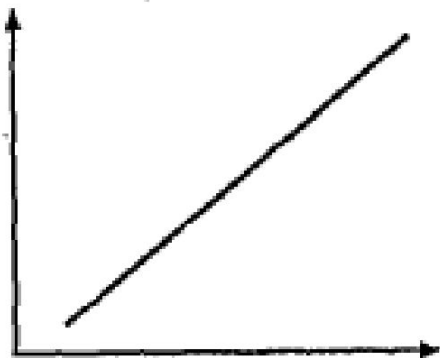


# КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

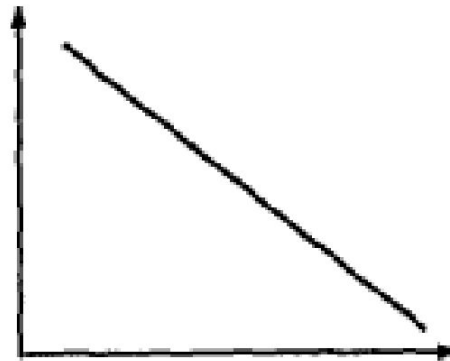
- **Корреляция** – согласованность изменений двух признаков

Линейная

Положительная связь



Отрицательная связь



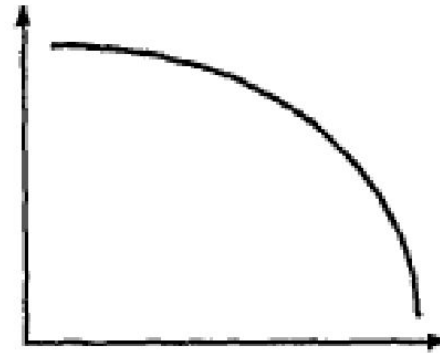
Если увеличение одной переменной связано с увеличением другой, то связь — *положительная (прямая)*; если увеличение одной переменной связано с уменьшением другой, то связь — *отрицательная (обратная)*.

Если изменение одной переменной на одну единицу всегда приводит к изменению другой переменной на одну и ту же величину, функция является *линейной* (график ее представляет прямую линию); любая другая связь — *нелинейная*.

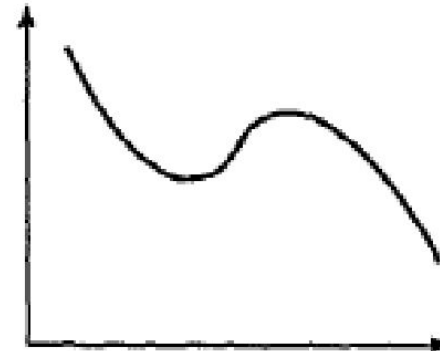
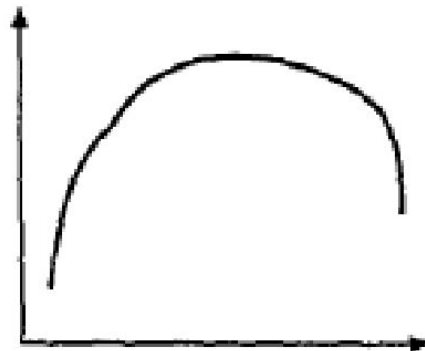
## Положительная связь

## Отрицательная связь

Нелинейная  
монотонная



Нелинейная  
немонотонная



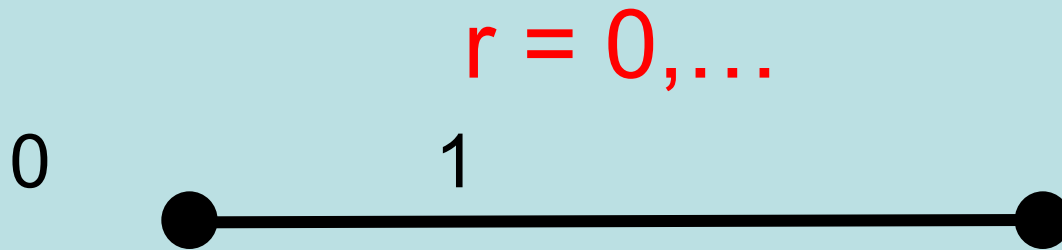
Если направление изменения одной переменной не меняется с возрастанием (убыванием) другой переменной, то такая функция — *монотонная*; в противном случае функцию называют *немонотонной*.

- **Корреляционный анализ** — это проверка гипотез о связях между переменными с использованием коэффициентов корреляции.
- **Коэффициент корреляции** — это количественная мера силы и направления связи.

## Задачи корреляционного анализа:

- измерение тесноты (силы) связи;
- **Показатель – эмпирическое значение.**
- установление направления (положительного или отрицательного) связи между признаками;
- **Показатель – знак коэф. корреляции.**
- проверка надежности связи.
- **Показатель – уровень значимости.**

- **Сила связи** определяется по абсолютной величине корреляции (меняется от 0 до 1).



