

«Компьютерное моделирование в науке и технике»,
ОАЭ (Дубай), 16-23 октября 2012 г.

Технические науки

**ИСПОЛЬЗОВАНИЕ МЕТОДА ГЛАВНЫХ
КОМПОНЕНТ И ЛАТЕНТНОГО АНАЛИЗА
В ДИАГНОСТИКИ АРТЕРИАЛЬНОЙ
ГИПЕРТЕНЗИИ**

¹Гольяпин В.В., ²Лобачев А.И.

¹Омский филиал ФГБУН Института математики
им. С.Л. Соболева СО РАН, Омск,
e-mail: golyapin@mail.ru;

²ГОУ «Омский государственный университет
им. Ф.М. Достоевского», Омск,
e-mail: a.lobachev@medteh-com.ru

В данной работе исследуется феномен артериальной гипертензии. Артериальная гипертензия является одной из наиболее значимых медико-социальных проблем Российской Федерации. Артериальная гипертензия – синдром повышенного артериального давления.

Гипертоническая болезнь одна из форм артериальной гипертензии – хроническое заболевание, основным клиническим признаком которого является длительное и стойкое повышение артериального давления. Согласно трех стадийной классификации гипертонической болезни первая стадия предполагает отсутствие поражения органов-мишеней, вторая стадия – присутствие изменений со стороны одного или нескольких органов-мишеней; диагноз третьей стадии устанавливается при наличии ассоциированных клинических состояний [1].

Первичной задачей исследования являлось выявление интегративных латентных характеристик, опирающихся на взаимозависимость измеряемых показателей. Для решения был выбран математический аппарат факторного анализа. Методы факторного анализа позволяют определять скрытые, неявные закономерности, объективно существующие в той или иной отрасли или науке. Как правило, эти данные не поддаются непосредственному изучению. Также факторный анализ помогает свести обширный статистический числовой материал к нескольким простым зависимостям [2].

Вторая задача данного исследования заключалась в формировании классов пациентов согласно выявленным латентно-интегративным характеристикам (факторам). Данная задача решалась посредством алгоритма латентного анализа. Латентный анализ – метод вероятностно-статистического моделирования, идея которого основана на предположении, что наблюдаемое поведение (например, ответы индивидов на вопросы теста или анкеты) есть внешнее проявление некоторой скрытой (латентной) харак-

теристики, присущей индивидам. Задача метода заключается в том, чтобы, изучив наблюдаемое поведение индивидов, вывести эту скрытую характеристику и классифицировать индивидов по сходству ее значений [3].

**Основы математического аппарата,
используемого в данном исследовании**

В качестве математического аппарата использовались модели и алгоритмы факторного и латентного анализа.

В факторном анализе требуется отыскание плоскостей или гиперплоскостей, проходящих через центр тяжести облака точек в-мерном пространстве и с условием, что сумма квадратов расстояний всех точек от этих плоскостей минимальна. Для метода главных факторов и метода главных компонент это максимизация дисперсии в одном направлении при выполнении ниже следующего ограничения [4, 5, 6]:

$$R - U^2 = A \cdot C \cdot A^T, \quad (1)$$

где A – матрица весовых нагрузок факторов размерности $m \times r$, C – матрица корреляции между факторными значениями размерности $r \times r$, а U^2 – диагональная матрица с общностями. Если принять $C = E$, то факторы должны быть ортогональны. Для ортогональных моделей равенство (1) для редуцированной матрицы R выглядит следующим образом:

$$R_h = A \cdot A^T. \quad (2)$$

Максимизация дисперсии достигается при выполнении ограничения (2), если в качестве матрицы положить A произведение двух первых матриц сингулярного разложения матрицы R_h . В принципе, можно вычислить столько главных компонент, сколько имеется переменных, но это неэкономично. Чтобы обобщить информацию, содержащуюся в исходных переменных, лучше выделить небольшое количество факторов. Для определения количества факторов предлагается несколько процедур: определение, основанное на предварительной информации; определение, основанное на собственных значениях факторов; критерий «каменистой осыпи»; определение на основе процента объясненной дисперсии; метод расщепления и критерии значимости. В нашем случае был использован критерий «каменистой осыпи». То есть если собственные значения матрицы R_h больше 1 и точка на графике лежит близко к основанию, то это означает о возможности исключения данного фактора. После

того, как найдена матрица A и производится расчет факторных значений для каждой группы показателей [7, 8].

