

# Описательные статистики

Описательные статистики выборки – это ЧИСЛА,  
которые характеризуют выборку

К основным описательным статистикам относятся:

- МОДА
- СРЕДНЕЕ ( $m$ )
- СТАНДАРТНОЕ ОТКЛОНЕНИЕ ( $s$ )

# МОДА

**Мода выборки** – это статистика (число) равная варианту выборки, частота которой наибольшая.

Выборку, в которой **только одна** варианта имеет наибольшую частоту, называют **унимодальной выборкой**.

**Примеры моды выборки:**

1 Самооценка по психологии	1	4	5	6	7	8	9
Частота	1	3	3	8	3	6	1

Мода выборки равна **6**

Так как **8** – наибольшая частота в данной выборке

Тип темперамента	Холерик	Флегматик	Меланхолик	Сангвиник
Частота	6	3	5	9

Мода выборки **Сангвиник**

Так как **9** – наибольшая частота в данной выборке

Выборка, в которой **только две смежные** варианты имеют наибольшую частоту, также является **унимодальной выборкой**.

2 Самооценка по логике	1	2	3	4	5	6	7
Частота	1	3	5	5	3	2	1

Мода выборки равна **3,5**  $((3+4)/2)$

Так как **5** – наибольшая частота в двух данных выборках

Выборка, в которой **две несмежные варианты** имеют наибольшую частоту, называют **бимодальной выборкой**.

3 Рост, см	165	166	167	168	169	170	171	172	173	174	175
Частота	3	3	4	8	2	5	3	8	2	1	1

Моды выборки

Так как **8** – наибольшая частота

В остальных случаях:  
**МОДЫ - НЕТ**

# СРЕДНЕЕ

Среднее выборки ( $m$ ) – это статистика (число) равная отношению суммы всех значений варианты к объёму выборки

Среднее ( $m$ ) – обозначает условный центр выборки.

Если выборка имеет *небольшой* объем – то **среднее вычисляют** по определению.

**Пример:** Вычислим среднее выборки «Самооценки по психологии», объем которой 9 студентов: 6, 8, 5, 6, 6, 7, 7, 8, 6.  
 $m = (6+8+5+6+6+7+7+8+6)/9 = 59/9 = 6,56$  баллов.

Если выборка имеет *большой* объем – то **среднее вычисляют** в Excel ( $f = \text{СРЗНАЧ}$ )

Если составлено **распределение частот выборки**, то для вычисления **среднего** используется формула:

$$m = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n}$$

где  $m$  – среднее выборки;

$x_1, x_2, \dots, x_k$  – варианты выборки;

$n_1, n_2, \dots, n_k$  – частоты вариант выборки;

$n$  – объем выборки.

Или вычисления также проводят в Excel

Пример вычисления среднего, если составлено распределение частот выборки											
Баллы ( $x_i$ )	62	69	73	75	80	82	84	85	92	99	сумма
Частота ( $n_i$ )	3	5	10	26	10	5	12	13	8	2	94
$x_i \cdot n_i$	186	345	730	1950	800	410	1008	1105	736	198	7468
СРЕДНЕЕ	=S5/S4										

Одной характеристики **СРЕДНЕЕ** недостаточно для описания выборки, так как варианты выборки могут находиться на *разных расстояниях от центра выборки*

НАИМЕНЬШИЙ							=СРЗНАЧ(C2:C26)		
	A	B	C	D	E	F	G	H	
	№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию			
1									
2	1	м	20	6	6	18			
3	2	м	21	8	3	19			
4	3	ж	23	5	5	20			
5	4	ж	22	6	3	20			
6	5	м	22	6	4	16			
7	6	ж	22	7	5	16			
8	7	ж	22	7	3	17			
9	8	м	21	8	4	17			
10	9	м	20	6	4	17			
11	10	ж	20	8	3	23			
12	11	м	21	8	4	25			
13	12	м	21	9	6	24			
14	13	м	21	8	2	24			
15	14	ж	20	9	1	25			
16	15	ж	19	4	1	16			
17	16	м	27	4	1	21			
18	17	ж	19	6	1	17			
19	18	ж	20	4	2	23			
20	19	ж	19	5	2	20			
21	20	м	18	6	2	18			
22	21	ж	18	6	2	18			
23	22	м	19	7	2	18			
24	23	ж	18	1	3	20			
25	24	ж	18	5	3	19			
26	25	ж	18	6	3	24			
27	среднее: =СРЗНАЧ(C2:C26)						19,8		

# СТАНДАРТНОЕ ОТКЛОНЕНИЕ

Стандартное отклонение ( $s$ ) – это статистика (число) обозначающая стандартный диапазон изменчивости (рассеяния) вариант от среднего ( $m$ )

Стандартным отклонением выборки ( $x_i$ ) объемом  $n$  со средним  $m$  называют число  $s$ , равное квадратному корню отношения суммы квадратов отклонений всех значений варианты от выборочного среднего к  $n - 1$ .

