

データの分析 まとめ

解答（全27問）

氏名 _____

得点 _____ / 27点

DRILL

まとめテスト（標準）

基本（1～15）

1

中央値（データ数が奇数）

9人のバドミントン部員のサーブ練習の成功本数：

15, 22, 19, 25, 18, 20, 24, 17, 21。中央値を求めよ。

中央値 20

小さい順に並べる：15, 17, 18, 19, 20, 21, 22, 24, 25

9個なので5番目が中央値。

中央値 20

2

四分位数（データ数が奇数・11個）

次の11個のデータについて、 $Q_1 \cdot Q_3 \cdot$ 四分位範囲を求めよ：

5, 9, 12, 15, 18, 20, 23, 26, 29, 31, 34

 $Q_1 = \underline{12}$ $Q_3 = \underline{29}$ 四分位範囲 $\underline{17}$

すでに小さい順。11個は奇数なので、中央値（6番目）20を除いて2組に分ける。

下組（1～5番目）：5, 9, 12, 15, 18 → 中央値（3番目） $Q_1 = 12$ 上組（7～11番目）：23, 26, 29, 31, 34 → 中央値（3番目） $Q_3 = 29$ 四分位範囲は $29 - 12 = 17$ $Q_1 = 12$ 、 $Q_3 = 29$ 、四分位範囲 17

3

四分位数（データ数が偶数・8個）

次の8個のデータについて、 $Q_1 \cdot Q_3 \cdot$ 四分位範囲を求めよ：4, 9, 11, 14, 16, 19, 21, 25 $Q_1 = \underline{10}$ $Q_3 = \underline{20}$ 四分位範囲 $\underline{10}$

8個は偶数なので、そのまま半分（4個ずつ）に分ける。

下組：4, 9, 11, 14 → $Q_1 = \frac{9 + 11}{2} = 10$ 上組：16, 19, 21, 25 → $Q_3 = \frac{19 + 21}{2} = 20$ 四分位範囲は $20 - 10 = 10$ $Q_1 = 10$ 、 $Q_3 = 20$ 、四分位範囲 10

4

平均値と中央値の違い

次の7個のデータについて、平均値と中央値を求めよ：3, 4, 5, 5, 6, 7, 40

平均値 $\underline{10}$ 中央値 $\underline{5}$

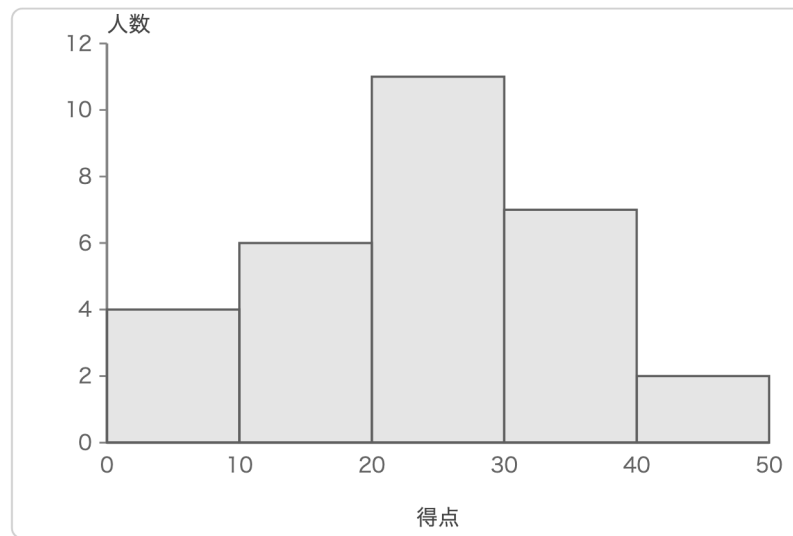
すでに小さい順。

平均値： $\frac{3 + 4 + 5 + 5 + 6 + 7 + 40}{7} = \frac{70}{7} = 10$

中央値（4番目）：5

平均値 10、中央値 5。最後の40という外れ値に平均値だけ引っばられている。

30人の小テストの得点の分布を表したヒストグラムである。最頻値が含まれる階級の階級値を求めよ。



階級値 25

度数がいちばん多い階級は「20以上30未満」で度数11。

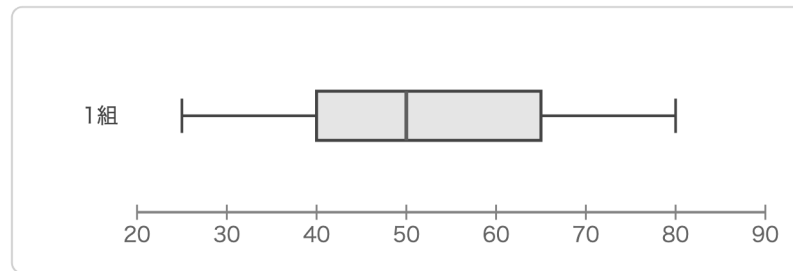
この階級の階級値は $\frac{20 + 30}{2} = 25$

階級値 25

6

箱ひげ図の読み取り

1組の小テスト結果を表した箱ひげ図である。中央値と第3四分位数を読み取れ。



中央値 50 第3四分位数 65

箱ひげ図から、箱の中の線が中央値、箱の右端が第3四分位数。目盛は10刻みなので、中央値は50の目盛線上、第3四分位数は60と70の目盛のちょうど中間の65と読み取れる。

