

Национальный исследовательский университет
“Высшая школа экономики”

Студентка второго курса
ОП “Бизнес-информатика”
Григорашенко Екатерина Андреевна
19.12.2020

ДОМАШНЕЕ ЗАДАНИЕ ПО КУРСУ
“ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА И ПРЕДСТАВЛЕНИЕ
СТАТИСТИЧЕСКИХ ДАННЫХ”
ЧАСТЬ №2. ДВУМЕРНЫЕ ДАННЫЕ

Содержание

Содержание	2
Используемые показатели для анализа	3
Описание показателей, выбранных для исследования.	4
Корреляционный анализ	4
Регрессионный анализ	7
Источники	14

1. Используемые показатели для анализа

у - курс доллара к рублю

х - цена одного барреля нефти Brent в долларах

Таблица 1. Показатели используемые для анализа

Дата	у	х
18.11.2020	44,13	76,25
19.11.2020	44,17	75,93
20.11.2020	45,24	76,26
21.11.2020	45,24	76,01
22.11.2020	45,13	76,01
23.11.2020	45,83	76,01
24.11.2020	47,91	75,76
25.11.2020	48,73	75,81
26.11.2020	47,77	75,47
27.11.2020	48,30	75,45
28.11.2020	48,30	75,86
29.11.2020	47,81	75,86
30.11.2020	47,76	75,86
01.12.2020	47,25	76,20
02.12.2020	47,97	76,32
03.12.2020	48,82	75,62
04.12.2020	49,03	75,20
05.12.2020	49,03	74,26
06.12.2020	49,06	74,26
07.12.2020	48,66	74,26

08.12.2020	48,85	74,25
09.12.2020	48,98	73,66
10.12.2020	50,43	73,31
11.12.2020	50,07	73,71
12.12.2020	50,07	73,12
13.12.2020	50,08	73,12
14.12.2020	50,36	73,12
15.12.2020	50,69	72,93
16.12.2020	51,16	73,45
17.12.2020	51,34	73,42
18.12.2020	51,35	72,98

2. Описание показателей, выбранных для исследования.

Показатель X является одномерным и временным. Он демонстрирует изменение цены одного барреля нефти Brent в долларах в течение 31 дня (с 18 ноября по 18 декабря 2020 года).

Показатель Y является одномерным и временным. Он демонстрирует изменение курса доллара к рублю в течение 31 дня (с 18 ноября по 18 декабря 2020 года).

Y является результативным признаком X, X же является независимой переменной в рамках данного исследования.

3. Корреляционный анализ

Используемые формулы:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y} \quad S_x = \sqrt{\overline{x^2} - (\bar{x})^2}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Корреляционный анализ по исходным данным

Таблица 2. Расчет характеристик Sx и Sy

ср. арифм. x^2	ср. квадр. x	ср. арифм. y^2	ср. квадр. y	Sx	Sy
2343,84	2339,81	5601,01	5599,53	2,01	1,22

Таблица 3. Расчет коэффициента корреляции

ср. арифм. xy	ср. арифм. x	ср. арифм. y	r
3617,61	48,37	74,83	-0,83

Коэффициент корреляции $r = -0,83$. Значение модуля равное 0,83 попадает в интервал от 0,7 до 0,9, следовательно между признаками X и Y присутствует сильная корреляция и курс доллара к рублю действительно зависит от цены на нефть Brent. Однако, так как сам коэффициент является отрицательным, зависимость между признаками является обратной.

- Корреляционный анализ по нормированным данным

Формулы использованные для нормирования:

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{\sigma_i}.$$

Таблица 4. Нормированные показатели используемые для анализа

Дата	y норм.	x норм.
18.11.2020	1,17	-2,11
19.11.2020	0,90	-2,09
20.11.2020	1,17	-1,56
21.11.2020	0,97	-1,56
22.11.2020	0,97	-1,61
23.11.2020	0,97	-1,27

24.11.2020	0,76	-0,23
25.11.2020	0,80	0,18
26.11.2020	0,53	-0,30
27.11.2020	0,51	-0,04
28.11.2020	0,85	-0,04
29.11.2020	0,85	-0,28
30.11.2020	0,85	-0,30
01.12.2020	1,12	-0,56
02.12.2020	1,22	-0,20
03.12.2020	0,65	0,22
04.12.2020	0,30	0,33
05.12.2020	-0,47	0,33
06.12.2020	-0,47	0,34
07.12.2020	-0,47	0,14
08.12.2020	-0,48	0,24
09.12.2020	-0,96	0,30
10.12.2020	-1,25	1,03
11.12.2020	-0,92	0,85
12.12.2020	-1,40	0,85
13.12.2020	-1,40	0,85
14.12.2020	-1,40	0,99
15.12.2020	-1,56	1,15
16.12.2020	-1,13	1,39
17.12.2020	-1,16	1,48
18.12.2020	-1,52	1,48

