УДК 519.237.7, 614.1

ФАКТОРНЫЕ И ЛАТЕНТНЫЕ МОДЕЛИ В ДИАГНОСТИКЕ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИИ

Гольтяпин В. В.[1], Лобачев А. И.[2]

[1]*Федеральное государственное бюджетное учреждение науки Институт математики им. С. Л. Соболева Сибирского отделения Российской академии наук, Новосибирск, Россия (630090, проспект академика Коптюга, 4)*
[2]*ФГБОУ ВПО Омский государственный университет им. Ф. М. Достоевского, Омск, Россия (644077, ул. пр. Мира, д. 55а)*

В данной работе исследуется феномен артериальной гипертензии. Цель данной работы – разработка метода диагностики артериальной гипертензии первой степени и первой стадии с помощью факторной и латентно-структурной моделей. Первичной задачей исследования являлось выявление интегративных латентных характеристик, опирающихся на взаимозависимость измеряемых показателей. Вторая задача данного исследования заключалась в формировании латентных классов пациентов согласно выявленным латентно-интегративным характеристикам (факторам). В качестве математического аппарата использовались модели и алгоритмы факторного и латентного анализа. Разработан, предложен и апробирован метод диагностики артериальной гипертензии 1-ой стадии и первой степени. В качестве объектов исследования выбирались пациенты с артериальной гипертензией и условно здоровые индивидуумы. Для каждой группы был произведен расчет факторных значений и найдены соответствующие латентные группы.

Ключевые слова: интегративные характеристики, артериальное давление, артериальная гипертензия, метод главных компонент, латентный анализ, факторная модель.

FACTOR AND LATENT MODELS IN THE DIAGNOSIS OF ARTERIAL HYPERTENSION

Goltyapin V. V.[1], Lobachev A. I.[2]

[1]*Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences Academy of Sciences, Novosibirsk, Russia (630090, 4 Acad. Koptyug avenue)*
[2]*Omsk State University of F. M. Dostoevsky, Omsk, Russia (644077, prospekt Mira street, 55a)*

In this article we investigate the phenomenon of hypertension. The purpose of of this work - the development of a method of diagnostics of arterial hypertension of the first degree and the first stage by the factor and latent structural models. The primary objective of the study was to identify latent integrative characteristics based on the interdependence of measurable indicators. The second objective of this study was to form latent class of patients according to the identified latent integrative characteristics (factors). As a mathematical tool used models and algorithms, and latent factor analysis. Designed, proposed and tested method of diagnosis of hypertension first stage and the first degree. The objects of study were selected patients with hypertension and relatively healthy individuals. For each group, we calculated the factor scores and find the corresponding latent group.

Key words: integrative characteristics, blood pressure, hypertension, principal component analysis, latency analysis, factor model.

Введение

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

Гипертоническая болезнь как одна из форм артериальной гипертензии – хроническое заболевание, основным клиническим признаком которого является длительное и стойкое повышение артериального давления. Согласно трехстадийной 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.

Цель данной работы – разработка метода диагностики артериальной гипертензии первой степени и первой стадии с помощью факторной и латентно-структурной моделей.

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 науки. Как правило, эти данные не поддаются непосредственному изучению. Также факторный анализ помогает свести обширный статистический числовой материал к нескольким простым зависимостям.

Вторая задача данного исследования заключалась в формировании латентных 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) характеристики, присущей индивидам. Задача метода заключается в том, чтобы, изучив наблюдаемое поведение индивидов, вывести эту скрытую характеристику и классифицировать индивидов по сходству ее значений.

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 m-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 ограничения

$$R - U^2 = A \cdot C \cdot A^T, \tag{1}$$

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_h выглядит следующим образом:

$$R_h = A \cdot A^T. \tag{2}$$

Максимизация дисперсии достигается при выполнении ограничения (2), если в качестве матрицы A положить произведение двух первых матриц сингулярного разложения матрицы R_h. В принципе, можно вычислить столько главных компонентов, сколько имеется переменных, но это неэкономично. Чтобы обобщить информацию, содержащуюся в исходных переменных, лучше выделить небольшое количество факторов. Для определения количества факторов предлагается несколько способов: анализ собственных значений корреляционной матрицы с применением критерия Гуттмана и критерия «каменистой осыпи»; определение процента объясненной дисперсии; метод расщепления и критерий значимости. В нашем случае были использованы критерий Гуттмана и критерий «каменистой осыпи». То есть если собственные значения матрицы R_h меньше 1 и точка на графике лежит близко к основанию, то это означает возможность исключения данного фактора. После того, как найдена матрица A, производится расчет факторных значений для каждой группы показателей.

