



# **Практическое занятие №4 «Дисперсионный анализ»**

**по дисциплине «Многомерный  
статистический анализ в  
социологических  
исследованиях»**

# План занятия

1. Дисперсионный анализ.
2. Решение задач.

# Дисперсионный анализ

- **Дисперсионный анализ** — метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях.
- В отличие от t-критерия, позволяет сравнивать средние значения трёх и более групп.
- Разработан Р. Фишером для анализа результатов экспериментальных исследований.
- В литературе также встречается обозначение ANOVA (от англ. *ANalysis Of VAriance*)

# Дисперсионный анализ

- Цель дисперсионного анализа - исследование наличия или отсутствия существенного влияния какого-либо качественного или количественного фактора на изменения исследуемого результативного признака.
- Для этого фактор, предположительно имеющий или не имеющий существенного влияния, разделяют *на классы градации* (говоря иначе, группы) и выясняют, *одинаково ли влияние фактора* путём исследования значимости между средними в наборах данных, соответствующих градациям фактора.

# Почему дисперсионный анализ называется дисперсионным?

- При дисперсионном анализе исследуется отношение **двух дисперсий**. Дисперсия – это характеристика рассеивания данных вокруг среднего значения.
- **Первая** - дисперсия, объяснённая влиянием фактора, которая характеризует рассеивание значений между градациями фактора (группами) вокруг средней всех данных.
- **Вторая** - необъяснённая дисперсия, которая характеризует рассеивание данных внутри градаций (групп) вокруг средних значений самих групп.
- Первую дисперсию можно назвать межгрупповой, а вторую - внутригрупповой.
- Отношение этих дисперсий называется фактическим отношением Фишера и сравнивается с критическим значением отношения Фишера. Если фактическое отношение Фишера больше критического, то средние классов градации отличаются друг от друга и исследуемый фактор существенно влияет на изменение данных. Если меньше, то средние классов градации не отличаются друг от друга и фактор не имеет существенного влияния.

# Виды дисперсионного анализа

- Суть дисперсионного анализа сводится к изучению влияния одной или нескольких независимых переменных, обычно именуемых факторами, на зависимую переменную. Зависимые переменные представлены значениями абсолютных шкал (шкала отношений). Независимые переменные являются номинативными (шкала наименований), то есть отражают групповую принадлежность, и могут иметь два или более значения (типа, градации или уровня). Градации, соответствующие независимым выборкам объектов, называются межгрупповыми, а градации, соответствующие зависимым выборкам, — внутригрупповыми.

## **В зависимости от типа и количества переменных различают:**

- однофакторный и многофакторный дисперсионный анализ (одна или несколько независимых переменных);
- одномерный и многомерный дисперсионный анализ (одна или несколько зависимых переменных);
- дисперсионный анализ с повторными измерениями (для зависимых выборок);
- дисперсионный анализ с постоянными факторами, случайными факторами, и смешанные модели с факторами обоих типов

# Допущения дисперсионного анализа

**Исходными положениями дисперсионного анализа являются**

- нормальное распределение значений изучаемого признака в генеральной совокупности;
- равенство дисперсий в сравниваемых генеральных совокупностях;
- случайный и независимый характер выборки.

Нулевой гипотезой в дисперсионном анализе является утверждение о равенстве средних значений:

При отклонении нулевой гипотезы принимается альтернативная гипотеза о том, что не все средние равны, то есть имеются, по крайней мере, две группы, отличающиеся средними значениями

# Однофакторный дисперсионный анализ

- Простейшим случаем дисперсионного анализа является одномерный однофакторный анализ для двух или нескольких независимых групп, когда все группы объединены по одному признаку. В ходе анализа проверяется нулевая гипотеза о равенстве средних.
- **При анализе двух групп** дисперсионный анализ тождественен двухвыборочному  $t$ -критерию Стьюдента для независимых выборок, и величина  $F$ -статистики равна квадрату соответствующей  $t$ -статистики.
- Для подтверждения положения о равенстве дисперсий обычно применяется критерий Ливена (*Levene's test*). В случае отвержения гипотезы о равенстве дисперсий основной анализ неприменим. Если дисперсии равны, то для оценки соотношения межгрупповой и внутригрупповой изменчивости применяется  $F$ -критерий Фишера
- Если  $F$ -статистика превышает критическое значение, то нулевая гипотеза не может быть принята (отвергается) и делается вывод о неравенстве средних. При анализе средних двух групп результаты могут быть интерпретированы непосредственно после применения критерия Фишера.
- **При наличии трёх и более групп** требуется попарное сравнение средних для выявления статистически значимых отличий между ними.