Для вычисления в Excel используется функция: ( $f_x = \text{СТАНДОТКЛОН}$ )

$$s = \sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + \dots + (x_k - m)^2}{n - 1}}$$

где  $x_1, x_2, \dots, x_k$  – варианты выборки;  
 $m$  – среднее выборки;  
 $n$  – объем выборки.

Если составлено **распределение частот выборки**, то для вычисления **стандартного отклонения** используется формула:

$$s = \sqrt{\frac{n \cdot B - A^2}{n(n - 1)}}$$

где  $n$  – объем выборки;  
 $B = x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k$   
 $A = x_1 n_1 + x_2 n_2 + \dots + x_k n_k$

Или вычисления также проводят в Excel

№	Пол	Возраст	Самостоятельно по психологии	Самостоятельно по математике	Сумма оценок за сессию	
1	м	20	6	6	18	
2	м	21	8	3	19	
3	ж	23	5	5	20	
4	ж	22	6	3	20	
5	м	22	6	4	16	
6	ж	22	7	5	16	
7	ж	22	7	3	17	
8	м	21	8	4	17	
9	м	20	6	4	17	
10	ж	20	8	3	23	
11	м	21	8	4	25	
12	м	21	9	6	24	
13	м	21	8	2	24	
14	ж	20	9	1	25	
15	ж	19	4	1	16	
16	м	27	4	1	21	
17	ж	19	6	1	17	
18	ж	20	4	2	23	
19	ж	19	5	2	20	
20	м	18	6	2	18	
21	ж	18	6	2	18	
22	м	19	7	2	18	
23	ж	18	1	3	20	
24	ж	18	5	3	19	
25	ж	18	6	3	24	
26	ж	18	6	3	24	
27	среднее:		20,4	6,2	3,0	19,8
28	станоткл:		=СТАНДОТКЛОН(C2:C26)			
29			СТАНДОТКЛОН(число1; [число2]; ...)			

Баллы (xi)	62	69	73	75	80	82	84	85	92	99	сумма
Частота (ni)	3	5	10	26	10	5	12	13	8	2	94
xi*ni	186	345	730	1950	800	410	1008	1105	736	198	7468
xi*xi	3844	4761	5329	5625	6400	6724	7056	7225	8464	9801	65229
xi*xi*ni	11532	23805	53290	1E+05	64000	33620	84672	93925	67712	19602	598408
стандоткл:	7,40										

# Статистический ВЫВОД

**Статистический критерий** – строгое математическое ПРАВИЛО, по которому принимается или отвергается та или иная статистическая гипотеза

Статистический вывод имеет  
*вероятностный характер*

К основным статистическим критериям относятся:

- $F$ -критерий Фишера;
- $L$ -критерий Колмогорова-Смирнова;
- $G$ -критерий знаков;
- $U$ -критерий Манна-Уитни.

# ВЫБОРКИ

ВЫБОРКИ бывают:

**связанные**  
множества значений **двух свойств**,  
полученные в **одной группе**  
респондентов

**несвязанные**  
множества значений **одного свойства**,  
полученные в **двух группах**  
респондентов

Пример:

№	Форма обучения	«Мотивация»	«Успешность»	«Общительность»
1	Очная	7	3	3
2	Очная	4	12	3
3	Очная	1	14	2
4	Очная	4	9	1
5	Очная	5	13	6
6	Очная	3	34	6
7	Заочная	9	23	7
8	Заочная	4	19	8
9	Заочная	4	18	9
10	Заочная	3	11	3
11	Заочная	5	23	4
12	Заочная	6	23	3
13	Заочная	6	23	3
14	Заочная	2	11	5

Выборки «Мотивация» и «Успешность» для всех студентов являются **связанными**

Выборка «Общительность» студентов очной формы обучения и выборка «Общительность» студентов заочной формы обучения являются **несвязанными**

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

**Статистическими гипотезами** называют **ПРЕДПОЛОЖЕНИЯ** о статистически значимых различиях выборок.

**ВОПРОС?**

«Значимо или не значимо  
отличаются выборки?»

**ГИПОТЕЗА 1:**

выборки статистически  
значимо не различаются

Обозначение –  $H_0$   
(*различий нет*)

**ГИПОТЕЗА 2:**

выборки статистически  
значимо различаются.

Обозначение –  $H_1$   
(*различия есть*)

**ОТВЕТ 1:**

выборки отличаются  
не значимо

**ОТВЕТ 2:**

выборки отличаются  
значимо

**Алгоритм проверки статистических гипотез:**

1. Проверяется гипотеза  $H_0$
2. Если  $H_0$  принимается, то  $H_1$  не рассматривается
3. Если  $H_0$  не принимается, тогда принимается  $H_1$

Принятие решения о  $H_1$  имеет **вероятностный характер**, поэтому указывается *уровень значимости принятия правильного решения* о  $H_1$  (**вероятность вывода**)

# ВЕРОЯТНОСТЬ ВЫВОДА

Так как сперва проверяется гипотеза  $H_0$ , то существует:

Вероятность  
ПРАВИЛЬНО ПРИНЯТЬ  $H_0$   
(т.е. вероятность того, что  
различий **НЕТ**)

Обозначения:

