КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ

Кафедра экономической теории и эконометрики

И. И. ИСМАГИЛОВ, Е.И. КАДОЧНИКОВА

РЕШЕНИЕ ЭКОНОМЕТРИЧЕСКИХ ЗАДАЧ В СРЕДЕ MS EXCEL

Учебное пособие для студентов,
обучающихся по направлению 38.03.01 «Экономика»

Казань 2019

УДК 330.43

ББК Ув631я73-1

Рекомендовано к публикации на заседании кафедры

экономической теории и эконометрики

Протокол №6 от 29 апреля 2019 года

Рецензенты:

доктор экономических наук,

заведующий кафедрой бизнес-статистики и

математических методов в экономике

Казанского национального исследовательского технологического университета

А. В. Аксянова

кандидат физико-математических наук,

доцент кафедры экономической теории и эконометрики ИУЭиФ КФУ

Л. К. Астафьева

Исмагилов И.И., Кадочникова Е.И. Решение эконометрических задач в среде MS Excel: учебное пособие для студентов, обучающихся по направлению 38.03.01 «Экономика» / И.И. Исмагилов, Е.И. Кадочникова – Казань: Казан. ун-т, 2019. – 80 с.

Данное учебное пособие предназначено для использования на практических занятиях по дисциплине «Эконометрика» для бакалавриата направления 38.04.01 «Экономика» и других направлений. Цель учебно-методического пособия – развить практические умения и навыки построения эконометрических моделей средствами MS Excel.

Содержание

Введение

Особенностью деятельности экономиста является работа в условиях недостатка информации и неполноты исходных данных. Анализ такой информации возможен с помощью эконометрических методов. В соответствии с Государственным образовательным стандартом высшего образования по направлению «Экономика» дисциплина «Эконометрика» входит в учебные планы подготовки бакалавров экономики в качестве общепрофессиональной дисциплины и является базовой дисциплиной современного экономического образования. Дисциплина «Эконометрика» опирается на курсы «Микроэкономика», «Макроэкономика», «Статистика», «Линейная алгебра», «Теория вероятностей и математическая статистика». Использование эконометрических методов позволяет осуществить проверку положений экономической теории.

Последовательность изложения материала в учебно-методическом пособии основана на наиболее традиционном понимании эконометрики как науки о связях экономических явлений. Большое место отводится регрессионному анализу: линейной модели парной регрессии, линейной модели множественной регрессии, проверке соблюдения предпосылок метода наименьших квадратов. Каждый из пяти разделов содержит решения теоретических задач, позволяющих лучше усвоить соответствующую тему, технологии проведения эконометрического моделирования средствами MS Excel. Статистический пакет Excel, включенный в MS Office, имеющийся в наличии фактически на каждом персональном компьютере, позволяет оценивать линейные уравнения регрессии и проводить необходимые эконометрические тесты. В приложении приведены некоторые статистические таблицы.

Материалы учебно-методического пособия могут быть использованы в курсе эконометрики в бакалавриате, а также могут быть самостоятельно использованы студентами для освоения в будущем магистерских курсов, требующих знания эконометрического инструментария, а также для выполнения эмпирической части выпускной квалификационной работы. Студентам также рекомендуется с целью более углубленного изучения рассматриваемых тем обратиться к учебникам, на которые авторы опирались при разработке данного учебно-методического пособия:

1. Демидова О. А. Эконометрика: учебник и практикум для прикладного бакалавриата / О. А. Демидова, Д. И. Малахов. – М.: Издательство Юрайт, 2018. – 334 с.
2. Доугерти К. Введение в эконометрику: Учебник. 3-е изд. / Пер. с англ. — М.: ИНФРА-М, 2009.
3. Эконометрика. Начальный курс. Магнус Я.Р., Катышев П.К., Пересецкий А.А. 6-е изд., перераб. и доп. - М.: Дело, 2004. — 576 с.

Раздел 1. Парная линейная регрессия и корреляция

В результате изучения данного раздела студент должен:

знать:

определение теоретической и выборочной парной регрессии;

суть МНК и его предпосылки для случая парной регрессии;

основные положения дисперсионного анализа зависимой переменной для случая парной регрессии (TSS=ESS+RSS);

определение и интерпретацию коэффициента детерминации для парной регрессии;

уметь:

находить МНК-оценки коэффициентов парной регрессии;

определять качество подгонки парной регрессии с помощью коэффициента детерминации;

проверять гипотезу о значимости коэффициента регрессии;

строить доверительный интервал для коэффициента регрессии;

строить точечные и интервальные прогнозы;

владеть:

навыками оценки и проверки качества парной регрессии в MS Excel.

Задачи с решениями

Задача 1.1. По совокупности 18 предприятий торговли изучается зависимость между ценой x на некоторый товар и прибылью y торгового предприятия. При оценке регрессионной модели были получены следующие результаты:

$$\sum (y-\hat{y})^2 = 23; \sum (y-\bar{y})^2 = 35.$$

Проверить статистическую значимость уравнения регрессии. Построить таблицу дисперсионного анализа.

Решение:

В условиях задачи n=18, остаточная сумма квадратов отклонений равна 23, а общая сумма квадратов отклонений составляет 35. Для расчета индекса корреляции воспользуемся выражением:

$$R = \sqrt{1-\frac{\sum (y-\hat{y})^2}{\sum (y-\bar{y})^2}} = \sqrt{1-\frac{23}{35}} = 0,586$$

$$R^2 = 0,343$$

Фактическое значение F-критерия рассчитаем с помощью выражения:

$$F = \frac{R^2}{1-R^2}\cdot(n-2) = \frac{0,343}{1-0,343}\cdot(18-2) = 8,35.$$

При проверке статистической значимости уравнения в целом воспользуемся F-статистикой и сравним ее с критическим значением для уровня значимости α=0,05. Табличное (критическое) значение при этом равно: $F_{0,05;1;18-2} = 4,49$.

Вывод: поскольку наблюдаемое значение F-статистики, равное 8,35, больше критического, нулевая гипотеза о статистической незначимости уравнения регрессии может быть отклонена, на уровне α=0,05 уравнение регрессии является статистически значимым в целом (адекватным), статистическая взаимосвязь между *y* и *x* подтверждается.

Однако, для α=0,01 $F_{0,01;1;16}$ = 8,53.

Вывод: поскольку наблюдаемое значение F-статистики, равное 8,35, меньше критического, то в этом случае нулевую гипотезу отклонить нельзя, на уровне α=0,01 уравнение не адекватно, статистическая взаимосвязь между *y* и *x* не подтверждается.

Замечание: Проверка статистической значимости уравнения парной регрессии эквивалентна проверке значимости коэффициента регрессии. Поэтому можно было провести проверку с использованием t-статистики, используя соотношение $F=t_b^2$.

Для построения таблицы дисперсионного анализа определим из балансового уравнения величину регрессионной суммы квадратов отклонений:

$$\sum(\hat{y}-\bar{y})^2 = \sum(y-\bar{y})^2 - \sum(y-\hat{y})^2 = 35-23=12.$$

Поскольку мы имеем дело с парной регрессионной зависимостью, число степеней свободы регрессионной суммы квадратов отклонений принимаем равным единице. С учетом этих условий таблица дисперсионного анализа выглядит следующим образом:

Источники вариации	Число степеней свободы	Сумма квадратов отклонений	Дисперсия на 1 степень свободы	F
Регрессионная	1	12	12	8,35
Остаточная	16	23	1,4375	
Общая	17	35		-

Задача 1.2. По 10 парам наблюдений получены следующие результаты:

$$\sum x_i = 100; \sum y_i = 200; \sum x_i y_i = 21000; \sum x^2{}_i = 12000; \sum y^2{}_i = 45000.$$

По МНК оцените коэффициенты уравнений регрессии *Y* на *X* и *X* на *Y*. Оцените коэффициент корреляции r_{yx} и коэффициент детерминации R^2.

Решение:

Для уравнения регрессии Y=a+b*x+e :

$$b = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{21000/10 - 10\cdot 20}{1200-10^2} = \frac{1900}{1100} = 1{,}73$$

$$a = \bar{y} - b\cdot\bar{x} = 20 - 1{,}73\cdot 10 = 2{,}7$$

Уравнение регрессии: Y=2,7+1,73*x+e. С увеличением X на 1 единицу собственного измерения Y возрастает в среднем на 1,73 единиц.

Для уравнения регрессии X=a+b*y+e:

$$b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{y^2} - \overline{y}^2} = \frac{21000/10 - 10 \cdot 20}{4500 - 20^2} = \frac{1900}{4100} = 0{,}46$$

$$a = \overline{x} - b \cdot \overline{y} = 10 - 0{,}46 \cdot 20 = 0{,}8$$

Уравнение регрессии: X=0,8+0,46*y+e. С увеличением Y на 1 единицу собственного измерения X возрастает в среднем на 0,46 единиц.

Отметим, что две оцененные регрессии различны.

Тесноту линейной связи измеряет линейный коэффициент парной корреляции:

$$r_{yx} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{1900}{\sqrt{1100 \cdot 4100}} = 0{,}895$$

$$\sigma_x = \overline{x^2} - \overline{x}^2; \sigma_y = \overline{y^2} - \overline{y}^2.$$

Величина линейного коэффициента парной корреляции показывает тесную линейную статистическую взаимосвязь рассматриваемых признаков.

Коэффициент детерминации обоих регрессий:

$R^2 = r_{yx}{}^2 = 0{,}801$, показывает, что 80,1% вариации признаков обусловлено их взаимным влиянием, оставшиеся 19,09% его величины – это доля вариации признаков за счет объясняющих переменных, не включенных в модель, и случайных ошибок в наблюдениях, если они имеются.

Задача 1.3. Зависимость объема продаж (*Y*) от расходов на рекламу (*X*) характеризуется по 12 предприятиям концерна следующим образом:

$$y = 10{,}6 + 0{,}6 \cdot x$$
$$\sigma_x = 4{,}7$$
$$\sigma_y = 3{,}4$$

Определите линейный коэффициент парной корреляции, регрессионную сумму квадратов отклонений, постройте таблицу дисперсионного анализа для оценки значимости уравнения в целом, определите F-статистику, t-статистику и доверительный интервал коэффициента регрессии.

Решение:

Для определения коэффициента корреляции применим формулу:

$$r_{yx} = b \cdot \frac{\sigma_x}{\sigma_y} = 0{,}6 \cdot \frac{4{,}7}{3{,}4} = 0{,}829$$

Значение коэффициента корреляции свидетельствует о тесной линейной взаимосвязи между объемом продаж и расходами на рекламу.

Коэффициент детерминации составит: $R^2 = r_{yx}{}^2 = 0{,}829^2 = 0{,}687$

Определим регрессионную сумму квадратов отклонений:

$$\sum(\hat{y}_x-\bar{y})^2=R^2\cdot\sum(y-\bar{y})^2=R^2\cdot\sigma^2{}_y\cdot(n-1)=0{,}687\cdot3{,}4^2\cdot(12-1)=87{,}359$$

Составим таблицу дисперсионного анализа и определим F-статистику Фишера.

Дисперсионный анализ результатов регрессии

Источники вариации	Число степеней свободы	Сумма квадратов отклонений	Дисперсия на одну степень свободы	F-статистика	
				факт.	табл., α=0,05
Регрессионная (объясненная)	1	87,359	87,359	21,949	4,96
Остаточная	10	39,801	3,9801		
Общая	11	127,16	11,56		

Поскольку $F_{факт}>F_{табл}$, то признается статистическая значимость уравнения регрессии на уровне значимости α=0,05.

Связь между F-статистикой Фишера, t-статистикой Стьюдента для коэффициента регрессии, t-статистикой Стьюдента для коэффициента корреляции выражается равенством:

$$t_r{}^2=t_b{}^2=F=21{,}949$$

Значит, $t_b=\sqrt{21{,}949}=4{,}685$. Знак статистики Стьюдента для коэффициента регрессии и знак собственно коэффициента регрессии совпадают. Табличное значение t-статистики для α=0,05, ν=10 составляет 2,2281. Поскольку $t_{факт}>t_{табл}$, то коэффициент регрессии *b* статистически значимо отличен от нуля.

Определим доверительный интервал для коэффициента регрессии b с надежностью 95%: $b\pm t\cdot m_b$

Стандартная ошибка коэффициента регрессии b определяется по формуле:

$$m_b=\sqrt{\frac{\sum(y-\hat{y}_x)^2/(n-2)}{\sum(x-\bar{x})^2}}=\frac{b}{t_b}=\frac{0{,}6}{4{,}685}=0{,}128$$

Для расчета доверительного интервала определим предельную ошибку:

$$\Delta=t\cdot m_b=2{,}2281\cdot0{,}128=0{,}285$$

Границы доверительного интервала, в котором с надежностью 95% лежит истинное значение коэффициента регрессии, составят:

$$b-\Delta\le\beta\le b+\Delta$$
$$0{,}315\le\beta\le0{,}885.$$

В границы интервала ноль не попадает, следовательно, оцениваемый параметр статистически значим на уровне α=0,05 и сформировался под влиянием систематически действующего фактора x.

Задача 1.4. Пусть имеется уравнение парной регрессии y = 5 - 6x + ε, построенное по 15 наблюдениям. При этом r=-0,7. Определить доверительный интервал, в который с надежностью 0,99 попадает коэффициент регрессии.

Решение:

Для построения доверительного интервала необходимо знать стандартную ошибку m_b коэффициента регрессии. Однако она не задана, и нужно определить ее косвенным путем. Для этого воспользуемся тем, что в парной регрессии существует связь между t- и F-статистиками: $t_b = \sqrt{F}$.

F-статистику определим из формулы : $F = \frac{(-0,7)^2}{1-(-0,7)^2} \cdot \frac{15-1-1}{1} = 12,5;$

$$t_b = \sqrt{12,5} = 3,53$$

Поскольку знак оцененного коэффициента b отрицательный, то указываем, что t_b= - 3,53.

$$m_b = \frac{b}{t_b} = \frac{-6}{-3,53} \approx 1,7\,;$$

Доверительный интервал с надежностью 0,99 имеет вид ($t_{табл}$(0,01;13)=3,01):

$$-6 - 1,7 \cdot 3,01 \le \beta \le -6 + 1,7 \cdot 3,01$$
$$-11,11 \le \beta \le -0,89$$

Задача 1.5. Уравнение регрессии потребления материалов от объема производства, построенное по 15 наблюдениям, имеет вид:Y = 5 + 5x + ε, t_b=4,0. Определить коэффициент детерминации для этого уравнения.

Решение:

Зная t-критерий для коэффициента регрессии, вычислим F-критерий для данного уравнения:

$F = t_b^2 = 4^2 = 16$. Далее воспользуемся уравнением, из которого определим коэффициент детерминации при n=15:

$$r^2 = \frac{F}{n-2+F} = \frac{16}{15-2+16} = 0,552\cdot$$

Упражнения с пояснениями

Имеется выборка по 16 филиалам торговой компании за 2017 год.

Объем продаж, тыс. руб.	126	137	148	191	274	370	432	445	367	367	321	307	331	345	364	384
Расходы на рекламу, тыс. руб.	4,0	4,8	3,8	8,7	8,2	9,7	14,7	18,7	19,8	10,6	8,6	6,5	12,6	6,5	5,8	5,7

Упражнение 1.1. Используя исходные данные из таблицы построить корреляционное поле при помощи Мастера диаграмм MS Excel и провести его визуальный анализ.

Решение. Для построения корреляционного поля (или диаграммы рассеяния) используем Мастер диаграмм MS Excel. Для этого выделим «мышью» исходные наблюдения переменных X и Y, затем в Главном меню MS Excel выберем: Вставка – Диаграммы – Точечная – Точечная с маркерами. Полученная диаграмма рассеяния представлена на рисунке 1.

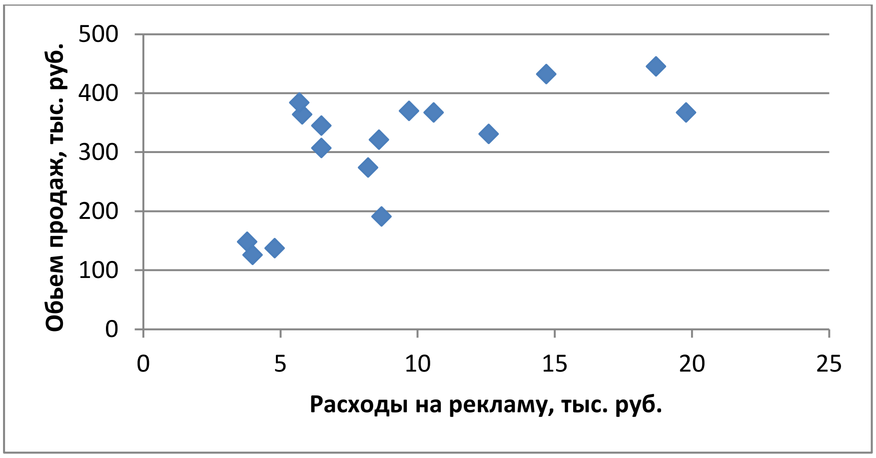


Рис. 1. Диаграмма рассеяния в MS Excel

Вывод 1: Как видно из рисунка 1, наблюдается прямая умеренная взаимосвязь между объемом продаж (Y) и расходами на рекламу (X). Точки на графике сосредоточены вокруг прямой линии, поэтому взаимосвязь между объемом продаж и расходами на рекламу уместно аппроксимировать прямой линией и применить линейную регрессию с одной объясняющей переменной.

Упражнение 1.2. Сформировать расчетную электронную таблицу заданной структуры, чтобы определить МНК-оценки коэффициентов линейной модели парной регрессии $Y = a + bx + \varepsilon$ «вручную» по соответствующим расчетным формулам.

Решение. Чтобы сформировать расчетную таблицу 1 (последовательно выполняя пункты 2-11 лабораторной работы), необходимо использовать Мастер формул MS Excel. В данном пункте достаточно заполнить столбцы *x, xy, xy, x^2, y^2*. Основы работы с пакетом MS Excel изучаются в курсах информатики, поэтому мы не будем подробно останавливаться на описании интерфейса данной программы, а несколько позже рассмотрим лишь надстройку *Анализ данных.*

Табл. 1

Подготовка данных для оценивания линейной регрессии объема продаж по расходам на рекламу

№ пп.	x	y	xy	x^2	y^2	$\hat{y}$	(y- $\hat{y}$)^2	(y-$y_{сред}$)^2	($\hat{y}$ -$y_{сред}$)^2	Abs(e/y)
1	4	126	504	16	15876	235,22	11928,89	32693,16	5125,56	0,87
2	4,8	137	657,6	23,04	18769	246,04	11889,44	28836,29	3693,45	0,80
3	3,8	148	562,4	14,44	21904	232,51	7142,72	25221,41	5520,17	0,57
4	8,7	191	1661,7	75,69	36481	298,78	11617,09	13412,54	64,48	0,56
5	8,2	274	2246,8	67,24	75076	292,02	324,74	1076,66	218,80	0,07
6	9,7	370	3589	94,09	136900	312,31	3328,52	3992,66	30,19	0,16
7	14,7	432	6350,4	216,09	186624	379,93	2711,60	15671,91	5345,73	0,12
8	18,7	445	8321,5	349,69	198025	434,02	120,49	19095,79	16182,58	0,02
9	19,8	367	7266,6	392,04	134689	448,90	6707,57	3622,54	20188,79	0,22
10	10,6	367	3890,2	112,36	134689	324,48	1808,09	3622,54	312,08	0,12
11	8,6	321	2760,6	73,96	103041	297,43	555,54	201,29	88,03	0,07

12	6,5	307	1995,5	42,25	94249	269,03	1441,75	0,04	1427,55	0,12
13	12,6	331	4170,6	158,76	109561	351,53	421,34	585,04	1999,34	0,06
14	6,5	345	2242,5	42,25	119025	269,03	5771,50	1458,29	1427,55	0,22
15	5,8	364	2111,2	33,64	132496	259,56	10907,13	3270,41	2232,54	0,29
16	5,7	384	2188,8	32,49	147456	258,21	15823,03	5957,91	2362,17	0,33
Сумма	148,7	4909					92499,43	158718,44	66219,00	4,60
Среднее	9,29	306,81	3157,46	109,00	104053,8		RSS	TSS	ESS	

Суммы и средние по столбцам в табл. 1 необходимо определить с помощью функций *СУММ(...) и СРЗНАЧ(...)*. Оценки коэффициентов модели $Y = a + bx + \varepsilon$ необходимо определить ниже табл. 1, используя ссылки на необходимые ячейки согласно формулам метода наименьших квадратов:

$$b = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} = \frac{3157,46 - 306,81 \cdot 9,29}{109 - 9,29 \cdot 9,29} = 13,52, \qquad a = \overline{y} - b\overline{x} = 306,81 - 13,52 \cdot 9,29 = 181,12.$$

Вывод 2: Запишем модель линейной регрессии: $Y = 181,12 + 13,52x + \varepsilon$. Экономическая интерпретация коэффициентов следующая. Увеличение расходов на рекламу (X) на 1 тыс. рублей приводит при прочих равных условиях к увеличению объема продаж (Y) в среднем на 13,52 тыс. рублей. Данный результат согласуется с экономической интуицией, так как при росте расходов на рекламу можно ожидать увеличения объема продаж. При отсутствии расходов на рекламу объем продаж может составить в среднем 181,12 тыс. руб.

Упражнение 1.3. «Вручную» найти значения выборочных дисперсий и средних квадратических отклонений для переменных *x* и *y*, а также выборочный коэффициент корреляции и средний коэффициент эластичности.

Решение. Значения дисперсий для *x* и *y*, а также средних квадратических отклонений *x, y* необходимо ниже табл. 1, используя ссылки на необходимые ячейки табл. 1 согласно формулам:

$$\sigma_x^2 = \overline{x^2} - \overline{x}^2 = 109 - 9,29 \cdot 9,29 = 22,63, \qquad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{22,63} = 4,76;$$

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2 = 104054 - 306,81 \cdot 306,81 = 9919,902, \qquad \sigma_y = \sqrt{\sigma_y^2} = \sqrt{9919,902} = 99,6.$$

Необходимо проверить результаты с помощью функций *ДИСПР(...)* и *СТАНДОТКЛОНП(...)*. Значение линейного коэффициента парной корреляции необходимо определить ниже табл. 1, используя ссылки на необходимые ячейк табл. 1 согласно формуле: $r = b\frac{\sigma_x}{\sigma_y} = 13,52 \cdot \frac{4,76}{99,6} = 0,65.$ Необходимо проверить результат с помощью функции *КОРРЕЛ(...)*. Значение среднего коэффициента эластичности необходимо определить ниже табл.1, используя ссылки на необходимые ячейки табл. 1 согласно формуле:

$$Э = b\frac{\overline{x}}{\overline{y}} = 13,52 \cdot \frac{9,29}{306,81} = 0,41.$$

Вывод 3: Согласно шкале Чеддока-Снедекора наблюдается умеренная прямая статистическая взаимосвязь между объемом продаж и расходами на рекламу. Увеличение расходов на рекламу на 1 % относительно среднего значения приводит в среднем к увеличению объема продаж на 0,41% относительно среднего.

Упражнение 1.4. Вычислить «вручную» предсказанные моделью значения $\hat{y}$ по построенному уравнению $\hat{y} = a + bx$. Вычислить остатки регрессии и их квадраты.

Решение. Чтобы вычислить предсказанные моделью значения $\hat{y}$ в табл. 1, необходимо в уравнение регрессии $\hat{y} = 181{,}12 + 13{,}52x$ вместо переменной *x*, последовательно, начиная с первого, подставить исходные наблюдаемые значения расходов на рекламу из табл 1. Полученные предсказанные моделью значения $\hat{y}$ указать в соответствующем столбце табл. 1. Остаток регрессии – это ошибка, несвязка (*discrepancy*) между наблюдаемым значением зависимой переменной y_i и предсказанными моделью значениями $\hat{y}$. Чтобы вычислить остатки регрессии и их квадраты в табл. 1 необходимо заполнить столбец: $(y-\hat{y})\wedge 2$ в табл. 1. Также необходимо заполнить в табл. 1 столбцы (y-y $_{\text{сред}}$)^2, $(\hat{y} - y_{\text{сред}})\wedge 2$.

Упражнение 1.5. Вычислить «вручную» регрессионную (ESS), остаточную (RSS) и общую (TSS) суммы квадратов отклонений, дисперсии на одну степень свободы и проверить балансовое соотношение для сумм квадратов отклонений.

Решение. Чтобы вычислить суммы квадратов отклонений (TSS, ESS, RSS), необходимо в табл. 1 в строке «Сумма» применить функцию *СУММ(...)* к стобцам (Y-Y_сред)^2, (Y_предск-Y_сред)^2, (Y-Yx)^2. Чтобы вычислить дисперсии на 1 степень свободы, надо использовать ссылки на необходимые ячейки согласно формулам:

$$D_{общ.} = \frac{\sum(y-\bar{y})^2}{n-1} = \frac{158718{,}438}{16-1} = 10581{,}23.$$

$$D_{\text{объясн..}} = \frac{\sum(\hat{y}_x-\bar{y})^2}{1} = \frac{66219{,}01}{1} = 66219{,}01.\ D_{ост.} = \frac{\sum(y-\hat{y}_x)^2}{n-2} = \frac{92499{,}4}{16-2} = 6607{,}1$$

Также необходимо ниже табл. 1 проверить балансовое соотношение для суммы квадратов отклонений:

$$\sum(y-\bar{y})^2 = \sum(\hat{y}_x-\bar{y})^2 + \sum(\hat{y}_x-\hat{y})^2.$$

Упражнение 1.6. Рассчитать «вручную» коэффициент детерминации и индекс корреляции с использованием сумм квадратов отклонений.

Решение. Значения коэффициента детерминации и индекса корреляции необходимо определить «вручную» ниже таблицы, используя ссылки на необходимые ячейки согласно формулам:

$$R^2 = 1-\frac{RSS}{TSS} = 1-\frac{92499{,}43}{158718{,}44} = 0{,}42;\ R^2 = \frac{ESS}{TSS} = \frac{66219{,}00}{158718{,}44} = 0{,}42;\ R = \sqrt{R^2} = \sqrt{0{,}42} = 0{,}65.$$

Правильность расчетов можно проверить с использованием коэффициента корреляции $R^2 = r^2$ и с помощью функции *КВПИРСОН(...).*

Вывод 4: Доля дисперсии объема продаж, объясненная с помощью расходов на рекламу, составляет 42%. Чем ближе R^2 к 1, тем лучше качество подгонки уравнения регрессии.

Упражнение 1.7. Используя тест Стьюдента, проверить статистическую значимость коэффициента регрессии β на уровнях 0,05 и 0,01.

Решение. Чтобы рассчитать стандартную ошибку коэффициента регрессии и значение статистики Стьюдента, необходимо использовать ссылки на необходимые ячейки согласно формулам:

$$m_b == \sqrt{\frac{D_{ост.}}{\sum (x-\bar{x})^2}} = \frac{S_e}{\sigma_x \sqrt{n}} = \frac{\sqrt{6607,1}}{4,76\cdot\sqrt{16}} = 4,27, \quad S_e = \sqrt{D_{ост.}}, \quad t_b = \frac{b}{m_b} = \frac{13,52}{4,27} = 3,17$$

Для нахождения критического значения статистики Стьюдента удобно использовать функцию *СТЬЮДРАСПОБР(...)*. В Главном меню MS Excel выберем: Формулы – Вставить функцию – Полный алфавитный перечень – СТЬЮДРАСПОБР(…).

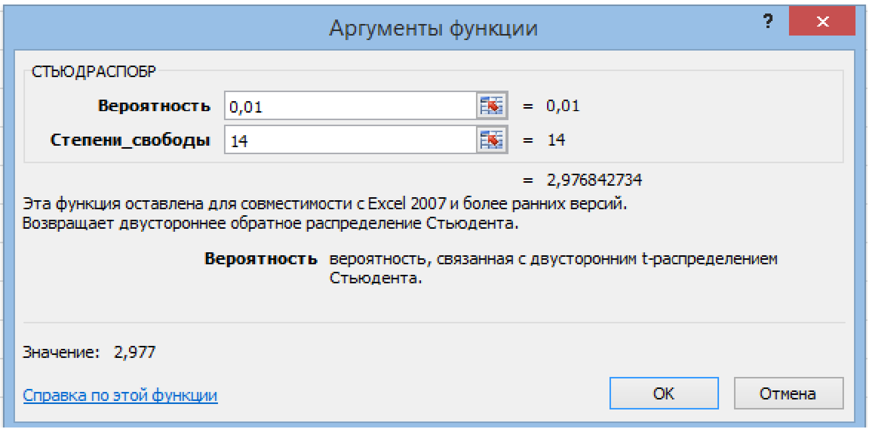


Рис. 2. Диалоговое окно функции СТЬЮДРАСПОБР(…).

Вывод 5: Поскольку 3,17 > 2,98, то гипотеза H_0: β = 0 отвергается, т.е. согласно тесту Стьюдента коэффициент регрессии β является значимым на уровне α=0,01. Это означает, что между переменными X (расходы на рекламу) и Y (объем продаж) существует значимая линейная связь.

Упражнение 1.8. Построить «вручную» доверительный интервал для коэффициента регрессии на уровне значимости 0,01.

Решение. Чтобы построить доверительный интервал для коэффициента регрессии на уровне значимости 0,01, необходимо, используя ссылки на необходимые ячейки, применить формулу:

$$b - m_b \cdot t_{\alpha/2,n-2} \le \beta \le b + m_b \cdot t_{\alpha/2,n-2}$$

$$13,52 - 4,27 \cdot 2,98 \le \beta \le 13,52 + 4,27 \cdot 2,98$$

$$0,80 \le \beta \le 26,24$$

Вывод 6: Диапазон границ доверительного интервала для коэффициента регрессии в модели с хорошим качеством подгонки обычно не превышает 3. В нашем случае, правая граница больше, чем левая, более чем в 3 раза, значит, качество подгонки модели рекомендуется улучшать.

Упражнение 1.9. Используя тест Фишера, проверить статистическую значимость уравнения регрессии на уровне 0,05. Отметим, что здесь эта проверка эквивалентна проверке значимости коэффициента регрессии.

Решение. Чтобы рассчитать значение статистики Фишера, необходимо использовать ссылки на необходимые ячейки согласно формуле: $F = D_{факт.} / D_{ост.} = 66219,00 / 6607,1 = 10,02.$ Чтобы проверить результат вычисления статистики с использованием коэффициента детерминации и статистики Стьюдента для β, необходимо использовать ссылки на необходимые ячейки согласно формулам:

$$F = \frac{R^2}{1-R^2} \cdot (n-2) = \frac{0,42}{1-0,42} \cdot (16-2) \approx 10,02, \; F = t_b^2 = 3,17 \cdot 3,17 \approx 10,02.$$

Для нахождения критического значения статистики Фишера удобно использовать функцию *FРАСПОБР(...)*. В Главном меню MS Excel выберем: Формулы – Вставить функцию – Полный алфавитный перечень – FРАСПОБР(…).

