ЛАБОРАТОРНАЯ РАБОТА № 6
ПРОВЕРКА СООТВЕТСТВИЯ УРАВНЕНИЯ ЛИНИИ РЕГРЕССИИ ПРЕДПОСЫЛКАМ МНК

Цель: изучить свойства параметров оценок уравнения регрессии с помощью проверки их на соответствие условиям Гаусса-Маркова.

Задание:

1 Построить линейную регрессионную модель зависимости курса евро от фактора времени;

2 Проверить статистическую значимость полученных оценок параметров уравнения регрессии и качество модели;

3 Проверить соответствие оценок параметров уравнения регрессии предпосылкам МНК и оценить последствия их невыполнимости.

Предпосылки МНК (Условия Гаусса-Маркова)

1 Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю:

$$M(\varepsilon_i) = 0, i = 1,2,..,n.$$

2 Гомоскедастичность (постоянство дисперсии отклонений). Дисперсия случайных отклонений ε_i постоянна:

$$D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2, \forall i, j.$$

3 Случайные отклонения ε_i и ε_j должны быть независимы друг от друга для всех $i \neq j$:

$$cov(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j \end{cases};$$

4 Случайное отклонение должно быть независимо от объясняющих переменных. Значение любой независимой переменной в каждом наблюдении должно быть полностью определяемым внешними причинами, не учитываемыми в уравнении регрессии

$$\sigma_{\varepsilon_i x_j} = 0.$$

5 Ошибки $\varepsilon_i, i = 1,2,..,n$ имеют нормальное распределение $\varepsilon_i \sim N(0, \sigma)$.

Теорема Гаусса-Маркова

Если предпосылки 1-5 выполнены, то оценки, полученные по МНК, обладают следующими свойствами:

1 Оценки являются несмещенными, то есть $M(b_0) = \beta_0, M(b_1) = \beta_1$. Это вытекает их того, что $M(\varepsilon_i) = 0$, и это говорит об отсутствии систематической ошибки в определении положения линии регрессии.

2 Оценки состоятельны, так как дисперсия оценок параметров при возрастании числа наблюдений n стремится к нулю:

$$D(b_0) \to 0, D(b_1) \to 0, n \to \infty$$

То есть при увеличении объема выборки надежность оценок увеличивается.

3 Оценки эффективны, то есть имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров.

Порядок выполнения лабораторной работы

Приступая к выполнению данной работы, студенту необходимо заранее подготовить исходные данные (курс валюты за любой месяц года). Наименование валюты выбирается согласно номеру своего варианта из таблицы 3.1. Номер варианта определяется по номеру студента по списку (уточняется у преподавателя).

Таблица 3.1 – Варианты заданий

Номер варианта	Наименование валюты	Обозначение
1	Австралийский доллар	AUD
2	Доллар США	USD
3	Евро	EUR
4	Аргентинское песо	ARS
5	Датская крона	DKK
6	Израильский шекель	ILS
7	Ирландский фунт	IEP
8	Канадский доллар	CAD
9	Швейцарский франк	CHF
10	Мексиканское песо	MXN

Рассмотрим пример.

Исходные данные для выполнения работы представлены в таблице 3.2.

Таблица 3.2 – Динамика курса ЕВРО

Дата	Курс евро, руб.	Дата	Курс евро, руб.
01.10.2016	70,93	17.10.2016	69,37
02.10.2016	70,93	18.10.2016	69,36
03.10.2016	70,93	19.10.2016	69,26
04.10.2016	70,24	20.10.2016	68,77
05.10.2016	69,76	21.10.2016	68,47
06.10.2016	70,08	22.10.2016	68,03
07.10.2016	69,85	23.10.2016	68,03
08.10.2016	69,23	24.10.2016	68,03
09.10.2016	69,23	25.10.2016	67,73
10.10.2016	69,23	26.10.2016	67,5
11.10.2016	69,73	27.10.2016	67,93
12.10.2016	69,12	28.10.2016	68,76
13.10.2016	69,12	29.10.2016	68,68
14.10.2016	69,73	30.10.2016	68,68
15.10.2016	69,37	31.10.2016	68,68
16.10.2016	69,37		

1 На основе данных из таблицы 3.1 необходимо построить линейную регрессионную модель зависимости курса иностранной валюты от фактора времени:

$$Y=b_0+b_1t, \tag{3.6}$$

Оценить коэффициенты регрессии для линейной модели можно, используя инструмент MS Excel «Регрессия» (меню *Данные→Анализ данных*), Результат расчетов представлен на рисунке 3.1.

Значения коэффициентов модели b_0 и b_1 рассчитаны в ячейках B17 и B18 соответственно (рисунок 3.1).

