

第5章 | データの分析



1 データの整理

A 度数分布表

B ヒストグラム

まとめ

1 変量

人の身長、体重や運動の記録などのように、ある特性を数量的に表すものを変量という。数学では、ある変量の測定値や観測値の集まりをデータという。

2 分布

データの散らばりの様子を分布という。

3 度数分布表

度数分布表において、区切られた各区間を階級、区間の幅を階級の幅、各階級に含まれる値の個数を度数という。また、各階級の真ん中の値を階級値という。データを度数分布表に整理すると、その傾向がわかりやすくなる。

4 ヒストグラム

度数分布表を、柱状のグラフで表したものをヒストグラムという。

ヒストグラムの各長方形の高さは、底辺部分の各階級の度数を表している。

5 章

データの分析

練習 1

教 p.169

教科書 168 ページの度数分布表について、次の問いに答えよ。

- (1) 度数が 9 である階級の階級値を求めよ。
- (2) 最高気温が 21°C 以上の日は何日あるか。
- (3) 最高気温が低い方から数えて 12 番目の日が含まれる階級をいえ。

指針 度数分布表と階級値

- (1) 階級値とは、度数分布表の各階級の真ん中の値。
- (2), (3) 度数分布表の度数に着目。

解答 (1) 度数 9 は、階級が 24°C 以上 27°C 未満のときの値であるから

$$\frac{24+27}{2}=25.5 (^{\circ}\text{C}) \quad \text{答}$$

- (2) $10+9+1=20$ (日) 答

- (3) $4+6<12<4+6+10$ であるから

21°C 以上 24°C 未満 答

練習
2

次のデータは、ある高校の1年生女子20人の、ハンドボール投げの記録である。

16	13	14	13	14	11	14	17	16	15
18	16	17	15	17	16	21	16	20	19

(単位はm)

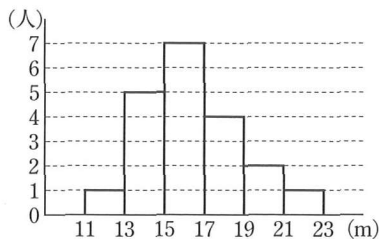
- (1) 11 m 以上 13 m 未満を階級の1つとして、どの階級の幅も2 m である度数分布表を作れ。
- (2) (1)で作った度数分布表をもとにして、ヒストグラムをかけ。

指針 度数分布表とヒストグラム 各階級の度数を調べるときは、もれや重複がないよう、必ず度数の合計が値の個数20と一致するかどうか確かめる。度数分布表の各階級の度数が、ヒストグラムの各長方形の高さである。

解答 (1) 答

階級 (m)	度数 (人)
11 以上 13 未満	1
13 ~ 15	5
15 ~ 17	7
17 ~ 19	4
19 ~ 21	2
21 ~ 23	1
計	20

(2) 答



2 データの代表値

A 平均値 B 最頻値 C 中央値

まとめ

1 代表値

データ全体の特徴を適当な1つの数値で表した値を、データの代表値という。

2 データの大きさ

データにおける測定値や観測値の個数を、そのデータの大きさという。

3 平均値

変量 x について、大きさ n のデータの値の総和を n で割った値を、このデータの平均値といい、 $\bar{x}$ で表す。

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

4 最頻値

データにおいて、最も個数の多い値を、そのデータの最頻値またはモードという。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を最頻値とする。

5 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を中央値またはメジアンという。データの大きさが偶数のときは、中央

データの大きさが奇数

小←データの値→大

○ ○ ○ ○ ○

↑

中央値

データの大きさが偶数

小←データの値→大

○ ○ ○ ○ ○ ○ ○ ○

↑↑

中央の2つの値の
平均が中央値

に2つの値が並ぶから、その平均をとって中央値とする。

データの中に、極端に離れた値がある場合、その値が全体の平均値に影響するため、データの代表値として適切ではない。このような場合は、中央値が代表値として適していることがある。

5章

データの分析

練習
3

教 p.171

次のデータは、ある生徒のある1週間における1日あたりの睡眠時間である。このデータの平均値を求めよ。

400, 410, 420, 390, 430, 450, 440 (分)

指針 平均値 平均値 $\bar{x}$ は、 $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ で求められる。

解答 $\frac{1}{7}(400 + 410 + 420 + 390 + 430 + 450 + 440) = \frac{2940}{7} = 420$ (分) 答

練習
4

教 p.171

教科書 168 ページの度数分布表において、最頻値を求めよ。

指針 最頻値の求め方 度数が最も大きい階級の階級値を求める。

解答 度数の最大値は 10 であり、その階級は、21℃ 以上 24℃ 未満である。
最頻値はこの階級の階級値であるから

$$\frac{21+24}{2}=22.5\text{ (}^{\circ}\text{C)} \quad \text{答 } 22.5\text{ }^{\circ}\text{C}$$

練習
5

教 p.172

次のデータは、8 人の生徒の右手の握力を測った結果である。その中央値を求めよ。

38, 56, 43, 41, 35, 49, 51, 31 (kg)

指針 中央値の求め方 データの大きさが 8 個で偶数であるから、中央の 2 つの値の平均を中央値とする。

解答 このデータを小さい順に並べると

31, 35, 38, 41, 43, 49, 51, 56

よって、このデータの中央値は $\frac{41+43}{2}=42\text{ (kg)}$ 答 42 kg

3 データの散らばりと四分位数

A 範囲

まとめ

範囲

データの最大値から最小値を引いた差を、データの 範囲 という。範囲は、データの散らばりの度合いを表す 1 つの値である。

練習
6

教 p.174

次のデータは、教科書例 4 と同じ年の B 市における月ごとの降水日数である。このデータの範囲を求めよ。また、データの散らばりの度合いが大きいのは A 市、B 市のどちらと考えられるか。得られたデータの範囲によって比較せよ。

19, 16, 12, 11, 6, 8, 21, 13, 10, 14, 18, 22 (日)

指針 範囲 まず、データの最大値と最小値を求める。その差がデータの範囲である。

解答 B市のデータの最大値は22日、最小値は6日である。

よって、B市のデータの範囲は $22-6=16$ (日) 答 16日

次に、教科書例4から、A市のデータの範囲は10日である。

よって、データの散らばりの度合いが大きいのは B市 答

B 四分位数

まとめ

1 四分位数

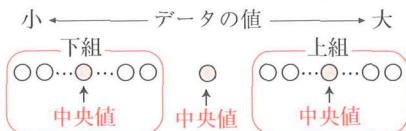
データを値の大きさに順に並べたとき、4等分する位置にくる値を四分位数という。四分位数は、小さい方から順に第1四分位数、第2四分位数、第3四分位数といい、順に Q_1 、 Q_2 、 Q_3 で表す。

