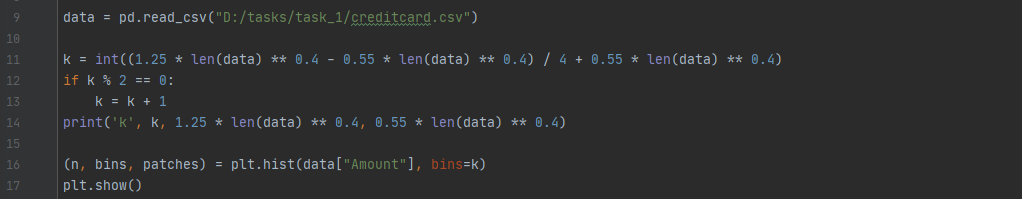
**Программирование в системах информационной безопасности  
 Лабораторный практикум**

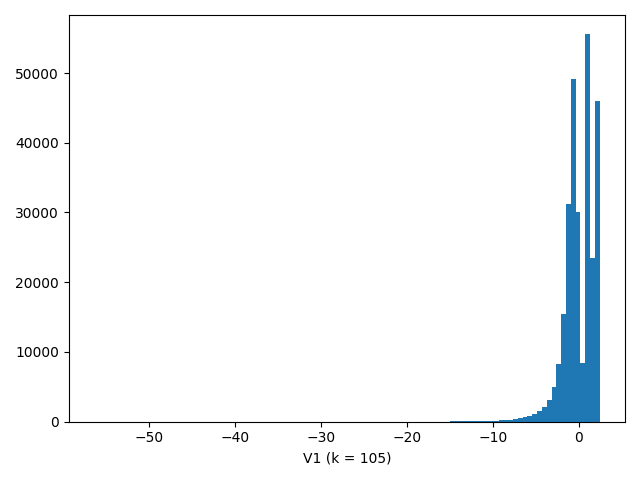
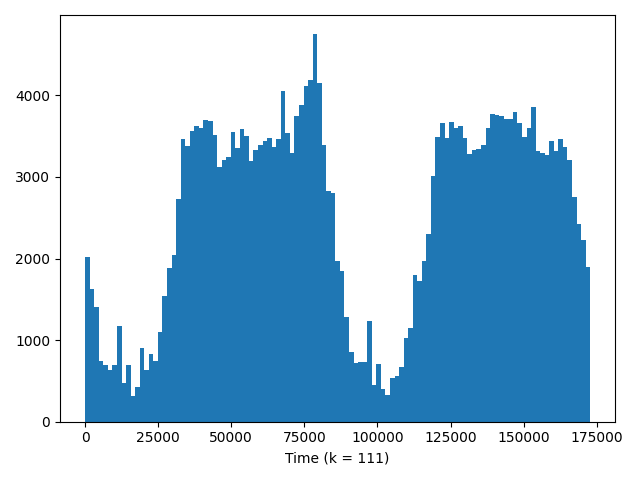
Васюткин Александр Олегович, группа 181-331

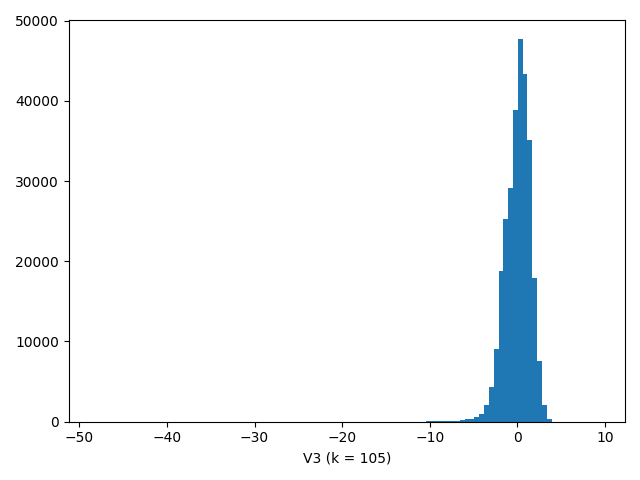
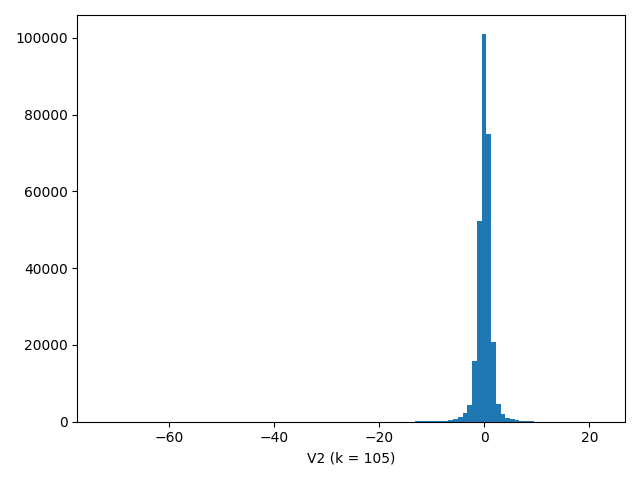
**Задание 1.2.**

