

# Отчет Шуплецов, НФИбд01-22, Вариант 15, ИДЗ\_1, 21.09.2024

1. Оценить линейное уравнение парной регрессии  $Y$  от  $X$ , проинтерпретировать коэффициенты регрессии.

Модель 1: МНК, использованы наблюдения 1-50  
Зависимая переменная:  $y$

	Коэффициент $m$	Ст. ошибка	$t$ - статистика	$p$ -значение	
const	-5,37828	0,724664	-7,422	<0,0001	***
$x$	1,02438	0,0652681	15,70	<0,0001	***
Среднее завис. перемен	5,640000	Ст. откл. завис. перемен	3,114679		
Сумма кв. остатков	77,52177	Ст. ошибка модели	1,270841		
<b>R-квадрат</b>	<b>0,836920</b>	Исправ. R-квадрат	0,833522		
$F(1, 48)$	246,3338	<b>R-значение (F)</b>	<b>1,56e-20</b>		
Лог. правдоподобие	-81,91032	Крит. Акаике	167,8206		
Крит. Шварца	171,6447	Крит. Хеннана-Куинна	169,2769		

$$Y = 1,02x - 5$$

$$Y = b_1 * x + b_0$$

При возрастании  $x$  на 1 ед. своего измерения  $y$  в среднем возрастает на **1,02 единиц** своего измерения

95% доверительный интервал

$$t(48, 0,025) = 2,011$$

	коэффициент	нижний	верхний
-----			
const	-5,37828	-6,83532	-3,92125
$x$	1,02438	0,893154	1,15562

При возрастании  $x$  на 1 ед. своего измерения  $y$  в среднем возрастает на величину заключенную в интервале с вер. (0,893154, 1,15562).

2. Проверить статистическую значимость уравнения в целом.

$H_0: b_1=0$  - уравнение незначимо в целом

$H_1: b_1 \neq 0$  - уравнение значимо в целом

Р-значение (F)

1,56e-20

**Р-значение <0,01 – принимаем гипотезу H1 с вер 99%**

3. Рассчитать линейный коэффициент парной корреляции и его стандартную ошибку. Дать интерпретацию коэффициентов корреляции и детерминации.

$$\text{corr}(x, y) = 0,91483325$$

**Коэффициент корреляции высокий, то есть с возрастанием x у тоже возрастает.**

Р-квадрат

0,836920

**Вывод: у объясняется x на ~84%, то есть существуют еще факторы, влияющие на у.**

4. Проверить значимость коэффициентов регрессии и корреляции на 1%, 5% и 10% уровне. На каком уровне достаточно провести проверку каждого из коэффициентов в вашем случае?

Р-значение (F)

1,56e-20

**Наша Р-значение крайне мало, поэтому коэффициенты регрессии и уравнение значимы на уровнях значимости 1%, 5%, 10%, причем достаточно было проверить значимость на уровне 1%.**

**Принимаем H1 с вероятностью 99%.**

5. Рассчитать коэффициент эластичности и проинтерпретировать его.

Описательная статистика, использованы наблюдения 1 - 50

Переменная	Среднее	Медиана	ст. откл.	Мин.	Макс.
x	10,8	10,8	2,78	4,80	18,4
y	5,64	5,45	3,11	-0,500	11,9

$$\varepsilon (X_i, Y_i) = \beta_1 * (X_{\text{ср}} / Y_{\text{ср}}) = 1,02438 * (10,8 / 5,64) \approx 1,96$$

**Если независимая переменная X возрастет на 1%, то зависимая переменная у возрастает на ~1,96%.**

**Так как значение эластичности больше 1, в нашем случае эластичная зависимость.**

6. Рассчитать среднюю ошибку аппроксимации и оценить по этому критерию качество модели.

Средняя ошибка (ME)

3,5261e-015

Корень из средней квадратичной ошибки (RMSE) 1,2452

Средняя абсолютная ошибка (MAE) 0,9806

Средняя процентная ошибка (MPE) -0,068354

**Средняя абсолютная процентная ошибка (MAPE) 37,361 – плохое качество модели**

U-статистика Тейла (Theil's U) 0,097784

Пропорция смещения, UM 0

Пропорция регрессии, UR 0

Пропорция возмущений, UD 1

*7. Выполнить прогноз  $\hat{Y}$  при прогнозном значении  $X$  - 107% от среднего уровня. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.*

$$\hat{X}_{pr} = 1,07 \cdot x_{sr} = 1,07 \cdot 10,8 = 11,556$$

$$\hat{Y}_{pr} = (-5,37828) + 1,02438 \cdot 11,556 \approx 6,46$$

Для 90% доверительных интервалов,  $t(48, 0,05) = 1,677$

Набл.	$y$	прогнозирование	ст. ошибка	90% доверительный интервал
51	не определено	6,45951	1,28455	(4,30503, 8,61398)

Для 95% доверительных интервалов,  $t(48, 0,025) = 2,011$

Набл.	$y$	прогнозирование	ст. ошибка	95% доверительный интервал
51	не определено	6,45951	1,28455	(3,87675, 9,04226)

Для 99% доверительных интервалов,  $t(48, 0,005) = 2,682$

Набл.	$y$	прогнозирование	ст. ошибка	99% доверительный интервал
51	не определено	6,45951	1,28455	(3,01409, 9,90493)

**Все три уровня доверия показывают, что прогноз  $y$  является достаточно надежным, учитывая стандартную ошибку 1,28455. Собранные данные и вычисленные интервалы подтверждают, что модель дает адекватный прогноз, и истинное значение переменной, скорее всего, находится в указанных интервалах.**

8. Представить результат оценки регрессии графически.

