

## ИССЛЕДОВАНИЕ ФИЗИОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ У БОЛЬНЫХ БРОНХИАЛЬНОЙ АСТМОЙ НА ОСНОВЕ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МЕТОДОВ

Бурцева А.Л.

*ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», Томск, Россия (634050, г. Томск, проспект Ленина, дом 30), e-mail: [anechkaby@mail.ru](mailto:anechkaby@mail.ru)*

В данной статье рассмотрен один из методов статистического анализа многомерных данных в медицине – кластерный анализ. Он позволяет найти сходства и различия между отдельными группами объектов и на основании этого выделить однородные группы. Методы многомерного анализа данных – это один из наиболее эффективных количественных инструментов для изучения процессов, описываемых большим числом характеристик. Для применения данного анализа на практике были рассмотрены различные программные продукты, и из огромного разнообразия статистических пакетов, разработанных для электронной обработки данных, в частности пакетов SAS, STATISTICA, Statgraphics, SPSS и т. д., для решения поставленной задачи был выбран пакет STATISTICA 8. Таким образом, использование кластерного анализа позволило определить группы пациентов с различными группами астмы и психогенной одышкой, которые схожи по структуре с исходным разбиением на четыре группы. Это может служить доказательством возможности разделения пациентов на четыре исходных группы.

Ключевые слова: бронхиальная астма (БА), STATISTICA 8, кластерный анализ, дендрограмма, классификация.

## THE STUDY OF PHYSIOLOGICAL CHARACTERISTICS OF PATIENTS WITH BRONCHIAL ASTHMA BASED ON THE BASIS MULTIVARIATE STATISTICAL METHODS

Burceva A.L.

*National Research Tomsk Polytechnic University, Russia (634050 Tomsk, Lenina Avenue, 30), e-mail: [anechkaby@mail.ru](mailto:anechkaby@mail.ru)*

This article discusses one of the methods of statistical analysis of multivariate data in medicine is cluster analysis. It allows to find the similarities and differences between individual groups of objects, and identify homogeneous groups. Methods of multivariate analysis is one of the most effective tool for studying the processes which are described by a variety of characteristics. For the application of this analysis were considered a variety of software products. Of the countless of electronic data processing packages, which are designed for electronic data processing, including packages SAS, STATISTICA, Statgraphics, SPSS solutions to this problem has been selected package STATISTICA 8. Thus, the use of cluster analysis allowed to determine groups of patients with different type of the asthma and group with dyspnea. These groups are similar by the structure with the original four groups. It is speaks about possibility of separating the initial four patients groups.

Keywords: bronchial asthma (BA), STATISTICA 8, cluster analysis, dendrogram, classification.

К.Г. Языков и Е.В. Немеров в процессе длительного динамического наблюдения анамнеза жизни выявили группу пациентов, у которых отмечалась высокая чувствительность к психотравмирующим жизненным ситуациям. В 2009 году они выдвинули гипотезу о том, что среди больных бронхиальной астмой существует определенная часть людей с особой психобиологической реактивностью. Дальнейшее обострение болезни у этих пациентов провоцирует психосоциальное стрессовое воздействие. В 2009–2012 годах они получили и опубликовали (в том числе в соавторстве с сотрудниками кафедры Прикладной математики ИК ТПУ) результаты, которые подтверждали данную гипотезу [1].

На основании результатов работы они предлагают новую классификацию бронхиаль-

ной астмы:

- БАПИ – бронхиальная астма психогенно-индуцированная;
- БАСП – бронхиальная астма сомато-психогенная;
- БАНП – бронхиальная астма непсихогенная;
- ПО – психогенная одышка.

**Цель данной работы:** выделение «типичных» групп больных из общей выборки, с помощью кластерного анализа.

**Объект исследования:** физиологические данные больных с различными формами астмы и группа больных с психогенной одышкой.

**Предмет исследования:** выявление особенностей в системе дыхания среди всей совокупности больных с целью разделения их на исходные группы.

**Материалы исследования:** данные традиционных показателей вентиляции легких и параметры механики дыхания для каждой из четырех групп.

