Лабораторная работа №1

Горбовский А.А.

30 11 2020

# Раздел I.

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), сделать маркеры полупрозрачными.
* Взаимосвязь показателей: графическое представление корреляционной матрицы с помощью функции corrplot(). Незначимые коэффициенты корреляции скрыть, остальные показать числами.

### Обозначения количественных показателей:

* *“Y.activity.2016.ps”* – Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по видам экономической деятельности: обрабатывающие производства
* *“INDEX.p.2016”* – Индексы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства numb.employees.2016 – Изменение среднегодовой численности занятых
* *“numb.employees.2016”* – Изменение среднегодовой численности занятых
* *“INDEX.VRP.2015”* – Индекс физического объема ВРП
* *“Sum.activity.2015”* – Сумма убытка организаций по отдельным видам экономической деятельности: обрабатывающие производства

# 1. Импорт данных

## Warning in `[.data.frame`(DF, as.numeric(row.names(DF)) < 1000, -1): в  
## результате преобразования созданы NA

Файл с данными содержит 96 строк и 7 столбцов

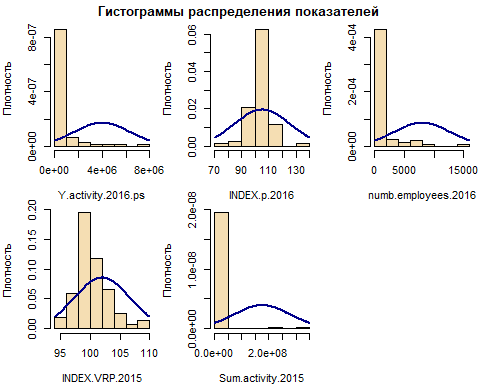
## 2. Описательные статистики

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.activity.2016.ps | INDEX.p.2016 | numb.employees.2016 | INDEX.VRP.2015 | Sum.activity.2015 |
| Среднее | 668790.9 | 103.4 | 1421.1 | 100.3 | 11987399.6 |
| Стандартное отклонение | 1224605.3 | 8.1 | 2242.8 | 2.8 | 40781506.5 |
| Коэффициент вариации, % | 183.1 | 7.8 | 157.8 | 2.8 | 340.2 |

**Вывод:** показатели не однородны, кроме *INDEX.p.2016*, *INDEX.VRP.2015*

## 3. Анализ распределения данных

# Гистограммы



**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“Y.activity.2016.ps”*, *“numb.employees.2016”*,*“Sum.activity.2015”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

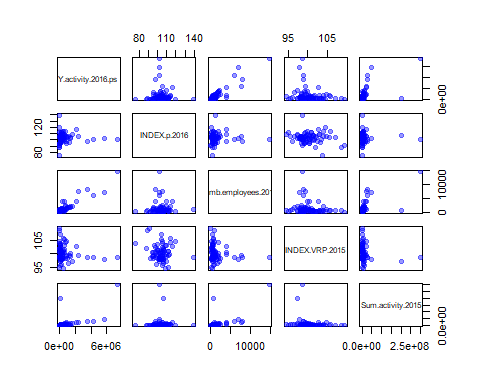
## Тесты на нормальность и крит.значения

|  |  |  |  |
| --- | --- | --- | --- |
|  | A | p | krit.zn |
| Y.activity.2016.ps.A | 12.3645 | 0e+00 | 4.7923 |
| INDEX.p.2016.A | 1.4620 | 8e-04 | 112.4443 |
| numb.employees.2016.A | 12.5423 | 0e+00 | 5.5603 |
| INDEX.VRP.2015.A | 1.2374 | 3e-03 | 312.6294 |
| Sum.activity.2015.A | 20.5894 | 0e+00 | 2.5793 |

**Вывод:**: По итогам теста Андерсона-Дарлинга на нормальность распредления переменных (при уровне значимости 0,05) отвергается нулевая гипотеза о нормальности распределения для трех показателей *“Y.activity.2016.ps”*, *“numb.employees.2016”*, *“Sum.activity.2015”*, т.к. P-значения < 0.05.