Коробчатая диаграмма графика остатков:

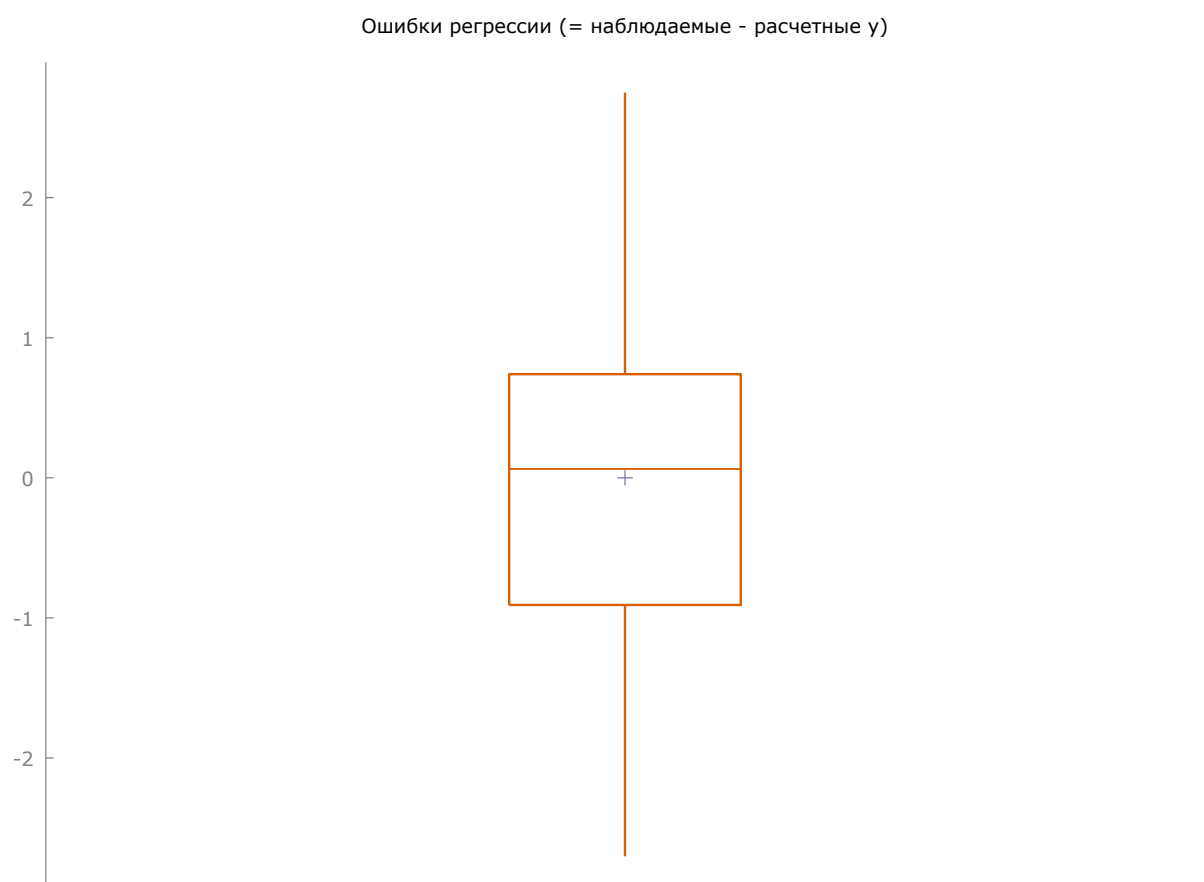


График остатков по номеру наблюдения:

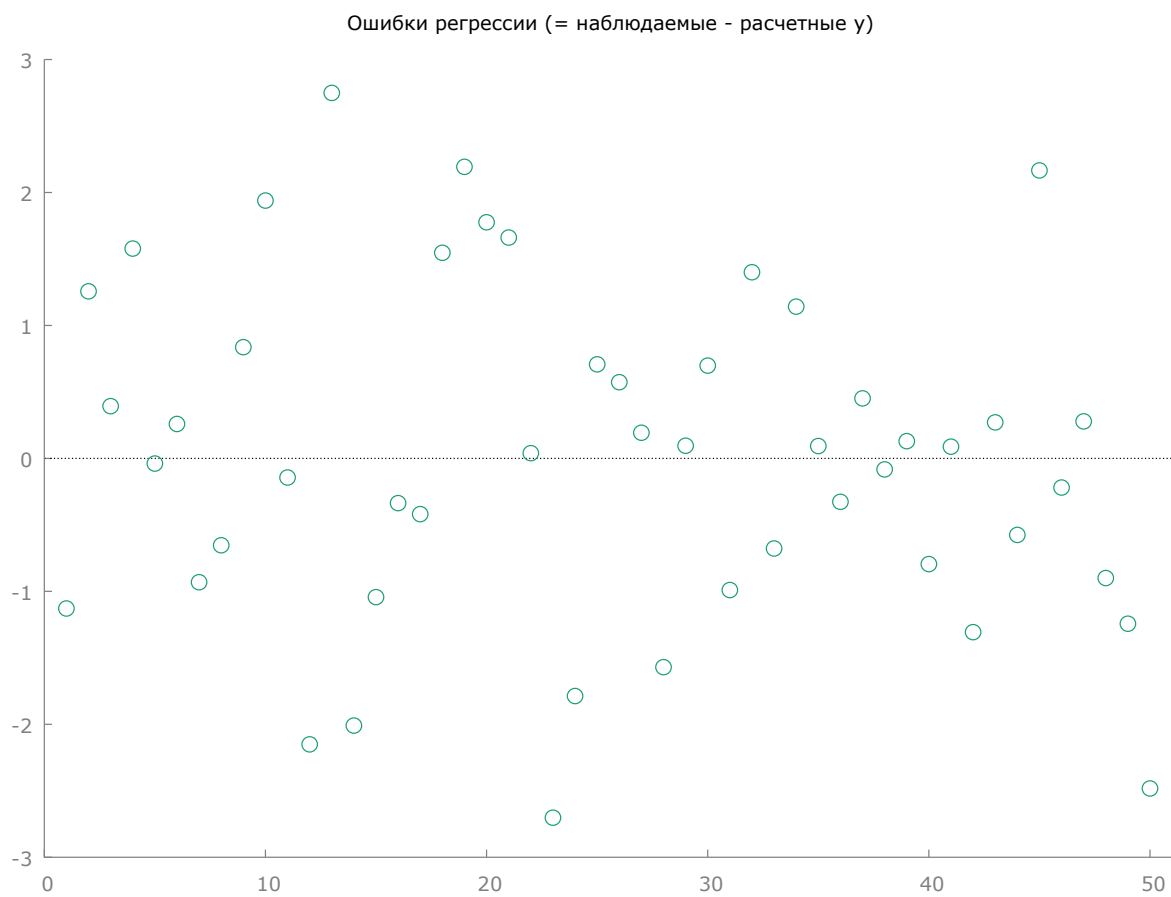
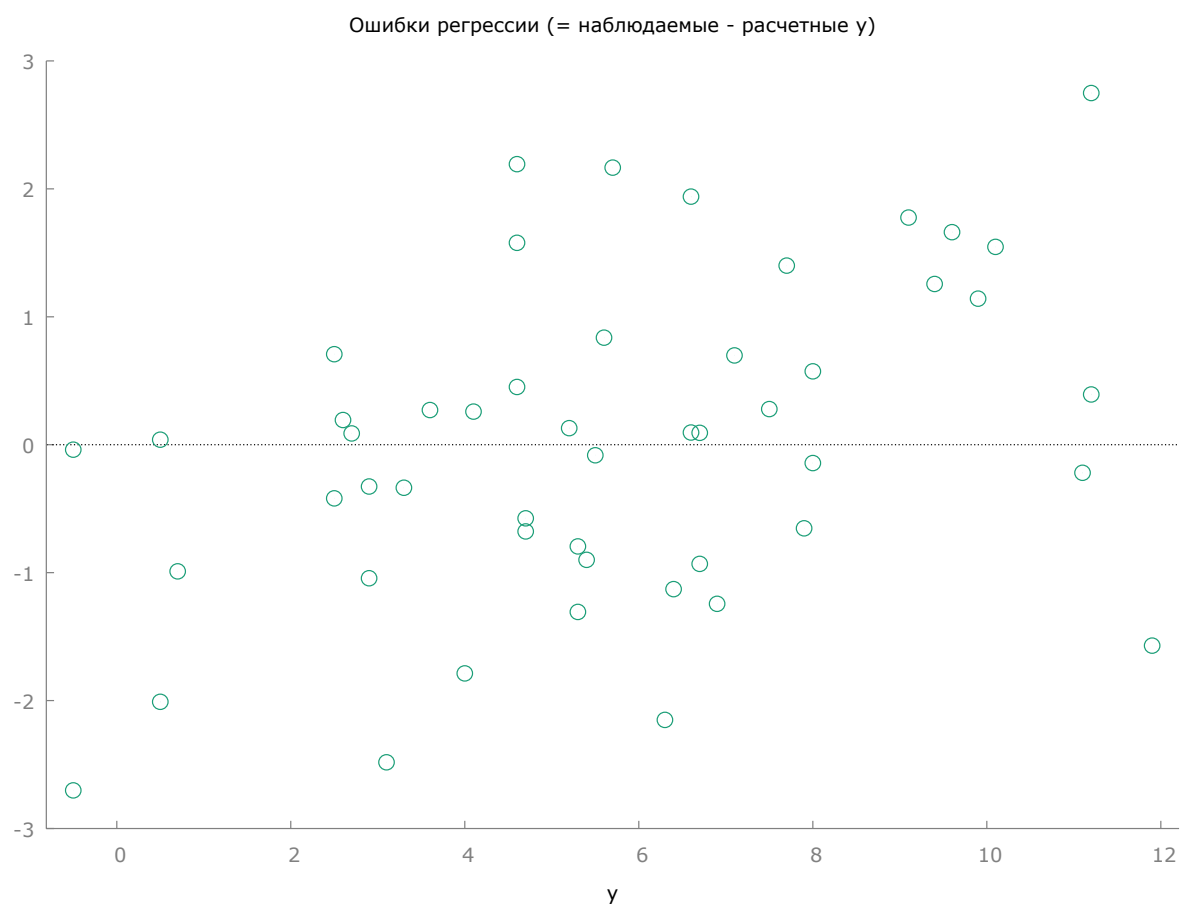


