

УДК 616/618+616.8

Мартиросян В.В., Крупская Ю.А.

**ИСПОЛЬЗОВАНИЕ КЛАСТЕРНОГО АНАЛИЗА
В ОЦЕНКЕ СТАТИСТИЧЕСКИ ЗНАЧИМЫХ СВЯЗЕЙ МЕЖДУ КОМПЛЕКСОМ
ЭКЗОГЕННЫХ ФАКТОРОВ И ХАРАКТЕРИСТИКАМИ ОСТРОГО НАРУШЕНИЯ
МОЗГОВОГО КРОВООБРАЩЕНИЯ**

ГБОУ ВПО Ростовский Государственный медицинский университет, г. Ростов-на-Дону

В данном исследовании подтверждается гипотеза о наличии статистически значимых связей между комплексом экзогенных факторов (GG-признаки) и признаками, отражающими характеристики острого нарушения мозгового кровообращения. На основании кластерного анализа произведена классификация, позволяющая описать дескриптивные характеристики каждого кластера с генерацией новой классификационной переменной CLUST_GG. Предложены 12 экзогенных факторов из числа GG-признаков и 9 признаков из числа эндогенных факторов риска, которые статистически значимо взаимосвязаны с зависимой переменной CLUST_GG.

Ключевые слова: инсульт, кластерный анализ, логистическая регрессия, солнечная активность.

Мартиросян В.В., Крупская Ю.А.

**ИСПОЛЬЗОВАНИЕ КЛАСТЕРНОГО АНАЛИЗА В ОЦЕНКЕ СТАТИСТИЧЕСКИ
ЗНАЧИМЫХ СВЯЗЕЙ МЕЖДУ КОМПЛЕКСОМ ЭКЗОГЕННЫХ ФАКТОРОВ
И ХАРАКТЕРИСТИКАМИ ОСТРОГО НАРУШЕНИЯ МОЗГОВОГО
КРОВООБРАЩЕНИЯ**

In the given research the hypothesis about presence statistically significant relationships between a complex of exogenous factors (GG-attributes) and attributes reflecting the characteristics of acute cerebrovascular accident proves to be true. On the basis of the cluster analysis the classification allowing to describe the descriptive characteristics of each cluster with generation new classification variable CLUST_GG is made. Are offered 12 of the exogenous factors from among GG-attributes and 9 attributes from number of endogenous risk factors, which statistically significantly are interconnected with dependent variable CLUST_GG.

Key words: cerebral stroke, cluster analysis, logistic regression, solar activity.

В формировании сосудистых заболеваний головного мозга прослеживается влияние многих факторов риска. Среди эндогенных это - наследственность, пол, возраст, артериальная гипертензия, атеросклероз, ожирение, среди экзогенных - эмоциональный стресс, гиподинамия, вредные привычки, метеогелиогеофизические факторы. [10]

В данной работе рассмотрена гипотеза о наличии статистически значимых связей между комплексом экзогенных факторов (GG-признаки) и признаками, отражающими характеристики заболеваний, в частно-

сти, характеристиками острого нарушения мозгового кровообращения (ОНМК).

Цель данного исследования:

1. Применив статистический метод кластерного анализа, получить обобщённый качественный признак, характеризующий локализацию комбинаций значений GG-признаков в отдельные группы наблюдений (кластеры);
2. Благодаря выделению таких группировок наблюдений, изучить их возможные связи со всем набором исходных качественных и количественных характеристик признаков ОНМК.

Материалы и методы

В основу исследования положены результаты ретроспективного анализа протоколов вскрытий лиц, умерших от мозговых инсультов (МИ) (всего - 1135 случаев) за годы высокой солнечной активности (СА) (2000-2002 гг.) и годы низкой СА (2008-2010 гг.) по материалам БСМП №1 г. Ростова-на-Дону, куда в основном госпитализировались лица с МИ.

Все случаи инсульта подразделялись по полу (мужчины, женщины), по возрасту (молодой (до 45 лет), средний (45-59), пожилой (60-74), старческий (75-89), долгожители (90 и старше)), по типу (ишемические (ИИ), геморрагические (ГИ) инсульты), по локализации (правое, левое полушарие и ствол головного мозга).

Ишемические инсульты в свою очередь подразделены на атеротромботические (АТБ) и неатеротромботические (неАТБ).