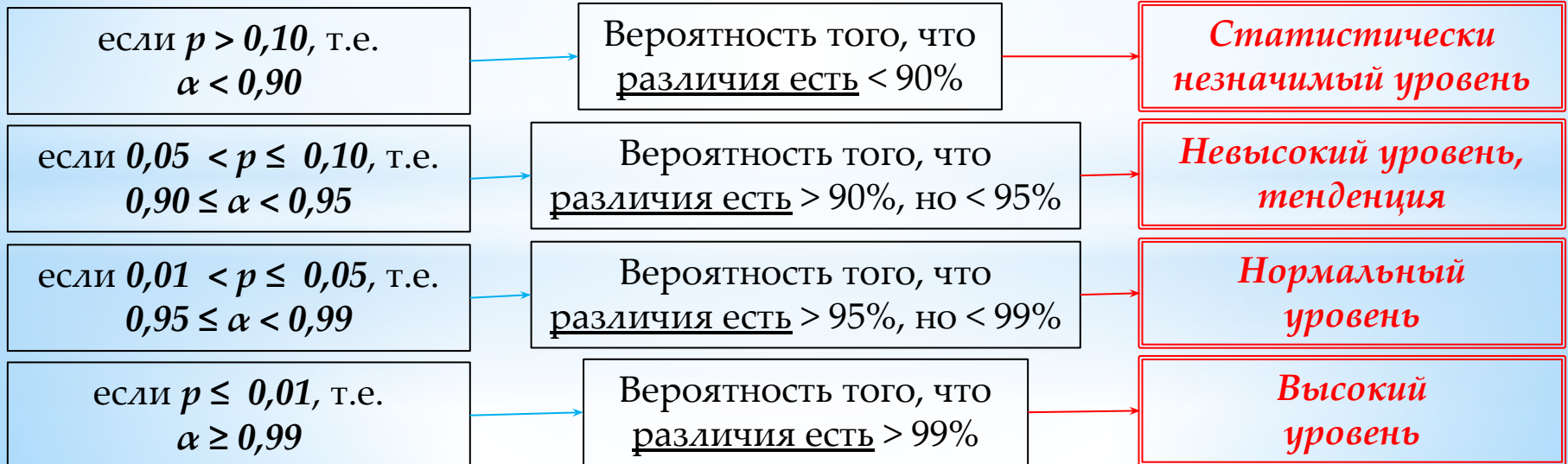
Вероятность  
ОШИБКИ ПРИНЯТИЯ  $H_0$   
(т.е. вероятность того, что  
различия **ЕСТЬ**)

$p$

$1-p = \alpha$

Вероятность ошибки принятия гипотезы  $H_0$  называется  
**уровнем статистической значимости ( $\alpha$ )**.

В психологии различают следующую **шкалу уровней статистической значимости**:





# АЛГОРИТМ ПРОВЕРКИ ГИПОТЕЗ

Статистические критерии (правила) представлены в форме *алгоритма проверки статистических гипотез* и содержат таблицы критических значений случайной величины.

Критерии имеют названия, как правило, связанные с именами авторов

## **ОБЩИЙ АЛГОРИТМ:**

1. Выбирается критерий (правило) сравнения

2. Вычисляется статистика (число) для сравниваемых выборок по правилу, соответствующему критерию (примем как  $C$ )

3. Находится предельное значение статистики (числа) ( $C_\alpha$ ) для установленного исследователем уровня значимости  $\alpha$

4. Сравниваются значения  $C$  и  $C_\alpha$ .

5. Исходя из того, какое значение больше, делается **статистический вывод** о том, принимается  $H_0$  или принимается  $H_1$

6. Формулируется **содержательный вывод**: **различия есть или различий нет**

# φ-критерий Фишера

Область применения:

С помощью φ -критерия Фишера устанавливается значимость различия **долей выраженности** одинакового свойства в двух выборках или двух разных свойств в одной выборке.

Под **долей выраженности свойства** понимается отношение числа респондентов, имеющих это психическое свойство, к объему выборки (например: на первом курсе 25% общительных студентов)

**Особенности применения:**

1. В каждой из сравниваемых выборок должно быть **не менее пяти** респондентов.
2. Выборки могут быть **связанными** или **несвязанными**.
3. Используется таблица φ-критерия Фишера (замены долей выраженности исследуемого свойства на  $\phi_1$  и  $\phi_2$ ).

**Алгоритм φ -критерия Фишера:**

1. Вычисленные доли (проценты) *выраженности* одинакового свойства в I и II выборках *заменяют на соответствующие им значения*  $\phi_1$  и  $\phi_2$  с помощью таблицы φ-критерия Фишера.

2. Вычисляют значение φ по формуле:

$$\varphi = |\varphi_1 - \varphi_2| \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

3. Статистический вывод:

Если  $\varphi < 1,29$ , то принимается гипотеза  $H_0$ .

Если  $1,29 \leq \varphi < 1,64$ , то принимается гипотеза  $H_1$  ( $p \leq 0,10$ ).

Если  $1,64 \leq \varphi < 2,31$ , то принимается гипотеза  $H_1$  ( $p \leq 0,05$ ).

