А. Н. Спирин, В. Н. Солнцев, В. Е. Назаров

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ СТЕПЕНИ ОДНОРОДНОСТИ БОЛЬШОЙ ВЫБОРКИ ПАЦИЕНТОВ С ОСТРЫМ ОТЕЧНЫМ ПАНКРЕАТИТОМ

Статья посвящена изучению возможностей методов многомерного статистического анализа для определения степени однородности большой выборки пациентов с острым отечным панкреатитом. Подробно описана методология проведения дискриминантного, факторного и кластерного анализов на большом статистическом материале. В результате проведенного исследования удалось большую группу пациентов разбить на четыре разнородных кластера и выделить дифференцирующие критерии каждого кластера. Данный факт является прямым доказательством неоднородности патогенеза острого отечного панкреатита. Это, в свою очередь, позволит проводить более дифференцированную патогенетическую терапию и даст возможности улучшить результаты лечения данного заболевания, снизить частоту рецидивов болезни.

Ключевые слова: острый отечный панкреатит, методы многомерного анализа, факторный, дискриминантный, кластерный анализы, патогенетическая терапия, дифференцирующие критерии, клинические варианты панкреатита.

A. Spirin, V. Solntsev, V. Nazarov

Application of Methods of Multidimensional Statistical Analysis for Determining the Degree of Homogeneity of the Large Sample of Patients with Acute Edematic Pancreatitis

The article describes a study of the possibilities of methods of multivariate analysis to determine the degree of homogeneity of a large sample of patients with acute edematous pancreatitis. Details concerning the methodology of the discriminant, factor and cluster analysis on a large statistical material are provided. A large group of patients was divided into four heterogeneous clusters and the differentiating criteria for each cluster were suggested. This fact is regarded as a direct evidence of heterogeneity of the pathogenesis of acute edematous pancreatitis. This, in turn, would allow for more differentiated pathogenetic therapy and provide opportunities to improve THE results of treatment of this disease, to reduce the frequency of relapses of the disease.

Keywords: acute edematous pancretitis, methods of multivariate analysis, factorial, discriminant, cluster analysis, pathogenetic therapy, differentiating criteria, clinical variants of pancreatitis.

Заболеваемость острым панкреатитом из года в год неуклонно растет и в настоящее время занимает лидирующие позиции в структуре «острого живота». Наблюдается также и увеличение числа больных, повторно госпитализированных в хирургические стационары.

Однако остается неудовлетворенность проводимой однотипной консервативной терапией на стациионарном и амбулаторном этапах, отсутствием дифференцированных методов диагностики и четких критериев к дифференцированной фармакотерапии. Все это создает предпосылки для углубленного изучения данной проблемы.

Целью нашей работы является определение степени однородности большой выборки пациентов с острым отечным панкреатитом (ООП) с использованием методов многомерного статистического анализа для дальнейшей разработки методов дифференцированной терапии.

Нами были изучены результаты обследования 182 больных острым отечным панкреатитом, поступивших в хирургическое отделение СПбГУЗ «Николаевская больница» в 2002 г. (109 мужчин и 73 женщины). У каждого пациента изучалась динамика 57 качественных и количественных показателей (данные анамнеза, клинической картины, лабораторного и инструментального исследования и т. д.).

Перед нами стояла задача — найти и выделить группы больных из большой выборки пациентов, различающихся по совокупности показателей, и определить характер различий между ними, а также диагностические критерии. Работа проводилась в несколько этапов: 1) разведочный анализ первичных данных, 2) детальный предварительный анализ, 3) основная процедура кластеризации, 4) описание выделенных этиопатогенетических групп.

Основные предпосылки кластерного анализа.

Как отмечено в работе [1, с. 6], только «когда определенного уровня достигли вычислительная база исследований и теоретические разработки многомерного статистического анализа, главной проблемой теории и практики классификации и снижения размерности стало развитие достаточно изощренного и эффективного в приложениях математического аппарата».

В основе практически всех подходов и процедур классификации лежат четыре генеральных идеи и методологических принципа [1, с. 14]. (1) Эффект существенной *многомерности*, т. е. необходимость «опираться на совокупность множества взаимосвязанных свойств с учетом структуры и характера их связей» [2]. (2) Возможность *лаконичного объяснения* природы многомерных структур, состоящего в предположении о существовании небольшого числа определяющих факторов, возможно, *латентных*, т. е. не наблюдаемых непосредственно, но восстанавливаемых по исходным данным [3]. (3) Максимальное использование «обучения» в настройке математической модели классификации, т. е. наличие исходной априорной *обучающей* информации (результатов обследования пациентов и, иногда, наличия установленного диагноза у них) [4]. (4) Оптимизационная формулировка задачи классификации, состоящая в требовании поиска *наилучшего* метода классификации путем оптимизации некоторого экзогенно заданного критерия качества метода.

