**Лабораторная работа №1**

«Первичный графический анализ статистических данных»

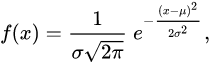
**Тема**: Динамика изменения оценок по теории вероятности и математической статистики за два года.

**Выполнил**: Николай Окуньков, 18ПИ-2.

**Цель работы:** Создание статистического ряда и изучение графических методов первичного анализа статистических данных с использованием встроенных в базовую версию пакета R функций.

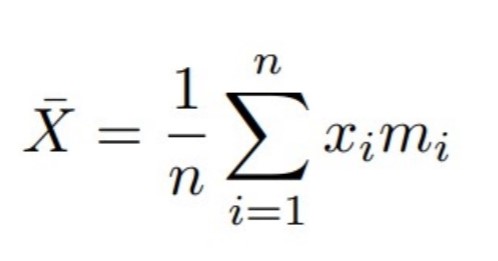
**Теоретическая часть**:

*Одномерное непрерывное равномерное распределение* – распределение случайной вещественной величины, принимающей значения, принадлежащие некоторому промежутку конечной длины, характеризующееся тем, что плотность вероятности на этом промежутке почти всюду постоянна.

*Нормальное распределение (распределение Гаусса)* – распределение вероятностей, которое в одномерном случае задаётся функцией плотности вероятности, совпадающей с функцией Гаусса:

где ***σ^2*** – стандартное отклонение, ***σ*** — среднеквадратическое отклонение распределения, ***μ*** — математическое ожидание (среднее значение), медиана и мода распределения.

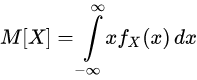
Пусть X1, ..., Xn – выборка из распределения вероятности, определенная на некотором вероятностном пространстве. Тогда её *выборочным средним* называется случайная величина:



где ***x(i)*** – элемент выборки под номером ***i***, ***m(i)*** – частота выборки данного элемента, а **n** – общий объем выборки.

*Относительная частота случайного события* – отношение числа появления данного события к общему числу проведенных одинаковых испытаний, в каждом из которых могло появиться или не появиться данное событие.

*Математическое ожидание* – среднее (взвешенное по вероятностям возможных значений) значение случайной величины. Для непрерывных случайных величин находится по следующей формуле:



*Дисперсия* – мера разброса значений случайной величины относительно её математического ожидания. Находится по формуле:



*Среднеквадратическое отклонение*– показатель рассеивания значений случайной величины относительно её математического ожидания. Находится по формуле:

*Гистограмма* – наглядное представление функции плотности вероятности некоторой случайной величины, построенное по выборке.

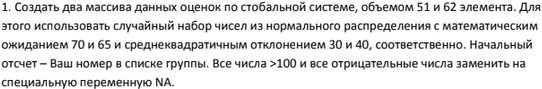
*Коробка с усами (диаграмма размаха*) – график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Границами ящика служат первый и третий квартили (25-й и 75-й процентили соответственно), линия в середине ящика — медиана (50-й процентиль). Концы усов — края статистически значимой выборки (без выбросов), минимальное и максимальное наблюдаемые значения данных по выборке (в этом случае выбросы отсутствуют).

*Квантиль* – значение, которое заданная случайная величина не превышает с фиксированной вероятностью. Если вероятность задана в процентах, то квантиль называется процентилем или перцентилем.

* 0,25 – квантиль называется первым или нижним квартилем;
* 0,50 – квантиль называется медианой или вторым квартилем;
* 0,75 – квантиль называется третьим или верхним квартилем.

*Процентиль* – это процентная доля элементов из выборки стандартизации, первичный результат которых ниже данного первичного показателя.

