МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

Кафедра системного анализа и управления

Н.А. ШУМИЛИНА
Р.Т. АБДРАШИТОВ
Т. Ш. НАСЫРОВА

РАЗРАБОТКА И ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПРИ ИССЛЕДОВАНИИ БИОЛОГИЧЕСКИХ ОБЪЕКТОВ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ЛАБОРАТОРНЫМ РАБОТАМ

Рекомендовано к изданию Редакционно–издательским советом
государственного образовательного учреждения
высшего профессионального образования
«Оренбургский государственный университет»

Оренбург 2006

УДК 510.67(076.5)
ББК 22.12 а73
Ш 96

Рецензент кандидат технических наук, доцент Денисов В.В.

Шумилина Н.А.
Разработка и использование математических моделей при иссле
Ш 96 довании биологических объектов: Методические указания к лабораторным работам/ Н.А. Шумилина., Р.Т.Абдрашитов, Т. Ш.Насырова,– Оренбург: ГОУ ОГУ, 2006 –19 с.

Практикум из двух лабораторных работ позволяет изучить основные этапы исследования и моделирования биологических систем, проведенные в диссертационной работе Насыровой Т.Ш.

Лабораторные работы включает: основные теоретические положения, рекомендации по этапам работ, а также контрольные вопросы для самопроверки.

Лабораторные работы предназначены для закрепления знаний по разделу «Построение моделей систем» курса «Теории Автоматического управления» для направления 220100 « Системный анализ и управление» и специальности 200402 «Инженерное дело в медико-биологической практике» раздел «Управление в медико-биологических системах» , а также для приобретения навыков проведения научного исследования и использования современных компьютерных способов анализа данных.

Закрепление основ системного похода и анализа происходит на материале представляющий особый интерес для человека – оценка здоровья, выявление зависимости функциональных показателей от антропометрических параметров индивида. При этом прививаются научные основы самонаблюдения, навыки самооценки, закладываются предпосылки постоянного научного анализа индивидуального физического состояния.

ББК 22.12а73

Содержание

Введение

Рассмотрение совокупности характеристик биологических объектов (системы показателей) позволяет получить представление о рассматриваемой особенности предмета исследования. Дальнейшее исследование систематизированного набора показателей (базы данных) с помощью современных компьютерных методов и средств обработки и анализа данных позволяет/10/ выявить неизвестные ранее закономерности, которые собственно и являются целью всякого научного исследования.

В лабораторной работе изучаются основные этапы разработки и исследования биологических моделей при построении модели физического состояния группы студентов. Постановка задач исследования, использование для этого индивидуальных показателей физического состояния каждого студента, критический анализ этих исходных данных, оценка закономерностей и перспектив развития на основе принципа «обобщенной инвариантности» приводит к формированию научного, системного подхода при рассмотрении биологических объектов.

Стандартным методом исследования биологических систем является моделирование ее структуры и закономерностей поведения. Модели, по выражению Н.М. Амосова /5/ – это «система, отражающая другую систему – объект. При моделировании биологическая система ее энергетическая и информационно – управляющая компоненты рассматриваются совместно, как единая, целостная система, в которой сохранность и функционирование метаболической системы поддерживаются и направляются механизмами регуляции. В любой сложной биологической системе – многие тысячи и даже миллионы взаимозависимостей. Соответственно можно получить массу их значений. К сожалению, этого недостаточно для построения более или менее полной математической модели. Для этого не хватает данных. Н.М. Амосов /5/ предложил метод эвристического моделирования – промежуточную ступень к реальным моделям сложных систем. Суть метода в том, что создается модель объекта на основе гипотезы о его структуре и функциях. При этом используются имеющиеся в литературе количественные данные и, исходя из качественной гипотезы, путем предположений добавляются недостающие.

Типовой план эвристического моделирования /5/ включает:

1)формулирование цели работы или назначения модели;

2)выбор уровня модели, т.к. все сложные системы построены по иерархическому принципу, поэтому выбирается тот нижний структурный уровень, который обеспечивает требуемую степень изучения объекта;

3)формирование качественной гипотезы о структуре и функциях объекта;

4)построение блок–схемы объекта: элементы, подсистемы и связи определяются гипотезой и выбранным нижним уровнем структур;

5) выбор значимых переменных (ограничение числа связей).

Завершающим этапом является исследование, анализ рассматриваемой совокупности данных с учетом ограничений, задаваемых математической моделью объекта.

Работа реализована в 2–х частях.В первой части работы рассматриваются первые два пункта эвристического моделирования, которые можно назвать – «Разработка информационной модели биологического объекта». В ней рассматриваются принципы формирования информационной модели биологического объекта на примере описания физического состояния студентов антропометрическими и функциональными показателями.

Во второй – рассмотрены остальные этапы. Формирование качественной гипотезы о структуре и функциях объекта, построение блок–схемы объекта: элементы, подсистемы и связи. Выбор значимых переменных (ограничение числа связей) произведен по математической модели, полученной при многомерном регрессионном анализе базы данных о физическом состоянии группы студенток. Основные закономерности влияния антропометрических и функциональных параметров на физическое состояние проводится графическими методами. Второй этап выполнен с помощью пакета статистических программ системы «Statistica».

При математическом моделировании следует помнить – не разумно ожидать идеального соответствия модели и данных /27/. Во–первых, рассматриваемые структурные модели описывают линейные зависимости и только приближенно соответствуют реальным явлениям. Истинность многих статистических предположений, накладываемых на проверяемую модель, остается под большим вопросом. На практике исследователя интересует "Согласуется ли модель достаточно хорошо, чтобы быть полезной для практического использования и разумного объяснения структуры наблюдаемых данных?" Во–вторых, идеальное соответствие модели данным не обязательно означает, что модель верна. Не возможно доказать, что модель верна – умение доказывать правильность модели эквивалентно умению предсказывать будущее. Например, вы можете сказать "Если Джо – кошка, то у Джо есть усы". Однако, из того, что "У Джо есть усы" не следует, что Джо – кошка. Аналогично, вы можете сказать, что "если определенная причинная модель верна, то она согласуется с наблюдаемыми данными". Однако, модель, согласующаяся с данными, не обязательно является верной. Возможно, существует другая модель, которая ничуть не хуже согласуется с теми же данными/ 27/.

В лабораторных работах реализована небольшая часть возможностей статистического пакета STATISTICA. Возбудить интерес к изучению других возможностей статистических методов и стремление освоить современные способы научного обоснования принимаемых решений основная цель этих работ.

1Разработка информационной модели биологического объекта

1.1 Некоторые представления о биологических системах

Биологическая система – совокупность функционально связанных элементов, образующих целостный биологический объект. Биологические системы обладают высокой устойчивостью к внешним возмущениям; при воздействии факторов среды в них возникают процессы, направленные на уменьшение эффекта этих возмущений. Они сохраняет свою специфичность в изменяющихся условиях окружающей среды и в определенных пределах обеспечивает гомеостаз внутренней среды. Биологические системы являются открытыми системами, т. е. в процессе жизнедеятельности обмениваются со средой веществом и энергией. Как известно /1,12,17,18/, биологические объекты, относятся к самым сложным системам.

Центральным понятием является понятие система /3/. Под системой подразумевается взаимосвязанное множество элементов, физически или мысленно выделенное из окружения, функционирование которой имеет определенную цель в рамках рассматриваемой задачи. При этом понятие конкретной системы связано с вполне определенной целью, а конкретная цель связана с определенной системой. Сложная система – система обладающая, по крайней мере, одним из признаков /3/: разбиение на подсистемы, изучение которых в рамках поставленной задачи имеет содержательный характер; система функционирует в условиях неопределенности, и воздействия среды обуславливают случайный характер изменения ее параметров и структуры; система осуществляет целенаправленный выбор своего поведения.

Другим понятием системного подхода является понятие структуры (организации). Структура выражает собою взаимосоответствие, взаимообусловленность свойств элементов систем и ее целостных характеристик. В дальнейшем это ведет к идее иерархии, к идее уровней в строении и детерминации систем. Под иерархией, под иерархическими системами понимают системы /20/, которые состоят из ансамбля последовательно "вложенных" одна в другую взаимодействующих и взаимообусловливающих подсистем. Понятие «иерархия» включает в себя представления о субординации, о взаимном соподчинении подсистем и уровней /20/. Структуры и функции процессов на вышележащих уровнях надстраиваются над структурами и функциями на нижележащих уровнях, чем и определяется характер их иерархических отношений. Реализация структуры осуществляется при помощи связей. Специфика сложных объектов заключена в характере связей и отношений между ними /18/. Каждая подсистема, каждый из уровней представлены своими структурными единицами и процессами, на каждом действуют специфические закономерности и решаются свои задачи.

Связи систем, как правило, классифицируются как /3/: постоянные и переменные; необходимые и случайные; устойчивые и неустойчивые. По высказыванию В.А. Энгельгардта, "более высоко лежащий иерархический уровень оказывает направляющее воздействие на уровень нижележащего порядка,

т.е., на подчиненный уровень. Это воздействие проявляется в том, что подчиненный член иерархии приобретает новые свойства, отсутствовавшие у него в изолированном состоянии. Из совокупности этих свойств, возникших в результате образования новой целостности, складывается специфический образ целого /20/. Можно выделить две предельных формы иерархической организации систем – иерархия, основанная на принципе жесткой детерминации, и иерархия, базирующаяся на принципах идеи вероятности и случайности. Реальные системы, реальные иерархии лежат между этими двумя предельными формами. Иерархия представляет собой и жесткие, однозначно определяемые зависимости, и лабильные, подвижные взаимоотношения, характеризующиеся наличием независимости в отношениях, наличием внутренней активности и элементов самоорганизации на каждом из уровней.

В иерархических отношениях живых систем ведущее значение приобретают информационные взаимодействия. "...В биологических иерархиях, – отмечает В.А. Энгельгардт /20/, – подчиненность в преобладающем числе случаев выступает в форме контроля, при осуществлении которого важная роль принадлежит действию обратных связей... Ведущими началами в биологических иерархиях являются элементы координирования и кооперации, а не доминирования и подчиненности". Иерархия направлена на обеспечение устойчивого функционирования как всей системы в целом, так и процессов на каждом структурном уровне /20/. Устойчивое функционирование систем в целом поддерживает (и опирается на) "оптимальное" функционирование подсистем и элементов, и, обратно, сбои на нижележащих уровнях приводят к ограничению эффективности деятельности систем в целом и даже к их разрушению.

Известны три вида описания системы, взаимно дополняющие друг друга, формирующих цельное, разностороннее представление о системе /3/:

1)функциональное описание, содержащее назначение системы, место и время функционирования, взаимоотношение с другими системами;

2)морфологическое описание, используемое для характеристики состава элементов их свойств и связи между элементами;

3)информационное описание, содержащее сведения о требованиях к параметрическим характеристикам системы, возможности их обеспечения и закономерностей возникновения погрешностей и ошибок.

Процесс функционирования и развития сложных и многообразных динамических связей между различными системами отражается через взаимодействие. Взаимодействия подразделяются на [4]: слабые и сильные; разрушающие и созидающие; интенсивные и вялотекущие и др. Процесс взаимодействия многогранен. Функции, которые выполняет каждый компонент в составе данной системы, определяют функциональность объекта. Функциональность компонентов по отношению к системе должны носить целесообразный характер и согласовываться во времени и пространстве, формируя систему как единое целое.

В разных сферах практической деятельности системные представления вместе с их теоретическими основами использования получали разные назва-

ния /2/, /18/: в инженерной деятельности – «методы проектирования», «методы инженерного творчества», «системотехника»; в военных и экономических вопросах – «исследование операций»; в административном и политическом управлении – «системный подход», «политология», «футурология»; в прикладных научных исследованиях – «имитационное моделирование», «методология эксперимента», «эвристическое моделирование» и т.д.

