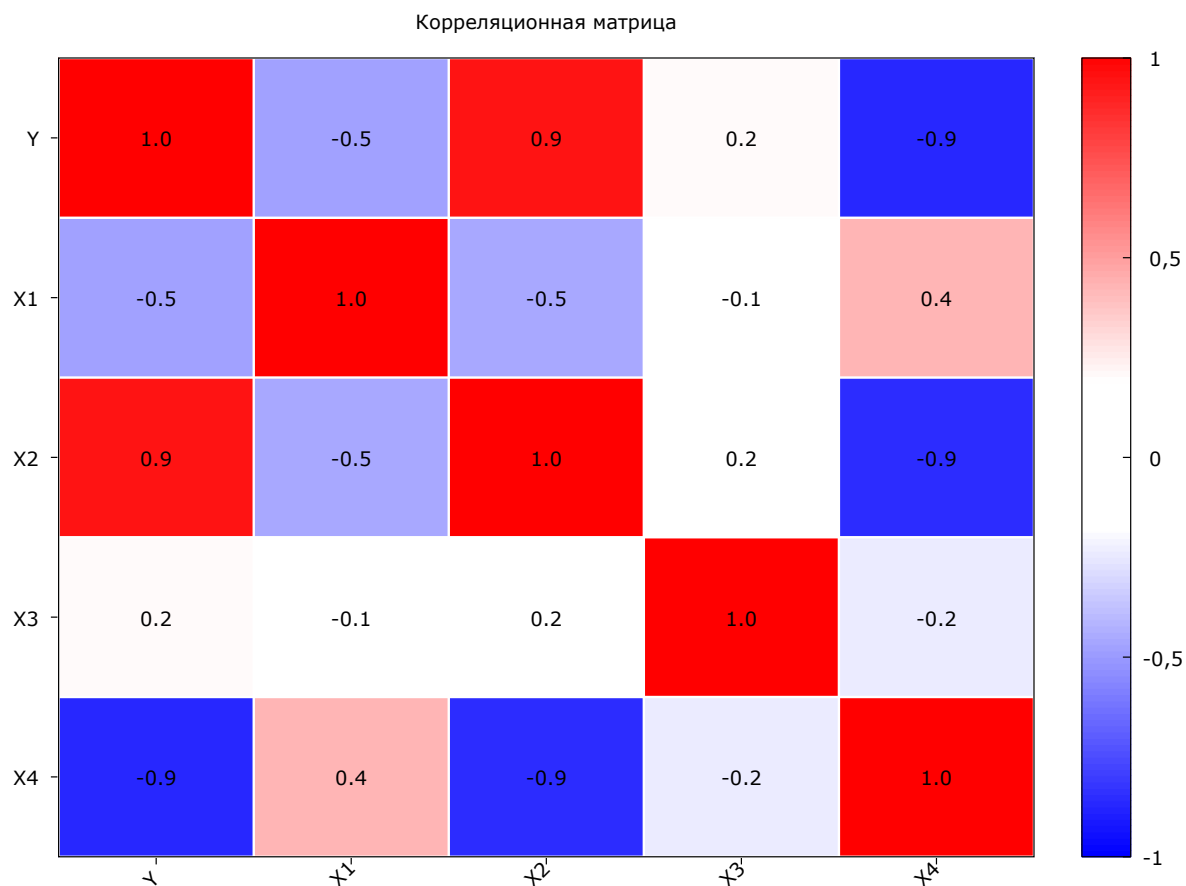


1. Определиться какие переменные войдут в модель – оставить модель как минимум с двумя переменными – корреляционная матрица, *vif* и тест на избыточные переменные

