Домашнее задание №2. Прогнозирование. Функции предсказания и сглаживания.

Задание 2.1.Прогнозирование экономических явлений.

Задание выполняется в MS Excel.

Отчет о выполнении задания оформить в документе Excel. Отчет должен содержать: задание, порядок выполнения задания, результаты выполнения, выводы.

Выполнить следующее задание – Пример 1.

Задача прогнозирования некоторого показателя (или нескольких показателей) на основании имеющихся данных является одной из важнейших в экономико-математическом моделировании. Существует несколько подходов к решению этой задачи.

Наиболее общим и достаточно простым является подход, использующий методы регрессионного анализа.

Рассмотрим прогнозирование по линейной регрессионной модели.

Регрессионная модель решает задачи определения связей между факторами, влияния одного фактора на другой. Рассмотрим два показателя X и Y. Предположим, что они зависимы, то есть изменение одного из них влечет за собой изменение другого. Если при этом, зная точно значение одного показателя можно точно определить значение другого, то связь между показателями называется *функциональной*. Однако на практике в большинстве встречаются зависимости иного вида, когда изменение одного показателя лишь в среднем приводит к изменению другого. Такие зависимости называются *статистическими*. При них, зная значение X, нельзя точно определить Y, так как на Y кроме X влияет еще множество неучтенных факторов. Поэтому, зная X можно лишь в среднем оценить значение Y.

Примеры таких зависимостей в экономике: зависимости между ценой на товар или услугу и объемом их потребления (функции спроса), зависимости между объемом инвестирования и полученным доходом, между издержками и рентабельностью и т.д.

Характер статистической зависимости изучается в регрессионном анализе, а сила статистической связи – в корреляционном анализе.

Предположим, что имеются два экономических показателя X и Y. Для выявления связи между ними проводится реальное одновременное их измерение в разных условиях. Получают выборки пар значений (x_1, y_1), (x_2, y_2),...,(x_n, y_n). Необходимо определить характер статистической зависимости между X и Y, то есть уравнение вида $y= f(x)$, которое позволяет по значению переменной x оценить в среднем значение y, спрогнозировав его. Это уравнение называется уравнением регрессии.

Рассмотрим простейший случай уравнения регрессии – линейную регрессию, когда уравнение регрессии имеет вид прямой линии: $y= ax + b$.

Пример 1. Торговая организация желает выяснить, как влияет количество вложенных в рекламную акцию денег - X (тыс. руб.) на количество проданного товара – Y (тыс. шт.). Для этого проводились наблюдения в разных городах региона и были получены следующие данные.

X	12	15	17	19	20	22	25	27	28	30	33	33
Y	34	42	45	49	53	55	61	68	67	71	75	74

Введем эту таблицу в ячейки A1-M2 электронной книги Excel. Просмотрим предварительно, как лежат точки на графике и какое уравнение регрессии лучше выбрать. Для этого строим график. Вызвав мастер диаграмм и выбрав тип диаграммы «Точечная» нажимаем «Далее» и поместив курсор в поле «Диапазон» обводим курсором данные Y (ячейки B2-M2). Переходим на закладку «Ряд» и в поле «Значения X» делаем ссылку на ячейки B1-M1, обводя их курсором. Нажимаем «Готово» Как видно из графика, точки хорошо укладываются на прямую линию, поэтому будем находить уравнение линейной регрессии вида $y= ax + b$.

Для нахождения коэффициентов a и b уравнения регрессии служат функции НАКЛОН и ОТРЕЗОК категории «Статистические».

Вводим в A5 подпись «a=» а в соседнюю ячейку B5 вводим функцию НАКЛОН, ставим курсор в поле «Изв_знач_y» задаем ссылку на ячейки B2-M2, обводя их мышью. Аналогично в поле «Изв_знач_x» даем ссылку на B1-M1. Результат: **1,923921**.

Найдем теперь коэффициент b. Вводим в A6 подпись «b=», а в B6 функцию ОТРЕЗОК с теми же параметрами, что и у функции НАКЛОН. Результат **12,78151**.

Следовательно, уравнение линейной регрессии есть **y=1,92 x + 12,78**.

Построим график уравнения регрессии. Для этого в третью строчку таблицы введем значения функции регрессии в заданных точках X (первая строка) - $y(x_i)$.

Для получения этих значений используется функция ТЕНДЕНЦИЯ категории «Статистические». Вводим в A3 подпись «Y(X)» и, поместив курсор в B3, вызываем функцию ТЕНДЕНЦИЯ. В полях «Изв_знач_y» и «Изв_знач_x» даем ссылку на B2-M2 и B1-M1. В поле «Нов_знач_x» вводим также ссылку на B1-M1. В поле «Константа» вводят 1, если уравнение регрессии имеет вид $y= ax + b$, и 0, если $y= ax$. В данном случае вводим единицу. Функция ТЕНДЕНЦИЯ является массивом, поэтому для вывода всех ее значений выделяем область B3-M3 и нажимаем F2 и Ctrl+Shift+Enter. Результат – значения уравнения регрессии в заданных точках.

