

4. Графическая визуализация результатов кластеризации.

5. Выбор оптимального варианта кластеризации

Таким образом, мы поняли: Классификация без обучения позволяет получить качественно однородные группы кластеров.

Изучаемые показатели позволяют формировать стратегию развития объектов диаметрально расхожих.

Учитывая весовую кластеризацию с применением евклидовой метрики, происходит выявление доминирующих показателей выявляющих субъекты, которые не удовлетворяют условиям расхожести.

Список использованных источников

1. Анализ больших наборов данных Джеффри Дэвид Ульман 29 января 2022 г.
2. Теоретический минимум по Big Data. Всё что нужно знать о больших данных 18 сентября 2018 г.

МНОГОМЕРНЫЙ ПОДХОД ПРИ ВЫЯВЛЕНИИ РАЗЛИЧИЙ И СТРУКТУРИЗАЦИЯ С ОБУЧЕНИЕМ

Н.А. Базанов

Научный руководитель А.Ю. Трусова

Самарский национальный исследовательский университет имени академика С.П. Королева

Актуальность работы. Актуальность обусловлена изучением факторов, призванных повысить уровень инновационного развития в ПФО, используя многомерные статистические методы анализа. Показатели инноваций формируют стратегию развития общества и определяют дальнейшие перспективы. Инновации играют важную роль в развитии общества.

Научная новизна заключается в использовании многомерных подходов при анализе показателей влияющих на инновации.

Практическая значимость состоит в том, что инновации являются активным звеном всех сфер жизни общества и вызывают в них прогрессивные изменения. Преимуществами инноваций являются появление новых профессий, интеллектуализация условий труда, повышение уровня образованности и культуры, однако недостатками являются разрушение такого источника экономического роста, как полная занятость [1].

Цель работы: выявить различия и структурировать многомерные данные в сфере инноваций в ПФО в период с 2017 – 2020 годы.

Для достижения цели были выполнены следующие задачи:

- 1) сформировать выборку по показателям инноваций;
- 2) проверить гипотезы о равенстве многомерных показателей многомерному постоянному вектору (стандарту);

3) проверить гипотезы о равенстве показателей инноваций в ПФО в разные периоды;

4) проверить гипотезы о сохранении тесноты взаимосвязи между показателями.

Социально-экономические процессы и явления зависят от большого числа параметров, их характеризующих, что обуславливает трудности, связанные с выявлением структуры взаимосвязей этих параметров.

Многомерный статистический анализ позволяет выявить взаимосвязи между различными факторами социально-экономических явлений и сформировать стратегию для улучшения или поддержания необходимых результатов в том или ином исследовании [2].

Для исследования были использованы данные с показателями по инновациям с 2017 по 2020 год.

Таблица 1 – Данные по инновациям в 2017 году

	Затраты	Разработанные	Использованные	ВРП	Ср. зп	Прож.мин.
Республика Башкортостан	29888,73	8	10026	1487892,20	30358	8318
Республика Марий Эл	1123,92	0	980	178193,20	25440	8742
Республика Мордовия	3894,29	10	2619	236090,80	24327	7824
Республика Татарстан	78404,64	57	7648	2264655,80	32324	8334
Удмуртская Республика	6550,75	21	5651	592037,60	28995	8453
Чувашская Республика	4750,13	0	3167	296505,80	24530	8236
Пермский край	33459,78	33	4216	1245826,90	32952	10098
Кировская область	6699,55	0	2449	331754,40	25215	9627
Нижегородская область	94979,17	30	8633	1387920,90	30387	8739
Оренбургская область	23093,14	0	1154	872857,10	27445	8234
Пензенская область	8473,46	0	1727	376076,20	26238	8326
Самарская область	31226,95	25	7506	1449005,70	30492	9400
Саратовская область	7895,63	11	7363	728946,40	24738	8222
Ульяновская область	6478,87	21	1850	374827,90	26254	9062

В исследовании выдвигается гипотеза о том, что средние значения показателей инноваций статистически значимо не различаются с показателями вектора стандарта. Данная гипотеза позволяет определить,

насколько средние значения экономических показателей соответствуют стандартным значениям таких показателей. В качестве обучающего вектора (вектора стандарта) были взяты средние значения показателей инновации по всей России.

Таблица 2 – Средние значения показателей инноваций с 2017 по 2020 год.

Фактор	2017	2018	2019	2020
Затраты	24065,64374	28380,29	31235,44	36353,48
Разработанные	15,42857143	18,71	14,50	21,43
Используемые	4642,071429	5444,86	5495,43	4990,00
ВРП	844470,7786	952199,79	1007410,27	959394,52
Ср. зп	27835,35714	30541,14	33049,43	35284,79
Прож.мин.	8686,785714	9066,21	9477,21	10150,43

Таблица 3 – Векторы стандартов по инновациям с 2017 по 2020 год.

