

Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное
учреждение
высшего образования
**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ПРОМЫШЛЕННЫХ ТЕХНОЛОГИЙ И ДИЗАЙНА»**

Кафедра информационных технологий

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

Методические указания к выполнению контрольной работы

для направления подготовки 38.03.06 – Торговля
непродовольственными товарами (РИНПО)

Составитель Т. А. Кравец

Санкт-Петербург

2023

УТВЕРЖДЕНО
на заседании кафедры

Рецензент

Учебное электронное издание сетевого распространения
Издано в авторской редакции

Системные требования:
электронное устройство с программным обеспечением для
воспроизведения файлов формата PDF

Режим доступа: http://publish.sutd.ru/tp_get_file.php?id= по паролю. – Загл.
с экрана.

Дата подписания к использованию

ФГБОУВО «СПбГУПТД»
Юридический и почтовый адрес: 191186, Санкт-Петербург, ул. Большая
Морская, 18.
<http://sutd.ru>

ВВЕДЕНИЕ

Контрольная работа по дисциплине «Информационные технологии в профессиональной деятельности» выполняется студентами направления 38.03.06 – «Торговля непродовольственными товарами» по результатам изученного материала, представленного в настоящих учебно-методических указаниях.

Материал методических указаний построен на основании полученных студентами знаний и навыков, сформированных в предыдущих семестрах, при изучении информационных технологий и математических дисциплин. Учебный план курса предусматривает выполнение контрольной работы и сдачу зачета.

В методических указаниях содержатся следующие материалы:

1. Аппроксимация экспериментальных данных линейной функцией.
2. Аппроксимация экспериментальных данных нелинейными функциями.
3. Требования к выполнению контрольной работы. Выбор варианта задания.
4. Список вопросов к зачету.
5. Список литературы для подготовки к зачету.

Для выполнения контрольной работы необходимо изучить теоретический материал, приведенный в методических указаниях, разобрать примеры выполнения расчетов, выполнить задание по индивидуальному варианту и оформить работу в соответствии с требованиями к контрольной работе.

Выполнение контрольной работы, изучение списка литературных источников в соответствии с предложенными вопросами к зачету позволит успешно освоить теоретический курс по изучаемой дисциплине.

1. Аппроксимация экспериментальных данных линейной функцией

1.1. Основные понятия и определения

Одной из часто встречающихся задач при обработке результатов эксперимента является подбор формулы для установления функциональной зависимости между экспериментальными данными, так называемого *уравнения регрессии*. Прежде чем приступить к подбору формулы, целесообразно нанести опытные данные на график и на глаз, от руки провести через полученные точки наиболее правдоподобную кривую. Очень часто общий вид кривой бывает известен из других соображений, что упрощает задачу, сводя ее к поиску числовых коэффициентов известной функциональной зависимости общего вида (например, прямолинейной, квадратичной, логарифмической и др.). При этом часто становятся видны те опытные данные, которые скорее всего содержат наибольшие погрешности. Кроме полученных экспериментальных точек существенным моментом при проведении кривой являются соображения общего характера: как ведет себя кривая вблизи нуля, пересекает ли она координатные оси, касается ли их, имеет ли асимптоты и т.д.

После того, как эта предварительная работа проделана, начинается собственно подбор формулы – уравнения регрессии. Решением задач, связанных с поиском уравнений регрессии занимается *регрессионный анализ*, а одним из его наиболее широко применяемых на практике алгоритмов является *метод наименьших квадратов*

В общем случае задача метода наименьших квадратов формулируется следующим образом. Пусть искомая функциональная зависимость величины y от величины x выражается формулой:

$$y = a_0 f_0(x) + a_1 f_1(x) + \dots + a_m f_m(x) \quad (1.1)$$

где $f_0(x), f_1(x), \dots, f_m(x)$ – заданные функции, $a_0, a_1, \dots, a_m$ – искомые параметры – коэффициенты уравнения (1.1).

Для получения оценок параметров $a_0, a_1, \dots, a_m$ проводят n экспериментов, результаты которых дают значения не этих параметров, а некоторой функции (1), зависящей от них линейно.

Метод наименьших квадратов состоит в том, что *оценки параметров* $a_0, a_1, \dots, a_m$ формулы (1) *определяются из условия: сумма квадратов отклонений достигает наименьшего значения.*

$$\Phi(a_0, a_1, \dots, a_m) = \sum_{i=1}^n [f(x_i, a_0, a_1, \dots, a_m) - y_i]^2 \rightarrow \min \quad (1.2)$$

$$\left\{ \begin{array}{l} \frac{\partial \Phi(a_0, a_1)}{\partial a_0} = 2 \sum_{i=1}^n [a_0 + a_1 x_i - y_i] \frac{\partial \phi}{\partial a_0} = 0 \\ \frac{\partial \Phi(a_0, a_1)}{\partial a_1} = 2 \sum_{i=1}^n [a_0 + a_1 x_i - y_i] x_i \frac{\partial \phi}{\partial a_1} = 0 \end{array} \right.$$

Проведя преобразования, система уравнений будет записана как (1.4):

$$\left\{ \begin{array}{l} a_0 \frac{n}{n} + a_1 \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i}{n} \\ a_0 \frac{\sum_{i=1}^n x_i}{n} + a_1 \frac{\sum_{i=1}^n x_i^2}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} \end{array} \right. \quad (1.4)$$

Решив систему с двумя неизвестными a_0, a_1 получим:

$$\begin{aligned} a_0 &= \frac{\sum_{i=1}^n y_i}{n} - a_1 \frac{\sum_{i=1}^n x_i}{n} \\ a_1 &= \frac{n(\sum_{i=1}^n x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2} \end{aligned} \quad (1.5)$$

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии (коэффициента a_1).

Если знак при коэффициенте регрессии - положительный, связь зависимой переменной с независимой будет положительной. В нашем случае знак коэффициента регрессии положительный, следовательно, связь также является положительной.

Если знак при коэффициенте регрессии - отрицательный, связь зависимой переменной с независимой является отрицательной (обратной).

Для анализа общего качества уравнения регрессии используют обычно множественный коэффициент детерминации R^2 , называемый также квадратом коэффициента множественной корреляции R . R^2 (мера определенности) всегда находится в пределах интервала $[0;1]$.

Если значение R^2 близко к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных. И

наоборот, значение R -квадрата, близкое к нулю, означает плохое качество построенной модели.

Коэффициент детерминации R^2 показывает, на сколько процентов ($R^2 \cdot 100\%$) найденная функция регрессии описывает связь между исходными значениями факторов x и y

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1.6)$$

где $(y_i^p - \bar{y})^2$ – объясненная вариация;

$(y_i - \bar{y})^2$ – общая вариация.

Соответственно, величина $(1 - R^2) \cdot 100\%$ показывает, сколько процентов вариации параметра y обусловлены факторами, не включенными в регрессионную модель. При высоком ($R^2 \geq 100\%$) значении коэффициента детерминации можно делать прогноз $y^* = f(x^*)$ для конкретного значения x^* .

Задача:

Фирма занимается поставками различных товаров на короткие расстояния внутри города. Оценить стоимость таких услуг, зависящую от затрачиваемого на поставку времени. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Были собраны исходные данные о десяти поставках (табл. 1.1)

Таблица 1.1 – Исходные данные для решения задачи

Переменные	№ опыта									
	1	2	3	4	5	6	7	8	9	10
Расстояние (X), км	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Время (Y), мин.	16	13	19	18	12	11	8	14	9	16

Найти:

1. Определить характер зависимости между расстоянием и затраченным временем, путем визуальной оценки распределения точек эксперимента на графике (использовать инструмент построения диаграмм *MS Excel*).
2. Провести расчеты и постройте уравнение регрессии, используя МНК.
3. Проанализировать силу регрессионной связи и сделать выводы.
4. Сделать прогноз времени поездки на 2 км. Посчитать и построить графически меру ошибки регрессионной модели используя табличный процессор *MS Excel*.
5. Провести регрессионный анализ, используя *MS Excel* и сравнить с результатами, полученными ранее.

1.3. Решение задачи линейной регрессии «вручную»

Покажем на диаграмме полученные результаты эксперимента (рис. 1.1) и определим, какая функция может быть выбрана в качестве линии регрессии. Поскольку помимо расстояния на время поставки влияют пробки на дорогах, время суток, дорожные работы, погода, квалификация водителя, вид транспорта, то построенные точки не находятся точно на линии. Это обусловлено описанными выше факторами. Но, на диаграмме хорошо видно, что точки собраны вокруг прямой линии, поэтому можно предположить линейную связь между параметрами. Все исходные точки равномерно распределены вдоль предполагаемой прямой линии, что позволяет применить метод наименьших квадратов.

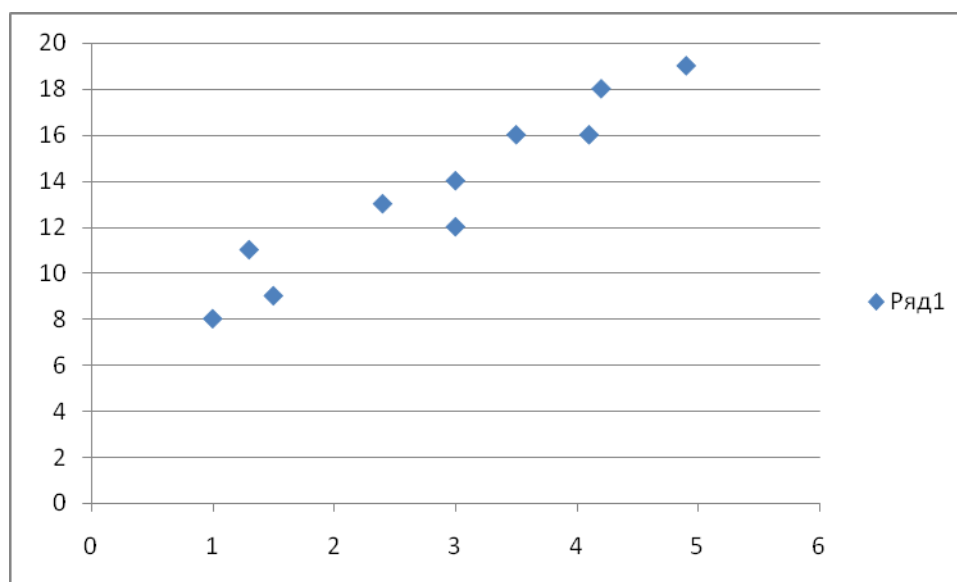


Рис.1.1 – Графическое представление исходных данных

Вычислим суммы, необходимые для расчета коэффициентов уравнения линейной регрессии и коэффициента детерминации R^2 с помощью вспомогательной таблицы. Сначала найдем значения x_i^2 , $x_i y_i$ и их суммы, результаты внесем в *табл. 1.2*.

