**Лабораторная работа № 6**

**Тема**: Линейная регрессия.

**Цель работы** *–* изучить на практике понятие корреляции, реализовать модели линейной регрессии из библиотеки *sklearn.*

1. Выберите любой доступный на просторах интернета набор данных подходящий для регрессионного анализа (как правило это прогнозирование цены, зарплаты и т.д.). Как и в прошлом задании это можно сделать, введя запрос на сайте kaggle.

**Требование к набору данных: минимум три признака, описывающих объект (можно больше).**

2. Визуализируйте матрицу корреляций в виде тепловой карты.

3. Постройте матрицу диаграмм рассеяния.

4. Проанализируйте коэффициенты корреляции и диаграммы рассеяния и выберите параметры для построения модели простой линейной регрессии.

5. Рассчитайте модель и визуализируйте график модели на диаграмме рассеяния параметра (который учтен в модели)

6. Оцените качество полученной модели.

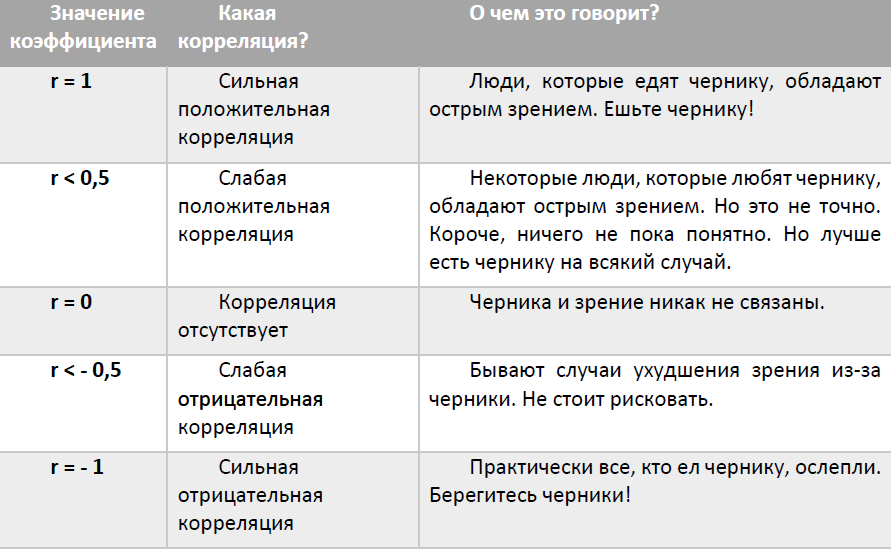
7. Добавьте в модель линейной регрессии еще несколько параметров, в той или иной мере влияющих на целевую переменную.

8. Рассчитайте модель с несколькими параметрами и оцените качество модели.

9. Сделайте вывод: какая модель лучше описывает зависимость целевой переменной от параметров.

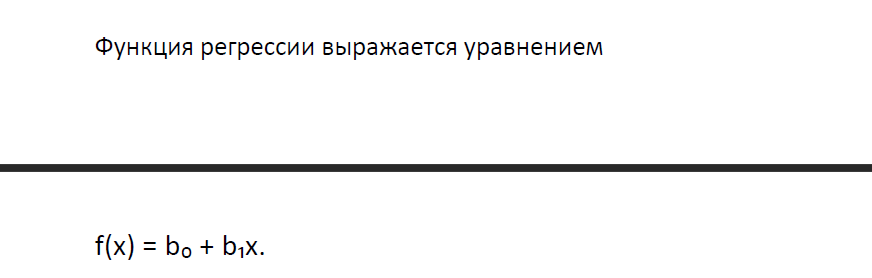
Вопросы:

1. Что такое корреляция? Корреля́ция (от лат. correlatio «соотношение»)— статистическая взаимосвязь двух или более случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми)
2. Что показывает коэффициент корреляции? Каким он может быть?

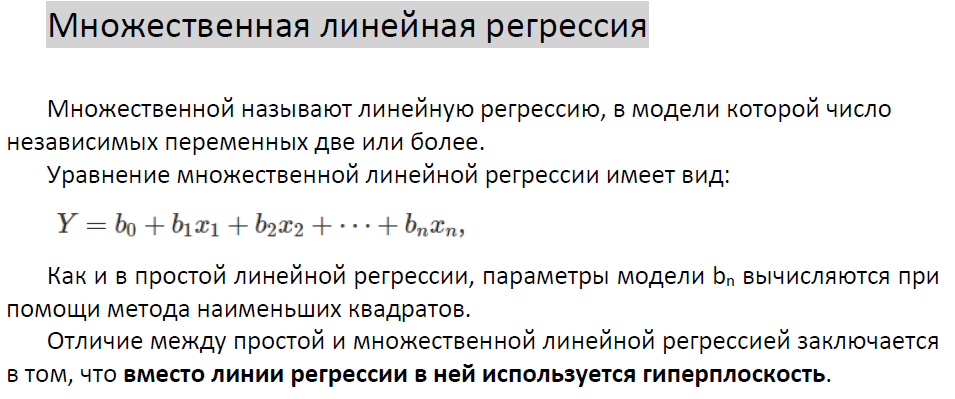
Математической мерой корреляции двух случайных величин служит **коэффициент корреляции**. 

1. Что представляет из себя модель простой линейной регрессии?

Простая или одномерная линейная регрессия – случай линейной регрессии с единственной независимой переменной x.



1. Чем множественная линейная регрессия отличается от простой линейной регрессии?



1. Что характеризует коэффициент детерминации?

Коэффициент детерминации характеризует долю вариации (дисперсии) результативного признака y, объясняемую регрессией, в общей вариации (дисперсии) y.

Коэффициент детерминации рассчитывается для оценки качества подбора уравнения регрессии. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50%. Модели с коэффициентом детерминации выше 80% можно признать достаточно хорошими. Значение коэффициента детерминации R2 = 1 означает функциональную зависимость между переменными.

1. Как сложность модели влияет на коэффициент детерминации?

При увеличении сложности модели, увеличивается значение коэффициента детерминации.

Основная проблема применения (выборочного) R^2 заключается в том, что его значение увеличивается (не уменьшается) от добавления в модель новых переменных, даже если эти переменные никакого отношения к объясняемой переменной не имеют! Поэтому сравнение моделей с разным количеством факторов с помощью коэффициента детерминации, вообще говоря, некорректно. Для этих целей можно использовать альтернативные показатели.