

第 II 部 統計

5 平均と分散と標準偏差

ポイント

- 総和
- 平均

- 分散
- 標準偏差
- 標準化

5.1 はじめに

- 邪念をぬぐい、余計な意見を混ぜない方法は数学です。
- データの特性を表す値を統計量といいます。
- 有名な統計量を紹介します。

5.2 総和

定義

n 個の数値があり、 x_i を i 番目の値とする。全ての x_i を足し合わせた値を『総和』と呼び、 $\sum_{i=1}^n x_i$ という記号であらわす。

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

5.2.1 定数 (c) の総和

n 個の数値があり、 x_i を i 番目の値とする。全ての x_i が定数 (c) に等しい場合、 x_i の総和を $\sum_{i=1}^n c$ という記号であらわす。

$$\sum_{i=1}^n c = c + c + \cdots + c \quad (2)$$

5.2.2 定数 (a) 倍した x_i の総和

n 個の数値があり、 x_i を i 番目の値とする。 x_i を定数 (a) 倍した値 (ax_i) の総和を $\sum_{i=1}^n ax_i$ という記号であらわす。

$$\sum_{i=1}^n ax_i = ax_1 + ax_2 + \cdots + ax_n \quad (3)$$

5.2.3 和の総和

二つの数値の組が n 個あり、 (x_i, y_i) を i 番目の組とする。 x_i と y_i の和 $(x_i + y_i)$ の総和を $\sum_{i=1}^n (x_i + y_i)$ という記号であらわす。

$$\sum_{i=1}^n (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n) \quad (4)$$

5.2.4 二乗の総和

n 個の数値があり、 x_i を i 番目の値とする。 x_i を二乗した値 (x_i^2) の総和を $\sum_{i=1}^n x_i^2$ という記号であらわす。

$$\sum_{i=1}^n x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \quad (5)$$

問題 II-5-1

以下の式が成り立つことを示しなさい。ただし、 c は定数とする。

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

解例 II-5-1

$$\sum_{i=1}^n c = c + c + \cdots + c$$

c は n 個あるので

$$= nc$$

問題 II-5-2

以下の式が成り立つことを示しなさい。ただし、 a は定数とする。

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

解例 II-5-2

$$\begin{aligned}\sum_{i=1}^n ax_i &= ax_1 + ax_2 + \cdots + ax_n \\ &= a(x_1 + x_2 + \cdots + x_n) \\ &= a \sum_{i=1}^n x_i\end{aligned}$$

問題 II-5-3

以下の式が成り立つことを示しなさい。

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

解例 II-5-3

$$\begin{aligned}\sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n) \\&= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\&= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n) \\&= \sum_{i=1}^n x_i + \sum_{i=1}^n y_i\end{aligned}$$

問題 II-5-4

以下の式が成り立つことを示しなさい。

$$\sum_{i=1}^n (ax_i + c) = a \sum_{i=1}^n x_i + nc$$

解例 II-5 - 4

$$\begin{aligned}\sum_{i=1}^n (ax_i + c) &= \sum_{i=1}^n ax_i + \sum_{i=1}^n c \\ &= a \sum_{i=1}^n x_i + nc\end{aligned}$$

5.2.5 総和の公式

確認された総和の公式を示す。

$$\sum_{i=1}^n c = nc \quad ; c \text{ は定数} \quad (6)$$

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad ; a \text{ は定数} \quad (7)$$

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \quad (8)$$

5.3 平均

定義

n 個の数値があり、 x_i を i 番目の値とする。 x_i の総和を n で割った値を『算術平均』『相加平均』または単に『平均』と呼び $\bar{x}$ という記号であらわす。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

