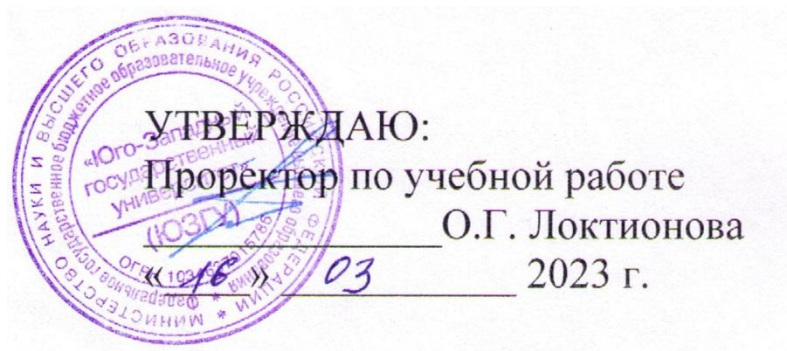


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 18.06.2023 15:30:38
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eab73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии



Лабораторный практикум по дисциплине «Математическая биология»

для студентов направления подготовки
12.03.04 «Биотехнические системы и технологии»

Курс 2023

УДК 50+51-7+57

Составители: М.В. Артеменко

Рецензент

Доктор биологических наук, профессор Привалова И.Л.

Лабораторный практикум по дисциплине «Математическая биология» / Юго-Зап. гос. ун-т; сост. М.В. Артеменко, 2023. –100с.: рисунков - 17 , таблиц – 3, Приложений - 6, библиография – стр. 26, 49, 61, 69, 82, 89, 100.

Лабораторный практикум содержит краткие теоретические сведения, порядок выполнения, содержание отчетов лабораторных работ по дисциплине «Медицинская биология» и соответствуют требованиям Федерального государственных образовательных стандартов высшего образования направлений подготовки 12.03.04 «Биотехнические системы и технологии». Рассматриваются разделы: представления информации о поведении биологических объектов инструментальными средствами офисных программ, статистическая обработка результатов лабораторных исследований, расчет показателей системной организации, анализ экологической ситуации и прогноз заболеваемости в регионе, выделение ритмических составляющих в биосигналах, применение корреляционного и автокорреляционного анализов в биомедицинских исследованиях. Экспериментальная часть работ основывается на применении компьютерных технологий для обработки результатов исследований.

Предназначено для студентов для студентов направления подготовки 12.03.04 «Биотехнические системы и технологии».

Текст печатается в авторской редакции

Подписано в печать

Формат 60x84x 1/16.

Усл.печ.л.

. Уч.-изд.л.

. Тираж 100 экз. Заказ.

Бесплатно.

Юго-Западный государственный университет.

305040, г.Курск, ул. 50 лет Октября,

СОДЕРЖАНИЕ

ЛАБОРАТОРНАЯ РАБОТА № 1. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ БИОМЕДИЦИНСКОГО ХАРАКТЕРА.....	4
ЛАБОРАТОРНАЯ РАБОТА №2. СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЛАБОРАТОРНЫХ ИССЛЕДОВАНИЙ ОБЩЕГО АНАЛИЗА КРОВИ	31
ЛАБОРАТОРНАЯ РАБОТА № 3. РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ АНАЛИЗЫ СЕРДЕЧНОГО РИТМА.	52
ЛАБОРАТОРНАЯ РАБОТА №4. РАСЧЕТ ПОКАЗАТЕЛЕЙ СИСТЕМНОЙ ОРГАНИЗАЦИИ БИООБЪЕКТА.....	63
ЛАБОРАТОРНАЯ РАБОТА №5. ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ ЗАБОЛЕВАЕМОСТИ В РЕГИОНЕ	70
ЛАБОРАТОРНАЯ РАБОТА №6. АНАЛИЗ ДИНАМИКИ ЭКОЛОГИЧЕСКОЙ СИТУАЦИИ В РЕГИОНЕ	84
ЛАБОРАТОРНАЯ РАБОТА №7. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ.....	91

ЛАБОРАТОРНАЯ РАБОТА № 1. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ БИОМЕДИЦИНСКОГО ХАРАКТЕРА

Цель работы: изучение гносеологических возможностей графических отображений биомедицинской информации в различных статических и динамических формах и овладение навыками использования графических средств стандартных инструментальных офисных пакетов.

Краткие теоретические сведения.

В процессе проведения биомедицинских и экологических исследований (включая системы мониторинга) осуществляется регистрация значений различных информационных показателей, характеризующих поведение исследуемого биообъекта как во времени так и пространстве. Для принятия решения о дальнейших наблюдениях и управлениях о биообъекте лицо принимающее решение (человек или некоторая автоматизированная система, обладающая свойствами и возможностями искусственного интеллекта) должно создать некоторый образ (модель) биообъекта. Простейший такой образ создается с помощью графических и аудио средств в динамике или статике, поскольку данные органы чувств у человека, как наиболее типичного лица принимающего решения в биомедицинских исследованиях, наиболее развиты.

Для искусственного интеллекта, применяемого в автоматизированных системах поддержки принятия решений, данные образы представляются многомерными кортежами данных или лингвистическими переменными.

Мало исследованными направлениями в данной области генерации новых знаний являются возможности создания и использования образов нацеленных на другие органы чувств (различных анализаторов человеческого мозга) человека (различных анализаторов головного и-или спинного мозгов) или использования «человеческих» образов как вторичной интегральной информации для компьютерных систем.

Визуальное представление биомедицинской информации в различных компьютеризированных системах осуществляется в следующих направлениях:

- дружественный интерфейс (в том числе графический и аудио);
- оптические изображения физических биообъектов, регистрируемые специализированной аппаратурой с помощью различных методов;
- статические изображения: графики, рисунки, «смайлики», диаграммы;
- изображения, отражающие результаты имитационного моделирования;
- таблицы с определенными образом выделенными информационными составляющими (структурами);
- динамические объекты изображений и другие мультимедийные средства.

Основной задачей визуализации данных является задача получения образа, однозначно соответствующего полученному сигналу. Обычно под *визуализацией биомедицинских данных* понимается построение графиков функций и поверхностей. В медицинских исследованиях в качестве индикатора общего состояния организма определена сердечно-сосудистая система, наиболее информативным показателем которой является ритм сердечных сокращений. Математический анализ сердечного ритма позволяет получить информацию, характеризующую состояние регуляторных механизмов, так как в ряде кардиосигналов содержится информация не только о деятельности регуляторных механизмов сердечно-сосудистой системы, но и о многочисленных функциях целостного организма. Наиболее перспективным диагностическим направлением является метод электрокардиографии.

Для отображения графической информации возможно использование готовых компонентов, содержащихся в используемой среде разработки. Такие компоненты, как правило, содержат множество свойств, методов и событий.

При этом все функции построения графиков выполняются средствами центрального процессора без использования

специализированных технологий, поддерживаемых видеоадаптером. Этот недостаток приводит к необходимости разработки собственных компонент эффективного отображения графической информации использующих возможности графических адаптеров.

Анализ информационных источников показывает, что визуализация значительного объема данных невозможна без применения специализированных графических технологий, поддерживаемых видеоадаптером . Среди существующих графических компонентов, отличающихся быстродействием можно выделить «MultipleLayer AFM Surface3D PRO» , «3D-Splot» , HOOPS Visualize .

Приведенные компоненты являются коммерческими продуктами, имеющими высокую стоимость. В этой связи обычно принимается решение о разработке собственного компонента на основе специализированных технологий, например, доступных в операционных системах Microsoft-Windows – DirectX3D или OpenGL.

Обобщенный алгоритм типовой визуализации представлен на рисунке 1.



Рис. 1. Алгоритм визуализации.

В соответствии с данным алгоритмом, программное представление результата вычислений, визуализацию которого должен осуществлять компонент, представляет собой двумерный массив вещественных значений. Таким образом, каждый элемент набора данных, подлежащий исследованию, характеризуется тремя величинами: индексом строки массива, индексом столбца массива и величиной, хранимой в ячейке массива. Естественным представлением множества точек, характеризуемых тремя величинами, является поверхность. Таким образом, визуализация заключается в построении изображения проекции трехмерного объекта. Дискретный характер определения зависимости параметров элементов набора данных не позволяет представить их в виде непрерывной поверхности. В качестве удовлетворительного приближения к идеальному визуальному представлению может быть использовано построение поверхности в виде множества плоских многоугольников, вершины которых расположены в

опорных точках, соответствующих представлению элементов набора данных. Возможность идентификации параметров элементов набора данных предоставляется за счет совмещения объекта поверхности с системой линий-отметок и цифровых и буквенных обозначений.

Для построения модели поверхности на основе двумерного массива значений массив интерпретируется как горизонтальная регулярная карта высот. Номера строк и номера столбцов массива отображаются в координаты на горизонтальной плоскости, а величина, хранящаяся в соответствующей ячейке массива, интерпретируется как высота точки над такой плоскостью. Интерпретация карты высот проиллюстрирована на рис. 2.

Строки массива входных данных обрабатываются попарно для формирования последовательности смежных треугольников, образующих элемент поверхности.

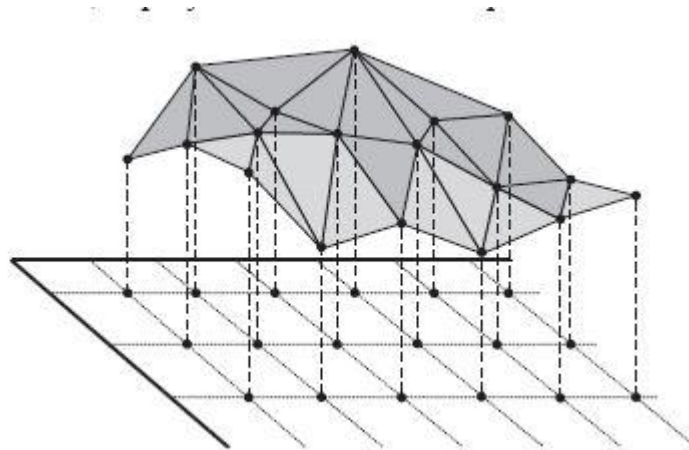


Рис. 2. Интерпретация исходных данных

Особенностью разработанного алгоритма и дополнительным способом визуализации одного из параметров элементов набора является присвоение точкам поверхности цвета в зависимости от величины одной из координат точки. Отображение величин в цветовые значения осуществляется по цветовым шкалам. Цветовые шкалы представляют собой упорядоченный набор цветовых значений, соответствующих пространству кодируемой величины. Пространство допустимых значений кодируемой величины разделено на смежные интервалы равной величины, и цвета шкалы соответствуют границам таких интервалов.

В большинстве случаев, в настоящее время в биомедицинских исследованиях используются пакеты статистической обработки данных со встроенными компонентами визуализации. Продолжительное время анализ медицинских данных был уделом специалистов, так как это требовало серьезной предварительной подготовки. С появлением и совершенствованием современных программ обработки данных статистическая обработка поднялась на новый уровень. Теперь исследователь-медик может и не иметь математической подготовки. Достаточно оперировать статистическими понятиями и, самое главное, правильно выбрать метод анализа. Все осуществимо благодаря компьютеру и новейшим программам.

Все программы статистической обработки данных можно разделить на профессиональные, полупрофессиональные (популярные) и специализированные. Статистические программы относятся к наукоемкому программному обеспечению, цена их часто недоступна индивидуальному пользователю.

Профессиональные пакеты имеют большое количество методов анализа, популярные пакеты - количество функций, достаточное для универсального применения.

Специализированные же пакеты ориентированы на какую-либо узкую область анализа данных. Создатели программных статистических пакетов заявляют, что их продукт превосходит аналоги. Отсутствие у большинства исследователей времени для освоения нескольких программ, делает непростым ее выбор. В данной статье приведена базовая информация о присутствующих на рынке основных полупрофессиональных программных пакетах пригодных для статистической обработки биомедицинских данных.

MS Excel. Самой часто упоминаемой (и используемой) в отечественных статьях является приложение MS Excel из пакета офисных программ компании Microsoft ? MS Office. Причины этого кроются в широком распространении этого программного обеспечения, наличии русскоязычной версии, тесной интеграцией с MS Word и PowerPoint. Однако, MS Excel - это электронная таблица с достаточно мощными математическими возможностями, где некоторые статистические функции являются просто

дополнительными встроенными формулами. Расчеты сделанные при ее помощи не признаются авторитетными биомедицинскими журналами. Также в MS Excel невозможно построить качественные научные графики. Безусловно, MS Excel хорошо подходит для накопления данных, промежуточного преобразования, предварительных статистических прикидок, для построения некоторых видов диаграмм. Однако окончательный статистический анализ необходимо делать в программах, которые специально созданы для этих целей. Существует макрос-дополнение XLSTAT-Pro <http://www.xlstat.com> для MS Excel который, включает в себя более 50 статистических функций, включая анализ выживаемости, которых в основных случаях достаточно для обычного применения. Пробную версию макроса можно взять на сайте производителя.

STADIA. Программа отечественной разработки с 16-и летней историей. Включает в себя все необходимые статистические функции. Она прекрасно справляется со своей задачей - статистическим анализом. Но. Программа внешне фактически не изменяется с 1996 года. Графики и диаграммы, построенные при помощи STADIA, выглядят в современных презентациях архаично. Цветовая гамма программы (красный шрифт на зеленом) очень утомляет в работе. К положительным качествам программы можно отнести русскоязычный интерфейс и наличие книг описывающих работу. Самый часто используемый пакет статистической обработки данных с более чем 30-и летней историей <http://www.spss.com> Отличается гибкостью, мощностью применим для всех видов статистических расчетов применяемых в биомедицине. Существует русскоязычное представительство компании <http://www.spss.ru>, которое предлагает полностью русифицированную версию SPSS 12.0.2 для Windows.

STATA. Профессиональный статистический программный пакет с data-management system, который может применяться для биомедицинских целей. Один из самых популярных в образовательных и научных учреждениях США наряду с SPSS. Официальный сайт <http://www.stata.com>. Программа хорошо документирована, издается специальный журнал для

пользователей системы. Однако возможности предварительного ознакомления с демо-версией нет.

STATISTICA. Производителем программы является фирма StatSoft Inc. (США) <http://www.statsoft.com> которая выпускает статистические приложения, начиная с 1985 года. STATISTICA включает большое количество методов статистического анализа (более 250 встроенных функций) объединенных следующими специализированными статистическими модулями: Основные статистики и таблицы, Непараметрическая статистика, Дисперсионный анализ, Множественная регрессия, Нелинейное оценивание, Анализ временных рядов и прогнозирование, Кластерный анализ, Факторный анализ, Дискриминантный функциональный анализ, Анализ длительностей жизни, Каноническая корреляция, Многомерное шкалирование, Моделирование структурными уравнениями и др. Несложный в освоении этот статистический пакет может быть рекомендован для биомедицинских исследований любой сложности.

JMR. Один из мировых лидеров в анализе данных. Развивает этот статистический пакет SAS Institute <http://www.jmp.com> который выкупил в конце 2002 года известную статистическую программу StatView. Однако особых преимуществ для медико-биологической статистики этот программный продукт не имеет.

SYSTAT Статистическая система для персональных компьютеров <http://systat.com> Последняя 11 версия обладает неплохим интуитивно понятным интерфейсом. Компания Systat Software также разрабатывает популярные у отечественных исследователей SigmaStat и SigmaPlot, которые являются соответственно, программой статистической обработки и программой построения диаграмм. При совместной работе становятся единым пакетом для статистической обработки и визуализации данных.

NCSS. Программа развивается с 1981 года и рассчитана на непрофессионалов в области статистической обработки. Интерфейс системы многооконный и как следствие этого явления - немного непривычный в использовании. Все действия пользователя сопровождаются подсказками. Сейчас доступна

версия 2004 г. С сайта <http://www.ncss.com> можно переписать полнофункциональную пробную версию, работающую 30 дней.

MINITAB 14. Статистический пакет MINITAB в настоящее время выпускается в версии 14. С сайта производителя <http://www.minitab.com> можно взять полнофункциональный пробный вариант программы, которая работает 30 дней. Это достаточно удобный в работе программный пакет, имеющий хороший интерфейс пользователя, хорошие возможности по визуализации результатов работы. Имеет подробную справку.

STATGRAPHICS PLUS. Довольно мощная статистическая программа. Содержит более 250 статистических функций, генерирует понятные, настраиваемые отчеты. Последняя доступная версия - 5.1. Ее можно получить на сайте <http://www.statgraphics.com>. Есть возможность скачать демо-версию. Следует отметить, что ранние версии этой программы были весьма популярны у отечественных исследователей.

PRISM. Эта программа создавалась специально для биомедицинских целей. Интуитивно понятный интерфейс позволяет в считанные минуты проанализировать данные и построить качественные графики. Программа содержит основные часто применяемые статистические функции, которых в большинстве исследований будет достаточно. Однако, как отмечают сами разработчики, программа не может полностью заменить серьезных статистических пакетов. На сайте <http://www.graphpad.com> помимо возможности ознакомления с демо-версией Prism можно получить справочник в формате PDF по биомедицинской статистике.

К рекомендациям выбора программ можно отнести:

- Если нужен мощный, общепризнанный пакет с простым и понятным даже начинающим пользователям интерфейсом, то лучше воспользоваться SPSS.
- Для начинающих и профессионалов, которым нужна подсказка и развитая документация на русском языке, можно рекомендовать STATISTICA. Это мощное приложение с профессиональными возможностями.

- Для непритязательных пользователей, которые ограничиваются в своих исследованиях стандартными статистическими методами можно рекомендовать англоязычную программу Prism.

Виды графиков.

Существует множество видов графических изображений (рис. 3 и 4). Классификация графиков основывается на ряде признаков:

1. способ построения графического образа;
2. геометрические знаки, изображающие статистические показатели;
3. задачи, решаемые с помощью графического изображения.



Рис. 3. Классификация статистических графиков по форме графического образа.

По способу построения статистические графики делятся на диаграммы и статистические карты.

Диаграммы - наиболее распространенный способ графических изображений. Это графики количественных отношений. Виды и способы их построения разнообразны. Диаграммы применяются для наглядного сопоставления в различных аспектах (пространственном, временном и др.) независимых друг от друга величин: территорий, населения и т. д. При этом сравнение исследуемых совокупностей производится по какому-либо существенному варьирующему признаку.



Рис. 4. Классификация статистических графиков по способу построения и задачам изображения

Статистические карты - графики количественного распределения по поверхности. По своей основной цели они близко примыкают к диаграммам и специфичны лишь в том отношении, что представляют собой условные изображения статистических данных на контурной географической карте, т. е. показывают пространственное размещение или пространственную распространенность статистических данных.

Геометрические знаки, как было сказано выше, - это либо точки, либо линии или плоскости, либо геометрические фигуры. В соответствии с этим различают графики точечные, линейные, плоскостные и пространственные (объемные).

При построении точечных диаграмм в качестве графических образов применяются совокупности точек; при построении линейных - линии. Основным принципом построения всех плоскостных диаграмм сводится к тому, что статистические величины изображаются в виде геометрических фигур и, в свою очередь, подразделяются на

- столбиковые,
- полосовые,
- круговые,
- квадратные,
- фигурные.

Статистические карты по графическому образу делятся на

- картограммы
- картодиаграммы.

В зависимости от круга решаемых задач выделяют

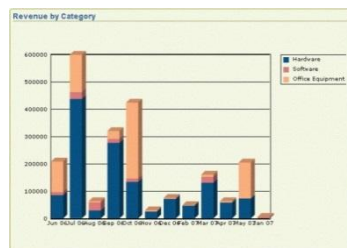
- диаграммы сравнения,
- структурные диаграммы
- диаграммы динамики.

Особым видом графиков являются диаграммы распределения величин, представленных вариационным рядом. Это гистограмма, полигон, огива, кумулята.

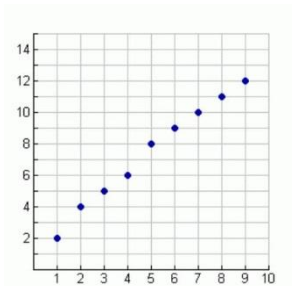
ВИДЫ ДИАГРАММ

Прежде чем составить какой либо график, необходимо определиться с вопросом о том, какие виды диаграмм вас именно интересуют. Рассмотрим основные из них.

Гистограмма. Само название этого вида позаимствовано из греческого языка. Дословный перевод – писать столбом. Это своеобразный столбчатый график. Диаграммы в Excel такого вида могут быть объемные, плоские, отображать вклады (прямоугольник в прямоугольнике) и т.д.

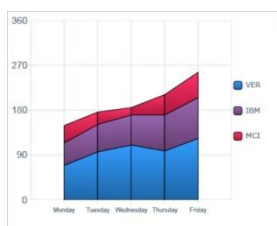


Точечная диаграмма показывает взаимную связь между числовыми данными в некотором количестве рядов и представляет собой пару групп цифр или чисел в виде единственного ряда точек в координатах. Виды диаграмм такого типа отображают кластеры данных, используются для научных целей. При предварительной подготовке к построению точечной диаграммы все данные, которые вы хотите расположить по оси ординат, следует расположить в одной строке/столбце, а значения по оси «абсцисс» - в смежной строке/столбце.

