第5章 データの分析

1 データの散らばり

例題 1　データの整理

右のデータは，男子生徒25人の握力検査の結果である．30kg以上35kg未満を階級の1つとし，どの階級の幅も5kgとして，次の問いに答えよ．

44	53	34	42	39
48	37	40	51	47
36	41	54	43	46
52	44	49	31	35
42	36	56	45	46

(1) 度数分布表を作れ．　(2) ヒストグラムをかけ．

(3) 50kg以上55kg未満の階級の相対度数を求めよ．

解 (1) 階級を30〜35，35〜40，… に分け，各階級に属するデータの個数(度数)を整理する．

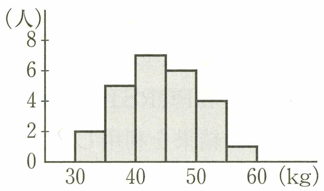
(2) 階級の幅を底辺，度数を高さとする長方形をかく．

(3) $\dfrac{\text{階級の度数}}{\text{度数の合計}}=\dfrac{4}{25}=\mathbf{0.16}$

(1)

階級(kg)	度数(人)
以上　未満 30 〜 35	2
35 〜 40	5
40 〜 45	7
45 〜 50	6
50 〜 55	4
55 〜 60	1
計	25

(2)



1 右のデータは，女子生徒20人の握力検査の結果である．16kg以上20kg未満を階級の1つとし，どの階級の幅も4kgとして，次の問いに答えよ．

30	23	29	33	25
29	21	34	28	24
37	25	20	26	35
27	18	31	34	27

(1) 度数分布表を作れ．

(2) ヒストグラムをかけ．

(3) 20kg以上24kg未満の階級の相対度数を求めよ．

2 右の表は，野球部員のハンドボール投げの記録を，度数分布表と相対度数の分布表にまとめたものである．次の問いに答えよ．

(1) 度数の合計を求めよ．

(2) 表中の**ア**〜**オ**にあてはまる数を求めよ．

階級(m)	度数(人)	相対度数
以上　未満 15 〜 20	2	0.08
20 〜 25	5	**ア**
25 〜 30	**イ**	0.32
30 〜 35	**ウ**	**エ**
35 〜 40	3	**オ**
計		1.00

●ポイント

① ある集団に属する人や物の性質を数で表した量(試験の得点や身長など)を**変量**といい，調査や実験などで得られた変量の観測値や測定値の集まりを**データ**という．

② 変量の値の範囲をいくつかの区間に区切ったとき，その各区間を**階級**，各区間の幅を**階級の幅**，階級の中央の値を**階級値**という．

③ 各階級に属するデータの個数を**度数**といい，各階級に度数を対応させた表を**度数分布表**という．

④ 階級の幅を底辺，度数を高さにとった柱状のグラフを**ヒストグラム**という．

⑤ 各階級の度数を度数の合計で割った値を**相対度数**という．

例題 2　平均値

次のデータ1は，高校生20人の1日の読書時間(分)を記したものであり，右のデータ2は，それを度数分布表にまとめたものである．次の問いに答えよ．ただし，答えは四捨五入して，小数第1位まで求めよ．

(データ1)

50,　60,　35,　90,　25,　50,　70,　95,　65,　15,

85,　110,　60,　30,　55,　10,　50,　75,　40,　45

(データ2)

階級(分)	度数(人)
以上　未満 0 〜 20	2
20 〜 40	3
40 〜 60	6
60 〜 80	5
80 〜 100	3
100 〜 120	1
計	20

(1) **データ1**から，読書時間の平均値$\overline{x}$を求めよ．

(2) **データ2**の度数分布表から，読書時間の平均値$\overline{x}$を求めよ．

解 (1) 上の20個の数値の総和をXとすると，$X=1115$ より，$\overline{x}=\dfrac{X}{n}=\dfrac{1115}{20}\fallingdotseq\mathbf{55.8}$ **(分)**

(2) 度数分布表の各階級の階級値は，10，30，50，70，90，110だから，

$$\overline{x}=\frac{1}{20}(10\cdot2+30\cdot3+50\cdot6+70\cdot5+90\cdot3+110\cdot1)=\frac{1}{20}\cdot1140=\mathbf{57.0}\ \text{(分)}$$

〔注〕 データから直接計算した平均値と，度数分布表による平均値は異なる場合が多いが，その差はそれほど大きくはならない．

3 右の**データ1**は，20個の卵の重さ(g)を調べた結果であり，**データ2**は，それを度数分布表にまとめたものである．次の問いに答えよ．ただし，答えは四捨五入して，小数第1位まで求めよ．

(1) **データ1**から，卵の重さの平均値$\overline{x}$を求めよ．

(2) **データ2**の度数分布表から，卵の重さの平均値$\overline{x}$を求めよ．

(データ1)

62	56	53	60
59	54	68	66
61	57	60	62
65	56	58	63
55	62	58	61

(データ2)

階級(g)	度数(個)
以上　未満 52 〜 55	2
55 〜 58	4
58 〜 61	5
61 〜 64	6
64 〜 67	2
67 〜 70	1
計	20

4 右のデータは，小テスト(5点満点)を受けた40人の生徒の得点を表にまとめたものである．平均値が3.2点のとき，表のx，yの値を求めよ．

得点(点)	0	1	2	3	4	5	計
人数(人)	0	2	x	y	9	6	40

●ポイント

① 変量xのn個のデータの値x_1，x_2，……，x_nの総和をnで割った値を**平均値**といい，$\overline{x}$で表す．

② **平均値(データから直接求める場合)**　$\overline{x}=\dfrac{1}{n}(x_1+x_2+\cdots\cdots+x_n)$

③ **度数分布表による平均値**　x_1，x_2，……，x_mを階級値，f_1，f_2，……，f_mを度数として，

$$\overline{x}=\frac{1}{n}(x_1f_1+x_2f_2+\cdots\cdots+x_mf_m)$$

④ データ全体の特徴を適当な1つの数値で表すとき，この値をデータの**代表値**という．よく用いられる代表値として，平均値，中央値，最頻値がある．

〔注〕 分布が偏っている場合は，平均値を代表値とすることは必ずしも適切とはいえない．

例題 3　中央値

次のデータは，ある都市における16日間の最高気温(℃)である．これから，中央値を求めよ．

25,　24,　26,　27,　27,　28,　28,　29,　30,　29,　29,　28,　28,　26,　25,　25

解　データを小さい順に並べると，中央にある2つの値は27，28より，$\dfrac{27+28}{2}=\mathbf{27.5}$ **(℃)**

5　次のデータは，あるクラスの生徒15人の家族の人数である．中央値を求めよ．

5,　2,　4,　3,　5,　4,　7,　4,　3,　4,　6,　4,　5,　4,　3

6　次のデータは，10人の生徒の身長(cm)である．中央値を求めよ．

158.9,　149.6,　159.7,　153.2,　164.7,　155.9,　161.5,　150.4,　154.5,　155.3

例題 4　最頻値

右の度数分布表から，最頻値(モード)を求めよ．

階級値	30	40	50	60	70	80	90	計
度数	1	3	3	7	8	5	3	30

解　最大の度数をもつ階級の階級値だから，**70**

7　次の問いに答えよ．

(1)　上の**5**のデータについて，最頻値を求めよ．

(2)　上の**6**のデータについて，147 cm以上150 cm未満を階級の1つとし，どの階級の幅も3 cmとして度数分布表を作り，その表から最頻値を求めよ．

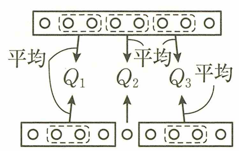
例題 5　四分位数

次のデータの第1四分位数Q_1，第2四分位数Q_2，第3四分位数Q_3を求めよ．

(1)　21,　23,　24,　24,　25,　27,　29,　30 (g)　　(2)　56,　57,　58,　60,　61,　62,　63,　64,　66 (kg)