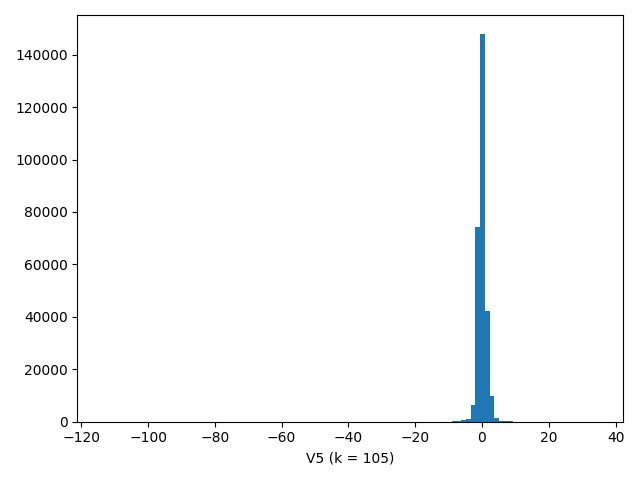
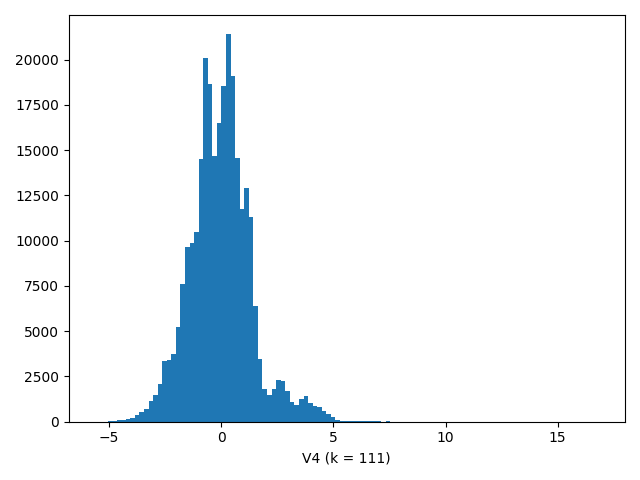
1. Построить гистограмму.

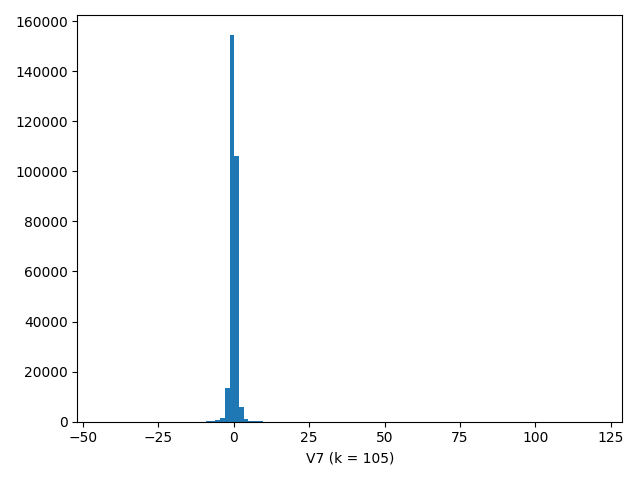
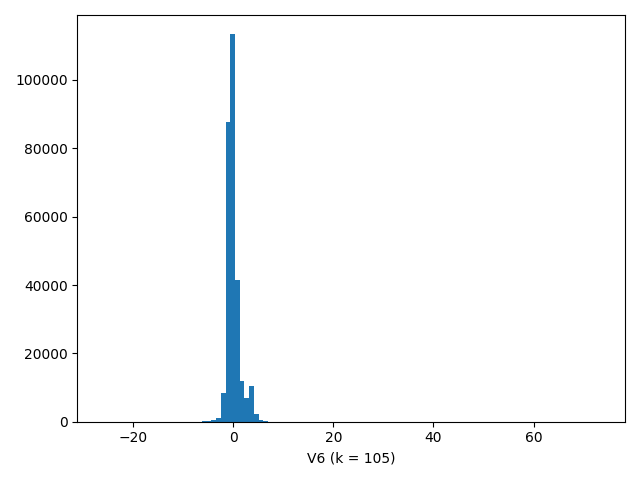


