УДК 330.43

Математическая модель для расчета плановых показателей приема в ВУЗ

© **Е. А. Черноиванова**[1]

Аннотация. В статье предложен метод расчета плановых цифр приема в ВУЗ на основе классического метода наименьших квадратов.

Ключевые слова: уравнение регрессии, метод наименьших квадратов для множественной регрессии, коэффициент детерминации

На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, план приема в ВУЗ связан с количеством выпускников в данном регионе, с уровнем доходов населения, востребованностью выпускников, с количеством выпускников СПО, НПО, с лицензионными показателями по формированию плана приема. В этом случае вместо парной регрессии $M(y/x) = f(x)$ рассматривается множественная регрессия

$$M(y/x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, x_m).$$

Задача оценки статистической взаимосвязи переменных Y и $X_1, X_2, \ldots, X_m$ формируется аналогично случаю парной регрессии. Уравнение множественной регрессии может быть представлено в виде

$$Y = f(\beta, X) + \varepsilon,$$

где $X = (X_1, X_2, \ldots, X_m)$ – вектор независимых переменных; β – вектор параметров; ε – случайная ошибка; Y – зависимая переменная. Предполагается, что для данной генеральной совокупности именно функция f связывает исследуемую переменную Y с вектором независимых переменных X.

Рассмотрим самую употребляемую и наиболее простую из моделей множественной регрессии – модель множественной линейной регрессии. Теоретическое линейное уравнение регрессии имеет вид:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon$$

или для индивидуальных наблюдений i, $i = 1, 2, \ldots, n$,

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i$$

Здесь $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$ вектор размерности $(m+1)$ неизвестных параметров. β_i, $i = 1, 2, \ldots, n$, называется i-м теоретическим коэффициентов регрессии. Он характеризуется чувствительностью величины Y к изменению X_j и отражает влияние на условное математическое ожидание $M(Y|x_1, x_2, \ldots, x_m)$ зависимой переменной Y объясняющей переменной X_j, при условии, что все другие объясняющие переменные модели остаются постоянными. β_0 – свободный член, определяющий значение Y в случае, когда все объясняющие переменные $X_i = 0$. После выбора линейной функции в качестве модели зависимости необходимо оценить параметры регрессии.

Пусть имеется n наблюдений вектора объясняющих переменных $x = (x_1, x_2, \ldots, x_m)$ и зависимой переменной Y:

$$(X_{i1}, X_{i2}, \ldots, X_{im}, y_i),\ i = 1, 2, \ldots, n.$$

[1] Доцент кафедры математики и физики, Саранский кооперативный институт, г. Саранск; elen.chernoivanova@yandex.ru

Для того, чтобы однозначно можно было решить задачу отыскания параметров $\beta_0, \beta_1, \ldots, \beta_m$, должно выполняться неравенство $n \geq m+1$. Если это неравенство не будет выполняться, то существует бесконечно много различных векторов параметров, при которых линейная формула связи между x и y будет абсолютно точно соответствовать имеющимся наблюдениям. При этом если $n = m+1$, то оценки коэффициентов вектора β рассчитываются единственным образом – путем решения системы $m+1$ линейного уравнения:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}, \ i = 1, 2, \ldots, m+1.$$

Число $k = n - m - 1$ называется числом степеней свободы.

Самым распространенным методом оценки параметров уравнения множественной линейной регрессии является метод наименьших квадратов (МНК).

При выполнении предпосылок МНК относительно ошибок ε_i оценки $b_0, b_1, \ldots, b_m$ параметров $\beta, \beta_2, \ldots, \beta_m$ множественной линейной регрессии по МНК являются несмещенными, эффективными и состоятельными. Отклонение e_i значения y_i зависимой переменной Y от модельного значения $\tilde{y}_i$, соответствующего уравнению регрессии в i-м наблюдении $(i = 1, 2, \ldots, n)$, рассчитываются по формуле:

$$e_i = y_i - b_0 - b_1 x_{i1} - \ldots - b_m x_{im}$$

Тогда по МНК для нахождения оценок $b_0, b_1, \ldots, b_m$ минимизируется следующая функция:

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij}))^2 \tag{9}$$

Данная функция является квадратичной относительно неизвестных величин b_j, $j = 0, 1, \ldots, m$. Она ограничена снизу, следовательно, имеет минимум, Необходимым условие минимума функции Q является равенство 0 всех её частных производных по b_j. Частные производные этой квадратической функции являются линейными функциями

$$\begin{cases} \dfrac{\partial Q}{\partial b_0} = -2 \sum\limits_{i=1}^{n} (y_i - (b_0 + \sum\limits_{j=1}^{m} b_j x_{ij})) \\ \dfrac{\partial Q}{\partial b_j} = -2 \sum\limits_{i=1}^{n} (y_i - (b_0 + \sum\limits_{j=1}^{m} b_j x_{ij})) x_{ij}, \ j = 1, 2, \ldots, m \end{cases} \tag{10}$$

