10

相関と回帰

10-1　相関係数

2つの変量 X, Y に関する n 対のデータ (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) が与えられたとき，これら n 個の点を平面上にプロットした図を**散布図**という．散布図をみれば，X と Y の間に直線的関係があるか，あるとすれば，それはどの程度かなどがわかる．

相関係数　X と Y の間の直線的関係の程度を表す統計的尺度を**相関係数**といい，

$$r=\frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y}$$

で定義する．ここで，$\bar{x}$, $\bar{y}$ は X, Y の平均，s_x, s_y は X, Y の標準偏差を表す．分子の $\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})$ を X と Y の**共分散**という．

r の計算　r の計算は，上の定義式より次の式が便利である．

$$r=\frac{n\sum x_i y_i-(\sum x_i)(\sum y_i)}{\sqrt{\{n\sum x_i^2-(\sum x_i)^2\}\{n\sum y_i^2-(\sum y_i)^2\}}}$$

r の性質　(i)　$-1\leqq r\leqq 1$ が成り立つ．r の絶対値が1に近いときは，X, Y の間に強い直線関係があることを意味し，r の絶対値が0に近いときは，X, Y の間に直線的関係がほとんどないことを意味する．$r>0$ ならば，**正の相関**があるといい，$r<0$ ならば**負の相関**があるという．$|r|=1$ のとき**完全相関**，$r=0$ のとき**無相関**という．

(ii)　$\dfrac{x-a}{c}=u$, $\dfrac{y-b}{d}=v$ (a, b, c, d は任意定数) によって，(X, Y) を (U, V) に変換しても，相関係数の値は変らない．すなわち，

$$r_{xy}=r_{uv}$$

最小2乗法　散布図から X, Y 間に直線的関係のあることがわかったとき，これらのデータに直線をあてはめる問題が考えられる．これは直線を $y=a+bx$ とするとき，

$$\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

を最小にするような a, b を求めるという考えで，直線を定める方法が使われる．このような方法を**最小2乗法**といい，この方法で求めた直線を**最小2乗直線**，または Y の X への**回帰直線**という．

順位相関係数　n 個の標本が2つの変量 X, Y についてそれぞれ順位がつけられているとき，これら2組の順位 $(x_1, x_2, \cdots, x_n)$ と $(y_1, y_2, \cdots, y_n)$ の間の相関係数は

$$r_S=1-\frac{6\sum d_i^2}{n(n^2-1)} \qquad (d_i=x_i-y_i)$$

となる．これを**スピアマンの順位相関係数**という．

10-2　相関係数の検定

母集団相関係数を ρ，標本の大きさを n，標本相関係数を r とする．

$\rho=0$ の検定　$H_0: \rho=0$

$H_1: \rho\neq 0$　（両側検定）

の検定は，検定統計量

$$t=\sqrt{\frac{r^2(n-2)}{1-r^2}}$$

が，H_0 の下で自由度 $n-2$ の t 分布に従うことから，有意水準 α％の棄却域は

$$|t|>t_{\frac{\alpha}{2}}(n-2)$$

で与えられる．

$\rho=\rho_0$ の検定（大標本）　$H_0: \rho=\rho_0$

$H_1: \rho\neq\rho_0$　（両側検定）

の検定は，

$$z=\frac{1}{2}\log\frac{1+r}{1-r}, \qquad \mu=\frac{1}{2}\log\frac{1+\rho}{1-\rho}$$

とおくとき，検定統計量 z が H_0 の下で近似的に $N\left(\mu, \frac{1}{n-3}\right)$ に従うことから，有意水準 α％の棄却域は

$$\frac{|z-\mu|}{\frac{1}{\sqrt{n-3}}} \geqq z_{\frac{\alpha}{2}}$$

で与えられる．

回帰直線　$y=a+bx$　または　$y-\bar{y}=b(x-\bar{x})$

ここで，$\bar{x}$，$\bar{y}$ は X，Y の平均で，$\bar{x}=\frac{1}{n}\sum x_i$，$\bar{y}=\frac{1}{n}\sum y_i$．また b，a は

$$b=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{n\sum x_iy_i-(\sum x_i)(\sum y_i)}{n\sum x_i^2-(\sum x_i)^2}$$

