

Дальнейшее развитие исследования может включать в себя улучшение точности детекции несовершенств кожи, расширение базы данных о косметических препаратах и анализа пользовательского опыта для более точных рекомендаций.

Таким образом, наше исследование вносит вклад в развитие методов искусственного интеллекта в области косметологии и ухода за кожей, а разработанный бот представляет собой перспективный инструмент для предоставления персонализированных рекомендаций по уходу за кожей через Telegram платформу.

### СПИСОК ИСТОЧНИКОВ

1. Гонсалес Р. Цифровая обработка изображений/ Р. Гонсалес // М.: Техносфера, 2005. 1072 с.
2. You only look once: Unified, real-time object detection / Redmon J. [et al] //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779-788
3. Redmon J., Farhadi A. YOLO9000: better, faster, stronger/ J. Redmon, A. Farhadi //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 7263-7271.
4. Redmon J., Farhadi A. YoloV3: An incremental improvement /J. Redmon, A. Farhadi //arXiv preprint arXiv:1804.02767. – 2018
5. Bochkovskiy A., Wang C. Y., Liao H. Y. M. YoloV4: Optimal speed and accuracy of object detection / A. Bochkovskiy, C. Y. Wang., H. Y. M. Liao. //arXiv preprint arXiv:2004.10934. – 2020
6. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems/ R. E. Kalman // Journal of Basic Engineering. 1960. Vol. 82, Is. 1. Pp. 35–45.

### Сведения об авторах

В.П. Иванова\* – врач-исследователь  
С.И. Богданов – доктор медицинских наук, доцент

### Information about the authors

V.P. Ivanova \*– врач исследователь  
S.I. Bogdanov – Doctor of Sciences (Medicine), Associate Professor

\*Автор, ответственный за переписку (Corresponding author):  
viktorya.schmeleva@yandex.ru

УДК: 613.6.027

## МЕТОДЫ ОПРЕДЕЛЕНИЯ СУЩЕСТВЕННОСТИ ДЕЙСТВУЮЩИХ ФАКТОРОВ НА СОСТОЯНИЕ ЗДОРОВЬЯ РАБОТНИКОВ МЕТАЛЛУРГИЧЕСКОГО ПРОИЗВОДСТВА

Кабакова Елизавета Александровна<sup>1,2</sup>, Панов Владимир Григорьевич<sup>2</sup>

<sup>1</sup>ФГАОУ ВО «УрФУ имени первого Президента России Б. Н. Ельцина»

<sup>2</sup>Институт промышленной экологии УрО РАН

Екатеринбург, Россия

### Аннотация

**Введение.** Болезни системы кровообращения занимают первое место в рейтинге общей заболеваемости в Российской Федерации, а также являются основной причиной смерти взрослого населения. Особое внимание следует уделять работникам вредных производств, таких как металлургические предприятия, поскольку условия производственной деятельности оказывают существенное влияние на показатели их здоровья, что напрямую влияет на эффективность и возможность выполнения рабочими профессиональных обязанностей. **Цель исследования** – методами математической статистики определить и ранжировать производственные и индивидуальные факторы риска для состояния системы кровообращения работников металлургического производства. Оценить возможность построения прогностической модели для предсказания состояния системы кровообращения на основе логистической регрессии. **Материал и методы.** Для определения и ранжирования факторов риска проведена оценка обезличенных показателей здоровья 1382 рабочих металлургического производственного предприятия. С помощью компьютерных программ проведен статистический анализ этих данных. **Результаты.** Определены факторы риска для состояния системы кровообращения, на основе которых можно оценить вероятность возникновения и развития патологии у сотрудников металлургического производства. Среди значимых факторов риска находятся следующие показатели: вес, уровень глюкозы в крови, холестерин, показатели артериального давления. Найденные наиболее важные переменные использованы для построения прогностической модели определения состояния системы кровообращения работников. Проведен ROC-анализ построенных моделей. **Выводы.** Выполненный статистический анализ данных работников металлургического производства показывает, что среди имеющихся показателей наиболее существенное влияние на состояние системы кровообращения оказывают индивидуальные факторы риска. Факторы производственного риска в данном случае играют заметно меньшую роль, что, возможно, объясняется слабо различающимися условиями производственного процесса испытуемых (отсутствием контрольной группы) или незначительным влиянием производственных условий именно на состояние системы кровообращения.

