

1. Построить модель, отражающую зависимость почасовой заработной платы от образования и умственных способностей (модель 1).

$$\ln(EARNINGS) = \beta_0 + \beta_1 S + \beta_2 ASVAVC + \varepsilon \quad (1)$$

Модель 1: МНК, использованы наблюдения 1-540  
Зависимая переменная: l\_EARNINGS

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	0,803680	0,134704	5,966	<0,0001	***
S	0,0795582	0,0105132	7,567	<0,0001	***
ASVABC	0,0175999	0,00289245	6,085	<0,0001	***
Среднее завис. перемен	2,796854	Ст. откл. завис. перемен		0,605414	
Сумма кв. остатков	138,9801	Ст. ошибка модели		0,508732	
R-квадрат	0,296507	Исправ. R-квадрат		0,293887	
F(2, 537)	113,1670	P-значение (F)		9,75e-42	
Лог. правдоподобие	-399,7725	Крит. Акаике		805,5450	
Крит. Шварца	818,4197	Крит. Хеннана-Куинна		810,5802	

2. Оценить полученное уравнение, проверить его качество, дать интерпретацию коэффициентам уравнения.

В полученной модели заметим, что p-значение у каждой переменной меньше 0,01 – переменные значимы с 99% вероятностью.

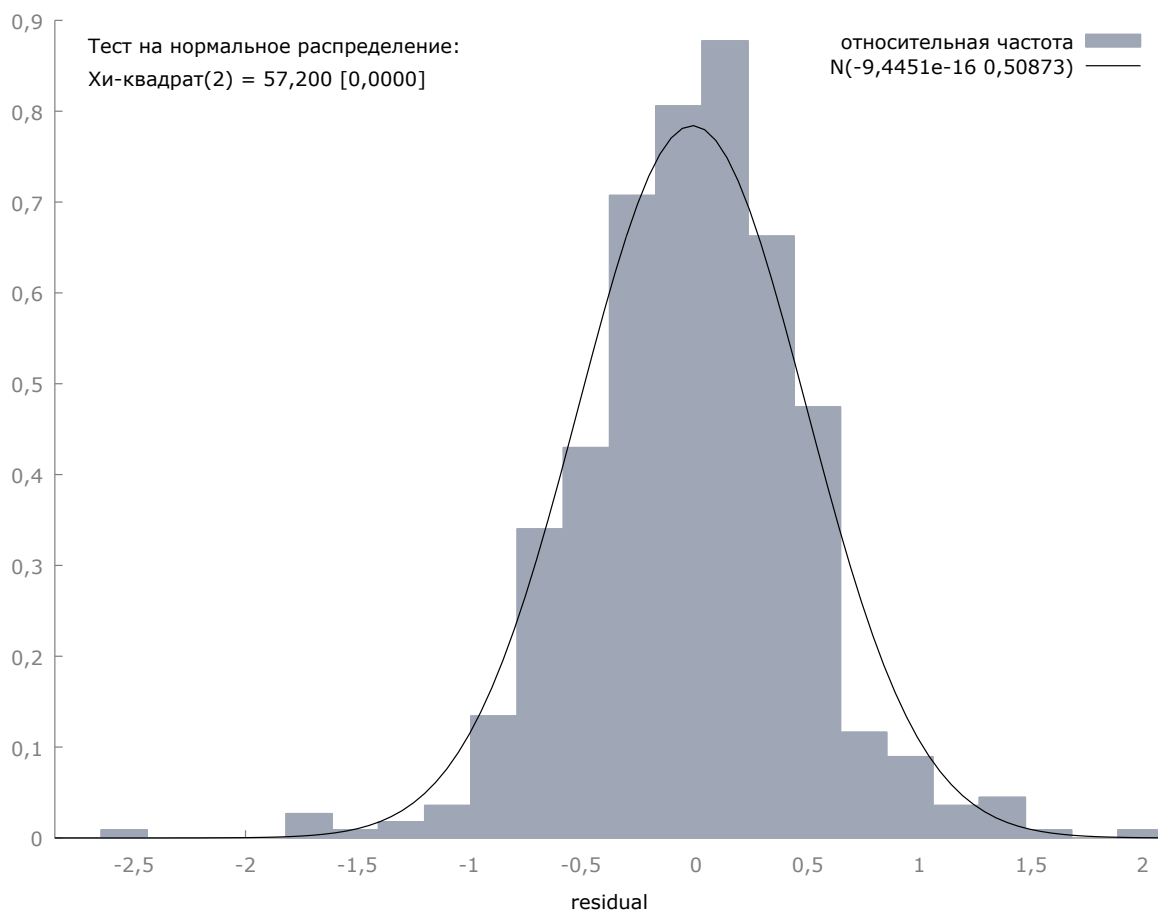
Увеличения числа завершенных лет обучения на один год приводит к увеличению текущей заработной платы на ~8%.

Улучшение результатов тестов на познавательные способности на 1 балл приводит к увеличению текущей заработной платы на ~1,8%.

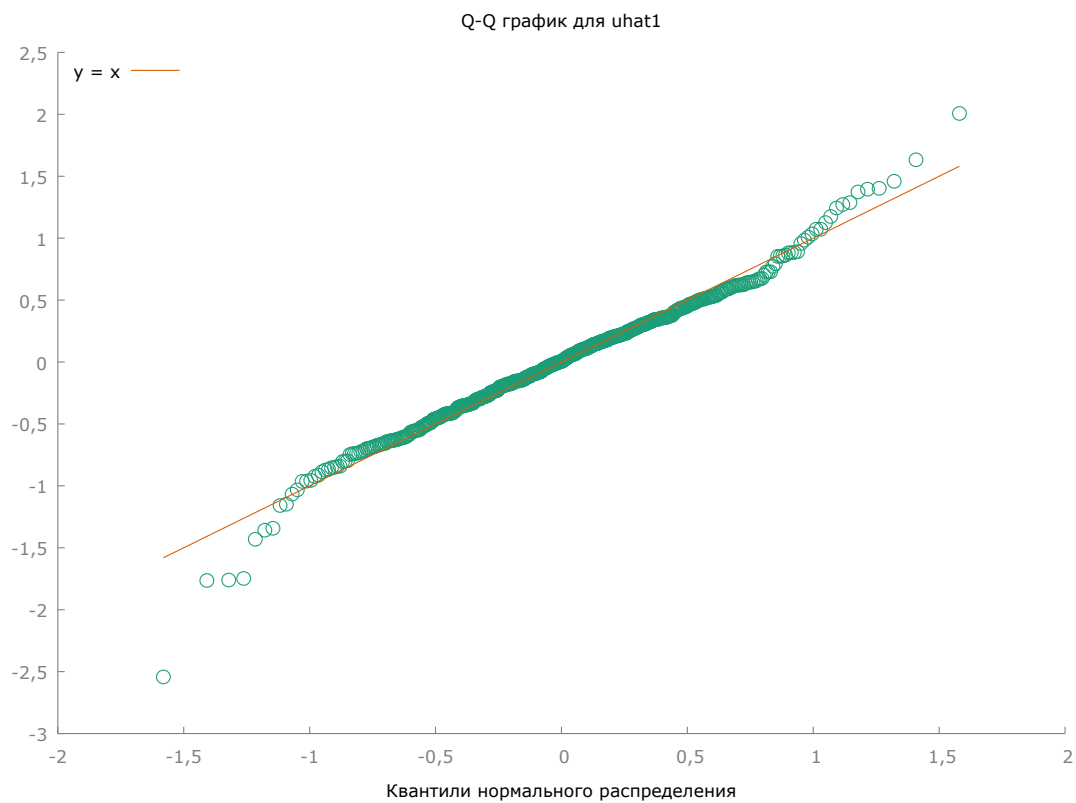
### 1. Нормальность.

H0: остатки имеют нормальное распределение

H1: распределение остатков отличается от нормального



Р-значение меньше 0,01 – распределение остатков отличается от нормального.



## 2. Гетероскедастичность.

H0 – гомоскедастичность

H1 – гетероскедастичность

Тест Вайта (White) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	p-значение
const	0,212835	0,678284	0,3138	0,7538
S	-0,0409145	0,0622970	-0,6568	0,5116
ASVABC	0,00859957	0,0252268	0,3409	0,7333
sq_S	-0,00232851	0,00311927	-0,7465	0,4557
X2_X3	0,00217845	0,00175719	1,240	0,2156
sq_ASVABC	-0,000347657	0,000341405	-1,018	0,3090

Неисправленный R-квадрат = 0,011045

Тестовая статистика:  $TR^2 = 5,964373$ ,

p-значение =  $P(\text{Chi-квадрат}(5) > 5,964373) = 0,309701$

Тест Вайта (White) на гетероскедастичность (только квадраты)

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	p-значение
const	-0,0640650	0,640776	-0,09998	0,9204
S	-0,00986317	0,0570687	-0,1728	0,8629
ASVABC	0,0102028	0,0252063	0,4048	0,6858
sq_S	0,000628719	0,00201093	0,3127	0,7547
sq_ASVABC	-6,94621e-05	0,000257427	-0,2698	0,7874

Неисправленный R-квадрат = 0,008199

Тестовая статистика:  $TR^2 = 4,427334$ ,

p-значение =  $P(\text{Chi-квадрат}(4) > 4,427334) = 0,351251$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	p-значение
const	-0,0713193	0,526463	-0,1355	0,8923
S	0,0284512	0,0410885	0,6924	0,4890
ASVABC	0,0132985	0,0113045	1,176	0,2400

Объясненная сумма квадратов = 17,0063

Тестовая статистика:  $LM = 8,503170$ ,

p-значение =  $P(\text{Chi-квадрат}(2) > 8,503170) = 0,014242$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	p-значение	
const	-0,275726	0,135496	-2,035	0,0423	**

S	0,00732249	0,0105750	0,6924	0,4890
ASVABC	0,00342263	0,00290945	1,176	0,2400

Объясненная сумма квадратов = 1,12649

Тестовая статистика: LM = 4,291554,  
p-значение = P(Chi-квадрат(2) > 4,291554) = 0,116977

В тесте Бройша-Пэгана p-значение оказалось меньше 0,05 и больше 0,01 – в модели с 95% вероятностью присутствует гетероскедастичность, тем не менее, остальные тесты показали отсутствие гетероскедастичности, однако все равно будем использовать модель с робастными стандартными ошибками, чтобы удостовериться в отсутствии гетероскедастичности.

Модель 3: МНК, использованы наблюдения 1-540

Зависимая переменная: l\_EARNINGS

Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>p-значение</i>	
	<i>m</i>				
const	0,803680	0,132345	6,073	<0,0001	***
S	0,0795582	0,0102975	7,726	<0,0001	***
ASVABC	0,0175999	0,00261492	6,731	<0,0001	***
Среднее завис. перемен	2,796854	Ст. откл. завис. перемен		0,605414	
Сумма кв. остатков	138,9801	Ст. ошибка модели		0,508732	
R-квадрат	0,296507	Исправ. R-квадрат		0,293887	
F(2, 537)	109,9826	P-значение (F)		9,25e-41	
Лог. правдоподобие	-399,7725	Крит. Акаике		805,5450	
Крит. Шварца	818,4197	Крит. Хеннана-Куинна		810,5802	

P-значение немного увеличилось, значимость уравнения немного снизилась.

