Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №3 по дисциплине “Математическая статистика”

**Боксплот Тьюки**

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

Оглавление

[1 Постановка задачи 3](#_Toc37882723)

[2 Теория 3](#_Toc37882724)

[3 Реализация 4](#_Toc37882725)

[4 Результаты 4](#_Toc37882726)

[5 Выводы 7](#_Toc37882727)

[6 Литература 7](#_Toc37882728)

[7 Приложения 7](#_Toc37882729)

Список иллюстраций и таблиц

[Cтандартное нормальное распределение](#стандартное_нормальное_распределение)…………………………………………………………………………………………..4

[Стандартное распределение Коши](#стандартное_распределение_коши)………………………………………………………………………………………………………5

[Распределение Лапласа](#распределение_лапласа)……………………………………………………………………………………………………………………….5

[Распределение Пуассона](#распределение_пуассона)………………………………………………………………………………………………………………………6

[Равномерное распределение](#равномерное_распределение)……………………………………………………………………………………………………………….6

[Таблица. Средние проценты выбросов](#средние_проценты_выбросов)………………………………………………………………………………………………..7

# 1 Постановка задачи

Сгенерировать выборки размером 20 и 100 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.

Распределения:

* Стандартное нормальное распределение:

|  |  |
| --- | --- |
|  | (1.1) |

* Стандартное распределение Коши:

|  |  |
| --- | --- |
|  | (1.2) |

* Распределение Лапласа:

|  |  |
| --- | --- |
|  | (1.3) |

* Распределение Пуассона:

|  |  |
| --- | --- |
|  | (1.4) |

* Равномерное распределение:

|  |  |
| --- | --- |
|  | (1.5) |

# **2 Теория**

Боксплот Тьюки – график, использующийся в описательной статистике, изображающий одномерное распределение вероятностей.

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

Выброс – результат, сильно выделяющийся из общей выборки.

Характеристики положения:

* Выборочное среднее:

|  |  |
| --- | --- |
|  | (2.1) |

* Выборочная медиана:

|  |  |
| --- | --- |
|  | (2.2) |

* Квартиль:

|  |  |
| --- | --- |
|  | (2.3) |

# 3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для генерации выборок с различными распределениями и математических расчётов, модуль *matplotlib* для построения и отображения боксплотов, модуль *pandas* для оптимального хранения статистических данных и функция *display* из модуля *IPython.display* для их корректного отображения в таблицах.

# 4 Результаты

Рис. 1 Стандартное нормальное распределение

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 2 Стандартное распределение Коши

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 3 Распределение Лапласа

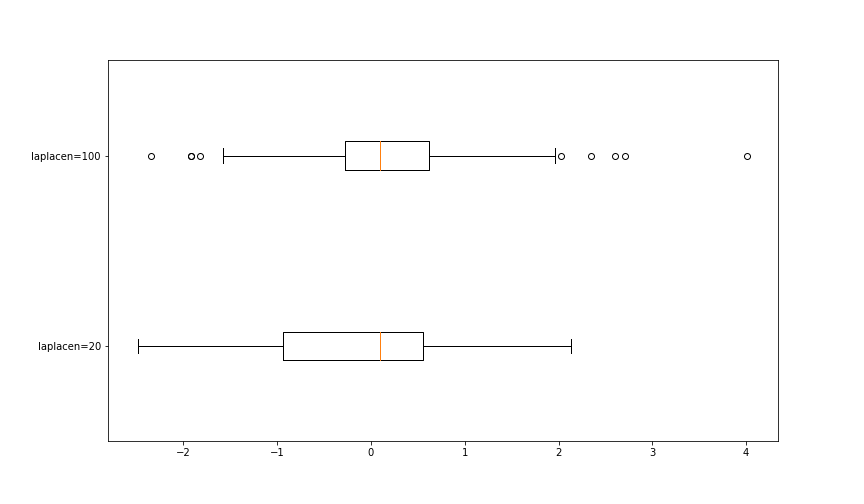


Рис. 4 Распределение Пуассона

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рис. 5 Равномерное распределение

Изображение выглядит как снимок экрана

Автоматически созданное описание

Таблица 1 Средние проценты выбросов

|  |  |
| --- | --- |
| Выборка | Средний процент выбросов |
| Стандартное нормальное распределение [(1.1)](#формула_1) |  |
| N = 20 | 2 |
| N = 100 | 1 |
| Стандартное распределение Коши [(1.2)](#формула_2) |  |
| N = 20 | 15 |
| N = 100 | 16 |
| Распределение Лапласа [(1.3)](#формула_3) |  |
| N = 20 | 8 |
| N = 100 | 6 |
| Распределение Пуассона [(1.4)](#формула_4) |  |
| N = 20 | 1 |
| N = 100 | 0 |
| Равномерное распределение [(1.5)](#формула_5) |  |
| N = 20 | 0 |
| N = 100 | 0 |

# 5 Выводы

Экспериментально полученные проценты выбросов, близки к теоретическим. Можно вывести соотношение между процентами выбросов у конкретных распределений:

(1.5) < (1.4) < (1.1) < (1.3) < (1.2)

По полученным данным видно, что наименьший процент выбросов у равномерного распределения, а наибольший у стандартного распределения Коши.

# 6 Литература

[Основы работы с *numpy* (отдельная глава курса)](https://stepik.org/course/401)

[Pandas обзор](https://www.dataquest.io/blog/pandas-cheat-sheet/)

# 7 Приложения

[Код лабораторной](https://github.com/MeShootIn/matstat/blob/master/lab_3/lab_3.ipynb)