



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение высшего  
образования

Российский химико-технологический университет имени Д. И. Менделеева

---



## Дополнительные главы математики

**Лекция № 5 на тему:**

**«Основы дисперсионного анализа»**

Лектор: кандидат технических наук  
доцент кафедры высшей математики

**Гордеева Елена Львовна**

2020 г.

# Разделы лекции

1. Основные понятия дисперсионного анализа
2. Однофакторный дисперсионный анализ
3. Двухфакторный дисперсионный анализ без повторений
4. Двухфакторный дисперсионный анализ с повторениями
5. Ранговые методы анализа
  - 5.1. Однофакторный анализ Краскела-Уоллиса
  - 5.2. Двухфакторный анализ Фридмана

# 1. Основные понятия дисперсионного анализа

**Дисперсионный анализ** – статистический метод, предназначенный для оценки влияния различных факторов на результаты эксперимента, а также для последующего планирования аналогичных экспериментов.

**Фактор (факторы)** – то, что, как предполагается, должно оказывать влияние на конечный результат.

Примеры: способ крепления детали при её обработке; режим функционирования прибора; технология производства продукта; методика обучения и т. д. Факторы носят *качественный* характер.

**Уровни фактора (градации)** – конкретная реализация фактора. Например, если требуется выяснить, какой вид удобрений наиболее эффективен для получения наибольшего урожая, то фактор – удобрение, а его уровни – виды удобрений.

**Отклик** – значения измеряемого признака, то есть величина результата. Например, урожайность при использовании различных видов удобрений.

**Задача дисперсионного анализа** – по результатам наблюдений над случайной величиной  $Y$  оценить зависимость её математического ожидания (среднего) от рассматриваемых факторов.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Модели дисперсионного анализа

Два вида факторов: со случайными уровнями и с фиксированными.

## 1. Модель с фиксированными уровнями факторов (модель I)

Все уровни фиксированы (остаются одними и теми же). Выводы применимы только к тем уровням факторов, которые привлекались для исследования.

## 2. Модель со случайными уровнями факторов (модель II)

Выбор уровней производится из бесконечной совокупности возможных уровней и сопровождается рандомизацией. Выводы по эксперименту можно распространить на всю генеральную совокупность.

## 3. Модель смешанного типа (модель III)

Часть факторов рассматривается на фиксированных уровнях, а уровни остальных выбираются случайным образом.

## Допущения дисперсионного анализа:

- 1) случайные ошибки эксперимента имеют нормальный закон распределения с нулевым средним и неизвестной дисперсией;
- 2) случайные ошибки эксперимента не коррелированы между собой и имеют одинаковую (неизвестную) дисперсию (эксперименты равноточны).

В зависимости от числа источников дисперсии различают **однофакторный** (*ANOVA – Analysis of Variance*) и **многофакторный дисперсионный анализ** (*MANOVA – Multivariate Analysis of Variance*).

