

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»

СОГЛАСОВАНО

Декан ФИТ

А.С. Авдеев

Рабочая программа дисциплины

Код и наименование дисциплины: **Б1.О.29 «Основы статистической обработки экспериментальных данных»**

Код и наименование направления подготовки (специальности): **09.03.01**

Информатика и вычислительная техника

Направленность (профиль, специализация): **Программно-техническое обеспечение автоматизированных систем**

Статус дисциплины: **обязательная часть**

Форма обучения: **очная**

Статус	Должность	И.О. Фамилия
Разработал	доцент	А.Н. Тушев
	профессор	Л.И. Сучкова
Согласовал	Зав. кафедрой «ИВТиИБ»	А.Г. Якунин
	руководитель направленности (профиля) программы	Л.И. Сучкова

г. Барнаул

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Содержание компетенции	Индикатор	Содержание индикатора
ОПК-1	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1	Применяет математический аппарат, методы математического анализа и моделирования для решения задач
		ОПК-1.3	Участвует в теоретических и экспериментальных исследованиях объектов профессиональной деятельности, в обработке их результатов
ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ОПК-2.1	Выбирает информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
		ОПК-2.2	Использует современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
ОПК-9	Способен осваивать методики использования программных средств для решения практических задач	ОПК-9.1	Использует программные средства для решения практических задач на основе существующих методик

2. Место дисциплины в структуре образовательной программы

Дисциплины (практики), предшествующие изучению дисциплины, результаты освоения которых необходимы для освоения данной дисциплины.	Интегралы и дифференциальные уравнения, Линейная алгебра и теория матриц, Математический анализ, Программирование приложений
Дисциплины (практики), для которых результаты освоения данной дисциплины будут необходимы, как входные знания, умения и владения для их изучения.	Выполнение и защита выпускной квалификационной работы, Основы принятия решений в информационных системах, Преддипломная практика, Современные технологии программирования

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающегося

Общий объем дисциплины в з.е. /час: 5 / 180

Форма промежуточной аттестации: Экзамен

Форма обучения	Виды занятий, их трудоемкость (час.)				Объем контактной работы обучающегося с преподавателем (час)
	Лекции	Лабораторные работы	Практические занятия	Самостоятельная работа	
очная	32	32	0	116	76

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Форма обучения: очная

Семестр: 4

Лекционные занятия (32ч.)

1. Применение математического аппарата теории вероятности и методов математического анализа для решения задач по обработке экспериментальных данных {с элементами электронного обучения и дистанционных образовательных технологий} (6ч.)[3,5,7] Понятия классической и геометрической вероятности, правила расчета и примеры задач. Условная вероятность, формулы полной вероятности, Байеса, Бернулли, примеры использования при обработке данных. Случайные величины, функция распределения, плотность распределения. Числовые характеристики случайных величин: математическое ожидание, дисперсия, стандартное отклонение, мода, медиана, асимметрия, эксцесс. Свойства численных характеристик и примеры их вычислений для решения задач обработки экспериментальных данных.

2. Использование современных информационных технологий и программных средств для решения задач статистической обработки экспериментальных данных. Применение Python для статистических вычислений. {лекция с разбором конкретных ситуаций} (2ч.)[2,3,4,6,7] Запись на Python разветвляющихся и циклических вычислительных процессов при обработке данных. Функции работы со строками в Python, примеры. Списки и их использование для обработки массивов. Применение библиотеки numpy для обработки одномерных и двумерных числовых массивов. Чтение из файла выборки, вычисление числовых характеристик выборки: среднее, дисперсия, стандартное отклонение, мода, медиана, асимметрия, эксцесс.

3. Критерии согласия Хи-квадрат и Колмогорова-Смирнова. Выборки с заданным законом распределения. Освоение программных средств для решения практических задач. {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.)[2,3,4,5,6,7] Выполнение оценки параметров распределений в соответствии с предположением о законе распределения выполнение экспоненциального, нормального, Парето. Применение графической библиотеки Python для построения эмпирических и

теоретических функций и плотностей распределения для визуальной оценки соответствия выборки соответствующему распределению. Расчет эмпирического значения Хи-квадрат и эмпирических величин K^+ , K^- для закона Колмогорова-Смирнова. Сравнение с табличными значениями. Вывод о соответствии выборки предполагаемому распределению генеральной совокупности.

