
第4 章

～ データの分析 ～

第1講 データの代表値と散らばり

1 データの管理 2 データの代表値

1 データ

人の身長、体重や運動の記録などのように、ある特性を表す数量を **変量** といい、ある変量の測定値や観測値の集まりを **データ** という。

2 度数分布

度数分布表 データのとり値をいくつかの区間に区切って階級を定め、各階級に度数を対応させた表。各階級の中央の値を **階級値** という。

ヒストグラム 度数分布表を柱状のグラフで表したもの。

3 代表値

平均値 変量 x のデータの値が $x_1, x_2, \dots, x_n$ であるとき、このデータの平均値 $\bar{x}$

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

最頻値 (モード) データにおいて、最も個数の多い値。データが度数分布表に整理されているときは、度数が最も大きい階級の階級値。

中央値 (メジアン) データを値の大きさの順に並べたとき、その中央の位置にくる値。データの大きさが偶数のときは、中央の2つの値の平均値。

3 データの散らばりと四分位数

1 四分位数

範囲 データの最大値から最小値を引いた差。

四分位数 データの値を大きさの順に並べたとき、4等分する位置の値。
小さい方から順に、**第1四分位数**、**第2四分位数**、**第3四分位数** といい、順に Q_1 、 Q_2 、 Q_3 で表す。第2四分位数はデータの中央値である。

四分位範囲 データの第3四分位数 Q_3 と第1四分位数 Q_1 の差 $Q_3 - Q_1$

四分位偏差 四分位範囲の半分の値。

2 箱ひげ図

箱ひげ図 データの分布を、次の5つの値で表した図。平均値を記入することもある。
最小値、第1四分位数 (Q_1)、中央値 (Q_2)、第3四分位数 (Q_3)、最大値



第1講 例題

1 ★☆☆

右の表は、Aさんの行きの通学時間を30日間にわたって調べた度数分布表である。

この度数分布表について、次の問いに答えよ。

- (1) 階級の幅をいえ。
- (2) 度数が2番目に大きい階級の階級値をいえ。
- (3) 56分以上かかった日数は全体の何%か。

通学時間の度数分布表

階級(分)	度数
50以上 52未満	2
52 ～ 54	5
54 ～ 56	11
56 ～ 58	7
58 ～ 60	4
60 ～ 62	1
計	30

2 ★☆☆

次のデータの平均値、最頻値、中央値を求めよ。

- (1) 3, 5, 9, 9, 10, 11, 11, 11, 12
- (2) 9, 3, 8, 12, 14, 12, 4, 13, 6, 12, 2, 7

3 ★★☆☆

次のデータは、ある8店舗での1kgあたりのみかんの価格である。ただし、 a の値は0以上の整数である。

525 550 498 560 550 555 500 a (単位は円)

- (1) このデータの平均値が535円であるとき、 a の値を求めよ。
- (2) a の値がわからないとき、このデータの中央値として何通りの値が考えられるか答えよ。

4 ★☆☆

次のデータの第1四分位数、第2四分位数、第3四分位数を求めよ。

- (1) 9, 15, 17, 22, 29, 31, 40, 43
- (2) 5, 7, 12, 19, 21, 30, 33, 36, 40
- (3) 8, 11, 20, 23, 24, 31, 42, 44, 55, 58
- (4) 10, 12, 18, 21, 31, 33, 42, 45, 48, 52, 56

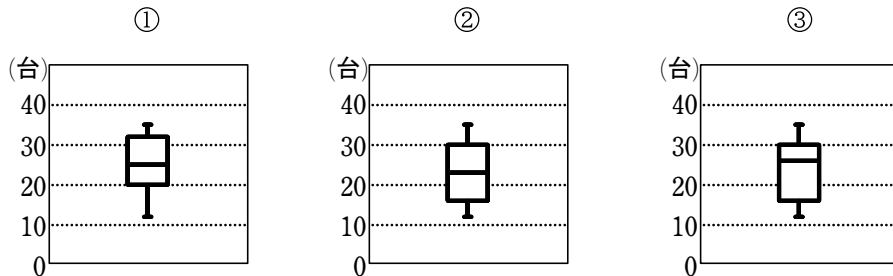
第 1 講 例題

5 ★☆☆

次のデータは、ある地点における、午前 8 時から午前 9 時までの 1 時間の車の交通量を 9 日間調べたものである。

29, 17, 23, 12, 15, 31, 35, 26, 21 (単位は台)

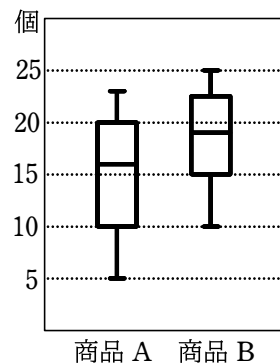
このデータの箱ひげ図を、下の ① ~ ③ から選べ。



6 ★★☆☆

右の図は、ある商店の商品 A と商品 B の 30 日間にわたる販売数のデータの箱ひげ図である。この箱ひげ図から読み取れることとして正しいものを、次の ① ~ ③ からすべて選べ。

- ① 商品 A は商品 B に比べて販売数の散らばりの度合いが大きい。
- ② 商品 A では販売数が 15 個以上の日が 15 日以上あった。
- ③ 商品 A, B ともに販売数が 10 個未満の日があった。



第1講 例題演習

1

右の表は、ある高校のクラス 40 人の通学時間の度数分布表である。この度数分布表について、次の問いに答えよ。

- (1) 階級の幅をいえ。
- (2) 度数が 9 である階級の階級値をいえ。
- (3) 40 分以上かかった人数は全体の何 % か。

通学時間の度数分布表

階級(分)	度数
0 以上 20 未満	2
20 ~ 40	13
40 ~ 60	15
60 ~ 80	9
80 ~ 100	1
計	40

2

次のデータの平均値、最頻値、中央値を求めよ。

- (1) 1, 1, 3, 5, 5, 6, 6, 6, 10, 10
- (2) 5, 8, 2, 12, 8, 5, 8, 3, 12, 5, 2, 13, 8

3

次のデータはある 10 人の生徒の数学のテストの得点である。ただし、 a の値は 0 以上の整数である。

60 74 66 62 82 38 45 41 67 a (点)

- (1) a の値がわからないとき、10 人の得点の中央値として何通りの値があり得るか。
- (2) 10 人の得点の平均値が 61.0 点のとき、 a の値を求めよ。

4

次のデータの第 1 四分位数、第 2 四分位数、第 3 四分位数を求めよ。

- (1) 3, 9, 12, 14, 34, 38, 41, 42, 45, 52, 54, 59, 64
- (2) 1, 4, 14, 20, 22, 25, 34, 38, 39, 50, 52, 58, 61, 65
- (3) 11, 26, 30, 34, 51, 52, 69, 74, 77, 81, 86
- (4) 7, 10, 16, 20, 26, 32, 39, 48, 49, 54, 56, 71

第1講 例題演習

5

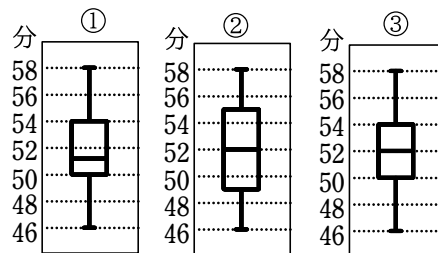
次のデータは、Aさんの15日間にわたる通勤時間である。(単位は分)

51, 49, 52, 54, 48, 55, 50, 51,

53, 46, 57, 54, 58, 52, 50

このデータを箱ひげ図に表したものを、右の

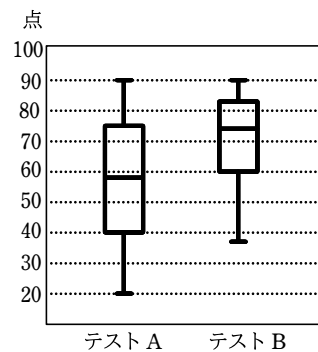
①～③から選べ。



6

右の図は、400人の生徒が受験したテストAとテストBの得点のデータの箱ひげ図である。この箱ひげ図から読み取れることとして正しいものを、次の①～④からすべて選べ。

- ① テストAはテストBに比べて得点の散らばりの度合いが大きい。
- ② テストAでは60点以上の生徒が200人より多い。
- ③ テストBでは80点以上の生徒が100人以上いる。
- ④ テストA, Bともに30点台の生徒がいる。



