Корреляционный анализ

Грауэр Л.В.

Коэффициент корреляции

(ξ, η)

$\rho(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}}$

Если ξ и η независимы $\Rightarrow \quad \rho(\xi, \eta) = 0$

$|\rho| = 1 \quad \Rightarrow \quad \eta = a + b\xi$

$(X, Y)_{[n]} = (X_1, Y_1), \ldots, (X_n, Y_n)$

Выборочный коэффициент корреляции Пирсона

$$r_{X,Y} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{D_X^* D_Y^*}}$$

$$D_X^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad D_Y^* = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Шкалой отношений называют такую шкалу с непрерывным множеством числовых значений, в которой о двух сопоставляемых объектах можно сказать не только, одинаковы они или различны, не только, в каком из них признак выражен сильнее, но и во сколько раз сильнее этот признак выражен.

Случай совместного нормального распределения

Пусть $(\xi, \eta) \sim N(a, K)$

$$\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

ξ и η независимы $\Leftrightarrow \rho(\xi, \eta)$.

Приближенный доверительный интервал для ρ

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

$n > 50$

$EZ \approx \frac{1}{2}\ln\frac{1+\rho}{1-\rho}, \qquad DZ \approx \frac{1}{n-3}, \qquad Z \approx N(EZ, \sqrt{DZ})$

$$\frac{1}{2}\ln\frac{1+r}{1-r} - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} < \frac{1}{2}\ln\frac{1+\rho}{1-\rho} < \frac{1}{2}\ln\frac{1+r}{1-r} + \frac{u_{1-\alpha/2}}{\sqrt{n-3}}$$

$$\mathrm{th}\left(\frac{1}{2}\ln\frac{1+r}{1-r} - \frac{u_{1-\alpha/2}}{\sqrt{n-3}}\right) < \rho < \mathrm{th}\left(\frac{1}{2}\ln\frac{1+r}{1-r} + \frac{u_{1-\alpha/2}}{\sqrt{n-3}}\right)$$

Проверка гипотезы $H_0 : \rho = \rho_0$

$$Z = \frac{Arth(r_{X,Y}) - Arth(\rho_0)}{1/\sqrt{n-3}}$$

H_1	V_k
$\rho > \rho_0$	
$\rho < \rho_0$	
$\rho \neq \rho_0$	

Проверка гипотезы $H_0 : \rho = 0$

$$t = \frac{r_{X,Y}\sqrt{n-2}}{\sqrt{1 - r_{X,Y}^2}} \sim T(n-2)$$

H_1	V_k
$\rho > 0$	
$\rho < 0$	
$\rho \neq 0$	

Порядковые шкалы

Шкалы, в которых существенен лишь взаимный порядок, в котором следуют результаты измерений, а не их количественные значения, называют *порядковыми или ординальными шкалами*.

$O_1, \ldots, O_n$

A и B: (X_i, Y_i), $1 \leqslant i \leqslant n$.

Коэффициент ранговой корреляции Спирмена

Коэффициент ранговой корреляции Кенделла

Коэффициент ранговой корреляции Спирмена

Проранжируем наблюдения

$$\begin{array}{lcl} X_1, \ldots, X_n & \Rightarrow & r_1, \ldots, r_n \\ Y_1, \ldots, Y_n & \Rightarrow & s_1, \ldots, s_n \end{array}$$

$$r_S = \frac{\frac{1}{n}\sum_{i=1}^n (r_i - \bar{r}_i)(s_i - \bar{s}_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^n (r_i - \bar{r}_i)^2 \frac{1}{n}\sum_{i=1}^n (s_i - \bar{s}_i)^2}} = 1 - \frac{6\sum_{i=1}^n (s_i - r_i)^2}{n^3 - n}$$

$$\sqrt{n-1}\, r_S \approx N(0,1)$$

Пример

эксперт 1	a	b	d	e	c
эксперт 2	b	a	e	c	d

Проверка гипотезы о независимости признаков

H_0: признаки A и B взаимно независимы.

H_1	таблицы	аппроксимация
">0" связь	$(c_1, 1]$	$\{\sqrt{n-1} r_S > u_{1-\alpha}\}$
"<0" связь	$[-1, c_2)$	$\{\sqrt{n-1} r_S < u_{\alpha}\}$
связь	$[-1, c_3) \bigcup (c_4, 1]$	$\{\lvert\sqrt{n-1} r_S\rvert > u_{1-\alpha/2}\}$

