

第4章 データの分析

35 データの代表値

1 データの代表値

- ① **平均値** \bar{x} 変量 x についてのデータの値が、 n 個の値 x_1, x_2, \dots, x_n であるとき、それらの総和を n で割ったもの

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

- ② **中央値** データを値の大きさの順に並べたとき、中央の位置にくる値
(メアン) データの大きさが偶数のときは、中央に 2 つの値が並ぶが、
その場合は、その 2 つの値の平均値とする。
③ **最頻値** データにおいて、最も個数の多い値
(モード)



355 次のデータは、ある 9 人がけんかを何回できたかを記録したものである。

3, 4, 4, 4, 7, 7, 8, 8, 9 (単位は回)

- (1) このデータの平均値を求めよ。
- (2) このデータの中央値を求めよ。
- (3) このデータの最頻値を求めよ。

*356 次の表は、ある野球チームが 20 試合したときのそれぞれの得点を、得点別に試合数をまとめたものである。この野球チームの得点について、下の問い合わせよ。

得点	0	1	2	3	4	5	6	計
試合数	2	3	5	2	3	3	2	20

- (1) このデータの平均値を求めよ。
- (2) このデータの中央値を求めよ。
- (3) このデータの最頻値を求めよ。



357 次のデータは、ある商店の 10 日間の弁当の販売数である。

72 70 77 65 86 63 65 59 79 a (単位は個)

このデータの平均値が 72 個であるとき、 a の値を求めよ。
また、このときのデータの中央値を求めよ。

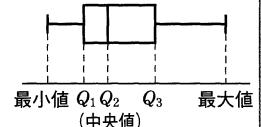
36 データの散らばり

1 データの散らばり

- ① **範囲** データの最大値と最小値の差

② **四分位数** データを値の大きさの順に並べたとき、4 等分する位置にある値。小さい方から第 1 四分位数、第 2 四分位数、第 3 四分位数といい、これらを順に Q_1, Q_2, Q_3 で表す。

- ③ **四分位範囲** $Q_3 - Q_1$ **四分位偏差** $\frac{Q_3 - Q_1}{2}$



- ④ **箱ひげ図** データの分布をみるための図

- ⑤ **分散** 偏差 (データの各値 x_k とその平均値 \bar{x} との差) の 2 乗の平均値

$$s^2 = \frac{1}{n} \{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}$$

$$= (x_k^2 \text{ のデータの平均値}) - (x_k \text{ のデータの平均値})^2$$

- ⑥ **標準偏差** 分散の正の平方根

$$s = \sqrt{\frac{1}{n} \{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}}$$



*358 次のデータは、P 市と Q 市のある 10 日間の降雨量を調べたものである。

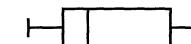
P 市	4	5	12	14	16	18	22	28	29	32
Q 市	16	70	54	62	14	28	18	20	76	12

(単位は mm)

- (1) P 市、Q 市のデータの範囲、四分位範囲、四分位偏差を求めよ。
- (2) P 市、Q 市のデータの箱ひげ図を並べてかけ。
- (3) P 市、Q 市のデータの平均値、分散、標準偏差を求めよ。ただし、小数第 1 位を四捨五入せよ。
- (4) P 市、Q 市のデータについて、四分位範囲、箱ひげ図、標準偏差から散らばりの度合を比較せよ。

*359 右の図は、ある高校 1 年生 320 人に行った実力

テストの得点のデータの箱ひげ図である。この 30 40 50 60 70 80 90 (点) 箱ひげ図から読み取ることを次の①～③から選べ。



- ① 30 点台の生徒は 80 人である。
- ② 50 点以上の生徒は 240 人以上いる。
- ③ 60 点未満の生徒は半数以上いる。

37

データの相関

1 データの相関

① 相関関係 2つの変量のデータにおいて、一方が増えると他方が増える（減る）傾向が認められるとき、2つの変量の間に正の（負の）相関関係があるという。

② 共分散 x の偏差と y の偏差の積 $(x_k - \bar{x})(y_k - \bar{y})$ の平均値

$$s_{xy} = \frac{1}{n} \{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})\}$$

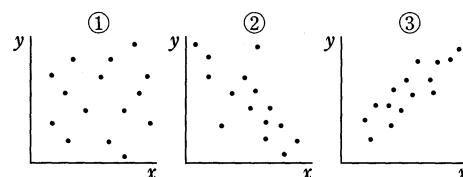
③ 相関係数 $r = \frac{s_{xy}}{(x \text{ の標準偏差})(y \text{ の標準偏差})} \quad (-1 \leq r \leq 1)$

正の相関関係があるとき $r > 0$ 、負の相関関係があるとき $r < 0$

A

- 360 右の①, ②, ③はある2つの変量 x , y のデータについての散布図である。データ①, ②, ③の x と y の相関係数は、

0.88, 0.03, -0.72 のいずれかである。各データの相関係数を答えよ。



- *361 次の表は10人の生徒の右手の握力と左手の握力を測定した結果である。

生徒の番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
右手の握力 (kg)	36	42	35	33	38	32	39	40	34	41
左手の握力 (kg)	27	39	35	25	41	23	43	31	29	37

