

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»

Кафедра автоматизированных систем управления

“УТВЕРЖДАЮ”
ДЕКАН АВТФ

к.т.н. Рева И. Л.

“ ____ ” _____ г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Компьютерные методы анализа данных и прогнозирования

Образовательная программа: 09.04.03 Прикладная информатика, магистерская программа:
Информационные технологии в моделировании и организации бизнес-процессов

Курс: 1, семестр : 2

Факультет автоматики и вычислительной техники,

		Семестр
№	Вид деятельности	2
1	Всего зачетных единиц (кредитов)	4
2	Всего часов	144
3	Всего занятий в контактной форме, час.	45
4	Лекции, час.	10
5	Практические занятия, час.	0
6	Лабораторные занятия, час.	20
7	из них в активной и интерактивной форме, час.	20
8	Аттестация, час.	2
9	Консультации, час.	13
10	Самостоятельная работа, час.	99
11	Виды самостоятельной работы (курсовой проект, курсовая работа, РГЗ, подготовка к контрольной работе)	КР
12	Вид аттестации	Э

Рабочая программа составлена на основании федерального государственного образовательного стандарта высшего образования по направлению (специальности): 09.04.03 Прикладная информатика

ФГОС введен в действие приказом №1404 от 30.10.2014 г. , дата утверждения: 28.11.2014 г.

Место дисциплины в структуре учебного плана: Блок 1, вариативная, по выбору студента

Рабочая программа разработана на основе компетентностной модели выпускника по направлению (специальности): 09.04.03 Прикладная информатика

Рабочая программа обсуждена на заседании кафедры

АСУ, протокол заседания кафедры №7 от 20.06.2017

Утверждена на совете факультета автоматике и вычислительной техники, протокол № 6 от 21.06.2017

Программу разработал:

профессор, д.т.н. Авдеенко Т. В.

Заведующий кафедрой:

профессор, д.т.н. Гриф М. Г.

Ответственный за образовательную программу:

профессор Мезенцев Ю. А.

1. Внешние требования

Таблица 1.1

Компетенция ФГОС: ПК.4 способность проводить научные эксперименты, оценивать результаты исследований; в части следующих результатов обучения:
з1. знать методы оценивания результатов исследований с использованием методов анализа данных
у1. уметь оценивать результаты исследований с использованием методов анализа данных
Компетенция ФГОС: ПК.8 способность анализировать данные и оценивать требуемые знания для решения нестандартных задач с использованием математических методов и методов компьютерного моделирования; в части следующих результатов обучения:
з1. знать методы и модели анализа данных для решения исследовательских задач
у1. уметь анализировать данные с использованием математических методов и методов компьютерного моделирования

2. Требования НГТУ к результатам освоения дисциплины

Таблица 2.1

Результаты изучения дисциплины по уровням освоения (иметь представление, знать, уметь, владеть)	Формы организации занятий
Компьютерные методы анализа данных и прогнозирования	
ПК.4.з1 знать методы оценивания результатов исследований с использованием методов анализа данных	
1. Основные подходы к анализу данных. Моделирование и идентификация. Основные этапы построения моделей по результатам экспериментальных данных. Современные подходы к анализу данных. Основные типы закономерностей, выявляемых с помощью методов Datamining. Классификация систем анализа данных. Профессиональные, универсальные, специализированные системы.	Лекции; Самостоятельная работа
2. Шкалы измерений: номинальная, порядковая, интервальная, относительная. Описательные статистики. Проверка гипотез о виде распределения. Диаграммы "ящик с усами" и "ствол с листьями". Выбросы. Методы очистки данных. Методы удаления пропущенных данных. t-критерий для одновыборочных, двухвыборочных, категоризованных данных. Группировка данных. Таблицы частот. Таблицы сопряженности. Проверка гипотез о независимости категориальных данных. Таблицы флагов и заголовков. Многомерные отклики и дихотомии.	Лекции; Лабораторные работы; Самостоятельная работа
3. Понятие временного ряда. Основные типы временных рядов. Строгая и слабая стационарность случайных процессов. Стационарность Гауссовских случайных процессов. Временной ряд как реализация случайного процесса. Автоковариационная, автокорреляционная и частная автокорреляционная функция (АКФ и ЧАКФ). Выборочные оценки среднего, АКФ и ЧАКФ. Коррелограммы. Белый шум и его характеристики.	Лекции; Самостоятельная работа
ПК.4.у1 уметь оценивать результаты исследований с использованием методов анализа данных	
4. Выбирать систему анализа данных и соответствующий модуль в соответствии с поставленными целями, типом данных и существующими рекомендациями.	Лабораторные работы; Самостоятельная работа
5. Определять тип шкалы отдельных переменных. Проводить разведочный анализ данных. Проводить анализ таблиц. Проверять гипотезы о форме и виде распределения, о независимости количественных и категориальных переменных, о равенстве основных статистик в различных и категоризованных выборках.	Лабораторные работы; Самостоятельная работа
6. Вычислять оценки среднего, дисперсии, АКФ и ЧАКФ. Строить коррелограммы. Идентифицировать процесс белого шума.	Лабораторные работы; Самостоятельная работа
ПК.8.з1 знать методы и модели анализа данных для решения исследовательских задач	

