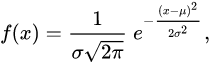
Лабораторная работа №1 «Первичный графический анализ статистических данных»  
Тема: Динамика изменения оценок по теории вероятности и математической статистике за два года.

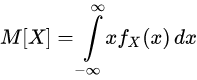
Выполнил: Саратовцев Артем 18Пи-1

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

ТЕОРИЯ:

**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 параметр μ — математическое ожидание (среднее значение), медиана и мода распределения, а параметр σ — среднеквадратическое отклонение (σ ² — дисперсия) распределения.

**Математическое ожидание** - среднее (взвешенное по вероятностям возможных значений) значение случайной величины. Для непрерывных случайных величин находится по 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;

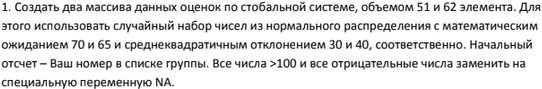
0,50 - квантиль называется медианой или вторым квартилем;

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 этого я воспользовался функцией R rnorm() и собственной написанной функцией insetrNA().

Функция rnorm(n, mean= , sd= ):

значения:

n - кол-во элементов выборки

mean - математическое ожидание

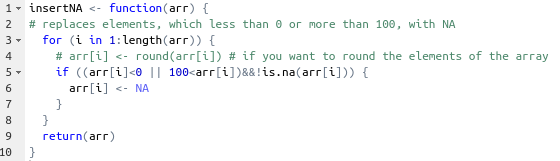
sd - среднеквадратическое отклонение

нужна для создания выборки, используя нормальное распределение.

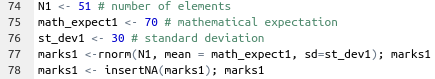
Функция insertNA(arr):

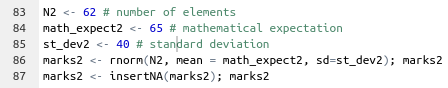
arr — выборка

нужна для замены всех отрицательных значений и значений, превышающих 100 на NA.

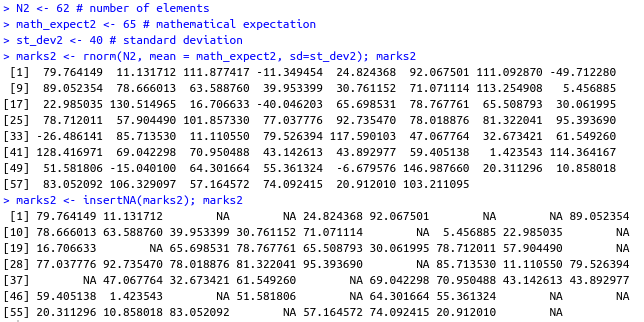
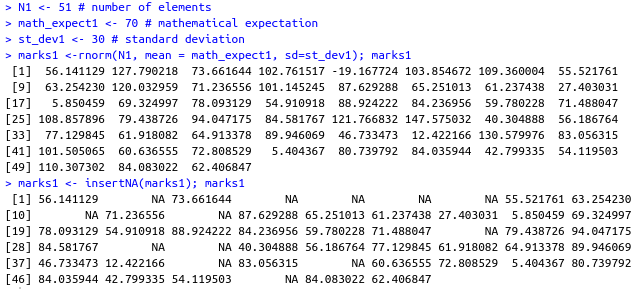


Код:

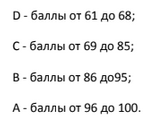
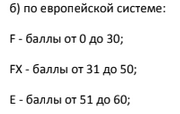
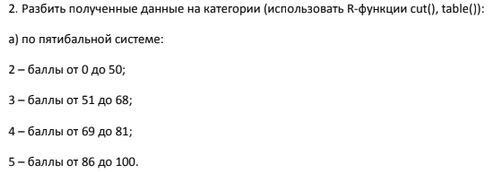




Результат выполнения (на следующей стр.):



Задание 2:

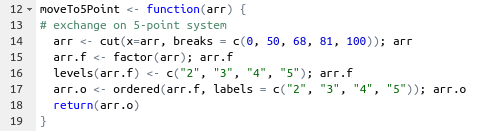


Для этого я использовал собственно написанные функции moveTo5Point() и moveToEUPoint() и функцию R table() для проверки корректности работы предыдущих функций.

