ОСНОВНЫЕ ПОЛОЖЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ

Цель работы

Целью работы является приобретение навыков использования основных приемов обработки экспериментальных статистических данных, построение эконометрических моделей и их решение в среде Microsoft Excel, осуществление эконометрического анализа исходной экономической задачи.

Руководство к выполнению лабораторной работы

Для выполнения лабораторной работы каждый студент обязан:

1) повторить теоретический материал, относящийся к данному занятию;

2) изучить технологию решения задач с помощью Excel, руководство к лабораторной работе, уяснить основную задачу занятия, методику и порядок ее выполнения;

3) по номеру своего варианта выбрать экономическую задачу и, согласно рассмотренным задачам, провести эконометрический анализ;

4) в процессе работы должен руководствоваться описанием лабораторной работы, строго придерживаясь рекомендованного порядка ее проведения;

5) после выполнения определенного объема работы, должен самостоятельно выполнить предложенное ему зачетное задание и сдать его в распечатанном виде с подробным эконометрическим анализом преподавателю.

План лабораторных занятий

	Тема	Кол-во часов
Лабораторная работа № 1	Простейшая обработка данных. Линейная регрессия. Коэффициент корреляции	2
Лабораторная работа № 2	Проверка качества уравнения линейной регрессии	2
Лабораторная работа № 3	Нелинейные модели. Коэффициент детерминации	3
Лабораторная работа № 4	Прогнозирование на основании линейной регрессии	2
Лабораторная работа № 5	Многофакторная линейная регрессия. Мультиколлинеарность	2
Лабораторная работа № 6	Построение линейного, логарифмического, полиномиального, степенного и экспоненциального трендов	2
Лабораторная работа № 7	Автокорреляция уровней временного ряда и выявление его структуры	2
Лабораторная работа № 8	Анализ сезонных колебаний	2

Лабораторная работа №1

Простейшая обработка данных. Линейная регрессия. Коэффициент корреляции. Его значимость

Цель: научиться находить коэффициент корреляции и определять его значимость; находить коэффициенты регрессии и строить уравнение регрессии.

Основные сведения

Парная регрессия – это уравнение связи двух переменных y и x:

$$y=f(x),$$

где y – зависимая переменная (результат, отклик);
x – независимая, объясняющая переменная (фактор).

Различают *линейные и нелинейные* регрессии.

Линейная регрессия: $\hat{y}_x = a + b \cdot x$.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют *метод наименьших квадратов* (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических y_x минимальна.

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и $в$:

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

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b \cdot \bar{x}, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)},$$

где

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x} \cdot \bar{y}$$ – ковариация признаков x и y,

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2$$ – дисперсия признака x и

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \ldots + y_n}{n},$$

$$\overline{y \cdot x} = \frac{1}{n}\sum y \cdot x, \quad \overline{x^2} = \frac{1}{n}\sum x^2$$

$$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2 = \sigma_x^2, \quad \operatorname{var}(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \overline{y^2} - \bar{y}^2 = \sigma_y^2,$$

$$\sigma_y = \sqrt{\frac{\sum_{k=1}^{n}\left(y_k - \bar{y}\right)^2}{n}} = \sqrt{\mathrm{var}(y)}, \qquad \sigma_x = \sqrt{\frac{\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}{n}} = \sqrt{\mathrm{var}(x)}.$$

Параметр b называется *коэффициентом регрессии*. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции r_{xy} для линейной регрессии ($-1 \le r_{xy} \le 1$):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\mathrm{cov}\left(x,\, y\right)}{\sigma_x \cdot \sigma_y} = \frac{\overline{yx} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\,\mathrm{var}(y)}}.$$

Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

Теснота связи	Значение коэффициента корреляции при наличии:	
	Прямой связи	Обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Весьма высокая	0,9–1	(–1)–(–0,9)

Положительное значение коэффициента корреляции говорит о положительной связи между *x* и *y*, когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, с ростом одной из переменных другая убывает, с убыванием одной из переменной другая растет.

Оценку статистической значимости коэффициента корреляции проводят с помощью *t*-критерия Стьюдента. Выдвигают гипотезу H_0 о статистически незначимом отличии коэффициента от нуля. Оценка значимости коэффициента корреляции с помощью *t*-критерия Стьюдента проводится путем сопоставления его значения с величиной случайной ошибки:

$$t_r = r/m_r.$$

Стандартная (случайная) ошибка коэффициента корреляции определяется по формуле:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}.$$

Сравнивая фактическое и табличное (критическое) значения *t*-статистики – $t_{табл}$ и $t_{факт}$ – принимает или отвергаем гипотезу H_0.

Если $t_{табл} < t_{факт}$, то гипотеза H_0 отклоняется, коэффициент корреляции не случайно отличается от. Если $t_{табл} > t_{факт}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования коэффициента корреляции.

Порядок выполнения работы.

По заданной выборке исследовать зависимость результата *y* от фактора *x*. Для этого

1. Создать таблицу данных.
2. Найти средние значения $\overline{x}, \overline{y}$, выборочные дисперсии S_x^2, S_y^2, исправленные средние квадратические отклонения $\overline{S}_x, \overline{S}_y$.
3. Найти коэффициент корреляции и проверить его значимость.
4. Найти коэффициенты линейного уравнения регрессии.
5. Построить график прямой регрессии.

Пример выполнения лабораторной работы.

В табл. 1.1 приведены данные об объеме производства *y* (тыс.ед.) в зависимости от численности занятых *x* (тыс.чел.) некоторой фирмы.

Таблица 1.1.

Исходные данные

x	11	13	15	18	20	22	24	25	27
y	15	17	21	20	28	33	34	32	29

1. В диапазоне B3:C11 подготовим исходные данные.
2. Вводим следующие формулы:

Ячейка	Формула	Примечание
D3	=B3*C3	Копируем в диапазон D3:D11
E3	=B3*B3	Копируем в диапазон E3:E11
F3	=C3*C3	Копируем в диапазон F3:F11
B12	=СРЗНАЧ(B3:B11)	Копируем в диапазон B12:F12
A17	=E12-B12*B12	Выборочная средняя фактора
B17	=F12-C12*C12	Выборочная средняя результата
A20	=СТАНДОТКЛОН(B3:B11)	Исправленное среднее квадратическое отклонение фактора
B20	=СТАНДОТКЛОН(C3:C11)	Исправленное среднее квадратическое отклонение результата

Получим следующие результаты (см. рис. 1.1).

	A	B	C	D	E	F
1	**Простейшая обработка данных**					
2		x	y	xy	x^2	y^2
3	1	11	25	275	121	625
4	2	13	27	351	169	729
5	3	15	31	465	225	961
6	4	18	30	540	324	900
7	5	20	38	760	400	1444
8	6	22	43	946	484	1849
9	7	24	44	1056	576	1936
10	8	25	42	1050	625	1764
11	9	27	49	1323	729	2401
12	среднее значение	19,44	36,56	751,78	405,89	1401,00
13						
14						
15	**Выборочные средние**					
16	S_x^2	S_y^2				
17	27,80	64,69				
18	**Исправленные средние квадратические отклонения**					
19	$S_{xиспр}$	$S_{yиспр}$				
20	5,59	8,53				
21						

Рис. 1.1. Результаты простейшей обработки данных

3. Для определения коэффициента корреляции воспользуемся формулой $r_{xy} = \dfrac{\overline{yx} - \overline{x}\cdot\overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}}$. Для этого в ячейку **E16** вводим формулу

=(D12-B12*C12)/КОРЕНЬ(A17*B17)

Из расчетов следует, что коэффициент корреляции r=0,97. Это свидетельствует о том, что связь между объемом выпуска продукции и численностью занятых весьма высокая и положительная.

4. Для проверки значимости коэффициента корреляции введем вспомогательные данные:

Ячейки

К16	9	число предприятий;
К17	0,05	уровень значимости.

5. Далее вводим следующие формулы:

H19	=КОРЕНЬ((1-E16*E16)/(K16-2))	Стандартная ошибка
H20	=E16/H19	t-статистика
H21	=СТЬЮДРАСПОБР(K17;K16-2)	Критическое значение t-статистики
H22	=ЕСЛИ(ABS(H20)>H21;"Значим";"Незначим")	Вывод

Таким образом, получим данные, представленные на рис. 1.2.

	C	D	E	F	G	H	I	J	K
13									
14									
15		**Коэффициент корреляции**					**Вспомогательные данные**		
16		r	0,97				n		9
17							уровень значимости		0,05
18		**Проверка значимости коэффициента корреляции**							
19		стандаратная ошибка				0,10			
20		t-статистика				9,91			
21		Критическое значение t-статистики				2,36			
22		Вывод				Значим			
23									

Рис. 1.2. Анализ значимости коэффициента корреляции

6. Для определения коэффициентов уравнения линейной регрессии на основе формул

$$в = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2}; \; a = \overline{y} - в\overline{x},$$

следует в ячейки **I3**, **I4** ввести соответственно следующие формулы:

=(D12-B12*C12)/A17;

=C12-I3*B12.

Уравнение регрессии y=7,9+1,47x.

Значение коэффициента b=1,47 говорит о том, что при увеличении численности занятых на 1 тыс.чел. объем продукции увеличится на 1,74 тыс.ед.

Результаты расчетов приведены на рис.1.3.

Простейшая обработка данных

	x	y	xy	x^2	y^2
1	11	25	275	121	625
2	13	27	351	169	729
3	15	31	465	225	961
4	18	30	540	324	900
5	20	38	760	400	1444
6	22	43	946	484	1849
7	24	44	1056	576	1936
8	25	42	1050	625	1764
9	27	49	1323	729	2401
среднее значение	19,44	36,56	751,78	405,89	1401,00

Коэффициенты регрессии

в	1,47
а	7,90

Выборочные средние

S_x^2	S_y^2
27,80	64,69

Коэффициент корреляции

r	0,97

Вспомогательные данные

n	9
уровень значимости	0,05

Исправленные средние квадратические отклонения

$S_{xиспр}$	$S_{yиспр}$
5,59	8,53

Проверка значимости коэффициента корреляции

стандаратная ошибка	0,10
t-статистика	9,91
Критическое значение t-статистики	2,36
Вывод	Значим

Рис. 1.3. Результаты расчетов

7. Для построения графика выделим диапазон B3:C11. Вызовем **Мастер диаграмм**. Чтобы ось отражала фактические данные, выберем тип диаграммы **Точечная.** После чего нажмем кнопку **Готово**. На построенной диаграмме выделим график функции, щелкнув по нему левой кнопкой мыши. Выделение обозначается светлыми маркерами на функции. Нажав правую кнопку мыши, выведем контекстно-зависимое меню, в котором выберем опцию **Добавить линию тренда**. В окне **Линия тренда** по вкладке **Тип** выберем тип функции **Линейная**, а во вкладке **Параметры** – установим флажок **показывать уравнение на диаграмме**. В результате на диаграмме появиться вид теоретической кривой – тренда и ее уравнение (рис.1.4).

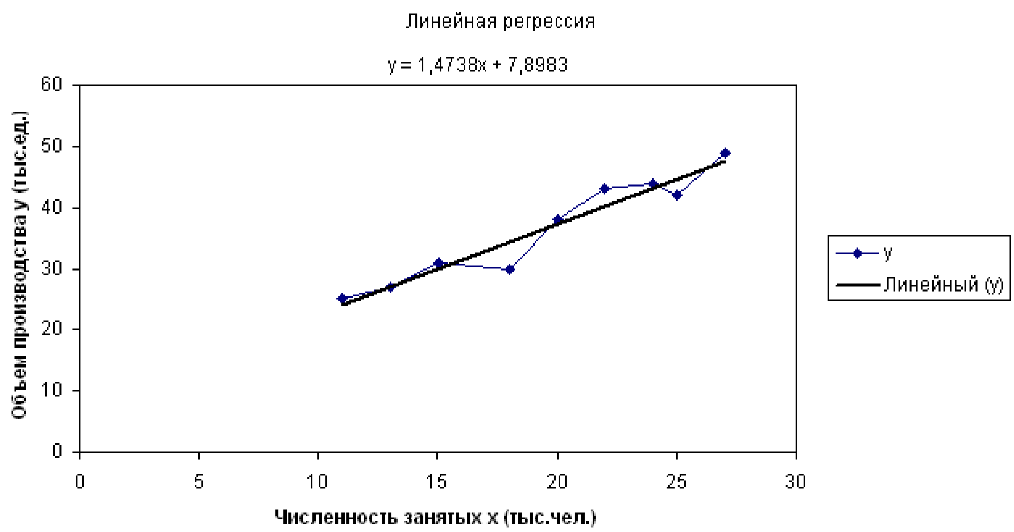


Рис. 1.4. Графики фактических данных и построенной регрессии

8. Вычисление параметров регрессии с помощью статистических функций Excel:

КОРРЕЛ(**массив1**;**массив2**) вычисляет коэффициент корреляции между двумя переменными; значения первой из них приведены в диапазоне массив1, значения второй – в диапазоне массив2;

НАКЛОН(известные_значения_y;известные_значения_x) служит для определения коэффициента *b*;

ОТРЕЗОК(известные_значения_y;известные_значения_x) служит для определения коэффициента *a*.

Вводим формулы:

C27	=КОРРЕЛ(B3:B11;C3:C11)	Коэффициент корреляции
C28	=НАКЛОН(C3:C11;B3:B11)	Коэффициент *b*
C29	=ОТРЕЗОК(C3:C11;B3:B11)	Коэффициент *a*

Встроенная статистическая функция **ЛИНЕЙН** определяет параметры линейной регрессии. Порядок вычислений следующий:

1) выделите область пустых ячеек 5x2 (5 строк, 2 столбца) с целью вывода результатов регрессионной статистики (A27:B3);

2) в главном меню выберите **Вставка/Функция**;

3) в строке **Категория** (рис.1.5) выберите **Статистические**, в окне **Функция** – **ЛИНЕЙН**. Щелкните **ОК**.

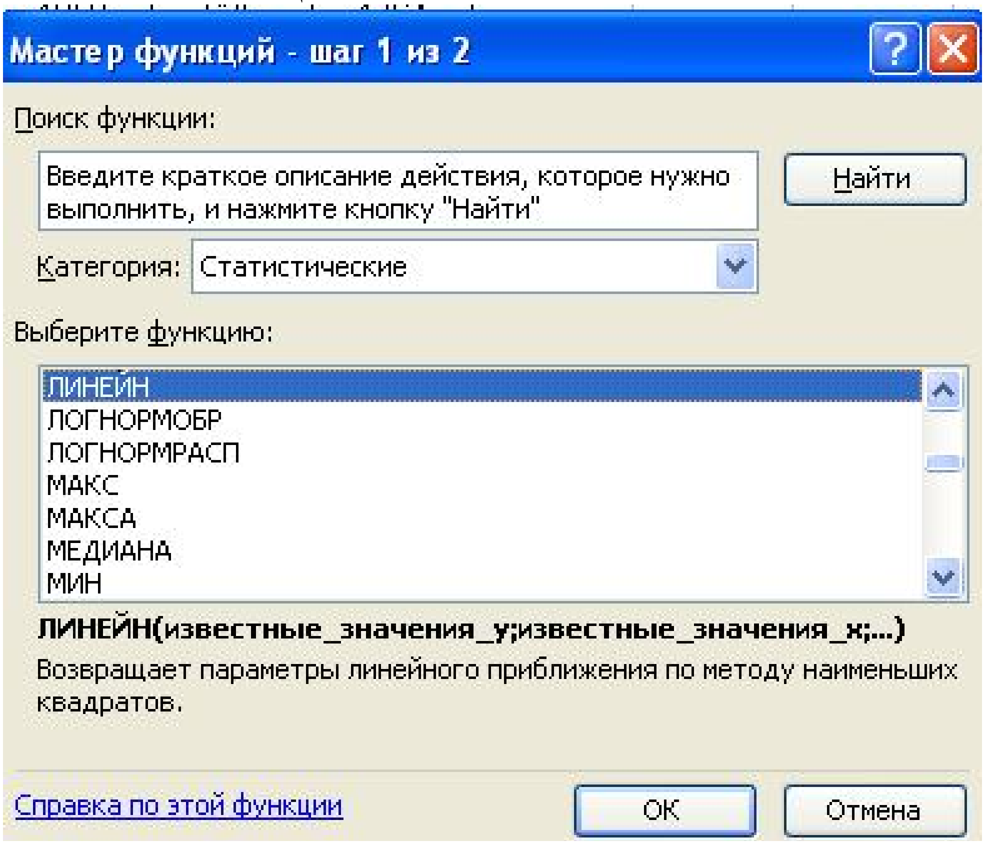


Рис. 1.5. Диалоговое окно «Мастер функций»

4) Заполните аргументы функции (рис.1.6.):

Известные_значения_y – диапазон, содержащий данные результативного признака;

Известные_значения_x – диапазон, содержащий данные факторов независимого признака;

Константа – логическое значение, которое указывает на наличие или на отсутствие свободного члена в уравнении; если *Константа* = 1, то свободный член рассчитывается обычным образом, если *Константа* = 0, то свободный член равен 0.

Статистика – логическое значение, которое указывает выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика* = 1, то дополнительная информация выводится, если *Статистика* = 0, то выводится только оценки параметров уравнения. Далее **ОК**.

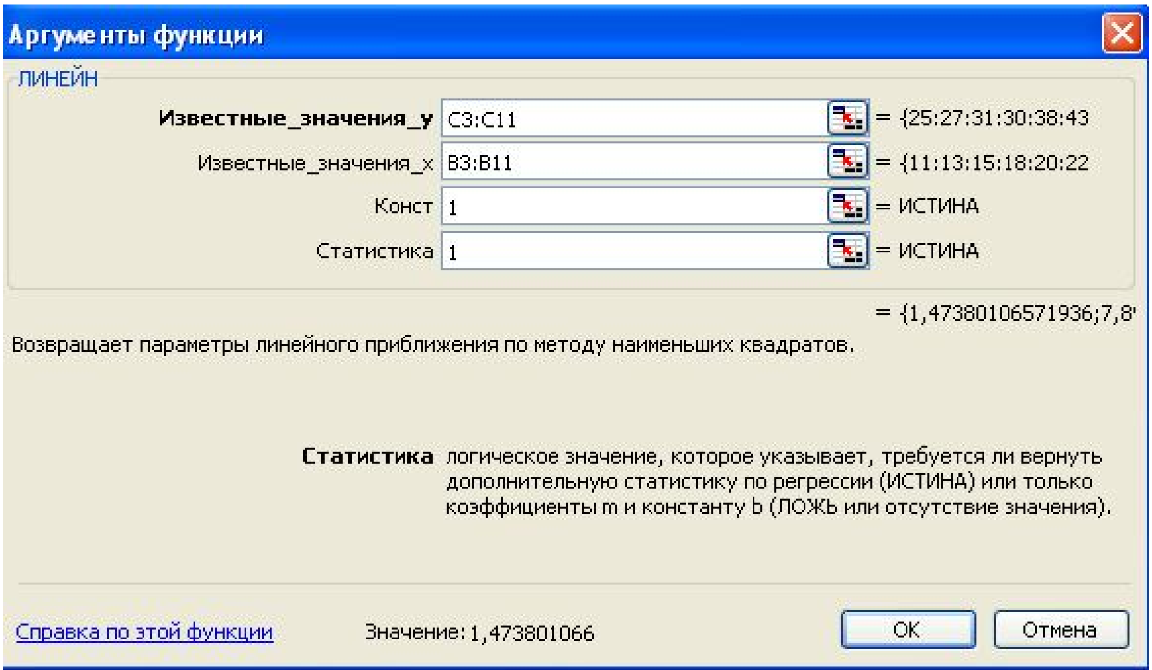


Рис.1.6. Диалоговое окно ввода аргументов функции **ЛИНЕЙН**

5) В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу **F2,** а затем – на комбинацию клавиш **CTRL+SHIFT+ENTER**. Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента *b*	Значение коэффициента *a*
Среднеквадратическое отклонение *b*	Среднеквадратическое отклонение *a*
Коэффициент детерминации R^2	Среднеквадратическое отклонение *y*
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов.

Результаты регрессионного анализа представлены на рис.1.7.

