

УДК 519.24:616.005

## КЛАССИФИКАЦИОННЫЕ ФУНКЦИИ ДЛЯ ЭКСПРЕСС-ДИАГНОСТИКИ СТЕПЕНИ ЗАБОЛЕВАНИЯ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИЕЙ



**О.Н. Будько**

*Доцент кафедры математического и  
информационного обеспечения  
экономических систем ГрГУ, кандидат  
физико-математических наук, доцент*



**Е.В. Сенько**

*Студентка ГрГУ*

*Гродненский государственный университет имени Янки Купалы, Республика Беларусь  
E-mail: budko\_on@mail.ru, helensenko@mail.ru*

**Аннотация.** В работе получены классификационные функции для пациентов-женщин и пациентов-мужчин, которые могут быть использованы для экспресс-диагностики степени заболевания артериальной гипертензией. При проведении расчетов использовались методы многомерного статистического анализа: факторный, кластерный и дискриминантный анализ. Исходными данными послужили результаты клинического обследования 300 пациентов-женщин и 300 пациентов-мужчин.

**Ключевые слова:** артериальная гипертензия, классификационные функции, многомерный статистический анализ.

**Введение.** Артериальная гипертензия (АГ) – одна из наиболее распространенных форм патологии в мире. Известно, что частота гипертензии составляет около 15%, а если включить в это число людей с пограничной артериальной гипертензией, то доля населения планеты с повышенным артериальным давлением возрастает до 25%, а среди людей в возрасте старше 65 лет – 50% и более. Актуальность проведения исследований обуславливается преобладанием артериальной гипертензии в структуре смертности и возрастающим удельным весом в структуре заболеваемости населения, широким распространением и огромной социальной значимостью этого заболевания [1].

Вопросам прогноза развития и клинического течения заболеваний в последние годы уделяется большое внимание. В большинстве случаев в медицинских исследованиях, как правило, используются статические математические прогнозные модели, построенные на данных, полученных при одномоментных исследованиях нескольких сравнительно небольших групп. Известны исследования, посвященные прогнозированию развития АГ и ишемической болезни сердца (ИБС) [2], развитию АГ у женщин массовых профессий в зависимости от возраста и характера профессиональной деятельности [3], развитию острого инфаркта миокарда у мужчин различных возрастных групп [4].

Исходя из исследования ряда работ, связанных с изучением и анализом АГ, можно сделать вывод, что по возрастным категориям зависимость проста и понятна – чем старше человек, тем больше вероятность появления у него патологий сосудистой системы.

Цель работы – построить классификационные функции для экспресс-диагностики степени заболевания артериальной гипертензией пациентов-женщин и пациентов-мужчин. Деление пациентов и классификационных функций на две группы по гендерному принципу обусловлено различием в нормах показателей для женщин и мужчин.

*Исходные данные и методика исследования.* Исходными данными послужили результаты обследований 600 пациентов – 300 мужчин и 300 женщин. В качестве системы показателей, характеризующих степень заболевания АГ, были взяты: 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) – всего 9 показателей.

Методика проведенного исследования состоит из следующих этапов:

- 1) формирование системы показателей и их предварительный анализ;
- 2) нормирование исходных данных;
- 3) построение интегрального показателя методом главных компонент факторного анализа;
- 4) кластеризация пациентов по степени заболевания (женщин и мужчин) методом k-средних кластерного анализа;
- 5) построение классификационных функций с использованием дискриминантного анализа.

Показатель Артериальное давление (систолическое и диастолическое – верхнее и нижнее) задавалось отдельно (САД и ДАД). Предварительно исходные данные были нормированы от 0 до 1 и приведены к одному направлению (чем больше, тем лучше) по формулам (1).

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad \text{или} \quad z_i = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}}. \quad (1)$$

Расчеты проводились в пакете Statistica.

*Полученные результаты.* Непосредственное разбиение на кластеры после нормировки не дало однозначных положительных результатов [5]. Поэтому сначала данные методом главных компонент факторного анализа [6, с. 47] были преобразованы в два главных фактора для пациентов-женщин и пациентов-мужчин (таблица 1).

