

**РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ ИМЕНИ
ПАТРИСА ЛУМУМБЫ**

Факультет физико-математических и естественных наук

Кафедра «Математического моделирования и искусственного интеллекта»

**Методы машинного обучения для анализа
временных рядов и панельных данных
Лабораторная работа №1
«Введение в регрессионный анализ и
анализ временных рядов»**

Студент

Шуплецов Александр Андреевич

Группа

НФИбд01-22

Москва

2024

Упражнение 1.1

Парная регрессия (линейная модель) $y = b_0 + b_1x$

Оцениваем значимость коэффициентов b_0, b_1 при $\alpha = 0.01$ (это значение в Gretl определено изначально разработчиками и его не надо специально вводить)

- H_0 : коэффициент незначим
- H_1 : коэффициент значим

Хлеб:

Модель 1: МНК, использованы наблюдения 1-190
Зависимая переменная: DWBREAD

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	3,88185	0,295710	13,13	<0,0001	***
PBREAD	-0,0751086	0,0139167	-5,397	<0,0001	***
Среднее завис. перемен	2,448684	Ст. откл. завис. перемен	1,922317		
Сумма кв. остатков	604,7198	Ст. ошибка модели	1,793487		
R-квадрат	0,134151	Исправ. R-квадрат	0,129545		
F(1, 188)	29,12782	R-значение (F)	2,02e-07		
Лог. правдоподобие	-379,5837	Крит. Акаике	763,1675		
Крит. Шварца	769,6615	Крит. Хеннана-Куинна	765,7981		

Уравнение регрессии имеет вид

$$DWBREAD = 3,8818 - 0,0751086 PBREAD$$

(0,295710) (0,0139167)

p-значение у коэффициентов b_0, b_1 <0,0001, поэтому принимаем гипотезу H_1 для b_0, b_1 (коэффициенты значимы с вероятностью 99%).

p-значение для F статистики $p = 2,02e - 07$.

Коэффициент детерминации $R^2=0,134151$, Исправ. $R^2= 0,129545$.

Картофель:

Модель 1: МНК, использованы наблюдения 1-111
Зависимая переменная: DPOTAT

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	8,32844	1,77700	4,687	<0,0001	***
PPOTAT	-0,282803	0,163483	-1,730	0,0865	*
Среднее завис. перемен	5,351351	Ст. откл. завис. перемен	4,705217		
Сумма кв. остатков	2370,227	Ст. ошибка модели	4,663175		
R-квадрат	0,026720	Исправ. R-квадрат	0,017791		
F(1, 109)	2,992416	R-значение (F)	0,086486		
Лог. правдоподобие	-327,3994	Крит. Акаике	658,7987		

Крит. Шварца 664,2178 Крит. Хеннана-Куинна 660,9971
Уравнение регрессии имеет вид

$$DPOTAT = 8,32844 - 0,282803 PPOTAT$$

$$(1,77700) \quad (0,163483)$$

p-значение у коэффициента $b_0 < 0,0001$, поэтому принимаем гипотезу H_1 для b_0 (коэффициент значим с вероятностью 99%), *p-значение* у коэффициента $b_1 = 0,0865$, коэффициент значим с вероятностью 90%

p-значение для F статистики $p = 0,086486$, модель значима с вероятностью 90%.

Коэффициент детерминации $R^2 = 0,026720$, Исправ. $R^2 = 0,017791$.

Модель имеет низкий R-квадрат. Это означает, что существуют другие факторы, которые сильно влияют на зависимую переменную.

Макароны:

Модель 1: МНК, использованы наблюдения 1-122
Зависимая переменная: DMACAR

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	1,90819	0,276090	6,911	<0,0001	***
PMACAR	-0,0259644	0,0111167	-2,336	0,0212	**
Среднее завис. перемен	1,306967	Ст. откл. завис. перемен		1,122595	
Сумма кв. остатков	145,8561	Ст. ошибка модели		1,102482	
R-квадрат	0,043482	Исправ. R-квадрат		0,035511	
F(1, 120)	5,455078	R-значение (F)		0,021171	
Лог. правдоподобие	-184,0051	Крит. Акаике		372,0101	
Крит. Шварца	377,6182	Крит. Хеннана-Куинна		374,2880	

Уравнение регрессии имеет вид

$$DMACAR = 1,90819 - 0,0259644 PMACAR$$

$$(0,276090) \quad (0,0111167)$$

p-значение у коэффициента $b_0 < 0,0001$, поэтому принимаем гипотезу H_1 для b_0 (коэффициент значим с вероятностью 99%), *p-значение* у коэффициента $b_1 = 0,0212$, коэффициент значим с вероятностью 95%

p-значение для F статистики $p = 0,021171$, модель значима с вероятностью 95%.

Коэффициент детерминации $R^2 = 0,043482$, Исправ. $R^2 = 0,035511$.

Модель имеет низкий R-квадрат. Это означает, что существуют другие факторы, которые сильно влияют на зависимую переменную.

Яйца:

Модель 1: МНК, использованы наблюдения 1-148
Зависимая переменная: DEGGS

	<i>Коэффициент</i> <i>t</i>	<i>Ст. ошибка</i>	<i>t-</i> <i>статистика</i>	<i>p-значение</i>	
const	28,3630	7,18152	3,949	0,0001	***
PEGGS	-3,98353	3,03115	-1,314	0,1908	
Среднее завис. перемен	19,00000	Ст. откл. завис. перемен	11,01823		
Сумма кв. остатков	17637,36	Ст. ошибка модели	10,99108		
R-квадрат	0,011691	Исправ. R-квадрат	0,004922		
F(1, 146)	1,727113	P-значение (F)	0,190841		
Лог. правдоподобие	-563,7645	Крит. Акаике	1131,529		
Крит. Шварца	1137,523	Крит. Хеннана-Куинна	1133,965		

$$\text{DEGGS} = 28,3630 - 3,98353 * \text{PEGGS}$$

$$(7,18152) \quad (3,03115)$$

Модель в целом незначима, так как P-значение = 0,190841, нет оснований отвергать гипотезу H0 (модель незначима).

Молоко:

Модель 1: МНК, использованы наблюдения 1-160
Зависимая переменная: DMILK

	<i>Коэффициент</i> <i>t</i>	<i>Ст. ошибка</i>	<i>t-</i> <i>статистика</i>	<i>p-значение</i>	
const	2,45359	1,02374	2,397	0,0177	**
PMILK	0,0104310	0,0494975	0,2107	0,8334	
Среднее завис. перемен	2,665625	Ст. откл. завис. перемен	2,382273		
Сумма кв. остатков	902,1074	Ст. ошибка модели	2,389464		
R-квадрат	0,000281	Исправ. R-квадрат	-0,006046		
F(1, 158)	0,044410	P-значение (F)	0,833364		
Лог. правдоподобие	-365,3949	Крит. Акаике	734,7899		
Крит. Шварца	740,9402	Крит. Хеннана-Куинна	737,2873		

$$DMILK = 2,45359 + 0,0104310 * PMILK$$

$$(1,02374) \quad (0,0494975)$$

Снова заметим, что модель в целом незначима, так как Р-значение = **0,833364**.

1.2. Оценим кривые Энгеля для каждого продукта

Кривые Энгеля отражают зависимость потребления определенного товара от дохода потребителя. Проверяем, соответствует ли кривая Энгеля для каждого товара уравнению прямой.

Хлеб:

Модель 1: МНК, использованы наблюдения 1-190
Зависимая переменная: DWBREAD

	Коэффициент <i>m</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	2,19572	0,161177	13,62	<0,0001	***
INCOME	1,16600e-05	3,93741e-06	2,961	0,0035	***
Среднее завис. перемен	2,448684	Ст. откл. завис. перемен		1,922317	
Сумма кв. остатков	667,2859	Ст. ошибка модели		1,883983	
R-квадрат	0,044567	Исправ. R-квадрат		0,039485	
F(1, 188)	8,769471	Р-значение (F)		0,003458	
Лог. правдоподобие	-388,9368	Крит. Акаике		781,8736	
Крит. Шварца	788,3676	Крит. Хеннана-Куинна		784,5042	

$$DWBREAD = 2,19572 + 1,16600e-05 * INCOME$$

$$(0,161177) \quad (3,93741e-06)$$

Р-значение уравнения = **0,003458**, уравнение в целом значимо с вероятностью 99%.

Р-значение коэффициента при b_1 = **0,0035**, коэффициент значим с вероятностью 99%.

Можно утверждать, что при увеличении дохода на 1 рубль спрос домохозяйств на хлеб увеличится на 1,16600e-05 кг.

Картофель:

Модель 1: МНК, использованы наблюдения 1-111
Зависимая переменная: DPOTAT

	Коэффициент <i>m</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	5,30380	0,507469	10,45	<0,0001	***
INCOME	2,17313e-06	1,08459e-05	0,2004	0,8416	
Среднее завис. перемен	5,351351	Ст. откл. завис. перемен		4,705217	
Сумма кв. остатков	2434,401	Ст. ошибка модели		4,725881	
R-квадрат	0,000368	Исправ. R-квадрат		-0,008803	

F(1, 109)	0,040146	Р-значение (F)	0,841569
Лог. правдоподобие	-328,8821	Крит. Акаике	661,7641
Крит. Шварца	667,1832	Крит. Хеннана-Куинна	663,9625

$$\text{DROTAT} = 5,30380 + 2,17313\text{e-}06 * \text{INCOME}$$

$$(0,507469) \quad (1,08459\text{e-}05)$$

Модель незначима в целом, так как Р-значение = 0,841569.

Р-значение при переменной INCOME также крайне высокое, доход не влияет на спрос домохозяйств на картофель.

Макаронны:

Модель 1: МНК, использованы наблюдения 1-122
Зависимая переменная: DMACAR

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	0,689011	0,126350	5,453	<0,0001	***
INCOME	3,08065e-05	4,57126e-06	6,739	<0,0001	***
Среднее завис. перемен	1,306967	Ст. откл. завис. перемен	1,122595		
Сумма кв. остатков	110,6202	Ст. ошибка модели	0,960123		
R-квадрат	0,274558	Исправ. R-квадрат	0,268512		
F(1, 120)	45,41631	Р-значение (F)	5,88e-10		
Лог. правдоподобие	-167,1375	Крит. Акаике	338,2750		
Крит. Шварца	343,8830	Крит. Хеннана-Куинна	340,5528		

$$\text{DMACAR} = 0,689011 + 3,08065\text{e-}05 * \text{INCOME}$$

$$(0,126350) \quad (4,57126\text{e-}06)$$

Р-значение уравнения = 5,88e-10, уравнение в целом значимо с вероятностью 99%.

Р-значение коэффициента при b1 = <0,0001, коэффициент значим с вероятностью 99%.

Можно утверждать, что при увеличении дохода на 1 рубль спрос домохозяйств на макароны увеличится на 3,08065e-05 кг.

Яйца:

Модель 1: МНК, использованы наблюдения 1-148
Зависимая переменная: DEGGS

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	17,4289	1,02846	16,95	<0,0001	***

INCOME	6,83244e-05	2,29717e-05	2,974	0,0034	***
Среднее завис. перемен	19,00000	Ст. откл. завис. перемен	11,01823		
Сумма кв. остатков	16826,46	Ст. ошибка модели	10,73544		
R-квадрат	0,057130	Исправ. R-квадрат	0,050672		
F(1, 146)	8,846373	P-значение (F)	0,003437		
Лог. правдоподобие	-560,2816	Крит. Акаике	1124,563		
Крит. Шварца	1130,558	Крит. Хеннана-Куинна	1126,999		

$$\text{DEGGS} = 17,4289 + 6,83244\text{e-}05 * \text{INCOME}$$

$$(1,02846) \quad (2,29717\text{e-}05)$$

P-значение уравнения = 0,003437, уравнение в целом значимо с вероятностью 99%.

