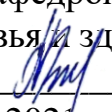


УТВЕРЖДАЮ
Зав. кафедрой общественного
здоровья и здравоохранения
 В.Л.Аджиенко
11.06.2021

Методические указания для студентов
второго курса ЛЕЧЕБНОГО факультета

к проведению практического занятия
по дисциплине «Медицинская информатика»

***Тема 11. Основы математического
моделирования. Методы регрессии и
аппроксимации***

Волгоград
2021

Ред 11.10.2021

ВОПРОСЫ ДЛЯ ПОДГОТОВКИ К ВЫПОЛНЕНИЮ ЗАДАНИЙ

1. С какой целью применяются методы моделирования?
2. Моделирование – определение?
3. Классификация моделей по способу представления?
4. Математическое моделирование – определение?
5. Классификация математических моделей по цели применения?
6. Назовите этапы решения задачи на основе компьютерного моделирования?
7. В таких задачах могут применяться два варианта вычислительного эксперимента?
8. Виды связей между признаками, примеры.
9. Функциональная зависимость (определение).
10. Какие явления могут быть смоделированы функциональной зависимостью?
11. Корреляционная зависимость (определение).
12. Способы обнаружения корреляционной связи.
13. Оценка силы и направления связи по величине коэффициента корреляции.
14. Методика расчета коэффициента линейной корреляции.
15. Как оценивается достоверность коэффициента корреляции?
16. Методика расчета рангового коэффициента корреляции.
17. Дайте определение термина «Регрессия».
18. Значение коэффициента регрессии и методика его расчета.
19. Создание уравнения регрессии.
20. Вычисление и оценка сигмы регрессии.
21. Метод аппроксимации – определение?
22. В чем заключаются задачи интерполирования и экстраполирования?
23. Какие виды функций аппроксимации позволяет выбирать программа Excel?
24. Каким действием в программе Excel производится добавление в диаграмму линии тренда?
25. Назначение коэффициента детерминации?

ОСНОВЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Моделирование - это замещение некоторого объекта *А* другим объектом *Б*. Замещаемый объект *А* называется оригиналом или объектом моделирования. Замещающий объект *Б* - моделью.

Моделью является представление исследуемого объекта в некоторой форме, отличной от формы его реального существования, выраженное различными средствами: математическим соотношением, числами, текстами, графиками, рисунками, словесным описанием или материальным объектом, изучение которых дает новые знания.

Создание модели производится с целью получения новой информации о моделируемом объекте путем проведения эксперимента с его моделью. Однако никакая модель не может учесть все свойства и поведение прототипа, поэтому полученный на основе модели результат соответствует реальности лишь приближённо. Создавая модель, человек, прежде всего, стремится отобрать наиболее существенные признаки объекта, пренебрегая теми, которые не оказывают заметного влияния на результат.

Моделирование применяется для изучения свойств:

- **Материальных тел:** архитектурных сооружений и художественных произведений, технических средств, Солнечной системы, земли (глобус), химического строения веществ, атомов и др. в том числе человека, его органов и систем.
- **Природных явлений:** модели грозового разряда, магнитных и электрических сил, землетрясений, цунами, климата и др.
- **Процессов:** модели развития Вселенной, экономических и экологических процессов.
- **Общественных явлений:** государственное развитие, социологические процессы и т.д.

Классификация моделей

В зависимости от поставленной задачи, способа создания модели и предметной области различают множество типов моделей:

1. По способу представления различают материальные и информационные (нематериальные, абстрактные) модели. Информационные модели, в свою очередь, подразделяются на знаковые (записанные специальным языком, в том числе математическими формулами) и вербальные (на естественном языке). Знаковые модели разделяются на компьютерные и некомпьютерные.

2. По форме представления встречаются геометрические, словесные, логические, специальные (ноты, химические формулы и т.п.), математические модели.

3. По области использования выделяют учебные, опытные, игровые, имитационные, научно-исследовательские модели.

4. По временному фактору - статические и динамические.

Одним из наиболее универсальных видов моделирования является математическое моделирование, которое описывает моделируемый физический процесс системой математических соотношений, решение которой позволяет получить ответ на вопрос о поведении объекта без создания физической модели, часто оказывающейся дорогостоящей и неэффективной.

***Математическое моделирование* - это средство изучения реального объекта, процесса или системы путем их замены математической моделью, более удобной для экспериментального исследования с помощью компьютера.**

В общем случае математическая модель реального объекта, процесса или системы представляется в виде системы функционалов

$$\Phi_i(X, Y, Z, t) = 0,$$

где X - вектор входных переменных, $X = [x_1, x_2, x_3, \dots, x_N]t$,

Y - вектор выходных переменных, $Y = [y_1, y_2, y_3, \dots, y_N]t$,

Z - вектор внешних воздействий, $Z = [z_1, z_2, z_3, \dots, z_N]t$,

t - координата времени.

Создание математической модели обычно начинается с построения и анализа простейшей, наиболее грубой математической модели рассматриваемого объекта. В дальнейшем, в случае необходимости, модель уточняется, делается ее соответствие объекту более полным.

Например, требуется определить площадь поверхности письменного стола. Обычно для этого измеряют его длину и ширину, а затем перемножают полученные числа. Такая элементарная процедура фактически обозначает следующее: реальный объект (поверхность стола) заменяется абстрактной математической моделью – прямоугольником. Она не учитывает его материал, вес, и другие характеристики, но позволяет оценить наиболее важный для нашей задачи параметр. Площадь прямоугольника, как известно из школьного курса геометрии, определяется через умножение длины и ширины. Эта формула является математической моделью поверхности стола, а его площадь приближенно равна площади его модели, т.е. прямоугольника. При этом могут не учитываться закругления углов стола и его изгибы.

Модель прямоугольника для письменного стола – это простейшая, наиболее грубая модель. Для более точного подхода к решению этой задачи такую модель необходимо проверить. Эту проверку можно выполнить, измерив длины противоположных сторон стола и длины его

диагоналей, а затем сравнить их между собой. Если, с требуемой степенью точности, длины противоположных сторон и длины диагоналей попарно равны между собой, то поверхность стола действительно можно рассматривать как прямоугольник. В противном случае модель прямоугольника придется отвергнуть и заменить ее другой моделью - четырехугольника общего вида. При более высоком требовании к точности может возникнуть необходимость в дополнительном уточнении этой модели, например, учесть закругления углов, а также изгиб края стола.

Классификация математических моделей по цели применения:

- **Оптимизационные** - связанные с нахождением точек минимума или максимума некоторых показателей. Например, расчет оптимального пути передвижения.

- **Описательные** - характеризующие поведение некоторой системы при изменении условий ее функционирования и не предназначенные для целей управления или оптимизации. Например, модель финансовых потоков государства.

Классификация математических моделей по способу оценки:

- **Детерминистские** - используют точную оценку переменных одним числом при конкретных значениях исходных данных (математические функции)

- **Стохастические** - описывают переменные несколькими параметрами на основе вероятностных характеристик (корреляция и регрессия)

Компьютерная модель – это математическая модель, выраженная средствами программной среды.

Основные пути создания компьютерных моделей:

- 1) Разработка компьютерной программы на одном из языков программирования (Бейсик, Делфи и др.)
- 2) С помощью прикладного программного обеспечения, например табличного редактора Excel.

Этапы решения задачи на основе компьютерного моделирования:

1-й этап: Постановка задачи. Поиск решения новой задачи начинается с её изучения и анализа. Результатом такого анализа должна стать четкая постановка задачи, включающая ответы на следующие вопросы. Что известно? Что требуется найти? Какие данные допустимы? Какие результаты возможны?

- 2-й этап: Разработка модели.** Определяется основная цель моделирования: познание окружающего мира, создание объекта с заданными свойствами, определение последствий воздействия на объект, эффективность управления объектом или процессом и др. Производится выбор оптимального вида моделирования.
- 3-й этап: Реализация модели** (компьютерный эксперимент) состоит в составлении алгоритма укрупненными блоками, детализации алгоритма каждого блока до уровня команд, написании компьютерной программы, её отладке и тестировании.
- 4-й этап: Анализ результатов моделирования:** сопоставление полученных результатов с ожидаемыми или возможными, принятие решений о дальнейшем использовании модели.

Для проверки адекватности разработанной математической модели (т.е. её соответствия реальному объекту, процессу или системе) результаты компьютерного (вычислительного) эксперимента сравниваются с данными, полученными на опытном натурном образце. Такая проверка используется для последующей корректировки математической модели или решения вопроса о её применимости к исследованию заданных объектов, процессов или систем.

Суть компьютерного эксперимента состоит в следующем: на основе математической модели с помощью ЭВМ проводится серия вычислительных экспериментов, т.е. исследуются свойства объектов или процессов, находятся их оптимальные параметры и режимы работы, а затем полученная модель повторно уточняется. Например, располагая уравнением, описывающим протекание того или иного процесса, можно изменяя его коэффициенты, начальные и граничные условия, исследовать, как будет вести себя объект. Более того, можно спрогнозировать поведение объекта в различных условиях.

Компьютерный эксперимент позволяет заменить дорогостоящий натурный эксперимент расчетами на ЭВМ. Он позволяет в короткие сроки и без значительных материальных затрат осуществить исследование большого числа вариантов проектируемого объекта или процесса для различных режимов его эксплуатации, что значительно сокращает сроки разработки сложных систем и их внедрение в производство.

Наиболее перспективным для проведения вычислительного эксперимента является его использование для решения крупных научно-технических и социально-экономических проблем. В том числе, проектирование реакторов для атомных электростанций, плотин гидроэлектростанций, магнетогидродинамических преобразователей энергии и др. В области экономики – составление сбалансированного плана работы отрасли, региона, для страны и др. В медицине – поиск новых лекарственных препаратов, прогнозирование развития эпидемий и патологических состояний, моделирование органов и систем человека и др.

