第 IX 部　母平均の差の検定

目次

2　正規分布

2.1　はじめに

- 様々な対象からデータをとると左右対称の釣り鐘型の分布をしていることがよく観測されます。
- また、二項分布において N を大きくすると同様の形に近づきます。
- この釣り鐘型の確率密度関数をもつ分布を正規分布といいます。

2.1.1　ポイント

- 組合せの数
- 二項分布
- 正規分布
- 仮説検定

2.1.2　ネイピア数

定義

以下で定義される無理数を e であらわし『**ネイピア数**』という。

$$\mathrm{e} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \tag{1}$$

ネイピア数 e は底とすることが多い。そこで

$$\mathrm{e}^x = \exp(x) \tag{2}$$

とあらわす。

ネイピア数は

$$\mathrm{e} = 2.718\,281\,828\,459\,045\,235\,36\cdots \tag{3}$$

とつづく無理数である。

2.1.3　順列

定義

異なったものを 1 列に並べた配列を『**順列**』という。n 個の異なったものから r 個を選んで並べるとき、異なる順列の個数を

$$_nP_r \tag{4}$$

であらわす。

$$_nP_r = n \times (n-1) \times \cdots \times (n-r+1) \tag{5}$$

である。特に

$$_nP_n = n \times (n-1) \times \cdots \times 3 \times 2 \times 1 \tag{6}$$

のとき『**$\boldsymbol{n}$ の階乗**』といい、

$$n! \tag{7}$$

であらわす。

$n=0$ のとき

$$0! = 1 \tag{8}$$

と定義する。これらの記号を使うと

$$ {}_nP_r = \frac{n!}{(n-r)!} \tag{9}$$

である。

問題 IX − 2 − 1

以下の値を求めなさい。

(i) ${}_8P_0$

(ii) ${}_8P_1$

(iii) ${}_8P_2$

(iv) ${}_8P_3$

(v) ${}_8P_4$

(vi) ${}_8P_5$

(vii) ${}_8P_6$

(viii) ${}_8P_7$

(ix) ${}_8P_8$

解例 IX − 2 − 1

(i) $\quad {}_8P_0 = \dfrac{8!}{8!} = 1$

(ii) $\quad {}_8P_1 = \dfrac{8!}{7!} = 8$

(iii) $\quad {}_8P_2 = \dfrac{8!}{6!} = 8 \times 7 = 56$

(iv) $\quad {}_8P_4 = \dfrac{8!}{5!} = 8 \times 7 \times 6 = 336$

(v) $\quad {}_8P_4 = \dfrac{8!}{4!} = 8 \times 7 \times 6 \times 5 = 1680$

(vi) $\quad {}_8P_5 = \dfrac{8!}{3!} = 8 \times 7 \times 6 \times 5 \times 4 = 6720$

(vii) $\quad {}_8P_6 = \dfrac{8!}{2!} = 8\times7\times6\times5\times4\times3 = 20\,160$

(viii) $\quad {}_8P_7 = \dfrac{8!}{1!} = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2$
$= 40\,320$

(ix) $\quad {}_8P_8 = \dfrac{8!}{0!} = 8! = 40\,320$

2.1.4 組合せ

定義

n 個の異なったものから r 個を選びだす選び方を『**組合せ**』といい、異なる組合せの総数を

$$ {}_nC_r \tag{10} $$

であらわす。

$r = 0$ のとき

$$ {}_nC_0 = 1 \tag{11} $$

と定義する。

順列の数 ${}_nP_r$ は n 個の異なるものから r 個とってきて、並べる総数である。n 個の異なったものから r 個を選びだす組合せの数は ${}_nC_r$ であり、選び出した r 個を全て並び替え

