Белорусский государственный технологический университет

Кафедра «Информационных систем и технологий»

Лабораторная работа № 9

**Прогнозирование**

Выполнил студент

3 курса 2 группы

Процукович К.М.

Проверил

Колесников В. Л.

Минск 2019

# **Цель работы**

Цель лабораторной работы – научиться строить прогнозы выбранных параметром и проанализировать результаты.

**2. Описание объекта исследования**

## **2.1 Сущность**

**Прогноз** – вероятностное суждение о будущем состоянии объекта.

**Прогнозирование** — исследование конкретных перспектив дальнейшего развития какого-либо процесса.

Сам же прогноз в первую очередь обусловлен желанием знать будущие, что невозможно принципе, из-за ограниченности текущих знаний, многовариантности следствий, статических ошибок.

Точность прогноза обусловлена следующими характеристиками:

* объемом верифицированных и неверифицированных данных;
* периодом сбора данных;
* методами прогнозирования;
* свойствами объекта прогнозирования.

Прогнозы условно можно разделить:

* по срокам: краткосрочные, долгосрочные;
* по масштабу: частные, региональные, глобальные.

**2.2 Классификация методов и моделей прогнозирования**

**Метод прогнозирования** представляет собой последовательность действий, которые нужно совершить для получения модели прогнозирования.

К основным методом прогнозированию относят:

* формализованные;
* интуитивные.

**Интуитивные методы** прогнозирования имеют дело с суждениями и оценками экспертов. На сегодняшний день они часто применяются в маркетинге, экономике, политике, так как система, поведение которой необходимо спрогнозировать, или очень сложна и не поддается математическому описанию, или очень проста и в таком описании не нуждается.

**Формализованные методы** — методы прогнозирования, в результате которых определяют такую математическую зависимость, которая позволяет вычислить будущее значение процесса, то есть сделать прогноз.

**Модель прогнозирования** – функциональное представление, адекватно описывающее исследуемый процесс и являющееся основой для получения его будущих значений.

В свою очередь формальные методы можно разграничить на **статистические** и **структурные модели**.

В статистических моделях зависимость будущего значения от прошлого задается в виде некоторого уравнения. К ним относятся:

* регрессионные модели (линейная регрессия, нелинейная регрессия);
* авторегрессионные модели (ARIMAX, GARCH, ARDLM);
* модель экспоненциального сглаживания(модель Хольта-Винтера);
* модель по выборке максимального подобия;

В структурных моделях зависимость будущего значения от прошлого задается в виде некоторой структуры и правил перехода по ней. К ним относятся:

* нейросетевые модели;
* модели на базе цепей Маркова;
* модели на базе классификационно-регрессионных деревьев.

Развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования, неравноценных по своему значению. По оценкам зарубежных и отечественных систематиков прогностики уже насчитывается свыше 100 методов прогнозирования.

**2.3 Модели экспоненциального сглаживания**

Можно выделить следующие модели ЭС:

* простое экспоненциальное сглаживание;
* двойное экспоненциальное сглаживание (модель Хольта);
* тройное экспоненциальное сглаживание (модель Хольта-Винтерса).

Рассмотрим **простое экспоненциальное сглаживание.** Отличительная черта это то, что ряд коэффициентов здесь выбирается совершенно определенным образом — их величина падает по экспоненциальному закону.

Пусть мы делаем прогноз на момент времени t+1 (на следующий период). Обозначим его как

http://scm-book.ru/sites/default/files/1image/BusinessFC_f10.gif

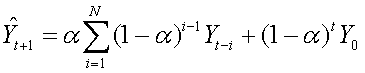
Здесь мы берем в качестве основы прогноза прогноз последнего периода, и добавляем поправку, связанную с ошибкой этого прогноза. Вес этой поправки будет определять, насколько «резко» наша модель будет реагировать на изменения. Очевидно, что **0 < a < 1**.

Считается, что для медленно меняющегося ряда лучше брать значение 0.1, а для быстро меняющегося — подбирать в районе 0.3-0.5.

