Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №4 по дисциплине “Математическая статистика”

**Выборочные коэффициенты корреляции и эллипсы рассеивания**

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 4](#_Toc42472627)

[2 Теория 4](#_Toc42472628)

[3 Реализация 4](#_Toc42472629)

[4 Результаты 4](#_Toc42472630)

[5 Выводы 15](#_Toc42472631)

[6 Литература 16](#_Toc42472632)

[7 Приложения 16](#_Toc42472633)

Список иллюстраций и таблиц

[Таблица 1 Двумерное стандартное нормальное распределение, n=20, r=0](#норм_20_0) ……………………………………4

[Таблица 2 Двумерное стандартное нормальное распределение, n=60, r=0](#норм_60_0)……………………………………5

[Таблица 3 Двумерное стандартное нормальное распределение, n=100, r=0](#норм_100_0)……………………………….…5

[Таблица 4 Двумерное стандартное нормальное распределение, n=20, r=0.5](#норм_20_05)…………………………………5

[Таблица 5 Двумерное стандартное нормальное распределение, n=60, r=0.5](#норм_60_05)…………………………………5

[Таблица 6 Двумерное стандартное нормальное распределение, n=100, r=0.5](#норм_100_05)……………………………….5

[Таблица 7 Двумерное стандартное нормальное распределение, n=20, r=0.9](#норм_20_09)…………………………………5

[Таблица 8 Двумерное стандартное нормальное распределение, n=60, r=0.9](#норм_60_09)…………………………………5

[Таблица 9 Двумерное стандартное нормальное распределение, n=100, r=0.9](#норм_100_09)……………………………….5

[Таблица 10 Смесь распределений, n=20](#смесь_20)……………………………………………………………………………………………….6

[Таблица 11 Смесь распределений, n=60](#смесь_60)……………………………………………………………………………………………….6

[Таблица 12 Смесь распределений, n=100](#смесь_100)…………………………………………………………………………………………….6

[*Рис. 1 Двумерное стандартное нормальное распределение для n=20, r=0*](#эллипс_20_0)*……………………………………..7*

[*Рис. 2 Двумерное стандартное нормальное распределение для n=60, r=0*](#норм_60_0)*……………………………………..7*

[*Рис. 3 Двумерное стандартное нормальное распределение для n=100, r=0*](#норм_100_0)*……………………………………8*

[*Рис. 4 Двумерное стандартное нормальное распределение для n=20, r=0.5*](#норм_20_05)*…………………………………..8*

[*Рис. 5 Двумерное стандартное нормальное распределение для n=60, r=0.5*](#норм_60_05)*…………………………………..9*

[*Рис. 6 Двумерное стандартное нормальное распределение для n=100, r=0.5*](#норм_100_05)*…………………………………9*

[*Рис. 7 Двумерное стандартное нормальное распределение для n=20, r=0.9*](#эллипс_20_09)*…………………………………10*

[*Рис. 8 Двумерное стандартное нормальное распределение для n=60, r=0.9*](#эллипс_60_09)*…………………………………10*

[*Рис. 9 Двумерное стандартное нормальное распределение для n=100, r=0.9*](#эллипс_100_09)*………………………………11*

[*Рис. 10 Смесь распределений для n=20*](#эллипс_смесь_20)*………………………………………………………………………………………………11*

[*Рис. 11 Смесь распределений для n=60*](#эллипс_смесь_60)*………………………………………………………………………………………………12*

[*Рис. 12 Смесь распределений для n=100*](#эллипс_смесь_100)*…………………………………………………………………………………………….12*

[*Рис. 13 Эллипс рассеивания для 2-х точек при r=0*](#эллипс_2_0)*……………………………………………………………………………13*

[*Рис. 14 Эллипс рассеивания для 2-х точек при r=0.5*](#эллипс_2_05)*…………………………………………………………………………13*

[*Рис. 15 Эллипс рассеивания для 2-х точек при r=0.9*](#эллипс_2_09)*…………………………………………………………………………14*

[*Рис. 16 Эллипс рассеивания для 3-х точек при r=0*](#эллипс_3_0)*……………………………………………………………………………14*

[*Рис. 17 Эллипс рассеивания для 3-х точек при r=0.5*](#эллипс_3_05)*…………………………………………………………………………15*

[*Рис. 18 Эллипс рассеивания для 3-х точек при r=0.9*](#эллипс_3_09)*…………………………………………………………………………15*

# 1 Постановка задачи

Сгенерировать двумерные выборки размерами 20, 60, 100 для нормального двумерного распределения . Коэффициент корреляции взять равным 0, 0.5, 0.9. Каждая выборка генерируется 1000 раз и для неё вычисляются: среднее значение, среднее значение квадрата и дисперсия коэффициентов корреляции Пирсона, Спирмена и квадрантного коэффициента корреляции. Повторить все вычисления для смеси нормальных распределений:

Изобразить сгенерированные точки на плоскости и нарисовать эллипс рассеяния.