### 3. MAPE.

Статистика для оценки прогноза использовано наблюдений - 540

Средняя ошибка (ME)	-1,3771e-015
Корень из средней квадратичной ошибки (RMSE)	0,50732
Средняя абсолютная ошибка (MAE)	0,38734
Средняя процентная ошибка (MPE)	-4,2699
<b>Средняя абсолютная процентная ошибка (MAPE)</b>	<b>15,755</b>
U-статистика Тейла (Theil's U)	0,089353

*MAPE < 10% отличное качество*

*10% < MAPE < 20% хорошее качество*

*20% < MAPE < 30% удовл.*

*Если больше 30% плохое качество модели*

Наша модель имеет хорошее качество.

#### **4. Значимость коэффициентов.**

H0 – коэффициент не значим

H1 – коэффициент значим

В нашей модели все р-значения при независимых переменных меньше 0,01 – все коэффициенты значимы с 99% вероятностью.

#### **5. Значимость модели в целом.**

H0 – модель незначима в целом

H1 – модель значима в целом

Р-значение (F) 9,25e-41

Р-значение меньше 0,01 – модель значима в целом с вероятностью 99%.

#### **6. Интерпретация коэффициентов.**

Увеличения числа завершенных лет обучения на один год приводит к увеличению текущей заработной платы на ~8%.

Улучшение результатов тестов на познавательные способности на 1 балл приводит к увеличению текущей заработной платы на ~1,8%.

#### **7. Доверительные интервалы для коэффициентов и их интерпретация.**

$$t(537, 0,05) = 1,648$$

	коэффициент	90 доверительный интервал
const	0,803680	[0,585616, 1,02174]
S	0,0795582	[0,0625910, 0,0965254]
ASVABC	0,0175999	[0,0132913, 0,0219085]

$$t(537, 0,025) = 1,964$$

коэффициент 95 доверительный интервал

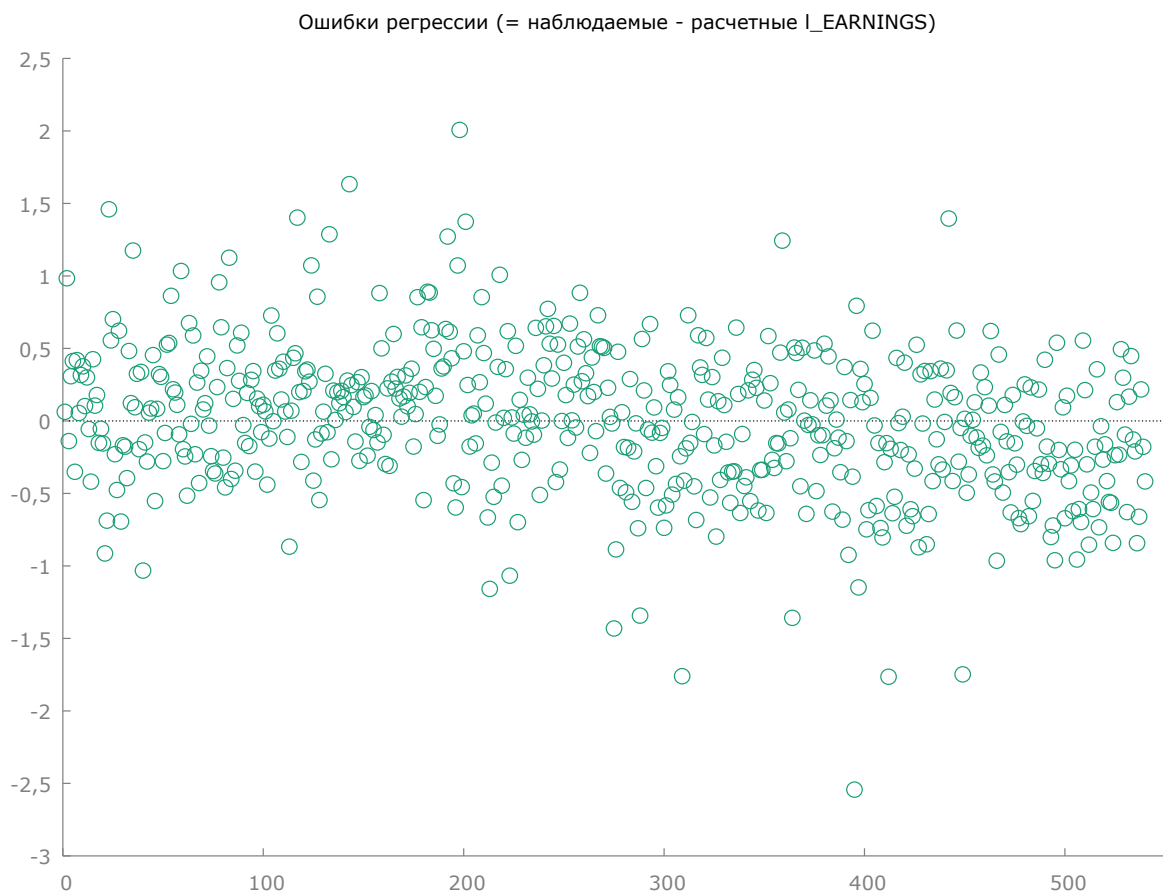
const	0,803680	[0,543703, 1,06366]
S	0,0795582	[0,0593298, 0,0997866]
ASVABC	0,0175999	[0,0124632, 0,0227367]

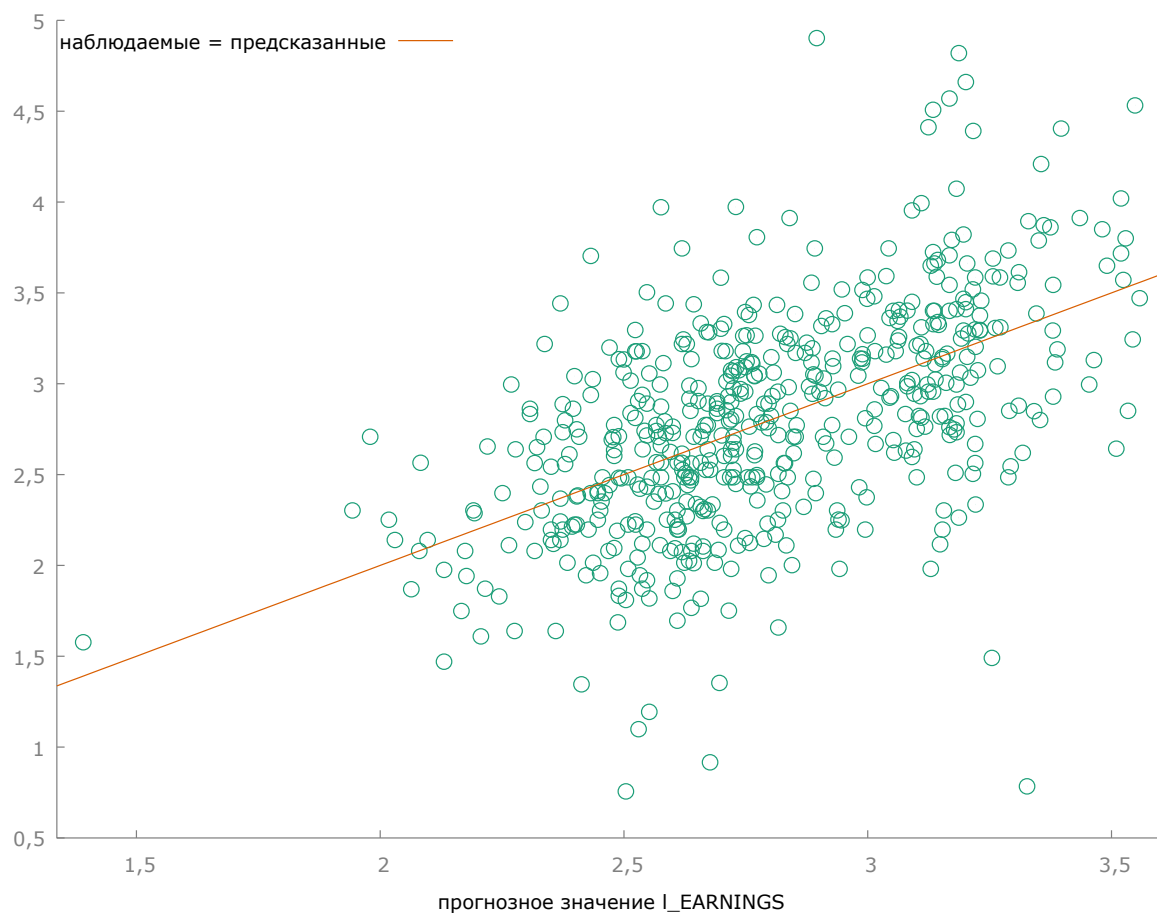
$$t(537, 0,005) = 2,585$$

	коэффициент	99 доверительный интервал
const	0,803680	[0,461566, 1,14579]
S	0,0795582	[0,0529389, 0,106177]
ASVABC	0,0175999	[0,0108403, 0,0243595]

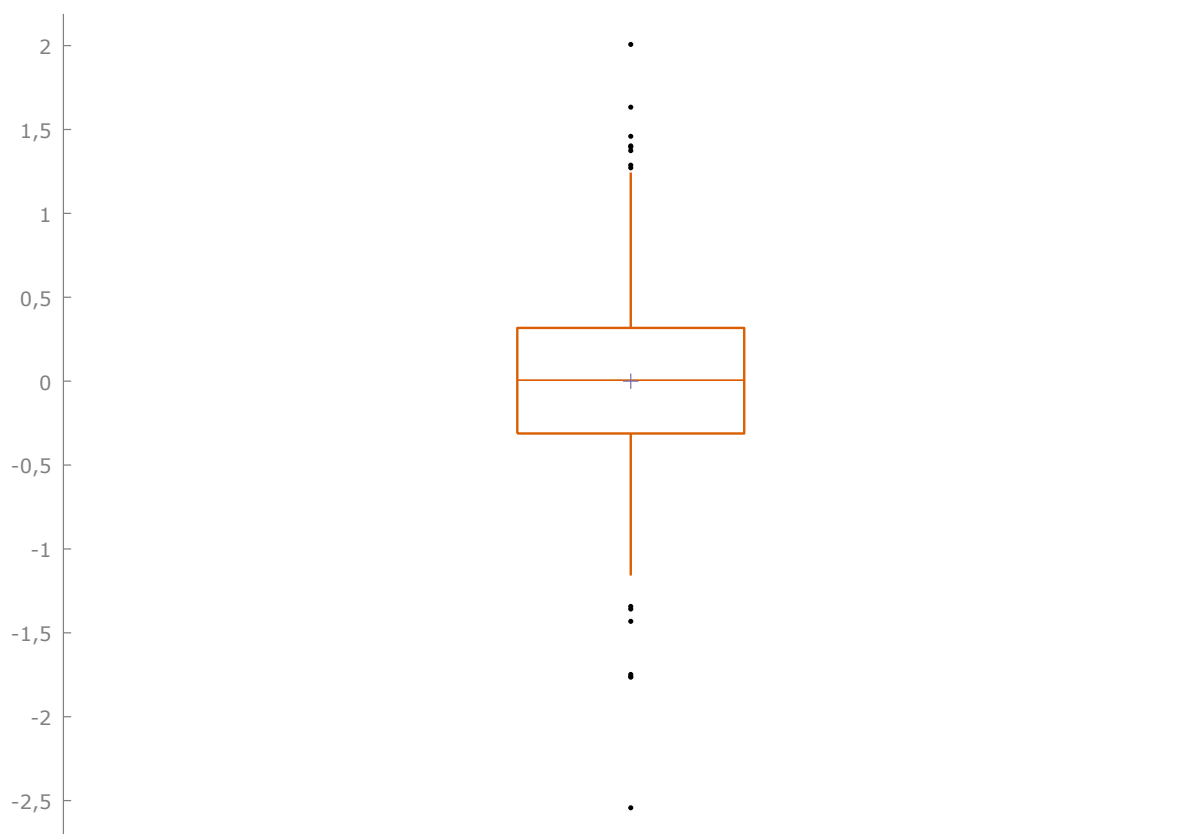
Как можно заметить, чем больше степень уверенности доверительного интервала, тем больше отрезок, в который попадают коэффициенты.

