

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет прикладной математики и информатики

“УТВЕРЖДАЮ”

Декан ФПМИ

профессор, д.т.н. Лемешко
Борис Юрьевич

“ ” г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дисциплины специализаций: Математическое обеспечение систем обработки
данных

ООП: специальность 010503.65 Математическое обеспечение и администрирование
информационных систем

Шифр по учебному плану: ДС.Ф.1.2

Факультет: прикладной математики и информатики очная форма обучения

Курс: 5, семестр: 9

Лекции: 12

Практические работы: - Лабораторные работы: 24

Курсовой проект: - Курсовая работа: - РГЗ: 9

Самостоятельная работа: 29

Экзамен: - Зачет: 9

Всего: 65

Новосибирск

2011

Рабочая программа составлена на основании Государственного образовательного стандарта высшего профессионального образования по направлению (специальности): 351500 Математическое обеспечение и администрирование информационных систем.(№ 72 мжд/сп от 10.03.2000)

ДС.Ф.1.2, дисциплины федерального компонента

Рабочая программа обсуждена на заседании кафедры Программных систем и баз данных протокол № 3 от 31.08.2011

Программу разработал

доцент, к.т.н.

Волкова Виктория Михайловна

Заведующий кафедрой

профессор, д.т.н.

Попов Александр Александрович

Ответственный за основную образовательную программу

профессор, д.т.н.

Попов Александр Александрович

1. Внешние требования

Таблица 1.1

Шифр дисциплины	Содержание учебной дисциплины	Часы
ДС.Ф.1.2	Концептуальная записка по специальности 010503.65 - "Математическое обеспечение и администрирование информационных систем". Основные модели данных. Методология разведочного анализа. Базовые модели нелинейной регрессии. Модели и методы факторного анализа. Модели стационарных и нестационарных временных рядов. Однородная нестационарность. Автокорреляционная и частная автокорреляционная функция. Основные этапы идентификации временных рядов. Проверка адекватности построенных моделей данных.	65

2. Особенности (принципы) построения дисциплины

Таблица 2.1

Особенности (принципы) построения дисциплины

Особенность (принцип)	Содержание
Основания для введения дисциплины в учебный план по направлению или специальности	Решение Ученого совета ФПМИ протокол №3 от 23.04.2007.
Адресат курса	Студенты специальности 010503.65 - "Математическое обеспечение и администрирование информационных систем".
Основная цель (цели) дисциплины	Изучение базовых математических методов обработки данных, формирование практических навыков работы с современными пакетами прикладных программ для решения задач анализа данных, с использованием таких методов, как методы линейного и нелинейного регрессионного анализа (и предшествующих ему разведочного и факторного анализа), методы анализа временных рядов.
Ядро дисциплины	Основные модели регрессионного анализа, основные модели описания временных рядов, методы разведывательного и факторного анализов.
Связи с другими учебными дисциплинами основной образовательной программы	Для изучения данного курса студент должен иметь начальную подготовку, предполагающую знания, умения и навыки, полученные при изучении таких курсов, как "Теория вероятностей и математическая статистика", "Статистические методы анализа данных", "Методы оптимизации". Освоение данной дисциплины является основой для последующего прохождения практики, подготовки к итоговой государственной аттестации.
Требования к первоначальному уровню	Для изучения данного курса студент должен иметь начальную подготовку, предполагающую знания, умения и навыки,

подготовки обучающихся	полученные при изучении таких курсов, как "Теория вероятностей и математическая статистика", "Статистические методы анализа данных", "Методы оптимизации".
Особенности организации учебного процесса по дисциплине	лабораторные работы проходят с использованием такого ПО как STATISTICA 6.0

3. Цели учебной дисциплины

Таблица 3.1

После изучения дисциплины студент будет

иметь представление	
1	о современных подходах к анализу данных
2	о современных пакетах прикладных программ, применяемых в анализе данных
3	о методологии Datamining
знать	
4	методологию предварительного (разведочного) анализа данных
5	методологию проведения линейного и нелинейного регрессионного анализа
6	базовые модели нелинейной регрессии
7	методику проведения факторного анализа
8	базовые модели стационарных и нестационарных временных рядов
9	методику проведения анализа временных рядов
уметь	
10	проводить разведочный анализ табличных данных
11	проводить пошаговый регрессионный анализ в соответствии с методологией регрессионного анализа
12	проводить анализ данных с использованием базовых моделей нелинейной регрессии
13	проводить классификацию и редукцию данных с использованием аппарата факторного анализа
14	использовать процедуру идентификации для построения пробной модели временного ряда
15	проводить анализ динамических данных с использованием моделей ARIMA