四分位数の求め方

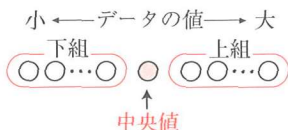
① データを値の大きさに順に並べ、中央値を求める。

② ①の中央値を境界としてデータの個数を2等分し、値が中央値以下の下組と、値が中央値以上の上組に分ける。ただし、データの大きさが奇数のとき、①の中央値は、下組と上組の両方に含めない。

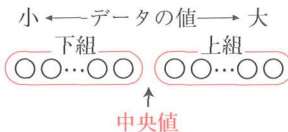
③ 下組の中央値、上組の中央値を求める。



データの大きさが奇数



データの大きさが偶数



2 四分位範囲、四分位偏差

データの第3四分位数 Q_3 と第1四分位数 Q_1 の差 $Q_3 - Q_1$ を四分位範囲といい、四分位範囲の半分を、四分位偏差という。

四分位範囲は、データの中に極端に離れた値がある場合でも、その影響を受けにくい。データの値が中央値の周りに集中しているほど、四分位範囲は小さくなる傾向にある。

5 章

データの分析

練習
7

次のデータ A, B のそれぞれについて、四分位範囲を求めよ。また、データの散らばりの度合いが大きいのは A, B のどちらと考えられるか。得られた四分位範囲によって比較せよ。

A 21, 29, 32, 36, 38, 40, 49, 53, 55, 68, 80

B 25, 31, 39, 42, 45, 46, 50, 53, 54, 65, 80

指針 四分位範囲 まず、データの中央値を境界として上組と下組の中央値を求める。下組の中央値を Q_1 、上組の中央値を Q_3 とする。

このとき四分位範囲は $Q_3 - Q_1$ である。

解答 データ A の第 2 四分位数 Q_2 は 40

データ 21, 29, 32, 36, 38 の中央値は 32

データ 49, 53, 55, 68, 80 の中央値は 55

よって、データ A の第 1 四分位数 Q_1 は 32, 第 3 四分位数 Q_3 は 55

したがって、四分位範囲は $55 - 32 = 23$ **答**

データ B の第 2 四分位数 Q_2 は 46

データ 25, 31, 39, 42, 45 の中央値は 39

データ 50, 53, 54, 65, 80 の中央値は 54

よって、データ B の第 1 四分位数 Q_1 は 39, 第 3 四分位数 Q_3 は 54

したがって、四分位範囲は $54 - 39 = 15$ **答**

データの散らばりの度合いが大きいほど四分位範囲は大きくなるから、
 $23 > 15$ より、データの散らばりの度合いが大きいのは

データ A **答**

考える

データの散らばり具合を知ることのよさについて考えてみよう。たとえば、教科書 174 ページのデータ A, B が各店で手作りするパンのチェーン店 A, B で作られた同じ商品 150 個の重さをヒストグラムにまとめたものとする。このとき、どちらのお店で買いたいと考えるか。理由も合わせて考えてみよう。

指針 ヒストグラムの縦軸はパン 1 個あたりの重さ、横軸は作ったパンの個数とみる。A, B のグラフを見比べると、A 店の方が B 店より重さにばらつきがない。

解答 基準となるパンの重量が大きい方がより大きいパンであり、買う側にとってはより大きいパンを買う方が得をするもの、という前提で考える。

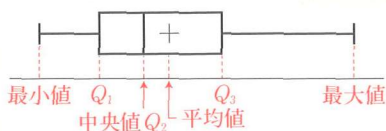
- ① 損はしたくないので、重さの散らばり具合が小さいA店で買いたい。
という考え方がある一方で
- ② 基準より少し小さいパンになる可能性もあるけれど、大きいパンになる可能性もあるので、B店で買いたい。
という考え方もある。 **答**

C 箱ひげ図

まとめ

1 箱ひげ図

データの分布の特徴を表す右のような図を、箱ひげ図という。平均値は省略することもある。



2 箱ひげ図のかき方

- ① 横軸にデータの値の目盛りをとるとする。
- ② Q_1 を左端、 Q_3 を右端とする箱をかき、箱の中に中央値 (Q_2) を示す縦線をかく。
- ③ 箱の左端から最小値まで、箱の右端から最大値まで線分を引く。箱ひげ図の箱の長さは、データの散らばりの度合いを表している。

注意 箱ひげ図は縦にかいてもよい。

5章

データの分析

練習 8

教 p.178

次のデータは、10 人の生徒に 100 点満点の数学、英語、国語のテストを行った結果である。単位は点である。

数学 68, 35, 86, 63, 30, 91, 50, 63, 46, 58

英語 75, 65, 90, 78, 52, 88, 70, 75, 59, 82

国語 63, 60, 73, 75, 58, 79, 68, 70, 66, 80

- (1) これらのデータの箱ひげ図を並べてかけ。
- (2) データの散らばりの度合いが最も大きいのは、数学、英語、国語のうちどれと考えられるか。(1) で得られた箱ひげ図を用いて比較せよ。

指針 **箱ひげ図によるデータの分析** まず、データを大きさの順に並べ、 Q_1 、 Q_2 、 Q_3 の値を求めて箱をかく。最大値、最小値を調べて、箱から線分を引く。箱ひげ図では、箱が長い方が、データの散らばりの度合いが大きい。ここでは、箱ひげ図を縦にかいてみよう。

解答 (1) それぞれのデータを、得点の低い方から順に並べる。

数学 30, 35, 46, 50, 58, 63, 63, 68, 86, 91

英語 52, 59, 65, 70, 75, 75, 78, 82, 88, 90

国語 58, 60, 63, 66, 68, 70, 73, 75, 79, 80

それぞれのデータについて、 Q_1 , Q_2 , Q_3 (点) を求める。

$$\text{数学 } Q_2 = \frac{58+63}{2} = 60.5$$

$$Q_1 = 46, Q_3 = 68$$

$$\text{英語 } Q_2 = \frac{75+75}{2} = 75$$

$$Q_1 = 65, Q_3 = 82$$

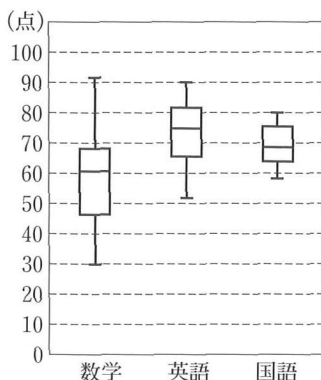
$$\text{国語 } Q_2 = \frac{68+70}{2} = 69$$