В некоторых процессах, где натурный эксперимент опасен для жизни и здоровья людей, вычислительный эксперимент является единственно возможным. К таким задачам относятся: термоядерный синтез, освоение космического пространства, проектирование и исследование химических производств, разработка новых лекарственных препаратов или методов лечения и т.д.

На практике, как правило, исследователь может получить зависимости между изучаемыми параметрами экспериментально. В этом случае проводится натурный эксперимент, измеряются значения параметров на входе и выходе системы. Результаты измерений заносятся в таблицу, которая выражается в форме табличной функции.

Затем над полученной табличной функцией выполняется вычислительный эксперимент с целью создания математической модели. Например, необходимо проинтегрировать или продифференцировать табличную функцию и др.

В таких задачах могут применяться два варианта вычислительного эксперимента:

1. Нахождение функции аппроксимации.
2. Интерполирование или экстраполирование.

МЕТОД КОРРЕЛЯЦИИ

При проведении исследования в биологии или медицине, как правило, регистрируются множество учетных признаков. Представляет интерес вопрос об их взаимном изменении, т.е. обнаружение зависимостей между ними. Выявление наличия таких взаимосвязей является одной из важнейших задач любой науки, в том числе и медицины.

Различают две формы количественных связей между явлениями или процессами: функциональную и корреляционную. Под ФУНКЦИОНАЛЬНОЙ понимают такую связь, при которой любому значению одного из признаков соответствует строго определенное значение другого. В точных науках, таких, как физика, химия и другие, может быть установлена функциональная взаимосвязь. Например, зависимость площади круга от длины окружности в геометрии, или в физике длина пути, пройденной телом в свободном падении, от времени. Наиболее известным видом функциональной зависимости является линейная, которая выражается математической формулой: $y = ax + b$.

В биологии и медицине установить функциональную зависимость, как правило, не удастся. Объекты этих исследований имеют большую изменчивость и зависят от огромного числа факторов, измерить которые просто невозможно. В этом случае определяется наличие

КОРРЕЛЯЦИОННОЙ связи, при которой значению каждой средней величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака. Например: связь между ростом и массой тела человека. У группы людей с одинаковым ростом наблюдается различная масса тела, однако она варьирует в определенных пределах вокруг средней величины. Поэтому такую зависимость нужно оценивать с использованием понятия случайной величины с привлечением подходов теории вероятности. Такую форму зависимостей называют «Корреляционной».

При поиске зависимости между признаками может быть обнаружена взаимосвязь, различная по направлению и силе:

- Прямая (при увеличении одного признака увеличивается второй);
- Обратная (при увеличении одного признака второй уменьшается).

Степень взаимосвязи признаков по силе (тесноте) принято обозначать как:

- Отсутствие;
- Слабая;
- Средняя;
- Сильная;
- Полная.

Способами выявления корреляционной взаимосвязи между признаками являются:

- Визуальные (таблицы и графики).
- Статистические (корреляция и регрессия).

Следует подчеркнуть, что обнаружение корреляции между двумя признаками еще не говорит о существовании причинной связи между ними, а лишь указывает на возможность таковой или на наличие фактора, определяющего изменение обеих переменных совместно.

Приёмы визуализации данных позволяют обнаружить корреляционную зависимость лишь при небольшом числе наблюдений и только приблизительно. Для обнаружения корреляционной взаимосвязи с помощью таблицы в ней располагают ранжированные вариационные ряды и затем определяют совместное изменение признаков. График более наглядно демонстрирует такую зависимость и позволяет оценить ее форму: линейная, параболическая, тригонометрическая и др.

Наиболее точным способом обнаружения взаимосвязи между признаками является вычисление коэффициента корреляции. В зависимости от природы обрабатываемых данных применяются параметрические или непараметрические методы вычисления этого коэффициента.

При вычислении коэффициента корреляции исследователь получает возможность судить о силе связи (степени сопряженности) и ее направлении, а также с требуемой долей вероятности делать вывод о проявлении этой связи в генеральной совокупности. Чем больше коэффициент корреляции, тем с большей степенью уверенности можно говорить о наличии корреляционной зависимости между признаками. Если каждому заданному значению одного признака соответствуют близкие друг к другу, тесно расположенные около средней величины значения другого признака, то связь является более тесной. Когда эти значения сильно варьируют, связь менее тесная. Таким образом, мера корреляции указывает, насколько тесно связаны между собой параметры.

Коэффициент корреляции может принимать значения от -1 до +1. Направление обнаруженной взаимосвязи определяют по знаку коэффициента корреляции. При его положительном значении обнаруженная связь является прямой, при отрицательном – обратной. Сила связи оценивается по модулю этого коэффициента. Условно выделяют следующие уровни корреляционной связи: отсутствие – 0; слабая – от 0 до 0,3; средняя – от 0,3 до 0,7; сильная – 0,7 и более; полная – 1. Однако обсуждать наличие корреляции имеет смысл только в тех случаях, когда она статистически значима ($p < 0,05$). Поэтому после вычисления коэффициента корреляции производится определение его ошибки репрезентативности и критерия достоверности.

Наиболее часто применяемыми в настоящее время методами обнаружения корреляции являются параметрический анализ по Пирсону и непараметрический анализ по Спирмену. Этими методами проверяется нулевая гипотеза (H_0) об отсутствии связи между параметрами. Если такая гипотеза отклоняется при заданном уровне значимости (p), можно говорить о наличии взаимосвязи между параметрами.

Корреляционный анализ по Пирсону используется при решении задачи исследования линейной связи двух нормально распределенных параметров. Кроме проверки на нормальность распределения каждого параметра, до проведения корреляционного анализа рекомендуется строить график в координатах оцениваемых параметров, чтобы визуально определить характер зависимости.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ПИРСОНА (r_{xy}) или коэффициент линейной корреляции, был разработан в 90-х годах XIX века Карлом Пирсоном, Френсисом Эджуортом и Рафаэлем Уэлдоном в Англии. Он рассчитывается по формуле:

$$r_{xy} = \frac{COV_{xy}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{X})(y - \bar{Y})}{\sqrt{\sum (x - \bar{X})^2 \sum (y - \bar{Y})^2}}$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 COV_{xy} – ковариация признаков x и y ;
 σ_x – среднее квадратическое отклонение признака x ;
 σ_y – среднее квадратическое отклонение признака y ;
 \bar{X} – средняя арифметическая признака x ;
 \bar{Y} – средняя арифметическая признака y .

В медицинской литературе встречается упрощенная запись этой формулы:

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}},$$

где: r_{xy} – коэффициент линейной корреляции Пирсона;
 d_x – отклонение каждой варианты признака x от средней этого признака: $d_x = x - M_x$,
 d_y – отклонение каждой варианты признака y от средней этого признака: $d_y = y - M_y$.

В программе Excel значение коэффициент линейной корреляции Пирсона может быть вычислено функцией =КОРРЕЛ(Диапазон ячеек 1-го ряда; Диапазон ячеек 2-го ряда).

Для прогнозирования уровня корреляции в генеральной совокупности определяют ошибку репрезентативности этого коэффициента m_r . Она вычисляется по формуле:

$$m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}},$$

где: m_r – ошибка репрезентативности коэффициента корреляции;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 n – число парных вариантов.

Достоверность коэффициента линейной корреляции оценивается по коэффициенту Стьюдента (t_r), который вычисляется с использованием его ошибки:

$$t_r = \left| \frac{r_{xy}}{m_r} \right|,$$

где: t_r – коэффициент достоверности Стьюдента;
 r_{xy} – коэффициент линейной корреляции Пирсона;
 m_r – ошибка репрезентативности коэффициента корреляции.

Если число парных вариантов $n > 30$, то при $t_r > 2$ связь считается достоверной при уровне значимости $p < 0,05$. Если число парных вариантов $n < 30$, то критическое значение $t_{r-Крит.}$ находят по таблице критических значений Стьюдента при степени свободы $df = n - 2$. В программе Excel это

значение вычисляется функцией =СТЮДРАСПОБР(Уровень значимости p ; Степени свободы df).

С целью уменьшения объема вычислений может применяться функция =КОРРЕЛ(Диапазон1; Диапазон2) или надстройка «Анализ данных» и ее модуль «Корреляционный анализ».

Отсутствие линейной корреляции еще не означает, что параметры полностью независимы. Связь между ними может быть нелинейной, или признаки, используемые в вычислениях, могут не подчиняться нормальному закону распределения. Поэтому, помимо вычисления коэффициента линейной корреляции, прибегают к использованию непараметрических коэффициентов корреляции. К ним относятся:

- Коэффициент ранговой корреляции Спирмена;
- Коэффициент ранговой корреляции Кендалла;
- Коэффициент корреляции знаков Фехнера;
- Коэффициент множественной ранговой корреляции (конкордации).

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ПО СПИРМЕНУ применяется для обнаружения взаимосвязи двух параметров, если распределение хотя бы одного из них отлично от нормального.

Каждому показателю x и y присваивается ранг. На основе полученных рангов рассчитываются их разности d . Затем вычисляется коэффициент корреляции (ρ) по формуле:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)},$$

где: ρ – коэффициент корреляции Спирмена;
 d – разность рангов;
 n – число парных вариантов.