Геморрагические инсульты классифицированы на субарахноидальные, геморрагическое пропитывание, гематомы и гематомы с прорывом крови в желудочки мозга.

По каждому наблюдению произведена фиксация даты начала МИ и даты летального исхода. Произведен подсчет индивидуального срока дожития, определены случаи летальности в 1-е сутки. Учтён фактор внезапности развития МИ (внезапно, постепенно).

Из факторов риска (ФР) развития инсульта учитывались: возраст (согласно классификации ВОЗ), артериальная гипертензия, атеросклероз сосудов головного мозга, сахарный диабет, заболевания сердца, ОНМК в анамнезе, заболевания почек, злоупотребление алкоголем.

Была произведена оценка комплексного воздействия факторов окружающей среды, на заболеваемость ОНМК, за указанные периоды с учетом СА. В анализе использовались 55 признаков экзогенных факторов (ЭФР):

- Подгруппа SA - всего 21 признак. Основные планетарные гелиофизические индексы, характеризующие СА, в т.ч.: визуальные индексы СА, Солнечный поток рентгеновского излучения, индексы, характеризующие вспышечную активность, Сол-

нечные компоненты космических лучей, параметры Солнечного ветра, компоненты межпланетного магнитного поля. Данные гелиофизических показателей получены из каталогов SpaceEnvironmentCenter (<http://sec.noaa.gov>) и NationalSpaceScience DataCenter (<http://nssdc.gsfc.nasa.gov>).

- Подгруппа GM - всего 6 признаков. Классические планетарные геофизические индексы. В исследовании использовалась группа индексов (имеющих среднее значение за одни сутки), которые рекомендуются для поиска связей геомагнитной активности с процессами и явлениями (в т.ч. медицинскими), происходящими на Земле. [3]

Ряды геофизических индексов получены из каталогов SPIDR (Space Physics InteractiveDataResource), которые находятся по адресу <http://spidr.ngdc.noaa.gov/spidr>.

- Подгруппа MD - всего 19 признаков. Среднесуточные значения 5-ти основных метеорологических показателей в г. Ростов/Дону за 2000-2002 и 2008-2010 гг., которые включают в себя: температуру воздуха, относительную влажность, атмосферное давление, скорость ветра и облачность. Также были рассчитаны их суточные амплитуды и их межсуточное изменение, среднесуточные значения местной Солнечной радиации (инсоляция).

Использовался архив погоды в Ростове-на-Дону за 2000-2002 и 2008-2010 гг. (метеостанция "Ростов-на-Дону"). Атм. давление на высоте метеостанции 75м над уровнем моря, частота измерений - каждые 3 часа. Каталоги архива находятся по адресу <http://rostovmeteo.ru/>

- Подгруппа IP - всего 9 признаков. Суточные значения суммарного метеорологического индекса патогенности и его компонент [6]. Расчет индекса производился на основании рекомендаций "Руководства по специализированному климатическому обслуживанию экономики" [5].

За период 2000-2010 гг. получена графическая демонстрация солнечной активности (СА) в относительных числах солнечных пятен (ISN) и значений потока радиоизлучения на волне 10.7 см (SFU) (Рисунок 1)