$$Q_1 = 63, Q_3 = 75$$

箱ひげ図は右の図のようになる。 答

(2) 箱が最も長いデータが、散らばりの度合いが最も大きいから

数学 答



D 箱ひげ図とヒストグラム E 外れ値

まとめ

1 外れ値

データの値のうち、他の値から極端に離れたものを外れ値という。

外れ値の基準はいくつかあり、たとえば、次の値を外れ値とする。

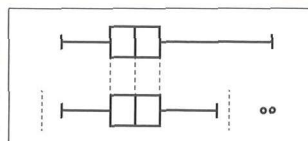
(第1四分位数 $-1.5 \times$ 四分位範囲) 以下の値

(第3四分位数 $+1.5 \times$ 四分位範囲) 以上の値

2 箱ひげ図と外れ値

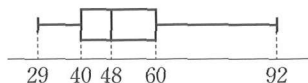
外れ値がある場合、箱ひげ図では外れ値を。で示す。

注意 外れ値を。で示す箱ひげ図をかく場合でも、四分位数は外れ値を除かないすべてのデータの四分位数であり、その値にもとづいて箱をかく。



練習
9

右の図はあるデータの箱ひげ図である。このデータの最大値 92, 最小値 29 は外れ値であるかを, 四分位範囲を利用して調べよ。



指針 **外れ値の判定** 外れ値の定義にしたがう。第 1, 第 3 四分位数を確認する。

解答 $Q_1=40$, $Q_3=60$ であるから, 四分位範囲は $Q_3-Q_1=20$

$$\text{ゆえに } Q_1-1.5 \times (Q_3-Q_1)=40-1.5 \times 20=10$$

$$Q_3+1.5 \times (Q_3-Q_1)=60+1.5 \times 20=90$$

よって, 四分位範囲を利用して外れ値を調べると, 最大値 92 は外れ値であり, 最小値 29 は外れ値ではない。答

4 分散と標準偏差

5
章

データの分析

A 分散, 標準偏差

まとめ

1 偏差

変量 x のデータの値 $x_1, x_2, \dots, x_n$ の平均値を $\bar{x}$ とするとき, 各値と平均値との差 $x_1-\bar{x}, x_2-\bar{x}, \dots, x_n-\bar{x}$ を, それぞれ平均値からの 偏差 といい, $x-\bar{x}$ で表す。偏差の総和は 0 になるから, 偏差の平均値も 0 である。

2 分散

データにおいて, 偏差の 2 乗の平均値を 分散 といい, 散らばりの度合いを表す。

3 標準偏差

分散の正の平方根を 標準偏差 といい, s で表す。

4 分散と標準偏差

変量 x のデータの値が $x_1, x_2, \dots, x_n$ で, その平均値が $\bar{x}$ のとき

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

$$\text{標準偏差} \quad s = \sqrt{\text{分散}}$$

練習
10

次のデータは, 10 人の生徒に計算テストを行った結果である。このデータの分散, 標準偏差を求めよ。

6, 10, 7, 7, 5, 4, 9, 10, 5, 7 (点)

指針 データの分散, 標準偏差 まず, 平均値 $\bar{x}$ を求める。次に, データの各値と平均値との差で偏差を求め, 分散と標準偏差を求める。

解答 10 人の計算テストの得点 x のデータの平均値 $\bar{x}$ を求めると

$$\bar{x} = \frac{6+10+7+7+5+4+9+10+5+7}{10} = \frac{70}{10} = 7 \text{ (点)}$$

よって, 各 x について $(x-\bar{x})^2$ を計算すると下の表のようになる。

x	6	10	7	7	5	4	9	10	5	7	計 70
$(x-\bar{x})^2$	1	9	0	0	4	9	4	9	4	0	計 40

したがって 分散 s^2 は $s^2 = \frac{1}{10} \times 40 = 4$ 答

標準偏差 s は $s = \sqrt{4} = 2$ (点) 答

B 分散と平均値の関係式

まとめ

分散と平均値の関係式

分散 $s^2 = \frac{1}{n} \{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\}$ の式を変形すると

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

変量 x の n 個のデータが $x_1, x_2, \dots, x_n$ のとき, $x_1^2, x_2^2, \dots, x_n^2$ を変量 x^2 の n 個のデータと考えることにすると

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

練習 11

教 p.182

次のデータについて, 分散, 標準偏差を求めよ。

5, 3, 6, 8, 5, 8, 5, 4, 6, 5

指針 データの分散, 標準偏差 x のデータの平均値を $\bar{x}$, x^2 のデータの平均値を $\overline{x^2}$ とすると, x のデータの分散は $s^2 = \overline{x^2} - (\bar{x})^2$ と表される。

標準偏差 s は, 分散 s^2 の正の平方根で求める。

解答 与えられたデータを x として表にすると, 次のようになる。

x	5	3	6	8	5	8	5	4	6	5	計 55
x^2	25	9	36	64	25	64	25	16	36	25	計 325

$$x \text{ のデータの平均値 } \bar{x} \text{ は } \bar{x} = \frac{1}{10} \times 55 = 5.5$$

$$x^2 \text{ のデータの平均値 } \overline{x^2} \text{ は } \overline{x^2} = \frac{1}{10} \times 325 = 32.5$$

$$\text{分散 } s^2 \text{ は } s^2 = \overline{x^2} - (\bar{x})^2 = 32.5 - 5.5^2 = 2.25 \quad \text{答}$$

$$\text{標準偏差 } s \text{ は } s = \sqrt{2.25} = 1.5 \quad \text{答}$$

研究 変量の変換

まとめ

変量の変換

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章

データの分析

練習 1

教 p.183

ある都市の日ごとの最高気温を摂氏度 ($^{\circ}\text{C}$) で計測し, 20 日分のデータを得た。その平均値は 15.0°C , 分散は 9.0 であった。このデータを華氏度 ($^{\circ}\text{F}$) に変更したときの, 平均値, 分散, 標準偏差を求めよ。ただし, 摂氏度が $x^{\circ}\text{C}$ のときの華氏度を $y^{\circ}\text{F}$ とすると, x と y には次の関係がある。

$$y = 1.8x + 32$$

指針 変量の変換 変量 x のデータから $y = ax + b$ (a, b は定数) 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

$$\bar{y} = a\bar{x} + b, \quad s_y^2 = a^2 s_x^2, \quad s_y = |a| s_x$$

が成り立つ。これを利用する。

解答 摂氏度が $x^{\circ}\text{C}$ のときの華氏度 $y^{\circ}\text{F}$ の, x と y の間には

$$y = 1.8x + 32$$

の関係があり, x のデータの平均値は 15.0°C , 分散は 9.0 であるから, 華氏度に変更したときの, 平均値, 分散, 標準偏差は

$$\text{平均値} \quad 1.8 \times 15.0 + 32 = 27.0 + 32 = 59.0 (^{\circ}\text{F}) \quad \text{答}$$

$$\text{分散} \quad 1.8^2 \times 9.0 = 29.16 \quad \text{答}$$

$$\text{標準偏差} \quad 1.8 \times \sqrt{9.0} = 1.8 \times 3.0 = 5.4 (^{\circ}\text{F}) \quad \text{答}$$