中央値 50、第3四分位数 65

7

分散・標準偏差（基本）

データ7, 7, 13, 13について、平均・分散・標準偏差を求めよ。

平均 10 分散 9 標準偏差 3

$$\text{平均} = \frac{7 + 7 + 13 + 13}{4} = \frac{40}{4} = 10$$

偏差は-3, -3, 3, 3。2乗の和は $9 + 9 + 9 + 9 = 36$ 。

$$\text{分散} = \frac{36}{4} = 9, \text{ 標準偏差} = \sqrt{9} = 3$$

平均 10、分散 9、標準偏差 3

8

公式「 $\overline{x^2} - (\bar{x})^2$ 」を使う

データ3, 3, 3, 11, 11, 11 (6個) の分散を、 $\overline{x^2} - (\bar{x})^2$ の公式で求め、標準偏差も答えよ。

分散 16 標準偏差 4

$$\text{平均} = \frac{3 + 3 + 3 + 11 + 11 + 11}{6} = \frac{42}{6} = 7$$

x^2 は9, 9, 9, 121, 121, 121。和は $9 \times 3 + 121 \times 3 = 27 + 363 = 390$ 。

$$\overline{x^2} = \frac{390}{6} = 65$$

$$V = \overline{x^2} - (\bar{x})^2 = 65 - 7^2 = 65 - 49 = 16、標準偏差 = \sqrt{16} = 4$$

分散 16、標準偏差 4

9

全員に同じ数を足す→分散はどうなる？

あるデータの平均は18、分散は9である。全員に7を加えると、新しいデータの平均・分散・標準偏差はいくらになるか。

平均 25 分散 9 標準偏差 3

平均は $18 + 7 = 25$ 。偏差の形は変わらないので分散は変わらず9、標準偏差も変わらず $\sqrt{9} = 3$ 。

平均 25、分散 9、標準偏差 3

10

全員をb倍する→分散は b^2 倍

あるデータの平均は8、分散は5である。全員を3倍すると、新しいデータの平均・分散・標準偏差はいくらになるか。

平均 24 分散 45 標準偏差 $3\sqrt{5}$

平均は $8 \times 3 = 24$ 。分散は $3^2 = 9$ 倍なので $5 \times 9 = 45$ 。標準偏差は3倍なので $\sqrt{5} \times 3 = 3\sqrt{5}$ (または $\sqrt{45} = \sqrt{9 \times 5} = 3\sqrt{5}$ と直接求めてもよい)。

平均 24、分散 45、標準偏差 $3\sqrt{5}$ (≈ 6.71)

4人の生徒の練習時間 x （時間）と、その後の小テストの得点 y （点）は $x = 1, 3, 5, 7$ 、 $y = 2, 5, 6, 11$ であった（生徒の順は対応している）。共分散 S_{xy} を求めよ。

$$S_{xy} = \underline{7}$$

$$\bar{x} = \frac{1+3+5+7}{4} = 4, \bar{y} = \frac{2+5+6+11}{4} = 6$$

$$x - \bar{x} : -3, -1, 1, 3$$

$$y - \bar{y} : -4, -1, 0, 5$$

$$\text{積} : 12, 1, 0, 15 \quad \text{合計} 28$$

$$S_{xy} = 28 \div 4 = 7$$

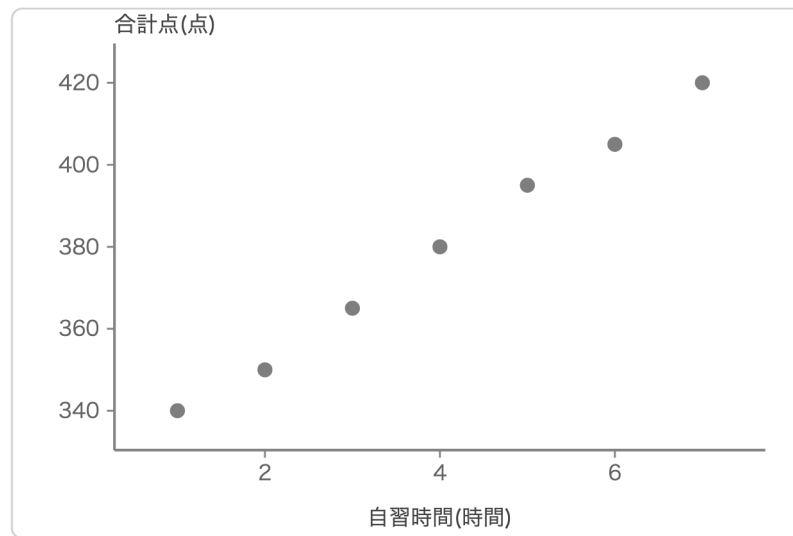
あるデータについて、 x の標準偏差 $S_x = 2.5$ 、 y の標準偏差 $S_y = 4$ 、共分散 $S_{xy} = -7$ であった。相関係数 r を求めよ。

