

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ НАЛИЧИИ КЛАССИФИКАЦИИ НАБЛЮДЕНИЙ

Е. С. АГЕЕВА, Ю. С. ХАРИН

Multiple regression model with classified observations is considered. Maximum likelihood estimators of the model parameters are constructed and conditions of their strong consistency and asymptotic normality are specified

Ключевые слова: регрессионный анализ, округленные данные, ОМП

В математической статистике широко используется регрессионная модель. Хорошо исследованы случаи, когда зависимые переменные наблюдаются с выбросами или с пропусками, а также ситуация, когда в выборке присутствуют цензурированные наблюдения; при этом построены робастные статистические выводы. В данной работе рассматривается ситуация, когда для множественной регрессионной модели вместо истинных значений зависимой переменной наблюдаются номера классов (интервалов), в которые попадают эти значения. Предлагаемая в статье новая модель множественной регрессии с классифицированными наблюдениями является обобщением известной модели с «округленными данными» (rounded data) [1–3].

Пусть на вероятностном пространстве (Ω, F, P) определена модель нелинейной множественной регрессии:

$$Y_t = F(X_t; \theta^0) + \xi_t, \quad t = 1, \dots, n,$$

где X_t – известный вектор регрессоров, $\xi_t \in R^1$ – случайная величина ошибок с нормальным распределением вероятностей $N(0, (\sigma^0)^2)$, $\alpha^0 = (\theta^0, (\sigma^0)^2) \in R^{m+1}$ – неизвестный вектор параметров.

Пусть задана непересекающаяся последовательность K интервалов $A_k = (a_{k-1}, a_k]$, $k \in K$, $-\infty = a_0 < a_1 < \dots < a_{K-1} < a_K = \infty$. Эта множество интервалов задает классификацию зависимой переменной Y_t : Y_t относится к классу v_t , если $Y_t \in A_{v_t}$, $v_t \in K = \{1, \dots, K\}$.

Задача заключается в том, чтобы по классифицированным наблюдениям v_1, \dots, v_n и значениям регрессоров X_1, \dots, X_n построить статистические оценки для неизвестного вектора параметров α^0 .

Для оценивания вектора параметров α^0 используем оценки максимального правдоподобия $\hat{\alpha}_n$. Для построенных оценок получаем условия для сильной состоятельности, а также условия для асимптотической нормальности и несмещенности [4].

Литература

1. Bai, Z., Zheng, S., Zhang, B., Hu, Z. Statistical Analysis for Rounded Data. J. Statist. Plann. Inference, 2009, 139, no. 8, 2526–2542.
2. Dempster, A.P. and Rubin, D.B. Rounding error in regression: the appropriateness of Sheppard corrections. J. Roy. Statist. Soc. Ser. B., 1983, 45, 51–59.
3. Sen Roy, S., Guriab, S. Estimation of regression parameters in the presence of outliers in response. Statistics, 2009, 43, no. 6, 531–539.
4. Wald, A. Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist., 1949, Vol. 20, no. 4, 595–601.