

ПРОГРАММНЫЙ ИНСТРУМЕНТАРИЙ РАСЩЕПЛЕНИЯ СМЕСИ РАСПРЕДЕЛЕНИЙ

И.В. Максимей, Н.Б. Осипенко, М.Н. Васенда

УО «Гомельский государственный университет им. Ф.Скорины»,
кафедра математических проблем управления
ул. Советская 106, г. Гомель, Беларусь
телефон: (0232) 60-42-37; e-mail: mpu@gsu.by

Описываются особенности, возможности и результаты практической апробации одномерного анализа данных на основе разработанного инструментария.

Ключевые слова – одномерный анализ данных, расщепление смеси распределений.

1 МЕТОДИКА ОДНОМЕРНОГО АНАЛИЗА ДАННЫХ

В работе описывается разработанное программное обеспечение Strand, предназначенное для автоматизации исследования активных систем (АС) [1] и реализующее концепцию конструктивного использования методов анализа данных и проверке их работоспособности при автоматизации выделения экспертом причинно-следственных связей в модели АС. В зависимости от размера выборки в Strand реализуется “ядерная” аппроксимация функции распределения параметра, проверка гипотез, процесс квантильного анализа. Имеется возможность для каждой выборки задать веса, которые будут учитываться при анализе (построении нормальной вероятностной бумаги на каждом этапе), что, например, позволяет корректно обрабатывать данные, полученные из различных источников (более надежные источники имеют больший вес). Часто встречающиеся на практике пропуски в данных могут в процессе анализа игнорироваться или учитываться как содержащие значения среднего арифметического по данному признаку.

Имеется возможность проводить автоматическое нормирование данных для дальнейшего анализа на основе полученной после работы программы информации о выборке, если теоретические параметры подобрать не удалось, но исследователь желает использовать признак в дальнейшем, нормирование проводится с использованием эмпирических характеристик.

Благодаря использованию на всем протяжении работы программы нормальной вероятностной бумаги имеется возможность работать со сжатыми выборками.

2 ОПИСАНИЕ РАЗРАБОТАННОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Программа Strand позволяет проводить анализ данных, а так же реализовать гибкие технологии анализа целевого

функционирования, позволяющие достигать оптимально возможной глубины охвата данных, знаний и находить наиболее точные прогнозные оценки, адекватные модели, классификации. Описание программы оформлено согласно ГОСТ 19.402-78 ЕСПД.

1 Общие сведения: для функционирования программы “Strand” (название происходит от сокращения “СТРАНД” – “Статистический и Регрессионный Анализ Данных”) версии 1.0 необходима операционная система Microsoft Windows XP/2000/Me/98/95/NT4.0 (с установленным Service Pack 3 или старше); программа написана на языке программирования C++, компилировалась с помощью интегрированной среды разработки Borland C++Builder.

2 Функциональное назначение: автоматизация использования анализа данных при выделении экспертом причинно-следственных связей, построения концептуальной модели формирования целевого функционирования и построения классификационных образов и регрессионных оценок целевых свойств объектов АС посредством: проведения одномерного анализа данных, классификации, построения регрессионных уравнений; визуального формирования фактор-признаковых описаний и деревьев взаимосвязей факторов; построения концептуальной модели с последующим преобразованием в визуальное представление в виде дерева взаимосвязей факторов.

3 Логическая структура: программа состоит из исполняемого файла, ini-файла конфигурации, dll-библиотек с математическими функциями и функциями работы с дополнительными форматами файлов данных; главное меню программы состоит из следующих пунктов: “Файл” - команды по работе с файлами, “Редактирование” - команды по редактированию данных, “Анализ” – команды по анализу данных, “Окно” - команды по упорядочению и быстрому переключению открытых окон. Каждый подпункт главного меню является полнофункциональным модулем; программа имеет многооконный интерфейс (MDI); в программе используются возможности работы с файлами данных в формате пакета STATISTICA.

4 Используемые технические средства: процессор не ниже Intel Pentium 90, либо аналогичный, не менее 32 МБ ОЗУ, операционная система Microsoft Windows

XP/2000/Me/98/95/NT4.0 (с установленным Service Pack 3 или старше), не менее 30 МВ свободного места на жестком диске.

5 Вызов и загрузка: вызов и загрузка производится из-под операционной системы, после проведения процесса установки приложения.

