

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Баламирзоев Азиза Абдулмичович
Должность: Врио ректора
Дата подписания: 03.06.2022 14:07:06
Уникальный программный ключ:
777029a1882856141bfb9e855f0a3c8b6eadae59e

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

КАФЕДРА ЕГОиСД

**Учебно-методические указания к выполнению
лабораторных работ по дисциплине
«Прогнозирование социально-экономических
процессов»
для студентов направления подготовки
бакалавров 09.03.03. – Прикладная информатика**

Дербент 2020

УДК:338.27

ББК :22.171.58

Методические указания к выполнению лабораторных работ по дисциплине «Прогнозирование социально-экономических процессов» /ДГТУ, Махачкала, ИПЦ ДГТУ, 2020, 48с.

Данные методические указания содержат цели и задачи курса по основам прогнозирования социально-экономических процессов и основные компетенции, которыми студент должен овладеть в процессе изучения дисциплины.

Предназначены для студентов очной и заочной форм обучения по направлению 09.03.03. – Прикладная информатика.

Составитель: ст.преподаватель Атаева Б.М.

Рецензенты:

1. Джумалиева Е.Р., ст. преподаватель филиала ДГТУ в г.Дербент.
2. Вурдиханов В.Р., научный сотрудник ДФ АГЭУ, к.т.н., доцент.

Печатается по решению Ученого совета ДГТУ.

Содержание

Введение	4
Лабораторная работа № 1 «Разработка прогноза с помощью методов экстраполяции. Прогноз уровня безработицы в республике Дагестан методом скользящей средней».....	6
Лабораторная работа № 2 «Прогнозы с применением метода скользящего среднего в программе для работы с электронными таблицами Microsoft Excel».....	8
Лабораторная работа № 3 «Разработка прогноза с помощью методов экстраполяции. Прогноз уровня безработицы в республике Дагестан на основе метода экспоненциального сглаживания».....	12
Лабораторная работа № 4 «Прогнозирование с применением метода экспоненциального сглаживания в Microsoft Excel».....	14
Лабораторная работа № 5 «Разработка прогноза с помощью методов экстраполяции. Прогноз уровня безработицы в республике Дагестан на основе метода наименьших квадратов».....	18
Лабораторная работа № 6 «Процедура графического прогнозирования в Microsoft Excel. Построение линии тренда».	21
Лабораторная работа № 7 «Прогнозирование с помощью функций регрессии Excel».....	26
Лабораторная работа № 8 «Прогнозирование на основе анализа трендов и сезонности».....	30
Лабораторная работа № 9 «Прогнозирование на основе парной регрессии».....	36
Лабораторная работа № 10 «Прогнозирование на основе множественной регрессии».....	41
Список использованной литературы	48

ВВЕДЕНИЕ

Масштаб стоящих перед российской экономикой проблем, а также качественный уровень развития современного научно-технического потенциала требует соответствующей теоретической и практической подготовки специалистов в области экономико-математического моделирования. Прогнозная информация, с одной стороны, необходима как основа планирования деятельности любого социально-экономического объекта, а с другой стороны - как предварительная оценка последствий принимаемых решений с целью их оптимизации. Отсюда ясна важность данной дисциплины для формирования специалиста в области математических методов и исследования операций в экономике.

В этой связи цель дисциплины «Прогнозирование социально-экономических процессов» вооружить студентов знаниями общих закономерностей составления научных прогнозов развития социально-экономических процессов; познакомить их с максимально широким инструментарием выработки прогнозов развития социально-экономических объектов, а также методиками его использования в практике прогнозирования; выработать в процессе обучения у студентов навыки грамотного использования аппарата математического моделирования посредством применения передовых информационных технологий.

Задачи курса: изучение методологических основ прогнозирования, а также приемов и методов прогнозирования социально-экономических процессов.

Дисциплина «Прогнозирование социально-экономических процессов» опирается на материал учебных дисциплин: «Математический анализ», «Теория вероятности и математическая статистика», «Экономическое моделирование», «Математические методы исследования операций», «Статистика» и других дисциплин.

Основная цель лабораторных занятий - углубленное изучение проблем, затронутых в лекционном курсе, и отработка навыков в применении изучаемых

методов и процедур прогнозирования с использованием современного программного обеспечения персональных компьютеров.

В качестве базового информационно-программного инструментария на лабораторных работах предлагается воспользоваться продуктами Excel. 5

Лабораторная №1 Разработка прогноза с помощью методов экстраполяции. Прогноз уровня безработицы в республике Дагестан методом скользящей средней

Цель работы: научиться строить прогнозы используя метод скользящей средней и уметь рассчитать ошибки полученных прогнозов.

Необходимый теоретический минимум:

Экстраполяция - это метод научного исследования, который основан на распространении прошлых и настоящих тенденций, закономерностей, связей на будущее развитие объекта прогнозирования. К методам экстраполяции относятся :

1. метод скользящей средней;
2. метод экспоненциального сглаживания;
3. метод наименьших квадратов.

Метод скользящих средних является одним из широко известных методов сглаживания временных рядов. Применяя этот метод, можно элиминировать случайные колебания и получить значения, соответствующие влиянию главных факторов.

Сглаживание с помощью скользящих средних основано на том, что в средних величинах взаимно погашаются случайные отклонения. Это происходит вследствие замены первоначальных уровней временного ряда средней арифметической величиной внутри выбранного интервала времени. Полученное значение относится к середине выбранного интервала времени (периода).

Затем период сдвигается на одно наблюдение, и расчет средней повторяется. При этом периоды определения средней берутся все время одинаковыми. Таким образом, в каждом рассматриваемом случае средняя центрирована, т.е. отнесена к серединной точке интервала сглаживания и представляет собой уровень для этой точки.

При сглаживании временного ряда скользящими средними в расчетах участвуют все уровни ряда. Чем шире интервал сглаживания, тем более плавным получается тренд. Сглаженный ряд короче первоначального на $(n-1)$ наблюдений, где n – величина интервала сглаживания.

При больших значениях n колеблемость сглаженного ряда значительно снижается. Одновременно заметно сокращается количество наблюдений, что создает трудности.

Выбор интервала сглаживания зависит от целей исследования. При этом следует руководствоваться тем, в какой период времени происходит действие, а следовательно, и устранение влияния случайных факторов.

Данный метод используется при краткосрочном прогнозировании. Его рабочая формула:

$$y_{t+1} = m_{t-1} + \frac{1}{n} \cdot (y_t - y_{t-1}), \text{ если } n = 3,$$

где $t + 1$ – прогнозный период; t – период, предшествующий прогнозному периоду (год, месяц и т.д.); Y_{t+1} – прогнозируемый показатель; m_{t-1} – скользящая средняя за два периода до прогнозного; n – число уровней, входящих в интервал сглаживания; Y_t – фактическое значение исследуемого явления за предшествующий период; Y_{t-1} – фактическое значение исследуемого явления за два периода, предшествующих прогнозному.

Задача. Имеются данные, характеризующие уровень безработицы в республике Дагестан, %

Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Постройте прогноз уровня безработицы в регионе на ноябрь, декабрь, январь месяцы, используя методы: скользящей средней, экспоненциального сглаживания, наименьших квадратов.
- Рассчитайте ошибки полученных прогнозов при использовании каждого метода.
- Сравните полученные результаты, сделайте выводы.

Решение методом скользящей средней

Для расчета прогнозного значения методом скользящей средней необходимо:

1. Определить величину интервала сглаживания, например равную 3 ($n = 3$).
2. Рассчитать скользящую среднюю для первых трех периодов
 $m_{\text{фев}} = (U_{\text{янв}} + U_{\text{фев}} + U_{\text{март}}) / 3 = (2,99 + 2,66 + 2,63) / 3 = 2,76$

Полученное значение заносим в таблицу в середину взятого периода.

Далее рассчитываем m для следующих трех периодов февраль, март, апрель.

$$m_{\text{март}} = (U_{\text{фев}} + U_{\text{март}} + U_{\text{апр}}) / 3 = (2,66 + 2,63 + 2,56) / 3 = 2,62$$

Далее по аналогии рассчитываем m для каждых трех рядом стоящих периодов и результаты заносим в таблицу.

3. Рассчитав скользящую среднюю для всех периодов, строим прогноз на ноябрь по формуле:

$$y_{t+1} = m_{t-1} + \frac{1}{n} \cdot (y_t - y_{t-1}), \text{ если } n = 3,$$

где $t + 1$ – прогнозный период; t – период, предшествующий прогнозному периоду (год, месяц и т.д.); U_{t+1} – прогнозируемый показатель; m_{t-1} – скользящая средняя за два периода до прогнозного; n – число уровней, входящих в интервал сглаживания; U_t – фактическое значение исследуемого явления за предшествующий период; U_{t-1} – фактическое значение исследуемого явления за два периода, предшествующих прогнозному.

$$U_{\text{ноябрь}} = 1,57 + 1/3 (1,42 - 1,56) = 1,57 - 0,05 = 1,52$$

Определяем скользящую среднюю m для октября.

$$m = (1,56 + 1,42 + 1,52) / 3 = 1,5$$

Строим прогноз на декабрь.

$$U_{\text{декабрь}} = 1,5 + 1/3 (1,52 - 1,42) = 1,53$$

Определяем скользящую среднюю m для ноября.

$$m = (1,42 + 1,52 + 1,53) / 3 = 1,49$$

Строим прогноз на январь.

$$U_{\text{январь}} = 1,49 + 1/3 (1,53 - 1,52) = 1,49$$

Заносим полученный результат в таблицу.

Месяцы	Уровень безработицы, Ут, %.	Скользящая средняя, т, %.	Расчет средней относительной ошибки, $\frac{Уф - Ур}{Уф} * 100, \%$
январь	2,99	-	-
февраль	2,66	2,76	$\frac{2,66-2,76}{2,66} * 100 = 3,76$
март	2,63	2,62	0,38
апрель	2,56	2,53	1,17
май	2,40	2,39	0,42
июнь	2,22	2,20	0,90
июль	1,97	1,97	0
август	1,72	1,75	1,74
сентябрь	1,56	1,57	0,64
октябрь	1,42	-	-
Итого:			9,01
Прогноз ноябрь	1,52		
Прогноз декабрь	1,53		
Прогноз январь	1,49		

Лабораторная работа №2 Прогнозы с применением метода скользящего среднего в программе для работы с электронными таблицами Microsoft Excel

Цель: научиться строить прогнозы с применением метода скользящего среднего на программе для работы с электронными таблицами Microsoft Excel.

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

Метод скользящего среднего применять достаточно просто, однако он не всегда позволяет сделать точный прогноз. При использовании этого метода прогноз любого периода представляет собой не что иное, как получение среднего показателя нескольких результатов наблюдений временного ряда. Например, если вы выбрали скользящее среднее за три месяца, прогнозом на май будет среднее значение показателей за февраль, март и апрель. Выбрав в качестве метода прогнозирования скользящее среднее за четыре месяца, вы сможете оценить майский показатель как среднее значение показателей за январь, февраль, март и апрель.

Вычисления с помощью этого метода довольно просты и достаточно точно отражают изменения основных показателей предыдущего периода. Иногда при составлении прогноза они эффективнее, чем методы, основанные на долговременных наблюдениях.

Например, вы составляете прогноз объема продаж давно и хорошо освоенной вашим предприятием продукции, причем средний показатель объема за последних несколько лет составляет 1000 единиц. Если ваша компания планирует значительное сокращение штата торговых агентов, логично предположить, что среднемесячный объем реализации будет сокращаться, по крайней мере, на протяжении нескольких месяцев.

Если, для прогнозирования объема продаж в будущем месяце вы воспользуетесь средним значением данного показателя за последние 24 месяца, то, вероятно, получите результат, несколько завышенный по сравнению с фактическим. Но если прогноз будет составлен на основании данных всего лишь за три последних месяца, то он намного точнее отразит последствия сокращения штата торговых агентов. В данном случае прогноз будет отставать по времени от фактических результатов на один-два месяца, как это показано на рисунке 1.

Разумеется, это происходит потому, что при применении скользящего среднего за три последних месяца каждый из трех показателей (за этот временной период) отвечает за одну треть значения прогноза. При 24-месячном скользящем среднем показатели этих же последних

месяцев отвечают только за 1/24 часть значения прогноза. Таким образом, чем меньше число результатов наблюдений, на основании которых вычислено скользящее среднее, тем точнее оно отражает изменения в уровне базовой линии.

Порядок выполнения лабораторной работы

Пример: обслуживание клиентов

Предположим, вы — менеджер отдела обслуживания клиентов фирмы, специализирующейся на разработке программного обеспечения. На днях вы получили от внештатной сотрудницы сообщение по электронной почте, в котором она известила вас, что в последнее время ей постоянно звонят клиенты с жалобами на новые программы вашей фирмы. Вы просите ее зарегистрировать все жалобы, поступающие в течение двух недель и сообщить вам результаты.

Полученный по истечении этого времени отчет включает ежедневное количество звонков с жалобами на конкретный программный продукт. Эти данные вы вводите в рабочий лист Excel, расположив их в ячейках A1:A10, как показано на рисунке 2. Чтобы понять, существует ли какая-либо определенная тенденция поступления жалоб, вы создаете на основе средних данных о полученных звонках скользящее среднее.

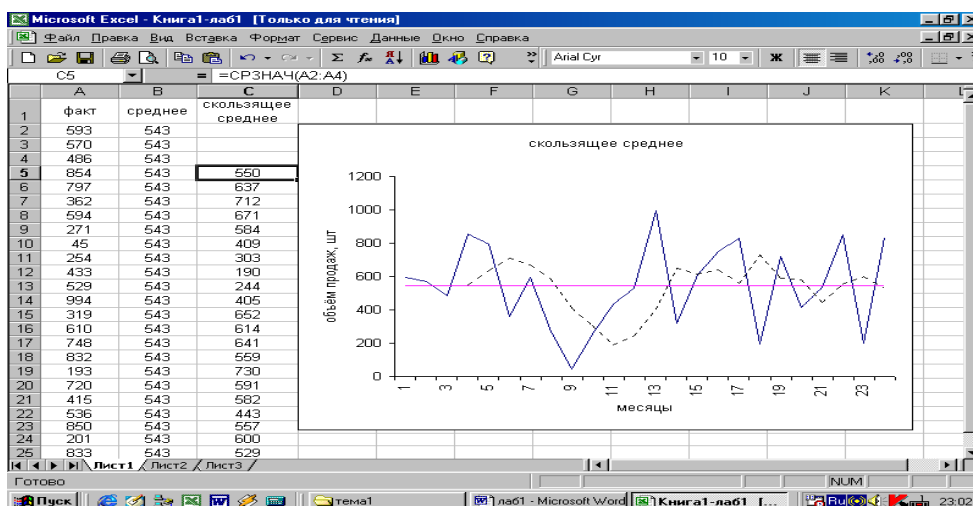


Рисунок 1 - Скользящее среднее объема продаж за три месяца позволяет отслеживать фактический объем продаж с большей точностью, нежели аналогичные наблюдения, но за длительный период времени

Вы решаете воспользоваться трехдневным скользящим средним. Почему за трехдневный период? Ответ на этот вопрос таков: скользящее среднее за меньший период может не отразить тенденцию, а за более продолжительный период слишком сгладит ее. Одним из способов создания скользящего среднего в Excel является прямое введение формулы.

Таким образом, чтобы получить трехдневное скользящее среднее количества телефонных звонков, вы вводите в ячейку B4:

=СРЗНАЧ(A1:A3) (Результат: 10,33).

Затем с помощью средства *Автозаполнение* копируете и вставляете эту формулу в ячейки B5:B10. В данном случае (и это видно из рисунка 2) показатель скользящего среднего действительно имеет тенденцию к увеличению, поэтому поставьте в известность о тревожной ситуации руководство отдела тестирования продукции вашей компании.

При использовании средства *Автозаполнение* для горизонтального или вертикального перетаскивания выбранной ячейки или диапазона ячеек, надо установить указатель в маркер заполнения, который представляет собой маленький крестик в нижнем правом углу выбранного диапазона.

	A	B	C
1	10		
2	11		
3	10		
4	12	10,33	
5	13	11,00	
6	13	11,67	
7	13	12,67	
8	10	13,00	
9	16	12,00	
10	17	13,00	
11			
12			

Рисунок 2 - Прогнозы с использованием скользящего среднего приводят к потере некоторых данных в начальном периоде базовой линии

Составление прогнозов с помощью надстроек скользящего среднего

Другим способом применения скользящего среднего является использование надстройки *Пакет анализа*. Установить эту надстройку можно следующим образом.

1. Выберите команду *Сервис* ⇒ *Надстройки*.
2. Появится диалоговое окно *Надстройки*. Установите флажок опции *Пакет анализа* и щелкните на кнопке *ОК*.
3. При необходимости активизируйте рабочий лист, содержащий данные о вашей базовой линии.
4. В меню *Сервис* вы обнаружите новую команду *Анализ данных*. Выберите команду *Сервис* ⇒ *Анализ данных*.
5. Появится диалоговое окно *Анализ данных*, в котором содержатся все доступные функции анализа данных. Из списка выберите инструмент анализа *Скользящее среднее* и щелкните на кнопке *ОК*.
6. Появится диалоговое окно *Скользящее среднее*, как на рисунке 3.
7. В поле *Входной интервал* наберите данные о вашей базовой линии либо укажите диапазон в рабочем листе, ссылка на него появится в этом поле.
8. В поле *Интервал* введите количество месяцев, которые хотите включить в подсчет скользящего среднего.
9. В поле ввода *Выходной интервал* введите адрес ячейки, с которой хотите начать вывод (либо просто щелкните на этой ячейке в рабочем листе).
10. Щелкните на кнопке *ОК*.

Рисунок 3 - Диалоговое окно *Скользящее среднее* позволяет определить период наблюдений, входящих в каждое среднее значение

Excel выполняет вместо вас работу по внесению значений в формулу для вычисления скользящего среднего. Значения скользящего среднего начинаются со значений #Н/Д, которые равны значению указанного вами интервала минус один. Это делается из-за недостаточного

количества данных для вычисления среднего значения нескольких первых результатов наблюдений. На рисунке 4 показан результат вычисления скользящего среднего для данных, приведенных на рисунке 1, с интервалом в 3 периода.

Как выполнить вычисления с использованием скользящего среднего

Как правило, прогноз с применением скользящего среднего рассматривается как прогноз на период, непосредственно следующий за периодом наблюдения. Например, вы вычисляете скользящее среднее доходов от продаж по результатам трех месяцев, и последние данные наблюдений в базовой линии включают результаты за январь, февраль и март. Среднее значение этих трех данных обычно считается скользящим средним для апреля, т.е. первого месяца, непосредственно следующего за результатами наблюдения.