Таким образом, регрессионная линейная модель зависимости курса доллара США от фактора времени выглядит следующим образом:

$$y = 3689{,}85 - 0{,}08 \cdot t, \tag{3.7}$$

где x – дата;

y – курс евро, руб.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,850119388							
5	R-квадрат	0,722702974							
6	Нормированный R-ква	0,713141008							
7	Стандартная ошибка	0,48618249							
8	Наблюдения	31							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*начимость F*			
12	Регрессия	1	17,86533875	17,86533875	75,5809991	1,4289E-09			
13	Остаток	29	6,854828992	0,236373414					
14	Итого	30	24,72016774						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
17	Y-пересечение	3689,848109	416,4704127	8,859808516	9,5455E-10	2838,07048	4541,626	2838,07	4541,626
18	Дата	-0,084875	0,009762779	-8,693733322	1,4289E-09	-0,1048421	-0,06491	-0,10484	-0,06491
19									
20									
21									
22	ВЫВОД ОСТАТКА								
23									
24	*Наблюдение*	*казанное Курс евро*	*Остатки*						
25	1	70,43860887	0,491391129						
26	2	70,35373387	0,576266129						
27	3	70,26885887	0,661141129						
28	4	70,18398387	0,056016129						

Рисунок 3.1 – Результаты регрессионного анализа

2 Оценить статистическую значимость полученных оценок параметров уравнения регрессии на основе критерия Стьюдента.

Статистическая значимость коэффициентов множественной линейной регрессии с n объясняющими переменными проверяется на основе t-статистики, имеющей в данной ситуации распределение Стьюдента с числом степеней свободы $\nu = n\text{-}m\text{-}1$ (n – объем выборки, m – количество независимых переменных). При требуемом уровне значимости наблюдаемое значение t-статистики сравнивается с критической точкой $t_{\frac{\alpha}{2},n-m-1}$ распределения Стьюдента. Если $|t| > t_{\frac{\alpha}{2},n-m-1}$, то коэффициент b_j считается статистически значимым. В противном случае ($|t| < t_{\frac{\alpha}{2},n-m-1}$) коэффициент b_j считается статистически незначимым (статистически близким к нулю). Это означает, что фактор X_j линейно не связан с зависимой переменной $У$.

Наблюдаемые значения t-статистик для коэффициентов b_0=3689,85 и b_1=-0,08 рассчитаны в ячейках D17 и D18 соответственно (рисунок 3.1).

Критическое значение $t_{\frac{\alpha}{2},n-m-1}$ можно найти с помощью встроенной статистической функции **СТЬЮДРАСПОБР** (рисунок 3.2).

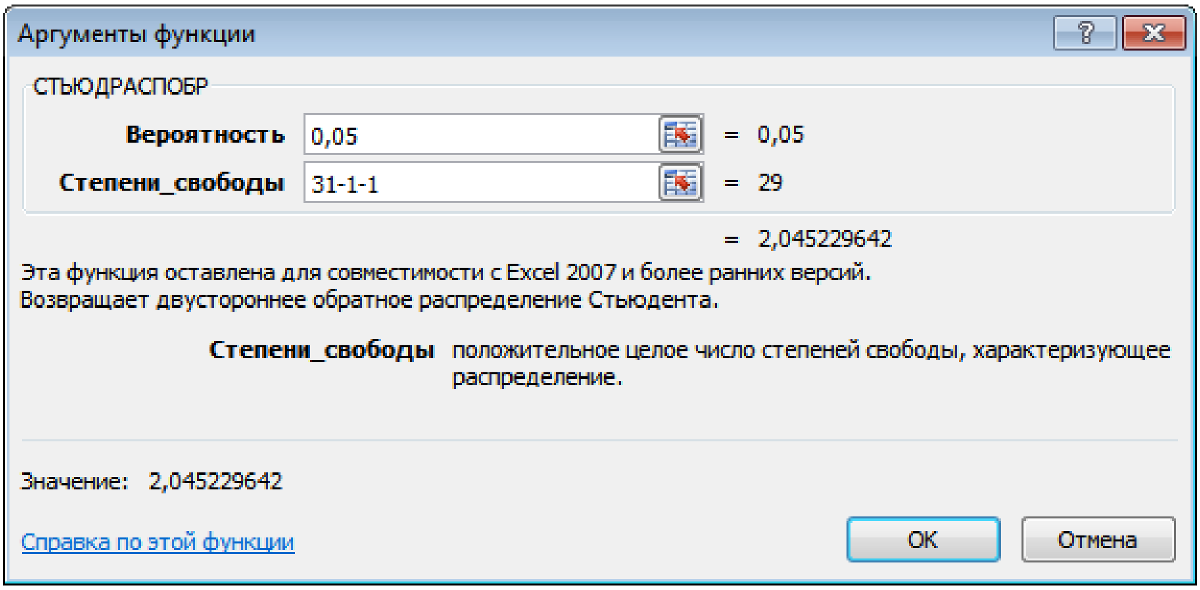

Рисунок 3.2 – Результат вычисления функции **СТЬЮДРАСПОБР**

Критическое значение $t_{\frac{\alpha}{2},n-m-1} = 2{,}045$, следовательно, оба коэффициента являются статистически значимыми ($\left|t_{b_0}\right| = 8{,}86 > 2{,}045$, $\left|t_{b_1}\right| = 8{,}69 > 2{,}045$).

3. С помощью критерия Фишера сделать выводы по качеству построенной модели.