Первая группа – БАПИ, исходно условно названная как БА психогенно индуцированная. В состав этой группы вошли пациенты, у которых первый приступ удушья развился после перенесенного эмоционального стресса. Последующее резкое ухудшение течения болезни было связано с какими-либо психологическими проблемами, имеющими негативный характер.

Вторая группа – БАНП – это группа больных с БА, исходно условно названной непсихогенной. В эту группу вошли лица с БА, главным образом, атопической формой заболевания. К обострению болезни, в свою очередь, приводили аллергия, вирусные инфекции, а также физические факторы (холод). Влияния психологических факторов не наблюдалось.

Третья группа (дополнительная) – БАСП, исходно условно обозначенная как БА сомато-психогенная. У пациентов данной группы «обычное» течение «обычной» болезни было нарушено жизненным стрессом, после которого психоэмоциональные триггеры (внешние раздражители) вызывали тяжелые приступы удушья, а также обострение болезни.

Четвертую группу, исходно условно отмеченную как «психогенная одышка» (ПО), образовали пациенты, направляемые к пульмонологу для исключения астмы, с жалобами на приступы удушья и одышку, связанные с психотравмирующими жизненными событиями, у которых многочисленными обследованиями была исключена бронхиальная обструкция и другие признаки астмы и органической патологии.

В исследовании принимали участие 83 пациента: БАПИ (24 человека), БАСП (18 человек), БАНП (29 человек), ПО (12 человек). Исходной информацией являются сведения о пациентах с одним из трех условных типов БА или с диагнозом ПО. Возраст пациентов от 15 до 56 лет.

В качестве исходных данных использовались не только традиционные показатели вентиляции легких, измеряемые методами спирографии и пневматихографии (зависят и от внелегочных причин, в том числе эмоциональных), но и показатели механики дыхания (отражают изменения внутрилегочных эластических и неэластических сопротивлений). Технология получения экспериментальных данных изложена в литературных источниках [2].

Оценивались такие традиционные показатели вентиляции легких, как: минутный объем дыхания (МОД), жизненная емкость легких (ЖЕЛ), объем форсированного выдоха (ОФВ-1), максимальная вентиляция легких (МВЛ), пиковая объемная скорость (ПОС), мгновенная объемная скорость (МОС-25, МОС-50, МОС-75) и параметры механики дыхания: общая работа дыхания при спонтанном дыхании ( $W_{\text{общ.}}$ ) и её фракции: неэластическая ( $W_{\text{н.эл.}}$ ) и эластическая ( $W_{\text{эл.}}$ ), удельная работа дыхания на литр вентиляции ( $W_{\text{уд.}}$ ), работа дыхания в условиях одинакового МОД, равного 10л/мин ( $W_{\text{МОД 10}}$ ), общая ( $W_{\text{МВЛ общ.}}$ ) и удельная ( $W_{\text{МВЛ уд.}}$ ), работа дыхания при МВЛ; а также динамическая ( $C_{\text{dyn.}}$ ) и статическая ( $C_{\text{stat.}}$ ) растяжимость легких и бронхиальное сопротивление на вдохе ( $R_{\text{вд.}}$ ) и на выдохе ( $R_{\text{выд.}}$ ), измеренные в условиях прерывания воздушного потока [1].

В данной работе, для решения поставленной задачи, была использована программа Statistica 8.0, возможности которой соответствуют требованиям цели работы и являются ориентированными на медицину. Система Statistica 8.0 может служить очень эффективным инструментом для научных исследований, особенно в области медицины. Statistica 8.0 является наиболее динамично развивающимся статистическим пакетом и по многочисленным рейтингам является мировым лидером на рынке статистического программного обеспечения [3].

**Метод исследования.** Для того чтоб добиться поставленной цели, необходимо прибегнуть к кластерному анализу, который предназначен для разбиения множества объектов на однородные группы (классы) на основании некоторого математического критерия качества классификации [4, 5].

Выбор количества классов может определяться на основе анализа специальных функционалов качества, на основе сравнения разбиений на различное количество классов, возможности из содержательной интерпретации и других критериев [4].

Каждый класс соответствует определенной группе, а наблюдения, попавшие в одну группу, характеризуются одинаковой вероятностью наступления приступа БА или ПО.