## 8. Графики.





Ошибки регрессии (= наблюдаемые - расчетные I\_EARNINGS)



3. Проверить, есть различие в уровне заработной платы для мужчин и женщин, введя фиктивную переменную Male (модель 2). Дать интерпретацию полученным результатам.

$$\ln(EARNINGS) = \beta_0 + \beta_1 S + \beta_2 ASVAVC + \delta \cdot Male + \varepsilon \quad (2)$$

Модель 4: МНК, использованы наблюдения 1-540

Зависимая переменная: l\_EARNINGS

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	0,668337	0,127892	5,226	<0,0001	***
S	0,0822500	0,00990599	8,303	<0,0001	***
ASVABC	0,0161575	0,00272943	5,920	<0,0001	***
MALE	0,344433	0,0413178	8,336	<0,0001	***
Среднее завис. перемен	2,796854	Ст. откл. завис. перем		0,605414	
Сумма кв. остатков	123,0294	Ст. ошибка модели		0,479095	
R-квадрат	0,377247	Исправ. R-квадрат		0,373762	
F(3, 536)	108,2315	P-значение (F)		8,58e-55	
Лог. правдоподобие	-366,8574	Крит. Акаике		741,7147	
Крит. Шварца	758,8810	Крит. Хеннана-Куинна		748,4284	

- Коэффициент: Значение 0,344433 показывает, что заработная плата мужчин в среднем больше на 34.44% по сравнению с женщинами, при прочих равных условиях (после учета других переменных S и ASVABC).

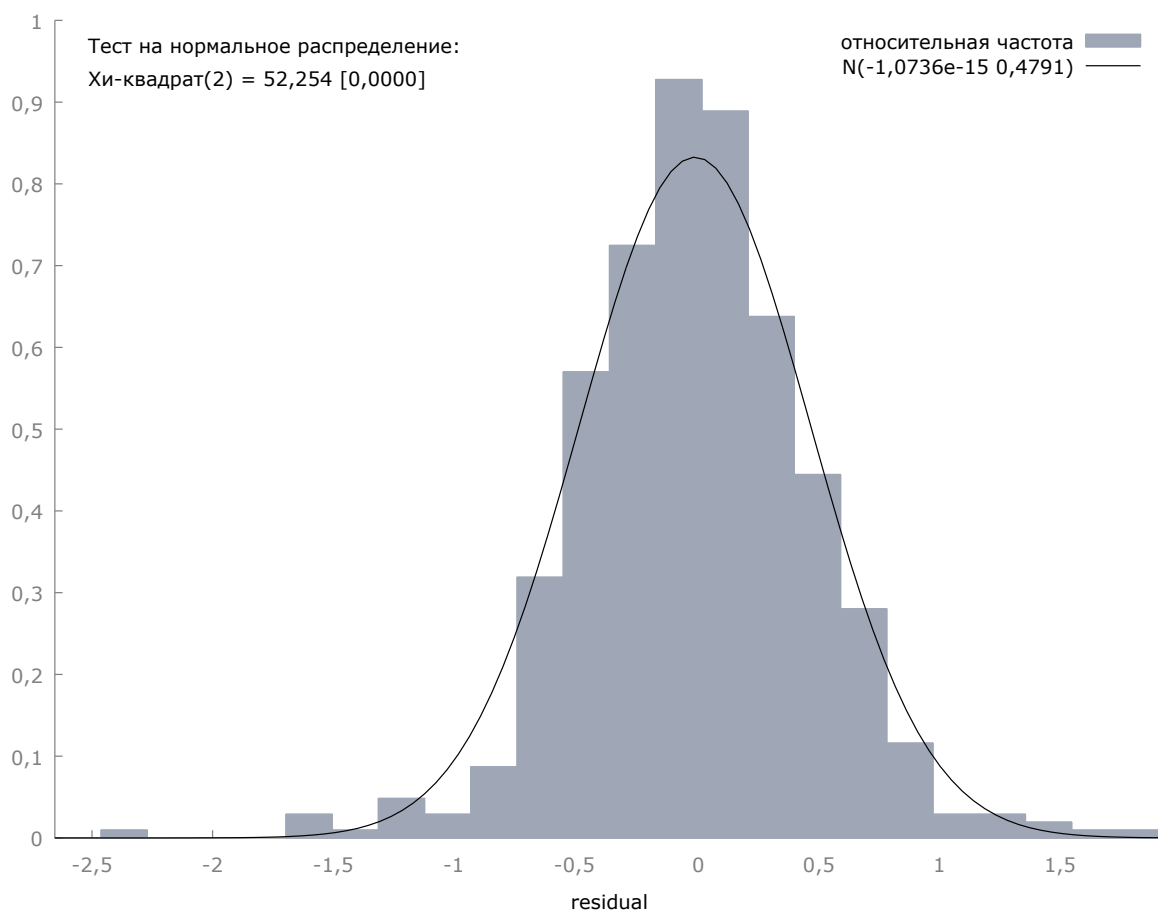
- Статистическая значимость: Низкое p-значение (<0,0001) говорит о том, что влияние этой переменной статистически значимо и мы можем с 99% уверенностью утверждать, что различие в заработной плате между мужчинами и женщинами действительно существует в данной выборке.

## 1. Нормальность.

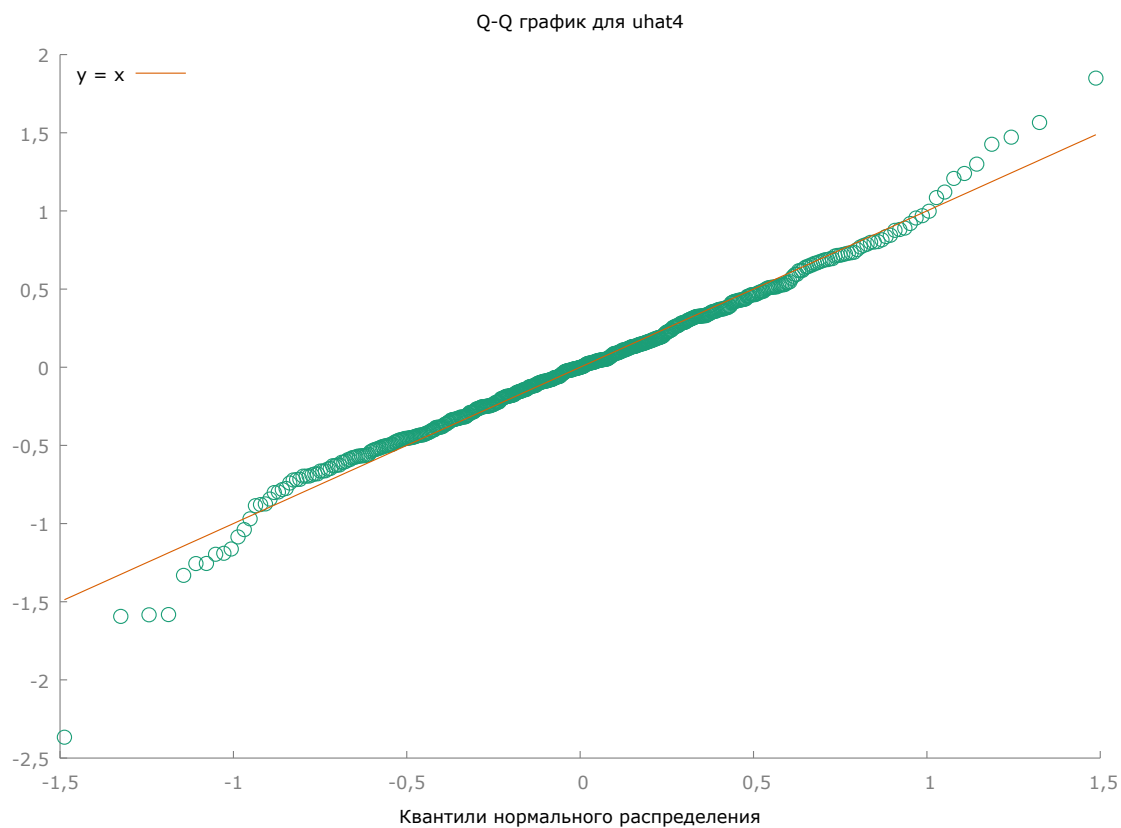
H0: остатки имеют нормальное распределение

H1: распределение остатков отличается от нормального





Р-значение меньше 0,01 – распределение остатков отличается от нормального.