Согласно этим принципам, процесс решения задачи классификации становится многоэтапным [1, с. 42]. На первом этапе на предметно-содержательном уровне формулируется задача, содержащая описание характера требуемых выводов, описание предмета и объектов исследования. На втором этапе задача переформулируется на статистическом языке и выбирается математическая модель. На третьем этапе база исходных данных вводится в компьютер. На четвертом этапе проводится разведочный статистический анализ данных с целью выявления специфики их вероятностной и геометрической природы и уточнения модели. На пятом этапе осуществляется уточнение математической постановки задачи и возможная корректировка данных. На шестом этапе проводится вычислительная реализация выбранного метода. И на заключительном этапе анализируются и интерпретируются

результаты проведенного статистического анализа и либо формулируются окончательные научные и/или прикладные выводы, либо проводятся уточнения задания и затем — возврат к одному из предыдущих этапов.

При наличии информации об априорной принадлежности исходных объектов к некоторому классу мы имеем дело с классификацией по обучающим выборкам, и задача состоит в построении некоторой формальной процедуры, позволяющей наиболее точно (в некотором смысле) по исходным данным отнести объект к одному из априори определенных классов. Построение таких процедур является задачей статистического дискриминантного анализа.

В случае отсутствия априорной информации о принадлежности объектов (пациентов) к определенному классу говорят о классификации без обучения. В этом случае [1, с. 145] основной задачей является разбиение исходной совокупности объектов на некоторое число классов с одновременным определением числа этих классов. Для решения этой задачи используются процедуры кластерного анализа. Основной целью кластерного анализа является разбиение исходного множества объектов на сравнительно небольшое число *однородных* в определенном смысле групп или классов.

Сложность процедур кластерного анализа выражается в том, что кроме неопределенности числа классов имеется неопределенность в понимании «однородности» классов. Основной подход к решению последней проблемы состоит в интерпретации каждого из исходных объектов (пациентов) в качестве точки в многомерном признаковом пространстве. Естественно предположить, что геометрическая близость двух или нескольких точек в этом пространстве означает близость «физических» состояний соответствующих объектов (или «клинических» состояний пациентов). Задача кластеризации состоит в разбиении анализируемой исходной совокупности точек-наблюдений таким образом, чтобы объекты, принадлежащие одному классу, находились в среднем ближе друг к другу, чем к точкам других классов. Полученные таким образом классы называются кластерами (таксонами, образами).

Следует отметить особенность кластерного анализа: наличие двух проблем: выбор меры близости точек-состояний, метрики и подбор алгоритма формирования кластеров. «*Выбор метрики* (*или меры близости*) *является узловым моментом исследования*, от которого решающим образом зависит окончательный вариант разбиения объектов на классы при заданном алгоритме разбиения» [1, с. 147]. При конструировании различных процедур классификации (кластер-процедур) в ряде ситуаций оказывается необходимым также введение понятия расстояния между целыми группами объектов.

Важной проблемой кластерного анализа является понятие «наилучшего» решения задачи, поскольку само понятие *качества* разбиения можно определить многими различными способами. Но даже при фиксации некоторого критерия качества как количественного показателя нахождение его экстремального значения является очень сложной и трудоемкой вычислительной задачей, имеющей несколько вариантов реализации.

Следует отметить один важный момент, связанный с ограничениями и условиями применения кластерного анализа. Во многих руководствах одним из таких условий называется нормальное распределение совокупности показателей. Это условие связано с неполным пониманием теории кластерного анализа теми людьми-практиками, которые применяют его в реальных задачах, и существованием одной из трактовок кластерного анализа как задачи разделения смеси распределений. Моделей многомерных распределений немного, и чаще всего предполагается нормальность. Однако проверить предположение о смеси нормальных распределений практически невозможно.

