

問題集 3 (後半)

北川梨津*

2022年5月9日

- 答えだけではなく、結果に至るまでの過程も示すこと。
- 手書きの答案は不可¹⁾。
- 締め切り：2022年5月8日 22:00

問 1 統計的仮説検定の理解 1

グループ A とグループ B におけるアウトカムについて、平均の差の検定を行った²⁾。すると、 $p = 0.04$ であった。グループ A, B の母平均をそれぞれ μ_A, μ_B とする。この結果の解釈として正しいものを、次から**すべて**選べ。正しいものがないときは、その旨を記して、正しい解釈を記述すること。

1. 帰無仮説 $H_0 : \mu_A = \mu_B$ が正しい確率は 4% であり、それは極めて稀であるので、帰無仮説は棄却される。
2. 帰無仮説 $H_0 : \mu_A = \mu_B$ が正しいとすると、今回の値が生じる確率は 4% であり、それは極めて稀であるので、帰無仮説は棄却される。
3. 帰無仮説 $H_0 : \mu_A = \mu_B$ であり、対立仮説が $H_1 : \mu_A \neq \mu_B$ であるところ、 $p = 0.04$ であれば有意水準 $\alpha = 0.05$ を下回るため、 $100 \times (1 - \alpha)\% = 95\%$ の確率で対立仮説が正しい。
4. 帰無仮説 $H_0 : \mu_A = \mu_B$ が正しいとすると、今回の値と同じかそれ以上に極端な結果が生じる確率は 4% であり、それは極めて稀であるので、帰無仮説は棄却される。

* 早稲田大学：ritsu.kitagawa@fuji.waseda.jp

1) Word の数式モードや L^AT_EX を使うこと。前者の使い方については、例えば「word で簡単に数式を書く」という記事 (<https://note.com/keisemi/n/na12bfeb77469>) を参考にするとよい。後者については、Cloud LaTeX というサービス (<https://cloudlatex.io/ja>) を使うのが最も簡便である。

2) 状況に適切な検定統計量を用いたと仮定してよい。

問 2 統計的仮説検定の理解 2

グループ A とグループ B におけるアウトカムについて、平均の差の検定を行った³⁾。すると、 $p = 0.06$ であった。グループ A, B の母平均をそれぞれ μ_A, μ_B とする。この結果の解釈として正しいものを、次からすべて選べ。正しいものがないときは、その旨を記して、正しい解釈を記述すること。

1. p 値が有意水準 $\alpha = 0.05$ よりも大きく、したがって帰無仮説 $H_0 : \mu_A = \mu_B$ はされず棄却されずに採択されるため、グループ A とグループ B の母平均に差はないと言える。
2. p 値が有意水準 $\alpha = 0.05$ よりも大きく、したがって帰無仮説は棄却 $H_0 : \mu_A = \mu_B$ されないため、グループ A とグループ B の母平均の差については何も言えない。
3. p 値が有意水準 $\alpha = 0.05$ よりも大きいがかかなり近い値なので、サンプルサイズをもう少し増やせば有意になるはずであるから、サンプルサイズを増やせないか検討するべきだ。
4. p 値が有意水準 $\alpha = 0.05$ よりも大きいがかかなり近い値なので、サンプルに含まれる測定バイアスを取り除けば有意になるかもしれないため、サンプルからランダムに選んだ観察値を取り除くべきだ。

3) 状況に適切な検定統計量を用いたと仮定してよい。

問 3 統計的仮説検定の理解 3

サンプルから母平均の推定を行ったところ、その推定値に関する 95% 信頼区間が [1.56, 3.98] と計算された。この結果の解釈として正しいものを、次からすべて選べ。正しいものがないときは、その旨を記して、正しい解釈を記述すること。

1. 真の値（母平均）が信頼区間 [1.56, 3.98] に含まれる確率は 95% である。
2. 信頼区間 [1.56, 3.98] はゼロを含んでいないので、真の値（母平均）がゼロではない確率は 95% 以上であると言える。
3. 真の値（母平均）が信頼区間 [1.56, 3.98] に含まれる確率は 0% か 100% である。
4. 信頼区間 [1.56, 3.98] はゼロを含んでいないので、統計学的に有意である。

問 4 統計的仮説検定の理解 4（第 105 回 医師国家試験より改題）

ある疾病を患う入院患者を無作為に処置群と統制群に割り当て、処置群には新薬を投与し、統制群には偽薬を投与して、入院期間をアウトカムとして新薬の効果を測定した。その結果を以下の表に示す。