## 2. Гетероскедастичность.

H0 – гомоскедастичность

H1 – гетероскедастичность

Тест Вайта (White) на гетероскедастичность

МНК, использованы наблюдения 1-540

Зависимая переменная:  $\hat{u}^2$

	коэффициент	ст. ошибка	t-статистика	p-значение
const	0,0299763	0,613431	0,04887	0,9610
S	-0,0387479	0,0544129	-0,7121	0,4767
ASVABC	0,0144663	0,0223944	0,6460	0,5186
MALE	0,0571123	0,238947	0,2390	0,8112
sq_S	-0,00190132	0,00274049	-0,6938	0,4881
X2_X3	0,00192082	0,00153590	1,251	0,2116
X2_X4	0,00459827	0,0187217	0,2456	0,8061
sq_ASVABC	-0,000369718	0,000299169	-1,236	0,2171
X3_X4	-0,00293319	0,00514724	-0,5699	0,5690

Неисправленный R-квадрат = 0,013643

Тестовая статистика:  $TR^2 = 7,367485$ ,

p-значение =  $P(\chi^2(8) > 7,367485) = 0,497552$

Тест Вайта (White) на гетероскедастичность (только квадраты)

МНК, использованы наблюдения 1-540

Зависимая переменная:  $\hat{u}^2$

	коэффициент	ст. ошибка	t-статистика	p-значение
const	-0,157390	0,564965	-0,2786	0,7807
S	-0,0115253	0,0497045	-0,2319	0,8167
ASVABC	0,0146546	0,0221976	0,6602	0,5094
MALE	-0,0275595	0,0389834	-0,7070	0,4799
sq_S	0,000786025	0,00175200	0,4486	0,6539
sq_ASVABC	-0,000127885	0,000226994	-0,5634	0,5734

Неисправленный R-квадрат = 0,010016

Тестовая статистика:  $TR^2 = 5,408537$ ,

p-значение =  $P(\chi^2(5) > 5,408537) = 0,368079$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

Зависимая переменная: масштабированное  $\hat{u}^2$

	коэффициент	ст. ошибка	t-статистика	p-значение
const	-0,00872535	0,522206	-0,01671	0,9867
S	0,0423411	0,0404480	1,047	0,2957
ASVABC	0,00967576	0,0111448	0,8682	0,3857
MALE	-0,135731	0,168708	-0,8045	0,4214

Объясненная сумма квадратов = 19,147

Тестовая статистика:  $LM = 9,573509$ ,

p-значение =  $P(\chi^2(3) > 9,573509) = 0,022562$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	p-значение	
const	-0,229820	0,118975	-1,932	0,0539	*
S	0,00964666	0,00921535	1,047	0,2957	
ASVABC	0,00220445	0,00253914	0,8682	0,3857	
MALE	-0,0309239	0,0384371	-0,8045	0,4214	

Объясненная сумма квадратов = 0,993874

Тестовая статистика: LM = 4,994050,

p-значение =  $P(\text{Хи-квадрат}(3) > 4,994050) = 0,172233$

В тесте Бройша-Пэгана p-значение оказалось меньше 0,05 и больше 0,01 – в модели с 95% вероятностью присутствует гетероскедастичность, тем не менее, остальные тесты показали отсутствие гетероскедастичности, однако все равно будем использовать модель с робастными стандартными ошибками, чтобы удостовериться в отсутствии гетероскедастичности.

Модель 5: МНК, использованы наблюдения 1-540

Зависимая переменная:  $\ln\_EARNINGS$

Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

	Коэффициент	Ст. ошибка	t-статистика	p-значение	
const	0,668337	0,125256	5,336	<0,0001	***
S	0,0822500	0,00979781	8,395	<0,0001	***
ASVABC	0,0161575	0,00245120	6,592	<0,0001	***
MALE	0,344433	0,0413905	8,322	<0,0001	***

Среднее завис. перемен	2,796854	Ст. откл. завис. перемен	0,605414
Сумма кв. остатков	123,0294	Ст. ошибка модели	0,479095
R-квадрат	0,377247	Исправ. R-квадрат	0,373762
F(3, 536)	107,5915	P-значение (F)	1,56e-54
Лог. правдоподобие	-366,8574	Крит. Акаике	741,7147
Крит. Шварца	758,8810	Крит. Хеннана-Куинна	748,4284

P-значение немного увеличилось, значимость уравнения немного снизилась.

### 3. MAPE.

Статистика для оценки прогноза использовано наблюдений - 540

Средняя ошибка (ME) -1,3417e-015

Корень из средней квадратичной ошибки (RMSE) 0,47732

Средняя абсолютная ошибка (MAE) 0,36256

Средняя процентная ошибка (MPE)	-3,7995
Средняя абсолютная процентная ошибка (MAPE)	14,673
U-статистика Тейла (Theil's U)	0,083991

*MAPE < 10% отличное качество*

*10% < MAPE < 20% хорошее качество*

*20% < MAPE < 30% удовл.*

*Если больше 30% плохое качество модели*

Наша модель имеет хорошее качество.

#### **4. Значимость коэффициентов.**

H0 – коэффициент не значим

H1 – коэффициент значим

В нашей модели все р-значения при независимых переменных и фиктивной переменной MALE меньше 0,01 – все коэффициенты значимы с 99% вероятностью.

#### **5. Значимость модели в целом.**

H0 – модель незначима в целом

H1 – модель значима в целом

Р-значение (F) 1,56e-54

Р-значение меньше 0,01 – модель значима в целом с вероятностью 99%.

#### **6. Интерпретация коэффициентов.**

Увеличения числа завершенных лет обучения на один год приводит к увеличению текущей заработной платы на ~8%.

Улучшение результатов тестов на познавательные способности на 1 балл приводит к увеличению текущей заработной платы на ~1,6%.

Значение 0,344433 показывает, что заработная плата мужчин в среднем больше на 34.44% по сравнению с женщинами, при прочих равных условиях (после учета других переменных S и ASVABC).

#### **7. Доверительные интервалы для коэффициентов и их интерпретация.**

$$t(536, 0,05) = 1,648$$

	коэффициент	90 доверительный интервал
const	0,668337	[0,461952, 0,874722]
S	0,0822500	[0,0661061, 0,0983938]
ASVABC	0,0161575	[0,0121187, 0,0201964]
MALE	0,344433	[0,276234, 0,412633]

$$t(536, 0,025) = 1,964$$

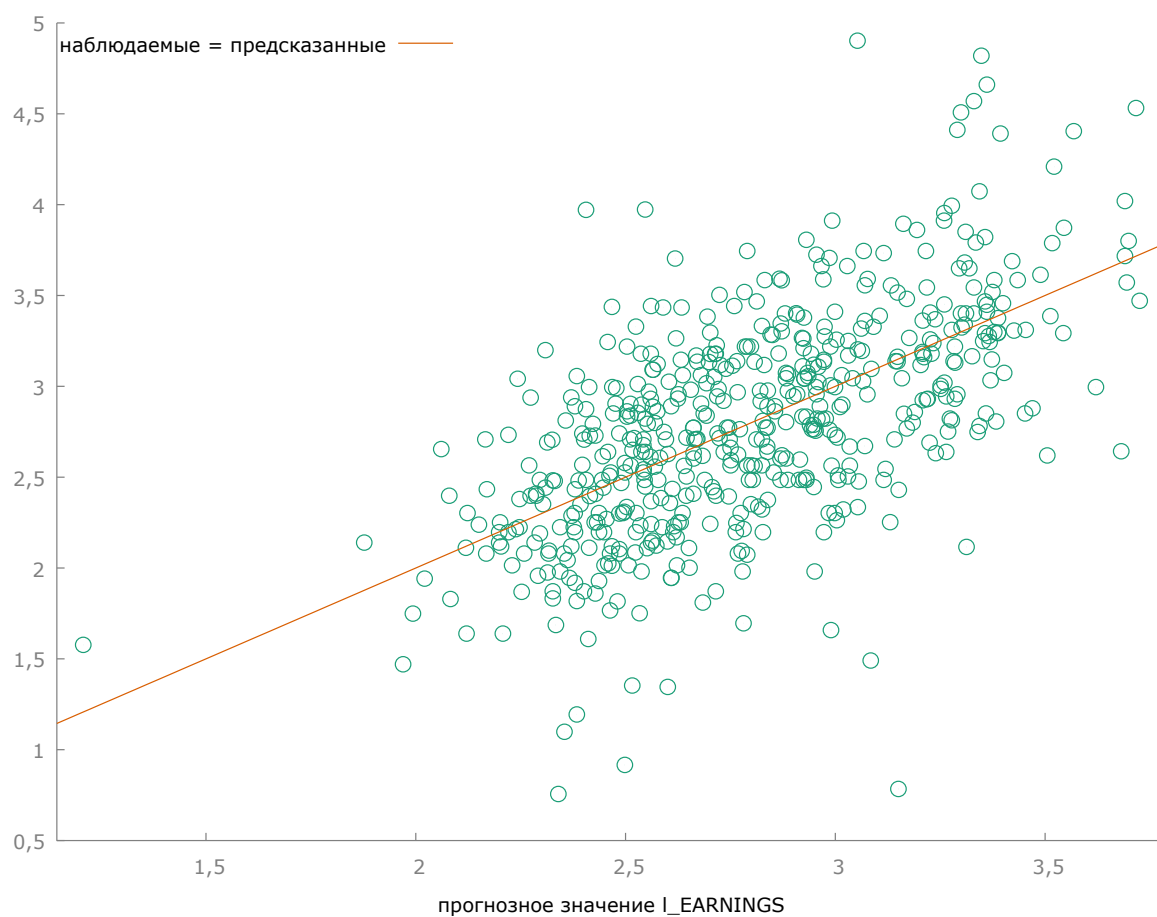
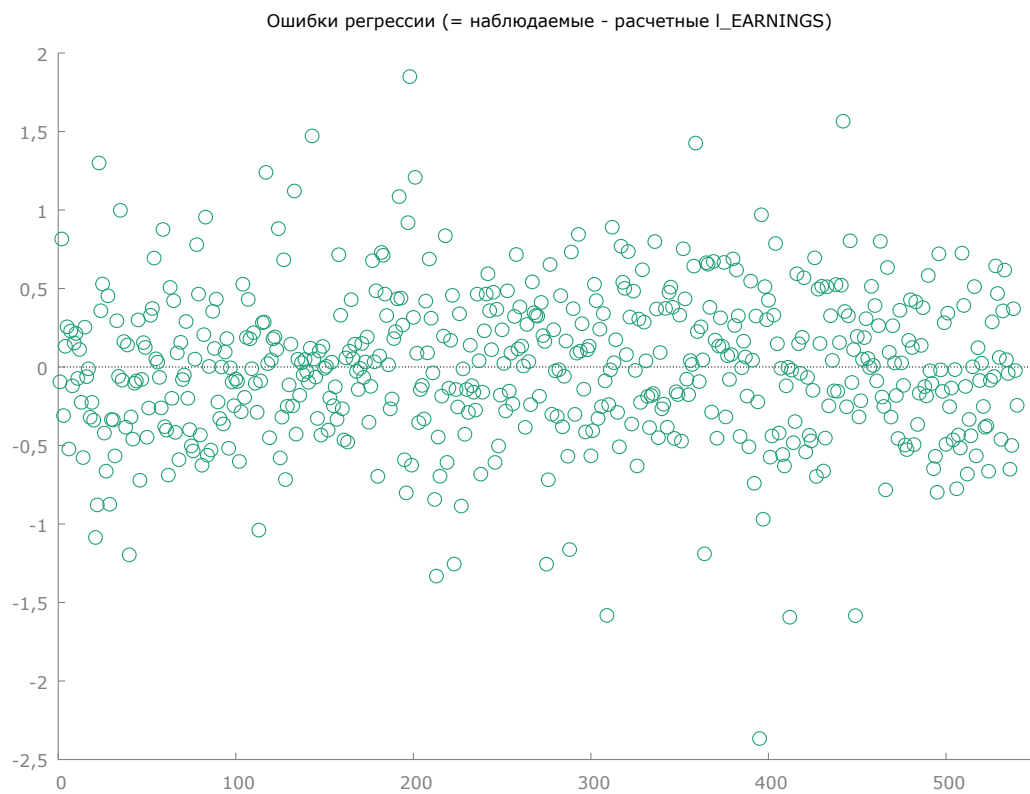
	коэффициент	95 доверительный интервал
const	0,668337	[0,422284, 0,914390]
S	0,0822500	[0,0630032, 0,101497]
ASVABC	0,0161575	[0,0113424, 0,0209727]
MALE	0,344433	[0,263126, 0,425741]