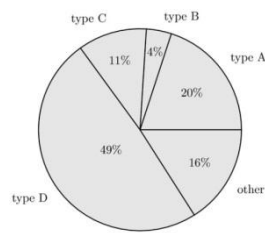


Линейчатая диаграмма и график. Диаграмма линейчатая описывает некое соотношение отдельных данных. На такой диаграмме значения располагаются по вертикальной оси, категории же – по горизонтальной. Из этого следует, что большее внимание такая диаграмма уделяет сопоставлению данных, нежели изменениям, происходящим с течением времени. Данный вид диаграмм существует с параметром «накопление», что позволяет показать взнос отдельных частей в общий конечный результат. График же отображает последовательность изменений числовых значений за абсолютно равные промежутки времени. Эти виды диаграмм наиболее часто используются для построений.

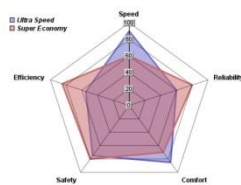
Диаграммы с областями. Основной целью такой диаграммы является акцент на величине изменения данных в течение некоторого периода, путем показа суммирования введенных значений. А также отображение доли отдельно взятых значений в общей сумме.



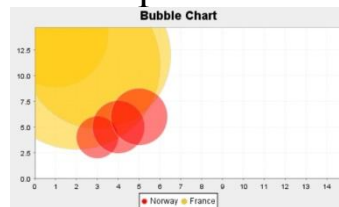
Кольцевая и круговая диаграммы. Данные виды диаграмм весьма схожи по целям. Обе они отображают роль каждого элемента в общей сумме. Их отличие заключается лишь в том, что диаграмма кольцевая имеет возможность содержать несколько рядов с данными. Каждое отдельное вложенное кольцо представляет собой индивидуальный ряд значений/данных.



Лепестковая. В данном случае каждая категория представляет индивидуальную координатную ось, исходящую от нулевой точки координат. Данный вид диаграмм позволяет сравнивать общие значения из некоторого количества введенных данных.



Пузырьковая. Одна из разновидностей точечной. Величина маркера зависит от величины третьей переменной. При предварительной подготовке располагать данные следует точно так же, как и при подготовке к построению точечной диаграммы.



Биржевая диаграмма. Использование таковой часто является неотъемлемым процессом при продаже акций или других ценных бумаг. Также возможно ее построение для наглядного определения изменения температурных режимов. Для трех и пяти значений такой вид графика может содержать в себе пару осей: первую – для столбиков, которые представляют интервал неких колебаний, вторую – для изменения ценовой категории.



Это лишь малая часть типов диаграмм, которые могут вам понадобиться. Выбор всегда зависит от целей.

Мультимедийные средства.

Понятие мультимедиа, вообще, и средств мультимедиа, в частности, с одной стороны тесно связано с компьютерной обработкой и представлением разнотипной информации и, с другой стороны, лежит в основе функционирования средств ИКТ, существенно влияющих на эффективность образовательного процесса. *Мультимедиа* - это:

- технология, описывающая порядок разработки, функционирования и применения средств обработки информации разных типов;
- информационный ресурс, созданный на основе технологий обработки и представления информации разных типов;
- компьютерное программное обеспечение, функционирование которого связано с обработкой и представлением информации разных типов;
- компьютерное аппаратное обеспечение, с помощью которого становится возможной работа с информацией разных типов;
- особый обобщающий вид информации, которая объединяет в себе как традиционную статическую визуальную (текст, графику), так и динамическую информацию разных типов (речь, музыку, видео фрагменты, анимацию и т.п.).

Таким образом, в широком смысле термин "мультимедиа" означает спектр информационных технологий, использующих различные программные и технические средства с целью наиболее эффективного воздействия на пользователя (ставшего одновременно и читателем, и слушателем, и зрителем).

Средства, используемые при создании мультимедийных продуктов:

- системы обработки статической графической информации;
- системы создания анимированной графики;
- системы записи и редактирования звука;
- системы видеомонтажа;
- системы интеграции текстовой и аудиовизуальной информации в единый проект.

Рассмотрим более подробно современные мультимедийные средства, перспективные для медицинского использования.

- *3D Очки.* Область применения – эффект присутствия для обучающихся во время реально проводимой операции, обследования больного.
- *Web-Камеры.* Web-камера - это стационарно установленная камера, имеющая встроенный web-сервер, сетевой интерфейс и подключающаяся непосредственно к LAN/ WAN/ Internet. Многие сетевые камеры имеют такие дополнительные средства как: детекторы движения, отправка сообщений по e-mail, работа с модемом, подключение внешних датчиков и пр. Пользователи могут обращаться к камере посредством стандартного web браузера.
- *мультимедиа компьютер, сканер, клавиатура.* В основе - светочувствительный сенсор - это своего рода сердце любой цифровой камеры. Именно он позволяет преобразовывать свет в электрические сигналы, доступные для дальнейшей электронной обработки. Основной принцип действия и ПЗС - и КМОП-сенсоров одинаков: под воздействием света в полупроводниковых материалах рождаются носители заряда, которые впоследствии преобразуются в напряжение. Различие между ПЗС - и КМОП-сенсорами заключается, прежде всего, в способе накопления и передачи заряда, а также в технологии преобразования его в аналоговое напряжение. Не вдаваясь в подробности конструкции различных типов сенсоров, отметим лишь, что КМОП-сенсоры являются значительно более дешевыми в производстве, но и более «шумными». Принцип работы Web-камеры схож с принципом работы любой цифровой камеры или фотоаппарата. Кроме оптического объектива и светочувствительного ПЗС - или КМОП-сенсора обязательным является наличие аналого-цифрового преобразователя (АЦП), основное назначение которого -- преобразовывать аналоговые сигналы светочувствительного сенсора, то есть напряжение в цифровой код. Кроме того, необходима система цветоформирования. Другим важным элементом камеры является схема, отвечающая за компрессию данных и подготовку к передаче в нужном формате. В Web-камерах видеоданные передаются в компьютер по USB-интерфейсу, то есть заключительной схемой камеры должен быть контроллер USB-интерфейса.

Сканер (англ. scanner) - устройство, которое, анализируя какой-либо объект (обычно изображение, текст), создаёт цифровую копию изображения объекта. Процесс получения этой копии называется сканированием. В большинстве сканеров для преобразования изображения в цифровую форму применяются светочувствительные элементы на основе приборов с зарядовой связью (ПЗС). По способу перемещения считывающей головки и изображения относительно друг друга сканеры подразделяются на ручные (англ. Handheld), рулонные (англ. Sheet-Feed), планшетные (англ. Flatbed) и проекционные. Разновидностью проекционных сканеров являются слайд-сканеры, предназначенные для сканирования фотопленок. В высококачественной полиграфии используются барабанные сканеры, в которых в качестве светочувствительного элемента используется фотоэлектронный умножитель (ФЭУ). Принцип работы однопроходного планшетного сканера состоит в том, что вдоль сканируемого изображения, расположенного на прозрачном неподвижном стекле, движется сканирующая каретка с источником света. Отраженный свет через оптическую систему сканера (состоящую из объектива и зеркал или призмы) попадает на три расположенных параллельно друг другу светочувствительных полупроводниковых элемента на основе ПЗС, каждый из которых принимает информацию о компонентах изображения.

Динамический диапазон сканера - это показатель технических возможностей сканеров, характеризующий интервал оптических плотностей, который воспринимается сканером. Основной характеристикой любого оригинала является его оптическая плотность, определяющаяся способностью оригинала отражать или пропускать свет. Оптическая плотность лежит в пределах от 0, что соответствует белому цвету, до 4, что соответствует черному цвету и обозначается OD (Optical Density) или просто D. Динамический диапазон (Dynamic Range), или диапазон плотности (Density Range), определяется как разница между самым светлым (D_{\min}) и самым темным (D_{\max}) участками оригинала и зависит от типа оригинала и его происхождения. Применительно к сканеру, динамический диапазон определяется как разница между самым светлым (D_{\min}) и самым темным (D_{\max}) участками оригинала,

которые сканер в состоянии обработать. С увеличением динамического диапазона сканера возрастает количество вводимых градаций яркости и, следовательно, плавность переходов в смежных тонах изображения. Недостаточный динамический диапазон сканера может привести к искажениям цветопередачи при сканировании изображений, содержащих плавные тоновые переходы (переходы яркости), наподобие фотоснимков голубого неба, заката, или к потере деталей в снимках светлых и темных предметов: цветов, белой одежды, облаков, “лунной дорожки” тень от здания и т.д. Напротив, сканер, имеющий высокий показатель динамического диапазона передает оригинал настолько "объемно", что, к примеру, отсканированные со слайда облака, кажется, движутся по экрану.

Мультимедийная компьютерная клавиатура, способная управлять громкостью звука и сетевым поведением компьютера. Многие современные компьютерные клавиатуры, помимо стандартного набора из ста четырёх клавиш, снабжаются дополнительными клавишами (как правило, другого размера и формы), которые предназначены для упрощённого управления некоторыми основными функциями компьютера:

- управление громкостью звука: громче, тише, включить или выключить звук;
- управление лотком в приводе для компакт-дисков: извлечь диск, принять диск;
- управление аудиопроигрывателем: играть, поставить на паузу, остановить воспроизведение, промотать аудиозапись вперёд или назад, перейти к следующей или предыдущей аудиозаписи;
- управление сетевыми возможностями компьютера: открыть почтовую программу, открыть браузер, показать домашнюю страницу, двигаться вперёд или назад по истории посещённых страниц, открыть поисковую систему;
- управление наиболее популярными программами: открыть калькулятор, открыть файловый менеджер;
- управление состоянием окон операционной системы: свернуть окно, закрыть окно, перейти к следующему или к предыдущему окну;

· управление состоянием компьютера: перевести в ждущий режим, перевести в спящий режим, пробудить компьютер, выключить компьютер.

Так как многие из этих функций (управление звуком и воспроизведением звукозаписей, управление компакт-дисками и т. п.) относятся к сфере мультимедиа, то такие клавиатуры часто называются «мультимедийными клавиатурами».

Виртуальная лазерная клавиатура. Идея реализации виртуальной клавиатуры без проводов и кнопок родилась несколько лет назад в стенах израильской компании Developer VKB Inc. Представленная на выставке CeBIT 2002 компанией Siemens Procurement Logistics Services первая виртуальная клавиатура без единого механического или электрического элемента стала первой практической реализацией этой идеи. Разработчики лазерного интерфейса виртуальной клавиатуры предполагали, что их разработка на практике может быть интегрирована в любое мобильное устройство - телефон, ноутбук, планшетный ПК и даже в стерильное медицинское оборудование. Принцип работы виртуальной лазерной клавиатуры прост и понятен без долгих объяснений. В конструкции используется два полупроводниковых диодных лазера - "красный" для создания проекции клавиатуры и невидимый инфракрасный с фотодетектором ИК-излучения для определения клавиши, к которой прикоснулся ваш палец. Пока вы непринужденно набираете текст по лазерной проекции клавиш - как на обычной клавиатуре, невидимый луч анализирует координаты положения пальцев и обрабатывает полученную информацию соответствующим образом. Добавляем к этой конструкции беспроводной интерфейс Bluetooth - и виртуальная клавиатура для любых типов стационарных и мобильных устройств - ПК, ноутбуков, карманных ПК или смартфонов, готова.

The Orbitouch. Данный агрегат выглядит как порождение злобного инопланетного разума, однако, на самом деле это тоже всего лишь клавиатура. Сразу же возникает вопрос - как же с ней работать? Ну, объяснить достаточно просто - выступы вращаются, а буквы набираются в соответствии с тем в каких позициях они стоят. Для каждой "ручки" есть восемь позиций, так что на числа и

буквы должно хватить. В клавиатуре есть встроенная мышь, так что тут покупатель может даже сэкономить.

Компьютерный руль. Компьютерный руль - игровой контроллер, имитирующий автомобильный руль. Применяется для игры в компьютерные игры - автосимуляторы. Помимо рулевого колеса и двух (трёх) педалей, в компьютерном руле могут быть такие органы управления. В медицине используется для изучения реакции человека как компонента эргатической системы. Компьютерный руль является потомком джойстика; первые рули действительно эмулировали двухосный джойстик. Существуют два рудимента того времени. Первый джойстик-руль для компьютерных игр появился в 1983 году. Это была обычная пластиковая коробка с баранкой диаметром 17 см и единственной гладкой педалью. Далее производители начали развивать идею. Постепенно они пришли к выводу, что какая-никакая отдача позволила бы игрокам прочувствовать дорогу намного лучше, да и интерактивность поднялась бы на несколько ступеней. Самый простой способ достижения подобного эффекта - установка вибромотора. Предположим, вы заехали одним колесом на обочину - игра посылает сигнал на джойстик, вибромотор начинает легко потряхивать баранку. Зачастую в прайсах модели с вибромотором обозначают как рули с обратной связью или Rumble Feedback.

Проектор. Проектор - световой прибор, перераспределяющий свет лампы с концентрацией светового потока на поверхности малого размера или в малом объёме. Проекторы являются в основном оптико-механическими или оптически-цифровыми приборами, позволяющими при помощи источника света проецировать изображения объектов на поверхность, расположенную вне прибора - экран. Появление проекционных аппаратов обусловило возникновение кинематографа, относящегося к проекционному искусству. Виды проекционных приборов:

- Диаскопический проекционный аппарат - изображения создаются при помощи лучей света, проходящих через светопроницаемый носитель с изображением. Это самый распространённый вид проекционных аппаратов. К ним относят такие приборы как:

кинопроектор, диапроектор, фотоувеличитель, проекционный фонарь, кодоскоп и др.

- Эпископический проекционный аппарат - создаёт изображения непрозрачных предметов путём проецирования отраженных лучей света. К ним относятся эпископы, мегаскоп.

- Эпидиаскопический проекционный аппарат - формирует на экране комбинированные изображения как прозрачных, так и непрозрачных объектов.

Мультимедийный проектор (также используется термин «Цифровой проектор») - с появлением и развитием цифровых технологий это наименование получили два, вообще говоря, различных класса устройств. На вход устройства подаётся видеосигнал в реальном времени (аналоговый или цифровой). Устройство проецирует изображение на экран. Возможно, при этом наличие звукового канала. Устройство получает на отдельном или встроенном в устройство носителе или из локальной сети файл или совокупность файлов («слайдшоу») - массив цифровой информации. Декодирует его и проецирует видеоизображение на экран, возможно, воспроизводя при этом и звук. Фактически, является сочетанием в одном устройстве мультимедийного проигрывателя и собственно проектора. Название «цифровой проектор» связано, прежде всего, с обычным ныне применением в таких проекторах цифровых технологий обработки информации и формирования изображения.

Лазерный проектор - выводит изображение с помощью луча лазера.

Области применения мультимедийных средств охватывают различные виды интеллектуальной деятельности: науку и технику, образование, культуру, бизнес, а также применяются в среде обслуживания при создании электронных гидов с погружением в реальную среду, мультитеках.

Одной из основных сфер применения систем мультимедиа является образование в широком смысле слова, включая и такие направления как видеоэнциклопедии, интерактивные путеводители, тренажеры, ситуационно-ролевые игры и др. Компьютер, снабженный платой мультимедиа, немедленно становится универсальным обучающим или информационным

инструментом по практически любой отрасли знания и человеческой деятельности. Очень большие перспективы перед мультимедиа в медицине: базы знаний, методики операций, каталоги лекарств и т.п. В сфере бизнеса фирма по продаже недвижимости уже используют технологию мультимедиа для создания каталогов продаваемых домов - покупатель может увидеть на экране дом в разных ракурсах, совершить интерактивную видеопрогулку по всем помещениям, ознакомиться с планами и чертежами. Технологические мультимедиа пользуется большим вниманием военных: так, Пентагон реализует программу перенесения на интерактивные видеодиски всей технической, эксплуатационной и учебной документации по всем системам вооружений, создания и массового использования тренажеров на основе таких дисков.

Весьма перспективными выглядят работы по внедрению элементов искусственного интеллекта в системе мультимедиа. Они обладают способностью "чувствовать" среду общения, адаптироваться к ней и оптимизировать процесс общения с пользователем; они подстраиваются под читателей, анализируют круг их интересов, помнят вопросы, вызывающие затруднения, и могут сами предложить дополнительную или разъясняющую информацию. Системы, понимающие естественный язык, распознаватели речи еще более расширяют диапазон взаимодействия с компьютером.

Еще одна быстро развивающаяся, совершенно уже фантастическая для нас область применения компьютеров, в которой важную роль играет технология мультимедиа - это системы виртуальной, или альтернативной реальности, а также близкие к ним системы "телеприсутствия". С помощью специального оборудования - система с двумя миниатюрными стереодисплеями, квадранаушниками, специальных сенсорных перчаток и даже костюма вы можете "войти" в сгенерированный или смоделированный компьютером мир, (а не заглянуть в него через плоское окошко дисплея) повернув голову, посмотреть налево или направо, пройти дальше, протянув руку вперед - и увидеть ее в этом виртуальном мире; можно даже взять какой либо виртуальный предмет (почувствовав при этом его тяжесть) и

переставить в другое место; можно таким образом строить, создавать этот мир изнутри.

Порядок выполнения работы

1. Изучите теоретический материал.
2. По таблице данных результатов анализа крови больных для каждой представленной характеристики постройте в Excel диаграммы различных типов и графики в разных формах (с моделями линий тренда).
3. По моделям, отражающим популяционные изменения и «хищник-жертва» постройте графики функций, изменяющиеся во времени (для модели «хищник-жертва» - график, отражающий зависимости численностей «жертвы» и «хищника»).
4. Изучите графическое отражение динамики игры «Жизнь» по электронной версии и зафиксируйте несколько «скрин-шотов» (см., например, <http://www.nature.air.ru/models/models.htm>).
5. По данным приложения составьте в Excel отчетную таблицу, содержащую: заголовок, наименование регистрируемых характеристик, среднее значение по каждой характеристике, диаграммы (полученные в п.2 – выборочно), выделение цветом и буквами строк, у которых большинство показателей выше средних значений (буквы формируются с помощью оператора выполнения условия. Постарайтесь совместно с цветовой гаммой использовать «смайлики»).
6. Осуществите представление динамически изменяющейся картины («слайдшоу»), отражающей функционирование любого органа человека (применять любые мультимедийные средства и графические редакторы – с указанием задействованных команд). Разрешается использовать сменяющиеся слайды.
7. Оформите отчет, включающий результаты выполнения работы и краткие ответы на контрольные вопросы.
8. Составьте аннотацию не менее 2 информационных источников, найденных в сети интернет по тематике лабораторной работы (объемом 150-350 слов).

Контрольные вопросы.

1. Что такое графическое представление информации (числовой и семантической)?
2. Какие мультимедийные средства применяются в операционной?
3. Какие мультимедийные средства применяются в системах прикроватного мониторинга?
4. Для чего предназначены «смайлики»?
5. Какие графики можно отображать в различных пакетах статистической обработки?
6. Какие диаграммы можно отображать в различных пакетах статистической обработки?
7. Как осуществляется приведение реальных графических изображений в окна определенного размера?
8. Каким образом отображаются трехмерные объекты на плоскостном экране монитора?
9. Каким образом осуществляется «слайдшоу»?
10. Каким образом используются графические изображения при имитационном моделировании?

Библиография

1. Биологические имитационные модели /URL: <http://www.nature.air.ru/models/models.htm>.
2. Биометрика /URL: <http://www.biometrica.tomsk.ru>
3. Диаграммы. /URL: <http://fb.ru/article/71308/vidyi-diagramm-i-ih-osobennosti>
4. Кулаичев А.П. Методы и средства анализа данных в среде Windows. - М: ИнКо, 2002. - 341 с. /URL: <http://www.protein.bio.msu.su/~akula/index.htm>
5. Разработка компонента визуализации биомедицинских данных на основе технологии OPENGL /URL: <http://ilab.xmedtest.net/?q=node/5870>
6. Современное программное обеспечение для статистической обработки биомедицинских исследований /URL: <http://www.disser.ru/library/31/440.htm>

Приложение 1

Пример математических моделей биологических систем

1. Игра «Жизнь».

Многие процессы, интересующих медиков (например, динамика возрастного состава в регионе) может быть с определенной точностью описано математической моделью. Самая простая модель известна под названием «Игра Жизнь».

Предполагается наличие прямоугольного клетчатого поля, в каждой клетке которого может «жить» существо. Если клетка пустая – то в ней никто не живет. Модель задается двумя параметрами: начальной конфигурацией (размером поля и расположением живых существ) и определенными «биологическими законами», регулирующими жизнь популяции существ. На каждой итерации осуществляется последовательный просмотр всех клеток с некоторой начальной (координаты клетки выбираются исследователем или случайным образом) с применением к ним биологических законов.

В качестве типовых законов, предлагаются, например, следующие:

1. Если выбранная клетка пуста, а в соседней с ней клетках находится более двух существ, то внутри клетки появляется существо («размножение»).
2. Если выбранная клетка непуста, а в соседних с ней клетках живет меньше трех или больше четырех существ, то клетка очищается (существо в ней погибает от одиночества или перенаселения).
3. Если правила 1 и 2 не выполняются, то ничего с клеткой не происходит. (Под соседними подразумеваются восемь окружающих клеток, за границей ореала – прямоугольного поля – существ нет).

