経済統計：期末試験

村澤 康友

2014 年 2 月 8 日

注意：3 問とも解答すること．

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) 帰無仮説 (b) p 値 (c) 回帰 (d) 通常の最小 2 乗法（OLS）
2. （30 点）連続テレビ小説「あまちゃん」の平均視聴率は，関東で 20%，関西で 15% だそうである．関東・関西の母集団における視聴率を p_X, p_Y，視聴率調査における大きさ n_X, n_Y の無作為標本の平均視聴率（＝標本平均）を $\hat{p}_X, \hat{p}_Y$ とする．次の片側検定問題を考える．

$$H_0 : p_X = p_Y \quad \text{vs} \quad H_1 : p_X > p_Y.$$

   （a） $\hat{p}_X, \hat{p}_Y$ の漸近分布を求めなさい．また $\hat{p}_X - \hat{p}_Y$ の漸近分布を求めなさい．
   （b） 検定統計量を与えなさい．それは H_0 の下でどのような分布に近似的に従うか？
   （c） 有意水準 5 %の検定の棄却域を定め，$n_X = n_Y = 625$ として検定を実行しなさい．
3. （50 点）ゴルトンは身長の遺伝を研究した．父親と息子の身長の無作為標本を $((x_1, y_1), \ldots, (x_n, y_n))$ とする（単位はインチ）．y_i の x_i 上への単回帰モデルは

$$\mathrm{E}(y_i|x_i) = \alpha + \beta x_i.$$

   回帰分析の結果は以下の通りであった．

```
Model 1: OLS, using observations 1-928
Dependent variable: child
             coefficient   std. error
  -----------------------------------
  const        23.9415      2.81088
  parent        0.646291    0.0411359
```

   β の OLS 推定量を b，その標準誤差を s とする．また $b \overset{a}{\sim} \mathrm{N}\left(\beta, s^2\right)$ とみなしてよい．
   （a） b の値は幾らか？ s の値は幾らか？
   （b） β の 95 %信頼区間を求めなさい．
   （c） β の t 値を求めなさい．
   （d） 身長の遺伝の有無の検定問題を定式化しなさい（問題意識を踏まえること）．
   （e） 有意水準 5 %の検定の棄却域を定め，検定を実行しなさい．

解答例

1. 統計学の基本用語
   (a) とりあえず真と想定する仮説.
   (b) H_0 の下で検定統計量が実現値以上になる確率.
      - 「H_0 の下で」がなければ 2 点.
   (c) $\mathrm{E}(Y|X)$ を求めることを，Y の X 上への回帰という.
      - 「$\mathrm{E}(Y|X)$」のみは説明不足なので 2 点.
   (d) 残差 2 乗和を最小にするように回帰係数を定める方法.
      - $\min_b \sum_{i=1}^n (y_i - bx_i)^2$ でも OK．ただし $\min_b$ がなければ 0 点.
2. 2 標本の母比率の差の検定
   (a)
$$\begin{aligned}\hat{p}_X &\overset{a}{\sim} \mathrm{N}\left(p_X, \frac{p_X(1-p_X)}{n_X}\right),\\ \hat{p}_Y &\overset{a}{\sim} \mathrm{N}\left(p_Y, \frac{p_Y(1-p_Y)}{n_Y}\right).\end{aligned}$$
   両者は独立なので
$$\hat{p}_X - \hat{p}_Y \overset{a}{\sim} \mathrm{N}\left(p_X - p_Y, \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right).$$
      - 各 5 点.
   (b) 検定統計量は
$$Z := \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n_X + \hat{p}_Y(1-\hat{p}_Y)/n_Y}}.$$
   H_0 の下で $Z \overset{a}{\sim} \mathrm{N}(0,1)$.
      - 検定統計量で 5 点．分布で 5 点.
      - 検定統計量に未知母数があったらダメ.
   (c) 棄却域は $[1.645, \infty)$．検定統計量の値は
$$\begin{aligned}Z &:= \frac{.2 - .15}{\sqrt{.2(1-.2)/625 + .15(1-.15)/625}}\\ &= \frac{.05}{\sqrt{(.16 + .1275)/625}}\\ &= \frac{.05 \cdot 25}{\sqrt{.2875}}\\ &= \frac{1.25}{\sqrt{.2875}}\\ &\approx 2.33.\end{aligned}$$
   したがって H_0 を棄却して H_1 を採択.
      - 棄却域で 5 点．検定統計量で 5 点.
3. 回帰分析
   (a) $b = .646291$，$s = .0411359$.

- 各 5 点.

(b) $b \overset{a}{\sim} \mathrm{N}(\beta, s^2)$ より

$$\frac{b-\beta}{s} \overset{a}{\sim} \mathrm{N}(0,1).$$

したがって

$$\Pr\left[-1.96 \le \frac{b-\beta}{s} \le 1.96\right] \approx .95,$$

または

$$\Pr[-1.96s \le b-\beta \le 1.96s] \approx .95,$$

または

$$\Pr[b-1.96s \le \beta \le b+1.96s] \approx .95.$$

したがって β の 95 %信頼区間は $[.566, .727]$.

(c)

$$\begin{aligned} t &= \frac{b}{s} \\ &= \frac{.646291}{.0411359} \\ &\approx 15.71. \end{aligned}$$

(d)

$$H_0 : \beta = 0 \ (\alpha \in \mathbb{R},\, \sigma^2 > 0) \quad \text{vs} \quad H_1 : \beta > 0 \ (\alpha \in \mathbb{R},\, \sigma^2 > 0).$$

- $H_1 : \beta \neq 0$ は 5 点. 遺伝は同方向に働く.

(e) 棄却域は $[1.645, \infty)$. t 値が棄却域に入るので H_0 を棄却して H_1 を採択. すなわち身長は遺伝すると言える.

- 棄却域で 5 点. 検定で 5 点.
- 検定問題と整合的なら OK.