第1講 レベルA

1

右の表は、ある店の1日の弁当の販売個数を30日間調べた結果の度数分布表である。

- (1) データの最頻値を求めよ。
- (2) この表から階級値を用いて、データの平均値を求めよ。
- (3) 階級値を用いなくて平均値を求めると、データの平均値はどのような値の範囲に入るか。

階級(個)	度数
100 以上 120 未満	3
120 ~ 140	5
140 ~ 160	11
160 ~ 180	8
180 ~ 200	3
計	30

2

右の2つの度数分布表は、あるクラスの10人について行われた漢字の「読み」と「書き取り」のテストの得点をそれぞれまとめたものである。ただし、度数分布表にまとめる前の得点はわからないものとして、度数分布表にまとめる前のテストの得点の中央値と平均値がどのような値であることを考える。

読 み	
階級(点)	人数
40 以上 49 以下	4
50 ~ 59	2
60 ~ 69	2
70 ~ 79	2
合 計	10

書 き 取 り	
階級(点)	人数
40 以上 49 以下	0
50 ~ 59	2
60 ~ 69	5
70 ~ 79	3
合 計	10

- (1) 「読み」の得点の中央値は、最も小さい値として^ア点の可能性がある、最も大きい値として^イ点の可能性がある。
- (2) 「書き取り」の得点の平均値は、最も小さい値として^ウ点の可能性がある、最も大きい値として^エ点の可能性がある。

3

次のデータは、ある町の6月から11月の間にゴミ集積場に集められたペットボトルのゴミの量である。

1.61, 4.12, 3.71, 2.87, 2.48, 2.13 (単位はt)

- (1) 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2.84tと3.02t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。

第1講 レベルA

4

次のデータは、A 君と B 君のある定期テストの得点である。

A 君 45, 67, 52, 96, 42, 72, 58, 83, 89

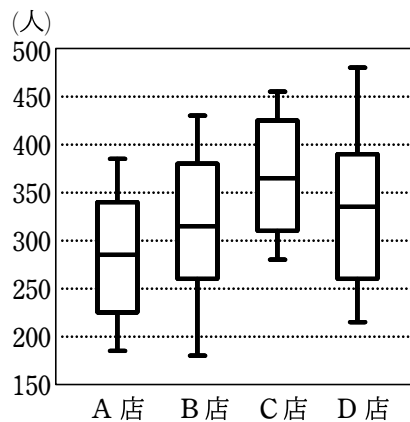
B 君 72, 60, 66, 82, 37, 70, 75, 81, 98 (単位は点)

A 君, B 君のデータの四分位範囲と四分位偏差を求めよ。また, 四分位範囲によってデータの散らばりの度合いを比較せよ。

5

右の図は, 同じ敷地内にある A 店, B 店, C 店, D 店の 1 日の入店者数を 31 日間調べたデータを, 箱ひげ図に表したものである。

- (1) 1 日の入店者数が 350 人を超えた日が 16 日以上あったのは, どの店か。
- (2) 1 日の入店者数が 250 人を下回る日が 8 日以上あったのは, どの店か。
- (3) B 店において 1 日の入店者数が 200 人を超えたのは, 最大で何日あった可能性があるか。



第2講 分散と標準偏差・データの相関・仮説検定

4 分散と標準偏差

① 分散, 標準偏差 $x_1, x_2, \dots, x_n$ を変量 x のデータの値とする。

偏差 データの各値と平均値 $\bar{x}$ との差 $x - \bar{x}$

分散 偏差の2乗の平均値。 s^2 で表す。

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

(x のデータの分散) = (x^2 のデータの平均値) - (x のデータの平均値)²

標準偏差 分散の正の平方根。 s で表す。

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

(x のデータの標準偏差) = $\sqrt{(\text{\textit{x}}^2 \text{ のデータの平均値}) - (\text{\textit{x}} \text{ のデータの平均値})^2}$

研究 変数の変換

a, b は定数とする。変量 x のデータから $y = ax + b$ によって新しい変量 y のデータが得られるとき, x, y のデータの平均値を $\bar{x}, \bar{y}$, 分散を s_x^2, s_y^2 , 標準偏差を s_x, s_y とすると $\bar{y} = a\bar{x} + b, \quad s_y^2 = a^2 s_x^2, \quad s_y = |a| s_x$

5 データの相関

① 散布図

散布図 2つの変量からなるデータを平面上に図示したもの。

正の相関 2つの変量からなるデータにおいて, 一方が増加すると他方も増加する傾向がみられるとき, 2つの変量には正の相関があるという。

負の相関 2つの変量からなるデータにおいて, 一方が増加すると他方が減少する傾向がみられるとき, 2つの変量には負の相関があるという。

散布図において, 正の相関があるとき, データを表す点は全体的に右上がりとなり, 負の相関があるとき, データを表す点は全体的に右下がりとなる。

② 相関係数

相関係数 相関の正負と強弱を表す値。2つの変量 x, y からなるデータにおいて, $(x - \bar{x})(y - \bar{y})$ の総和を偏差積といい, この平均値を変量 x と変量 y の共分散といい s_{xy} で表す。変量 x の標準偏差を s_x , 変量 y の標準偏差を s_y とするとき, 次の値 r を x と y の 相関係数 という。

$$r = \frac{s_{xy}}{s_x s_y} \quad (-1 \leq r \leq 1)$$

相関係数 r が1に近いほど正の相関が強く, -1 に近いほど負の相関が強い。また, 相関がないとき, r は0に近い値をとる。

6 仮説検定

1 仮説検定の考え方

得られたデータをもとに，母集団に対する仮説を立て，それが正しいかどうかを判断する手法を仮説検定という。

2 仮説検定の手順

ある手段が正しいかどうか判断するための仮説検定は，次のような手順で行う。

- ① 正しいかどうか判断したい主張に対し，その主張に反する仮説を立てる。
- ② 基準となる確率を定め，立てた仮説のもとで，得られたデータがどの程度の確率で起こるかを求める。
- ③ 仮説が正しいかどうかをもとに，主張が正しいか判断する。

第2講 例題

1 ★☆☆

右の表は、ある製品を成型できる2台の工作機械X, Yの1時間あたりのそれぞれの不良品の数 x , y を5時間にわたって調べたものである。(単位は個)

x	5	4	8	12	6
y	6	9	8	5	7

- (1) x , y のデータの平均値, 分散, 標準偏差をそれぞれ求めよ。ただし, 小数第2位を四捨五入せよ。
- (2) x , y のデータについて, 標準偏差によってデータの平均値からの散らばりの度合いを比較せよ。

2 ★☆☆

次のデータは, 5人の生徒の数学のテストの得点 x である。

50, 70, 90, 80, 50 (点)

- (1) このデータの平均値 $\bar{x}$ を求めよ。
- (2) このデータの各値の2乗の平均値 $\overline{x^2}$ を求めよ。
- (3) (2)の値を用いてこのデータの分散, 標準偏差を求めよ。

3 ★★☆☆

15個の値からなるデータがあり, そのうちの10個の値の平均値は9, 分散は3, 残り5個の値の平均値は6, 分散は9である。この15個のデータの平均値と分散を求めよ。

4 ★★☆☆

次の変数 x のデータは, ある地域の6つの山の高さである。

1008, 992, 980, 1008, 984, 980 (単位はm)

- (1) 仮平均 x_0 を1000として, 変数 x のデータの平均値 $\bar{x}$ を求めよ。
- (2) $u = \frac{x-1000}{4}$ とおくことにより, 変数 x のデータの標準偏差, 分散を求めよ。

第2講 例題

5 ★☆☆

次のような2つの変数 x , y からなるデータがある。これらについて、散布図をかき、 x と y の間に相関があるかどうかを調べよ。また、相関がある場合には、正か負のどちらの相関であるかをいえ。

(1)

x	3	8	5	2	6	4	9	5	1	7
y	4	9	8	1	7	2	8	3	2	5

(2)

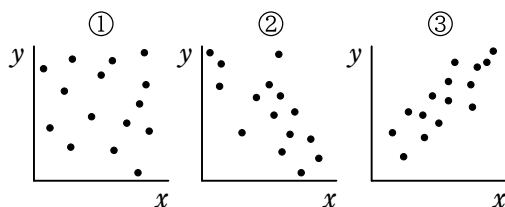
x	78	63	86	54	92	57	95	69	81	73
y	32	29	4	48	2	37	13	41	26	15

(3)

x	3.7	2.8	4.2	1.9	5.1	1.6	5.4	2.3	4.7	3.5
y	5.3	7.6	3.2	4.8	6.5	6.9	4.1	3.4	7.3	5.7