P-значение коэффициента при b1 = 0,0034, коэффициент значим с вероятностью 99%.

Можно утверждать, что при увеличении дохода на 1 рубль спрос домохозяйств на яйца увеличится на 6,83244e-05 кг.

Молоко:

Модель 1: МНК, использованы наблюдения 1-160
Зависимая переменная: DMILK

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	2,51186	0,219551	11,44	<0,0001	***
INCOME	6,88912e-06	5,09206e-06	1,353	0,1780	
Среднее завис. перемен	2,665625	Ст. откл. завис. перемен	2,382273		
Сумма кв. остатков	892,0271	Ст. ошибка модели	2,376077		
R-квадрат	0,011452	Исправ. R-квадрат	0,005195		
F(1, 158)	1,830376	P-значение (F)	0,178016		
Лог. правдоподобие	-364,4960	Крит. Акаике	732,9920		
Крит. Шварца	739,1423	Крит. Хеннана-Куинна	735,4894		

$$\text{DMILK} = 2,51186 + 6,88912\text{e-}06 * \text{INCOME}$$

$$(0,219551) \quad (5,09206\text{e-}06)$$

Модель незначима в целом, так как P-значение = 0,178016.

P-значение при переменной INCOME также крайне высокое, доход не влияет на спрос домохозяйств на молоко.

Упражнение 1.3. Оценка модели CAPM по американским данным

$r_j = \text{IBM}$, $r_f = \text{RKFREE}$, $r_m = \text{MARKET}$

Продолжаем использовать модель парной регрессии:

$$y = b_0 + b_1 \cdot x$$

Построим модель с помощью метода наименьших квадратов, где $Y = \text{IBM} - \text{RKFREE}$, $X = \text{MARKET} - \text{RKFREE}$:

Модель 1: МНК, использованы наблюдения 1-120

Зависимая переменная: Y

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	-0,00048959	0,00464003	-0,1055	0,9161	
	<i>4</i>				
X	0,456821	0,0675477	6,763	<0,0001	***
Среднее завис. перемен	0,002778	Ст. откл. завис. перем		0,059299	
Сумма кв. остатков	0,301558	Ст. ошибка модели		0,050553	
R-квадрат	0,279333	Исправ. R-квадрат		0,273226	
F(1, 118)	45,73725	P-значение (F)		5,49e-10	
Лог. правдоподобие	188,9045	Крит. Акаике		-373,8090	
Крит. Шварца	-368,2340	Крит. Хеннана-Куинна		-371,5450	

Уравнение регрессии имеет вид

$$Y = -0,000489594 + 0,456821 X$$

(0,00464003) (0,0675477)

Уравнение модели CAMP имеет вид:

$$\text{IBM-RKFREE} = -0,000489594 + 0,456821(\text{MARKET-RKFREE})$$

(0,00464003) (0,0675477)

p-значение при переменной X <0,0001 – переменная значима с вероятностью 99%.

p-значение для F статистики $p = 5,49e - 10$, модель значима с вероятностью 99%.

p-значение при константе больше 0,1 – коэффициент незначим.

Коэффициент детерминации $R^2 = 0,279333$, Исправ. $R^2 = 0,273226$

Так как b_1 является значимым, то доходность бумаги компании IBM зависит от доходности общего рыночного портфеля ценных бумаг.

Экономически можно интерпретировать так: коэффициент b_1 в нашем случае меньше 1, это значит, что акции IBM достаточно стабильны к изменениям на рынке. Они не реагируют на рыночные колебания сильнее, чем средний актив.

Упражнение 1.4. Оценка кривой Филлипса

1.4.1

$$X1 = 1/X$$

Модель 2: МНК, использованы наблюдения 1958-1969 (T = 12)
Зависимая переменная: Y

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	-0,259437	1,00864	-0,2572	0,8022	
X1	20,5879	4,67948	4,400	0,0013	***
Среднее завис. перемен	4,066667	Ст. откл. завис. перемен		1,271601	
Сумма кв. остатков	6,058842	Ст. ошибка модели		0,778386	
R-квадрат	0,659360	Исправ. R-квадрат		0,625296	
F(1, 10)	19,35654	R-значение (F)		0,001336	
Лог. правдоподобие	-12,92694	Крит. Акаике		29,85387	
Крит. Шварца	30,82368	Крит. Хеннана-Куинна		29,49481	
параметр rho	0,513773	Стат. Дарбина-Уотсона		0,639368	

Уравнение регрессии имеет вид: $Y = -0,259437 + 20,5879 * 1/X$
(1,00864) (4,67948)

Коэффициент b1 является значимым с 99% вероятностью, так как *p*-значение = 0,0013.
Уравнение в целом также значимо с 99% вероятностью, так как *p*-значение = 0,001336.
3. Дайте экономическую интерпретацию полученному результату.

Экономически результаты можно интерпретировать так: при уменьшении уровня безработицы на один процент, на 20,5879% увеличится почасовая заработная плата. Результаты подтверждают теорию Филиппса, увеличение часовой заработной платы зависит от уровня безработицы обратно пропорционально, то есть существует краткосрочная обратная корреляция между инфляцией и безработицей.

4. Снижался ли, согласно полученному результату, уровень инфляции при увеличении уровня безработицы?

Как можно заметить, согласно полученному уравнению, увеличение уровня безработицы приводит к замедлению роста заработной платы, что указывает на снижение инфляции в экономике США в период с 1958 по 1969 годы.

1.4.2

Модель 1: МНК, использованы наблюдения 1-29
Зависимая переменная: Y

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	111,948	1,98500	56,40	<0,0001	***
X1	-71,4192	16,2827	-4,386	0,0002	***
Среднее завис. перемен	103,3138	Ст. откл. завис. перемен		1,766903	
Сумма кв. остатков	51,04351	Ст. ошибка модели		1,374955	
R-квадрат	0,416075	Исправ. R-квадрат		0,394448	
F(1, 27)	19,23881	R-значение (F)		0,000158	
Лог. правдоподобие	-49,34727	Крит. Акаике		102,6945	
Крит. Шварца	105,4291	Крит. Хеннана-Куинна		103,5510	

Уравнение регрессии имеет вид: $Y = 111,948 - 71,4192 * 1/X$
 (1,98500) (16,2827)

Коэффициент b1 является значимым с 99% вероятностью, так как р-значение = 0,0002. Уравнение в целом также значимо с 99% вероятностью, так как р-значение = 0,000158.

3. Дайте экономическую интерпретацию полученному результату.

Экономически результаты можно интерпретировать так: Результаты подтверждают теорию Филипса, существует краткосрочная обратная корреляция между инфляцией и безработицей. Чем больше уровень безработицы, тем медленнее растет инфляция.

4. Снижался ли, согласно полученному результату, уровень инфляции при увеличении уровня безработицы?

В данном случае уровень инфляции не снижается при увеличении уровня безработицы, так как увеличение X приводит к увеличению Y и к росту инфляции.

Упражнение 2.1

1. Оцените параметры уравнения множественной регрессии

PRICE = $\beta_0 + \beta_1 \text{ GREEN} + \beta_2 \text{ NOSTALKS} + \beta_3 \text{ DISPERSE} + \epsilon$.

Модель 1: МНК, использованы наблюдения 1-200
 Зависимая переменная: PRICE

	Коэффициент <i>t</i>	Ст. ошибка	<i>t</i> - статистика	р-значение	
const	40,7613	5,32784	7,651	<0,0001	***
GREEN	0,137598	0,00709935	19,38	<0,0001	***
NOSTALKS	-1,35726	0,150822	-8,999	<0,0001	***
DISPERSE	-0,345283	0,129656	-2,663	0,0084	***
Среднее завис. перемен	90,09500	Ст. откл. завис. перемен		29,47439	
Сумма кв. остатков	47230,75	Ст. ошибка модели		15,52331	
R-квадрат	0,726799	Исправ. R-квадрат		0,722617	
F(3, 196)	173,8070	Р-значение (F)		5,71e-55	
Лог. правдоподобие	-830,2360	Крит. Акаике		1668,472	
Крит. Шварца	1681,665	Крит. Хеннана-Куинна		1673,811	

PRICE = 40,7613 + 0,137598 * GREEN – 1,35726 * NOSTALKS – 0,345283 * DISPERSE
 (5,32784) (0,00709935) (0,150822) (0,129656)

2. Проверьте адекватность регрессии в целом и значимость коэффициентов регрессии по отдельности.

р-значение уравнения < 0,01 – уравнение значимо в целом с 99% вероятностью.

р-значения всех переменных также меньше 0,01 – все переменные значимы с 99% вероятностью.

Также у модели достаточно высокий исправленный R-квадрат (0,722617).

Средняя абсолютная процентная ошибка (MAPE) 14,721 – хорошее качество модели

Тест Вайта (White) на гетероскедастичность -

Нулевая гипотеза: гетероскедастичность отсутствует

Тестовая статистика: LM = 13,5837

p-значение = $P(\chi^2(9) > 13,5837) = 0,137925$

Как можно заметить, гетероскедастичность не наблюдается в нашей модели.

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

GREEN 1,017

NOSTALKS 1,141

DISPERSE 1,159

Мультиколлинеарности также нет.

Вывод: модель адекватна, коэффициенты значимы.

3. Дайте интерпретацию полученным результатам.

Если длина стеблей спаржи в одном пучке (GREEN) увеличивается на одну сотую дюйма, то относительная цена (PRICE) увеличивается в среднем на 0,137598 условных единиц.

Если количество стеблей спаржи в одном пучке (NOSTALKS) увеличивается на 1, то относительная цена (PRICE) уменьшается в среднем на 1,35726 условных единиц.

Если разброс в размере стеблей (DISPERSE) увеличивается на одну единицу, то относительная цена пучка уменьшается в среднем на 0,345283 единиц.

Упражнение 2.2. Зависимость заработной платы от способностей индивида, его образования и образования его родителей в США (линейная модель)

2.2.1. Зависимость длительности образования от способностей и длительности обучения родителей

1. С помощью данных файла data 2.2.gdt оцените параметры уравнения множественной регрессии $S = \beta_0 + \beta_1 \text{ASVABC} + \beta_2 \text{SM} + \varepsilon$.

Модель 1: МНК, использованы наблюдения 1-540

Зависимая переменная: S

	Коэффициент <i>m</i>	Ст. ошибка	<i>t</i> - статистика	<i>p</i> -значение	
const	4,35718	0,577870	7,540	<0,0001	***
ASVABC	0,128631	0,0106034	12,13	<0,0001	***
SM	0,227710	0,0388214	5,866	<0,0001	***

Среднее завис. перемен	13,67222	Ст. откл. завис. перемен	2,555863
Сумма кв. остатков	2346,569	Ст. ошибка модели	2,090401
R-квадрат	0,333547	Исправ. R-квадрат	0,331065
F(2, 537)	134,3793	P-значение (F)	4,81e-48
Лог. правдоподобие	-1162,895	Крит. Акаике	2331,789
Крит. Шварца	2344,664	Крит. Хеннана-Куинна	2336,825

$$S = 4,35718 + 0,128631 * ASVABC + 0,227710 * SM + \epsilon.$$

$$(0,577870) \quad (0,0106034) \quad (0,0388214)$$

2. Проверьте адекватность регрессии в целом и значимость коэффициентов регрессии по отдельности.

p-значение уравнения < 0,01 – уравнение значимо в целом с 99% вероятностью.

p-значения всех переменных также меньше 0,01 – все переменные значимы с 99% вероятностью.

