

# Тема

Статистическое изучение  
взаимосвязи социально-  
экономических явлений

# Методы изучения взаимосвязи

1. Графический
2. Параллельных данных
3. Корреляционный анализ
4. Регрессионный анализ

# 1. по степени причинно-следственной определенности

- *Функциональная* — значение результивного признака «у» полностью определяется значением факторного признака «х».
- *Корреляционная* — значение результивного признака «у» в большей или меньшей степени определяется значением факторного признака «х»; при этом имеет место влияние прочих, неучтенных, факторов.

## 2. По направлению:

- Прямая
- Обратная

### 3. По аналитическому выражению:

- Линейная
- Нелинейная

## 4. По количеству факторных признаков:

- Парная
- Множественная

# Графический анализ взаимосвязи

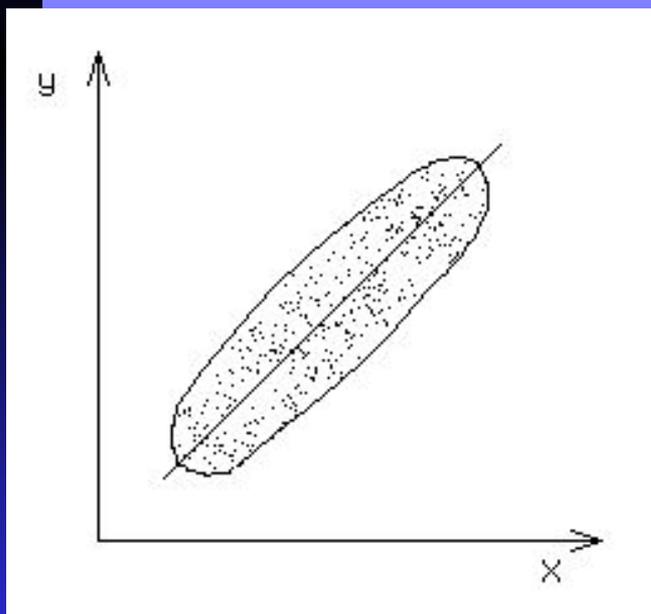
## Исходные данные для построения поля корреляции (точечной диаграммы)

№ единицы совокупности	Значения факторного признака	Значения результативного признака
$1$	$x_1$	$y_1$
$2$	$x_2$	$y_2$
$3$	$x_3$	$y_3$
$\bar{x}$	$\bar{x}$	$\bar{y}$
$n$	$x_n$	$y_n$

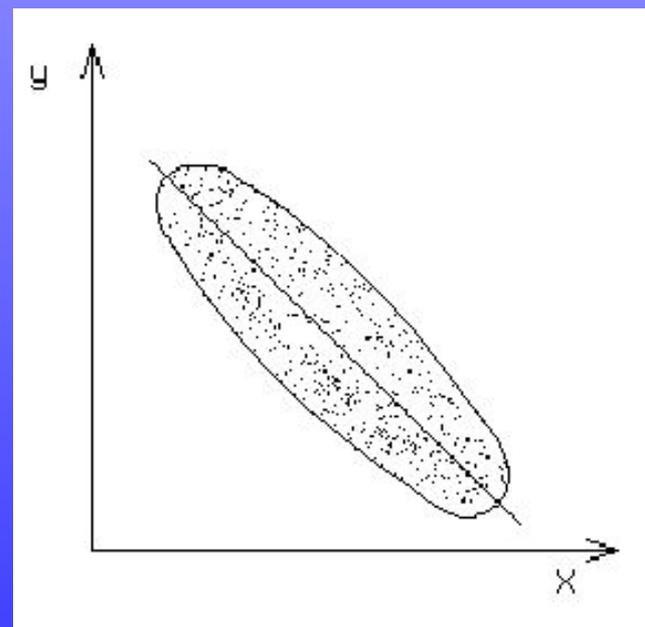
Каждая единица совокупности представлена на приведенных ниже диаграммах в виде отдельной точки. Совокупность точек формирует графический образ.