Более того, многомерная нормальность фактически обязана теоретическим исследованиям середины XX века, где она просто постулировалась [2]. Но почти все часто используемые многомерные методы могут быть обоснованы и без использования гипотезы нормальности [5]. Опыт биомедицинской статистики свидетельствует, что ни эмпирически, ни даже теоретически априорные утверждения о нормальности не только многомерных, но даже одномерных распределений показателей строго обоснованы быть не могут. При этом из одномерной нормальности всех показателей совсем не следует нормальность многомерной совокупности (хотя из многомерной нормальности следует одномерная). Кроме того, в биологии и медицине крайне сложно представить себе ситуацию, модель, из которой следовала бы одновременно нормальность десятков (а то и сотен) разнородных показателей.

Вместо математической модели в кластерном анализе обычно формулируются априорные предположения, связанные с содержательной постановкой задачи. Так, в медицине предположения связаны с конкретной болезнью, определяемой через ее проявления, при этом опираются на сходство максимально большого количества симптомов. Важным условием корректного решения задачи кластерного анализа является однородность исходной совокупности, отсутствие аномальных случаев, выбросов (outliers). Поиск и исключение выбросов осуществляется с помощью разведочного анализа, который является обязательным этапом статистического исследования.

Необходимость обеспечить однородность выборки больных в клинических исследованиях привела к формированию большого набора правил отбора пациентов, правил «включения» и «исключения», которые приводят к созданию «правильной» выборки, к отсечению выбросов. Но большинство таких правил работают вслепую и не всегда приводят к однородной совокупности в математическом смысле. Поэтому в биомедицинской статистике практически все методические рекомендации содержат этап разведочного анализа, проверки одномерных и многомерных распределений показателей [3; 6].

Разведочный анализ.

В соответствии с описанными выше рекомендациями анализ однородности группы больных ООП начинался с разведочного анализа, целями которого являлись: выбор информативных признаков, характеристик состояния пациентов, измеренных при первичном обследовании, проверка однородности всей исследуемой совокупности (отсутствие резко выделяющихся наблюдений); подготовка рабочей базы данных для дальнейшего анализа.

Первоначально были проанализированы только количественные показатели, так как при максимальной информативности они наиболее удобны для анализа. Вычисления первичных статистических характеристик для исходных показателей показало, что больше половины показателей имеют сильно несимметричное распределение. Критерием качества распределения на данном этапе рассматривались: 1) близость друг другу трех характеристик положения (средних арифметического и геометрического и медианы), 2) близость к нулю коэффициентов асимметрии и эксцесса и 3) форма гистограммы распределения. Несколько показателей оказались малоинформативными и были исключены из дальнейшего анализа. У оставшихся показателей не было выявлено индивидуальных значений, резко отличающихся от основной массы их значений. Однако это не исключает отсутствия многомерных выбросов, например, когда имеются аномальные сочетания значений показателей.

Для выявления многомерных выбросов был использован анализ главных компонент, вариант факторного анализа, основанный на использовании корреляций между показателями для выделения наиболее информативных их комбинаций. При этом появляется возможность выявления и исключения показателей, дублирующих друг друга: у таких показателей коэффициенты корреляции очень близки к 1 или к –1, и тогда один из показателей

должен быть исключен из исследования. Анализ корреляционной матрицы коэффициентов корреляции Пирсона показал, что имеется только несколько относительно больших коэффициентов корреляции, например, между тремя размерами поджелудочной железы (головкой, телом и хвостом), от 0,64 до 0,71, между уровнями гемоглобина и эритроцитов, 0,77, между уровнями АЛТ и АСТ, 0,59. Эти корреляции естественны, но не настолько велики, чтобы исключать из исследования какой-либо из показателей. Остальные коэффициенты корреляции не превосходили по абсолютной величине 0,4.

Анализ главных компонент (АГК) на основании таблицы парных коэффициентов корреляции (корреляционной матрицы) формирует новые показатели, концентрирующие в себе всю основную вариабельность исходной совокупности. При этом такие новые показатели позволяют проводить их визуальное изучение, чтобы понять, с чем связана их вариабельность: с концентрацией естественной вариабельности исходных показателей или с неблагоприятным экстремальным сочетанием нескольких показателей, образующим «многомерный выброс». При большом числе показателей в АГК могут включаться не только количественные, непрерывные показатели, но и дискретные, качественные показатели, грубо описывающие количественные характеристики (например, степень выраженности патологии). При этом такой метод позволяет численно характеризовать информативность создаваемых им новых показателей, выражая ее в процентах от общей информативности всей совокупности.

