第 IX 部　母平均の差の検定

目次

3　標本平均の分布

3.1　はじめに

- 母集団から無作為に復元抽出を行います。
- 選ばれた標本から標本平均を求めます。
- 標本平均の期待値は、サンプルサイズに関わりなく母平均に一致します。
- 標本平均の分散は、母分散とサンプルサイズに影響を受けます。

3.1.1　ポイント

- 無作為抽出
- 復元抽出
- 母平均と母分散
- 標本平均の期待値と分散

3.2　サンプリング

3.2.1　母集団と標本

- 調査対象全体の集合 Ω を考える。
- Ω から観測結果としていくつかの要素を取り出す。
- Ω から測定のために要素を取り出すことを『**抽出**』といい、抽出された部分集合を『**標本**』という。標本が抽出されたとき Ω を『**母集団**』という。
- 母集団に属する要素の数を『**サイズ**』といい、標本に属する要素の数を『**サンプル・サイズ**』という。

3.2.2 抽出方法

- 要素を１つ選び出した後、その要素を元に戻し、次の要素を選び出す抽出法を『**復元抽出**』という。
- 要素を１つ選び出した後、その要素を元に戻さず、次の要素を選び出す抽出法を『**非復元抽出**』という。

3.2.3 ランダム・サンプリングの仮定

- 母集団から要素を抽出するとき、それぞれの要素が標本の要素として選ばれる確率が等しい場合、『**無作為抽出**』または『**ランダム・サンプリング**』という。
- 十分大きなサイズの標本を無作為に抽出すると、標本の相対度数は母集団の分布を反映すると仮定する。
- この仮定を『**ランダム・サンプリングの仮定**』という。

3.3　母平均と母分散

- 確率変数 $X = x$ からなる母集団を考える。
- 母集団における x の平均を『**母平均**』といい μ であらわす。
- 母集団における x の分散を『**母分散**』といい σ^2 であらわす。

3.4 標本平均

- 確率変数 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 に属する x_{ij} の総和を $\displaystyle\sum_{x_{ij} \in A_i}^{n} x_{ij}$ であらわす。
- 標本 A_i に属する x_{ij} の平均を『**標本平均**』といい $\bar{X}_i$ であらわす。

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

- $\bar{X}$ の期待値 $\mathsf{E}(\bar{X})$ は、母平均 μ に一致する。

$$\mathsf{E}(\bar{X}) = \mu \tag{2}$$

3.4.1 標本平均の分布

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それぞれ 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と母分散を計算する。

$$\mu = \frac{(-8)+8}{2} = 0 \tag{3}$$

$$\sigma^2 = \frac{(8-0)^2+(-8-0)^2}{2} = 64 \tag{4}$$

3.4.2 標本平均の値

サンプルサイズ n のとき標本平均 $\bar{X}_i$ は

$$\bar{X}_i = \frac{1}{n}\sum_{x_{ij}\in A_i}^{n} x_{ij} \tag{5}$$

である。

$n = 2$ の場合

標本点は裏裏, 表裏, 表表だからそれぞれの標本平均は

$$\bar{X}_1 = \frac{(-8) + (-8)}{2} = -8 \tag{6}$$

$$\bar{X}_2 = \frac{(-8) + (8)}{2} = 0 \tag{7}$$

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

である。

そしてそれぞれの出現確率 p_i は、二項分布 $B(2, 0.5)$ に従うから、

$$p_1 = {}_2C_0(0.5)^0(0.5)^{(2-0)} = 1 \times 0.5^2 = 0.25 \tag{9}$$

$$p_2 = {}_2C_1(0.5)^1(0.5)^{(2-1)} = 2 \times 0.5^2 = 0.50 \tag{10}$$

$$p_3 = {}_2C_2(0.5)^2(0.5)^{(2-2)} = 1 \times 0.5^2 = 0.25 \tag{11}$$

である。

よって、$\bar{X}_i$ は確率変数である。ここで $\bar{X}_i$ 期待値 $\mathsf{E}\left(\bar{X}_i\right)$ をもとめると、

$$\mathsf{E}\left(\bar{X}\right) = (-8) \times (0.25) + (0) \times (0.50) + (8) \times (0.25) = 0 \tag{12}$$