## Линейная зависимость

а) прямая

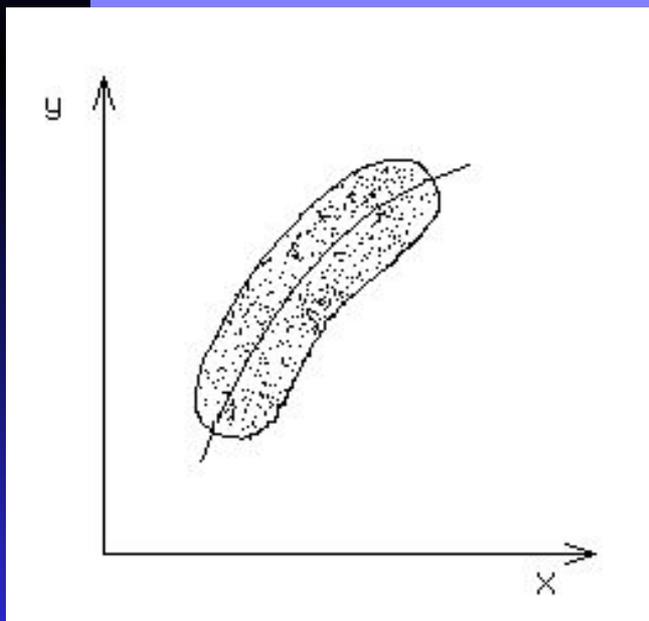


б) обратная

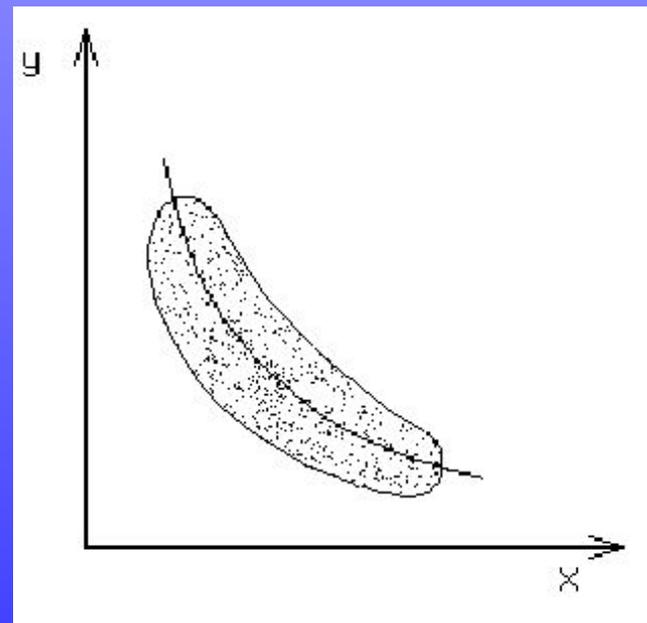


## Нелинейная зависимость

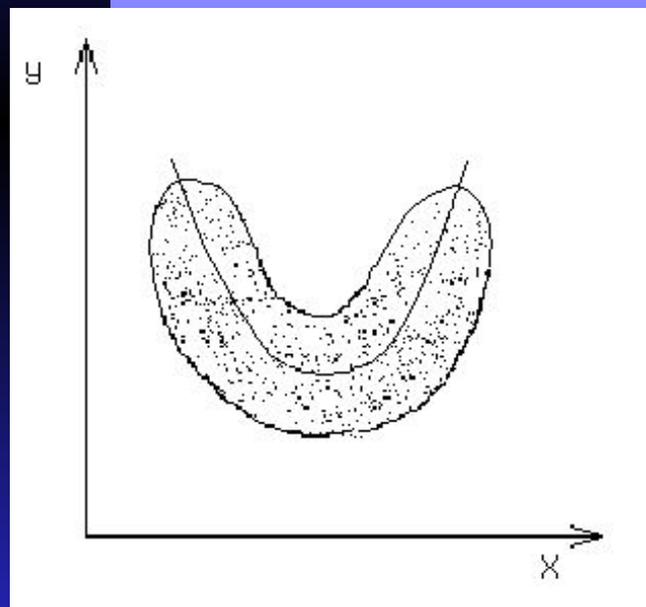
а) прямая



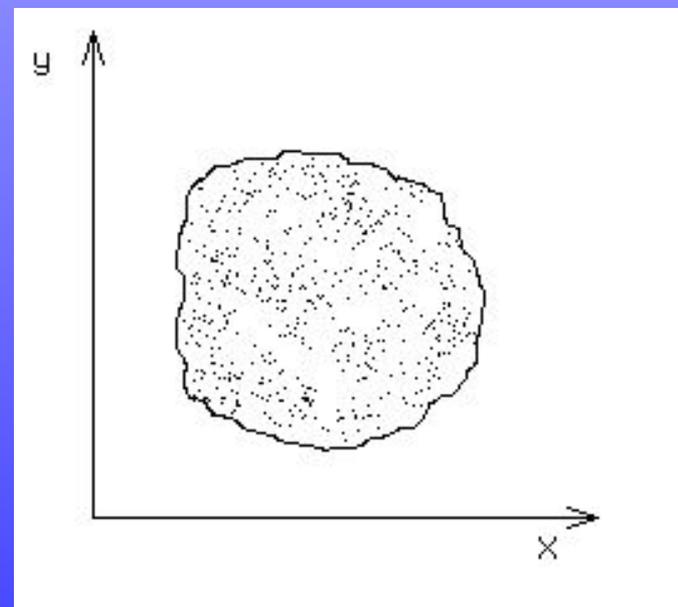
б) обратная



в) разнонаправленная



Отсутствие зависимости



# Парная линейная зависимость

$$\bar{y}_x = a_0 + a_1 x$$

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum yx \end{cases}$$

# Парный линейный коэффициент корреляции Пирсона

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$$

$$\overline{xy} = \frac{\sum xy}{n} \quad \bar{x} = \frac{\sum x}{n} \quad \bar{y} = \frac{\sum y}{n}$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} \quad \sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2}$$

$$-1 \leq r_{xy} \leq 1$$

$|r_{xy}| < 0,3$  - очень слабая связь,  
практически отсутствует

$0,3 < |r_{xy}| < 0,5$  - связь слабая

$0,5 < |r_{xy}| < 0,7$  - связь средняя

$|r_{xy}| > 0,7$  - связь тесная

Например по данным о стоимости основных фондов и объеме произведенной продукции оцените тесноту связи и постройте линейное уравнение взаимосвязи

