Лабораторная работа №3

**Прогнозирование в Excel**

**с помощью регрессионного анализа**

**Цель работы:** научиться выполнять прогнозирование экономических

 параметров с помощью одномерного и многомерного

 регрессионного анализа

**Содержание работы:**

 1 Линейный регрессионный анализ.

 2 Экспоненциальный регрессионный анализ.

 3 Линейный многомерный регрессионный анализ

 **1 Линейный регрессионный анализ**

Одним из методов, используемых для прогнозирования, является регрессионный анализ.

***Регрессия -*** это статистический метод, который позволяет найти уравнение, наилучшим образом описывающее совокупность данных, заданных таблицей.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **х** | **x1** | **x2** | **...** | **хi** | **...** | **хn** |
| **y** | **x1** | **y2** | **...** | **yi** | **...** | **yn** |

На графике эти данные отображаются точками. Регрессия позволяет подобрать к этим точкам кривую у=f(x), которая вычисляется по методу наименьших квадратов и даёт максимальное приближение к табличным данным.

 •

#  у у •

 • •

 •

 • •

 • • • •

 **b** • •

 **х** **х**

Рисунок 1 Линейная регрессия Рисунок 2 Нелинейная регрессия

По полученному уравнению можно вычислить (сделать прогноз) значение функции **у** для любого значения **х** , как внутри интервала изменения **х** из таблицы(интерполяция), так и вне его (экстраполяция).

## Линейная регрессия

Линейная регрессия дает возможность наилучшим образом провести прямую линию через точки одномерного массива данных. Уравнение с одной независимой переменной, описывающее прямую линию, имеет вид:

 **y=mx+b,** (1)

где:

 **х** - независимая переменная ;

 **у** -зависимая переменная;

 **m** – характеристика наклона прямой;

 **b** - точка пересечения прямой с осью **у**.

 Например, имея данные о реализации товаров за год с помощью линейной регрессии можно получить коэффициенты прямой (1) и, предполагая дальнейший линейный рост, получить прогноз реализации на следующий год.

 **2 Экспоненциальный регрессионный анализ.**

Нелинейная регрессия позволяет подбирать к табличным данным нелинейное уравнение – параболу, гиперболу и др. Excel реализует нелинейность в виде экспоненты, т.е. подбирает кривую вида

 **y=b·mx** (2)

которая позволяет наилучшим образом провести экспоненциальную

кривую по точкам данных, которые изменяются нелинейно.

Так, например, данные о росте населения почти всегда лучше описываются не прямой линией, а экспоненциальной кривой. При этом нужно помнить, что достоверное прогнозирование возможно только на участках подъёма или спуска кривой (при отрицательных значениях х), т.к. сама кривая (2) изменяется монотонно, без точек перегиба. Например, делать экспоненциальный прогноз для функции, изменяющейся синусоидально, можно только на участках подъёма или спуска функции, для чего её разбивают на соответствующие интервалы.

***Множественная регрессия***

Множественная регрессия представляет собой анализ более одного набора данных аргумента х и дает более реалистичные результаты. Множественный регрессионный анализ также может быть как линейным, так и экспоненциальным. Уравнения регрессии (1) и(2) примут соответственно вид (3) и (4):

 **у=mlxl + m2x2 + ... + mnxn + b** (3)

 **у = b·mlxl· m2x2 · ... · mnxn** (4)

где: **xl, x2, ..., xn** - независимые переменные.

С помощью множественной регрессии, например, можно оценить стоимость дома в некотором районе, основываясь на данных его площади, размерах участка земли, этажности, вида из окон и т.д.

***Использование функций регрессии***

В Excel имеется 5 функций для линейной регрессии (ЛИНЕЙН(...), ТЕНДЕНЦИЯ(...), ПРЕДСКАЗ(...), НАКЛОН(...), СТОШУХ(...)) и 2 функции дня экспоненциальной регрессии - ЛГРФПРИБЛ(...) и POCT(...). Рассмотрим некоторые из них.

1 Функция **=ЛИНЕЙН(**изв.\_знач.**\_y;**изв. \_знач. **x;конст;стат)** (5)

вычисляет коэффициент **m** и постоянную **b** для уравнения прямой (1).

Известные\_значения\_**y** и известные\_значения\_**x** - это множество значений y и необязательное множество значений x (их вводить необязательно), которые уже известны для соотношения (1).

**Константа**-это логическое значение , которое указывает требуется ли, чтобы константа **b** была равна 0. Если константа имеет значение ИСТИНА или опущено,то **b** вычисляется обычным образом.

**Статистика-**это логическое значение, которое указывает, требуется ли вывести дополнительную статистику по регрессии.

Если статистика имеет значение ЛОЖЬ(или 0),то функция ЛИНЕЙН возвращает только значения коэффициентов **m** и **b**, в противном случае выводится дополнительная регрессионная статистика в виде табл.1:

 Таблица 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| mn | mn-1 | ... | m2 | m1 | b |
| sen | sen-1 | ... | se2 | se1 | seb |
| r2 | sev | ... | #Н/Д | #Н/Д | #Н/Д |
| F | df | ... | #Н/Д | #Н/Д | #Н/Д |
| ssreg | ssresid | ... | #Н/Д | #Н/Д | #Н/Д |

где

se1,se2,...,sen - стандартные значения ошибок для коэффициентов m1,m2,...,mn.

seb - стандартное значение ошибки для постоянной b (seb равно #Н/Д, т.е. «нет допустимого значения», если **конст**. имеет значение ЛОЖЬ).

r2 - коэффициент детерминированности. Сравниваются фактические значения **y** и значения, получаемые из уравнения прямой; по результатам сравнения вычисляется коэффициент детерминированности, нормированный от 0 до 1. Если он равен 1, то имеет место полная корреляция с моделью, т.е. нет различия между фактическим и оценочным значениями **y.** В противоположном случае, если коэффициент детерминированности равен 0, то уравнение регрессии неудачно для предсказания значений **y**.

sey - стандартная ошибка для оценки **y** (предельное отклонение для **у**).

