

Российской Федерации и в России есть субъекты, показатели инновации которых схожи с Самарской областью. Цель работы достигнута. Были найдены пар схожих объектов методом множественного сравнения Шеффе. Также можно сказать, что на уровне министерств можно разрабатывать стратегии развития региональных экономик на основе метода множественного сравнения Шеффе и LSD – критерия.

**Список использованных источников:**

1. Конституция Российской Федерации. Глава 3. Федеративное устройство: [Электронный ресурс]: - <http://www.constitution.ru/10003000/10003000-5.htm>
2. Метод множественных сравнений Шеффе: [Электронный ресурс]: - <http://www.machinelearning.ru/wiki/index.php>

**ОЦЕНКА ВЫБОРОЧНОЙ СРЕДНЕЙ ПО ДАННЫМ, СОДЕРЖАЩИМ  
«ЗАСОРЕНИЯ»**

**У.С. Куцебо**

Научный руководитель А.Ю. Трусова

Самарский национальный исследовательский университет имени академика  
С.П. Королева

**Актуальность.** При обработке экономической информации чрезвычайно важно выявлять резко выделяющиеся наблюдения, т. к. грубые ошибки могут существенно исказить значения характеристик распределения.

**Научная новизна.** Использование методов исчисления устойчивых статистических оценок Хубера.

**Практическая значимость.** Если очистить данные от ложной информации, то можно получить более точные наблюдения и составить верный прогноз.

**Цель.** Проанализировать экономические показатели регионов России.

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

Выборка из генеральной совокупности может содержать значения, существенно отклоняющиеся от основного массива данных. Такие наблюдения называются аномальными (неправдоподобными, резко выделяющимися) или грубыми ошибками («выбросами») [1]. Наличие аномальных объектов искажает структуру совокупности (является одной из причин неоднородности данных), приводит к смещению оценок параметров распределения и к отклонению закона распределения от теоретического.

**Основными причинами** наличия аномальных наблюдений являются:

- специфические особенности отдельных элементов изучаемой совокупности;

- неправильное причисление данных к исследуемой совокупности, например, ошибки группировки, типологической классификации и пр.;
- ошибки при регистрации и обработки данных[2].

В таблице приведены данные с Росстата о величине прожиточного минимума по регионам России за 2022 год.

Таблица 1. Величина прожиточного минимума с 1 июня 2022 года.

Регионы	Исходные данные	Усеченная совокупность данных	Винзорированные данные
Белгородская область	11692	-	12388
Республика Татарстан	11832	-	12388
Омская область	12436	12436	12436
Ставропольский край	12528	12528	12528
Республика Алтай	13085	13085	13085
Самарская область	13085	13085	13085
Чеченская республика	13363	13363	13363
Краснодарский край	13363	13363	13363
Республика Крым	13501	13501	13501
Тюменская область	13919	13919	13919
г. Санкт-Петербург	14476	14476	14476
Республика Бурятия	15172	15172	15172
Красноярский край	15451	15451	15451
Еврейская автономная область	18758	18758	18758
Ямало-ненецкий авт. округ	19832	19832	19832
г. Москва	20585	20585	20585
Мурманская область	22250	22250	22250
Магаданская область	24220	24220	24220
Ненецкий автономный округ	25419	25419	25419
Чукотский автономный округ	31736	-	25419
Итого	336703	281443	331638
Средняя величина	16835,15	16555,47	16581,9
Дисперсия	28890436,93	-	-
Ср. квадратическое отклонение	5374,98	-	-

Данные были систематизированы по возрастанию. В исходной совокупности выделяются значения 11692, 11832 и 31736 (Белгородская область, республика Татарстан и Чукотский автономный округ). Также были посчитаны средние по исходным данным, по усеченной совокупности и винзорированным данным по формуле, представленной ниже. Выдвигается гипотеза о том, что эти показатели могут являться грубыми ошибками.

Чтобы проверить данные на засорения, необходимо применить критерий Граббса, используя формулу для всех 3-х значений

$$(1) \quad T = \frac{x - (\bar{x})}{\sigma},$$

для  $T_{кр}$  было взято табличное значение (критерий Смирнова-Граббса при  $\alpha=0,1$ ).

Таблица 2 – Подсчет T-критерия.

Критерий Граббса	
T1	0,957
T2	0,93
T3	2,772
$T_{кр}$	2,45

При расчетах следует, что к грубым ошибкам можно отнести только значение Чукотского автономного округа 31736, так как в иных случаях  $T_{кр} > T_{набл}$ .

Но следует отметить, что хоть 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].

Уточнить результаты проверки можно при помощи более чувствительного критерия Титьена и Мура, при этом для начала нужно отнести все три значения к числу ошибок, посчитав E-критерий с помощью формулы

$$(2) \quad E = \frac{\sum_{i=3}^{17} (x_i - \bar{x}_k)^2}{\sum_1^{20} (x_i - \bar{x})^2},$$

для  $E_{кр}$  было взято табличное значение (E-критерий Титьена и Мура при  $\alpha=0,05$ ).

Таблица 3 – Подсчет E-критерия.