Модель 1: МНК, использованы наблюдения 1-50
Зависимая переменная: Y

| | Коэффициент | Ст. ошибка | t- статистика | p-значение | |
|-------|-------------|------------|------------------|------------|-----|
| | <i>m</i> | | | | |
| const | −4,79059 | 11,4212 | −0,4194 | 0,6769 | |
| X1 | −1,31094 | 1,19424 | −1,098 | 0,2782 | |
| X2 | 4,17534 | 0,567987 | 7,351 | <0,0001 | *** |
| X3 | 0,0358755 | 0,393538 | 0,09116 | 0,9278 | |
| X4 | −0,436544 | 0,158282 | −2,758 | 0,0084 | *** |

| | | | |
|------------------------|-----------|--------------------------|----------|
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 |
| Сумма кв. остатков | 2357,363 | Ст. ошибка модели | 7,237806 |
| R-квадрат | 0,902411 | Исправ. R-квадрат | 0,893737 |
| F(4, 45) | 104,0295 | P-значение (F) | 3,89e-22 |
| Лог. правдоподобие | −167,2788 | Крит. Акаике | 344,5576 |
| Крит. Шварца | 354,1178 | Крит. Хеннана-Куинна | 348,1982 |



Как можно заметить из корреляционной матрицы, X2 и X4 имеют высокую степень корреляции с Y, в то время как X1 и X3 нет, они требуют дополнительных проверок.

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

| | |
|----|-------|
| X1 | 1,276 |
| X2 | 4,068 |
| X3 | 1,071 |
| X4 | 3,999 |

$VIF(j) = 1/(1 - R(j)^2)$, где $R(j)$ - это коэффициент множественной корреляции между переменной j и другими независимыми переменными

Диагностика коллинеарности Белсли-Ку-Велша (Belsley-Kuh-Welsch):

разложение дисперсии

| lambda | cond | const | X1 | X2 | X3 | X4 |
|--------|--------|-------|-------|-------|-------|-------|
| 4,758 | 1,000 | 0,000 | 0,001 | 0,001 | 0,002 | 0,002 |
| 0,184 | 5,090 | 0,004 | 0,023 | 0,008 | 0,011 | 0,147 |
| 0,036 | 11,426 | 0,008 | 0,090 | 0,022 | 0,858 | 0,001 |
| 0,017 | 16,833 | 0,010 | 0,181 | 0,595 | 0,000 | 0,744 |
| 0,005 | 29,834 | 0,978 | 0,705 | 0,374 | 0,129 | 0,107 |

lambda = собственные значения обратной матрицы ковариаций (маленький 0,00534529)

cond = условный индекс

примечание: сумма столбцов с пропорциями дисперсии равна 1.0

Согласно ВКВ, условие ≥ 30 указывает на "сильную" (близкую к линейной) зависимость, и условие между 10 и 30 "умеренно сильную". Оценки параметров, чья дисперсия в основном связана с проблемными усл. значениями, могут считаться проблемными сами по себе.

Количество индексов состояния ≥ 30 : 0

Количество индексов состояния ≥ 10 : 3

Пропорции дисперсии ≥ 0.5 , связанные с условием ≥ 10 :

| const | X1 | X2 | X3 | X4 |
|-------|-------|-------|-------|-------|
| 0,996 | 0,976 | 0,991 | 0,987 | 0,851 |

Тестирование модели 7:

Нулевая гипотеза: параметры регрессии нулевые

X1, X3

Тестовая статистика: $F(2, 45) = 0,611969$, p-значение 0,546733

Как видно по VIF, все значения меньше 10, мультиколлинеарности нет. Также тест на избыточные переменные подтвердил, что X1 и X3 – незначимые.

Модель 3: МНК, использованы наблюдения 1-50

Зависимая переменная: Y

| | <i>Коэффициент</i> <i>t</i> | <i>Ст. ошибка</i> | <i>t-</i> <i>статистика</i> | <i>p-значение</i> | |
|------------------------|--------------------------------|--------------------------|--------------------------------|-------------------|-----|
| const | -15,2583 | 6,29990 | -2,422 | 0,0194 | ** |
| X2 | 4,31089 | 0,555636 | 7,758 | <0,0001 | *** |
| X3 | 0,0541032 | 0,394062 | 0,1373 | 0,8914 | |
| X4 | -0,444079 | 0,158485 | -2,802 | 0,0074 | *** |
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 | | |
| Сумма кв. остатков | 2420,488 | Ст. ошибка модели | 7,253916 | | |
| R-квадрат | 0,899798 | Исправ. R-квадрат | 0,893263 | | |
| F(3, 46) | 137,6908 | P-значение (F) | 5,48e-23 | | |
| Лог. правдоподобие | -167,9395 | Крит. Акаике | 343,8789 | | |
| Крит. Шварца | 351,5270 | Крит. Хеннана-Куинна | 346,7914 | | |

