

第 II 部 統計

11 母集団と標本

ポイント

- 母集団と標本
- 復元抽出と非復元抽出

- 母平均と標本平均
- 母分散と不偏分散と標本分散

11.1 はじめに

- 全部を知ることは困難です。
- 一部であれば知ることはできます。
- 一部を通じて全部を知る方法は魅力的です。

11.2 母集団を把握する方法

- 調査対象を全て調べることは困難な場合が多い。
 - 調査対象が大きすぎる場合
 - 調査により調査対象を棄損する場合
- 対処法として
 1. 調査対象から一部を選びだす。
 2. その一部を分析する。
 3. 調査対象全体について推測を行う。

11.2.1 標本抽出

- 調査対象に含まれる要素の数を N であらわす。
- 調査対象から分析のために選びだされた要素の集合を『標本』という。
- 標本を選び出したとき、調査対象全体を『母集団^{ぼしゅうだん}』という。
- 標本として選ばれる要素の数を n であらわす。この n を『標本の大きさ』または『サンプルサイズ』という。
- 標本を母集団から選びだすことを『標本抽出』という。
- 選びだされた標本を直接調査し、母集団の分布を把握することを『統計的推測』と呼ぶ。

11.3 母平均と母分散

母集団の要素の数を N であらわし、要素を識別する添え字に j を用いて各要素を x_j とあらわす。母集団の要素の平均を『母平均』^{ぼへいきん} といい μ であらわす。また母集合の要素の分散を『母分散』^{ぼぶんさん} とよび σ^2 であらわす。通常 μ, σ^2 は知られていない。

$$\mu = \frac{1}{N} \sum_{j=1}^N x_j = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} \quad (1)$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \mu)^2 = \frac{1}{N} \sum_{j=1}^N x_j^2 - \mu^2 \quad (2)$$

μ はギリシャ文字ミューの小文字である。

11.3.1 $N = 6$ の母集団

$N = 6$ の極めて規模の小さな母集団を使い母平均, 母分散を考察する。

表 1 6 個の要素からなる母集団

j	1	2	3	4	5	6
要素	x_1	x_2	x_3	x_4	x_5	x_6

11.3.2 $N = 6$ の母集団の母平均 μ

母集団における平均が母平均 μ である。したがって母平均 μ は、

$$\mu = \frac{1}{6} \sum_{j=1}^6 x_j \quad (3)$$

$$= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \quad (4)$$

である。

11.3.3 $N = 6$ の母集団の母分散 σ^2

母集団における分散が母分散 σ^2 である。したがって、

$$\sigma^2 = \frac{1}{6} \sum_{j=1}^6 (x_j - \mu)^2 = \frac{1}{6} \sum_{j=1}^6 x_j^2 - \mu^2 \quad (5)$$

$$= \frac{1}{6} \sum_{j=1}^6 x_j^2 - \left(\frac{1}{6} \sum_{j=1}^6 x_j \right)^2 \quad (6)$$

$$= \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2}{6} - \left(\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \right)^2 \quad (7)$$

である。

これを計算すると、

$$\begin{aligned}\sigma^2 = \frac{1}{36} & \left(5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_4^2 + 5x_5^2 + 5x_6^2 \right. \\ & - 2x_1x_2 - 2x_1x_3 - 2x_1x_4 - 2x_1x_5 - 2x_1x_6 \\ & - 2x_2x_3 - 2x_2x_4 - 2x_2x_5 - 2x_2x_6 - 2x_3x_4 \\ & \left. - 2x_3x_5 - 2x_3x_6 - 2x_4x_5 - 2x_4x_6 - 2x_5x_6 \right) \quad (8)\end{aligned}$$

を得る。

11.4 復元抽出と非復元抽出

- 要素を一つ選び出す度に、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。
- 要素を一つ選び出した後、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。