Функция moveTo5Point(arr):

arr — выборка

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



Функция moveToEUPoint(arr):

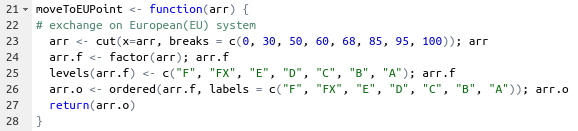
arr — выборка

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

Функция table(x):

x — набор значений

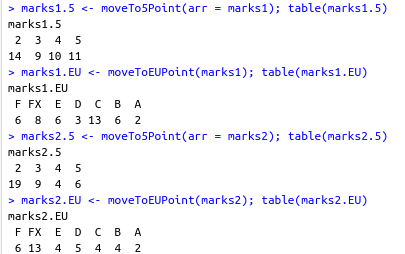
возвращает таблицу с частотами встречаемости каждого значения х.

Код:





Результат выполнения:



Задание 3:

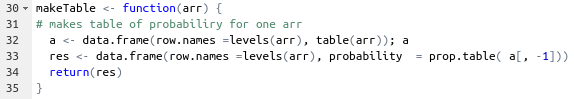


Для этого я использовал собственно написанные функции makeTable() и makeDoubleTable().

Функция makeTable(arr):

arr — выборка

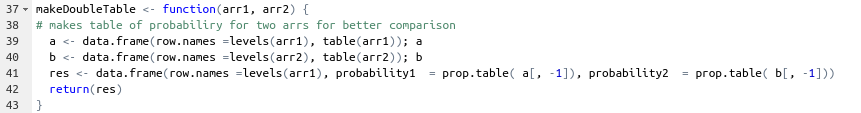
нужна для создания таблицы относительных частот в выборке arr.

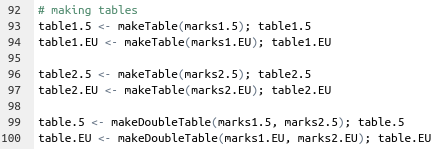


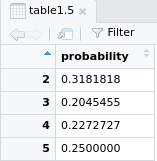
Функция makeTable(arr1, arr2):

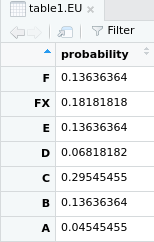
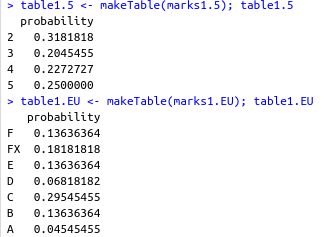
arr — выборка

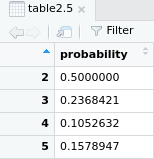
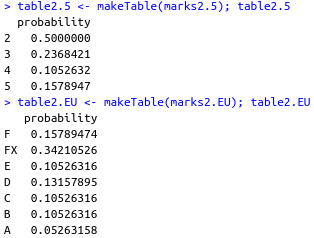
нужна для создания таблицы относительных частот в выбороках arr1 и arr2 для удобного их сравнения.

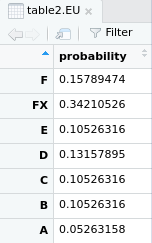
Код:

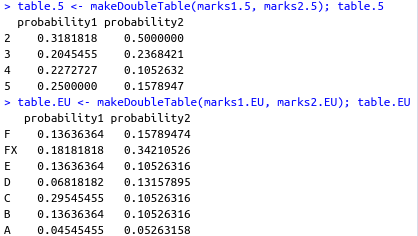


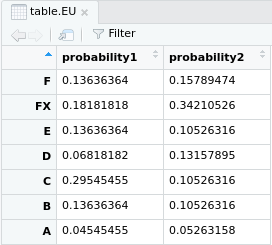
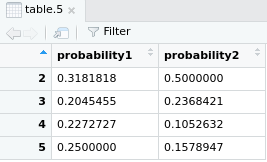
Результат выполнения:











Задание 4



Для этого я использовал собственно написанные функции makePlot5() и makePlotEU(), в которых также использовались такие функции, как par(), hist(), lines(), density().

Функция par():

нужна для того, чтобы разделить пространство, на котором будут графики, на некое определенное количество маленьких пространств, для того, чтобы нарисовать сразу несколько графиков на одном экране.

Функция hist(x, breaks = ):

x — переменная

breaks — кол-во столбцов

нужна для создания гистограмм частот значений переменной х; аргумент breaks = можно использовать, чтобы изменить принятое по умолчанию количество столбцов.