	A	B	C	D	E	F
25						
26		**Линейн**				
27	1,4738011	7,89831261	0,97	коэффициент корреляции		
28	0,1486756	2,99532023	1,47	b		
29	0,9335011	2,35181208	7,90	a		
30	98,264891	7				
31	543,50508	38,7171403				

Рис. 1.7.Результаты регрессионного анализа

Индивидуальное задание к лабораторной работе

По предприятиям легкой промышленности региона получена информация, характеризующая зависимость объема выпуска продукции (*y*, млн. руб.) от объема капиталовложений (*x*, млн. руб.)

№		1	2	3	4	5	6	7	8	9	10
1	*x*	66	58	73	82	81	84	55	67	81	59
	y	133	107	145	162	163	170	104	132	159	116
2	*x*	72	52	73	74	76	79	54	68	73	64
	y	121	84	119	117	129	128	102	111	112	98
3	*x*	38	28	27	37	46	27	41	39	28	44
	y	69	52	46	63	73	48	67	62	47	67
4	*x*	36	28	43	52	51	54	25	37	51	29
	y	104	77	117	137	143	144	82	101	132	77
5	*x*	31	23	38	47	46	49	20	32	46	24
	y	38	26	40	45	51	49	34	35	42	24
6	*x*	33	17	23	17	36	25	39	20	13	12
	y	43	27	32	29	45	35	47	32	22	24
7	*x*	36	28	43	52	51	54	25	37	51	29
	y	85	60	99	117	118	125	56	86	115	68
8	*x*	17	22	10	7	12	21	14	7	20	3
	y	26	27	22	19	21	26	20	15	30	13
9	*x*	12	4	18	27	26	29	1	13	26	5
	y	21	10	26	33	34	37	9	21	32	14
10	*x*	26	18	33	42	41	44	15	27	41	19
	y	43	28	51	62	63	67	26	43	61	33

Отчет по лабораторной работе

1. Запишите уравнение линейной парной регрессии для своего варианта и поясните экономическую сущность параметров уравнения.
2. Что является показателем тесноты связи в парной линейной регрессии?
3. Каково значение коэффициента корреляции?
4. Каково значение коэффициента детерминации и что он характеризует?
5. Как оценивается значимость коэффициента корреляции?
6. Является ли коэффициент корреляции для вашего варианта значимым и почему?

Лабораторная работа №2

Проверка качества уравнения линейной регрессии

Цель: научиться проверять статистическую значимость коэффициентов и общего качества уравнения линейной регрессии.

Основные сведения

Оценку качества построенной модели дает коэффициент (индекс) детерминации r_{xy}^2 (ρ_{xy}^2), а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

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

Допустимый предел значений средней ошибки аппроксимации – не более 8–10%.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения $\overline{y}$ раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum\left(y - \overline{y}\right)^2 = \sum\left(\hat{y}_x - \overline{y}\right)^2 + \sum\left(y - \hat{y}_x\right)^2,$$

где $\sum\left(y - \overline{y}\right)^2$ – общая сумма квадратов отклонений; $\sum\left(\hat{y}_x - \overline{y}\right)^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum\left(y - \hat{y}_x\right)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

$$\sum\left(y - \overline{y}\right)^2 = n\sigma_y^2; \ \sum\left(\hat{y}_x - \overline{y}\right)^2 = n\sigma_y^2 R^2 = b^2 n\sigma_x^2; \ \sum\left(y - \hat{y}_x\right)^2 = n\sigma_y^2(1 - R^2).$$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F-критерия Фишера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2}.$$

Фактическое значение F-критерия Фишера (1.9) сравнивается с табличным значением $F_{\text{табл}}\left(\alpha; k_1; k_2\right)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F-критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m=1$, поэтому

$$F=\frac{S^2_{\text{факт}}}{S^2_{\text{ост}}}=\frac{\sum\left(\hat{y}_x-\overline{y}\right)^2}{\sum\left(y-\hat{y}_x\right)^2}\cdot(n-2).$$

Величина F-критерия связана с коэффициентом детерминации r^2_{xy}, и ее можно рассчитать по следующей формуле:

$$F=\frac{r^2_{xy}}{1-r^2_{xy}}\cdot(n-2).$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b=\sqrt{\frac{S^2_{\text{ост}}}{\sum\left(x-\overline{x}\right)^2}}=\frac{S_{\text{ост}}}{\sigma_x\cdot\sqrt{n}},$$

где $S^2_{\text{ост}}=\dfrac{\sum\left(y-\hat{y}_x\right)^2}{n-2}$ – остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с t-распределением Стьюдента при $n-2$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение t-критерия Стьюдента: $t_b=\dfrac{b}{m_b}$ которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Доверительный интервал для коэффициента регрессии определяется как $b\pm t_{\text{табл}}\cdot m_b$.

Стандартная ошибка параметра a определяется по формуле:

$$m_a=\sqrt{S^2_{\text{ост}}\cdot\frac{\sum x^2}{n\cdot\sum\left(x-\overline{x}\right)^2}}=S_{\text{ост}}\cdot\frac{\sqrt{\sum x^2}}{\sigma_x\cdot n}.$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется t-критерий: $t_a=\dfrac{a}{m_a}$, его величина сравнивается с табличным значением при $n-2$ степенях свободы. Доверительный интервал для коэффициента регрессии определяется как $a \pm t_{\text{табл}} \cdot m_a$.

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, т.к. он не может одновременно принимать и положительное, и отрицательное значения.

Порядок выполнения лабораторной работы

Используя данные к лабораторной работе №1, найти уравнение линейной регрессии и проверить:

I. Значимость коэффициента *b*. Для этого надо найти:

1. Сумму квадратов остатков:

$$\sum e_i^2 = \sum (y - y_x)^2 .$$

2. Найти сумму квадратов отклонений:

$$\sum (x - \overline{x})^2 .$$

3. Стандартную ошибку параметра *b*:

$$m_b = \sqrt{\frac{\sum (y - y_x)^2}{(n-2)\sum (x - \overline{x})^2}}.$$

4. Наблюдаемой значение *t*-статистики параметра *b*:

$$t_b = b/m_b.$$

5. Число степеней свободы $k=n-2$ и критическое значение $t_{табл} = t_{\alpha;k}$.

6. Сделать вывод о значимости коэффициента *b*: если $t_{табл} < |t_b|$, то параметр регрессии *b* статистически значим, а в противном случае статистически незначим.

II. Значимость коэффициента *a*. Для этого надо найти:

1. Сумму квадратов:

$$\sum x^2 .$$

2. Стандартную ошибку параметра *a*:

$$m_a = m_b \sqrt{\frac{\sum x^2}{n}}.$$

3. Наблюдаемой значение *t*-статистики параметра *a*:

$$t_a = a/m_a.$$

4. Сделать вывод о значимости коэффициента *a*: если $t_{табл} < |t_a|$, то параметр регрессии *a* статистически значим, а в противном случае статистически незначим.

III. Общее качество уравнения регрессии. Для этого надо найти:

1. Сумму квадратов отклонений:

$$\sum (y-\overline{y})^2 .$$

2. Коэффициент детерминации:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y-\overline{y})^2}.$$

3. Наблюдаемое значение F-статистики:

$$F_{факт} = \frac{R^2}{1-R^2} \cdot (n-2).$$

4. Число степеней свободы критерия Фишера-Снедекора:
k_1=1; k_2= n–2 и критическое значение этого критерия $F_{табл}$= $F_{\alpha;k_1;k_2}$.

5. Сделать вывод о значимости уравнения регрессии: если $F_{факт}$ > $F_{табл}$, то уравнение регрессии статистически значимо и надежно, если $F_{факт}$ < $F_{табл}$ признается статистическая незначимость, ненадежность уравнения регрессии.

IV. Общее качество уравнения регрессии с помощью средней ошибки аппроксимации. Для этого надо найти:

1. Отклонения:

$$y - y_x .$$

2. A_i:

$$A_i = \left| \frac{y - y_x}{y} \right| \cdot 100\%.$$

3. Среднюю ошибку аппроксимации $\overline{A}$:

$$\overline{A} = \frac{1}{n} \cdot \sum A_i .$$

4. Сделать вывод о качестве уравнения регрессии: если $\overline{A}$ не превышает предела значений 8-10%, то качество модели хорошее.

V. Представить результаты с помощью инструмента анализа данных **Регрессия** ППП Excel.

Пример выполнения лабораторной работы

1. В диапазоне А2:С11 подготовить исходные.
2. Введем вспомогательные данные:

Ячейка	Формула	Примечание
C16	9	Число предприятий
C17	0,05	Уровень значимости
C18	=ОТРЕЗОК(C3:C11;B3:B11)	Коэффициент *a*
C19	=НАКЛОН(C3:C11;B3:B11)	Коэффициент *b*
C20	=СРЗНАЧ(B3:B11)	Среднее значение фактора
C21	=СРЗНАЧ(C3:C11)	Среднее значение результата

Проверка значимости коэффициента *b*.

1) Для расчетов сумм квадратов отклонений введем формулы:

Ячейка	Формула	Примечание
D3	=$C$18+$C$19*B3	Копируем в диапазон D3:D11
E3	=(C3-D3)^2	Копируем в диапазон E3:E11
F3	=(B3-$C$20)^2	Копируем в диапазон F3:F11
E12	=СУММ(E3:E11)	$\sum e_i^2 = \sum (y - y_x)^2$
F12	=СУММ(F3:F11)	$\sum (x - \overline{x})^2$

2) Стандартная ошибка параметра *b* определяется по формуле:

$$m_b = \sqrt{\frac{\sum (y - y_x)^2}{(n-2)\sum (x - \overline{x})^2}},$$

поэтому введем в ячейку D24 формулу:

=(E12/((C16-2)*F12))^0,5.

3) В ячейке D25 рассчитана *t*-статистика параметра *b* как отношение величины этого параметра к его стандартной ошибке:

=C19/D24.

4) Критическое значение *t*-статистики определим в ячейке D26 с помощью функции СТЬЮДРАСПОБР, у которой первым аргументом является пороговая значимость или вероятность (в нашем случае примем ее равной 0,05), а вторым – число степеней свободы (*n*–2=9–2=7). Таким образом, формула, введенная в D26, должна иметь вид:

=СТЬЮДРАСПОБР($C$17;$C$16-2).

5) Для того чтобы автоматически был получен вывод о значимости параметра *b* построим в ячейке D27 формулу:

=ЕСЛИ(ABS(D25)>D26;"Значим";"Незначим").

6) Для расчета доверительного интервала определяем предельную ошибку в ячейке D28:

=D26*D24.

7) Нижняя граница доверительного интервала в ячейке D29: **=C19-D28.**

8) Верхняя граница доверительного интервала в ячейке D30:

=C19+D28.

Таким образом, доверительный интервал параметра *b* имеет вид (1,12; 1,83).

Проверка значимости коэффициента *a*.

Вводим формулы:

Ячейка	Формула	Примечание
G3	=B3*B3	Копируем в диапазон G3:G11
G12	=СУММ(E3:E11)	$\sum x^2$
D33	=D24*КОРЕНЬ(G12/C16)	$m_a = m_b \sqrt{\frac{\sum x^2}{n}}.$
D34	=C18/D33	t-статистика параметра a
D35	=СТЬЮДРАСПОБР($C$17;$C$16-2)	Критическое значение t-статистики
D36	=ЕСЛИ(ABS(D34)>D35;"Значим";"Незначим")	
D37	=D35*D33	Предельная ошибка
D38	=C18-D37	Нижняя граница доверительного интервала
D39	=C18+D37	Верхняя граница доверительного интервала

Проверка общего качества уравнения регрессии с помощью F-теста.

Вводим формулы:

Ячейка	Формула	Примечание
D42	=(КОРРЕЛ(B3:B11;C3:C11)^2)	Коэффициент детерминации
D43	=D42*(C16-2)/(1-D42)	F-статистика
D44	=FРАСПОБР(C17;1;C16-2)	Критическое значение F-статистики
D45	=ЕСЛИ(D43>D44;"Значимо";"Незначимо")	

Результаты расчетов приведены на рис. 1.2.

	A	B	C	D	E	F	G
1	**Проверка качества уравнения регрессии**						
2	№ предприятия	x	y	y_x	e^2	$(x\text{-}xcp)^2$	x^2
3	1	11	25	24,110	0,792	71,309	121
4	2	13	27	27,058	0,003	41,531	169
5	3	15	31	30,005	0,989	19,753	225
6	4	18	30	34,427	19,596	2,086	324
7	5	20	38	37,374	0,391	0,309	400
8	6	22	43	40,322	7,172	6,531	484
9	7	24	44	43,270	0,534	20,753	576
10	8	25	42	44,743	7,526	30,864	625
11	9	27	49	47,691	1,714	57,086	729
12				сумма	**38,717**	**250,222**	**3653,000**
13							
14							
15	**Вспомогательные данные**						
16	n		9				
17	уровень значимости		0,05				
18	a		7,90				
19	b		1,47				
20	$x_{ср}$		19,44				
21	$y_{ср}$		36,56				
22							
23	**Значимость b**						
24	стандартная ошибка			0,1487			
25	t-статистика			9,9129			
26	Критическое значение t-статистики			2,3646			
27	Вывод			Значим			
28	предельная ошибка			0,35			
29	нижняя граница			1,12			
30	верхняя граница			1,83			
31							
32	**Значимость a**						
33	стандартная ошибка			2,9953			
34	t-статистика			2,6369			
35	Критическое значение t-статистики			2,3646			
36	Вывод			Значим			
37	предельная ошибка			7,08			
38	нижняя граница			0,82			
39	верхняя граница			14,98			
40							
41	**Значимость уравнения регрессии**						
42	коэффициент детерминации			0,9335			
43	F-статистика			98,2649			
44	Критическое значение t-статистики			5,5914			
45	Вывод			Значимо			

Рис. 1.2. Анализ значимости уравнения регрессии

Проверка общего качества уравнения регрессии с помощью средней ошибки аппроксимации.

Вводим формулы:

Ячейка	Формула	Примечание
H3	=C3-D3	
I3	=ABS(H3/C3)*100	
I12	=СРЗНАЧ(I3:I11)	Средняя ошибка аппроксимации

В среднем расчетные значения отклоняются от фактических на 4,5%. Качество построенной модели оценивается как хорошее, так как средняя ошибка аппроксимации не превышает 8-10% (см. рис. 2.2).

	A	B	C	D	E	F	G	H	I
1	**Проверка качества уравнения регрессии**								
2	№ предприятия	x	y	y_x	e^2	$(x\text{-}xcp)^2$	x^2	$y\text{-}y_x$	A_i
3	1	11	25	24,110	0,792	71,309	121	0,890	3,56
4	2	13	27	27,058	0,003	41,531	169	-0,058	0,21
5	3	15	31	30,005	0,989	19,753	225	0,995	3,21
6	4	18	30	34,427	19,596	2,086	324	-4,427	14,76
7	5	20	38	37,374	0,391	0,309	400	0,626	1,65
8	6	22	43	40,322	7,172	6,531	484	2,678	6,23
9	7	24	44	43,270	0,534	20,753	576	0,730	1,66
10	8	25	42	44,743	7,526	30,864	625	-2,743	6,53
11	9	27	49	47,691	1,714	57,086	729	1,309	2,67
12				сумма	**38,717**	**250,222**	**3653,000**		**4,50**
13									

Рис. 2.2. Анализ общего уравнения регрессии с помощью ошибки аппроксимации

Представим результаты с помощью инструмента анализа данных **Регрессия** ППП Excel.

Для этого:

В главном меню выберите **Сервис – Анализ данных – Регрессия – ОК.**

Заполнить диалоговое окно ввода данных и параметров вывода (см. рис. 3.2).

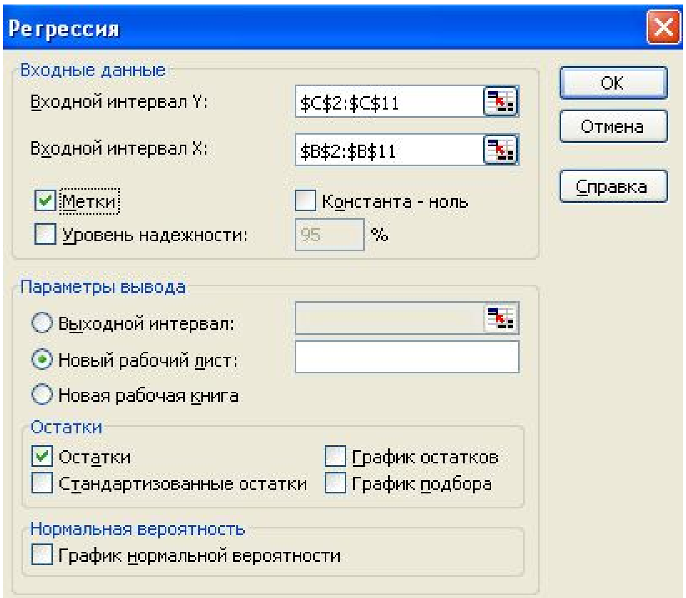


Рис. 3.2. Окно **Регрессия**

Входной интервал Y – диапазон, содержащий данные результативного признака;

Входной интервал X – диапазон, содержащий данные факторов независимого признака;

Метки – флажок, который указывает, содержит ли первая строка название столбцов или нет;

Константа – ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

На новом рабочем листе появляются данные регрессионного анализа (см. рис. 4.2.)

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,966178604							
5	R-квадрат	0,933501095							
6	Нормированный R-квадрат	0,924001251							
7	Стандартная ошибка	2,351812077							
8	Наблюдения	9							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*Значимость F*			
12	Регрессия	1	543,5050819	543,5050819	98,264891	2,26642E-05			
13	Остаток	7	38,71714032	5,531020046					
14	Итого	8	582,2222222						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
17	Y-пересечение	7,898312611	2,995320227	2,63688421	0,033575093	0,815505763	14,98111946	0,815505763	14,98111946
18	x	1,473801066	0,148675592	9,912864924	2,26642E-05	1,122239154	1,825362977	1,122239154	1,825362977
19									
20									
21									
22	ВЫВОД ОСТАТКА								
23									
24	*Наблюдение*	*Предсказанное y*	*Остатки*						
25	1	24,11012433	0,889875666						
26	2	27,05772647	-0,057726465						
27	3	30,0053286	0,994671403						
28	4	34,42673179	-4,426731794						
29	5	37,37433393	0,625666075						
30	6	40,32193606	2,678063943						
31	7	43,26953819	0,730461812						
32	8	44,74333925	-2,743339254						
33	9	47,69094139	1,309058615						
34									

Рис. 4.2. Результаты регрессионного анализа

Отчет по лабораторной работе

1. Запишите уравнение линейной парной регрессии для своего варианта.
2. Как оценивается значимость параметров уравнения регрессии?
3. Являются ли параметры уравнения регрессии для вашего варианта значимыми и почему?
4. Запишите доверительные интервалы для параметров уравнения регрессии для вашего варианта.
5. Каким образом осуществляется проверка значимости уравнения в целом.
6. Значимо ли уравнение регрессии для вашего варианта и почему?
7. Каким образом осуществляется проверка качества уравнения регрессии?
8. В чем смысл средней ошибки аппроксимации и каково ее значение для вашего варианта?
9. Сравнить полученные результаты с результатами применения инструмента **Регрессия.**

Лабораторная работа №3

Нелинейные модели. Коэффициент детерминации

Цель: научиться строить нелинейные модели и находить коэффициент детерминации.

Основные сведения

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают два класса нелинейных регрессий:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например

– полиномы различных степеней – $\hat{y}_x = a + b \cdot x + c \cdot x^2$;

– равносторонняя гипербола – $\hat{y}_x = a + b/x$;

– полулогарифмическая функция – $\hat{y}_x = a + b \cdot \ln x$.

2. Регрессии, нелинейные по оцениваемым параметрам, например

– степенная – $\hat{y}_x = a \cdot x^b$;

– показательная – $\hat{y}_x = a \cdot b^x$;

– экспоненциальная – $y=ae^{bx}$.

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью метода наименьших квадратов.

Несколько иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: нелинейные модели внутренне линейные (приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием) и нелинейные модели внутренне нелинейные (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например, степенная функция – $\hat{y}_x = a \cdot x^b$, показательная – $\hat{y}_x = a \cdot b^x$, экспоненциальная – $\hat{y}_x = a \cdot e^{bx}$, обратная – $\hat{y}_x = \dfrac{1}{a + b \cdot x}$.