4. Описательная статистика и построение доверительных интервалов для нормального закона распределения. Применение существующих методик при разработке программного обеспечения на Python. {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.) [2,3,4,5,6,7] Построение статистического отчета описательной статистики программным кодом Python, содержащего частоты, накопленные частоты, относительные частоты и накопленные относительные частоты. Использование библиотеки matplotlib системы программирования Python для построения гистограммы и полигона частот выборки. Использование библиотеки scipy.stats для расчета доверительных интервалов для истинного среднего по функции вычисления квантилей распределения Стьюдента и доверительных интервалов истинного значения дисперсии по функции вычисления квантилей распределения Хи-квадрат.

5. Проверка статистических гипотез о равенстве средних и дисперсий нормальных генеральных совокупностей по двум выборкам. {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.) [2,3,4,5,6,7] Знакомство с библиотекой Tkinter для построения графических интерфейсов пользователя системы программирования Python. Элементы графического интерфейса: поле вывода Label, поле ввода Entry, кнопка Button, радиокнопка Radiobutton, флажок Checkbutton. Построение меню с помощью библиотеки Tkinter. Примеры. Оценивание средних, дисперсий и стандартных отклонений двух выборок нормальных генеральных совокупностей. Проверка нулевых гипотез о равенстве средних и дисперсий генеральных совокупностей. Вычисление по функциям библиотеки scipy.stats теоретических значений распределений Стьюдента и Хи-квадрат для подтверждения или опровержения нулевых гипотез.

6. Математический аппарат корреляционного анализа. Применение современных информационных технологий и программных средств для решения задач корреляционного анализа при обработке данных эксперимента {с элементами электронного обучения и дистанционных образовательных технологий} (4ч.) [2,3,4,5,6,7] Объектно-ориентированное программирование на Python. Классы и объекты. Атрибуты и методы. Разграничение доступа, пометка private полей и методов через двойное подчеркивание. Конструктор, его использование. Класс с атрибутами вспомогательного класса. Коэффициент парной корреляции Пирсона. Вычисление коэффициента парной корреляции и построение диаграммы рассеяния функцией scatter библиотеки matplotlib. проверка значимости коэффициента парной корреляции по распределению Стьюдента. Корреляционное отношение как обобщение парной корреляции. Коэффициенты частной и множественной

корреляции для нескольких выборок. Коэффициент ранговой корреляции Спирмена.

7. Регрессионный анализ. Решение задач профессиональной деятельности при обработке данных эксперимента. {с элементами электронного обучения и дистанционных образовательных технологий} (4ч.)[2,3,4,5,6,7] Производные классы в объектно-ориентированном программировании на Python. Особенность конструктора производного класса с параметрами. Линейная регрессия двух переменных. Построение прямой регрессии методом наименьших квадратов. Коэффициенты детерминации и множественной корреляции, их вычисление на Python. Варианты регрессионных зависимостей, сводящихся к линейной регрессии: гиперболическая, экспоненциальная, логарифмическая, степенная. Полиномиальная регрессия. Исследование значимости уравнения регрессии в целом по распределению Фишера и значимости отдельных коэффициентов по распределению Стьюдента в предположении нормальности выборок. Построение доверительного коридора прогнозных значений.

8. Математический аппарат факторного анализа. Применение современных информационных технологий и программных средств для решения задач факторного анализа при обработке данных эксперимента {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.)[2,3,4,5,6,7] Построение главных компонент корреляционной матрицы. Определение числа факторов по количеству собственных чисел больших единицы корреляционной матрицы и по величине в процентах накопленной дисперсии. Построение матрицы факторных нагрузок варимаксным методом через собственные вектора корреляционной матрицы. Интерпретация факторов. Вычисление количества каждого фактора у наблюдаемых переменных.

9. Дисперсионный анализ. Решение задач профессиональной деятельности при обработке данных эксперимента. {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.)[2,3,4,5,6,7] Таблица значений измеряемой величины при наборах уровней одного или двух неизмеримых факторов. Вычисление межгрупповой и внутригрупповой дисперсии. Проверка значимости влияния факторов и их взаимодействия на измеряемую величину по критерию Фишера.

