Лабораторная работа 1

Зраева Екатерина

01 12 2020

Показатели:

* y - ВРП 2015
* x1 - Инвестиции в основной капитал 2014
* x2 - Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику 2014
* x3 - Использование информационных и коммуникационных технологий в организациях: персональные компьютеры 2014
* x4 - Внутренние затраты на научные исследования и разработки 2014

Источник данных: <https://www.gks.ru/folder/210/document/13204>

Исходные данные для работы хранятся в файле Data\_Zraeva.csv.

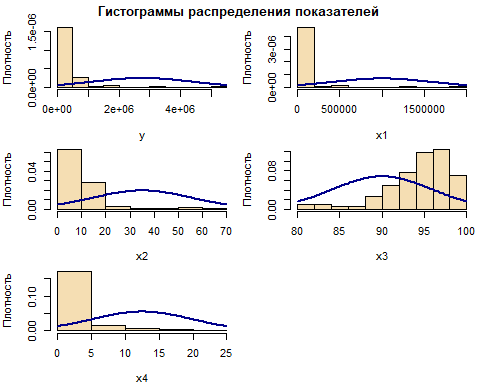
Дата сбора данных: 30.11.20

Количество наблюдений:

В исходной таблице количество строк, столбцов - 96, 8 соответственно. Оставим исходную таблицу без изменений, создав фрейм «reg.df», в который войдут только нужные нам строки и столбцы, а также фрейм «reglog.df», в который войдут логарифмированные данные.

Проведем анализ собранных данных:

Посмотрим, насколько нормально распределение исходных показателей. Для этого построим гистограммы плотности распределения и проведём статистический тест Шапиро-Уилка на нормальность.

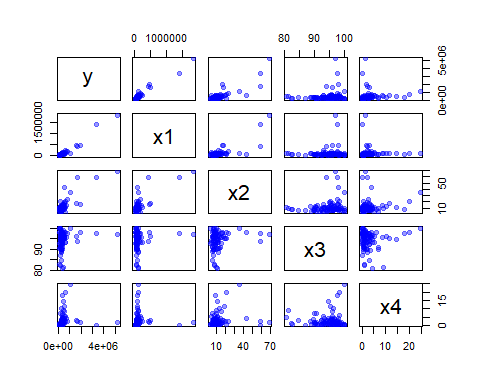


|  |  |
| --- | --- |
|  | peremennaia |
| y.W | 0.41 |
| x1.W | 0.31 |
| x2.W | 0.55 |
| x3.W | 0.90 |
| x4.W | 0.63 |

Все переменные не подчиняются нормальному распределению. Судя по гистограммам, распределения всех показателей сильно скошены (асимметричны) по сравнению с нормальным законом. Кроме того, показатели имеют более островершинные распределения по сравнению с нормальным законом.

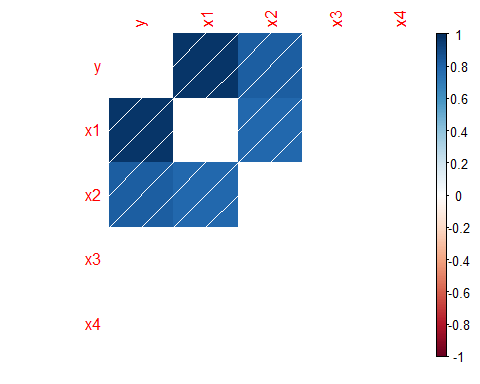
Протестируем показатели на нормальность. Нулевая гипотеза теста Шапиро-Уилка: распределение соответствует нормальному закону. Альтернативная: распределение ненормально. Проверим гипотезу для переменной y на уровне значимости 0,05. p.value=1.83e-17<0,05, следовательно, нулевая гипотеза отвергается: распределение ненормально.

Для выявления взаимосвязей между парами показателей удобно использовать графики взаимного разброса.



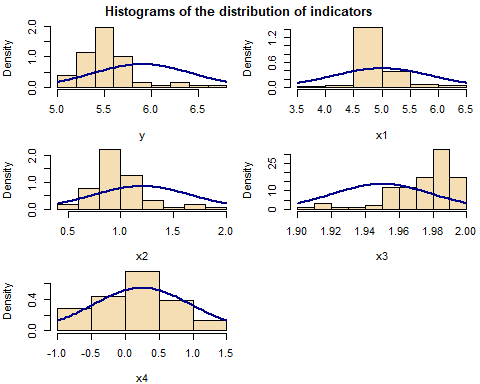
Линейная положительная корреляция заметна между переменными y(ВРП) и x1(Инвестиции в основной капитал), y(ВРП) и x2(Расходы на национальную экономику)

Чтобы подтвердить или опровергнуть эти предположения, рассчитаем корреляционную матрицу и проверим на значимость коэффициенты корреляции.



Сильная линейная связь между всеми парами перменных: У, х1, х2, между остальными парами связь незначима, поэтому на графике не отображается.

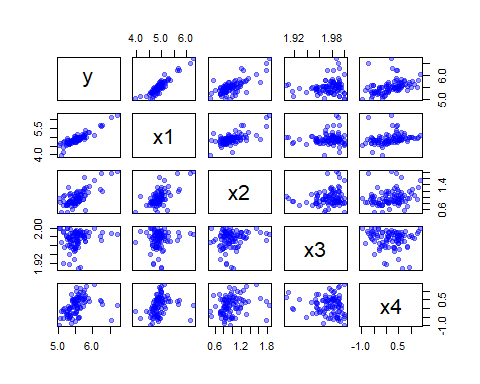
Проведем анализ логарифмированных данных:

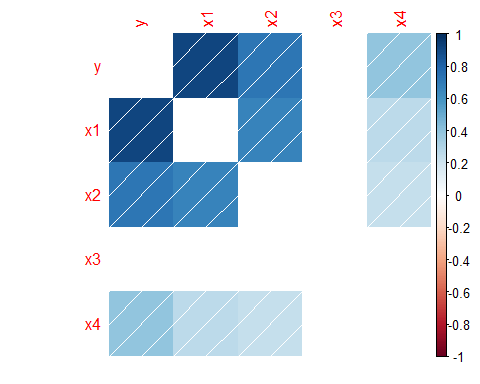


Переменная х4 стала наиболее выражено подчиняться нормальному закону распределения, другие показатели также стали чуточку ближе к нормальному распределению, но они сильно сдвинуты.

|  |  |
| --- | --- |
|  | peremennaialog |
| y.W | 0.89 |
| x1.W | 0.86 |
| x2.W | 0.90 |
| x3.W | 0.88 |
| x4.W | 0.99 |

p.value=1.37e-06<0,05, нулевая гипотеза по-прежнему отвергается: распределение ненормально.



Четко отслеживается положительная линейная взаимосвязь между у и х1, связь между у и х2, у и х3 также стала заметнее, возможна связь между у и х4. Чтобы подтвердить или опровергнуть эти предположения, рассчитаем корреляционную матрицу и проверим на значимость коэффициенты корреляции. 

Корреляционная матрица выделила 3 показателя оказывающих наиболее сильное влияние на результативный признак у. Это показатели х1, х2, х4. Также заметна связь между пказателями х1 и х2, х1 и х4, ч2 и ч4.