

Задание 1

Имеется файл `BindingEnergy.txt`, содержащий информацию в csv-формате. Напишите программу для чтения данных из этого файла. Найдите и выведите на экран 5 строк, которые имеют наибольшую сумму квадратов отклонений относительно первого столбца – контроля. Напишите функцию для поиска моды в столбцах (без использования сторонних библиотек).

Задание 2

Найдите и загрузите одну запись последовательности из genbank (можно без использования python). Идентификатор записи `HE805982.1`. Эта запись содержит информацию об участке ДНК, кодирующем HBx, многофункциональный вирусный белок гепатита В. Напишите программу, которая выделяет из genbank информацию об особенностях (features) и последовательность белка (из cds) и сохраняет в разные файлы. Оформите чтение и запись в файл как функции. Функции запрашивают имя файла как входной параметр, вводимый с клавиатуры; сами файлы хранятся в папке `./parsing_genbank` где `.` – ваша текущая рабочая папка. Сделать assert проверку на существование файла (через `import os`; функцию `os.path.exists (path)`)

Задание 3

Прочитайте содержимое текстового файла `cox1.fasta`, который содержит набор последовательностей. Выведите на экран считанные данные. Напишите код для преобразования `cox1.fasta` файла в текстовый табличный файл, с табуляцией в качестве разделителя между ячейками. В файле должны быть ID последовательности, её длина и сама последовательность.

Результат (пример результата, у вас может отличаться):

ID length sequence

O78750 230 MAYPMQLGFQDATSPIMEELLHINDSGETVEFE...

P00411 256 MGLLFNNLIMNFDAPSPW...

P00412 764 IVTSADVLHSWA...

Напишите функцию, которая бы разделяла мультифасту `cox1.fasta` на отдельные fasta-файлы. В качестве имени новых файлов используйте информацию из заголовков (например, `O78750_COX2_SHEEP.fasta`).