$$r = \underline{-0.7}$$

$$r = \frac{S_{xy}}{S_x \times S_y} = \frac{-7}{2.5 \times 4} = \frac{-7}{10}$$

$$r = -0.7$$

次の散布図は、7人の1日の自習時間（時間）と定期テストの合計点（点）の関係を表したものである。相関の種類を「正」「負」「なし」のいずれかで答えよ。



相関の種類 (**正**) / 負 / なし

自習時間が長い生徒ほど、合計点が高い傾向にある。点がおおむね右上がりに分布している。

正の相関

次の値0.95, -1.2, 0, 1.05, -0.6のうち、相関係数の値としてありえないものを、小さい順にコンマ区切りで全て答えよ。

ありえない値 -1.2, 1.05

相関係数は必ず $-1 \leq r \leq 1$ の範囲におさまる。

-1.2はこの範囲より小さく、1.05はこの範囲より大きいのでありえない。0.95, 0, -0.6はどれも範囲内なのであり得る。

ありえない値は-1.2, 1.05

友達が「このコインは公正だ（表が出る確率0.5）」と言うので、20回投げてみたところ表が17回出た。公正なコインを20回投げる実験を1000回シミュレーションした結果（の一部）を下の表に示す。「20回中17回以上表が出る」割合（%）を求め、5%を基準に「このコインは公正である」という仮説を棄却できるか判断せよ。

表の枚数	10	11	12	13	14	15	16	17	18	19	20
1000回中の回数	176	160	120	74	37	15	5	1	0	0	0

割合 0.1 % 判断 (**棄却できる** / 棄却できない)

表から、17回・18回・19回・20回が出た回数を足し合わせる。

$$1 + 0 + 0 + 0 = 1 \text{ (1000回中1回)}$$

割合にすると $1 \div 1000 = 0.001$ 、つまり**0.1%**。

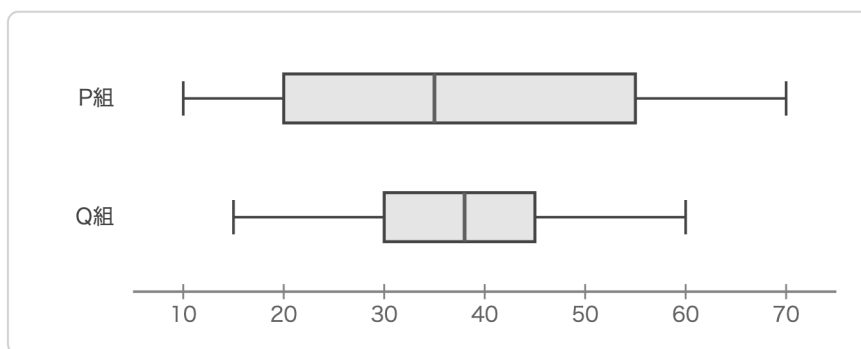
0.1% < 5%なので、「公正なコインで17回以上表が出る」ことは非常にめずらしいと判断できる。したがって、「このコインは公正である」という仮説は疑わしいとして**棄却できる**。

17回以上表が出る割合は0.1% (5%より小さい) → 仮説を棄却できる

16

箱ひげ図の比較・四分位範囲

P組・Q組の小テスト結果を表した箱ひげ図である。それぞれの四分位範囲を求め、どちらの組の方が成績が「そろっている」（散らばりが小さい）と言えるか、P・Qの記号で答えよ。