11.4.1 非復元抽出時の要素の選び方の数

非復元抽出では N 個の母集団から順番に n 個の標本を選ぶとき、最初の一つはどれでもいいので N 個ある。二つ目はそれぞれの N に対して $(N - 1)$ がある。したがって二つを順番に選ぶときの選び方の総数は

$$N \times (N - 1) \tag{9}$$

である。そしてこれは n 個まで繰り返されるので n 番目の可能な個数は $(N - n + 1)$ なので n 個を順番に選ぶときの選び方の総数は

$$N \times (N - 1) \times (N - 2) \times \cdots \times (N - n + 1) \tag{10}$$

である。

11.4.2 復元抽出時の要素の選び方の数

復元抽出では N 個の母集団から n 回要素を選ぶことになるので、要素の選びかたの総数は

$$N^n \tag{11}$$

である。

n に比べて N が十分に大きい場合、復元抽出による標本の選び方の総数と非復元抽出による標本の選び方の総数の差はほとんどない。

表 2 $N = 6$ の母集団から非復元抽出による $n = 2$ の抽出

(x_1, x_2)	(x_1, x_3)	(x_1, x_4)	(x_1, x_5)	(x_1, x_6)
(x_2, x_1)	(x_2, x_3)	(x_2, x_4)	(x_2, x_5)	(x_2, x_6)
(x_3, x_1)	(x_3, x_2)	(x_3, x_4)	(x_3, x_5)	(x_3, x_6)
(x_4, x_1)	(x_4, x_2)	(x_4, x_3)	(x_4, x_5)	(x_4, x_6)
(x_5, x_1)	(x_5, x_2)	(x_5, x_3)	(x_5, x_4)	(x_5, x_6)
(x_6, x_1)	(x_6, x_2)	(x_6, x_3)	(x_6, x_4)	(x_6, x_5)

表3 $N = 6$ の母集団から復元抽出による $n = 2$ の抽出

(x_1, x_1)	(x_1, x_2)	(x_1, x_3)	(x_1, x_4)	(x_1, x_5)	(x_1, x_6)
(x_2, x_1)	(x_2, x_2)	(x_2, x_3)	(x_2, x_4)	(x_2, x_5)	(x_2, x_6)
(x_3, x_1)	(x_3, x_2)	(x_3, x_3)	(x_3, x_4)	(x_3, x_5)	(x_3, x_6)
(x_4, x_1)	(x_4, x_2)	(x_4, x_3)	(x_4, x_4)	(x_4, x_5)	(x_4, x_6)
(x_5, x_1)	(x_5, x_2)	(x_5, x_3)	(x_5, x_4)	(x_5, x_5)	(x_5, x_6)
(x_6, x_1)	(x_6, x_2)	(x_6, x_3)	(x_6, x_4)	(x_6, x_5)	(x_6, x_6)

表 4 $N = 6, n = 2$ の非復元抽出による標本と確率

標本 _{<i>i</i>}	要素の組	p_i	標本 _{<i>i</i>}	要素の組	p_i	標本 _{<i>i</i>}	要素の組	p_i
A_1	$\{x_1, x_2\}$	$\frac{1}{15}$	A_6	$\{x_2, x_3\}$	$\frac{1}{15}$	A_{11}	$\{x_3, x_5\}$	$\frac{1}{15}$
A_2	$\{x_1, x_3\}$	$\frac{1}{15}$	A_7	$\{x_2, x_4\}$	$\frac{1}{15}$	A_{12}	$\{x_3, x_6\}$	$\frac{1}{15}$
A_3	$\{x_1, x_4\}$	$\frac{1}{15}$	A_8	$\{x_2, x_5\}$	$\frac{1}{15}$	A_{13}	$\{x_4, x_5\}$	$\frac{1}{15}$
A_4	$\{x_1, x_5\}$	$\frac{1}{15}$	A_9	$\{x_2, x_6\}$	$\frac{1}{15}$	A_{14}	$\{x_4, x_6\}$	$\frac{1}{15}$
A_5	$\{x_1, x_6\}$	$\frac{1}{15}$	A_{10}	$\{x_3, x_4\}$	$\frac{1}{15}$	A_{15}	$\{x_5, x_6\}$	$\frac{1}{15}$