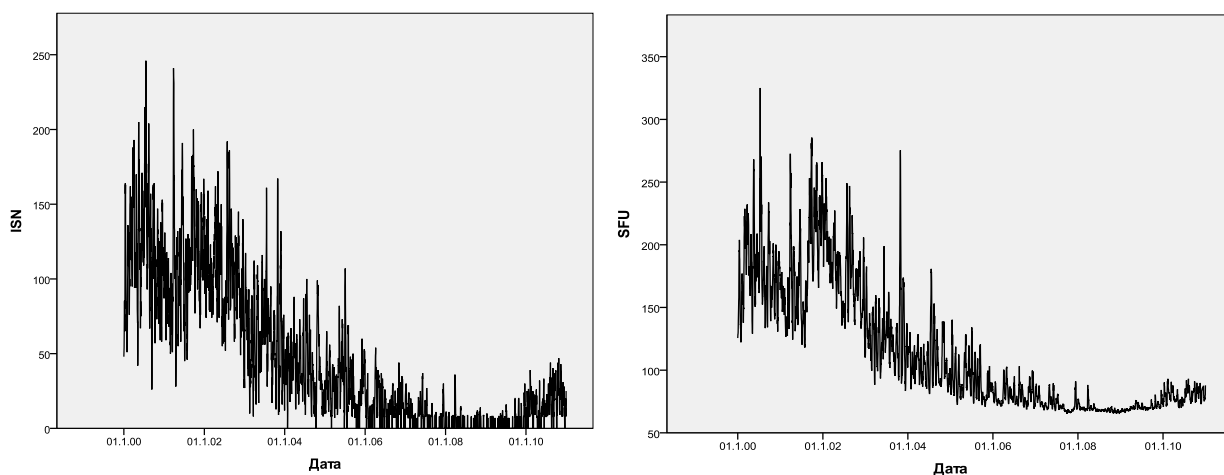


Рис 1. Графическая демонстрация солнечной активности (СА) за период 2000-2010 гг. в относительных числах солнечных пятен (ISN) и значений потока радиоизлучения на волне 10.7 см (SFU)

Для подтверждения либо опровержения исходной гипотезы, исследование решено ограничить периодами 2000-2002 гг. - года высокой СА и 2008-2010 гг. - года низкой СА.

Статистический анализ данных выполнен в Центре БИОСТАТИСТИКА под руководством доцента факультета информатики Томского государственного университета, к.т.н., Леонова В.П. Процедуры статистического анализа выполнялись с помощью статистических пакетов SAS 9.2, STATISTICA 10 и SPSS-20. Критическое значение уровня статистической значимости при проверке нулевых гипотез принималось равным 0,05. В случае превышения достигнутого уровня значимости статистического критерия этой величины, принималась нулевая гипотеза.

Массив данных содержал 2261 наблюдение, включавших 81 признак. По массиву проводился анализ, основной целью которого было выявление латентных групп наблюдений с последующим созданием новых группирующих признаков. С этой целью проводился кластерный анализ с использованием всех количественных признаков. Учитывая существенную разномасштабность шкал анализируемых признаков, все количественные признаки перед проведением кластерного анализа предварительно стандартизовались. В результате этого все признаки в новых шкалах имели нулевое среднее и единичное стандартное

отклонение. Для проведения кластерного анализа с этими новыми признаками вначале использовали иерархический алгоритм кластеризации. Процедура иерархического кластерного анализа, с построением дендрограмм, позволяла визуально определить количество выделяемых кластеров, идентификация которых производилась далее с помощью метода k-средних. В обоих случаях использовали алгоритм Варда и Евклидову метрику [8,9].

Проверка нормальности распределения количественных признаков в отдельных группах сравнения проводилась с использованием критериев Колмогорова-Смирнова Шапиро-Уилка, Крамера-фон-Мизеса и Андерсона-Дарлинга. Алгоритм принятия решения о нормальности распределения принимался при совпадении положительных результатов по не менее трем применяемым критериям. Для сравнения центральных параметров групп использовались параметрические и непараметрические методы: дисперсионный анализ, в т.ч. с критерием Краскела-Уоллиса и ранговыми метками Вилкоксона, медианный критерий и критерий Ван-дер-Вардена [1,8]. При неоднозначных выводах за основу брался результат, полученный посредством применения критерия Ван-дер-Вардена. Для всех количественных признаков в сравниваемых группах производилась оценка средних

арифметических и среднеквадратических (стандартных) ошибок среднего, а также коэффициента вариации. Deskриптивные статистики в тексте представлены как $M \pm m$, где M - среднее, а m - ошибка среднего.

Исследование взаимосвязи между парами дискретных качественных признаков проводилось с использованием анализа парных таблиц сопряжённости. Помимо оценок критерия Пирсона Хи-квадрат и достигнутого уровня статистической значимости этого критерия, вычислялись и оценки интенсивности связи анализируемых признаков, такие как коэффициент Фи, коэффициент контингенции и V-коэффициент Крамера. [1,7] В случае наличия статистически значимых связей между парой качественных признаков проводился углублённый анализ распределения частот в клетках таблицы сопряжённости, позволявший установить структуру выявленной взаимосвязи на уровне сочетаний отдельных градаций обоих признаков.