Если  $2,31 \leq \varphi$ , то принимается гипотеза  $H_1$  ( $p \leq 0,01$ ).

### Таблица $F$ -критерия Фишера

$0, \alpha$	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
,00	0,000	0,200	0,284	0,348	0,403	0,451	0,495	0,536	0,574	0,609
,10	0,644	0,676	0,707	0,738	0,767	0,795	0,823	0,850	0,876	0,902
,20	0,927	0,952	0,976	1,000	1,024	1,047	1,070	1,093	1,115	1,137
,30	1,159	1,182	1,203	1,224	1,245	1,266	1,287	1,308	1,328	1,349
,40	1,369	1,390	1,410	1,430	1,451	1,471	1,491	1,511	1,531	1,551
,50	1,571	1,591	1,611	1,631	1,651	1,671	1,691	1,711	1,731	1,752
,60	1,772	1,793	1,813	1,834	1,855	1,875	1,897	1,918	1,939	1,961
,70	1,982	2,004	2,026	2,049	2,071	2,094	2,118	2,141	2,165	2,190
,80	2,214	2,240	2,265	2,292	2,319	2,246	2,375	2,404	2,434	2,465
,90	2,498	2,532	2,568	2,606	2,647	2,691	2,739	2,793	2,858	2,941

# λ-критерий Колмогорова-Смирнова

## Область применения:

С помощью λ -критерия Колмогорова-Смирнова устанавливается уровень статистической значимости различий **распределений частот** одинакового свойства в двух выборках

**Особенности применения** или двух разных свойств в одной выборке.

1. В каждой из сравниваемых выборок должно быть **не менее 50** респондентов.
2. Выборки могут быть **связанными** или **несвязанными**.

## Алгоритм λ -критерия Колмогорова-Смирнова:

1. Составляют процентильные распределения частот исследуемого свойства для I и II выборок в общей таблице.

2. Вычисляют модуль разности процентильного распределения

3. Вычисляют λ по формуле:

$$\lambda = d \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

, где  $d$  – наибольшая разность процентильного распределения;  
 $n_1$  – число респондентов в выборке I;

$n_2$  – число респондентов в выборке II;

4. Статистический вывод:

Если  $\lambda < 1,22$ , то принимается гипотеза  $H_0$ .

Если  $1,22 \leq \lambda < 1,36$ , то принимается гипотеза  $H_1$  ( $p \leq 0,10$ ).

Если  $1,36 \leq \lambda < 1,63$ , то принимается гипотеза  $H_1$  ( $p \leq 0,05$ ).

Если  $1,63 \leq \lambda$ , то принимается гипотеза  $H_1$  ( $p \leq 0,01$ ).

# G-критерий знаков

## Область применения:

С помощью **G-критерия знаков** устанавливается уровень статистической значимости различий свойства A и свойства B у респондентов **одной выборки**.

## Особенности применения:

1. В выборке должно быть **не менее пяти** респондентов.
2. Выборки свойств A и B должны быть связанными.
3. Свойства A и B должны быть измерены в одной шкале или ранжированы.
4. Используется таблица **G-критерия знаков**

## Алгоритм G-критерия знаков:

1. К протоколу свойств A и B, измеренных в одной шкале или ранжированных, добавляют столбец «**Знак A – B**» и заполняют его.

2. Вводят обозначения:

$a$  – число «плюсов» в столбце «**Знак A – B**»;

$b$  – число «минусов» в столбце «**Знак A – B**»;

$n$  – сумма  $a$  и  $b$ ;

$G$  – число, равное меньшему из чисел  $a$  и  $b$ .

3. Находят в таблице G-критерия знаков, в строке  $n$  соответствующие значения  $G_{0,10}$ ,  $G_{0,05}$  и  $G_{0,01}$ .

Если  $G > G_{0,10}$ , то принимается гипотеза  $H_0$ .

Если  $G_{0,05} < G \leq G_{0,10}$ , то принимается гипотеза  $H_1$  ( $p \leq 0,10$ ).

Если  $G_{0,01} < G \leq G_{0,05}$ , то принимается гипотеза  $H_1$  ( $p \leq$

4. Статистический

вывод:

## Таблица G-критерия знаков

n	α			n	α			n	α			n	α		
	0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01
5	1	0	0	27	9	8	7	49	19	18	16	92	38	37	34
6	1	1	0	28	10	9	7	50	19	18	16	94	39	38	35
7	1	1	0	29	10	9	8	52	20	19	17	96	40	38	35
8	2	1	1	30	11	10	8	54	21	20	18	98	41	39	36
9	2	2	1	31	11	10	8	56	22	21	18	100	42	40	37
10	2	2	1	32	11	10	9	58	23	22	19	110	46	45	42
11	3	2	1	33	12	11	9	60	24	22	20	120	51	49	46
12	3	3	2	34	12	11	10	62	25	23	21	130	56	54	50
13	4	3	2	35	13	12	10	64	25	24	22	140	60	58	55
14	4	3	2	36	13	12	10	66	26	25	23	150	65	63	59
15	4	4	3	37	14	13	11	68	27	26	23	160	70	68	64
16	5	4	3	38	14	13	11	70	28	27	24	170	74	72	68
17	5	5	3	39	14	13	12	72	29	28	25	180	79	77	73
18	6	5	4	40	15	14	12	74	30	29	26	190	84	82	77
19	6	5	4	41	15	14	12	76	31	29	27	200	88	86	82
20	6	6	4	42	16	15	13	78	32	30	28	220	98	95	91
21	7	6	5	43	16	15	13	80	33	31	29	240	107	105	100
22	7	6	5	44	17	16	14	82	34	32	29	260	117	114	109
23	8	7	5	45	17	16	14	84	34	33	30	280	126	124	118
24	8	7	6	46	17	16	14	86	35	34	31	300	136	133	128
25	8	8	6	47	18	17	15	88	36	35	32				
26	9	8	7	48	18	17	15	90	37	36	33				