表 5 $N = 6, n = 2$ の復元抽出による標本と確率

標本 _i	要素の組	p_i	標本 _i	要素の組	p_i	標本 _i	要素の組	p_i
B_1	$\{x_1, x_1\}$	$\frac{1}{36}$	B_8	$\{x_2, x_3\}$	$\frac{2}{36}$	B_{15}	$\{x_3, x_6\}$	$\frac{2}{36}$
B_2	$\{x_1, x_2\}$	$\frac{2}{36}$	B_9	$\{x_2, x_4\}$	$\frac{2}{36}$	B_{16}	$\{x_4, x_4\}$	$\frac{1}{36}$
B_3	$\{x_1, x_3\}$	$\frac{2}{36}$	B_{10}	$\{x_2, x_5\}$	$\frac{2}{36}$	B_{17}	$\{x_4, x_5\}$	$\frac{2}{36}$
B_4	$\{x_1, x_4\}$	$\frac{2}{36}$	B_{11}	$\{x_2, x_6\}$	$\frac{2}{36}$	B_{18}	$\{x_4, x_6\}$	$\frac{2}{36}$
B_5	$\{x_1, x_5\}$	$\frac{2}{36}$	B_{12}	$\{x_3, x_3\}$	$\frac{1}{36}$	B_{19}	$\{x_5, x_5\}$	$\frac{1}{36}$
B_6	$\{x_1, x_6\}$	$\frac{2}{36}$	B_{13}	$\{x_3, x_4\}$	$\frac{2}{36}$	B_{20}	$\{x_5, x_6\}$	$\frac{2}{36}$
B_7	$\{x_2, x_2\}$	$\frac{1}{36}$	B_{14}	$\{x_3, x_5\}$	$\frac{2}{36}$	B_{21}	$\{x_6, x_6\}$	$\frac{1}{36}$

11.5 標本平均

A_i の要素を x_j とする。 A_i に含まれる要素の数を n_i とする。 A_i に属する全ての要素の総和を以下の記号であらわす。

$$\sum_{x_j \in A_i}^{n_i} x_j \quad (12)$$

標本 A_i の要素の平均を『標本平均』とよ

び $\bar{X}_i$ であらわす。

$$\bar{X}_i = \frac{1}{n_i} \sum_{x_j \in A_i}^{n_i} x_j \quad (13)$$

標本平均 $\bar{X}_i$ は確率変数である。

表 6 $N = 6, n = 2$ の非復元抽出による標本平均

A_i	$\bar{X}_i$	p_i	A_i	$\bar{X}_i$	p_i	A_i	$\bar{X}_i$	p_i
A_1	$\frac{x_1 + x_2}{2}$	$\frac{1}{15}$	A_6	$\frac{x_2 + x_3}{2}$	$\frac{1}{15}$	A_{11}	$\frac{x_3 + x_5}{2}$	$\frac{1}{15}$
A_2	$\frac{x_1 + x_3}{2}$	$\frac{1}{15}$	A_7	$\frac{x_2 + x_4}{2}$	$\frac{1}{15}$	A_{12}	$\frac{x_3 + x_6}{2}$	$\frac{1}{15}$
A_3	$\frac{x_1 + x_4}{2}$	$\frac{1}{15}$	A_8	$\frac{x_2 + x_5}{2}$	$\frac{1}{15}$	A_{13}	$\frac{x_4 + x_5}{2}$	$\frac{1}{15}$
A_4	$\frac{x_1 + x_5}{2}$	$\frac{1}{15}$	A_9	$\frac{x_2 + x_6}{2}$	$\frac{1}{15}$	A_{14}	$\frac{x_4 + x_6}{2}$	$\frac{1}{15}$
A_5	$\frac{x_1 + x_6}{2}$	$\frac{1}{15}$	A_{10}	$\frac{x_3 + x_4}{2}$	$\frac{1}{15}$	A_{15}	$\frac{x_5 + x_6}{2}$	$\frac{1}{15}$