Для различных типов биологических систем характерны различные энергетические и информационно – управляющие процессы, но в любой биологической системе обе компоненты тесно связаны между собой и играют существенную роль в поддержании ее существования, совместно обеспечивая сохранение стационарного неравновесного состояния всех подсистем и самой целостной биологической системы.

Для понимания сущности обмена веществ и энергии в живой клетке нужно учитывать ее энергетическое своеобразие. Все части клетки имеют примерно одинаковую температуру, т. е. клетка по существу изотермична. Различные части клетки мало отличаются и по давлению. Это значит, что клетки не способны использовать тепло в качестве источника энергии, т. к. работа при постоянном давлении может совершаться лишь при переходе тепла от более нагретой зоны к менее нагретой. Таким образом, живые клетки не похожи на обычные тепловые или электрические двигатели. Живую клетку можно рассматривать как изотермическую химическую машину /19/.

Живые организмы представляют собой открытые системы, так как они обмениваются с окружающей средой, как энергией, так и веществом, и при этом преобразуют и то, и другое /19/. Однако они не находятся в равновесии с окружающей средой и поэтому могут быть названы *неравновесными открытыми системами*. Тем не менее, при наблюдении в течение определенного отрезка времени в химическом составе организма видимых изменений не происходит. Но это не значит, что химические вещества, составляющие организм, не подвергаются никаким превращениям. Напротив, они постоянно и достаточно интенсивно обновляются. Кажущееся постоянство химического состава объясняется, так называемым стационарным состоянием, т. е. таким положением вещей, при котором скорость переноса вещества и энергии из среды в систему точно уравновешивается скоростью переноса из системы в среду. Таким образом, живая клетка представляет собой неравновесную открытую стационарную систему. Одним из направлений развития биологической системы является эволюция.

И.С. Моросанов /14/ отметил, что «эволюция – не пассивный естественный отбор из независимых случайностей, как представляли Дарвин и Пригожин, а активный процесс (самопроизвольный и, следовательно, неизбежный) направленный в сторону иерархического симбиоза под контролем внешней и внутренней сред». Окружающая среда сама по себе не может изменить генетическую информацию. Однако, если внешние условия таковы, что возникает неравновероятность конечных состояний, то должна появиться перенормировка, то есть появляется мутация – равновесный результат неравновесного про-

цесса восстановления повреждений, вызванных нарушением баланса энергии в системе.

Фундаментальным методологическим приемом исследования биологических систем является принцип системной организованности , /17/, согласно которому любой биологический объект представляет собой биологическую систему, способную к регулированию как внутренних соотношений между своими подсистемами, так и соотношений целостного объекта со средой. Изучение биологических объектов в таких методологических рамках представляет собой системный анализ биологических систем. При системном анализе биологический объект представляется в виде биологической системы – множества функционально связанных элементов, реализующих энергетические и информационные процессы системы. Совокупность существенных связей между этими элементами биологическими системами определяет ее структуру /6/.

1.2 Информационная модель физического состояния студентов

В работе/15/, целью является выявление закономерностей воздействия морфологических и функциональных показателей на физическое состояние при двигательной нагрузке. Для описания модели принят метод оценки физического состояния по доступным для самоконтроля антропометрическим параметрам и функциональным характеристикам сердечно–сосудистой, дыхательной и мышечной систем.

На основе литературных источников сформирована гипотетическая структура и построена блок–схема (модель индивидуальной изменчивости) объекта: с указанием элементов и подсистем, рисунок 1.

Рисунок 1 – Морфологическая структура объекта, характеризующая индивидуальность человека

Различные возможности тканей в обмене энергии, варианты сочетаний состояний подсистем являются биологической основой организации и связей, представляющих самоорганизующие возможности биологических систем.

Бланк, разработанной системы данных, был использован для сбора информации. Выборка данных из этих бланков по 33 переменным позволила сформировать матрицу размером 33 х 75. Лабораторная версия этой базы данных используется в примере (fiz – DATE – laborat. STA 21v x 75).

В таблице 1 приведены исследуемые переменные, характеризующие антропометрические и функциональные параметры испытуемой группы студентов. Для определения значений этих переменных не требуются сложные и специальные средства измерения. Необходимы – сантиметровая лента, эспандер и скоба для определения толщины кожной складки.

Таблица 1 – Переменные, используемые для формирования базы данных

Код	Наименование переменной
1	2
1 ROST	Рост - в момент заполнения, см
2 VES	Вес в момент заполнения, кг
3 O_TALII	Окружность талии, в см
4 O_GRUDI	Окружность груди, в см
5 O_BEDER	Окружность бедер, см
6 O_ZAPJAS	Окружность запястья, см
7 O_GOLENI	Окружность голени, см
8 O_BICEPS	Окружность бицепса, см
9 O_IKRON	Окружность икроножной мышцы, см
10 OK_BEDRA	Окружность одного бедра,см
11 O_KOLENI	Окружность сомкнутых коленей, см
12 O_GOLOVI	Окружность головы, см
13 SL_HVAT	Сила хвата(левой), в кг

Продолжение таблицы 1

1	2
14 SP_HVAT	Сила хвата(правой), в кг
15 TS_LOPAT	Толщина кожной складки в области угла лопатки, см
16 TS_Z_PLE	Толщина кожной складки на 1/3 задней поверхности плеча
17 TS_P_PLE	Толщина кожной складки на передней поверхности плеча, см
18 TS_Z_P_P	Толщина кожной складки на задней поверхности предплечья, см
19 TS_P_P_P	Толщина кожной складки на передней стороне предплечья, см
20 TS_Z_GOL	Толщина кожной складки на задней стороне предплечья, см
21 TS_P_GOL	Толщина кожной складки на передней поверхности голени
22 TS_P_BED	Толщина кожной складки на передней поверхности бедра, см
23 TS_N_JGO	Толщина нижней ягодичной кожной складки, см
24 TS_T_KIS	Толщина кожной складки на тыльной стороне кисти, см
25 TS_JIVOT	Толщина кожной складки на нижней части живота, см
26 CHAST_DI	Частота дыхания, раз/в мин
27 ZAD_VDOH	Задержка дыхания на вдохе, с
28 ZAD_VIDO	Задержка дыхания на выдохе, с
29 PULS_POK	Пульс в покое (до занятий)
30 PULS_NAG	Пульс после нагрузки (после занятий)
31 T_PRIJKI	Тест -60 прыжков (со скакалкой в мах темпе) - частота пульса в мин
32 T_PRISED	Тест -20 приседаний. Частота пульса после приседаний за 1 мин.
33 T_NAKLON	Тест -30 наклонов. Частота пульса после наклонов в 1 мин.

1.3 Методика выполнения первой части работы

1.3.1 Проводим сбор данных (замер и запись в индивидуальный бланк каждым студентом группы – приложение А) по основным:

–антропометрическим параметрам (ROST, VES, O_TALII, O_GRUDI, O_BEDER, O_ZAPJAS, O_BICEPS);

–силовым характеристикам кистей рук (SL_HVAT, SP_HVAT);

–толщине кожных складок (TS_Z_PLE, TS_P_BED, TS_T_KIS, TS_JIV-OT);

–характеристикам дыхательной системы (CHAST_DI, ZAD_VDOH, ZAD_VIDO);

–характеристике сердечно–сосудистой системы в покое (PULS_POK);

–характеристикам сердечно–сосудистой системы при различной нагрузке (T_PRIJKI, T_PRISED, T_NAKLON)

–характеристике сердечно–сосудистой системы после окончания занятий (PULS_NAG) – не менее 15 – 20 мин.

Полученные значения данных всей группы студентов (10–15 чел) вносим в исходную таблицу статистического пакета «Statistica» (Noname.sta) и создаем файл с именем соответствующим шифру группы.

1.4 Контрольные вопросы

1 Каково определение «системы» по Амосову?

2 Почему моделирование биологических систем представляет собой проблему?

3 В чем идея «эвристического» моделирования Н.М. Амосова?

4 Какие основные этапы «эвристического» моделирования вы можете назвать?

5 Почему невозможно идеальное соответствие математической модели имеющимся данным?

6 Как ставится вопрос приемлемости модели при моделировании сложных объектов?

7 Какие основные характеристики входят в понятия «сложная система», «биологическая система»?

8 Какие основные понятия определяют «системный подход»?

9 Назовите виды описания систем?

10 Назовите иерархическую структуру категорий, определяющих сложную систему?

11 Сколько вариантов названий системного подхода вы можете назвать?

12 Какие особенности биологических объектов позволяют классифицировать их как открытые системы и отличать живые клетки от тепловых или электрических машин?

13 Каков механизм протекания мутации?

2 Разработка математических моделей биологических объектов и их использование при анализе данных

В практической работе рассмотрены этапы, включающие/5/:

–формирование качественной гипотезы о структуре и функциях объекта;

–построение блок–схемы объекта: элементы, подсистемы и связи

–выбор значимых переменных (ограничение числа связей).

Этапы выполнены на основе многомерной, многофакторной регрессионной модели объекта. В основе методов многофакторного анализа лежит гипотеза /7/, /10/, /26/: наблюдаемые параметры являются лишь косвенными характеристиками изучаемого, существуют скрытые, внутренние (латентные, фундаментальные) параметры или свойства, число которых мало и которые определяют значения наблюдаемых признаков, при этом скрытые параметры называются факторами. Таким образом, математической моделью является регрессионная зависимость, которая качественно описывает всю совокупность элементов (представляет структуру), связывает их функции с функцией всей системы, ограничивает число связей (элементов), т.к. в регрессионное уравнение не включаются малозначимые элементы. Получение регрессионного уравнения позволяет решить все задачи моделирования.

Использование многомерных методов, позволяет исследовать различные формы ассоциации (близости, связи, подобия) между несколькими разнородными переменными и/или объектами /2/, /10, /26/. Многомерными зависимости можно исследовать методами факторного, кластерного, дискриминантного и дисперсионного анализа. Многочисленность методов многомерного анализа/10/, обусловлена: 1) различиями в постановке задач; 2) пригодностью того или иного метода для решения поставленных задач, в зависимости от свойств анализируемых данных.

Целью практической работы является изучение методов:

– построения математических моделей – регрессионных зависимостей функциональных характеристик (характеристик работы сердечно–сосудистой системы) от антропометрических и ресурсных параметров организма;

- выявления наиболее достоверных зависимостей характеристик работы сердечно–сосудистой системы от антропометрических и ресурсных параметров организма.

2.1 Основные положения разработки математической модели

Известно использование двух методов разработки регрессионной модели /1/, /2/, /10/, /26/: методом главных компонент, основан на информативности дисперсии системы признаков, и факторным анализом, объясняющем имеющиеся между признаками корреляции. Таким образом, с помощью этих методов решается проблема «сжатия» информации /7/, /13/, /22/, /25/, /26/, /27/, содержащейся в данных и вычленения «существенной» информации, которая затемнена и искажена разного рода данными, не имеющими отношения к сути изучаемого явления. Наблюдая большое число измеряемых параметров, выяв-

ляют небольшое число факторов, находящихся на втором плане, которые в основном определяют поведение измеряемых параметров /9/. В большинстве случаев эти два метода приводят к весьма близким результатам. Однако анализ главных компонент часто более предпочтителен как метод сокращения данных, в то время как анализ главных факторов лучше применять с целью определения структуры данных /26/.

Основным показателем качества регрессионного уравнения является коэффициент детерминации R^2, объясняющий какая доля дисперсии (разброса) объясняется построенной функцией регрессии и коэффициент детерминации, скорректированный с учетом степеней свободы (adjusted R–squared).

