УТВЕРЖДАЮ
Проректор-директор ИК
_____Захарова А. А.
«___»_____2014 г.

БАЗОВАЯ РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

НАПРАВЛЕНИЕ ООП **09.04.02 Информационные системы и технологии**

ПРОФИЛЬ ПОДГОТОВКИ **Геоинформационные системы**
Системная инженерия программного обеспечения
КВАЛИФИКАЦИЯ (СТЕПЕНЬ) **магистр**
БАЗОВЫЙ УЧЕБНЫЙ ПЛАН ПРИЕМА **2014 г.**

КУРС **1** СЕМЕСТР **2**
КОЛИЧЕСТВО КРЕДИТОВ **3 кредита ECTS**
КОД ДИСЦИПЛИНЫ **М1.В3**

Виды учебной деятельности	Временной ресурс по очной форме обучения
Лекции, ч.	24
Лабораторные занятия, ч.	24
Аудиторные занятия, ч.	48
Самостоятельная работа, ч.	60
ИТОГО, ч.	108

ВИД ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ **зачет**

ОБЕСПЕЧИВАЮЩЕЕ ПОДРАЗДЕЛЕНИЕ **кафедра ВТ**

ЗАВЕДУЮЩИЙ КАФЕДРОЙ _____ **Марков Н.Г., профессор**

РУКОВОДИТЕЛЬ ООП _____ **Марков Н.Г., профессор**

ПРЕПОДАВАТЕЛЬ _____ **Кацман Ю.Я., доцент**

2014 г.

1. Цели освоения дисциплины

Основными целями дисциплины являются: ознакомление студентов с основными принципами анализа случайных данных; получение навыков работы и исследований с применением профессиональных математических пакетов.

При изучении данной дисциплины основное внимание уделяется методам первичной обработки данных (точечное и интервальное оценивание), проверке статистических гипотез, методам корреляционного и регрессионного анализа. Для специалистов данного направления особый интерес представляют методы кластерного и факторного анализа, анализа временных рядов, изучению которых уделяется большое внимание.

Поставленные цели полностью соответствуют целям (Ц1-Ц5) ООП.

2. Место дисциплины в структуре ООП

Дисциплина «Статистические методы обработки данных» (М1.В3) является дисциплиной вариативной части общенаучного цикла (М1).

Для её успешного усвоения необходимы **знания,** базовых понятий математического анализа, линейной алгебры и аналитической геометрии, дискретной математики, информатики, программирования, вычислительной математики, теории вероятностей, математической статистики, теории оптимизации, систем искусственного интеллекта, полученные при обучении в бакалавриате.

Содержание разделов дисциплины «Статистические методы обработки данных» согласовано с содержанием дисциплин, изучаемых параллельно (КОРРЕКВИЗИТЫ):

- «Теория графов» (М1.Б1.2).
- «Современные проблемы информационных систем и технологий» (М1.В2)

3. Результаты освоения дисциплины

В соответствии с требованиями ООП освоение дисциплины «Статистические методы обработки данных» направлено на формирование у студентов следующих компетенций (результатов обучения), в т. ч. в соответствии с ФГОС:

Таблица 1

Составляющие результатов обучения, которые будут получены при изучении данной дисциплины

Результаты обучения (компетенции из ФГОС)	Составляющие результатов обучения					
	Код	Знание	Код	Умение	Код	Владение опытом
Р1 (ОК-1, ПК-2, ПК-16)	З1.3	Методов регрессионного и дискриминантного анализа. Методов идентифицирования модели, оценки качества и параметров модели.	У1.3	Проводить точечное и интервальное оценивание экспериментальных данных. Анализировать исходные данные, выдвигать и проверять гипотезы (параметрические и непараметрические).	В1.3	

В результате освоения дисциплины «Статистические методы обработки данных» студентами должны быть достигнуты следующие результаты (табл. 2):