Известно, что в факторном пространстве однородные факторные значения не превышают трех. Это связано с тем, что математическое ожидание равно нулю, а дисперсия и, следовательно, стандартные отклонения единичные. 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 интервал (-3; 3). Если факторное значение попадает в этот интервал, то ставим «$+$» при формировании соответствующей номинальной шкалы, и «$-$» в противоположном случае. В том случае, если, например, выделяется 2 фактора, то возможно выделение следующих групп: «$+,+$», «$+,-$», «$-,+$», «$-,-$». Все индивидуумы с двумя плюсами считаются условно здоровыми, а 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 таким образом таблица с данными передается для обработки латентным анализом, который рассматривает её как некий тест или анкету.

Суть латентного анализа состоит в обработке теста или анкеты, состоящей из rвопросов, которые относятся к изучаемой скрытой характеристике. Выделенные вопросы называют явными переменными, а скрытую характеристику – латентной переменной или структурой. В теории тестов скрытая характеристика интерпретируется как одномерный латентный континуум (непрерывная латентная переменная).

В латентном анализе вводится функция вопросов, которую будем обозначать за $f_i(x)$. Это вероятность положительного ответа индивида на i-й вопрос, при условии, если индивид находится в точке x латентного 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, введем следующие величины:

$f_i(x)$ – 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 p_i.

Отсюда основное расчетное уравнение латентного анализа:

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

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, находящегося в точке x данного континуума, вероятность положительного ответа на оба вопроса равна произведению вероятностей положительно ответить на каждый вопрос:

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

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

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

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

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

где σ – все наборы индексов i, j ...

Общего решения эта система уравнений не имеет. В зависимости от условий, налагаемых на функции, получаются 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 $\upsilon^{\alpha}, \alpha = 1, \dots, k$. Вместо непрерывного графика i-го вопроса получаются 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 . Следовательно, формируем следующую систему уравнений:

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

Итак, в общем случае получаем: в 1-й строке – одно 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 $\left(C_r^2 = \frac{r(r-1)}{2}\right)$; и т.д. То есть в каждой строке –C_r^i уравнений.

Всего r строк, и, следовательно, общее число уравнений равно сумме биноминальных 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 [6,7]. Таким образом, необходимое условие разрешимости модели латентных классов соблюдается при $k(r+1) = 2^r$.

В нашем случае, преобразовав полученные факторные значения в номинальную шкалу и вычислив значения маргиналов для классов, можно найти вероятности $p_1^1, p_2^1, p_3^1, p_1^2, p_2^2, p_3^2$ и частоты υ^1 и υ^2, решив следующую систему уравнений:

$$\begin{cases} p_1 = p_1^1\upsilon^1 + p_1^2\upsilon^2 \\ p_2 = p_2^1\upsilon^1 + p_2^2\upsilon^2 \\ p_3 = p_3^1\upsilon^1 + p_3^2\upsilon^2 \\ p_{12} = p_{12}^1\upsilon^1 + p_{12}^2\upsilon^2 \\ p_{13} = p_{13}^1\upsilon^1 + p_{13}^2\upsilon^2 \\ p_{23} = p_{23}^1\upsilon^1 + p_{23}^2\upsilon^2 \\ p_{123} = p_{123}^1\upsilon^1 + p_{123}^2\upsilon^2 \\ 1 = \upsilon^1 + \upsilon^2 \end{cases}.$$

Получив эти значения и применив формулу Байеса, можно разделить исследуемых респондентов на два латентных класса.

Экспериментальная часть работы. Результаты и выводы

В качестве объектов исследования выбирались пациенты с артериальной гипертензией 1-ой стадии 1 степени и контрольная группа, состоящая из условно здоровых индивидуумов. Исходные данные из контрольной группы были сформированы в таблицу типа «объект – свойство» с объемом выборки 35 и размерности выборки 11.

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). На базе этих измеряемых параметров была вычислена соответствующая матрица коэффициентов корреляций между переменными, представленная на таблице 1.