# U-критерий Манна-Уитни

## Область применения:

С помощью **U-критерия Манна-Уитни** устанавливается уровень статистической значимости различий одного свойства у респондентов

## Особенности применения: двух выборок (I и II).

1. В выборке должно быть **не менее четырех-пяти** респондентов.
2. Выборки одного свойства (I и II) должны быть **несвязанными**.
3. Используется таблица **U-критерия Манна-Уитни**.

## Алгоритм U-критерия Манна-Уитни:

1. Проводят ранжирование общей выборки, в которую входят сравниваемые выборки (I и II).

2. Находят сумму рангов вариант выборки I и сумму рангов вариант выборки II.

3. Вычислите значение **U** по формуле:  $U = n_1 * n_2 + 0,5 * n_R * (n_R + 1) - R$ .

где  $n_1$  – объем выборки I;

$n_2$  – объем выборки II;  $n_R$  – объем выборки, имеющей большую сумму рангов;

$R$  – значение большей суммы рангов.

4. Находят в таблице U-критерия Манна-Уитни на пересечении строки и столбца  $n_1$  и  $n_2$  значения  $U_{0,05}$  и  $U_{0,01}$

## 5. Статистический вывод:

Если  $U > U_{0,05}$ , то принимается гипотеза  $H_0$ .

Если  $U_{0,01} < U \leq U_{0,05}$ , то принимается гипотеза  $H_1$  ( $p \leq 0,05$ ).

Если  $U \leq U_{0,01}$ , то принимается гипотеза  $H$  ( $p \leq 0,01$ ).

**Таблица критерия Манна-Уитни ( $\alpha = 0,05$ )**

$n_1$ $n_2$	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
5	2	4																									
6	3	5	7																								
7	4	6	8	11																							
8	5	8	10	13	15																						
9	6	9	12	15	18	21																					
10	7	11	14	17	20	24	27																				
11	8	12	16	19	23	27	31	34																			
12	9	13	17	21	26	30	34	38	42																		
13	10	15	19	24	28	33	37	42	47	51																	
14	11	16	21	26	31	36	41	46	51	56	61																
15	12	18	23	28	33	39	44	50	55	61	66	72															
16	14	19	25	30	36	42	48	54	60	65	71	77	83														
17	15	20	26	33	39	45	51	57	64	70	77	83	89	96													
18	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109												
19	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123											
20	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138										
21	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146	154									
22	20	28	36	44	52	60	69	77	85	94	102	111	119	128	136	145	154	162	171								
23	21	29	37	46	55	63	72	81	90	99	107	116	125	134	143	152	161	170	180	189							
24	22	31	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169	179	188	198	207						
25	23	32	41	50	60	69	79	89	98	108	118	128	137	147	157	167	177	187	197	207	217	227					
26	24	33	43	53	62	72	82	93	103	113	123	133	143	154	164	174	185	195	206	216	226	237	247				
27	25	35	45	55	65	75	86	96	107	118	128	139	150	160	171	182	193	203	214	225	236	247	258	268			
28	26	36	47	57	68	79	89	100	111	122	133	144	156	167	178	189	200	212	223	234	245	257	268	279	291		
29	27	38	48	59	70	82	93	104	116	127	139	150	162	173	185	196	208	220	232	243	255	267	278	290	302	314	
30	28	39	50	62	73	85	96	108	120	132	144	156	168	180	192	204	216	228	240	252	265	277	289	301	313	326	338