К внутренне нелинейным моделям можно, например, отнести следующие модели: $\hat{y}_x = a + b \cdot x^c$, $\hat{y}_x = a \cdot \left(1 - \dfrac{1}{1 - x^b}\right)$.

Приведем формулы для расчета параметров наиболее часто используемых типов уравнений регрессии (табл. 1.3):

Таблица 1.3

Вид функции, y	Линеаризация	Параметры уравнения регрессии	Искомое уравнение
Степенная $y=ax^b$	$X=\ln x$, $Y=\ln y$, $A=\ln a$, $B=b$	$b=\dfrac{\overline{YX}-\overline{Y}\cdot\overline{X}}{\overline{X^2}-(\overline{X})^2}$ $A=\overline{Y}-b\overline{X}$	$y=e^A\cdot x^b$
Показательная $y=ab^x$	$X=x$, $Y=\ln y$, $B=\ln b$, $A=\ln a$,	$B=\dfrac{\overline{Yx}-\overline{Y}\cdot\overline{x}}{\overline{x^2}-(\overline{x})^2}$ $A=\overline{Y}-B\overline{x}$	$y=(e^A)\cdot(e^B)^x$
Обратная $y=\dfrac{1}{a+b\cdot x}$	$X=x$, $Y=1/y$, $A=a$, $B=b$	$b=\dfrac{\overline{Yx}-\overline{Y}\cdot\overline{x}}{\overline{x^2}-(\overline{x})^2}$ $a=\overline{Y}-b\overline{x}$	$y=\dfrac{1}{a+b\cdot x}$
Полулогарифмическая $y=a+b\cdot\ln x$	$X=\ln x$, $Y=y$, $A=a$, $B=b$	$b=\dfrac{\overline{yX}-\overline{y}\cdot\overline{X}}{\overline{X^2}-(\overline{X})^2}$ $a=\overline{y}-b\overline{X}$	$y=a+b\cdot\ln x$
Гиперболическая $y=a+b/x$	$X=1/x$, $Y=y$, $A=a$, $B=b$	$b=\dfrac{\overline{yX}-\overline{y}\cdot\overline{X}}{\overline{X^2}-(\overline{X})^2}$ $a=\overline{y}-b\overline{X}$	$y=a+b/x$
Экспоненциальная $y=ae^{bx}$	$X=x$, $Y=\ln y$, $A=\ln a$, $B=b$	$b=\dfrac{\overline{Yx}-\overline{Y}\cdot\overline{x}}{\overline{x^2}-(\overline{x})^2}$ $A=\overline{Y}-b\overline{x}$	$y=e^A\cdot e^{bx}$

В случае нелинейной зависимости тесноту связи между величинами оценивают по величине корреляционного отношения:

$$\rho_{xy}=\sqrt{1-\frac{\sum(y-y_x)^2}{\sum(y-\overline{y})^2}}=\sqrt{1-\frac{\sigma^2_{ост}}{\sigma^2_y}}.$$

Интервал изменения корреляционного отношения $0\le\rho_{xy}\le 1$.

Оценку качества построенной модели дает индекс детерминации ρ^2_{xy}.

Коэффициент детерминации $R^2=\rho^2_{xy}$ – квадрат индекса корреляции – характеризует долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака y.

$$R^2 = 1 - \frac{\sum (y_x - y)^2}{\sum (y - \overline{y})^2}.$$

Чем ближе коэффициент детерминации к 1, тем выше качество уравнения регрессии, тем в большей мере оно объясняет поведение отклика.

Порядок выполнения работы.

Используя данные лабораторной работы №1, построить линейную, степенную, показательную, экспоненциальную, полулогарифмическую, гиперболическую и обратную модели и с помощью коэффициента детерминации сравнить эти модели. Для чего необходимо:

1. Найти уравнение регрессии.
2. Найти общую сумму квадратов отклонений и остаточную сумму квадратов отклонений.
3. Найти коэффициент детерминации.
4. Найти параметры регрессии с помощью статистической функции **ЛИНЕЙН**.

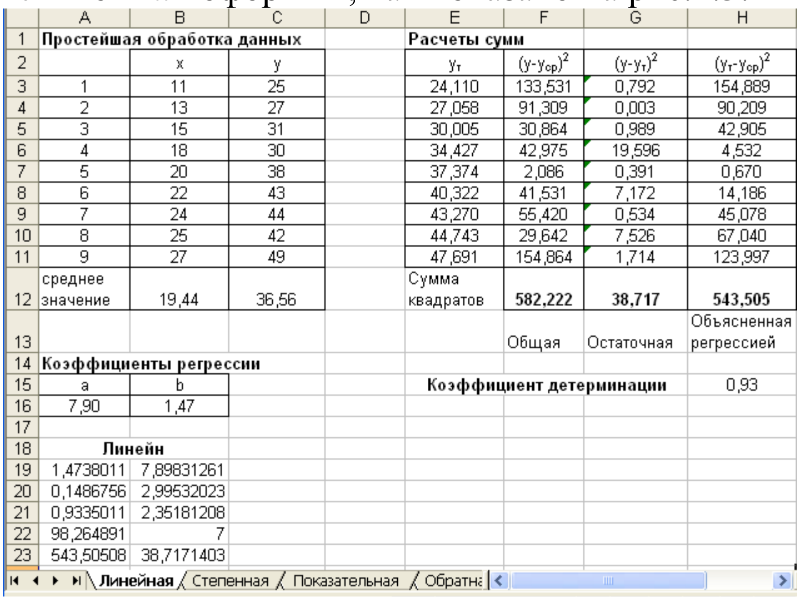
Пример выполнения лабораторной работы.

Создадим новую рабочую книгу с семью листами.

Название листа	Назначение
Линейная	Для анализа линейной модели
Степенная	Для анализа степенной модели
Показательная	Для анализа показательной модели
Обратная	Для анализа обратной модели
Полулогарифмическая	Для анализа полулогарифмической модели
Гиперболическая	Для анализа гиперболической модели
Экспоненциальная	Для анализа экспоненциальной модели

Будем использовать данные из лабораторной работы №1.

1. Лист **Линейная** оформим, как показано на рис.1.3:



	A	B	C	D	E	F	G	H
1	**Простейшая обработка данных**				**Расчеты сумм**			
2		x	y		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25		24,110	133,531	0,792	154,889
4	2	13	27		27,058	91,309	0,003	90,209
5	3	15	31		30,005	30,864	0,989	42,905
6	4	18	30		34,427	42,975	19,596	4,532
7	5	20	38		37,374	2,086	0,391	0,670
8	6	22	43		40,322	41,531	7,172	14,186
9	7	24	44		43,270	55,420	0,534	45,078
10	8	25	42		44,743	29,642	7,526	67,040
11	9	27	49		47,691	154,864	1,714	123,997
12	среднее значение	19,44	36,56		Сумма квадратов	**582,222**	**38,717**	**543,505**
13						Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**							
15	a	b			**Коэффициент детерминации**			0,93
16	7,90	1,47						
17								
18	**Линейн**							
19	1,4738011	7,89831261						
20	0,1486756	2,99532023						
21	0,9335011	2,35181208						
22	98,264891	7						
23	543,50508	38,7171403						

Линейная / Степенная / Показательная / Обратна

Рис. 1.3. Лист **Линейная**

На этом листе коэффициенты линейной регрессии определяются с помощью статистических функций (см. лабораторную работу №1).

Для расчета сумм, которые понадобятся при определении коэффициента детерминации (и при выполнении следующей лабораторной работы), введем формулы:

Ячейка	Формула	Примечание
E3	=$A$16+$B$16*B3	Расчет теоретических значений результата y_m. Копируем в диапазон E3:E11
F3	=(C3-$C$12)^2	Копируем в диапазон F3:F11
G3	=(C3-E3)^2	Копируем в диапазон G3:G11
H3	=(E3-$C$12)^2	Копируем в диапазон H3:H11
F12	=СУММ(F3:F11)	Копируем в диапазон F12:H12

Замечание. В приведенных формулах неоднократно используется абсолютная адресация, содержащая знак «$». Это необходимо для того, чтобы при копировании формул данный адрес не изменялся. Для того чтобы превратить относительный адрес A16 в абсолютный ($A$16), достаточно нажать клавишу F4 в то время, когда курсор находится на ячейке A16.

Для вычисления коэффициента детерминации в ячейку H15 введем формулу:

=1-G12/F12.

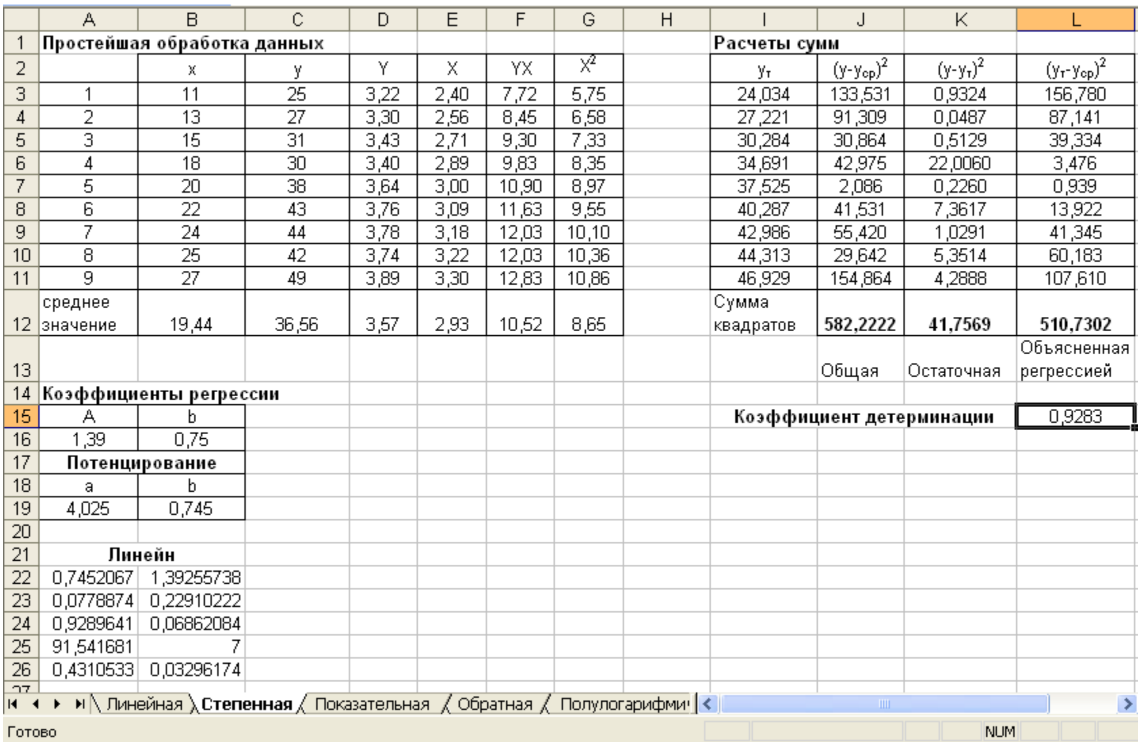
2. Регрессия в виде степенной функции имеет вид: $y=ax^b$.

Для нахождения параметров регрессии $y=ax^b$ необходимо провести ее линеаризацию:

$$Y=A+bX,$$

где $Y=\ln y$, $X=\ln x$, $A=\ln a$.

Составляем вспомогательную таблицу для преобразованных данных (рис. 2.3):



	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	Y	X	YX	X^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	3,22	2,40	7,72	5,75		24,034	133,531	0,9324	156,780
4	2	13	27	3,30	2,56	8,45	6,58		27,221	91,309	0,0487	87,141
5	3	15	31	3,43	2,71	9,30	7,33		30,284	30,864	0,5129	39,334
6	4	18	30	3,40	2,89	9,83	8,35		34,691	42,975	22,0060	3,476
7	5	20	38	3,64	3,00	10,90	8,97		37,525	2,086	0,2260	0,939
8	6	22	43	3,76	3,09	11,63	9,55		40,287	41,531	7,3617	13,922
9	7	24	44	3,78	3,18	12,03	10,10		42,986	55,420	1,0291	41,345
10	8	25	42	3,74	3,22	12,03	10,36		44,313	29,642	5,3514	60,183
11	9	27	49	3,89	3,30	12,83	10,86		46,929	154,864	4,2888	107,610
12	среднее значение	19,44	36,56	3,57	2,93	10,52	8,65		Сумма квадратов	**582,2222**	**41,7569**	**510,7302**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	b							**Коэффициент детерминации**			0,9283
16	1,39	0,75										
17	**Потенцирование**											
18	a	b										
19	4,025	0,745										
20												
21	**Линейн**											
22	0,7452067	1,39255738										
23	0,0778874	0,22910222										
24	0,9289641	0,06862084										
25	91,541681	7										
26	0,4310533	0,03296174										

Линейная / **Степенная** / Показательная / Обратная / Полулогарифми

Готово NUM

Рис. 2.3 Лист **Степенная**

Вводим формулы:

Ячейка	Формула	Примечание
D3	=LN(C3)	Y=ln y Копируем в диапазон D3:D11
E3	=LN(B3)	X=ln x Копируем в диапазон E3:E11
F3	=D3*E3	Копируем в диапазон F3:F11
G3	=E3^2	Копируем в диапазон G3:G11
D12	=СРЗНАЧ(D3:D11)	Копируем в диапазон D12:G12

Для вычисления коэффициентов регрессии введем следующие формулы

Ячейка	Формула	Примечание
B16	=(F12-E12*D12)/(G12-E12^2)	b
A16	=D12-B16*E12	A

После потенцирования находим искомые коэффициенты регрессии:

Ячейка	Формула	Примечание
A19	=EXP(A16)	a
B19	=B16	b

Тогда уравнение регрессии будет иметь вид: y=4,02$x^{0,75}$.

Для расчета сумм введем формулы:

Ячейка	Формула	Примечание
I3	=$A$19*B3^$B$19	Расчет теоретических значений результата y_m. Копируем в диапазон I3:I11
J3	=(C3-$C$12)^2	Копируем в диапазон J3:J11
K3	=(C3-E3)^2	Копируем в диапазон K3:K11
L3	=(E3-$C$12)^2	Копируем в диапазон L3:L11
J12	=СУММ(J3:J11)	Копируем в диапазон J12:L12

Для вычисления коэффициента детерминации в ячейку L15 введем формулу:

=1-K12/J12.

Проведем расчеты параметров регрессии с помощью статистической функции **ЛИНЕЙН**.

Выделим диапазон A22:B26. введем формулу

=ЛИНЕЙН(D3:D11;E3:E11;1;1).

В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу **F2,** а затем – на комбинацию клавиш **CTRL+SHIFT+ENTER**.

3. Расчеты на остальных листах во многом повторяют расчеты, произведенные на листе **Степенная**, поэтому остальные листы лучше всего получить копированием листа **Степенная**.

Для этого необходимо:

- находясь на листе **Степенная**, выделить его полностью, щелкнув мышью на пересечении названий столбцов и строк; с помощью кнопки  (**Копировать**) скопировать лист в **Буфер обмена**;
- перейти на следующий лист и выделив ячейку A1,щелкнуть мышью по кнопке  (**Вставить**).

Получим следующие результаты (рис. 3.3-7.3):

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	Y	x	Yx	x^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	3,22	11,00	35,41	121,00		25,056	133,531	0,0031	132,250
4	2	13	27	3,30	13,00	42,85	169,00		27,237	91,309	0,0562	86,833
5	3	15	31	3,43	15,00	51,51	225,00		29,609	30,864	1,9356	48,258
6	4	18	30	3,40	18,00	61,22	324,00		33,559	42,975	12,6652	8,980
7	5	20	38	3,64	20,00	72,75	400,00		36,481	2,086	2,3079	0,006
8	6	22	43	3,76	22,00	82,75	484,00		39,657	41,531	11,1740	9,621
9	7	24	44	3,78	24,00	90,82	576,00		43,110	55,420	0,7917	42,964
10	8	25	42	3,74	25,00	93,44	625,00		44,948	29,642	8,6902	70,432
11	9	27	49	3,89	27,00	105,08	729,00		48,862	154,864	0,0192	151,438
12	среднее значение	19,44	36,56	3,57	19,44	70,65	405,89		Сумма квадратов	**582,2222**	**37,6430**	**550,7818**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	B							**Коэффициент детерминации**			0,9353
16	2,76	0,04										
17	**Потенцирование**											
18	a	b										
19	15,83	1,04										
20												
21	**Линейн**											
22	0,0417435	2,76191628										
23	0,003998	0,08054714										
24	0,9396628	0,06324257										
25	109,01471	7										
26	0,4360177	0,02799736										

Линейная / Степенная / **Показательная** / Обратная / Полулогарифми

Рис. 3.3. Лист **Показательная**

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	Y	x	Yx	x^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	0,04	11,00	0,44	121,00		25,608	133,531	0,3695	119,853
4	2	13	27	0,04	13,00	0,48	169,00		27,306	91,309	0,0939	85,545
5	3	15	31	0,03	15,00	0,48	225,00		29,246	30,864	3,0748	53,422
6	4	18	30	0,03	18,00	0,60	324,00		32,735	42,975	7,4804	14,596
7	5	20	38	0,03	20,00	0,53	400,00		35,563	2,086	5,9389	0,985
8	6	22	43	0,02	22,00	0,51	484,00		38,926	41,531	16,5990	5,618
9	7	24	44	0,02	24,00	0,55	576,00		42,991	55,420	1,0181	41,415
10	8	25	42	0,02	25,00	0,60	625,00		45,360	29,642	11,2865	77,510
11	9	27	49	0,02	27,00	0,55	729,00		50,977	154,864	3,9067	207,965
12	среднее значение	19,44	36,56	0,03	19,44	0,53	405,89		Сумма квадратов	**582,2222**	**49,7678**	**606,9100**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	B							**Коэффициент детерминации**			0,9145
16	0,0524	-0,0012										
17	**Потенцирование**											
18	a	b										
19	0,0524	-0,0012										
20												
21	**Линейн**											
22	-0,0012146	0,05241122										
23	0,0001213	0,00244362										
24	0,9347494	0,00191864										
25	100,27871	7										
26	0,0003691	2,5768E-05										

Степенная / Показательная / **Обратная** / Полулогарифмическая / Ги

Рис. 4.3. Лист **Обратная**

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	y	X	yX	X^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	25,00	2,40	59,95	5,75		22,769	133,531	4,9780	190,073
4	2	13	27	27,00	2,56	69,25	6,58		27,124	91,309	0,0153	88,959
5	3	15	31	31,00	2,71	83,95	7,33		30,854	30,864	0,0213	32,506
6	4	18	30	30,00	2,89	86,71	8,35		35,607	42,975	31,4391	0,900
7	5	20	38	38,00	3,00	113,84	8,97		38,354	2,086	0,1251	3,233
8	6	22	43	43,00	3,09	132,91	9,55		40,838	41,531	4,6731	18,342
9	7	24	44	44,00	3,18	139,83	10,10		43,107	55,420	0,7983	42,915
10	8	25	42	42,00	3,22	135,19	10,36		44,171	29,642	4,7120	57,991
11	9	27	49	49,00	3,30	161,50	10,86		46,177	154,864	7,9694	92,572
12	среднее значение	19,44	36,56	36,56	2,93	109,24	8,65		Сумма квадратов	**582,2222**	**54,7315**	**527,4907**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	B							**Коэффициент детерминации**			0,9060
16	-39,74	26,07										
17	**Потенцирование**											
18	a	b										
19	-39,74	26,07										
20												
21	**Линейн**											
22	26,068659	-39,741061										
23	3,1738096	9,33561316										
24	0,9059955	2,79620874										
25	67,464555	7										
26	527,49074	54,7314833										

Полулогарифмическая / Гиперболическая / Экспоненциальная

Рис. 5.3. Лист **Полулогарифмическая**

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	y	z	yz	z^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	25,00	0,09	2,27	0,01		21,702	133,531	10,8800	220,643
4	2	13	27	27,00	0,08	2,08	0,01		27,657	91,309	0,4312	79,190
5	3	15	31	31,00	0,07	2,07	0,00		32,024	30,864	1,0482	20,537
6	4	18	30	30,00	0,06	1,67	0,00		36,755	42,975	45,6281	0,040
7	5	20	38	38,00	0,05	1,90	0,00		39,120	2,086	1,2553	6,578
8	6	22	43	43,00	0,05	1,95	0,00		41,056	41,531	3,7799	20,252
9	7	24	44	44,00	0,04	1,83	0,00		42,669	55,420	1,7724	37,370
10	8	25	42	42,00	0,04	1,68	0,00		43,378	29,642	1,8998	46,550
11	9	27	49	49,00	0,04	1,81	0,00		44,640	154,864	19,0101	65,357
12	среднее значение	19,44	36,56	36,56	0,06	1,92	0,00		Сумма квадратов	**582,2222**	**85,7050**	**496,5172**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	B							**Коэффициент детерминации**			0,8528
16	60,41	-425,79										
17	**Потенцирование**											
18	a	b										
19	60,41	-425,7945										
20												
21	**Линейн**											
22	-425,79454	60,4101097										
23	66,863162	3,92330082										
24	0,8527968	3,499081										
25	40,553314	7										
26	496,51725	85,7049748										

Обратная / Полулогарифмическая / **Гиперболическая** / Экспоненциа...

