

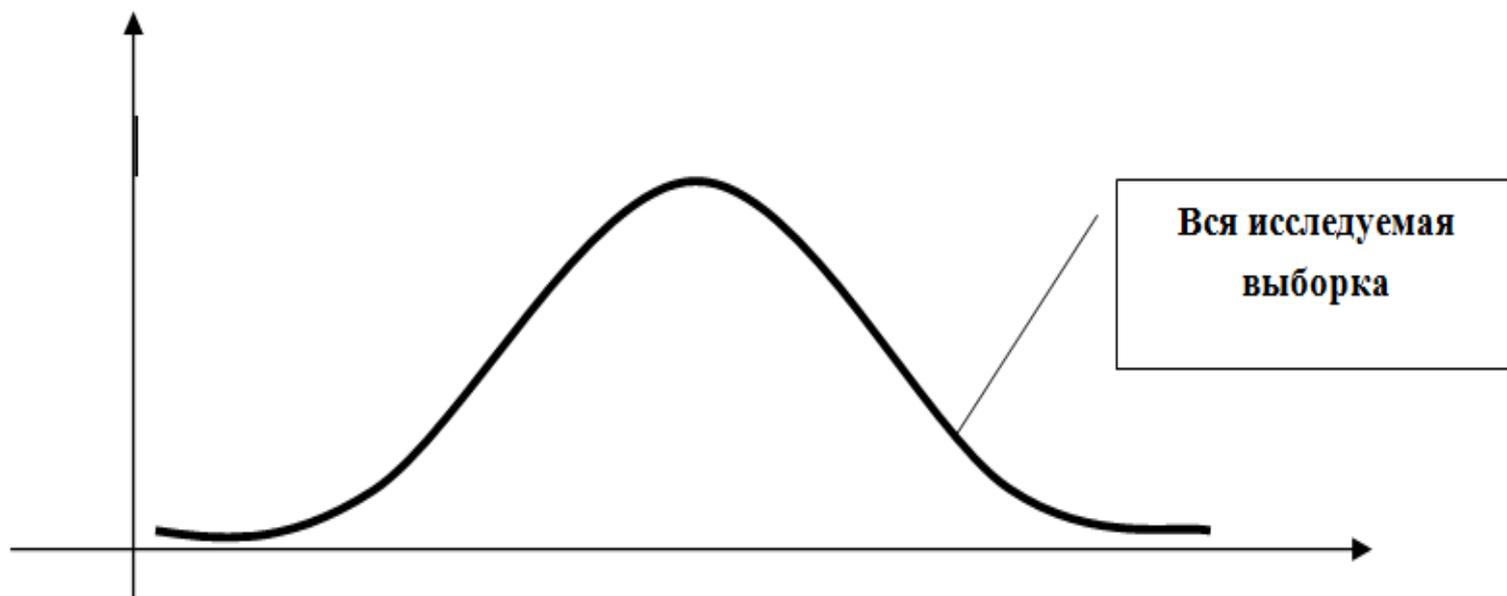


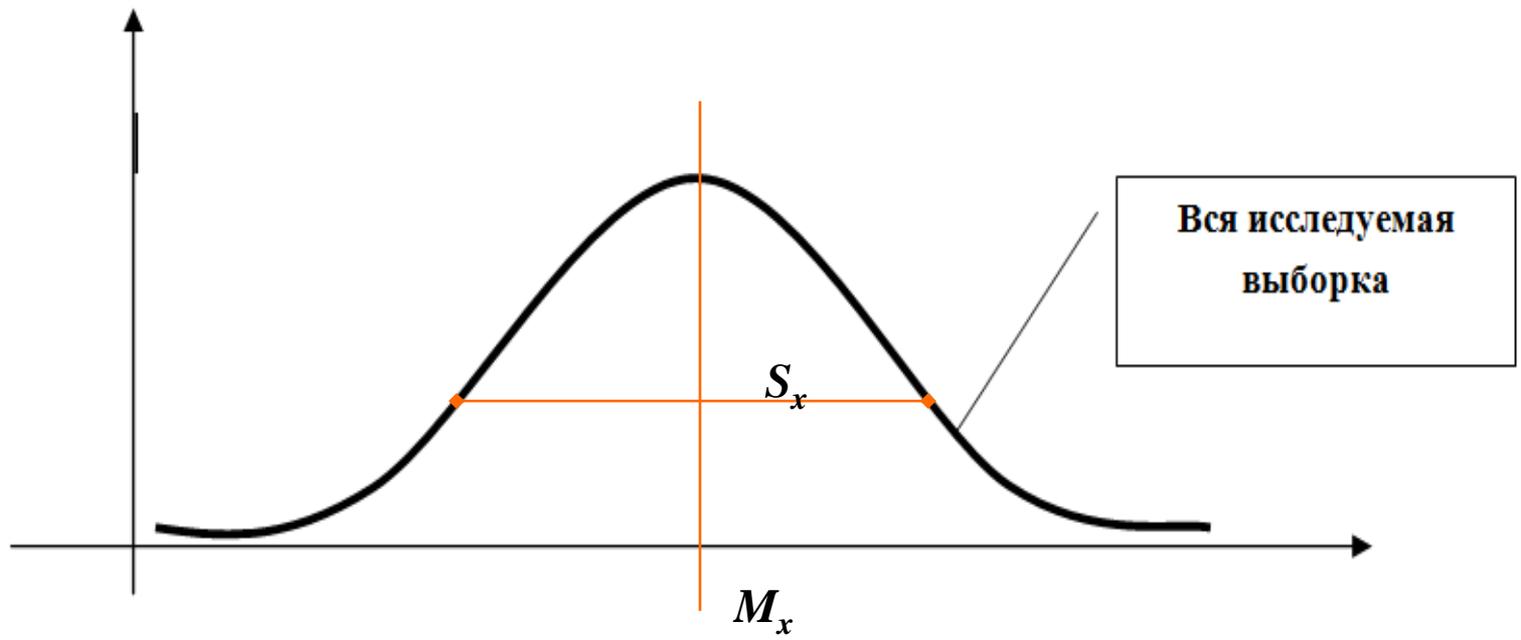
# **Однофакторный дисперсионный анализ**