Рассчитаем значения коэффициентов a_0 и a_1 по выведенным ранее формулам (1.5).

$$a_1 = \frac{n(\sum_{i=1}^n x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2} = \frac{10 \cdot 435,30 - 28,9 \cdot 136}{10 \cdot 99,41 - (28,9)^2} = 2,660$$

$$a_0 = \frac{\sum_{i=1}^n y_i}{n} - a_1 \frac{\sum_{i=1}^n x_i}{n} = \frac{136}{10} - 2,66 \cdot \frac{28,9}{10} = 5,913$$

Таким образом, искомая регрессионная зависимость будет иметь вид:

$$y^p = 5,913 + 2,660 \cdot x$$

Таблица 1.2 – Вспомогательные вычисления для определения уравнения регрессии

№	x_i	y_i	x_i^2	$x_i y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	12,25	56,00	15,223	2,634129	5,76
2	2,4	13	5,76	31,2	12,297	1,697809	0,36
3	4,9	19	24,01	93,1	18,947	28,59041	29,16
4	4,2	18	17,64	75,60	17,085	12,14523	19,36
5	3,0	12	9,00	36,00	13,893	0,085849	2,56
6	1,3	11	1,69	14,30	9,371	17,88444	6,76
7	1,0	8	1,00	8,00	8,573	25,27073	31,36
8	3,0	14	9,00	42,00	13,893	0,085849	0,16
9	1,5	9	2,25	13,50	9,903	13,66781	21,16
10	4,1	16	16,81	65,60	16,819	10,36196	5,76
Сумма:	28,9	136	99,41	435,30	—	112,4242	122,4

Рассчитаем значения y_i^p при каждом значении x_i , определим среднее значение $\bar{y}$ и вычислим $(y_i^p - \bar{y})^2$, $(y_i - \bar{y})^2$. Результаты расчета внесем в *табл. 1.2*.

Вычислим коэффициент детерминации для линейной регрессии (1.6):

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{112,424}{122,400} \cdot 100\% = 0,918 \cdot 100\% = 91,8\%$$

Поскольку полученное значение R^2 близко к единице (0,918), то построенная модель имеет хорошее качество и на 91,8% описывает влияние x на y .

Решим задачу прогнозирования времени поездки на 2 км ($x^* = 2 \text{ км}$). Поскольку коэффициент детерминации R^2 имеет достаточно высокое значение и расстояние 2 км, для которого надо сделать прогноз, находится в пределах диапазона исходных данных (табл. 1.1), то мы можем использовать полученное уравнение линейной регрессии для прогнозирования.

$$y^*(x^*) = 5,913 + 2,660 \cdot 2 = 11,2 \text{ мин.}$$

При прогнозах на расстояния, не входящие в диапазон исходных данных, нельзя гарантировать справедливость полученной модели. Это объясняется тем, что связь между временем и расстоянием может изменяться по мере увеличения расстояния. На время дальних перевозок могут влиять новые факторы такие, как использование скоростных шоссе, остановки на отдых, обед и т.п.

1.4. Решение задачи с использованием пакета *MS Excel*

Проведем регрессионный анализ с использованием режима **Регрессия** *MS Excel*. Значения параметров, установленных в одноименном диалоговом окне, представлены на *рис. 1.2*, результаты вычислений показаны на *рис. 1.3*.

Величина R -квадрат, называемая также мерой определенности, характеризует качество полученной регрессионной прямой. Это качество выражается степенью соответствия между исходными данными и регрессионной моделью (расчетными данными). Мера определенности всегда находится в пределах интервала $[0;1]$. В нашем примере мера определенности равна 0,91829, что говорит о хорошем описании исходных данных регрессионной прямой. Данный коэффициент детерминации совпадает с R^2 , вычисленным по формуле.

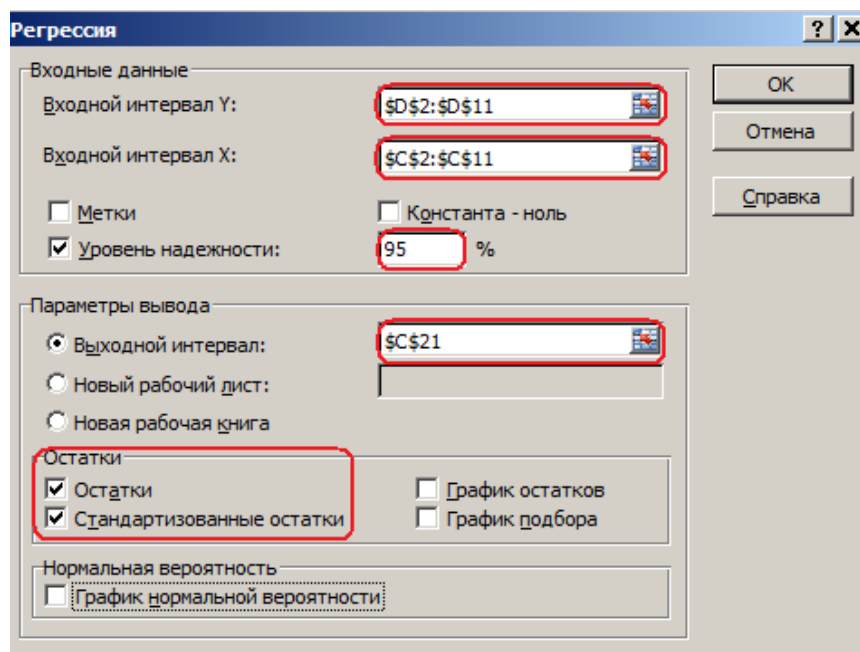


Рис. 1.2. Настройки окна «Регрессия»

Таким образом, линейная модель объясняет 91,8% вариации времени доставки, что означает правильность выбора фактора (расстояния). Не объясняется $100\% - 91,8\% = 8,2\%$ вариации времени поездки, которые обусловлены остальными факторами, влияющими на время поставки, но не включенными в линейную модель регрессии.

Рассчитанный уровень значимости $\alpha_p = 1,26E - 05 < 0,05$ (показатель значимость F в таблице Дисперсионный анализ) подтверждает значимость R^2 .

Множественный R - коэффициент множественной корреляции R - выражает степень зависимости независимых переменных (x) и зависимой переменной (y) и равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы. В простом линейном регрессионном анализе множественный R равен коэффициенту корреляции Пирсона. Действительно, множественный R в нашем случае равен коэффициенту корреляции Пирсона (0,95827), который вычисляется по формуле:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}$$

Регрессионная статистика

Множественный R	0,958275757
R-квадрат	0,918292427
Нормированный R-квадрат	0,90807898
Стандартная ошибка	1,11809028
Наблюдения	10

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	112,398993	112,398993	89,91014059	1,26072E-05
Остаток	8	10,00100699	1,250125873		
Итого	9	122,4			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	5,913462144	0,884389599	6,686489927	0,00015485
Переменная X 1	2,65970168	0,280497238	9,482095791	1,26072E-05

ВЫВОД ОСТАТКА

<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
		-	
5	13,89256718	1,892567185	-1,795356486
6	9,371074328	1,628925672	1,545256778
		-	
7	8,573163824	0,573163824	-0,543723571
8	13,89256718	0,107432815	0,101914586
		-	
9	9,903014664	0,903014664	-0,8566318
		-	
10	16,81823903	0,818239033	-0,776210624

Рис. 1.3. Результаты вычислений

В нижней части таблицы регрессионного анализа отображены коэффициент регрессии a_1 (2,65970168) и смещения по оси ординат, т.е. константа a_0 (5,913462144).

Полученные в ходе расчетов коэффициенты позволяют записать уравнение регрессии:

$$y^p = 5,91346 + 2,65970 \cdot x$$

Видно, что это уравнение, совпадает с уравнением, полученным при расчете по МНК «вручную» с точностью до ошибки округления.

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии a_1 . В нашем случае знак коэффициента регрессии положительный, следовательно, связь также является положительной.

Проверим значимость коэффициентов регрессии: a_0 и a_1 . Сравнивая попарно значения столбцов «Коэффициенты» и «Стандартная ошибка» на *рис. 3*, видно, что абсолютные значения коэффициентов больше чем их стандартные ошибки. К тому же эти коэффициенты являются значимыми, о чем можно судить по значениям показателя P -значение, которые меньше заданного уровня значимости $\alpha = 0,05$.

Отклонения каждой точки исходных данных от построенной линии регрессии содержатся в колонке отчета «Остатки». Наибольшее абсолютное значение остатка в нашем случае - 1,89256, наименьшее - 0,05399. Для лучшей интерпретации этих данных воспользуемся графиком исходных данных и построенной линией регрессии, представленными на *рис. 1.4*. Как видим, линия регрессии хорошо "подогнана" под значения исходных данных.

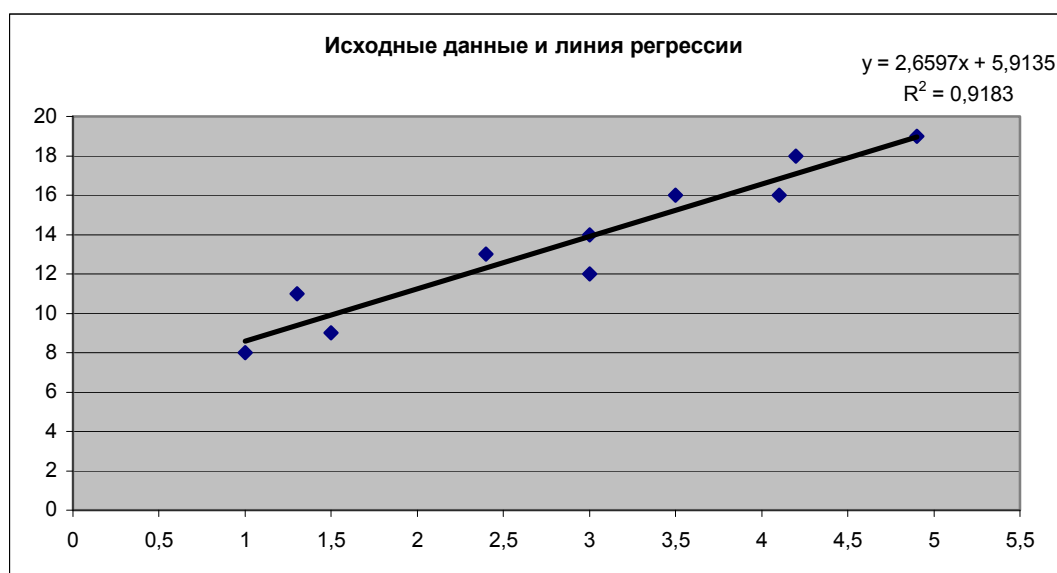


Рис. 1.4 – Диаграмма исходных данных и найденной линии регрессии

Приблизительным, но самым простым и наглядным способом проверки удовлетворительности регрессионной модели является графическое представление отклонений (рис. 1.5)

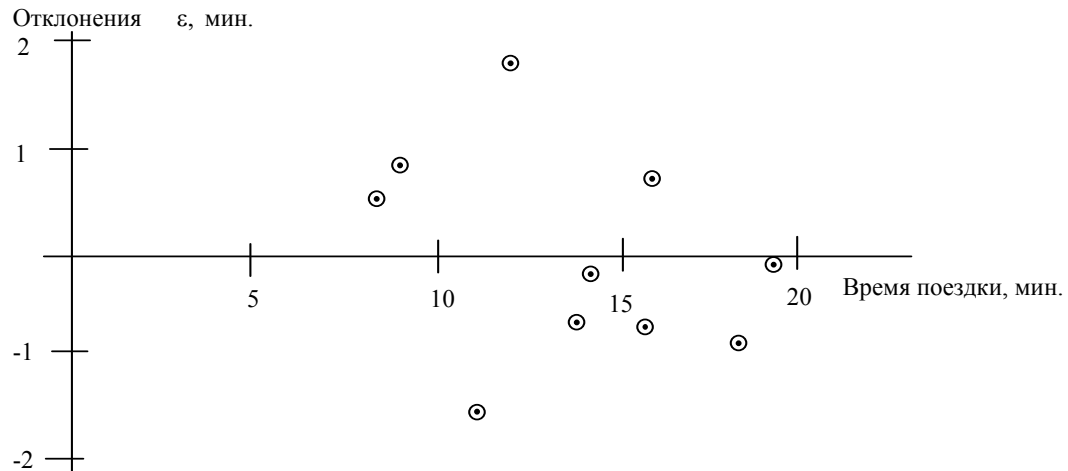


Рис. 1.5. График отклонений экспериментального значения от расчетного

Отложим отклонения $(y_i^p - y_i)$ по оси y , для каждого значения y_i (рис. 1.5). Если регрессионная модель близка к реальной зависимости, то отклонения будут носить случайный характер и их сумма будет близка к нулю. В рассмотренном примере $\sum_{i=1}^n (y_i^p - y_i) = 0,004$.