解　(1)　$Q_1=\dfrac{23+24}{2}=\mathbf{23.5\,(g)}$，$Q_2=\dfrac{24+25}{2}=\mathbf{24.5\,(g)}$，

$Q_3=\dfrac{27+29}{2}=\mathbf{28\,(g)}$



(2)　$Q_1=\dfrac{57+58}{2}=\mathbf{57.5\,(kg)}$，$Q_2=\mathbf{61\,kg}$，$Q_3=\dfrac{63+64}{2}=\mathbf{63.5\,(kg)}$

●**ポイント**

①　データ(変量)を大きさの順に並べたとき，中央にある値を**中央値(メジアン)**という．データの個数が偶数の場合は，中央にある2つの値の平均値を中央値とする．

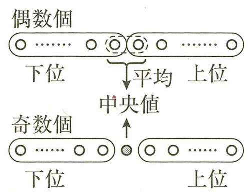
②　データにおいて，最も個数の多い値を**最頻値(モード)**という．度数分布表から求める場合は，最大の度数をもつ階級の階級値を最頻値とする．

③　データの値を大きさの順に並べたとき，4等分する位置にくる値を**四分位数**といい，小さい順に**第1四分位数**，**第2四分位数**，**第3四分位数**という．これらをQ_1，Q_2，Q_3と表すことがある．(本書では以後Q_1，Q_2，Q_3を用いる)　四分位数は次のようにして求める．

(i)　中央値を求める．この値がQ_2となる．

(ii)　中央値を境界に，データを等分する．データの個数が奇数のときは，中央値はどちらのデータにも含めない．

(iii)　小さい方，大きい方それぞれのデータの中央値をQ_1，Q_3とする．



〔注〕　四分位数の定義は他にもいくつかある．

8　次のデータの第1四分位数Q_1，第2四分位数Q_2，第3四分位数Q_3を求めよ．

(1)　23,　25,　27,　30,　31,　33,　35,　36,　38,　41

(2)　67,　71,　72,　72,　76,　78,　80,　81,　83,　84,　87

(3)　23,　31,　19,　22,　36,　28,　18,　25,　21,　29,　35,　26

(4)　34,　43,　37,　48,　33,　41,　44,　38,　53,　51,　31,　40,　48

例題 6　四分位範囲，四分位偏差

次のデータは，10人の生徒の垂直跳びの記録である．範囲，四分位範囲，四分位偏差を求めよ．

48,　50,　54,　57,　58,　60,　62,　63,　64,　69 (cm)

解　範囲は，$69-48=\mathbf{21\,(cm)}$　　また，$Q_1=54$，$Q_3=63$より，

四分位範囲は，$Q_3-Q_1=63-54=\mathbf{9\,(cm)}$　　四分位偏差は，$\dfrac{Q_3-Q_1}{2}=\dfrac{9}{2}=\mathbf{4.5\,(cm)}$

9　次のデータは，あるテスト(100点満点)の得点を男女別に並べたものである．

男子　58,　63,　72,　80,　60,　77,　55,　83,　75,　85,　68,　70 (点)

女子　76,　70,　82,　68,　87,　84,　72,　60,　90,　75,　80,　95 (点)

(1)　男子，女子のデータの範囲，四分位範囲，四分位偏差を求めよ．

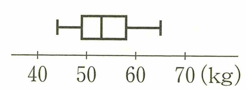
(2)　男子，女子のデータについて，四分位範囲によって中央値付近の散らばりの度合いを比較せよ．

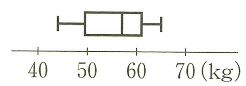
例題 7　箱ひげ図①

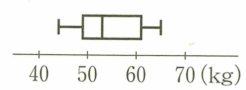
右のデータは，12人の生徒の体重(kg)である．

このデータを箱ひげ図に表したものを，次の①～③から選べ．

50,　52,　44,　62,　54,　65,
63,　48,　57,　60,　52,　46

①


②


③


解　$Q_1=49$，$Q_2=53$，$Q_3=61$より，箱の両端は49，61，箱の中の線は53なので，③

10　次のデータは，15人の生徒の通学時間(分)である．このデータの箱ひげ図をかけ．

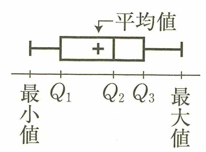
35,　18,　33,　25,　20,　40,　35,　15,　23,　38,　28,　45,　30,　36,　42

●**ポイント**

①　データの最大値と最小値の差を**範囲(レンジ)**という．範囲は簡単に求められるが，極端に離れた値の影響を受けやすく，散らばりの度合いを比較する量としては適切でない場合もある．

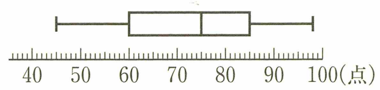
②　Q_3-Q_1を**四分位範囲**，$\dfrac{Q_3-Q_1}{2}$を**四分位偏差**という．四分位範囲や四分位偏差は，範囲に比べて極端に離れた値の影響を受けにくい．

③　右図のように，最小値，四分位数，最大値を箱と線(ひげ)で表した図を**箱ひげ図**という．ひげの両端までの長さは範囲，箱の長さは四分位範囲を表している．また，箱ひげ図に平均値を記入することもある．



例題 8　箱ひげ図②

右の箱ひげ図は，100人の生徒が受けたテストの得点の分布を表している．次の①，②について，正しいといえるものには○，いえないものには×をつけよ．



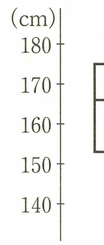
① 得点が最も高い生徒と最も低い生徒の得点の差は25点である．

② 得点が60点以上の生徒は75人以上いる．

解 ① 得点が最も高い生徒は98点，最も低い生徒は45点なので，得点の差は，98−45＝53（点）
よって，×

② $Q_1=60$（点）より，高い方から数えて75番目の生徒は60点以上である．よって，○

11 右の図は，１年生の生徒120人の身長のデータを箱ひげ図に表したものである．この箱ひげ図から読み取れることとして正しいといえるものを次の①～③から１つ選べ．

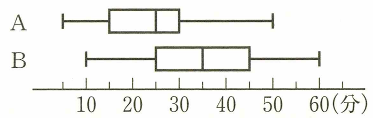


① 身長が160 cm未満の生徒は60人より多い．

② 身長が140 cm台の生徒は30人である．

③ 身長が170 cm以上の生徒は30人以上いる．

12 右の２つの箱ひげ図は，生徒100人がA問題とB問題を解くのに要した時間の分布を表している．次の問いに答えよ．



(1) それぞれのデータについて，範囲，四分位範囲，四分位偏差を求めよ．

(2) 次の①～③について，正しいといえるものには○，いえないものには×をつけよ．

① 20分以内で解いた生徒はA問題では25人以上，B問題では25人以下である．

② 30分以上要した生徒はどちらも半数未満である．

③ A問題を20分，B問題を30分で解いた生徒は，どちらも早い方から数えて50番目以内である．

13 次のデータは，A，B２つのにわとり小屋で１日にとれる卵の個数を10日間にわたって調べたものである．小屋ごとの箱ひげ図を並べてかけ．また，下の①，②について，最も適切なものをそれぞれ１つ選べ．

A　47，65，63，80，57，73，54，55，40，70（個）

B　36，56，63，48，30，52，70，47，44，60（個）

① 四分位範囲は（**ア** Aの方が大きい．　**イ** Bの方が大きい．　**ウ** 同じである．）

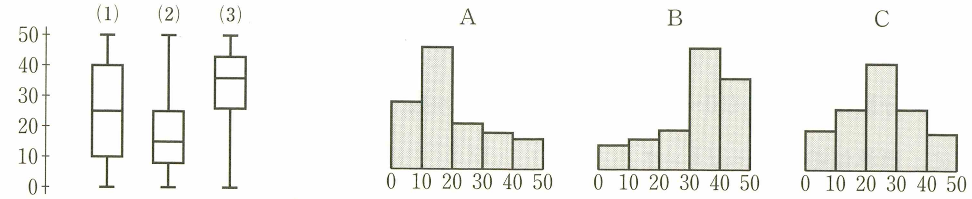
② Aの方が（**カ** 個数が多い方に分布している．　**キ** 散らばりの度合が大きい．　**ク** 散らばりの度合が小さい．）