Проверку качества модели осуществляют с помощью критерия Фишера (F-статистика). Расчетное значение F-статистики (рисунок 3.1, ячейка E12) сравнивают с критическим значением ($F_{\text{крит.}}$), которое можно найти по таблицам при заданном уровне значимости α и числе степеней свободы $v1=m$, $v2=n-m-1$ или с помощью статистической функции **FРАСПОБР** (рисунок 3.3). Если $F_{расч.} > F_{крит.}$, связь признается существенной, а модель адекватной опытным данным.

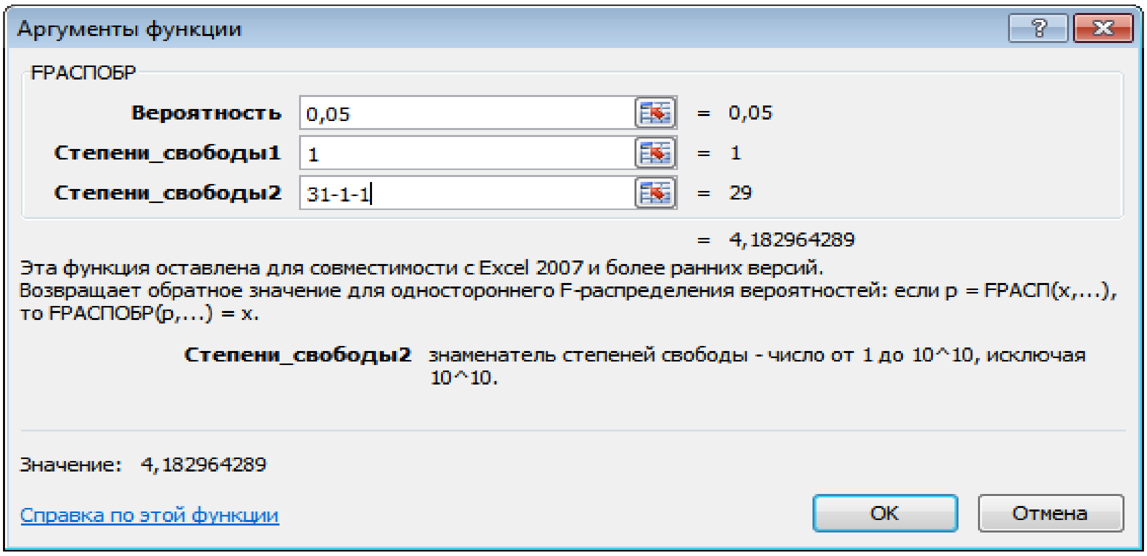

Рисунок 3.3 – Результат вычисления функции **FРАСПОБР**

$F_{расч.}=75{,}58 > F_{крит.}=4{,}183$, следовательно, связь можно признать существенной, а модель адекватной опытным данным.

4 Оценить соответствие коэффициентов регрессии **предпосылкам МНК.**

Проверка выполнения 1-ого условия. Фактически если уравнение регрессии включает постоянный член, то первое условие выполняется автоматически, так как роль константы состоит в определении любой систематической тенденции в зависимой переменной, которую не учитывают объясняющие переменные, включенные в уравнение регрессии.

Проверка выполнения 2-ого условия. Во втором условии Гаусса - Маркова утверждается, что дисперсия остатков в каждом наблюдении должна быть постоянной. Если дисперсия постоянна для всех наблюдений, то это явление называют гомоскедастичности, а если нет – гетероскедастичностью.

Выдвинем гипотезу H_0 об отсутствии гетероскедастичности.

Для обнаружения гетероскедастичности необходимо использовать тест Голфелда-Квандта.

Порядок проведения этого теста таков:

1 Все n наблюдений упорядочиваются по величине X. Упорядочить по дате можно с помощью меню *Данные→Сортировка*.

2 Вся упорядоченная выборка разбивается на 3 подвыборки размерностей k, $(n-2k)$, k соответственно. В примере это будут 3 подвыборки размерностью 10, 11 и 10 наблюдений.

3 Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) (рисунок 3.4) и для третьей подвыборки (k последних наблюдений) (рисунок 3.5).

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

O12 f_x

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,948398506							
5	R-квадрат	0,899459726							
6	Нормированный R-квадр	0,886892192							
7	Стандартная ошибка	0,237828879							
8	Наблюдения	10							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*ачимость F*			
12	Регрессия	1	4,048189394	4,048189394	71,57010337	2,91E-05			
13	Остаток	8	0,452500606	0,056562576					
14	Итого	9	4,50069						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
17	Y-пересечение	9517,329939	1116,712499	8,522632236	2,76137E-05	6942,186	12092,474	6942,186	12092,47
18	Дата	-0,221515152	0,026184098	-8,459911546	2,91392E-05	-0,2819	-0,161135	-0,2819	-0,16113
19									
20	ВЫВОД ОСТАТКА								
21									
22	*Наблюдение*	*казанное Курс евро*	*Остатки*	e_i^2					
23	1	71,03781818	-0,107818182	0,01162476					
24	2	70,81630303	0,11369697	0,012927001	$S_1 = \sum_{i=1}^{k} e_i^2 =$		0,4525006		
25	3	70,59478788	0,335212121	0,112367166					
26	4	70,37327273	-0,133272727	0,01776162					
27	5	70,15175758	-0,391757576	0,153473998					
28	6	69,93024242	0,149757576	0,022427331					
29	7	69,70872727	0,141272727	0,019957983					
30	8	69,48721212	-0,257212121	0,066158075					
31	9	69,26569697	-0,03569697	0,001274274					
32	10	69,04418182	0,185818182	0,034528397					
33	итого			0,452500606					
34									