Задача отбора переменных в уравнении множественной регрессии может решаться методами последовательного увеличения или последовательного уменьшения их числа. Чем меньше количество переменных, тем легче интерпретировать содержание регрессионных моделей

(Подробнее о системе STATISTICA и регрессионном анализе см Приложение Б)

2.2 Методы анализа и интерпретация данных после моделирования

Анализ и предметная интерпретация полученной системы факторов должна пролить свет на внутреннюю природу исследуемой системы объектов или явлений. При этом содержательная интерпретация новых факторов в предметных (физических) терминах является творческой, неформальной задачей исследователя.

Чаще всего для анализа данных используют статистические методы исследования. Зародившиеся в XVII-XVIII веках методы до 90–х годов XX века использовалась как традиционная математическая статистика, основывающаяся на концепции усреднения по выборке, приводящей к операциям над фиктивными величинами («средняя температура в больнице»). В настоящее время бурно развивается направление исследований называемых «Data mining» – «раскопка» знаний. Подробнее об этом направлении в источнике / /.

Рассмотрим использование статистических методов разведочного анализа данных, позволяющих визуально и численно исследовать структуру данных.

Эффективным методом исследования структуры данных является, так называемый, «анализ соответствия». Метод анализа разработан в конце 1960-х годов. Существуют другими названия: «оптимальное шкалирование», «взаимное усреднение», «оптимальная оцифровка», «квантификационный метод» или «анализ однородности». Таким образом решается главная цель – представление в упрощенном виде (пространстве меньшей размерности) информации, содержащейся в больших разнородных массивах. Для этого предварительно проводят разделение т.е. «категоризацию» непрерывной последовательности значений данных. Идея состоит в том, чтобы разбить множество значений разнородных наблюдений на однородные группы /7/, по каким–либо, осмысленным исследователем, признакам.

Категоризацию (разделение), как правило, проводят на 2 подгруппы (по принципу – признак: есть или нет) или на 3 подгруппы (значение признака – высокое, среднее или низкое). В принципе возможно разделение на любое число подгрупп, например, по шкале 10 бальной (100 бальной) оценки. Чем больше подгрупп, тем сложнее интерпретировать табличные или графические результаты применяемого метода. Исследуются парные взаимовлияния переменных. Число сравниваемых пар 6 (шесть) – процедура «Crosstabulation & Stub-and-Banner Tables» в меню «Basic Statistics and Tables»/22 / подробнее см. Приложение Б.

В некоторых случаях, для некоторых наборов данных даже выявление возможности группирования различных данных, полученных в исследовании, подсказывают направление дальнейшего анализа. Для этого используют методы кластерного и дискриминантного анализа.

Кластерный анализ, предназначен для группирования множества объектов на заданное или неизвестное число классов на основании некоторого математического критерия (в программном пакете около 10 критериев). Кластерный анализ организует наблюдаемые данные в наглядные структуры. В общем, всякий раз, когда необходимо классифицировать информацию к пригодным для дальнейшей обработки группам, кластерный анализ оказывается весьма полезным и эффективным /10/, /22/, . В связи с тем, что кластерный анализ не содержит вычислительного механизма проверки гипотезы об адекватности получаемых классификаций, результаты кластеризации обосновываются с помощью дискриминантного анализа.

Дискриминантный анализ используется для выявления закономерности взаимодействия при нелинейных связях критериального показателя и исходных показателей (признаков, параметров) и состоит в том, чтобы на основе анализа комплекса характеристик классифицировать (разделить) на заданное или неопределенное число групп некоторым оптимальным способом. Под оптимальным способом подразумевается либо минимум потерь (математическое ожидание потерь) или минимум вероятности ложной классификации

Одним из направлений развития кластерного анализа является иерархическое группирование позволяющее получить наглядное представление о структуре всей исследуемой совокупности объектов в виде дерева (дендрограммы). Это отдельная процедура, включается в структуру кластерного анализа.

Различные методы обработки, включая и перечисленные представлены в различных компьютерных статистических пакетах программ /10/, /22/. Достаточно широкое распространение получил пакет «Statistica» – хорошо сбалансированный пакет по соотношению «мощность/удобство», с широким спектром функциональных алгоритмов, достаточно автономных и связанных в единый блок из 25 модулей – в работе рассмотрены только небольшая часть методов анализа.

Сильной стороной является графика пакета«Statistica», двумерного (типа 2D) и – трехмерного (3D). Развитие компьютерных методов обработки

данных, привело к выделению визуальных методов обработки в отдельное, эффективное направление исследования зависимостей /10/. Возможность использования различных методов аппроксимации поверхностей, позволяет оценить наличие отклонений от прогнозируемого поведения, при этом метод не имеет ограничений, кроме невозможности графического представления многофакторных (более трех) переменных. Графики поверхности используются для анализа данных, для наглядного представления результатов анализа, позволяющего обнаружить сложные, динамичные нелинейные взаимозависимости между переменными.

Они позволяет определить минимизацию значений и значения близкие к экстремуму, линии уровня поверхности показывают расслоение значений переменных, интенсивность цветов показывает значения переменных (при сравнении с цветовой меткой). Графические методы, как правило, подтверждают обоснованность вывода, позволяют определить динамику процесса.

2.3 Методика выполнения работы

Работа выполняется в 2 этапа.

На первом – разрабатывается математическая модель зависимости функциональных характеристик человека от антропометрических и ресурсных параметров. Эта часть работы выполняется по матрице данных о физическом состоянии по анкетам (форма – Приложение А) конкретной группы, введенных в программный статистический пакет (Statistika, Statgraf и др.) в ходе выполнения работы 1, или используется готовый блок данных, подготовленный по анкетам другой группы студентов.

При этом рекомендуется следующая последовательность решения задач:

1) разработать математическую модель зависимости ЧСС от антропометрических и ресурсных параметров группы студентов используя процедуру компьютерного пакета «Statistica» – многомерный, многофакторный регрессионный анализ («Multiple Regression» – для английской версии пакета). При этом исследуемой выходной функциональной характеристикой может быть – ЧСС после прыжков, ЧСС после приседаний или ЧСС после наклонов, а также ЧСС после окончания занятий

2) оценить: достоверность модели, определить значимые факторы, ранговые характеристики параметров, влияющих на физическое состояние студентов.

На втором этапе – описать основные закономерности влияния антропометрических и функциональных параметров на физическое состояние. Для этого по ранговым характеристикам выбрать наиболее значимые параметры. После этого используя процедуру «Stats 3D Sequential Graphs» в графическом меню «Graphs Gallery» – для английской версии пакета, визуализировать (получить график парных влияний на исследуемую выходную характеристику) основных антропометрических и ресурсных переменных.

–двумерный визуальный анализ данных (гистограммы, таблицы частот, сопряженностей, диаграммы диапазонов, размаха);

–трехмерный визуальный анализ данных (графики поверхностей 3-х переменных)

Полученные графики позволяют описать изменение переменных, их взаимозависимость и влияние на ЧСС.

2.4 Содержание отчета

В отчет включается:

–регрессионное уравнение, описывающее зависимость ЧСС от антропометрических и ресурсных параметров

–ранговые характеристики значимых переменных,

– 2–3 графика поверхностей 3-х переменных.

- Описание 2–3 зависимостей.

2.5 Контрольные вопросы

1 Какая проблема решается при использовании статистических методов исследования?

2 Какие методы разработки регрессионных моделей Вам известны?

3 Почему содержательная интерпретация математической модели наиболее сложная часть исследования?

4 В чем перспективность метода «Data mining» – «раскопка» знаний?

5 Как проводится «анализ однородности» данных?

6 Назовите назначение и методы кластерного анализа?

7 Назовите неазначение дискриминантного анеализа?

8 Назовите преимущество и недостатки графических методов анализа?

Список использованных источников

1 Абдрашитов, Р. Т. О синтезе систем управления сложными системами. Социокультурная динамика региона. Наука. Культура. Образование: материалы конференции. Ч.1. / Р. Т. Абдрашитов. Оренбург : ИПК ОГУ, 2000.-35-44с.

2 Абдрашитов, Р.Т.. Компьютерная поддержка анализа и прогнозирования социально-экономической и экологической ситуации в регионах/ И.В. Влацкая, В.В.Каштанов [и др] Оренбург: ОГУ, 1996. – 19с.

3 Аверьянов, А.Н. Системное познание мира: методологические проблемы / **А.Н. Аверьянов** -М.: Политиздат,1985.-263с

4 Александров, В.В. Анализ данных на ЭВМ. [Текст]/ В.В. Александров -М.: Финансы и статистика, 1990. – 192с.

5 Амосов, Н.М. Энциклопедия Амосова. Алгоритм здоровья [Текст]/ Н.М.Амосов- М.: Издательство АСТ; Донецк: «Сталкер», 2002.– 590с.

6 Анохин, П.К. Избранные труды. Философский аспект теории функциональной системы[Текст]/ П.К. Анохин- М.: Наука, 1978.-400с.

7 Боровиков, В.П. Популярное введение в программу STATISTICA. [Текст]/ В.П Боровиков- М.: Компьютер Пресс, 1998. - 267 с.

8 Боровиков, В.П. STATISTICA: Искусство анализа данных на компьютере. Для профессионалов./ В.П. Боровиков -СПб.: Питер, 2001.–656с.

9 Браверман, Э. М. Структурные методы обработки эмпирическихданных. [Текст]/ Э. М. Браверман, И. Б. Мучник— М.: Наука. 1983.—464 с.

10 Дюк, В.. Data mining: учебный курс [Текст]/ Самойленко А. – СПб: Питер, 2001. – 368с.

11 Елисеева, И.И. Логика прикладного статистического анализа. [Текст]/ И.И.Елисеева, Рукавишников В.О. - М:. Финансы и статистика. 1982.– 192с

12 Ивахненко, А.Г. Самообучающиеся системы распознавания и автоматического управления. [Текст]/ А.Г. Ивахненко- К.1969Техника,.–392с

13 Кимбл, Г. Как правильно пользоваться статистикой. [Т.Кимбл]/ Пер с англ. Б.И.Клименко.- М. Финансы и статистика,1982.-294с.

14 Моросанов, И.С. Эволюционная концепция теории систем. Системные исследования. Методологические проблемы. Ежегодник 1998. Ч.1 с.33–43. [Текст]/ Под ред. Д.М.Гришиани [и др.] М.:Эдиториал УРСС, 1999.– 360с.

15 Новосельцев, В.Н. Организм как элемент большой системы (Проблемы моделирования). В материалах конференции «Управление большими системами» [Текст]/ Под. ред В.Н. Буркова, Д.А. Новикова. (Серия «Информатизация России на пороге 21 века») – М.: СИНТЕГ, 1997. – с 295.

16 Перегудов Ф.И. Основы системного анализа [Текст]/ Ф.П.Тарасенко -Томск: Изд-во «НТЛ», 1997. – 396 с.

17 Розенгарт, В.И. Обмен веществ и энергии. [Текст]/ В.И. Розенгарт**,** Заремба̀ский Р.А., Фролкис В.В. Обзорная статья. Большая медицинская энциклопедия – М.: Советская энциклопедия.Т.17, С.121 – 130.

18 Сачков, Ю.В. Вероятностная революция в науке (Вероятность, случайность, независимость, иерархия). [Текст]/ Ю.В. Сачков– М.: Научный мир, 1999. – 144 с.

19 Спицнадель, В.Н. Основы системного анализа[Текст]: учебное пособие. / В.Н.Спицнадель – СПб.: Изд. дом. «Бизнес – пресса», 2000. – 326с.

20 Кулаичев, А. Статистический пакет «Стадия» [Текст]/ - М.: Мир ПК" 1/95.

21 под ред. Енюкова, И.С. .Факторный, дискриминантный и кластерный анализ. [Текст]/ пер с англ. - М:.Финансы и статистика, 1989.- 215с.

22 Учебник по промышленной статистике. [Электронный ресурс]/ - Режим доступа:.http://www.StatSoft.ru.