6 ★☆☆

右の①, ②, ③はある2つの変数 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。



7 ★★☆☆

下の表は、10人の生徒に10点満点の2種類のテストA, Bを行った結果である。A, Bの得点の相関係数を求めよ。また、これらの間にはどのような相関があると考えられるか。

生徒番号	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
テストA	5	3	5	5	3	8	10	5	8	8
テストB	6	9	6	5	7	2	3	8	3	1

第2講 例題

8 ★☆☆

ある企業が発売している製品を改良し、20人にアンケートを実施したところ、15人が「品質が向上した」と回答した。この結果から、製品の品質が向上したと判断してよいか。仮説検定の考え方をを用い、基準となる確率を0.05として考察せよ。ただし、公正なコインを20枚投げて表が出た枚数を記録する実験を200回行ったところ、次のようになったとし、この結果を用いよ。

表の枚数	4	5	6	7	8	9	10	11	12	13	14	15	16	17
度数	1	3	8	14	24	30	37	32	23	16	8	3	0	1

9 ★★☆☆

AとBがあるゲームを9回行ったところ、Aが7回勝った。この結果から、AはBより強いと判断してよいか。仮説検定の考え方をを用い、基準となる確率を0.05として考察せよ。ただし、ゲームに引き分けはないものとする。

第2講 例題演習

1

次の変数 x , y のデータは、2組の歌手グループ X, Y の CD 購入者 10 名ずつに、曲を聴いた満足度を 10 点満点で評価してもらった結果である。

x	5	6	5	5	8	6	3	5	3	4
y	4	0	6	3	4	2	10	3	7	1

- (1) 変数 x , y の平均値, 分散, 標準偏差を, それぞれ求めよ。ただし, 小数第 2 位を四捨五入せよ。
- (2) 変数 x , y について, 標準偏差によってデータの平均値からの散らばりの度合いを比較せよ。

2

次のデータは、5 人の生徒の数学のテストの得点 x (点) である。

50, 70, 90, 80, 50

- (1) このデータの平均値 $\bar{x}$ を求めよ。
- (2) このデータの各値の 2 乗の平均値 $\overline{x^2}$ を求めよ。
- (3) (2) の値を用いてこのデータの分散 s^2 , 標準偏差 s を求めよ。

3

20 個の値からなるデータがあり, そのうちの 8 個の値の平均値は 3, 分散は 4, 残りの 12 個の値の平均値は 8, 分散は 9 である。

- (1) このデータの平均値を求めよ。
- (2) このデータの分散を求めよ。

4

- (1) 次の変数 x のデータについて, 次の問いに答えよ。

844, 893, 872, 844, 830, 865 (単位は点)

(ア) 仮平均 x_0 を 830 とし, 変数 x のデータの平均値 $\bar{x}$ を求めよ。

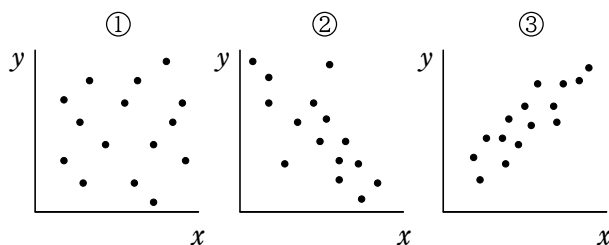
(イ) $u = \frac{x - 830}{7}$ とおくことにより, 変数 x のデータの標準偏差, 分散を求めよ。

- (2) 変数 x の平均を $\bar{x}$, 標準偏差を s_x とする。 $z = \frac{x - \bar{x}}{s_x}$ において得られる新しい変数 z の平均 $\bar{z}$ と標準偏差 s_z を求めよ。

第2講 例題演習

5

右の①, ②, ③は, ある
2つの変数 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, 0.87 , -0.04 ,
 -0.71 のいずれかである。
各データの相関係数を答えよ。



6

次のような変数 x , y のデータがある。これらについて, 散布図をかき, x と y の間に相
関関係があるかどうかを調べよ。また, 相関関係がある場合には, 正・負のどちらである
かをいえ。

(1)

x	3.7	2.8	4.2	1.9	5.1	1.6	5.4	2.3	4.7	3.5
y	5.3	7.6	3.2	4.8	6.5	6.9	4.1	3.4	7.3	5.7

(2)

x	4.5	3.3	5.7	2.6	6.4	2.9	6.1	3.8	5.2	4.7
y	5.3	6.7	4.6	7.9	5.2	8.1	4.8	8.3	6.4	7.5

(3)

x	7.7	6.8	8.2	5.9	9.1	5.6	9.4	6.3	8.7	7.5
y	3.4	2.7	4.2	0.8	3.5	1.3	4.1	0.9	2.5	1.6

7

下の表は, 10 人の生徒に 50 点満点の 2 種類のテスト A, B を行った得点の結果である。
テスト A, B の得点をそれぞれ x , y とするとき, x と y の相関係数 r を求めよ。ただ
し, 小数第 3 位を四捨五入せよ。

生徒の番号	1	2	3	4	5	6	7	8	9	10
x	43	41	43	38	39	42	42	39	41	42
y	49	42	44	36	40	44	45	42	42	46

第2講 例題演習

8

X 地区における政党 A の支持率は $\frac{2}{3}$ であった。政党 A がある政策を掲げたところ、支持率が変わったのではないかと考え、アンケート調査を行うことにした。30 人に対しアンケートをとったところ、25 人が政党 A を支持すると回答した。

この結果から、政党 A の支持率は上昇したと判断してよいか。仮説検定の考え方をを用い、次の各場合について考察せよ。ただし、公正なさいころを 30 個投げて、1 から 4 までのいずれかの目が出た個数を記録する実験を 200 回行ったところ、次のようになったとし、この結果を用いよ。

(1) 基準となる確率を 0.05 とする

(2) 基準となる確率を 0.01 とする

1~4 の個数	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	計
度数	1	0	2	5	9	14	22	27	32	29	24	17	11	4	2	1	200

9

1 枚のコインを 8 回投げたところ、裏が 7 回出た。この結果から、このコインは裏が出やすいと判断してよいか。仮説検定の考え方をを用い、基準となる確率を 0.05 として考察せよ。

レベル A

1

大豆を 1 粒ずつ箸でつかみ 30 秒間で隣の器へ移す個数を競う大会が行われた。

以下のデータは、A ～ J の 10 選手が、30 秒間で隣の器へ移した大豆の個数 x (個) である。

ただし、 x のデータの平均値を $\bar{x}$ で表し、 $x < 25$ とする。

	A	B	C	D	E	F	G	H	I	J
x	22	21	16	a	13	15	24	b	12	14
$(x - \bar{x})^2$	16	9	4	c	25	9	d	4	36	16

- (1) $\bar{x}$ の値を求めよ。 (2) a, b, c, d の値を求めよ。
 (3) x の分散と標準偏差を求めよ。ただし、小数第 2 位を四捨五入せよ。

2

右の表は、30 個のボールを 3 つの組 A, B, C に分けて重さを量ったときの結果である。30 個のボールについて、重さの平均値、分散を求めよ。

	個数	平均値 (g)	標準偏差
A	15	20	1
B	5	18	5
C	10	15	2

3

ある 2 つの変量 x, y のデータについて、 $(x - \bar{x})(y - \bar{y})$ の総和が -270 、 $(x - \bar{x})^2$ の総和が 250 、 $(y - \bar{y})^2$ の総和が 810 であった。 x と y の相関係数を求めよ。ただし、 $\bar{x}$ と $\bar{y}$ はそれぞれ x のデータの平均値、 y のデータの平均値である。

4

野球において、打者の評価の指標の 1 つに「打率」がある。ここでは、打席に立った時はヒットを打つか打たないかのいずれかとし、打率を $\frac{\text{ヒットを打った回数}}{\text{打席に立った回数}}$ で定義する。一般に、この打率が高いほど、打者としての評価は高いといわれる。以下では、打率をヒットを打つ確率とする。

野球選手 A のヒットを打つ確率は、前シーズンまで $\frac{1}{4}$ であった。今年、A 選手はシーズン前のキャンプで猛練習を積み、今シーズンの開幕直後の 2 試合に出場した。次の各場合について、A 選手の打者としての評価が高まったかどうか、仮説検定の考え方をを用い、基準となる確率を 0.05 として考察せよ。

