Санкт-Петербургский Политехнический Университет
им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №2 по дисциплине “Математическая статистика”

**Характеристики положения выборки**

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

# Оглавление



# Список таблиц

## 1 Постановка задачи

Любыми средствами сгенерировать выборки с мощностями 10, 100 и 1000 элементов для 5 распределений. Для каждой выборки вычислить следующие характеристики положения:

$avrg\,(выборочное\ среднее), med\,x, Z_R, Z_Q, Z_{tr}, при\ r \approx \frac{n}{4}$. Построить по ним таблицы.

Распределения:

- Стандартное нормальное распределение:

$$N(x,0,1)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \tag{1.1}$$

- Стандартное распределение Коши:

$$C(x,0,1)=\frac{1}{\pi(1+x^2)} \tag{1.2}$$

- Распределение Лапласа:

$$L\left(x,0,\frac{1}{\sqrt{2}}\right)=\frac{1}{\sqrt{2}}e^{-\sqrt{2}|x|} \tag{1.3}$$

- Распределение Пуассона:

$$P(k,10)=\frac{10^k}{k!}e^{-10} \tag{1.4}$$

- Равномерное распределение:

$$U(x,-\sqrt{3},\sqrt{3})=\begin{cases}\frac{1}{2\sqrt{3}}, при\ |x|\le\sqrt{3}\\ 0, при\ |x|>\sqrt{3}\end{cases} \tag{1.5}$$

## 2 Теория

Характеристики положения:

- Выборочное среднее:

$$avrg=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{2.1}$$

- Выборочная медиана:

$$med\, x = \begin{cases} x_{k+1}, n=2k+1 \\ \frac{1}{2}(x_k + x_{k+1}), n=2k \end{cases} \qquad (2.2)$$

- Полусумма экстремальных значений:

$$Z_R = \frac{1}{2}(x_1 + x_n) \qquad (2.3)$$

- Полусумма квартилей:

$$Z_Q = \frac{1}{2}\left(Z_{\frac{1}{4}} + Z_{\frac{3}{4}}\right) \qquad (2.4)$$

- Усечённое среднее:

$$Z_{tr} = \frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_i \qquad (2.5)$$

# 3 Реализация

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

После вычисления характеристик положения 1000 раз, для каждой характеристики находятся их средние значения и дисперсии:

$$E(z) = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad (3.1)$$

$$D(z) = E(z^2) - E^2(z) \qquad (3.2)$$

# 4 Результаты

*Таблица 1 Стандартное нормальное распределение*

| n=10 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
|---|---|---|---|---|---|
| $E(z)$ | 0.01 | 0.01 | 0.02 | 0 | 0.28 |
| $D(z)$ | 0.097848 | 0.12874 | 0.190887 | 0.11015 | 0.109265 |
| n=100 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 0 | 0 | 0 | -0 | 0.03 |
| $D(z)$ | 0.009964 | 0.015673 | 0.093959 | 0.012457 | 0.011819 |
| n=1000 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 0 | 0 | -0 | 0 | 0 |
| $D(z)$ | 0.00094 | 0.001552 | 0.062108 | 0.00116 | 0.001157 |

*Таблица 2 Стандартное распределение Коши*

| n=10 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
|---|---|---|---|---|---|
| $E(z)$ | -12.31 | 0 | -61.48 | -0.02 | 0.65 |
| $D(z)$ | 161834.047069 | 0.276286 | 4045976.36516 | 0.750808 | 0.918042 |
| n=100 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | -0.52 | -0 | -24.32 | -0.01 | 0.03 |
| $D(z)$ | 875.221257 | 0.02564 | 2090211.365057 | 0.05555 | 0.028177 |
| n=1000 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | -0.4 | 0 | -198.59 | -0 | 0 |
| $D(z)$ | 520.951391 | 0.002476 | 128259873.764481 | 0.004587 | 0.002515 |

*Таблица 3 Распределение Лапласа*

| n=10 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
|---|---|---|---|---|---|
| $E(z)$ | 0 | 0.01 | 0 | 0 | 0.23 |
| $D(z)$ | 0.098106 | 0.066493 | 0.405711 | 0.085741 | 0.078464 |
| n=100 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 0 | 0 | -0 | 0 | 0.02 |
| $D(z)$ | 0.010295 | 0.005968 | 0.419241 | 0.010031 | 0.006175 |
| n=1000 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 0 | 0 | 0.02 | 0 | 0 |
| $D(z)$ | 0.00102 | 0.000534 | 0.404961 | 0.00103 | 0.000639 |

*Таблица 4 Распределение Пуассона*

| n=10 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
|---|---|---|---|---|---|
| $E(z)$ | 9.98 | 9.82 | 10.29 | 9.89 | 10.74 |
| $D(z)$ | 0.939362 | 1.378308 | 1.653056 | 1.114103 | 1.202277 |
| n=100 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 10.01 | 9.85 | 11 | 9.91 | 9.95 |
| $D(z)$ | 0.098454 | 0.202596 | 0.987738 | 0.156002 | 0.121359 |
| n=1000 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 10 | 10 | 11.66 | 9.99 | 9.87 |
| $D(z)$ | 0.010283 | 0.002991 | 0.689351 | 0.004562 | 0.010918 |

*Таблица 5 Равномерное распределение*

| n=10 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
|---|---|---|---|---|---|
| $E(z)$ | -0.03 | -0.03 | -0.01 | -0.03 | 0.29 |
| $D(z)$ | 0.09994 | 0.228635 | 0.048268 | 0.138834 | 0.156282 |
| n=100 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |
| $E(z)$ | 0 | 0 | 0 | 0 | 0.04 |
| $D(z)$ | 0.009809 | 0.028060 | 0.000538 | 0.014772 | 0.019582 |
| n=1000 | *avrg* | *med x* | $Z_R$ | $Z_Q$ | $Z_{tr}$ |

| | | | | | |
|---|---|---|---|---|---|
| $E(z)$ | 0 | -0 | -0 | 0 | 0 |
| $D(z)$ | 0.00096 | 0.00287 | 0.000006 | 0.001504 | 0.001887 |

## 5 Обсуждение

При вычислении средних значений пришлось отбрасывать некоторое число знаков после запятой, так как получаемая дисперсия не могла гарантировать получаемое точное значение.

Иными словами, дисперсия может гарантировать порядок точности среднего значения только до первого значащего знака после запятой в дисперсии включительно.

Единственным исключением (в отбрасывании знаков после запятой) стало стандартное распределение Коши, так как оно имеет бесконечную дисперсию, а значит может гарантировать сколь угодно большую точность.

## 6 Выводы

В процессе работы вычислены значения характеристик положения для каждого из 5 распределений на выборках фиксированных мощностей и получены следующее ранжирование характеристик положения:

1. Стандартное нормальное распределение:
   $Z_R < Z_{tr} < med\,x < avrg < Z_Q$
2. Стандартное распределение Коши:
   $Z_R < avrg < Z_Q < med\,x < Z_{tr}$
3. Распределение Лапласа:
   $med\,x < avrg < Z_Q < Z_{tr} < Z_R$
4. Распределение Пуассона:
   $Z_Q < med\,x < Z_{tr} < avrg < Z_R$
5. Равномерное распределение:
   $med\,x < Z_R < Z_Q < avrg < Z_{tr}$

## 7 Литература


Основы работы с *numpy* (отдельная глава курса)

Pandas обзор


## 8 Приложения

Код лабораторной