Таблица 5. Расчет характеристик S_x и S_y для нормированных показателей

ср. арифм. x^2	ср. квадр. x	ср. арифм. y^2	ср. квадр. y	S_x	S_y
1	2,89E-29	1	2,29E-28	1	1

Таблица 6. Расчет коэффициента корреляции для норм. показателей

ср. арифм. xy	ср. арифм. x	ср. арифм. y	r
-0.83	5,37E-15	1,51E-14	-0,83

Коэффициент корреляции нормированных показателей $r_{\text{норм.}} = -0,83$.
 Значение модуля равное 0,83 попадает в интервал от 0,7 до 0,9,
 следовательно между нормированными признаками X и Y присутствует
 сильная корреляция и курс доллара к рублю действительно зависит от
 цены на нефть Brent. Однако, так как сам коэффициент является
 отрицательным, зависимость между признаками является обратной.

4. Регрессионный анализ

Используемые формулы:

$$\hat{y} = b_0 + b_1 \cdot x$$

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x^2}$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

- Построение линейной регрессионной модели по исходным данным

$$b_1 = (3617,61 - 48,37 * 74,83) / 2,01^2 = -0,5$$

$$b_0 = 74,83 - (-0,5) * 48,37 = 99,25$$

Уравнение регрессии для исходных данных: $\hat{y} = 99,25 - 0,5x$

- Построение линейной регрессионной модели по стандартизованным данным

Аналогичные расчеты

$$b_1 = -0,83$$

$$b_0 = 1,96E-14$$

Уравнение регрессии для стандартизованных данных: $\hat{y} = 1,96E-14 - 0,83x$

- Оценка качества полученных моделей

Найдем значения \hat{y}_i .

Таблица 7. Расчет значений \hat{y}_i для исходных данных

Дата	y	x	$\hat{y} = 99,25 - 0,5x$
18.11.2020	76,25	44,13	76,97
19.11.2020	75,93	44,17	76,95
20.11.2020	76,26	45,24	76,41
21.11.2020	76,01	45,24	76,41
22.11.2020	76,01	45,13	76,47
23.11.2020	76,01	45,83	76,11
24.11.2020	75,76	47,91	75,06
25.11.2020	75,81	48,73	74,65
26.11.2020	75,47	47,77	75,13
27.11.2020	75,45	48,30	74,87
28.11.2020	75,86	48,30	74,87
29.11.2020	75,86	47,81	75,11
30.11.2020	75,86	47,76	75,14
01.12.2020	76,20	47,25	75,40
02.12.2020	76,32	47,97	75,03

03.12.2020	75,62	48,82	74,60
04.12.2020	75,20	49,03	74,50
05.12.2020	74,26	49,03	74,50
06.12.2020	74,26	49,06	74,48
07.12.2020	74,26	48,66	74,68
08.12.2020	74,25	48,85	74,59
09.12.2020	73,66	48,98	74,52
10.12.2020	73,31	50,43	73,79
11.12.2020	73,71	50,07	73,97
12.12.2020	73,12	50,07	73,97
13.12.2020	73,12	50,08	73,97
14.12.2020	73,12	50,36	73,83
15.12.2020	72,93	50,69	73,66
16.12.2020	73,45	51,16	73,42
17.12.2020	73,42	51,34	73,33
18.12.2020	72,98	51,35	73,33

Таблица 7. Расчет значений y_i^{\wedge} для нормированных данных

Дата	y	x	$y^{\wedge} = 1,9E-14 - 0,83x$
18.11.2020	1,17	-2,11	1,76
19.11.2020	0,90	-2,09	1,74
20.11.2020	1,17	-1,56	1,30
21.11.2020	0,97	-1,56	1,30
22.11.2020	0,97	-1,61	1,34
23.11.2020	0,97	-1,27	1,05
24.11.2020	0,76	-0,23	0,19