# Многофакторный дисперсионный анализ

- Многофакторный анализ позволяет проверить влияние нескольких факторов на зависимую переменную.
- В отличие от однофакторной модели, где имеется одна межгрупповая сумма квадратов, модель многофакторного анализа включает суммы квадратов для каждого фактора в отдельности и суммы квадратов всех взаимодействий между ними. Так, в двухфакторной модели межгрупповая сумма квадратов раскладывается на сумму квадратов фактора А, сумму квадратов фактора В и сумму квадратов взаимодействия факторов А и В.



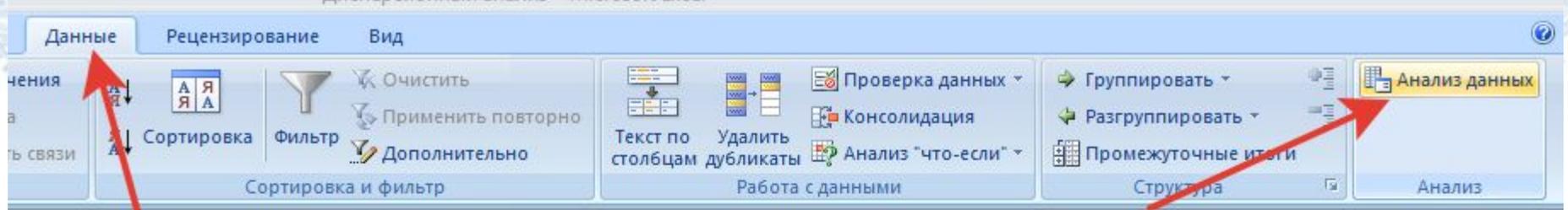
**Решение задач с помощью  
метода дисперсионного  
анализа**

# Однофакторный дисперсионный анализ

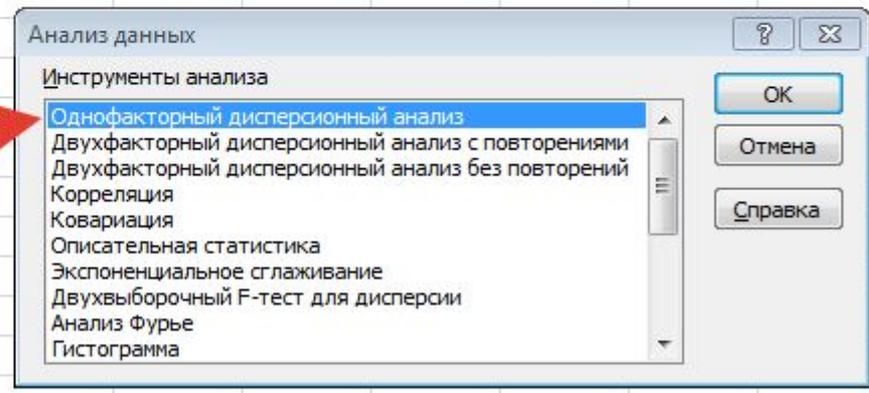
- ПРИМЕР
- Психолог фирмы проанализировал с помощью специальной методики стратегии поведения сотрудников в конфликтной ситуации.
- Предполагается, что на поведение влияет уровень образования (1 – среднее, 2 – среднее специальное, 3 – высшее).