Ошибка репрезентативности коэффициента корреляции Спирмена определяется по формуле:

$$m_\rho = \sqrt{\frac{1 - \rho^2}{n - 2}},$$

а коэффициент достоверности Стьюдента:

$$t_\rho = \left| \frac{\rho}{m_\rho} \right|,$$

где: t_ρ – коэффициент достоверности Стьюдента;
 ρ – коэффициент корреляции Спирмена;
 m_ρ – ошибка репрезентативности коэффициента корреляции Спирмена.

Оценка коэффициента корреляции Спирмена и его достоверности выполняется так же, как и коэффициента линейной корреляции Пирсона.

МЕТОД РЕГРЕССИИ

МЕТОД РЕГРЕССИИ - это статистический способ поиска функции, которая позволяет по величине одного коррелируемого признака судить о величине другого. С помощью метода регрессии решается задача построения модели позволяющей выполнить прогноз количественного изменения одной величины (зависимой переменной) при изменении другой величины (независимой переменной) на единицу. Для выполнения такого прогноза требуется определить коэффициент корреляции Пирсона, с использованием которого вычисляют коэффициент регрессии ($R_{y/x}$). Он участвует в создании линейной регрессионной функции вида $y=ax+b$, которая применяется для прогнозирования требуемых параметров.

Коэффициент регрессии вычисляется по формуле:

$$R_{y/x} = r_{x/y} \frac{\sigma_y}{\sigma_x},$$

где: $R_{y/x}$ – коэффициент регрессии;

$r_{x/y}$ – коэффициент корреляции Пирсона;

σ_x – среднее квадратическое отклонение признака x ;

σ_y – среднее квадратическое отклонение признака y .

Среднее квадратическое отклонение (сигма) вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}},$$

а в программе Excel функцией =СТАНДОТКЛОН(Диапазон ячеек).

Значение коэффициента регрессии ($R_{y/x}$) в программе Excel может быть вычислено функцией =НАКЛОН(Диапазон_y; Диапазон_x).

Формула определения значения зависимого признака:

$$y = M_y + R_{y/x} (x - M_x),$$

где: y – зависимая переменная;

M_y – средняя признака y ;

$R_{y/x}$ – коэффициент регрессии;

x – значение измеренного признака;

M_x – средняя арифметическая признака x .

В программе Excel значение зависимой переменной (y) при заданном значении x может быть вычислено функцией =ПРЕДСКАЗ(x ; Диапазон_у; Диапазон_х).

После получения прогнозируемого значения (y) выполняется определение его доверительного интервала с целью экстраполяции данных на генеральную совокупность с уровнем значимости $p < 0,05$. Для этого вычисляется сигма регрессии $\sigma_{Ry/x}$, которая показывает меру вариабельности зависимого признака, вычисленного по уравнению регрессии, в генеральной совокупности.

Она определяется по формуле:

$$\sigma_{Ry/x} = \sigma_y \sqrt{1 - r_{xy}^2}.$$

Вычисление значения σ_y может производиться функцией =СТАНДОТКЛОН(Диапазон_у).

Доверительные границы колебания зависимого признака в генеральной совокупности, как правило, вычисляют для уровня значимости $p < 0,05$. Для определения нижнего значения доверительной границы от значения зависимого признака вычитают удвоенную сигму регрессии, а верхней граница – прибавляют удвоенную сигму.

Пример решения задачи на выявление корреляционной зависимости

Условие задачи: выполнены измерения уровня запыленности на рабочих местах работников с учетом температуры в помещении (таблица 1).

Таблица 1. Результаты измерений запыленности на рабочем месте и температуры воздуха

Измерение на рабочем месте	Температура воздуха С°	Запыленность мг/м ³
1. Слесарь	20	0,2
2. Электрик	21	0,25
3. Сварщик	21	0,24
4. ...	19	0,08
5. ...	19	0,08
6. ...	19	0,07
7. ...	22	0,3
8. ...	22	0,28
9. ...	25	0,33
10....	24	0,31
11....	21	0,26
12....	21	0,27

Задание: определите силу и направление зависимости между температурой окружающей среды и уровнем запыленности помещения с помощью:

- таблицы;
- графического изображения взаимосвязи между признаками;

в) коэффициента корреляции Пирсона;

Решение: необходимо запустить программу Excel, создать новый лист, на этом листе ввести данные в таблицу, создать диаграмму и выполнить расчеты, как показано ниже, сохранить файл.

а) Для определения связи между признаками требуется поместить вариационные ряды рядом друг с другом и выполнить совместное ранжирование по одному из признаков, как показано в таблице 2, а затем проанализировать совместное возрастание или убывание значений признаков.

Таблица 2. Ранжирование данных в таблице.

№ варианты	Параметр x (температура воздуха, $^{\circ}\text{C}$)	Параметр y (запыленность, $\text{мг}/\text{м}^3$)
1	19	0,07
2	19	0,08
3	19	0,08
4	20	0,2
5	21	0,24
6	21	0,25
7	21	0,26
8	21	0,27
9	22	0,3
10	22	0,28
11	24	0,31
12	25	0,33

Вывод: в таблице наблюдается совместное прямое изменение двух изучаемых параметров.

б) С помощью графика может быть выполнена визуальная оценка формы взаимосвязи между признаками. Для этого в программе Excel необходимо создать точечную диаграмму с прямыми или гладкими отрезками между маркерами, как показано на рис. 1.

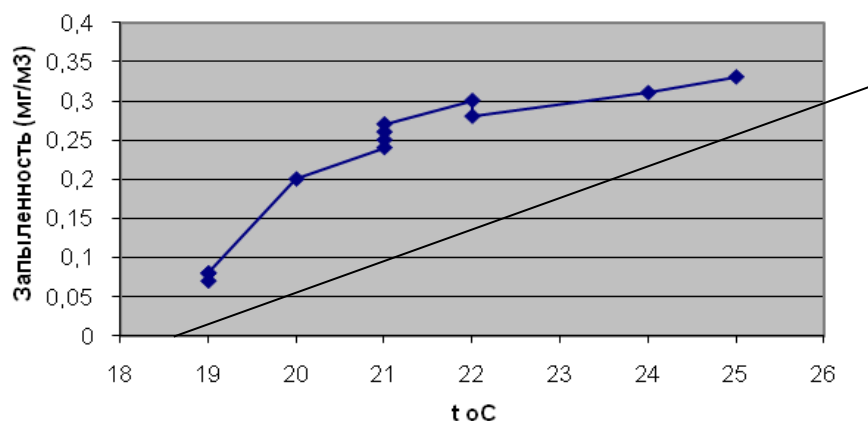


Рис. 1. Зависимость уровня запыленности от температуры воздуха в помещении.

Вывод: график зависимости совместного изменения двух изучаемых параметров показывает наличие взаимосвязи, которая приближенно оценивается как линейная.

в) Вычисление и оценка коэффициента корреляции методом Пирсона (таблица 3).

Таблица 3. Вычисление отклонений вариант от средней арифметической

Варианта №	Температура воздуха (x)	Запыленность мг/м ³ (y)	$d_x = x - M_x$	$d_y = y - M_y$	$d_x * d_y$	d_x^2	d_y^2
1	19	0,07	-2,2	-0,153	0,330	4,7	0,0233
2	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
3	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
4	20	0,2	-1,2	-0,023	0,026	1,4	0,0005
5	21	0,24	-0,2	0,018	-0,003	0,0	0,0003
6	21	0,25	-0,2	0,028	-0,005	0,0	0,0008
7	21	0,26	-0,2	0,038	-0,006	0,0	0,0014
8	21	0,27	-0,2	0,048	-0,008	0,0	0,0023
9	22	0,3	0,8	0,078	0,065	0,7	0,0060
10	22	0,28	0,8	0,058	0,048	0,7	0,0033
11	24	0,31	2,8	0,088	0,248	8,0	0,0077
12	25	0,33	3,8	0,108	0,412	14,7	0,0116
Средняя (M) = 21,2 n= 12			0,223	Сумма (Σ) = 1,725 39,7 0,0976			

Коэффициент корреляции вычисляется по формуле:

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{1,725}{\sqrt{39,7 \times 0,0976}} = \underline{\underline{0,88}}.$$

В программе Excel может использоваться функция =КОРРЕЛ(Диапазон1;Диапазон2) или модуль «Корреляция» из пакета анализа, который вызывается командой «Данные» - «Анализ данных». Он производит создание таблицы «Корреляционная матрица». В ней по диагонали указаны значения коэффициента корреляции между одинаковыми признаками, выражающиеся цифрой 1, и между разными признаками в нижней части матрицы. При использовании последнего способа расчета имеется возможность вычислить коэффициенты корреляции одновременно для нескольких признаков, если выделенный диапазон данных будут включены более двух столбцов. Результат вычислений, выполненный с помощью указанного модуля, приведен в таблице 4.

Таблица 4. Вычисление корреляционной матрицы модулем «Корреляция» в программе Excel

	Температура воздуха (x)	Запыленность мг/м ³ (y)
Температура воздуха (x)	1	
Запыленность мг/м ³ (y)	0,876588407	1

Оценка достоверности коэффициента корреляции выполняется с помощью ошибки репрезентативности корреляции (m_r) и критерия Стьюдента (t_r).

Ошибка репрезентативности корреляции вычисляется формулой:

$$m_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0,88^2}{12-2}} = \underline{0,152}, \text{ где: } n - \text{число парных вариантов.}$$

Критерий достоверности Стьюдента для коэффициента корреляции рассчитывается по формуле:

$$t_r = \left| \frac{r_{xy}}{m_r} \right| = \frac{0,88}{0,152} = \underline{5,8} > 2.$$

Сравнивая полученное значение критерия Стьюдента с критическим значением 2 можно оценить достоверность установленной корреляционной зависимости. При величине $t_r > 2$ она считается достоверной с уровнем значимости $p < 0,05$.

