ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

ЦЕНТР ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Кафедра «Управление качеством»

ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ

Методические указания к практическим занятиям
по дисциплине "Системный анализ"

Авторы:
Димитров Валерий Петрович
Борисова Людмила Викторовна
Зубрилина Елена Михайловна

Ростов-на-Дону, 2014

Аннотация

Методические указания предназначены для бакалавров по направлению подготовки 221700 Стандартизация и метрология.
Печатается по решению методической комиссии факультета «Приборостроение и техническое регулирование».

Авторы

Заведующий кафедрой «Управление качеством» ДГТУ д.т.н., профессор Димитров Валерий Петрович;
Профессор кафедры «Управление качеством» д.т.н. Борисова Людмила Викторовна;
Доцент кафедры «Управление качеством» ДГТУ, к.т.н, Зубрилина Елена Михайловна.

Оглавление

ВВЕДЕНИЕ

К задачам регрессионного анализа относятся:

1) установление формы зависимости; 2) определение уравнения регрессии т.е. определение неизвестных коэффициентов модели; 3) оценка неизвестных значений зависимой переменной.

Различают линейную и нелинейную связи.

Линейная связь имеет место, когда с возрастанием (или убыванием) значений X значения Y увеличиваются (или уменьшаются) более или менее равномерно.

Математически линейная связь может быть выражена уравнением прямой, которое называется линейным *уравнением регрессии:*

$$Y_{теор} = b_0 + b_1 \cdot X , \qquad (1)$$

где X- факторный признак; $Y_{теор}$ – результативный признак; b_0, b_1- коэффициенты уравнения.

Если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, степенной, показательной, экспоненциальной и т. д.), то такую связь называют **нелинейной.** На практике часто пользуются *уравнением параболы второго порядка:*

$$Y_{теор} = b_0 + b_1 \cdot X + b_2 \cdot X^2 \qquad (2)$$

Уравнение нелинейной связи может быть выражено и в виде *уравнения гиперболы:*

$$Y_{теор} = b_0 + b_1/X$$

или показательной функции:

$$Y_{теор} = b_0 \cdot b_1^X$$

После определения формы связи, т.е. вида уравнения регрессии, по эмпирическим данным определяют коэффициенты искомого уравнения.

1 РАСЧЕТ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ РЕГРЕССИИ

Коэффициенты b_1 и b_0 уравнения (1) определяются по формулам:

$$b_1 = \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i - n \cdot \sum_{i=1}^{n} x_i \cdot y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \cdot \sum_{i=1}^{n} x_i^{\,2}},$$

$$b_0 = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} y_i - b_1 \cdot \sum_{i=1}^{n} x_i\right), \text{ или } b_0 = \overline{y} - b_1 \cdot \overline{x}.$$

Для экспоненциальной (степенной) зависимости

$$y = b_0 \cdot e^{b_1 \cdot x}$$

коэффициенты b_1 и b_0 определяются по формулам

$$b_1 = \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} \ln y_i - n \cdot \sum_{i=1}^{n} x_i \cdot \ln y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \cdot \sum_{i=1}^{N} x_i^{\,2}},$$

$$b_0 = \exp\left[\frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln y_i - b_1 \cdot \sum_{i=1}^{n} x_i\right)\right] \quad (14)$$

Для параболическая зависимости

$$y = b_0 + b_1 x + b_2 x^2 .$$

Коэффициенты b_0, b_1, b_2 определяются при решении системы из трех уравнений (например, методом Гаусса):

$$\begin{cases} b_0 \cdot n + b_1 \cdot \sum_{i=1}^{n} x_i + b_2 \cdot \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} y_i; \\ b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \cdot \sum_{i=1}^{n} x_i^{\,2} + b_2 \cdot \sum_{i=1}^{n} x_i^{\,3} = \sum_{i=1}^{n} x_i \cdot y_i; \\ b_0 \cdot \sum_{i=1}^{n} x_i^{\,2} + b_1 \cdot \sum_{i=1}^{n} x_i^{\,3} + b_2 \cdot \sum_{i=1}^{n} x_i^{\,4} = \sum_{i=1}^{n} x_i^{\,2} \cdot y_i \end{cases}$$

$$D=\begin{vmatrix} N & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{vmatrix},$$

$$D_1=\begin{vmatrix} \sum y & \sum x & \sum x^2 \\ \sum xy & \sum x^2 & \sum x^3 \\ \sum x^2 y & \sum x^3 & \sum x^4 \end{vmatrix}$$

$$D_2=\begin{vmatrix} N & \sum y & \sum x^2 \\ \sum x & \sum xy & \sum x^3 \\ \sum x^2 & \sum x^2 y & \sum x^4 \end{vmatrix},$$

$$D_3=\begin{vmatrix} N & \sum x & \sum y \\ \sum x & \sum x^2 & \sum xy \\ \sum x^2 & \sum x^3 & \sum x^2 y \end{vmatrix}$$

$$D_1=\sum y\sum x^2\sum x^4-\sum y\left(\sum x^3\right)^2-\sum x\sum xy\sum x^4+$$

;

$$\sum x\sum x^2 y\sum x^3+\sum x^2\sum xy\sum x^3-\left(\sum x^2\right)^2\sum x^2 y$$

$$D_2=N\sum xy\sum x^4-N\sum x^2 y\sum x^3-\sum x\sum y\sum x^4+$$

;

$$\sum y\sum x^2\sum x^3+\sum x\sum x^2\sum x^2 y-\left(\sum x^2\right)^2\sum xy$$

$$D_3=N\sum x^2\sum x^2 y-N\sum x^3\sum xy-\left(\sum x\right)^2\sum x^2 y+$$

;

$$\sum x\sum x^2\sum x\ y+\sum x\sum y\sum x^3-\left(\sum x^2\right)^2\sum y$$

$$D=N\sum x^2\sum x^4-N\left(\sum x^3\right)^2-\left(\sum x\right)^2\sum x^4-\left(\sum x^2\right)^3+2\sum x\sum x^2\sum x^3$$

