

Нахождение уравнения множественной регрессии и коэффициента корреляции на примере решения задачи в системе POMWIN

Пронина Ольга Юрьевна

*Приамурский государственный университет им.Шолом-Алейхема
студент*

Баженов Руслан Иванович

*Приамурский государственный университет им.Шолом-Алейхема
к.п.н., доцент, зав.кафедрой информационных систем, математики и
методик обучения*

Аннотация

В статье рассматривается решение задачи на нахождение уравнения множественной регрессии и коэффициента корреляции в программной системе POMWIN, а также нахождение значения индекса детерминации. Приведены иллюстрации работы в программе.

Ключевые слова: уравнение множественной регрессии, коэффициент корреляции, индекс детерминации, POMWIN.

Finding the multiple regression equation and the correlation coefficient for the example of solving a problem in the POMWIN system

Pronina Olga Yurievna

*Sholom-Aleichem Priamursky State University
student*

Bazhenov Ruslan Ivanovich

*Sholom-Aleichem Priamursky State University
Candidate of pedagogical sciences, associate professor, Head of the Department
of Information Systems, Mathematics and teaching methods*

Abstract

The article deals with the solution of the problem of finding the multiple regression equation and the correlation coefficient in the POMWIN software system, as well as finding the value of the determination index. The illustrations of the work in the program are given.

Keywords: multiple regression equation, correlation coefficient, determination index, POMWIN.

Программа POMWIN содержит в себя огромное количество возможностей, в том числе и метод нахождения коэффициента корреляции, индекса детерминации и уравнения регрессии.

Исследование о реализации объектов совместным множественным уравнением регрессии провел Ю.П.Ванин [1]. О нахождении уравнения множественной регрессии для фракционных концентраций атмосферного аэрозоля и спектральных значений коэффициента ослабления написали М.М.Кугейко и С.А.Лысенко [2]. О спецификации модели (уравнения множественной регрессии) провела исследование З.В.Шилова [3]. Анализ стоимости квартир на рынке недвижимости г.Набережные Челны на основе уравнения множественной регрессии провели Р.В.Спеваков и Э.М.Ибрагимова [4]. Методы множественной регрессии в эконометрике описал Л.И.Ивашнёв [5]. Уравнение множественной регрессии для прогнозной оценки стоимости квартир выявили Н.С.Булгакова и М.Саадаева [6]. Исследования о нахождении уравнений множественной регрессии, коэффициенте корреляции и индекса детерминации проводят также и зарубежные ученые [7-10].

Объектом исследования является решение задачи на нахождении уравнения регрессии и коэффициента корреляции в программной системе POMWIN.

Имеются следующие показатели по десяти предприятиям некоторой отрасли [11] (рис.1).

| Номер предприятия | Стоимость промышленно-производственных основных фондов, тыс. руб. | Валовая продукция в оптовых ценах предприятия, тыс. руб. | Среднесписочная численность промышленно-производственного персонала, чел. | Среднесписочная численность рабочих, чел. |
|-------------------|-------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------------------------------------|-------------------------------------------|
| 1 | 4999 | 5349 | 420 | 331 |
| 2 | 6929 | 6882 | 553 | 486 |
| 3 | 6902 | 7046 | 570 | 498 |
| 4 | 10097 | 7248 | 883 | 789 |
| 5 | 8097 | 5256 | 433 | 359 |
| 6 | 11116 | 14090 | 839 | 724 |
| 7 | 4880 | 3525 | 933 | 821 |
| 8 | 7355 | 5431 | 526 | 428 |
| 9 | 10066 | 7680 | 676 | 607 |
| 10 | 7884 | 8226 | 684 | 619 |

Рисунок 1- Данные задачи

Приняв стоимость основных промышленно – производственных основных фондов за результативный признак, а остальные показатели – за факторные признаки, необходимо:

1. вычислить множественный коэффициент корреляции и сделать выводы о форме и силе корреляционной зависимости;
2. вычислить значение общего индекса детерминации;
3. найти уравнение линейной модели множественной регрессии.

Для начала, необходимо запустить программу POMWIN, выбрать модуль Forecasting (прогнозирование) и создать новый проект наименьших квадратов простой и множественной регрессии (Least Squares Simple and Multiple Regression).

Появившееся окно необходимо заполнить данными и нажать ОК (рис.2).