Обычно мерой ошибки регрессионной модели служит среднее квадратическое отклонение:

$$\sigma_{\varepsilon} = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{(n-2)}} = \sqrt{\frac{\sum_{i=1}^n (y_i^p - y_i)^2}{(n-2)}}$$

Для нормально распределенных процессов приблизительно 67% точек находится в пределах одного отклонения σ_{ε} от линии регрессии и 95% - в пределах $2\sigma_{\varepsilon}$ (на рис. 1.6 ближе к линии регрессии указано отклонение σ_{ε} и дальше – отклонение $2\sigma_{\varepsilon}$).

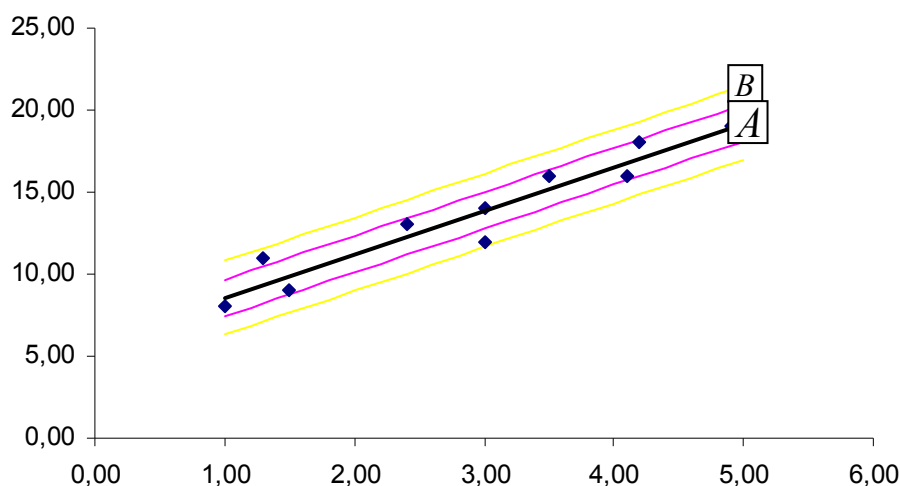


Рис.1.6. Отклонение σ_{ε} и $2\sigma_{\varepsilon}$

1.5. Порядок выполнения работы

1. Запустить пакет *MS Excel*, на вкладке **Данные** в группе **Анализ** нажать кнопку **Анализ данных** (рис. 1.7). Если кнопка **Анализ данных** недоступна, необходимо загрузить надстройку «Пакет анализа». Для этого необходимо:

- щелкнуть значок **Кнопка Microsoft Office** , а затем **Параметры Excel**;
- перейти на вкладку **Надстройки**, а затем в поле **Управление** выбрать **Надстройки Excel**;
- нажать кнопку **Перейти** и в поле **Доступные надстройки** установить флажок **Пакет анализа**, затем нажать кнопку **ОК**.

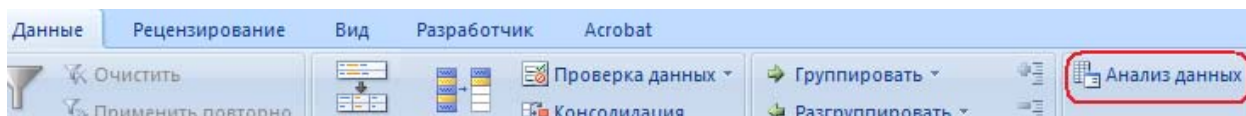


Рис. 1.7 – Панель анализа данных

2. На листе в *MS Excel* создать макет табл. 1.2 «Вспомогательные вычисления» и ввести в нее исходные данные в соответствии с индивидуальным заданием.

3. Провести вычисления x_i^2 , $x_i y_i$, рассчитать суммы для x_i , y_i , x_i^2 , $x_i y_i$ и внести эти данные в *табл. 1.2*.
4. По формуле (1.5) рассчитать коэффициенты a_1 , a_0 и построить уравнение линейной регрессии $y^P = a_0 + a_1 x$, подставив найденные a_1 , a_0 .
5. Вычислить среднее значение $\bar{y}$, рассчитать значения y_i^P , $(y_i^P - \bar{y})^2$, $(y_i - \bar{y})^2$ и заполнить *табл. 1.2* найденными значениями.
6. Вычислить коэффициент детерминации по формуле (1.6). Сделать вывод о построенной модели регрессии.
7. Исследовать исходные данные с помощью режима **Регрессия** в *MS Excel* и сравнить с результатами, рассчитанными «вручную».
8. Сделать прогноз расчетного значения функции y^* для x^* , заданного в индивидуальном задании.

2. Аппроксимация экспериментальных данных нелинейными функциями

2.1. Виды функций нелинейной регрессии

В ряде случаев аппроксимация линейной функцией может иметь неудовлетворительное решение, тогда необходимо использовать нелинейные функции регрессии.

Рассмотрим наиболее простые случаи нелинейной регрессии: экспоненту, степенную функцию и параболу. При нахождении коэффициентов экспоненты и степенной функции используют прием приведения нелинейной регрессионной зависимости к линейному виду. Это позволяет для нахождения коэффициентов аппроксимирующей функции перейти к системе линейных уравнений (1.4) и, решив ее, получить уравнения регрессии для нелинейных функций.

Экспонента

Для приведения к линейному виду уравнения экспоненты $y = a_0 e^{a_1 x}$ проведем логарифмирование:

$$\ln y = \ln(a_0 e^{a_1 x}) = \ln a_0 + \ln(e^{a_1 x}) = \ln a_0 + a_1 x$$

Введем переменные $b_0 = \ln a_0$ и $b_1 = a_1$, тогда получим линейное уравнение $\ln y = b_0 + b_1 x$.

Тогда система уравнений (1.4) будет иметь вид (2.1).

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n \ln y_i \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \ln y_i \end{cases} \quad (2.1)$$

От численных значений коэффициентов b_0 и b_1 , найденных при решении системы, необходимо перейти к коэффициентам a_0 и a_1 . Исходя из введенных обозначений и определения логарифма, получаем.

$$a_0 = e^{b_0}, \quad a_1 = b_1$$

Степенная функция

Для нахождения коэффициентов уравнения степенной функции $y = a_0 \cdot x^{a_1}$ необходимо также провести логарифмирование.

$$\ln y = \ln(a_0 x^{a_1}) = \ln a_0 + \ln(e^{a_1 x}) = \ln a_0 + a_1 \ln x$$

Введем переменные $b_0 = \ln a_0$, $b_1 = a_1$ тогда $\ln y = b_0 + b_1 \cdot \ln x$, откуда следует, что можно перейти к системе уравнений (1.4), в которой вместо значений x_i , y_i необходимо использовать $\ln x_i$, $\ln y_i$ соответственно.

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^n \ln x_i = \sum_{i=1}^n \ln y_i \\ b_0 \sum_{i=1}^n \ln x_i + b_1 \sum_{i=1}^n (\ln x_i)^2 = \sum_{i=1}^n \ln x_i \ln y_i \end{cases} \quad (2.2)$$

Полученные коэффициенты b_0 и b_1 приводим к коэффициентам a_0 и a_1 . Исходя из введенных обозначений и определения логарифма, получаем.

$$a_0 = e^{b_0}, \quad a_1 = b_1$$

Парабола

Для нахождения коэффициентов уравнения параболы $y = a_0 + a_1 x + a_2 x^2$ необходимо записать выражение (2.3) и устремить его к минимуму.

$$\Phi(a_0, a_1, a_2) = \sum_{i=1}^n [a_0 + a_1 x_i + a_2 x_i^2 - y_i]^2 \rightarrow \min \quad (2.3)$$

Минимум функции можно получить, взяв частные производные по a_0 , a_1 , a_2 и приравняв их к нулю:

$$\begin{cases} \frac{\partial \Phi(a_0, a_1, a_2)}{\partial a_0} = 2 \sum_{i=1}^n [a_0 + a_1 x_i + a_2 x_i^2 - y_i] \frac{\partial \phi}{\partial a_0} = 0 \\ \frac{\partial \Phi(a_0, a_1, a_2)}{\partial a_1} = 2 \sum_{i=1}^n [a_0 + a_1 x_i + a_2 x_i^2 - y_i] x_i \frac{\partial \phi}{\partial a_1} = 0 \\ \frac{\partial \Phi(a_0, a_1, a_2)}{\partial a_2} = 2 \sum_{i=1}^n [a_0 + a_1 x_i + a_2 x_i^2 - y_i] x_i^2 \frac{\partial \phi}{\partial a_2} = 0 \end{cases}$$

Проведя преобразования, получим систему уравнений (2.4), решив которую найдем коэффициенты регрессии a_0, a_1, a_2 :

$$\left\{ \begin{array}{l} a_0 n + a_1 \left(\sum_{i=1}^n x_i \right) + a_2 \left(\sum_{i=1}^n x_i^2 \right) = \sum_{i=1}^n y_i \\ a_0 \left(\sum_{i=1}^n x_i \right) + a_1 \left(\sum_{i=1}^n x_i^2 \right) + a_2 \left(\sum_{i=1}^n x_i^3 \right) = \sum_{i=1}^n x_i y_i \\ a_0 \left(\sum_{i=1}^n x_i^2 \right) + a_1 \left(\sum_{i=1}^n x_i^3 \right) + a_2 \left(\sum_{i=1}^n x_i^4 \right) = \sum_{i=1}^n x_i^2 y_i \end{array} \right. \quad (2.4)$$

Сила регрессионной связи R^2 для параболы определяется по той же формуле что и для линейной модели (1.6).

При вычислении коэффициента детерминации для экспоненты и степенной функции все значения параметра y (исходные, регрессионные, среднее) необходимо заменить на их логарифмы. Например, для экспоненциальной функции при расчете R^2 необходимо использовать формулу (1.6), подставляя значения по следующим зависимостям:

$$y_i^P = b_0 + b_1 \cdot x_i;$$

$$\bar{y}_i = \frac{\sum_{i=1}^n \ln y_i}{n},$$

$$y_i = \ln y_i$$

Если функция регрессии определена, интерпретирована и обоснована, и оценка точности регрессионного анализа соответствует требованиям, можно считать, что построенная модель и прогнозные значения обладают достаточной надежностью.

