

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені Тараса Шевченка  
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  
Кафедра програмних систем і технологій

Дисципліна  
«Ймовірнісні основи програмної інженерії»

Лабораторна робота № 3

Виконав:	Мельничук Дмитро Олегович	Перевірив:	Вечерковська А.С.
Група	ІПЗ-24(1)	Дата перевірки	
Форма навчання	денна	Оцінка	
Спеціальність	121		
2022			

# ДВОВИМІРНА СТАТИСТИКА

Мета – навчитись використовувати на практиці набуті знання про міри в двовимірній статистиці.

## Завдання

1. Намалюйте діаграму розсіювання для даних. Укажіть, чи існує тренд у даних. Якщо так, то вкажіть, чи є це негативним трендом, чи позитивним.
2. Знайдіть центр ваги і коваріацію.
3. Знайти рівняння лінії регресії у від х.
4. Розрахуйте коефіцієнт кореляції між даними.
5. Зробити висновок про залежності

Математична модель:

Завдання 1:

Для побудови діаграми розсіювання необхідно відокремити суму покупки та час проведений в магазині і відповідно підставити на осі ОХ та ОУ.

Для знаходження тренду необхідно проаналізувати графік і побачити чи перший елемент відрізняється від останнього, якщо він буде більший від останнього то тренд спадаючий (негативний), якщо менший то спадаючий(позитивний)

Завдання 2:

Для знаходження центру ваги скористаємося формулою:

$$\vec{r}_c = \frac{\sum_i m_i \vec{r}_i}{\sum_i m_i},$$

Де:

$r_c$  – Центр ваги

$m_i$  – поточний елемент X

$r_i$  – поточний елемент Y

Коваріація - це величина, яка відображає, наскільки дві випадкові величини спільно змінюються щодо їхніх значень.

Для її розрахунку скористаємося формулою:

$$Cov(X, Y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$Cov(X, Y)$  – коваріація

$X_i$  – поточний елемент  $X$

$Y_i$  – поточний елемент  $Y$

$\bar{X}$  – Середнє значення  $X$

$\bar{Y}$  – Середнє значення  $Y$

$n$  – кількість елементів

Завдання 3:

Для знаходження рівняння лінії регресії  $y$  від  $x$  використаємо формулу:

$$y - \bar{y} = b(x - \bar{x})$$

Знайдемо  $b$  за формулою:

$$b_{yx} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

Завдання 4:

Для знаходження кореляції між даними використаємо формулу:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Де:

$r$  - кореляція

Завдання 5:

Проаналізувавши всі дані можна сказати, що залежність  $y$  від  $x$  є прямопропорційною тобто зі збільшенням значення  $x$  значення  $y$  збільшується, відповідно і зі збільшення  $y$  збільшується  $x$

Псевдокод алгоритмів:

Тренд:

- Викликається функція `trand`, в яку передаються значення вхідного масиву даних `data`
- Виконується перевірка:
  - чи час проведений в магазині в початку масиву більший ніж час проведений в магазині в кінці масиву, відповідно виводиться що існує негативний тренд,
  - чи час проведений в магазині в початку масиву менший ніж час проведений в магазині в кінці масиву, відповідно виводиться що існує позитивний тренд,
  - якщо час однаковий, то виводиться повідомлення про відсутність тренду

Результат на консолі:

```
Positive trand  
[[1.5, 12], [2.4, 19], [2.6, 29], [2.8, 34], [3.6, 26], [3.8, 32], [3.9, 45], [4.1, 30], [4.4, 28], [4.6, 40], [5.4, 43], [7.6, 82]]
```

Центр ваги:

- Викликається функція `gravitycenter`, в яку передається значення масиву даних `data`
- Створюються допоміжні змінні:
  - `chis`, що відповідає за чисельник, встановлюється початкове значення 0
  - `znam`, що відповідає за знаменник, встановлюється початкове значення 0
- Створюється цикл в якому перебираються кількість всіх пар даних масиву `data`, і встановлює відповідне значення тимчасової змінної `i` від 0 до останньої пари елементів
  - З кожною ітерацією додає до змінної `chis` добуток `i` – тої пари елементів масиву `data`
  - З кожною ітерацією додає до змінної `znam` значення першого елемента в `i` – тої пари елементів масиву `data`
- Повертає значення різниці `chis` та `znam`

Результат на консолі:

```
Gravitycenter = 40.9100642398287
```

## Коваріація:

- Викликається функція `covariance`, в яку передається значення масиву даних `data`
- Створюється додаткова змінна `covariance` зі значенням 0
- Викликається функція `middleXY`, в яку передається значення масиву `data`, в якій розраховуються змінні `middleX`, та `middleY`
- Створюється цикл в якому перебираються кількість всіх пар даних масиву `data`, і встановлює відповідне значення тимчасової змінної `i` від 0 до останньої пари елементів
  - З кожною ітерацією циклу додає до змінної `covariance` добуток двох елементів, які складаються з :
    - Різниці першого елемента `i`-тої пари масиву `data`, та середнього значення `x` зі змінної `middle`
    - Різниці другого елемента `i`-тої пари масиву `data`, та середнього значення `y` зі змінної `middleY`
- Повертає значення різниці змінної `covariance` та кількості пар елементів масиву `data`

## Результат на консолі:

```
Covariance = 23.0
```

## Рівняння лінії регресії:

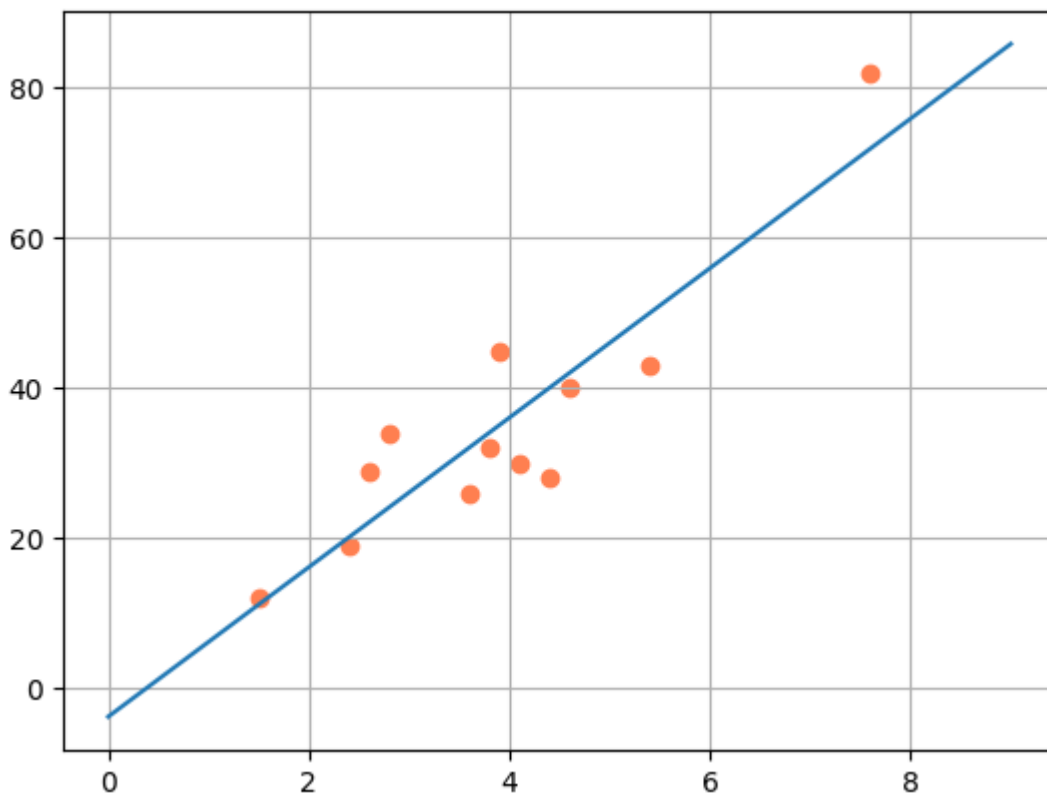
- Викликається функція `regretionline`, в яку передається значення масиву даних `data`
- Створюються допоміжні змінні:
  - `dobXY`, яка рівна 0 – для добутку `x` та `y`
  - `sumX`, яка рівна 0 – для суми елементів `x`
  - `sumY`, яка рівна 0 – для суми елементів `y`
  - `sumdobXX`, яка рівна 0 – для суми піднесення `y` в квадрат елементів `x`
- Створюється цикл в якому перебираються кількість всіх пар даних масиву `data`, і встановлює відповідне значення тимчасової змінної `i` від 0 до останньої пари елементів
  - З кожною ітерацією додає до змінної `dobXY`, добуток першого та другого елементів `i`-тої пари елементів масиву `data`
  - З кожною ітерацією додає до змінної `sumX`, перший елемент `i`-тої пари елементів масиву `data`
  - З кожною ітерацією додає до змінної `sumY`, другий елемент `i`-тої пари елементів масиву `data`
  - З кожною ітерацією додає до змінної `sumdobXX`, перший елемент `i`-тої пари елементів масиву `data` піднесений у квадрат
- Створюється змінна `b`, яка використовується для збереження різниці елементів:
  - Кількість елементів масиву помножена на змінну `dobXY` відняти добуток змінних `sumX` та `sumY`