表7 $N = 6, n = 2$ の復元抽出による標本平均

標本	標本平均	確率	標本	標本平均	確率	標本	標本平均	確率
B_1	$\frac{x_1 + x_1}{2}$	$\frac{1}{36}$	B_8	$\frac{x_2 + x_3}{2}$	$\frac{2}{36}$	B_{15}	$\frac{x_3 + x_6}{2}$	$\frac{2}{36}$
B_2	$\frac{x_1 + x_2}{2}$	$\frac{2}{36}$	B_9	$\frac{x_2 + x_4}{2}$	$\frac{2}{36}$	B_{16}	$\frac{x_4 + x_4}{2}$	$\frac{1}{36}$
B_3	$\frac{x_1 + x_3}{2}$	$\frac{2}{36}$	B_{10}	$\frac{x_2 + x_5}{2}$	$\frac{2}{36}$	B_{17}	$\frac{x_4 + x_5}{2}$	$\frac{2}{36}$
B_4	$\frac{x_1 + x_4}{2}$	$\frac{2}{36}$	B_{11}	$\frac{x_2 + x_6}{2}$	$\frac{2}{36}$	B_{18}	$\frac{x_4 + x_6}{2}$	$\frac{2}{36}$
B_5	$\frac{x_1 + x_5}{2}$	$\frac{2}{36}$	B_{12}	$\frac{x_3 + x_3}{2}$	$\frac{1}{36}$	B_{19}	$\frac{x_5 + x_5}{2}$	$\frac{1}{36}$
B_6	$\frac{x_1 + x_6}{2}$	$\frac{2}{36}$	B_{13}	$\frac{x_3 + x_4}{2}$	$\frac{2}{36}$	B_{20}	$\frac{x_5 + x_6}{2}$	$\frac{2}{36}$
B_7	$\frac{x_2 + x_2}{2}$	$\frac{1}{36}$	B_{14}	$\frac{x_3 + x_5}{2}$	$\frac{2}{36}$	B_{21}	$\frac{x_6 + x_6}{2}$	$\frac{1}{36}$

11.5.1 非復元抽出による標本平均の期待値

表 6 を用いて非復元抽出による標本平均の期待値を求める。

$$\begin{aligned}
 E(\bar{X}) &= \sum_{i=1}^{15} X_i p_i & (14) \\
 &= \frac{x_1 + x_2}{2} \frac{1}{15} + \frac{x_1 + x_3}{2} \frac{1}{15} + \frac{x_1 + x_4}{2} \frac{1}{15} + \frac{x_1 + x_5}{2} \frac{1}{15} + \frac{x_1 + x_6}{2} \frac{1}{15} \\
 &\quad + \frac{x_2 + x_3}{2} \frac{1}{15} + \frac{x_2 + x_4}{2} \frac{1}{15} + \frac{x_2 + x_5}{2} \frac{1}{15} + \frac{x_2 + x_6}{2} \frac{1}{15} + \frac{x_3 + x_4}{2} \frac{1}{15} \\
 &\quad + \frac{x_3 + x_5}{2} \frac{1}{15} + \frac{x_3 + x_6}{2} \frac{1}{15} + \frac{x_4 + x_5}{2} \frac{1}{15} + \frac{x_4 + x_6}{2} \frac{1}{15} + \frac{x_5 + x_6}{2} \frac{1}{15} & (15)
 \end{aligned}$$

共通因子をはき出すと

$$= \frac{1}{30} (5x_1 + 5x_2 + 5x_3 + 5x_4 + 5x_5 + 5x_6) \quad (16)$$

さらに括弧中から共通因子をはき出し

$$= \frac{1}{6} (x_1 + x_2 + x_3 + x_4 + x_5 + x_6) \quad (17)$$

$$= \mu \quad (18)$$

この例において、非復元抽出された標本平均の期待値が母平均に一致することが確認出来る。

問題 II-11 - 1

表 7 を用いて、 $N = 6$, $n = 2$ の復元抽出における標本平均の期待値を求めなさい。

解例 II-11-1

$$\begin{aligned}
 E(\bar{X}_i) &= \frac{x_1 + x_1}{2} \frac{1}{36} + \frac{x_1 + x_2}{2} \frac{2}{36} + \frac{x_1 + x_3}{2} \frac{2}{36} + \frac{x_1 + x_4}{2} \frac{2}{36} + \frac{x_1 + x_5}{2} \frac{2}{36} + \frac{x_1 + x_6}{2} \frac{2}{36} \\
 &+ \frac{x_2 + x_2}{2} \frac{1}{36} + \frac{x_2 + x_3}{2} \frac{2}{36} + \frac{x_2 + x_4}{2} \frac{2}{36} + \frac{x_2 + x_5}{2} \frac{2}{36} + \frac{x_2 + x_6}{2} \frac{2}{36} + \frac{x_3 + x_3}{2} \frac{1}{36} \\
 &+ \frac{x_3 + x_4}{2} \frac{2}{36} + \frac{x_3 + x_5}{2} \frac{2}{36} + \frac{x_3 + x_6}{2} \frac{2}{36} + \frac{x_4 + x_4}{2} \frac{1}{36} + \frac{x_4 + x_5}{2} \frac{2}{36} + \frac{x_4 + x_6}{2} \frac{2}{36} \\
 &+ \frac{x_5 + x_5}{2} \frac{1}{36} + \frac{x_5 + x_6}{2} \frac{2}{36} + \frac{x_6 + x_6}{2} \frac{1}{36} \\
 &= \frac{2}{36} \frac{1}{2} (6x_1 + 6x_2 + 6x_3 + 6x_4 + 6x_5 + 6x_6) \\
 &= \frac{1}{6} (x_1 + x_2 + x_3 + x_4 + x_5 + x_6) \\
 &= \mu
 \end{aligned}$$