Коэффициент ранговой корреляции Кенделла

$$X_1, \ldots, X_n \quad \Rightarrow \quad r_1, \ldots, r_n$$
$$Y_1, \ldots, Y_n \quad \Rightarrow \quad s_1, \ldots, s_n$$

$$r_K = \frac{C - D}{C + D} = \frac{2(C - D)}{n(n - 1)}$$

$$\sqrt{\frac{9n(n+1)}{2(2n+5)}} r_K \approx N(0, 1)$$

Пример

эксперт 1	a	b	d	e	c
эксперт 2	b	a	e	c	d

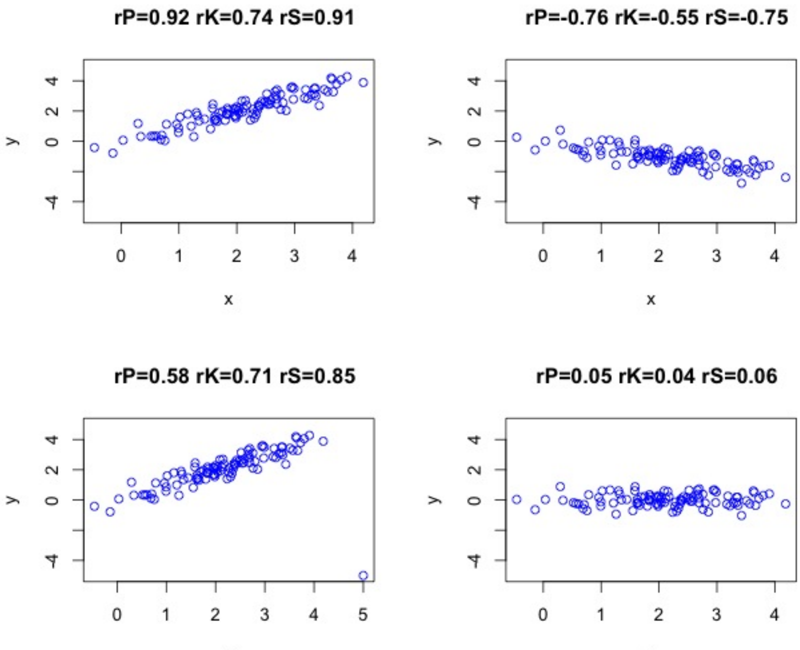
(a,b), (a,d), (a,e), (a,c), (b,d), (b,e), (b,c), (d,e), (d,c), (e,c)

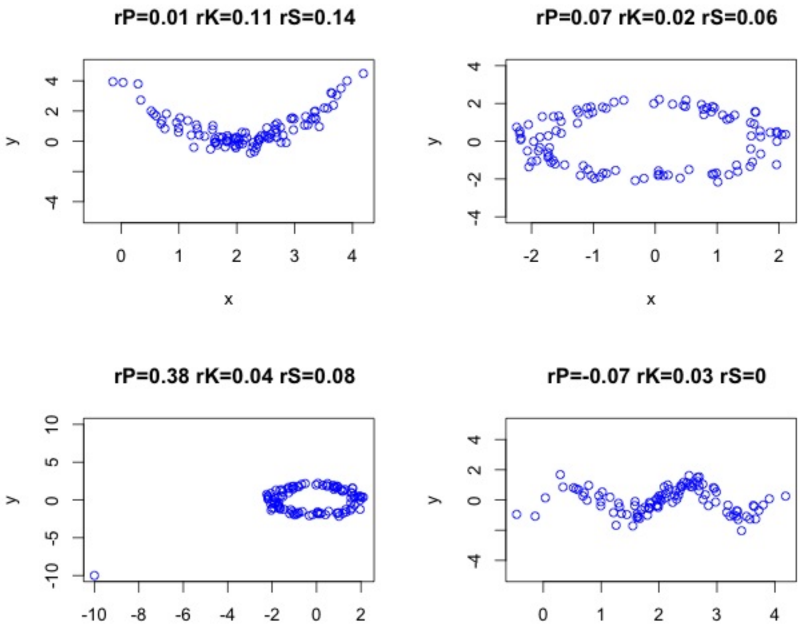
(b,a), (b,e), (b,c), (b,d), (a,e), (a,c), (a,d), (e,c), (e,d), (c,d)

Проверка гипотезы о независимости признаков

H_0: признаки A и B взаимно независимы.

H_1	таблицы	аппроксимация
“>0” связь	$(c_1, 1]$	$\left\{\sqrt{\frac{9n(n+1)}{2(2n+5)}}r_K > u_{1-\alpha}\right\}$
“<0” связь	$[-1, c_2)$	$\left\{\sqrt{\frac{9n(n+1)}{2(2n+5)}}r_K < u_{\alpha}\right\}$
связь	$[-1, c_3) \bigcup (c_4, 1]$	$\left\{\left\|\sqrt{\frac{9n(n+1)}{2(2n+5)}}r_K\right\| > u_{1-\alpha/2}\right\}$


rP=0.92 rK=0.74 rS=0.91
rP=-0.76 rK=-0.55 rS=-0.75
rP=0.58 rK=0.71 rS=0.85
rP=0.05 rK=0.04 rS=0.06
x
y


rP=0.01 rK=0.11 rS=0.14
rP=0.07 rK=0.02 rS=0.06
rP=0.38 rK=0.04 rS=0.08
rP=-0.07 rK=0.03 rS=0
x
y