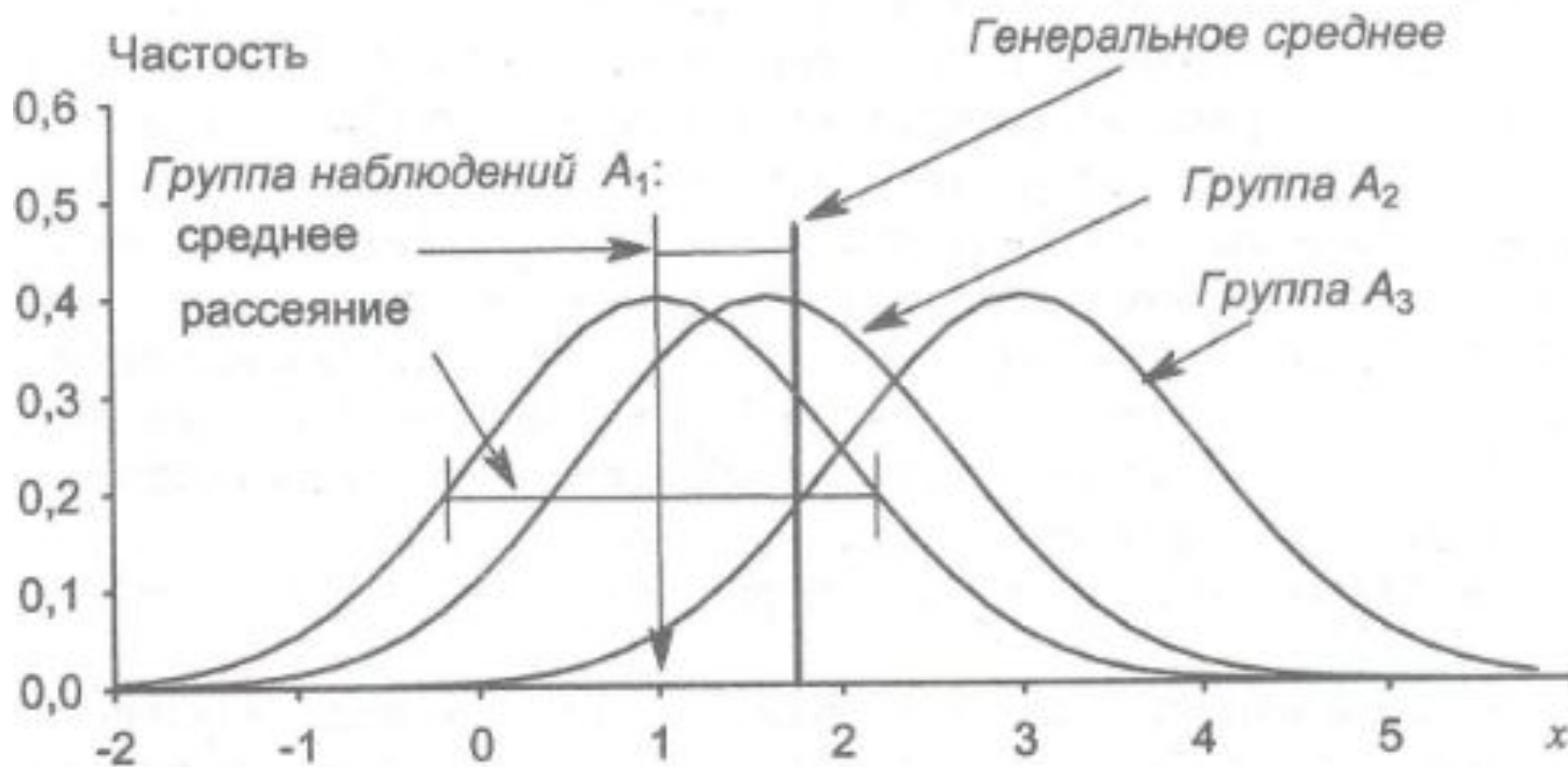
## 2. Однофакторный дисперсионный анализ

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

*Исходные данные для однофакторного анализа*

Уровень фактора $A$	Результативный признак $Y$
$A_1$	$y_{11}, y_{12}, \dots, y_{1n_1}$
$A_2$	$y_{21}, y_{22}, \dots, y_{2n_2}$
...	...
$A_m$	$y_{m1}, y_{m2}, \dots, y_{mn_m}$

# Влияние уровней фактора на распределение случайной величины





# Математическая модель

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Расчётные формулы

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Схема однофакторного дисперсионного анализа

Источник дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты
Фактор А	$SS_{\text{факт}} = \sum_{i=1}^m (\bar{y}_i - \bar{y})^2 \cdot n_i$	$m - 1$	$S_{\text{факт}}^2 = \frac{SS_{\text{факт}}}{m - 1}$
Остаток	$SS_{\text{ост}} = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$N - m$	$S_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{N - m}$
Общая сумма	$SS_{\text{общ}} = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$N - 1$	$S_{\text{общ}}^2 = \frac{SS_{\text{общ}}}{N - 1}$

# Проверка гипотезы в однофакторном дисперсионном анализе

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Пример 1

Исследовалось влияние трёх видов катализаторов  $A_1$ ,  $A_2$ ,  $A_3$  на выход целевого продукта (в т). Данные представлены в таблице:

Вид катализатора	Результаты наблюдений $y_{ij}$ – выход продукта в т				
$A_1$	3,2	3,1	3,1	2,8	
$A_2$	2,6	3,1	2,7	2,9	2,7
$A_3$	3,7	3,4	3,2	3,3	3,5

Требуется выяснить, влияет ли вид катализатора на выход продукта. Принять уровень значимости  $\alpha = 0,05$ .