Таблица 1

Факторные нагрузки и дисперсии

| Показатели               | Пациенты-женщины |               | Пациенты-мужчины |               |
|--------------------------|------------------|---------------|------------------|---------------|
|                          | Фактор 1         | Фактор 2      | Фактор 1         | Фактор 2      |
| Вес, кг                  | 0,2832           | <b>0,9158</b> | <b>0,7865</b>    | 0,4088        |
| ИМТ, кг/м                | 0,4909           | <b>0,7926</b> | <b>0,8085</b>    | 0,5710        |
| ЧД                       | <b>0,7036</b>    | 0,4208        | <b>0,8777</b>    | 0,4514        |
| КСР, мм                  | <b>0,8146</b>    | 0,4175        | <b>0,8557</b>    | 0,4973        |
| КСО, мл                  | <b>0,8273</b>    | 0,4324        | <b>0,8533</b>    | 0,4759        |
| КДР, мм                  | <b>0,8534</b>    | 0,3736        | <b>0,7752</b>    | 0,5654        |
| КДО, мл                  | <b>0,8766</b>    | 0,3007        | 0,4264           | <b>0,8359</b> |
| САД (мм рт. ст.)         | <b>0,8440</b>    | 0,3838        | 0,5564           | <b>0,7825</b> |
| ДАД (мм рт. ст.)         | <b>0,8424</b>    | 0,3529        | 0,5404           | <b>0,7921</b> |
| Сохраняемая дисперсия, % | 76,87            | 7,46          | 87,77            | 4,63          |
| Собственные числа        | 6,919            | 0,671         | 7,899            | 0,417         |

Для пациентов-женщин первый главный фактор сохраняет 76,88% дисперсии, второй – 7,46%; накопленная дисперсия составляет 84,33%. Первый главный фактор составляют показатели: ЧД, КСР, КСО, КДР, КДО, САД и ДАД, второй – вес и ИМТ.

Для пациентов-мужчин первый и второй фактор сохраняют 87,77% и 4,63% дисперсии соответственно, накопленная дисперсия – 92,40%. Наибольший вклад в образование первого

главного фактора вносят показатели: вес, ИМТ, ЧД, КСР, КСО, КДР, второго – КДО, САД, ДАД.

Для построения интегрального показателя по каждой группе рассчитывались рейтинговые числа путем умножения значений главного фактора (*factor scores*) на собственные числа, соответствующие главным факторам по формулам (2).

$$R_{жен} = 6,919 \cdot F_{1,жен} + 0,671 \cdot F_{2,жен}, \quad R_{муж} = 7,899 \cdot F_{1,муж} + 0,417 \cdot F_{2,муж}. \quad (2)$$

Так как количество пациентов достаточно большое: 300 женщин и 300 мужчин, представить интегральные показатели по ним не представляется возможным. Интегральные показатели можно использовать следующим образом:

– построить рейтинг пациентов по группам по степени заболевания артериальной гипертензией;

– использовать для дальнейших исследований, например, для формирования групп пациентов (кластеров) со схожим диагнозом.

Для реализации последнего применим метод k-средних кластерного анализа. Пациенты-женщины и пациенты мужчины были сгруппированы в 3 кластера, допускающие однозначную интерпретацию.

В таблице 2 представлены средние значения показателей в кластерах для пациентов-женщин, их интерпретация содержится в таблице 3. В первую группу вошли пациенты-женщины с нормальными значениями показателей, во вторую – имеющие незначительные нарушения (нн) показателей, а третью группу составили пациенты-женщины с заболеванием АГ (умеренные нарушения). Группы пациентов со значительными нарушениями показателей не выявлено. Средние значения показателей в кластерах достоверно различны по критерию Шеффе ( $p < 0,05$ ).

