

# ДИСПЕРСИОННЫЙ АНАЛИЗ

## ANOVA

ANOVA - это акроним от ANalysis Of VAriance (дисперсионный анализ). Дисперсионный анализ был введен Фишером - английским учёным, сделавшим огромный вклад в развитие науки. ANOVA в статистике - это мощный инструмент для определения влияния различных групп наблюдений между собой.

### | ПРИМЕР

Предположим, Вы хотите эмпирическим методом провести исследование бензина на качество, для этого вы заправляете бак на одной заправке и проезжаете  $n$  километров, повторяете такой эксперимент, скажем, пять раз, затем проводите такой же эксперимент, только на другой заправке. У Вас два набора данных - заправка А и заправка В. Разумеется, цифры разбегаются, но всё же есть некоторая зависимость, так вот, что бы определить, влияет ли заправка на расход бензина (или данные не связаны между собой) Вы используете дисперсионный анализ.

Дисперсионный анализ позволяет определить какой из факторов влияет больше, внутригрупповой или межгрупповой. В примере выше Вы сможете определить, насколько влияет на расход бензина выбор заправки. В этом суть дисперсионного анализа: узнать, является ли выбранный фактор значимым для выбранных наблюдений.

В некотором смысле, дисперсионный анализ похож на регрессионный и корреляционный анализы, т.к. позволяет определить влияние переменных друг на друга.

## АНАЛИЗ

В теории, для анализа дисперсии выстраивается простая модель, схожая с изучаемой в [анализе временных рядов](#).

### | МОДЕЛЬ

Модель дисперсионного анализа включает в себя среднее значение, эффект эксперимента и случайную ошибку:

$$y = \mu + \tau + \varepsilon$$

$\tau$  - эффект эксперимента,  $\varepsilon$  - случайная ошибка

## ОДНОФАКТОРНЫЙ

Однофакторный дисперсионный анализ рассматривает влияние одного критерия, делается это так: мы проводим два эксперимента, в одном из них включаем дополнительный фактор и анализируем, внёс ли этот фактор изменения. В качестве исходных данных рассмотрим результаты ряда экспериментов:

N	E <sub>1</sub>	E <sub>2</sub>	E <sub>3</sub>	E <sub>4</sub>
1	33	60	88	50
2	41	55	121	58
3	49	38	84	47
4	40	56	87	48
5	43	49	124	30
$\mu_i$	<b>41.2</b>	<b>51.6</b>	<b>100.8</b>	<b>46.6</b>

$$\mu = (41.2 + 51.6 + 100.8 + 46.6) / 4 = 60.05$$

Квадрат ошибок внутри групп (Square Sum within group):

$$SS_w = \sum_i \sum_j (y_{ij} - \mu_i)^2 = 2428$$

Квадрат ошибок между группами (Square Sum between group):

$$SS_b = \sum_i (\mu_i - \mu)^2 = 2268.19$$

Учитывая степени свободы, ожидаемое среднее:

$$MS_w = SS_w / a(n-1) = 161.87$$

$$MS_b = SS_b / a-1 = 567.05$$

Значение  $F_{\text{крит}}$  :

$$F_0 = MS_b / MS_w = 3.503$$

Тест Фишера: если значение  $F_0$  окажется больше чем значение  $F_{\lambda,4,15}$ , значит фактор оказывает влияние.

$$\text{Для } n = 20 \text{ и } a = 5, F_{\lambda,n-a,a-1} = F_{\lambda,15,4} = 5,86$$

Поскольку  $F_0 = 3.503 < 5.86$ , то принимаем, что введённый фактор **не** оказал

влияния на результаты эксперимента.

## ДВУХФАКТОРНЫЙ

При двухфакторном анализе выдвигаются три гипотезы на проверку:

- Факторы А и В не оказывают влияния на результат
- Фактор А не оказывает влияния на результат
- Фактор В не оказывает влияния на результат

Для проведения двухфакторного анализа необходимо составить группы результатов: несколько измерений для всех значения каждого из факторов, т.е.:

	A <sub>1</sub>	A <sub>2</sub>
B <sub>1</sub>	X <sub>1,a1,b1</sub> ...X <sub>N,a1,b1</sub>	X <sub>1,a1,b2</sub> ...X <sub>N,a1,b2</sub>
B <sub>2</sub>	X <sub>1,a1,b2</sub> ...X <sub>N,a1,b2</sub>	X <sub>1,a1,b2</sub> ...X <sub>N,a1,b2</sub>

Далее подсчитывается среднее значение для каждого значения факторов, т.е. среднее для A<sub>1</sub>, среднее для B<sub>1</sub> и т.д. Затем подсчитывается общее среднее для всех результатов. Зададимся количеством критериев: k = 2 (количество критериев А) и m = 2 (количество критериев В).

$$T = \sum \sum \sum x_{ijk}$$

Сумма элементов под влиянием фактора А:

$$T_{Ai} = \sum x_{i \cdot k}$$

Сумма элементов под влиянием фактора В:

$$T_{Bj} = \sum x_{\cdot jk}$$

Сумма элементов под влиянием фактора АВ:

$$T_{AiBj} = \sum x_{ij \cdot}$$

$$SST = \sum x_{ijk}^2 - T^2/N$$

$$SSA = \sum T_{Ai}^2/n \cdot m - T^2/N$$

$$SSB = \sum T_{Bj}^2/n \cdot k - T^2/N$$

$$SSAB = \sum \sum T_{AiBj}^2/n - SSA - SSB - T^2/N$$

$$SSE = \sum \sum \sum x_{ijk}^2 - \sum \sum T_{AiBj}^2/n$$

$$SST = SSA + SSB + SSAB + SSE$$

$$MSE = SSE/(n-1) \cdot m \cdot k$$

$$MSA = SSA/k-1$$

$$MSB = SSB/m-1$$

$$MSAB = SSAB/(m-1) \cdot (k-1)$$

Тест "Критерий А **не** оказывает влияние на результат",  $\nu_1 = k-1$ :

