第 IX 部　母平均の差の検定

目次

4　不偏分散

4.1　はじめに

- 通常母集団の情報はありません。
- 『標本の分散』を使って母集団の分散を推定します。
- 『標本の分散』と『標本分散』は異なる意味を持ちます。
- 具体的な数字を用いて確認します。

4.1.1　ポイント

- 不偏分散
- 不偏分散の期待値

4.1.2　総和

定義

n 個の数値があり、i 番目の値を x_i であらわす。n 個ある数値を全て足し合わせることを『**総和**』といい、$\displaystyle\sum_{i=1}^{n} x_i$ であらわす。

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{1}$$

定義から以下の公式を導き出すことができる。ただし $a,\ c$ を定数とする。

$$\sum_{i=1}^{n} c = nc \tag{2}$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i \tag{3}$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{4}$$

4.1.3 平均

定義

総和を n で割った値を『**平均**』といい、$\bar{x}$ であらわす。

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

x_i と平均 $\bar{x}$ の差

$$x_i - \bar{x} \tag{6}$$

を『**平均からの偏差**』という。

4.1.4 分散

定義

「『平均からの偏差』の二乗」の平均を『**分散**』といい、V_x であらわす。

$$V_x = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{7}$$

4.2 不偏分散

定義

標本 A_i に属する j 番目の要素を x_{ij} とあらわす。標本 A_i に属する要素 $x_{ij} \in A_i$ を全て足し合わせた値を $\sum_{x_{ij} \in A_i}^{n} x_{ij}$ とあらわす。

- 確率変数 $X = x$ からなる母集団から無作為に復元抽出されたサンプル・サイズ n の標本 A_i を考える。
- 抽出は何度も繰り返すことを考え、抽出を識別する添え字に i を採用する。
- 抽出された標本 A_i における要素を識別する添え字に j を採用し標本 A_i の要素を $x_{ij} \quad ; j = 1, 2, \cdots, n$ とする。
- 標本 A_i における x_{ij} の平均を『**標本平均**』といい $\bar{X}_i$ であらわす。

$$\bar{X}_i = \frac{1}{n} \sum_{x_{ij} \in A_i}^{n} x_{ij} \tag{8}$$

定義

標本 A_i において以下の式で定義される値を『**不偏分散**』といい ${s_i}^2$ であらわす。

$$ {s_i}^2 = \frac{1}{n-1} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i \right)^2 \tag{9} $$

この時の $n-1$ を『**自由度**』という。

4.2.1 数値例

コインを投げ、表が出たら $x = 8$, 裏が出たら $x = -8$ とする。このコインは表と裏が出る確率は 0.5 であるとする。

表 1 コインの表裏と値

事象	裏	表
x_i	-8	8
p_i	0.5	0.5

表 1 を母集団とし、母平均 μ と母分散 σ^2 を計算する。

$$\mu = \frac{-8+8}{2} = 0 \tag{10}$$

$$\sigma^2 = \frac{(8-0)^2 + (-8-0)^2}{2} = 64 \tag{11}$$

4.2.2 $n = 2$ の不偏分散の値

$n = 2$ のとき不偏分散 ${s_i}^2$ は

$$ {s_i}^2 = \frac{1}{2-1} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i\right)^2 \tag{12}$$

である。

標本 $\boldsymbol{A_i}$ の分布

$A_1 = \{$ 裏裏 $\}$, $A_2 = \{$ 裏表 $\}$, $A_3 = \{$ 表表 $\}$ とすると、それぞれの標本 A_i の出現確率 p_i は $B\,(2,\ 0.5)$ に従うから、

$$p_1 = {}_2C_0 \times (0.5)^0 \times (1-0.5)^{2-0} = \frac{2!}{(2-0)!0!} \times 0.5^2 = 0.25 \tag{13}$$

$$p_2 = {}_2C_1 \times (0.5)^1 \times (1-0.5)^{2-1} = \frac{2!}{(2-1)!1!} \times 0.5^2 = 0.50 \tag{14}$$

$$p_3 = {}_2C_2 \times (0.5)^2 \times (1-0.5)^{2-2} = \frac{2!}{(2-2)!2!} \times 0.5^2 = 0.25 \tag{15}$$