Таблица 2

Средние значения показателей в кластерах для пациентов-женщин

| Кластеры | Кол-во пациентов | Вес, кг | ИМТ, кг/м | ЧД   | КСР, мм | КСО, мл | КДР, мм | КДО, мл | САД (мм рт. ст.) | ДАД (мм рт. ст.) |
|----------|------------------|---------|-----------|------|---------|---------|---------|---------|------------------|------------------|
| 1        | 73               | 67,5    | 20,1      | 16,3 | 29,2    | 33,3    | 45,6    | 79,35   | 132,1            | 84,9             |
| 2        | 72               | 78,6    | 25,6      | 18,7 | 36,8    | 56,2    | 56,1    | 112,3   | 150,7            | 94,9             |
| 3        | 155              | 96,3    | 31,9      | 19,7 | 39,8    | 66,3    | 60,2    | 125,5   | 173,1            | 106,0            |

Таблица 3

Интерпретация средних значений показателей пациентов-женщин

| Класс-теры | Кол-во пациентов | Вес, кг | ИМТ, кг/м | ЧД    | КСР, мм | КСО, мл | КДР, мм | КДО, мл | САД (мм рт. ст.) | ДАД (мм рт. ст.) |
|------------|------------------|---------|-----------|-------|---------|---------|---------|---------|------------------|------------------|
| 1          | 73               | норма   | норма     | норма | норма   | норма   | норма   | норма   | норма            | норма            |
| 2          | 72               | нн      | нн        | нн    | нн      | нн      | нн      | нн      | нн               | нн               |
| 3          | 155              | болен   | болен     | болен | болен   | болен   | болен   | болен   | болен            | болен            |

Заметим, что около половины пациентов-женщин попали в третий кластер «болен».

Аналогичные результаты были получены после классификации пациентов-мужчин по интегральному показателю (таблица 4 – таблица 5). Средние значения показателей в кластерах также достоверно различны по критерию Шеффе ( $p < 0,05$ ).

Из таблицы 5 следует, что около половины пациентов-мужчин попали в первый кластер с нормальными значениями показателей, 2-ой кластер составили пациенты с незначительными нарушениями и 3-ий кластер – с умеренными и значительными нарушениями (болен).

На основании проведенных классификаций пациентов методом пошагового включения дискриминантного анализа были построены классификационные функции [6, с. 48], коэффициенты которых представлены в таблице 6.

Таблица 4

Средние значения показателей в кластерах для пациентов-мужчин

| Клас-теры | Кол-во па-циен-тов | Вес, кг | ИМТ, кг/м | ЧД   | КСР, мм | КСО, мл | КДР, мм | КДО, мл | САД (мм рт.ст.) | ДАД (мм рт.ст.) |
|-----------|--------------------|---------|-----------|------|---------|---------|---------|---------|-----------------|-----------------|
| 1         | 146                | 69,0    | 20,3      | 16,6 | 29,6    | 40,3    | 49,8    | 110,8   | 129,8           | 84,4            |
| 2         | 74                 | 90,7    | 26,6      | 19,4 | 38,5    | 63,6    | 61,1    | 157,9   | 145,1           | 92,9            |
| 3         | 80                 | 111,6   | 33,2      | 21,4 | 43,8    | 79,8    | 69,1    | 196,2   | 180,8           | 111,9           |

Таблица 5

Интерпретация средних значений показателей пациентов-мужчин

| Клас-теры | Кол-во па-циен-тов | Вес, кг | ИМТ, кг/м | ЧД    | КСР, мм | КСО, мл | КДР, мм | КДО, мл | САД (мм рт.ст.) | ДАД (мм рт.ст.) |
|-----------|--------------------|---------|-----------|-------|---------|---------|---------|---------|-----------------|-----------------|
| 1         | 146                | норма   | норма     | норма | норма   | норма   | норма   | норма   | норма           | норма           |
| 2         | 74                 | нн      | нн        | нн    | нн      | нн      | нн      | нн      | нн              | нн              |
| 3         | 80                 | болен   | болен     | болен | болен   | болен   | болен   | болен   | болен           | болен           |