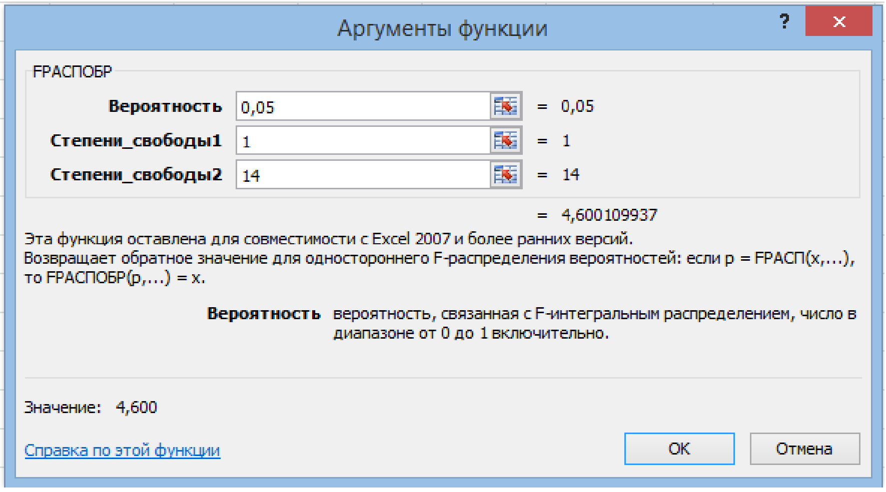


Рис. 3. Диалоговое окно функции FРАСПОБР(…).

Вывод 7: Поскольку 10,02 > 4,60, то гипотеза H_0: $\beta = 0$ отвергается, т.е. согласно тесту Фишера регрессия адекватна. Это означает, что между переменными X (расходы на рекламу) и Y (объем продаж) с надежностью 95% существует значимая линейная связь.

Комментарий: В линейной модели парной регрессии не возникает вопрос о совокупном влиянии нескольких переменных, который актуален для множественной регрессии. Для того, чтобы определить, есть ли линейная зависимость между переменными X (расходы на рекламу) и Y (объем продаж), достаточно проверить значимость коэффициента β при переменной X по тесту Стьюдента. Поэтому для проверки адекватности парной регрессии достаточно выполнить только тест Стьюдента.

Упражнение 1.10. Проверить качество уравнения (а именно, точностные свойства) по средней относительной ошибке аппроксимации.

Решение. Чтобы проверить качество регрессии по средней относительной ошибке аппроксимации, необходимо в таблице 1 заполнить столбец Abs (e/y) и использовать ссылки на необходимые ячейки согласно формуле: $\bar{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right|\cdot 100\%. = \frac{1}{16}\cdot 4,60 = 28,74.$

Вывод 8: Допустимый интервал средней ошибки аппроксимации – от 4% до 7%. Для построенной линейной парной регрессии средняя ошибка аппроксимации составила 28,74%, значит, качество модели рекомендуется улучшать, возможно, путем расширения модели до множественной регрессии через включение в нее дополнительных регрессоров.

Упражнение 1.11. С помощью встроенного инструмента **Регрессия** в Пакете анализа MS Excel построить линейную регрессию объема продаж на расходы на рекламу, спрогнозировать расходы на рекламу в 1 квартале следующего года - x_p, если увеличить их на 5% от среднего значения. Получить прогноз объема продаж в 1 квартале следующего года.

Решение. Чтобы работать с надстройкой *Анализ данных* MS Excel, сначала ее надо подключить. Для этого нажмите *Файл – Параметры*. Далее в меню слева выберите *Надстройки.* Затем в разделе *Управление* выберите *Надстройки Excel* и нажмите кнопку *Перейти* (рис.4).

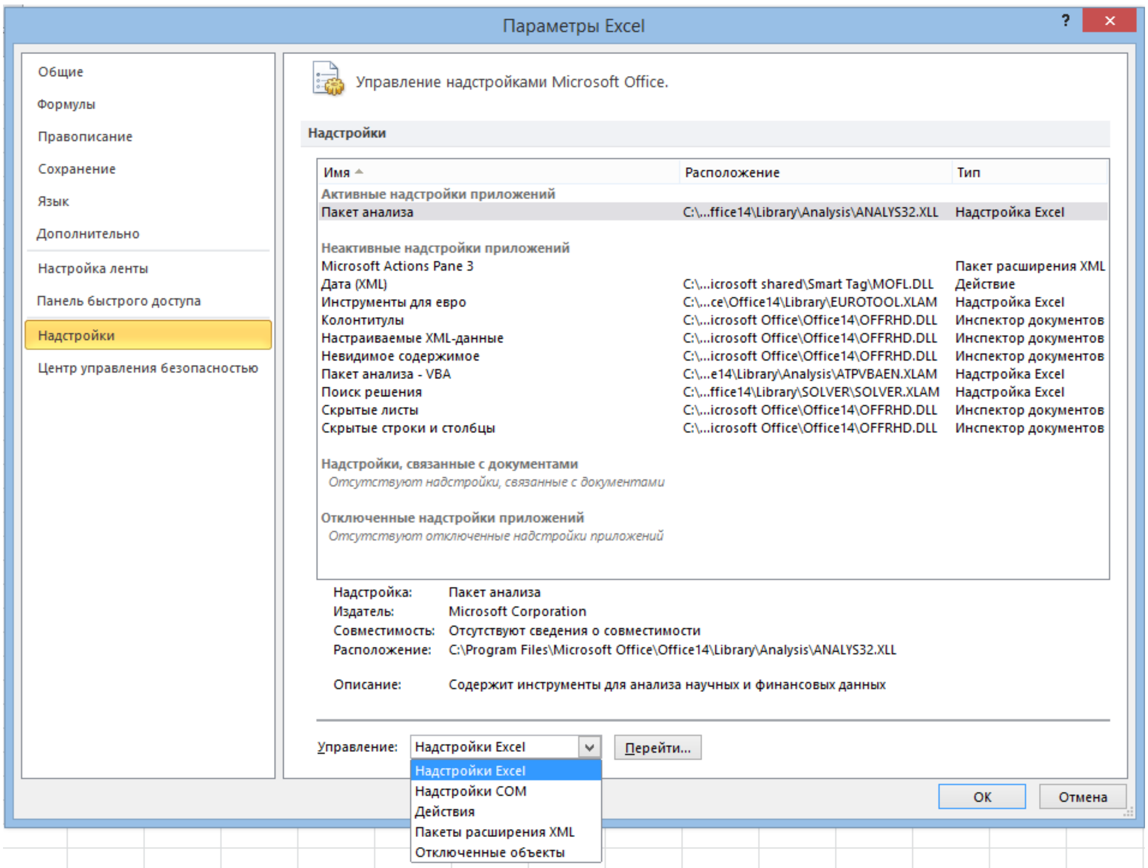


Рис. 4. Включение надстроек в Excel

В появившемся окне выберите надстройку *Пакет анализа*, поставив галочку напротив названия этого расширения (рис.5).

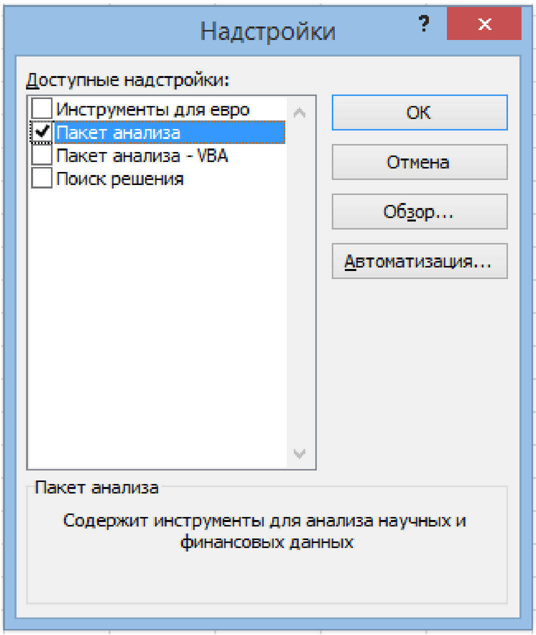


Рис. 5. Включение анализа данных в Excel

Теперь в главном меню программы на вкладке *Данные* (в верхней строке) появилась опция *Анализ данных (рис.6)*. Ее мы и будем в дальнейшем использовать.

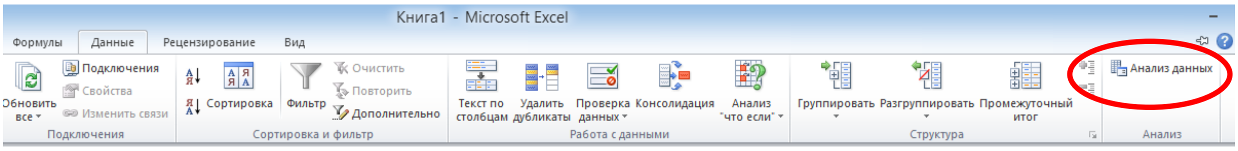


Рис.6. Кнопка *Анализ данных* в Excel

Нажав на кнопку Анализ данных, следует выбрать опцию Регрессия (рис.7).

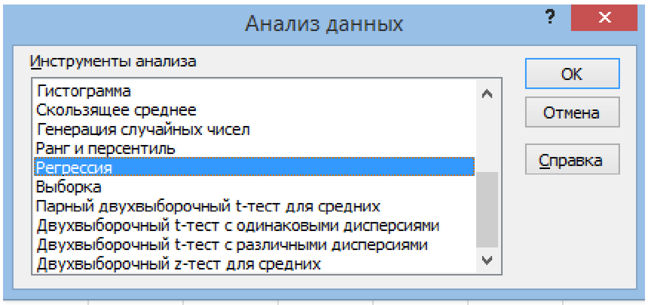


Рис. 7. Выбор регрессионного анализа в Excel

На экране появится новое окно с опциями регрессионного анализа (рис.8).

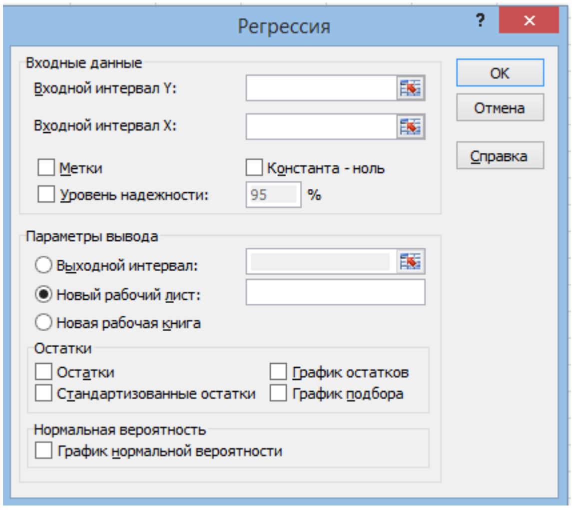


Рис. 8. Окно параметров регрессии в Excel

В поле *Входной интервал Y* необходимо выбрать столбик, который будет отвечать за значения регрессанта – зависимой переменной (это может быть лишь один столбик). В графе *Входной интервал X* необходимо выбрать столбики, которые содержат значения регрессоров. Опция *Метки,* если в ней указать «галочку», трактует первую строку данных как названия переменных. Опция *Константа – ноль* позволяет строить модели без константы. Опция *Уровень надежности* позволяет установить значение *1-α* для данной модели (где α – уровень значимости). В разделе *Параметры вывода* можно устанавливать, куда выводятся результаты регрессии (обычно выводят результаты на этот же лист, так как удобно держать данные и результаты оценки модели на одном листе). С помощью опции *Остатки модели* можно получить ряд остатков модели. С помощью опции *Стандартизированные остатки* вы можете получить ряд остатков, деленных на свое стандартное отклонение. *График остатков* отражает зависимость значения остатков от значений каждого регрессора. *График подбора* показывает зависимость фактических значений регрессанта от каждого регрессора и зависимость предсказанных значений регрессанта от каждого регрессора. Таким образом, благодаря этому графику можно судить о качестве подгонки модели. При построении *графика нормальной вероятности* по оси ординат откладываются значения регрессанта, а по оси абсцисс - процентили нормального распределения.

Чтобы оценить нашу регрессию в Excel в качестве регрессанта выберем переменную Y – объем продаж, а в качестве регрессора выберем переменную X – расходы на рекламу, укажем «галочку» в опции Метки, чтобы трактовать первую строку данных как названия переменных (рис.9).

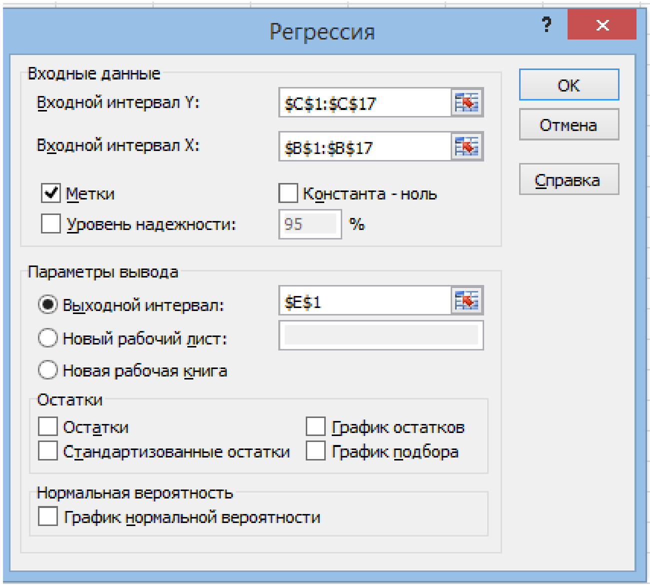
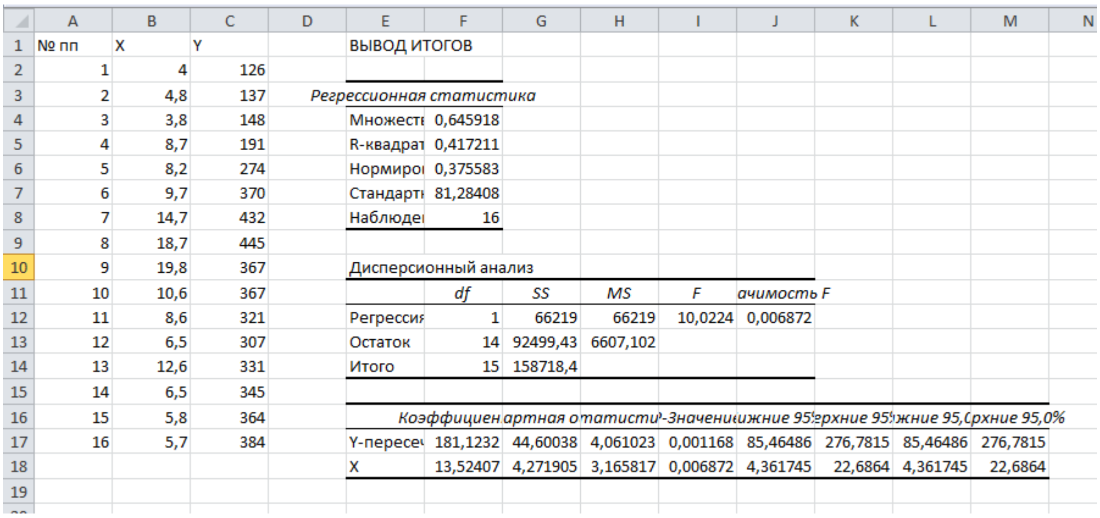


Рис. 9. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.10).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	№ пп	X	Y		ВЫВОД ИТОГОВ								
2	1	4	126										
3	2	4,8	137		*Регрессионная статистика*								
4	3	3,8	148		Множеств	0,645918							
5	4	8,7	191		R-квадрат	0,417211							
6	5	8,2	274		Нормиро	0,375583							
7	6	9,7	370		Стандарт	81,28408							
8	7	14,7	432		Наблюде	16							
9	8	18,7	445										
10	9	19,8	367		Дисперсионный анализ								
11	10	10,6	367			*df*	*SS*	*MS*	*F*	*ачимость F*			
12	11	8,6	321		Регрессия	1	66219	66219	10,0224	0,006872			
13	12	6,5	307		Остаток	14	92499,43	6607,102					
14	13	12,6	331		Итого	15	158718,4						
15	14	6,5	345										
16	15	5,8	364			*Коэффициен*	*артная о*	*татисти*	*Р-Значени*	*ижние 95*	*ерхние 95*	*жние 95,(*	*рхние 95,0%*
17	16	5,7	384		Y-пересе	181,1232	44,60038	4,061023	0,001168	85,46486	276,7815	85,46486	276,7815
18					X	13,52407	4,271905	3,165817	0,006872	4,361745	22,6864	4,361745	22,6864
19													

Рис.10. Вывод результатов оценивания регрессии в Excel

Более подробно рассмотрим результаты, переведя их в обычные таблицы.

Регрессионная статистика						
Множественный R	0,64592					
R-квадрат	0,417211					
Нормированный R-квадрат	0,375583					
Стандартная ошибка	81,28408					
Наблюдения	16					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	66219	66219	10,0224	0,006872	
Остаток	14	92499,43	6607,102			
Итого	15	158718,4				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	181,1232	44,60038	4,061023	0,001168	85,46486	276,7815
X	13,52407	4,271905	3,165817	0,006872	4,361745	22,6864

В подтаблице «Регрессионная статистика» *Множественный R* – это коэффициент корреляции между регрессантом и регрессором. Строкой ниже Excel выдает значение R^2, равное 0,42. Чуть ниже можно увидеть значение *нормированного* (или, по-другому, скорректированного (adjusted) R^2. Еще ниже указано значение *Стандартной ошибки* регрессии (MSE) – квадратного корня из остаточной дисперсии (пересечение столбца *MS* и строки *Остаток* по второй подтаблице), равное 81,28. В конце подтаблицы дано значение количества *Наблюдений*, равное 16.

В подтаблице «Дисперсионный анализ» во втором столбце *df* даны числа степеней свободы, в третьем столбце даны значения *ESS, RSS* и *TSS* (соответственно 66219; 92499,43; 158718,4). Четвертый столбик содержит отношения ESS/df_{ESS}, RSS/df_{RSS} соответственно. *F* – это значение F-статистики теста Фишера для проверки значимости регрессии. В самом правом столбце дано значение *p-value* для этой статистики. Значение *p-value* позволяет провести тест (в данном случае тест Фишера), не пользуясь критическими значениями статистики. Если *p-value* для статистики меньше, чем α =0,1, то нулевая гипотеза отвергается с надежностью 90%. Если *p-value* для статистики меньше, чем α =0,05, то нулевая гипотеза отвергается с надежностью 95%. Если *p-value* для статистики меньше, чем α =0,01, то нулевая гипотеза отвергается с надежностью 99%. В нашем случае *p-value* для статистики Фишера составило 0,007, что меньше, чем α =0,01. Это означает, что с надежностью 99% отвергается нулевая гипотеза о незначимости уравнения регрессии. Согласно тесту Фишера регрессия адекватна, между переменными X и Y существует значимая линейная связь.

В нижней подтаблице в столбце «Коэффициенты» содержатся *коэффициент регрессии*, равный 13,52, а также *свободный член*, равный 181,12 на пересечении со строкой «Y-пересечение». В третьем столбце указаны стандартные ошибки коэффициентов, далее – значение t-статистики теста Стьюдента, затем дано значение *p-value* для этой статистики. В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии составило 0,007, что меньше, чем α =0,01. Это означает, что коэффициент регрессии β является значимым, между переменными X (расходы на рекламу) и Y (объем продаж) существует значимая линейная связь. Также *p-value* для статистики Стьюдента для свободного коэффициента составило 0,001, что меньше, чем α =0,01. Это означает, что свободный коэффициент α является значимым. В двух самых правых столбцах даны доверительные границы коэффициентов. В нашем случае, границы значение «ноль» не включают, что еще раз подтверждает статистическую значимость (отличие от нуля) коэффициентов уравнения регрессии.

Чтобы спрогнозировать расходы на рекламу в 1 квартале следующего года, надо среднее значение X увеличить на 5 % : 9,29*1,05 = 9,75 тыс. руб. Получим прогноз объема продаж в 1 квартале следующего года: Yp = 181,12+13,52*9,75=312,94 тыс.руб.

Упражнение 1.12. С надежностью 95% построить доверительный интервал прогноза объема продаж в 1 квартале следующего года.

Решение. Чтобы построить 95% -й интервал прогноза объема продаж в 1 квартале следующего года, необходимо использовать ссылки на необходимые ячейки согласно формуле:

$$Y_p - m_{\hat{y}} \cdot t_{\alpha/2,n-2} \le y^* \le Y_p + m_{\hat{y}} \cdot t_{\alpha/2,n-2}$$

В данной формуле $m_{\hat{y}}$ - стандартная ошибка прогноза. Ее можно определить для среднего прогнозного значения и для индивидуального прогнозного значения:

a) для среднего прогнозного значения:

$$m_{\hat{y}_x} = S_e \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}}$$

$$m_{\hat{y}_x} = 81,28 \cdot \sqrt{\frac{1}{16} + \frac{(9,75 - 9,29)^2}{16 \cdot 22,63}} = 20,42$$

Тогда доверительный интервал прогноза среднего объема продаж с надежностью 95% составит:

$$312,94 - 20,42 \cdot 2,14 \le y^* \le 312,94 + 20,42 \cdot 2,14$$

$$269,31 \le y^* \le 356,89$$

b) для индивидуального прогнозного значения:

$$m_{\hat{y}_x} = S_e \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}}$$

$$m_{\hat{y}_x} = 81,28 \cdot \sqrt{1 + \frac{1}{16} + \frac{(9,75 - 9,29)^2}{16 \cdot 22,63}} = 83,81$$

Тогда доверительный интервал прогноза индивидуального объема продаж с надежностью 95% составит:

$$312,94 - 83,81 \cdot 2,14 \le y^* \le 312,94 + 83,81 \cdot 2,14$$

$$133,34 \le y^* \le 492,85$$

Раздел 2. Парная нелинейная регрессия и корреляция

В результате изучения данного раздела студент должен:
знать:
основные типы нелинейных зависимостей в парной регрессии;
основные способы линеаризации нелинейных зависимостей;
определение и интерпретацию индекса корреляции для нелинейной парной регрессии;
уметь:
находить МНК-оценки коэффициентов нелинейной парной регрессии;
определять качество подгонки нелинейной парной регрессии;
владеть:
навыками оценки нелинейной парной регрессии в MS Excel.

Задачи с решениями

Задача 2.1. По семи областям региона известны следующие данные:

Номер региона	Расходы на покупку продовольственных товаров, % к общему объему расходов, *y*	Средний заработок 1 работающего за неделю, тыс. ден.ед., *x*
1	68,8	4,5
2	58,3	5,9
3	62,6	5,7
4	52,1	7,2
5	54,5	6,2
6	57,1	6,0
7	51,0	7,8

Для характеристики зависимости доли расходов на покупку продовольственных товаров от доходов рассчитайте параметры следующих функций: линейной, степенной, экспоненты, показательной, равносторонней гиперболы, обратной. Найти показатели тесноты связи для каждой модели.
Решение:
Параметры a и b линейной регрессии $\hat{y} = a + b \cdot x$ рассчитываются в результате решения системы нормальных уравнений относительно a и b:

$$\begin{cases} n \cdot a + b \cdot \sum x = \sum y, \\ a \cdot \sum x + b \cdot \sum x^2 = \sum y \cdot x \end{cases}$$

По исходным данным рассчитываем $\sum y, \sum x, \sum y \cdot x, \sum x^2, \sum y^2$.

Номер региона	X	Y	XY	X^2	Y^2
1	4,5	68,8	309,6	20,25	4733,44
2	5,9	58,3	343,97	34,81	3398,89
3	5,7	62,2	356,82	32,49	3918,76
4	7,2	52,1	375,12	51,84	2714,41
5	6,2	54,5	337,9	38,44	2970,25
6	6	57,1	342,6	36	3260,41
7	7,8	51	397,8	60,84	2601,00
Среднее	6,186	57,77	351,97	39,24	3371,02

Применим следующие формулы для нахождения параметров:

$$b=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{x^2}-\bar{x}^2}=\frac{351{,}97-6{,}186\cdot 57{,}77}{39{,}24-6{,}186^2}=-6{,}0897,$$

$$a=\bar{y}-b\cdot\bar{x}=57{,}77+6{,}0897\cdot 6{,}186=95{,}465.$$

Получаем линейное уравнение регрессии: $\hat{y}=95{,}465-6{,}0897\cdot x$.

Линейное уравнение регрессии дополняется расчетом линейного коэффициента корреляции.

$$r_{xy}=b_1\frac{\sigma_x}{\sigma_y}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sqrt{(\overline{x^2}-\bar{x}^2)\cdot(\overline{y^2}-\bar{y}^2)}}=\frac{351{,}97-6{,}186\cdot 57{,}77}{\sqrt{(39{,}24-6{,}186^2)\cdot(3371{,}02-57{,}77^2)}}=-0{,}9421.$$

2. Регрессия в виде степенной функции имеет вид: $\hat{y}=a\cdot x^b$. Для оценки параметров модели линеаризуем модель путем логарифмирования: $\ln y=\ln a+b\cdot\ln x.$

Обозначим $\ln y=Y,\ln a=A,\ln x=X.$ Тогда получим: $Y=A+b\cdot X.$ Для расчетов составим таблицу.

Номер региона	X	Y	XY	X^2	Y^2	$\hat{Y}$	$(Y-\hat{Y})^2$
1	1,504	4,231	6,364	2,262	17,901	68,6	0,04
2	1,775	4,066	7,217	3,151	16,532	58,6	0,09
3	1,740	4,137	7,198	3,029	17,115	59,9	7,29
4	1,974	3,953	7,803	3,897	15,626	52,3	0,04
5	1,825	3,948	7,296	3,329	15,984	56,9	5,76
6	1,792	4,045	7,249	3,211	16,362	58,1	1,00
7	2,054	3,932	8,076	4,219	15,461	49,9	1,21
Сумма	12,664	28,312	51,203	23,098	114,981	404,3	15,43
Среднее	1,809	4,052	7,315	3,300	16,426	-	-

Применим следующие формулы для нахождения параметров:

$$b=\frac{\overline{XY}-\bar{X}\cdot\bar{Y}}{\overline{X^2}-\bar{X}^2}=\frac{7{,}315-1{,}809\cdot 4{,}052}{3{,}3-1{,}809^2}=-0{,}5770,$$

$$A=\bar{Y}-b\cdot\bar{X}=4{,}052+0{,}5770\cdot 1{,}809=5{,}0956.$$

Уравнение регрессии: $\ln y=5{,}0956-0{,}5770\cdot\ln x.$

Выполнив потенцирование, получим: $y=e^{5{,}0956}\cdot x^{-0{,}5770}=163{,}302\cdot x^{-0{,}5770}.$

Показателем тесноты связи выступает индекс корреляции:

$$R=\sqrt{1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}}=\sqrt{1-\frac{\sum e_i^2}{n\sigma_y^2}}=\sqrt{1-\frac{15{,}43}{7*(16{,}426-4{,}052^2)}}=0{,}9665.$$

3.Регрессия в виде экспоненты имеет вид: $\hat{y}=ae^{bx}$

Для оценки параметров приведем уравнение к линейному виду: $\ln y=\ln a+b\cdot x.$

Обозначим $\ln y=Y,\ln a=A.$ Тогда получим: $Y=A+b\cdot x.$ Для расчетов составим таблицу.

	Y	x	Y*x	x^2
1	4,231	4,5	19,039	20,25
2	4,066	5,9	23,989	34,81
3	4,137	5,7	23,581	32,49

4	3,953	7,2	28,462	51,84
5	3,948	6,2	24,478	38,44
6	4,045	6	24,27	36
7	3,932	7,8	30,669	60,84
Среднее	4,045	6,186	24,927	39,239

Применим следующие формулы для нахождения параметров:

$$b = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - \bar{x}^2} = \frac{24,927 - 6,186 \cdot 4,045}{39,239 - 6,186^2} = -0,094008,$$

$$A = \bar{Y} - b \cdot \bar{x} = 4,045 + 0,09361 \cdot 6,186 = 4,62608.$$

Уравнение регрессии: $\ln y = 4,62608 - 0,094008 \cdot x.$

После потенцирования получаем: $\hat{y} = e^{4,62608} \cdot e^{-0,094008x}$ или $\hat{y} = 102,11303e^{-0,094008x}$.

4. Регрессия в виде показательной кривой $y = a \cdot b^x$ линеаризуется в следующий вид: $\ln y = \ln a + x \ln b.$

Обозначим $\ln y = Y, \ln a = A, \ln b = B.$ Тогда получим: $Y = A + B \cdot x.$ Для расчетов составим таблицу.

	Y	x	Y*x	x^2
1	4,231	4,5	19,039	20,25
2	4,066	5,9	23,989	34,81
3	4,137	5,7	23,581	32,49
4	3,953	7,2	28,462	51,84
5	3,948	6,2	24,478	38,44
6	4,045	6	24,27	36
7	3,932	7,8	30,669	60,84
Среднее	4,045	6,186	24,927	39,239

Применим следующие формулы для нахождения параметров:

$$B = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - \bar{x}^2} = \frac{24,927 - 6,186 \cdot 4,045}{39,239 - 6,186^2} = -0,094008,$$

$$A = \bar{Y} - b \cdot \bar{x} = 4,045 + 0,09361 \cdot 6,186 = 4,62608.$$

Уравнение регрессии: $\ln y = 4,62608 - 0,094008 \cdot x.$

После потенцирования получаем: $\hat{y} = e^{4,62608} \cdot (e^{-0,094008})^x$ или $\hat{y} = 102,11303 \cdot 0,91028^x$.

5. Регрессия в виде равносторонней гиперболы имеет вид: $\hat{y} = a + \frac{b}{x}$. Чтобы оценить параметры a и b, приведем модель к линейному виду, заменив $\frac{1}{x} = z$. Тогда $y = a + b \cdot z.$ Для расчета параметров составим таблицу.

	y	z=1/x	zy	z^2	$\hat{y}$	$(y-\hat{y})^2$
1	68,8	0,222	15,289	0,049	69,9	1,21
2	58,3	0,170	9,881	0,029	58,5	0,04
3	62,6	0,175	10,983	0,031	59,8	7,84
4	52,1	0,139	7,236	0,019	51,9	0,01
5	54,5	0,161	8,790	0,026	56,7	4,84

6	57,1	0,167	9,517	0,028	57,9	0,64
7	51,0	0,128	6,538	0,016	49,6	1,96
Сумма	404,4	1,162	68,234	0,198	404,4	16,54
Среднее	57,771	0,166	9,748	0,028	57,757	2,363

Применим следующие формулы для нахождения параметров:

$$b = \frac{\overline{zy} - \overline{z} \cdot \overline{y}}{\overline{z^2} - \overline{z}^2} = \frac{9,748 - 0,166 \cdot 57,771}{0,028 - 0,166^2} = 216,053,$$

$$a = \overline{y} - b \cdot \overline{z} = 57,771 - 216,053 \cdot 0,166 = 21,907.$$

Уравнение регрессии: $\hat{y} = 21,907 + \dfrac{216,053}{x}.$

Для оценки тесноты связи найдем индекс корреляции.

