.d}}{\sim} B(\pi_2, 1) \ \ (j = 1, \dots, m) \qquad \sum_{j=1}^{m} Y_j \sim B(\pi_2, m)$$

$$\bar{\pi}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{n \to \infty}{\sim} N\left(\pi_1, \frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n}\right) \qquad \bar{\pi}_2 = \frac{1}{m}\sum_{j=1}^{m} Y_j \underset{m \to \infty}{\sim} N\left(\pi_2, \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}\right)$$

$\bar{\pi}_1 \perp \bar{\pi}_2$であることから $\bar{\pi}_1 - \bar{\pi}_2 \sim N\left(\pi_1 - \pi_2, \frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}\right)$

Unconditional approach (2)

$H_0: \pi_1 = \pi_2$のもとで

$$\bar{\pi}_1 - \bar{\pi}_2 \sim N\left(0, \frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}\right), \quad Z \equiv \frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}} \sim N(0,1)$$

$H_1: \pi_1 \neq \pi_2$のもとで

$$\bar{\pi}_1 - \bar{\pi}_2 \sim N\left(\pi_1 - \pi_2, \frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}\right),$$

$$Z \equiv \frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}} \sim N\left(\frac{\pi_1 - \pi_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}, 1\right)$$

Unconditional approach (3)

棄却域Wを次のようにする。

$$W = \left\{ (x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m) \;\middle|\; |Z| > z\left(\frac{\alpha}{2}\right) \right\}$$

H_0: $\pi_1 = \pi_2$のもとで

$$\beta_W(\mu_1, \mu_2) = P_{\theta_0}\big((X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_m) \in W\big) = \alpha$$

第1種の過誤確率

$$\Leftrightarrow \int_{|Z| > z\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) dZ = \alpha$$

この棄却域Wに基づく検定は相似検定である

Unconditional approach (4)

$H_1 : \pi_1 \neq \pi_2$のもとで

$$\beta_W(\mu_1, \mu_2) = P_{\theta_1}\left((X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_m) \in W\right)$$

検出力

$$\Leftrightarrow \int_{|Z|>z\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(Z - \frac{\pi_1 - \pi_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}\right)^2\right) dZ$$

Conditional approach (1)

$$X_i \underset{\text{i.i.d}}{\sim} B(\pi_1, 1) \ \ (i = 1, \dots, n)$$

$$\sum_{i=1}^{n} X_i \sim B(\pi_1, n)$$

$\perp$

$$Y_j \underset{\text{i.i.d}}{\sim} B(\pi_2, 1) \ \ (i = 1, \dots, m)$$

$$\sum_{i=1}^{m} Y_i \sim B(\pi_2, m)$$

$$\bar{\pi} = \frac{\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i}{n + m}$$

$H_0 : \pi_1 = \pi_2 \ (= \pi)$ のもとで

$$\bar{\pi}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \underset{n \to \infty}{\sim} N\left(\pi_1, \frac{\bar{\pi}(1 - \bar{\pi})}{n}\right) \qquad \bar{\pi}_2 = \frac{1}{m} \sum_{i=1}^{m} Y_i \underset{m \to \infty}{\sim} N\left(\pi_2, \frac{\bar{\pi}(1 - \bar{\pi})}{m}\right)$$

$\bar{\pi}_1 \perp \bar{\pi}_2$であることから $\quad \bar{\pi}_1 - \bar{\pi}_2 \sim N\left(\pi_1 - \pi_2, \left(\frac{1}{n} + \frac{1}{m}\right) \bar{\pi}(1 - \bar{\pi})\right)$

Conditional approach (2)

$H_0: \pi_1 = \pi_2\ (= \pi)$ のもとで

$$\bar{\pi}_1 - \bar{\pi}_2 \sim N\left(0, \left(\frac{1}{n} + \frac{1}{m}\right)\bar{\pi}(1-\bar{\pi})\right), \qquad Z \equiv \frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\bar{\pi}(1-\bar{\pi})}} \sim N(0,1)$$

$H_1: \pi_1 \neq \pi_2$ のもとで

$$\bar{\pi}_1 - \bar{\pi}_2 \sim N\left(\pi_1 - \pi_2, \frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}\right),$$

$$Z \equiv \frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}} \sim N\left(\frac{\pi_1 - \pi_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}, 1\right)$$