F - F-статистика, или F-наблюдаемое значение. F-статистика используется для определения того, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет.

df - cтепени свободы. Степени свободы полезны для нахождения F-критических значений в статистической таблице. Для определения уровня надежности модели нужно сравнить значения в таблице с F-статистикой, возвращаемой функцией ЛИНЕЙН.

ssreg - регрессионая сумма квадратов.

ssresid - oстаточная сумма квадратов.

#Н/Д – ошибка, означающая “Нет доступного значения”.

· Любую прямую можно задать ее наклоном **m**  и **y**-пересечением:

*Наклон (****m****):*

Для того, чтобы определить наклон прямой, обычно обозначаемый через **m,** нужно взять две точки прямой (x1,y1) и (x2,y2); тогда наклон равен m = (y2 - y1)/(x2 - x1).

***y****-пересечение (****b****):*

**y**-пересечением прямой, обычно обозначаемым через **b**, является значение **y** для точки, в которой прямая пересекает ось **y**.

Уравнение прямой имеет вид **y = mx + b**. Если известны значения **m** и **b**, то можно вычислить любyю точку на прямой, подставляя значения **y** или **x** в уравнение. Можно также использовать функцию ТЕНДЕНЦИЯ (см. ниже).

· Если для функции **у** имеется только одна независимая переменная **x**, можно получить наклон и **y**-пересечение непосредственно, используя следующие формулы:

2 Наклон **m**:

**ИНДЕКС(ЛИНЕЙН(изв\_знач\_y;изв\_знач\_x);1)**

3 y-пересечение **b**:

**ИНДЕКС(ЛИНЕЙН(изв\_знач\_y;изв\_знач\_x);2)**

· Точность аппроксимации с помощью прямой, вычисленной функцией ЛИНЕЙН, зависит от степени разброса данных. Чем ближе данные к прямой, тем более точными являются модель, используемая функцией ЛИНЕЙН, и значения, получаемые из уравнения прямой.

4 В случае экспоненциальной регрессии аналогом функции (5) является

функция **=ЛГРФПРИБЛ(изв\_знaч\_у; изв\_ знач\_х; конст;стат),** (6) которая отличается лишь тем, что вычисляет коэффициенты **m** и  **b** для экспоненциальной кривой (2).

5 Функция

 =**ТЕНДЕНЦИЯ(**изв\_знач**\_y;** изв\_знач**\_x;**нов\_знач**\_x;конст)** (7)

вращает числовые значения, лежащие на прямой линии, наилучшим образом аппроксимирующей известные табличные данные.

Новые\_значения\_**x -** это те, для которых необходимо вычислить соответствующие значения **y.**

Если параметр новые\_значения\_**x** пропущен,то считается,что он совпадает с известными **x**. Назначение остальных параметров функции ТЕНДЕНЦИЯ совпадает с описанными выше.

6. В случае экспоненциальной регрессии аналогом функции (7) является функция **=РОСТ(изв\_знач\_y; изв\_знач\_x; конст)** (8)

7. функция **=CTOШУX(**изв**\_**знач\_**y;**изв**\_**знач\_**х)**  (9)

возвращает стандартную погрешность регрессии - меру погрешности

предсказываемого значения **у** для заданного значения **х.**

***Правила ввода функций:***

Формулы (5)-(8) являются табличными, т.е. они заменяют собой несколько обычных формул и возвращают не один результат, а массив результатов. Поэтому необходимо соблюдать следующие правила:

1 ) перед вводом одной из формул (5)-(8) выведите блок ячеек, точно совпадающей по размеру с величиной возвращаемого формулой массива результатов. Например, при использовании функции **ЛИНЕЙН** с выводом статистики нужно выделить массив ячеек, равный табл. 1, если параметр **статистика** равен **ЛОЖЬ,** достаточно выделить одну строку табл.1;

2) наберите функцию в строке формул. При этом слова на русском языке можно набирать строчными буквами, т.к. они являются ключевыми и при вводе Excel автоматически переведёт их в заглавные. Имена ячеек обязательно вводятся латинским шрифтом. Вместо слова ИСТИНА можно вводить числа от 1 до 9 (не 0), а вместо слова ЛОЖЬ – число 0. Если в результате выполнения функции выводится одно число, можно вводить формулы не вручную, а использовать аппарат Мастера функций.

3) одновременно нажмите клавиши **Shift+Ctri+Enter**. Результаты вычислений заполнят выделенные ячейки.

***Линия тренда***

Excel позволяет наглядно отображать тенденцию изменения данных с помощью линии тренда, которая представляет собой интерполяционную кривую, описывающую отложенные на диаграмме данные.

Для того, чтобы дополнить диаграмму исходных данных линией тренда, необходимо выполнить следующие действия :

1) выделить на диаграмме ряд данных, для которого требуется построить линию тренда;

2) в меню **Диаграмма** выбрать команду **Добавить линию тренда**,

3) в открывшемся окне задать метод интерполяции (линейный, полиномиальный, логарифмический и т.д.), а также через вкладку ***Параметры*** - другие параметры (например, вывод уравнения кривой тренда, коэффициента детерминированности **r2**, направление и количество периодов для экстраполяции (прогноза) и др.);

 4) нажать кнопку ОК.