В данном исследовании АГК был выполнен трижды. Сначала в анализ было включено максимальное число (52 показателя) количественных и качественных показателей, полученных как при первичном обследовании пациентов, так и в процессе лечения, чтобы исключить аномальное течение болезни. Выделились две четверки наиболее информативных ГК, включающие 26,7% и 17,4% общей информации (всего 44,1%). Затем малоинформативные показатели были исключены. На этом этапе информативность понималась как участие соответствующего показателя в формировании первых главных компонент. На втором этапе были исключены 14 показателей, затем, на третьем, еще пять показателей. Основным результатом такого анализа можно считать подтверждение структуры наиболее информативных ГК: первые две четверки ГК остались устойчивыми во всех трех вычислениях; при этом их общая информативность возросла сначала до 55,2%, а затем — и до 56,6%.

Поскольку главные компоненты имеют смысл новых показателей, индексов, содержащих в себе максимум информации об исходных показателях, можно визуально проанализировать значения этих показателей для всех индивидов в четырех плоскостях, определяемых последовательными парами ГК (1–2, 3–4, 5–6 и 7–8). Анализ построенных диаграмм рассеяния точек, соответствующих всем пациентам, включенным в исследование, позволил сделать вывод, что исходный набор показателей достаточно хорошо согласован, отсутствуют грубые ошибки в данных. Данные не «распадаются» на четкие подгруппы, но можно предположить, что некоторая внутренняя неоднородность может быть выявлена при более тщательном анализе.

Для исходных данных, без какого-либо отбора и преобразований основных показателей, сделаны три варианта кластерного анализа.

1. Кластерный анализ по переменным во многом продублировал результаты анализа корреляций и показал пары и группы показателей, наиболее сильно статистически связанных между собой.

2. Кластерный анализ по индивидам показал, что выделяется группа из 20 пациентов, очень далеко отстоящих от всех остальных, которые разделяются на две довольно близкие

группы. Из этого следовало, что эти выделенные большие группы далее могут быть разбиты на более мелкие подгруппы.

3. Кластерный анализ с разбиением на фиксированное число групп выделил группы из 63, 12 и 107 пациентов, причем во вторую группу вошли 12 из 20 пациентов, выделенных предыдущим анализом как самые далеко отстоящие от остальных (рис. 1).

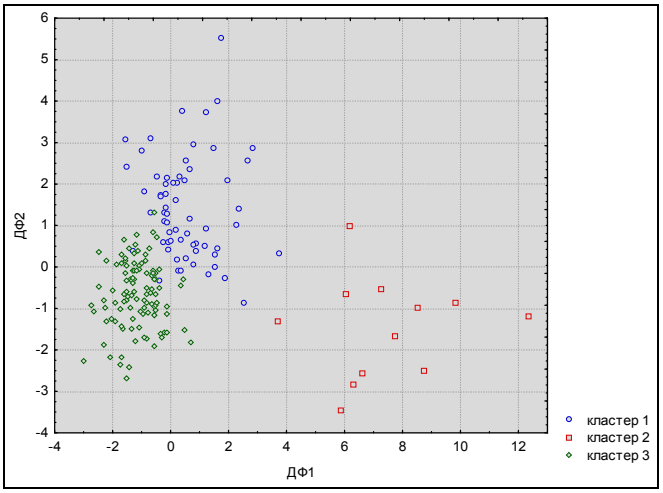

Рис. 1. Результат кластерного анализа с разбиением на три группы, визуализированный с помощью дискриминантного анализа

Однако процедуры кластерного анализа не позволяют «увидеть» наглядно результат разбиения. Таким инструментом может служить дискриминантный анализ, в котором имеется возможность построить наиболее наглядное расщепление групп. Необходимо подчеркнуть, что все эти результаты получены для первичных показателей, большинство из которых имеет сильно асимметричное распределение. Но разведочный анализ подтвердил, что данные однородны, не имеют грубых ошибок и имеют внутреннюю структуру.

Детальный предварительный анализ.

Для проведения детального анализа необходимо выполнить симметризацию показателей с помощью логарифмирования. Основная причина преобразования показателей — стремление сделать распределение значений показателя как можно более «нормальным», т. е. симметричным и одномодальным. Реально имеет смысл требовать, чтобы распределение было «похоже на нормальное», что означает требование однородности и типичности изучаемой совокупности (большинство индивидов похожи друг на друга, а отклонения от некоего «среднего», типичного индивида тем реже, чем больше величина отклонения, без различия, в какую сторону). Во многих биомедицинских исследованиях используются логарифмированные показатели, например, pH (логарифм концентрации ионов водорода в растворе).