2. Модель «хищник» - «жертва».

В системе хищник-жертва ситуация моделируется следующим образом. В случае конкурирующих популяций исчезновение одной означает выигрыш для другой в борьбе за дополнительные ресурсы. Обозначим через C численность популяции хищника, N – популяцию жертвы. Наиболее популярная модель, отражающая колебания численности имеет вид:

$$\begin{aligned} N_{i+1} &= N_i + (r \cdot N_i - a \cdot N_i \cdot C_i) \cdot \Delta t, \\ C_{i+1} &= C_i + (-q \cdot C_i - a \cdot f \cdot N_i \cdot C_i) \cdot \Delta t. \end{aligned}$$

Согласно первому уравнению при $C=0$ численность жертв быстро растет со скоростью r , поскольку модель не учитывает внутривидовой конкуренции. Скорость роста числа жертв ($\frac{\Delta N}{\Delta t}$) уменьшается тем больше, чем чаще происходят встречи особей видов (тогда a - коэффициент эффективности поиска).

Второе уравнение показывает, что в отсутствии жертв численность хищников быстро убывает со скоростью q : положительное слагаемое в правой части уравнения компенсирует эту убыль, f – коэффициент эффективности перехода пищи к потомству хищников.

3. Внутривидовая конкуренция в популяции с дискретным размножением.

Для популяций с дискретным размножением (некоторые виды растений, насекомые) поколения дифференцированно разнесены во времени и особи разных поколений вместе не сосуществуют. Численность подобной популяции характеризуется числом N_t , а время t – дискретная величина – можно, в первом приближении, считать номером популяции. Тогда одна из моделей межвидовой конкуренции может быть описана уравнением:

$$N_{t+1} = \frac{N_t \cdot R}{1 + (a \cdot N_t)^b}$$

где R – скорость воспроизводства популяции в отсутствие внутривидовой конкуренции (математически это соответствует $a=0$); a – параметр, характеризующий интенсивность внутривидовой конкуренции, при $b=1$ осуществляется выход численности популяции на стационарное значение при любых значениях других параметров модели.

Знаменатель в уравнении отражает наличие конкуренции, делающей скорость роста тем меньше, чем больше численность популяции. Данная модель описывает четыре вида эволюции:

1. монотонное установление стационарной численности популяции;
2. колебательное установление стационарной численности популяции;
3. устойчивые предельные циклы изменения численности популяций;
4. случайные изменения численности популяции без наличия явных закономерностей.

4. Внутривидовая конкуренция в популяции с непрерывным размножением.

В данном случае численность популяции $N(t)$ является непрерывной функцией во времени. В начале эволюционного процесса численность популяции невелика, а ее удельная скорость не зависит от численности:

$\frac{1}{N} \cdot \frac{\Delta N}{\Delta t} = r$ – скорость роста численности популяции в отсутствие

конкуренции. Далее, по мере роста численности, скорость роста начинает уменьшаться и при достижении определенного критического значения K обращается в ноль. Таким образом, в первом приближении, математическая модель имеет вид:

$$N_{i+1} = N_i + r \cdot N_i \cdot \left(\frac{K - N_i}{K}\right) \cdot \Delta t .$$

Приложение 2

Показатели крови пациентов отделения гастроэнтрологии.

эритроциты	гемоглобин	Цветовой показатель	лейкоциты	эозинофилы	палочкоядерные	лимфоциты	моноциты
4,6	142	0,93	7,05	1	14	38	2
4,2	115	0,82	7,8	2	4	10	2
3,4	107	0,94	8	1	6	27	2
3	87	0,87	7,3	4	1	36	4
3,4	100	0,88	6,15	0	4	17	2
5,0	170	1,02	7,1	2	5	30	4
5	150	0,90	3,7	6	1	21	1
3,6	105	0,88	17,6	0	7	26	2
5,3	108	0,61	4,2	4	1	46	2
4,16	132	0,95	5,1	2	4	47	4
4,3	145	1,01	4,55	4	5	36	3
4,0	128	0,96	7,6	4	2	29	7
5,0	160	0,96	6,8	0	4	23	1
4,46	146	0,98	6,5	4	1	27	2
5,4	176	0,98	7,8	2	2	29	0
3,1	108	1,05	3,7	1	2	31	4
4,0	124	0,93	4,4	2	7	22	3
5,1	164	0,96	7,9	0	5	60	6
4,25	138	0,97	6,2	2	7	28	4
4,8	160	1,00	6,7	2	2	25	1
4,3	150	1,05	12	1	1	30	9
4,0	113	0,85	5,85	6	4	37	5
4,8	153	0,96	7,1	6	1	30	2
3,6	114	0,95	13,1	0	2	24	1
3,35	103	0,92	9,1	1	3	27	17

ЛАБОРАТОРНАЯ РАБОТА №2. СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЛАБОРАТОРНЫХ ИССЛЕДОВАНИЙ ОБЩЕГО АНАЛИЗА КРОВИ

Цель работы: овладение навыками использования «Пакета анализа» в Excel для проведения разведочного статистического анализа медицинских данных в рамках методологии доказательной медицины на примере результатов лабораторных исследований общего анализа крови.

Краткие теоретические сведения

Доказательная медицина (Evidence-based medicine — медицина, основанная на доказательствах) — это подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их эффективности и безопасности, а такие доказательства подвергаются поиску, сравнению, обобщению и широкому распространению для использования в интересах пациентов.

В основе доказательной медицины лежит проверка эффективности и безопасности методик диагностики, профилактики и лечения в клинических исследованиях. Под практикой доказательной медицины понимают использование данных, полученных из клинических исследований, в повседневной клинической работе врача.

Некоторые правила проведения клинических исследований изложены в стандарте GCP (Good Clinical Practice, «Надлежащая клиническая практика»), правила производства лекарственных средств — в стандарте GMP, правила выполнения лабораторных исследований — в стандарте GLP.

В начале 1990-х годов разработана рейтинговая система оценки клинических исследований, где с возрастанием порядкового номера доказательности качество клинических исследований снижается. Уровни принято обозначать римскими цифрами (I, II, III, IV) или буквами латинского алфавита (A, B, C, D). Цифры обозначают уровень доказательности

результатов научных исследований. Буквы обозначают уровень доказательности принятых рекомендаций.

Класс (уровень) I (A): большие двойные слепые плацебо контролируемые исследования, а также данные, полученные при мета-анализе нескольких рандомизированных контролируемых исследований.

Класс (уровень) II (B): небольшие рандомизированные контролируемые исследования, в которых статистические расчёты проводятся на ограниченном числе пациентов.

Класс (уровень) III (C): нерандомизированные клинические исследования на ограниченном количестве пациентов.

Класс (уровень) IV (D): выработка группой экспертов консенсуса по определённой проблеме.

Общим для всего направления является использование принципа доказательности на любом уровне принятия решений — от государственной программы до назначения индивидуальной терапии.

Таким образом, доказательный подход к медицинской практике подразумевает использование лечебных, профилактических и диагностических действий с доказанной эффективностью, что предполагает поиск информации, сравнение, проведение исследований и мета-анализов. Эффективность доказывается не собственным опытом или личным мнением.

В узком понимании подразумевается такой способ медицинской практики отдельного врача, когда он использует в своей работе только то, что имеет качественную доказательную базу эффективности (к этому не относится собственный опыт и личное мнение, они ставятся на последнее место — когда всё доказанное кончилось или случай неординарный, тогда и личный опыт подойдёт). Ясно, что объективность сама по себе субъективна, особенно в медицине, но личный опыт субъективнее всего остального, поэтому к нему так и относятся. Подход к лечению только на основе опыта и впечатлений называется «импрессионистская медицина».

В основе ДМ лежит клиническая эпидемиология, которая изучает распространение заболеваний и разрабатывает методологию клинических испытаний таким образом, чтобы они

приводили к научно обоснованным выводам, минимизируя случайные и систематические ошибки, а также заинтересованность участников. В целом это такой свод принципов для опровержения или подтверждения научных данных.

Противники доказательной медицины любят разглагольствовать в стиле, мол, если не доказано действие, ещё не значит, что его нет. Такой подход отсылает нас к области веры, а не науки: именно в религии не нужно ничего доказывать. Так что не стоит обращать внимания на такие логические ошибки, это просто аргумент адептов астрала.

Последнее время контроль нередко проводится не с плацебо, а с конвенциональным лечением (уже утверждённым и многократно проверенным), в тех случаях, когда нельзя пациентов оставить совсем без лечения. Тогда одной группе дают стандартную терапию, а другой стандартную + исследуемую, и сравнивают различия. Каждое исследование должно иметь подтверждения своих выводов в массе подобных, иначе практические выводы сделать нельзя; для анализа и оценки статей есть толпы экспертов, которые в итоге выдают простым смертным врачам клинические рекомендации (гайдлайны, «guidelines») на основе лучших доказательств. При изучении работ надо строго разделять показатели процесса (любые изменения параметров) и показатели собственно результата (именно они имеют клиническую значимость), к которым приводят те изменения. Во время чтения публикаций или спора со сторонником какого-либо метода нужно придерживаться этого разделения, поскольку не составляет труда показать действие возможного фактора на процесс, а вот выяснение достоверного результата и его положительной связи с тем фактором требует серьёзной работы. Исследовательские работы можно отнести к тому или иному виду по весу доказательности, что зависит от его структуры (по уменьшению крутизны):

Систематический обзор посредством мета-анализа: формируется множество пачка схожих клинических испытаний одного метода, высчитываются их общие и различные параметры, анализируется совпадение/несовпадение результатов.

РКИ (рандомизированное клиническое исследование, «randomized clinical trial», «RCT») является основой доказательности доказательности - различия, что же явилось следствием воздействия, а что — случайностью. Заключается в динамическом наблюдении профилактического – диагностического - лечебного вмешательства, которые применяются к случайно сформированным (рандомизированным) группам из конкретной выборки пациентов. Все возможные факторы одинаково действуют на группы подопытных, только в одной это будет полностью плацебо-эффект, а во второй — непосредственно эффект медицинского вмешательства, из которого можно вычистить первый и получить кристаллизованную достоверность в виде подтверждения/опровержения изначальной гипотезы.

Популяционное (проспективное, когортное, продольное) исследование осуществляется следующим образом: выделяются два отряда населения (когорты), например, которые подвергались фактору риска и не подверженные ему, затем за ними длительно наблюдают, обследуют и сравнивают данные. Используется для определения прогноза и причин заболеваний, их факторов риска и уровня заболеваемости, что весьма трудоёмко из-за необходимости больших выборок (новые случаи заболеваний могут оказаться слишком редкими) и длительности наблюдения этих больших групп.

Аналитическое одномоментное исследование (поперечное): используется для оценки эффективности диагностики, распространённости исходов и течения заболеваний практически в реальном времени — по сути это срез базы данных по каким-либо критериям.

Исследование случай—контроль (ретроспективное): осуществляется статистический анализ архива историй болезни, что позволяет получить относительно точные данные (без внешнего влияния — ведь все наблюдения произошли уже до анализа), на основе которых вполне можно выдвинуть какую-либо гипотезу.

Описание серии случаев: широко используется, но по сути это тот самый «личный многолетний опыт», доказательной ценности имеющий совсем чуть-чуть, поскольку если человек хочет что-то

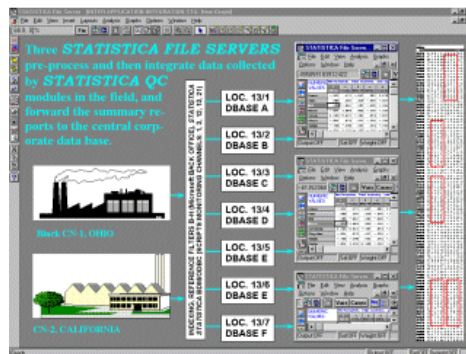
видеть, то он увидит это и на первый, и на десятый, и на тысячный раз.

Обзор ресурсов доказательной медицины.

Кокрановское сотрудничество. «Cochrane Collaboration» - крупнейшая международная организация для предоставления пруфов или опровержения лекарств/методов с помощью метаанализов, объединяя многие разнородные данные в общие темы. Выводы мета-анализов Кокрейна намного более важны, чем чьи-либо личные сведения о многолетнем опыте применения какого-либо фуфломицина.

NCBI - The National Center for Biotechnology Information - Национальный (американский, естественно) центр биотехнологической информации. Нельзя утверждать, что там есть всё, но базы данных крупнее, чем у них, попросту не существует.

"Data Mining". StatSoft определяет понятие "Data Mining" как процесс аналитического исследования больших массивов информации (обычно экономического характера) с целью выявления определенных закономерностей и систематических взаимосвязей между переменными, которые затем можно применить к новым совокупностям данных. Этот процесс включает три основных этапа: исследование, построение модели или структуры и ее проверку. В идеальном случае, при достаточном количестве данных можно организовать итеративную процедуру для построения устойчивой (робастной) модели. В то же время, в реальной ситуации практически невозможно проверить экономическую модель на стадии анализа и поэтому начальные результаты имеют характер эвристик, которые можно использовать в процессе принятия решения (например, "Имеющиеся данные свидетельствуют о том, что у женщин частота приема снотворных средств увеличивается с возрастом быстрее, чем у мужчин.").



Методы Data Mining приобретают все большую популярность в качестве инструмента для анализа экономической информации, особенно в тех случаях, когда предполагается, что из имеющихся данных можно будет извлечь знания для принятия решений в условиях неопределенности. Хотя в последнее время возрос интерес к разработке новых методов анализа данных, специально предназначенных для сферы бизнеса (например, Деревья классификации), в целом системы Data Mining по-прежнему основываются на классических принципах разведочного анализа данных (РАД) и построения моделей и используют те же подходы и методы.

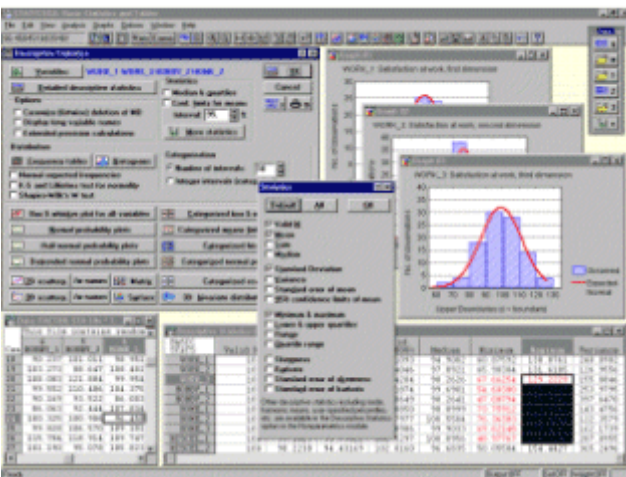
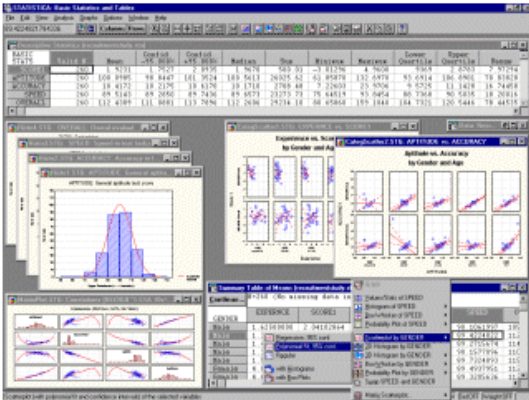
Имеется, однако, важное отличие процедуры Data Mining от классического разведочного анализа данных (РАД) : системы Data Mining в большей степени ориентированы на практическое приложение полученных результатов, чем на выяснение природы явления. Иными словами, при Data Mining нас не очень интересует конкретный вид зависимостей между переменными задачи. Выяснение природы участвующих здесь функций или конкретной формы интерактивных многомерных зависимостей между переменными не является главной целью этой процедуры.

Основное внимание уделяется поиску решений, на основе которых можно было бы строить достоверные прогнозы. Таким образом, в области Data Mining принят такой подход к анализу данных и извлечению знаний, который иногда характеризуют словами "черный ящик". При этом используются не только классические приемы разведочного анализа данных, но и такие методы, как нейронные сети, которые позволяют строить достоверные прогнозы, не уточняя конкретный вид тех зависимостей, на которых такой прогноз основан.

Data Mining часто рассматривается как естественное развитие концепции хранилищ данных.

Система STATISTICA предлагает пользователю широкий выбор методов разведочного анализа данных. Программа вычисляет практически все используемые описательные статистики общего характера: медиану, моду, квартили, заданные пользователем процентиля, среднее значение и стандартное отклонение, квартильный размах, доверительные интервалы для среднего, асимметрию и эксцесс (и их стандартные ошибки), гармоническое и геометрическое среднее, а также многие специальные описательные статистики. Как и во всех других модулях системы STATISTICA, проведение разведочного анализа данных поддерживают разнообразные графики и диаграммы, в т.ч. различные виды диаграмм размаха и гистограмм, гистограммы двумерных распределений (трехмерные и категоризованные), двух- и трехмерные диаграммы рассеяния с помеченными подмножествами данных, нормальные и полунормальные вероятностные графики и графики с исключенным трендом, К-К и В-В графики и т.д. Имеется набор критериев для подгонки нормального распределения к данным (критерии Колмогорова-Смирнова, Лиллиефорса и Шапиро-Уилкса).

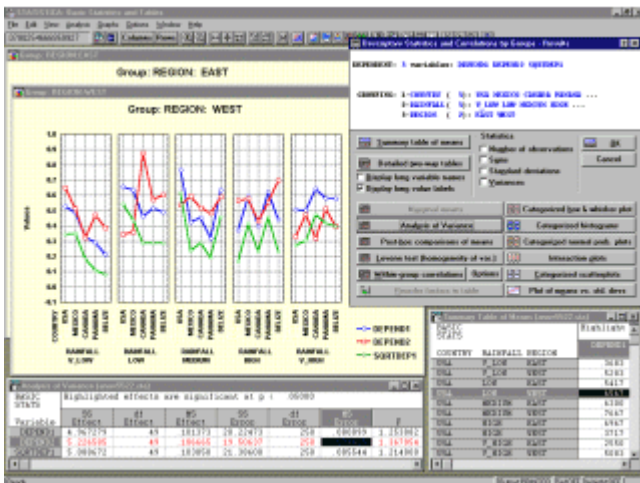
Процедуры для подгонки многих других типов распределений можно найти также в описании анализа процессов и графических возможностей системы. Практически все описательные статистики и графики могут быть построены для данных, категоризованных (сгруппированных) по значениям одной или нескольких группирующих переменных. Например, с помощью нескольких щелчков



мыши можно сгруппировать имеющиеся данные о людях по полу и возрасту и затем просмотреть категоризованные гистограммы, диаграммы размаха, графики на нормальной вероятностной бумаге, диаграммы рассеяния и т.д. В случае, если было выбрано более двух категоризирующих переменных, автоматически будет построен каскад соответствующих графиков.

Имеется возможность производить категоризацию по числовым (непрерывным) переменным, например, можно потребовать, чтобы значения переменной были разбиты на заданное число интервалов; с помощью средства перекодировки в реальном времени можно задать конкретный специальный способ перекодировки переменной (возможности практически сколь угодно сложной перекодировки доступны в любой момент, причем перекодировка может быть задана через соотношения между любыми переменными файла данных). В дополнение к этому в системе имеется специализированная процедура иерархической группировки, позволяющая осуществлять категоризацию данных по многим (до шести) переменным и строить различные категоризованные графики, описательные статистики и корреляционные матрицы для подгрупп (пользователь может в интерактивном режиме сделать неучитываемыми некоторые из факторов в полной таблице группировок и изучать статистики для маргинальных таблиц).

Многочисленные возможности форматирования и расстановки меток позволяют получать таблицы и отчеты презентационного качества, содержащие длинные метки и



описания переменных. При этом важно отметить, что процедура группировки выполняется для чрезвычайно больших объемов данных (например, по одной категоризирующей переменной можно построить до 300 групп), а ее результаты содержат все соответствующие статистики дисперсионного

анализа (включая полные таблицы ANOVA, критерии проверки гипотез типа критерия Левена однородности дисперсии, семь различных апостериорных (post-hoc) критериев и т.д.).