- (1) 開幕直後の 1 試合目では、5 打席中 3 打席ヒットを打った。1 試合目の成績から、A 選手の打者としての評価は高まったと判断してよいか。
- (2) 続く 2 試合目では、10 打席中 6 打席 ヒットを打った。開幕直後の 2 試合の成績から、A 選手の打者としての評価は高まったと判断してよいか。ただし、「公正なコインを 2 枚同時に投げる」という操作を 10 回繰り返したとき、「2 枚とも表が出る」ことがちょうど k 回起こる確率 p_k ($0 \leq k \leq 5$) は、次の表の通りである。

k	0	1	2	3	4	5
p_k	0.056	0.188	0.282	0.250	0.146	0.058

章末問題A

1 [2014 センター本試]

次の表は、生徒 10 人の国語と英語のテスト (各 10 点満点) の得点をまとめたものである。ただし、テストの得点は整数値をとり、 $C < D$ である。

生徒番号	1	2	3	4	5	6	7	8	9	10	平均値	分散
国語	6	5	6	4	4	9	6	3	9	8	A	B
英語	6	C	8	7	D	7	5	7	8	8	7	1

- (1) 10人の国語の得点の平均値 A は ^ア 点である。また、国語の得点の分散 B の値は ^イ である。さらに、国語の得点の中央値は ^ウ 点である。
- (2) 10 人の英語の得点の平均値が 7 点、分散が 1 であることから、 $C + D =$ ^エ,
- $(C - 7)^2 + (D - 7)^2 =$ ^オ が成り立つ。条件 $C < D$ により、 C, D の値は、それぞれ ^カ 点, ^キ 点である。
- (3) 国語と英語の得点の相関係数の値は ^ク である。

章末問題A

2 [2016 福岡大]

次の表は2つの変数 x , y からなるデータを記録したものであるが, * 印の数値が汚れて見えなくなってしまった。ただし, $\bar{x}$ は x の平均値であり, $\bar{y}$ は y の平均値である。このとき変数 y のデータの値の合計は $a = \boxed{}$ である。 x と y の相関係数 r を, 小数第3位を四捨五入して, 小数第2位まで計算すると $r = \boxed{}$ である。ただし, $\sqrt{299} = 17.3$ とする。

データの 番号	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	5	6	-1	-1	1	1	1
2	7	9	1	2	1	*	*
3	4	6	-2	*	*	*	*
4	5	4	-1	*	*	*	*
5	8	9	*	*	*	*	*
6	3	*	*	*	*	*	*
7	8	*	*	*	*	*	*
8	7	*	*	*	*	*	1
9	6	*	*	*	*	1	0
10	*	*	*	*	*	9	3
合計	*	a	*	*	*	46	28

章末問題A

③[2018 上智大]

受験者数が100人の試験が実施され、この試験を受験した智子さんの得点は84(点)であった。また、この試験の得点の平均値は60(点)であった。

なお、得点の平均値が m (点)、標準偏差が s (点)である試験において、得点が x (点)である受験者の偏差値は $50 + \frac{10(x-m)}{s}$ となることを用いてよい。

(1) 智子さんの偏差値は62であった。したがって、100人の受験者の得点の標準偏差は ア (点)である。

(2) この試験において、得点が x (点)である受験者の偏差値が65以上であるための必要十分条件は $x \geq \text{イ}$ である。

(3) 後日、この試験を新たに50人が受験し、受験者数は合計で150人となった。その結果、試験の得点の平均値が62(点)となり、智子さんの偏差値は60となった。したがって、150人の受験者の得点の標準偏差は ウ (点)である。また、新たに受験した50人の受験者の得点について、平均値は エ (点)であり、標準偏差は オ $\sqrt{\text{カ}$ (点)である。

章末問題A

4 [16センター追試]

次の表1は、あるクラスの生徒30人に行った科目Xと科目Yのテストの得点であり、これらの平均値、標準偏差、共分散をまとめたものが下の表2である。

表1 科目Xと科目Yの得点

科目X	63	76	58	71	75	56	81	80	84	77	76	63	63	59	63
科目Y	47	78	60	46	58	63	73	59	66	49	62	58	65	50	42

科目X	77	78	68	59	72	68	79	67	79	73	77	67	63	78	76
科目Y	82	66	40	55	42	69	77	57	63	52	49	45	55	84	56

表2

	平均値	標準偏差
科目X	70.9	7.81
科目Y	58.9	11.74

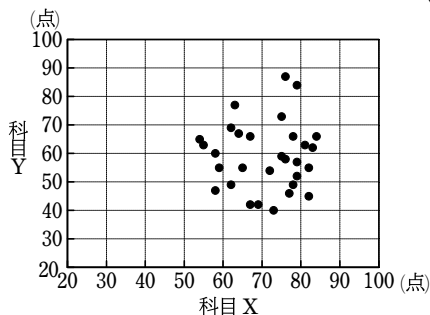
科目Xと科目Yの得点の共分散 36.89

(共分散とは、科目Xの得点の偏差と科目Yの得点の偏差の積の平均値である)

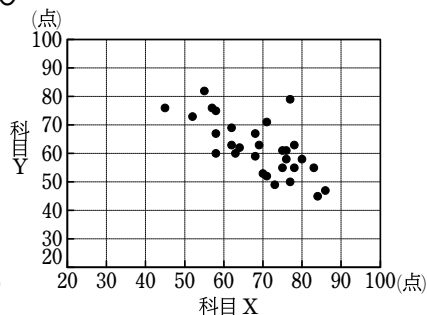
(1) 次の に当てはまるものを、下の①～③のうちから一つ選べ。

科目Xと科目Yの得点を散布図にしたものは である。

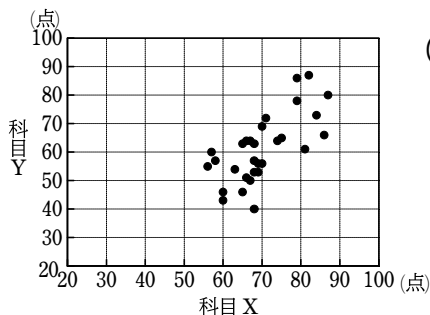
①



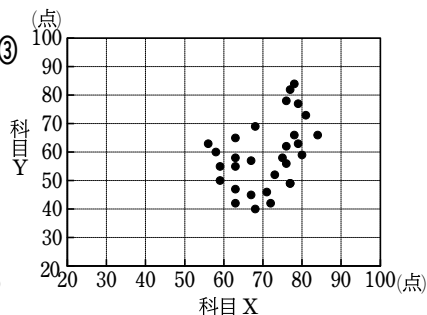
②



③



④



章末問題 A

(2) 次の に当てはまるものを、下の ㉠～㉣ のうちから一つ選べ。

表 1 の得点を $\frac{1}{2}$ にして 50 点満点の得点に換算した。例えば、62 点であった場合は得

点を 2 で割った値である 31 点とし、63 点であった場合は 31.5 点とする。

このとき、科目 X の得点の偏差と科目 Y の得点の偏差は、換算後、それぞれもとの得点の偏差の $\frac{1}{2}$ になる。

したがって、科目 X についてもとの標準偏差と換算後の標準偏差を比較し、さらにもとの共分散と換算後の共分散を比較すると、。

㉠ 換算後の標準偏差と共分散の値はともに、もとの値の $\frac{1}{2}$ になる

㉡ 換算後の標準偏差と共分散の値はともに、もとの値の $\frac{1}{4}$ になる

㉢ 換算後の標準偏差の値はもとの値の $\frac{1}{2}$ になり、共分散の値はもとの値の $\frac{1}{4}$ になる

㉣ 換算後の標準偏差の値はもとの値の $\frac{1}{4}$ になり、共分散の値はもとの値の $\frac{1}{2}$ になる

章末問題B

1 [16センター追試]

