

Київський національний університет імені Тараса Шевченка

Факультет кібернетики

Кафедра системного аналізу та теорії прийняття рішень

Звіт до лабораторної роботи №2

На тему:

«Кореляційний аналіз 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 кореляційний аналіз обраних для обробки скалярних змінних, яких потрібно взяти не менше трьох.

На основі результатів попередньої обробки обраного набору даних визначитися, які характеристики статистичного зв'язку потрібно використати при подальшому їх кореляційному аналізі.

Провести аналіз істотності парних статистичних зв'язків для усіх пар скалярних змінних.

Провести аналіз істотності множинних статистичних зв'язків між кожною обраною в якості залежної скалярною змінною та множиною усіх інших скалярних змінних.

Опис вхідної інформації

В обраному датасеті розкладається статистика населення США за 55 років.

Змінні, з якими я буду працювати

1. Population;
2. Urban Population;
3. World Population.

Аналіз істотності парних статистичних зв'язків

1. Коефіцієнт кореляції та рівень значущості

```
cor.test(Population, Urban.Population)
cor.test(Population, World.Population)
cor.test(Urban.Population, World.Population)
```

```
Pearson's product-moment correlation

data: Population and Urban.Population
t = 127.38, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9986449 0.9998209
sample estimates:
      cor
0.9995073
```

```
Pearson's product-moment correlation

data: Population and World.Population
t = 83.011, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9968145 0.9995786
sample estimates:
      cor
0.9988411
```

```
Pearson's product-moment correlation

data: Urban.Population and World.Population
t = 59.39, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9937929 0.9991779
sample estimates:
      cor
0.9977396
```

- 1.(Population;Urban. Population)
- 2.(Population;World. Population)
- 3.(Urban.Population;World. Population)

Для пари кожної пари рівень значущості менше $2.2e-16$, хоча за означенням розподіл значущості рівномірний на відріжку (0, 1).

Аналіз істотності множинних статистичних зв'язків

```
populationModel <- lm(Population ~ Urban.Population + World.Population)
urbanPopulationModel <- lm(Urban.Population ~ Population + World.Population)
worldPopulationModel <- lm(World.Population ~ Urban.Population + Population)

cor.test(populationModel$model$Population, populationModel$fitted.values)
cor.test(urbanPopulationModel$model$Urban.Population, urbanPopulationModel$fitted.values)
cor.test(worldPopulationModel$model$World.Population, worldPopulationModel$fitted.values)
```

```
Pearson's product-moment correlation

data: populationModel$model$Population and populationModel$fitted.values
t = 194.41, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9994179 0.9999231
sample estimates:
      cor 
0.9997884

Pearson's product-moment correlation

data: urbanPopulationModel$model$Urban.Population and urbanPopulationModel$fitted.values
t = 139.21, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9988653 0.9998500
sample estimates:
      cor 
0.9995875

Pearson's product-moment correlation

data: worldPopulationModel$model$World.Population and worldPopulationModel$fitted.values
t = 90.736, df = 16, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9973326 0.9996472
sample estimates:
      cor 
0.9990297
```

1.Population;

2.Urban Population;

3.World Population.

Множинний статистичний зв'язок неможливий.

Висновок

В цій лабораторній я навчився користуватись основними інструментами кореляційного аналізу, проводити аналіз істотності парних статистичних зв'язків та істотності множинних статистичних зв'язків.

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.