$$a=\bar{y}-b\bar{x}$$

である．この直線を **Y の X への回帰直線**といい，a, b を**回帰係数**という．

10-3 線形回帰

2変数 X，Y 間の構造に関して線形回帰モデルとよばれる確率モデルを仮定し，モデルの式に含まれている母数の推定や検定の問題を考える．

線形回帰モデル　$x_1, x_2, \cdots, x_n$ を X の選ばれた値，$y_1, y_2, \cdots, y_n$ をこれら X の値に対する観測値とするとき，

$$y_i=\alpha+\beta x_i+\varepsilon_i \qquad (i=1, 2, \cdots, n)$$

と書ける．ここで，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ は互いに独立で同一の正規分布 $N(0, \sigma^2)$ に従う誤差変量である．このモデルを**線形回帰モデル**という．このとき，X を**独立変数**，Y を**従属変数**という．

線形回帰モデルの下において，種々の推測問題が考えられる．

誤差分散 σ^2 の推定値　σ^2 の不偏推定値は

$$s_e{}^2=\frac{1}{n-2}\sum(y_i-y_i')^2 \qquad (y_i'=\bar{y}+b(x_i-\bar{x}))$$

$$=\frac{1}{n-2}\{\sum(y_i-\bar{y})^2-b\sum(x_i-\bar{x})(y_i-\bar{y})\}$$

$$=\frac{1}{n-2}\{\sum y_i{}^2-a\sum y_i-b\sum x_iy_i\}$$

β の信頼限界　母集団回帰係数 β の $100(1-\alpha)$%信頼限界は

$$b\pm t_{\frac{\alpha}{2}}(n-2)\frac{s_e}{\sqrt{\sum(x_i-\bar{x})^2}}$$

$\beta=0$ の検定　$H_0: \beta=0$

$H_1: \beta\neq0$　（両側検定）

の検定は，検定統計量

$$t=\frac{b}{\dfrac{s_e}{\sqrt{\sum(x_i-\bar{x})^2}}}$$

が H_0 の下で自由度 $n-2$ の t 分布に従うことから，棄却域

$$|t|>t_{\frac{\alpha}{2}}(n-2)$$

で与えられる．

特定の x の値に対する Y の母平均の信頼限界　$X=x_0$ に対する Y の母平均の $100(1-\alpha)$％信頼限界は，

$$a+bx_0\pm t_{\frac{\alpha}{2}}(n-2)\,s_e\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

で与えられる．

X の Y への回帰直線　この場合は，Y を独立変数，X を従属変数と考える．

$$x=a'+b'y \quad \text{または} \quad x-\bar{x}=b'(y-\bar{y})$$

ここで x，y は X，Y の平均，

$$b'=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(y_i-\bar{y})^2}=\frac{n\sum x_iy_i-(\sum x_i)(\sum y_i)}{n\sum y_i{}^2-(\sum y_i)^2}, \qquad a'=\bar{x}-b'\bar{y}$$

この直線を X の Y への**回帰直線**といい，a', b' を**回帰係数**という．

例　題

例題 1　**(相関係数)**

次に示す 10 人の学生の身長 (X) と体重 (Y) のデータから，相関係数を求めよ．

身長 (x cm)	166	172	182	183	174	152	170	173	160	165
体重 (y kg)	60.5	62.9	65.0	64.8	59.2	56.9	61.0	61.9	59.3	58.3

解　計算を簡略化するため，$u=x-170$，$v=y-60$ とおく．

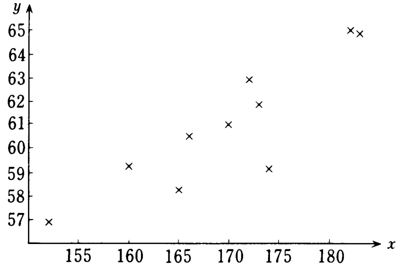
x	y	$u=x-170$	$v=y-60$	u^2	v^2	uv
166	60.5	-4	0.5	16	0.25	-2.0
172	62.9	2	2.9	4	8.41	5.8
182	65.0	12	5.0	144	25.00	60.0
183	64.8	13	4.8	169	23.04	62.4
174	59.2	4	-0.8	16	0.64	-3.2
152	56.9	-18	-3.1	324	9.61	55.8
170	61.0	0	1.0	0	1.00	0
173	61.9	3	1.9	9	3.61	5.7
160	59.3	-10	-0.7	100	0.49	7.0
165	58.3	-5	-1.7	25	2.89	8.5
計		-3	9.8	807	74.94	200

