Лабораторная работа №1

Бессмертная М.М.

13 11 2020

# Раздел 1.

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

# **Необходимо построить:**

* Фактические плотности распределения переменных с наложенными теоретическими плотностями нормального закона (разместить графики на одном полотне).
* Тест Шапиро-Уилка на нормальность: таблица со статистикой теста, P-значением и выводом по каждой переменной.
* Точечные графики взаимного разброса, только над главной диагональю, сделать маркеры полупрозрачными.
* Взаимосвязь показателей: графическое представление корреляционной матрицы. Незначимые коэффициенты корреляции скрыть, остальные показать квадратами.

## Обозначение показателей:

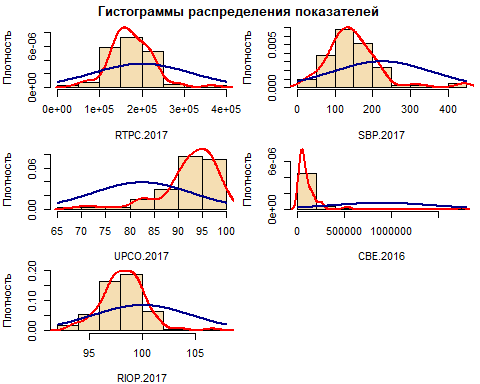
* *RTPC.2017* - Оборот розничной торговли на душу населения в 2017 году, рублей.
* *SBP.2017* - Число малых предприятий на 10000 человек населения в 2017 году, шт.
* *UPCO.2017* - Использование информационных и коммуникационных технологий в организациях: персональные компьютеры в 2017 году, в процентах от общего числа обследованных организаций.
* *CBE.2016* - Расходы консолидированных бюджетов субъектов Российской Федерации в 2016 году: всего, млн.руб.
* *RIOP.2017* - Реальные денежные доходы населения в 2017 году, в процентах к предыдущему году.

## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | RTPC.2017 | SBP.2017 | UPCO.2017 | CBE.2016 | RIOP.2017 |
| Среднее | 174108.9 | 147.5 | 92.4 | 116860.8 | 98.3 |
| Стандартное отклонение | 49348.9 | 72.7 | 5.8 | 200709.7 | 2.1 |
| Коэффициент вариации, % | 28.3 | 49.3 | 6.3 | 171.8 | 2.1 |

## Анализ распределния данных

Построим гистограммы с их фактическими плотностями распределений и наложим сверху теоретическую нормальную плотность распределения.



**Вывод:** При визуальной оценке графиков распределения показателей мы можем предположить, что переменные *‘RTPC.2017’*, *‘SBP.2017’*, *‘UPCO.2017’*, *‘RIOP.2017’*, *‘CBE.2016’* не имеют нормального распределения.

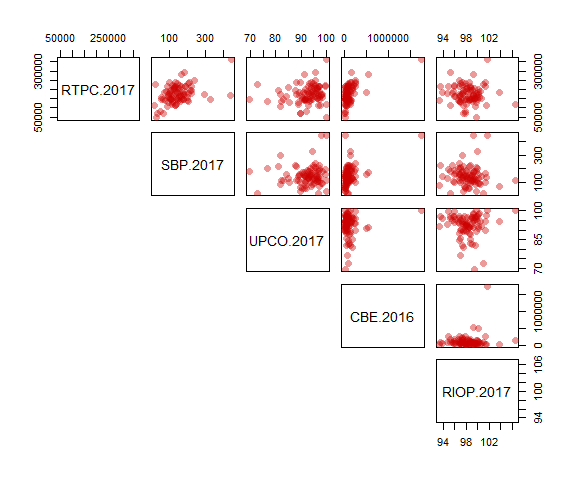
## Тест Шапиро-Уилка на нормальность

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | RTPC.2017 | SBP.2017 | UPCO.2017 | CBE.2016 | RIOP.2017 |
| Статистика | 0.96760 | 0.88177 | 0.86875 | 0.383 | 0.96279 |
| P-значение | 0.03096 | 0.00000 | 0.00000 | 0.000 | 0.01495 |

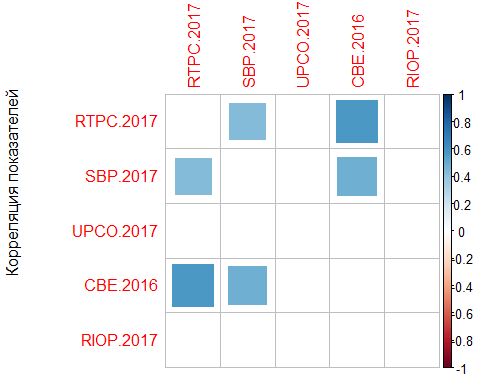
**Вывод:** По итогам теста Шапиро-Уилка на нормальность распредления переменных (при уровне значимости 0,05) отвергается нулевая гипотеза о нормальности распределения для всех пяти показателей *‘RTPC.2017’*, *‘SBP.2017’*, *‘UPCO.2017’*, *‘RIOP.2017’*, *‘CBE.2016’*, т.к. P-значения < 0.05. Следовательно, данные не имеют нормального распределения.