仮平均

変量 x のデータの平均値や分散を求めるとき、 x_0 を与えて、 $x - x_0$ のデータを考えると見通しよく計算できることがある。この x_0 を仮平均という。

練習
2

教 p.184

変量 x のデータが次のように与えられている。

750, 740, 720, 770, 750, 740

いま、 $c=10$, $x_0=740$, $u = \frac{x-x_0}{c}$ として新しい変量 u を作る。

- (1) 変量 u のデータの平均値と標準偏差を求めよ。
- (2) 変量 x のデータの平均値と標準偏差を求めよ。

指針 変量の変換と平均値, 標準偏差

(1) $u = \frac{x-740}{10}$ により、 x のデータを u のデータに変換する。

(2) $x=10u+740$ の関係を用いて変換する。

解答 (1) 変量 u のデータ, 変量 u^2 のデータの値は, それぞれ次の表のようになる。

x	750	740	720	770	750	740	計
u	1	0	-2	3	1	0	3
u^2	1	0	4	9	1	0	15

上の表から, 変量 u , u^2 のデータの平均 $\bar{u}$, $\bar{u^2}$ はそれぞれ

$$\bar{u} = \frac{1}{6} \times 3 = \frac{1}{2}, \quad \bar{u^2} = \frac{1}{6} \times 15 = \frac{5}{2}$$

よって, 変量 u のデータの分散 s_u^2 は

$$s_u^2 = \frac{5}{2} - \left(\frac{1}{2}\right)^2 = \frac{10-1}{4} = \frac{9}{4}$$

変量 u のデータの標準偏差 s_u は $s_u = \frac{3}{2}$

答 平均値 $\frac{1}{2}$, 標準偏差 $\frac{3}{2}$ (平均値 0.5, 標準偏差 1.5)

(2) $x=cu+x_0$, $c=10$, $x_0=740$ であるから

変量 x のデータの平均値 $\bar{x}$ は $\bar{x} = 10 \times \frac{1}{2} + 740 = 745$

変量 x のデータの標準偏差 s_x は $s_x = 10 \times \frac{3}{2} = 15$

答 平均値 745, 標準偏差 15



5 2つの変量の間関係

A 散布図

まとめ

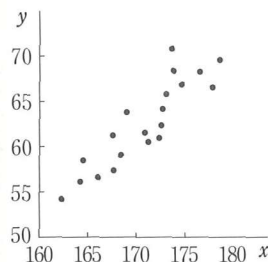
1 相関

2つの変量からなるデータの間に、一方が増加すればそれにしたがって他方も増加する、または他方が減少するという傾向がみられるとき、2つの変量の間に相関がある、または相関関係があるという。

2 散布図

右の図のように、2つの変量からなるデータを平面上に図示したものを散布図という。

散布図をかくと、データにおける2つの変量の間の関連性を視覚的にとらえることができる。



5章

データの分析

B 正の相関、負の相関

まとめ

1 正の相関・負の相関

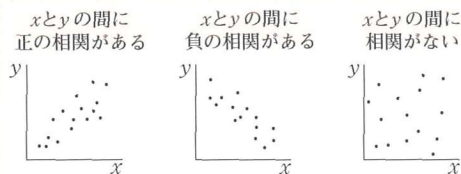
2つの変量からなるデータにおいて一方が増加すると他方も増加する傾向がみられるとき、2つの変量には正の相関があるという。

また、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。

どちらの傾向もみられないときは相関がないまたは相関関係がないという。

2 散布図の分類

散布図には、次の3つの場合がある。



3 相関の強弱

2つの変量の間に相関があるとき、散布図における点の分布の様子が1つの直線に接近しているほど相関が強いといい、散らばっているほど相関が弱いという。

練習
12

下の表は、各地点の緯度 x (度) と 2018 年 4 月の平均気温 y ($^{\circ}\text{C}$) を調べた結果である。

地点	札幌	青森	仙台	東京	長野	大阪	高知	鹿児島
x	43.1	40.8	38.3	35.7	36.7	34.7	33.6	31.6
y	8.2	9.6	12.5	17.0	13.1	16.9	17.3	18.5

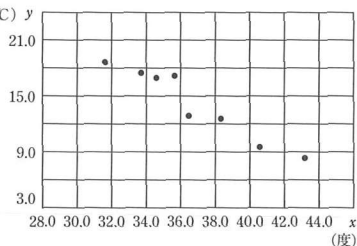
(気象庁ホームページより作成)

- (1) 2 つの変数 x , y の散布図をかけ。
- (2) x と y の間には、正、負どちらの相関があると考えられるか。

指針 2 つの変数の間の関係

- (2) (1) でかいた散布図から、 x の値が増えると y の値が増えるのか減るのか、どちらともいえないかを判断する。

解答 (1) 答 (1) $(^{\circ}\text{C})$ y



- (2) (1) の散布図より、 x の値が増えると、 y の値は減る傾向があるから、 x と y の間には、負の相関があると考えられる。 答

C 相関係数

まとめ

1 共分散, 相関係数

x と y の間の相関関係を調べるのに, x の偏差と y の偏差の積 $(x_1 - \bar{x})(y_1 - \bar{y}), (x_2 - \bar{x})(y_2 - \bar{y}), \dots, (x_n - \bar{x})(y_n - \bar{y})$ について, その平均値 $\frac{1}{n} \{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})\}$ を x と y の共分散といい, s_{xy} で表す。 x と y の間に, 正の相関があるとき共分散は正となり, 負の相関があるとき共分散は負となる。

相関の強弱をみるために, 共分散 s_{xy} を, x の標準偏差 s_x と y の標準偏差 s_y の積 $s_x s_y$ で割った値を考える。この値を x と y の相関係数といい, r で表す。

$$r = \frac{s_{xy}}{s_x s_y} \quad \leftarrow \frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$$

$$= \frac{\frac{1}{n} \{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})\}}{\sqrt{\frac{1}{n} \{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\}} \sqrt{\frac{1}{n} \{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\}}}$$

$$= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} \{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\}}}$$

