Лабораторная работа №1

Трапезникова А.А.

13 11 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 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.2012* – оборот розничной торговли на душу населения.
* *KMPred.2012* – число малых предприятий на 10000 человек населения.
* *Pers.comp.2012* – использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* *RKBS.2011* – расходы консолидированных бюджетов субъектов Российской Федерации: всего.
* *RDDN.2012* – реальные денежные доходы населения.

## Импорт данных:

Файл с данными содержит 92 строк и 8 столбцов.

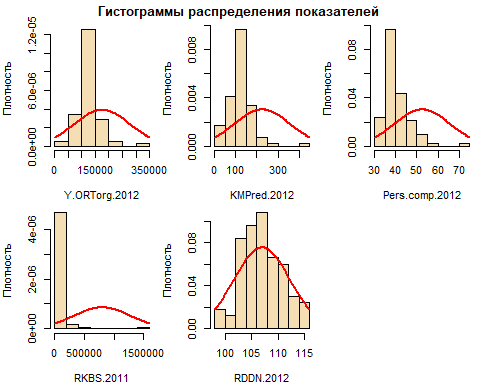
## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.ORTorg.2012 | KMPred.2012 | Pers.comp.2012 | RKBS.2011 | RDDN.2012 |
| Среднее | 125039.9 | 123.6 | 40.8 | 92480.5 | 107.1 |
| Стандартное отклонение | 38604.3 | 54.6 | 6.3 | 160879.5 | 3.7 |
| Коэффициент вариации, % | 30.9 | 44.2 | 15.4 | 174.0 | 3.5 |

**Вывод:** показатели неоднородны.

## Анализ распределения данных

Построим гистограммы с теоретической нормальной плотностью распределения.



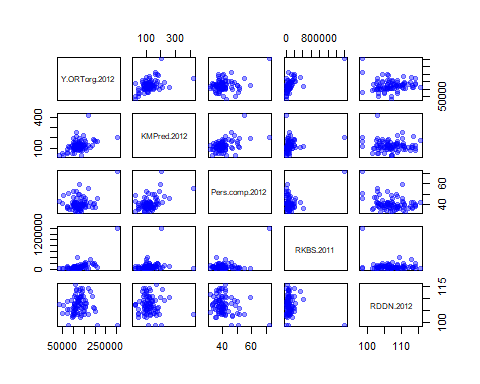
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“Y.ORTorg.2012”*, *“KMPred.2012”*, *“Pers.comp.2012”* и *“RKBS.2011”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Шапиро-Уилка на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| Y.ORTorg.2012 | 0.0000 | 0.9153 |
| KMPred.2012 | 0.0000 | 0.8588 |
| Pers.comp.2012 | 0.0000 | 0.8337 |
| RKBS.2011 | 0.0000 | 0.3759 |
| RDDN.2012 | 0.4715 | 0.9854 |

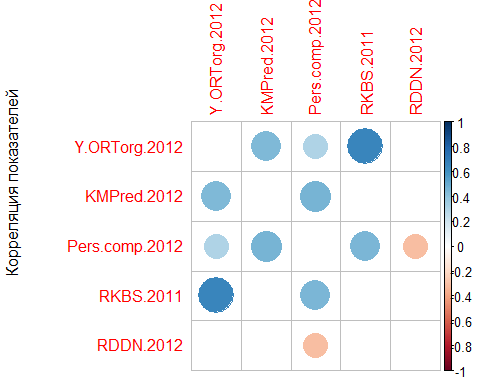
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“RDDN.2012”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *Y.ORTorg.2012* и остальными показателями. Вероятно, что связь отсутствует со следующими показателями: *KMPred.2012* и *Pers.comp.2012*. С *RKBS.2011* и *RDDN.2012* возможны слабые положительные линейные связи. Чтобы проверить эти предположения, рассчитаем корреляционную матрицу и проверим коэффициенты на значимость.

## Корреляционная матрица



**Вывод:** Взаимосвязь *Y.ORTorg.2012* значима с показателями *KMPred.2012*, *Pers.comp.2012* и *RKBS.2011*. Между этими показателями присутствует значимая связь, кроме как у *RKBS.201*1 и *KMPred.2012*. У *RDDN.2012* и *Pers.comp.2012* обнаружена отрицательная связь.

# Раздел II.

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

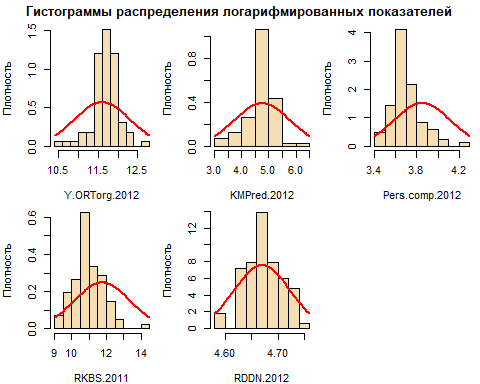
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y.ORTorg.2012 | KMPred.2012 | Pers.comp.2012 | RKBS.2011 | RDDN.2012 |
| Среднее | 11.689 | 4.725 | 3.698 | 10.965 | 4.673 |
| Стандартное отклонение | 0.321 | 0.451 | 0.138 | 0.875 | 0.035 |
| Коэффициент вариации, % | 2.746 | 9.545 | 3.732 | 7.980 | 0.749 |

**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.



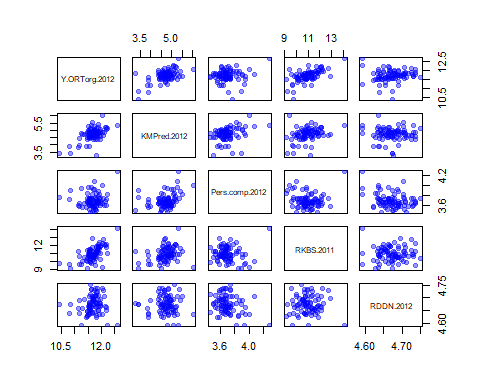
**Вывод:** Судя по гистограммам, распределения показателей *Pers.comp.2012* и *RKBS.2011* ассиметричны по сравнению с нормальным законом. Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Шапиро-Уилка на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| Y.ORTorg.2012 | 0.0002 | 0.9282 |
| KMPred.2012 | 0.0002 | 0.9282 |
| Pers.comp.2012 | 0.0000 | 0.9105 |
| RKBS.2011 | 0.1560 | 0.9776 |
| RDDN.2012 | 0.4957 | 0.9858 |

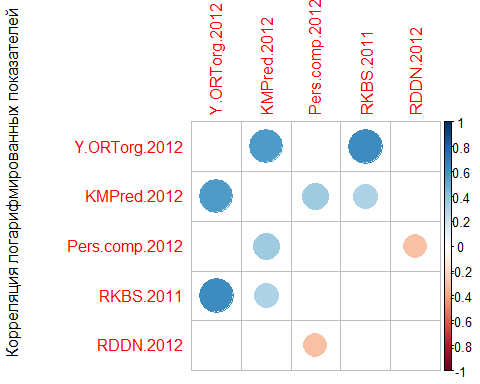
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“RDDN.2012”* и *“RKBS.2011”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *Y.ORTorg.2012* 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.2012* и *RKBS.2011*, и при этом она стала сильнее.