

第 161 講 標本平均

学習目標

この講では、次のことができるようになることを目標とする。

- ・ 観測した標本の値から標本平均を計算できる。
- ・ 母平均、観測値、標本平均の違いを説明できる。
- ・ 標本平均を求めるとき、観測した個数で割る理由を理解できる。
- ・ 同じ標本の大きさでも、抽出された値によって標本平均が変わることを理解できる。

導入

ある集団全体の特徴を調べたいとき、すべての対象を調べるのが難しい場合がある。例えば、ある工場で作られた製品全体の重さの平均を調べる場合、すべての製品を測定できるとは限らない。そのようなとき、全体から一部を取り出して調べ、その結果から全体の特徴を考える。ここで大切なのは、母平均と標本平均は異なる概念であるという点である。母平均は母集団全体の平均として決まった値であり、標本平均は取り出した標本によって決まる値である。標本平均が母平均と一致することはあるが、常に一致するとは限らない。一部として取り出したデータの平均を標本平均と呼び、この講では標本平均の意味と計算方法を学ぶ。以後の数値や抽出例は学習用に設定した架空のものである。

母集団・標本・観測値

この講では、有限個の対象からなる母集団を扱う。母集団に含まれるそれぞれの対象には、1つの数値が対応しているものとする。

調べたい対象全体を母集団と呼ぶ。

母集団に含まれるすべての対象について、それぞれの数値を1個ずつ数えて平均した値を母平均と呼び、記号 μ で表す。

例えば、母集団の対象が4個あり、それぞれの値が2, 4, 6, 8である場合、母平均は

$$\mu = \frac{2 + 4 + 6 + 8}{4}$$

で求める。

母集団から取り出して調べる対象の集まりを標本と呼ぶ。

標本から得られた数値を観測値と呼ぶ。

標本の大きさ n は、抽出された対象の延べ個数である。復元抽出では、同じ対象が複数回選ばれることがあるため、同じ対象に対応する観測値が複数回現れる場合がある。その場合も、それぞれを1回の観測として数える。

標本平均とは何か

標本から得られた観測値の平均を標本平均と呼ぶ。

標本の大きさを正整数 n とし、観測値を $x_1, x_2, \dots, x_n$ とする。ただし、各 x_i は実数である。

標本平均 $\bar{x}$ は、観測値の合計を観測数で割った値であり、

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

で表される。

ここで、 n は 1 つの標本に含まれる観測数であり、標本全体を何回取り直したかを表す数ではない。
例えば、標本が 3, 3, 5, 7 の 4 個の観測値である場合、

$$\begin{aligned}\bar{x} &= \frac{3 + 3 + 5 + 7}{4} \\ &= \frac{18}{4} \\ &= 4.5\end{aligned}$$

となる。

同じ値が複数回現れていても、それぞれを観測した回数だけ数える。

また、 $n = 1$ の場合でも定義はそのまま使え、

$$\bar{x} = \frac{x_1}{1} = x_1$$

となる。

確率変数として見る場合

抽出する前の段階では、まだどの値が選ばれるか決まっていない。

このとき、標本の値を確率変数として $X_1, X_2, \dots, X_n$ と表す。

標本平均も確率変数であり、

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

と表す。

一方、実際に抽出して得られた値を $x_1, x_2, \dots, x_n$ とすると、それらは決まった数値であり、標本平均も

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

という数値になる。

抽出前は大文字、観測後は小文字で区別する。

母平均と標本平均の違い

母平均 μ は、母集団が固定されていれば決まった値である。

一方、標本平均 $\bar{x}$ は、どの標本を取り出すかによって変化する。

ただし、標本平均が毎回必ず異なるわけではない。同じ値になる場合もある。

また、ある標本平均が母平均と一致したからといって、その抽出方法が無作為であると断定することはできない。逆に一致しなかったからといって、無作為でないと断定することもできない。

標本平均は、標本から得られた平均値として扱う必要がある。

標本平均を求める手順

標本平均を求めるときは、次の手順で考える。

1. 観測した値の個数 n を確認する。
2. 観測値をすべて足して標本全体の合計を求める。
3. 合計を観測した個数 n で割る。

特に、分母は「値の種類数」ではなく「観測数」であることに注意する。

例題 1

標本の観測値が 2, 4, 9 である。

この標本の大きさ n と標本平均 $\bar{x}$ を求めよ。

例題 1 の解答・解説

観測値は 3 個あるため、標本の大きさは

$n = 3$ である。

標本平均は、観測値の合計を観測数で割って求める。

$$\begin{aligned}\bar{x} &= \frac{2 + 4 + 9}{3} \\ &= \frac{15}{3} \\ &= 5\end{aligned}$$