Задача:

Используем исходные данные (табл. 1.1) для подбора нелинейных функций регрессии: экспоненты, степенной функции и параболы.

Найти:

1. Провести расчеты «вручную» и построить уравнения регрессии для экспоненты, степенной функции и параболы методом наименьших квадратов. Исследования каждой функции проводить на отдельном листе. Лист

предварительно переименовать в соответствии с проводимым расчетом функции регрессии.

2. Проанализировать силу регрессионной связи для каждой функции.

3. Провести регрессионный анализ для экспоненты, степенной функции, используя табличный процессор *MS Excel*.

4. На каждом листе исследуемых функций построить точечные диаграммы исходных данных задачи. Добавить линии тренда и вывести их параметры (уравнение регрессии и коэффициент детерминации R^2) на диаграммы в соответствии с исследуемой функцией. Сравнить правильность результатов, вычисленных «вручную» с параметрами линии тренда на диаграмме.

5. Среди функций: линейной, экспоненты, степенной функции и параболы выбрать наиболее точно описывающую данные эксперимента.

2.2. Решение задачи «вручную»

Экспонента

Вычислим суммы, необходимые для расчета коэффициентов b_0, b_1 уравнения регрессии $\ln y = b_0 + b_1 x$ и коэффициента детерминации R^2 с помощью вспомогательной таблицы (табл. 2.1).

Таблица 2.1 – Вспомогательные вычисления для определения уравнения регрессии

№	x_i	y_i	x_i^2	$\ln y_i$	$x_i \cdot \ln y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	12,25	2,7726	9,7041	2,6995	0,0157	0,0393
2	2,4	13	5,76	2,5649	6,1559	2,4738	0,0101	0,0001
3	4,9	19	24,01	2,9444	14,4278	2,9868	0,1701	0,1370
4	4,2	18	17,64	2,8904	12,1396	2,8432	0,0723	0,0999
5	3	12	9	2,4849	7,4547	2,5969	0,0005	0,0080
6	1,3	11	1,69	2,3979	3,1173	2,2480	0,1065	0,0311
7	1	8	1	2,0794	2,0794	2,1865	0,1504	0,2449
8	3	14	9	2,6391	7,9172	2,5969	0,0005	0,0042
9	1,5	9	2,25	2,1972	3,2958	2,2891	0,0814	0,1422
10	4,1	16	16,81	2,7726	11,3676	2,8227	0,0617	0,0393
Σ	28,9	136	99,41	25,7435	77,6593	25,7434	0,6692	0,7460
Среднее:	2,89	13,6	9,941	2,5743	7,7659	2,6995	0,0157	0,0393

Подставив результаты расчетов в систему уравнений (2.1) получим:

$$\begin{cases} 10 \cdot b_0 + 28,9 \cdot b_1 = 25,7435 \\ 28,9 \cdot b_0 + 99,41 \cdot b_1 = 77,6593 \end{cases}$$

Для решения системы нормальных уравнений используем надстройку «Поиск решения» пакета *MS Excel* (**Данные-Поиск решения**). Составим исходную табличную модель (рис. 2.1).

В блок «Переменные» в первую строку записываем буквенные обозначения искомых коэффициентов системы алгебраических уравнений b_0, b_1 . Во вторую строку записываем произвольные числовые значения (удобнее в качестве числовых значений поставить номера коэффициентов, т. е. 0 и 1), в дальнейшем, после выполнения команды **Поиск решения**, в этих ячейках табличный процессор подберет значения искомых коэффициентов b_0, b_1 .

В блок «Матрица исходной системы» записываем числовые значения системы линейных уравнений при коэффициентах b_0, b_1

Переменные	
b0	b1
0	1

Матрица исходной системы	
10	28,9
28,9	99,41

Значения левых частей уравнений
28,9
99,41

=
=

Свободные члены исходной системы
25,7434
77,6592

Рис. 2.1. Табличная модель решения системы линейных уравнений средствами Excel

«Значения левых частей уравнений» в верхнюю ячейку вводим формулу: =СУММПРОИЗВ(«диапазон числовых значений переменных b_0, b_1 (0 и 1)»; «диапазон первой строки матрицы исходной системы (10 и 28,9)»). Автоматически заполняем весь столбец «Значения левых частей уравнений».

В блок «Свободные члены исходной системы» в столбец записываем значения правой части исходной системы и вызываем форму «Поиск решения» *рис. 2.2.*

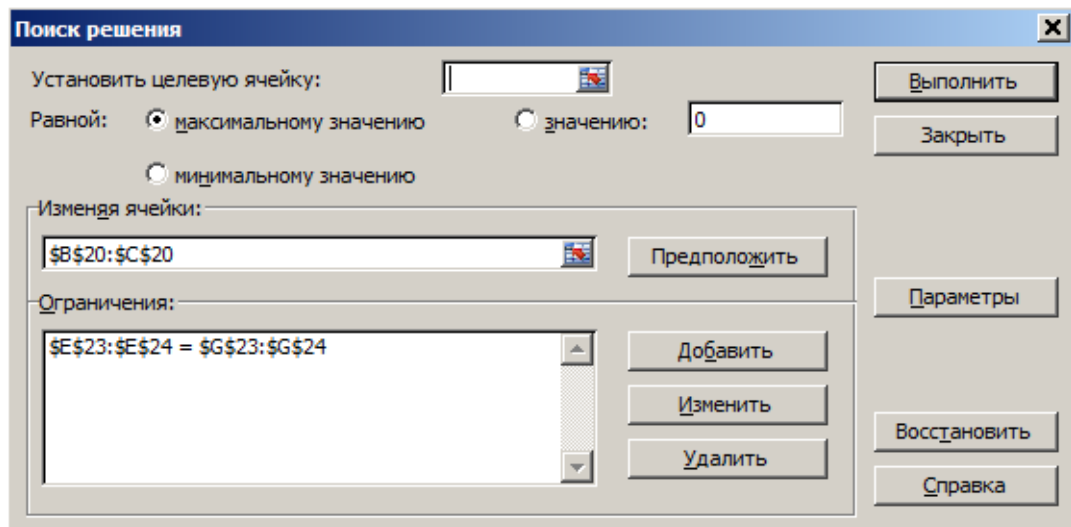


Рис. 2.2. Окно «Поиск решения»

Необходимо заполнить форму «Поиск решения» следующими данными:

- установить целевую ячейку – ничего не ставить;
- равной – максимальному значению;
- изменяя ячейки – диапазон строки значений переменных (в нашем случае выделить значения 0 и 1);
- ограничения – диапазон «Значения левых частей уравнений» = диапазон «Свободные члены исходной системы»;
- нажать «Выполнить»;
- в форме «Результаты поиска решения» (*рис. 2.3*) выбрать «Сохранить найденное решение»;
- нажать **ОК**.

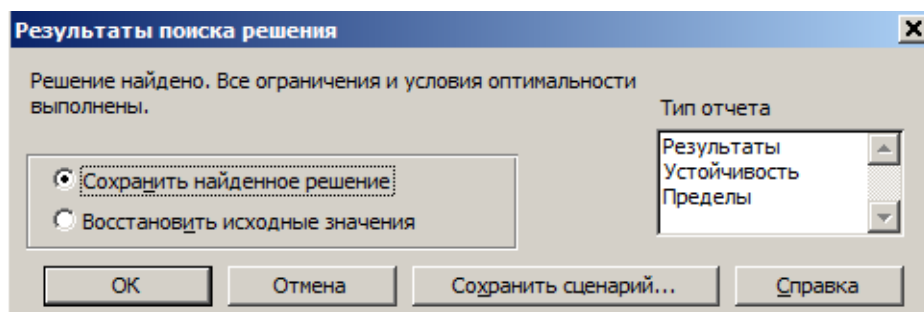


Рис. 2.3. Окно «Результаты поиска решения»

Результат выполнения команды «Поиск решения» для коэффициентов b_0, b_1 представлен на рис. 2.4.

Переменные	
b0	b1
1,981248	0,205222

Матрица коэффициентов исходной системы	
10	28,9
28,9	99,41

Значения левых частей уравнений
25,74339973
77,659199

Свободные члены исходной системы
25,7434
77,6592

Рис. 2.4. Результаты выполнения команды «Поиск решения» для экспоненциального распределения

Для определения экспоненциального распределения необходимо вычислить коэффициенты $a_0 = e^{b_0} = 7,2518$ и $a_1 = b_1 = 0,2052$.

Тогда уравнение регрессии примет вид $y = a_0 e^{a_1 x} = 7,2518 e^{0,2052 x}$

Определим коэффициент детерминации R^2 . Для этого необходимо рассчитать и заполнить пустые столбцы табл. 2.1 данными. При расчетах

необходимо использовать формулы: $y_i^p = b_0 + b_1 \cdot x_i$, $\bar{y}_i = \frac{\sum_{i=1}^n \ln y_i}{n}$, $y_i = \ln y_i$, тогда

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{0,6692}{0,7460} \cdot 100\% = 89,70 \%$$

Степенная функция

Вычислим суммы, необходимые для расчета коэффициентов b_0, b_1 уравнения регрессии $\ln y = b_0 + b_1 \ln x$ и коэффициента детерминации R^2 с помощью вспомогательной таблицы (табл. 2.2).

Таблица 2.2 – Вспомогательные вычисления для определения уравнения регрессии

№	x_i	y_i	$\ln x_i$	$(\ln x_i)^2$	$\ln y_i$	$\ln x_i \cdot \ln y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	1,2528	1,5694	2,7726	3,4734	2,7294	0,0240	0,0393
2	2,4	13	0,8755	0,7664	2,5649	2,2455	2,5406	0,0011	0,0001
3	4,9	19	1,5892	2,5257	2,9444	4,6794	2,8977	0,1046	0,1370
4	4,2	18	1,4351	2,0595	2,8904	4,1479	2,8206	0,0606	0,0999
5	3	12	1,0986	1,2069	2,4849	2,7299	2,6523	0,0061	0,0080
6	1,3	11	0,2624	0,0688	2,3979	0,6291	2,2339	0,1159	0,0311
7	1	8	0,0000	0,0000	2,0794	0,0000	2,1026	0,2225	0,2449
8	3	14	1,0986	1,2069	2,6391	2,8993	2,6523	0,0061	0,0042
9	1,5	9	0,4055	0,1644	2,1972	0,8909	2,3055	0,0723	0,1422
10	4,1	16	1,4110	1,9909	2,7726	3,9121	2,8085	0,0548	0,0393
Σ	28,9	136	9,4286	11,5590	25,7435	25,6076		0,6681	0,7460
Среднее:	2,89	13,6	0,9429	1,1559	2,5743	2,5608			

Подставив результаты расчетов в систему уравнений (2.1) получим:

$$\begin{cases} 10 \cdot b_0 + 9,4229 \cdot b_1 = 25,7435 \\ 9,4229 \cdot b_0 + 11,5590 \cdot b_1 = 25,6076 \end{cases}$$

Для решения системы нормальных уравнений используем надстройку «Поиск решения» пакета *MS Excel* (**Данные-Поиск решения**). Составим исходную табличную модель (рис. 2.5). Все действия выполняются аналогично поиску коэффициентов для экспоненциального распределения.

