

## МАТЕМАТИЧЕСКИЕ МЕТОДЫ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ ИНФОРМАЦИОННЫХ СИСТЕМ

© 2016 С. И. Рыжков, С. А. Харченко, К. С. Ласточкина

*Российский новый университет  
Воронежский институт высоких технологий  
Первый Московский государственный медицинский университет им. И. М. Сеченова*

*В данной работе рассматриваются математические методы, которые применяются в медико-биологических информационных системах. Отмечено, что весьма важным является проверка однородности выбранных групп наблюдения, в том числе контрольных.*

*Ключевые слова: информационная система, медицина, биология, математический метод.*

Введение. В настоящее время большое внимание уделяется разработке устройств для проведения учебно-исследовательских работ в области образования и науки, особенно таких наукоемких отраслях, как экспериментальная медицина, биология, фармакология, физиология и т. п.

Существовавшая ранее практика использования для этих целей типовых медицинских приборов (кардиографы, миографы и т. д.) ныне невозможна по экономическим соображениям, да и технические характеристики этих приборов, ориентированные на организм человека, не всегда удовлетворяют требованиям работы с экспериментальными биологическими объектами, поскольку параметры жизнедеятельности человека отличаются от аналогичных параметров животных.

Исходя из практики проведения анализа различных результатов, в качестве необходимого условия можно считать оценку того, насколько они достоверны и подтверждаются подобными данными или при помощи контрольных групп, или при решении задач, связанных с классификацией для того, чтобы проводились процессы, связанные с диагностикой, прогнозированием для разной лечебной практики. В последние годы для этого широко используются отдельные программы или пакеты прикладных программ, реализованные на персональных компьютерах.

Но использование соответствующих методов, связанных с обработкой данных, можно рассматривать как заключительный

этап при анализе данных. Весьма важно использовать четкую постановку задачи и готовить материал для математической обработки.

Постановка задачи. В зависимости от поставленной задачи, объема и характера материала, вида данных и их связей определяется выбор методов математической обработки на этапах как предварительного (для оценки характера распределения в исследуемой выборке), так и окончательного анализа в соответствии с целями исследования. Крайне важным является проверка однородности выбранных групп наблюдения, в том числе контрольных, что может быть проведено или экспертным путем, или методами многомерной статистики (например, с помощью кластерного анализа). Но первый этап связан с тем, что составляется вопросник (или тезаурус), в нем предусмотрено проведение стандартизованного описания признаков, где требуется единый подход в том, как понимаются одни и те же симптомы различными врачами, в том числе и при учете диапазонов их изменений (говорят о степенях выраженности). Когда есть существенные различия в процессах, связанных с регистрацией исходных данных (при субъективном характере оценки со стороны разных специалистов) и нет возможностей привести их к единому виду для сбора информации, может быть затем осуществлена так называемая коррекция ковариант, которая предполагает нормализацию переменных, т. е. устранение ненормальностей показателей в матрице данных. «Согласование мнений» осуществляется при учете специальностей и опыта врачей, что дает возможности проводить сравнение полученных ими результатов обследований между собой. Для

---

Рыжков Сергей Иванович – РосНОУ, студент, e-mail: vut6r5e20ghtvv@yandex.ru.

Харченко Сергей Александрович – ВИБТ АНОО ВО, аспирант, e-mail: harrtsssaaa1990w@yandex.ru.

Ласточкина Ксения Сергеевна – Первый МГМУ, студентка 4 курса лечебного факультета.

этого могут использоваться многомерный дисперсионный и регрессионный анализы.

Во многих случаях делают попытки анализа для большого количества признаков, это, как считают исследователи, может вести к тому, что повышается информативность представленных выборок. Но проведение выбора полезной информации, когда отбираются признаки, можно рассматривать как необходимую операцию, так как при решении любых классификационных задач требуется отбирать полезную, с точки зрения данных задач, информацию. В тех случаях, когда в силу обстоятельств исследователи это не сделали самостоятельно или отсутствуют достаточно обоснованные критерии для снижения размерности пространства признаков по содержательным соображениям, борьба с избыточностью информации осуществляется уже формальными методами на основе того, что проводится оценка информативности. После того, как определен перечень признаков, для того, чтобы проводить математическую обработку, переходят к этапу формирования формализованных документов, в которых для каждого значения признака или симптома ставят в соответствие некоторое кодовое слово. Кодирование не ведет к изменению семантической силы и смысла в исходных данных.

Представление информации в квантованной форме обеспечивает фиксацию степени выраженности или состояний качественных признаков и диагностически значимых интервалов количественных признаков.

Если говорят о двух и более классах заболеваний, то проводить операции численного кодирования необходимо таким образом, чтобы сохранилось максимальное различие среди изображений разных классов.

Это связано с тем, что признаки несут некоторый содержательный смысл, что очень важно, так как при кодировании происходит замена реальных объектов числами (векторами).

В качестве примера можно привести выборку, составленную из реальных объектов, в которой одному классу соответствуют объекты со значением признаков, не превосходящими 4, а другому – те, которые имеют значение такого параметра, лежащего между 5 и 9. Для произвольного кодирования есть возможность сделать его для объектов так, что такая важная с точки зрения распознавания объектов характеристика будет утеряна.