Абсолютная – без учета знака

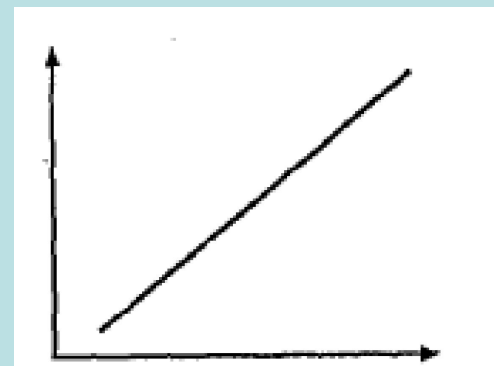
# Классификация коэффициентов корреляции по силе

<b>Сильная</b>	<b><math>r &gt; 0,7</math></b>
<b>Средняя</b>	<b><math>0,5 &lt; r &lt; 0,69</math></b>
<b>Умеренная</b>	<b><math>0,3 &lt; r &lt; 0,49</math></b>
<b>Слабая</b>	<b><math>0,2 &lt; r &lt; 0,29</math></b>
<b>Очень слабая</b>	<b><math>r &lt; 0,19</math></b>

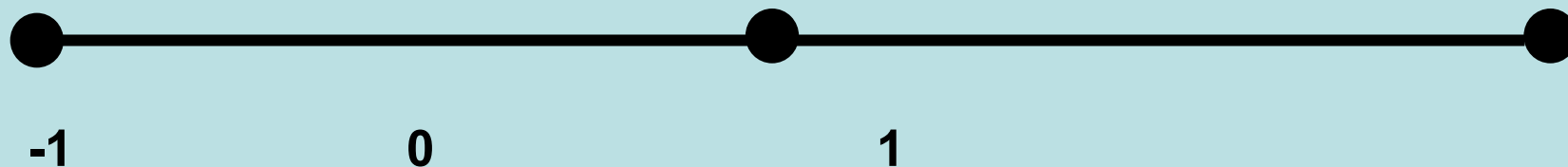
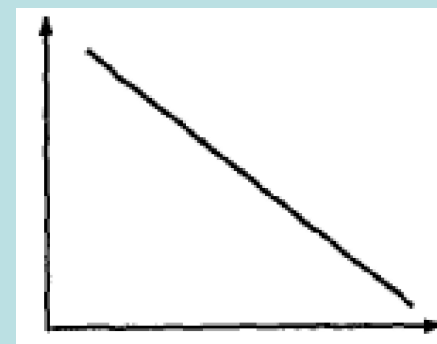


# Направление связи определяется по знаку корреляции:

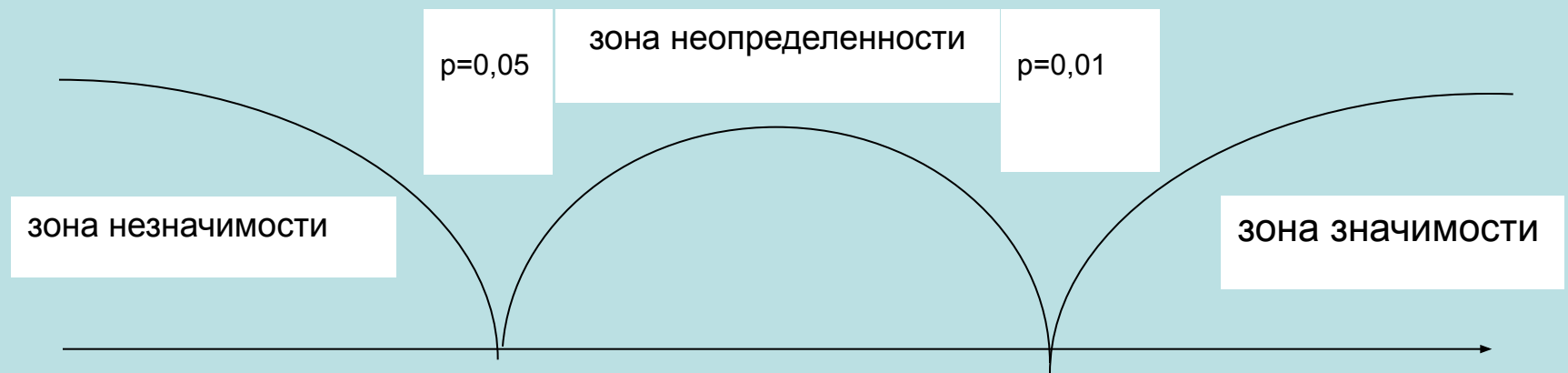
положительный — связь  
прямая;



отрицательный — связь  
обратная.



- Надежность связи определяется  $p$ -уровнем статистической значимости (чем меньше  $p$ -уровень, тем выше статистическая значимость, достоверность связи).