В блок «Переменные» в первую строку записываем буквенные обозначения искомых коэффициентов системы алгебраических уравнений b_0, b_1 . Во вторую строку записываем произвольные числовые значения (удобнее в качестве числовых значений поставить номера коэффициентов, т. е. 0 и 1), в дальнейшем, после выполнения команды Поиск решения, в этих ячейках получим решение системы (значения коэффициентов b_0, b_1).

В блок «Матрица исходной системы» записываем числовые значения системы линейных уравнений при коэффициентах b_0, b_1 .

Переменные	
b0	b1
0	1

Матрица исходной системы	
10	9,4285
9,4285	11,559

Значения левых частей уравнений	
9,4285	=
11,559	=

Свободные члены исходной системы	
25,7434	
25,6076	

Рис. 2.5. Табличная модель решения системы линейных уравнений средствами Excel

«Значения левых частей уравнений» в верхнюю ячейку вводим формулу: =СУММПРОИЗВ(«диапазон числовых значений переменных b_0, b_1 (0 и 1)»; «диапазон первой строки матрицы исходной системы (10 и 9,4285)»). Автоматически заполняем весь столбец «Значения левых частей уравнений».

В блок «Свободные члены исходной системы» в столбец записываем значения правой части исходной системы и вызываем форму «Поиск решения» рис. 2.6.

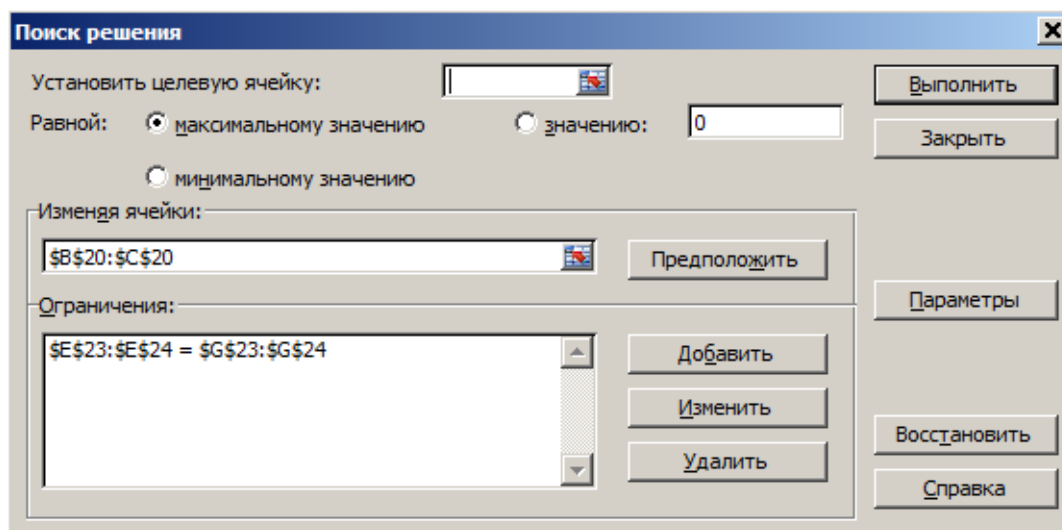


Рис. 2.6. Окно «Поиск решения»

- Необходимо заполнить форму «Поиск решения» следующими данными:
- установить целевую ячейку – ничего не ставить;
 - равной – максимальному значению;
 - изменяя ячейки – диапазон строки значений переменных (в нашем случае выделить значения 0 и 1);
 - ограничения – диапазон «Значения левых частей уравнений» = диапазон «Свободные члены исходной системы»;
 - нажать «Выполнить»;
 - в форме «Результаты поиска решения» (рис. 2.7) выбрать «Сохранить найденное решение»;
 - нажать ОК.

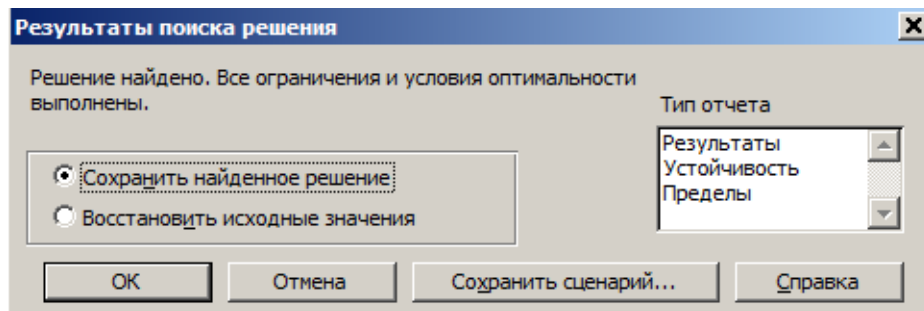


Рис. 2.7. Окно «Результаты поиска решения»

Результат выполнения команды «Поиск решения» для коэффициентов b_0, b_1 представлен на рис. 2.8.

Переменные	
b0	b1
2,1026	0,5003

Матрица коэффициентов исходной системы	
10	9,4285
9,4285	11,559

Значения левых частей уравнений
25,7434
25,6076

Свободные члены исходной системы
25,7434
25,6076

Рис. 2.8. Результаты выполнения команды «Поиск решения» для степенной функции

Полученные коэффициенты b_0 и b_1 приводим к коэффициентам a_0 и a_1 . Исходя из введенных обозначений и определения логарифма, получаем.
 $a_0 = e^{b_0}$, $a_1 = b_1$.

Для определения экспоненциального распределения необходимо вычислить коэффициенты $a_0 = e^{b_0} = 8,1878$ и $a_1 = b_1 = 0,5003$.

Тогда уравнение регрессии примет вид $y = a_0 \cdot x^{a_1} = 8,1878 \cdot x^{0,5003}$

Определим коэффициент детерминации R^2 . Для этого необходимо рассчитать и заполнить пустые столбцы *табл. 2.2* данными. При расчетах

необходимо использовать формулы: $y_i^p = b_0 + b_1 \cdot \ln x_i$, $\bar{y}_i = \frac{\sum_{i=1}^n \ln y_i}{n}$,
 $y_i = \ln y_i$, тогда

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{0,6681}{0,7460} \cdot 100\% = 89,52 \%$$

Парабола

Для нахождения коэффициентов уравнения параболы $y = a_0 + a_1 x + a_2 x^2$ необходимо решить систему уравнений (2.4).

Построим вспомогательную *табл. 2.3* и вычислим суммы, необходимые для дальнейшего расчета коэффициентов a_0 , a_1 , a_2 и коэффициента детерминации R^2 .

Подставим результаты расчетов в систему уравнений (2.4) и получим:

$$\begin{cases} 10 \cdot a_0 + 28,9 \cdot a_1 + 99,41 \cdot a_2 = 136 \\ 28,9 \cdot a_0 + 99,41 \cdot a_1 + 377,93 \cdot a_2 = 435,3 \\ 99,41 \cdot a_0 + 377,93 \cdot a_1 + 1524,38 \cdot a_2 = 1594,39 \end{cases} \quad (2.4)$$

Для решения системы нормальных уравнений используем надстройку «Поиск решения» пакета *MS Excel* (**Данные-Поиск решения**). Составим исходную табличную модель (*рис. 2.9*).

В блок «Переменные» в первую строку записываем буквенные обозначения искомых коэффициентов системы алгебраических уравнений a_0 , a_1 , a_2 . Во вторую строку записываем произвольные числовые значения

(удобнее в качестве числовых значений поставить номера коэффициентов, т. е. 0, 1, 2), в дальнейшем, после выполнения команды Поиск решения, в этих ячейках получим решение системы (значения коэффициентов a_0, a_1, a_2).

Таблица 2.3 – Вспомогательные вычисления для определения уравнения регрессии

№	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i \cdot y_i$	$x_i^2 \cdot y_i$	y_i^p	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3,5	16	12,25	42,875	150,063	56	196	15,1397	2,3708	5,76
2	2,4	13	5,76	13,824	33,1776	31,2	74,88	12,1989	1,9632	0,36
3	4,9	19	24,01	117,65	576,48	93,1	456,19	19,1299	30,5795	29,16
4	4,2	18	17,64	74,088	311,17	75,6	317,52	17,1002	12,2514	19,36
5	3	12	9	27	81	36	108	13,7818	0,0330	2,56
6	1,3	11	1,69	2,197	2,8561	14,3	18,59	9,4289	17,3981	6,76
7	1	8	1	1	1	8	8	8,7031	23,9794	31,36
8	3	14	9	27	81	42	126	13,7818	0,0330	0,16
9	1,5	9	2,25	3,375	5,0625	13,5	20,25	9,9198	13,5437	21,16
10	4,1	16	16,81	68,921	282,576	65,6	268,96	16,8159	10,3420	5,76
Σ	28,9	136	99,41	377,93	1524,38	435,3	1594,39		112,4942	122,40
Среднее:	2,89	13,6								

В блок «Матрица исходной системы» записываем числовые значения системы линейных уравнений при коэффициентах a_0, a_1, a_2 .

Переменные		
a0	a1	a2
0	1	2

Матрица исходной системы		
10	28,9	99,41
28,9	99,41	377,93
99,41	377,93	1524,38

Значения левых частей уравнений
136
435,299
1594,39

=
=
=

Свободные члены исходной системы
136
435,3
1594,39

Рис. 2.9. Табличная модель решения системы линейных уравнений средствами *Excel*

«Значения левых частей уравнений» в верхнюю ячейку вводим формулу: =СУММПРОИЗВ(«диапазон числовых значений переменных a_0, a_1, a_2 (0, 1, 2)»; «диапазон первой строки матрицы исходной системы (10, 28,9, 99,41)»). Автоматически заполняем весь столбец «Значения левых частей уравнений». В блок «Свободные члены исходной системы» в столбец записываем значения правой части исходной системы и вызываем форму «Поиск решения» *рис. 2.10*.