●**ポイント**

① 箱ひげ図を並べると，複数のデータの分布を比較するのに便利である．箱ひげ図の箱やひげの長さで散らばりの様子を視覚的にとらえることができる．

例題 9　データの分布と箱ひげ図の関係

次の箱ひげ図(1)，(2)，(3)について，それぞれ対応するヒストグラムをA，B，Cの中から選べ．



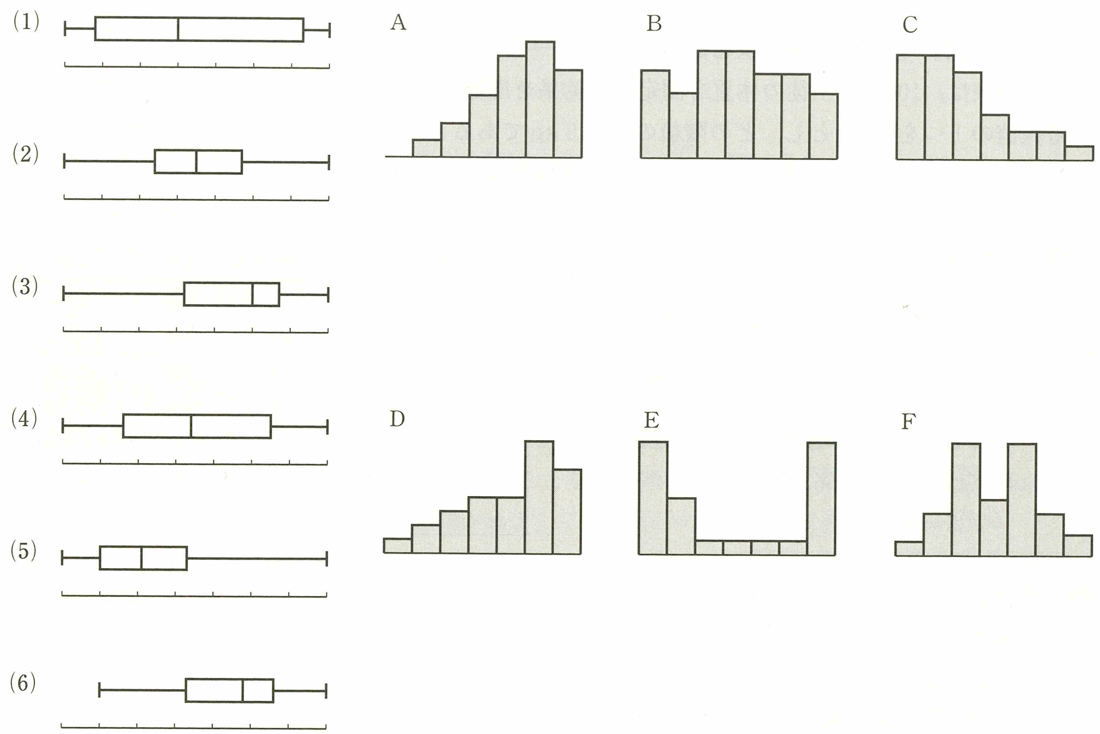
解 １つの山のヒストグラムの場合，ヒストグラムの山の高い部分に箱ひげ図の箱が対応し，山のすその部分に箱ひげ図のひげが対応する．

(1) 箱が範囲の中央にあり，上下のひげの長さが等しい．よって，対応するヒストグラムは，山が中央にある**C**．

(2) 箱が範囲の下側にあり，箱の下のひげが短い．よって，対応するヒストグラムは，山が左にある**A**．

(3) 箱が範囲の上側にあり，箱の上のひげが短い．よって，対応するヒストグラムは，山が右にある**B**．

14 次の箱ひげ図(1)～(6)に対応するヒストグラムを，A～Fから選べ．



●**ポイント**

① 箱ひげ図はヒストグラム同様，データの分布を表現するのに適している．ヒストグラムは，度数分布表のすべての階級の度数が必要となるのに対し，箱ひげ図は最小値，第１四分位数Q_1，中央値Q_2，第３四分位数Q_3，最大値の５つの数値がわかればかくことができる．

例題 10　分散，標準偏差

６つの数 0，2，2，3，5，6 について，分散 s^2，標準偏差 s を求めよ．

解　平均値は，$\frac{1}{6}(0+2+2+3+5+6)=3$

よって，分散は，$s^2=\frac{1}{6}\{(0-3)^2+(2-3)^2+(2-3)^2+(3-3)^2+(5-3)^2+(6-3)^2\}=\mathbf{4}$

ゆえに，標準偏差は，$s=\sqrt{4}=\mathbf{2}$

15 次のデータは，あるクラスの男子10人の小テストの結果である．分散，標準偏差を求めよ．

6，5，4，10，1，6，0，9，3，6（点）

16 変量 x が n 個の値 x_1，x_2，……，x_n をとるとするとき，分散 s^2 について，$s^2=\overline{x^2}-(\overline{x})^2$ が成り立つことを証明せよ．また，これを用いて上の**15**のデータの分散を求めよ．

17 次のデータは，あるクラスの女子10人の小テストの結果である．分散，標準偏差を求めよ．また，上の**15**のデータと散らばりの度合いを比較せよ．ただし，$\sqrt{6}=2.4$ とする．

7，8，1，6，5，10，6，6，3，8（点）

例題 11　度数分布表による標準偏差

次の表は，100人の生徒の体重を測定した結果を度数分布表にまとめたものである．ただし，階級値の１つを45kgとし，どの階級の幅も５kgであるとする．この表について，生徒100人の体重の平均値と標準偏差を求めよ．

階級値(kg)	45	50	55	60	65	70	計
度数(人)	2	10	20	32	26	10	100

解　度数分布表より，右のような表を作る．

x の平均値 $\overline{x}$ は，$\overline{x}=\frac{6000}{100}=\mathbf{60\,(kg)}$

x^2 の平均値 $\overline{x^2}$ は，$\overline{x^2}=\frac{363600}{100}=3636$

よって，x の標準偏差を s とすると，

$s=\sqrt{\overline{x^2}-(\overline{x})^2}=\sqrt{3636-60^2}$

$=\sqrt{36}=\mathbf{6\,(kg)}$

階級値 x (kg)	度数 f	xf	x^2f
45	2	90	4050
50	10	500	25000
55	20	1100	60500
60	32	1920	115200
65	26	1690	109850
70	10	700	49000
計	100	6000	363600

●ポイント

① 変量 x_1，x_2，……，x_n の平均値を $\overline{x}$ とするとき，$x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ を，それぞれ x_1，x_2，……，x_n の平均値からの**偏差**という．

② 偏差の２乗の平均値を**分散**といい，$\boldsymbol{s^2}$ で表す．分散の正の平方根を**標準偏差**といい，$\boldsymbol{s}$ で表す．これらは，変量の平均値のまわりの散らばりの度合いを表している．（s の単位は変量の単位と同じ）

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

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

18 次の表は，ある塾の生徒20人の通塾時間について調べた結果を度数分布表にまとめたものである．この表をもとに，生徒20人の通塾時間の標準偏差を求めよ．ただし，階級値の１つを５分とし，どの階級の幅も５分とする．

階級値(分)	5	10	15	20	25	計
度数（人）	2	2	6	6	4	20

19 次の表は，20人の高校生が１週間に部活動をする時間を調べ，度数分布表にまとめたものである．この表をもとに，20人の高校生の活動時間の標準偏差を求めよ．ただし，$\sqrt{5}=2.2$ とする．

