

Корреляционный анализ

Основные задачи

- 1) Существует ли связь между факторами и откликом?
- 2) Оценить силу связи
- 3) Выявить факторы, оказывающие наибольшее влияние на отклик

Основные числовые характеристики

- 1) Выборочный коэффициент корреляции Пирсона
- 2) Ранговые коэффициенты корреляции
- 3) Выборочное корреляционное отношение
- 4) Множественный и частный коэффициенты корреляции

Шкалы измерений

Количественные признаки

- 1) Абсолютная шкала (начало отсчета+единицы измерения)
- 2) Шкала отношений (во сколько раз)
- 3) Интервальная шкала (на сколько единиц)

Качественные признаки

- 1) Порядковая шкала (ординальная)
- 2) Номинальная (шкала наименований)

Коэффициент корреляции Пирсона

ТВ

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

1) X, Y независимы \Rightarrow

$$\Rightarrow r_{xy} = 0$$

2) $|r_{xy}| \leq 1$

3) $Y = a + bX \Leftrightarrow$
 $\Leftrightarrow |r_{xy}| = 1$

МС

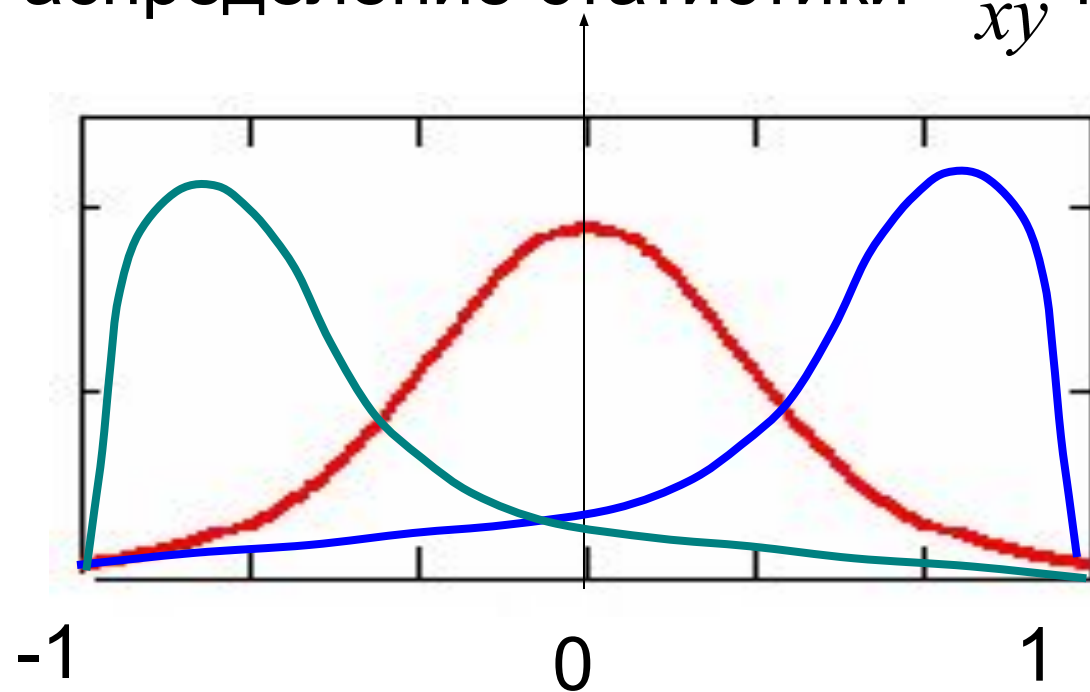
$$\tilde{r}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Для 2НГС оценка
асимптотически
несмещенная,
асимптотически
эффективная

$$\tilde{r}_{xy} = \frac{\frac{1}{n} \sum x_i y_i - \bar{x} \cdot \bar{y}}{S_X S_Y}$$

Коэффициент корреляции Пирсона

Распределение статистики \tilde{r}_{xy} при различных r_{xy}



$$\left. \begin{aligned} & \text{---} f_{\tilde{r}} \left(t \mid r_{xy} = 0 \right) \\ & \text{---} f_{\tilde{r}} \left(t \mid r_{xy} = 0.8 \right) \\ & \text{---} f_{\tilde{r}} \left(t \mid r_{xy} = -0.8 \right) \end{aligned} \right\}$$

Наибольшее значение статистики \tilde{r}_{xy} при значении

$r_{xy} = 0$ ($\alpha = 0.05$)

$n - 2$	5	10	20	30	50
\tilde{r}_{\max}	0.75	0.58	0.42	0.35	0.27

Коэффициент корреляции Пирсона

Задача 1 (о значимости связи)

(2НГС+некоррелированность) \equiv (независимость)

Гипотеза	$H_0 : r_{xy} = 0$
Требования	2НГС, $n \geq 50$
Оценки	$\bar{x}, \bar{y}, s_x, s_y, \tilde{r}_{xy}$
Статистика	$\frac{\tilde{r}_{xy} \sqrt{n-2}}{\sqrt{1-\tilde{r}_{xy}^2}}$
Распределение	$T_{(n-2)}_5$