Исправ. R-квадрат (0,331065) не является высоким, значит существуют другие факторы, влияющие на зависимую переменную.

Средняя абсолютная процентная ошибка (MAPE) 12,202 – хорошее качество модели

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

$$ASVABC \quad 1,152$$

$$SM \quad 1,152$$

Мультиколлинеарность отсутствует.

Тест Вайта (White) на гетероскедастичность -

Нулевая гипотеза: гетероскедастичность отсутствует

Тестовая статистика: LM = 10,337

$$p\text{-значение} = P(\text{Хи-квадрат}(5) > 10,337) = 0,0662316$$

Присутствует гетероскедастичность с 90% вероятностью.

Модель 2: МНК, использованы наблюдения 1-540

Зависимая переменная: S

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

	Коэффициент	Ст. ошибка	t-статистика	p-значение	
	<i>m</i>				
const	4,35718	0,606997	7,178	<0,0001	***
ASVABC	0,128631	0,0108685	11,84	<0,0001	***
SM	0,227710	0,0413820	5,503	<0,0001	***

Среднее завис. перемен	13,67222	Ст. откл. завис. перемен	2,555863
Сумма кв. остатков	2346,569	Ст. ошибка модели	2,090401
R-квадрат	0,333547	Исправ. R-квадрат	0,331065
F(2, 537)	119,6031	P-значение (F)	1,09e-43
Лог. правдоподобие	-1162,895	Крит. Акаике	2331,789
Крит. Шварца	2344,664	Крит. Хеннана-Куинна	2336,825

Как можно заметить, поправка на гетероскедастичность незначительно увеличила р-значение, тем не менее, можно сделать общий вывод, что модель в целом адекватна и все переменные в ней – значимые.

3. Исходя из полученных результатов сделайте вывод, какие факторы влияют на длительность обучения индивида. Дайте интерпретацию полученным результатам.

Способности школьника (ASVABC) и количество лет обучения матери (SM) влияют на длительность обучения индивида.

При увеличении способностей школьника на 1 условную единицу кол-во лет обучения в среднем увеличивается на 0,128631 лет.

При увеличении количества лет обучения матери индивида кол-во лет обучения индивида уменьшается в среднем на 0,22771 лет.

При условии нулевых значений независимых переменных кол-во лет обучения индивида в среднем составит 4,35718 лет.

2.2.2. Зависимость заработной платы от способностей и образования индивида

С помощью данных файла data 2.2.gdt оцените параметры уравнения множественной регрессии $EARN = \beta_0 + \beta_1 S + \beta_2 ASVABC + \varepsilon$.

Модель 3: МНК, использованы наблюдения 1-540
Зависимая переменная: EARNINGS

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	-20,5329	3,59502	-5,711	<0,0001	***
S	1,76657	0,262150	6,739	<0,0001	***
ASVABC	0,311208	0,0735111	4,233	<0,0001	***
Среднее завис. перемен	19,71924	Ст. откл. завис. перемен			
Сумма кв. остатков	92146,36	Ст. ошибка модели			
R-квадрат	0,198148	Исправ. R-квадрат			
F(2, 537)	66,34997	P-значение (F)			
Лог. правдоподобие	-2153,909	Крит. Акаике			
Крит. Шварца	4326,693	Крит. Хеннана-Куинна			

$$EARN = -20,5329 + 1,76657 * S + 0,311208 * ASVABC + \varepsilon$$

$$(3,59502) \quad (0,262150) \quad (0,0735111)$$

2. Проверьте адекватность регрессии в целом и значимость коэффициентов регрессии по отдельности.

р-значение уравнения $< 0,01$ – уравнение значимо в целом с 99% вероятностью.

р-значения всех переменных также меньше 0,01 – все переменные значимы с 99% вероятностью.

Исправ. R-квадрат (0,195162) не является высоким, значит существуют другие факторы, влияющие на зависимую переменную.

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

S 1,410

ASVABC 1,410

Мультиколлинеарность отсутствует

Средняя абсолютная процентная ошибка (MAPE) 55,294 – неудовлетворительное качество модели

Тест Вайта (White) на гетероскедастичность -

Нулевая гипотеза: гетероскедастичность отсутствует

Тестовая статистика: $LM = 32,2606$

р-значение = $P(\text{Хи-квадрат}(5) > 32,2606) = 5,27537e-06$

Также в модели присутствует гетероскедастичность

Модель 4: МНК, использованы наблюдения 1-540

Зависимая переменная: EARNINGS

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

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	-20,5329	4,10869	-4,997	<0,0001	***
S	1,76657	0,358034	4,934	<0,0001	***
ASVABC	0,311208	0,0744864	4,178	<0,0001	***
Среднее завис. перемен	19,71924	Ст. откл. завис. перемен		14,60151	
Сумма кв. остатков	92146,36	Ст. ошибка модели		13,09942	
R-квадрат	0,198148	Исправ. R-квадрат		0,195162	
F(2, 537)	41,53051	P-значение (F)		1,70e-17	
Лог. правдоподобие	-2153,909	Крит. Акаике		4313,818	
Крит. Шварца	4326,693	Крит. Хеннана-Куинна		4318,854	

Как можно заметить, поправка на гетероскедастичность незначительно увеличила r -значение, тем не менее, можно сделать общий вывод, что модель в целом значима и все переменные в ней – значимые, тем не менее, качество модели – неудовлетворительное.

3. Исходя из полученных результатов сделайте вывод, какие факторы влияют на заработную плату индивида. Дайте интерпретацию полученным результатам.

Так как все коэффициенты значимы и само уравнение регрессии значимо, заработная плата индивида (EARNINGS) зависит от количества лет обучения индивида (S) и от интеллектуальных способностей индивида по совокупности тестов (ASVABC).

При увеличении обучения индивида на 1 год его почасовая заработная плата увеличивается в среднем на 1,76657 единиц её измерения.

При увеличении уровня интеллектуальных способностей индивида по результатам совокупности тестов на одну единицу почасовая заработная плата индивида увеличивается в среднем на 0,311208 единиц её измерения.

Упражнение 2.3. Зависимость потребления основных видов товаров от их цен и доходов домохозяйств (по данным RLMS)

1. По данным файла data 2.3.gdt оцените зависимость потребления одного из видов товаров Y от его цены P и дохода домохозяйства I : $Y = \beta_0 + \beta_1 P + \beta_2 I + \epsilon$.

Я выберу молоко.

$$\text{buymilk_b} = b_0 + b_1 * \text{pr_milk} + b_2 * \text{inc}$$

Модель 1: МНК, использованы наблюдения 1-3335 ($n = 2127$)

Исключено пропущенных или неполных наблюдений: 1208

Зависимая переменная: `buymilk_b`

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>p-значение</i>	
	<i>m</i>				
const	3,34220	0,0725235	46,08	<0,0001	***
pr_milk	-0,00213311	0,000839833	-2,540	0,0112	**
inc	5,18039e-06	3,16305e-06	1,638	0,1016	
Среднее завис. перемен	3,368806	Ст. откл. завис. перемен		2,682983	
Сумма кв. остатков	15237,92	Ст. ошибка модели		2,678462	
R-квадрат	0,004305	Исправ. R-квадрат		0,003367	
F(2, 2124)	4,591247	P-значение (F)		0,010241	
Лог. правдоподобие	-5112,193	Крит. Акаике		10230,39	
Крит. Шварца	10247,37	Крит. Хеннана-Куинна		10236,60	

Уравнение регрессии имеет вид:

$$\text{buymilk_b} = 3,34220 - 0,00213311 * \text{pr_milk} + 5,18039e-06 * \text{inc}$$

$$(0,0725235) \quad (0,000839833) \quad (3,16305e-06)$$

2. Проверьте адекватность регрессии в целом и значимость коэффициентов регрессии по отдельности

R-значение - 0,010241, в целом модель значима с 95% вероятностью.

R-значение при переменной pr_milk - 0,0112, переменная значима с 95% вероятностью

R-значение при переменной inc - 0,1016, переменная незначима.

Исправленный R-квадрат - 0,003367, что крайне мало.

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

pr_milk 1,000

inc 1,000

$VIF(j) = 1/(1 - R(j)^2)$, где $R(j)$ - это коэффициент множественной корреляции

между переменной j и другими независимыми переменными

Мультиколлинеарности нет.

Тест Вайта (White) на гетероскедастичность -

Нулевая гипотеза: гетероскедастичность отсутствует

Тестовая статистика: LM = 6,10866

p-значение = $P(\chi^2(5) > 6,10866) = 0,295789$

Гетероскедастичности тоже нет.

В целом можно сделать вывод, что модель адекватна, цена молока влияет на его покупки, а вот доход семьи – нет.

3. Оцените вашу регрессию по более однородной выборке. Например, можно выбрать наблюдения, относящиеся к одному федеральному округу, и/или одной первичной единице отбора, и/или одному типу населенного пункта. Дайте экономическую интерпретацию полученным результатам. Можно включить в регрессию цену товаров — субституты или комплименты.

Сделаем выборку ограниченной только центральным регионом.

Модель 2: МНК, использованы наблюдения 1-322 ($n = 213$)

Исключено пропущенных или неполных наблюдений: 109

Зависимая переменная: buymilk_b

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	3,19490	0,330092	9,679	<0,0001	***
pr_milk	0,000600258	0,00983197	0,06105	0,9514	
inc	1,24364e-05	1,35972e-05	0,9146	0,3614	
Среднее завис. перемен	3,401408	Ст. откл. завис. перемен		2,665671	
Сумма кв. остатков	1500,379	Ст. ошибка модели		2,672950	

R-квадрат	0,004016	Исправ. R-квадрат	-0,005469
F(2, 210)	0,423431	P-значение (F)	0,655354
Лог. правдоподобие	-510,1412	Крит. Акаике	1026,282
Крит. Шварца	1036,366	Крит. Хеннана-Куинна	1030,358

$$\text{buymilk_b} = 3,19490 + 0,000600258 * \text{pr_milk} + 1,24364\text{e-}05 * \text{inc}$$

$$(0,330092) \quad (0,00983197) \quad (1,35972\text{e-}05)$$

Как можно заметить, данная модель стала плохой – Все P-значения стали значительно больше 0,1, а исправленный R-квадрат и вовсе ушел в минус. Это означает, что независимые переменные не объясняют зависимую.

Упражнение 2.4. Моделирование продаж одежды

1. Оцените параметры уравнения множественной регрессии $\text{sales} = \beta_0 + \beta_1 \text{hoursw} + \beta_2 \text{ssize} + \varepsilon$.

Модель 1: МНК, использованы наблюдения 1-400
Зависимая переменная: sales

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	5133,59	321,693	15,96	<0,0001	***
hoursw	37,5284	2,83722	13,23	<0,0001	***
ssize	-22,1446	1,62507	-13,63	<0,0001	***
Среднее завис. перемен	6334,751	Ст. откл. завис. перемен		3739,344	
Сумма кв. остатков	3,54e+09	Ст. ошибка модели		2985,371	
R-квадрат	0,365804	Исправ. R-квадрат		0,362609	
F(2, 397)	114,4947	P-значение (F)		5,51e-40	
Лог. правдоподобие	-3766,662	Крит. Акаике		7539,323	
Крит. Шварца	7551,297	Крит. Хеннана-Куинна		7544,065	

$$\text{sales} = 5133,59 + 37,5284 * \text{hoursw} - 22,1446 * \text{ssize}$$

$$(321,693) \quad (2,83722) \quad (1,62507)$$

P-значение модели очень низкое, модель значима с 99% вероятностью. P-значения независимых переменных также низкие – все переменные значимы с 99% вероятностью.