Модель 4: МНК, использованы наблюдения 1-50

Зависимая переменная: Y

| | <i>Коэффициент</i> <i>t</i> | <i>Ст. ошибка</i> | <i>t-</i> <i>статистика</i> | <i>p-значение</i> | |
|------------------------|--------------------------------|--------------------------|--------------------------------|-------------------|-----|
| const | -4,33999 | 10,1846 | -0,4261 | 0,6720 | |
| X2 | 4,17167 | 0,560422 | 7,444 | <0,0001 | *** |
| X4 | -0,439129 | 0,154034 | -2,851 | 0,0065 | *** |
| X1 | -1,31554 | 1,18024 | -1,115 | 0,2708 | |
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 | | |
| Сумма кв. остатков | 2357,798 | Ст. ошибка модели | 7,159363 | | |
| R-квадрат | 0,902393 | Исправ. R-квадрат | 0,896027 | | |
| F(3, 46) | 141,7594 | P-значение (F) | 3,00e-23 | | |
| Лог. правдоподобие | -167,2834 | Крит. Акаике | 342,5669 | | |
| Крит. Шварца | 350,2150 | Крит. Хеннана-Куинна | 345,4793 | | |

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

| | |
|----|-------|
| X2 | 4,048 |
| X4 | 3,871 |
| X1 | 1,274 |

$VIF(j) = 1/(1 - R(j)^2)$, где $R(j)$ - это коэффициент множественной корреляции между переменной j и другими независимыми переменными

Диагностика коллинеарности Белсли-Ку-Велша (Belsley-Kuh-Welsch):

разложение дисперсии

| lambda | cond | const | X2 | X4 | X1 |
|--------|--------|-------|-------|-------|-------|
| 3,803 | 1,000 | 0,001 | 0,001 | 0,003 | 0,001 |
| 0,175 | 4,665 | 0,007 | 0,005 | 0,145 | 0,033 |
| 0,017 | 15,049 | 0,012 | 0,597 | 0,768 | 0,182 |
| 0,006 | 25,153 | 0,980 | 0,396 | 0,083 | 0,784 |

lambda = собственные значения обратной матрицы ковариаций (маленький 0,00601014)

cond = условный индекс

примечание: сумма столбцов с пропорциями дисперсии равна 1.0

Согласно BКW, условие ≥ 30 указывает на "сильную" (близкую к линейной) зависимость, и условие между 10 и 30 "умеренно сильную". Оценки параметров, чья дисперсия в основном связана с проблемными усл. значениями, могут считаться проблемными сами по себе.

Количество индексов состояния ≥ 30 : 0

Количество индексов состояния ≥ 10 : 2

Пропорции дисперсии ≥ 0.5 , связанные с условием ≥ 10 :

| const | X2 | X4 | X1 |
|-------|-------|-------|-------|
| 0,992 | 0,993 | 0,851 | 0,966 |

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

| | |
|----|-------|
| X2 | 3,876 |
| X4 | 3,992 |
| X3 | 1,069 |

$VIF(j) = 1/(1 - R(j)^2)$, где $R(j)$ - это коэффициент множественной корреляции между переменной j и другими независимыми переменными

Диагностика коллинеарности Белсли-Ку-Велша (Belsley-Kuh-Welsch):

разложение дисперсии

| lambda | cond | const | X2 | X4 | X3 |
|--------|--------|-------|-------|-------|-------|
| 3,826 | 1,000 | 0,002 | 0,001 | 0,003 | 0,003 |
| 0,135 | 5,329 | 0,041 | 0,005 | 0,199 | 0,067 |
| 0,028 | 11,642 | 0,222 | 0,139 | 0,114 | 0,742 |
| 0,011 | 18,701 | 0,736 | 0,854 | 0,684 | 0,188 |

lambda = собственные значения обратной матрицы ковариаций (маленький 0,0109398)

cond = условный индекс

примечание: сумма столбцов с пропорциями дисперсии равна 1.0

Согласно BКW, условие ≥ 30 указывает на "сильную" (близкую к линейной) зависимость, и условие между 10 и 30 "умеренно сильную". Оценки параметров, чья дисперсия в основном связана с проблемными усл. значениями, могут считаться проблемными сами по себе.