Conditional approach (3)

棄却域Wを次のようにする。

$$W = \left\{ (x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m) \;\middle|\; |Z| > z\left(\frac{\alpha}{2}\right) \right\}$$

$H_0 : \pi_1 = \pi_2$のもとで

$$\beta_W(\mu_1, \mu_2) = P_{\theta_0}\left((X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_m) \in W\right) = \alpha$$

第1種の過誤確率

$$\Leftrightarrow \int_{|Z| > z\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) dZ = \alpha$$

この棄却域Wに基づく検定は相似検定である

Conditional approach (4)

$H_1 : \pi_1 \neq \pi_2$のもとで

$$\beta_W(\mu_1, \mu_2) = P_{\theta_1}\left((X_1, X_2, \dots, X_n; Y_1, Y_2, \dots, Y_m) \in W\right)$$

検出力

$$\Leftrightarrow \int_{|Z|>z\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(Z - \frac{\pi_1 - \pi_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}\right)^2\right) dZ$$

Unconditional approachの場合

- 次式が成立するときに帰無仮説 H_0 を有意水準αで棄却

$$\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}\right| \geq z_{\alpha/2}$$

$z_{\alpha/2}$：標準正規分布の上側$\alpha/2$％点

- サンプルサイズ設計（検出力の計算）
  - α, r, π_1, π_2はある値に固定する

検出力の計算 (1)

$r = \frac{m}{n}$ $\kappa = \frac{r}{r+1}$ $\frac{1-\kappa}{\kappa} = \frac{1}{r}$

- 対立仮説 $H_1\colon \varepsilon \neq 0$ のもとで，検出力はnの関数になる

$$\phi(n) = P\left(\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}}\right| \geq z_{\alpha/2}\right)$$

$\sqrt{\frac{\bar{\pi}_1(1-\bar{\pi}_1)}{n} + \frac{\bar{\pi}_2(1-\bar{\pi}_2)}{m}}$を見積値に置き換える

$$\approx P\left(\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_1}\right| \geq z_{\alpha/2}\right)$$

$$\sigma_1 = \sqrt{\frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2)}{m}} = \sqrt{\frac{\kappa\pi_1(1-\pi_1) + (1-\kappa)\pi_2(1-\pi_2)}{n\kappa}}$$

$$= P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_1} \geq z_{\alpha/2}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_1} \leq -z_{\alpha/2}\right)$$

$$= P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \geq z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \leq -z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right)$$

検出力の計算 (2)

$\Phi(\cdot)$: 標準正規分布の累積分布関数

- 対立仮説 $H_1: \varepsilon \neq 0$ のもとで，検出力は$\boldsymbol{n}$の関数になる

$$\phi(n) \approx P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \geq z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \leq -z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right)$$

$$= \left(1 - \Phi\left(z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right)\right) + \Phi\left(-z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right)$$

$$= \Phi\left(\frac{\varepsilon}{\sigma_1} - z_{\alpha/2}\right) + \Phi\left(-z_{\alpha/2} - \frac{\varepsilon}{\sigma_1}\right)$$

$$\approx \Phi\left(\frac{\varepsilon}{\sigma_1} - z_{\alpha/2}\right) \qquad \left(\because \frac{\varepsilon}{\sigma_1} \gg 0\right)$$

サンプルサイズ設計

- 標準正規分布の上側β%点をz_βとする
  - $1-\beta=\Phi(z_\beta)$

- 検出力$1-\beta$を満たすnは，次式の解として与えられる

$$\frac{\varepsilon}{\sigma_1}-z_{\alpha/2}=z_\beta \Leftrightarrow \frac{\varepsilon}{\sqrt{\frac{\kappa\pi_1(1-\pi_1)+(1-\kappa)\pi_2(1-\pi_2)}{n\kappa}}}-z_{\alpha/2}=z_\beta$$

$$\Leftrightarrow n=\frac{\left(z_{\alpha/2}+z_\beta\right)^2}{\varepsilon^2}\frac{\kappa\pi_1(1-\pi_1)+(1-\kappa)\pi_2(1-\pi_2)}{\kappa} \qquad m=rn$$

nに小数点以下の単数が含まれるので，切り上げることで
試験に必要な参加者数を得る。

検出力が$1-\beta$以上となる最小の自然数として得られる。

Conditional approachの場合

- 次式が成立するときに帰無仮説H_0を有意水準αで棄却

  $H_0: \pi_1 = \pi_2\ (= \pi)$ のもとで

$$\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\bar{\pi}(1 - \bar{\pi})}}\right| \geq z_{\alpha/2}$$

