

## **ФОРМИРОВАНИЕ ОБУЧАЮЩИХ ВЫБОРОК МЕТОДОМ МНОЖЕСТВЕННОГО СРАВНЕНИЯ ШЕФФЕ**

**Васильева А.А.** (Самарский национальный исследовательский университет им.  
С.П. Королева)

Научный руководитель – доцент, кандидат физико-математических наук, -  
**Трусова А.Ю.** (Самарский национальный исследовательский университет им. С.П.  
Королева)

**Введение.** Аналитика больших информационных потоков предполагает формирование обучающих выборок с целью последующего перераспределения будущих данных. В настоящее время, все сферы человеческой деятельности технические, гуманитарные, социально-экономические охвачены масштабными информационными данными. В связи с чем интерес представляет расширение возможностей формирования групп, которые характеризуются однородными свойствами и могут быть использованы как эталоны.

Любой информационный поток данных имеет ключевой показатель или группу показателей, среди которых, например может быть показатель времени, а также любой количественный или качественный показатель [1]. Таким образом выбрав показатель инновации в качестве ключевого, решались задачи формирования групп по фактору времени, а также по фактору субъекта федерации.

**Основная часть.** Актуальность работы заключается в изучении возможностей на региональном уровне выделять эталонные субъекты или эталонные временные периоды. Научная новизна исследования заключается в применении метода множественного сравнения Шеффе в анализе многомерного массива. Практическая значимость заключается в реализации альтернативных ветвей, которые могут быть ориентирами для последующих информационных потоков данных, характеризующих инновации.

Целью данного исследования является формирование эталонных групп методом множественного сравнения Шеффе.

Для достижения поставленной цели были определены следующие задачи:

- изучить теорию метода множественного сравнения Шеффе и LSD-критерия;
- сформировать массивы;
- рассчитать значения критериев метода множественного сравнения Шеффе и LSD;
- сформировать эталонные группы.

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. Объектами исследования первого типа являются регионы РФ. Фактором, формирующим эталонные группы, является временной период, характеризующий уровень развития показателей инновации с 2010-го по 2020-й годы. Второй тип задачи предполагает поиск регионов, которые формируют обучающие выборки по показателям инновации. В исследовании использовались массивы затрат, по используемым ресурсам и по разработанным ресурсам.

Для решения задач была найдена внутригрупповая дисперсия, которая отражает случайную вариацию, сформирована таблица для попарного сравнения средних значений, а также сформулированы нулевая гипотеза и конкурирующая гипотеза, рассчитаны наблюдаемые и критические значения.

При решении первого типа задач сформулирована гипотеза, что объем затрат, используемых ресурсов и разработанных ресурсов, совпадают или не совпадают по времени. В результате решения первого типа задач результаты были сформированы в следующие временные периоды: с 2010 по 2012 год, в 2017, в 2019, в 2020 годах был схож объем затрат, потребляемый всеми регионами. Объем используемых затрат был схож во всех годах, кроме 2020-го года. В 2010 и 2018 годах был схож объем разработанных затрат во всех регионах.

Во втором типе задач выдвинута нулевая гипотеза о наличии схожести субъектов РФ по затратам со значениями Самарской области, и конкурирующая гипотеза - в Российской Федерации отсутствуют регионы, затраты, используемые и разработанные ресурсы которых схожи со значениями Самарской области.

По результатам расчётов были сформированы эталонные группы регионов и эталонные временные периоды по показателям инновации.

**Выводы.** В работе предложена методика обучающих выборок методом множественного сравнения Шеффе. В качестве показателя группировки использовался показатель времени и показатель субъекта Российской Федерации. Дополнительно изучены особенности эталонных групп схожих по показателям инновации с Самарской областью.

#### **Список использованных источников:**

1. Мхитарян В.С., Архипова М.Ю., Дубровина Т.А., Миронкина Ю.Н., Сиротин В.П. Анализ данных // Учебник для вузов. – 2022. – С. 19-21.

2. Метод множественных сравнений Шеффе: [Электронный ресурс]: - <http://www.machinelearning.ru/wiki/index.php>

Васильева А.А. (автор)

Подпись

Трусова А.Ю. (научный руководитель)

Подпись