Лабораторная работа №1

Людуховский В.В.

13 11 2020

# Раздел 1.

В этом разделе необходимо собрать данные по социально-экономическим показателям в регионах РФ и изучить их взаимосвязи.

# **Необходимо построить:**

* Коробчатые диаграммы (совместить коробки для всех переменных на одном графике);
* Тест Шапиро-Уилка на нормальность: таблица со статистикой теста, критическим значением статистики и выводом по каждой переменной.
* Точечные графики взаимного разброса с полупрозрачными маркерами;
* Взаимосвязь показателей: графическое представление корреляционной матрицы.

## Обозначение показателей:

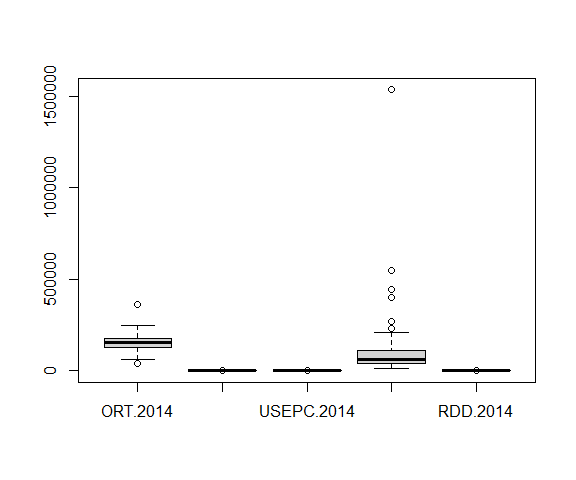
* *ORT.2014* - Оборот розничной торговли на душу населения (руб.)
* *KMP.2014* - Число малых предприятий на 10000 человек населения (шт.)
* *USEPC.2014* - Использование информационных и коммуникационных технологий в организациях: персональные компьютеры ()
* *Budg.2013* - Расходы консолидированных бюджетов субъектов Российской Федерации: всего
* *RDD.2014* - Реальные денежные доходы населения

## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ORT.2014 | KMP.2014 | USEPC.2014 | Budg.2013 | RDD.2014 |
| Среднее | 153849.0 | 129.8 | 94.4 | 111163.1 | 101.2 |
| Стандартное отклонение | 46172.4 | 55.1 | 4.3 | 180990.8 | 3.8 |
| Коэффициент вариации, % | 30.0 | 42.4 | 4.6 | 162.8 | 3.8 |

## Анализ распределния данных

Построим коробчатые диаграммы (совместим коробки для всех переменных на одном графике)



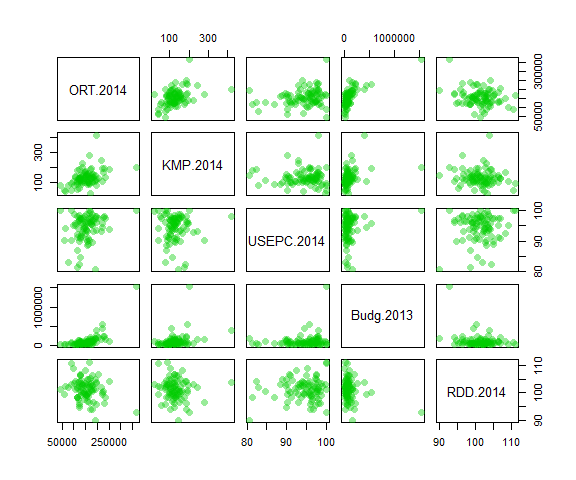
## Тест Шапиро-Уилка на нормальность

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ORT.2014 | KMP.2014 | USEPC.2014 | Budg.2013 | RDD.2014 |
| Статистика | 0.93923 | 0.87247 | 0.89216 | 0.41768 | 0.99009 |
| Крит.значение | 0.00058 | 0.00000 | 0.00000 | 0.00000 | 0.76992 |

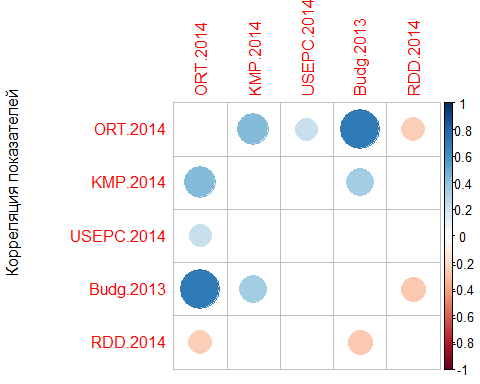
**Вывод:** По итогам теста Шапиро-Уилка на нормальность распредления, при уровне значимости 0,05, отвергаются все, кроме *“RDD.2014”*,т.к. все остальные P-значения значения < 0.05