$n=10$, $\sum u_i=-3$, $\sum v_i=9.8$, $\sum u_i{}^2=807$, $\sum v_i{}^2=74.94$, $\sum u_iv_i=200$ を r の計算式に代入すれば

$$r_{xy}=r_{uv}=\frac{n\sum u_iv_i-(\sum u_i)(\sum v_i)}{\sqrt{\{n\sum u_i{}^2-(\sum u_i)^2\}\{n\sum v_i{}^2-(\sum v_i)^2\}}}$$

$$=\frac{10\times 200-(-3)\times 9.8}{\sqrt{(10\times 807-(-3)^2)(10\times 74.94-(9.8)^2)}}=\mathbf{0.88}$$

散布図を示すと



例題 2 （特別な散布図と相関係数）

次のデータに対して，相関係数 r を求めよ．それぞれについて，散布図をかき，得られた r の値について論評せよ．

（a）

X	1	2	3	4	5
Y	2	4	6	8	10

(b)

X	1	1	−1	−1
Y	1	−1	1	−1

(c)

X	−3	−2	−1	0	1	2	3
Y	8	3	0	−1	0	3	8

解

(a)　$n=5,\ \sum x_i=15,\ \sum y_i=30,\ \sum x_i^2=55,\ \sum y_i^2=220,\ \sum x_i y_i=110.$

$$r=\frac{n\sum x_i y_i-(\sum x_i)(\sum y_i)}{\sqrt{\{n\sum x_i^2-(\sum x_i)^2\}\{n\sum y_i^2-(\sum y_i^2)\}}}$$

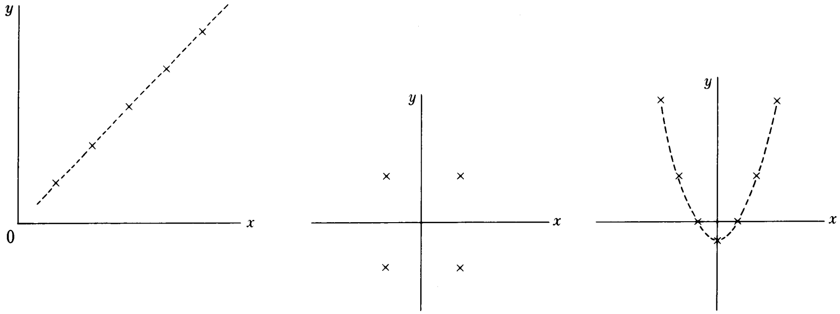
$$=\frac{5\times110-15\times30}{\sqrt{(5\times55-15^2)(5\times220-30^2)}}=\mathbf{1}$$

(b)　$n=4,\ \sum x_i=0,\ \sum y_i=0,\ \sum x_i^2=4,\ \sum y_i^2=4,\ \sum x_i y_i=0.$

$$r=\frac{4\times0-0\times0}{\sqrt{(4\times4-0^2)(4\times4-0^2)}}=\mathbf{0}$$

(c)　$n=7,\ \sum x_i=0,\ \sum y_i=-1,\ \sum x_i^2=28,\ \sum y_i^2=147,\ \sum x_i y_i=0.$

$$r=\frac{7\times0-0\times(-1)}{\sqrt{(7\times28-0^2)(7\times147-(-1)^2)}}=\mathbf{0}$$



(a)　　(b)　　(c)

(a)は各観測点が直線 $y=2x$ の上にのっているので，$r=+1$ になるのは当然.

(b)は，4つの点が原点に関して対称な位置にあって，直線的関係は全くないので，$r=0$ となっている.

(c)は，7個の点が放物線 $y=x^2-1$ の上にのっており，直線関係以外の放物線的関係なので，直線的傾向しか測り得ない相関係数は0となる.