階級(時間)	11以上～13未満	13～15	15～17	17～19	19～21	計
度数（人）	1	2	8	4	5	20

例題 12　合併されたデータの分析

10個の値からなるデータがある．そのうちの４個の値の平均値は13，分散は１であり，残りの６個の値の平均値は８，分散は２である．このとき，このデータの平均値 $\overline{x}$，分散 s^2 を求めよ．

解　平均値は，$\overline{x}=\frac{13\cdot4+8\cdot6}{10}=\mathbf{10}$

４個の値の２乗の和を a，６個の値の２乗の和を b とすると，

$\frac{1}{4}a-13^2=1$，$\frac{1}{6}b-8^2=2$ より，$a=680$，$b=396$

よって，10個の値からなるデータの分散は，$s^2=\frac{680+396}{10}-10^2=\mathbf{7.6}$

20 20人のけん垂の回数のデータがある．男子８人のデータの平均値は11回，分散は２であり，女子12人のデータの平均値は６回，分散は３である．20人のデータの平均値 $\overline{x}$，分散 s^2 を求めよ．

例題 13　仮平均①

変量 x のデータが，48，53，51，45，58のとき，仮平均を $x_0=50$，変量 u を $u=x-x_0$ とする．このとき，x の平均値 $\overline{x}$ は，$\overline{x}=x_0+\overline{u}$ となることを確かめよ．また，u のデータの分散 $s_u{}^2$ が x のデータの分散 $s_x{}^2$ に等しくなることを確かめよ．

解　$\overline{x}=51$，$s_x{}^2=19.6$

x	48	53	51	45	58	計
u	−2	3	1	−5	8	5
u^2	4	9	1	25	64	103

右の表より，$\overline{u}=\frac{1}{5}\cdot5=1$，$\overline{u^2}=\frac{1}{5}\cdot103=20.6$

よって，$x_0+\overline{u}=50+1=51=\overline{x}$，$s_u{}^2=\overline{u^2}-(\overline{u})^2=20.6-1=19.6=s_x{}^2$

●ポイント

① 度数分布表による標準偏差は，$s=\sqrt{\overline{x^2}-(\overline{x})^2}$ で計算する．例題11のような表を作り，xf，x^2f を計算して，$\overline{x}$，$\overline{x^2}$ を求めるとよい．

② m 個の値，n 個の値からなるデータを合わせた平均値を $\overline{x}$，m 個の値の２乗の和を a，n 個の値の２乗の和を b とすると，２つのデータを合わせた分散は，$s^2=\frac{a+b}{m+n}-(\overline{x})^2$ で求められる．

21 次のデータは，5人の生徒の身長である．仮平均を適当に決め，このデータの平均値，分散，標準偏差を求めよ．ただし，$\sqrt{3}=1.7$ とする．

172，166，173，170，164 (cm)

例題 14 仮平均②

仮平均を x_0 とし，正の値 c を用いて，$u=\dfrac{x-x_0}{c}$ とおき，変量 x を変量 u に変換する．このとき，x，u の平均値を $\overline{x}$，$\overline{u}$，標準偏差を s_x，s_u とすると，$\overline{x}=x_0+c\overline{u}$，$s_x=cs_u$ となることが知られている．これを用いて，次の変量 x の平均値 $\overline{x}$，標準偏差 s_x を求めよ．

342，282，318，306，288

解 仮平均 $x_0=300$，$c=6$，$u=\dfrac{x-300}{6}$ として，右のような表を作る．

x	u	u^2
342	7	49
282	−3	9
318	3	9
306	1	1
288	−2	4
計	6	72

これより，$\overline{u}=\dfrac{6}{5}=1.2$

よって，$\overline{x}=x_0+c\overline{u}=300+6\cdot1.2=\mathbf{307.2}$

$s_u=\sqrt{\overline{u^2}-(\overline{u})^2}=\sqrt{\dfrac{72}{5}-\left(\dfrac{6}{5}\right)^2}=\sqrt{\dfrac{324}{25}}=\dfrac{18}{5}=3.6$ より，

$s_x=6\cdot3.6=\mathbf{21.6}$

22 次のデータは，5人の生徒の走り幅跳びの記録である．平均値，標準偏差を求めよ．

465，425，455，470，445 (cm)

23 次の表は，60人の生徒のテストの得点を変量 x として，度数分布表にまとめたものである．ただし，階級値の1つを15点とし，どの階級の幅も10点であるとする．$u=\dfrac{x-55}{10}$ として，次の問いに答えよ．ただし，$\sqrt{6}=2.45$ とする．

階級値 x (点)	15	25	35	45	55	65	75	85	95	計
度数(人)	2	3	5	9	12	10	10	6	3	60

(1) 変量 u の平均値 $\overline{u}$，変量 x の平均値 $\overline{x}$ を求めよ．

(2) 変量 u の標準偏差 s_u，変量 x の標準偏差 s_x を求めよ．

●ポイント

① $u=x-x_0$ とすると，平均値 $\overline{x}$，$\overline{u}$，分散 s_x^2，s_u^2，標準偏差 s_x，s_u について，
$\overline{x}=x_0+\overline{u}$，$s_x^2=s_u^2$，$s_x=s_u$ が成り立つ．

② 正の値 c を用いて，$u=\dfrac{x-x_0}{c}$ とすると，平均値 $\overline{x}$，$\overline{u}$，標準偏差 s_x，s_u について，
$\overline{x}=x_0+c\overline{u}$，$s_x=cs_u$ が成り立つ．

③ 度数分布表では，最大の度数をもつ階級の階級値を x_0，階級の幅を c として変量変換すると，計算が簡単になることが多い．

例題 15 外れ値

データの第1四分位数を Q_1，第3四分位数を Q_3 とし，以下にあてはまるデータを外れ値とする．

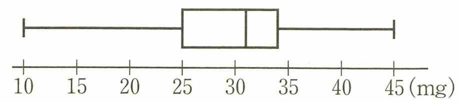
$Q_1-1.5\times(Q_3-Q_1)$ 以下，または，$Q_3+1.5\times(Q_3-Q_1)$ 以上

このとき，次のデータの箱ひげ図をかけ．また，外れ値を答えよ．

(1) 32，27，25，33，10，30，34，45，19，37 (mg)

(2) 8，25，15，23，50，0，12，20，22 (点)

解 (1) 箱ひげ図は右のようになる．

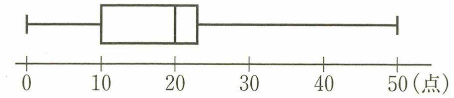


また，$Q_1=25$mg，$Q_3=34$mg より，

$Q_1-1.5\times(Q_3-Q_1)=11.5$ (mg)，

$Q_3+1.5\times(Q_3-Q_1)=47.5$ (mg) だから，外れ値は**10mg**．

(2) 箱ひげ図は右のようになる．



また，$Q_1=10$点，$Q_3=24$点より，

$Q_1-1.5\times(Q_3-Q_1)=-11$ (点)

$Q_3+1.5\times(Q_3-Q_1)=45$ (点) だから，外れ値は**50点**．

24 例題15で定めた範囲にあてはまるデータを外れ値とするとき，次のデータの外れ値を答えよ．

(1) 15，42，13，31，20，44，10，60，25，88，12，23，30 (℃)

(2) 33，24，15，30，22，33，25，40，1，20，35，28 (分)

(3) 20，65，48，93，55，61，10，58，45，50，100 (mL)