Известно, что факторном пространстве однородные факторные значения не превышают трех. Это связано с тем, что математическое ожидание равно нулю, а дисперсия и, следовательно, стандартные отклонения единичные. А любая выборка факторных значений симметрическая. Поэтому в качестве метода распознавания образов можно использовать латентный анализ на базе построенной номинальной шкалы, используя для ее построения информацию о попадании вычисленных факторных значений у диагностируемого индивидуума в $[-3; 3]$. Если факторное значение попадает в этот интервал, то ставим «+», при формировании соответствующей номинальной шкалы, и «-» в обратном случае. В том случае, если, например, выделяется 2 фактора, то возможно выделение следующих групп: «+, +», «+, -», «-, +», «-, -». Все индивидуумы у которых оба плюса считаются условно здоровыми. Все остальные имеют соответствующую патологию по одному фактору или по обоим.

На последнем этапе факторного исследования объединяем все полученные факторные значения групп в одну таблицу и преобразуем в элементы номинальной шкалы по принципу «+», «-». Построенная таблица с данными передается для обработки латентному анализу, который рассматривает её как некий тест или анкету.

Суть латентного анализа состоит обработки теста или анкеты, состоящей из вопросов, которые относятся к изучаемой скрытой характеристике. Выделенные вопросы называют явными переменными, а скрытую характеристику – латентной переменной. В теории тестов скрытая характеристика интерпретируется как одномерный латентный континуум (непрерывная латентная переменная) [9, 10, 11].

Вводится так функция вопросов-го вопроса $f_i(x)$. Это вероятность положительного ответа индивида на i -й вопрос, при условии, если индивид находится в точкелатентного континуума. Далее вводится так называемый маргинал i -го вопроса – p_i . Это доля лиц, которые положительно ответили на i -й вопрос. Наконец, поскольку задача вероятностная, необходимо найти закон распределения лиц на континууме, т.е. плотность вероятности $\varphi(x)$.

Таким образом, нами введены следующие величины:

$f_i(x)f_j(x)$ – функции вопросов;
 p_i – маргиналы вопросов;
 $\varphi(x)$ – закон распределения лиц на латентном континууме;
 $\varphi(x)dx$ – число лиц в интервале x и $x + dx$;
 $f_i(x)\varphi(x)dx$ – число лиц в интервале x и $x + dx$, которые положительно ответили на i -й вопрос;

$\int_{-\infty}^{\infty} f_i(x)\varphi(x)dx$ – число лиц на всем континууме, которые положительно ответили на i -й вопрос, т.е. это число равно маргиналу – известной величине.

Отсюда основное расчетное уравнение латентного анализа:

$$p_i = \int_{-\infty}^{\infty} f_i(x)\varphi(x)dx. \quad (3)$$

Слева – исходные эмпирические переменные, справа – латентные переменные, которые нам неизвестны. Цель исследования – нахождение функции $\varphi(x)$.

Вводится основное математическое допущение, «условие локальной независимости». Оно заключается в том, что если взяты два вопроса, то для индивида в точке X вероятность положительно ответить одновременно на оба вопроса, которую обозначим $f_{ij}(x)$, равна произведению вероятностей положительного ответа на каждый вопрос:

$$f_{ij}(x) = f_i(x)f_j(x).$$

В общем виде, если взято вопросов, уравнение принимает вид:

$$f_{ij}(x) = f_i f_j \dots f_k = \prod_{i=1}^k f_i(x).$$

В случае уравнения (3) мы для k вопросов получим следующую систему уравнений:

$$\int_{-\infty}^{\infty} f_i(x)\varphi(x)dx = p_{\sigma},$$

где σ – все наборы индексов $i, j \dots$

Общего решения эта система уравнений не имеет. В зависимости от условий, налагаемых на функции, получают те или иные модификации основного расчетного уравнения, которые называются моделями латентного анализа.

Вместо непрерывной функции плотности будем иметь k частот, которые соответствуют относительным объемам латентных классов. Обозначим их v^{α} , $\alpha = 1, \dots, k$, $\alpha = 1, \dots, k$. . Вместо непрерывного графика k -го вопроса получают отдельные вероятности для каждого класса, которые обозначим за p_i^{α} . Поскольку важной стороной модели латентных классов является число эмпирических данных и число латентных переменных, то необходимым условием существования решения системы латентных уравнений является тот факт, что число неизвестных должно быть не больше числа уравнений равных 2^r .

Имеем

$$\begin{cases} 1 = v^1 + \dots + v^k; \\ p_i = p_i^1 v^1 + \dots + p_i^k v^k; \\ p_{ij} = p_{ij}^1 v^1 + \dots + p_{ij}^k v^k. \end{cases}$$

В 1-й строке – 1 уравнение ($C_r^0 = 1$);
 во 2-й строке – r уравнений ($C_r^1 = r$);
 в 3-й строке – $\frac{r(r-1)}{2}$ уравнений
 $\left(C_r^2 = \frac{r(r-1)}{2} \right)$;

.....
 В i -й строке – C_r^i уравнений.