Функция lines(x, col = , lwb = ):

x — значения

col — цвет

lwd — толщина линии

нужна для создания линий на графиках.

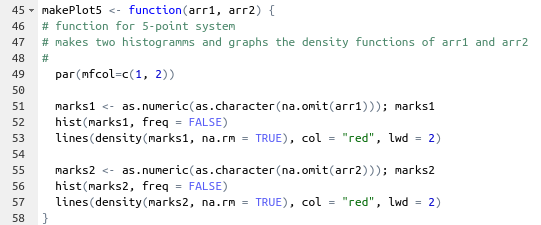
Функция density():

нужна для нахождения ядерных плотностей вероятностей (ядерная плотность вероятности — оценка случайной величины).

Функция makePlot5(arr1, arr2):

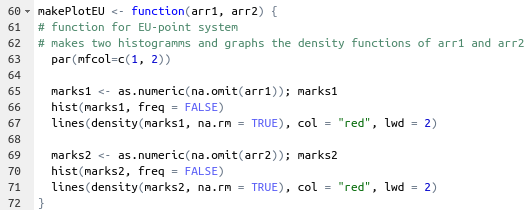
arr1, arr2 -выборки

нужна для создания гистограмм двух выборок в 5-балльной системе счисления.



Функция makePloEU(arr1, arr2):

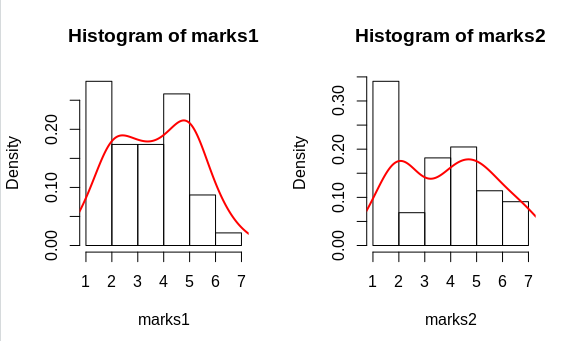
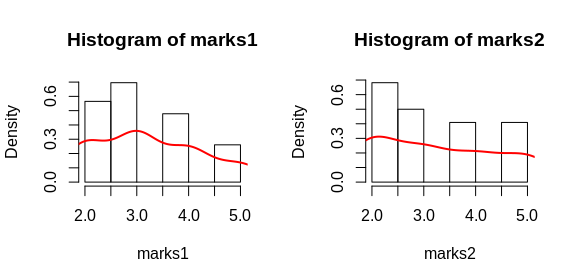
arr1, arr2 -выборки

 нужна для создания гистограмм двух выборок в 7-балльной системе счисления.

Код:



Результат выполнения:



Задание 5:



Для этого использовал функцию boxplot() и dev.off().

Функция dev.off():

нужна для очистки окна вывода.

Функция boxplot(x, main = , ylab = ):

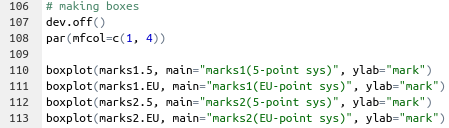
x — выборка

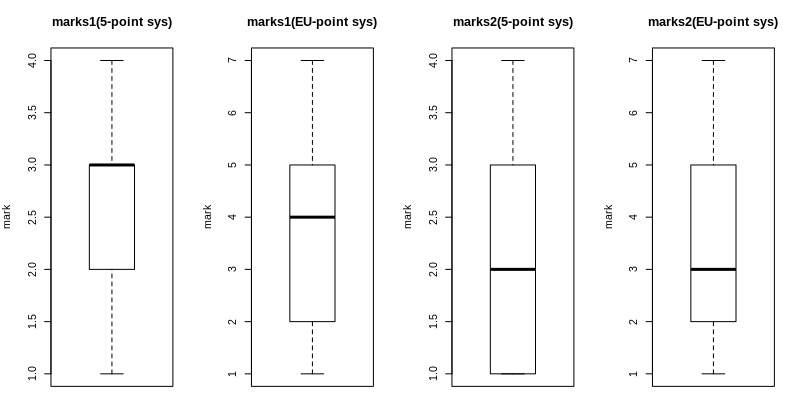
main — название графика

ylab — название оси OY

нужна для построения диаграмм размахов («коробок с усами»).

Код:



Результат выполнения:

Вывод:

Проведя данную лабораторную работу, я укрепил свои знания в теоретической части, а также получил опыт использования встроенных в R и написания собственных функций в Rstudio.

Конец.