例題 3 **(最小 2 乗法)**

ある計算アルゴリズムは，変数の個数 x と計算時間 y との間に

$$y=a+b\log_{10}x$$

の関係があるという．実際のデータは

個数 x	10	50	100	500	1000
時間 y	7.2	8.5	9.1	10.3	11.0

このとき，a と b の値を最小 2 乗法で求めよ．また散布図の上に求めた直線をかき入れよ．

解　いま $X=\log_{10}x$，$Y=y$ とおき，

$$Y=a+bX$$

の形にして，最小 2 乗法で，a，b の値を求める．

	X	Y	XY	X^2
	1	7.2	7.2	1
	1.7	8.5	14.45	2.89
	2	9.1	18.2	4
	2.7	10.3	27.81	7.29
	3	11.0	33.0	9
計	10.4	46.1	100.66	24.18

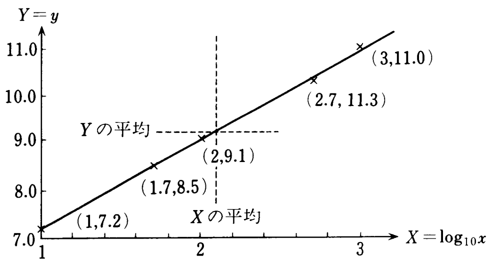
$$\overline{X}=\frac{10.4}{5}=2.08,\qquad \overline{Y}=\frac{46.1}{5}=9.22$$

$$b=\frac{n\sum X_iY_i-(\sum X_i)(\sum Y_i)}{n\sum X_i^2-(\sum X_i)^2}=\frac{5\times100.66-10.4\times46.1}{5\times24.18-(10.4)^2}=\mathbf{1.87}$$

$$a=\overline{Y}-b\overline{X}$$
$$=9.22-1.87\times2.08$$
$$=\mathbf{5.33}$$

よって，Y の X への回帰直線は

$$Y=5.33+1.87X$$



例題 4（相関係数，回帰直線の計算）

10人の生徒について，身長とひじの長さを測定して次のデータを得た（単位：cm）．

身長 (x)	152	145	164	153	132	148	149	138	142	150
ひじの長さ (y)	22	21	25	23	20	21	22	21	22	24

身長とひじの長さの相関係数を求めよ．また，Y の X への回帰直線を求めよ．身長が 150 cm の生徒のひじの長さはどのくらいか．

解　$x-150=u$, $y-20=v$ とおき，

$$(x \text{ と } y \text{ の相関係数})=(u \text{ と } v \text{ の相関係数})$$

$$(y \text{ の } x \text{ への回帰係数})=(v \text{ の } u \text{ への回帰係数})$$

を利用して，計算する．

x	y	u	v	u^2	v^2	uv
152	22	2	2	4	4	4
145	21	-5	1	25	1	-5
164	25	14	5	196	25	70
153	23	3	3	9	9	9
132	20	-18	0	324	0	0
148	21	-2	1	4	1	-2
149	22	-1	2	1	4	-2
138	21	-12	1	144	1	-12
142	22	-8	2	64	4	-16
150	24	0	4	0	16	0
計		-27	21	771	65	46

$$n=10,\quad \bar{u}=-2.7,\quad \bar{v}=2.1$$

$$\therefore\quad \bar{x}=\bar{u}+150=147.3$$

$$\bar{y}=\bar{v}+20=22.1$$

相関係数は

$$r_{xy}=r_{uv}=\frac{n\sum u_iv_i-(\sum u_i)(\sum v_i)}{\sqrt{\{n\sum u_i{}^2-(\sum u_i)^2\}}\sqrt{\{n\sum v_i{}^2-(\sum v_i)^2\}}}$$

$$=\frac{10\times 46-(-27)\times 21}{\sqrt{\{10\times 771-(-27)^2\}}\sqrt{\{10\times 65-(21)^2\}}}=\mathbf{0.85}$$

回帰係数は

$$b=\frac{n\sum u_iv_i-(\sum u_i)(\sum v_i)}{n\sum u_i^2-(\sum u_i)^2}=\frac{10\times 46-(-27)\times 21}{10\times 771-(-27)^2}=\mathbf{0.147}$$

