Лабораторная работа №1

Полянцев И.С.

27 12 2020

# Раздел I.

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

**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.Wage”* – Среднедушевые денежные доходы населения
* *“X1.GRP”* – ВРП на душу населения
* *“X2.Retail”* – Оборот розничной торговли на душу населения
* *“X3.Social.Policy”* – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику на душу населения
* *“X4.Small.Businesses”* – Число малых предприятий на 10000 человек населения

# 1. Импорт данных

Файл с данными содержит 94 строк и 7 столбцов.

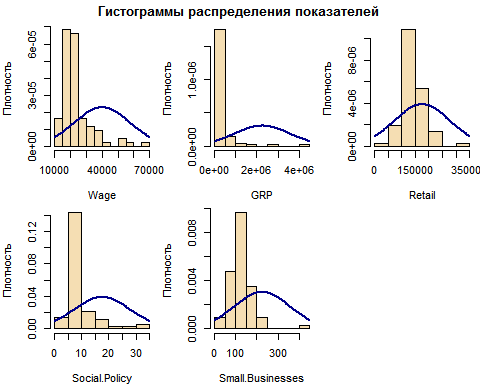
## 2. Описательные статистики

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Wage | GRP | Retail | Social.Policy | Small.Businesses |
| Среднее | 23774.9 | 401865.4 | 139471.0 | 9.1 | 129.9 |
| Стандартное отклонение | 9846.8 | 537241.2 | 42663.3 | 5.4 | 54.0 |
| Коэффициент вариации, % | 41.4 | 133.7 | 30.6 | 59.3 | 41.6 |

**Вывод:** Некоторые данные (*Wage*, *Retail*, *Small.businesses*) относительно однородны (Коэффициенты варианты не сильно превышают (либо не превышают) 33%), другие данные (*GRP*, *Social.Policy*) не однородны

## 3. Анализ распределения данных

# Гистограммы



**Вывод:**Распределения всех показателей, кроме *Retail*, ассиметричны по сравнению с нормальным законом. Кроме того, все показатели имеют более островершинные распределения по сравнению с нормальным законом.

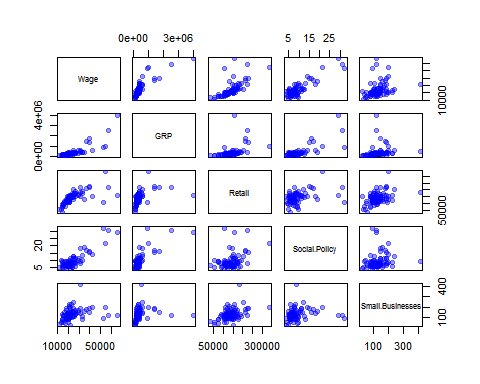
## Тест Андерсона-Дарлинга на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | W | p |
| Wage.A | 6.2791 | 0.0000 |
| GRP.A | 15.6777 | 0.0000 |
| Retail.A | 1.2547 | 0.0027 |
| Social.Policy.A | 8.1524 | 0.0000 |
| Small.Businesses.A | 1.3793 | 0.0013 |

**Вывод:**: По итогам теста Андерсона-Дарлинга на нормальность распредления переменных (при уровне значимости 0,05) отвергается нулевая гипотеза о нормальности распределения, т.к. P-значения < 0.05.

## 4. Анализ линейных взаимосвязей

# Графики разброса

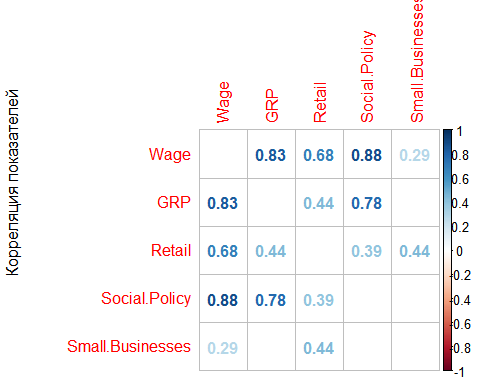


**Вывод:** Попытаемся сделать вывод о взаимосвязи *Wage* c остальными показателями. Судя по графикам, между показателем *Wage* и остальными показателями присутствует прямая линейная связь, но с различными степенями разброса данных (Наиболее сильная связь с показателем *Social.Policy*, наиболее слабая - с *Small.Businesses*).

