

相関係数・仮説検定の考え方

解答（全20問）

氏名 _____

得点 _____ / 20点

DRILL

練習問題（全15問）

基本（1～6）

1

共分散の計算

4人の生徒の勉強時間 x （分）と小テストの得点 y （点）は次の通りである。 $x = 10, 20, 30, 40$ 、 $y = 60, 65, 75, 80$ 。共分散 S_{xy} を求めよ。

$$S_{xy} = \underline{87.5}$$

$$\bar{x} = (10 + 20 + 30 + 40) \div 4 = 25, \bar{y} = (60 + 65 + 75 + 80) \div 4 = 70$$

$$x - \bar{x} : -15, -5, 5, 15$$

$$y - \bar{y} : -10, -5, 5, 10$$

$$\text{積} : 150, 25, 25, 150 \quad \text{合計 } 350$$

$$S_{xy} = 350 \div 4 = 87.5$$

2

公式に代入するだけ

あるデータについて、 x の標準偏差 $S_x = 2$ 、 y の標準偏差 $S_y = 3$ 、共分散 $S_{xy} = 4.8$ であった。相関係数 r を求めよ。

$$r = \underline{0.8}$$

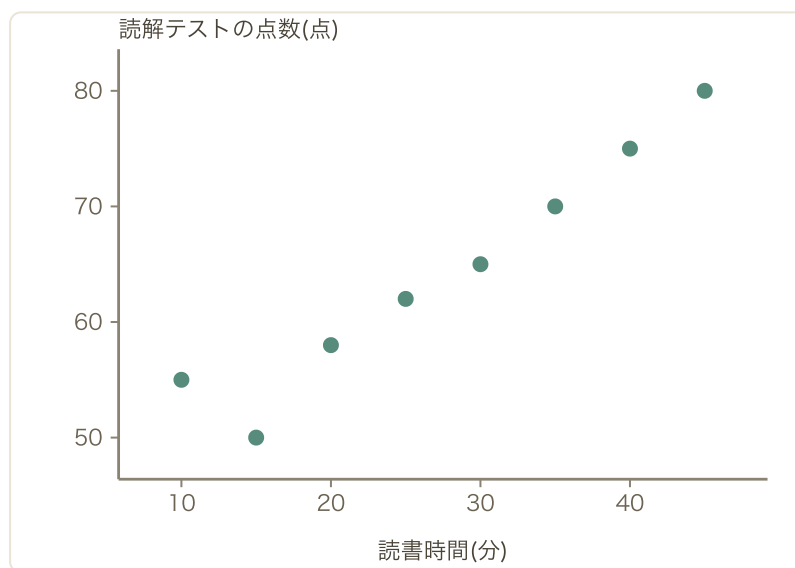
$$r = \frac{S_{xy}}{S_x \times S_y} = \frac{4.8}{2 \times 3} = \frac{4.8}{6}$$

$$r = 0.8$$

3

散布図の読み取り

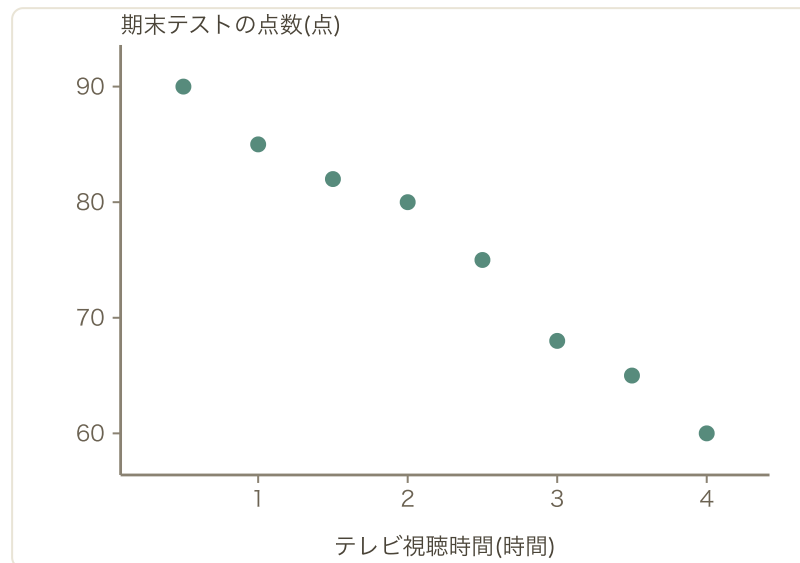
次の散布図は、5人の1日の読書時間（分）と読解テストの点数の関係を表したものである。相関の種類を「正」「負」「なし」のいずれかで答えよ。