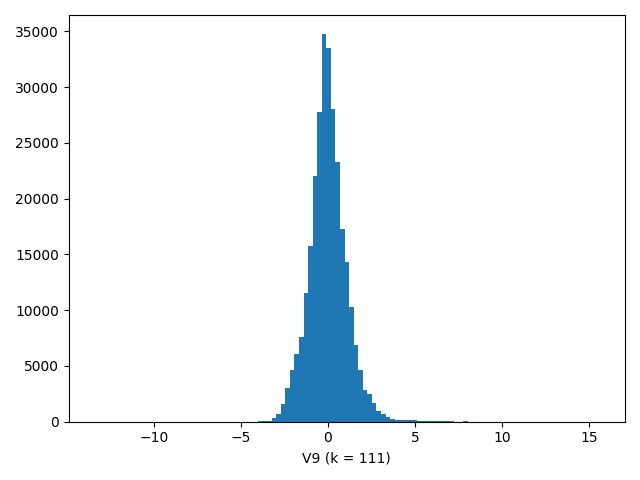
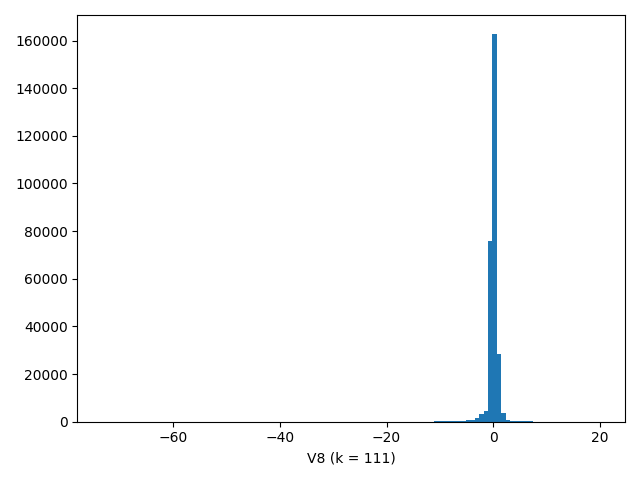
Чтобы построить гистограмму, необходимо определить оптимальное число интервалов k из промежутка , что мы и делаем в 12 строке, объединив формулы нахождения минимального и максимального значений интервала. Так как k должно быть нечетным, мы проверяем его четность, и, если это необходимо, прибавляем единицу. Для каждого признака число интервалов подгонялось вручную, чтобы гистограмма была наиболее подробной. Наконец, найдя количество интервалов, мы можем построить гистограмму (при помощи библиотеки matplotlib), записав в переменную n количество значений случайной величины, входящих в каждый интервал, a в bins – значения границ каждого интервала. Полученные гистограммы:

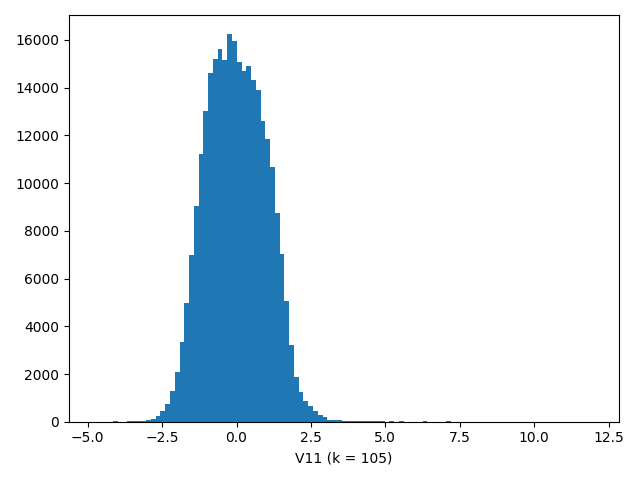
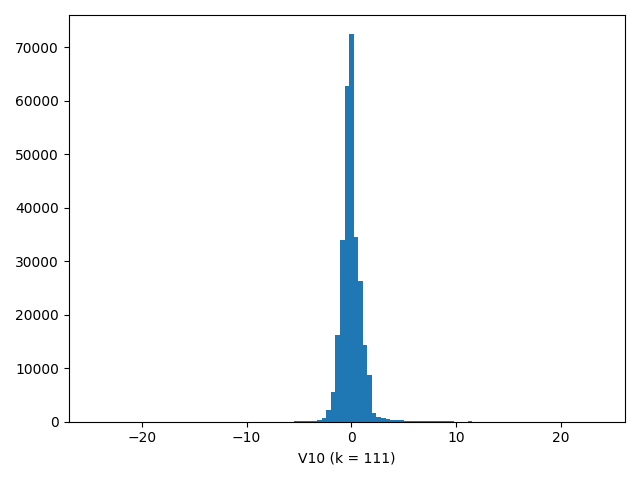


