

Секция «Теория вероятностей и математическая статистика»

**Применение адаптированного к данным критерия сумм для выявления генных маркеров сердечно-сосудистых заболеваний**

**Куценко Владимир Александрович**

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова,  
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия  
*E-mail: vlakutsenko@ya.ru*

Возникновение атеросклероза, возможно, сопряжено с высоким уровнем окисленных липопротеинов низкой плотности (окЛПНЛ) в плазме крови человека. В связи с этим основная задача данной работы – с помощью модификации критерия сумм, предложенной в [1], выяснить, является ли повышенный уровень окЛПНП следствием комбинации некоторых однонуклеотидных мутаций. В выборку включены 726 пациентов. Для каждого из них определены некоторые группы генов, т.е. наборы однонуклеотидных полиморфизмов (мутаций), и уровень окЛПНЛ. Пациенты разделяются на две группы – 1 группа это люди в зоне риска с повышенным уровнем окЛПНЛ и им присваивается значение маркера 1, вторая группа – остальные, им присваивается значение маркера 0. Наборы однонуклеотидных полиморфизмов (ОМП) кодировались определенным способом: 0 – дикий тип ОМП, 1 – гетерозиготная мутация, 2 – гомозиготная мутация. Наборы закодированных последовательностей тестировались на наличие ассоциации с уровнем окЛПНП. В качестве нулевой гипотезы принималось предположение об отсутствии ассоциации (т.е. нулевым коэффициенте в регрессионной модели). Вначале производится адаптация данных: для каждого ОМП строится модель логистической регрессии. Если полученный коэффициент отрицателен и р-значение меньше заданного уровня, то кодировка ОМП меняется на противоположную. Затем строится модель логистической регрессии, в которой в качестве независимой переменной берётся сумма кодировок ОМП из рассматриваемого набора. Используется тест множителей Лагранжа [1]. Если нулевая гипотеза верна, то функция вклада (U-статистика), которая есть производная функции максимального правдоподобия распределена как стандартное нормальное распределение. Потом рассматриваются выборки с перемешанным статусом маркера [1], к ним применяются вышеприведённые операции. По полученным U(i)-статистикам строится итоговая статистика, имеющая при H<sub>0</sub> распределение  $\chi^2$  с количеством степеней свободы d, где d – размерность U. Проблема анализа данной выборки, и вообще задач подобного типа заключается в том, что, как правило, связи между патологиями и обычными мутациями, слабы. Вследствие этого зачастую приходится проводить дополнительный анализ мутаций, встречаемость которых в популяции меньше 5% (редкие мутации). Идея модификации теста для редких мутаций состоит в комбинировании редких мутаций с минорными аллельными частотами меньше 1% в отдельный ген и тестировании получившегося набора вышеуказанным способом. В качестве дополнительных источников данных в качестве независимых можно использовать непрерывные физиологические переменные, разбив их предварительно на два класса. В итоге с помощью этого теста выявляется зависимость между некоторыми группами мутаций и повышенным уровнем окЛПНП. В работе использованы данные, полученные в лаборатории молекулярной генетики ФГБУ «Государственный научно-исследовательский центр профилактической медицины» МЗ РФ

**Источники и литература**

- 1) Fang Han Wei Pan. A Data-Adaptive Sum Test for Disease Association with Multiple Common or Rare Variants // Human Heredity 2010;70:42–54