**МАТЕМАТИЧЕСКИЕ МЕТОДЫ**

**ПРОГНОЗИРОВАНИЯ ПРОДАЖ**

**Лабораторный практикум**

# прогнозирование по Парной линейной регрессионной модели

***Цель****: Освоить методы построения линейного уравнения парной регрессии с помощью ЭВМ, научиться получать и анализировать основные характеристики регрессионного уравнения.*

Рассмотрим методику построения регрессионного уравнения на примере.

***ПРИМЕР.*** 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 А1-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 В2-М2). Переходим на закладку «Ряд» и в поле «Значения Х» делаем ссылку на ячейки В1-М1, обводя их курсором. Нажимаем «Готово» Как видно из графика, точки хорошо укладываются на прямую линию, поэтому будем находить уравнение линейной регрессии вида .

Для нахождения коэффициентов *а* и *b* уравнения регрессии служат функции НАКЛОН и ОТРЕЗОК. категории «Статистические». Вводим в А5 подпись «*а*=» а в соседнюю ячейку В5 вводим функцию НАКЛОН, ставим курсор в поле «Изв\_знач\_у» задаем ссылку на ячейки В2-М2, обводя их мышью. Аналогично в поле «Изв\_знач\_х» даем ссылку на В1-М1. Результат 1,923921. Найдем теперь коэффициент b. Вводим в А6 подпись «b=», а в В6 функцию ОТРЕЗОК с теми же параметрами, что и у функции НАКЛОН. Результат 12,78151. Следовательно, уравнение линейной регрессии есть .

Построим график уравнения регрессии. Для этого в третью строчку таблицы введем значения функции регрессии в заданных точках Х (первая строка) - . Для получения этих значений используется функция ТЕНДЕНЦИЯ категории «Статистические». Вводим в А3 подпись «Y(X)» и, поместив курсор в В3, вызываем функцию ТЕНДЕНЦИЯ. В полях «Изв\_знач\_у» и «Изв\_знач\_х» даем ссылку на В2-М2 и В1-М1. В поле «Нов\_знач\_х» вводим также ссылку на В1-М1. В поле «Константа» вводят 1, если уравнение регрессии имеет вид , и 0, если . В нашем случае вводим единицу. Функция ТЕНДЕНЦИЯ является массивом, поэтому для вывода всех ее значений выделяем область В3-М3 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 В3-М3. Переходим на закладку «Ряд» и в поле «Значения Х» вводим ссылку на В1-М1, 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 В2-М3. Переходим на закладку «Ряд» и в поле «Подписи оси Х» вводим ссылку на В1-М1, нажимаем «Готово». Результат – две линии (Синяя – исходные данные, красная – уравнение регрессии). Видно что линии мало различаются между собой.

Для вычисления коэффициента корреляции служит функция ПИРСОН. Размещаем графики так, чтобы они располагались выше 25 строки, и в А25 делаем подпись «Корреляция», в В25 вызываем функцию ПИРСОН, в полях которой «Массив 1» и «Массив 2» вводим ссылки на исходные данные В1-М1 и В2-М2. Результат 0,993821. Коэффициент детерминации – это квадрат коэффициента корреляции . В А26 делаем подпись «Детерминация», а в В26 – формулу «=В25\*В25». Результат 0,987681.

Однако, в Excel существует одна функция, которая рассчитывает все основные характеристики линейной регрессии. Это функция ЛИНЕЙН. Ставим курсор в В28 и вызываем функцию ЛИНЕЙН, категории «Статистические». В полях «Изв\_знач\_у» и «Изв\_знач\_х» даем ссылку на В2-М2 и В1-М1. 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 нашем случае ставим туда единицу. Функция возвращает массив размером 2 столбца и 5 строк. После ввода выделяем мышью ячейки В28-С32 и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица значений, числа в которой имеют следующий смысл:

|  |  |
| --- | --- |
| Коэффициент *а* | Коэффициент *b* |
| Стандартная ошибка  | Стандартная ошибка  |
| Коэффициент детерминации  | Среднеквадратическое отклонение *у* |
| *F* – статистика | Степени свободы *п*-2 |
| Регрессионная сумма квадратов  | Остаточная сумма квадратов  |

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 о адекватности регрессионной модели. Данное число нудно 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 нашем случае это условие выполняется, что говорит о хорошей регрессии.

# прогнозирование по Парной нелинейной регрессионной модели с линейным ядром

***Цель****: Освоить методы построения основных видов нелинейных уравнений парной регрессии с помощью ЭВМ (внутренне линейные модели), научиться получать и анализировать показатели качества регрессионных уравнений.*

Рассмотрим случай, когда нелинейные модели с помощью преобразования данных можно свести к линейным (внутренни линейные модели).

***ПРИМЕР.*** Некоторая организация желает исследовать зависимость полученной прибыли Y (сотни тыс. руб.) от вложения средств в научные разработки выпускаемой продукции Х (тыс. руб.). Для этого рассматриваются 4 регрессионных уравнения: линейное: , гиперболическое , экспоненциальное  и степенное . В результате наблюдений, получены данные:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Прибыль Y | 5 | 6 | 8 | 11 | 16 | 22 | 29 | 35 | 44 | 57 | 83 |
| Вложения Х | 2 | 4 | 7 | 9 | 10 | 12 | 15 | 16 | 20 | 22 | 25 |

 Введем данные в таблицу вместе с подписями (ячейки А1-L2). Оставим свободными три строчки ниже таблицы для ввода преобразованных данных, выделим первые пять строк, проведя по левой серой границе по числам от 1 до 5 и выбрать какой либо цвет (светлый – желтый или розовый) раскрасить фон ячеек. Далее, начиная с A6, выводим параметры линейной регрессии. Для этого в ячейку А6 делаем подпись «Линейная» и в соседнюю ячейку В6 вводим функцию ЛИНЕЙН (категория «Статистические», см. предыдущую лабораторную работу). В полях «Изв\_знач\_у» и «Изв\_знач\_х» даем ссылку на В1-L1 и В2-L2, следующие два поля принимают значения по единице. Далее обводим область ниже в 5 строчек и левее в 2 строки (ячейки В6-С10) и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица с параметрами регрессии, из которых наибольший интерес представляет коэффициент детерминации в первом столбце третий сверху. В нашем случае он равен R1=0,90627178. Значение F-критерия, позволяющего проверить адекватность модели F1=87,02230833 (четвертая строка, первый столбец). Уравнение регрессии равно  (коэффициенты *а* и *b* приведены в ячейках В6 и С6).

