Лабораторная работа №1: Предварительный анализ данных

Айгунов Гаджи-Мурад

27 10 2020

*Раздел I*

В этом разделе собираются данные по социально-экономическим показателям в регионах РФ и изучаются их взаимосвязи.

**Необходимо построить:** \*Коробчатые диаграммы на одном полотне;

\*Таблицу со статистикой теста Шапиро-Уилка на нормальность;

\*Точечные графики взаимного разброса;

\*Графическое представление корреляционной матрицы.

Обозначения количественных показателей:

\*X1 - Валовой региональный продукт 2017г;

\*X2 - Инвестиции в основной капитал 2016г;

\*X3 - Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику 2016г;

\*X4 - Использование информационных и коммуникационных технологий в организациях: персональные компьютеры 2016г;

\*X5 - Внутренние затраты на научные исследования и разработки 2016г;

Файл с данными содержит 87 строк и 6 столбцов.

*Описательная статистика*

## X1 X2 X3 X4 X5   
## 968452.7 199820.6 23787.5 93.0 11288.2

## X1 X2 X3 X4   
## 1928924.6 329605.3 53980.1 5.6

## Warning in sds/mns: длина большего объекта не является произведением длины  
## меньшего объекта

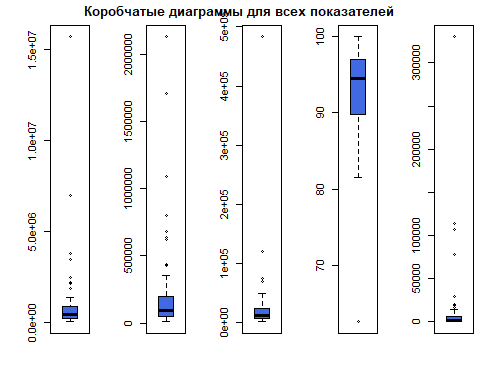
## X1 X2 X3 X4 X5   
## 199.2 165.0 226.9 6.0 17088.0

## Warning in rbind(mns, sds, coef.vars): number of columns of result is not a  
## multiple of vector length (arg 2)

## X1 X2 X3 X4 X5  
## Среднее 968452.7 199820.6 23787.5 93.0 11288.2  
## Стандартное отклонение 1928924.6 329605.3 53980.1 5.6 1928924.6  
## Коэффициент вариации, % 199.2 165.0 226.9 6.0 17088.0

**Вывод:** показатели неоднородны.

*Анализ распределения данных*

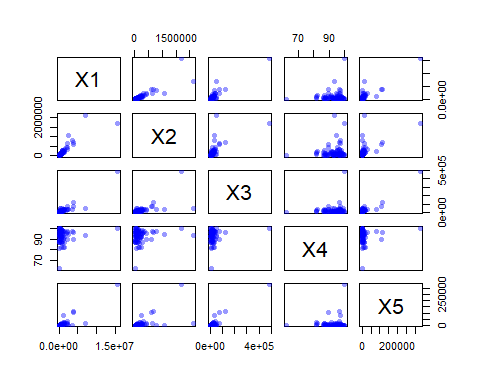


*Тест Шапиро-Уилка на нормальность распределения*

## X1.W X2.W X3.W X4.W X5.W  
## Статистика теста 0.4033 0.5181 0.2935 0.826 0.2701  
## Критическое значение 0.0000 0.0000 0.0000 0.000 0.0000

**Вывод:** На уровне значимости 0,05 мы можем отвергнуть нулевую гиппотезу о нормальности закона распределения у всех показателей, поскольку все p-значения меньше 0,05.

*Графики разброса*

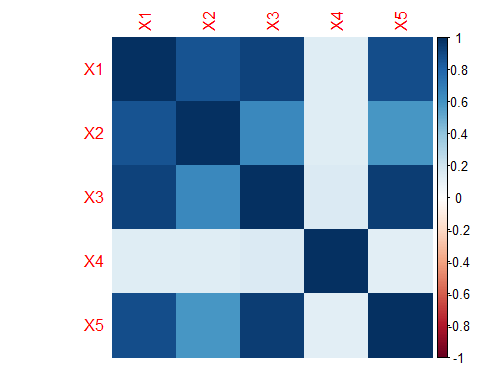


*Корреляционная матрица*

## X1 X2 X3 X4 X5  
## X1 1.00 0.87 0.92 0.14 0.88  
## X2 0.87 1.00 0.64 0.13 0.58  
## X3 0.92 0.64 1.00 0.16 0.94  
## X4 0.14 0.13 0.16 1.00 0.13  
## X5 0.88 0.58 0.94 0.13 1.00  
##   
## n= 84   
##   
##   
## P  
## X1 X2 X3 X4 X5   
## X1 0.0000 0.0000 0.2081 0.0000  
## X2 0.0000 0.0000 0.2354 0.0000  
## X3 0.0000 0.0000 0.1509 0.0000  
## X4 0.2081 0.2354 0.1509 0.2500  
## X5 0.0000 0.0000 0.0000 0.2500

## X1 X2 X3 X4 X5  
## X1 1.0000000 0.8650892 0.9221475 0.1387613 0.8812591  
## X2 0.8650892 1.0000000 0.6446075 0.1308644 0.5820467  
## X3 0.9221475 0.6446075 1.0000000 0.1580832 0.9449917  
## X4 0.1387613 0.1308644 0.1580832 1.0000000 0.1269139  
## X5 0.8812591 0.5820467 0.9449917 0.1269139 1.0000000

## X1 X2 X3 X4 X5  
## X1 NA 0.000000e+00 0.000000e+00 0.2080983 0.000000e+00  
## X2 0.0000000 NA 3.672884e-11 0.2354126 6.359250e-09  
## X3 0.0000000 3.672884e-11 NA 0.1509469 0.000000e+00  
## X4 0.2080983 2.354126e-01 1.509469e-01 NA 2.499729e-01  
## X5 0.0000000 6.359250e-09 0.000000e+00 0.2499729 NA



**Вывод:** Значимая взаимосвязь наблюдается между переменными X1, X2, X3, X5.

*Раздел II* В этом разделе необходимо прологарифмировать данные из первого раздела и провести с новыми данными те же махинации.