Рис. 6.3. Лист **Гиперболическая**

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Простейшая обработка данных**								**Расчеты сумм**			
2		x	y	Y	x	Yx	x^2		$y_т$	$(y-y_{ср})^2$	$(y-y_т)^2$	$(y_т-y_{ср})^2$
3	1	11	25	3,22	11,00	35,41	121,00		25,056	133,531	0,0031	132,250
4	2	13	27	3,30	13,00	42,85	169,00		27,237	91,309	0,0562	86,833
5	3	15	31	3,43	15,00	51,51	225,00		29,609	30,864	1,9356	48,258
6	4	18	30	3,40	18,00	61,22	324,00		33,559	42,975	12,6652	8,980
7	5	20	38	3,64	20,00	72,75	400,00		36,481	2,086	2,3079	0,006
8	6	22	43	3,76	22,00	82,75	484,00		39,657	41,531	11,1740	9,621
9	7	24	44	3,78	24,00	90,82	576,00		43,110	55,420	0,7917	42,964
10	8	25	42	3,74	25,00	93,44	625,00		44,948	29,642	8,6902	70,432
11	9	27	49	3,89	27,00	105,08	729,00		48,862	154,864	0,0192	151,438
12	среднее значение	19,44	36,56	3,57	19,44	70,65	405,89		Сумма квадратов	**582,2222**	**37,6430**	**550,7818**
13										Общая	Остаточная	Объясненная регрессией
14	**Коэффициенты регрессии**											
15	A	b							**Коэффициент детерминации**			0,9353
16	2,76	0,04										
17	**Потенцирование**											
18	a	b										
19	15,83	0,0417										
20												
21	**Линейн**											
22	0,0417435	2,76191628										
23	0,003998	0,08054714										
24	0,9396628	0,06324257										
25	109,01471	7										
26	0,4360177	0,02799736										

Полулогарифмическая / Гиперболическая / **Экспоненциальная**

Рис. 7.3. Лист **Экспоненциальная**

Выберем наилучшую модель, для чего объединим результаты построения парных регрессий в одной таблице (табл. 2.1).

Все уравнения регрессии достаточно хорошо описывают исходные данные. Некоторое предпочтение можно отдать показательной или экспоненциальной функции, для которых значение коэффициента детерминации наибольшее.

Вид регрессии	Уравнение регрессии	Коэффициент детерминации
Линейная	$y=7{,}9+1{,}47x$	0,9335
Степенная	$y=4{,}03x^{0,75}$	0,9283
Показательная	$y=15{,}83\cdot 1{,}04^{x}$	0,9353
Обратная	$y=1/(0{,}05–0{,}0012x)$	0,9145
Полулогарифмическая	$y= –39{,}74+26{,}07\ln x$	0,9060
Гиперболическая	$y=60{,}41–425{,}79/x$	0,8528
Экспоненциальная	$y=15{,}83e^{0,0417x}$	0,9353

Отчет по лабораторной работе

1. Запишите все виды моделей, нелинейных относительно включаемых переменных и оцениваемых параметров.
2. Как осуществляется линеаризация модели?
3. Назовите показатели корреляции, используемые при нелинейных соотношениях рассматриваемых признаков.
4. Записашите уравнения линейной, степенной, показательной, экспоненциальной, полулогарифмической, гиперболической и обратной моделей и с помощью коэффициента детерминации сравнить эти модели.

Лабораторная работа №4

Прогнозирование на основании линейной регрессии

Цель: научиться прогнозировать индивидуальные значения зависимой переменной на основании линейной регрессии; уметь определять точность прогноза.

Краткие теоретические сведения.

Пусть по заданной выборке объема n найдено выборочное уравнение линейной регрессии

$$y=a+bx$$

С помощью этого уравнения можно прогнозировать значение результата y_p при определенном прогнозном значении фактора x_p.

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии $y=a+bx$ соответствующего прогнозного значения x_p.

Точное уравнение регрессии нам неизвестно. Поэтому мы не можем сделать точный прогноз. Можно только утверждать, что прогнозное значение результата y_p при данном x_p с вероятностью γ попадет в доверительный интервал γ_p. Вероятность γ называется уровнем надежности.

Ошибка прогноза составляет:

$$m_p = S\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{\sum(x-\bar{x})^2}},$$

где $S=\sqrt{\dfrac{\sum y^2 - a\sum y - b\sum xy}{n-2}}=\sqrt{\dfrac{\sum(y-y_x)^2}{n-2}}$ – стандартная ошибка регрессии (дисперсия ошибки или остаточная дисперсия).

Предельная ошибка прогноза, составит:

$$\Delta_p = t_{табл}\cdot m_p.$$

Доверительный интервал прогноза:

$$\gamma_p = y_p \pm \Delta_p.$$

Точность прогноза можно оценить с помощью относительной ошибки прогноза:

$$\delta_p = \frac{\Delta_p}{|y_p|}\cdot 100\%.$$

Порядок выполнения работы.

Используя данные к лабораторной работе №1 при x_p=20:

1. найти уравнение регрессии;
2. рассчитать доверительный интервал прогноза при значениях уровня надежности 80%, 90%, 95%;
3. найти относительную ошибку прогноза;
4. построить графики линии регрессии с доверительными границами.

Пример выполнения лабораторной работы.

Расчеты для каждого из уровней надежности производить на отдельных листах, которые назовем , соответственно: 80%, 90%, 95%.

I. Лист 80%.

9. В диапазоне A2:C11 подготовим исходные данные.
10. В ячейку B12 запишем значение $x_{\text{р}}$=15,5, для которого необходимо спрогнозировать значение результата $y_{\text{р}}$.
11. Вводим следующие формулы:

Ячейка	Формула	Примечание
C16	=ОТРЕЗОК(C3:C11;B3:B11)	Определяем коэффициенты регрессии
C17	=НАКЛОН(C3:C11;B3:B11)	
C18	=СРЗНАЧ(B3:B11)	$\overline{x}$
C19	=СТОШYX(C3:C11;B3:B11)	Стандартная ошибка регрессии
C20	9	Количество предприятий
C21	=80/100	Уровень надежности
C22	=1−C21	Уровень значимости
C23	=C20−2	Число степеней свободы
C24	=СТЬЮДРАСПОБР(C22;C23)	критическое значение t-статистики
D3	=(B3−$C$18)^2	Формулу скопируем вниз до D11
D12	=СУММ(D3:D11)	Находим $\sum(x-\overline{x})^2$
C27	=C19*КОРЕНЬ(1+1/C20+(B12-C18)^2/D12)	Стандартная ошибка прогноза
C28	=C24*C27	Предельная ошибка прогноза
C29	=C16+C17*B12	Прогнозное значение результата
C30	=C29−C28	Нижняя граница доверительного интервала
C31	=C29+C28	Верхняя граница доверительного интервала
C32	=C28*100/C29	Относительная ошибка прогноза

12. Для графического представления полученных результатов:
- Вводим следующие формулы:

E3	=$C$16+$C$17*B3	Копируем вниз до E11
F3	=$C$19*КОРЕНЬ(1+1/$C$20+(B3-$C$18)^2/$D$12)	Копируем вниз до F11
G3	=F3*$C$24	Копируем вниз до G11
H3	=E3−G3	Копируем вниз до H11
I3	=E3+G3	Копируем вниз до I11

Таким образом получим данные, представленные на рис. 1.4.

- Выделим одновременно диапазоны B2:C11, E2:E11, H2:I11 (поскольку эти диапазоны несмежные, при этом должна быть нажата клавиша Ctrl);
- Вызовем **Мастер диаграмм**. Чтобы ось отражала фактические данные, выберем тип диаграммы **Точечная**;
- Для добавления на диаграмму прогнозируемых значений в **Мастере диаграмм** на шаге 2 перейдем на вкладку **Ряд** (рис. 2.4). Щелкнем по кнопке **Добавить** и введем с помощью левой кнопки мыши: **Имя** − Прогноз, **Значения X** – B12, **Значения Y** – C29. Щелкнув по кнопке **Готово**, получим диаграмму, представленную на рисунке 3.4.

Отформатируем диаграмму. Для этого щелкнем дважды по фону и выберем заливку **прозрачная**, затем щелкнем дважды по линии регрессии и выберем тип линии, цвет и толщину, а переключатель маркера поставим в положение **отсутствует**. Аналогичным образом форматируются линии, представляющие границы доверительных интервалов, и точки, отображающие прогнозируемые значения. В итоге получим диаграмму, представленную на рис. 4.4.

II. **Лист 90% и 95%.**

Чтобы получить расчеты для уровней надежности 90% и 95%, достаточно скопировать лист 80% на листы 90% и 95% и ввести на них в ячейку C21 соответственно значения 0,9 и 0,95. При этом диаграммы, полученные при таком копировании, следует удалить и построить заново на основе расчетов, полученных на листах 90% и 95% (рис. 5.4 и 6.4).

Исходные данные

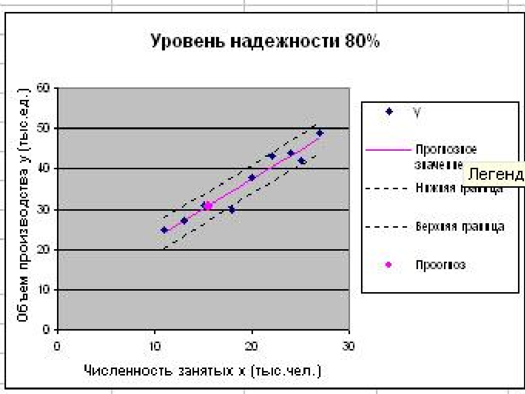
№ предприятия	x	y	$(x-x_{ф})^2$	Прогнозное значение	Стандартная ошибка проноза	Прельная ошибка прогноза	Нижняя граница	Верхняя граница
1	11	25	71,3086	24,110	2,779	3,932	20,178	28,042
2	13	27	41,5309	27,058	2,658	3,761	23,297	30,818
3	15	31	19,7531	30,005	2,566	3,630	26,375	33,635
4	18	30	2,0864	34,427	2,488	3,521	30,906	37,948
5	20	38	0,3086	37,374	2,480	3,510	33,865	40,884
6	22	43	6,5309	40,322	2,508	3,549	36,773	43,871
7	24	44	20,7531	43,270	2,570	3,636	39,633	46,906
8	25	42	30,8642	44,743	2,613	3,697	41,046	48,441
9	27	49	57,0864	47,691	2,722	3,851	43,840	51,542
	15,5	Сумма	250,2222					

Вспомогательные данные

a	7,8983
$в$	1,4738
$\bar{x}$	19,444
S	2,3518
n	9
уровень надежности	0,8
уровень значимости	0,2
число степеней свободы	7
критическое значение t-статистики	1,4149

Завершающие данные

m_p	2,5474
Δ_p	3,6044
y_p	30,742
нижняя граница	27,138
верхняя граница	34,347
δ_p	11,725



80% Лист2 Лист3

Готово

Рис. 1.4. Прогнозирование на основании линейной модели при уровне надежности 80%

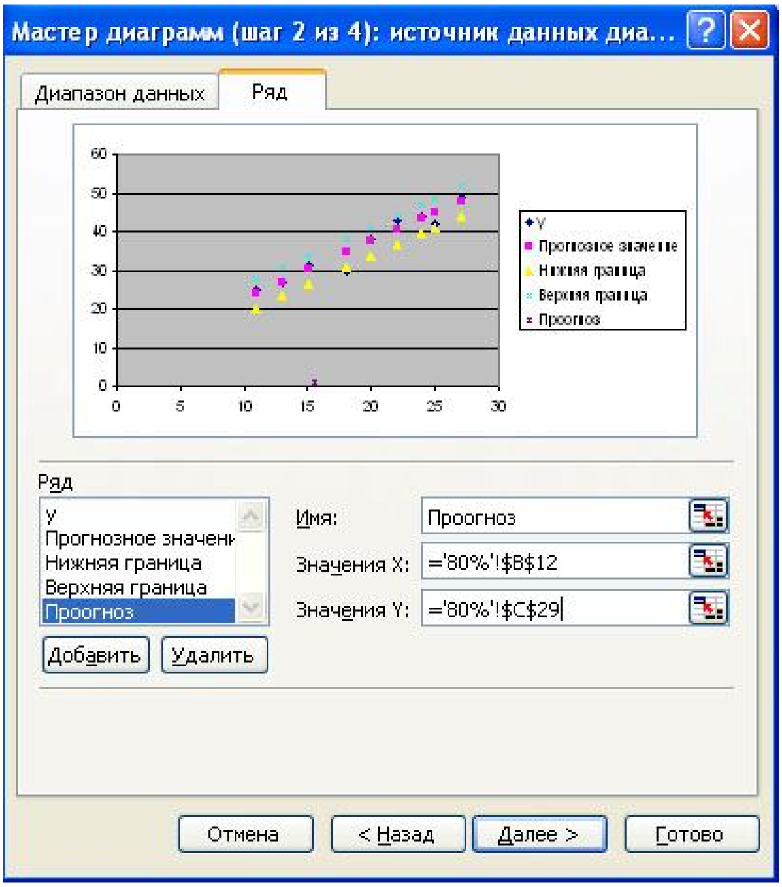


Рис. 2.4. Шаг 2 Мастера диаграмм

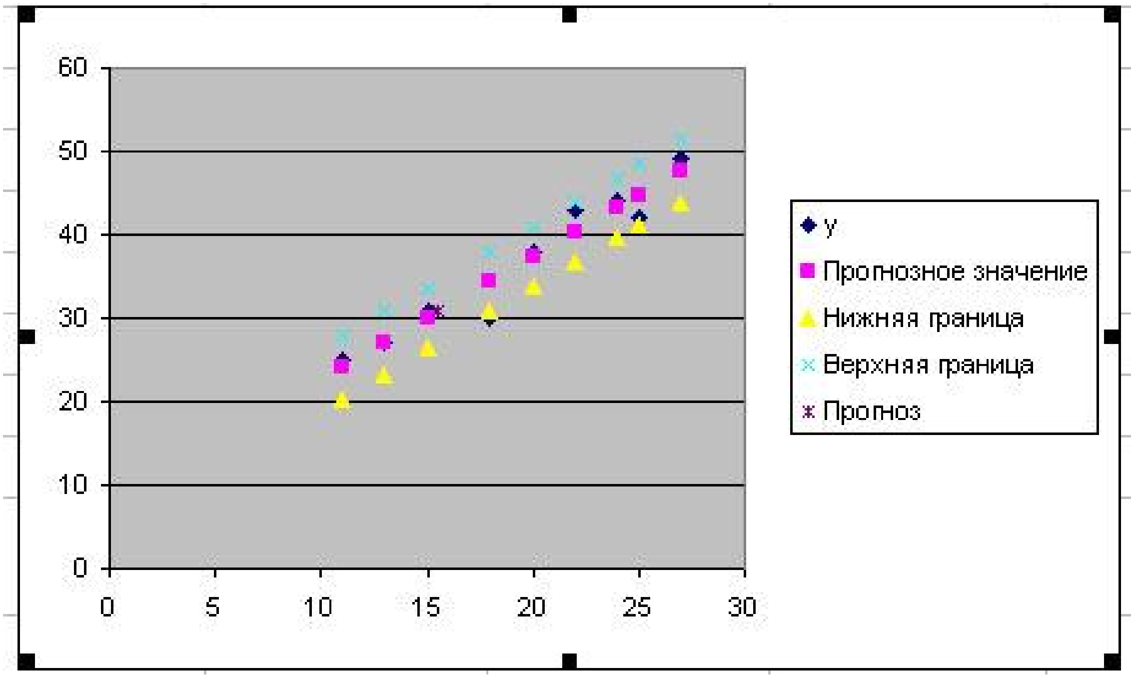


Рис. 3.4. Диаграмма, построена с помощью **Мастера**

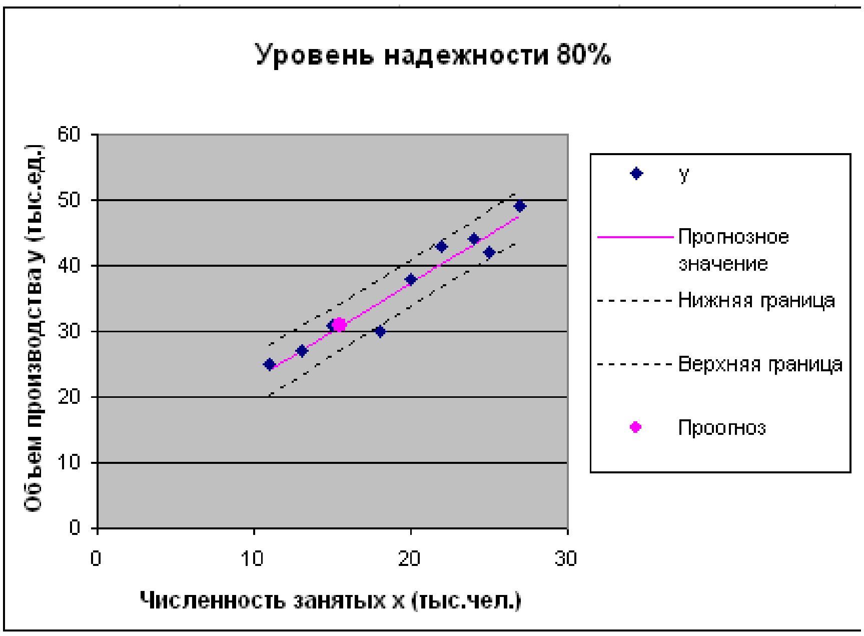
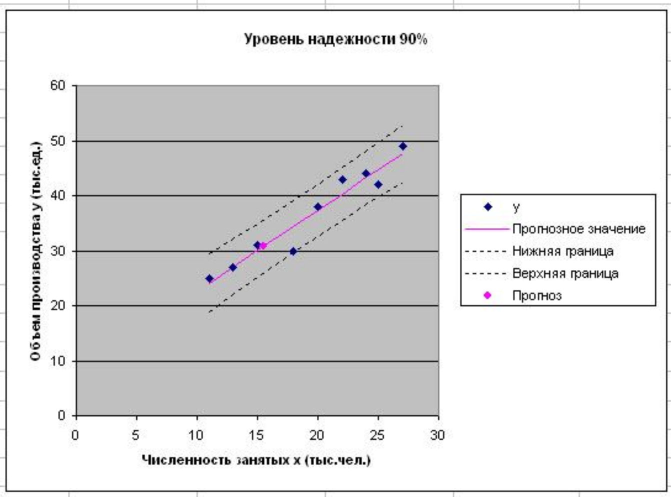


Рис. 4.4. Итоговый вид диаграммы при уровне надежности 80%

	A	B	C	D	E	F	G	H	I
1	**Исходные данные**								
2	№ предприятия	x	y	$(x - x_{ср})^2$	Прогнозное значение	Стандартная ошибка проноза	Прельная ошибка прогноза	Нижняя граница	Верхняя граница
3	1	11	25	71,3086	24,110	2,779	5,265	18,845	29,375
4	2	13	27	41,5309	27,058	2,658	5,035	22,022	32,093
5	3	15	31	19,7531	30,005	2,566	4,861	25,145	34,866
6	4	18	30	2,0864	34,427	2,488	4,714	29,712	39,141
7	5	20	38	0,3086	37,374	2,480	4,699	32,675	42,074
8	6	22	43	6,5309	40,322	2,508	4,752	35,570	45,073
9	7	24	44	20,7531	43,270	2,570	4,869	38,401	48,138
10	8	25	42	30,8642	44,743	2,613	4,951	39,793	49,694
11	9	27	49	57,0864	47,691	2,722	5,156	42,535	52,847
12		**15,5**	Сумма	250,2222					

