

ИССЛЕДОВАНИЕ ЛИНЕЙНОЙ ВЗАИМОСВЯЗИ МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ НА ОСНОВЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ПИРСОНА

К.Д. Сухих, И.А. Ботыгин
Томский политехнический университет
kds7@tpu.ru

Введение

На сегодняшний день стремительно развиваются технологии обработки метеорологических данных, помогая прогнозировать климатические изменения. Одним из широко используемых инструментов анализа метеорологических данных является вероятностно-статистический подход. В рамках этого подхода достаточно эффективным средством исследования метеорологических рядов наблюдений является корреляционная теория.

В данной работе в качестве метода исследования взаимосвязи основных метеорологических параметров был выбран корреляционный анализ, который использовался для выявления их статистической зависимости за установленный промежуток времени на данных по Томской области.

Корреляция Пирсона

Для выявления меры корреляции между двумя метеорологическими параметрами был использован коэффициент корреляции Пирсона. Заметим, что данный подход позволяет определить лишь наличие линейной взаимосвязи между исследуемыми параметрами.

На входные метеоданные были наложены следующие ограничения:

- показатели должны быть измерены в количественной шкале;
- на вход должно подаваться не более двух типов метеоданных – исследовалась только парная корреляция.

Расчет коэффициента корреляции Пирсона для выборки наблюдений для пар переменных (x, y) производится по следующей формуле:

$$r_{xy} = \frac{\sum(d_x \times d_y)}{\sqrt{(\sum d_x^2 \times \sum d_y^2)}}$$

Отметим свойства коэффициента корреляции [1]:

- коэффициент измеряется в интервале от -1 до +1;
- положительный коэффициент показывает, как одна переменная увеличивается по мере того, как увеличивается другая;
- отрицательный коэффициент показывает, как одна переменная уменьшается по мере того, как увеличивается другая;
- корреляция между величинами не обязательно означает соотношение причины и следствия;
- величина коэффициента будет обоснована только в диапазоне значений выборки;
- коэффициент корреляции безразмерен.

Используемые библиотеки Python

Для подсчета коэффициента корреляции Пирсона была использована библиотека `scipy` Python, из которой была выбрана функция `scipy.stats.linregress()`, которая вычисляет регрессию наименьших квадратов для двух наборов измерений. На выходе функция возвращает несколько параметров, например: наклон линии регрессии и необходимый нам коэффициент корреляции Пирсона.

Для вывода диаграммы рассеяния и построения графиков была использована библиотека `matplotlib.pyplot.plot`.

Для анализа входных данных была использована библиотека `pandas`.

Для визуализации статистических данных была использована библиотека `seaborn`.

Для математических вычислений была использована библиотека `numpy`.

Реализация программы

Для реализации программы был выбран язык высокого уровня Python, написание программы проводилось в онлайн среде Colaboratory от Google. Данная среда была выбрана по нескольким важным параметрам:

1. Возможность запуска программы на любых устройствах с доступом в интернет;
2. Простой и понятный интерфейс;
3. Множество встроенных библиотек, которые не нужно дополнительно устанавливать.

В качестве входных данных был использован ресурс, который за указанный период времени предоставляет информацию о метеорологических и радиационных данных по Томской области [2]. Изображение используемого ресурса представлено на рисунке 1.

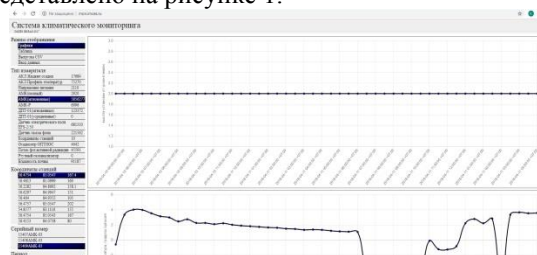


Рис. 1. Изображение сайта «Система климатического мониторинга»

В качестве регистратора была выбрана ультразвуковая метеостанция. С данного ресурса за период с 10.04.2018 по 11.04.2018 были получены следующие метеорологические параметры:

1. Время измерения.
2. Индекс группы на момент измерения.

3. Серийный номер.
4. Высота установки от уровня земли, м.
5. Средняя температура, градусы Цельсия.
6. Средняя скорость горизонтального ветра, м/с.
7. Среднее направление горизонтального ветра, градусы.
8. Минимальная скорость мгновенного горизонтального ветра, м/с.
9. Максимальная скорость мгновенного горизонтального ветра, м/с.
10. Средняя скорость вертикального ветра, м/с.
11. Атмосферное давление, мм. рт. ст.
12. Относительная влажность воздуха, %.
13. Температура точки росы, градусы Цельсия.
14. Упругость (давление) водяного пара, гПа.
15. Абсолютная влажность, г/м³.
16. Плотность воздуха, кг/м³.
17. Скорость звука в воздухе, м/с.