	処置群	統制群	P 値
観察数	198 人	201 人	
入院期間（平均）	8.1 日	9.6 日	0.036

この結果の解釈として正しいものを次から選べ。

1. 処置群は統制群に比べて 3.6%入院期間が短い。
2. 処置群の方が統制群に比べて入院期間が短くなる確率は 3.6%である。
3. 平均の差の誤差は 3.6%以下である。
4. 処置群と統制群とで入院期間の平均に差がないのに誤って差があるとする確率は 3.6%である。

問5 平均処置効果と処置群平均処置効果

以下の問題に答えよ。表記は教科書 (*The Mixtape*) に従う。

1. 次の等式が正しいことを示せ。

$$ATE = ATT - (1 - \pi)(ATT - ATU).$$

(ヒント： $ATE = \pi ATT + (1 - \pi)ATU$ であることを利用するとよい。)

2. セレクションバイアスがゼロであるとき、SDO が ATT に等しいことを示せ。つまり、

$$E[Y^0 | D = 1] - E[Y^0 | D = 0] = 0 \Rightarrow SDO = ATT,$$

であることを示せ。

3. $Y^0 \perp\!\!\!\perp D$ であるとき、SDO は ATE, ATT, ATC のそれぞれを識別できるかどうかを答え、その主張が正しいことを示せ。ただし、 $Y^1 \perp\!\!\!\perp D$ とは限らないことに注意すること。またそのような状況（
4. $Y^0 \perp\!\!\!\perp D$ だが $Y^1 \perp\!\!\!\perp D$ ではない）の例を挙げよ。
5. $Y^1 \perp\!\!\!\perp D$ であるとき、SDO は ATE, ATT, ATC のそれぞれを識別できるかどうかを答え、その主張が正しいことを示せ。ただし、 $Y^0 \perp\!\!\!\perp D$ とは限らないことに注意すること。またそのような状況（
6. $Y^1 \perp\!\!\!\perp D$ だが $Y^0 \perp\!\!\!\perp D$ ではない）の例を挙げよ。
7. $(Y^0, Y^1) \perp\!\!\!\perp D$ であるとき、SDO は ATE, ATT, ATC のそれぞれを識別できるかどうかを答え、その主張が正しいことを示せ。
8. 教科書では、処置効果の異質性によるバイアスよりもセレクションバイアスの方が問題적であるという旨の記述がある。問4の3, 4, 5を踏まえて、なぜ問題적なのか、その理由を述べよ。

問 6 Potential outcomes model と回帰分析 (任意)

Potential outcomes model (POM) と回帰分析の関係を理解するために、以下の問いに答えよ。

1. 定義から、 $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$ であることが確認できる。この等式を Y_i を従属変数、 D_i を独立変数とする線形単回帰モデルとして次のように書き直すことができる。

$$Y_i = \alpha + \rho D_i + \varepsilon_i.$$

ただし、ここでは $Y_i^1 - Y_i^0 = \rho, \forall i$ と仮定している。このとき、パラメータ α, ρ と誤差項 ε_i が何に相当するかを POM の表記を使って説明せよ。

2. 上の回帰モデルを用いて、潜在アウトカム Y_i^0 と Y_i^1 をそれぞれ表現せよ。
3. ゼミでは、 $SDO = \mathbf{E}[Y_i^1 - Y_i^0 \mid D_i = 1] + \mathbf{E}[Y_i^0 \mid D_i = 1] - \mathbf{E}[Y_i^0 \mid D_i = 0]$ であることを確認した。このとき、 $(Y_i^0, Y_i^1) \perp\!\!\!\perp D_i$ が満たされるとすると、 $\mathbf{E}[Y_i^1 - Y_i^0 \mid D_i = 1] = \mathbf{E}[Y_i^1 - Y_i^0] = ATE$ であり、また、 $\mathbf{E}[Y_i^0 \mid D_i = 1] - \mathbf{E}[Y_i^0 \mid D_i = 0] = 0$ であり、したがって、 $SDO = ATE$ となることも確認した。同様のことを回帰モデルの枠組みで示せ。
4. 次に、 $Y_i^1 - Y_i^0 = \rho, \forall i$ の仮定をゆるめて、個人について処置効果が異なることを許容して (i.e., $Y_i^1 - Y_i^0 = \rho_i$) , 1, 2, 3 を解け。
5. ゼミでは、回帰分析における文脈では、誤差項の (説明変数についての) 条件付き期待値がゼロであることが OLS 推定量が不偏推定量であるための重要な条件であることを確認した。条件付き期待値ゼロの仮定のもとで、ATE, ATT, ATU のそれぞれが識別されるかを答えよ。

以上