Лабораторная работа №1

Самышко П.О.

11 Ноябрь 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;
* таблицу со статистикой теста Шапиро-Уилка на нормальность;
* точечные графики взаимного разброса с полупрозрачными маркерами;
* графическое представление корреляционной матрицы.

### Обозначения количественных показателей:

* *Y.ORTorg.2013* – оборот розничной торговли на душу населения.
* *KMPred.2013* – число малых предприятий на 10000 человек населения.
* *Pers.comp.2013* – использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* *RKBS.2012* – расходы консолидированных бюджетов субъектов Российской Федерации: всего.
* *RDDN.2013* – реальные денежные доходы населения.

## Импорт данных:

Файл с данными содержит 92 строк и 8 столбцов.

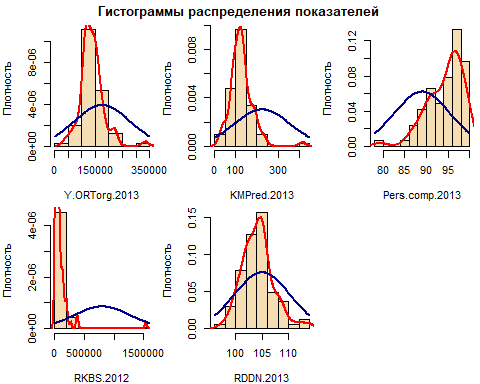
## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.ORTorg.2013 | KMPred.2013 | Pers.comp.2013 | RKBS.2012 | RDDN.2013 |
| Среднее | 138420.6 | 129.1 | 94.5 | 100480.9 | 104.2 |
| Стандартное отклонение | 42390.2 | 54.3 | 4.0 | 175995.0 | 3.0 |
| Коэффициент вариации, % | 30.6 | 42.1 | 4.2 | 175.2 | 2.9 |

**Вывод:** показатели неоднородны.

## Анализ распределения данных

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



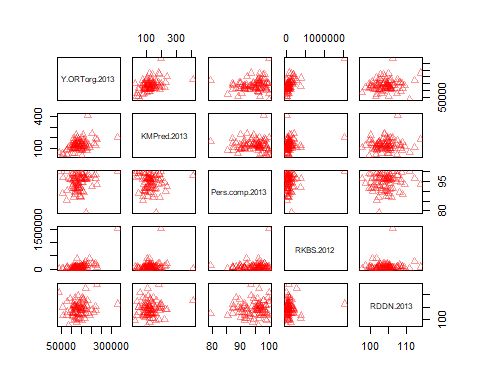
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“KMPred.2013”*, *“Pers.comp.2013”* и *“RKBS.2012”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Шапиро-Уилка на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| Y.ORTorg.2013 | 0.0001 | 0.9239 |
| KMPred.2013 | 0.0000 | 0.8823 |
| Pers.comp.2013 | 0.0002 | 0.9305 |
| RKBS.2012 | 0.0000 | 0.3707 |
| RDDN.2013 | 0.1013 | 0.9747 |

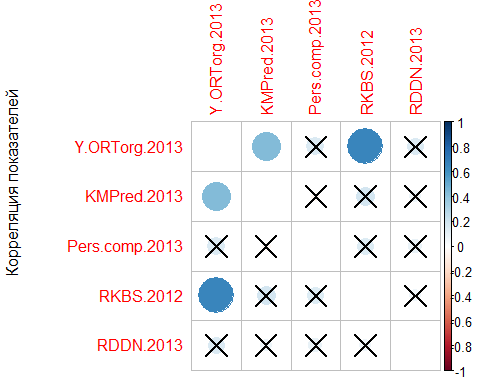
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“RDDN.2013”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *Y.ORTorg.2013* и остальными показателями. Вероятно, что связь отсутствует со следующими показателями: *KMPred.2013* и *RKBS.2012*. С *Pers.comp.2013* и *RDDN.2013* возможны слабые положительные линейные связи. Чтобы проверить эти предположения, рассчитаем корреляционную матрицу и проверим коэффициенты на значимость.

## Корреляционная матрица



**Вывод:** Взаимосвязь *Y.ORTorg.2013* значима с показателями *KMPred.2013* и *RKBS.2012*.

# Раздел II.

В этом разделе необходимо прологарифмировать данные из первого раздела и провести с новыми данными те же махинации.

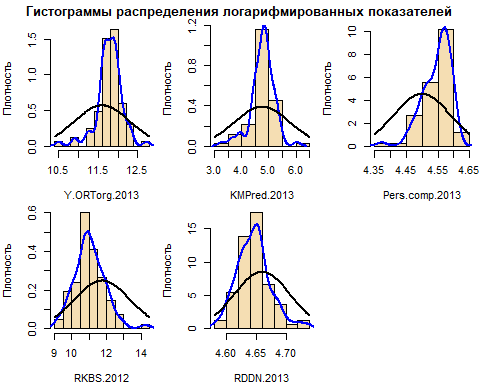
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.ORTorg.2013 | KMPred.2013 | Pers.comp.2013 | RKBS.2012 | RDDN.2013 |
| Среднее | 11.791 | 4.775 | 4.547 | 11.050 | 4.646 |
| Стандартное отклонение | 0.321 | 0.433 | 0.044 | 0.872 | 0.028 |
| Коэффициент вариации, % | 2.722 | 9.068 | 0.968 | 7.891 | 0.603 |

**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.



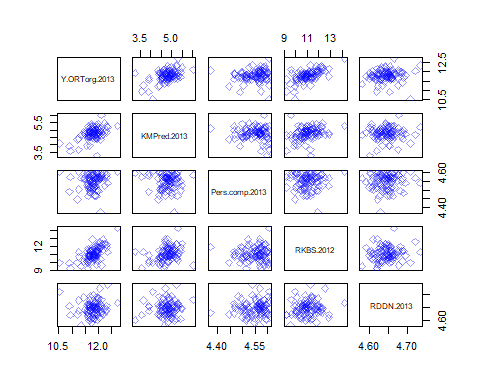
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у “Pers.comp.2013” и “Y.ORTorg.2013”) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Шапиро-Уилка на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| Y.ORTorg.2013 | 0.0002 | 0.9303 |
| KMPred.2013 | 0.0010 | 0.9425 |
| Pers.comp.2013 | 0.0001 | 0.9173 |
| RKBS.2012 | 0.1657 | 0.9780 |
| RDDN.2013 | 0.2347 | 0.9803 |

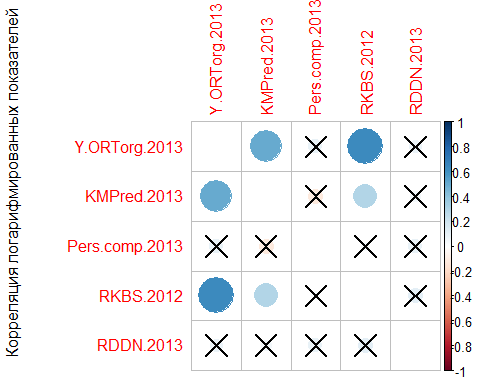
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“RDDN.2013”* и *“RKBS.2012”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между Y.ORTorg.2013 и остальными показателями. Можно предположить, что со всеми показателями существует слабая положительная связь. Либо нелинейная. Чтобы проверить эти предположения, рассчитаем корреляционную матрицу и проверим коэффициенты на значимость.

## Корреляционная матрица



**Вывод:** Взаимосвязь *Y.ORTorg.2013* значима с показателями *KMPred.2013* и *RKBS.2012*, и при этом она стала сильнее. Однако также появилась значимая связь между самими показателями.