第 IX 部　母平均の差の検定

目次

6　2 つの母平均の差の検定

6.1　はじめに

- 異なる二つの母集団から得られたデータがあります。
- 母集団の分散が分かっている場合、母集団の平均が等しいかどうかを考えます。

6.1.1　ポイント

- $\mathsf{E}(\bar{X}-\bar{Y})$
- $\mathsf{V}(\bar{X}-\bar{Y})$
- $\bar{X}-\bar{Y}$ の標準化

6.1.2 標準化

『平均からの偏差』を『標準偏差』で割ることを**『標準化』**といい、標準化された値を z であらわす。μ_x を平均とし、σ_x を標準偏差とすると、標準化された確率変数 z は

$$z = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

である。標準化すると z の平均は 0, 標準偏差は 1 になる。

6.1.3　正規分布

確率変数 $X=x$ が標準正規分布 $Z=z$ を用いて

$$x=\mu+\sigma z \qquad ;\text{ただし}\sigma\neq 0 \tag{2}$$

とあらわされるとき X は『**平均 $\boldsymbol{\mu}$, 分散 $\boldsymbol{\sigma^2}$ の正規分布**』であるといい $N\left(\mu,\ \sigma^2\right)$ であらわす。

X が正規分布 $N\left(\mu,\ \sigma^2\right)$ に従うとき、$X=x$ を標準化した

$$z=\frac{x-\mu}{\sigma} \tag{3}$$

は標準正規分布 $N(0,\ 1)$ に従う。

6.1.4 信頼区間

標準正規分布において z が

$$-1.96 \leqq z \leqq 1.96 \tag{4}$$

の区間に収まる確率は

$$\Pr(-1.96 \leqq z \leqq 1.96) = 0.950\,004\cdots \tag{5}$$

である。

(4) の区間を『**95% 信頼区間**』という。

6.1.5　仮説検定

前提となる状況を『**帰無仮説**』という。帰無仮説の下で稀な現象が発生した場合、統計学では「帰無仮説は妥当ではなく、帰無仮説が妥当ではない状況において普通の現象が観測された。」と考える。

- 帰無仮説が妥当ではないと判断することを、帰無仮説を『**棄却する**』という。
- 帰無仮説が妥当であると判断することを、帰無仮説を『**受容する**』または『**棄却しない**』という[1)]。
- 帰無仮説の妥当性に対し統計的基準を用いて判断することを『**統計的検定**』という。
- 妥当性に対する基準を『**有意水準**』という。

1)　通常、帰無仮説は棄却することが前提なので、妥当だと判断することはない。

6.1.6 有意水準

- 帰無仮説が正しいと仮定したときの観測された現象の珍しさの程度を『**有意水準**』といい α を使ってあらわされることが多い。
- 有意水準 α は正値であり、5%, 1% がよく使われる。
- $\Pr(Z \geqq t) \leqq \alpha$ を満たす最小の t を『**臨界値**』という。
- 標準正規分布において $\alpha = 5\%$ のとき、臨界値はおよそ $t = 1.96$ である。
- 信頼区間の外側の区間を『**棄却域**』という。
- $X = x \sim N\left(\mu,\ \sigma^2\right)$ の $\alpha = 5\%$ の棄却域 w は

$$w = \left\{x \middle| x < \mu - 1.96\sigma, x > \mu + 1.96\sigma\right\}$$

である。

6.1.7　大数の法則

母平均を μ, 母分散を σ^2 とする。この母集団からサンプル・サイズ n の標本を復元抽出する。標本平均を $\bar{X}$ とすると、標本平均の期待値と分散は

$$\mathsf{E}\left(\bar{X}\right) = \mu \tag{6}$$

$$\mathsf{V}\left(\bar{X}\right) = \frac{\sigma^2}{n} \tag{7}$$

になる。

6.1.8　独立な X と Y の和の期待値と分散

X と Y を独立な確率変数とし、$a,\ b$ を任意の定数とする。

$$\begin{aligned} \mathsf{E}\left(aX+bY\right) &= \mathsf{E}\left(aX\right)+\mathsf{E}\left(bY\right) && (8)\\ &= a\mathsf{E}\left(X\right)+b\mathsf{E}\left(Y\right) && (9)\\ \mathsf{V}\left(aX+bY\right) &= \mathsf{V}\left(aX\right)+\mathsf{V}\left(bY\right) && (10)\\ &= a^2\mathsf{V}\left(X\right)+b^2\mathsf{V}\left(Y\right) && (11) \end{aligned}$$

