

Лабораторная работа_2

1. Типы данных и создание новых переменных
 - Определите и классифицируйте типы данных каждого столбца в наборе данных. Какие из них категориальные? Какие числовые?
 - Создайте новую переменную `цена_за_балл`, разделив цену на оценку, чтобы анализировать соотношение цены и качества вин.
2. Генеральная совокупность и выборка, частотные таблицы и распределения
 - Сформируйте частотную таблицу для переменной 'страна', чтобы увидеть количество отзывов по каждой стране.
 - Выберите случайную выборку из 100 строк и сравните среднюю цену с ценой по всему набору данных. Хорошо ли выборка представляет собой всю совокупность?
3. Описательные статистики
 - Рассчитайте описательные статистики для числовых переменных (среднее, медиана, мода, размах, дисперсия и стандартное отклонение).
 - Обсудите, какая мера центральной тенденции лучше всего описывает переменные оценки и цена и почему.
4. Z-оценка и выбросы
 - Рассчитайте Z-баллы для переменной `цена`, чтобы определить выбросы. Какой порог вы бы использовали и почему?
 - Как выбросы влияют на набор данных? Следует ли их удалять или оставить для анализа?
5. Корреляция
 - Вычислите матрицу корреляции для числовых переменных. Какие пары показывают наивысшую корреляцию?
6. Создайте гистограмму для переменной оценки и диаграмму размаха для переменной `цена`.
7. Постройте диаграмму рассеяния, чтобы визуализировать взаимосвязь между ценой и оценками. Используйте цветовую кодировку, чтобы различать страны. Введение в тестирование гипотез, непараметрические критерии
8. Линейная регрессия
Постройте модель линейной регрессии для прогнозирования цены на основе оценок. Каково значение R^2 и как оно интерпретируется?
9. Логистическая регрессия
Используя логистическую регрессию, предскажите, является ли вино из США на основе его оценок и цены. Обсудите точность модели.