7. Линейная регрессионная модель. МНК-оценки параметров и их свойства. Проверка адекватности линейной регрессионной модели. Проверка гипотез об исходных предположениях линейной регрессии. Понятие толерантности регрессоров. Устранение мультиколлинеарности удалением избыточных регрессоров. Методы пошаговой регрессии.	Лекции; Лабораторные работы; Самостоятельная работа
8. Основные типы нелинейных моделей. Сведение нелинейных моделей к линейным. Логит-регрессия и пробит-регрессия. Кусоч-но-линейная регрессия. Регрессионные модели с точками разрыва. Общая модель роста. Методы нелинейного оценивания. Функция потерь.	Лекции; Лабораторные работы; Самостоятельная работа
9. Математическая модель факторного анализа. Факторные нагрузки. Собственные значения. Критерий Кайзера и критерий каменистой осыпи для выбора количества факторов. Структурная модель данных.	Лекции; Лабораторные работы; Самостоятельная работа
10. Основные модели стационарных временных рядов: авторегрессионная модель (AR), модель скользящего среднего (MA), авторегрессионная модель со скользящим средним (ARMA). Свойства моделей $AR(1)$, $AR(2)$, ..., $AR(p)$, $MA(1)$, $MA(2)$, ..., $MA(p)$, $AR-MA(1,1)$, ..., $ARMA(p,q)$. Условия стационарности и обратимости данных моделей. Их коррелограммы.	Лекции; Лабораторные работы; Самостоятельная работа
11. Понятие однородной нестационарности. Модель $ARIMA(p,d,q)$. Модель случайного блуждания со сдвигом и без сдвига.	Лекции; Лабораторные работы; Самостоятельная работа
12. Основные этапы построения модели временного ряда. Основные этапы идентификации.	Лабораторные работы; Самостоятельная работа
13. Общая мультипликативная $ARIMA$ - модель для описания сезонных временных рядов. Функции описания интервенций. Модель $ARIMA$, включающая интервенции.	Лабораторные работы; Самостоятельная работа
ПК.8.y1 уметь анализировать данные с использованием математических методов и методов компьютерного моделирования	
14. Обоснованно использовать аппарат множественной регрессии для анализа данных. Грамотно проводить линейный регрессионный анализ, проверять гипотезы о значимости регрессии, о значимости параметров, о нормальности, гомоскедастичности и независимости остатков.	Лабораторные работы; Самостоятельная работа
15. Обосновывать необходимость использования нелинейной регрессии и выбор типа нелинейной функции. Проводить оценивание параметров и анализ остатков построенной регрессии.	Лекции; Лабораторные работы; Самостоятельная работа
16. Использовать методику проведения факторного анализа в задачах классификации и снижения размерности. Строить структурную модель данных на основе результатов факторного анализа и проводить проверку адекватности построенной модели.	Лабораторные работы; Самостоятельная работа
17. Идентифицировать стационарные временные ряды для моделей $AR(1)$, $AR(2)$, ..., $AR(p)$, $MA(1)$, $MA(2)$, ..., $MA(p)$, $ARMA(1,1)$, ..., $ARMA(p,q)$.	Лабораторные работы; Самостоятельная работа
18. Устанавливать и подтверждать наличие однородной нестационарности временного ряда. Различать ситуации случайного блуждания со сдвигом и без сдвига.	Лабораторные работы; Самостоятельная работа
19. Проводить идентификацию и последующее моделирование временных рядов по данным наблюдений. Проводить анализ остатков временного ряда, проверку наличия детерминированного тренда. Осуществлять прогнозирование значений временного ряда.	Лабораторные работы; Самостоятельная работа
20. Идентифицировать сезонность и интервенции во временных рядах. Строить общую $ARIMA$ модель временного ряда, включающего сезонность и интервенции. Осуществлять прогнозирование сезонных временных рядов.	Лабораторные работы; Самостоятельная работа

3. Содержание и структура учебной дисциплины

Таблица 3.1

Темы лекций	Активные формы, час.	Часы	Ссылки на результаты обучения	Учебная деятельность
Семестр: 2				
Дидактическая единица: Основные сведения и разведочный анализ данных				

2. Разведочный (предварительный) анализ данных. Проверка гипотез о виде распределения, об однородности выборок. Анализ таблиц.	1	1	1, 2	Лекция-дискуссия
8. Классический и современный анализ данных. Классификация современных систем обработки данных.	1	1	1	Лекция-дискуссия
Дидактическая единица: Методы регрессионного анализа				
3. Линейный регрессионный анализ. Основные этапы проведения регрессионного анализа. Мультиколлинеарность. Проверка адекватности модели.	2	2	7	Лекция-дискуссия
4. Нелинейный регрессионный анализ. Построение нелинейных регрессионных моделей.	2	2	15, 8	Лекция-дискуссия
Дидактическая единица: Методы многомерного анализа данных				
5. Классификация и редукция многомерных данных с использованием аппарата факторного анализа	2	2	9	Лекция-дискуссия
Дидактическая единица: Методы анализа временных рядов				
6. Анализ стационарных временных рядов. Модели AR, MA, ARMA.	1	1	10, 3	Лекция-дискуссия
7. Анализ нестационарных временных рядов. Однородная нестационарность. Модели ARIMA.	1	1	10, 11, 3	Лекция-дискуссия