.

$$b_0 = \frac{D_1}{D}; \qquad b_1 = \frac{D_2}{D}; \qquad b_2 = \frac{D_3}{D}$$

Для функции y = f(x), имеющей вид:

$$y = b_0 + b_1/X$$

система уравнений для определения коэффициентов уравнения регрессии имеет вид:

$$\begin{cases} nb_0 + b_1 \cdot \sum \dfrac{1}{x_i} = \sum y_i \\ b_0 \sum \dfrac{1}{x_i} + b_1 \sum \dfrac{1}{x^2} = \sum \dfrac{x_i}{y_i} \end{cases}$$

Для функции y = f(x), имеющей вид:

$$y = b_0 \cdot b_1^x$$

система уравнений для определения коэффициентов уравнения имеет вид:

$$\begin{Bmatrix} \sum \lg y = n \lg b_0 + \lg b_1 \sum x \\ \sum x \lg y = \lg b_0 \sum x + \lg b_1 \sum x^2 \end{Bmatrix}.$$

2 ПРИМЕР РАСЧЕТА КОЭФФИЦИЕНТОВ УРАВНЕНИЯ РЕГРЕССИИ

Имеются статистические данные о зависимости рентабельности производства продукции (%) по ряду предприятий, производящих одноименную продукцию, от выработки (в стоимостных показателях) на одного среднесписочного работника производственно-промышленного персонала. Полученные данные представлены в табл. 1:

Таблица 1 – Статистические данные по предприятиям

Номер	X	Y
1	907	11,20
2	926	11,05
3	506	6,84
4	741	9,21
5	789	9,42
6	889	10,08
7	874	9,45

8	510	6,73
9	529	7.24
10	420	6,12
11	679	7,63
12	872	9,43
13	924	9,46
14	607	7,64
15	452	6,92
16	729	8,95
17	794	9,33
18	844	10,23
19	1010	11,77
Итого	14623	176,11

Для прогноза результирующего признака Y применим модель парной регрессии, в которой используется только одна факторная переменная — *X*.

Анализ табличных данных показывает наличие прямой линейной зависимости между факторным X (выработки продукции) и результативным признаком Y (рентабельностью производства). Тесноту и направление связи между факторным и результативным признаками определим с помощью коэффициентом корреляции r.

$$r = \frac{n \cdot \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sqrt{n \cdot \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2}}$$

где X_i и Y_i - значения факторного и результативного признаков соответственно; n – объем выборки (число пар исходных данных).

Для рассматриваемого примера значение коэффициента корреляции составляет:

$$r = \frac{20 \times 134127{,}9 - 14623 \times 176{,}11}{\sqrt{20 \times 11306109 - 14623^2} \times \sqrt{20 \times 1602{,}0971 - 176{,}11^2}} =$$

0,955

Значение коэффициента корреляции показывает на наличие довольно сильной связи.

Рассчитаем параметры уравнения регрессии: $Y_{теор} = b_0 + b_1 \cdot X$.

Подставив данные из табл. 1 в расчетные формулы получим значения коэффициентов: $b_0 = 2{,}423$; $b_1 = 0{,}00873$.

Таким образом, линейное уравнение регрессии имеет вид:

$$Y_{теор} = 2{,}423 + 0{,}00873 \cdot X .$$

Коэффициент b_1 характеризует наклон линии регрессии. Значение $b_1 = 0{,}00873$ и это означает, что при увеличении *X* на единицу ожидаемое значение Y возрастет на 0,00873. Отсюда b_1 может быть интерпретирован как прирост нормы рентабельности, который варьируется в зависимости от средней выручки. Свободный член уравнения $b_0 = 2{,}423$ у.е.. Это значение Y при *X,* равном нулю. Поскольку маловероятно значение выработки, равное нулю, то можно интерпретировать b_0 как меру влияния на величину рентабельности других факторов, не включенных в данное уравнение регрессии.

Регрессионная модель может быть использована для прогноза уровня рентабельности, (который будет на предприятии, например, где средняя выработка на одного работника составит 600 руб.)

Для того, чтобы определить прогнозируемое значение, следует X = 600 подставить в уравнение:

$$Y = 2{,}423 + 0{,}00873 \bullet 600 = 7{,}661.$$

Расчеты показали, что прогнозируемый уровень рентабельности для предприятия со средней выработкой 600 рублей на одного рабочего ППП составит 7,661 %.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Введение в системный анализ: учеб.пособие / В.П. Димитров, Л.В. Борисова, Б.Б. Жмайлов – 2-е изд., испр. и доп. – Ростов н/Д: Издательский центр ДГТУ, 2013. – 77 с.
2. Системный анализ и принятие решений: учебное пособие / С.А. Баркалов, И.С. Суровцев, А.И. Половинкина ; науч.ред. В.Н. Бурков. – Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2010. – 652 с.

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Задачи регрессионного анализа.
2. Что называется линейным уравнением регрессии?
3. Записать уравнение нелинейной связи.
4. Что называется коэффициентом регрессии?
5. Записать уравнение для определения коэффициентов b_0, b_1, $b_{2.}$
6. Уравнение для определения коэффициента корреляции.