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}} = \sqrt{1 - \frac{\sum e_i^2}{n\sigma_y^2}} = \sqrt{1 - \frac{16,54}{234,42}} = 0,9641.$$

6. Регрессия в виде обратной функции имеет вид: $\hat{y} = \dfrac{1}{a + bx}.$ Приведем ее к линейному виду, заменив $Z = \dfrac{1}{y}.$ Тогда $Z = a + b \cdot x$. Для расчета параметров составим таблицу.

	Z=1/y	x	Z*x	x^2	$\hat{y}$	$(y-\hat{y})^2$
1	0,01454	4,5	0,0654	20,25	67,7	1,21
2	0,01715	5,9	0,1012	34,81	58,8	0,25
3	0,01597	5,7	0,0910	32,49	59,9	7,29
4	0,01919	7,2	0,1382	51,84	52,3	0,04
5	0,01835	6,2	0,1138	38,44	57,1	6,76
6	0,01751	6	0,1051	36	58,2	1,21
7	0,01961	7,8	0,1529	60,84	49,8	1,44
Сумма	0,122	43,300	0,768	274,670	403,800	18,200
Среднее	0,017	6,186	0,110	39,239	57,686	2,600

Применим следующие формулы для нахождения параметров:

$$b = \frac{\overline{Zx} - \overline{Z} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} = \frac{0,110 - 0,017 \cdot 6,186}{39,239 - 6,186} = 0,002,$$

$$a = \overline{Z} - b \cdot \overline{x} = 0,017 - 0,002 \cdot 6,186 = 0,007.$$

Уравнение регрессии: $\hat{y} = \dfrac{1}{0,007 + 0,002x}.$

Для оценки тесноты связи найдем индекс корреляции.

$$R = \sqrt{1 - \frac{18,20}{234,42}} = \sqrt{0,92236} = 0,9604.$$

Задача 2.2. Исследуя спрос на продукцию фирмы, аналитический отдел собрал данные по 20 торговым точкам компании и представил их в виде:

$$\ln y = 6,8 - 0,6 \ln x + \varepsilon,$$
$$(2,7) \quad (-2,8)$$

где y – объем спроса,

x – цена единицы продукции.
В скобках приведены фактические значения t – критерия.
Ранее предполагалось, что увеличение цены на 1% приводит к уменьшению спроса на 1,2%. Можно ли утверждать, что приведенные результаты подтверждают это предположение?

Решение:
Уравнение регрессии в прологарифмированном виде. Судя по форме записи, уравнение имеет степенной вид и записывается так:

$$\hat{y} = e^{6,8} \cdot x^{-0,6}$$

Надо проверить предположение о том, что эластичность спроса по цене равна –1,2. В степенной зависимости эластичность равна показателю степени b, поэтому оценка эластичности равна –0,6. Таким образом, задача сводится к проверке статистической гипотезы H_0: b=-1,2 при альтернативной H_1:b≠-1,2. Критическая область двусторонняя, поэтому проверка гипотезы может быть заменена построением доверительного интервала для b и, если проверяемое значение b=-1,2 попадает в него, то нуль-гипотеза не отклоняется; в противном случае принимается альтернативная гипотеза.
Доверительный интервал строится по формуле:

$$- 0,6- m_b*t_{таб} \leq \beta \leq - 0,6+ m_b*t_{табл}.$$

Определим стандартную ошибку параметра b из формулы:

$$m_b = \frac{b}{t_b} = \frac{-0,6}{-2,8} = 0,2143$$

Для определения $t_{табл}$ зададим уровень значимости, равный 0,05, следовательно:$t_{табл}(\alpha;\ n\text{-}2) = t_{табл}(0,05;18) = 2,1$(используем таблицу критических точек распределения Стьюдента для двустороннего α=0,05).
Доверительный интервал равен:

$$-0,6-0,2143*2,1 \leq \beta \leq -0,6+0,2143*2,1$$

или $-1,05 \leq \beta \leq -0,15$.
Значение, равное –1,2, в интервал не попадает, следовательно, предположение о значении коэффициента эластичности на уровне значимости 0,05 следует отклонить. Однако, если задать значимость на уровне 0,01, то $t_{табл}$=2,88, и интервал будет таким: -1,217 < b < 0,017
Следовательно, на уровне значимости 0,01 первоначальное предположение не может быть отклонено, поскольку значение –1,2 попадает в доверительный интервал.
Можно проверить статистическую гипотезу напрямую, вычислив t–статистику для разницы между гипотетическими и вычисленным значениями b:

$$t_{b\text{-}bo} = \frac{b-b_0}{m_b} = \frac{-0,6-(-1,2)}{0,2143} = 2,8.$$

Сравним полученную статистику по абсолютной величине с критическим значением на заданном уровне значимости. На уровне α=0,05: $\left|t_{b-b_0}\right| = 2,8 > \left|t_{табл}\right| = 2,1$;

Нуль-гипотеза отклоняется, эластичность спроса по цене не может быть равна –1,2. На уровне α=0,01: $\left|t_{b-b_0}\right| = 2,8 \leq \left|t_{табл}\right| = 2,88$;нуль-гипотеза не отклоняется, эластичность может быть равна –1,2.

Задача 2.3. Для двух видов продукции А и Б зависимости удельных постоянных расходов от объема выпускаемой продукции выглядят следующим образом:

$$\hat{y}_A = 15 + 8\ln x,$$

$$\hat{y}_Б = 25x^{0,3}.$$

Сравнить эластичность затрат по каждому виду продукции при x=50 и определить объемы продукции обоих видов, при котором эластичности будут одинаковы.

Решение:

Регрессионная зависимость для продукции А является полулогарифмической, и для вычисления эластичности воспользуемся формулой:

$$Э_А = \frac{b}{a + b\ln x} = \frac{8}{15 + 8\ln 50} = 0{,}173.$$

Для продукции Б регрессионная зависимость является степенной, где коэффициент эластичности равен показателю степени при любых значениях независимой переменной, следовательно:$Э_Б$ = 0,3. Теперь определим точку, в которой эластичности по обоим видам продукции одинаковы. Для продукции Б подходит любой объем, т.к. эластичность постоянна, а для определения объема выпуска продукции Б составим и решим уравнение: $\frac{8}{15 + 8\ln x} = 0{,}3$; отсюда $X_А$ = 4,3 единиц.

Таким образом, при объеме производства продукции А, равном 4,3, эластичности удельных постоянных расходов обоих видов продукции по объему выпуска одинаковы и равны 0,3.

Задача 2.4. Зависимость среднемесячной производительности труда от возраста рабочих характеризуется моделью: $y = a + bx + cx^2$. Ее использование привело к результатам, представленным в таблице:

№ п/п	Производительность труда рабочих, тыс.руб. (y)		№ п/п	Производительность труда рабочих, тыс. руб. (y)	
	фактическая	расчетная		фактическая	расчетная
1	16	15	6	18	16
2	13	14	7	11	13
3	15	14	8	12	12
4	12	10	9	14	14
5	14	16	10	15	17

Оценить качество модели, определив ошибку аппроксимации, индекс корреляции.

Решение:

Средняя ошибка аппроксимации рассчитывается по формуле:

$$\overline{A} = \frac{1}{n}\Sigma\left|\frac{y - \hat{y}}{y}\right| * 100\%;$$

и характеризует среднее отклонение расчетных значений от фактических. Это значение считается приемлемым, если оно не превышает 8-10%.

Для приведенных в таблице данных имеем:

$$\overline{A} = \frac{1}{10}\left[\left|\frac{16-15}{16}\right| + \left|\frac{13-14}{13}\right| + \ldots + \left|\frac{15-17}{15}\right|\right] * 100\% = 9{,}42\%,$$

что оказывается в допустимых границах и говорит о приемлемой точности аппроксимации регрессионной модели.

Индекс корреляции рассчитаем по формуле , предварительно определив общую и остаточную СКО:

$$\Sigma(y-\bar{y})^2 = \Sigma y^2 - n\bar{y}^2,\ \bar{y} = \frac{1}{n}\Sigma y = 14,$$

$$\Sigma(y-\bar{y})^2 = 16^2 + 13^2 + \ldots + 15^2 - 10 * 14^2 = 2000 - 1960 = 40,$$

$$\Sigma(y-\hat{y})^2 = (16-15)^2 + (13-14)^2 + \ldots + (15-17)^2 = 23,$$

$$R = \sqrt{1 - \frac{23}{40}} = 0{,}652,$$ R^2=0,425.

Задача 2.5. Для следующих уравнений регрессии:

$$y = 3{,}7 + 0{,}0024x + e$$

$$y = 3{,}7 + 0{,}0024\ln x + e$$

$$\ln y = 3{,}7 + 0{,}0024\ln x + \ln e$$

$$\ln y = 3{,}7 + 0{,}0024x + \ln e$$

Определить коэффициенты эластичности при значении фактора, равном 85.

Решение:

Уравнение регрессии является линейным, поэтому коэффициент эластичности равен $Э = y'\frac{x}{y} = \frac{bx}{a+bx} = \frac{0{,}0024 \cdot 85}{3{,}7 + 0{,}0024 \cdot 85} = 0{,}052 \cdot$

Здесь имеем дело с полулогарифмической зависимостью: $Э = y'\frac{x}{y} = \frac{b}{a+b\ln x} = \frac{0{,}0024}{3{,}7 + 0{,}0024 \cdot \ln 85} = 0{,}000647 \cdot$

Это преобразованная (путем логарифмирования) степенная зависимость; её коэффициент эластичности постоянен и равен показателю степени, т.е. 0,0024.

В данном случае зависимость показательная (или экспоненциальная), в преобразованном виде логарифмируется только зависимая переменная. В любой из трех форм записи экспоненциальной регрессии коэффициент эластичности равен произведению коэффициента при факторе на значение самого фактора, т.е. $Э = y'\frac{x}{y} = 0{,}0024 \cdot 85 = 0{,}204 \cdot$

Упражнения с пояснениями

Предлагается оценить пять нелинейных регрессий (степенная, экспоненциальная, полулогарифмическая, обратная, гиперболическая) для аппроксимации зависимости объема реализации продукции (Y, млн. руб.) от цены (X, тыс. руб.):

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	23	18	27	29	43	23	55	47	35	38	14	51	20	39	35
X	37	33	15	36	26	24	15	33	44	34	63	8	44	43	31

Для осуществления выбора парной нелинейной регрессии с наиболее хорошим качеством подгонки в каждом упражнении выполняется 5 шагов:

линеаризация исходной парной нелинейной регрессии;

оценивание линеаризованного уравнения с помощью инструмента Регрессия надстройки Пакет анализа MS Excel;
получение предсказанных значений по исходному нелинейному уравнению и расчет *ESS, RSS, TSS* в расчетной таблице;
вычисление индекса детерминации и индекса корреляции;
вычисление среднего коэффициента эластичности и средней относительной ошибки аппроксимации.

Упражнение 2.1. Оценить параметры степенной регрессии $\hat{y} = ax^b$. Вычислить индексы детерминации и корреляции, средний коэффициент эластичности, среднюю относительную ошибку аппроксимации.

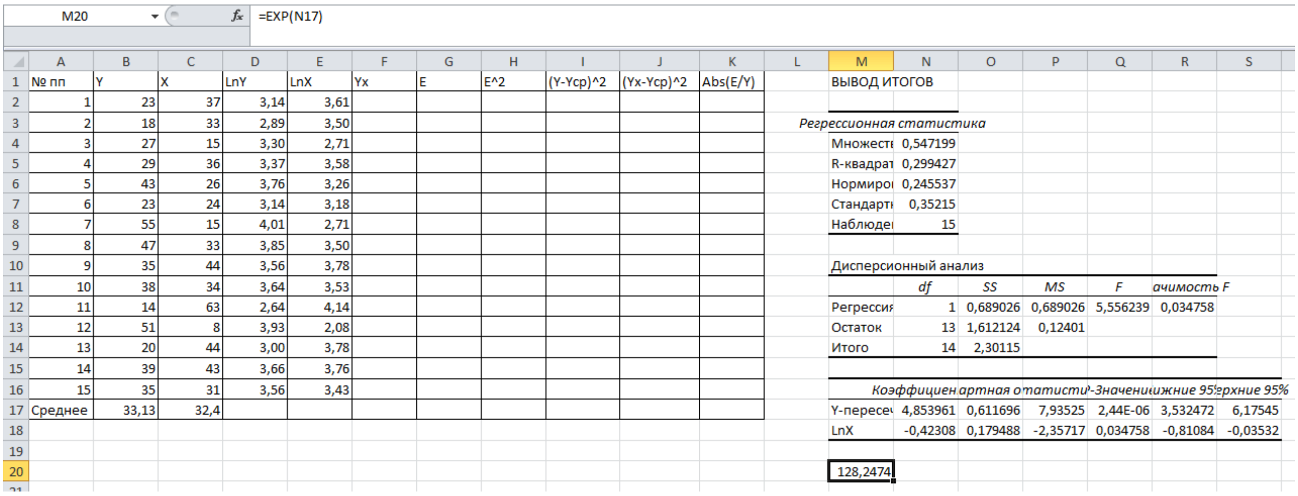
Решение. Чтобы оценить коэффициенты a и b исходной степенной регрессии достаточно заполнить столбцы y, x, lny, lnx в расчетной табл. 1, используя Мастер формул MS Excel.

Табл. 1

Подготовка данных для оценивания нелинейной степенной регрессии объема продаж по цене

№ пп	y	x	lny	lnx	$\hat{y}$	e	e^2	(y-$y_{ср}$)^2	($\hat{y}$ -$y_{ср}$)^2	Abs(e/y)
1	23	37	3,14	3,61						
2	18	33	2,89	3,50						
3	27	15	3,30	2,71						
4	29	36	3,37	3,58						
5	43	26	3,76	3,26						
6	23	24	3,14	3,18						
7	55	15	4,01	2,71						
8	47	33	3,85	3,50						
9	35	44	3,56	3,78						
10	38	34	3,64	3,53						
11	14	63	2,64	4,14						
12	51	8	3,93	2,08						
13	20	44	3,00	3,78						
14	39	43	3,66	3,76						
15	35	31	3,56	3,43						
Среднее	33,13	32,4								

Чтобы оценить коэффициенты a и b нелинейной исходной степенной регрессии $\hat{y} = ax^b$, ее необходимо линеаризовать через логарифмирование: $\ln y = A + b\ln x$, где $a = \exp(A)$. В инструменте *Регрессия* надстройки *Пакет анализа* MS Excel в поле *Входной интервал Y* необходимо выбрать столбик lny, в поле *Входной интервал X* необходимо выбрать столбик lnx. Далее, нажав на кнопку Ок, получаем результаты (рис 1).



M20 =EXP(N17)

№ пп	Y	X	LnY	LnX	Yx	E	E^2	(Y-Ycp)^2	(Yx-Ycp)^2	Abs(E/Y)
1	23	37	3,14	3,61						
2	18	33	2,89	3,50						
3	27	15	3,30	2,71						
4	29	36	3,37	3,58						
5	43	26	3,76	3,26						
6	23	24	3,14	3,18						
7	55	15	4,01	2,71						
8	47	33	3,85	3,50						
9	35	44	3,56	3,78						
10	38	34	3,64	3,53						
11	14	63	2,64	4,14						
12	51	8	3,93	2,08						
13	20	44	3,00	3,78						
14	39	43	3,66	3,76						
15	35	31	3,56	3,43						
Среднее	33,13	32,4								

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множеств	0,547199
R-квадрат	0,299427
Нормиро	0,245537
Стандарт	0,35215
Наблюде	15

Дисперсионный анализ

	df	SS	MS	F	ачимость F
Регрессия	1	0,689026	0,689026	5,556239	0,034758
Остаток	13	1,612124	0,12401		
Итого	14	2,30115			

	Коэффициен	артная о	татисти	-Значени	ижние 95	ерхние 95%
Y-пересеч	4,853961	0,611696	7,93525	2,44E-06	3,532472	6,17545
LnX	-0,42308	0,179488	-2,35717	0,034758	-0,81084	-0,03532

128,2474

Рис. 1. Вывод результатов оценивания промежуточной вспомогательной регрессии в Excel для степенной зависимости

Чтобы восстановить оценку свободного коэффициента *a* ниже вывода итогов необходимо взять экспоненту для A (4,853…) с помощью функции *EXP()* в MS Excel. Теперь можно продолжить заполнение расчетной табл. 1 до конца. Чтобы заполнить столбец $\hat{y}$ надо использовать исходную степенную регрессию $\hat{y} = ax^b$ и ссылки на нужные ячейки в Мастере формул MS Excel. Например, для первого значения:

$\hat{y}$ =128,24…*$37^{-0,42\ldots}$=27,83. В столбец *e* надо поместить остатки регрессии (e = y - $\hat{y}$).

Табл. 1

Подготовка данных для оценивания нелинейной степенной регрессии объема продаж по цене

№ пп	y	x	lny	lnx	$\hat{y}$	e	e^2	(y-$y_{ср}$)^2	($\hat{y}$ - $y_{ср}$)^2	Abs(e/y)
1	23	37	3,14	3,61	27,83	-4,83	23,36	102,68	28,09	0,21
2	18	33	2,89	3,50	29,21	-11,21	125,75	229,02	15,36	0,62
3	27	15	3,30	2,71	40,78	-13,78	189,93	37,62	58,50	0,51
4	29	36	3,37	3,58	28,16	0,84	0,71	17,08	24,75	0,03
5	43	26	3,76	3,26	32,31	10,69	114,18	97,35	0,67	0,25
6	23	24	3,14	3,18	33,43	-10,43	108,73	102,68	0,09	0,45
7	55	15	4,01	2,71	40,78	14,22	202,16	478,15	58,50	0,26
8	47	33	3,85	3,50	29,21	17,79	316,34	192,28	15,36	0,38
9	35	44	3,56	3,78	25,87	9,13	83,43	3,48	52,81	0,26
10	38	34	3,64	3,53	28,85	9,15	83,77	23,68	18,37	0,24
11	14	63	2,64	4,14	22,22	-8,22	67,60	366,08	119,06	0,59
12	51	8	3,93	2,08	53,21	-2,21	4,87	319,22	402,94	0,04
13	20	44	3,00	3,78	25,87	-5,87	34,41	172,48	52,81	0,29
14	39	43	3,66	3,76	26,12	12,88	165,92	34,42	49,20	0,33
15	35	31	3,56	3,43	30,00	5,00	25,03	3,48	9,84	0,14
Среднее	33,13	32,4				Сумма	1546,21	2179,73	906,35	4,61

Чтобы вычислить индекс детерминации, необходимо, используя ссылки на необходимые ячейки, применить формулу:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$$

$$R^2 = 1 - \frac{1546,21}{2179,73} = 0,29$$

Для определения индекса корреляции достаточно взять квадратный корень из индекса детерминации: $R = \sqrt{0,29} = 0,54$.

Вывод 1: Степенная регрессия объема продаж по цене объясняет 29% его вариации вокруг своего среднего значения. Корреляция умеренная.

Чтобы вычислить средний коэффициент эластичности, необходимо, используя ссылки на необходимые ячейки, применить формулу: $Э = (a \cdot x^b)' \cdot \frac{x}{y} = a \cdot b \cdot x^{b-1} \cdot \frac{x}{a \cdot x^b} = \frac{a \cdot b \cdot x^{b-1}}{a \cdot x^{b-1}} = b$, *Э=b= -0,42.*

Вывод 2: Средний коэффициент эластичности показывает, что увеличение цены на 1% приводит к уменьшению объема продаж на 0,42%.

Чтобы вычислить среднюю относительную ошибку аппроксимации, необходимо, используя сумму по последнему столбцу расчетной таблицы и ссылки на необходимые ячейки, применить формулу:

$$A = \frac{1}{n} \cdot \sum \left| \frac{y - y_x}{y} \right| \cdot 100$$

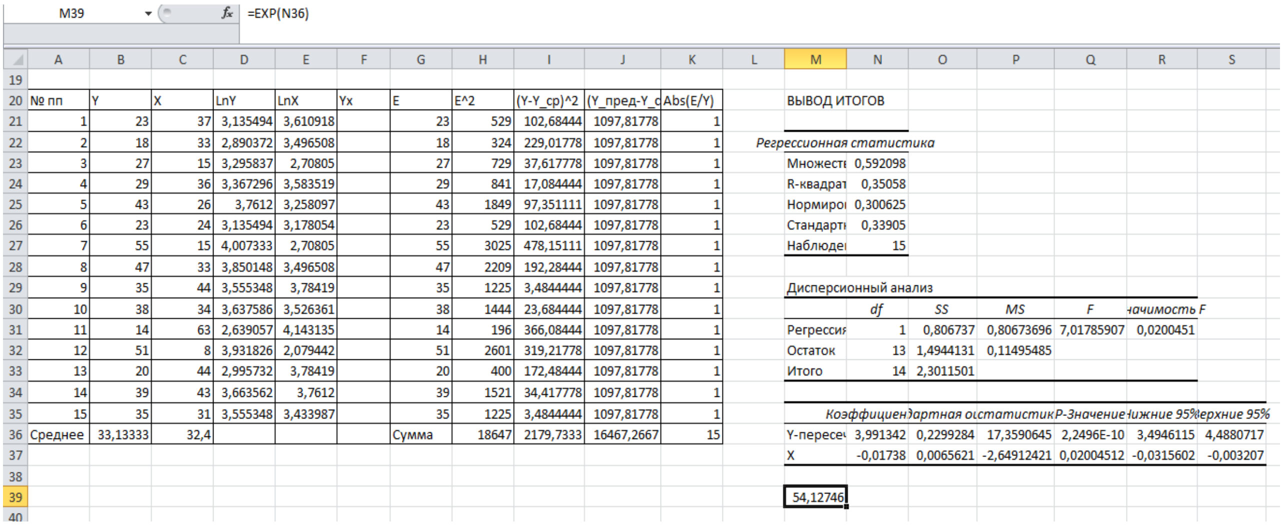
$$A = \frac{1}{16} \cdot 4,61 \cdot 100 = 30,74\%$$

Вывод 3: Допустимый интервал средней ошибки аппроксимации – от 4% до 7%. Для построенной парной нелинейной степенной регрессии средняя ошибка аппроксимации составила 30,74%, значит, качество подгонки модели рекомендуется улучшать, возможно, путем использования других нелинейных зависимостей.

Упражнение 2.2. Оценить параметры экспоненциальной регрессии $\hat{y} = ae^{bx}$. Вычислить индексы детерминации и корреляции, средний коэффициент эластичности, среднюю относительную ошибку аппроксимации.

Решение. Для построения экспоненциальной и последующих нелинейных регрессий удобнее всего скопировать полностью заполненную результатами степенной зависимости расчетную таблицу 1 и вставить ее ниже на этом же листе MS Excel в режиме вставки формул. Для этого после копирования в контекстном меню (по щелчку правой кнопки мыши) выбрать *Специальная вставка – Вставка – Формулы.* Либо сразу в контекстном меню нажать на значок f_x. После вставки в расчетной табл. 1 надо удалить расчетные значения зависимой переменной - содержимое столбца $\hat{y}$. Так мы получили «заготовку» для расчетной табл. 2.

Затем надо оценить коэффициенты a и b нелинейной исходной экспоненциальной регрессии $\hat{y} = ae^{bx}$, ее необходимо линеаризовать через логарифмирование: $\ln y = A + b \cdot x$, где $a = \exp(A)$. В инструменте *Регрессия* надстройки *Пакет анализа* MS Excel в поле *Входной интервал Y* необходимо выбрать столбик *lny,* в поле *Входной интервал X* необходимо выбрать столбик *x*. Далее, нажав на кнопку Ок, получаем результаты (рис 2).



M39 =EXP(N36)

№ пп	Y	X	LnY	LnX	Yx	E	E^2	(Y-Y_ср)^2	(Y_пред-Y_с	Abs(E/Y)
1	23	37	3,135494	3,610918		23	529	102,68444	1097,81778	1
2	18	33	2,890372	3,496508		18	324	229,01778	1097,81778	1
3	27	15	3,295837	2,70805		27	729	37,617778	1097,81778	1
4	29	36	3,367296	3,583519		29	841	17,084444	1097,81778	1
5	43	26	3,7612	3,258097		43	1849	97,351111	1097,81778	1
6	23	24	3,135494	3,178054		23	529	102,68444	1097,81778	1
7	55	15	4,007333	2,70805		55	3025	478,15111	1097,81778	1
8	47	33	3,850148	3,496508		47	2209	192,28444	1097,81778	1
9	35	44	3,555348	3,78419		35	1225	3,4844444	1097,81778	1
10	38	34	3,637586	3,526361		38	1444	23,684444	1097,81778	1
11	14	63	2,639057	4,143135		14	196	366,08444	1097,81778	1
12	51	8	3,931826	2,079442		51	2601	319,21778	1097,81778	1
13	20	44	2,995732	3,78419		20	400	172,48444	1097,81778	1
14	39	43	3,663562	3,7612		39	1521	34,417778	1097,81778	1
15	35	31	3,555348	3,433987		35	1225	3,4844444	1097,81778	1
Среднее	33,13333	32,4				Сумма	18647	2179,7333	16467,2667	15

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множеств	0,592098
R-квадрат	0,35058
Нормиро	0,300625
Стандарт	0,33905
Наблюде	15

Дисперсионный анализ

	df	SS	MS	F	ачимость F
Регрессия	1	0,806737	0,80673696	7,01785907	0,0200451
Остаток	13	1,4944131	0,11495485		
Итого	14	2,3011501			

	Коэффициен	дартная о	статистик	P-Значение	ижние 95%	ерхние 95%
Y-пересе	3,991342	0,2299284	17,3590645	2,2496E-10	3,4946115	4,4880717
X	-0,01738	0,0065621	-2,64912421	0,02004512	-0,0315602	-0,003207

54,12746

Рис. 2. «Заготовка» расчетной таблицы 2 после удаления столбца Yx и вывод результатов оценивания промежуточной вспомогательной регрессии в Excel для экспоненциальной зависимости

Чтобы восстановить оценку свободного коэффициента *a* ниже вывода итогов необходимо взять экспоненту для A (3,991…) с помощью функции *EXP()* в MS Excel.

Теперь можно завершить заполнение расчетной таблицы 2. Чтобы заполнить столбец $\hat{y}$ надо использовать исходную экспоненциальную регрессию $\hat{y} = ae^{bx}$ и ссылки на нужные ячейки в Мастере формул MS Excel. Например, для первого значения:

$\hat{y}$ =54,12…*e$^{-0,01…*37}$=28,45.

После заполнения столбца Yx, благодаря скопированным формулам, значения в других столбцах расчетной табл. 2 меняются автоматически:

Табл. 2

Подготовка данных для оценивания нелинейной экспоненциальной регрессии объема продаж в зависимости от цены

№ пп	y	x	lny	lnx	$\hat{y}$	e	e^2	(y-y$_{ср}$)^2	($\hat{y}$ - y$_{ср}$)^2	Abs (e/y)
1	23	37	3,14	3,61	28,45	-5,45	29,70	102,68	21,93	0,24
2	18	33	2,89	3,50	30,50	-12,50	156,21	229,02	6,94	0,69
3	27	15	3,30	2,71	41,70	-14,70	216,19	37,62	73,45	0,54
4	29	36	3,37	3,58	28,95	0,05	0,00	17,08	17,51	0,00
5	43	26	3,76	3,26	34,45	8,55	73,19	97,35	1,72	0,20
6	23	24	3,14	3,18	35,66	-12,66	160,37	102,68	6,40	0,55
7	55	15	4,01	2,71	41,70	13,30	176,80	478,15	73,45	0,24
8	47	33	3,85	3,50	30,50	16,50	272,30	192,28	6,94	0,35
9	35	44	3,56	3,78	25,19	9,81	96,23	3,48	63,09	0,28

10	38	34	3,64	3,53	29,97	8,03	64,43	23,68	9,99	0,21
11	14	63	2,64	4,14	18,10	-4,10	16,85	366,08	225,86	0,29
12	51	8	3,93	2,08	47,10	3,90	15,21	319,22	195,07	0,08
13	20	44	3,00	3,78	25,19	-5,19	26,94	172,48	63,09	0,26
14	39	43	3,66	3,76	25,63	13,37	178,70	34,42	56,27	0,34
15	35	31	3,56	3,43	31,58	3,42	11,71	3,48	2,42	0,10
Среднее	33,13	32,4				Сумма	1494,84	2179,73	824,14	4,38

Для вычисления индекса детерминации и индекса корреляции в экспоненциальной и других зависимостях необходимо применять те же формулы, что и для степенной зависимости.

Для вычисления средней ошибки аппроксимации в экспоненциальной и других зависимостях необходимо применять те же формулы, что и для степенной зависимости.

Для вычисления среднего коэффициента эластичности в экспоненциальной и других зависимостях необходимо использовать формулы, которые указаны в табл. 3:

Табл. 3

Средний коэффициент эластичности для линейной и нелинейных зависимостей

Зависимость	Коэффициент эластичности
$y = a + bx + \varepsilon$	$\dfrac{bx}{a+bx}$
$y = ax^b\varepsilon$	b
$y = ae^{bx}\varepsilon$	$b \cdot x$
$y = a + b\ln x + \varepsilon$	$\dfrac{b}{a+b\ln x}$
$y = \dfrac{1}{a+bx+\varepsilon}$	$\dfrac{-b \cdot x}{a + b \cdot x}$
$y = a + \dfrac{b}{x} + \varepsilon$	$\dfrac{-b}{ax+b}$

Аналогично степенно й и экспоненциальной зависимостям самостоятельно предлагается составить расчетные таблицы и оценить коэффициенты a и b для парных полулогарифмической $\hat{y} = a + b\ln x$, обратной $\hat{y} = \dfrac{1}{a+bx}$ и гиперболической регрессий $\hat{y} = a + \dfrac{b}{x}$. Затем результаты надо обобщить в сводной табл. 4.

Табл. 4

Сводная таблица оценивания нелинейных регрессий объема продаж в зависимости от цены

Зависимость	Регрессия	R^2	R	A	Э
Степенная	$\hat{y} = 128,25x^{-0,42}\varepsilon$	0,29	0,53	30,74	-0,42
Экспоненциальная	$\hat{y} = 54,12e^{-0,02x}\varepsilon$	0,31	0,56	29,21	-0,56
Полулогарифмическая	$\hat{y} = 79,01 - 13,61\ln x + \varepsilon$	0,33	0,57	31,46	-0,43

Обратная	$\hat{y} = \dfrac{1}{0,0134 + 0,0007x + \varepsilon}$	0,25	0,50	29,10	-0,62
Гиперболическая	$\hat{y} = 23,52 + \dfrac{240,28}{x} + \varepsilon$	0,29	0,53	33,48	-0,24

Вывод: Исходя из индекса детерминации и средней ошибки аппроксимации нельзя выбрать нелинейную регрессию с хорошим качеством подгонки. Рекомендуется прежде всего увеличить выборку наблюдений.

