

АЛГОРИТМ ГРУППОВОЙ ФАКТОРНОЙ МОДЕЛИ ДИСПЛАЗИИ СОЕДИНИТЕЛЬНОЙ ТКАНИ

Глотов А.В., Гольяпин В.В.,
Ложников Е.В.

Омский государственный университет
им. Ф.М. Достоевского, Омск
[e-mail avglotov@omsu.ru](mailto:e-mail_avglotov@omsu.ru)

1. Введение

Развитие современной медицины обусловлено появлением различных электронных приборов и устройств, которые позволяют получать большой объем информации, количественных и качественных показателей. В связи с этим появилось большое число методов, с помощью которых обрабатывается информативный массив данных. Наибольший интерес представляют методы многомерного факторного анализа, являющиеся аналитическими методами обработки биологических и медицинских результатов.

Использование в исследованиях понятия факторных нагрузок представляется актуальным для анализа многочисленных параметров, характеризующих состояние больных с генетически детерминированными заболеваниями, обусловленными дисплазией соединительной ткани (ДСТ) [1].

Дисплазия соединительной ткани – группа генетически гетерогенных и клинически полиморфных патологических состояний, объединенных нарушением формирования соединительной ткани в эмбриональном и постнатальном периодах [2].

Актуальность проведенной работы состоит в том, что проблемы диагностики дисплазии соединительной ткани (ДСТ) обусловлены трудностями правильной оценки типичных нарушений метаболизма компонентов соединительной ткани, липидного обмена, иммунитета и эндотелиальной функции, сопутствующих этому синдрому, а также необходимостью дифференцировать эти нарушения с использованием понятия факторной структуры.

Целью исследования является создание модели группового факторного анализа, которая способна обнаружить основные скрытые закономерности, характеризующие состояния больных дисплазией соединительной ткани, а также позволяющей выявить нарушение метаболизма и эндотелиальной дисфункции при дисплазии соединительной ткани.

Для реализации данной цели были поставлены следующие задачи: формирование базы исходных показателей, характеризующих ДСТ, освоение математического аппарата, используемого для построения факторных моделей, разработка алгоритма построения групповой факторной модели, реализация полученного алгоритма в диагностике ДСТ, получение научно значимых результатов работы алгоритма и их адекватная интерпретация.

2. Материалы и методы

Для построения математической модели использованы биохимические, гематологические параметры, полученные при обследовании пациентов с ДСТ. Проведено исследование гематологических показателей, характеризующих тромбоцитарно-эндотелиальную дисфункцию (понтанная агрегация тромбоцитов (САТ), индуцированная аденозиндифосфата и коллаген на агрегометре «Биола-0 231LA» по методу J. Born в модификации З.А. Габбасова).

Непрямой иммунофлюоресцентный анализ использовался для определения адгезивных макромолекул – рецептора CD44 лимфоцитов. Свободный оксипролин определялся по методике R.E. Neuman и M.A. Logan в модификации П.Н. Шараева (2003 г.) [3]. Уровень фибронектина определялся в плазме с помощью стандартной иммуноферментной тест-системы (ЦНИВС им. И.И. Мечникова, Россия), проколлаген III типа – с помощью иммуноферментной тест-системы «Labsystems» (Финляндия), холестерин липопротеидов различной плотности и триглицериды – иммуноферментным методом с помощью стандартных тест-систем.

3. Теоретические основы построения групповой факторной модели и алгоритм реализации

Основная цель факторного анализа состоит в выявлении гипотетических величин, или факторов, по большому числу экспериментальных данных, а задачей факторного анализа является нахождение простой структуры, которая бы достаточно точно отражала и воспроизводила реальные, существующие в природе зависимости.

Пусть данные распределены по нормальному закону распределения и записаны в виде матрицы $X = (x_{ij})$, где индекс $i = 1, 2, \dots, m$ относится к переменным, а $j = 1, 2, \dots, n$ к индивидуумам. Для приведения такого типа переменных к одному масштабу применяют классическую стандартизацию. Матрицу стандартизованных данных будем обозначать за $Z = (z_{ij})$, тогда все

средние значения переменных Z равны нулю, а все дисперсии равны единице [4]. Из этого сле-

дует, что для корреляционной и ковариационной матриц R и S имеет место соотношение:

$$\frac{1}{n-1}ZZ' = R = S. \quad (1)$$

Групповой метод отличается от всех других методов выделения факторов тем, что одновременно на одном шаге получают сразу несколько факторов. Основная отличительная особенность метода состоит в том, что вначале определяются группы переменных, которые тесно коррелируют между собой, и затем выделяется ряд факторов, из которых каждый соответствует такой группе. В данном методе каждый раз перед выделением факторов необходима априорная груп-

пировка переменных. Кроме этого, необходимо оценить диагональные элементы корреляционной матрицы – общности, которые представляют собой части единичных дисперсий переменных, которые обозначим за вектор столбец $\vec{h}^2 = (h_1^2, \dots, h_m^2)$ размерностью m . Один из самых распространенных методов оценки общностей является усреднение коэффициентов корреляции данного параметра:

$$h_i^2 = \frac{1}{m-1} \sum_{\substack{i \in \bar{z}_j \\ i \neq j}} r_{ij}. \quad (2)$$

При необходимости, можно получить точные общности, используя метод минимальных остатков [5].