P組の四分位範囲 35 Q組の四分位範囲 15 そろっている組 Q

$$\text{P組} : Q_3 - Q_1 = 55 - 20 = 35$$

$$\text{Q組} : Q_3 - Q_1 = 45 - 30 = 15$$

四分位範囲が短いほど、箱の中（真ん中50%）のデータが密集していて散らばりが小さい。
Q組の方が四分位範囲がずっと短いので、Q組の方が成績がそろっている。

P組の四分位範囲 35、Q組の四分位範囲 15。そろっているのはQ組。

あるデータの平均は12、分散は8である。全員を -2 倍してから5を引いた新しいデータ ($y = -2x - 5$) の平均・分散・標準偏差を求めよ。

平均 -29 分散 32 標準偏差 $4\sqrt{2}$

平均は「 -2 倍してから5を引く」をそのまま平均に適用する： $12 \times (-2) - 5 = -24 - 5 = -29$ 。

分散は「 -2 倍」の部分だけが効き、「 -5 」(定数を引く)の部分は無関係。倍率の符号に関係なく2乗するので $(-2)^2 = 4$ 倍： $8 \times 4 = 32$ 。

標準偏差は $\sqrt{32} = \sqrt{16 \times 2} = 4\sqrt{2}$ (標準偏差は必ず正の値になる)。

平均 -29 、分散 32 、標準偏差 $4\sqrt{2}$ (≈ 5.66)

生徒15人のグループAの小テストの平均点は72点、生徒25人のグループBの平均点は84点であった。AとBを合わせた全体の平均点を求めよ。

全体の平均点 79.5

合計人数は $15 + 25 = 40$ 人。平均点は「人数の重みつき平均」で計算する (人数が違うグループを単純に $\frac{72 + 84}{2}$ と平均してはいけない)。

合計点はAの合計 + Bの合計 = $15 \times 72 + 25 \times 84 = 1080 + 2100 = 3180$ 点。
 全体の平均点 = $\frac{3180}{40} = 79.5$ 点。

全体の平均点 79.5点

生徒6人のグループAは平均5点・分散3、生徒4人のグループBは平均10点・分散3であった。AとBを合わせた10人全体の平均点と分散を求めよ。

全体の平均点 7 全体の分散 9

$$\text{全体の平均点} = \frac{6 \times 5 + 4 \times 10}{10} = \frac{30 + 40}{10} = \frac{70}{10} = 7 \text{点。}$$

分散は「各グループの分散をそのまま平均する」だけでは求まらない。各グループの平均が全体の平均からずれている分も加味する必要があるので、 $\overline{x^2}$ を使う方法で確認する。公式 $V = \overline{x^2} - (\bar{x})^2$ を逆に使うと $\overline{x^2} = V + (\bar{x})^2$ 。

$$\text{Aグループ: } \overline{x^2}_A = 3 + 5^2 = 3 + 25 = 28, \text{ 合計} = 6 \times 28 = 168$$

$$\text{Bグループ: } \overline{x^2}_B = 3 + 10^2 = 3 + 100 = 103, \text{ 合計} = 4 \times 103 = 412$$

$$\text{全体の } \overline{x^2} = \frac{168 + 412}{10} = \frac{580}{10} = 58$$

$$\text{全体の分散} = \overline{x^2} - (\bar{x})^2 = 58 - 7^2 = 58 - 49 = 9$$

全体の平均点 7、全体の分散 9

5人の生徒の1回目のテスト得点 x （点）と2回目のテスト得点 y （点）は $x = 10, 20, 30, 40, 50$ 、 $y = 30, 10, 50, 20, 40$ であった（生徒の順は対応している）。共分散 S_{xy} と相関係数 r を求めよ。

$$S_{xy} = \underline{60} \quad r = \underline{0.3}$$

x, y はどちらも $\{10, 20, 30, 40, 50\}$ を並べ替えただけの、同じ数の集まりなので、平均も分散も等しい。

$$\bar{x} = \bar{y} = 30$$

$$x - \bar{x} : -20, -10, 0, 10, 20$$

$$y - \bar{y} : 0, -20, 20, -10, 10$$

$$\text{積} : 0, 200, 0, -100, 200 \quad \text{合計} 300$$

$$S_{xy} = 300 \div 5 = 60$$

x の分散 $= (400 + 100 + 0 + 100 + 400) \div 5 = 200$ 。 x, y は同じ数の集まりなので y の分散も200、よって $S_x = S_y = \sqrt{200}$ 。

$$r = \frac{60}{\sqrt{200} \times \sqrt{200}} = \frac{60}{200}$$

$S_{xy} = 60$ 、 $r = 0.3$ （正の相関だが、1回目と2回目の得点の並びがかなり入れ替わっている
ので、弱い相関にとどまっている）

相関係数についての次の記述のうち、正しいものを全て選び、番号を小さい順にコンマ区切りで答えよ。

1. 単位を変えると相関係数の値は必ず変化する
2. 相関係数は常に -1 以上 1 以下の値をとる
3. 外れ値が1つ混ざっても相関係数の値は変化しない
4. 相関係数の絶対値が 1 に近いほど、直線的な関連が強い