11.6 不偏分散

標本 A_i の『不偏分散』を s_i^2 であらわし以下の式で定義する。

$$s_i^2 = \frac{1}{n_i - 1} \sum_{x_j \in A_i}^{n_i} (x_j - \bar{X}_i)^2 \quad (19)$$

この時の $n_i - 1$ を自由度と呼ぶ。

11.6.1 標本 $A_i = \{x_j, x_k\}$ の不偏分散 s_i^2

標本 A_i における不偏分散 s_i^2 を求めると、

$$s_i^2 = \frac{1}{2-1} \left(\left(x_j - \frac{x_j + x_k}{2} \right)^2 + \left(x_k - \frac{x_j + x_k}{2} \right)^2 \right) \quad (20)$$

$$= \left(\frac{x_j - x_k}{2} \right)^2 + \left(\frac{x_k - x_j}{2} \right)^2 \quad (21)$$

$$= \frac{1}{2}(x_j^2 + x_k^2 - 2x_jx_k) \quad (22)$$

である。この s_i^2 を、非復元抽出による全ての A_i に対して求めたものが表 8 であり、復元抽出による全ての B_i に対して求めたものが表 9 である。

表 8 非復元抽出時の不偏分散 s_i^2

A_i	s_i^2	p_i
A_1	$\frac{x_1^2 + x_2^2 - 2x_1x_2}{2}$	$\frac{1}{15}$
A_2	$\frac{x_1^2 + x_3^2 - 2x_1x_3}{2}$	$\frac{1}{15}$
A_3	$\frac{x_1^2 + x_4^2 - 2x_1x_4}{2}$	$\frac{1}{15}$
A_4	$\frac{x_1^2 + x_5^2 - 2x_1x_5}{2}$	$\frac{1}{15}$
A_5	$\frac{x_1^2 + x_6^2 - 2x_1x_6}{2}$	$\frac{1}{15}$

A_i	s_i^2	p_i
A_6	$\frac{x_2^2 + x_3^2 - 2x_2x_3}{2}$	$\frac{1}{15}$
A_7	$\frac{x_2^2 + x_4^2 - 2x_2x_4}{2}$	$\frac{1}{15}$
A_8	$\frac{x_2^2 + x_5^2 - 2x_2x_5}{2}$	$\frac{1}{15}$
A_9	$\frac{x_2^2 + x_6^2 - 2x_2x_6}{2}$	$\frac{1}{15}$
A_{10}	$\frac{x_3^2 + x_4^2 - 2x_3x_4}{2}$	$\frac{1}{15}$

A_i	s_i^2	p_i
A_{11}	$\frac{x_3^2 + x_5^2 - 2x_3x_5}{2}$	$\frac{1}{15}$
A_{12}	$\frac{x_3^2 + x_6^2 - 2x_3x_6}{2}$	$\frac{1}{15}$
A_{13}	$\frac{x_4^2 + x_5^2 - 2x_4x_5}{2}$	$\frac{1}{15}$
A_{14}	$\frac{x_4^2 + x_6^2 - 2x_4x_6}{2}$	$\frac{1}{15}$
A_{15}	$\frac{x_5^2 + x_6^2 - 2x_5x_6}{2}$	$\frac{1}{15}$

表 9 復元抽出時の不偏分散 s_i^2

B_i	s_i^2	p_i
B_1	$\frac{x_1^2 + x_1^2 - 2x_1x_1}{2}$	$\frac{1}{36}$
B_2	$\frac{x_1^2 + x_2^2 - 2x_1x_2}{2}$	$\frac{2}{36}$
B_3	$\frac{x_1^2 + x_3^2 - 2x_1x_3}{2}$	$\frac{2}{36}$
B_4	$\frac{x_1^2 + x_4^2 - 2x_1x_4}{2}$	$\frac{2}{36}$
B_5	$\frac{x_1^2 + x_5^2 - 2x_1x_5}{2}$	$\frac{2}{36}$
B_6	$\frac{x_1^2 + x_6^2 - 2x_1x_6}{2}$	$\frac{2}{36}$
B_7	$\frac{x_2^2 + x_2^2 - 2x_2x_2}{2}$	$\frac{1}{36}$