# Данные имеют вид

	A	B	C	D	E	F	G
1	№	Образ.	Соперн.	Сотрудн.	Избег.	Присп.	Компром.
2	1	1	10	8	6	5	5
3	2	1	5	7	9	4	6
4	3	1	5	8	8	5	7
5	4	1	6	6	7	5	8
6	5	1	9	8	4	5	5
7	6	2	5	6	5	7	4
8	7	2	8	6	4	6	7
9	8	2	7	5	4	6	6
10	9	2	5	7	6	7	7
11	10	2	6	5	7	4	8
12	11	3	5	5	10	6	5
13	12	3	4	4	8	7	7
14	13	3	6	5	6	8	6
15	14	3	10	5	4	6	5
16	15	3	8	4	6	9	4



	E	F	G	H	I	J	K	L	M	N	O	P
тво	Избегание	Приспособление	Компромисс									
	6	5	5									
	9	4	6									
	8	5	7									
	7	5	8									
	4	5	5									
	5	7	4									
	4	6	7									
	4	6	6									
	6	7	7									
	7	4	8									
	10	6	5									
	8	7	7									



В	С	D	E	F	G	H	I	J	K	L	M	N
Образование	Соперничество	Сотрудничество	Избегание	Приспособление	Компромисс							
1	10	8	6	5	5							
1	5	7	9	4	6							
1	5	8	8	5	7							
1	6	6	7	5	8							
1	9	8	4	5	5							
2	5	6	5	7	4							
2	8	6	4	6	7							
2	7	5	4	6	6							
2	5	7	6	7	7							
2	6	5	7	4	8							
3	5	5	10	6	5							
3	4	4	8	7	7							
3	6	5	6	8	6							
3	10	5	4	6	5							
3	8	4	6	9	4							

Однофакторный дисперсионный анализ

Входные данные  
Входной интервал:

Группирование:  
 по столбцам  
 по строкам

Метки в первой строке  
Альфа:

Параметры вывода  
 Выходной интервал:  
 Новый рабочий дист:  
 Новая рабочая книга

OK  
Отмена  
Справка

Выбираем весь диапазон данных

Однофакторный дисперсионный анализ

ИТОГИ

Группы	Счет	Сумма	Среднее	Дисперсия
Столбец 1	15	30	2	0,714285714
Столбец 2	15	99	6,6	3,828571429
Столбец 3	15	89	5,933333	1,923809524
Столбец 4	15	94	6,266667	3,638095238
Столбец 5	15	90	6	2
Столбец 6	15	90	6	1,714285714

Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	220,9333	5	44,18667	19,18511371	0,00000000000011	2,323126496
Внутри групп	193,4667	84	2,303175			
Итого	414,4	89				

- Уровень значимости (р-значение) больше 0,05.
- Если эмпирическое значение больше ли равно критическому, то принимается альтернативная гипотеза, если меньше - то нулевая.
- Поведение в конфликтной ситуации зависит от уровня образования.

# Задача 1

- С помощью однофакторного дисперсионного анализа выявить, влияет ли уровень адаптационных способностей на развитие личностных качеств.
- (Переменную «адаптационные способности» перекодировать в три уровня градации: 7-8 баллов – высокий, 6 баллов – средний, 1-3 балла – низкий)

## Задача 2

С помощью однофакторного дисперсионного анализа выявить, влияет ли

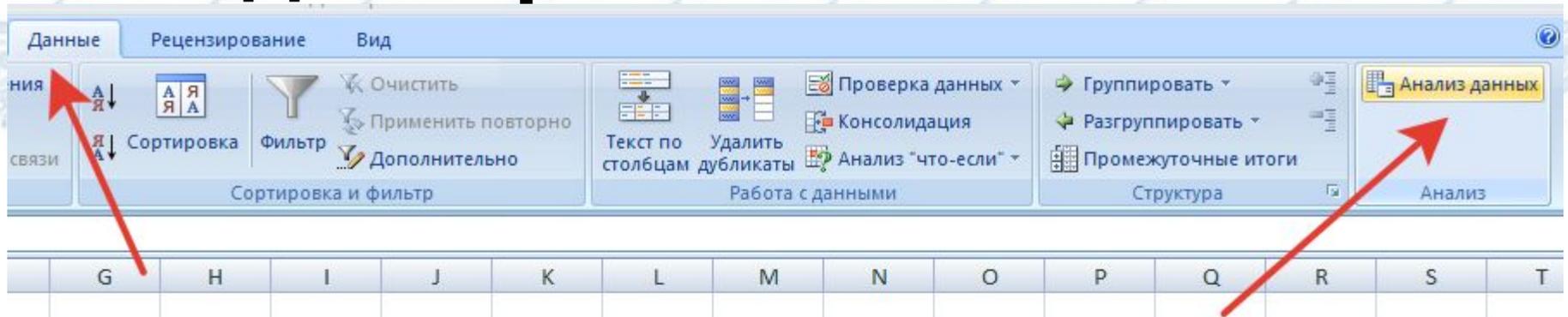
- возраст молодой семьи на доход семьи;
- возраст молодой семьи – на удовлетворенность доходом;
- пол респондента на оценку дохода семьи;
- пол респондента – на удовлетворенность доходом.