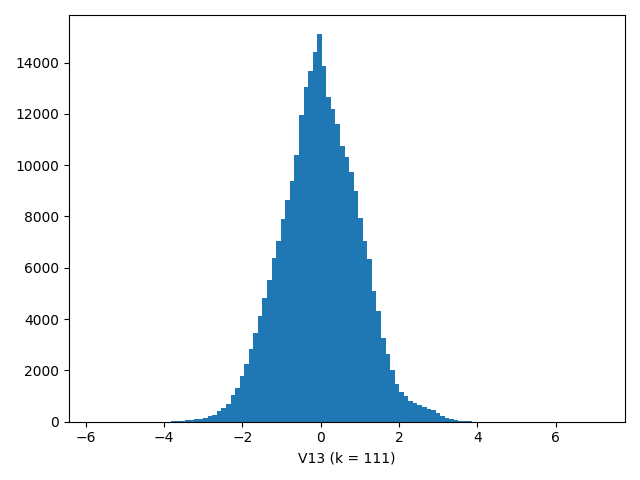
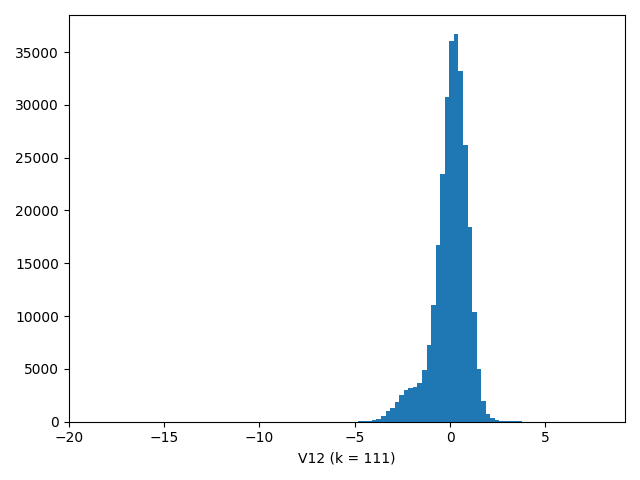


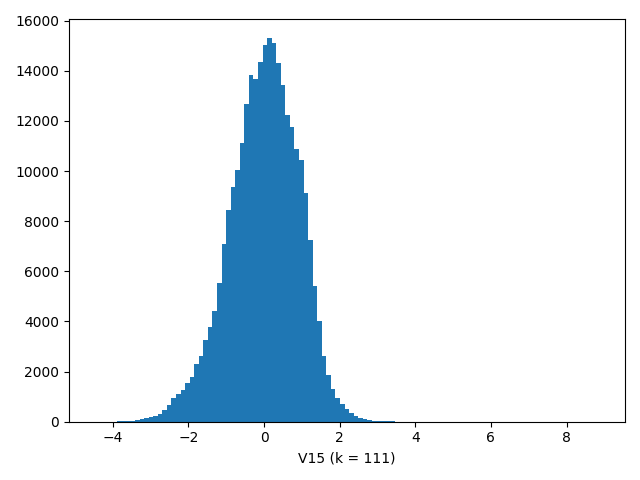
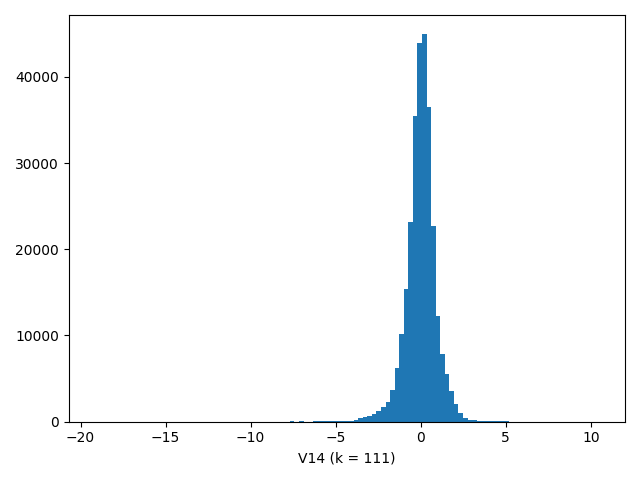


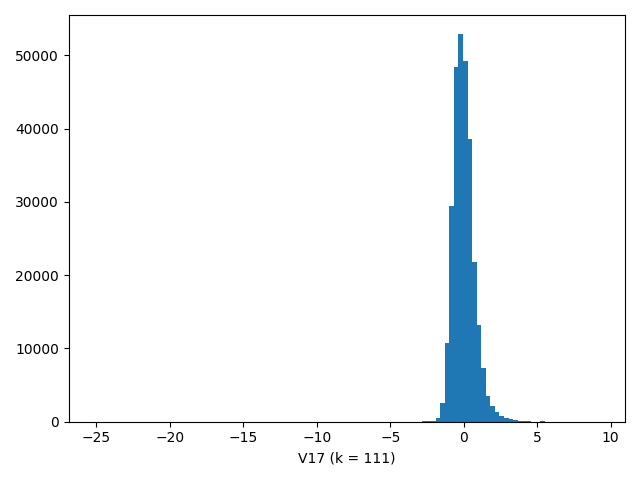
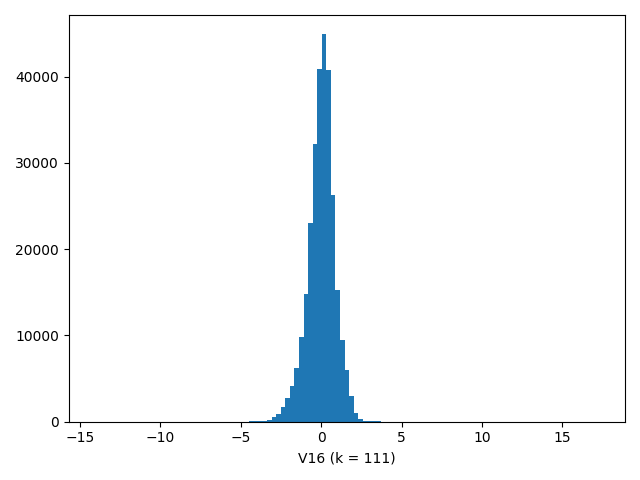