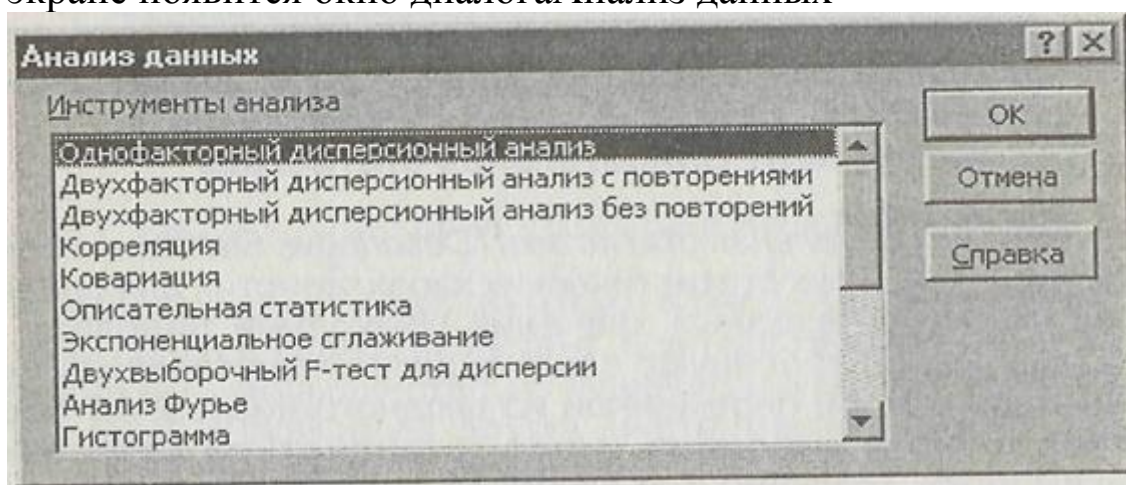
Как и во всех других модулях системы STATISTICA, для достижения высокой - не имеющей аналогов в сравнении с другими пакетами - точности результатов здесь можно производить вычисления с повышенной точностью (если нужно - с четырехкратной). Благодаря интерактивному характеру системы изучение данных становится очень простым делом. Например, графики и диаграммы для разведочных статистик можно получать непосредственно из данных любых выходных таблиц, просто указав мышью на отдельные ячейки или группы ячеек. Одним щелчком мыши можно получать каскады графиков (в том числе сложных, например, со множественными категориями), которые затем можно просматривать в режиме подобном демонстрации слайдов, просто нажимая кнопку **Продолжить**. В дополнение к большому числу готовых статистических графиков пользователь может самостоятельно задавать различные типы визуализации исходных данных, описательных статистик, взаимосвязей между статистиками, группировок и категоризаций с помощью средств прямого доступа (point-and-click), что позволяет существенно уменьшить требуемое количество действий мышью. Средства графического разведочного анализа объединены с собственно статистическими процедурами, что существенно облегчает визуальный анализ данных (например, в интерактивном режиме можно удалять выбросы, выделять подмножества данных, осуществлять сглаживание и подгонку функций, а богатые средства работы с кистью позволяют легко выявлять и/или выделять нужные данные).

Разведочный статистический анализ данных средствами Excel. Microsoft Excel предлагает средства для анализа статистических данных. Такие встроенные функции, как СРЗНАЧ (AVERAGE), МЕДИАНА (MEDIAN) и МОДА (MODE), могут использоваться для проведения анализа данных. Если встроенных статистических функций недостаточно, необходимо обратиться к пакету Анализ данных.

Пакет *Анализ данных*, являющийся надстройкой, содержит коллекцию функций и инструментов, расширяющих встроенные аналитические возможности Excel. В частности, пакет Анализ данных можно использовать для создания гистограмм, ранжирования данных, извлечения случайных или периодических выборок из набора данных, проведения регрессионного анализа, получения основных статистических характеристик выборки, генерации случайных чисел с различным распределением, а также для обработки данных с помощью преобразования Фурье и других преобразований.

Пакет Анализ данных доступен при каждом запуске Excel. Функции пакета Анализ данных можно использовать точно так же, как и любые другие функции Excel, а чтобы получить к ним доступ, выполните описанные ниже действия:

1. Выберите в меню Сервис команду Анализ данных. При первом выборе этой команды Excel загружает файл с диска. Затем на экране появится окно диалога Анализ данных



2. Чтобы использовать какой-либо из инструментов анализа, выберите его имя в списке и нажмите кнопку ОК.

3. Заполните открывшееся окно диалога. В большинстве случаев это означает задание входного диапазона с данными, которые вы собираетесь анализировать, задание выходного диапазона, куда должны быть помещены результаты, и выбор нужных параметров. Если команда *Анализ данных* отсутствует в меню *Сервис* или формула, содержащая функцию из пакета анализа, возвращает ошибочное значение `#ИМЯ? (#NAME?)`, выберите в меню Сервис команду *Надстройки*, затем *Пакет анализа* в списке

надстроек, после чего нажмите кнопку ОК. Если Пакет анализа отсутствует в списке надстроек, вы должны установить его, запустив программу Setup.

При анализе данных часто возникает необходимость определения различных статистических характеристик или параметров распределения.

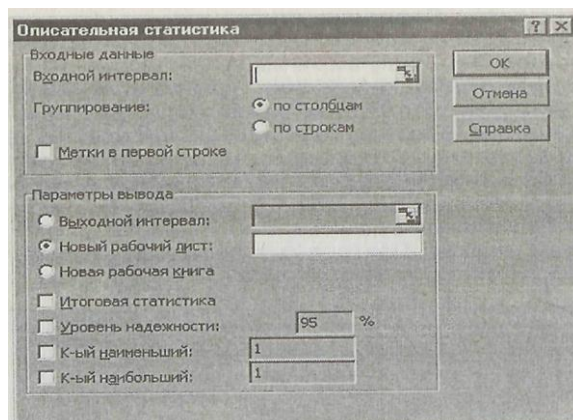
С помощью Microsoft Excel можно анализировать распределение, используя несколько инструментов:

- встроенные статистические функции,
- функции для оценки разброса данных,
- инструмент *Описательная статистика* (Descriptive Statistics), который предоставляет удобные сводные таблицы основных параметров распределения,
- инструменты *Гистограмма* (Histogram),
- *Ранг и перцентиль* (Rank and Percentile).

Встроенные статистические функции Microsoft Excel применяются при проведении статистического анализа данных. Кроме них Excel также предлагает более сложные функции *ЛИНЕЙН* (LINEST), *ЛГРФПРИБЛ* (LOGEST), *ТЕНДЕНЦИЯ* (TREND) и *РОСТ* (GROWTH), которые работают с числовыми массивами.

Описательная статистика (Descriptive Statistics) позволяет создать таблицу основных статистических характеристик для одного или нескольких множеств входных значений. Выходной диапазон содержит таблицу со статистическими характеристиками для каждой переменной входного диапазона: среднее, стандартная ошибка, медиана, мода, стандартное отклонение и дисперсия выборки, коэффициент эксцесса, коэффициент асимметрии, размах, минимальное значение, максимальное значение, сумма, количество значений, k-е наибольшее и наименьшее значения (для любого заданного k) и доверительный интервал для среднего.

Для использования режима *Описательная статистика* в меню *Сервис* выберите команду *Анализ данных*, затем в списке *Инструменты анализа* окна диалога *Анализ данных* выберите инструмент *Описательная статистика* и нажмите кнопку *ОК*. Появится окно диалога,



Инструмент *Описательная статистика* требует задания входного диапазона, который может содержать одну или несколько переменных, и выходного диапазона. Необходимо указать, как расположены переменные в столбцах или в строках. Установите флажок *Метки* в первой строке, если первая строка во входном диапазоне содержит названия столбцов. Excel использует эти метки для создания заголовков в выходной таблице.

Чтобы получить представленную выше таблицу статистических характеристик, установите флажки в области *Параметры вывода*.

Подобно другим инструментам пакета анализа, *Описательная статистика* создает *таблицу констант*. Если эта таблица не устраивает исследователя, можно получить большинство из перечисленных ниже статистических характеристик с помощью других инструментов пакета анализа или формул с использованием встроенных функций Excel.

Анализ данных можно осуществить и через «*Диспетчер сценариев*». Чтобы провести анализ данных в Excel 2010 необходимо активировать «Диспетчер сценариев». Это идет по схеме: вкладка «Данные», активируем кнопочку «Работа с данными» - «Анализ «Что если»» - «Сценарии». После увидите окошко, в котором необходимо задействовать команду «Добавить» и появится следующее окошко «Создание сценария».

В случае если у вас уже есть готовые сценарии, то их можно изменить. Здесь необходимо указать имя сценария. Называйте его так, чтобы можно было быстро определить, для чего он применяется, скажем, через несколько месяцев. Поле «Изменяемые

ячейки» сообщает сценарию, откуда необходимо брать исходные данные, так что указывайте адреса ячеек, опираясь на собственные нужды. Они могут не быть смежными и тогда их адреса указываются через запятую (не более 32 изменяемых ячеек на сценарий). Примечание самостоятельно заполняется, получая сведения об авторе и время создания сценария. Его можно дополнять по необходимости. После указания всех параметров сохраните изменения (кнопка «Ок»). В результате появится окошко «Значение ячеек», где будут отображены все внесенные изменения. После заполнения данной формы также сохраните изменения. Вы автоматически вернетесь к «Диспетчеру». Теперь осталось только воспользоваться готовым сценарием. Чтобы вывести результаты расчета сценария, необходимо указать его название в окне «Диспетчера» и активировать функцию «Вывести». Изменения вносятся в сценарий с помощью функции «Изменить». Также можно просмотреть отчет, активировав одноименную функцию. В новом окошке укажите его тип (это может быть сводная таблица или структурированный список) и ячейки результата (в них будут находиться формулы, результаты которых и нужно подвергнуть анализу).

Отчет представляет все изменяемые значения для любого сценария («Изменяемые») и значения формул, которые были вычислены при помощи этих значений («Результат»). Поскольку мы сначала присвоили имена для всех изменяемых ячеек и ячеек, в которых будет сохранен результат, то и при создании самого сценария идет вывод не адресов этих секторов, а их имена. Сам же отчет, в результате, выглядит максимально понятно. Это дает возможность проанализировать самое разное количество возможных вариантов. Разные пользователи могут хранить необходимые для анализа сведения в совершенно отдельных файлах-«книгах», но их можно собрать вместе и объединить в один сценарий, что еще больше ускорит работу. В результате можно легко получить необходимый отчет, опирающийся даже на информацию полученную таким путем.

Визуальный анализ. Такой анализ данных в Excel2010 хорошо подходит для создания отчетов, что позволяет эффективно систематизировать, скажем, данные относительно совершенно

любой деятельности за определенный временной отрезок. Допустим, у вас уже есть необходимые данные, но их необходимо подготовить для анализа.

Заходим на вкладку «Вставка» и приступаем к созданию сводной таблицы. Новый лист предложит макет стандартной такой таблицы. Все параметры, фигурировавшие в начальной таблице, будут перечислены справа. С помощью мыши перетаскиваем их в графу «Название строк». Делаем это с «Датами» и «Менеджерами», но у вас, скорее всего, будут иные наименования. В графу «Значения» помещаем «Объемы продаж», «Выручку» и «Прибыль». После этого таблица самостоятельно отформатируется и станет «лентой». Кстати, размещение элементов в «Названии строк» играет существенную роль. В случае, если «Менеджеры» выше «Дат», то все данные будут разбиваться соответственно имен этих работников. Если выше «Даты» - то соответственно календарных дат.

Теперь необходимо провести оформление созданной таблицы. Форматируем ее, как таблицу (вкладка «Главная» - «Форматировать как таблицу»). Появится список самых разных шаблонов – необходимо выбрать тот, который более удобен именно для вас. После этого Excel самостоятельно определит границы, но их всегда можно отрегулировать и вручную.

Сохраняем параметры (кнопка «Ок») и смотрим на то, что получилось. Теперь уже можно сортировать параметры. Однако визуально просматривать значения и сравнивать с плановыми может оказаться весьма сложным делом. Особенно, если таких значений очень много. Допустим, что месячная выручка каждого нашего менеджера не должна быть меньше 100 000. Просматривать все самостоятельно не нужно – это займет слишком много времени и сил. Поэтому просто проводим условное форматирование (вкладка «Вставка» - «Условное форматирование» - «Набор значков») по понравившемуся шаблону. Например, «Светофор». Создаем правила форматирования (вводим показатели напротив значков), что позволит автоматически оценить работу сотрудника, как отличную, стабильную и неудовлетворительную. Показатели вводим напротив каждого кружочка в «Значение», в «Тип» устанавливаем «Числа», а не «Процент». Мы установили такой

показатель, как 100 000 и 90 000, а третий выставиться автоматически, чтобы подключить оставшиеся значения. Сохраняем. Согласитесь, что теперь можно гораздо быстрее проанализировать данные и определить, кто из менеджеров хорошо справляется со своей задачей, а кого можно уже и увольнять.

Однако это только самые простые из возможностей современного Excel2010. В нем появились дополнительные элементы, называемые «Цветовыми шкалами» и «Гистограммами». Давайте попробуем использовать именно их. Итак, выделяем значения в ячейках и форматируем их (вкладка «Вставка» - «Условное форматирование» - «Гистограммы»). Выпадающее меню демонстрирует список шаблонов (доступен предварительный просмотр при наведении курсора мыши на наименование).

Выбираем удобную цветовую схему. В результате получаем ячейки, залитые горизонтальными столбцами разной величины. Они в графическом плане отображают присутствующее в ячейке значение. Теперь можно уже даже скользнув взглядом по таблице понять, насколько плохо выполняет свои обязанности наш «Менеджер 5». Примечательно, если значение уйдет ниже минуса, то график сместится в противоположную от ячейки сторону, явно намекая на отрицательный результат. При использовании компонента «Цветовые шкалы» происходит заливка ячейки соответствующим цветом, который полностью соответствует результату. В результате наименьшее значение получит красный цвет, среднее – желтый, а высокое – зеленый.

Естественно, такую схему можно подобрать самостоятельно. Это более наглядный пример, чем применение «Набора значков», однако суть у них одна. Впрочем, есть еще один способ, который основан на применении срезов. Допустим, наши менеджеры проработали в компании уже не один год. Естественно, что дат станет гораздо больше и просмотреть такой документ, даже с применением форматирования, будет крайне сложно. Не говоря уже об анализе данных. В этом случае добраться до интересующей нас даты можно одним из двух способов.

При построении любой сводной таблицы в правой части располагаются элементы, которые мы можем расположить в любые

удобные поля. Если обратится к «Датам» и нажать на кнопочку с изображением стрелочки, то вызовется выпадающее меню. Остается только отфильтровать информацию по дате. Появляется большой список, предлагающий всевозможные варианты форматирования.

Воспользуемся помесечной сортировкой. Для этого необходимо открыть «Все даты за период» и выбрать нужный месяц. Для нас это «Октябрь». Это позволит значительно сократить нашу таблицу, оставив только интересующие нас значения.

Теперь давайте рассмотрим использование среза. Данный инструмент анализа великолепно подходит для цифровых данных. Открываем вкладку «Вставка» и ищем там «Срез». В результате открывается меню «Вставка среза», где необходимо отметить показатель, согласно которому и будет идти выборка интересующих нас значений. В данном случае это будет колонка «Даты», хотя можно выбрать любую из интересующих.

Подтверждаем и смотрим на страничку с рамкой и имеющимися значениями. Осталось только переместить ее в любое удобное место и отрегулировать размер, чтобы было видно все значения. При необходимости можно поменять цвет среза, чтобы сделать его более понятным (шаблоны находятся на верхней панели).

Теперь можно одним кликом выбрать необходимую дату и посмотреть на результаты работы наших менеджеров. Естественно, что благодаря своей гибкости использование среза более удобно, чем применение фильтра по дате. К тому же можно указать несколько значений для выборки.

В Excel2010 можно применять инфокривые графики. Для этого необходимо выделить ячейку напротив строки с данными и сделать ее активной. Далее вновь используем вкладку «Вставка», раздел «Инфокривые» (он еще может носить название «Сперклайны»). Выделяем нашу строку, как диапазон данных и подтверждаем. В результате в выбранной ячейке появится небольшой график. Сделаем это же для результатов всех наших сотрудников. Теперь растянем каждую ячейку на другие строчки. Для этого можно просто потянуть за край, отмеченный точкой или же сделать на нем двойной клик. Как и при работе с иными

полезными функциями, можно самостоятельно подбирать стиль инфокривой (верхняя панель, режим «конструктор»). Такой график хорошо отражает тренд и, если значений достаточно много, то визуально легко определяются взлеты и падения, а также начало возможного систематического падения или роста.

Сама инфокривая может быть одного из 3 типов: график, столбец – он отображает обрабатываемые данные маленькими столбиками. Чем больше данных, тем тоньше будут столбики, но они способны наглядно продемонстрировать минимальное и максимальное значение; - «Выигрыш / Проигрыш» - ячейка условно делится на две части. При положительном результате квадратики помещаются в верхнюю часть, а при отрицательном – в нижнюю. «Ноль» в этом случае вообще не отображается.

Порядок выполнения работы

1. Изучите теоретический материал.
2. Составьте множество исходных данных для двух классов заболеваний по материалам Приложения к лабораторной работе №2 (классы выбираются по вариантам 1-2, 2-3, 1-2 согласно порядковому номеру в журнале).
3. Для каждого класса составьте выполните описательную статистику (программа STATISTICA или «Пакет анализа»). Сделайте выводы по отличительным особенностям для каждого класса.
4. Постройте гистограммы распределения по каждому признаку по разным классам на одном графике. Сделайте вывод о диагностических возможностях (специфичности) признаков.
5. В каждом классе постройте граф связности между признаками согласно парным коэффициентам корреляции (положительную корреляцию отобразите сплошной линией, отрицательную – пунктиром, значимость коэффициента корреляции (модульную величину) отобразите цветом). Сделайте выводы по результатам анализа различий между графами.
6. Оформите отчет, включающий результаты выполнения работы, ответы на контрольные вопросы, аннотацию каждого из приведенных в библиографии информационных источников (100-350 слов).

Контрольные вопросы

1. Что показывает гистограмма?
2. Что такое нормальное распределение случайной величины?
3. Какие диаграммы целесообразно использовать только для альтернативных вопросов?
4. Как называется разница между наибольшим и наименьшим значениями?
5. Какая статистика показывает, в какую сторону относительно среднего сдвинуто большинство значений распределения?
6. Кем была создана программа SPSS? Для чего она предназначена?
7. Альтернативный вопрос предполагает: А) выбор респондентом одного варианта ответа Б) выбор респондентом нескольких вариантов ответа В) предложение респондентом своего варианта ответа?
8. Какое действие позволяет управлять расположением данных внутри ячейки: А) выравнивание Б) шкалирование В) расширение?
9. Для чего нужен контроль данных: А) для исправления ошибок Б) для добавления новых переменных В) чтобы избежать повторений?
10. Какая процедура позволяет строить статистические ряды распределений: А) «частоты» Б) «статистики» В) «анализ» ?
11. Какой из типов переменной чаще всего используется как основание для деления объектов на группы: А) количественный Б) порядковый В) категориальный?
12. Для чего выполняют агрегирование данных?
13. Какой из типов столбчатых диаграмм используется для визуализации таблиц сопряженности: А) простые Б) составные В) кластерные ?
14. Что является первым этапом статистического анализа данных: А) описательный анализ Б) сравнительный анализ В) частотный анализ Г) факторный анализ?
15. Какие типы диаграмм целесообразно использовать для визуализации перекрестных таблиц: А) простые и кластерные Б) кластерные и состыкованные В) простые и состыкованные?
16. Приведите классическое определение разведочного анализа.