Вывод: зависимость изменения двух изучаемых параметров является сильной прямой и статистически достоверной при уровне значимости $p < 0,05$.

Пример прогнозирования значения признака по известному значению другого с помощью модели регрессии

Условие задачи: на основе данных, характеризующих уровень запыленности рабочих мест (см. предыдущий пример), необходимо выполнить прогноз уровня пыли при температуре воздуха 23С°.

Задание: получить уравнение регрессии для зависимости между температурой окружающей среды и уровнем запыленности помещения, построить регрессионную функцию и вычислить значение уровня пыли при температуре воздуха 23С°. Определить сигму регрессии и доверительный интервал для прогнозируемого значения уровня пыли.

Решение: необходимо запустить программу Excel, создать новый лист, на этом листе ввести данные и выполнить требуемые этапы вычислений, как показано на рис. 2, сохранить файл.

	A	B	C	D	E	F	G	H	I	J	K	L
1	1) Вычисление коэффицента корреляции Пирсона											
2	Варианта	Температура воздуха (x)	Запыленность мг/м3 (y)	d _x	d _y	d _x *d _y	d _x ²	d _y ²				
3	1	19	0,07	2,2	0,153	0,330	4,7	0,0233				
4	2	19	0,08	2,2	0,143	0,309	4,7	0,0203				
5	3	19	0,08	2,2	0,143	0,309	4,7	0,0203				
6	4	20	0,2	1,2	0,023	0,026	1,4	0,0005				
7	5	21	0,24	0,2	-0,018	-0,003	0,0	0,0003				
8	6	21	0,25	0,2	-0,028	-0,005	0,0	0,0008				
9	7	21	0,26	0,2	-0,038	-0,006	0,0	0,0014				
10	8	21	0,27	0,2	-0,048	-0,008	0,0	0,0023				
11	9	22	0,3	-0,8	-0,078	0,065	0,7	0,0060				
12	10	22	0,28	-0,8	-0,058	0,048	0,7	0,0033				
13	11	24	0,31	-2,8	-0,088	0,248	8,0	0,0077				
14	12	25	0,33	-3,8	-0,108	0,412	14,7	0,0116				
15	Средняя (M) =	21,2	0,223	Сумма (Σ) =		1,725	39,7	0,0976				
16	n=	12										
17												
18	$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$		$r_{xy} =$	0,88	=КОРРЕЛ(B3:B14;C3:C14)=			0,876588				
19	2) Вычисление коэффицента регрессии:											
20		$\sigma = \sqrt{\frac{\sum d^2}{n-1}}$	$\delta_x =$	1,90	=КОРЕНЬ(39,7/(12-1))=			1,9	=СТАНДОТКЛОН(B3:B14)=		1,9	
21			$\delta_y =$	0,09	=КОРЕНЬ(0,0976/(12-1))=			0,094	=СТАНДОТКЛОН(C3:C14)=		0,094	
22		$R_{x/y} = r_{xy} * \delta_y / \delta_x$	$R_{x/y} =$	0,043	=0,88*0,094/1,9=			0,044	=D18*D21/D20=		0,043	
23	3) Вычичление значения зависимого признака:											
24	$y = M_y + R_{x/y}(x - M_x)$		x =	23								
25			y =	0,302	=0,223+0,04(23-21,2)=				=ПРЕДСКАЗ(D24; y1:yn; x1:xn)		0,302227	
26	4) Вычисление доверительных границ зависимого признака:											
27	$\delta_{R_{x/y}} = \delta_y * \sqrt{1 - r_{xy}^2}$		$\delta_{R_{x/y}} =$	0,045	=0,094*КОРЕНЬ(1-0,88^2)=			0,045	=D21*КОРЕНЬ(1-D18^2)=		0,045	
28	Доверительные границы (26):		0,212	0,393								
29												
30	Вывод: при температуре воздуха 23С° запыленность составит 0,302±0,045											

Рис. 2. Регрессионный анализ в программе Excel

Вычисления выполняются в четыре этапа:

1) уравнение регрессии может быть получено после вычисления коэффициента корреляции Пирсона с помощью таблицы отклонений (таблица 5) или функцией =КОРРЕЛ(Диапазон1;Диапазон2). При этом Диапазон1 должен включать адреса ячеек, где расположены значения переменной x , а Диапазон2 - переменной y , и отделяться друг от друга точкой с запятой.

Таблица 5. Вычисление коэффициента корреляции Пирсона

Варианта	Температура воздуха (x)	Запыленность мг/м ³ (y)	$d_x = x - M_x$	$d_y = y - M_y$	$d_x * d_y$	d_x^2	d_y^2
1	19	0,07	-2,2	-0,153	0,330	4,7	0,0233
2	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
3	19	0,08	-2,2	-0,143	0,309	4,7	0,0203
4	20	0,2	-1,2	-0,023	0,026	1,4	0,0005
5	21	0,24	-0,2	0,018	-0,003	0,0	0,0003
6	21	0,25	-0,2	0,028	-0,005	0,0	0,0008
7	21	0,26	-0,2	0,038	-0,006	0,0	0,0014
8	21	0,27	-0,2	0,048	-0,008	0,0	0,0023
9	22	0,3	0,8	0,078	0,065	0,7	0,0060
10	22	0,28	0,8	0,058	0,048	0,7	0,0033
11	24	0,31	2,8	0,088	0,248	8,0	0,0077
12	25	0,33	3,8	0,108	0,412	14,7	0,0116

Средняя(M) = 21,2 0,223 Сумма (Σ) = 1,725 39,7 0,0976
n= 12

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{1,725}{\sqrt{39,7 \times 0,0976}} = \text{КОРРЕЛ}(x_1:x_n; y_1:y_n) = \underline{0,88}.$$

2) вычисление коэффициента регрессии ($R_{y/x}$) производится с использованием среднего квадратического отклонения (σ) и вычисленного ранее коэффициента корреляции ($r_{x/y}$):

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}}.$$

$$\sigma_x = \text{КОРЕНЬ}(39,7/(12-1)) = \text{СТАНДОТКЛОН}(x_1:x_n) = 1,9.$$

$$\sigma_y = \text{КОРЕНЬ}(0,0976/(12-1)) = \text{СТАНДОТКЛОН}(y_1:y_n) = 0,094.$$

Подставив значения в формулу вычисления коэффициента регрессии:

$$R_{y/x} = r_{x/y} \frac{\sigma_y}{\sigma_x},$$

получим:

$$R_{y/x} = 0,88 \times 0,094 / 1,9 = \underline{0,04}.$$

3) вычисление величины зависимого признака (y) при температуре 23C^0 :

$$y = M_y + R_{y/x}(x - M_x) \text{ При } x = 23\text{C}^0$$

$$y = 0,223 + 0,04(23 - 21,2) = \text{ПРЕДСКАЗ}(x; y_1:y_n; x_1:x_n) = \underline{0,30} \text{ мг/м}^3$$

4) вычисление сигмы регрессии ($\sigma_{Ry/x}$) и доверительных границ колебаний зависимого признака в генеральной совокупности с уровнем значимости $p < 0,05$ ($y \pm 2\sigma_{Ry/x}$).

$$\sigma_{Ry/x} = \sigma_y \sqrt{1 - r_{xy}^2} \quad \sigma_{Ry/x} = 0,094 \times \text{КОРЕНЬ}(1 - 0,88^2) = \underline{0,045} \text{ мг/м}^3.$$

Доверительные границы прогнозируемого значения y при $p < 0,05$ определяются как $y \pm 2\sigma_{Ry/x}$. То есть $0,3 \pm 0,09 \text{ мг/м}^3$, или от $0,3 - 2 \times 0,045 = \underline{0,21}$ до $0,3 + 2 \times 0,045 = \underline{0,39} \text{ мг/м}^3$.

Вывод: при температуре воздуха 23C^0 запыленность составит от $0,21 \text{ мг/м}^3$ до $0,39 \text{ мг/м}^3$.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ АППРОКСИМАЦИИ

Для создания модели аппроксимации требуется подготовить таблицу, в которой для некоторых дискретных значений аргумента x_i , расположенных в порядке возрастания, заданы соответствующие значения функции y_i . На графике, построенном по этим данным, точки с координатами (x_i, y_i) называются узловыми точками или узлами.

В графическом виде такая табличная функция представляется как совокупность узловых точек, как показано на рисунке 3.

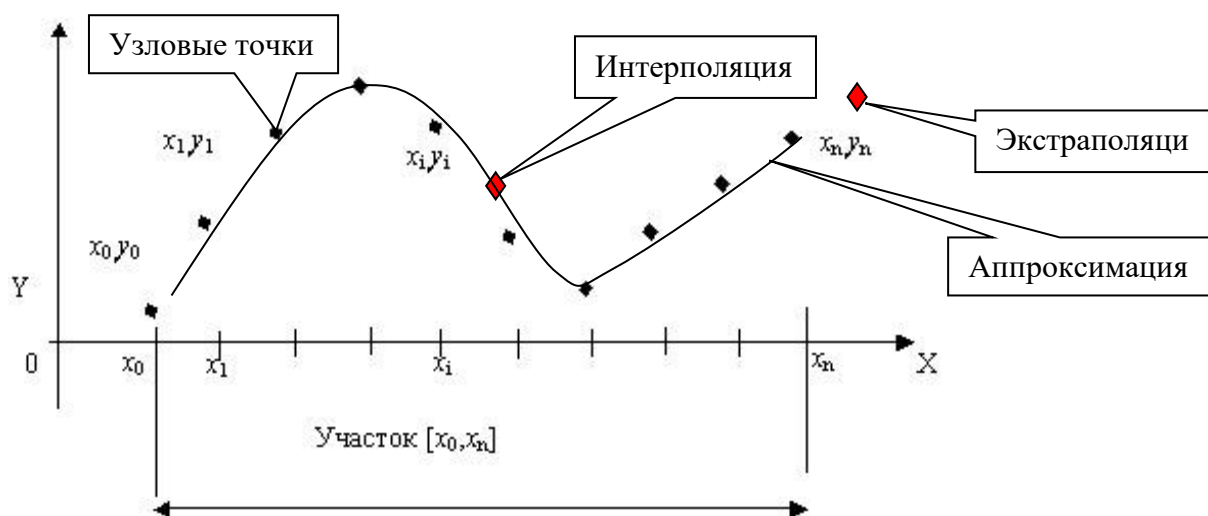


Рис. 3. Диаграмма узловых точек

В практической деятельности достаточно часто возникают задачи нахождения значений табличной функции для аргументов (узлов), которые отсутствуют в таблице. Такие задачи называются задачами интерполирования или экстраполирования.