Рисунок 3.4 – Результаты регрессионного анализа для первой подвыборки

Если предположение о пропорциональности дисперсий отклонений значениям X верно, то дисперсия регрессии по первой подвыборке (сумма квадратов отклонений $S_1 = \sum_{i=1}^{k} e_i^2$) будет существенно меньше дисперсии регрессии по третьей подвыборке (суммы квадратов отклонений $S_3 = \sum_{i=n-k+1}^{n} e_i^2$).

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,690250768
R-квадрат	0,476446123
Нормированный R-квадрат	0,411001888
Стандартная ошибка	0,350301602
Наблюдения	10

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	1	0,893360303	0,893360303	7,280184814	0,02715
Остаток	8	0,981689697	0,122711212		
Итого	9	1,87505			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	-4372,00903	1645,631923	-2,65673567	0,028951693	-8166,84	-577,17501	-8166,84	-577,175
Переменная X 1	0,104060606	0,038566937	2,698181761	0,027150375	0,01513	0,192996122	0,01513	0,193

ВЫВОД ОСТАТКА

Наблюдение	Предсказанное Y	Остатки	e_i^2
1	67,73672727	0,293272727	0,086008893
2	67,84078788	0,189212121	0,035801227
3	67,94484848	0,085151515	0,007250781
4	68,04890909	-0,318909091	0,101703008
5	68,1529697	-0,652969697	0,426369425
6	68,2570303	-0,327030303	0,106948819
7	68,36109091	0,398909091	0,159128463
8	68,46515152	0,214848485	0,046159871
9	68,56921212	0,110787879	0,012273954
10	68,67327273	0,006727273	4,52562E-05
итого			0,981689697

$$S_3 = \sum_{i=n-k+1}^{n} e_i^2 = 0{,}981689697$$

Рисунок 3.5– Результаты регрессионного анализа для третьей подвыборки

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 * t_i \tag{3.8}$$

Случайные отклонения e_i рассчитываются в столбце С (рисунки 3.4 и 3.4), графа «Остатки». Поэтому чтобы рассчитать S_1 и S_3 достаточно возвести в квадрат остатки и просуммировать получившиеся квадраты отклонений (рисунки 3.4 и 3.4, ячейка D33).

4 Для сравнения соответствующих дисперсий строится следующая *F*-статистика:

$$F = \frac{S_3/(k-m-1)}{S_1/(k-m-1)} = \frac{S_3}{S_1}, \qquad (3.9)$$

где *(k-m-1)* – число степеней свободы соответствующих выборочных дисперсий (*m* – количество объясняющих переменных в уравнении регрессии).

Таким образом, для конкретного примера $F = \frac{0{,}4525}{0{,}9817} = 2{,}169$.

5 Если $F_{набл.} = \frac{S_3}{S_1} > F_{крит.} = F_{\alpha;\nu_1;\nu_2}$, то гипотеза об отсутствии гетероскедастичности отклоняется.

Критическое значение критерия Фишера ($F_{крит.}$) определяется с помощью статистической функции **FРАСПОБР** (рисунок 3.6).