Библиография

1. Васильев, Алексей Николаевич .Числовые расчёты в Excel [Текст] : учебное пособие / А. Н. Васильев. - Санкт-Петербург: Лань, 2014. - 608 с
2. Доказательная медицина //URL: <http://encyclopatia.ru/>
3. Доказательная медицина для всех//URL: <http://medspecial.ru/>
4. Доказательная медицина. Презентация //URL: http://www.kurskmed.com/uvr_docmed/uploads/9b3c40a.pdf
5. Пакет Анализа в Excel //URL: http://studopedia.su/12_127574_analiz-dannih-v-Excel.html
6. Разведочный анализ в СТАТИСТИКА //URL: http://www.exponenta.ru/SOFT/STATIST/statistica5_5/1/1.asp
7. Разведочный анализ. Лекция //URL: <http://www.slideshare.net/DEVTYPE/ss-53792094>
8. Разведочный анализ в Excel //URL: <http://advanceduser.ru/microsoft-excel/provodim-analiz-dannykh-v-excel-2010.html>

Приложение к лабораторной работе №2

эр	Нб	ЦП	лейк	эоз	пал	сегм	лимф	мон	заболевание
4,6	142	0,93	7,05	1	14	45	38	2	рак желудка
4,2	115	0,82	7,8	2	4	81	10	2	рак желудка
3,4	107	0,94	8	1	6	64	27	2	рак желудка
3	87	0,87	7,3	4	1	55	36	4	рак желудка
3,4	100	0,88	6,15	0	4	77	17	2	рак желудка
5,0	170	1,02	7,1	2	5	59	30	4	рак желудка
5	150	0,90	3,7	6	1	71	21	1	рак желудка
3,6	105	0,88	17,6	0	7	64	26	2	рак желудка
5,3	108	0,61	4,2	4	1	47	46	2	рак желудка
4,16	132	0,95	5,1	2	4	43	47	4	рак желудка
4,3	145	1,01	4,55	4	5	52	36	3	рак желудка
4,0	128	0,96	7,6	4	2	57	29	7	рак желудка
5,0	160	0,96	6,8	0	4	72	23	1	рак желудка
4,46	146	0,98	6,5	4	1	66	27	2	рак желудка
5,4	176	0,98	7,8	2	2	67	29	0	рак желудка
3,1	108	1,05	3,7	1	2	62	31	4	рак желудка
4,0	124	0,93	4,4	2	7	66	22	3	рак желудка
5,1	164	0,96	7,9	0	5	29	60	6	рак желудка
4,25	138	0,97	6,2	2	7	59	28	4	рак желудка
4,8	160	1,00	6,7	2	2	70	25	1	рак желудка
4,3	150	1,05	12	1	1	59	30	9	рак желудка
4,0	113	0,85	5,85	6	4	48	37	5	рак желудка
4,8	153	0,96	7,1	6	1	61	30	2	рак желудка
3,6	114	0,95	13,1	0	2	73	24	1	рак желудка
3,35	103	0,92	9,1	1	3	58	27	17	рак желудка
4,5	150	1,00	8	4	1	56	36	3	рак желудка
эр	Нб	ЦП	лейк	эоз	пал	сегм	лимф	мон	
4,6	130	0,85	6,7	1	3	44	38	14	Цирроз
3,7	117	0,95	3,6	5	3	50	32	10	цирроз
3,3	110	1,00	4,4	1	3	70	23	3	цирроз
4,4	140	0,95	13,2	3	2	63	30	2	цирроз
4,5	150	1,00	7	1	6	50	34	9	цирроз
4,4	132	0,90	4,4	0	5	60	33	2	цирроз
4,1	136	1,00	5	1	1	70	24	4	цирроз
4	130	0,98	4,9	3	2	71	20	4	цирроз
3,5	115	0,99	7,8	1	6	51	38	4	цирроз
4,2	132	0,94	6	3	4	67	20	6	цирроз
4,3	140	0,98	4,3	2	2	71	17	8	цирроз
4	118	0,89	8,3	0	5	61	26	8	цирроз
4,5	144	0,96	6,4	2	1	60	33	4	цирроз
4,60	150	0,98	5,6	7	4	46	36	7	цирроз
4	130	0,98	7,0	5	6	54	26	9	цирроз

3,9	130	1,00	9,6	1	19	48	15	5	цирроз
4,0	112	0,84	15,8	4	18	61	13	4	цирроз
4,6	150	0,98	5,4	1	2	52	37	8	цирроз
4,6	150	0,98	5,4	2	5	51	34	8	цирроз
4,6	145	0,95	5,2	1	8	68	20	3	цирроз
4,6	150	0,98	6,6	2	3	64	42	12	цирроз
3,2	108	1,01	3	2	5	48	49	6	цирроз
4,5	141	0,94	8,4	2	4	58	24	12	цирроз
4,2	140	1,00	7,1	1	3	60	33	3	цирроз
4	130	0,98	7,0	5	6	54	26	9	цирроз
3,7	112	0,91	3,9	2	1	61	32	4	цирроз
4,0	130	0,98	4	3	1	57	35	4	цирроз
3,7	114	0,92	3,8	3	2	50	41	4	цирроз
4,4	144	0,98	7	2	12	48	26	12	цирроз
эр	Нв	ЦП	лейк	эоз	пал	сегм	лимф	мон	
3,9	109	0,84	4,6	3	3	67	24	3	желчКамБол
4,0	132	0,99	7	3	5	84	10	1	желчКамБол
4,4	144	0,98	9,1	4	7	72	20	1	желчКамБол
3,6	126	1,05	4,4	2	8	54	37	1	желчКамБол
3,5	106	0,91	11,9	1	4	73	18	5	желчКамБол
3,0	100	1,00	6,0	3	1	60	30	6	желчКамБол
3,9	139	1,07	6,4	2	4	59	34	1	желчКамБол
3,8	127	1,00	5,9	3	4	40	50	6	желчКамБол
3,1	103	1,00	10,2	4	1	68	17	10	желчКамБол
3,8	120	0,95	6,8	1	1	58	37	3	желчКамБол
4,1	134	0,98	8,9	2	5	58	36	1	желчКамБол
4,0	120	0,90	7,2	3	1	78	19	2	желчКамБол
3,5	126	1,08	5,3	1	1	51	46	2	желчКамБол
3,5	114	0,98	5,2	3	3	59	33	2	желчКамБол
3,9	126	0,97	9,3	1	11	66	20	2	желчКамБол
4,35	138	0,95	19,0	2	8	86	3	1	желчКамБол
3,8	124	0,98	7,2	2	7	69	27	2	желчКамБол
3,8	128	1,01	10,2	1	8	77	14	1	желчКамБол
3,5	110	0,94	9,4	3	6	68	24	2	желчКамБол
3,5	114	0,98	14,9	2	3	79	15	3	желчКамБол
3,9	126	0,97	9,9	4	5	76	18	1	желчКамБол
3,9	120	0,92	17,9	3	5	67	24	4	желчКамБол
4,5	140	0,93	12,7	1	8	87	3	1	желчКамБол
3,3	110	1,00	3,2	2	1	66	27	4	желчКамБол
4,2	140	1,00	9,6	1	4	67	25	3	желчКамБол
4,1	130	0,95	7,8	3	10	58	26	3	желчКамБол

ЛАБОРАТОРНАЯ РАБОТА № 3. РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ АНАЛИЗЫ СЕРДЕЧНОГО РИТМА

Цель работы: овладение навыками использования для исследований в структурных связях между регистрируемыми характеристиками состояния биообъекта корреляционного и регрессионного анализов на примере сердечного ритма.

Краткие теоретические сведения

Как правило, медико-биологических исследований представляют собой либо вектор, характеризующий изменения биологических сигналов во времени – ЭКГ, реограммы и т.п. (временной ряд) либо их совокупность либо вектор, характеризующий значения различных характеристик зарегистрированных практически одномоментно.

При моделировании временных рядов модели в виде эмпирических формул, полиномов или рядов Фурье являются функциональными моделями, так как представляют собой аналитическое выражение или функцию, которая отражает функциональную связь между зависимым признаком (амплитудой сигнала или артериальным давлением) и независимым признаком – временем.

Отличительной чертой биосистем является многообразие признаков, характеризующих эту биосистему (возраст, пол, рост, вес, пульс, давление и так далее). Часто между вариациями этих признаков существует связь, (например: обычно чем больше рост, тем больше вес, но могут быть и исключения).

Эта связь не является строго функциональной, так как возможны отклонения от общей тенденции, она называется корреляционной (от английского correlation – соотношение, взаимосвязь, взаимодействие). Корреляция между двумя признаками (с точки зрения статистических догматов) состоит в том, что среднее значение одного признака (результативного или зависимого) изменяется в зависимости от изменения среднего значения другого признака (факторного или независимого). Поэтому, если нужно построить модель для связи средних значений каких-либо признаков (например, среднегодовым числом сердечно-сосудистых заболеваний и среднегодовым значением

метео-осадков), то для этого используют методы корреляционного анализа. Предметом корреляционного анализа является выявление корреляционной связи между признаками.

Корреляционный анализ включает в себя четыре этапа:

1. выявление наличия корреляционной связи,
2. определение формы корреляционной связи,
3. вычисление тесноты корреляционной связи,
4. оценка статистической достоверности результатов п.3.

Пусть биосистемой является пациент. У него измерено два признака: вес (Р) и артериальное давление (АД). Данные измерений занесены в таблицу. Цель исследования – найти корреляционную связь между этими признаками.

Применим поэтапно корреляционный анализ.

Этап 1. Выявление наличия корреляционной связи. Этот этап проводится при помощи построения корреляционного поля или диаграммы рассеивания. Чтобы построить корреляционное поле, нужно в координатной плоскости признаков нанести все объекты. Каждый объект будет изображен в виде точки с координатами, соответствующими значениям признаков. Множество таких точек и будет *корреляционным полем* или диаграммой рассеивания.

Для построения диаграммы рассеивания необходимо использовать в электронной таблице режим «Диаграмма точечная». Чтобы выявить наличие или отсутствие корреляционной связи, нужно проанализировать форму корреляционного поля (диаграммы рассеивания). При наличии корреляционной связи форма корреляционного поля близка к эллипсоидной. Оси эллипса расположены по диагонали по отношению к осям координат в координатной плоскости. Главная ось эллипса наклонена к оси абсцисс по острым углом. Если наклон «направо», то корреляция положительная, «налево» - отрицательная.

Этап 2. Определение формы корреляционной связи. Если ранжировать выборку из двух признаков по возрастанию одного из них и отобразить график зависимости между признаками, то, наблюдая за полученной кривой можно предположить вид корреляционной функции между признаками. Вид этой функции называется линией регрессии. (Заметим, что если каждому значению одного признака соответствует только одно значение

другого – то связь называется функциональной). Полученную линию регрессии можно промоделировать аналитической функцией, уравнение которой называется уравнением регрессии. Оно определяет форму корреляционной связи. Если линия регрессии моделируется прямой, то корреляционная связь между признаками называется линейной регрессией. Это частный, но весьма распространённый случай. Для выявления уравнения регрессии в Excel на графике, отражающем связь между характеристиками необходимо построить линию тренда, отобразить на графике коэффициент детерминации (квадрат парной корреляции в данном случае) и график зависимости, выбрать регрессионное уравнение соответствующее максимальному значению коэффициента детерминации.

Этап 3. Вычисление тесноты корреляционной связи. Тесноту или степень связи можно вычислить при помощи коэффициентов корреляции. В случае линейной регрессии между признаками x и y тесноту связи характеризует линейный коэффициент парной корреляции $r_{y/x}$. Величина $r_{y/x}$ изменяется в пределах от -1 до 1 ($-1 < r_{y/x} < 1$). Если $r_{y/x} > 0$, то корреляционная связь между признаками прямая, причём, чем ближе к 1 , тем связь теснее, тем больше она приближается к строгой (функциональной) линейной связи. Если $r_{y/x} < 0$, то связь обратная, причём, чем ближе $r_{y/x}$ к -1 , тем теснее обратная связь. Если $r_{y/x} = 1$ (или $r_{y/x} = -1$), то связь между признакам - строгая функциональная (прямая или обратная).

Оценка статистической значимости полученного значения коэффициента парной корреляции (с применением соответствующей функции в Excel) осуществляется, например, путем вычисления коэффициента Стьюдента и оценки его статистической значимости путем вычисления значения ошибки первого рода и сравнения с пороговым значением. Коэффициент

Стьюдента вычисляется по формуле:
$$t = \frac{ry/x^2}{2\sqrt{1-|ry/x|}} \sqrt{n-1}$$
. Где n – количество значений пар x, y в выборке.

Изменение регистрируемой характеристики во времени представляется в виде одномерного массива, называемого «временным рядом».

Временной ряд является нестационарным, если он содержит такие систематические составляющие как тренд и цикличность. Нестационарные временные ряды характеризуются тем, что значения каждого последующего уровня временного ряда корреляционно зависят от предыдущих значений.

Для исследования временного ряда на первоначальном этапе осуществляется построение и изучение автокорреляционной функции. Автокорреляцией уровней временного ряда называется корреляционная зависимость между настоящими и прошлыми значениями уровней данного ряда.

Лагом называется величина сдвига между рядами наблюдений. Лаг временного ряда определяет порядок коэффициента автокорреляции. Например, если уровни временного ряда x_t и x_{t-1} корреляционно зависимы, то величина временного лага равна единице. Следовательно, данная корреляционная зависимость определяется коэффициентом автокорреляции первого порядка между рядами наблюдений $x_1 \dots x_{n-1}$ и $x_2 \dots x_n$. Если лаг между рядами наблюдений равен двум, то данная корреляционная зависимость определяется коэффициентом автокорреляции второго порядка и т. д.

При увеличении величины лага на единицу число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается на единицу. Поэтому максимальный порядок коэффициента автокорреляции рекомендуется брать равным $n/4$, где n – количество уровней временного ряда.

Автокорреляция между уровнями временного ряда оценивается с помощью выборочного коэффициента автокорреляции.

Анализ структуры временного ряда с помощью коэффициентов автокорреляции стоит на следующих правилах:

- 1) исследуемый временной ряд содержит только трендовую компоненту, если наибольшим является значение коэффициента автокорреляции первого порядка r_{1-1} ;
- 2) исследуемый временной ряд содержит трендовую компоненту и колебания периодом 1, если наибольшим является коэффициент автокорреляции порядка 1. Эти колебания могут быть как циклическими, так и сезонными;

- 3) если ни один из коэффициентов автокорреляции $r_l(l=1,L)$ не окажется значимым, то делается один из двух возможных выводов:
- а) данный временной ряд не содержит трендовой и циклической компонент, а его колебания вызваны воздействием случайной компоненты, т. е. ряд представляет собой модель случайного тренда;
 - б) данный временной ряд содержит сильную нелинейную тенденцию, для выявления которой необходимо провести его дополнительный анализ.

Графическим способом анализа структуры временного ряда является построение графиков автокорреляционной и частной автокорреляционной функций.

Автокорреляционной функцией называется функция оценки коэффициента автокорреляции в зависимости от величины временного лага между исследуемыми рядами. Графиком автокорреляционной функции является коррелограмма.

Частная автокорреляционная функция отличается от автокорреляционной функции тем, что при её построении устраняется корреляционная зависимость между наблюдениями внутри лагов.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (1)$$

$$\text{где: } \bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}.$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и y_{t-1} .

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по

формуле:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}} \quad (2)$$

где

$$\bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют лагом. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило – максимальный лаг должен быть не больше $n/4$.

Свойства коэффициента автокорреляции.

По коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется коррелограммой. Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда

наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ , то ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты.

Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют аналитическим выравниванием временного ряда.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

линейный тренд	$\hat{y}_t = a + b \cdot t$;
гипербола:	$\hat{y}_t = a + \frac{b}{t}$;
экспоненциальный тренд:	$\hat{y}_t = e^{a+bt}$
(или $\hat{y}_t = a \cdot b^t$);	степенная функция: $\hat{y}_t = a \cdot t^b$;

полиномы различных степеней: $\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_m \cdot t^m$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, \dots, n$, а в качестве зависимой переменной – фактические уровни временного ряда \hat{y}_t . Для нелинейных

трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени.

В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда.

Если временной ряд имеет линейную тенденцию, то его соседние уровни \hat{y}_t и \hat{y}_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким.

Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Порядок выполнения.

1. Изучить теоретический материал.
2. Выделить из исходного массива данных (см. Приложение) два вектора пульсограмм (до и после нагрузки) длиной 64 измерений, начиная с номера, определенного датчиком случайных чисел в интервале от 1 до вашего номера в студенческом журнале.

3. Построить карелограммы и определить функциональные зависимости между значениями ее координат. Оценить адекватность полученных моделей.
4. Построить автокорреляционные функции (порядка 10-15 первых значений) для пульсограмм до и после нагрузки (см. Приложение). Сделать выводы (обратить внимание на первый переход автокорреляционных функций через «0»).
5. Постройте спектры Фурье (модуль, сдвиг фаз, мнимую и действительные части, годограф) до и после нагрузки, сделайте выводы.
6. Оформите отчет, включающий в себя результаты выполнения (скрин-шоты), выводы, ответы на контрольные вопросы (не менее пяти, из них №№12,13,14 – обязательны).

Контрольные вопросы.

1. Чем отличается функциональная и корреляционная связь между признаками?
2. Что такое временной ряд биофизиологического сигнала?
3. Что такое ранжирование выборки?
4. В каком случае регрессия будет линейной?
5. В каком случае линии регрессии совпадают?
6. Что можно сказать о корреляционной связи между признаками X и Y если значение коэффициента корреляции равно 0,3?
7. Что можно сказать о корреляционной связи между признаками, если корреляционное поле имеет форму круга?
8. Какой метод применяется для нахождения коэффициентов уравнения линейной регрессии?
9. Что такое автокорреляционная функция? Как она определяется?
10. Каким образом оценивается значимость коэффициента корреляции?
11. Как строится карелограмма?
12. Каким образом в электронной таблице осуществляется корреляционный анализ?
13. Как применяются результаты корреляционного анализа в медицине?
14. Как применяется автокорреляция в медицине?
15. Что характеризует интервалограмма сердечной деятельности?

16. Каким образом может быть построен физиологический портрет по регистрируемым показателям?
17. Каким образом в контурах биологически обратной связи может быть использован физиологический портрет?
18. В каких случаях для осуществления спектрального анализа биосигнала рекомендуется использование преобразование Фурье?
19. В каких случаях для осуществления спектрального анализа биосигнала рекомендуется использование преобразование Уолша?
20. В каких случаях для осуществления спектрального анализа биосигнала рекомендуется использование самоорганизационных методов типа алгоритмов МГУА?
21. Что характеризует автокорреляционная функция интерваллограммы?
22. Что характеризует индекс Баевского?

Библиография

1. Автокорреляция. Презентация 1. /URL: <http://www.myshared.ru/slide/82380/>
2. Автокорреляция. Презентация 2. /URL: <http://present5.com/prezentaciya-kafedra-avtomatizacii-obrabotki-informacii2/>
3. Автокорреляция уровней временного ряда. /URL: <http://be5.biz/ekonomika/e011u/80.htm>
4. Баевский Р.М. Прогнозирование состояний на грани нормы и патологии. М.: Изд-во: Медиал. , 2012 г.- 295 с.
5. Корреляция. Презентация. /URL: <http://informslice.ru/korrelyacionnye-zavisimosti/>
6. Корреляционные зависимости. Презентация. /URL: <http://www.myshared.ru/slide/84314/>
7. Статистическая обработка данных. Презентация. /URL: <http://informslice.ru/statisticheskaya-obrabotka-dannyx/>

Приложение к лабораторной работе № 3:

Интерваллограмма (пульсограмма)

до нагрузки

0,900,920,960,960,960,920,900,840,880,880,860,900,82
 0,760,740,720,730,820,94 1,000,980,960,940,960,980,98
 0,960,96 1,020,980,900,96 1,000,900,900,980,90 1,04 1,02

0,960,94 1,00 1,00 0,96 1,04 1,02 1,02 1,04 1,02 1,02 1,02 0,96
 0,860,80 0,82 0,94 0,90 0,98 0,94 0,94 0,96 0,96 0,98 1,02 1,00
 0,98 0,92 1,08 0,98 0,92 1,02 1,02 0,96 1,02 1,10 0,98 1,06 1,04

после нагрузки 1

0,64 0,72 0,74 0,76 0,78 0,76 0,80 0,76 0,74 0,72 0,72 0,72 0,76
 0,78 0,76 0,80 0,74 0,78 0,76 0,74 0,74 0,84 0,80 0,80 0,82 0,74
 0,76 0,80 0,78 0,80 0,78 0,74 0,74 0,72 0,70 0,64 0,62 0,64 0,66
 0,74 0,76 0,72 0,78 0,76 0,72 0,74 0,70 0,70 0,74 0,76 0,74 0,80
 0,84 0,76 0,78 0,72 0,72 0,70 0,72 0,72 0,76 0,80 0,74 0,80 0,80
 0,74 0,76 0,74 0,74 0,74 0,76 0,76 0,80 0,82 0,78 0,80 0,80 0,78

после нагрузки 2

0,76 0,78 0,82 0,80 0,80 0,82 0,80 0,78 0,72 0,74 0,76 0,78 0,78
 0,84 0,86 0,82 0,84 0,76 0,72 0,68 0,72 0,74 0,76 0,78 0,76 0,80
 0,82 0,78 0,82 0,82 0,78 0,82 0,76 0,87 0,80 0,76 0,78 0,78 0,74
 0,80 0,78 0,78 0,76 0,78 0,74 0,80 0,78 0,76 0,80 0,78 0,78 0,80
 0,76 0,82 0,78 0,76 0,80 0,78 0,74 0,81 0,86 0,80 0,80 0,82 0,80
 0,82 0,80 0,78 0,82 0,82 0,80 0,82 0,84 0,74 0,68 0,66 0,74 0,84

ЛАБОРАТОРНАЯ РАБОТА №4. РАСЧЕТ ПОКАЗАТЕЛЕЙ СИСТЕМНОЙ ОРГАНИЗАЦИИ БИООБЪЕКТА

Цель работы: Овладение навыками построения и анализа показателей системной организации биообъектов с целью диагностики изменений их состояний в процессе развития или управления: максимального градиента функциональных различий и оценки напряженности функционального состояния.

Краткие теоретические сведения

Состояние биообъекта или системы в процессе ее функционирования меняется – реакция на окружающую среду (внешние воздействия), что характеризуется изменениями синхронности взаимодействия разнообразных функциональных систем. Поскольку данный процесс отслеживается путем регистрации определенных характеристических параметров, то следует предположить, что изменение синхронизации отражается как количественных сдвигом величин данных параметров, так и изменением значений величин, прямо или косвенно отображающих парную или множественную связи между «временными рядами» параметров. В статистике, в качестве таких величин, могут использоваться коэффициент парной корреляции и коэффициент корреляционного отношения, отражающей нелинейную регрессию.

В этом случае, предлагается использовать для количественной оценки происходящих корреляционных сдвигов Максимальный Градиент Функциональных Различий, который вычисляется следующим образом. Он вычисляется как наибольшее значение межранговых отклонений упорядоченных по убыванию связей показателей, выбранных для характеристики ФС (например, значения различных корреляционных коэффициентов). Показатель, имеющий наибольшую связность называется системообразующим для данного класса. В качестве меры связности рассматривается обычно значения коэффициента парной корреляции, превышающие порог определенного уровня статистической значимости и ранжированные определенным образом (например, в алфавите – от 0 до 9).

Например, пусть два некоторых класса имеют структуру представленную на рисунке 1.