Чтобы отобразить на графике (гистограмме и др.) новые, прогнозируемые в результате регрессионного анализа данные, нужно:

1) определить их с помощью функций ТЕНДЕНЦИЯ, РОСТ или другим способом,

2) выделить на диаграмме нужную кривую, щелкнув по ней мышью,

3) в меню Диаграмма выбрать команду Добавить данные..., в появившемся окне ввести диапазон ячеек с новыми данными вручную или протащив по ним курсор при нажатой левой клавише мыши, нажать ОК.

На диаграмме появится продолжение кривой, построенной по новым данным.

***Простая линейная регрессия***

***Пример 1***. Функция ТЕНДЕНЦИЯ.

а) Предположим, что фирма желает приобрести земельный участок в июле. Фирма собирает информацию о ценах за последние 12 месяцев, начиная с марта, на типичный земельный участок. Номера месяцев с 1 по 12 (известные значения **х**) записаны в ячейки А2...А13. Известные значения **y** содержат множество известных значений (133 890 руб., 135 000 руб., 135 790 руб., 137 300 руб., 138 130 руб., 139 100 руб., 139900 руб., 141 120 руб., 141 890руб" 143 230 руб., 144 000 руб., 145 290 руб.), которые находятся в ячейках В2:В13 соответственно (данные условные). Новые значения **х**, т.е. числа 13, 14, 15, 16, 17 введём в ячейки А14...А18. Для того чтобы определить ожидаемые значения цен на март, апрель, май, июнь, июль, выделим любой интервал ячеек, например, С2:С6 (по одной ячейке для каждого месяца) и в строке формул введем функцию:

=ТЕНДЕНЦИЯ (В2:В13;А2:А13;А14:А18;С2:С6). (10)

После нажатия клавиш **Ctrl+Shift+Enter** данная функция будет введена как формула вертикального массива, а в ячейках С2:С6 появится результат: {146172:147190:148208:149226:150244}.

Таким образом, в июле фирма может ожидать цену около 150 244 руб.

б) Тот же результат будет получен, если вводить в формулу не все массивы переменных х и у, а использовать часть массивов, которые предусматриваются автоматически по умолчанию. Тогда формула (10) примет вид:

=ТЕНДЕНЦИЯ (В2:В13;;{13:14:15:16:17}). (11)

В формуле (11) используется массив по умолчанию (1:2:3:4:5:6:7:8:9:10: 11:12) для аргумента «известные \_значения\_х», соответствующий 12 месяцам, для которых имеются данные по продажам. Он должен был бы быть помещен в формуле(11) между двумя знаками  **;;**.

Массив (13:14:15:16:17) соответствует следующим 5 месяцам, для которых и получен массив результатов (146172:147190:148208:149226:150244).

Элементы массивов разделяет знак " : " , который указывает на то, что они расположены по столбцам.

в) Аргумент «новые значения х» можно задать другим массивом ячеек, например, В14:В18, в которые предварительно записаны те же номера месяцев 13, 14, 15, 16, 17. Тогда вводимая в строку формул функция примет вид =ТЕНДЕНЦИЯ (В2:В13;;В14:В18).

***Пример 2*** а) Функция ЛИНЕЙН. Дана таблица изменения температуры в течение шести часов, введенная в ячейки D2:E7 (табл. 2). Требуется определить температуру во время восьмого часа.

 Таблица 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | ... | D | E |
| 1 |  | **х -**№ часа | **у -**t○, град. |
| 2 |  | 1 | 2 |
| 3 |  | 2 | 3 |
| 4 |  | 3 | 4 |
| 5 |  | 4 | 7 |
| 6 |  | 5 | 12 |
| 7 |  | 6 | 18 |

Выделим ячейки D8:E12 для вывода результата (в соответствии с табл. 1), введём в строку ввода формулу =ЛИНЕЙН(E2:E7;D2:D7;1;1), нажмём клавиши **Ctrl+Shift+Enter**, в выделенных ячейках появится результат:

|  |  |
| --- | --- |
| 3,1428571 | -3,3333333 |
| 0,5408484 | 2,106302 |
| 0,8940887 | 2,2625312 |
| 33,767442 |  4 |
| 172,85714 |  20,47619 |

Таким образом, коэффициент **m =** 3,143 со стандартной ошибкой 0,541, а свободный член **b = -**3,333 со стандартной ошибкой 2,106, т.е. функция, описывающая данные табл. 2, имеет вид

 **у =** 3,143∙**х -**3,333 (12)

Стандартные ошибки показывают максимально возможное отклонение параметра от рассчитанной величины . Для **у** оно составляет 2,263, т.е. реальное значение **у** может лежать в пределах **у±2,263.**

Точность приближения к табличным данным (коэффициент детерминированности r2) cоставляет 0,894 или 89,4%, что является высоким

показателем. При х = 8 получим: **у** = 3,143∙**8** – 3,333 = **21,81** град.

 б)Тот же результат можно получить, использовав функцию =ТЕНДЕНЦИЯ (E2:E7;;G2:G5) для, например, следующих четырёх часов, предварительно введя в ячейки G2:G5 числа с 7 до 10. Выделив ячейки Н2:Н5, введя в строку формул эту функцию и нажав Ctrl+Shift+Enter , получим в выделенных ячейках массив {18,667; 21,80952; 24,95238; 28,09524}, т.е. для восьмого часа значение у = 21,809 ≈21,81град.

в) Функция ПРЕДСКАЗ – позволяет предсказывать значение **у** для нового значения **х** по известным значениям **х** и **у** , используя линейное приближение зависимости **у = f(x)** . Синтаксис функции**:**

**=ПРЕДСКАЗ(**нов­,\_знач.\_**х;** изв.\_знач.\_ **у**;изв.\_ знач.\_**х)**

Для данных примера 2 ввод формулы =ПРЕДСКАЗ(8;E2:E7;D2:D7) выводит в заранее выделенной ячейке результат 21,809. Новое значение **х** может быть задано не числом, а ячейкой, в которую записано это число.