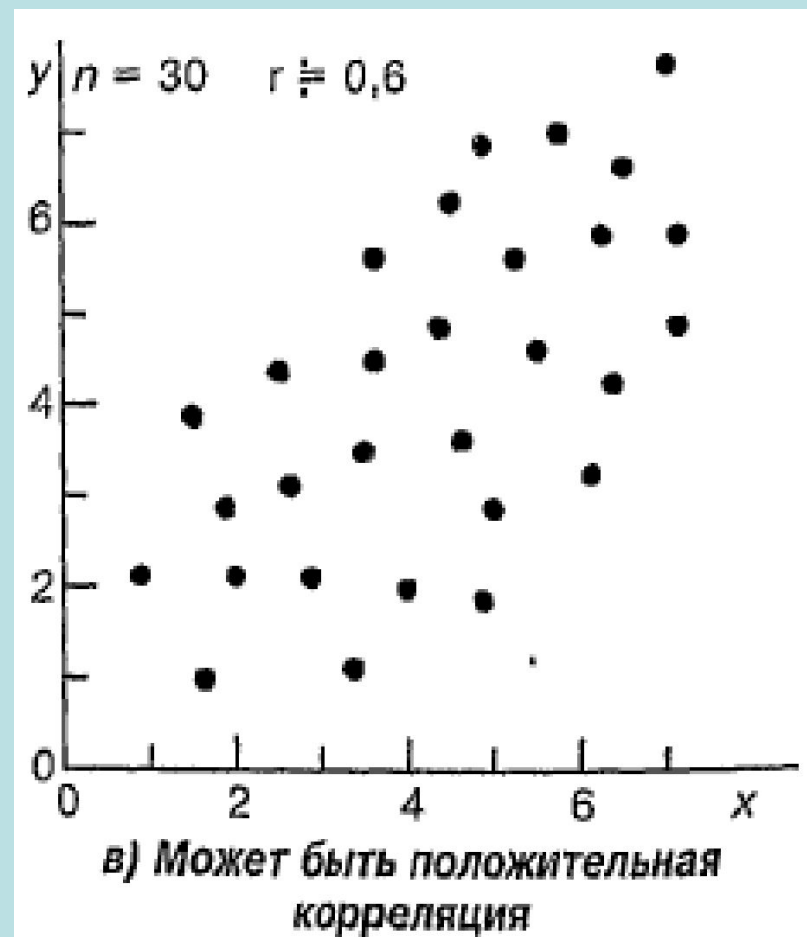
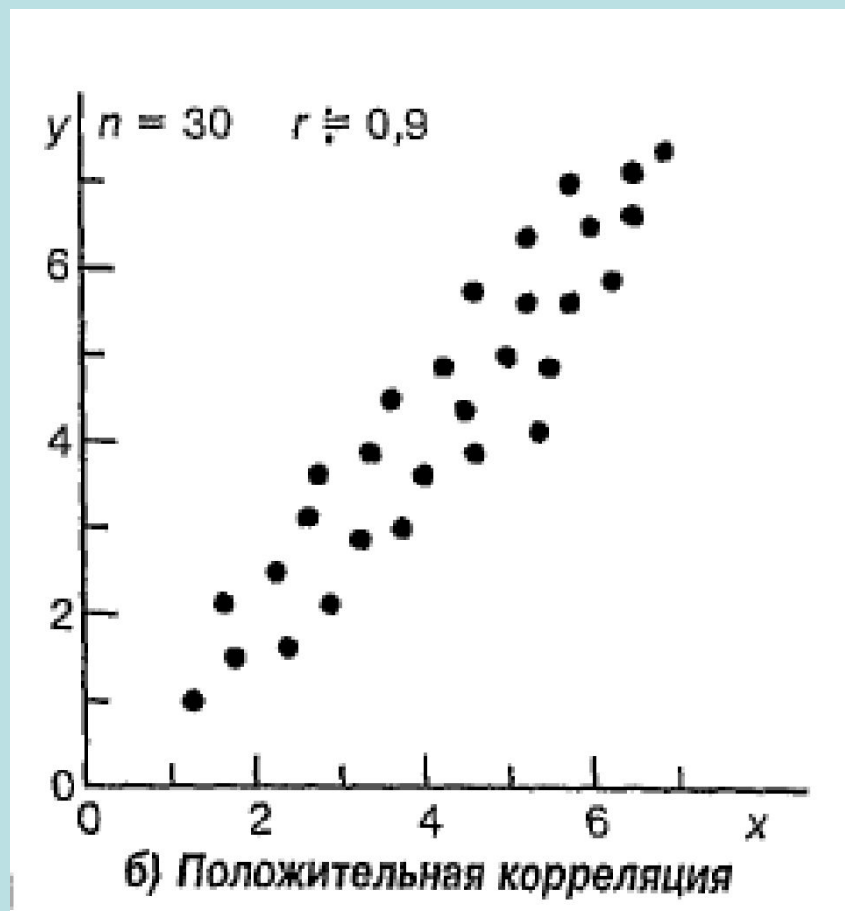
# Классификация коэффициентов корреляции по значимости

