

Секция «Теория вероятностей и математическая статистика»
ROC-анализ и поиск пороговых значений в биогенетических исследованиях
Алексей Орловский Александрович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

E-mail: leha-ork@yandex.ru

Важным этапом в биогенетических исследованиях является выявление взаимосвязей между генетическими данными и уровнями показателей, связанных с заболеванием. При этом может быть использован широкий спектр вероятностно-статистических методов. В представленной работе пороговые значения уровня изучаемого показателя устанавливаются с помощью ROC-анализа [4] на основе генетических данных. В отличие от многочисленных работ (см., напр., [1,2]), где ROC-анализ применяется для нахождения порогового значения непрерывного показателя с использованием одной ROC-кривой, в настоящей работе при анализе генов используются две ROC-кривые с выделением двух отрезных точек с целью выявления генетических мутаций, ассоциированных с болезнью. Как известно [3], повышенный уровень окисленных липопротеидов низкой плотности (oxLDL) в плазме крови человека является фактором риска развития атеросклероза. **Цель работы:** изучить для каждого гена из некоторой группы связь с уровнем oxLDL в плазме крови для выборки пациентов ($n=726$), имеющих низкий риск развития смертельного сердечно-сосудистого заболевания по шкале SCORE. Группа генов была исследована методом Cardio-MetaboChip (компания "Illumina") на наличие мутаций. При этом каждому гену каждого пациента были сопоставлены статусы 0 (нет мутации), 1 (гетерозигота) или 2 (гомозигота). Относительно каждого гена все пациенты разбиваются на две группы двумя способами: первый - $\{0\}$ против $\{1,2\}$, второй - $\{0;1\}$ против $\{2\}$. Далее для каждого разбиения относительно уровня oxLDL (непрерывного параметра) строится ROC-кривая. По площади под кривой (AUC) делается вывод о существовании зависимости между уровнем oxLDL и принадлежностью к одному из классов разбиения. При этом производится поиск отрезного порога непрерывного параметра, который определяет критерий принадлежности к одному из классов в зависимости от уровня oxLDL. Таким образом, для некоторых генов удалось получить две различные отрезные точки по уровню oxLDL и рассмотреть три класса разбиения, каждый из которых ассоциирован с одним из трех генных статусов. Предложенный подход позволил выявить гены, статус которых связан с уровнем oxLDL в плазме крови пациентов. Полученные результаты представляются важными для профилактики развития атеросклероза на основе генного профиля у пациентов с низким риском развития сердечно-сосудистых заболеваний. Представленные данные получены в лаборатории молекулярной генетики ФГБУ «Государственный научно-исследовательский центр профилактической медицины» МЗ РФ. Анализ данных проведен с помощью пакета прикладных статистических программ SPSS v.14.

Источники и литература

- 1) Рыбачков В.В., Четверикова Е.Н., Шубин Л.Б., Кабанов Е.Н. ПРОГНОЗИРОВАНИЕ ЭФФЕКТИВНОСТИ ОПЕРАТИВНЫХ МЕТОДОВ ЛЕЧЕНИЯ ОБЛИТЕРИРУЮЩЕГО АТЕРОСКЛЕРОЗА АРТЕРИЙ НИЖНИХ КОНЕЧНОСТЕЙ // Современные проблемы науки и образования.–2015.– №2-1;
- 2) Рубаненко А. О., Шукин Ю. В. Факторы, ассоциированные с развитием тромбоза ушка левого предсердия у больных с постоянной формой фибрилляции предсердий // Клиническая медицина – № 11 / том 92 / 2014;
- 3) Tsimikas S., Brilakis E.S., Miller E.R. et al. Oxidized Phospholipids, Lp(a) Lipoprotein, and Coronary Artery Disease. N Engl J Med. July 7, 2005;353:46-57;

- 4) Tom Fawcett. ROC Graphs: Notes and Practical Considerations for Data Mining Researchers// Intelligent Enterprise Technologies Laboratory – January 7, 2003; HPL-2003-4.

Слова благодарности

Автор благодарит рук. лаборатории молекулярной генетики ФГБУ «Государственный научно-исследовательский центр профилактической медицины» МЗ РФ к.м.н. А.Н. Мешкова за постановку задачи и предоставленные данные и научного руководителя доцента кафедры теории вероятностей Е. Б. Яровую за помощь и постоянное внимание к работе.