である。

そして標本平均はそれぞれ、

$$\bar{X}_1 = \frac{(-8)+(-8)}{2} = -8 \tag{16}$$

$$\bar{X}_2 = \frac{(-8)+8}{2} = 0 \tag{17}$$

$$\bar{X}_3 = \frac{8+8}{2} = 8 \tag{18}$$

である。

よって、それぞれの不偏分散は

$$ {s_1}^2 = ((-8) - (-8))^2 + ((-8) - (-8))^2 = 0 \tag{19} $$

$$ {s_2}^2 = ((8) - (0))^2 + ((-8) - (0))^2 = 128 \tag{20} $$

$$ {s_3}^2 = ((8) - (8))^2 + ((8) - (8))^2 = 0 \tag{21} $$

である。

ここで、不偏分散 ${s_i}^2$ の期待値 $\mathsf{E}(s^2)$ を求めると

$$ \mathsf{E}\left(s^2\right) = 0 \times 0.25 + 128 \times 0.50 + 0 \times 0.25 = 64 \tag{22} $$

である。

4.2.3　$n=4$ の不偏分散の値

$n=4$ のとき不偏分散 ${s_i}^2$ は

$$ {s_i}^2 = \frac{1}{4-1} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i \right)^2 \tag{23} $$

である。

$A_1 = \{$ 裏裏裏裏 $\}$, $A_2 = \{$ 裏裏裏表 $\}$, $A_3 = \{$ 裏裏表表 $\}$, $A_4 = \{$ 裏表表表 $\}$,
$A_5 = \{$ 表表表表 $\}$ とする。

すると、A_i の出現確率 p_i は、$B(4, 0.5)$ に従うから

$$p_1 = {}_4C_0 \times 0.5^0 \times 0.5^{4-0} = \frac{4!}{(4-0)!0!} \times 0.5^4 = 1 \times \left(\frac{1}{2}\right)^4 = \frac{1}{16} \tag{24}$$

$$p_2 = {}_4C_1 \times 0.5^1 \times 0.5^{4-1} = \frac{4!}{(4-1)!1!} \times 0.5^4 = 4 \times \left(\frac{1}{2}\right)^4 = \frac{4}{16} \tag{25}$$

$$p_3 = {}_4C_2 \times 0.5^2 \times 0.5^{4-2} = \frac{4!}{(4-2)!2!} \times 0.5^4 = 6 \times \left(\frac{1}{2}\right)^4 = \frac{6}{16} \tag{26}$$

$$p_4 = {}_4C_3 \times 0.5^3 \times 0.5^{4-4} = \frac{4!}{(4-3)!3!} \times 0.5^4 = 4 \times \left(\frac{1}{2}\right)^4 = \frac{4}{16} \tag{27}$$

$$p_5 = {}_4C_4 \times 0.5^4 \times 0.5^{4-4} = \frac{4!}{(4-4)!4!} \times 0.5^4 = 1 \times \left(\frac{1}{2}\right)^4 = \frac{1}{16} \tag{28}$$

それぞれの標本平均は

$$\bar{X}_1 = \frac{(-8)+(-8)+(-8)+(-8)}{4} = -8 \tag{29}$$

$$\bar{X}_2 = \frac{(-8)+(-8)+(-8)+(8)}{4} = -4 \tag{30}$$

$$\bar{X}_3 = \frac{(-8)+(-8)+(8)+(8)}{4} = 0 \tag{31}$$

$$\bar{X}_4 = \frac{(-8)+(8)+(8)+(8)}{4} = 4 \tag{32}$$

$$\bar{X}_5 = \frac{(8)+(8)+(8)+(8)}{4} = 8 \tag{33}$$

である。

よって不偏分散は

$$
{s_1}^2 = \frac{1}{3}\left(\left((-8)-(-8)\right)^2 + \left((-8)-(-8)\right)^2 + \left((-8)-(-8)\right)^2 + \left((-8)-(-8)\right)^2\right) = \frac{0}{3} \tag{34}
$$

$$
{s_2}^2 = \frac{1}{3}\left(\left((-8)-(-4)\right)^2 + \left((-8)-(-4)\right)^2 + \left((-8)-(-4)\right)^2 + \left((8)-(-4)\right)^2\right) = \frac{192}{3} \tag{35}
$$

$$
{s_3}^2 = \frac{1}{3}\left(\left((-8)-(0)\right)^2 + \left((-8)-(0)\right)^2 + \left((8)-(0)\right)^2 + \left((8)-(0)\right)^2\right) = \frac{256}{3} \tag{36}
$$

$$
{s_4}^2 = \frac{1}{3}\left(\left((-8)-(4)\right)^2 + \left((8)-(4)\right)^2 + \left((8)-(4)\right)^2 + \left((8)-(4)\right)^2\right) = \frac{192}{3} \tag{37}
$$

$$
{s_5}^2 = \frac{1}{3}\left(\left((8)-(8)\right)^2 + \left((8)-(8)\right)^2 + \left((8)-(8)\right)^2 + \left((8)-(8)\right)^2\right) = \frac{0}{3} \tag{38}
$$