Строим график. Ставим курсор в любую свободную клетку, вызываем мастер диаграмм, выбираем категорию «Точечная», вид графика – линия без точек (в нижнем правом углу), нажимаем «Далее», в поле «Диапазон» вводим ссылку на B3-M3. Переходим на закладку «Ряд» и в поле «Значения X» вводим ссылку на B1-M1, нажимаем «Готово». Результат – прямая линия регрессии.

Посмотрим, как различаются графики опытных данных и уравнения регрессии. Для этого ставим курсор в любую свободную ячейку, вызываем мастер диаграмм, категория «График», вид графика – ломаная линия с точками (вторая сверху левая), нажимаем «Далее», в поле «Диапазон» вводим ссылку на вторую и третью строки B2-M3. Переходим на закладку «Ряд» и в поле «Подписи оси X» вводим ссылку на B1-M1, нажимаем «Готово». Результат – две линии (Синяя – исходные данные, красная – уравнение регрессии). Видно, что линии мало различаются между собой.

Для вычисления коэффициента корреляции r_{xy} служит функция ПИРСОН. Размещаем графики так, чтобы они располагались выше 25 строки, и в A25 делаем подпись «Корреляция», в B25 вызываем функцию ПИРСОН, в полях которой «Массив 1» и «Массив 2» вводим ссылки на исходные данные B1-M1 и B2-M2. Результат **0,993821**.

Коэффициент детерминации R_{xy} – это квадрат коэффициента корреляции r_{xy}. В A26 делаем подпись «Детерминация», а в B26 – формулу «=B25*B25». Результат **0,987681.**

В Excel существует одна функция, которая рассчитывает все основные характеристики линейной регрессии. Это функция ЛИНЕЙН. Ставим курсор в B28 и вызываем функцию ЛИНЕЙН, категории «Статистические». В полях «Изв_знач_y» и «Изв_знач_x» даем ссылку на B2-M2 и B1-M1. Поле «Константа» имеет тот же смысл, что и в функции ТЕНДЕНЦИЯ, у нас она равна 1. Поле «Стат» должно содержать 1, если нужно вывести полную статистику о регрессии. В данном случае ставим туда единицу. Функция возвращает массив размером 2 столбца и 5 строк. После ввода выделяем мышью ячейки B28-C32 и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица значений, числа в которой имеют следующий смысл:

Коэффициент a	Коэффициент b
Стандартная ошибка m_a	Стандартная ошибка m_b
Коэффициент детерминации R_{xy}	Среднеквадратичное отклонение y
F - статистика	Степени свободы n-2
Регрессионная сумма квадратов S^2_b	Остаточная сумма квадратов S^2_a

Анализ результата: в первой строчке – коэффициенты уравнения регрессии, сравните их с рассчитанными функциями НАКЛОН и ОТРЕЗОК. Вторая строчка – стандартные ошибки коэффициентов. Если одна из них по модулю больше чем сам коэффициент, то коэффициент считается нулевым. Коэффициент детерминации характеризует качество связи между факторами. Полученное значение 0,987681 говорит об очень хорошей связи факторов. F – статистика проверяет гипотезу об адекватности регрессионной модели.

Данное число нужно сравнить с критическим значением. для его получения вводим в Е33 подпись «F-критическое», а в F33 функцию FРАСПОБР, аргументами которой вводим соответственно «0,05» (уровень значимости), «1» (число факторов Х) и «10» (степени свободы). Видно, что *F* – статистика больше, чем *F*– критическое, значит - регрессионная модель адекватна.

В последней строке приведены регрессионная сумма квадратов и остаточные суммы квадратов. Важно, чтобы регрессионная сумма (объясненная регрессией) была намного больше остаточной (не объясненная регрессией, вызванная случайными факторами). В данном случае это условие выполняется, что говорит о хорошей регрессии.

Задание 2.2. Работа с функциями предсказания и сглаживания.

Задание выполняется в пакете MathCad.

Отчет о выполнении задания оформить в документе Word, вставив в отчет результаты выполнения в MathCad («скриншоты»). Отчет должен содержать: задание, порядок выполнения задания, результаты выполнения, выводы.

Выполнить следующие задания.

Пример 2. С помощью функции **pridict** предскажите поведение функции f(x)=sin(2*x) на отрезке [2π, 4π], если предположить, что она задана на отрезке [0, 2π].