# Решение примера 1

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

Оценки дисперсий:

$$S_{\text{факт}}^2 = \frac{SS_{\text{факт}}}{m - 1} = \frac{0,972}{3 - 1} = 0,486;$$

$$S_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{N - m} = \frac{0,398}{14 - 3} = 0,036.$$

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Решение примера 1 в MS Excel (инструмент «Однофакторный дисперсионный анализ»)

	A	B	C	D	E	F	G
1	Вид катализатора	Результаты наблюдений $y_{ij}$					
2							
3	A1	3,2	3,1	3,1	2,8		
4	A2	2,6	3,1	2,7	2,9	2,7	
5	A3	3,7	3,4	3,2	3,3	3,5	
6							
7	Однофакторный дисперсионный анализ						
8							
9	ИТОГИ						
10	<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
11	A1	4	12,2	3,05	0,03		
12	A2	5	14	2,8	0,04		
13	A3	5	17,1	3,42	0,037		
14							
15							
16	Дисперсионный анализ						
17	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
18	Между группами	0,971	2	0,486	13,422	0,001	3,982
19	Внутри групп	0,398	11	0,036			
20							
21	Итого	1,369	13				

## Линейные контрасты

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Пример 2

• **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

• **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

### 3. Двухфакторный дисперсионный анализ без повторений

● **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

Уровни фактора A	Уровни фактора B						$\bar{y}_{i*}$
	$B_1$	$B_2$	...	$B_j$	...	$B_m$	
$A_1$	$y_{11}$	$y_{12}$	...	$y_{1j}$	...	$y_{1m}$	$\bar{y}_{1*}$
$A_2$	$y_{21}$	$y_{22}$	...	$y_{2j}$	...	$y_{2m}$	$\bar{y}_{2*}$
...	...	...	...	...	...	...	...
$A_i$	$y_{i1}$	$y_{i2}$	...	$y_{ij}$	...	$y_{im}$	$\bar{y}_{i*}$
...	...	...	...	...	...	...	...
$A_n$	$y_{n1}$	$y_{n2}$	...	$y_{nj}$	...	$y_{nm}$	$\bar{y}_{n*}$
$\bar{y}_{*j}$	$\bar{y}_{*1}$	$\bar{y}_{*2}$	...	$\bar{y}_{*j}$	...	$\bar{y}_{*m}$	$\bar{y}$

Требуется оценить влияние факторов на результат.



# Математическая модель

● **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Схема двухфакторного дисперсионного анализа без повторений

Источник дисперсии	Суммы квадратов	Число степеней свободы	Средние квадраты
Фактор А	$SS_A = m \cdot \sum_{i=1}^n (\bar{y}_{i*} - \bar{y})^2$	$n - 1$	$S_A^2 = \frac{SS_A}{n - 1}$
Фактор В	$SS_B = n \cdot \sum_{j=1}^m (\bar{y}_{*j} - \bar{y})^2$	$m - 1$	$S_B^2 = \frac{SS_B}{m - 1}$
Остаток	$SS_{\text{ост}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y})^2$	$(n - 1)(m - 1)$	$S_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{(n - 1)(m - 1)}$
Общая сумма	$SS_{\text{общ}} = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y})^2$	$N - 1$	$S_{\text{общ}}^2 = \frac{SS_{\text{общ}}}{N - 1}$

## Гипотезы в двухфакторном дисперсионном анализе без повторений

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Проверка гипотез в двухфакторном дисперсионном анализе без повторений

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Пример 3

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних

<i>A</i> \ <i>B</i>	<i>B</i> <sub>1</sub>	<i>B</i> <sub>2</sub>	<i>B</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	1	2	3
<i>A</i> <sub>2</sub>	5	6	10

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Решение примера 3

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Решение примера 3 (продолжение)

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## 4. Двухфакторный дисперсионный анализ с повторениями