Поскольку изначально число классов неизвестно, то нужно обратиться к иерархическим кластер-процедурам. В пакете Statistica 8.0 реализованы агломеративные кластер-процедуры, т. е. каждое наблюдение начинается в собственном кластере, и пары кластеров объединяются по мере продвижения вверх по иерархии.

Исходные данные представлены матрицей «объект-свойство», строки которой пред-

ставляют объекты, а столбцы – характеризующие эти объекты признаки. Кластеризоваться в данном случае будут объекты – больные.

Поскольку нет информации о том, что какой-то признак более важен для классификации, чем остальные, то будем учитывать различия по каждому признаку. Выберем обычное евклидово расстояние, а в качестве алгоритма кластеризации выберем метод Уорда в этой серии экспериментов. Этот метод очень эффективен, т.к. он дает довольно компактные и хорошо разделенные кластеры.

**Результаты и их обсуждение.** Результаты иерархической классификации нагляднее всего представить в виде дендрограммы (рис. 1).

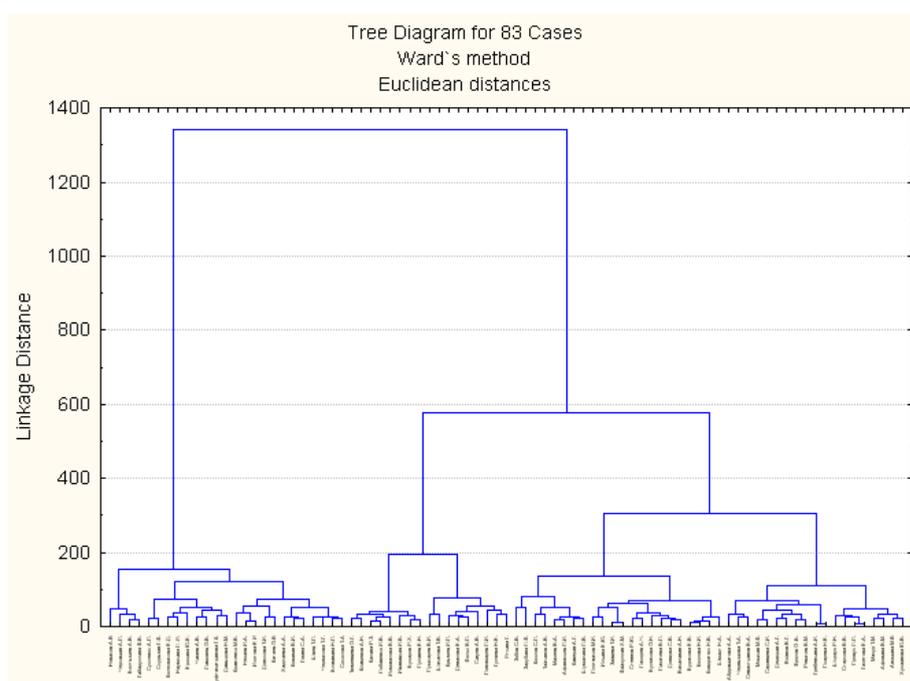


Рис.1. Дендрограмма, полученная методом Варда

Отделимость классов оценивается сравнением внутрикластерных и межкластерных расстояний на качественном уровне.

На рис. 1 видно, что всех пациентов можно разделить на 4 группы. В каждом из четырех кластеров находятся объекты со схожим диагнозом. Для того чтоб выбрать «правильное» количество групп, можно использовать пороговое расстояние. При пороговом расстоянии 200 выделяется 4 класса (рис. 1).

Для того чтоб подтвердить правильность выбора количества кластеров, воспользуемся графиком объединения объектов в классы.



Рис. 2. График объединения объектов в классы

На рис. 2 видно, что на графике находится точка «перелома» и номер шага  $m$ , на котором произошел «перелом»; тогда количество классов равно  $n-m$ , где  $n$  – количество объектов в выборке. В нашем случае в качестве точки перелома можно рассматривать шаг под номером 79, откуда получаем  $83 - 79 = 4$ , т.е. четыре класса.

