

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

(факультет / институт / филиал)

УТВЕРЖДЕНО
решением ученого совета ННГУ
протокол от
« 30 » _ноября___ 2022 г. № _13_

Рабочая программа дисциплины

Прикладная математическая статистика

(наименование дисциплины (модуля))

Уровень высшего образования

магистратура

(бакалавриат / магистратура / специалитет)

Направление подготовки / специальность

02.04.02 Фундаментальная информатика и информационные технологии

(указывается код и наименование направления подготовки / специальности)

Направленность образовательной программы

Вероятностное моделирование и анализ данных

(указывается профиль / магистерская программа / специализация)

Форма обучения

очная

(очная / очно-заочная / заочная)

Нижний Новгород

2023 год

1. Место дисциплины в структуре ОПОП

Дисциплина относится к обязательной части

Б1.О.05 Прикладная математическая статистика

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	Блок 1. Дисциплины (модули) Обязательная часть	Дисциплина Б1.О.05, «Прикладная математическая статистика» относится к обязательной части ООП направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии».

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции* (код, содержание индикатора)	Результаты обучения по дисциплине**	
ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	ОПК-3.1. Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей.	Знать следующие понятия математической статистики: 1) выборка 2) выборочное распределение и выборочные характеристики 3) точечная оценка параметра 4) простая и сложная статистические гипотезы 5) критерии согласия Колмогорова–Смирнова и Пирсона 6) таблица сопряженности, критерии независимости 7) кривая регрессии	Собеседование
	ОПК-3.2. Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем.	Уметь; 1) оценивать параметры распределений стандартными методами 2) формулировать и проверять статистические гипотезы сообразно с характером задачи	Задачи (практические задания)
	ОПК-3.3. Имеет практический опыт применения разработки программного	Владеть навыками выполнения стандартных статистических процедур с помощью профессионального программного обеспечения	Задачи (практические задания)

	<i>обеспечения и тестирования программных продуктов.</i>		
ПК-3. Способен эксплуатировать, разрабатывать и развивать системное и прикладное программное обеспечение, новые информационные технологии на основе анализа современного состояния науки и информационных технологий, и управлять такой эксплуатацией и разработкой в области профессиональной деятельности.	ПК-3.1. Знает основы информационных технологий и имеет навыки анализа современного состояния науки и информационных технологий в области профессиональной деятельности.	Знать основные особенности профессионального статистического программного обеспечения	Собеседование
	ПК-3.2. Умеет применять навыки проектирования, разработки и развития решений в области информационных технологий на основе анализа современного состояния науки и информационных технологий в области профессиональной деятельности.	Уметь реализовывать алгоритмы математической статистики на языке высокого уровня	Задачи (практические задания)
	ПК-3.3. Имеет практический опыт управления разработкой и развитием решений в области информационных технологий на основе анализа современного состояния науки и информационных технологий в области профессиональной деятельности	Владеть навыками отладки программного кода на языке программирования R	Задачи (практические задания)

3. Структура и содержание дисциплины

3.1. Трудоемкость дисциплины

	Очная форма обучения
Общая трудоемкость	4 ЗЕТ
Часов по учебному плану	144
в том числе	
аудиторные занятия (контактная работа):	66
- занятия лекционного типа	32
- занятия семинарского типа	32
- занятия лабораторного типа	0
- текущий контроль (КСР)	2
самостоятельная работа	42
Промежуточная аттестация – экзамен	36

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы) Очная	В том числе				
		Контактная работа (работа во взаимодействии с преподавателем), часы. Из них				Самостоятельная работа обучающегося, часы Очная
		Занятия лекционного типа Очная	Занятия семинарского типа Очная	Занятия лабораторного типа Очная	Всего Очная	
Введение в R. Виды данных. Генерация псевдослучайных чисел с различным законом распределения. Метод Монте-Карло.	8	4	2		6	2
Выборочное распределение и выборочные статистики. Ядерные оценки плотности.	14	4	4		8	6
Статистические гипотезы и виды ошибок. Критерии согласия. Проверка гипотез нормальности и экспоненциальности.	18	4	6		10	8
Доверительные интервалы. Методы получения точечных оценок.	20	6	6		12	8
Проверка гипотез о параметрах распределений. Лемма Неймана-Пирсона. t -распределение и F -распределение	12	4	4		8	4
Непараметрические гипотезы однородности, независимости. Ранговые критерии	20	6	6		12	8
Регрессионный анализ	8	2	2		4	4
Однофакторный дисперсионный анализ	6	2	2		4	2
Текущий контроль (КСР)	2	0	0		2	0
Промежуточная аттестация – экзамен/зачет	36	0	0		36	0
Итого	144	32	32		102	42

Текущий контроль успеваемости реализуется в формах опросов на занятиях семинарского типа

Промежуточная аттестация проходит в традиционных формах (экзамен)

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа заключается в чтении литературы из списка основной литературы, решения практических заданий и подготовке к промежуточной аттестации.

Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

Для обеспечения самостоятельной работы обучающихся используется электронный курс «Прикладная математическая статистика» (<https://e-learning.unn.ru/course/view.php?id=6164>), созданный в системе электронного обучения ННГУ – <https://e-learning.unn.ru/>.

5. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), включающий:

5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности компетенций						
	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	Не зачтено		Зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки.	Минимально допустимый уровень знаний. Допущено много негрубых ошибок.	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько незначительных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки.	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме.	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения, решены все основные задачи с отдельными незначительными недочетами, выполнены все задания в полном	Продemonстрированы все основные умения, решены все основные задачи. Выполнены все задания, в полном объеме без недочетов

				недочетами.		объеме.	
<u>Навыки</u>	Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки.	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами.	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продемонстрирован творческий подход к решению нестандартных задач.

Шкала оценивания при собеседовании

Результат ответов	Оценка
Студент дает верные развернутые ответы на вопросы преподавателя	зачтено
Студент отвечает только на часть вопросов, или отвечает с ошибками, или не дает развернутого ответа на вопросы	не зачтено

Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	Отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо»
	Удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	Неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.2. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

5.2.1 Контрольные вопросы

вопросы	Код формируемой компетенции
1. Выборка, выборочное распределение одномерной случайной величины. Построение эмпирической плотности и эмпирической функции распределения в пакете R.	ОПК-3
2. Выборочные числовые характеристики случайных величин. Оценка основных выборочных числовых характеристик в пакете R.	ОПК-3
3. Типы статистических данных (числовые, порядковые, номинальные). Частотные распределения (одномерные и многомерные). Маргинальные частотные распределения. Проблема нахождения выборочных числовых характеристик по группированным данным. Ее решение в пакете R.	ПК-3
4. Графические методы представления и анализа данных: гистограммы, графики «ящик с усами», диаграммы рассеяния, матрицы диаграмм рассеяния. Использование цвета для выявления группировок.	ОПК-3
5. Метод получения выборки значений случайной величины с помощью генератора псевдослучайных чисел. Метод обратной функции, метод исключения.	ОПК-3
6. Вычисление интегралов методом Монте – Карло в пакете R. Неоднозначность в разложении подынтегральной функции и ее влияние на эффективность метода.	ОПК-3
7. Точечная оценка параметра. Метод максимального правдоподобия. Оценки максимального правдоподобия для нормального и экспоненциального распределений в пакете R.	ОПК-3
8. Точечная оценка параметра. Метод максимального правдоподобия. Оценки максимального правдоподобия для геометрического и пуассоновского распределений в пакете R.	ОПК-3
9. Точечная оценка параметра. Метод максимального правдоподобия. Оценки максимального правдоподобия для смесей распределений в пакете R.	ОПК-3
10. Формула Байеса. Условное распределение. Байесовское оценивание вероятности успеха с дискретным и непрерывным априорным распределением.	ОПК-3
11. Формула Байеса для плотностей. Условное распределение. Байесовское оценивание дисперсии нормального закона.	ОПК-3
12. Односторонние и двухсторонние доверительные интервалы для случайной величины. Построение доверительных интервалов с помощью квантилей. Процедура построения в пакете R.	ОПК-3
13. Простые и сложные гипотезы. Ошибки первого и второго рода. Критерий согласия Колмогорова – Смирнова. Применение критерия в пакете R.	ОПК-3
14. Простые и сложные гипотезы. Ошибки первого и второго рода. Критерий согласия хи-квадрат Пирсона. Применение критерия в пакете R.	ОПК-3
15. Методы проверки нормальности данных при неизвестных параметрах распределения (вероятностная бумага, критерии Лилли – Фора).	ОПК-3
16. Методы проверки нормальности данных при неизвестных параметрах распределения (критерий хи-квадрат Пирсона, критерий Шапиро – Уилкса).	ОПК-3
17. Построение наиболее мощных критериев для проверки простых параметрических гипотез на основе леммы Неймана – Пирсона (лемма без доказательства). Примеры критериев с реализацией в пакете R.	ОПК-3
18. Задача о сравнении средних в двух нормальных совокупностях. Критерий Стьюдента. Применение в пакете R.	ОПК-3
19. Задача об обнаружении эффекта обработки в предположении о совместном нормальном распределении. Критерий Стьюдента для проверки соответствующей гипотезы. Проведение теста в пакете R.	ОПК-3

20. Задача о сравнении центров распределения в двух выборках при неизвестных законах распределения. Непараметрические критерии. Критерий Манна – Уитни.	ОПК-3
21. Задача о сравнении центров распределения в нескольких выборках при неизвестных законах распределения. Критерий Краскалла – Уоллиса. Работа с критерием в пакете R.	ОПК-3
22. Таблицы сопряженности признаков. Критерий независимости хи-квадрат для проверки независимости двух переменных. Оценки максимального правдоподобия при нулевой гипотезе. Работа с критерием в пакете R.	ОПК-3
23. Непараметрические критерии независимости. Ранговый критерий независимости Спирмена. Работа с критерием в пакете R.	ОПК-3
24. Модель Гаусса – Маркова простой линейной регрессии. Оценка параметров и анализ остатков. Решение данных задач в пакете R.	ОПК-3
25. Модель Гаусса – Маркова простой линейной регрессии. Проверка значимости регрессии, изменение модели и прогноз новых значений в пакете R.	ОПК-3
26. Модель линейной множественной регрессии. Ее применение для выявления нелинейных зависимостей для одной и нескольких переменных. Работа с этими моделями в пакете R.	ОПК-3

5.2.2. Типовые задания/задачи для оценки сформированности компетенции «ОПК-3»

1. Что делают следующие функции в среде R: `mean(x)`, `var(x)`, `sd(x)`, `median(x)`, `ecdf(x)`? Приведите расчетные формулы.