Стоимость основных производственных фондов, млн.руб. $x$	Объем валовой продукции, млн.руб. $y$
1	20
2	25
3	31
4	31
5	40
6	56
7	52
8	60
9	60
10	70

$x$	$y$	$xy$	$x^2$	$y^2$	$\bar{y}_x$
1	20				
2	25				
3	31				
4	31				
5	40				
6	56				
7	52				
8	60				
9	60				
10	70				
$\Sigma=55$	$\Sigma=445$				

$x$	$y$	$xy$	$x^2$	$y^2$	$\bar{y}_x$
1	20	20			
2	25	50			
3	31	93			
4	31	124			
5	40	200			
6	56	336			
7	52	364			
8	60	480			
9	60	540			
10	70	700			
$\Sigma=55$	$\Sigma=445$	$\Sigma=2\ 907$			

$x$	$y$	$xy$	$x^2$	$y^2$	$\bar{y}_x$
1	20	20	1		
2	25	50	4		
3	31	93	9		
4	31	124	16		
5	40	200	25		
6	56	336	36		
7	52	364	49		
8	60	480	64		
9	60	540	81		
10	70	700	100		
$\Sigma=55$	$\Sigma=445$	$\Sigma=2\ 907$	$\Sigma=385$		

$x$	$y$	$xy$	$x^2$	$y^2$	$\bar{y}_x$
1	20	20	1	400	
2	25	50	4	625	
3	31	93	9	961	
4	31	124	16	961	
5	40	200	25	1 600	
6	56	336	36	3 136	
7	52	364	49	2 704	
8	60	480	64	3 600	
9	60	540	81	3 600	
10	70	700	100	4 900	
$\Sigma=55$	$\Sigma=445$	$\Sigma=2\,907$	$\Sigma=385$	$\Sigma=22\,487$	

