1. **Введение**

Дисперсионный анализ (ANOVA от англ. ANalysis Of VAriance) — это статистический метод, который используется для сравнения средних значений двух или более выборок. Он позволяет определить, различаются ли средние значения между группами, или же различия случайны. Суть 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.

1. **Основные понятия**

*Зависимая переменная (показатель)* – переменная, значения которой определяются с помощью измерений в ходе эксперимента.

*Независимая переменная (фактор)* – переменная, которой можно управлять при проведении эксперимента.

Например, если в эксперименте сравнивается число лейкоцитов (WCC) в крови мужчин и женщин, то Пол можно назвать независимой переменной, а WCC - зависимой переменной.

*Общая дисперсия* –мера вариации признака по всей совокупности под влиянием всех факторов, обусловивших эту вариацию.

Чем дисперсия больше, тем дальше разбросаны значения относительно общей средней, и наоборот, чем дисперсия меньше, тем они к средней ближе.

*Межгрупповая дисперсия* характеризует систематическую вариацию, т.е. различия в величине изучаемого признака, возникающие под влиянием признака-фактора, положенного в основание группировки.

*Внутригрупповая дисперсия* отражает случайную вариацию, т.е. часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основание группировки.

*Правило сложения дисперсий –* общая дисперсия включает в себя внутригрупповую и межгрупповую дисперсию.

1. **Дисперсионный план.**

Для проведения однофакторного дисперсионного анализа с кодированными значения составляют план эксперимента. Введем следующие обозначения:

– исследуемый фактор

– максимальное число разных уровней фактора

– номер уровня фактора

– конкретное значение (качественное или количественное) уровня фактора

– номер повторного опыта при одном значении фактора

– общее число опытов при эксперименте

Классическая форма дисперсионного плана:

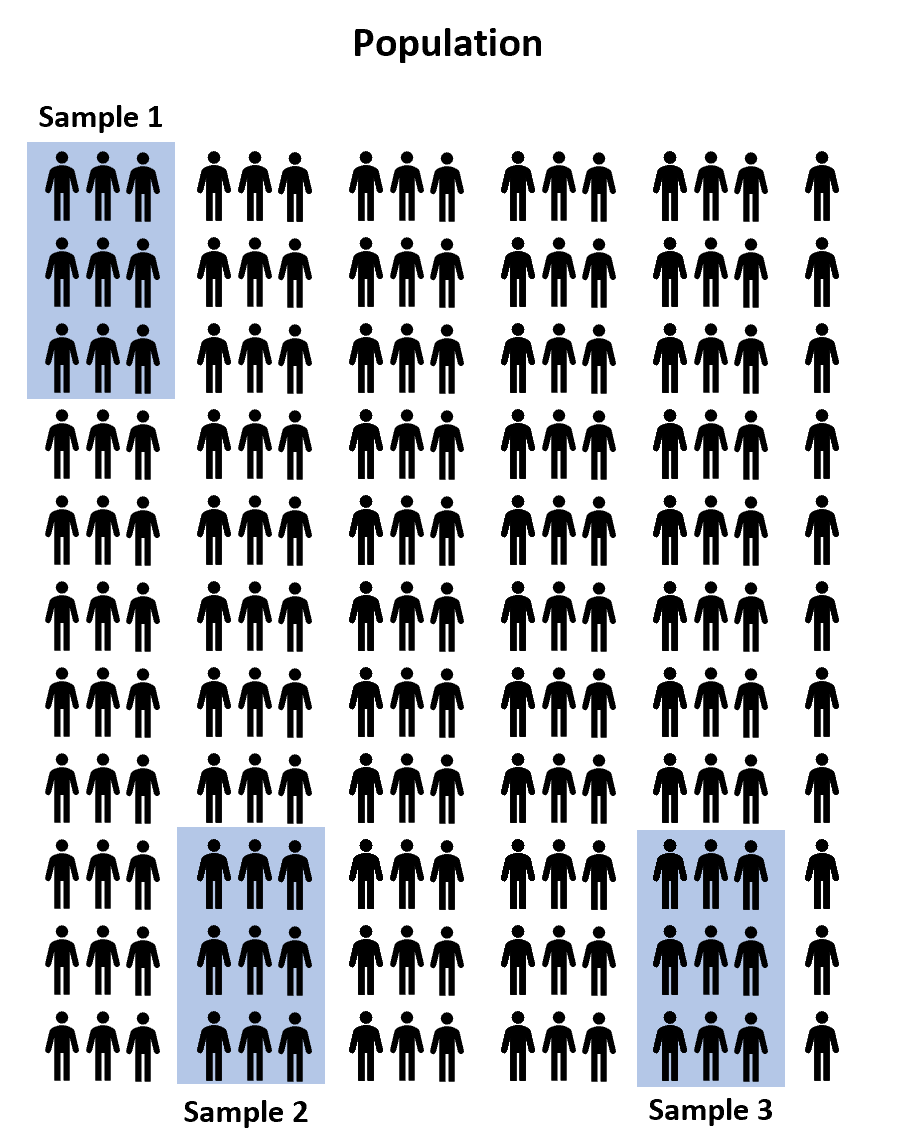
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер повторного опыта | Значения при уровне фактора | | | | | | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Пример задачи дисперсионного анализа:

Предположим, мы хотим узнать, приводят ли три разные программы подготовки к экзаменам к разным средним баллам на вступительном экзамене в колледж. Поскольку по всей стране проживают миллионы старшеклассников, было бы слишком много времени и денег, чтобы обратиться к каждому ученику и позволить им использовать одну из программ подготовки к экзаменам.

Вместо этого мы могли бы выбрать три случайные выборки из 100 студентов из населения и позволить каждой выборке использовать одну из трех программ подготовки к экзамену для подготовки к экзамену. Затем мы могли бы записывать баллы для каждого студента после сдачи экзамена.

Практически гарантировано, что средний балл за экзамен между тремя выборками будет хотя бы немного отличаться. Однофакторный дисперсионный анализ позволит ответить на вопрос, является ли эта разница статистически значимой.



1. **Шаги проведения однофакторного дисперсионного анализа**
2. Формулировка нулевой и альтернативной гипотез.

*Нулевая гипотеза ( -* принимаемое по умолчанию предположение о том, что не существует связи между двумя наблюдаемыми событиями.

*Альтернативная гипотеза () -* единственное утверждение, являющееся логическим отрицанием нулевой гипотезы. Часто альтернативная гипотеза означает наличие связи между изучаемыми переменными

1. Выбор уровня значимости.

*Уровень значимости* показывает вероятность того, что различия между группами случайны. Обычно принимают равным 0.05

1. Расчет суммы квадратов отклонений внутри группы:

– элемент в -й строке и -м столбце, – среднее значение в -й групппе.

1. Расчет суммы квадратов отклонений между средними каждой группы и общим средним:

– количество наблюдений в -й групе, – среднее значение в -й группе, – среднее значение во всей выборке.

1. Определение степеней свободы.

*Степень свободы* – количество наблюдений, которые могут быть спокойно изменены в каждой группе данных.

межгрупповые = , где – количество групп;

внутригрупповые = , где – общее количество наблюдений.

1. Расчет средних квадратов.

(Mean Square Between), средний квадрат – оценка расстояния между средними в группах.

(Mean Square Within) – оценка общей дисперсии по разбросу внутри групп. Сумма квадратов стандартных отклонений внутри групп.

