Лабораторная работа № 1

Спиридонова А.Д.

13 11 2020

## Раздел I

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

**Необходимо построить:**

* Коробчатые диаграммы (совместить коробки для всех переменных на одном графике);
* 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 (?corrplot). Незначимые коэффициенты корреляции скрыть, остальные показать кругами.

### Обозначения количественных показателей:

* *IPI.2014* – Индексы промышленного производства.
* *PIM.2014* – Индесы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства.
* *DDFA.2014* – Степень износа основных фондов.
* *FCI.2013* – Инвестиции в основвной капитал на душу населения.
* *DLR.2014* – Задолжность по кредитам в рублях, предоставленым кредитными организациями юридическим лицам.

## Импорт данных:

Файл с данными содержит 92 строк и 8 столбцов.

## Описательная статистика

## IPI.2014 PIM.2014 DDFA.2014 FCI.2013 DLR.2014   
## 103.4 102.9 2694.7 116911.9 216425.7

## IPI.2014 PIM.2014 DDFA.2014 FCI.2013 DLR.2014   
## 7.3 8.2 10524.4 196695.4 744750.7

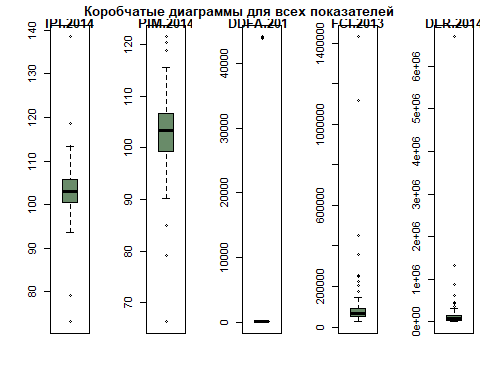
## IPI.2014 PIM.2014 DDFA.2014 FCI.2013 DLR.2014   
## 7.1 8.0 390.6 168.2 344.1

## IPI.2014 PIM.2014 DDFA.2014 FCI.2013 DLR.2014  
## Среднее 103.4 102.9 2694.7 116911.9 216425.7  
## Стандартное отклонение 7.3 8.2 10524.4 196695.4 744750.7  
## Коэффициент вариации, % 7.1 8.0 390.6 168.2 344.1

**Вывод:** показатели неоднородны.

## Анализ распределения данных

# Коробчатые диаграммы



**Вывод:** Можно заметить, что каждый показатель имеет выбросы

# Тесты на нормальность

# Тест Андерсона-Дарлинга

##   
## Anderson-Darling normality test  
##   
## data: reg.df$IPI.2014  
## A = 3.7509, p-value = 1.953e-09

## $IPI.2014  
##   
## Anderson-Darling normality test  
##   
## data: newX[, i]  
## A = 3.7509, p-value = 1.953e-09  
##   
##   
## $PIM.2014  
##   
## Anderson-Darling normality test  
##   
## data: newX[, i]  
## A = 1.3681, p-value = 0.001424  
##   
##   
## $DDFA.2014  
##   
## Anderson-Darling normality test  
##   
## data: newX[, i]  
## A = 29.355, p-value < 2.2e-16  
##   
##   
## $FCI.2013  
##   
## Anderson-Darling normality test  
##   
## data: newX[, i]  
## A = 17.662, p-value < 2.2e-16  
##   
##   
## $DLR.2014  
##   
## Anderson-Darling normality test  
##   
## data: newX[, i]  
## A = 21.273, p-value < 2.2e-16

## List of 4  
## $ statistic: Named num 3.75  
## ..- attr(\*, "names")= chr "A"  
## $ p.value : num 1.95e-09  
## $ method : chr "Anderson-Darling normality test"  
## $ data.name: chr "reg.df$IPI.2014"  
## - attr(\*, "class")= chr "htest"

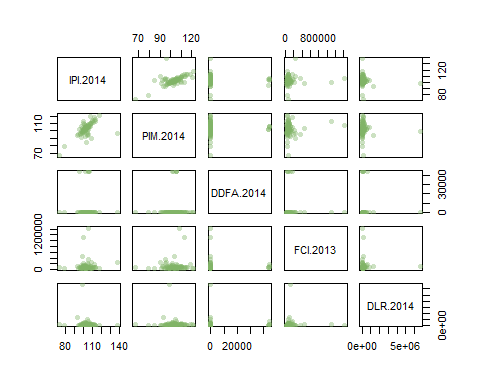
## IPI.2014 PIM.2014 DDFA.2014 FCI.2013 DLR.2014   
## 3.75 1.37 29.35 17.66 21.27