Таблица 3.2

Темы лабораторных работ	Активные формы, час.	Часы	Ссылки на результаты обучения	Учебная деятельность
Семестр: 2				
Дидактическая единица: Основные сведения и разведочный анализ данных				
1. Разведочный анализ данных с использованием модуля "Основные статистики и таблицы"	2	4	2, 4, 5	Получение навыков работы в системе Statistica. Выдвижение гипотез на основе основных статистик, анализа распределений и таблиц.
Дидактическая единица: Методы регрессионного анализа				
2. Анализ и редукция данных с использованием модуля "Множественная регрессия"	2	4	14, 7	Построение регрессионных моделей с использованием технологии линейного регрессионного анализа. Проверка адекватности модели. Анализ остатков.
3. Анализ данных с использованием типовых нелинейных регрессионных моделей в модуле "Нелинейное оценивание".	1	2	15, 8	Проверка предположения о линейности. Получение навыков выдвижения гипотез о типе нелинейности. Анализ адекватности нелинейных моделей. Работа с моделями бинарных откликов.

Дидактическая единица: Методы многомерного анализа данных				
4. Анализ данных с использованием факторного анализа.	1	2	16, 9	Получение навыков работы с многомерными данными. Построение модели факторного анализа. Определение количества главных факторов методом Кайзера и каменистой осыпи. Построение структурной модели методом подтверждающего факторного анализа
Дидактическая единица: Методы анализа временных рядов				
5. Исследование моделей стационарных и нестационарных временных рядов	1	2	10, 11, 12, 13, 17, 18, 19, 20, 6	Получение модельных данных на основе моделей заданного вида. Исследование характеристик стационарных и нестационарных моделей на основе смоделированных данных. Изучение АКФ и ЧАКФ исследуемых моделей.
6. Построение моделей стационарных, нестационарных, сезонных временных рядов, рядов с интервенциями.	3	6	12, 13, 17, 18, 19	Проведение основных этапов идентификации модели временного ряда на основе анализа АКФ и ЧАКФ. Проверка адекватности модели временного ряда. Анализ остатков. Выдвижение гипотез о модификации модели временного ряда. Диагностика модели.

Таблица 3.3

Темы для самостоятельного изучения	Активные формы, час.	Часы	Ссылки на результаты обучения	Учебная деятельность
Семестр: 2				
Дидактическая единица: Основные сведения и разведочный анализ данных				
10. Общая схема проверки статистических гипотез. Основные типы статистических гипотез	0	2	1	Самостоятельное изучение теоретического материала
Дидактическая единица: Методы анализа временных рядов				
1. Математические модели стационарных и нестационарных временных рядов	0	4	10, 11	Самостоятельное изучение теоретического материала, решение задач.
8. Основные этапы построения моделей временных рядов. Идентификация временных рядов.	0	4	10, 11, 12	Самостоятельное изучение методики идентификации моделей временных рядов.
9. Анализ сезонных временных рядов. Анализ интервенций.	0	8	13, 20	Самостоятельное изучение теоретического материала

4. Самостоятельная работа обучающегося

№	Виды самостоятельной работы	Ссылки на результаты обучения	Часы на выполнение	Часы на консультации
---	-----------------------------	-------------------------------	--------------------	----------------------

Семестр: 2				
1	Курсовая работа	1, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 2, 20, 3, 4, 5, 6, 7, 8, 9	30	6
Целью КР является изучение и усвоение студентами дополнительного материала по предлагаемой тематике, а также навыков анализа данных с использованием современных методов. Предполагается сбор реальных данных (автоматизированное или ручное анкетирование, сбор данных с сайтов, ...). Примерный дополнительный инструментария анализа данных для изучения: 1. Прогнозирование стационарных и нестационарных временных рядов. 2. Оценивание параметров модели ARIMA методом условного максимального правдоподобия. 3. Использование метода обратного прогнозирования в задаче оценивания параметров модели ARIMA. 4. Аппарат нейронных сетей в задаче прогнозирования временных рядов. 5. Использование аппарата деревьев классификации в задачах анализа данных. 6. Использование аппарата нейронных сетей в задачах классификации. 7. Метод безусловного максимального правдоподобия для оценивания параметров модели ARIMA. 8. Спектральный анализ временных рядов. 9. Векторные ARIMA-модели. 10. Динамические факторные модели. 11. Линейн: 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://elibrary.nstu.ru/source?bib_id=vtls000162364. - Загл. с экрана. Попов А. А. Статистические методы анализа данных: методические указания к расчетно-графическому заданию для студентов IV курса ФПМИ всех направлений и специальностей [Электронный ресурс] : учебно-методическое пособие / А. А. Попов ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа: http://ciu.nstu.ru/fulltext/unofficial/2012/lib_856_1326361424.pdf. - Загл. с экрана.				
2	Подготовка к занятиям	10, 11, 12, 13, 2, 7, 8, 9	32	5
: 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://elibrary.nstu.ru/source?bib_id=vtls000162364 . - Загл. с экрана.				
3	Подготовка к аттестации	1, 10, 11, 12, 13, 2, 3, 7, 8, 9	19	2
: 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://elibrary.nstu.ru/source?bib_id=vtls000162364 . - Загл. с экрана.				
4	Самостоятельное изучение теоретического материала	1, 10, 11, 12, 13, 20	18	0
Студент изучает темы, приведенные в таблице 3.3 : Попов А. А. Статистические методы анализа данных: методические указания к расчетно-графическому заданию для студентов IV курса ФПМИ всех направлений и специальностей [Электронный ресурс] : учебно-методическое пособие / А. А. Попов ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа: http://ciu.nstu.ru/fulltext/unofficial/2012/lib_856_1326361424.pdf . - Загл. с экрана.				

