

# *Статистическое изучение взаимосвязей*

- 1. Понятие и виды статистических связей**
- 2. Методы оценки статистических связей между качественными признаками**
- 3. Методы оценки статистических связей между количественными признаками**
- 4. Понятие и методика регрессионного анализа**

В статистике преимущественно рассматривают следующие виды связей:

- **функциональная связь** или **полная корреляция** – связь, при которой каждому значению факторного признака соответствует строго определенное значение результативного признака.
- **стохастическая связь** – это связь, при которой одному значению факторного признака соответствует группа значений результативного признака;
- **корреляционная связь** – это связь, при которой с изменением значений факторного признака изменяются средние значения результативного признака;

# Виды связей

По числу взаимосвязанных признаков различают:

- **парные** связи, когда анализируется взаимосвязь только двух признаков: факторного и результативного;
- **множественные** связи, когда характеризуется влияние нескольких факторных признаков на один результативный;

По механизму взаимодействия различают:

- **непосредственные** связи, когда причина прямо влияет на следствие;
- **косвенные** связи, когда между причиной и следствием существуют промежуточные признаки

По **направлению** связи подразделяют на:

- **прямые** связи, когда значения факторного и результирующего признаков изменяются в одном направлении;
- **обратные** связи, когда их значения изменяются в разных направлениях;

По **аналитическому выражению** выделяют:

- **прямолинейные** связи, которые выражаются уравнением прямой линией;
- **криволинейные** связи, которые можно выразить уравнением параболы, гиперболы, полулогарифмической кривой и т.д.;

По степени тесноты связи её классифицируют по величине значений коэффициентов корреляции, представленным в таблице Чеддока

<b>Теснота связи</b>	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99	1,0
<b>Характер связи</b>	Слабая	Умеренная	Заметная	Тесная	Очень тесная	Функциональная

# Матрица взаимного распределения частот определения коэффициентов ассоциации и КОНТИНГЕНЦИИ

<b>1 признак 2 признак</b>	<b>Да</b>	<b>Нет</b>	<b>Итого</b>
<b>Да</b>	a	b	a+b
<b>Нет</b>	c	d	c+d
<b>Итого</b>	a+c	b+d	a+b+c+d

- Коэффициент **ассоциации** определяется по формуле:

$$K_A = \frac{ad - bc}{ad + bc}$$

- Коэффициент контингенции:

$$K_K = \frac{ad - bc}{\sqrt{(a + b)(b + d)(d + c)(a + c)}}$$

Зависимость между полом и фактом совершения покупки посетителями магазина

1 признак	М	Ж	Итого
2 признак			
Купил	24	32	56
Не купил	16	28	44
Итого	40	60	100



- Коэффициент взаимной сопряженности признаков Пирсона определяется по формуле:

$$K_{\Pi} = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}$$

- Коэффициент взаимной сопряженности признаков Чупрова:

$$K_{\text{Ч}} = \sqrt{\frac{\varphi^2}{(k_1 - 1)(k_2 - 1)}}$$

$\varphi^2$  - показатель взаимной сопряженности признаков, который рассчитывается на основе матрицы взаимного распределения частот ( $\varphi^2 = L_1 + L_2 + L_3 - 1$ ,  $L_i = \left( \sum \frac{s_{ij}^2}{m_i} \right) : n_j$ )

## Матрица взаимного распределения частот

	<b>1 гр.</b>	<b>2 гр.</b>	<b>3 гр.</b>	<b>Итого</b>
<b>1 гр.</b>	$s_{11}$	$s_{12}$	$s_{13}$	$n_1$
<b>2 гр.</b>	$s_{21}$	$s_{22}$	$s_{23}$	$n_2$
<b>3 гр.</b>	$s_{31}$	$s_{32}$	$s_{33}$	$n_3$
<b>Итого</b>	$m_1$	$m_2$	$m_3$	

# Зависимость между величиной магазина и формой обслуживания

	Самообслуживание	Традиционное	Итого
Мелкие магазины	12	45	57
Средние	19	10	29
Крупные	14	4	18
Итого	45	59	104

### 3. Методы оценки статистических связей между количественными признаками

- Коэффициент Фехнера:

$$K_{\phi} = \frac{\sum c - \sum n}{\sum c + \sum n} = \frac{\sum c - \sum n}{n}$$

- Коэффициент корреляции рангов Спирмена:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

# Взаимосвязь между фондовооруженностью и производительностью труда

Фондовооруженность, тыс. руб. $x$	Производительность, тыс. руб. $y$	$x - \bar{x}$	$y - \bar{y}$	$c/n$
2	3	-3	-3	C
5	6	0	0	C
3	4	-2	-2	C
7	6	2	0	C
2	4	-3	-2	C
6	8	1	2	C
4	6	-1	0	H
9	9	4	3	C
8	9	3	3	C
4	5	-1	-1	C

