



Лекция 10. Однофакторный дисперсионный анализ

10-1. Задача дисперсионного анализа

10-2. Однофакторный дисперсионный анализ

10-3. Решение в SPSS

Пример данных



Имеется ли разница в среднем возрасте учителей, администрации и обслуживающего персонала школы? Взяты выборки из трех генеральных совокупностей.

Учителя	Администрация	Обслуживающий персонал
24	59	34
27	35	29
26	29	35
50	40	31
48	39	40
40	54	45
	56	



10-1.

Задача дисперсионного анализа

Однофакторный и двухфакторный дисперсионный анализ

Дисперсионный анализ (Analysis of Variance)



F -критерий, который мы использовали при сравнении дисперсий, может применяться для сравнения трех и более средних.

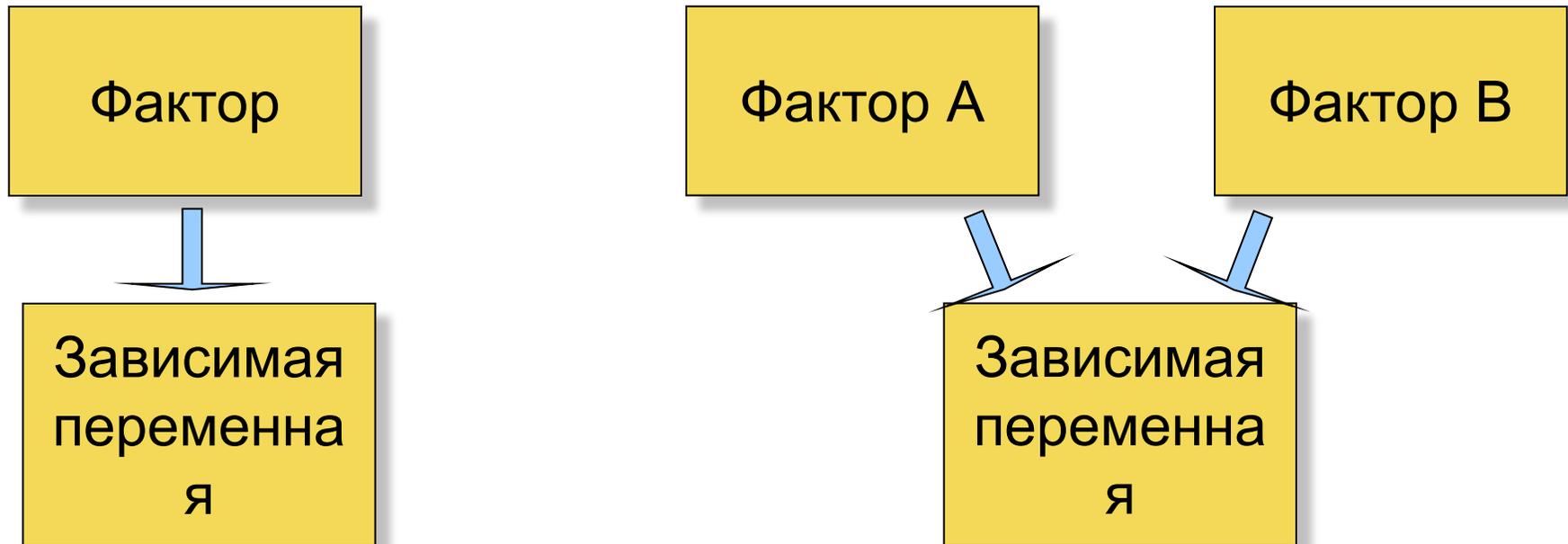
Этот метод называется **дисперсионным анализом** или в англоязычной аббревиатуре *ANOVA* (Analysis of Variance).

F -критерий можно использовать при сравнении двух средних. Но в этом случае он становится идентичным t -критерию.

Однофакторный и двухфакторный анализ



Дисперсионный анализ, который рассматривает только одну переменную называется **однофакторным дисперсионным анализом** (One-Way ANOVA). Дисперсионный анализ может также применяться в случае двух переменных - это **двухфакторный дисперсионный анализ** (Two-Way ANOVA).





10-2. Однофакторный дисперсионный анализ

Постановка задачи

Описание метода

Пример

Признак, фактор и уровни фактора



Исследуется **только один признак** или переменная: возраст сотрудников.

Рассматривается **только один фактор**: категория персонала.

Три уровня фактора: учителя, администрация, обслуживающий персонал.

Представление данных



Данные удобно представлять в виде таблицы. Выборки не обязаны иметь одинаковый объем.

Уровни фактора	Уровень 1	Уровень 2	...	Уровень k
Измерения признака	x_{11}	x_{21}	...	x_{k1}
	x_{12}	x_{22}	...	x_{k2}
		x_{23}	...	
			...	
Объемы выборок	n_1	n_2		n_k

Имеется k уровней.

Всего проведено N измерений.



1. Генеральные совокупности, из которых формируются выборки, должны быть нормально распределены.
2. Выборки должны быть независимы.
3. Дисперсии генеральных совокупностей должны быть равны.