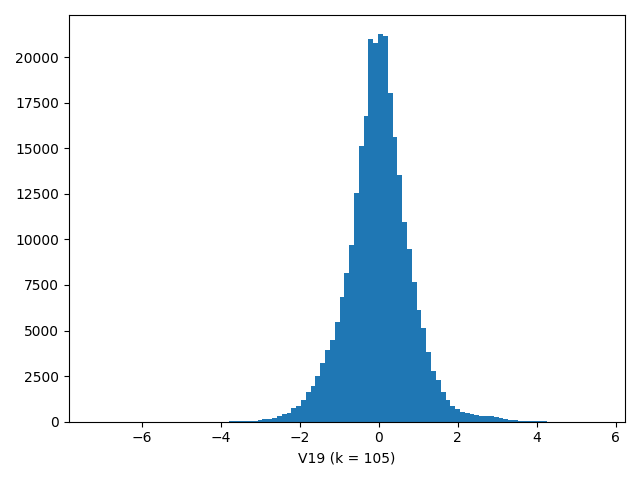
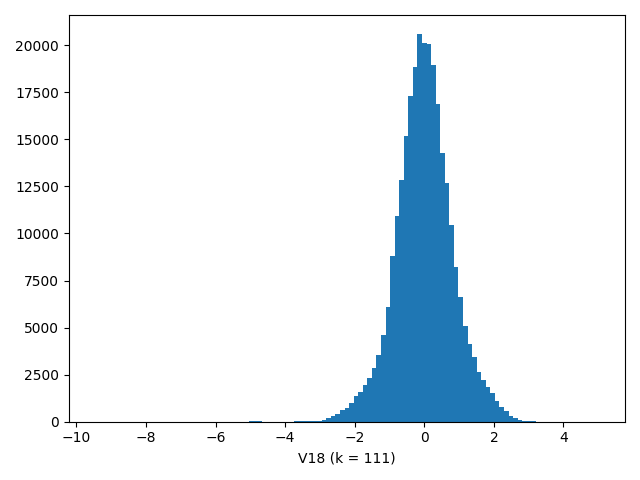


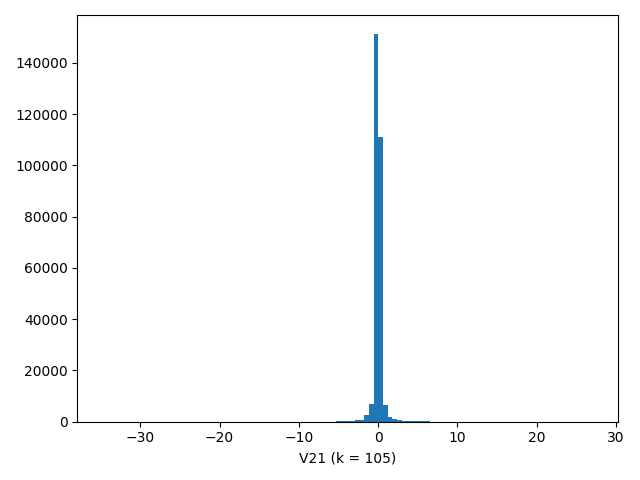
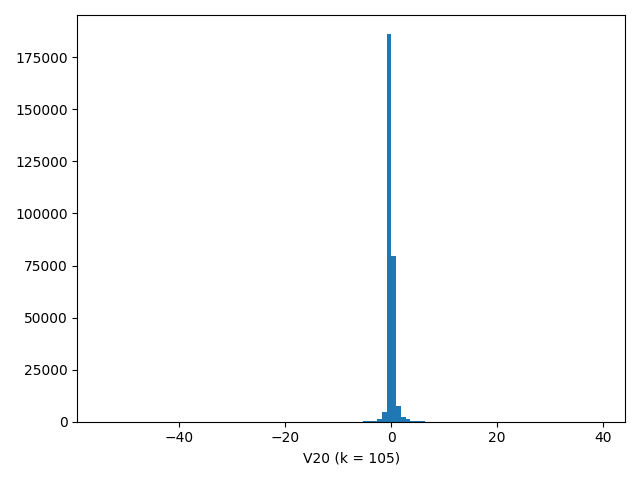