- (1) 散布図をかけ。
(2) 相関係数 r を求めよ。ただし、小数第3位を四捨五入せよ。
(3) 右手の握力と左手の握力の間には、どのような相関関係があるか。

Aのまとめ

- 362 次の表は、5人の生徒に漢字と英単語のテストを行った得点の結果である。相関係数を求めて、漢字と英単語の得点の間にはどのような相関関係があるかをいえ。

生徒の番号	①	②	③	④	⑤
漢字 (点)	4	6	8	5	9
英語 (点)	2	6	7	7	6

38

第4章 演習問題

B

- *363 次のデータは、ある家庭の1週間ごとに集まるゴミの重量を6週間記録したものである。 6.4 2.4 4.6 4.0 5.4 4.8 (単位は kg)

- (1) このデータの中央値と平均値を求めよ。
(2) 上記の6個の数値のうち、1個が誤りであることがわかった。正しい数値に基づく中央値と平均値は、それぞれ 5.1 kg と 4.8 kg であるという。誤っている数値を選び、正しい数値を求めよ。

- *364 ある高校1年生の期末試験の結果が次の通りであった。

1組は20人受け、得点の平均値は60点、標準偏差は10点

2組は30人受け、得点の平均値は50点、標準偏差は20点

このとき、1組と2組を合わせた全体の得点の平均値と標準偏差を求めよ。ただし、小数第1位を四捨五入して整数で答えよ。

- *365 次のデータは、ある地域の6日間の降雨量を記録したものである。

14, 11, 10, 18, 16, 9 (単位は mm)

- (1) このデータの平均値を求めよ。
(2) このデータには記録ミスがあり、18 mm は正しくは 17 mm, 9 mm は正しくは 10 mm であった。この誤りを修正したときのデータの平均値、分散は、修正前から増加する、減少する、変化しないを答えよ。

- 366 右の表は、10人の生徒について行った数

階級 (点) 以上	数学 (人)	英語 (人)
30~39	0	2
40~49	3	3
50~59	4	3
60~69	3	2
合計	10	10

学と英語のテストの得点のデータを、度数分布表にまとめたものである。また、下の表は、10人の生徒それぞれについて、テ

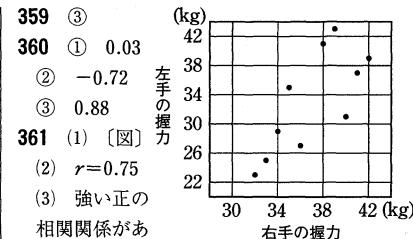
ストの得点のデータをまとめたものである。

ただし、 $a < b$, $c < d$ とする。

生徒の番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	平均値
数学 (点)	41	a	61	57	63	43	b	59	54	50	54
英語 (点)	39	47	35	c	67	d	53	65	55	48	51

- (1) 数学の得点のデータの範囲が25点であるとき、 a , b の値を求めよ。

- (2) 英語の得点のデータの中央値を求めよ。



$$(2) r = \frac{16.4}{\sqrt{11} \times \sqrt{44}}$$

362 相関係数 0.57, 漢字と英単語の得点の間には弱い正の相関関係があると考えられる

$$\left[\frac{1.96}{\sqrt{3.44} \sqrt{3.44}} \right]$$

- 363 (1) 中央値 4.7 kg, 平均値 4.6 kg
 (2) 誤りの数値 4.6 kg, 正しい数値 5.8 kg
 [(2) 平均が 4.6 kg から 4.8 kg と 0.2 kg 増加するから全体として 1.2 kg 増加]

- 364 平均値 54 点, 標準偏差 17 点
 [全体の平均値は $(60 \times 20 + 50 \times 30) \div 50$
 1組と2組の得点の2乗の平均値をそれぞれ
 a, b とすると $a - 60^2 = 10^2, b - 50^2 = 20^2$
 50人の得点の2乗の和は $a \times 20 + b \times 30$]

- 365 (1) 13 mm
 (2) 平均値は変化しない, 分散は減少する

- 366 (1) $a=46, b=66$ (2) 50.5 点
 [(1) $a+b=112$, 度数分布表と得点の表から $40 \leq a \leq 49, 60 \leq b \leq 69$ 数学の得点のデータの最小値は 40, 41 のいずれかである
 (2) $c+d=101, 40 \leq c \leq 49, 50 \leq d \leq 59$
 中央値は, 小さい方から 5 番目の値と 6 番目の値の平均値である。すなわち, 40~49 の階級の最大値 M と 50~59 の階級の最小値 m の平均値である。 M の値は 48, 49 のいずれかである]

- 355 (1) 6 回 (2) 7 回 (3) 4 回
 356 (1) 2.9 点 (2) 2.5 点 (3) 2 点
 357 $\alpha=84$, 中央値 71 個
 358 (1) 範囲, 四分位範囲, 四分位偏差の順
 P市…28 mm, 16 mm, 8 mm
 Q市…64 mm, 46 mm, 23 mm (2) [図]
 (3) 平均値, 分散, 標準偏差の順
 P市…18 mm, 85, 9 mm
 Q市…37 mm, 585, 24 mm
 (4) 四分位範囲, 箱ひげ図, 標準偏差のいざれをみても, P市よりもQ市の方が散らばりの度合が大きい。

