



Применение дискриминантного анализа в профилактической медицине

Зубов С. А., Тихонова О. А.

ФГБУ «Государственный научный Центр Российской Федерации – Федеральный медицинский биофизический центр имени А. И. Бурназяна», Москва, Российская Федерация

Аннотация

Актуальность. Дискриминантный анализ (от латинского *discriminatio* — различение) позволяет классифицировать объекты по пороговому значению (константе дискриминации). В статье представлен клинический пример прогнозирования точности системы переменных-предикторов с помощью модели анализа дискриминантной функции.

Цель. Обосновать целесообразность использования дискриминантного анализа в качестве статистического инструмента для мониторинга здоровья и экспертизы трудоспособности работников радиационно-опасных производств.

Материалы и методы. В выборке 232 человек с верифицированным токсическим пылевым бронхитом и бериллиозом методом линейного дискриминантного анализа изучали предикторы: уровень фибриногена, форсированная жизненная ёмкость лёгких, уровень лактатдегидрогеназы, время прохождения электрического импульса от синусового узла до желудочков на электрокардиограмме (интервал PQ). У 124 пациентов значения предикторов выходили за пределы референсных значений (группа «больные»), у 108 человек результаты лабораторно-инструментальных исследований были нормальными (группа «нормальные»). Дополнительно для проверки рабочей гипотезы рассмотрены 23 клинических случая без уточнённой категории (N=23): «патология»/«норма». Чтобы определить принадлежность нового пациента к той или иной группе, по формуле $V_{new} = New \times V$ рассчитывали значения дискриминантных функций для новых объектов (пациентов) используя входные данные (результаты тестов) с помощью вектора оценок дискриминации. Полученный результат сравнивали с константой.

Результаты. Константа дискриминации, полученная в результате расчётов, позволила отнести тестируемых лиц (N=23) к категориям «больные» или «норма» с точностью 91,3 %.

Выводы. Анализ дискриминантной функции целесообразно использовать в здравоохранении и клинической практике для решения задач в чётко определённых группах с заранее выбранными параметрами. Оправдано его применение и в комбинации с другими статистическими методами, поскольку он позволяет эффективно классифицировать диагнозы пациентов. Ограничением метода является нормальность распределения анализируемой выборки, сложность интерпретации результата в выбранной модели при работе с новыми данными.

Ключевые слова: дискриминантный анализ; модель прогнозирования статистические методы; профилактическая медицина; частота выявления

Для цитирования: Зубов С. А., Тихонова О. А. Применение дискриминантного анализа в профилактической медицине. *Реальная клиническая практика: данные и доказательства*. 2026;6(3):15–20. <https://doi.org/10.37489/2782-3784-myrd-108>. EDN: EPIBFI.

Поступила: 23.07.2026. В доработанном виде: 25.08.2026. Принята к печати: 23.09.2026. Опубликовано: 05.09.2026.

Discriminant analysis as a forecasting tool in disease prevention

Sergei A. Zubov, Olga A. Tikhonova

Research Center – Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russian Federation

Abstract

Relevance. Discriminant analysis (from the Latin *discriminatio* — “distinction”) allows objects to be classified according to a threshold value (the discrimination constant). The article presents a clinical example of predicting the accuracy of a system of predictor variables using a discriminant function analysis model.

Objective. To substantiate the feasibility of using discriminant analysis as a statistical tool for monitoring health and assessing work capacity among workers in radiation-hazardous industries.

Materials and methods. In a sample of 232 individuals with verified toxic dust bronchitis and berylliosis, predictors were studied using linear discriminant analysis: fibrinogen level, forced vital capacity (FVC), lactate dehydrogenase (LDH) level, and the time required for an electrical impulse to travel from the sinoatrial node to the ventricles on the electrocardiogram (PQ interval). In 124 patients, the predictor values were outside the reference ranges (the “patients” group); in 108 individuals, the results of laboratory and instrumental tests were within the normal range (the “non-patients” group). Additionally, to test the working hypothesis, 23 clinical cases without a specified category (N=23) were considered: “pathology”/“normal”. To determine which group a new patient belongs to, the values of the discriminant functions for new objects (patients) were calculated using the formula $V_{new} = New \times V$, applying the input data (test results) with the help of the discrimination score vector. The obtained result was compared with the constant.

Results. The discrimination constant obtained through calculations enabled the classification of the tested individuals (N=23) into the “patients” or “normal” categories with an accuracy of 91.3 %.