# **2 Теория**

1. Двумерное стандартное нормальное распределение:
2. Коэффициент корреляции Пирсона:
3. Коэффициент корреляции Спирмена:
4. Квадрантный коэффициент корреляции:

# 3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для вычисления описательных статистик, модуль *scipy* для генерации выборок на основе двумерного нормального распределения и расчёта коэффициентов корреляции, модуль *matplotlib* для построения и отображения графиков, модуль *pandas* для хранения статистических данных в таблицах и функция *display* из модуля *IPython.display* для их корректного отображения.

# 4 Результаты

Таблица 1 Двумерное стандартное нормальное распределение, n=20, r=0

|  |  |  |  |
| --- | --- | --- | --- |
| **n=20** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.012 | 0.011 | 0.003 |
| **E(z^2)** | 0.048 | 0.05 | 0.051 |
| **D(z)** | 0.048 | 0.049 | 0.051 |

Таблица 2 Двумерное стандартное нормальное распределение, n=60, r=0

|  |  |  |  |
| --- | --- | --- | --- |
| **n=60** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.004 | 0.004 | 0.003 |
| **E(z^2)** | 0.016 | 0.017 | 0.017 |
| **D(z)** | 0.016 | 0.017 | 0.017 |

Таблица 3 Двумерное стандартное нормальное распределение, n=100, r=0

|  |  |  |  |
| --- | --- | --- | --- |
| **n=100** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.001 | 0.001 | -0.001 |
| **E(z^2)** | 0.01 | 0.01 | 0.011 |
| **D(z)** | 0.01 | 0.01 | 0.011 |

Таблица 4 Двумерное стандартное нормальное распределение, n=20, r=0.5

|  |  |  |  |
| --- | --- | --- | --- |
| **n=20** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.489 | 0.46 | 0.322 |
| **E(z^2)** | 0.271 | 0.246 | 0.148 |
| **D(z)** | 0.032 | 0.035 | 0.044 |

Таблица 5 Двумерное стандартное нормальное распределение, n=60, r=0.5

|  |  |  |  |
| --- | --- | --- | --- |
| **n=60** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.496 | 0.476 | 0.331 |
| **E(z^2)** | 0.255 | 0.237 | 0.124 |
| **D(z)** | 0.009 | 0.01 | 0.014 |

Таблица 6 Двумерное стандартное нормальное распределение, n=100, r=0.5

|  |  |  |  |
| --- | --- | --- | --- |
| **n=100** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.497 | 0.477 | 0.331 |
| **E(z^2)** | 0.253 | 0.233 | 0.118 |
| **D(z)** | 0.005 | 0.006 | 0.009 |

Таблица 7 Двумерное стандартное нормальное распределение, n=20, r=0.9

|  |  |  |  |
| --- | --- | --- | --- |
| **n=20** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.896 | 0.867 | 0.696 |
| **E(z^2)** | 0.806 | 0.756 | 0.513 |
| **D(z)** | 0.002 | 0.004 | 0.029 |

Таблица 8 Двумерное стандартное нормальное распределение, n=60, r=0.9

|  |  |  |  |
| --- | --- | --- | --- |
| **n=60** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.898 | 0.883 | 0.707 |
| **E(z^2)** | 0.808 | 0.78 | 0.508 |
| **D(z)** | 0.001 | 0.001 | 0.009 |

Таблица 9 Двумерное стандартное нормальное распределение, n=100, r=0.9

|  |  |  |  |
| --- | --- | --- | --- |
| **n=100** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | 0.899 | 0.886 | 0.708 |
| **E(z^2)** | 0.809 | 0.786 | 0.507 |
| **D(z)** | 0 | 0.001 | 0.005 |

Таблица 10 Смесь распределений, n=20

|  |  |  |  |
| --- | --- | --- | --- |
| **n=20** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | -0.08 | -0.078 | -0.05 |
| **E(z^2)** | 0.061 | 0.061 | 0.056 |
| **D(z)** | 0.054 | 0.055 | 0.054 |

Таблица 11 Смесь распределений, n=60

|  |  |  |  |
| --- | --- | --- | --- |
| **n=60** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | -0.092 | -0.086 | -0.06 |
| **E(z^2)** | 0.025 | 0.024 | 0.021 |
| **D(z)** | 0.016 | 0.016 | 0.017 |