**Ключевые слова:** болезни системы кровообращения, факторы риска, методы математической статистики, ROC-анализ, логистическая регрессия.

## **METHODS FOR DETERMINING IMPORTANCE OF THE FACTORS ACTING ON THE HEALTH STATUS OF METALLURGICAL PLANT WORKERS**

Kabakova Elizaveta Aleksandrovna<sup>1,2</sup>, Panov Vladimir Grigorievich<sup>2</sup>

<sup>1</sup>Ural Federal University named after the First President of Russia B.N. Yeltsin

<sup>2</sup>Institute of Industrial Ecology

Ekaterinburg, Russia

### **Abstract**

**Introduction.** The circulatory system diseases occupy the first place in the rating of general morbidity in the Russian Federation and are the main cause of death in the adult population. Special attention should be paid to the workers of harmful industries, such as metallurgical plants, because the industrial activity conditions have a significant impact on their health, which directly affects the efficiency and ability of workers to perform their professional duties. **The aim of the study** is to determine and rank industrial and individual risk factors for circulatory system status of metallurgical plant workers using the methods of mathematical statistics. To evaluate the possibility of constructing a prognostic model for predicting the status of circulatory system with help of logistic regression. **Material and Methods.** To determine and rank the risk factors, the anonymized health indicators of 1382 workers of a metallurgical plant were evaluated. Statistical analysis of these data was carried out using computer programs. **Results.** Risk factors for circulatory system status have been determined, and they are used for estimation probability of occurrence and development of a pathology in workers of the metallurgical production. Among the significant risk factors are the following: weight, blood glucose level, cholesterol, blood pressure indices. The found most important variables were used to build prognostic model for determining the status of workers' circulatory system. ROC-analysis of the constructed models was carried out. **Conclusion.** The performed statistical analysis of metallurgical plant workers data shows that among the available indices individual risk factors have the most significant influence on the circulatory system status. The industrial risk factors in this case play a noticeably smaller role, which may be explained by the weakly differing conditions of the production process of the subjects (absence of the control group) or insignificant influence of the production conditions on the circulatory system status.

**Keywords:** circulatory diseases, risk factors, mathematical statistics methods, ROC-analysis, logistic regression.

### **ВВЕДЕНИЕ**

Болезни системы кровообращения (БСК), соответствующие классу IX по классификации ВОЗ МКБ-10, на протяжении последних десятилетий стабильно занимают первое место в рейтинге общей заболеваемости, а также являются преобладающей причиной смерти населения Российской Федерации [1]. Задачи, направленные на выявление факторов, оказывающих влияние на возникновение и развитие упомянутых патологий у населения в целом и у представителей различных профессий, сохраняют свою актуальность, несмотря на интенсивные исследования в этой области [2-4] вследствие увеличения смертности от данного типа заболеваний [4]. Отдельно стоит рассмотреть определение факторов, влияющих на состояние здоровья работников вредных производств, таких как металлургические предприятия, поскольку условия производственной деятельности имеют существенное влияние на показатели здоровья работников. В свою очередь, от состояния их здоровья напрямую зависят возможность и качество выполнения производственных задач. Таким образом, определение производственных факторов совместно с выявлением факторов индивидуального риска по отношению к различным заболеваниям, в частности, заболеваниям системы кровообращения, имеет также и экономическую составляющую.

В качестве примера анализа факторов риска работников промышленных предприятий была рассмотрена задача определения и ранжирования таких факторов по степени их воздействия на заболеваемость болезнями системы кровообращения для работников металлургического предприятия.

**Цель исследования** – методами математической статистики определить и ранжировать производственные и индивидуальные факторы риска для состояния системы кровообращения работников металлургического производства. Оценить возможность построения прогностической модели для предсказания состояния системы кровообращения на основе логистической регрессии.

