

К. А. Осипенко, Н. Б. Осипенко
(ГГУ им. Ф. Скорины, Гомель)
ЭКСПРЕСС-ДИАГНОСТИКА КОЛИЧЕСТВЕННОГО
ПРИЗНАКА НА ОСНОВЕ КЛАСТЕРНОЙ РЕГРЕССИИ

Удачность статистического исследования зависимостей в случае применения регрессионного анализа данных зависит от того, насколько верно выбран класс параметрического семейства моделей F , в котором будет осуществлен поиск аппроксимирующей модели $\tilde{f}(x)$ для прогноза значений результирующего показателя y . Стандартных рекомендаций и методов для этого выбора нет, но есть некоторые моменты, которые необходимо учитывать при выборе общего вида регрессии: максимальное использование априорной информации о содержательной (физической, экономической, социологической и др.) сущности анализируемой зависимости; предварительный анализ геометрической структуры данных вида $\{(x_i, y_i)\}$, ($i = \overline{1, N}$, где N – объем выборки), по которым конструируется зависимость; различные статистические приемы обработки данных, улучшающие выбор одного из нескольких вариантов.

На практике не всегда удается найти подходящий аналитический вид регрессионной зависимости. Использование стандартных классов функций может привести к заметной систематической ошибке. Во избежание этого используют методы локального оценивания регрессии (так называемые непараметрические) или же разбивают область возможных значений на несколько частей и строят свое аналитическое описание регрессионной зависимости. В частности, в настоящей работе описывается пример детализации второго метода для предварительной экспресс-диагностики возможной продолжительности жизни человека на основе паспортных сведений [1] и факторов увеличивающих и уменьшающих её. Данный подход является частным случаем метода кластерной регрессии. Идея метода состоит в том, что вся выборка по значениям результирующего признака y разбивается на группы (в нашем случае две: больших и малых значений y) и в каждой отдельной группе строится своя прогнозная модель $y = \tilde{f}(x)$, где x - вектор информативных признаков (для каждой группы является уникальным).

В результате такого прогноза удалось определить группы более значимых в каждом случае экзогенных переменных, что может послужить целям предварительной экспресс-диагностики эндогенной переменной. Результаты проведенного пробного исследования показывают, что паспортные данные вполне пригодны для включения их в список признаков при экспресс-диагностике возможной продолжительности жизни человека.

Литература

1 Осипенко, К.А. Метод регрессионного моделирования продолжительности жизни по дате рождения / К.А.Осипенко, Н.Б.Осипенко // Творчество молодых 2012: сборник научных работ студентов и аспирантов УО «ГГУ им. Ф. Скорины»: в 2 ч. / Гомельский гос. ун-т им. Ф.Скорины; отв. ред. О.М. Демиденко. – Гомель, 2012. –Ч. 1. –С.194-197.