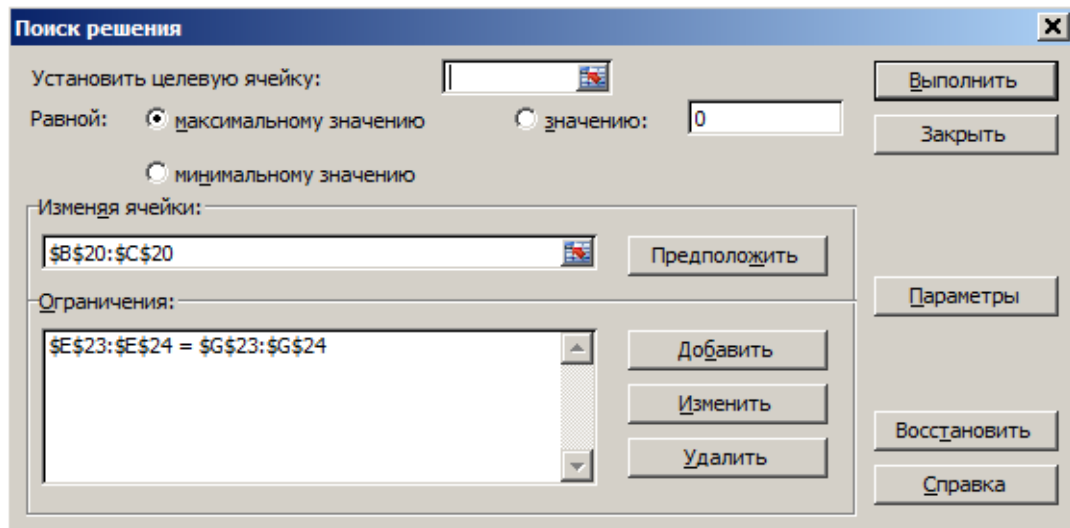


Рис. 2.10. Окно «Поиск решения»

- Необходимо заполнить форму «Поиск решения» следующими данными:
- установить целевую ячейку – ничего не ставить;
 - равной – максимальному значению;
 - изменяя ячейки – диапазон строки значений переменных (в нашем случае выделить значения 0, 1, 2);
 - ограничения – диапазон «Значения левых частей уравнений» = диапазон «Свободные члены исходной системы»;
 - нажать «Выполнить»;
 - в форме «Результаты поиска решения» (*рис. 2.11*) выбрать «Сохранить найденное решение»;
 - нажать ОК.

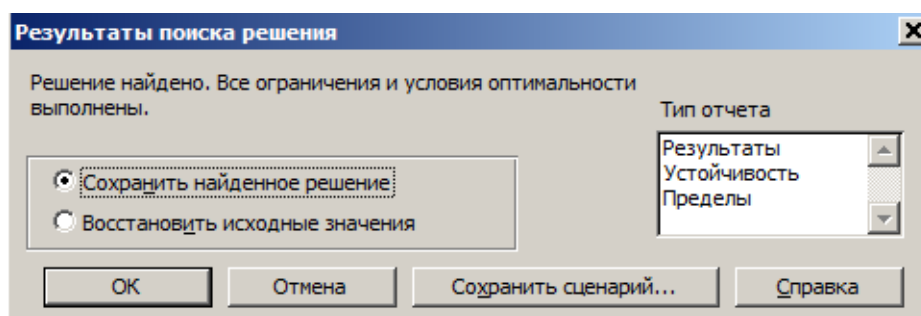


Рис. 2.11. Окно «Результаты поиска решения»

Результат выполнения команды «Поиск решения» для коэффициентов a_0, a_1, a_2 представлен на рис. 2.12.

Переменные		
a0	a1	a2
6,3756	2,2568	0,0706

Матрица исходной системы		
10	28,9	99,41
28,9	99,41	377,93
99,41	377,93	1524,38

Значения левых частей уравнений
136
435,2999
1594,39

Свободные члены исходной системы
136
435,3
1594,39

Рис. 2.12. Результаты выполнения команды «Поиск решения» для параболы

Полученные коэффициенты a_0, a_1, a_2 подставляем в уравнение регрессии и получаем

$$y = 6,3756 + 2,2568 \cdot x + 0,0706 \cdot x^2$$

Определим коэффициент детерминации R^2 . Для этого необходимо рассчитать и заполнить пустые столбцы табл. 2.3 данными. При расчетах необходимо использовать формулу: $y_i^p = a_0 + a_1 x + a_2 x^2$.

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{112,4942}{122,40} \cdot 100\% = 91,90\%$$

2.3. Решение задачи с использованием пакета *MS Excel*

Проведем регрессионный анализ с использованием режима **Регрессия** *MS Excel*. Пакет позволяет оценить линейную модель регрессии. Поэтому можно провести анализ только для экспоненциального и степенного распределения. Исследование параболы с применением средств *MS Excel* невозможно.

Экспонента

Исходные данные для проведения оценки необходимо выбирать из *табл. 2.1*. Поскольку *MS Excel* позволяет проводить анализ только линейных моделей регрессии, то из *табл. 2.1* необходимо извлекать данные для входного интервала y как $\ln y_i$. Входной интервал для x выбирается из столбца со значениями x_i . После выполнения команды **Данные-Анализ данных-Регрессия** открывается окно (*рис. 2.12*) «Регрессия» где и задаются оба интервала. Уровень надежности устанавливается 95 % и задается выходной интервал – ячейка, в которую будет записан результат решения.

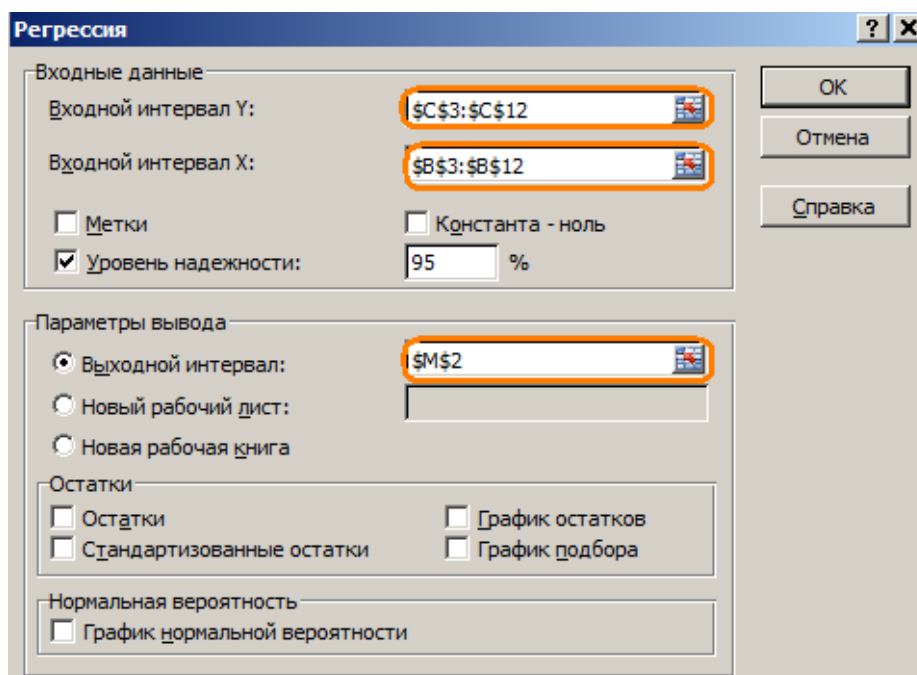


Рис. 2.12. Настройки окна «Регрессия»

Результаты по регрессионной статистике, полученные при расчете, представлены на *рис. 2.13*.

Регрессионная статистика	
Множественный R	0,947090055
R-квадрат	0,896979572
Нормированный R-квадрат	0,884102019
Стандартная ошибка	0,098013712
Наблюдения	10

Дисперсионный анализ					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	0,669149	0,669149045	69,65450171	3,21575E-05
Остаток	8	0,076854	0,009606688		
Итого	9	0,746003			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,981269415	0,077527	25,5558273	5,89199E-09
Переменная X 1	0,205216912	0,024589	8,345927253	3,21575E-05

Рис. 2.13. Результаты расчета

Видно, что коэффициенты экспоненциальной регрессии b_0 и b_1 являются значимыми, поскольку показатель P -значение меньше заданного уровня значимости $\alpha = 0,05$.

Значение $R^2 = 0,8969$ говорит о хорошем описании исходных данных регрессионной кривой. А рассчитанный уровень значимости $\alpha_p = 3,21575E - 05 < 0,05$ (показатель значимость F в таблице Дисперсионный анализ) подтверждает значимость R^2 .

Выполненные «вручную» расчеты полностью совпадают с расчетами в пакете *MS Excel*.

Степенная функция

Исходные данные для проведения оценки необходимо выбирать из табл. 2.2. Поскольку *MS Excel* позволяет проводить анализ только линейных моделей регрессии, то из табл. 2.2 необходимо извлекать данные для входного интервала y как $\ln y_i$. Входной интервал для x выбирается из столбца со значениями $\ln x_i$. После выполнения команды **Данные-Анализ данных-Регрессия** открывается окно (рис. 2.12) «Регрессия» где и задаются оба интервала. Уровень надежности устанавливается 95 % и задается выходной интервал – ячейка, в которую будет записан результат решения.

Результаты по регрессионной статистике, полученные при расчете, представлены на рис. 2.14.

Регрессионная статистика	
Множественный R	0,9461794
R-квадрат	0,89525547
Нормированный R-квадрат	0,8821624
Стандартная ошибка	0,09883047
Наблюдения	10

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	Значимость <i>F</i>
Регрессия	1	0,667863	0,667863	68,3763	3,44E-05
Остаток	8	0,07814	0,009767		
Итого	9	0,746003			

	Коэффициенты	Стандартная ошибка	t- статистика	P- Значение
Y-пересечение	2,10271616	0,065037	32,33094	9,13E-10
Переменная X 1	0,50021266	0,060493	8,268996	3,44E-05

Рис. 2.14. Результаты расчета

Видно, что коэффициенты экспоненциальной регрессии b_0 и b_1 являются значимыми, поскольку показатель P -значение меньше заданного уровня значимости $\alpha = 0,05$.

Значение $R^2 = 0,8952$ говорит о хорошем описании исходных данных регрессионной кривой. А рассчитанный уровень значимости

$\alpha_p = 3,44E - 05 < 0,05$ (показатель значимости F в таблице Дисперсионный анализ) подтверждает значимость R^2 .

Выполненные «вручную» расчеты полностью совпадают с расчетами в пакете *MS Excel*.

Построение линии тренда

На каждом листе исследуемых функций построить точечные диаграммы исходных данных задачи. Добавить линии тренда. Для этого необходимо щелкнуть правой кнопкой мыши по данным на диаграмме и выбрать команду «Добавить линию тренда» (рис. 2.15). В открывшемся окне (рис. 2.16) выбрать параметры линии тренда в соответствии с исследуемой функцией и отметить флажками вывод уравнения линии тренда и коэффициента детерминации на график.