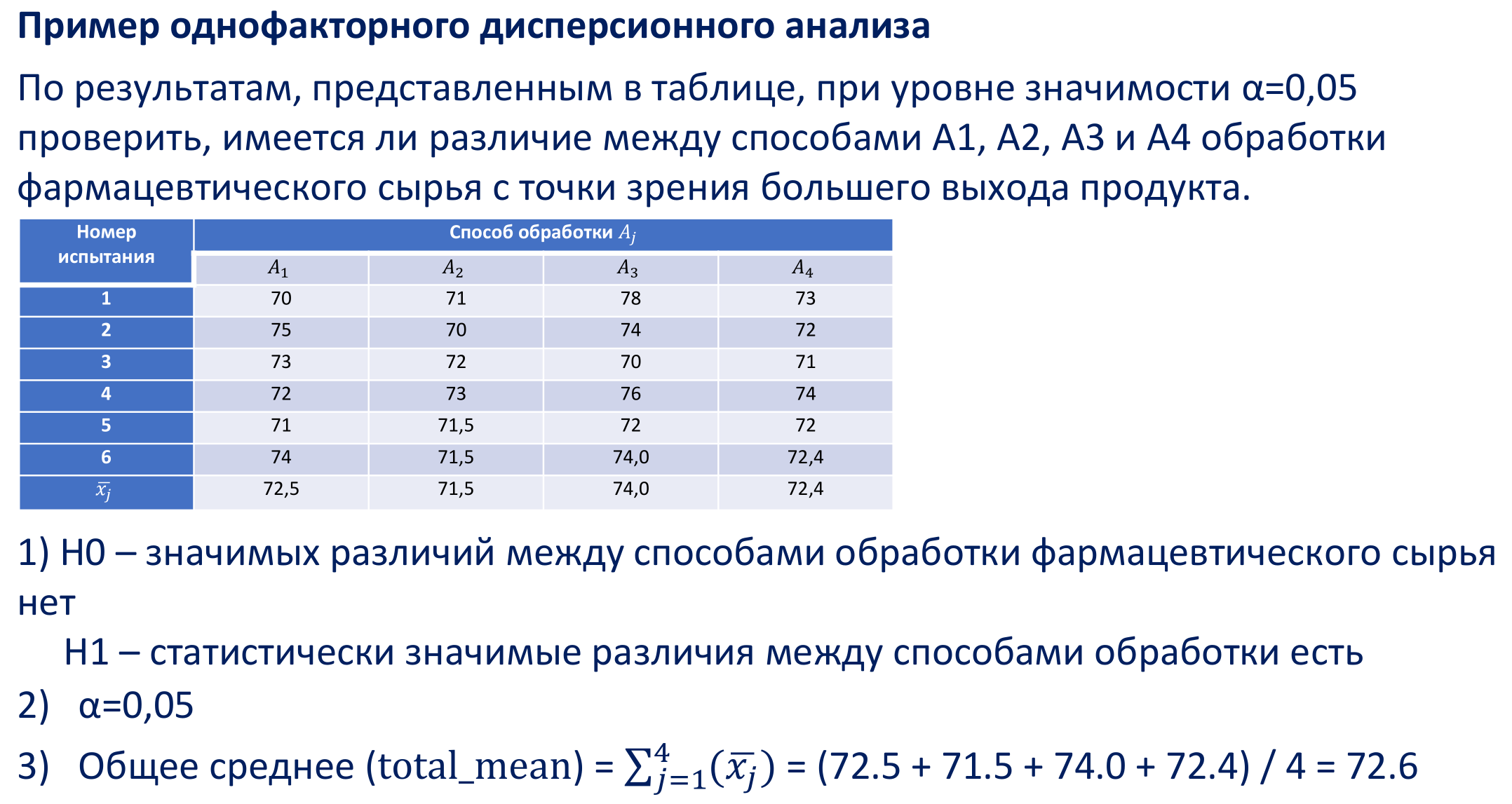
1. Расчет -статистики.