Conclusions. It is advisable to use discriminant function analysis in healthcare and clinical practice to address problems in clearly defined groups with pre-selected parameters. Its application is also justified in combination with other statistical methods, as it allows for the effective classification of patients’ diagnoses. The limitations of the method include the requirement for a normal distribution of the analyzed sample and the complexity of interpreting the results in the selected model when working with new data.

Keywords: discriminant analysis; forecasting model; statistical methods; preventive medicine; detection rate

For citation: Zubov SA, Tikhonova OA. Discriminant analysis as a forecasting tool in disease prevention. *Real-World Data & Evidence*. 2026;6(3):15–20. <https://doi.org/10.37489/2782-3784-myrd-108>. EDN: EPIBFI.

Received: 23.07.2026. **Revision received:** 25.08.2026. **Accepted:** 23.09.2026. **Published:** 05.09.2026.

Введение / Introduction

Прогнозирование рисков тех или иных событий в профилактической медицине — это не просто констатация фактов, а инструмент планирования ресурсов здравоохранения [1]. В связи с этим, совершенствование методологий предиктивной аналитики с использованием комплекса медико-социальных детерминант не теряет актуальности и расширяет возможности персонализированной профилактики. Вместе с этим неизменной проблемой остаётся поиск инструментов для построения прогнозных моделей. Примером предсказательных подходов являются мультиклассификаторы [2], комбинированная модель на основе независимых предикторов [3] и другие инструменты мультимедийного анализа, сочетающие современные цифровые технологии и давно известный дискриминантный анализ [4, 5].

В статье на примере пациентов с верифицированным диагнозом токсического пылевого бронхита и бериллиоза представлена модель анализа дискриминантной функции для прогнозирования точности системы переменных-предикторов. Известно, что традиционные технологии ядерной энергетики сопряжены с воздействием электромагнитного поля сверхвысокой частоты, малых доз ионизирующего излучения, интенсивными умственными и физическими нагрузками. Часто хронические заболевания у работников радиационно-опасных производств не имеют выраженных клинических проявлений [6, 7, 8]. Показания к лечебно-профилактическим и восстановительным мероприятиям в большинстве случаев определяются врачом-профпатологом,

при этом редко используют количественные значения уровней сохранности здоровья. Для разработки отличительного признака, позволяющего на этапе обязательного периодического осмотра быстро определить групповую принадлежность пациента, нуждающегося или не нуждающегося в дополнительном медицинском обследовании и восстановительном лечении, предлагаем использовать описанный ниже дискриминантный анализ (ДА).

Цель / Objective

Обосновать целесообразность использования ДА в качестве статистического инструмента для мониторинга здоровья и экспертизы трудоспособности работников особо опасных производств.

Материалы и методы / Materials and methods

Для статистического анализа ретроспективно применяли результаты обследования лиц, контактировавших с промышленными аэрозолями фиброгенного действия [Свидетельство о государственной регистрации базы данных № 2024625156. База данных больных, страдающих хронической бронхолегочной патологией, вызванной воздействием промышленных аэрозолей. Авторы: Тихонова О. А., Петрова И. А., Бушманов А. Ю., Царев А. Н. Дата регистрации 14.11.2024 года]. В качестве исходных данных для ДА выбраны следующие показатели: фибриноген (X1), форсированная жизненная ёмкость лёгких (ФЖЕЛ) (X2), лактатдегидрогеназа (ЛДГ) (X3), длительность интервала PQ на ЭКГ (X4)¹ Далее в тексте эти показатели указаны как

¹ Далее по тексту результаты тестов

«результаты теста». Общий размер выборки описываемого примера отражает итог обследования 232 чел., из которых у 124 чел. уровни фибриногена, ФЖЕЛ, ЛДГ и PQ (переменных-предикторов) выходили за референсные пределы. У 108 человек результаты теста были нормальными, клинические проявления болезни отсутствовали (ремиссия). Первую группу квалифицировали как «больных», вторую как «небольных». Дополнительно были рассмотрены 23 клинических случая без уточнённой категории: «патология»/«норма», которые использованы для проверки рабочей гипотезы. Для математической обработки данных применяли DA, предназначенный для изучения различий между двумя или большим количеством групп объектов с использованием данных о разнообразии нескольких признаков, отличающих эти объекты друг от друга².

При DA, в отличие от близкого метода исследования — классификации, число групп задают заранее. Основной задачей DA в нашем случае будет ответ на вопрос: «Если дан объект (пациент) с набором клиничко-лабораторных и инструментальных данных, полученных в результате медосмотра (тестирования), к какой когорте он относится: нуждающихся («больные») или не нуждающихся («небольные») в лечебно-профилактических мероприятиях?