Задача интерполирования (интерполяции) функции состоит в том, чтобы найти значения y_k табличной функции в любой промежуточной точке x_k , расположенной внутри интервала $[x_0, x_n]$.

Задача экстраполирования (экстраполяции) функции состоит в том, чтобы найти значения y_1 табличной функции в точке x_1 , которая не входит в интервал $[x_0, x_n]$. Такую задачу часто называют задачей прогноза.

Эти задачи решаются путем подбора эмпирической формулы $P(x)$ для опытной зависимости $F(x)$, что называется аппроксимацией. Для этого на основе опытных данных строится график функции $F(x)$ и к нему подбирается аппроксимирующая функция $P(x)$, которая с максимальной степенью близости приближается к опытной зависимости. График этой функции называется «Линия тренда».

Степень близости подбираемой линии тренда к экспериментальной зависимости оценивается коэффициентом детерминации « R^2 ». В модели парной линейной регрессии коэффициент детерминации R^2 равен квадрату обычного коэффициента корреляции r между переменными y и x . Его рассматривают как универсальную меру связи одной случайной величины от множества других. Если нет других теоретических соображений, то выбирают функцию с коэффициентом R^2 , стремящимся к 1. Полученная

линия тренда позволяет установить как вид эмпирической формулы, так и численные значения неизвестных параметров.

В современных условиях все вычисления могут выполняться в любом табличном редакторе. В частности, может использоваться программа Microsoft Excel. В ней используется аппроксимация данных по методу наименьших квадратов (МНК, OLS, Ordinary Least Squares). Метод наименьших квадратов - это один из базовых методов регрессионного анализа основанный на минимизации суммы квадратов остатков регрессии.

Для аппроксимации зависимостей с одним неизвестным в программе Excel используются графики, а для зависимостей со многими неизвестными – пары функций из группы Статистические ЛИНЕЙН() и ТЕНДЕНЦИЯ(), ЛГРФПРИБЛ() и РОСТ().

Программа Excel позволяет выбирать 5 видов функций аппроксимации:

- **Линейную** – $y = cx + b$. Это простейшая функция, отражающая рост и убывание данных с постоянной скоростью.

- **Полиномиальную** – $y = c_0 + c_1x^1 + c_2x^2 + \dots + c_6x^6$. Функция описывает попеременно возрастающие и убывающие данные. Полином 2-ой степени может иметь один экстремум (min или max), 3-ей степени – до 2-х экстремумов, 4-ой степени – до 3-х и т.д.

- **Логарифмическую** – $y = c \ln x + b$. Эта функция описывает быстро возрастающие (убывающие) данные, которые затем стабилизируются.

- **Степенную** – $y = cx^b$, ($x > 0$ и $y > 0$). Функция отражает данные с постоянно увеличивающейся (убывающей) скоростью роста.

- **Экспоненциальную** – $y = ce^{bx}$, (e – основание натурального логарифма). Функция описывает быстро растущие (убывающие) данные, которые затем стабилизируются.

При выборе функции аппроксимации может оцениваться ее форма, которая должна максимально приближаться к экспериментальной, а также наивысшее значение коэффициента детерминации R^2 . С целью вычисления прогнозируемых значений с использованием уравнения аппроксимации в задачах экстраполяции требуется оценить вид кривой за пределами экспериментальных точек в большую (вперёд) или меньшую (назад) сторону. Форма кривой должна отражать моделируемый процесс в соответствии с данными, полученными на практике в ранее выполненных экспериментах или известных по литературным источникам.

Пример 1. Построение математической модели объема продаж лекарства. Рассмотрим зависимость объема продаж лекарственного препарата от затрат на его рекламу, заданную данными табличной функции (таблица 6).

Таблица 6. Объем продаж лекарственного препарата в зависимости от затрат на рекламу

Реклама (тыс. руб.)	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
Продажи (тыс. руб.)	30	130	250	350	400	450	480	500	510	?

Требуется найти функцию, наилучшим образом отражающую эту зависимость. Кроме того, необходимо спрогнозировать объем продаж для рекламных вложений в 6 тысяч руб.

Решение. Необходимо запустить программу Excel, внести название таблицы и данные из условия задачи в диапазон адресов ячеек В3:К4. Используя этот диапазон, создается **точечная** диаграмма. Затем щелчком правой клавиши мыши на любом узле выполняется добавление в диаграмму линии тренда, как показано на рис. 4.

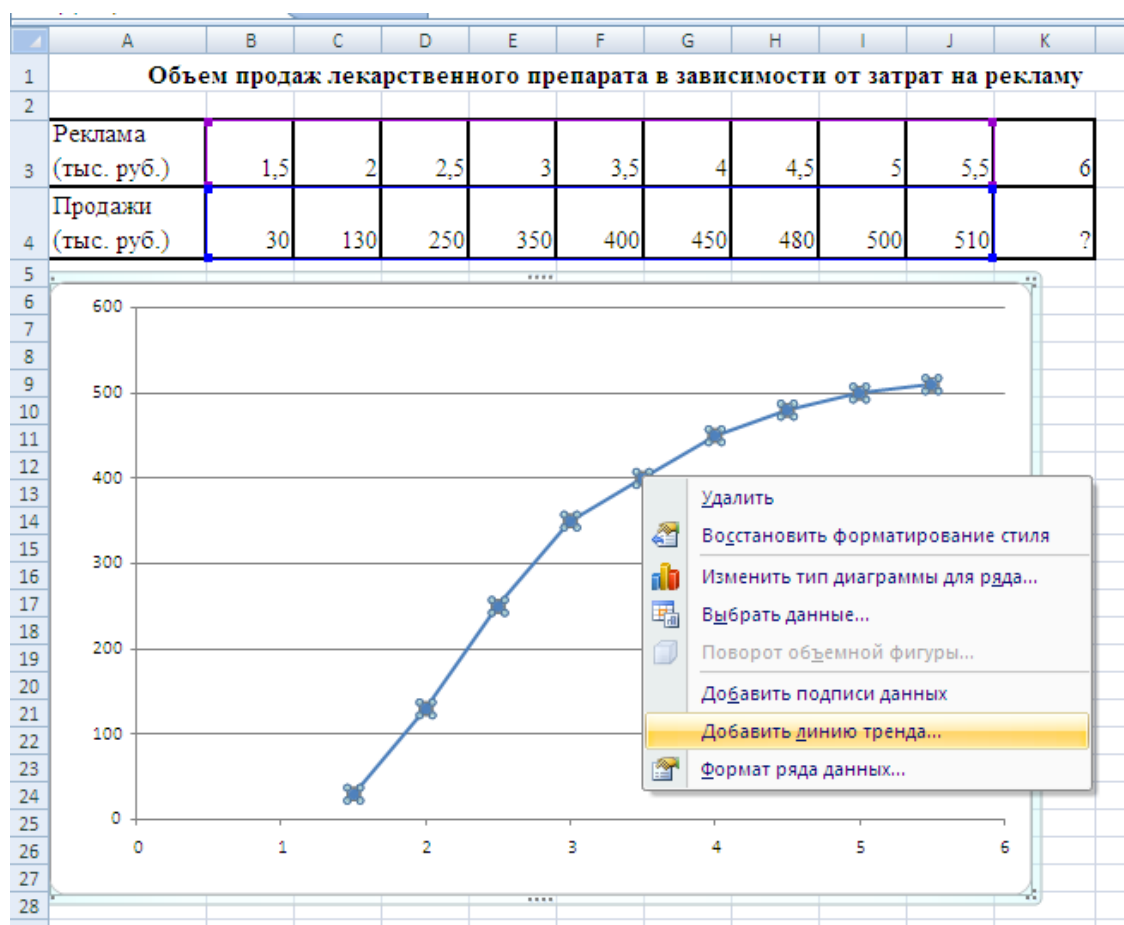


Рис. 4. Добавление линии тренда в диаграмму

Для выполнения следующих действий требуется переместить открывшееся окно «Формат линии тренда» так, чтобы рисунок диаграммы

был виден полностью (рис. 5), а затем последовательно изменить отметку вариантов вида линии тренда, обращая внимание на ее форму. Например, при выборе «Логарифмическая» и установке флажков, отображающих на графике уравнение функции аппроксимации и коэффициент детерминации (величина достоверности аппроксимации), на рисунке будут показаны эти сведения.