Однако настройка для вычисления скользящего среднего (линия тренда) связывает прогноз с конечными результатами наблюдений в конкретном среднем значении. Сравните, например, рисунок 1, на котором скользящее среднее получено путем введения формул непосредственно в рабочий лист, с рисунком 4, составленным с помощью настройки для вычисления скользящего среднего.

Заметьте, что каждый показатель скользящего среднего на рисунке 4 сдвинут на одну позицию по сравнению с рисунком 1. Это значит, что в данном примере прогнозом для третьего периода является значение 550, т.е. первое вычисленное скользящее среднее.

Не логично рассматривать некое число в качестве прогноза на период, на основе которого составлялся этот прогноз. В стандартной интерпретации значение 550 является прогнозом не на третий, а на четвертый месяц.

Удобнее всего вводить формулы вручную. Однако существует и другое решение: когда будет вычислено скользящее среднее, выделите все эти значения и сместите их на одну строку листа вниз. Это действие присоединит прогнозы именно к тем периодам, с которыми они связаны.

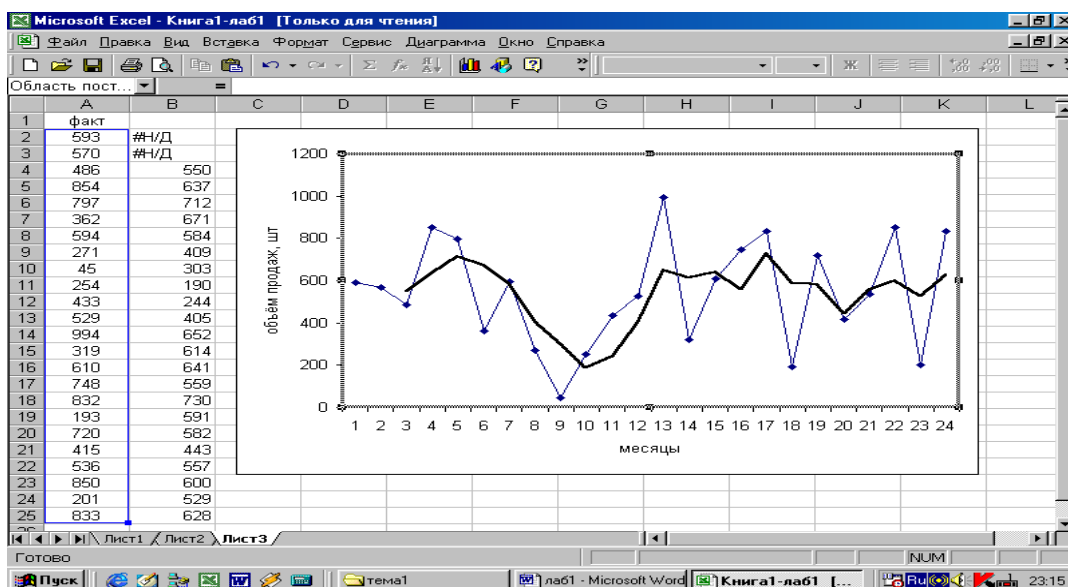


Рисунок 4 - Поскольку скользящее среднее базируется на данных предыдущих наблюдений, они имеют тенденцию запаздывать по сравнению с изменениями в базовых линиях, лежащих в их основе

Отчет по лабораторной работе должен включать:

1. Исходные данные и задание.
2. Расчётные формулы для проведения необходимых расчетов исходя из задания лабораторной работы.
3. Графическое представление полученных результатов расчета (при необходимости).
4. Результаты расчёта показателей, оформленные в виде таблиц.

5. Основные выводы.

Выводы: Выводы, сделанные в ходе проведенных расчетов должны быть четко и полно аргументированы.

Вопросы к защите лабораторной работы:

1. Какой процесс вы прогнозировали?
2. При помощи какого метода вы прогнозировали процесс? Его особенности.
3. Какие выводы можно сделать выводы?

Лабораторная работа №3 Разработка прогноза с помощью методов экстраполяции.
Прогноз уровня безработицы в республике Дагестан на основе метода экспоненциального сглаживания

Цель работы: научиться строить прогнозы, используя метод экспоненциального сглаживания и уметь рассчитать ошибки полученных прогнозов.

Необходимый теоретический минимум:

Экстраполяция - это метод научного исследования, который основан на распространении прошлых и настоящих тенденций, закономерностей, связей на будущее развитие объекта прогнозирования. **К методам экстраполяции относятся** метод скользящей средней, метод экспоненциального сглаживания, метод наименьших квадратов.

Метод экспоненциального сглаживания наиболее эффективен при разработке среднесрочных прогнозов. Он приемлем при прогнозировании только на один период вперед. Его основные достоинства простота процедуры вычислений и возможность учета весов исходной информации. Рабочая формула метода экспоненциального сглаживания:

$$U_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot U_t,$$

где t – период, предшествующий прогнозируемому; $t+1$ – прогнозный период; U_{t+1} – прогнозируемый показатель; α - параметр сглаживания; U_t - фактическое значение исследуемого показателя за период, предшествующий прогнозируемому; U_t - экспоненциально взвешенная средняя для периода, предшествующего прогнозируемому.

При прогнозировании данным методом возникает два затруднения:

- выбор значения параметра сглаживания α ;
- определение начального значения U_0 .

От величины α зависит, как быстро снижается вес влияния предшествующих наблюдений. Чем больше α , тем меньше сказывается влияние предшествующих лет. Если значение α близко к единице, то это приводит к учету при прогнозе в основном влияния лишь последних наблюдений. Если значение α близко к нулю, то веса, по которым взвешиваются уровни временного ряда, убывают медленно, т.е. при прогнозе учитываются все (или почти все) прошлые наблюдения.

Таким образом, если есть уверенность, что начальные условия, на основании которых разрабатывается прогноз, достоверны, следует использовать небольшую величину параметра сглаживания ($\alpha \rightarrow 0$). Когда параметр сглаживания мал, то исследуемая функция ведет себя как средняя из большого числа прошлых уровней. Если нет достаточной уверенности в начальных условиях прогнозирования, то следует использовать большую величину α , что приведет к учету при прогнозе в основном влияния последних наблюдений.

Точного метода для выбора оптимальной величины параметра сглаживания α нет. В отдельных случаях автор данного метода профессор Браун предлагал определять

величину α , исходя из длины интервала сглаживания. При этом α вычисляется по формуле:

$$\alpha = \frac{2}{n+1},$$

где n – число наблюдений, входящих в интервал сглаживания.

Задача выбора U_0 (экспоненциально взвешенного среднего начального) решается следующими способами:

- если есть данные о развитии явления в прошлом, то можно воспользоваться средней арифметической и приравнять к ней U_0 ;
- если таких сведений нет, то в качестве U_0 используют исходное первое значение базы прогноза Y_1 .

Также можно воспользоваться экспертными оценками.

Отметим, что при изучении экономических временных рядов и прогнозировании экономических процессов метод экспоненциального сглаживания не всегда «срабатывает». Это обусловлено тем, что экономические временные ряды бывают слишком короткими (15-20 наблюдений), и в случае, когда темпы роста и прироста велики, данный метод не «успевает» отразить все изменения.

Задача. Имеются данные, характеризующие уровень безработицы в регионе, %

Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Постройте прогноз уровня безработицы в регионе на ноябрь, декабрь, январь месяцы, используя методы: скользящей средней, экспоненциального сглаживания, наименьших квадратов.
- Рассчитайте ошибки полученных прогнозов при использовании каждого метода.
- Сравните полученные результаты, сделайте выводы.

Решение методом экспоненциального сглаживания

1) Определяем значение параметра сглаживания по формуле:

$$\alpha = \frac{2}{n+1},$$

где n – число наблюдений, входящих в интервал сглаживания. $\alpha = 2 / (10+1) = 0,2$

2) Определяем начальное значение U_0 двумя способами:

I способ (средняя арифметическая) $U_0 = (2,99 + 2,66 + 2,63 + 2,56 + 2,40 + 2,22 + 1,97 + 1,72 + 1,56 + 1,42) / 10 = 22,13 / 10 = 2,21$

II способ (принимается первое значение базы прогноза) $U_0 = 2,99$

3) Рассчитываем экспоненциально взвешенную среднюю для каждого периода, используя формулу

$$U_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot U_t,$$

где t – период, предшествующий прогнозному; $t+1$ – прогнозный период; U_{t+1} – прогнозируемый показатель; α – параметр сглаживания; U_t – фактическое значение

исследуемого показателя за период, предшествующий прогнозному; U_t - экспоненциально взвешенная средняя для периода, предшествующего прогнозному.

Например:

$$U_{\text{фев}} = 2,99 \cdot 0,2 + (1-0,2) \cdot 2,21 = 2,37 \text{ (I способ)}$$

$$U_{\text{март}} = 2,66 \cdot 0,2 + (1-0,2) \cdot 2,37 = 2,43 \text{ (I способ) и т.д.}$$

$$U_{\text{фев}} = 2,99 \cdot 0,2 + (1-0,2) \cdot 2,99 = 2,99 \text{ (II способ)}$$

$$U_{\text{март}} = 2,66 \cdot 0,2 + (1-0,2) \cdot 2,99 = 2,92 \text{ (II способ)}$$

$$U_{\text{апр}} = 2,63 \cdot 0,2 + (1-0,2) \cdot 2,92 = 2,86 \text{ (II способ) и т.д.}$$

4) По этой же формуле вычисляем прогнозное значение

$$U_{\text{ноябрь}} = 1,42 \cdot 0,2 + (1-0,2) \cdot 2,08 = 1,95 \text{ (I способ)}$$

$$U_{\text{ноябрь}} = 1,42 \cdot 0,2 + (1-0,2) \cdot 2,18 = 2,03 \text{ (II способ)}$$

Результаты заносим в таблицу.

Месяцы	Уровень безработицы, U_t , %	Экспоненциально взвешенная средняя U_t		Расчет средней относительной ошибки, $\frac{ U_{\text{ф}} - U_{\text{р}} }{U_{\text{ф}}} \cdot 100, \%$	
		I способ	II способ	I способ	II способ
январь	2,99	2,21	2,99	26,09	0
февраль	2,66	2,37	2,99	10,90	12,41
март	2,63	2,43	2,92	7,60	11,03
апрель	2,56	2,47	2,86	3,52	11,72
май	2,40	2,49	2,80	3,75	16,67
июнь	2,22	2,47	2,72	11,26	22,52
июль	1,97	2,42	2,62	22,84	32,99
август	1,72	2,33	2,49	35,47	44,77
сентябрь	1,56	2,21	2,34	41,67	50
октябрь	1,42	2,08	2,18	46,48	53,52
Итого:				209,58	255,63
Прогноз ноябрь		1,95	2,03		

5) Рассчитываем среднюю относительную ошибку по формуле:

$$\varepsilon = \frac{1}{n} \cdot \sum_{i=1}^n \left[\frac{|y_{\text{ф}} - y_{\text{р}}|}{y_{\text{ф}}} \cdot 100 \right]$$

$$\varepsilon = 209,58/10 = 20,96\% \text{ (I способ)}$$

$$\varepsilon = 255,63/10 = 25,56\% \text{ (II способ)}$$

В каждом случае точность прогноза является удовлетворительной поскольку средняя относительная ошибка попадает в пределы 20-50%.

Лабораторная работа №4: Прогнозирование с применением метода экспоненциального сглаживания в Microsoft Excel

Цель: научиться строить прогнозы с применением метода экспоненциального сглаживания на программе для работы с электронными таблицами Microsoft Excel.

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

Кроме метода скользящего среднего для устранения колебаний в динамическом ряде используется метод экспоненциального сглаживания. Каждое сглаженное значение рассчитывается путём сочетания предыдущего сглаженного значения и текущего значения временного ряда. В этом случае текущее значение временного ряда взвешивается с учётом сглаживающей константы. Расчёт производится по формуле:

$$S_t = ay_t + (1-a) S_{t-1}$$

где S_t - значение экспоненциальной средней в момент времени t ;

a – параметр сглаживания, $a < \text{const}$, $0 < a < 1$;

y_t – текущее значение временного ряда;

S_{t-1} – предыдущее значение экспоненциальной средней

Порядок выполнения лабораторной работы

Пример: прогнозирование объемов продаж

Используя данные об объемах продаж на одном из предприятий химической промышленности за 1989-2003 годы (рисунок 1), необходимо выравнять (сгладить) данный временной ряд и сделать прогноз на следующий год.

Для этого в ячейку C2 просто переносим значение 1989 года, так как неизвестно предыдущее значение объема продаж. Введём в ячейку C3 следующую формулу:

$$=0,1 * B3 + 0,9 * C2 \text{ (Результат: 165,00).}$$

В диапазон ячеек C4:C16 скопируем данную формулу.

Так как значения параметра a могут изменяться от 0 до 1, примем изначально минимальное значение параметра сглаживания, равное 0,1. Однако при этом возникают определённые сложности. Основной недостаток состоит в том, что между изменениями в исходном ряду значений и соответствующими изменениями в ряду сглаженных значений отмечается лаг (или запаздывание). Так, мы видим, что анализируемые данные демонстрируют восходящий тренд объема продаж. Однако скользящие средние «медленно» обозначают тренд – все сглаженные значения за последние пять лет находятся под фактическими значениями объема продаж.

В целом, чем меньше значение a , тем менее оно чувствительно к изменениям тренда в данном временном ряду. Чтобы решить эту проблему, мы можем взять большее значение a . Рассмотрим, например, значение сглаживающей константы, равное $a=0,3$. На рисунке 2 в столбце C приведены сглаженные значения, рассчитанные по этой константе.

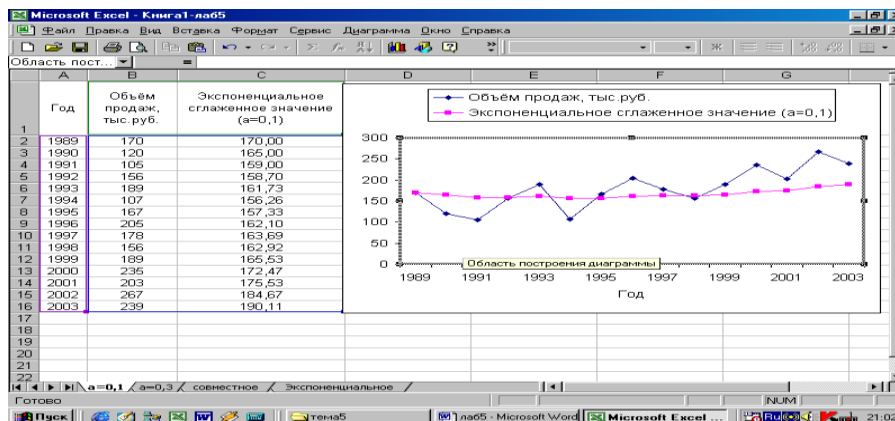


Рисунок 1 – При параметре сглаживания $a = 0,1$ сглаженные уровни за 1999-2003 гг. ниже фактических данных, что требует увеличения параметра a

Введём в ячейку C3 следующую формулу:

$$=0,3 * B3 + 0,7 * C2 \text{ (Результат: 155,00),}$$

и скопируем эту формулу в остальные ячейки столбца C.

При анализе расхождений результатов применения двух сглаживающих констант при выделении тренда следует обратить внимание на два момента.

Во-первых, временной лаг, который очевиден при $a = 0,1$, гораздо менее выражен при $a = 0,3$. В целом, чем больше значение при вычислении сглаженных значений, тем последние более чувствительны к изменениям в последних значениях временного ряда. То есть в этом случае сглаженные значения отстают от значений временного ряда не столь сильно, как это происходит при более малых значениях сглаживающей константы.

Этот фактор не играет ни какой роли, если отсутствует существенное изменение в общем тренде временного ряда. Однако он крайне важен при составлении прогнозов, когда отмечается значимое восхождение или нисхождение общего тренда временного ряда.

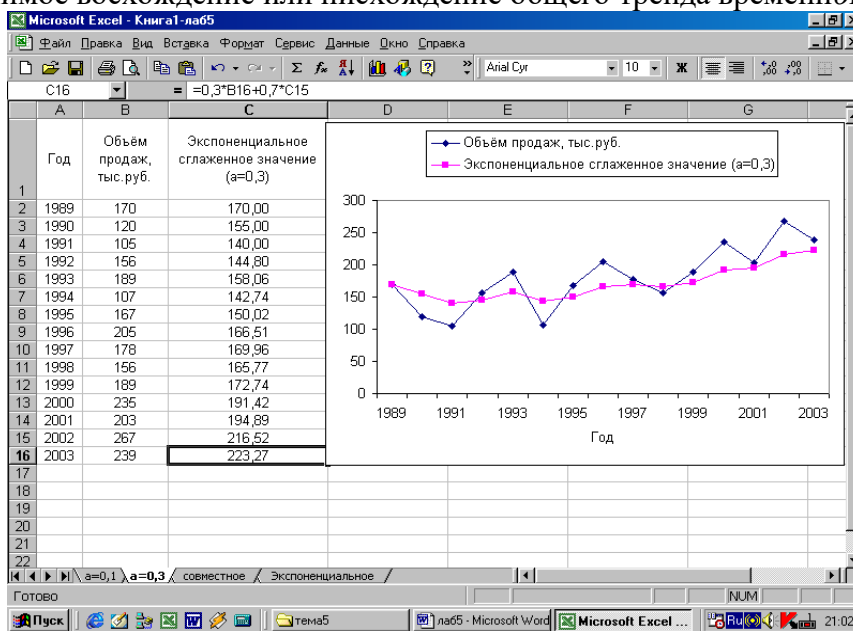


Рисунок 2 - Сглаженные уровни при параметре сглаживания $a=0,3$ лучше отражают общий тренд, то есть фактические данные, чем при $a=0,1$. Значения, полученные при $a=0,3$ лучше отражают общий тренд, чем те, которые рассчитаны при $a=0,1$.

Во-вторых, необходимо учитывать то, что при более низких значениях достигается большее сглаживание данных, а это позволяет выделять тренд с большей точностью. Ряд значений, полученных при сглаживающей константе $a=0,3$, при относительной сглаженности все же выказывает гораздо больше отклонений, чем ряд, полученный при $a=0,1$ (см. рисунок 3).

Для каждого конкретного случая придется выбирать наиболее приемлемое значение сглаживающей константы. Малое значение приводит к большему сглаживанию значений, а большое значение более точно отражает изменения тренда. В большинстве случаев значение сглаживающей константы лежит в пределах от 0,1 до 0,3, однако в ряде случаев возможно использование и других значений a , находящихся вне этого диапазона.