Для анализа взаимосвязи между одним качественным признаком, выступающим в роли зависимого, результирующего показателя, и подмножеством количественных и качественных признаков, использовалась модель логистической регрессии с пошаговым алгоритмом включения и исключения предикторов. Результаты оценки уравнений логистической регрессии представлены набором коэффициентов регрессии, достигнутыми уровнями значимости для каждого коэффициента, а также оценкой показателя согласия (Concordant) фактической принадлежности пациента к той или иной из групп, и теоретической принадлежности, полученной по уравнению логит-регрессии. Всего было получено порядка ста уравнений логит-регрессии, из которых производился отбор уравнений имеющих самые высокие значения этого показателя. Ранжирование выделенных предикторов по степени связи с зависимой переменной производилось путём сортировки предикторов по модулю стандартизованных коэффициентов регрессии [7] В связи с отсутствием дополнительного массива данных кросс-проверка не проводилась. Качество моде-

ли оценивалось по моделирующей выборке. Учитывая значительный объём наблюдений (более 2-х тысяч), можно декларировать высокую устойчивость полученных оценок показателя Concordant.

Результаты и обсуждение

Для выделения на основе собранных данных GG-признаков - ЭФР (т.е. подгрупп SA, GM, MD и IP), ряда реально существующих взаимосвязей признаков и порождения на этой основе типологии использовалась процедура кластерного анализа.

Задача кластерного анализа состоит в разбиении исходной совокупности объектов на группы схожих, близких между собой объектов. Эти группы называются кластерами. Объединение объектов в кластеры происходит на основе вычисляемой меры близости. Другими словами, кластерный анализ - это один из способов кластеризации объектов по их признакам. Из множества алгоритмов кластеризации нами избран иерархический кластерный анализ и кластеризация методом k-средних.

Учитывая тот факт, что часть GG-переменных имеет как весьма маленькие значения, порядка $5,70E-07$, так и весьма большие значения, порядка $5,90E+06$ и $1,30E+08$, очевидно, что при оценке межкластерных расстояний будут доминировать переменные имеющие эти большие значения. Таким образом, истинная структура кластеров будет искажаться за счёт неравных весов переменных разных масштабов. Поэтому изначально была произведена обычная процедура стандартизации всех количественных переменных, сделав для каждой переменной средней значение равное 0, и стандартное отклонение сделав равным 1. После чего был проведён иерархический кластерный анализ с использованием набора количественных GG-признаков, т.е. подгрупп SA, GM, MD и IP. Ниже приведён график с дендрограммой, полученной с использованием новых, стандартизованных переменных. (Рис. 2)

Как видим, достаточно большее количество наблюдений и кластеров имеют ненулевые расстояния между собой, что выражается в наличии большего количества

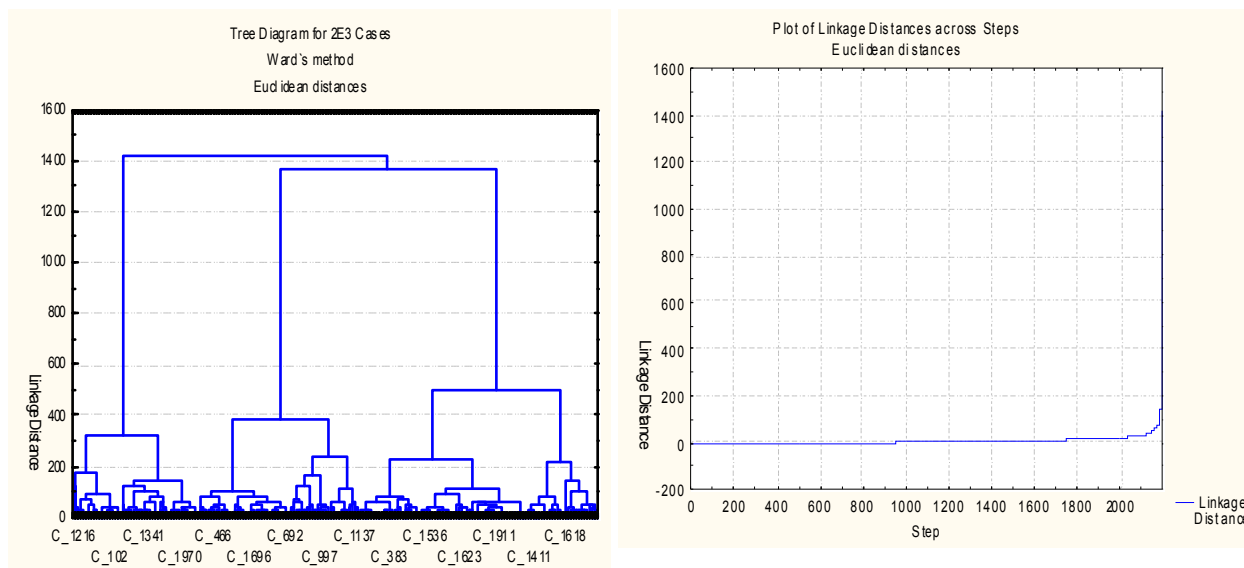


Рис. 2 Графическая демонстрация дендрограммы, полученной с использованием стандартизованных переменных.