Таблица 12 Смесь распределений, n=100

|  |  |  |  |
| --- | --- | --- | --- |
| **n=100** | **Pearson** | **Spearman** | **quadrant** |
| **E(z)** | -0.097 | -0.092 | -0.063 |
| **E(z^2)** | 0.019 | 0.019 | 0.014 |
| **D(z)** | 0.01 | 0.01 | 0.01 |

Рис. 1 Двумерное стандартное нормальное распределение для n=20, r=0Изображение выглядит как внутренний, текст, карта, фотография

Автоматически созданное описание

Рис. 2 Двумерное стандартное нормальное распределение для n=60, r=0Изображение выглядит как фотография, стол, компьютер, сидит

Автоматически созданное описание

Рис. 3 Двумерное стандартное нормальное распределение для n=100, r=0Изображение выглядит как фотография, компьютер, стол, белый

Автоматически созданное описание

Рис. 4 Двумерное стандартное нормальное распределение для n=20, r=0.5Изображение выглядит как карта, текст

Автоматически созданное описание

Рис. 5 Двумерное стандартное нормальное распределение для n=60, r=0.5Изображение выглядит как фотография, стол, полный, белый

Автоматически созданное описание

Рис. 6 Двумерное стандартное нормальное распределение для n=100, r=0.5

Изображение выглядит как фотография, стол, другой, компьютер

Автоматически созданное описание

Рис. 7 Двумерное стандартное нормальное распределение для n=20, r=0.9Изображение выглядит как фотография, мужчина, другой, стол

Автоматически созданное описание

Рис. 8 Двумерное стандартное нормальное распределение для n=60, r=0.9Изображение выглядит как фотография, стол, мужчина, снег

Автоматически созданное описание

Рис. 9 Двумерное стандартное нормальное распределение для n=100, r=0.9Изображение выглядит как фотография, мужчина, стол, снег

Автоматически созданное описание

Рис. 10 Смесь распределений для n=20Изображение выглядит как карта

Автоматически созданное описание

Рис. 11 Смесь распределений для n=60Изображение выглядит как фотография, стол, белый, компьютер

Автоматически созданное описание

Рис. 12 Смесь распределений для n=100

Изображение выглядит как внутренний, фотография, компьютер, стол

Автоматически созданное описание

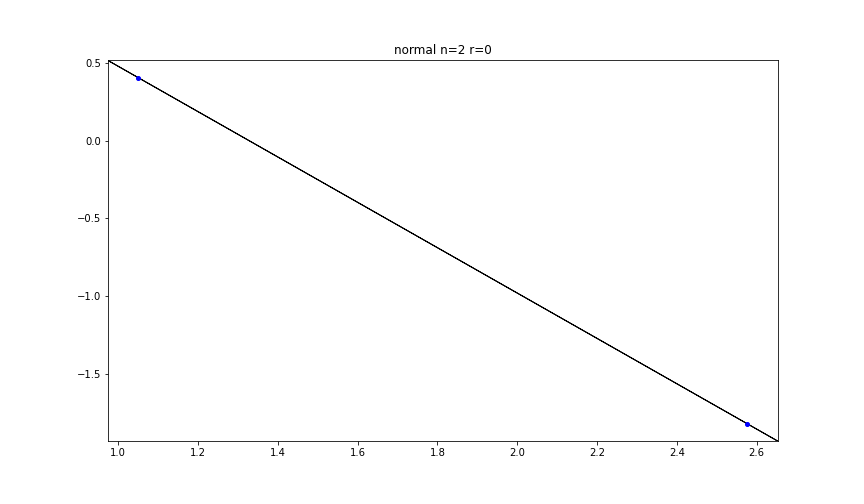
Рис. 13 Эллипс рассеивания для 2-х точек при r=0

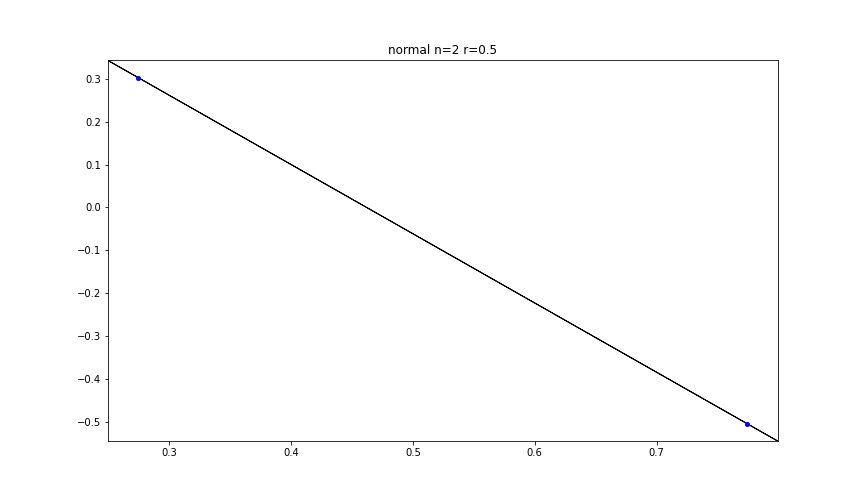
Рис. 14 Эллипс рассеивания для 2-х точек при r=0.5