Отличие функции ПРЕДСКАЗ от функции ТЕНДЕНЦИЯ заключается в том, что ПРЕДСКАЗ прогнозирует значение функции линейного приближения только для *одного* нового значения **х**.

***Экспоненциальная регрессия***

***Пример 3.*** а) Функция ЛГРФПРИБЛ.

Условие примера 2.

Поскольку функция в табл. 2 носит явно нелинейный характер, целесообразно искать её приближение в виде не прямой линии, как в примере 2, а в виде нелинейной кривой. Из всех видов нелинейности (гипербола, парабола и др.) Excel реализует только экспоненциальное приближение вида у = b∙mx c помощью функции ЛГРФПРИБЛ, которая рассчитывает для этого уравнения значения b и m.

Выделим для результата блок ячеек F8:G12, введём в строку формул

Функцию =ЛГРФПРИБЛ(E2:E7;D2:D7;1;1), нажмём клавиши Ctrl+Shift+ Enter, в выделенных ячейках появится результат:

|  |  |
| --- | --- |
| 1,56628015 | 1,196513 |
| 0,02038299 | 0,07938 |
| 0,99181334 | 0,085268 |
| 484,599687 | 4 |
| 3,52335921 | 0,029083 |

Таким образом, коэффициент m = 1,556, а b = 1,197, т.е. уравнение приближающей кривой имеет вид :

 **у = 1,197∙(1,556 х)** (13)

со стандартными ошибками для m, b и y равными 0,02, 0,07 и 0,08 соответ-ственно. Коэффициент детерминированности **r2** = 0,992, т.е.

полученное уравнение даёт совпадение с табличными данными с вероятностью 99,2%.

Поскольку интерполяция табл. 2 экспоненциальной кривой даёт более точное приближение (99,2%) и с меньшими стандартными ошибками для m,b и y, в качестве приближающего уравнения принимаем уравнение (13).

При х = 8 получим **у** = 1,197**∙**34,363 = **41,131** град.

б) Функция РОСТ вычисляет прогнозируемое по экспоненциальному приближению значения **у** для новых значений **х** , имеет формат:

**=РОСТ**(изв\_знач\_**у**;изв\_знач\_**х**;нов\_знач\_**х**;**константа**).

Выделим блок ячеек F14:F17, введём формулу

=РОСТ(E2:E7;D2:D7;G2:G5;ИСТИНА), в выделенных ячейках появится массив чисел {27,6696434; 43,3384133; 67,8800967; 106,319248}, т.е. при х=8 значение функции **у = 43,34** град. Это значение немного отличается от вычисленного в п. а), поскольку функция РОСТ использует для расчетов линию экспоненциального тренда.

*Примечание.* При выборе экспоненциальной приближающей кривой следует учитывать, что интерполировать ею можно только участки, где функция монотонно возрастает или убывает (при отрицательном аргументе **х**), т.е. функцию, имеющую точки перегиба (например, параболу, синусоиду , кривую рис. 2 – т..А и др.) следует разбить на участки монотонного изменения от одной точки перегиба до другой и каждый участок интерполировать отдельно. Для рис. 2 функцию нужно разбить на 2 участка – от начала до т..А и от т.А до конца кривой.

 **3 Линейный многомерный регрессионный анализ**

***Пример 4.*** Предположим, что коммерческий агент рассматривает возможность закупки небольших зданий под офисы в традиционном деловом районе. Агент может ис­пользовать множественный регрессионный анализ для оценки цены здания под офис на основе следующих переменных:

**y**-оценочная цена здания под офис;

**x1**-общая площадь в квадратных метрах;

**x2**-количество офисов ;

**x3**-количество входов;

**x4**-время эксплуатации здания в годах.

Агент наугад выбирает 11 зданий из имеющихся 1500 и получает сле­дующие данные:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 | х1 -площадь, м2 | х2 -офисы | х3 - входы | х4- срок, лет | Цена, у.е. |
| 2 | 2310 | 2 | 2 | 20 | 42000 |
| 3 | 2333 | 2 | 2 | 12 | 144000 |
| 4 | 2356 | 3 | 1.5 | 33 | 151000 |
| 5 | 2379 | 3 | 2 | 43 | 151000 |
| 6 | 2402 | 2 | 3 | 53 | 139000 |
| 7 | 2425 | 4 | 3 | 23 | 169000 |
| 8 | 2448 | 2 | 1.5 | 99 | 126000 |
| 9 | 2471 | 2 | 2 | 34 | 142000 |
| 10 | 2494 | 3 | 3 | 23 | 163000 |
| 11 | 2517 | 4 | 4 | 55 | 169000 |
| 12 | 2540 | 2 | 3 | 22 | 149000 |

"Пол-входа" означает вход только для доставки корреспонденции.

В этом примере предполагается, что существует линейная зависимость между каждой независимой переменной (x1,x2,x3,x4) и зависимой переменной (y) ,т.е. ценой здания под офис в данном рай­оне.

