

Лабораторная работа №1

Задание 1

Первое задание представлено в четырёх вариантах. Сами варианты.

1. В файле [iris.csv](#)¹ 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 (более точно – оценки 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, проданных в некотором автосалоне за 93 год. 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 американских и не американских авто. 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.

Задание 2

Представлено в пяти вариантах. Везде требуется найти оценку указанным методом, смещение, дисперсию, среднеквадратическую ошибку (**теоретические**) и указать свойства оценок. Также провести эксперимент при указанных параметрах по следующей схеме:

1. Задайте массив объемов выборки
2. Для каждого объема выборки n сгенерируйте m выборок из вашего распределения и для каждой сгенерированной выборки посчитайте оценку параметра согласно полученной формуле
3. Обработайте результаты (посчитайте выборочные характеристики для разницы между оценкой и реальным параметром для каждого объема выборки, количество выборок, для которых оценка отличается от реального параметра более чем на заданный вами порог и т.п.), визуализируйте результат.

Сами варианты:

¹датасеты взяты с открытых источников, в частности с сайта для соревнований по Data Science и Machine Learning [Kaggle.com](https://www.kaggle.com)

1. Методом моментов найти оценку параметра θ равномерного распределения на $[-\theta, \theta]$. Эксперимент при $\theta = 10$. **Подсказка:** в зависимости от выбранной функции $g(x)$ можно оценить не θ , а θ^2 (и написать выкладки относительно оцениваемой функции θ^2).
2. Методом моментов найти оценку квадрата масштабирующего параметра θ распределения Лапласа (сдвиг считать нулевым). Эксперимент при $\theta = 0.5$. **Указание:** для плотности используйте параметризацию $f_\theta(x) = \frac{1}{2\theta} \exp\{-\frac{|x|}{\theta}\}$.
3. Методом максимального правдоподобия найти оценку параметра θ биномиального распределения $\text{Bin}(m, \theta)$, считая m известным. Эксперимент при $m = 4$, $\theta = 1/5$.
4. Найти оценку максимального правдоподобия параметра θ для распределения с плотностью

$$f_\theta(x) = \frac{3x^2}{\sqrt{2\pi}} \exp\left(-\frac{(\theta - x^3)^2}{2}\right).$$

Эксперимент при $\theta = 5$. **Подсказка:** здесь распределение не стандартное, так что имеет смысл генерировать величины по схеме $X = F^{-1}(Y)$, где $U \sim U[0, 1]$, F^{-1} – обратная к функции распределения (в данном случае она выражается через квантильную функцию стандартного нормального закона и арифметические операции).

5. С помощью метода моментов найти оценку параметра θ распределения с плотностью

$$f_\theta(x) = \frac{1}{(k-1)!\theta^k} x^{k-1} e^{-x/\theta} \mathbf{1}(x > 0),$$

если $k \in \mathbb{N}$ – известный параметр. Эксперимент при $\theta = 2$, $k = 3$.