В нашем исследовании повторный анализ распределений прологарифмированных показателей показал, что логарифмирование существенно улучшило их. После этого был проведен АГК по всей совокупности количественных показателей, в которой показатели с асимметричными распределениями были заменены на их логарифмы. Повторное применение АГК показало, что общая структура главных компонент сохранилась: основная информация, 54,8%, содержится в первых 8 ГК, и они группируются в две четверки, очень похожие на полученные при первичном анализе. Однако пространственное распределение то-

чек-пациентов в этих новых ГК несколько изменилось. Большинство точек, выделявшихся при первичном анализе, после логарифмирования слились с основной группой.

Далее был проведен повторный кластерный анализ, который показал, что структура совокупности пациентов в новой системе показателей тоже несколько изменилась. Теперь все пациенты разделились на четыре группы, не так сильно отличные друг от друга, как при предварительном анализе. Численности кластеров: 42, 48, 87 и 5 пациентов.

Дискриминантный анализ, проведенный для этих четырех групп пациентов, подтвердил, что такое разбиение может иметь место. Пошаговый отбор информативных показателей, проведенный двумя разными методами (метод последовательного включения и метод последовательного исключения), оставил 17 информативных показателей. Визуальный анализ результатов дискриминантного анализа показал, что четвертая группа достаточно далеко отстоит от трех первых (рис. 2).

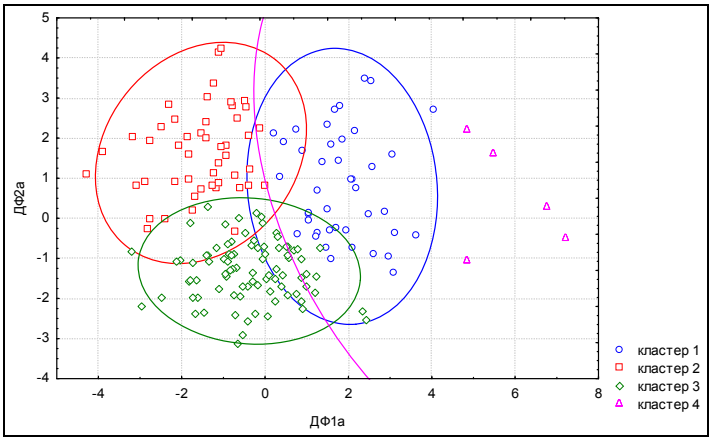

Рис. 2. Результат повторного кластерного анализа

Был проведен также анализ связи выделенных групп с дискретными показателями, который подтвердил оправданность такого разбиения. Полученные данные позволили приступить к следующему этапу исследования.

Основная процедура кластеризации.

В результате двух этапов первичного анализа для разбиения всей исследуемой группы, состоящей из 182 пациентов, на кластеры было отобрано 17 первичных количественных показателей: 1) возраст, 2) размер печени, 3) гемоглобин при поступлении, 4) эритроциты при поступлении, 5) сегментоядерные лейкоциты, 6) логарифм эозинофилов, 7) СОЭ, 8) общий белок, 9) логарифм головки П.Ж., 10) логарифм тела П.Ж., 11) логарифм хвоста П.Ж., 12) логарифм холедоха, 13) логарифм толщины стенки желчного пузыря, 14) логарифм лейкоцитов в динамике, 15) логарифм палочкоядерных лейкоцитов, 16) логарифм лимфоцитов в динамике, 17) логарифм моноцитов.

Поскольку все эти показатели измерены в самых различных единицах, для кластерного анализа была выполнена их стандартизация (из значений каждого показателя было вычтено его среднее значение, и результат разделен на величину стандартного отклонения). Для того чтобы приступить к основному этапу кластеризации, сначала был выполнен АГК для полученного массива 17 стандартизированных показателей.

Анализ показал, что первые четыре главные компоненты содержат более 50% всей вариации, т. е. информации о всей совокупности данных, первые шесть — почти 2/3, а первые восемь — почти 75%. В первую ГК входят с наибольшим весом логарифмы размеров

головки, тела и хвоста ПЖ, а также значение СОЭ и логарифм количества лимфоцитов. Во вторую — другие показатели крови. Ротация показателей в плоскости первой и второй ГК позволила получить их более четкую интерпретацию: первая ГК определяется исключительно тремя размерами поджелудочной железы, а вторая ГК — показателями крови (СОЭ, количеством сегментоядерных нейтрофилов и логарифмами количеств лейкоцитов, лимфоцитов, палочкоядерных нейтрофилов и эозинофилов), при этом количество лимфоцитов является ведущим показателем второй ГК.