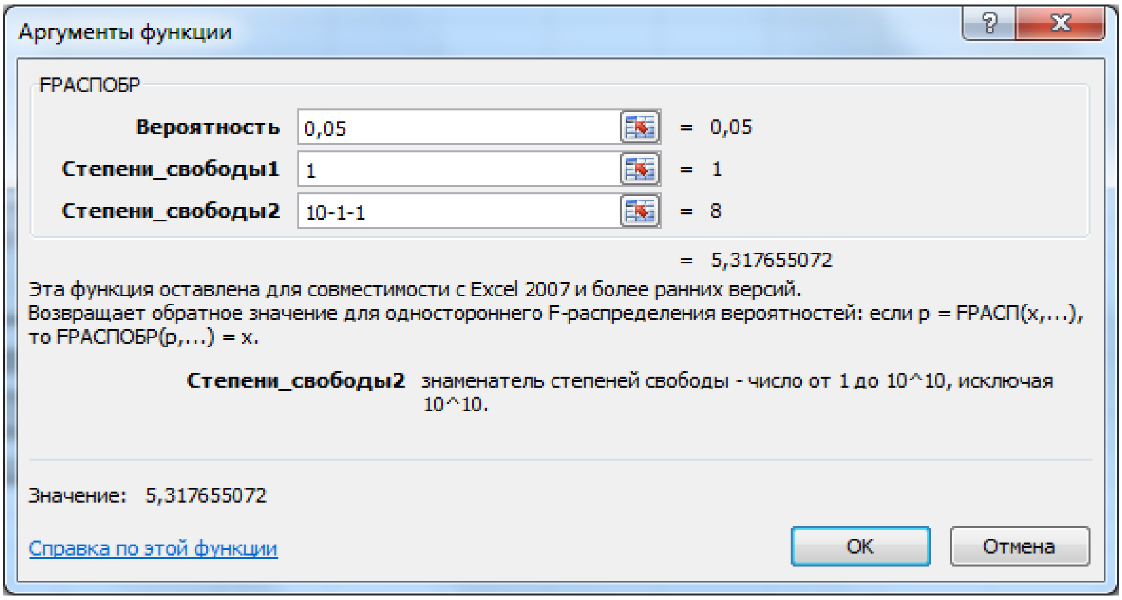

Рисунок 3.6 – Результат вычисления функции **FРАСПОБР**

$F_{набл.} = 2{,}169 < F_{крит.} = 5{,}18$, следовательно, гипотеза об отсутствии гетероскедастичности подтверждается. Таким образом, второе условие Гаусса-Маркова выполняется.

Проверка выполнения 3-его условия. Для проверки соответствия остатков третьему условию должна быть рассчитана статистика Дарбина-Уотсона, показывающая наличие автокорреляции в данной выборке:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}, \tag{3.10}$$

где DW – статистика Дарбина-Уотсона;

e_t – случайные отклонения.

По таблице критических точек Дарбина-Уотсона (приложение Д) определяются два числа d_1 и d_u и осуществляют выводы по правилу:

$0 \le DW < d_1$ – существует положительная автокорреляция;

$d_1 \le DW < d_u$ – вывод о наличии автокорреляции не определен;

$d_u \le DW < 4 - d_u$ – автокорреляция отсутствует;

$4 - d_u \le DW < 4 - d_1$ – вывод о наличии автокорреляции не определен;

$4 - d_1 \le DW \le 4$ – существует отрицательная автокорреляция.

Для расчета критерия Дарбина-Уотсона необходимо произвести дополнительные вычисления (таблица 3.2).

Таблица 3.2 – Расчет критерия Дарбина-Уотсона

Наблюдение	*Предсказанное Курс евро, руб.*	*Остатки*	*(e_t-e_{t-1})2*	*e^2*
1	70,43860887	0,491391129		0,24146524
2	70,35373387	0,576266129	0,007203766	0,33208265
3	70,26885887	0,661141129	0,007203766	0,43710759
4	70,18398387	0,056016129	0,366176266	0,00313781
5	70,09910887	-0,339108871	0,156123766	0,11499483
6	70,01423387	0,065766129	0,163923766	0,00432518
7	69,92935887	-0,079358871	0,021061266	0,00629783
8	69,84448387	-0,614483871	0,286358766	0,37759043
9	69,75960887	-0,529608871	0,007203766	0,28048556
10	69,67473387	-0,444733871	0,007203766	0,19778822
11	69,58985887	0,140141129	0,342078766	0,01963954
12	69,50498387	-0,384983871	0,275756266	0,14821258

13	69,42010887	-0,300108871	0,007203766	0,09006533
14	69,33523387	0,394766129	0,482851266	0,1558403
15	69,25035887	0,119641129	0,075693766	0,014314
16	69,16548387	0,204516129	0,007203766	0,04182685
17	69,08060887	0,289391129	0,007203766	0,08374723
18	68,99573387	0,364266129	0,005606266	0,13268981
19	68,91085887	0,349141129	0,000228766	0,12189953
20	68,82598387	-0,055983871	0,164126266	0,00313419
21	68,74110887	-0,271108871	0,046278766	0,07350002
22	68,65623387	-0,626233871	0,126113766	0,39216886
23	68,57135887	-0,541358871	0,007203766	0,29306943
24	68,48648387	-0,456483871	0,007203766	0,20837752
25	68,40160887	-0,671608871	0,046278766	0,45105848
26	68,31673387	-0,816733871	0,021061266	0,66705422
27	68,23185887	-0,301858871	0,265096266	0,09111878
28	68,14698387	0,613016129	0,836996266	0,37578877
29	68,06210887	0,617891129	2,37656E-05	0,38178945
30	67,97723387	0,702766129	0,007203766	0,49388023
31	67,89235887	0,787641129	0,007203766	0,62037855
Итого			3,761075469	6,85482899

На основе данных таблицы 3.2 вычислим критерий Дарбина-Уотсона:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{3{,}761}{6{,}855} = 0{,}549.$$

По таблице критических точек Дарбина-Уотсона (приложение Г) определяем d_1=1,35 и d_u=1,49. Так как $0 \le 0{,}549 < 1{,}49$, следовательно существует положительная автокорреляция. Таким образом, третье условие Гаусса-Маркова не выполняется.

Проверка выполнения 4-ого условия. Проверка гипотезы о наличии связи между случайной составляющей и переменными должна быть осуществлена посредством расчета показателя ковариации. Данный показатель рассчитывает с помощью функции **КОВАР** (категория «Статистические»). В качестве массивов 1 и 2 вводятся значения независимой переменной x (графа «Дата») и случайных отклонений e_i (графа «Остатки») (рисунок 3.7).