Приравнивая их к 0, мы получаем систему $m+1$ линейных уравнений с $m+1$ неизвестными:

$$\begin{cases} \sum\limits_{i=1}^{n} (y_i - (b_0 + \sum\limits_{j=1}^{m} b_j x_{ij})) = 0 \\ \sum\limits_{i=1}^{n} (y_i - (b_0 + \sum\limits_{j=1}^{m} b_j x_{ij})) x_{ij} = 0, \ j = 1, 2, \ldots, m \end{cases} \tag{11}$$

Такая система имеет обычно единственное решение. Эта система называется системой нормальных уравнений.

Представим данные наблюдений и соответствующие коэффициенты в матричной форме.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{pmatrix}, B = \begin{pmatrix} b_0 \\ b_1 \\ \ldots \\ b_m \end{pmatrix}, e = \begin{pmatrix} e_0 \\ e_1 \\ \ldots \\ e_n \end{pmatrix}$$

Здесь Y – n-мерный вектор-столбец наблюдений зависимой переменной Y; – матрица размерности $n \times (m+1)$, в которой i строка $(i = 1, 2, \ldots, n)$ представляет наблюдения вектора значений независимых переменных $x_1, x_2, \ldots, x_m$; – вектор-столбец размерности $(m+1)$ параметров уравнения регрессии; – вектор-столбец размерности n отклонений выборочных значений y_i зависимой переменной Y от значений $\tilde{y}_i$, получаемых по уравнению регрессии

$$\tilde{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_m x_{im}.$$

Необходимым условием экстремума функции Q является равенство 0 её частных производных $\frac{\partial Q}{\partial b_j}$ по всем направлениям b_j, $j = 0, 1, \ldots, m$.

Приравнивая $\partial Q / \partial B$ к 0, получим

$$B = (X^T X)^{-1} X^T Y.$$

Полученные общие соотношения справедливы для уравнений регрессии с произвольным количеством m объясняющих переменных.

Знание дисперсий и стандартных ошибок позволяет анализировать точность оценок, строить доверительные интервалы для теоретических коэффициентов, проверять соответствующие гипотезы.

Общее качество уравнения регрессии можно проверить с использованием коэффициента детерминации R^2, который в общем случае рассчитывается по формуле:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \overline{y})^2}.$$

Если $0.75 \leq R^2 \leq 1$, то уравнение регрессии составлено верно.

Пусть нам дан временной ряд, характеризующий зависимость плана приема в некоторый ВУЗ от года. Этот ряд задан в виде таблицы.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Время t	1	2	3	4	5	6	7	8	9	10	11
План приема в ВУЗ (y)	97	105	135	152	160	195	223	297	310	318	326

К формированию линейного тренда привлекают все тот же метод наименьших квадратов, полагая в качестве независимой переменной время t, а результирующей переменной – величину y. Тогда оценка параметров тренда определяется формулой

$$\acute{A} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = (T' \cdot T)^{-1} \cdot T \cdot Y, \text{ где } Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \; T = \begin{pmatrix} 1t_1 \\ 1t_2 \\ \ldots \\ 1t_n \end{pmatrix}.$$

По этой формуле вычислим значение матрицы

$$A = \begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix} \cdot \begin{pmatrix} \sum y_i \\ \sum t_i \cdot y_i \end{pmatrix} =$$

$$= \frac{1}{n \sum t_i^2 - (\sum t_i)} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & n \end{pmatrix} \cdot \begin{pmatrix} \sum y_i \\ \sum t_i \cdot y_i \end{pmatrix} =$$

$$= \frac{1}{5566 - 4356} \begin{pmatrix} 506 & -66 \\ -66 & 11 \end{pmatrix} \cdot \begin{pmatrix} 2318 \\ 16783 \end{pmatrix} = \begin{pmatrix} 53,914 \\ 26,136 \end{pmatrix}.$$

Запишем уравнение тренда $y_t = a_0 + a_1 t = 53,914 + 26,136t$. При $t = 12$ значение $y_{12} = 367,55$, то есть план приема в ВУЗ в 2012 году был равен 367.

Список литературы

1. Айвазян С.А., *Основы эконометрики*, «ЮНИТИ», М., 2001.

The mathematical method of calculation plan of education

© **E. A. Chernoivanova**[2]

Abstract. In article propose method of calculation plan of education on classical method of minimal square used.

Key Words: regression curve, coefficient of regression, coefficient of determination, method of minimal square

[2] Associate Professor, Department of Mathematics and Physics, Saransk Cooperative Institute, Saransk; elen.chernoivanova@yandex.ru