Лабораторная 1

Шевцова Я А

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

## Обозначение показателей:

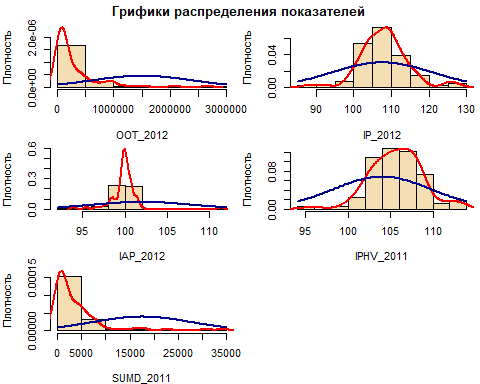
* *OOT\_2012* - Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по видам экономической деятельности: обрабатывающие производства
* *IP\_2012* - Индексы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства
* *IAP\_2012* - Изменение среднегодовой численности занятых
* *IPHV\_2011* -Индекс физического объема ВРП
* *SUMD\_2011* - Сумма убытка организаций по отдельным видам экономической деятельности: обрабатывающие производства

## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | OOT\_2012 | IP\_2012 | IAP\_2012 | IPHV\_2011 | SUMD\_2011 |
| Среднее | 308777.0 | 108.0 | 99.9 | 105.7 | 4127.9 |
| Стандартное отклонение | 453317.3 | 6.3 | 1.8 | 3.0 | 6170.3 |
| Коэффициент вариации, % | 146.8 | 5.8 | 1.8 | 2.8 | 149.5 |

## Анализ распределния данных

Построим график



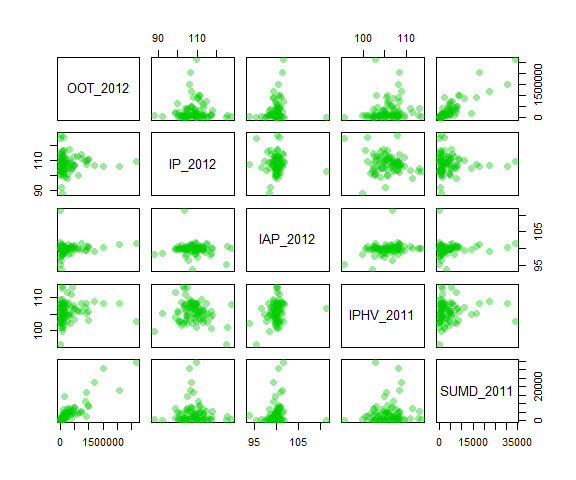
## Тест Андерсона-Дарлинга на нормальность

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | OOT\_2012.A | IP\_2012.A | IAP\_2012.A | IPHV\_2011.A | SUMD\_2011.A |
| Статистика теста | 8.7775 | 0.7847 | 6.1746 | 0.4025 | 8.8441 |
| p-значение | 0.0000 | 0.0401 | 0.0000 | 0.3503 | 0.0000 |

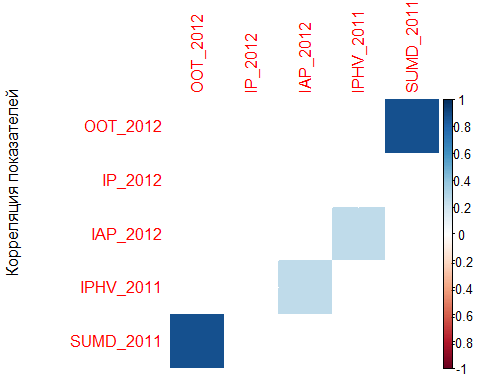
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все показатели кроме IPHV\_2011.

## Анализ взаимосвязей показателей

### Графики разброса



### Корреляционная матрица



**Вывод:** Сильная корреляция наблюдается между *‘OOT\_2012’* и *‘SUMD\_2011’* и значимая корреляция наблюдается между *‘IPHV\_2011’* и *‘IAP\_2012’*.