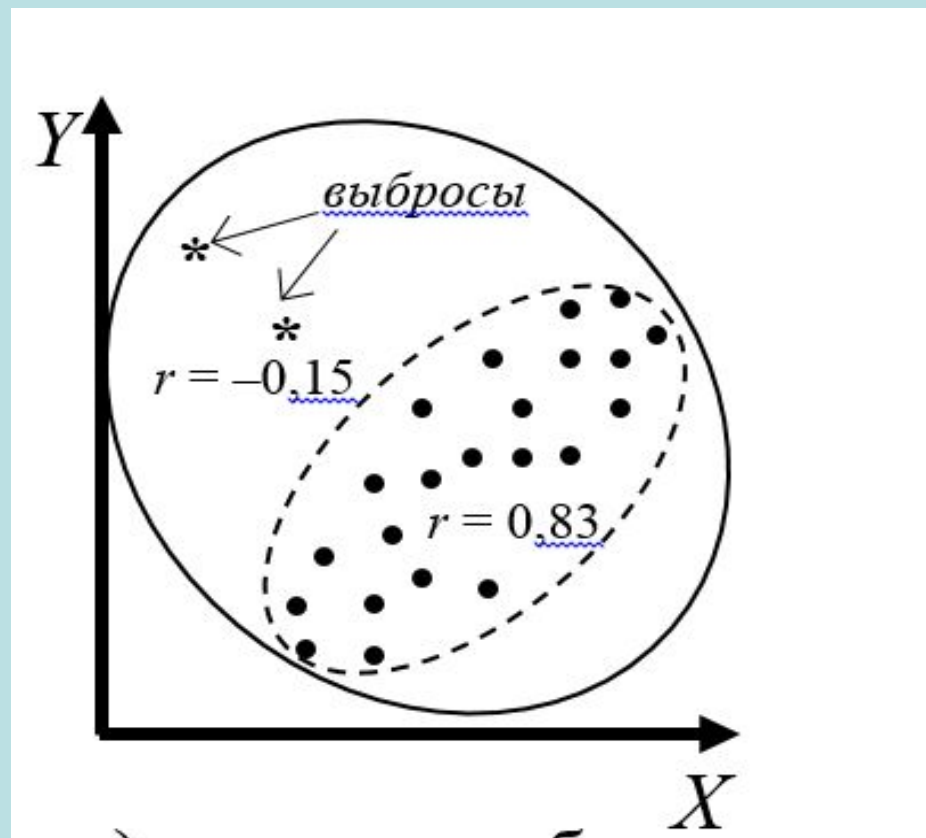
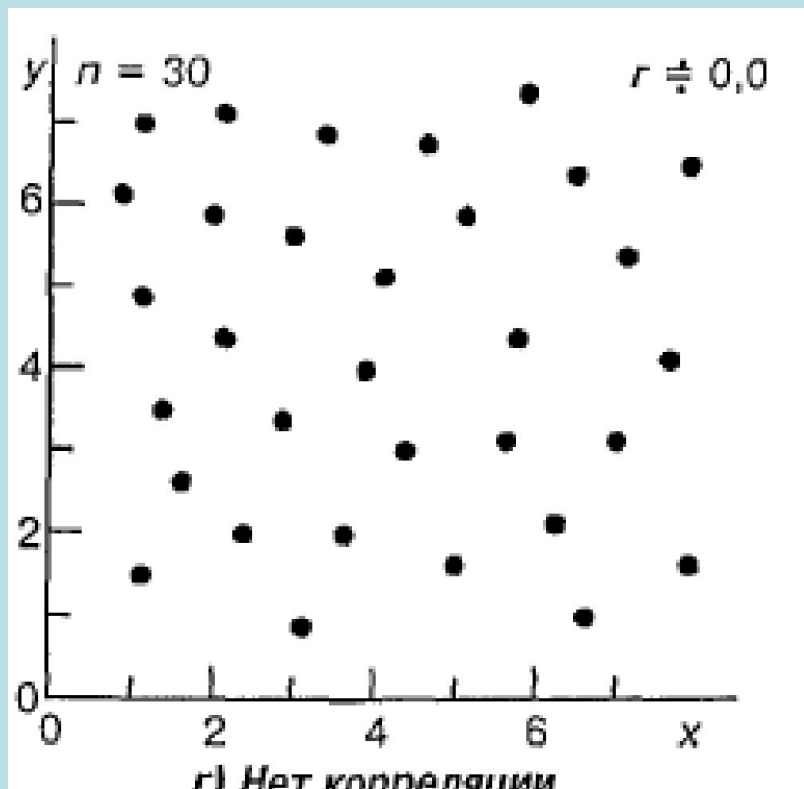
<b>Высокозначимая корреляция</b>	<b><math>p \leq 0,01</math></b>
<b>Статистически значимая корреляция</b>	<b><math>p \leq 0,05</math></b>
<b>Незначимая корреляция</b>	<b><math>p &gt; 0,05</math></b>