正しい記述の番号 2,4

1は誤り：単位を変えても（正の数倍・平行移動しても）点の並び方の「形」は変わらないので、相関係数の値は変わらない。

2は正しい：相関係数は必ず $-1 \leq r \leq 1$ の範囲におさまる。

3は誤り：相関係数は外れ値の影響を受けやすい。傾向から大きく外れた点が1つ混ざるだけで、値は大きく変わってしまう。

4は正しい： $r = 1$ は完全な正の相関（点が右上がりの一直線）、 $r = -1$ は完全な負の相関を表すので、絶対値が 1 に近いほど点は一直線に近づく（直線的な関連が強い）。

正しい記述は2, 4

友達が持ってきたサイコロについて「6の目が出やすい」と主張したので、40回投げてみたところ6の目が11回出た。次の表は、公正なサイコロ（6の目が出る確率 $\frac{1}{6}$ ）で40回投げる実験を1000回シミュレーションした結果（の一部）である。「40回中11回以上6の目が出る」割合（%）を求め、5%を基準に「このサイコロは公正である」という仮説を棄却できるか判断せよ。

6の 目の 回数	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1000 回中 の回 数	167	162	134	95	59	32	16	7	3	1	0	0	0	0	0

割合 5.9 % 判断 （棄却できる / **棄却できない**）

表から、11回以上（11回～20回）が出た回数を足し合わせる。

$$32 + 16 + 7 + 3 + 1 + 0 + 0 + 0 + 0 + 0 = 59 \text{ (1000回中59回)}$$

割合は $59 \div 1000 = 0.059$ 、つまり**5.9%**。

5.9% > 5%なので、「公正なサイコロで11回以上6の目が出る」ことは、それほどめずらしいとは言えない。したがって、「このサイコロは公正である」という仮説は**棄却できない**（＝今回の40回・11回という結果だけでは、友達の主張を裏づける十分な根拠があるとは言えない）。

11回以上6の目が出る割合は5.9%（5%より大きい） → 仮説を棄却できない

あるゲームアプリのユーザー13人の1週間のプレイ時間（分、小さい順）：

12, 15, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 75。 Q_1 ・ Q_3 ・上側の基準値を求め、75分が外れ値と言えるか判定せよ。

$Q_1 = \underline{19}$ $Q_3 = \underline{33}$ 上側の基準値 $\underline{54}$ 判定 $\underline{\text{外れ値}}$

すでに小さい順。13個は奇数なので、中央値（7番目）26を除いて2組に分ける。

下組（1～6番目）：12, 15, 18, 20, 22, 24 → 6個（偶数）なので3番目と4番目の平均

$$Q_1 = \frac{18 + 20}{2} = 19$$

上組（8～13番目）：28, 30, 32, 34, 36, 75 → 3番目と4番目の平均 $Q_3 = \frac{32 + 34}{2} = 33$

四分位範囲： $IQR = 33 - 19 = 14$

上側の基準値： $Q_3 + 1.5 \times IQR = 33 + 1.5 \times 14 = 33 + 21 = 54$

75は上側の基準値54より大きい。

上側の基準値は54で、75分はこれを超えているので外れ値と判定される

あるクラス20人の小テストの平均点は50点、標準偏差は6点であった。全員の点数を1.5倍にしたあと、10点を一律に加える特別採点を行った（先に1.5倍、その後に+10）。最終的な平均点・分散・標準偏差を求めよ。

最終的な平均点 85 分散 81 標準偏差 9

もとの点数を x とすると、最終的な点数は $y = 1.5x + 10$ （先に1.5倍、その後に+10すると、この順に式になる）。

平均： $1.5 \times 50 + 10 = 75 + 10 = 85$ 点。

分散：もとの分散は標準偏差6の2乗で $V = 6^2 = 36$ 。「+10」の部分は分散に無関係で、「1.5倍」の部分だけが $1.5^2 = 2.25$ 倍として効く。

$$36 \times 2.25 = 81$$

標準偏差： $\sqrt{81} = 9$ 。

最終的な平均点 85点、分散 81、標準偏差 9

9人が参加したマラソン大会の記録（分、すでに小さい順に並んでいる）：

38, 40, 42, 44, 46, 48, 50, 52, 90。(1) 平均値、(2) Q_1 ・中央値・ Q_3 ・四分位範囲、(3) $1.5 \times \text{IQR}$ 基準で90分の記録が外れ値と言えるかを求めよ。

平均値 50 $Q_1 =$ 41 中央値 46 $Q_3 =$ 51 四分位範囲 10 上側の基準値 66 判定 外れ値

$$(1) \text{ 平均値 : } \frac{38 + 40 + 42 + 44 + 46 + 48 + 50 + 52 + 90}{9} = \frac{450}{9} = 50 \text{ 分}$$