Пример 3. С помощью функции rnd (random – r n d) введите 50 случайных чисел из отрезка [0, 2]. Постройте функции сглаживания данных (с помощью различных встроенных функций).

Сделать выводы.

1) Работа с функцией предсказания.

В MathCad имеется функция предсказания **pridict (data,k,N)**, которая по ряду заданных равномерно расположенных точек позволяет рассчитать некоторое число N последующих точек. Функция предсказания обеспечивает высокую точность при монотонных исходных функциях или функциях, представляемых полином невысокой степени.

pridict (data,k,N), где:

data – вектор данных;

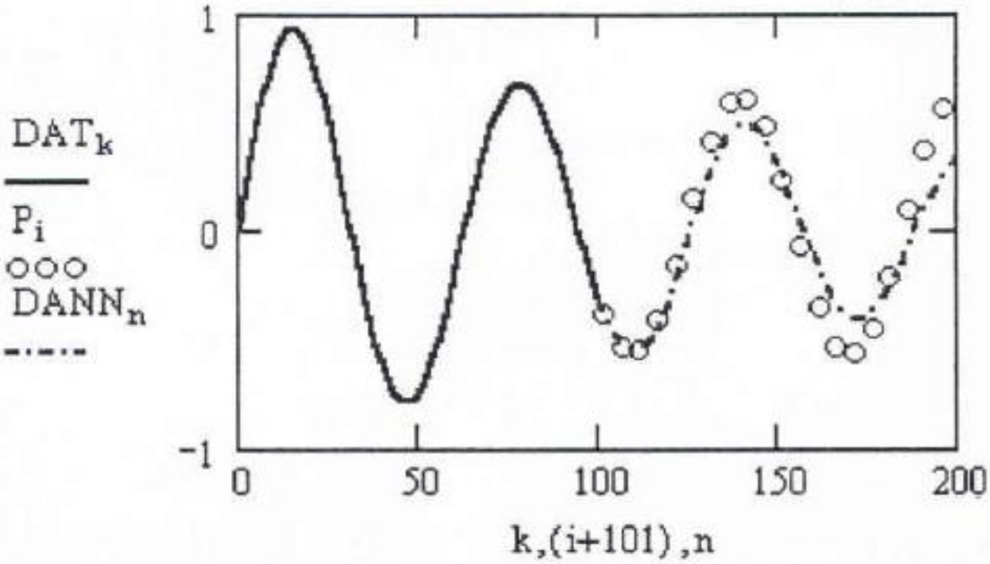
k – число последних точек существующих данных, на основе которых происходит расчет предсказываемых точек;

N - число точек, в которых необходимо предсказать данные.

$$k := 0..100 \qquad DAT_k := \exp\left(\frac{-k}{200}\right)\cdot\sin\left(\frac{k}{10}\right)$$

$$P := \mathrm{predict}(DAT, 3, 100) \qquad i := 0, 5..99$$

$$n := 100, 105..200 \qquad DANN_n := \exp\left(\frac{-n}{200}\right)\cdot\sin\left(\frac{n}{10}\right)$$



2) Работа с функцией сглаживания.

Данные большинства экспериментов имеют случайные составляющие погрешности. Поэтому часто возникает необходимость статистического сглаживания данных. Ряд функций MathCAD предназначен для выполнения операций сглаживания данных различными методами.

Перечень этих функций:

medsmooth(VY, n) – для вектора с *m* действительными числами возвращает m-мерный вектор сглаженных данных по методу скользящей медианы, параметр *n* задает ширину окна сглаживания (*n* должно быть нечетным числом, меньшим *m*);

ksmooth(VX, VY, b) – возвращает *n*-мерный вектор сглаженных *VY*, вычисленных на основе распределения Гаусса. *VX*, *VY* - *n*-мерные векторы действительных чисел. Параметр *b* (полоса пропускания) задает ширину окна сглаживания (*b* должно в несколько раз превышать интервал между точками по оси *x*);

supsmooth(VX, VY) – возвращает *n*-мерный вектор сглаженных *VY*, вычисленных на основе использования процедуры линейного сглаживания методом наименьших квадратов по правилу *k*-ближайших соседей с адаптивным выбором *k*. *VX*, *VY* - *n*-мерные векторы действительных чисел. Элементы вектора *VX* должны идти в порядке возрастания.

$i := 1..50 \quad VX_i := i \quad VY_i := rnd(1)$

$n := rows(VY) \quad n = 51$

$SM1 := medsmooth(VY, 9) \quad SM2 := ksmooth(VX, VY, 5) \quad SM3 := supsmooth(VX, VY)$

$i := 0..n-1$