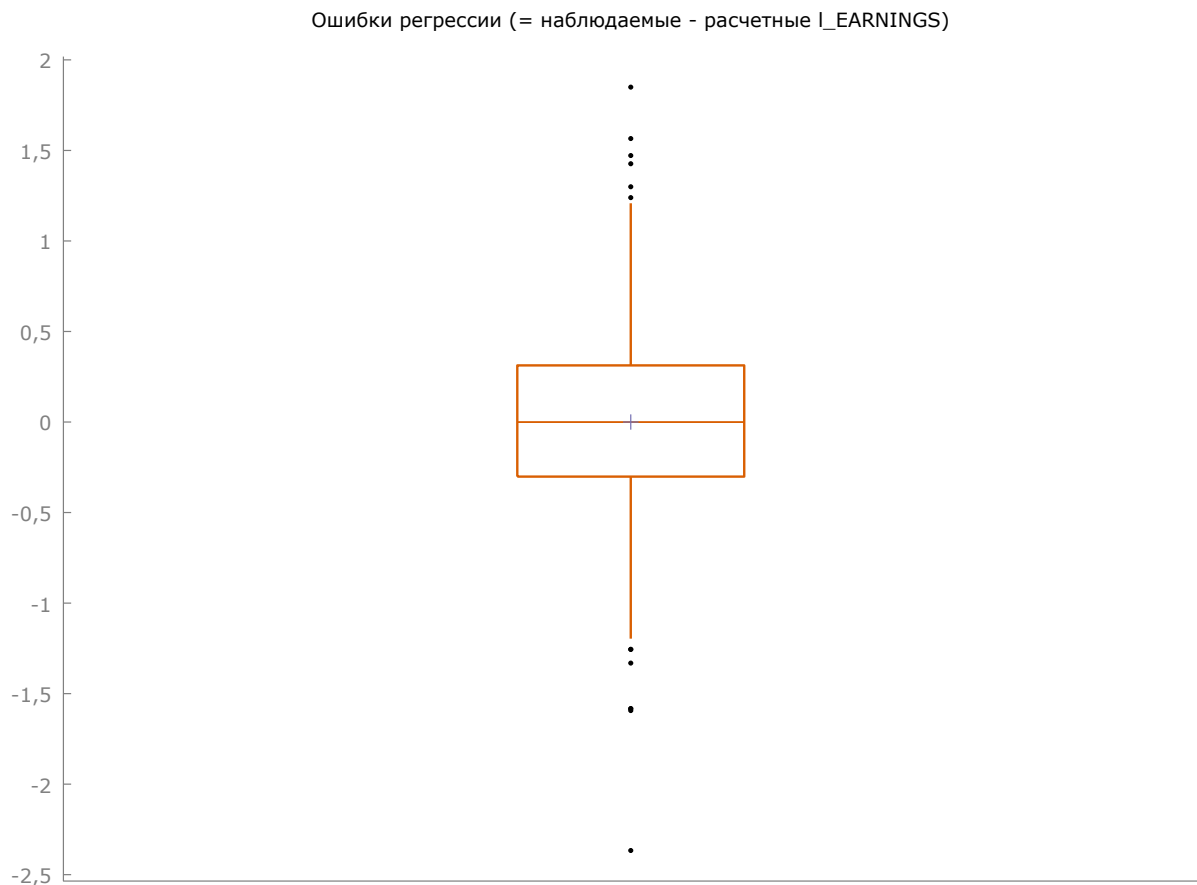
$$t(536, 0,005) = 2,585$$

	коэффициент	99 доверительный интервал
const	0,668337	[0,344546, 0,992128]
S	0,0822500	[0,0569223, 0,107578]
ASVABC	0,0161575	[0,00982109, 0,0224940]
MALE	0,344433	[0,237438, 0,451429]

Как можно заметить, чем больше степень уверенности доверительного интервала, тем больше отрезок, в который попадают коэффициенты.

## 8. Графики.





4. Проверить, влияет ли на уровень заработной платы этническое происхождение. Построить модель с фиктивными переменными ETHBLACK и ETHHISP, выбрав в качестве эталонной переменной ETHWHITE (модель 3). Проинтерпретировать коэффициенты при фиктивных переменных.

$$\ln(EARNINGS) = \beta_0 + \beta_1 S + \beta_2 ASVAVC + \delta \cdot Male + \gamma_1 ETHBLACK + \gamma_2 ETHHISP + \varepsilon \quad (3)$$

Модель 6: МНК, использованы наблюдения 1-540  
Зависимая переменная: I\_EARNINGS

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	0,659993	0,140041	4,713	<0,0001	***
S	0,0819166	0,0100996	8,111	<0,0001	***
ASVABC	0,0163784	0,00304340	5,382	<0,0001	***
MALE	0,344517	0,0414384	8,314	<0,0001	***
ETHBLACK	0,0128639	0,0721992	0,1782	0,8587	

ETHHISP	0,00231189	0,0968797	0,02386	0,9810
---------	------------	-----------	---------	--------

Среднее завис. перемен	2,796854	Ст. откл. завис. перемен	0,605414
Сумма кв. остатков	123,0220	Ст. ошибка модели	0,479977
R-квадрат	0,377284	Исправ. R-квадрат	0,371454
F(5, 534)	64,70682	P-значение (F)	9,20e-53
Лог. правдоподобие	-366,8413	Крит. Акаике	745,6826
Крит. Шварца	771,4320	Крит. Хеннана-Куинна	755,7531

$\ln Y = b_1 + b_2 * S + b_3 * ASVAVC + b_4 * MALE + b_5 * ETHBLACK + b_6 * ETHHISP + e$

$H_0: b_5 = b_6 = 0$

$H_1: b_5 \neq 0$  или  $b_6 \neq 0$

Множество ограничений

1:  $b[ETHBLACK] = 0$

2:  $b[ETHHISP] = 0$

Тестовая статистика: Робастный  $F(2, 534) = 0,0163792$ , с p-значением = 0,983755

Оценка с учетом ограничений:

	коэффициент	ст. ошибка	t-статистика	p-значение
const	0,668337	0,127892	5,226	2,49e-07 ***
S	0,0822500	0,00990599	8,303	8,31e-016 ***
ASVABC	0,0161575	0,00272943	5,920	5,76e-09 ***
MALE	0,344433	0,0413178	8,336	6,49e-016 ***
ETHBLACK	0,000000	0,000000	NA	NA
ETHHISP	0,000000	0,000000	NA	NA

Стандартная ошибка регрессии = 0,479095

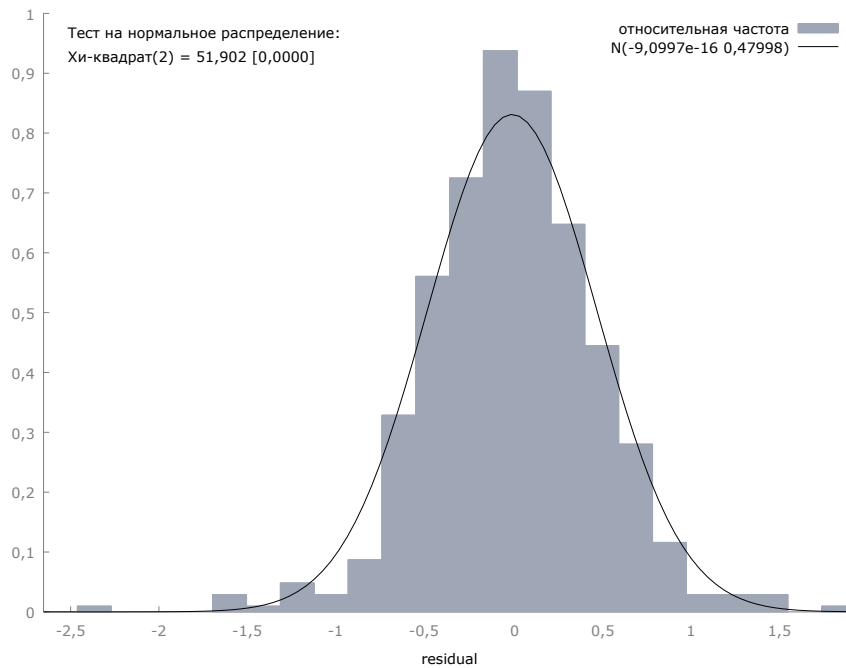
$H_0$  принимаем, происхождение не влияет на заработную плату.



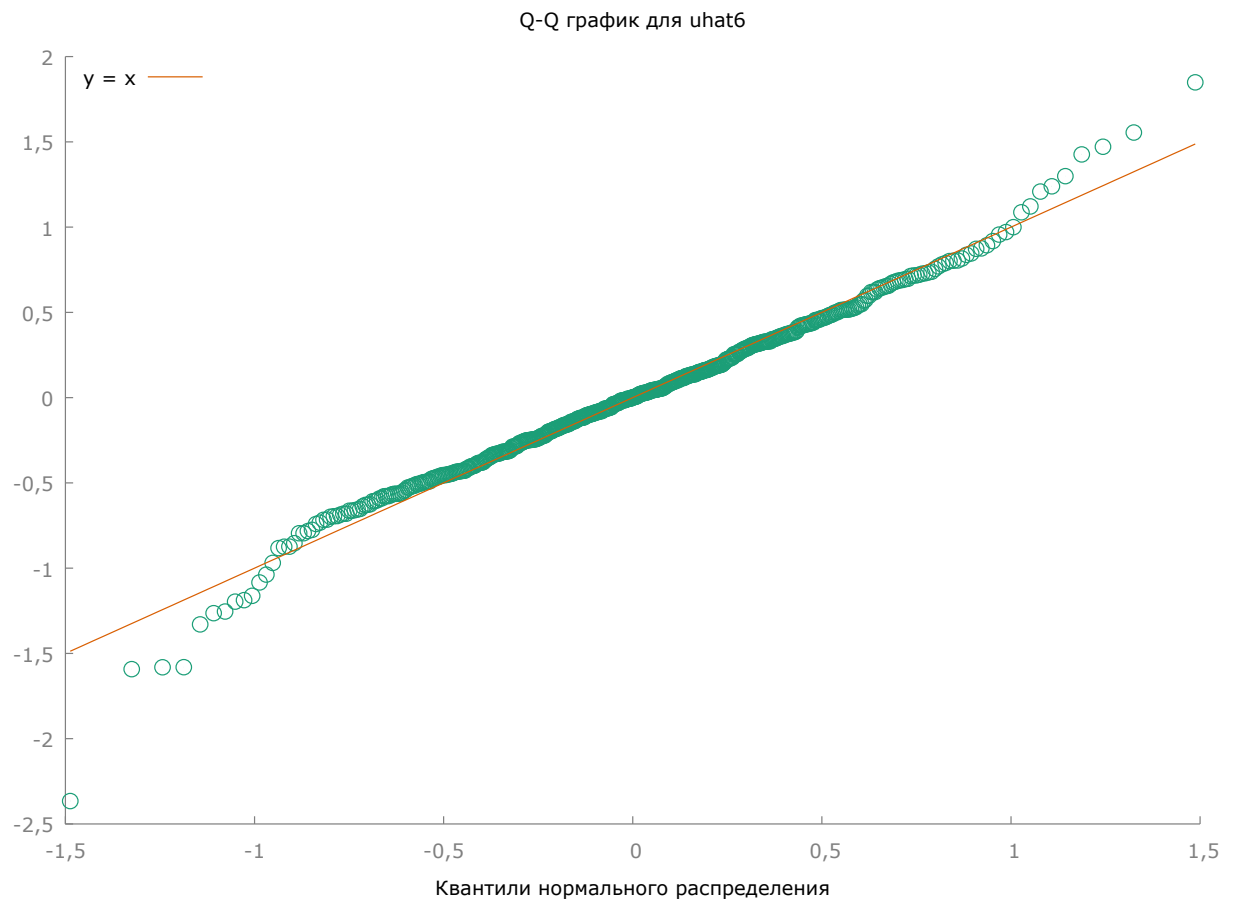
## 1. Нормальность.