# Раздел 2.

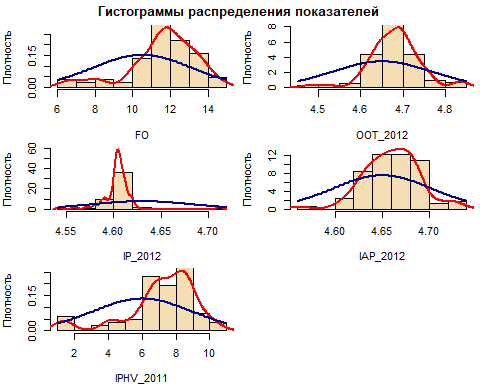
В этом разделе необходимо произвести те же самые действия, с логарифмированными исходными данными.

## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | OOT\_2012 | IP\_2012 | IAP\_2012 | IPHV\_2011 | SUMD\_2011 |
| Среднее | 11.6 | 4.7 | 4.6 | 4.7 | 7.2 |
| Стандартное отклонение | 1.8 | 0.1 | 0.0 | 0.0 | 2.1 |
| Коэффициент вариации, % | 15.5 | 2.1 | 0.0 | 0.0 | 29.2 |

## Анализ распределния логарифмированных данных

Строим фактическую плотность распределения переменных с наложенными теоретическими плотностями нормального закона



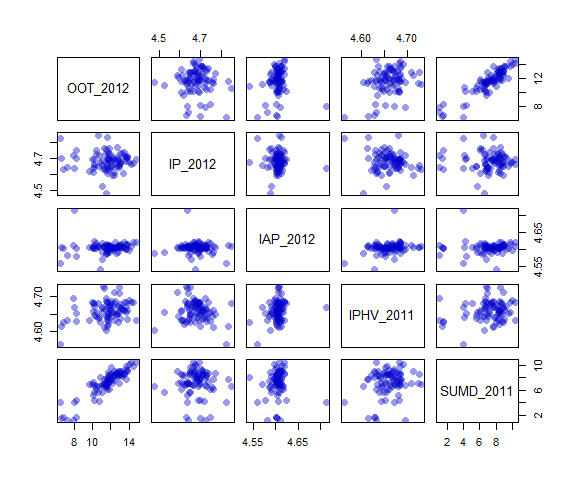
## Тест Андерсона-Дарлинга на нормальность для логарифмированных данных

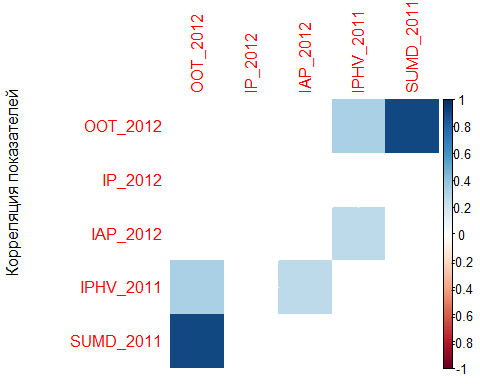
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | OOT\_2012.A | IP\_2012.A | IAP\_2012.A | IPHV\_2011.A | SUMD\_2011.A |
| Статистика теста | 1.8294 | 0.7142 | 5.9725 | 0.4034 | 2.7762 |
| p - значение | 0.0001 | 0.0601 | 0.0000 | 0.3487 | 0.0000 |

**Вывод:** По итогам теста Андерсона-Дарлинга на нормальность распредления логарифмированных переменных, при уровне значимости 0,05, отвергаются все переменные кроме *‘IP\_2012’* и *‘IPHV\_2011’*.

## Анализ взаимосвязей логарифмированных показателей

### Графики разброса логарифмированных величин





**Вывод:** Наблюдается сильная корреляционная связь между показателями *‘OOT\_2012’* и *‘SUMD\_2011’*, и значимая корреляция наблюдается между показателями *‘OOT\_2012’* и *‘IPHV\_2011’*, *‘IPHV\_2011’* и *‘IAP\_2012’*.