A 組から D 組の各組 30 人の生徒に対して理科のテストを行った。次の図 1 は、各組ごとに理科のテストの得点を箱ひげ図にしたものである。

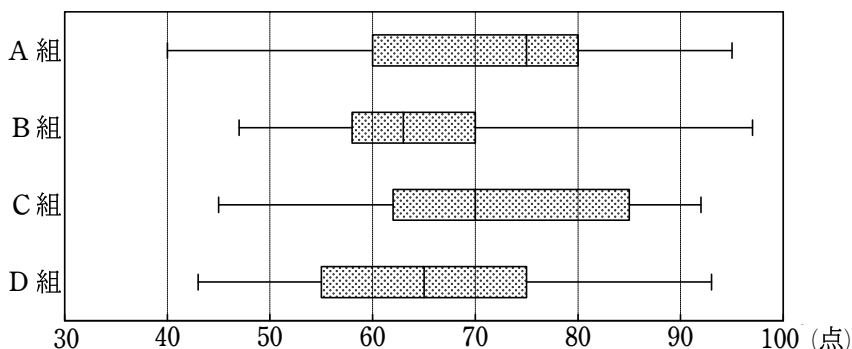


図 1 A 組から D 組の理科のテストの箱ひげ図

- (1) 次の , に当てはまるものを、下の ①～⑤ のうちから一つずつ選べ。
ただし、解答の順序は問わない。

図 1 の箱ひげ図について述べた文として誤っているものは と である。

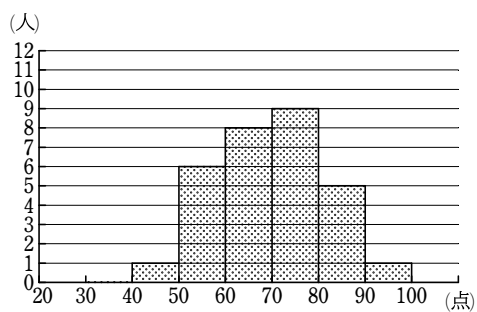
- ① A, B, C, D の 4 組全体の最高点の生徒がいるのは B 組である。
- ② A, B, C, D の 4 組で比べたとき、四分位範囲が最も大きいのは A 組である。
- ③ A, B, C, D の 4 組で比べたとき、第 1 四分位数と中央値の差が最も小さいのは B 組である。
- ④ A 組では、60 点未満の人数は 80 点以上の人数よりも多い。
- ⑤ A 組と C 組で 70 点以下の人数を比べたとき、C 組の人数は A 組の人数以上である。

章末問題B

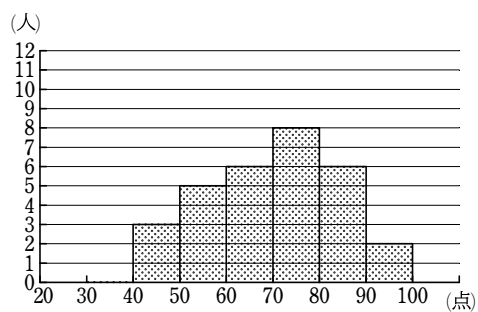
(2) 次の に当てはまるものを，下の ①～③ のうちから一つ選べ。

図1のC組の箱ひげ図のもとになった得点をヒストグラムにしたとき，対応するものは である。

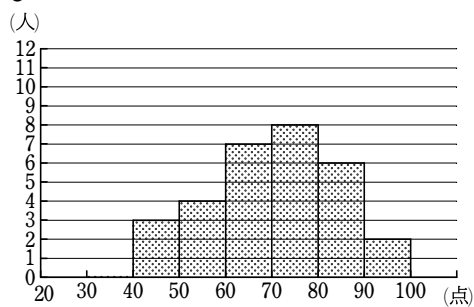
①



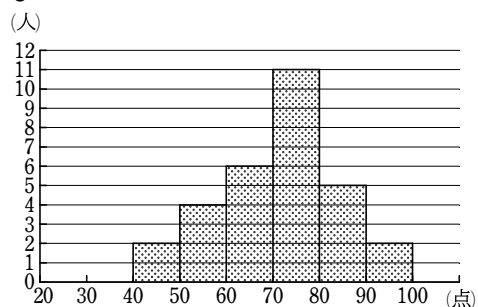
②



③



④



章末問題B

2 [2015 センター本試]

ある高校3年生1クラスの生徒40人について、ハンドボール投げの飛距離のデータを取った。次の図1は、このクラスで最初にとったデータのヒストグラムである。

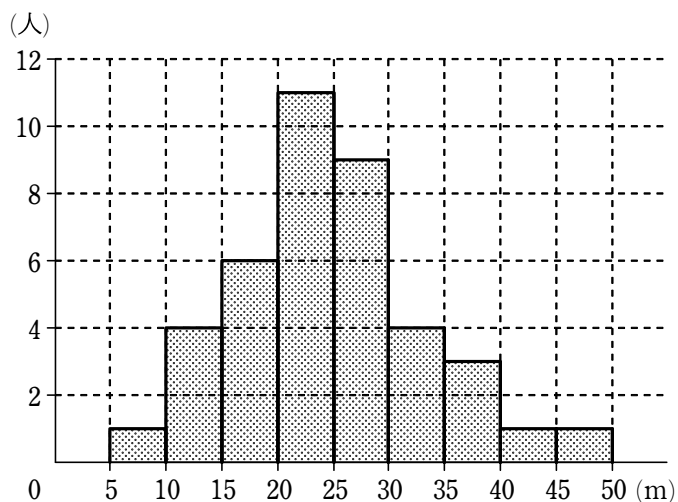


図1 ハンドボール投げ

- (1) 次の に当てはまるものを、下の①～⑧のうちから一つ選べ。

この40人のデータの第3四分位数が含まれる階級は、 である。

- | | |
|-------------------|-------------------|
| ① 5 m 以上 10 m 未満 | ⑤ 10 m 以上 15 m 未満 |
| ② 15 m 以上 20 m 未満 | ⑥ 20 m 以上 25 m 未満 |
| ③ 25 m 以上 30 m 未満 | ⑦ 30 m 以上 35 m 未満 |
| ④ 35 m 以上 40 m 未満 | ⑧ 40 m 以上 45 m 未満 |
| ⑤ 45 m 以上 50 m 未満 | |

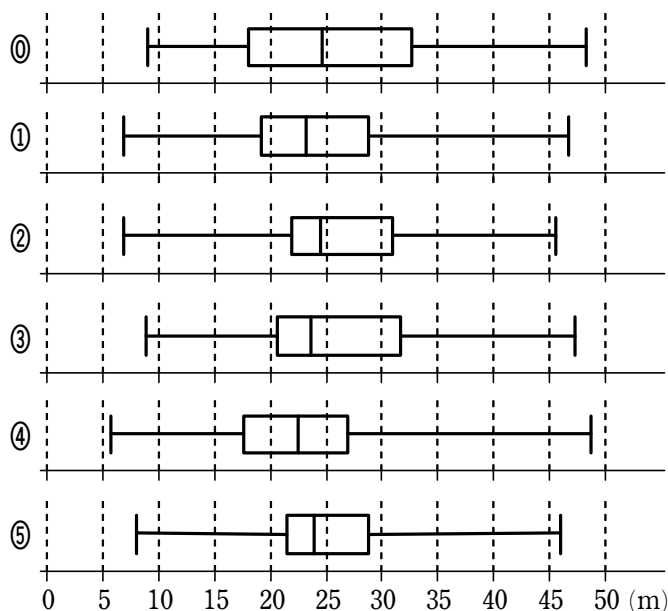
- (2) 次の ～ に当てはまるものを、下の①～⑤のうちから一つずつ選べ。

ただし、 ～ の解答の順序は問わない。

このデータを箱ひげ図にまとめたとき、図1のヒストグラムと矛盾するものは、

, , , である。

章末問題B



- (3) 次の文章中の , に入れるものとして最も適当なものを、下の①～③のうちから一つずつ選べ。ただし、 , の解答の順序は問わない。

後日、このクラスでハンドボール投げの記録を取り直した。次に示した A～D は、最初にとった記録から今回の記録への変化の分析結果を記述したものである。a～d の各々が今回取り直したデータの箱ひげ図となる場合に、①～③の組合せのうち分析結果と箱ひげ図が矛盾するものは、 , である。