Кроме того, результаты анализа ГК позволили проверить внутреннюю однородность совокупности. Очень больших выбросов, аналогичных тем, что были выявлены при первичном анализе всех количественных показателей, обнаружено не было. Однако было отмечено несколько пациентов, первичные показатели которых в совокупности достаточно далеко отстоят от остальной группы. Эта информация была использована при кластерном анализе.

Третья ГК полностью определилась парой связанных между собой показателей — количествами гемоглобина и эритроцитов, а четвертая — балансом между сегментоядерными и палочкоядерными нейтрофилами (так как они входят в четвертую ГК с разными знаками), а также возрастом пациента. Пятую ГК определили логарифмы размеров холедоха и толщины стенки желчного пузыря. В шестой и седьмой ГК ведущими показателями оказались уровень белка и размеры печени, а в восьмой ГК — логарифм количества моноцитов. Таким образом, все 17 показателей оказались в разной степени информативными при описании разнообразия в группе из 182 пациентов. Полученные данные позволили приступить к окончательному проведению кластеризации.

Поскольку весь предыдущий анализ был направлен на формирование максимально однородной, «шарообразной» совокупности объектов (пациентов), естественным явился выбор в качестве расстояния между пациентами евклидова расстояния (соответствующего естественным геометрическим представлениям). Методом построения «дерева» расстояний был выбран метод Ворда (Ward) как максимально последовательный и структурированный (рис. 3). Результат этого шага анализа показал, что из всей совокупности довольно четко может быть выделена группа из 55 пациентов (левая ветвь дерева), которая в принципе может быть разбита на две примерно равные по численности подгруппы. В оставшейся группе опять выделяется относительно небольшая подгруппа из 24 пациентов, а остальная часть, скорее всего, может быть разбита на две или три подгруппы.

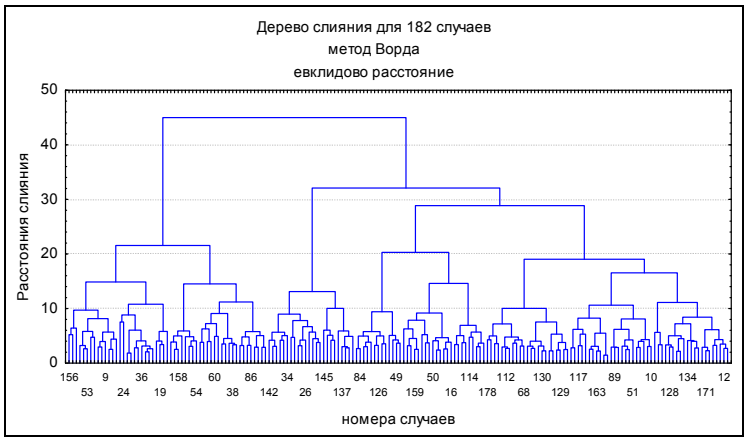

Рис 3. Основной этап кластеризации 182 пациентов,
«дерево», построенное методом Ворда на основе евклидова расстояния

Этот анализ послужил основой для выбора числа подгрупп в следующем методе — методе *k*-средних. Этот метод ориентируется на заранее заданное число *k* подгрупп, но формирование подгрупп может происходить с помощью двух несколько различных процедур. Таким образом, было проанализировано шесть вариантов: по два варианта для 3-й, 4-й и 5-й групп, а также два варианта, построенных непосредственно по «дереву» расстояний — из 4-й и 6-й групп. Попарное сравнение вариантов показало, что они довольно хорошо согласуются между собой. При этом чаще всего варианты с большим числом групп «получались расщеплением» более крупных групп в вариантах с меньшим числом групп. Наилучшее соответствие между собой показали три варианта из четырех групп. Они и были взяты за основу разбиения, классификации больных. Поскольку ни в одной из пар вариантов группы не совпадали полностью, были выделены пациенты, входящие, по крайней мере, одновременно в две из аналогичных групп разных вариантов.

Описание выделенных групп.