* выделим блок ячеек А14:Е18 (в соответствии с табл. 1),
* введем формулу =ЛИНЕЙН (E2:E12;A2:D12;ИСТИНА;ИСТИНА),
* нажмём клавиши Ctrl+Shift+ Enter,
* в выделенных ячейках появится результат:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 14 | -234.237 | 2553.210 | 12529.7682 | 27.6413 | 52317.830 |
| 15 | 13.2680 | 530.66915 | 400.066838 | 5.42937 | 12237.361 |
| 16 | 0.99674 | 970.57846 | #H/Д | #H/Д | #H/Д |
| 17 | 459.753 |  6 | #H/Д | #H/Д | #H/Д |
| 18 | 1732393319 | 5652135.3 | #H/Д | #H/Д | #H/Д |

Уравнение множественной регрессии y=ml**∙**xl+m2**∙**x2+m3**∙**x3+m4**∙**x4+b теперь может быть получено из строки 14:

 y=27,64**∙**x1+12530**∙**x2+2553**∙**x3-234,24**∙**x4+52318 (14)

Теперь агент может определить оценочную стоимость здания под офис

в том же районе, которое имеет площадь 2500 кв. м , три офиса, два входа, зда­нию 25 лет, используя следующее уравнение:

у=27,64**∙**2500+12530**∙**3+2553**∙**2-234,24**∙**25+52318=158261 у.е.

Это значение может быть вычислено с помощью функции ТЕНДЕНЦИЯ:

 =ТЕНДЕНЦИЯ (Е2:Е12; A2:D12; {2500;3;2;25}).

При интерполировании с помощью функции

 =ЛГРФПРИБЛ(E2:E12;A2:D12;ИСТИНА;ИСТИНА)

для получения уравнения множественной экспоненциальной регрессии выводится результат:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0,99835752 | 1,0173792 | 1,0830186 | 1,0001704 | 81510,335 |
| 0,00014837 | 0,0065041 | 0,0048724 | 6,033E-05 | 0,1365601 |
| 0,99158875 | 0,0105158 | #Н/Д | #Н/Д | #Н/Д |
| 176,832548 | 6 | #Н/Д | #Н/Д | #Н/Д |
| 0,07821851 | 0,0006635 | #Н/Д | #Н/Д | #Н/Д |
| #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |

Коэффициент детерминированности здесь составляет 0,992 (99,2%), т.е. меньше, чем при линейной интерполяции, поэтому в качестве основного следует оставить уравнение множественной регрессии (14).

Таким образом, функции ЛИНЕЙН, ЛГРФПРИБЛ, НАКЛОН определяют коэффициенты , свободные члены и статистические параметры для уравнений одномерной и множественной регрессии, а функции ТЕНДЕНЦИЯ, ПРЕДСКАЗ, РОСТ позволяют получить прогноз новых значений без составления уравнения регрессии по значениям тренда.

**Контрольные вопросы**

1 Сущность регрессионного анализа, его использование для прогнозирования функций.

2 Как получить уравнение одномерной линейной регрессии, каков синтаксис функций линейного приближения?

3 Как получить уравнение многомерной линейной регрессии, каков синтаксис функции?

4 Как получить уравнение одномерной экспоненциальной регрессии, каков синтаксис функции экспоненциального приближения?

5 Как получить уравнение многомерной экспоненциальной регрессии, каков синтаксис функции экспоненциального приближения?

6 Что выполняют функции ЛИНЕЙН, ТЕНДЕНЦИЯ, РОСТ, ЛГРФПРИБЛ, ПРЕДСКАЗ?

7 Каковы правила ввода и использования табличных формул?

8 Как на гистограмме исходных данных добавить линию тренда?

9 Как с помощью линии тренда отобразить прогнозируемые величины?

**Задание**

Вариант задания к данной лабораторной работе включает две задачи. Для каждой из них необходимо составить и определить:

1. Таблицу исходных данных, а также значений, полученных методами линейной и экспоненциальной регрессии.

2. Коэффициенты в уравнениях прямой и экспоненциальной кривой (функции ЛИНЕЙН и ЛГРФПРИБЛ), напишите уравнения прямой и экспоненциальной кривой для простой и множественной регрессии..

3. Погрешности (ошибки) прямой и экспоненциальной кривой. вычислений для коэффициентов и функций, коэффициенты детерминированности. Оценить, какой тип регрессии наилучшим образом подходит для вашего варианта задания.

4. Прогноз изменения данных, выполненный с использованием линей­ной и экспоненциальной регрессии (функции ТЕНДЕНЦИЯ, ПРЕДСКАЗ, РОСТ).

5. Построить гистограмму (или график) исходных данных для задачи 1 (одномерная регрессия), отобра­зить на ней линию тренда, а также соответствующее ей уравнение и коэффициент детерминированности.

***Варианты заданий.***

***Вариант 1***

 1 На рынке наблюдается стойкое снижение цен на компьютеры. Сде­лать прогноз, сколько необходимо будет снизить цену на компьютеры в следующем месяце в Вашей фирме, чтобы как минимум сравнять ее с ценой на ана­логичные компьютеры в конкурирующей фирме, если известна динамика изменения цен на них в конкурирующей фирме за последние 12 месяцев.

Для выполнения задания нужно ввести ряд из 12 ячеек с ценами конкурирующей фирмы , сделать прогноз цены на следующий месяц и др.(см. ***Задание***).

2 Известна структура расходов фирмы на рекламу в газетах, на радио, в журналах, на телевидении, на наружную рекламу ( в процентах от общей суммы ), а также оборот фирмы в каждом за последние 6 месяцев. Какой оборот можно ожидать в следующем месяце, если предполагается следующая структура расходов на рекламу: газеты-40%, журналы-40%, радио-5%, телевидение-14%, наружная реклама-1%.

Для выполнения задания нужно составить таблицу со столбцами вида:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Месяц | Х1 – газеты, % | Х2 –журналы, % | Х3 – радио, % | Х4 – телевид.,% | Х5 – наружн. рекл.% | Оборот, $. |
| Январь | 40 | 35 | 10 | 10 | 5 | 410000 |
| **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| Июнь | 38 | 42 | 8 | 15 | 7 | 425000 |

и сделать множественный регрессионный прогноз .(см. ***Задание***).