Всего r строк, и, следовательно, общее число уравнений равно сумме биномиальных коэффициентов:

$$1 + C_r^1 + \dots + C_r^r = 2^r.$$

Число неизвестных латентных параметров равно $k(r+1)$, поскольку r – число латентных вероятностей и k – число латентных частот в классах. Таким образом, необходимое (но недостаточное) условие разрешимости модели латентных классов соблюдается при $k(r+1) \leq 2r$. Вычислив значения маргиналов для классов, т.е. $p_1^1, p_2^1, p_3^1, p_1^2, p_2^2, p_3^2$ и т.д., можно получить частоты вариантов ответов для классов [9].

Экспериментальная часть работы.

Результаты и выводы

В качестве объектов исследования были выбраны пациенты с артериальной гипертензией 1-й стадии 1 степени и контрольная группа, состоящая из условно здоровых индивидуумов. Исходные данные из контрольной были сформированы в таблицу типа «объект – свойство» с объемом выборки 130 и размерности выборки 11. В результате были выделены следующие параметры объектов: сегментоядерные нейтрофилы (С), лимфоциты (Л), конечно-систолический размер левого желудочка (КСР), конечно-систолический объем левого желудочка (КСО), конечно-диастолический размер левого желудочка (КДР), конечно-диастолический объем левого желудочка (КДО), ударный объем (УО), минутный объем сердца (МОС), общее периферическое сосудистое сопротивление (ОПСС), фракция выброса левого желудочка (ФВ), фракция укорочения левого желудочка (ФУ).

Далее была вычислена соответствующая матрица коэффициентов корреляций между переменными, представленная на табл. 1.

Таблица 1

Матрица коэффициентов корреляций между переменными

| | С | Л | КСР | КСО | КДР | КДО | УО | МОС | ОПСС | ФВ | ФУ |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| С | 1,000 | -0,923 | 0,182 | 0,282 | 0,180 | 0,236 | 0,171 | 0,284 | -0,293 | -0,194 | 0,036 |
| Л | -0,923 | 1,000 | -0,239 | -0,318 | -0,271 | -0,262 | -0,187 | -0,266 | 0,198 | 0,299 | 0,124 |
| КСР | 0,182 | -0,239 | 1,000 | 0,716 | 0,737 | 0,830 | 0,848 | 0,777 | -0,631 | -0,357 | -0,422 |
| КСО | 0,282 | -0,318 | 0,716 | 1,000 | 0,786 | 0,934 | 0,775 | 0,736 | -0,600 | -0,401 | -0,418 |
| КДР | 0,180 | -0,271 | 0,737 | 0,786 | 1,000 | 0,871 | 0,850 | 0,627 | -0,612 | -0,182 | -0,266 |
| КДО | 0,236 | -0,262 | 0,830 | 0,934 | 0,871 | 1,000 | 0,949 | 0,864 | -0,721 | -0,167 | -0,247 |
| УО | 0,171 | -0,187 | 0,848 | 0,775 | 0,850 | 0,949 | 1,000 | 0,885 | -0,750 | 0,047 | -0,077 |
| МОС | 0,284 | -0,266 | 0,777 | 0,736 | 0,627 | 0,864 | 0,885 | 1,000 | -0,664 | 0,015 | -0,016 |
| ОПСС | -0,293 | 0,198 | -0,631 | -0,600 | -0,612 | -0,721 | -0,750 | -0,664 | 1,000 | 0,023 | 0,061 |
| ФВ | -0,194 | 0,299 | -0,357 | -0,401 | -0,182 | -0,167 | 0,047 | 0,015 | 0,023 | 1,000 | 0,799 |
| ФУ | 0,036 | 0,124 | -0,422 | -0,418 | -0,266 | -0,247 | -0,077 | -0,016 | 0,061 | 0,799 | 1,000 |

С помощью метода главных факторов найдено факторное отображение, представленное матрицей весовых нагрузок в табл. 2. Число факторов определялось по критерию Гутмана и критерию каменистой оси.