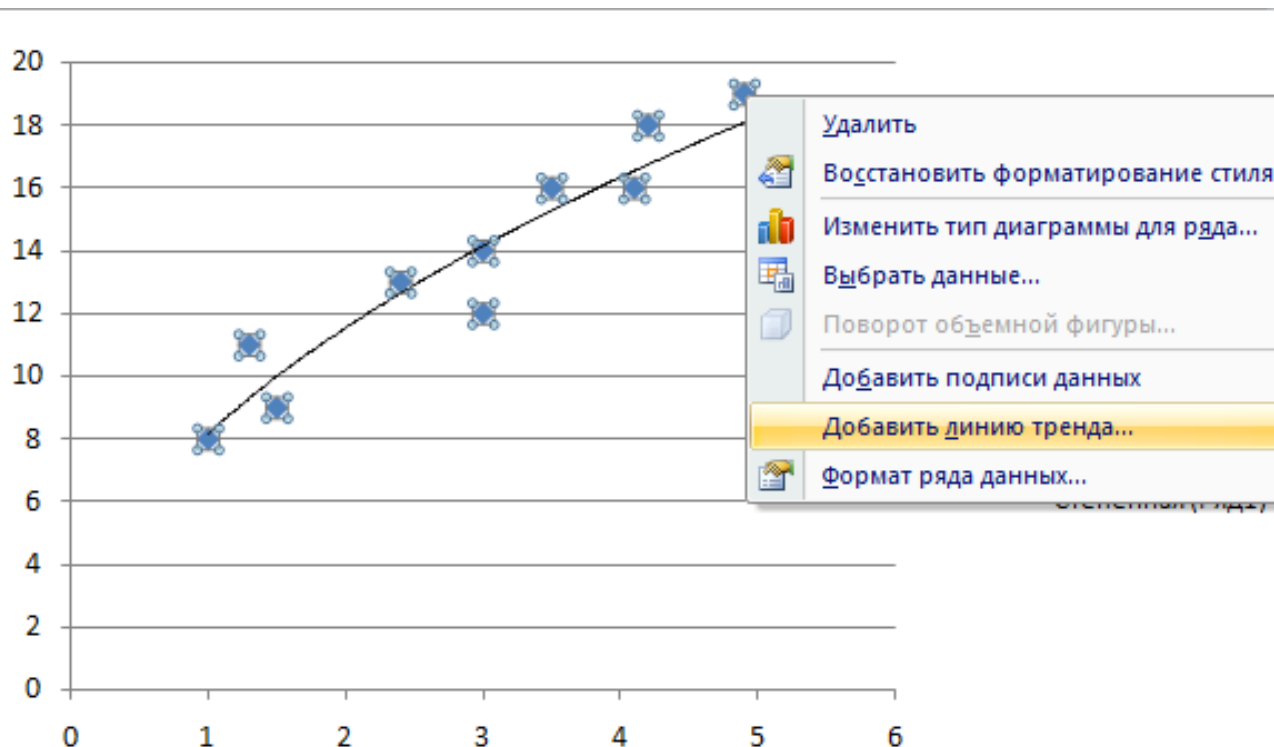


Рис. 2.15. Добавление линии тренда на диаграмму

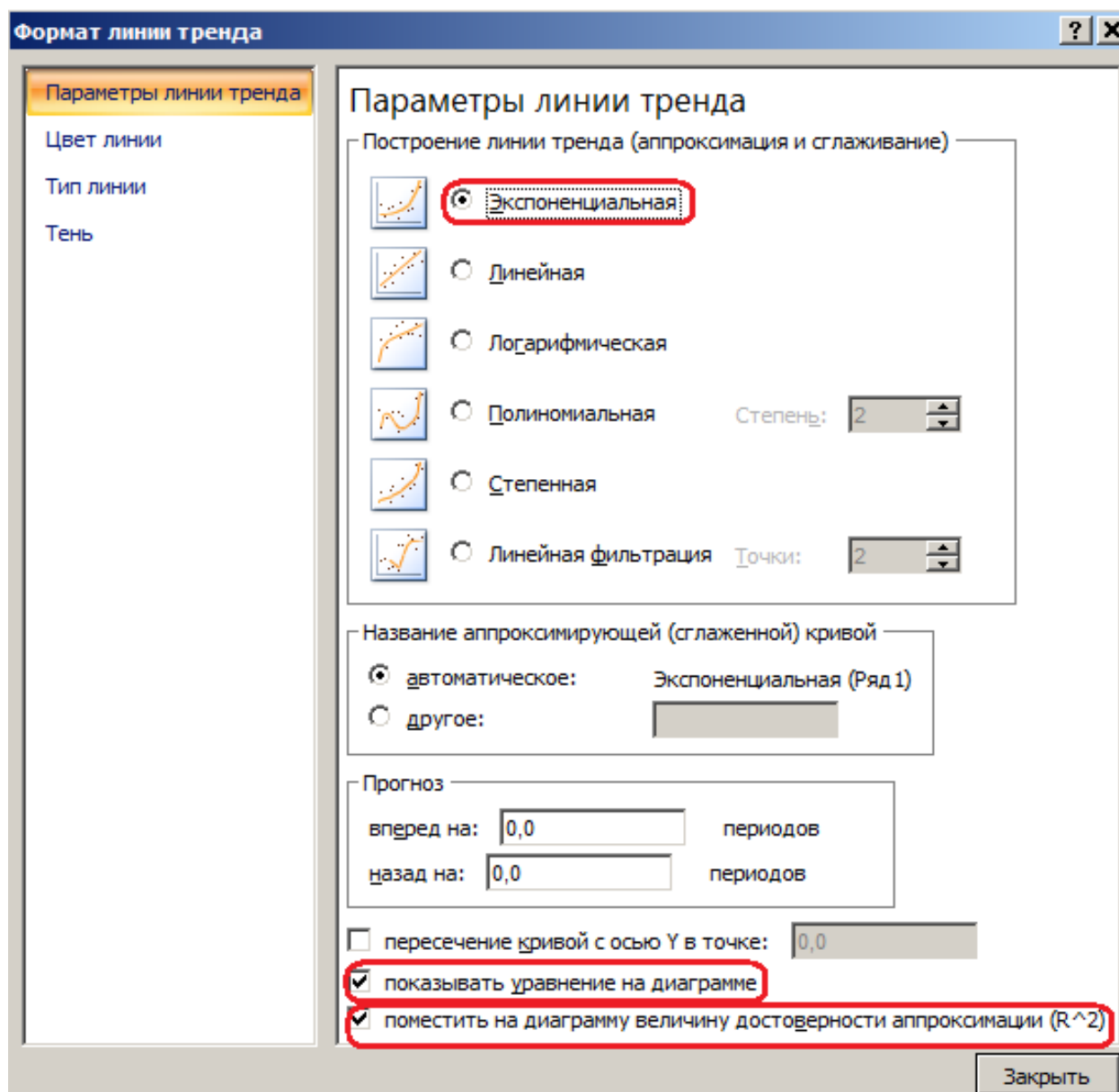


Рис. 2.16. Определение параметров линии тренда

На рис. 2.17-2.19 представлены диаграммы экспоненциальной, степенной и параболической аппроксимаций исходных данных.

