

2. «Математикалық ұғымдарды оқыту негіздері» әдістемесі. Ә. Кенеш. Алматы-1999.

3. «Математика және логика» ғылыми әдістемелік журнал. 2014 ж. 4-ші басылым.

4. «Алгебра және анализ бастамалары» 11 сынып оқулығы Алматы «Атамұра» 2011.

5. Интернет www.wikipedia.kz

Калжанов М.У.¹, Едрисова А.С.²

1. Научный руководитель, кандидат физико-математических наук, доцент

2. Студент 4 курс, кафедра физико-математических и общетехнических дисциплин, специальность «Математика»

ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА В ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ

В математической статистике существует множество различных методов для описания, систематизации и анализа данных. Один из таких методов это кластерный анализ. В век стремительного развития технологий, компьютеры стали неотложной частью человеческой деятельности. Они так же играют большую роль в математической статистике. Некоторые методы описательной статистики так же подразумевают собой использование самых новейших и современных компьютеров для решения ряда задач. К таким методам математической статистики относится и кластерный анализ, который направлен на сжатие большого массива информации и систематизацию объектов по их схожим признакам [1, стр.1].

Кластеризацией люди занимались с древнейших времен, до того момента когда появились первые компьютеры. Люди не осознавая того, что занимаются кластеризацией, группировали растения на съедобные и несъедобные, животных, по признаку опасности и так далее. Это было важно для сохранения жизни. Уже тогда начались зарождаться предпосылки к кластерному анализу. Р.Р. Сокэл считал, что кластеризация это высший уровень интеллектуальной деятельности [2, стр. 7-19].

По мере развития современного мира, растет и количество поступающей информации. Объем этой информации может до таких размеров, которые люди не способны сами систематизировать и анализировать. Систематизация в подобных случаях бывает проблематичной. Так же информация поступает с невероятно высокой скоростью. Нередко, люди попросту на физическом уровне не могут рассматривать информацию в темпе их поступления. При данных обстоятельствах кластерный анализ подходит для выполнения данной задачи, потому что готов в достаточной мере, моментально осуществлять сортировку объектов, принимая во внимание все без исключения нужные характеристики, описанные в этом методе.

Вышеуказанные позиции определили актуальность и позволили уточнить проблему исследования, заключающуюся в поиске путей для сокращения и сжимания массивных информационных, применяемых в обучении и образовании.

Кластерный анализ – это способ группировки многомерных объектов, основанный на представлении результатов отдельных наблюдений точками подходящего геометрического пространства с последующим выделением групп как «сгустков» этих точек. В переводе с английского языка «кластер» означает «скопление» или «сгусток» [3, стр.4].

С помощью кластерного анализа можно проводить разбиение объекту по нескольким признакам. Чтобы провести анализ данных применяют «меры сходства» [4, стр. 176]. Мера сходства - это вычисление расстояние между объектами. Самая популярная мера сходства это евклидово расстояние. Оно вычисляется по формуле:

$$p = \sqrt{\sum (x_{ik} - x_{jk})^2}$$

В случае использования дихотомических (имеющих всего два значения) качественных признаков широко используется расстояние Хемминга

$$d_{ij} = \sum_{k=1}^v |x_{ik} - x_{jk}|$$

равное числу несовпадений значений соответствующих признаков для рассматриваемых i-того и j-того объектов.

Методы кластерного анализа можно разделить на два вида: иерархические и неиерархические. Каждая из этих групп, включает в себя множество различных подходов и алгоритмов. Наиболее популярный и часто используемый метод: метод ближнего соседа. В методе ближнего соседа или по-другому одиночной связи, расстояние между кластерами определяется расстоянием между близкими объектами в разных кластерах.

Пример кластерного анализа. Метод ближнего соседа.

Рассматривается малая группа учеников из 8 человек. У которых x_1 –это характеристика оценок по предмету алгебра, x_2 – это характеристика оценок учащихся по предмету геометрия. Данные приведены в таблице 1.

Таб. 1

	1 Аетова А.	2 Айсина М.	3 Аманбаев А.	4 Ахметова А.	5 Белова В.	6 Васильева И.	7 Галуза В.	8 Зиннатуллина А.
x_1	3	5	4	4	5	3	4	4
x_2	3	4	4	4	4	3	4	4

Найдем расстояние между объектами. Воспользуемся этой формулой:

$$p = \sqrt{\sum (x_{ik} - x_{jk})^2}$$

Получим слеующую матрицу расстояний, представленную в таблице 2.