E-критерий Титьена и Мура	
E	0,524
$E_{кр}$	0,302

Также необходимо посчитать L-критерий, взяв значение Чукотского автономного округа, как аномальное значение, и рассчитать по формуле

$$(3) \quad L = \frac{\sum_{i=1}^{19} (x_i - \bar{x}_k)^2}{\sum_1^{20} (x_i - \bar{x})^2},$$

для  $L_{кр}$  было взято табличное значение (L-критерий Титьена и Мура при  $\alpha=0,05$ ).

Таблица 4 – Подсчет L-критерия.

L-критерий Титьена и Мура	
L	0,604
$L_{кр}$	0,639

Результаты проверки представлены в таблице 5.

Таблица 5 – Проверка гипотезы.

Проверка гипотезы		
E	Выборка однородна и не содержит грубых ошибок	$E_{набл} > E_{кр}$
L	Выборка неоднородна	$L_{набл} < L_{кр}$

Итак, можно сделать вывод, что данные по Чукотскому автономному округу действительно можно отнести к аномальным. Так, необходимо провести расчет оценки Хубера, который предусматривает итеративно повторяющиеся вычислительные процедуры.

На первом шаге в качестве исходной оценки выбрана простая арифметическая средняя: 16835,15. При одной ошибке в выборочной совокупности найдено значение параметра, равное 1,399. В таблице 6 представлены данные, расформированные на 3 класса: максимально приближенные к средней величине, значительно больше и меньше нее.

Таблица 6 – Первая модификация данных.

	1 класс	2 класс $x_i > 16836,55$	3 класс $x_i < 16833,75$
Исходные значения		18758, 19832, 20585, 22250, 24220, 25419, 31736	11692, 11832, 12436, 12528, 13085, 13085, 13363, 13363, 13501, 13919, 14476, 15172, 15451
Модифицированные значения		18757, 19831, 20584, 22249, 24219, 25418, 31735	11693, 11833, 12437, 12529, 13086, 13086, 13364, 13364, 13502, 13920, 14477, 15173, 15452

Для первой модификации посчитана устойчивая оценка по формуле

$$(4) \quad \hat{\theta} = \frac{1}{n} (\sum_{|x_i - \hat{\theta}| < k} x_i + (n_2 - n_1)k),$$

равная 16835,99.

Таблица 7 – Вторая модификация данных.

	1 класс	2 класс $x_i > 16837,39$	3 класс $x_i < 16834,59$
Исходные значения		18757, 19831, 20584, 22249, 24219, 25418, 31735	11693, 11833, 12437, 12529, 13086, 13086, 13364, 13364, 13502, 13920, 14477, 15173, 15452
Модифицированные значения		18755, 19829, 20582, 22247, 24217, 25416, 31733	11695, 11835, 12439, 12531, 13088, 13088, 13366, 13366, 13504, 13922, 14479, 15175, 15454

В таблице 7 произведена повторная модификация данных, где посчитана вторая устойчивая оценка, равная 16836,41.

Исходя из вышесказанного, можно сделать вывод, что данные Росстата о величине прожиточного минимума по регионам России за 2022 год содержат одно аномальное наблюдение. Данное отличие может быть связано как с географическим расположением субъекта, уровнем инфраструктуры, так и с ошибками расчетов.

#### **Список использованных источников**

1. Большаков, А. А. Методы обработки многомерных данных и временных рядов: учеб. пособие для вузов / А. А. Большаков, Р. Н. Каримов. – М.: Горячая линия – Телеком, 2007. – 522 с.
2. Чудинова, О.С. Методы робастного оценивания: методические указания / О.С. Чудинова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018. – 33 с

3. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. М., Финансы и статистика, 2003. – С 285

## **ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА СТАТИСТИКУ ЗАБОЛЕВАЕМОСТИ**

**А.А. Матвеева**

Научный руководитель А.Ю.Трусова

Самарский национальный исследовательский университет имени академика С.П. Королева

Современный мир подвержен глобальным изменениям со стороны внешней политики, внутренней политики, сферы культуры, социальной сферы. Происходит постоянное развитие во всех областях жизни человека. Появляются инновационные технологии, которые усовершенствуют промышленность. Для дальнейшего развития в области экономики требуется множество природных ресурсов, затрачиваемых производством, для комфортного и обеспеченного существования населения. Но у развития экономической промышленности есть и минусы. К сожалению, с годами леса уничтожаются, озера пересыхают, происходят глобальные химические и другого рода выбросы в атмосферу, что приводит к плохой экологии, а следовательно к исчерпанию природных запасов.

Экологическая проблема стала видимой относительно недавно. С появлением пластиковых пакетов, нефтепродукции, разных уходовых средств возросло количество загрязнений планеты. Данные продукты за множество лет не успевают разложиться, это одна из причин, по которой происходят глобальные катаклизмы на Земле.

В данной статье будут рассмотрены показатели формирования отходов по видам экономической деятельности, которые являются одними из факторов экологического загрязнения.

Первичный анализ будет проводиться на данных временных рядов показателей формирования отходов, взятых с сайта Росстат [1]. Целью первичного анализа является наглядное представление о динамике экологических показателей (рис.1).

По графику видно, что с годами динамика показателей только возрастает. Поэтому существует тенденция к увеличению отходов, что негативно повлияет на окружающую среду.

Для определения точного направления динамики показателей в работе проверяются гипотезы о наличии тренда, а именно критерия «восходящих и нисходящих» серий и медианного метода, которые показывают, что тренд у данных значений присутствует.