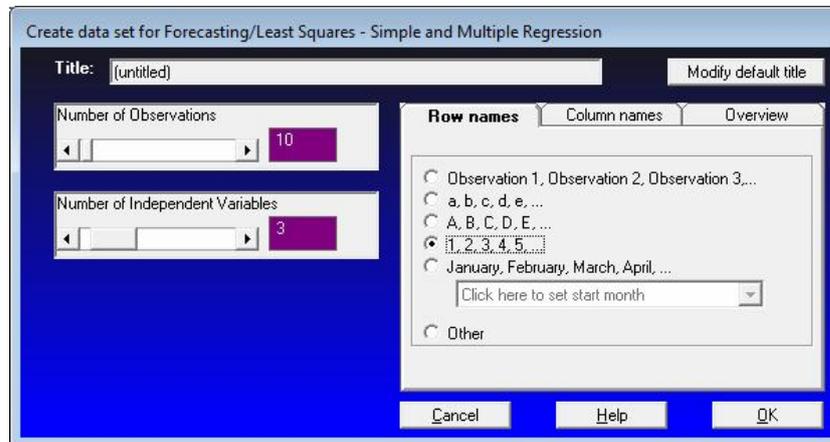


Рисунок 2- Создание нового проекта

Далее заполняем таблицу исходными данными (рис.3).

| | Dpndnt var, Y | X1 | X2 | X3 |
|----|---------------|-------|-----|-----|
| 1 | 4999 | 5349 | 420 | 331 |
| 2 | 6929 | 6882 | 553 | 486 |
| 3 | 6902 | 7046 | 570 | 498 |
| 4 | 10097 | 7248 | 883 | 789 |
| 5 | 8097 | 5256 | 433 | 359 |
| 6 | 11116 | 14090 | 839 | 724 |
| 7 | 4880 | 3525 | 933 | 821 |
| 8 | 7355 | 5431 | 526 | 428 |
| 9 | 10066 | 7680 | 676 | 607 |
| 10 | 7884 | 8226 | 684 | 619 |

Рисунок 3- Заполнение таблицы исходными данными

Далее нажимаем кнопку Solve, тем самым запускаем процесс расчетов.

Появляется таблица результатов прогнозирования (Forecasting Results) (рис.4).

| Measure | Value |
|------------------------------------|-----------|
| Error Measures | |
| Bias (Mean Error) | 0 |
| MAD (Mean Absolute Deviation) | 1157,185 |
| MSE (Mean Squared Error) | 1485358,0 |
| Standard Error (denom=n-2-2=6) | 1573,403 |
| MAPE (Mean Absolute Percent Error) | ,159 |
| Regression line | |
| Dpndnt var, Y = 4380,063 | |
| + ,533 * X1 | |
| -22,273 * X2 | |
| + 25,075 * X3 | |
| Statistics | |
| Correlation coefficient | ,791 |
| Coefficient of determination (r^2) | ,625 |

Рисунок 4- Рассчитанные данные

По полученным данным видно, что множественный коэффициент корреляции (Correlation coefficient) равен 0,791. Данное значение обозначает высокую степень тесноты и линейности корреляционной зависимости

Коэффициент детерминации (Coefficient determination) равен 0,625. Данный коэффициент означает, что факторные признаки задачи влияют на результативные в пределах 62,5%.

Следующим этапом рассмотрим таблицу Details and Error Analysis (рис.5).