ると、${}_rP_r$ である。よって

$$ {}_nP_r = {}_nC_r \times {}_rP_r \tag{12} $$

$$ \frac{{}_nP_r}{{}_rP_r} = {}_nC_r \tag{13} $$

これを階乗を使ってあらわすと

$$ {}_nC_r = \frac{n!}{r!\,(n-r)!} \tag{14} $$

である。

問題 IX − 2 − 2

次の値を求めなさい。

(i) ${}_8C_0$

(ii) ${}_8C_1$

(iii) ${}_8C_2$

(iv) ${}_8C_3$

(v) ${}_8C_4$

(vi) ${}_8C_5$

(vii) ${}_8C_6$

(viii) ${}_8C_7$

(ix) ${}_8C_8$

解例 IX − 2 − 2

(i) $${}_8C_0 = \frac{8!}{8!0!} = 1$$

(ii) $${}_8C_1 = \frac{8!}{7!1!} = 8$$

(iii) $${}_8C_2 = \frac{8!}{6!2!} = 28$$

(iv) $${}_8C_3 = \frac{8!}{5!3!} = 56$$

(v) $${}_8C_4 = \frac{8!}{4!4!} = 70$$

(vi) $${}_8C_5 = \frac{8!}{3!5!} = 56$$

(vii) $${}_8C_6 = \frac{8!}{2!6!} = 28$$

(viii) $${}_8C_7 = \frac{8!}{1!7!} = 8$$

(ix) $${}_8C_8 = \frac{8!}{0!8!} = 1$$

2.2　二項分布

定義

ある試行において事象 E の起こる確率を $\Pr(E) = p$、その補事象の確率を $\Pr\left(E^C\right) = q$ とする。この試行を独立に n 回繰り返すとき、事象 E の起こる回数を X とすれば、X の確率分布は

$$\Pr\left(X = x\right) = {}_nC_x\ p^x q^{n-x} \tag{15}$$

である。この確率分布を、確率 p に対する次数 n の『**二項分布**』といい、$B(n, p)$ であらわす。

$B(n, p)$ の期待値, 分散ならびに標準偏差は

$$\mathsf{E}(X) = np \tag{16}$$

$$\mathsf{V}(X) = npq \tag{17}$$

$$\mathsf{D}(X) = \sqrt{npq} \tag{18}$$

である。

問題 IX − 2 − 3

X は $B\left(8,\ \dfrac{1}{2}\right)$ に従う確率変数とする。以下の確率を求めなさい。

(i) $\Pr(X = 0)$

(ii) $\Pr(X = 1)$

(iii) $\Pr(X = 2)$

(iv) $\Pr(X = 3)$

(v) $\Pr(X = 4)$

(vi) $\Pr(X = 5)$

(vii) $\Pr(X = 6)$

(viii) $\Pr(X = 7)$

(ix) $\Pr(X = 8)$

解例 IX − 2 − 3

(i) $\Pr(X=0)=\dfrac{1}{256}$

(ii) $\Pr(X=1)=\dfrac{8}{256}$

(iii) $\Pr(X=2)=\dfrac{28}{256}$

(iv) $\Pr(X=3)=\dfrac{56}{256}$

(v) $\Pr(X=4)=\dfrac{70}{256}$

(vi) $\Pr(X=5)=\dfrac{56}{256}$

(vii) $\Pr(X=6)=\dfrac{28}{256}$

(viii) $\Pr(X=7)=\dfrac{8}{256}$

(ix) $\Pr(X=8)=\dfrac{1}{256}$

問題 IX − 2 − 4

X は $B\left(8,\ \dfrac{1}{2}\right)$ に従う確率変数とする。以下の値を求めなさい。

$$\mathsf{E}\left(X\right)=\sum_{i=0}^{n}x_ip_i$$

問題 IX − 2 − 5

X は $B\left(8,\ \dfrac{1}{2}\right)$ に従う確率変数とする。以下の値を求めなさい。

$$\mathsf{V}\left(X\right)=\sum_{i=0}^{n}\left(x_i-\mathsf{E}\left(X\right)\right)^2p_i$$

解例 IX − 2 − 4

$$
\begin{aligned}
&0\times\frac{1}{256}+1\times\frac{8}{256}+2\times\frac{28}{256}+3\times\frac{56}{256}+4\times\frac{70}{256}+5\times\frac{56}{256}+6\times\frac{28}{256}+7\times\frac{8}{256}+8\times\frac{1}{256}\\
&=\frac{0+8+56+168+280+280+168+56+8}{256}\\
&=4
\end{aligned}
$$

解例 IX − 2 − 5

$$
\begin{aligned}
&(0-4)^2\times\frac{1}{256}+(1-4)^2\times\frac{8}{256}+(2-4)^2\times\frac{28}{256}+(3-4)^2\times\frac{56}{256}\\
&+(4-4)^2\times\frac{70}{256}+(5-4)^2\times\frac{56}{256}+(6-4)^2\times\frac{28}{256}+(7-4)^2\times\frac{8}{256}+(8-4)^2\times\frac{1}{256}\\
&=2
\end{aligned}
$$