**Назначение и общие понятия (ANOVA)  
однофакторного дисперсионного анализа**

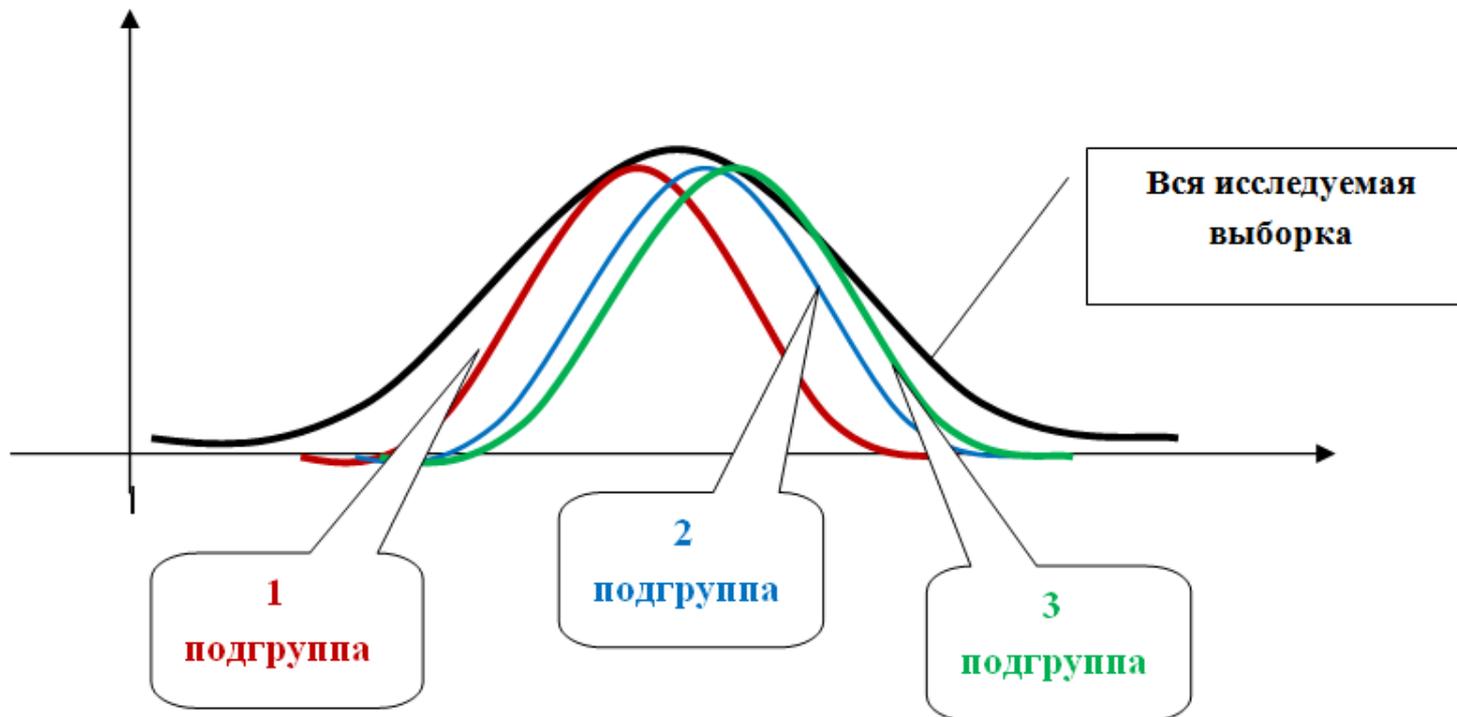
Дисперсионный анализ – это анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов. В зарубежной литературе дисперсионный анализ часто обозначается как ANOVA, что переводится как анализ вариативности (Analysis of Variance). Автором метода является Р. А. Фишер (Fisher RA.. 1918, 1938).