Раздел 3. Множественная регрессия и корреляция

В результате изучения данного раздела студент должен:

знать:

формулу для нахождения МНК-оценок коэффициентов множественной регрессии;

определение и интерпретацию коэффициента множественной детерминации;

уметь:

проверять гипотезы о значимости коэффициентов множественной регрессии;

строить доверительные интервалы для коэффициентов множественной регрессии;

проверять гипотезы об адекватности множественной регрессии;

находить МНК-оценки коэффициентов множественной регрессии;

определять качество подгонки множественной регрессии с помощью коэффициента множественной детерминации;

владеть:

навыками оценки множественной регрессии в MS Excel;

навыками интерпретации основных результатов оценки множественной регрессии, содержащихся в таблице дисперсионного анализа.

Задачи с решениями

Задача 3.1. Уравнение регрессии, построенное по 17 наблюдениям, имеет вид:

$$\begin{array}{lcccc} y = & ?+ & 0,36x_1 - & 9,6x_2 + & ?\,x_3 \\ m_{b_j} & (3) & (\;) & (3,0) & (5,0) \\ t_{b_j} & (1,4) & (1,5) & (\;) & (2,4) \end{array}$$

Расставить пропущенные значения, а также построить доверительный интервал для b_2 с надежностью 0,99.

Решение: Пропущенные значения определяем с помощью формулы:

$$a = t_a \cdot m_a = 1,4 \cdot 3,0 = 4,2; \quad m_{b_1} = \frac{b_1}{t_{b_1}} = \frac{0,36}{1,5} = 0,24; \qquad t_{b_2} = \frac{b_2}{m_{b_2}} = \frac{-9,6}{3,0} = -3,2;$$

$$b_3 = t_{b_3} \cdot m_{b_3} = 2,4 \cdot 5,0 = 12,0.$$

Таким образом, уравнение регрессии со статистическими характеристиками выглядит так:

$$\begin{array}{lcccc} y = & 4,2 + & 0,36x_1 - & 9,6x_2 + & 12x_3 \\ m_{b_j} & (3) & (0,24) & (3,0) & (5,0) \\ t_{b_j} & (1,4) & (1,5) & (-3,2) & (2,4) \end{array}$$

Для доверительного интервала для b_2 уровень значимости равен 0,01, а число степеней свободы равно $n - p - 1 = 17 - 3 - 1 = 13$, где $n = 17$ – объём выборки, $p = 3$ – число факторов в уравнении регрессии. Отсюда $t\ (0,01;13) = 3,0123; \quad m_{b_2} = 3,0; \quad b_2 = -9,6.$

$-9,6 - 3,0123 \cdot 3,0 \le \beta_2 \le -9,6 + 3,0123 \cdot 3,0$, или $\beta_2 \in (-18,64; -0,56)$. Этот доверительный интервал накрывает истинное значение параметра β_2 с надежностью, равной 0,99.

Задача 3.2. По некоторым переменным имеются следующие статистические данные: $\bar{y} = 15,0;$ $\bar{x}_1 = 6,5;$ $\bar{x}_2 = 12,0;$ $\sigma_y = 4,0;$ $\sigma_{x_1} = 2,5;$ $\sigma_{x_2} = 3,5;$ $r_{yx_1} = 0,63;$ $r_{yx_2} = 0,78;$ $r_{x_1x_2} = 0,52.$

Построить уравнение регрессии в стандартизованном и натуральном масштабах.

Решение: Поскольку изначально известны коэффициенты парной корреляции между переменными, начать следует с построения уравнения регрессии в стандартизованном масштабе. Для этого надо решить систему нормальных уравнений (14), которая в случае двух факторов имеет вид:

$$\begin{cases} \beta_1 + r_{x_1x_2}\beta_2 = r_{yx_1}, \\ r_{x_1x_2}\beta_1 + \beta_2 = r_{yx_2}, \end{cases}$$

или, после подстановки исходных данных:

$$\begin{cases} \beta_1 + 0,52\beta_2 = 0,63, \\ 0,52\beta_1 + \beta_2 = 0,78. \end{cases}$$

Решаем эту систему любым способом, получаем: $\beta_1 = 0,3076$, $\beta_2 = 0,62$.

Запишем уравнение регрессии в стандартизованном масштабе:

$$\hat{t}_y = 0,3076 t_{x_1} + 0,62 t_{x_2}.$$

Теперь перейдем к уравнению регрессии в натуральном масштабе, для чего используем формулы:

$$b_1 = 0,3076 \cdot \frac{4,0}{2,5} = 0,4922; \quad b_2 = 0,62 \cdot \frac{4,0}{3,5} = 0,7086;$$

$$a = 15 - 0,4922 \cdot 6,5 - 0,7086 \cdot 12 = 3,298.$$

Уравнение регрессии в натуральном масштабе имеет вид:

$$\hat{y} = 3,298 + 0,4922 x_1 + 0,7086 x_2.$$

Задача 3.3. Уравнение регрессии в стандартизованных переменных выглядит так:

$$t_y = -0,82 t_{x_1} + 0,65 t_{x_2} - 0,43 t_{x3}.$$

При этом вариации всех переменных равны следующим величинам:

$$V_y = 32\%; V_{x1} = 38\%; V_{x2} = 43\%; V_{x3} = 35\%.$$

Сравнить факторы по степени влияния на результирующий признак и определить значения частных коэффициентов эластичности.

Решение: Стандартизованные уравнения регрессии позволяют сравнивать факторы по силе их влияния на результат. При этом, чем больше по абсолютной величине коэффициент при стандартизованной переменной, тем сильнее данный фактор влияет на результирующий признак. В рассматриваемом уравнении самое сильное воздействие на результат оказывает фактор x_1, имеющий коэффициент – 0,82, самое слабое – фактор x_3 с коэффициентом, равным – 0,43.

В линейной модели множественной регрессии обобщающий (средний) коэффициент частной эластичности определяется выражением, в которое входят средние значения переменных и коэффициент при соответствующем факторе уравнения регрессии натурального масштаба. В условиях задачи эти величины не заданы. Поэтому воспользуемся выражениями для вариации по переменным:

$$V_y = \frac{\sigma_y}{|\bar{y}|} \cdot 100\%; \quad V_{x_j} = \frac{\sigma_y}{|\bar{x}_j|} \cdot 100\%.$$

Коэффициенты b_j связаны со стандартизованными коэффициентами β_j соотношением (15), которое подставим в формулу:

$$\bar{Э}_j = b_j \frac{\bar{x}_j}{\bar{y}} = \beta_j \frac{\sigma_y}{\sigma_{x_j}} \cdot \frac{\bar{x}_j}{\bar{y}} = \beta_j \frac{V_y}{V_{x_j}}.$$

При этом знак коэффициента эластичности будет совпадать со знаком β_j:

$$\bar{Э}_1 = -0,82 \cdot \frac{32}{38} = -0,691; \quad \bar{Э}_2 = 0,65 \cdot \frac{32}{43} = 0,484; \quad \bar{Э}_3 = -0,43 \cdot \frac{32}{35} = -0,393.$$

Задача 3.4. По 32 наблюдениям получены следующие данные:

$$\hat{y} = a + 1{,}864x_1 - 2{,}56x_2 + 2{,}86x_3; \quad R^2 = 0{,}58;$$

$$\bar{y} = 110; \bar{x}_1 = 80; \bar{x}_2 = 140; \bar{x}_3 = 130.$$

Определить значения скорректированного коэффициента детерминации, частных коэффициентов эластичности и параметра a.

Решение: Значение скорректированного коэффициента детерминации определим по одному из равенств:

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} = 1 - (1 - 0{,}58)\frac{32-1}{32-3-1} = 0{,}535.$$

Частные коэффициенты эластичности (средние по совокупности) вычисляем по формулам:

$$\bar{Э}_1 = 1,864 \cdot \frac{80}{110} = 1,356; \quad \bar{Э}_2 = -2,56 \cdot \frac{140}{110} = -3,258; \quad \bar{Э}_3 = 2,86 \cdot \frac{130}{110} = 3,38.$$

Поскольку линейное уравнение множественной регрессии выполняется при подстановке в него средних значений всех переменных, определяем параметр a :

$$a = 110 - 1{,}864 \cdot 80 + 2{,}56 \cdot 140 - 2{,}86 \cdot 130 = -52{,}52.$$

Задача 3.5. По совокупности 30 предприятий концерна изучается зависимость прибыли y (млн. руб.) от выработки продукции на одного работника x_1 (ед.) и индекса цен на продукцию x_2 (%). Данные приведены в таблице:

Признак	Среднее значение	Среднее квадратическое отклонение	Линейный коэффициент парной корреляции
y	250	38	ryx1=0,68
x1	54	12	ryx2=0,63
x2	112	21	rx1x2=0,42

Постройте уравнение множественной регрессии в стандартизованной и натуральной форме. Определите показатели частной и множественной корреляции. Найдите частные коэффициенты эластичности и сравните их с β-коэффициентами.

Решение:

Линейное уравнение множественной регрессии y от x_1 и x_2 имеет вид: $y=a+b_1*x_1+b_2*x_2+e$. Для расчета его параметров применим метод стандартизации переменных и построим искомое уравнение в стандартизованном масштабе: $t_y=\beta_1*t_{x1}+\beta_2*t_{x2}$.

Расчет β-коэффициентов выполним по формулам:

$$\beta_1 = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{1 - r^2{}_{x1x2}} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{1 - 0{,}42^2} = \frac{0{,}4154}{0{,}8236} = 0{,}5044.$$

$$\beta_2 = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{1 - r^2{}_{x1x2}} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{1 - 0{,}42^2} = \frac{0{,}3444}{0{,}8236} = 0{,}4182.$$

Получим уравнение $t_y = 0{,}5044 t_{x1} + 0{,}4182 t_{x2}$.

Для построения уравнения в натуральной форме рассчитаем b_1 и b_2, используя формулы для перехода от βi к bi:

$$\beta_i = b_i \cdot \frac{\sigma_{xi}}{\sigma_y}; b_i = \beta_i \cdot \frac{\sigma_y}{\sigma_{xi}}$$

$$b_1 = \frac{0{,}5044 \cdot 38}{12} = 1{,}5973; b_2 = \frac{0{,}4182 \cdot 38}{21} = 0{,}7567.$$

Значение a определим из соотношения:

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2 = 250 - 1{,}5973 \cdot 54 - 0{,}7567 \cdot 112 = 250 - 86{,}2542 - 84{,}7504 = 78{,}9954.$$

Получим уравнение: $y=79{,}00+1{,}60*x_1+0{,}76*x_2+e$.

Линейные коэффициенты частной корреляции рассчитываются по рекуррентной формуле:

$$r_{yx1\cdot x2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{\sqrt{(1 - r^2_{yx2})(1 - r^2_{x1x2})}} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{\sqrt{(1 - 0{,}63^2)(1 - 0{,}42^2)}} = \frac{0{,}4154}{\sqrt{0{,}4967}} = 0{,}5894;$$

$$r_{yx2\cdot x1} = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{\sqrt{(1 - r^2_{yx1})(1 - r^2_{x1x2})}} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{\sqrt{(1 - 0{,}68^2)(1 - 0{,}42^2)}} = \frac{0{,}3444}{\sqrt{0{,}4428}} = 0{,}5176;$$

$$r_{x1x2\cdot y} = \frac{r_{x1x2} - r_{yx1} \cdot r_{yx2}}{\sqrt{(1 - r^2_{yx1})(1 - r^2_{yx2})}} = \frac{0{,}42 - 0{,}68 \cdot 0{,}63}{\sqrt{(1 - 0{,}68^2)(1 - 0{,}63^2)}} = \frac{-0{,}0084}{\sqrt{0{,}3242}} = -0{,}0148.$$

Если сравнить значения коэффициентов парной и частной корреляции, то приходим к выводу, что из-за умеренной межфакторной связи (r_{x1x2}=0,42) коэффициенты парной корреляции оказались завышены.

Расчет линейного коэффициента множественной корреляции выполним по формуле:

$$R_{yx1x2} = \sqrt{r_{yx1} \cdot \beta_1 + r_{yx2} \cdot \beta_2} = \sqrt{0{,}68 \cdot 0{,}5044 + 0{,}63 \cdot 0{,}4182} = \sqrt{0{,}6065} = 0{,}7788.$$

Зависимость y от x_1 и x_2 характеризуется как тесная, в которой 61% вариации прибыли определяются вариацией выработки продукции и индекса цен на продукцию. Прочие факторы, не включенные в модель, определяют соответственно 39% от общей вариации y.

Общий F-критерий проверяет гипотезу о статистической значимости уравнения регрессии и показателя тесноты связи:

$$F = \frac{R^2_{yx1x2}}{1 - R^2_{yx1x2}} : \frac{m}{n - m - 1} = \frac{0{,}6065}{1 - 0{,}6065} \cdot \frac{27}{2} = 20{,}82.$$

$$F_{\alpha,v1,v2} = 3{,}35.$$

Сравнивая 20,82>3,35, с надежностью 1-α=0,95 делаем заключение о статистической значимости уравнения в целом и показателя тесноты связи R_{yx1x2}, которые сформировались под неслучайным воздействием факторов x_1 и x_2.

3. Для характеристики относительной силы влияния x_1 и x_2 на y рассчитаем частные коэффициенты эластичности.

$$Э_{yxj} = b_j \cdot \frac{\bar{x}_j}{\bar{y}};$$

$$Э_{yx1} = \frac{1{,}5973 \cdot 54}{250} = 0{,}3450\%$$

$$Э_{yx2} = \frac{0{,}7567 \cdot 112}{250} = 0{,}3390\%.$$

С увеличением выработки продукции одним работником x_1 на 1% от ее среднего уровня прибыль y в среднем возрастает на 0, 35% от своего среднего уровня; при повышении индекса цен x_2 на 1% прибыль y в среднем возрастает на 0,34% от своего среднего уровня. Очевидно, что сила влияния выработки на размер прибыли немного выше, чем индекса цен. К аналогичным выводам о силе влияния факторов приходим при сравнении модулей значений β_1 и β_2.

Задача 3.6. При построении линейной множественной регрессии $y = f(x_1, \ldots, x_{10})$ по 48 измерениям коэффициент детерминации составил 0,578. После исключения факторов x_3, x_7 и x_8 коэффициент детерминации уменьшился до 0,495.

Обоснованно ли было принятое решение об изменении состава влияющих переменных на уровнях значимости 0,1, 0,05 и 0,01?

Решение: Пусть R_1^2 - коэффициент детерминации уравнения регрессии при первоначальном наборе факторов, R_2^2 - коэффициент детерминации после исключения трех факторов. Выдвигаем гипотезы:

$$\mathrm{H}_0 : R_1^2 - R_2^2 = 0;$$

$$\mathrm{H}_1 : R_1^2 - R_2^2 > 0$$

Основная гипотеза предполагает, что уменьшение величины R^2 было несущественным, и решение об исключении группы факторов было правильным. Альтернативная гипотеза говорит о правильности принятого решения об исключении.

Для проверки нулевой гипотезы используем следующую статистику:

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - p - 1}{k},$$

где $n = 48$, $p = 10$ – первоначальное количество факторов, $k = 3$ – количество исключаемых факторов. Тогда

$$F_{набл} = \frac{0{,}578 - 0{,}495}{1 - 0{,}578} \cdot \frac{48 - 10 - 1}{3} = 2{,}426$$

Сравним полученное значение с критическим $F(\alpha; 3; 39)$ на уровнях 0,1; 0,05 и 0,01:

$F(0{,}1; 3; 37) = 2{,}238$;

$F(0{,}05; 3; 37) = 2{,}86$;

$F(0{,}01; 3; 37) = 4{,}36$.

На уровне $\alpha = 0{,}1$ $F_{набл} > F_{кр}$, нулевая гипотеза отвергается, исключение данной группы факторов не оправдано, на уровнях 0,05 0,01 нулевая гипотеза не может быть отвергнута, и исключение факторов можно считать оправданным.

Упражнения с пояснениями

Имеется выборка по 15 филиалам торговой компании в 2017 году: объем реализации продукции (***Y***, ден.ед.), цена за единицу продукции (***X1***, ден.ед.), расходы на рекламу (***X2***, ден.ед.).

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	23	18	27	29	43	23	55	47	35	38	14	51	20	39	35
X1	37	33	15	36	26	24	15	33	44	34	63	8	44	43	31
X2	39	40	35	48	53	42	54	54	50	53	46	50	43	55	51

Упражнение 3.1. Используя исходные данные из таблицы построить матрицу линейных коэффициентов парной корреляции.

Решение. Чтобы провести корреляционный анализ, из расчетной таблицы 1 необходимо рядом скопировать столбцы Y, X1, X2, а затем применить инструмент **КОРРЕЛЯЦИЯ** надстройки **Пакет анализа**. Нажав на кнопку Анализ данных, следует выбрать опцию Корреляция (рис.1).

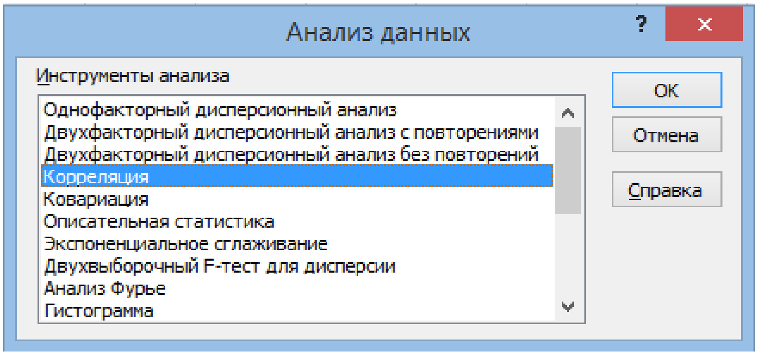


Рис. 1. Выбор корреляционного анализа в Excel

На экране появится новое окно с опциями корреляционного анализа (рис.2).

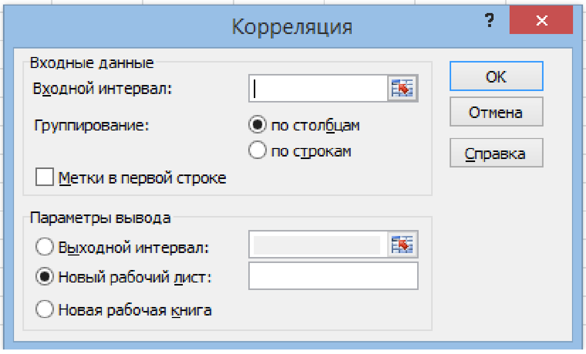


Рис.2. Окно параметров корреляции в Excel

В поле *Входной интервал* необходимо выбрать таблицу со всеми исходными для корреляционного анализа. Опция *«Группирование»* дает возможность группировать переменные в таблице по столбцам (как принято традиционно) или по строкам. Опция *Метки,* если в ней указать «галочку», трактует первую строку данных как названия переменных. В разделе *Параметры вывода* можно устанавливать, куда выводятся результаты корреляции (обычно выводят результаты на этот же лист, так как удобно держать данные и результаты корреляции на одном листе). Чтобы оценить корреляцию в Excel во входном интервале укажем диапазон переменных Y, X1, X2, укажем «галочку» в опции Метки, чтобы трактовать первую строку данных как названия переменных (рис.3).

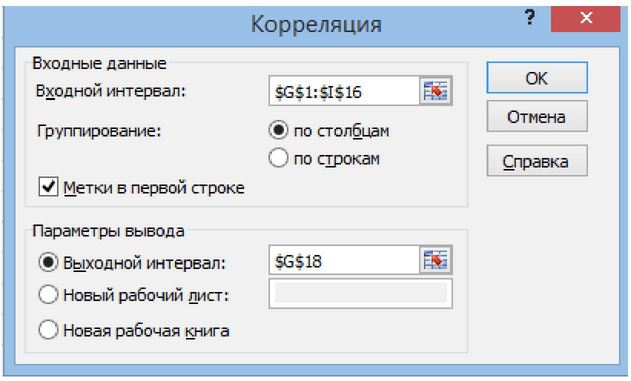


Рис. 3. Окно параметров корреляции с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.4).

	A	B	C	D	E	F	G	H	I	J
1	i	Y	X0	X1	X2		Y	X1	X2	
2	1	23	1	37	39		23	37	39	
3	2	18	1	33	40		18	33	40	
4	3	27	1	15	35		27	15	35	
5	4	29	1	36	48		29	36	48	
6	5	43	1	26	53		43	26	53	
7	6	23	1	24	42		23	24	42	
8	7	55	1	15	54		55	15	54	
9	8	47	1	33	54		47	33	54	
10	9	35	1	44	50		35	44	50	
11	10	38	1	34	53		38	34	53	
12	11	14	1	63	46		14	63	46	
13	12	51	1	8	50		51	8	50	
14	13	20	1	44	43		20	44	43	
15	14	39	1	43	55		39	43	55	
16	15	35	1	31	51		35	31	51	
17	Среднее	33,13333		32,4	47,53333					
18	СКО	12,05469		13,34066	6,130434			*Y*	*X1*	*X2*
19							Y	1		
20							X1	-0,57987	1	
21							X2	0,749597	0,025107	1

Рис. 4. Вывод результатов оценивания корреляции в Excel

В столбце Y на пересечении со строками X1, X2 указаны линейные коэффициенты парной корреляции каждого регрессора (X1, X2) с регрессантом (Y). Наличие умеренной и тесной статистической взаимосвязи говорит о целесообразности включения регрессанта во множественную регрессию.

Комментарий: При наличии слабой взаимосвязи регрессанта с регрессором или ее практическом отсутствии коэффициент регрессии при таком регрессанте скорее всего будет незначим. Обычно из множественной регрессии исключают регрессоры (факторы), коэффициенты при которых незначимы, чтобы не терять эффективность оценок, но незначимую константу принято оставлять, чтобы избежать смещения оценок и иметь возможность интерпретировать R^2.

Вывод 1: Согласно шкале Чеддока-Снедекора наблюдается умеренная обратная статистическая взаимосвязь между объемом продаж и ценой (R_{yx1}= -0,58), а также тесная прямая статистическая взаимосвязь между объемом продаж и расходами на рекламу (R_{yx2}= 0,75). В столбцах X1, X2 на пересечении со строками X1, X2 указан линейный коэффициент парной корреляции между регрессорами (X1, X2) – межфакторная корреляция. Согласно шкале Чеддока-Снедекора практически отсутствует статистическая взаимосвязь между факторами (R_{x1x2} = 0,03), что показывает о соблюдении одного из требований к построению множественной регрессии: в уравнение включаются регрессоры, которые между собой не взаимосвязаны.

Упражнение 3.2. Сформировать расчетную таблицу заданной структуры, чтобы «вручную» матрично-векторным способом определить МНК-оценки коэффициентов линейного уравнения множественной регрессии $\hat{y} = a + b_1 x_1 + b_2 x_2$.

Решение. Чтобы «вручную» матрично-векторным способом определить МНК-оценки коэффициентов линейного уравнения множественной регрессии достаточно заполнить столбцы y, x0, x1, x2 в расчетной табл. 1.

Табл. 1

Подготовка данных для оценивания линейной регрессии с двумя регрессорами

i	y	x0	x1	x2	$\hat{y}$	e	e^2	(y-$y_{сред}$)^2	($\hat{y}$ - $y_{сред}$)^2
1	23	1	37	39	17,81	5,19	26,91	102,68	234,72
2	18	1	33	40	21,48	-3,48	12,12	229,02	135,76
3	27	1	15	35	23,71	3,29	10,84	37,62	88,84
4	29	1	36	48	31,89	-2,89	8,33	17,08	1,56

5	43	1	26	53	44,82	-1,82	3,30	97,35	136,51
6	23	1	24	42	29,36	-6,36	40,46	102,68	14,23
7	55	1	15	54	52,28	2,72	7,42	478,15	366,42
8	47	1	33	54	42,53	4,47	19,97	192,28	88,33
9	35	1	44	50	30,56	4,44	19,69	3,48	6,61
10	38	1	34	53	40,49	-2,49	6,18	23,68	54,07
11	14	1	63	46	14,26	-0,26	0,07	366,08	356,07
12	51	1	8	50	50,05	0,95	0,90	319,22	286,18
13	20	1	44	43	20,04	-0,04	0,00	172,48	171,49
14	39	1	43	55	38,62	0,38	0,14	34,42	30,12
15	35	1	31	51	39,10	-4,10	16,84	3,48	35,64
Среднее	33,13		32,4	47,53		Сумма	173,17	2179,73	2006,56
СКО	12,05		13,34	6,13					

Суммы и средние по столбцам в расчетной табл. 1 необходимо определить с помощью функций *СУММ(…) и СРЗНАЧ(…).*

Чтобы оценить параметры линейного уравнения множественной регрессии $\hat{y} = a + b_1 x_1 + b_2 x_2$ «вручную» матрично-векторным способом, необходимо пошагово применить формулу $\boldsymbol{B=(X'X)^{-1} X'Y}$, используя встроенные в Excel функции для операций с матрицами.

Шаг 1: получение матрицы ***X'*** путем транспонирования структурной матрицы X (диапазон переменных X0, X1, X2). Под расчетной таблицей 1 выделяем пустой массив ячеек размера 3 строки и 15 столбцов. Затем в Главном меню MS Excel выберем: Формулы – Вставить функцию – Ссылки и массивы – *ТРАНСП (…).* В окне параметров функции *ТРАНСП (…)* в поле *Массив* необходимо указать диапазон переменных (X0, X1, X2), которые надо транспонировать. Далее, нажав на кнопку Ок, важно правильно раскрыть результат транспонирования. Для этого необходимо нажать на клавишу F2 (или в Excel поместить курсор мыши в командную строку формул), а затем — на сочетание клавиш CTRL + SHIFT + ВВОД.

Шаг 2: вычисление произведения матриц ***X'X***. Под расчетной табл. 1 выделяем пустой массив ячеек размера 3 строки и 3 столбца. Затем в Главном меню MS Excel выберем: Формулы – Вставить функцию – Математические – *МУМНОЖ* (…).В окне параметров функции *МУМНОЖ (…)* в поле *Массив 1* необходимо указать диапазон матрицы ***X'***, в поле *Массив 2* необходимо указать диапазон матрицы ***X*** (переменные X0, X1, X2). Далее, нажав на кнопку Ок, важно правильно раскрыть результат произведения матриц. Для этого необходимо нажать на клавишу F2 (или в Excel поместить курсор мыши в командную строку формул), а затем — на сочетание клавиш CTRL + SHIFT + ВВОД.

Шаг 3: вычисление произведения ***X'Y.*** Под расчетной табл. 1 выделяем пустой массив ячеек размера 3 строки и 1 столбец. Затем в Главном меню MS Excel выберем: Формулы – Вставить функцию – Математические – *МУМНОЖ* (…).В окне параметров функции *МУМНОЖ (…)* в поле *Массив 1* необходимо указать диапазон матрицы ***X'***, в поле *Массив2* необходимо указать диапазон столбца ***Y***. Далее, нажав на кнопку Ок, важно правильно раскрыть результат произведения. Для этого необходимо нажать на клавишу F2 (или в Excel поместить курсор мыши в командную строку формул), а затем — на сочетание клавиш CTRL + SHIFT + ВВОД.

Шаг 4: вычисление обратной матрицы $\boldsymbol{(X'X)^{-1}}$. Под расчетной табл. 1 выделяем пустой массив ячеек размера 3 строки и 3 столбца. Затем в Главном меню MS Excel выберем: Формулы – Вставить функцию – Математические – *МОБР* (…). В окне параметров функции *МОБР (…)* в поле *Массив* необходимо указать диапазон произведения матриц X'X, полученного на 2 шаге. Далее, нажав на кнопку Ок, важно правильно раскрыть результат обратной матрицы. Для этого необходимо нажать на клавишу F2 (или в Excel поместить курсор мыши в командную строку формул), а затем — на сочетание клавиш CTRL + SHIFT + ВВОД.

Шаг 5: определение вектора коэффициентов ***B* = *(X'X)*$^{-1}$ *X'Y***. Под расчетной таблицей 1 выделяем пустой массив ячеек размера 3 строки и 1 столбец. Затем в Главном меню MS Excel выберем: Формулы – Вставить функцию – Математические – *МУМНОЖ* (…).В окне параметров функции *МУМНОЖ (…)* в поле *Массив 1* необходимо указать диапазон обратной матрицы ***(X'X)***$^{-1}$, в поле *Массив 2* необходимо указать диапазон произведения ***X'Y***. Далее, нажав на кнопку Ок, важно правильно раскрыть результат произведения. Для этого необходимо нажать на клавишу F2 (или в Excel поместить курсор мыши в командную строку формул), а затем — на сочетание клавиш CTRL + SHIFT + ВВОД. Результат «ручной» реализации матрично-векторного способа представлен на рисунке 5.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	i	Y	X0	X1	X2		Y	X1	X2						
2	1	23	1	37	39		23	37	39						
3	2	18	1	33	40		18	33	40						
4	3	27	1	15	35		27	15	35						
5	4	29	1	36	48		29	36	48						
6	5	43	1	26	53		43	26	53						
7	6	23	1	24	42		23	24	42						
8	7	55	1	15	54		55	15	54						
9	8	47	1	33	54		47	33	54						
10	9	35	1	44	50		35	44	50						
11	10	38	1	34	53		38	34	53						
12	11	14	1	63	46		14	63	46						
13	12	51	1	8	50		51	8	50						
14	13	20	1	44	43		20	44	43						
15	14	39	1	43	55		39	43	55						
16	15	35	1	31	51		35	31	51						
17	Среднее	33,13333		32,4	47,53333										
18	СКО	12,05469		13,34066	6,130434			**Y**	**X1**	**X2**					
19							**Y**	**1**							
20							**X1**	**-0,57987**	**1**						
21							**X2**	**0,749597**	**0,025107**	**1**					
22	X`														
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	37	33	15	36	26	24	15	33	44	34	63	8	44	43	31
25	39	40	35	48	53	42	54	54	50	53	46	50	43	55	51
26															
27	X`X				X`Y		(X`X)-1				B				
28	15	486	713		497		4,407547	-0,01117	-0,08371	a	-20,797				
29	486	18416	23132		14704		-0,01117	0,000375	-2E-05	b1	-0,54132				
30	713	23132	34455		24455		-0,08371	-2E-05	0,001775	b2	1,503558				
31															

Рис. 5. МНК-оценки множественной регрессии, полученные «вручную»

Экономическая интерпретация МНК-оценок. Запишем линейное уравнение множественной регрессии

$$\hat{y} = -20,797 - 0,541x_1 + 1,504x_2 .$$

Вывод 2: Полученные МНК-оценки коэффициентов можно интерпретировать следующим образом: при увеличении цены на одну денежную единицу объем продаж при прочих равных условиях снизится в среднем на 0,541 денежных единиц, если расходы на рекламу возрастают на 1 ден. единицу, то объем продаж увеличивается в среднем на 1,503 ден. единиц при прочих равных условиях.