- ① A-a ② B-b ③ C-c ④ D-d

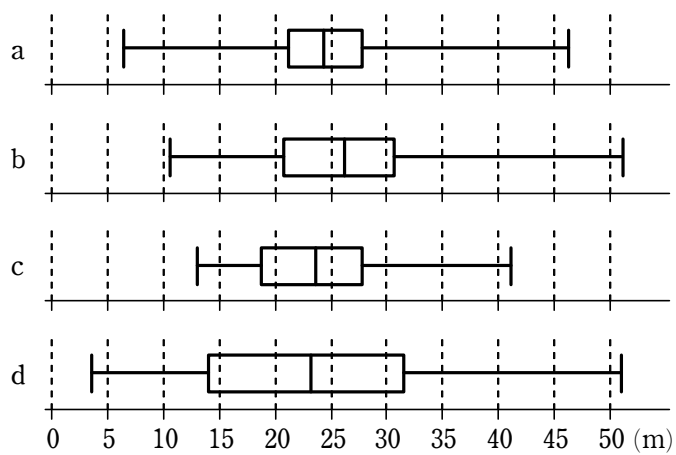
A：どの生徒の記録も下がった。

B：どの生徒の記録も伸びた。

C：最初にとったデータで上位 $\frac{1}{3}$ に入るすべての生徒の記録が伸びた。

D：最初にとったデータで上位 $\frac{1}{3}$ に入るすべての生徒の記録は伸び、下位 $\frac{1}{3}$ に入るすべての生徒の記録は下がった。

章末問題B



章末問題B

3

次の表は、五つの教科について、好きな順に1から5の順位を同じ順位がないように、10人の生徒につけてもらった結果をまとめたものである。

	教科1	教科2	教科3	教科4	教科5
生徒1	4	3	1	2	5
生徒2	5	4	2	3	1
生徒3	1	3	5	4	2
生徒4	4	5	2	3	1
生徒5	3	2	4	1	5
生徒6	4	3	2	5	1
生徒7	1	4	5	3	2
生徒8	2	3	5	1	4
生徒9	1	5	2	3	4
生徒10	4	5	2	1	3
中央値	A	3.5	2.0	3.0	2.5
平均値	2.9	3.7	3.0	2.6	B
分散	2.09	C	2.20	1.64	2.36

以下、小数の形で解答する場合、指定された桁数の一つ下の桁を四捨五入し、解答せよ。途中で割り切れた場合、指定された桁まで⑩にマークすること。

- (1) 教科1について10人の順位の中央値Aは ア イ , 教科5について10人の順位の平均値Bは ウ エ である。また、教科2について10人の順位の分散Cの値は オ カキ である。

- (2) j, k を相異なる5以下の自然数とする。教科 j と教科 k に対して、教科 j の順位が教科 k の順位より上位である生徒の人数を変数 w で表す。まず、 $j < k$ を満たす10通りの (j, k) について w をまとめると次の表になる。

(j, k)	(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	(3,4)	(3,5)	(4,5)
w	6	D	4	6	4	E	3	5	4	5

表中のD, Eの値はそれぞれ ク ケ である。また、 $\alpha \neq \beta$ としたとき、 $j = \alpha, k = \beta$ のときの w の値 w_1 と $j = \beta, k = \alpha$ のときの w の値 w_2 について、関係式 コ が成り立つから、 $j > k$ のときの w の値は上の表から求めることができる。

コ に当てはまるものを、次の①～③のうちから一つ選べ。

① $w_1 = w_2$ ② $w_1 = -w_2$ ③ $w_1 + w_2 = 10$ ④ $w_1 - w_2 = 10$

さらに、教科 j の中央値から教科 k の中央値を引いた差を変数 u 、教科 j の平均値から

章末問題B

教科 k の平均値を引いた差を変量 v で表す。たとえば, $j=2, k=3$ のとき

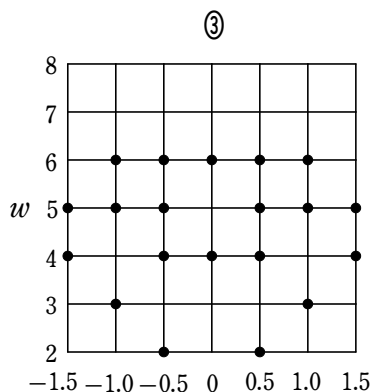
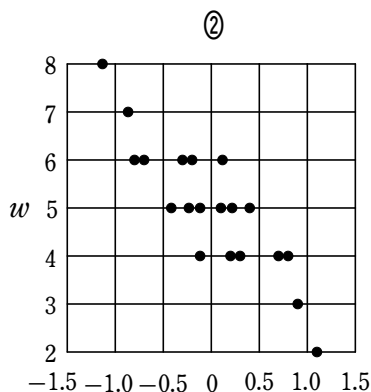
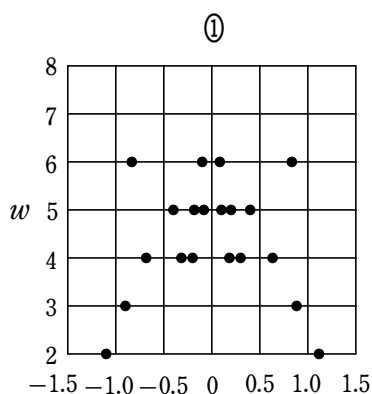
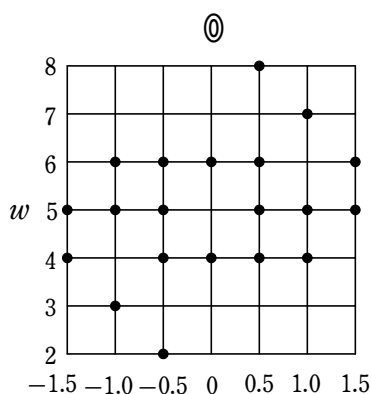
$u = \boxed{\text{サ}}, \boxed{\text{シ}}, v = \boxed{\text{ス}}, \boxed{\text{セ}}$ であり, $j=2, k=4$ のとき $u=0.5$,

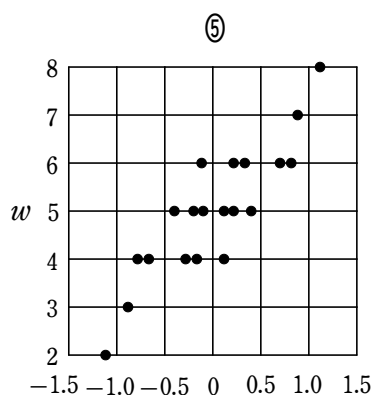
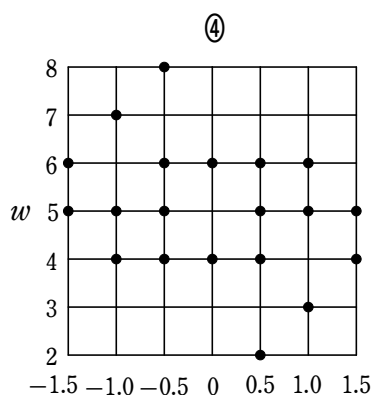
$v=1.1$ である。

以上から, $j \neq k$ を満たす 20 通りの (j, k) について, u と w の相関図 (散布図) は

$\boxed{\text{ソ}}, v$ と w の相関図は $\boxed{\text{タ}}$ となる。したがって, 変量 u と変量 w の相関係数の値を r_1 , 変量 v と変量 w の相関係数の値を r_2 とするとき, $\boxed{\text{チ}}$ が成り立つ。

$\boxed{\text{ソ}}, \boxed{\text{タ}}$ に当てはまるものを, 次の ①～⑤ のうちから一つずつ選べ。ただし, 以下の相関図では, 横軸が変量 u あるいは変量 v を表している。





チ に当てはまるものを，次の ①～⑤ のうちから一つ選べ。

① $r_2 < 0 < r_1$

② $r_1 < 0 < r_2$

③ $0 < r_2 < r_1$

④ $0 < r_1 < r_2$

⑤ $r_1 < r_2 < 0$

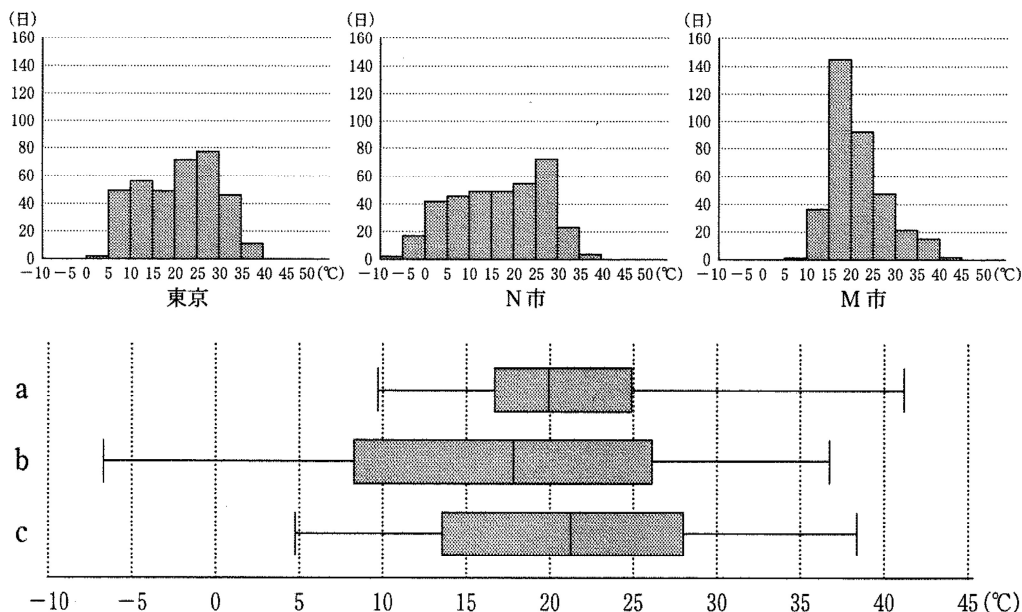
⑥ $r_2 < r_1 < 0$

章末問題C

1 [16センター本試]