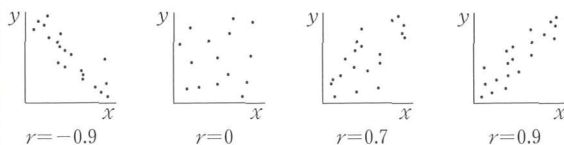
2 相関係数の値の範囲

相関係数 r について, $-1 \leq r \leq 1$ である。

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。

相関がないとき, r は 0 に近い値をとる。

x と y の相関係数 r と散布図



練習 13

教 p.188

教科書 185 ページのデータについて, x の標準偏差は 4.40, y の標準偏差は 4.71, x と y の共分散は 18.22 である。これらの数値を用いて, x と y の相関係数を計算せよ。ただし, 小数第 3 位を四捨五入せよ。

指針 相関係数 (相関係数) = $\frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$ を利用。

解答 求める相関係数は

$$\frac{18.22}{4.40 \times 4.71} = 0.879 \cdots \doteq 0.88 \quad \text{答}$$

練習
14

教 p.189

下の表は、6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。

	①	②	③	④	⑤	⑥
テストA	5	7	5	4	3	6
テストB	4	1	3	5	9	2

(単位は点)

指針 相関係数 テストAの変量を x 、平均値を $\bar{x}$ 、テストBの変量を y 、平均値を $\bar{y}$ とすると、相関係数 r は、次の式から求められる。

$$r = \frac{(x - \bar{x})(y - \bar{y}) \text{ の総和}}{\sqrt{(x - \bar{x})^2 \text{ の総和}} \times \sqrt{(y - \bar{y})^2 \text{ の総和}}}$$

解答 テストAの変量を x 、テストBの変量を y で表すと

$$\bar{x} = \frac{30}{6} = 5, \quad \bar{y} = \frac{24}{6} = 4$$

ゆえに、次の表のようにまとめられる。

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
①	5	4	0	0	0	0	0
②	7	1	2	-3	-6	4	9
③	5	3	0	-1	0	0	1
④	4	5	-1	1	-1	1	1
⑤	3	9	-2	5	-10	4	25
⑥	6	2	1	-2	-2	1	4
計	30	24			-19	10	40

よって、 x と y の相関係数 r は

$$r = \frac{-19}{\sqrt{10} \times \sqrt{40}} = -\frac{19}{20} = -0.95 \quad \text{答}$$

また、相関係数は負で -1 に近いから x と y の間には強い負の相関があると考えられる。 答

D 相関関係と因果関係

まとめ

相関関係と因果関係

相関関係とは、一方が増加するとき、他方が増加もしくは減少する傾向が認められるという2つの量の関係であり、因果関係とは、原因とそれによって生ずる結果との関係をいい、一般に、2つの変量の間に相関関係があるといって、必ずしも因果関係があるとはいえない。

E 質的データをとる2つの変量の間の関係

まとめ

1 量的データ、質的データ

身長や体重のように数値として得られるデータを量的データという。

また、A組、B組とか都道府県別に調べたデータのように、数値としてではなく分類として測定できるデータを質的データという。

2 分割表

複数の分類の質的データによって測定値を集計する表を分割表、あるいはクロス集計表という。

5章

データの分析

練習
15

教 p.191

教科書の表3をもとに、次の問いに答えよ。

- 表4の空らん適切な数を入れよ。
- 教材Bを使用した者、使用していない者のそれぞれにおいて、合格者、不合格者の占める割合を計算して教科書の表2のようにまとめよ。

表4	合	否	計
B：有	37		
B：無			
計			100

指針 分割表(クロス集計表)

- 表4の37という数が表3のどの数と関係しているかをまず調べる。
- 合格者、不合格者それぞれにおける割合であることに注意する。問題文に断りがなければ、教科書の表2の数値にならって、%の小数点以下を四捨五入した整数値を答える。

解答 (1) $37=6+31$ であるから

B：有で不合格者は $1+9=10$

以下同様に調べると、右の表のようになる。 答

- 教材Bを使用していて合格した者は、
(1)の表より37人であるから、使用していた47人において、37人は

	合	否	計
B：有	37	10	47
B：無	14	39	53
計	51	49	100

$$\frac{37}{47} \times 100 = 78.7\ldots \div 79 (\%)$$

よって、教材Bを使用していて不合格になった者は

$$100 - 79 = 21 (\%)$$

また、教材Bを使用しないで合格した者は、(1)の表より14人であるから、使用していなかった53人において、14人は

$$\frac{14}{53} \times 100 = 26.4\ldots \div 26 (\%)$$

よって、教材Bを使用せずに不合格になった者は

$$100 - 26 = 74 (\%)$$

以上から、右の表のようになる。 答

	合	否
B : 有	79 %	21 %
B : 無	26 %	74 %

深める

教 p.191

教科書の表1～表4(練習15で求めた表)で得られたデータから、教材A、Bのどちらの方が、この試験の合否により影響を及ぼしていると予想できるだろうか。

指針 教材を使用して実力が落ちるということは考えにくいから、教材を使用して合格した割合が大きいのはどちらであるかを考える。

解答 使用している受験者の合格する割合は、Aの方が64%、Bの方が79%で、Bの方が多い。よって、Bの方がより影響を及ぼしていると予想できる。 答

コラム 回帰分析

練習

教 p.193

教科書のコラムの図(下の図)のように回帰直線を引くことで、考えられるようになる事柄を説明してみよう。

解答 回帰直線を引くことによって、たとえば、175 cm以上の身長の子の体重は、ほぼ身長に比例していることがわかる。 答

6 仮説検定の考え方

A 仮説検定の考え方

まとめ

仮説検定

ある集団に対して何らかの調査を行う場合、調べたい集団の全体のデータを集めることは困難であることが多く、実際には調べたい集団から一部を抜き出して、そのデータから集団全体の状況を推測することがある。その推測が正しいかどうかを判断する1つの考え方として、ある集団に関してある主張を立て、その主張に反する主張を仮定し、その仮定が正しいかどうかを判断する。このような手法を **仮説検定** という。

練習 16

教 p.196

ある地域の水道局が、水道水の品質改良に取り組んでいる。無作為に選んだ地域の住民 20 人に以前に比べて水道水がおいしくなったと感じるか回答してもらったところ、15 人が以前よりおいしくなったと回答した。この回答のデータから、以前に比べて水道水がおいしくなったと住民から評価されていると判断してよいか。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 回投げて表の出た回数を記録する実験を 200 セット行ったら、次の表のようになったとし、この結果を用いよ。

