

УДК 004

**ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО АНАЛИЗА
В ОБРАБОТКЕ МЕДИЦИНСКИХ ДАННЫХ**

Т.А. Былина, М.В. Шин

Научный руководитель: О.В. Марухина, к.т.н., доцент ТПУ

Национальный исследовательский Томский политехнический университет

E-mail: marishapicke@gmail.com, bylina_1995@mail.ru

Abstracts. *Medical data frequently represent multidimensional datasets as investigated factors and clinical and laboratory parameters coverage is huge. This research area is very important in terms of practical applications. We were given monthly lipid metabolism and hormonal status data of children (including children suffering from obesity) of Siberian region during a year. In this article some research results appear.*

Keywords: multidimensional data, data analysis, statistical methods.

Многомерные методы обработки данных. Накопленные в архивах медицинские данные актуальны для изучения механизмов ожирения среди детского населения, поскольку содержат огромный запас информации, сведений о различных случаях заболеваний, ходе и тяжести лечения, значения разнообразных клинических и лабораторных показателей.

Подобные числовые массивы показателей оформлены в виде больших числовых таблиц, поскольку охват исследуемых факторов и клинико-лабораторных показателей велик. Характерное число строк (объектов) в таблицах может достигать до нескольких десятков тысяч, а характерное число столбцов (признаков) – несколько сотен [1].

Выражение статистической информации в виде многомерных матриц оправдано, а показатели, содержащиеся в массивах, являются многомерными данными.

Анализ многомерных данных представляет собой область знаний, в которой исследуются явления, процессы и системы, характеризующие совокупностью величин. Эта область исследований является весьма важной с точки зрения практических приложений, поскольку большинство важных систем описывается набором переменных. Таким образом, проблемой является определение совокупности технического, программного и организационного обеспечения, персонала, предназначенного для того, чтобы своевременно обеспечивать медицинский персонал надлежащей информацией о состоянии пациента, а также иметь возможность проводить исследования влияния различных факторов пациентов на состояние показателей анализов и определение взаимосвязи между комплексами оздоровительных мероприятий и клинико-лабораторными показателями.

Для анализа нам были предоставлены данные среднемесячных показателей липидного обмена и гормонального статуса детей (в том числе и детей с ожирением) Сибирского региона в течение года.

Многомерные методы обработки данных, выбор инструментария. Методы многомерного статистического анализа – методы математической статистики, используемые для построения оптимальных планов сбора, систематизации и обработки многомерных статистических данных, направленные на выявление характера и структуры взаимосвязей между компонентами исследуемого многомерного признака и предназначенные для получения научных и практических выводов.

Существует множество программ для обработки статистических данных. Из наиболее популярных программных продуктов для обработки статистических данных нами был выбран пакет *STATISTICA* (разработка компании StatSoft). Пакет реализует функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов [2].

Результаты работы

Исходные данные представляют собой выборку из 250 пациентов, характеризующиеся более чем 15 признаками (клинико-лабораторными показателями). Первоначальные показатели, которые были получены в результате исследований: среднее (арифметическое) значение, медиана, мода, частота моды, стандартное отклонение.

На рис. 1 изображён график зависимости среднего значения массы тела до и после лечения от группы ожирения (для пяти пациентов).

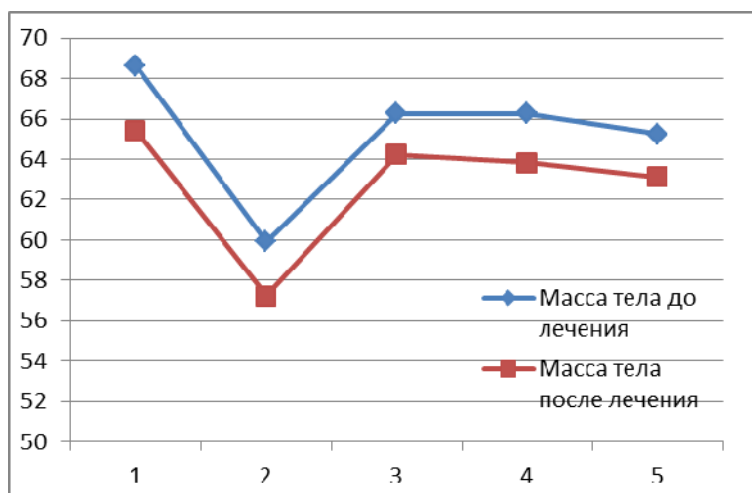


Рис. 1. Зависимость среднего значения массы тела до и после лечения от группы ожирения

Проведенный корреляционный анализ (рис. 2) показал наличие обратной связи между массой тел пациента до лечения (root1) и уровнем холестерина после лечения (var6).

Factor Structure, right set (canonical.sta)			
CANONICAL ANALYSIS	Root 1	Root 2	Root 3
VAR5	-,687636	-,725975	-,010831
VAR6	-,999690	,024603	,003881
VAR7	-,998791	,030386	-,038634

Рис. 2. Матрица корреляции

Заключение

Для решения задач анализа объектов, характеризующихся множеством признаков, применение методов анализа многомерных данных наиболее закономерно. Так, в докладе показано, что решением проблемы заболевания ожирением может послужить анализ клинико-лабораторных показателей, отражающих состояние и клиническую картину детей и нахождение закономерностей, помогающих корректировать лечение.

Исследования частично поддержаны РФФИ, проект 14-06-00026

Список литературы

1. Дюк В., Обработка данных на ПК в примерах – СПб: Питер, 1997.
2. Боровиков В. Statistica. Искусство анализа данных на компьютере. – Изд-во Питер. 2003 г. – 2-е изд. – 688 стр.