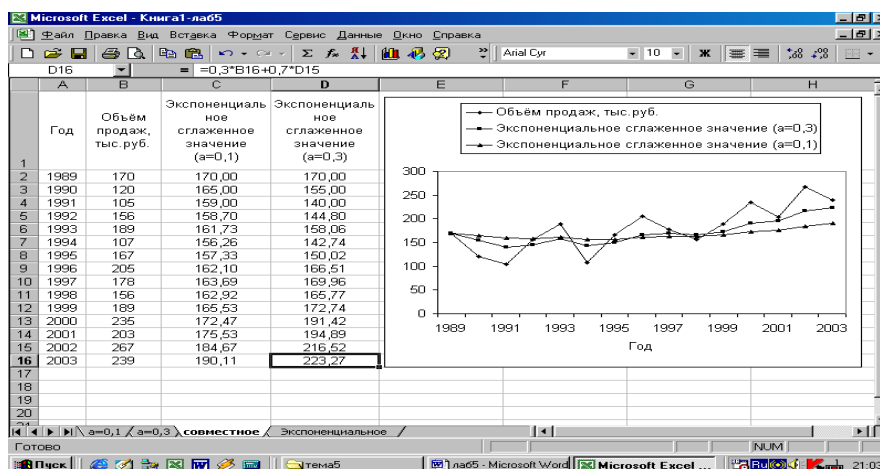


Рисунок 3 – Ряд значений, полученных при сглаживающей константе $a=0,3$ более точно характеризует изменение фактических данных, но менее сглажен, то есть сильнее отражает колебания по сравнению с рядом, полученным при сглаживающей константе $a= 0,1$

Методы прогнозирования под названием «сглаживание» учитывают эффекты выброса функций намного лучше, чем способы, использующие регрессионный анализ. Excel непосредственно поддерживает один из таких методов с помощью средства *Экспоненциальное сглаживание* в надстройке *Пакет анализа*.

Активизировать средство *Экспоненциальное сглаживание* можно выбрав команду *Сервис*⇒*Анализ данных* после загрузки надстройки *Пакет анализа*. Если *Пакет анализа* не установлен необходимо выбрать *Сервис*⇒*Надстройки*⇒*Пакет анализа*. Если и после этого *Пакет анализа* не работает следует переустановить Excel.

После того, когда активизировано средство *Экспоненциальное сглаживание*, необходимо заполнить диалоговое окно с одноименным названием так, как показано на рисунке 4.

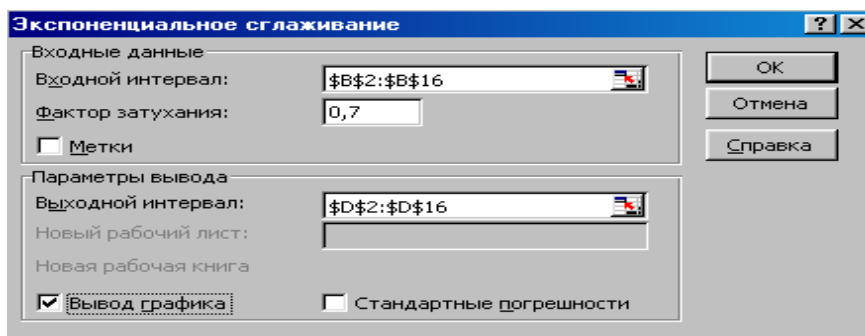


Рисунок 4 - Средство *Экспоненциальное сглаживание* не требует заполнения всех опций (например, не обязательными являются опции *Метки* и *Стандартные погрешности*)

Входные данные отражают диапазон ячеек с исходным динамическим рядом.

Фактор затухания – это показатель, рассчитываемый как разница между 1 и параметром сглаживания a . Для случая при сглаживающей константе $a= 0,3$ фактор затухания равен 0,7 (1-0,3). Следует избегать использования параметра *фактор затухания*, который меньше значения 0,7. Если у вас создается впечатление, что при большем значении константы сглаживания средство *Экспоненциальное сглаживание* действует значительно лучше, то, вероятнее всего, это происходит благодаря высокому уровню автокорреляции во временном ряду.

Выходной интервал отражает данные, рассчитанные при экспоненциальном сглаживании Excel.

Вывод графика позволяет автоматически выводить диаграмму в рабочий лист Excel.

Результаты применения средства *Экспоненциальное сглаживание* представлены на рисунке 5. Кроме того, на нем показано прогнозное значение на 2004 год, которое рассчитано на основе сглаженного экспоненциального тренда. При этом использовано фактическое значение за последний период времени (2003 год - ячейка B16) и последнее сглаженное значение (2003 год - ячейка D16).

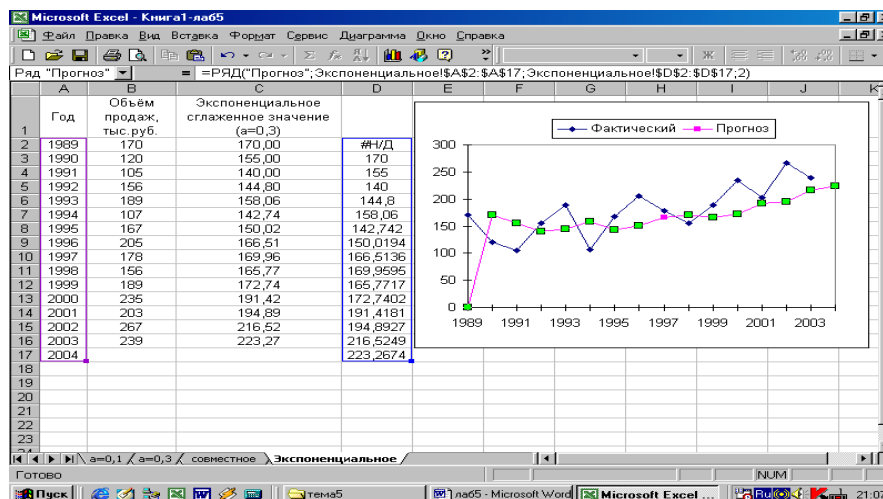


Рисунок 5 – Результаты применения средства *Экспоненциальное сглаживание*. Сравнивая значения ячеек в столбцах С и D, можно заметить, что их значения совпадают, но со сдвигом на одну ячейку вниз. Ячейка D2 содержит ошибку #Н/Д, вызванную тем, что отсутствует предыдущее фактическое значение ряда за 1988 год, которое необходимо для расчёта.

Форма отчета: По окончании выполнения лабораторных работ оформляется отчет в печатном виде в соответствии с требованиями ГОСТа в формате А4, шрифт 14, интервал 1,5. Для оформления отчета по каждой из лабораторных работ следует детально проработать предлагаемые методические рекомендации, внимательно изучить исходную информационную базу и провести необходимые расчеты.

Отчет по лабораторной работе должен включать:

1. Исходные данные и задание.
2. Расчётные формулы для проведения необходимых расчетов исходя из задания лабораторной работы.
3. Графическое представление полученных результатов расчета (при необходимости).
4. Результаты расчёта показателей, оформленные в виде таблиц.

Выводы: Выводы, сделанные в ходе проведенных расчетов должны быть четко и полно аргументированы.

Вопросы к защите лабораторной работы:

1. Какой процесс вы прогнозировали?
2. При помощи какого метода вы прогнозировали процесс? Его особенности.
3. Какие выводы можно сделать?

Лабораторная работа №5 Разработка прогноза с помощью методов экстраполяции. Прогноз уровня безработицы в республике Дагестан на основе метода наименьших квадратов

Цель: научиться строить прогнозы с применением метода наименьших квадратов.

Необходимый теоретический минимум:

Сущность *метода наименьших квадратов* состоит в минимизации суммы квадратических отклонений между наблюдаемыми и расчетными величинами. Расчетные величины находятся по подобранному уравнению – уравнению регрессии. Чем меньше расстояние между фактическими значениями и расчетными, тем более точен прогноз, построенный на основе уравнения регрессии.

Теоретический анализ сущности изучаемого явления, изменение которого отображается временным рядом, служит основой для выбора кривой. Иногда принимаются во внимание соображения о характере роста уровней ряда. Так, если рост выпуска продукции ожидается в арифметической прогрессии, то сглаживание производится по прямой. Если же оказывается, что рост идет в геометрической прогрессии, то сглаживание надо производить по показательной функции.

Рабочая формула метода наименьших квадратов:

$$Y_{t+1} = a \cdot X + b,$$

где $t + 1$ – прогнозный период; Y_{t+1} – прогнозируемый показатель; a и b – коэффициенты; X – условное обозначение времени.

Расчет коэффициентов a и b осуществляется по следующим формулам:

$$a = \frac{\sum_{i=1}^n (Y_{\phi} * X) - (\sum_{i=1}^n X * \sum_{i=1}^n Y_{\phi}) / n}{\sum_{i=1}^n X^2 - (\sum_{i=1}^n X)^2 / n}$$

$$b = \frac{\sum_{i=1}^n Y_{\phi}}{n} - a * \frac{\sum_{i=1}^n X}{n}$$

где, Y_{ϕ} – фактические значения ряда динамики; n – число уровней временного ряда;

Сглаживание временных рядов методом наименьших квадратов служит для отражения закономерности развития изучаемого явления. В аналитическом выражении тренда время рассматривается как независимая переменная, а уровни ряда выступают как функция этой независимой переменной.

Развитие явления зависит не от того, сколько лет прошло с отправного момента, а от того, какие факторы влияли на его развитие, в каком направлении и с какой интенсивностью. Отсюда ясно, что развитие явления во времени выступает как результат действия этих факторов.

Правильно установить тип кривой, тип аналитической зависимости от времени – одна из самых сложных задач предпрогнозного анализа.

Подбор вида функции, описывающей тренд, параметры которой определяются методом наименьших квадратов, производится в большинстве случаев эмпирически, путем построения ряда функций и сравнения их между собой по величине среднеквадратической ошибки, вычисляемой по формуле:

$$S = \sqrt{\frac{\sum_{i=1}^n (y_{\phi} - y_p)^2}{n - p - 1}},$$

где Y_{ϕ} – фактические значения ряда динамики; Y_p – расчетные (сглаженные) значения ряда динамики; n – число уровней временного ряда; p – число параметров, определяемых в формулах, описывающих тренд (тенденцию развития).

Недостатки метода наименьших квадратов:

- при попытке описать изучаемое экономическое явление с помощью математического уравнения, прогноз будет точен для небольшого периода времени и уравнение регрессии следует пересчитывать по мере поступления новой информации;
- сложность подбора уравнения регрессии, которая разрешима при использовании типовых компьютерных программ.

Задача. Имеются данные, характеризующие уровень безработицы в республике Дагестан, %

Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Постройте прогноз уровня безработицы в регионе на ноябрь, декабрь, январь месяца, используя методы: скользящей средней, экспоненциального сглаживания, наименьших квадратов.
- Рассчитайте ошибки полученных прогнозов при использовании каждого метода.
- Сравните полученные результаты, сделайте выводы.

Решение методом наименьших квадратов

Для решения составим таблицу, в которой будем производить необходимые расчеты:

Месяцы	Уровень безработицы, Уф, %.	Условное обозначение времени, X	Уф*Х	Х ²	Ур	Расчет средней относительной ошибки, $\frac{Уф - Ур}{Уф} * 100, \%$
1	2	3	4	5	6	7
январь	2,99	1	2,99	1	-0,17*1+3,15=2,98	$\frac{2,99-2,98}{2,99} * 100 = 0,33$
февраль	2,66	2	5,32	4	-0,17*2+3,15=2,81	5,64
март	2,63	3	7,89	9	-0,17*3+3,15=2,64	0,38
апрель	2,56	4	10,24	16	-0,17*4+3,15=2,47	3,52
май	2,40	5	12	25	-0,17*5+3,15=2,3	4,17
июнь	2,22	6	13,32	36	-0,17*6+3,15=2,13	4,05
июль	1,97	7	13,79	49	-0,17*7+3,15=1,96	0,51
август	1,72	8	13,76	64	-0,17*8+3,15=1,79	4,07
сентябрь	1,56	9	14,04	81	-0,17*9+3,15=1,62	3,85
октябрь	1,42	10	14,2	100	-0,17*10+3,15=1,45	2,11
Итого:	22,13	55	107,55	385	-	28,63
Прогноз ноябрь	1,28	11				
Прогноз декабрь	1,11	12				
Прогноз январь	0,94	13				

Определим условное обозначение времени как последовательную нумерацию периодов базы прогноза (графа 3). Рассчитаем графы 4 и 5. Расчетные значения ряда Ур определим по формуле $У_{t+1} = a * X + b$, где $t + 1$ – прогнозный период; $У_{t+1}$ – прогнозируемый показатель; а и b – коэффициенты; X - условное обозначение времени.

Коэффициенты а и b определим по следующим формулам:

$$a = \frac{\sum_{i=1}^n (Y\phi * X) - (\sum_{i=1}^n X * \sum_{i=1}^n Y\phi) / n}{\sum_{i=1}^n X^2 - (\sum_{i=1}^n X)^2 / n}$$

$$b = \frac{\sum_{i=1}^n Y\phi}{n} - \frac{a * \sum_{i=1}^n X}{n}$$

где, Уф – фактические значения ряда динамики; n – число уровней временного ряда.

$$a = [107,55 - (55 * 22,13) / 10] / [385 - 55^2 / 10] = -0,17$$

$$b = 22,13 / 10 - (-0,17) * 55 / 10 = 3,15$$

Далее определяем прогнозное значение:

$$У \text{ ноябрь} = -0,17 * 11 + 3,15 = 1,28$$

$$У \text{ декабрь} = -0,17 * 12 + 3,15 = 1,11$$

$$У \text{ январь} = -0,17 * 12 + 3,15 = 0,94$$

Рассчитываем среднюю относительную ошибку по формуле:

$$\varepsilon = \frac{1}{n} \cdot \sum_{i=1}^n \left[\frac{|y\phi - y_p|}{y\phi} \cdot 100 \right]$$

$$\varepsilon = 28,63 / 10 = 2,86\% < 10\% \text{ - точность прогноза высокая.}$$

Вывод:

Сравнивая результаты, полученные при расчетах методом скользящей средней, методом экспоненциального сглаживания и методом наименьших квадратов, можно сказать, что средняя относительная ошибка при расчетах методом экспоненциального сглаживания попадает в пределы 20-50%. Это значит, что точность прогноза в данном случае является лишь удовлетворительной.

В первом и третьем случае точность прогноза является высокой, поскольку средняя относительная ошибка менее 10%. Но метод скользящих средних позволил получить более достоверные результаты (прогноз на ноябрь – 1,52%, прогноз на декабрь – 1,53%, прогноз на январь – 1,49%), так как средняя относительная ошибка при использовании этого метода наименьшая – 1,13%.

Лабораторная работа № 6 Процедура графического прогнозирования в Microsoft Excel.

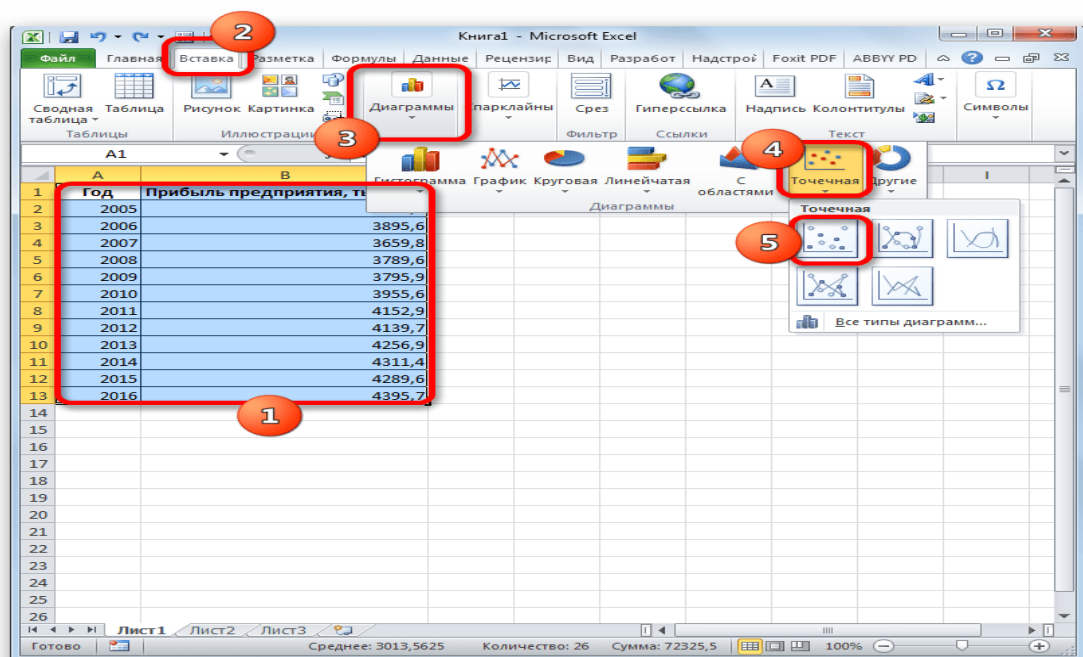
Построение линии тренда

Цель: Построение линии тренда в Microsoft Excel.