Исправленный R-квадрат не очень большой – существуют другие факторы, влияющие на зависимую переменную.

Средняя абсолютная процентная ошибка (MAPE) 45,705 – качество модели – удовлетворительное

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

hoursw 1,496

ssize 1,496

$VIF(j) = 1/(1 - R(j)^2)$, где $R(j)$ - это коэффициент множественной корреляции

между переменной j и другими независимыми переменными

Мультиколлинеарности нет.

Можно сделать вывод, что модель достаточно качественная, так как все переменные значимые и MAPE достаточно низкая.

3. Влияет ли общее число отработанных часов и размер магазина на продажи в расчете на квадратный метр (и каким образом)?

Общее число отработанных часов и размер магазина влияют на продажи в расчете на квадратный метр таким образом:

Продажи магазина в расчете на квадратный метр в среднем уменьшаются на 22,1446 гильдена, при увеличении площади магазина на один квадратный метр.

При увеличении общего числа отработанных часов продажи в расчете на квадратный метр увеличиваются в среднем на 37,5284 гильдена.

4. Улучшится ли качество подгонки регрессии при включении в модель переменной poww переменной npart ? Поэкспериментируйте и с включением других переменных.

Модель 3: МНК, использованы наблюдения 1-400
Зависимая переменная: sales

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	4175,02	411,993	10,13	<0,0001	***
hoursw	37,0192	2,79793	13,23	<0,0001	***
ssize	-23,8549	1,66811	-14,30	<0,0001	***
npart	816,719	224,372	3,640	0,0003	***
Среднее завис. перемен	6334,751	Ст. откл. завис. перемен	3739,344		
Сумма кв. остатков	3,42e+09	Ст. ошибка модели	2940,352		
R-квадрат	0,386337	Исправ. R-квадрат	0,381688		
F(3, 396)	83,10161	P-значение (F)	1,02e-41		
Лог. правдоподобие	-3760,079	Крит. Акаике	7528,158		
Крит. Шварца	7544,124	Крит. Хеннана-Куинна	7534,481		
Средняя абсолютная процентная ошибка (MAPE)		45,553			

Как можно заметить, P-значение и MAPE немного уменьшились – качество модели увеличилось, также исправленный R-квадрат стал больше – тоже хороший показатель.

Теперь давайте добавим poww :

Модель 5: МНК, использованы наблюдения 1-400
Зависимая переменная: sales

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	5451,14	387,504	14,07	<0,0001	***
hoursw	39,5100	3,13938	12,59	<0,0001	***
ssize	-22,6041	1,65275	-13,68	<0,0001	***
nown	-380,265	259,534	-1,465	0,1437	
Среднее завис. перемен	6334,751	Ст. откл. завис. перемен		3739,344	
Сумма кв. остатков	3,52e+09	Ст. ошибка модели		2981,069	
R-квадрат	0,369223	Исправ. R-квадрат		0,364445	
F(3, 396)	77,26587	P-значение (F)		2,30e-39	
Лог. правдоподобие	-3765,580	Крит. Акаике		7539,160	
Крит. Шварца	7555,126	Крит. Хеннана-Куинна		7545,483	
Средняя абсолютная процентная ошибка (MAPE) 45,725					

Как можно заметить, пусть Р-значение и MAPE также стали меньше, а R-квадрат увеличился, тем не менее, с переменной part модель оказалась лучше.

Попробуем включить в модель переменную nfull, логично, если с этой переменной модель станет лучше.

Модель 11: МНК, использованы наблюдения 1-400
Зависимая переменная: sales

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	4457,21	362,809	12,29	<0,0001	***
hoursw	32,4565	3,09176	10,50	<0,0001	***
ssize	-22,5292	1,60131	-14,07	<0,0001	***
nfull	651,818	171,168	3,808	0,0002	***
Среднее завис. перемен	6334,751	Ст. откл. завис. перемен		3739,344	
Сумма кв. остатков	3,41e+09	Ст. ошибка модели		2935,866	
R-квадрат	0,388208	Исправ. R-квадрат		0,383573	
F(3, 396)	83,75947	P-значение (F)		5,56e-42	
Лог. правдоподобие	-3759,468	Крит. Акаике		7526,937	
Крит. Шварца	7542,903	Крит. Хеннана-Куинна		7533,260	
Средняя абсолютная процентная ошибка (MAPE) 46,544					

R² снова увеличился, но MAPE стала меньше, тем не менее, с 99% вероятностью nfull влияет на зависимую переменную.

7. Если Radj 2 при включении дополнительной переменной увеличивается, то модель при включении этой переменной становится лучше?

Исправленный R² (adjusted) учитывает количество переменных. Если этот показатель увеличивается, модель становится лучше с добавлением новых переменных.

Упражнение 2.5. Моделирование продаж мороженого

1. Оцените параметры уравнения множественной регрессии $\text{cons} = \beta_0 + \beta_1 \text{income} + \beta_2 \text{price} + \beta_3 \text{temp} + \varepsilon$.

Модель 1: МНК, использованы наблюдения 1-30
Зависимая переменная: cons

	Коэффициент	Ст. ошибка	t- статистика	p-значение	
	<i>m</i>				
const	0,197315	0,270216	0,7302	0,4718	
income	0,00330776	0,00117142	2,824	0,0090	***
price	-1,04441	0,834357	-1,252	0,2218	
temp	0,00345843	0,000445547	7,762	<0,0001	***
Среднее завис. перемен	0,359433	Ст. откл. завис. перемен		0,065791	
Сумма кв. остатков	0,035273	Ст. ошибка модели		0,036833	
R-квадрат	0,718994	Исправ. R-квадрат		0,686570	
F(3, 26)	22,17489	P-значение (F)		2,45e-07	
Лог. правдоподобие	58,61944	Крит. Акаике		-109,2389	
Крит. Шварца	-103,6341	Крит. Хеннана-Куинна		-107,4459	

$$\text{cons} = 0,197315 + 0,00330776 * \text{income} - 1,04441 * \text{price} + 0,00345843 * \text{temp}$$

(0,270216) (0,00117142) (0,834357) (0,000445547)

Как можно заметить, модель значима с 99 вероятностью, так как Р-значение меньше 0.01, также переменные income и temp значимы с 99 вероятностью, а вот переменная price незначима.

Также у модели достаточно высокий исправленный R².

Средняя абсолютная процентная ошибка (MAPE) 6,8982 – отличное качество модели.

Метод инфляционных факторов

Минимальное возможное значение = 1.0

Значения > 10.0 могут указывать на наличие мультиколлинеарности

income 1,144

price 1,036

temp 1,144

Мультиколлинеарности нет.

В целом можно сделать вывод, что модель адекватна, доход и температура влияют на потребление мороженого, а вот его цена скорее всего нет.

Если средний еженедельный доход семьи увеличивается на 1 доллар, то потребление мороженого на одного покупателя увеличивается на 0,00330776 пинты.

Если средняя температура в градусах фаренгейта увеличивается на 1, то потребление мороженого на одного покупателя увеличивается на 0,00345843 пинты.

Упражнение 3.1. Анализ временных рядов.

В файлах Alum 2012_22.gdt, Brent 2012_22.gdt, Gas 2012_22.gdt, Gold 2012_22.gdt,

Silver 2012_22.gdt представлены данные (в формате Gretl.gdt) котировок на момент закрытия дневных сессий алюминия, нефти, газа, золота и серебра из базы данных мировых фондовых рынков с 2012 по 2022 г.г.

Для всех котировок из представленных файлов выполнить следующее:

1. Построить график временного ряда и визуально проанализировать его;
2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;
3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели;
4. Построить прогноз динамики временного ряда с использованием наилучшей модели.
5. Оформить отчет по результатам упражнения 3.1 в Word-файле.

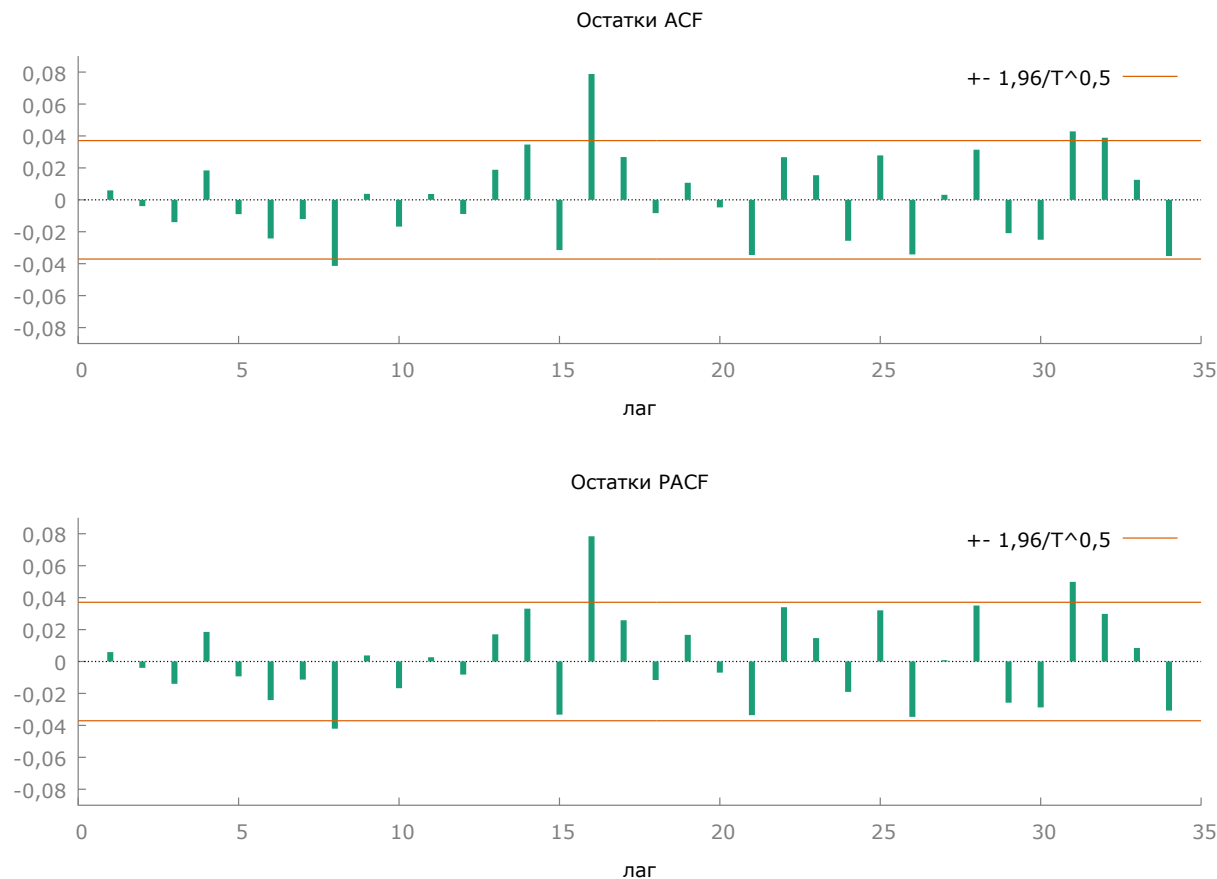
Alum (алюминий)

1. Построить график временного ряда и визуально проанализировать его;



На графике есть признаки цикличности, что может быть обусловлено сезонными факторами.