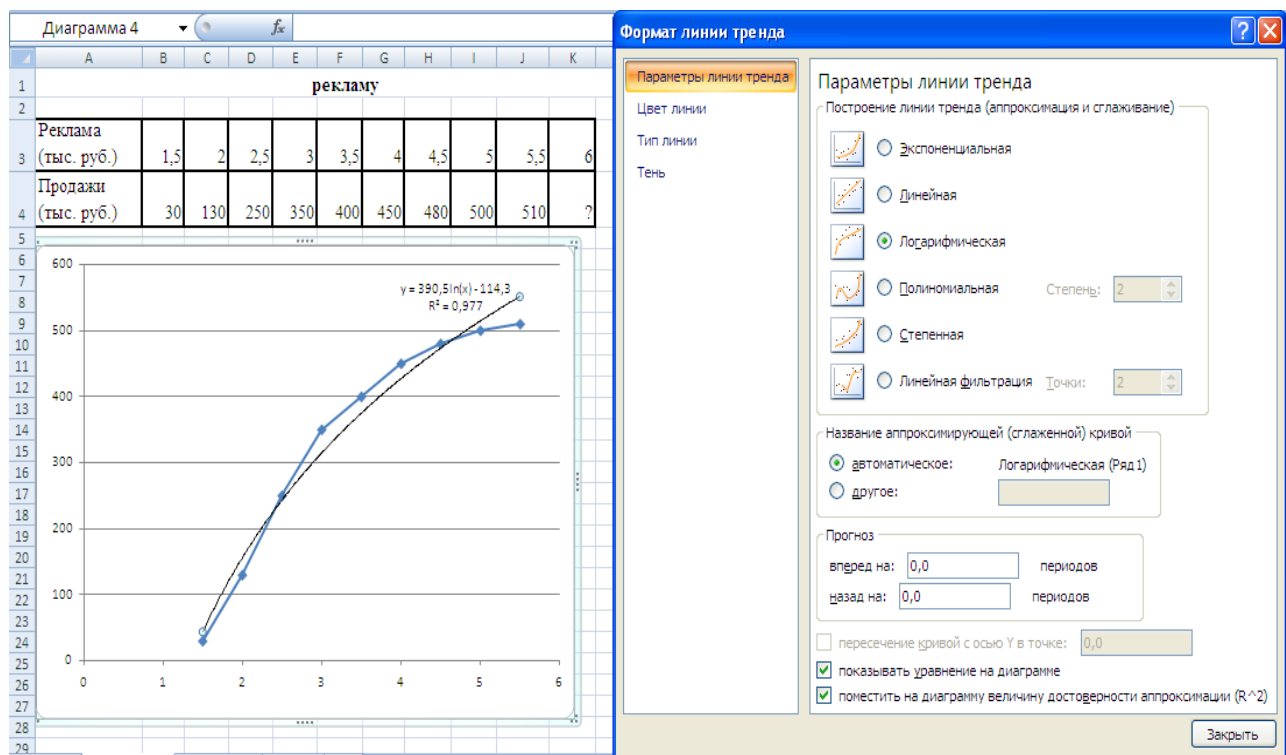


Рис. 5. Выбор вида функции аппроксимации

На рисунке диаграммы отображаются уравнение аппроксимации и коэффициент детерминации (достоверности аппроксимации). Его величина $R^2=0,977$ является достаточно высокой степенью близости выбранной функции аппроксимации к экспериментальным данным. Вид функции аппроксимации выбирается по критериям близости кривой к экспериментальным точкам, а также максимального значения коэффициента детерминации R^2 .

Прогноз величины продаж выполняется двумя способами. Приблизительную оценку можно получить с помощью продолжения линии тренда на несколько периодов по оси X, как в большую сторону (вперёд), так и в меньшую (назад). Например, в области окна «Прогноз» в поле «вперед на:» можно ввести цифру 5. После щелчка мыши в следующем поле на рисунке будет показано продолжение кривой аппроксимации. Она отражает дальнейшее изменение продаж при увеличении затрат на рекламу (рис. 6) в следующих 5 периодах.

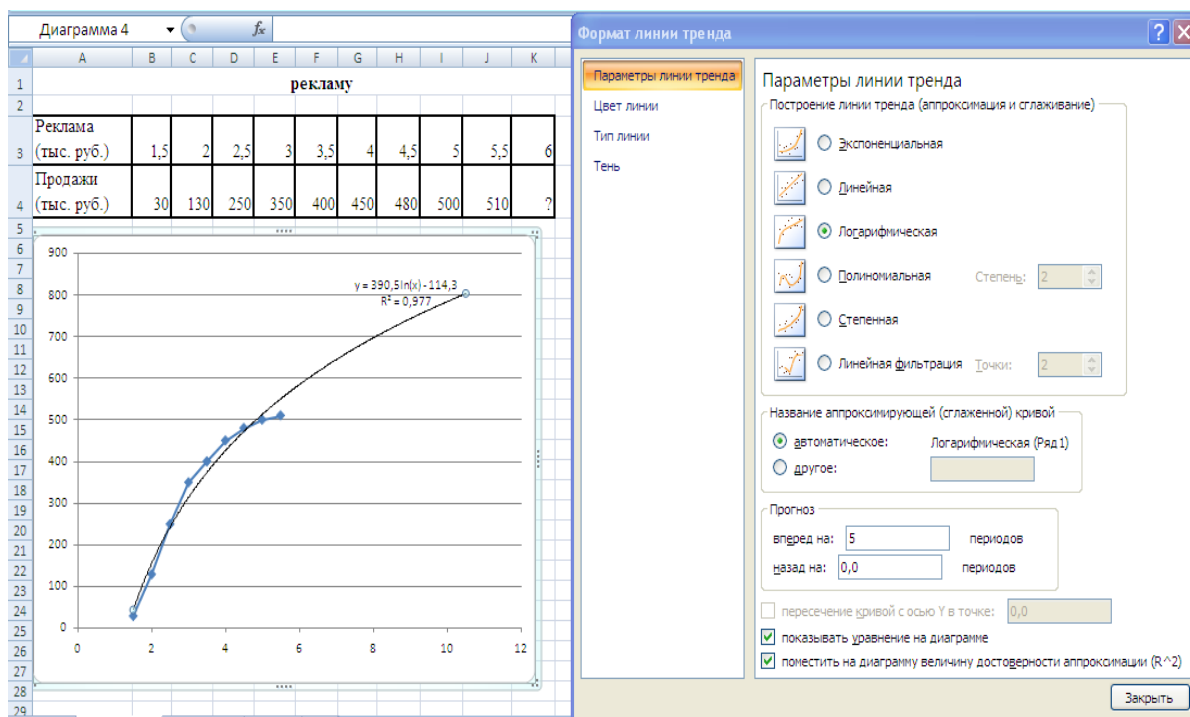


Рис. 6. Продолжение кривой аппроксимации

Вычисление объема продаж при вложении в рекламу 6 тысяч рублей производится в ячейке K4, где требуется создать формулу на основе указанного в диаграмме уравнения аппроксимации « $= 390,5 * \ln(K3) - 114,3$ ». В этой ячейке появится результат 585,4 тысячи рублей, как показано на рис. 7.

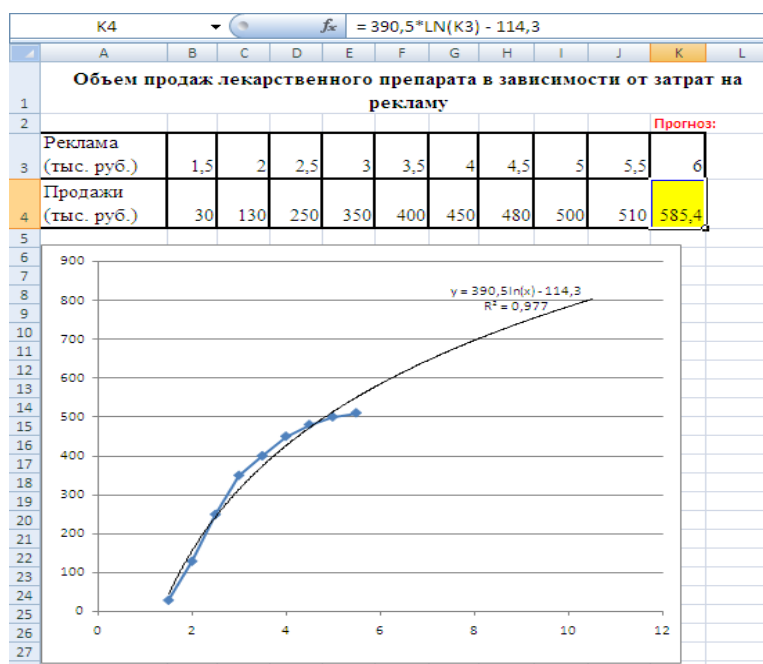


Рис. 7. Прогноз продаж при затратах на рекламу 6 тысяч рублей

В некоторых задачах при выборе линии тренда возможно использование модели линейной зависимости вида $y = ax + b$. Тогда для

определения прогнозируемых значений может применяться функция ПРЕДСКАЗ(х; ряд значений у; ряд значений х), которая вычисляет для величины х прогнозируемое значение Y по существующим рядам значений х и у с использованием уравнения линейной регрессии.

Изменение типа линии тренда может быть выполнено щелчком правой клавиши мыши на этой линии с последующей модификацией параметров аппроксимации в диалоговом окне Формат линии тренда, как показано на рисунке 8. В нашем примере при выборе более грубой математической модели линейной функции аппроксимации $y = ax + b$ будет показано уравнение $y = 119,67x - 74,389$ и значение $R^2 = 0,9024$. Низкое значение $R^2 = 0,9024$ указывает на недостаточную для медицинских задач вероятность прогноза, которая составляет $R^2 > 0,95$, и может использоваться только для приблизительной оценки. В то же время линейная функция упрощает вычисления прогнозируемых значений.

