数I

データの分析

・ 度数分布表 … データのちらばりをまとめた表

階級 … 度数分布において区切られた区間

階級の幅 … 区間の幅

度数 … 各階級に含まれる値の個数

階級値 … 各階級のまん中の値

次のデータは，東京の 2018 年 4 月の日ごとの最高気温である。

21.9	24.5	23.4	26.2	15.3	22.4	21.8	16.8	19.9	19.1
21.9	25.9	20.9	18.8	22.1	20.0	15.0	16.0	22.2	26.4
26.0	28.3	18.7	21.3	22.5	25.0	22.0	26.1	25.6	25.7

（気象庁ホームページより作成，単位は °C）

最高気温の度数分布表

階級（°C）	度数（日）
15 以上 18 未満	4
18 〜 21	6
21 〜 24	10
24 〜 27	9
27 〜 30	1
計	30

練習 1 前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。

(1) 度数が 9 の階級は 24°C 以上 27°C 未満なので

$$\frac{24+27}{2} = 25.5$$

25.5 °C

(2) 最高気温が 21°C 以上の日は

$10+9+1=20$

20日

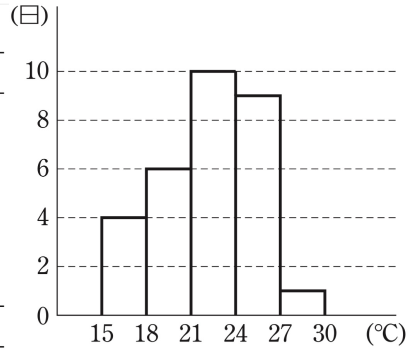
(3) 階級の低い順から度数を数え, 12に達するのは

21°C 以上 24°C 未満

・ ヒストグラム … 度数分布表を柱状に表したもの

最高気温の度数分布表

階級 (°C)	度数 (日)
15 以上 18 未満	4
18 ～ 21	6
21 ～ 24	10
24 ～ 27	9
27 ～ 30	1
計	30



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)

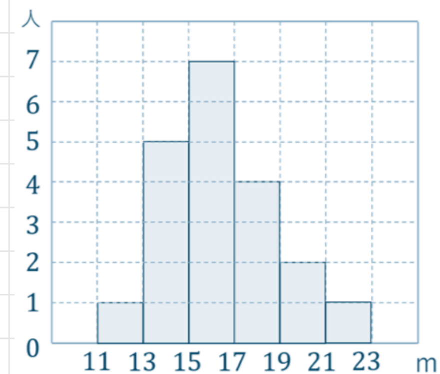
(1) 11 m 以上 13 m 未満を階級の 1 つとして，どの階級の幅も 2 m である度数分布表を作れ。

(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。

(1)

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

(2)



平均値

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

$$\overline{x}=\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)$$

練習 3 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）

$$\frac{400+410+420+390+430+450+440}{7}=420$$

420分

- 最頻値（モード）… データにおいて最も個数が多い値

度数分布表においては度数の最も大きい階級の階級値を用いる.

練習 4 168 ページの度数分布表において，最頻値を求めよ。

最高気温の度数分布表

階級（°C）	度数（日）
15 以上 18 未満	4
18 ～ 21	6
21 ～ 24	10
24 ～ 27	9
27 ～ 30	1
計	30

左の度数分布表において度数の最も大きい値の階級は

21°C 以上 24°C 未満 であるので

求める最頻値はこの階級の階級値

$$\frac{21+24}{2} = 22.5$$

22.5°C

・ 中央値(メジアン)… データを値の大きさ順に並べたとき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2つの値の平均値とする.

