Лабораторная работа №1

Блинов И.И.

13 11 2020

# Раздел 1.

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

# **Необходимо построить:**

* Гистограммы распределения с наложенными теоретическими кривыми плотности нормального закона, разместить графики на одном полотне.
* Тест Андерсона-Дарлинга на нормальность: таблица со статистикой теста, критическим значением и выводом по каждой переменной.
* Точечные графики взаимного разброса (?pairs) только под главной диагональю, сделать маркеры полупрозрачными.
* Взаимосвязь показателей: графическое представление корреляционной матрицы с помощью функции corrplot(). Незначимые коэффициенты корреляции скрыть, остальные показать числами.

## Обозначение показателей:

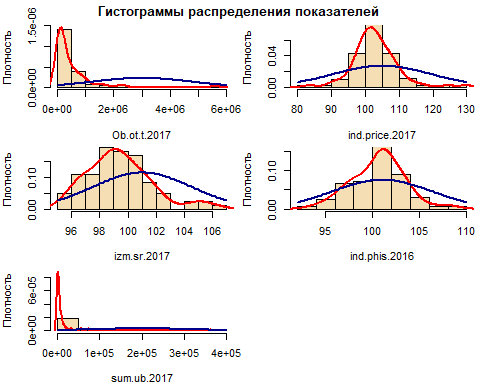
* *Ob.ot.t.2017* - Оборот розничной торговли на душу населения
* *ind.price.2017* - Число малых предприятий на 10000 человек населения
* *izm.sr.2017* - Использование информационных и коммуникационных технологий в организациях: персональные компьютеры
* *ind.phis.2016* - Расходы консолидированных бюджетов субъектов Российской Федерации: всего
* *sum.ub.2017* - Реальные денежные доходы населения

## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Ob.ot.t.2017 | ind.price.2017 | izm.sr.2017 | ind.phis.2016 | sum.ub.2017 |
| Среднее | 466588.5 | 103.0 | 99.4 | 100.6 | 14521.0 |
| Стандартное отклонение | 744975.0 | 6.8 | 2.3 | 2.8 | 45042.5 |
| Коэффициент вариации, % | 159.7 | 6.6 | 2.3 | 2.8 | 310.2 |

## Анализ распределния данных

Построим гистограмму



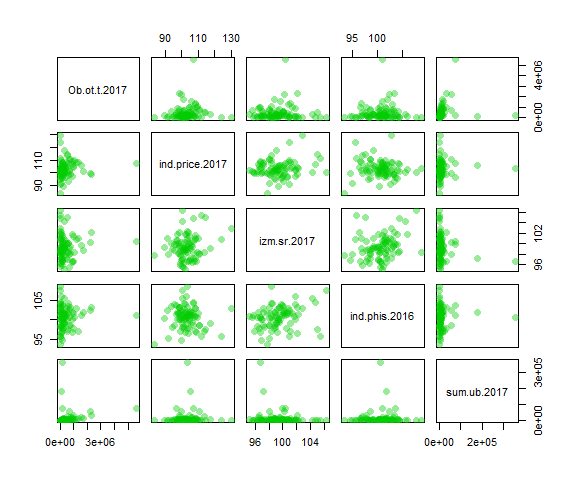
## Тест Андерсона-Дарлинга на нормальность

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Ob.ot.t.2017.A | ind.price.2017.A | izm.sr.2017.A | ind.phis.2016.A | sum.ub.2017.A |
| Статистика теста | 9.0534 | 1.2568 | 0.6950 | 0.4010 | 19.5527 |
| Критическое значение | 0.0000 | 0.0027 | 0.0671 | 0.3534 | 0.0000 |

**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“izm.sr.2017, ind.phis.2016”*, так как остальные p-значения < 0.05

## Анализ взаимосвязей показателей

### Графики разброса



### Корреляционная матрица



**Вывод:** Единственная значимая корреляция наблюдается между *‘izm.sr.2017’* и *‘ind.phis.2016’*.

# Раздел 2.

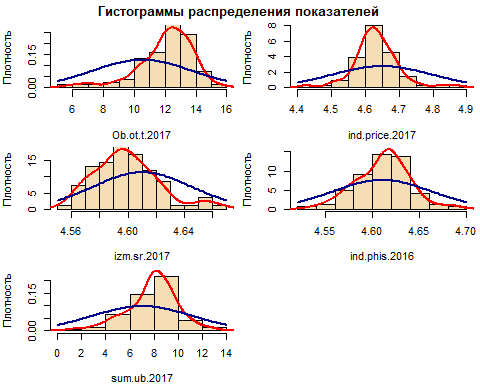
В этом разделе необходимо произвести теже самые действия, с логарифмированными исходными данными.

## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Ob.ot.t.2017 | ind.price.2017 | izm.sr.2017 | ind.phis.2016 | sum.ub.2017 |
| Среднее | 12.1 | 4.6 | 4.6 | 4.6 | 7.9 |
| Стандартное отклонение | 1.7 | 0.1 | 0.0 | 0.0 | 2.1 |
| Коэффициент вариации, % | 14.0 | 2.2 | 0.0 | 0.0 | 26.6 |

## Анализ распределния логарифмированных данных

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



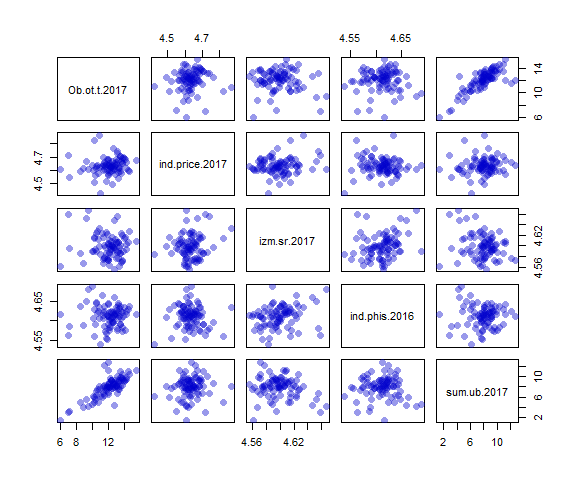
## Тест Андерсона-Дарлинга на нормальность для логарифмированных данных

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Ob.ot.t.2017.A | ind.price.2017.A | izm.sr.2017.A | ind.phis.2016.A | sum.ub.2017.A |
| Статистика теста | 1.4729 | 1.0213 | 0.5977 | 0.4394 | 0.9272 |
| Критическое значение | 0.0008 | 0.0103 | 0.1172 | 0.2860 | 0.0177 |

**Вывод:** По итогам теста Андерсона-Дарлинга на нормальность распредления логарифмированных переменных, при уровне значимости 0,05, отвергаются первые три переменные *‘Ob.ot.t.2017’*, *‘ind.price.2017’*, *‘sum.ub.2017’*, т.к. P-значения < 0.05. А переменные *‘ind.phis.2016’* и *‘izm.sr.2017’* не отвергаются.

## Анализ взаимосвязей логарифмированных показателей

### Графики разброса логарифмированных величин





**Вывод:** Наблюдается сильная корреляционная связь между показателями *‘Ob.ot.t.2017’* и *‘sum.ub.2017’*.