5. Технология обучения

Для организации и контроля самостоятельной работы обучающихся, а также проведения консультаций применяются информационно-коммуникационные технологии (табл. 5.1).

Таблица 5.1

Деятельность	Информационно-коммуникационные технологии
Информирование	e-mail
Консультирование	e-mail
Контроль	e-mail; ЭБС
Размещение учебных материалов	ЭБС

6. Правила аттестации обучающихся по учебной дисциплине

Для аттестации обучающихся по дисциплине используется балльно-рейтинговая система (БРС), позволяющая выставять оценки по традиционной шкале и 15-уровневой ECTS. Краткая информация о БРС приведена в табл. 6.1.

Таблица 6.1

Оцениваемые виды деятельности обучающихся	Мин. балл	Максимальный балл
Семестр: 2		
<i>Лекция:</i>	0	
Контролирующие материалы (список вопросов) приводятся в "Попов А. А. Статистические методы анализа данных: методические указания к расчетно-графическому заданию для студентов IV курса ФПМИ всех направлений и специальностей [Электронный ресурс] : учебно-методическое пособие / А. А. Попов ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа: http://ciu.nstu.ru/fulltext/unofficial/2012/lib_856_1326361424.pdf . - Загл. с экрана."		
<i>Лабораторная:</i>	30	60
Контролирующие материалы (список вопросов) приводятся в "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://elibrary.nstu.ru/source?bib_id=vtls000162364 . - Загл. с экрана."		
<i>Курсовая работа: Итого</i>	0	100
Контролирующие материалы (список вопросов) приводятся в "Попов А. А. Статистические методы анализа данных: методические указания к расчетно-графическому заданию для студентов IV курса ФПМИ всех направлений и специальностей [Электронный ресурс] : учебно-методическое пособие / А. А. Попов ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа: http://ciu.nstu.ru/fulltext/unofficial/2012/lib_856_1326361424.pdf . - Загл. с экрана."		
<i>Экзамен:</i>	20	40
Контролирующие материалы (список вопросов) приводятся в "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://elibrary.nstu.ru/source?bib_id=vtls000162364 . - Загл. с экрана."		

В таблице 6.2 представлено соответствие форм контроля заявляемым требованиям к результатам освоения дисциплины.

Таблица 6.2

Коды компетенций ФГОС	Результаты обучения	Формы контроля	
		Защита КР/КР	Экзамен
ПК.4	з1. знать методы оценивания результатов исследований с использованием методов анализа данных		+
	у1. уметь оценивать результаты исследований с использованием методов анализа данных		+
ПК.8	з1. знать методы и модели анализа данных для решения исследовательских задач	+	+
	у1. уметь анализировать данные с использованием математических методов и методов компьютерного моделирования	+	+

Фонд оценочных средств по дисциплине представлен в приложении № 1 к рабочей программе.

7. Литература

Основная литература

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. Цильковский И. А. Методы анализа знаний и данных : конспект лекций / И. А. Цильковский, В. М. Волкова ; Новосиб. гос. техн. ун-т. - Новосибирск, 2010. - 66, [2] с. : ил.. - Режим доступа: http://elibrary.nstu.ru/source?bib_id=vtls000134804

Дополнительная литература

1. Уэс Маккинли Python и анализ данных [Электронный ресурс]/ Маккинли Уэс— Электрон. текстовые данные.— Саратов: Профобразование, 2017.— 482 с.— Режим доступа: <http://www.iprbookshop.ru/64058.html>.— ЭБС «IPRbooks»

Интернет-ресурсы

1. Бизнес-информатика [Электронный ресурс] : научный журнал. - НИУ ВШЭ, 1993–2017. - Режим доступа: <https://bijournal.hse.ru/>. - Загл. с экрана.
2. ЭБС НГТУ : <http://elibrary.nstu.ru/>
3. ЭБС «Издательство Лань» : <https://e.lanbook.com/>
4. ЭБС IPRbooks : <http://www.iprbookshop.ru/>
5. ЭБС "Znaniyum.com" : <http://znaniyum.com/>
6. :

8. Методическое и программное обеспечение

8.1 Методическое обеспечение

1. Попов А. А. Статистические методы анализа данных: методические указания к расчетно-графическому заданию для студентов IV курса ФПМИ всех направлений и специальностей [Электронный ресурс] : учебно-методическое пособие / А. А. Попов ; Новосиб. гос. техн. ун-т. - Новосибирск, [2011]. - Режим доступа: http://ciu.nstu.ru/fulltext/unofficial/2012/lib_856_1326361424.pdf. - Загл. с экрана.
2. 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://elibrary.nstu.ru/source?bib_id=vtls000162364. - Загл. с экрана.