Коэффициент корреляции Пирсона

Задача 1 (о значимости)

(2НГС+некоррелированность) \equiv (независимость)

Г	$H_0 : r_{xy} = 0$
Т	2НГС, $n \geq 50$
О	$\bar{x}, \bar{y}, s_x, s_y, \tilde{r}_{xy}$
С	$\frac{\tilde{r}_{xy} \sqrt{n-2}}{\sqrt{1-\tilde{r}_{xy}^2}}$
Р	$T_{(n-2)}$

Пример. $n=66, \tilde{r}_{xy} = 0.4$

1) $\alpha = 0,05$

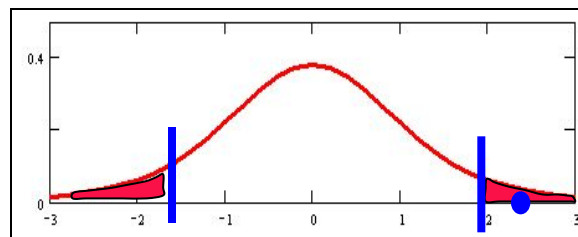
$$H_0 : r_{xy} = 0$$

$$H_1 : r_{xy} \neq 0$$

2) $t_{табл} = t(0.025; 60) = 2.0$

3) $t_{набл} = \frac{0,4 \cdot 8}{\sqrt{1-0,16}} \approx 3,6$

4) $|t_{набл}| \geq t_{табл} \Rightarrow H_1$



Коэффициент корреляции Спирмена

- а) ГС не является нормальной
- б) X, Y – в порядковой шкале
- в) $n < 50$

$$\tilde{r}_{xy} = \frac{\frac{1}{n} \sum x_i y_i - \bar{x} \cdot \bar{y}}{S_x S_y}$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

$$d_i = x'_i - y'_i$$

$$\Leftrightarrow \left\{ \begin{array}{l} x_i \otimes x'_i \\ y_i \otimes y'_i \end{array} \right\} \Leftrightarrow \downarrow$$

1) X, Y независимы $\Rightarrow r_s = 0$

2) $|r_s| \leq 1$

3) X, Y – монотонная связь

$$\Leftrightarrow |r_s| = 1$$

$$\left[\begin{array}{l} H_0: r_s = 0 \\ H_1: r_s \neq 0 \end{array} \right]$$

H_0 : связь незначима
 H_1 : связь значима

$$\frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \sim T_{(n-2)}$$

Коэффициент корреляции Спирмена

Пример. Y – вес монеты (грамм), X – возраст (лет)

x_i	5	9	14	17	23	31	35	42	46	50
y_i	2.82	2.85	2.80	2.80	2.79	2.78	2.77	2.79	2.75	2.72

x'_i	1	2	3	4	5	6	7	8	9	10
y'_i	9	10	7.5	7.5	5.5	4	3	5.5	2	1
d_i	-8	-8	-4.5	-3.5	-0.5	2	4	2.5	7	9

$$d_i = x'_i - y'_i \quad \sum d_i^2 = 317 \quad r_s = -0.92$$

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

Коэффициент корреляции Спирмена

Пример. Y – вес монеты (грамм), X – возраст (лет)

x_i	5	9	14	17	23	31	35	42	46	50
y_i	2.82	2.85	2.80	2.80	2.79	2.78	2.77	2.79	2.75	2.72

Вывод: между возрастом и весом монеты значимая монотонная связь (обратная)

H_0 : связь незначима

H_1 : связь значима

$$t_{\text{набл}} = \frac{-0.92 \cdot \sqrt{8}}{\sqrt{1 - 0,85}} \approx -6.17$$

$$t_{\text{табл}} = t(0.025; 8) = 2.31$$

$$r_S = -0.92$$

$$\frac{r_S \sqrt{n-2}}{\sqrt{1-r_S^2}} \sim T_{(n-2)}$$

Корреляционное отношение

Пример. Корреляционная таблица 1

X \ Y	3	5	7	n_x	\bar{y}_x
2	8	4	0	12	3.67
4	6	0	3	9	4.33
6	4	3	2	9	4.56
n_y	18	7	5	$n = 30$	$\bar{y} = 4.13$

$$\tilde{\eta}_{yx}^2 = \frac{s_{\text{межгр}}^2}{s_{\text{общ}}^2}$$

$$s_{\text{межгр}}^2 = \frac{1}{n} \sum n_x (\bar{y}_x - \bar{y})^2$$

$$s_{\text{общ}}^2 = \frac{1}{n} \sum n_y (y - \bar{y})^2$$

Корреляционное отношение

ТВ

$$\eta_{yx}^2 = \frac{\sigma^2(y|x)}{\sigma_y^2}$$

$$\tilde{\eta}_{yx} = \sqrt{\frac{s_{\text{межгр}}^2}{s_{\text{общ}}^2}}$$

МС

$$\tilde{\eta}_{yx}^2 = \frac{s_{\text{межгр}}^2}{s_{\text{общ}}^2}$$

Свойства корреляционного отношения

1) $\eta_{yx}^2 \neq \eta_{xy}^2$

2) $0 \leq \eta_{yx} \leq 1$

3) $\eta_{yx} \geq |r_{yx}|$

4) $\eta_{yx} = 0 \Rightarrow Y, X$ независимы

5) $\eta_{yx} = 1 \Rightarrow$ связь функциональная

6) $\eta_{yx} = |r_{yx}| \Rightarrow$ связь линейная

Корреляционное отношение

Пример. Корреляционная таблица 2

X \ Y	3	5	7	n_x	\bar{y}_x
2	6	3	0	9	3.67
4	0	5	2	7	5.57
6	0	5	9	14	6.29
n_y	6	13	11	$n = 30$	$\bar{y} = 5.18$