График остатков в зависимости от  $y$ :



### График остатков в зависимости от x:

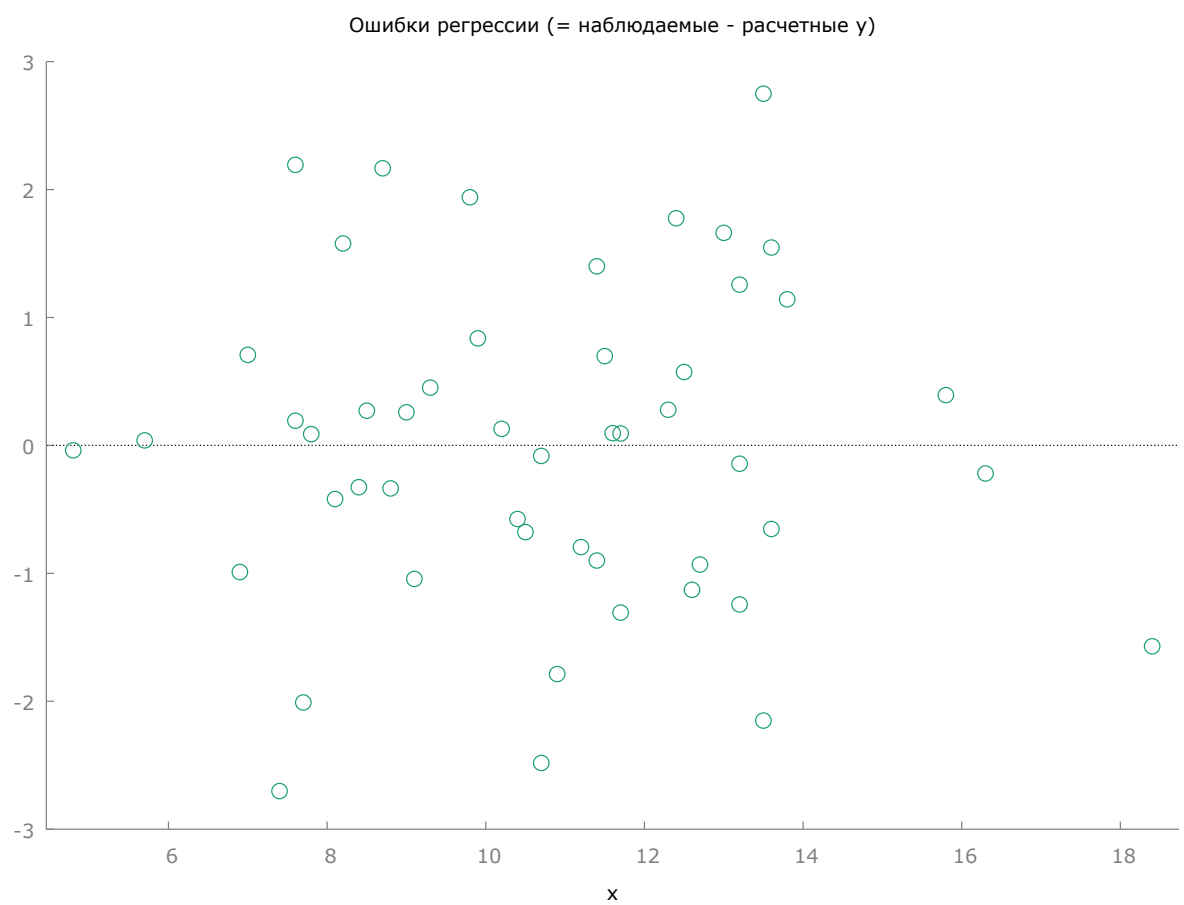


График наблюдаемых и расчетных значений по номеру наблюдений:

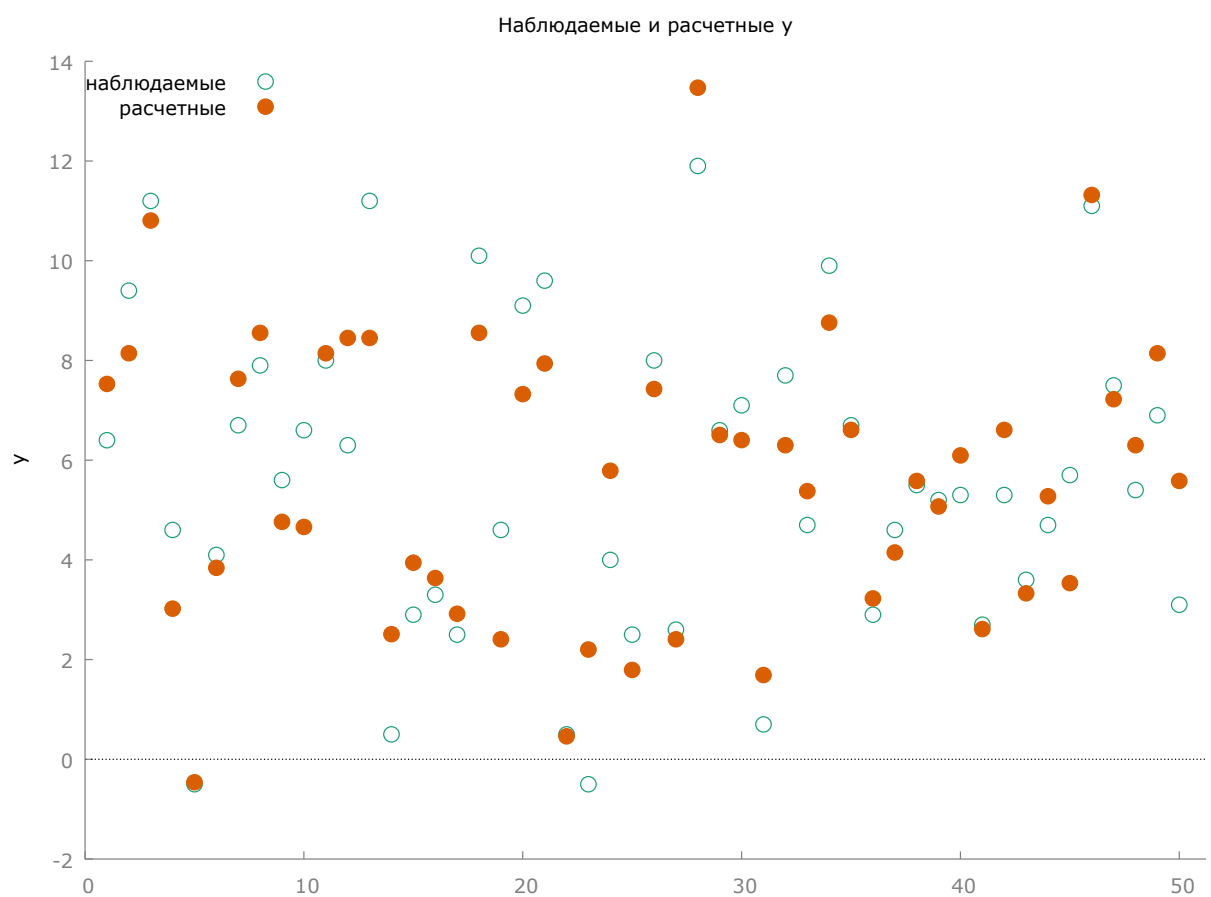




График наблюдаемых и расчетных значений в зависимости от  $x$ :