相関の種類 正

読書時間が長い生徒ほど、読解テストの点数が高い傾向にある。点がおおむね右上がりに分布している。

正の相関

次の散布図は、1日のテレビ視聴時間（時間）と期末テストの点数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 **負**

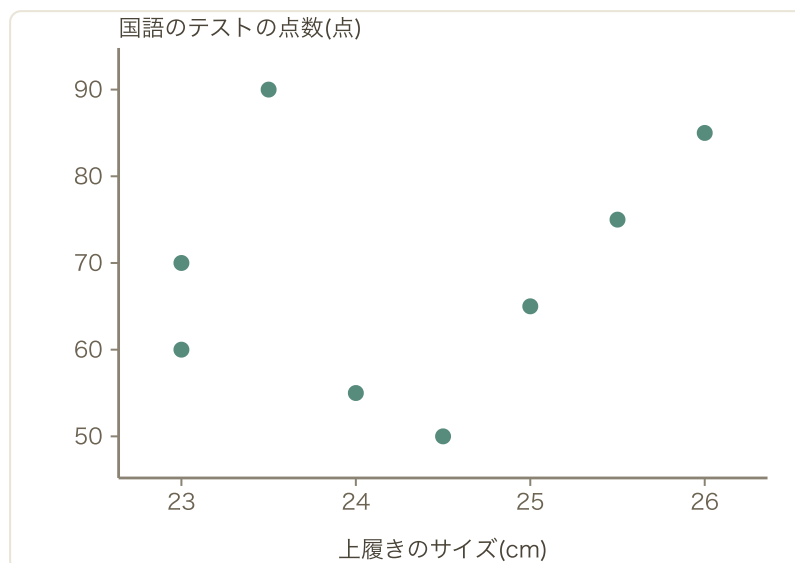
テレビ視聴時間が長い生徒ほど、テストの点数が低い傾向にある。点がおおむね右下がりに分布している。

負の相関

5

散布図の読み取り

次の散布図は、生徒の上履きのサイズ（cm）と国語のテストの点数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 なし

上履きのサイズが大きくても小さくても、国語の点数は高い場合も低い場合もあり、右上がり・右下がりのどちらの傾向も見えない。サイズと国語の点数は本来無関係な量なので、これは自然な結果といえる。

相関なし（無相関）

6

相関係数の範囲

次の値 -1.5 , 0.3 , 2 , -0.9 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 $-1.5, 2$

相関係数は必ず $-1 \leq r \leq 1$ の範囲におさまる。

-1.5 はこの範囲より小さく、 2 はこの範囲より大きいのでありえない。 0.3 と -0.9 はどちらも範囲内なのであり得る。

ありえない値は $-1.5, 2$

- 7 5人の生徒が受けた2回の小テスト（各5点満点）の得点は、1回目 $x = 1, 2, 3, 4, 5$ 、2回目 $y = 3, 1, 2, 5, 4$ 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。

$$S_{xy} = \underline{1.2} \quad r = \underline{0.6}$$

$$\bar{x} = \bar{y} = 3$$

$$x - \bar{x} : -2, -1, 0, 1, 2$$

$$y - \bar{y} : 0, -2, -1, 2, 1$$

$$\text{積} : 0, 2, 0, 2, 2 \quad \text{合計 } 6$$

$$S_{xy} = 6 \div 5 = 1.2$$

分散はどちらも $(4 + 1 + 0 + 1 + 4) \div 5 = 2$ なので $S_x = S_y = \sqrt{2}$ 。

$$r = \frac{1.2}{\sqrt{2} \times \sqrt{2}} = \frac{1.2}{2}$$

$$S_{xy} = 1.2、r = 0.6$$

5人について、通学時間の順位 $x = 1, 2, 3, 4, 5$ （早く着く順）と朝食にかける時間の順位 $y = 4, 5, 2, 3, 1$ （長い順）を調べたところ次の対応になった（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。

$$S_{xy} = \underline{-1.6} \quad r = \underline{-0.8}$$

$$\bar{x} = \bar{y} = 3$$

$$x - \bar{x} : -2, -1, 0, 1, 2$$

$$y - \bar{y} : 1, 2, -1, 0, -2$$

$$\text{積} : -2, -2, 0, 0, -4 \quad \text{合計} -8$$

$$S_{xy} = -8 \div 5 = -1.6$$

分散はどちらも 2 なので $S_x = S_y = \sqrt{2}$ 。

$$r = \frac{-1.6}{\sqrt{2} \times \sqrt{2}} = \frac{-1.6}{2}$$

$$S_{xy} = -1.6, r = -0.8$$

あるデータで x の単位を「cm」から「mm」に変える（すべての値を10倍する）と、相関係数 r の値はどうなるか。「10倍になる」「10分の1になる」「変わらない」のいずれかで答えよ。

