

Санкт-Петербургский Политехнический Университет Петра Великого

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

**Отчет по курсовой работе по математической
статистике**

Выполнил:
Группа:
Преподаватель:

Корьев Максим
5030102/10101
Баженов Александр Николаевич

Санкт-Петербург
2024 год

Содержание

Список иллюстраций	2
1 Введение	3
2 Теория	3
2.1 Двумерное нормальное распределение	3
2.2 Корреляционный момент и коэффициент корреляции	3
2.3 Выборочные коэффициенты корреляции	3
2.3.1 Выборочный коэффициент корреляции Пирсона	4
2.3.2 Выборочный квадрантный коэффициент корреляции	4
2.3.3 Выборочный коэффициент ранговой корреляции Спирмена	4
2.4 Эллипсы рассеивания	5
3 Методология	5
3.1 Описание данных	5
3.2 Аналитические методы	5
4 Результаты	5
4.1 Диаграммы рассеивания с эллипсами	5
4.2 Коэффициенты корреляции	7
5 Выводы	7
6 Реализация	8

Список иллюстраций

1	0.05Vsp953.dat	6
2	0.15Vsp93.dat	6
3	-0.15Vsp831.dat	6

1 Введение

Данный отчет посвящен анализу характеристик оценок при изменении данных или методов оценивания. Основное внимание уделяется эллипсам рассеивания и коэффициентам корреляции.

2 Теория

2.1 Двумерное нормальное распределение

Двумерная случайная величина (X, Y) называется распределённой нормально (или просто нормальной), если её плотность вероятности определена формулой:

$$N(x, y, \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\} \quad (1)$$

где:

- μ_X, μ_Y — математические ожидания X и Y ,
- σ_X, σ_Y — среднеквадратичные отклонения X и Y ,
- ρ — коэффициент корреляции X и Y .

2.2 Корреляционный момент и коэффициент корреляции

Ковариация K между X и Y определяется как:

$$K = \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \quad (2)$$

Коэффициент корреляции ρ двух случайных величин X и Y определяется как:

$$\rho = \frac{K}{\sigma_X\sigma_Y} \quad (3)$$

2.3 Выборочные коэффициенты корреляции

Коэффициенты корреляции используются для оценки степени линейной зависимости между двумя переменными. Существует несколько типов выборочных коэффициентов корреляции.

2.3.1 Выборочный коэффициент корреляции Пирсона

Выборочный коэффициент корреляции Пирсона r определяется как:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2 \cdot \frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{K}{s_X s_Y} \quad (4)$$

где:

- x_i, y_i — наблюдаемые значения X и Y ,
- \bar{x}, \bar{y} — выборочные средние X и Y ,
- s_X, s_Y — выборочные среднеквадратичные отклонения X и Y .

Коэффициент корреляции Пирсона позволяет определить степень линейной связи между двумя переменными. Его значения лежат в диапазоне от -1 до 1. Значение $r = 1$ указывает на идеальную положительную линейную связь, $r = -1$ — на идеальную отрицательную линейную связь, а $r = 0$ означает отсутствие линейной связи.

2.3.2 Выборочный квадрантный коэффициент корреляции

Квадрантный коэффициент корреляции r_Q основывается на медианных значениях переменных:

$$r_Q = \frac{(n_1 + n_3) - (n_2 + n_4)}{n} \quad (5)$$

где:

- n_1, n_2, n_3, n_4 — количество точек в каждом квадранте, определенном медианами X и Y ,
- n — общее количество наблюдений.

Квадрантный коэффициент корреляции является более устойчивым к выбросам, чем коэффициент корреляции Пирсона, так как основывается на медианах. Он полезен при анализе данных с ненормальным распределением или при наличии выбросов.

2.3.3 Выборочный коэффициент ранговой корреляции Спирмена

Коэффициент ранговой корреляции Спирмена r_S определяется на основе рангов наблюдений:

$$r_S = \frac{\frac{1}{n} \sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\frac{1}{n} \sum (u_i - \bar{u})^2 \cdot \frac{1}{n} \sum (v_i - \bar{v})^2}} \quad (6)$$

где:

- u_i, v_i — ранги значений X и Y ,
- \bar{u}, \bar{v} — средние ранги.

Коэффициент Спирмена используется для измерения монотонной связи между двумя переменными. Он не требует нормального распределения данных и устойчив к выбросам.

2.4 Эллипсы рассеивания

Эллипсы рассеивания используются для визуального представления ковариации между двумя переменными. Уравнение эллипса на плоскости xOy выглядит следующим образом:

$$\frac{(x - \bar{x})^2}{\sigma_x^2} - 2\rho \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} = \text{const} \quad (7)$$

где:

- \bar{x}, \bar{y} — средние значения переменных X и Y ,
- σ_x, σ_y — стандартные отклонения переменных X и Y ,
- ρ — коэффициент корреляции.

Эллипсы с равной вероятностью покрывают 95% наблюдений при нормальном распределении и дают наглядное представление о форме и направлении распределения данных. Эллипсы рассеивания помогают визуализировать линейную зависимость между переменными и выявить направление и силу корреляции. Чем сильнее корреляция, тем более вытянутым будет эллипс.

3 Методология

3.1 Описание данных

Данные содержат несколько переменных, измеренных на различных интервалах. Анализ включает построение диаграмм рассеивания с эллипсами и вычисление коэффициентов корреляции.