Приложение А (справочное)

А.1 Некоторые страницы «help» – файла системы STATISTICA

А.1.1 Окно «Таблица исходных данных»

Система STATISTICA хранит данные на диске в файлах, которые оптимизированы для быстрого доступа и эффективного хранения. Общая организация этих файлов схожа с той, которая используется программами управления базами данных; проще всего представить себе каждый файл как "таблицу", в которой строки представляют собой записи (например, пациенты, предметы, модели) и каждая запись состоит из одного и того же числа полей (например, пяти типов данных [измерений], известных о каждом пациенте).

Данные системы STATISTICA делятся на наблюдения и переменные. Если такая запись вам не знакома, можно считать наблюдения эквивалентом записей в программе управления базами данных (или строками электронной таблицы), а переменные - эквивалентом полей (столбцов электронной таблицы). Каждое наблюдение состоит из набора значений переменных.

Первый столбец в файле может содержать имена наблюдений (этот параметр не является обязательным. Имена наблюдений используются только как метки (а не текстовые значения), поэтому они могут содержать любые печатаемые символы (также как и метки значений). Это диалоговое окно можно вызвать, выбрав пункт «Имена наблюдений» из меню «Правка» или дважды щелкнув на имени наблюдения, которое нужно изменить.

Кнопка «Параметры». Эта кнопка открывает диалоговое окно «Параметры» имен наблюдений, в котором можно изменить длину имени наблюдения, получить имена наблюдений из другой переменной или произвести с именами наблюдений какие-либо операции через буфер обмена.

1.2 Двойная запись значений

Система STATISTICA поддерживает двойную запись значений, в которой каждое значение конкретной переменной может одновременно иметь и текстовую и числовую запись. Переключаться между отображением текстовых или числовых значений или меток значений (если они есть) можно с помощью кнопки «ABC». Текстовым значениям в формате "двойной записи", система STATISTICA присваивает числовые эквиваленты. Поэтому такие переменные можно использовать во всех процедурах статистического анализа данных, как если бы они были числовыми.

1.3 Рабочая книга

Каждый файл данных системы STATISTICA может также содержать следующую текстовую информацию:

одну строку информации (отображается в области заголовка таблицы исходных данных), которую можно использовать для пометки отчетов,

многострочные комментарии или примечания о файле (в виде абзаца текста), а также список имен дополнительных файлов (например, графиков,

таблиц результатов, текстовых/графических отчетов, групп или программ для преобразования данных), связанных с текущим набором данных (см. информацию по Рабочим книгам).

Они доступны по двойному щелчку на области заголовка таблицы исходных данных (или по нажатию кнопки «Рабочая книга» панели инструментов, а также при выборе пункта «Заголовок» выпадающего меню Правка).

Файлы данных системы STATISTICA (расширение *.sta) могут рассматриваться как "Рабочие книги" файлов, поскольку они содержат (автоматически сохраняют) информацию обо всех дополнительных файлах (например, графиках, отчетах, программах), которые используются с текущим набором данных. Список этих файлов можно посмотреть, нажав кнопку «Рабочая книга» панели инструментов или дважды щелкнув на области названия (заголовке) таблицы исходных данных.

1.4 Разделение прокрутки

Таблицы исходных данных можно разделять на две или четыре секции (окна) перетаскиванием полосы разделения (черного прямоугольника в верхней части вертикальной полосы прокрутки или в левой части горизонтальной полосы прокрутки). Это может оказаться полезным, если таблица исходных данных большая и нужно посмотреть результаты в разных частях таблицы. Если подвести курсор мыши к полосе разделения, появляются значки разделения. Далее, чтобы установить разделение, нужно нажать левую кнопку мыши и перетащить полосу на нужное место. Чтобы изменить положение разделения, нужно переместить на новое место полосу разделения (которая теперь расположена между окнами). Обратите внимание, что окна, разделенные по вертикали, прокручиваются вместе при вертикальном прокручивании, а окна, разделенные по горизонтали - при горизонтальном прокручивании.

2 Модули системы STATISTICA

(список – в алфавитном порядке)
1 Анализ выживаемости
2 Анализ процессов
3 Анализ соответствий
4 Временные ряды и прогнозирование
5 Деревья классификации
6 Дискриминантный анализ
7 Дисперсионный анализ
8 Добыча данных (Data Mining)
9 Канонический анализ
10 Кластерный анализ
11 Компоненты дисперсии
12 Контроль качества
13 Логлинейный анализ
14 Менеджер мегафайлов
15 Множественная регрессия

16 Многомерное шкалирование
17 Моделирование структурными уравнениями
18 Надежность и позиционный анализ
19 Нелинейное оценивание
20 Непараметрическая статистика
21 Нейронные сети
22 Основные статистики и таблицы
23 Планирование эксперимента
24 Сервер файлов системы STATISTICA
25 Управление данными
26 Факторный анализ
27 Элементарные понятия

3 Основные статистики и таблицы

3.1 Описательные статистики

Описательные статистики вычисляются в диалоговом окне «Основные статистики и таблицы» отдельно для каждой переменной в файле данных и обеспечивают исследователя:

–основной описательной информацией относительно распределения переменной (среднее, минимальное и максимальное значение, мода, медиана)

–различные меры изменчивости – вариабельности (дисперсия, среднее отклонение, стандартное отклонение, стандартную ошибку),

–характеристики формы распределения (асимметрия, эксцесс).

Кроме того, доступны различные критерии нормальности.

Статистики вычисляются для каждой переменной.

Для вычисления описательных статистик нужно выбрать переменные в текущем файле данных (переменные можно выбрать щелчком кнопки). STATISTICA построит таблицу описательных статистик, расположив их в отдельной строке для каждой переменной.

3.2 Процентили

Процентиль распределения (термин введен Галтоном в 1885г.) - это число x_p, значения *p*-й части совокупности меньше или равны x_p. Например, 25-я процентиль (называется квантиль 0.25 или нижняя квартиль) переменной - это такое значение (x_p), что 25% значений переменной (p) попадают ниже этого значения. Аналогичным образом вычисляется 75-я процентиль (квантиль 0.75 или верхняя квартиль) - значение, ниже которого попадают 75% значений переменной.

4 Исследование зависимости переменных

При различных видах статистического анализа обычно различают независимые и зависимые переменные. Зависимые переменные, это те переменные, поведение которых исследователь пытается "объяснить", то есть исследователь предполагает, что эти переменные зависят от независимых переменных

и хочет эту зависимость (связь) оценить. Выявление и оценка степени этих зависимостей и является целью любого статистического исследования. Если исследуется зависимость переменной от другой(их) одной природы, то исследование одномерное, а если от переменных различной природы, то многомерное.

4.1 Разведочный анализ данных

Разведочный анализ данных (РАД) применяется для нахождения систематических связей между переменными в ситуациях, когда отсутствуют (или имеются недостаточные) априорные представления о природе этих связей. Как правило, при разведочном анализе учитывается и сравнивается большое число переменных, при этом для поиска закономерностей используются самые разные методы.

4.2 Основные методы разведочного статистического анализа.

К основным методам разведочного статистического анализа относится процедура анализа распределений переменных (например, чтобы выявить сильно несимметричные или не нормально распределенные, в том числе, бимодальные переменные), просмотр больших корреляционных матриц на предмет отыскания коэффициентов, превосходящих по величине определенные пороговые значения, и анализ многовходовых таблиц частот (например, последовательный "послойный" анализ комбинаций уровней управляющих переменных) и методы многомерного разведочного анализа

4.3 Методы многомерного разведочного анализа.

Методы многомерного разведочного анализа специально разработаны для поиска закономерностей в многомерных данных (или последовательностях одномерных данных). К ним относятся: кластерный анализ, факторный анализ, дискриминантный анализ, многомерное шкалирование, логлинейный анализ, каноническая корреляция, пошаговая линейная и нелинейная (например, логит) регрессия, анализ соответствий, анализ временных рядов и деревья классификации.

4.4 Кросстабуляция и многомерный анализ соответствий

Эффективным методом разведочного анализа данных является кросстабуляция входящий в модуль «Основные статистики и таблицы» позволяющий анализировать и получать графическое отображение состояния данных в различных типах многомерных таблиц. При этом непрерывную цепочку данных приходится разбивать на некоторое число групп (табулировать). Сравнительный анализ таких групп лежит в основе метода.

Например, в медицине можно табулировать частоты различных симптомов заболевания по возрасту и полу пациентов; в области образования можно табулировать число учащихся, покинувших среднюю школу в зависимости от возраста, пола и этнического происхождения; экономист может табулировать число банкротств в зависимости от вида промышленности, региона и начального капитала; исследователь спроса может табулировать предпочтения потребителя в зависимости от вида товара, возраста и пола и т.д. Одним из способов

классификации (табулирования) является использование процентилей распределения (см п. 3.2).

Во всех этих случаях результаты представляются в виде многовходовых (многомерных) таблиц частот, то есть в виде таблиц сопряженности с двумя или более факторами. STATISTICA включает модуль специально предназначенный для проведения анализа соответствий двухвходовых и многовходовых таблиц частот (т.е. для выполнения многомерного анализа соответствий).

Анализ соответствий является описательным / разведочным методом, созданным для анализа сложных таблиц, содержащих некоторые меры соответствий между переменными - столбцами и переменными - строками. Получаемые результаты содержат информацию, похожую по своей природе на результаты Факторного анализа. Они позволяют изучить структуру категориальных переменных, включенных в таблицу.

Более глубокие методы исследования этих таблиц в системе STATISTICA проводится в модуле «Логлинейный анализ». Термин логлинейный (или логарифмически-линейный) происходит из-за того, что с помощью логарифмического преобразования можно переформулировать задачу анализа многомерных таблиц частот в терминах дисперсионного анализа. В частности, многовходовую таблицу частот можно рассматривать как отражение различных главных и взаимодействующих влияний, которые складываются вместе линейным образом.

4.5 Кластерный анализ

Кластерный анализ включает: алгоритм древовидной кластеризации, метода двувходового объединения и метода «К средних.»

Принципы объединения и анализа древовидной кластеризации заключается в том, что диаграмма начинается с каждого объекта в классе (слева на право или снизу вверх диаграммы). Представим себе, что постепенно (очень малыми шагами) вы ослабляете ваш критерий о том, какие объекты являются уникальными (самостоятельными, независимыми), а какие нет. Другими словами, вы понижаете порог, относящийся к решению об объединении двух или более объектов в один кластер. В результате, вы связываете вместе всё большее и большее число объектов и агрегируете (объединяете) все больше и больше кластеров, состоящих из все сильнее различающихся элементов. Окончательно, на последнем шаге все объекты объединяются вместе.

На этих диаграммах горизонтальные оси представляют расстояние объединения (в вертикальных древовидных диаграммах вертикальные оси представляют расстояние объединения). Так, для каждого узла в графе (там, где формируется новый кластер) вы можете видеть величину расстояния, для которого соответствующие элементы связываются в новый единственный кластер. Когда данные имеют ясную структуру в терминах кластеров объектов, сходных между собой, тогда эта структура, скорее всего, должна быть отражена в иерархическом дереве различными ветвями. В результате успешного анализа методом объединения появляется возможность обнаружить кластеры (ветви) и интерпретировать их.

Принцип двувходового объединения заключается в двунаправленноой одновременной кластеризации наблюдений и переменных.

Принцип классификации на заданное число кластеров заключается в задании числа кластеров при условии «чтобы они были настолько различны, насколько это возможно».

4.6 Дискриминантный анализ

Дискриминантный анализ - это очень полезный инструмент (1) для поиска переменных, позволяющих относить наблюдаемые объекты в одну или несколько реально наблюдаемых групп; (2) для классификации наблюдений в различные группы.

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие (или определенные априори) совокупности (группы).

В дискриминантном анализе, как правило, одновременно рассматривается более одной независимой переменной и определяются типы (классы) значений этих переменных. Говоря несколько технически, в дискриминантном анализе находят такие линейные комбинации зависимых переменных, которые наилучшим образом определяют принадлежность наблюдения к определенному классу, причем число классов известно заранее.

