

Санкт-Петербургский политехнический университет Петра Великого  
Институт прикладной математики и механики  
**Кафедра «Прикладная математика»**

**ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ  
ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ  
СТАТИСТИКА»**

Выполнил  
студент группы 3630102/70201

Крупкина Дарья

Проверил  
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург  
2020

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>2</b>
1.1	Задание . . . . .	2
<b>2</b>	<b>Теория</b>	<b>2</b>
2.1	Распределения . . . . .	2
2.2	Боксплот Тьюки . . . . .	3
2.2.1	Определение . . . . .	3
2.2.2	Описание . . . . .	3
2.2.3	Построение . . . . .	3
2.3	Теоретическая вероятность выбросов . . . . .	3
<b>3</b>	<b>Реализация</b>	<b>4</b>
<b>4</b>	<b>Результаты</b>	<b>4</b>
4.1	Боксплот Тьюки . . . . .	4
4.2	Доля выбросов . . . . .	7
4.3	Теоретическая доля выбросов . . . . .	7
<b>5</b>	<b>Обсуждение</b>	<b>7</b>
<b>6</b>	<b>Приложения</b>	<b>7</b>

# Список иллюстраций

1	Нормальное распределение . . . . .	4
2	Распределение Коши . . . . .	5
3	Распределение Лапласа . . . . .	5
4	Распределение Пуассона . . . . .	6
5	Равномерное распределение . . . . .	6

# Список таблиц

1	Доля выбросов . . . . .	7
2	Теоретическая вероятность выбросов . . . . .	7

# 1 Постановка задачи

Для 5 распределений:

1.  $N(x, 0, 1)$  – нормальное распределение
2.  $C(x, 0, 1)$  – распределение Коши
3.  $L(x, 0, \frac{1}{\sqrt{2}})$  – распределение Лапласа
4.  $P(k, 10)$  – распределение Пуассона
5.  $U(x, -\sqrt{3}, \sqrt{3})$  – равномерное распределение

## 1.1 Задание

Сгенерировать выборки размером 20 и 1000 элементов.

Построить для них боксплот Тьюки.

Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000 раз, и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически.

# 2 Теория

## 2.1 Распределения

1. Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (1)$$

2. Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad (2)$$

3. Распределение Лапласа

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (3)$$

4. Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (4)$$

5. Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \quad (5)$$

## 2.2 Боксплот Тьюки

### 2.2.1 Определение

Боксплот (англ. box plot) — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей[1].

### 2.2.2 Описание

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Несколько таких ящичков можно нарисовать бок о бок, чтобы визуально сравнивать одно распределение с другим; их можно располагать как горизонтально, так и вертикально. Расстояния между различными частями ящичка позволяют определить степень разброса (дисперсии) и асимметрии данных и выявить выбросы.

### 2.2.3 Построение

Границами ящичка служат первый и третий квартили, линия в середине ящичка — медиана. Концы усов — края статистически значимой выборки (без выбросов). Длину «усов» определяют разность первого квартиля и полутора межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид

$$X_1 = Q_1 + \frac{3}{2}(Q_3 - Q_1), X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \quad (6)$$

где  $X_1$  - нижняя граница уса,  $X_2$  - верхняя граница уса,  $Q_1$  - первый квартиль,  $Q_3$  - третий квартиль.

Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков.

## 2.3 Теоретическая вероятность выбросов

По формуле (6) можно вычислить теоретические нижнюю и верхнюю границы уса ( $X_1^T$ ,  $X_2^T$  соответственно). Выбросами считаются величины  $x$ , такие что:

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (7)$$

Теоретическая вероятность выбросов для непрерывных распределений

$$P^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)), \quad (8)$$

где  $F(X) = P(x \leq X)$  - функция распределения.

Теоретическая вероятность выбросов для дискретных распределений

$$P^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)), \quad (9)$$

где  $F(X) = P(x \leq X)$  - функция распределения.

### 3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. Используются библиотеки `numpy` для генерации выборки, `matplotlib` для построения боксплотов и `tabulate` для удобного представления табличных данных. Исходный код лабораторной работы приведён в приложении в виде ссылки на репозиторий GitHub.