(2) 9個は奇数なので、中央値（5番目）46を除いて2組に分ける。

$$\text{下組 (1~4番目) : } 38, 40, 42, 44 \rightarrow Q_1 = \frac{40 + 42}{2} = 41$$

$$\text{上組 (6~9番目) : } 48, 50, 52, 90 \rightarrow Q_3 = \frac{50 + 52}{2} = 51$$

$$\text{四分位範囲 : } 51 - 41 = 10$$

$$(3) \text{ 上側の基準値 : } Q_3 + 1.5 \times \text{IQR} = 51 + 1.5 \times 10 = 51 + 15 = 66$$

90は上側の基準値66より大きいので、外れ値と判定される。

この章で学んだ「平均値は外れ値に引っぱられる（中央値46より平均値50の方が高い）」「四分位範囲・外れ値の基準は極端な値の影響を受けにくい」という2つの見方が、同じ1組のデータの中で同時に確認できる。

平均値 50、 $Q_1 = 41$ 、中央値 46、 $Q_3 = 51$ 、四分位範囲 10、上側の基準値 66。90分は外れ値と判定される。

記述チャレンジ

答えだけでなく「なぜそうなるか」を文章と式で書く練習。模範解答と見比べて、書き方の型を身につけよう。

記1 平均値・中央値・分散・標準偏差は、どれも「データの特徴を1つの数値にまとめる」道具である。これらを「外れ値に対する強さ」という観点から比較し、どのような場面でどれを使うべきかを説明せよ。

平均値は、全データの値をそのまま足して個数で割るので、極端に大きい・小さい値（外れ値）が1つ混ざるだけで、その値に大きく引っばられてしまう。実際にこの章の練習問題でも、3, 4, 5, 5, 6, 7, 40のようなデータでは、中央値が5であるのに対し平均値は10まで押し上げられていた。

中央値は、データを小さい順に並べたときの「順位」だけで決まる値である。外れ値がどれだけ極端な値であっても、それが最大側・最小側にあるかぎり順位そのものは動かないので、中央値はほとんど影響を受けない。四分位数・四分位範囲も同じ理由で外れ値に強い（真ん中50%だけを見るため）。

分散・標準偏差はどうかというと、これらは平均値からの偏差を2乗して平均する量なので、平均値と同じ弱点を受け継いでいる。むしろ偏差を2乗する分、外れ値の影響はより強く出る（外れ値の偏差は大きいので、2乗するとさらに大きくなる）。四分位範囲（ $Q_3 - Q_1$ ）は分散・標準偏差と同じ「散らばりの指標」でありながら、外れ値の影響をほとんど受けない点で対照的である。

したがって、データにどんな外れ値が混ざっているかわからない、あるいは外れ値の影響を避けて「典型的な実態」をつかみたい場面では中央値・四分位範囲を使うべきであり、逆に、外れ値も含めた全データの情報をそのまま反映させたい場面（合計に直結する量を扱うときなど）では平均値・分散・標準偏差を使うのが適切である。

平均値・分散・標準偏差は全データを使う分、外れ値に弱い。中央値・四分位範囲は順位や真ん中50%だけで決まるので外れ値に強い。外れ値を避けて実態をつかみたいなら中央値・四分位範囲、全データの情報を反映したいなら平均値・分散・標準偏差を使う。

相関係数と仮説検定は、一見すると別々の技術に見えるが、どちらも「見た目の印象だけで判断せず、数値の基準で判断する」という共通の考え方にもとづいている。それぞれ具体例を1つずつ挙げながら、この共通点を説明せよ。

散布図を見るだけでも「点が右上がりに並んでいるから、正の相関がありそうだ」という印象は得られる。しかし「かなり強い相関」なのか「弱い相関」なのかは、見た目の印象だけでは人によって判断がぶれてしまう。そこで相関係数 r を計算すると、この関連の強さが必ず -1 以上 1 以下のただ1つの数値として表される。例えば5人のテストの得点を比べた例題で、 $r = 0.8$ と計算できれば、「かなり強い正の相関」だと、印象ではなく数値の根拠をもって言い切ることができる。

仮説検定も同じ発想である。「20回コインを投げて16回も表が出た」という結果を見ると、直感的には「なんだか多い気がする」という印象を持つ。しかしその「気がする」をそのまま結論にはしない。公正なコインで実際に20回投げる実験を1000回シミュレーションし、「16回以上表が出る」ことが1000回中何回起こったか（この章の例では6回、つまり0.6%）を数値で確認し、あらかじめ決めておいた基準（5%）と比べてはじめて、「公正だ」という仮説を棄却できる」と判断する。

このように、相関係数は「関連の強さ」という印象を、仮説検定は「結果のめずらしさ」という印象を、それぞれ具体的な数値（ r の値、割合（%））に置きかえ、あらかじめ決めた基準と比べることで結論を出す。感覚や思い込みではなく、計算された数値と基準にもとづいて判断するという姿勢は、この章全体を貫く共通の考え方である。

