**Анализ данных продаж: Прогнозирование воздействия скидок и стоимости доставки**

**Аннотация**

В современном мире, где объемы данных постоянно растут, умение извлекать из них ценную информацию приобретает решающее значение. Данная статья посвящена анализу данных о продажах с помощью языка программирования R и метода линейной регрессии. Наша основная цель - изучить влияние скидок и стоимости доставки на объем продаж в компании.

**Введение**

В современном мире объемы данных растут с каждым днем, и умение извлекать из них ценную информацию становится ключевым навыком. В этой статье мы сосредоточимся на анализе данных о продажах с помощью языка программирования R и метода линейной регрессии. Наша цель - исследовать, как скидки и стоимость доставки могут влиять на объем продаж в компании.

**Материалы и методы и исходные данные**

Для начала мы загрузим исходные данные, представленные в формате CSV. В данной работе мы будем использовать набор данных "superstore.csv", содержащий информацию о продажах, скидках и стоимости доставки. Данные были разделены на обучающий и тестовый наборы, чтобы оценить работу модели на новых данных.

**Линейная регрессия**

Чтобы проанализировать влияние скидок и стоимости доставки на продажи, мы воспользуемся линейной регрессией. Вот как мы строим модель:

Здесь "Продажи" - зависимая переменная, а "Скидка" и "Стоимость доставки" - независимые переменные. Мы предполагаем, что эти факторы могут влиять на объем продаж (см. Рисунок 1).



Картинка - 1

**Результаты**

После построения линейной регрессионной модели мы можем оценить ее результаты и интерпретировать влияние скидок и стоимости доставки на объем продаж. Для этого мы воспользуемся следующим кодом:

Этот код предоставит нам статистическую сводку по модели, включая коэффициенты регрессии, p-значения и другие важные статистические данные. Например, если коэффициент для переменной "Скидка" положительный и значимый, это может означать, что увеличение скидок положительно влияет на объем продаж.

Мы также можем использовать модель для прогнозирования продаж на тестовом наборе данных:

Эти прогнозы можно сравнить с фактическими значениями из тестового набора данных, чтобы оценить точность модели

Был проведен анализ данных, и была построена линейная регрессионная модель, предсказывающая количество проданных товаров. Однако точность прогноза оказалась недостаточной, что открывает широкие возможности для улучшения модели.

Очень важно изучить коэффициенты регрессии для каждой независимой переменной. Если коэффициент положительный, это может свидетельствовать о положительном влиянии на продажи, а если отрицательный - о негативном. Р-значения помогут определить статистическую значимость этих влияний.

Также учитывайте коэффициент детерминации (R-квадрат), который измеряет объясненную дисперсию зависимой переменной. Высокий R-квадрат может указывать на то, что модель хорошо объясняет изменения в продажах.

**Заключение**

Анализ данных о продажах с помощью линейной регрессии в R - мощный инструмент для выявления влияния различных факторов на объем продаж. Однако важно помнить, что корреляция не всегда означает причинно-следственную связь, и результаты моделирования следует интерпретировать с осторожностью. Помимо линейной регрессии, для более глубокого понимания сложных взаимосвязей можно использовать и другие методы анализа данных.

Доступность кода и данных [https://github.com/seryibirka]

**Библиография**

<https://www.kaggle.com/datasets>

<https://github.com/>

<https://doc.arcgis.com/ru/insights/latest/analyze/regression-analysis.htm>

<http://statistica.ru/theory/osnovy-lineynoy-regressii/>