10. Дискриминантный анализ. {с элементами электронного обучения и дистанционных образовательных технологий} (2ч.)[2,3,4,5,6,7] Построение канонических линейных дискриминантных функций для двух и более классов. Вычисление коэффициентов дискриминантных функций методом наименьших квадратов. Статистический анализ классификации выборки на основе статистики Уилкса и частного F-отношения.

11. Кластерный анализ. Решение задач профессиональной деятельности при обработке данных эксперимента. {лекция с разбором конкретных ситуаций} (2ч.)[2,4,6,7] Методы кластерного анализа выборки K внутригрупповых средних (K -mean) и максиминное расстояние. Особенности их использования. Библиотека ImageDraw системы программирования Python для отображения графических образов и их обработки. Применение библиотеки ImageDraw для графического

отображения построенных кластеров методами кластерного анализа.

12. Метод Arima для статистического анализа данных, заданных в виде временного ряда. {лекция с разбором конкретных ситуаций} (2ч.)[2,4,6,7] Модели авторегрессии, проинтегрированного скользящего среднего для анализа выборки в виде временного ряда. Смешанные модели. Автокорреляционная функция и частная автокорреляционная функция, их оценки библиотеками Python. Линейный метод наименьших квадратов для оценки параметров модели авторегрессии и метод Марквардта для оценки параметров скользящего среднего. Сезонные и несезонные разности модели. Построение прогнозов по модели с заданными вероятностными пределами.

Лабораторные работы (32ч.)

1. Применение математического аппарата теории вероятности и математического анализа для решения задач по обработке экспериментальных данных {творческое задание} (6ч.)[1,3,5,7] Применение математического аппарата теории вероятности и методов математического анализа для задач расчета классической, геометрической вероятностей по варианту задания. Расчет функций распределения и плотностей распределения генеральных совокупностей и их числовых характеристик по варианту задания.

2. Использование языка Python для решения практических задач по созданию и обработке структур данных на основе существующих методик, используемых в статистической обработке экспериментальных данных {творческое задание} (2ч.)[1,2,3,4,6,7] Разработать и реализовать 5 программ на Python по индивидуальному варианту. Программы включают циклический алгоритм вычисления функции, заданной в виде степенного ряда с заданной точностью или числом слагаемых, программу по обработке строк, построению и обработке одномерного массива, заданного в виде списка, использованию библиотеки numpy для построения двумерных массивов, чтение из файла выборки размерностью 20-30 элементов, вычисление ее числовых характеристик средствами стандартных команд Python и запись результатов в файл.

3. Критерии согласия Хи-квадрат и Колмогорова-Смирнова выборки с заданным законом распределения. Применение информационных технологий и программных средств для работы с данными эксперимента. {творческое задание} (2ч.)[1,2,3,4,5,6,7] Выборку по варианту задания проверить на соответствие одному из трех распределений экспоненциального, нормального и распределения Парето с оценкой соответствующих параметров, по критериям Хи-квадрат и Колмогорова-Смирнова, используя программный код на языке Python.

4. Описательная статистика и построение доверительных интервалов для нормального закона распределения. {творческое задание} (2ч.)[1,2,3,4,5,6,7] Для выборки по варианту, используя программный код на языке Python, построить таблицу описательной статистики с помощью библиотеки matplotlib

построить гистограмму и полигон частот выборки. Для выборки из нормальной генеральной совокупности по варианту найти доверительные интервалы для оценок среднего и дисперсии на уровне значимости 0,05 и 0,99.

5. Проверка статистических гипотез о равенстве средних и дисперсий нормальных генеральных совокупностей по двум выборкам. {творческое задание} (2ч.)[1,2,3,4,5,6,7] По варианту задания построить графический интерфейс пользователя на основе библиотеки Tkinter для вычисления кусочно-аналитической функции, включающий Label, Entry, Button, Radiobutton, Chckbutton. Для двух выборок из нормальной генеральной совокупности подтвердить или опровергнуть гипотезы о равенстве средних и дисперсий генеральных совокупностей на уровне значимости 0,95 и 0,99.

