

**Київський національний університет імені Тараса Шевченка**

**Факультет кібернетики**

**Кафедра системного аналізу та теорії прийняття рішень**

**Звіт до лабораторної роботи №3**

**На тему:**

**«Побудова математичної моделі методом аналізу даних»**

студента 3 курсу

групи САТР-3

Каплана Іллі

## Зміст

1. Постановка задачі
2. Опис вхідної інформації
3. Аналіз роботи
4. Висновок
5. Література

## Постановка задачі

Потрібно провести побудову математичної моделі методами аналізу даних.

### Опис вхідної інформації

Данні були взяті з сайту Kaggle (URL: <https://www.kaggle.com/datasets>).

Назва dataset — Population of USA

(посилання: <https://www.kaggle.com/datasets/anandhuh/population-data-usa>).

Було обрано 3 скалярні змінні:

1. Population;
2. Yearly Change;
3. Migrants.

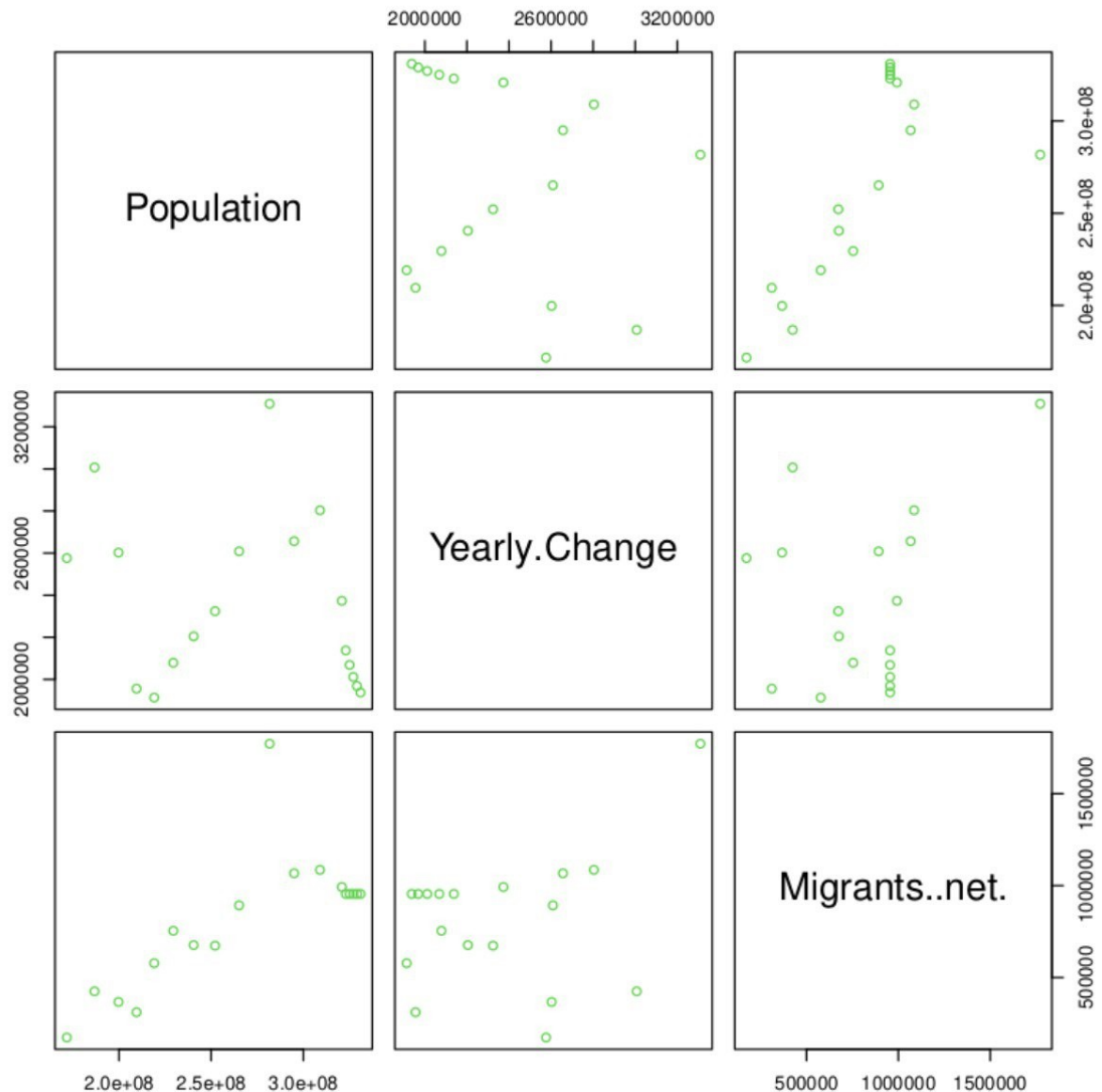
```
data <- read.csv('./data/population_usa.csv')
attach(data)

newData <- data.frame(Population, Yearly.Change, Migrants..net.)
```