したがって、標本の大きさは $n = 3$ 、標本平均は $\bar{x} = 5$ である。

ここで 3 で割っているのは、異なる値が 3 種類あるからではなく、観測した値が 3 個あるからである。

例題 2

値 2, 4, 6, 8 をもつ 4 個の対象から、各対象が等確率で選ばれ、選んだ対象を戻してから次を選ぶ独立な復元抽出を行う。

この母集団の母平均を求めよ。

また、同じ標本の大きさ $n = 2$ の抽出例として、次の 3 回の観測結果を考える。

標本	観測値	n	標本平均
A	(2, 4)	2	求めよ
B	(4, 6)	2	求めよ
C	(6, 8)	2	求めよ

それぞれの標本平均を求め、母平均と比較せよ。

例題 2 の解答・解説

母集団の値は 2, 4, 6, 8 である。

母平均は、すべての対象の値の平均であるため、

$$\begin{aligned}
 \mu &= \frac{2 + 4 + 6 + 8}{4} \\
 &= \frac{20}{4} \\
 &= 5
 \end{aligned}$$

となる。

次に、それぞれの標本平均を求める。

標本 A では、

$$\begin{aligned}
 \bar{x} &= \frac{2 + 4}{2} \\
 &= \frac{6}{2} \\
 &= 3
 \end{aligned}$$

である。

標本 B では、

$$\begin{aligned}
 \bar{x} &= \frac{4 + 6}{2} \\
 &= \frac{10}{2} \\
 &= 5
 \end{aligned}$$

である。

標本 C では、

$$\begin{aligned}
 \bar{x} &= \frac{6 + 8}{2} \\
 &= \frac{14}{2} \\
 &= 7
 \end{aligned}$$

である。

結果を整理すると、次のようになる。

標本	観測値	n	平均
A	(2, 4)	2	3
B	(4, 6)	2	5
C	(6, 8)	2	7

母平均は 5 であり、標本 A、標本 B、標本 C の標本平均はそれぞれ 3、5、7 である。

標本 A では $3 < 5$ 、標本 B では $5 = 5$ 、標本 C では $7 > 5$ となる。

このように、同じ標本の大きさ $n = 2$ でも、取り出された値によって標本平均は異なる場合がある。ただし、この 3 つは大きさ 2 の標本を 3 組示したものであり、6 個の観測値からなる 1 つの標本ではない。

この例は標本平均が変化しうることを示しているが、すべての母集団において標本平均が必ず変化するとは限らない。

なお、ここでは各対象が等確率で選ばれ、復元して次の抽出を行う独立な復元抽出を考えている。

例題 3

互いに重ならない 2 つの群について、次の観測データがある。
第 1 群は観測数が 3 個で平均が 4、第 2 群は観測数が 2 個で平均が 9 である。
2 つの群を合わせた全 5 個の観測データの平均を求めよ。
ただし、2 つの群の平均をそのまま平均してはいけない。

例題 3 の解答・解説

群ごとの平均だけを平均すると、

$$\frac{4 + 9}{2} = 6.5$$

となる。

しかし、2 つの群の観測数は異なるため、この計算では正しい全体平均にならない。

全体の平均を求めるには、まず各群の合計を求める。

第 1 群の合計は、 $3 \times 4 = 12$ である。

第 2 群の合計は、 $2 \times 9 = 18$ である。

したがって、全 5 個の観測値の合計は、 $12 + 18 = 30$ となる。

観測数は $3 + 2 = 5$ であるため、全体の平均は

$$\bar{x} = \frac{30}{5} = 6$$

である。

全体平均を求めるときは、群平均だけを見るのではなく、観測数を考えて総和に戻してから計算する。

演習問題**練習問題 1**

観測値が 6, 8, 10, 12 である。
標本の大きさ n と標本平均 $\bar{x}$ を求めよ。

練習問題 1 の解答・解説

観測値は 4 個あるため、標本の大きさは

$n = 4$ である。

標本平均は、観測値の合計を観測数で割って求める。

$$\begin{aligned}\bar{x} &= \frac{6 + 8 + 10 + 12}{4} \\ &= \frac{36}{4} \\ &= 9\end{aligned}$$