6. Математический аппарат корреляционного анализа. {творческое задание} (4ч.)[1,2,3,4,5,6,7] По варианту сделать на Python описание класса. Описание включает инкапсуляцию в виде private атрибутов и public методов, конструктора. Методы включают в себя ввод всех атрибутов класса, вывод значений атрибутов на консоль, инициализацию всех атрибутов и метод, индивидуальный для каждого варианта. Сделать на Python описание двух классов, где в качестве полей основного класса записаны атрибуты вспомогательного класса. Вычислить коэффициент парной корреляции двух выборок, проверить его значимость, построить диаграмму рассеяния функцией scatter библиотеки matplotlib, вычислить корреляционное отношение и сравнить с коэффициентом парной корреляции. Для трех выборок вычислить частный коэффициент корреляции и множественный коэффициент корреляции. Для баллов результатов студенческой группы вычислить ранговый коэффициент корреляции Спирмена.

7. Регрессионный анализ. Решение задач профессиональной деятельности при обработке данных эксперимента. {творческое задание} (4ч.)[1,2,3,4,5,6,7] Класс в лабораторной 6 дополнить производным классом (subclass) по варианту. Атрибуты в суперклассе сделать protected. Создать конструктор производного класса с параметрами с явным вызовом конструктора суперкласса. По двум выборкам построить линейное, гиперболическое, экспоненциальное, логарифмическое, степенное уравнение регрессии. Построить диаграммы рассеяния и уравнения регрессии на диаграммах. Вычислить коэффициенты детерминации и множественной корреляции для каждого случая. Построить уравнение регрессии третьей степени. Проверить значимость всего уравнения по критерию Фишера и значимость каждого коэффициента уравнения по критерию Стьюдента. Построить прогноз с коридором на уровне доверия 0,95.

8. Математический аппарат факторного анализа. Применение современных информационных технологий и программных средств для решения задач факторного анализа при обработке данных эксперимента {творческое задание} (2ч.)[1,2,3,4,5,6,7] По заданной аттестации или экзамена в баллах студенческой группы выполнить факторный анализ. Вычисления включают вычисление собственных значений и векторов корреляционной матрицы, построение главных компонент. По выводу массива собственных значений и накопленных дисперсий выполняется ввод числа факторов. Вычисление матрицы

факторных нагрузок и значений факторов. Интерпретация факторов.

9. Дисперсионный анализ. {творческое задание} (2ч.)[1,2,3,4,5,6,7] По заданной таблице значений величины при различных уровнях двух факторов проверить значимость каждого фактора и их взаимодействия. Программа включает вычисления межгрупповой и внутригрупповой дисперсий и проверку значимости по критерию Фишера. Вычисление значения распределения Фишера выполняется через функцию библиотеки `scipy.stats`.

10. Дискриминантный анализ. {творческое задание} (2ч.)[1,2,3,4,5,6,7] По выборке построить каноническую дискриминантную линейную функцию для разбиения векторов выборки на 2 класса. На основе статистики Уилкса и частного F-отношения выполнить вероятностную оценку качества выполненного разбиения.

11. Кластерный анализ. {творческое задание} (2ч.)[1,2,4,6,7] По выборке, содержащей набор 36-мерных векторов выполнить кластерный анализ методами K-внутригрупповых средних и максимальным расстоянием. Считая вектора бинарными образами 6×6 точек, построить кластеры в виде полутоновых изображений с помощью библиотеки `ImageDraw` системы программирования Python.

12. Метод Arima для статистического анализа данных, заданных в виде временного ряда. {творческое задание} (2ч.)[1,2,4,6,7] По выборке, заданной в виде временного ряда построить автокорреляционную и частную автокорреляционную функции. Оценить порядок взятия сезонных и несезонных разностей. По полученной информации оценить число параметров авторегрессии p и скользящего среднего q . Найти коэффициенты модели. Построить прогноз модели с вероятностным коридором на уровне 0,95.

Самостоятельная работа (116ч.)