вертикальных ветвей в нижней части дендрограммы. Если разрезать полученную дендрограмму примерно на уровне 600 единиц, что более чем в 2 раза больше межкластерного расстояния на последнем шаге, то получим 3 ветви. Качество полученной дендрограммы оценивается визуально с помощью графика, приведённого справа. На графике отражено изменение межкластерного расстояния на каждом шаге. Чем ближе полученная кривая приближается своей формой к прямому углу, тем выше качество. Т.е. чем однороднее выделенные группировки наблюдений внутри каждого кластера, тем больше различие этих группировок по своим дескриптивным показателям между собой. Судя по графику, приведённому выше, качество кластеризации отличное.

После оценки числа кластеров методом иерархической классификации использовался метод k-средних для выделения данных кластеров с генерацией новой классификационной переменной `CLUST_GG`, содержащей значения номера кластера, к которому относится каждое конкретное наблюдение. Название `CLUST_GG` означает, что процесс кластеризации основан на использовании количественных переменных, отражающих GG - признаки.

Ниже на рисунке показаны расположения средних значений по каждому GG-при-

знаку для 3-х кластеров, где кривая каждого кластера обозначена ломаной кривой одного цвета. В левой части рисунка видно, что примерно до признака `VAR42= 'компонента Vz'` вместо 3-х кривых имеем 2-е кривые, т.к. для кластеров 1 и 2 средние практически идентичны. Отметим, что наименьшие различия в средних у 3-х кластеров наблюдается для признака `VAR37= 'ProtonFluence > 100 MeV'`. По мере продвижения в правую часть рисунка наблюдается разделение кривых на три отдельных линии. (Рисунок 3)

На следующем этапе исследования осуществлен поиск взаимосвязи между классификационной переменной `CLUST_GG` и GG-признаками (подгрупп SA, GM, MD и IP) участвующими в анализе. Целью многомерного анализа было выявить факторы, из 55 GG-признаков, которые несут основную ответственность в объяснении различий 3-х кластеров. Для установления статистически значимых связей использовали метод логистической регрессии.

В результате проведенного исследования было выявлено 12 экзогенных факторов из числа GG-признаков, что свидетельствует о том, что различие между структурой 3-х кластеров разделяется достаточно большим сектором показателей.