Таким образом, мы ответили на первый вопрос – по схожести структуры больших целесообразно разделить на четыре класса.

Анализируя порядок следования пациентов сверху вниз на дендрограмме, полученной методом Варда, заметим, что пациенты разделились по классам следующим образом:

	A	B	C	D	E	F	G	H	I	J	K
1	Пашин	Группа	Кластер	1	Пашин	Группа	Кластер	1	Пашин	Группа	Кластер
2	Афон	VANP	4	Васили	VARI	3	Мошки	VASP	4		
3	Болты	VANP	1	Газето	VARI	4	Наушки	VASP	1		
4	Булга	VANP	3	Гигади	VARI	1	Новики	VASP	1		
5	Бурак	VANP	3	Гогин	VARI	1	Файзул	VASP	3		
6	Бухар	VANP	2	Давид	VARI	1	Черны	VASP	1		
7	Гребен	VANP	4	Дончо	VARI	3	Черны	VASP	4		
8	Дедю	VANP	4	Заруби	VARI	3	Чигаев	VASP	1		
9	Зайше	VANP	3	Изотоя	VARI	1	Шакал	VASP	2		
10	Зобов	VANP	3	Кагаев	VARI	1	Шкраул	VASP	4		
11	Ильин	VANP	3	Козлов	VARI	3	Сигуев	PD	2		
12	Кавер	VANP	3	Псарен	VARI	4	Денис	PD	2		
13	Козло	VANP	3	Ревиче	VARI	4	Иванчи	PD	2		
14	Колес	VANP	1	Саранд	VARI	4	Исмаги	PD	2		
15	Кривч	VANP	1	Севаст	VARI	4	Килина	PD	2		
16	Лазор	VANP	3	Сокол	VARI	1	Климен	PD	2		
17	Лихач	VANP	3	Старк	VARI	4	Крыме	PD	2		
18	Мазур	VANP	4	Тайчи	VARI	3	Лобан	PD	2		
19	Миге	VANP	3	Халжа	VARI	1	Поном	PD	2		
20	Муфт	VANP	1	Хусан	VARI	4	Пушка	PD	2		
21	Нехае	VANP	1	Шмици	VARI	3	Тюмен	PD	2		
22	Плотн	VANP	3	Шуми	VARI	3	Фольк	PD	2		
23	Проиц	VANP	4	Ятзин	VARI	2					
24	Пулин	VANP	2	Абдра	VASP	4					
25	Скури	VANP	1	Благин	VASP	1					
26	Стегни	VANP	3	Войно	VASP	4					
27	Сусле	VANP	1	Гуляев	VASP	2					
28	Фролд	VANP	4	Донск	VASP	1					
29	Черка	VANP	1	Ерохин	VASP	1					
30	Швец	VANP	1	Кондр	VASP	3					
31	Анош	VARI	4	Конин	VASP	1					
32	Афан	VARI	3	Лимо	VASP	1					

Рис. 3. Таблица распределения пациентов по кластерам

Получим средние значения признаков в каждом из выделенных классов, как показано

на рис. 4, и определим, чем же классы отличаются друг от друга.

Variable	Кластер			
	1	2	3	4
МОД1	8,34760	8,4324	8,03818	8,65526
ЖЕЛ1	67,25320	100,5424	92,13636	81,98421
ФЖЕЛ1	60,00760	97,0188	87,68909	73,63158
ОФВ1_1	54,04520	93,3894	81,11364	66,90000
ОФВ1/ЖЕП 1	69,83520	91,5141	83,43500	77,39000
МВЛ1	46,56360	86,1194	83,47182	62,14211
ПОС1	37,75360	79,6353	64,55773	54,23158
МОС25_1	36,22160	77,4782	60,94091	52,05789
МОС50_1	30,80640	70,1353	52,92273	46,09474
МОС75_1	26,45880	65,0294	44,94545	43,00000
Cdyn1	0,09164	0,1494	0,10759	0,11295
Cstat1	0,13236	0,1689	0,15100	0,13400
Rвд1	5,59040	3,0724	4,58000	4,77105
Rвыд1	7,35720	3,7629	6,08682	6,51526
Wобщ1	0,71626	0,3876	0,46509	0,50611
Wуд1	0,08628	0,0467	0,05605	0,05989
Wн.эл1	0,65544	0,3136	0,41200	0,42963
Wэл1	0,33360	0,2097	0,23586	0,24016
W МОД10_1	0,84724	0,4538	0,56309	0,57342
W МОД15_1	1,41960	0,7390	0,98641	0,97332
W МВЛобщ1	15,15604	20,0155	26,52432	18,33763
WМВЛуд1	0,32116	0,2360	0,30495	0,29758