В частности, коэффициент корреляции Пирсона рассчитывался для следующих двух пар параметров: «средней скорости вертикального ветра» и «температуры» за сутки, и для «атмосферного давления» и «температуры» также за сутки.

Первый шаг программы – загрузка текстового файла с метео данными, и, по разделителю «;», реструктуризация на столбцы по значениям.

Далее были рассчитаны коэффициенты корреляции Пирсона для двух указанных выше пар значений. Диаграммы рассеяния, построенные по значениям коэффициентов и периодам, за которые коэффициент рассчитывался, представлены на рисунке 2 и рисунке 3.

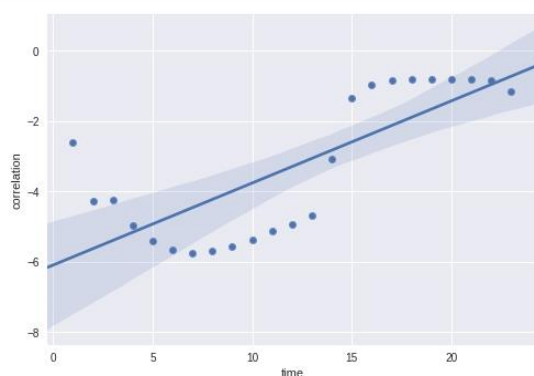


Рис. 2. Диаграмма рассеяния коэффициента корреляции для температуры и атмосферного давления (на оси x показано время в часах, по оси y – коэффициент корреляции)

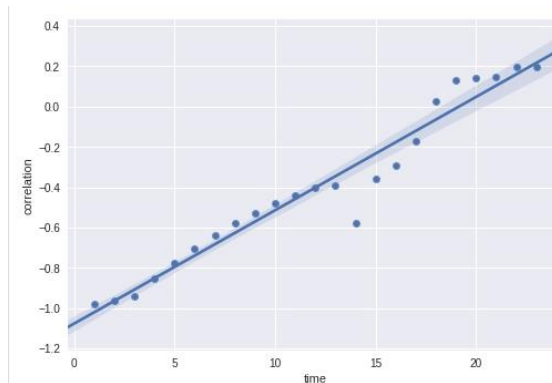


Рис. 3. Диаграмма рассеяния коэффициента корреляции для температуры и средней скорости вертикального ветра (на оси x показано время в часах, по оси y – коэффициент корреляции)

Следующий шаг – построение графиков корреляционных зависимостей параметров. На рисунке 4 изображен пример такого графика.

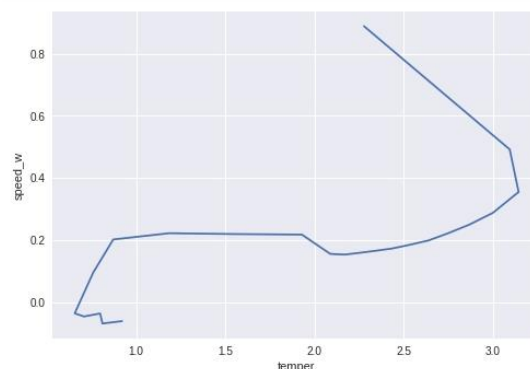


Рис. 4. График корреляционной зависимости средней скорости ветра и температуры за сутки

Заключение

В результате проделанной работы была написана программа на Python, которая рассчитывает коэффициент корреляции Пирсона для заданных метеорологических параметров за определенный период и строит графики корреляционной зависимости этих параметров.

Данная программа может быть расширена для дальнейших исследований метеорологических данных.

Список использованных источников

1. Коэффициент корреляции // Портал знаний. URL <http://statistica.ru/theory/koeffitsient-korrelyatsii/> (дата обращения: 20.11.2018).
2. Система климатического контроля // IMCES SB RAS. URL <http://mon.imces.ru> (дата обращения: 19.11.2018).
3. Matplotlib // Учебники. URL <https://matplotlib.org/tutorials/index.html> (дата обращения: 20.11.2018).
4. Scipy.stats.linregress // Scipy.org. URL <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html#scipy.stats.linregress> (дата обращения: 20.11.2018).