Таблица 2

Планируемые результаты освоения дисциплины

№ п.п	Результат
РД1	Знать методы получения "хороших оценок"; статистические методы проверки параметрических и непараметрических гипотез; методы регрессионного и дисперсионного анализа. Уметь проводить точечное и интервальное оценивание экспериментальных данных.
РД2	Знать методы идентифицирования модели, оценки качества и параметров модели. Уметь проанализировать исходные данные, выдвигать и проверять гипотезы (параметрические и непараметрические.
РД3	Владеть статистическими пакетами для обработки и анализа экспериментальных данных, в том числе пакетом Statistica

4. Структура и содержание дисциплины

Раздел 1. *Элементы математической статистики, точечное и интервальное оценивание*

Основные понятия и задачи статистики. Выборочное оценивание. Требование "хороших" оценок: несмещенность, эффективность и состоятельность. Доверительные интервалы для выборочного среднего и выборочной дисперсии. Распределения Стьюдента и "хи - квадрат".

Перечень лабораторных работ по разделу:

Лабораторная работа №1 "Первичная обработка эмпирических данных"

Раздел 2. *Проверка статистических гипотез*

Ошибки первого и второго рода. Гипотеза о равенстве двух выборочных средних, двух выборочных дисперсий. Критерий Пирсона. Колмогорова, Колмогорова – Смирнова.

Перечень лабораторных работ по разделу:

Лабораторная работа №2 "Проверка параметрических и непараметрических гипотез"

Раздел 3. *Обработка данных в рамках линейной регрессионной модели.*

Линейный корреляционный анализ. Линейная регрессия, подбор параметров прямой методами наименьших квадратов и максимального правдоподобия. Множественная регрессия.

Перечень лабораторных работ по разделу:

Лабораторная работа №3 "Решение задачи линейного корреляционного и регрессионного анализа"

Раздел 4. *Факторный анализ*

Основы факторного анализа (ранговый и дисперсионный анализ). Однофакторный анализ. Метод минимальных остатков в факторном анализе.

Лабораторная работа №4 "Исследование влияния уровней фактора в рамках однофакторного анализа".

Раздел 5. *Кластерный анализ, анализ временных рядов и прогнозирование*

Методы иерархической классификации. Многомерное шкалирование. Экспоненциальное сглаживание и прогнозирование. Спектральный (Фурье) анализ. Анализ распределенных лагов.

5. Организация и учебно-методическое обеспечение самостоятельной работы студентов

5.1 Самостоятельную работу студентов (СРС) можно разделить на текущую и творческую.

Текущая СРС – работа с лекционными и учебно-методическими материалами, включая образовательные ресурсы, представленные в сети Internet и Intranet.

Творческая проблемно-ориентированная исследовательская самостоятельная работа – заключается в поиске информации (учебники, монографии, интернет ресурсы); изучении методов статистического анализа и представления данных; проведении различных компьютерных экспериментов (исследований) в среде математического пакета **Statistica**.

5.2 Содержание самостоятельной работы студентов по дисциплине направлено на углубление и закрепление знаний, развитие практических умений и навыков и включает:

- подготовку к лабораторным работам;
- изучение теоретических учебно-методических материалов при опережающей самостоятельной работе;
- работу с литературой (интернет) при изучении тем, вынесенных на самостоятельную проработку;
- подготовку к защите лабораторных работ;
- подготовку к экзамену.

5.3 Контроль самостоятельной работы

Оценка результатов самостоятельной работы организуется как единство двух форм: самоконтроль и контроль со стороны преподавателя.

С целью самоконтроля каждая тема (раздел) в учебно-методических материалах завершается набором контрольных вопросов. Аналогичный подход реализован при выдаче индивидуальных лабораторных заданий (4 работы).

По результатам текущего контроля формируется допуск студента к экзамену. Экзамен проводится в письменной форме и оценивается преподавателем.

5.4 Учебно-методическое обеспечение самостоятельной работы студентов

Для самостоятельной работы студенты в компьютерных классах кафедры обеспечены неограниченным доступом к сетевым ресурсам Internet, учебно-методические и справочные материалы размещены в сети по адресу: ftp://ftp.vt.tpu.ru/study/Katsman/public/Statistica/.