6 Входные данные: обучающие примеры и тестовые данные в файлах различных форматов; ручной ввод необходимых приложению данных пользователем; файлы описания данных с расширением "desc"; файлы с диаграммами деревьев взаимосвязей; файлы данных в формате пакета STATISTICA.

7 Выходные данные: вывод данных в виде таблиц результатов, графиков и диаграмм; журнала работы (Log), в который записываются все действия пользователя, а также все ошибки при обработке данных; файлы описания данных с расширением "desc"; файлы с диаграммами деревьев взаимосвязей; файлы данных в формате пакета STATISTICA; файлы с векторными изображениями графиков ядерных аппроксимаций, нормальной вероятностной бумагой, а также подобранной функцией плотности распределения в формате "wmf".

3 ИССЛЕДОВАНИЕ АЦЕТИЛЯТОРНОГО ФЕНОТИПА У ЖИТЕЛЕЙ ГОМЕЛЬСКОЙ ОБЛАСТИ ПРИМЕНИТЕЛЬНО К ЗАДАЧЕ ИНДИВИДУАЛИЗАЦИИ ФОРМАКОТЕРАПИИ

Посредством Strand был проведен анализ данных, полученных при обследовании 129 практически здоровых добровольцев контрольной группы из европеоидной популяции г. Гомеля и Гомельской области.

Клиническое обследование лиц контрольной группы включало: сбор и анализ жалоб, анамнеза, определение фенотипа ацетилирования. Определение фенотипа N-ацетилирования проводилось с помощью метода высокоэффективной жидкостной хроматографии с ультрафиолетовым обнаружением на аппарате «Agelent 1100» с помощью тестового препарата изониазида. Лица контрольной группы натошак принимали изониазид в дозе 10 мг/кг. Клинический материал (венозная кровь) собирали через 3 часа после приема изониазида. Определяли концентрацию изониазида и его ацетилированного метаболита. Фенотип N-ацетилирования определялся как скорость ацетилирования изониазида и рассчитывался как отношение ацетизониазида к изониазиду. В группу здоровых добровольцев входили лица в возрасте от 25 до 58 лет: 45 (35%) мужчин и 84 (65%) женщины. Было выделено четыре одномодальных компонента смеси с точкой расщепления медленных и быстрых ацетиляторов 0,28. Соотношение медленных и быстрых ацетиляторов составило 66% и 34% соответственно, что согласуется с данными большинства регионов Европы. Особенности аппроксимации исходного распределения смесью из четырех одномодальных нормальных компонент представлены на рисунке 1.

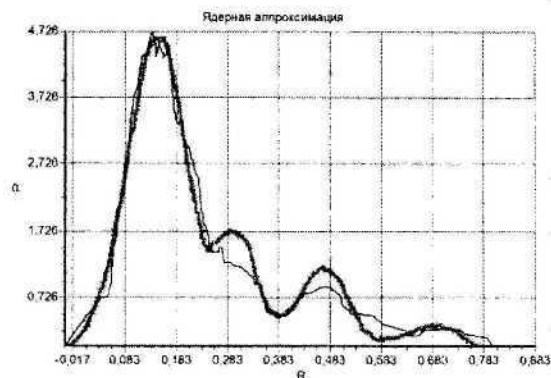


Рис.1. Оценка эмпирического распределения отношений (R) концентраций AcINH и INH теоретической смесью из четырех нормальных распределений.

С целью практического использования полученной оценки было выделено три интервала отношения R: (0 - 0,28]; (0,28 - 0,37]; [0,37 - 1). При этом левый интервал относится к достоверно медленным, правый интервал - к достоверно быстрым, а средний - к промежуточным или предбыстрым ацетиляторам. Глубокое и детальное изучение ферментативных систем биотрансформации ксенобиотиков в конкретной популяции позволит приблизиться к индивидуализированной терапии, подразумевающей эффективное и безопасное применение лекарственных средств. Для решения этой задачи была предложена формула определения дозы назначаемого лекарственного препарата с учетом предварительно выявленного у пациента отношения концентраций AcINH и INH в сыворотке крови.

ЛИТЕРАТУРА

- [1] Осипенко, А.Н. Метод и средства автоматизации моделирования активных систем: Автореф. дис... канд.техн.наук: ГГУ, Гомель(1997).
- [2] Осипенко, Н.Б. Методические и программно-технологические средства оценки и анализа сезонной динамики доз внутреннего облучения жителей населенных пунктов / Н.Б. Осипенко и др. // Известия Гомельского государственного университета имени Ф. Скорины. - 2004. - 6 (27). - С.171-176.