Упражнение 3.3. Вычислить «вручную» предсказанные линейным уравнением множественной регрессии значения объема продаж - $\hat{y}$, остатки регрессии и их квадраты.

Решение. Чтобы вычислить предсказанные значения $\hat{y}$ в расчетной табл. 1, необходимо в уравнение множественной регрессии вместо x_1 и x_2 последовательно, начиная с первого, подставить исходные наблюдаемые значения переменных *x1* и *x2* из расчетной табл. 1. Полученные предсказанные значения $\hat{y}$ указать в соответствующем столбце расчетной табл. 1.

Остаток регрессии – это ошибка, несвязка (*discrepancy*) между наблюдаемым значением зависимой переменной y_i и предсказанным значением $\hat{y}_i$. Чтобы вычислить остатки регрессии и их квадраты в расчетной табл. 1 необходимо заполнить столбцы: e = (y- $\hat{y}$), e^2 = (y- $\hat{y}$)^2.

Необходимо заполнить в табл. 1 столбцы (y-$y_{сред}$)^2, ($\hat{y}$ -$y_{сред}$)^2. Чтобы вычислить суммы квадратов отклонений (TSS, ESS, RSS), необходимо в расчетной табл. 1 применить функцию *СУММ(...)* к столбцам (y-$y_{сред}$)^2, ($\hat{y}$ -$y_{сред}$)^2, (y- $\hat{y}$)^2.

Упражнение 3.4. Определить «вручную» значения коэффициента множественной детерминации, индекса множественной корреляции, скорректированного коэффициента детерминации, средних коэффициентов эластичности. Дать количественную интерпретацию полученным значениям показателей.

Решение. Чтобы рассчитать значения коэффициента множественной детерминации и индекса множественной корреляции, скорректированного коэффициента детерминации, необходимо «вручную» ниже расчетной таблицы 1 использовать ссылки на необходимые ячейки согласно формулам:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum\left(y_i - \bar{y}\right)^2} = 1 - \frac{173,18}{2179,73} = 0,921$$

$$R = \sqrt{R^2} = \sqrt{0,921} = 0,959$$

$$R_{adj}^2 = \bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1} = 1 - (1 - 0,921)\frac{15-1}{15-2-1} = 0,907$$

Вывод 3: Коэффициент множественной детерминации R^2 показывает, что данная множественная регрессия объясняет 92% вариации объема продаж. Индекс множественной корреляции отражает наличие тесного совместного влияния цены и расходов на рекламу на регрессант - объем продаж. Скорректированный коэффициент детерминации R^2_{adj} всегда немного меньше обычного коэффициента детерминации.

Комментарий: Отметим, что регрессии с разными зависимыми переменными ни по R^2, ни по R^2_{adj} сравнивать нельзя (из-за разных выборочных дисперсий зависимых показателей). Коэффициент R^2_{adj} является одним из показателей, с помощью которых можно сравнивать регрессии с одинаковыми зависимыми переменными и разным набором независимых. Обычно лучшей считается та модель, в которой R^2_{adj} наибольший. Однако при этом не следует увлекаться увеличением этого показателя в ущерб возможности дать модели осмысленную экономическую интерпретацию.

Чтобы рассчитать значения средних коэффициентов эластичности для каждого регрессора, необходимо «вручную» ниже расчетной табл. 1 использовать ссылки на необходимые ячейки согласно формулам:

$$\bar{Э}_1 = b_1\frac{\bar{x}_1}{\bar{y}} = -0,541\frac{32,4}{33,13} = -0,53$$

$$\bar{Э}_2 = b_2\frac{\bar{x}_2}{\bar{y}} = 1,50\frac{47,53}{33,13} = 2,16$$

Вывод 4: Увеличение цены на 1 % приводит в среднем к снижению объема продаж на 0,53%, а увеличение расходов на рекламу на 1% приводит к увеличению объема продаж в среднем на 2,16%. Следовательно, расходы на рекламу сильнее, чем цена, влияют на изменение объема продаж.

Упражнение 3.5. С помощью встроенного инструмента **Регрессия** в Пакете анализа MS Excel определить МНК-оценки коэффициентов линейного уравнения множественной регрессии $\widehat{y} = a + b_1x_1 + b_2x_2$. Проверить качество модели по результатам инструмента Регрессия надстройки Пакет анализа.

Решение. Чтобы оценить нашу регрессию в Excel с помощью инструмента **Регрессия** надстройки **Пакет анализа** в главном меню программы Excel на вкладке *Данные* (в верхней строке) выберем опцию *Анализ данных*, а в списке *Инструментов анализа* выберем инструмент *Регрессия*. В качестве регрессанта выберем переменную *y* – объем продаж, а в качестве регрессоров укажем диапазон переменных *x1* – цена, *x2* – расходы на рекламу, укажем «галочку» в опции Метки, чтобы трактовать первую строку данных как названия переменных (рис.6).

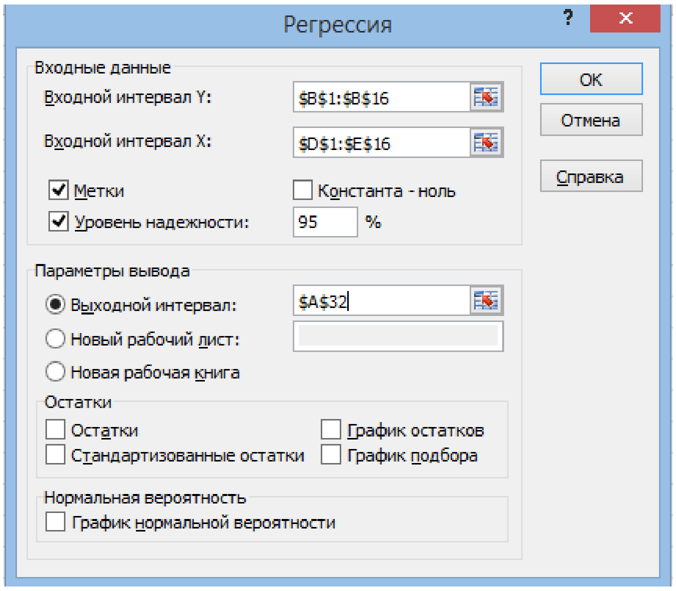
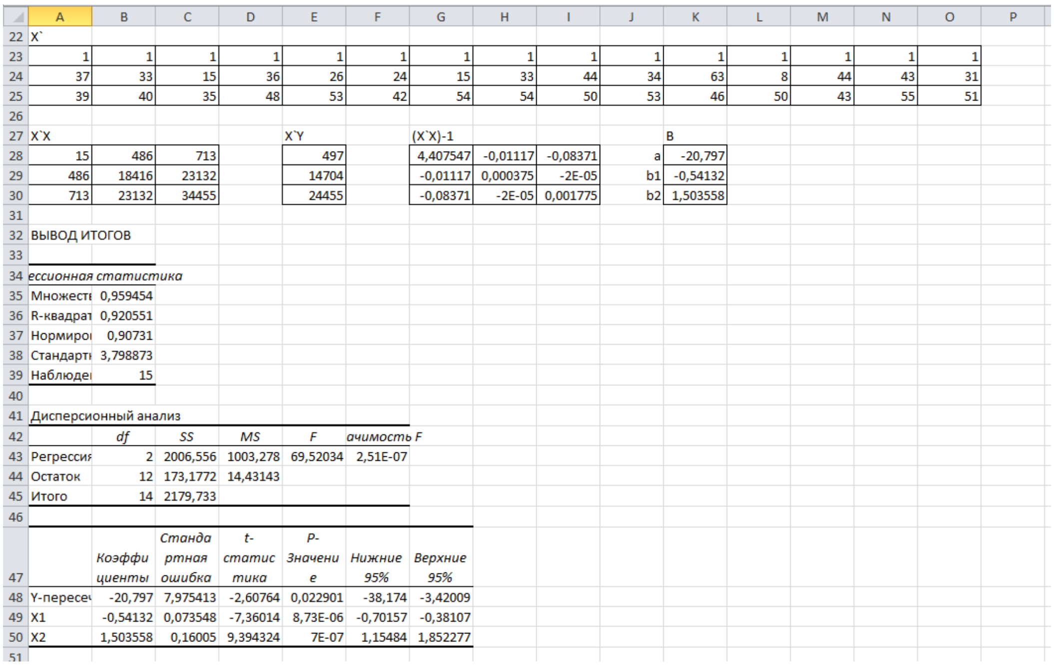


Рис. 6. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.7).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
22	X`														
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	37	33	15	36	26	24	15	33	44	34	63	8	44	43	31
25	39	40	35	48	53	42	54	54	50	53	46	50	43	55	51
26															
27	X`X				X`Y		(X`X)-1				B				
28	15	486	713		497		4,407547	-0,01117	-0,08371	a	-20,797				
29	486	18416	23132		14704		-0,01117	0,000375	-2E-05	b1	-0,54132				
30	713	23132	34455		24455		-0,08371	-2E-05	0,001775	b2	1,503558				
31															
32	ВЫВОД ИТОГОВ														
33															
34	ессионная статистика														
35	Множести	0,959454													
36	R-квадрат	0,920551													
37	Нормиро	0,90731													
38	Стандарт	3,798873													
39	Наблюде	15													
40															
41	Дисперсионный анализ														
42		df	SS	MS	F	ачимость F									
43	Регрессия	2	2006,556	1003,278	69,52034	2,51E-07									
44	Остаток	12	173,1772	14,43143											
45	Итого	14	2179,733												
46															
47		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%								
48	Y-пересеч	-20,797	7,975413	-2,60764	0,022901	-38,174	-3,42009								
49	X1	-0,54132	0,073548	-7,36014	8,73E-06	-0,70157	-0,38107								
50	X2	1,503558	0,16005	9,394324	7E-07	1,15484	1,852277								

Рис.7. Вывод результатов оценивания регрессии в Excel

Более подробно рассмотрим результаты, переведя их в обычные таблицы.

Регрессионная статистика	
Множественный R	0,959454
R-квадрат	0,920551
Нормированный R-квадрат	0,90731
Стандартная ошибка	3,798873
Наблюдения	15
Дисперсионный анализ	

	df	SS	MS	F	Значимость F	
Регрессия	2	2006,556	1003,278	69,52034	2,51E-07	
Остаток	12	173,1772	14,43143			
Итого	14	2179,733				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	-20,797	7,975413	-2,60764	0,022901	-38,174	-3,42009
X1	-0,54132	0,073548	-7,36014	8,73E-06	-0,70157	-0,38107
X2	1,503558	0,16005	9,394324	7E-07	1,15484	1,852277

В подтаблице «Регрессионная статистика» *Множественный R* – это индекс корреляции между регрессантом и регрессорами. Строкой ниже Excel выдает значение R^2, равное 0,92. Чуть ниже можно увидеть значение *нормированного* (или, по-другому, скорректированного (adjusted) R^2. Еще ниже указано значение *Стандартной ошибки* регрессии (MSE) – квадратного корня из остаточной дисперсии на одну степень свободы (пересечение столбца *MS* и строки *Остаток* по второй подтаблице), равное 3,799. В конце подтаблицы дано значение количества Наблюдений, равное 15.

В подтаблице «Дисперсионный анализ» во втором столбце df даны числа степеней свободы, в третьем столбце даны значения ESS, RSS и TSS (соответственно 2006,556;173,1772; 2179,733). Четвертый столбик содержит отношения ESS/df_{ESS}, RSS/df_{RSS} соответственно. *F* – это значение F-статистики теста Фишера для проверки адекватности регрессии. В самом правом столбце дано значение *p-value* для этой статистики.

Комментарий: В модели парной регрессии, для того чтобы определить, есть ли линейная зависимость между переменными X и Y, достаточно проверить значимость коэффициента регрессии при переменной X по тесту Стьюдента. Однако в случае множественной регрессии дело обстоит сложнее. Здесь возникает вопрос о совокупном влиянии нескольких переменных. Эта задача не сводится к проверке значимости каждого из коэффициентов регрессии. Каждый из коэффициентов регрессии в отдельности может быть незначим, но тем не менее включенные в уравнение регрессии регрессоры (факторы) в совокупности могут оказывать значимое влияние на зависимую переменную – регрессант. Для модели $Y = \beta_0 + \beta_1 X_1 + ... + \beta_2 X_2 + \varepsilon$ проверка гипотезы $H_0 : \beta_1 = \beta_2 = ... = \beta_m = 0$ при альтернативной гипотезе: $H_1 : \exists \beta_j \neq 0, j = 1,...,m$, то есть когда хотя бы один коэффициент отличен от нуля, называется проверкой гипотезы об адекватности регрессии (статистической значимости в целом). Гипотеза об адекватности проверяется всеми статистическими пакетами. Выдается соответствующее значение F-статистики (тест Фишера) и *p-value* для этой F-статистики.

Повторимся, что значение *p-value* позволяет провести тест (в данном случае тест Фишера), не пользуясь критическими значениями статистики. Если *p-value* для статистики меньше, чем α =0,1, то нулевая гипотеза отвергается с надежностью 90%. Если *p-value* для статистики меньше, чем α =0,05, то нулевая гипотеза отвергается с надежностью 95%. Если *p-value* для статистики меньше, чем α =0,01, то нулевая гипотеза отвергается с надежностью 99%.

Вывод 5: В нашем случае *p-value* для статистики Фишера составило 2,51E-07, что меньше, чем α =0,01. Это означает, что с надежностью 99% отвергается нулевая гипотеза о неадекватности уравнения регрессии. Согласно тесту Фишера регрессия адекватна, выбранный набор независимых переменных (X1 – цена, X2 – расходы на рекламу) оказывает линейное влияние на зависимую переменную Y – объем продаж.

В нижней подтаблице в столбце «Коэффициенты» содержатся *коэффициенты регрессии для регрессоров X1 и X2*, равные -0,541 и 1,503 соответственно, а также *свободный член*, равный -20,797 на пересечении со строкой «Y-пересечение». В третьем столбце указаны стандартные ошибки коэффициентов, далее – значение t-статистики теста Стьюдента, затем дано значение *p-value* для этой статистики. Гипотезы о значимости коэффициента регрессии (тест Стьюдента) формулируются и проверяются для каждого коэффициента регрессии в отдельности аналогично случаю парной регрессии. Единственное отличие состоит в числе степеней свободы:

$$H_0 : \beta_j = 0;$$
$$H_1 : \beta_j \neq 0.$$

t-статистика и *p-value* для t -статистики Стьюдента для проверки гипотезы о значимости каждого из коэффициентов множественной регрессии выдаются статистическими пакетами.

Вывод 6: В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии при X1 составило 8,73E-06, что меньше, чем α =0,01. Это означает, что коэффициент регрессии β_1 является значимым с надежностью 99%, между переменными X1 (цена) и Y (объем продаж) существует значимая линейная связь. Также *p-value* для статистики Стьюдента для коэффициента регрессии при X2 составило 7E-07, что меньше, чем α =0,01. Это означает, что коэффициент регрессии β_2 является значимым, между переменными X2 (расходы на рекламу) и Y (объем продаж) существует значимая линейная связь. *p-value* для статистики Стьюдента для свободного коэффициента составило 0,023, что меньше, чем α =0,05. Это означает, что свободный член α является значимым с надежностью 95%. В двух самых правых столбцах даны доверительные границы коэффициентов. В нашем случае, границы значение «ноль» не включают, что еще раз подтверждает статистическую значимость (отличие от нуля) коэффициентов уравнения регрессии.

Раздел 4. Гетероскедастичность и автокоррелированность остатков регрессии. Обобщенный метод наименьших квадратов.

В результате изучения данного раздела студент должен:
знать:
тест Голдфелда-Квандта и тест Дарбина-Уотсона;
метод рядов для обнаружения автокррреляции в остатках регрессии;
методы устранения гетероскедастичности и автокоррелированности остатков регресии;
уметь:
проводить диагностику гетероскедастичности и автокоррелированности остатков регрессии;
владеть:
навыками проведения тестов Голдфелда-Квандта и Дарбина-Уотсона в MS Excel;
навыками оценки параметров уравнения регрессии в MS Excel при гетероскедастичности и автокоррелированности остатков регрессии.

Задачи с решениями

Задача 1.1. При анализе данных на гетероскедастичность вся выборка была после упорядочения по одному из факторов разбита на три подвыборки. Затем по результатам трехфакторного регрессионного анализа было определено, что остаточная сумма квадратов отклонений в первой подвыборке составила 180, а в третьей – 63. Подтверждается ли наличие гетероскедастичности, если объем данных в каждой подвыборке равен 20?

Решение: Рассчитаем F – статистику для проверки нулевой гипотезы о гомоскедастичности по тесту Голдфелда – Квандта:

$$F_{набл} = \frac{180}{63} = 2{,}857 \cdot$$

Найдем критические значения статистики Фишера:

$$F(0{,}1; k-p-1; k-p-1) = F(0{,}1; 20-3-1; 20-3-1) = 1{,}93;$$
$$F(0{,}05;16;16) = 2{,}33;$$
$$F(0{,}01;16;16) = 3{,}38.$$

Следовательно, на уровнях значимости 0,1 и 0,05 $F_{набл} > F_{кр}$, и гетероскедастичность имеет место, а на уровне 0,01 $F_{набл} < F_{кр}$, и гипотезу о гомоскедастичности отклонить нельзя.

Задача 1.2. По следующим данным проверьте нулевую гипотезу об отсутствии автокорреляции в остатках, применив критерий Дарбина-Уотсона.

Наблюдение	*Предсказанное Y*	*Остатки*	e_i-e_{i-1}	e^2
1	142,2467	-16,2467		263,9565
2	124,6969	12,30313	815,0949	151,367
3	159,2365	-11,2365	554,1143	126,259
4	242,3533	-51,3533	1609,361	2637,166
5	247,0209	26,97914	6135,978	727,874
6	307,0568	62,94318	1293,413	3961,844
7	361,2	70,79997	61,72901	5012,635
8	416,8019	28,19815	1814,915	795,1356
9	424,1765	-57,1765	7288,836	3269,156
10	350,3247	16,67529	5454,091	278,0653
11	345,3655	-24,3655	1684,344	593,6761
12	334,7235	-27,7235	11,27654	768,5939
13	386,7897	-55,7897	787,7102	3112,491
14	352,0517	-7,05169	2375,394	49,72629
15	353,2302	10,76977	317,6042	115,9879
16	361,7251	22,27488	132,3677	496,1704
Итого			30336,23	22360,1

Решение:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\ 2}} = \frac{30336{,}23}{22360{,}1} = 1{,}36$$

Сравним наблюдаемое значение DW=1,36 с табличными:

D_l=0,98, D_2=1,54. В данном случае 0,98<1,36<1,54 – наблюдаемое значение находится в области неопределенности. Поэтому окончательный вывод об автокорреляции остатков по критерию DW сделать нельзя.

Задача 1.3. По следующим данным проверьте нулевую гипотезу об отсутствии автокорреляции, применив метод рядов.

Наблюдение	*Предсказанное Y*	*Остатки*	Знак остатка
1	142,2467	-16,2467	-
2	124,6969	12,30313	+
3	159,2365	-11,2365	-
4	242,3533	-51,3533	-
5	247,0209	26,97914	+
6	307,0568	62,94318	+
7	361,2	70,79997	+
8	416,8019	28,19815	+
9	424,1765	-57,1765	-
10	350,3247	16,67529	+
11	345,3655	-24,3655	-
12	334,7235	-27,7235	-
13	386,7897	-55,7897	-
14	352,0517	-7,05169	-
15	353,2302	10,76977	+
16	361,7251	22,27488	+

Решение: Последовательность одинаковых знаков называется рядом. Количество рядов (К) составляет 8. Количество плюсов (N1) равно 8. Количество минусов (N2) тоже 8. Случайная величина К при отсутствии автокорреляции имеет асимптотически нормальное распределение, в котором

$$M(k) = \frac{2 \cdot n_1 \cdot n_2}{n_1 + n_2} + 1;$$

$$D(k) = \frac{2 \cdot n_1 \cdot n_2 \cdot (2 \cdot n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \cdot (n_1 + n_2 - 1)}.$$

Тогда, если $M(k) - u_{\alpha/2} \cdot D(k) < k < M(k) + u_{\alpha/2} \cdot D(k)$, то гипотеза об отсутствии автокорреляции не отклоняется. Если $k \le M(k) - u_{\alpha/2} \cdot D(k)$, то констатируется положительная автокорреляция; в случае $k \ge M(k) + u_{\alpha/2} \cdot D(k)$ признается наличие отрицательной автокорреляции.

В данной задаче:

$$M(k)=\frac{2\cdot 8\cdot 8}{8+8}+1=\frac{128}{16}+1=9.$$

$$D(k)=\frac{2\cdot 8\cdot 8\cdot(2\cdot 8\cdot 8-8-8)}{(8+8)^2\cdot(8+8-1)}=\frac{14336}{3840}=3{,}73.$$

Тогда с надежностью 95% :

$$M(k)-u_{\alpha/2}\cdot D(k)=9-1{,}96\cdot 3{,}73=1{,}69.$$

$$M(k)+u_{\alpha/2}\cdot D(k)=9+1{,}96\cdot 3{,}73=16{,}31.$$

В данной задаче: 1,69<8<16,31, следовательно, нулевая гипотеза об отсутствии автокорреляции не отклоняется.

Упражнения с пояснениями

Имеются данные о заработной плате (y, ден.ед.) и количестве сотрудников (x, чел.) по 30 предприятиям отрасли.

X	100	100	100	100	100	100	200	200	200	200	200	200
Y	75,5	75,5	77,5	78,5	80	81	80,5	82	84,5	85	85,5	86,5
X	300	300	300	300	300	300	400	400	400	400	400	400
Y	85,5	88,5	90	91	95	96	93	93,5	97,5	99	102,5	105
X	500	500	500	500	500	500						
Y	102	105,5	107	110,5	115	118,5						

Упражнение 4.1. По данным о заработной плате сотрудников (y, ден.ед.) и количестве сотрудников (x, чел.) необходимо оценить линейное уравнение парной регрессии заработной платы сотрудников и выполнить графический анализ остатков регрессии.

Решение. Чтобы оценить регрессию в Excel с помощью инструмента **Регрессия** надстройки **Пакет анализа** в главном меню программы Excel на вкладке *Данные* (в верхней строке) выберем опцию *Анализ данных*, а в списке *Инструментов анализа* выберем инструмент *Регрессия*. В качестве регрессанта выберем переменную Y – заработная плата сотрудников, а в качестве регрессоров укажем диапазон переменной X – количество сотрудников, укажем «галочку» в опции *Метки*, чтобы трактовать первую строку данных как названия переменных, также укажем «галочку» в опции *Остатки*, чтобы вывести таблицу остатков, укажем «галочку» в опции *График остатков*, чтобы вывести график остатков (рис.1).

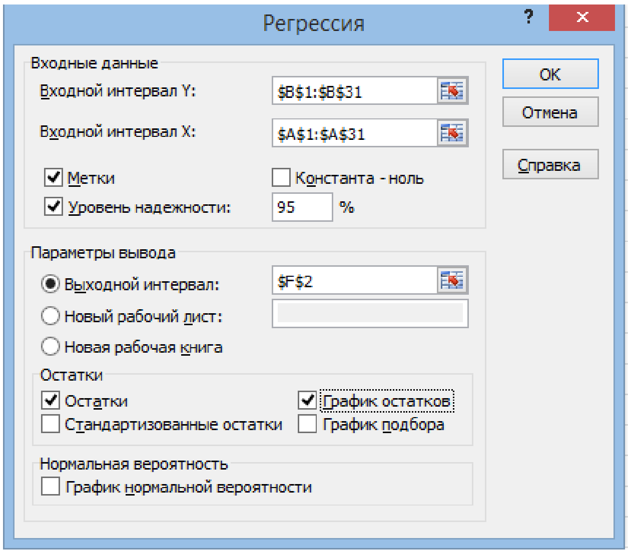


Рис. 1. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.2).

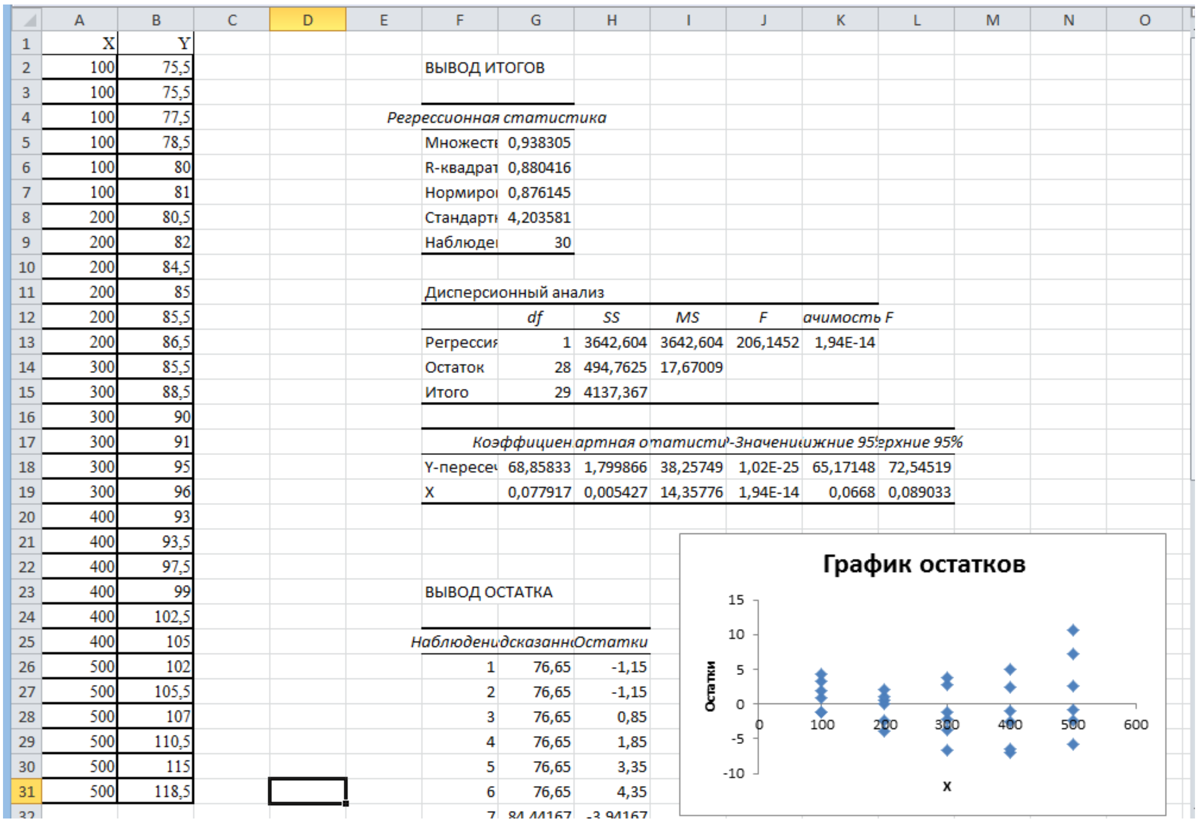


Рис. 2. Вывод результатов оценивания регрессии в Excel

Более подробно рассмотрим результаты, переведя их в обычные таблицы.

Регрессионная статистика	
Множественный R	0,938305
R-квадрат	0,880416
Нормированный R-квадрат	0,876145
Стандартная ошибка	4,203581
Наблюдения	30

Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	3642,604	3642,604	206,1452	1,94E-14	
Остаток	28	494,7625	17,67009			
Итого	29	4137,367				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	68,85833	1,799866	38,25749	1,02E-25	65,17148	72,54519
X	0,077917	0,005427	14,35776	1,94E-14	0,0668	0,089033

Вывод 1: В нашем случае *p-value* для статистики Фишера составило 1,94E-14, что меньше, чем α =0,01. Это означает, что с надежностью 99% отвергается нулевая гипотеза о неадекватности уравнения регрессии. Согласно тесту Фишера регрессия адекватна, количество сотрудников (X) линейно влияет на их заработную плату (Y). В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии при X составило 1,94E-14, что меньше, чем α =0,01. Это означает, что коэффициент регрессии β является значимым с надежностью 99%, между переменными X (количество сотрудников) и Y (заработная плата) существует значимая линейная связь. Доверительные границы коэффициентов значение «ноль» не включают, что еще раз подтверждает статистическую значимость (отличие от нуля) коэффициентов уравнения регрессии. Запишем исходную линейную модель парной регрессии:

$$Y = 68{,}858+0{,}078X+\varepsilon.$$

Комментарий: Если в предыдущих лабораторных работах по преимуществу обсуждались вопросы адекватности модели регрессии в целом и ее отдельных коэффициентов на основании тестов Фишера и Стьюдента, то здесь речь пойдет о случайных ошибках регрессии (для выборки – остатков регрессии). Надежность МНК-оценок зависит от свойств ошибок регрессии, которые принято называть предпосылками МНК. В теореме Гаусса-Маркова делается достаточно сильное предположение об одинаковой дисперсии всех случайных ошибок (гомоскедастичность). При нарушении предположения о гомоскедастичности случайных ошибок (их гетероскедастичности) формулы расчета F-статистики и t-статистики не годятся (дисперсии случайных ошибок разные и не имеют общую оценку RSS/n-m-1), поэтому могут быть сделаны неверные качественные выводы. Статистические пакеты F-статистику и t-статистики рассчитывают по традиционным формулам, без учета гетероскедастичности. Поэтому в оцениваемых регрессионных моделях обязательно надо проверять условие о гомоскедастичности ошибок, и, если оно не выполняется, применять взвешенный метод наименьших квадратов.

Если на графике остатков модули остатков достаточно сильно различаются, то это может свидетельствовать об их гетероскедастичности.

Вывод 2: В нашем случае на графике остатков с увеличением переменной X модули остатков тоже увеличиваются. График остатков скорее всего свидетельствует об их гетероскедастичности.