Количество индексов состояния ≥ 30 : 0

Количество индексов состояния ≥ 10 : 2

Пропорции дисперсии ≥ 0.5 , связанные с условием ≥ 10 :

| const | X2 | X4 | X3 |
|-------|----|----|----|
|-------|----|----|----|

0,957 0,994 0,798 0,930

Будем изучать модель с переменными X2 и X4. X1 и X3 незначимы.

2. Интерпретация коэффициентов в модели. Доверительные интервалы для коэффициентов

Модель 6: МНК, использованы наблюдения 1-50
Зависимая переменная: Y

| | Коэффициент | Ст. ошибка | t- статистика | p-значение | |
|------------------------|-------------|--------------------------|------------------|------------|-----|
| | <i>m</i> | | | | |
| const | -14,6329 | 4,30676 | -3,398 | 0,0014 | *** |
| X2 | 4,30607 | 0,548708 | 7,848 | <0,0001 | *** |
| X4 | -0,448024 | 0,154223 | -2,905 | 0,0056 | *** |
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 | | |
| Сумма кв. остатков | 2421,480 | Ст. ошибка модели | 7,177802 | | |
| R-квадрат | 0,899757 | Исправ. R-квадрат | 0,895491 | | |
| F(2, 47) | 210,9300 | P-значение (F) | 3,35e-24 | | |
| Лог. правдоподобие | -167,9497 | Крит. Акаике | 341,8994 | | |
| Крит. Шварца | 347,6355 | Крит. Хеннана-Куинна | 344,0837 | | |

Наша модель имеет следующий вид: $Y = b_0 + b_1 * X_2 + b_2 * X_4 = -14,6329 + 4,30607 * X_2 - 0,448024 * X_4$

При возрастании X2 на 1 единицу своего измерения Y в среднем возрастает на 4,31 единиц своего измерения.

При возрастании X4 на 1 единицу своего измерения Y в среднем убывает на 0,45 единиц своего измерения.

$$t(47, 0,025) = 2,012$$

| | коэффициент | 95 доверительный интервал |
|-------|-------------|---------------------------|
| const | -14,6329 | [-23,2970, -5,96885] |
| X2 | 4,30607 | [3,20221, 5,40993] |
| X4 | -0,448024 | [-0,758281, -0,137768] |

3. Значимость модели в целом

H0: модель незначима в целом $b_1 = b_2 = 0$

H1: модель значима в целом $b_1 \neq 0$ $b_2 \neq 0$

P-значение (F) 3,35e-24

Принимаем H1 с вероятностью 99%.

4. Значимость коэффициентов в модели

H0: $b_1 = 0$ – незначим

H1: $b_1 \neq 0$

$p < 0,0001$ – значим на 99%

H0: $b_2 = 0$ незначим

H1: $b_2 \neq 0$

$p = 0,0056$ – значим на 99%

| Описательная статистика, использованы наблюдения 1 - 50 | | | | | | |
|---|---------|---------|-----------|-------|-------|--|
| Переменная | Среднее | Медиана | ст. откл. | Мин. | Макс. | |
| Y | 52,9 | 58,5 | 22,2 | 11,5 | 96,0 | |
| X2 | 13,0 | 13,6 | 3,67 | 5,59 | 21,1 | |
| X4 | -26,1 | -27,6 | 13,1 | -63,1 | -5,70 | |

5. Эластичности

$$\varepsilon(X_2, y) = b_1 * X_{2cp} / Y_{cp} = 4,30607 * 13 / 52,9 \approx 1,058$$

$$\varepsilon(X_4, y) = b_2 * X_{4cp} / Y_{cp} = -0,448024 * -26,1 / 52,9 \approx 0,221$$

При увеличении X2 на 1%, зависимая переменная Y увеличивается на 1,058%.

При увеличении X4 на 1%, зависимая переменная Y увеличивается на 0,221%.