H0: остатки имеют нормальное распределение

H1: распределение остатков отличается от нормального



P-значение меньше 0,01 – распределение остатков отличается от нормального.



## 2. Гетероскедастичность.

$H_0$  – гомоскедастичность

$H_1$  – гетероскедастичность

Тест Вайта (White) на гетероскедастичность

МНК, использованы наблюдения 1-540

Зависимая переменная: uhat<sup>2</sup>

	коэффициент	ст. ошибка	t-статистика	p-значение
const	-0,269364	0,759623	-0,3546	0,7230
S	-0,0240203	0,0755268	-0,3180	0,7506
ASVABC	0,0216770	0,0264542	0,8194	0,4129
MALE	0,00147122	0,266225	0,005526	0,9956
ETHBLACK	0,444883	0,449133	0,9905	0,3224
ETHHISP	0,327708	0,492961	0,6648	0,5065
sq_S	-0,00123872	0,00327699	-0,3780	0,7056
X2_X3	0,00137330	0,00164839	0,8331	0,4052
X2_X4	0,00287515	0,0192806	0,1491	0,8815
X2_X5	-0,0514240	0,0391899	-1,312	0,1900
X2_X6	-0,00598049	0,0483579	-0,1237	0,9016
sq_ASVABC	-0,000372204	0,000340111	-1,094	0,2743
X3_X4	-0,00157335	0,00580029	-0,2713	0,7863
X3_X5	0,00579397	0,0100733	0,5752	0,5654
X3_X6	-0,00742960	0,0125068	-0,5940	0,5527
X4_X5	0,0454951	0,137516	0,3308	0,7409
X4_X6	0,0911477	0,188344	0,4839	0,6286

Неисправленный R-квадрат = 0,020024

Тестовая статистика:  $TR^2 = 10,812757$ ,

р-значение =  $P(\chi^2(16) > 10,812757) = 0,820893$

Тест Вайта (White) на гетероскедастичность (только квадраты)

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	р-значение
const	-0,144634	0,588353	-0,2458	0,8059
S	-0,0141466	0,0504410	-0,2805	0,7792
ASVABC	0,0148676	0,0229561	0,6477	0,5175
MALE	-0,0263862	0,0390518	-0,6757	0,4995
ETHBLACK	0,0110716	0,0683162	0,1621	0,8713
ETHHISP	-0,0250514	0,0917141	-0,2731	0,7848
sq_S	0,000867134	0,00177131	0,4895	0,6247
sq_ASVABC	-0,000129381	0,000231920	-0,5579	0,5772

Неисправленный R-квадрат = 0,010236

Тестовая статистика:  $TR^2 = 5,527626$ ,

р-значение =  $P(\chi^2(7) > 5,527626) = 0,595854$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	р-значение
const	0,0198516	0,571202	0,03475	0,9723
S	0,0416799	0,0411946	1,012	0,3121
ASVABC	0,00933879	0,0124135	0,7523	0,4522
MALE	-0,130818	0,169020	-0,7740	0,4393
ETHBLACK	0,0141250	0,294488	0,04796	0,9618
ETHHISP	-0,121574	0,395156	-0,3077	0,7585

Объясненная сумма квадратов = 19,4664

Тестовая статистика: LM = 9,733211,

р-значение =  $P(\chi^2(5) > 9,733211) = 0,083157$

Тест Бройша-Пэгана (Breusch-Pagan) на гетероскедастичность

МНК, использованы наблюдения 1-540

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

	коэффициент	ст. ошибка	t-статистика	р-значение
const	-0,223296	0,130130	-1,716	0,0868 *
S	0,00949546	0,00938491	1,012	0,3121
ASVABC	0,00212755	0,00282803	0,7523	0,4522
MALE	-0,0298028	0,0385060	-0,7740	0,4393
ETHBLACK	0,00321794	0,0670899	0,04796	0,9618
ETHHISP	-0,0276969	0,0900239	-0,3077	0,7585

Объясненная сумма квадратов = 1,01033

Тестовая статистика: LM = 5,087620,

р-значение =  $P(\chi^2(5) > 5,087620) = 0,405281$

В тесте Бройша-Пэгана р-значение оказалось меньше 0,1 и больше 0,05 – в модели с 90% вероятностью присутствует гетероскедастичность, тем не менее, остальные тесты показали отсутствие гетероскедастичности, однако

все равно будем использовать модель с робастными стандартными ошибками, чтобы удостовериться в отсутствии гетероскедастичности.

Модель 7: МНК, использованы наблюдения 1-540

Зависимая переменная: l\_EARNINGS

Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>p-значение</i>	
	<i>m</i>				
const	0,659993	0,132983	4,963	<0,0001	***
S	0,0819166	0,0101608	8,062	<0,0001	***
ASVABC	0,0163784	0,00278693	5,877	<0,0001	***
MALE	0,344517	0,0414481	8,312	<0,0001	***
ETHBLACK	0,0128639	0,0710775	0,1810	0,8564	
ETHHISP	0,00231189	0,0819593	0,02821	0,9775	
Среднее завис. перемен	2,796854	Ст. откл. завис. перемен		0,605414	
Сумма кв. остатков	123,0220	Ст. ошибка модели		0,479977	
R-квадрат	0,377284	Исправ. R-квадрат		0,371454	
F(5, 534)	64,43668	P-значение (F)		1,40e-52	
Лог. правдоподобие	-366,8413	Крит. Акаике		745,6826	
Крит. Шварца	771,4320	Крит. Хеннана-Куинна		755,7531	

P-значение немного увеличилось, значимость уравнения немного снизилась.

### 3. MAPE.

Статистика для оценки прогноза использовано наблюдений - 540

Средняя ошибка (ME)	-1,3565e-015
Корень из средней квадратичной ошибки (RMSE)	0,4773
Средняя абсолютная ошибка (MAE)	0,3627
Средняя процентная ошибка (MPE)	-3,7987
<b>Средняя абсолютная процентная ошибка (MAPE)</b>	<b>14,678</b>
U-статистика Тейла (Theil's U)	0,083989

*MAPE < 10% отличное качество*

*10% < MAPE < 20% хорошее качество*

$20\% < MAPE < 30\%$  удовл.

Если больше 30% плохое качество модели

Наша модель имеет хорошее качество.

#### 4. Значимость коэффициентов.

H0 – коэффициент не значим

H1 – коэффициент значим

В нашей модели все р-значения при независимых переменных и фиктивной переменной MALE меньше 0,01 – эти коэффициенты значимы с 99% вероятностью. Фиктивные переменные ETHBLACK и ETHHISP имеют высокие р-значения, они незначимы.

#### 5. Значимость модели в целом.

H0 – модель незначима в целом

H1 – модель значима в целом

Р-значение (F) 1,40e-52

Р-значение меньше 0,01 – модель значима в целом с вероятностью 99%.

#### 6. Интерпретация коэффициентов.

Увеличения числа завершенных лет обучения на один год приводит к увеличению текущей заработной платы на ~8%.

Улучшение результатов тестов на познавательные способности на 1 балл приводит к увеличению текущей заработной платы на ~1,6%.

Значение 0,34517 показывает, что заработная плата мужчин в среднем больше на 34.52% по сравнению с женщинами, при прочих равных условиях (после учета других переменных S и ASVABC).

Коэффициенты ETHBLACK и ETHHISP незначимы.

#### 7. Доверительные интервалы для коэффициентов и их интерпретация.

$$t(534, 0,05) = 1,648$$

	коэффициент	90 доверительный интервал
const	0,659993	[0,440876, 0,879111]
S	0,0819166	[0,0651745, 0,0986586]
ASVABC	0,0163784	[0,0117864, 0,0209705]

MALE	0,344517	[0,276222, 0,412811]
ETHBLACK	0,0128639	[-0,104251, 0,129979]
ETHHISP	0,00231189	[-0,132733, 0,137357]