よって，Y の X への回帰直線は

$$y-22.1=0.147(x-147.3)$$

$$\therefore\quad \boldsymbol{y=0.45+0.147x}$$

$x=150$ に対する y の値は

$$y=0.45+0.147\times 150=\mathbf{22.5}$$

例題 5　（順位相関係数）

（a）　出品された 8 個の作品を 2 人の審査員が独立に評価して，次の順位を与えた．スピアマンの順位相関係数を求めよ．

作品の番号	1	2	3	4	5	6	7	8
審査員 A	7	4	3	1	6	8	5	2
審査員 B	5	8	1	6	2	7	3	4

（b）　6 人の学生の数学と英語の成績が 5 段階評価で次のように与えられた．数学の成績と英語の成績のスピアマン順位相関係数を求めよ．

学生の番号	1	2	3	4	5	6
数学（X）	A	B	B	E	D	C
英語（Y）	C	B	A	E	E	B

解　（a）

A	B	$d=A-B$	d^2
7	5	2	4
4	8	-4	16
3	1	2	4
1	6	-5	25
6	2	4	16
8	7	1	1
5	3	2	4
2	4	-2	4
計		0	74

順位相関係数の公式から

$$r_S=1-\frac{6\sum d_i{}^2}{n(n^2-1)}=1-\frac{6\times 74}{8(8^2-1)}=\mathbf{0.12}$$

（b）　数学では，学生番号の 2 番と 3 番が同順位なので，平均順位 2.5 を与え，英語では，2 番と 6 番および 4 番と 5 番が同順位なので，前者には平均順位 2.5 を与え，後者には平均順位 5.5 を与える．

X	Y	$d=X-Y$	d^2
1	4	-3	9
2.5	2.5	0	0
2.5	1	1.5	2.25
6	5.5	0.5	0.25
5	5.5	-0.5	0.25
4	2.5	1.5	2.25
計		0	14.0

$$r_S=1-\frac{6\times 14.0}{6(6^2-1)}=\mathbf{0.6}$$

例題 6　**（相関係数の検定）**

A 地区のどぶネズミの体重と尾の長さの相関係数は，これまでの観測から，$\rho=0.75$ であることが知られている．いま B 地区の 12 匹のどぶネズミの体重と尾の長さの標本相関係数を求めたら，$r=0.62$ となった．この地区のどぶネズミの体重と尾の長さの相関係数は A 地区のそれと同じであるといえるか．

解　帰無仮説を $H_0:\rho=0.75$，対立仮説を $H_1:\rho\neq 0.75$ とし，有意水準 $\alpha=0.05$ の両側検定を行う．検定統計量

$$z=\frac{1}{2}\log_e\frac{1+r}{1-r}\qquad (r\text{ は標本相関係数})$$

は，H_0 の下で $z\sim N\left(\frac{1}{2}\log_e\frac{1+\rho}{1-\rho},\ \frac{1}{n-3}\right)$ であるから，$\alpha=0.05$ の棄却域は

$$\frac{\left|z-\frac{1}{2}\log_e\frac{1+\rho}{1-\rho}\right|}{\frac{1}{\sqrt{n-3}}}>1.96$$

$n=12, \rho=0.75$ を代入すれば，この棄却域は

$$z<0.32 \quad \text{または} \quad z>1.63$$

$r=0.62$ に対する $z=\dfrac{1}{2}\log_e\dfrac{1+0.62}{1-0.62}=0.725$ は棄却域に入らないから，H_0 は棄却されない．したがって B 地区の相関係数は，A 地区と比べて**差があるとはいえない**．

例題 7　**（予測値の信頼限界）**

スプリングの変位を測定して次のデータを得た．

荷重（x）	0	10	20	30	40
延び（y）	18.2	22.3	27.0	31.3	34.2

（a）　Y の X への回帰直線を求めよ．この直線を散布図上にかき入れよ．

（b）　$x=35$ のとき，y の値に対する 95％信頼区間を求めよ．

解　データから

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	0	18.2	-20	-8.4	400	70.56	168
	10	22.3	-10	-4.3	100	18.49	43
	20	27.0	0	0.4	0	0.16	0
	30	31.3	10	4.7	100	22.09	47
	40	34.2	20	7.6	400	57.76	152
計	100	133	0	0	1000	169.06	410

（a）　$n=5$，$\bar{x}=20$，$\bar{y}=26.6$，$\sum(x_i-\bar{x})^2=1000$，$\sum(y_i-\bar{y})^2=169.06$，$\sum(x_i-\bar{x})(y_i-\bar{y})=410$

であるから回帰係数は

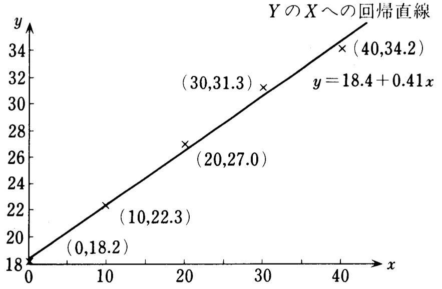
$$b=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{410}{1000}=0.41$$