練習 5 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)

データを値の大きさ順に並べると

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

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

↳ この平均値が中央値

$$\frac{43+41}{2} = 42$$

42 kg

・ 範囲…データのちらばりの度合いを表す値

範囲が大きいほどデータのちらばりの度合いが大きい

例 4 データの範囲

次のデータは，ある年のA市における月ごとの降水日数である。

7，4，9，7，10，13，14，7，4，12，13，5 （日）

このデータの範囲は　　14－4＝10 （日）　終

練習 6 次のデータは，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）

このデータの範囲は

26－6＝16 (日)

A市の範囲が10日に対してB市の範囲は16日なのでB市の方がデータのちらばりの度合いが大きい

- 四分位数 … データを大きさの順に並べたとき，データを4等分する位置にくる値

小さい方から第1四分位数 Q_1，第2四分位数 Q_2，第3四分位数 Q_3

Q_1 … 下位のデータの中央値

Q_2 … データ全体の中央値

Q_3 … 上位のデータの中央値

データの大きさが 9 のとき

$\boxed{2 \quad 6 \quad 8 \quad 9}\quad 13 \quad \boxed{16 \quad 19 \quad 21 \quad 29}$

$Q_1=7$　$Q_2=13$　$Q_3=20$

データの大きさが 8 のとき

$\boxed{2 \quad 6 \quad 8 \quad 9}\quad \boxed{13 \quad 16 \quad 19 \quad 21}$

$Q_1=7$　$Q_2=11$　$Q_3=17.5$

- 四分位範囲 … （第3四分位数 Q_3 − 第1四分位数 Q_1）の値

- 四分位偏差 … 四分位範囲を2で割った値

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

Q_1（32）　Q_2（40）　Q_3（55）

B　25，31，39，42，45，46，50，53，54，65，80

Q_1（39）　Q_2（46）　Q_3（54）

$Q_3 - Q_1 = 55 - 32 = 23$

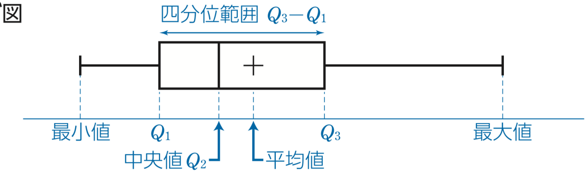
$Q_3 - Q_1 = 54 - 39 = 15$

Aの方が四分位範囲が大きいのでデータのちらばりの度合いが大きいのはA

・ 箱ひげ図 … データの分布を見るための図

最小値, 第1四分位数, 第2四分位数(中央値), 第3四分位数, 最大値を表現した図

箱ひげ図



箱ひげ図は複数のデータの分布を比較するときに便利な図である。

練習 8 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

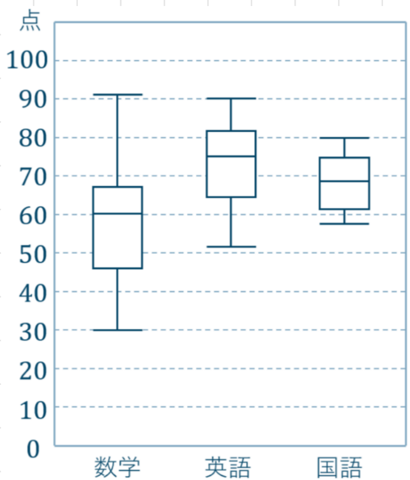
英語　75，65，90，78，52，88，70，75，59，82

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。

(1)



(2)

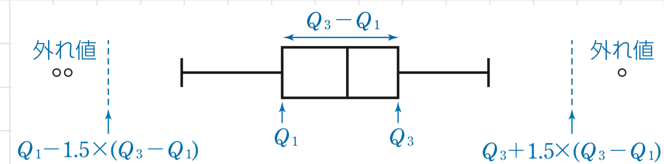
数学

・ 外れ値…データの中の他の値から極端にはなれた値

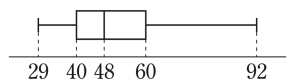
(第1四分位数－1.5×四分位範囲)以下の値

(第3四分位数＋1.5×四分位範囲)以上の値

箱ひげ図への図示



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。



四分位範囲は $60-40=20$

29が $Q_1-1.5\times20$ 以下であれば29は外れ値である.