6. Уравнение в стандартизированной форме и выводы о влиянии переменных

Стандартизированные коэффициенты

$$t(X_2) = b_1 * sX_2 / sY = 4,30607 * 3,67 / 22,2 \approx 0,719$$

$$t(X_4) = b_2 * sX_4 / sY = -0,448024 * 13,1 / 22,2 \approx -0,264$$

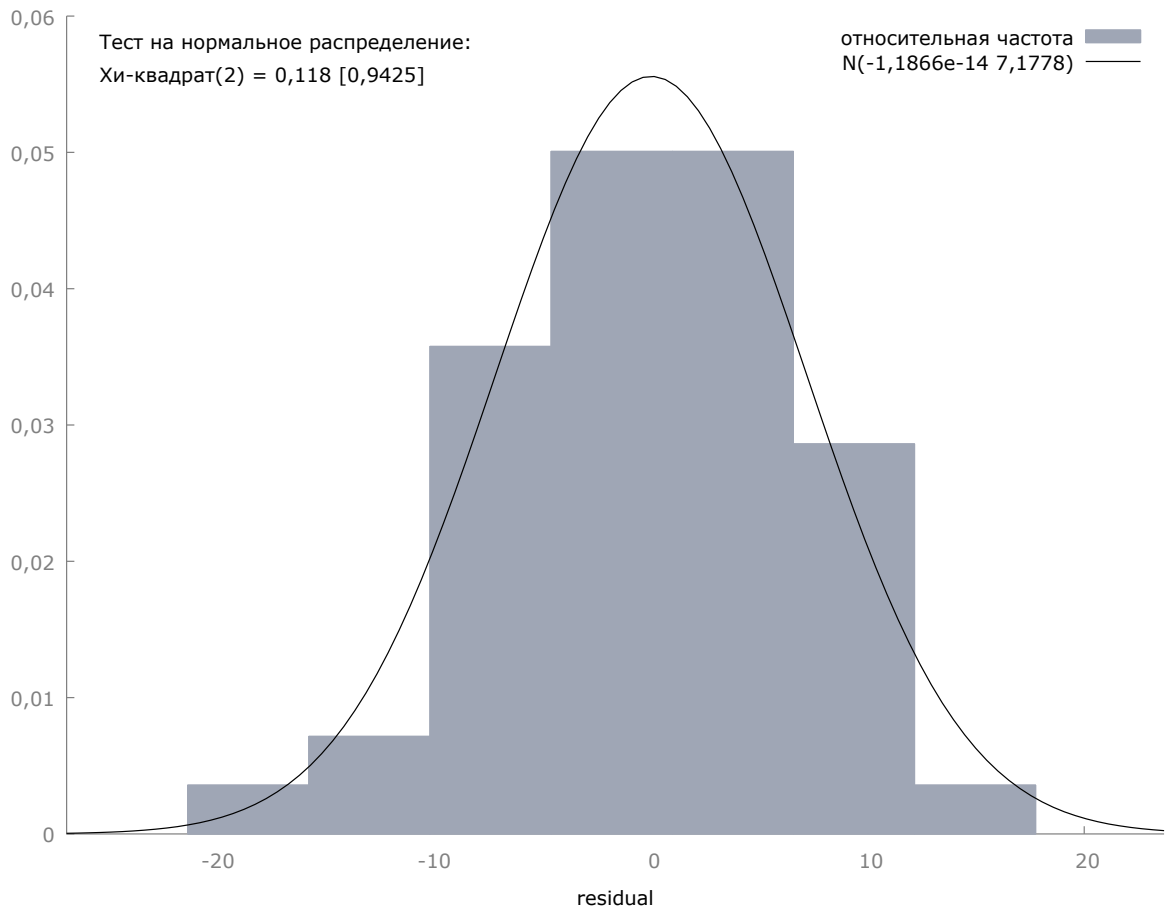
$Y = t(X_2) * X_2 + t(X_4) * X_4 = 0,719 * X_2 - 0,264 * X_4$ – уравнение в стандартизированных коэффициентах

Влияние X2 больше влияния X4, так как модуль стандартизированного коэффициента больше.