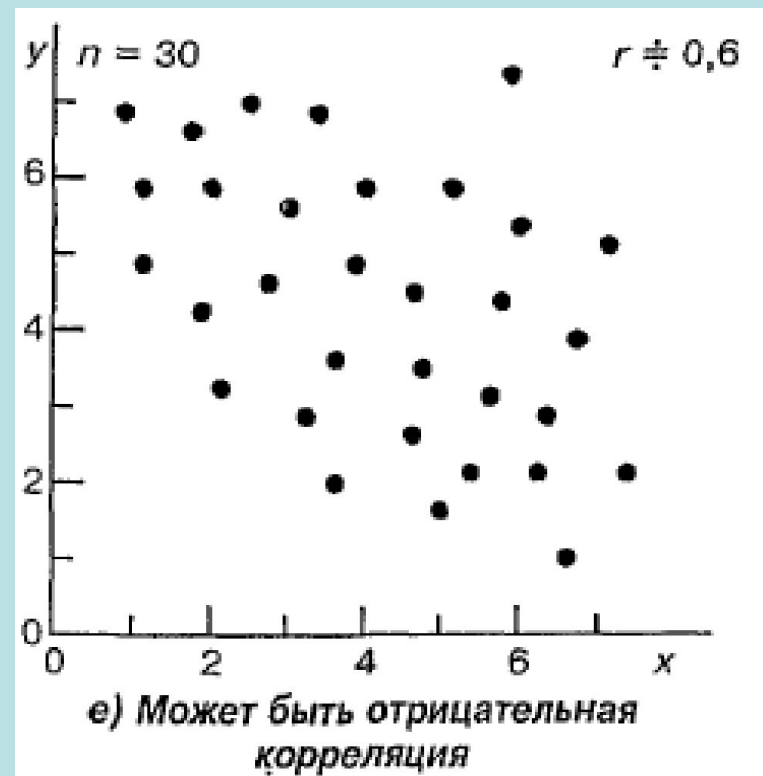
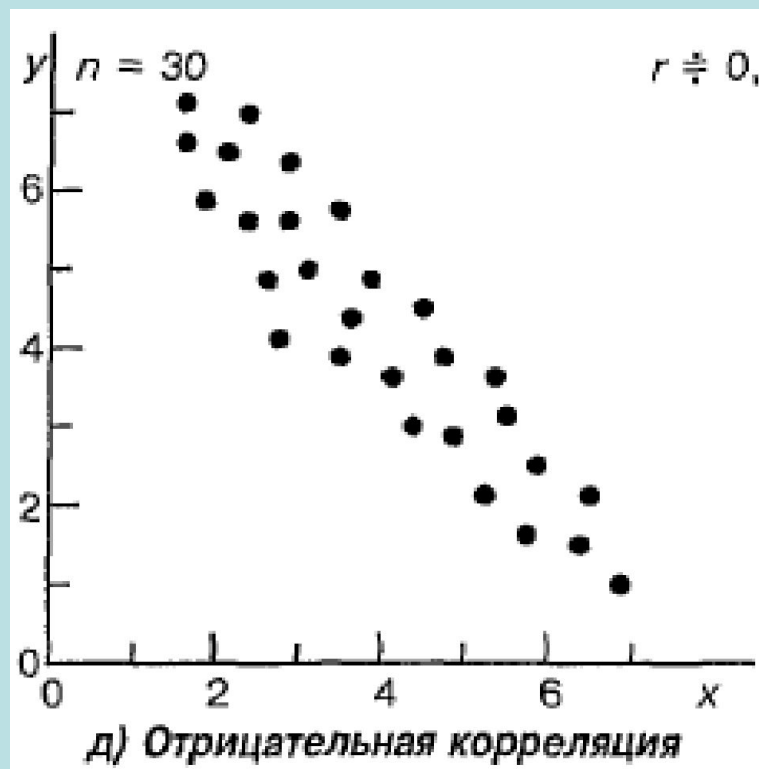
**Статистическая значимость  
коэффициента корреляции тем выше  
( $p$ -уровень меньше):**

- чем больше его абсолютная величина (при одном и том же объеме выборки)
- чем больше объем выборки (при одном и том же значении корреляции).

# Типы корреляционных связей







## Выбор коэффициента корреляции в зависимости от типа измерительной шкалы

Тип шкалы		Характер распределени я	Мера связи
Переменная X	Переменная Y		
Интервальная, абсолютная	Интервальная, абсолютная	Нормальное	Коэффициент корреляции r-Пирсона
Интервальная, абсолютная	Интервальная, абсолютная	Отличается от нормального	Коэффициент корреляции r-Спирмена
Ранговая, интервальная, абсолютная	Ранговая, интервальная, абсолютная	-	Коэффициент корреляции r-Спирмена
Ранговая	Ранговая	-	$\tau$ -Кендалла



## **Основная статистическая гипотеза**

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

показатель корреляции значимо не отличается от нуля; взаимосвязь статистически недостоверна

## **Альтернативная статистическая гипотеза**

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

показатель корреляции значимо отличается от нуля; взаимосвязь статистически достоверна

# Проверка значимости коэффициента корреляции

- Проверка гипотез осуществляется путем сравнения полученных эмпирических коэффициентов с табличными критическими значениями.
- $r_{\text{ЭМП.}} \geq r_{\text{кр.}} \Rightarrow H_1$
- Обнаружена статистически значимая связь между показателями.
- $r_{\text{ЭМП.}} < r_{\text{кр.}} \Rightarrow H_0$
- Не установлено наличие достоверной связи между показателями.