*Таблица критерия Манна-Уитни ( $\alpha = 0,01$ )*

$n_1$ $n_2$	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
5	0	1																									
6	1	2	3																								
7	1	3	4	6																							
8	2	4	6	7	9																						
9	3	5	7	9	11	14																					
10	3	6	8	11	13	16	19																				
11	4	7	9	12	15	18	22	25																			
12	5	8	11	14	17	21	24	28	31																		
13	5	9	12	16	20	23	27	31	35	39																	
14	6	10	13	17	22	26	30	34	38	43	47																
15	7	11	15	19	24	28	33	37	42	47	51	56															
16	7	12	16	21	26	31	36	41	46	51	56	61	66														
17	8	13	18	23	28	33	38	44	49	55	60	66	71	77													
18	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88												
19	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101											
20	10	16	22	28	34	40	47	53	60	67	73	81	87	93	100	107	114										
21	10	16	22	29	35	42	49	56	63	70	77	84	91	98	105	113	120	127									
22	10	17	23	30	37	45	52	59	66	74	81	89	96	104	111	119	127	134	142								
23	11	18	25	32	39	47	55	62	70	78	86	94	102	109	117	125	133	141	150	158							
24	12	19	26	34	42	49	57	66	74	82	90	98	107	115	123	132	140	149	154	166	174						
25	12	20	27	35	44	52	60	69	77	86	95	103	112	121	130	138	147	156	165	174	183	192					
26	13	21	29	37	46	54	63	72	81	90	99	108	117	126	136	145	154	163	173	182	191	201	210				
27	14	22	30	39	48	57	66	75	85	94	103	113	122	132	142	151	161	171	180	190	200	209	219	229			
28	14	23	32	41	50	59	69	78	88	98	108	118	128	138	148	158	168	178	188	198	208	218	229	239	249		
29	15	24	33	42	52	62	72	82	92	102	112	123	133	143	154	164	175	185	196	206	217	227	238	249	259	270	
30	15	25	34	44	54	64	75	85	95	106	117	127	138	149	160	171	182	192	203	214	225	236	247	258	270	281	292

# Корреляция

Корреляция (англ. *correlation*) – взаимосвязь, соответствие, взаимозависимость, связь.

**Корреляционным отношением свойств** называют взаимную связь свойств.

Математическим методом выявления силы связей свойств является **анализ корреляции** (через **коэффициент корреляции**).

## Коэффициент корреляции ( $r$ )

**Коэффициент корреляции** – двумерная статистика (характеристика) об уровне связи ( $r$ ) и

**Пример:** уровне значимости ( $p$ ) связи между связанными свойствами.

Способности школьников понимать учителя статистически значимо связаны с их способностями понятно выражать свои мысли ( $r = 0,56; p < 0,05$ )

### **Свойства уровня связи переменных ( $r$ ):**

1. Уровень связи  $|r| \leq 1$  вычисляется для **связанных выборок**
2. Если  $r$  положительное число, то связь свойств **прямая**, то есть большему значению одного свойства соответствует большее значение другого.
3. Если  $r$  отрицательное число, то связь свойств **обратная**, то есть большему значению одного свойства соответствует меньшее значение другого.
4. Если  $r$  близко к нулю, то связь свойств **отсутствует**

Для вычисленного значения  $r$  устанавливается **уровень его статистической значимости**

# Корреляция

Для определения *уровня значимости связи* переменных используется **таблица критических значений коэффициентов корреляции**

**Таблица  $r$ -критерия Спирмена ( $r$ -критерия Пирсона)**

n	$\alpha$			n	$\alpha$			n	$\alpha$		
	0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01
5	0,805	0,878	0,959	20	0,378	0,444	0,561	35	0,283	0,334	0,430
6	0,729	0,811	0,917	21	0,369	0,433	0,549	36	0,279	0,329	0,424
7	0,669	0,754	0,875	22	0,360	0,423	0,537	37	0,275	0,325	0,418
8	0,621	0,707	0,834	23	0,352	0,413	0,526	38	0,271	0,320	0,413
9	0,582	0,666	0,798	24	0,344	0,404	0,515	39	0,267	0,316	0,408
10	0,549	0,632	0,765	25	0,337	0,396	0,505	40	0,264	0,312	0,403
11	0,521	0,602	0,735	26	0,330	0,388	0,496	42	0,257	0,304	0,393
12	0,497	0,576	0,708	27	0,323	0,381	0,487	44	0,251	0,297	0,384
13	0,476	0,553	0,684	28	0,317	0,374	0,479	46	0,246	0,291	0,374
14	0,458	0,532	0,661	29	0,311	0,367	0,471	50	0,235	0,279	0,361
15	0,441	0,514	0,641	30	0,306	0,361	0,463	60	0,214	0,254	0,330
16	0,426	0,497	0,623	31	0,301	0,355	0,452	70	0,198	0,235	0,306
17	0,412	0,482	0,606	32	0,296	0,349	0,449	80	0,185	0,220	0,286
18	0,400	0,468	0,590	33	0,291	0,344	0,442	90	0,174	0,207	0,270
19	0,389	0,456	0,575	34	0,287	0,339	0,436	100	0,165	0,197	0,256

## Шкала уровней связи переменных ( $r$ )

- $0,70 \leq r \leq 1$  – **сильная прямая** корреляция;
- $0,50 \leq r < 0,70$  – **средняя** прямая корреляция;
- $0,30 \leq r < 0,50$  – **умеренная** прямая корреляция;
- $0,20 \leq r < 0,30$  – **слабая** прямая корреляция;
- $-0,20 < r < 0,20$  – **корреляция отсутствует**;
- $-0,30 < r \leq -0,20$  – **слабая обратная** корреляция;
- $-0,50 < r \leq -0,30$  – **умеренная обратная** корреляция;
- $-0,70 < r \leq -0,50$  – **средняя обратная** корреляция;
- $-1 \leq r \leq -0,70$  – **сильная обратная** корреляция.