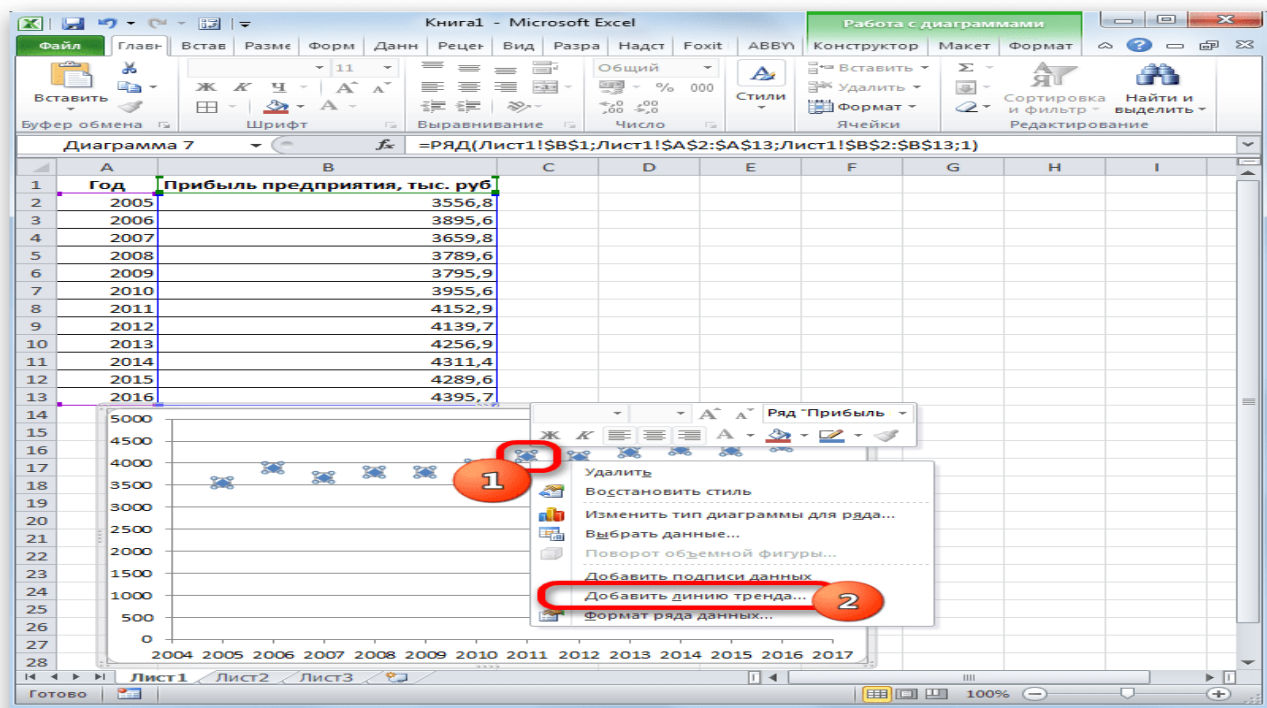
Оборудование: ПК, оснащенный необходимым программным обеспечением.

Задача: Попробуем предсказать сумму прибыли предприятия через 3 года на основе данных по этому показателю за предыдущие 12 лет.

1. Строим график зависимости на основе табличных данных, состоящих из аргументов и значений функции. Для этого выделяем табличную область, а затем, находясь во вкладке «Вставка», кликаем по значку нужного вида диаграммы, который находится в блоке «Диаграммы». Затем выбираем подходящий для конкретной ситуации тип. Лучше всего выбрать точечную диаграмму. Можно выбрать и другой вид, но тогда, чтобы данные отображались корректно, придется выполнить редактирование, в частности убрать линию аргумента и выбрать другую шкалу горизонтальной оси.



2. Теперь нам нужно построить линию тренда. Делаем щелчок правой кнопкой мыши по любой из точек диаграммы. В активированном контекстном меню останавливаем выбор на пункте «Добавить линию тренда».



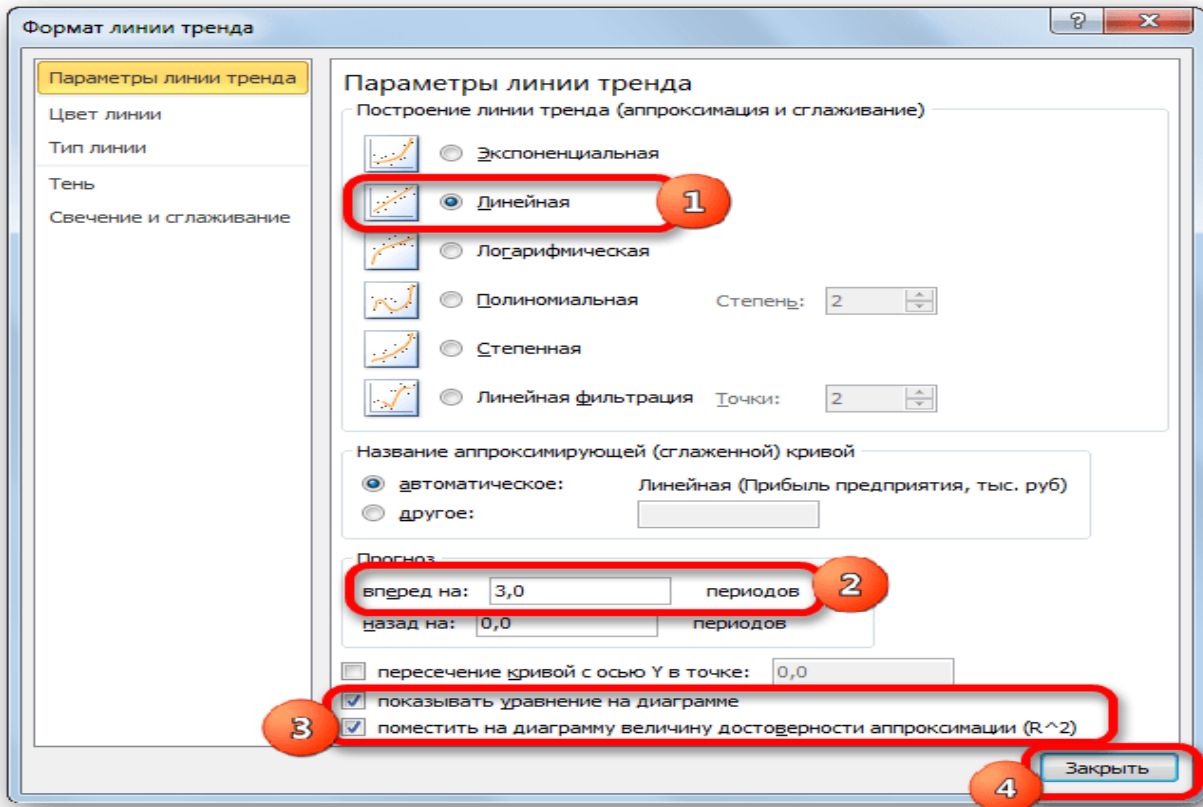
3. Открывается окно форматирования линии тренда. В нем можно выбрать один из шести видов аппроксимации:

- **Линейная;**
- **Логарифмическая;**
- **Экспоненциальная;**
- **Степенная;**
- **Полиномиальная;**
- **Линейная фильтрация.**

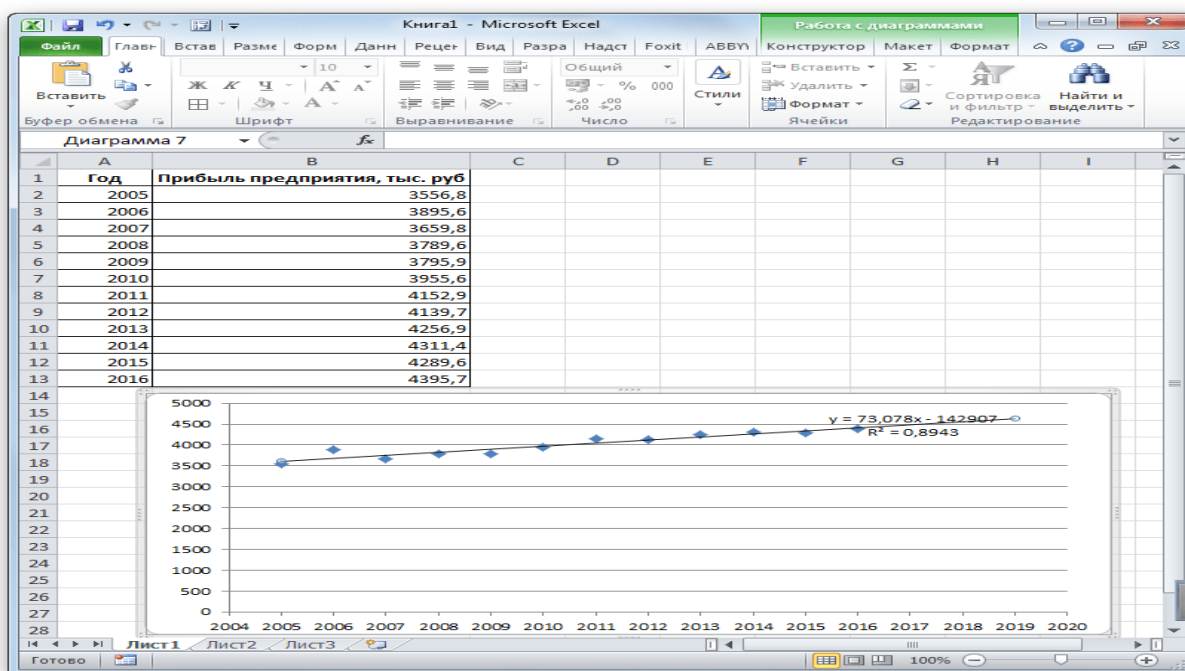
Давайте для начала выберем линейную аппроксимацию.

В блоке настроек «Прогноз» в поле «Вперед на» устанавливаем число «3,0», так как нам нужно составить прогноз на три года вперед. Кроме того, можно установить галочки около настроек «Показывать уравнение на диаграмме» и «Поместить на диаграмме величину достоверности аппроксимации (R^2)». Последний показатель отображает качество линии

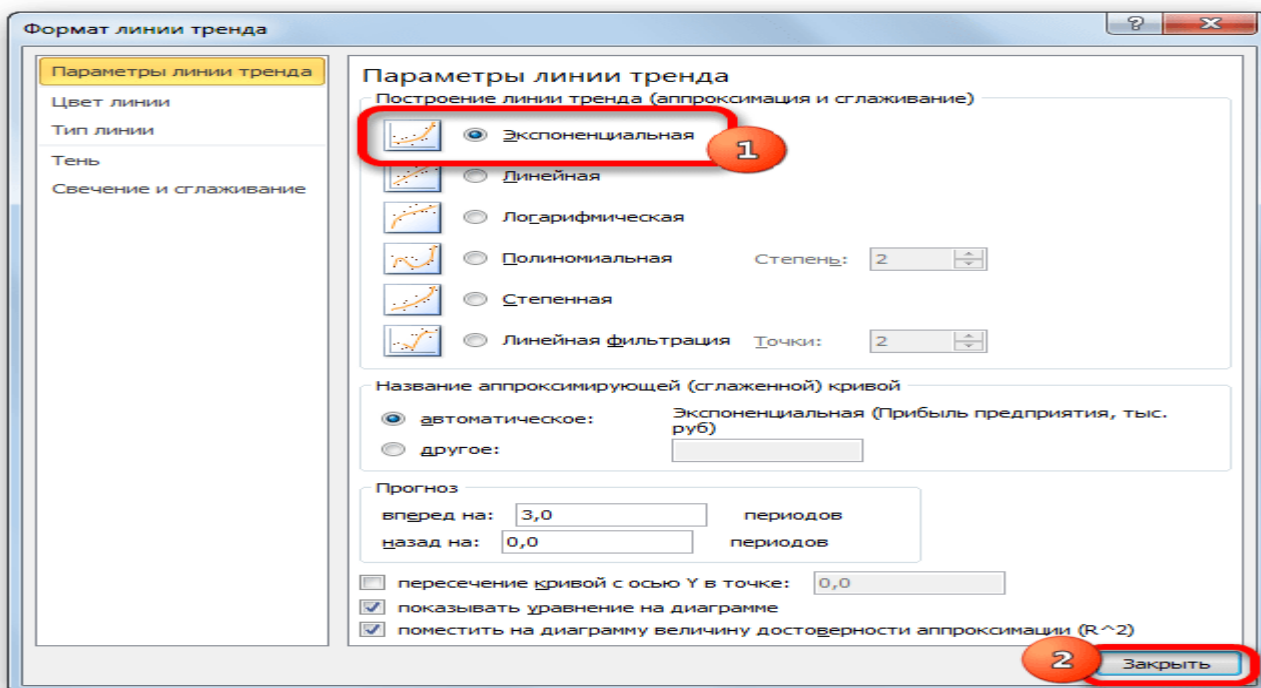
тренда. После того, как настройки произведены, ждем на кнопку «Закреть».



4. Линия тренда построена и по ней мы можем определить примерную величину прибыли через три года. Как видим, к тому времени она должна перевалить за 4500 тыс. рублей. Коэффициент R^2 , как уже было сказано выше, отображает качество линии тренда. В нашем случае величина R^2 составляет 0,89. Чем выше коэффициент, тем выше достоверность линии. Максимальная величина его может быть равной 1. Принято считать, что при коэффициенте выше 0,85 линия тренда является достоверной.



5. Если же вас не устраивает уровень достоверности, то можно вернуться в окно формата линии тренда и выбрать любой другой тип аппроксимации. Можно перепробовать все доступные варианты, чтобы найти наиболее точный.



Нужно заметить, что эффективным прогноз с помощью экстраполяции через линию тренда может быть, если период прогнозирования не превышает 30% от анализируемой базы периодов. То есть, при анализе периода в 12 лет мы не можем составить эффективный прогноз более чем на 3-4 года. Но даже в этом случае он будет относительно достоверным, если за это время не будет никаких форс-мажоров или наоборот чрезвычайно благоприятных обстоятельств, которых не было в предыдущих периодах.

Лабораторная работа № 7: Прогнозирование с помощью функций регрессии Excel

Цель: научиться строить прогнозы с помощью функций регрессии Excel

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

Простое скользящее среднее является быстрым, но довольно неточным способом выявления общих тенденций временного ряда. Если вы достаточно разобрались с примерами составления прогнозов на основе скользящего среднего, то, очевидно, обратили внимание на то, что они не дают прогноза выходящего за пределы, в которых данные уже известны. Передвинуть границу оценки в будущее по временной оси можно с помощью одной из функций регрессии Excel.

Каждый из методов регрессии оценивает взаимосвязь между фактическими данными наблюдений и другими параметрами, которые зачастую являются показателями того, когда были сделаны эти наблюдения. Это могут быть как числовые значения каждого результата наблюдения во временном ряду, так и дата наблюдения.

Составление линейных прогнозов: функция тенденция

Использование функции рабочего листа ТЕНДЕНЦИЯ — это самый простой способ регрессионного анализа. Предположим, результаты ваших наблюдений внесены в ячейки A2:A11, а дни месяца расположены в ячейках B2:B11, как на рисунке 1.

Выделите ячейки C2:C11 и введите следующую формулу, используя формулу массива: = ТЕНДЕНЦИЯ (A2:A11;B2:B11).

Для этого необходимо:

- 1) Запустить Мастер функций.
- 2) В появившемся диалоговом окне выбрать функции Статистические⇒Тенденция.
- 3) Вести нужные диапазоны ячеек.
- 4) Когда рассчиталось одно значение из выделенного диапазона необходимо щёлкнуть мышью на формульную строку (после знака равенства).
- 5) Для ввода формулы массива нажмите комбинацию клавиш <Ctrl+Shift+Enter>

Рассматривая данный метод прогнозирования, следует обратить особое внимание на следующие моменты.

1. Каждый результат в ячейках C2:C11 получается на основе одной и той же формулы массива, внутри которой «спрятано» более сложное выражение – линейное уравнение регрессии. В данном случае формула имеет следующий вид:

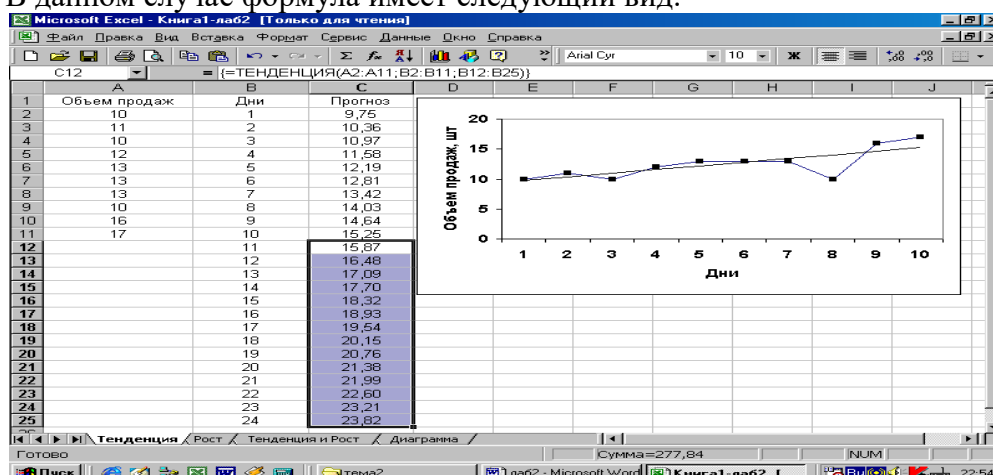


Рисунок 1 - С помощью функции ТЕНДЕНЦИЯ можно прогнозировать базовую линию (линейный тренд) результатов наблюдений, а также значения показателя, выходящие за пределы конца базовой линии

Ячейка C2: = 9,13 + 0,61*1

Ячейка C3: = 9,13 + 0,61*2

Ячейка C4: = 9,13 + 0,61*3 и так далее.

Значение 9,13 представляет собой длину отрезка, отсекаемого на оси ординат линией прогноза, т.е. значение прогноза в начальный момент. Это свободный член уравнения регрессии. Значение 0,61 равно угловому коэффициенту линии прогноза, другими словами, значения прогноза изменяются в результате изменений дат проведения наблюдений. Этот параметр уравнения регрессии носит название коэффициента регрессии. Знак перед коэффициентом регрессии говорит о направлении связи: прямая или обратная. В данном случае с увеличением времени от 1 до 10 дней ежедневный средний прирост объемов продаж составляет 0,61 – связь прямая.

2. Поскольку все значения прогноза составляются на основе одних и тех же показателей отрезка, отсекаемого на оси ординат, и углового коэффициента, прогноз не отражает происходящих изменений во временном ряду.

Например, данные ряда резко изменяются между восьмым (10) и девятым результатами наблюдений (16). Это изменение влияет на все значения прогноза, даже значение прогноза второго временного отрезка (11), хотя и располагается на шесть результатов наблюдений раньше, чем это изменение фактически произошло.

3. В данном примере функция ТЕНДЕНЦИЯ вычисляет прогноз, основанный на связи между фактическими результатами наблюдений и числами 1-10, которые могут отражать либо первых десять дней месяца, либо первых десять месяцев года. Excel выражает первый аргумент как аргумент изв_знач_у (известные значения у) функции ТЕНДЕНЦИЯ, а второй — как аргумент функции изв_знач_x (известные значения x).

Регрессионный анализ позволяет производить перспективную оценку более удаленного будущего. Однако регрессионный прогноз распространяется за пределы данных самого последнего фактического наблюдения. На практике желательно составить прогноз хотя бы на первый, следующий за этим, период временного ряда (т.е. на тот, для которого еще нет результатов наблюдения).

Введите в ячейку B12 число 11, а в ячейку C12 — следующее:

= ТЕНДЕНЦИЯ (A2:A11;B2:B11;B12)

Первый аргумент - A2:A11 - определяет данные наблюдений базовой линии (изв_знач_у); второй аргумент - B2:B11 - определяет временные моменты, в которые эти данные были получены (изв_знач_x). Значение 11 в ячейке B12 является нов_знач_x (новым значением x) и определяет время, которое связывается с перспективной оценкой.