## 4. Анализ линейных взаимосвязей

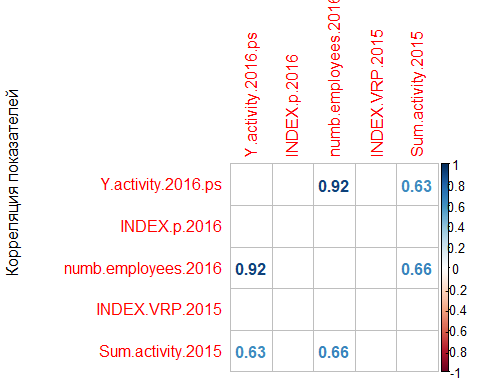
# Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *Y.activity.2016.ps* и остальными показателями. Скорее всего связь отсутствует со всеми показателями, но возможны слабые связи.

# Корреляционная матрица

# Визуализация корреляционной матрицы



**Вывод:** Взаимосвязь *Y.activity.2016.ps* значима с показателями *numb.employees.2016* и *Sum.activity.2015*.

# Раздел II

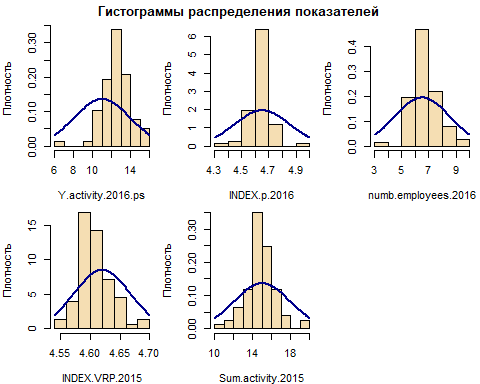
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.activity.2016.ps | INDEX.p.2016 | numb.employees.2016 | INDEX.VRP.2015 | Sum.activity.2015 |
| Среднее | 12.440 | 4.636 | 6.659 | 4.608 | 14.878 |
| Стандартное отклонение | 1.500 | 0.078 | 1.026 | 0.028 | 1.563 |
| Коэффициент вариации, % | 12.058 | 1.682 | 15.408 | 0.608 | 10.505 |

**Вывод:** *Y.activity.2016.ps*, *numb.employees*, *Sum.activity.2015* показатели не однородны

## Анализ распределения данных

# Гистограммы



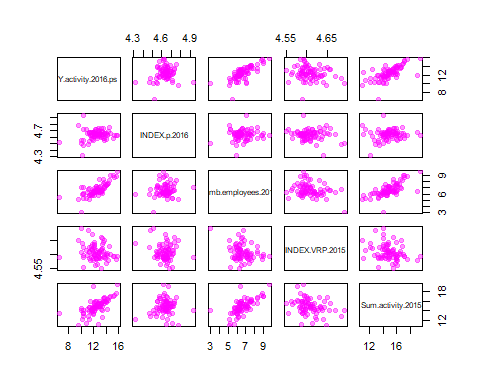
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом

## Тесты на нормальность и крит.значения

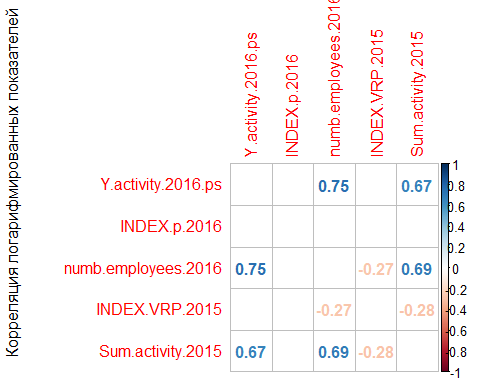
|  |  |  |  |
| --- | --- | --- | --- |
|  | A | p | krit.zn |
| Y.activity.2016.ps.A | 0.4188 | 0.3200 | 72.7818 |
| INDEX.p.2016.A | 1.5455 | 0.0005 | 519.7003 |
| numb.employees.2016.A | 1.4201 | 0.0011 | 56.9536 |
| INDEX.VRP.2015.A | 1.0924 | 0.0069 | 1456.2793 |
| Sum.activity.2015.A | 0.4824 | 0.2242 | 83.5456 |

**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *Y.activity.2016.ps* и *Sum.activity.2015* так как остальные p-значения < 0.05 .

# Графики разброса

 **Вывод:** по первой строке графиков попытаемся определить наличие связи между *Y.activity.2016.ps* и остальными показателями. Можно предположить, что есть положительная связь между *Y.activity.2016.ps* и *numb.employees.2016*, *Sum.activity.2015*.

# Корреляционная матрица



**Вывод:** Взаимосвязь *Y.activity.2016.ps* значима с показателями *numb.employees.2016* и *Sum.activity.2015*