# Двухфакторный дисперсионный анализ

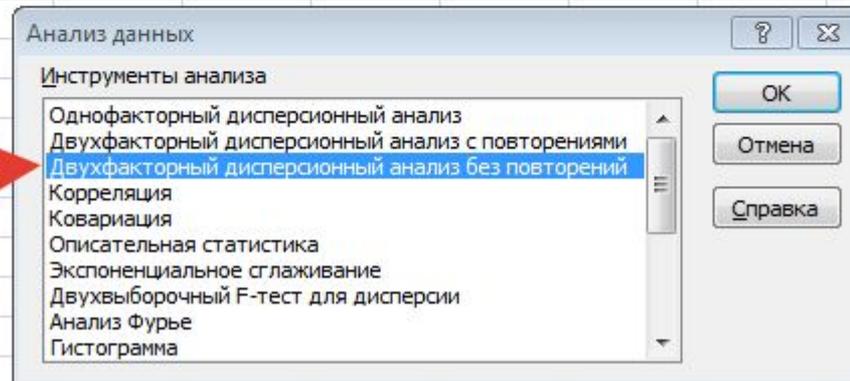
- Показывает, как влияет два фактора на изменение значения случайной величины.
- ПРИМЕР
- У мужчин и женщин с разным уровнем образования изучали уровень дохода. Влияют ли два данных фактора на уровень материал

	образование		
пол	среднее	СПО	ВО
мужчины	21 523	22 545	35 500
женщины	19 150	22 122	33 453

# Двухфакторный дисперсионный анализ



	образование		
пол	среднее	СПО	ВО
мужчины	21 523	22 545	35 500
женщины	19 150	22 122	33 453



## Двухфакторный дисперсионный анализ без повторений

ИТОГИ	Счет	Сумма	Среднее	Дисперсия
Строка 1	3	79568	26522,67	60705506,33
Строка 2	3	74725	24908,33	56966692,33
Столбец 1	2	40673	20336,5	2815564,5
Столбец 2	2	44667	22333,5	89464,5
Столбец 3	2	68953	34476,5	2095104,5

### Дисперсионный анализ

Фактор вариации	SS	df	MS	F	P-Значение	F критическое
Строки	3909108	1	3909108	7,165934735	0,115805022	18,51282051
Столбцы	2,34E+08	2	1,17E+08	214,7093792	0,004635867	19
Погрешность	1091025	2	545512,7			

Уровень значимости (p-значение) больше 0,05.

Если эмпирическое значение больше или равно критическому, то принимается альтернативная гипотеза, если меньше - то нулевая. Следовательно, из двух факторов решающим является фактор образования.

Доход не зависит от пола.

Доход зависит от уровня образования.

# Задача 3

- С помощью двухфакторного дисперсионного анализа определить степень влияния факторов пола и возраста на общую удовлетворенность жизнью.
- (Промежуточные операции: нужно подсчитать среднее арифметическое в группах отдельно по полу, отдельно по возрасту; составить таблицу 2 на 2 по количеству градаций признаков).

## Задача 4

- С помощью двухфакторного дисперсионного анализа определить степень влияния факторов девиантности и уровня конфликтности на выбор стратегии поведения в конфликте у подростков.

# Задача 5

- С помощью двухфакторного дисперсионного анализа определить степень влияния факторов пола и успеваемости на развитие организаторских способностей старшеклассников.
- (Промежуточные операции: перекодировать переменную «средний балл» в переменную «успеваемость» в соответствии с примечанием)



**Благодарю за работу!**