2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;



Большинство лагов находятся внутри доверительных интервалов, что говорит о слабой автокорреляции.

3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели;

Подберем разные параметры модели ARIMA

1, 0, 0:

Модель 4: ARMA, использованы наблюдения 1-2793
Зависимая переменная: CLOSE

Коэффициент	Ст. ошибка	z	p-значение
t			

phi_1	0,999915	не определено	не определено	не определено
Среднее завис. перемен	1979,474	Ст. откл. завис. перемен	374,4363	
Среднее инноваций	0,334315	Ст. откл. инноваций	28,47667	
R-квадрат	0,994221	Исправ. R-квадрат	0,994221	
Лог. правдоподобие	-13321,43	Крит. Акаике	26646,86	
Крит. Шварца	26658,73	Крит. Хеннана-Куинна	26651,15	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	1,0001	0,0000	1,0001	0,0000

Средняя абсолютная процентная ошибка (MAPE) 0,92677 – отличное качество модели

Отвергаем эту модель, так как показатели при переменной не определены.

1, 1, 0:

Модель 5: ARIMA, использованы наблюдения 2-2793 (T = 2792)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>
	<i>m</i>			
phi_1	-0,0208712	0,0189316	-1,102	0,2703

Среднее завис. перемен	0,157593	Ст. откл. завис. перемен	28,48184
Среднее инноваций	0,160896	Ст. откл. инноваций	28,47098
R-квадрат	0,994225	Исправ. R-квадрат	0,994225
Лог. правдоподобие	-13311,76	Крит. Акаике	26627,53
Крит. Шварца	26639,40	Крит. Хеннана-Куинна	26631,81

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-47,9130	0,0000	47,9130	0,5000

Средняя абсолютная процентная ошибка (MAPE) 0,92683 – отличное качество модели

Тем не менее, также отвергаем эту модель, так как p-значение при переменной больше 0,1.

0, 1, 1:

Модель 6: ARIMA, использованы наблюдения 2-2793 (T = 2792)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>
	<i>m</i>			
theta_1	-0,0202897	0,0186764	-1,086	0,2773

Среднее завис. перемен	0,157593	Ст. откл. завис. перемен	28,48184
Среднее инноваций	0,160873	Ст. откл. инноваций	28,47115
R-квадрат	0,994225	Исправ. R-квадрат	0,994225
Лог. правдоподобие	-13311,78	Крит. Акаике	26627,56
Крит. Шварца	26639,43	Крит. Хеннана-Куинна	26631,85

		<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
МА					
	Корень 1	49,2862	0,0000	49,2862	0,0000

Снова заметим, что р-значение больше 0,1 – отвергаем эту модель.

1, 1, 1:

Модель 7: ARIMA, использованы наблюдения 2-2793 (Т = 2792)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,713217	0,190748	-3,739	0,0002	***
theta_1	0,687590	0,197360	3,484	0,0005	***

Среднее завис. перемен	0,157593	Ст. откл. завис. перемен	28,48184
Среднее инноваций	0,159690	Ст. откл. инноваций	28,45818
R-квадрат	0,994230	Исправ. R-квадрат	0,994228
Лог. правдоподобие	-13310,51	Крит. Акаике	26627,02
Крит. Шварца	26644,82	Крит. Хеннана-Куинна	26633,45

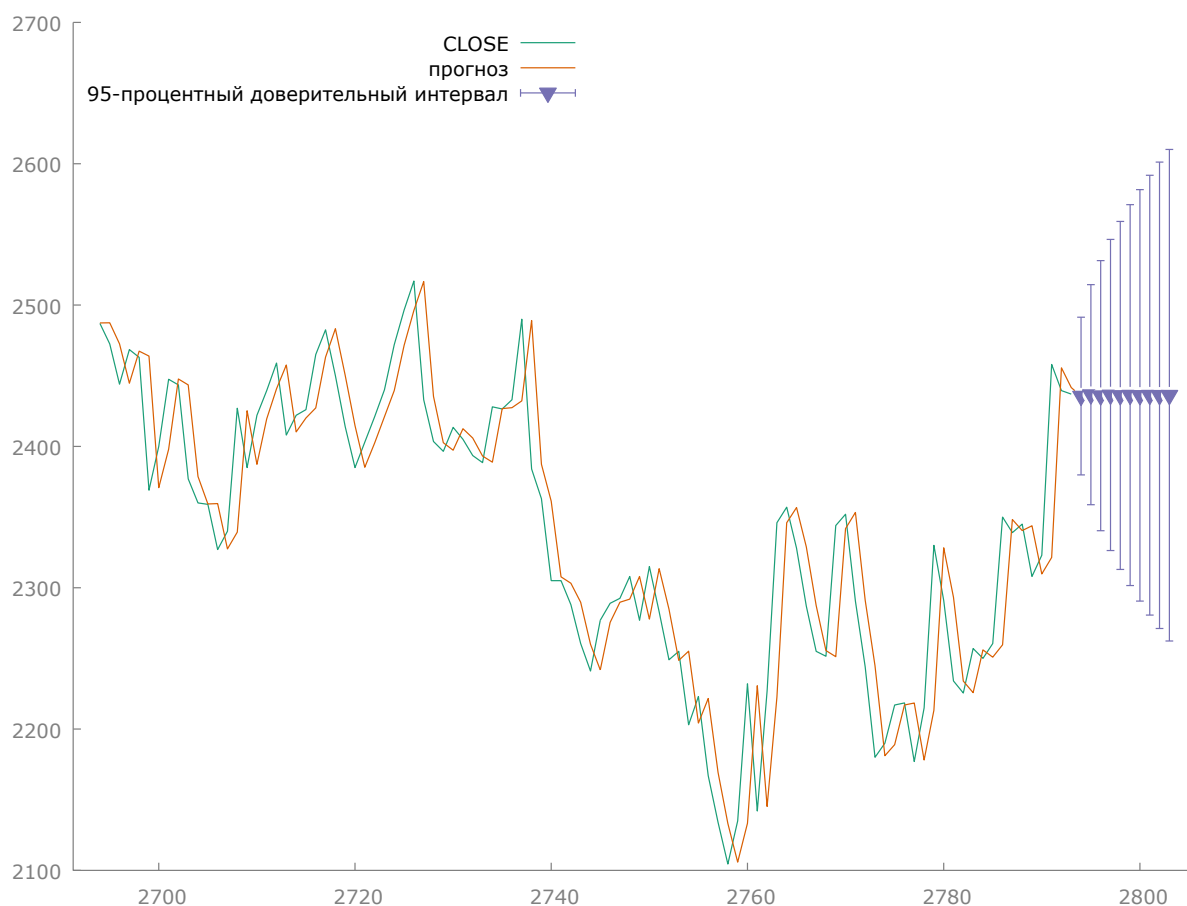
		<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR					
	Корень 1	-1,4021	0,0000	1,4021	0,5000
МА					
	Корень 1	-1,4544	0,0000	1,4544	0,5000

Средняя абсолютная процентная ошибка (MAPE) 0,92672 – отличное качество

р-значения при переменных низкие, они значимы с 99% вероятностью, модель удовлетворительна.

Модель 1, 1, 1 – лучшая, так как в ней значимы переменные и средняя абсолютная процентная ошибка показала отличное качество.

4. Построить прогноз динамики временного ряда с использованием наилучшей модели



Для 95% доверительных интервалов, $z(0,025) = 1,96$

Набл.	CLOSE	прогнози- рование	ст. ошибка	95% доверительный интервал
2794	не определено	2435,62	28,4582	(2379,84, 2491,39)
2795	не определено	2436,60	39,7336	(2358,73, 2514,48)
2796	не определено	2435,90	48,7521	(2340,35, 2531,45)
2797	не определено	2436,40	56,1601	(2326,33, 2546,47)
2798	не определено	2436,04	62,8169	(2312,92, 2559,16)
2799	не определено	2436,30	68,7560	(2301,54, 2571,06)
2800	не определено	2436,12	74,2721	(2290,55, 2581,69)
2801	не определено	2436,25	79,3721	(2280,68, 2591,81)
2802	не определено	2436,15	84,1864	(2271,15, 2601,16)
2803	не определено	2436,22	88,7245	(2262,32, 2610,12)

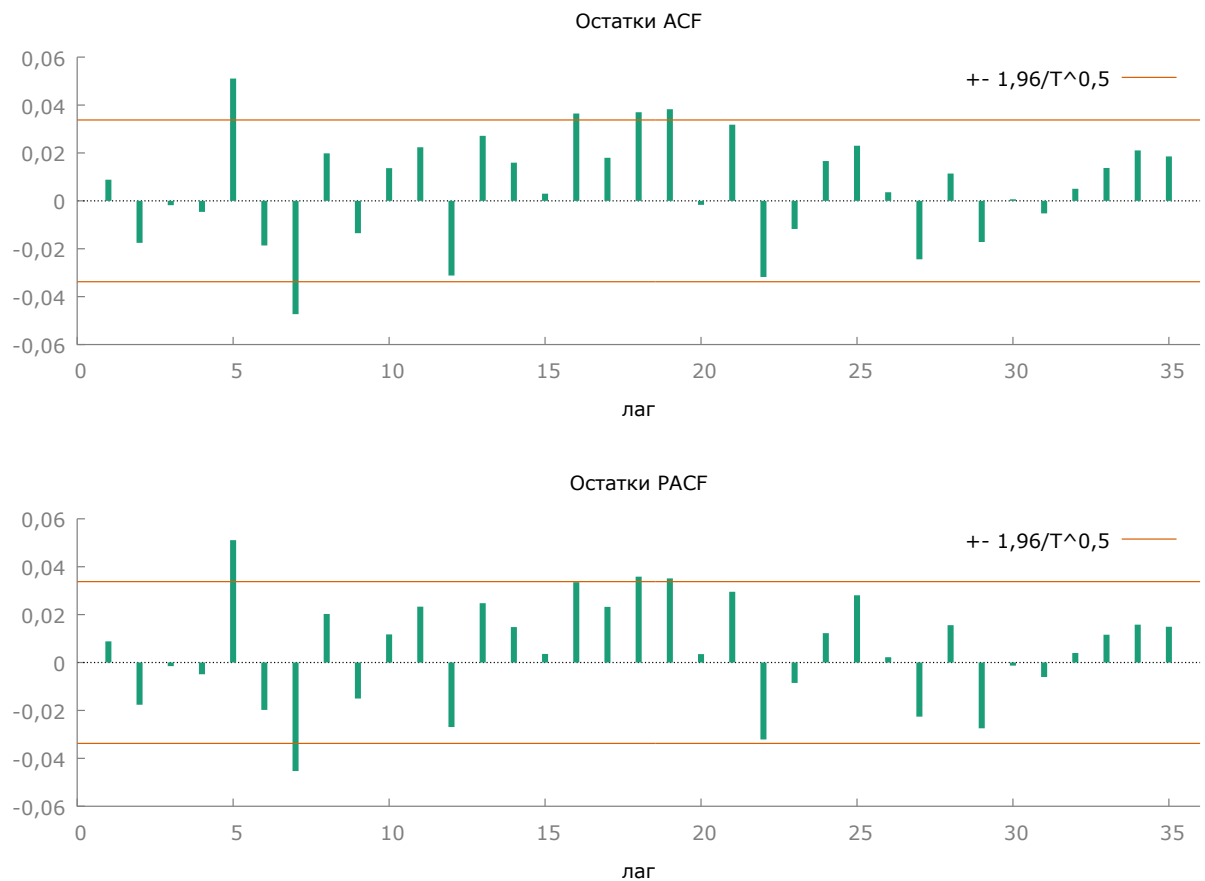
Brent (нефть)

1. Построить график временного ряда и визуально проанализировать его;



График демонстрирует значительные колебания цены на нефть, что является характерным для этого рынка.