## IPI.2014.A PIM.2014.A DDFA.2014.A FCI.2013.A DLR.2014.A  
## Статистика теста 3.7509 1.3681 29.3548 17.6622 21.2725  
## Критическое значение 0.0000 0.0014 0.0000 0.0000 0.0000

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

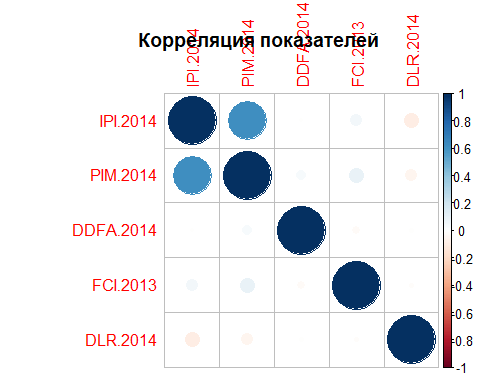
# Анализ линейных взаимосвязей

# Точечные графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *IPI.2014* и остальными показателями. У *IPI.2014* и *PIM.2014* положительная связь. А с остальными у *IPI.2014* связи нет

#4.2 Корреляционная матрица #коэффициенты Пирсона с Р-значениями

 **Вывод:** Взаимосвязь *IPI.2014* значима с показателем *PIM.2013* .

## Раздел II

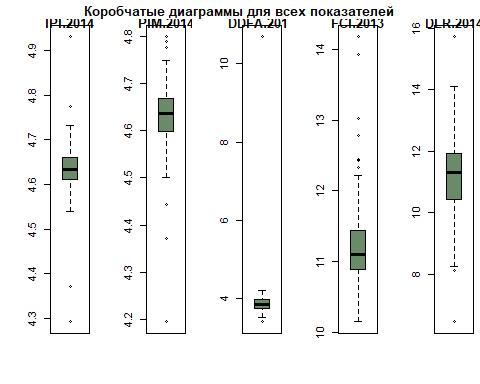
В этом разделе необходимо прологарифмировать данные из первого раздела и провести с новыми данными те же махинации. ## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | IPI.2014 | PIM.2014 | DDFA.2014 | FCI.2013 | DLR.2014 |
| Среднее | 4.636 | 4.630 | 4.234 | 11.281 | 11.213 |
| Стандартное отклонение | 0.072 | 0.086 | 1.653 | 0.693 | 1.368 |
| Коэффициент вариации, % | 1.553 | 1.857 | 39.041 | 6.143 | 12.200 |

**Вывод:** показатели считаются не полностью однородными, т.к. коэффициенты вариации не превышают 10%, кроме *DLR.2014*.

## Анализ распределения данных

Построим коробчатые диаграммы.



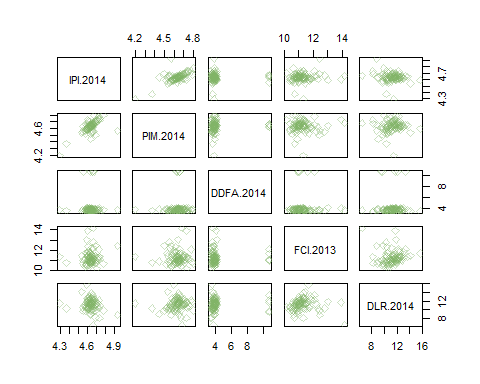
**Вывод:** Каждый признак имеет выбросы.

## Тест Андерсона-Дарлинга на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| IPI.2014 | 0.0000 | 4.0214 |
| PIM.2014 | 0.0000 | 2.0579 |
| DDFA.2014 | 0.0000 | 23.2005 |
| FCI.2013 | 0.0000 | 3.3580 |
| DLR.2014 | 0.0833 | 0.6573 |

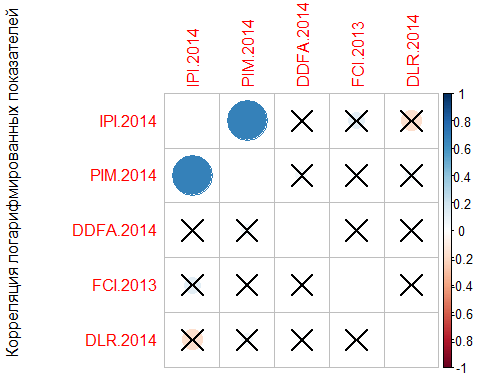
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“DDFA.2014”* и *“DLR.2014”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *IPI.2014* и остальными показателями. У *IPI.2014* и *PIM.2014* положительная связь. А с остальными у *IPI.2014* связи нет.

## Корреляционная матрица



**Вывод:** Взаимосвязь *IPI.2014* значима с показателем *PIM.2014*, и при этом она стала сильнее.