### **МАТЕРИАЛ И МЕТОДЫ**

Анализ проведен на основе медико-биологических данных, содержащих обезличенные сведения о состоянии здоровья 1382 работников металлургического предприятия на момент прохождения диспансеризации. В базу включены индивидуальные показатели здоровья, учитывается наличие ряда заболеваний в семейном анамнезе, образ жизни работников, условия производственной деятельности, их специальность и осуществляемый род деятельности. До начала реализации статистического анализа база была очищена от текстовых меток и данных, при введении которых были допущены ошибки, а для исследования отобраны и проверены на нормальность данные сотрудников, занятых во вредных условиях труда.

В процессе работы использовалось прикладное программное обеспечение Statistica.10 и Wolfram Research *Mathematica* v.11.3.

## РЕЗУЛЬТАТЫ

На первом этапе был выполнен анализ однофакторных эффектов посредством расчета критерия Манна-Уитни, отношения шансов и  $V$  Крамера с целью выявления переменных, имеющих статистически значимое влияние на индикаторную переменную БСК. Это позволило из непрерывных и дискретных переменных отобрать те, для которых распределения значений в выборках с выявленными заболеваниями системы кровообращения и их отсутствием имеют статистически значимые различия. Полученный список включает 37 переменных, из которых предстоит выделить наиболее значимые.

Далее был оценен размер эффекта отобранных переменных для определения степени их влияния на интересующие нас патологии путем расчета величин  $dCor$  и  $d$  Коэна с последующим предварительным ранжированием переменных. Результаты вычислений позволили сократить перечень возможных факторов риска, который теперь содержит возраст, стаж во вредных условиях, вес, окружность талии, уровень глюкозы в крови, уровень холестерина в крови, систолическое артериальное давление, диастолическое артериальное давление, отношение окружности талии к росту и их закодированные варианты в виде дискретных переменных. Также было решено добавить в список класс условий труда (КУТ) по тяжести трудового процесса: несмотря на отсутствие статистической значимости при проверке распределения этого показателя в группах БСК = 0 и БСК = 1 ( $p = 0,19$  при стандартном уровне значимости  $\alpha = 0,05$ ), данный параметр может быть важен для последующих расчетов [3,5].

Для выявления возможной линейной связи между двумя непрерывными переменными и оценки её силы был выполнен корреляционный анализ, включающий в себя расчет коэффициентов Пирсона и Спирмена для различных комбинаций пар переменных в упомянутых выборках с диагностированными БСК и без них. Данная процедура позволила определить, какие переменные взаимосвязаны или дублируют друг друга. Наиболее сильной, что очевидно, оказалась связь между систолическим и диастолическим показателями артериального давления. Также было решено опустить дискретные переменные в виду большей информативности и точности расчетов при работе с их непрерывными аналогами.

Чтобы определить, как именно предложенные факторы риска влияют на возникновение БСК, и иметь возможность предсказать отклик со стороны данного параметра по известным значениям выбранных переменных, были построены модели логистической регрессии: линейные, с перекрестными произведениями и квадратичные относительно логита отклика. Качество полученных моделей оценивалось по величине  $pR^2$  Макфаддена (McFadden), значение которого в пределах 0.2–0.4 считается превосходным («excellent model fit»). Оказалось, что переход от более простых моделей к сложным или к моделям с большим числом переменных не приводит к существенным улучшениям, а потому для дальнейших работ достаточно линейной модели, включающей три переменных. Для каждой из рассматриваемых линейных моделей найдены вероятности отсека из условия равенства чувствительности и специфичности предполагаемого диагностического теста. Также рассчитаны обычные характеристики бинарного классификатора: вероятности истинно-положительных, истинно-отрицательных, ложноположительных, ложноотрицательных результатов, индекс Юдена, площадь под ROC-кривой.

## ОБСУЖДЕНИЕ

В результате выполненного анализа была определена оптимальная модель логистической регрессии (Таблица 1, Рисунок 1), включающая в себя такие переменные, как возраст, окружность талии и систолическое артериальное давление. Данные переменные являются индивидуальными факторами риска для работников металлургического предприятия по отношению к заболеваниям сердечно-сосудистой системы. Располагая известными значениями данных переменных, можно предсказать статус БСК с вероятностью в 0,8.