Таблица 6

Коэффициенты функций классификации

| Показатели      | Для пациентов-женщин |           |           | Для пациентов-мужчин |           |           |
|-----------------|----------------------|-----------|-----------|----------------------|-----------|-----------|
|                 | Кластер 1            | Кластер 2 | Кластер 3 | Кластер 1            | Кластер 2 | Кластер 3 |
| ИМТ, кг/м       | 5,4                  | 19,9      | 11,8      | 8,6                  | -28,6     | -1,8      |
| КСР, мм         | 20,4                 | 49,5      | 42,1      | -7,8                 | 94,2      | 63,2      |
| САД (мм рт.ст.) | 1,1                  | 12,4      | -0,8      | 0,2                  | 3,9       | 25,1      |
| КДО, мл         | 21,4                 | 52,9      | 48,1      | 11,9                 | 23,8      | 27,6      |
| ЧД              | 4,9                  | 17,5      | 14,5      | -4,6                 | -31,9     | -16,0     |
| КСО, мл         | 10,1                 | 35,8      | 29,2      | 16,6                 | -5,3      | -7,8      |
| ДАД (мм рт.ст.) | -11,7                | -19,4     | -20,6     | -5,0                 | -4,9      | 10,1      |
| вес, кг         | 3,1                  | 9,6       | 5,7       | 3,6                  | 9,7       | 15,0      |
| КДР, мм         | 22,4                 | 65,4      | 56,8      | 6,9                  | 16,4      | 16,8      |
| Constant        | -13,0                | -103,4    | -67,4     | -5,4                 | -28,0     | -57,2     |

Классификационные функции для пациентов-женщин они имеют вид (1)-(3).

$$KF1_{жен} = 5,4 \cdot ИМТ + 20,4 \cdot КСР + 1,1 \cdot САД + 21,4 \cdot КДО + 4,9 \cdot ЧД + 10,1 \cdot КСО - 11,78 \cdot ДАД + 3,1 \cdot Вес + 22,4 \cdot КДР - 13,0; \quad (3)$$

$$KF2_{жен} = 19,9 \cdot ИМТ + 49,5 \cdot КСР + 12,4 \cdot САД + 52,9 \cdot КДО + 17,5 \cdot ЧД + 35,8 \cdot КСО - 19,4 \cdot ДАД + 9,6 \cdot Вес + 65,4 \cdot КДР - 103,4; \quad (4)$$

$$KF3_{жен} = 11,8 \cdot ИМТ + 42,1 \cdot КСР - 0,8 \cdot САД + 48,1 \cdot КДО + 14,5 \cdot ЧД + 29,2 \cdot КСО - 20,6 \cdot ДАД + 5,7 \cdot Вес + 56,8 \cdot КДР - 67,4; \quad (5)$$

Для проверки качества дискриминации использована статистика лямбда Уилкса (чем ближе к 0, тем меньше вероятность ошибочного разделения). Так как значение лямбды Уилкса равно 0,042, следовательно, данная классификация корректна.

Из значений матрицы классификации следует, что построенные модели (функции) правильно определяют экспертную оценку с точностью 98,3 %. При этом лучше всего она определяет оценку для второго кластера (100%), хуже для первого (98,6%) и третьего (94,4%).

Классификационные функции для пациентов-мужчин имеют вид (6)-(8):

$$KF1_{муж} = 8,6 \cdot ИМТ - 7,8 \cdot КСР + 0,2 \cdot САД + 11,9 \cdot КДО - 4,6 \cdot ЧД + 16,6 \cdot КСО - 5,0 \cdot ДАД + 3,6 \cdot Вес + 6,9 \cdot КДР - 5,4; \quad (6)$$

$$KF2_{муж} = -28,6 \cdot ИМТ + 94,2 \cdot КСР + 3,9 \cdot САД + 23,8 \cdot КДО - 31,9 \cdot ЧД - 5,3 \cdot КСО - 4,9 \cdot ДАД + 9,7 \cdot Вес + 16,4 \cdot КДР - 28,0; \quad (7)$$

$$KF3_{муж} = -1,8 \cdot ИМТ + 63,2 \cdot КСР + 25,1 \cdot САД + 27,6 \cdot КДО - 16,0 \cdot ЧД - 7,8 \cdot КСО + 10,1 \cdot ДАД + 16,0 \cdot Вес + 16,8 \cdot КДР - 57,2; \quad (8)$$

Построенные модели правильно определяют экспертную оценку с точностью 98%. Самую высокую точность, равную 100 %, имеет функция классификации третьего кластера, 99,3 % – второго, 93,2% – третьего.