	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

Таблица 1. Матрица коэффициентов корреляций между переменными

С помощью метода главных факторов найдено факторное отображение, представленное матрицей весовых нагрузок в таблице 2. Число факторов определялось по критерию Гуттмана и критерию каменистой осыпи. 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. Нарушения взаимосвязи этих показателей лежат в основе

изменений уровня АД [1,8]. Вместе с тем, изменение уровня артериального давления взаимосвязано с модуляцией сердца, за которую отвечают параметры КСР, КСО, КДР, КДО;
2) фактор, составленный из параметров «Фракция выброса левого желудочка» и «Фракция укорочения левого желудочка», можно считать важным для непосредственной оценки контрактильной функции левого желудочка. Этот фактор определяет объемную ресурсоемкость левого желудочка. Он показывает, насколько использованы объемные резервы самого сердца для поддержания уровня артериального давления;
3) иммунологический фактор, который может отражать психосоматическое состояние индивида, поскольку этот фактор активируется в стрессовых состояниях. Основной вклад в формирование этого фактора вносят сегментоядерные нейтрофилы и лимфоциты.

	Факторы			
		F_1	**F_2**	**F_3**
Исходные переменные	**С**	0,370776	-0,41245	**-0,81652**
	Л	-0,404	0,528124	**0,714934**
	КСР	**0,890653**	-0,03196	0,202028
	КСО	**0,907448**	-0,11513	0,125641
	КДР	**0,868244**	0,081827	0,109605
	КДО	**0,970558**	0,146025	0,07668
	УО	**0,923965**	0,351747	0,027032
	МОС	**0,860275**	0,289195	-0,11113
	ОПСС	**-0,76617**	-0,2399	0,097026
	ФВ	-0,28428	**0,853501**	-0,32593
	ФУ	-0,32528	**0,70249**	-0,55446

Таблица 2. Факторное отображение.

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 при формировании патофизиологических взаимоотношений в системе кровообращения у пациентов с гипертонической болезнью. 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 130 респондентов с данной патологией и проведен латентный анализ. Результаты моделирования по поиску латентных структур представлены в таблице 3.

Маргиналы		Частоты и вероятности		Вариант ответа респондента	Вероятность принадлежности респондента к 1классу.
p_1	0,717558	v^1	0,826432	+++	0,994056
p_2	0,534351	v^2	0,173568	++-	0,966446
p_3	0,59542	p_1^1	0,763251	-++	0,981088
p_{12}	0,40458	p_2^1	0,631575	+-+	0,882417
p_{13}	0,442749	p_3^1	0,666667	+--	0,563771
p_{23}	0,351145	p_1^2	0,499992	--+	0,699497
p_{123}	0,267176	p_2^2	0,071425	-+-	0,899334
		p_3^2	0,256185	---	0,286154

Таблица 3. Основные результаты латентно-структурного моделирования.

Для каждой группы респондентов, попавших в соответствующий латентный класс, были предложены свои лечебные рекомендации в зависимости от степени влияния того или иного фактора.

Список литературы

1. Иберла К. Факторный анализ. – М.: Статистика, 1980. – С. 397.
2. Осипов Г. В. Методы измерения в социологии. – М.: Наука, 2003.
3. Фофанов П. Н. Учебное пособие по механокардиографии. – Ленинград: Военно-медицинская ордена Ленина Краснознаменная академия имени С. М. Кирова,1977.
4. Харман Г. Современный факторный анализ / Харман Г. – М.: Статистика, 1972. С. 483.
5. Чазова И. Е. и др. Диагностика и лечение артериальной гипертензии // Москва: Рекомендации Российского медицинского общества по артериальной гипертонии и Всероссийского научного общества кардиологов – 2008.
6. Bollen K. A. Structural Equations with Latent Variables. – N.Y.: John Wiley & Sons, 1989.
7. Lazarsfeld P. F. The logical and mathematical foundation of latent structure analysis // Measurement and Prediction. – N. Y., 1950.
8. Thurstone L. L. Multiple factor analysis / Thurstone L. L. – 6th. ed. – Chicago, 1961.

Рецензенты:

Нечаева Галина Ивановна, доктор медицинских наук, профессор, заведующий кафедрой внутренних болезней и семейной медицины ГБОУ ВПО ОмГМА Минздравсоцразвития России, г. Омск.

Топчий Валентин Алексеевич, доктор физико-математических наук, директор Омского филиала ФГБУН Института математики им. С. Л. Соболева СО РАН, г. Омск.