

УДК 519.237(075.8)

ВЫЯВЛЕНИЕ АНОМАЛЬНЫХ ЗНАЧЕНИЙ В ВЫБОРКЕ И ПОЛУЧЕНИЕ УСТОЙЧИВЫХ ОЦЕНОК СРЕДНЕЙ

Куцебо У.С. (Самарский университет),

Научный руководитель – доцент, кандидат физико-математических наук,
Трусова А.Ю. (Самарский университет)

Введение. Современные информационные данные, так как являются достаточно объемными, с большой долей вероятности, могут содержать засорения. Их необходимо оперативно выявлять, т. к. грубые ошибки могут существенно исказить значения характеристик распределения.

Основными причинами наличия аномальных наблюдений являются:

- специфические особенности отдельных элементов изучаемой совокупности;
- неправильное отнесение данных к исследуемой совокупности, например, ошибки группировки, типологической классификации и пр.;
- ошибки при регистрации и обработке данных [1].

Робастные (устойчивые) методы оценивания – это методы оценивания, учитывающие наличие «грубых ошибок» и позволяющие при этом достаточно точно определить оценки параметров.

Основная часть. Актуальность. Данные методы работают в широком спектре наук: инженерно-технических, естественно-научных, экономических, социальных и т. д.

Научная новизна. Применение инструмента робастного оценивания для получения устойчивых оценок средней.

Практическая значимость. Во многих научных направлениях используются данные с целью прогноза, выявления устойчивых тенденций, поэтому важно, чтобы качество выборочной совокупности было высоким.

Целью работы является выявление аномально больших и малых данных по величине прожиточного минимума в регионах России, установленных 1 июня 2022 года [2].

Для достижения цели решались следующие задачи:

- 1) изучение методов робастной статистики;
- 2) формирование исходного массива данных;
- 3) проверка гипотез с помощью критериев Граббса, Титьена и Мура и Хубера;
- 4) анализ полученных результатов.

Для анализа использовались данные с Росстата о величине прожиточного минимума (руб.) по 20 регионам России за 2022 год [2].

В исходной совокупности выделялись значения 11692, 11832 и 31736 (Белгородская область, республика Татарстан и Чукотский автономный округ). В работе анализировались данные по исходной выборке и усеченной выборке, средние значения по ним составили 16835,15, 16555,47 и 16581,9.

Выдвигалась гипотеза о том, что эти показатели могли являться грубыми ошибками. Для проверки данных на засорения, применялся критерий Граббса. В ходе вычислений наблюдаемые значения составили 0,957, 0,93 и 2,772, соответственно. Для критического было взято табличное значение 2,45 на уровне значимости 0,1.

При расчетах следует, что к грубым ошибкам можно отнести только значение Чукотского автономного округа 31736, так как в иных случаях наблюдаемые значения не превышали критического.

Но следует отметить, что хотя критерий Граббса прост и легко применим в анализе, все же имеет существенные недостатки. В частности, исследователи обращают внимание на его недостаточную точность (часто он дает весьма грубые оценки) [3].

Для уточнения результатов проверки использовался более чувствительный критерий Титьена и Мура, при этом для начала все три значения были отнесены к числу ошибок.

Критическое значение составило 0,302 на уровне значимости 0,05. В результате вычислений наблюдаемое значение составило 0,524, таким образом, по данному критерию выборка не содержала грубых ошибок.

Далее был взят показатель Чукотского автономного округа, как единственное аномальное значение. Наблюдаемое и критическое значения составили 0,604 и 0,639. Следовательно, данные по Чукотскому автономному округу действительно можно отнести к аномальным.

В исследовании был проведен расчет оценки Хубера, который предусматривал итеративно повторяющиеся вычислительные процедуры.

На первом шаге в качестве исходной оценки была выбрана простая арифметическая средняя: 16835,15. При одной ошибке в выборочной совокупности найдено значение параметра, равное 1,399.

Данные были расформированы на 3 класса: максимально приближенные к средней величине, значительно больше и меньше нее. Значения из первого класса остались неизменными, к значениям третьего класса был прибавлен параметр, равный 1,399, из значений второго класса данный параметр вычитался. Для первой модификации посчитана устойчивая оценка, равная 16835,99.

Аналогично проводилась вторая модификация, в результате чего оценка составила 16836,41.

Выводы. Таким образом, математические методы робастной статистики способствуют выявлению аномальных данных, что позволяет получать качественные оценки показателей по выборочным данным.

Список использованных источников:

1. Чудинова, О.С. Методы робастного оценивания: методические указания / О.С. Чудинова; Оренбургский гос. ун-т.– Оренбург: ОГУ, 2018. – 33 с
2. Росстат. Федеральная служба государственной статистики [Электронный ресурс]. — URL: <https://rosstat.gov.ru/vpm> (дата обращения 10.11.2022)
3. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. М., Финансы и статистика, 2003. – С 285