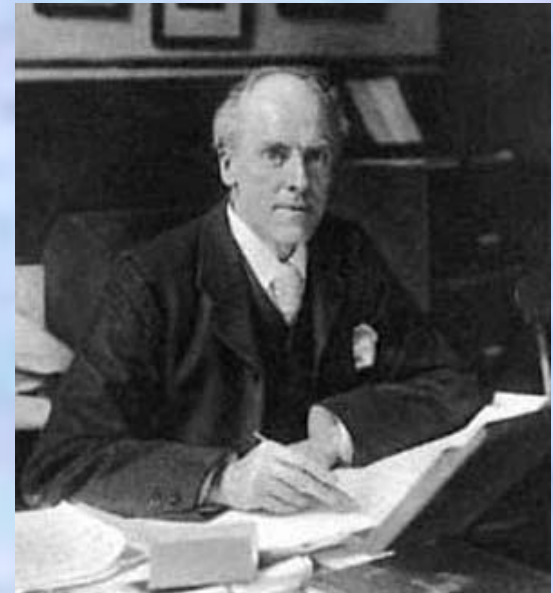
# КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ $r$ -ПИРСОНА ( $r$ -СПИРМЕНА)

(для проверки ненаправленных альтернатив,  $n$  — объем выборки)

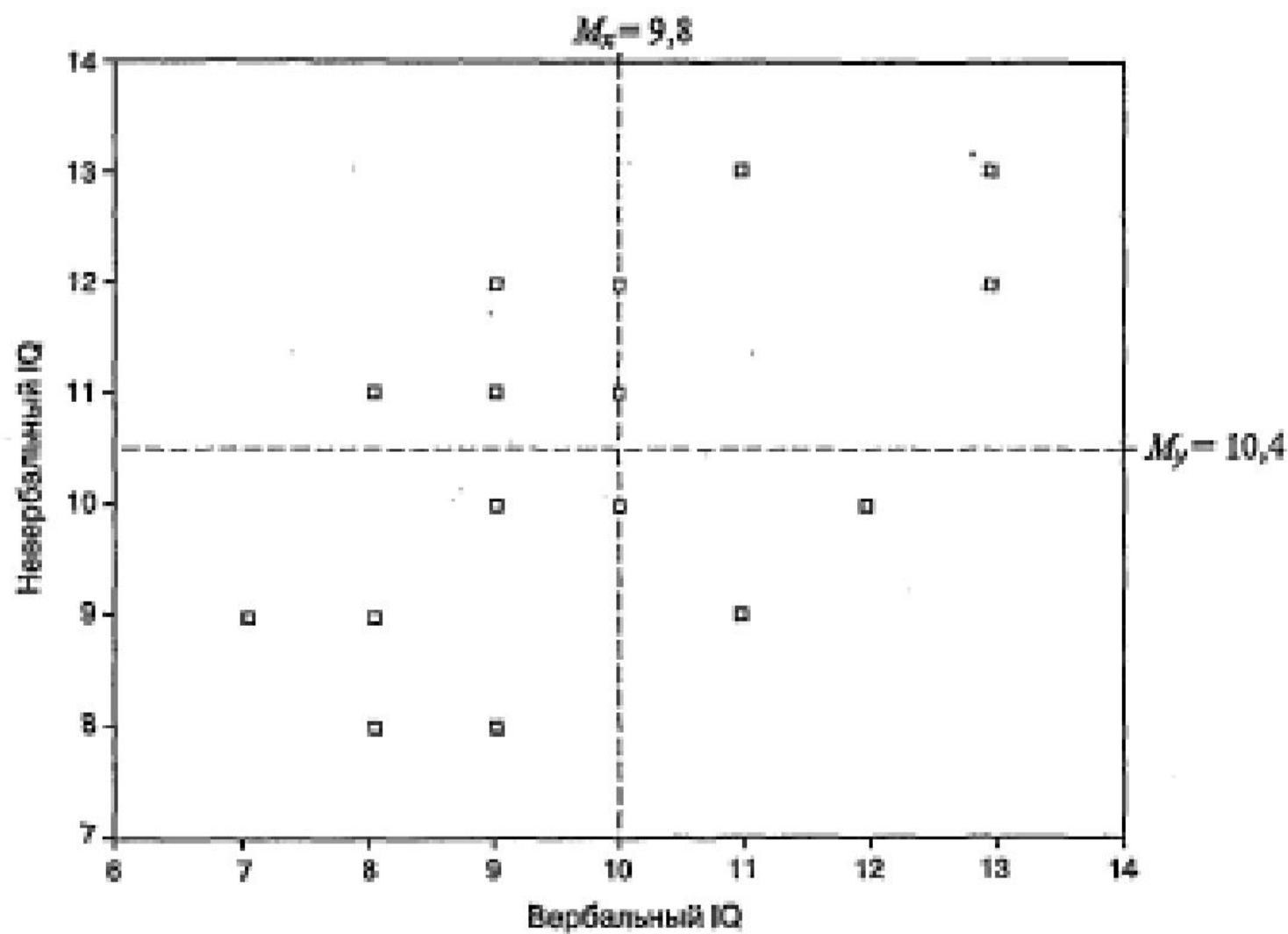
$n$	$p$				$n$	$p$			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428

# Коэффициент корреляции $r$ -Пирсона

- Применяется для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке.
- **Ограничение:**
- не менее 5 испытуемых.