Результаты и обсуждение / Results and discussion

В результате вычислений определяли, к какой генеральной совокупности (группе) отнести вновь обследованного пациента с известными нам переменными: к группе «больных» или условно здоровых («небольных»)? В используемом наборе данных зависимой переменной являлся статус человека (I — нуждается в лечении, II — не нуждается в лечении (стадия ремиссии)), а независимыми переменными были результаты тестов: фибриноген (X1), ФЖЕЛ (X2), ЛДГ (X3) и PQ (X4). Таким образом, все данные были сгруппированы в две первоначальные матрицы: D размера 124×4 и H размера 108×4³. Так как зависимая переменная имеет две категории (группы), мы рассматриваем пример двух-группового DA.

Все независимые переменные оказались распределёнными нормально⁴. На первом этапе мы рассчитывали средние значения по каждому признаку (независимой переменной) для двух групп по формуле:

$$\bar{x}_{jk} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ij}, \quad \text{где}$$

x_{ij} — значение j -того признака (результат тестов) для i -того объекта (пациента) в рассматриваемой группе (группа I «больные» и группа II «небольные»);

$\bar{x}_{jk}$ — среднее значение j -того признака (результат теста) для данной группы k («больные» или «небольные»);

n_k — общее количество объектов (пациентов) в группе k , где $k=1$ — группа I «больные», $k=2$ — группа II «небольные»;

$i=1,2,\dots,n_k$ — количество объектов в группах. В данном кейсе в группе I оказалось 124 объекта, $n_D=124$; а в группе II оказалось 108 объектов, $n_H=108$; $j=1,2,\dots,m$ — количество признаков (анализов) в рассматриваемом примере $m=4$ (фибриноген (X1), ФЖЕЛ (X2), ЛДГ (X3) и PQ (X4)).

Таким образом получили векторы средних по матрицам групп. На следующем шаге для каждой группы рассчитывали квадратные ковариационные матрицы размерностью 4×4 (по количеству признаков (результатов тестов)). Для этого рассчитали центрированные матрицы, где из каждого элемента каждого признака вычитают среднее значение по данному признаку. Формула расчёта:

$$M_{kc} = \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1m} - \bar{x}_m \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2m} - \bar{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n1} - \bar{x}_2 & \dots & x_{nm} - \bar{x}_m \end{pmatrix}, \quad \text{где}$$

M_{kc} — центрированные матрицы для двух групп («больные» и «небольные»);

x_{ij} — результаты тестов (признаки) для каждого пациента (объекта);

$\bar{x}_j$ — средние значения по каждому признаку (анализы).

И далее рассчитывают квадратные ковариационные матрицы по формуле:

$$M_k = \frac{1}{n_k - 1} (M_{kc}^T \times M_{kc}), \quad \text{где}$$

M_k — ковариационная матрица для группы k ;

n_k — общее количество объектов для группы k ;

M_{kc} — центрированная матрица;

M_{kc}^T — транспонированная центрированная матрица.

В итоге получают ковариационные матрицы для двух групп.

² Расчёты проведены в программе MS Excel

³ D — disease; H — health (англ.)

⁴ Использовался тест Шапиро-Уилка

На следующем этапе вычисляют несмещенную оценку суммарной ковариационной матрицы для двух групп:

$$M_{sum} = \frac{1}{n_D + n_H - 2} (n_D \times M_D + n_H \times M_H), \text{ где}$$

M_{sum} — несмещенная оценка суммарной ковариационной матрицы;

Полученную матрицу преобразуют в обратную матрицу ($M_{D,H}^{-1}$)⁵.

Далее находят вектор оценок коэффициентов дискриминации, рассчитываемый как разница между векторами средних по матрицам обеих групп («больные» и «небольные») умноженная на обратную матрицу, полученную на предыдущем шаге:

$$V = M_{D,H}^{-1} (D - H), \text{ где}$$

V — вектор оценок коэффициентов дискриминации; $M_{D,H}^{-1}$ — обратная матрица;

D — средние значения по признакам (результатам тестов) для группы I «больные»;

H — средние значения по признакам (результатам тестов) для группы II «небольные».

Таким образом, дискриминантная функция для рассматриваемого примера имеет следующий общий вид:

$$F_{1,2} = v_1 \times x_{i1} + v_2 \times x_{i2} + v_3 \times x_{i3} + v_4 \times x_{i4}, \text{ где}$$

x_{ij} — значение j -того признака (результат теста) для i -того объекта (пациента);

v_1, v_2, v_3, v_4 — значения вектора оценок коэффициентов дискриминации V .

