

Анализ достоверности и полноты данных в эмпирических таблицах на основе непараметрических методов статистики

Рыбаков К.А., Черепанов Е.В.

**Московский авиационный институт
(государственный технический университет)
125993, Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4,
+7 499 158-48-11, dep805@mai.ru**

**Институт экономики и комплексных проблем связи
117209, Москва, ул. Зюзинская, д. 6, корп. 2,
+7 495 331-05-11, mail@nii-ecos.ru**

В эконометрических, социально- и технико-экономических исследованиях, как правило, базой для анализа информации служат таблицы эмпирических данных [1, 3]. При этом, как свидетельствуют многочисленные отечественные и зарубежные авторы, такие эмпирические таблицы часто оказываются неполными (содержат пропуски значений показателей для некоторых наблюдений) и обладают заметной недостоверностью (часть данных неточны – случайные ошибки, ложные сведения). В этой связи проблема выявления недостающей и недостоверной информации в эмпирических матрицах данных может считаться неотъемлемой частью первичной статистической обработки практически во всех прикладных статистических работах.

Первым среди отечественных специалистов с решением этой проблемы выступил Н.Г. Загоруйко [2], совместно с сотрудниками ИМ СО АН разработавший алгоритм «ЗЭТ» («заполнение эмпирических таблиц»), который послужил хорошим инструментом в прикладных исследованиях. Идея алгоритма «ЗЭТ» состоит в том, что таблицы данных являются избыточными, что позволяет оценить недостающие значения матрицы данных.

В данной работе развивается и обобщается методология анализа эмпирических таблиц с целью выявления недостающей и ложной информации. По сути метод опирается на то, что, во-первых, числовые показатели, как правило, коррелированы, и, во-вторых, наблюдения в таблице обладают мерами подобия, что также поддается формализации [5, 7]. Существенна и предпосылка о том, что значения показателей в эмпирических таблицах измерены в интервальных шкалах, что делает допустимыми любые монотонно возрастающие преобразования.

Отечественный и зарубежный опыт свидетельствует о том, что решение задачи выявления недостающей и недостоверной информации в матрицах данных, основанное на прямом использовании числовой информации таблицы, как правило, дает очень грубые и неточные результаты. Связано это с тем, что классические оценки обладают весьма низкой робастностью. Так, выборочное среднее крайне чувствительно к асимметрии эмпирического распределения, а стандартное отклонение – к «ненормальностям» эксцесса. В такой ситуации требуется использование более тонких алгоритмов анализа данных. Предлагается использовать различные подходы, в частности, переход к ранговым переменным, использование робастной оценки Ходжеса-Лемана, основанной на ранговом критерии Уилкоксона, так как использование ранговых статистик является одним из наиболее радикальных методов непараметрической статистики,

дающих стабильность и относительно высокую эффективность. Кроме того, предлагается использовать полиграммные методы оценивания [4, 6], для которых в ряде социально-экономических исследований установлена высокая стабильность и хорошая точность.

Список литературы

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998.
2. Загоруйко Н.Г., Елкина В.Н., Тимеркаев В.С. Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм «ZET») // Вычислительные системы. Вып. 61. Эмпирическое предсказание и распознавание образов. – Новосибирск, 1975. – С. 3–27.
3. Орлов А.И. Эконометрика. – М.: Экзамен, 2004.
4. Рыбаков К.А., Черепанов Е.В. Полиграммные оценки моментов непрерывных случайных величин в социально-экономических исследованиях // 3-й Всероссийский социологический конгресс (ВСК-3). Москва. 2008: Тр. конф. – М.: Институт социологии РАН, 2008.
5. Сютюренко О.В., Черепанов Е.В. Информатика: анализ данных и эконометрия // Средства связи, 1986. – № 4. С. 36–44.
6. Тарасенко Ф.П., Черепанов Е.В. Полиграммные оценки линейных функционалов // Математическая статистика и ее приложения. Вып. X. – Томск: Изд-во ТГУ, 1986. – С. 204–211.
7. Черепанов Е.В. Анализ полноты и достоверности информации в таблицах эмпирических данных // Анализ социально-экономических и политических процессов и систем. Вып. 4. – М.: Академия менеджмента инноваций, 2007. – С. 147–153.