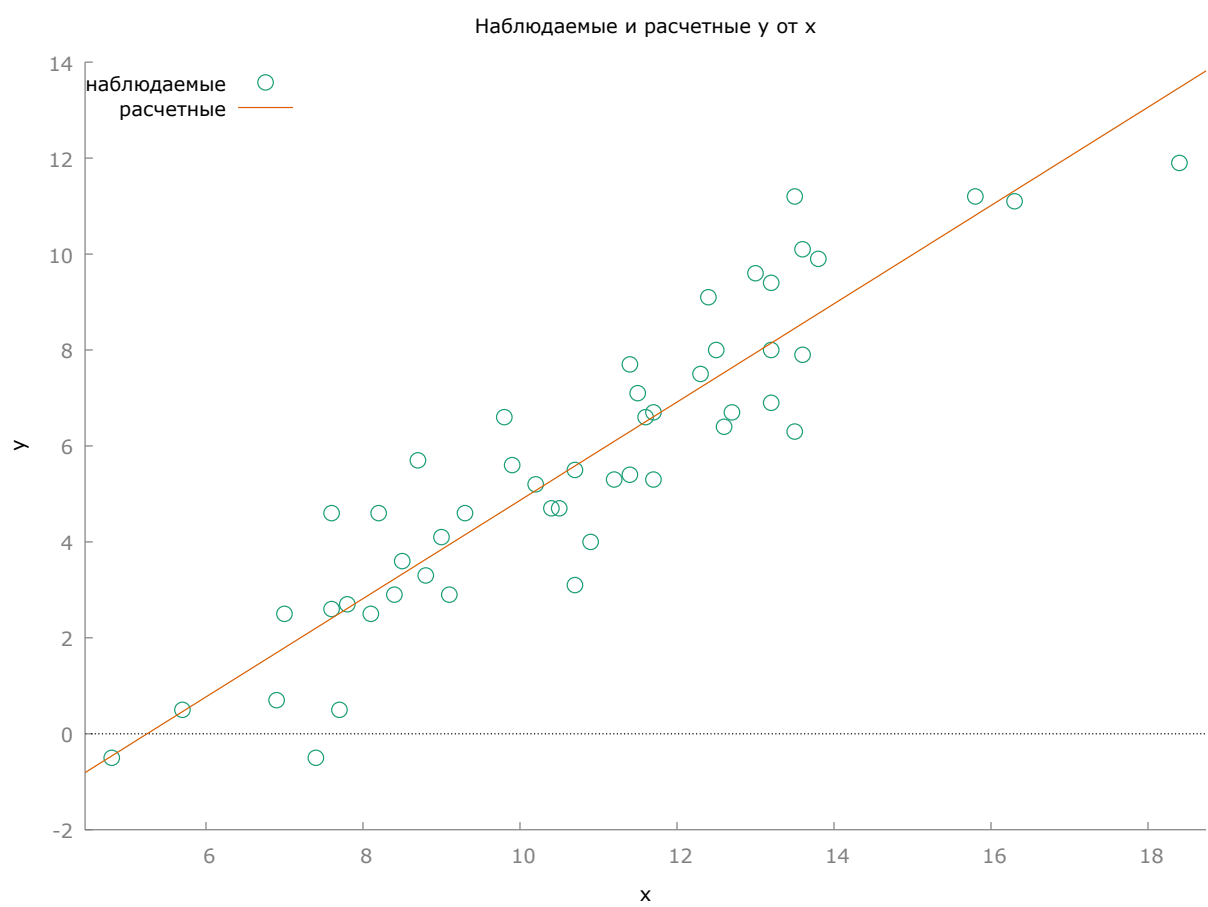
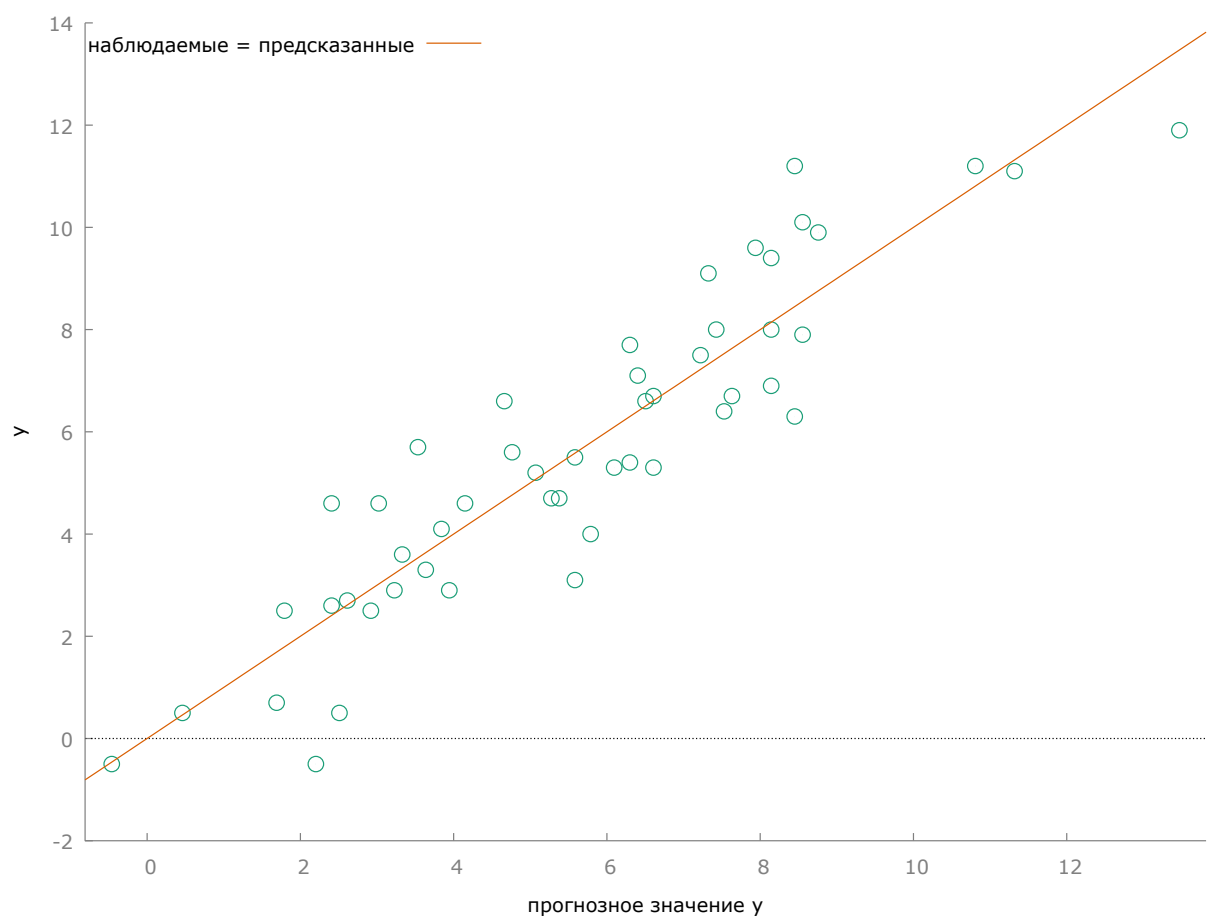