7. Нормальность остатков – 5 критериев, нормальная вероятностная бумага, асимметрия

H0: остатки имеют нормальное распределение

H1: распределение остатков отличается от нормального



$p = 0,9425$ – нет оснований отвергнуть H_0

Тест на нормальное распределение uhat6:

Тест Дурника-Хансена (Doornik-Hansen) = 0,118356, p-значение 0,942539

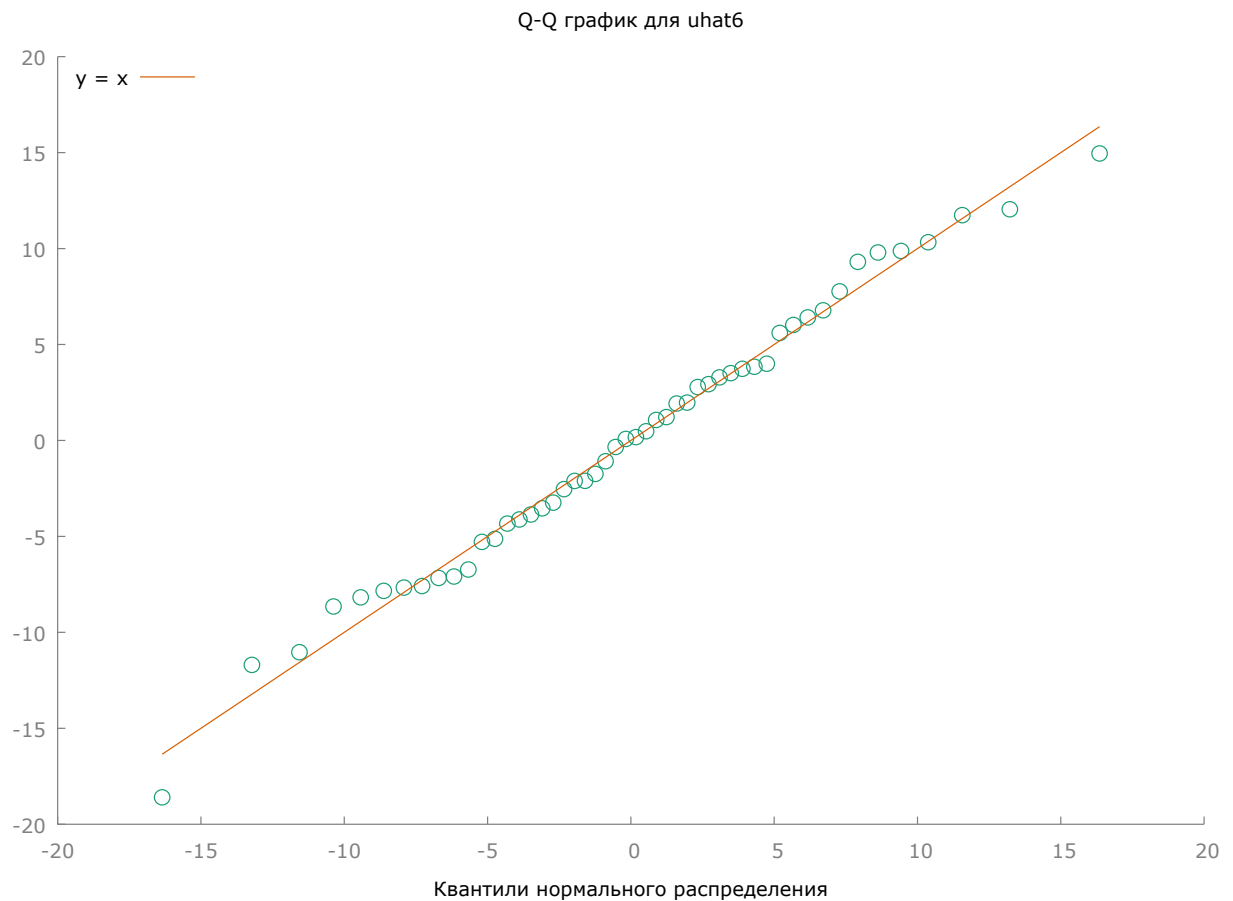
Тест Шапиро-Уилка (Shapiro-Wilk W) = 0,989608, p-значение 0,936593

Тест Лиллифорса (Lilliefors) = 0,0506472, p-значение ≈ 1

Тест Жарка-Бера (Jarque-Bera) = 0,181848, p-значение 0,913087

Все тесты показывают, что нет оснований отвергнуть H_0 , то есть в модели нет систематической ошибки.

