



Санкт-Петербургский государственный университет
Кафедра системного программирования

Сравнение эффективности трёх методов обучения с использованием дисперсионного анализа

Юсуп Амин Турмуди, группа 24.M71-мм

Санкт-Петербург
2025

Что такое однофакторный дисперсионный анализ (One-Way ANOVA)?

- Однофакторный дисперсионный анализ — это статистический тест, который используется для проверки того, имеют ли три и более независимые группы значимо разные средние значения.
- Предположения:
 - ▶ Независимость наблюдений : Каждое значение данных должно принадлежать разным людям или единицам, которые действуют независимо. Один результат не должен влиять на другой. Пример: Студенты из разных классов, проходящие один и тот же тест.
 - ▶ Нормальное распределение данных в каждой группе : Данные каждой группы должны примерно соответствовать нормальному (колоколообразному) распределению, без сильного скошения влево или вправо.
 - ▶ Однородность дисперсий : Во всех группах разброс (вариативность) данных должен быть примерно одинаковым. Сильные различия в разбросе могут повлиять на результаты ANOVA.

- Сравнение средних значений более чем двух групп.
- Пример: Сравнение средних оценок студентов, обучавшихся по трём разным методам преподавания.

- Нулевая гипотеза (H_0): Средние значения всех групп равны, то есть

$$\mu_1 = \mu_2 = \mu_3$$

- Альтернативная гипотеза (H_1): По крайней мере одно среднее значение группы отличается, то есть

$$\exists i, j \text{ такие, что } \mu_i \neq \mu_j$$

- **Таблица ANOVA:** Структурированное представление, которое делит общую вариацию данных на:
 - ▶ **Межгрупповая вариация (SSB):** Различия, обусловленные средними значениями групп, $SSB = \sum n_i(\bar{X}_i - \bar{X})^2$
 - ▶ **Внутригрупповая вариация (SSW):** Различия внутри каждой группы (случайная изменчивость), $SSW = \sum (\text{оценка} - \bar{X}_i)^2$
 - ▶ **Средние квадраты (MSB, MSW):** Дисперсии, полученные делением SS на степени свободы, $MSB = SSB/df_{\text{межгрупп}}$, $MSW = SSW/df_{\text{внутригрупп}}$
 - ▶ **F-статистика:** Отношение MSB к MSW, $F = MSB/MSW$, используемое для проверки значимости различий средних значений групп
 - ▶ **p-value:** Вероятность получить наблюдаемое значение F случайно, если нулевая гипотеза H_0 верна

- Использовать пример набора данных с **тремя группами**
- Выполнить анализ в **Google Colab**
- Реализовать рабочий процесс с помощью языка программирования **R**
- Провести **ANOVA**, построить визуализации и проверить статистические допущения

Заключение: язык программирования R.

- ▶ Мощные инструменты для статистического анализа.
- ▶ Отличные возможности визуализации данных.
- ▶ Надёжный и воспроизводимый рабочий процесс анализа данных.