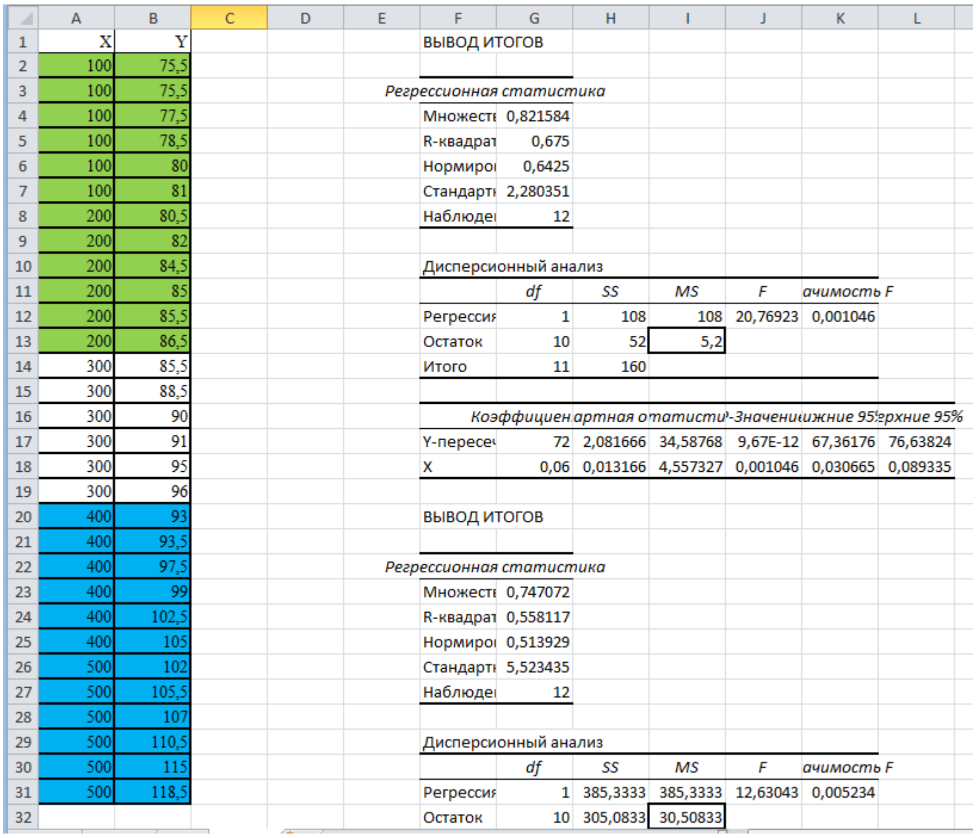
Упражнение 4.2. Проверьте наличие гетероскедастичности в остатках регрессии, используя тест Голдфелда-Квандта, k=12.

Решение. Для четкого ответа на вопрос, имеет ли место гетероскедастичность остатков, необходимо проводить специальные тесты.

Тест Голдфелда-Квандта имеет предположение о том, что дисперсии случайных ошибок линейно пропорциональны переменной X. Тест состоит из нескольких шагов.

Шаг 1: исходные наблюдения упорядочиваем по модулю переменной X и делим на три подвыборки (приблизительно по трети в каждой), в первой и третьей подвыборках количество наблюдений одинаково (по n_1= n_3= 12).

Шаг 2: по первым n_1 и последним n_3 наблюдениям оцениваем отдельные регрессии с помощью инструмента **Регрессия** надстройки **Пакет анализа.** Наблюдения, входящие в среднюю подвыборку, не используем. Определяем остаточные дисперсии $S_{RSS}{}^{2}{}_{3}$ и $S_{RSS}{}^{2}{}_{1}$ (в подтаблице «Дисперсионный анализ» на пересечении строки «Остаток» и столбца «MS»), рис. 3.



	A	B	C	D	E	F	G	H	I	J	K	L
1	X	Y				ВЫВОД ИТОГОВ						
2	100	75,5										
3	100	75,5				Регрессионная статистика						
4	100	77,5				Множеств	0,821584					
5	100	78,5				R-квадрат	0,675					
6	100	80				Нормиро	0,6425					
7	100	81				Стандарт	2,280351					
8	200	80,5				Наблюде	12					
9	200	82										
10	200	84,5				Дисперсионный анализ						
11	200	85					df	SS	MS	F	ачимость F	
12	200	85,5				Регрессия	1	108	108	20,76923	0,001046	
13	200	86,5				Остаток	10	52	5,2			
14	300	85,5				Итого	11	160				
15	300	88,5										
16	300	90					Коэффициен	артная о	татисти	Р-Значени	ижние 95	ерхние 95%
17	300	91				Y-пересе	72	2,081666	34,58768	9,67E-12	67,36176	76,63824
18	300	95				X	0,06	0,013166	4,557327	0,001046	0,030665	0,089335
19	300	96										
20	400	93				ВЫВОД ИТОГОВ						
21	400	93,5										
22	400	97,5				Регрессионная статистика						
23	400	99				Множеств	0,747072					
24	400	102,5				R-квадрат	0,558117					
25	400	105				Нормиро	0,513929					
26	500	102				Стандарт	5,523435					
27	500	105,5				Наблюде	12					
28	500	107										
29	500	110,5				Дисперсионный анализ						
30	500	115					df	SS	MS	F	ачимость F	
31	500	118,5				Регрессия	1	385,3333	385,3333	12,63043	0,005234	
32						Остаток	10	305,0833	30,50833			

Рис. 3. Вывод результатов оценивания остаточных дисперсий в Excel

Шаг 3: гипотеза H_0 сводится к проверке гипотезы о равенстве дисперсий первых n_1 и последних n_3 наблюдений с помощью F-статистики:

$$F = \frac{S_{RSS}{}^{2}{}_{3}}{S_{RSS}{}^{2}{}_{1}} = \frac{30,508}{5,2} = 5,867$$

Шаг 4: сравниваем наблюдаемое значение F-статистики с критическим $F_{\alpha;\ n1-m-1;\ n2-m-1}$. Для α=0,05 $F_{0,05;\ 12-1-1;\ 12-1-1}$=2,978.

Вывод 3: Согласно тесту Голдфелда-Квандта наблюдаемое значение F-статистики превышает критическое, значит, H_0 отвергается, остатки гетероскедастичны.

Комментарий: Зависимость дисперсии случайных ошибок от X может быть необязательно линейной. Более разнообразные формы функциональной зависимости проверяются в *тесте Глейзера.* В случае зависимости случайных ошибок от нескольких факторов X ответ на вопрос, а имеет ли место гетероскедастичность достаточно общего вида дает *тест Уайта*.

Упражнение 4.3. Для устранения гетероскедастичности примените взвешенный метод наименьших квадратов, предполагая, что дисперсии остатков пропорциональны значениям X. Проведите графический анализ остатков регрессии после применения взвешенного метода наименьших квадратов.

Решение. Чтобы применить для устранения гетероскедастичности взвешенный метод наименьших квадратов в предположении, что дисперсии пропорциональны значениям X, надо сформулировать новые переменные: Y/ корень(X), 1/корень(X), корень(X). Необходимо построить по ним новое уравнение регрессии с использованием инструмента **Регрессия** надстройки **Пакет анализа**. Обязательно поставьте «флажок» в опциях *Константа, Остатки* диалогового окна инструмента **Регрессия** (рис.4) .

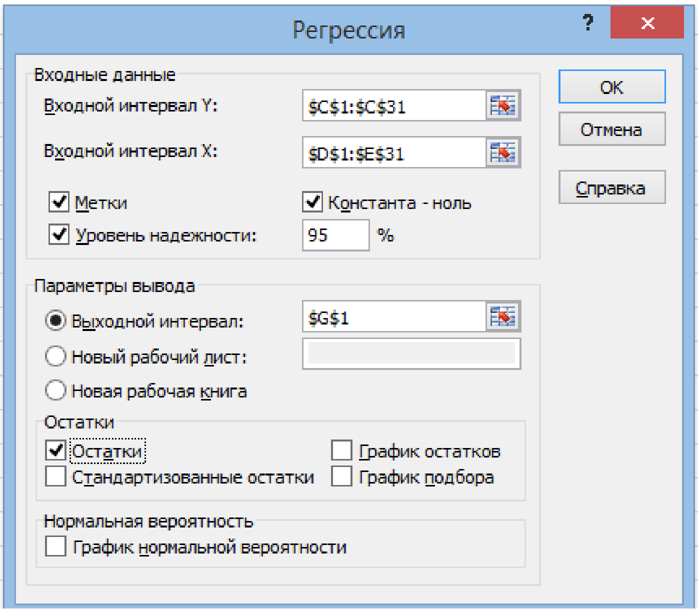
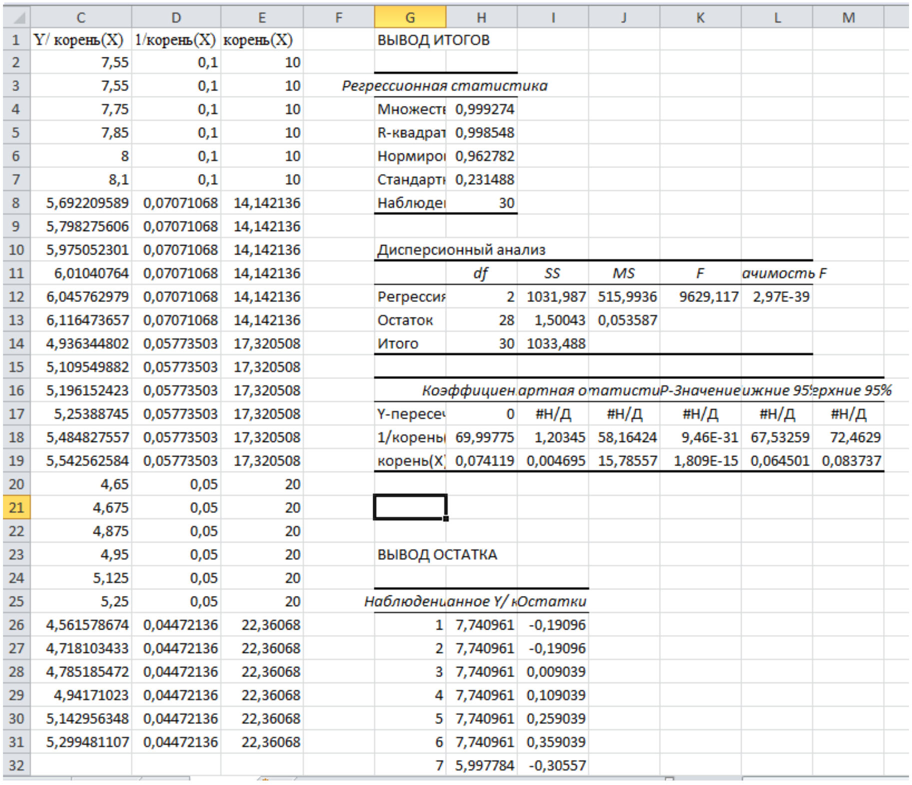


Рис. 4. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты взвешенного метода наименьших квадратов (рис.5).



	C	D	E	F	G	H	I	J	K	L	M
1	Y/ корень(X)	1/корень(X)	корень(X)		ВЫВОД ИТОГОВ						
2	7,55	0,1	10								
3	7,55	0,1	10		*Регрессионная статистика*						
4	7,75	0,1	10		Множеств	0,999274					
5	7,85	0,1	10		R-квадрат	0,998548					
6	8	0,1	10		Нормиро	0,962782					
7	8,1	0,1	10		Стандарт	0,231488					
8	5,692209589	0,07071068	14,142136		Наблюде	30					
9	5,798275606	0,07071068	14,142136								
10	5,975052301	0,07071068	14,142136		Дисперсионный анализ						
11	6,01040764	0,07071068	14,142136			*df*	*SS*	*MS*	*F*	*ачимость F*	
12	6,045762979	0,07071068	14,142136		Регрессия	2	1031,987	515,9936	9629,117	2,97E-39	
13	6,116473657	0,07071068	14,142136		Остаток	28	1,50043	0,053587			
14	4,936344802	0,05773503	17,320508		Итого	30	1033,488				
15	5,109549882	0,05773503	17,320508								
16	5,196152423	0,05773503	17,320508			*Коэффициен*	*артная о*	*татисти*	*P-Значение*	*ижние 95*	*ерхние 95%*
17	5,25388745	0,05773503	17,320508		Y-пересеч	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
18	5,484827557	0,05773503	17,320508		1/корень	69,99775	1,20345	58,16424	9,46E-31	67,53259	72,4629
19	5,542562584	0,05773503	17,320508		корень(X	0,074119	0,004695	15,78557	1,809E-15	0,064501	0,083737
20	4,65	0,05	20								
21	4,675	0,05	20								
22	4,875	0,05	20								
23	4,95	0,05	20		ВЫВОД ОСТАТКА						
24	5,125	0,05	20								
25	5,25	0,05	20		*Наблюдени*	*анное Y/ к*	*Остатки*				
26	4,561578674	0,04472136	22,36068		1	7,740961	-0,19096				
27	4,718103433	0,04472136	22,36068		2	7,740961	-0,19096				
28	4,785185472	0,04472136	22,36068		3	7,740961	0,009039				
29	4,94171023	0,04472136	22,36068		4	7,740961	0,109039				
30	5,142956348	0,04472136	22,36068		5	7,740961	0,259039				
31	5,299481107	0,04472136	22,36068		6	7,740961	0,359039				
32					7	5,997784	-0,30557				

Рис. 5. Вывод результатов оценивания регрессии взвешенным методом наименьших

квадратов в Excel

Вывод 4: Для случая взвешенных наименьших квадратов *p-value* для статистики Фишера составило 2,97E-39, что меньше, чем α =0,01. Это означает, что с надежностью 99% отвергается нулевая гипотеза о неадекватности уравнения регрессии. Согласно тесту Фишера регрессия адекватна, переменная $\sqrt{X}$ и переменная $1/\sqrt{X}$ влияют на переменную $Y/\sqrt{X}$. В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии при $\sqrt{X}$ составило 1,81E-15, для коэффициента регрессии при

$1/\sqrt{X}$ составило 9,46E-31,что меньше, чем $\alpha = 0{,}01$. Это означает, что коэффициенты регрессии, полученные взвешенным методом наименьших квадратов, являются значимыми с надежностью 99%. Доверительные границы коэффициентов значение «ноль» не включают, что еще раз подтверждает статистическую значимость (отличие от нуля) коэффициентов уравнения регрессии. Запишем преобразованную линейную модель парной регрессии:

Y = 69,998+0,074X+ε.

В силу теоремы Гаусса-Маркова соответствующие оценки коэффициентов являются эффективными.

Чтобы построить корреляционное поле преобразованных Остатков (по оси Y) и количества сотрудников (по оси X) используем Мастер диаграмм MS Excel. Для этого в отдельное поле скопируем и выделим «мышью» исходные наблюдения переменной X - количество сотрудников и переменной Y – столбец «Остатки» из подтаблицы *Вывод остатка* (рис.5), затем в Главном меню MS Excel выберем: Вставка – Диаграммы – Точечная – Точечная с маркерами. Полученная диаграмма рассеяния представлена на рисунке 6.

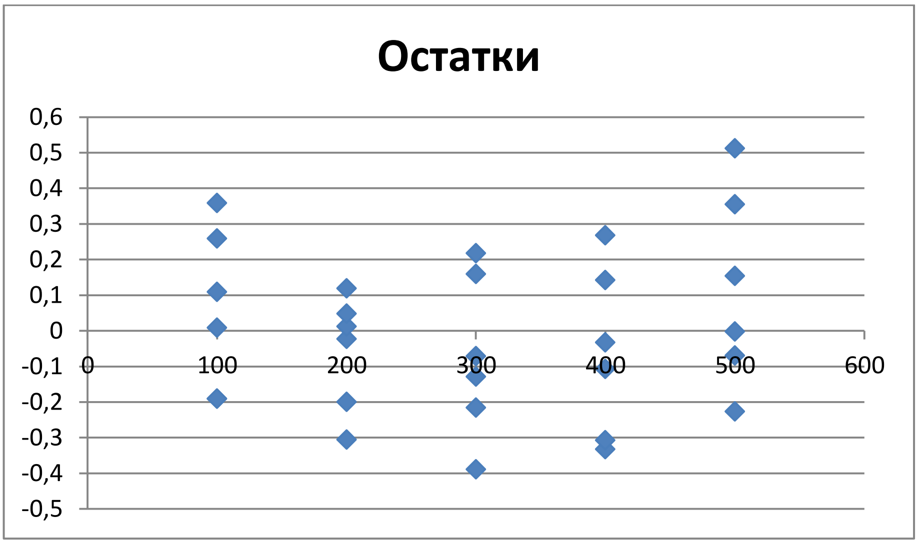


Рис.6. График остатков в MS Excel

Вывод 5: В нашем случае на графике остатков с увеличением переменной X модули остатков почти не изменяются. График остатков скорее всего свидетельствует об их гомоскедастичности.

Упражнение 4.4. Необходимо оценить линейное уравнение парной регрессии импорта и выполнить графический анализ остатков регрессии на основе ежегодных данных об импорте (Y, млн. ден. ед.) и экспорте (X, млн. ден. ед.) промышленной корпорации.

Y	11,07	11,5	12,01	12,28	13,16	13,43	13,28	13,5	15,32	15,62	17,44	16,14
X	12,47	12,65	12,89	12,97	13	13,31	13,25	12,65	14,49	14,47	14,74	14,62
Y	16,14	16,08	16,55	15	18,72	17,8	16,64	17,39	18,7	18,02	17,46	16,96
X	17,6	17,7	16,6	15,26	19,49	19,08	18,69	18,65	19,33	19,11	18,62	18,4
Y	15,06	16,01	16,63	17,86	14,56	15,64	16,45	17,42	14,3	14,59	14,66	14,95
X	16,15	16,58	17,6	18,48	15,36	15,25	15,61	15,93	14,38	14,3	14,75	15,58

Решение: Чтобы оценить регрессию в Excel с помощью инструмента **Регрессия** надстройки **Пакет анализа** в главном меню программы Excel на вкладке *Данные* (в верхней строке) выберем опцию *Анализ данных*, а в списке *Инструментов анализа* выберем инструмент *Регрессия*. В качестве регрессанта выберем переменную Y – импорт, а в качестве регрессоров укажем диапазон переменной X – экспорт, укажем «галочку» в опции *Метки*, чтобы трактовать первую строку данных как названия переменных, также укажем «галочку» в опции *Остатки*, чтобы вывести таблицу остатков, укажем «галочку» в опции *График остатков*, чтобы вывести график остатков (рис.1).

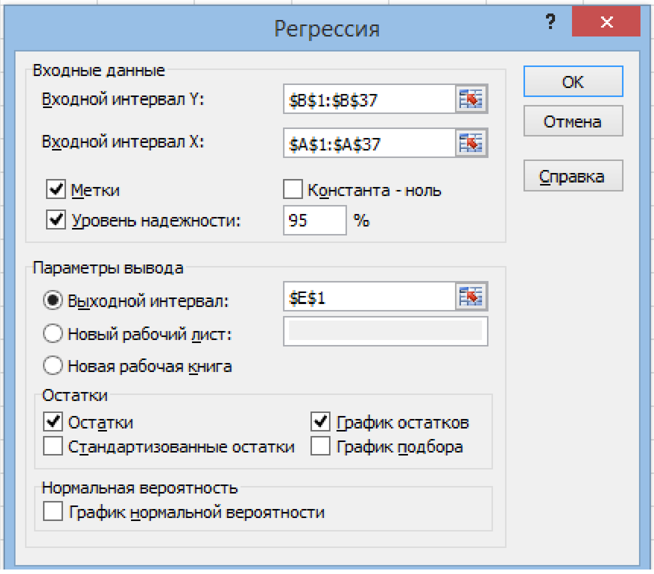
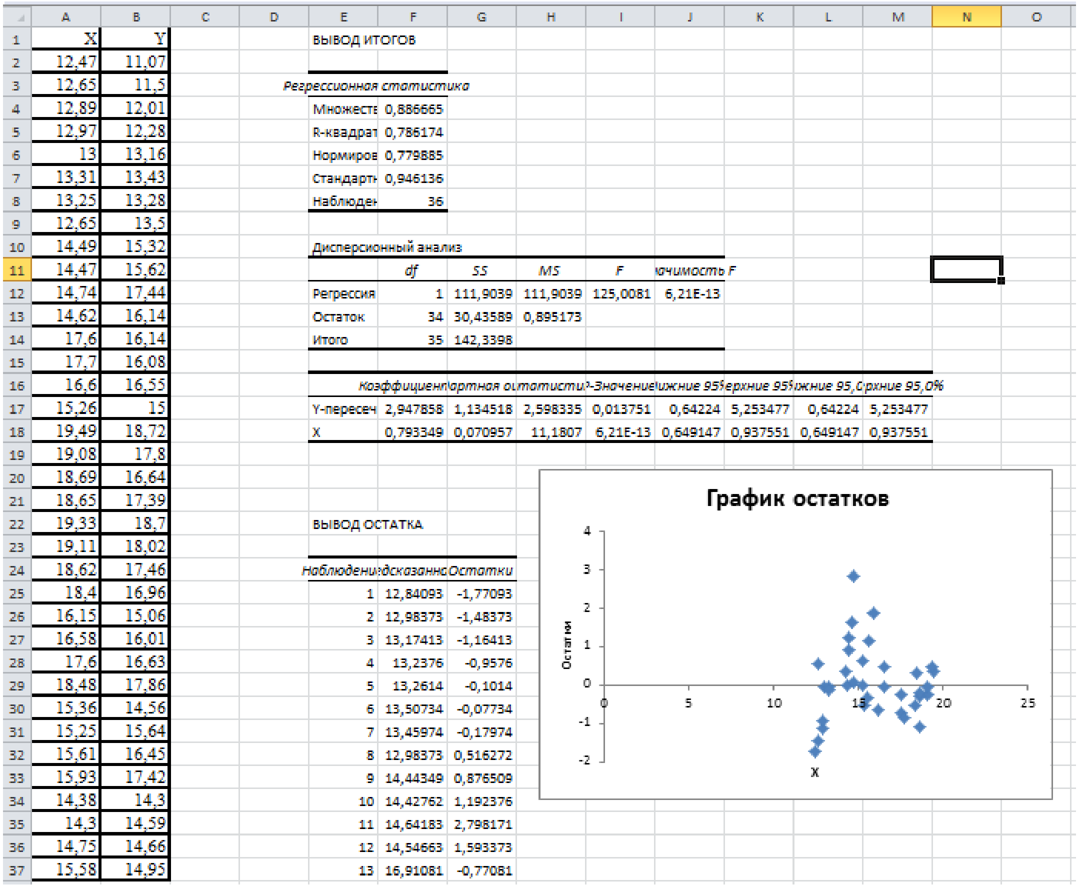


Рис.1. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.2).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	X	Y			ВЫВОД ИТОГОВ								
2	12,47	11,07											
3	12,65	11,5			Регрессионная статистика								
4	12,89	12,01			Множеств	0,886665							
5	12,97	12,28			R-квадрат	0,786174							
6	13	13,16			Нормиров	0,779885							
7	13,31	13,43			Стандартн	0,946136							
8	13,25	13,28			Наблюден	36							
9	12,65	13,5											
10	14,49	15,32			Дисперсионный анализ								
11	14,47	15,62				df	SS	MS	F	ачимость F			
12	14,74	17,44			Регрессия	1	111,9039	111,9039	125,0081	6,21E-13			
13	14,62	16,14			Остаток	34	30,43589	0,895173					
14	17,6	16,14			Итого	35	142,3398						
15	17,7	16,08											
16	16,6	16,55				Коэффициент	артная оц	татистик	-Значение	ижние 95%	ерхние 95%	ижние 95,0%	рхние 95,0%
17	15,26	15			Y-пересеч	2,947858	1,134518	2,598335	0,013751	0,64224	5,253477	0,64224	5,253477
18	19,49	18,72			X	0,793349	0,070957	11,1807	6,21E-13	0,649147	0,937551	0,649147	0,937551
19	19,08	17,8											
20	18,69	16,64											
21	18,65	17,39											
22	19,33	18,7			ВЫВОД ОСТАТКА								
23	19,11	18,02											
24	18,62	17,46			Наблюдение	дсказанное	Остатки						
25	18,4	16,96			1	12,84093	-1,77093						
26	16,15	15,06			2	12,98373	-1,48373						
27	16,58	16,01			3	13,17413	-1,16413						
28	17,6	16,63			4	13,2376	-0,9576						
29	18,48	17,86			5	13,2614	-0,1014						
30	15,36	14,56			6	13,50734	-0,07734						
31	15,25	15,64			7	13,45974	-0,17974						
32	15,61	16,45			8	12,98373	0,516272						
33	15,93	17,42			9	14,44349	0,876509						
34	14,38	14,3			10	14,42762	1,192376						
35	14,3	14,59			11	14,64183	2,798171						
36	14,75	14,66			12	14,54663	1,593373						
37	15,58	14,95			13	16,91081	-0,77081						

Рис 2. Вывод результатов оценивания регрессии в Excel

Более подробно рассмотрим результаты, переведя их в обычные таблицы.

Регрессионная статистика	
Множественный R	0,886665

R-квадрат						0,786174
Нормированный R-квадрат						0,779885
Стандартная ошибка						0,946136
Наблюдения						36
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	1	111,9039	111,9039	125,0081	6,21E-13	
Остаток	34	30,43589	0,895173			
Итого	35	142,3398				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	2,947858	1,134518	2,598335	0,013751	0,64224	5,253477
X	0,793349	0,070957	11,1807	6,21E-13	0,649147	0,937551

Вывод 1: В нашем случае *p-value* для статистики Фишера составило 6,21E-13, что меньше, чем $\alpha = 0{,}01$. Это означает, что с надежностью 99% отвергается нулевая гипотеза о неадекватности уравнения регрессии. Согласно тесту Фишера регрессия адекватна, экспорт (X) линейно влияет на импорт (Y). В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии при X составило 6,21E-13, что меньше, чем $\alpha = 0{,}01$. Это означает, что коэффициент регрессии β является значимым с надежностью 99%, между переменными X (экспорт) и Y (импорт) существует значимая линейная связь. Доверительные границы коэффициентов значение «ноль» не включают, что еще раз подтверждает статистическую значимость (отличие от нуля) коэффициентов уравнения регрессии. Запишем исходную линейную модель парной регрессии:

$$Y = 2{,}948+0{,}793X+\varepsilon.$$

Комментарий: В теореме Гаусса-Маркова делается достаточно сильное предположение об отсутствии взаимосвязи между случайными ошибками (отсутствие автокорреляции остатков). При нарушении предположения об отсутствии автокорреляции остатков формулы расчета F-статистики и t-статистики не годятся (дисперсии случайных ошибок разные и не имеют общую оценку RSS/n-m-1), поэтому могут быть сделаны неверные качественные выводы. Статистические пакеты F-статистику и t-статистики рассчитывают по традиционным формулам, без учета автокорреляции. Поэтому в оцениваемых регрессионных моделях обязательно надо проверять условие об отсутствии автокорреляции ошибок, и, если оно не выполняется, применять авторегрессионное преобразование.

Если на графике остатки серийно меняют свой знак и достаточно сильно различаются, то это может свидетельствовать об их автокоррелированности.

Вывод 2: В нашем случае на графике остатков с увеличением переменной X остатки серийно меняют свой знак и увеличиваются. График остатков скорее всего свидетельствует об их автокоррелированности.

Упражнение 4.5. Вычислите значение DW - статистики Дарбина-Уотсона и на ее основе проанализируйте наличие автокорреляции. На основе полученных результатов ответьте на вопрос: Будет ли отклоняться гипотеза о положительной зависимости между объемами экспорта и импорта?

Решение: Чтобы вычислить значение DW - статистики Дарбина-Уотсона, необходимо в подтаблице «Вывод остатка» рядом со столбцом «Остатки» сформировать дополнительные расчетные графы (E_i-E_{i-1})^2, E_i^2 и определить по ним суммы (рис.3).

ВЫВОД ОСТАТКА

Наблюдени	*редсказанное*	*Остатки*	(E_i-E_{i-1})^2	E_i^2
1	12,84092555	-1,77093		3,136177
2	12,98372844	-1,48373	0,082482	2,20145
3	13,1741323	-1,16413	0,102142	1,355204
4	13,23760025	-0,9576	0,042655	0,916998
5	13,26140074	-0,1014	0,733078	0,010282
6	13,50733905	-0,07734	0,000579	0,005981
7	13,45973809	-0,17974	0,010486	0,032306
8	12,98372844	0,516272	0,484429	0,266536
9	14,44349136	0,876509	0,129771	0,768267
10	14,42762437	1,192376	0,099772	1,42176
11	14,64182871	2,798171	2,57858	7,829763
12	14,54662678	1,593373	1,451538	2,538838
13	16,91080803	-0,77081	5,589353	0,594145
14	16,99014297	-0,91014	0,019414	0,82836
15	16,11745862	0,432541	1,802801	0,187092
16	15,05437041	-0,05437	0,237083	0,002956
17	18,41023842	0,309762	0,132592	0,095952
18	18,08496516	-0,28497	0,3537	0,081205
19	17,77555889	-1,13556	0,72351	1,289494
20	17,74382491	-0,35382	0,611108	0,125192
21	18,28330251	0,416697	0,593705	0,173637
22	18,10876564	-0,08877	0,255493	0,007879
23	17,72002443	-0,26002	0,02933	0,067613
24	17,54548756	-0,58549	0,105926	0,342796
25	15,76045138	-0,70045	0,013217	0,490632
26	16,10159163	-0,09159	0,37071	0,008389
27	16,91080803	-0,28081	0,035803	0,078853
28	17,60895551	0,251044	0,282867	0,063023
29	15,13370535	-0,57371	0,680212	0,329138
30	15,04643691	0,593563	1,362516	0,352317
31	15,3320427	1,117957	0,274989	1,249829
32	15,58591451	1,834085	0,51284	3,36387
33	14,35622292	-0,05622	3,573266	0,003161
34	14,29275497	0,297245	0,12494	0,088355
35	14,64976221	0,010238	0,082373	0,000105
36	15,30824222	-0,35824	0,135778	0,128337
		Сумма	23,61904	30,43589

Рис.3. Подтаблица «Вывод остатка» в Excel

Для расчета DW – статистики Дарбина -Уотсона под подтаблицей «Вывод остатка», используя ссылки на необходимые ячейки, необходимо применить формулу:

$$DW = \frac{\sum_{n=2}^{N}(e_i - e_{i-1})^2}{\sum_{n=1}^{N} e_i^{\,2}} = \frac{23,619}{30,436} = 0,776$$

Критические значения статистики Дарбина –Уотсона в нашем случае составляют d_1=1,4 и d_2=1,52.

Вывод 3: Наблюдаемое значение DW – статистики, равное 0,076, меньше, чем d_1=1,4, значит в остатках регрессии присутствует положительная автокорреляция. Поскольку в условиях автокоррелированности остатков регрессии могут быть сделаны неверные качественные выводы о наличии линейной взаимосвязи между регрессором и регрессантом на основе F-статистики и t-статистики, то на основе полученных результатов нельзя утверждать о положительной зависимости между объемами экспорта и импорта.

Упражнение 4.6. Для устранения автокорреляции необходимо выполнить авторегрессионное преобразование, в котором ρ=1. Выполнить графический анализ остатков и расчет DW – статистики Дарбина – Уотсона после применения авторегрессионного преобразования.

Решение: Чтобы выполнить авторегрессионное преобразование, необходимо сформировать новые переменные-приращения: для импорта: $Y^*=Y_i-Y_{i-1}$, для экспорта: $X^*=X_i-X_{i-1}$. Затем с использованием инструмента **Регрессия** надстройки **Пакет анализа** необходимо построить регрессию приращения импорта на приращение экспорта. Обязательно поставьте «флажок» в опции *Константа* и в опции *Остатки* диалогового окна инструмента **Регрессия** (рис. 4).