Определим аналогичные характеристики для других регрессий и в результате сравнения коэффициентов детерминации найдем лучшую регрессионную модель. Рассмотрим гиперболическую регрессию. Для ее получения преобразуем данные. В третьей строек в ячейку А3 введем подпись «1/х» а в ячейку В3 введем формулу «=1/В2». Растянем автозаполнением данную ячейку на область В3-L3. Получим характеристики регрессионной модели. В ячейку А12 введем подпись «Гипербола», а в соседнюю функцию ЛИНЕЙН. В полях «Изв\_знач\_у» и «Изв\_знач\_х» даем ссылку на В1-L1 и преобразованные данные аргумента х – В3-L3, следующие два поля принимают значения по единице. Далее обводим область ниже в 5 строчек и левее в 2 строки и нажимаем F2 и Ctrl+Shift+Enter. Получаем таблицу параметров регрессии. Коэффициент детерминации в данном случае равен R2=0,345994664, что намного хуже, чем в случае линейной регрессии. F-статистика равна F2=4,761355604. Уравнение регрессии равно .

Рассмотрим экспоненциальную регрессию. Для ее линеаризации получаем уравнение  , где , , . Видно, что надо сделать преобразование данных - у заменить на ln y. Ставим курсор в ячейку А4 и делаем заголовок «ln y». Ставим курсор в В4 и вводим формулу LN (категория «Математические»). В качестве аргумента делаем ссылку на В1. Автозаполнением распространяем формулу на четвертую строку на ячейки В4-L4. Далее в ячейке F6 задаем подпись «Экспонента» и в соседней G6 вводим функцию Линейн , аргументами которой будут преобразованные данные В4-L4 (в поле «Изв\_знач\_у»), а остальные поля такие же как и для случая линейной регрессии (В2-L2, 1, 1). Далее обводим ячейки G6-H10 и нажимаем F2 и Ctrl+Shift+Enter. Результат R3= 0,979276, F3= 425,2748, что говорит об очень хорошей регрессии. Для нахождения коэффициентов уравнения регрессии ставим курсор в J6 и делаем заголовок «а=», а в соседней К6 формулу «=EXP(H6)», в J7 даем заголовок «b=» а в К7 формулу «=G6». Уравнение регрессии есть .

Рассмотрим степенную регрессию. Для ее линеаризации получаем уравнение , где , , , . Видно, что надо сделать преобразование данных - у заменить на ln y и х заменить на ln x. Строчка с ln y у нас уже есть. Преобразуем переменные х. В ячейку А5 даем подпись «ln x», а в В5 и вводим формулу LN (категория «Математические»). В качестве аргумента делаем ссылку на В2. Автозаполнением распространяем формулу на пятую строку на ячейки В5-L5. Далее в ячейке F12 задаем подпись «Степенная» и в соседней G12 вводим функцию Линейн , аргументами которой будут преобразованные данные В4-L4 (в поле «Изв\_знач\_у»), и В5-L5 (в поле «Изв\_знач\_х»), остальные поля – единицы. Далее обводим ячейки G12-H16 и нажимаем F2 и Ctrl+Shift+Enter. Результат R4= 0,895786, F4= 77,36103, что говорит об хорошей регрессии. Для нахождения коэффициентов уравнения регрессии ставим курсор в J12 и делаем заголовок «а=», а в соседней К12 формулу «=EXP(H12)», в J13 даем заголовок «b=» а в К13 формулу «=G12». Уравнение регрессии есть .

 Проверим, все ли уравнения адекватно описывают данные. Для этого нужно сравнить F-статистики каждого критерия с критическим значением. Для его получения вводим в А21 подпись «F-критическое», а в В21 функцию FРАСПОБР, аргументами которой вводим соответственно «0,05» (уровень значимости), «1» (число факторов Х в строке «Уровень значимости 1») и «9» (степень свободы 2 = n-2). Результат 5,117357. Видно, что *F* – статистика для первой третьей и четвертой регрессионной модели больше, чем *F* – критическое, значит эти модели адекватны. А гиперболическая регрессия неадекватна, т.к.  . Для того, чтобы определить, какая модель наилучшим образом описывает данные, сравним индексы детерминации для каждой модели . Наибольшим является R3= 0,979276. Значит, опытные данные лучше описывать моделью .

1. **ПРОГНОЗИРОВАНИЕ С ПОМОЩЬЮ ПОЛИНОМИАЛЬНОЙ МОДЕЛИ**

***Цель:*** *По опытным данным построить уравнение регрессии вида* .

 ***ПРИМЕР.*** Исследуется зависимость объема проданного товара *Y* от доходов населения *X*. Так, как зависимость спроса от дохода определяется качеством товара и разная для разных слоев населения, то эта зависимость может иметь сложный нелинейный характер с несколькими экстремумами. Для описания таких зависимостей лучше использовать полиномиальную регрессионную модель. Опытные данные приведены в таблице.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Х | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 | 23 |
| Y | 5 | 7 | 12 | 13 | 11 | 8 | 5 | 3 | 2 | 4 | 6 | 9 |

Введем эти данные в электронную таблицу вместе с подписями в ячейки А1-М2. Построим график. Для этого обведем данные Y (ячейки В2-М2), вызываем мастер диаграмм, выбираем тип диаграммы «График», вид диаграммы – график с точками (второй сверху левый), нажимаем «Далее», переходим на закладку «Ряд» и в поле «Подписи оси Х» делаем ссылку на В2-М2, нажимаем «Готово». Видно, что график имеет 2 экстремума и один перегиб, поэтому его можно приблизить полиномом 3 степени . Для нахождения коэффициентов  нужно решить систему уравнений:



Рассчитаем суммы. Для этого в ячейку А3 вводим подпись «X^2», а в В3 вводим формулу «=В1\*В1» и автозаполнением переносим ее на всю строку В3-М3. В ячейку А4 вводим подпись «X^3», а в В4 формулу «=В1\*В3» и автозаполнением переносим ее на всю строку В4-М4. В ячейку А5 вводим «X^4», а в В5 формулу «=В4\*В1» , автозаполняем строку. В ячейку А6 вводим «X^5», а в В6 формулу «=В5\*В1» , автозаполняем строку. В ячейку А7 вводим «X^6», а в В7 формулу «=В6\*В1» , автозаполняем строку. В ячейку А8 вводим «X\*Y», а в В8 формулу «=В2\*В1» , автозаполняем строку. В ячейку А9 вводим «X^2\*Y», а в В9 формулу «=В3\*В2» , автозаполняем строку. В ячейку А10 вводим «X^3\*Y», а в В10 формулу «=В4\*В2» , автозаполняем строку. Теперь считаем суммы. Выделяем другим цветом столбец N, щелкнув по заголовку и выбрав цвет. В ячейку N1 помещаем курсор и щелкнув по кнопке автосуммы со значком Σ, вычисляем сумму первой строки. Автозаполнением переносим формулу на ячейки N1-N10.