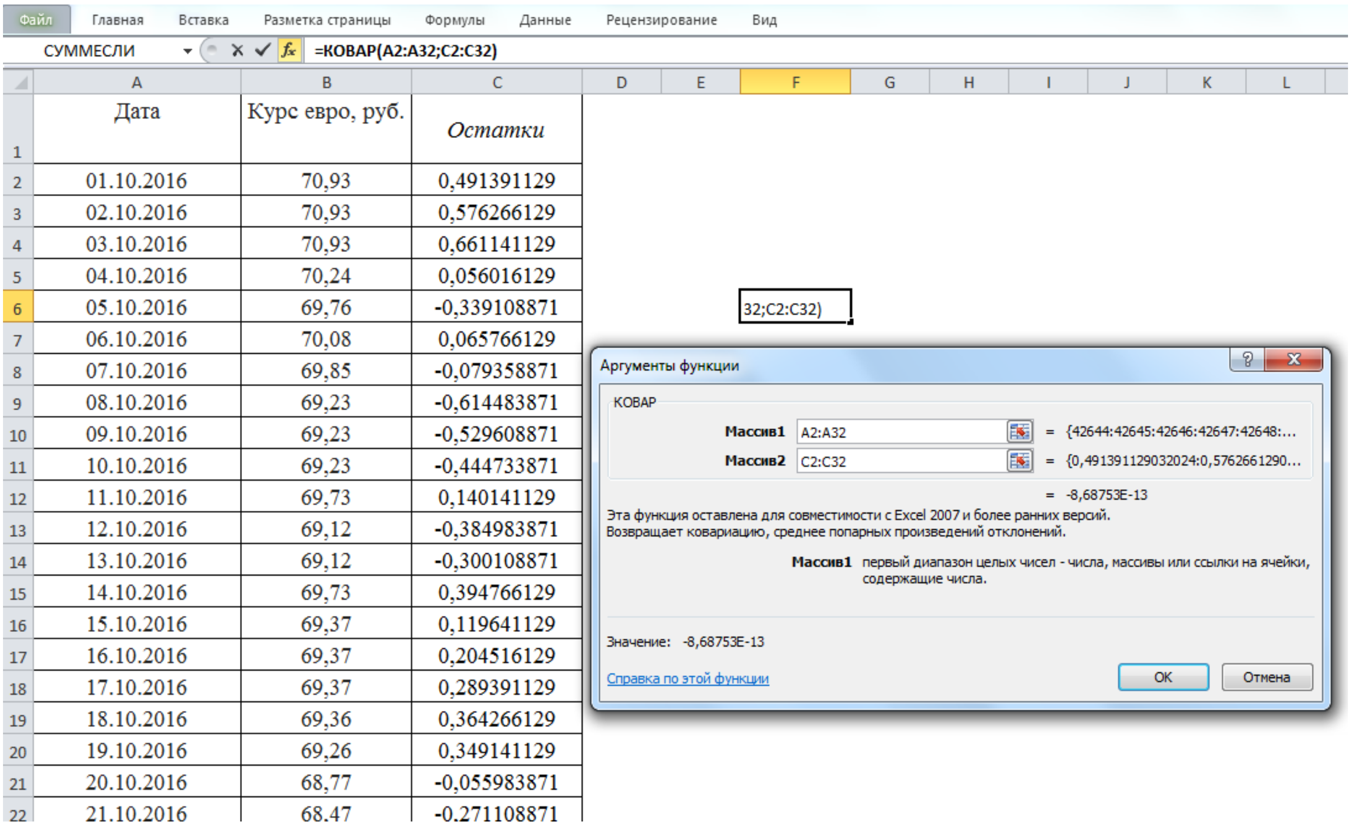

Рисунок 3.7 – Использование функции **КОВАР**

Показатель ковариации составил -0,000000000000868, что является чрезвычайно малым значением, следовательно, можно сделать вывод об отсутствии взаимосвязи между случайными отклонениями e_i и значениями независимой переменной x. Таким образом, четвертое условие Гаусса-Маркова выполняется.

Проверка выполнения 5-ого условия. Наряду с условиями Гаусса-Маркова во множественной регрессии предполагается, что остатки распределены нормально (т. е. подчиняются закону нормального распределения). Дело в том, что если остатки нормально распределены, то также будут распределены и коэффициенты регрессии. Хотя большинство тестов довольно робастны (устойчивы) по отношению к отклонениям от этого предположения, всегда, прежде чем сделать окончательные выводы, стоит рассмотреть распределения представляющих интерес переменных.

Проверка гипотезы о нормальном распределении остатков e_i будет проводиться на основе критерия Пирсона χ^2. Для этого сравниваются эмпирические (полученные по данным выборки) частоты f_i и теоретические (вычисленные в предположении нормального распределения) частоты $f_i^{'}$. Критерий Пирсона построен так, что: если эмпирические и теоретические частоты f_i и $f_i^{'}$ различаются незначимо, то с гипотезой о нормальном распределении генеральной совокупности соглашаются; если эмпирические и теоретические частоты различаются значимо, то с гипотезой о нормальном распределении не соглашаются, т. е. ее отвергают.

