

1. Корреляционный анализ

Интервальные оценки коэффициента корреляции

-более **адекватно** отражают реальное положение вещей.

-необходимо знать *закон распределения* значений *коэффициента корреляции*.

$$r - k_{\beta} \cdot \sigma_r < \rho < r + k_{\beta} \cdot \sigma_r$$

При **нормальном** распределении

$$M(r) = \rho + O\left(\frac{1}{n}\right), \quad \left(M(r) = r - \frac{r(1-r^2)}{2n} \right),$$

$$D(r) = \frac{(1-\rho^2)^2}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

1. Корреляционный анализ

Интервальная оценка

$$r - u_{\beta} \cdot \frac{(1 - r^2)}{\sqrt{n}} < \rho < r + u_{\beta} \cdot \frac{(1 - r^2)}{\sqrt{n}},$$

u_{β} – параметр функции *Лапласа* при заданной доверительной вероятности β

При малом числе n и значениях близких к ± 1 весьма **грубо** отражает реальность.

при $\rho = 0$: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ t -распределение *Стьюдента* с

$n - 2$ степенями свободы

1. Корреляционный анализ

Приближение *интервальной* оценки:

$$\left(r + \frac{r(1-r^2)}{2n} \right) - u_{q/2} \cdot \frac{1-r^2}{\sqrt{n}} < \rho < \left(r + \frac{r(1-r^2)}{2n} \right) + u_{q/2} \cdot \frac{1-r^2}{\sqrt{n}}$$

Достоинства, недостатки.

1. Корреляционный анализ

z -функция Фишера

$$r = th(z), \rightarrow z = \frac{1}{2} \cdot \ln \left(\frac{1+r}{1-r} \right)$$

Распределение z не зависит от значений ρ и n , при $n > 10$ быстро сходится к *нормальному закону* с параметрами

$$M(z) \approx \frac{1}{2} \cdot \ln \left(\frac{1+\rho}{1-\rho} \right) + \frac{\rho}{2(n-1)}, \quad D(z) \approx \frac{1}{n-3}.$$

1. Корреляционный анализ

Доверительный интервал $[z_1, z_2]$ для математического ожидания $M(z)$

$$z_{1,2} = \frac{1}{2} \cdot \ln \left(\frac{1+r}{1-r} \right) \mp \left(\frac{u_{q/2}}{\sqrt{n-3}} - \frac{r}{2(n-1)} \right) = \operatorname{arcth}(z) \mp \left(\frac{u_{q/2}}{\sqrt{n-3}} - \frac{r}{2(n-1)} \right)$$

истинное значение коэффициента корреляции ρ с доверительной вероятностью $\beta = 1 - q$ заключено в пределах

$$\operatorname{th}(z_1) < \rho < \operatorname{th}(z_2)$$

Варианты упрощения без смещения.

1. Корреляционный анализ

Значимость статистической связи

-сводится к проверке статистической значимости коэффициента корреляции.

Общий случай: проверяется нулевая гипотеза с какой либо альтернативной, например

$$H_0 : R^2 = 0,$$

$$H_1 : R^2 > 0$$

F-тест Фишера: общий разброс разлагается на объясненную и не объясненную составляющие

$$y^2 = k^2 + e^2.$$

Переход к дисперсиям.

1. Корреляционный анализ

F-статистика:

$$F = \frac{\left(\frac{k^2}{p} \right)}{\left(\frac{e^2}{m} \right)}$$

$p = n - 1$ – число объясняющих переменных (для парной регрессии 1); $m = n - k$ – число наблюдений без количества оцениваемых коэффициентов (для парной регрессии $n - 2$).

$F > F_{\beta; p, q}$ - нулевая гипотеза **отвергается**, R^2 значимо

1. Корреляционный анализ

Значимость *парного* выборочного коэффициента корреляции $p = 1$, $m = n - 2$:

$$F = \frac{k^2(n-2)}{e^2} = \frac{r_{12}^2 \sigma_y^2 (n-2)}{\sigma_y^2 (1-r_{12}^2)} = \frac{r_{12}^2 (n-2)}{(1-r_{12}^2)}$$

$p = 1$, то $F = t^2$:

$$\sqrt{F} = t = r_{12} \sqrt{\frac{(n-2)}{1-r^2}}$$

если $t < t_{\beta, q}$, нулевая гипотеза $H_0 : r = 0$ принимается с доверительной вероятностью β .

2. Дисперсионный анализ

- значимость** влияния факторов;
- выбор** наиболее важных факторов;
- оценка** их влияния.

Основная идея:

- разложение общей дисперсии случайной величины на независимые случайные характеризующие слагаемые,*
- сравнение этих дисперсий для оценки существенности влияния факторов на исследуемую величину.*

Виды дисперсионного анализа.

2. Дисперсионный анализ

Однофакторный дисперсионный анализ

-на результаты эксперимента действует только **один** фактор. Отслеживают процесс в m серий по n элементов в каждой:

Число элементов n Число серий m	1	2	...	j	...	n	$\bar{x}$
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1n}	$\bar{x}_{1j}$
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2n}	$\bar{x}_{2j}$
...	...	...	...	...	...	...	...
i	x_{i1}	x_{i2}	...		...	x_{in}	$\bar{x}_{ij}$
...	...	...	...	...	...	...	...
m	x_{m1}	x_{m2}	...	x_{mj}	...	x_{mn}	$\bar{x}_{mj}$

2. Дисперсионный анализ

Проверим гипотезу о равенстве средних β_i в серии с *нулевой гипотезой*

$$H_0: \beta_1 = \beta_2 = \dots = \beta_i = \dots = \beta_m$$

сравнением внутрисерийных и общей дисперсий по F -критерию.

-расхождение незначительно, то нулевая гипотеза **принимается**.

- расхождение значительно, значит значительно **действие** исследуемого фактора.

2. Дисперсионный анализ

Для этого:

- Найдем сумму квадратов отклонений всех элементов от общего среднего Q ;
- Разложим эту величину на *девиаты*: сумму квадратов отклонений между сериями (Q_1 - рассеиванием по факторам) и сумму квадратов отклонений внутри серии (Q_2 - остаточное рассеивание):

$$Q = Q_1 + Q_2 \qquad F = \frac{Q_1 / (m - 1)}{Q_2 / m(n - 1)}$$

Сила проявления **влияния** факторов: если $F < F_q$ - гипотеза об отсутствии достаточно сильного влияния фактора **принимается**