Другой главной целью применения Дискриминантного анализа является проведение классификации. Как только модель установлена и получены дискриминирующие функции, возникает вопрос о том, как хорошо они могут предсказывать, к какой совокупности принадлежит конкретный образец?

Модуль «Дискриминантный анализ» автоматически вычисляет функции классификации. Последние не следует путать с дискриминирующими функциями. Функции классификации предназначены для определения того, к какой группе наиболее вероятно может быть отнесен каждый объект. Имеется столько же функций классификации, сколько групп. Каждая функция позволяет вам для каждого образца и для каждой совокупности вычислить Веса классификации.

Как только вы вычислили показатели классификации для наблюдения, легко решить, как производить классификацию наблюдений. В общем случае наблюдение считается принадлежащим той совокупности, для которой получен наивысший показатель классификации.

4.7 Множественная регрессия

В общественных и естественных науках процедуры множественной регрессии чрезвычайно широко используются в исследованиях. Общее назначение множественной регрессии (этот термин был впервые использован в работе Пирсона - Pearson, 1908) состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной.

Специалисты по кадрам обычно используют процедуры множественной регрессии для определения вознаграждения адекватного выполненной работе.

Можно определить некоторое количество факторов или параметров, таких, как "размер ответственности" (Resp) или "число подчиненных" (No_Super), которые, как ожидается, оказывают влияние на стоимость работы. Кадровый аналитик затем проводит исследование размеров окладов (Salary) среди сравнимых компаний на рынке, записывая размер жалования и соответствующие характеристики (т.е. значения параметров) по различным позициям. Эта информация может быть использована при анализе с помощью множественной регрессии для построения регрессионного уравнения в следующем виде:

Salary =0.5*Resp + 0.8*No_Super (1)

После того как, так называемая, линия регрессии определена, аналитик оказывается в состоянии построить график ожидаемой (предсказанной) оплаты труда и реальных обязательств компании по выплате жалования. Таким образом, аналитик может определить, какие позиции недооценены (лежат ниже линии регрессии), какие оплачиваются слишком высоко (лежат выше линии регрессии), а какие оплачены адекватно.

Общая вычислительная задача, которую требуется решать при анализе методом множественной регрессии, состоит в подгонке прямой линии к некоторому набору точек. В простейшем случае, когда имеется одна зависимая и одна независимая переменная, это можно увидеть на диаграмме рассеяния. (диаграммы рассеяния можно автоматически построить для таблиц результатов, содержащих корреляции).

Прежде всего, предполагается, что связь между переменными является линейной. На практике это предположение, в сущности, никогда не может быть подтверждено; к счастью, процедуры множественного регрессионного анализы в незначительной степени подвержены воздействию малых отклонений от этого предположения. Нелинейный режим модуля «Множественная регрессия» позволяет вводить различные нелинейные члены в уравнение регрессии (другие опции нелинейной регрессии доступны в модуле Нелинейное оценивание).

4.7.1 Ограничения

Основное концептуальное ограничение всех методов регрессионного анализа состоит в том, что они позволяют обнаружить только числовые зависимости, а не лежащие в их основе причинные (causal) связи. Например, можно обнаружить сильную положительную связь (корреляцию) между разрушениями, вызванными пожаром, и числом пожарных, участвующих в борьбе с огнем. Следует ли заключить, что пожарные вызывают разрушения? Конечно, наиболее вероятное объяснение этой корреляции состоит в том, что размер пожара (внешняя переменная, которую забыли включить в исследование) оказывает влияние, как на масштаб разрушений, так и на привлечение определенного числа пожарных (т.е. чем больше пожар, тем большее количество пожарных вызывается на его тушение). Хотя этот пример довольно прозрачен, в реальности при исследовании корреляций альтернативные причинные объяснения часто даже не рассматриваются.

4.7.2 Процедура определения модели

В диалоговом окне «Множественный регрессионный анализ» –Multiple Regression выбираются независимые и зависимые переменные, определяется метод и условия проведения анализа и запускается расчет. Этапы расчета показаны на примере в Приложении В.

При выборе нескольких зависимых переменных, программа запускается автоматически для каждой из них.

4.7.3 Результаты регрессионного анализа

Результаты регрессионного анализа появляются в табличной форме в специальном окне «Результат регрессионного анализа». Диалоговое окно текущего множественного регрессионного анализа содержит следующие элементы и функциональные блоки:

а) информационное поле;
б) поле параметров модели;
в) итоговую таблицу регрессии;
г) дисперсионный анализ;
д) дисперсионный анализ, скорректированный на среднее;
ж) ковариации коэффициентов;
з) текущая матрица выметания;
и) частные корреляции;
к) итоги по шагам;
л) предсказанные зависимые переменные;
м) избыточность;
н) корреляции и описательные статистики;
о) выделяемый уровень значимости.

4.7.4 Описание и оценка сообщений в окне «Результат регрессионного анализа»:

а) информационное поле:

В этой части диалогового окна выводятся:

Зависимая переменная. Здесь отображается имя зависимой переменной.

Число наблюдений. Отображается число наблюдений, принятых к обработке (N) (зависит от выбранного способа обработки пропущенных данных).

Множественный R. Отображается коэффициент множественной корреляции между зависимой и независимыми переменными, равный положительному корню из R-квадрат (коэффициента детерминации, см. Остаточная дисперсия и R-квадрат).

R-квадрат. Отображается коэффициент множественной детерминации, определяемый как:

R-квадрат = 1 - (SS остатков/Общая SS) (2)

Равен квадрату коэффициенту множественной корреляции, показывает какую долю дисперсии зависимой переменной объясняет уравнение регрессии.

Скорректированный R-квадрат. R-квадрат корректируется делением суммы квадратов ошибок и общей суммы квадратов на соответствующее им число степеней свободы.

R-квадрат(скорректированный)=1-[(SSостатков/сс)/(ОбщаяSS/сс)] (3)

Стандартная ошибка оценки. Эта величина измеряет рассеяние наблюдаемых значений относительно линии регрессии.

Свободный член. Если выбрана регрессия, включающая свободный член (см. Определение модели), то будет приведена оценка свободного члена.

Стандартная ошибка. Отображается стандартная ошибка оценки свободного члена.

t (число ст.св.) и p-значение. t-статистика и соответствующее ей p-значение используются для проверки гипотезы о равенстве нулю свободного члена в уравнении регрессии.

F-критерий, ст.свободы и p-значение: Статистика F-критерия и соответствующий уровень p используются в качестве общего F-критерия для проверки гипотезы о зависимости предикторов и отклика.

Имеем:

F = MS регрессии/MS остатков (4)

б) поле параметров модели

В этой части окна отображаются и выделяются статистически значимые коэффициенты регрессии (бета - коэффициенты) для переменных, включенных в анализ. Критерий статистической значимости (выделяемый уровень значимости - альфа) может быть выбран из интервала (0.0001–0.5). По умолчанию он равен 0.05 (см. ниже).

в) итоговая таблица регрессии

Эта опция строит таблицу результатов со стандартизованными (бета) и нестандартизованными (B) регрессионными коэффициентами (весами), их стандартными ошибками и уровнями значимости. [Стандартизованные коэффициенты оцениваются по стандартизованным данным, имеющим выборочное среднее 0 и дисперсию 1.] Итоговые статистики регрессионного анализа (в том числе R, R-квадрат и др.) отображаются в заголовке таблицы.

г) дисперсионный анализ

Нажатие этой кнопки строит полную таблицу дисперсионного анализа (ANOVA) для данного регрессионного уравнения. (Подробнее см. в разделе Дисперсионный анализ).

д) дисперсионный анализ, скорректированный на среднее

Опция доступна только, когда свободный член не включен в модель. В этом случае можно вычислить значение множественного R-квадрата, основываясь или на вариации относительно нуля или относительно среднего. По умолчанию, значение R-квадрата, приводимое в информационном поле, соответствует первому случаю, т.е. равно доле вариации зависимой переменной от-

носительно нуля, объясняемой предикторами. Если вы нажмете эту кнопку, STATISTICA построит таблицу дисперсионного анализа, скорректированного на среднее (при вычислении сумм квадратов будет вычитаться выборочное среднее). По поводу альтернативных способов вычисления R- квадрата смотрите работу Kvalseth (1985).

ж) ковариации коэффициентов

Опция строит две таблицы результатов: (1) таблицу с корреляциями регрессионных коэффициентов. (2) таблицу с выборочными дисперсиями (на диагонали) и ковариациями регрессионных коэффициентов.

з) текущая матрица выметания

Эта опция строит таблицу результатов с текущей матрицей выметания. Обращение матрицы в процедуре Множественной регрессии производится методом выметания. Матрица выметания для всех независимых переменных, включенных на данном этапе в регрессионное уравнение, равна -1, умноженной на матрицу, обратную к корреляционной матрице этих переменных (умножение на -1 означает, что знак каждого элемента матрицы изменен на противоположный). Диагональные элементы для переменных, невключенных в уравнение, можно интерпретировать как значения (1 - R-квадрат), при рассмотрении соответствующей переменной в качестве зависимой и использовании всех текущих независимых переменных в уравнении.

Множители, увеличивающие дисперсию: Отметим, что диагональные элементы матрицы, обратной к матрице корреляций (т.е. -1 умноженной на диагональные элементы матрицы выметания, показываемые с помощью этой опции) для переменных, входящих в уравнение иногда также называют множителями, увеличивающими дисперсию (VIF; см., например, работу Neter, Wasserman, Kutner, 1985). Эта терминология обязана своим появлением тому факту, что дисперсии стандартизованных коэффициентов регрессии можно вычислить как произведение дисперсии остатков (для модели с преобразованными корреляциями) на соответствующие диагональные элементы обратной к матрице корреляции. Если предикторы не коррелированы, диагональные элементы обратной к корреляционной матрице равны 1.0; поэтому для коррелированных предикторов эти элементы представляют "увеличивающие множители" дисперсий коэффициентов регрессии, создаваемые избыточностью предикторов.

и) частные корреляции

Нажатие этой кнопки открывает таблицу результатов, содержащую для каждой переменной:

(1) Коэффициент бета (в) (стандартизованный коэффициент для соответствующей переменной, если бы она была включена в уравнение регрессии как независимая переменная);

(2) Частную корреляцию (между соответствующей переменными и зависимой переменной после учета влияния всех остальных независимых переменных в уравнении);

(3) Получастную корреляцию (корреляция между нескорректированной зависимой переменной и соответствующей переменной после учета влияния всех остальных независимых переменных в уравнении);

(4) Толерантность (определяется как 1 минус квадрат множественной корреляции между соответствующей переменной и всеми независимыми переменными в уравнении регрессии);

(5) Коэффициент детерминации R-квадрат (квадрат коэффициента множественной корреляции между данной переменной и всеми остальными переменными, входящими в регрессионное уравнение);

(6) t-значения, ассоциированные с этими статистиками;

(7) уровень значимости для t-значений.

Эти статистики отображаются отдельно для переменных, невключенных в текущее уравнение регрессии, и для переменных, включенных в уравнение (если такие имеются).

к) итоги по шагам

Эта опция доступна в диалоговом окне «Результаты множественной регрессии» только в том случае, если (1) пользователь выбрал регрессию с пошаговым включением или исключением предикторов, или если (2) в предыдущем стандартном регрессионном анализе исследовалась та же самая зависимая переменная и прежний список независимых переменных являлся подмножеством текущего списка независимых переменных или наоборот. Таким образом, в последнем случае можно вычислить приращения R-квадрата, вызванное исключением или включением нескольких переменных на одном шаге (иерархический анализ).