25.11.2020	0,80	0,18	-0,15
26.11.2020	0,53	-0,30	0,25
27.11.2020	0,51	-0,04	0,03
28.11.2020	0,85	-0,04	0,03
29.11.2020	0,85	-0,28	0,23
30.11.2020	0,85	-0,30	0,25
01.12.2020	1,12	-0,56	0,46
02.12.2020	1,22	-0,20	0,17
03.12.2020	0,65	0,22	-0,19
04.12.2020	0,30	0,33	-0,27
05.12.2020	-0,47	0,33	-0,27
06.12.2020	-0,47	0,34	-0,29
07.12.2020	-0,47	0,14	-0,12
08.12.2020	-0,48	0,24	-0,20
09.12.2020	-0,96	0,30	-0,25
10.12.2020	-1,25	1,03	-0,85
11.12.2020	-0,92	0,85	-0,70
12.12.2020	-1,40	0,85	-0,70
13.12.2020	-1,40	0,85	-0,71
14.12.2020	-1,40	0,99	-0,82
15.12.2020	-1,56	1,15	-0,96
16.12.2020	-1,13	1,39	-1,15
17.12.2020	-1,16	1,48	-1,23
18.12.2020	-1,52	1,48	-1,23

Формула средней ошибки аппроксимации:

$$\overline{A_i} = \frac{1}{n} \cdot \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100$$

Средняя ошибка аппроксимации для модели построенной по исходным данным равна 0,79%. Так как значение не превышает 12-15%, то значит модель является адекватной.

Средняя ошибка аппроксимации для модели построенной по нормированным данным равна 23,54%. Так как значение превышает 12-15%, то значит модель не является адекватной.

- Интерпретация моделей

Коэффициент регрессии в модели построенной по исходным данным равен -0,5. Следовательно, с увеличением цены одного барреля нефти Brent курс доллара к рублю уменьшается на 0,5 рубля.

Коэффициент регрессии в модели построенной по нормированным данным равен -0,83. Следовательно, с увеличением цены одного барреля нефти Brent курс доллара к рублю уменьшается на 0,83 рубля.

- Интерпретация коэффициента детерминации

Коэффициент регрессии в модели построенной по исходным данным равен 69,17%. Следовательно, изменение курс доллара к рублю можно на 69,17% объяснить изменением цены одного барреля нефти Brent.

Коэффициент регрессии в модели построенной по нормированным данным равен 69,17%. Следовательно, изменение курс доллара к рублю можно на 69,17% объяснить изменением цены одного барреля нефти Brent.

- Выбор оптимальной модели

Модель построенная по исходным данным будет более оптимальной, так как значения средней ошибки аппроксимации показали, что модель по норм. данным не является адекватной. Также из-за более маленького

коэффициента регрессии, первая модель более четко отображает зависимость признаков.