8.2 Специализированное программное обеспечение

- 1 Statistica

9. Материально-техническое обеспечение

Презентационное оборудование

№	Наименование	Назначение
1	Презентационное оборудование (мультимедиа-проектор, экран, компьютер для управления)	Лекции, презентации теоретического материала

Компьютерный класс

№	Наименование	Назначение
1	Компьютерный класс (Компьютеры объединены в локальную сеть с выходом в Internet)	Проведение лабораторных и практических занятий

Порядок определения рейтинга студента по дисциплине

Рейтинг студента по дисциплине является основой для выставления итоговой оценки по дисциплине в «буквенной» форме в соответствии с 15-уровневой шкалой оценок ECTS, а также в традиционной форме (четырёхуровневая шкала либо «зачтено»). Итоговая оценка в двух формах проставляется в ведомость и зачетную книжку студента.

Характеристика работы студента	Диапазон баллов рейтинга	Оценка ECTS	Традиционная (4-уровневая) шкала оценки			
«Отлично» – работа высокого качества, уровень выполнения отвечает всем требованиям, теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному	98–100	A+	ОТЛИЧНО	зачтено		
	93–97	A				
	90–92	A–				
«Очень хорошо» – работа хорошая, уровень выполнения отвечает большинству требований, теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному	87–89	B+				
	83–86	B				
	80–82	B–				
«Хорошо» – уровень выполнения работы отвечает всем основным требованиям, теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые из выполненных заданий, возможно, содержат ошибки	77–79	C+				
	73–76	C				
	70–72	C–	удовлетворительно			

Характеристика работы студента	Диапазон баллов рейтинга	Оценка ECTS	Традиционная (4-уровневая) шкала оценки	
«Удовлетворительно» – уровень выполнения работы отвечает большинству основных требований, теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые виды заданий выполнены с ошибками	67–69	D+	удовлетворительно	зачтено
	63–66	D		
	60–62	D–		
«Посредственно» – работа слабая, уровень выполнения не отвечает большинству требований, теоретическое содержание курса освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные программой обучения учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному	50–59	E		
«Неудовлетворительно» (с возможностью пересдачи) – теоретическое содержание курса освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий	25–49	FX	неудовлетворительно	незачтено
«Неудовлетворительно» (без возможности пересдачи) – теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, все выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над материалом курса не приведет к какому-либо значимому повышению качества выполнения учебных заданий	0–24	F		

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»

Кафедра автоматизированных систем управления

“УТВЕРЖДАЮ”
ДЕКАН АВТФ
к.т.н., доцент И.Л. Рева
“ ” _____ Г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

УЧЕБНОЙ ДИСЦИПЛИНЫ

Компьютерные методы анализа данных и прогнозирования

Образовательная программа: 09.04.03 Прикладная информатика, магистерская программа:
Информационные технологии в моделировании и организации бизнес-процессов

1. Обобщенная структура фонда оценочных средств учебной дисциплины

Обобщенная структура фонда оценочных средств по дисциплине Компьютерные методы анализа данных и прогнозирования приведена в Таблице.

Таблица

Формируемые компетенции	Показатели сформированности компетенций (знания, умения, навыки)	Темы	Этапы оценки компетенций	
			Мероприятия текущего контроля (курсовой проект, РГЗ(Р) и др.)	Промежуточная аттестация (экзамен, зачет)
ПК.4/НИ способность проводить научные эксперименты, оценивать результаты исследований	з1. знать методы оценивания результатов исследований с использованием методов анализа данных	Анализ нестационарных временных рядов. Однородная нестационарность. Модели ARIMA. Анализ стационарных временных рядов. Модели AR, MA, ARMA. Классический и современный анализ данных. Классификация современных систем обработки данных. Общая схема проверки статистических гипотез. Основные типы статистических гипотез Разведочный анализ данных с использованием модуля "Основные статистики и таблицы" Разведочный (предварительный) анализ данных. Проверка гипотез о виде распределения, об однородности выборок. Анализ таблиц.	Защита курсовой работы, вопросы 5–6 Защита лабораторных работ 1, 5-6	Экзамен Вопросы 1-31
ПК.4/НИ	у1. уметь оценивать результаты исследований с использованием методов анализа данных	Исследование моделей стационарных и нестационарных временных рядов Разведочный анализ данных с использованием модуля "Основные статистики и таблицы"	Защита курсовой работы, Вопросы 5-6 Защита лабораторной работы 1	Экзамен Вопросы 1-31
ПК.8/А способность анализировать данные и оценивать требуемые знания для решения нестандартных задач с использованием математических методов и методов компьютерного моделирования	з1. знать методы и модели анализа данных для решения исследовательских задач	Анализ данных с использованием типовых нелинейных регрессионных моделей в модуле "Нелинейное оценивание". Анализ данных с использованием факторного анализа. Анализ и редукция данных с использованием модуля "Множественная регрессия" Анализ нестационарных временных рядов. Однородная нестационарность. Модели ARIMA. Анализ сезонных временных рядов. Анализ интервенций. Анализ стационарных временных рядов. Модели AR, MA, ARMA. Исследование моделей стационарных и нестационарных временных рядов Классификация и	Защита курсовой работы, вопросы 3-4 Защита лабораторных работ 2-6	Экзамен Вопросы 1-31