## 4 Результаты

### 4.1 Боксплот Тьюки

Для каждого распределения представлен боксплот Тьюки.

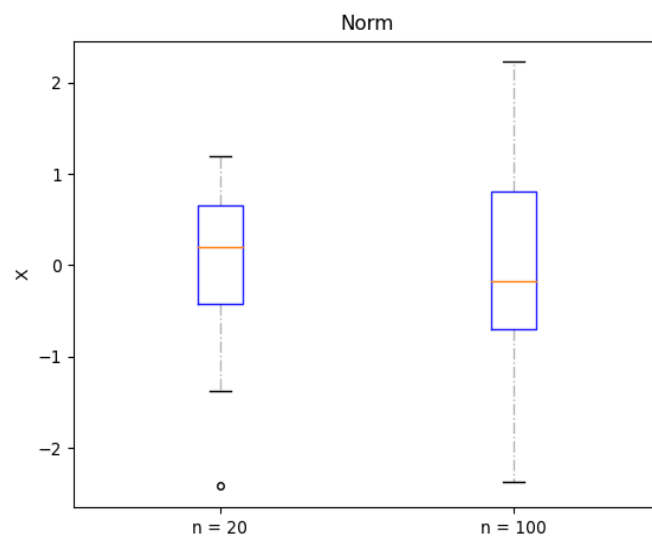


Рис. 1: Нормальное распределение

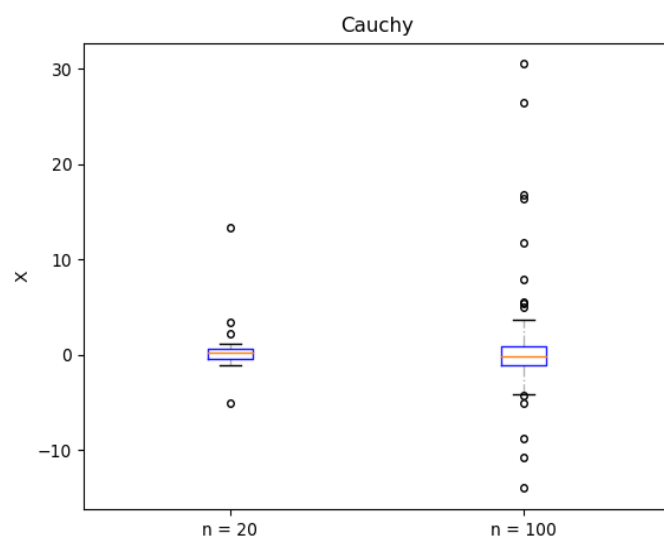


Рис. 2: Распределение Коши

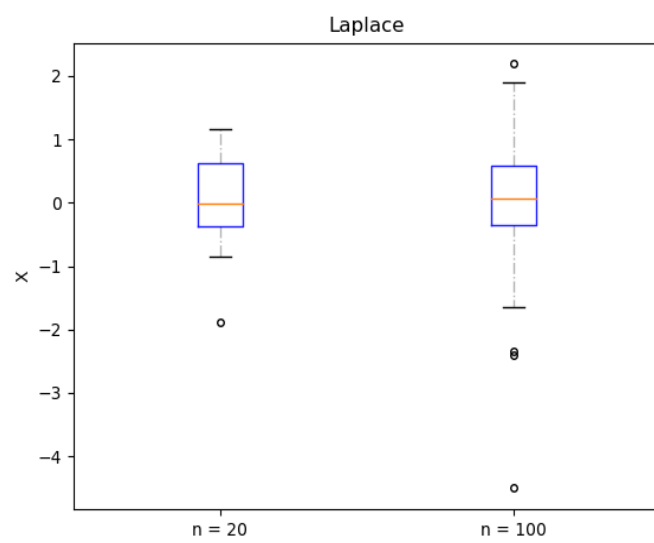


Рис. 3: Распределение Лапласа

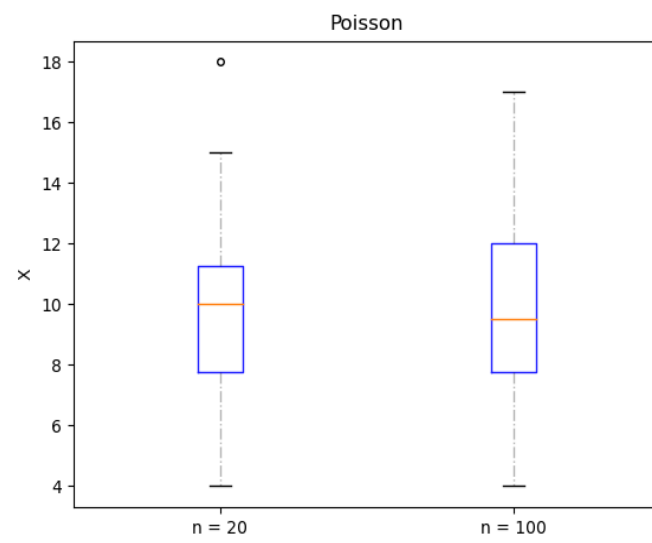


Рис. 4: Распределение Пуассона

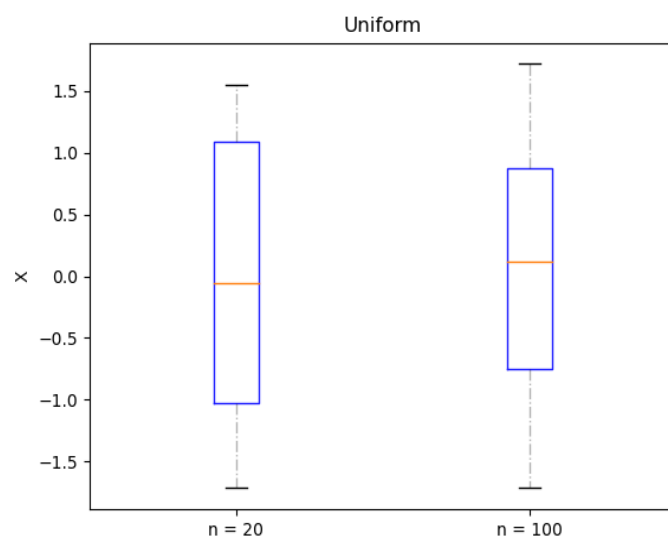


Рис. 5: Равномерное распределение

## 4.2 Доля выбросов

Выборка	Доля выбросов
Norm, n = 20	0.02
Norm, n = 100	0.01
Cauchy, n = 20	0.15
Cauchy, n = 100	0.16
Laplace, n = 20	0.07
Laplace, n = 100	0.06
Poisson, n = 20	0.02
Poisson, n = 100	0.01
Uniform, n = 20	0
Uniform, n = 100	0

Таблица 1: Доля выбросов

## 4.3 Теоретическая доля выбросов

Распределение	$Q_1^t$	$Q_2^T$	$X_1^T$	$X_2^T$	$P^T$
Нормальное распределение	-0.674	0.674	-2.698	2.698	0.007
Распределение Коши	-1	1	-4	4	0.156
Распределение Лапласа	-0.490	0.490	-1.961	1.961	0.063
Распределение Пуассона	8	12	2	18	0.008
Равномерное распределение	-0.866	0.866	-3.464	3.464	0

Таблица 2: Теоретическая вероятность выбросов

## 5 Обсуждение

По полученным результатам можно сделать вывод, что доля выбросов, полученная на практике, и теоретическая вероятность выбросов совпадают для равномерного распределения - выбросы в этом распределении не наблюдаются и вероятность соответственно равна 0. Доли выбросов для нормального распределения и распределения Пуассона оказались ниже теоретической вероятности, в то время как результаты для распределений Коши и Лапласа оказались близки к теоретическим.

При этом для всех распределений результаты, полученные для выборки из 100 элементов, ближе к теоретическим.

## 6 Приложения

Код программы на GitHub, URL: <https://github.com/DariaKrup/Statistics>



## Список литературы

- [1] Box plot. URL: [https://en.wikipedia.org/wiki/Box\\_plot](https://en.wikipedia.org/wiki/Box_plot)