表の回数	5	6	7	8	9	10	11	12	13	14	15	16	17	計
度数	1	7	23	20	30	35	35	18	20	7	3	0	1	200

5 章 データの分析

指針 仮説検定 仮説検定の考え方で考察する。まず、「水道水の品質は以前と変わっていない」という仮定を立てる。この仮定が否定できるかできないかを、コイン投げの表から判断する。「15 回表が出る相対度数」ではなく、「15 回以上表が出る相対度数」に着目する。

解答 「水道水の品質が以前と変わっていない」という仮定を立てる。

ここで、コイン投げの実験結果を利用すると、コインを 20 回投げて 15 回以上表が出る場合の相対度数は $\frac{3+0+1}{200} = \frac{4}{200} = 0.02$

これは、コインを 20 回投げて 15 回以上表が出る確率が 0.02 と考えてよいから、この確率は基準の確率 0.05 より小さい。

すなわち「水道水の品質が以前と変わっていない」という仮定のもとで、15人以上が「水道水はおいしくなった」という確率の小さいことが起きたのであるから「水道水の品質は以前と変わっていない」という仮定が正しくなかったと考えられる。

よって「水道水の品質は以前と変わっていない」という仮定は否定される。したがって、水道水は以前に比べておいしくなったと評価されていると判断してよい。 

発展 仮説検定と反復試行の確率

まとめ

1 試行, 事象

その結果が偶然によって決まる実験や観測を **試行** といい、試行の結果として起こる事柄を **事象** という。

2 反復試行の確率

1回の試行で事象Aの起こる確率を p とする。この試行を n 回繰り返し行うとき、事象Aがちょうど r 回起こる確率は

$${}_nC_r p^r (1-p)^{n-r} \quad \text{ただし } 0 \leq r \leq n$$

練習
1

教 p.197

1枚のコインを6回投げたところ、表が5回出た。このコインは表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。

解答 「このコインの表の出る確率は0.5である」という仮定を立てる。このとき、1枚のコインを6回投げて、5回以上表が出る確率は

$${}_6C_1 \left(\frac{1}{2}\right)^5 \cdot \frac{1}{2} + \left(\frac{1}{2}\right)^6 = \frac{6}{64} + \frac{3}{64} = 0.09375$$

これは0.05より大きいから、「このコインの表の出る確率は0.5である」という仮定のもとで、基準の確率0.05より大きい確率で「このコインは表が出やすい」という事象が起きたのであるから、「このコインの表の出る確率は0.5である」という仮定は否定できない。

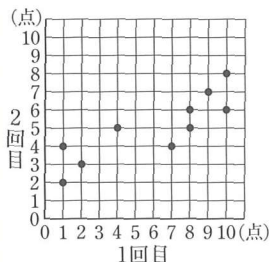
したがって、このコインは表が出やすいとは判断できない。 

第5章

問題

教 p.198

- 1 右の図は、10人の生徒に漢字テストを2回行い、得点のデータを散布図にしたものである。1回目のデータを横軸に、2回目のデータを縦軸にとっている。次の値を求めよ。なお、得点は整数である。



- (1) 1回目のデータの平均値
- (2) 1回目のデータの中央値
- (3) 1回目のデータと2回目のデータの相関係数

指針 平均値, 中央値, 相関係数 (1)~(3) 表にして計算するとよい。

(2) それぞれのデータの大きさが偶数であるから、中央2つの平均をとる。

解答 1回目のデータを x , 2回目のデータを y とすると

$$\bar{x} = \frac{60}{10} = 6 \text{ (点)}, \quad \bar{y} = \frac{50}{10} = 5 \text{ (点)}$$

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
①	1	2	-5	-3	15	25	9
②	1	4	-5	-1	5	25	1
③	2	3	-4	-2	8	16	4
④	4	5	-2	0	0	4	0
⑤	7	4	1	-1	-1	1	1
⑥	8	5	2	0	0	4	0
⑦	8	6	2	1	2	4	1
⑧	9	7	3	2	6	9	4
⑨	10	6	4	1	4	16	1
⑩	10	8	4	3	12	16	9
計	60	50			51	120	30

(1) 平均値 6点 答 (2) 中央値 $\frac{7+8}{2} = 7.5$ (点) 答

(3) 相関係数は $\frac{51}{\sqrt{120} \times \sqrt{30}} = 0.85$ 答

教 p.198

- 2 5個の値 2, 3, a , 8, 12 からなるデータの平均値が 6 であるとき, a の値を求めよ。また, このデータの分散を求めよ。

5章

データの分析

指針 平均値, 分散 変量 x のデータの値が $x_1, x_2, \dots, x_n$ で, その平均値が $\bar{x}$ のときの分散は

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

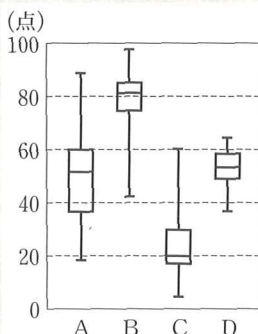
解答 5 個の値の平均値が 6 であるから $\frac{1}{5}(2+3+a+8+12)=6$
 $25+a=30$ から $a=5$ 答

$$\begin{aligned} \text{分散は} & \frac{1}{5} \{(2-6)^2 + (3-6)^2 + (5-6)^2 + (8-6)^2 + (12-6)^2\} \\ & = \frac{1}{5} (16+9+1+4+36) = \frac{1}{5} \times 66 = 13.2 \quad \text{答} \end{aligned}$$

教 p.198

3 右の図は, ある学校で行った 4 種類のテスト A, B, C, D についての, 300 人の得点の箱ひげ図である。なお, どのテストも 100 点満点である。

- (1) 80 点以上の生徒が半数以上いたのはどのテストか。
- (2) 40 点以下の生徒がいて, その人数が 75 人未満である可能性があるのはどのテストか。



指針 箱ひげ図によるデータの分析 箱ひげ図では, データの分布の大まかな様子がわかる。箱の上端は Q_3 , 下端は Q_1 , 箱の中の線は Q_2 を表している。