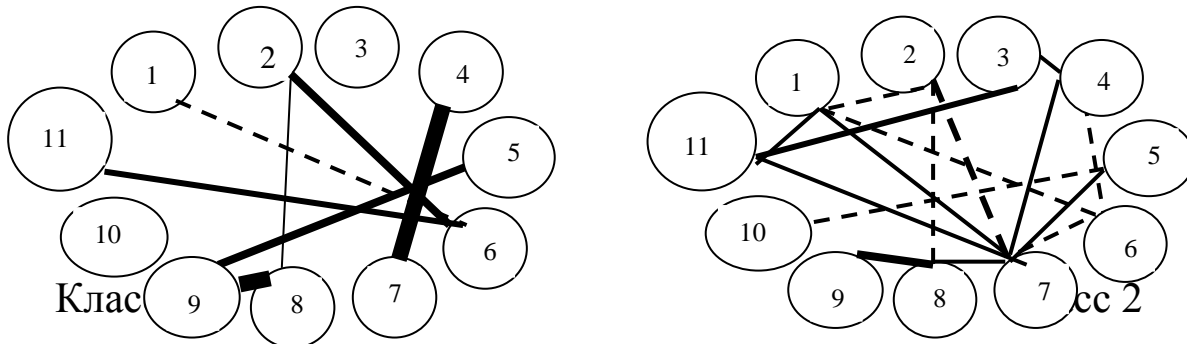


Рис. 1 Корреляционные связи между показателями 1-11 (толщина линии отражает величину коэффициента корреляции, пунктир – обратная связь).

Для рассматриваемого примера: системообразующими показателями являются: в Классе 1 – 6 показатель, в Классе 2 – 7. МГФР определяется следующим образом. Упорядочиваем показатели по мере связности – соответственно получаем: в Классе 1- 6,9,8,7,4,2,5,11,1,(3,10); в Классе 2 – 7,1,11,8,6,2,4,5,9,3,10. Таким образом, смещение показателей (или градиент – в терминологии Завьялова А.В.) составило в абсолютном значении: для 1 показателя –7, для 2 – 0, для 3 – 0, для 4 – 5, для 5 – 1, для 6 – 4, для 7 – 3, для 8 – 1, для 9 – 7, для 10 – 0 и для 11 – 5. Т.о., МГФР Класса 2 относительно Класса 1 равен 7.

Чем больше значение МГФР, тем более произошедшие в процессе терапии изменения в синхронности работы функциональных систем организма.

Классификация системной организации функций различных состояний биообъектов или систем в процессе их эволюции и-или в процессе управления (например, терапевтического воздействия) характеризуется некоторым отклонением от «базового состояния» - точки отсчета - представлено в Таблице 1.

Таблица 1

№ уровня состоя ния	Характеристика уровня состояния	Работа организма (биообъекта)	Домин ирован ие обратн ых связей
1. Состоя ние нормы	Отсутствие сдвигов функционального состояния. Высокостабильные взаимоотношения функций системы. Нормальный уровень напряжения регуляторных механизмов межсистемной интеграции.	Гомеостаз	Отрица тельны е
2. Групп а риска	Незначительные изменения функционального состояния. Незначительное повышение уровня напряженности регуляторных механизмов межсистемной интеграции.	Самооргани зация	Отрица тельны е
3. Состоя ние 1 степен и напря женно сти	Умеренное повышение уровня напряжения и рассогласования регуляторных механизмов межсистемной интеграции.	Самовосстан овление, мал ые флуктуации- «рыскание»	Неусто йчивая «точка» баланса отрицат ельных и положи тельны х
4. Состоя ние 2 степен и напря женно сти	Выраженные изменения функционального состояния, напряжения и рассогласования регуляторных механизмов межсистемной интеграции.	Переход в новое состояние с возможностя ми восстановле ния, бифуркации, большие флуктуации	Полож ительн ые
5. Состоя ние 3 степен и напря женно сти	Резко выраженные изменения функционального состояния. Высокий уровень напряжения и рассогласования регуляторных механизмов межсистемной интеграции.	Переход в новое состояние без самовозмож ности восстановле ния, бифуркации практически отсутствуют , флуктаций нет	Отрица тельны е – «закреп ляющи е» новые состоян ия

Оценку напряженности функционального состояния предлагается осуществлять следующими методами.

Первый подход заключается в применении интегрального показателя, характеризующего «удаленность» текущего состояния от некоторой нормы - условно оптимально работающей функциональной системы. Из классической теории распознавания образов известно, что одним из лучших таких показателей является нормированное по дисперсии расстояние в факторном пространстве. Аналогом этого показателя в медицине является показатель стабильности системной организации функций (ПССОФ) (ПСОФ) и его различные модификации. Обобщая проведенные исследования в данном направлении, в качестве классификатора биообъекта по стратам напряженности функциональных состояний, предлагается применение следующего интегрального показателя функционирования биообъекта (ИПФБ) — формула (1):

$$ИПФБ = \frac{1}{\sum_i \alpha_i} \cdot \sum_i \frac{\alpha_i \cdot (X_i - X_{i,норм})^2}{\sigma_{i,норм}^2} \quad (1)$$

где X_i - значение характеристики i , $X_{i,норм}$, $\sigma_{i,норм}$ - номинальное (модальное) значение и СКО (среднеквадратичное отклонение) X , α_i - степень значимости (определяется экспертным путем или экспериментально, например, с помощью аппарата искусственных нейронных сетей). Степень значимости позволяет для различных биологических (функциональных) систем организма получить характерные для каждой из них множества значений (векторы) весовых коэффициентов α_i для множества регистрируемых характеристик X , позволяющее производить наилучшую диагностику (разделение классов).

Применение данного показателя для построения решающих правил идентификации классов состояний (нозологий) предполагает следующую технологию.

На первом этапе осуществляется сбор фактологического репрезентативного материала и отбрасываются артефакты. Общая выборка случайным образом делится на три части – обучающую, настраивающую и экзаменационную. На первой выборке

рассчитываются показатели системной организации (по формуле (1) или обобщенной формуле (2) – последняя учитывает корреляцию между регистрируемыми характеристиками). На второй - определяются функции принадлежности. На третьей выборке определяется эффективность идентифицированных диагностических решающих правил. Соотношения статистических мощностей указанных выборок рекомендуется выбирать исходя из взаимной репрезентативности (они должны подчиняться подобным законам распределения). Опыт доказывает, что этого можно достичь, используя случайную сортировку по равномерному закону распределения и принцип «золотого сечения», т.е. примерное соотношение объемов указанных выборок - 0,46:0,32:0,22.

Второй подход анализа системных сдвигов заключается в следующем. Он основан на том, что различные функциональные состояния биообъекта различаются корреляционными связями между информационными (в том числе, системообразующими) характеристиками. Это различие наблюдается как в топологическом аспекте (одни связи исчезают, другие появляются), так и информационно-динамическом (меняется сила влияния, ее направление, характер – линейный или нелинейный, детерминированный, вероятностный или хаотический, дискретный или аналоговый, с запаздыванием или без, парный или множественный и т.п.).

При анализе результатов исследований состояний определенной системы, не имеющих непрерывный временной характер, непосредственное применение данного подхода (основанного на анализе матриц связности) затруднено, поскольку в этом случае невозможно вычисление значений корреляционных функций для конкретного объекта в определенный момент времени.

В связи с этим предлагается применять специальный показатель функциональных сдвигов (ПФС), количественно характеризующий «отклонение» матриц связности от некоторого базового (нормативного) состояния.

ПФС вычисляется следующим образом. Пусть, для некоторого базового класса (например, «нормальное»

функционирование) были идентифицированы различные математические зависимости, отражающие взаимосвязь значений регистрируемых показателей – множество $\{F\}$. Предполагаем, что упомянутые выше корреляционные сдвиги вызывают «сдвиги» и между структурами и параметрами математических (в простейшем случае – алгебраических) моделей. Тогда формулу (1) можно использовать для вычисления ПФС, осуществляя переход от исходного факторного пространства к виртуальному и рассматривая в качестве X_i (в формуле (1)) значения квадрата относительного отклонения «истинного» (зарегистрированного) значения характеристики i от вычисленного по идентифицированным ранее (на этапе обучения) вышеуказанным математическим моделям. Аналогичным образом может быть применена и формула (2).

Порядок выполнения работы

1. Изучить теоретический материал.
2. В среде EXCEL сформировать таблицу данных протокола мониторинга (по Приложению к лабораторной работе №2).
3. Построить корреляционную матрицу.
4. Отобразить графически (в виде связанных графов, вершинами которых являются регистрируемые характеристики, дугами – статистически значимые корреляционные связи, количество дуг между двумя вершинами – ранг связности, определяемый по формуле $rangs_{xy} = round(6 * \pi * (\cosh(R_{xy}) - 1))$, пунктир дуги означает «отрицательную» связь) построенные матрицы.
5. Рассчитать значение множественного градиента функциональных различий.
6. Рассчитать значения показателей ИПФБ и ПФС для каждого класса состояний, считая равными единице весовые коэффициенты.
7. Рассчитать значения показателей ИПФБ и ПФС для каждого класса, оценив (до второго знака после запятой) значения весовых коэффициентов как отношения среднего значения анализируемой характеристики к нормативному значению (если нормативное значение не задано, то в качестве него принимается модальное значение одного из класса – состояний)

8. Оценить напряженность состояния биообъекта путем соотнесения к определенному уровню напряженности по логарифмической шкале, а именно – натуральный логарифм значения показателя после округления до целых значений (уровень 1 – менее 2, уровень 2 – [2-3], уровень 3 – менее [4-7], уровень 4 – менее [7-10], уровень 5 – более 10).

9. Оформите отчет, включающий результаты выполнения работы и действия к ним приводящие, выводы и краткие ответы на контрольные вопросы, аннотацию не менее 2 информационных источников, найденных в сети интернет по тематике лабораторной работы (объемом 150-350 слов).

Примечания: для расчетов и их графической интерпретации используйте любые офисные средства.

Контрольные вопросы.

1. Что такое функциональное состояние биосистемы?
2. Как определяется системообразующий показатель?
3. Что показывает граф связности регистрируемых показателей?
4. Каким образом можно оценить напряженность состояния биообъекта?
5. Какое состояние биообъекта рекомендуется выбирать в качестве базового?
6. Какие недостатки (в смысле классификационного процесса) имеет показатель ИПФБ?

Библиография.

1. Дронова Т.А. Диагностика заболеваний органов пищеварения с применением традиционных приемов и способов оценки системных нарушений функций разного биологического качества/ Т.А. Дронова А.В. Завьялов // Курск, КГМУ, 2012. — 44 с.
2. Завьялов, А.В. Системная организация физиологических функций в норме и патологии: Актовая речь на заседании Ученого совета Курского государственного медицинского университета 8 февраля 2005 г. - Курск, КГМУ, 2005.-20 с.

ЛАБОРАТОРНАЯ РАБОТА №5. ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ ЗАБОЛЕВАЕМОСТИ В РЕГИОНЕ

Цель работы: овладение навыками методами идентификации и анализа прогностических моделей уровня заболеваемости в регионе на основе данных многолетних наблюдений.

Краткие теоретические сведения.

Экстраполяционные методы прогнозирования основываются на концепции сохранения в будущем тенденций прошлых закономерностей с учетом текущей ситуации (закономерностей) - различные методы представлены, например, в работах.

В условиях невозможности детерминированного (или плохо формализуемого) применения временных рядов для прогнозирования и большой неопределенности объекта (нечеткости или грубости результатов мониторингования) применяются методы экспертных оценок.

Методы имитационного моделирования предполагают разработку математических или логических моделей будущего функционирования объекта, в том числе с применением лингвистических переменных. В этом случае, удачно применяют достижения искусственного интеллекта на основе: формального аппарата математической логики, индукции и дедукции, теория вероятностей и статистические методы, теория распознавания образов, теория нечетких множеств и нечеткого логического вывода, искусственные нейронные и иммунные сети, методы построения информационно-аналитических и концептуальных моделей, мягкие вычисления, теорию субъективного анализа и т.д.

В обобщенном виде схема прогнозирующей системы показана на рисунке 1.

К наиболее распространенным методам прогнозирования, применяемым в настоящее время, относятся:

1. Регрессионные модели синтезируются с использованием специальных методов подбора вида экстраполирующей функции и определения значений её параметров (<http://prognoz.org/lib/metody-prognozirovaniya>);

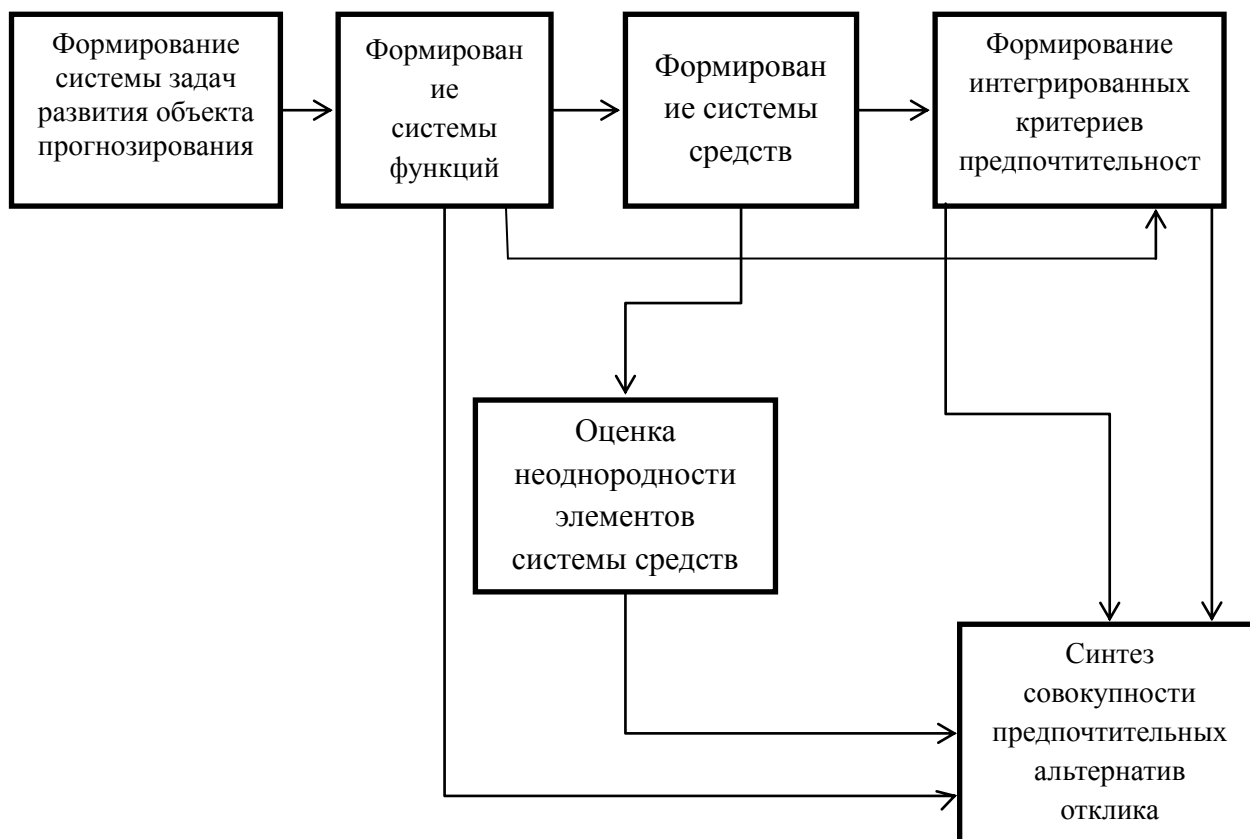


Рисунок 1 - Схема прогнозирующей системы.

2. Адаптивное сглаживание предполагает итерационный процесс пересмотра выбранных значений весовых коэффициентов: веса пересматриваются по завершении каждого прогнозного периода и селектируются те значения, при котором прогноз был бы наименее ошибочным (бизнес-учебники.рф/logist/metodyi-pronozirovaniya.html);

3. Факторный анализ позволяет на основе натуральных данных экспериментального наблюдения за изменениями значений характеризующих биообъект признаков сформировать определенное (конечное) множество группы показателей (например, латентных), определяющих корреляционную взаимосвязь между признаками (<http://prognoz.org/lib/faktorny-analiz>);

4. Многомерная фильтрация;

5. В процессе имитационного моделирования синтезируются структуры, описывающие объекты или процессы с определенными коэффициентами подобия (как во времени, так и в пространстве) действительности с задаваемыми ограничениями

(http://ru.wikipedia.org/wiki/Имитационное_моделирование);

6. Метод группового учета аргументов (МГУА) использует семейство различных индуктивных алгоритмов для математического моделирования мультипараметрических данных, классификации, оценки информативности, решения задач интерполяции и экстраполяции временного ряда на основе различных аппроксимационных функций. Метод основан на рекурсивном селективном отборе моделей по иному от идентификации модели критериям и иной выборке, на основе которых постепенно усложняются структуры моделей, до момента «вырождения» значений критериев селекции.

(http://ru.wikipedia.org/wiki/Метод_группового_учета_аргументов);

7. Экспоненциальное сглаживание тренда – моделирование путем использования экспоненциальных функций, обладающих свойствами непрерывности производных, в окне задаваемого пользователем размера анализируемого временного ряда (http://ru.wikipedia.org/wiki/Экспоненциальное_сглаживание);

8. Спектральные методы – методы, основанные на изучении спектров излучения, поглощения и рассеивания (не путать с Фурье-анализом временного ряда);

9. Метод скользящей средней применяется для выравнивания временного ряда на основе вычисления средневзвешанных характеристик «окна» определенного размера;

10. Применение сплайн-функций предполагает идентификацию некоторых элементарных математических функций, область определения которых разбита на конечное число, как правило, равноотстоящих отрезков, на каждом из которых сплайн-аппроксимация совпадает с алгебраическим полиномом или гармонической функцией (<http://ru.wikipedia.org/wiki/Сплайн>);

11. Оптимальные фильтры;

12. Метод Бокса-Дженкинса предполагает прогнозирование путем применения авторегрессионных структур моделей интегрированного скользящего среднего (<http://business-gruppa.ru/box-jenkins-metod-boksa-dzhenkinsa>);

13. Метод Марковских цепей основывается на анализе последовательности случайных событий с конечным или счётным числом исходов при фиксированном настоящем зависимым от

конкретного «прошлого» (обычно с единичным временным запаздыванием (http://ru.wikipedia.org/wiki/Цепь_Маркова));

14. Модели разностных уравнений используются, как правило, для исследования динамических характеристик импульсных систем (например, стабилизаторов напряжения, цепочек импульсов, передаваемых в нейронных сетях)

15. Авторегрессионная модель временных рядов основывается на автокорреляции – т.е., на построении линейных (или нелинейных) структур с запаздыванием (зависящих от «прошлых» значений ряда с определенным временным шагом) (http://ru.wikipedia.org/wiki/Авторегрессионная_модель);

16. Вероятностный метод позволяет с приемлемой (задаваемой) точностью определить, в каких пределах будет изменяться искомая величина и-или с какой вероятностью следует ожидать наступление определенного события (значения временного ряда) (<http://pictoris.ru/1/4/index.html>).

В качестве статических критериев, характеризующих адекватность и приемлемость модели, предлагается применять: тип распределения, оценку надежности связи между системообразующими поведением объекта характеристиками, однородность и репрезентативность динамического ряда, оценку уровня (в определенном интервале) гармонических (или иных ритмических) волновых составляющих.

При анализе поведения системы синтезируются модели, отражающие динамику поведения каждого ее элемента во времени и связей между ними, по которым осуществляется прогноз перехода системы в определенное. Для чего используются иногда функции когерентности.

При этом, возникают проблемы оценки качества прогноза до его реализации и оценка достоверности прогноза, который еще не осуществлен (аналогично проблеме необходимости наличия постаприорной вероятности в формулах Байеса при распознавании образов).

Применяемые в настоящее время методы верификации прогноза в основном оперируют статическими процедурами оценки доверительных интервалов прогнозных значений. Ошибки

возникают в двух случаях: информационные ошибки описания объекта и ошибки применяемого метода прогнозирования.

К наиболее эффективным относят следующие принципы прогнозирования:

1. Активное управление активно при пассивном прогнозе – эффективность прогноза реализуется механизмами управления;
2. Эффективность прогноза определяется мерой ее детерминированности;
3. Эффективность прогноза функционально зависит от параметров описания объекта, системно связанных друг с другом;
4. Эффективность прогноза для различных уровней описания объекта структурирования ограничена неуправляемыми факторами развития объекта, которые невозможно учесть на начальном этапе моделирования.

Обобщаемая структура взаимосвязи методов прогнозирования с исходной информацией показана на рисунке 2.

Экстраполяция, проводимая в будущее, - это перспектива, а в прошлое, - ретроспектива. К предпосылкам использования экстраполяционной методологии относятся:

- высокая вероятность того, что развитие исследуемого явления, характеризуемого рассматриваемым временным рядом, в целом описывать некоторой плавной кривой (дифференцируемой и интегрируемой);
- общая тенденция развития явления в прошлые и текущие моменты времени с высокой степенью вероятности не изменяются в ближайшем будущем.

При этом могут использоваться различные методы анализа структур данных и моделирования временных рядов в зависимости от исходной информации.