2.3　標準正規分布

確率変数 X が二項分布 $B(n,p)$ に従うとき確率変数 X を標準化した変数 Z

$$z = \frac{x - np}{\sqrt{np(1-p)}} \tag{19}$$

は n を十分に大きくすると、そのヒストグラムは、左右対称な釣り鐘型の図形に近づいてゆく。

このとき、ヒストグラムの上部を点で結んだ曲線は

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \tag{20}$$

に近似する。(20) を『**正規曲線**』という。この正規曲線と z 軸で囲われる範囲の面積は 1 である。正規曲線は $z=0$ の時、最大値

$$f(z) = 0.39894228\cdots \tag{21}$$

をとる f 軸に対して対称な単峰型の正値関数である。

図 1　$B(100, 0.5)$ の Z の分布

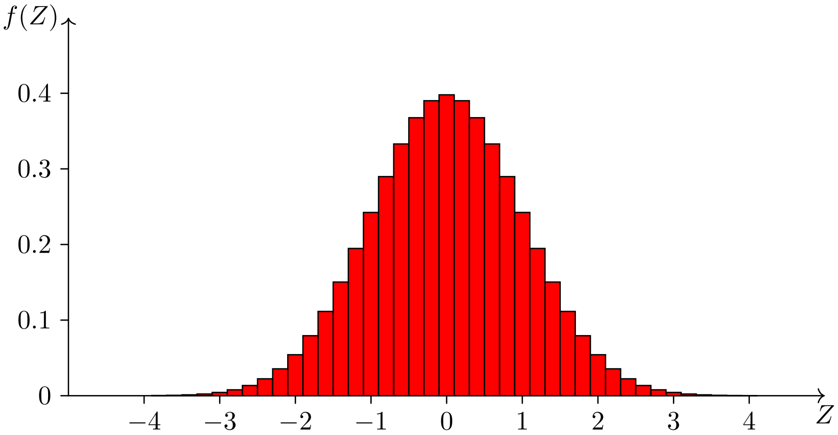
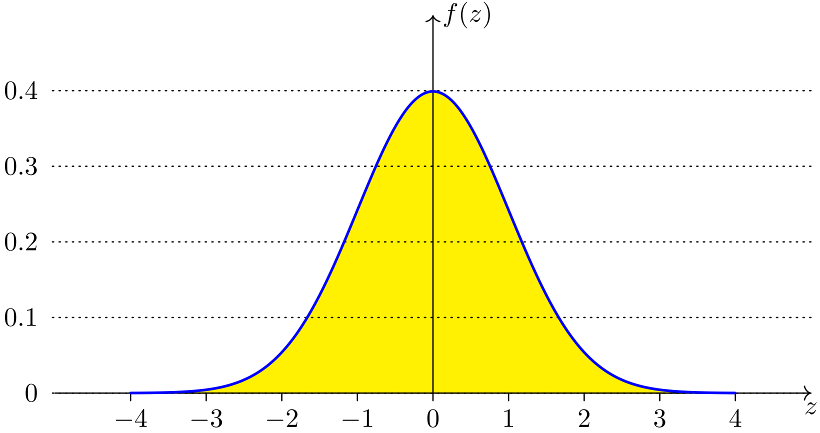


図 2　正規曲線



2.3.1　標準正規分布

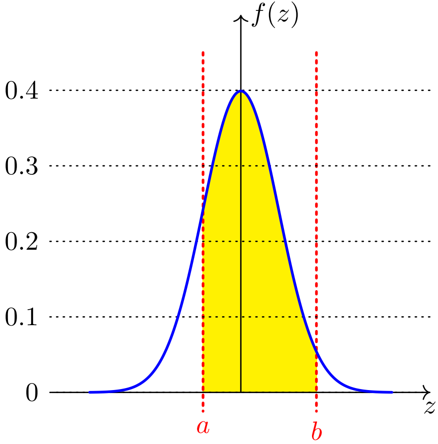
定義

確率変数 Z が a と b の間にある確率 $\Pr(a \leqq z \leqq b)$ が区間 $[a,\ b]$ において、**正規曲線**と z 軸に挟まれた領域の面積として与えられるとき、確率変数 Z の分布は『**標準正規分布**』であるといい