- Графически представить кривые исходных данных и по регрессионным моделям

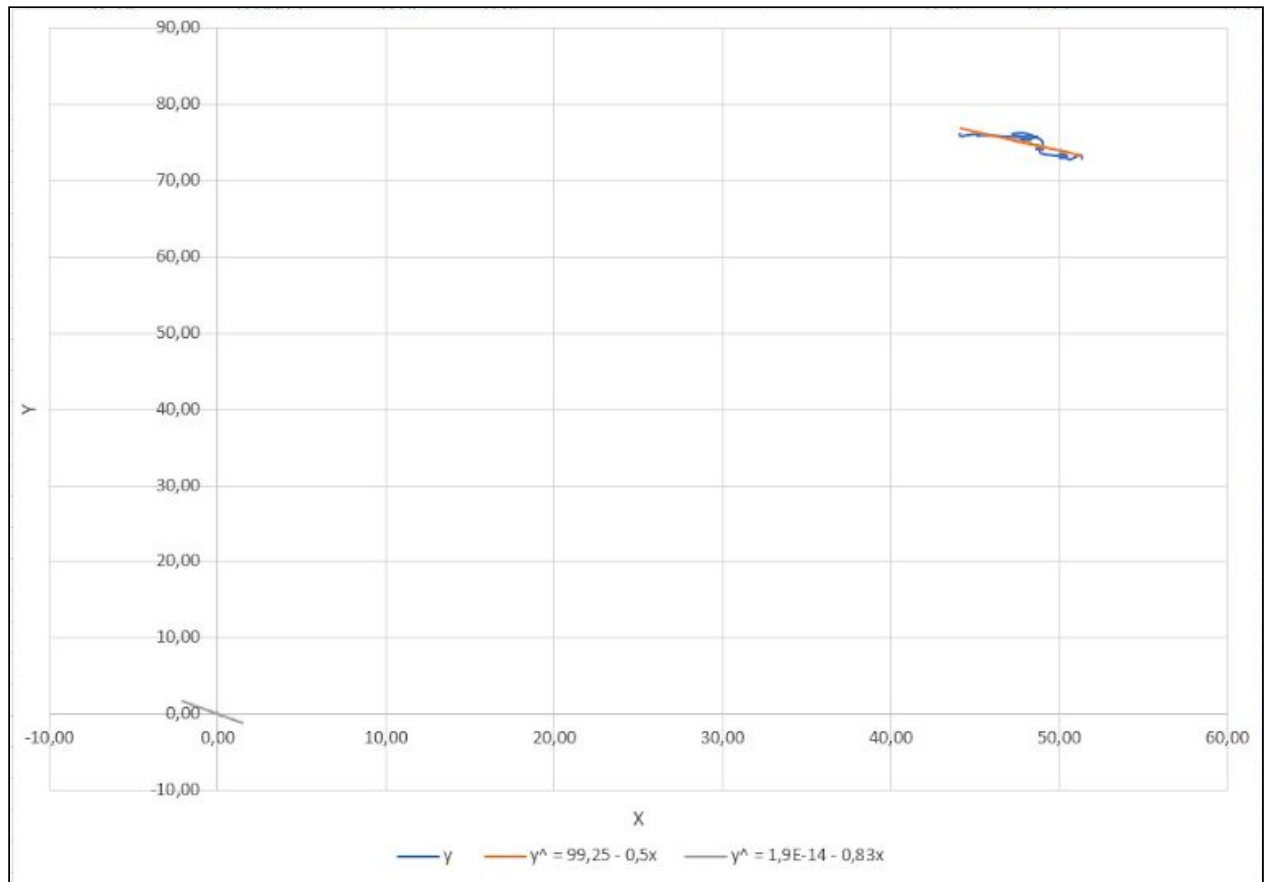


Рис.1 Графически представленные исходные данные и две регрессионные модели.

В задание было указано построить все 3 кривых на одном графике, однако, так как в моей работе нормированные данные очень сильно отличаются от исходных, ниже будут продемонстрированы отдельно кривые связанные с исходными данными и кривые связанные с нормированными данными.