Например, если в первом анализе переменные с номерами от 1 до 5 были выбраны в качестве независимых переменных, а в последующем анализе с той же зависимой переменной пользователь задал переменные с номерами от 1 до 3 в качестве независимых переменных, то выбор этой опции приведет к построению таблицы со значениями R-квадрата, приращениями R-квадрата, F-исключить и числом переменных, удаленных на этом единственном шаге (две переменные в нашем примере).

Если была задана регрессия с пошаговым включением или исключением предикторов, таблица результатов будет содержать приращения R-квадрата на каждом шаге.

л) предсказать зависимые переменные

Опция открывает окно, в котором можно ввести значения независимых переменных и вычислить (предсказать) значение зависимой переменной с помощью построенного уравнения регрессии.

Для предсказанного значения также вычисляются доверительные границы (обозначаемые в таблице результатов ДГ) или границы для предсказания (обозначаемые ГП), в зависимости от того, какая из опций «Вычислить доверительные границы для среднего» или «Вычислить границы для предсказания» выбрана. Вы можете ввести соответствующий альфа:

-уровень для доверительных границ или границ для предсказания в поле ввода Уровень значимости. Программа вычисляет 1-альфа доверительные границы для ожидаемого значения (среднего) зависимой переменной или 1-альфа границы для конкретного предсказаного значения зависимой переменной (подробнее см. в работе Neter, Wasserman, & Kutner, 1985).

м) избыточность

Эта опция строит таблицу с различными показателями избыточности независимых переменных (включенных или не включенных на данном этапе в уравнение регрессии). Для каждой переменной вычисляется: (1) толерантность (определенную как 1 - R-квадрат соответствующей переменной со всеми остальными переменными, включенными в уравнение), (2) R-квадрат (между текущей переменной и всеми остальными переменными уравнения), (3) частную корреляцию (между соответствующей переменной и зависимой переменной после учета влияния всех остальных независимых переменных в уравнении), (4) получастную корреляцию (корреляцию между нескорректированной зависимой переменной и соответствующей переменной после учета влияния всех независимых переменных в уравнении).

н) корреляции и описательные статистики

Кнопка открывает диалоговое окно Просмотр описательных статистик. В этом диалоговом окне вы можете выбрать для просмотра таблицы средних и стандартных отклонений, корреляционную и ковариационную матрицы или сохранить корреляционную матрицу в формате матричного файла STATISTICA.

о) выделяемый уровень значимости

По умолчанию выделяемый уровень значимости (альфа) равен 0.05. Если p-уровень меньше альфа, то этот предиктор выделяется цветом в таблице результатов как значимый. В соответствующем поле диалогового окна вы можете изменить значение выделяемого уровня (0.0001 < альфа <0.5). Нажмите затем кнопку «Применить», чтобы построить таблицу с выделением результатов на новом уровне.

4.8 Графические методы визуализации системы STATISTICA

Графические методы визуализации позволяют находить зависимости, тренды и смещения, "скрытые" в неструктурированных наборах данных.

В системе STATISTICA используются два типа графиков: статистические и пользовательские. Различие состоит в том, что статистические графики отображают статистические результаты или дают другие представления исходных данных для выбранных переменных в текущем файле данных, а пользовательские графики являются общим инструментом наглядного представления числовых значений текущего выделенного блока.

4.8.1 Пользовательские 3М диаграммы рассеяния и поверхности

Эти типы пользовательских графиков вызываются через панели инструментов таблицы результатов или таблицы исходных данных или выпадающее меню «Графика».

Здесь можно задать трехмерную диаграмму рассеяния или поверхность для визуализации сложных интерактивных соотношений между переменными (используя комбинации значений из строк и/или столбцов текущей таблицы результатов или таблицы исходных данных, или подмножества этих значений).

Ниже приведен список зависимостей, доступных в этом диалоговом окне:

–диаграмма рассеяния;

–трассировочный график;

–пространственный график;

–график поверхности;

–спектральная диаграмма;

–карта линий уровня;

–диаграмма отклонений;

Можно выбрать отображение поверхности (или карты линий уровня), представляющей результат сглаживания данных с помощью одной из процедур преобразования/подгонки:

–линейное сглаживание;

–отрицательное экспоненциально-взвешенное сглаживание;

–квадратичное сглаживание;

–сплайны;

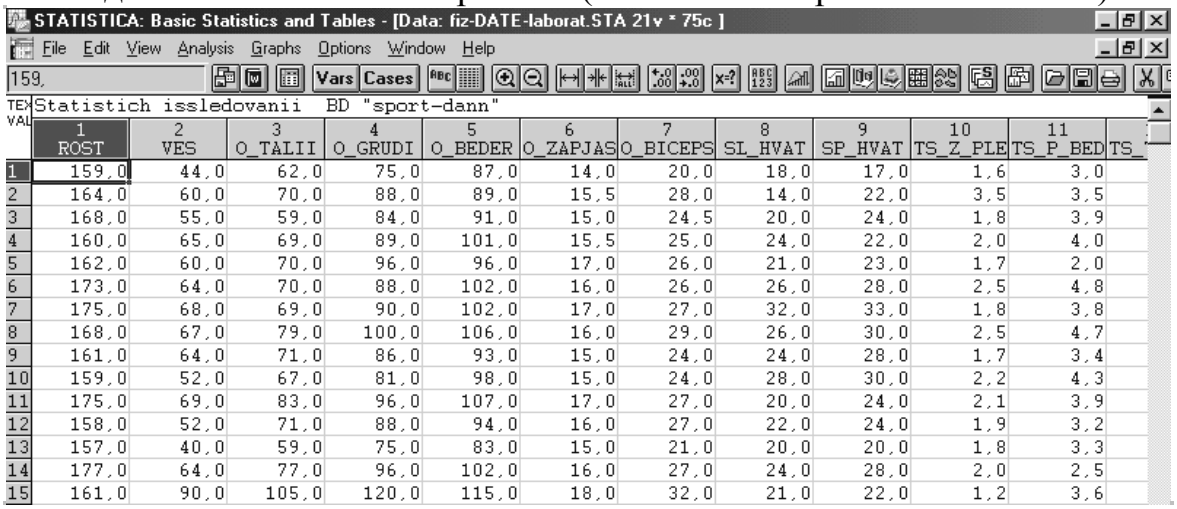
–наименьшие квадраты;

–функция пользователя;

Для вращения или изменения перспективы трехмерных графиков можно использовать интерактивный режим Перспектива и вращение (он вызывается с помощью кнопки «Вращение» или с помощью соответствующей команды графического выпадающего меню «Вид»). В интерактивном окне Перспектива и вращение можно предварительно просмотреть все сделанные изменения, перед тем как принять окончательное решение (нажать кнопку OK). Если выбрать команду «Вращать» из меню «Вид» или нажать кнопку, то сначала появится упрощенное представление графического окна (для сокращения времени перерисовывания).

Приложение Б (рекомендуемое)

- Последовательность этапов работы (английская версия ПП Statistica)



	1 ROST	2 VES	3 O_TALII	4 O_GRUDI	5 O_BEDER	6 O_ZAPJAS	7 O_BICEPS	8 SL_HVAT	9 SP_HVAT	10 TS_Z_PLE	11 TS_P_BED
1	159,0	44,0	62,0	75,0	87,0	14,0	20,0	18,0	17,0	1,6	3,0
2	164,0	60,0	70,0	88,0	89,0	15,5	28,0	14,0	22,0	3,5	3,5
3	168,0	55,0	59,0	84,0	91,0	15,0	24,5	20,0	24,0	1,8	3,9
4	160,0	65,0	69,0	89,0	101,0	15,5	25,0	24,0	22,0	2,0	4,0
5	162,0	60,0	70,0	96,0	96,0	17,0	26,0	21,0	23,0	1,7	2,0
6	173,0	64,0	70,0	88,0	102,0	16,0	26,0	26,0	28,0	2,5	4,8
7	175,0	68,0	69,0	90,0	102,0	17,0	27,0	32,0	33,0	1,8	3,8
8	168,0	67,0	79,0	100,0	106,0	16,0	29,0	26,0	30,0	2,5	4,7
9	161,0	64,0	71,0	86,0	93,0	15,0	24,0	24,0	28,0	1,7	3,4
10	159,0	52,0	67,0	81,0	98,0	15,0	24,0	28,0	30,0	2,2	4,3
11	175,0	69,0	83,0	96,0	107,0	17,0	27,0	20,0	24,0	2,1	3,9
12	158,0	52,0	71,0	88,0	94,0	16,0	27,0	22,0	24,0	1,9	3,2
13	157,0	40,0	59,0	75,0	83,0	15,0	21,0	20,0	20,0	1,8	3,3
14	177,0	64,0	77,0	96,0	102,0	16,0	27,0	24,0	28,0	2,0	2,5
15	161,0	90,0	105,0	120,0	115,0	18,0	32,0	21,0	22,0	1,2	3,6

Рисунок Б1 – Вид экрана с таблицей данных ПП «Statistica»

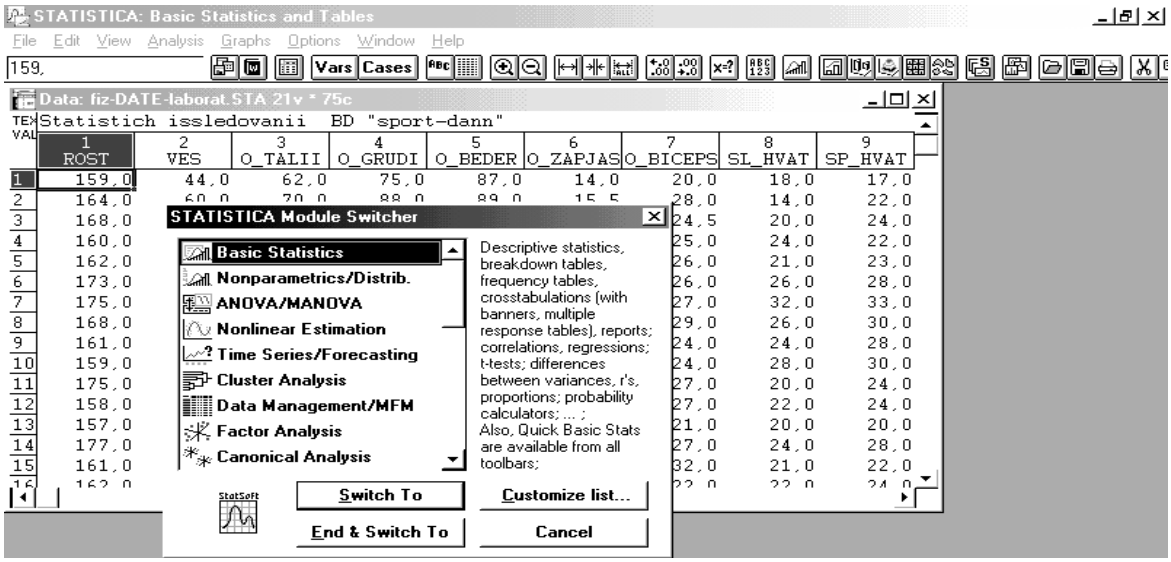


Рисунок Б 2 – Окно переключения статистических модулей

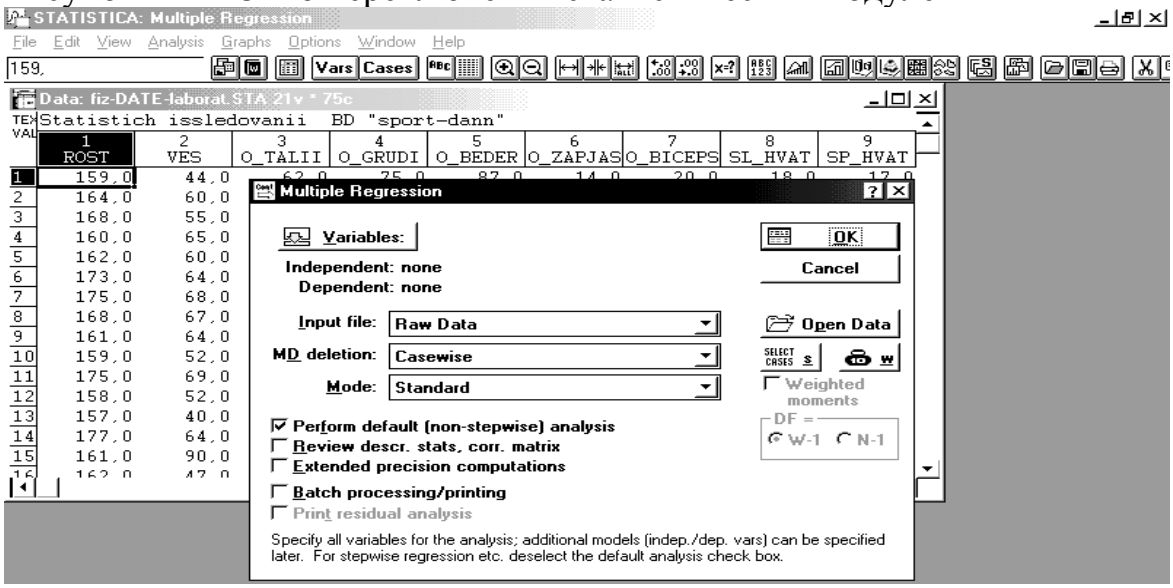


Рисунок Б 3 – Стартовая панель модуля «Множественная регрессия»

После запуска модуля Множественной регрессии на экране появляется стартовая панель модуля «множественная регрессия», в которой задаются переменные для анализа, тип файла данных и включает опции определяющие ход проведения анализа, рисунок Б3.