$z_{\alpha/2}$：標準正規分布の上側$\alpha/2$％点

- サンプルサイズ設計（検出力の計算）
  - α, r, π_1, π_2はある値に固定する

検出力の計算 (1)

$r = \dfrac{m}{n}$　　$\kappa = \dfrac{r}{r+1}$　　$\dfrac{1-\kappa}{\kappa} = \dfrac{1}{r}$

- 対立仮説 $H_1 : \varepsilon \neq 0$ のもとで，検出力はnの関数になる

$$\phi(n) = P\left(\left| \frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) \bar{\pi}(1 - \bar{\pi})}} \right| \geq z_{\alpha/2} \right)$$

$\bar{\pi}$を見積値　$\pi = \dfrac{n\pi_1 + m\pi_2}{n+m}$　に置き換える

$$\frac{1}{n} + \frac{1}{m} = \frac{1}{n} + \frac{1}{nr} = \frac{1}{n}\left(1 + \frac{1}{r}\right) = \frac{1}{n}\frac{r+1}{r} = \frac{1}{n\kappa}$$

$$\sigma_0^2 = \left(\frac{1}{n} + \frac{1}{m}\right)\pi(1-\pi) = \frac{1}{n\kappa}\pi(1-\pi)$$

$$\sigma_1^2 = \frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2)}{m} = \frac{\kappa\pi_1(1-\pi_1) + (1-\kappa)\pi_2(1-\pi_2)}{n\kappa}$$

検出力の計算 (2)

- 対立仮説 $H_1 : \varepsilon \neq 0$ のもとで，検出力はnの関数になる

$$\phi(n) = P\left(\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\bar{\pi}(1-\bar{\pi})}}\right| \geq z_{\alpha/2}\right)$$

$$\approx P\left(\left|\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_0}\right| \geq z_{\alpha/2}\right)$$

$$= P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_0} \geq z_{\alpha/2}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2}{\sigma_0} \leq -z_{\alpha/2}\right)$$

$$= P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \geq z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \leq -z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right)$$

検出力の計算 (3)

$\Phi(\cdot)$: 標準正規分布の累積分布関数

- 対立仮説 $H_1: \varepsilon \neq 0$ のもとで，検出力はnの関数になる

$$\phi(n) \approx P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \geq z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right) + P\left(\frac{\bar{\pi}_1 - \bar{\pi}_2 - \varepsilon}{\sigma_1} \leq -z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right)$$

$$= \left(1 - \Phi\left(z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right)\right) + \Phi\left(-z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right)$$

$$= \Phi\left(\frac{\varepsilon}{\sigma_1} - z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right)\right) + \Phi\left(-z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\varepsilon}{\sigma_1}\right)$$

$$\approx \Phi\left(\frac{\varepsilon}{\sigma_1} - z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right)\right) \qquad \left(\because \frac{\varepsilon}{\sigma_1} \gg 0\right)$$

サンプルサイズ設計

- 標準正規分布の上側β%点をz_βとする
  - $1-\beta=\Phi(z_\beta)$

- 検出力$1-\beta$を満たすnは，次式の解として与えられる

$$\frac{\varepsilon}{\sigma_1}-z_{\alpha/2}\left(\frac{\sigma_0}{\sigma_1}\right)=z_\beta$$

$$\Leftrightarrow \frac{\varepsilon}{\sqrt{\frac{\kappa\pi_1(1-\pi_1)+(1-\kappa)\pi_2(1-\pi_2)}{n\kappa}}}-z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{\kappa\pi_1(1-\pi_1)+(1-\kappa)\pi_2(1-\pi_2)}}=z_\beta$$

$$\Leftrightarrow n=\frac{1}{\varepsilon^2}\left(z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{\kappa}}+z_\beta\sqrt{\frac{\kappa\pi_1(1-\pi_1)+(1-\kappa)\pi_2(1-\pi_2)}{\kappa}}\right)^2$$

$$m=rn$$