中級統計学：定期試験

村澤 康友

2021 年 7 月 27 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) 統計的仮説 (b) t 検定 (c) p 値 (d)（回帰の）誤差項
2. （30 点）ベルヌーイ母集団 $\mathrm{Bin}(1,p)$ から抽出した大きさ n の無作為標本の標本比率を $\hat{p}$ とする．次の片側検定問題を考える．
$$H_0 : p = \frac{1}{2} \quad \text{vs} \quad H_1 : p > \frac{1}{2}$$
   （a）検定統計量を与えなさい．
   （b）有意水準 5 %の検定の棄却域を定めなさい．
   （c）$n = 100$，$\hat{p} = .6$ として検定統計量の値と p 値を求めなさい．
3. （50 点）高校での成績が大学での成績の予測に役立つかどうかを検証したい．そこで大学での GPA（colgpa）を高校での GPA（hsgpa）と英語（vsat）・数学（msat）の入試得点で説明する重回帰モデルを推定し，次の結果を得た．

モデル 1: **最小二乗法**（OLS），**観測**: 1-427
従属変数: colgpa

	係数	**標準誤差**
const	0.423249	0.219749
hsgpa	0.398349	0.0605865
vsat	0.000737453	0.000280682
msat	0.00101521	0.000293603

高校での GPA の偏回帰係数を β 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）β の点推定値・標準誤差・t 値は幾らか？
（c）古典的正規線形回帰モデルを仮定する．β の t 値は H_0 の下でどのような分布に従うか？
（d）有意水準 5 %の検定の棄却域を定め，検定を実行しなさい．
（e）β の 95 %信頼区間を求めなさい（漸近分布で近似してよい）．

解答例

1. 統計学の基本用語
   (a) 母集団分布に関する仮説.
   (b) t 統計量を用いる検定.
   (c) H_0 の下で検定統計量が実現値以上になる確率.
   (d) $U := Y - \mathrm{E}(Y|X)$.
2. 母比率の検定
   (a) $\mathrm{Bin}(1,p)$ の平均は p，分散は $p(1-p)$ なので，中心極限定理より
   $$\hat{p} \overset{a}{\sim} \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$
   標準化すると
   $$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$
   検定統計量は
   $$Z := \frac{\hat{p}-1/2}{\sqrt{(1/2)(1-1/2)/n}}$$
   - $H_0 : p = 1/2$ を代入しなければ 0 点.

   (b) 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
   $$\Pr[Z \geq 1.65] \approx .05$$
   したがって近似的な棄却域は $[1.65, \infty)$.
   (c) $n = 100$，$\hat{p} = .6$ なら
   $$\begin{aligned} Z &:= \frac{.6-1/2}{\sqrt{(1/2)(1-1/2)/100}} \\ &= \frac{.1}{\sqrt{1/400}} \\ &= 2 \end{aligned}$$
   標準正規分布表より p 値は.022750.
   - 検定統計量の値で 5 点，p 値で 5 点.
3. 回帰分析
   (a)
   $$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0$$
   - 予測に役立つかどうかの検証なら両側検定.

   (b) 点推定値は.398349，標準誤差は.0605865，t 値は.398349/.0605865=6.575.
   - 点推定値と標準誤差は各 3 点，t 値は 4 点.

   (c) t 値は H_0 の下で t(423) にしたがう.

（d）t 分布表より H_0 の下で

$$\Pr[|t| \geq 1.960] \approx .05$$

したがって近似的な棄却域は $(-\infty, -1.960] \cup [1.960, \infty)$．$t = 6.575$ より H_0 を棄却して H_1 を採択.

（e）β の OLS 推定量を b，その標準誤差を s とすると，$b \overset{a}{\sim} \mathrm{N}\left(\beta, s^2\right)$ より

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

したがって

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

または

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

または

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

$b = .398349$，$s = .0605865$ より β の 95 %信頼区間は $[.279261, .517437]$.

- $[b-1.96s, b+1.96s]$ で 5 点.