25 次の問いに答えよ．

(1) 平均値，中央値，最頻値のうち，外れ値の影響を最も受けやすいものを選べ．

(2) 四分位範囲と分散のうち，外れ値の影響を受けやすい方を選べ．

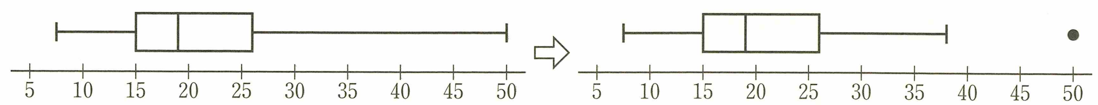
●ポイント

① データの中で，他の値から極端にかけ離れた値を**外れ値**という．外れ値の目安としては，例えば，

(外れ値) $\leqq Q_1-1.5\times(Q_3-Q_1)$，または，$Q_3+1.5\times(Q_3-Q_1)\leqq$ (外れ値)

のように，Q_1 から小さい方(Q_3 から大きい方)へ四分位範囲の1.5倍以上離れているものとすることがある．

② 以下のように，外れ値が50のとき，外れ値を点●などで記入し，外れ値を除いたデータの中で最も大きな値をひげの右端にとって箱ひげ図をかくことがある．



〔注〕 測定ミスや記入ミスなどが明らかな値は**異常値**という．異常値とは違い，外れ値は必ずしも除外すべきものではない．

混合問題

A

1 右のデータは5人の生徒の垂直跳びの結果である． 58，52，67，61，56 (cm)
ただし，データのうち1個の値は誤りであるという．正しいデータによると，中央値は61 cm，平均値は59.6 cm である．誤りの値，正しい値をそれぞれ求めよ．

2 ある商店では，毎日80個の弁当を仕入れて販売している．
右のデータは，最近の20日間について，1日に売れた弁当の個数を記録したものである．
67，73，50，70，58，51，65，43，60，63，
53，74，77，47，67，63，59，75，72，48（個）
(1) このデータの四分位数 Q_1，Q_2，Q_3 を求めよ．また，箱ひげ図をかけ．
(2) 弁当の売り切れが生じる日数を25％程度にするには，1日に何個の弁当を仕入れるのが適当か．上のデータにもとづいて答えよ．

3 右の変量 x，y のデータについて，次の問いに答えよ．

x	10	8	7	6	10	10	7	11	6	5
y	2	11	4	7	8	4	12	7	9	6

(1) 変量 x，y それぞれの分散 $s_x{}^2$，$s_y{}^2$ を求めよ．
(2) 標準偏差によって，データの平均値のまわりの散らばりの度合いを比較せよ．

4 50人のテスト（30点満点）の得点のデータがある．A組の20人の平均値は22点，分散は6であり，B組の30人の平均値は17点，分散は11である．50人のデータの平均値 $\overline{x}$，分散 s^2 を求めよ．

B

5 右の表は50人の生徒のテストの得点を度数分布表にまとめたものである．ただし，階級値の1つを15点とし，どの階級の幅も10点であるとする．このとき，次の問いに答えよ．ただし，$\sqrt{11}=3.3$ とし，標準偏差は四捨五入して整数で求めよ．

階級値 x (点)	15	25	35	45	55	65	75	85	95	計
度数（人）	2	4	5	5	10	11	7	4	2	50

(1) 適当な変量変換により，得点の平均値 $\overline{x}$，標準偏差 s_x を求めよ．
(2) $t=\dfrac{10(x-\overline{x})}{s_x}+50$ で与えられる変量を偏差値という．54点，84点の偏差値を求めよ．

6 5，7，9，a，b の平均値が $a+4$ のとき，分散が最小となる a，b の値と，そのときの分散の値を求めよ．

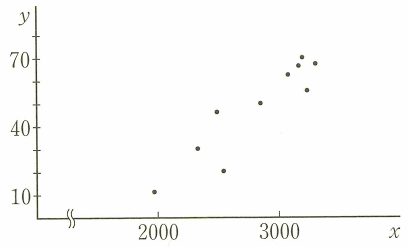
■ヒント
6 b を a を用いて表し，分散を a だけの式で表して考える．

2 データの相関・仮説検定

例題 1 散布図

下のデータは，ある地域10世帯の1人1日当たりのカロリー摂取量を x kcal，食品摂取量に対するたんぱく質の割合を y％として調べたものである．x と y についての散布図をかき，どのような相関関係があるか答えよ．

x	3199	3168	3079	3308	2850	3238	2490	2330	2542	1971
y	70.3	66.6	62.7	67.5	50.2	55.7	46.4	30.3	20.4	11.4

解


左の散布図より，カロリー摂取量が増えるとたんぱく質摂取の割合も増える傾向にあるので，2つの変量の間には**正の相関関係がある**といえる．

1 下のような2つの変量 x，y がある．散布図をかき，相関関係を調べよ．

(1)

x	35	43	12	23	40	18	32	28
y	38	34	16	22	43	14	26	30

(2)

x	2.8	4.6	7.2	1.8	3.4	5.4	4.0	6.3	2.2
y	21	16	8	36	32	18	26	13	28

例題 2 相関係数

右の表は，2つの変量 x，y についての5個のデータの値である．x，y の相関係数 r を求め，どのような相関関係があるか答えよ．ただし，$\sqrt{5}=2.24$ とし，答えは四捨五入して小数第2位まで求めよ．

x	10	4	8	6	2
y	9	3	6	7	0

解 x，y の平均値を $\overline{x}$，$\overline{y}$ とすると，$\overline{x}=\dfrac{30}{5}=6$，$\overline{y}=\dfrac{25}{5}=5$

$(x_1-\overline{x})^2+\cdots\cdots+(x_5-\overline{x})^2=4^2+(-2)^2+2^2+0^2+(-4)^2=40$

$(y_1-\overline{y})^2+\cdots\cdots+(y_5-\overline{y})^2=4^2+(-2)^2+1^2+2^2+(-5)^2=50$

$(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_5-\overline{x})(y_5-\overline{y})=4\cdot4+(-2)\cdot(-2)+2\cdot1+0\cdot2+(-4)\cdot(-5)=42$

よって，$r=\dfrac{42}{\sqrt{40}\sqrt{50}}=\dfrac{21\sqrt{5}}{50}\fallingdotseq\mathbf{0.94}$

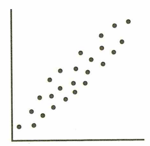
相関係数 r が1に近いので，2つの変量 x，y には，**強い正の相関関係がある．**

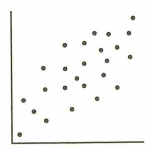
●ポイント

① 2つの変量の値の組を座標とする点を平面上にとり，これらの関係を表した図を**散布図**という．

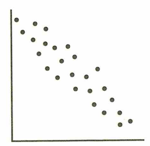
② 相関関係と散布図を示すと次の(ア)〜(ウ)のようになる．

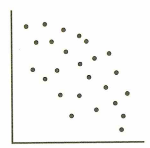
(ア)**正の相関関係がある**

Ⓐ強い


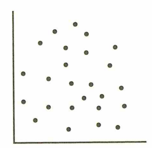
Ⓑ弱い


(イ)**負の相関関係がある**

Ⓒ強い


Ⓓ弱い


(ウ)**相関関係がない**

Ⓔない


2 下の表は，５人の生徒a，b，c，d，eの身長x(cm)と体重y(kg)の測定結果である．表の空欄をうめ，身長と体重の相関係数rを求めよ．ただし，答えは四捨五入して小数第２位まで求めよ．

	x	y	$x-\overline{x}$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$	$(x-\overline{x})^2$	$(y-\overline{y})^2$
a	169	59					
b	171	56					
c	175	61					
d	166	55					
e	174	59					
計	855	290					

