Лабораторная работа №1

Иванова Ксения

28 12 2020

## Раздел I.

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

**Необходимо построить:**

* гистограммы на одном полотне, отразив теоретические плотности распределения;
* таблицу со статистикой теста Андерсона-Дарлинга на нормальность;
* точечные графики взаимного разброса с полупрозрачными маркерами;
* графическое представление корреляционной матрицы.

### Обозначения количественных показателей:

* *PMI.2015* – Среднедушевые денежные доходы населения
* *GRPpc.2015* – ВРП на душу населения
* *RTT.2015* – Оборот розничной торговли на душу населения
* *ECB.2014* – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику
* *NSB.2015* – Число малых предприятий на 10000 человек населения

## Импорт данных:

#Файл с данными содержит 92 строк и 8 столбцов.

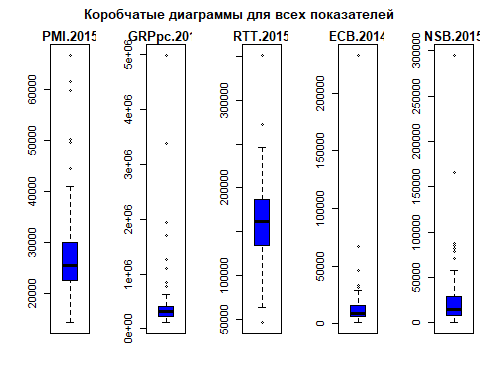
## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PMI.2015 | GRPpc.2015 | RTT.2015 | ECB.2014 | NSB.2015 |
| Среднее | 27653.7 | 476974.1 | 162172.5 | 14355.2 | 26498.4 |
| Стандартное отклонение | 9785.6 | 668480.9 | 46253.5 | 26401.5 | 39414.4 |
| Коэффициент вариации, % | 35.4 | 140.2 | 28.5 | 183.9 | 148.7 |

**Вывод:** показатели *GRPpc.2015*, *ECB.2014* и *NSB.2015* неоднородны, т.к. коэффициент вариации > 33%

## Анализ распределения данных

# Коробчатые диаграммы



**Вывод:** Можно заметить, что каждый показатель имеет выбросы (свои аномальные точки)

# Тесты на нормальность

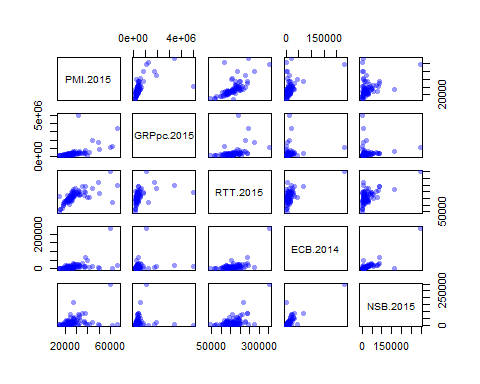
# Тест Андерсона-Дарлинга

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PMI.2015.A | GRPpc.2015.A | RTT.2015.A | ECB.2014.A | NSB.2015.A |
| Статистика теста | 4.7271 | 16.01 | 0.7056 | 14.6318 | 9.8188 |
| Критическое значение | 0.0000 | 0.00 | 0.0631 | 0.0000 | 0.0000 |

**Вывод:**Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все факторы,так как их p-значения < 0.05.

# Анализ линейных взаимосвязей

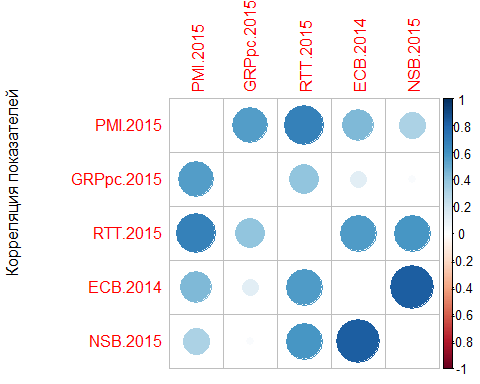
# Точечные графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *PMI.2015* и остальными показателями. У *PMI.2015* наблюдается положительная связь о всеми поазателями.