Математические методы. Для множественного регрессионного анализа рассматри-

вают связь одной переменной, которая называется зависимой, и определенным числом других, которые называют независимыми. Такую связь представляют на основе математической модели, это уравнение, связывающее зависимую переменную и независимые. Для функции может быть свойство линейности – тогда говорят о линейной модели регрессии, в других случаях – о нелинейной. Таким путем могут быть установлены зависимости заболеваемости от характера и уровня химического загрязнения окружающей среды, позволяющие оценивать вклад различных веществ и их сочетаний, в том числе выбросов конкретных предприятий в реализацию заболеваний, что и было продемонстрировано. Однако нужно иметь в виду, что если независимые переменные (аргументы) сильно коррелированы и, соответственно, полученные коэффициенты также сильно коррелированы, то нельзя менять один отдельный фактор, используя соответствующий ему коэффициент.

Дисперсионный анализ позволяет определить влияние разных факторов (условий) на исследуемый признак (явление), что достигается путем разложения совокупной изменчивости (дисперсии, выраженной в сумме квадратов отклонений от общего среднего) на отдельные компоненты, вызванные влиянием различных источников изменчивости. Методы анализа данных нашли широкое распространение, в том числе в исследованиях по изучению угрозы заболевания при наличии факторов риска. Используются и специально разработанные математические модели. В концепции относительного риска рассматривают отношения среди пациентов, имеющих определенную болезнь и пациентами, которые ее не имеют. Величина относительного риска дает возможность определить, во сколько раз увеличивается вероятность заболеть при его наличии. Выявлять факторы, способствующие возникновению (проявлению) заболевания, мы можем на основе разных способов, например, на основе того, что проводится оценка информативности, а затем ранжируются признаки. Это не дает указание на общее действие тех параметров, которые отобрали, есть отличие от применения регрессионного и факторного анализов, способов, связанных с теорией распознавания образов, дающих возможности для того, чтобы получились «симптомо-комплексы» риск-факторов. Кроме того, более сложные методы позволяют анализировать и непрямые связи между факторами

риска и заболеваниями. Дискриминантный анализ является обобщенным названием методов, связанных с классификацией объектов, они представлены на основе многомерных наблюдений. Его применяют при медицинских исследованиях для целей дифференциальной диагностики, проведения процессов прогнозирования осложнений или протекания заболеваний. Для тех случаев, когда его используют для существующих у исследователей выборок, более часто применяют методы, связанные с распознаванием образов на базе вероятностных процедур или алгоритмов детерминистской логики. В качестве основы в первых подходах рассматривают процедуру Байеса, позволяющую сделать выбор одной из нескольких гипотез по вычислению вероятностей заболеваний на базе вероятностей обнаруженных признаков у больных людей.

Таким путем достигается различение проверяемых гипотез с заданной точностью при значительно меньшем числе наблюдений. Для выявления оптимального сочетания признаков, имеющих диагностическое (прогностическое) значение могут применяться различные методы теории распознавания образов, основанные на вычислении расстояний между признаками, характеризующими больных сравниваемых классов.

Последнее, на что представляется целесообразным обратить внимание, является визуальная форма представления результатов. Это могут быть таблицы, графики, гистограммы, геометрическая интерпретация классификации многомерных наблюдений. Имеются методы, а priori предполагающие такой вид выходных форм, например, главные компоненты, кластер-анализ. В других случаях можно ставить такую задачу и реализовывать ее с помощью специально написанного программного обеспечения. В ряде случаев можно использовать стандартные пакеты программ, позволяющие обеспечить графическое представление данных.

Вывод. Таким образом, рассмотренные методы в информационных системах дают возможности проведения медико-биологических исследований, позволяющих использовать статистические методы в медицине, ошибки в их применении и способы, позволяющие избежать эти ошибки.

#### ЛИТЕРАТУРА

1. Болгов С. В. Прогнозирование стоматологической заболеваемости по медико-биологическим и социально-гигиеническим

факторам риска / С. В. Болгов, К. А. Разинкин, О. Н. Чопоров // Врач-аспирант. – 2011. – Т. 49. – № 6.2. – С. 294-301.

2. Махер Х. А. Разработка и использование моделей для прогнозирования качества жизни беременных по их медико-социальным характеристикам / Х. А. Махер, Н. В. Наумов, Г. Я. Клименко, О. Н. Чопоров // Системный анализ и управление в биомедицинских системах. – 2011. – Т. 10. – № 4. – С. 789-793.

3. Бугакова Е. Н. Анализ медико-социальных факторов риска развития аллергических дерматитов / Е. Н. Бугакова, Г. Я. Клименко, О. Н. Чопоров // Системный анализ и управление в биомедицинских системах. – 2009. – Т. 8. – № 3. – С. 795-798.

4. Чопоров О. Н. Оптимизация управления функционированием медицинских систем различного уровня / О. Н. Чопоров, И. Я. Львович, К. А. Разинкин, А. А. Рындин // Системы управления и информационные технологии. – 2013. – Т. 53. – № 3. – С. 100-104.