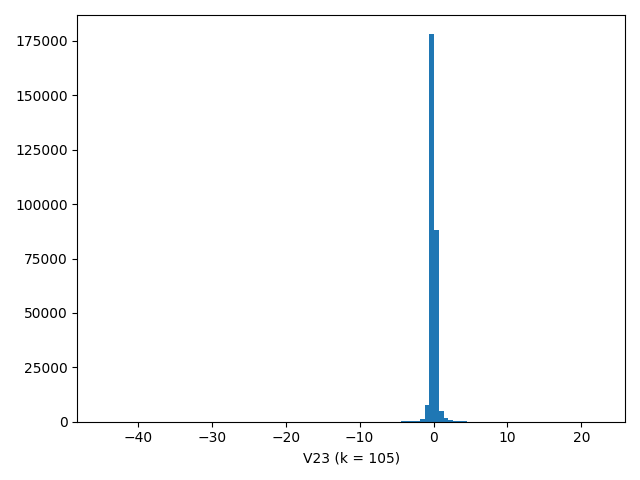
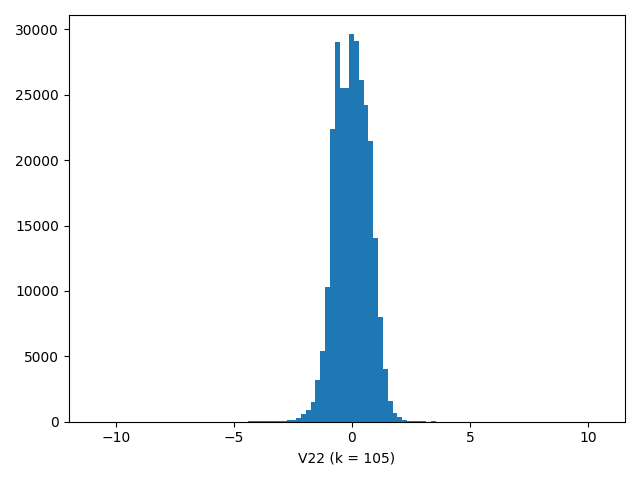


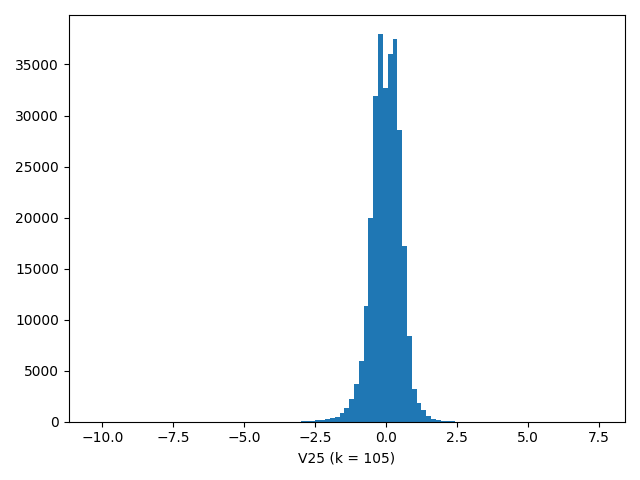
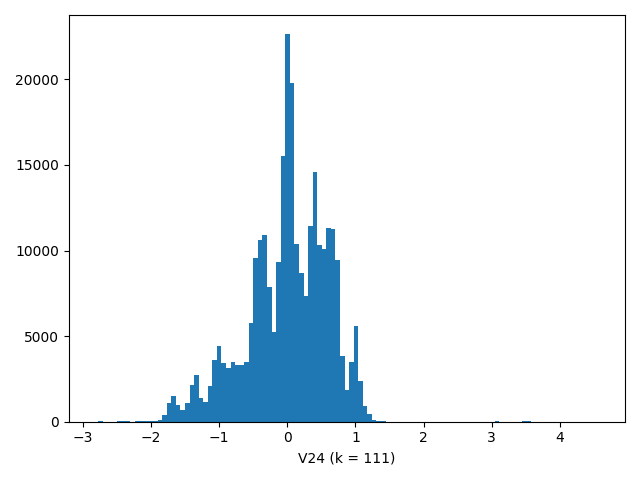


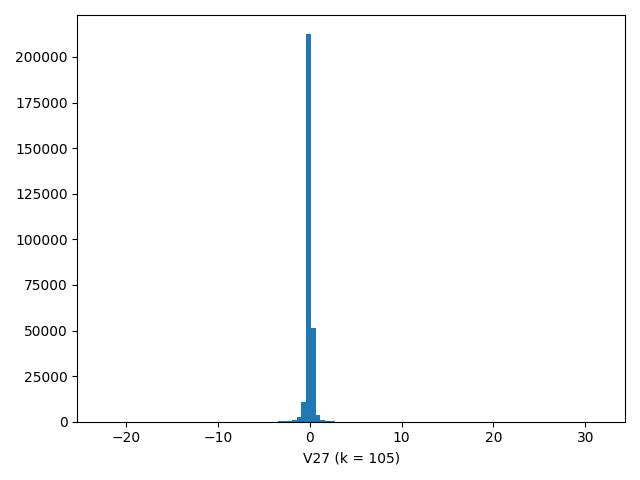
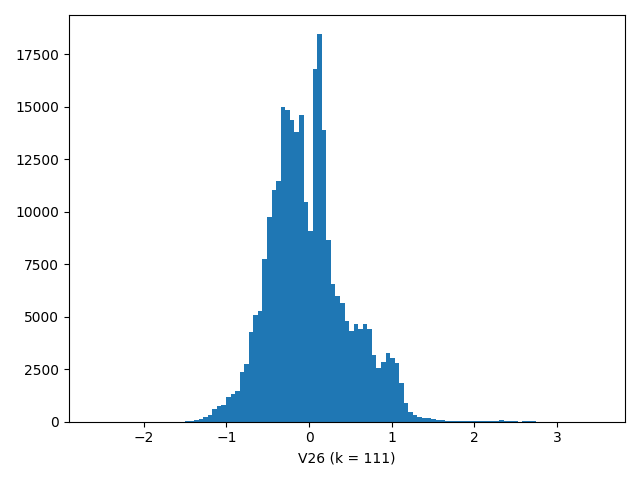