# Взаимосвязь между товарооборотом и уровнем издержек обращения в магазинах

Однодневный товарооборот, тыс. руб. $x$	Издержки в % к товарообороту $y$	Ранги		$d = R_x - R_y$	$d^2$
		$R_x$	$R_y$		
18	20,5	1	4	-3	9
23	23,4	2	6	-4	16
29	21,2	3	5	-2	4
45	18,9	4	2	2	4
78	19,2	5	3	2	4
93	17,5	6	1	5	25
Всего	-	-	-	-	62

# Формулы коэффициентов корреляции

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[ \sum x^2 - \frac{(\sum x)^2}{n} \right] \left[ \sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$

$$r = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sigma_x \sigma_y}$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Если определена форма корреляционной связи и коэффициент регрессии , то коэффициент корреляции можно рассчитать по формуле:

$$r = a_1 \frac{\sigma_x}{\sigma_y}$$



# Расчет коэффициента корреляции

Фондовооруженность, тыс. руб. $x$	Производительность, тыс. руб. $y$	$x^2$	$y^2$	$xy$
2	3	4	9	6
5	6	25	36	30
3	4	9	16	12
7	6	49	36	42
2	4	4	16	8
6	8	36	64	48
4	6	16	36	24
9	9	81	81	81
8	9	64	81	72
4	5	16	25	20
50	60	304	400	343

# Расчет коэффициента корреляции

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[ \sum x^2 - \frac{(\sum x)^2}{n} \right] \left[ \sum y^2 - \frac{(\sum y)^2}{n} \right]}} = \frac{343 - \frac{50 * 60}{10}}{\sqrt{\left[ 304 - \frac{50^2}{10} \right] \left[ 400 - \frac{60^2}{10} \right]}} = 0,925$$

Значимость линейного коэффициента корреляции проверяется на основе  $t$  – критерия Стьюдента.

$$t_p = \frac{|r|\sqrt{n}}{\sqrt{1-r^2}} = \frac{0,925\sqrt{10}}{\sqrt{1-0,925^2}} = \frac{2,925}{\sqrt{0,144}} = 7,7$$

Входными параметрами для отыскания табличного значения являются:  $\alpha$  (0.05; 0.01) и число степеней свободы  $d.f. = n - 2$ .

Если  $t_p > t_{\text{табл}}$ , то коэфф. корреляции статистически значим

7,7 > 2,3060 (при уровне значимости 0,05 и числе степеней свободы 8)

Формула множественного коэффициента  
корреляции:

$$R_{y_{x_1 x_2}} = \sqrt{\frac{r_{y_{x_1}}^2 + r_{y_{x_2}}^2 - 2r_{y_{x_1}} r_{y_{x_2}} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}}$$

# Уравнение линейной регрессии:

$$y_x = a_0 + a_1x$$

Параметры уравнения прямой определяются путем решения системы нормальных уравнений:

$$na_0 + a_1 \sum x = \sum y$$

$$a_0 \sum x + a_1 \sum x^2 = \sum xy$$

$$a_0 = \frac{\overline{xy} - \bar{x} * \bar{y}}{x^2 - (\bar{x})^2}$$

$$a_1 = \bar{y} - a_0 \bar{x}$$

## Расчет параметров уравнения регрессии

Фондовооруженность, тыс. руб. $x$	Производительность, тыс. руб. $y$	$x^2$	$xy$	$y_x$
2	3	4	6	3,61
5	6	25	30	6,0
3	4	9	12	4,41
7	6	49	42	7,59
2	4	4	8	3,61
6	8	36	48	6,80
4	6	16	24	5,20
9	9	81	81	9,19
8	9	64	72	8,38
4	5	16	20	5,20
50	60	304	343	60

1. Для определения параметров уравнения регрессии подставим в систему нормальных уравнений фактические данные из таблицы:

$$10a_0 + 50a_1 = 60$$

$$50a_0 + 304a_1 = 343$$

2. разделим каждый член первого уравнения на 10, а каждый член второго уравнения на 50:

$$a_0 + 5a_1 = 6$$

$$a_0 + 6,08a_1 = 6,86$$

3. вычтем из второго уравнения первое и получим:

$$1,08a_1 = 0,86 \quad \text{Отсюда} \quad a_1 = 0,796$$

4. подставим значение  $a_1$  в первое уравнение, получим:

$$a_0 = 2,02$$

$$y_x = 2,02 + 0,796x$$

Параметр  $a_0$  показывает усредненное влияние на результирующий признак неучтенных, т.е. не выделенных для исследования факторных признаков;

Параметр  $a_1$  – это **коэффициент регрессии**, который показывает, насколько изменяется значение результирующего признака при изменении факторного признака на единицу его собственного измерения.