世界4都市(東京、O市、N市、M市)の2013年の365日の各日の最高気温のデータについて考える。

- (1) 次のヒストグラムは、東京、N市、M市のデータをまとめたもので、この3都市の箱ひげ図は下のa、b、cのいずれかである。

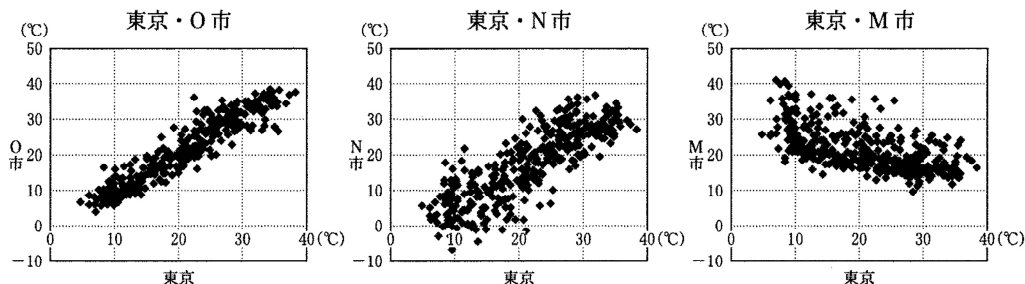


出典：『過去の気象データ』(気象庁 Web ページ) などにより作成

次の に当てはまるものを、下の ① ~ ⑤ のうちから一つ選べ。

都市名と箱ひげ図の組合せとして正しいものは、 である。

- ① 東京—a, N市—b, M市—c ② 東京—a, N市—c, M市—b
 ③ 東京—b, N市—a, M市—c ④ 東京—b, N市—c, M市—a
 ⑤ 東京—c, N市—a, M市—b ⑥ 東京—c, N市—b, M市—a
- (2) 次の3つの散布図は、東京、O市、N市、M市の2013年の365日の各日の最高気温のデータをまとめたものである。それぞれ、O市、N市、M市の最高気温を縦軸にとり、東京の最高気温を横軸にとってある。



出典：『過去の気象データ』(気象庁 Web ページ) などにより作成

章末問題C

次の , に当てはまるものを、下の ① ~ ④ のうちから一つずつ選べ。

ただし、解答の順序は問わない。

これらの散布図から読み取れることとして正しいものは、 と である。

- ① 東京と N 市、東京と M 市の最高気温の間にはそれぞれ正の相関がある。
- ② 東京と N 市の最高気温の間には正の相関、東京と M 市の最高気温の間には負の相関がある。
- ③ 東京と N 市の最高気温の間には負の相関、東京と M 市の最高気温の間には正の相関がある。
- ④ 東京と O 市の最高気温の間の相関の方が、東京と N 市の最高気温の間の相関より強い。
- ⑤ 東京と O 市の最高気温の間の相関の方が、東京と N 市の最高気温の間の相関より弱い。

- (3) 次の , , に当てはまるものを、下の ① ~ ⑨ のうちから一つずつ選べ。ただし、同じものを繰り返し選んでもよい。

N 市では温度の単位として摂氏 (°C) のほかに華氏 (°F) も使われている。華氏 (°F) での温度は、摂氏 (°C) での温度を $\frac{9}{5}$ 倍し、32 を加えると得られる。例えば、摂氏 10 °C は、 $\frac{9}{5}$ 倍し 32 を加えることで華氏 50 °F となる。

したがって、N 市の最高気温について、摂氏での分散を X 、華氏での分散を Y とすると、 $\frac{Y}{X}$ は になる。

東京 (摂氏) と N 市 (摂氏) の共分散を Z 、東京 (摂氏) と N 市 (華氏) の共分散を W とすると、 $\frac{W}{Z}$ は になる (ただし、共分散は 2 つの変量のそれぞれの偏差の積の平均値)。

東京 (摂氏) と N 市 (摂氏) の相関係数を U 、東京 (摂氏) と N 市 (華氏) の相関係数を V とすると、 $\frac{V}{U}$ は になる。

- | | | | | |
|--------------------|------------------|--------|------------------|--------------------|
| ① $-\frac{81}{25}$ | ② $-\frac{9}{5}$ | ③ -1 | ④ $-\frac{5}{9}$ | ⑤ $-\frac{25}{81}$ |
| ⑥ $\frac{25}{81}$ | ⑦ $\frac{5}{9}$ | ⑧ 1 | ⑨ $\frac{9}{5}$ | ⑩ $\frac{81}{25}$ |

章末問題C

2 [19センター本試]

全国各地の気象台が観測した「ソメイヨシノ(桜の種類)の開花日」や、「モンシロチョウの初見日(初めて観測した日)」, 「ツバメの初見日」などの日付を気象庁が発表している。気象庁発表の日付は普通の月日形式であるが, この問題では該当する年の1月1日を「1」とし, 12月31日を「365」(うるう年の場合は「366」)とする「年間通し日」に変更している。例えば, 2月3日は, 1月31日の「31」に2月3日の3を加えた「34」となる。

(1) 図1は全国48地点で観測しているソメイヨシノの2012年から2017年までの6年間の開花日を, 年ごとに箱ひげ図にして並べたものである。図2はソメイヨシノの開花日の年ごとのヒストグラムである。ただし, 順番は年の順に並んでいるとは限らない。なお, ヒストグラムの各階級の区間は, 左側の数値を含み, 右側の数値を含まない。次の , に当てはまるものを, 図2の ㊸ ~ ㊾ のうちから一つずつ選べ。

- ・ 2013年のヒストグラムは である。
- ・ 2017年のヒストグラムは である。

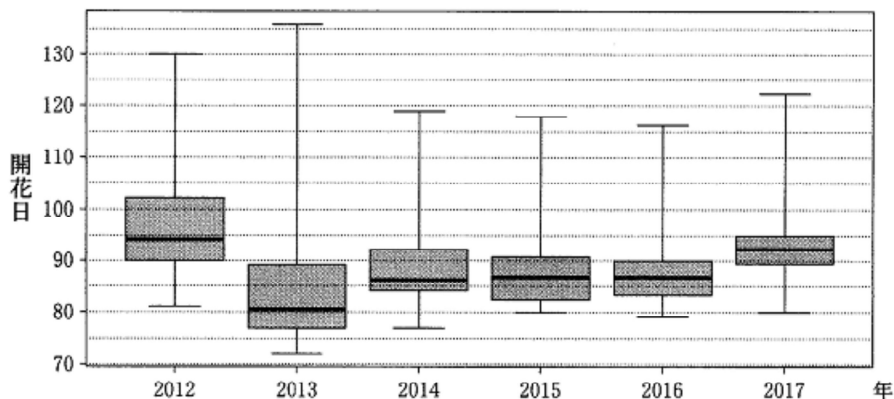


図1 ソメイヨシノの開花日の年別の箱ひげ図

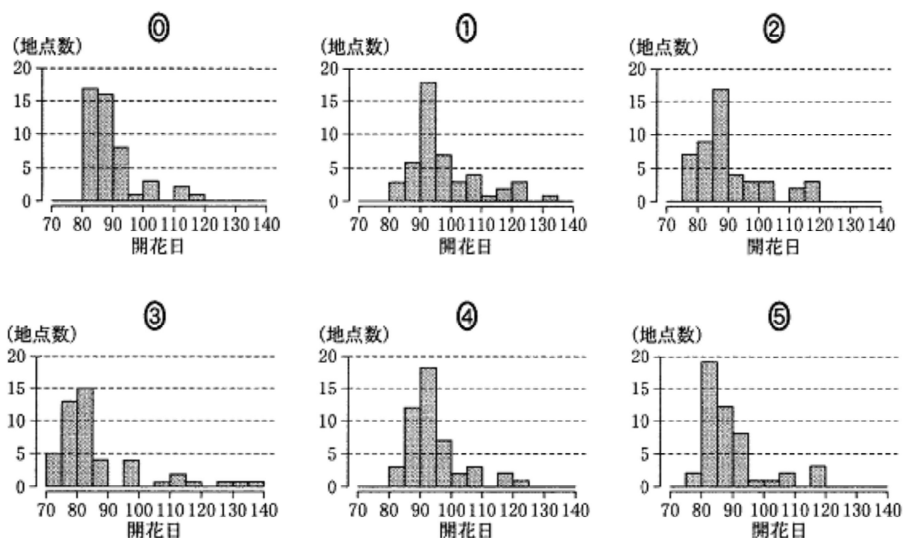


図2 ソメイヨシノの開花日の年別のヒストグラム

(出典：図1、図2は気象庁「生物季節観測データ」Webページにより作成)

