

第 IX 部 統計的検定

6 サンプルング

- ランダム・サンプリング
- 母集団
- 母平均
- 標本平均

6.1 はじめに

- 観測結果の背後には巨大な集団があります。
- 背後の集団の特徴を把握するために観測が行われます。
- 観測結果と背後の集団の関係を確認します。

6.2 母集団

- 調査対象の全体の集合 Ω を考える。
- Ω から測定結果としていくつかの要素を取り出す。
- 調査対象全体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』とよぶ。
- 標本に属する要素の数を『標本数』または『サンプルサイズ』とよぶ。
- 母集団のサイズが有限個の場合『有限母集団』といい、サイズが無限個の場合『無限母集団』という。

6.2.1 抽出方法

- 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。

6.3 ランダム・サンプリング

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。

6.3.1 トランプの例

1組のトランプを考える。トランプ1組は52枚のカードからなり、4つのマークごとに、13枚のカードが割り振られている。マークは♠ ♢ ♣ ♡の4種類であり、♠ ♣を黒札、♢ ♡を赤札とよぶ。カードにはAJQKの4つのアルファベットまたは2から10の数字が付されており、AJQKのカードを絵札とよび、2から10のカードを数札とよぶ。

6.3.2 母集団の分布

ラプラスの定義に従い、母集団におけるマークの分布を考える。シャッフルされた1組のトランプから1枚のカードが選ばれるとき、それぞれのカードの選ばれるやすさは同様に確からしい。カードは全部で52枚であって、それぞれのマークのは13枚ある。 $\spadesuit$ が選ばれる確率は

$$\Pr(\spadesuit) = \frac{13}{52} \quad (1)$$

である。他のマークも同様である。

これはカードに占める $\spadesuit$ の比率である。標本空間における $\spadesuit$ の事象の数の比であり、相対度数である。

表1 マークの分布

マーク					合計
枚数	13	13	13	13	52
構成比	$\frac{13}{52}$	$\frac{13}{52}$	$\frac{13}{52}$	$\frac{13}{52}$	$\frac{52}{52}$
確率	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

黒札は 26 枚、赤札は 26 枚である。

表 2 色札の分布

色	黒札	赤札	合計
枚数	26	26	52
確率	$\frac{1}{2}$	$\frac{1}{2}$	1

数札は 36 枚、絵札は 16 枚である。

表 3 数札と絵札の分布

色	数札	絵札	合計
枚数	36	16	52
確率	$\frac{9}{13}$	$\frac{4}{13}$	1

6.3.3 復元抽出による観測結果に関する仮定

- トランプから無作為に 1 枚選び、マーク・色・数字を確認し、カードを元に戻す試行を繰り返す。
- 十分多くの回数試行を繰り返せば、その観測結果の相対度数は、それぞれの確率に近づいていくことが予想される。
- 十分大きなサイズの抽出を無作為に復元抽出すると、サンプルの相対度数は母集団の分布を反映すると考えることができる。
- この仮定を『ランダム・サンプリングの仮定』という。

6.4 母集団の分布

♠ の数札のみの母集団を考える。数札のそれぞれの数値を x であらわす。 x と確率を対応させると x は確率変数である。

表 4 数札の分布

x	2	3	4	5	6	7	8	9	10	合計
枚数	1	1	1	1	1	1	1	1	1	9
$\Pr(X = x)$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	1

6.4.1 母平均

母集団における x の平均を『母平均』という。母平均を μ であらわすと、

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i = \frac{2+3+4+5+6+7+8+9+10}{9} = 6 \quad (2)$$

X の期待値は

$$E(X) = 2 \times \frac{1}{9} + 3 \times \frac{1}{9} + 4 \times \frac{1}{9} + 5 \times \frac{1}{9} + 6 \times \frac{1}{9} + 7 \times \frac{1}{9} + 8 \times \frac{1}{9} + 9 \times \frac{1}{9} + 10 \times \frac{1}{9} = 6 \quad (3)$$