Формула фактически говорит о следующем: «Если известно, каким образом у-значения в диапазоне A2:A11 соотносятся с x-значениями в диапазоне B2:B11, то какой результат у-значения мы получим, зная новое x-значение временного момента, равное 11?». Полученное значение 15,87 является прогнозом на основе фактических данных на пока еще не наступивший одиннадцатый временной отсчет.

Введя в ячейку B12 большее значение, вы сможете прогнозировать данные более позднего временного момента, чем непосредственно следующий за текущим. Предположим, что результаты наблюдений в ячейках A2:A11 отражают ежемесячные объемы продаж с января по октябрь 2002 года. В этом случае число 24, введенное в ячейку B12, будет определять 24 месяц, т.е. декабрь 2003 года. Выполняя вычисления с помощью функции ТЕНДЕНЦИЯ, получаем результат 23,82, который и будет отражать прогнозируемый объем продаж на декабрь 2003 года, полученный на основе фактических результатов наблюдений за период с января по октябрь 2002 года.

Кроме того, существует возможность одновременного прогнозирования данных для нескольких новых временных моментов. Например, введите числа 11-24 в ячейки B12:B25, а затем выделите ячейки C12:C25 и введите с помощью формулы массива следующее:

= ТЕНДЕНЦИЯ (A2:A11;B2:B11;B12:B25)

Excel вернет в ячейки C12:C25 прогноз на временные моменты с 11 по 24. Данный прогноз будет базироваться на связи между данными наблюдений базовой линии диапазона A2:A11 и временными моментами базовой линии с 1 по 10, указанными в ячейках B2:B11.

Составление нелинейного прогноза: функция РОСТ

Функция ТЕНДЕНЦИЯ вычисляет прогнозы, основанные на линейной связи между результатом наблюдения и временем, в которое это наблюдение было зафиксировано. Предположим, что вы составляете линейный график данных, на вертикальной оси которого отмечаете результаты наблюдений, а на горизонтальной фиксируете временные моменты их получения. Если эта взаимосвязь носит линейный характер, то линия на графике будет либо прямой, либо слегка наклоненной в одну или другую сторону, либо горизонтальной. Это и будет лучшей подсказкой о том, что взаимосвязь является линейной, и поэтому в данном случае функция ТЕНДЕНЦИЯ — самый удобный способ регрессионного анализа.

Однако, если линия резко изгибается в одном из направлений, то это означает, что взаимосвязь показателей носит нелинейный характер. Существует большое количество типов данных, которые изменяются во времени нелинейным способом. Некоторыми примерами таких данных являются объем продаж новой продукции, прирост населения, выплаты по основному кредиту. В случае нелинейной взаимосвязи функция Excel РОСТ поможет вам подучить более точную картину направления развития вашего бизнеса, чем при использовании функции ТЕНДЕНЦИЯ.

Пример: книжная торговля

Представим, что менеджер по закупкам отдела «Книга-почтой» недавно разослал клиентам новый каталог, рекламирующий роман, получивший очень высокую оценку критиков. Менеджер считает, что следует заранее заказать дополнительное количество экземпляров, чтобы не оказаться в ситуации, когда книга закончится раньше, чем перестанут приходить заявки на нее. Менеджер начал отслеживать ежедневные заказы на роман, и регистрировать объемы продаж, как это показано на рисунке 2.

Здесь демонстрируется, каким образом фактические и прогнозируемые данные фиксируются на стандартном линейном графике. Поскольку линия имеющихся в наличии товаров резко изгибается вверх, менеджер принимает решение составить прогноз с использованием функции РОСТ. Как и при использовании функции ТЕНДЕНЦИЯ, пользователь в данном случае может генерировать прогнозы, просто подставляя нов_знач_x. Чтобы спрогнозировать результаты 11-13 недель, следует ввести эти числа в ячейки B12:B14, а затем с помощью формулы массива в диапазон ячеек C12:C14 ввести следующее:

= РОСТ(A2:A11;B2:B11;B12:B14)

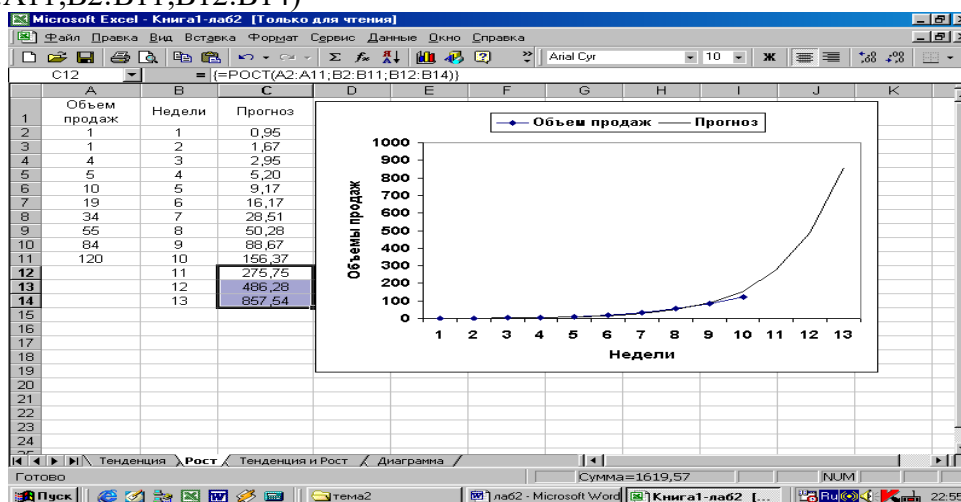


Рисунок 2 - Функция РОСТ очень удобна при прогнозировании нелинейных базовых линий (нелинейных трендов)

В ячейках C12:C14 приведены значения предварительной оценки количества заказов, которое может ожидать менеджер в последующие три недели при условии, что текущая тенденция роста останется неизменной. Однако следует учитывать, что такой оптимистичный

прогноз на практике, вероятно, претерпит определенные изменения. Если при вычислении прогноза количество планируемых заказов превысит количество клиентов, от него, скорее всего, следует просто отказаться.

А что было бы, если бы в вышеприведенном примере вместо функции РОСТ использовалась функция ТЕНДЕНЦИЯ? В этом случае, поскольку аргумент `изв_знач_x` носит линейный характер, функция ТЕНДЕНЦИЯ выдаст линейные значения.

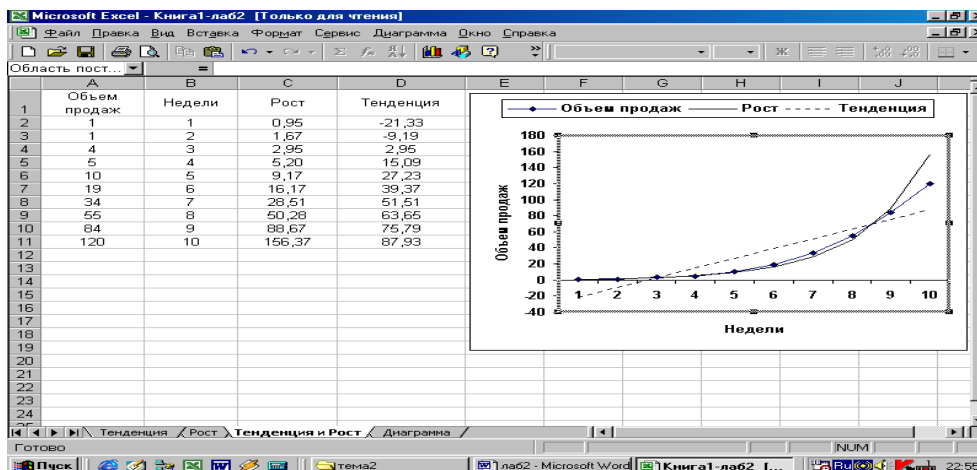


Рисунок 3 - Линия тренда, построенная с помощью функции РОСТ, дает на основе нелинейной базовой линии намного более точный прогноз, чем линия тренда, построенная с помощью функции тенденция

Из рисунка 3 видно, что ряд ТЕНДЕНЦИЯ в столбце D описывает прямую графика; кривая РОСТ в столбце С намного точнее отражает тенденцию первых десяти результатов наблюдений, чем линия ТЕНДЕНЦИЯ.

На практике бывает лучше воспользоваться не функцией РОСТ, а функцией ТЕНДЕНЦИЯ, даже когда базовая линия носит нелинейный характер.

Регрессионный анализ с помощью диаграмм

Иногда возникает необходимость провести регрессионный анализ непосредственно на графике, без введения в рабочий лист значений для прогноза. Это можно сделать с помощью графической линии тренда методом, во многом сходным с методом получения прогноза с применением скользящего среднего на основе графика.

Постройте диаграмму на основе данных, содержащихся в ячейках A2:A11 (рисунок 4). Щелкнув мышью на диаграмме, вы получите возможность ее редактировать. Щелкните на ряде нужных данных для его выбора.

После этого выполните следующие шаги.

1. Выберите команду Диаграмма⇒Добавить линию тренда.
2. Выберите тип линии тренда Линейная.
3. Щелкните на корешке вкладки Параметры.
4. В поле Прогноз⇒Вперед_на введите количество желаемых периодов, на протяжении которых линия тренда будет проложена вперед.

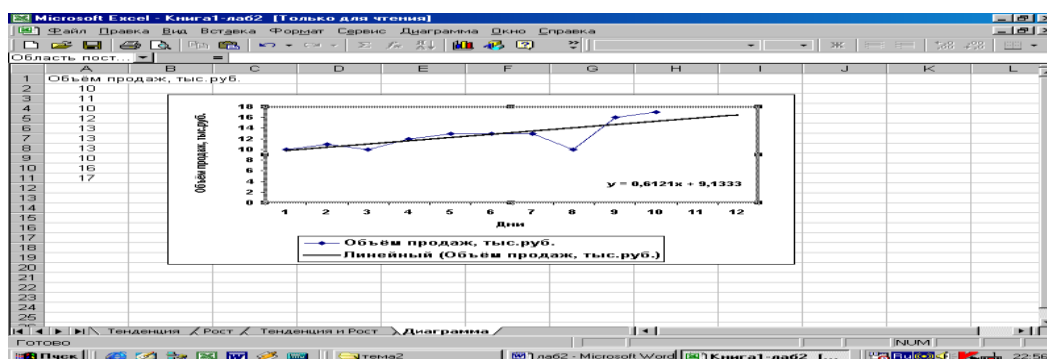


Рисунок 4. С помощью линий тренда можно создавать прогнозы, основанные на регрессии, непосредственно на диаграмме

5. При желании, можете установить флажок опции Показывать уравнение на диаграмме. В результате уравнение для прогноза разместится на графике в виде текста. Excel может расположить уравнение таким образом, что оно перекроет некоторые данные графика или линии тренда (либо, частично, само уравнение). В этом случае выделите уравнение, щелкнув на нем мышью, а затем перетащите его в другое, более удобное место.

6. Щелкните на кнопке ОК.

В отличие от линии тренда Скользящее среднее, с помощью линии тренда Линейная можно вернуть значения прогноза и, если специально указано, показать их на диаграмме.

Форма отчета: По окончании выполнения лабораторных работ оформляется отчет в печатном виде в соответствии с требованиями ГОСТа в формате А4, шрифт 14, интервал 1,5. Для оформления отчета по каждой из лабораторных работ следует детально проработать предлагаемые методические рекомендации, внимательно изучить исходную информационную базу и провести необходимые расчеты.

Отчет по лабораторной работе должен включать:

1. Исходные данные и задание.
2. Расчётные формулы для проведения необходимых расчетов исходя из задания лабораторной работы.
3. Графическое представление полученных результатов расчета (при необходимости).
4. Результаты расчёта показателей, оформленные в виде таблиц.
5. Основные выводы.

Выводы: Выводы, сделанные в ходе проведенных расчетов должны быть четко и полно аргументированы.

Вопросы к защите лабораторной работы:

1. Какой процесс вы прогнозировали?
2. При помощи какого метода вы прогнозировали процесс? Его особенности.
3. Какие выводы можно сделать выводы?

Лабораторная работа №8: Прогнозирование на основе анализа трендов и сезонности

Цель: научиться строить прогнозы на основе анализа трендов и сезонности на программе для работы с электронными таблицами Microsoft Excel.

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

Анализ трендов и сезонности представляет собой непосредственный, интуитивный подход к оцениванию четырех базовых компонентов помесечных или поквартальных временных рядов: долгосрочный тренд (тенденция), сезонность, циклическая вариация и нерегулярный компонент. Базовая модель временного ряда представляет числа в этом ряде в виде произведения, получаемого путем умножения перечисленных компонентов (мультипликативная модель).

1. Долгосрочный тренд (тенденция) указывает долгосрочное поведение временного ряда — как правило, в виде прямой линии или экспоненциальной кривой. Это бывает полезно в случае, если требуется увидеть картину в целом.

2. Точно повторяющийся сезонный компонент определяет влияние времени года. Например, потребность в обогревателях высока в зимние месяцы, а объемы продаж сельскохозяйственной продукции высоки в период сбора урожая.

3. Среднесрочный циклический компонент состоит из последовательных повышений и понижений, которые не повторяются каждый год и поэтому исключаются из сезонного компонента.

4. Краткосрочный нерегулярный (случайный) компонент представляет остаточную вариацию, которую невозможно объяснить. В нем проявляется действие тех однократных событий, которые происходят с течением времени случайно, а не систематически.

Эти четыре базовых компонента временного ряда можно оценивать с помощью метода, который называется «отношением к скользящему среднему».

Порядок выполнения лабораторной работы

Пример: продажа автомобилей

На рисунке 1 представлены поквартальные объемы продаж автомобилей компании А. Этот временной ряд демонстрирует ярко выраженные сезонные колебания. Объемы продаж, как правило, достигают пика во втором квартале, о чем свидетельствует график временного ряда для соответствующих данных. Затем они в целом нарастают в течение последующих трех кварталов. Поскольку этот сезонный сценарий не повторяется в точности каждый год, рассматриваемый временной ряд характеризуется также некоторой циклическостью и нерегулярностью поведения. Советуем обратить внимание и на долгосрочную тенденцию, выражающуюся в общем росте продаж с течением времени.

Тренд и циклический компонент: скользящее среднее

Наша цель заключается в том, чтобы выделить четыре базовых компонента временного ряда. Начнем с усреднения данных за год, чтобы избавиться от сезонного компонента и уменьшить нерегулярный компонент.

Скользящее среднее представляет собой новый ряд, полученный путем усреднения соседних наблюдений временного ряда и перехода к следующему периоду времени — в итоге получается более гладкий ряд.

Выполняя усреднение данных за целый год, мы приходим к тому, что вклад сезонных компонентов — независимо от времени года — остается практически одинаковым.

Найти скользящее среднее значение на рисунке 2 для поквартальных данных за третий квартал 1996 г. можно следующим образом. В ячейку D4 введите следующую формулу:
$$= (1/2 * C2 + C3 + C4 + C5 + 1/2 * C6) / 4$$
 (Результат: 18453).

Взвешивая крайние точки коэффициентом 1/2, вы гарантируете, что этот квартал учтен в скользящем среднем точно так же, как и другие кварталы. Такое взвешенное среднее необходимо для того, чтобы интервал по обе стороны от базового периода времени был симметричным и вместе с тем охватывал в точности данные за один год. Скользящее среднее отсутствует для первых двух и последних двух кварталов ряда. В диапазон ячеек D5:D27 скопируйте формулу из ячейки D4, протянув мышью вниз и удерживая нажатой её левую кнопку.

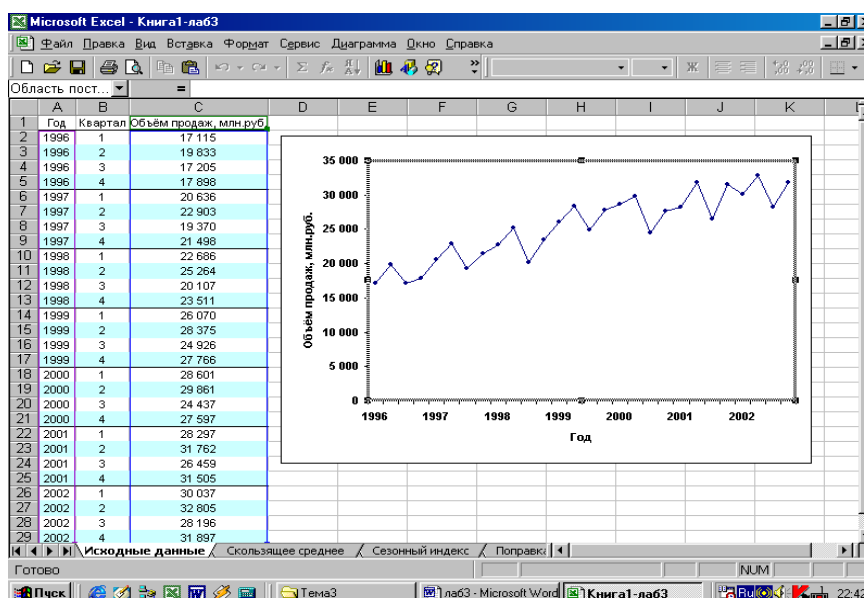


Рисунок 1 - График временного ряда поквартальных продаж автомобилей компании А.

Чтобы выделить сезонное поведение, следует получить отношение исходных значений к скользящему среднему. Полученный результат будет включать сезонный и нерегулярный компоненты, поскольку скользящее среднее исключает из данных тренд и циклический компонент.

Затем, чтобы устранить нерегулярный компонент, вы усредняете эти значения для каждого сезона. Сезонный компонент проявляется, поскольку он присутствует ежегодно, тогда как нерегулярный компонент, как правило, удается усреднить. Конечные результаты включают сезонный индекс для каждого времени года — фактор, который указывает, насколько большим (или меньшим) бывает рассматриваемый показатель в этот конкретный период времени в сравнении с типичным периодом на протяжении года.

Например, сезонный индекс за первый квартал, равный 1,3, свидетельствует о том что рассматриваемый показатель в первом квартале, как правило, на 30% больше, чем в типичном квартале. С другой стороны, сезонный индекс за третий квартал, равный 0,74, свидетельствует о том, что рассматриваемый показатель в третьем квартале, как правило, на 26% ниже, чем в типичном квартале.

Первое значение отношения к скользящему среднему за третий квартал 1996г. содержится в ячейке E4 (рисунок 3), рассчитывается по формуле:

=C4/D4 (Результат: 0,9324).