Опция – Файл ввода (Imput file), рисунок Б 3

В качестве входных данных используют обычные исходные данные (наблюдения и переменные Рисунок Б1) или корреляционную матрицу. Корреляционную матрицу можно предварительно создать в самом модуле Множественная регрессия или вычислить в других модулях пакета STATISTICA

Опция – Variables – «Переменные»

Кнопка вызывает стандартное диалоговое окно выбора переменных, в котором можно выбрать переменные для регрессионного анализа (корреляционная матрица будет вычислена для всех выбранных переменных). Рисунок Б4.

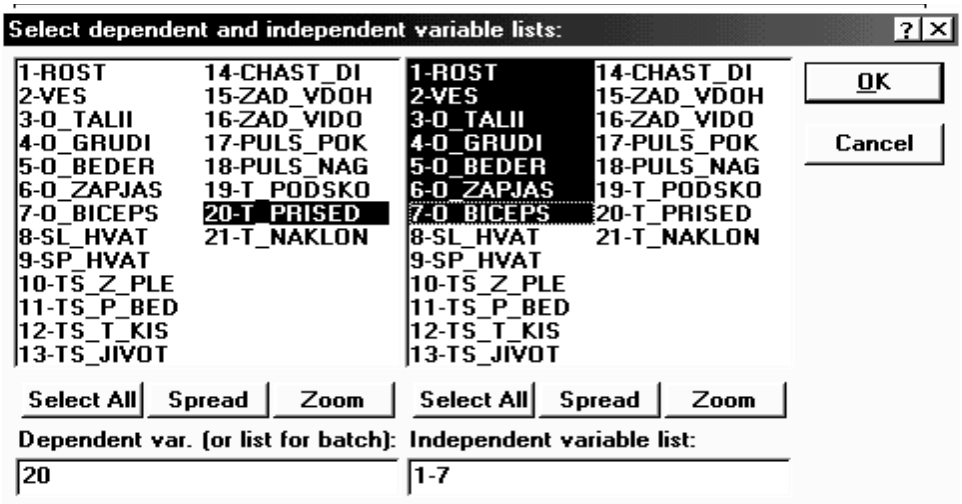


Рисунок Б4 – Окно «Выбор зависимой и независимых переменных»

Опция –Weighted moments – «Взвешенные моменты (B-1: N-1)», рисунок В3

При опции «Взвешенные моменты», вклад каждого наблюдения умножается на значение его веса. Статистики моментов основаны на сумме весов, если выбрана опция B-1, или на числе (невзвешенных) наблюдений, если выбрана опция N-1.

Опция – MD delitions – «Удаление пропущенных данных», рисунок В3

Это поле активно только при использовании файла исходных данных. Пропущенные данные могут удаляться построчно, заменяться средними значениями или удаляться попарно при выборе соответствующей опции в поле.

Построчное удаление пропущенных данных. При выборе этой опции при проведении анализа используются только те наблюдения, которые не имеют пропущенных значений во всех выбранных для анализа переменных.

Замена средним. При выборе этой опции пропущенные значения в каждой переменной заменяются средним, вычисленным по имеющимся наблюдениям соответствующей переменной (только при проведении анализа, а

не в самом файле данных).

Попарное удаление пропущенных данных. Если выбрана эта опция, то при вычислении парных корреляций удаляются наблюдения, имеющие пропущенные значения в соответствующих парах переменных. При проведении последующего анализа все статистические критерии будут основываться на минимальном числе допустимых наблюдений, обнаруженном во всех выбранных переменных.

Опция – Mode –«Режим», рисунок Б3

Пользователь может выбрать стандартную или фиксированную нелинейную регрессию.

Опция – Perform default (non–stepwise) analysis – «Провести анализа по умолчанию», рисунок Б3

После выбора этой опции и нажатия кнопки OK стартовой панели, программа использует для анализа установки, принятые по умолчанию (т.е. Стандартная регрессионная модель, включающая свободный член) в диалоговом окне Определение модели, и перейдет в диалоговое окно Результаты регрессии. Если эта опция отменена, то при щелчке мышью на кнопке OK стартовой панели откроется диалоговое окно «Определение модели», в котором выбирается как тип регрессионного анализа (например, пошаговый, гребневый и др.), так и другие опции.

Опция – Review descry. stats, corr matrix –«Показывать описательные статистики и корреляционную матрицу» , рисунок Б 3.

Опция – Extended precision computations – «Вычисления с повышенной точностью», рисунок Б 3 – опцию следует установить, если анализируемые переменные имеют малую относительную дисперсию

Опция – Batch processing/printing – «Пакетная обработка и печать», рисунок В3–выбор способа печати выходных результатов регрессии.

Опция – Print resulting analyis – «Печать результатов анализа остатков», рисунок Б 3 – опция доступна если установлен переключатель «Провести анализ по умолчанию».

Используем опцию «Провести анализа по умолчанию» – после выбора этой опции и нажатия кнопки OK стартовой панели (рисунок Б3), программа использует для анализа установки, принятые по умолчанию (т.е. Стандартная регрессионная модель, включающая свободный член) и перейдет в диалоговое окно «Результаты регрессии», рисунок Б5.

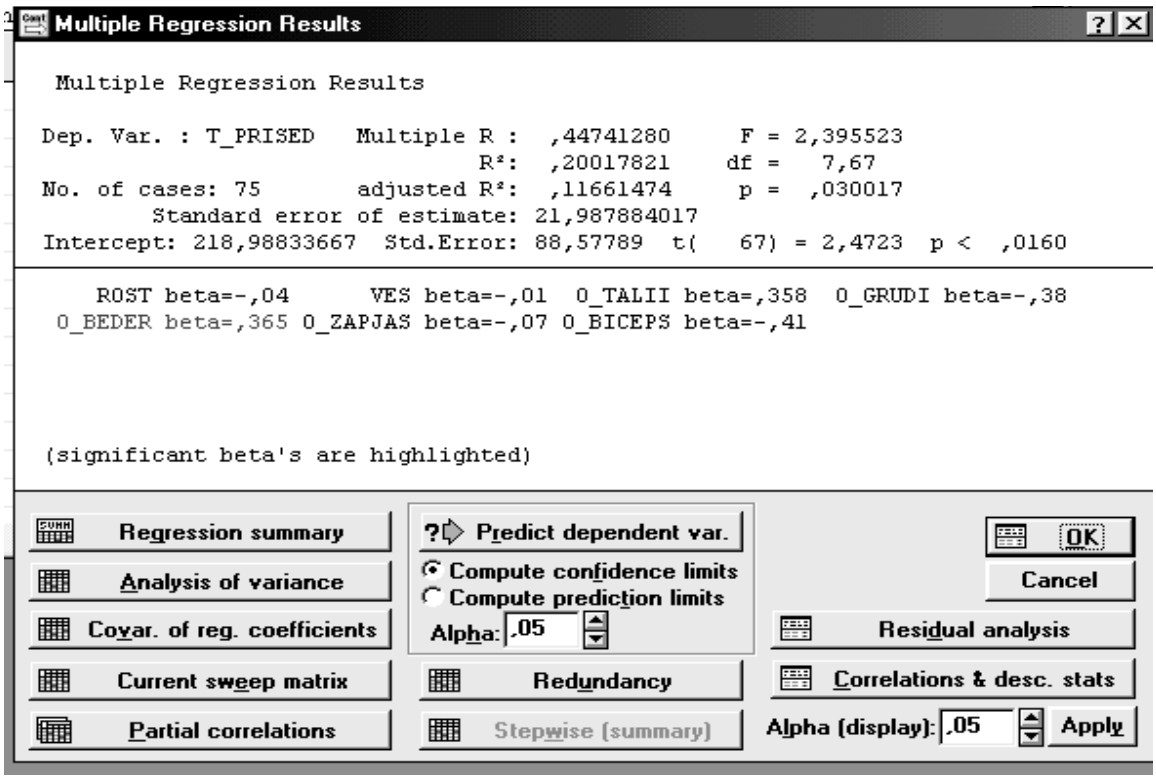


Рисунок Б5 – Окно «Результаты регрессионного анализа»

Если опция Perform default (non–stepwise) analysis –«Провести анализа по умолчанию», рисунок Б3, отменена, то при щелчке мышью на кнопке OK стартовой панели откроется диалоговое окно «Определение модели», рисунок Б6, в котором выбирается тип регрессионного анализа (например, пошаговый, гребневый и др.) и другие опции.

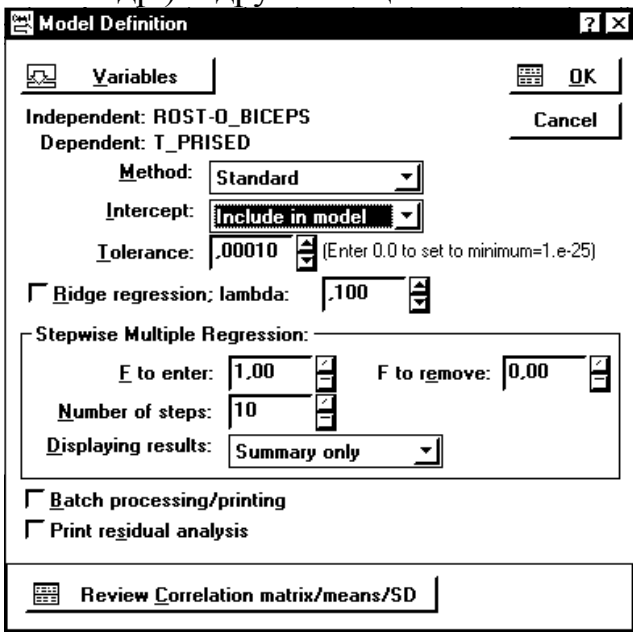


Рисунок Б6 – Диалоговое окно «Определение модели»

Опция –Variabies – «Переменные», рисунок Б6 – выбор зависимой(ых) и независимых переменных.

Опция – Method – «Процедура», рисунок Б 6 – выбор типа регрессионного анализа.

Стандартная регрессия. При выборе этой опции все переменные будут включены в уравнение регрессии одним блоком (т.е. на одном шаге итерации).