		редукция многомерных данных с использованием аппарата факторного анализа Линейный регрессионный анализ. Основные этапы проведения регрессионного анализа. Мультиколлинеарность. Проверка адекватности модели. Математические модели стационарных и нестационарных временных рядов Нелинейный регрессионный анализ. Построение нелинейных регрессионных моделей. Основные этапы построения моделей временных рядов. Идентификация временных рядов. Построение моделей стационарных, нестационарных, сезонных временных рядов, рядов с интервенциями.		
ПК.8/А	у1. уметь анализировать данные с использованием математических методов и методов компьютерного моделирования	Анализ данных с использованием типовых нелинейных регрессионных моделей в модуле "Нелинейное оценивание". Анализ данных с использованием факторного анализа. Анализ и редукция данных с использованием модуля "Множественная регрессия" Исследование моделей стационарных и нестационарных временных рядов Нелинейный регрессионный анализ. Построение нелинейных регрессионных моделей. Построение моделей стационарных, нестационарных, сезонных временных рядов, рядов с интервенциями.	Защита курсовой работы, вопросы 1-3 Защита лабораторных работ 2-6	Экзамен Вопросы 1-31

2. Методика оценки этапов формирования компетенций в рамках дисциплины.

Промежуточная аттестация по дисциплине проводится в 2 семестре - в форме экзамена, который направлен на оценку сформированности компетенций ПК.4/НИ, ПК.8/А.

Кроме того, сформированность компетенций проверяется при проведении мероприятий текущего контроля, указанных в таблице раздела 1.

В 2 семестре обязательным этапом текущей аттестации является курсовая работа. Требования к выполнению курсовой работы, состав и правила оценки сформулированы в паспорте курсовой работы.

Общие правила выставления оценки по дисциплине определяются балльно-рейтинговой системой,

приведенной в рабочей программе учебной дисциплины.

На основании приведенных далее критериев можно сделать общий вывод о сформированности компетенций ПК.4/НИ, ПК.8/А, за которые отвечает дисциплина, на разных уровнях.

Общая характеристика уровней освоения компетенций.

Ниже порогового. Уровень выполнения работ не отвечает большинству основных требований, теоретическое содержание курса освоено частично, пробелы могут носить существенный характер, необходимые практические навыки работы с освоенным материалом сформированы не достаточно, большинство предусмотренных программой обучения учебных заданий не выполнены или выполнены с существенными ошибками.

Пороговый. Уровень выполнения работ отвечает большинству основных требований, теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые виды заданий выполнены с ошибками.

Базовый. Уровень выполнения работ отвечает всем основным требованиям, теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые из выполненных заданий, возможно, содержат ошибки.

Продвинутый. Уровень выполнения работ отвечает всем требованиям, теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра автоматизированных систем управления

Паспорт экзамена

по дисциплине «Компьютерные методы анализа данных и прогнозирования», 2 семестр

1. Методика оценки

Экзамен проводится в устной форме, по билетам. Билет формируется по следующему правилу: первый вопрос выбирается из диапазона вопросов 1–16, второй вопрос из диапазона вопросов 17–31 (список вопросов приведен ниже). В ходе экзамена преподаватель вправе задавать студенту дополнительные вопросы из общего перечня (п. 4).

Форма экзаменационного билета

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет АВТФ

Билет № _____

к экзамену по дисциплине «Компьютерные методы анализа данных и
прогнозирования»

1. Разведочный анализ данных. Основные понятия, инструменты и методы разведочного анализа данных. Формулирование и проверка простейших гипотез (10 баллов)
2. Трехпараметрическое экспоненциальное сглаживание (10 баллов)
3. Задача. (20 баллов) Провести исследование данных методом нелинейного регрессионного анализа.

Утверждаю: зав. кафедрой _____ должность, ФИО
(подпись) _____
(дата)

2. Критерии оценки

- Ответ на экзаменационный билет (тест) считается **неудовлетворительным**, если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, при решении задачи допускает принципиальные ошибки, оценка составляет *0-10 баллов*.
- Ответ на экзаменационный билет (тест) засчитывается на **пороговом** уровне, если студент при ответе на вопросы дает определение основных понятий, может показать

причинно-следственные связи явлений, при решении задачи допускает непринципиальные ошибки, например, вычислительные, оценка составляет 10-20 баллов.

- Ответ на экзаменационный билет (тест) билет засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, законы, дает характеристику процессов, явлений, проводит анализ причин, условий, может представить качественные характеристики процессов, не допускает ошибок при решении задачи, оценка составляет 20-30 баллов.
- Ответ на экзаменационный билет (тест) билет засчитывается на **продвинутом** уровне, если студент при ответе на вопросы проводит сравнительный анализ подходов, проводит комплексный анализ, выявляет проблемы, предлагает механизмы решения, способен представить количественные характеристики определенных процессов, приводит конкретные примеры из практики, не допускает ошибок и способен обосновать выбор метода решения задачи, оценка составляет 30-40 баллов.