3 右の表の２つの変量x，yについて，相関係数rを求めよ．ただし，$\sqrt{3}=1.73$とし，答えは四捨五入して小数第２位まで求めよ．

x	14	13	11	9	5	8
y	10	8	11	13	14	16

4 次の変量x，yの相関係数rを求めよ．また，
$|r|<0.3$のとき「ほとんど相関関係がない」，
$0.3\leqq|r|<0.9$のとき「正(負)の相関関係がある」，
$|r|\geqq0.9$のとき「強い正(負)の相関関係がある」
とするとき，xとyの間にはどのような相関関係があるか答えよ．
　ただし，$\sqrt{2}=1.41$，$\sqrt{6}=2.45$とし，答えは四捨五入して小数第２位まで求めよ．

(1)

x	7	10	12	9	14	10	13	5
y	11	6	8	5	8	9	6	11

(2)

x	7	14	8	13	3	6	10	11
y	9	15	10	14	8	9	10	13

(3)

x	8	9	6	3	10	6	9	5
y	5	10	9	4	2	5	4	9

●ポイント

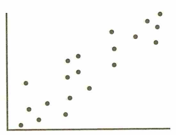
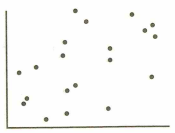
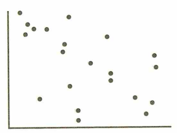
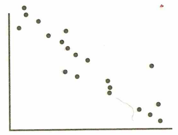
① xの偏差とyの偏差の積$(x_k-\overline{x})(y_k-\overline{y})$の平均値
$\dfrac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$を$x$と$y$の**共分散**といい，$\boldsymbol{s_{xy}}$で表す．

② 共分散s_{xy}を，標準偏差s_x，s_yの積s_xs_yで割った量を**相関係数**といい，$\boldsymbol{r}$で表す．

$r=\dfrac{s_{xy}}{s_xs_y}$　$s_x=\sqrt{\dfrac{1}{n}\{(x_1-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}}$，$s_y=\sqrt{\dfrac{1}{n}\{(y_1-\overline{y})^2+\cdots\cdots+(y_n-\overline{y})^2\}}$ より，

$$r=\frac{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})}{\sqrt{(x_1-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2}\sqrt{(y_1-\overline{y})^2+\cdots\cdots+(y_n-\overline{y})^2}}$$

③ 相関係数は$-1\leqq r\leqq1$の値をとり，$|r|$が１に近いほど，２つの変量x，yの相関関係が強く，０に近いほど相関関係は弱い．rの値が１に近いほど，正の相関が強く，散布図の点は右上がりに分布し，rの値が-1に近いほど，負の相関が強く，散布図の点は右下がりに分布する．

$r=0.9$

$r=0.5$

$r=-0.5$

$r=-0.9$


例題 3　相関表と相関係数

あるクラスの32人について，兄弟姉妹の数(x)と，いとこの数(y)を調べたところ，右の相関表を得た．次の問いに答えよ．

x＼y	1	2	3	4
2	0	1	4	3
1	4	4	4	4
0	3	4	1	0

(1) 変量yの平均値$\overline{y}$と，分散s_y^2を求めよ．
(2) 変量xと変量yの相関係数rを求めよ．

解 (1) 右の相関表を縦に見て人数を足していくと各階級の度数がわかるので，

平均値は，$\overline{y}=\dfrac{1}{32}(7\cdot1+9\cdot2+9\cdot3+7\cdot4)=\mathbf{2.5}$

分散は，$s_y^2=\dfrac{1}{32}\{7\cdot(1-2.5)^2+9\cdot(2-2.5)^2+9\cdot(3-2.5)^2+7\cdot(4-2.5)^2\}=\dfrac{36}{32}=\dfrac{9}{8}=\mathbf{1.125}$

(2) 変量xの平均値，分散をそれぞれ$\overline{x}$，s_x^2とすると，(1)と同様にして，$\overline{x}=1$，$s_x^2=0.5$
xとyの相関係数をrとする．$x-\overline{x}$を固定させて共分散を計算すると，$\overline{x}=1$より，

$$r=\frac{1}{s_xs_y}\cdot\frac{1}{32}[(0-1)\{3\cdot(1-2.5)+4\cdot(2-2.5)+1\cdot(3-2.5)+0\cdot(4-2.5)\}$$
$$+(2-1)\{0\cdot(1-2.5)+1\cdot(2-2.5)+4\cdot(3-2.5)+3\cdot(4-2.5)\}]$$
$$=\frac{1}{32}\cdot\frac{12}{s_xs_y}=\frac{1}{32}\cdot\frac{12}{\sqrt{0.5}\sqrt{1.125}}=\mathbf{0.5}$$

5 下の表は，生徒20人の身長x(cm)と体重y(kg)を調べた結果である．x，yの相関表を作れ．ただし，xについては階級の幅を６とし，階級は150から区切り始めるものとする．yについては階級の幅を４とし，階級は50から区切り始めるものとする．

x	166	163	173	158	167	170	172	152	178	167
y	55.5	65.5	62.5	53.5	59.5	57.5	65.5	59.0	68.5	60.5

x	169	165	155	173	161	164	166	160	171	157
y	61.5	57.0	51.5	63.0	57.0	63.5	59.0	61.5	65.5	60.0

6 右の表は，20人の生徒が受けた小テスト(５点満点)の，Ａ教科の得点(x)と，Ｂ教科の得点(y)の相関表である．次の問いに答えよ．
(1) 表中のaの値を求めよ．
(2) 変量xの平均値$\overline{x}$，分散s_x^2を求めよ．
(3) 変量xと変量yの相関係数rを求めよ．

x＼y	0	1	2	3	4	5
5				1	1	
4		1	2	3		
3		1	3	1		
2		2	a			
1		1				
0	1					

●ポイント

① ２つの変量の相互の関連を表す度数分布表を**相関表**という．
② 相関表では，それぞれの変量について，各階級の度数が分かるので，相関係数を求める際に，平均値，分散を求める計算が楽になる場合がある．
③ 相関表において相関係数を求める場合，一方の変量を固定させて，
共分散$s_{xy}=\dfrac{1}{n}\{(x_1-\overline{x})(y_1-\overline{y})+\cdots\cdots+(x_n-\overline{x})(y_n-\overline{y})\}$を計算する．

例題 4　仮説検定の考え方

Ｐ社では，これまで販売していた商品Ａを改良した商品Ｂを作り，20人に商品Ａ，Ｂのどちらが使いやすいかを比べてもらったところ，16人が商品Ｂの方が使いやすいと答えた．このとき，商品Ｂは，商品Ａより使いやすくなったと判断してよいかを考える．

このことについて考えるため，表と裏の出方が同様に確からしい20枚のコインを同時に投げる作業を200回行ったところ，表の出た枚数は以下のようになった．

枚数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
度数	0	1	0	1	3	7	14	24	31	35	32	24	15	7	3	1	0	1	1	0

上記の結果を用いて，次の問いに答えよ．

(1)　基準となる確率が0.05のとき，商品Ｂは商品Ａより使いやすくなったと判断できるか．

(2)　基準となる確率が0.01のとき，商品Ｂは商品Ａより使いやすくなったと判断できるか．

解　「ＢはＡより使いやすい」という仮説Ｈを立てる．仮説Ｈが起こらないこととして，「ＡとＢは同等」がどの程度起こるかを調べる．ここで，ＡとＢは同等だとすると，Ａの方がよい場合とＢの方がよい場合が起こる可能性は半々だから，コイン投げに例えることができる．

コインを投げ，表が出た場合をＢの方が使いやすいとする．コイン20枚を同時に投げる作業を200回行い，16枚以上が表になった回数の相対度数を p とする．p が基準となる確率以下であれば，仮説Ｈが起こらないことを棄却でき，仮説Ｈが正しいと判断できる．

(1)　$p=(1+0+1+1+0)\div200=0.015<0.05$ より，**使いやすくなったと判断できる．**

(2)　$p=(1+0+1+1+0)\div200=0.015>0.01$ より，**使いやすくなったとは判断できない．**

7　例題４のＰ社では，さらに改良した商品Ｃを作り，20人に商品Ａ，Ｃのどちらが使いやすいかを比べてもらったところ，15人が商品Ｃの方が使いやすいと答えた．次の問いに答えよ．