相関係数は「関連の強さ」という印象を -1 から 1 の数値に、仮説検定は「結果のめずらしさ」という印象を割合（%）の数値に置きかえ、あらかじめ決めた基準と比べて判断する。どちらも「印象ではなく数値の基準で判断する」という共通の考え方にもとづいている。

EXAM

共通テスト形式チャレンジ

大問1 目標時間 8分/配点 20点

花子さんのクラス（Aクラス）では、文化祭の来場者アンケートで「満足度スコア」（0点から100点）を聞いている。ランダムに選んだ5人分の記録は72, 78, 80, 84, 86点であった。先生「これを使って、データの基本的な特徴を調べてみよう。」

- (1) Aクラスの5人分のスコアについて、平均値 $\bar{x}$ = アイ点、偏差の2乗の和はウエオ、分散は V = カキである。また、標準偏差は s = クである（根号を用いて表せ）。

アイ 80 ウエオ 120 カキ 24 ク $2\sqrt{6}$

$$\text{平均値: } \bar{x} = \frac{72 + 78 + 80 + 84 + 86}{5} = \frac{400}{5} = 80 \text{点}$$

$$\begin{aligned} \text{偏差: } & 72 - 80 = -8, 78 - 80 = -2, 80 - 80 = 0, 84 - 80 = 4, 86 - 80 = 6 \\ \text{偏差の2乗: } & 64, 4, 0, 16, 36 \quad \text{和は } 64 + 4 + 0 + 16 + 36 = 120 \end{aligned}$$

$$\text{分散: } V = \frac{120}{5} = 24$$

$$\text{標準偏差: } s = \sqrt{24} = \sqrt{4 \times 6} = 2\sqrt{6}$$

$$\bar{x} = 80 \text{点、偏差の2乗の和は120、} V = 24, s = 2\sqrt{6} (\approx 4.90)$$

- (2) 太郎「後日、この5人には一律5点のボーナス点を追加することになりました。」

ボーナス点を追加した後の平均点はケコ点、分散はサシである。

ケコ 85 サシ 24

全員に一律5点を加えるのは「+5」の平行移動なので、平均は $80 + 5 = 85$ 点に変わるが、偏差の形（散らばり方）は変わらないので分散は変わらず24のままである。

- (3) 花子「隣のBクラスでも同じアンケートを5人からとったところ、75, 79, 80, 81, 85点でした。」

Bクラスの平均値は スセ 点、分散は ソタ である。また、Aクラス（ボーナス点追加前、分散24）とBクラスでは、どちらのスコアがより「そろっている」（散らばりが小さい）と言えるか、チ に $A \cdot B$ の記号で答えよ。

スセ 80 ソタ 10.4 チ B

$$\text{Bクラスの平均} : \frac{75 + 79 + 80 + 81 + 85}{5} = \frac{400}{5} = 80 \text{点}$$

$$\text{偏差} : -5, -1, 0, 1, 5 \quad \text{2乗の和} : 25 + 1 + 0 + 1 + 25 = 52$$

$$\text{分散} : \frac{52}{5} = 10.4$$

Aクラス（ボーナス点追加前）の分散は24、Bクラスの分散は10.4。分散が小さいBクラスの方が、スコアのばらつきが小さく、成績がそろっている。

Bクラスの平均値 80点、分散 10.4。そろっているのはBクラス。

大問2 目標時間 8分／配点 20点

太郎さんは、文化祭の模擬店の準備として、5日間の気温 x (°C) と、模擬店の来場者数 y (人) を記録した。 $x = 20, 22, 24, 26, 28$ 、 $y = 34, 37, 43, 40, 46$ であった（日の順は対応している）。先生「気温と来場者数に関係がありそうか、数値で確かめてみよう。」

- (1) $\bar{x} =$ アイ °C、 $\bar{y} =$ ウエ 人である。共分散は $S_{xy} =$ オカ、相関係数は $r =$ キ である。

アイ 24 ウエ 40 オカ 10.8 キ 0.9

$$\bar{x} = \frac{20 + 22 + 24 + 26 + 28}{5} = \frac{120}{5} = 24^{\circ}\text{C}$$

$$\bar{y} = \frac{34 + 37 + 43 + 40 + 46}{5} = \frac{200}{5} = 40 \text{人}$$

$$x - \bar{x} : -4, -2, 0, 2, 4$$

$$y - \bar{y} : -6, -3, 3, 0, 6$$

積：24, 6, 0, 0, 24 合計54

$$S_{xy} = 54 \div 5 = 10.8$$

$$x \text{の分散} = (16 + 4 + 0 + 4 + 16) \div 5 = 8$$

$$y \text{の分散} = (36 + 9 + 9 + 0 + 36) \div 5 = 18$$

$$r = \frac{10.8}{\sqrt{8} \times \sqrt{18}} = \frac{10.8}{\sqrt{144}} = \frac{10.8}{12}$$