解答 (1) B のテストの中央値が 80 点より大きいから, B のテストでは, 半数以上の生徒が 80 点以上であったと考えられる。

また, テスト A では, Q_3 が 60 点であるから, 80 点以上の生徒は 25 % 以下, テスト C, D では 80 点以上の生徒はいない。

よって, 条件を満たすテストは B 答

- (2) 40 点以下の生徒がいたテストは A, C, D

このうち, A は箱の下端, すなわち第 1 四分位数が 40 点より小さいから,

生徒の人数が $\frac{1}{4}$ 以上, すなわち, 75 人以上の生徒が 40 点以下であると考えられる。

また, C については, Q_1, Q_2, Q_3 が 40 点より小さいから, これも 75 人以上の生徒が 40 点以下であると考えられる。

D は第 1 四分位数が 40 点より大きいから, 40 点以下の生徒は 75 人未満であると考えられる。

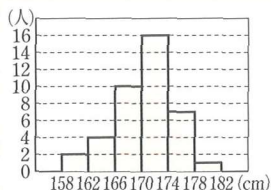
よって, 条件を満たすテストは D 答

第5章 章末問題 A

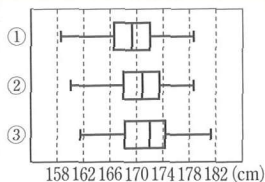
教 p.199

1. 下の図 [1] は、40 人の生徒の身長データをヒストグラムにしたものである。ただし、各階級は 158 cm 以上 162 cm 未満のように区切っている。
- (1) 170 cm 以上の人は何人いるか。
- (2) このデータの箱ひげ図として矛盾するものを、下の図 [2] の ①～③ からすべて選べ。

[1]



[2]



指針 ヒストグラムと箱ひげ図

- (1) ヒストグラムから、170 cm 以上の階級の度数を足し合わせる。
- (2) ヒストグラムから、データの最小値、第 1 四分位数、中央値、第 3 四分位数、最大値が入る階級を求めて箱ひげ図と比較する。

解答 (1) 170 cm 以上 174 cm 未満、174 cm 以上 178 cm 未満、178 cm 以上 182 cm 未満の各階級の度数を足し合わせて

$$16 + 7 + 1 = 24 \text{ (人)} \quad \text{答}$$

- (2) ヒストグラム [1] から、データの最小値、第 1 四分位数、中央値、第 3 四分位数、最大値が入る階級は次のようになる。

最小値	158 cm 以上 162 cm 未満	……	(a)
第 1 四分位数	166 cm 以上 170 cm 未満	……	(b)
中央値	170 cm 以上 174 cm 未満	……	(c)
第 3 四分位数	170 cm 以上 174 cm 未満	……	(d)
最大値	178 cm 以上 182 cm 未満	……	(e)

箱ひげ図 ① は (c) に矛盾する。

箱ひげ図 ③ は (d) に矛盾する。

箱ひげ図 ② は (a)～(e) のどれにも矛盾しない。

よって、矛盾する箱ひげ図は ①, ③ **答**

5 章

データの分析

2. 生徒7人のテストの得点が次のようになった。

57, 46, 37, 38, 54, 48, 35 (点)

- (1) このデータの平均値, 中央値, 分散を求めよ。
 (2) 生徒全員の得点に10点ずつ足したときの平均値および分散を求めよ。

指針 平均値, 中央値, 分散

- (1) データの大きさは奇数であるから, 中央値は, データを大きさの順に並べたとき, 中央の位置にくる値をとる。

分散は, データにおいて, 偏差の2乗の平均値を計算する。

解答 データを値の大きさの順に並べると, 次のようになる。

35, 37, 38, 46, 48, 54, 57

- (1) 平均値は

$$\begin{aligned} & \frac{1}{7}(35+37+38+46+48+54+57) \\ &= \frac{1}{7} \times 315 = 45 \text{ (点)} \quad \text{答} \end{aligned}$$

中央値は 46 (点) 答

データの値と偏差の2乗の値は, 次の表のようになる。

x	35	37	38	46	48	54	57
$x - \bar{x}$	-10	-8	-7	1	3	9	12
$(x - \bar{x})^2$	100	64	49	1	9	81	144

したがって, 分散は

$$\frac{1}{7}(100+64+49+1+9+81+144) = \frac{1}{7} \times 448 = 64 \quad \text{答}$$

- (2) 全員の得点に10点ずつ足したとき, データの値の総和は 10×7 (点) 増加する。

よって, 平均値は

$$\frac{1}{7} \times (315 + 10 \times 7) = \frac{1}{7} \times 385 = 55 \text{ (点)} \quad \text{答}$$

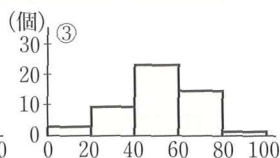
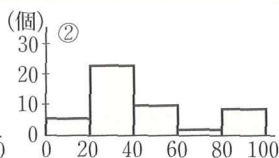
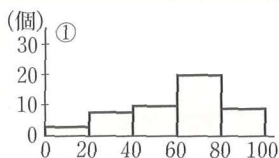
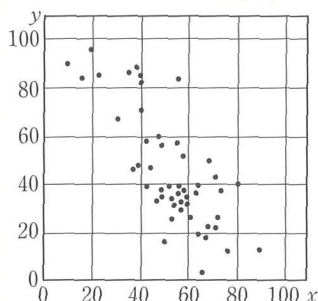
全員の得点に10点ずつ足したとき, データの各値も平均値も10点だけ増加するから, 偏差はもとのデータと変わらない。

したがって, 分散は 64 答

注意 変量 x, y のデータの間に $y = ax + b$ (a, b は定数) の関係があるとき, x, y のデータの分散を s_x^2, s_y^2 とすると $s_y^2 = a^2 s_x^2$ が成り立つ。

3. 右の図は、ある2つの変量 x , y のデータについての散布図である。 x のとる値についてのヒストグラムを下の①～③から選べ。また、 x と y の相関係数は、次のうちのどれに近いか。

0.3, 0.8, -0.8, -1



指針 散布図, 相関係数 散布図から、2つの変量の間の関連性を読みとる。

解答 散布図における点の分布の様子から、 x と y の間に負の相関がある。

また、散布図で、 x は 40～60 の間に集中しているから、ヒストグラムは

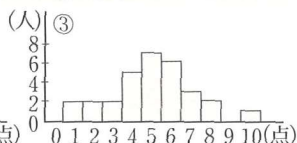
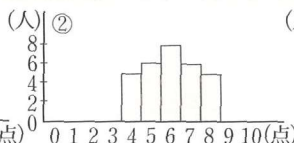
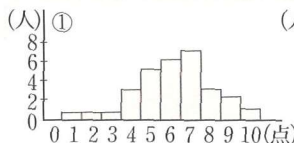
③ **答**

相関係数は -0.8 か -1 のいずれかであるが、散布図で、ある程度の散らばりがあることから、相関係数は -0.8 に近い。 **答**

第5章 章末問題 B

4. 3つのクラス A, B, C で、10点満点の小テストを行った。右の表は、各クラスの成績の結果である。それぞれのクラスの成績を表すヒストグラムを、下の①～③から選べ。