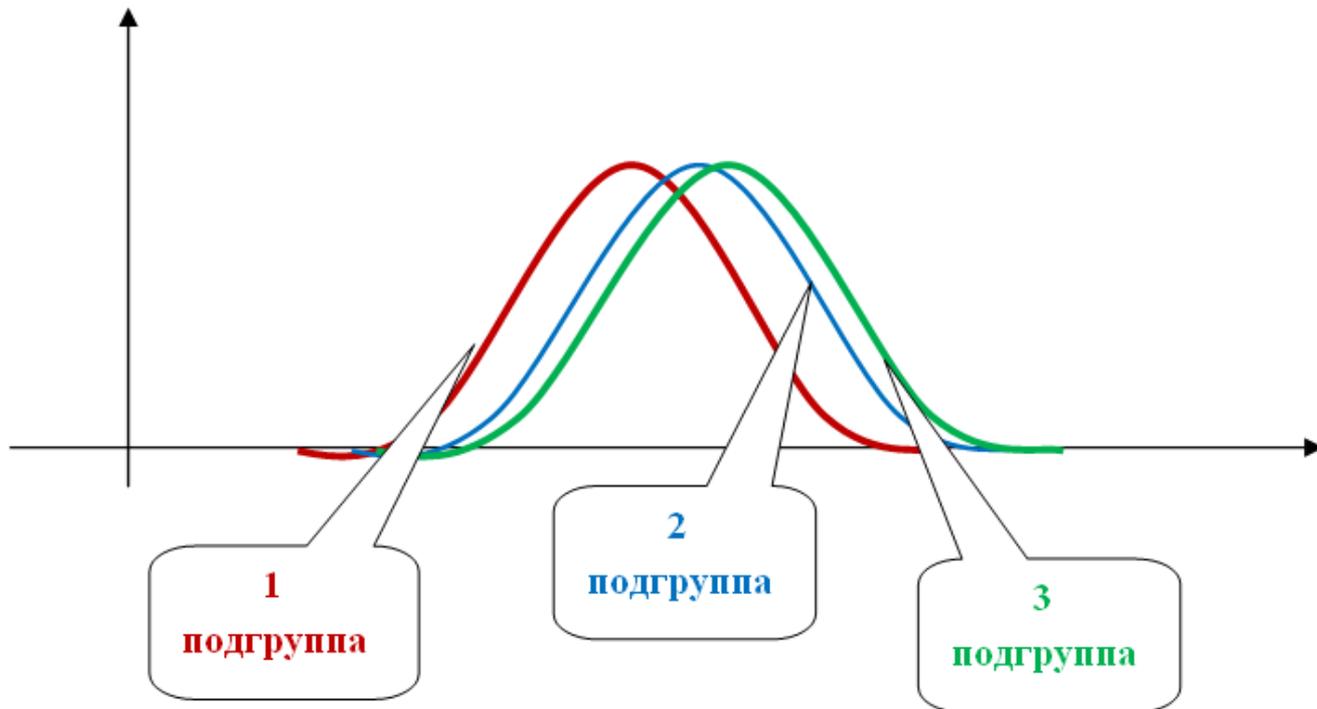




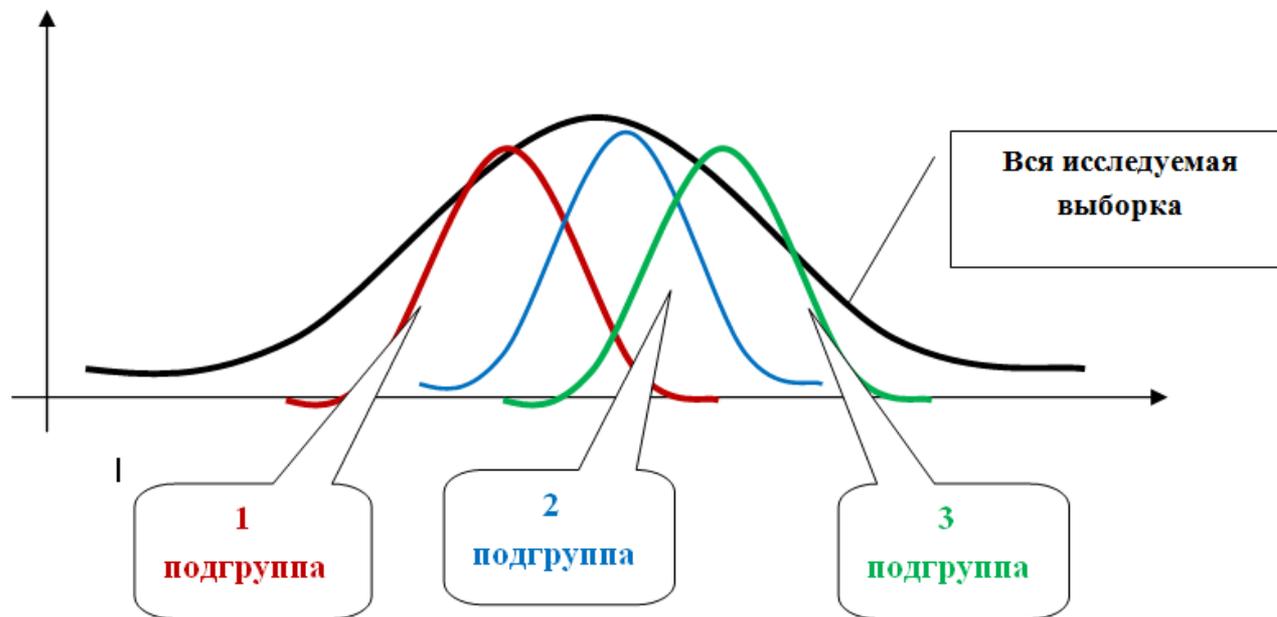


# ФАКТОР ДЕЛИТ ВСЮ ИССЛЕДУЕМУЮ ВЫБОРКУ НА ПОДГРУППЫ





# ФАКТОР ДЕЛИТ ВСЮ ИССЛЕДУЕМУЮ ВЫБОРКУ НА ПОДГРУППЫ



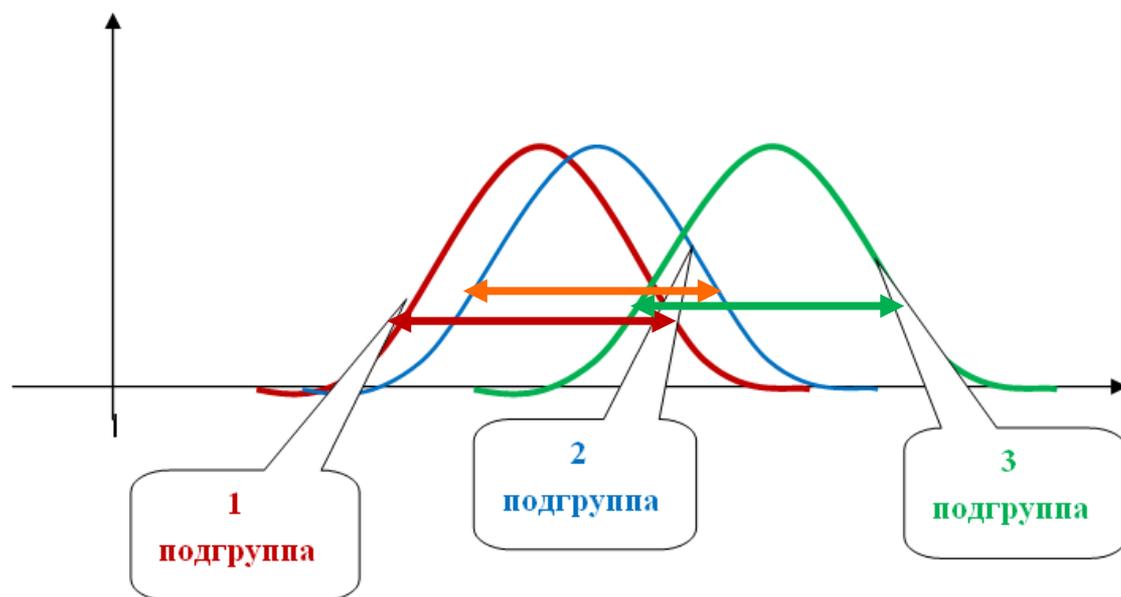
# ЗАДАЧА ДИСПЕРСИОННОГО АНАЛИЗА

состоит в том, чтобы из общей вариативности признака вычленить вариативность троякого рода:

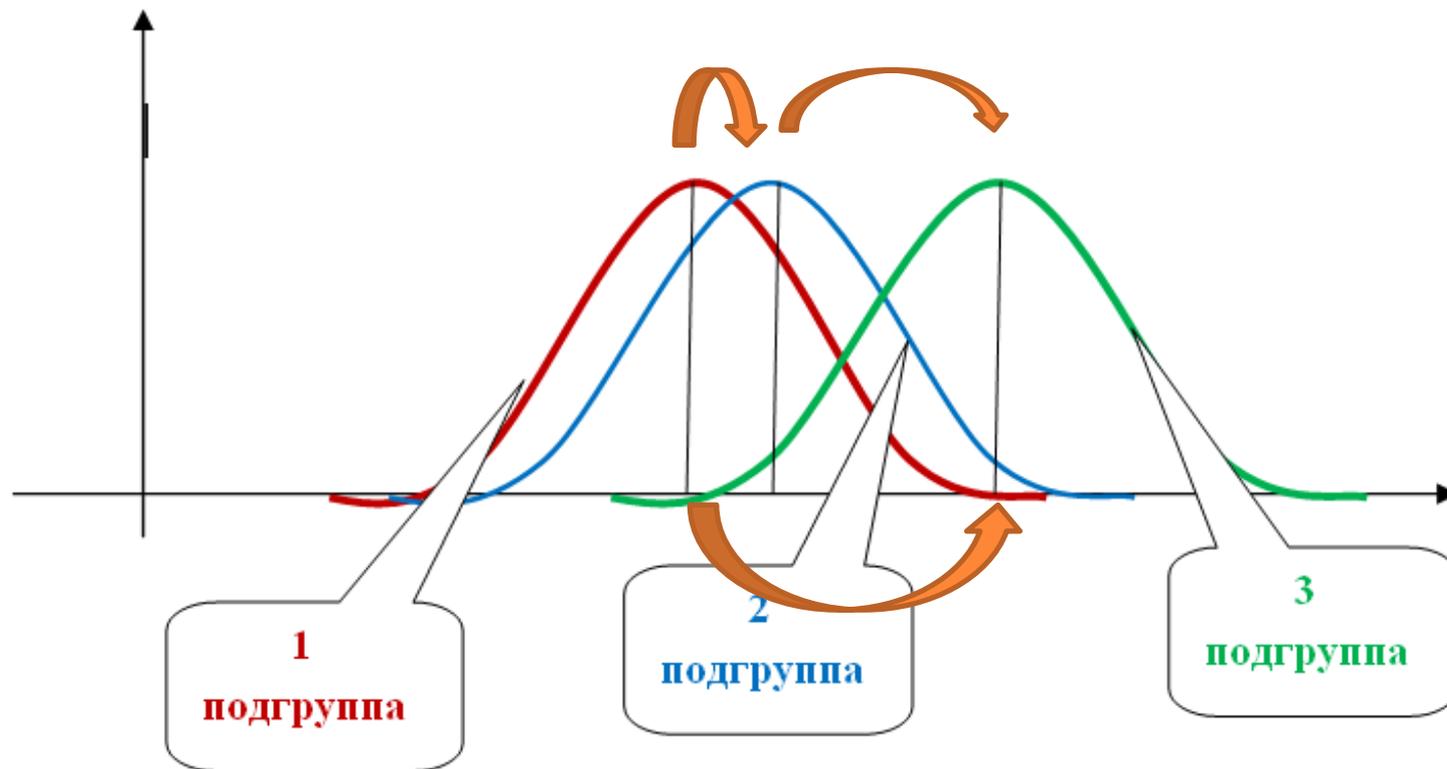
- а) вариативность, обусловленную действием каждой из исследуемых независимых переменных;
- б) вариативность, обусловленную взаимодействием исследуемых независимых переменных;
- в) случайную вариативность, обусловленную всеми другими неизвестными переменными.