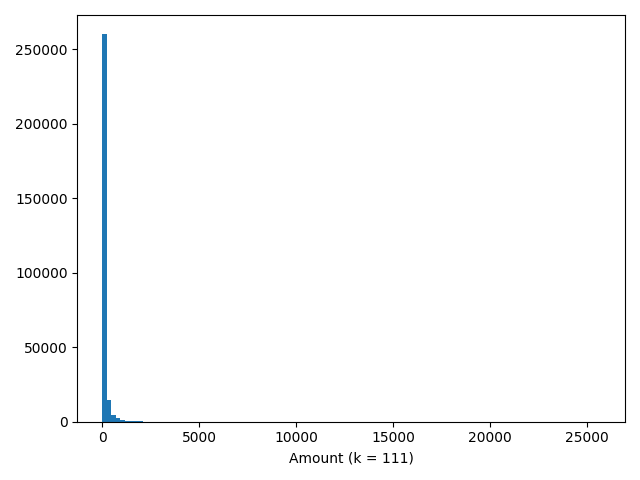
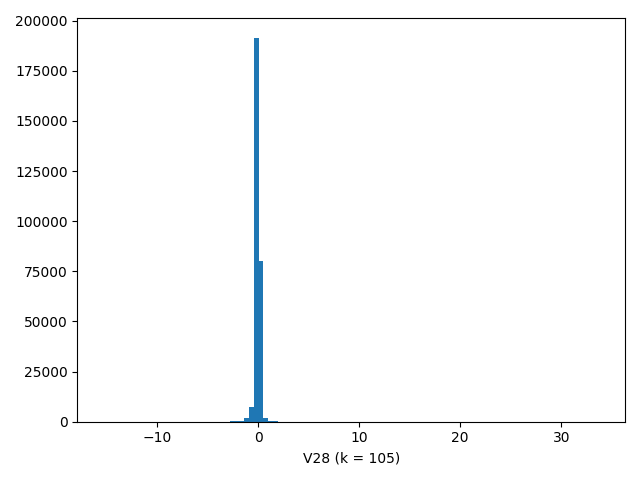




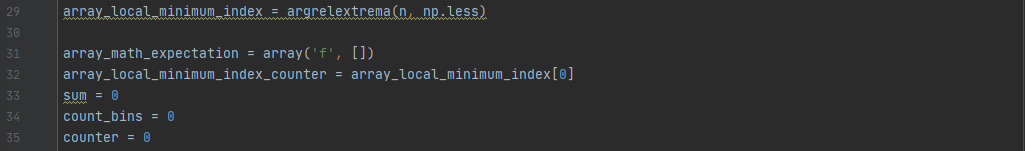








1. Вычислить значения признака, соответствующие локальным минимумам плотности частоты h.

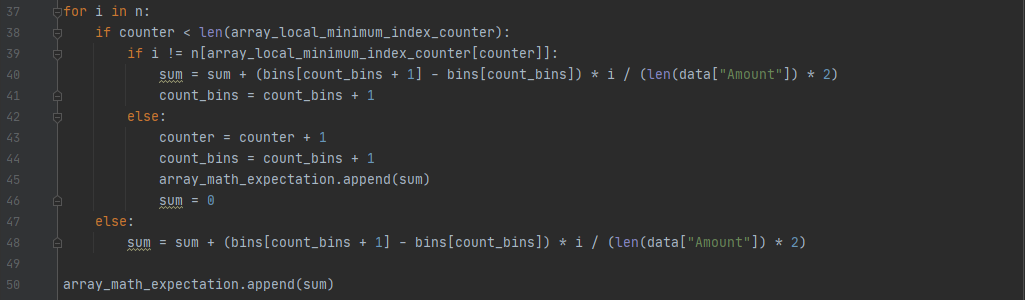


Локальные минимумы ищем в 29 строке при помощи метода из библиотеки ScyPi. В 31 строке создаем массив, в который будем записывать значения математических ожиданий каждой области однородности. В строке 32 мы создаем переменную, которую будем использовать для перебора индексов локальных минимумов. Далее объявляем три переменные, необходимые для нахождения математического ожидания для каждой области однородности, перебора значений границ интервалов (те самые интервалы, количество которых равно k) и определения конца массива с индексами локальных минимумов.

1. По вычисленным значениям разбить область значений признака на непересекающиеся области однородности.

