Лабораторная работа

Степанов Г.А.

16 11 2020

# Раздел I.

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

**Необходимо построить:**

* Фактические плотности распределения переменных с наложенными теоретическими плотностями нормального закона (разместить графики на одном полотне).
* Тест Андерсона-Дарлинга на нормальность: таблица со статистикой теста, P-значением и выводом по каждой переменной.
* Точечные графики взаимного разброса, только над главной диагональю, сделать маркеры полупрозрачными.
* Взаимосвязь показателей: графическое представление корреляционной матрицы. Незначимые коэффициенты корреляции скрыть, остальные показать квадратами.

### Обозначения количественных показателей:

* SREDDENDOXOD2012 – Среднедушевые денежные доходы населения.
* VRP2012 – ВРП на душу населения.
* Y.ORTorg.2012 –Оборот розничной торговли на душу населения.
* RKBS.2011 – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику.
* KMPred.2012 – Число малых предприятий на 10000 человек населения.

## Импорт данных:

Файл с данными содержит 92 строк и 8 столбцов.

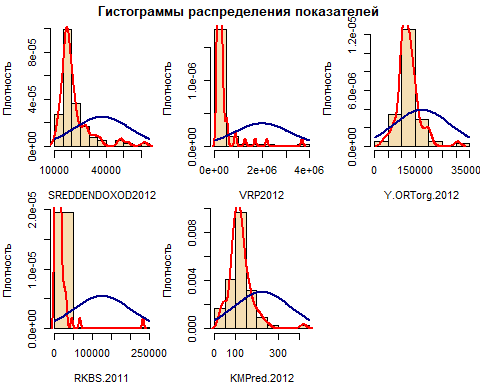
## Описательная статистика

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SREDDENDOXOD2012 | VRP2012 | Y.ORTorg.2012 | RKBS.2011 | KMPred.2012 |
| Среднее | 21244.4 | 360615.1 | 125039.9 | 14355.2 | 124.2 |
| Стандартное отклонение | 9033.3 | 488439.3 | 38604.3 | 26401.5 | 55.5 |
| Коэффициент вариации, % | 42.5 | 135.4 | 30.9 | 183.9 | 44.7 |

**Вывод:** показатели неоднородны.

## Анализ распределения данных

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



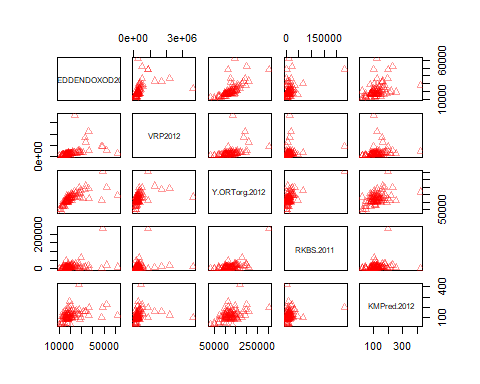
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“SREDDENDOXOD2012”*, *“VRP2012”* и *“RKBS.2011”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Андерсона-Дарлинга на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| SREDDENDOXOD2012 | 0e+00 | 6.4446 |
| VRP2012 | 0e+00 | 15.5629 |
| Y.ORTorg.2012 | 5e-04 | 1.5443 |
| RKBS.2011 | 0e+00 | 14.6318 |
| KMPred.2012 | 0e+00 | 2.0262 |

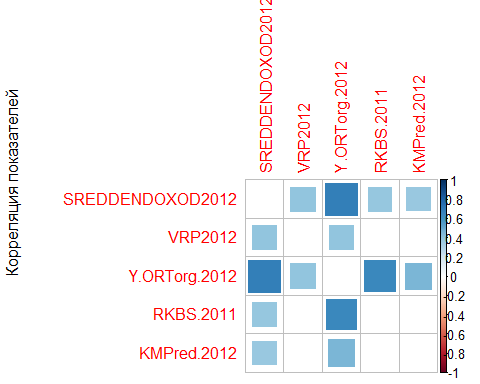
**Вывод:**: По итогам теста Андерсона-Дарлинга на нормальность распредления переменных (при уровне значимости 0,05) отвергается нулевая гипотеза о нормальности распределения для всех пяти показателей *“SREDDENDOXOD2012”*, *“VRP2012”*, *“RKBS.2011”*, *“KMPred.2012”*, *“Y.ORTorg.2012”*, т.к. P-значения < 0.05. Следовательно, данные не имеют нормального распределения.

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *“SREDDENDOXOD2012”* и остальными показателями. Вероятно, что связь сильная есть со следующими показателями: *“Y.ORTorg.2012”*. С *“KMPred.2012”*, *“VRP2012”*, *“RKBS.2011”* возможны средние положительные линейные связи. Чтобы проверить эти предположения, рассчитаем корреляционную матрицу и проверим коэффициенты на значимость.