Таблица 1.

Параметры оптимальной линейной модели логистической регрессии

Параметр	Значение
Истинно-положительные результаты (TP)	382
Ложноположительные результаты (FP)	185
Истинно-отрицательные результаты (TN)	721
Ложноотрицательные результаты (FN)	94
Чувствительность теста (Sen или TPR)	0,803
Специфичность теста (Spe или TNR)	0,796
Вероятность ложноотрицательных результатов (FNR)	0,197
Вероятность ложноположительных результатов (FPR)	0,204
Точность теста (Acc)	0,798
Диагностическое отношение шансов (DOR)	15,838
Индекс Юдена (J)	0,598
Площадь под ROC-кривой (AUROC)	0,867
Точка отсечения	0,350

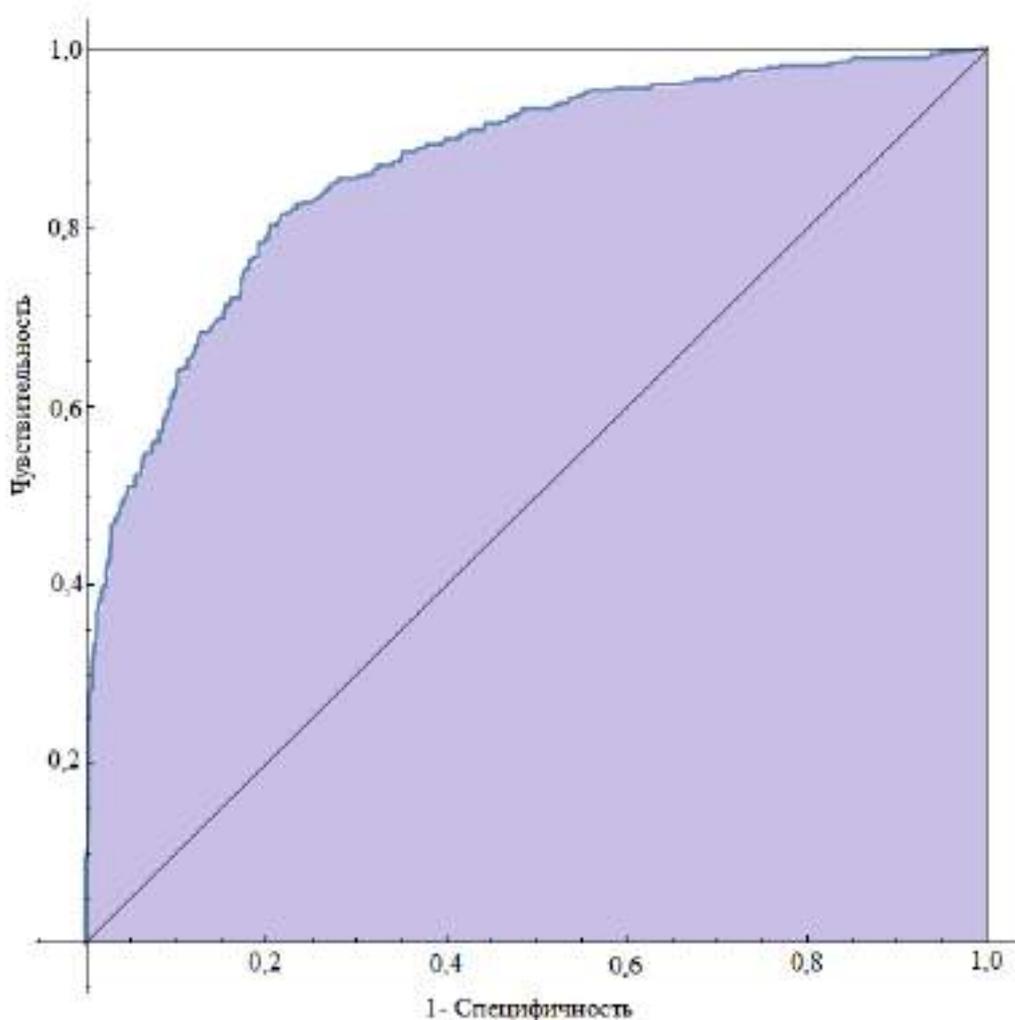


Рис. 1 ROC-кривая оптимальной модели

Полученный перечень факторов риска согласуется с результатами других работ [2,3,5-7]. Но имеются и различия.