4. Содержание и структура учебной дисциплины

Лекционные занятия

Таблица 4.1

(Модуль), дидактическая единица, тема	Часы	Ссылки на цели
Семестр: 9		
Модуль: Анализ статистических данных		
Дидактическая единица: основные модели данных		
Классические и современные методы анализа данных. Классификация современных систем обработки данных.	1	1, 2, 3
Основные этапы проведения регрессионного анализа.	1	11, 5
Дидактическая единица: Методология разведочного анализа		
Разведочный (предварительный) анализ данных.	2	10, 4

Дидактическая единица: Базовые модели нелинейной регрессии		
Построение нелинейных регрессионных моделей.	1	12, 5, 6
Дидактическая единица: Модели и методы факторного анализа		
Классификация и редукция многомерных данных с использованием аппарата факторного анализа.	1	13, 7
Модуль: Анализ временных рядов		
Дидактическая единица: Модели временных рядов (Модели стационарных и нестационарных временных рядов; однородная нестационарность)		
Анализ стационарных временных рядов. Модели AR, MA, ARMA.	2	8, 9
Анализ нестационарных рядов. Однородная нестационарность. Модели ARIMA.	2	14, 8
Дидактическая единица: Основные этапы идентификации временных рядов и используемые для этого характеристики (автокорреляционная и частная автокорреляционная функция; основные этапы идентификации временных рядов; проверка адекватности построенных моделей данных)		
Основные этапы построения моделей временных рядов. Идентификация временных рядов.	1	14, 15, 9
Анализ сезонных временных рядов. Анализ интервенций.	1	14, 15, 8, 9

Лабораторная работа

Таблица 4.2

(Модуль), дидактическая единица, тема	Учебная деятельность	Часы	Ссылки на цели
Семестр: 9			
Модуль: Анализ статистических данных			
Разведочный анализ данных с использованием модуля "Основные статистики и таблицы"	Решая задачи, студент: знакомится с возможностями модуля "Основные статистики и таблицы"; изучает методологию разведочного анализа данных; проводит разведочный анализ данных.	4	10, 4
Анализ и редукция данных с использованием модуля "Множественная регрессия"	Решая задачи, студент: знакомится с возможностями модуля "Множественная регрессия"; изучает методологию	4	11, 5

	регрессионного анализа; проводит анализ и редукцию данных с использованием пошаговой регрессии.		
Анализ данных с использованием типовых нелинейных регрессионных моделей в модуле "Нелинейное оценивание".	Решая задачи, студент: знакомится с возможностями модуля "Нелинейное оценивание"; изучает базовые нелинейные регрессионные модели и методику их использования для анализа данных; проводит анализ данных с использованием нелинейной регрессии.	2	12, 5, 6
Анализ данных с использованием факторного анализа. Построение структурной модели методом подтверждающего факторного анализа.	Решая задачи, студент: знакомится с возможностями модуля "Факторный анализ"; изучает математическую модель факторного анализа и методику его проведения; проводит редукцию данных с использованием факторного анализа.	4	13, 7
Модуль: Анализ временных рядов			
Исследование моделей стационарных и однородно нестационарных временных рядов.	Решая задачи, студент: изучает базовые модели AR, MA, ARMA стационарных временных рядов и моделей ARIMA однородно нестационарных временных рядов с использованием генерирования соответствующих выборок и их последующего исследования; проводит сравнение	4	8

	теоретических результатов с результатами, полученными в результате эмпирического исследования.		
Построение моделей стационарных, нестационарных, сезонных временных рядов, рядов с интервенциями.	Решая задачи, студент: знакомится с возможностями модуля "Временные ряды"; изучает методику анализа и прогнозирования временных рядов; проводит анализ динамических данных с использованием методологии Бокса-Дженкинса.	6	14, 15, 8, 9

5. Самостоятельная работа студентов

Семестр- 9, Подготовка к зачету

Самостоятельная работа предполагает работу студентов с конспектами лекций и литературой. На подготовку к зачету выделяется 5 часов самостоятельной работы.