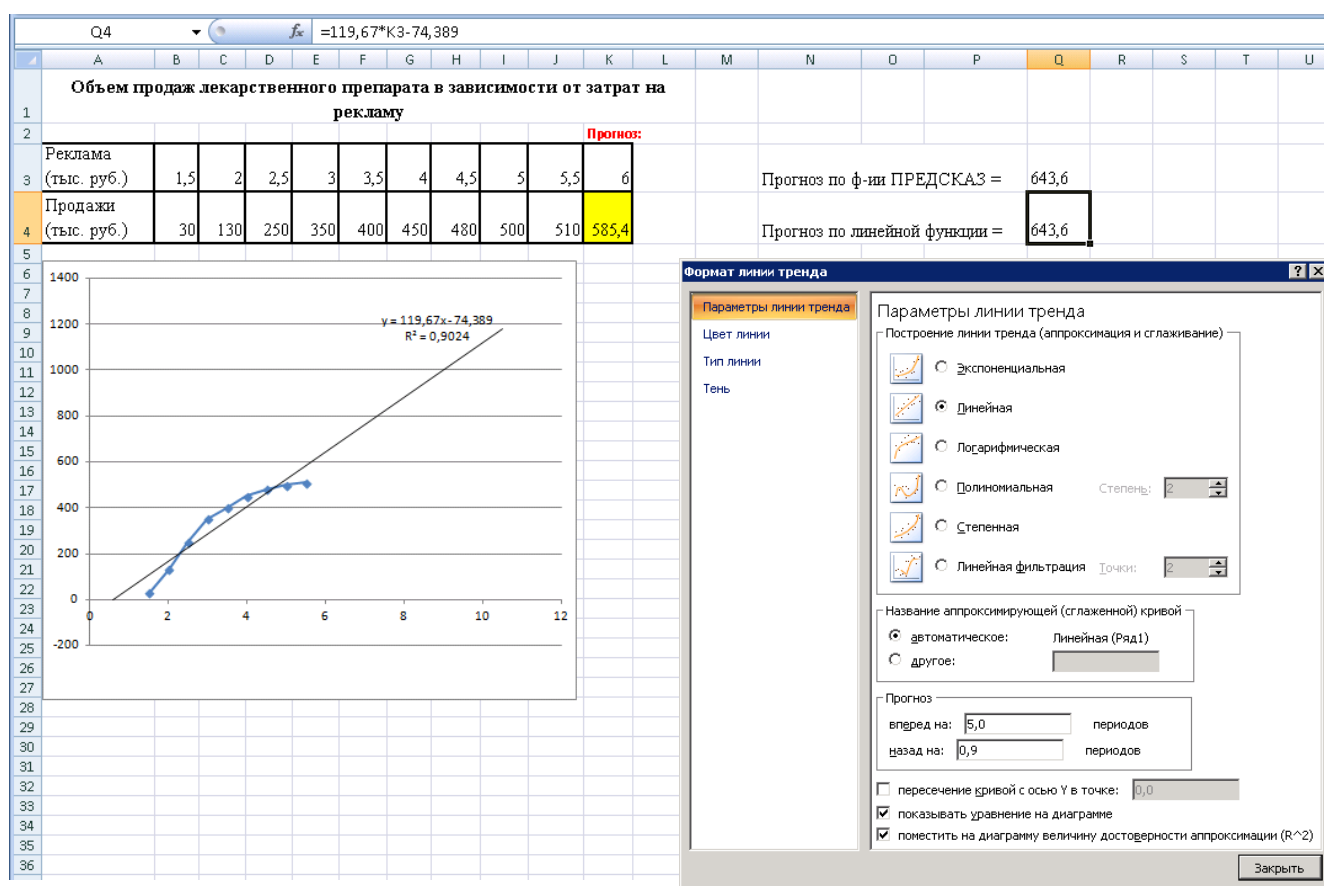


Рис. 8. Прогноз продаж на основе линейной зависимости

Формула вычисления прогноза по математической модели линейной функции тренда для объема продаж при 6 тысячах рублей вложений запишется как =119,67*K3-74,389 или =ПРЕДСКАЗ(K3;B4:J4;B3:J3). Результат вычисления составит 643,6 тысячи рублей, что значительно выше, чем прогноз по логарифмической модели (584,4). Значит линейная

модель не может использоваться в представленной задаче для вычисления прогнозируемых значений.

Вывод: зависимость объема продаж лекарственного препарата от затрат на его рекламу может быть описано математической моделью логарифмической функции с достаточной вероятностью ($R^2 > 0,95$). Прогноз объема продаж лекарства при затратах 6 тысяч рублей составляет 585,4 тысячи рублей.

Пример 2. Создание математической модели эпидемического процесса и прогноз его завершения.

Получены экспериментальные данные о распространенности заболевания N выражающегося числом заболевших за неделю наблюдения по срокам его развития (Таблица 7).

Таблица 7. Число заболевших по периодам наблюдения

Период наблюдения, неделя	Число заболевших лиц
1	2850
2	2450
3	2121
4	1545
5	1124
6	950
7	841
8	812
9	751

Требуется определить вид функциональной зависимости эпидемического процесса и найти период, когда число заболевших будет близким к 0.

Решение. Необходимо запустить программу Excel, внести краткое описание задачи в первую строку и данные из условия задачи в диапазон адресов ячеек A2:B8.

Внимание! Определение уравнения функции аппроксимации в программе Excel производится на основе данных близких к единице. Если размерность значений большая (более сотни) или маленькая (сотые или тысячные) необходимо, предварительно привести их к требуемым величинам. Поэтому в нашем примере необходимо выразить число заболевших единицей измерения «Тысяч человек». С этой целью в таблицу требуется добавить столбец с вычислением в нем значений частного от деления числа заболевших на 1000, как показано в таблице 8.

Таблица 8. Число заболевших по периодам наблюдения

Период наблюдения, неделя	Число заболевших, тыс.чел.	Число заболевших лиц
1	2,85	2850
2	2,45	2450
3	2,121	2121
4	1,545	1545
5	1,124	1124
6	0,95	950
7	0,841	841
8	0,812	812
9	0,751	751

Для построения диаграммы в программе Excel необходимо разместить данные в рядом расположенных столбцах. Требуется перенести данные диапазона ячеек B2:B8 в диапазон C2:C8, а затем вычислить требуемые значения в диапазоне ячеек B2:B8.

Затем, используя диапазон A2:B8, создается **точечная** диаграмма в которой щелчком правой клавиши мыши на любом узле выполняется добавление линии тренда, как показано на рисунке 9.

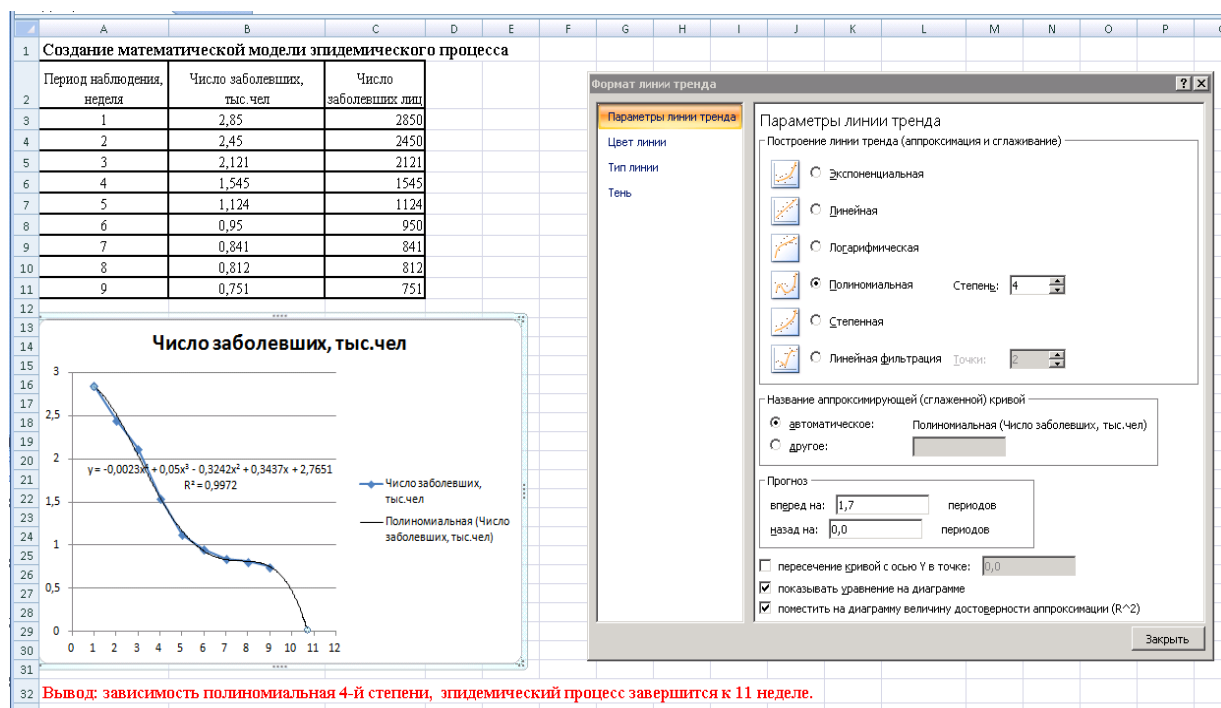


Рис. 9. Прогноз периода завершения распространения заболевания

В появившемся диалоговом окне **Формат линии тренда** необходимо установить отметки для отображения на диаграмме уравнения аппроксимации и коэффициента детерминации R^2 (достоверности аппроксимации) соответствующими отметками в нижней части окна. Для подбора функции аппроксимации изменяя выбор типа функции в верхней

части этого окна подбирают форму кривой, максимально совпадающую с экспериментальными данными, а также учитывающую реальную информацию о моделируемом процессе. При этом необходимо оценивать значение коэффициента детерминации R^2 , которое показывает степень соответствия линии тренда и данных эксперимента. Как видно из рисунка 7 наиболее близко отражает развитие эпидемического процесса полиномиальная функция 4-й степени. А также, величина коэффициент детерминации $R^2=0,9972$ соответствует высокой степени близости выбранной функции аппроксимации к экспериментальным данным.