 Решаем теперь систему уравнений. Для этого вводим основную матрицу системы. В ячейку А13 вводим подпись «А=», а в ячейки матрицы В13-Е16 вводим ссылки, отраженные в таблице

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *В* | *С* | *D* | *Е* |
| *13* | **=N7** | **=N6** | **=N5** | **=N4** |
| *14* | **=N6** | **=N5** | **=N4** | **=N3** |
| *15* | **=N5** | **=N4** | **=N3** | **=N1** |
| *16* | **=N4** | **=N3** | **=N1** | **12** |

Вводим также правые части системы уравнений. В G13 вводим подпись «В=», а в Н13-Н16 вводим, соответственно ссылки на ячейки «=N10», «=N9», «=N8», «=N2». Решаем систему матричным методом. Из высшей математики известно, что решение равно . Находим обратную матрицу. Для этого в ячейку I13 вводим подпись «А обр.» и, поставив курсор в J13 задаем формулу МОБР (категория «Математические»). В качестве аргумента «Массив» даем ссылку на ячейки B13:E16. Результатом также должна быть матрица размером 4х4. Для ее получения обводим ячейки J13-M16 мышью, выделяя их и нажимаем F2 и Ctrl+Shift+Enter. Результат – матрица  . Найдем теперь произведение этой матрицы на Столбец В (ячейки Н13-Н16). Вводим в ячейку А18 подпись «Коэффициенты» и в В18 задаем функцию МУМНОЖ (категория – «Математические»). Аргументами функции «Массив 1» служит ссылка на матрицу  (ячейки J13-M16), а в поле «Массив 2»даем ссылку на столбец В (ячейки H13-H16). Далее выделяем В18-В21 и нажимаем F2 и Ctrl+Shift+Enter. Получившийся массив – коэффициенты уравнения регрессии . В результате получаем уравнение регрессии вида:

 .

Построим графики исходных данных и полученных на основании уравнения регрессии. Для этого в ячейку А11 вводим подпись «Регрессия» и в В11 вводим формулу «=$B$18\*B4+$B$19\*B3+$B$20\*B1+$B$21». Автозаполнением, переносим формулу в ячейки В11-М11. Для построения графика выделяем ячейки В11-М11 и, удерживая клавишу Ctrl, выделяем также ячейки В2-М2. Вызываем мастер диаграмм, выбираем тип диаграммы «График», вид диаграммы – график с точками (второй сверху левый), нажимаем «Далее», переходим на закладку «Ряд» и в поле «Подписи оси Х» делаем ссылку на В2-М2, нажимаем «Готово». Видно, что кривые почти совпадают.

# Прогнозирование по Линейной множественной регрессионной модели

***Цель:*** *По опытным данным построить уравнение множественной линейной регрессии и определить его характеристики.*

 ***ПРИМЕР.*** Некоторая организация занимается торговлей компьютерами. Она определила, что на количество продаж основное влияние оказывают следующие факторы: цена товара , затраты на рекламу  и число конкурирующих организаций в регионе . Результаты наблюдений приведены в таблице

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Х1 | 20 | 20 | 18 | 17 | 17 | 19 | 18 | 16 | 16 | 16 | 15 | 15 | 14 | 14 |
| Х2 | 37 | 38 | 36 | 42 | 47 | 55 | 53 | 54 | 49 | 50 | 52 | 52 | 51 | 54 |
| Х3 | 3 | 4 | 3 | 5 | 4 | 3 | 2 | 3 | 2 | 2 | 1 | 2 | 1 | 3 |
| Y | 112 | 132 | 129 | 134 | 132 | 137 | 139 | 139 | 138 | 143 | 141 | 146 | 148 | 150 |

Построим линейное уравнение множественной регрессии. Для этого предварительно исследуем матрицу парных коэффициентов корреляции. Вводим исходные данные вместе с подписями в ячейки А1-О4. Для построения матрицы парной корреляции вызываем меню «Сервис/Анализ данных» (если пункт меню отсутствует, то вызываем «Сервис/Надстройки» и ставим галочку напротив строки «Пакет анализа»). Выбираем пункт «Корреляция». В появившемся окне в поле «Входной интервал» задаем ссылку на таблицу – А1-О4. Указываем группирование «По строкам». Ставим флажок в «Метки в первом столбце» (так как в ссылках на таблицу указаны подписи строк). В области «Параметры вывода» ставим флажок напротив «Выходной интервал» и напротив в поле даем ссылку на какую-либо ячейку, откуда будет осуществляться вывод данных, например А7. Нажимая «ОК», получаем нижнюю половину матрицы парной корреляции. Для общей оценки мультиколлинеарности факторов и адекватности регрессионной модели рассчитаем определители матриц . Сформируем полную матрицу парных коэффициентов корреляции. В С8 задаем формулу «=В9», в D8 ссылку «=B10», в D9 – «=С10», в Е8 – «=В11», в Е9 – «=С11», в Е10 – «D11». Далее, для вычисления определителей в ячейку А13 вводим заголовок «=» и в В13 ставим курсор и задаем функцию «МОПРЕД» (категория «Математические»), в которой аргумент «Массив» является ссылкой на ячейки B8:E11. В ячейку А14 вводим заголовок «=» и в В14 ставим курсор и задаем функцию «МОПРЕД» с аргументом «Массив» - ссылкой на B8:D10. Результат - 0,427104 (ближе к 0, чем к 1), что говорит о достаточно высокой общей мультикорреляции. Найдем теперь коэффициент множественной корреляции. В А15 вводим «Rх=», а в В15 формулу «=КОРЕНЬ(1-В13/В14)». Результат 0,883788 говорит о достаточно высокой связи между фактором и функцией отклика.

Проведем теперь отбор факторов. Рассмотрим матрицу коэффициентов парной корреляции. Видно, что . Факторы можно считать коллинеарными (интеркоррелированными), если их парный коэффициент по модулю больше 0,7. В нашем случае таких пар факторов нет. Все факторы можно считать независимыми друг от друга и использовать в регрессионном уравнении. Определим теперь влияние каждого фактора  на функцию отклика *Y*. Для этого рассмотрим коэффициенты парной корреляции . Видно, что третий коэффициент -0,45 намного меньше по модулю, чем примерная граница 0,7, поэтому влияние третьего фактора Х3 на результат мало и его можно отбросить из рассмотрения.

Копируем теперь на А18-О20 функцию отклика и значимые факторы (1, 2 и 4 строки, соответственно А1-О2 копируем в А18-О19, а А4-О4 – в А20-О20). В ячейку А22 вводим заголовок «Линейная» и в соседнюю В22 вводим функцию, определяющую параметры линейной регрессии «=ЛИНЕЙН» (категория «Статистические»). Аргументы функции: «Изв\_знач\_у» - B20:O20 (значения функции), «Изв\_знач\_х» - B18:O19 (значения двух значимых аргументов), «Константа» – 1 (расчет свободного члена), «Стат» - 1 (вывод дополнительных характеристик регрессии). Результат вычислений функции ЛИНЕЙН в случае функции нескольких переменных  имеет вид:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Коэффициент *аk* | Коэффициент *ak-*1 | … | Коэффициент  | Коэффици-ент  |
| Стандартная ошибка  | Стандартная ошибка  | … | Стандартная ошибка  | Стандартная ошибка  |
| Коэффициент детерминации  | Среднеквадратическое отклонение *у* | … | *нет данных* | *нет данных* |
| *F* – статистика | Степень свободы *п*-*k*-1 | … | *нет данных* | *нет данных* |
| Регрессионная сум-ма квадратов  | Остаточная сумма квадратов  | … | *нет данных* | *нет данных* |

В нашем случае имеем два фактора , поэтому обводим 5 строк и три столбца В22-D26 и нажимаем F2 и Ctrl+Shift+Enter.

Первая строка результата – значения параметров регрессионного уравнения  - числа . Следовательно, уравнение регрессии есть . Вторая строка – стандартные ошибки коэффициентов. Все они меньше самих коэффициентов, это значит, что коэффициенты значимы. В ячейках D24-D26 стоят значения «#Н/Д» (нет данных), как и должно быть в соответствии с таблицей.

Коэффициент детерминации равен 0,7783. Видно, что F-критерий регрессионной модели равен 19,3097. Проверим модель на адекватность. Вычислим критическое значение статистики. Вводим в ячейку Н22 подпись «F-критическое» а в I22 вводим функцию FРАСПОБР, имеющую аргументы: «Вероятность» – уровень значимости, если он не задан в условии (как в нашем случае), то обычно его принимаем 0,05. Аргумент «Степени\_свободы\_1» - число независимых переменных (у нас их две - ). Аргументом «Степени\_свободы\_2» служит число, показанное в ячейке С25 (в данном примере – 11). Видно, что F-статистика больше ее критического значения, поэтому модель адекватна.

# Прогнозирование по неЛинейной множественной регрессионной модели

***Цель****: По опытным данным построить уравнение множественной нелинейной регрессии и определить его характеристики.*

 Рассмотрим случай, когда нелинейное уравнение с помощью преобразования данных можно привести к линейному виду. В связи с этим, данная лабораторная работа будет аналогична предыдущей.

Рассмотрим следующий пример.

***ПРИМЕР.*** Предприятие выпускает продукцию, количество которой за месяц Y (тыс. шт.) зависит от затрат материальных ресурсов  (т.), трудозатрат  (тыс. час.) и энергозатрат  (млн. кВт). При расширении производства наблюдалась следующая эмпирическая зависимость между выпуском Y и затратами ресурсов .

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Х1 | 16 | 20 | 18 | 22 | 21 | 24 | 27 | 26 | 28 | 31 | 35 | 34 | 33 | 34 |
| Х2 | 50 | 55 | 58 | 50 | 57 | 59 | 62 | 64 | 59 | 64 | 59 | 62 | 65 | 70 |
| Х3 | 7 | 6 | 7 | 8 | 10 | 8 | 9 | 7 | 11 | 10 | 12 | 11 | 9 | 13 |
| Y | 45,0 | 50,3 | 54,1 | 55,1 | 60,8 | 65,6 | 68,8 | 66,6 | 73,2 | 81,9 | 91,8 | 86,1 | 83,1 | 93,1 |

Из теории производственных функций известно, что зависимость результирующего признака (функции откликов) от факторов имеет вид . Вводим исходные данные вместе с подписями в ячейки А1-О4. Чтобы привести уравнение к линейному виду нужно прологарифмировать уравнение . Вводим вместо исходных данных их логарифмы. Для этого в ячейки А5-А8 вводим подписи «Ln X1», «Ln X2», «LnX3», «LnY». Ставим курсор в ячейку В5 и вводим функцию LN (категория математические) с аргументом «Число» В1, которое отобразится в строке формул в виде «=LN(В1)», затем переносим формулу на все данные, автозаполняя ячейки В5-О8. После этого исследуем матрицу парных коэффициентов корреляции. Для построения матрицы вызываем меню «Сервис/Анализ данных» и выбираем пункт «Корреляция». В появившемся окне в поле «Входной интервал» задаем ссылку на преобразованные данные – А5-О8. Указываем группирование «По строкам». Ставим флажок в «Метки в первом столбце» (так как в ссылках на таблицу указаны подписи строк). В области «Параметры вывода» ставим флажок напротив «Выходной интервал» и напротив в поле даем ссылку на какую-либо ячейку, откуда будет осуществляться вывод данных, например А10 и нажимаем «ОК». Для общей оценки мультиколлинеарности факторов и адекватности регрессионной модели рассчитаем определители матриц . Сформируем полную матрицу парных коэффициентов корреляции. В С11 задаем формулу «=В12», в D11 ссылку «=B13», в D12 – «=С13», в Е11 – «=В14», в Е12 – «=С14», в Е13 – «D14». Далее, для вычисления определителей в ячейку А16 вводим заголовок «=» и в В16 ставим курсор и задаем функцию «МОПРЕД» (категория «Математические»), в которой аргумент «Массив» является ссылкой на ячейки B11:E114. В ячейку А17 вводим заголовок «=» и в В17 ставим курсор и задаем функцию «МОПРЕД» с аргументом «Массив» - ссылкой на B11:D13. Результат - 0,163303, он близок к нулю, что говорит о достаточно сильной общей мультикорреляции факторов между собой. Найдем теперь коэффициент множественной корреляции. В А18 вводим «Rх=», а в В18 формулу «=КОРЕНЬ(1-В16/В17)». Результат 0,993187 говорит о достаточно высокой связи между фактором и функцией отклика.

Проведем теперь отбор факторов. Видно, что первый фактор сильно связан и со вторым и с третьим, поэтому его выводим их регрессионной модели. Одновременно видно, что влияние второго и третьего фактора на функцию Y достаточно сильно, поэтому, принимаем к рассмотрению регрессионную модель . В строках с номерами 20-22 копируем значимые факторы. Для этого в А20-А22 вводим подписи «Ln X2, LnX3, LnY», а в В20 вводим функцию «=LN(B2)» и автозаполняем ее В20-О22. В ячейку А24 вводим заголовок «Линейная» и в соседнюю В24 вводим функцию, определяющую параметры линейной регрессии «=ЛИНЕЙН». Аргументы функции: «Изв\_знач\_у» - B22:O22, «Изв\_знач\_х» - B20:O21, «Константа» = 1, «Стат» = 1. Далее обводим 5 строк и три столбца В24-D28 (т.к. находим параметры функции Y и двух факторов ), и нажимаем F2 и Ctrl+Shift+Enter. Первая строка результата – значения параметров преобразованного регрессионного уравнения. Чтобы получить данные исходного уравнения, вводим в G10, G11 и G12 подписи «», а в соседние ячейки Н10, Н11 и Н12 формулы “=C24”, “=B24” и “=EXP(D24)”. В результате уравнение регрессии есть . Вторая строка – стандартные ошибки коэффициентов. Все они меньше самих коэффициентов, это значит, что коэффициенты значимы. Коэффициент детерминации равен 0,892. Видно, что F-критерий регрессионной модели равен 45,359. Проверим модель на адекватность. Вычислим критическое значение статистики. Вводим в ячейку Н24 подпись «F-критическое» а в I24 вводим функцию FРАСПОБР, имеющую аргументы: «Вероятность» – 0,05 (т.к. ), «Степени\_свободы\_1» - 2 (число независимых переменных равно двум - ). Аргументом «Степени\_свободы\_2» служит число, показанное в ячейке С27 (в данном примере – 11). Видно, что F-статистика больше ее критического значения, поэтому модель адекватна.

# Модели прогнозирования с фиктивными факторами

***Цель:*** *Научиться составлять множественные регрессионные модели, включающие в себя фиктивные переменные, проводить анализ этих моделей.*

Фиктивные факторы вводятся в случае, если показатель, влияющий на функцию отклика, нельзя измерить количественно. Рассмотрим их применение на следующем примере.

 ***ПРИМЕР.*** Строительная организация продает облицовочную плитку в трех городах: Воронеже, Липецке и Курске. Маркетинговая служба хочет определить влияние отчислений на рекламу X (тыс. р.) на количество проданной продукции Y (млн. шт.). При этом предполагается, что зависимость фактора Х на функцию Y линейная и степень влияния факторов друг на друга (коэффициент *а* уравнения регрессии) во всех городах примерно одинаков, но различный спрос на продукцию (свободный член уравнения). Организация желает включить в регрессионную модель такой фактор как «город». Имеются следующие статистические данные.

**г. Воронеж**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 25 | 14 | 19 | 27 | 33 | 31 | 12 | 16 | 28 |
| Y | 37 | 24 | 25 | 39 | 42 | 43 | 22 | 27 | 27 |

 **г. Липецк**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 13 | 18 | 19 | 24 | 21 | 17 | 31 | 29 | 16 | 27 | 22 | 21 |
| Y | 30 | 33 | 33 | 41 | 35 | 31 | 45 | 45 | 30 | 40 | 33 | 32 |

**г. Курск**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 16 | 15 | 11 | 19 | 27 | 31 | 29 | 22 | 19 | 26 |
| Y | 22 | 20 | 18 | 25 | 28 | 35 | 32 | 27 | 26 | 31 |

Введем фиктивные переменные

 

В результате получаем регрессионную функцию трех переменных , а результаты наблюдений можно записать как

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Y | 37 | 24 | 25 | 39 | 42 | 43 | 22 | 27 | 27 | 30 | 33 | 33 | 41 | 35 | 31 | 45 |
| X | 25 | 14 | 19 | 27 | 33 | 31 | 12 | 16 | 28 | 13 | 18 | 19 | 24 | 21 | 17 | 31 |
| Z1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Z2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Продолжение таблицы с данными:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Y | 45 | 30 | 40 | 33 | 32 | 22 | 20 | 18 | 25 | 28 | 35 | 32 | 27 | 26 | 31 |
| X | 29 | 16 | 27 | 22 | 21 | 16 | 15 | 11 | 19 | 27 | 31 | 29 | 22 | 19 | 26 |
| Z1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Z2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Вводим эти данные в электронную таблицу. В А1-А4 вводим подписи “Y”, “X”, “Z1”, “Z2”. В ячейки В1-AF4 вводим данные из таблицы (вводится одна длинная таблица). Данные вводятся в 4 строки и 31 столбец, не считая заголовки. Посмотрим, имеются ли мультиколлинеарные факторы. Для этого находим матрицу коэффициентов интеркорреляции. Вызываем меню «Сервис/Анализ данных» и выбираем пункт «Корреляция». В появившемся окне в поле «Входной интервал» задаем ссылку на данные – факторы: А2-AF4. Указываем группирование «По строкам». Ставим флажок в «Метки в первом столбце» (так как в ссылках на таблицу указаны подписи строк). В области «Параметры вывода» ставим флажок напротив «Выходной интервал» и напротив в поле даем ссылку на какую-либо ячейку, откуда будет осуществляться вывод данных, например А7 и нажимаем «ОК». Ставим курсор в С8 и делаем ссылку «=В9», в D8 – ссылку «=В10», в D9 – ссылку «=С10». Ставим курсор в ячейку G7, вводим заголовок «=» и в H7 ставим курсор и задаем функцию «МОПРЕД» с аргументом «Массив» - ссылкой на B8:D10. Результат - 0,7352005, близок к единице, что говорит о достаточно слабой зависимости факторов между собой. Все факторы оставляем в регрессионной модели. вычисляем параметры регрессии. Ставим курсор в А12 и даем заголовок «Параметры регрессии», а затем переводим курсор в А13 и вызываем функцию «ЛИНЕЙН», аргументы которой: «Изв\_знач\_у» - B1:AF1, «Изв\_знач\_х» - B2:AF4, «Константа» = 1, «Стат» = 1. Далее обводим 5 строк и 4 столбца А13-D17, и нажимаем F2 и Ctrl+Shift+Enter. В первой строке – коэффициенты регрессионного уравнения, следовательно, уравнение регрессии есть . Вторая строка – стандартные ошибки коэффициентов. Все они меньше самих коэффициентов, это значит, что коэффициенты значимы.

Проверим модель на адекватность. Видно, что F-критерий регрессионной модели равен 63,703. Вычислим критическое значение статистики. Вводим в ячейку G13 подпись «F-критическое» а в Н13 вводим функцию FРАСПОБР, имеющую аргументы: «Вероятность» – 0,05 (произвольно примем по умолчанию), «Степени\_свободы\_1» - 3 (число независимых переменных равно трем - ). Аргументом «Степени\_свободы\_2» служит число, показанное в ячейке В16 (в данном примере – 27). Видно, что F-статистика больше ее критического значения, поэтому модель адекватна. Построим уравнение регрессии для каждого города в отдельности. Вводим в G15 подпись «Воронеж», в Н15 подпись «*а*=», в I15 – формулу «=C13» , в Н16 – «b=», в I16 - «=D13+B13». Уравнение для Воронежа есть . Аналогично рассчитываем для Липецка. Вводим в G17 подпись «Липецк», в Н17 подпись «*а*=», в I17 – формулу «=C13» , в Н18 – «b=», в I18 - «=D13+А13». Результат . Для Курска вводим в G19 подпись «Курск», в Н19 подпись «*а*=», в I19 – формулу «=C13» , в Н20 – «b=», в I20 - «=D13». Результат .

# МОДЕЛИРОВАНИЕ ВРЕМЕННОГО РЯДА

 ***Цель:*** *Изучить методы выявления структуры временного ряда, методы моделирования тенденции, и циклической компоненты ряда.*

***ПРИМЕР.*** Имеются данные о ценах на съемное жилье в курортной зоне за восемь сезонов.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Квартал |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Цена | 15 | 5 | 10 | 35 | 26 | 19 | 23 | 46 | 38 | 31 | 34 | 58 | 51 | 41 | 46 | 70 |
| Квартал |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Цена | 63 | 53 | 58 | 82 | 75 | 67 | 70 | 94 | 86 | 77 | 84 | 105 | 98 | 89 | 94 | 117 |

Простроить модель временного ряда.

**1 этап. Выявление структуры ряда**

Построим график зависимости цены на жилье от квартала. Для этого вводим в а1 подпись «квартал», в ячейки а2 и а3 числа «1» и «2», затем обводим мышкой а2 и а3, выделяя их и зацепив мышью за маркер автозаполнения, опускаем его до а33. В в1 вводим подпись «цена», а в столбец в2-в33 вводим значения строки «цена» из таблицы данных. Переходим на «лист 2», нажимаем на кнопку  «Мастер диаграмм» и выбираем категорию «График» и тип «График с маркерами…», второй сверху левый. Нажимаем «далее», в поле «диапазон» обводим на листе 1 ячейки «b2-b33», переходим на закладку «ряд», обводим в поле «подписи оси х» ячейки листа 1 «а2-а33». Нажимаем «готово». Видно, что график имеет явно выраженные линейную трендовую составляющую и циклическую компоненту. Однако, для более полного анализа ряда, построим коррелограмму. Для этого в а1 (лист 2) вводим подпись «лаг», и в а2 и в а3 вводим «1» и «2». Обводим, выделяя, а2-а3 и автозаполнением переводим данные на а2-а9. Результат – последовательность чисел от 1 до 8. Далее ставим курсор в в1, вводим «корреляция», переносим курсор в в2, вызываем формулу «=ПИРСОН» (вычисляет парные коэффициент корреляции, категория «статистические»). Аргументом «массив 1» будет служить ссылка на данные «цена» кроме последнего значения (ссылка на в2-в32, лист 1). Аргумент «массив 2» – эти же данные, но без первого аргумента (ссылка на в3-в33). Далее аналогично находим коэффициенты автокорреляции, но со сдвижкой (лагом) на 2, 3, …, 8 значений. Заполняем ячейки в3-в9 в соответствии с таблицей:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ячейка | Функция | Массив 1 | Массив 2 | Ячейка | Функция | Массив 1 | Массив 2 |
| В3 | ПИРСОН | В2-В31 | В4-В33 | В7 | ПИРСОН | В2-В27 | В8-В33 |
| В4 | ПИРСОН | В2-В30 | В5-В33 | В8 | ПИРСОН | В2-В26 | В9-В33 |
| В5 | ПИРСОН | В2-В29 | В6-В33 | В9 | ПИРСОН | В2-В25 | В10-В33 |
| В6 | ПИРСОН | В2-В28 | В7-В33 |

 Далее обводим ячейки В2-В9 мышью и вызываем мастер диаграмм, выбираем категорию «График» верхний левый график, нажимаем «Готово». Видно, что график имеет всплески при четвертом и восьмом лаге (коэффициент автокорреляции близок к единице), что говорит о наличии циклической составляющей с периодом 4. Проведем теперь моделирование временного ряда, выделив в нем тренд, циклическую и случайные компоненты.

**2 этап. Модель тенденции и циклической компоненты ряда**

Перейдем на лист 1. Все дальнейшие вычисления будут проводиться на этом листе. Проводим в третьем столбце сглаживание данных скользящей средней. Для этого в ячейку С1 вводим подпись «Сглаживание» а в С2 вводим формулу «=(B2+B3+B4+B5)/4» и автозаполняем ячейку на С2-С30. Посчитаем теперь в 4 столбце центрированное скользящее среднее, вводим в D1 подпись «Центрированное», а в D4 вводим формулу «=(C2+C3)/2». Автозаполняем область D4-D31. Вычисляем теперь оценку сезонной компоненты S. Для этого в пятый столбец вводим разность между показателем (столбец 2) и сглаженным значением (столбец 4). Вводим в Е1 подпись «Оценка S», а в Е4 вводим формулу «=B4-D4» и автозаполняем ее на Е4-Е31. Для моделирования циклической компоненты S . Для этого выводим оценку сезонной компоненты по кварталам года. Вводим в А35 подпись «Квартал», в ячейки В35-Е35 числа 1,2,3 и 4, а в ячейки В36-Е43 ссылки в соответствии с таблицей:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ячейка | D36 | E36 | B36 | C36 | D37 | E37 | B37 |
| Ссылка | =E4 | =E5 | =Е6 | =Е7 | =Е8 | =Е9 | =Е10 |
| Ячейка | C37 | D38 | E38 | B38 | C38 | D39 | E39 |
| Ссылка | =Е11 | =Е12 | =Е13 | =Е14 | =Е15 | =Е16 | =Е17 |
| Ячейка | В39 | C39 | D40 | E40 | B40 | C40 | D41 |
| Ссылка | =Е18 | =Е19 | =Е20 | =Е21 | =Е22 | =Е23 | =Е24 |
| Ячейка | E41 | B41 | C41 | D42 | E42 | B42 | C42 |
| Ссылка | =Е25 | =Е26 | =Е27 | =Е28 | =Е29 | =Е30 | =Е31 |

Вводим в А43 подпись «Среднее» и в В43 функцию «СРЗНАЧ», в поле аргумента «Число 1» дать ссылку на В36-В42. Автозаполняем данные на В43-Е43. Вводим в F42 подпись «Сумма», а в F43 формулу «=СУММ(B43:E43)». Видно, что сумма среднего сезонного воздействия отличается от нуля, однако, суммарное воздействие циклической компоненты должно быть нулевым. Для расчета циклической компоненты, рассчитаем ее поправку, которую отнимем от полученных средних данных. Для этого в G42 вводим подпись «Поправка» и в G43 формулу «=F43/4», в А44 вводим «S=», а в В44 вводим «=B43-$G$43», автозаполняем на В44-E44. Получили циклическую компоненту за 4 квартала: 2,759; -8,831; -7,258; 13,33. Вводим эти числа в ячейки F2-F5, введя ссылки: в F2 ссылка на «=В44» после ввода ссылки нажимаем F4, получаем «=$B$44», аналогично, в F3 дается ссылка на «=$С$44», в F4 ссылка на «=$D$44», в F5 ссылка на «=$E$44». Обведем четыре введенные ячейки F2-F5 курсором, и, автозаполнением, скопируем эти четыре ячейки на F2-F33. При этом в ячейку F1 вводим подпись «Циклическое S». Исключим теперь Циклическую компоненту из временного ряда. Для этого вводим в G1 подпись «Т+Е=у-S», а в G2 формулу «=B2-F2» и автозаполняем ячейку на G2-G33. Вычисляем теперь трендовую компоненту (тенденцию временного ряда). Введем в столбец Н трендовую компоненту *Т* в виде линейной функции , для этого в Н1 вводим подпись «Тренд Т», а в Н2 вводим функцию «ТЕНДЕНЦИЯ» категория «Статистические», аргументы которой «Изв\_знач\_у» – ссылка на В2-В33, «Изв\_знач\_х» - ссылка на А2-А33, «Нов\_знач\_х» - вновь ссылка на А2-А33, «Константа» – единица. Далее выделяем обводя курсором ячейки Н2-Н33, и нажимаем F2 и Ctrl+Shift+Enter. В следующем столбце I будет находиться модель временного ряда, состоящая из суммы циклической компоненты *S* и тренда *Т*. Вводим в I1 заголовок «Модель ряда», а в I2 вводим формулу «=H2+F2». Автозаполняем результат на I2-I33. Получим график значений временного ряда, тренда и его модели. Для этого ставим курсор в любую свободную ячейку, вызываем мастер диаграмм и выбираем тип «График» и первый верхний график слева, нажимаем «Далее». В поле «Диапазон» вводим ссылку на В2-В33, обводя их, затем удерживая Ctrl обводим еще области G2-G33 и I2-I33. График синего цвета – временной ряд практически совпадает со своей моделью - желтая линия. В центре линия красного цвета – трендовая составляющая ряда. Получим теперь случайную составляющую временного ряда – остатки *Е*. Для этого в J1 вводим подпись «Остатки Е», а в J2 формулу «=B2-I2». Автозаполняем на J3-J33. По полученным данным можно построить график остатков. Обводим, выделяя, ячейки J2-J33, вызываем мастер диаграмм, выбираем тип «Точечная», верхний график, нажимаем «Готово. График остатков говорит о случайном их расположении. Для проверки качества модели, рассчитаем остаточную сумму квадратов остатков  и остаточную дисперсию (дисперсию адекватности). Для этого в К1 вводим подпись «Е^2», а в К2 вводим формулу «=J2\*J2». Автозаполнением переносим формулу на К2-К33. Вычисляем оценку дисперсии адекватности. Вводим в J35 подпись «Dа=», а в К35 формулу «=СУММ(K2:K33)/32». Результат 1,768711. Вычисляем теперь оценку полную дисперсии показателя. Вводим в J36 подпись «Dy», а в К36 вводим функцию ДИСПР (категория «Статистические»), аргументом которой «Число 1» является ссылка на значения признака B2:B33. Видно, что оценка дисперсия адекватности  намного меньше оценки полной дисперсии , которая равна 860,7937, что говорит о хорошем качестве модели. Оценка парного коэффициента корреляции вычисляется по формуле . Для его получения вводим в J37 подпись «Корреляция», а в К37 формулу «=КОРЕНЬ(1-K35/K36)». Результат близок к единице, что еще раз подтверждает хорошее качество модели.

1. **Моделирование тенденции временного ряда при наличии структурных изменений**

***Цель:*** *научиться определять целесообразность построения кусочно-линейной модели тенденции временного ряда, учитывающие его структурные изменения, с помощью теста Чоу, и строить соответствующие модели.*

 **Пример.** Развивающееся предприятие «Альфа» в течении 13 месяцев своего существования постоянно увеличивало свою прибыль, которая за это время выросла почти вдвое. Однако, на 14-м месяце существования удалось получить дополнительное инвестирование и закупить современное оборудование, после чего темпы роста прибыли заметно увеличились. Имеется временной ряд прибыли предприятия за 25 месяцев. С помощью теста Чоу проверить на уровне значимости предположение о том, какая модель тенденции лучше описывает временной ряд: общая линейная модель тенденции, построенная по всем 25 месяцем ряда, или кусочно-линейная, состоящая из двух линейных моделей, построенных по первым 14 и последующим 11 периодом времени. Методами регрессионного анализа построить эти модели.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Месяц | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| **Прибыль** | **38** | **40** | **41** | **43** | **48** | **49** | **50** | **53** | **55** | **57** | **58** | **61** | **61** |
| Месяц | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| **Прибыль** | **65** | **66** | **68** | **79** | **82** | **88** | **89** | **94** | **96** | **99** | **101** | **105** |

 Решение. Открываем новую книгу EXCEL, вводим в А1 подпись «t=», а в ячейку В2 подпись «Y=», затем в ячейки А2-А26 вводим номера месяцев 1,2,3,…,25, а в ячейки В2-В26 – значения прибыли из приведенной выше таблицы. Для удобства построения кусочно-линейной модели, выделим ячейки А2-D15 в какой-нибудь свет, например в желтый, а ячейки А16-D26 в другой цвет, например в розовый. Проверим по критерию Чоу целесообразность построения кусочно-линейной модели. Для этого с помощью функции ЛИНЕЙН рассчитаем параметры моделей.

 Сначала рассчитаем параметры общей линейной модели. Для этого вводим в F1 подпись «Общая линейная» и ниже, в F2 вводим функцию **=ЛИНЕЙН(B2:B26;A2:A26;1;1).** Выделяем мышью ячейки от F2 до G6 и нажимаем сначала F2 а затем Ctrl+Shift+Enter. Получаем таблицу из 2 столбцов и 5 строк параметров модели. Нас интересуют значения коэффициентов общего линейного уравнения тенденции , которые записаны в первой строке. В результате, общее уравнение линейной тенденции имеет вид: . Кроме того, для критерия Чоу нужно знать суммы квадратов остатков регрессионной модели. Эти значения записаны в нижнем правом углу матрицы, выдаваемой функцией линейн. Для общей модели остаточная сумма равна .

 Находим параметры первой и второй части кусочно-линейной модели. Вводим в ячейку F8 подпись «Кусочно-линейная 1» и в ячейку F9 вводим формулу **=ЛИНЕЙН(B2:B15;A2:A15;1;1).** Выделяем мышью ячейки от F9 до G13 и нажимаем сначала F2 а затем Ctrl+Shift+Enter. Видно, что уравнение регрессии есть , а остаточная сумма . Затем вводим в ячейку F15 подпись «Кусочно-линейная 2» и в ячейку F16 вводим формулу **=ЛИНЕЙН(B16:B26;A16:A26;1;1).** Выделяем мышью ячейки от F16 до G20 и нажимаем сначала F2 а затем Ctrl+Shift+Enter. Видно, что уравнение регрессии второй части кусочно-линейной модели есть , а остаточная сумма . Статистика критерия Чоу для парной регрессионной модели вычисляется по формуле: , где *п* – число уровней ряда (в данном случае – число месяцев, равно 25). Вводим в ячейку I1 подпись «Статистика», а в G1 – формулу **=(G6-G13-G20)/(G13+G20)\*21/2**. Критическое значение равно значению обратного распределения Фишера, полученного по параметрам: - уровень значимости, указан в условии задачи; *k*=2 – степени свободы 1, равные числу параметров модели (у нас из 2: *a* и *b*, т.к. уравнение регрессии ); = 21 – степени свободы 2, равные 21. Вводим в I2 подпись «Критическое», а в G2 формулу **=FРАСПОБР(0,05;2;21).** Видно, что статистика больше критического значения, что говорит о том, что кусочно-линейная функция лучше описывает временной ряд, чем общая модель.

 Строим кусочно-линейную модель. Вводим в С1 подпись «Линейная», а в С2 вводим функцию **=ТЕНДЕНЦИЯ(B2:B26; A2:A26; A2:A26; 1),** выделяем диапазон С2-С26 и нажимаем F2 а затем Ctrl+Shift+Enter. Вводим в ячейку D1 подпись «Кусочно-линейная», а в ячейку D2 вводим формулу **=ТЕНДЕНЦИЯ(B2:B15;A2:A15;A2:A15;1)**, выделяем диапазон D2-D15 и нажимаем F2 а затем Ctrl+Shift+Enter. Затем, для построения второй ветви линейного уравнения вводим в ячейку D16 вводим формулу **=ТЕНДЕНЦИЯ(B16:B26;A16:A26;A16:A26;1)**, выделяем диапазон D16-D26 и нажимаем F2 а затем Ctrl+Shift+Enter. Построим график по полученным данным. Ставим курсор в свободную ячейку, вызываем мастер функций, выбираем тип «График», вид график без точек в верхнем левом углу, нажимаем «Далее», переводим курсор в поле «Диапазон» и обводим ячейки В2-D26. переходим на закладку «Ряд», щелкаем мышкой по надписи «Ряд 1» в поле «Ряд» и переводим курсор в поле «Имя» и вводим в нем текст «Данные», затем щелкаем мышкой по надписи «Ряд 2» в поле «Ряд» и переводим курсор в поле «Имя» и вводим в нем текст «Линейная», после чего щелкаем мышкой по надписи «Ряд 3» и в поле «Имя» и вводим текст «Кусочно-линейная», нажимаем «Готово».

 Рассмотрим другой метод построения модели с переменной структурой. Для этого воспользуемся фиктивной переменной. Пусть *Z* – фиктивная переменная, которая принимает значения:



 Тогда общая регрессионная модель примет вид: . Для определения параметров модели , сформируем исходные данные в следующем виде. Переходим на **Лист 2**. В ячейки A1, B1, C1, D1 вводим подписи «Y», «t», «Z», «Zt» (кавычки не вводить).

 В первый столбец копируем значения уровней временного ряда. Для этого переходим на **Лист 1**, выделяем ячейки В2-В26, выполняем ПРАВКА/КОПИРОВАТЬ. Затее переходим обратно на **Лист 2**, ставим курсор в А2 и даем команду ПРАВКА/ВСТАВИТЬ.

 Во второй столбец **Листа 2** (ячейки В2-В26) копируем ячейки А2-А26 из **Листа 1**. В столбец С **Листа 2** вводим значения переменной Z. В ячейки С2-С15 вводим число 0. В ячейки С16-С26 вводим число 1. В столбец D вводим произведение переменных . Ставим курсор в D2 и вводим формулу **=B2\*C2**. Автозаполняем формулу на D2-D26. Строим линейную регрессионную модель. Для этого в Е2 вводим формулу (категория «Статистические»): **=ТЕНДЕНЦИЯ(A2:A26;B2:D26;B2:D26;1)**

и обводим диапазон Е2-Е26, нажимаем клавишу F2, затем одновременно Ctrl+Shift+Enter. В результате получаем модель линейной регрессии. Вычислим ее числовые характеристики. Для этого в G2 вводим функцию:

**=ЛИНЕЙН(A2:A26;B2:D26;1;1)**

выделяя, обводим ячейки G2-J6, нажимаем F2, затем Ctrl+Shift+Enter. Проверяем адекватность полеченной модели. Видно, что коэффициент детерминации равен 0,99 (ячейка G4), что говорит об очень высокой качестве регрессии.

 Строим график регрессионной модели. Ставим курсор в свободную ячейку, вызываем мастер функций, выбираем тип «График», вид график без точек в верхнем левом углу, нажимаем «Далее», переводим курсор в поле «Диапазон» и обводим ячейки A2-A26, затем, удерживая Ctrl, обводим еще и диапазон Е2-Е26. Переходим на закладку «Ряд», щелкаем мышкой по надписи «Ряд 1» в поле «Ряд» и переводим курсор в поле «Имя» и вводим в нем текст «Данные», затем щелкаем мышкой по надписи «Ряд 2» в поле «Ряд» и переводим курсор в поле «Имя» и вводим в нем текст «Модель», нажимаем «Готово».