(1)　基準となる確率が0.05のとき，商品Ｃは商品Ａより使いやすくなったと判断できるか．

(2)　基準となる確率が0.01のとき，商品Ｃは商品Ａより使いやすくなったと判断できるか．

●ポイント

①　ある仮説を立て，その仮説が正しいかどうか(妥当かどうか)を判定する統計的手法を**仮説検定**といい，次のような手順で行う．

(i)　主張したい仮説 H_1 と仮説 H_1 の内容に対立する仮説 H_0 を立てる．

(ii)　基準となる確率を決める．

(iii)　仮説 H_0 が起こる確率を求める．

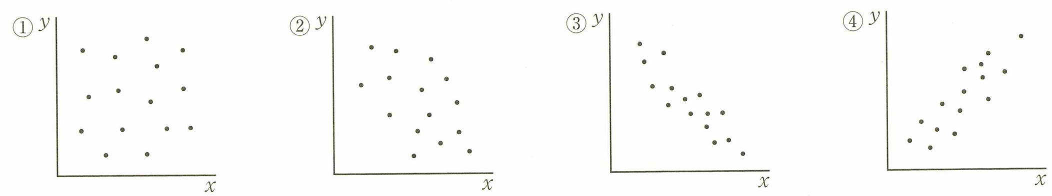
(iv)　(iii)で求めた確率が(ii)で決めた確率よりも小さければ，仮説 H_0 を棄却する．
仮説 H_0 を棄却されれば，仮説 H_1 を採択する．

〔注〕　基準となる確率は0.05や0.01を用いることが多い．

混合問題

Ａ

1　下の①，②，③，④は，ある２つの変量 x と y のデータについての散布図である．①，②，③，④の x と y の相関係数は，0.82，0.12，-0.57，-0.92 のいずれかである．①，②，③，④のデータの相関係数をそれぞれ答えよ．



2　右のデータは，Ａ～Ｆの６人が受けた小テスト(20点満点)のＰ教科の得点(x)とＱ教科の得点(y)である．x，y の相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，答えは四捨五入して小数第２位まで求めよ．

	A	B	C	D	E	F
x	14	16	13	14	15	18
y	12	15	12	13	15	17

3　右の表は，10人が行ったＡのゲームの得点(x)とＢのゲームの得点(y)の相関表である．次の問いに答えよ．

(1)　変量 x の平均値 $\overline{x}$，分散 s_x^2 を求めよ．

(2)　変量 x と変量 y の相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，四捨五入して小数第２位まで求めよ．

x \ y	0	1	2	3	4
3				1	2
2			1	2	2
1			1		
0		1			

4　右の表は，10人の生徒が受けたＡテストの得点(x)とＢテストの得点(y)である．変量 x と変量 y の相関表を作れ．また，相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，四捨五入して小数第２位まで求めよ．

x	6	9	8	9	9	6	9	6	8	10
y	6	7	6	6	8	6	7	6	8	10

Ｂ

5　右の表は，５人の生徒の数学のテストの得点(x)と国語のテストの得点(y)である．次の問いに答えよ．ただし，a は整数とする．

数学(x)	9	7	5	9	a
国語(y)	10	9	7	7	7

(1)　相関係数が $\dfrac{3\sqrt{2}}{8}$ であるとき，a の値を求めよ．ただし，共分散の計算は次のことを用いてよい．

● ２つの変量 x，y の共分散 s_{xy} について，$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$

(2)　a の値が(1)で求めた値であるとき，変量 x が９点，変量 y が７点となっている生徒の変量 x の値は誤りであることがわかり，正しい値の４点に修正した．修正前と修正後では，相関関係はどのように変わるか．

■ヒント

5　(1)　$\overline{x}$ を a で表し，$s_x^2=\overline{x^2}-(\overline{x})^2$，$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$ を利用して，相関係数を a の式で表す．

例題 4　仮説検定の考え方

Ｐ社では，これまで販売していた商品Ａを改良した商品Ｂを作り，20人に商品Ａ，Ｂのどちらが使いやすいかを比べてもらったところ，16人が商品Ｂの方が使いやすいと答えた．このとき，商品Ｂは，商品Ａより使いやすくなったと判断してよいかを考える．

このことについて考えるため，表と裏の出方が同様に確からしい20枚のコインを同時に投げる作業を200回行ったところ，表の出た枚数は以下のようになった．

枚数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
度数	0	1	0	1	3	7	14	24	31	35	32	24	15	7	3	1	0	1	1	0

上記の結果を用いて，次の問いに答えよ．

(1)　基準となる確率が0.05のとき，商品Ｂは商品Ａより使いやすくなったと判断できるか．

(2)　基準となる確率が0.01のとき，商品Ｂは商品Ａより使いやすくなったと判断できるか．

解　「ＢはＡより使いやすい」という仮説Ｈを立てる．仮説Ｈが起こらないこととして，「ＡとＢは同等」がどの程度起こるかを調べる．ここで，ＡとＢは同等だとすると，Ａの方がよい場合とＢの方がよい場合が起こる可能性は半々だから，コイン投げに例えることができる．

コインを投げ，表が出た場合をＢの方が使いやすいとする．コイン20枚を同時に投げる作業を200回行い，16枚以上が表になった回数の相対度数を p とする．p が基準となる確率以下であれば，仮説Ｈが起こらないことを棄却でき，仮説Ｈが正しいと判断できる．

(1)　$p=(1+0+1+1+0)\div200=0.015<0.05$ より，**使いやすくなったと判断できる．**

(2)　$p=(1+0+1+1+0)\div200=0.015>0.01$ より，**使いやすくなったとは判断できない．**

7　例題４のＰ社では，さらに改良した商品Ｃを作り，20人に商品Ａ，Ｃのどちらが使いやすいかを比べてもらったところ，15人が商品Ｃの方が使いやすいと答えた．次の問いに答えよ．

(1)　基準となる確率が0.05のとき，商品Ｃは商品Ａより使いやすくなったと判断できるか．

(2)　基準となる確率が0.01のとき，商品Ｃは商品Ａより使いやすくなったと判断できるか．

●ポイント

①　ある仮説を立て，その仮説が正しいかどうか(妥当かどうか)を判定する統計的手法を**仮説検定**といい，次のような手順で行う．

(i)　主張したい仮説 H_1 と仮説 H_1 の内容に対立する仮説 H_0 を立てる．

(ii)　基準となる確率を決める．

(iii)　仮説 H_0 が起こる確率を求める．

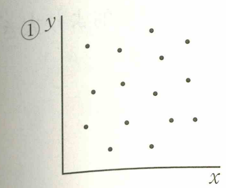
(iv)　(iii)で求めた確率が(ii)で決めた確率よりも小さければ，仮説 H_0 を棄却する．
仮説 H_0 を棄却されれば，仮説 H_1 を採択する．

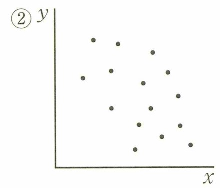
〔注〕　基準となる確率は0.05や0.01を用いることが多い．

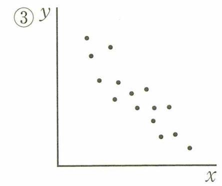
混合問題

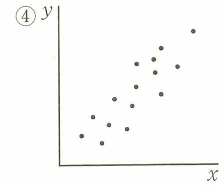
A

1　下の①，②，③，④は，ある２つの変量 x と y のデータについての散布図である．①，②，③，④の x と y の相関係数は，0.82，0.12，-0.57，-0.92 のいずれかである．①，②，③，④のデータの相関係数をそれぞれ答えよ．