Я написал программу таким образом, что данный пункт объединяется со следующим – разбил область значений признака на непересекающиеся области однородности прям внутри условия.

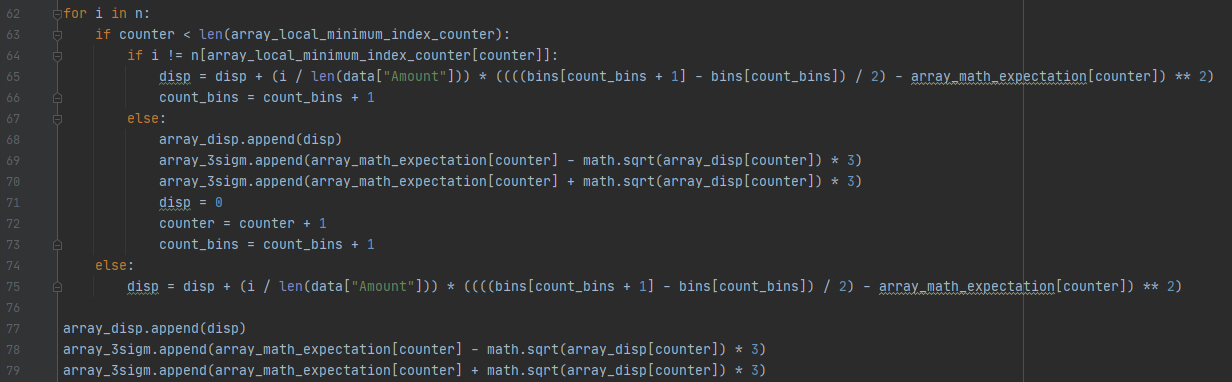
1. Каждую область однородности представить математическим ожиданием μ и границами доверительного интервала ±3σ.



На данном фрагменте программа находит математическое ожидание для каждой области однородности, руководствуясь следующей формулой: В цикле мы поочередно обрабатываем все значения массива n, полученные из гистограммы, сравнивая каждое значение со значением локального минимума. Пока не встретится значение, равное значению первого локального минимума, мы суммируем произведения значений середин интервалов, входящих в область однородности, и их вероятностей (находим середину интервала, поделив пополам разность значений правой и левой границ интервала, и умножаем полученное значение на частное значения, соответствующего количеству значений случайной величины, входящих в интервал, и значения, равного количеству всех значений случайной величины). Здесь же мы переходим к следующему значению в массиве с границами интервалов. Когда встретится значение из массива n, равное значению, которое соответствует локальному минимуму, мы записываем полученную сумму в массив, обнуляем переменную, где хранилось сумма, и переходим к следующему локальному минимуму, чтобы продолжить обрабатывать значения, соответствующие количеству значений случайной величины, входящих в интервал. Часть кода (47-50 строки) необходима для обработки значений массива n, входящих в интервалы последней области однородности (когда все локальные минимумы закончились).

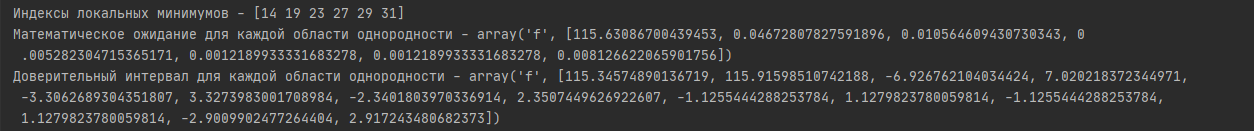


Объявляем два массива, в которые будем записывать значения дисперсии и границ доверительного интервала соответственно, а также три переменные, которые выполняют аналогичные объявленным перед предыдущим условием переменным функции.



На данном фрагменте программа находит значения границ доверительного интервала для каждой области однородности. В цикле мы поочередно обрабатываем все значения массива n, полученные из гистограммы, сравнивая каждое значение со значением первого локального минимума. Пока не встретится значение, равное значению локального минимума, мы рассчитываем дисперсию для каждой области однородности по следующей формуле Здесь же мы переходим к следующему значению в массиве с границами интервалов. Как только среди обрабатываемых значений встречается значение, равное значению локального минимума, мы записываем полученную дисперсию в массив, высчитываем границы доверительного интервала () для рассматриваемой области однородности и записываем полученные значения в массив, обнуляем переменную, где хранилось значение дисперсии, и переходим к значению следующего локального минимума, чтобы продолжить работу цикла. Часть кода (74-79 строки) необходима для обработки значений, соответствующих количеству значений случайной величины, входящих в интервалы последней области однородности (когда все локальные минимумы закончились).

В результате работы программы мы получаем два массива: массив с математическими ожиданиями и массив с границами доверительных интервалов каждой области однородности.



**Задание 1.3.** Для каждого качественного признака по частотам значений построить столбчатую диаграмму.

Так как из 31 признака в датасете 30 являются количественными (количество секунд, пройденное с момента осуществления первой транзакции, сумма транзакции и 28 признаков, преобразованных при помощи метода PCA), а метка не является признаком, получается, что датасет не содержит качественных признаков.