Далее оценивают дискриминантную функцию путем умножения первоначальных данных на вектор оценок коэффициентов дискриминации (V), по которым рассчитывают средние значения. Из полученных значений вычисляют константу дискриминации как среднее значение по двум средним значениям векторов оценок дискриминантной функции.

$$V_D = D \times V; V_H = H \times V; \text{Const} = \frac{V_{Dcp} + V_{Hcp}}{2}, \text{ где}$$

D — первоначальная матрица по группе I «больные»;

H — первоначальная матрица по группе II «небольные»;

V_D — оценки дискриминантной функции по группе I «больные»;

V_H — оценки дискриминантной функции по группе II «небольные»;

V_{Dcp} — среднее значение всех элементов вектора V_D ;

V_{Hcp} — среднее значение всех элементов вектора V_H ;

Const — константа дискриминации.

Чтобы определить принадлежность нового пациента к группе «больные» или «небольные», берут подлежащие дискриминации входные данные (результаты тестов) и с помощью вектора оценок дискриминации рассчитывают значения дискриминантных функций для новых объектов (пациентов).

$$V_{new} = New \times V$$

Полученный результат сравнивают с константой и проводят распределение по следующему правилу:

- если $V_{new} > \text{Const}$, то при $V_{Dcp} > V_{Hcp}$ объект относится к группе I, а при $V_{Dcp} < V_{Hcp}$ к группе II;
- если $V_{new} < \text{Const}$, то при $V_{Dcp} > V_{Hcp}$ объект относится к группе II, а при $V_{Dcp} < V_{Hcp}$ к группе I.

В нашем исследовании результаты проведенного ДА совпали с результатами клинической практики на 91,3 % (21 случай из 23).

Заключение / Conclusion

Предложена интерпретация результатов ДА, используемого для прогнозирования точности системы переменных-предикторов, в упрощенной форме с использованием клинического примера для лучшего понимания. Представленный пример двухгруппового дискриминантного анализа данных, позволяющего с достаточно высокой долей уверенности распределять пациентов на две группы в зависимости от статуса: нуждающихся в назначении немедикаментозной восстановительной терапии с целью здоровьесбережения и не нуждающихся в дополнительном медицинском осмотре и лечении. Независимыми переменными в данном случае были результаты клинико-инструментальных тестов: фибриноген (X1), ФЖЕЛ (X2), ЛДГ (X3) и РQ (X4). Целесообразность применения дискриминантного анализа ограничивается нормальностью распределения анализируемой выборки, требованием заранее выбранных параметров групп, сложностью интерпретации результата в выбранной модели при работе с новыми данными. Таким образом, дискриминантный анализ уместен для решения задач в четко определенных группах, кроме того, оправдано его применение и в комбинации с другими статистическими методами.

⁵ Преобразование с использованием методов линейной алгебры (метод Гаусса, метод Крамера и др.)

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов в связи с данной публикацией.

Этическая экспертиза

При проведении работы авторы придерживались этических принципов медицинских исследований, описанных в Хельсинкской декларации Всемирной медицинской ассоциации «Этические принципы проведения медицинских исследований с участием человека в качестве объекта» последней редакции (октябрь 2024 г.).

Участие авторов

Все авторы внесли значимый вклад в проведение исследования и подготовку статьи, прочли и одобрили финальную версию статьи перед публикацией, выразили согласие нести ответственность за все аспекты работы, подразумевающую надлежащее изучение и решение вопросов, связанных с точностью или добросовестностью любой части работы. Zubov S. A. — идея работы, разработка методологии и обработка данных, создание черновика рукописи; Tikhonova O. A. — поиск литературы, концепция исследования, общее руководство, финальное редактирование.

Финансирование

Исследование выполнено при финансовой поддержке ФМБА России, гос. задание № 122032200139–5 «Здоровьесбережение».

СВЕДЕНИЯ ОБ АВТОРАХ

Зубов Сергей Александрович — к. э. н., доцент, старший научный сотрудник лаборатории Мультидисциплинарных клинических исследований отдела Медицинской физики и лучевых технологий Управления радиационной медицины, ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр имени А. И. Бурназяна», Москва, Российская Федерация

e-mail: tehwas@yandex.ru

ORCID ID: 0009-0006-2918-0369

РИНЦ SPIN-код: 8800-9504

Тихонова Ольга Александровна — к. м. н., зав. лабораторией Мультидисциплинарных клинических исследований клинического отдела Медицинской физики и лучевых технологий Управления радиационной медицины, ФГБУ «Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр имени А. И. Бурназяна», Москва, Российская Федерация

Автор, ответственный за переписку

e-mail: ttx_2001@mail.ru

ORCID ID: 0000-0002-2255-4667

РИНЦ SPIN-код: 5904-4013

ADDITIONAL INFORMATION

Conflict of interests

The authors declare that there is no conflict of interest in connection with this publication.