Семестр- 9, РГЗ

На выполнение, подготовку отчета и подготовку к защите выделяется 6 часов самостоятельной работы. Выполняется в соответствии с методическими указаниями по выполнению РГЗ (см. пункт 8.1).

Семестр- 9, Подготовка к занятиям

На данный вид работы выделяется 18 часов. Подготовка к занятиям заключается в изучении теоретического материала по конспектам лекций и литературе, подготовке отчетов к лабораторным работам, используя методические указания (см. пункт 8.1).

6. Правила аттестации студентов по учебной дисциплине

По дисциплине учебным планом предусмотрен зачет.

В таблице приведена структура, содержание и оценка деятельности студентов по бально-рейтинговой системе обучения и контроля (БРСО).

Структура, содержание и оценка деятельности студентов

Вид деятельности	Срок выполнения (неделя)	Минимальная оценка (балл)	Максимальная оценка (балл)
2. Выполнить задания лабораторных работ, оцениваются 6 работ.			
Лабораторная работа №1.	4	6	6,5
Лабораторная работа №2.	5	5	5,5
Лабораторная работа №3.	6	5,5	6
Лабораторная работа №4.	8	5,5	6
Лабораторная работа №5.	9	6	6,5
Лабораторная работа №6.	10	5	5,5
		Итого 33	Итого 36
3. Выполнить и защитить РГЗ	11 неделя сдача и защита	6	10
4. Поощрения за выполнение графика работ При сдаче лабораторной работы.	При сдаче в срок ставится максимальный поощрительный балл, равный – 5. На каждой следующей неделе этот поощрительный балл становится – на 0,5 баллов меньше, пока не уменьшится до нуля	+0	+30
При сдаче РГЗ.	При сдаче в срок ставится поощрительный балл, равный 4. На каждой следующей неделе этот поощрительный балл становится – на 1 балл меньше, пока не уменьшится до нуля	+0	+4
		Итого 0	Итого 34
6. Допуск к зачёту (не менее)		39	

7. Максимум к зачёту			80
8. Сдать устный зачет по дисциплине, получить зачёт по дисциплине	12	11	20
9. Минимальный рейтинг по дисциплине		50	
10. Максимальный рейтинг по дисциплине			100

Примечания:

1. Максимальный балл за лабораторную работу (или РГЗ) студент получает, если она выполнена в полном соответствии с заданием и студент дал по ней исчерпывающие ответы на все вопросы, продемонстрировал знание теоретического материала, необходимого для выполнения работы, минимальный – работа выполнена в целом в соответствии с заданием (с небольшими замечаниями), но студент дал ответы на половину вопросов. Сдача работы предусматривает собой демонстрацию оформленного отчета (в электронном или бумажном виде). Сдача происходит побригадно, в присутствии всех членов бригады.
2. Студент допускается к сдаче зачёта по дисциплине, если выполняет все указанные в таблице виды деятельности и набирает не менее 39 баллов.
3. Студент получает зачёт по дисциплине, если по результатам проведения устного зачета он набирает хотя бы 11 баллов, что соответствует ситуации, когда студент дал основные определения, но не понимает сути рассказанного. Максимальный балл за зачет студент получает, если он на свои вопросы дал исчерпывающие ответы.
4. Итоговая оценка в разных шкалах приведена в таблице ниже:

Диапазон баллов рейтинга	Оценка ECTS	Традиционная (4-уровневая) шкала оценки	
97-100+	A+	отлично	зачтено
93-96	A		
90-92	A-		
87-89	B+		
83-86	B	хорошо	
80-82	B-		
77-79	C+		
73-76	C		
70-72	C-	удовлетворительно	
67-69	D+		
63-66	D		
60-63	D-		
50-59	E		
25-49	FX	неудовлетворительно	не зачтено
0-24	F		

7. Список литературы

7.1 Основная литература

В печатном виде

1. Боровиков В. П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учебное пособие для вузов по специальности "Прикладная математика" / В. П. Боровиков, Г. И. Ивченко. - М., 2006. - 367, [1] с. : ил. - Рекомендовано МО.
2. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл.
3. Халафян А. А. STATISTICA 6. Статистический анализ данных : [учебное пособие для вузов по специальности "Статистика" и другим экономическим специальностям] / А. А. Халафян. - М., 2007. - 503, [5] с. : ил. - Рекомендовано МО.
4. Халафян А. А. STATISTICA 6. Статистический анализ данных : учебное пособие для вузов / А. А. Халафян. - М., 2008. - 503, [5] с. : ил. - Рекомендовано МО.
5. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл.