答え 変わらない

単位を変えても、点の並び方の「形」（右上がり・右下がりの傾向の強さ）そのものは変わらない。相関係数は単位に依存しない量である。

変わらない

ほとんどの点が右上がりの一直線に近い形で分布しているデータに、傾向から大きく外れた点を1つ追加すると、相関係数の絶対値は一般にどうなるか。「大きくなる」「小さくなる」「変わらない」のいずれかで答えよ。

答え 小さくなる

相関係数は外れ値の影響を受けやすい。一直線に近い並びの中に、傾向から離れた点が1つ混ざるだけで、一直線からのズレが大きくなり、相関係数の絶対値は小さくなる（1や-1から離れる）方向に動く。

小さくなる

ある年の統計で「アイスクリームの売上」と「水の事故件数」には正の相関があった。しかし、アイスクリームが直接水の事故を引き起こしているとは考えにくい。この2つの背後で、両方に共通して影響していると考えられる要因を答えよ。

共通の要因 気温

夏で気温が高い時期には、アイスクリームがよく売れる一方、海や川で泳ぐ人が増えて水の事故も増える。つまり「気温（夏かどうか）」という第三の要因が、両方の数値を同時に押し上げている。相関があるからといって、一方が他方の原因とは限らない。

気温（夏の暑さ）

友達が「このコインは公正だ」と言うので20回投げてみたところ、表が14回出た。STEP4の表を参考に、「20回中14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。

割合 5.8 % 判断 棄却できない

STEP4の表から、14回・15回・16回・17回・18回・19回・20回が出た回数を足し合わせる。

$$37 + 15 + 5 + 1 + 0 + 0 + 0 = 58 \text{ (1000回中58回)}$$

割合は $58 \div 1000 = 0.058$ 、つまり**5.8%**。

5.8% > 5% なので、「公正なコインで14回以上表が出る」ことはそれほどめずらしいとは言えない。したがって、「このコインは公正である」という仮説は**棄却できない**（＝仮説が正しいと証明されたわけではないが、疑うだけの根拠もない）。

14回以上表が出る割合は5.8%（5%より大きい） → 仮説を棄却できない

13

単位が違う実データでの計算

5人の生徒の学習時間 x (分) とテストの得点 y (点) は $x = 10, 20, 30, 40, 50$ 、 $y = 60, 70, 90, 80, 100$ であった (生徒の順は対応している)。相関係数 r を求めよ。

$$r = \underline{0.9}$$

$$\bar{x} = (10 + 20 + 30 + 40 + 50) \div 5 = 30, \bar{y} = (60 + 70 + 90 + 80 + 100) \div 5 = 80$$

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

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

$$\text{積} : 400, 100, 0, 0, 400 \quad \text{合計 } 900$$

$$S_{xy} = 900 \div 5 = 180$$

$$x \text{ の分散} = (400 + 100 + 0 + 100 + 400) \div 5 = 200, \text{ よって } S_x = \sqrt{200} = 10\sqrt{2}$$

$$y \text{ の分散} = (400 + 100 + 100 + 0 + 400) \div 5 = 200, \text{ よって } S_y = \sqrt{200} = 10\sqrt{2}$$

$$r = \frac{180}{10\sqrt{2} \times 10\sqrt{2}} = \frac{180}{200}$$

$$r = 0.9 \text{ (学習時間が長い生徒ほど得点が高い、強い正の相関)}$$

5人について、1週間の漢字テスト復習回数 x (回) と、テストの誤答率 y (%) は $x = 2, 4, 6, 8, 10$ 、 $y = 25, 15, 20, 10, 5$ 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。

$$S_{xy} = \underline{-18} \quad r = \underline{-0.9}$$

$$\bar{x} = (2 + 4 + 6 + 8 + 10) \div 5 = 6, \bar{y} = (25 + 15 + 20 + 10 + 5) \div 5 = 15$$