Пошаговая регрессия с включением предикторов. При выборе этой опции независимые переменные будут по отдельности включаться или исключаться из модели на каждом шаге регрессии (если выбрано F - включить) до тех пор, пока не будет получена "наилучшая" регрессионная модель.

Пошаговая регрессия с исключением предикторов. При выборе этой опции независимые переменные будут исключаться из модели по одной на каждом шаге (если выбрано F - исключить) до тех пор, пока не будет получена "наилучшая" регрессионная модель.

Опция –Пошаговая регрессия по блокам (иерархическая)
(см электронное руководство)

Опция –Intercept – «свободный член»

Эта опция позволяет задать регрессионное уравнение, включающее свободный член (если выбрана опция– Добавить в модель–«Include in model») или не содержащее свободный член (свободный член приравнивается к нулю–«set to zero» и регрессия проходит через начало координат). В большинстве приложений (в частности, в общественных и естественных науках) интересующие переменные измеряются в более или менее произвольных шкалах и нулевая отметка не имеет особого значения. Поэтому, по умолчанию, в поле Свободный член выбрано значение «Добавить в модель».

Опция –Tolerance – «Толерантность»

Толерантность определяется как 1 минус квадрат множественной корреляции переменной с другими независимыми переменными уравнения регрессии. Поэтому, чем меньше толерантность переменной, тем в большей степени ее вклад в регрессию является избыточным (т.е. она является избыточной при заданных значениях других независимых переменных).

Если толерантность переменной на входе в регрессионное уравнение меньше установленного по умолчанию значения толерантности (0,0001), то это означает, что эта переменная на 99,99 процентов является излишней (идентичной) для переменных, уже включенных в уравнение. Принудительное включение чрезвычайно избыточных переменных в регрессионное уравнение не только сомнительно с точки зрения уместности получаемых результатов, но и приводит к очень ненадежным оценкам.

Опции –F to enter –«включить»; F to remove– «исключить»

Опции доступны только в случае выбора пошаговой регрессии. Заданное значение «F-включить» определяет насколько значимым должен быть вклад переменной в регрессию, чтобы она была добавлена в уравнение. Значение F-исключить, определяет насколько "незначимым" должен быть вклад переменной в уравнение регрессии, чтобы она могла быть исключена из регрессионного уравнения.

Если в регрессии с пошаговым включением предикторов желательно принудительно включить все (или почти все) переменные в уравнение (по од-

ной на каждом шаге), то значение «F-включить» должно быть установлено минимальным (0,0001), а значение «F-исключить» должно быть установлено на свой минимум (0,0); «F-исключить» должно всегда быть меньше «F-включить».

Если в регрессии с пошаговым исключением предикторов желательно удалить все переменные из уравнения (по одной на каждом шаге), то значение «F- включить» должно быть установлено равным очень большому числу (например, 999), а значение «F-исключить» должно быть установлено на сходное по величине значение (например, 998; напомним, что значение «F-включить» всегда должно быть меньше, чем значение «F-исключить»).

Опция – Number of steps – «Число шагов»

Опция – Displaying –«Отображение результатов»

Можно выбрать для просмотра только окончательные (итоговые) результаты пошагового регрессионного анализа («итоги») или результаты каждого шага анализа («На каждом шаге»).

Опция –ridge regression – «Гребневой анализ»

Гребневая (или ридж) регрессия используется, когда независимые переменные очень сильно коррелируют друг с другом, и, поэтому, устойчивые оценки регрессионных коэффициентов не могут быть получены с помощью обычного метода наименьших квадратов. Гребневая регрессия искусственно занижает коэффициенты корреляции, так, что могут быть вычислены более устойчивые (хотя и смещенные) оценки (бета-коэффициенты).

Пакетная обработка и печать – см ранее

Печатать результаты анализа остатков– см ранее

Просмотреть описательные статистики– см ранее

После определения модели запускается программа расчета и выводится окно «Таблица результатов регрессионного анализа», рисунок В5. В этом диалоговом окне приводится сводка результатов текущего регрессионного анализа и предоставляет возможность специфических дополнительных оценок.

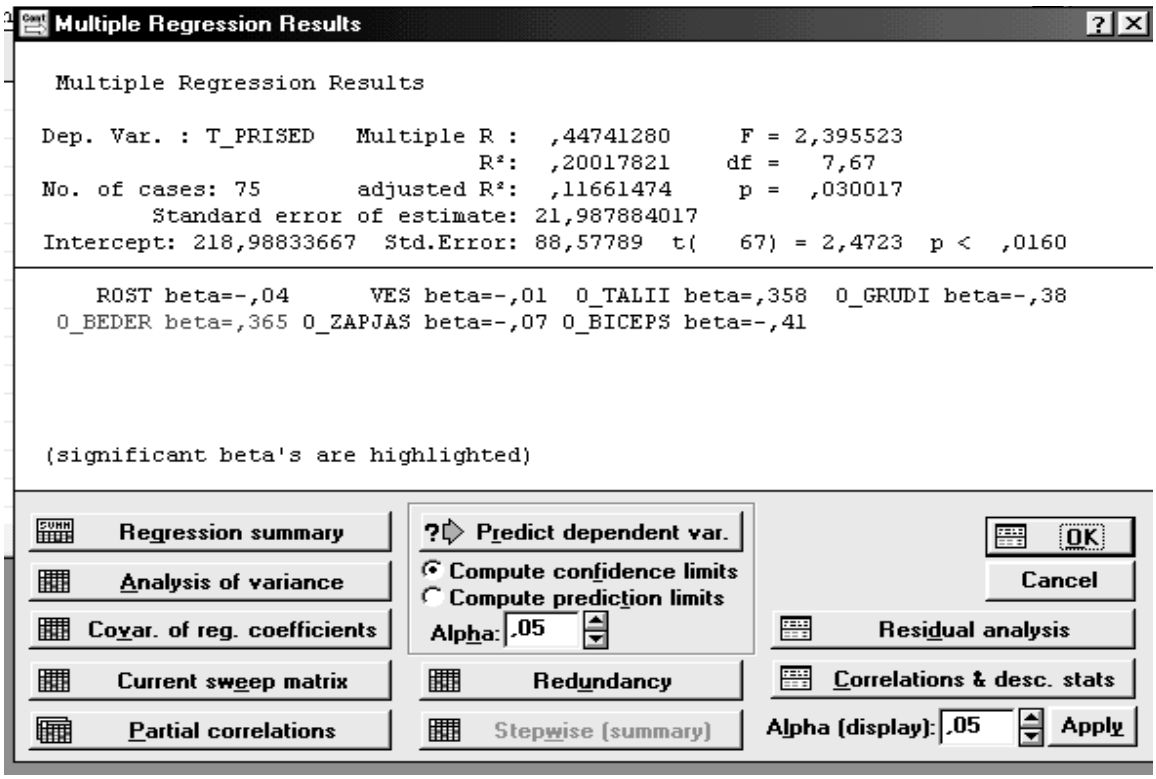


Рисунок Б7 – Окно «Результаты регрессионного анализа» (повторно)

а) информационное поле (рисунок Б7):

Зависимая переменная – T–PRISED

Число наблюдений, принятых к обработке (N) –75

Коэффициент множественной корреляции между зависимой и независимыми переменными R – 0,447

Коэффициент множественной детерминации R-квадрат–0,200

Скорректированный R-квадрат– 0,117

Стандартная ошибка оценки– 21,988

Свободный член – 218,988

Стандартная ошибка–88,578

t-статистика (p-значение) – 2,472 (0,016)

F-критерий (уровень p) – F(7,67)= 2,396

б) поле параметров модели:

статистически значимые коэффициенты регрессии (бета – коэффициенты):

O_BEDER –0,365

в) регрессионная зависимость:

Regression Summary for Dependent Variable: T_PRISED

MULTIPLE REGRESS. R= ,44741280 RI= ,20017821 Adjusted RI= ,11661474
F(7,67)=2,3955 p<,03002 Std.Error of estimate: 21,988

N=75	BETA	St. Err. of BETA	B	St. Err. of B	t(67)	p-level
Intercpt			218,99	88,58	2,47	,02
ROST	-,04	,13	-,14	,48	-,30	,76
VES	-,01	,27	-,02	,69	-,02	,98
O_TALII	,36	,19	,99	,53	1,85	,07
O_GRUDI	-,38	,27	-1,09	,78	-1,41	,16
O_BEDER	,36	,18	1,15	,56	2,06	,04
O_ZAPJAS	-,07	,16	-1,68	4,04	-,42	,68
O_BICEPS	-,41	,21	-3,69	1,94	-1,91	,06

Рисунок Б8 – Итоговая таблица регрессии

Математическая модель частоты пульса приседаний (T_PRISED) по итоговой таблице регрессии:

T_PRISED = 214,51 + 0,36 O_BEDER

(4)

Остальные переменные имеют уровень значимости меньше допустимого уровня 0,05: O_BICEPS – 0,06; O_TALII – 0,07

г) дисперсионный анализ

Analysis of Variance; DV: T_PRISED (fiz-date-laborat.sta)

Continue...	Sums of Squares	df	Mean Squares	F	p-level
Regress.	8107,09	7	1158,156	2,396	,0300
Residual	32392,29	67	483,467		
Total	40499,39				

Рисунок Б9 – таблица дисперсионного анализа

ж) ковариация коэффициентов

Covariances of Regression Coefficients B; DV:

Continue...	ROST	VES	O_TALII	O_GRUDI	O_BEDER	O_ZAPJAS	O_BICEPS
ROST	,228	-,13	,02	,051	-,023	-,34	,150
VES	-,127	,48	-,15	-,171	-,098	-,46	-,096
O_TALII	,024	-,15	,28	-,098	-,011	,12	,005
O_GRUDI	,051	-,17	-,10	,603	-,004	-1,15	-,718
O_BEDER	-,023	-,10	-,01	-,004	,313	,21	-,397
O_ZAPJAS	-,344	-,46	,12	-1,147	,215	16,29	,505
O_BICEPS	,150	-,10	,01	-,718	-,397	,51	3,749

Рисунок Б10 – Таблица ковариации коэффициентов

з) получение матриц выметания

Current Status of Sweep Matrix; DV: T_PRISED

Continue...	ROST	VES	O_TALII	O_GRUDI	O_BEDER	O_ZAPJAS	O_BICEPS	T_PRISED
ROST	-1,45	1,12	-,20	-,41	,17	,32	-,38	-,04
VES	1,12	-5,89	1,74	1,93	1,00	,59	,34	-,01
O_TALII	-,20	1,74	-3,12	1,05	,10	-,14	-,02	,36
O_GRUDI	-,41	1,93	1,05	-6,21	,04	1,34	2,34	-,38
O_BEDER	,17	1,00	,10	,04	-2,61	-,23	1,16	,36
O_ZAPJAS	,32	,59	-,14	1,34	-,23	-2,16	-,19	-,07
O_BICEPS	-,38	,34	-,02	2,34	1,16	-,19	-3,86	-,41
T_PRISED	-,04	-,01	,36	-,38	,36	-,07	-,41	,80

Рисунок Б11 - Текущая матрица выметания

д) дисперсионный анализ, скорректированный на среднее

Variables currently in the Equation; DV: T_PRISED

Continue...	Beta in	Partial Cor.	Semipart Cor.	Tolernce	R-square	t(67)	p-level
ROST	-,040	-,037	-,033	,69	,31	-,30	,76
VES	-,006	-,003	-,002	,17	,83	-,02	,98
O_TALII	,358	,221	,203	,32	,68	1,85	,07
O_GRUDI	-,383	-,169	-,154	,16	,84	-1,41	,16
O_BEDER	,365	,244	,226	,38	,62	2,06	,04
O_ZAPJAS	-,067	-,051	-,045	,46	,54	-,42	,68
O_BICEPS	-,409	-,227	-,208	,26	,74	-1,91	,06

Рисунок Б12–Таблица дисперсионного анализа, скорректированного на среднее