$\bar{x} = 24^{\circ}\text{C}$ 、 $\bar{y} = 40$ 人、 $S_{xy} = 10.8$ 、 $r = 0.9$ （気温が高い日ほど来場者数が多い、強い正の相関）

(2) 花子「相関係数が0.9ということは、かなり強い正の相関ですね。ここで、 x の単位を $^{\circ}\text{C}$ からF（華氏）に変える（ $x' = 1.8x + 32$ という変換）と、 r の値はどうなるでしょうか。」

相関係数の性質を用いて、単位を変換した後の相関係数は $r' = \boxed{\text{ク}}$ であることを答えよ。

ク 0.9

相関係数には「単位を変えても r は変わらない」という性質がある（1. 学習の順路の3「相関係数・仮説検定の考え方」の要点整理で学んだ性質2）。

$x' = 1.8x + 32$ は、 x を1.8倍してから32を足す変換である。「1.8倍」の部分は偏差も1.8倍にするので、共分散は1.8倍、 x の標準偏差も1.8倍になり、相関係数の式で1.8がちょうど約分されて消える。「+32」の部分は全データを平行移動させるだけなので、そもそも偏差そのものを変えない（平均も32だけ動くので、データー平均の値は変わらない）。

したがって、 $^{\circ}\text{C}$ をF（華氏）に変換しても、相関係数の値はまったく変わらない。

$r' = 0.9$ （変換前と同じ。単位を変える一次変換では、倍率の部分は約分されて消え、平行移動の部分は偏差そのものを変えないため）

(3) 先生「実は、このアンケートには6日目の記録もありました。ただしこの日は大雨で模擬店を早めに閉めてしまい、気温は 30°C と高かったのに、来場者数はわずか10人でした。」

花子「(1)では $r = 0.9$ というかなり強い正の相関でしたよね。この6日目のデータ ($x = 30, y = 10$) を加えても、同じように強い相関になるのでしょうか。」

6日分のデータ $x = 20, 22, 24, 26, 28, 30, y = 34, 37, 43, 40, 46, 10$ について、新しい平均は $\bar{x} = \boxed{\text{ケコ}}^{\circ}\text{C}$ 、 $\bar{y} = \boxed{\text{サシ}}$ 人、共分散は $S_{xy} = \boxed{\text{ス}}$ である (S_{xy} は符号もつけて入力せよ)。また、この6日分で相関係数を計算し直すと、符号は $\boxed{\text{セ}}$ (正・負・なしのいずれか) になる。

ケコ 25 サシ 35 ス -16 セ (正 / 負 / なし)

$$\text{新しい平均: } \bar{x} = \frac{20 + 22 + 24 + 26 + 28 + 30}{6} = \frac{150}{6} = 25^{\circ}\text{C}, \bar{y} = \frac{34 + 37 + 43 + 40 + 46 + 10}{6} = \frac{210}{6} = 35\text{人}$$

$$x - \bar{x} : -5, -3, -1, 1, 3, 5$$

$$y - \bar{y} : -1, 2, 8, 5, 11, -25$$

$$\text{積} : 5, -6, -8, 5, 33, -125 \quad \text{合計} 5 - 6 - 8 + 5 + 33 - 125 = -96$$

$$S_{xy} = -96 \div 6 = -16$$

x の分散 $= (25 + 9 + 1 + 1 + 9 + 25) \div 6 = 70 \div 6 \approx 11.67$ 、 y の分散 $= (1 + 4 + 64 + 25 + 121 + 625) \div 6 = 840 \div 6 = 140$ 。どちらも正の値なので S_x, S_y は必ず正になり、相関係数 $r = \frac{S_{xy}}{S_x \times S_y}$ の符号は S_{xy} の符号と一致する。 $S_{xy} = -16 < 0$ なので、相関係数は負になる (実際に計算すると $r = \frac{-16}{\sqrt{11.67} \times \sqrt{140}} \approx -0.40$)。

(1)では5日間で $r = 0.9$ という強い正の相関があったのに、6日目のデータを1つ加えただけで r は負の値にまで変わってしまった。これは、この章の要点整理で学んだ相関係数の性質「外れ値がひとつ混ざるだけで大きく変わることがある」の実例である。6日目は気温が高いのに来場者数が極端に少ないという、それまでの5日間の傾向から大きく外れた点 (外れ値) であり、この1点が相関係数全体を大きく動かしてしまった。

$\bar{x} = 25^{\circ}\text{C}$ 、 $\bar{y} = 35$ 人、 $S_{xy} = -16$ 、相関係数の符号は負 ($r \approx -0.40$)。外れ値が1つ混ざるだけで相関係数が大きく変わるという性質の実例。