Упрощенные приемы целесообразны при недостаточной информации о предыстории развития явления (нет достаточно «длинного» ряда – у исследователя имеется обучающая выборка небольшого объема с точки зрения статистической репрезентативности рассматриваемого процесса или поведения объекта – эта ситуация характерна для биообъекта).

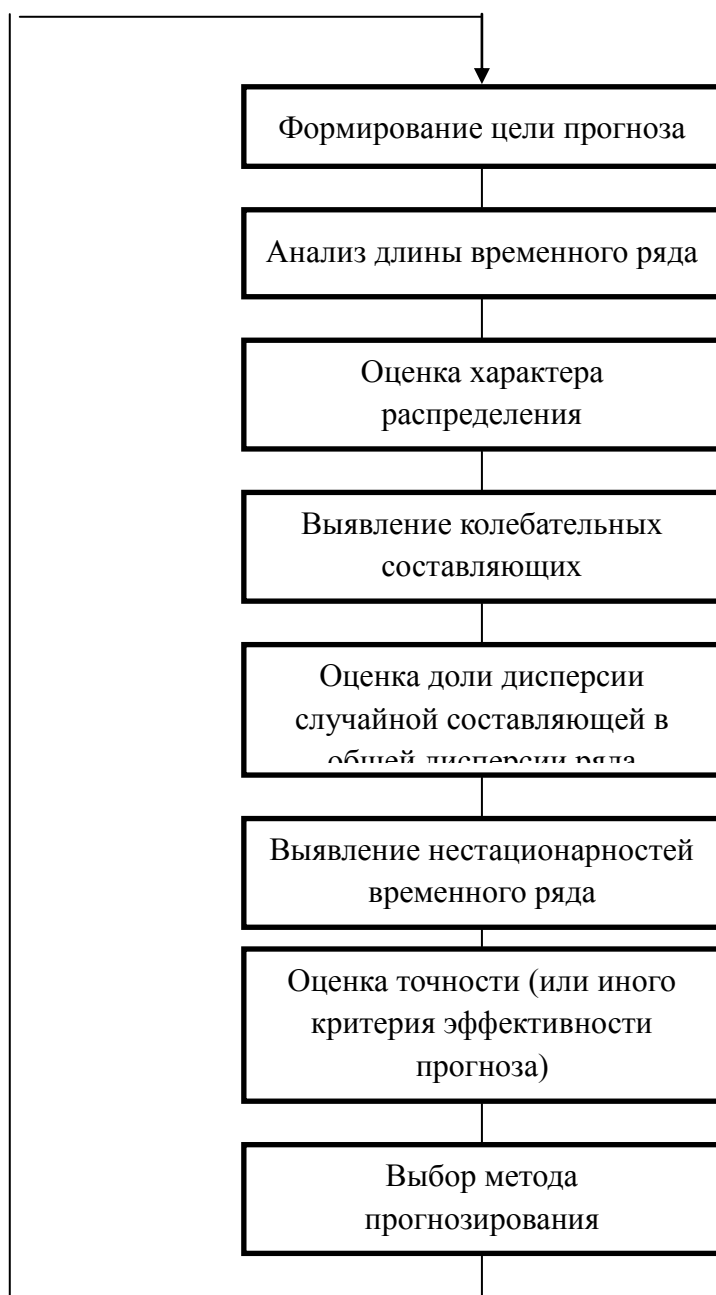


Рисунок 2 - Обобщенная структура связи исходной информации и метода прогнозирования.

Применяемы в этом случае, упрощенные приемы математического моделирования основываются на усредненных значениях показателей рассматриваемого динамического ряда. Например, для эконометрических показателей (подобными во многом показателям биообъектов, если в качестве последних рассматриваются группы людей, заболеваемость и т.п. в регионе), выделяются:

1. *Метод среднего абсолютного прироста.* Для нахождения искомого аналитического выражения, определяющего тенденцию на любое время (дату), определяется средний абсолютный прирост и осуществляется его накопление (последовательное прибавление его значения к последнему уровню ряда столько раз, на сколько периодов осуществляется экстраполяция ряда). Применение в экстраполяции среднего абсолютного прироста предполагает, что развитие явления происходит по арифметической прогрессии и относится в прогнозировании к классу «наивных» моделей, ибо чаще всего развитие явления следует по иному пути, чем арифметическая прогрессия. Поэтому данный метод относится к процедурам «разведочного» анализа и находит применение в качестве предварительного прогноза, когда у модельера нет достаточно репрезентативного динамического ряда (например, имеется только информация в начале и конце рассматриваемого периода (например, данные на начало и окончание года)).

2. *Метод среднего темпа роста.* Данный метод осуществляется в случае, когда общая тенденция характеризуется показательной кривой (т.е. степенной или экспоненциальной или логарифмической).

3. *Выравнивание рядов по определенной аналитической формуле.* Определение формулы в данном случае обычно осуществляется с помощью экспертного анализа.

4. *Аналитическое сглаживание* позволяет определить общую тенденцию изменения явления на определенном временном отрезке и выполнить расчеты для периодов, в отношении которых нет исходных данных. Характерной особенностью данного метода является неглубокая величина прогноза (достаточно хорошо он работает при краткосрочном прогнозе).

5. *Адаптивные методы* используются в условиях большой колебательности значений временного ряда и позволяют при изучении тенденции учитывать степень влияния предыдущих уровней на последующие значения. К адаптивным методам относятся, например, алгоритмы, реализующие ранее упомянутые процедуры скользящих и экспоненциальных средних, метод гармонических весов, методы авторегрессионных преобразований.

Основная цель адаптивных методов заключается в структурно-параметрической идентификации самонастраивающихся моделей, способных учитывать информационную ценность (веса) различных членов (термов) математических структур моделей временного ряда и осуществлять оценки с приемлемой точности прогнозируемым значениям ряда.

6. Особенность метода *экспоненциального сглаживания* заключается в том, что в процедуре выравнивания каждого наблюдения используется только значения предыдущих уравнений, взятых с определенным весом. Смысл экспоненциальных средних состоит в нахождении таких средних, в которых влияние прошлых наблюдений затухает по мере удаления от момента, для которого определяется средние («прошлое забывается по мере отдаления от настоящего»). Главный недостаток, который проявляется особенно при анализе биомедицинских сигналов, это не учет ритмических составляющих, присущих большинству саморганизуемых объектов и-или процессов в силу наличия у них автоколебаний.

Любой статистический прогноз носит приближенный характер, поэтому целесообразно определение доверительных интервалов прогноза.

Для временных рядов главный интерес представляет описание или моделирование их структуры. Цель таких исследований, как правило, шире моделирования, хотя некоторую информацию можно получить и непосредственно из модели, делая выводы о выполнении определенных физических или биологических законов и проверяя различные гипотезы. Построенные модели могут использоваться для статистического моделирования длинных рядов наблюдений при исследовании больших систем, для которых временной ряд рассматривается как входная информация.

Основной тенденцией развития процесса или поведения объекта называется плавное и устойчивое изменение уровня характеризующего наблюдаемого показателя во времени, свободное от случайных колебаний. В этом случае задача состоит в выявлении общей тенденции в изменении уровней ряда, освобожденной от действия различных факторов.

Изучение тренда включает два основных этапа:

- ряд динамики проверяется на наличие тренда;
- производится выравнивание временного ряда и непосредственно выделение тренда с экстраполяцией полученных результатов.

С этой целью временные ряды обрабатываются методами укрупнение интервалов, скользящей средней и-или аналитического выравнивания:

1. *Метод укрупнения интервалов.* Одним из наиболее простых способов изучения общей тенденции временного ряда является укрупнение интервалов - он основывается на укрупнении периодов, к которым относятся уровни временного ряда динамики.

2. *Метод скользящей средней.* Выявление общей тенденции ряда динамики осуществляется путем сглаживания значений временного с помощью скользящей средней (рассматривался ранее). Скользящая средняя – это «подвижное» значение некоторой средней динамической величины, рассчитываемой по массиву элементов временного ряда путем последовательного передвижения, как правило, на один временной интервал. То есть, первоначально вычисляется некоторый средний (или медианный) уровень из определенного числа первых по порядку уровней ряда, затем - средний уровень из такого же числа членов, начиная со второго. Таким образом, «скользящая средняя» как бы передвигается по вектору значений временного ряда от начала к концу, последовательно раз отбрасывая один уровень в начале и добавляя один в следующий. При этом посредством осреднения эмпирических данных индивидуальные колебания погашаются, и общая тенденция развития явления выражается в виде определенной плавной линии.

«Скользящая средняя» обладает достаточной гибкостью. Однако, существенным недостатком метода является укорачивание сглаженного ряда по сравнению с фактическим, что ведет к потере информации. В связи с этим, «скользящая средняя» не позволяет получить аналитического выражения для тренда и, следовательно, осуществлять имитационное моделирование.

Период скользящей может быть четным и нечетным. Практически удобнее использовать нечетный период, так как в

этом случае скользящая средняя будет отнесена к середине периода скольжения. Полученные средние соотносятся к соответствующему срединному интервалу.

Особенность сглаживания по четному числу уровней состоит в том, что каждая из численных (например, четырехчленных) средних относится к соответствующим промежуткам между смежными периодами. Для получения значений сглаженных уровней соответствующих периодов необходимо произвести центрирование расчетных средних.

Недостатком способа сглаживания рядов динамики является то, что полученные средние не дает теоретических рядов, в основе которых лежала бы математически выраженная закономерность.

3. *Метод аналитического выравнивания.* Более совершенным приемом изучения общей тенденции в рядах динамики является аналитическое выравнивание. При изучении общей тенденции методом аналитического выравнивания исходят из того, что изменения уровней ряда динамики могут быть с той или иной степенью точности приближения выражены определенными математическими функциями. Вид уравнения определяется характером динамики развития конкретного явления. Логический анализ при выборе вида уравнения может быть основан на рассчитанных показателях динамики, а именно:

- если относительно стабильны абсолютные приросты (первые разности уровней приблизительно равны), сглаживание может быть выполнено по прямой;
- если абсолютные приросты равномерно увеличиваются (вторые разности уровней приблизительно равны), можно принять параболу второго порядка;
- при ускоренно возрастающих или замедляющихся абсолютных приростах - параболу третьего порядка;
- при относительно стабильных темпах роста-показательную функцию.

Для аналитического выравнивания наиболее часто используются следующие виды трендовых моделей: прямая (линейная), парабола второго порядка, показательная (логарифмическая) кривая, гиперболическая.

Цель аналитического выравнивания - определение аналитической или графической зависимости. На практике по имеющемуся временному ряду задают вид и находят параметры функции, а затем анализируют поведение отклонений от тенденции. Чаще всего при выравнивании используются следующие зависимости; линейная, параболическая и экспоненциальная.

После выяснения характера кривой развития необходимо определить ее параметры, что можно сделать различными методами:

1. решением системы уравнений по известным уровням ряда динамики;

2. методом средних значений (линейных отклонений), который заключается в следующем: ряд расчленяется на две примерно равные части, и вводятся преобразования, чтобы сумма выровненных значений в каждой части совпала с суммой фактических значений, например, в случае выравнивания прямой линии

$$\sum(\gamma - a_0 - a_1 t) = 0; \quad (1)$$

3. выравниванием ряда динамики с помощью метода конечных разностей;

4. методом наименьших квадратов: это некоторый прием получения оценки детерминированной компоненты $f(t)$, характеризующих тренд или ряд изучаемого явления.

Во многих случаях моделирование рядов динамики с помощью полиномов или экспоненциальной функции не дает удовлетворительных результатов, так как в рядах динамики содержатся заметные периодические колебания вокруг общей тенденции. В таких случаях следует использовать гармонический анализ.

Порядок выполнения работы.

1. Самостоятельно изучите теоретический материал.

2. Из приведенного в приложении временного ряда сформируйте последовательность, начиная с номера A_1 в количестве A_2 измерений. A_1 определяется как остаток от деления порядкового

номера в группе на семь плюс 1. A2 определяется как утроенное значение количества букв в Вашей фамилии.

3. С помощью инструментария Excel идентифицируйте трех аппроксимантов полученных в п.2 временных трендов: лучшая по критерию детерминированности модель в режиме «построить линию тренда»; авторегрессионную модель первого порядка; гармоническую модель по двум-трем частотам (частоты определите путем подбора соответствующих периодов циклов, полученных по анализу первой производной временного тренда согласно численному дифференцированию).

4. Постройте графики временного тренда и полученных моделей на одной плоскости.

4* (повышенной сложности). Постройте модель спектра Фурье.

5. Оцените средние значения относительных ошибок аппроксимации: интерполяции, экстраполяции (до и после интерполяционного интервала). Сделайте выводы.

6. Оформите отчет, включающий в себя результаты работы (возможны скриншоты), выводы, краткие ответы на контрольные вопросы, аннотацию одного из информационных источников, указанных в библиографии или иных (найденного самостоятельно).

Контрольные вопросы:

1. Что называется интерполяцией?
2. Что называется экстраполяцией?
3. Каким образом строятся гармонические модели?
4. Охарактеризуйте виды прогнозов (сиюминутный, краткосрочный, среднесрочный, долгосрочный)?
5. Как осуществляется проверка качества прогностической модели?
6. Могут ли прогностические модели быть логическими?
7. Как осуществляется прогноз во времени и пространстве?
8. Для чего необходимо прогнозировать заболеваемость в регионе?
9. Какие заболевания населения носят ритмический характер?
10. Какие природные циклы оказывают влияние на региональную заболеваемость (и почему)?

Библиография

1. Демографические прогнозирование. Презентация.
<http://900igr.net/prezentatsii/geografija/Demograficheskoe-prognozirovanie/008-Vidy-i-metody-prognozirovaniya.html>
2. Дуброва, Т. А. Прогнозирование социально-экономических процессов [Текст]: учебное пособие / Т. А. Дуброва. - 2-е изд., испр. и доп. - М.: Маркет ДС, 2010. - 192 с.
3. Методы прогнозирования. Презентация.
<http://nashaucheba.ru/v24021>
4. Методы прогнозирования. Презентация. <http://ppt-online.org/5165>

Приложение

Динамика психических заболеваний в регионе

t	0	1	2	3	4	5	6	7	8	9	10
Ур-нь	19,4	24,09	27,95	32,61	36,89	40,33	44,8	48,7	50,88	52,78	54,56
11	12	13	14	15	16	17	18	19	20		
56,42	58,15	58,94	60,03	60,39	58,68	57,43	56,83	55,37	53,70		
21	22	23	24	25	26	27	28	29	30		
51,49	48,67	45,33	40,75	36,89	34,00	29,59	24,72	20,06	14,38		

ЛАБОРАТОРНАЯ РАБОТА №6. АНАЛИЗ ДИНАМИКИ ЭКОЛОГИЧЕСКОЙ СИТУАЦИИ В РЕГИОНЕ

Цель работы: овладение навыками математического моделирования экологической ситуации в регионе как одного из основных факторов влияющих на заболеваемость на основе результатов биометрического мониторинга.

Краткие теоретические сведения.

Ухудшение экологической обстановки и социальной среды существенно отражается на состоянии здоровья человека. Здоровье человека и биосферы неразделимо связаны и определяются множеством компонент. Взаимодействуя с миром в ходе своей деятельности, человек ощущает на себе ответную реакцию окружающей среды.

Кроме неизбежных природных явлений (таких как изменение солнечной активности), на здоровье человека могут влиять экологические факторы, вызываемые им самим в ходе своей деятельности. Указанное обуславливает необходимость управления деятельностью основных загрязнителей окружающей среды на основе математических методов (моделей), позволяющих достаточно адекватно оценить влияние различных факторов на рассматриваемый класс заболеваний с целью прогноза и (или) управления динамикой последнего.

Региональные экологические проблемы, сформировавшиеся в результате загрязнения окружающей среды из-за деятельности человека, требуют для своего решения использования региональных информационных систем (РИС). Одна из задач таких систем должна состоять в своевременном определении воздействия загрязняющих веществ на здоровье человека на основании анализа накопленной информации о состоянии окружающей среды и медико-биологической информации и прогнозирования уровня региональной заболеваемости. Исследование и прогноз выполняются на основе имитационного моделирования.

Для характеристики уровня экологической напряженности региона используют понятие экологической нормы, которое отражает определенные параметры сохранения приспособительных структур и функций экосистемы определенного иерархического

уровня. Такое определение нормы может указывать на степень максимально допустимого воздействия человека и общества на окружающую среду, которое обеспечивает функционирование и сохранение структуры и динамических качеств экосистемы в целом.

Поражение городского населения возможно главным образом через атмосферу, экологических природных систем - через все природные среды.

Таким образом, выделяются два аспекта проблемы для изучения влияния антропогенной деятельности человека на заболеваемость:

а) воздействие на импактном уровне (на относительно небольшой территории);

б) массовое воздействие на природу, природные экосистемы на фоновом (как правило, невысоком) уровне, но на обширных территориях, практически по всей территории земного шара.

В концептуальном моделировании принято рассматривать три этапа:

- сбор и анализ априорной информации о предметной области и проблемной среде;

- концептуальный анализ предметной области с учетом требований пользователей;

- концептуальный синтез или собственно построение концептуальной модели предметной области.

Общая технология экологического управления в регионе состоит из трех этапов.

Целью первого этапа является получение информации о фактическом загрязнении сред региона. Учитывается как анализ источников антропогенного загрязнения региона, так и анализ естественных процессов, приводящих к фоновым концентрациям загрязняющих веществ в средах региона.

На втором этапе оценивается влияние состояния среды на заболеваемость населения.

На третьем этапе строится прогноз заболеваемости населения в зависимости от состояния среды и изменение самой среды, с последующей выдачей рекомендаций планирующим, природоохранным и хозяйственным органам.

Разработка пути возможного оздоровления и профилактика уровня заболеваемости в регионе в автоматизированной системе основывается на оценке влияния выбросов отдельных предприятий на ту или иную заболеваемость с последующей выдачей рекомендаций планирующим, природоохранным и хозяйственным органам о проведении мероприятий, призванных скорректировать выбросы соответствующих предприятий. Таблица связей, полученная при помощи автоматизированной системой моделирования, призвана ставить в соответствие предприятиям региона мероприятия по снижению выбросов.

В общем виде методика анализа вклада выбросов отдельного предприятия на уровень заболеваемости населения выглядит следующим образом:

- 1) Определяется список экологических факторов, обусловленных выбросами $V_I - V_p$ в окружающую среду, влияющих на уровень заболеваемости по нозологии N_k .
- 2) Используя полученную математическую модель влияния факторов окружающей среды на уровень заболеваемости осуществляется прогноз о показателях заболеваемости в конкретном регионе.
- 3) На основании полученных показателей заболеваемости оценивается вклад каждого выброса в рост уровня заболеваемости населения по определенной нозологии в регионе.
- 4) Определяется вклад каждого выброса конкретного предприятия административного района в рост уровня заболеваемости путем нахождения доли выброса этого предприятия к общим выбросам по району.
- 5) Определяется общий вклад предприятия в рост уровня заболеваемости по данной нозологии.
- 6) Определяется экономический ущерб U_k , вызванный влиянием деятельности предприятия на рост данной заболеваемости.
- 7) Шаги 1-6 повторяются для других нозологии, тем самым определяется общий экономический ущерб, вызванный влиянием деятельности предприятия на рост всей заболеваемости населения и определяется размер штрафа R_m для этого предприятия (T), эквивалентный сумме ущерба.

Далее из списка мероприятий для уменьшения показателей выбросов автоматически выбираются конкретные мероприятия, которые необходимо провести на данных предприятиях для уменьшения показателей выбросов, что в свою очередь должно привести к уменьшению заболеваемости системы кровоснабжения у населения.

Методика определения снижения значений выбросов загрязняющих веществ конкретными предприятиями состоит в следующем.

На первом этапе строится прогноз заболеваемости населения региона на следующий год при помощи автоматизированной системы математического моделирования.

Далее, определяется рост заболеваемости в процентном отношении для каждого района и для каждой нозологии.

На третьем этапе осуществляется коррекция значений выбросов предприятий региона с целью уменьшения уровня заболеваемости по следующему алгоритму. Задается порог роста заболеваемости h_i , при котором считается целесообразным принятие мер. Также определяется шаг снижения выбросов предприятием h .

После определения заболеваемости и района, в котором рост заболеваемости ожидается выше порогового уровня, определяются выбросы, значимо влияющие на эту заболеваемость по заданному критерию. Затем определяются предприятия региона, вносящие наибольший вклад в сброс этих веществ, и моделируется снижение уровня выбросов на заболеваемость.

Далее прогнозные значения уровня заболеваемости пересчитываются с новыми значениями выбросов загрязняющих веществ и проверяется условие роста заболеваемости в этом районе. Операция повторяется до тех пор, пока рост заболеваемости не станет ниже h_i .

Порядок выполнения работы.

1. Определить вариант задания по формуле $N_{вар} = N \bmod (4) + 1$, где N - порядковый номер в группе. Согласно приложению и $N_{вар}$ сформировать протокол мониторинга экологической ситуации города $\{X_{N_{вар}}, Y_1, Y_2, Y_3\}$, где X - показатель уровня

заболеваемости населения, Y - показатель загрязненности города определенным веществом.