Для выявления различия между тремя и более средними, выдвигаются следующие гипотезы:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_m$$

H_1 : не все средние равны



Вычисляются две оценки: **межгрупповая дисперсия** и **внутригрупповая дисперсия**.

Если нет разницы в средних, то оценки межгрупповой и внутригрупповой дисперсий приблизительно равны и значение F -критерия близко к 1, поэтому нулевая гипотеза принимается.

Если различие в средних значительно, межгрупповая дисперсия будет гораздо больше, чем внутригрупповая. Значение F -критерия будет значительно больше 1 и нулевая гипотеза будет отвергнута.

Поскольку при проверке гипотезы мы сравниваем дисперсии, метод и получил название *дисперсионный анализ*.

Степени свободы и критическая область



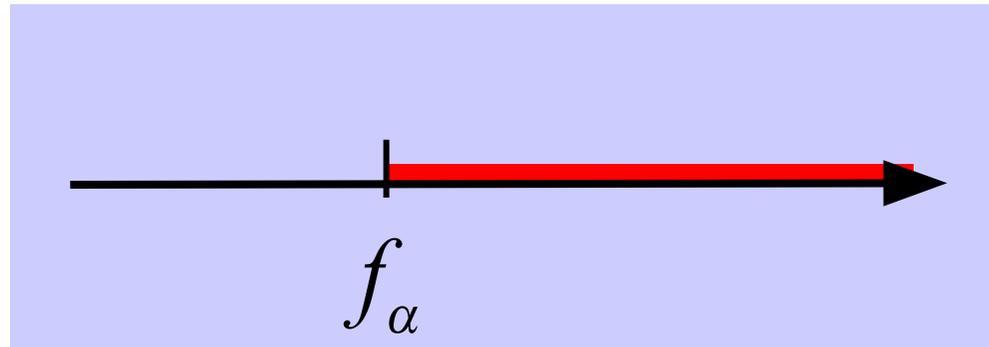
Степени свободы F-распределения задаются двумя значениями:

Числителя: $df = k - 1$

Знаменателя: $df = N - k$

Уравнение критической области (правосторонняя):

$$P(F > f_{\alpha}) = \alpha$$



Суммы квадратов отклонений



Межгрупповая сумма квадратов отклонений:

$$SS_b = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$$

Sum **S**quare
Between Groups

Внутригрупповая сумма квадратов отклонений:

$$SS_w = \sum (x - \bar{x}_i)^2$$

Sum **S**quare
Within Groups

Общая сумма квадратов отклонений:

$$SS = \sum (x - \bar{x})^2 = SS_b + SS_w$$

Sum **S**quare

Факторная и остаточная дисперсия. Критерий



Межгрупповая (факторная) дисперсия:

$$MS_B = \frac{SS_B}{k - 1}$$

Mean **S**quare
Between Groups

Внутригрупповая (остаточная) дисперсия:

$$MS_W = \frac{SS_W}{N - k}$$

Mean **S**quare
Within Groups

F-критерий:

$$F = \frac{MS_B}{MS_W}$$

Таблица результатов



Результаты вычислений принято представлять в виде следующей таблицы:

	Сумма квадратов	df	Среднее квадратичное	F
Между группами	SS_B	$k - 1$	MS_B	F -значение
Внутри групп	SS_W	$N - k$	MS_W	
Итого	$SS_B + SS_W$	$N - 1$	$MS_B + MS_W$	

Пример



Учителя	Администраци я	Обслуживающ ий персонал
24	59	34
27	35	29
26	29	35
50	40	31
48	39	40
40	54	45
	56	

Шаг 1. Гипотезы: $H_0 : \mu_1 = \mu_2 = \dots = \mu_m$

$H_1 : \text{не все средние равны}$

Шаг 2. Критическая область



Найдем критическое значение по таблице критических точек распределения Фишера.

Уровень значимости $\alpha = 0,05$.

Так как $k = 3$ и $N = 19$, то

$$\text{числитель} \quad df = k - 1 = 3 - 1 = 2$$

$$\text{знаменатель} \quad df = N - k = 19 - 3 = 16$$

Критическое значение равно 3,633.