Ethical review

In carrying out the work, the authors adhered to the ethical principles of medical research described in the World Medical Association Declaration of Helsinki "Ethical Principles for Medical Research Involving Human Subjects", latest edition (October 2024).

Authors' participation

All authors made a significant contribution to the research and preparation of the manuscript, read and approved the final version of the article before publication, and agreed to be responsible for all aspects of the work, including proper study and resolution of issues related to the accuracy or integrity of any part of the work. Zubov S. A. — study idea, methodology development and data processing, manuscript drafting; Tikhonova O. A. — literature search, study concept, general supervision, final editing.

Financing

The study was financially supported by the Federal Medical and Biological Agency of Russia, state task No. 122032200139–5 "Health saving".

ABOUT THE AUTHORS

Sergey A. Zubov — Cand. Sci. (Econ.), Associate Professor, Senior Researcher, Laboratory of Multidisciplinary Clinical Research, Department of Medical Physics and Radiation Technologies, Department of Radiation Medicine, Research Center — Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russian Federation

e-mail: tehwas@yandex.ru

ORCID ID: 0009-0006-2918-0369

RSCI SPIN-code: 8800-9504

Olga A. Tikhonova — Cand. Sci. (Med.), Head of the Laboratory of Multidisciplinary Clinical Research, Clinical Department of Medical Physics and Radiation Technologies, Department of Radiation Medicine, Research Center — Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russian Federation

Corresponding author

e-mail: ttx_2001@mail.ru

ORCID ID: 0000-0002-2255-4667

RSCI SPIN-code: 5904-4013

Литература / References

1. Галиуллин А.Н., Гайфуллина Р.Ф., Галиуллин Д.А., и др. Моделирование и прогнозирование хронических неинфекционных заболеваний по медико-социальным факторам риска у лиц старше трудоспособного возраста. *Профилактическая медицина*. 2025;28(4):67–74. Doi: 10.17116/profmed20252804167. [Galiullin AN, Gaifullina RF, Galiullin DA, et al. Modeling and predicting chronic non-communicable diseases based on medical and social risk factors in older then working age people. *Russian Journal of Preventive Medicine*. 2025;28(4):67-74. (In Russ.)].
2. Zhao T, Lin G, Chen W, et al. Predicting symptomatic carotid artery plaques with radiomics-based carotid perivascular adipose tissue characteristics: a multicenter, multiclassifier study. *BMC Med Imaging*. 2025 Aug 19;25(1):337. doi: 10.1186/s12880-025-01876-x.
3. de Almeida-Neto PF, Baxter-Jones ADG, Arais RF, et al. Enhancement of a Mathematical Model for Predicting Puberty Stage in Boys: A Cross-Sectional Study. *Am J Hum Biol*. 2025 Jan;37(1):e24193. doi: 10.1002/ajhb.24193.
4. Li Z, Zhang W, Huang J, et al. Machine learning and discriminant analysis model for predicting benign and malignant pulmonary nodules. *BMC Med Inform Decis Mak*. 2025 Jul 18;25(1):272. doi: 10.1186/s12911-025-03067-8.
5. Piotrowski T, Ryczkowski A, Kalendralis P, et al. Forecasting model for qualitative prediction of the results of patient-specific quality assurance based on planning and complexity metrics and their interrelations. Pilot study. *Rep Pract Oncol Radiother*. 2024 Jul 22;29(3):318-328. doi: 10.5603/rpor.101093.
6. Manenti G, Coppeta L, Kirev IV, et al. Low-Dose Occupational Exposure to Ionizing Radiation and Cardiovascular Effects: A Narrative Review. *Healthcare (Basel)*. 2024 Jan 18;12(2):238. doi: 10.3390/healthcare12020238.
7. Hamada N. Cancer and Non-Cancer Effects Following Ionizing Irradiation. *Cancers (Basel)*. 2024 Mar 13;16(6):1141. doi: 10.3390/cancers16061141.
8. Guo L, Wu B, Wang X, et al. Long-term low-dose ionizing radiation induced chromosome-aberration-specific metabolic phenotype changes in radiation workers. *J Pharm Biomed Anal*. 2022 May 30;214:114718. doi: 10.1016/j.jpba.2022.114718.