$n$  – объем выборки;  $\alpha$  – уровень значимости

## Шкала уровней значимости связи переменных ( $a$ )

- $0,10 < p$  – **связь** статистически **не значимая**;
- $0,05 < p \leq 0,10$  – связь статистически **значимая** (**тенденция**);
- $0,01 < p \leq 0,05$  – связь статистически значимая (**достоверная**);
- $p \leq 0,01$  – связь статистически значимая (**высокая**).

# r-критерий Спирмена

Область применения:

*r-критерий Спирмена* является критерием ранговой корреляции, который применяется для переменных (свойств), измеренных, как правило, в **шкале порядка**.

Алгоритм *r-критерий Спирмена*:

1. Заменяют варианты (значения) выборок *A* и *B* рангами  $r_A$  и  $r_B$ ;

2. Вычисляют значение  $D$  по формуле:

$$D = \sum (r_A - r_B)^2$$

3. Определяют коэффициент  $r$ :

$$r = 1 - \frac{6 * D}{n(n^2 - 1)} \quad n - \text{объем выборки.}$$

4. Находят в строке « $n$ » таблицы *r-критерия Спирмена* значения:  $r_{0,10}$ ;  $r_{0,05}$ ;  $r_{0,01}$ .

5. Статистический вывод:

Если  $|r| < r_{0,10}$ , то принимается гипотеза  $H_0$ .

Если  $r_{0,10} \leq |r| < r_{0,05}$ , то принимается гипотеза  $H_1$  ( $r, p \leq 0,10$ ).

Если  $r_{0,05} \leq |r| < r_{0,01}$ , то принимается гипотеза  $H_1$  ( $r, p \leq 0,05$ ).

Если  $r_{0,01} \leq |r|$ , то принимается гипотеза  $H_1$  ( $r, p \leq 0,01$ ).

# r-критерий Пирсона

## Область применения:

*r-критерий Пирсона* является критерием линейной корреляции, который применяется для переменных (свойств), измеренных в **шкале интервалов** или **шкале отношений**.

## Алгоритм r-критерий Пирсона :

1. Разместите таблицу выборок *A* и *B* в Excel.

2. Курсор поставьте на пустую ячейку таблицы.

3. Последовательно выполните операции:

- нажмите клавишу со знаком «fx»;

- в появившемся окне «Мастер функций» в ячейке «Поиск функции» наберите

**КОРРЕЛ**. Нажмите кнопку «Найти» и «ОК»;

- в появившееся окно «Аргументы функции» впишите:

- в строку «Массив 1» – код первой ячейки выборки *A*: код последней ячейки выборки *A*;

- в строку «Массив 2» – код первой ячейки выборки *B*: код последней ячейки выборки *B*, нажмите «ОК».

В ячейке появится значение ***r*** (уровня связи значений выборок *A* и *B*)

4. Находят в строке «*n*» таблицы *r-критерия Пирсона* значения:  $r_{0,10}$ ;  $r_{0,05}$ ;  $r_{0,01}$ .

## 5. Статистический вывод:

Если  $|r| < r_{0,10}$ , то принимается гипотеза  $H_0$ .

Если  $r_{0,10} \leq |r| < r_{0,05}$ , то принимается гипотеза  $H_1 (r, p \leq 0,10)$ .

Если  $r_{0,05} \leq |r| < r_{0,01}$ , то принимается гипотеза  $H_1 (r, p \leq 0,05)$ .

Если  $r_{0,01} \leq |r|$ , то принимается гипотеза  $H_1 (r, p \leq 0,01)$ .

# Z-критерий Фишера

**Область применения:**

Z-критерий Фишера является критерием, который применяется для сравнения коэффициентов корреляции двух переменных (свойств), полученных на разных выборках.

**Пример:**

Есть ли статистически значимые различия связей показателей общительности и академической успеваемости у студентов очной ( $r = 0,31, p \leq 0,05$ ) и

заочной ( $r = 0,63, p \leq 0,05$ ) форм обучения?

**Алгоритм Z-критерия Фишера:**

1. В Excel занесите вычисленные значения  $r_1, n_1, r_2, n_2$ , где  $r_1$  – коэффициент корреляции переменных в выборке I;  $n_1$  – число респондентов в выборке I;  $r_2$  – коэффициент корреляции переменных в выборке II;  $n_2$  – число респондентов в выборке II.

2. Найдите с помощью функции **ФИШЕР** в Excel значения:  $Z_1 = Z(r_1); Z_2 = Z(r_2)$ .

3. Вычислите Z по формуле:

$$Z = \frac{|Z_1 - Z_2|}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

4. Найдите с помощью функции **НОРМСТРАСП** в Excel значение  $P(Z)$

5. Вычислите уровень значимости  $\alpha$  по формуле:

$$\alpha = 2 * (1 - P(Z)).$$

6. Статистический

Если  $\alpha > 0,10$ , то принимается  $H_0$ .

Если  $\alpha \leq 0,10$ , то принимается  $H_1$  ( $p \leq \alpha$ ).

