Лабораторная работа №1

Ларшина М.А.

21 12 2020

# Раздел I.

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

**Необходимо построить:**

* графики на одном полотне, отразив фактические и теоретические плотности распределения;
* Тест Андерсона-Дарлинга на нормальность: таблица со статистикой теста, критическим значением и выводом по каждой переменной;
* точечные графики взаимного разброса с полупрозрачными маркерами;
* графическое представление корреляционной матрицы.

### Обозначения количественных показателей:

* *Y.Srmonpop.2014* – оборот розничной торговли на душу населения.
* *VRPpop.2014* – число малых предприятий на 10000 человек населения.
* *ORTpop.2014* – использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* *RKBSsocial.2013* – расходы консолидированных бюджетов субъектов Российской Федерации: всего.
* *NumSmEnt.2014* – реальные денежные доходы населения.

## Импорт данных:

Файл с данными содержит 92 строк и 8 столбцов.

## Warning in Ops.factor(factor(DF$FO)): '-' not meaningful for factors

## [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## [51] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## [76] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA

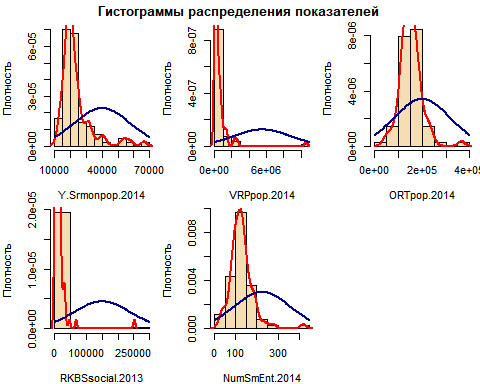
## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.Srmonpop.2014 | VRPpop.2014 | ORTpop.2014 | RKBSsocial.2013 | NumSmEnt.2014 |
| Среднее | 23592.9 | 601445.3 | 152739.9 | 15811.1 | 129.1 |
| Стандартное отклонение | 9862.7 | 1224463.6 | 46017.7 | 28515.6 | 55.4 |
| Коэффициент вариации, % | 41.8 | 203.6 | 30.1 | 180.4 | 42.9 |

**Вывод:** показатели неоднородны.

## Анализ распределения данных

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



**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“VRPpop.2014”*, *“ORTpop.2014”* и *“RKBSsocial.2013”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

# Тесты на нормальность

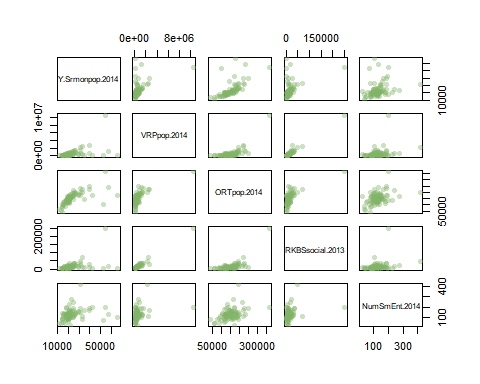
# Тест Андерсона-Дарлинга

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.Srmonpop.2014.A | VRPpop.2014.A | ORTpop.2014.A | RKBSsocial.2013.A | NumSmEnt.2014.A |
| Статистика теста | 6.5906 | 14.1895 | 0.8515 | 14.1254 | 1.8852 |
| Критическое значение | 0.0000 | 0.0000 | 0.0273 | 0.0000 | 0.0001 |

**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

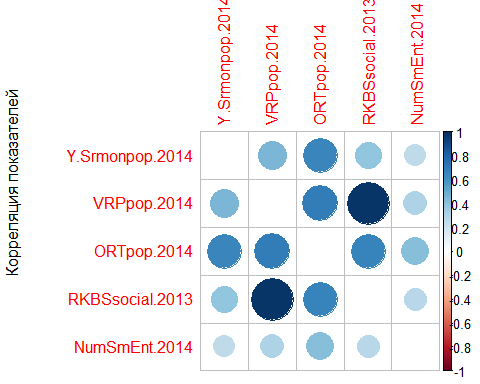
# Точечные графики разброса



**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

# 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.

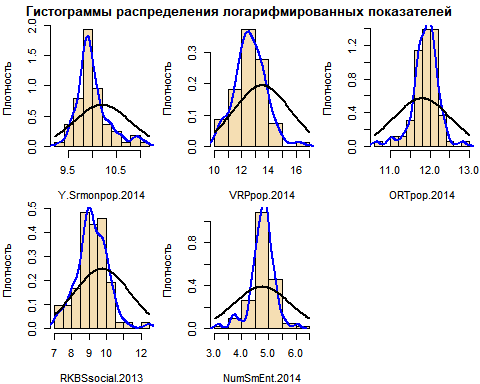
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.Srmonpop.2014 | VRPpop.2014 | ORTpop.2014 | RKBSsocial.2013 | NumSmEnt.2014 |
| Среднее | 10.007 | 12.640 | 11.890 | 9.193 | 4.775 |
| Стандартное отклонение | 0.332 | 1.091 | 0.318 | 0.895 | 0.434 |
| Коэффициент вариации, % | 3.318 | 8.631 | 2.675 | 9.736 | 9.089 |

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

## Анализ распределения данных

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



**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у “ORTpop.2014” и “Y.Srmonpop.2014”) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

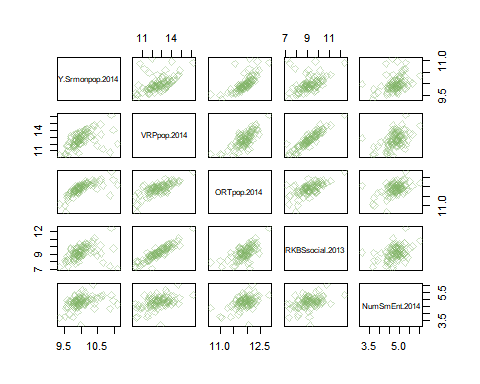
## Тест Андерсона-Дарлинга на нормальность распределения

## Y.Srmonpop.2014 VRPpop.2014 ORTpop.2014 RKBSsocial.2013 NumSmEnt.2014   
## FALSE FALSE FALSE FALSE FALSE

|  |  |  |
| --- | --- | --- |
|  | p | W |
| Y.Srmonpop.2014 | 0.0000 | 2.4951 |
| VRPpop.2014 | 0.0000 | 0.2136 |
| ORTpop.2014 | 0.0273 | 1.2604 |
| RKBSsocial.2013 | 0.0000 | 0.4777 |
| NumSmEnt.2014 | 0.0001 | 1.4809 |

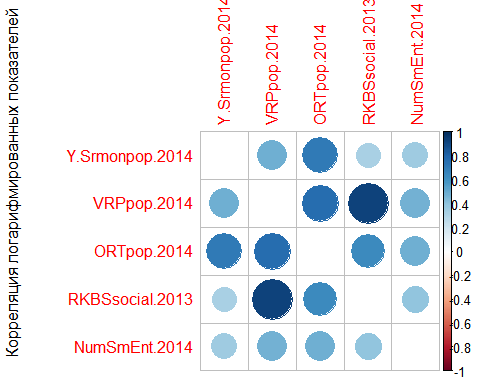
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“DDFA.2014”* и *“DLR.2014”*, так как остальные p-значения < 0.05 .

## 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.