

НАУЧНАЯ СТАТЬЯ О ПЕРЕРАБОТКЕ И АНАЛИЗЕ ДАННЫХ

Алламуратова Венера Жумамуратовна

*Каракалпакский государственный университет имени Бердаха
Ассистент*

ВВЕДЕНИЕ

Переработка и анализ данных играют важную роль в современных социальных, экономических и научных областях. Каждый день мы собираем данные из различных источников, но для того чтобы эффективно проанализировать эти данные и получить полезные результаты, их необходимо переработать. В данной статье рассматриваются процессы переработки и анализа данных, методы и практические примеры.

1. Процесс переработки данных

Процесс переработки данных включает в себя несколько этапов:

- **Сбор данных:** Первый этап – это сбор данных. На этом этапе данные собираются из различных источников (опросы, эксперименты, архивные данные и др.). Например, в одном исследовании мог быть проведен опрос о методах обучения студентов.

- **Очистка данных:** Собранные данные часто содержат ошибки или недостатки. Процесс очистки включает в себя исправление и приведение данных в нужный формат для дальнейшего анализа. Например, некоторые ответы студентов могут отсутствовать или быть указаны неверно.

- **Изменение данных:** На этом этапе данные могут быть преобразованы в новый формат, например, сгруппированы, нормализованы или вычислены какие-либо показатели. Например, оценки студентов могут быть переведены в 100-балльную систему.

- **Хранение данных:** После переработки данные необходимо эффективно сохранить в базе данных или файле. Этот процесс позволяет удобно получать доступ к данным и обновлять их в будущем.

2. Методы анализа данных

Процесс анализа данных включает в себя несколько методов:

- **Описательная статистика:** Описательная статистика используется для вычисления общих характеристик данных, таких как среднее, медиана, мода и стандартное отклонение. Например, при анализе методов обучения студентов можно определить их средние оценки и разброс этих оценок.

- **Корреляционный анализ:** Этот метод помогает изучить связь между двумя переменными. Например, мы можем выяснить корреляцию между временем, проведенным студентами на учебе, и их оценками. Если коэффициент корреляции близок к +1, это указывает на сильную положительную связь между переменными.

- **Регрессионный анализ:** Регрессионный анализ позволяет изучать зависимость одной переменной (например, оценки) от другой (например, времени, потраченного на учебу). Это дает возможность выявлять взаимосвязи между переменными и делать прогнозы.

- **Классификация и кластеризация:** Процесс группировки и классификации данных. Например, мы можем разделить студентов на разные группы в зависимости от их оценок, что поможет нам спланировать, как мы можем помочь им в образовательном процессе.

3. Практические примеры

3.1. Анализ методов обучения студентов

Например, в одном исследовании были собраны данные о 100 студентах, их методах обучения и оценках. Собранные данные могут выглядеть следующим образом:

ID Студента	Время на учебу (часы)	Оценки
1	2	85
2	3	90
3	1	78
4	4	95
5	3	88

Для анализа этих данных мы можем сначала рассчитать описательную статистику:

- Средняя оценка: $(85 + 90 + 78 + 95 + 88) / 5 = 87.20$
- Среднее время на учебу: $(2 + 3 + 1 + 4 + 3) / 5 = 2.6$ часа

Кроме того, мы можем провести корреляционный анализ. В результате этого анализа мы можем обнаружить, что существует положительная корреляция между временем на учебу и оценками.

3.2. Анализ процесса приема на работу

В экономической сфере анализ данных при приеме на работу является важным. Например, компания могла бы собрать данные о 500 человек, которые были приняты на работу:

ID Работника	Опыт работы (годы)	Зарплата (\$)	Оценка
1	5	55000	85
2	3	45000	78
3	8	60000	90
4	2	40000	70

Анализируя эти данные, мы можем изучить, как опыт работы и зарплата влияют на оценки. Мы можем использовать регрессионный анализ, чтобы определить связь между опытом работы и оценками.

4. Визуализация данных

При анализе данных важно представить результаты в графической форме. Этот процесс помогает сделать результаты более ясными и понятными. Представление данных с помощью графиков создает удобство для студентов или исследователей.

Например, оценки студентов можно представить с помощью гистограммы.

Если у нас есть оценки 100 студентов, мы можем использовать гистограмму, чтобы показать распределение оценок. С помощью гистограммы мы можем разделить оценки на группы и показать количество студентов в каждой группе.

5. Заключение

Переработка и анализ данных являются необходимыми процессами для современных исследований. В этих процессах важны этапы сбора, очистки, изменения и анализа данных. Статистические методы анализа, такие как описательная статистика, корреляция, регрессия и кластеризация, позволяют исследователям эффективно использовать данные. Кроме того, представление данных в графической форме облегчает понимание результатов. Процессы переработки и анализа данных имеют значительное значение для принятия решений в социальных, экономических и других областях.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА:

1. Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.
2. Pallant, J. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. Open University Press.
3. Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics*. Pearson.
4. Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press.