- Кількість елементів масиву помножена на змінну `sumdobXX`, відняти змінну `sumX` у квадраті
- Використаємо бібліотеку `sympy` для знаходження рівняння:
- Створимо символи `X`, `Y` для відображення, у формулах будемо використовувати їх аналоги `x` та `y`
- Створимо змінну `line`, в якій буде записано рівняння за допомогою функції `equal`, яка прирівнює 2 значення:
  - `y - middley` [середнє значення] `y`,
  - `b * (x - middlex` [середнє значення] `x`)
- виведемо значення рівняння відносно `y` за допомогою функції `solve`, в яку передамо значення змінної `line` та `x`
- виведемо значення рівняння відносно `x` за допомогою функції `solve`, в яку передамо значення змінної `line` та `y`

Результат на консолі:

```
Y = [9.95341848234409*X - 3.7353869271224]
X = [0.100467995169082*Y + 0.375286835748786]
```

Виведення графіку:



### Кореляція:

- Викликається функція `correlation`, в яку передається значення масиву даних `data`
- Створюються допоміжні змінні:
  - `chis`, яка рівна 0 – числівника
  - `sumXX`, яка рівна 0 – для різниці поточного елемента `x` та середнього значення `x`
  - `sumYY`, яка рівна 0 – для різниці поточного елемента `y` та середнього значення `y`
- Створюється цикл в якому перебираються кількість всіх пар даних масиву `data`, і встановлює відповідне значення тимчасової змінної `i` від 0 до останньої пари елементів
  - З кожною ітерацією додає до змінної `chis`, добуток різниці першого елемента з `i`-тої пари масиву `data` і середнього значення `x` та другого елемента з `i`-тої пари масиву `data` і середнього значення `y`
  - З кожною ітерацією додає до змінної `sumXX`, піднесену у квадрат різницю першого елемента `i`-тої пари елементів масиву `data` та середнього значення `x`
  - З кожною ітерацією додає до змінної `sumYY`, піднесену у квадрат різницю першого елемента `i`-тої пари елементів масиву `data` та середнього значення `y`
- Повертає різницю числівника та квадратного корня добутку змінних `sumXX` та `sumYY`

### Результат на консолі:

```
Correlation = 0.9010014623100246
```

### Висновок:

Навчився на практиці використовувати набуті знання про міри в двовимірній статистиці, будувати діаграму розподілу та аналізувати її