Решение:

$$\begin{cases} 10a_0 + 55a_1 = 445 \\ 55a_0 + 385a_1 = 2907 \end{cases}$$

$$\begin{cases} a_0 = 13,8 \\ a_1 = 5,6 \end{cases}$$

$$\bar{y}_x = 13,8 + 5,6x$$

$x$	$y$	$xy$	$x^2$	$y^2$	$\bar{y}_x$
1	20	20	1	400	19,4
2	25	50	4	625	25,0
3	31	93	9	961	30,6
4	31	124	16	961	36,2
5	40	200	25	1 600	41,8
6	56	336	36	3 136	47,4
7	52	364	49	2 704	53,0
8	60	480	64	3 600	58,6
9	60	540	81	3 600	64,2
10	70	700	100	4 900	69,8
$\Sigma=55$	$\Sigma=445$	$\Sigma=2\ 907$	$\Sigma=385$	$\Sigma=22\ 487$	$\Sigma=446$

$$r_{xy} = 0,98$$

$$\overline{xy} = \frac{2907}{10} = 290,7$$

$$\bar{x} = \frac{55}{10} = 5,5 \quad \bar{y} = \frac{445}{10} = 44,5$$

$$\sigma_y = \sqrt{2248,7 - 44,5^2} = 16,38$$

$$\sigma_x = \sqrt{38,5 - 5,5^2} = 2,87$$

$$r_{xy} = \frac{290,7 - 5,5 \cdot 44,5}{16,38 \cdot 2,87} = 0,98$$

# Парная нелинейная зависимость

$$\bar{y}_x = a_0 + \frac{a_1}{x}$$

$$\left\{ \begin{array}{l} na_0 + a_1 \sum \frac{1}{x} = \sum y \\ a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{y}{x} \end{array} \right.$$

# Корреляционное отношение

$$\eta = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}}$$

$$0 \leq \eta \leq 1$$

$$\sigma_{ост}^2 = \frac{\sum (y_i - \bar{y}_x)^2}{n}$$

# Ранговый коэффициент корреляции Спирмена

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n - T}$$

$$d^2 = (R_y - R_x)^2$$

$$T = \frac{1}{2} \left[ \sum (t_i^3 - t_i) \right]$$

$$-1 \leq \rho \leq 1$$

# Ранговый коэффициент корреляции Спирмена

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n}$$

$$d^2 = (R_y - R_x)^2$$

$$-1 \leq \rho \leq 1$$

$|\rho| < 0,3$  - очень слабая связь,  
практически отсутствует

$0,3 < |\rho| < 0,5$  - связь слабая

$0,5 < |\rho| < 0,7$  - связь средняя

$|\rho| > 0,7$  - связь тесная

Стоимость основных фондов, $x_i$	Производство продукции, $y_i$	$R_x$	$R_y$	$d^2$
6,8	5,4	3	1	4
9,0	10,9	5	7	4
8,0	6,8	4	3	1
9,9	8,5	6	4	4
6,5	9,3	2	5	9
10,2	9,8	7,5	6	2,25
5,4	6,5	1	2	1
12,0	15,6	9	9	0
10,2	12,9	7,5	8	0,25
14,5	16,4	10	10	0
				$\Sigma=25,5$

$$T = \frac{1}{2}(2^3 - 2) = 3$$

$$\rho = 1 - \frac{6 \cdot 25,5}{10^3 - 10 - 3} = 0,84$$

# Ранговый коэффициент корреляции Кенделла

$$\tau = \frac{2S}{n^2 - n}$$

$$\tau = \frac{S}{\sqrt{\left[ \frac{(n^2 - n)}{2} - V_x \right] \cdot \left[ \frac{(n^2 - n)}{2} - V_y \right]}}$$

$$V_{xy} = \frac{1}{2} \sum (t^2 - t)$$

# Алгоритм расчета рангового коэффициента корреляции Кенделла

1. Значения  $X$  ранжируются в порядке возрастания или убывания
2. Значения  $Y$  располагаются в порядке соответствующем значениям  $X$
3. Для каждого значения  $Y$  определяется количество рангов следующих за ним и превышающих его величину. Сумма данных значений обозначается -  $P$

# Алгоритм расчета рангового коэффициента корреляции Кенделла

4. Для каждого значения  $U$  определяется количество рангов следующих за ним и меньших по величине. Сумма данных значений обозначается –  $Q$ . И фиксируется со знаком минус.
5. Определяется  $S=P+Q$