$$t(534, 0,025) = 1,964$$

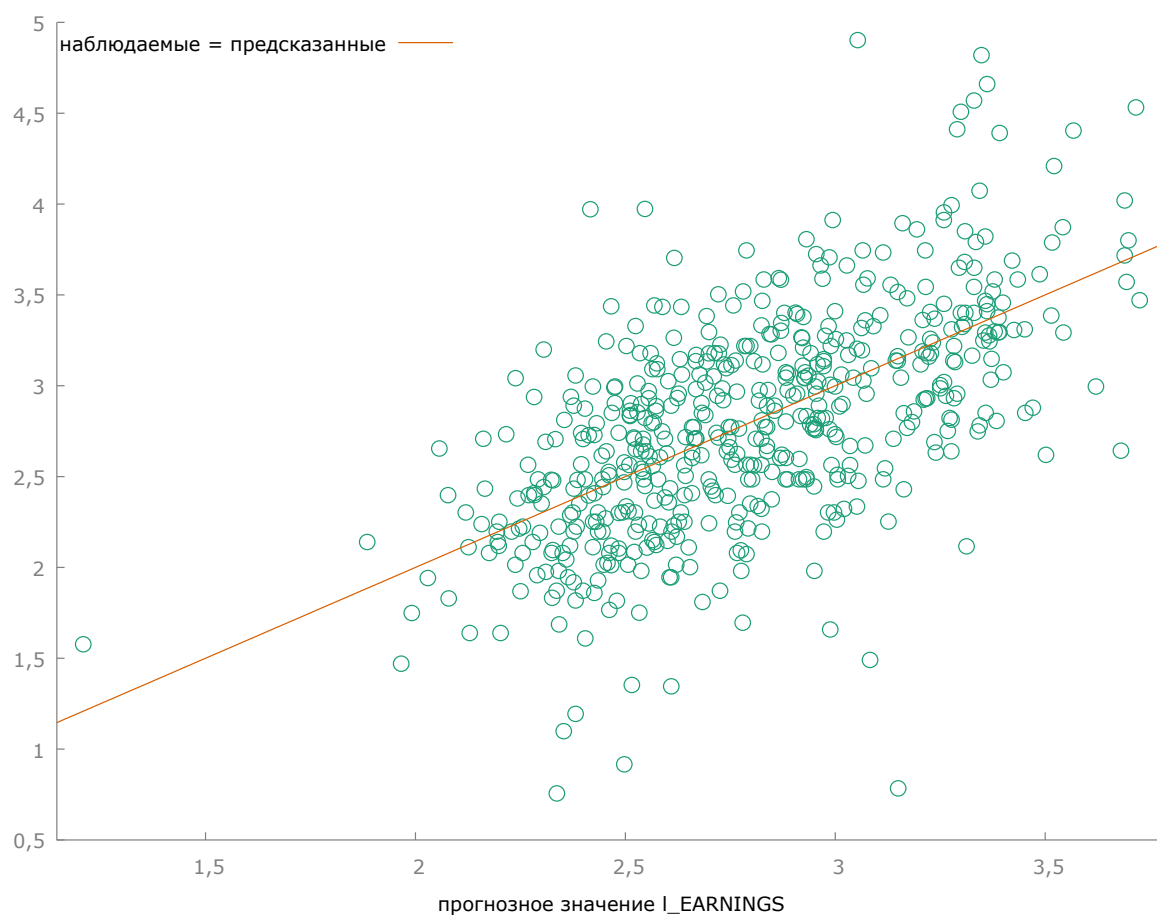
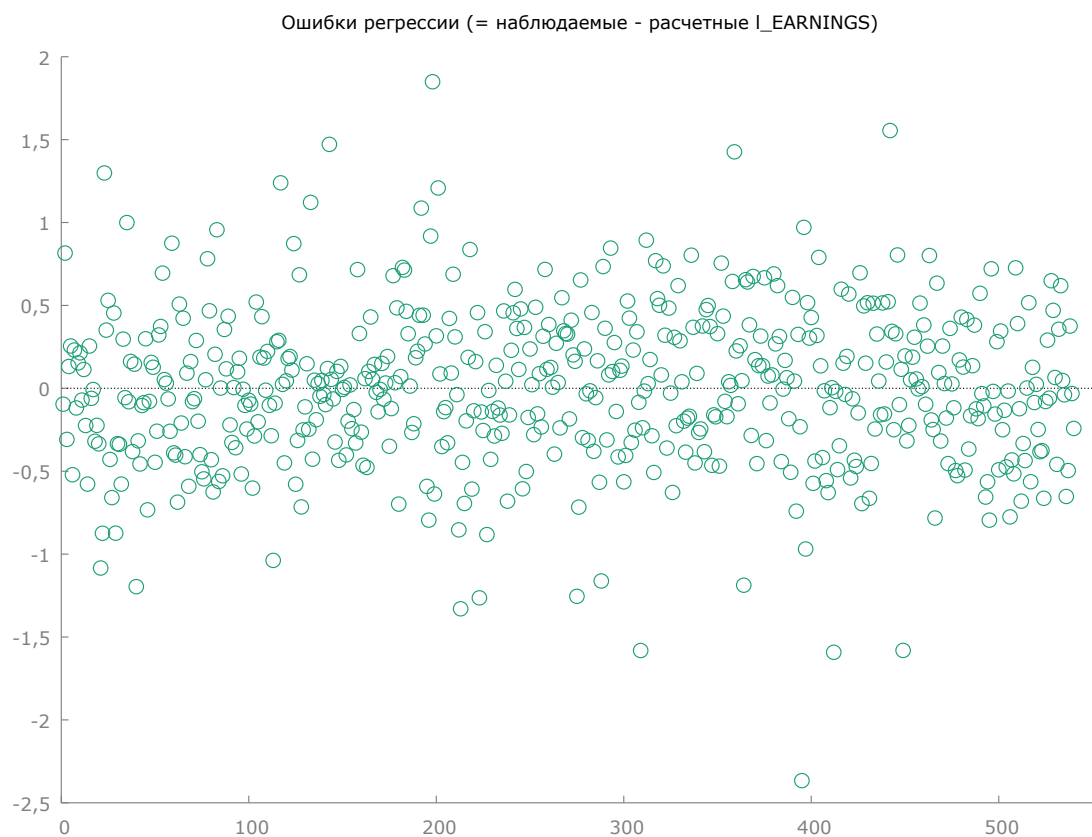
	коэффициент	95 доверительный интервал
const	0,659993	[0,398760, 0,921227]
S	0,0819166	[0,0619565, 0,101877]
ASVABC	0,0163784	[0,0109037, 0,0218531]
MALE	0,344517	[0,263095, 0,425938]
ETHBLACK	0,0128639	[-0,126762, 0,152490]
ETHHISP	0,00231189	[-0,158690, 0,163314]

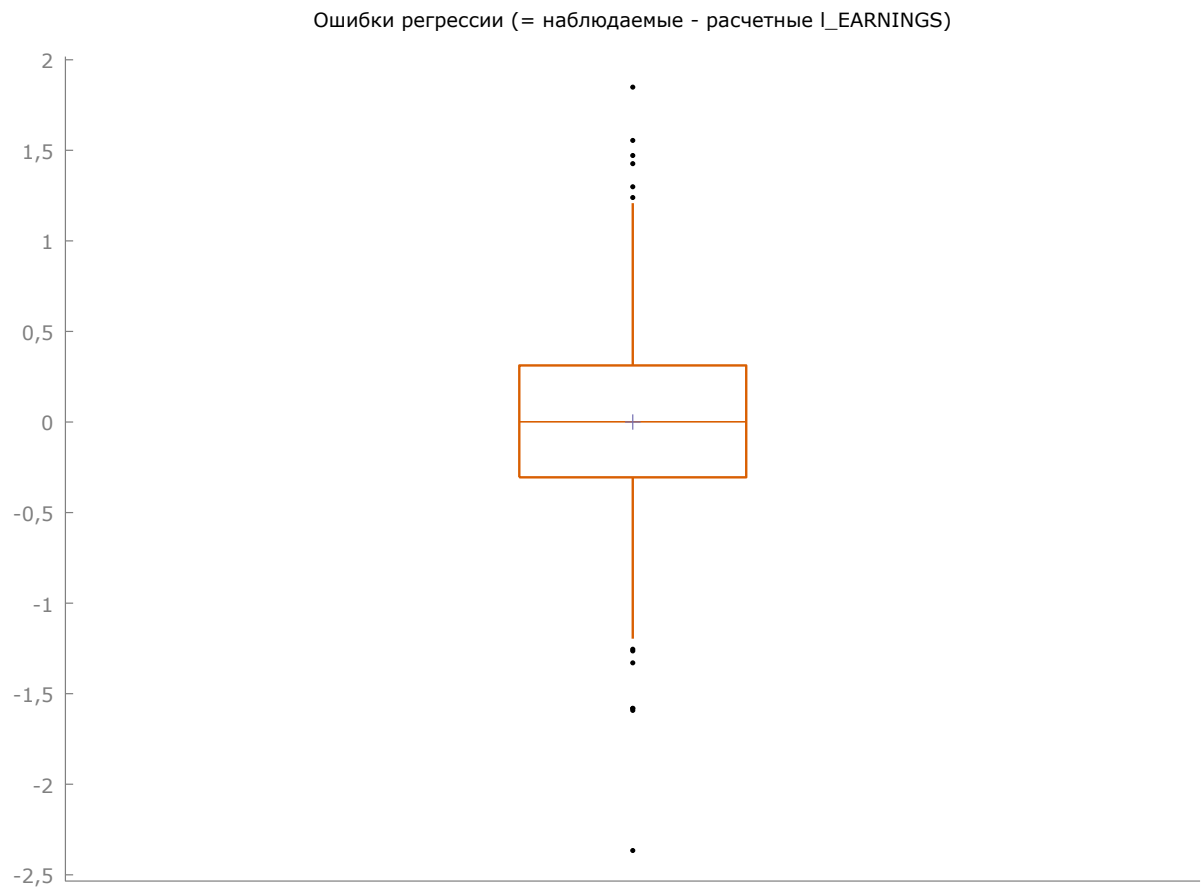
$$t(534, 0,005) = 2,585$$

	коэффициент	99 доверительный интервал
const	0,659993	[0,316224, 1,00376]
S	0,0819166	[0,0556502, 0,108183]
ASVABC	0,0163784	[0,00917401, 0,0235828]
MALE	0,344517	[0,237371, 0,451663]
ETHBLACK	0,0128639	[-0,170876, 0,196604]
ETHHISP	0,00231189	[-0,209558, 0,214182]

Как можно заметить, чем больше степень уверенности доверительного интервала, тем больше отрезок, в который попадают коэффициенты.