Так по итогам нашего исследования производственные факторы, среди которых класс условий труда, играют заметно меньшую роль (Таблица 2) относительно индивидуальных, тогда как в ряде работ [3,6,7] отмечается их значимость. Это можно объяснить слабо различающимися условиями производственного процесса испытуемых (отсутствием контрольной группы) или незначительным влиянием производственных условий именно на состояние системы кровообращения. Также мы ввели дополнительную переменную в виде отношения окружности талии к росту, которой не уделено внимание в других исследованиях. Было обнаружено, что она значительно влияет на проявление БСК.

Если сравнивать с наиболее близким к нашей работе исследованием [7], можно заметить, что в нем рассматривались показатели здоровья у представителей кузнечных цехов, тогда как у нас охвачен весь рабочий персонал мужского пола, непосредственно занятый в условиях металлургического производства. В работе [6] рассматривается производство другого типа: несмотря на то, что условия труда имеют общие черты, это не гарантирует полного повторения возможных факторов риска. В работе [3] в качестве индикаторной патологии выбраны кардиореспираторные заболевания, что увеличивает количество возможных переменных за счет исследования влияния факторов на дыхательную систему человека.

Таблица 2.

Значимые переменные, ранжированные по величине отношения шансов на интервал изменения переменных

Переменная	Значение отношения шансов
Систолическое артериальное давление	8983,824
Уровень глюкозы в крови	6137,580
Диастолическое артериальное давление	3361,931
Возраст	163,257
Окружность талии/рост	98,103
Окружность талии	62,159
Стаж во вредных условиях	32,319
Вес	14,047
Уровень холестерина в крови	4,552
Класс условий труда по тяжести	1,235

Стоит отметить, что полученную в данной работе модель целесообразно использовать для прогнозирования откликов со стороны БСК у представителей рабочих специальностей, занятых на конкретном предприятии, т. к. любые отличия производственных условий способны повлиять на результаты ранжирования факторов по величине эффекта, а также ввести в итоговый перечень важнейших переменных дополнительные показатели. При этом продемонстрированный нами метод можно применить и по отношению к другим заболеваниям, использовать для определения отклика той или иной системы организма у рабочих различных цехов и т.д.

## ВЫВОДЫ

В ходе выполнения данной работы методами математической статистики был произведен анализ медико-биологических данных и определены факторы риска для состояния системы кровообращения работников металлургического предприятия: в плане значимости индивидуальные факторы преобладают над производственными. На основе полученных данных был отобран список переменных, которые были использованы для построения моделей логистической регрессии для предсказания состояния системы кровообращения. Оптимальную модель, являющуюся линейной и включающую в себя возраст, окружность талии и систолическое артериальное давление, можно использовать для прогнозирования состояния здоровья представителей металлургического предприятия, а именно оценить отклик со стороны болезней системы кровообращения.

