УДК 51(077)

К вопросу проверки качества уравнения регрессии

Минченкова Л.П.
Белорусский национальный технический университет

Рассмотрим, как проводится прогнозирование индивидуальных значений зависимой переменной. На практике очень часто более важно знать дисперсию Y, чем ее средние значения или доверительные интервалы для условных математических ожиданий. Это позволяет определить допустимые границы для конкретного значения Y.

Пусть нас интересует некоторое возможное значение y_0 переменной Y при определенном значении x_p объясняющей переменной X. Прогнозируемое по уравнению регрессии значение Y при $X = x_p$ равно y_p. Если рассматривать значение y_0 как случайную величину Y_0, а y_p – как случайную величину Y_P, то можно сказать, что $Y_0 \sim N\left(\beta_0 + \beta_1 x_p, \sigma^2\right)$, а $Y_p \sim N\left(\beta_0 + \beta_1 x_p, \sigma^2\left[\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum(x_i - \bar{x})^2}\right]\right)$. Случайные величины Y_0 и Y_p являются независимыми, а следовательно, случайная величина $U = Y_0 - Y_p$ имеет нормальное распределение с $M(U) = 0$ и $D(U) = \sigma^2\left[1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum(x_i - \bar{x})^2}\right]$.

Но тогда можно сказать, что случайная величина $\frac{U}{S_U} = \frac{Y_0 - Y_p}{S\sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum(x_i - \bar{x})^2}}}$ имеет распределение Стьюдента с числом степеней свободы $v = n - 2$. На основании этого можно сделать вывод, что доверительный интервал $\left(b_0 + b_1 x_p \pm t_{\frac{\alpha}{2}, n-2} S\sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{\sum(x_i - \bar{x})^2}}\right)$ определяет границы, за пределами которых может оказаться не более $100\alpha\,\%$ точек.