$$t_{a_0} = \frac{\hat{a}_0}{S_{a_0}} = \frac{2,02}{0,282} = 7,2 \quad t_{a_1} = \frac{a_1}{S_{a_1}} = \frac{0,796}{0,123} = 6,47$$

Где  $S_{a_0}$  стандартная ошибка параметра  $a_0$

$$S_{a_0} = \frac{\sigma_y \sqrt{1-r^2}}{\sqrt{n-2}} = \frac{2,1\sqrt{1-0,925^2}}{\sqrt{10-2}} = \frac{0,797}{2,828} = 0,282$$

$S_{a_1}$  стандартная ошибка параметра  $a_1$

$$S_{a_1} = \frac{\sigma_y \sqrt{1-r^2}}{\sigma_x \sqrt{n-2}} = \frac{2,1\sqrt{1-0,925^2}}{2,3\sqrt{10-2}} = \frac{0,797}{6,504} = 0,123$$

- Фактические значения  $t$ -критерия сравниваются с табличными (с учетом уровня значимости  $\alpha$  и числа степеней свободы ( $d.f. = n - k - 1$ )). Параметры признаются статистически значимыми, т.е. сформированными под воздействием неслучайных факторов, если  $t_{\text{факт}} > t_{\text{табл}}$ .

Значимость уравнения в целом оценивается на основе  $F$ -критерия Фишера:

$$F = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{k} : \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - k - 1}$$

- где  $k$  – число степеней свободы факторной дисперсии, равное числу независимых переменных (признаков-факторов) в уравнении регрессии;
- $n-k-1$  - число степеней свободы остаточной дисперсии.

$$F = \frac{r_{yx}^2}{1 - r_{yx}^2} \frac{n - k - 1}{k} = \frac{0,925^2}{1 - 0,925^2} \frac{10 - 1 - 1}{1} = \frac{0,856}{1 - 0,856} * \frac{8}{1} = 47,6$$

Расчетное значение критерия сопоставляется с табличным (с учетом числа степеней свободы:  $d.f. = k$  и  $d.f. = n - k - 1$ ).

Если  $F_{\text{расч.}} \geq F_{\text{табл.}}$ , то делается вывод о статистической значимости уравнения в целом.

$$F_{\text{расч.}} \geq F_{\text{табл.}}$$
$$47,6 \geq 5,32$$

Уравнение параболы второго порядка:

$$y_x = a_0 + a_1x + a_2x^2$$

Система нормальных уравнений:

$$na_0 + a_1 \sum x + a_2 \sum x^2 = \sum y$$

$$a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy$$

$$a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum yx^2$$

Уравнение гиперболы:

$$y_x = a_0 + a_1 \frac{1}{x}$$

Система уравнений:

$$na_0 + a_1 \sum \frac{1}{x} = \sum y$$

$$a_0 \sum \frac{1}{x} + a_1 \sum \left( \frac{1}{x} \right)^2 = \sum y \frac{1}{x}$$

Замена переменных:  $\frac{1}{x} = x_1$

Система нормальных уравнений примет следующий вид:

$$na_0 + a_1 \sum x_1 = \sum y$$

$$a_0 \sum x_1 + a_1 \sum x_1^2 = \sum x_1 y$$

Уравнение множественной регрессии  
(линейное уравнение с двумя  
переменными):

$$\acute{o}_{\delta} = \acute{a}_0 + \acute{a}_1 \tilde{o}_1 + \acute{a}_2 x_2$$

Система нормальных уравнений:

$$\acute{a}_0 \sum 1 + \acute{a}_1 \sum \tilde{o}_1 + \acute{a}_2 \sum x_2 = \sum \acute{o}$$

$$\acute{a}_0 \sum \tilde{o}_1 + \acute{a}_1 \sum \tilde{o}_1^2 + \acute{a}_2 \sum \tilde{o}_1 x_2 = \sum \tilde{o}_1 \acute{o}$$

$$\acute{a}_0 \sum x_2 + \acute{a}_1 \sum \tilde{o}_1 x_2 + \acute{a}_2 \sum x_2^2 = \sum x_2 \acute{o}$$



# Матрица парных коэффициентов корреляции

Признак к	$y$	$x_1$	$x_2$	$x_3$	...	$x_k$
$y$	1	$r_{yx_1}$	$r_{yx_2}$	$r_{yx_3}$	...	$r_{yx_k}$
$x_1$		1	$r_{x_1x_2}$	$r_{x_1x_3}$	...	$r_{x_1x_k}$
$x_2$			1	$r_{x_2x_3}$	...	$r_{x_2x_k}$
$x_3$				1	...	$r_{x_3x_k}$
					1	
...						1
$x_k$						1
...						