**Курс «Теория вероятностей и статистика в Машинном Обучении»**

**Лабораторная работа №1:** Описательная статистика, визуализация, предварительная обработка данных

Часть 1

1. Найти на сайте РосСтат <https://rosstat.gov.ru/free_doc/new_site/zdor22/PublishSite_2022/index.html> данные, распределение которых было бы близко к

a) нормальному (визуально выглядит как кривая Гаусса - «колокол»),

б) равномерному.

1. Найти распределение со смещенной медианой относительно среднего (~15% размаха) и несмещенной.
2. Посчитать описательные характеристики распределения выбранных данных (среднее, мода, медиана), дать визуальное представление данных (условное форматирование Excel, построить график рассеивания/гистограмму/круговую диаграмму).

Часть 2

1. Найти датасет с аномальными значениями, которые можно выявить, построив диаграмму размаха «ящик с усами». Попробовать выдвинуть гипотезу о причинах аномальности, которую можно подтвердить дальнейшими исследованиями, используя синтетические признаки (или корреляционный анализ и т.п.).

<https://statanaliz.info/excel/diagrammy/diagramma-yashhik-s-usami-boxplot-v-excel-2016/>

<https://excel2.ru/articles/blochnaya-diagramma-v-ms-excel>