3.2 Аналитические методы

Анализ данных осуществляется с использованием Python. Скрипт читает данные из файла, выполняет статистический анализ и строит диаграммы.

4 Результаты

4.1 Диаграммы рассеивания с эллипсами

Рис. 1 Эллипс рассеивания

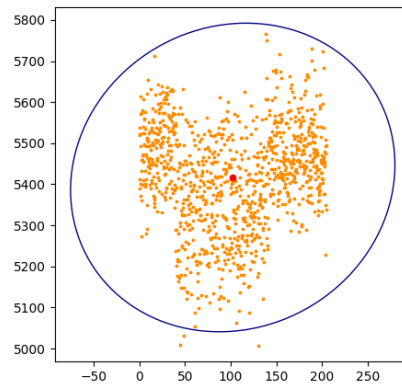


Рис. 1: 0.05Vsp953.dat

Рис. 1 Эллипс рассеивания

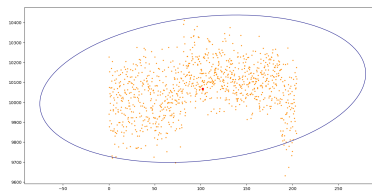


Рис. 2: 0.15Vsp93.dat

Рис. 1 Эллипс рассеивания

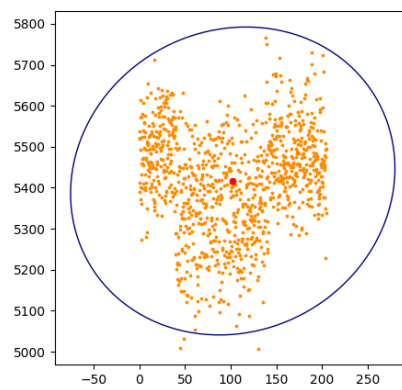


Рис. 3: -0.15Vsp831.dat

4.2 Коэффициенты корреляции

Таблица 1: Коэффициенты для файла 0.05Vsp953.dat

Metric	Pearson	Spearman	Quadrant
Mean	0.082	0.06	0.054
Mean of Squares	0.007	0.005	0.004
Variance	0.001	0.001	0.001

Таблица 2: Коэффициенты для файла 0.15Vsp93.dat

Metric	Pearson	Spearman	Quadrant
Mean	0.206	0.217	0.23
Mean of Squares	0.044	0.048	0.054
Variance	0.001	0.001	0.001

Таблица 3: Коэффициенты для файла-0.15Vsp831.dat

Metric	Pearson	Spearman	Quadrant
Mean	0.29	0.273	0.281
Mean of Squares	0.085	0.076	0.08
Variance	0.001	0.001	0.001

5 Выводы

При анализе представленных таблиц коэффициентов корреляции для трех файлов данных (0.05Vsp953.dat, 0.15Vsp93.dat, 0.15Vsp831.dat) можно сделать следующие выводы:

Файл 0.05Vsp953.dat: Среднее значение (Mean): Значения коэффициентов корреляции (Pearson: 0.082, Spearman: 0.06, Quadrant: 0.054) показывают слабую положительную корреляцию. Среднее квадратичное значение (Mean of Squares): Значения (Pearson: 0.007, Spearman: 0.005, Quadrant: 0.004) также показывают очень слабую положительную корреляцию. Дисперсия (Variance): Все коэффициенты равны 0.001, что указывает на отсутствие корреляции или на крайне слабую корреляцию.

Файл 0.15Vsp93.dat: Среднее значение (Mean): Значения коэффициентов корреляции (Pearson: 0.206, Spearman: 0.217, Quadrant: 0.23) указывают на умеренную положительную корреляцию. Среднее квадратичное значение (Mean of Squares): Значения (Pearson: 0.044, Spearman: 0.048, Quadrant: 0.054) демонстрируют слабую положительную корреляцию. Дисперсия (Variance): Все коэффициенты равны 0.001, что свидетельствует о крайне слабой корреляции.

Файл 0.15Vsp831.dat: Среднее значение (Mean): Значения коэффициентов корреляции (Pearson: 0.29, Spearman: 0.273, Quadrant: 0.281) указывают на умеренную положительную корреляцию. Среднее квадратичное значение (Mean of Squares): Значения (Pearson: 0.085, Spearman: 0.076, Quadrant: 0.08) показывают слабую положительную корреляцию. Дисперсия (Variance): Все коэффициенты равны 0.001, что указывает на крайне слабую корреляцию.

Общие наблюдения

Увеличение значений среднего: В среднем, коэффициенты корреляции для метрики "Mean" показывают более высокую корреляцию в сравнении с другими метриками, особенно для файлов с большими значениями (0.15Vsp93.dat и 0.15Vsp831.dat). Это указывает на то, что при увеличении среднего значения данных, корреляция становится более выраженной.

Стабильность дисперсии: Независимо от файла и типа коэффициента корреляции, значения для дисперсии остаются одинаковыми (0.001). Это свидетельствует о стабильности и низком уровне корреляции в отношении вариации данных.

Сравнение методов оценки: Pearson, Spearman и Quadrant показывают схожие тенденции корреляции, однако Quadrant метод немного чаще показывает чуть более высокие значения корреляции. Это может свидетельствовать о более чувствительном подходе этого метода к специфике данных.

6 Реализация

<https://github.com/1Qwix1/MathStatistics>