# Корреляционная матрица

## Wage GRP Retail Social.Policy Small.Businesses  
## Wage 1.00 0.83 0.68 0.88 0.29  
## GRP 0.83 1.00 0.44 0.78 0.16  
## Retail 0.68 0.44 1.00 0.39 0.44  
## Social.Policy 0.88 0.78 0.39 1.00 0.16  
## Small.Businesses 0.29 0.16 0.44 0.16 1.00  
##   
## n= 85   
##   
##   
## P  
## Wage GRP Retail Social.Policy Small.Businesses  
## Wage 0.0000 0.0000 0.0000 0.0071   
## GRP 0.0000 0.0000 0.0000 0.1534   
## Retail 0.0000 0.0000 0.0002 0.0000   
## Social.Policy 0.0000 0.0000 0.0002 0.1514   
## Small.Businesses 0.0071 0.1534 0.0000 0.1514

# Визуализация корреляционной матрицы



**Вывод:** Взаимосвязь *Y.Wage* значима со всеми показателями и сильна со всеми, кроме *X4.Small.Businesses*

# Раздел II

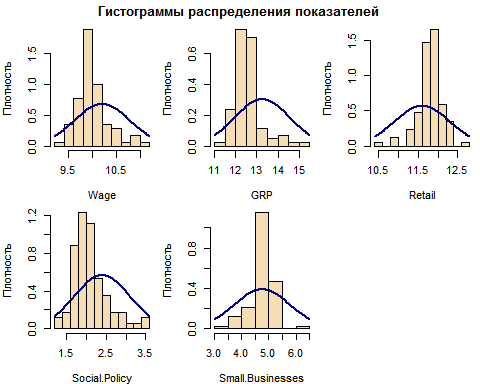
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Wage | GRP | Retail | Social.Policy | Small.Businesses |
| Среднее | 10.014 | 12.587 | 11.798 | 2.095 | 4.783 |
| Стандартное отклонение | 0.333 | 0.659 | 0.322 | 0.433 | 0.430 |
| Коэффициент вариации, % | 3.325 | 5.236 | 2.729 | 20.668 | 8.990 |

**Вывод:** Показатели *Social.Policy* и *Small.Businesses* не однородны

## Анализ распределения данных

# Гистограммы



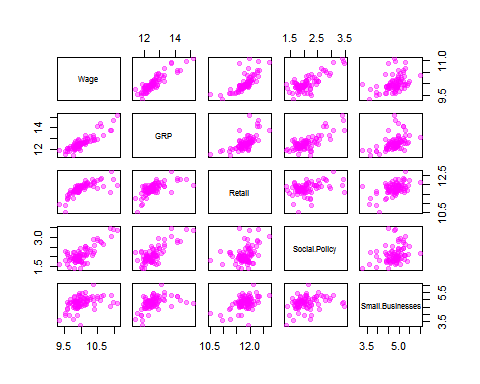
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным распределением

## Тест Андерсона-Дарлинга на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | W | p |
| Wage.A | 2.3118 | 0e+00 |
| GRP.A | 3.1401 | 0e+00 |
| Retail.A | 1.4902 | 7e-04 |
| Social.Policy.A | 2.4175 | 0e+00 |
| Small.Businesses.A | 1.4518 | 9e-04 |

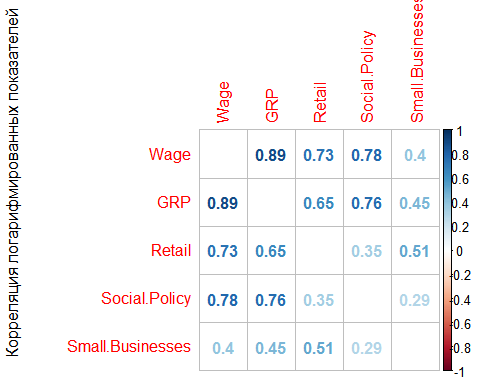
**Вывод:** Мы отвергаем гипотезу о нормальном распределении для всех показателей, т.к. p-значения < 0.05

# Графики разброса

 **Вывод:** По графикам видно, что показатель *Wage* имеет некоторую прямую линейную связь со всеми остальными показателями

# Корреляционная матрица

## Wage GRP Retail Social.Policy Small.Businesses  
## Wage 1.00 0.89 0.73 0.78 0.40  
## GRP 0.89 1.00 0.65 0.76 0.45  
## Retail 0.73 0.65 1.00 0.35 0.51  
## Social.Policy 0.78 0.76 0.35 1.00 0.29  
## Small.Businesses 0.40 0.45 0.51 0.29 1.00  
##   
## n= 85   
##   
##   
## P  
## Wage GRP Retail Social.Policy Small.Businesses  
## Wage 0.0000 0.0000 0.0000 0.0001   
## GRP 0.0000 0.0000 0.0000 0.0000   
## Retail 0.0000 0.0000 0.0011 0.0000   
## Social.Policy 0.0000 0.0000 0.0011 0.0067   
## Small.Businesses 0.0001 0.0000 0.0000 0.0067



**Вывод:** Взаимосвязь *Wage* значима со всеми показателями