Если переписать эту формулу в другом виде, получается

http://scm-book.ru/sites/default/files/1image/BusinessFC_f12.gif

Мы получили так называемое рекуррентное соотношение — когда последующий член выражается через предыдущий. Теперь мы прогноз прошлого периода выражаем тем же способом через позапрошлое значение ряда и так далее. В итоге удается получить формулу прогноза



В качестве иллюстрации продемонстрируем сглаживание при разных значениях постоянной сглаживания (рис. 2.1)

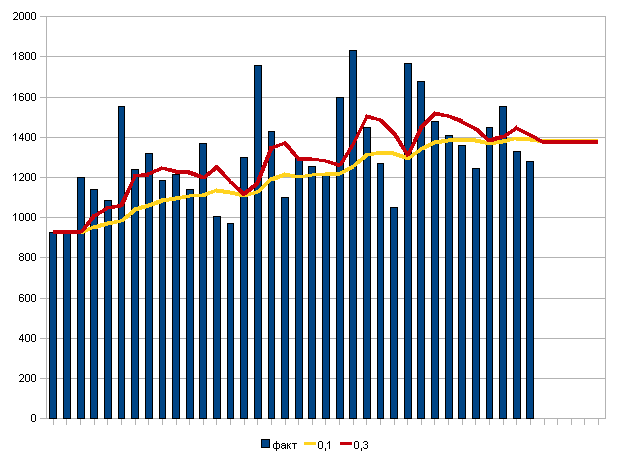


Рис. 2.1 – Сглаживание при разных значениях постоянной сглаживания

**2.4 Модели регрессионного анализа**

Регрессионный анализ помогает понять, как меняется типичное значение зависимой переменной, если одна из независимых переменных изменяется, в то время как другие независимые переменные остаются фиксированными.

Модели регрессионного анализа включают следующие переменные:

* неизвестные параметры, обозначенные как β (бета), которые могут представлять собой скаляр или вектор.;
* независимые переменные, X;
* зависимые переменные, Y.

Принято выделять две регрессионные модели анализа: **линейный** и **нелинейный** **регрессионный анализ**.

**2.4.1 Линейная регрессионный анализ**

В линейной регрессии (рис. 2.2) особенностью является то, что зависимая переменная, которой является yi, представляет собой линейную комбинацию параметров. Например, в простой линейной регрессии для моделирования n-точек используется одна независимая переменная, xi, и два параметра, β0 и β1.

Имеет вид:

**y = a+b\*x+e**

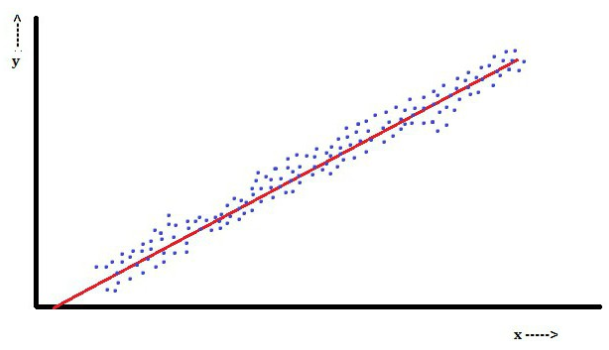


Рис. 2.2 – График линейной регрессии

При случайной выборке из популяции ее параметры позволяют получить образец модели линейной регрессии.

**2.4.2 Нелинейная регрессионный анализ**

Нелинейные регрессии делятся на два класса:

* регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам,
  + полиномы разных степеней: **y = a+b1**\***x+b2\*x2+b3\*x3+e,**
  + равносторонняя гипербола **y = a+ b/x + e,**
* регрессии, нелинейные по оцениваемым параметрам,
  + степенная **y = a\*xb \*e**
  + показательная **y = a\*bx \*e**

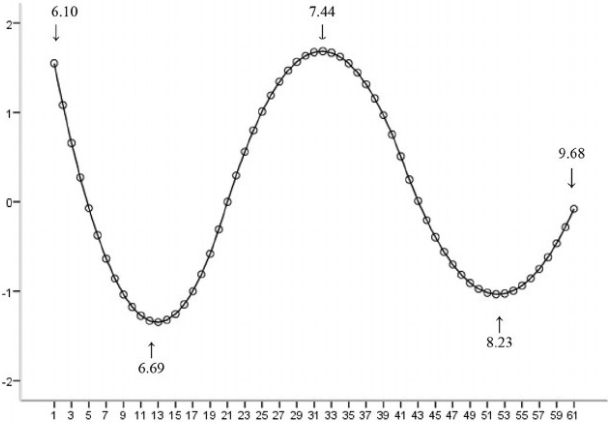


Рис. 2.3 – График нелинейной регрессии

**2.4.3 Метод наименьших квадратов (МНК)**

Методом наименьших квадратов называют такой метод, при котором нахождение оптимальных параметров линейной регрессии, имеет в сумме квадратов регрессионных остатков минимальное количество ошибок.