## Анализ взаимосвязей показателей

### Графики разброса



### Корреляционная матрица



**Вывод:** На основании корредяционной матрицы мы можем сделать вывод, что между показателями *‘RTCP.2017’* и *‘CBE.2016’* наблюдается сильная прямая корредяционная зависимость. Между показателями *‘CBE.2016’* и *‘SBP.2017’* наблюдается заметная прямая корреляционная зависимость. Между показателями *‘RTCP.2017’* и *‘SBP.2017’* прямая умеренная связь.

# Раздел 2.

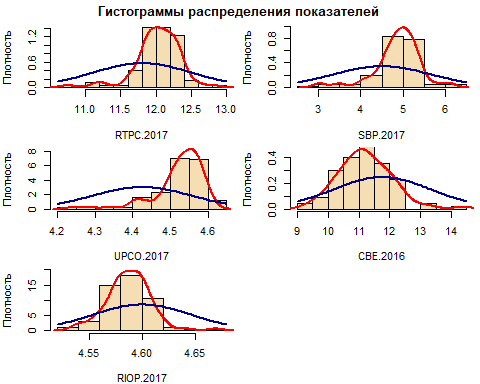
В этом разделе необходимо произвести теже самые действия, с логарифмированными исходными данными.

## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | RTPC.2017 | SBP.2017 | UPCO.2017 | CBE.2016 | RIOP.2017 |
| Среднее | 12.0 | 4.9 | 4.5 | 11.2 | 4.6 |
| Стандартное отклонение | 0.3 | 0.5 | 0.1 | 0.9 | 0.0 |
| Коэффициент вариации, % | 2.5 | 10.2 | 2.2 | 8.0 | 0.0 |

## Анализ распределния логарифмированных данных

Построим гистограммы логарифмированных данных с их фактическими плотностями распределений и наложим сверху теоретическую нормальную плотность распределения.



**Вывод:** При визуальном анализе графиков мы можем предположить, что нормированные данные по показателям *‘RIOP.2017’* и *‘CBE.2016’* имеют распределение, близкое к нормальному.

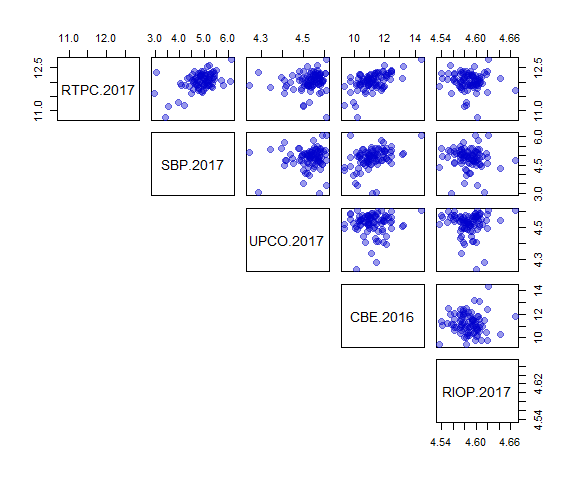
## Тест Шапиро-Уилка на нормальность для логарифмированных данных

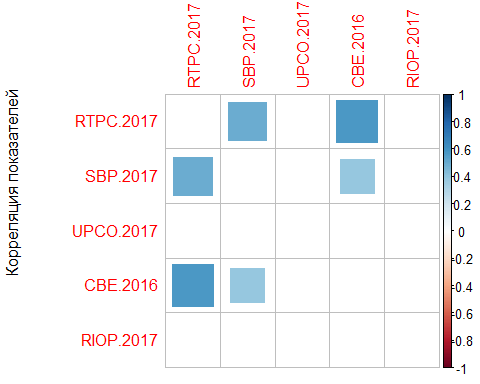
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | RTPC.2017 | SBP.2017 | UPCO.2017 | CBE.2016 | RIOP.2017 |
| Статистика | 0.92537 | 0.91695 | 0.83425 | 0.97658 | 0.96721 |
| P-значение | 0.00011 | 0.00004 | 0.00000 | 0.12449 | 0.02916 |

**Вывод:** По итогам теста Шапиро-Уилка на нормальность распредления логарифмированных переменных (при уровне значимости 0,05) отвергается нулевая гипотеза о нормальности распределения для четырех переменных *‘RTPC.2017’*, *‘SBP.2017’*, *‘UPCO.2017’*, *‘RIOP.2017’* т.к. P-значения < 0.05. А для переменной *‘CBE.2016’* нулевая гипотеза принимается. Следовательно, нормированные данные по показателю имеют нормальное распределение.

## Анализ взаимосвязей логарифмированных показателей

### Графики разброса логарифмированных величин





**Вывод:** На основании корредяционной матрицы мы можем сделать вывод, что между нормарованными показателями *‘RTCP.2017’* и *‘CBE.2016’* наблюдается сильная прямая корредяционная зависимость. Между показателями *‘CBE.2016’* и *‘SBP.2017’* наблюдается умеренная прямая корреляционная зависимость. Между показателями *‘RTCP.2017’* и *‘SBP.2017’* заметная прямая связь.