2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;



Большинство лагов находятся внутри доверительных интервалов, что говорит о слабой автокорреляции.

3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели;

1, 1, 0:

Модель 2: ARIMA, использованы наблюдения 2-3368 ($T = 3367$)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	Коэффициент	Ст. ошибка	z	p-значение
	<i>m</i>			
phi_1	-0,0256707	0,0172302	-1,490	0,1363

Среднее завис. перемен	-0,005426	Ст. откл. завис. перемен	1,402769
Среднее инноваций	-0,005576	Ст. откл. инноваций	1,402109
R-квадрат	0,997057	Исправ. R-квадрат	0,997057
Лог. правдоподобие	-5915,536	Крит. Акаике	11835,07
Крит. Шварца	11847,32	Крит. Хеннана-Куинна	11839,45

	Действительная часть	Мнимая часть	Модуль	Частота
AR				

Корень 1 -38,9550 0,0000 38,9550 0,5000
P-значение выше 0,1 – модель неудовлетворительна.

0, 1, 1:

Модель 6: ARIMA, использованы наблюдения 2-3368 (T = 3367)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>
	<i>m</i>			
theta_1	-0,0280208	0,0180078	-1,556	0,1197
Среднее завис. перемен	-0,005426	Ст. откл. завис. перемен		1,402769
Среднее инноваций	-0,005593	Ст. откл. инноваций		1,402067
R-квадрат	0,997057	Исправ. R-квадрат		0,997057
Лог. правдоподобие	-5915,435	Крит. Акаике		11834,87
Крит. Шварца	11847,11	Крит. Хеннана-Куинна		11839,25

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
МА				
Корень 1	35,6878	0,0000	35,6878	0,0000

P-значение выше 0,1 – модель неудовлетворительна.

1, 1, 1:

Модель 7: ARIMA, использованы наблюдения 2-3368 (T = 3367)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	0,652921	0,146681	4,451	<0,0001	***
theta_1	-0,689496	0,140022	-4,924	<0,0001	***
Среднее завис. перемен	-0,005426	Ст. откл. завис. перемен		1,402769	
Среднее инноваций	-0,006041	Ст. откл. инноваций		1,400937	
R-квадрат	0,997062	Исправ. R-квадрат		0,997061	
Лог. правдоподобие	-5912,724	Крит. Акаике		11831,45	
Крит. Шварца	11849,81	Крит. Хеннана-Куинна		11838,01	

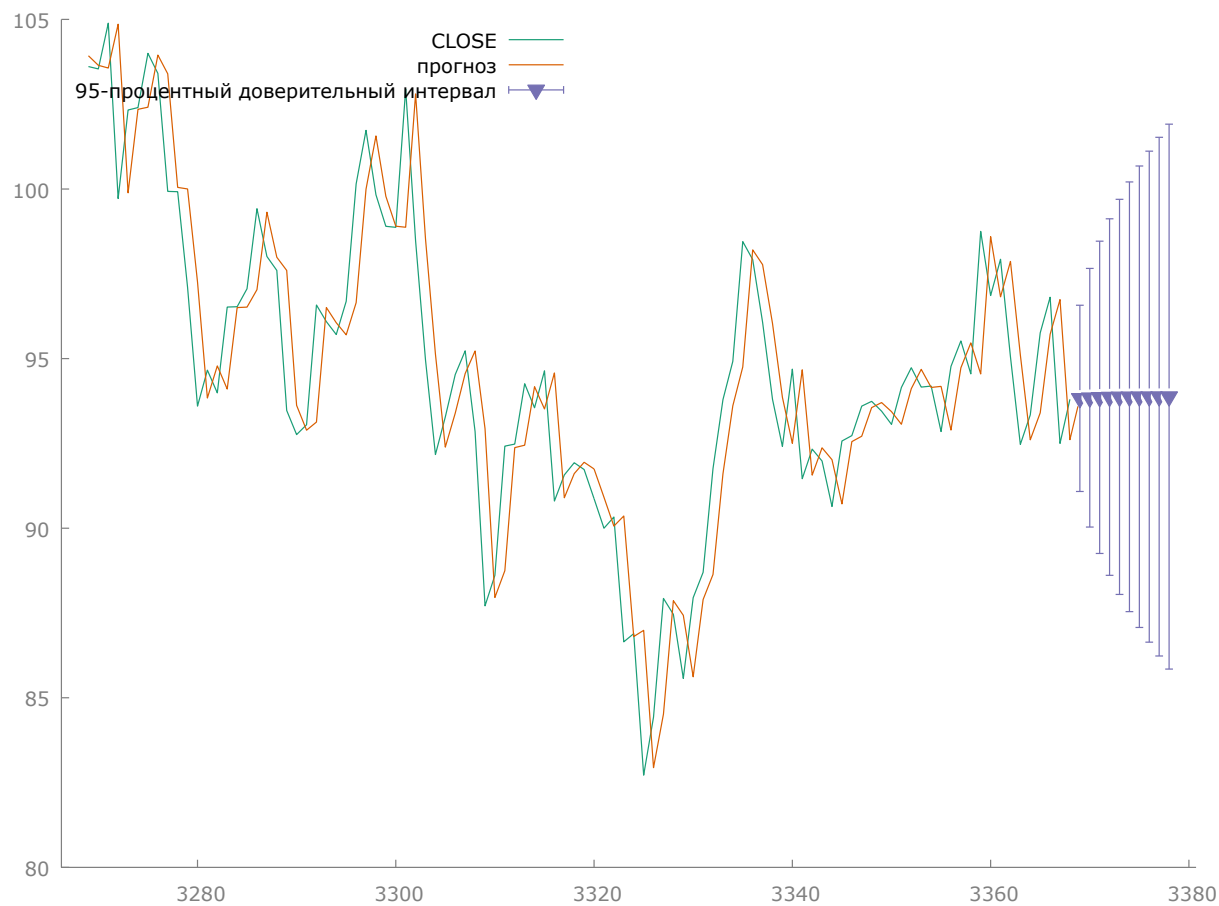
	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	1,5316	0,0000	1,5316	0,0000
МА				
Корень 1	1,4503	0,0000	1,4503	0,0000

Средняя абсолютная процентная ошибка (MAPE) 1,3856 – отличное качество

Все переменные значимы с 99% вероятностью. Выбираем эту модель.

4. Построить прогноз динамики временного ряда с использованием наилучшей

модели.



Для 95% доверительных интервалов, $z(0,025) = 1,96$

Набл.	CLOSE	прогнози- рование	ст. ошибка	95% доверительный интервал
3369	не определено	93,8279	1,40094	(91,0821, 96,5736)
3370	не определено	93,8460	1,94533	(90,0333, 97,6588)
3371	не определено	93,8579	2,34879	(89,2544, 98,4615)
3372	не определено	93,8657	2,68184	(88,6094, 99,1220)
3373	не определено	93,8707	2,97171	(88,0463, 99,6952)
3374	не определено	93,8740	3,23203	(87,5394, 100,209)
3375	не определено	93,8762	3,47068	(87,0738, 100,679)
3376	не определено	93,8776	3,69258	(86,6403, 101,115)
3377	не определено	93,8785	3,90105	(86,2326, 101,524)
3378	не определено	93,8791	4,09841	(85,8464, 101,912)

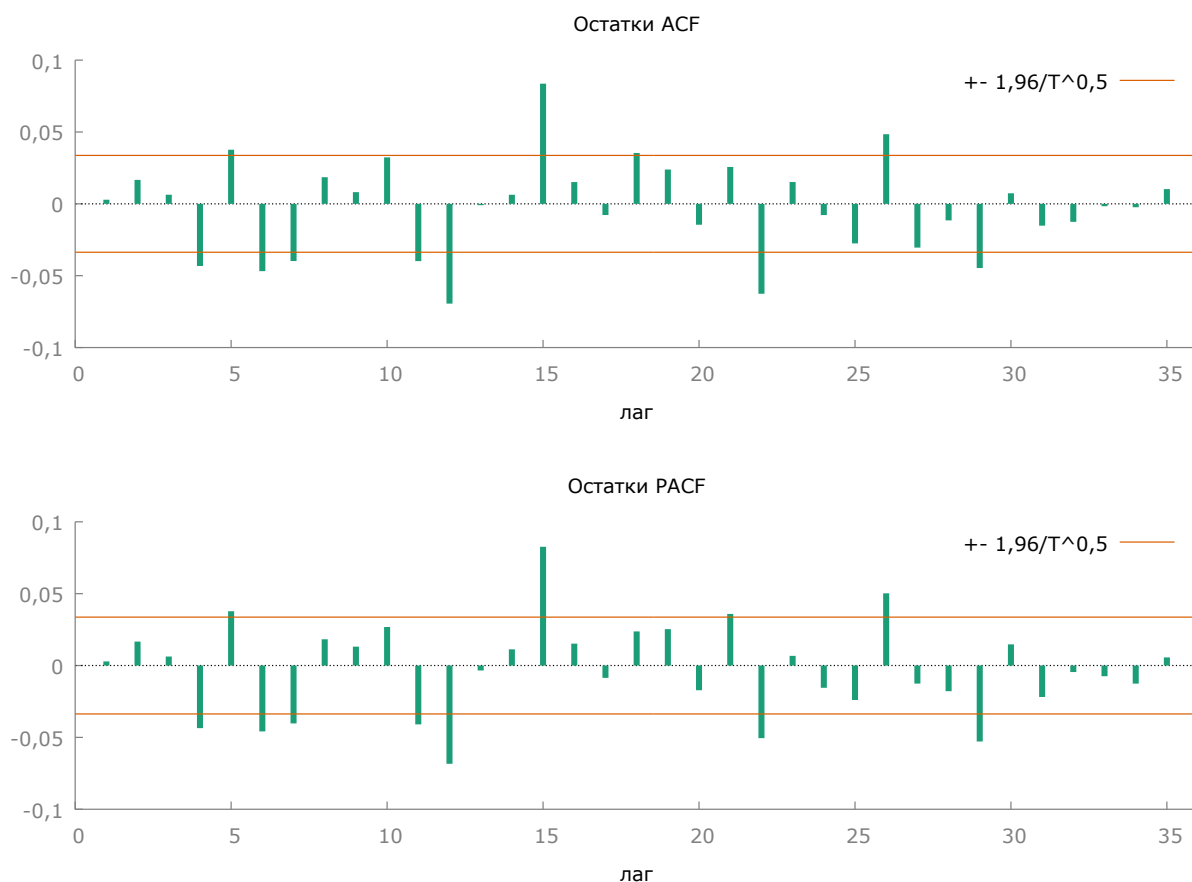
Gas (газ)

1. Построить график временного ряда и визуально проанализировать его;



На графике можно заметить циклы резкого роста и падения цен, а также колебания, предположительно связанные с сезонностью.

2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;



Большинство лагов находятся внутри доверительных интервалов, что говорит о слабой автокорреляции.