Рис. 4. Вид файла с данными по средним значениям признаков в классах

Анализ этих данных позволяет нам дать интерпретацию классам и, наконец, выяснить, что же за классы были нами выделены.

В первый кластер попали пациенты с самыми высокими значениями параметров дыхания (кроме динамической и статической растяжимости легких) и низкими традиционными показателями вентиляции легких. Для второго кластера характерно высокое значение традиционных показателей (кроме МОД) и низкое значение параметров механики дыхания (кроме динамической и статической растяжимости легких). Третий и четвертый кластер не имеют четко выраженных особенностей (табл.1).

Таблица 1. Распределение по кластерам в процентном соотношении

Группа/Кластер	1	2	3	4
БАСП	50%	11,10%	11,10%	27,80%
БАПИ	29,20%	4,20%	33,30%	33,30%
БАНП	31%	6,90%	41,40%	20,70%
ПО	0	100%	0	0

В 1-м кластере преимущественно оказались пациенты с диагнозом БАСП и БАНП, а также 7 человек с БАПИ. Во второй кластер попали все больные с ПО, по 2 человека с диагнозом БАСП и БАНП и 1 человек с БАПИ. Третий кластер объединил преимущественно людей с диагнозом БАНП. Также сюда попали 2 человека с БАСП, 8 человек с БАПИ. В четвертом кластере преимущественно оказались люди с БАПИ, а также 4 с БАСП и 5 с БАНП.

Можно выдвинуть предположение, что:

1 Кластер – БАСП;

2 Кластер – ПО;

3 Кластер – БАНП;

4 Кластер – БАПИ.

**Заключение.** Таким образом, использование кластерного анализа позволило определить группы пациентов с БА и ПО, которые схожи по структуре с исходным разбиением на четыре группы. Это может служить доказательством возможности разделения пациентов на четыре исходных группы.

*Работа выполнена в рамках проекта №1957 Госзадания «Наука» Министерства образования и науки РФ.*

### Список литературы

- 1 Немеров Е.В., Языков К.Г. К вопросу изучения личностных свойств в психофизиологической реактивности больных бронхиальной астмой на аудиовизуальную стимуляцию // Вестник ТГПУ. – 2011. – Вып. 6 (108). – С. 134 – 137.
- 2 Платонов, А.Е. Статистический анализ в медицине и биологии: задачи, терминология, логика, компьютерные методы. – М.: Изд-во РАМН, 2000. – 52 с.
- 3 Красильников В.В., Тоискин В.С. Математические методы в психолого-педагогических исследованиях: учебно-методическое пособие. – Ставрополь: Изд-во СГПИ, 2008. – 84 с.
- 4 Берестнева, О.Г. Компьютерный анализ данных: учеб. пособие / О.Г. Берестнева, Е.А. Муратова, А.М. Уразаев. – Томск: Изд-во ТПУ, 2003. – 204 с.
- 5 Воловоденко В.А., Берестнева О.Г., Немеров Е.В., Осадчая И.А. Применение методов визуализации при исследовании структуры экспериментальных многомерных данных // Известия Томского политехнического университета. – 2012. – Т. 320, № 5: Управление, вычислительная техника и информатика. – С. 125-130.

### Рецензенты:

Жигинас Н.В., д.псх.н., профессор, заведующий кафедрой психологии развития личности ФГБОУ ВПО Томский государственный педагогический университет (ТГПУ), г. Томск.

Уразаев А.М., д.б.н., профессор, профессор Института теории образования, ФГБОУ ВПО Томский государственный педагогический университет (ТГПУ), г. Томск.