2. Построить линейные и-или нелинейные регрессионные модели вида (отобрать лучшие по критерию детерминированности):

2.1. $X=f(Y_1)$; $X=f(Y_2)$; $X=f(Y_3)$; $X=f(Y_1, Y_2)$; $X=f(Y_1, Y_3)$; $X=f(Y_2, Y_3)$; $X=f(Y_1, Y_2, Y_3)$;

2.2 Повторить пункт 2.1, используя в качестве Y_i , $i=1..3$ значения Y_{ij} с нарастающим шагом, т.е.

$$Y'_{ij} = \sum_{k=1}^{j-1} Y_{ik}$$

3. На основе информационных источников изучить, что влияет на уменьшение Y_i .

4. На основании анализа математических моделей, идентифицированных в пункте 2 и результатов «экспертного» анализа пункта 3, сформулировать предложения управляющего воздействия на факторы Y с целью улучшения показателя отклика X . Оцените доминантность влияния уровня определенного загрязнения на параметр здоровья. Сделайте вывод о структуре наиболее адекватной модели по критериям корреляции и СКО. Сравните частоты ритмической модели (если она адекватна) с внешними космогеологическими частотами (см. Приложение) и сделайте выводы. Ритмическую модель в данной работе предлагается получить путем анализа автокорреляционной функции или визуального анализа динамики заболеваемости во времени или визуального анализа поведения первой производной, вычисленной численными методами.

5. Оформите отчет, включающий в себя результаты работы (возможны скриншоты), выводы, краткие ответы на контрольные вопросы, аннотацию одного из информационных источников, указанных в библиографии или иных (или найденного самостоятельно).

Контрольные вопросы:

1. С какими природными циклами наиболее коррелирует динамика определенных заболеваний?

2. Каким образом связаны между собой уровни заболеваемости населения и уровни антропогенного воздействия на окружающую среду (на примере уровней загрязнителей)?
3. Почему антропогенное воздействие следует учитывать с нарастающим эффектом?
4. Каким образом используются регрессионные и авторегрессионные математические модели для прогнозирования заболеваний?
5. Как осуществляется прогнозирование в Excel с помощью линии тренда?
6. Каким образом можно прогнозировать ритмические тенденции региональной заболеваемости?
7. Каким образом можно использовать логические функции (модели) для прогнозирования заболеваний?
8. Можно ли использовать искусственные нейронные сети для прогнозирования заболеваемости?
9. Каким образом можно использовать прогностические модели для удаления артефактов и восстановления пропущенных значений в мониторинге заболеваемости или состояния пациента в процессе терапевтического воздействия?

Библиография

1. Артеменко М.В., Протасова В.В. Методы и средства моделирования влияния экологической напряженности региона на здоровье населения: (монография) /Монография. КурскГТУ, 2009, - 225с.
4. Статистика [Текст]: учебник для бакалавров / под ред. И. И. Елисеевой; Санкт-Петерб. гос. экон. ун-т. - 3-е изд., перераб. и доп. - Москва: Юрайт, 2014. - 558 с.
5. Статистические методы, математическое моделирование и прогнозирование в социологических исследованиях [Электронный ресурс]: методические рекомендации по изучению дисциплины для направления подготовки 39.04.01 «Социология» / Юго-Зап. гос. ун-т ; сост. О. А. Гримов. - Электрон. текстовые дан. (295 КБ). - Курск : ЮЗГУ, 2016. - 28 с

Приложение к лабораторной работе №6

Таблица 2. уровни показателей здоровья городского населения

Годы	Смертность	рождаемость	Прививки от дифтерии	Прививки от столбняка	Прививки от кори	Прививки от гепатита
1	13,2	14,99	10,2	24,6	8,9	15,8
2	11,9	13,36	11,0	30,5	7,7	34,3
3	12,8	11,4	9,5	29,3	9,6	24,9
4	13,5	12,23	10,3	27,7	10,6	28,5
5	13,4	12,92	11,3	43	11,7	18,9
6	13,0	11,67	11,2	44,7	11,5	18,2
7	13,7	11,55	12,4	47,5	10,7	13,0
8	13,0	13,3	12,6	44,1	11,0	19,1
9	14,2	14,7	11,7	44,4	11,9	24,2
10	14,6	14,1	12,1	46,6	18,2	28,2
11	13,0	14,6	13,2	39,5	16,6	12,3
12	13,3	14,8	13,0	40,0	16,3	19,51
13	13,6	13,8	13,5	38,7	17,4	10,12
14	13,6	11,8	14,5	40,9	26,8	19,91
15	14,6	10,8	13,8	34,2	25,1	12,54
16	14,7	9,7	11,6	28,3	12,1	4,2
17	16,5	9,0	11,0	31,1	11,0	85,7
18	18,0	9,3	7,3	16,6	5,6	72,2
19	16,7	8,6	11,5	38,7	3,7	2,02
20	17,0	8,3	11,1	52,5	11,4	22,6

Таблица 2. -уровни загрязнения города (условные единицы)

годы	Пыль	Оксид углерода	Диоксид азота	фенол	формальдеги д	марганец
1	6	1,5	2,667	8	3,5	0,8
2	4	0,667	2,222	8	2,75	0,16
3	6	1,45	1,333	6	2,25	1,7
4	2	1,067	4,444	6	2,75	1,7
5	4	1,833	0,889	6	3,25	2,5
6	4	1,833	0,889	4	3,25	1,1
7	6	1,667	1,333	6	3,75	2,0
8	6	1,667	1,333	10	3,0	0,7
9	8	2	1,156	8	2,75	0,6
10	4	1,667	0,889	7,4	2,0	1,2
11	4	1,667	1,333	8	3,0	3,5
12	2	1,667	1,33	8	3,75	1,4
13	2	1,667	1,778	10	3,5	3,2
14	2	1,667	1,333	6	3,25	1,4
15	2	1,663	2,222	4	3,0	0,2
16	2	1,667	1,333	4	3,0	0,6
17	4	1,667	1,333	2	2,5	0,4
18	5	2,333	1,333	8	2,25	0,5
19	1,8	2,5	1,333	4	3,0	0,6
20	1,8	2,67	1,3	4	3,25	0,1

ЛАБОРАТОРНАЯ РАБОТА №7. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Цель работы: овладение навыками имитационного моделирования, включая анимацию, для решения медико-биологических задач

Краткие теоретические сведения.

Модель — это создаваемое человеком подобие изучаемого объекта (макет, изображение, схема, карта, словесное описание, математическое представление и т.п.). Метод моделирования состоит в исследовании объекта, явления или процесса путем построения моделей и их изучения. Модель всегда проще реального объекта, но она позволяет выделить главное, не отвлекаясь на детали. Необходимость моделирования объясняется принципиальной невозможностью исследования многих объектов или большой ресурсоемкостью их изучения.

Различают биофизические, физические, электрические, ситуационные, информационные, математические и другие модели.

Информационная модель — модель объекта, процесса или явления, в которой представлены информационные аспекты моделируемого объекта, процесса или явления. Среди информационных моделей особое место занимают модели представления знаний. Математическая модель — приближенное описание объекта, явления или процесса с помощью математической символики. Эта модель представляет собой систему математических соотношений: формул, функций, уравнений, систем уравнений, описывающих те или иные стороны изучаемого объекта, явления или процесса. Математическое моделирование — мощное средство познания, прогнозирования и управления. Анализ математической модели помогает проникнуть в суть изучаемого объекта или явления.

Математические модели строятся на основе данных эксперимента или умозрительно, описывают гипотезу, теорию или зако-

номерность того или иного феномена и требуют дальнейшей проверки на практике. Различные варианты проводимых экспериментов выявляют границы применения математической модели и создают условия для ее дальнейшей коррекции. Математическое моделирование часто позволяет предвидеть характер изменения исследуемого процесса в условиях, трудно воспроизводимых в эксперименте, а в отдельных случаях позволяет предсказать ранее неизвестные явления и процессы.

Процесс математического моделирования принято делить на несколько этапов.

1. *Постановка задачи.* Необходимо отметить, что построение модели подразумевает наличие у специалиста хорошего уровня знаний предметной области, в рамках которой осуществляется моделирование. В постановку задачи входят определение цели исследования, выделение объекта исследования, определение параметров исследуемого объекта, выявление взаимосвязей между параметрами. Этап завершается записью модели в математическом виде.

2. *Проведение модельных экспериментов.* На этом этапе осуществляется решение прямой задачи, для которой предназначена математическая модель, т. е. получение выходных данных для дальнейшего сопоставления с результатами наблюдений изучаемых явлений. Исследователь сознательно изменяет условия функционирования модели, регистрирует ее «поведение» в разных условиях. Важная роль при проведении модельных экспериментов принадлежит вычислительной технике. Именно она обеспечивает возможность обсчета многочисленных модельных экспериментов. Итогом второго этапа моделирования является множество результатов модельных экспериментов. При математическом моделировании разных процессов и явлений может использоваться один и тот же математический аппарат. Это упрощает задачу моделирования, дает возможность выбора из полученных вариантов.

3. *Оценка реализованной модели.* Выясняют, удовлетворяет ли созданная математическая модель критерию практики, т.е. согласуются ли результаты наблюдений с теоретическими (гипотетическими, модельными) данными в пределах заданной точности. Достижение такого результата означает, что положения, лежащие в основе модели, правильны и модель пригодна для исследования выбранного объекта или явления.

4. *Анализ модели на основе накопленных данных об изучаемом объекте, модернизация первоначально построенной модели.* С получением новых научных данных знания об исследуемом объекте уточняются, и наступает момент, когда результаты, получаемые на основании существующей модели, перестают им соответствовать. Возникает необходимость уточнения данной модели или построения новой. Между моментами построения исходной и последующей моделей проходят разные промежутки времени в зависимости от сути изучаемого явления, уровня и скорости исследования данной предметной области, характера полученных новых знаний и данных.

В медицине модели применяются для исследования структур, функций и процессов на разных уровнях организации живого организма: атомарно-молекулярном, субклеточном, клеточно-тканевом, органно-системном, организменном, биоценоотическом.

В медицине, как и в биологии, используются в большинстве случаев биологические, физико-химические, математические модели. Исторически сложилось, что в медицине до сих пор широко распространены словесные описания объектов и процессов (например, заболеваний), а в последние десятилетия все чаще применяются информационные модели.

Биологические модели в медицине применяются для воспроизводства на лабораторных животных заболеваний или состояний, встречающихся у человека. Таким образом, в эксперименте исследуются механизмы возникновения заболевания, его этиология,

патогенез, течение, изучаются варианты воздействия на протекание болезни, сравнивается эффективность применения различных лечебных пособий. В эксперименте, например, моделируются ишемические нарушения и гипертоническая болезнь, злокачественные новообразования и генетические заболевания, инфекционные процессы и др.

Для реализации биологических моделей экспериментальным животным вводят токсины, заражают их микробами, перевязывают сосуды, исключают из пищи определенные вещества, помещают в искусственно создаваемую среду обитания и др. Подобные экспериментальные модели применяются в нормальной и патологической физиологии, генетике, фармакологии, хирургии, реаниматологии. Физико-химические модели имитируют сложные акты поведения, например формирование условного рефлекса.

Удачным следует признать опыт построения электронных схем, моделирующих биоэлектрические потенциалы в нервной клетке и синапсе на основе данных электрофизиологических исследований.

В настоящее время в медицине самое широкое распространение получили математические модели. Они используются практически во всех ее областях. Математические модели применяются для изучения сложных физиологических процессов, диагностики патологических состояний, исследования взаимодействия систем организма в норме и патологии, при изучении эпидемических процессов, в клинической иммунологии, фармакокинетике.

Из математических моделей, известных в физиологии, следует упомянуть модель возбуждения нервного волокна, предложенную А.Ходжкином и А.Хаксли.

Модель сердечной деятельности Ван дер Пола и Ван дер Марка, основанная на теории релаксационных колебаний, позволила

предсказать возможность особого нарушения сердечного ритма, впоследствии обнаруженного у человека.

Ярким примером использования математической модели для обобщения накопленных экспериментальных знаний является модель кровообращения Ф. Гродинза. Построением и исследованием моделей кровообращения, применяющихся в практике российской сердечно-сосудистой хирургии, занимается В.А.Лищук.

В медицинской информатике широко используется моделирование, особенно часто математическое и информационное. Математические модели используются для расчета клинически значимых показателей при обработке сигналов и изображений, для описания заболеваний и состояний при вычислительной диагностике и прогнозировании. Информационное моделирование все чаще применяется при описании деятельности ЛПУ и их подразделений.

Когда осуществляют процессный анализ в ЛПУ с последующим выходом на планирование эксперимента и принятие управленческих решений, то всегда наталкиваемся на одну и ту же проблему разработки и апробации модели управления процессом обеспечения КМП в условиях реального функционирования ЛПУ. Одна из важных особенностей управления качеством производства медицинских услуг — принципиальная невозможность проведения реальных работ по управлению КМП до завершения эксперимента. Возможным выходом является использование имитационных моделей. Сущность метода имитационного моделирования состоит в построении так называемой имитационной модели исследуемого ЛПУ или его подразделения и целенаправленном экспериментировании с разрабатываемой моделью управления для получения ответов на те или иные вопросы. Говоря о методе имитационного моделирования, как правило, имеют в виду метод, ориентированный на применение вычислительной техники, хотя в

принципе могут использоваться любые средства, включая лист бумаги и карандаш.

Другой важный аспект - использование имитационных моделей в процессе эксплуатации информационных технологий управления для принятия управленческих решений по качеству. Такие модели создаются в процессе проектирования, чтобы их можно было непрерывно модернизировать и корректировать в соответствии с изменяющимися условиями работы пользователей. Эти же модели могут быть использованы для обучения персонала ЛПУ перед вводом в действие разработанных технологий в эксплуатацию или для проведения деловых игр.

Принципиальные возможности метода имитационного моделирования весьма велики, он позволяет при необходимости исследовать системы любой сложности и назначения с любой степенью детализации. Ограничениями являются лишь мощность используемой ПЭВМ и трудоемкость подготовки сложного комплекса программ. Методы имитационного моделирования развиваются в основном в направлении исследования степени подобия имитационных моделей реальным системам и разработки типовых методов и приемов создания имитационных моделей.

Имитационное моделирование в медицине используется в основном по следующим направлениям:

1. при исследовании сложных внутренних и внешних взаимодействий ЛПУ с целью оптимизации их функционирования. Для этого на модели ЛПУ изучают закономерности взаимосвязи переменных, вносят в модель ЛПУ изменения и наблюдают их влияние на поведение системы производства медицинских услуг;
2. для прогнозирования поведения ЛПУ в будущем на основе моделирования развития самого ЛПУ и его внешней среды;
3. в целях обучения персонала ЛПУ, которое может быть двух типов:
 - первый тип - индивидуальное обучение оператора, управляющего неким технологическим процессом или медицинской аппаратурой;

- второй тип - обучение группы персонала, осуществляющей коллективное управление сложным объектом по производству медицинских услуг.

Рассмотрим простейшие примеры математического имитационного моделирования применяемого в медико-биологических исследованиях.

1.. Игра «Жизнь».

Многие процессы, интересующих медиков (например, динамика возрастного состава в регионе) может быть с определенной точностью описано математической моделью. Самая простая модель известна под названием «Игра Жизнь».

Предполагается наличие прямоугольного клетчатого поля, в каждой клетке которого может «жить» существо. Если клетка пустая – то в ней никто не живет. Модель задается двумя параметрами: начальной конфигурацией (размером поля и расположением живых существ) и определенными «биологическими законами», регулирующими жизнь популяции существ. На каждой итерации осуществляется последовательный просмотр всех клеток с некоторой начальной (координаты клетки выбираются исследователем или случайным образом) с применением к ним биологических законов.

В качестве типовых законов, предлагаются, например, следующие:

1. Если выбранная клетка пуста, а в соседней с ней клетках находится более двух существ, то внутри клетки появляется существо («размножение»).
2. Если выбранная клетка непуста, а в соседних с ней клетках живет меньше трех или больше четырех существ, то клетка очищается (существо в ней погибает от одиночества или перенаселения).
3. Если правила 1 и 2 не выполняются, то ничего с клеткой не происходит.

(Под соседними подразумеваются восемь окружающих клеток, за границей ореала – прямоугольного поля – существ нет).

2. Модель «хищник» - «жертва».

В системе хищник-жертва ситуация моделируется следующим образом. В случае конкурирующих популяций исчезновение одной означает выигрыш для другой в борьбе за дополнительные ресурсы. Обозначим через C численность популяции хищника, N – популяцию жертвы. Наиболее популярная модель, отражающая колебания численности имеет вид:

$$\begin{aligned} N_{i+1} &= N_i + (r \cdot N_i - a \cdot N_i \cdot C_i) \cdot \Delta t, \\ C_{i+1} &= C_i + (-q \cdot C_i - a \cdot f \cdot N_i \cdot C_i) \cdot \Delta t. \end{aligned}$$

Согласно первому уравнению при $C=0$ численность жертв быстро растет со скоростью r , поскольку модель не учитывает внутривидовой конкуренции. Скорость роста числа жертв ($\frac{\Delta N}{\Delta t}$) уменьшается тем больше, чем чаще происходят встречи особей видов (тогда a - коэффициент эффективности поиска).

Второе уравнение показывает, что в отсутствии жертв численность хищников быстро убывает со скоростью q : положительное слагаемое в правой части уравнения компенсирует эту убыль, f – коэффициент эффективности перехода пищи к потомству хищников.

3. Внутривидовая конкуренция в популяции с дискретным размножением.

Для популяций с дискретным размножением (некоторые виды растений, насекомые) поколения дифференцированно разнесены во времени и особи разных поколений вместе не сосуществуют. Численность подобной популяции характеризуется числом N_t , а время t - дискретная величина – можно, в первом приближении, считать номером популяции. Тогда одна из моделей межвидовой конкуренции может быть описана уравнением:

$$N_{t+1} = \frac{N_t \cdot R}{1 + (a * N_t)^b}$$

где R – скорость воспроизводства популяции в отсутствии внутривидовой конкуренции (математически это соответствует $a=0$); a – параметр, характеризующий

интенсивность внутривидовой конкуренции, при $b=1$ осуществляется выход численности популяции на стационарное значение при любых значениях других параметров модели.

Знаменатель в уравнении отражает наличие конкуренции, делающей скорость роста тем меньше, чем больше численность популяции. Данная модель описывает четыре вида эволюции:

1. монотонное установление стационарной численности популяции;
2. колебательное установление стационарной численности популяции;
3. устойчивые предельные циклы изменения численности популяций;
4. случайные изменения численности популяции без наличия явных закономерностей.

4. Внутривидовая конкуренция в популяции с непрерывным размножением.

В данном случае численность популяции $N(t)$ является непрерывной функцией во времени. В начале эволюционного процесса численность популяции невелика, а ее удельная скорость не зависит от численности: $\frac{1}{N} \cdot \frac{\Delta N}{\Delta t} = r$ - скорость роста численности популяции в отсутствие конкуренции. Далее, по мере роста численности, скорость роста начинает уменьшаться и при достижении определенного критического значения K обращается в ноль. Таким образом, в первом приближении, математическая модель имеет вид:

$$N_{i+1} = N_i + r \cdot N_i \cdot \left(\frac{K-N_i}{K}\right) \cdot \Delta t.$$

Порядок выполнения работы.

1. Изучите теоретические сведения.
2. Изучите возможности имитационного моделирования в биологии на примере игры «Жизнь» (в различных режимах) на портале <http://www.nature.air.ru/models/models.htm>.
3. Проведите с помощью MathCad и/или Excel моделирование динамики в системе «хищник-жертва» при значениях параметров модели: $r=8$, $a=0.15$, $q=2$, $f=0.5$, $N_0=175$, $C_0=45$. Сделайте выводы.

4. Изучите функционирование аналогичной программы, представленной на сайте <http://www.nature.air.ru/models/models.htm>.
5. Исследуйте зависимость амплитуд колебаний численности хищников от амплитуд колебаний численности жертв в зависимости от параметров a и f . Значения остальных параметров фиксируйте по своему усмотрению.
6. Проведите с помощью MathCad и Excel моделирование в четырех видах эволюции с $R=2$ (1 и 2 режимы) и $R=3$ (3 и 4 режимы). На фазовой плоскости (b, R) найдите границы зон, разделяющих разные режимы эволюции изучаемой системы. Сделайте выводы.
7. Составьте и отладьте программу моделирующую внутривидовую конкуренцию в популяции с непрерывным размножением при нескольких значениях входящих в модель параметров на языке высокого уровня и предложите ее реализацию в Excel. Сделайте выводы.
8. Постройте (или найдите в информационных источниках) альтернативную модель внутривидовой конкуренции и исследуйте предсказываемые режимы.
9. Изучите интерфейс и функциональные возможности программ, представленных на сайте <http://www.nature.air.ru/models/models.htm>. Результаты изучения представьте в табличном виде.
10. Оформите отчет, включающий результаты выполненных действий и презентации к минидокладу на тему «Роль имитационного моделирования в медико-биологических исследованиях».
11. Придумайте контрольные вопросы (не менее 5) в виде мини тестов по указанной в п.10 тематике. (Конкретное гносеологическое насыщение вопросов – на Ваше усмотрение).

Библиография

1. Имитационное моделирование в биологии. Презентация. <http://900igr.net/prezentatsii/biologija/Modelirovanie-v-biologii/Modelirovanie-v-biologii.html>