Для получения корреляционной групповой матрицы нужно сгруппировать элементы матрицы с общностями на главной диагонали. Груп-

пировку проводим априорно, группируем m параметров по r группам G_p ($p = 1, \dots, r$).

Для вычисления дисперсии факторов и коэффициента ковариации между ними удобно посчитать сначала некоторые суммы коэффициентов корреляции:

$$w_{ip} = \sum_{k \in G_p} r_{ik}, \quad (3)$$

где $i = 1, 2, \dots, m$, а p – номер группы.

$$W_{pq} = \sum_{i \in G_p} w_{iq}, \quad (4)$$

где p и q – номера соответствующих выделенных групп.

С помощью этих сумм легко получить корреляционную групповую матрицу $\Phi = (\phi_{pq})$. Для этого воспользуемся формулой:

$$\phi_{pq} = \frac{W_{pq}}{\sqrt{W_{pp} W_{qq}}}. \quad (5)$$

Далее выразим через суммы исходных коэффициентов корреляции значения коэффициентов корреляции между параметрами и факто-

рами, то есть элементы косоугольной факторной структуры. Эти элементы будут выражаться следующим образом:

$$v_{ip} = \frac{w_{ip}}{\sqrt{W_{pp}}}. \quad (6)$$

В итоге получаем матрицу $V = (v_{ip})$ косоугольной факторной структуры, которая имеет размерность $m \times r$ [6].

4. Алгоритм построения групповой факторной модели

1. Стандартизация входных параметров $X = (x_{ij})$.
2. Нахождение корреляционной матрицы по формуле (1) и оценка общностей по формуле (2).
3. Вычисление суммы коэффициентов корреляции каждого параметра со всеми параметрами

каждой из групп по формуле (3) и расчет матрицы суммы коэффициентов корреляции между всеми параметрами группы G_p и всеми параметрами группы G_q по формуле (4).

4. Получение групповой корреляционной матрицы, используя формулу (5).

5. Расчет матрицы $V = (v_{ip})$ косоугольной факторной структуры, используя (6).

6. Результаты и выводы

В результате исследовательской работы получена следующая факторная структура, представленная на таблице.

Факторная структура с группами факторов, влияющих на ДСТ

	V_1	V_2	V_3	V_4
Спонтанная агрегация	0,88	0,08	0,18	0,14
АДФ-индуцированная агрегация	-0,85	-0,26	-0,28	-0,07
Коллаген-индуцированная агрегация	-0,94	-0,18	-0,27	-0,30
Холестерин	-0,06	-0,95	-0,07	-0,14
Триглицериды	-0,05	-0,86	-0,15	-0,15
α -Холестерин	-0,20	-0,81	-0,28	-0,29
Оксипролин мочи	0,23	0,17	0,75	-0,10
Оксипролин ИДС	0,24	0,25	0,89	-0,15
Оксипролин крови	-0,18	-0,07	0,93	-0,18
Фибронектин	0,05	0,35	0,14	-0,94
СД-44	0,38	0,20	0,37	0,85
Коллаген III типа	-0,26	-0,25	-0,26	-0,74

Анализ полученной факторной структуры позволяет сделать вывод:

Первая группа параметров относится к агрегациям. Коллаген-индуцированная агрегация обусловлена повреждением сосудистой стенки, АДФ-индуцированная агрегация показывает появление факторов в сосудистом русле, спонтанная агрегация – это самоиндуцированная агрегация тромбоцитов.

Вторая группа параметров (холестерин, триглицериды, α -Холестерин) относится к фактору нарушения липидного обмена, то есть накопления липидов в плазме и снижение их соединений в мембране клетки. Таким образом, мембрана тромбоцитов теряет способность к текучести, становится ломкой. Тромбоцитарные дисфункции связаны с нарушением адгезии к коллагену, транспорта внутриклеточного кальция и реакции высвобождения.

Третья группа параметров отвечает за фактор разрушения стенки сосуда, поскольку избыток оксипролина, входящего в состав коллагена,

указывает на то, что средняя стенка сосуда дефективная.