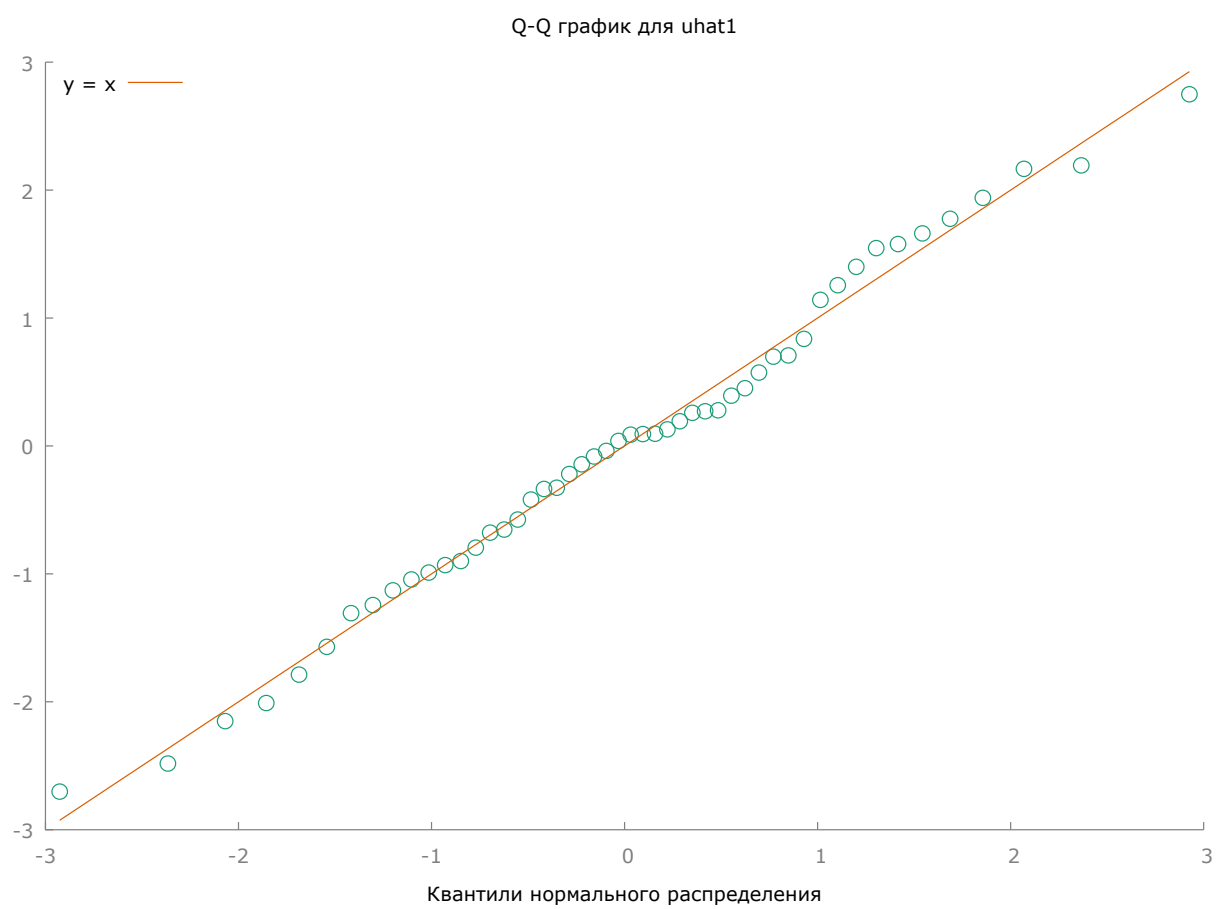


