

В сравнении с планом на основе МРТ-изображений постановка дополнительных интерстициальных игл позволила существенно увеличить объем охвата опухоли 100 % изодозой (V_{100}), при этом не превысив толерантные дозы на критические органы. Суммарная доза на объем мишени за весь курс лучевой терапии составила 86,7 Гр, на мочевого пузыря – 84,2 Гр, на прямую и сигмовидную кишку – 71 Гр.

Проведенный анализ плана облучения данного пациента позволил сделать следующие предварительные выводы о возможности и целесообразности применения ультразвука в режиме реального времени при проведении внутритканевой брахитерапии рака шейки матки. Применение внутритканевых игл под контролем УЗИ в режиме реального времени позволяет проводить внедрение игл более безопасно, а также регулировать глубину и расположение игл для увеличения охвата мишени и недопущения перфорации здоровых органов и тканей, что, в свою очередь позволяет достичь оптимального дозового распределения и облучать мишень большими значениями поглощенной дозы. После проведения дополнительного анализа большего значения клинических случаев пациентов, облученных с использованием данной методики, представляется возможным выявление наиболее значимых достоинств и недостатков и выработка рекомендаций для реализации такого облучения в других онкологических учреждениях.

ЛИТЕРАТУРА

1. *Parkin, D. M.* Global cancer statistics / Bray F, Ferlay J, Pisani P., 2002. // *CA Cancer J Clin.* – 2005. No. 55 (2). P. 74–108.
2. *Kamangar, F.* Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world / G. M. Dores, W. F. Anderson // *J Clin. Oncol.* – 2006. – No. 24 (14). – P. 2137–2150.
3. *Lanciano, R. M.* Tumor and treatment factors improving outcome in stage III-B cervix cancer / K. Martz, L. R. Coia, G. E. Hanks // *Int. J. Radiat. Oncol. Biol. Phys.* – 1991. – No. 20. – P. 95–100.
4. *Van Dyk, S.* Conformal brachytherapy planning for cervical cancer using transabdominal ultrasound / K. Narayan, R. Fisher, D. Bernshaw // *Int. J. Radiat. Oncol. Biol. Phys.* – 2009. – No. 75. – P. 64–70.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПОСТРОЕНИЯ МНОЖЕСТВ РЕГРЕССИОННЫХ МОДЕЛЕЙ НА ВЫБОРКАХ ОГРАНИЧЕННОГО ОБЪЕМА В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

SOFTWARE FOR BUILDING OF SETS OF REGRESSION MODELS ON SAMPLES OF LIMITED SIZE IN MEDICAL RESEARCHES

А. В. Копыцкий, В. Н. Хильманович
A. Kapytski, V. Khilmanovich

Гродненский государственный медицинский университет, г. Гродно, Республика Беларусь
fizika@grsmu.by
Grodno State Medical University, Grodno, Republic of Belarus

Обработка данных, полученных в результате научных исследований в области медицины, неразрывно связана с применением методов математической статистики (в частности, построением регрессионных моделей). Одной из часто возникающих проблем в исследованиях (особенно медицинских) является ограниченность объема выборки, для которой собирается статистика при большом числе изучаемых показателей. Традиционные методы определения статистически значимых оценок коэффициентов регрессионных моделей (такие, как пошаговое включение или исключение) в данном случае неприменимы. Нами предлагается реализация алгоритма построения множеств регрессионных моделей, получаемых перебором возможных сочетаний предикторов, для случаев, когда объем выборки сопоставим с числом предикторов.

Processing of data obtained in medicine researches is associated with application of statistical methods (in partially, with building of regression models). One of the often appearing problems in researches (especially medical) is limited sample size of sample, for which statistical data is collected, with a large number of variables. Usual methods of determination of statistically significant estimations of regression model coefficients (such as forward selection and backward elimination) are not applicable in this case. We propose an implementation of an algorithm of building of regression models sets, got with enumeration of all possible combinations of predictors, for cases when sample size is comparable with a number of predictors.

Ключевые слова: программное обеспечение, медицинские исследования, выборки малого объема, регрессионные модели.

Keywords: software, medical researches, small samples size, regression models.

Основная идея доказательной медицины, сложившейся в конце 80-х годов прошлого столетия, – применение в практике только тех методов диагностики и лечения, эффективность которых доказана на основе строгих научных принципов в результате клинических исследований. Одним из основных инструментов доказательной медицины является математическая статистика. В настоящее время статистическое описание и количественный анализ клинических явлений присутствуют, как правило, во всех зарубежных и отечественных научных публикациях.