本講座では、基本的に『平均』とよぶ。 $\bar{x}$ は、「エックス・バー」と読む。

5.3.1 平均と総和の関係

ここで、平均

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

の両辺に n をかけると、

$$n\bar{x} = \cancel{n} \frac{1}{\cancel{n}} \sum_{i=1}^n x_i \quad (10)$$

$$n\bar{x} = \sum_{i=1}^n x_i \quad (11)$$

である。

数値例

表 1 の値は、ある小学校の児童の体重のデータである。この値を x_i とすると、

$$\sum_{i=1}^6 x_i = 21.3 + 24.2 + 27.7 + 31.1 + 34.4 + 39.2 \quad (12)$$

$$= 177.9 \quad (13)$$

$$\bar{x} = \frac{177.9}{6} = 29.65 \quad (14)$$

表 1 小学生の体重

i	x_i
1	21.3
2	24.2
3	27.7
4	31.1
5	34.4
6	39.2

5.4 平均からの偏差

四則演算の確認

$$(a - b)c = ac - bc \quad (15)$$

$$(a - b)(c - d) = a(c - d) - b(c - d) \quad (16)$$

$$= ac - ad - bc + bd \quad (17)$$

$$(a - b)^2 = aa - ab - ba + bb \quad (18)$$

$$= a^2 - 2ab + b^2 \quad (19)$$

定義

x_i と $\bar{x}$ との差を『平均からの偏差』という。

$$x_i - \bar{x} \quad (20)$$

『平均からの偏差』が

正の場合、平均より大きな値であることを意味し

負の場合、平均より小さな値であることを意味する。

5.4.1 『平均からの偏差』の総和

ここで、『平均からの偏差』の総和を求めてみる。

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i + \sum_{i=1}^n (-\bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} \quad (21)$$

$$= \sum_{i=1}^n x_i - n\bar{x} \quad (22)$$

$$= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0 \quad (23)$$

数値例

表 1 をもとに平均からの偏差を求めると、

$$x_1 - \bar{x} = 21.3 - 29.65 = -8.35 \quad (24)$$

$$x_2 - \bar{x} = 24.2 - 29.65 = -5.45 \quad (25)$$

$$x_3 - \bar{x} = 27.7 - 29.65 = -1.95 \quad (26)$$

$$x_4 - \bar{x} = 31.1 - 29.65 = 1.45 \quad (27)$$

$$x_5 - \bar{x} = 34.4 - 29.65 = 4.75 \quad (28)$$

$$x_6 - \bar{x} = 39.2 - 29.65 = 9.55 \quad (29)$$

$$\sum_{i=1}^6 (x_i - \bar{x}) = 0 \quad (30)$$

表 2 平均からの偏差

i	x_i	$x_i - \bar{x}$
1	21.3	-8.35
2	24.2	-5.45
3	27.7	-1.95
4	31.1	1.45
5	34.4	4.75
6	39.2	9.55

『平均からの偏差』は符号を持つため、足し合わせると正負が打ち消しあい、総和は必ず 0 になる。このような問題を『符号の問題』という。符号の問題により、『平均からの偏差』の総和では、ばらつきの尺度とすることはできない。そこで、『平均からの偏差』を二乗し平均をとることでばらつきの尺度とする。

定義

「『平均からの偏差』の二乗」の平均を『分散』と呼び V_x という記号であらわす。

$$V_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (31)$$

数値例

表 1 をもとに『平均からの偏差』の二乗を求めると、

$$(x_1 - \bar{x})^2 = (-8.35)^2 = 69.7225 \quad (32)$$

$$(x_2 - \bar{x})^2 = (-5.45)^2 = 29.7025 \quad (33)$$

$$(x_3 - \bar{x})^2 = (-1.95)^2 = 3.8025 \quad (34)$$

$$(x_4 - \bar{x})^2 = (1.45)^2 = 2.1025 \quad (35)$$

$$(x_5 - \bar{x})^2 = (4.75)^2 = 22.5625 \quad (36)$$

$$(x_6 - \bar{x})^2 = (9.55)^2 = 91.2025 \quad (37)$$

表 3 『平均からの偏差』の二乗

i	x_i	$(x_i - \bar{x})^2$
1	21.30	69.7225
2	24.20	29.7025
3	27.70	3.8025
4	31.10	2.1025
5	34.40	22.5625
6	39.20	91.2025
Total		219.0950
V_x		36.515 833

問題 II-5-5

次の式を展開しなさい。

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

解例 II-5-5

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

数値例

$$V_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (38)$$

$$= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (39)$$

である。この結果を利用して表 1 をもとに分散を求めると、

$$V_x = 915.64 - 29.65^2 \quad (40)$$

$$= 36.515\,833 \quad (41)$$

表 4 x_i の二乗

i	x_i	x_i^2
1	21.3	453.69
2	24.2	585.64
3	27.7	767.29
4	31.1	967.21
5	34.4	1183.36
6	39.2	1536.64
平均	29.65	915.64