В этом методе ключевым моментом выступает, минимизация евклидова расстояния:

[МНК](http://edufuture.biz/index.php?title=%D0%A4%D0%B0%D0%B9%D0%BB:11kl_MNK01.jpg)

Между вектором восстановленных значений зависимой переменной и вектором фактических значений зависимой переменной.

**2.4.4 Другие методы**

Несмотря на то что параметры регрессионной модели, как правило, оцениваются с использованием метода наименьших квадратов, существуют и другие методы, которые используются гораздо реже. К примеру, это следующие методы:

* Байесовские методы (например, байесовский метод линейной регрессии).
* Процентная регрессия, использующаяся для ситуаций, когда снижение процентных ошибок считается более целесообразным.
* Наименьшие абсолютные отклонения, что является более устойчивым в присутствии выбросов, приводящих к квантильной регрессии.
* Непараметрическая регрессия, требующая большого количества наблюдений и вычислений.
* Расстояние метрики обучения, которая изучается в поисках значимого расстояния метрики в заданном входном пространстве

**2.5 Временные ряды**

Временной ряд – это множество наблюдений, генерируемых последовательно во времени. Если время непрерывно, временно ряд называется непрерывным. Если время изменяется дискретно, временной ряд дискретен.

Наблюдения дискретного временного ряда,  сделанные в моменты времени http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image001.gif могут быть обозначены черезhttp://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image002.gif. Когда имеется http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image004.gif последовательных значений такого ряда, доступных для анализа, мы пишем http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image005.gif, обозначая так наблюдения, сделанные в равноотстоящие моменты времени http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image006.gif. Во многих случаях значения http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image007.gif и http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image003.gif не важны, но если необходимо точно определить времена наблюдений, нужно указать эти два значения. Если мы принимаем http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image007.gif за начало и http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image003.gif за единицу времени, мы можем рассматривать http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image008.gif как наблюдение в момент времени http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image009.gif.

Существуют три основные систематические компоненты временного ряда: *тренд*, *сезонность*, *цикличность*.

**Тренд** - это систематическая линейная или нелинейная компонента, которая изменяется во времени.

**Сезонность** - это периодические колебания уровней временного ряда внутри года.

**Цикличность** - это периодические колебания, выходящие за рамки одного года. Промежуток времени между двумя соседними вершинами или впадинами в масштабах года считается длиной цикла.

Дискретные временные ряды могут появляться двумя путями.

1) Выборкой из непрерывных временных рядов

2) Накоплением переменной в течение некоторого периода времени; примерами могут служить дождевые осадки, которые обычно накапливаются за такие периоды, как день или месяц, или выход партий продукта, накапливающегося за время цикла. Например, на рис. 2.4 показан временной ряд, состоящий из значений выхода 70 последовательных партий продукта химического процесса.

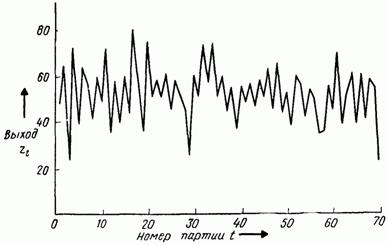


Рис. 2.4 Выход 70 последовательных партий продукта химического процесса**.**

**2.5.1 Детерминированные и случайные временные ряды**

Если будущие значения временного ряда точно определены какой-либо математической функцией, например, такой, как

http://www.sernam.ru/archive/arch.php?path=../htm/book_boks1/files.book&file=boks_13.files/image011.gif,

То временной ряд называют *детерминированным*. Если будущие значения могут быть описаны только с помощью распределения вероятностей, временной ряд называют *недетерминированным*, или просто случайным. Данные о партиях продукта на рис. 2.4 – это пример случайного временного ряда. Хотя в этом ряду имеется отчетливая тенденция к чередованию «вверх-вниз», невозможно точно предсказать выход следующей партии.