ВЫВОД:

# ЗАДАНИЯ

1. Найдите **моду** выборок по таблицам распределения частот:

Таблица А

Распределение частот исследования темперамента

Тип темперамента	Меланхолик	Сангвиник	Флегматик	Холерик
Частота	11	23	6	21

Таблица Б

Распределение частот исследования «Рост студентов»

Рост, см	162	164	165	168	169	170	176	177	181
Частота	3	6	11	3	4	9	11	3	1

Таблица В

Распределение оценок по политологии

Оценка	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично
Частота	5	5	5	5

2. В таблице представлено **распределение частот** оценок студентов по психологии:

Оценка, в баллах	51	58	63	66	68	69	70	71	75	80
Частота	1	3	5	9	10	11	8	7	4	2

- вычислите **среднее выборки** ( $m$ );
- вычислите **стандартное отклонение выборки** ( $s$ ).

3. В таблице представлены результаты тестирования школьников по истории в баллах.

Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл
АА	51	ДА	66	КА	68	ОА	69	РА	70	ТА	71
АВ	58	ДВ	66	КВ	68	ОВ	69	РВ	70	ТВ	71
БВ	58	ЕВ	66	КВ	68	ПВ	69	СВ	70	ФВ	71
БН	58	ЕН	66	КН	68	ПН	69	СН	70	ФН	71
ВИ	63	ЖИ	66	ЛИ	68	ПИ	69	СИ	70	ХИ	75
ВР	63	ЖР	66	МР	68	ПР	69	СР	70	ХР	75
ВТ	63	ЗТ	66	МТ	68	ПТ	69	СТ	70	ЦТ	75
ГП	63	ЗП	66	МП	68	ПП	69	СП	71	ЧП	75
ГН	63	ИН	68	НН	69	РН	69	ТН	71	ШН	80
ГЛ	66	КЛ	68	НЛ	69	РЛ	70	ТЛ	71	ЯЛ	80

- найдите **моду** выборки;
- вычислите **среднее выборки** ( $m$ );
- вычислите **стандартное отклонение выборки** ( $s$ ).

# ЗАДАНИЯ

4. Сформулируйте статистические гипотезы  $H_0$  и  $H_1$  о различиях выборок школьников:

- а) оценок по математике мальчиков и девочек;
- б) оценок школьников по литературе и по физике.

5. В таблице приведены распределения частот рангов ценности «Работа» у студентов до (I выборка) и после (II выборка) производственной практики.

Ранг ценности «Работа»	1	2	3	4	5	6
Частота I выборки	4	8	9	6	12	16
Частота II выборки	13	6	14	9	3	10

**Установите уровень статистической значимости** различий распределений частот рангов ценности «Работа» у студентов до и после производственной практики

6. По данным метеорологов вероятность прогноза погоды на один день равна 95%, на три дня – 90%.

**Определите вероятность ошибки** прогноза погоды: а) на один день; б) на три дня.



# ЗАДАНИЯ

7. В протоколе приведены самооценки студентов по психологии (А) и самооценки по математике (В).

Имя	А	В	Имя	А	В	Имя	А	В
Алексей	4	1	Рая	7	4	Николай	7	2
Борис	4	1	Роза	5	5	Олег	1	3
Вова	5	7	Тарас	8	6	Юля	8	7
Леонид	6	1	Света	5	5	Ольга	5	3
Марина	4	2	Сергей	7	5	Петр	6	3
Мария	5	2	Стас	7	6	Рита	6	3
Михаил	6	2	Голя	7	6	Роман	8	9
Настя	6	6	Таня	7	6	Таня	7	6
Юра	7	8	Федя	8	6	Вася	7	8

**Установите уровень статистической значимости** различий самооценок по психологии и самооценок по математике у студентов

8. В первой группе, состоящей из 24 студентов, у шести из них выявлен высокий уровень общительности. Во второй группе, состоящей из 28 студентов, у двенадцати выявлен высокий уровень общительности.

**Есть ли статистически значимые различия** долей студентов одной группы от студентов другой группы, имеющих высокий уровень общительности?

9. В таблице приведены результаты теста «Логические способности», проведенного среди школьников.

Оценка в баллах	17	19	20	21	22	24	25
Частота	6	3	8	1	5	3	2

**Проведите ранжирование вариант выборки.**

# ЗАДАНИЯ

10. В протоколе приведены результаты измерения *логических способностей (А)* и *уровней образного мышления (В)* у школьников.

Имя	А	В
Алексей	42	2,2
Борис	44	2,8
Леонид	18	5,2
Марина	16	4,8
Мария	14	5,6
Михаил	38	2,2
Настя	44	2,2

Имя	А	В
Николай	28	4,6
Олег	28	4,4
Ольга	44	2,6
Петр	36	2,2
Рита	44	2,2
Роман	24	4,2
Степан	24	6,0

Имя	А	В
Стас	28	2,2
Тарас	18	2,4
Татьяна	16	2,6
Ульяна	44	2,2
Федор	36	2,8
Юрий	34	5,2
Яна	38	4,6

**Найдите уровни связи и ее статистической значимости** показателей логических способностей и уровней образного мышления у школьников.