Совпадение факта и расчета по градациям классификационной переменной

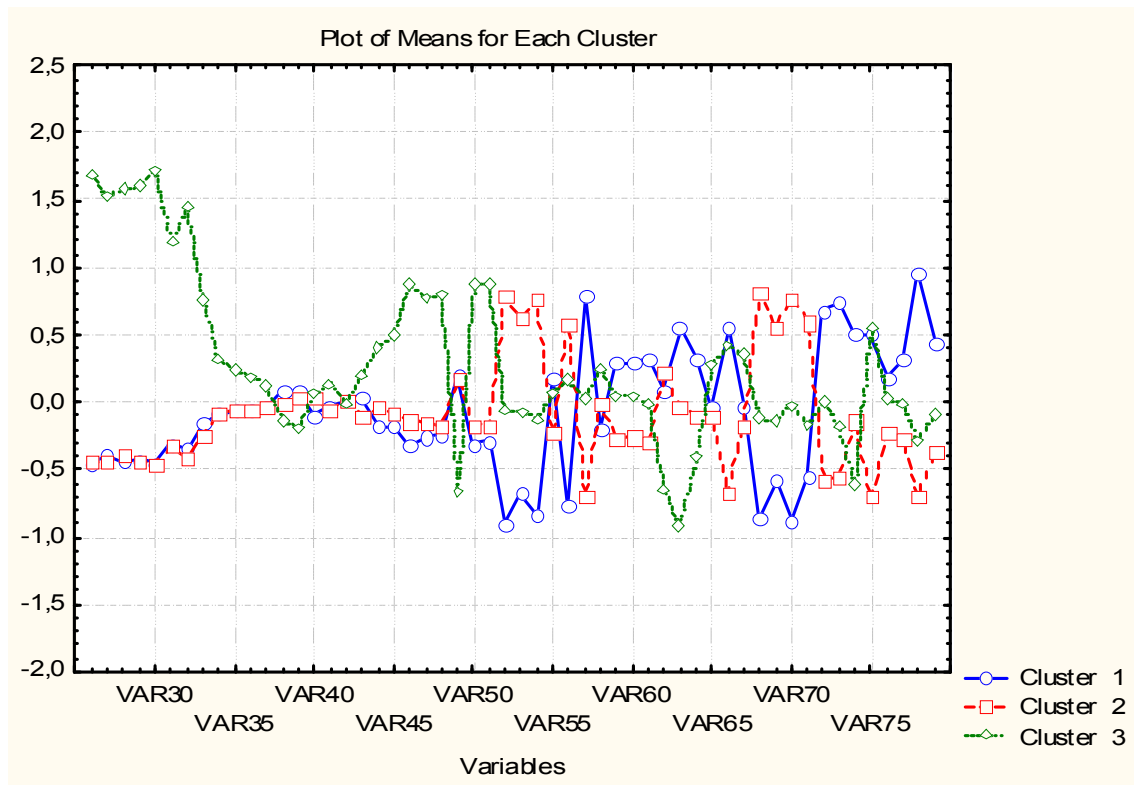


Рис. 3. Графическая демонстрация расположения средних значений по каждому GG-признаку для 3-х кластеров

CLUST_GG признаков "кластер 1" и "кластер 2" и "кластер 3" в группах исследования было обеспечено в 97,6% случаев (Concordant), коэффициент связи D-Зоммера (Somers'D) составили 0.952. Полученные данные представлены в таблице №. 1

Произведена оценка дескриптивных статистик (среднее, ошибка среднего, коэффициент вариации) для всех 12 выявленных количественных GG-признаков отдельно в подгруппах, образованных признаком CLUST_GG, для интерпретации каждого кластера на предмет того, какие из этих GG-признаков в том или ином кластере (группировке наблюдений) принимали максимальные или минимальные значения. Выделены наборы количественных признаков статистически значимо взаимосвязанных с признаком CLUST_GG.

На завершающем этапе исследования проанализирована, с помощью метода логистической регрессии, взаимосвязь между зависимой переменной CLUST_GG и всеми качественными и количественными признаками, отражающими характеристики заболеваний ОНМК.

В целях поиска "пускового" механизма к участию в анализе были подключены эндогенные факторы. В результате проведенного исследования (методом логистической регрессии) было выявлено 9 признаков из числа ФР, которые статистически значимо взаимосвязаны с зависимой переменной CLUST_GG.

Совпадение факта и расчета по градациям классификационной переменной CLUST_GG признаков "кластер 1" и "кластер 2" и "кластер 3" в группах исследования было обеспечено в 60,3% случаев (Concordant), коэффициент связи D-Зоммера (Somers'D) составил 0.212. Полученные данные представлены в таблице №.2

Для интерпретации структуры логит-линейного уравнения использовались результаты анализа таблиц сопряженности и результаты сравнения центральных мер для групп зависимого признака CLUST_GG и выявленных признаков из числа ФР.

Интерпретация результатов.

На основании произведенного кластерного анализа и анализа влияний комбинаций GG-признаков на признаки, отражающие харак-

Таблица 1

№	Предиктор	Коэффициент (В)	Стд. Ошибка	Статистика Вальда	Pr>ChiSq
	Intercept кластер 1 Intercept кластер 2	-35.8161 -29.1271	13.6977 13.6851	6.8369 4.5300	0.0089 0.0333
1	ISN совокупное число солнечных пятен	-0.0208	0.00734	8.0172	0.0046
2	SFU поток солн.излуч. част -1 2800 МГц	-0.0622	0.00997	38.8886	<.0001
3	X-RayBkgdFlux сут рентг поток 1-8 А	-2099763	0.041	25.9984	<.0001
4	Flares X-Ray класс С малые вспышки	-0.3058	0.0431	50.3475	<.0001
5	DST осн. хар-ка геомагнитн. Бури и Т	1.2587	0.2667	22.2706	<.0001
6	Ap планет. средн. инд. геомагн. акт-ти	-0.2669	0.0391	46.5792	<.0001
7	Ср_сутТ температура	0.1060	0.0263	16.2568	<.0001
8	Ср_сут V	0.3663	0.1331	7.5717	0.0059
9	Индекс патогенности Т	0.1513	0.0484	9.7899	0.0018
10	Индекс патогенности V	-0.2253	0.0652	11.9376	0.0006
11	Суммарный индекс патогенности I	0.1830	0.0289	40.1622	<.0001
12	Индекс Солнечной активности	3.7259	0.6027	38.2230	<.0001