よって Y の X への回帰直線は

$$y=\bar{y}+b(x-\bar{x})=26.6+0.41(x-20)$$

すなわち，

$$\boldsymbol{y=18.4+0.41x}$$



(b)　σ^2 の推定値 $s_e{}^2$ は

$$s_e{}^2=\frac{1}{n-2}\{\sum(y_i-\bar{y})^2-b\sum(x_i-\bar{x})(y_i-\bar{y})\}$$

$$=\frac{1}{3}\{169.06-0.41\times 410\}=0.32$$

ゆえに，$s_e=0.566$.

$x=35$ に対する y の予測値の 95%信頼限界は，$n=5$, $x_0=35$, $t_{0.025}(3)=3.182$ より

$$a+bx_0\pm t_{\frac{\alpha}{2}}(n-2)s_e\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

$$=18.4+0.41\times 35\pm 3.182\times 0.566\sqrt{\frac{1}{5}+\frac{(35-20)^2}{1000}}$$

$$=32.75\pm 1.17$$

$$=31.58,\ 33.92$$

よって求める信頼区間は **(31.58, 33.92)**.

例題 8　(最小 2 乗法)

ある微生物の成長モデルは時間 x と大きさ y の関係が

$$y=Ae^{Bx}$$

であるという．いま x と y について 5 個の観測値

x	1	2	3	4	5
y	5.5	14.9	39.9	110.2	300.1

を得たとき，A, B の値を最小 2 乗法で求めよ．

解　モデルの式で両辺の対数をとれば，$\log y=\log A+Bx$ となる．$Y=\log y$，$X=x$，$a=\log A$，$b=B$ とすれば，1次式

$$Y=a+bX$$

を得る．最小2乗法によって a と b を推定する．データから

$X=x$	1	2	3	4	5
$Y=\log y$	1.70	2.70	3.69	4.70	5.70

であるから，

$$n=5,\quad \sum X_i=15,\quad \sum Y_i=18.49,\quad \sum X_iY_i=65.47,\quad \sum X_i^2=55$$

したがって

$$b=\frac{5\times65.47-15\times18.49}{5\times55-15^2}=1.0$$

$$a=\overline{Y}-b\overline{X}=3.7-1.0\times3=0.7$$

よって求める1次式は，$Y=0.7+X$．

$$A=e^{0.7}=\mathbf{2.0},\quad B=b=\mathbf{1.0}$$

ゆえに，この成長モデルの式は

$$y=2e^x$$

例題 9　(2つの回帰直線)

8人の男子学生の身長(X)と右足(Y)の大きさを測って，次の値を得た．

身長 (x cm)	174	156	165	178	164	162	166	171
右足の大きさ (y cm)	24.5	22.5	25.0	26.5	23.0	24.0	25.0	26.0

このデータより，Y の X への回帰直線と X の Y への回帰直線を求めよ．

適切な回帰直線によって次を求めよ．

(a)　身長が 168 cm の男子学生の右足の大きさ．

(b)　右足の大きさが 25.5 cm の男子学生の身長．

解　計算の簡略化のため，$x-165=u$，$y-25=v$ とおく．両方の回帰直線を求めるには $\sum u_i$，$\sum v_i$，$\sum u_i^2$，$\sum v_i^2$，$\sum u_iv_i$ の値が必要である．

x	y	u	v	u^2	v^2	uv
174	24.5	9	-0.5	81	0.25	-4.5
156	22.5	-9	-2.5	81	6.25	22.5
165	25.0	0	0	0	0	0
178	26.5	13	1.5	169	2.25	19.5
164	23.0	-1	-2.0	1	4.0	2
162	24.0	-3	-1.0	9	1.0	3
166	25.0	1	0	1	0	0
171	26.0	6	1.0	36	1.0	6
計		16	-3.5	378	14.75	48.5