# ФЭМП А= ВАРИАТИВНОСТЬ, ОБУСЛОВЛЕННАЯ ПЕРЕМЕННОЙ А СЛУЧАЙНАЯ ВАРИАТИВНОСТЬ



**ФЭМП Б= ВАРИАТИВНОСТЬ, ОБУСЛОВЛЕННАЯ ПЕРЕМЕННОЙ Б  
СЛУЧАЙНАЯ ВАРИАТИВНОСТЬ**



ГЭМП АБ=

ВАРИАТИВНОСТЬ, ОБУСЛОВЛЕННАЯ ВЗАИМОДЕЙСТВИЕМ ПЕРЕМЕННЫХ А И Б  
СЛУЧАЙНАЯ ВАРИАТИВНОСТЬ

---



В дисперсионном анализе исследователь **исходит из предположения**, что

одни переменные могут рассматриваться как **причины**, а другие – как следствия.

Переменные первого рода считаются **факторами**, а переменные второго рода – **результативными признаками**.

В этом отличие дисперсионного анализа от прямолинейного корреляционного анализа, в котором мы исходим из предположения, что изменения одного признака просто сопровождаются определенными изменениями другого.



два пути **разделения** всех исследуемых переменных на независимые переменные (**факторы**) и зависимые переменные (**результативные признаки**).

мы **совершаем** какие-либо **воздействия** на испытуемых или учитываем какие-либо не зависящие от нас воздействия на них, и **именно эти воздействия** считаем **независимыми** переменными, или **факторами**,

исследуемые признаки рассматриваем как **зависимые** переменные, или **результативные** признаки

мы, не совершая никаких воздействий, считаем, что **при разных уровнях развития** одних **психологических признаков** другие проявляются тоже **по-разному**

По тем или иным причинам мы решаем, что одни признаки могут рассматриваться скорее как **факторы**, а другие - как **результат** действия этих факторов

# ПОСЛЕДОВАТЕЛЬНОСТЬ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА

Однофакторный (One-Way) ANOVA позволяет проверить гипотезу о том, что изучаемый фактор оказывает влияние на зависимую переменную (средние значения, соответствующие разным градациям фактора, различаются).



# МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОДНОФАКТОРНОГО ANOVA

предполагает выделение в общей изменчивости зависимой переменной двух ее составляющих.

Межгрупповая (факторная) составляющая изменчивости обусловлена различием средних значений под влиянием фактора.

Внутригрупповая (случайная) составляющая изменчивости обусловлена влиянием неучтенных причин.

Соотношение первой и второй из указанных составляющих изменчивости и есть основной показатель, определяющий статистическую значимость влияния фактора (различия средних значений групп, соответствующих уровням фактора).



# МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОДНОФАКТОРНОГО ANOVA

- Нулевая статистическая гипотеза содержит утверждение о равенстве средних значений.
- При ее отклонении принимается альтернативная гипотеза о том, что, по крайней мере, два средних значения различаются.



## ИСХОДНЫЕ ПРЕДПОЛОЖЕНИЯ:

Распределение зависимой переменной в сравниваемых генеральных совокупностях характеризуется **нормальным законом** и одинаковыми дисперсиями.

Выборки являются случайными и **независимыми**.

Проверка исходных предположений сводится к проверке однородности дисперсии в сравниваемых выборках в случае, если они заметно различаются по численности.



# ПРИМЕР

## У старшеклассников проводили исследование стиль учебной деятельности

- Автономный стиль учебной деятельности подразумевает проявление развитого самоконтроля, уверенности в своих силах, стремление к независимому выполнению работы.
- Зависимый стиль учебной деятельности, напротив, предполагает либо отсутствие, либо малое количество решительных качеств, такие люди при исполнении работы полагаются на советы, рекомендации и подсказки со стороны

В анкете 1 из вопросов:

Где Вы чувствуете себя более уверенно при решении школьного задания:

- дома
- В школе
- нигде



автономность-зависимость		
дома	нигде	В школе
3	6	7
4	7	7
4	7	11
5	11	11
5	11	11
5	11	11
5	13	11
5	13	11
6	13	11
6	13	13
7	13	13
7		14
7		14
11		17
11		
11		
12		
12		
12		
12		
12		
16		

- **Вопрос:**  
связан ли стиль учебной деятельности с тем, где они чувствуют себя более уверенно?
- есть ли различия в стиле учебной у учеников с разным выбором ответа?
- **Фактор:** выбор ответа на вопрос:  
Где Вы чувствуете себя более уверенно при решении школьного задания:  
дома  
В школе  
нигде
- **Зависимая переменная :**  
стиль учебной деятельности (автономность-зависимость)



# РЕШЕНИЕ

Анализ данных

Инструменты анализа

- Однофакторный дисперсионный анализ
- Двухфакторный дисперсионный анализ с повторениями
- Двухфакторный дисперсионный анализ без повторений
- Корреляция
- Ковариация
- Описательная статистика
- Экспоненциальное сглаживание
- Двухвыборочный F-тест для дисперсии
- Анализ Фурье
- Гистограмма

	A	B	C
1	авто-зависимость		
2	дом	нет	школа
3	3	6	7
4	4	7	7
5	4	7	11
6	5	11	11
7	5	11	11
8	5	11	11
9	5	13	11
10	5	13	11
11	6	13	11
12	6	13	13
13	7	13	13
14	7		14
15	7		14
16	11		17
17	11		
18	11		
19	12		
20	12		
21	12		
22	12		
23	16		

Однофакторный дисперсионный анализ

Входные данные

Входной интервал:

Группирование:

- по столбцам
- по строкам

Метки в первой строке

Дельфа:

Параметры вывода

Выходной интервал:

- Новый рабочий лист:
- Новая рабочая книга

	A	B	C	D	E	F	G	H	I
1	авто-зависимость								
2	дом	нет	школа						
3	3	6	7						
4	4	7	7						
5	4	7	11						
6	5	11	11						
7	5	11	11						
8	5	11	11						
9	5	13	11						
10	5	13	11						
11	6	13	11						
12	6	13	13						
13	7	13	13						
14	7		14						
15	7		14						
16	11		17						
17	11								
18	11								
19	12								
20	12								
21	12								
22	12								
23	16								