№	Вербальный IQ (x)	Невербальный IQ (y)
1	13	12
2	9	11
3	8	8
4	9	12
5	7	9
6	9	11
7	8	9
8	13	13
9	11	9
10	12	10
11	8	9
12	9	8
13	10	10
14	10	12
15	12	10
16	10	10
17	8	11
18	9	10
19	10	11
20	11	13
Средние:	9,8	10,4



- Положение каждой точки:  
 $(x_i - M_x)$  и  $(y_i - M_y)$
  - Если произведение отклонений  $(x_i - M_x)(y_i - M_y)$  положительное, то данные  $i$ -испытуемого говорят о прямой взаимосвязи, если отрицательное – то об обратной.
- Общий показатель силы и направления связи:

$$\sum_{i=1}^N (x_i - M_x)(y_i - M_y)$$

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{(N-1)\sigma_x\sigma_y}$$

или

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\frac{\sum_i (x_i - M_x)^2}{N-1}} \sqrt{\frac{\sum_i (y_i - M_y)^2}{N-1}} (N-1)}$$



No	X	Y	$(x_i - M_x)$	$(y_i - M_y)$	$(x_i - M_x)^2$	$(y_i - M_y)^2$	$(x_i - M_x)(y_i - M_y)$
1	159	47	-7	-11	49	121	77
2	160	49	-6	-9	36	81	54
3	172	65	6	7	36	49	42
4	160	57	-6	-1	36	1	6
5	171	68	5	10	25	100	50
6	163	50	-3	-8	9	64	24
7	164	59	-2	1	4	1	-2
8	166	68	0	10	0	100	0
9	175	63	9	5	81	25	45
10	170	54	4	-4	16	16	-16
<b><math>\Sigma</math></b>	<b>1660</b>	<b>580</b>			<b>292</b>	<b>558</b>	<b>280</b>

- $M_x = 1660/10 = \mathbf{166}$      $M_y = 580/10 = \mathbf{58}$

$$\sigma_x = \sqrt{\frac{292}{10-1}} \approx 5,59$$

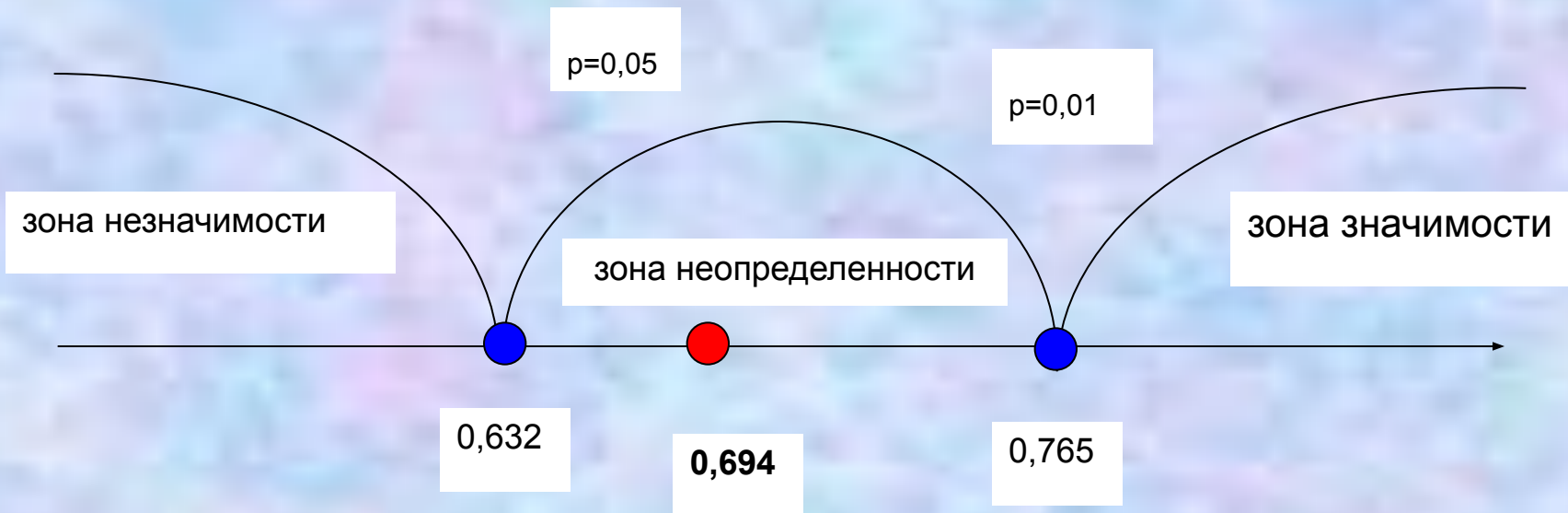
$$\sigma_y = \sqrt{\frac{558}{10-1}} \approx 7,87$$

$$r = \frac{280}{(10-1) \times 5,69 \times 7,87} \approx 0,694$$

# КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ $r$ -ПИРСОНА ( $r$ -СПИРМЕНА)

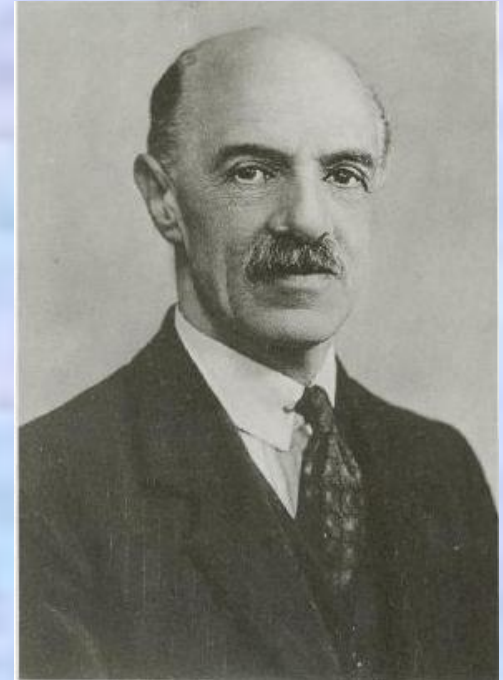
(для проверки ненаправленных альтернатив,  $n$  — объем выборки)

