

МОДЕЛИРОВАНИЕ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИИ МЕТОДАМИ ДИСКРИМИНАНТНОГО АНАЛИЗА

студентка Е.В. СЕНЬКО

(Гродненский государственный университет им. Янки Купалы, Беларусь)

Введение. При помощи математических методов возможно построение моделей закономерностей каких-либо процессов или явлений в медицине. Математическое моделирование систем является вторым кардинальным направлением применения математики в медицине.

Здоровье для каждого – неоценимое богатство человека, это условие счастья. Всемирная Организация Здравоохранения (ВОЗ) определяет здоровье как состояние полного физического, духовного и социального благополучия, а не только отсутствие болезней и физических дефектов.

Артериальная гипертензия (АГ) – часто встречающееся заболевание в современном мире. Это заболевание имеет генетическую предрасположенность, характеризуется стойким хроническим повышением систолического и диастолического артериального давления.

Цель работы – построение классификационных функций для определения состояния здоровья человека на предмет артериальной гипертензии. Объект исследования – результаты пациентов-мужчин, находящихся на клиническом обследовании, а предмет – артериальная гипертензия.

Исходные данные и методика исследования. На начальном этапе исследования была сформирована система показателей, характеризующих пациентов-мужчин с АГ. После этого были применены методы прикладной статистики – факторный анализ: метод главных компонент, кластерный анализ: метод k-средних и дискриминантный анализ с пошаговым включением в статистическом пакете Statistica 8 [1, с. 22]. На следующем этапе была проведена классификация пациентов-мужчин и получены функции классификации.

Для анализа пациентов с артериальной гипертензией были использованы данные по 300 пациентам-мужчинам. В качестве системы показателей, характеризующих степень заболевания АГ, были взяты 9 показателей: вес, индекс массы тела (ИМТ), частота дыхания (ЧД), конечно-систолический размер левого желудочка (КСР), конечно-систолический объем левого желудочка (КСО), конечно-диастолический размер левого желудочка (КДР), конечно-диастолический объем левого желудочка (КДО) и артериальное давление: систолическое и диастолическое ($АД = САД/ДАД$).

Сначала был применен факторный анализ, с помощью которого пациенты были проранжированы. Далее в ходе кластерного анализа пациенты-мужчины были разбиты на 3 кластера [2]. Так, с помощью факторного и кластерного анализа категории респондентов были классифицированы по однородным кластерам. Дискриминантный анализ позволил выявить, какие переменные и как разделили входящий поток данных на группы [3, с. 48]. Классификационные функции, выведенные посредством дискриминантного анализа, представляет собой линейную комбинацию независимых переменных, с помощью которой можно наилучшим образом различить категории зависимой переменной.

Полученные результаты. Результаты разбиения пациентов-мужчин на кластеры следующие: кластер 1 составили пациенты, имеющие заболевание АГ, кластер 2 –

пациенты с незначительными нарушениями, кластер 3 – пациенты с нормальными медицинскими показателями. Для анализа и дальнейшего построения классификационных функций использовались: группирующая переменная – кластер, независимые переменные – показатели, характеризующие АГ. Был применен метод с пошаговым включением. В таблице 1 представлены коэффициенты функций классификации при переменных-показателях.

Таблица 1. – Коэффициенты функций классификации

Показатели	Кластер 1	Кластер 2	Кластер 3
Вес, кг	22,962	9,543	1,288
ИМТ, кг/м ²	-1,984	-8,862	20,354
ЧД	33,972	17,552	-4,333
КСР, мм	11,509	11,928	-37,722
КСО, мл	40,652	35,293	18,935
КДР, мм	10,628	5,982	5,243
КДО, мл	-16,525	-7,942	12,051
САД (мм рт. ст.)	-25,004	-16,150	-3,162
ДАД (мм рт. ст.)	-18,421	-13,104	-7,557
Константа	-27,247	-12,813	-4,261