3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели;

1, 1, 0:

Модель 3: ARIMA, использованы наблюдения 2-3385 ($T = 3384$)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,0648790	0,0171532	-3,782	0,0002	***
Среднее завис. перемен	0,000941	Ст. откл. завис. перемен	0,122057		
Среднее инноваций	0,001000	Ст. откл. инноваций	0,121785		
R-квадрат	0,991549	Исправ. R-квадрат	0,991549		
Лог. правдоподобие	2323,311	Крит. Акаике	-4642,622		
Крит. Шварца	-4630,368	Крит. Хеннана-Куинна	-4638,241		
	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>	
AR					
Корень 1	-15,4133	0,0000	15,4133	0,5000	

Средняя абсолютная процентная ошибка (MAPE) 2,0505 – отличное качество

0, 1, 1:

Модель 4: ARIMA, использованы наблюдения 2-3385 (T = 3384)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
theta_1	-0,0608405	0,0165874	-3,668	0,0002	***
Среднее завис. перемен	0,000941	Ст. откл. завис. перемен	0,122057		
Среднее инноваций	0,001000	Ст. откл. инноваций	0,121801		
R-квадрат	0,991546	Исправ. R-квадрат	0,991546		
Лог. правдоподобие	2322,865	Крит. Акаике	-4641,730		
Крит. Шварца	-4629,476	Крит. Хеннана-Куинна	-4637,349		

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
MA				
Корень 1	16,4364	0,0000	16,4364	0,0000

Средняя абсолютная процентная ошибка (MAPE) 2,0498 – отличное качество модели

1, 1, 1:

Модель 5: ARIMA, использованы наблюдения 2-3385 (T = 3384)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,302059	0,147046	-2,054	0,0400	**
theta_1	0,236071	0,149335	1,581	0,1139	
Среднее завис. перемен	0,000941	Ст. откл. завис. перемен	0,122057		
Среднее инноваций	0,000990	Ст. откл. инноваций	0,121751		
R-квадрат	0,991554	Исправ. R-квадрат	0,991551		
Лог. правдоподобие	2324,262	Крит. Акаике	-4642,525		
Крит. Шварца	-4624,144	Крит. Хеннана-Куинна	-4635,954		

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-3,3106	0,0000	3,3106	0,5000
MA				
Корень 1	-4,2360	0,0000	4,2360	0,5000

P-значение theta_1 показывает, что переменная незначима, модель неудовлетворительна

Сравним модели, определим, что лучшие показатели имеет модель 0, 1, 1 (более низкая MAPE, ниже стандартные ошибки)

4. Построить прогноз динамики временного ряда с использованием наилучшей модели.



Для 95% доверительных интервалов, $z(0,025) = 1,96$

Набл.	CLOSE	прогнози- рование	ст. ошибка	95% доверительный интервал
3386	не определено	6,15493	0,121801	(5,91620, 6,39365)
3387	не определено	6,15493	0,167095	(5,82743, 6,48243)
3388	не определено	6,15493	0,202499	(5,75804, 6,55182)
3389	не определено	6,15493	0,232575	(5,69909, 6,61077)
3390	не определено	6,15493	0,259184	(5,64694, 6,66292)
3391	не определено	6,15493	0,283305	(5,59966, 6,71020)
3392	не определено	6,15493	0,305527	(5,55611, 6,75375)
3393	не определено	6,15493	0,326240	(5,51551, 6,79435)
3394	не определено	6,15493	0,345713	(5,47734, 6,83251)

3395 не 6,15493 0,364147 (5,44121, 6,86864)

определено

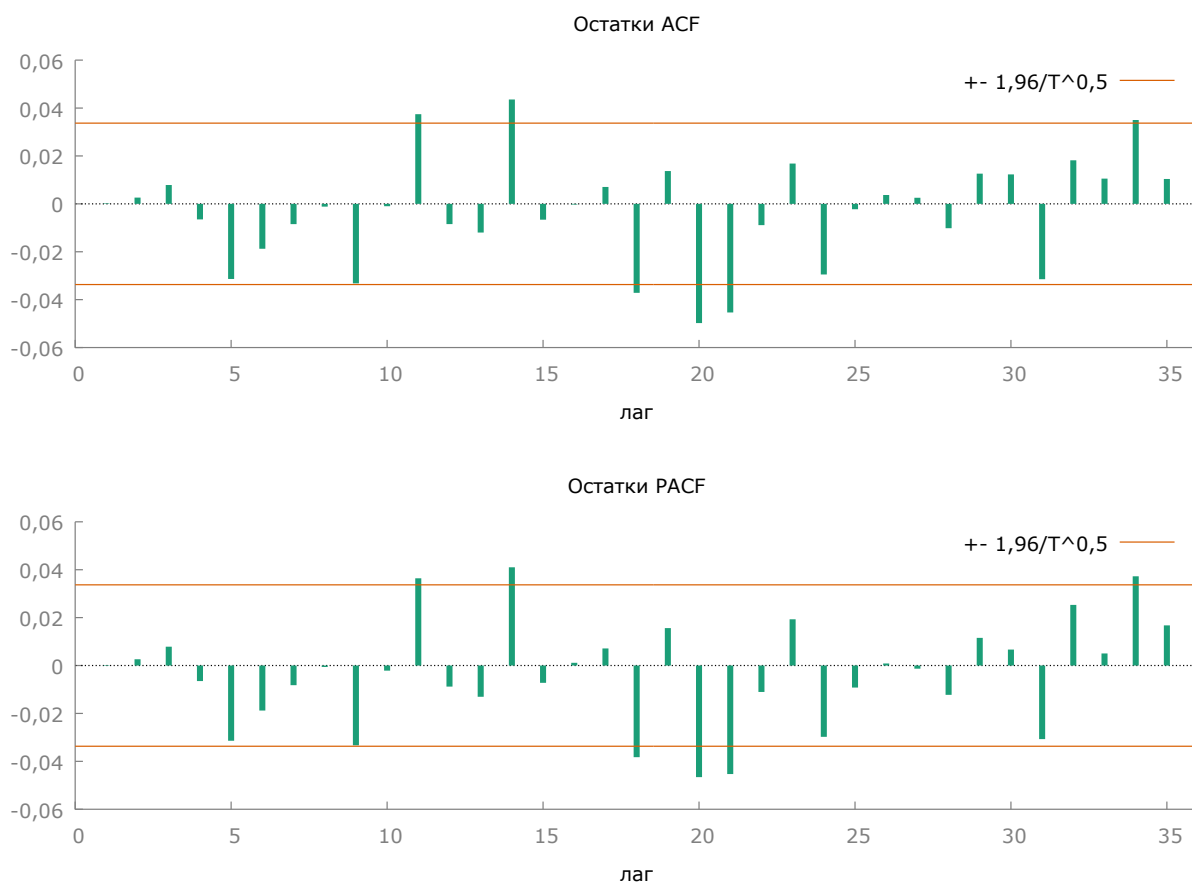
Gold(золото):

1. Построить график временного ряда и визуально проанализировать его;



Визуально можно заметить сильные колебания, предположительно, они связаны с сезонностью.

2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;



Большинство лагов находятся внутри доверительных интервалов, что говорит о слабой автокорреляции.

3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели;

1, 1, 0:

Модель 2: ARIMA, использованы наблюдения 2-3385 ($T = 3384$)

Зависимая переменная: $(1-L)$ CLOSE

Стандартные ошибки рассчитаны на основе Гессiana

	Коэффициент	Ст. ошибка	z	p-значение
<i>m</i>				
phi_1	-0,0247012	0,0171868	-1,437	0,1507
Среднее завис. перемен	0,058126	Ст. откл. завис. перемен	13,34810	
Среднее инноваций	0,059495	Ст. откл. инноваций	13,34218	
R-квадрат	0,997238	Исправ. R-квадрат	0,997238	
Лог. правдоподобие	-13569,40	Крит. Акаике	27142,79	
Крит. Шварца	27155,05	Крит. Хеннана-Куинна	27147,17	
	Действительная часть	Мнимая часть	Модуль	Частота
AR				
Корень 1	-40,4839	0,0000	40,4839	0,5000

Средняя абсолютная процентная ошибка (MAPE) 0,60479 – отличное качество модели

P-значение больше 0,1 – модель неудовлетворительная.

0, 1, 1:

Модель 3: ARIMA, использованы наблюдения 2-3385 (T = 3384)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гессииана

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>
	<i>m</i>			
theta_1	-0,0243751	0,0170642	-1,428	0,1532
Среднее завис. перемен	0,058126	Ст. откл. завис. перемен	13,34810	
Среднее инноваций	0,059510	Ст. откл. инноваций	13,34223	
R-квадрат	0,997238	Исправ. R-квадрат	0,997238	
Лог. правдоподобие	-13569,41	Крит. Акаике	27142,82	
Крит. Шварца	27155,07	Крит. Хеннана-Куинна	27147,20	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
MA				
Корень 1	41,0255	0,0000	41,0255	0,0000

Средняя абсолютная процентная ошибка (MAPE) 0,6048 – качество модели отличное

P-значение больше 0,1 – модель неудовлетворительная.

1, 1, 1:

Модель 4: ARIMA, использованы наблюдения 2-3385 (T = 3384)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гессииана

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>
	<i>m</i>			
phi_1	-0,175202	0,540691	-0,3240	0,7459
theta_1	0,150423	0,542854	0,2771	0,7817
Среднее завис. перемен	0,058126	Ст. откл. завис. перемен	13,34810	
Среднее инноваций	0,059328	Ст. откл. инноваций	13,34202	
R-квадрат	0,997238	Исправ. R-квадрат	0,997238	
Лог. правдоподобие	-13569,36	Крит. Акаике	27144,72	
Крит. Шварца	27163,10	Крит. Хеннана-Куинна	27151,29	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-5,7077	0,0000	5,7077	0,5000
MA				
Корень 1	-6,6479	0,0000	6,6479	0,5000

Средняя абсолютная процентная ошибка (MAPE) 0,60481 – качество модели отличное

P-значения сильно больше 0,1 – модель неудовлетворительная.

2, 2, 0:

Модель 6: ARIMA, использованы наблюдения 3-3385 (T = 3383)

Зависимая переменная: (1-L)^2 CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,691125	0,0161612	-42,76	<0,0001	***
phi_2	-0,340820	0,0161606	-21,09	<0,0001	***

Среднее завис. перемен	-0,001685	Ст. откл. завис. перемен	19,10944
Среднее инноваций	-0,005998	Ст. откл. инноваций	15,39102
R-квадрат	0,996337	Исправ. R-квадрат	0,996336
Лог. правдоподобие	-14048,94	Крит. Акаике	28103,88
Крит. Шварца	28122,26	Крит. Хеннана-Куинна	28110,45

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-1,0139	-1,3806	1,7129	-0,3508
Корень 2	-1,0139	1,3806	1,7129	0,3508

Средняя абсолютная процентная ошибка (MAPE) 0,73507 – качество модели отличное

Здесь p-значения меньше 0,01 – переменные значимы с 99% вероятностью, эту модель мы и будем считать оптимальной.

4. Построить прогноз динамики временного ряда с использованием наилучшей модели.