Учебная дисциплина «Математическая статистика в медицине» занимает важное место в системе подготовки врача по всем специальностям, так как позволяет специалисту описывать группы объектов, выявлять значимые различия между группами, классифицировать объекты и явления по их числовым характеристикам, по имеющемуся экспериментальному материалу, делать выводы об изучаемых объектах и предсказывать их поведение. Она позволяет сформировать основы грамотного применения статистических методов обработки результатов экспериментов и измерений в медицинских исследованиях; научить правильно интерпретировать встречающиеся в специальной литературе термины и результаты статистических исследований.

В Гродненском государственном медицинском университете эта дисциплина, как компонент УВО, преподается с 2012 г. С 2017 г. силами кафедры медицинской и биологической физики были организованы курсы по основам математической статистики для профессорско-преподавательского состава.

Изучение основ статистики позволяет сформировать представление о ее принципах и методах; научить планировать процесс сбора и статистической обработки экспериментальных данных; научить регистрировать, группировать и представлять экспериментальные данные; овладеть методами анализа и обработки статистических данных для научных и практических целей; научить использованию компьютерных программ для группировки, анализа и обработки экспериментальных данных. Между тем, математическая статистика в медицине является комплексной дисциплиной, имеющей широкие междисциплинарные связи. Изучение этой дисциплины призвано обеспечить формирование академических, социально-личностных и профессиональных компетенций: умение применять базовые научно-теоретические знания для решения теоретических и практических задач; владение системным и сравнительным анализом, исследовательскими навыками; умение работать самостоятельно и креативно; владение междисциплинарным подходом при решении проблем; умение применять навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером; умение повышать свою квалификацию в течение всей жизни, самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности.

К социально-личностным компетенциям можно отнести такие компетенции, как способность к социальному взаимодействию; способность к межличностным коммуникациям; к критике и самокритике (критическое мышление); умение работать в команде.

К профессиональным компетенциям: осуществлять оценку состояния здоровья у пациентов; изучать распространенность психических и поведенческих расстройств, уровень суицидального поведения населения; собирать аналитическую информацию и разрабатывать на ее основе психообразовательные программы; планировать и проводить научные исследования, обобщать и представлять их результаты; уметь создавать и использовать в своей деятельности объекты интеллектуальной собственности; применять методы математического анализа и моделирования, теоретического и экспериментального исследования; использовать оптимальные виды, методы и средства обучения; осуществлять методическую работу; анализировать и оценивать собранные данные; готовить доклады, материалы к презентациям и представлять их; пользоваться глобальными информационными ресурсами.

Однако на практике применение статистических методов обработки медицинских исследований сопряжено с рядом трудностей. Одной из проблем в медицинских научных исследованиях является ограниченность объемов выборок, для которых собирается статистика, при большом числе изучаемых показателей. Зачастую встречаются ситуации, когда число показателей сопоставимо с объемом выборки. Одной из задач статистического анализа данных является задача построения регрессионных моделей, которые позволяют изучать в упрощенном виде взаимосвязи между переменными [1–5]. Однако при ограниченных объемах наблюдений такие широко используемые и хорошо известные методы построения регрессионных моделей, как пошаговое включение и исключение, неприменимы, что не позволяет определить наилучшее сочетание предикторов для данной переменной-отклика. Поэтому остается только ручной подбор множества предикторов. Таким образом, актуальным является разработка метода построения множества регрессионных моделей, из которого исследователь в дальнейшем выбирает наиболее подходящую модель. Задача подобного рода относится к так называемой науке о данных («Data Science»), согласно которой ручной подбор множеств может быть заменен автоматизированным перебором с использованием специализированного программного обеспечения.

Анализ разработки проблемы в республике и за рубежом показывает, что, несмотря на то, что данная проблема является актуальной, доступны её решения только для линейных моделей: пакет расширения «leaps» (США, Новая Зеландия) языка «R»; решение, предложенное И. М. Митасовым и А. Н. Завьяловым (Россия); решение, приводимое С. Э. Мастицким и В. К. Шитиковым (Беларусь, Россия)). Решения для обобщенных линейных моделей, имеющиеся в пакетах расширения «bestglm» (Канада) и «glmulti» (Франция), также имеют ряд ограничений при малых объемах выборок. Кроме того, использование описанных решений подразумевает наличие навыков программирования на языках «R» или «C» со стороны исследователя-медика или биолога. Популярными коммерческими статистическими пакетами «Statistica» (США), «SPSS» (США) также не имеют готовых имплементированных решений поиска

оптимальной модели при ограниченном объеме выборки. Существующие на сегодняшний день решения данной задачи эффективно работают только в случае, когда число предикторов превышает объем наблюдений. Такими решениями являются пакеты языка программирования «R»: «bestglm», «glmulti». Встречаются также решения задач выбора модели из множества, но для частных случаев: например, только для линейной модели.