	人数	平均値	標準偏差
A	30	6.0	1.3
B	30	5.0	2.0
C	30	6.0	2.0



指針 データの散らばりの度合いとヒストグラム ヒストグラムが平均値の周りに集中していると、データの散らばりの度合いが小さいと考えられる。

解答 データの値が平均値の周りに集中しているほど、標準偏差は小さくなる。よって、標準偏差の小さいクラスAのヒストグラムは、平均値 6.0 の周りに集中しているから ②

クラスBとクラスCは、平均値の違いから、Bは③、Cは①

答 A ②, B ③, C ①

教 p.200

5. 25 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 4, 分散は 14, 残りの 15 個の値の平均値は 9, 分散は 19 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。

指針 分割されたデータの平均値, 分散ともとのデータの平均値, 分散

$$(1) (\text{平均値}) = \frac{(\text{データの値の総和})}{(\text{データの大きさ})}$$

データの値の総和は $4 \times 10 + 9 \times 15$

データの大きさは 25

$$(2) (x \text{ のデータの分散}) = (x^2 \text{ のデータの平均値})$$

$$- (x \text{ のデータの平均値})^2$$

を利用する。

解答 (1) 平均値は $\frac{1}{25}(4 \times 10 + 9 \times 15) = 7$ **答**

(2) 10 個の値の 2 乗の平均値を a とすると

$$a - 4^2 = 14 \quad \text{よって} \quad a = 30$$

残りの 15 個の値の 2 乗の平均値を b とすると

$$b - 9^2 = 19 \quad \text{よって} \quad b = 100$$

よって、25 個の値の 2 乗の和は

$$a \times 10 + b \times 15 = 30 \times 10 + 100 \times 15 = 1800$$

したがって、25 個の値の分散は

$$\frac{1800}{25} - 7^2 = 72 - 49 = 23 \quad \text{答}$$

6. ある種子Aの発芽する確率は $\frac{2}{3}$ である。種子Aを改良した種子Bに

ついて、発芽試験を行ったところ、30個の種子のうち、25個が発芽した。この結果から種子BはAに比べ、発芽する確率が高いかどうかを仮説検定の考え方をを用いて考察したい。

- (1) 「種子Bが発芽する確率は種子Aと変わらない」という仮定のもとで30個中25個以上が発芽する確率を実験を通して考える。このとき、次の実験を用いることが適切でない理由を述べよ。

公正なコインを30枚投げて表が出た枚数の記録を500回行う。

- (2) 公正なさいころを30個投げて3以上の目が出る個数を記録する実験を500回行ったところ下の表のようになった。

3以上の目が出た個数	13	14	15	16	17	18	19	20	21	22	23	24	25	26	計
度数	4	9	13	26	33	55	62	70	81	61	42	27	11	6	500

基準となる確率を0.05とすると、種子Bの発芽する確率はAに比べて高いと判断してよいかをこの実験結果を用いて考察せよ。

指針 仮説検定

- (1) 公正なコインを投げて表が出る確率は、種子Aの発芽する確率とは異なる。
 (2) 3以上の目が25個以上出る相対度数(確率)を求める。

解答 「種子Bの発芽する確率は種子Aと変わらない」と仮定する。

- (1) この仮説検定に必要となる実験は、確率が $\frac{2}{3}$ であるような実験である。

よって、公正なコインを1枚投げて表が出る確率は $\frac{1}{2}$ であり、確率 $\frac{2}{3}$ と異なるから、この仮説検定のための実験にはならない。 **答**

- (2) さいころ投げの実験結果を利用すると、25個以上のさいころで3以上の目が出る場合の相対度数は $\frac{11+6}{500} = \frac{17}{500} = 0.034$

これは0.05より小さいから、「種子Bの発芽する確率は種子Aと変わらない」という仮定は否定される。

よって、種子Bの発芽する確率はAに比べて高いと判断してよい。 **答**

総合問題

1. ※ 問題文は、教科書 201 ページを参照

指針 まず、何を x とするかであるが、ここでは当選に必要な最低得票数を x 票とする。このとき、(すべての投票数) $-4x$ は、落選した立候補者全員の得票数であるから、もしもこの票すべてが落選した立候補者 1 人の得票数としたとき、この得票数と x との間にどのような関係が成り立つかを考える。

解答 当選に必要な、1 人の立候補者の最低得票数を x 票とする。

投票率が 50 % であるから、投票数は $100000 \times \frac{50}{100} = 50000$ (票)

ここで、当選した立候補者全員がこの最低得票数で当選したと仮定する。

このとき、残りの $(50000 - 4x)$ 票がすべて落選した立候補者のうちの 1 人だけに投票されていたとすると、この $(50000 - 4x)$ 票は x 票より少ないから

$$50000 - 4x < x$$

これを解くと $x > 10000$

よって、開票率 40 % の時点での開票数は $50000 \times 0.4 = 20000$ (票) であり、そのうち立候補者 A の得票数は 10001 票以上であったと考えられる。

答 10001 票以上

2. ※ 問題文は、教科書 201 ページを参照

指針 (1) (i) $x \in X$ ならば $x \in Y$ が成り立つ $\Leftrightarrow X \subset Y$,
 $X \subset Y \Leftrightarrow \overline{X} \supset \overline{Y}$ を利用。

(ii), (iv) (エ) から $C \cap E \neq \emptyset$ であることに着目。

(2) ① は (1) の (i) から。② は条件 (イ), (ウ) から。③~⑤ は条件 (ウ) から。

解答 (1) (i) $x \in D$ とすると条件 (ウ) より $x \in B$ である。すなわち $D \subset B$
 また、条件 (ア) より、 $\overline{A} \subset \overline{B}$ であるから $B \subset A$
 ゆえに、 $D \subset A$ が成り立つから、 $x \in D$ ならば $x \in A$ は成り立つ。
 よって、(i) は正しい。

(ii) 条件 (エ) から $C \cap E \neq \emptyset$ であるから、 $x \in C \cap E$ とすると、条件 (イ) より
 $x \in B$

よって、(ii) は正しくない。

(iii) (ア)~(エ) の 4 つの条件からは判断できない。

(iv) 条件 (エ) から $C \cap E \neq \emptyset$ であるから、 $x \in C \cap E$ とすると、 x は E の要素である。このとき、条件 (イ) より $x \in B$
 条件 (ウ) より、 $D \subset B$ であるから $x \in D$
 よって、 $x \in D$ を満たすような E の要素 x があるから、(iv) は正しい。