

Лабораторная работа 2-1

Лабораторная работа 2-1 рассчитана на два занятия и работу дома. Её целью является изучение основ корреляционного анализа данных в Python.

В качестве задания лабораторной работы студентам предлагается найти в интернете набор реальных статистических данных. Затем сформировать из него Pandas DataFrame и выполнить корреляционный анализ заданных пар переменных.

Для выполнения лабораторной работы студентам потребуется самостоятельно познакомиться с построением графика линейной регрессии.

Могут быть полезны следующие ссылки:

https://cbr.ru/currency_base/dynamics/

<https://bhom.ru/commodities>

<https://www.finam.ru/profile/tovary/brent/export>

Вариант 1

1. Соберите в интернете набор данных о курсах американского доллара, канадского доллара и евро не менее, чем за последние 50 месяцев (как минимум одно наблюдение на месяц). Курсы валют должны быть приведены к одинаковым единицам измерения. Например, количество рублей за единицу валюты или количество фунтов стерлингов за единицу валюты и т.д.
2. Загрузите собранную информацию в набор Pandas DataFrame, назвав переменные USD, CAD, EUR соответственно.
3. Выполните графическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным при помощи графиков P-P и Q-Q. При необходимости выполните нормализацию или стандартизацию данных.
4. Проверьте приблизительное равенство моды, медианы и среднего арифметического значения исследуемого набора данных. Постройте их на гистограмме исследуемого набора данных.
5. Выполните аналитическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным. Обосновано выберите один из методов: Шапиро-Уилка, Андерсона-Дарлинга или Колмогорова-Смирнова.
6. Удалите наиболее заметные выбросы и повторите исследование.
7. Для каждой из пар переменных выполните графическое исследование данной пары на предмет взаимной корреляции.
8. Для каждой из пар переменных выполните оценку с помощью четырёх методов (Пирсона, Спирмена, Кендалла). Объясните, какой из методов было лучше использовать в вашем случае. Дайте интерпретацию полученных результатов.
9. Для каждой из пар переменных постройте график, отображающий все точки наблюдений и функцию линейной регрессии.

Вариант 2

1. Соберите в интернете набор данных о цене барреля нефти BRENT и курсах российского рубля и нигерийской найры не менее, чем за последние 50 месяцев (как минимум одно наблюдение на месяц). Курсы валют и цена барреля должны быть приведены к одинаковым единицам измерения. Например, количество долларов за единицу валюты или баррель.
2. Загрузите собранную информацию в набор Pandas DataFrame, назвав переменные BAR, RUB, NGN соответственно.
3. Выполните графическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным при помощи графиков P-P и Q-Q. При необходимости выполните нормализацию или стандартизацию данных.
4. Проверьте приблизительное равенство моды, медианы и среднего арифметического значения исследуемого набора данных. Постройте их на гистограмме исследуемого набора данных.
5. Выполните аналитическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным. Обосновано выберите один из методов: Шапиро-Уилка, Андерсона-Дарлинга или Колмогорова-Смирнова.
6. Удалите наиболее заметные выбросы и повторите исследование.
7. Для каждой из пар переменных выполните графическое исследование данной пары на предмет взаимной корреляции.
8. Для каждой из пар переменных выполните оценку с помощью четырёх методов (Пирсона, Спирмена, Кендалла). Объясните, какой из методов было лучше использовать в вашем случае. Дайте интерпретацию полученных результатов.
9. Для каждой из пар переменных постройте график, отображающий все точки наблюдений и функцию линейной регрессии.

Вариант 3

1. Соберите в интернете набор данных о ставках рефинансирования и курсах национальных валют России и Казахстана не менее, чем за последние 50 месяцев (как минимум одно наблюдение на месяц). Курсы валют должны быть приведены к одинаковым единицам измерения. Например, количество долларов за единицу валюты.
2. Загрузите собранную информацию в набор Pandas DataFrame.
3. Выполните графическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным при помощи графиков P-P и Q-Q. При необходимости выполните нормализацию или стандартизацию данных.
4. Проверьте приблизительное равенство моды, медианы и среднего арифметического значения исследуемого набора данных. Постройте их на гистограмме исследуемого набора данных.
5. Выполните аналитическое исследование для каждой из переменных на предмет совпадения её распределения с нормальным. Обосновано выберите один из методов: Шапиро-Уилка, Андерсона-Дарлинга или Колмогорова-Смирнова.
6. Удалите наиболее заметные выбросы и повторите исследование.
7. Для каждой из пар переменных выполните графическое исследование данной пары на предмет взаимной корреляции.
8. Для каждой из пар переменных выполните оценку с помощью четырёх методов (Пирсона, Спирмена, Кендалла). Объясните, какой из методов было лучше использовать в вашем случае. Дайте интерпретацию полученных результатов.
9. Для каждой из пар переменных постройте график, отображающий все точки наблюдений и функцию линейной регрессии.