| (untitled) Solution | | | | | | | | | |
|---------------------|---------------|--------|---------|--------|----------|-----------|----------|-----------|-----------|
| | Dpndnt var, Y | X1 | X2 | X3 | Forecast | Error | Error | Error^2 | Pct Error |
| 1 | 4999 | 5349 | 420 | 331 | 6176,458 | -1177,458 | 1177,458 | 1386406,0 | 23,554% |
| 2 | 6929 | 6882 | 553 | 486 | 7917,897 | -988,897 | 988,897 | 977917,3 | 14,272% |
| 3 | 6902 | 7046 | 570 | 498 | 7927,573 | -1025,573 | 1025,573 | 1051801,0 | 14,859% |
| 4 | 10097 | 7248 | 883 | 789 | 8360,476 | 1736,524 | 1736,524 | 3015517 | 17,198% |
| 5 | 8097 | 5256 | 433 | 359 | 6539,419 | 1557,581 | 1557,581 | 2426057,0 | 19,237% |
| 6 | 11116 | 14090 | 839 | 724 | 11357,93 | -241,935 | 241,935 | 58532,34 | 2,176% |
| 7 | 4880 | 3525 | 933 | 821 | 6064,575 | -1184,575 | 1184,575 | 1403217,0 | 24,274% |
| 8 | 7355 | 5431 | 526 | 428 | 6291,455 | 1063,545 | 1063,545 | 1131128,0 | 14,46% |
| 9 | 10066 | 7680 | 676 | 607 | 8637,723 | 1428,277 | 1428,277 | 2039976,0 | 14,189% |
| 10 | 7884 | 8226 | 684 | 619 | 9051,49 | -1167,49 | 1167,49 | 1363034,0 | 14,808% |
| TOTALS | 78325 | 70733 | 6517 | 5662 | | 0 | 11571,85 | 14853580 | 1,59 |
| AVERAGE | 7832,5 | 7073,3 | 651,7 | 566,2 | | 0 | 1157,185 | 1485358,0 | ,159 |
| Betas | 4380,063 | ,533 | -22,273 | 25,075 | | (Bias) | (MAD) | (MSE) | (MAPE) |
| | | | | | | | Std err | 1573,403 | |

Рисунок 5- Таблица Details and Error Analysis

По данной таблице можно выявить уравнение множественной регрессии. Для этого нужно обратить внимание на строку Betas. Уравнение множественной регрессии имеет вид: $Y=4380,063+0,533*X1-22,273*X2+25,075*X3$. А также данное уравнение множественной регрессии можно увидеть в таблице Forecasting Results в поле Regression line.

В ходе решения задачи, были рассмотрены варианты определения коэффициента корреляции, коэффициента детерминации и уравнения множественной регрессии с помощью программной системы POMWIN.

Библиографический список

1. Ванин Ю.П. Реализация объектов совместным множественным уравнением регрессии // Наука и знание: инновационные подходы к решению практических профессиональных задач по формированию конкурентоспособных специалистов Материалы XVI научно-практической конференции Новороссийского филиала Московского гуманитарно-экономического института. Москва: 2014. С. 232-236.
2. Кугейко М.М., Лысенко С.А. Уравнения множественной регрессии для фракционных концентраций атмосферного аэрозоля и спектральных значений коэффициента ослабления // Журнал прикладной спектроскопии. 2006. №6. С. 807-812.
3. Шилова З.В. О спецификации модели (уравнения множественной регрессии) // Математический вестник педвузов и университетов Волго-Вятского региона. 2016. №18. С. 220-227.
4. Спеваков Р.В., Ибрагимова Э.М. Анализ стоимости квартир на рынке недвижимости г. набережные челны на основе уравнения множественной регрессии // Тенденции и закономерности развития современного российского общества: экономика, политика, социально-культурная и правовая сферы материалы Всероссийской научно-практической конференции с международным участием: в 2-х частях. 2016. С. 38-39.
5. Ивашнёв Л.И. Методы множественной регрессии в эконометрике // Проблемы современного социально-экономического развития национальной экономики. Политехнический институт имени В.С. Черномырдина Московского государственного машиностроительного университета . Маши: 2014. С. 194-200.
6. Булгакова Н.С., Саадаева М. Уравнение множественной регрессии для прогнозной оценки стоимости квартир // Вестник Махачкалинского филиала МАДИ. 2014. №14. С. 113-116.
7. Averkin A.N., Ярушев С.А. Review of studies on time series forecasting based on hybrid methods, neural networks and multiple regression // Программные продукты и системы. 2016. №1. С. 75-82.
8. Huberty C.J. Multiple correlation versus multiple regression // Educational and Psychological Measurement. 2003. №2. С. 271-278.
9. Li W., Shinde T., Fujimura H., Miyajima Ch., Nishino T., Itou K., Takeda K., Itakura F. Multiple regression of log spectra for in-car speech recognition using multiple distributed microphones // IEICE Transactions on Information and Systems. 2005. №3. С. 384.
10. Kreuger J., Tornqvist L. Multiple regression analysis of pesticide occurrence in streamflow related to pesticide properties and quantities applied // Chemosphere. 1998. №2. С. 189-207.

11. Решения задач – Множественная регрессия // goodstudents.ru URL:
<http://www.goodstudents.ru/statistika-zadachi/1308-regressiya.html> (дата
обращения: 27.04.2017).