Таким образом, актуальным является разработка программного обеспечения, позволяющего автоматически перебирать все возможные регрессионные модели для заданного набора предикторов, не требующего навыков программирования со стороны исследователя-медика.

Идея нашей работы состоит в создании программного обеспечения для автоматизированного перебора всех возможных регрессионных моделей, которые могут быть получены для данного набора предикторов при данной целевой переменной-отклике. Универсальные модели, описывающие широкий круг зависимостей, могут быть получены в рамках обобщенной линейной регрессии, включающей в себя линейную, логит-, пробит-регрессии и т. д. Так как перебор всех возможных комбинаций предикторов является экстенсивным («жадным») алгоритмом, то количество возможных комбинаций может быть существенно уменьшено, если исходить из того, что число предикторов в модели не должно быть слишком велико. Для этого мы используем следующие основания: во-первых, модель должна быть относительно простой (т. е. должна быть легко интерпретируемой), во-вторых, большинство оценок качества подгонки моделей, исходя из первого основания, вводит штраф за избыточное количество предикторов в модели. Кроме того, уменьшение времени перебора достигается анализом только тех моделей, где все коэффициенты регрессии статистически значимые; также возможно ограничение множества сочетаний путем введения предикторов, которые обязательно (с точки зрения исследователя) должны быть включены в модель. Уменьшение времени перебора также может быть достигнуто за счёт параллелизации расчетов и/или использования графических ускорителей («GPU»).

Полученное программное обеспечение впервые в Беларуси позволит строить множества обобщенных линейных моделей регрессии на ограниченных выборках в медицинских исследованиях. Данная возможность особенно актуальна для пилотных исследований в медицине и биологии, после которых исследователи будут лучше представлять взаимосвязи изучаемых показателей и смогут определять направления дальнейших исследований.

На сегодняшний день нами уже разработан пакет расширения языка программирования «R», позволяющий определять характеристики обобщенных линейных моделей с различными функциями связи. Эти модели строятся для всех возможных сочетаний из *m* предикторов из множества из *n* предикторов. В табл. приведен пример результата работы данного пакета для обобщенной линейной модели с бинарным откликом и пробит-функцией связи, где бинарная переменная «v36» рассматривается как функция линейной комбинации заданного числа предикторов. В данной таблице используются следующие обозначения: «Short.formula» – стандартный способ записи регрессионной модели; «N» – объем выборки (после удаления записей с пропущенными данными), «VIF» – индикатор высоких значений фактора инфляции дисперсии, хотя бы одного из предикторов (наличие отметки в ячейке сигнализирует о возможном наличии проблемы мультиколлинеарности), «AIC» – значение информационного критерия Акаике, «R²» – значение псевдо-R², «Deviance» – значение остаточного девианса модели, «P.cutoff» – значение порога отсека по вероятности, при котором определялись чувствительность «Se» и специфичность «Sp». По умолчанию, данные в итоговой таблице отсортированы по возрастанию величины критерия Акаике. Для построения и анализа моделей нами использованы следующие пакеты расширения языка программирования «R»: «openxlsx» – для чтения и записи данных из таблиц и в таблицы формата «*.xlsx», «ROCR» – для проведения ROC-анализа получаемых моделей, стандартный пакет «stats», пакет «car» – для получения значений факторов инфляции дисперсии.

Таблица – Пример результата работы программы

Short.formula	N	VIF	AIC	R ²	Deviance	P.cutoff	Se	Sp
v36 ~ v1 + v2 + v3 + v46	158		266.18	0.345	252.18	0.5052	0.75	0.7537
v36 ~ v1 + v3 + v46	158		267.11	0.343	253.11	0.4726	0.75	0.7537
v36 ~ v1 + v3	158		267.46	0.342	253.46	0.4994	0.7917	0.791
v36 ~ v1 + v46	158		267.25	0.342	253.25	0.4757	0.75	0.7537

На основании уже имеющихся разработок перед нами поставлена цель разработки программного обеспечения для построения множеств регрессионных моделей для выборок ограниченного объема. Для достижения поставленной цели предлагается решить следующие задачи:

- проектирование и разработка программного обеспечения для построения множеств обобщенных линейных моделей в медицинских исследованиях на выборках ограниченного объема;
- проектирование и разработка программного обеспечения для построения множеств мультиномиальных логистических моделей в медицинских исследованиях на выборках ограниченного объема;
- проектирование и разработка программного обеспечения для построения множеств моделей выживаемости в медицинских исследованиях на выборках ограниченного объема;
- проектирование и разработка методов ускорения перебора моделей с помощью параллелизации вычислений и/или с помощью графического ускорителя;

• проектирование и разработка графического интерфейса пользователя программы, методов инкапсуляции программы в динамическую библиотеку или в пакет расширения некоторых языков программирования.