1. Подготовка к выполнению и защите лабораторных работ. {с элементами электронного обучения и дистанционных образовательных технологий} (80ч.)[1,2,3,4,5,6,7]

2. Подготовка к сдаче экзамена. {с элементами электронного обучения и дистанционных образовательных технологий} (36ч.)[2,3,4,5,6,7]

5. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине

Для каждого обучающегося обеспечен индивидуальный неограниченный доступ к электронно-библиотечным системам: Лань, Университетская библиотека он-лайн, электронной библиотеке АлтГТУ и к электронной информационно-образовательной среде:

1. Тушев А.Н. Основы статистической обработки экспериментальных данных. Методические указания к выполнению лабораторных работ

[Электронный ресурс]: Методические указания.— Электрон. дан.— Барнаул: АлтГТУ, 2021.— Режим доступа: http://elib.altstu.ru/eum/download/ivtib/Tushev_OSOED_mu.pdf, авторизованный

6. Перечень учебной литературы

6.1. Основная литература

2. Самойленко, А. П. Информационные технологии статистической обработки данных : учебное пособие / А. П. Самойленко, О. А. Усенко. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2017. — 126 с. — ISBN 978-5-9275-2521-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <http://www.iprbookshop.ru/87418.html>. — Режим доступа: для авторизир. пользователей

3. Подопригора, И. В. Общая теория статистики : учебное пособие / И. В. Подопригора. — Томск : Томский государственный университет систем управления и радиоэлектроники, 2015. — 110 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/72144.html>. — Режим доступа: для авторизир. пользователей

4. Гребенникова, И. В. Методы математической обработки экспериментальных данных : учебно-методическое пособие / И. В. Гребенникова. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2015. — 124 с. — ISBN 978-5-7996-1456-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <http://www.iprbookshop.ru/66551.html>. — Режим доступа: для авторизир. пользователей

6.2. Дополнительная литература

5. Лихачев, А. В. Введение в теорию вероятностей и математическую статистику : учебное пособие / А. В. Лихачев. — Новосибирск : Новосибирский государственный технический университет, 2019. — 102 с. — ISBN 978-5-7782-3903-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <http://www.iprbookshop.ru/98696.html>. — Режим доступа: для авторизир. пользователей.

6. Тазиева, Р. Ф. Обработка экспериментальных данных. В 2 частях. Ч.2 : учебное пособие / Р. Ф. Тазиева, А. Н. Титов. — Казань : Казанский национальный исследовательский технологический университет, 2018. — 136 с. — ISBN 978-5-7882-2260-8, 978-5-7882-2262-2 (ч.2). — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/100573.html>. — Режим доступа: для авторизир. пользователей

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

7. www.exponenta.ru

8. Фонд оценочных материалов для проведения текущего контроля успеваемости и промежуточной аттестации

Содержание промежуточной аттестации раскрывается в комплекте контролирующих материалов, предназначенных для проверки соответствия уровня подготовки по дисциплине требованиям ФГОС, которые хранятся на кафедре-разработчике РПД в печатном виде и в ЭИОС.

Фонд оценочных материалов (ФОМ) по дисциплине представлен в приложении А.

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем

Для успешного освоения дисциплины используются ресурсы электронной информационно-образовательной среды, образовательные интернет-порталы, глобальная компьютерная сеть Интернет. В процессе изучения дисциплины происходит интерактивное взаимодействие обучающегося с преподавателем через личный кабинет студента.

№пп	Используемое программное обеспечение
1	Acrobat Reader
1	LibreOffice
2	Windows
3	PyCharm Community Edition
3	Антивирус Kaspersky
4	Python

№пп	Используемые профессиональные базы данных и информационные справочные системы
1	Бесплатная электронная библиотека онлайн "Единое окно к образовательным ресурсам" для студентов и преподавателей; каталог ссылок на образовательные интернет-ресурсы (http://Window.edu.ru)
2	Национальная электронная библиотека (НЭБ) — свободный доступ читателей к фондам российских библиотек. Содержит коллекции оцифрованных документов (как открытого доступа, так и ограниченных авторским правом), а также каталог изданий, хранящихся в библиотеках России. (http://нэб.рф/)

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Наименование специальных помещений и помещений для самостоятельной работы
учебные аудитории для проведения учебных занятий

Наименование специальных помещений и помещений для самостоятельной работы
помещения для самостоятельной работы

Материально-техническое обеспечение и организация образовательного процесса по дисциплине для инвалидов и лиц с ограниченными возможностями здоровья осуществляется в соответствии с «Положением об обучении инвалидов и лиц с ограниченными возможностями здоровья».