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

$$y - \bar{y} : 10, 0, 5, -5, -10$$

$$\text{積} : -40, 0, 0, -10, -40 \quad \text{合計 } -90$$

$$S_{xy} = -90 \div 5 = -18$$

$$x \text{ の分散} = (16 + 4 + 0 + 4 + 16) \div 5 = 8, \text{ よって } S_x = \sqrt{8} = 2\sqrt{2}$$

$$y \text{ の分散} = (100 + 0 + 25 + 25 + 100) \div 5 = 50, \text{ よって } S_y = \sqrt{50} = 5\sqrt{2}$$

$$r = \frac{-18}{2\sqrt{2} \times 5\sqrt{2}} = \frac{-18}{20}$$

$$S_{xy} = -18, r = -0.9 \text{ (復習回数が多いほど誤答率が下がる、強い負の相関)}$$

あるデータについて、 x の標準偏差 $S_x = 1.5$ 、 y の標準偏差 $S_y = 2$ 、相関係数 $r = -0.5$ であることが分かっている。共分散 S_{xy} を求めよ。

$$S_{xy} = \underline{-1.5}$$

$$r = \frac{S_{xy}}{S_x \times S_y} \text{ を } S_{xy} \text{ について解くと } S_{xy} = r \times S_x \times S_y。$$

$$S_{xy} = (-0.5) \times 1.5 \times 2 = (-0.5) \times 3$$

$$S_{xy} = -1.5$$

応用問題（プリント限定）

ここから先はPDF限定の腕試し。相関係数の性質と仮説検定の考え方を、別の場面で使ってみよう。

応1

相関係数の不変性

STEP2の例題のデータ（1回目の得点 $x = 1, 2, 3, 4, 5$ 、2回目の得点 $y = 1, 3, 2, 5, 4$ 、相関係数 $r = 0.8$ ）について、 x を「点」から「10点満点中の点数の10倍（つまりmm的な単位変換にあたる $x' = 10x$ ）」に、 y を $y' = 20y$ に変換したとする。このとき相関係数 r' はいくつになるか。

$$r' = \underline{0.8}$$

相関係数は「単位を変えても変わらない」という性質を持つ（STEP3・性質2）。

$x' = 10x$ 、 $y' = 20y$ はどちらも正の数をかけるだけの変換なので、点の並び方の形は変わらない。

実際に計算しても確かめられる。 $x' = 10, 20, 30, 40, 50$ 、 $y' = 20, 60, 40, 100, 80$ 。

$$\bar{x}' = 30, \bar{y}' = 60$$

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

$$y' - \bar{y}' : -40, 0, -20, 40, 20$$

$$\text{積} : 800, 0, 0, 400, 400 \quad \text{合計 } 1600 \quad S_{x'y'} = 1600 \div 5 = 320$$

$$x' \text{ の分散} = (400 + 100 + 0 + 100 + 400) \div 5 = 200, S_{x'} = \sqrt{200} = 10\sqrt{2}$$

$$y' \text{ の分散} = (1600 + 0 + 400 + 1600 + 400) \div 5 = 800, S_{y'} = \sqrt{800} = 20\sqrt{2}$$

$$r' = \frac{320}{10\sqrt{2} \times 20\sqrt{2}} = \frac{320}{400} = 0.8$$

$$r' = 0.8 \text{（変換前と同じ）}$$

5人の生徒について、勉強時間の順位 $x = 1, 2, 3, 4, 5$ と、あるテストの得点の順位 $y = 2, 3, 4, 5, 1$ を調べた（生徒の順は対応している）。まず5人全員のデータで相関係数 r_1 を求めよ。次に、明らかに他の4人の傾向から外れている5人目（ $x = 5, y = 1$ ）を除いた4人だけで相関係数 r_2 を求め、外れ値を除くと相関係数はどう変化したか、「大きくなった」「小さくなった」「変わらなかった」のいずれかで答えよ。

$r_1 = \underline{0}$ $r_2 = \underline{1}$ 変化 大きくなった

【5人全員】 $\bar{x} = 3, \bar{y} = 3$

$x - \bar{x} : -2, -1, 0, 1, 2$

$y - \bar{y} : -1, 0, 1, 2, -2$

積：2, 0, 0, 2, -4 合計 0 $S_{xy} = 0 \div 5 = 0$ なので $r_1 = 0$ （分散はどちらも2で0でないため、 $r_1 = 0 \div 2 = 0$ ）。