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

Уровни фактора $A$	Уровни фактора $B$						$\bar{y}_{i*}$
	$B_1$	$B_2$	...	$B_j$	...	$B_m$	
$A_1$	$y_{111}, y_{112}, \dots, y_{11k}$	$y_{121}, y_{122}, \dots, y_{12k}$	...	$y_{1j1}, y_{1j2}, \dots, y_{1jk}$	...	$y_{1m1}, y_{1m2}, \dots, y_{1mk}$	$\bar{y}_{1**}$
$A_2$	$y_{211}, y_{212}, \dots, y_{21k}$	$y_{221}, y_{222}, \dots, y_{22k}$	...	$y_{2j1}, y_{2j2}, \dots, y_{2jk}$	...	$y_{2m1}, y_{2m2}, \dots, y_{2mk}$	$\bar{y}_{2**}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	...
$A_i$	$y_{i11}, y_{i12}, \dots, y_{i1k}$	$y_{i21}, y_{i22}, \dots, y_{i2k}$	...	$y_{ij1}, y_{ij2}, \dots, y_{ijk}$	...	$y_{im1}, y_{im2}, \dots, y_{imk}$	$\bar{y}_{i**}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
$A_n$	$y_{n11}, y_{n12}, \dots, y_{n1k}$	$y_{n21}, y_{n22}, \dots, y_{n2k}$	...	$y_{nj1}, y_{nj2}, \dots, y_{nj k}$	...	$y_{nm1}, y_{nm2}, \dots, y_{nmk}$	$\bar{y}_{n**}$
$\bar{y}_{*j*}$	$\bar{y}_{*1*}$	$\bar{y}_{*2*}$	...	$\bar{y}_{*j*}$	...	$\bar{y}_{*m*}$	$\bar{y}$



# Математическая модель

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Схема двухфакторного дисперсионного анализа с повторениями

Источник дисперсии	Суммы квадратов	Число степеней свободы	Средние квадраты
Фактор А	$SS_A = mk \sum_{i=1}^n (\bar{y}_{i**} - \bar{y})^2$	$n - 1$	$S_A^2 = \frac{SS_A}{n - 1}$
Фактор В	$SS_B = nk \sum_{j=1}^m (\bar{y}_{*j*} - \bar{y})^2$	$m - 1$	$S_B^2 = \frac{SS_B}{m - 1}$
Взаимодействие	$SS_{AB} = k \sum_{i=1}^n \sum_{j=1}^m (\bar{y}_{ij*} - \bar{y}_{i**} - \bar{y}_{*j*} + \bar{y})^2$	$(n - 1)(m - 1)$	$S_{AB}^2 = \frac{SS_{AB}}{(n - 1)(m - 1)}$
Остаток	$SS_{\text{ост}} = \sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k (y_{iju} - \bar{y}_{ij*})^2$	$nm(k - 1)$	$S_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{nm(k - 1)}$
Общая сумма	$SS_{\text{общ}} = \sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k (y_{iju} - \bar{y})^2$	$N - 1$	$S_{\text{общ}}^2 = \frac{SS_{\text{общ}}}{N - 1}$

# Гипотезы в двухфакторном дисперсионном анализе с повторениями

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## Пример 4

Исследовалось влияние на процесс органического синтеза двух типов растворителей (A1 и A2) и трёх галогеналкилов (B1, B2 и B3). При каждом сочетании типа растворителя и галогеналкила сделано четыре параллельных опыта. Результаты (выход полимера в процентах) представлены в таблице. Требуется оценить значимость влияния типа растворителя и галогеналкила на процесс синтеза при уровне значимости  $\alpha=0,05$ .

Растворитель	Галогеналкил		
	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
A <sub>1</sub>	14,8	14,4	12,4
	15,9	15,4	12,9
	14,8	15,3	12,8
	14,9	14,5	11,9
A <sub>2</sub>	14,3	17,5	16,9
	15,1	16,8	18,8
	14,7	17,3	16,8
	15,9	15,2	17,9

## Решение примера 4

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

## 5. Однофакторный анализ Краскела-Уоллиса

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Вычисление критерия Краскела-Уоллиса при наличии связанных рангов

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{набл} < F_{кр}$ , то влияние фактора считается незначимым. Если же  $F_{набл} > F_{кр}$  то фактор влияет на изменчивость средних

$$H'_{набл} = \frac{N_{набл}}{1 - \frac{\sum_{j=1}^g T_j}{n^3 - n}}$$

**Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{набл} < F_{кр}$ , то влияние фактора считается незначимым. Если же  $F_{набл} > F_{кр}$  то фактор влияет на изменчивость средних.



## Пример 5

В таблице приведены данные о содержании иммуноглобулина IgA в сыворотке крови (в мг %) у больных четырёх возрастных групп.

Группа	содержание IgA (в мг %)										
1	83	85	82	82	84	-	-	-	-	-	-
2	84	85	85	86	86	87	-	-	-	-	-
3	86	87	87	87	88	88	88	88	88	89	90
4	89	90	90	91	91	-	-	-	-	-	-

Проверить гипотезу о том, что содержание иммуноглобулина у всех возрастных групп совпадает. Принять  $\alpha=0,05$ .