6. Средства текущей и итоговой оценки качества освоения дисциплины

Текущий контроль по каждому разделу дисциплины осуществляется преподавателем, при этом учитывается своевременность и полноту проведенных исследований при выполнении каждой лабораторной работы. При защите работы обязательно оценивается исследовательская составляющая и знание теории.

Оценка качества освоения дисциплины производится по результатам следующих контролирующих мероприятий:

Контролирующие мероприятия	Результаты обучения по дисциплине
Выполнение лабораторной работы №1	РД1
Выполнение лабораторной работы №2	РД1 – РД2
Выполнение лабораторной работы №3	РД1 – РД3
Выполнение лабораторной работы №4	РД1 – РД3
Зачет	РД1 – РД3

Для оценки качества освоения дисциплины при проведении контролирующих мероприятий предусмотрены следующие средства:

1. **База данных контрольных вопросов для самостоятельной подготовки и подготовки к защите лабораторных работ:**
   - Свойства "хороших оценок".
   - Что такое гистограмма, полигон, кумулята, "ящик с усами"?
   - Что такое уровень значимости критерия?
   - Что показывает коэффициент детерминации?
   - Ошибка первого рода при проверке гипотезы?
   - Ошибка второго рода при проверке гипотезы?
   - Критерий Стьюдента, ограничения?
   - Критерий Фишера?
2. **Варианты индивидуальных лабораторных заданий (№1 – 4) по математической статистике**

7. Рейтинг качества освоения дисциплины

Оценка качества освоения дисциплины в ходе текущей и промежуточной аттестации обучающихся осуществляется в соответствии с «Руководящими материалами по текущему контролю успеваемости, промежуточной и итоговой аттестации студентов Томского политехнического университета», утвержденными приказом ректора № 77/од от 29.11.2011 г.

В соответствии с «Календарным планом изучения дисциплины»:

- текущая аттестация (оценка качества усвоения теоретического материала (ответы на вопросы и др.) и результаты практической деятельности (решение задач, выполнение заданий, решение проблем и др.) производится в течение семестра (оценивается в баллах (максимально 60 баллов), к моменту завершения семестра студент должен набрать не менее 33 баллов);
- промежуточная аттестация (экзамен, зачет) производится, как правило, в письменном виде в конце семестра (оценивается в баллах (максимально 40 баллов), на экзамене (зачете) студент должен набрать не менее 22 баллов).

Итоговый контроль проводится, как правило, в письменном виде. Итоговый рейтинг по дисциплине определяется суммированием баллов, полученных в ходе текущей и промежуточной аттестаций. Максимальный итоговый рейтинг соответствует 100 баллам.

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература

1. Справочник по прикладной статистике. В 2-х т. Т. 1, 2: Пер. с англ. / Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. М.: Финансы и статистика, 1989, 1990.
2. Ефимова М. Р. Общая теория статистики : учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. — 2-е изд., испр. и доп. — Москва: Инфра-М, 2013. — 416 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика [Электронный ресурс] : учебное пособие для бакалавров / В. Е. Гмурман. — 12-е изд.. — Мультимедиа ресурсы (10 директорий; 100 файлов; 740MB). — Москва: Юрайт, 2013. — 1 Мультимедиа CD-ROM. — Электронные учебники издательства Юрайт. — Электронная копия печатного издания. — Доступ из корпоративной сети ТПУ.
4. Халафян А.А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник – М.: ООО «Бином-Пресс», 2007 г. – 512 с.
5. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. 2-е изд. (+CD). – СПб.: Питер, 2003. – 688с.
6. Боровков А. А. Математическая статистика : учебник / А. А. Боровков. — 3-е изд., испр. — Москва: Физматлит, 2007. — 704 с.
7. Кремер Н.Ш. Эконометрика: учебник / Н. Ш. Кремер, Б. А. Путко; Под ред. Н. Ш. Кремера. — 3-е изд., перераб. и доп.. — Москва: ЮНИТИ-ДАНА, 2010. — 328 с.
8. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников.- М.: ФИЗМАТЛИТ, 2006,-816 с.