Для получения эмпирических и теоретических необходимо провести группировку исходных данных (остатков). Весь процесс построения группировки можно разбить на ряд этапов:

1 Определяется количество групп, на которые, надо разбить совокупность. Число групп можно подсчитать математическим путем с помощью **формулы Стерджесса:**

$$n = 1 + 3{,}322 \cdot logN, \tag{3.11}$$

где n – число групп;

N – число единиц совокупности.

Для данного примера: $n = 1 + 3{,}322 \cdot log(31) = 5{,}96 \approx 6$ групп.

2 Определяется интервал группировки. **Интервал** – это значение варьирующего признака, лежащее в определенных границах. Каждый интервал имеет свою величину, **нижнюю** (наименьшее значение признака в интервале) и **верхнюю** (наибольшее) границы или хотя бы одну из них. Если вариация признака проявляется в узких границах и распределение носит равномерный характер, то строят группировку с равными интервалами. **Величина равного интервала** определяется по формуле:

$$i = \frac{(x_{max} - x_{min})}{n}, \qquad (3.12)$$

где i – ширина интервала;

x_{max} – максимальное значение признака в совокупности;

x_{min} – минимальное значение признака в совокупности;

n – число групп.

Максимальное и минимальное значения признака в совокупности можно выявить с помощью встроенных статистических функций **МАКС** и **МИН**. Таким образом, ширина интервалов равна:

$$i = \frac{(-0{,}816733871 - 0{,}787641129)}{6} = 0{,}267.$$

3 После определения ширины интервалов проводится группировка. Схема проведения группировки представлена на рисунке 3.8.

Группы	Интервалы		Число единиц в группе
1 группа	X_{min}	$Xmin+i$	f_1
2 группа	$X_{min}+i$	$Xmin+i+i$	f_2
3 группа	$X_{min}+i+i$	$Xmax$	f_3

Рисунок 3.8 – Методика проведения группировки

Группировка остатков (e_i) приведена на рисунке 3.9.

	Интервалы		Частота, f
7			
8	Интервалы		Частота, f
9	-0,816733871	-0,549338	5
10	-0,549338038	-0,281942	7
11	-0,549338038	-0,281942	3
12	-0,549338038	-0,281942	5
13	-1,098676075	-0,83128	5
14	-1,38061828	-1,113222	6
15			31

Рисунок 3.9 – Группировка остатков

4 Далее необходимо рассчитать теоретические частоты нормального распределения. Все вспомогательные расчеты приведены на рисунке 3.10. Методика расчета представлена ниже.

4.1 Для сгруппированных данных находят среднюю арифметическую:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} \qquad (3.13)$$

	C	D	E	F	G	H	I	J	K	L
1	MIN	-0,8167339								
2	MAX	0,78764113								
3	Средняя арифметическая	-0,7082643								
4	Среднее квадратическое отклонение (σ)	0,32912534								
5	Количество групп	5,95430355								
6	Ширина интервала	0,26739583								
7										
8	Интервалы		Частота, f	Середина интервала, x	$x_i * f_i$	$(x_i - x_{ср.})^2 * f$	$(x_i - x_{ср.})/\sigma$	$\varphi(t)$	f'	$(f - f')^2/f'$
9	-0,816733871	-0,549338	5	-0,683035954	-3,415	0,003182357	0,076652823	0,1588	3,9987	0,25072848
10	-0,549338038	-0,281942	7	-0,415640121	-2,909	0,599402537	0,889096599	0,1072	2,7011	6,84180837
11	-0,549338038	-0,281942	3	-0,415640121	-1,247	0,256886802	0,889096599	0,1072	2,7011	0,03307386
12	-0,549338038	-0,281942	5	-0,415640121	-2,078	0,428144669	0,889096599	0,1072	2,7011	1,95656766
13	-1,098676075	-0,83128	5	-0,964978159	-4,825	0,329509922	-0,77998801	0,1175	2,9586	1,4085325
14	-1,38061828	-1,113222	6	-1,246920363	-7,482	1,740901863	-1,636628842	0,0417	1,0509	23,3077586
15			31		-21,96	3,35802815				33,7984694

Рисунок 3.10 – Расчет частот нормального распределения (f')

В качестве вариант (x_i) принимают среднее арифметическое границ интервалов (рисунок 3.10, столбец F).

Таким образом, средняя арифметическая для данного примера равна: $\bar{x} = -21{,}96/31 = -0{,}708$ (рисунок 3.10, ячейка D3).

4.2 Следующий этап – вычисление среднеквадратического отклонения:

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}} = \sqrt{\frac{3{,}358}{31}} = 0{,}329. \qquad (3.14)$$

4.3 Далее рассчитывают нормированное отклонение от средней (рисунок 3.10, столбец I):

$$t = \frac{x_i - \bar{x}}{\sigma}. \qquad (3.15)$$

4.4 На основе нормированного отклонения определяют величину стандартной плотности нормального распределения *φ(t)* (рисунок 3.10, столбец J):

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot exp^{-\frac{t^2}{2}}. \qquad (3.16)$$

4.5 Расчет теоретических частот (рисунок 3.10, столбец K) производят по следующей формуле:

$$f' = \frac{N \cdot i}{\sigma} \cdot \varphi(t), \qquad (3.17)$$

где f' – искомая теоретическая частота;

N – число наблюдений, или сумма всех частот вариационного ряда;

i – величина интервала в группах (классах);

φ(t) – величина стандартной плотности нормального распределения.