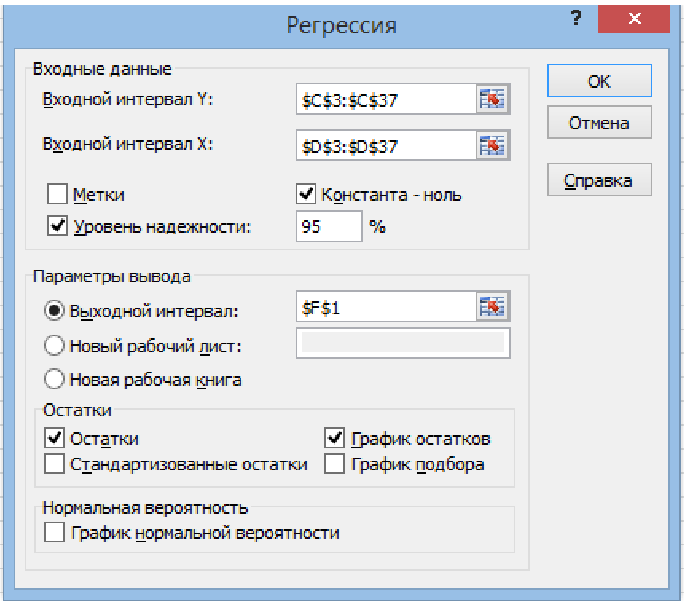
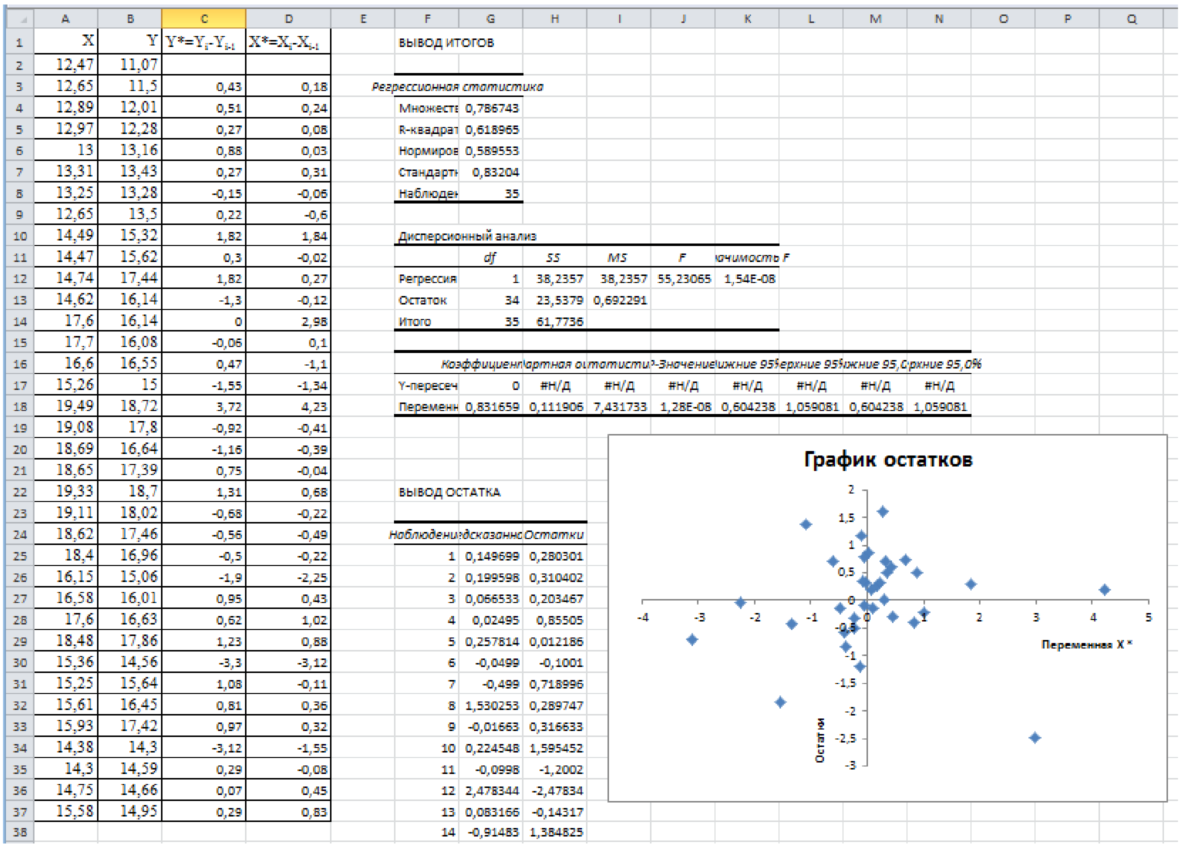


Рис.4. Окно параметров регрессии с введенными значениями в Excel

Далее, нажав на кнопку Ок, получаем результаты (рис.5).



	A	B	C	D
1	X	Y	Y*=Yi-Yi-1	X*=Xi-Xi-1
2	12,47	11,07		
3	12,65	11,5	0,43	0,18
4	12,89	12,01	0,51	0,24
5	12,97	12,28	0,27	0,08
6	13	13,16	0,88	0,03
7	13,31	13,43	0,27	0,31
8	13,25	13,28	-0,15	-0,06
9	12,65	13,5	0,22	-0,6
10	14,49	15,32	1,82	1,84
11	14,47	15,62	0,3	-0,02
12	14,74	17,44	1,82	0,27
13	14,62	16,14	-1,3	-0,12
14	17,6	16,14	0	2,98
15	17,7	16,08	-0,06	0,1
16	16,6	16,55	0,47	-1,1
17	15,26	15	-1,55	-1,34
18	19,49	18,72	3,72	4,23
19	19,08	17,8	-0,92	-0,41
20	18,69	16,64	-1,16	-0,39
21	18,65	17,39	0,75	-0,04
22	19,33	18,7	1,31	0,68
23	19,11	18,02	-0,68	-0,22
24	18,62	17,46	-0,56	-0,49
25	18,4	16,96	-0,5	-0,22
26	16,15	15,06	-1,9	-2,25
27	16,58	16,01	0,95	0,43
28	17,6	16,63	0,62	1,02
29	18,48	17,86	1,23	0,88
30	15,36	14,56	-3,3	-3,12
31	15,25	15,64	1,08	-0,11
32	15,61	16,45	0,81	0,36
33	15,93	17,42	0,97	0,32
34	14,38	14,3	-3,12	-1,55
35	14,3	14,59	0,29	-0,08
36	14,75	14,66	0,07	0,45
37	15,58	14,95	0,29	0,83
38				

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множеств	0,786743
R-квадрат	0,618965
Нормиров	0,589553
Стандартн	0,83204
Наблюден	35

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	1	38,2357	38,2357	55,23065	1,54E-08
Остаток	34	23,5379	0,692291		
Итого	35	61,7736			

	Коэффициент	*артная ош*	*татисти*	*Р-Значение*	*ижние 95%*	*ерхние 95%*	*ижние 95,0%*	*рхние 95,0%*
Y-пересеч	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Переменн	0,831659	0,111906	7,431733	1,28E-08	0,604238	1,059081	0,604238	1,059081

ВЫВОД ОСТАТКА

Наблюдени	*едсказанно*	*Остатки*
1	0,149699	0,280301
2	0,199598	0,310402
3	0,066533	0,203467
4	0,02495	0,85505
5	0,257814	0,012186
6	-0,0499	-0,1001
7	-0,499	0,718996
8	1,530253	0,289747
9	-0,01663	0,316633
10	0,224548	1,595452
11	-0,0998	-1,2002
12	2,478344	-2,47834
13	0,083166	-0,14317
14	-0,91483	1,384825

Рис 5. Вывод результатов авторегрессионного преобразования в Excel

Вывод 4: После авторегрессионного преобразования на графике остатков с увеличением переменной X остатки в своем поведении не проявляют очевидной закономерности. График остатков скорее всего свидетельствует об отсутствии автокорреляции в остатках регрессии.

Чтобы вычислить значение DW - статистики Дарбина-Уотсона после авторегрессионного преобразования, необходимо в подтаблице «Вывод остатка» рядом со столбцом «Остатки» (рис.5) сформировать дополнительные расчетные графы $(E_i\text{-}E_{i-1})$^2, E_i^2 и определить по ним суммы (рис.6).

38					
39	ВЫВОД ОСТАТКА				
40					
41	Наблюдение	Предсказанное	*Остатки*	$(E_i\text{-}E_{i-1})$^2	E_i^2
42	1	0,149699	0,28030135		0,078569
43	2	0,199598	0,3104018	0,000906037	0,096349
44	3	0,066533	0,20346727	0,011434994	0,041399
45	4	0,02495	0,85505022	0,424560352	0,731111
46	5	0,257814	0,01218566	0,710420679	0,000148
47	6	-0,0499	-0,10010045	0,01260817	0,01002
48	7	-0,499	0,7189955	0,670918177	0,516955
49	8	1,530253	0,28974713	0,184254164	0,083953
50	9	-0,01663	0,31663318	0,00072286	0,100257
51	10	0,224548	1,59545202	1,635377628	2,545467
52	11	-0,0998	-1,2002009	7,815675273	1,440482
53	12	2,478344	-2,47834432	1,633650609	6,142191
54	13	0,083166	-0,14316592	5,453058187	0,020496
55	14	-0,91483	1,38482509	2,334756503	1,917741
56	15	-1,11442	-0,43557671	3,313862712	0,189727
57	16	3,517918	0,20208172	0,406608274	0,040837
58	17	-0,34098	-0,57901974	0,610119487	0,335264
59	18	-0,32435	-0,83565292	0,065860591	0,698316
60	19	-0,03327	0,78326637	2,620899671	0,613506
61	20	0,565528	0,74447177	0,001505021	0,554238
62	21	-0,18297	-0,49703498	1,541339006	0,247044
63	22	-0,40751	-0,15248701	0,118713307	0,023252
64	23	-0,18297	-0,31703498	0,027076036	0,100511
65	24	-1,87123	-0,02876687	0,083098505	0,000828
66	25	0,357613	0,59238656	0,385831581	0,350922
67	26	0,848292	-0,22829235	0,673513873	0,052117
68	27	0,73186	0,49813993	0,527703863	0,248143
69	28	-2,59478	-0,70522339	1,448083293	0,49734
70	29	-0,09148	1,17148251	3,522025044	1,372371
71	30	0,299397	0,5106027	0,436762122	0,260715
72	31	0,266131	0,70386907	0,037351889	0,495432
73	32	-1,28907	-1,83092829	6,425197629	3,352298
74	33	-0,06653	0,35653273	4,784985723	0,127116
75	34	0,374247	-0,30424663	0,436629362	0,092566
76	35	0,690277	-0,40027711	0,009221854	0,160222
77			Сумма	48,36473247	23,5379

Рис. 6. Подтаблица «Вывод остатка» после авторегрессионного преобразования в Excel

Для расчета DW – статистики Дарбина -Уотсона под подтаблицей «Вывод остатка» после авторегрессионного преобразования, используя ссылки на необходимые ячейки, необходимо применить формулу:

$$DW = \frac{\sum_{n=2}^{N}(e_i - e_{i-1})^2}{\sum_{n=1}^{N} e_i^2} = \frac{48,364}{23,538} = 2,05$$

Критические значения статистики Дарбина –Уотсона в нашем случае составляют d_1=1,39 и d_2=1,51.

Вывод 5: Наблюдаемое значение DW – статистики после авторегрессионного преобразования, равное 2,05, больше, чем d_1=1,51, значит, в остатках регрессии отсутствует автокорреляция.

После авторегрессионного преобразования *p-value* для статистики Фишера составило 1,54E-08, что меньше, чем $\alpha = 0{,}01$. Это означает, что с надежностью 99% отвергается нулевая гипотеза о незначимости уравнения регрессии в целом. Согласно тесту Фишера регрессия статистически значима, приращение экспорта (X*) линейно влияет на приращение импорта (Y*). В нашем случае *p-value* для статистики Стьюдента для коэффициента регрессии при X* составило 1,28E-08, что меньше, чем $\alpha = 0{,}01$. Это означает, что коэффициент регрессии β является значимым с надежностью 99%, между переменными X* (приращение экспорта) и Y* (приращение импорта) существует значимая линейная связь. Доверительные границы коэффициента значение «ноль» не включают, что еще раз подтверждает статистическую значимость коэффициента регрессии. Запишем линейную модель парной регрессии после авторегрессионного преобразования:

$$Y = 2{,}948+0{,}832X+\varepsilon.$$

Если исходная модель регрессии с автокоррелированными остатками показывала, что увеличение экспорта на 1 ден. единицу приводит к увеличению импорта на 0,793 ден. единиц, то после авторегрессионного преобразования эффективная (с наименьшей дисперсией) оценка коэффициента регрессии показывает, что увеличение экспорта на 1 ден. единицу приводит к увеличению импорта на 0,832 ден. единиц.

Раздел 5. Тренд-сезонные модели нестационарных временных рядов

В результате изучения данного раздела студент должен:

знать:

особенности анализа временных рядов;

тренд-сезонные модели анализа временных рядов;

компоненты тренд-сезонных моделей;

уметь:

отличать модели с тренд-сезонной компонентой от стационарных рядов;

определять скорректированную сезонную компоненту;

владеть:

навыками декомпозиции временного ряда с трендовой и сезонной компонентами;

навыками прогнозирования по тренд-сезонным моделям.

Задачи с решениями

Задача 1.1. Имеются следующие условные данные о средних расходах на конечное потребление за 8 лет:

t	1	2	3	4	5	6	7	8
Y_t	7	8	8	10	11	12	14	16

В предположении, что расходы на конечное потребление в текущем году зависят от расходов на конечное потребление предыдущих лет, определите коэффициенты автокорреляции первого и второго порядков.

Решение:

Коэффициент автокорреляции уровней ряда первого порядка характеризует зависимость между соседними уровнями ряда y_t и y_{t-1} и определяется по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2 \cdot}}, \text{ где}$$

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{8+8+10+11+12+14+16}{7} = \frac{79}{7} = 11{,}29;$$

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12+14}{7} = \frac{70}{7} = 10.$$

Коэффициент автокорреляции уровней второго порядка измеряет зависимость между уровнями ряда y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}, \text{ где}$$

$$\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2} = \frac{8+10+11+12+14+16}{6} = \frac{71}{6} = 11{,}83;$$

$$\bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2} = \frac{7+8+8+10+11+12}{6} = \frac{56}{6} = 9{,}33.$$

Составим расчетные таблицы:

Расчет коэффициента автокорреляции первого порядка

t	y_t	y_{t-1}	$y_t-\bar{y}_1$	$y_{t-1}-\bar{y}_2$	$(y_t-\bar{y}_1)*(y_{t-1}-\bar{y}_2)$	$(y_t-\bar{y}_1)^2$	$(y_{t-1}-\bar{y}_2)^2$
1	7	-	-	-	-	-	-
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0,00	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1

7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Итого	86	70	-0,03	0	44,0	53,4287	38

$$r_1 = \frac{44}{\sqrt{53{,}43 \cdot 38}} = 0{,}976.$$

Полученное значение свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и предшествующего годов и, следовательно, о наличии во временном ряде сильной линейной тенденции.

Расчет коэффициента автокорреляции второго порядка

t	y_t	y_{t-2}	$y_t-\bar{y}_3$	$y_{t-2}-\bar{y}_4$	$(y_t-\bar{y}_3)$*$(y_{t-2}-\bar{y}_4)$	$(y_t-\bar{y}_3)^2$	$(y_{t-2}-\bar{y}_4)^2$
1	7	-	-	-	-	-	-
2	8	-	-	-	-	-	-
3	8	7	-3,83	-2,33	8,9239	14,6689	5,4289
4	10	8	-1,83	-1,33	2,4339	3,3489	1,7689
5	11	8	-0,83	-1,33	1,1039	0,6889	1,7689
6	12	10	0,17	0,67	0,1139	0,0289	0,4489
7	14	11	2,17	1,67	3,6239	4,7089	2,7889
8	16	12	4,17	2,67	11,1339	17,3889	7,1289
Итого	86	56	0,02	0,02	27,3334	40,8334	19,3334

$$r_2 = \frac{27{,}3334}{\sqrt{40{,}8334 \cdot 19{,}3334}} = 0{,}973.$$

Полученные результаты подтверждают вывод о том, что временной ряд расходов на конечное потребление содержит линейную тенденцию.

Задача 1.2. На основе квартальных данных объемов продаж предприятия за 2012-2017 гг. была построена аддитивная модель временного ряда, трендовая компонента которой имеет вид:

$$T = 200 + 3 \cdot t \quad (t = 1, 2, \ldots).$$

Показатели за 2016 г. приведены в таблице:

Квартал	Фактический объем продаж	Компонента аддитивной модели		
		трендовая	сезонная	случайная
1	2	3	4	5
1	200			- 11
2			15	+ 5
3	250		32	
4				

Определить недостающие в таблице данные, зная, что объем продаж за 2016 г. составил 1000 тыс. руб.

Решение: В первую очередь определим все значения трендовой компоненты. Чтобы использовать имеющееся уравнение тренда, надо определить моменты времени, относящиеся к 2016 г. Поскольку модель относится к периоду 2012 – 2017 гг., т.е. охватывает 6 лет, квартальные временные отметки изменяются от 1 до 24. В этом случае 2016 г. (предпоследний в исследуемом периоде) соответствует моментам времени 17, 18, 19 и 20.

Подставим в уравнение тренда, получим:

$$T_1 = 200 + 3 \cdot 17 = 251;$$

$$T_2 = 200 + 3 \cdot 18 = 254;$$

$$T_3 = 200 + 3 \cdot 19 = 257;$$

$$T_4 = 200 + 3 \cdot 20 = 260.$$

Далее недостающие величины для первого, второго и третьего кварталов вычисляем по балансу из уравнения (1) для аддитивной модели временного ряда:

$$S_1 = y_1 - T_1 - E_1 = 200 - 251 - (-11) = -40;$$

$$y_2 = T_2 + S_2 + E_2 = 254 + 15 + 5 = 274;$$

$$E_3 = y_3 - T_3 - S_3 = 250 - 257 - 32 = -39.$$

Осталось определить только величины для четвертого квартала, где известно только значение трендовой компоненты. В условиях задачи задан общий объем продаж за год. Поскольку известны продажи за три первых квартала, четвертый определяется легко:

$$y_4 = 1000 - (y_1 + y_2 + y_3) = 1000 - (200 + 274 + 250) = 276.$$

Для расчета сезонной компоненты за 4 – й квартал воспользуемся тем, что в аддитивной модели сумма сезонных компонент за один период должны равняться нулю:

$$S_4 = -(S_1 + S_2 + S_3) = -(40 + 15 + 32) = -7.$$

Последнее значение в таблице – случайную компоненту за 4 – й квартал – вычисляем по балансу из формулы (1), поскольку все остальные компоненты уже известны:

$$E_4 = y_4 - T_4 - S_4 = 276 - 260 + 7 = 23.$$

Квартал	Фактический объем продаж	Компонента аддитивной модели		
		трендовая	сезонная	случайная
1	2	3	4	5
1	200	251	- 40	-11
2	274	254	15	+ 5
3	250	257	32	- 39
4	276	260	- 7	+ 23

Задача 1.3. На основе поквартальных данных за 9 последних лет была построена мультипликативная модель некоторого временного ряда. Уравнение тренда в этой модели имеет вид:

$$T_1 = 10{,}8 + 0{,}1 \cdot t.$$

Скорректированные значения сезонной компоненты равны: в 1 – м квартале – 1,5; в 3 – м квартале – 0,6; в 4 – м квартале – 0,8.

Определить сезонную компоненту за 2 – й квартал и прогноз моделируемого показателя за 2 – й и 3 – й кварталы следующего года.

Решение: В мультипликативной модели сумма скорректированных сезонных компонент за один период должны равняться количеству этих коэффициентов, т.е. четырем. Отсюда находим недостающую сезонную компоненту за 2 – й квартал:

$$S_2 = 4 - (S_1 + S_3 + S_4) = 4 - (1{,}5 + 0{,}6 + 0{,}8) = 1{,}1.$$

Для прогнозирования по мультипликативной модели воспользуемся соотношением (2), в котором не будем учитывать случайную компоненту. При этом следует иметь в виду, что 2 – й и 3 – й кварталы будущего года будут относиться в рамках рассматриваемой модели соответственно к 38 – й и 39 – й отметкам времени соответственно:

$$\hat{y}_{38} = (10{,}8 + 0{,}1 \cdot 38) \cdot 1{,}1 = 16{,}06;$$

$$\hat{y}_{39} = (10{,}8 + 0{,}1 \cdot 39) \cdot 0{,}6 = 8{,}82.$$

Задача 1.4. На основе помесячных данных за последние 5 лет была построена аддитивная временная модель потребления тепла в районе. Скорректированные значения сезонной компоненты приведены в таблице

Январь	+ 27	Май	- 20	Сентябрь	- 10
Февраль	+ 22	Июнь	- 34	Октябрь	+ 12
Март	+ 15	Июль	- 42	Ноябрь	+20
Апрель	- 2	Август	- 18	Декабрь	?

Уравнение тренда выглядит так:

$$T = 300 + 1{,}1 \cdot t.$$

Определить значение сезонной компоненты за декабрь, а также точечный прогноз потребления тепла на 2 – й квартал следующего года.

Решение: В аддитивной модели временного ряда сумма скорректированных сезонных компонент за один период, в данном случае за год, должна равняться нулю. Отсюда значение сезонной компоненты за декабрь:

$$S_{12} = 0 - \sum_{i=1,(i \neq 12)}^{12} S_i = 0 - (27 + 22 + 15 - 2 - 20 - 34 - 42 - 18 - 10 + 12 + 20) = -30.$$

Прогноз потребления тепла рассчитывается по формуле (1), в которой не учитывается случайная составляющая, поскольку она не прогнозируется. Здесь для расчета трендовой компоненты следует иметь в виду, что второму кварталу следующего года (апрель, май, июнь) соответствуют отметки времени 64, 65 и 66. Прогноз за весь второй квартал складывается из прогнозов за апрель, май и июнь.

$$\hat{y}(\textit{апрель}) = (300 + 1{,}1 \cdot 64) - 2 = 368{,}4;$$

$$\hat{y}(\textit{май}) = (300 + 1{,}1 \cdot 65) - 20 = 351{,}5;$$

$$\hat{y}(\textit{июнь}) = (300 + 1{,}1 \cdot 66) - 34 = 338{,}6;$$

$$\hat{y}(2 - \textit{й квартал}) = 368{,}4 + 351{,}5 + 338{,}6 = 1058{,}5.$$

Упражнения с пояснениями

По предприятию имеются поквартальные данные о выручке от продаж за 4 года, млн. руб.:

t	1	2	3	4	5	6	7	8
Yt	6	4,4	5	9	7,2	4,8	6	10
t	9	10	11	12	13	14	15	16
Yt	8	5,6	6,4	11	9	6,6	7	10,8

Упражнение 5.1. По имеющимся исходным данным определить скорректированную сезонную компоненту S_t.

Решение: Выполним графическую визуализацию исходного временного ряда дохода предприятия. Используем из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами.

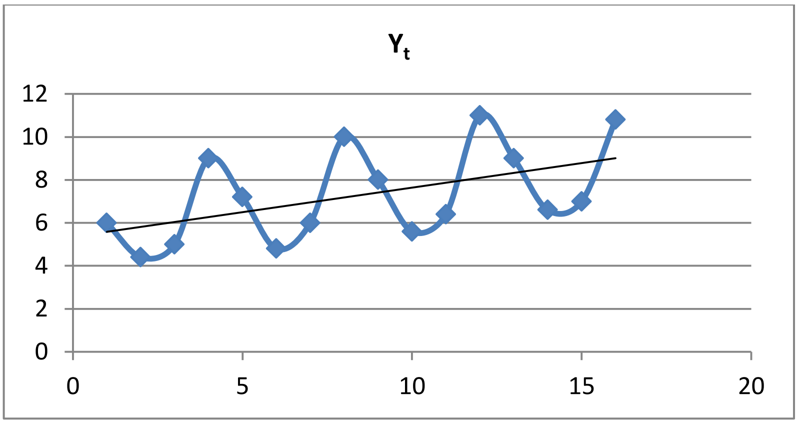


Рис. 1. Динамика дохода предприятия за 4 года

Вывод 1: График временного ряда демонстрирует линейный тренд с периодически повторяющейся сезонной волной примерно одинаковой амплитуды. В силу наличия тренда математическое ожидание дохода зависит от времени, и, следовательно, процесс является нестационарным. Для прогноза представляется возможным применить аддитивную (в силу примерно одинаковой амплитуды в сезонной волне) тренд-сезонную модель:

$$Y_t = T_t + S_t + E_t$$

Чтобы выполнить расчет скользящей средней за 4 квартала в расчетной таблице 1 надо последовательно просуммировать уровни ряда за каждые четыре квартала со сдвигом на один момент времени и затем разделить каждую полученную сумму на 4 согласно формулам:

$$\bar{Y}_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{6 + 4,4 + 5 + 9}{4} = 6,1$$

$$\bar{Y}_4 = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{4,4 + 5 + 9 + 7,2}{4} = 6,4$$

$$\bar{Y}_5 = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} = \frac{5 + 9 + 7,2 + 4,8}{4} = 6,5$$

и т. д.

Чтобы выполнить расчет центрированной скользящей средней в расчетной таблице 1 необходимо определить средние значения из двух последовательных скользящих средних:

$$\bar{Y}_3^* = \frac{\bar{Y}_3 + \bar{Y}_4}{2} = \frac{6,1+6,4}{2} = 6,25$$

$$\bar{Y}_4^* = \frac{\bar{Y}_4 + \bar{Y}_5}{2} = \frac{6,4+6,5}{2} = 6,45$$

и т. д.

Чтобы рассчитать оценку сезонной вариации *Is* необходимо найти разность между уровнями и центрированными скользящими средними:

$$Is_3 = Y_3 - \bar{Y}_3^* = 5 - 6,25 = -1,25$$

$$Is_4 = Y_4 - \bar{Y}_4^* = 9 - 6,45 = 2,55$$

и т. д.

Табл. 1

Подготовка данных для построения аддитивной тренд-сезонной модели

t	Y_t	Скользящая средняя за 4 квартала	Центрированная скольз. средняя	Оценка сезонной вариации	S_t	Y_t-S_t = T_t+E_t	T_t	T_t+S_t	E_t=Y_t-(T_t+S_t)
1	6				0,58				
2	4,4				-1,98				
3	5	6,1	6,25	-1,25	-1,29				
4	9	6,4	6,45	2,55	2,69				
5	7,2	6,5	6,625	0,575	0,58				
6	4,8	6,75	6,875	-2,075	-1,98				
7	6	7	7,1	-1,1	-1,29				
8	10	7,2	7,3	2,7	2,69				
9	8	7,4	7,45	0,55	0,58				
10	5,6	7,5	7,625	-2,025	-1,98				
11	6,4	7,75	7,875	-1,475	-1,29				
12	11	8	8,125	2,875	2,69				
13	9	8,25	8,325	0,675	0,58				
14	6,6	8,4	8,375	-1,775	-1,98				
15	7	8,35			-1,29				
16	10,8				2,69				

Расчет скорректированной сезонной компоненты S_t необходимо выполнить в расчетной таблице 2, которую можно заполнить в 4 шага.

Шаг 1: оценку сезонной вариации *Is* необходимо разместить согласно соответствующему году и соответствующему кварталу в расчетной табл. 2;

Шаг 2: суммы и средние по столбцам в расчетной табл. 2 необходимо определить с помощью функций *СУММ(…) и СРЗНАЧ(…).*

Комментарий: сумма средних сезонных компонент в аддитивной модели равна нулю;

Шаг 3: необходимо определить сумму по строке «Среднее» в расчетной таблице 2. Она, к сожалению, не равна нулю. Поэтому определим коэффициент корректировки:

$$k = \frac{0,075}{4} = 0,01875$$

Шаг 4: из средней сезонной компоненты для каждого квартала вычтем коэффициент корректировки, равный 0,01875. Полученный результат из последней строки расчетной табл. 2 перенесем в столбец S_t - Сезонная компонента, расчетной табл. 1.

Табл. 2

Скорректированная сезонная компонента

год	№квартала				
	1	*2*	*3*	*4*	
1			-1,25	2,55	
2	0,575	-2,075	-1,1	2,7	
3	0,55	-2,025	-1,475	2,875	
4	0,675	-1,775			
Итого	1,8	-5,875	-3,825	8,125	
Среднее	0,6	-1,9583	-1,275	2,70833	0,075
S_t	0,58125	-1,9770	-1,2937	2,68958	0

Упражнение 5.2. Определить трендовую компоненту Tt, случайную компоненту Et и представить их графически.

Решение: В расчетной таблице 1 необходимо заполнить столбец $Y_t - S_t = T_t + E_t$.
В расчетной таблице 1 необходимо заполнить столбец Tt – Трендовая компонента, применив статистическую функцию *Тенденция* к данным из столбца $Y_t - S_t = T_t + E_t$. В Главном меню MS Excel выберем: Формулы – Вставить функцию – Статистические – ТЕНДЕНЦИЯ(…).

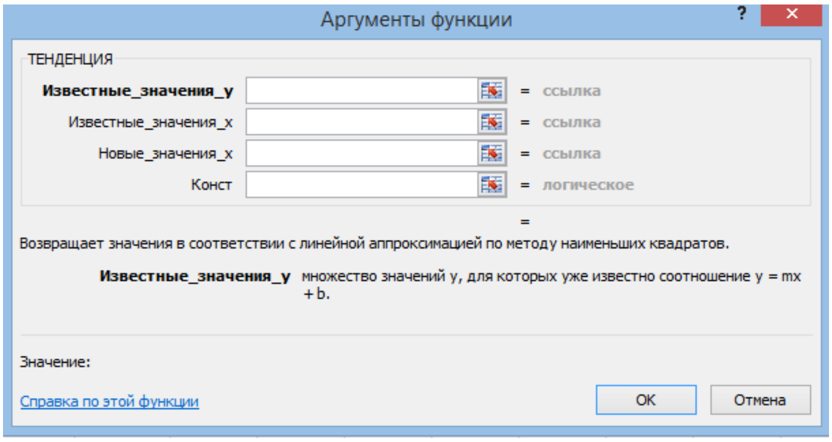


Рис. 2. Окно параметров функции ТЕНДЕНЦИЯ (…) в Excel

Чтобы определить трендовую компоненту T_t в поле *Известные значения Y* необходимо выбрать столбец $Y_t - S_t = T_t + E_t$. В поле *Известные значения X* необходимо выбрать столбец t, который содержит сквозной порядковый номер времени (в нашем случае квартала). Оба поля надо зафиксировать значком $. В поле *Новые значения X* необходимо выбрать ячейку для t=1. В поле Константа необходимо ввести 1 (что соответствует логическому значению *Истина*), чтобы уравнение тренда вычислить обычным образом, со свободным коэффициентом (рис 3).