## Анализ взаимосвязей показателей

### Графики разброса



### Корреляционная матрица



**Вывод:** Наблюдается сильная корреляционная связь между показателями *‘Budg.2013’* и *‘ORT.2014’*.

# Раздел 2.

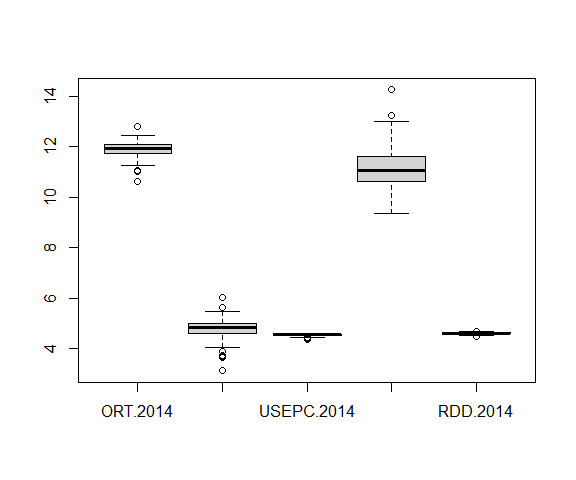
В этом разделе необходимо произвести теже самые действия, с логарифмированными исходными данными.

## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ORT.2014 | KMP.2014 | USEPC.2014 | Budg.2013 | RDD.2014 |
| Среднее | 11.9 | 4.8 | 4.5 | 11.2 | 4.6 |
| Стандартное отклонение | 0.3 | 0.4 | 0.0 | 0.9 | 0.0 |
| Коэффициент вариации, % | 2.5 | 8.3 | 0.0 | 8.0 | 0.0 |

## Анализ распределния логарифмированных данных

Построим коробчатые диаграммы (совместим коробки для всех переменных на одном графике)



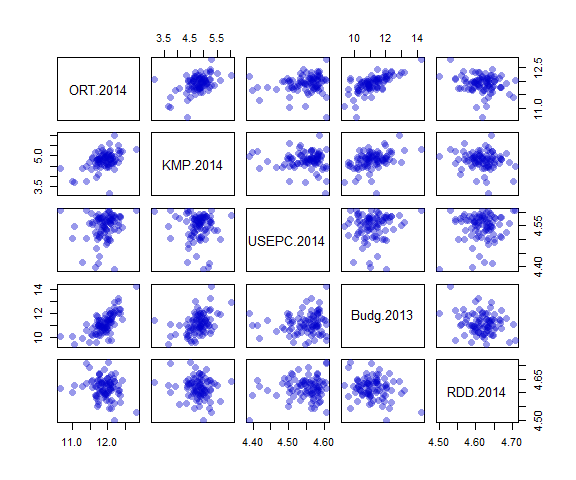
## Тест Шапиро-Уилка на нормальность для логарифмированных данных

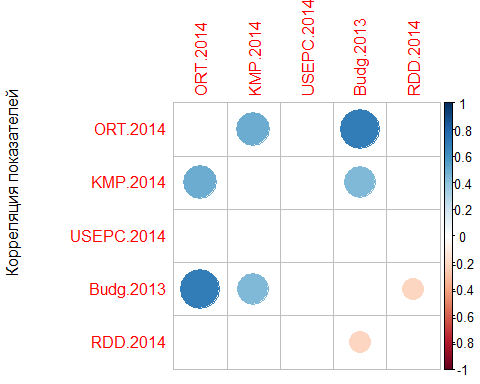
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ORT.2014 | KMP.2014 | USEPC.2014 | Budg.2013 | RDD.2014 |
| Статистика | 0.93758 | 0.93769 | 0.8743 | 0.97648 | 0.98784 |
| Крит.значение | 0.00047 | 0.00048 | 0.0000 | 0.12245 | 0.61295 |

**Вывод:** По итогам теста Шапиро-Уилка на нормальность распредления логарифмированных переменных, при уровне значимости 0,05, отвергаются первые три переменные *‘ORT.2014’*, *‘KMP.2014’*, *‘USEPC.2014’*, т.к. P-значения < 0.05. А переменные *‘Budg.2013’* и *‘RDD.2014’* не отвергаются.

## Анализ взаимосвязей логарифмированных показателей

### Графики разброса логарифмированных величин





**Вывод:** Наблюдается сильная корреляционная связь между показателями *‘Budg.2013’* и *‘ORT.2014’*.