***Вариант 2***

1 Имеются данные о динамике продаж в расчете на душу населения по хлебобулочным продуктам и молочным изделиям, а также динамика изменения среднедушевого годового дохода за последние 10 лет. Для каждой группы товаров построить регрессионные модели, описывающие зависимость объемов продаж от размера доходов. Сделать прогноз об объемах продаж и размерах доходов на следующий год.

Для выполнение задания нужно составить таблицу вида:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Годы | 1 | 2 | **. . .** | 10 | 11 |
| х1 – хлеб, кг | 0,5 | 26,7 | **. . .** | 42,8 |  |
| х2 –молоко, л | 0,45 | 22 | **. . .** | 39,5 |  |
| у – доход, р. | 6600 | 7200 | **. . .** | 18250 |  |

и получить два уравнения – у =f (x1) и у= f(х2) , сделать прогноз на следующий год для рядов х1, х2, у и др. .(см. ***Задание***).

2 Руководство фирмы провело оценку качеств пяти рекламных агентов по следующим признакам: х1- эрудиция, х2- энергичность, х3- умение работать с людьми, х4 - внешность, х 5- знание предметной области. Полученные средние оценки, нормированные от 0 до 1, были сопоставлены с оценками эффективности деятельности агентов ( % успешных сделок от количества возможных). Определить, какую эффективность можно ожидать от рекламного агента, обладающего усредненными качествами. Сравнить ее со средней эффективностью упомянутых 5 агентов.

Исходные данные нужно ввести в таблицу вида:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G |
| 1 |  | х1-Эруд. | х2 -Энер | х3-Люди  | х4-Вн. | х 5-Зн. | Эф-ть |
| 2 | Агент 1 | 0,8 | 0,2 | 0,4 | 0,6 | 1,0 | 76% |
| **.** | **. . .** |  **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 6 | Агент 5 | 0,5 | 0,7 | 0,3 | 0,4 | 0,74 | 81% |
| 7 | Средняя эффективность пяти агентов |  |
| 8 | Ср. агент | 0,5 | 0,5 | 0,5 | 0,5 | 0,5 |  |

Массив ячеек B2-F6 заполняется произвольными числами от 0 до 1,

столбец G2-G6 – процентами удачных сделок по принципу «Чем выше уровень качеств агента, тем выше эффективность его работы», в ячейке G7 должна быть формула для вычисления среднего значения содержимого ячеек G2-G6, в ячейке G8 нужно вычислить значение эффективности для среднего агента по формуле, полученной в результате множественного регрессионного анализа работы пяти агентов. Остальные пункты – см. ***Задание***.

***Вариант 3***

1 Автомобильный салон имеет данные о количестве проданных автомобилей "Мерседес" и "БМВ" за последние 4 квартала. Учитывая тенденцию изменения объема продаж, определить, каких автомобилей необходимо закупить больше ( "Мерседес" или "БМВ" ) в следующем квартале?

Для выполнения задания нужно составить и заполнить таблицу вида

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 квартал | 2 квартал | 3 квартал | 4 квартал | Новый квартал |
| Мерседес | 10 | 12 | 15 | 18 |  |
| БМВ | 9 | 10 | 14 | 17 |  |

сделать прогноз продаж на новый квартал и выполнить другие пункты ***Задания.***

2 Известны следующие данные о 5 недавно проданных подержанных автомобилях: х1 - стоимость продажи, х2 - стоимость аналогичного нового автомобиля, х3 - год выпуска, х4 - пробег, х5- кол-во капитальных ремонтов, х6- экспертные заключения о состоянии кузова и техническом состоянии автомобилей ( по 10-бальной шкале ). Определить, сколько может стоить автомобиль с соответствующими характеристиками: 20 000 руб., 34 000 руб., 1990 г. , 140000 км. , 0, 6 – см. пример 4.

  ***Вариант 4*** 1 Определить минимально необходимый тираж ежемесячного журнала " Speed-Info " и возможный доход от размещения в нем рекламы в следующем месяце, если известны данные об объемах продаж этого журнала и доходах от размещения рекламы за прошедшие 12 месяцев (считать, что расценки на рекламу не менялись ).

Для выполнения задания нужно составить таблицу вида

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Месяц | 1 | 2 | **. . .** | 12 | 13 |
| Тираж | 100000 | 120000 | **. . .** | 145000 |  |
| Доход  | 128000 | 135000 | **. . .** | 172000 |  |

и заполнить ячейки за 12 месяцев условными данными. По этим данным нужно сделать линейный и экспоненциальный прогноз и др. (см. ***Задание***).

2 В целях привлечения покупателей и увеличения оборота фирма проводит стратегию ежемесячного снижения цен на свой товар. На основании

данных о динамике изменения цен, объемов продаж в данной фирме и еще в 3 конкурирующих фирмах за последние 12 месяцев сделать прогноз о том, возрастет ли объем продаж у данной фирмы при очередном снижении цен в следующем месяце, если предположить, что цены и объемы у конкурентов в следующем месяце будут средние за рассматриваемый период.

Для выполнения задания нужно составить таблицу вида

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | T | F | G | H | I |
| 1 | мес | Фирма | Конкурент 1 | Конкурент 2 | Конкурент 3 |
| 2 | 1 | У1- объём  | Х1- цена  |  Х2- объём | Х3- цена  | Х4- объём  | Х5 -цена  | Х6 -объём | Х7 -цена  |
| 3 | 2 | 10000 | 1875 | 12000 | 1720 | 12500 | 1710 | 11970 | 1700 |
| **. .**  | **. . .** | **. . .** | **. . .**  | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 13 | 12 | 13470 | 1510 | 14020 | 1488 | 13900 | 1490 | 14000 | 1475 |
| 14 | 13 |  |  |  |  |  |  |  |  |

***Вариант 5***

1 На основании данных о курсе американского доллара и немецкой марки в первом полугодии сделать прогноз о соотношении данных валют на второе полугодие. Во что будет выгоднее вкладывать деньги в конце года?

