

Проверка гипотез показала, что все матрицы ковариаций статистически не отличаются друг от друга.

В ходе исследования было выяснено, что степени связи между изучаемыми показателями с течением периода сохраняются, также они не меняют направление. Однако, инновациям стало уделяться больше внимания с течением времени, это объясняет различия в средних показателях.

Таким образом можно сделать вывод, что инновации имеют стабильную тенденцию к развитию и оказывают влияние не только на свою сферу применения, но и на другие экономические показатели такие как прожиточный минимум, ВРП и среднюю заработную плату.

Список использованных источников

1. Степанова Ю.Н., Лесникова М.С. Роль инноваций в современном развитии российского общества // Международный студенческий научный вестник. – 2017. – № 6.;

2. Трусова А.Ю. Многомерные статистические методы: учебное пособие для студентов факультета экономики и управления: в 2 ч. Ч.1. / А.Ю. Трусова; Федер. агенство по образованию. – Самара: Изд-во «Самарский университет» 2008. – 67с.

3. Медведев Ю.И., Ивченко Г.И. Математическая статистика. Москва, 1984.

ПОИСК ПАР СХОЖИХ ОБЪЕКТОВ МЕТОДОМ МНОЖЕСТВЕННОГО СРАВНЕНИЯ ШЕФФЕ

А.А.Васильева

Научный руководитель: А.Ю.Трусова

Самарский национальный исследовательский университет имени академика С.П. Королева

В России более 80 регионов и в каждом из них разный уровень развития показателей инновации [1]. Поэтому перед работой были поставлены несколько вопросов:

1) Найти схожие временные периоды по уровню развития показателей инновации?

2) Найти схожие субъекты в Российской Федерации в сравнении с Самарской областью?

Целью данной работы является поиск пар схожих объектов методом множественного сравнения Шеффе. Будут рассмотрены массивы с затратами, использованными ресурсами и разработанными ресурсами временных периодов схожие по уровню развития показателей инновации и субъектов Российской Федерации в сравнении с Самарской областью.

Метод множественных сравнений Шеффе - это модификация t-критерия Стьюдента. Это параметрический тест, который выявляет наличие статистически значимых различий между средними для нормально распределенных связанных групп на основе дисперсионного анализа. Объемы выборок могут различаться [2].

В работе были рассмотрены два типа задач. В первом типе будут рассмотрены регионы РФ для нахождения схожих временных периодов по уровню развития показателей инновации. Во втором типе будут найдены регионы, схожие по уровню развития показателей инновации с Самарской областью.

Рассмотрим массив данных, состоящий из трех федеральных округов Приволжского, Северо-Западного и Центрального. Общее количество 40 регионов для первой задачи: Белгородская область, Брянская область, Владимирская область, Воронежская область, Ивановская область, Калужская область, Костромская область, Курская область, Липецкая область, Московская область, Орловская область, Рязанская область, Смоленская область, Тамбовская область, Тверская область, Тульская область, Ярославская область, Республика Карелия, Республика Коми, Архангельская область, Вологодская область, Калининградская область, Ленинградская область, Мурманская область, Новгородская область, Псковская область, Республика Башкортостан, Республика Марий Эл, Республика Мордовия, Республика Татарстан, Удмуртская республика, Чувашская республика, Пермский край, Кировская область, Нижегородская область, Оренбургская область, Пензенская область, Самарская область, Саратовская область, Ульяновская область. Во второй задаче рассмотрим массив без Московской и Ленинградской областей и республики Татарстан в количестве 37 регионов. Общий массив представлен в таблице 1.

Таблица 1. Общий массив исходных данных

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Белгородская область	1215	1030	1614	1421	1837	1970	2335	2408	2444	2536	3401
Калужская область	4858	2316	2094	2057	2130	2446	2387	3176	4473	4639	3453
Костромская область	1623	1069	1302	1541	1634	1608	1666	1668	1678	1560	1491
...											
Самарская область	6189	6870	6688	7291	7769	8630	7568	7506	7852	8037	7425
Саратовская область	5911	4359	4630	4664	5087	5437	7529	7363	7628	7734	5499
Ульяновская область	1488	1685	1798	1717	1679	1821	2220	1850	2618	2081	1729

Для решения задач была найдена внутригрупповая дисперсия S^2 , которая отражает случайную вариацию. Формула расчета внутригрупповой дисперсии:

$$(1) S_{\text{вн}}^2 = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

Помимо вычисления внутригрупповой дисперсии, произведены также промежуточные расчеты. Сформирована таблица для попарного сравнения средних значений. Посчитаны S наблюдаемое и $S_{\text{критическое}}$. Если значение S набл. было больше $S_{\text{кр.}}$, то принималась гипотеза H_0 , и если $S_{\text{набл.}} < S_{\text{кр.}}$, то верна гипотеза H_1 .

Для первой задачи были рассмотрены следующие гипотезы:

H_0 : Во всех годах затраты были одинаковые;

H_1 : В каждый год затраты были разные.

При решении этой задачи были найдены следующие пары групп: в 2010 и 2017 годах, в 2011 и 2019, а также в 2012 и 2020 годах был схож объем затрат, потребляемый всеми регионами.

По аналогичному принципу решаются другие две задачи первого типа: используемые ресурсы и разработанные ресурсы. В результате были найдены следующие пары. В 2010 и 2014 годах, в 2011 и 2013, в 2012 и 2015, 2016 и 2017, а также в 2018 и 2019 годах был схож объем используемых затрат, потребляемый всеми регионами. В 2010 и 2018 годах был схож объем разработанных затрат, потребляемый всеми регионами. Таким образом верная гипотеза H_1 .

Аналогично решается второй тип задачи. Пусть будут выдвинуты гипотезы:

H_0 : Существуют регионы Российской Федерации, затраты, используемые и разработанные ресурсы которых совпадают со значениями Самарской области;

H_1 : В Российской Федерации не существуют регионов, затраты, используемые и разработанные ресурсы которых совпадают со значениями Самарской области.

При формировании пар групп сравниваем значение Самарской области с остальными регионами. И вывод, представленный в таблице 3, получается исходя из итогового значения S набл.:

1) объем затрат в Пермском крае и в республике Башкортостан схож с объемом затрат Самарской области. В остальных регионах объем затрат не совпадает с затратами Самарской области;

2) объем используемых затрат Самарской области схож с объемами используемых затрат в республике Башкортостан, Удмуртской республике, Пермском крае и Саратовской области. В остальных регионах объем затрат не совпадает с затратами Самарской области;

3) объем разработанных ресурсов Самарской области схож с объемом разработанных ресурсов Владимирской области, Смоленской области и Нижегородской области. В остальных регионах объем затрат не совпадает с затратами Самарской области. Таким образом во втором типе задач была верна гипотеза H_0 .

Подводя итог, можно сказать, что в этой работе были найдены схожие временные периоды по уровню развития показателей инновации в регионах

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;