### Аналіз роботи

Побудуємо матричну діаграму даних

```
pairs(newData, col=3)
```



Спостерігається лінійна залежність Population від інших двох регресорів. Будуємо моделі множинної регресії, де залежною змінною є Population, а незалежними - Yearly Change, Migrants. Потім дивимось на summary.

```
rm <- lm(Population ~ Migrants..net. + Yearly.Change, data = newData)
summary(rm)
```

```

Call:
lm(formula = Population ~ Migrants..net. + Yearly.Change, data = newData)

Residuals:
    Min       1Q   Median       3Q      Max
-50304183 -14680828  9464584 16526970 34034204

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.180e+08  4.062e+07   7.828 1.12e-06 ***
Migrants..net.  1.321e+02  1.955e+01   6.759 6.43e-06 ***
Yearly.Change -6.650e+01  1.765e+01  -3.767  0.00186 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28380000 on 15 degrees of freedom
Multiple R-squared:  0.7664,    Adjusted R-squared:  0.7352
F-statistic: 24.6 on 2 and 15 DF,  p-value: 1.836e-05

```

Аналіз результатів

Отже, модель має вигляд

Population =  $3.180 \times 10^8 + 1.321 \times 10^2 \cdot \text{Migrants} + (-6.650 \times 10^1) \cdot \text{Yearly Change}$

Коефіцієнт детермінації і скоригований коефіцієнт детермінації:

$R^2 = 0.7664$

*Adjusted*  $R^2 = 0.7352$

Аналіз залишків.

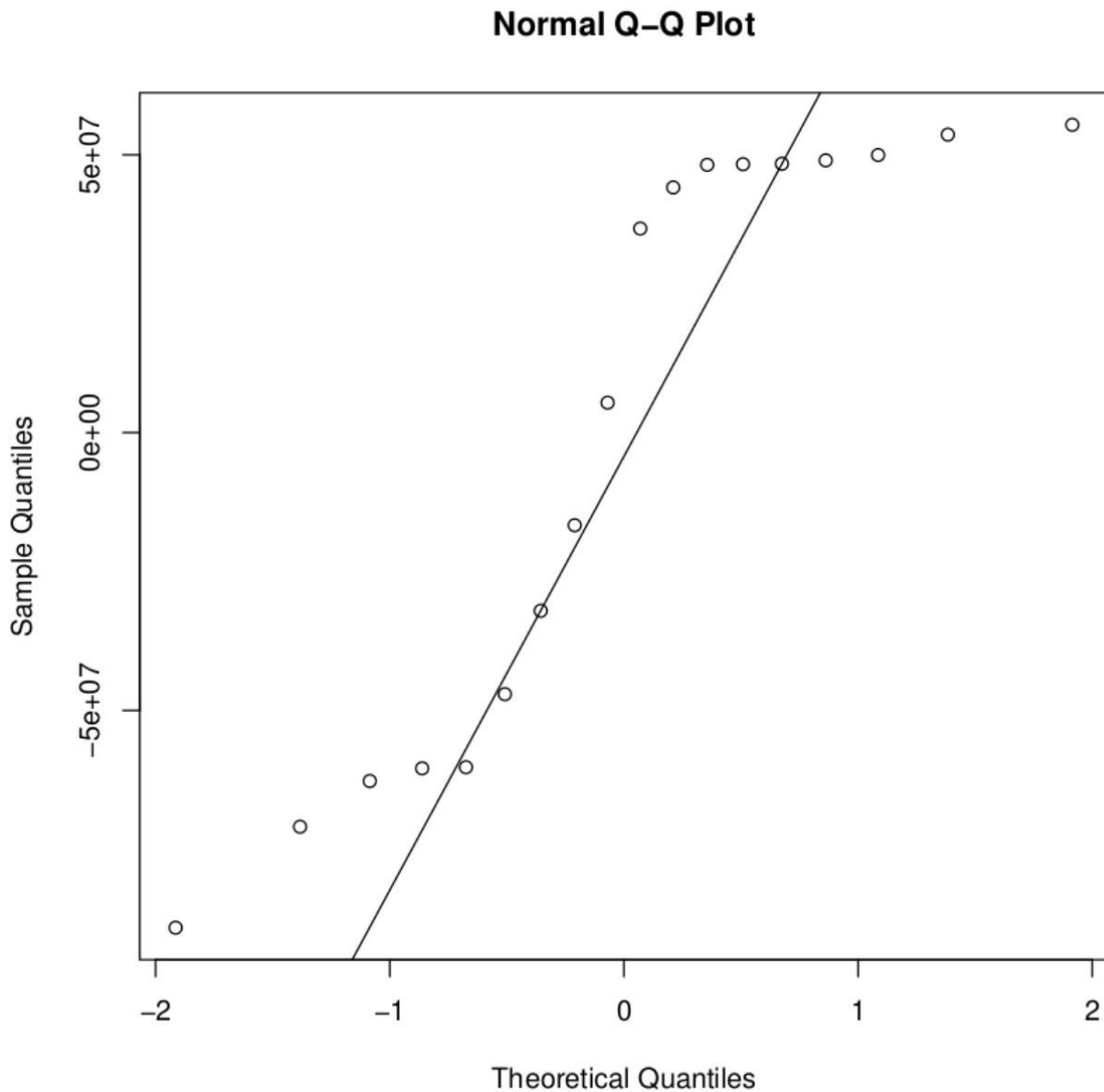
Будуємо q-q-діаграму на порівняння залишків з нормальним розподілом.