Для 95% доверительных интервалов, $z(0,025) = 1,96$

Набл.	CLOSE	прогнози- рование	ст. ошибка	95% доверительный интервал
3386	не определено	1784,17	15,3910	(1754,00, 1814,33)
3387	не определено	1790,30	25,3516	(1740,61, 1839,99)
3388	не определено	1796,09	37,0400	(1723,49, 1868,69)
3389	не определено	1800,83	51,6764	(1699,55, 1902,12)
3390	не определено	1806,41	67,1291	(1674,84, 1937,99)
3391	не определено	1811,77	83,9746	(1647,19, 1976,36)
3392	не определено	1817,00	102,209	(1616,68, 2017,33)
3393	не определено	1822,40	121,484	(1584,29, 2060,50)
3394	не определено	1827,72	141,874	(1549,65, 2105,79)
3395	не определено	1833,04	163,312	(1512,95, 2153,12)

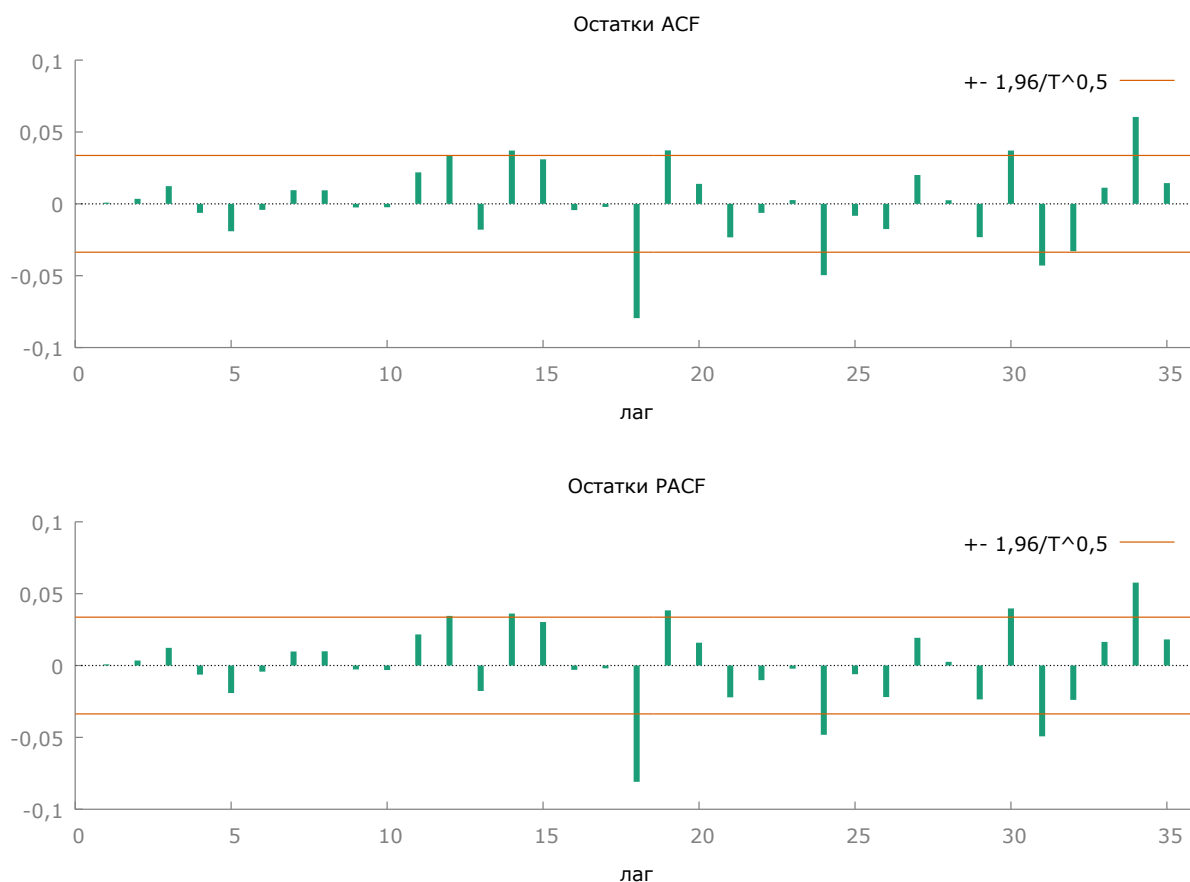
Silver (серебро)

1. Построить график временного ряда и визуально проанализировать его;



На графике можно заметить циклы резкого роста и падения цен, а также колебания, предположительно связанные с сезонностью.

2. Построить коррелограмму и проанализировать поведение функций ACF, PACF;



Большинство лагов находятся внутри доверительных интервалов, что говорит о слабой автокорреляции.

3. Подобрать наилучшую модель ARIMA(p,k,q) и обосновать выбор этой наилучшей модели
1, 1, 0:

Модель 2: ARIMA, использованы наблюдения 2-3391 (T = 3390)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,104807	0,0171342	-6,117	<0,0001	***
Среднее завис. перемен	-0,001854	Ст. откл. завис. перемен		0,363177	
Среднее инноваций	-0,002039	Ст. откл. инноваций		0,361140	
R-квадрат	0,995236	Исправ. R-квадрат		0,995236	
Лог. правдоподобие	-1357,526	Крит. Акаике		2719,051	
Крит. Шварца	2731,308	Крит. Хеннана-Куинна		2723,433	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-9,5413	0,0000	9,5413	0,5000
Средняя абсолютная процентная ошибка (MAPE) 1,0826 – отличное качество модели				

Модель удовлетворительна.

0, 1, 1:

Модель 3: ARIMA, использованы наблюдения 2-3391 (T = 3390)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
theta_1	-0,100944	0,0167138	-6,040	<0,0001	***
Среднее завис. перемен	-0,001854	Ст. откл. завис. перемен		0,363177	
Среднее инноваций	-0,002053	Ст. откл. инноваций		0,361212	
R-квадрат	0,995234	Исправ. R-квадрат		0,995234	
Лог. правдоподобие	-1358,200	Крит. Акаике		2720,400	
Крит. Шварца	2732,657	Крит. Хеннана-Куинна		2724,782	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
MA				
Корень 1	9,9065	0,0000	9,9065	0,0000
Средняя абсолютная процентная ошибка (MAPE)	1,0822 – отличное качество модели			

Модель удовлетворительна.

1, 1, 1:

Модель 4: ARIMA, использованы наблюдения 2-3391 (T = 3390)

Зависимая переменная: (1-L) CLOSE

Стандартные ошибки рассчитаны на основе Гесса

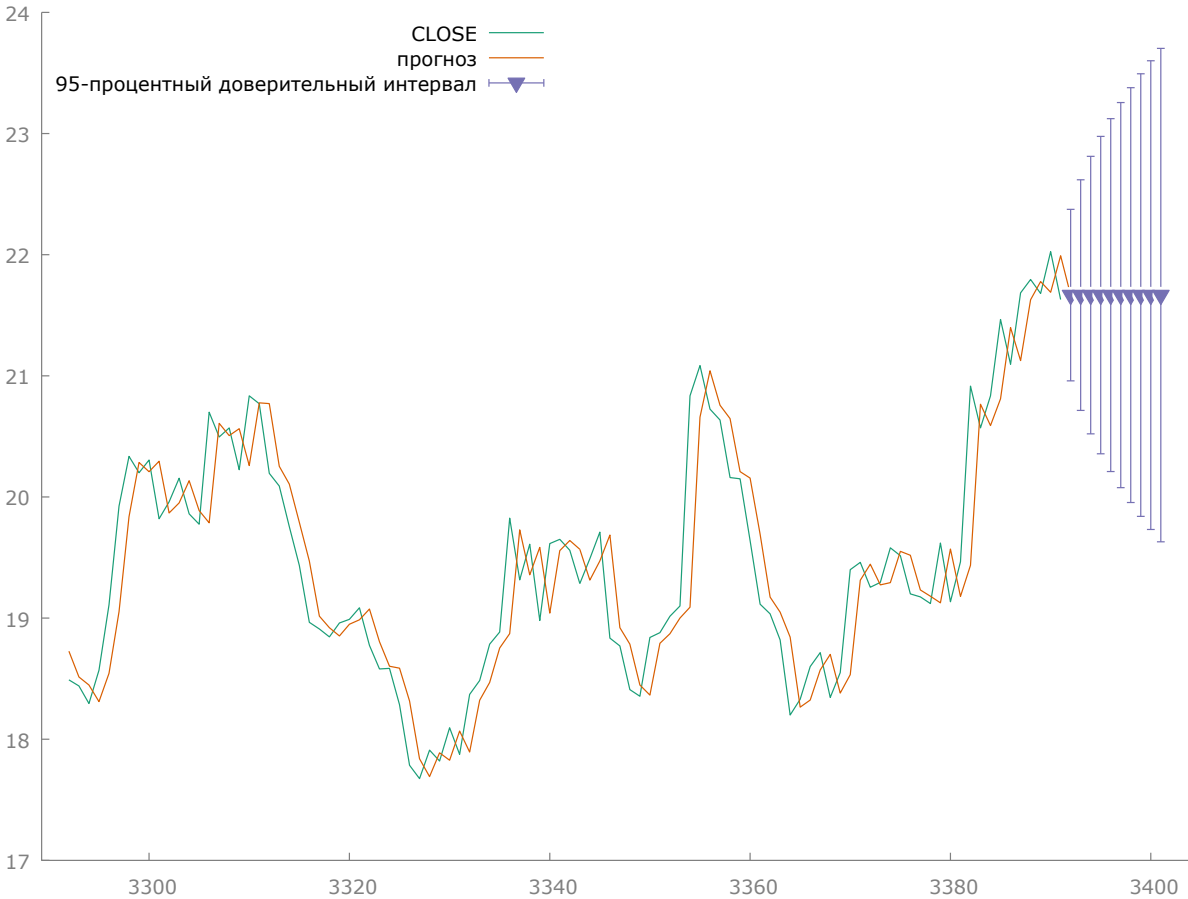
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>z</i>	<i>p-значение</i>	
	<i>m</i>				
phi_1	-0,198791	0,143715	-1,383	0,1666	
theta_1	0,0949021	0,145753	0,6511	0,5150	
Среднее завис. перемен	-0,001854	Ст. откл. завис. перемен		0,363177	
Среднее инноваций	-0,002020	Ст. откл. инноваций		0,361117	
R-квадрат	0,995237	Исправ. R-квадрат		0,995236	
Лог. правдоподобие	-1357,316	Крит. Акаике		2720,632	
Крит. Шварца	2739,018	Крит. Хеннана-Куинна		2727,204	

	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
AR				
Корень 1	-5,0304	0,0000	5,0304	0,5000
MA				
Корень 1	-10,5372	0,0000	10,5372	0,5000

P-значения переменных больше 0,1- модель неудовлетворительна.

Сравнив результаты трех моделей, выберем модель 0, 1, 1.

4. Построить прогноз динамики временного ряда с использованием наилучшей модели.



Для 95% доверительных интервалов, $z(0,025) = 1,96$

Набл.	CLOSE	прогнози- рование	ст. ошибка	95% доверительный интервал
3392	не определено	21,6665	0,361212	(20,9585, 22,3744)
3393	не определено	21,6665	0,485733	(20,7144, 22,6185)
3394	не определено	21,6665	0,584293	(20,5213, 22,8117)
3395	не определено	21,6665	0,668476	(20,3563, 22,9766)
3396	не определено	21,6665	0,743184	(20,2098, 23,1231)
3397	не определено	21,6665	0,811040	(20,0768, 23,2561)
3398	не определено	21,6665	0,873640	(19,9542, 23,3788)
3399	не определено	21,6665	0,932046	(19,8397, 23,4932)
3400	не определено	21,6665	0,987002	(19,7320, 23,6009)
3401	не определено	21,6665	1,03905	(19,6299, 23,7030)