問題 IX − 6 − 1

X と Y を独立な確率変数とし、$a,\ b$ を任意の定数とする。

$aX - bY$ の期待値と分散を求めなさい。

解例 IX − 6 − 1

$$
\begin{aligned}
\mathsf{E}\left(aX - bY\right) &= \mathsf{E}\left(aX + (-b)Y\right) \\
&= \mathsf{E}\left(aX\right) + \mathsf{E}\left((-b)Y\right) \\
&= a\mathsf{E}\left(X\right) - b\mathsf{E}\left(Y\right) \\
\mathsf{V}\left(aX - bY\right) &= \mathsf{V}\left(aX + (-b)Y\right) \\
&= \mathsf{V}\left(aX\right) + \mathsf{V}\left((-b)Y\right) \\
&= a^2\mathsf{V}\left(X\right) + b^2\mathsf{V}\left(Y\right)
\end{aligned}
$$

6.1.9　異なる 2 つの正規母集団から抽出された確率変数

2 つの独立な確率変数 X と Y を考える。

$N(\mu_x,\ {\sigma_x}^2)$ に従う正規母集団から抽出された n 個の確率変数を X とする。X の要素を識別する添え字に i を採用し x_i とあらわす。$1 \leqq i \leqq n$ である。

$$X = \{x_1,\ x_2,\ \cdots\ ,\ x_n\} \tag{12}$$

$N(\mu_y,\ {\sigma_y}^2)$ に従う正規母集団から抽出された m 個の確率変数を Y とする。Y の要素を識別する添え字に j を採用し y_j とあらわす。$1 \leqq j \leqq m$ である。

$$Y = \{y_1,\ y_2,\ \cdots\ ,\ y_m\} \tag{13}$$

ここで、X と Y の差 $X - Y$ は正規分布に従うことが知られている。

任意の定数 a と b を用いて、

$$aX - bY \tag{14}$$

を作る。

$aX - bY$ の期待値は

$$\begin{aligned}\mathsf{E}\left(aX - bY\right) &= \mathsf{E}\left(aX\right) - \mathsf{E}\left(bY\right) &(15)\\ &= a\mathsf{E}\left(X\right) - b\mathsf{E}\left(Y\right) &(16)\\ &= a\mu_x - b\mu_y &(17)\end{aligned}$$

であって、

分散は

$$
\begin{aligned}
\mathsf{V}\left(aX - bY\right) &= \mathsf{V}\left(aX\right) + \mathsf{V}\left(-bY\right) && (18)\\
&= a^2\mathsf{V}\left(X\right) + b^2\mathsf{V}\left(Y\right) && (19)\\
&= a^2{\sigma_x}^2 + b^2{\sigma_y}^2 && (20)
\end{aligned}
$$

である。

よって $aX - bY$ は、$N\left(a\mu_x - b\mu_y,\ a^2{\sigma_x}^2 + b^2{\sigma_y}^2\right)$ に従う。

$$
aX - bY \sim N\left(a\mu_x - b\mu_y,\ a^2{\sigma_x}^2 + b^2{\sigma_y}^2\right) \tag{21}
$$

6.2 $\boldsymbol{\sigma}$ が既知の場合の母平均の差の検定

$$\bar{X} = \frac{1}{n} \sum_{x_i \in A}^{n} x_i \tag{22}$$

$$\bar{Y} = \frac{1}{m} \sum_{y_j \in B}^{n} y_j \tag{23}$$

とする。

$\bar{X} - \bar{Y}$ の期待値は

$$\mathsf{E}\left(\bar{X} - \bar{Y}\right) = \mathsf{E}\left(\bar{X}\right) - \mathsf{E}\left(\bar{Y}\right) \tag{24}$$

$$= \mu_x - \mu_y \tag{25}$$

である。

そして分散は、

$$\mathsf{V}\left(\bar{X}-\bar{Y}\right)=\mathsf{V}\left(\bar{X}\right)+\mathsf{V}\left(\bar{Y}\right) \tag{26}$$

$$=\mathsf{V}\left(\frac{1}{n}\sum_{x_i\in A}^{n}x_i\right)+\mathsf{V}\left(\frac{1}{m}\sum_{y_j\in B}^{m}y_j\right) \tag{27}$$

$$=\frac{1}{n^2}\mathsf{V}\left(\sum_{x_i\in A}^{n}x_i\right)+\frac{1}{m^2}\mathsf{V}\left(\sum_{y_j\in B}^{m}y_j\right) \tag{28}$$

$$=\frac{\mathsf{V}\left(x_1\right)+\mathsf{V}\left(x_2\right)+\cdots+\mathsf{V}\left(x_n\right)}{n^2}$$
$$+\frac{\mathsf{V}\left(y_1\right)+\mathsf{V}\left(y_2\right)+\cdots+\mathsf{V}\left(y_m\right)}{m^2} \tag{29}$$

$$=\frac{n{\sigma_x}^2}{n^2}+\frac{m{\sigma_y}^2}{m^2} \tag{30}$$

$$=\frac{{\sigma_x}^2}{n}+\frac{{\sigma_y}^2}{m} \tag{31}$$

である。

よって

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y,\ \frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}\right) \tag{32}$$