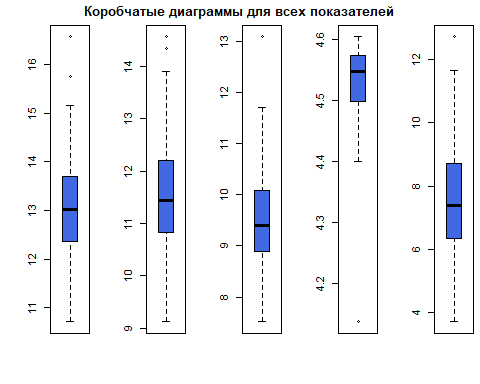
*Описательная статистика логарифмированных данных:*

## X1 X2 X3 X4 X5  
## Среднее 13.081 11.528 9.485 4.531 7.542  
## Стандартное отклонение 1.100 1.117 0.946 0.066 1.783  
## Коэффициент вариации, % 8.409 9.689 9.974 1.457 23.641

**Вывод:** показатели не полностью однородны.

*Анализ распределения*

Построим коробчатые диаграммы.



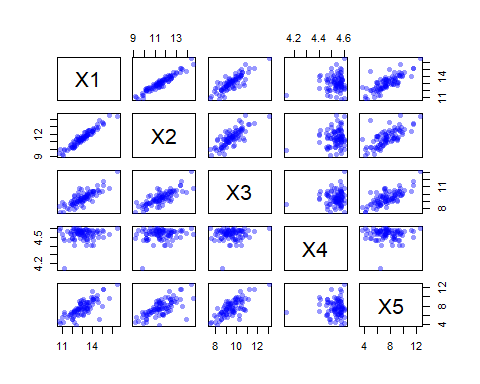
*Тест Шапиро-Уилка на нормальность распределения.*

## $X1  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.98614, p-value = 0.5084  
##   
##   
## $X2  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.98781, p-value = 0.6193  
##   
##   
## $X3  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.9749, p-value = 0.09971  
##   
##   
## $X4  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.76592, p-value = 3.028e-10  
##   
##   
## $X5  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.98768, p-value = 0.6103

## X1.W X2.W X3.W X4.W X5.W  
## Статистика теста 0.4033 0.5181 0.2935 0.826 0.2701  
## Критическое значение 0.0000 0.0000 0.0000 0.000 0.0000

**Вывод:** На уровне значимости 0.05 мы отвергаем гипотезу о нормальности распределения показателей.

*Графики разброса*

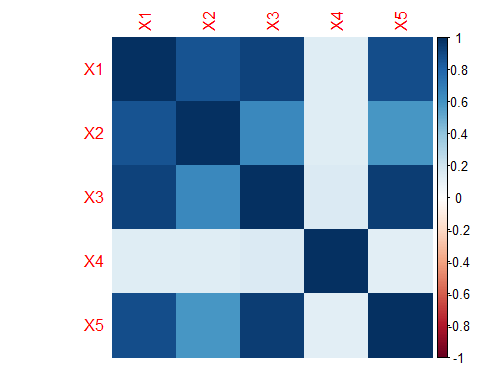


*Корреляционная матрица*

## X1 X2 X3 X4 X5  
## X1 1.00 0.96 0.83 0.14 0.77  
## X2 0.96 1.00 0.80 0.17 0.69  
## X3 0.83 0.80 1.00 0.12 0.74  
## X4 0.14 0.17 0.12 1.00 -0.01  
## X5 0.77 0.69 0.74 -0.01 1.00  
##   
## n= 84   
##   
##   
## P  
## X1 X2 X3 X4 X5   
## X1 0.0000 0.0000 0.2159 0.0000  
## X2 0.0000 0.0000 0.1316 0.0000  
## X3 0.0000 0.0000 0.2809 0.0000  
## X4 0.2159 0.1316 0.2809 0.9305  
## X5 0.0000 0.0000 0.0000 0.9305

## X1 X2 X3 X4 X5  
## X1 1.0000000 0.9633509 0.8283583 0.136441223 0.772665532  
## X2 0.9633509 1.0000000 0.7990359 0.165874056 0.691400990  
## X3 0.8283583 0.7990359 1.0000000 0.119005491 0.738245777  
## X4 0.1364412 0.1658741 0.1190055 1.000000000 -0.009666498  
## X5 0.7726655 0.6914010 0.7382458 -0.009666498 1.000000000

## X1 X2 X3 X4 X5  
## X1 NA 0.000000e+00 0.000000e+00 0.2158778 0.000000e+00  
## X2 0.0000000 NA 0.000000e+00 0.1315676 3.326228e-13  
## X3 0.0000000 0.000000e+00 NA 0.2809457 1.110223e-15  
## X4 0.2158778 1.315676e-01 2.809457e-01 NA 9.304573e-01  
## X5 0.0000000 3.326228e-13 1.110223e-15 0.9304573 NA



**Вывод:** Значимая взаимосвязь наблюдается между переменными X1, X2, X3, X5.