Факторы, образованные выделенными группами переменных проинтерпретированы следующим образом:

1) главный фактор можно интерпретировать как гемодинамический фактор, включающий параметры, описывающие центральную и периферическую гемодинамику. Переменные УО, МОС, ОПСС определяют уровень артериального давления. В норме, изменениям минутного объема циркуляции должна соответствовать адекватная по величине и направлению реакция прекапиллярного русла, которая бы нивелировала эти изменения и сохраняла среднее давление

на нормальном уровне. Например, если МО снижен, то артериолы должны сузиться. Если МО увеличен, то артериолы должны расшириться. Нарушения взаимосвязи этих показателей лежат в основе изменений уровня АД [12]. Вместе с тем изменение уровня артериального давления взаимосвязано с модуляцией сердца, за которую отвечают параметры КСР, КСО, КДР, КДО;

2) иммунологический фактор, который может отражать психосоматическое состояние индивида, поскольку этот фактор активируется в стрессовых состояниях. Основной вклад в формирования этого фактора вносят сегментоядерные нейтрофилы и лимфоциты;

3) фактор, составленный из параметров Фракция выброса левого желудочка и Фракция укорочения ЛЖ, можно считать важным для непосредственной оценки контрактильной (со-

кратительной, нагнетательной) функции левого желудочка. Этот фактор определяет объемную ресурсоемкость ЛЖ. Он показывает, насколько

использованы объемные резервы самого сердца для поддержания уровня артериального давления.

Таблица 2

Факторное отображение

| | Факторы | | | |
|---------------------|---------|----------------|----------------|----------------|
| | | F ₁ | F ₂ | F ₃ |
| Исходные переменные | С | 0,307808 | -0,76225 | -0,12315 |
| | Л | -0,35731 | 0,673376 | 0,272011 |
| | КСР | 0,884942 | -0,03066 | -0,03746 |
| | КСО | 0,911006 | -0,03558 | -0,14912 |
| | КДР | 0,86562 | 0,00654 | 0,115139 |
| | КДО | 0,972613 | -0,00016 | 0,124654 |
| | УО | 0,923934 | 0,029811 | 0,342506 |
| | МОС | 0,848804 | -0,10817 | 0,258131 |
| | ОПСС | -0,74 | 0,160976 | -0,31675 |
| | ФВ | -0,27441 | 0,170288 | 0,834728 |
| | ФУ | -0,35026 | -0,12602 | 0,770578 |

Выделенные факторы позволяют указать группу параметров, на которую нужно воздействовать, чтобы получить максимальный эффект от лечения. Например, для стабилизации уровня артериального давления следует воздействовать на всю группу признаков, описывающих гемодинамический фактор. При этом следует учитывать ремоделирование сердца (структурно-геометрическое состояние) при формировании патофизиологических взаимоотношений в системе кровообращения у пациентов с гипертонической болезнью [1]. Исключение стрессовых ситуаций пациентом позволит улучшить показатели, формирующие иммунологический фактор, а также нормализовать фактор, характеризующий уровень слаженности работы сердца и легких.

На базе выше представленной матрицы факторного отображения получены факторные значения 300 респондентов и при помощи латентного анализа были сформированы 8 латентных классов по соответствующим степеням влияния выделенных факторов. Для каждого из полученных латентных классов были предложены свои лечебные рекомендации в зависимости от степени влияния того или иного фактора.

Список литературы

1. Чазова И.Е. и др. Диагностика и лечение артериальной гипертензии. – М.: Рекомендации Российского медицинского общества по артериальной гипертонии и Всероссийского научного общества кардиологов. – 2008.
2. Иберла К. Факторный анализ: пер. с нем. В.М. Ивановой. – М.: Статистика, 1980. – С. 397.
3. Lazarsfeld P.F. The logical and mathematical foundation of latent structure analysis. – In: Measurement and Prediction. N. Y., 1950.
4. Thurstone L.L. Multiple factor analysis. – 6th. ed. – Chicago, 1961.
5. Zurmuhl R. Matrizen und ihre technischen Anwendungen. – Berlin, 1964.
6. Harman H. Modern factor analysis. – Chicago, 1967; Русский перевод: Харман Г. Современный факторный анализ. – М.: Статистика, 1972. – С. 483.
7. Hotelling H. Analysis of a complex of statistical variables into principal components / Hotelling H. // J. educ. Psychol. – 1993. – P. 24, 417–441., 498–520.
8. Wulsten A.R. Komponentenanalyse. Diss. Staatswirtsch. Fak. d. Univ. Munchen, 1960.
9. Bollen K.A. Structural Equations with Latent Variables. N.Y.: John Wiley & Sons, 1989.
10. Duncan O.D. Path Analysis: Sociological Examples // The American Journal of Sociology. – July 1966. – Vol. 72. – № 1. – P. 1–16.
11. Осипов Г.В. Методы измерения в социологии. – М.: Наука, 2003.
12. Фофанов П.Н. Учебное пособие по механокардиографии. – Л.: Военно-медицинская академия имени С.М. Кирова, 1977.