$n$	$p$				$n$	$p$			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428



Подтверждается гипотеза  $H_1$ .  
 Имеется значимая  
 корреляционная связь  
 между показателями X и Y

## Коэффициент корреляции r-Спирмена



Если обе переменные представлены в порядковой (ранговой) шкале или одна в порядковой, а другая – в метрической, используется **коэффициент ранговой корреляции r-Спирмена**.

$$r_s = 1 - \frac{6 \sum_i d_i^2}{N(N^2 - 1)}$$

где  $d_i$  – разность рангов для испытуемого с номером  $i$

**Ограничение:** не менее 5 испытуемых.

№	X	Y	Ранги X	Ранги Y	$d_i$	$d_i^2$
1	122	4,7	7	2	5	25
2	105	4,5	10	4	6	36
3	100	4,4	11	5	6	36
4	145	3,8	5	9	-4	16
5	130	3,7	6	10	-4	16
6	90	4,6	12	3	9	81
7	162	4	3	8	-5	25
8	172	4,2	1	6	-5	25
9	120	4,1	8	7	1	1
10	150	3,6	4	11	-7	49
11	170	3,5	2	12	-10	100
12	112	4,8	9	1	8	64
$\Sigma$	-	-	78	78	0	<b>474</b>

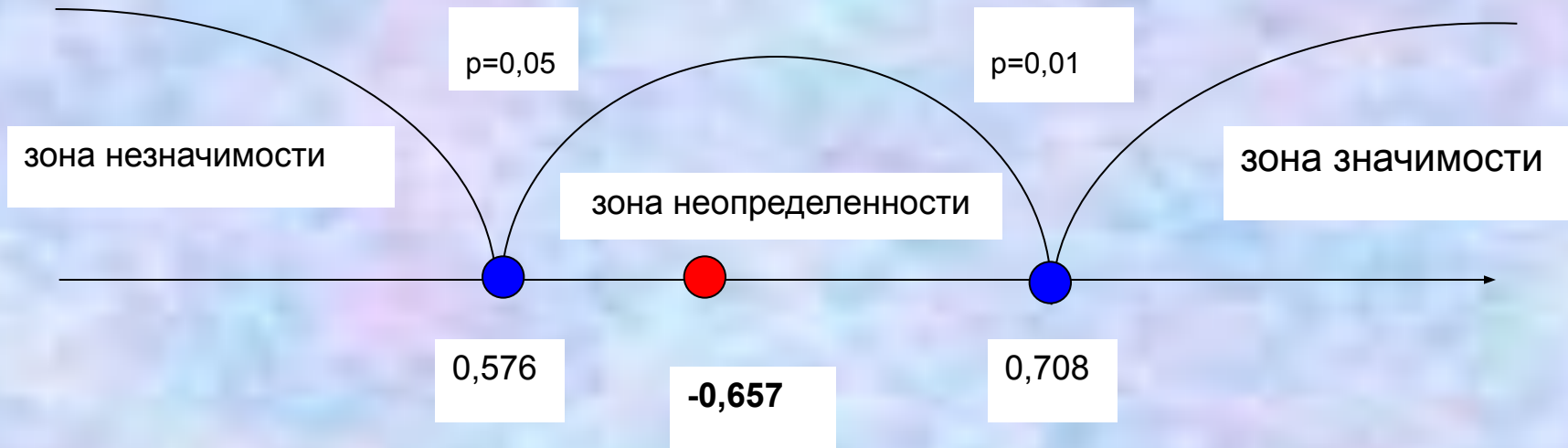
$$r_s = 1 - \frac{6 \times 474}{12(144 - 1)} = -0,657$$



# КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ $r$ -ПИРСОНА ( $r$ -СПИРМЕНА)

(для проверки ненаправленных альтернатив,  $n$  — объем выборки)

$n$	$p$				$n$	$p$			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428



Подтверждается гипотеза  $H_1$ .  
 Имеется значимая обратная  
 корреляционная связь  
 между показателями X и Y

$$r_s = 1 - \frac{6 \sum_i d_i^2 + T_a + T_b}{N(N^2 - 1)}$$

$$T_a = \Sigma(a^3 - a) / 12$$

$$T_b = \Sigma(b^3 - b) / 12$$

a – объем каждой группы одинаковых рангов в ранговом ряду А

b – объем каждой группы одинаковых рангов в ранговом ряду В

- Если две группы (или более) одинаковых рангов

$$T_a = \Sigma \left[ (a_1^2 - a_1) + (a_2^2 - a_2) \right] / 12$$

$$T_b = \Sigma \left[ (b_1^2 - b_1) + (b_2^2 - b_2) \right] / 12$$