Таким образом, были получены следующие классификационные функции:

$$F1 = -27,247 + 22,962 \cdot \text{вес} - 1,984 \cdot \text{ИМТ} + 33,972 \cdot \text{ЧД} + 11,509 \cdot \text{КСР} + 20,652 \cdot \text{КСО} + 10,628 \cdot \text{КДР} - 16,525 \cdot \text{КДО} - 25,004 \cdot \text{САД} - 18,421 \cdot \text{ДАД}; \quad (1)$$

$$F2 = -12,813 + 9,543 \cdot \text{вес} - 8,862 \cdot \text{ИМТ} + 17,552 \cdot \text{ЧД} + 11,928 \cdot \text{КСР} + 35,293 \cdot \text{КСО} + 5,982 \cdot \text{КДР} - 7,942 \cdot \text{КДО} - 16,150 \cdot \text{САД} - 13,104 \cdot \text{ДАД}; \quad (2)$$

$$F3 = -4,261 + 1,288 \cdot \text{вес} + 20,354 \cdot \text{ИМТ} - 4,333 \cdot \text{ЧД} - 37,722 \cdot \text{КСР} + 18,935 \cdot \text{КСО} + 5,243 \cdot \text{КДР} + 12,051 \cdot \text{КДО} - 3,162 \cdot \text{САД} - 7,557 \cdot \text{ДАД}. \quad (3)$$

С помощью функций классификации можно определить принадлежность классифицируемых наблюдений к определенному кластеру. Наблюдение будет отнесено к тому кластеру, для которого функция классификации примет наибольшее значение.

Классификационная матрица (таблица 2) показала, что построенная модель классификации правильно определяют экспертную оценку с точностью 95%. При этом лучше всего она определяет оценку для 3 кластера (98,7%), хуже для первого и второго, 95,1% и 91,8% соответственно. Кластер 3 составили 80 пациентов, 2 из них были классифицированы неверно; кластер 2 – 97, кластер 3 – 123 пациента.

Таблица 2. – Классификационная матрица

Номер кластера	Точность (%)	Кластер 1	Кластер 2	Кластер 3
1	95,122	117	6	0
2	91,837	6	90	2
3	98,734	0	1	78
Точность/кол-во пациентов	95	123	97	80

Классификационные функции (1)-(3) можно использовать следующим образом. Пусть некоторый пациент имеет значения показателей (наблюдаемые), приведенные в таблице 3. Пронормируем их также как и исходные данные и подставим в функции классификации (1)-(3).

Таблица 3. – Медицинские показатели пациента-мужчины

Показатели	Вес	ИМТ	ЧД	КСР	КСО	КДР	КДО	САД	ДАД
Наблюдаемые	105	26,783	19	39	60	62	161	158	98
Нормированные	0,753	0,753	0,553	0,714	0,714	0,841	0,710	0,750	0,387

Проведя вычисления, получаем: $F1 = -13,431$; $F2 = 13,3301$; $F3 = 3,89487$.

Так как $F2$ – максимальное значение, то пациент относится ко второму кластеру. Исходя из разбиения пациентов на кластеры, у рассматриваемого пациента имеются незначительные нарушения.

Заключение. В работе исследованы медицинские показатели пациентов-мужчин на предмет заболевания артериальной гипертензией методом пошагового включения дискриминантного анализа. Построено три однородные группы (кластера) пациентов-мужчин и соответствующие классификационные функции. Полученные результаты позволяют определить принадлежность новых пациентов к определенному кластеру, а значит, установить степень заболевания артериальной гипертензией.

Литература

1. Буреева, Н.Н. Многомерный статистический анализ с использованием ППП «STATISTICA» / Н.Н. Буреева. – Нижний Новгород, 2007. – 112 с.
2. Сенько, Е.В. Классификация пациентов с артериальной гипертензией методами прикладной статистики / Е.В. Сенько, О.Н. Будько // Матеріали VI Міжнародної науково-практичної конференції студентів, аспірантів та молодих вчених. – Херсон : ПП Вишемирський В. С., 2017. – С. 146–148.
3. Дронов, С.В. Многомерный статистический анализ : учеб. пособие / С.В. Дронов. – Барнаул, 2013. – 213 с.