В электронном виде

1. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл.. - Режим доступа: <http://www.ciu.nstu.ru/fulltext/textbooks/2008/avdeenko.pdf>. - Инновационная образовательная программа НГТУ «Высокие технологии».
2. Авдеенко Т. В. Компьютерные методы анализа временных рядов и прогнозирования : учебное пособие / Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, 2008. - 270, [1] с. : ил., табл.. - Режим доступа: <http://www.ciu.nstu.ru/fulltext/textbooks/2008/avdeenko.pdf>. - Инновационная образовательная программа НГТУ «Высокие технологии».

7.2 Дополнительная литература

В печатном виде

1. Трауберг Л. З. Дэвид Уорк Гриффит / Леонид Трауберг. - М., 1981. - 208, [1] с., [24] л. ил. : портр.
2. Боровиков В. П. STATISTICA: искусство анализа данных на компьютере / Владимир Боровиков. - СПб., 2003. - 688 с. : ил. + 1 электрон. опт. диск (CD-ROM).
3. Боровиков В. П. STATISTICA: искусство анализа данных на компьютере : Для профессионалов. - СПб., 2001. - 650 с. : ил.. - В прилож.: CD-ROM.
4. Боровиков В. П. STATISTICA: Статистический анализ и обработка данных в среде Windows / В. П. Боровиков, И. П. Боровиков. - М., 1997. - 608 с.
5. Боровиков В. П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учебное пособие для вузов по специальности "Прикладная математика" / В. П. Боровиков, Г. И. Ивченко. - М., 2000. - 384 с. : ил.
6. Бериков В. Б. Анализ статистических данных с использованием деревьев решений : учебное пособие [для 4-5 курсов ИДО (направление 0617-Статистика)] / В. Б. Бериков ; Новосиб. гос. техн. ун-т. - Новосибирск, 2002. - 60 с. : ил.

7. Боровиков В. П. Программа STATISTICA для студентов и инженеров. - М., 2001. - 300 с. : ил.
8. Маркушевич А. И. Теория аналитических функций : учебное пособие для вузов / А. И. Маркушевич. - М., 1950. - 703 с. : ил. - Рекомендовано МО.
9. Загоруйко Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. - Новосибирск, 1999. - 269 с. : ил. - Библиогр.: с. 247-260. - Предм. указ.: с. 261-263.
10. Дрейпер Н. Р. Прикладной регрессионный анализ : пер. с англ. / Норман Р. Дрейпер, Гарри Смит. - М. [и др.], 2007. - 911 с. : ил., табл.
11. Бокс Д. Анализ временных рядов прогноз и управление. Вып. 1 / Дж. Бокс, Г. Дженкинс ; пер. с англ. А. Л. Левшина, под ред. В. Ф. Писаренко. - М., 1974. - 406 с. : ил., табл.
12. Дубров А. М. Многомерные статистические методы для экономистов и менеджеров : учебник для экон. спец. вузов / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. - М., 2000. - 350 с. : ил.
13. Бокс Д. Анализ временных рядов прогноз и управление. Вып. 2 / Дж. Бокс, Г. Дженкинс ; пер. с англ. А. Л. Левшина, под ред. В. Ф. Писаренко. - М., 1974. - 197 с. : ил., табл.
14. Харман Г. Современный факторный анализ / Гарри Харман ; пер. с англ. В. Я. Лумельского ; ред. и вступ. статья Э. М. Бравермана. - М., 1972. - 485, [1] с. : табл.
15. Лбов Г. С. Анализ данных и знаний : учебное пособие [для 3-5 курсов ФПМИ (специальность 5102) дневного отделения] / Г. С. Лбов ; Новосиб. гос. техн. ун-т. - Новосибирск, 2001. - 86 с. : ил.
16. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров; Под ред. В. Э. Фигурнова. - М., 2003. - 544 с. : ил.