したがって、標本の大きさは $n = 4$ 、標本平均は $\bar{x} = 9$ である。

分母の 4 は、異なる値の種類数ではなく、観測した値の個数である。

練習問題 2

観測値が $-3, 1, 2$ である。

標本平均 $\bar{x}$ を求めよ。

また、負の値を含む場合でも、なぜそのまま足して計算するのか説明せよ。

練習問題 2 の解答・解説

観測した値は 3 個であるため、

$n = 3$ である。

標本平均は、すべての観測値の合計を観測数で割る。

$$\begin{aligned}\bar{x} &= \frac{-3 + 1 + 2}{3} \\ &= \frac{0}{3} \\ &= 0\end{aligned}$$

したがって、標本平均は $\bar{x} = 0$ である。

平均は、観測した値全体の大きさの中心を求める計算であり、絶対値の大きさを表すものではない。

したがって、符号を保った観測値の総和を求め、その合計を観測数で割るという標本平均の定義に基づいて計算する。負の値も観測された値として、そのまま計算に含める。

-3 を正の値として扱うことは、元の観測値を変えてしまうため正しくない。

練習問題 3

観測値が $1.2, 1.5, 1.8$ である。

標本平均 $\bar{x}$ を求めよ。

練習問題 3 の解答・解説

観測値は 3 個なので、

$n = 3$ である。

標本平均を計算する。

$$\begin{aligned}\bar{x} &= \frac{1.2 + 1.5 + 1.8}{3} \\ &= \frac{4.5}{3} \\ &= 1.5\end{aligned}$$

したがって、標本平均は $\bar{x} = 1.5$ である。

観測値が小数であっても、標本平均の定義は変わらない。観測した個数で割る。

練習問題 4

ある標本の大きさは $n = 5$ であり、観測値の総和は 50 である。

この標本の平均 $\bar{x}$ を求めよ。

練習問題 4 の解答・解説

標本平均は、観測値の総和を標本の大きさで割る。
観測値の総和が 50、標本の大きさが 5 なので、

$$\begin{aligned}\bar{x} &= \frac{50}{5} \\ &= 10\end{aligned}$$

したがって、標本平均は $\bar{x} = 10$ である。
平均を求めるときは、まず合計を考え、その合計を観測した個数 n で割る。

練習問題 5

次の 2 つの場合について、それぞれ標本の大きさ n と標本平均 $\bar{x}$ を求めよ。

- (1) 観測値が 7 だけである場合。
- (2) 観測値が 7, 7, 7 である場合。

練習問題 5 の解答・解説

(1) 観測値は 1 個なので、
 $n = 1$ である。
標本平均は、

$$\begin{aligned}\bar{x} &= \frac{7}{1} \\ &= 7\end{aligned}$$

したがって、 $n = 1$, $\bar{x} = 7$ である。
(2) 観測値は 3 個なので、 $n = 3$ である。
同じ値が 3 回観測されているが、観測数は 3 個である。

$$\begin{aligned}\bar{x} &= \frac{7 + 7 + 7}{3} \\ &= \frac{21}{3} \\ &= 7\end{aligned}$$

したがって、 $n = 3$, $\bar{x} = 7$ である。
2 つの場合は標本平均が同じ 7 になるが、標本の大きさは異なる。
値の種類数ではなく、実際に観測した回数を n とすることが重要である。

練習問題 6

架空の小規模モデルとして、値 0, 2, 4, 6 を持つ 4 対象を母集団とする。
各対象が等確率で選ばれ、選んだ対象を戻してから次を選ぶ独立な復元抽出を 2 回行ったところ、観測結果が 0, 0 となった。
母平均 μ と標本平均 $\bar{x}$ を求めよ。
また、「標本平均が母平均と一致しないため、この抽出は無作為抽出ではない」という主張が正しいか判定せよ。

練習問題 6 の解答・解説

母集団の値は 0, 2, 4, 6 である。

母平均は、母集団に含まれるすべての値の平均である。

$$\begin{aligned}\mu &= \frac{0 + 2 + 4 + 6}{4} \\ &= \frac{12}{4} \\ &= 3\end{aligned}$$

したがって、母平均は $\mu = 3$ である。

観測結果は 0, 0 であり、標本の大きさは $n = 2$ である。

標本平均は、

$$\begin{aligned}\bar{x} &= \frac{0 + 0}{2} \\ &= \frac{0}{2} \\ &= 0\end{aligned}$$