15	**Вспомогательные данные**	
16	a	7,8983
17	$в$	1,4738
18	$\bar{x}$	19,444
19	S	2,3518
20	n	9
21	уровень надежности	0,9
22	уровень значимости	0,1
23	число степеней свободы	7
24	критическое значение t-статистики	1,8946
26	**Завершающие данные**	
27	m_p	2,5474
28	Δ_p	4,8263
29	y_p	30,742
30	нижняя граница	25,916
31	верхняя граница	35,569
32	δ_p	15,699



80% 90% 95% Лист2 Лист3

Рис. 5.4. Прогнозирование на основании линейной модели при уровне надежности 90%

Исходные данные

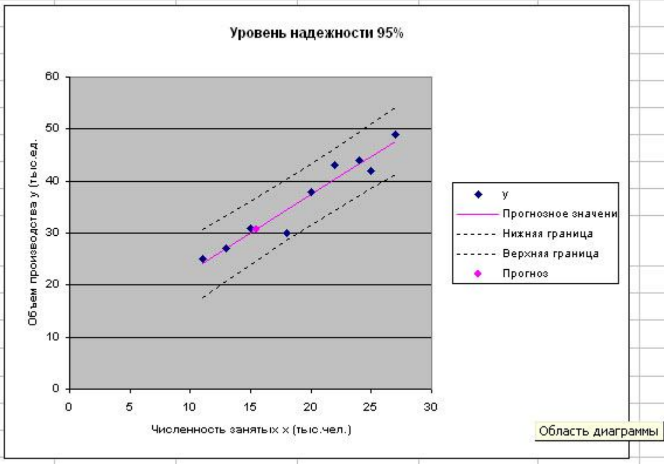
№ предприятия	x	y	$(x - x_{ср})^2$	Прогнозное значение	Стандартная ошибка проноза	Прельная ошибка прогноза	Нижняя граница	Верхняя граница
1	11	25	71,3086	24,110	2,779	6,571	17,539	30,681
2	13	27	41,5309	27,058	2,658	6,285	20,773	33,342
3	15	31	19,7531	30,005	2,566	6,067	23,939	36,072
4	18	30	2,0864	34,427	2,488	5,884	28,543	40,311
5	20	38	0,3086	37,374	2,480	5,865	31,509	43,240
6	22	43	6,5309	40,322	2,508	5,930	34,392	46,252
7	24	44	20,7531	43,270	2,570	6,077	37,193	49,346
8	25	42	30,8642	44,743	2,613	6,179	38,565	50,922
9	27	49	57,0864	47,691	2,722	6,436	41,255	54,127
	15,5	Сумма	250,2222					

Вспомогательные данные

a	7,8983
$в$	1,4738
$\bar{x}$	19,444
S	2,3518
n	9
уровень надежности	0,95
уровень значимости	0,05
число степеней свободы	7
критическое значение t-статистики	2,3646

Завершающие данные

$m_р$	2,5474
$\Delta_р$	6,0238
$y_р$	30,742
нижняя граница	24,718
верхняя граница	36,766
$\delta_р$	19,594



80% / 90% / 95% / Лист2 / Лист3
Готово

Рис. 6.4. Прогнозирование на основании линейной модели при уровне надежности 95%

Сравним относительные погрешности прогнозов при различных уровнях надежности, для x_p=15,5:

Уровень надежности	80%	90%	95%
Относительная погрешность	11,7%	15,7%	19,6%

Повышение уровня надежности с 80% до 95% снижает точность прогноза в 19,6/11,7≈1,68 раза.

Отчет по лабораторной работе

1. Запишите завершающие данные: прогнозное значение результата, стандартная и предельная ошибки прогноза, доверительный интервал прогноза, относительная погрешность прогноза.
2. Сравнить относительные погрешности прогнозов при различных уровнях надежности, например, для x_p=20.

Лабораторная работа №5

Многофакторная линейная регрессия. Мультиколлинеарность

Цель: научиться проверять факторы на мультиколлинеарность; находить уравнение многофакторной линейной регрессии, проверять его качество; находить средние коэффициенты эластичности.

Краткие теоретические сведения.

Множественная регрессия – уравнения связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, ..., x_p),$$

где y – зависимая переменная (результативный признак);

$x_1, x_2, \ldots, x_p$ – независимые переменные (факторы).

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов* (МНК).

Линейное уравнение регрессии имеет вид: $y=b_1x_1+b_2x_2+a$.

Средний коэффициент эластичности $\overline{Э}$ показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины при изменении фактора x_j на 1% от своего среднего значения. Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле:

$$\overline{Э}_{yx_j} = b_j \frac{\overline{x}_j}{\overline{y}}.$$

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y - \overline{y})^2}.$$

F-тест – оценивание качества уравнения регрессии – состоит в проверке гипотезы H_0 *о статистической незначимости уравнения регрессии и показателя тесноты связи*. Для этого выполняется сравнение фактического $F_{факт}$ и критического(табличного) $F_{табл}$ значений *F-критерия Фишера*. $F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}.$$

Число степеней свободы критерия Фишера-Снедекора: $k_1=m$; $k_2= n-m-1$ и критическое значение этого критерия $F_{кр}= F_{\alpha;k_1;k_2}$.

Если $F_{факт} > F_{табл}$, то гипотеза H_0 о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{факт} < F_{табл}$, то гипотеза H_0 не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Оценка значимости коэффициентов регрессии с помощью *t*-критерия Стьюдента сводится к вычислению наблюдаемого значения *t*-статистики:

$$t_{j\ набл} = \frac{b_j}{m_{b_j}};$$

$$t_{a\ набл} = \frac{a}{m_a}.,$$

где

m_{b_i} – средняя квадратическая ошибка коэффициента регрессии b_i, она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \dots x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \dots x_p}}} \cdot \frac{1}{\sqrt{n - m - 1}}.$$

m_a – средняя квадратическая ошибка коэффициента регрессии *a*.

Критическое значение *t*-статистики: $t_{кр}=t_{\alpha;n-m-1}$ где $k=n-m-1$ – число степеней свободы, *m* – число факторов;

Если $|t_{набл}|>t_{кр}$, то коэффициент регрессии статистически значим; в противном случае – статистически незначим.

При построении уравнений множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов, их тесной линейной связанности.

Считая, что две переменные явно *коллинеарны*, то есть находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0,7$.

Статистическая функция ЛИНЕЙН

Анализ многофакторных моделей в Excel удобно проводить с помощью статистической функции ЛИНЕЙН, которая на основе МНК рассчитывает массив данных, описывающих уравнение линейной многофакторной регрессии $y=b_1x_1+b_2x_2+a$.

Синтаксис

ЛИНЕЙН(известные_значения_y;известные_значения_x;конст;статистика)

Известные_значения_y – множество значений *y*.

Известные_значения_x – диапазон известных значений факторов.

Конст – логическое значение, которое указывает, требуется ли, чтобы константа (свободный член) *a* была равна 0.

Если конст имеет значение 1 или опущено, то *a* вычисляется обычным образом.

Если аргумент конст имеет значение 0, то свободный член полагается равным нулю.

Статистика – логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии.

Если аргумент статистика имеет значение 1 то функция ЛИНЕЙН возвращает дополнительную регрессионную статистику.

Если аргумент статистика имеет значение 0 или опущен, то функция ЛИНЕЙН не возвращает дополнительную регрессионную статистику.

В случае, когда значение аргумента **стат** равно 1, таблица результатов, выводимых функцией ЛИНЕЙН имеет вид:

b_2	b_1	a
m_{b2}	m_{b1}	m_a
R^2	S	
$F_{набл}$	k	
$S_{факт}$	$S_{остат}$	

Величина	Описание
m_{b2} m_{b1} m_a	Стандартные значения ошибок для коэффициентов.
R^2	Коэффициент детерминации.
S	Стандартная ошибка регрессии.
$F_{набл}$	F-статистика, или F-наблюдаемое значение. F-статистика используется для определения того, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет.
k	Степени свободы. $k=n–m–1$.
$S_{факт}$	Регрессионная сумма квадратов.
$S_{остат}$	Остаточная сумма квадратов.

Регрессия и Excel

Excel позволяет при построении уравнения линейной регрессии большую часть работы сделать очень быстро. Важно понять, как интерпретировать полученные результаты. Воспользуемся надстройкой *Пакет анализа.*

Сервис – Анализ данных – Регрессия – ОК. Появляется диалоговое окно, которое нужно заполнить. В графе *Входной интервал Y:* указывается ссылка на ячейки, содержащие значения результативного признака *y.* В графе *Входной интервал X:* указывается ссылка на ячейки, содержащие значения факторов x_1, ..., x_m (m < 16). Если первые из ячеек содержат пояснительный текст, то рядом со словом *Метки* нужно поставить «галочку». *Уровень надежности* (доверительная вероятность) по умолчанию предполагается равным 95%. Если исследователя это значение не устраивает, то рядом со словами *Уровень надежности* нужно поставить «галочку» и указать требуемое значение. Поставив «галочку» рядом со словом *константа-ноль,* исследователь получит $a = 0$ по умолчанию. Если нужны значения остатков e_t и их график, то нужно поставить «галочки» рядом со словами *Остатки* и *График остатков. ОК.* Появляется итоговое окно.

Если число в графе *Значимость F* превышает 1 – *Уровень надежности,* то принимается гипотеза $R^2 = 0$. Иначе принимается гипотеза $R^2 \neq 0$.

Р-значение – это значения уровней значимости, соответствующие вычисленным t-статистикам.

Нижние 95% и Верхние 95% – это нижние и верхние границы 95% доверительных интервалов для коэффициентов теоретического уравнения линейной регрессии. Если исследователь согласился с принятым по умолчанию значением доверительной вероятности 95%, то последние два столбца будут дублировать два предыдущих столбца. Если исследователь вводил свое значение доверительной вероятности p, то последние два столбца содержат значения соответственно нижней и верхней границы p-процентных доверительных интервалов.

Порядок выполнения работы.

1. **Проверить факторы на мультиколлинеарность**. Поскольку их всего два, то следует проверить их на коллинеарность по значению парного коэффициента корреляции. Поэтому необходимо:
   1) Найти коэффициент корреляции между факторами r_{12}.
   2) Проверить статистическую значимость полученного коэффициента корреляции при α=0,05. Для этого:
      - найти наблюдаемое значение t-статистики:

$$t_{набл} = r_{12}\sqrt{\frac{n-2}{1-r_{12}^2}};$$

      - найти критическое значение t-статистики: $t_{кр}=t_{\alpha;n-2}$ где $к=n-2$ – число степеней свободы;
      - сделать вывод о значимости коэффициента корреляции r_{12} и, значит о наличие коллинеарности: если $|t_{набл}|>t_{кр}$, то коэффициент корреляции статистически значим и факторы коллинеарны; в противном случае – факторы неколлинеарны.

2. **Найти линейное уравнение регрессии $y=b_1x_1+b_2x_2+a$ и проверить его качество**. Для этого необходимо:
   1) Найти коэффициенты уравнения b_1, b_2, a.

2) Для проверки статистической значимости коэффициентов уравнения b_1, b_2, a при α=0,05 необходимо:
   - найти наблюдаемые значения t-статистик:

$$t_{j\ набл} = \frac{b_j}{m_{b_j}};$$

$$t_{a\ набл} = \frac{a}{m_a}.$$

   - найти критическое значение t-статистики: $t_{кр}=t_{\alpha;n-m-1}$ где $k=n-m-1$ – число степеней свободы, m – число факторов;
   - сделать вывод о значимости каждого из коэффициентов b_1, b_2, a: если $|t_{набл}|>t_{кр}$, то коэффициент регрессии статистически значим; в противном случае – статистически незначим.

3) Для проверки общего качества уравнения регрессии проверить значимость коэффициента детерминации. Для этого:
   - найти коэффициент детерминации;
   - найти наблюдаемое значение F-статистики

$$F_{факт} = \frac{R^2}{1-R^2} \cdot (n-2);$$

   - найти критическое значение F-статистики. Число степеней свободы критерия Фишера-Снедекора: $k_1=1$; $k_2= n-m-1$ и критическое значение этого критерия $F_{кр}= F_{\alpha;k_1;k_2}$.
   - сделать вывод о значимости уравнения регрессии: если $F_{факт} > F_{кр}$, то уравнение регрессии статистически значимо и надежно, если $F_{факт} < F_{кр}$ признается статистическая незначимость, ненадежность уравнения регрессии.

3. **Найти средние коэффициенты эластичности** по переменным x_1 и x_2 по формуле:

$$\bar{Э}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}.$$

Пример выполнения лабораторной работы

Дана зависимость объема выпускаемой продукции (млн. ден. ед.) от соответствующих затратах на рабочую силу x_1 (млн. ден. ед.) и средств производства x_2 (млн. ден. ед.) для десяти регионов страны. Необходимо в соответствии со сформулированным выше заданием построить и проанализировать линейную двухфакторную модель.

№ п/п	x_1	x_2	y
1	1,83	1,41	12,2
2	1,41	1,43	7,89
3	1,31	1,16	8,36
4	0,83	1,54	6,9
5	0,86	1,01	5,74
6	1,18	1,84	9,77
7	1,75	1,78	10,99
8	1,13	1,16	6,95
9	0,95	1,99	9,97
10	1,4	1,68	9,16

I. Проверка факторов на мультиколлинеарность

1. Подготовим исходные данные, как показано на рис 1.5.

2. Вводим следующие формулы:

Ячейка	Формула	Примечание
B15	=КОРРЕЛ(B3:B12;C3:C12)	Коэффициент корреляции между факторами
B16	=B15*((A12–2)/(1–B15*B15))^0,5	Наблюдаемое значение t-статистики коэффициента корреляции
B17	=СТЬЮДРАСПОБР(0,05;A12-2)	Критическое значение t-статистики коэффициента корреляции
B18	=ЕСЛИ(ABS(B16)>B17;"факторы мультиколлинеарны";"факторы не мультиколлинерны")	Выводы о проверке на мультиколлинеарность

	A	B	C	D	E	F
1	**Линейная многофакторная регрессия**					
2	№ п/п	x_1	x_2	y		
3	1	1,83	1,41	12,2		
4	2	1,41	1,43	7,89		
5	3	1,31	1,16	8,36		
6	4	0,83	1,54	6,9		
7	5	0,86	1,01	5,74		
8	6	1,18	1,84	9,77		
9	7	1,75	1,78	10,99		
10	8	1,13	1,16	6,95		
11	9	0,95	1,99	9,97		
12	10	1,4	1,68	9,16		
13						
14	**Проверка мультиколлинеарности**					
15	r_{12}	0,1261				
16	$t_{набл}$	0,3595				
17	$t_{кр}$	2,306				
18	Вывод	факторы не мультиколлинерны				

Рис. 1.5. Проверка мультиколлинеарности

II. Уравнение линейной регрессии

1. Найдем массив данных, описывающих уравнение линейной многофакторной регрессии (см. рис. 2.5). Для этого:

- выделим мышью весь диапазон C22:E26;
- вызовем функцию ЛИНЕЙН, в окне которой введем следующие значения параметров: **Изв_знач_*y*** – D3:D12, **Изв_знач_*x*** – B3:C12, **Конст** – 1, **Статистика** – 1;
- после ввода параметров вместо кнопки **ОК** нажмем одновременно три клавиши: <Shift>+<Ctrl>+<Enter>.

	A	B	C	D	E
19					
20	**Анализ уравнения регрессии**				
21			b_2	b_1	a
22	Коэффициенты модели		3,177	4,042	-1,085
23	Стандартные ошибки коэффициентов		0,939	0,885	1,710
24	R^2	s	0,840	0,909	#Н/Д
25	F-статистика	k	18,397	7	#Н/Д
26	$S_{факт}$	$S_{остат}$	30,425	5,788	#Н/Д
27					
28	t-статистики коэффициентов		3,382	4,569	0,635
29	Значимость коэффициентов		Значим	Значим	Незначим
30	$t_{кр}$		2,365		
31	$F_{кр}$		4,737		
32	Значимость уравнения		Значимо		

Рис. 2.5. Анализ уравнения регрессии

Из полученных данных следует, что уравнение линейной регрессии имеет вид:

$$y = 4{,}042x_1 + 3{,}177x_2 - 1{,}085.$$
$$(0{,}885) \quad (0{,}939) \quad (1{,}71)$$

В скобках указаны стандартные ошибки соответствующих коэффициентов.

2. Вводим следующие формулы:

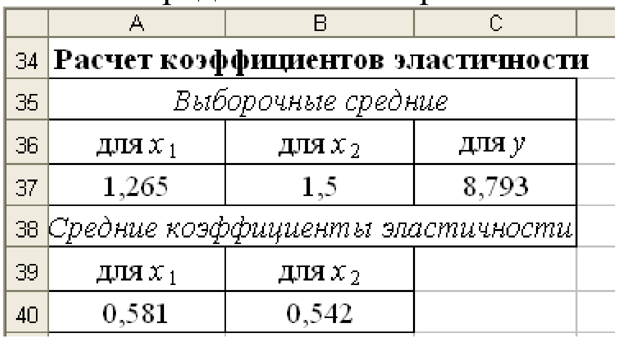
Ячейка	Формула	Примечание
C28	=ABS(C22)/C23	*t*-статистика коэффициентов. Копируем до E28.
C30	=СТЬЮДРАСПОБР(0,05;A12-2-1)	Критическое значение *t*-статистики коэффициента корреляции.
C31	=FРАСПОБР(0,05;A12-1-D25;D25)	Критическое значение *F*-статистики.
C29	=ЕСЛИ(C28>$C$30;"Значим";"Незначим")	Выводы о значимости коэффициентов. Копируем до E29.
C32	=ЕСЛИ(C25>C31;"Значимо";"Незначимо")	Вывод о значимости уравнения регрессии

III. Средние коэффициенты эластичности

Вводим следующие формулы:

Ячейка	Формула	Примечание
A37	=СРЗНАЧ(B3:B12)	Средние значения фактора и результата. Копируем до C37.
A40	=D22*A37/C37	Средний коэффициент эластичности для x_1
B40	=C22*B37/C37	Средний коэффициент эластичности для x_2

Результаты вычислений представлены на рис. 5.3.



	A	B	C
34	**Расчет коэффициентов эластичности**		
35	*Выборочные средние*		
36	для x_1	для x_2	для y
37	1,265	1,5	8,793
38	*Средние коэффициенты эластичности*		
39	для x_1	для x_2	
40	0,581	0,542	

Рис. 3.5. Расчет средних коэффициентов эластичности

Следовательно при увеличении затрат на рабочую силу на 1% объем выпускаемой продукции увеличится в среднем на 0,581%, при увеличении затрат на средства производства на 1% объем выпускаемой продукции увеличится в среднем на 0,542%.

IV. Регрессия и Excel.

1. В главном меню выберите **Сервис** – **Анализ данных** – **Регрессии**. Щелкните по кнопке **ОК**.

2. Заполните диалоговое окно ввода данных и параметров вывода (рис. 4.5):

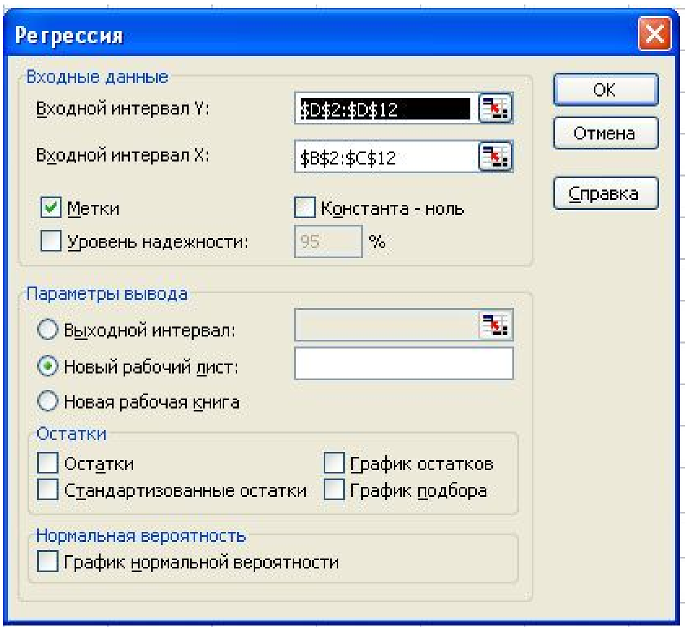


Рис. 4.5. Диалоговое окно ввода параметров инструмента **Регрессия**

3. Результаты регрессионного анализа представлены на рис. 5.5.