5. Чопоров О. Н. Рационализация управления региональными системами на основе использования методов системного анализа, информационных и ГИСТехнологий / О. Н. Чопоров, Н. А. Гладских, С. С. Пронин, М. И. Чудинов, С. Н. Семенов, К. Л. Матюшевский // Прикладные информационные аспекты медицины. – 2007. – Т. 10. – № 2. – С. 15-19.

6. Гладских Н. А. Применение статистических методов прогнозирования и ГИСТехнологий для мониторинга системы регионального здравоохранения / Н. А. Гладских, В. А. Голуб, С. Н. Семенов, О. Н. Чопоров // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2008. – № 1. – С. 111-116.

7. Клименко Г. Я. Исследование качества жизни беременных женщин как новый интегральный показатель оценки состояния их здоровья / Г. Я. Клименко, В. И. Стародубов, С. В. Говоров, Н. Б. Костюкова, О. Н. Чопоров // Успехи современного естествознания. – 2010. – № 9. – С. 131-132.

8. Коротких Н. Г. Изучение диагностической ценности характеристик стоматологических заболеваний хирургического профиля / Н. Г. Коротких, О. Ю. Шалаев, О. Н. Чопоров, Л. В. Бут // Российский стоматологический журнал. – 2008. – № 2. – С. 22-25.

9. Косолапов В. П. Медико-социальные особенности образа жизни и здоровья детей

школьного возраста / В. П. Косолапов, И. Э. Есауленко, П. Е. Чесноков // Вестник Волгоградского государственного медицинского университета. – 2010. – № 4. – С. 45-48.

10. Косолапов В. П. Влияние социально-экономических факторов и образа жизни на здоровье населения в воронежской области / В. П. Косолапов, Л. И. Летникова, Г. В. Сыч, М. В. Фролов, А. В. Сыч // Системный анализ и управление в биомедицинских системах. – 2015. – Т. 14. – № 4. – С. 820-828.

11. Косолапов В. П. Особенности репродуктивного здоровья населения воронежской области на фоне ЦЧР / В. П. Косолапов, П. Е. Чесноков, Г. Я. Клименко, О. Н. Чопоров // Системный анализ и управление в биомедицинских системах. – 2010. – Т. 9. – № 3. – С. 649-655.

12. Чесноков П. Е. Результаты исследования медико-социальных характеристик родильниц / П. Е. Чесноков, В. П. Косолапов, Г. Я. Клименко, Г. А. Шемаринов // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2007. – № 6. – С. 10.

13. Косолапов В. П. К вопросу о формировании приоритетных направлений развития системы охраны материнства и детства на региональном уровне / В. П. Косолапов, П. Е. Чесноков, Г. Я. Клименко // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2011. – № 2. – С. 28-32.

14. Клименко Г. Я. Методика и результаты преобразования лингвистических характеристик в численные оценки факторов

риска / Г. Я. Клименко, В. П. Косолапов, О. Н. Чопоров // Сибирский Консилиум. 2001. – № 4. – С. 25.

15. Клименко Г. Я. Использование балльной оценки для формирования интегрального показателя состояния здоровья населения / Г. Я. Клименко, И. Э. Есауленко, О. Н. Чопоров, В. П. Косолапов, Г. А. Шемаринов // Бюллетень Национального научно-исследовательского института общественного здоровья им. Н. А. Семашко. – 2003. – № 9. – С. 18-22.

16. Клименко Г. Я. Методика и результаты преобразования лингвистических характеристик в численные оценки факторов риска / Г. Я. Клименко, В. П. Косолапов, О. Н. Чопоров // Консилиум. – 2001. – № 4. – С. 25.

17. Чопоров О. Н. Интегральное оценивание и прогностическое моделирование состояния здоровья беременных, рожениц и родильниц с учетом их медико-социальных характеристик / О. Н. Чопоров, В. П. Косолапов, Н. В. Наумов, Х. А. Гацайниева // Вестник Воронежского института высоких технологий. – 2012. – № 9. – С. 91-95.

18. Калаев В. Н. Регрессионный анализ в биологических исследованиях / В. Н. Калаев, Е. А. Калаева, А. П. Преображенский, О. В. Хорсева // Системный анализ и управление в биомедицинских системах. – 2007. – Т. 6. – № 3. – С. 755-759.

19. Калаев В. Н. Применение кластерного анализа в биологических исследованиях / В. Н. Калаев, Е. А. Калаева, В. Г. Артюхов, А. П. Преображенский // Системный анализ и управление в биомедицинских системах. – 2007. – Т. 6. – № 4. – С. 1008-1014.

## THE MATHEMATICAL METHODS IN BIOMEDICAL RESEARCH INFORMATION SYSTEMS

© 2016 S. I. Ryzhkov, S. A. Harchenko, K. S. Lastochkina

*Russian New University  
Voronezh institute of high technologies  
First Moscow State Medical University named after I. M. Sechenov*

*This paper discusses the mathematical methods used in biomedical information systems. Noted that is very important is to check the homogeneity of the selected study groups, including control.*

*Keywords: information system, medicine, biology, mathematical method.*