Прогноз снижения числа заболевших до 0 выполняется с помощью продолжения линии тренда. Для этого указывается число периодов, на которые должна быть продолжена линия аппроксимации. Например, в области окна «Прогноз» в поле «вперед на:» можно ввести цифру 1,7 которая покажет на рисунке точку пересечения кривой линии тренда с горизонтальной осью рядом с значением 11 недель.

Однако, при продолжении линии тренда в противоположном направлении (назад) на 1 период обнаруживается снижение линии аппроксимации до 2,7 тыс. чел. в начале эпидемического процесса (рис. 10), что противоречит литературным сведениям о законах развития эпидемий. Значит, функция полинома 4 степени может использоваться только для задач интерполяции, она не применима в случае прогноза за пределами экспериментальных значений.

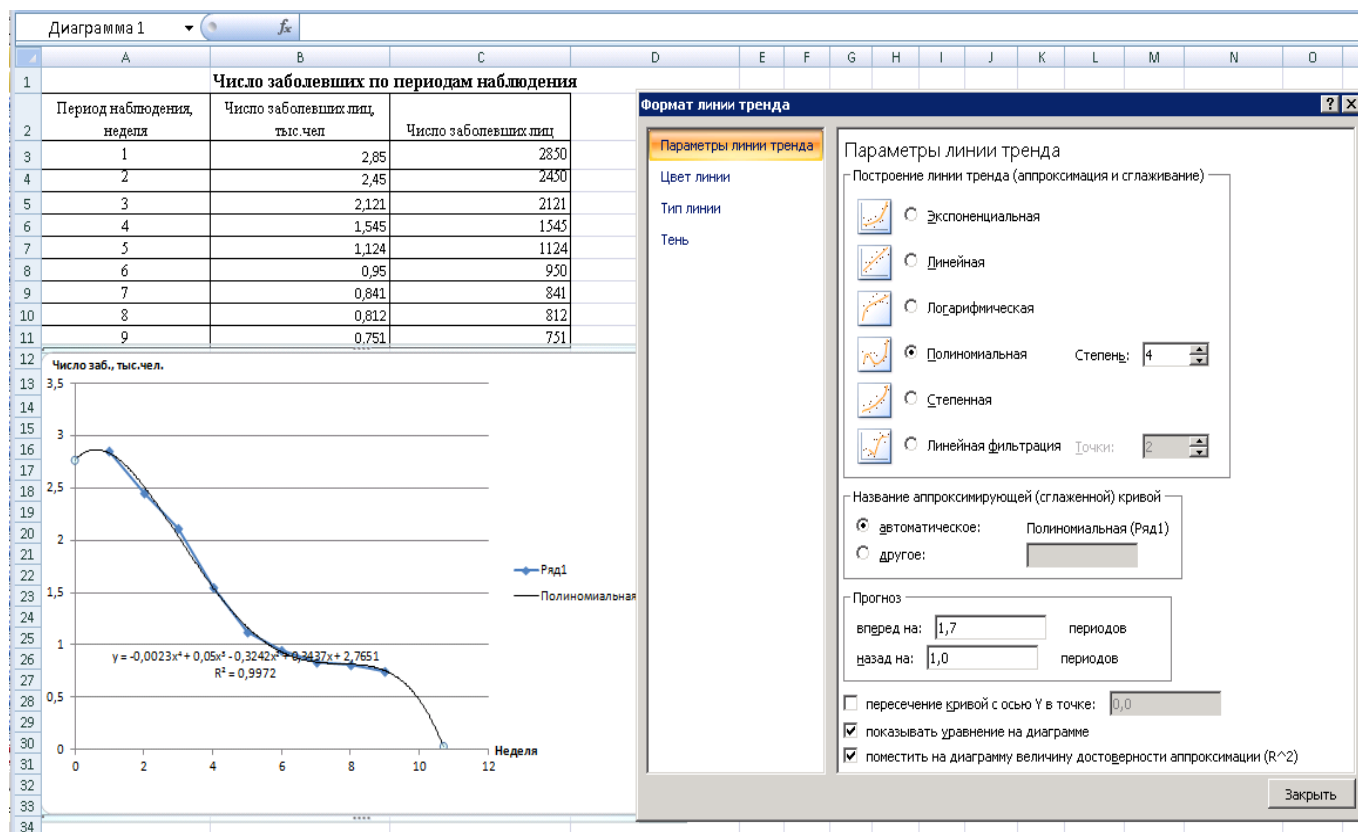


Рис. 10. Прогноз числа заболевших в начале распространения заболевания

В случае выбора экспоненциальной функции аппроксимации коэффициент детерминации будет ниже, но вполне достаточным для уверенного прогноза $R^2 = 0,9503$ (рис. 11). Предварительная оценка на основе диаграммы прогноза периода завершения распространения заболевания по экспоненциальному уравнению может выражаться значением 30 недель. Более точное значение можно получить с помощью уравнения, в котором требуется вычислить значение x приближающегося к 0. В нашем примере формула для вычисления этого значения принимает вид: $= 3,3077 * \text{EXP}(1)^{(-0,183 * A14)}$. Тогда число недель выражается значением 50, которое вычислено в ячейке C14.

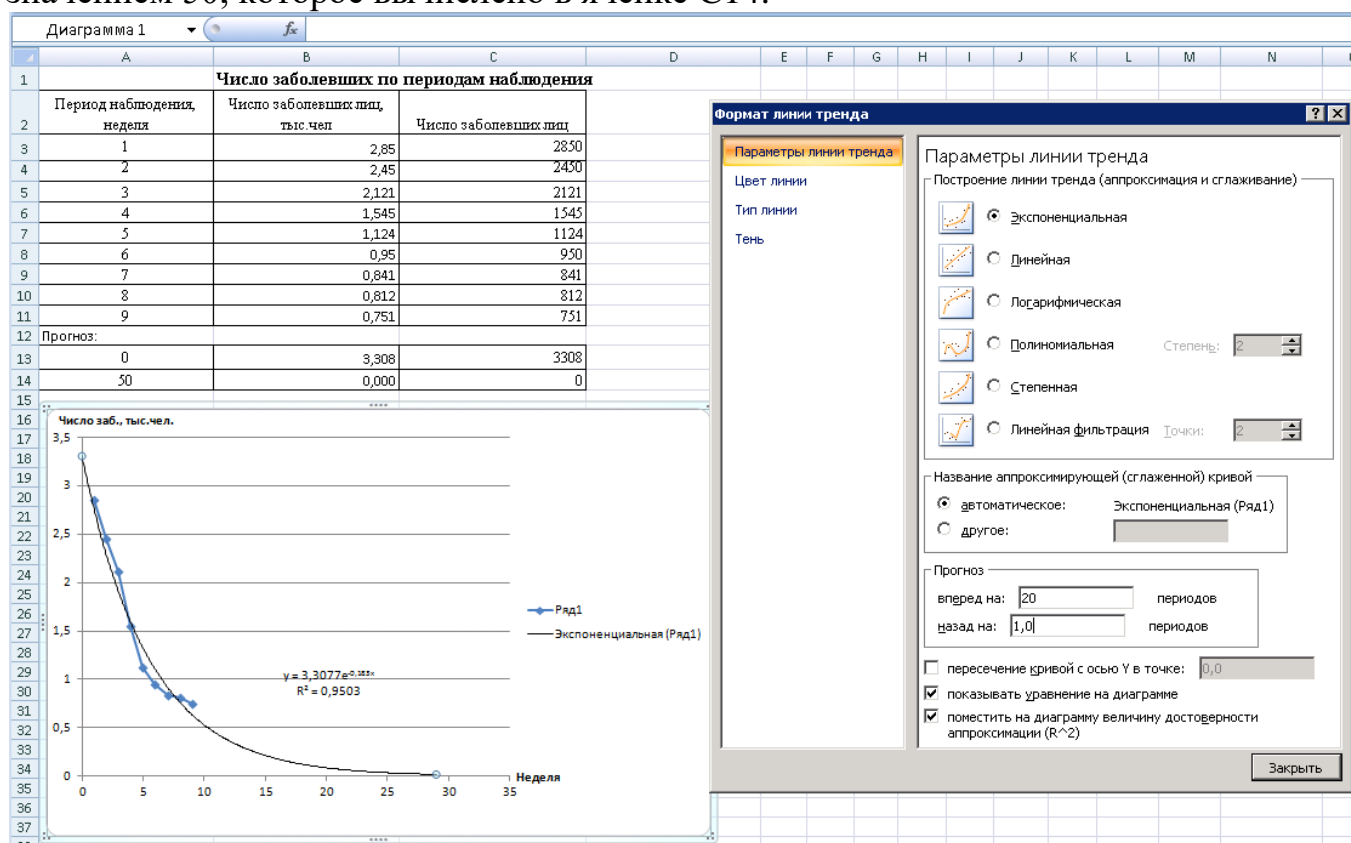


Рис. 11. Прогноз периода завершения распространения заболевания на основе экспоненциальной функции

Вывод: модель эпидемического процесса заболевания N выражается экспоненциальной функцией с вероятностью прогноза 0,95, распространение заболевания полностью завершиться к 50 неделе.