Для выполнения задания нужно составить таблицу вида

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Месяц | 1 | 2 | 3 | 4 | 5 | 6 | 7 | **. . .** | 12 |
| Доллар  | 24,.5 | 24,9 | 25,7 | 26,9 | 28,0 | 28,8 |  |  |  |
| Марка  | 72,1 | 76,3 | 79,6 | 85,3 | 89,7 | 90,9 |  |  |  |

и сделать линейный прогнозы на следующие 6 месяцев и др.(см. ***Задание***).

2 Известны данные за последние 6 месяцев о том, сколько раз выходила реклама фирмы, занимающейся недвижимостью, на телевидении – х1, радио – х2, в газетах и журналах – х3, а также количество звонков – у1  и количество совершенных сделок у2. Какое соотношение количества совершенных сделок к количеству звонков у (в %) можно ожидать в следующем месяце, если известно, сколько раз выйдет реклама в каждом из перечисленных средств массовой информации.

Для выполнения задания нужно составить и заполнить таблицу вида

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 |  | х1 | х2 | х3 | y = y2/y1**·**100% |
| 2 | Январь | 15 | 10 | 24 | 78% |
| **. .**  | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 7 | Июнь | 22 | 14 | 20 | 89% |
| 8 | Июль |  |  |  |  |

и выполнить применительно к таблице пункты ***Задания.***

***Вариант 6***

1 Для некоторого региона известен среднегодовой доход населения, а также данные о структуре расходов ( тыс. руб. в год ) за последние 5 лет по следующим статьям: питание – х1, жилье – х2, одежда – х3, здоровье – х4, транспорт –х5, отдых – х6, образование – х7. На основании известных данных провести анализ потребительского кредита ( или накопления ) в следующем году.

Для выполнения задания нужно составить таблицу вида

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Годы  | х1 | х2 | х3 | х4 | х5 | х6 | х7 |  Σхi | Доход  | Кредит  |
| 1 | 5 | 2 | 1 | 1 | 0,3 | 5 | 4 | 18,3 | 21,4 | 3,1 |
| **. . .** | **. .**  | **. .**  | **. .** | **. .** | **. . .** | **. .**  | **. . .** | **. . .** | ... | ... |
| 5 | 7 | 3 | 1 | 2 | 1,2 | 4 | 6,5 | 24,7 | 26,2 | 1,5 |

В ячейках столбца **Σхi**  должны быть записаны формулы, вычисляющие суммы всех расходов х1 + х2+...+ х7 в каждом году, в ячейках столбца **Доход** – **Доход** соответствующие среднегодовые доходы, в ячейках столбца **Кредит –** формулы разности содержимого ячеек с ежегодными доходами и затратами, т.е. **Кредит = Доход - Σхi**.. Затем для столбца **Кредит** нужно выполнить регрессионный прогноз на следующий год и другие пункты ***Задания.***

2 Для 10 однокомнатных квартир, расположенных в одном районе, известны следующие данные: общая площадь - х1, жилая площадь – х2, площадь кухни – х3, наличие балкона – х4, телефона – х5, этаж – х6, а также стоимость – х7. Определить, сколько может стоить однокомнатная квартира в этом районе без балкона, без телефона, расположенная на 1-ом этаже, общей площадью 28 кв. м, жилой- 16 кв. м, с кухней 6 кв. м.

***Вариант 7***

1 Определить возможный прирост населения ( кол-во человек на 1000 населения ) в 2005 году, если известны данные о кол-ве родившихся и умерших на 1000 населения в 1991-2000 годах.

2 После некоторого спада наметился рост объемов продаж матричных

принтеров. Используя данные об объемах продаж, ценах на матричные, струйные и лазерные принтеры, а также на их расходные материалы за последние 6 месяцев, определить возможный спрос на матричные принтеры в следующем месяце. Проанализируйте, связано ли увеличение спроса на матричные принтеры с уменьшением спроса на струйные и лазерные.

Для выполнения задания нужно составить и заполнить таблицу вида

|  |  |  |  |
| --- | --- | --- | --- |
|  | Матричные принтеры | Струйные принтеры | Лазерные принтеры |
|  | Спрос ,у1  | Цена, x1 | Р.мат z1 | Спрос,y2 | Цена,x2 | Р.мат.z2 | Спрос, y3 | Ценаx3 | Р.мат.z3 |
| 1 | 56 | 4172 | 174 | 26 | 2384 | 558 | 8 | 12517 | 1558 |
| **. .**  | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 6 | 75 | 4318 | 213 | 18 | 2543 | 768 | 6 | 14587 | 2200 |

сделать прогноз на седьмой месяц по уравнению у1= f(x1,z1), получить уравнение у1 = f(y2, x2, z2, y3, x3, z3) и проанализировать его. Если слагаемые y2 и y3 входят в регрессионное уравнение со знаком “-“ , то уменьшение спросов y2 и y3 ведёт к увеличению спроса у1. Выполнить другие пункты ***Задания.***

***Вариант 8***

1 Построить прогноз развития спроса населения региона на телевизоры, если известна динамика продаж телевизоров ( тыс. шт. ) и динамика численности населения данного региона ( тыс. чел. ) за последние 10 лет.