定義

分散の正の平方根を『標準偏差』と呼び σ_x という記号であらわす。

$$\sigma_x = \sqrt{V_x} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (42)$$

σ はギリシャ文字シグマの小文字である。

分散の正の平方根が標準偏差なので、標準偏差の二乗は分散である。

$$\sigma_x^2 = V_x \quad (43)$$

問題 II-5-6

$z_i = \frac{x_i - \bar{x}}{\sigma_x}$, $\sigma_x > 0$ とする。以下の値を求めよ。ただし、 $\bar{x}$ は x の平均, σ_x は x の標準偏差とする。

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$$

解例 II-5-6

$$\begin{aligned}\bar{z} &= \frac{1}{n} \sum_{i=1}^n z_i \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \\ &= \frac{1}{n} \frac{1}{\sigma_x} \sum_{i=1}^n (x_i - \bar{x}) \\ &= 0\end{aligned}$$

問題 II-5-7

$z_i = \frac{x_i - \bar{x}}{\sigma_x}$, $\sigma_x > 0$ とする。以下の値を求めよ。ただし、 $\bar{x}$ は x の平均, σ_x は x の標準偏差とする。

$$V_z = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2$$

解例 II-5 - 7

$$\begin{aligned} V_z &= \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2 = \frac{1}{n} \sum_{i=1}^n z_i^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right)^2 = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma_x^2} \\ &= \frac{1}{\sigma_x^2} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{V_x} V_x \\ &= 1 \end{aligned}$$

5.4.2 標準化

『平均からの偏差』を標準偏差で割ることを標準化といい、標準化された変数を z_i であらわす。

$$z_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (44)$$

標準化すると z_i の平均は 0、標準偏差は 1 になる。

値を標準化すると、それぞれの値のおおよその位置が把握可能となる。

表 5 標準化した表 1 の値

i	x_i	z_i
1	21.3	-1.38
2	24.2	-0.90
3	27.7	-0.32
4	31.1	0.24
5	34.4	0.79
6	39.2	1.58

問題 II-5-9

表 6 はある小学校の児童の身長データである。身長を x_i として、平均, 分散, 標準偏差を求め、それぞれの値を標準化しなさい。小数点以下の処理は各自の判断に任せる。

算出に際して、何らかのアプリを使用することは自由である。

表 6 小学生の身長

i	身長
1	116.5
2	122.6
3	128.5
4	133.5
5	138.6
6	145.0

表7 値と平均からの偏差と標準化された値

i	x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	z_x	z_y
1	21.3	116.5	-8.35	-14.28	-1.38	-1.50
2	24.2	122.6	-5.45	-8.18	-0.90	-0.86
3	27.7	128.5	-1.95	-2.28	-0.32	-0.24
4	31.1	133.5	1.45	2.72	0.24	0.28
5	34.4	138.6	4.75	7.82	0.79	0.82
6	39.2	145	9.55	14.22	1.58	1.49

問題 II-5-9

表 8 はある小学校の児童の 50m 走のタイムのデータである。タイムを x_i として、平均, 分散, 標準偏差を求め、それぞれの値を標準化しなさい。小数点以下の処理は各自の判断に任せる。

算出に際して、何らかのアプリを使用することは自由である。

表 8 50m 走のタイム

i	タイム
1	11.6
2	10.7
3	10.0
4	9.7
5	9.4
6	9.0

解例 II-5-9

$$\bar{x} = 10.066\,667$$

$$V_x = 0.745\,556$$

$$\sigma_x = 0.863\,456$$

表 9 50m 走のタイムの z 値

i	x_i	z_i
1	11.6	1.775 810
2	10.7	0.733 487
3	10.0	-0.077 209
4	9.7	-0.424 650
5	9.4	-0.772 091
6	9.0	-1.235 346

5.5 まとめ

- $\sum_{i=1}^n x_i$ を総和という。
- 平均 ($\bar{x}$) は総和を n で割った商である。

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

- 「『平均からの偏差 ($x_i - \bar{x}$) 』の二乗」の平均を分散 (V_x) という。

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

- 分散 (V_x) の正の平方根を標準偏差 (σ_x) という。