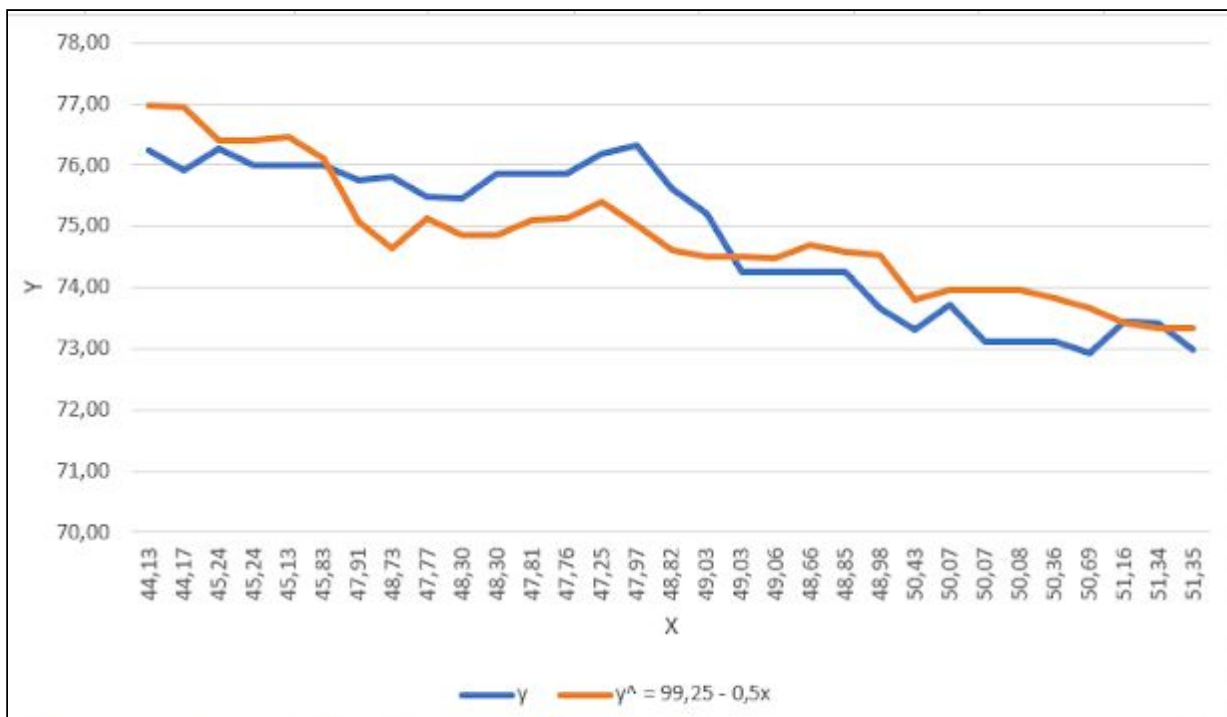


Рис.2 Графически представленные исходные данные и регрессионная модель, основанная на них.

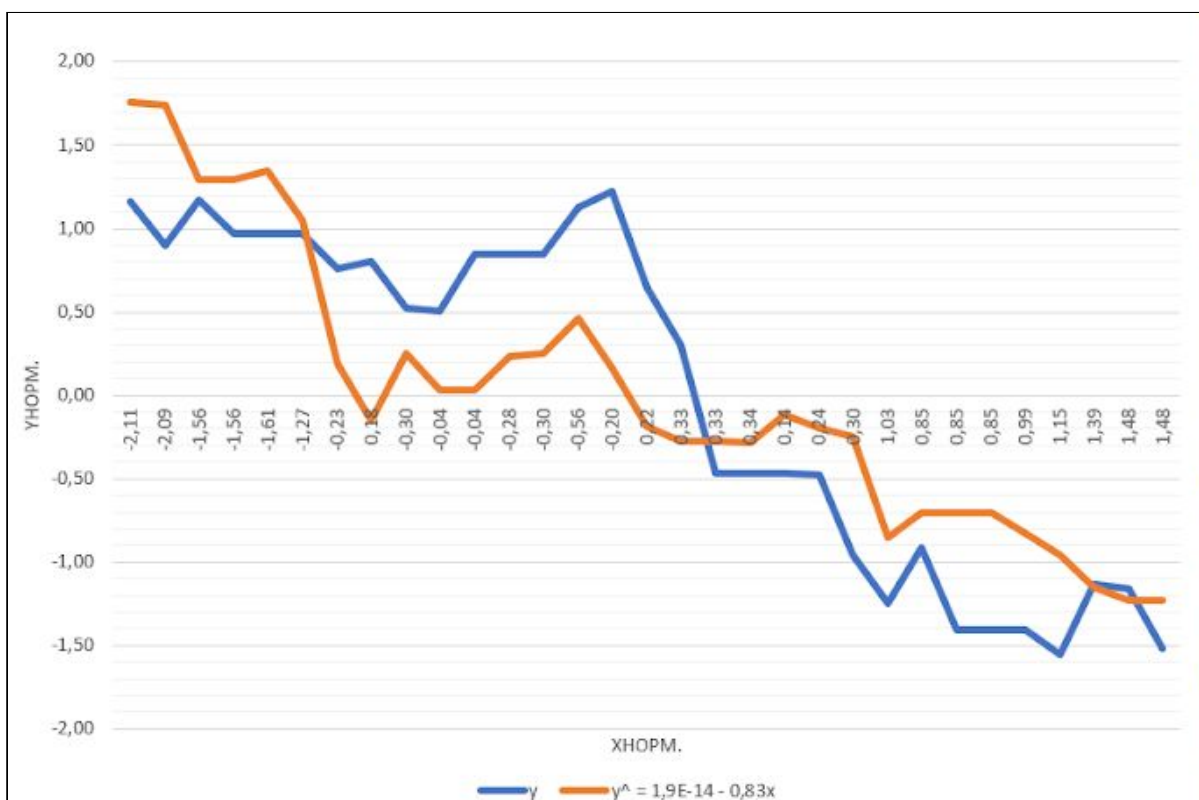


Рис.3 Графически представленные нормированные данные и регрессионная модель, основанная на них.

Источники

- <https://www.banki.ru/products/currency/rub/>
- https://cbr.ru/currency_base/daily/?UniDbQuery.Posted=True&UniDbQuery.To=18.11.2020
- <https://wiki.loginom.ru/articles/linear-normalization.html>
- <http://masters.donntu.org/2005/kita/tokarev/library/linreg.htm>