$40-1.5\times20=10$ なので29は外れ値ではない

92が $Q_3+1.5\times20$ 以上であれば29は外れ値である.

$60+1.5\times20=90$ なので92は外れ値である.

よって

最大値92は外れ値

最小値29は外れ値でない

・ 偏差 … 変量 x のデータの値を $x_1, x_2, \cdots, x_n$, 平均値を $\bar{x}$ とするときの $x_1-\bar{x}, x_2-\bar{x}, \cdots, x_n-\bar{x}$ のこと

・ 分散 … 偏差の2乗の平均値, s^2 で表す.

・ 標準偏差 … 分散の正の平方根, s で表す.

分散と標準偏差

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

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

標準偏差 $s=\sqrt{分散}$

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）

合計点は 70点

平均点は 7点

分散 $$S^2=\frac{1}{10}\{(6-7)^2+(10-7)^2+(7-7)^2+(7-7)^2+(5-7)^2+(4-7)^2+(9-7)^2+(10-7)^2+(5-7)^2+(7-7)^2\}=\frac{1}{10}\times 40=4$$

標準偏差 $s=\sqrt{4}=2$

分散 4

標準偏差 2

分散の求め方(2)

$$(x\text{のデータの分散}) = (x^2\text{のデータの平均値}) - (x\text{のデータの平均値})^2 \quad \cdots\cdots ①$$

練習 11 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

合計 55　平均 5.5

x^2のデータ　25, 9, 36, 64, 25, 64, 25, 16, 36, 25　合計 325　平均 32.5

分散 $s^2 = (x^2\text{のデータの平均値}) - (x\text{のデータの平均値})^2 = 32.5 - (5.5)^2 = 32.5 - 30.25 = 2.25$

標準偏差 $s = \sqrt{2.25} = 1.5$

分散　2.25

標準偏差　1.5

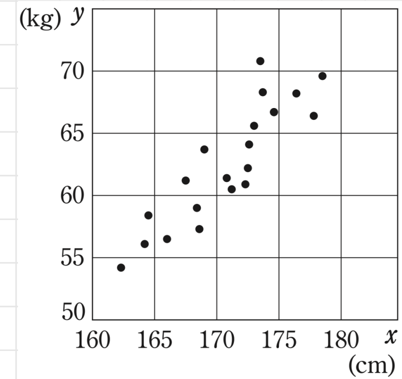
・**散布図** … 2つの変量からなるデータを点として平面上に図示したもの

（x の単位は cm，y の単位は kg）

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
x	168.4	164.5	171.2	173.0	162.3	170.8	172.5	164.2	169.0	168.6
y	59.0	58.4	60.5	65.6	54.2	61.4	62.2	56.1	63.7	57.3

	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳
x	172.6	166.0	173.7	176.4	178.5	167.5	177.8	174.6	172.3	173.5
y	64.1	56.5	68.3	68.2	69.6	61.2	66.4	66.7	60.9	70.8

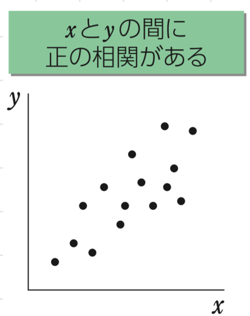
散布図にすると ⟶

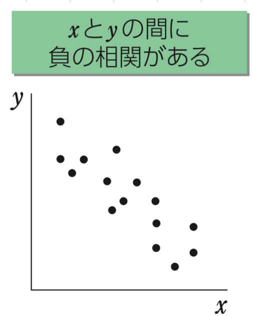


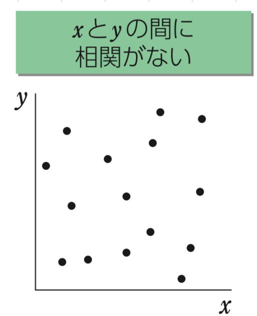
・**正の相関** … 2つの変量からなるデータにおいて，一方が増えると他方も増える傾向