# Корреляционная матрица

# коэффициенты Пирсона с Р-значениями

 **Вывод:** Взаимосвязь *PMI.2015* с показателями ослабевает. Также наблюдается прямая связь и между самими показателями.

## Раздел II

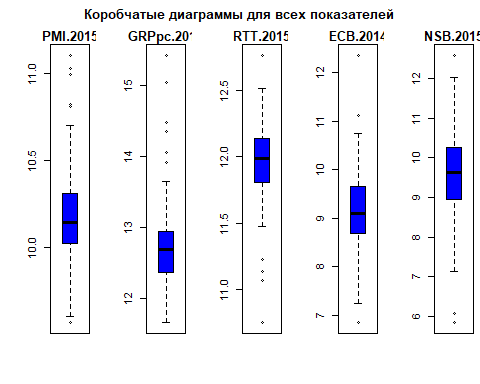
В этом разделе необходимо прологарифмировать данные из первого раздела и провести с новыми данными те же махинации. ## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PMI.2015 | GRPpc.2015 | RTT.2015 | ECB.2014 | NSB.2015 |
| Среднее | 10.179 | 12.755 | 11.953 | 9.087 | 9.559 |
| Стандартное отклонение | 0.301 | 0.655 | 0.309 | 0.906 | 1.187 |
| Коэффициент вариации, % | 2.957 | 5.135 | 2.585 | 9.970 | 12.418 |

**Вывод:** показатели однородны.

## Анализ распределения данных

#Построим коробчатые диаграммы.



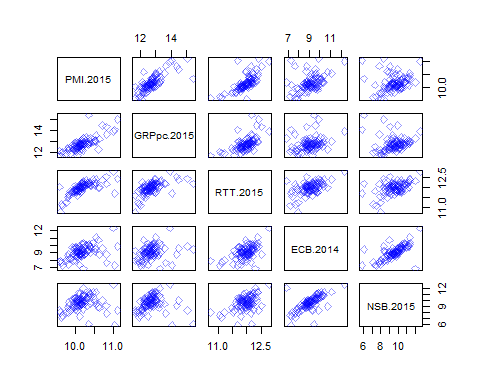
**Вывод:** Каждый признак имеет выбросы.

## Тест Андерсона-Дарлинга на нормальность распределения

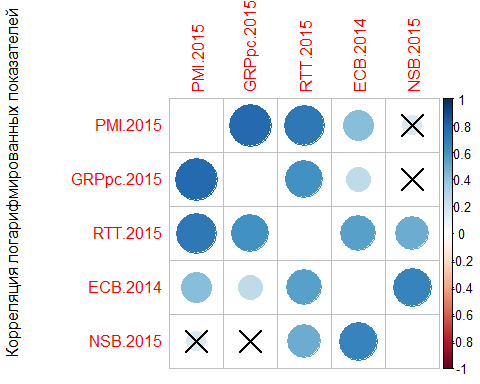
|  |  |  |
| --- | --- | --- |
|  | p | W |
| PMI.2015 | 0.0003 | 1.6707 |
| GRPpc.2015 | 0.0000 | 2.8409 |
| RTT.2015 | 0.0011 | 1.4164 |
| ECB.2014 | 0.1253 | 0.5828 |
| NSB.2015 | 0.1918 | 0.5101 |

**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“ECB.2014”* и *NSB.2015*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *PMI.2015* и остальными показателями. У *PMI.2015* положительная связь со всеми показателями, кроме *NSB.2015*.



**Вывод:** Взаимосвязь *PMI.2015* имеет взаимосвязь со всеми показателями, кроме *NBS.2015* и показателями ослабевает. Также наблюдается прямая связь и между самими показателями..