$$N(0, 1) \tag{22}$$

であらわす。

図 3　$\Pr(a \leqq z \leqq b)$

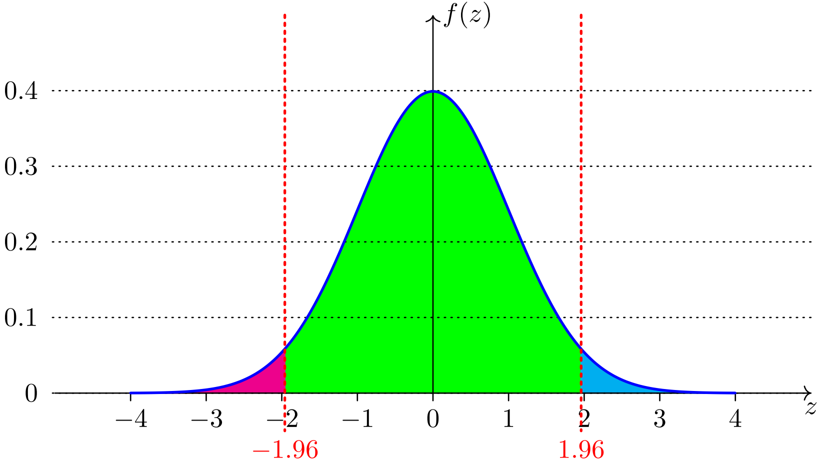


2.3.2　標準正規分布の特徴

確率変数 Z が標準正規分布に従うとき、以下の特徴がある。

- $\Pr\left(z \leqq -1.959\,964\right) = 0.024\,997\cdots$ である。
- $\Pr\left(z \geqq 1.959\,964\right) = 0.024\,997\cdots$ である。
- $\Pr\left(-1.959\,964 \leqq z \leqq 1.959\,964\right) = 0.950\,004\cdots$ である。

図 4 標準正規分布の特徴



定義

確率変数 $X = x$ が標準正規分布 Z を用いて

$$x = \mu + \sigma z \quad ;\text{ただし}\sigma \neq 0 \tag{23}$$

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

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

$$z = \frac{x - \mu}{\sigma} \tag{24}$$

は標準正規分布 $N(0, 1)$ に従う。

平均を μ, 分散を $\sigma^2 > 0$ とすると、正規分布の確率密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{25}$$

で与えられる。

2.3.3　信頼区間

標準正規分布において z が $-1.96 \leqq z \leqq 1.96$ 区間に収まる確率は

$$\Pr\left(-1.96 \leqq z \leqq 1.96\right) = 0.950\,004\cdots \tag{26}$$

である。この区間 $-1.96 \leqq z \leqq 1.96$ を『**95% 信頼区間**』という。

2.3.4　信頼区間の意味

- 観測される $Z = z$ が 95 % 信頼区間の範囲に収まる確率は 0.95 である。
- 範囲に含まれない確率は 0.05 である。
- 信頼区間の幅を広げれば、範囲に含まれる確率は大きくなる。
- 標準正規分布では定義域は $(-\infty,\ +\infty)$ なので、$-1.96 \leqq z \leqq 1.96$ は相対的にほんのわずかな区間である。
- 信頼区間の範囲を決めるのは分析者である。
- 同じ確率を持つならば、信頼区間は狭い範囲であること好ましい。
- 標準正規分布は左右対称であって、対称の軸で最大値を持つ単峰形をしているので、同じ確率を持つならば $z = 0$ を中心にした対称の区間が最も好ましい。

2.3.5 一般的な正規分布の 95% 信頼区間

$X = x$ が平均 μ, 標準偏差 σ の正規分布 $N(\mu,\ \sigma^2)$ に従う場合

$$x = \mu + \sigma z \tag{23}$$

なので

$$z = \frac{x - \mu}{\sigma} \tag{24}$$

である。従って、x の 95% 信頼区間は

$$\mu - 1.96\ \sigma \leqq \quad x \quad \leqq \mu + 1.96\ \sigma \tag{27}$$

である。

2.4　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「帰無仮説は不当であり、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。
- 帰無仮説できないと判断することを、帰無仮説を『**受容する**』という。
- 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。

2.4.1　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 であらわす。
- 有意水準 α は正値であり、5%, 1% が良くとられる。
- $\Pr(Z \geqq t) \leqq \alpha$ 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 (27) の外側の区間を『**棄却域**』という。
- $X = x \sim N(\mu,\ \sigma^2)$ の $\alpha = 5\%$ の棄却域 w は

$$w = \left\{ x \;\middle|\; x < \mu - 1.96\,\sigma,\ x > \mu + 1.96\,\sigma \right\} \text{である。} \tag{28}$$

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