Таблица 2

№	Предиктор	Коэффициент (В)	Стд. Ошибка	Статистика Вальда	Pr > ChiSq
	Intercept кластер 1 Intercept кластер 2	-3.6114 -1.6774	0.4028 0.3962	80.3760 17.9207	<.0001 <.0001
1	Полных лет	0.0487	0.0101	23.3052	<.0001
2	Пол	-0.1550	0.0825	3.5268	0.0604
3	Возрастная группа	-0.4325	0.1425	9.2164	0.0024
4	Инсульг в анамнезе	0.3359	0.1597	4.4237	0.0354
5	Тип МИ	0.2576	0.1018	6.3982	0.0114
6	Выживаемость дней	-0.0215	0.00476	20.4343	<.0001
7	Мерцательная аритмия	0.3487	0.1425	5.9854	0.0144
8	Вторичная артериальная гипертензия	0.3732	0.1258	8.7964	0.0030
9	ИБС	0.5442	0.1305	17.3938	<.0001

теристики заболевания ОНМК произведена классификация, позволяющая описать дескриптивные характеристики каждого кластера:

Кластер №1 (812 наблюдений).

Состоит из:

- группы GG-признаков, которые объединены характерными отличительными признаками:

- Низкий совокупный индекс Солнечной активности. Преимущественный тип погоды по классификации В.Г. Бокша - острый. Высокое значение среднесуточной скорости ветра, низкая среднесуточная температура (менее 0 С), как следствие - высокое значение совокупного индекса патогенности в основном за счет своих

компонент: индекса патогенности температуры и скорости ветра.

- группы пациентов, умерших от МИ, которые объединены характерными отличительными признаками: мужской пол, тип МИ - геморрагический (класс: субарахноидальные, геморрагическое пропитывание и гематомы), как правило, сопутствующие факторы риска: пожилой возраст, артериальная гипертензия и ИБС. Атеросклероз сосудов головного мозга 3-4 ст.

Кластер №2 (923 наблюдений).

Состоит из:

- группы GG-признаков, которые объединены характерными отличительными признаками:

- Низкий совокупный индекс Солнечной активности. Преимущественный тип погоды по классификации В.Г. Бокша - оптимальный. Высокое значение среднесуточной температуры (свыше $18,5^{\circ}\text{C}$), которая характерна своим постоянством (наименьшие значения межсуточного изменения температуры) в совокупности с низким значением среднесуточной облачности, как следствие - высоким значением Солнечной радиации: toa_dwn - инсоляция верхних слоев атмосферы.

- группы пациентов, умерших от МИ, которые объединены характерными отличительными признаками: женский пол, тип МИ - ишемический (класс: неатеротромботические), как правило, сопутствующие факторы риска: пожилой либо старческий возраст, инсульт в анамнезе и мерцательная аритмия, атеросклероз сосудов головного мозга 3-4 ст.

Кластер №3 (464 наблюдений).

Состоит из:

- группы GG-признаков, которые объединены характерными отличительными признаками:

- Высокий совокупный индекс Солнечной активности, как следствие - большое значение индексов: ISN - (International Sunspot Number), совокупное число наблюдаемых солнечных пятен; SFU - поток солнечного радиоизлучения на частоте 2800 МГц (10.7 см); X-RayBkgdFlux - суточное фоновое значение рентгеновского потока;

C-класс вспышки - общее количество малых рентгеновских вспышек наблюдаемых в течение дня; Ap - планетарный среднесуточный индекс. Низкое значение Dst-эквивалентный индекс экваториального магнитного возмущения, что характерно для магнитных бурь. Преимущественный тип погоды по классификации В.Г. Бокша - раздражающий.