$$F_A = MS_A/MS_E$$

Тест "Критерий В **не** оказывает влияние на результат",  $\nu_1 = m-1$ :

$$F_B = MS_B/MS_E$$

Тест "Критерии А и В **не** оказывают влияние на результат",  $\nu_1 = (k-1)(m-1)$ :

$$F_{int} = MS_{AB}/MS_E$$

Для каждого F, если  $F > F_{\alpha, \nu_1, \nu_2}$ , то гипотеза отвергается.  $\nu_2 = N-mk$

## МНОГОФАКТОРНЫЙ

Многофакторный анализ аналогичен двухфакторному - проводятся те же операции, но критерии группируются и итеративно находится влияние каждого из факторов.

## С ПОВТОРНЫМИ ИЗМЕРЕНИЯМИ

Дисперсионный анализ с повторными измерениями означает, что для каждого критерия производилось несколько замеров случайной величины для получения более точного результата (поскольку в ANOVA) используется внутригрупповая сумма квадратов.

## ПРИМЕНЕНИЕ

Дисперсионный анализ применяют в самых различных отраслях науки и производства тогда, когда необходимо изучить зависимость критериев на различие средних значений, при этом сравнивается не среднее значение, а разброс результатов вокруг среднего значения, т.е. дисперсию.

## РЕШЕНИЕ ЗАДАЧ

В качестве примера приведём задачу из метрологии. На заводе размещены пять станков, на которых производят валы. Необходимо определить, влияет ли выбор станка или подготовка работника на результат производства. Для анализа производят замеры для каждого станка и работника, в результате получается таблица:

### Оператор 1

<b>M1</b>	30.435	30.652	30.591	30.353	30.405	30.458	30.505	30.595	30.506	30.501
<b>M2</b>	30.381	30.391	30.38	30.345	30.318	30.4	30.3	30.351	30.387	30.325
<b>M3</b>	30.387	30.331	30.371	30.312	30.313	30.304	30.3	30.387	30.318	30.326
<b>M4</b>	30.224	30.281	30.202	30.205	30.294	30.221	30.214	30.215	30.276	30.295
<b>M5</b>	30.592	30.502	30.565	30.309	30.46	30.512	30.433	30.343	30.561	30.488

### Оператор 2

<b>M1</b>	30.333	30.392	30.36	30.302	30.393	30.337	30.322	30.392	30.334	30.329
<b>M2</b>	30.397	30.326	30.301	30.321	30.384	30.375	30.372	30.35	30.378	30.311
<b>M3</b>	30.368	30.358	30.338	30.333	30.392	30.371	30.325	30.351	30.374	30.317
<b>M4</b>	30.117	30.106	29.663	29.949	29.983	29.694	29.736	29.967	29.523	29.868
<b>M5</b>	30.394	30.381	30.306	30.304	30.303	30.374	30.362	30.341	30.307	30.354

Воспользуемся методом двухфакторного анализа, фактор А - оператор, фактор В - станок. Рассчитаем суммы квадратов, для этого необходимо рассчитать значение среднего для каждой из групп:

$T$	$T_{A1}$	$T_{A2}$	$T_{B1}$	$T_{B2}$	$T_{B3}$	$T_{B4}$	$T_{B5}$
3031.688	1519.12	1512.568	608.495	607.093	606.876	601.033	608.191

$$SSA = 0.429$$

$$SSB = 1.854$$

$$SSAB = 0.506$$

$$SSE = 0.583$$

$$MSA = 0.429$$

$$MSB = 0.464$$

$$MSAB = 0.127$$

$$MSE = 0.146$$

$$F_A = 2.938$$

$$F_B = 3.178$$

$$F_{AB} = 0.87$$

Критические значения для теста Фишера:

$$F_{\text{crit A}} = F_{0.1, 1, 90} = 2.77$$

$$F_{\text{crit B}} = F_{0.1, 4, 90} = 2.01$$

$$F_{\text{crit AB}} = F_{0.1, 4, 90} = 2.01$$

Таблица результатов:

<b>Влияние станка на результат</b>	Нет 2.938 > 2.77
<b>Влияние квалификации работника на результат</b>	Нет 3.178 > 2.01
<b>Взаимное влияние квалификации работника и выбора станка на результат</b>	Да 0.87 < 2.01

## В EXCEL/OPEN CALC

Для решения дисперсионного анализа в электронной таблице Вам потребуются следующие формулы:

sumproduct Сумма произведений, используется для нахождения суммы квадратов  
finv Обратное значение распределения F - критерий Фишера

Таблица для скачивания в форматах [ods](#) и [xls](#).

---

### **УДК: ГРНТИ:**

**Автор статьи:** Телятников З.А.

**Дата написания статьи:** 30.06.2017

**Адрес статьи в интернете:**

[http://k-tree.ru/articles/statistika/analiz\\_dannyh/dispersionnii\\_analiz\\_anova](http://k-tree.ru/articles/statistika/analiz_dannyh/dispersionnii_analiz_anova)

**Дата формирования документа:** 22.10.2017 04:04

---

Все материалы данного файла являются объектами авторского права (в том числе дизайн).  
Запрещается копирование, распространение (в том числе путем копирования на другие сайты и ресурсы в Интернете) или любое иное использование информации и объектов без предварительного согласия правообладателя.