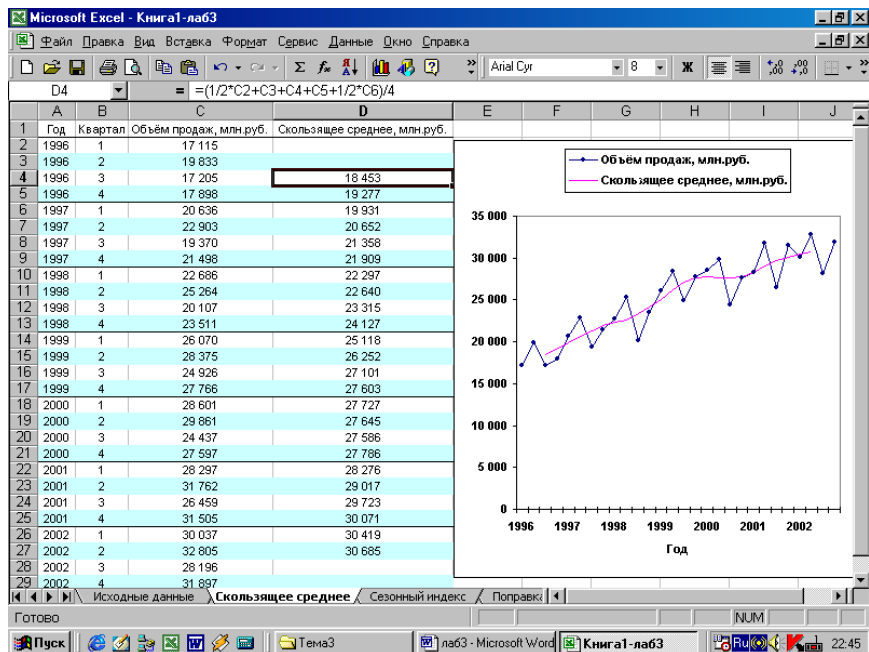


Рисунок 2 - Скользящее среднее объемов продаж автомобилей компании А.

Из графика удалось устранить сезонные и нерегулярные колебания объемов продаж; остались лишь тренд и циклический компонент.

Сезонный индекс: отношения к скользящему среднему отражают сезонное поведение. Сезонный индекс за третий квартал определяется путем усреднения значений этих отношений за третий квартал по всем рассматриваемым годам.

Для этого в ячейку F4 введите формулу: $=CPЗНАЧ(E4;E8;E12;E16;E20;E24)$ (Результат: 0,8996).

Аналогично можно рассчитать сезонный индекс за четвертый (в ячейке F5), первый (в ячейке F6) и второй (в ячейке F7) кварталы. После того как вычислен каждый сезонный индекс, его можно использовать везде для расчёта данных столбца F - даже там, где нельзя вычислить скользящее среднее (F2, F3, F28, F29) поскольку, по определению, сезонные колебания в точности повторяются каждый год.

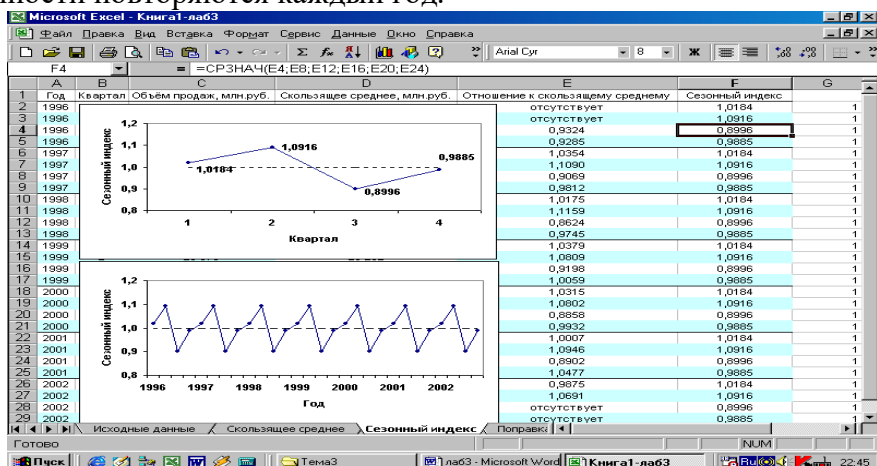


Рисунок 3 - Сезонные индексы показывают, что объемы продаж достигают пика во втором квартале, падают до минимума в третьем квартале, а затем снова повышаются вплоть до следующего второго квартала (верхняя диаграмма). Сезонный компонент объемов продаж, извлеченный из исходного ряда, в точности повторяется каждый год (нижняя диаграмма).

Поправка на сезон: деление ряда на сезонный индекс

Что означают слова «с поправкой на сезонные колебания» и может ли наблюдаться снижение показателя, вычисленного с поправкой на сезонные колебания, даже в том случае, когда его фактическое значение повышается?

Поправка на сезонные колебания устраняет из результатов измерения ожидаемый сезонный компонент, что позволяет нам непосредственно сравнивать один квартал с другим, выявляя, таким образом, те или иные скрытые тенденции.

Чтобы найти некоторое значение с поправкой на сезонные колебания, достаточно разделить исходные данные на сезонный индекс для соответствующего квартала. Например, объем продажи автомобилей во втором квартале 2002 г. — с поправкой на сезонные колебания — вычисляется как фактический объем продажи (32 805 млн.руб.), деленный на сезонный индекс второго квартала (1,0916).

Введите в ячейку E27 формулу:
 $=C27/D27$ (Результат: 30 052).

Аналогично рассчитываются другие ячейки столбца E на рисунке 4. Почему результат с поправкой на сезонные колебания оказался меньше фактического объема продаж? Дело в том, что объем продажи во втором квартале, как правило, выше по сравнению с типичным кварталом года.

В сущности, вы заранее рассчитываете на то, что объем продажи во втором квартале будет примерно на 9,2% выше (исходя из сезонного индекса, равного 1,0916). Деление на сезонный индекс нивелирует влияние этой ожидаемой сезонной флуктуации, приводя объем продажи во втором квартале в соответствие с типичным кварталом года (т.е. снижая его).

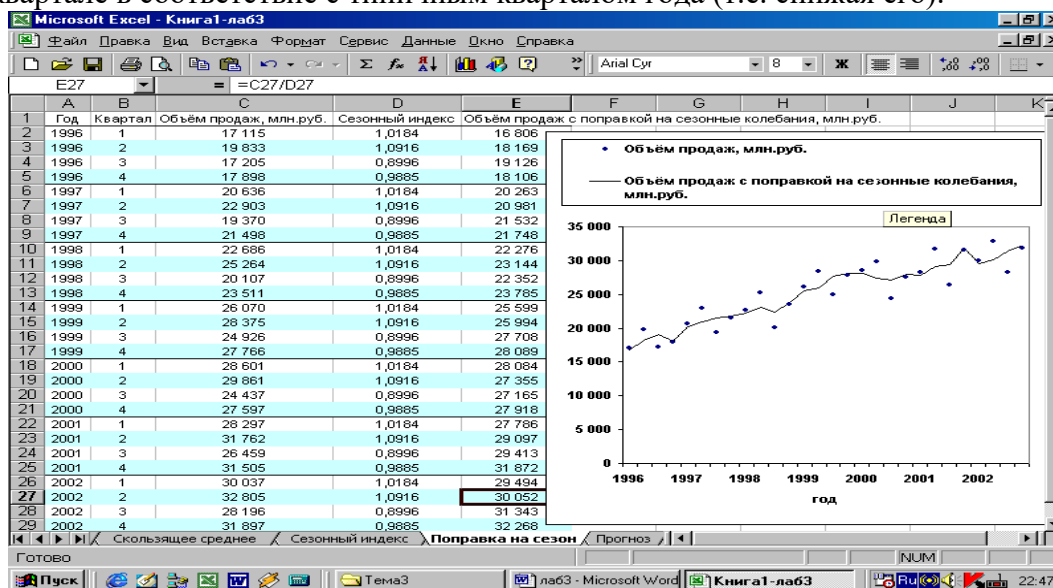


Рисунок 4 - Ряд, отражающий объемы продаж с поправкой на сезонные колебания, позволяет сравнивать один квартал с другим. Избавляясь от ожидаемых сезонных флуктуаций, мы получаем более четкую картину развития

На рисунке 4 отражены объемы продажи с поправкой на сезонные колебания для всего временного ряда. В графическом виде эти данные представлены вместе с исходными данными. Ряд, в котором учитывается поправка на сезонные колебания, оказывается несколько более гладким, чем исходные данные, поскольку в первом случае нам удалось избавиться от сезонных флуктуаций.

Однако и в этом случае остаются немалые «шероховатости», поскольку, помимо тренда, в нем по-прежнему присутствуют нерегулярный и циклический компоненты.

Долгосрочный тренд и прогноз с поправкой на сезонные колебания

Когда временной ряд демонстрирует долгосрочную линейную тенденцию к нарастанию или снижению, для оценки этой тенденции и прогнозирования будущего можно воспользоваться регрессионным анализом. Для прогнозирования ряда, в котором учитывается

поправка на сезонные колебания (переменная Y), используется период времени (переменная X). Результирующее уравнение регрессии будет представлять долгосрочный тренд. Подставляя будущие временные периоды в качестве новых значений X, вы получите возможность экстраполировать эту долгосрочную тенденцию на будущее.

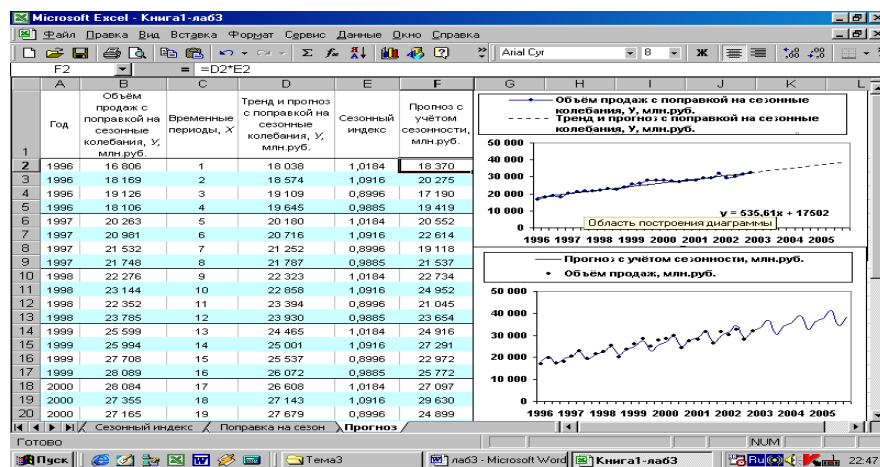


Рисунок 5 - Построенная методом наименьших квадратов линия регрессии используется для прогнозирования на основании временного ряда объемов продажи с поправкой на сезонные колебания. Эту линию можно продолжить вправо, чтобы получить прогнозы с поправкой на сезонные колебания (верхняя диаграмма). Прогнозирование может выполняться путем умножения линии тренда на сезонный индекс. Полученный результат включает тренд и сезонный компонент, но не циклическое и нерегулярное поведение ряда (нижняя диаграмма).

Этого можно добиться, воспользовавшись числами 1, 2, 3, ... для представления X непосредственно в виде номера временного периода (квартала или месяца). На рисунке 5 в столбце В и С представлены данные для регрессионного анализа, цель которого заключается в выявлении долгосрочного тренда. Применяя функцию ТЕНДЕНЦИЯ и формулу массива, в ячейки D2:D29 введём следующую формулу: $\{= \text{ТЕНДЕНЦИЯ}(B2:B29;C2:C29)\}$

Построенное методом наименьших квадратов линейное уравнение регрессии, имеет следующий вид:
 Долгосрочный тренд = 17 502 + 535,61*(период времени).

Это выражение указывает на то, что объемы продажи автомобилей увеличиваются в среднем на 535,61 (млн.руб.) за квартал. Этот долгосрочный тренд нетрудно прогнозировать, подставляя в уравнение регрессии соответствующий временной период.

Например, чтобы найти значение тренда для первого квартала 2003 г., используйте значение X=29, которое будет представлять период времени, следующий за окончанием нашего временного ряда.

Прогнозируемое значение тренда для первого квартала 2003 г.
 $=17\ 502 + 535,61 \cdot (\text{период времени}) = 17\ 502 + 535,61 \cdot 29 = 33\ 035$ (млн.руб.).

Спрогнозируем объемы продаж на период 2003-2005 годы. В диапазон ячеек C30:C41 последовательно занесем значения номеров кварталов 29, 30, 31...40. Прогнозные значения рассчитываются в ячейках D30:D41, куда введём формулу:

$\{= \text{ТЕНДЕНЦИЯ}(B2:B29;C2:C29; C30:C41)\}$

Прогноз: тренд с учетом сезонности

Все, что вам теперь требуется сделать, чтобы прогнозировать будущее, - это учесть сезонность в долгосрочном тренде, вернув ему ожидаемую сезонную вариацию. Для этого достаточно умножить значение тренда на значение сезонного индекса для того периода времени, который вы прогнозируете. Этот процесс является обратным по отношению к

внесению поправки на сезонные колебания. Результирующий прогноз включает долгосрочный тренд и сезонную вариацию.

Чтобы предсказать объемы продажи автомобилей компании А за первый квартал 2003г., достаточно умножить значение тренда, равное 33 035 на сезонный индекс для первого квартала, равный 1,0184.

Прогноз объема продаж за первый квартал 2003г.

$=33\ 035 * 1,0184 = 33\ 643$ (млн.руб.).

Аналогично можно рассчитать все ячейки в столбце F (рисунок 5). Учитывающий сезонность тренд отражает рассматриваемый ряд и экстраполируется вправо, обеспечивая достаточно надежные прогнозы, включающие ожидаемое сезонное поведение объемов продаж.

Лабораторная работа № 9: Прогнозирование на основе парной регрессии

Цель: научиться строить прогнозы с применением на основе парной регрессии на программе для работы с электронными таблицами Microsoft Excel.

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

Парная регрессия позволяет предсказывать одну переменную на основании другой с использованием прямой линии, характеризующей взаимосвязь между этими двумя переменными. Переменную, поведение которой прогнозируется, принято обозначать буквой Y; переменную, которая используется для такого прогнозирования, принято обозначать буквой X. Очень важно, что определяется как X, а что как Y, поскольку X предсказывает Y, и Y предсказывается с помощью X.

Порядок выполнения лабораторной работы

Пример. Количество произведенных изделий и затраты

Каждую неделю на заводе выпускается определенная партия изделий. Между количеством выпускаемых изделий и затратами на их производство предполагается достаточно сильная взаимосвязь. В те недели, когда завод работает с полной нагрузкой, выпускается большое количество изделий, для производства которых требуется большие объемы исходных материалов. Соответственно возрастают и затраты. Однако в имеющихся данных присутствует выброс – резко отклоняющееся значение. Несмотря на то, что с увеличением количества выпускаемой продукции затраты действительно возрастают, корреляция между изучаемыми признаками отрицательная (рисунок 1).

В ячейке B21 рассчитывается коэффициент корреляции по формуле:

$=КОРРЕЛ(A2:A20;B2:B20)$ (Результат: -0,623).

Результат появления в имеющихся данных выброса нивелирует имеющуюся положительную связь. Выброс стал следствием пожара, который произошел на заводе. Значительная часть исходных материалов была уничтожена огнем – именно это и стало причиной резкого скачка затрат на этой неделе. Выпуск продукции резко упал из-за остановки производства.

На рисунке 2 отражены те же данные, но без выброса. Как видно, корреляция, рассчитанная в ячейке B22 положительная и тесная ($r=0,869$). Если к имеющимся данным добавить линию тренда и выбрать в диалоговом окне Формат линии тренда⇒Параметры⇒Прогноз назад на 22 единицы, получим точку пересечения тренда с осью Y. Эта точка является параметром а для линейного тренда (также его называют «отрезком» или «сдвигом») и рассчитывается в ячейке F20 по формуле:

$=ОТРЕЗОК(C2:C19;B2:B19)$ (Результат: 2272,07).

Параметр b (коэффициент регрессии или «наклон») характеризует крутизну наклона линии регрессии и рассчитывается в ячейке F21: $=НАКЛОН(C2:C19;B2:B19)$ (Результат: 51,66).

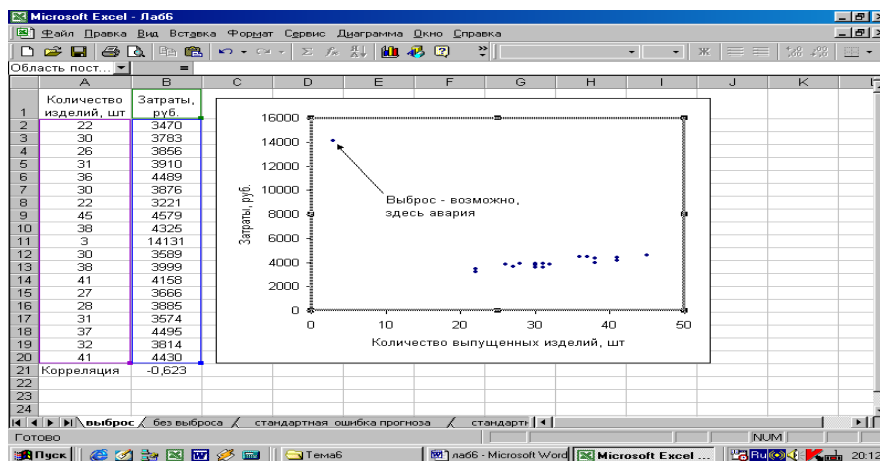


Рисунок 1 – Резко отклоняющееся значение нарушило корреляцию. Вместо того, чтобы выявить в целом взаимосвязь роста между объёмом произведённой продукции и затратами, коэффициент корреляции, $r = -0,623$, указывает на наличие не прямой, а обратной взаимосвязи, при которой более высоким объёмам производства соответствуют меньшие затраты.

В ячейке B20 рассчитывается среднее значение количества изделий по формуле: $=СРЗНАЧ(B2:B19)$ (Результат: 32,5), а в ячейке C20 - среднее значение затрат по формуле: $=СРЗНАЧ(C2:C19)$ (Результат: 3951,06).