**2.6 Метод скользящих средних**

***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).

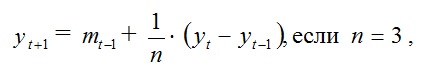
Затем период сдвигается на одно наблюдение, и расчет средней повторяется. При этом периоды определения средней берутся все время одинаковыми. Таким образом, в каждом рассматриваемом случае средняя центрирована, т.е. отнесена к серединной точке интервала сглаживания и представляет собой уровень для этой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



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

1. **Ход работы**

Для составления прогноза нами будут использованы казуальные модели.

Первым этапом будет выбор нами параметра для прогноза. Параметр будет выбран из выходных параметров виртуального комплекса, в нашем случае мы составим прогноз для всех выходных параметров (прочность, пластичность, влагопрочность, загрязнение воды, загрязнение атмосферы, затраты на производство) из них два будет выбрано для анализа, это будут параметры «Затраты на производство» и «Загрязнение атмосферы», наш выбор пал на эти параметры, так как они затрагивают самые важные проблемы, а именно экологии и получение прибыли, что напрямую завит от затрат на продукцию.

Прогноз будет осуществляется с помощью опции «Лист прогноза» в программе Excel. Что показано на рисунке 3.1.

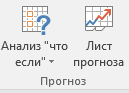


Рис. 3.1 – опция «Лист прогноза»

Эта опция имеет параметры определения сезонности, что будет использовано при прогнозировании по затратам и загрязнению, как можно увидеть на рисунке 3.2, при создании базы данных было симулировано сезонные показатели дебета реки, где год равен 60 записям, что и будет использовано при будущих прогнозах.

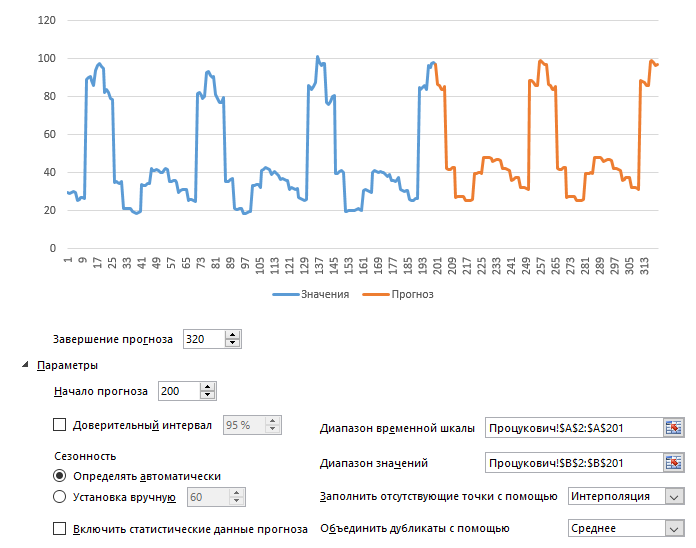


Рис. 3.2 –прогноз дебета реки

Программа учитывает динамику значений и на основе этой динамики и сезонности строит прогноз. Как можно увидеть из таблицы, был произведен прогноз всех выходных параметров с установленной сезонностью.

Таблица 1. Прогноз выходных параметров

|  |  |
| --- | --- |
| Прочность |  |
| Пластичность |  |
| Влагопрочность |  |
| Загрязнение воды |  |
| Загрязнение атмосферы |  |
| Затраты на производство |  |

**Выводы**

С помощью специальных средств Excel был сделан прогноз всех выходных параметров виртуального производственного комплекса. С сезонностью в 60 записей, так как это значение было выбрано при формировании базы данных. И имея базу данных состоящую из двухсот записей.

Для анализа были выбраны параметры затрат и загрязнения атмосферы, как можно заметить на графиках присутствует цикличность (3 полных цикла), что обусловлено сезонностью и количеством записей, и как можно увидеть прогноз в общих чертах повторяет предшествующую динамику.