ランダムサンプリングの仮定より、試行の回数が十分に大きければ $E(X)$ は母平均に一致する。

6.4.2 母分散

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。

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (4)$$

問題 IX-6-1

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

を展開しなさい。

解例 IX-6-1

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) \\&= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n 2\mu x_i + \frac{1}{n} \sum_{i=1}^n \mu^2 \\&= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\mu \frac{1}{n} \sum_{i=1}^n x_i + n \frac{1}{n} \mu^2 \\&= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\mu^2 + \mu^2 \\&= \frac{1}{n} \sum_{i=1}^n x_i^2 - \mu^2\end{aligned}$$

問題 IX-6-2

♠ の数札 9 枚を母集団とし、数札の数値を x とする。 x の分散を求めなさい。

解例 IX-6-2

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{2^2 + 3^2 + 4^2 + 5^2 + 6 + 7^2 + 8^2 + 9^2 + 10^2}{9} = 42.666 \dots \quad (5)$$

なので、

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \mu^2 = 42.666 - 36 = 6.666 \quad (6)$$

6.5 標本平均の分布

6.5.1 標本平均

標本 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 とし、標本 A_i に含まれる要素 x_j の総和を

$$\sum_{x_j \in A_i}^{n_i} x_j \quad (7)$$

とあらわす。

標本 A_i の要素の平均を『標本平均』といい、 $\bar{X}_i$ とあらわす。

$$\bar{X}_i = \frac{1}{n_i} \sum_{x_j \in A_i}^{n_i} x_j \quad (8)$$

6.5.2 $n_i = 2$ の標本

♠ の数札のみ 9 枚を母集団とし、数札の数値を x_j とする。 $n_i = 2$ の復元抽出における標本平均を考える。

$n_i = 2$ の復元抽出なので、得られる標本は $9^2 = 81$ 通りある。

問題 IX-6-3

♠ の数札 9 枚から $n_i = 2$ の復元抽出をするとき得られる要素の組を全て書きなさい。

表 5 $n_i = 2$ の復元抽出時の要素の組

$\{2, 2\}$	$\{2, 3\}$	$\{2, 4\}$	$\{2, 5\}$	$\{2, 6\}$	$\{2, 7\}$	$\{2, 8\}$	$\{2, 9\}$	$\{2, 10\}$
$\{3, 2\}$	$\{3, 3\}$	$\{3, 4\}$	$\{3, 5\}$	$\{3, 6\}$	$\{3, 7\}$	$\{3, 8\}$	$\{3, 9\}$	$\{3, 10\}$
$\{4, 2\}$	$\{4, 3\}$	$\{4, 4\}$	$\{4, 5\}$	$\{4, 6\}$	$\{4, 7\}$	$\{4, 8\}$	$\{4, 9\}$	$\{4, 10\}$
$\{5, 2\}$	$\{5, 3\}$	$\{5, 4\}$	$\{5, 5\}$	$\{5, 6\}$	$\{5, 7\}$	$\{5, 8\}$	$\{5, 9\}$	$\{5, 10\}$
$\{6, 2\}$	$\{6, 3\}$	$\{6, 4\}$	$\{6, 5\}$	$\{6, 6\}$	$\{6, 7\}$	$\{6, 8\}$	$\{6, 9\}$	$\{6, 10\}$
$\{7, 2\}$	$\{7, 3\}$	$\{7, 4\}$	$\{7, 5\}$	$\{7, 6\}$	$\{7, 7\}$	$\{7, 8\}$	$\{7, 9\}$	$\{7, 10\}$
$\{8, 2\}$	$\{8, 3\}$	$\{8, 4\}$	$\{8, 5\}$	$\{8, 6\}$	$\{8, 7\}$	$\{8, 8\}$	$\{8, 9\}$	$\{8, 10\}$
$\{9, 2\}$	$\{9, 3\}$	$\{9, 4\}$	$\{9, 5\}$	$\{9, 6\}$	$\{9, 7\}$	$\{9, 8\}$	$\{9, 9\}$	$\{9, 10\}$
$\{10, 2\}$	$\{10, 3\}$	$\{10, 4\}$	$\{10, 5\}$	$\{10, 6\}$	$\{10, 7\}$	$\{10, 8\}$	$\{10, 9\}$	$\{10, 10\}$