B_i	s_i^2	p_i
B_8	$\frac{x_2^2 + x_3^2 - 2x_2x_3}{2}$	$\frac{2}{36}$
B_9	$\frac{x_2^2 + x_4^2 - 2x_2x_4}{2}$	$\frac{2}{36}$
B_{10}	$\frac{x_2^2 + x_5^2 - 2x_2x_5}{2}$	$\frac{2}{36}$
B_{11}	$\frac{x_2^2 + x_6^2 - 2x_2x_6}{2}$	$\frac{2}{36}$
B_{12}	$\frac{x_3^2 + x_3^2 - 2x_3x_3}{2}$	$\frac{1}{36}$
B_{13}	$\frac{x_3^2 + x_4^2 - 2x_3x_4}{2}$	$\frac{2}{36}$
B_{14}	$\frac{x_3^2 + x_5^2 - 2x_3x_5}{2}$	$\frac{2}{36}$

B_i	s_i^2	p_i
B_{15}	$\frac{x_3^2 + x_6^2 - 2x_3x_6}{2}$	$\frac{2}{36}$
B_{16}	$\frac{x_4^2 + x_4^2 - 2x_4x_4}{2}$	$\frac{1}{36}$
B_{17}	$\frac{x_4^2 + x_5^2 - 2x_4x_5}{2}$	$\frac{2}{36}$
B_{18}	$\frac{x_4^2 + x_6^2 - 2x_4x_6}{2}$	$\frac{2}{36}$
B_{19}	$\frac{x_5^2 + x_5^2 - 2x_5x_5}{2}$	$\frac{1}{36}$
B_{20}	$\frac{x_5^2 + x_6^2 - 2x_5x_6}{2}$	$\frac{2}{36}$
B_{21}	$\frac{x_6^2 + x_6^2 - 2x_6x_6}{2}$	$\frac{1}{36}$

11.6.2 非復元抽出時の不偏分散 s_i^2 の期待値

表 8 をもとに s_i^2 の期待値を求めてみる。

$$\begin{aligned} E(s^2) = \sum_{i=1}^{15} s_i^2 p_i &= \frac{1}{30} (5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_4^2 + 5x_5^2 + 5x_6^2 \\ &\quad - 2x_1x_2 - 2x_1x_3 - 2x_1x_4 - 2x_1x_5 - 2x_1x_6 \\ &\quad - 2x_2x_3 - 2x_2x_4 - 2x_2x_5 - 2x_2x_6 - 2x_3x_4 \\ &\quad - 2x_3x_5 - 2x_3x_6 - 2x_4x_5 - 2x_4x_6 - 2x_5x_6) \end{aligned} \quad (23)$$

この例において、非復元抽出時の s_i^2 の期待値は母分散に一致せず、母分散より大きな値となる。

問題 II-11 - 2

表 9 を用いて、 $N = 6$, $n = 2$ の復元抽出における不偏分散 s_i^2 の期待値を求めなさい。

11.6.3 復元抽出時の不偏分散の期待値

表 9 をもとに s_i^2 の期待値を求めてみると

$$\begin{aligned} E(s^2) &= \sum_{i=1}^{21} s_i^2 p_i = \frac{1}{36} (5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_4^2 + 5x_5^2 + 5x_6^2 \\ &\quad - 2x_1x_2 - 2x_1x_3 - 2x_1x_4 - 2x_1x_5 - 2x_1x_6 \\ &\quad - 2x_2x_3 - 2x_2x_4 - 2x_2x_5 - 2x_2x_6 - 2x_3x_4 \\ &\quad - 2x_3x_5 - 2x_3x_6 - 2x_4x_5 - 2x_4x_6 - 2x_5x_6) \end{aligned} \quad (24)$$