さらに $\bar{X}_i$ の分散 $\mathsf{V}\left(\bar{X}_i\right)$ を求めると

$$\mathsf{V}\left(\bar{X}\right) = (-8-0)^2 \times (0.25) + (0-0)^2 \times (0.50) + (8-0)^2 \times (0.25) = 32 \tag{13}$$

である。

表 2　$n = 2$ のときの $\bar{X}_i$ の分布

	$\bar{X}_i$	p_i
裏裏	-8	0.25
表裏	0	0.50
表表	8	0.25

$n = 4$ の場合

標本点は裏裏裏裏, 裏裏裏表, 裏裏表表, 裏表表表, 表表表表だからそれぞれの標本平均は

$$\bar{X}_1 = \frac{(-8)+(-8)+(-8)+(-8)}{4} = -8 \tag{14}$$

$$\bar{X}_2 = \frac{(-8)+(-8)+(-8)+(8)}{4} = -4 \tag{15}$$

$$\bar{X}_3 = \frac{(-8)+(-8)+(8)+(8)}{4} = 0 \tag{16}$$

$$\bar{X}_4 = \frac{(-8)+(8)+(8)+(8)}{4} = 4 \tag{17}$$

$$\bar{X}_5 = \frac{(8)+(8)+(8)+(8)}{4} = 8 \tag{18}$$

である。

そしてそれぞれの出現確率は

$$p_1 = {}_4C_0 \times 0.5^2 = 0.0625 \tag{19}$$

$$p_2 = {}_4C_1 \times 0.5^2 = 0.2500 \tag{20}$$

$$p_3 = {}_4C_2 \times 0.5^2 = 0.3750 \tag{21}$$

$$p_4 = {}_4C_3 \times 0.5^2 = 0.2500 \tag{22}$$

$$p_5 = {}_4C_4 \times 0.5^2 = 0.0625 \tag{23}$$

である。

よって期待値と分散は

$$\mathsf{E}(\bar{X}) = 0 \tag{24}$$

$$\mathsf{V}(\bar{X}) = 16 \tag{25}$$

問題 IX$-\mathbf{3}-\mathbf{1}$

　コインを投げ、裏が出たら $x=-8$、表が出たら $x=8$ とする。このコインは裏と表が出る確率がそれぞれ 0.5 であるとする。このコインを 8 回投げたときの標本平均 $\bar{X}$ の期待値と分散を求めなさい。

表 3　$n=8$ のときの事象

裏裏裏裏裏裏裏裏	裏裏裏裏裏裏裏表
裏裏裏裏裏裏表表	裏裏裏裏裏表表表
裏裏裏裏表表表表	裏裏裏表表表表表
裏裏表表表表表表	裏表表表表表表表
表表表表表表表表	

$$\mathsf{E}(\bar{X}) = 0 \tag{26}$$

$$\mathsf{V}(\bar{X}) = 8 \tag{27}$$

表 4　$n = 8$ の標本平均の分布

i	$\bar{X}_i$	p_i	$\bar{X}_i^2$
1	-8	0.0039	64
2	-6	0.0313	36
3	-4	0.1094	16
4	-2	0.2188	4
5	0	0.2734	0
6	2	0.2188	4
7	4	0.1094	16
8	6	0.0313	36
9	8	0.0039	64

$n = 16$ の場合

$$\mathsf{E}(\bar{X}) = 0 \tag{28}$$

$$\mathsf{V}(\bar{X}) = 4 \tag{29}$$

$n = 32$ の場合

$$\mathsf{E}(\bar{X}) = 0 \tag{30}$$

$$\mathsf{V}(\bar{X}) = 2 \tag{31}$$

$n = 64$ の場合

$$\mathsf{E}(\bar{X}) = 0 \tag{32}$$

$$\mathsf{V}(\bar{X}) = 1 \tag{33}$$

であることは容易に確かめることができる。

これまでの結果をまとめたものが表 5 である。表から、サンプル数に関わらず、標本平均は母平均に一致していることが読み取れる。

$$\mathsf{E}\left(\bar{X}\right) = \mu \tag{34}$$

また、標本平均の分散 $\mathsf{V}\left(\bar{X}\right)$ と母分散の関係は

$$\mathsf{V}\left(\bar{X}\right) = \frac{\sigma^2}{n} \tag{35}$$

であることが予想される。

表 5　母集団と標本平均の関係

母集団	$\mu = 0$	$\sigma^2 = 64$
$n = 2$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 32$
$n = 4$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 16$
$n = 8$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 8$
$n = 16$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 4$
$n = 32$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 2$
$n = 64$	$\mathsf{E}(\bar{X}) = 0$	$\mathsf{V}(\bar{X}) = 1$