6.5.3 標本の出現確率

81 組の標本の出現確率が等しいと仮定すると、それぞれの標本の出現確率は $\frac{1}{81}$ である。

表 6 $n_i = 2$ の標本平均の度数分布

$\bar{X}_i$	$\frac{4}{2}$	$\frac{5}{2}$	$\frac{6}{2}$	$\frac{7}{2}$	$\frac{8}{2}$	$\frac{9}{2}$	$\frac{10}{2}$	$\frac{11}{2}$	$\frac{12}{2}$	$\frac{13}{2}$	$\frac{14}{2}$	$\frac{15}{2}$	$\frac{16}{2}$	$\frac{17}{2}$	$\frac{18}{2}$	$\frac{19}{2}$	$\frac{20}{2}$
$\Pr(\bar{X})$	$\frac{1}{81}$	$\frac{2}{81}$	$\frac{3}{81}$	$\frac{4}{81}$	$\frac{5}{81}$	$\frac{6}{81}$	$\frac{7}{81}$	$\frac{8}{81}$	$\frac{9}{81}$	$\frac{8}{81}$	$\frac{7}{81}$	$\frac{6}{81}$	$\frac{5}{81}$	$\frac{4}{81}$	$\frac{3}{81}$	$\frac{2}{81}$	$\frac{1}{81}$

問題 IX-6-3

表 6 に基づき $\bar{X}$ の期待値と分散を求めなさい。

6.5.4 標本平均の期待値と分散

♠ の数札を母集団とする $n_i = 2$ の標本平均 $\bar{X}_i$ の期待値と分散は

$$E(\bar{X}) = 6 \quad (9)$$

$$V(\bar{X}) = 3.333333 \dots \quad (10)$$

である。

$n_i = 3$ の標本平均 $\bar{X}_i$ の期待値と分散は

$$E(\bar{X}) = 6 \quad (11)$$

$$V(\bar{X}) = 2.222222 \dots \quad (12)$$

である。

$n_i = 4$ の標本平均 $\bar{X}_i$ の期待値と分散は

$$E(\bar{X}) = 6 \quad (13)$$

$$V(\bar{X}) = 1.666666 \dots \quad (14)$$

である。

標本平均の期待値は母平均に一致し、サンプルサイズが大きくなるに従い、標本平均の分散は小さくなってゆく。

この性質を『大数の法則』^{たいすう}という。

大数の法則

1つの母集団からサンプルサイズ n_i の標本を復元抽出し、その標本平均を $\bar{X}_i$ とする。このとき、 n_i が大きければ大きいほど、標本平均 $\bar{X}_i$ は母平均 μ に近い値をとりやすくなる。

6.6 まとめ

- 調査対象から選ばれる要素の集合を『標本』という。
- 標本を選び出すことを『抽出』という。
- 標本を抽出したとき、調査対象全体を『母集団』という。
- 標本の相対度数分布は母集団の分布を反映する。
- 母集団の平均を『母平均』といい分散を『母分散』という。
- 標本の平均を『標本平均』といい分散を『標本分散』という。
- 標本平均 $\bar{X}$ の期待値 $E(\bar{X})$ は母平均に一致し、標本平均の分散 $V(\bar{X})$ はサンプルサイズが大きくなるほど小さくなる。この性質を『大数の法則』という。

6.6.1 参考文献

- (1) 「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
- (2) 「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