・**負の相関** … 2つの変量からなるデータにおいて，一方が増えると他方が減る傾向

（どちらの傾向も見られなければ相関はない）







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(°C) を調べた結果である。

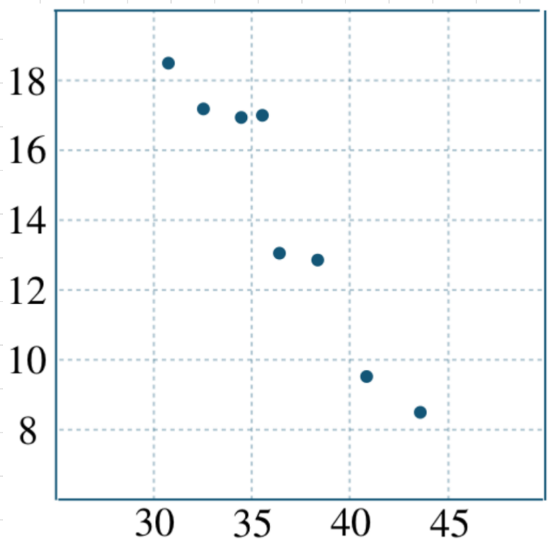
地点	札幌	青森	仙台	東京	長野	大阪	高知	鹿児島
x	43.1	40.8	38.3	35.7	36.7	34.7	33.6	31.6
y	8.2	9.6	12.5	17.0	13.1	16.9	17.3	18.5

(気象庁ホームページより作成)

(1) 2 つの変量 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。

(1)



(2) 負の相関がある.

・ 相関係数… x と y の共分散を x の標準偏差と y の標準偏差の積で割ったもの. r で表す.

$$\boxed{r = \frac{x\text{と}y\text{の共分散}}{(x\text{の標準偏差}) \times (y\text{の標準偏差})}}$$

ここで r は $-1 \leqq r \leqq 1$ であり, 正の相関が強いほど r は 1 に近づく

負の相関が強いほど r は -1 に近づく

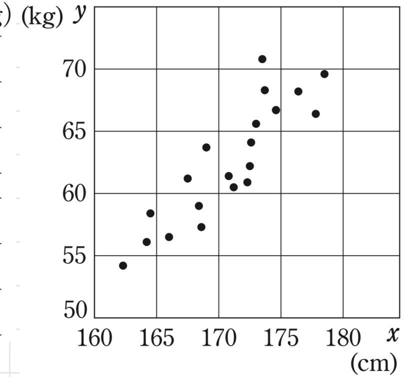
相関がないときは r は 0 に近い値となる.

・ 共分散 … x と y の偏差の積の平均 → $\frac{1}{n}\left\{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\right\}$

（x の単位は cm，y の単位は kg）

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
x	168.4	164.5	171.2	173.0	162.3	170.8	172.5	164.2	169.0	168.6
y	59.0	58.4	60.5	65.6	54.2	61.4	62.2	56.1	63.7	57.3

	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳
x	172.6	166.0	173.7	176.4	178.5	167.5	177.8	174.6	172.3	173.5
y	64.1	56.5	68.3	68.2	69.6	61.2	66.4	66.7	60.9	70.8



練習 13 185 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。

$$r = \frac{x\text{と}y\text{の共分散}}{(x\text{の標準偏差}) \times (y\text{の標準偏差})} = \frac{18.22}{4.40 \times 4.71} = 0.88$$

$\underline{0.88}$

練習 14 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)

テストAのデータを x，テストBのデータを y とする.

$\bar{x} = \frac{1}{6} \times (5+7+5+4+3+6) = 5$, $\bar{y} = \frac{1}{6} \times (4+1+3+5+9+2) = 4$

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

相関係数 r は

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

$r = -0.95$ より強い負の相関がある.