**Ход работы:**

Задание 1:

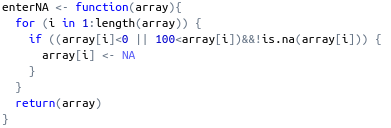
Для решения задачи была использована функция **R rnorm()** и собственная функция **enterNA()**:

**rnorm(n, mean= , sd= )** – функция для создания выборки, используя нормальное распределение;

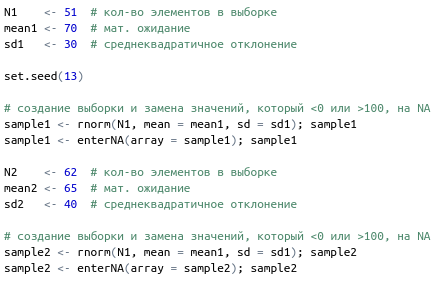
(Параметры: **n** – количество элементов выборки; **mean** –математическое ожидание; **sd** – среднеквадратическое отклонение.)

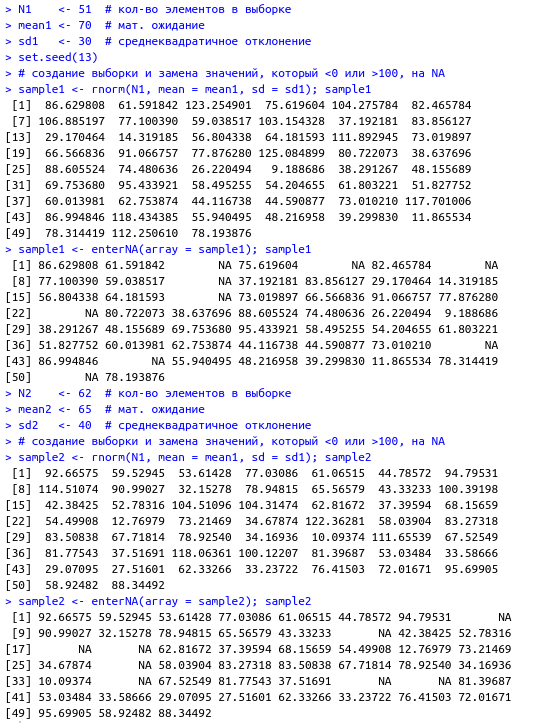
**enterNA(arr)** – нужна для замены всех отрицательных значений и значений, превышающих 100 на NA.

(Параметры: **arr** — выборка.)

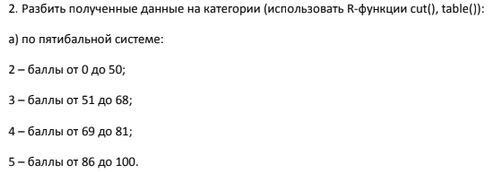


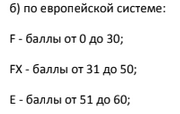
Код:

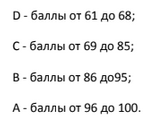


Результат:

Задание 2:



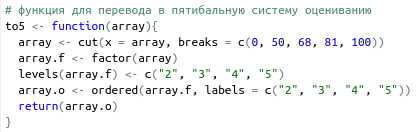




Для решения задачи были использованы собственные функции **to5()** и **toEU()**:

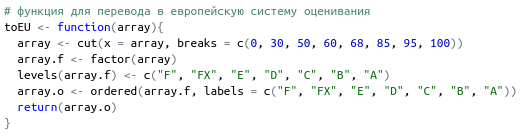
**to5(array)** – переводит оценки из 100-бальной системы в 5-бальную посредством разбиения выборки с помощью функции **cut()**, которая получает на вход вектор и делит их на равные или заранее заданные интервалы;

(Параметры: **array** — выборка.)



**toEU(array)** – переводит оценки из 100-бальной системы в 7-бальную посредством разбиения выборки с помощью функции **cut()**, которая получает на вход вектор и делит их на равные или заранее заданные интервалы.