9. Куценко А.И. Статистический анализ технологических процессов в среде Statistica и Excel: учебное пособие / А.И. Куценко, Е.М. Черёмушкина, В.А. Щеглов, А.А. Усольцев; Новокузнецкий филиал ГОУ ВПО "Национальный исследовательский Томский политехнический университет". Томск: Изд-во Томского политехнического университета, 2010. – 277 с.

Дополнительная литература

1. Айвазян С.А. Прикладная статистика в задачах и упражнениях: учебник / С.А. Айвазян, В.С. Мхитарян. – Москва: ЮНИТИ ДАНА, 2001. – 270 с.
2. Green “Econometric Analysis”, 3d edition, Prentice Hall, 1997.
3. Hardle, W. and O. Linton (1994) Applied Nonparametric Methods, in Handbook of Econometrics, Vol. 4, Elsevier Science, North Holland.
4. Hardle, W. (1990) Applied Nonparametric Regression. Cambridge University Press.
5. Franses, P. and D. van Dijk (2000) Nonlinear Time Series Models in Empirical Finance, Chapter 5, Cambridge University Press.
6. Pagan, A. and A. Ullah (1999) Nonparametric Econometrics, Cambridge University Press.
7. Mee R.W. Simultaneous tolerance intervals for normal population with common variance //Technometrics. 1990. V. 32. № 1, P. 83-92.
8. Prescott P. An approximate test for outlier in linear models//Technometrics, 1975. V. 17, №1.P. 129-132.
9. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA® в среде Windows. Основы теории и интенсивная практика на компьютере: Учеб. пособие. - М.: Финансы и статистика, 1999. - 384 с.
10. Кацман Ю.Я. Теория вероятностей, математическая статистика и случайные процессы/Учебник (Гриф УМО). – Томск: Изд-во ТПУ, - 2013. – 130с.

Internet-ресурсы

1. http://ru.wikipedia.org/wiki
2. http://www.exponenta.ru Раздел Statistica.
3. http://www.nnspu.ru/Exponenta_Ru/educat/systemat/kabanov/main.asp.htm Проведение регрессионного анализа при помощи модуля Multiple Regressions
4. http://www.cultinfo.ru/fulltext/1/001/008/096/007.htm
5. http://lanserv2.kemsu.ru/faculty/kmc/book/matekon/Chapter9/par9_3.html Линейный регрессионный анализ
6. ftp://ftp.vt.tpu.ru/study/Katsman/public/Statistica/Lectures/New/ Кацман Ю.Я. Курс лекций по математической статистике.
7. ftp://ftp.vt.tpu.ru/study/Katsman/public/Statistica/ Кацман Ю.Я. Методические указания к лабораторным работам__1.docx.

Используемое программное обеспечение:

• StatSoft, Inc. (2003). STATISTICA (data analysis software system), version 6. www.statsoft.com.

• Borland Delphi 7 (Pascal)
• Microsoft Visual C++
• JAVA 1.6.0

9. Материально-техническое обеспечение дисциплины

При изучении данной дисциплины, самостоятельной работе студентов и выполнении лабораторных работ в компьютерных классах кафедры (ауд. 403а, б – 10 корпуса) используются современные персональные компьютеры, оснащенные лицензионным программным обеспечением и доступом в Internet.

№ п/п	Наименование оборудования	Корпус, ауд., количество установок
1.	Мультимедийное оборудование (компьютер, проектор, экран)	10 корпус ТПУ, 410 ауд.
2.	ПК Intel Core 2 E6320, 1,86 GHz.	10 корпус ТПУ, 403А ауд., 9 рабочих мест)
3.	Сетевая версия программного обеспечения серии STATISTICA 6 Ru	10 корпус ТПУ, 403А ауд., 9 рабочих мест

Программа составлена на основе Стандарта ООП ТПУ в соответствии с требованиями ФГОС по направлению 09.04.02 «Информационные системы и технологии» и профилю подготовки «Геоинформационные системы», «Системная инженерия программного обеспечения»

Программа одобрена на заседании кафедры вычислительной техники

(протокол № ____ от «___» _____ 2014 г.).

Автор Кацман Ю.Я.

Рецензент(ы) _____