3. Шкала оценки

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины.

4. Вопросы к экзамену по дисциплине «Компьютерные методы анализа данных и прогнозирования»

1. Разведочный анализ данных. Основные понятия, инструменты и методы разведочного анализа данных. Формулирование и проверка простейших гипотез.
2. Общая схема проверки статистических гипотез. Основная и альтернативная гипотезы. Ошибка 1 и 2 рода. Статистический критерий. Мощность критерия. Процедура проверки гипотезы.
3. Проверка гипотез о виде распределения. Критерий Хи-квадрат и критерий Колмогорова. Основные виды распределений. Используемых в анализе данных.
4. t-критерий для одно- и двухвыборочных данных, для категоризованных данных.
5. Математическая постановка задачи линейного регрессионного анализа. Возможности применения регрессионного анализа в экспериментальных исследованиях. Предположения регрессионного анализа.
6. МНК-оценки параметров модели. Свойства оценок (несмещенность, состоятельность, эффективность). Понятие функции потерь. Функция потерь для МНК.
7. Экспериментальная проверка предположений регрессионного анализа. Проверка гипотез об адекватности модели (значимость регрессии) и о значимости коэффициентов регрессии. Коэффициент детерминации. Анализ остатков регрессии.
8. Исследование предположений регрессионного анализа. Проблема мультиколлинеарности и гетероскедастичности. Анализ независимости остатков. Критерий Дарбина-Уотсона.
9. Множественная регрессия. Проблема мультиколлинеарности. Вычисление толерантности. Устранение избыточности.
10. Множественная регрессия. Понятие избыточного регрессора. Методы устранения избыточных регрессоров.
11. Основные статистики, характеризующие математическое ожидание выборки (среднее, медиана, мода), и формулы для их вычисления. Основные статистики,

- характеризующие разброс выборки (дисперсия, с.к.о., размах), и формулы для их вычисления.
12. Асимметрия и эксцесс. Проверка симметричности и крутизны распределения. Понятие квантили, процентиля, квартили. Медиана. Использование квантилей при проверке гипотез.
 13. Основные типы нелинейных регрессионных моделей. Возможности сведения нелинейных моделей к линейным. Существенно нелинейные модели. Технология нелинейного регрессионного анализа в компьютерных системах анализа данных.
 14. Логит-регрессия и пробит-регрессия и проверка их адекватности.
 15. Регрессионные модели с точками разрыва. Основные типы кусочно-линейной регрессии.
 16. Математическая постановка задачи факторного анализа. Вращение факторов. Решение задачи классификации.
 17. Методы сглаживания временных рядов (скользящее среднее и скользящая медиана, метод prior). Окно сглаживания.
 18. Основные компоненты временного ряда. Методы выделения трендов.
 19. Декомпозиция временных рядов (метод Census I).
 20. Однопараметрическое экспоненциальное сглаживание.
 21. Двухпараметрическое экспоненциальное сглаживание.
 22. Трехпараметрическое экспоненциальное сглаживание.
 23. Проверка качества модели сглаживания. Критерии качества. Прогнозирование временных рядов.
 24. Математическая модель факторного анализа. Технология факторного анализа в компьютерных системах анализа данных.
 25. Модель $ARMA(p,q)$. Модель $ARMA(1,1)$ и ее свойства. Теоретические АКФ и ЧАКФ для модели $ARMA(1,1)$.
 26. Модель авторегрессии 1-ого, 2-ого, p - порядка ($AR(1)$, $AR(2)$, $AR(p)$). Теоретические АКФ и ЧАКФ этих моделей. Модель скользящего среднего 1-ого, 2-ого, q - порядка ($MA(1)$, $MA(2)$, $MA(q)$). Теоретические АКФ и ЧАКФ этих моделей.
 27. Понятия автоковариационной и автокорреляционной функций (АКФ). Частная автокорреляционная функция (ЧАКФ). Коррелограмма. Представление случайного процесса в виде скользящего среднего (МА) и авторегрессии (АР).
 28. Идентификация модели временного ряда. Основные этапы идентификации.
 29. Нестационарность временного ряда по дисперсии и автоковариациям. Преобразования временных рядов, стабилизирующие дисперсию.
 30. Однородная нестационарность временного ряда. Модель $ARIMA(p,d,q)$. Операция взятия разности.
 31. Расширение методологии $ARIMA$ на анализ сезонных временных рядов. Модель $ARIMA(p,d,q) \times (P,D,Q)_s$

Паспорт курсовой работы

по дисциплине «Компьютерные методы анализа данных и прогнозирования», 2 семестр

1. Методика оценки.

Задание:

Целью КР является изучение и усвоение студентами дополнительного материала по предлагаемой тематике, а также навыков анализа данных с использованием современных методов.

Структура:

Работа состоит из двух частей: сбор реальных данных для исследования (автоматизированное или ручное анкетирование, сбор данных с сайтов, ...) и собственно анализ данных с использованием дополнительных методов анализа, изученных студентом самостоятельно.