【5人目を除いた4人】 $x = 1, 2, 3, 4, y = 2, 3, 4, 5$ 。

$\bar{x} = 2.5, \bar{y} = 3.5$

$x - \bar{x} : -1.5, -0.5, 0.5, 1.5$

$y - \bar{y} : -1.5, -0.5, 0.5, 1.5$

偏差の並びが完全に一致している（ x と y が同じだけ増える完全な一直線）ので $r_2 = 1$ 。

5人目（ $x = 5, y = 1$ ）は他の4人が作る右上がりの直線から大きく外れた点だった。この1点があるだけで、相関係数は完全な正の相関（ $r_2 = 1$ ）から無相関（ $r_1 = 0$ ）にまで下がってしまう。

$r_1 = 0, r_2 = 1 \rightarrow$ 外れ値を除くと相関係数は大きくなった

「このさいころは公正で、1の目が出る確率は $\frac{1}{6}$ である」という仮説のもとで、さいころを60回投げる実験を1000回シミュレーションしたところ、1の目が出た回数について次の結果（の一部）が得られた。

1の目の回数	13	14	15	16	17	18	19	20
1000回中の回数	75	50	31	17	9	4	2	1

実際に60回投げたところ1の目が18回出た。「18回以上1の目が出る」割合（%）を求め、5%を基準にこの仮説を棄却できるか判断せよ。

割合 0.7 % 判断 棄却できる

表から、18回・19回・20回が出た回数を足し合わせる。

$$4 + 2 + 1 = 7 \text{ (1000回中7回)}$$

割合は $7 \div 1000 = 0.007$ 、つまり**0.7%**。

0.7% < 5% なので、「公正なさいころで18回以上1の目が出る」ことは非常にめずらしいと判断できる。したがって、「このさいころは公正である」という仮説は**棄却できる**。

18回以上1の目が出る割合は0.7%（5%より小さい） → 仮説を棄却できる

ある工場で、機械の点検回数 x (回/月) と故障件数 y (件/月) を5か月分調べたところ $x = 2, 4, 6, 8, 10$ 、 $y = 25, 15, 20, 10, 5$ であった (月の順は対応している)。相関係数 r を求め、この結果からどのようなことが言えそうか、最も適切なものを次から選べ：1. 点検回数を増やすと必ず故障がゼロになる／2. 点検回数が多い月ほど故障件数が少ない傾向がある／3. 点検回数と故障件数には関係がない

$r =$ -0.9 適切な記述の番号 2

計算方法は練習問題14とまったく同じ (数値も同じ)。

$$\bar{x} = 6, \bar{y} = 15$$

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

$$y - \bar{y} : 10, 0, 5, -5, -10$$

$$\text{積} : -40, 0, 0, -10, -40 \quad \text{合計} -90 \quad S_{xy} = -90 \div 5 = -18$$

$$S_x = \sqrt{8} = 2\sqrt{2}, S_y = \sqrt{50} = 5\sqrt{2}$$

$$r = \frac{-18}{2\sqrt{2} \times 5\sqrt{2}} = \frac{-18}{20} = -0.9$$

記述1は「必ず」「ゼロになる」が言い過ぎ (相関は全体の傾向であって、因果関係を保証するものでも、個々のケースを断定するものでもない)。記述3は相関があるのに「関係がない」としており誤り。記述2は「傾向がある」という相関の性質に沿った、控えめで正確な表現なので適切。

$r = -0.9$ 、適切な記述は2

ある会社で社員50人について、残業時間 x と仕事の満足度アンケートの点数 y を調べたところ、相関係数は $r = -0.75$ だった。次の記述のうち、この結果から確実に言えることを全て選び、番号を小さい順にコンマ区切りで答えよ。

1. 残業時間を減らせば満足度は必ず上がる（因果関係が証明された）
2. 残業時間と満足度の間には、線形的に見て負の傾向がある
3. 残業時間が多い社員は、全員例外なく満足度が低い
4. この $r = -0.75$ という値は、相関係数として妥当な範囲におさまっている

正しい記述の番号 2,4

1は誤り：相関があるからといって因果関係が証明されたことにはならない（STEP3・性質4）。残業時間と満足度の両方に影響する別の要因（例えば職場の人間関係など）がある可能性もある。

2は正しい：相関係数が負であることは、 x が増えると y が減るという線形的な傾向があることをそのまま意味する。

3は誤り：相関はデータ全体の傾向を表すものであり、「全員例外なく」のように個々の社員すべてに当てはまることを保証するものではない（外れ値も存在しうる）。

4は正しい：相関係数は必ず $-1 \leq r \leq 1$ の範囲におさまる（STEP3・性質1）。 -0.75 はこの範囲内なので妥当な値である。

正しい記述は2, 4