Четвертая группа параметров образует фактор, который характеризует степень разрушения волокон соединительной ткани. Фибронектин указывает на степень склеивания клеток. Ключевым в миграции клеточных элементов крови при ДСТ являются СД44 – глюкокоронат хряща. Процессы миграции клеточных элементов при ДСТ нарушены, что существенно влияет на состояние адаптивного иммунитета. Коллаген показывает на фактор поражения стенки сосудов. Чем выше содержание коллагена, тем больше поражения сосудистой стенки.

Полученная матрица косоугольной факторной структуры позволяет выявить существенные взаимосвязи между биохимическими и гематологическими параметрами у больных ДСТ. Разработанная методика может применяться в диагностике и выявлении гематологических, биохимических и иммунологических отклонений у пациентов с ДСТ.

СПИСОК ЛИТЕРАТУРЫ

1. Яковлев В.М., Нечаева Г.И., Готов А.В. Терминология, определение с позиции, классификация врожденной дисплазии соединительной ткани // Врожденная дисплазия соединительной ткани: тезисы регионального симпозиума. – Омск, 1990. – С. 3–5.
2. Яковлев В.М., Готов А.В., Ягода А.В. Иммунопатологические синдромы при наследственной дисплазии соединительной ткани. – Ставрополь, 2005.
3. Шараев П.Н., Иванов В.Г., Гаврилов А.Л. Методы лабораторного исследования показателей обмена коллагена в биологических жидкостях: Информационное письмо для врачей клинической лабораторной диагностики. – Ижевск, 2003. – С. 19.
4. Иберла К. Факторный анализ. – М.: Статистика, 1980.
5. Гольяпин В.В. Вычислительные аспекты метода минимальных остатков при разрешении варианта Хейвуда. // Сибирский журнал промышленной математики. – 2005. – Том VIII, №3(23). – С. 145–151.
6. Харман Г. Современный факторный анализ. – М.: Статистика, 1972.

ФАКТОРИЗОВАННЫЕ МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ В КОМПЬЮТЕРНОЙ ДИАГНОСТИКЕ ДИСПЛАЗИИ СОЕДИНИТЕЛЬНОЙ ТКАНИ

Готов А.В., Гольяпин В.В.,
Лобачев А.И.

Омский государственный университет
им. Ф.М. Достоевского, Омск

Введение

В развитии синдрома соединительнотканной дисплазии (СТД) принимают участие многие факторы. Как правило, в них задействованы эндогенные и экзогенные механизмы.

Увеличение числа случаев СТД, наблюдающееся в настоящее время, связано с патогенными воздействиями в онтогенезе, связанными с ухудшением экологической обстановки. Так же немаловажными факторами являются плохое питание и стрессы.

На сегодняшний день одной из задач современной медицины является исследование дисплазии соединительной ткани. Это состояние, которое зависит от молекулярных, генетически детерминированных дефектов биосинтеза компонентов соединительной ткани (каллогена, эластина, фибрина).

В связи с увеличением случаев СТД и, соответственно, с возросшей актуальностью данной проблемы, основной целью стала реализация факторной модели средовых воздействий и создание программного комплекса на базе факторизованных методов распознавания образов, решающего диагностическую задачу, посвященную дисплазии соединительной ткани подростков. В целях проведения соответствующего научного исследования специалистами была составлена анкета для опроса детей школьного возраста, характеризующая социальный статус индивидуумов с СТД и без него.

Факторизованные методы распознавания образов

Прежде чем приступить к изложению факторизованных методов распознавания образов, сформулируем основные понятия факторного анализа. Одной из основных целей факторного анализа является извлечение на поверхность фактора, который бы наиболее точно позволил воспроизвести наблюдаемые корреляции.

Для любого факторного исследования исходные данные записываются в виде матрицы $Y = \{y_{ij}\}$, где индекс $i = 1, 2, \dots, m$ относится к переменным (показателям), а индекс $j = 1, 2, \dots, n$ к объектам (индивидуумам). Если исходные данные распределены по нормальному закону распределения, то имеется возможность проведения стандартизации матрицы Y . Тогда для стандартизованных данных имеет место соотношение:

$$R = \frac{1}{n-1} Z Z^T, \quad (1)$$

где $Z = \{z_{ij}\}$ – матрица стандартизованных исходных данных; R – корреляционная матрица.

Целью любого метода факторного анализа является представление величины z_{ij} , то есть

элемента матрицы Z , в виде линейной комбинации нескольких гипотетических переменных, или факторов. Т.е. значение z_{ij} может быть выражено в виде линейной комбинации факторов:

$$z_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \dots + a_{ir}f_{rj}, \quad (2)$$