## СПИСОК ИСТОЧНИКОВ

1. Открытые данные Федеральной службы государственной статистики – официальный сайт. URL: <https://rosstat.gov.ru/folder/12781> (дата обращения: 15.02.2023). Текст: электронный.
2. Бабанов С. Профессиональные поражения сердечно-сосудистой системы / С. Бабанов, Р. Бараева // Врач. – 2015. – №. 3. – С. 7-10.
3. Власова Е. М. Особенности формирования кардиореспираторной патологии у работников титаномагневых производств / Е.М. Власова, А.А. Воробьева, Т.А. Пономарева // Медицина труда и промышленная экология. – 2017. – №. 9. – С. 38-38.
4. Улумбекова Г. Э. Факторы влияния на смертность от новообразований и болезней системы кровообращения в РФ с 2019 по 2020 г / Г.Э. Улумбекова, А.Б. Гинойн, И.В. Петрачков // ОРГЗДРАВ: Новости. Мнения. Обучение. Вестник ВШОУЗ. – 2021. – №. 3 (25). – С. 4-23.
5. Шляпников Д. М. Риск развития связанных с работой заболеваний системы кровообращения у работников титаномагневых производств / Д.М. Шляпников, Е.М. Власова // Санитарный врач. – 2015. – №. 10. – С. 25-36.
6. Производственная среда автомобилестроения как один из факторов риска развития болезней системы кровообращения у работников / Э.Т. Валеева, Р.Р. Галимова, А.А. Дистанова [и др.] // Анализ риска здоровью. – 2023. – №. 2. – С. 95-103.
7. Ведущие факторы риска формирования патологии системы кровообращения и костно-мышечной системы у работников металлургического предприятия / Е.Л. Базарова, А.Н. Вараксин, Т.А. Маслакова [и др.] // Здоровье населения и среда обитания–ЗНисО. – 2023. – Т. 31. – №. 11. – С. 50-57.

### Сведения об авторах

Е.А. Кабакова\* – студент магистратуры, инженер

В.Г. Панов – кандидат физ.-мат. наук, ведущий научный сотрудник

### Information about the authors

E.A. Kabakova\* – M.S. student, engineer

V.G. Panov – Candidate of Sciences (Physics and Mathematics), leading researcher

\*Автор, ответственный за переписку (Corresponding author):

Liza.Kabakowa@yandex.ru

УДК: 51-7.57.087

## ТЕОРИЯ И МЕТОДОЛОГИЯ ПРЕПОДАВАНИЯ БИОСТАТИСТИКИ ДЛЯ СТУДЕНТОВ НАО «МЕДИЦИНСКИЙ УНИВЕРСИТЕТ АСТАНА»

Калиева Жаныл Ахметовна, Султанова Жанылсын Джиенгельдиевна

Кафедра биostatистики, биоинформатики и информационных технологии

НАО «Медицинский университет Астана»,

Астана, Казахстан.

### Аннотация

**Введение.** Статистическая обработка результатов медико-биологических экспериментов и данных повседневной медицинской практики сложна, многокомпонентна. В связи с этим на сегодня актуальным является вопрос о том, что закономерность и постановка новых научных гипотез при изучении состояния больного или заболевания, в качестве диагностических целей необходим комплекс математического аппарата с использованием статистических программ, основанных на методах и этапах статистического анализа. **Цель исследования** – сформировать у студентов теоретические знания и практические навыки решения аналитических и клинических задач с помощью статистических методов. **Материал и методы.** В качестве материала были использованы мультимедийные презентации, логическая структура занятия, ситуационные задачи по биostatистике, технические средства обучения: ПК, мультимедийный проектор, табличный проектор Microsoft Excel. **Результаты.** Изучение каждой темы сопровождается примерами решения конкретных медицинских проблем, например, определение референтных величин, установление взаимосвязи между физиологическими процессами в норме и при патологии, сравнительная оценка заболеваемости в различных популяциях, прогнозирование выживаемости при различных методах лечения. **Выводы.** Главным требованием к учебному процессу на кафедре биostatистики, биоинформатики и информационных технологии является подготовка врачей-специалистов, на способности анализировать литературные данные для постоянного самосовершенствования, что требует от врача быть в курсе всего нового в непрерывном потоке информации.

**Ключевые слова:** статистика, наглядная медицинская статистика, статистическая обработка, биоинформационные технологии, программы для обработки статистической информации SPSS Statistics, MS Excel.

## THEORY AND METHODOLOGY OF TEACHING BIostatISTICS FOR STUDENTS OF NJSC «ASTANA MEDICAL UNIVERSITY»

Kalievа Zhanyl Akhmetovna<sup>1</sup>, Sultanova Zhanylsyn.Dzhiengeldievna<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Bioinformatics and Information Technology

Astana Medical University

Astana, Kazakhstan.