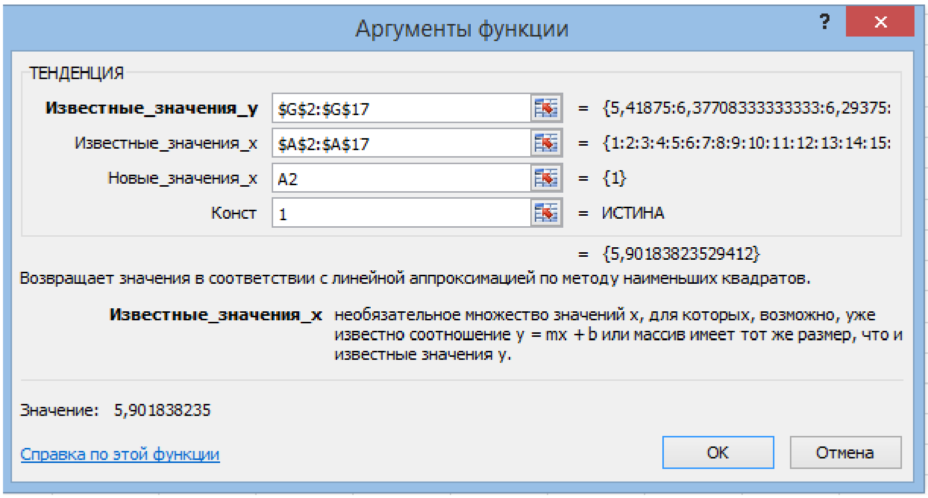


Рис.3. Окно параметров функции ТЕНДЕНЦИЯ (…) с введенными значениями в Excel

В расчетной таблице 1 необходимо заполнить столбец T+S и столбец E=Y-(T+S) для t=1…16.

Табл. 1

Подготовка данных для построения аддитивной тренд-сезонной модели

t	Y	Скользя-щая средняя за 4 квартала	Центрирован-ная скольз. средняя	Оценка сезонной вари-ации	S	Y-S =T+E	T	T+S	E=Y-(T+S)
1	6				0,58	5,42	5,90	6,48	-0,48
2	4,4				-1,98	6,38	6,09	4,11	0,29
3	5	6,1	6,25	-1,25	-1,29	6,29	6,27	4,98	0,02
4	9	6,4	6,45	2,55	2,69	6,31	6,46	9,15	-0,15
5	7,2	6,5	6,625	0,575	0,58	6,62	6,65	7,23	-0,03
6	4,8	6,75	6,875	-2,075	-1,98	6,78	6,83	4,86	-0,06
7	6	7	7,1	-1,1	-1,29	7,29	7,02	5,73	0,27
8	10	7,2	7,3	2,7	2,69	7,31	7,21	9,90	0,10
9	8	7,4	7,45	0,55	0,58	7,42	7,39	7,97	0,03
10	5,6	7,5	7,625	-2,025	-1,98	7,58	7,58	5,60	0,00
11	6,4	7,75	7,875	-1,475	-1,29	7,69	7,77	6,47	-0,07
12	11	8	8,125	2,875	2,69	8,31	7,95	10,64	0,36
13	9	8,25	8,325	0,675	0,58	8,42	8,14	8,72	0,28
14	6,6	8,4	8,375	-1,775	-1,98	8,58	8,33	6,35	0,25
15	7	8,35			-1,29	8,29	8,51	7,22	-0,22
16	10,8				2,69	8,11	8,70	11,39	-0,59

Используя из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами, представим графически компоненты: сезонную - S_t, трендовую - T_t, случайную - E_t.

Рис. 4. Сезонная компонента в доходе предприятия за 4 года

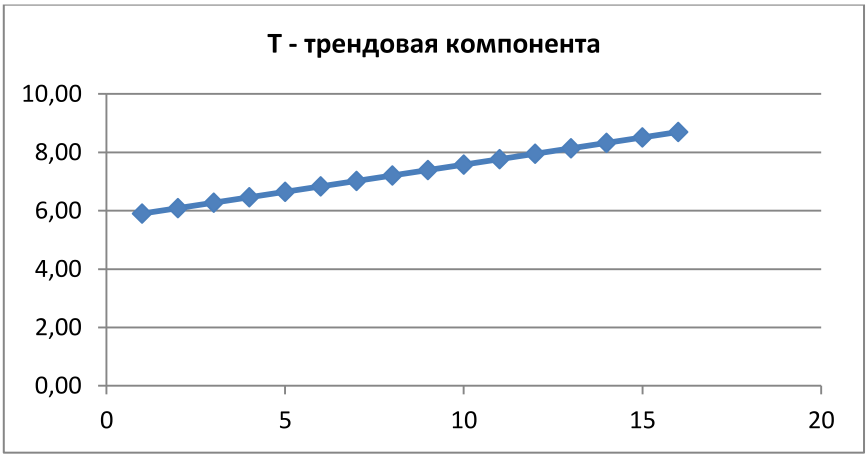


Рис. 5. Трендовая компонента в доходе предприятия за 4 года

Рис. 6. Случайная компонента в доходе предприятия за 4 года

Упражнение 5.3. Получить прогноз выручки от продаж в следующем году.

Решение: Необходимо продлить переменную t значениями 17, 18, 19, 20. Затем надо получить прогноз тренда T, продлив функцию Тенденция на значение t=17, затем на значение t=18, t=19, t=20. Получить прогноз дохода, просуммировав $T_{17} + S_1$, $T_{18} + S_2$, $T_{19} + S_3$, $T_{20} + S_4$ соответственно.

Табл. 1

Подготовка данных для построения аддитивной тренд-сезонной модели

t	Y	Скользящая средняя за 4 квартала	Центрированная скольз.средняя	Оценка сезонной вариации	S	Y-S =T+E	T	T+S	E=Y-(T+S)
1	6				0,58	5,42	5,90	6,48	-0,48
2	4,4				-1,98	6,38	6,09	4,11	0,29
3	5	6,1	6,25	-1,25	-1,29	6,29	6,27	4,98	0,02
4	9	6,4	6,45	2,55	2,69	6,31	6,46	9,15	-0,15
5	7,2	6,5	6,625	0,575	0,58	6,62	6,65	7,23	-0,03
6	4,8	6,75	6,875	-2,075	-1,98	6,78	6,83	4,86	-0,06
7	6	7	7,1	-1,1	-1,29	7,29	7,02	5,73	0,27
8	10	7,2	7,3	2,7	2,69	7,31	7,21	9,90	0,10
9	8	7,4	7,45	0,55	0,58	7,42	7,39	7,97	0,03
10	5,6	7,5	7,625	-2,025	-1,98	7,58	7,58	5,60	0,00
11	6,4	7,75	7,875	-1,475	-1,29	7,69	7,77	6,47	-0,07
12	11	8	8,125	2,875	2,69	8,31	7,95	10,64	0,36
13	9	8,25	8,325	0,675	0,58	8,42	8,14	8,72	0,28
14	6,6	8,4	8,375	-1,775	-1,98	8,58	8,33	6,35	0,25
15	7	8,35			-1,29	8,29	8,51	7,22	-0,22
16	10,8				2,69	8,11	8,70	11,39	-0,59
17					0,58		8,88	9,47	
18					-1,98		9,07	7,09	
19					-1,29		9,26	7,96	
20					2,69		9,44	12,13	

Используя из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами, представим графически прогноз дохода предприятия на 4 квартала следующего года (красным цветом).

Вывод 2: Предварительное выделение сезонной и трендовой компоненты позволило спрогнозировать снижение дохода в 1 и 2 квартале и сохранить в прогнозе сезонную волну и растущий тренд.

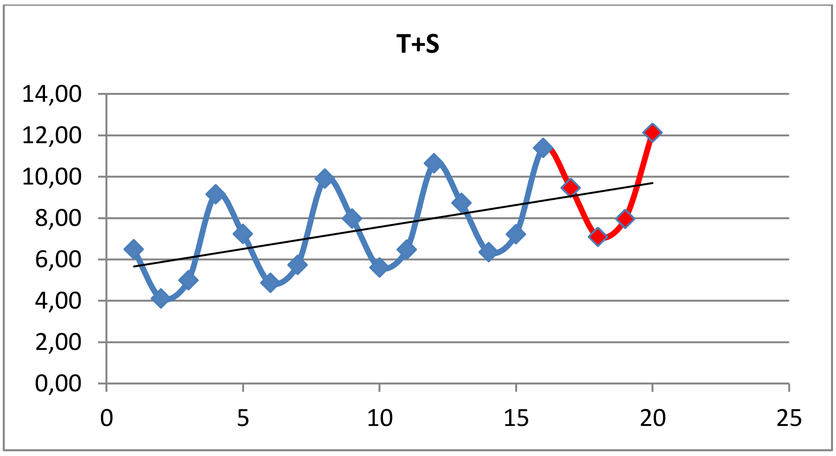


Рис. 7. Прогноз дохода предприятия на 4 квартала следующего года

Упражнение 5.4. По имеющимся поквартальные данным о прибыли предприятия за 4 года, млн. руб., определить скорректированную сезонную компоненту S_t.

t	1	2	3	4	5	6	7	8
Yt	40	50	60	70	60	80	100	110
t	9	10	11	12	13	14	15	16
Yt	50	70	80	130	30	50	60	70

Решение: Выполним графическую визуализацию исходного временного ряда дохода предприятия. Используем из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами.

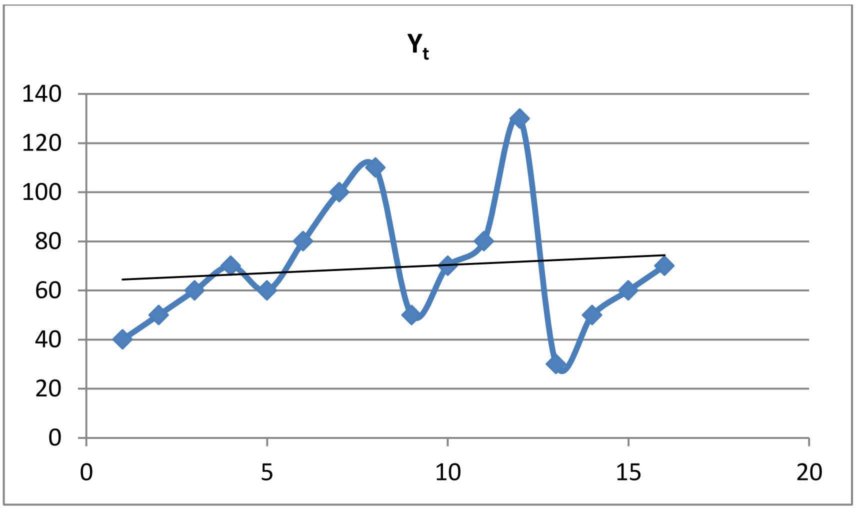


Рис. 1. Динамика прибыли предприятия за 4 года

Вывод 1: График временного ряда демонстрирует линейный тренд с сезонной волной растущей амплитуды. В силу наличия тренда и растущей амплитуды сезонных колебаний математическое ожидание прибыли зависит от времени, и, следовательно, процесс является нестационарным. Для прогноза представляется возможным применить мультипликативную (в силу растущей амплитуды в сезонной волне) тренд-сезонную модель:

$$Y_t = T_t * S_t * E_t$$

Чтобы выполнить расчет скользящей средней за 4 квартала в расчетной табл. 1 надо последовательно просуммировать уровни ряда за каждые четыре квартала со сдвигом на один момент времени и затем разделить каждую полученную сумму на 4 согласно формулам:

$$\bar{Y}_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{40 + 50 + 60 + 70}{4} = 55$$

$$\bar{Y}_4 = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{50 + 60 + 70 + 60}{4} = 60$$

$$\bar{Y}_5 = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} = \frac{60 + 70 + 60 + 80}{4} = 67{,}5$$

и т. д.

Чтобы выполнить расчет центрированной скользящей средней в расчетной табл. 1 необходимо определить средние значения из двух последовательных скользящих средних:

$$\overline{Y}_3^* = \frac{\overline{Y}_3 + \overline{Y}_4}{2} = \frac{55+60}{2} = 57,5$$

$$\overline{Y}_4^* = \frac{\overline{Y}_4 + \overline{Y}_5}{2} = \frac{60+67,5}{2} = 63,75$$

и т. д.

Чтобы рассчитать оценку сезонной вариации *Is* необходимо разделить исходные уровни на центрированные скользящие средние:

$$Is_3 = Y_3 / \overline{Y}_3^* = 60 / 57,5 = 1,043$$

$$Is_4 = Y_4 / \overline{Y}_4^* = 70 / 63,75 = 1,098$$

и т. д.

Табл. 1

Подготовка данных для построения мультипликативной тренд-сезонной модели

t	Yt	Скользящая средняя за 4 квартала	Центрированная скольз. средняя	Оценка сезонной вариации	St	Yt/St = Tt*Et	Tt	Tt*St	Et = Yt/(Tt*St)
1	40				0,631956				
2	50				0,899657				
3	60	55	57,5	1,043478	1,075981				
4	70	60	63,75	1,098039	1,392405				
5	60	67,5	72,5	0,827586	0,631956				
6	80	77,5	82,5	0,969697	0,899657				
7	100	87,5	86,25	1,15942	1,075981				
8	110	85	83,75	1,313433	1,392405				
9	50	82,5	80	0,625	0,631956				
10	70	77,5	80	0,875	0,899657				
11	80	82,5	80	1	1,075981				
12	130	77,5	75	1,733333	1,392405				
13	30	72,5	70	0,428571	0,631956				
14	50	67,5	60	0,833333	0,899657				
15	60	52,5			1,075981				
16	70				1,392405				

Расчет скорректированной сезонной компоненты S_t необходимо выполнить в расчетной табл. 2, которую можно заполнить в 4 шага.

Шаг 1: оценку сезонной вариации *Is* необходимо разместить согласно соответствующему году и соответствующему кварталу в расчетной табл.2;

Шаг 2: суммы и средние по столбцам в расчетной табл. 2 необходимо определить с помощью функций *СУММ(...) и СРЗНАЧ(...).*

Комментарий: сумма средних сезонных компонент в мультипликативной модели равна количеству циклов в исследуемом периоде (нашем случае равна 4, т.к. в году четыре квартала);

Шаг 3: необходимо определить сумму по строке «Среднее» в расчетной табл. 2. Она, к сожалению, не равна 4. Поэтому определим коэффициент корректировки:

$$k = \frac{4}{3,969} = 1,008$$

Шаг 4: каждую среднюю сезонную компоненту для каждого квартала умножим на коэффициент корректировки, равный 1,008. Полученный результат из последней строки расчетной табл. 2 перенесем в столбец S_t - Сезонная компонента, расчетной табл. 1.

Табл. 2

Скорректированная сезонная компонента

	№квартала				
год	1	2	3	4	
1			1,043478	1,098039	
2	0,827586	0,969697	1,15942	1,313433	
3	0,625	0,875	1	1,733333	
4	0,428571	0,833333			
Средняя	0,627053	0,892677	1,067633	1,381602	3,968964
Скорр S	0,631956	0,899657	1,075981	1,392405	4

Упражнение 5.5. Определить трендовую компоненту Tt, случайную компоненту Et и представить их графически.

Решение. В расчетной таблице 1 необходимо заполнить столбец Y_t / S_t = T_t* E_t. В расчетной табл. 1 необходимо заполнить столбец Tt – Трендовая компонента, применив статистическую функцию *Тенденция* к данным из столбца Y_t / S_t = T_t* E_t. В Главном меню MS Excel выберем: Формулы – Вставить функцию – Статистические – ТЕНДЕНЦИЯ(…).

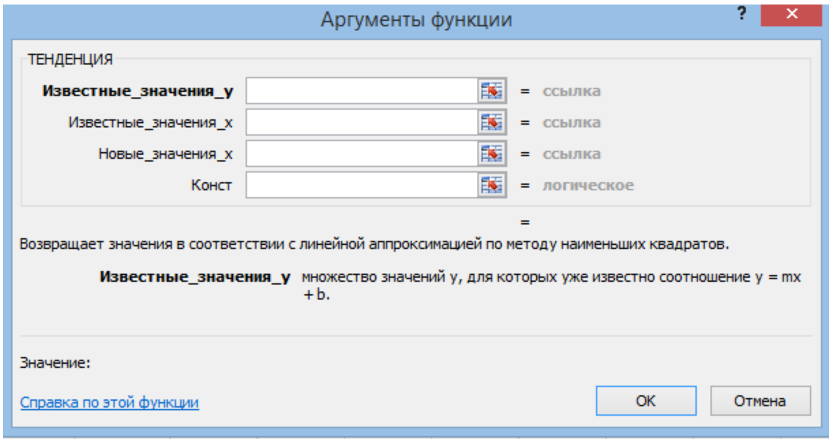


Рис. 2. Окно параметров функции ТЕНДЕНЦИЯ (…) в Excel

Чтобы определить трендовую компоненту T_t в поле *Известные значения Y* необходимо выбрать столбец Y_t / S_t = T_t* E_t. В поле *Известные значения X* необходимо выбрать столбец t, который содержит сквозной порядковый номер времени (в нашем случае квартала). Оба поля надо зафиксировать значком $. В поле *Новые значения X* необходимо выбрать ячейку для t=1. В поле Константа необходимо ввести 1 (что соответствует логическому значению *Истина*), чтобы уравнение тренда вычислить обычным образом, со свободным членом (рис 3).

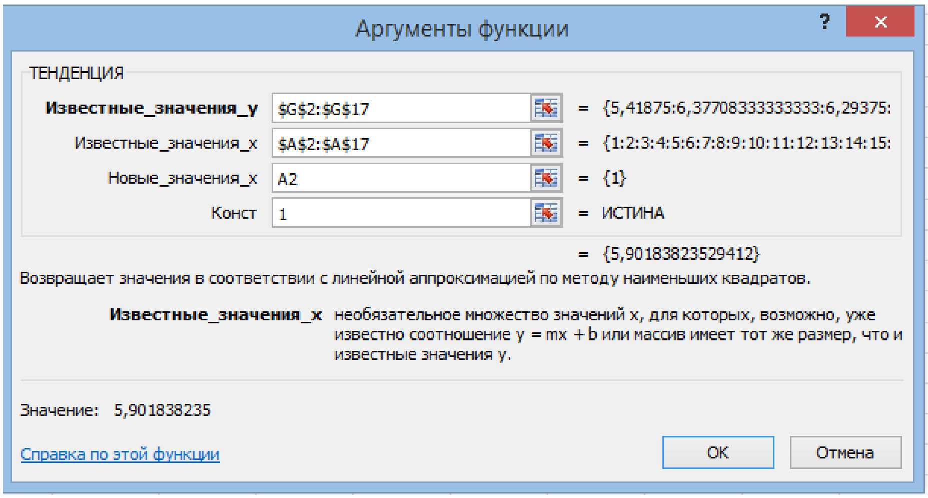


Рис.3. Окно параметров функции ТЕНДЕНЦИЯ (…) с введенными значениями в Excel

В расчетной табл. 1 необходимо заполнить столбец T*S и столбец E=Y/(T*S) для t=1…16.

Табл.1

Подготовка данных для построения мультипликативной тренд-сезонной модели

t	Yt	Скользящая средняя за 4 квартала	Центрированная скольз. средняя	Оценка сезонной вариации	St	Yt/St = Tt*Et	Tt	Tt*St	Et = Yt/ (Tt*St)
1	40				0,631956	63,2955	73,5057	46,45238	0,861097
2	50				0,899657	55,5767	72,9919	65,6677	0,761409
3	60	55	57,5	1,043478	1,075981	55,7630	72,4781	77,9851	0,769377
4	70	60	63,75	1,098039	1,392405	50,2727	71,9643	100,203	0,698578
5	60	67,5	72,5	0,827586	0,631956	94,9433	71,4506	45,1536	1,328797
6	80	77,5	82,5	0,969697	0,899657	88,9227	70,9368	63,8188	1,253549
7	100	87,5	86,25	1,15942	1,075981	92,9384	70,4230	75,7738	1,319716
8	110	85	83,75	1,313433	1,392405	78,9999	69,9092	97,3420	1,130036
9	50	82,5	80	0,625	0,631956	79,1194	69,3954	43,8548	1,140124
10	70	77,5	80	0,875	0,899657	77,8074	68,8816	61,9699	1,12958
11	80	82,5	80	1	1,075981	74,3507	68,3679	73,5626	1,087509
12	130	77,5	75	1,733333	1,392405	93,3636	67,8541	94,4804	1,375946
13	30	72,5	70	0,428571	0,631956	47,4716	67,3403	42,5561	0,704951
14	50	67,5	60	0,833333	0,899657	55,5767	66,8265	60,1210	0,831656
15	60	52,5			1,075981	55,7630	66,3127	71,3513	0,840909
16	70				1,392405	50,27271	65,7990	91,6189	0,764035

Используя из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами, представим графически компоненты: сезонную - S_t, трендовую - T_t, случайную - E_t.

Рис. 4. Сезонная компонента в прибыли предприятия за 4 года

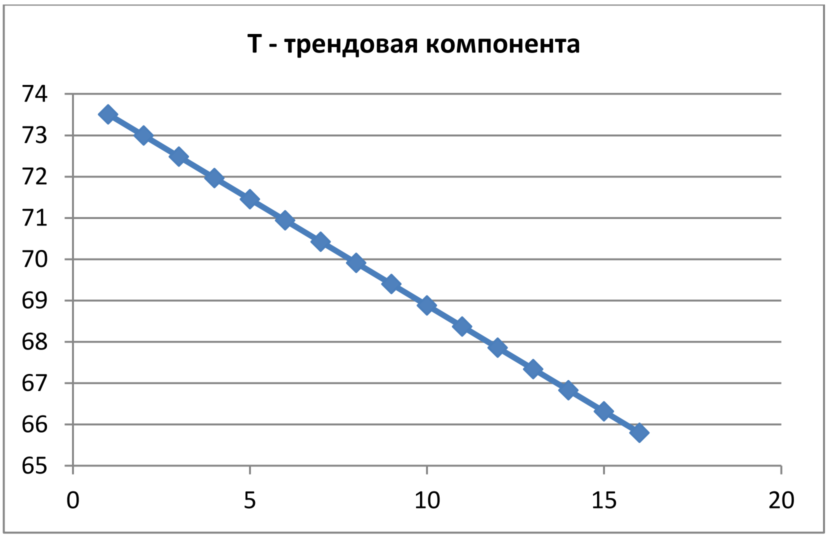


Рис. 5. Трендовая компонента в прибыли предприятия за 4 года

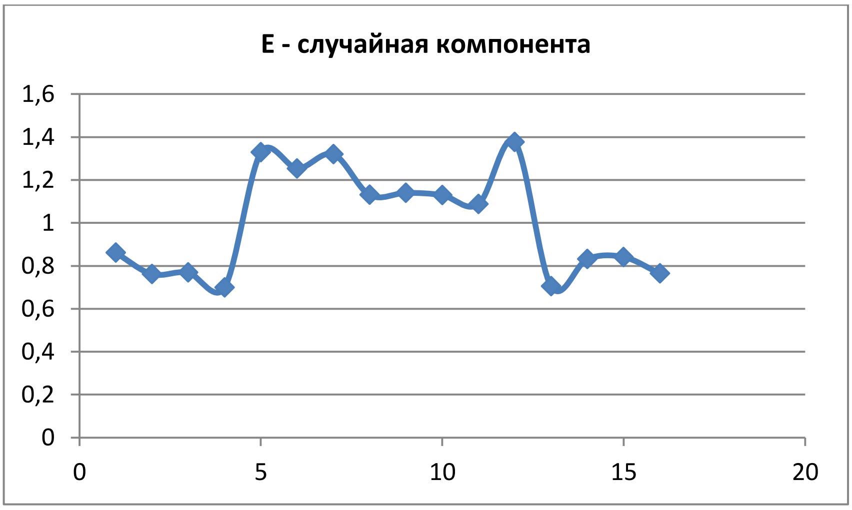


Рис. 6. Случайная компонента в прибыли предприятия за 4 года

Упражнение 5.6. Получить прогноз прибыли предприятия в следующем году.

Необходимо продлить переменную t значениями 17, 18, 19, 20. Затем надо получить прогноз тренда Т, продлив функцию Тенденция на значение t=17, затем на значение t=18, t=19, t=20. Получить прогноз дохода, умножив $T_{17} * S_1$, $T_{18} * S_2$, $T_{19} * S_3$, $T_{20} * S_4$ соответственно.

Табл. 1

Подготовка данных для построения мультипликативной тренд-сезонной модели

t	Yt	Скользящая средняя за 4 квартала	Центрированная скольз. средняя	Оценка сезонной вариации	St	Yt/St = Tt*Et	Tt	Tt*St	Et = Yt/(Tt*St)
1	40				0,631956	63,2955	73,5057 2	46,4523	0,861097
2	50				0,899657	55,5767	72,9919	65,6677	0,761409
3	60	55	57,5	1,043478	1,075981	55,7630	72,4781	77,9851	0,769377
4	70	60	63,75	1,098039	1,392405	50,2727	71,9643	100,203	0,698578
5	60	67,5	72,5	0,827586	0,631956	94,9433	71,4506	45,1536	1,328797
6	80	77,5	82,5	0,969697	0,899657	88,9227	70,9368	63,8188	1,253549
7	100	87,5	86,25	1,15942	1,075981	92,9384	70,4304	75,7738	1,319716
8	110	85	83,75	1,313433	1,392405	78,9999	69,9092	97,3420	1,130036
9	50	82,5	80	0,625	0,631956	79,1194	69,3954	43,8548	1,140124
10	70	77,5	80	0,875	0,899657	77,8074	68,8816	61,9699	1,12958
11	80	82,5	80	1	1,075981	74,3507	68,3679	73,5626	1,087509
12	130	77,5	75	1,733333	1,392405	93,3636	67,8541	94,4804	1,375946
13	30	72,5	70	0,428571	0,631956	47,4716	67,3403	42,5561	0,704951
14	50	67,5	60	0,833333	0,899657	55,5767	66,8265	60,1210	0,831656
15	60	52,5			1,075981	55,7630	66,3127	71,3513	0,840909
16	70				1,392405	50,2727	65,7990	91,6189	0,764035
17					0,631956		65,2852	41,2573	
18					0,899657		64,7714	58,2721	
19					1,075981		64,2576	69,1400	
20					1,392405		63,7438	88,7573	

Используя из Главного меню MS Excel: Вставка – Диаграммы – Точечная – Точечная с гладкими кривыми и маркерами, представим графически прогноз прибыли предприятия на 4 квартала следующего года (красным цветом).

Вывод 2: Предварительное выделение сезонной и трендовой компоненты позволило спрогнозировать снижение прибыли в 1 квартале, сохранить в прогнозе сезонную волну.

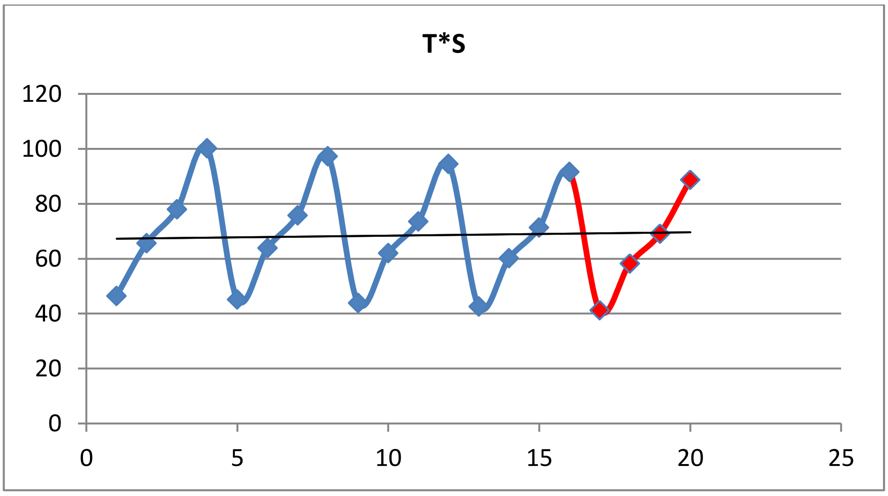


Рис. 7. Прогноз прибыли предприятия на 4 квартала следующего года

Рекомендуемая литература

1. Эконометрика: Учеб. пособие / Л.Е. Басовский. М.: РИОР: ИНФРА-М, 2017. 48 с. (ВО: Бакалавриат).http://znanium.com/catalog/product/559446

2. Эконометрика : учебник / В.А. Колемаев. ? М. : ИНФРА-М, 2017. 160 с. (Высшее образование: Бакалавриат). http://znanium.com/catalog/product/768143

3. Эконометрика: теоретические основы: Учебное пособие / Г.А. Соколов. - М.: ИНФРА-М, 2012. - 216 с.: 60x90 1/16 + CD-ROM. - (Высшее образование). (переплет, cd rom) ISBN 978-5-16-004180- http://znanium.com/catalog/product/243046

4. Эконометрика. Практикум: Учебное пособие / С.А. Бородич. - М.: НИЦ ИНФРА-М; Мн.: Нов. знание, 2014. - 329 с.: ил.; 60x90 1/16. - (Высшее образование: Бакалавриат). (переплет) ISBN 978-5-16-009429-8. http://znanium.com/catalog/product/440758

5. Эконометрика : теория и практика : учеб. пособие / Л.И. Ниворожкина, С.В. Арженовский, Е.П. Кокина. М. : РИОР : ИНФРА-М, 2018. 207 с. (Высшее образование). http://znanium.com/catalog/product/907587

6. Эконометрика и эконометрическое моделирование: учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. - М. : Вузовский учебник : ИНФРА-М, 2018. - 385 с. : ил. (Высшее образование: Бакалавриат). http://znanium.com/catalog/product/968797

7. Практическая эконометрика в кейсах : учеб. пособие / В.П. Невежин, Ю.В. Невежин. М. : ИД 'ФОРУМ' : ИНФРА-М, 2017. 317 с. + Доп. материалы [Электронный ресурс; Режим доступа: http://www.znanium.com]. (Высшее образование). http://znanium.com/catalog/product/752452

8. Исмагилов И. И., Кадочникова Е. И., Костромин А. В., Бадриева Л. Д., Хасанова С. Ф. Эконометрика. Конспект лекций. КФУ, Казань, 2014 г. 236 с. https://dspace.kpfu.ru/xmlui/bitstream/handle/net/21939/72_182_kl-000832.pdf

9. Исмагилов И. И., Кадочникова Е. И., Костромин А. В., Бадриева Л. Д. Эконометрика. Учебно-методическое пособие для самостоятельной работы. КФУ, Казань, 2014 г. 51 с. https://dspace.kpfu.ru/xmlui/bitstream/handle/net/21880/72_200_A5-00781.pdf

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

Таблица значений *F*-критерия Фишера при уровне значимости $\alpha = 0{,}05$

k_1 / k_2	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	234,52
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65

29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,65	1,31
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Критические значения *t*-критерия Стьюдента при уровне значимости 0,10; 0,05;0,01 (двухсторонний)

Число степеней свободы d.f.	0,10	0,05	0,01	Число степеней свободы d.f.	0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Статистика Дарбина-Уотсона: d_1 и d_2 для 5% уровня значимости

n – число наблюдений,

k – число объясняющих переменных (без учета постоянного члена)

n	k=1		k=2		k=3		k=4		k=5	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77

60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78