Для выполнения задания нужно составить таблицу из двух рядов (продаж телевизоров и численности населения по годам и сделать прогноз по обоим рядам на следующий год. Выполнить другие пункты ***Задания.***

2 Размещая рекламу в 4-х изданиях, фирма собрала сведения о поступивших на нее откликов - у и сопоставила их с данными об изданиях: х1- стоимость издания, х2 - стоимость одного блока рекламы, х3 - тираж, х4 - объем аудитории, х5- периодичность, х6- наличие телепрограммы. Какое количество откликов можно ожидать на рекламу в издании со следующими характеристиками: 15000 руб. , 10$ , 1000 экз. , 25000 чел. , 4 раза в месяц , без телепрограммы.

Для выполнения задания нужно составить и заполнить таблицу вида

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Данные  | Отклики-у | х1 | х2 | х3 | х4 | х5 | х6 |
| Издание 1 | 108 | 10000 | 13 | 700 | 15000 | 4 | 1 |
| **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| Издание 4 | 124 | 17850 | 11 | 1200 | 32000 | 26 |  1 |
| Прогноз  |  | 15000 | 10 | 1000 | 25000 | 4 | 0 |

сделать прогноз при заданных характеристиках и выполнить другие пункты ***Задания.***

***Вариант 9.*** 1 Размещая свою рекламу в двух печатных изданиях одновременно, фирма собрала сведения о количестве поступивших звонков и количестве совершенных сделок по объявлениям в каждом из указанных изданий за последние 12 месяцев. Определить, в каком из изданий и насколько эффективность размещения рекламы в следующем месяце будет больше?

Для выполнения задания нужно составить таблицу вида:

|  |  |  |
| --- | --- | --- |
|  | Издание 1 | Издание 2 |
| Месяцы | Звонки | Сделки | Звонки | Сделки |
| 1 | 98 | 66 | 112 | 79 |
| 2 | 105 | 72 | 143 | 85 |
| **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 12 | 139 | 88 | 155 | 97 |
| 13 лин. |  |  |  |  |
| 13-эксп. |  |  |  |  |

Эффективность определяется как **сделки /звонки**. Сделать линейный и экспоненциальный прогнозы по обоим изданиям, выполнить другие пункты ***Задания.***

2 Пусть комплект мягкой мебели ( диван + 2 кресла ) характеризуется следующими признаками ( 1- есть, 0- нет ) : х1- деревянные подлокотники, х2- велюровое покрытие, х3 - кресло-кровать, х4 - угловой диван, х5- раскладывающийся диван, х6 - место для хранения белья. На основании данных о стоимости 5 комплектов мягкой мебели, для которых известны перечисленные признаки, сделать вывод о возможной стоимости комплекта с обычным раскладывающимся диваном , с местом для белья, без деревянных подлокотников, с обычными креслами.

Для выполнения задания нужно составить таблицу

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Признаки | х1 | х2 | х3 | х4 | х5 | х6 | у1 - стоимость |
| Комплект 1 | 1 | 1 | 1 | 0 | 1 | 1 | 12560 р. |
| **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **...** |
| Комплект 5 | 0 | 1 | 0 | 1 | 1 | 1 | 10980 р. |
| Прогноз | 0 | 0 | 0 | 0 | 1 | 1 |  |

сделать прогноз и выполнить другие пункты ***Задания.***

***Вариант 10***

1Для двух радиостанций известны данные об изменении объема аудитории и динамике роста цен за 1 минуту эфирного времени за последние 12 месяцев. Определить, для какой радиостанции стоимость одного контакта со слушателем в следующем месяце будет меньше?

Для выполнения задания нужно составить и заполнить таблицу вида

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 |  | Радиостанция 1 | Радиостанция 2 |
| 2 | Месяц | Аудитория | Цена 1 мин. | Аудитория | Цена 1 мин. |
| 3 | 1 | 250000 | 8000 | 300000 | 7560 |
| 4 | 2 | 540000 | 6500 | 450000 | 6340 |
| **...** | **...** | **...** | **...** | **...** | **...** |
| 14 | 12 | 1108000 | 4700 | 1025000 | 4540 |
| 15 | Прогноз |  |  |  |  |
| 16 | Контакт |  |  |  |  |

В строке «Контакт» в ячейках С8 и D8 должны быть записаны формулы =С7/B7 и =E7/D7 соответственно, вычисляющие стоимость 1 мин. эфира для одного слушателя в прогнозируемом месяце. Прогноз нужно выполнить для линейного и экспоненциального приближений и выбрать более достоверный, а также сделать другие пункты ***Задания.***

2 На основании данных ежемесячных исследований известна динамика рейтинга банка ( в условных единицах ) за последние 6 месяцев в следующих сферах:

а) менеджмент и технология – х1;

б) менеджеры и персонал – х2;

в) культура банковского обслуживания – х3;

 г) имидж банка на рынке финансовых услуг – х4;

д) реклама банка – х5.

 Определить возможное изменение количества вкладчиков данного банка в следующем месяце, если известны значения сфер рейтинга и количество вкладчиков в каждом из рассматриваемых 6 месяцев.

Для выполнения задания нужно составить и заполнить таблицу

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G |
| 1 | Месяц | х1 | х2 | х3 | х4 | х5 | Кол-во вкладчиков |
| 2 | 1 | 4 | 5 | 8 | 7 | 6 | 130000 |
| **. .**  | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** | **. . .** |
| 7 | 6 | 7 | 7 | 9 | 9 | 6 | 148000 |
| 8 | Прогноз |  |  |  |  |  |  |

и выполнить другие пункты ***Задания.***

**Содержание отчёта**

1Название, цель, содержание работы

2 Задание своего варианта

3 Письменные ответы на контрольные вопросы

4 Выводы по работе

В отчете по лабораторной работе должны быть записаны все выполненные пункты **Задания** (таблицы, уравнения, прогнозируемые значения, стандартные ошибки, коэффициенты детерминированности, графики или гистограммы с приближающими кривыми и линиями тренда и др.