2　右のデータは，Ａ～Ｆの６人が受けた小テスト(20点満点)のＰ教科の得点(x)とＱ教科の得点(y)である．x，y の相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，答えは四捨五入して小数第２位まで求めよ．

	A	B	C	D	E	F
x	14	16	13	14	15	18
y	12	15	12	13	15	17

3　右の表は，10人が行ったＡのゲームの得点(x)とＢのゲームの得点(y)の相関表である．次の問いに答えよ．

(1)　変量 x の平均値 $\overline{x}$，分散 $s_x{}^2$ を求めよ．

(2)　変量 x と変量 y の相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，四捨五入して小数第２位まで求めよ．

x＼y	0	1	2	3	4
3				1	2
2			1	2	2
1			1		
0		1			

4　右の表は，10人の生徒が受けたＡテストの得点(x)とＢテストの得点(y)である．変量 x と変量 y の相関表を作れ．また，相関係数 r を求めよ．ただし，$\sqrt{5}=2.24$ とし，四捨五入して小数第２位まで求めよ．

x	6	9	8	9	9	6	9	6	8	10
y	6	7	6	6	8	6	7	6	8	10

B

5　右の表は，５人の生徒の数学のテストの得点(x)と国語のテストの得点(y)である．次の問いに答えよ．ただし，a は整数とする．

数学(x)	9	7	5	9	a
国語(y)	10	9	7	7	7

(1)　相関係数が $\dfrac{3\sqrt{2}}{8}$ であるとき，a の値を求めよ．ただし，共分散の計算は次のことを用いてよい．

● ２つの変量 x，y の共分散 s_{xy} について，$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$

(2)　a の値が(1)で求めた値であるとき，変量 x が９点，変量 y が７点となっている生徒の変量 x の値は誤りであることがわかり，正しい値の４点に修正した．修正前と修正後では，相関関係はどのように変わるか．

■ヒント

5　(1)　$\overline{x}$ を a で表し，$s_x{}^2=\overline{x^2}-(\overline{x})^2$，$s_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}$ を利用して，相関係数を a の式で表す．

章末問題 A

1 次のデータは，10人の高校生があるパズルを解くまでにかかった時間を調べたものである．

9，11，6，13，12，10，7，9，18，5（分）

(1) このデータの平均値を求めよ．

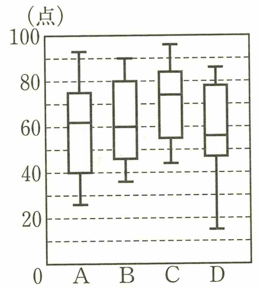
(2) データには一部誤りがあった．正しくは，９分のうちの１つが８分，７分が12分，18分が14分である．この誤りを修正すると，データの平均値，中央値，分散は，それぞれ修正前より大きくなるか，小さくなるか，または修正前と等しいか，答えよ．

2 次のデータの四分位数Q_1，Q_2，Q_3，四分位範囲，四分位偏差を求めよ．

(1) 174，163，181，171，158，168，166，161，177，167（cm）

(2) 57，65，41，54，38，62，47，40，42，51，59，63（kg）

3 右の図は，160人の生徒が受けた４種類のテストA，B，C，Dの得点のデータを箱ひげ図に表したものである．この図から読み取れることとして正しいといえるものを，次の①～④からすべて選べ．



① 20点台の生徒は，Aにはいるが，Bにはいない．

② 40点以上の生徒が最も多いのはCである．

③ 全てのテストが70点であった生徒は，どのテストでも上位80位以内である．

④ 50点以上の生徒は，Bでは120人以下，Cでは120人以上である．

4 次のデータは，ある数学の問題を解いた10人の生徒が，正解を出すまでにかかった時間x（分）を記録したものである．ただし，xの平均値を$\overline{x}$で表している．また，10人全員が制限時間の20分以内で正解を出している．次の問いに答えよ．

x	11	17	a	16	9	12	7	11	b	5
$(x-\overline{x})^2$	1	25	9	c	9	0	25	1	d	49

(1) $\overline{x}$を求めよ．

(2) a，b，c，dの値を求めよ．

(3) このデータの標準偏差を求めよ．

5 ５人の生徒が，A，B２種類のゲームを行った．次のデータは，ゲームAの得点x（点）とゲームBの得点y（点）を記録したものである．ただし，表中のaの値は整数である．

x	7	6	a	10	4
y	5	6	9	7	3

(1) 変量xの標準偏差は２点であった．aの値を求めよ．

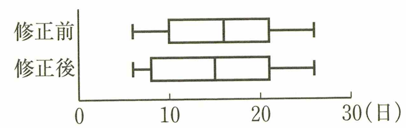
(2) aの値が(1)で求めた値のとき，２つの変量xとyの相関係数rを求めよ．

章末問題 B

1 次のデータは，11人の生徒が１か月間に部活動に参加した日数を調べたものである．

23，19，10，16，6，15，20，8，12，21，26（日）

このデータには１つだけ誤りがある．この誤りを修正すると，修正前と比べて，平均値は１日減少し，四分位偏差は１日増加する．また，修正前と修正後の箱ひげ図は，それぞれ右図のようになった．



誤りのデータはどれか．修正前の値を答えよ．

2 次の問いに答えよ．

(1) 変量xのN個の値の中で，x_1がf_1個，x_2がf_2個，……，x_nがf_n個あるとする．a，bを定数として，変量uを$u=ax+b$，xの平均値を$\overline{x}$，uの平均値を$\overline{u}$とすると，$\overline{u}=a\overline{x}+b$が成り立つことを示せ．

(2) (1)において，変量xの標準偏差をs_x，変量uの標準偏差をs_uとすると，$s_u=|a|s_x$が成り立つことを示せ．

(3) ２つの変量x，yのn個の値の組$(x_1,\ y_1)$，$(x_2,\ y_2)$，……，$(x_n,\ y_n)$について，a，b，c，dを定数としてそれぞれ，$z=ax+b$，$w=cy+d$と変換する．$ac>0$のとき，z，wの相関係数r_{zw}と，x，yの相関係数r_{xy}について，$r_{zw}=r_{xy}$が成り立つことを示せ．

(4) ２つの変量x，yのn個の値の組$(x_1,\ y_1)$，$(x_2,\ y_2)$，……，$(x_n,\ y_n)$が座標平面上において，直線$y=ax+b$（a，bは定数で，$a\neq0$）上にあるとき，相関係数をrとすると，$a>0$ならば$r=1$，$a<0$ならば$r=-1$であることを示せ．

3 次の問いに答えよ．

(1) 変量x_k（$k=1,\ 2,\ \cdots\cdots,\ n$）の平均値が$-3$，分散が５，変量$y_k=ax_k+b$（$a>0$）の平均値が０，変量${y_k}^2$の平均値が10であるとき，定数$a$，$b$の値を求めよ．

(2) 10個の数があり，全体の平均は５，標準偏差は$\sqrt{15}$であるが，その中の６個については，平均は３，標準偏差は３であるという．このとき，残りの４個の平均と標準偏差を求めよ．

(3) ８個のデータx_1，x_2，……，x_8があり，平均は４，標準偏差は２であるという．これにデータ$x_9=11$，$x_{10}=3$を付け加えたとき，10個のデータx_1，x_2，……，x_{10}の平均と標準偏差を求めよ．

4 ３つの変量x_k，y_k，z_k（$k=1,\ 2,\ \cdots\cdots,\ n$）があり，$y_k=2{x_k}^2-3x_k$，$z_k=4{x_k}^2-5x_k$とする．変量$y_k$の平均値が６，変量$z_k$の平均値が14のとき，変量$x_k$の平均値$\overline{x}$および分散${s_x}^2$を求めよ．

5 ２つの変量x，yのn個の値の組$(x_1,\ y_1)$，$(x_2,\ y_2)$，……，$(x_n,\ y_n)$について，xとyの相関係数をrとすると，$|r|\leqq1$となることを示せ．