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．

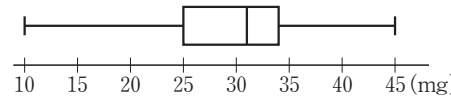
$Q_1-1.5\times(Q_3-Q_1)$ 以下，または，$Q_3+1.5\times(Q_3-Q_1)$ 以上

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．

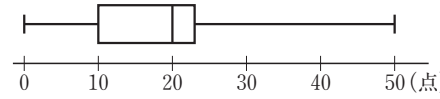


また，$Q_1=25$ mg，$Q_3=34$ mg より，

$Q_1-1.5\times(Q_3-Q_1)=11.5$ (mg)，

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

(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．

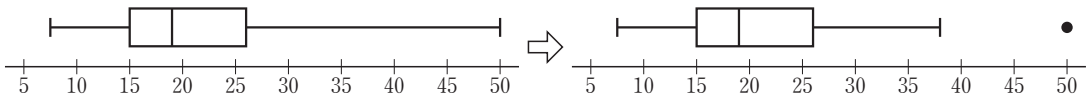
●ポイント

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．

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)　$p=(1+0+1+1+0)\div 200=0.015<0.05$ より，**使いやすくなったと判断できる．**

(2)　$p=(1+0+1+1+0)\div 200=0.015>0.01$ 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．

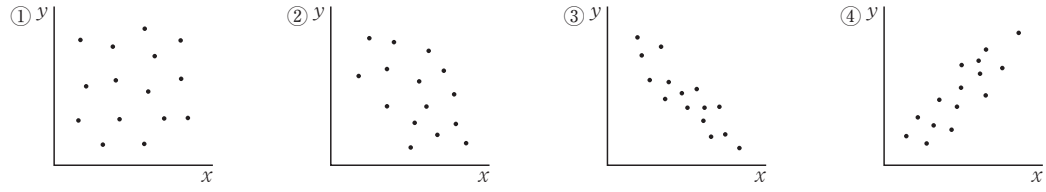
仮説H_0を棄却されれば，仮説H_1を採択する．

〔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）

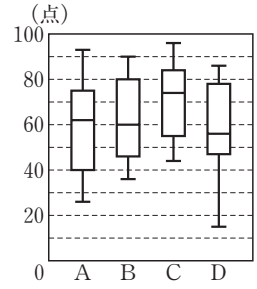
3 右の図は，160人の生徒が受けた4種類のテストA，B，C，Dの得点のデータを箱ひげ図に表したものである．この図から読み取れることとして正しいといえるものを，次の①～④からすべて選べ．

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，B 2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．

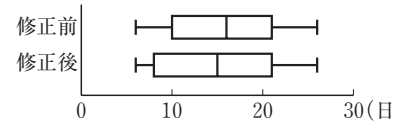
章末問題 B

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，$a\neq 0$）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，$y_k=2x_k^2-3x_k$，$z_k=4x_k^2-5x_k$ 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，$|r|\leqq 1$ となることを示せ．