	A	B	C	D	E	F	G
1	**ВЫВОД ИТОГОВ**						
2							
3	*Регрессионная статистика*						
4	Множественный R	0,917					
5	R-квадрат	0,840					
6	Нормированный R-квадрат	0,794					
7	Стандартная ошибка	0,909					
8	Наблюдения	10					
9							
10	**Дисперсионный анализ**						
11		*df*	*SS*	*MS*	*F*	*Значимость F*	
12	Регрессия	2	30,425	15,212	18,397	0,002	
13	Остаток	7	5,788	0,827			
14	Итого	9	36,213				
15							
16		*Коэффициенты*	*Стандартная ошибка*	*t-статисти*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
17	*a*	-1,085	1,710	-0,635	0,546	-5,127	2,957
18	b_1	4,042	0,885	4,569	0,003	1,950	6,134
19	b_2	3,177	0,939	3,382	0,012	0,956	5,398

Рис. 5.5. Результат применения инструмента **Регрессия**

Индивидуальное задание к лабораторной работе

Исследуется зависимость между стоимостью грузовой автомобильной перевозки, весом груза и расстоянием.

№ компании	Стоимость грузовой автомобильной перевозки *y* (тыс. руб.)									
	№ варианта									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	51,0	56,1	20,4	84,2	38,8	127,5	71,4	182,2	83,8	89,8
2	16,0	17,6	6,4	26,4	12,2	40,0	22,4	57,2	26,3	28,2
3	74,0	81,4	29,6	122,1	56,2	185,0	103,6	264,3	121,6	130,2
4	7,5	8,3	3,0	12,4	5,7	18,8	10,5	26,8	12,3	13,2
5	33,0	36,3	13,2	54,5	25,1	82,5	46,2	117,9	54,2	58,1
6	26,0	28,6	10,4	42,9	19,8	65,0	36,4	92,9	42,7	45,8
7	11,5	12,7	4,6	19,0	8,7	28,8	16,1	41,1	18,9	20,2
8	52,0	57,2	20,8	85,8	39,5	130,0	72,8	185,7	85,4	91,5
9	15,8	17,4	6,3	26,1	12,0	39,5	22,1	56,4	26,0	27,8
10	8,0	8,8	3,2	13,2	6,1	20,0	11,2	28,6	13,1	14,1
11	26,0	28,6	10,4	42,9	19,8	65,0	36,4	92,9	42,7	45,8
12	6,0	6,6	2,4	9,9	4,6	15,0	8,4	21,4	9,9	10,6
13	5,8	6,4	2,3	9,6	4,4	14,5	8,1	20,7	9,5	10,2
14	13,8	15,2	5,5	22,8	10,5	34,5	19,3	49,3	22,7	24,3
15	6,2	6,8	2,5	10,2	4,7	15,5	8,7	22,1	10,2	10,9
16	7,9	8,7	3,2	13,0	6,0	19,8	11,1	28,2	13,0	13,9
17	5,4	5,9	2,2	8,9	4,1	13,5	7,6	19,3	8,9	9,5

18	56,0	61,6	22,4	92,4	42,6	140,0	78,4	200,0	92,0	98,6
19	25,5	28,1	10,2	42,1	19,4	63,8	35,7	91,1	41,9	44,9
20	7,1	7,8	2,8	11,7	5,4	17,8	9,9	25,4	11,7	12,5
	Вес груза x_1 (тонн)									
	№ варианта									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	35,0	45,5	14,0	54,6	26,6	91,0	54,6	125,0	61,3	72,8
2	16,0	20,8	6,4	25,0	12,2	41,6	25,0	57,2	28,0	33,3
3	18,0	23,4	7,2	28,1	13,7	46,8	28,1	64,3	31,5	37,4
4	2,0	2,6	0,8	3,1	1,5	5,2	3,1	7,1	3,5	4,2
5	14,0	18,2	5,6	21,8	10,6	36,4	21,8	50,0	24,5	29,1
6	33,0	42,9	13,2	51,5	25,1	85,8	51,5	117,9	57,8	68,6
7	20,0	26,0	8,0	31,2	15,2	52,0	31,2	71,4	35,0	41,6
8	25,0	32,5	10,0	39,0	19,0	65,0	39,0	89,3	43,8	52,0
9	13,0	16,9	5,2	20,3	9,9	33,8	20,3	46,4	22,8	27,0
10	2,0	2,6	0,8	3,1	1,5	5,2	3,1	7,1	3,5	4,2
11	21,0	27,3	8,4	32,8	16,0	54,6	32,8	75,0	36,8	43,7
12	11,0	14,3	4,4	17,2	8,4	28,6	17,2	39,3	19,3	22,9
13	3,0	3,9	1,2	4,7	2,3	7,8	4,7	10,7	5,3	6,2
14	3,5	4,6	1,4	5,5	2,7	9,1	5,5	12,5	6,1	7,3
15	2,8	3,6	1,1	4,4	2,1	7,3	4,4	10,0	4,9	5,8
16	17,0	22,1	6,8	26,5	12,9	44,2	26,5	60,7	29,8	35,4
17	3,4	4,4	1,4	5,3	2,6	8,8	5,3	12,1	6,0	7,1
18	24,0	31,2	9,6	37,4	18,2	62,4	37,4	85,7	42,0	49,9
19	9,0	11,7	3,6	14,0	6,8	23,4	14,0	32,1	15,8	18,7
20	4,5	5,9	1,8	7,0	3,4	11,7	7,0	16,1	7,9	9,4
	Расстояние x_2 (тыс. км)									
	№ варианта									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2,0	2,4	0,6	3,8	1,1	5,4	2,2	5,1	2,5	3,8
2	1,1	1,3	0,3	2,1	0,6	3,0	1,2	2,8	1,4	2,1
3	2,6	3,1	0,8	4,9	1,5	6,9	2,8	6,5	3,1	4,9
4	1,7	2,0	0,5	3,3	1,0	4,6	1,8	4,4	2,1	3,3
5	2,4	2,9	0,7	4,6	1,4	6,5	2,6	6,2	3,0	4,6
6	1,6	1,9	0,5	3,0	0,9	4,2	1,7	4,0	1,9	3,0
7	0,6	0,7	0,2	1,2	0,3	1,6	0,6	1,5	0,7	1,2
8	2,3	2,8	0,7	4,4	1,3	6,2	2,5	5,9	2,8	4,4
9	1,4	1,7	0,4	2,7	0,8	3,8	1,5	3,6	1,7	2,7
10	2,1	2,5	0,6	4,0	1,2	5,7	2,3	5,4	2,6	4,0
11	1,3	1,6	0,4	2,5	0,7	3,5	1,4	3,3	1,6	2,5
12	0,4	0,4	0,1	0,7	0,2	0,9	0,4	0,9	0,4	0,7
13	1,7	2,0	0,5	3,2	0,9	4,5	1,8	4,2	2,0	3,2

14	2,9	3,5	0,9	5,6	1,7	7,8	3,1	7,4	3,6	5,6
15	0,8	0,9	0,2	1,4	0,4	2,0	0,8	1,9	0,9	1,4
16	0,6	0,7	0,2	1,2	0,3	1,6	0,6	1,5	0,7	1,2
17	0,9	1,1	0,3	1,7	0,5	2,4	1,0	2,3	1,1	1,7
18	2,5	3,0	0,8	4,8	1,4	6,8	2,7	6,4	3,1	4,8
19	2,2	2,6	0,7	4,2	1,3	5,9	2,4	5,6	2,7	4,2
20	1,0	1,1	0,3	1,8	0,5	2,6	1,0	2,4	1,2	1,8

Отчет по лабораторной работе

1. Сформулируйте требования, предъявляемые к факторам для включения их в модель множественной регрессии.
2. Мультиколлинеарны ли факторы для вашего варианта? Почему?
3. Запишите уравнение линейной множественной регрессии для вашего варианта и интерпретируйте оценки параметров регрессии.
4. Как оценивается значимость параметров уравнения регрессии?
5. Являются ли параметры уравнения регрессии для вашего варианта значимыми и почему?
6. Запишите доверительные интервалы для параметров уравнения регрессии для вашего варианта.
7. Каким образом осуществляется проверка значимости уравнения в целом.
8. Значимо ли уравнение регрессии для вашего варианта и почему?
9. Найти частные уравнения регрессии.
10. Найти средние коэффициенты эластичности. Сделать выводы.
11. Сравните полученные результаты с результатами, полученными с помощью инструмента анализа данных **Регрессия**.

Лабораторная работа №6

Построение линейного, логарифмического, полиномиального, степенного и экспоненциального трендов

Цель: научиться проводить расчет параметров линейного, логарифмического, полиномиального, степенного и экспоненциального трендов, строить графики ряда динамики и трендов. Уметь выбирать наилучший вид трендов на основании графического изображения и значения коэффициента детерминации.

Порядок выполнения работы.

1. Провести расчет параметров линейного, логарифмического, полиномиального, степенного и экспоненциального трендов;
2. Построить графики ряда динамики и трендов;
3. Выбрать наилучший вид трендов на основании графического изображения и значения коэффициента детерминации.

Пример выполнения лабораторной работы

Динамика выпуска продукции некоторой страны характеризуется данными (усл. ед.), представленными в таблице:

Год	Выпуск продукции		Год	Выпуск продукции		Год	Выпуск продукции
1961	1054		1973	3837		1985	13617
1962	1104		1974	5490		1986	16356
1963	1149		1975	5502		1987	20037
1964	1291		1976	6342		1988	21748
1965	1427		1977	7665		1989	23298
1966	1505		1978	8570		1990	26570
1967	1513		1979	11172		1991	23080
1968	1635		1980	14150		1992	23981
1969	1987		1981	14004		1993	23446
1970	2306		1982	13088		1994	29658
1971	2367		1983	12518		1995	39573
1972	2913		1984	13471		1996	38435

В диапазоне A4:C39 (рис. 1.6) введем исходные данные.

Для определения параметров линейного тренда по методу наименьших квадратов воспользуемся статистической функцией **ЛИНЕЙН**. В качестве зависимой переменной в данном примере выступает время (t = 1, 2,…, n). Порядок вычисления был рассмотрен в лабораторной работе №1.

	A	B	C	D	E	F	G
1	**Построение трендов**						
2							
3	Год, *x*	Время, *t*	Выпуск продукции, *y*				
4	1961	1	1054				
5	1962	2	1104				
6	1963	3	1149		**ЛИНЕЙН**		
7	1964	4	1291		977,11982	-5969,5222	
8	1965	5	1427		60,6780848	1287,4149	
9	1966	6	1505		0,88408467	3782,05096	
10	1967	7	1513		259,317553	34	
11	1968	8	1635		3709254808	486332922	
12	1969	9	1987				
13	1970	10	2306				
14	1971	11	2367				
15	1972	12	2913				
16	1973	13	3837				
17	1974	14	5490				
18	1975	15	5502				
19	1976	16	6342				
20	1977	17	7665				
21	1978	18	8570				
22	1979	19	11172				
23	1980	20	14150				
24	1981	21	14004				
25	1982	22	13088				
26	1983	23	12518				
27	1984	24	13471				
28	1985	25	13617				
29	1986	26	16356				
30	1987	27	20037				
31	1988	28	21748				
32	1989	29	23298				
33	1990	30	26570				
34	1991	31	23080				
35	1992	32	23981				
36	1993	33	23446				
37	1994	34	29658				
38	1995	35	39573				
39	1996	36	38435				

Рис. 1.6 Исходные данные

Запишем уравнения линейного тренда, используя данные рис.1.6:

$$y_t = -5969{,}52 + 977{,}12t.$$

Выделим диапазон A4:A39 и, нажав **Ctrl**, выделим диапазон C4:C39, выбрав в **Мастере диаграмм** тип – **График**, построим график зависимости выпуска продукции от времени (см. рис. 2.6):

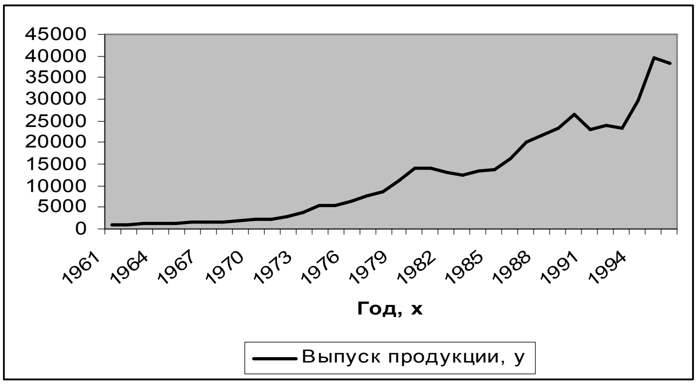


Рис.2.6. Динамика выпуска продукции.

В ППП MS Excel линия тренда может быть добавлена в диаграмму с областями гистограммы или в график. Для этого:

1. выделите область построения диаграммы; в главном меню выберите **Диаграмма/ Добавить линию тренда**;
2. в появившемся диалоговом окне (рис.3.6.) выберите вид линии тренда и задайте соответствующие параметры. Для полиномиального тренда необходимо задать степень аппроксимирующего полинома, для скользящего среднего – количество точек усреднения.

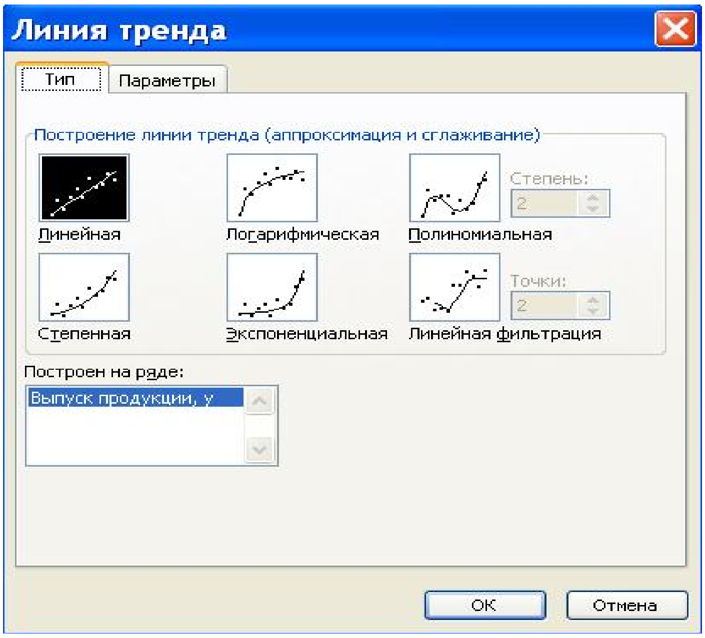


Рис. 3.6. Диалоговое окно типов линий тренда

В качестве дополнительной информации на диаграмме можно отобразить уравнение регрессии и значение среднеквадратического отклонения, установив соответствующие значки на закладке Параметры (рис.4.6.). Щелкните по кнопке **ОК**.

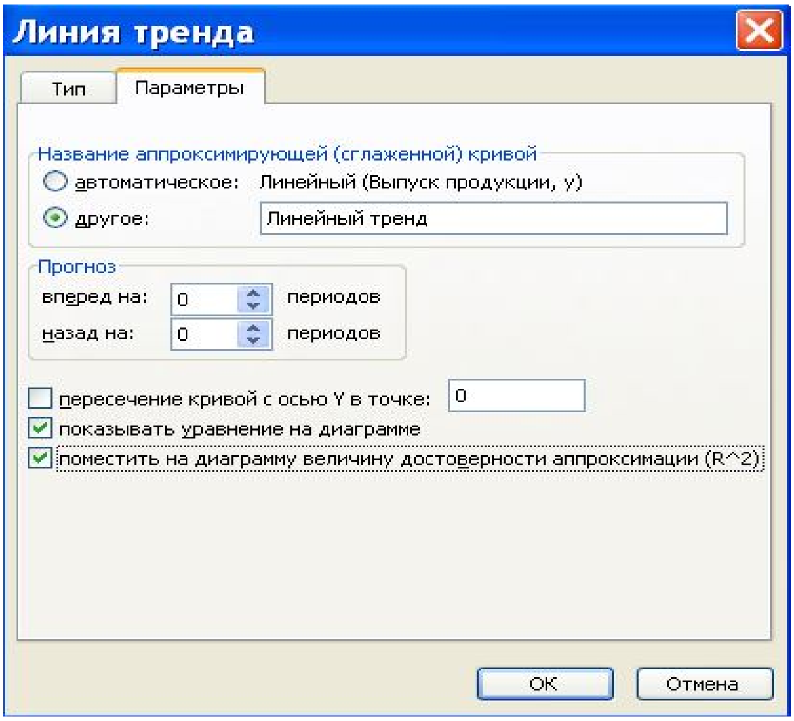


Рис. 4.6. Диалоговое окно параметров линии тренда

На рис.5.6.– 9.6. представлены различные виды трендов, описывающие исходные данные задачи.