## Однофакторный дисперсионный анализ

### ИТОГИ

<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>
дом	21	166	7,905	13,390
нет	11	118	10,727	7,618
школа	14	162	11,571	6,879

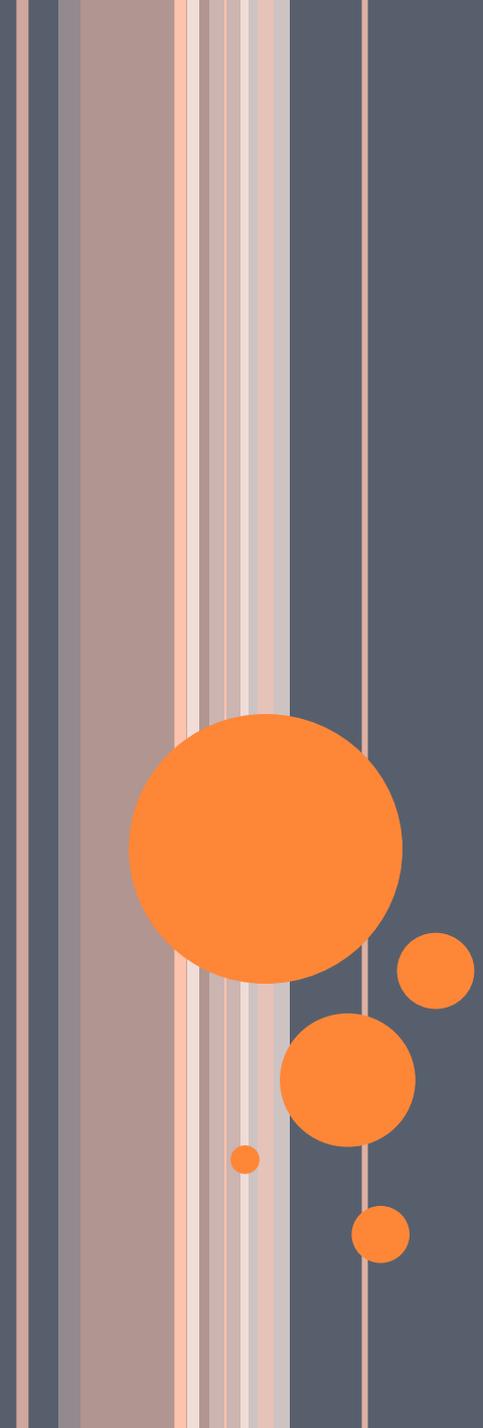
### Дисперсионный анализ

<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
Между группами	128,319	2	64,160	6,365	0,004	3,214
Внутри групп	433,419	43	10,080			
Итого	561,739	45				



- Ответ:
- Стиль учебной деятельности у школьников чувствующих себя более уверенно дома, в школе и нигде значимо различается (критерий  $F = 6,365$ , уровень значимости различий  $0,003$ ).





# СРАВНЕНИЕ НЕСКОЛЬКИХ ВЫБОРОК

2 способ

(если распределение не является нормальным)

# ДИСПЕРСИОННЫЙ АНАЛИЗ

- Исходным материалом для дисперсионного анализа служат данные исследования **трех и более выборок**, которые могут быть как равными, так и неравными по численности, как связными, так и несвязными.
- Сущность дисперсионного анализа заключается в расчленении общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак.
- Сравнивая компоненты дисперсии друг с другом посредством  $F$ - критерия Фишера, можно определить, какая доля общей вариативности результативного признака обусловлена действием регулируемых факторов.



# КРИТЕРИЙ КРУСКАЛА-УОЛЛИСА (KRUSKAL W.H.&WALLIS W.A.)

- является непараметрическим аналогом однофакторного дисперсионного анализа
- Данный критерий позволяет оценить различия между несколькими выборками по степени выраженности какого-либо признака. Критерий  $H$  применяется для оценки различий по степени выраженности анализируемого признака одновременно между тремя, четырьмя и более выборками
- Он позволяет выявить степень изменения признака в выборках, не указывая, однако, на направление этих изменений



- Критерии основан на том принципе, что чем меньше взаимопересечение выборок, тем выше уровень значимости Нэмп.
- в выборках может быть разное количество испытуемых.
- Работа с данными начинается с того, что все выборки условно объединяются по порядку встречающихся величин в одну выборку и значениям этой объединенной выборки проставляются ранги.



# 1. ФОРМУЛИРУЕМ ГИПОТЕЗЫ

- $H_0$ : Уровень исследуемого признака в выборках можно считать одинаковым (значимых различий нет).
- $H_a$ : Выборки значимо различаются по уровню исследуемого признака.



## 2. ВЫБИРАЕМ КРИТЕРИЙ

- Критерий H Краскала-Уоллиса
- 3 и более выборок.



### 3. ПОКАЗАТЕЛИ ВЫБОРКИ, НЕОБХОДИМЫЕ ДЛЯ ВЫЧИСЛЕНИЯ КРИТЕРИЯ

- $n_1, n_2, \dots, n_k$  – объём 1, 2 и т.д. выборок
- $k$  – количество выборок
- $N$  – общее количество человек во всех выборках ( $N = \sum n_i$ ).
- Проранжировать все измерения, объединив результаты двух выборок. Правила ранжирования следующие:
  - Наименьшему значению из всех присваивается ранг 1, наибольшему -  $N$ .
  - Если два или большее количество значений равны, то для них рассчитывается средний ранг.
  - Например, если три наименьших значения в выборке равны, то их средний ранг был бы  $(1+2+3)/3=2$ . А если равны 10-е и 11-е значения, то их средний ранг  $(10+11)/2=10,5$ .
  - Подсчитать сумму рангов отдельно для каждой  $R_i$  выборки. Проверить, совпадает ли общая сумма рангов с расчетной по формуле



## 4. ЭМПИРИЧЕСКОЕ ЗНАЧЕНИЕ КРИТЕРИЯ

$$H_{\text{эмп}} = \left[ \frac{12}{N * (N + 1)} * \sum \frac{R_i^2}{n_i} \right] - 3 * (N + 1)$$



## 5. ТАБЛИЧНЫЕ ЗНАЧЕНИЯ

- используется та же таблица, что и для критерия  $\chi^2$  (*хи-квадрат*) Пирсона
- Число степеней свободы  $\nu = k - 1$ , где  $k$  – количество сравниваемых групп.