В результате детальной процедуры кластеризации в 1-ю группу попали пациенты из самой первой, левой группы «дерева разбиений» (42 человека), во 2-ю — пациенты из самой маленькой группы (23 человека), в 3-ю — из следующей, средней группы (27 человек) и в 4-ю — из самой правой группы (46 человек). Остальные 44 пациента остались как бы «между группами», т. е. это те, которых нельзя с уверенностью отнести ни к одной из четырех основных групп.

После такого формирования групп была оценена информативность каждого из 17 исходных показателей относительно этого разбиения. Малоинформативными оказались только три показателя: размер печени (он «работает» только при разбиении на пять групп, но в таком разбиении хуже работают другие показатели), логарифм количества моноцитов и уровень белка. Кроме того, была проверена информативность разбиения относительно 10 качественных показателей (в основном, анамнестических). Значимыми оказались только три из них: пол ($p = 0{,}003$), наличие сердечно-сосудистых заболеваний ($p = 0{,}0001$) и употребление алкоголя ($p = 0{,}03$).

Анализ результатов этого исследования показывает, что у количественных показателей первого обследования различий между группами нет только по нескольким: по размеру печени, по числу моноцитов, по биохимическим показателям (кроме уровня билирубина). Значимые попарные различия по критерию Шеффе выявлены у 13 показателей. В то же время из показателей второго обследования слабо значимые различия оказались только по двум: по эритроцитам и СОЭ. Полученное различие результатов двух обследований можно объяснить тем, что правильная терапия почти полностью устраняет различия между подгруппами пациентов; первоначальное их различие остается только в двух показателях крови.

Дополнительно было проанализировано, проявляется ли различие между группами на изменениях количественных показателей между двумя обследованиями. Анализ показал, что различие между группами относительно изменений проявляется у четырех показателей: у лейкоцитов, лимфоцитов, палочкоядерных и билирубина. Однако различие между группами оказывается не настолько сильным, чтобы можно было относить пациентов к той или иной группе. В то же время можно, используя разработанные механизмы дискриминантного анализа, отобрать наиболее информативные (с точки зрения различения групп) показатели и построить из них новые, дискриминантные индексы, используя которые можно классифицировать больных.

Для отнесения пациента к одной из четырех подгрупп необходимо использовать 10 исходных показателей, из которых были построены три дискриминантных функции (индекса). При этом получилась иерархическая процедура классификации, когда каждая из

дискриминантных функций последовательно определяет одну из подгрупп. В таблице приведены коэффициенты полученных дискриминантных функций, позволяющих отнести пациентов к одной из четырех групп.

Показатель	ДФ1	ДФ2	ДФ3
Возраст	–1,94	–2,71	–1,94
Логарифм головки	14,6	–3,02	14,7
Логарифм тела	2,27	—	—
Логарифм хвоста	13	–7,37	7,25
Логарифм холедоха	–17,9	–12,4	–26,7
Эритроциты	–6,2	192,6	14,6
СОЭ	1,93	–3,26	–2,9
Логарифм лимфоцитов	–6,8	–2,24	6,48
Логарифм лейкоцитов	0,47	3,15	–16,7
Сегментоядерные	–0,42	2,89	1,12
Константа	–367,56	-375,01	–215,12

Для вычисления значения индекса значения показателей умножаются на соответствующие коэффициенты, складываются и добавляется константа. Последующий алгоритм прост:

если ДФ1 < 0, то это — 1-я группа, иначе

если ДФ2 < 0, то это — 3-я группа, иначе

если ДФ3 < 0, то это — 2-я группа, иначе это — 4-я группа.

Алгоритм может быть легко реализован в виде простой вычислительной программы. Тестирование этого алгоритма на исходных данных показало, что он дает только 5,3% ошибочных решений. При этом алгоритм распределяет по четырем группам и 44 пациентов, первоначально оказавшихся вне кластеров.

Представляют большой интерес сопоставление полученного разбиения пациентов на четыре группы с качественными данными анамнеза и обследование пациентов, которое показало, что группы значимо различаются по полу, наличию желчно-каменной болезни, сердечно-сосудистых заболеваний, наличию рвоты и структуре поджелудочной железы на УЗИ. Это позволяет достаточно хорошо идентифицировать полученные субдиагнозы по их этиологии и сформулировать рекомендации к терапии.