Таб. 2

	1	2	3	4	5	6	7	8
1	0	2,23	1,4	1,4	2,23	0	1,4	1,4

2		0	1	1	0	2,23	1	1
3			0	0	1	1,4	0	0
4				0	1	1,4	0	0
5					0	2,23	1	1
6						0	1,4	1,4
7							0	0
8								0

Шаг 1. На первом шаге объект представляет собой отдельный кластер: $|1|, |2|, |3|, |4|, |5|, |6|, |7|$ и $|8|$. Согласно критерию кластеризации, объединение проходит среди кластеров, расстояние между ближайших представителей которых наименьшее: кластеры $|3,4,5|$ и $|7,8|$. Расстояние на котором произошло объединение – 1. Необходимо произвести перерасчет матрицы расстояний с учетом нового кластера. Новые данные представлены в таблице 3.

Таб. 3

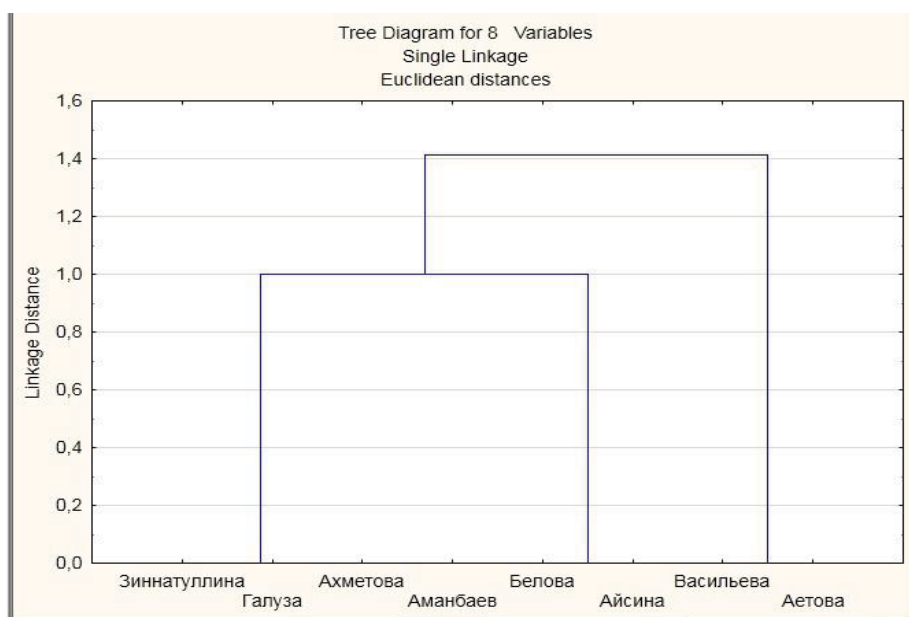
	1	2,3,4,5	6,7,8
1	0	2,23	0
2		0	2,23
3,4,5			1,4
6,7,8			0

Шаг 2. Кластеры на данном шаге $|1|, |2,3,4,5|$ и $|6,7,8|$. Согласно новой матрицы расстояний, кластеры $|2,3,4,5|$ и $|6,7,8|$ наиболее близки. Расстояние объединения – 1,4. Проведенный перерасчет матрицы расстояний с учетом нового кластера представлен в таблице 4.

Таб. 4

	1	2,3,4,5,6,7,8
1	0	2,23
2,3,4,5,6,7,8		0

Шаг 3. Расстояние между кластерами – 2,23. Образование кластеров закончено. Результат кластеризации методом ближнего соседа представлен в виде дендограммы:



Дендограмма получена использованием статистического пакета Statistica.

Разобраться в значении кластеров помогают кластерные профили, которые представляют собой средние значения переменных, которые включены в анализ, распределенные по кластерной принадлежности. Средние значения оценок учащихся по кластерам представлены в таблице 5.

Таб.5

	Кластер 1	Кластер 2
Алгебра	4,3	3
Геометрия	4	3

Вывод: При использовании метода ближнего соседа, получено два кластера. В первый кластер вошли 6 человека (Зиннатуллина А., Галуза В., Ахметова А., Аманбаев А., Белова В., Айсина М.). Во второй 2 человека (Васильева И., Аетова А.). В первый кластер вошли ученики, у которых средний бал по предмету больше 4. Во второй кластер вошли учащиеся, у которых средний бал по предмету равен 3.

Список использованных источников

1. Интернет ресурсы: [https://ru.wikipedia.org/wiki/Математическая статистика](https://ru.wikipedia.org/wiki/Математическая_статистика)
2. Сокэл Р.Р. Кластерный анализ и классификация: предпосылки и основные направления. В кн: Классификация и кластер /Под ред. Дж.Вэн Райзина М: Мир, 1980.
3. Мандель И.Д. Кластерный анализ, 1988.
4. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика, 1988