Построенные функции классификации можно использовать для экспресс-анализа степени заболевания артериальной гипертензией. После подстановки значений показателей нового пациента, он относится к тому кластеру, для которого функция классификации примет наибольшее значение.

Пусть некоторый пациент имеет значения показателей, приведенные в таблице 7.

Таблица 7

Медицинские показатели пациента-женщины

| ИМТ   | КСР, мм | САД, мм рт.ст. | КДО, мл | ЧД | КСО, мл | ДАД, мм рт.ст. | Вес, кг | КДР, мм |
|-------|---------|----------------|---------|----|---------|----------------|---------|---------|
| 17,43 | 32      | 133            | 69      | 15 | 25      | 80             | 59      | 53      |

Подставляя эти значения в классификационные функции (3)-(5) пациентов-женщин, получаем:

$$KF1_{жен} = 3\,116,922; \quad KF2_{жен} = 10\,764,857; \quad KF3_{жен} = 9\,108,274.$$

Так как  $KF2_{жен}$  имеет максимальное значение, то пациент относится ко второму кластеру. Исходя из таблицы 3, у рассматриваемого пациента имеются незначительные нарушения.

**Заключение.** Таким образом, используя технику методов многомерного статистического анализа (факторного, кластерного и дискриминантного анализа) построены классификационные функции для пациентов-женщин и пациентов мужчин, которые могут служить для экспресс-диагностики степени заболевания артериальной гипертензией.

#### Список литературы

- [1]. Статистика ВОЗ по распространенности заболевания артериальной гипертензией [Электронный ресурс]. – Режим доступа: <http://vashfleblog.ru/arterial-pressure/hypertension/arterialnaya-gipertenziya-statistika-voz.html>. – Дата доступа – 30.03.2018.
- [2]. Мелехов А.В., Рязанцева Е.Е. Сочетание артериальной гипертензии и ишемической болезни сердца: улучшение прогноза возможно // Лечебное дело. 2014. №1. С. 49–54.
- [3]. Артамонова Г.В., Максимов С.А., Индукаева Е.В. и др. Прогнозирование артериальной гипертензии у женщин в зависимости от возраста и характера трудовой деятельности // Бюллетень сибирской медицины. 2011. Т. 10, №4. С. 141–145.
- [4]. Особенности клинического течения острого инфаркта миокарда у мужчин разного возраста [Электронный ресурс]. – Режим доступа: <http://www.docme.ru/doc/1410029.pdf>. – Дата доступа – 30.03.2018.
- [5]. Сенько, Е.В. Классификация пациентов с артериальной гипертензией методами прикладной статистики / Е.В. Сенько, О.Н. Будько // Молодь у світі сучасних технологій за тематикою: Економіко-математичні моделі та інформаційні технології розвитку регіону: Матеріали VI Міжнародної науково-практичної конференції студентів, аспірантів та молодих вчених (Херсон 1-2, червня 2017 р.) / за заг. ред. Н.А. Соколової. Херсонський національний технічний університет. – Херсон: ПП Вишемирський В. С., 2017. – С. 146–148.
- [6]. Факторный, дискриминантный и кластерный анализ: пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка [и др.]; под ред. И.С. Енюкова. М.: Финансы и статистика, 1989. – 215 с.

## CLASSIFICATION FUNCTIONS FOR EXPRESS DIAGNOSTICS OF THE DEGREE OF DISEASE WITH ARTERIAL HYPERTENSION

***O.N. BUDKO, PhD***

*Associate Professor of the Department of Mathematical and Information Support of Economic Systems, Associate Professor*

***E.V. SENKO***

*Student Yanka Kupala State University of Grodno*

*Yanka Kupala State University of Grodno, Republic of Belarus*  
*E-mail: budko\_on@mail.ru, helensenko@mail.ru*

**Abstract.** The classification functions for female patients and male patients were obtained, which can be used for express diagnostics of the degree of arterial hypertension. For the calculations, multidimensional statistical analysis methods were used: factor, cluster and discriminant analysis. The initial data were the results of a clinical examination of 300 female patients and 300 male patients.

**Key words:** arterial hypertension, classification functions, multivariate statistical analysis.