8. Методическое и программное обеспечение

8.1 Методическое обеспечение

В электронном виде

1. Волкова В. М. Методические указания к выполнению лабораторных работ по предмету "Математическое обеспечение систем обработки данных" [Электронный ресурс] : учебно-методическое пособие / В. М. Волкова, Т. В. Авдеенко ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа:
http://ciu.nstu.ru/fulltext/unofficial/2012/lib_21752_1325843383.rar. - Загл. с экрана.
2. Волкова В. М. Методические указания к выполнению РГЗ по предмету "Математическое обеспечение систем обработки данных" [Электронный ресурс] : учебно-методическое пособие / В. М. Волкова ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа:
http://ciu.nstu.ru/fulltext/unofficial/2012/lib_21752_1325854052.rar. - Загл. с экрана.

8.2 Программное обеспечение

1. Statsoft , Statistica , ПО – комплект всех основных инструментов STATISTICA

9. Контролирующие материалы для аттестации студентов по дисциплине

Вопросы к теоретическому зачету

1. Базовые понятия анализа данных. Описательные статистики. Основные типы шкал измерений. Процентили и квартили. Ящик с усами. Доверительное оценивание и проверка статистических гипотез. Предварительный разведочный анализ данных.
2. t -критерий для одно- и двухвыборочных данных, для категоризованных данных. Однофакторный дисперсионный анализ.
3. Основные различия между методами главных компонент и факторным анализом. Виды и особенности методов главных компонент.
4. Основные понятия и этапы факторного анализа. Этапы проведения подтверждающего факторного анализа
5. Методы вращения факторов, их основные особенности. Методы определения числа значимых факторов.
6. Математическая постановка задачи регрессионного анализа. Возможности применения регрессионного анализа в экспериментальных исследованиях. МНК-оценки параметров модели. Свойства оценок. Проверка гипотез об адекватности модели и о значимости коэффициентов регрессии. Анализ остатков.
7. Исследование предположений регрессионного анализа. Проблема мультиколлинеарности и гетероскедастичности. Анализ независимости остатков. Критерий Дарбина-Уотсона.
8. Основные типы нелинейных регрессионных моделей. Возможности сведения нелинейных моделей к линейным. Логит-регрессия и пробит-регрессия и проверка их адекватности. Регрессионные модели с точками разрыва.
9. Построение нелинейной регрессионной модели и проверка ее адекватности. Общая модель роста. Методы нелинейного оценивания (вид функции потерь). Основные методы оптимизации, применяемые для минимизации целевой функции. Проблема мультиколлинеарности в нелинейном оценивании.
10. Понятия автоковариационной и автокорреляционной функций (АКФ). Частная автокорреляционная функция (ЧАКФ). Коррелограмма. Представление случайного процесса в виде скользящего среднего (МА) и авторегрессии (АР).
11. Оценки основных статистик, используемых при анализе временных рядов. Выборочное среднее и дисперсия, выборочные АКФ и ЧАКФ. Свойства оценок.
12. Модель авторегрессии 1-ого, 2-ого, p - порядка ($AR(1)$, $AR(2)$, $AR(p)$). Теоретические АКФ и ЧАКФ этих моделей.
13. Модель скользящего среднего 1-ого, 2-ого, q - порядка ($MA(1)$, $MA(2)$, $MA(q)$). Теоретические АКФ и ЧАКФ этих моделей.
14. Смешанная авторегрессионная модель скользящего среднего $ARMA(p,q)$ и ее свойства.

15. Модель $ARMA(1,1)$ и ее свойства. Теоретические АКФ и ЧАКФ для этой модели.
16. Нестационарность временного ряда по дисперсии и автоковариациям. Преобразования временных рядов, стабилизирующие дисперсию.
17. Идентификация модели временного ряда. Основные этапы идентификации.
18. Метод наименьших квадратов в анализе временных рядов. Его недостатки при оценивании параметров модели, содержащей скользящее среднее.
19. Сезонные ARIMA-модели и их идентификация.
20. Модели интервенций и их идентификация.