# Решение примера 5

Исходные данные:

	A	B	C	D
1	Группа 1	Группа 2	Группа 3	Группа 4
2	83	84	86	89
3	85	85	87	90
4	82	85	87	90
5	82	86	87	91
6	84	86	88	91
7		87	88	
8			88	
9			88	
10			88	
11			89	
12			90	

	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	№ группы	IgA (в мг %)	№ группы	IgA (в мг %)	R								
2	1	83	1	82	1,5	R1=	17,5	n1=	5				
3	1	85	1	82	1,5	R2=	52	n2=	6				
4	1	82	1	83	3	R3=	186	n3=	11				
5	1	82	1	84	4,5	R4=	122,5	n4=	5				
6	1	84	2	84	4,5			n=	27				
7	2	84	1	85	7	Ннабл=	21,69						
8	2	85	2	85	7								
9	2	85	2	85	7								
10	2	86	2	86	10								
11	2	86	2	86	10	g=	9						
12	2	87	3	86	10	j=1	t1=	2	T1=	6			
13	3	86	2	87	13,5	j=2	t2=	2	T2=	6			
14	3	87	3	87	13,5	j=3	t3=	3	T3=	24			
15	3	87	3	87	13,5	j=4	t4=	3	T4=	24			
16	3	87	3	87	13,5	j=5	t5=	4	T5=	60			
17	3	88	3	88	18	j=6	t6=	5	T6=	120			
18	3	88	3	88	18	j=7	t7=	2	T7=	6			
19	3	88	3	88	18	j=8	t8=	3	T8=	24			
20	3	88	3	88	18	j=9	t9=	2	T9=	6			
21	3	88	3	88	18					Сумма=	276		
22	3	89	3	89	21,5					Ннабл=	21,995		
23	3	90	4	89	21,5	$\chi^2(0,05;4-1)=$	7,81			$=\chi^2.ОБР.ПХ(0,05;3)$			
24	4	89	3	90	24								
25	4	90	4	90	24								
26	4	90	4	90	24								
27	4	91	4	91	26,5								
28	4	91	4	91	26,5								

**Вывод:** так как  $Ннабл < \chi^2(21,995 < 7,81)$ , то гипотеза  $H_0$  отклоняется: данные свидетельствуют о различном содержании иммуноглобулина в крови больных разных возрастных групп.

## 6. Двухфакторный анализ Фридмана

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Коэффициент конкордации Кендалла

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

$$F' = \frac{\sum_{j=1}^n \left( \sum_{i=1}^k R_{ij} - \frac{1}{2} k(n+1) \right)^2}{\frac{1}{12} kn(n+1) - \frac{1}{n-1} \sum_{i=1}^k T_i}$$

## Пример 6

Киноплёнка четырёх видов была представлена трём экспертам для определения лучшей из них. Каждому эксперту предложили упорядочить плёнки по степени предпочтения. Баллы (ранги), проставленные экспертами, приведены в таблице. Наибольший балл соответствует плёнке самого лучшего качества.

Эксперты, k	Вид плёнки, n			
	1	2	3	4
1	2	1	3	4
2	2	1	4	3
3	2	1	4	3
$\Sigma$	6	3	11	10

Требуется определить, различаются ли виды плёнок и согласованы ли оценки экспертов.

# Решение примера

- **Основная идея дисперсионного анализа:** выборочная дисперсия, вызванная воздействием факторов, сравнивается с выборочной дисперсией, обусловленной случайными причинами. Если различия значимы, то считают, что фактор оказывает существенное влияние. Проверка значимости дисперсий проводится по критерию Фишера. Если  $F_{\text{набл}} < F_{\text{кр}}$ , то влияние фактора считается незначимым. Если же  $F_{\text{набл}} > F_{\text{кр}}$  то фактор влияет на изменчивость средних.

# Заключение

1. Рассмотрены основные понятия дисперсионного анализа.
2. Приведены расчетные формулы для проведения однофакторного и двухфакторного дисперсионного анализа, а также для проведения множественных сравнений.
3. Рассмотрены однофакторный анализ Краскела-Уоллиса и анализ Фридмана.
4. Приведены примеры проведения дисперсионного анализа.

# Литература

1. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя. – СПб. : Питер, 2008. – 608 с.
2. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента в химической технологии. Учеб. пособие для хим.-технол. спец. вузов.-М., Высш. шк., 1985.-327 с.
3. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL.: учебное пособие. – М., Форум, 2008. – 464 с.



**Спасибо за внимание!**