

# **Сторителлинг: как KNIME помог увидеть закономерности в учебных данных**

## **Выполнили:**

Ибрагимов Р.А.

Мархаев А.М.

Салимов Д.Р

## **Контекст:**

Современные цифровые технологии генерируют массивы данных, которые невозможно анализировать вручную. Особенно это актуально в образовании, где важно понимать, что влияет на успеваемость студентов: количество часов подготовки? Предыдущий уровень знаний? Вовлечённость?

В рамках курсовой работы стояла задача не просто обработать данные, а найти статистически значимые взаимосвязи между количеством часов подготовки и итоговыми баллами за экзамен, используя прикладной инструментарий. Для этого был выбран визуальный аналитический пакет KNIME.

## **Цель:**

Основная цель проекта — оценить степень взаимосвязи между количеством часов подготовки и результатом экзамена, применяя методы корреляционного анализа и визуализируя данные с помощью KNIME. Это позволило бы не только проверить рабочую гипотезу, но и продемонстрировать практическое применение платформы в анализе реальных образовательных данных.

## **Инструмент: KNIME:**

KNIME (Konstanz Information Miner) — визуальная среда анализа данных, позволяющая собирать рабочие процессы из узлов, каждый из которых выполняет конкретную задачу: загрузка данных, трансформация, вычисления, визуализация.

Выбор этой платформы был не случаен:

- она гибка в обработке данных,
- воспроизводима (все шаги сохраняются в виде схемы),
- и позволяет работать без программирования — что важно для скоростной сборки прототипов.

## **Структура исследования**

Анализ состоял из нескольких этапов:

**1. Подготовка данных**

- Набор из 100 строк: Study\_Hours (часы подготовки) и Exam\_Score (баллы).
- Предварительная очистка от пропущенных значений.

**2. Визуализация**

- Построение обычных и полулогарифмических диаграмм рассеяния.
- Наблюдение за формой связи между переменными.

**3. Статистические расчёты**

- Математическое ожидание и стандартные отклонения.
- Коэффициент ковариации.
- Коэффициенты корреляции Пирсона и Спирмена.

**4. Сравнение и интерпретация**

- Анализ различий между линейной и монотонной связью.
- Проверка устойчивости модели к выбросам.

**Основные находки:**

- **Диаграммы рассеяния** показали явную положительную зависимость между количеством часов подготовки и результатами экзамена. Это проявлялось в том, что точки на графике располагались близко к прямой линии с положительным наклоном.
- **Полулогарифмические диаграммы** выявили интересный эффект: зависимость не строго линейная, а логарифмическая. Начальные часы подготовки влияли на результат сильнее, чем последующие.
- **Коэффициенты корреляции:**
  - Пирсона: 0.89 — высокая линейная положительная связь.
  - Спирмена: близок к Пирсона — связь монотонна и почти линейна.
  - Ковариация положительная, подтверждающая зависимость.
- **Математическое ожидание и отклонения:**  
Среднее количество часов — около 27, средний результат — 74.  
Стандартные отклонения — 6.07 часов и 9.6 баллов соответственно.

**Практическое значение:**

Результаты анализа позволяют сделать обоснованный вывод: увеличение времени подготовки действительно связано с более высокими результатами. Но также важно понимать, что влияние времени подготовки не бесконечно линейно — эффективность часов постепенно снижается, что нужно учитывать при планировании учебного процесса.

**Почему KNIME — это удобно:**

KNIME не просто выполнял вычисления. Он позволил:

- визуализировать все этапы, сделать анализ наглядным;
- автоматизировать расчёты без риска человеческой ошибки;
- быстро проверять разные гипотезы, меняя параметры без переписывания кода;
- повторно использовать шаблон анализа для других наборов данных.

### **Вывод:**

Проект подтвердил, что KNIME — мощный инструмент для анализа количественных данных в учебной среде.

Использование методов корреляционного анализа позволяет принимать более осознанные решения в образовании, планировании и научных исследованиях.

Методология, реализованная в этой курсовой работе, может быть масштабирована и применена в других контекстах — от маркетинга до медицины — везде, где важно понимать взаимосвязь между данными, а не просто их фиксировать.