よって，$n=8$, $\sum u_i=16$, $\sum v_i=-3.5$, $\sum u_i^2=378$, $\sum v_i^2=14.75$,

$\sum u_iv_i=48.5$

(x, y) から (u, v) へ変換しても回帰係数は変らないから，Y の X への回帰係数を b, X の Y への回帰係数を b' とすれば

$$b=\frac{n\sum u_iv_i-(\sum u_i)(\sum v_i)}{n\sum u_i^2-(\sum u_i)^2}=\frac{8\times48.5-16\times(-3.5)}{8\times378-16^2}=0.16$$

$$b'=\frac{n\sum u_iv_i-(\sum u_i)(\sum v_i)}{n\sum v_i^2-(\sum v_i)^2}=\frac{8\times48.5-16\times(-3.5)}{8\times14.75-(-3.5)^2}=4.20$$

$$\bar{x}=165+\bar{u}=165+\frac{16}{8}=167$$

$$\bar{y}=25+\bar{v}=25+\frac{3.5}{8}=24.5625$$

よって，Y の X への回帰直線は

$$y-24.56=0.16(x-167)$$

$$\boldsymbol{y=0.16x-2.16} \qquad (1)$$

同様にして，X の Y への回帰直線は

$$\boldsymbol{x=4.20\,y+63.8} \qquad (2)$$

（a）(1)より，$x=168$ に対する y の値は

$$y=0.16\times168-2.16=\mathbf{24.72}$$

（b）(2)より，$y=25.5$ に対する x の値は

$$x=4.2\times25.5+63.8=\mathbf{170.9}$$

10 章の問題

10.1　次の表は 10 人の生徒の化学と物理の得点を示す．化学と物理の両得点の間の相関係数 r と順位相関係数 r_S を求めよ．

	生徒									
	A	B	C	D	E	F	G	H	I	J
化学 (x)	83	91	79	77	71	68	43	46	30	39
物理 (y)	77	89	95	71	59	55	59	43	35	48

10.2　2 つの離散変量 X と Y は，それぞれ 2 通りの可能な値 $\{p, q\}$ と $\{r, s\}$ をとるとする．いま，(X, Y) についての n 個のデータが，次のように 2×2 分割表で与えられるとき，X と Y の相関係数を求めよ．

		x の値		計
		p	q	
y の値	r	a	b	$a+b$
	s	c	d	$c+d$
	計	$a+c$	$b+d$	$n=a+b+c+d$

10.3　A, B 2 人の OL に 5 種類のオーディコロンの試作品を与えて，100 点満点で点数をつけてもらった．

	①	②	③	④	⑤
A	50	80	75	70	60
B	60	100	100	80	50

この 2 人の好みは，かなり近いといえるだろうか．

10.4　2 変量のデータ (x_i, y_i), $i=1, 2, \cdots, n$ が，順位データであるときの相関係数はスピアマンの順位相関係数

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

になることを示せ．ただし $d_i = x_i - y_i$ とする．

10.5 美人コンテストで 3 人の審査員が 8 人の出場者に次の順位をつけた.

	出場者							
	A	B	C	D	E	F	G	H
審査員 1	2	5	6	7	8	1	4	3
審査員 2	5	1	4	3	6	2	7	8
審査員 3	8	5	6	4	7	3	2	1

各審査員の間の順位相関係数を求めよ.

あなたが主任審査員であるとすれば, このコンテストの第 1 位, 第 2 位, 第 3 位を誰にするか.

10.6 国語 (X) と英語 (Y) の試験を 10 人の生徒に行ったところ, 次のデータを得た.

生徒	1	2	3	4	5	6	7	8	9	10
国語	39	62	68	61	57	74	41	47	43	65
英語	43	71	67	67	70	89	46	69	58	76

(X, Y) は正規分布に従うとして, 帰無仮説 $H_0 : \rho = 0$ を対立仮説 $H_1 : \rho \neq 0$ に対して 1% 有意水準で検定せよ.

10.7 理科の実験で, 針金からバネばかりを作った. これを用いておもりをいろいろ変えたときのバネの長さを測定し, 次のデータを得た.

おもり (x g)	10	12	14	16	18	20
バネの長さ (y cm)	20.2	21.3	21.9	23.0	24.2	25.1

(a) Y の X への回帰直線を求めよ.

(b) 15 g のおもりをのせたら, バネの長さは何 cm になるだろうか.

10.8 ある化学実験での反応は, 温度の 1 次式であるという. いま 3 水準の温度で 2 回ずつ 6 回測定を行ったときの反応値を示す.