Нормальная вероятностная бумага (Q-Q график):



В целом, остатки модели близки к нормальному распределению, большинство точек лежит очень близко к прямой к линии.

Описательная статистика, использованы наблюдения 1 - 50
для переменной uhat6 (50 наблюдений)

| Среднее | Медиана | Минимум | Максимум |
|--------------|-------------|-----------------------|-------------------|
| -1,1866e-014 | 0,12714 | -18,598 | 14,957 |
| Ст. откл. | Вариация | Асимметрия | Эксцесс |
| 7,0298 | 5,9243e+014 | -0,077132 | -0,25197 |
| 5% Проц. | 95% Проц. | Межквартильный размах | Пропущенные набл. |
| -11,333 | 11,879 | 9,5687 | 0 |

Асимметрия остатков близка к нулю, распределение остатков достаточно симметрично.

8. Гетероскедастичность +поправки

H0: гомоскедастичность остатков (дисперсия ошибок постоянна)

H1: гетероскедастичность остатков (дисперсия ошибок изменяется с изменением значений предикторов)

Тест Вайта (White) на гетероскедастичность
МНК, использованы наблюдения 1-50
Зависимая переменная: uhat^2

| | коэффициент | ст. ошибка | t-статистика | p-значение | |
|-------|-------------|------------|--------------|------------|-----|
| const | -36,3661 | 81,8583 | -0,4443 | 0,6590 | |
| X2 | 25,2918 | 17,2381 | 1,467 | 0,1494 | |
| X4 | 5,39064 | 4,75714 | 1,133 | 0,2633 | |
| sq_X2 | 0,312752 | 1,03885 | 0,3011 | 0,7648 | |
| X2_X3 | 1,15469 | 0,641799 | 1,799 | 0,0789 | * |
| sq_X4 | 0,327967 | 0,104276 | 3,145 | 0,0030 | *** |

Неисправленный R-квадрат = 0,424379

Тестовая статистика: $TR^2 = 21,218940$,
p-значение = $P(\chi^2(5) > 21,218940) = 0,000736$

В нашей модели есть гетероскедастичность, нужно сделать поправку.

Модель 9: С поправкой на гетероскедастичность, использованы наблюдения 1-50

Зависимая переменная: Y
Без квадратов в уравнении дисперсии

| | Коэффициент <i>m</i> | Ст. ошибка | t- статистика | p-значение | |
|-------|-------------------------|------------|------------------|------------|-----|
| const | -11,4252 | 4,44991 | -2,568 | 0,0135 | ** |
| X2 | 4,44004 | 0,532404 | 8,340 | <0,0001 | *** |
| X4 | -0,281577 | 0,132912 | -2,119 | 0,0394 | ** |

Статистика, полученная по взвешенным данным:

| | | | |
|--------------------|-----------|----------------------|----------|
| Сумма кв. остатков | 142,2165 | Ст. ошибка модели | 1,739507 |
| R-квадрат | 0,919092 | Исправ. R-квадрат | 0,915649 |
| F(2, 47) | 266,9531 | P-значение (F) | 2,18e-26 |
| Лог. правдоподобие | -97,08012 | Крит. Акаике | 200,1602 |
| Крит. Шварца | 205,8963 | Крит. Хеннана-Куинна | 202,3446 |

Статистика, полученная по исходным данным:

| | | | |
|------------------------|----------|--------------------------|----------|
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 |
| Сумма кв. остатков | 2593,029 | Ст. ошибка модели | 7,427707 |

0,532404
0,132912

0,548708
0,154223

Как можно заметить, стандартные ошибки уменьшились.

9. R^2 и его интерпретация

Исправ. R-квадрат 0,915649

R квадрат достаточно высокий, модель хорошо объясняет вариацию зависимой переменной.