(Параметры: **array** — выборка)

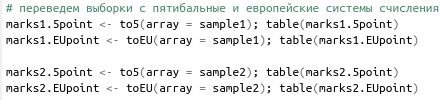


Для проверки корректности работы этих функций я использовал функцию R **table()**:

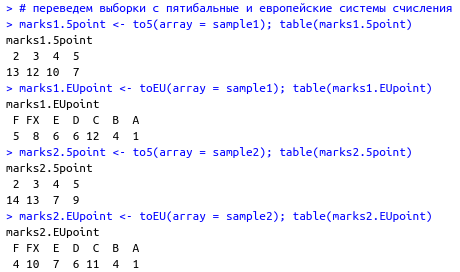
**table(x)** – возвращает таблицу с частотами встречаемости каждого значения **х**.

(Параметры: **x** — набор значений.)

Код:



Результат:



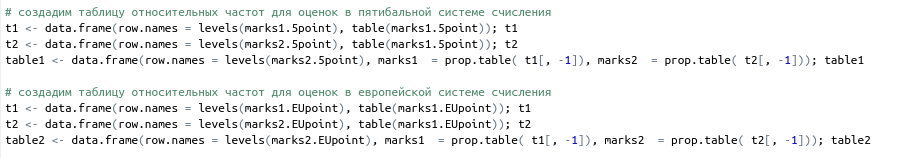
Задание 3:



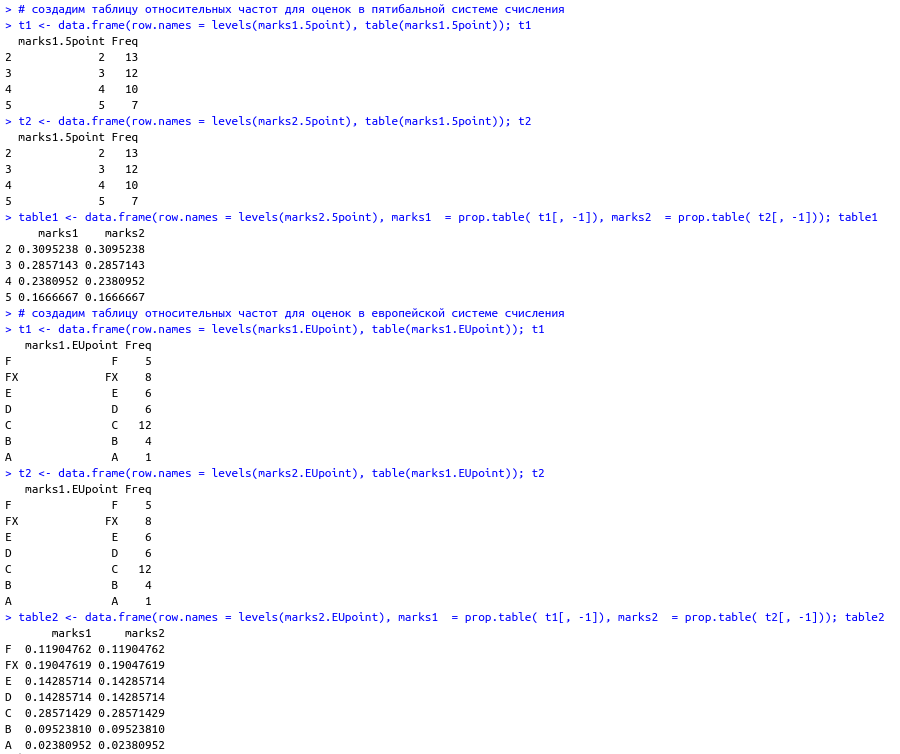
Для решения задачи были использованы собственные функция **data.frame()**:

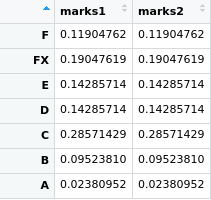
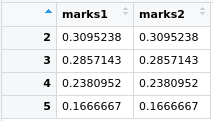
**data.frame() -** создает таблицу данных из поименованных или непоименованных аргументов;

Код:



Результат:





Задание 4:



Для решения задачи были использованы собственные функции **par()**, **hist()**, **lines()**, **density()**:

**par()** – функция для разделения пространства, на котором будут графики, на некое определенное количество маленьких пространств, для того, чтобы нарисовать сразу несколько графиков на одном экране;

**hist(x, breaks = )** – функция для создания гистограмм частот значений переменной **х** (аргумент **breaks =** можно использовать, чтобы изменить принятое по умолчанию количество столбцов);

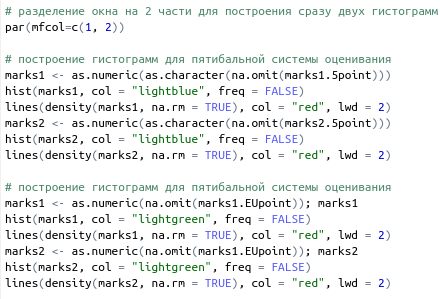
(Параметры: **x** — переменная; **breaks** — количество столбцов.)

**lines(x, col = , lwb = )** – функция для создания линий на графиках;

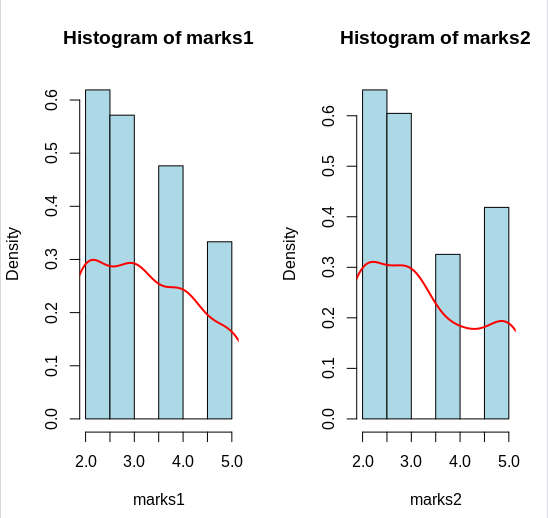
(Параметры: **x** — значения; **col** — цвет; **lwd** — толщина линии.)

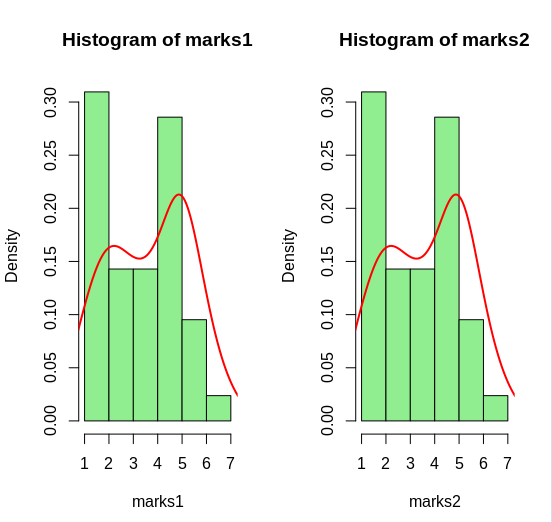
**density()** – функция для нахождения ядерных плотностей вероятностей (ядерная плотность вероятности — оценка случайной величины);

Код:



Результат:





Задание 5:



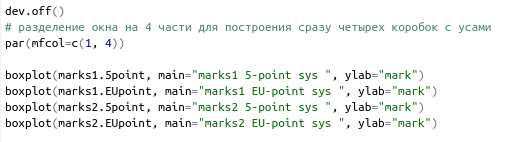
Для решения задачи были использованы функции **boxplot()** и **dev.off()**:

**dev.off()** – функция для очистки окна вывода.

**boxplot(x, main = , ylab = )** – функция для построения диаграмм размахов («коробок с усами»).

(Параметры: **x** — выборка; **main** — название графика; **ylab** — название осиOY.)

Код:



Результат:

