МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ

Лабораторная работа № 2

по дисциплине

«Дискретный анализ»

по теме

«Линейный регресионный анализ»

вариант № 3

Выполнила:

ст.гр ОИ-071

Бойко Маргарита

Одеса 2010

Формулировка задачи

Результаты наблюдений функционирования отрасли (Y –валовый выпуск, Х1-фондовооруженность, Х2-производительность труда) приведены в таблице. Предполагая, что валовый выпуск зависит от фондовооруженности и производительности труда построить линейную регрессионную модель и сделать анализ модели.

|  |  |  |
| --- | --- | --- |
| У | X1 | X2 |
| 1133 | 4,61 | 8,1 |
| 1160 | 4,56 | 6,9 |
| 1402 | 4,67 | 8,7 |
| 1524 | 4,94 | 9,2 |
| 1595 | 5,08 | 9,2 |
| 1556 | 5,41 | 9,5 |
| 1679 | 5,59 | 9,8 |
| 1843 | 5,74 | 10,2 |

Теоретические основания к поставленной задаче

Рассмотрим случай, когда число переменных больше 2. В этом случае линейная модель имеет вид

, (1)



где  - вектор ошибок наблюдений.

Введем следующие матричные обозначения :  
где -вектор-столбец выборочных значений результирующего признака;

-матрица значений переменных , включая единичный столбец, отвечающий свободному члену;

- вектор –столбец всех параметров регрессии;

-вектор-столбец выборочных реализаций случайной составляющей,

M() = 0, cov

 -независимы и имеют нормальное распределение с параметрами (0,).

В матричном виде модель записывается так: 

Y = X + , (2)

Вектор параметров регрессии находят при условии минимизации ее ошибки  по формуле:

, (3)

штрих здесь и далее означает транспонирование .

Остаточная сумма квадратов *Qe* вычисляется по формуле

*Qe* =  (4)

 (5)

Проверка гипотезы Н0:  позволяет установить, находятся ли переменные  во взаимосвязи с Y. Статистикой критерия для проверки гипотезы Н0 является соотношение

z=, (6)

если выборочное значение этой статистики > , то гипотеза

Н0 отклоняется; в противном случае следует считать , что взаимосвязи Y с переменными  нет.

Границы доверительных интервалов для параметров *β*j  определяются по формуле

 , *j*=1,2,…,*m* (7)

где - диагональный элемент матрицы .

При использовании модели (1) для представления данных необходимо решить вопрос целесообразности включения переменных  в модель. для этого проверяются гипотезы  Эти гипотезы могут быть проверены непосредственно по доверительным интервалам (7). Если доверительный интервал для  накрывает нуль, то гипотеза

принимается и соответствующую переменную  не целесообразно включать в модель, в противном случае отклоняется.

Коэффициент множественной корреляции определяется по формуле

*R* = . (8)

Таблица 2.

Дополнительная регрессионная статистика.

|  |  |
| --- | --- |
| **Величина** | **Описание** |
| se1,se2,...,sen | Стандартные значения ошибок для коэффициентов m1,m2,...,mn. |
| seb | Стандартное значение ошибки для постоянной b (seb = #Н/Д, если конст имеет значение ЛОЖЬ). |
| r2 | Коэффициент детерминации. Сравниваются фактические значения y и значения, получаемые из уравнения прямой; по результатам сравнения вычисляется коэффициент детерминации, 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 детерминации равен 0, то уравнение регрессии неудачно для предсказания значений y. |
| sey | Стандартная ошибка для оценки y. |
| 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. |
| d | Степени 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. |
| SSрег | Регрессионная сумма квадратов. |
| SSост | Остаточная сумма квадратов. |

Данные при использовании Excel выводятся в следующую таблицу :

Таблица 3.

Результирующая таблица.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | А | В | С | D | E | F |
| 1 |  |  | … |  |  | *b* |
| 2 |  |  | … |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 | *F* |  |  |  |  |  |
| 5 |  |  |  |  |  |  |

Компьютерная реализация однофакторного дисперсионного анализа

|  |  |  |  |
| --- | --- | --- | --- |
|  | А | B | C |
| 1 | У | X1 | X2 |
| 2 | 1133 | 4,61 | 8,1 |
| 3 | 1160 | 4,56 | 6,9 |
| 4 | 1402 | 4,67 | 8,7 |
| 5 | 1524 | 4,94 | 9,2 |
| 6 | 1595 | 5,08 | 9,2 |
| 7 | 1556 | 5,41 | 9,5 |
| 8 | 1679 | 5,59 | 9,8 |
| 9 | 1843 | 5,74 | 10,2 |

Рисунок 1. Исходные данные

|  |  |  |  |
| --- | --- | --- | --- |
|  | А | B | C |
| 10 | 131,1707 | 223,6413 | -822,457 |
| 11 | 67,57019 | 153,947 | 381,9195 |
| 12 | 0,905532 | 89,17151 | #Н/Д |
| 13 | 23,96412 | 5 | #Н/Д |
| 14 | 381104,2 | 39757,79 | #Н/Д |

Рисунок 2.Результирующие данные

= 131,1707

 = 67,57019

 = 0,905532

*F =* 23,96412

= 381104,2

 = 223,6413

 = 153,947

 = 89,17151

= 5

 = 39757,79

= --822,457

=381,9195

**Использование статистик F и r2**

Уравнение множественной регрессии *y = m1\*x1 + m2\*x2 + m3\*x3 + m4\*x4 + b* теперь может быть получено из строки 10:

Коэффициент детерминации r2 равен 0,905532 (, что указывает на сильную зависимость между независимыми переменными и продажной ценой. Можно использовать F-статистику, чтобы определить, является ли этот результат (с таким высоким значение r2 ) случайным.

Предположим, что на самом деле нет взаимосвязи между переменными, просто статистический анализ вывел сильную взаимозависимость по взятой равномерной выборке 8 отраслей.

Если F-наблюдаемое больше, чем F-критическое, то взаимосвязь между переменными имеется. F-критическое можно получить из таблицы F-критических значений в любом справочнике по математической статистике. Для того, чтобы найти это значение, используя односторонний тест, положим уровень значимости 0,05, а для числа степеней свободы (обозначаемых обычно v1 и v2), положим

v1 = k = 2 и v2 = n - (k + 1) = 8 - (2 + 1) = 7

, где k - это число переменных, а n - число точек данных.

Из таблицы справочника F-критическое равно 4,74.

F-наблюдаемое равно 23,96412 (ячейка A18), что заметно больше чем F-критическое (4,74). Следовательно, полученное регрессионное уравнение полезно для предсказания валовой продукции

**Вычисление t-статистики**

Определим, имеет ли каждый коэффициент статистическую значимость для оценки валового выпуска: t-критическое с 7 степенями свободы и = 0,05 равно 1,895

tтабл0,95(7)=1,895

Полученные наблюдаемые t-значения для каждой из независимых переменных приводим в таблице :

Таблица 4.

Вычисление t-статистики

|  |  |
| --- | --- |
| **Переменная** | **t-наблюдаемое значение** |
| фондоворуженность | 1,452715624 |
| производительность | -2,153483551 |

t = m4 / se4 = 223,6413/153,947= 1,452715624

Поскольку абсолютная величина t, равная 1,452715624 , меньше, чем 1,895, фондовооруженность — это не важная переменная для валового выпуска продукции. Аналогичным образом можно протестировать все другие переменные на статистическую значимость. Абсолютная величина для производительности = 2,153483551 , это больше, чем 1,895 значит производительность полезная длопределения валового выпуска продукции.

Литература.

* 1. Попов А.А.

Excel: Практическое руководство, ДЕСС КОМ.-М.-2000.

1. Дьяконов В.П., Абраменкова И.В. Mathcad7 в математике, физике и в Internet. Изд-во « Номидж», М.-1998, раздел 2.13. Выполнение регрессии.
2. Л.А. Сошникова, В.Н. Томашевич и др. Многомерный статистический анализ в экономике под ред. В.Н. Томашевича.- М. –Наука, 1980.
3. Колемаев В.А., О.В. Староверов, В.Б. Турундаевский Теория вероятностей и математическая статистика. –М. – Высшая школа- 1991.
4. К Иберла. Факторный анализ.-М. Статистика.-1980.