График наблюдаемые от расчетных:



Q-Q график остатков:

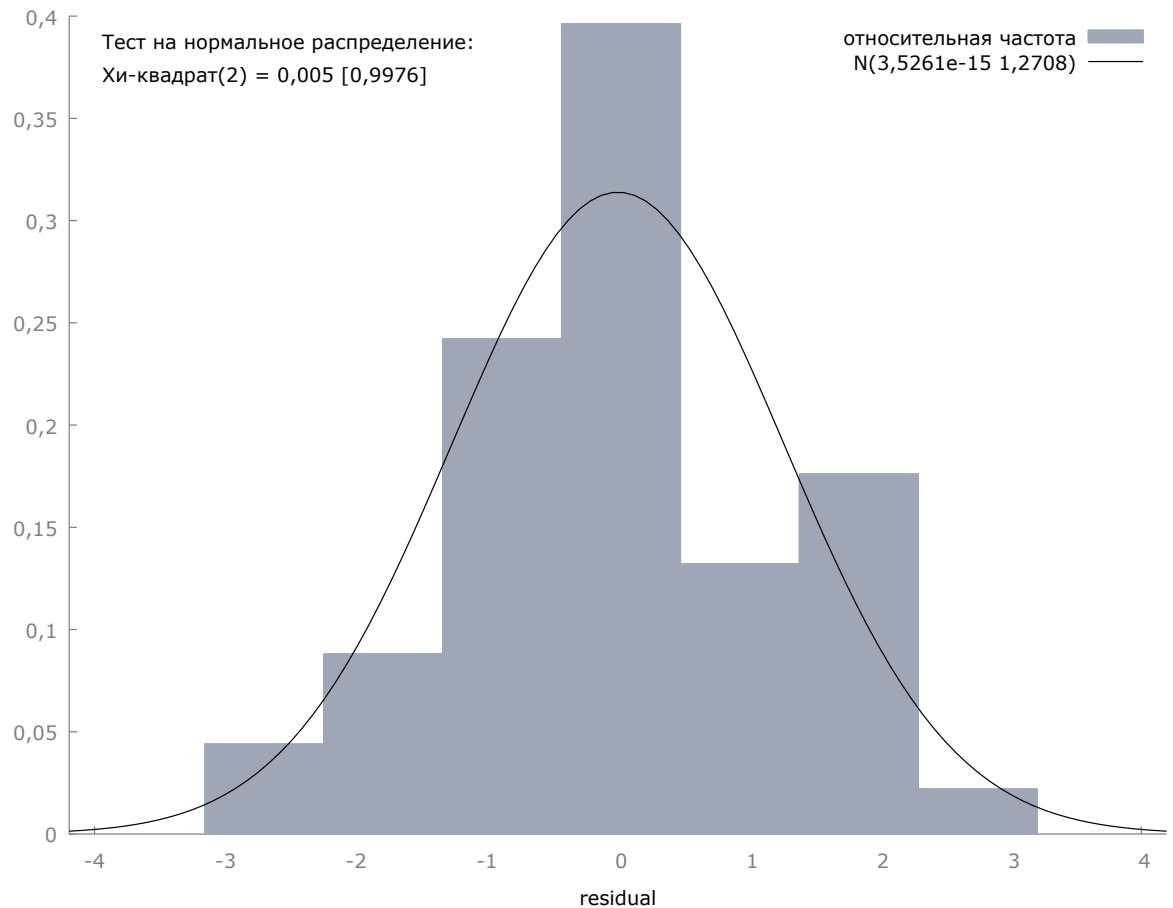


9. Проверить остатки на нормальность. Графически и по критериям.

H0: остатки имеют нормальное распределение

H1: распределение остатков отличается от нормального

По хи- квадрат – 0,9976 - принимаем H0



Тест на нормальное распределение у:

Тест Дурника-Хансена (Doornik-Hansen) = 0,137709, p-значение 0,933463

Тест Шапиро-Уилка (Shapiro-Wilk W) = 0,980326, p-значение 0,565777

Тест Лиллифорса (Lilliefors) = 0,0566961, p-значение  $\approx 1$

Тест Жарка-Бера (Jarque-Bera) = 0,594731, p-значение 0,742772

**Все тесты показывают, что нам следует принять Н0.**

	Ст. откл.	Вариация	Асимметрия	Экссесс
x	2,7816	0,25861	0,23902	-0,027511
y	3,1147	0,55225	0,051111	-0,52443

**Асимметрия по модулю меньше 1, значит модель завышает у не сильно.**

*10. Проверить остатки на гетероскедастичность.*

Тест Вайта (White) на гетероскедастичность

МНК, использованы наблюдения 1-50

Зависимая переменная:  $\text{uhat}^2$

	коэффициент	ст. ошибка	t-статистика	p-значение
-----				
const	0,596707	3,33505	0,1789	0,8588
x	0,150635	0,611365	0,2464	0,8065
sq_x	-0,00540669	0,0271774	-0,1989	0,8432

Неисправленный R-квадрат = 0,002712

Тестовая статистика:  $TR^2 = 0,135619$ ,

p-значение =  $P(\text{Хи-квадрат}(2) > 0,135619) = 0,934438$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-50

Зависимая переменная: масштабированное  $\text{uhat}^2$

	коэффициент	ст. ошибка	t-статистика	p-значение
-----				
const	0,786550	0,734228	1,071	0,2894
x	0,0198447	0,0661295	0,3001	0,7654

Объясненная сумма квадратов = 0,149304

Тестовая статистика: LM = 0,074652,

p-значение =  $P(\chi^2(1) > 0,074652) = 0,784680$

Тест Бройша-Пагана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-50

Зависимая переменная: масштабированное  $\hat{u}^2$  (робастный вариант Коенкера (Koenker))

	коэффициент	ст. ошибка	t-статистика	p-значение
-----				
const	-0,330940	1,13837	-0,2907	0,7725
x	0,0307680	0,102530	0,3001	0,7654

Объясненная сумма квадратов = 0,358904

Тестовая статистика: LM = 0,093630,

p-значение =  $P(\chi^2(1) > 0,093630) = 0,759612$

**Во всех трех тестах на гетероскедастичность (тест Уайта, тест Бройша-Пагана и его коенкерский вариант) p-значения значительно превышают стандартные уровни значимости (обычно 0,01, 0,05 или 0,10). Это означает, что у нас нет оснований для отклонения нулевой гипотезы об отсутствии гетероскедастичности. Таким образом, по результатам проведенных тестов можно сделать вывод, что гетероскедастичность в остатках нашей модели не обнаружена. Модель, вероятно, обладает постоянной дисперсией ошибок, что подтверждает предположения классической линейной модели.**