4.6 Далее на основе теоретических частот рассчитывается критерий Пирсона (χ^2). Критерий Пирсона представляет собой сумму отношений квадратов расхождений между эмпирических и теоретических частот к теоретическим частотам (рисунок 3.10, ячейка L15):

$$\chi^2 = \sum \frac{(f_i - f_i')^2}{f_i'} = 33{,}798. \tag{3.18}$$

Фактическое значение χ^2 необходимо сравнить с критическим, определяемым по специальным таблицам в зависимости от принимаемого уровня значимости и числа степеней свободы или с помощью встроенной функции **ХИ2.ОБР** (рисунок 3.11).

Уровень значимости, α – вероятность допуска ошибки в утверждении гипотетического закона (характера) распределения – принимается равным 5 % (α = 0,05).

Число степеней свободы *k*, рассчитывается как число групп *n* в ряду распределения минус единица и минус число параметров эмпирического распределения, использованных для нахождения теоретических частот. Поскольку при расчете теоретических частот используется два параметра эмпирического распределения среднеквадратическое отклонение и среднюю арифметическую, то число степеней свободы равно:

$$k = n - 1 - 2, \tag{3.19}$$

где *k* – число степеней свободы;

n – число групп в ряду распределения.

Если фактическое χ^2 оказывается меньше табличного (критического), то расхождения между эмпирическими и теоретическими частотами можно считать случайными.

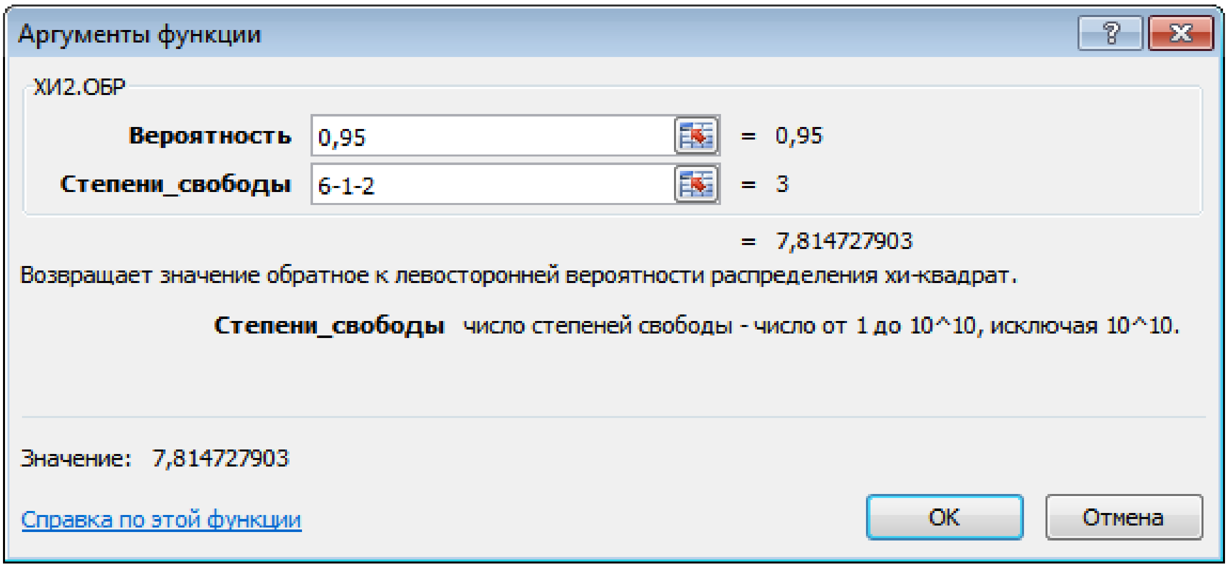

Рисунок 3.11 – Результат вычисления функции **ХИ2.ОБР**

Так как $\chi^2_{\text{расч.}} = 33{,}798 > \chi^2_{\text{крит.}} = 7{,}814$, то расхождения между эмпирическими и теоретическими частотами нельзя считать случайными, а распределение остатков – нормальным. Следовательно, пятое условие Гаусса-Маркова не выполняется.

5 На основе теоремы Гаусса-Маркова необходимо сделать вывод о свойствах полученных оценок на основе предпосылок МНК. В нашем случае, выполняются не все условия Гаусса-Маркова, следовательно, оценки, полученные по МНК, не обладают свойствами несмещенности, эффективности и состоятельности. Другими словами, полученные оценки параметров не являются наилучшими из всех возможных оценок.

Контрольные вопросы для защиты лабораторной работы

1 В чем заключается суть метода наименьших квадратов (МНК)?

2 Перечислите предпосылки МНК.

3 Каковы последствия их выполнимости либо невыполнимости?

4 В чем суть наилучших несмещенных оценок?

5 В чем заключается суть эффективных и состоятельных оценок?

6 Что такое автокорреляция?