Дополнительные методы анализа данных для изучения самостоятельно:

1. Прогнозирование стационарных и нестационарных временных рядов.
2. Оценивание параметров модели ARIMA методом условного максимального правдоподобия.
3. Использование метода обратного прогнозирования в задаче оценивания параметров модели ARIMA.
4. Аппарат нейронных сетей в задаче прогнозирования временных рядов.
5. Использование аппарата деревьев классификации в задачах анализа данных.
6. Использование аппарата нейронных сетей в задачах классификации.
7. Метод безусловного максимального правдоподобия для оценивания параметров модели ARIMA.
8. Спектральный анализ временных рядов.
9. Векторные ARIMA-модели.
10. Динамические факторные модели.
11. Линейный дискриминантный анализ Фишера.
12. Канонический дискриминантный анализ.
13. Планирование эксперимента в пакете Statistica 6.
14. Анализ соответствий.
15. Многомерное шкалирование.
16. Распознавание образов без обучения.
17. Распознавание образов с обучением.

Этапы выполнения и защиты:

- 1) сбор реальных данных для исследования
- 2) анализ данных с использованием дополнительных методов анализа, изученных студентом самостоятельно

- 3) Оформление отчета
- 4) Защита курсовой работы

Оцениваемые позиции:

- 1) Глубина теоретического изучения материала
- 2) Практическая ценность решаемой задачи анализа данных
- 3) Адекватность и достаточность выбора моделей и методов анализа данных
- 4) Полученные результаты, самостоятельность постановки целей исследования
- 5) Качество интерпретации результатов и выводов.

2. Критерии оценки.

- работа считается **не выполненной**, если поставленная цель не имеет практической значимости, данные и результаты являются вторичными, новый теоретический материал не изучен, оценка составляет 0–49 баллов.
- работа считается выполненной **на пороговом** уровне, если поставленная цель имеет слабую практическую значимость, данные являются вторичными, однако результаты обладают определенной новизной, изучение нового теоретического материала носит поверхностный характер, оценка составляет 50–69 баллов.
- работа считается выполненной **на базовом** уровне, если поставленная цель имеет практическую значимость, данные являются оригинальными, результаты обладают определенной новизной, изучение нового теоретического материала носит достаточный характер, оценка составляет 70–86 баллов.
- работа считается выполненной **на продвинутом** уровне, если поставленная цель имеет большую практическую значимость, данные являются оригинальными, проведены глубокие исследования на основе изучения новых методов, использование которых обоснованно, результаты являются новыми и хорошо оформлены, оценка составляет 87–100 баллов.

3. Шкала оценки.

В общей оценке по дисциплине баллы за работы учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины. Оценка за курсовую работу ставится по 100-балльной системе, пересчет в буквенную оценку ECTS, а также в пятибалльную систему проводится в соответствии со стандартными правилами пересчета, представленным в файле БРС

4. Примерный перечень тем курсовой работы.

1. Статистический анализ имущества на примере предприятия строительства трубопроводных магистралей нефтегазового комплекса
2. Статистический анализ товарных запасов предприятия
3. Статистический анализ данных возрастной периодизации и уровня инновационного производства ФО
4. Статистический анализ данных финансовой отчетности и построение прогноза активов предприятия
5. Сегментация клиентов Интернет-провайдера, которые работают с ним от 301 до 999 дней
6. Построение краткосрочного прогноза на основе статистических данных посещаемости сайта компании
7. Статистический анализ текущей ситуации на зарегистрированном рынке труда Сибирского федерального округа

8. Классификация пользователей с «серыми» IP адресами в телекоммуникационной компании
9. Классификация пользователей с «белыми» IP адресами в телекоммуникационной компании
10. Сегментация пользователей Интернет услуг со скоростью более 4000 килобит/с
11. Анализ з/п сотрудников на основе статистических данных
12. Анализ роста российских предприятий с использованием компьютерных методов анализа данных
13. Анализ данных российского мониторинга об удовлетворенности работой
14. Сравнительный анализ данных профиля пользователя социальных сетей "Вконтакте", "Одноклассники", "FaceBook", "Мой мир"
15. Статистический анализ рейтинга инновационного развития субъектов РФ
16. Анализ данных поведенческих факторов пользователей сайта обучающей тематики
17. Анализ данных пользователей интернет магазина
18. Компьютерный анализ данных использования информационных систем
19. Компьютерный анализ данных показателей эффективности инвестиционных проектов в существующих онлайн-сервисах
20. Компьютерный анализ данных профиля пользователя социальной сети Вконтакте
21. Компьютерный анализ временных рядов по продажам молочной компании
22. Анализ данных посетителей экстрим центра
23. Статистический анализ данных, количество индивидуальных посещений сайтов за апрель 2014 по аудитории города Новосибирска

5. Перечень вопросов к защите курсовой работы.

1. Обоснование актуальности поставленной цели исследования
2. Обоснование данных для анализа.
3. Обоснование выбранных методов исследования
4. Математические модели и алгоритмы используемых методов анализа данных
5. Сравнение планируемых и полученных результатов.
6. Оценка и интерпретация полученных результатов.