

Е.Е. ПРОЛИСКО, М.В. КОЗЛОВА

КОРРЕКЦИЯ БУТСТРАПОВСКОЙ ИНТЕРВАЛЬНОЙ ОЦЕНКИ ПАРАМЕТРОВ РЕГРЕССИИ

В основе бутстраповского подхода лежит та идея, что истинное распределение данных можно хорошо приблизить эмпирическим, т.е. тем, как данные легли в выборке. На самом деле, эмпирическое распределение – единственный источник информации об истинном распределении данных, что у исследователя есть помимо модели.

Объем бутстраповской выборки обычно рекомендуется брать равным объему основной выборки [1]. Но проведенные имитационные эксперименты по получению интервальных оценок параметров регрессии (использовался стандартный Эфроновский доверительный интервал) показали, что полученные на основе бутстраповского подхода эти оценки являются заниженными. Так, например, вместо ожидаемого доверительного интервала 95%, реально получался доверительный интервал около 92%. Те же исследования показали, что ширина полученного доверительного интервала меняется при изменении объема бутстраповской выборки.

Была поставлена задача получить эмпирическим методом рекомендуемые объемы бутстраповской выборки, обеспечивающие наиболее точную инференцию результатов регрессионного анализа.

Используя методы имитационного моделирования, была исследована парная линейная регрессия при различных ее параметрах (объем выборки, значения параметров регрессии и видов распределения ее ошибок). Обобщая полученные результаты можно сделать следующие выводы:

- 1) оптимальный объем бутстраповской выборки не зависит от параметров регрессии;
- 2) не зависимо от объема исходной выборки оптимальный объем бутстраповской выборки меньше ее на 3-4 единицы;
- 3) отсутствует зависимость объема бутстраповской выборки от вида распределения ошибок регрессии, при условии, что эти распределения – симметричные относительно нуля. Для несимметричных распределений данное утверждение справедливо для коэффициента регрессии, но не для ее константы.

1. Анатолев, С. Основы бутстрапирования / С. Анатолев, // Квантиль. – 2007. – № 3. – С. 1-12.