Число степеней свободы	0,05	0,01
1	3,841	6,635
2	5,991	9,210
3	7,813	11,341
4	9,488	13,277
5	11,070	15,086
6	12,592	16,812
7	14,067	18,475
8	15,507	20,090
9	16,919	21,666
10	18,307	23,209
11	18,675	24,725
12	21,026	26,217
13	22,362	27,688
14	23,685	29,141
15	24,996	30,578
16	26,296	32,000
17	27,587	33,409
18	28,869	34,805
19	30,144	36,191



# КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ КРУСКАЛА-УОЛЛИСА 3 ВЫБОРКИ (K=3)

n1	n2	n3	H, 0,05	n1	n2	n3	H, 0,05	n1	n2	n3	H, 0,05
3	2	2	4,714	5	4	2	5.273	6	4	4	5.681
3	3	2	5,361	5	4	3	5,656	6	5	2	5,338
3	3	3	5,600	5	4	4	5.657	6	5	3	5,602
4	2	2	5,333	5	5	2	5,338	6	5	4	5,661
4	3	2	5,444	5	5	3	5,705	6	5	5	5,729
4	3	3	5,791	5	5	4	5,666	6	6	2	5,410
4	4	2	5,455	5	5	5	5,780	6	6	3	5,625
4	4	3	5,598	6	2	2	5.345	6	6	4	5,724
4	4	4	5,692	6	3	2	5.348	6	6	5	5,765
5	2	2	5,160	6	3	3	5,615	6	6	6	5,801
5	3	2	5.251	6	4	2	5.340	7	7	7	5,819
5	3	3	5,648	6	4	3	5.610	8	8	8	5,805



# КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ КРУСКАЛА-УОЛЛИСА

4 ВЫБОРКИ (K=4)

n1	n2	n3	n4	H, 0,05	n1	n2	n3	n4	H, 0,05
2	2	2	2	6.167	4	3	3	2	6.795
3	2	2	2	6.333	4	3	3	3	6.984
3	3	2	2	6.527	4	4	2	2	6.731
3	3	3	2	6.727	4	4	2	2	6.874
3	3	3	3	7.000	4	4	4	2	6,957
4	2	2	2	6.545	4	4	4	3	7.142
4	3	2	2	6.621	4	4	4	4	7.235



## 6. ВЫБОР ГИПОТЕЗЫ

- Если  $N_{эмн} < N_{0.05}$  верна  $H_0$ 
  - - различия не обнаружены
- Если  $N_{эмн} > N_{0.01}$  верна  $H_a$ 
  - - различия значимы (уровень значимости 0,01)
- Если  $N_{0.05} \leq N_{эмн} \leq N_{0.01}$  нужны дополнительные исследования.
  - различия значимы (уровень значимости 0,05)



## ПРИМЕР:

Проводилось изучение синдрома выгорания у работников административных служб.

В исследовании принимало участие 30 женщин, со стажем работы от 1 до 12 лет.

Группа исследуемых была разделена на 3 подгруппы в зависимости от стажа работы:

- 11 человек – стаж работы 1-3 года;
- 9 человек – стаж работы 5-7 лет;
- 10 человек – стаж работы 9-12 лет.

Эмоциональное выгорание исследовалось по методике Бойко В.В. «Диагностика уровня эмоционального выгорания».



№ п/п	Эмоциональное истощение, баллы		
	Группа 1 Стаж 1-3 г.	Группа2: Стаж 5-7л.	Группа3: Стаж 9-12л
1	19	48	71
2	24	30	56
3	11	52	65
4	12	47	63
5	33	38	59
6	28	57	72
7	38	37	55
8	21	31	67
9	25	21	81
10	20		64
11	27		



# 1. ФОРМУЛИРУЕМ ГИПОТЕЗЫ

- $H_0$ : Уровень эмоционального истощения в выборках женщин с разным стажем можно считать одинаковым (значимых различий нет).
- $H_a$ : Выборки женщин с разным стажем значимо различаются по уровню эмоционального истощения



## 2. ВЫБИРАЕМ КРИТЕРИЙ

- Критерий Н Краскала-Уоллиса
- 3 выборки.



### 3. ПОКАЗАТЕЛИ ВЫБОРКИ, НЕОБХОДИМЫЕ ДЛЯ ВЫЧИСЛЕНИЯ КРИТЕРИЯ

- $n_1=11$   $n_2=9$ , ...,  $n_k=10$
- $k=3$  – количество выборок
- $N=30$  – общее количество человек во всех выборках ( $N=\sum n_i$ ).