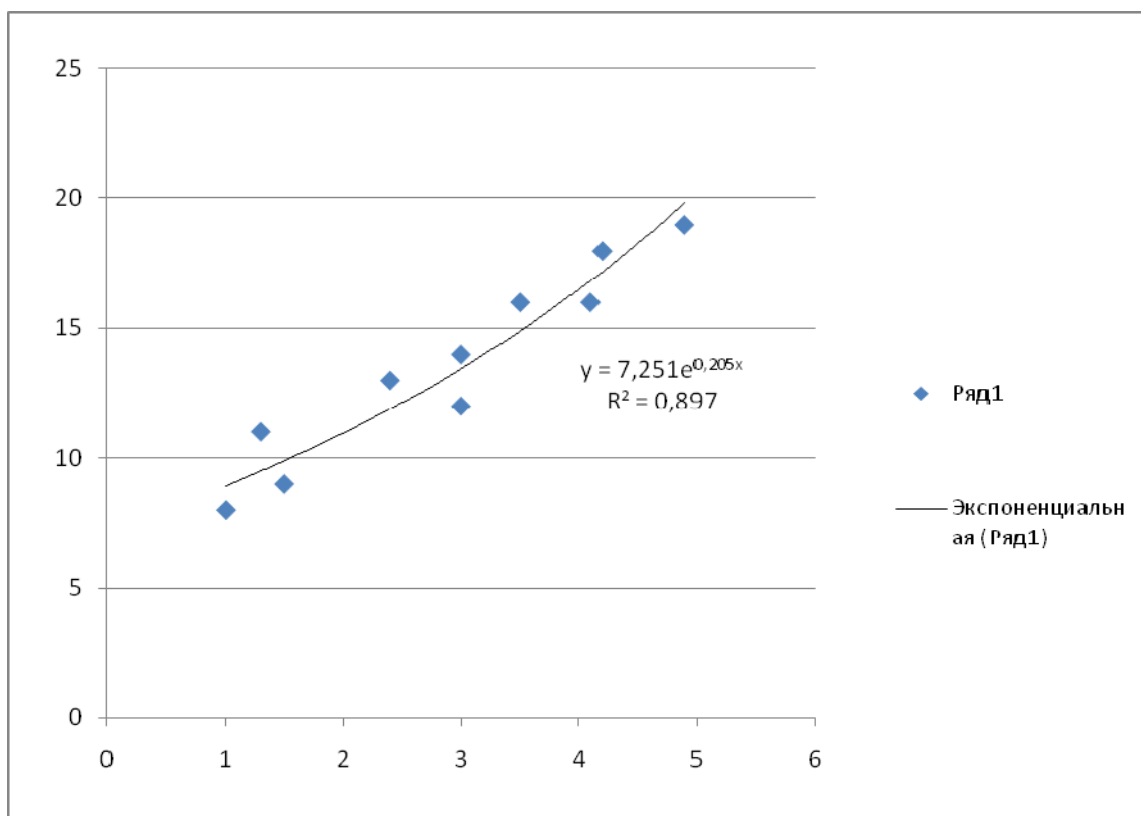


Рис. 2.17. Экспоненциальная аппроксимация

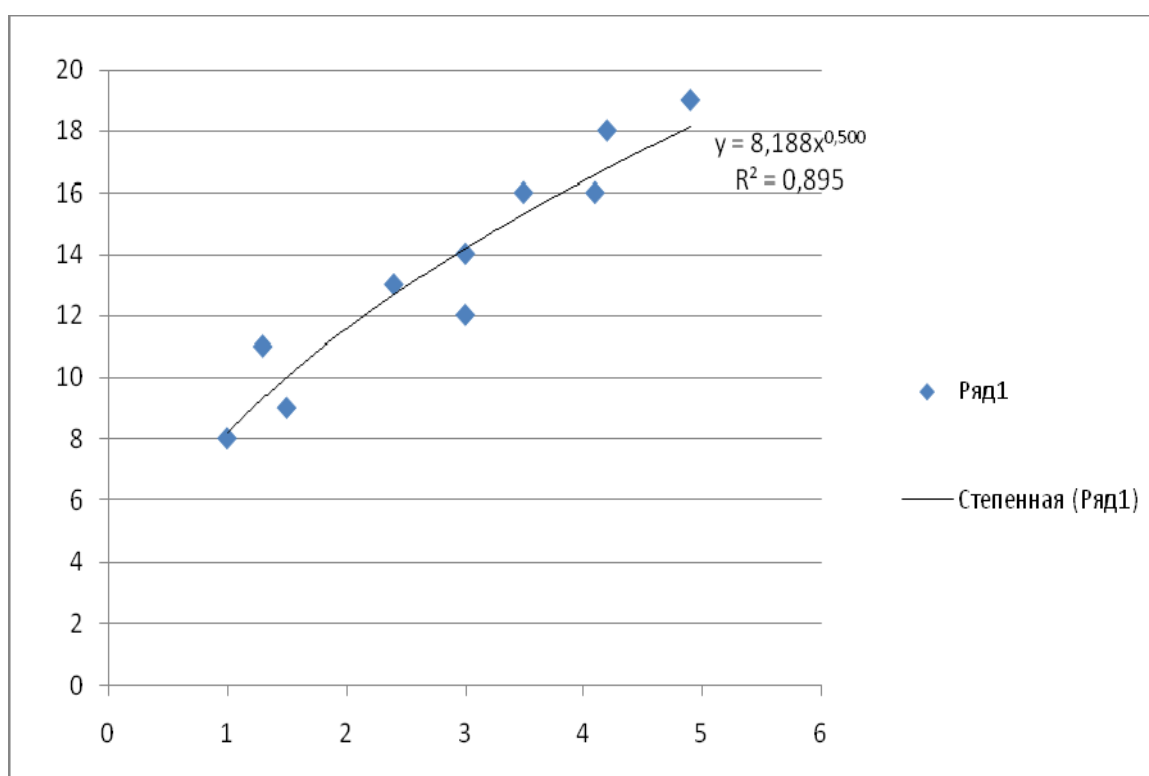


Рис. 2.18. Степенная аппроксимация

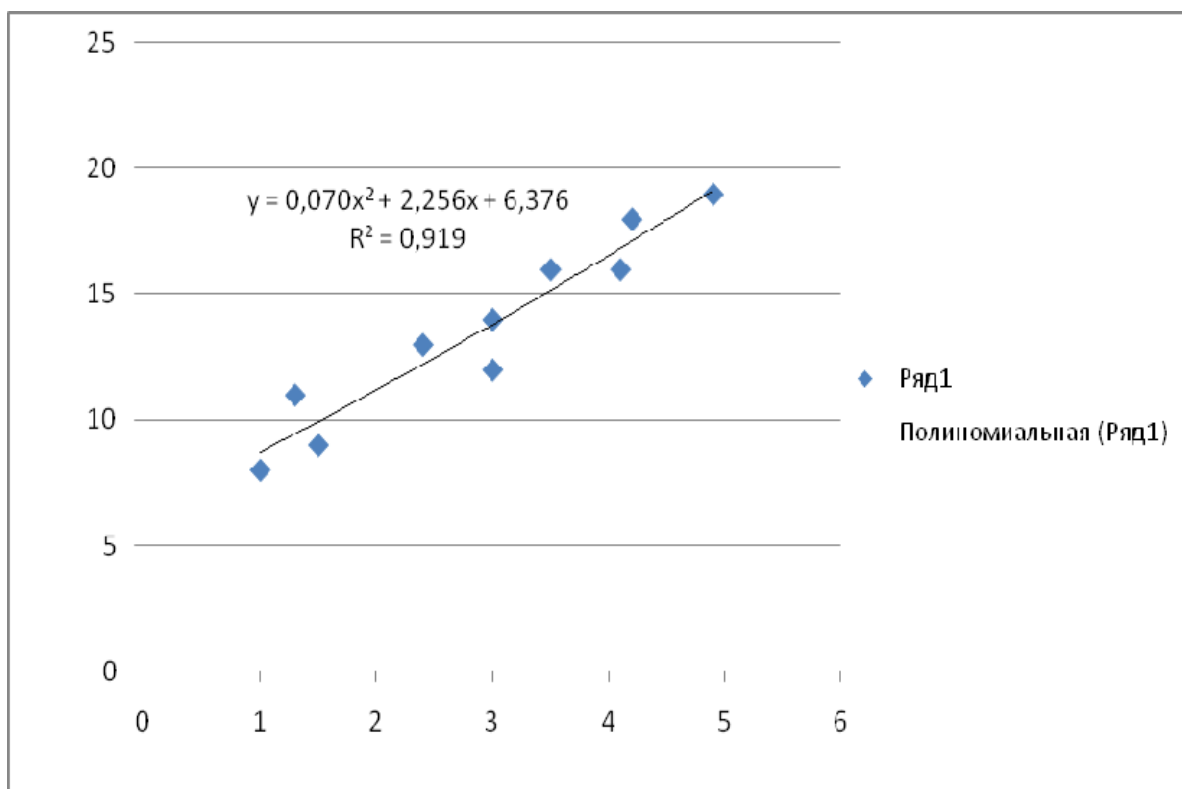


Рис. 2.19. Параболическая аппроксимация

Сравнивая результаты, вычисленные «вручную» с параметрами линий тренда на диаграммах видно, что вычисленные коэффициенты рассчитаны правильно.

Таким образом, линейная функция имеет коэффициент детерминации 0,918, экспоненциальная – 0,897, степенная – 0,895, парабола – 0,919. Наилучшее описание исходных данных имеет линейная и параболическая функции регрессии.

2.4. Порядок выполнения работы

1. Запустить пакет *MS Excel* и проверить наличие кнопки **Поиск решения** на вкладке **Данные** в группе **Анализ**. Если такая кнопка недоступна, необходимо загрузить надстройку «Пакет анализа». Для этого необходимо:

- щелкнуть значок **Кнопка Microsoft Office** , а затем **Параметры Excel**;
- перейти на вкладку **Надстройки**, а затем в поле **Управление** выбрать **Надстройки Excel**;

– нажать кнопку **Перейти** и в поле **Доступные надстройки** установить флажок **Поиск решения**, а затем нажать кнопку **ОК**.

2. На отдельных листах в *MS Excel* создать макеты табл. 2.1-2.3 для расчетов функций регрессии «вручную». Ввести в них исходные данные и рассчитать необходимые суммы.

3. Найти коэффициенты линий регрессии, используя функцию «Поиск решения».

4. Рассчитать коэффициенты детерминации для каждой зависимости.

5. Используя средства *MS Excel* провести регрессионный анализ для экспоненциальной и степенной функций.

6. Построить диаграммы для каждой зависимости и добавить линию тренда исследуемой функции.

7. Сделать выводы, сравнив коэффициенты детерминации линейной, экспоненциальной, степенной, параболической функций.

3. Требования к выполнению контрольной работы. Выбор варианта задания

Студент выполняет контрольную по индивидуальному заданию. В соответствии с последней цифрой в номере зачетки необходимо выбрать вариант и данные индивидуального задания в *табл. 3.1*.

В контрольной работе необходимо провести расчеты методом наименьших квадратов для линейной регрессии и одного вида нелинейной регрессии (экспонента, степенная, парабола) по выбору студента. Необходимо провести расчеты «вручную» и с использованием пакета *MS Excel*. По итогам расчетов выбрать функцию, которая наилучшим образом описывает экспериментальные данные, подтвердить свой выбор количественной оценкой, в конце работы написать выводы.

Контрольную работу необходимо представить в виде пояснительной записки в формате *MS Word* и файла с расчетами в формате *MS Excel*.

Таблица 3.1 – Варианты индивидуальных заданий

№ варианта	Переменные	№ опыта				X^*
		1	2	3	4	
0	X	5	3	6	4	1.5
	Y	7.5	7	5	3.5	?
1	X	1	2	3	4	1.6
	Y	30	7	8	1	?
2	X	1	2	3	4	2.3
	Y	25	7	7	2	
3	X	9	5	2	3	2.9
	Y	25	7	7	2	?
4	X	1	2	3	4	2.6
	Y	15	10	7	0.5	?
5	X	10	3	6	4	8
	Y	25	7	7	2	?
6	X	9	5	2	3	2.5
	Y	15	8.5	7.5	5	?
7	X	2	3	7	8	7.5
	Y	11	8.5	6.5	5	?
8	X	10	3	6	4	9
	Y	15	7	8	6	?
9	X	2	3	4	5	4.5
	Y	13	9	8	7	?

4. Список вопросов к зачету

1. Общая схема исследования зависимостей.
2. Типы переменных.
3. Этапы статистического анализа.
4. Области применения прикладной статистики.
5. Случайные события. Операции над событиями.
6. Законы распределения случайных величин.
7. Числовые характеристики законов распределения случайных величин.
8. Основные понятия непараметрической статистики.
9. Наглядные методы представления выборок. Гистограмма.
10. Точечные статистики.
11. Интервальное оценивание параметров.
12. Корреляционный анализ. Измерение тесноты парной связи между количественными переменными. Диаграмма рассеяния. Коэффициент корреляции.
13. Проверка значимости парного коэффициента корреляции. Коэффициент детерминации. Анализ множественных связей. Частный (условный) коэффициент корреляции.
14. Ранговая корреляция.
15. Статистические методы Data Mining.
16. Выбор общего вида или класса функции регрессии. Оценка параметров регрессионной модели методом наименьших квадратов.
17. Анализ точности уравнения регрессии. Проверка адекватности уравнения регрессии.
18. Анализ множественной регрессии.
19. Анализ временных рядов. Индексы. Базовые модели временных рядов.
20. Сглаживание временного ряда. Разложение временного ряда.
21. Дискретные динамические модели: модели скользящего среднего и авторегрессии.
22. Оценка параметров уравнения авторегрессии.
23. Модели объектов и методы их исследования.
24. Проверка гипотез о законе распределения.
25. Проверка гипотез о параметрах распределения.

5. Список литературы для подготовки к зачету

Автор	Заглавие	Издательство	Год издания	Ссылка
Основная учебная литература				
Рожков Н. Н.	Прикладная статистика	Санкт-Петербург: СПбГУПТД	2021	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2021192
Федорова, Е. П.	Социально-экономическая статистика	Саратов: Вузовское образование	2021	http://www.iprbookshop.ru/107930.html
Мельниченко, А. С.	Математическая статистика и анализ данных	Москва: Издательский Дом МИСиС	2018	http://www.iprbookshop.ru/78563.html
Казаков В.Г., Громова Е.Н.	Планирование экспериментальных исследований и статистическая обработка данных. Основы научных исследований в промышленной	Санкт-Петербург: СПбГУПТД	2020	http://publish.sutd.ru/tp_ext_inf_publish.php?id=20205098
Глебов, В. И., Криволапов, С. Я.	Практикум по математической статистике. Проверка гипотез с использованием	Москва: Прометей	2019	http://www.iprbookshop.ru/94504.html
Сальникова, К. В.	Практические основы статистики и эконометрического	Саратов: Ай Пи Ар Медиа	2020	http://www.iprbookshop.ru/91121.html
Башина, О. Э., Ярных, Э. А., Башина, О. Э.	Статистика торгового дела	Москва: Московский гуманитарный	2015	http://www.iprbookshop.ru/50674.html
Дополнительная учебная литература				
Пименов В. И.	Обработка информации	СПб.: СПбГУПТД	2017	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2017669
Кравец Т. А., Пименов В. И.	Математические методы и компьютерные технологии в науке и	СПб.: СПбГУПТД	2018	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2018373
Пименов В. И., Пименов И. В.	Интеллектуальный анализ данных	СПб.: СПбГУПТД	2017	http://publish.sutd.ru/tp_ext_inf_publish.php?id=201748

Содержание

ВВЕДЕНИЕ	3
1. Аппроксимация экспериментальных данных линейной функцией	4
1.1. Основные понятия и определения	4
1.2. Линейная регрессия.....	5
1.3. Решение задачи линейной регрессии «вручную».....	8
1.4. Решение задачи с использованием пакета <i>MS Excel</i>	10
1.5. Порядок выполнения работы.....	15
2. Аппроксимация экспериментальных данных нелинейными функциями.....	17
2.1. Виды функций нелинейной регрессии	17
2.2. Решение задачи «вручную»	20
2.3. Решение задачи с использованием пакета <i>MS Excel</i>	31
2.4. Порядок выполнения работы.....	37
3. Требования к выполнению контрольной работы. Выбор варианта задания	39
4. Список вопросов к зачету	40
5. Список литературы для подготовки к зачету	41