- группы пациентов, умерших от МИ, которые объединены характерными отличительными признаками: мужской и женский пол присутствует достаточно однородно, при этом - более молодой возраст (среднестатистический возраст на 5 лет моложе, чем в иных группах), наиболее жестокий тип МИ - геморрагический (класс гематомы с прорывом крови в желудочки мозга), либо ишемический (класс: атеротромботический). Как правило, сопутствующие факторы риска отсутствуют. Атеросклероз сосудов головного мозга не выражен.

Вывод

Представленные выше данные позволяют сформулировать следующие основные выводы:

1. Посредством применения статистического метода кластерного анализа удалось классифицировать участвующие в анализе экзогенные факторы риска (GG признаки) на 3 типовые группы. В первой группе сконцентрировались GG признаки, несущие ответственность за формирование острого типа погоды (по классификации В.Г. Бокша), во второй группе - оптимального типа погоды, в третьей группе сгруппированы характерные признаки для высокой геогелиомагнитной активности.
2. Выявлено четко выраженное разделение лиц, умерших от ОНМК, по признакам, отражающим характеристики заболевания ОНМК, а также признакам, характеризующим тяжесть течения заболевания, которые в свою очередь демонстрируют статистически значимую взаимосвязь ($p < 0.05$) с уникальными комплексами GG признаков, классифицированными по 3-м типологическим группам.
3. Необходимо отметить, что при обработ-

ке собранного массива данных кластерный анализ оказался мощным методом, позволяющий прийти к выводам, к которым было бы затруднительно прийти, применяя иные методы статистического анализа, не связанные с классификацией. Данный метод непараметрической статистики вполне можно использовать в клинической медицине как составную часть системного анализа, если есть необходимость оценки комплексного влияния нескольких факторов (клинических, биологических, погодных и др.) на тяжесть течения заболевания, эффективность лечения и т.д.

4. Сформулированная в начале статьи гипотеза о наличии статистически значимых связей между комплексом экзогенных факторов (GG-признаки) и признаками, отражающими характеристики заболевания, (ОНМК) подтверждена.

Выражаем благодарность заведующему патологоанатомическим отделением Ю.Г.Кириченко, заведующей неврологическим отделением О.Г.Стокановой ГБ №1 г. Ростов-на-Дону за предоставленные материалы исследований, доценту факультета информатики Томского государственного университета, к.т.н., В.П. Леонову за качественно проведенный статистический анализ данных.

ЛИТЕРАТУРА:

1. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. - М.: Мир, 1982. - 488 с.
2. Воропай Н.Г. Клинические и конституциональные особенности пациентов с ишемическими инсультами в метеорологических условиях г Новосибирска Дис. канд. мед. наук Новосибирск 2003. - 199 с.
3. Заболотная Н.А. Индексы геомагнитной активности: Справочное пособие. Изд.2 2007. - С - 88.
4. Иерусалимский А.П., Фейгин В.Л., Альперин Л.Б. Опыт прогнозирования и факторы риска мозгового инсульта в Новосибирске // Журнал неврологии и психиатрии 1988. № 1.- С 44-48.
5. Кобышева Н.В. Руководства по специализированному климатическому обслуживанию экономики // Главная геофизическая обсерватория им. А.И. Воейкова.// - СПб., 2008.- С. 336.
6. Латышев Г.Д., Бокша В.Г. К вопросу медицинской оценки погод (индекс погоды и реакция больных) // Вопросы курортологии. - 1965 - № 4. - С. 345-351.
7. Леонов В.П. Логистическая регрессия в медицине и биологии. URL: http://www.biometrica.tomsk.ru/logit_1.htm.
8. Ллойд Э., Ледерман У., Тюрин Ю.Н. Справочник по прикладной статистике. В 2-х т. Т.1: Пер. с англ./ Финансы и статистика, 1989; - 510с. Т. 2: 1990; - 526 с.
9. Мьюллер Ч.У., Дж.-О. Ким, Клекк У.Р. и др. Под ред. И.С. Енюкова Факторный, дискриминантный и кластерный анализ: Пер. с англ./ - М.: Финансы и статистика, 1989. - 215 с.
10. Трошин В.Д., Густов А.В., Трошин О.В. Острые нарушения мозгового кровообращения Н. Новгород, 2000.- 440 с.