Рис. 5.6. Линейный тренд

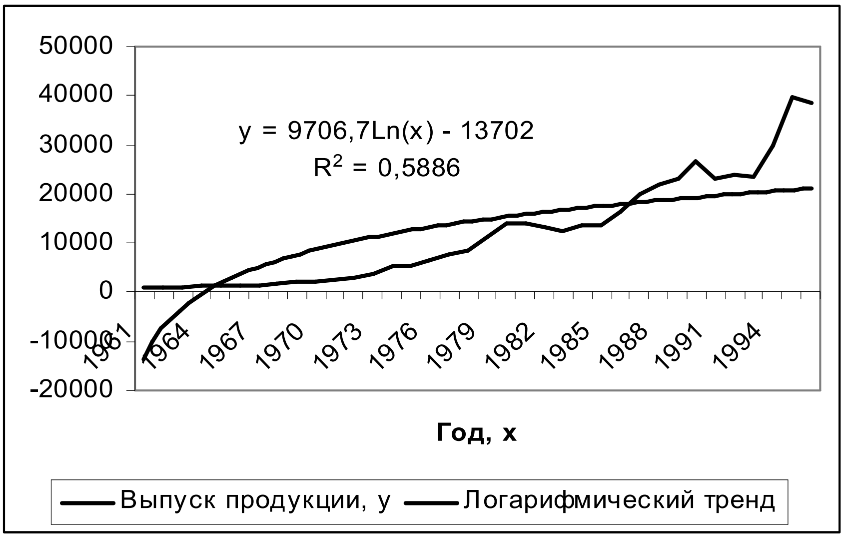


Рис.6.6. Логарифмический тренд

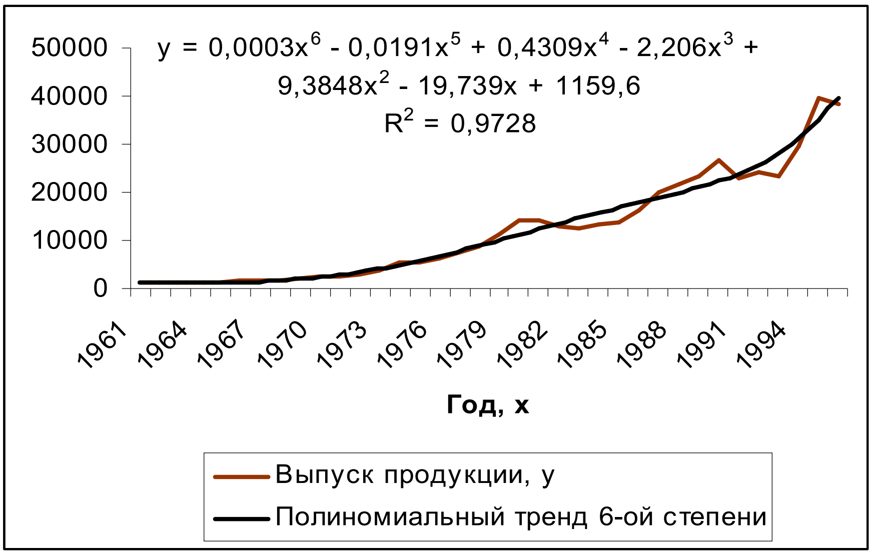


Рис.7.6. Полиномиальный тренд

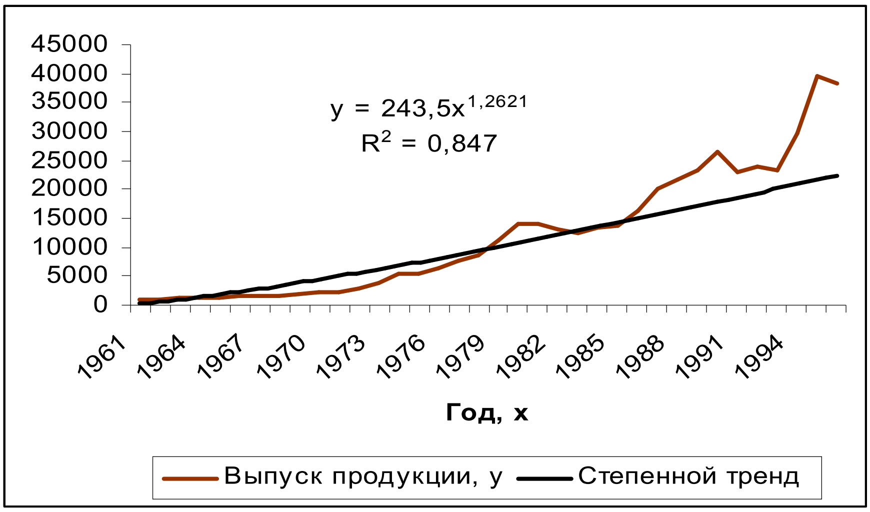


Рис.8.6. Степенной тренд

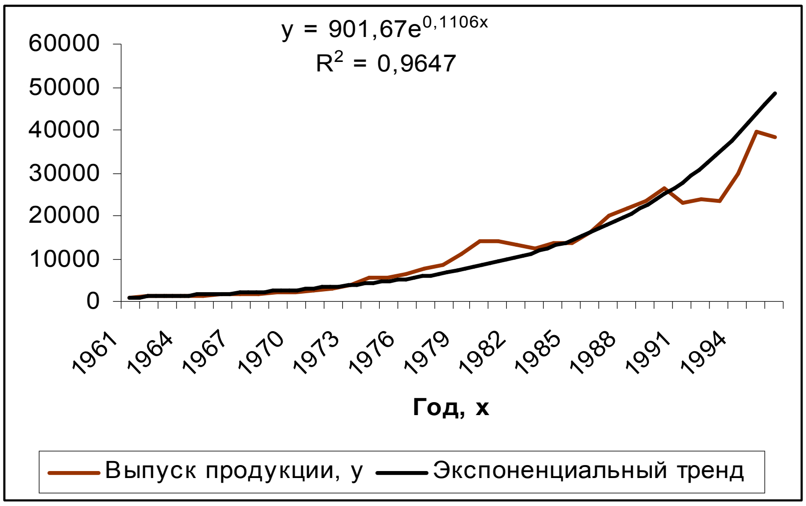


Рис.9.6. Экспоненциальный тренд

Сравним значения коэффициентов детерминации по разным уравнениям трендов:

Полиномиальный 6-й степени	0,9738;
Экспоненциальный	0,9647;
Линейный	0,8841;
Степенной	0,847;
Логарифмический	0,5886.

Исходные данные лучше всего описывает полином 6-й степени. Следовательно, в рассматриваемом примере для расчета прогнозных значений следует использовать полиномиальное уравнение.

Индивидуальное задание к лабораторной работе

№	Год									
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
1	25	27	30	29	30	35	33	40	40	42
2	29,4	23,5	26,2	48,5	73,4	56,6	77	78,4	81,2	82
3	183,5	153,5	140,7	107,1	87,5	68,3	83,1	84,2	79,6	80
4	105,3	94,9	92	83,9	72,7	56,9	49,1	46,5	39,7	35,5
5	130	150	296	542	254	274	272	262	369	334
6	41	42	49	64	53	44	52	51	71	92
7	252	217	210	229	302	320	270	287	291	237
8	97	89	77	81	87	94	90	90	93	87
9	18,5	15,1	30,8	34,4	25,4	35,1	36,9	44,7	57,3	64,6
10	143	130,3	149,9	296,6	541,5	363,2	254,1	272,4	368,4	334,3

Отчет по лабораторной работе

1. Записать уравнения линейного, логарифмического, полиномиального, степенного и экспоненциального трендов.
2. Выбрать наилучший вид трендов на основании значения коэффициента детерминации.

Лабораторная работа №7

Автокорреляция уровней временного ряда и выявление его структуры

Цель: научиться находить коэффициенты автокорреляции.

Краткие теоретические сведения.

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов), называются *моделями временных рядов*.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов.

Каждый уровень временного ряда формируется из *трендовой* (T), *циклической (сезонной)* (S) и случайной (E) компонент.

Автокорреляция уровней ряда – это корреляционная зависимость между последовательными уровнями временного ряда:

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

где $\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}$; $\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}$ – коэффициент автокорреляции уровней ряда первого порядка;

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

где $\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}$; $\bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}$ – коэффициент автокорреляции уровней ряда второго порядка.

Формулы для расчета коэффициентов автокорреляции старших порядков легко получить из формулы линейного коэффициента корреляции.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют *автокорреляционной функцией* временного ряда, а график зависимости ее значений от величины лага (число периодов, по которым рассчитывается коэффициент автокорреляции) – *коррелограммой*.

Порядок выполнения работы.

Для заданного временного ряда:

1. Найти коэффициенты автокорреляции до пятого порядка.
2. Построить коррелограмму.

Пример выполнения лабораторной работы

Представлены поквартальные данные о валовом объеме продаж (млн.шт.) за последние четыре года:

Квартал	Год			
	1	2	3	4
I	6	7,2	8	9
II	4,4	4,8	5,6	6,6
III	5	6	6,4	7
IV	9	10	11	10,8

1. В диапазоне B6:E9 (рис. 1.7) введем исходные данные.

	A	B	C	D	E
1	**Анализ временного ряда**				
2					
3	**Исходные данные**				
4	Квартал	Год			
5		1	2	3	4
6	I	6	7,2	8	9
7	II	4,4	4,8	5,6	6,6
8	III	5	6	6,4	7
9	IV	9	10	11	10,8
10					
11					
12	**Коэффициент автокорреляции**				
13	1-го порядка	0,1652			
14	2-го порядка	-0,5669			
15	3-го порядка	0,1136			
16	4-го порядка	0,9830			
17	5-го порядка	0,1187			

	G	H	I	J	K	L	M	N
4	Квартал	Время	Временной ряд	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}
5	I	1	6					
6	II	2	4,4	6				
7	III	3	5	4,4	6			
8	IV	4	9	5	4,4	6		
9	I	5	7,2	9	5	4,4	6	
10	II	6	4,8	7,2	9	5	4,4	6
11	III	7	6	4,8	7,2	9	5	4,4
12	IV	8	10	6	4,8	7,2	9	5
13	I	9	8	10	6	4,8	7,2	9
14	II	10	5,6	8	10	6	4,8	7,2
15	III	11	6,4	5,6	8	10	6	4,8
16	IV	12	11	6,4	5,6	8	10	6
17	I	13	9	11	6,4	5,6	8	10
18	II	14	6,6	9	11	6,4	5,6	8
19	III	15	7	6,6	9	11	6,4	5,6
20	IV	16	10,8	7	6,6	9	11	6,4

Рис. 1.7. Исходные данные

Для анализа временного ряда перепишем поквартальные данные в один столбец. Для этого:

- в столбце G запишем номера кварталов;
- в столбце H запишем числа 1,2,…,16;
- чтобы перенести данные для временного ряда, в столбцы запишем формулы согласно таблице:

Ячейка	Формула	Примечание
I5	=B6	Копируем в диапазон I6:I8
I9	=C6	Копируем в диапазон I10:I12
I13	=D6	Копируем в диапазон I14:I16
I17	=E6	Копируем в диапазон I18:I20

2. В столбцах J, K, L, M, N подготовим сдвинутые соответственно 1, 2, 3, 4, 5 уровней значений временного ряда. Для этого запишем в каждую из ячеек: J6, K7, L8, M9, N10 – одну и ту же формулу:

=I5,

которую затем скопируем вниз до строки 20.

Найдем коэффициенты автокорреляции с помощью следующих формул:

Ячейка	Формула	Коэффициент автокорреляции
B13	=КОРРЕЛ(I6:I20;J6:J20)	1-го порядка
B14	=КОРРЕЛ(I7:I20;K7:K20)	2-го порядка
B15	=КОРРЕЛ(I8:I20;L8:L20)	3-го порядка
B16	=КОРРЕЛ(I9:I20;M9:M20)	4-го порядка
B17	=КОРРЕЛ(I10:I20;N10:N20)	5-го порядка

3. Выделим диапазон A13:B17 и, выбрав в **Мастере диаграмм** тип – **Гистограмма**, построим коррелограмму (рис. 2.7).

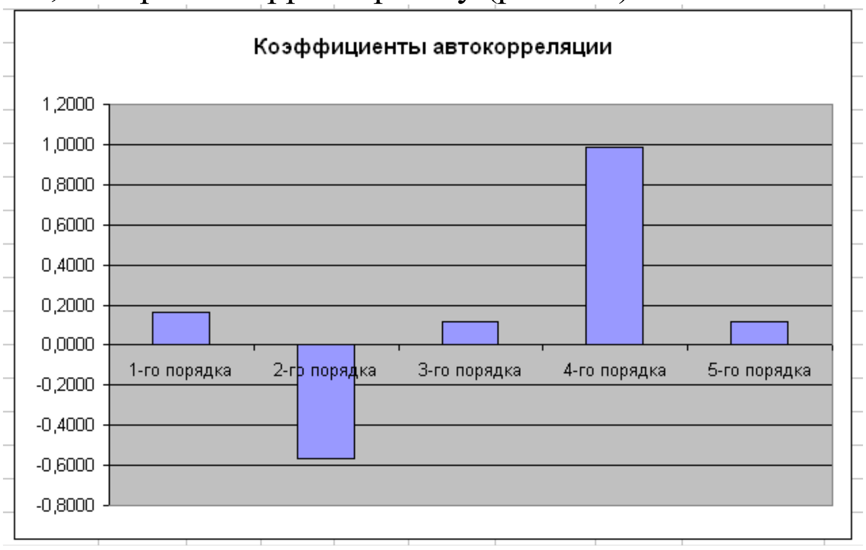


Рис. 2.7. Коррелограмма

Таким образом, коэффициенты автокорреляции временного ряда:

Порядок	1	2	3	4	5
Коэффициент автокорреляции	0,1652	– 0,567	0,1136	0,983	0,1187

Наибольшее значение имеет коэффициент автокорреляции 4-го порядка. Следовательно, временной ряд имеет тенденцию и сезонную компоненту с периодом, равным четырем кварталам.

Индивидуальное задание к лабораторной работе

Вариант 1

Квартал	Год			
	1	2	3	4
I	6,0	8,5	9,5	5,7
II	5,8	5,3	8,6	6,2
III	8,5	5,3	5,9	6,8
IV	9,9	7,9	9,7	7,0

Вариант 2

Квартал	Год			
	1	2	3	4
I	6,6	5,7	7,5	9,4
II	9,8	8,6	8,3	8,2
III	7,8	5,2	5,1	6,7
IV	6,7	6,3	5,7	8,7

Вариант 3

Квартал	Год			
	1	2	3	4
I	8,7	8,3	8,6	5,3
II	6,2	9,1	7,3	6,6
III	7,2	7,6	7,8	8,4
IV	7,0	9,4	8,6	9,6

Вариант 4

Квартал	Год			
	1	2	3	4
I	5,5	8,2	9,1	8,5
II	5,7	8,0	9,7	7,5
III	5,5	6,6	5,9	8,7
IV	6,6	6,2	5,6	6,1

Вариант 5

Квартал	Год			
	1	2	3	4
I	9,3	9,8	8,7	6,6
II	9,9	9,0	8,0	7,1
III	7,9	8,8	9,9	9,9
IV	7,5	6,1	5,7	8,0

Вариант 6

Квартал	Год			
	1	2	3	4
I	8,3	5,9	9,9	6,7
II	7,5	7,5	6,3	7,0
III	8,8	5,4	8,5	6,6
IV	5,1	6,2	8,7	6,5

Вариант 7

Квартал	Год			
	1	2	3	4
I	6,3	7,0	5,5	9,5
II	8,2	9,8	5,3	8,0
III	8,1	6,8	8,0	8,3
IV	6,0	5,2	5,4	8,5

Вариант 8

Квартал	Год			
	1	2	3	4
I	9,1	8,9	5,7	9,6
II	5,2	8,4	9,2	9,1
III	7,2	8,0	5,6	7,5
IV	8,1	7,5	9,5	5,6

Вариант 9

Квартал	Год			
	1	2	3	4
I	8,9	5,6	8,8	7,6
II	6,9	6,6	9,2	8,3
III	5,4	6,9	8,1	6,6
IV	5,0	8,0	8,0	9,8

Вариант 10

Квартал	Год			
	1	2	3	4
I	9,9	9,6	5,0	9,7
II	7,7	7,6	5,1	8,0
III	9,8	7,7	9,2	6,2
IV	7,9	8,7	5,5	9,1

Отчет по лабораторной работе

1. Что такое автокорреляция уровней временного ряда и как ее можно оценить количественно?
2. Записать значения коэффициентов автокорреляции до пятого порядка. Сделать вывод.

Лабораторная работа №8

Анализ сезонных колебаний

Цель: научиться строить аддитивную и мультипликативную модели временных рядов.

Краткие теоретические сведения.

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов), называются *моделями временных рядов*.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов.

Каждый уровень временного ряда формируется из *трендовой* (T), *циклической (сезонной)* (S) и случайной (E) компонент.

Модели, в которых временной ряд представлен как сумма перечисленных компонент, – *аддитивные модели*, как произведение – *мультипликативные модели* временного ряда.

Аддитивная модель имеет вид: $Y=T+S+E$.

Мультипликативная модель: $Y=T\cdot S\cdot E$.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S, и E для каждого уровня ряда.

Построение модели включает следующие шаги:

1) выравнивание ряда методом скользящей средней;
2) расчет значений сезонной компоненты S;
3) устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной (T+E) или в мультипликативной (T·E) модели;
4) аналитическое выравнивание уровней (T+E) или (T·E) и расчет значений T с использованием полученного уравнения тренда;
5) расчет полученных по модели значений (T+S) или (T·S);
6) расчет абсолютных и/или относительных ошибок.

Порядок выполнения работы.

Для каждого из двух заданных временных рядов:

1. Построить график временного ряда.
2. Определить вид модели и период сезонных колебаний.
3. Произвести выравнивание исходного ряда методом скользящей средней.
4. Найти значения сезонной компоненты S.
5. Устранить сезонную компоненту из исходных уровней ряда и получить выровненные данные в аддитивной (T+E) или мультипликативной (T·E) модели.
6. Произвести аналитическое выравнивание уровней (T+E) или (T·E) и рассчитать значения T с использованием полученного

уравнения тренда.
7. Получить расчетные значения: (T+S) или (T·S).
8. Определить абсолютные и относительные ошибки.
9. Найти прогноз двух последующих уровней временного ряда.

Пример выполнения лабораторной работы

Представлены поквартальные данные о валовом объеме продаж (млн.шт.) за последние четыре года:

Квартал	Год			
	1	2	3	4
I	20	28	34	31
II	31	43	43	41
III	46	53	56	55
IV	52	64	66	70

Для выполнения лабораторной работы выделим в рабочей книге два листа:

Название листа	Назначение
Анализ 1	Для анализа временного ряда 1
Анализ 2	Для анализа временного ряда 2

Временной ряд 1.

1. В диапазоне B2:E5 (рис. 1.8) введем исходные данные.

I37 fx

	A	B	C	D	E
1	**месяц**	**2001**	**2002**	**2003**	**2004**
2	*1 квартал*	20	28	34	31
3	*2 квартал*	31	43	43	41
4	*3 квартал*	46	53	56	55
5	*4 квартал*	52	64	66	70

Рис. 1.8. Исходные данные

Для анализа временного ряда перепишем поквартальные данные в один столбец. Для этого:

- в столбце A запишем числа 1,2,…,16;
- чтобы перенести данные для временного ряда, в столбцы запишем формулы согласно таблице:

Ячейка	Формула	Примечание
B8	=B2	Копируем в диапазон B9:B11
B12	=C2	Копируем в диапазон B13:B15
B16	=D2	Копируем в диапазон B17:B19
B20	=E2	Копируем в диапазон B21:B23

2. По данным диапазона B8:B23 построим график временного ряда 1 (рис. 2.8).

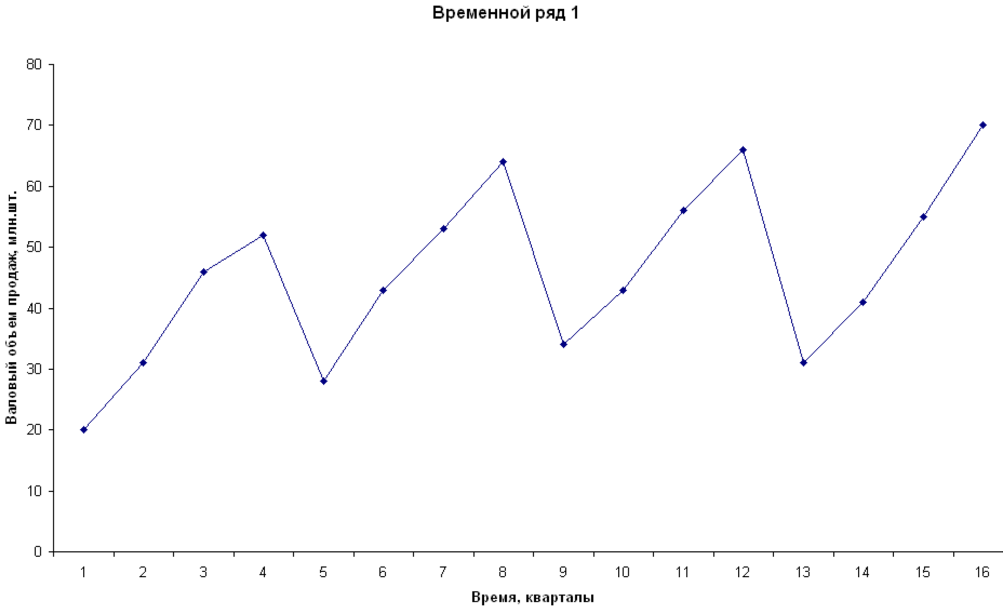


Рис. 2.8. График временного ряда 1

Как следует из графического представления временного ряда, изображен возрастающий тренд, содержащий сезонные колебания, период которых равен четырем кварталам. Объем продаж во втором полугодии (кварталы III, IV) значительно выше, чем в первом (кварталы I, II). С увеличением t не происходит заметного увеличения сезонных колебаний. Следовательно, для данного временного ряда лучше подходит аддитивная модель.

3. Произведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

а) найдем скользящие средние за 4 квартала (столбец C). С этой целью в ячейку C10 введем формулу:

=СРЗНАЧ(B8:B11),

которую затем скопируем вниз до строки C22.

б) поскольку скользящие средние за 4 квартал следует относительно середине между двумя кварталами, то их центрируют. Для этого в столбце D найдем средние значения для двух последовательных скользящих средних, с этой целью в ячейку D10 введем формулу:

=СРЗНАЧ(C10:C11),

которую затем скопируем вниз до строки D22.

4. Найдем оценки сезонной компоненты как разности между фактическими уровнями ряда и центрированными скользящими средними. Для этого в ячейку E10 введем формулу:

=B10–D10,

которую скопируем в диапазон E11:E21.

Для расчета сезонной компоненты удобно переписать полученные оценки в виде таблицы (рис. 3.8).