問題 IX − 4 − 1

表 1 を母集団とする。$n=4$ のとき不偏分散 ${s_i}^2$ の期待値 $\mathsf{E}(s^2)$ を求めなさい。

よって不偏分散の期待値 $\mathsf{E}\left(s^2\right)$ は

$$\begin{aligned}
\mathsf{E}\left(s^2\right) &= \frac{0}{3}\times\frac{1}{16}+\frac{192}{3}\times\frac{4}{16}+\frac{256}{3}\times\frac{6}{16}+\frac{192}{3}\times\frac{4}{16}+\frac{0}{3}\times\frac{1}{16} && (39)\\
&= 0+16+32+16+0 && (40)\\
&= 64 && (41)
\end{aligned}$$

問題 IX − 4 − 2

表 1 を母集団とする。$n = 8$ のとき不偏分散 ${s_i}^2$ の期待値 $\mathsf{E}\left(s^2\right)$ を求めなさい。

表 2　$n=8$ の標本

$A_1 = \{$ 裏裏裏裏裏裏裏裏 $\}$	$A_2 = \{$ 裏裏裏裏裏裏裏表 $\}$
$A_3 = \{$ 裏裏裏裏裏裏表表 $\}$	$A_4 = \{$ 裏裏裏裏裏表表表 $\}$
$A_5 = \{$ 裏裏裏裏表表表表 $\}$	$A_6 = \{$ 裏裏裏表表表表表 $\}$
$A_7 = \{$ 裏裏表表表表表表 $\}$	$A_8 = \{$ 裏表表表表表表表 $\}$
$A_9 = \{$ 表表表表表表表表 $\}$	

表 3　$n = 8$ のときの A_i の統計量

標本	p_i	$\bar{X}_i$	${s_i}^2$	${s_i}^2 \times p_i$
A_1	${}_8C_0 \times 0.5^8 = 1/2^8$	-8	$0/7$	0
A_2	${}_8C_1 \times 0.5^8 = 8/2^8$	-6	$224/7$	1
A_3	${}_8C_2 \times 0.5^8 = 28/2^8$	-4	$384/7$	6
A_4	${}_8C_3 \times 0.5^8 = 56/2^8$	-2	$480/7$	15
A_5	${}_8C_4 \times 0.5^8 = 70/2^8$	0	$512/7$	20
A_6	${}_8C_5 \times 0.5^8 = 56/2^8$	2	$480/7$	15
A_7	${}_8C_6 \times 0.5^8 = 28/2^8$	4	$384/7$	6
A_8	${}_8C_7 \times 0.5^8 = 8/2^8$	6	$224/7$	1
A_9	${}_8C_8 \times 0.5^8 = 1/2^8$	8	$0/7$	0

4.2.4　$n = 8$ の不偏分散の値

$n = 8$ のとき不偏分散 ${s_i}^2$ は

$$ {s_i}^2 = \frac{1}{8-1} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i \right)^2 \tag{42} $$

標本を表 2 とする。すると各種統計量は表 3 である。ここから $\mathsf{E}\left(s^2\right)$ を求めると

$$ \mathsf{E}\left(s^2\right) = 0 + 1 + 6 + 15 + 20 + 15 + 6 + 1 + 0 = 64 \tag{43} $$

である。

4.2.5　標本分散と不偏分散

サンプルサイズ n の標本 A_i における分散を『**標本分散**』といい、$V_{\bar{X}_i}$ であらわす。

$$V_{\bar{X}_i} = \frac{1}{n} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i\right)^2 \tag{44}$$

(44) 両辺に n を掛けると

$$nV_{\bar{X}_i} = \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i\right)^2 \tag{45}$$

ここで (9) 不偏分散の両辺に $n-1$ をかけると

$$s_i{}^2 = \frac{1}{n-1} \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i\right)^2 \tag{9}$$

$$(n-1)s_i{}^2 = \sum_{x_{ij} \in A_i}^{n} \left(x_{ij} - \bar{X}_i\right)^2 \tag{46}$$

(45) 左辺と (46) 左辺はそれぞれ右辺を介して等しいので

$$(n-1)s_i{}^2 = nV_{\bar{X}_i} \tag{47}$$

$$s_i{}^2 = \frac{n}{n-1} V_{\bar{X}_i} \tag{48}$$

4.3　まとめ

- 不偏分散の期待値は母分散に一致する。
- ここで示した以外でも同様に不偏分散の期待値は母分散に一致することが知られている。
- 標本分散を $n/(n-1)$ で調整した値が不偏分散である。

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