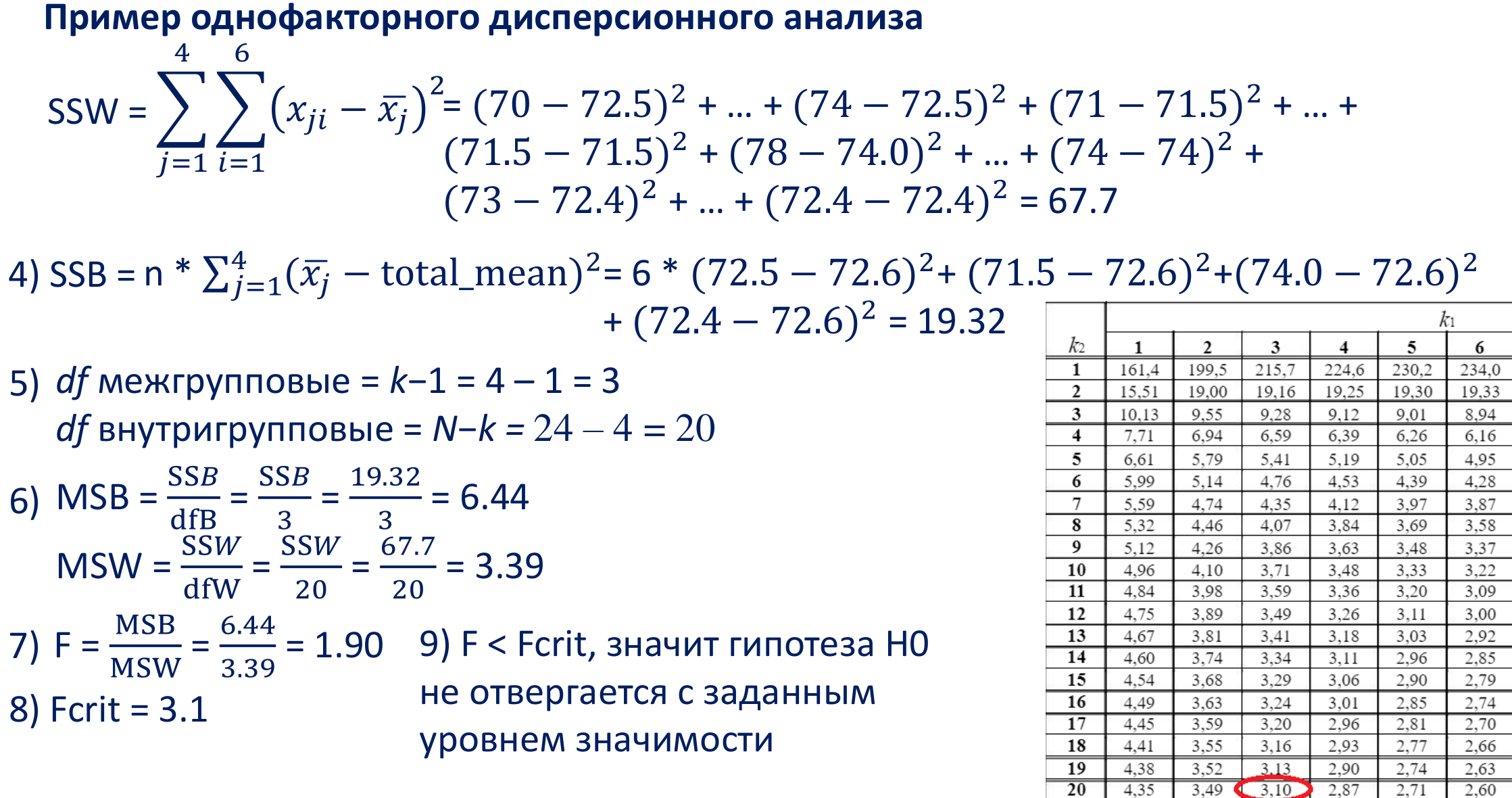
Отношение между средними значениями в группах и дисперсий внутри групп. Если большое, то это указывает на статистически значимые различия между группами.

1. Определение критического значения.

По таблице квантили распределения Фишера для заданного уровня значимости ( - межгрупповые, - внутригрупповые).

1. Если вычисленное значение превышает критическое значение, то мы отклоняем нулевую гипотезу





1. Преимущества использования однофакторного дисперсионного анализа.

Дисперсионный анализ позволяет изучать каждый фактор, управляя значениями остальных факторов. Это, в действительности, и является основной причиной его большей статистической мощности (для получения значимых результатов требуются меньшие объемы выборок). По этой причине дисперсионный анализ даже на небольших выборках дает статистически более значимые результаты, чем простой t-критерий.

1. Условия применимости.

- Выборки независимы между собой.

- Нормальное распределение признака в статистических совокупностях, из которых извлечены выборки.

- Равенство (гомогенность) дисперсий изучаемого признака в статистических совокупностях, из которых извлечены выборки

- Наблюдения в каждой из выборок независимы.

1. Заключение

ANOVA очень важен для статистического анализа данных и исследований. Он помогает находить значимые различия и определить факторы, влияющие на результаты исследования. Однако, для более точных результатов, необходимо учитывать все факторы влияния и применять другие методы анализа, если это необходимо.

1. Литература.

Шеффе Г. Дисперсионный анализ, пер. с англ. — М., 1963.

Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. — 2 изд.. — М., 1965.

Дудина А. А. Дисперсионный Анализ – Башкирский государственный аграрный университет, 2017.