	A	B	C	D	E	F	G	H	I	J	K	L	M
25													
26					Расчет сезонной компоненты								
27													
28			номер квартала в году										
29			1	2	3	4							
30		2001			7,75	11,25							
31		2002	-15,13	-2,50	5,25	15,50							
32		2003	-14,88	-6,50	6,63	17,25					к=сумма/4	-0,04	
33		2004	-17,38	-7,75							Sск=Sср-к		
34	итого		-47,38	-16,75	19,63	44,00							
35	средняя оценка сезонной компоненты Sср		-15,79	-5,58	6,54	14,67	-0,17	сумма					
36	скорректированная сезонная компонента Sск		-15,75	-5,54	6,58	14,71	0,00	сумма					
37													

Рис. 3.8. Расчет сезонной компоненты

Определим средние значения оценок сезонной компоненты по каждому кварталу, для этого в ячейку C35 введем формулу:

=СРЗНАЧ(C31:C33).

которую скопируем до столбца F35.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. *В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем месяцам должна быть равна нулю*. Для данной модели имеем:

$$-15{,}79-5{,}58+6{,}54+14{,}67= -0{,}17.$$

Определим корректирующий коэффициент в ячейке L32 с помощью формулы:

=СРЗНАЧ(C35:F35).

В диапазоне C36:F36 найдем скорректированные значения сезонной компоненты как разности между соответствующими средними оценками и корректирующим коэффициентом. Для этого в ячейку C36 введем формулу:

=C35–$L$32,

которую скопируем в диапазон D36:F36.

Для проверки условия равенства нулю суммы значений сезонной компоненты в ячейке G36 введем формулу

=СУММ(C36:F36).

Значение 0, полученное в ячейке G36, указывает на то, что упомянутое условие выполняется.

Скорректированные значения сезонной компоненты перенесем в столбец F.

Для графического представления полученных результатов выделим диапазон B8:B23, содержащий уровни исходного временного ряда и (при нажатой клавише **Ctrl**) диапазон F8:F23, содержащий значения сезонной компоненты и воспользуемся **Мастером диаграмм**. Результат представлен на рис. 4.8.

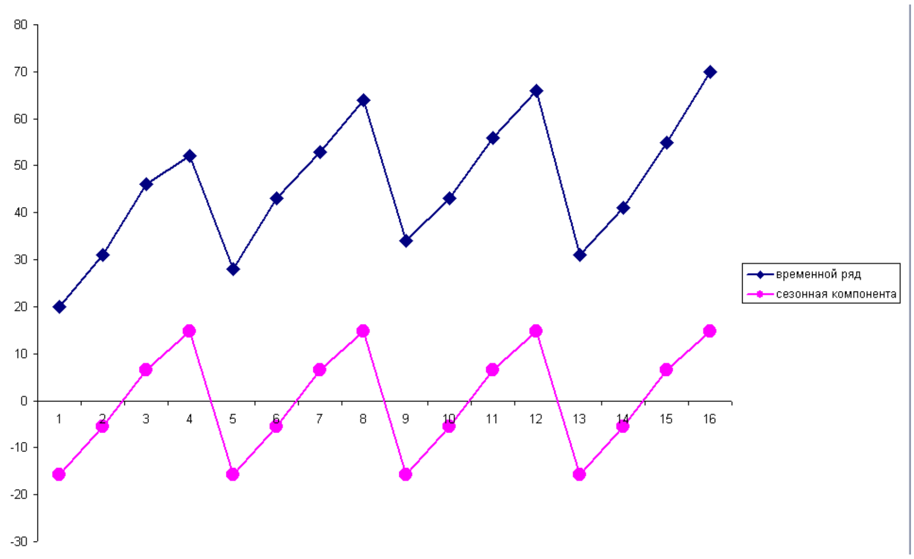


Рис. 4.8. Временной ряд и его сезонная компонента

5. Устраним влияние сезонной компоненты, вычитая ее значения из каждого уровня исходного временного ряда. Для этого в ячейке G8 запишем формулу:

=B8–F8,

которую скопируем вниз до строки 24.

6. Определим компоненту T для данной модели. Для этого произведем аналитическое выравнивание ряда (T+E) с помощью линейного тренда.

Ячейка	Формула	Примечание
К43	=ОТРЕЗОК(G8:G23;A8:A23)	Коэффициент *a*
К44	=НАКЛОН(G8:G23;A8:A23)	Коэффициент *b*

Тогда уравнение тренда будет иметь вид:

$$T=37{,}38+0{,}99t.$$

Найдем уровни тренда для каждого квартал. Для этого в ячейку H8 введем формулу:

=$K$44*A8+$K$43,

которую скопируем до строки 24.

7. В столбце I найдем значения уровней ряда, полученные по аддитивной модели. Для этого для каждого квартала найдем сумму тренда и сезонной компоненты. С этой целью в ячейку I8 запишем формулу:

=F8+H8

которую скопируем до строки 24.

8. В столбце J произведем расчет абсолютных ошибок. Для этого в ячейке J8 запишем формулу:

=B8–I8

которую скопируем до строки 24.

Относительная ошибка определяется как отношение суммы абсолютных ошибок к сумме квадратов отклонений уровней ряда от его среднего значения:

$$\delta = \frac{\sum e_t^2}{\sum (y_t - \overline{y})^2}.$$

Для того, чтобы воспользоваться этой формулой, последовательно найдем:

Ячейка	Формула	Примечание
K8	=B8–СРЗНАЧ($B$8:$B$23)	Копируем в диапазон K9:K23
K60	=СУММКВ(J8:J23)	$\sum e_t^2$
K61	=СУММКВ(K8:K23)	$\sum (y_t - \overline{y})^2$
K62	=K60/K61*100	δ

9. Прогнозное значение в аддитивной модели есть сумма трендовой и сезонной компонент. Расчет прогнозных значений в I и II кварталах пятого года выполним в строках 60-61 (рис. 5.8).

	A	B	C	D	E
59			**T**	**S**	**F**
60		17	54,25	-15,75	38,50
61		18	55,24	-5,54	49,70

Рис. 5.8. Получение прогнозных значений

Для этого в столбце B запишем очередные номера кварталов: 17 и 18, а в остальных столбцах необходимые формулы:

Ячейка	Формула	Примечание
C60	=$K$43+$K$44*B60	Копируем в C61
D60	=C36	
D61	=D36	
E60	=C60+D60	Копируем в E61

В итоге 38,5 и 49,7 – прогнозные значения валового объема продаж в I и II кварталах пятого года.

Временной ряд 2

Представлены поквартальные данные о валовом объеме продаж некоторого продукта за последние четыре года:

Квартал	Год			
	1	2	3	4
I		27	39	46
II	18	24	29	33
III	31	38	46	59
IV	45	53	72	

На листе **Анализ 2** подготовим исходные данные, как показано на рис. 6.8.

	A	B	C	D	E
1	**Анализ временного ряда**				
2					
3	**Исходные данные**				
4	**квартал**	**2001**	**2002**	**2003**	**2004**
5	*1*		27	39	46
6	*2*	18	24	29	33
7	*3*	31	38	46	59
8	*4*	45	53	72	
9					

Рис. 8.6. Исходные данные на листе **Анализ 2**.

Перепишем исходные данные в столбцы A и B, как показано на рис. 7.8.

	A	B
10	Время	Временной ряд
11	1	
12	2	18
13	3	31
14	4	45
15	5	27
16	6	24
17	7	38
18	8	53
19	9	39
20	10	29
21	11	46
22	12	72
23	13	46
24	14	33
25	15	59
26	16	
27		

Рис. 7.8. Данные для расчета мультипликативной модели

По данным диапазона B12:B25 построим график временного ряда (рис. 8.8.).

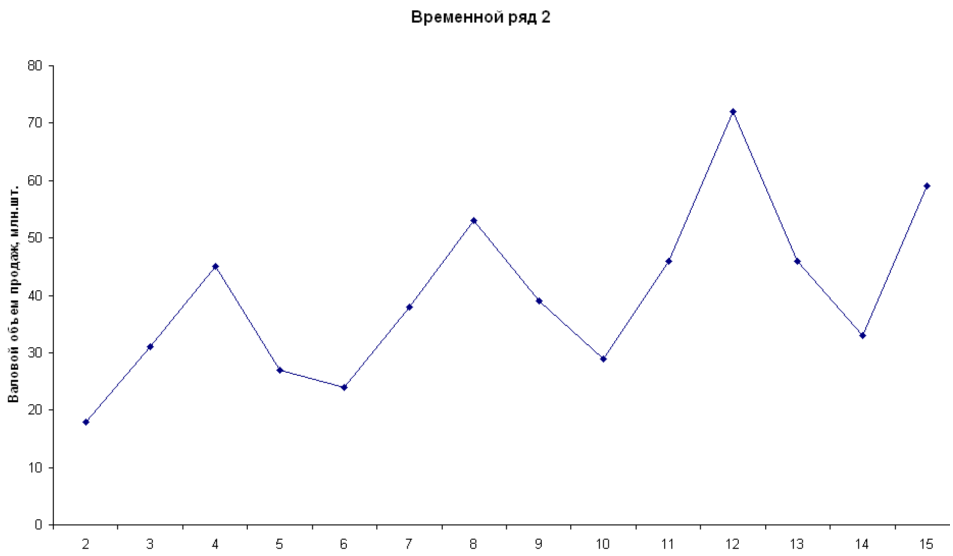


Рис. 8.8. График временного ряда 2

Как и в предыдущем примере, валовой объем продаж рассматриваемого продукта, подвержен сезонным колебаниям и значения в первом полугодии (кварталы I-II) значительно меньше, чем во втором полугодии (кварталы III-IV). Однако размах вариации фактических значений относительно линии тренда постоянно возрастает. К таким данным следует применять модель с мультипликативной компонентой.

Для выравнивания исходных уровней ряда методом скользящей средней введем формулы:

Ячейка	Формула	Примечание
C14	=СРЗНАЧ(B12:B15)	Копируем в диапазон C15:C24
D14	=СРЗНАЧ(C14:C15)	Копируем в диапазон D15:D23
E14	=B14/D14	Копируем в диапазон E15:E23

Найдем значения сезонной компоненты.

Перепишем полученные оценки в виде таблицы (рис. 9.8).

	A	B	C	D	E	F	G	H	I	J	K	L	M
28													
29			1	2	3	4							
30		2001				1,45							
31		2002	0,83	0,70	1,03	1,35					к - корректирующий коэффициент		
32		2003	0,96	0,66	0,97	1,48					к=4/сумма=	1,0001	
33		2004	0,90								Sск=Sср*к		
34	средняя оценка сезонной компоненты Sср		0,8963	0,6764	0,9990	1,4277	3,9994	сумма					
35	скорректированная сезонная компонента Sск		0,8964	0,6765	0,9991	1,4279	4,00	сумма					

Рис. 9.8. Расчет сезонной компоненты в мультипликативной модели

Определим средние значения оценок сезонной компоненты по каждому кварталу, для этого в ячейку C34 введем формулу:

=СРЗНАЧ(C30:C33).

которую скопируем до столбца F34.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. *В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов (среднее арифметическое ее значений по всем кварталам должно равняться единице)*. Для данной модели имеем в ячейке G34:

0,8963+0,6764+0,999+1,4277=3,9994.

Определим корректирующий коэффициент в ячейке L32 с помощью формулы:

=4/G34.

В диапазоне C35:F35 найдем скорректированные значения сезонной компоненты. Для этого в ячейку C35 введем формулу:

=C34*$L$32,

которую скопируем в диапазон D35:F35.

Для проверки условия равенства четырем суммы значений сезонной компоненты в ячейке G35 введем формулу

=СУММ(C35:F35).

Значение 4, полученное в ячейке G35, указывает на то, что упомянутое условие выполняется.

Скорректированные значения сезонной компоненты перенесем в столбец F.

Приведем итоговый вид формул:

Ячейка	Формула	Примечание
G12	=B12/F12	Копируем до строки 26
E37	=НАКЛОН(G12:G25;A12:A25)	Коэффициент b
F37	=ОТРЕЗОК(G12:G25;A12:A25)	Коэффициент a
H12	=$E$37*A12+$F$37	Копируем до строки 26
I12	=F12*H12	Копируем до строки 26
J12	=B12–I12	Копируем до строки 26
B40	16	
B41	17	
C40	=$F$37+$E$37*B40	Копируем вниз
D40	=F35	
D41	=C35	
E40	=C40*D40	Копируем вниз
J39	=СУММКВ(J12:J25)	$\sum e_t^2$
J40	=СУММКВ(K12:K25)	$\sum (y_t - \bar{y})^2$
J41	=J39/J40*100	δ

Результаты представлены на рис.10.8.

M24

	A	B	C	D	E	F	G	H	I	J	K	L	
10	Время	Временной ряд	Скользящая средняя	Центрированная скользящая средняя	Оценка сезонной компоненты	S	y/S	T	TS	E=y-(TS)	y-y$_{ср}$		
11	1					0,90							
12	2	18				0,68	26,61	26,40	17,86	0,14	-22,00		
13	3	31				1,00	31,03	28,63	28,60	2,40	-9,00		
14	4	45	30,25	31,00	1,45	1,43	31,51	30,85	44,05	0,95	5,00		
15	5	27	31,75	32,63	0,83	0,90	30,12	33,07	29,64	-2,64	-13,00		
16	6	24	33,50	34,50	0,70	0,68	35,47	35,29	23,88	0,12	-16,00		
17	7	38	35,50	37,00	1,03	1,00	38,03	37,51	37,48	0,52	-2,00		
18	8	53	38,50	39,13	1,35	1,43	37,12	39,74	56,74	-3,74	13,00		
19	9	39	39,75	40,75	0,96	0,90	43,51	41,96	37,61	1,39	-1,00		
20	10	29	41,75	44,13	0,66	0,68	42,87	44,18	29,89	-0,89	-11,00		
21	11	46	46,50	47,38	0,97	1,00	46,04	46,40	46,36	-0,36	6,00		
22	12	72	48,25	48,75	1,48	1,43	50,42	48,63	69,43	2,57	32,00		
23	13	46	49,25	50,88	0,90	0,90	51,32	50,85	45,58	0,42	6,00		
24	14	33	52,50			0,68	48,78	53,07	35,90	-2,90	-7,00		
25	15	59				1,00	59,05	55,29	55,24	3,76	19,00		
26	16												
27													
28													
29			1	2	3	4							
30		2001				1,45							
31		2002	0,83	0,70	1,03	1,35					к - корректирующий коэфс		
32		2003	0,96	0,66	0,97	1,48					к=4/сумма=	1,0001	
33		2004	0,90								Sск=Sср*к		
34	средняя оценка сезонной компоненты Sср		0,8963	0,6764	0,9990	1,4277	3,9994	сумма					
35	скорректированная сезонная компонента Sск		0,8964	0,6765	0,9991	1,4279	4,00	сумма					
36										*уравнение тренда*			
37				**b**	2,22	21,96	**a**		*T=*	***21,96***	**+**	***2,22***	***t***
38													

Рис. 10.8. Итоговые расчеты для мультипликативной модели

Прогнозные значения для двух очередных кварталов (IV квартал 2004 года и I квартал 2005 года) в ячейках E40 и E41. Ошибка приближения – в ячейке J41 (рис. 11.8).

M79

	A	B	C	D	E	F	G	H	I	J
39			**T**	**S**	**F**					60,0945
40		16	57,51	1,43	82,13					2816
41		17	59,74	0,90	53,55				ошибка	2,13
42										

Рис. 11.8. Прогнозные значения и ошибка прогноза

Построим график на основе данных диапазонов: B12:B25 (исходный ряд), H12:H25 (тренд), I12:I25 (приближение ряда с помощью мультипликативной модели) (рис. 12.8).

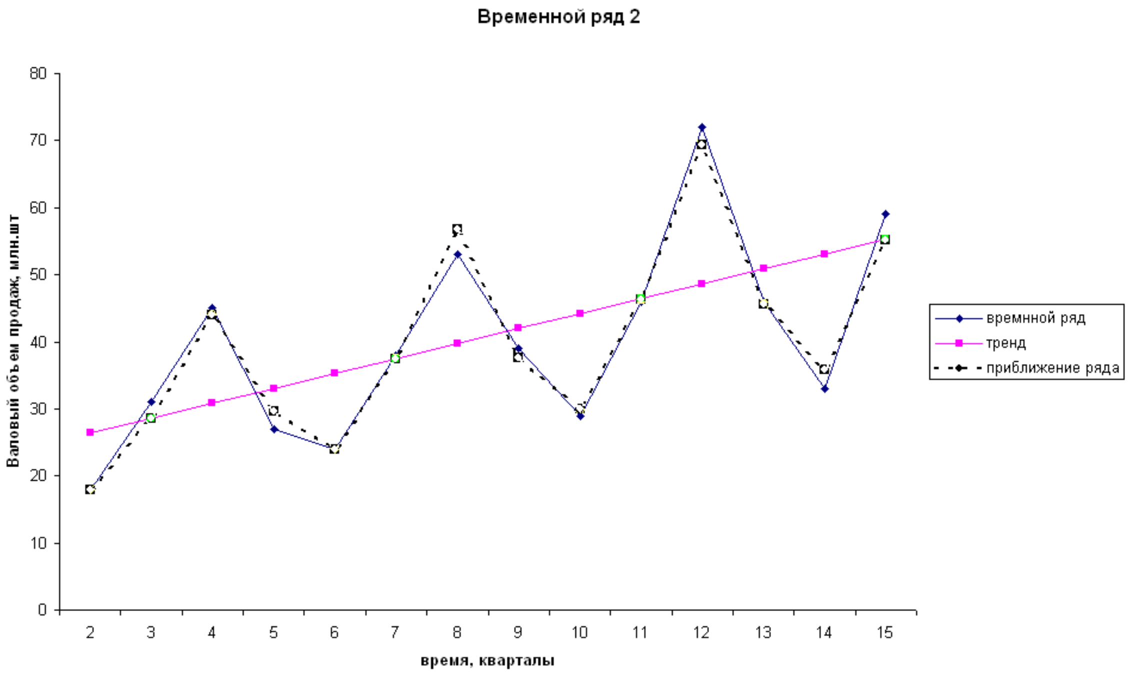


Рис. 12.8. Графики временного ряда, тренда и приближения временного ряда с помощью мультипликативной модели

Индивидуальное задание к лабораторной работе

Временной ряд 1.

Для 10 предприятий по данным внутренней отчетности получены данные о квартальных товарооборотах за 4 последних года в условных единицах.

Квартал	Номер варианта									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
I	11	12	32	77	19	32	102	316	100	92
II	14	25	34	79	22	58	108	322	115	94
III	17	28	34	78	25	64	110	320	130	93
IV	14	32	29	73	21	71	93	300	113	88
I	16	40	26	66	23	87	84	270	123	81
II	22	45	28	59	29	97	92	244	153	74
III	22	40	27	65	29	87	87	268	153	80
IV	21	40	23	62	28	88	77	254	148	77
I	23	47	21	55	31	101	69	228	160	70
II	27	50	22	55	34	108	72	226	178	70
III	28	44	22	60	36	96	72	246	185	75
IV	26	51	16	55	33	109	54	226	173	70
I	29	58	14	47	36	124	48	194	188	62
II	34	63	17	43	42	134	59	180	215	58
III	34	64	14	40	41	135	50	168	213	55
IV	32	62	11	27	40	132	41	116	205	42

Временной ряд 2.

По данным внутренней отчетности для 10 фирм получены данные об объемах заказов фирмы за 4 последних года в условных единицах.

Квартал	Номер варианта									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
I	413	1148			1213	146			581	313
II	728	1225	1086		2159	251	417		619	550
III	613	1504	913	2250	1813	213	510	313	759	463
IV	598	890	891	1329	1769	208	305	306	452	452
I	573	961	853	1435	1693	200	329	293	487	433
II	988	1022	1476	1527	2939	338	349	501	518	745
III	813	1248	1213	1866	2413	280	425	413	631	613
IV	778	733	1161	1093	2309	268	253	396	373	587
I	733	778	1093	1161	2173	253	268	373	396	553
II	1248	813	1866	1213	3719	425	280	631	413	940
III	1022	988	1527	1476	3041	349	338	518	501	770
IV	961	573	1435	853	2857	329	200	487	293	724
I	890	598	1329	891	2645	305	208	452	306	671
II	1504	613	2250	913	4487	510	213	759	313	1132
III	1225		1831	1086	3649		251	619	371	
IV			1711	598			143	579		

Отчет по лабораторной работе

Для каждого из двух заданных временных рядов:

1. Записать уравнение тренда.
2. На основе полученной модели сделать прогноз объема продаж на следующие два квартала с учетом выявленной сезонности.