

# Лабораторная работа №1

## Задание 1

Выберите распределение, у которого существуют первые четыре момента, и экспериментально убедитесь в асимптотической нормальности выборочного среднего, выборочной дисперсии, выборочной квантили порядка 0.5 для данного распределения. Также экспериментально убедитесь в том, что  $nF(X_{(2)}) \rightarrow U_1 \sim \Gamma(2, 1)$  и  $n(1 - F(X_{(n)})) \rightarrow U_2 \sim \Gamma(1, 1) = \text{Exp}(1)$ .

*Указание:* сгенерируйте достаточно большое количество выборок достаточно большого объема, для каждой сгенерированной выборки вычислите соответствующие статистики (функции от выборок), постройте гистограммы результатов для каждой статистики, для наглядности рядом с гистограммой можно нарисовать соответствующую плотность (пока это метод "на глаз", но в дальнейшем мы разберем статистическую процедуру, позволяющую проверить согласованность распределения выборки с заданным вероятностным законом), также можно помимо гистограммы вывести мат. ожидание, дисперсию (или стандартное отклонение) и медиану.

## Задание 2

Второе задание представлено в четырёх вариантах.

1. В файле *iris.csv* представлены данные о параметрах различных экземплярах цветка ириса. Какой вид в датасете представлен больше всего, какой – меньше? Рассчитайте выборочное среднее, выборочную дисперсию, выборочную медиану и выборочную квантиль порядка  $2/5$  для суммарной площади чашелистика и лепестка всей совокупности и отдельно для каждого вида. Построить график эмпирической функции распределения, гистограмму и box-plot суммарной площади чашелистика и лепестка для всей совокупности и каждого вида.
2. В файле *sex\_bmi\_smokers.csv* приведены данные (пол, ИМТ, курит/не курит) о более 1000 испытуемых. Сравните количество курящих мужчин и некурящих женщин. Рассчитайте выборочное среднее, выборочную дисперсию, выборочную медиану и выборочную квантиль порядка  $3/5$  ИМТ всех наблюдателей и отдельно для каждой возможной комбинации пол-курение. Построить график эмпирической функции распределения, гистограмму и box-plot ИМТ для всех наблюдателей и отдельно для каждой возможной комбинации пол-курение.
3. В файле *cars93.csv* представлены данные об автомобилях. Какие типы автомобилей представлены в датасете? Какой тип наиболее распространен, какой – менее? Рассчитайте выборочное среднее, выборочную дисперсию, выборочную медиану и межквартильный размах мощности для всей совокупности автомобилей и отдельно для каждого типа автомобиля. Построить график эмпирической функции распределения, гистограмму и box-plot мощности для всей совокупности и отдельно для каждого типа авто.
4. В файле *mobile\_phones.csv* приведены данные о мобильных телефонах. В сколько моделей можно вставить 2 сим-карты, сколько поддерживают 3-G, каково наибольшее число ядер у процессора? Рассчитайте выборочное среднее, выборочную дисперсию, выборочную медиану и выборочную квантиль порядка  $2/5$ , построить график эмпирической функции распределения, гистограмму и box-plot для емкости аккумулятора для всей совокупности и в отдельности для поддерживающих/не поддерживающих Wi-Fi.