$|\tau| < 0,3$  - очень слабая связь,  
практически отсутствует

$0,3 < |\tau| < 0,5$  - связь слабая

$0,5 < |\tau| < 0,7$  - связь средняя

$|\tau| > 0,7$  - связь тесная

$$\tau = \frac{2}{3} \rho$$



$$S=39+(-6)=33$$

$$V_x = \frac{1}{2} [(2^2 - 2)] = 1$$

$$\tau = \frac{33}{\sqrt{[(10^2 - 10) / 2 - 1] \cdot [(10^2 - 10) / 2 - 0]}} = 0.74$$

# Ранговый коэффициент конкордации

$$W = \frac{12S}{m^2 (n^3 - n)}$$

$m$  – количество факторов

$n$  – число наблюдений

$S$  – отклонение суммы квадратов рангов от  
средней квадратов рангов

# Оцените тесноту связи

Стоимость основных фондов, $x_i$	Производство продукции, $y_i$	Численность работающих, $z_i$	$R_x$	$R_y$	$R_z$	Сумма строк	Квадраты сумм
6,8	5,4	21	3	1	1	5	25
9,0	10,9	35	5	7	7	19	361
8,0	6,8	25	4	3	2	9	81
9,9	8,5	30	6	4	4	14	196
6,5	9,3	36	2	5	8	15	225
10,2	9,8	32	7	6	5	18	324
5,4	6,5	29	1	2	3	6	36
12,0	15,6	37	9	9	9	27	729
10,3	12,9	33	8	8	6	22	484
14,5	16,4	40	10	10	10	30	900
						165	3361

$$S = 3361 - \frac{165^2}{10} = 336,1$$

$$W = \frac{12 \cdot 336,1}{3^2 (10^3 - 10)} = 0,45$$

# Коэффициент знаков Фехнера

$$K_{\phi} = \frac{a - b}{a + b}$$

где  $a$  – число совпадений знаков отклонений значений признака от их средней величины

$b$  – число несовпадений знаков отклонений значений признака от их средней величины

$$-1 \leq K_{\phi} \leq 1$$

$K_{\phi}$  со знаком “+” – связь прямая

$K_{\phi}$  со знаком “-” – связь обратная

$|K_{\phi}| < 0,3$  - очень слабая связь,  
практически отсутствует

$0,3 < |K_{\phi}| < 0,5$  - связь слабая

$0,5 < |K_{\phi}| < 0,7$  - связь средняя

$|K_{\phi}| > 0,7$  - связь тесная

# Пример: Оцените тесноту связи между

$x_i$	$y_i$	Знак отклонения $x_i$ от $\bar{x}$	Знак отклонения $y_i$ от $\bar{y}$	$a/b$
6,8	5,4			
9,0	10,9			
8,0	6,8			
9,9	8,5			
6,5	9,3			
10,2	9,8			
5,4	6,5			
12,0	15,6			
10,2	12,9			
14,5	16,4			
$\Sigma=92,5$	$\Sigma=102,1$			

$x_i$	$y_i$	Знак отклонения $x_i$ от $\bar{x}$	Знак отклонения $y_i$ от $\bar{y}$	$a/b$
6,8	5,4	-		
9,0	10,9	-		
8,0	6,8	-		
9,9	8,5	+		
6,5	9,3	-		
10,2	9,8	+		
5,4	6,5	-		
12,0	15,6	+		
10,2	12,9	+		
14,5	16,4	+		
$\Sigma=92,5$	$\Sigma=102,1$			

$$\bar{x} = 9,25$$

$x_i$	$y_i$	Знак отклонения $x_i$ от $\bar{x}$	Знак отклонения $y_i$ от $\bar{y}$	$a/b$
6,8	5,4	-	-	
9,0	10,9	-	+	
8,0	6,8	-	-	
9,9	8,5	+	-	
6,5	9,3	-	-	
10,2	9,8	+	-	
5,4	6,5	-	-	
12,0	15,6	+	+	
10,2	12,9	+	+	
14,5	16,4	+	+	
$\Sigma=92,5$	$\Sigma=102,1$			

$$\bar{x} = 9,25$$

$$\bar{y} = 10,21$$

$x_i$	$y_i$	Знак отклонения $x_i$ от $\bar{x}$	Знак отклонения $y_i$ от $\bar{y}$	$a/b$
6,8	5,4	-	-	a
9,0	10,9	-	+	b
8,0	6,8	-	-	a
9,9	8,5	+	-	b
6,5	9,3	-	-	a
10,2	9,8	+	-	b
5,4	6,5	-	-	a
12,0	15,6	+	+	a
10,2	12,9	+	+	a
14,5	16,4	+	+	a
$\Sigma=92,5$	$\Sigma=102,1$			

$$\bar{x} = 9,25$$

$$\bar{y} = 10,21$$

$$a=7$$

$$b=3$$

$$K_{\phi} = \frac{7-3}{10} = 0,4$$

# Оценка тесноты связи между качественными признаками

- коэффициент ассоциации
- коэффициент контингенции
- коэффициент взаимной  
сопряженности Пирсона
- коэффициент взаимной  
сопряженности Чупрова

# Коэффициент ассоциации

$$A = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}$$

Где  $a, b, c, d$  - частоты внутри  
таблицы сопряженности

$$-1 \leq A \leq 1$$

$|A| > 0,5$  - связь существует

A со знаком “+” – связь прямая

A со знаком “-” – связь обратная

# Коэффициент контингенции

$$K = \frac{a \cdot d - b \cdot c}{\sqrt{(a + b) \cdot (a + c) \cdot (b + d) \cdot (c + d)}}$$

$$-1 \leq K \leq 1$$

$|K| > 0,3$  - связь существует

# Коэффициент взаимной сопряженности Пирсона

$$c_n = \sqrt{\frac{\phi^2}{1 + \phi^2}}$$

Где  $\phi^2$  - это показатель взаимной сопряженности

$$\phi^2 = \sum \frac{n_{xy}^2}{n_x \cdot n_y} - 1$$

# Коэффициент взаимной сопряженности Чупрова

$$c_{\text{ч}} = \sqrt{\frac{\varphi^2}{\sqrt{(\kappa_1 - 1) \cdot (\kappa_2 - 1)}}}$$

$$0 \leq c \leq 1$$

$c \rightarrow 0$  - связь слабая

$c \rightarrow 1$  - связь тесная

Пример: определите тесноту  
связи между цветом глаз  
матерей и дочерей

Цвет глаз дочерей	Цвет глаз матерей		<i>Итого</i>
	светлые	темные	
светлые	471 $a$	148 $b$	619
темные	151 $c$	230 $d$	381
<i>Итого</i>	622	378	1000

$$A = \frac{471 \cdot 230 - 151 \cdot 148}{471 \cdot 230 + 151 \cdot 148} = \frac{85982}{130678} = 0,66$$

$$K = \frac{471 \cdot 230 - 151 \cdot 148}{\sqrt{619 \cdot 622 \cdot 378 \cdot 381}} = 0,37$$

Пример: оцените тесноту связи между цветом а/м и социальным статусом автовладельца

Социальное положение	цвет			<i>Итого</i>
	красный	синий	черный	
рабочий	19	12	9	40
служащий	7	18	15	40
Предприниматель	4	10	26	40
<i>Итого</i>	30	40	50	120

$$\begin{aligned} \varphi^2 = & \frac{19^2}{40 \cdot 30} + \frac{12^2}{40 \cdot 40} + \frac{9^2}{40 \cdot 50} + \frac{7^2}{40 \cdot 30} + \\ & + \frac{18^2}{40 \cdot 40} + \frac{15^2}{40 \cdot 50} + \frac{4^2}{40 \cdot 30} + \frac{10^2}{40 \cdot 40} + \frac{26^2}{40 \cdot 50} - 1 = 0,204 \end{aligned}$$

$$c_n = \sqrt{\frac{0,204}{1 + 0,204}} = 0,41$$

$$c_y = \sqrt{\frac{0,204}{\sqrt{(3-1) \cdot (3-1)}}} = 0,32$$

# Проверка значимости коэффициентов корреляции

# Проверка значимости парного линейного коэффициента корреляции.

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

Например:

$$r_{xy} = \frac{290,7 - 5,5 \cdot 44,5}{16,38 \cdot 2,87} = 0,98$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

# Проверка значимости рангового коэффициента корреляции Спирмена.

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

# Проверка значимости коэффициента конкордации

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$

$$t_{\text{расч}} = -$$

$$t_{\text{расч}} = \frac{r}{\sigma_r}$$