$$= \sigma^2 \quad (25)$$

を得る。この例において、復元抽出時の『不偏分散』の期待値は母分散に一致することが確認できる。

11.7 標本分散

標本 A_i の標本分散を V_i であらわし以下の式で定義する。

$$V_i = \frac{1}{n_i} \sum_{x_j \in A_i}^{n_i} (x_j - \bar{X}_i)^2 \quad (26)$$

11.8 標本 $A_i = \{x_j, x_k\}$ の標本分散 V_i

標本 A_i の標本分散 V_i を求めると

$$V_i = \frac{1}{2} \left(\left(x_j - \frac{x_j + x_k}{2} \right)^2 + \left(x_k - \frac{x_j + x_k}{2} \right)^2 \right) \quad (27)$$

$$= \frac{1}{4} (x_j^2 + x_k^2 - 2x_j x_k) \quad (28)$$

である。

11.8.1 非復元抽出時の V_i の期待値

非復元抽出時の標本分散 V_i の期待値を求めると、

$$\begin{aligned} E(V) = \sum_{i=1}^{15} V_i p_i = & \frac{1}{60} (5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_4^2 + 5x_5^2 + 5x_6^2 \\ & - 2x_1x_2 - 2x_1x_3 - 2x_1x_4 - 2x_1x_5 - 2x_1x_6 \\ & - 2x_2x_3 - 2x_2x_4 - 2x_2x_5 - 2x_2x_6 - 2x_3x_4 \\ & - 2x_3x_5 - 2x_3x_6 - 2x_4x_5 - 2x_4x_6 - 2x_5x_6) \end{aligned} \quad (29)$$

を得る。この例において非復元抽出時の『標本分散』の期待値は母分散とは一致せず、母分散より小さな値になる。

11.8.2 復元抽出時の V_i の期待値

非復元抽出時の標本分散 V_i の期待値を求めると、

$$\begin{aligned} E(V) = \sum_{i=1}^{21} V_i p_i = & \frac{1}{72} (5x_1^2 + 5x_2^2 + 5x_3^2 + 5x_4^2 + 5x_5^2 + 5x_6^2 \\ & - 2x_1x_2 - 2x_1x_3 - 2x_1x_4 - 2x_1x_5 - 2x_1x_6 \\ & - 2x_2x_3 - 2x_2x_4 - 2x_2x_5 - 2x_2x_6 - 2x_3x_4 \\ & - 2x_3x_5 - 2x_3x_6 - 2x_4x_5 - 2x_4x_6 - 2x_5x_6) \end{aligned} \quad (30)$$

を得る。この例において復元抽出時の『標本分散』の期待値は母分散とは一致せず、母分散より小さな値になる。

11.9 『不偏分散』と『標本分散』の関係

(19)(26) において

$$\alpha = \sum_{j=1}^n (x_j - \bar{X}_i)^2 \quad (31)$$

とおくと、

$$s_i^2 = \frac{1}{n_i - 1} \alpha \quad (32)$$

$$V_i = \frac{1}{n_i} \alpha \quad (33)$$

(33) を α に関して解くと

$$\alpha = n_i V_i \quad (34)$$

この (34) を (32) に代入すると

$$s_i^2 = \frac{n_i}{n_i - 1} V_i \quad (35)$$

ここで (35) 右辺の期待値を取ると

$$E\left(\frac{n_i}{n_i - 1} V\right) = \frac{n_i}{n_i - 1} E(V) \quad (36)$$

同様に (35) 左辺の期待値を取ると

$$E(s^2) = \sigma^2 \quad (37)$$

同じ確率変数の期待値をとっているので両式は等しいので

$$\frac{n_i}{n_i - 1} E(V) = \sigma^2 \quad (38)$$

$\frac{n_i}{n_i - 1}$ は標本分散 (V_i) の期待値を母分散に一致させるための係数である。

11.10 まとめ

- 把握したい母集団の統計量は『母平均』と『母分散』である。
- 標本の抽出方法には復元抽出と非復元抽出がある。
- 復元・非復元抽出にかかわらず、『標本平均』の期待値は母平均に一致する。
- 復元抽出であれば『不偏分散』の期待値は母分散に一致する。
- 『標本分散』の期待値を母分散に一致させるための係数の分母を自由度 $(n - 1)$ という。
- N が n に比べて十分大きければ非復元抽出は復元抽出と差は無く、 n が十分に大きければ標本分散は不偏分散に近づく。
- ここでは、 $N = 6$, $n = 2$ のきわめて小さな母集団と標本を用いて確認したが、それ以外の場合であっても成り立つ。