10. Прогноз на средние значения и доверительный интервал прогноза

$$X2_{pr} = X2_{sr}$$

$$X4_{pr} = X4_{sr}$$

Средняя абсолютная процентная ошибка (MAPE) - 18,042

Вывод: MAPE \approx 18%, модель хорошая по качеству

Для 95% доверительных интервалов, $t(47, 0,025) = 2,012$

| Набл. | Y | прогнози- вание | ст. ошибка | 95% доверительный интервал |
|-------|------------------|--------------------|------------|-------------------------------|
| усг | не определено | 53,6445 | 2,00858 | (49,6037, 57,6852) |

$Y_{pr} = 53,6445$ своего измерения

Доверительный интервал (49,6037; 57,6852) с вероятностью 95% прогнозное значение попадает в этот интервал.

11. Сравним две парные модели и одну множественную по качеству

Модель 11: МНК, использованы наблюдения 1-50

Зависимая переменная: Y

| | Коэффициен т | Ст. ошибка | t- статистика | p-значение | |
|------------------------|-----------------|--------------------------|------------------|------------|-----|
| const | -20,7362 | 4,04041 | -5,132 | <0,0001 | *** |
| X2 | 5,67820 | 0,300125 | 18,92 | <0,0001 | *** |
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 | | |
| Сумма кв. остатков | 2856,279 | Ст. ошибка модели | 7,714001 | | |
| R-квадрат | 0,881757 | Исправ. R-квадрат | 0,879294 | | |
| F(1, 48) | 357,9446 | P-значение (F) | 6,79e-24 | | |
| Лог. правдоподобие | -172,0782 | Крит. Акаике | 348,1564 | | |
| Крит. Шварца | 351,9805 | Крит. Хеннана-Куинна | 349,6127 | | |

Модель 12: МНК, использованы наблюдения 1-50
Зависимая переменная: Y

| | <i>Коэффициент</i> <i>t</i> | <i>Ст. ошибка</i> | <i>t-</i> <i>статистика</i> | <i>p-значение</i> | |
|------------------------|--------------------------------|--------------------------|--------------------------------|-------------------|-----|
| const | 14,0175 | 3,43626 | 4,079 | 0,0002 | *** |
| X4 | -1,48984 | 0,118056 | -12,62 | <0,0001 | *** |
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | | 22,20317 | |
| Сумма кв. остатков | 5594,425 | Ст. ошибка модели | | 10,79586 | |
| R-квадрат | 0,768405 | Исправ. R-квадрат | | 0,763580 | |
| F(1, 48) | 159,2583 | P-значение (F) | | 7,37e-17 | |
| Лог. правдоподобие | -188,8845 | Крит. Акаике | | 381,7690 | |
| Крит. Шварца | 385,5930 | Крит. Хеннана-Куинна | | 383,2252 | |

Модель 13: С поправкой на гетероскедастичность, использованы наблюдения 1-50
Зависимая переменная: Y

| | <i>Коэффициент</i> <i>t</i> | <i>Ст. ошибка</i> | <i>t-</i> <i>статистика</i> | <i>p-значение</i> | |
|-------|--------------------------------|-------------------|--------------------------------|-------------------|-----|
| const | -14,2724 | 4,15000 | -3,439 | 0,0012 | *** |
| X4 | -0,354383 | 0,144444 | -2,453 | 0,0179 | ** |
| X2 | 4,52545 | 0,475173 | 9,524 | <0,0001 | *** |

Статистика, полученная по взвешенным данным:

| | | | |
|--------------------|-----------|----------------------|----------|
| Сумма кв. остатков | 136,8598 | Ст. ошибка модели | 1,706432 |
| R-квадрат | 0,896647 | Исправ. R-квадрат | 0,892249 |
| F(2, 47) | 203,8769 | P-значение (F) | 6,86e-24 |
| Лог. правдоподобие | -96,12028 | Крит. Акаике | 198,2406 |
| Крит. Шварца | 203,9766 | Крит. Хеннана-Куинна | 200,4249 |

Статистика, полученная по исходным данным:

| | | | |
|------------------------|----------|--------------------------|----------|
| Среднее завис. перемен | 52,86680 | Ст. откл. завис. перемен | 22,20317 |
| Сумма кв. остатков | 2472,540 | Ст. ошибка модели | 7,253084 |

Как можно заметить, модель парной регрессии X2 и Y лучше, чем модель парной регрессии X4 и Y. Множественная регрессия с участием этих переменных показывает наилучший результат.