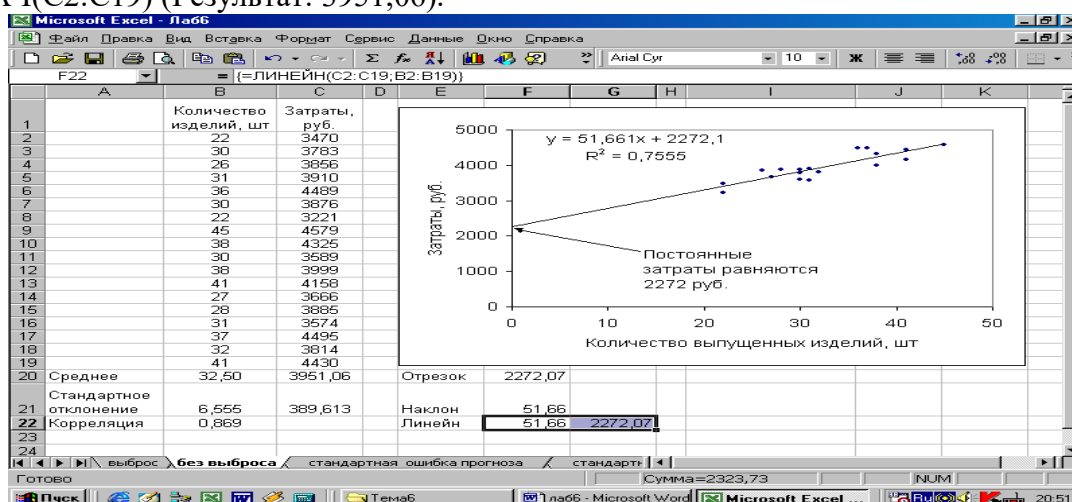


Рисунок 2 – Исходная совокупность данных без выброса иллюстрирует прямую взаимосвязь между объёмом произведённой продукции и затратами для «обычных» недель (без чрезвычайных происшествий). Коэффициент корреляции, $r = 0,869$, в этом случае имеет положительное значение и приближается к 1, что указывает на тесную связь.

В ячейке B21 рассчитывается среднее квадратическое отклонение (стандартное отклонение) количества изделий по формуле: $=СТАНДОТКЛОН(B2:B19)$ (Результат: 6,555), а в ячейке C21 – среднее квадратическое отклонение (стандартное отклонение) затрат по формуле:

$=СТАНДОТКЛОН(C2:C19)$ (Результат: 389,613).

Если выделить диапазон ячеек F22:G22 и ввести с помощью формулы массива следующее выражение:

$\{=ЛИНЕЙН(C2:C19;B2:B19)\}$ (Результат: 51,66; 2272,07), то таким образом рассчитаем одновременно оба параметра уравнения регрессии. Формула массива вводится постепенным нажатием клавиш $Ctrl \Rightarrow Shift \Rightarrow Enter$.

Стандартная ошибка (S_e) является приближённым показателем величины ошибок остатков для имеющихся данных, измеряется в тех же единицах, что и Y . Она показывает величину отклонения в большую и меньшую сторону от значений линии регрессии с определенной

долей вероятности. Так, если ошибки прогноза имеют нормальное распределение, то можно ожидать, что примерно 2/3 точек данных будут находиться на расстоянии не более величины S_e выше или ниже линии регрессии. При 95% вероятности значения будут сосредоточены вокруг линии регрессии на расстоянии $\pm 2S_e$ и при вероятности близкой к 1 - $\pm 3S_e$.

На рисунке 3 стандартная ошибка прогноза рассчитывается в ячейке H1 по формуле: $=СТОШУХ(B2:B19;A2:A19)$ (Результат: 198,582).

Тренд (столбец C) рассчитывается по формуле: $\{=ТЕНДЕНЦИЯ(B2:B19;A2:A19)\}$.

Нижняя и верхняя границы рассчитываются как разница и сумма тренда и стандартных ошибок прогноза соответственно. Рассчитаем границы с вероятностью 2/3 (около 67%).

Нижняя граница в ячейке D2: $=C2-\$H\1 (Результат: 3210).

Остальные ячейки столбца D копируем.

Верхняя граница в ячейке E2: $=C2+\$H\1 (Результат: 3607).

Остальные ячейки столбца E копируем.

Чтобы построить диаграмму, как на рисунке 3 необходимо выполнить следующие действия:

1. выделить диапазон ячеек A1:B19;
2. выбрать Вставка⇒Диаграмма⇒Точечная и оформить в соответствии с рисунком;
3. Диаграмма⇒Добавить линию тренда⇒Линейная;
4. Диаграмма⇒Исходные данные⇒Ряд⇒Добавить;
5. В качестве значения X выбрать диапазон ячеек A2:A19, значения Y – диапазон ячеек D2:D19
6. щёлкнув на диаграмме по появившимся ниже линии тренда точкам и выбрав Формат рядов данных ⇒Линия⇒ Обычная⇒ Маркер⇒ Отсутствует, получите на диаграмме нижнюю границу прогноза.
- 7.

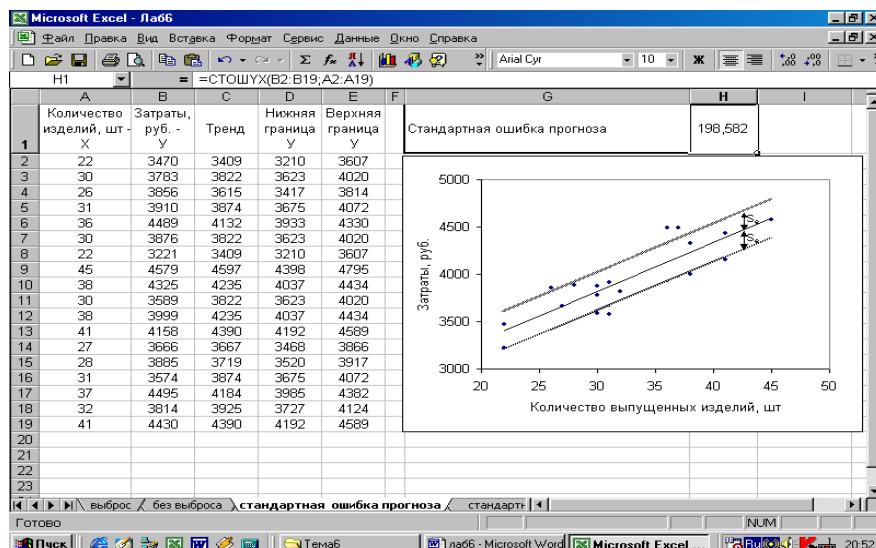


Рисунок 3 – Стандартная ошибка прогноза (S_e) показывает приблизительно ошибку какой величины можно допустить, когда вместо фактического значения Y используется прогнозируемое значение Y. Можно ожидать, что в случае обычной линейной связи примерно 2/3 точек данных будут находиться между верхней и нижней границами прогноза.

Аналогично строится верхняя граница прогноза.

На рисунке 4 показан расчёт стандартных ошибок параметров уравнения регрессии. Введем следующие формулы:

в ячейку E2: =СТОШУХ(B2:B19;A2:A19)	в ячейку H2: =СТОШУХ(B2:B19;A2:A19)
в ячейку E3: =1/СЧЁТ(B2:B19)	в ячейку H3: =СТАНДОТКЛОН(A2:A19)
в ячейку E4: =(СРЗНАЧ(A2:A19))^2	в ячейку H4: =СЧЁТ(B2:B19)-1
в ячейку E5: =(СТАНДОТКЛОН(A2:A19))^2	в ячейку H5: =КОРЕНЬ(H4)
в ячейку E6: =СЧЁТ(B2:B19)-1	в ячейку H1: =H2/(H3*H5)
в ячейку E7: =КОРЕНЬ(E3+E4/(E5*E6))	
в ячейку E1: =E2*E7	

	A	B	C	D	E	F	G	H
1	Количество изделий, шт - X	Затраты, руб. - Y	Стандартная ошибка свободного члена уравнения регрессии (параметр a)		243,33	Стандартная ошибка коэффициента регрессии (параметр b)		7,35
2	22	3470	Стандартная ошибка прогноза		198,58	Стандартная ошибка прогноза		198,58
3	30	3783	1/n		0,06	Стандартное отклонение X		6,56
4	26	3856	Среднее значение X в квадрате		1056,25	n-1		17,00
5	31	3910	Стандартное отклонение X в квадрате		42,97	корень		4,12
6	36	4489	n-1		17,00			
7	30	3876	корень		1,23			
8	22	3221	Расчёт доверительных интервалов параметров уравнения регрессии					
9	45	4579	Отрезок (параметр a)		2272,067	Наклон (параметр b)		51,86119
10	38	4325	t-критерий Стьюдента		2,119905	t-критерий Стьюдента		2,119905
11	30	3589	Нижняя граница		1756,23	Нижняя граница		36,09
12	38	3999	Верхняя граница		2787,91	Верхняя граница		67,24
13	41	4158						
14	27	3666						
15	28	3885						
16	31	3574						
17	37	4495						
18	32	3814						
19	41	4430						
20								
21								
22								

Рисунок 4 – Для расчёта доверительных интервалов параметров уравнения регрессии используются стандартные ошибки параметров a и b. Они показывают вероятное отклонение соответствующих параметров в генеральной и выборочной совокупностях.

Расчёт доверительных интервалов осуществляется вводом следующих формул:

в ячейку E10: =ОТРЕЗОК(B2:B19;A2:A19)	в ячейку H10: =НАКЛОН(B2:B19;A2:A19)
в ячейку E11: =СТЮДРАСПОБР(0,05;16)	в ячейку H11: =СТЮДРАСПОБР(0,05;16)
в ячейку E12: =E10-E11*E1	в ячейку H12: =H10-H11*H1
в ячейку E13: =E10+E11*E1	в ячейку H13: =H10+H11*H1

Функция СТЬЮДРАСПОБР позволяет определить t-статистику Стьюдента, которая придаёт доверительным интервалам вероятностный характер. В данном случае параметр вероятность, равный 0,05, говорит о 95% вероятности, а параметр степени_свободы, равный 16, о том, что из 18 степеней свободы 2 степени свободы теряются из-за двух параметров в уравнении регрессии.

В Excel для расчёта параметров уравнения регрессии в Пакете анализа существует средство Регрессия. Его можно активировать следующим образом: выбрать Сервис⇒Анализ данных⇒Регрессия. Появится диалоговое окно, показанное на рисунке 5.

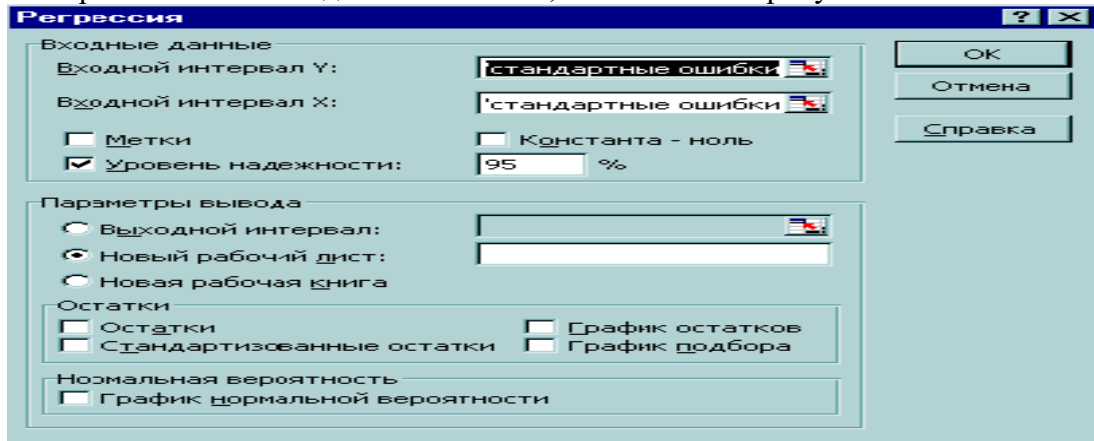


Рисунок 5 – Средство Пакета анализа «Регрессия» позволяет рассчитать большинство параметров регрессионной статистики и дисперсионного анализа

Если заполнить диалоговое окно Регрессия так, как показано на рисунке, в новом рабочем листе будет выведен перечень основных показателей регрессионной статистики и дисперсионного анализа (рисунок 6).

Вывод итогов							
Регрессионная статистика							
Множественный R							0,87
R-квадрат							0,76
Нормированный R-квадрат							0,74
Стандартная ошибка							198,58
Наблюдения							18
Дисперсионный анализ							
		df	SS	MS	F	Значимость F	
Регрессия		1	1949615,86	1949615,86	49,44	0,0000028	
Остаток		16	630957,09	39434,82			
Итого		17	2580572,94				
		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение		2272,07	243,33	9,34	0,00000007	1756,23	2787,91
Переменная X 1		51,86	7,35	7,03	0,00000284	36,09	67,24

Рисунок 6 – Результаты компьютерных вычислений для данных о производственных затратах и объёме производства продукции

В ячейке В4 рассчитан множественный коэффициент корреляции (множественный R). Это же значение показано на рисунке 2 – ячейка В22.

В ячейке В5 рассчитан множественный коэффициент детерминации (R-квадрат). Это же значение можно увидеть на диаграмме рисунка 2.

В ячейке В6 рассчитан скорректированный множественный коэффициент детерминации (нормированный R-квадрат), который содержит поправку на число степеней свободы и всегда меньше множественного коэффициента детерминации (R-квадрат).

В ячейке B7 рассчитана стандартная ошибка прогноза. Её значение соответствует результатам, полученным на рисунках 3 (ячейка H1) и 4 (ячейки E2 и H2).

В ячейке B8 содержится количество наблюдений (объём изучаемой совокупности).

В ячейках B11:B14 (df) содержатся данные о количестве степеней свободы.

Одна степень свободы в ячейке B12 говорит о непредсказуемости направления связи (прямая или обратная), число 16 в ячейке B13 представляет количество наблюдений минус число параметров уравнения регрессии.

В ячейках C11:C14 (SS – сумма квадратов) рассчитаны: объяснённая (факторная) – ячейка C12, остаточная – ячейка C13, и общая – ячейка C14, дисперсии.

В ячейках D11:D13 (MS – сумма квадратов на одну степень свободы) рассчитаны объяснённая (факторная) и остаточная дисперсии на 1 степень свободы.

В ячейке E12 рассчитана F-статистика Фишера (путём деления ячейки D12 на ячейку D13), а в ячейке F12 её значимость. Полученные данные позволяют с очень высокой степенью вероятности говорить о значимости построенной регрессионной модели. Лишь в 28 случаях из 10000000 модель не значима.

В ячейках B17:B18 рассчитаны коэффициенты уравнения регрессии, они равны соответствующим значениям на рисунках 2 и 4. В ячейках C17:C18 рассчитаны стандартные ошибки, соответствующие значениям E1 и H1 на рисунке 4.

Если разделить коэффициент уравнения регрессии на стандартную ошибку, то будет получено значение t-статистики. Так, если B17 разделить на C17, получим ячейку D17. Значения ячеек D17:D18 превосходят табличные значения t-статистики. Это означает, что коэффициенты уравнения регрессии значимо отличаются от нуля. Р-значения в ячейках E17:E18 говорят об очень незначительной вероятности получения нулевого результата.

В ячейках F17:G18 содержатся доверительные интервалы коэффициентов уравнения регрессии при 95% уровне вероятности, которые совпадают с данными рисунка 4.

Форма отчета: По окончании выполнения лабораторных работ оформляется отчет в печатном виде в соответствии с требованиями ГОСТа в формате А4, шрифт 14, интервал 1,5. Для оформления отчета по каждой из лабораторных работ следует детально проработать предлагаемые методические рекомендации, внимательно изучить исходную информационную базу и провести необходимые расчеты.

Отчет по лабораторной работе должен включать:

1. Исходные данные и задание.
2. Расчётные формулы для проведения необходимых расчетов исходя из задания лабораторной работы.
3. Графическое представление полученных результатов расчета (при необходимости).
4. Результаты расчёта показателей, оформленные в виде таблиц.
5. Основные выводы.

Выводы: Выводы, сделанные в ходе проведенных расчетов должны быть четко и полно аргументированы.

Вопросы к защите лабораторной работы:

1. Какой процесс вы прогнозировали?
2. При помощи какого метода вы прогнозировали процесс? Его особенности.
3. Какие выводы можно сделать выводы?

Лабораторная работа №10: Прогнозирование на основе множественной регрессии

Цель: научиться строить прогнозы на основе множественной регрессии на программе для работы с электронными таблицами Microsoft Excel.

Оборудование: ПК, оснащенный необходимым программным обеспечением.

Необходимый теоретический минимум:

В подавляющем большинстве реальных экономических задач приходится рассматривать данные более чем об одном или двух факторах. Прогнозирование единственной переменной Y на основании двух или нескольких переменных X называется множественной регрессией.

Порядок выполнения лабораторной работы

Пример: Смета на рекламу и цены

На предприятии рассматривается возможность увеличения сметы расходов на рекламу конкретной продукции (18 различных видов) и сокращения цен реализации с целью увеличения объёмов продаж в натуральном выражении. Необходимо выявить связь между данными факторами, а при её наличии сделать прогноз объёмов продаж продукции.

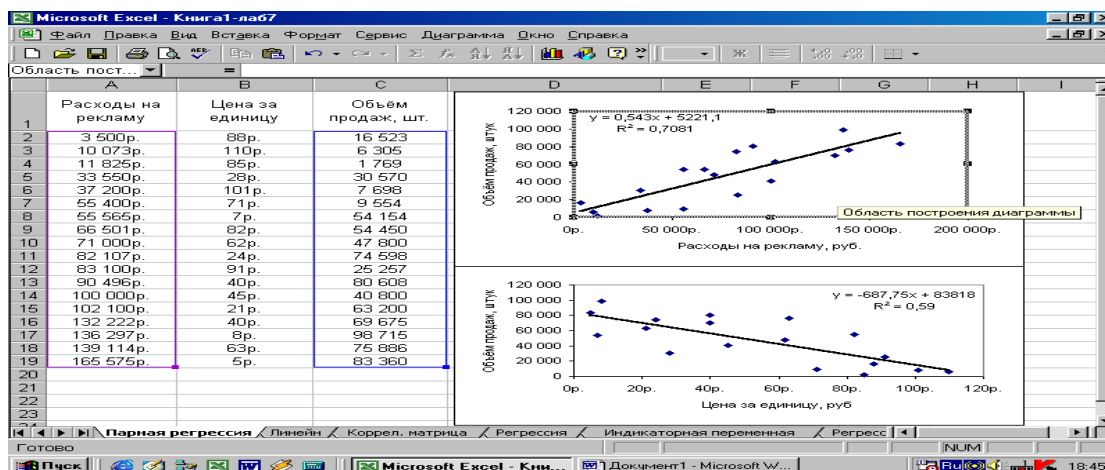


Рисунок 1 – По мере увеличения расходов на рекламу и понижения цен реализации объёмы продаж возрастают

При рассмотрении парных зависимостей между: 1) объёмом продаж (Y) и расходами на рекламу (X₁) и 2) между объёмом продаж и ценами за единицу продукции (X₂) выявлены, соответственно, прямая и обратная связь: с увеличением расходов на рекламу продажи увеличиваются, а при увеличении цен, напротив, сокращаются (рисунок 1).