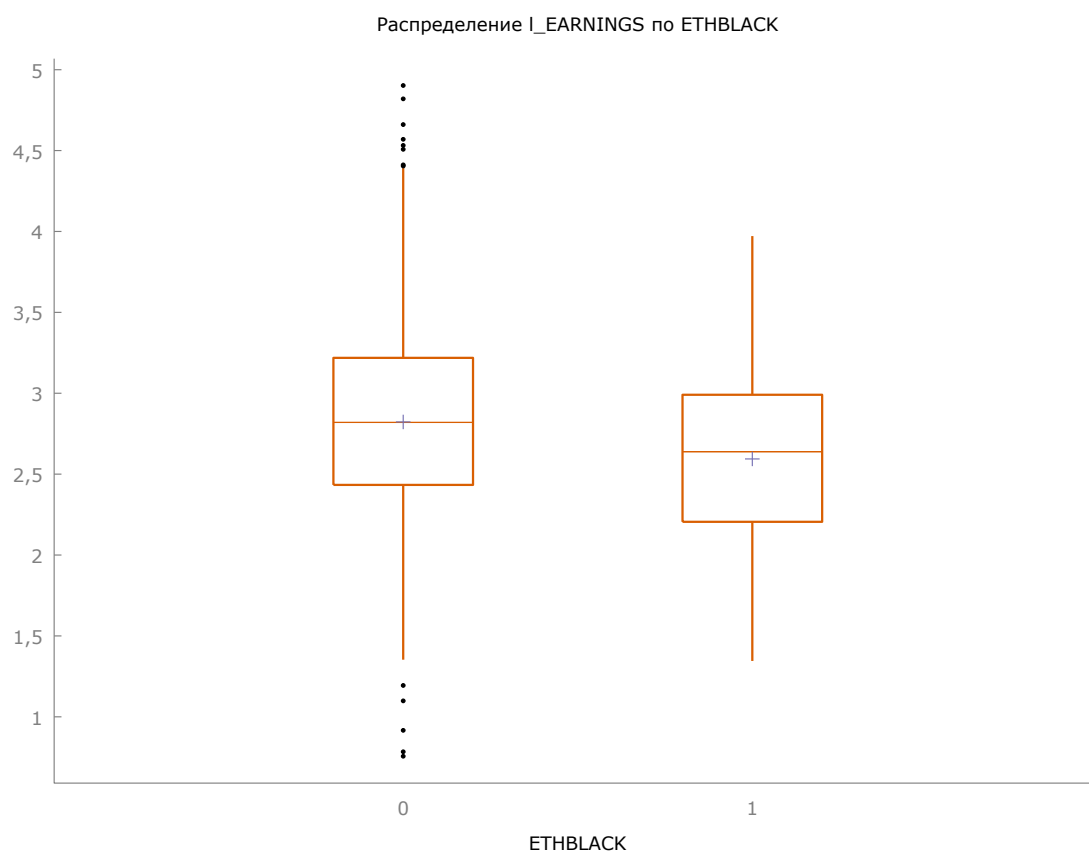
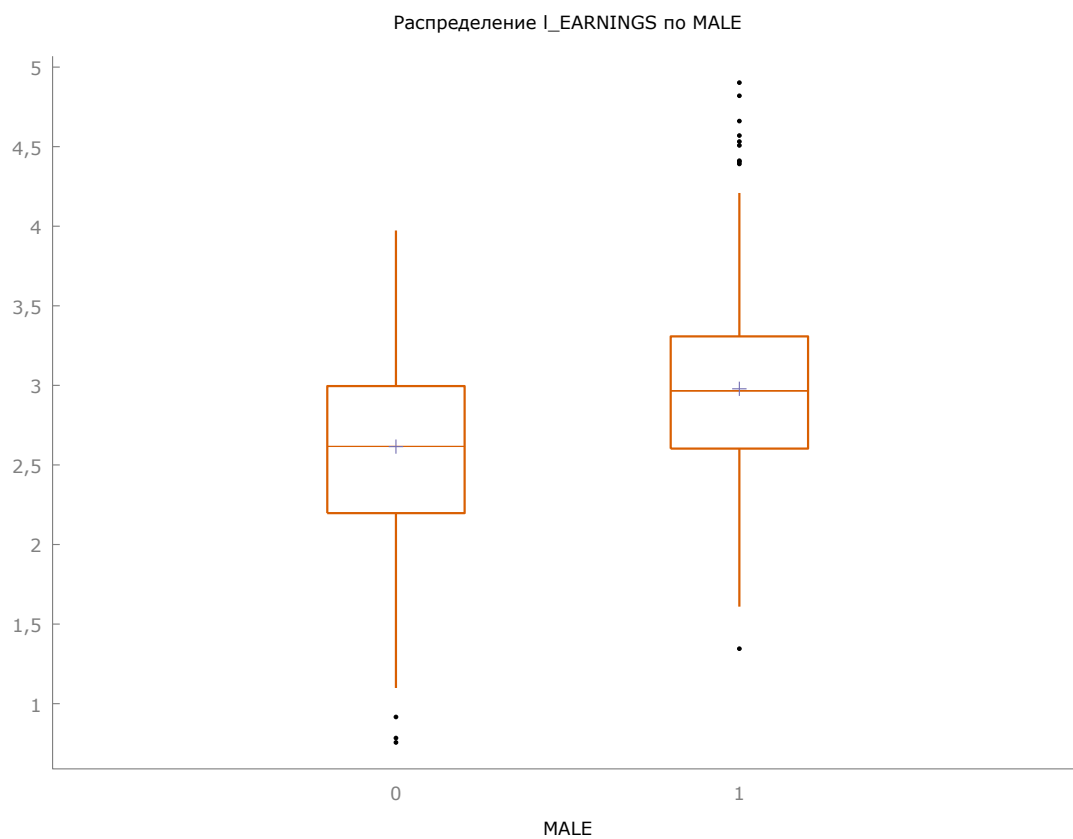
## 8. Графики.



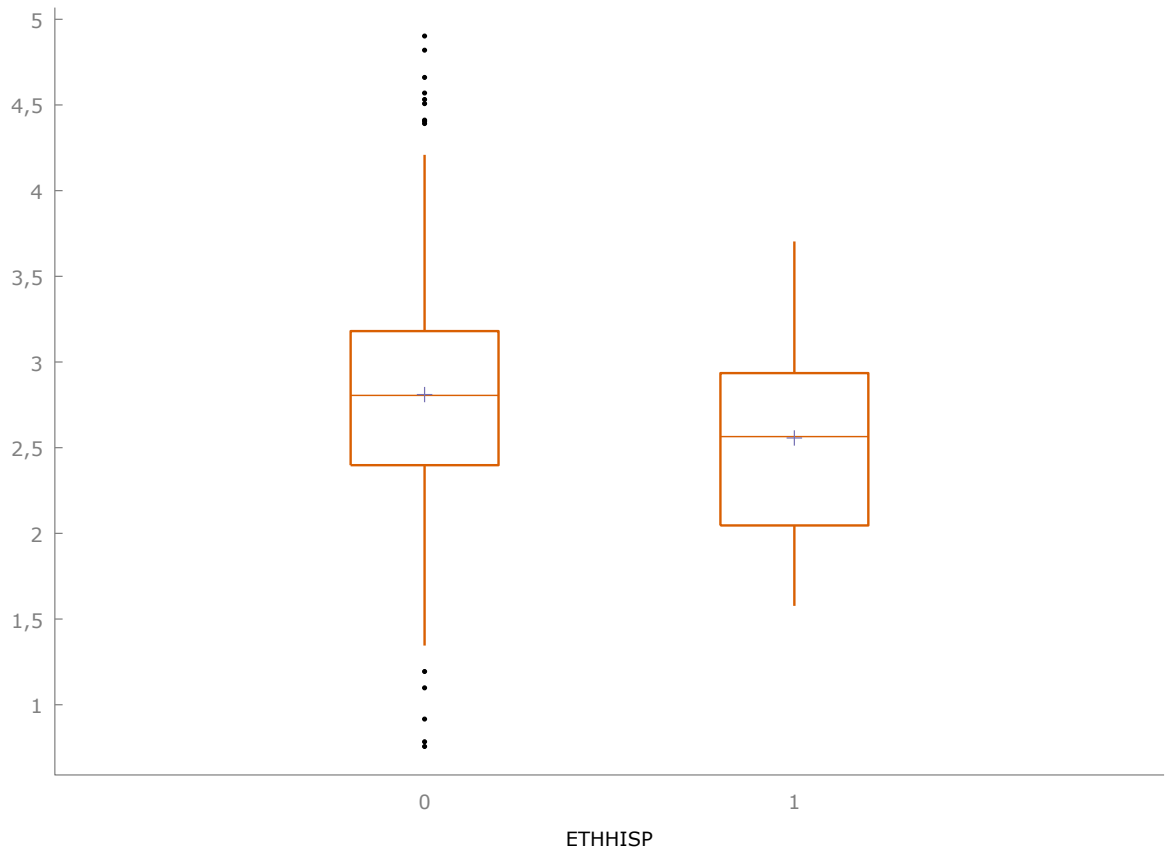


Посмотрим влияние различных фиктивных переменных на зависимую переменную:

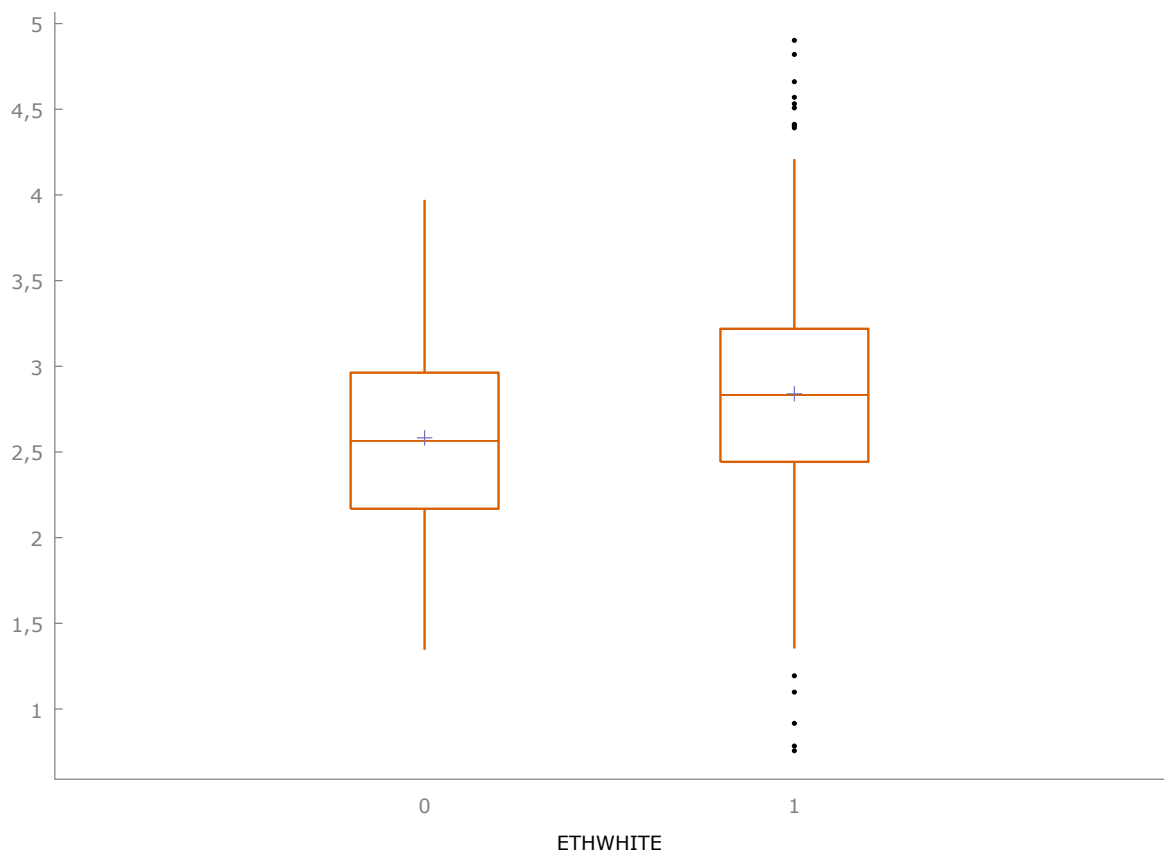




Распределение I\_EARNINGS по ETHHISP



Распределение I\_EARNINGS по ETHWHITE



Сравним три модели по  $R^2$  adj, MAPE, критериям Акаике и Шварца:

	$R^2$ adjusted	MAPE	Критерий Акаике	Критерий Шварца
$\ln(\text{EARNINGS})$ $= b_1 + b_2 * S +$ $b_3 * \text{ASVABC} +$ $e$	0,293887	15,755	805,5450	818,4197
$\ln(\text{EARNINGS})$ $= b_1 + b_2 * S +$ $b_3 * \text{ASVABC} +$ $b_4 * \text{MALE} + e$	0,373762	14,673	741,7147	758,8810
$\ln(\text{EARNINGS})$ $= b_1 + b_2 * S +$ $b_3 * \text{ASVABC} +$ $b_4 * \text{MALE} +$ $b_5 * \text{ETHBLACK}$ $+ b_6 * \text{ETHISP} +$ $e$	0,371454	14,678	745,6826	771,4320

Как можно заметить, лучшая модель оказалась модель номер 2,  
 $\ln(\text{EARNINGS}) = b_1 + b_2 * S + b_3 * \text{ASVABC} + b_4 * \text{MALE} + e$ , у нее самый  
 высокий  $R^2$  adjusted, самые низкие MAPE, критерии Акаике и Шварца.