Фактор	стандарт 2017	стандарт 2018	стандарт 2019	стандарт 2020
Затраты	16276,21875	17045,12	20698,24	23637,28
Разработанные	15,075	17,98	14,58	17,69
Используемые	3389,425	3815,48	3919,48	3647,56
ВРП	693292,5225	770635,06	820948,21	831204,74
Ср. зп	30176,675	33378,50	36100,75	38686,13
Прож.мин.	9411,575	9874,63	10326,43	10997,44

Были выдвинуты гипотезы:

$$(1) H_0: \bar{X} = \mu; H_1: \bar{X} \neq \mu$$

При выполнении проверки гипотез используется матрица ковариации, элементы которой вычисляются при помощи соотношения:

$$(2) S = \frac{1}{n-1} (Z^T Z),$$

Далее выполняется проверка гипотезы. Наблюдаемое и критическое значение критической статистики вычисляется с помощью Т-критерия Хотеллинга:

$$(3) T_{набл}^2 = n(\bar{X} - \mu)^T S^{-1} (\bar{X} - \mu)$$

$$(4) T_{кр}^2 = (\alpha; k_1; k_2) = \frac{m(n-1)}{n-m} F(\alpha; k_1; k_2)$$

В случае принятия конкурирующей гипотезы, рассчитывается частный критерий Хотеллинга, так как необходимо выявить какие именно показатели значимо различаются.

Таблица 4 – Результат проверки гипотезы о равенстве средних значений вектору μ

Результаты проверки				
Год	Результат	T^2 набл.	T^2 кр	Частный критерий Хотеллинга
2017	H1	62,94352543	34,905	Все значения несущественно отличаются друг от друга
2018	H1	116,0498188	34,905	Все значения несущественно отличаются друг от друга
2019	H1	198,4044328	34,905	X6 существенно отличается от X6 постоянного вектора

2020	H1	394,2264829	34,905	X6 существенно отличается от X6 постоянного вектора
------	----	-------------	--------	---

В ходе исследования было выяснено, что средние показатели отличаются от обучающего вектора. В основном различия несущественные, что говорит о стабильном уровне развития инноваций. Также, при помощи частного критерия Хотеллинга было выяснено, что прожиточный минимум в 2019 и 2020 годах уже существенно отличается от обучающего вектора.

Далее была проверены гипотезы о равенстве средних значений показателей попарно между каждым годом.

Для проверки гипотез также используется Т-критерий Хотеллинга.

Таблица 5 – Результаты проверки гипотез о равенстве средних значений

Результаты				
	T ² набл.	T ² крит.	Результат	Частный критерий Хотеллинга
2017 и 2018	16,15504417	19,11157219	H0	
2017 и 2019	78,99442374	19,11157219	H1	Все несущественно отличаются
2017 и 2020	180,5956659	19,11157219	H1	Все несущественно отличаются
2018 и 2019	22,74188786	19,11157219	H1	Все несущественно отличаются
2018 и 2020	114,4874017	19,11157219	H1	Все несущественно отличаются
2019 и 2020	47,47670269	19,11157219	H1	Все несущественно отличаются

Средние показатели в 2017 и 2018 году статистически не различимы, а в остальные года все значения несущественно отличаются друг от друга.

Для выявления взаимосвязей и проверки гипотезы о сохранении зависимости с течением времени были выдвинуты гипотезы о равенстве матриц ковариаций.

Учет ковариаций изучаемого комплекса признаков и проверка равенства матриц ковариаций значительно снижают возможность появления ошибки в выводах. Это происходит из-за малой вероятности совпадения одновременно большого числа сложных характеристик связей признаков.

Выдвигаются гипотезы:

$$(5) H_0: \Sigma_1 = \Sigma_2, H_1: \Sigma_1 \neq \Sigma_2$$

Наблюдаемое и критическое значение статистики определяется отношением:

$$(6) W_{набл} = b * \ln(v)$$

$$(7) W_{кр} = \chi^2(\alpha; k), k = \frac{m(m+1)}{2}$$

Таблица 6 – Результаты проверки гипотез о равенстве матриц ковариаций

Результаты			
Год	Wнабл	Wкр	Результат
2017 и 2018	21,83912	32,67057	H0
2017 и 2019	22,76297	32,67057	H0
2017 и 2020	30,76123	32,67057	H0
2018 и 2019	23,83584	32,67057	H0
2018 и 2020	29,57423	32,67057	H0
2019 и 2020	15,55558	32,67057	H0

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. агентство 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.