Учитывая тесную связь в данных зависимостях (величина R² позволяет сделать вывод о существенной связи в обоих случаях) попытаемся оценить совместное влияние изменений сметы на рекламу и цен реализации на объёмы продаж. Для этого воспользуемся функцией рабочего листа ЛИНЕЙН. В диапазоне E8:G12 на рисунке 2 приведён результат вычисления функции ЛИНЕЙН, которая вводится с помощью формулы массива:

`.}=ЛИНЕЙН(С2:С19;А2:В19;ИСТИНА;ИСТИНА{`

Тогда как рассчитано первое значение в ячейке E8, следует щёлкнуть мышью в формульную строку после знака «=» и постепенно набрать клавиши Ctrl⇒Shift⇒Enter.

Результат, возвращённый функцией, очень богат информацией. Он обеспечивает нас уравнением регрессии для получения наиболее точной оценки, с помощью которого можно предсказать объёмы продаж на основе расходов на рекламу и цен реализации:

$$Y = 36779,49 + 0,38 * X_1 - 358,14 * X_2$$

где Y – объём продаж в единицах продукции,

X₁ - расходы на рекламу в руб.,

X₂ – цена реализации в руб.

Свободный член уравнения регрессии 36779,49 показывает величину Y при нулевых значениях X₁ и X₂. Функция ЛИНЕЙН возвращает коэффициенты регрессии 0,38 и -358,14 не в том порядке, в котором расположены изменяемые переменные в рабочем листе. Это одно из неудобств функции ЛИНЕЙН, которое никак нельзя обойти.

Во второй строке результатов вычисления (ячейки E8:G8) рассчитываются стандартные ошибки соответствующих параметров уравнения регрессии, которые подтверждают их статистическую значимость. Если разделить каждый коэффициент регрессии на его

стандартную ошибку, то будет получен t-критерий, называемый стандартизованной, или нормированной переменной.

Например, в ячейке G17 рассчитан t-критерий для коэффициента регрессии при X_1 по формуле:

=F8/F9 (Результат: 4,097).

	А	В	С	Д	Е	Ф	Г	Н	И	Ж
	Расходы на рекламу, руб.	Цена за единицу, руб.	Объем продаж, шт.						Прогнозный объем продаж, шт.	
1										
2	3 500	88	16 523						6 603	
3	10 073	110	6 305						1 240	
4	11 825	85	1 769						10 865	
5	33 550	28	30 570						39 596	
6	37 200	101	7 698						14 849	
7	55 400	71	9 554						32 561	
8	55 565	7	54 154		-368,14	0,38	36779,49		55 545	
9	66 501	82	54 450		129,66	0,09	13165,54		32 871	
10	71 000	62	47 800		0,8065	14348,62	#Н/Д		41 756	
11	82 107	24	74 598		31,27	15,00	#Н/Д		59 618	
12	83 100	91	25 257		12 875 046 967	3 088 244 387	#Н/Д		36 003	
13	90 496	40	80 608						57 099	
14	100 000	45	40 800						58 947	
15	102 100	21	63 200						68 347	
16	132 222	40	69 675				0,8065		73 074	
17	136 297	8	98 715				4,097203		86 095	
18	139 114	63	75 886				0,0005		67 475	
19	165 575	5	83 360						98 378	
20										
21										
22										

Рисунок 2 – Функция рабочего листа ЛИНЕЙН возвращает статистические данные, описывающие связь между несколькими факторными и резульативной переменными

Полученное значение показывает, что коэффициент регрессии более чем в 4 раза превышает стандартную ошибку. Для того, чтобы считать статистически значимым данный параметр, необходимо воспользоваться функцией СТЬЮДРАСП и ввести в ячейку G18 формулу:

=СТЬЮДРАСП(G17;15;1) (Результат: 0,0005).

Рассмотрим аргументы данной функции:

G17 - t-критерий для коэффициента регрессии;

15 – число степеней свободы, рассчитанное как разница между 18 результатами наблюдений и количеством параметров в уравнении регрессии – 3 (18-3=15);

1 – количество возвращаемых «хвостов» (вводим 1, если заранее известно направление связи и 2, если не известно, положительная или отрицательная связь будет наблюдаться между факторами).

В нашем случае логично предположить, что связь между объёмами продаж и расходами на рекламу будет прямой, поэтому выбираем 1.

Функция СТЬЮДРАСП возвращает значение 0,0005, то есть в рассматриваемом случае вероятность получения t-критерия, равного 4,097 при условии равенства коэффициента регрессии 0, составляет 5 шансов из 10000, то есть коэффициент регрессии с большой вероятностью отличен от 0.

Значение 0,8065 в третьей строке и первом столбце результатов вычисления функции ЛИНЕЙН является значением R^2 (коэффициентом детерминации) между объёмом продаж и самым удачным сочетанием сметы на рекламу и цен реализации. Это значит, что данными двумя факторами объяснено 80,65% вариации объёма продаж. Это же значение можно получить, воспользовавшись функцией КВПИРСОН. Введём в ячейку G16 формулу:

=КВПИРСОН(C2:C19;I2:I19) (Результат: 0,8065)

Но что более важно, величина 0,8065 больше числа 0,7081, которое представляет собой коэффициент детерминации между значениями объёма продаж и расходов на рекламу. Это означает, что, объединяя расходы на рекламу и цены реализации, можно объяснить

происхождение около 10% изменчивости (80,65-70,81) возникающих дополнительно в показателе объёмов продаж.

В ячейке F10 рассчитывается стандартная ошибка прогнозирования, равная 14348,62. В ячейке E11 находится F-критерий Фишера (31,27), ячейке F11 – число степеней свободы (15), а в ячейках E12 и F12 – факторная и остаточная дисперсии (12875046967 и 3088244387). Прогнозный объём продаж может быть получен из уравнения регрессии. Введём в ячейку I2 формулу:

= $\$G\$8+(\$E\$8*B2)+(\$F\$8*A2)$ (Результат: 6603).

Для этой формулы использовалась абсолютная адресация ячеек (знаки \$ перед буквенным и цифровым форматом ячейки) для того, чтобы при копировании в ячейки I2:I19, процессор обращался к одним и тем же значениям - параметрам уравнения регрессии.

Те же данные можно получить, используя средство Пакета анализа Регрессия:

выберем Сервис⇒Анализ данных⇒Регрессия;

заполним диалоговое окно Регрессия (подобно лабораторной работе №6).

Результаты вычислений представлены на рисунке 3.

Для выяснения тесноты взаимосвязи между результативным и факторными признаками строится корреляционная матрица с помощью средства Корреляция из Пакета анализа. Для этого:

выберем Сервис⇒Анализ данных⇒Корреляция;

Заполним диалоговое окно Корреляция как на рисунке 4.

Регрессионная статистика						
Множественный R	0,8981					
R-квадрат	0,8065					
Нормированный R-квадрат	0,7807					
Стандартная ошибка	14348,62					
Наблюдения	18,00					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	2	12 875 046 967	6 437 523 484	31,27	0,00000446	
Остаток	15	3 088 244 387	205 882 959			
Итого	17	15 963 291 354				
Коэффициенты						
Y-пересечение	36779,49	13165,54	2,79	0,01363	8717,79	64841,20
Переменная X 1	0,38	0,09	4,10	0,00095	0,18	0,58
Переменная X 2	-358,14	129,66	-2,76	0,01452	-634,50	-81,78

Рисунок 3 - Результаты компьютерных вычислений для данных о расходах на рекламу (X_1), ценах за единицу продукции (X_2) и объёмах продаж продукции (Y)

Анализируя рассчитанную компьютером корреляционную матрицу (рисунок 5), можно сказать, что наиболее тесная положительная взаимосвязь обнаруживается между расходами на рекламу и объёмом продаж (коэффициент корреляции $r=0,84$). Соответственно, наименьшая отрицательная связь – между расходами на рекламу и ценой реализации (коэффициент корреляции $r=-0,62$).

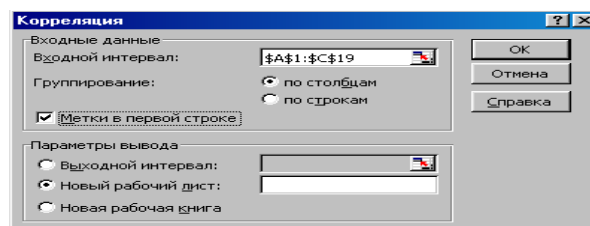


Рисунок 4 – Средство Пакета анализа Корреляция

	A	B	C	D
1		Расходы на рекламу, руб.	Цена за единицу, руб.	Объем продаж, шт.
2	Расходы на рекламу, руб.	1		
3	Цена за единицу, руб.	-0,620464619	1	
4	Объем продаж, шт.	0,841508566	-0,768136283	1
5				

Рисунок 5 – Корреляционная матрица показывает парные корреляции между факторными и результативными показателями

Часто при построении множественной регрессии применяют, так называемые, фиктивные (индикаторные) переменные — это переменные, принимающие лишь два возможных значения (0 или 1). Они используются в качестве независимой (объясняющей) X-переменной для представления качественных категориальных данных.

Количество фиктивных переменных для замены качественной переменной должно быть на единицу меньше количества категорий. Оставшаяся категория определяет базу. Базовая категория представляется в результирующем уравнении регрессии постоянным членом. Вместо использования фиктивных переменных можно находить отдельные уравнения регрессии для каждой из категорий. Это приводит к более гибкой модели с различными коэффициентами регрессии для каждой из X-переменных по каждой категории.

Пример: Пол работников, как один из факторов множественной регрессии

Руководство фирмы озабочено возможностью обвинений в дискриминации сотрудников по признаку пола. Возникают подозрения, что сотрудники-мужчины зарабатывают больше, чем женщины. Краткий анализ заработной платы 24 мужчин и 26 женщин показывает, что в среднем мужчина получает за год на 4214 рублей больше женщины (таблица 1).

Таблица 1 – Заработная плата, стаж работы и пол сотрудников

№	Годовая заработная плата (ТГ.), У	Стаж работы (лет), X ₁	Пол (1 - женщины, 0 - мужчины), X ₂	№	Годовая заработная плата (ТГ.), У	Стаж работы (лет), X ₁	Пол (1 - женщины, 0 - мужчины), X ₂
1	39700	16	0	27	37750	44	0
2	28500	2	1	28	31350	2	1
3	30650	2	1	29	27350	0	1
4	31000	3	1	30	35700	19	1
5	33700	25	0	31	32250	7	0
6	33250	15	0	32	25200	0	1
7	35050	16	1	33	35900	15	1
8	22800	0	1	34	36700	14	0
9	36300	33	0	35	32050	4	1
10	35600	29	1	36	38050	33	0
11	32350	3	1	37	36100	19	0
12	31800	16	0	38	35200	20	1
13	26900	0	1	39	34800	24	0
14	37250	19	0	40	26550	3	0
15	30450	1	1	41	26550	0	1
16	31350	2	1	42	32750	17	0
17	38200	32	0	43	39200	19	0
18	38200	21	1	44	30450	0	1
19	28950	0	1	45	38800	21	0
20	33950	34	0	46	41000	31	0
21	34100	8	1	47	29900	6	0

22	32900	11	1	48	40400	35	0
23	30150	5	1	49	37400	20	0
24	30800	1	0	50	35500	23	0
25	31300	11	1	Среднее значение	33313	13,98	52,0%
26	33550	18	1	Стандартное отклонение	4188	11,87	

Означает ли это, что дискриминация сотрудников по признаку пола действительно имеет место в данной фирме?

Исходные данные для множественной регрессии между годовой заработной платой (Y), стажем работников (X₁) и фиктивной переменной – полом работников (X₂) представлены также и на рисунке 6.

Указанные результаты не обязательно являются свидетельством дискриминации. Предполагается, что на заработную плату оказывают влияние и другие факторы, например, стаж работы. Для выяснения этого пригодной является множественная регрессия, так как коэффициент регрессии при фиктивной переменной, представляющей пол сотрудника, даст ожидаемую разницу в уровне заработной платы между мужчиной и женщиной с одинаковым рабочим стажем.

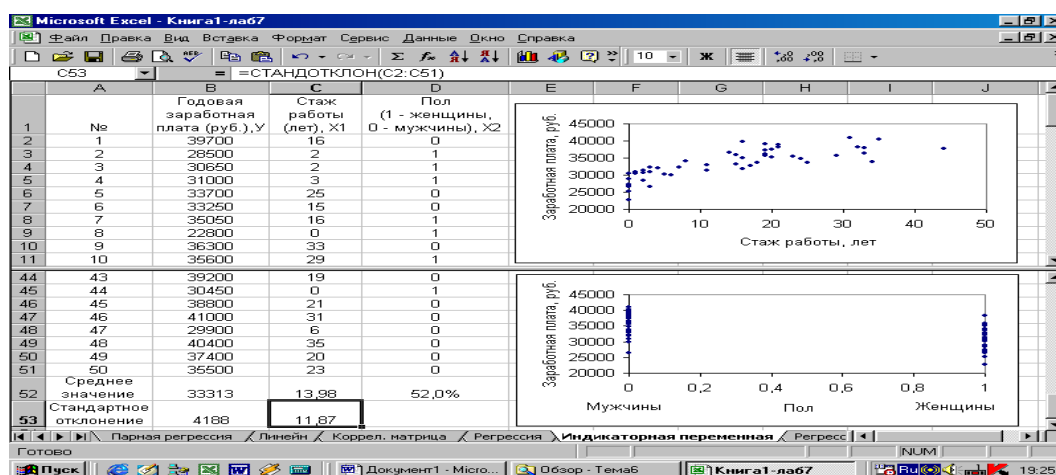


Рисунок 6 – На верхнем графике диаграмма рассеяния значений заработной платы в зависимости от рабочего стажа свидетельствует о наличии сильной прямой связи, тогда как нижний график трудно интерпретировать, так как пол является фиктивной переменной

Применив средство Пакета анализа Регрессия, получим основные параметры множественной регрессии (рисунок 7).

Уравнение регрессии будет иметь следующий вид:

$$\text{Заработная плата} = 29776 + 271,1 * \text{стаж работы} - 488,1 * \text{пол}$$

Стандартная ошибка прогноза, равная 2539 указывает на типичную величину ошибок прогнозирования для данной совокупности. Значение R-квадрат, равное 64,7% указывает, какая часть дисперсии заработной платы объясняется данной регрессионной моделью.

Статистический вывод на уровне 5% относительно заработной платы действительно объясняет значимую долю вариации (F=43,1).

Коэффициент регрессии для пола, -488,1, указывает, что ожидаемая разница в уровне заработной платы между мужчиной и женщиной с одинаковым рабочим стажем равна 488,1 рубля, причём труд женщин оплачивается ниже, чем труд мужчин. Это объясняется тем, что увеличение фиктивной переменной X₂ на единицу приводит к переходу от 0 (мужчина) к 1 (женщина), результатом чего является ожидаемое отрицательное изменение (-488,1 рублей) заработной платы.

Регрессионная статистика						
Множественный R	0,80464					
R-квадрат	0,64745					
Нормированный R-квадрат	0,63245					
Стандартная ошибка	2539					
Наблюдения	50					
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	2	556 323 981	278 161 991	43,15720	0,000000000023	
Остаток	47	302 930 069	6 445 321			
Итого	49	859 254 050				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	29776,1	948,9	31,4	0,00000	27867,3	31685,0
Стаж работы, X1	271,1	37,6	7,2	0,00000	196,5	346,8
Пол, X2	-488,1	885,3	-0,6	0,58402	-2269,0	1292,9

Рисунок 7 - Результаты компьютерных вычислений для данных о годовой заработной плате (Y), стаже работников (X₁) и поле работников (X₂). При значимости множественной модели регрессии незначимой может оказаться одна из переменных, например, как в данном случае – пол работников

Если посмотреть на последнюю строку рисунка 7, в ячейках F19:G19 можно увидеть нижнюю и верхнюю границы коэффициента регрессии при 95% вероятности. То что, нижняя граница представляет собой отрицательное число, а верхняя – положительное, говорит о том, что коэффициент регрессии при фиктивной переменной не значим, так как 0 попадает в диапазон между этими границами. Это подтверждает и P-значение (ячейка E19).

Полученный результат свидетельствует о том, что с поправкой на рабочий стаж не наблюдается сколько-нибудь значимой разницы между средними уровнями заработной платы мужчин и женщин. Очевидные различия в уровнях заработной платы представителями разных полов можно объяснить различиями в их стаже работы. Итак, мы получили убедительное доказательство того, что если в данной фирме и есть дискриминация, она основана на стаже работы, но не на признаке пола.

Форма отчета: По окончании выполнения лабораторных работ оформляется отчет в печатном виде в соответствии с требованиями ГОСТа в формате А4, шрифт 14, интервал 1,5. Для оформления отчета по каждой из лабораторных работ следует детально проработать предлагаемые методические рекомендации, внимательно изучить исходную информационную базу и провести необходимые расчеты.

Отчет по лабораторной работе должен включать:

1. Исходные данные и задание.
2. Расчётные формулы для проведения необходимых расчетов исходя из задания лабораторной работы.
3. Графическое представление полученных результатов расчета (при необходимости).
4. Результаты расчёта показателей, оформленные в виде таблиц.
5. Основные выводы.

Выводы: Выводы, сделанные в ходе проведенных расчетов должны быть четко и полно аргументированы.

Вопросы к защите лабораторной работы:

1. Какой процесс вы прогнозировали?
2. При помощи какого метода вы прогнозировали процесс? Его особенности.
3. Какие выводы можно сделать выводы?

Список использованной литературы

1. Бабич, Т., Н. Прогнозирование и планирование в условиях рынка: Учебное пособие / Т. 2. Н. Бабич, И. А. Козьева, Ю. В. Вертакова. - М.: Инфра-М, 2018. - 256 с.
3. Коротаев, А.В. Законы истории: Математическое моделирование и прогнозирование мирового и регионального развития / А.В. Коротаев, Д.А. Халтурина, А.С. Малков и др. - М.: Издательство ЛКИ, 2019. - 344 с.
4. Политический анализ и прогнозирование: Учебное пособие. Стандарт третьего поколения. Для бакалавров / Под ред. В. Семенова, В. Колесникова. - СПб.: Питер, 2017. 432с.
5. Политический анализ и прогнозирование. Учебное пособие. Стандарт третьего поколения. Для бакалавров / Под ред. Семенова В.А.. - СПб.: Питер, 2017. - 16 с.