А тепер будуємо просто графік залишків – залежність між залишками і передбаченими регресією значеннями Population.

```

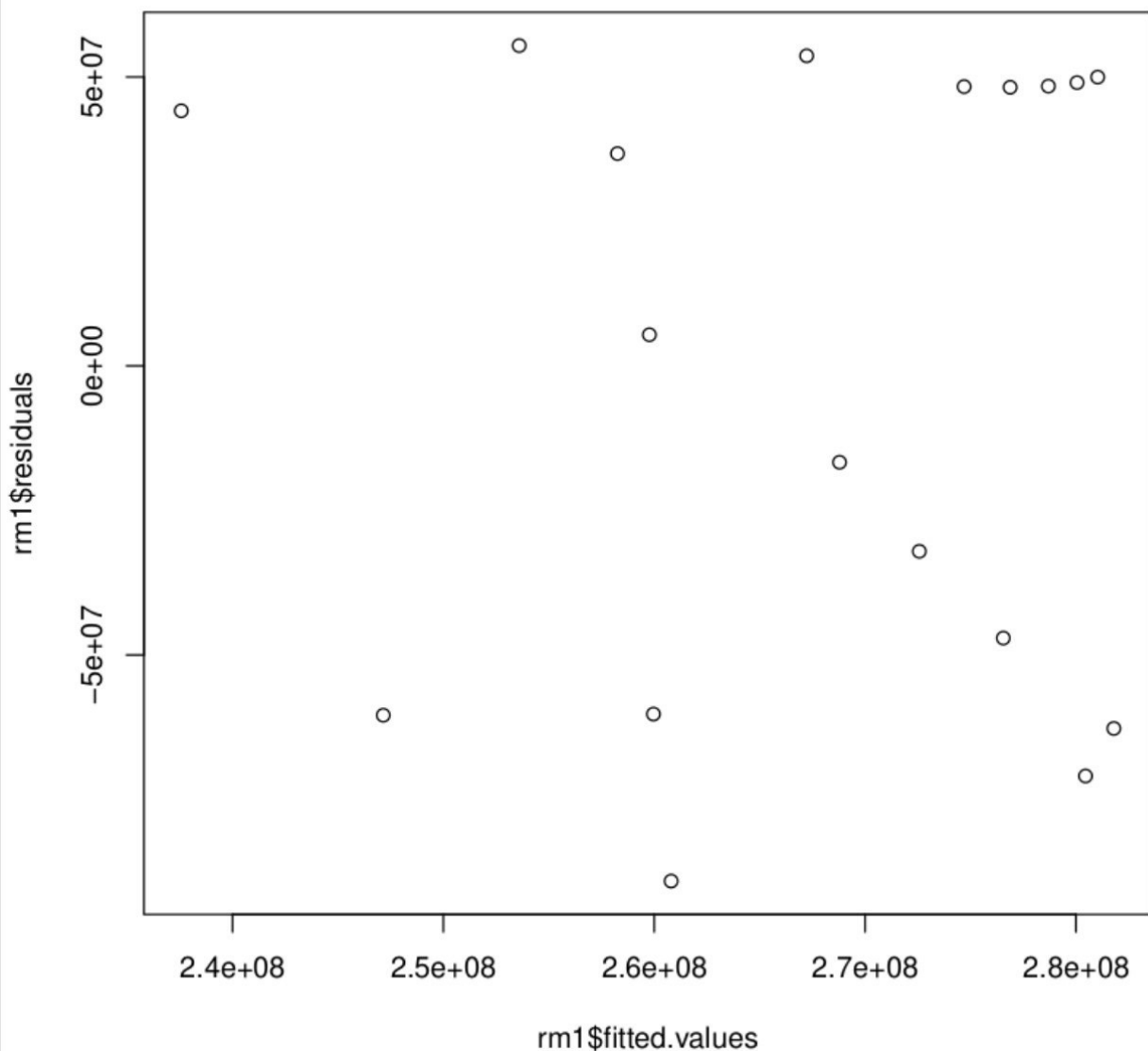
qqnorm(rm1$residuals)
qqline(rm1$residuals)

```



А тепер будемо просто графік залишків – залежність між залишками і передбаченими регресією значеннями Population.

```
plot(rm1$fitted.values, rm1$residuals)
```



Бачимо, що залежність відсутня, і це добре характеризує нашу модель.

#### Висновок

В ході роботи я навчився будувати математичну модель методами аналізу даних.

#### Література

1. Майборода Р.Є. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019.
2. Майборода Р.Є. Аналіз даних за допомогою пакета R: Навчальний посібник/ Р. Є. Майборода, О.В. Сугакова. – К. : КНУ, 2015.