	$s_{\text{общ}}^2$	$s_{\text{межгр}}^2$	η_{yx}^2	r_{yx}^2
Пример 1	2.32	0.15	0.066	0.061
Пример 2	2.18	1.29	0.594	0.588

Корреляционное отношение

Задача 2. Проверка линейности

Гипотеза	$H_0: \eta_{yx} = r_{yx}$
Требования	2НГС, $n \geq 50$
Оценки	r_{yx}, η_{yx}
Статистика	$\frac{(\eta_{yx}^2 - r_{yx}^2)(n - r)}{(1 - \eta_{yx}^2)(r - 2)}$
Распределение	$F_{(r-2, n-r)}$

Корреляционное отношение

Задача 2. Проверка линейности

Г	$H_0: \eta_{yx} = r_{yx} $
Т	2НГС, $n \geq 50$
О	r_{yx}, η_{yx}
С	$\frac{(\eta_{yx}^2 - r_{yx}^2)(n - r)}{(1 - \eta_{yx}^2)(r - 2)}$
Р	$F(r-2, n-r)$

$$n = 30, r = 3$$

$$\eta_{yx}^2 = 0.594$$

$$r_{yx}^2 = 0.588$$

$$H_0: \eta_{yx} = |r_{yx}|$$

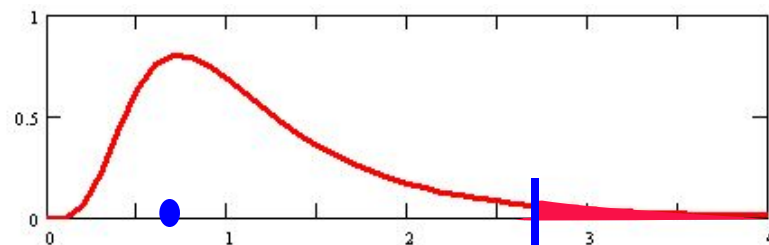
$$H_1: \eta_{yx} > |r_{yx}|$$

1) $\alpha = 0.05$

2) $F_{табл} = F(0.05; (2; 27)) = 3.35$

3) $F_{набл} = \frac{0,006 \cdot 27}{0.934 \cdot 1} \approx 0,173$

4) $F_{набл} < F_{табл} \Rightarrow H_0$



Многомерный корреляционный анализ

X_1, X_2, \dots, X_p

1) Парные коэффициенты корреляции

$$Q = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix}$$

2) Множественный коэффициент корреляции

$$R_{i,(1,\dots,i-1,i+1,\dots,p)} = \sqrt{1 - \frac{|Q|}{Q_{ii}}} \quad 0 \leq R \leq 1$$

Q_{ii} - алгебраическое дополнение к r_{ii}

3) Частные коэффициенты корреляции

$$\frac{R^2(n-p)}{(1-R^2)(p-1)} \sim F_{(p-1; n-p)}$$

$$-1 \leq r_{ij,(1\dots p)} \leq 1$$

$$r_{ij,(1,\dots,p)} = \frac{-Q_{ij}}{\sqrt{Q_{ii} \cdot Q_{jj}}}$$

Многомерный корреляционный анализ

X_1, X_2, X_3

1) Парные коэффициенты корреляции

2) Множественный коэффициент корреляции

$$Q = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$R_{3,(1,2)} = \sqrt{\frac{r_{13}^2 - 2r_{12}r_{13}r_{23} + r_{23}^2}{1 - r_{12}^2}}$$

$$0 \leq R_{3,(1,2)} \leq 1$$

3) Частные коэффициенты корреляции

$$r_{13(2)} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1 - r_{12}^2) \cdot (1 - r_{32}^2)}}$$

$$-1 \leq r_{13(2)} \leq 1$$

Многомерный корреляционный анализ

10 менеджеров оценивались 15 экспертами по 5-балльной шкале. Как связаны психологические характеристики между собой?

Результаты теста

Испытуемые п/п	тактичность	требовательность	критичность
1	70	18	36
2	60	17	29
3	70	22	40
4	46	10	12
5	58	16	31
6	69	18	32
7	32	9	13
8	62	18	35
9	46	15	30
10	62	22	36

Корреляционная матрица

	тактичность	требовательность	критичность
тактичность	1		
требовательность	0,86	1	
критичность	0,85	0,95	1

Частные коэффициенты корреляции

	тактичность	требовательность	критичность
тактичность	1		
требовательность	0,32	1	
критичность	0,21	0,81	1