である。仮定により母分散 ${\sigma_x}^2$, ${\sigma_y}^2$ が分かっているので、標準化すると

$$w = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}}} \sim N(0,\ 1) \tag{33}$$

である。

6.2.1 帰無仮説と対立仮説

ここで検定する帰無仮説 H_0 は、『それぞれの母平均には差がない』であり、

$$H_0 : \quad \mu_x - \mu_y = 0 \tag{34}$$

対立仮説は

$$H_1 : \quad \mu_x - \mu_y \neq 0 \tag{35}$$

である。

帰無仮説 H_0 の下では、

$$\mu_x - \mu_y = 0 \tag{36}$$

なので

$$w_0 = \frac{\bar{X} - \bar{Y} - (0)}{\sqrt{\frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}}} \tag{37}$$

は標準正規分布 $N(0,\ 1)$ に従う。

$N(0,\ 1)$ の 95% 信頼区間の臨界値は 1.96 なので

$$-1.96 \leqq w_0 \leqq 1.96 \tag{38}$$

の範囲にない場合は H_0 を棄却し、範囲内の時は H_0 を棄却しない。

$\boldsymbol{\sigma}$ が既知の場合の母平均の差の検定

仮説

$$H_0: \quad \mu_x - \mu_y = 0 \tag{34}$$

$$H_1: \quad \mu_x - \mu_y \neq 0 \tag{35}$$

検定統計量

$$w = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}}} \tag{39}$$

H_0 の下で

$$w_0 = \frac{\bar{X} - \bar{Y} - (0)}{\sqrt{\frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}}} \tag{37}$$

は $N(0,\ 1)$ に従う。そこで有意水準 $\alpha = 5\%$ とすると、

- $|w_0| \geqq 1.96$ のとき　H_0 を棄却する。
- $|w_0| < 1.96$ のとき　H_0 を棄却しない。

6.3　まとめ

- 独立な 2 つの確率変数 $X,\ Y$ の差 $X-Y$ の期待値と分散は

$$\mathsf{E}\left(X-Y\right)=\mathsf{E}\left(X\right)-\mathsf{E}\left(Y\right)$$
$$\mathsf{V}\left(X-Y\right)=\mathsf{V}\left(X\right)+\mathsf{V}\left(Y\right)$$

- $X\sim N(\mu_x,\ {\sigma_x}^2)$ ， $Y\sim N(\mu_y,\ {\sigma_y}^2)$ ならば

$$X-Y\sim N\left(\mu_x-\mu_y,\ {\sigma_x}^2+{\sigma_y}^2\right)$$

- すると、$\bar{X}-\bar{Y}$ は

$$\bar{X}-\bar{Y}\sim N\left(\mu_x-\mu_y,\ \frac{{\sigma_x}^2}{n}+\frac{{\sigma_y}^2}{m}\right)$$

なので、この分布を使って $\bar{X}-\bar{Y}$ を標準化し、

$$H_0:\quad \mu_x-\mu_y=0$$

を検定するのが『**$\boldsymbol{\sigma}$ が既知のときの母平均の差の検定**』である。

参考文献

1.『経済数学教室 1 巻』小山昭雄著「岩波書店」1994 年 5 月 30 日
2.『新体系 大学数学 入門の教科書 上下』吉沢光雄著「講談社」2022 年 6 月 16 日
3.『技術者のための高等数学 7 確率と統計（原書代 8 版)』E・クライツィグ著　田栗正章訳「培風館」2004 年 11 月 30 日第 8 版
4.『データ解析のための数理統計入門』久保川達也著「共立出版」2023 年 10 月 15 日初版
5.『統計学入門』東京大学教養学部統計学教室篇「東京大学出版会」1991 年 7 月 9 日初版
6.『行動科学における統計解析法」芝祐順・南風原朝和著　「東京大学出版会」1990 年 3 月 20 日　初版
7.『計量経済学の理論』A.S. ゴールドバーガー著, 福地崇生・森口親司訳「東洋経済新報社」

1970 年 9 月 10 日 第 1 刷

8.『計量経済学序説』R.J. ウォナコット・T.H. ウォナコット著, 国府田恒夫・田中一盛訳「培風館」 1975 年 6 月 10 日初版
9.『マーケティング・サイエンス』片平秀貴著 「東京大学出版会」1987 年 4 月 20 日初版
10.『回帰分析』佐和光男著 「朝倉書店」1979 年 4 月 20 日初版
11.『完全独習　統計学入門』小島寛之著「ダイヤモンド社」2006 年 9 月 28 日　初版
12.『分散分析の基礎』高橋敬子著「プレアデス出版」2009 年 10 月 1 日