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
19		48		71	
24		30		56	
11		52		65	
12		47		63	
33		38		59	
28		57		72	
38		37		55	
21		31		67	
25		21		81	
20				64	
27					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11		21		55	
12		30		56	
19		31		59	
20		37		63	
21		38		64	
24		47		65	
25		48		67	
27		52		71	
28		57		72	
33				81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21		55	
12	2	30		56	
19	3	31		59	
20	4	37		63	
21		38		64	
24		47		65	
25		48		67	
27		52		71	
28		57		72	
33				81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	
12	2	30		56	
19	3	31		59	
20	4	37		63	
21	5,5	38		64	
24		47		65	
25		48		67	
27		52		71	
28		57		72	
33				81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	
12	2	30		56	
19	3	31		59	
20	4	37		63	
21	5,5	38		64	
24	7	47		65	
25	8	48		67	
27	9	52		71	
28	10	57		72	
33				81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	
12	2	30	11	56	
19	3	31	12	59	
20	4	37		63	
21	5,5	38		64	
24	7	47		65	
25	8	48		67	
27	9	52		71	
28	10	57		72	
33				81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	
12	2	30	11	56	
19	3	31	12	59	
20	4	37	14	63	
21	5,5	38		64	
24	7	47		65	
25	8	48		67	
27	9	52		71	
28	10	57		72	
33	13			81	
38					



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	20
12	2	30	11	56	21
19	3	31	12	59	
20	4	37	14	63	
21	5,5	38	15,5	64	
24	7	47	17	65	
25	8	48	18	67	
27	9	52	19	71	
28	10	57		72	
33	13			81	
38	15,5				



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	20
12	2	30	11	56	21
19	3	31	12	59	23
20	4	37	14	63	24
21	5,5	38	15,5	64	25
24	7	47	17	65	26
25	8	48	18	67	27
27	9	52	19	71	28
28	10	57	22	72	29
33	13			81	30
38	15,5				



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	20
12	2	30	11	56	21
19	3	31	12	59	23
20	4	37	14	63	24
21	5,5	38	15,5	64	25
24	7	47	17	65	26
25	8	48	18	67	27
27	9	52	19	71	28
28	10	57	22	72	29
33	13			81	30
38	15,5				



Группа 1 Стаж 1-3 г.		Группа2: Стаж 5-7л		Группа3: Стаж 9-12л	
баллы	ранг	баллы	ранг	баллы	ранг
11	1	21	5,5	55	20
12	2	30	11	56	21
19	3	31	12	59	23
20	4	37	14	63	24
21	5,5	38	15,5	64	25
24	7	47	17	65	26
25	8	48	18	67	27
27	9	52	19	71	28
28	10	57	22	72	29
33	13			81	30
38	15,5				
Сумма рангов	<b>78</b>	Сумма рангов	<b>134</b>	Сумма рангов	<b>253</b>

$$\sum R_i = \frac{N * (N + 1)}{2} \quad 78 + 134 + 253 = \frac{30 * 31}{2}$$



## 4. ЭМПИРИЧЕСКОЕ ЗНАЧЕНИЕ КРИТЕРИЯ

$$H_{\text{эмп}} = \left[ \frac{12}{N * (N + 1)} * \sum \frac{R_i^2}{n_i} \right] - 3 * (N + 1)$$



## 4. ЭМПИРИЧЕСКОЕ ЗНАЧЕНИЕ КРИТЕРИЯ

$$H_{\text{эмп}} = \left[ \frac{12}{N * (N + 1)} * \sum \frac{R_i^2}{n_i} \right] - 3 * (N + 1)$$
$$= \left[ \frac{12}{30 * 31} * \left( \frac{78^2}{11} + \frac{134^2}{9} + \frac{253^2}{10} \right) \right] - 3 * 31 = 22,472$$



## 5. ТАБЛИЧНЫЕ ЗНАЧЕНИЯ

- используется та же таблица, что и для критерия  $\chi^2$  (*хи-квадрат*) Пирсона
- Число степеней свободы  $\nu = k - 1 = 3 - 1 = 2$ , где  $k$  – количество сравниваемых групп.

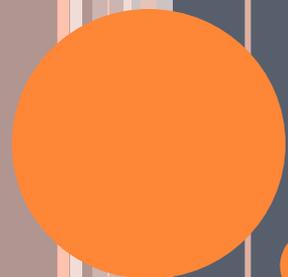
Число степеней свободы	0,05	0,01
1	3,841	6,635
<b>2</b>	<b>5,991</b>	<b>9,210</b>
3	7,813	11,341
4	9,488	13,277
5	11,070	15,086
6	12,592	16,812
7	14,067	18,475
8	15,507	20,090
9	16,919	21,666
10	18,307	23,209
11	18,675	24,725
12	21,026	26,217
13	22,362	27,688
14	23,685	29,141
15	24,996	30,578
16	26,296	32,000
17	27,587	33,409
18	28,869	34,805
19	30,144	36,191



○ Ответ:

*Группы работников административных служб с различным стажем работы достоверно ( $H = 22.472$ .  $p < 0,001$ ) различаются по такому компоненту эмоционального выгорания как «Истощение».*





# ВИДЫ ЗАДАЧ НА СРАВНЕНИЕ

Вид задачи	Условия	Метод решения
Выявление различия в уровне исследуемого признака	Две независимые выборки испытуемых	$t$ - критерий Стьюдента $U$ - критерий Манна-Уитни*
	три или более независимых выборок	Дисперсионный анализ Критерий Краскела-Уоллисса
Установление различия дисперсий признака	Две независимые выборки испытуемых	F - критерий Фишера
Оценка сдвига значений исследуемого признака	два замера на одной и той же выборке испытуемых	$t$ - критерий Стьюдента $T$ - критерий Вилкоксона для попарных сравнений
Сравнение относительных частот встречаемости признака	Две независимые выборки испытуемых	критерий $\phi^*$ - угловое преобразование Фишера $Z$ критерий
Выявление различий в распределении признаков	Сопоставление двух эмпирических распределений между собой	$\chi^2$ - критерий Пирсона