3.4.3 母集団と標本平均の統計量の関係

- 数値例において

$$\mathsf{E}\left(\bar{X}\right) = \mu \tag{34}$$

$$\mathsf{V}\left(\bar{X}\right) = \frac{\sigma^2}{n} \tag{35}$$

  の傾向があることが確認できる。
- この傾向は、他の分布、サンプル・サイズでも確認できることが知られている。
- サンプル・サイズ n が十分に大きいとき、標本平均 $\bar{X}$ は、母平均 μ に近い値をとる可能性が高くなると考えられる。
- サンプル・サイズ n が大きくなるほどこの傾向は強まる。この性質を『**大数の法則**』という。
- 母集団の分布がどのようなものであっても n が十分に大きいとき標本平均が正規分布に従うことは『**中心極限定理**』で示されている。

3.4.4　正規母集団

- 正規分布に従う母集団を『**正規母集団**』という。
- 正規母集団ならば、母集団と標本平均の分布は、ともに正規分布である。
- 正規母集団の母平均を 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と、標本平均 $\bar{X}$ の期待値 $\mathsf{E}\left(\bar{X}\right)$ は μ に一致し、分散 $\mathsf{V}\left(\bar{X}\right)$ は $\dfrac{\sigma^2}{n}$ になる。

3.5　母平均の区間推定

3.5.1　95% 信頼区間

$N\left(\mu,\ \sigma^2\right)$ に従う x の 95% 信頼区間は

$$\mu - 1.96\ \sigma \leqq x \leqq \mu + 1.96\ \sigma \tag{36}$$

である。観測される x は 95% の確率で (36) の範囲に入ることを意味する。

3.5.2　正規母集団からの標本平均の 95% 信頼区間

$N\left(\mu,\ \sigma^2\right)$ に従う正規母集団からサンプル・サイズ n の標本を抽出したときの標本平均 $\bar{X}$ は $N\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ に従う。よって、$N\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ の 95% 信頼区間は、

$$\mu - 1.96\ \frac{\sigma}{\sqrt{n}} \leqq \bar{X} \leqq \mu + 1.96\ \frac{\sigma}{\sqrt{n}} \tag{37}$$

(37) を変形すると

$$-1.96 \leqq \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \leqq 1.96 \tag{38}$$

を得る。

3.5.3　標本平均 $\bar{X}$ による μ の区間推定

母集団が正規分布に従い、母分散が σ^2 であることが既知であるとする。この母集団から、サンプル・サイズ n の標本を抽出すると「標本平均 $\bar{X}$」の分散 $\mathsf{V}\left(\bar{X}\right)$ は

$$\mathsf{V}\left(\bar{X}\right) = \frac{\sigma^2}{n} \tag{39}$$

なので標準偏差 $\mathsf{D}\left(\bar{X}\right)$ は

$$\mathsf{D}\left(\bar{X}\right) = \frac{\sigma}{\sqrt{n}} \tag{40}$$

とあらわすことができる。

このとき未知である母平均を μ^* であらわせば 95% 信頼区間は

$$-1.96 \leqq \frac{\bar{X} - \mu^*}{\left(\frac{\sigma}{\sqrt{n}}\right)} \leqq 1.96 \tag{41}$$

である。(41) を整理すると

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leqq \mu^* \leqq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \tag{42}$$

である。

これはサンプル・サイズ n の標本平均が $\bar{X}$ ならば、母分散が σ^2 の正規母集団の母平均 μ^* は 95% の確率で (42) の区間に含まれることを意味する。

(42) の区間を『**母平均 $\boldsymbol{\mu}$ の 95% 信頼区間**』という。サンプルから得られた標本平均 $\bar{X}$ からこの区間を求めることを『**母平均の区間推定**』という。

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