Следует отметить одну важную особенность процедур статистического анализа (особенно регрессионного и дискриминантного): если качество процедуры оценивается по той же выборке, по которой строилась процедура, то это качество оказывается несколько завышенным, а вероятность ошибочных решений — заниженной [5; 8]. Поэтому при решении практических задач статистическими методами всегда рекомендуется либо использовать дополнительные более сложные методы для оценки реальной надежности метода (например, методы перепроверки, или бутстрепа [7], либо проводить дополнительное исследование на независимой выборке пациентов.

Вывод. Таким образом, используя методы многофакторного анализа, мы доказали, что острый отечный панкреатит — это неоднородное и полиэтиологическое заболевание, состоящее из четырех этиопатогенетических групп. Такое разделение нам в последующем позволит выделить диагностические критерии каждой группы, определить ту или иную причину развития острого панкреатита, а зная причину, мы сможем проводить дифференцированную терапию данного заболевания и предотвращать частые рецидивы болезни.

СПИСОК ЛИТЕРАТУРЫ

1. *Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Классификация и снижение размерности: Справочное издание / Под ред. С. А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.

2. *Андерсон Т.* Введение в многомерный статистический анализ. М.: Физматгиз, 1963. 500 с.

3. *Гланц С.* Медико-биологическая статистика / Пер. с англ. М.: Практика, 1998. 459 с.

4. *Ланг Т. А., Сесик М.* Как описывать статистику в медицине: Аннотированное руководство для авторов, редакторов и рецензентов / Пер. с англ. М.: Практическая медицина, 2010. 485 с.

5. *Рао С.Р.* Линейные статистические методы и их приложения. М.: Наука, Физматгиз, 1968. 548 с.

6. *Реброва О. Ю.* Статистический анализ медицинских данных. С помощью пакета программ «Статистика». М.: МедиаСфера, 2001.

7. *Эфрон Б.* Нетрадиционные методы многомерного статистического анализа: Сб. статей / Пер. с англ. М.: Финансы и статистика, 1988. 263 с.

8. *Krzanovski W. J.* Principles of multivariate analysis: a user's perspective. Oxford: Claredon press, 1993.

REFERENCES

1. *Ajvazjan S. A., Buhshtaber V. M., Enjukov I. S., Meshalkin L. D.* Prikladnaja statistika: Klassifikatsja i snizhenie razmernosti: Spravochnoe izdanie / Pod red. S. A. Ajvazjana. M.: Finansy i statistika, 1989. 607 s.

2. *Anderson T.* Vvedenie v mnogomernyj statisticheskij analiz. M.: Fizmatgiz, 1963. 500 s.

3. *Glants S.* Mediko-biologicheskaja statistika / Per. s angl. M.: Praktika, 1998. 459 s.

4. *Lang T. A., Sesik M.* Kak opisyvat' statistiku v meditsine. Annotirovannoe rukovodstvo dlja avtorov, redaktorov i retsenzentov / Per. s angl. M.: Prakticheskaja meditsina, 2010. 485 s.

5. *Rao S. R.* Linejnye statisticheskie metody i ih prilozhenija. M.: Nauka, Fizmatgiz, 1968. 548 s.

6. *Rebrova O. Ju.* Statisticheskij analiz meditsinskih dannyh. S pomoshch'ju paketa programm «Statistika». M.: MediaSfera, 2001.

7. *Jefron B.* Netraditsionnye metody mnogomernogo statisticheskogo analiza: Sb. statej / Per. s angl. M.: Finansy i statistika, 1988. 263 s.

8. *Krzanovski W. J.* Printsiples of multivariate analysis: a user's perspective. Oxford: Claredon press, 1993.

Д. А. Соколов, А. И. Ходанович

ПРОГРАММНО-АППАРАТНЫЙ КОМПЛЕКС МОНИТОРИНГА ФОРМИРОВАНИЯ ИНФОРМАЦИОННО-ПРАВОВОЙ КОМПЕТЕНТНОСТИ

Современная информационно-коммуникативная действительность требует новых методологических подходов к ее изучению и соответственно новых методов и критериев оценки педагогических достижений. Необходимо систематизировать материалы в области инновационных образовательных технологий. В связи с углублением процессов интеграции науки и образования особую актуальность сегодня приобретают теоретико-экспериментальные исследования компетентностного подхода в системе непрерывного образования. В статье рассмотрены концептуальные вопросы разработки и применения программно-аппаратного комплекса мониторинга формирования информационно-правовой компетентности.

Ключевые слова: программно-аппаратный комплекс, мониторинг, инновации, информационно-правовая компетентность.