Рис. 15 Эллипс рассеивания для 2-х точек при r=0.9Изображение выглядит как снимок экрана

Автоматически созданное описание

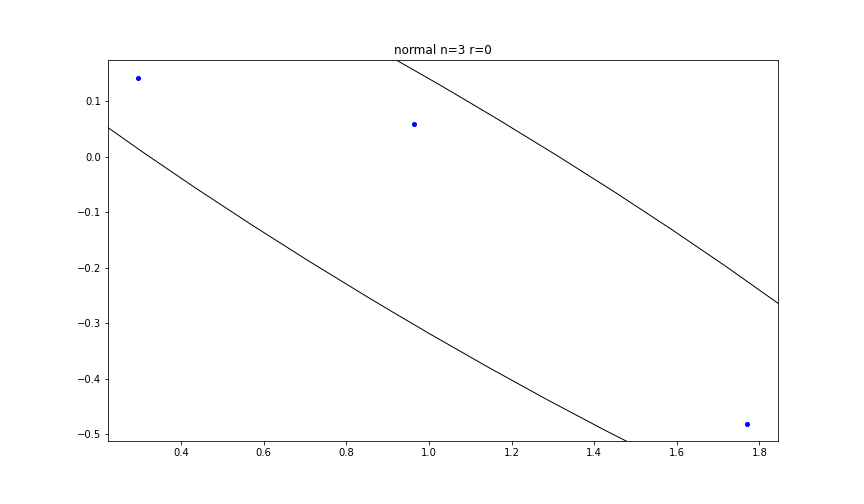
Рис. 16 Эллипс рассеивания для 3-х точек при r=0

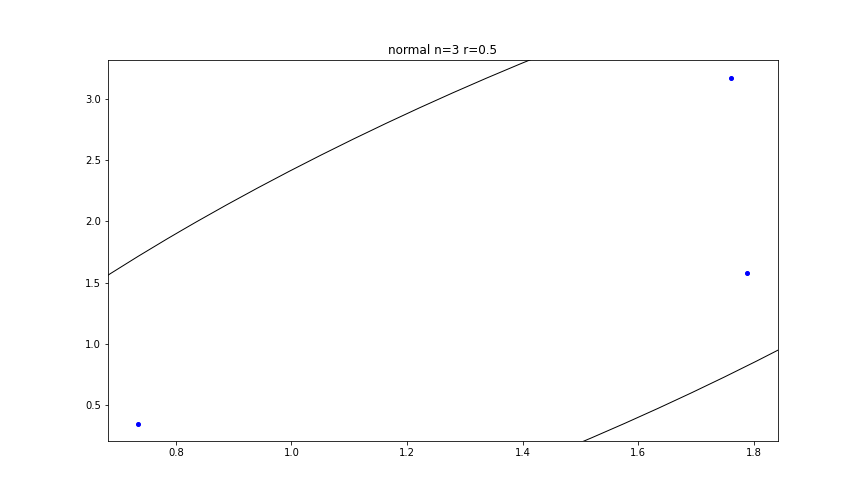
Рис. 17 Эллипс рассеивания для 3-х точек при r=0.5

Рис. 18 Эллипс рассеивания для 3-х точек при r=0.9

Изображение выглядит как снимок экрана

Автоматически созданное описание

# 5 Выводы

Ближе всего к теоретическому коэффициенту корреляции находится коэффициент Пирсона.

По графикам видно, что

* при увеличении объёма выборки коэффициенты корреляции стремятся к теоретическим
* при уменьшении корреляции эллипс рассеивания стремится к окружности, а при увеличении – вырождается в прямую с углом наклона в против часовой стрелки
* для построения эллипса рассеивания нужно минимум 3 точки, а при 2-х точках эллипс вырождается в прямую под определённым углом

# 6 Литература

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

[Документация по *scipy*](https://docs.scipy.org/doc/scipy/reference/index.html)

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

# 7 Приложения

[Код лабораторной](https://github.com/MeShootIn/matstat/blob/master/lab_5/lab_5.ipynb)