(2) 図3と図4は、モンシロチョウとツバメの両方を観測している41地点における、2017年の初見日の箱ひげ図と散布図である。

散布図の点には重なった点が2点ある。なお、散布図には原点を通り傾き1の直線(実線)、切片が-15および15で傾きが1の2本の直線(破線)を付加している。

次の 、 に当てはまるものを、下の①～⑦のうちから一つずつ選べ。ただし、解答の順序は問わない。

図3、図4から読み取れることとして正しくないものは、、 である。

- ① モンシロチョウの初見日の最小値はツバメの初見日の最小値と同じである。
- ② モンシロチョウの初見日の最大値はツバメの初見日の最大値より大きい。
- ③ モンシロチョウの初見日の中央値はツバメの初見日の中央値より大きい。

章末問題C

- ③ モンシロチョウの初見日の四分位範囲はツバメの初見日の四分位範囲の3倍より小さい。
- ④ モンシロチョウの初見日の四分位範囲は15日以下である。
- ⑤ ツバメの初見日の四分位範囲は15日以下である。
- ⑥ モンシロチョウとツバメの初見日が同じ所が少なくとも4地点ある。
- ⑦ 同一地点でのモンシロチョウの初見日とツバメの初見日の差は15日以下である。

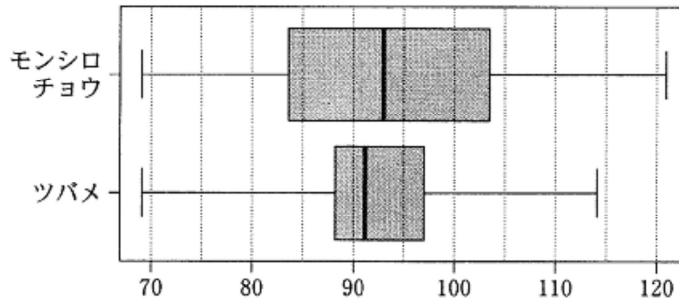


図3 モンシロチョウとツバメの初見日(2017年)の箱ひげ図

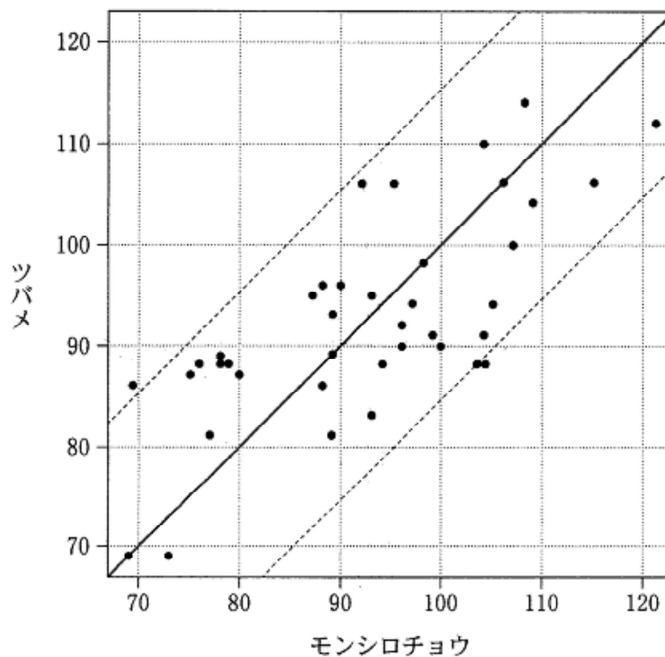


図4 モンシロチョウとツバメの初見日(2017年)の散布図

(出典：図3、図4は気象庁「生物季節観測データ」Webページにより作成)

章末問題C

(3) 一般に n 個の数値 $x_1, x_2, \dots, x_n$ からなるデータ X の平均値を $\bar{x}$, 分散を s^2 ,

標準偏差を s とする。各 x_i に対して $x'_i = \frac{x_i - \bar{x}}{s}$ ($i=1, 2, \dots, n$) と変換した $x'_1,$

$x'_2, \dots, x'_n$ をデータ X' とする。ただし, $n \geq 2, s > 0$ とする。

次の , , に当てはまるものを, 下の ① ~ ⑧ のうちから一つずつ選べ。ただし, 同じものを繰り返し選んでもよい。

・ X の偏差 $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$ の平均値は である。

・ X' の平均値は である。

・ X' の標準偏差は である。

① 0 ② 1 ③ -1 ④ $\bar{x}$ ⑤ s

⑥ $\frac{1}{s}$ ⑦ s^2 ⑧ $\frac{1}{s^2}$ ⑨ $\frac{\bar{x}}{s}$

図 4 で示されたモンシロチョウの初見日のデータ M とツバメの初見日のデータ T について上の変換を行ったデータをそれぞれ M', T' とする。

次の に当てはまるものを, 図 5 の ① ~ ④ のうちから一つ選べ。

変換後のモンシロチョウの初見日のデータ M' と変換後のツバメの初見日のデータ T' の散布図は, M' と T' の標準偏差の値を考慮すると である。

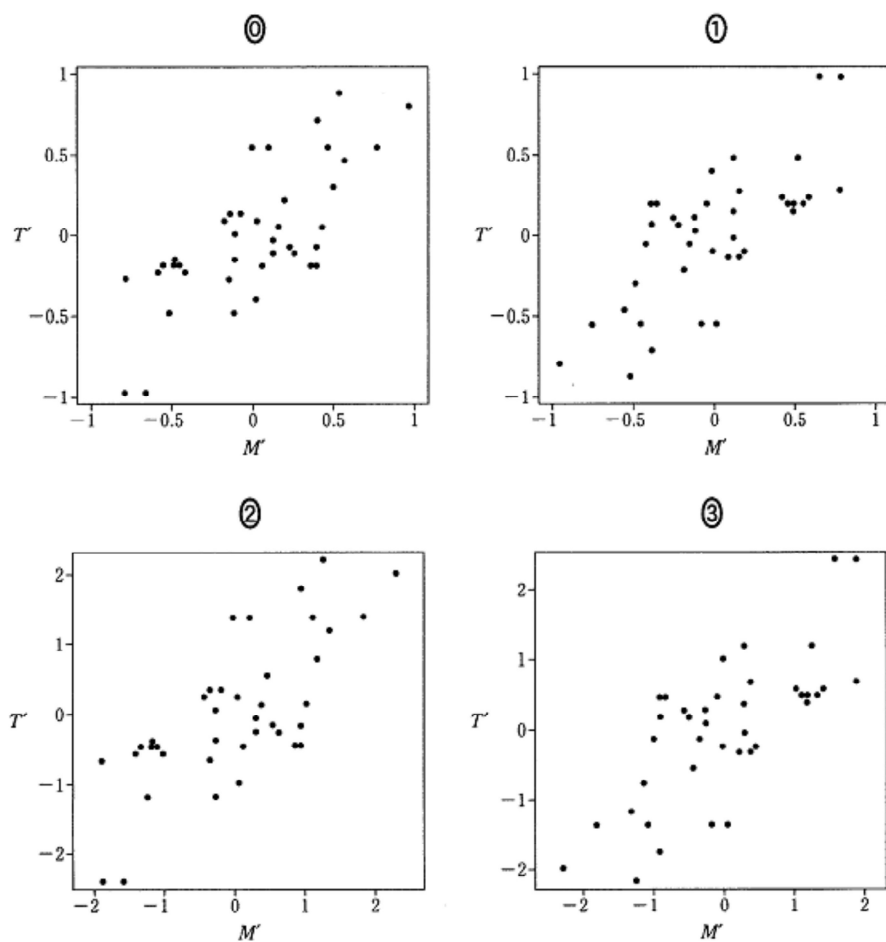


図5 四つの散布図