Мы выделили следующие этапы разработки:

• проектирование, написание программного кода, отладка и тестирование программного обеспечения для построения множеств обобщенных линейных моделей на выборках ограниченного объема в медицинских исследованиях;

• проектирование, написание программного кода, отладка и тестирование программного обеспечения для построения множеств мультиномиальных логистических моделей на выборках ограниченного объема в медицинских исследованиях;

• проектирование, написание программного кода, отладка и тестирование программного обеспечения для построения множеств моделей выживаемости в медицинских исследованиях на выборках ограниченного объема;

• проектирование, написание программного кода, отладка и тестирование программного обеспечения для построения множеств моделей множественной регрессии с использованием метода параллелизации вычислений, а также с использованием графического ускорителя;

• проектирование и создание графического интерфейса пользователя, реализация динамической библиотеки и пакета расширения на основании уже разработанного программного обеспечения для инкапсуляции кода.

Полученные результаты позволят впервые в практике медицинских научных исследований в Беларуси получить компьютерную программу (или модуль, который может служить расширением другой программы), позволяющую методом перебора проанализировать все возможные модели множественной регрессии в случае, когда количество предикторов больше объема выборки.

Разработанную программу предполагается свободно распространять как самостоятельный продукт, как динамическую библиотеку и как пакет расширения. Программа будет полезна исследователям в области медицинской статистики при анализе данных пилотных исследований.

ЛИТЕРАТУРА

1. Андреева, И. В. Разработка методики прогнозирования с использованием корреляционно-регрессионного анализа / И. В. Андреева, М. Ю. Червякова // Экономический анализ: теория и практика. – 2013. – № 3 (19). – С. 58–61.
2. Черкашина, Ю. А. Применение регрессионного анализа в задаче диагностирования состояния здоровья детей / Ю. А. Черкашина // Современные проблемы науки и образования. – 2015. – № 1 (1). – С. 123–127.
3. Чернов, Н. А. Выявление и оценка зависимости показателей социально-экономического развития территорий с помощью методов корреляционно-регрессионного анализа // Новая наука: проблемы и перспективы. – 2016. – № 121 (1). – С. 254–257.
4. Bollerslev, T. Generalized autoregressive conditional heteroscedasticity // Journal of Econometrics. – 1986. – Vol. 31. – P. 307–327.
5. Сиднеева, Л. В. Применение методов регрессионного анализа в биометрических исследованиях / Л. В. Сиднеева, Н. Н. Медведева, В. Г. Николаев, Н. Н. Стрелкович, И. И. Орлова // Вестник новых медицинских технологий, Тула. 2013. – № 2. – С. 216–219.

ЭФФЕКТ ШТАРКА И НАНОЧАСТИЦЫ НА ОСНОВЕ CdSe/ZnS ДЛЯ СУБКЛЕТОЧНОГО ЗОНДИРОВАНИЯ ЛОКАЛЬНОГО pH

STARK EFFECT AND CdSe / ZnS NANOPARTICLES FOR SUBCELLULAR SENSING OF LOCAL pH

И. Г. Мотевич¹, Н. Д. Стрекаль², С. А. Маскевич²

I. Motevich¹, N. Strekal², S. Maskevich²

¹Гродненский государственный университет им. Янки Купалы,
г. Гродно, Республика Беларусь,
i.motevich@grsu.by

²Белорусский государственный университет, МГЭИ им. А. Д. Сахарова БГУ,
г. Минск, Республика Беларусь

¹Grodno State University, Grodno, Republic of Belarus

²Belarusian State University, ISEI BSU, Minsk, Republic of Belarus

Предложен оригинальный метод анализа параметров флуоресценции наночастиц CdSe/ZnS в модельных водных системах и тканях толстого кишечника на основании эффекта Штарка, создаваемого локальными электрическими полями ближайшего (двойной слой Гельмгольца) микроокружения. Обсуждается механизм отклика наночастиц в водных средах с выраженными щелочными или кислотными свойствами,