## Корреляционная матрица



**Вывод:** На основании корредяционной матрицы мы можем сделать вывод, что между показателями *“SREDDENDOXOD2012”* и *“Y.ORTorg.2012”* наблюдается сильная прямая корредяционная зависимость, также как между *“Y.ORTorg.2012”* и *“RKBS.2011”*. Между показателями *“SREDDENDOXOD2012”* и *“VRP2012”* наблюдается заметная прямая корреляционная зависимость, также как между *“Y.ORTorg.2012”* и *“KMPred.2012”*. Между показателями *“SREDDENDOXOD2012”* и *“RKBS.2011”;* *“SREDDENDOXOD2012”* и *“KMPred.2012”;* *“VRP2012”* и *“Y.ORTorg.2012”* прямая умеренная связь.

# Раздел II.

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

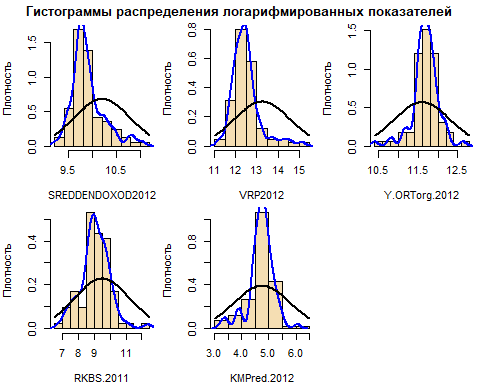
## Описательная статистика логарифмированных данных:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SREDDENDOXOD2012 | VRP2012 | Y.ORTorg.2012 | RKBS.2011 | KMPred.2012 |
| Среднее | 9.900 | 12.483 | 11.689 | 9.087 | 4.728 |
| Стандартное отклонение | 0.337 | 0.653 | 0.321 | 0.906 | 0.455 |
| Коэффициент вариации, % | 3.404 | 5.231 | 2.746 | 9.970 | 9.624 |

**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.



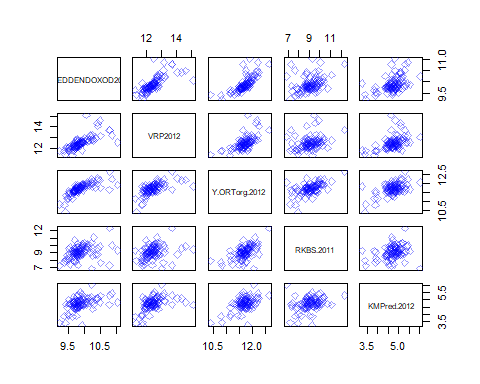
**Вывод:** Судя по гистограммам, распределения всех показателей ассиметричны по сравнению с нормальным законом.(Особенно это заметно у *“KMPred.2012”* и *“Y.ORTorg.2012”*) Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

## Тест Андерсона-Дарлинга на нормальность распределения

|  |  |  |
| --- | --- | --- |
|  | p | W |
| SREDDENDOXOD2012 | 0.0000 | 2.4520 |
| VRP2012 | 0.0000 | 2.7098 |
| Y.ORTorg.2012 | 0.0004 | 1.6123 |
| RKBS.2011 | 0.1253 | 0.5828 |
| KMPred.2012 | 0.0001 | 1.8498 |

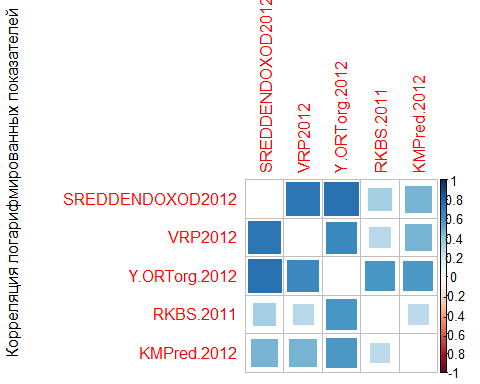
**Вывод:** Проверяя гипотезу нормального распределения для переменных на уровне значимости 0.05, мы отвергаем все, кроме *“RKBS.2011”*, так как остальные p-значения < 0.05 .

## Графики разброса



**Вывод:** по первой строке графиков попытаемся определить наличие связи между *“SREDDENDOXOD2012”* и остальными показателями. Можно предположить, что со всеми показателями существует прямая положительная связь,чтобы проверить эти предположения, рассчитаем корреляционную матрицу и проверим коэффициенты на значимость.

## Корреляционная матрица



**Вывод:** Сильная взаимосвязь *“SREDDENDOXOD2012”* значима с показателями *“VRP2012”* и *“Y.ORTorg.2012”*.Между другими показателями заметно увеличилась связь.