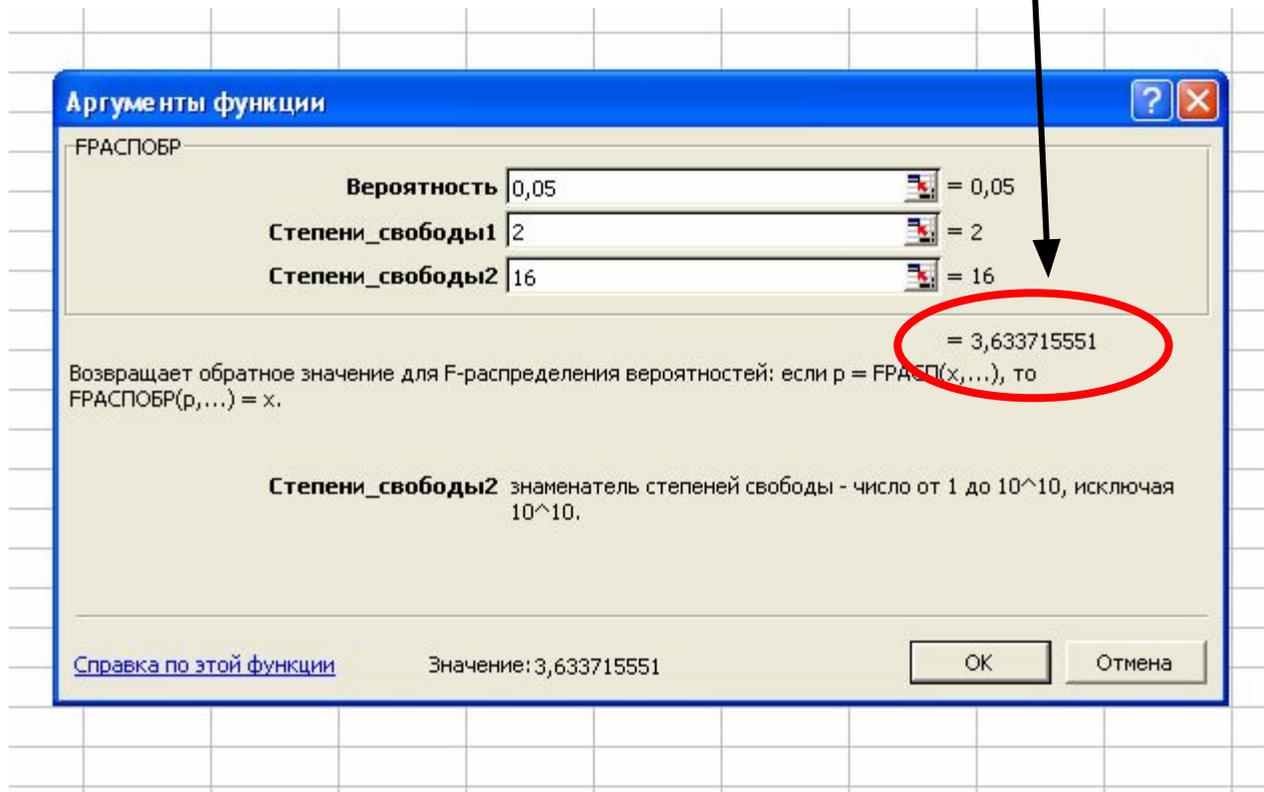
Критическая область $F > 3,633$

Нахождение F-значения в Excel



Критическое значение можно найти, используя функцию в Excel:

$$\text{FРАСПОБР}(0,05; 2; 16) = 3,633\dots$$



Шаг 3. Вычисление статистики F



Учителя	Администрация	Обслуживающий персонал
24	59	34
27	35	29
26	29	35
50	40	31
48	39	40
40	54	45
	56	

Шаг 3а. Подсчет средних

$$\bar{x}_1 = 35,8 \quad n_1 = 6$$

$$\bar{x}_2 = 44,6 \quad n_2 = 7$$

$$\bar{x}_3 = 35,7 \quad n_3 = 6$$

$$\bar{\bar{x}} = 39$$

$$N = n_1 + n_2 + n_3 = 19$$

Шаг 3б. Расчет отклонений



$$\begin{aligned}SS_b &= \sum n_i (\bar{x}_i - \bar{\bar{x}})^2 = \\ &= 6 \cdot (35,8 - 39)^2 + 7 \cdot (44,6 - 39)^2 + 6 \cdot (35,7 - 39)^2 = \\ &= 344,1\end{aligned}$$

$$\begin{aligned}SS_w &= \sum (x - \bar{x}_i)^2 = \\ &= (24 - 35,8)^2 + (27 - 35,8)^2 + \dots + (48 - 35,8)^2 + (40 - 35,8)^2 + \\ &+ (59 - 44,6)^2 + (35 - 44,6)^2 + \dots + (54 - 44,6)^2 + (56 - 44,6)^2 + \\ &+ (34 - 35,7)^2 + (29 - 35,7)^2 + \dots + (40 - 35,7)^2 + (45 - 35,7)^2 = \\ &= 1669,9\end{aligned}$$

Шаг 3с. Расчет дисперсий



$$MS_B = \frac{SS_B}{k-1} = \frac{344,1}{2} = 172,06$$

$$MS_W = \frac{SS_W}{N-k} = \frac{1669,9}{16} = 104,37$$

Шаг 3д. Расчет статистики



$$F = \frac{MS_B}{MS_W} = \frac{172,06}{104,37} = 1,649$$

Шаг 4-5. Получение выводов, ответ



$$1,649 < 3,633$$

Полученное значение статистики не попало в критическую область. У нас нет оснований думать, что средние значения отличаются.

Ответ.

Средний возраст рассматриваемых категорий персонала не различается значимо.



10-3. Решение задачи в SPSS

Ввод данных

Анализ

Отчет



ANOVA

VCE					
Between groups	Sum of Squares	5	115,000	23,000	253
Within groups	Sum of Squares	10	892,400	89,240	89,240
Total	Sum of Squares	15	1007,400	112,240	112,240