したがって、標本平均は $\bar{x} = 0$ である。

母平均と標本平均は一致していない。

しかし、「一致しないから無作為抽出ではない」という判断は誤りである。

無作為な抽出でも、取り出された標本によって標本平均は母平均と異なることがある。

この例では、各対象が等確率で選ばれ、復元抽出によって同じ対象が再び選ばれる可能性がある条件で考えている。

練習問題 7

互いに重ならない 2 つの群がある。

第 1 群は標本の大きさが $n = 2$ 、平均が 3 である。

第 2 群は標本の大きさが $n = 6$ 、平均が 7 である。

2 つの群を合わせた全体の平均を求めよ。

また、 $\frac{3+7}{2}$ と計算する誤りについて説明せよ。

練習問題 7 の解答・解説

群ごとの平均をそのまま平均すると、

$$\frac{3+7}{2} = 5$$

となる。

ただし、今回のように群平均が 3 と 7 で異なる場合には、観測数が等しいときだけ単純平均が使えるが、一般には群平均が等しい場合も単純平均の値が一致する例外があるため、原則として各群の総和を求めて総観測数で割る。

正しい方法では、まず各群の合計を求める。

第 1 群の合計は、 $2 \times 3 = 6$ である。

第 2 群の合計は、 $6 \times 7 = 42$ である。

全体の合計は、 $6 + 42 = 48$ となる。

全体の観測数は、 $2 + 6 = 8$ である。

したがって、全体の平均は、

$$\begin{aligned}\bar{x} &= \frac{48}{8} \\ &= 6\end{aligned}$$

である。

群平均だけを平均すると、観測数が少ない第 1 群と多い第 2 群を同じ重さで扱ってしまうため誤りである。

練習問題 8

架空の小規模モデルとして、値 2, 4, 6, 8 を持つ 4 対象の母集団を考える。

各対象を等確率で選び、選んだ対象を戻してから次を選ぶ独立な復元抽出を 3 回行う試行を 2 回行った。

その結果、標本 A では (2, 4, 6)、標本 B では (2, 2, 8) が得られた。

(1) 母平均 μ を求めよ。

(2) 標本 A、標本 B の標本平均 $\bar{x}_A$ 、 $\bar{x}_B$ をそれぞれ求めよ。

(3) 「抽出し直すと標本平均は必ず変わる」という主張が正しいか判定せよ。

また、抽出前の $\bar{X}$ と観測後の $\bar{x}_A$ 、 $\bar{x}_B$ の違いを説明せよ。

練習問題 8 の解答・解説

(1) 母集団の値は 2, 4, 6, 8 である。

母平均は、

$$\begin{aligned}\mu &= \frac{2+4+6+8}{4} \\ &= \frac{20}{4} \\ &= 5\end{aligned}$$

である。

したがって、母平均は $\mu = 5$ である。

(2) 標本 A は (2, 4, 6) である。

標本の大きさは、 $n = 3$ である。

標本平均は、

$$\begin{aligned}\bar{x}_A &= \frac{2 + 4 + 6}{3} \\ &= \frac{12}{3} \\ &= 4\end{aligned}$$

である。

標本 B は $(2, 2, 8)$ である。

同じ値 2 が 2 回含まれているが、観測した個数は 3 個である。

$$\begin{aligned}\bar{x}_B &= \frac{2 + 2 + 8}{3} \\ &= \frac{12}{3} \\ &= 4\end{aligned}$$

である。

したがって、 $\bar{x}_A = 4$ 、 $\bar{x}_B = 4$ である。

(3) 「抽出し直すと標本平均は必ず変わる」という主張は誤りである。

標本 A と標本 B は異なる標本であるが、どちらも標本平均は 4 で一致している。

抽出し直したとき、標本平均が変化する場合もあれば、同じになる場合もある。

また、同じ平均であっても同じ標本であるとは限らない。

抽出前は、まだどの値が選ばれるか決まっていないため、標本平均は確率変数として $\bar{X}$ と表す。

一方、実際に観測した後の標本 A、標本 B の平均は決まった数値なので、 $\bar{x}_A, \bar{x}_B$ と表す。

ここで、復元抽出では同じ対象が再び選ばれることがあるため、標本 B のように同じ値が複数回現れることがある。