2. Напишите код, который сгенерирует выборку из стандартного нормального распределения, постройте гистограмму распределения и наложите на нее график нормальной плотности с параметрами, найденными по выборке.

3. При вызове команды

```
ks.test(subset(iris, iris$Species=='setosa')$Sepal.Length, function(x)
pnorm(x, 5, 0.4))
```

был получен следующий вывод:

```
One-sample Kolmogorov-Smirnov test
```

```
data: subset(iris, iris$Species == "setosa")$Sepal.Length
D = 0.12129, p-value = 0.4537
alternative hypothesis: two-sided
```

Поясните, какая задача решалась и как следует интерпретировать полученный результат.

4. Объясните, что делает следующая последовательность команд:

```
x <- apply(matrix( runif( 1200, -0.5, 0.5 ), ncol=12), 1, mean)
```

Какой закон распределение должен хорошо приближать выборку x?

5. Напишите, как следует подобрать модель регрессии для переменных Girth (обхват) и Volume (Объем) в кадре данных trees (деревья) и как можно оценить ее параметры с помощью команд R.

6. Что делает следующая последовательность команд:

```
mydf <- function(x, a1, s1, a2, s2, p) {
  0.5*(dnorm(x, a1, s1)+dnorm(x, a2, s2))
};
fitdistr( data, mydf, list (a1=0, s1=1, a=1, s2=1, p=0.5))
```


7. Предположим, что кадр данных `sera` содержит две переменные: переменная `x` указывает процент содержания серы в образце стали, а переменная `y` – имя одного из двух поставщиков стали. Напишите, как с помощью функции `t.test` проверить гипотезу об одинаковом содержании серы в обеих группах продукции против альтернативы, что у первого поставщика содержание серы ниже.

8. В кадре данных `iris` содержатся сведения о четырех обмерах (переменные `Sepal.Width`, `Sepal.Length`, `Petal.Width`, `Petal.Length`) цветков ириса трех разных сортов. Как можно визуализировать диаграммы рассеяния для всех пар переменных с раскраской по сорту цветка? Как нарисовать графики типа «ящик с усами» для каждой переменной в зависимости от сорта цветка?

5.2. 4. Типовые задачи для оценки сформированности компетенции «ПК-3»:

1. Напишите реализацию функции для проверки гипотезы о величине дисперсии стандартного закона на основании леммы Неймана–Пирсона

5.2.5 Вопросы для собеседования для оценки компетенции «ОПК-3»:

1. Какие виды категориальных переменных вы знаете?
2. Что такое выборочное математическое ожидание, выборочная дисперсия?
3. Что такое ошибка первого рода?
4. Что такое сложная гипотеза?
5. Как получить выборку из показательного распределения методом обратной функции?
6. Что такое критерий согласия?

5.2.6. Вопросы для собеседования для оценки компетенции «ПК-3»:

1. Как представляются данные разных типов в статистических программах?
2. Какие требования разумно предъявлять к профессиональной среде статистической обработки данных?
3. Что такое кадр данных? Какие с ним операции желательно иметь?
4. Что такое группирующая переменная?

6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

1. Федоткин М.А. Основы прикладной теории вероятностей и статистики. – М.: Высшая школа, 2006. – 368 с. (Библиотека ННГУ – 185 экз.)
2. Зорин А.В., Федоткин М.А. Введение в прикладной статистический анализ в пакете R: Учебно-методическое пособие. — Нижний Новгород: ННГУ, 2010. — 50 с.
<http://www.unn.ru/pages/e-library/methodmaterial/files/3.pdf> (Фонд Учебно-методических материалов, подготовленных в рамках реализации проекта "Нижегородский государственный университет им. Лобачевского - Национальный исследовательский университет")

б) дополнительная литература:

1. Ивченко Г.И., Медведев Ю.И. Математическая статистика: [учеб. пособие для вузов]. – М.: Высшая школа, 1984 г. – 248 с. (Библиотека ННГУ – 24 экз.)
2. Бочаров П. П., Печинкин А. В. Теория вероятностей. Математическая статистика: учеб. пособие. – М.: ФИЗМАТЛИТ, 2005. – 296 с. (Адрес интернет: <http://znanium.com/catalog/product/405754>)

в) программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

- Сайт проекта R для установки свободного программного продукта <http://www.r-project.org>

7. Материально-техническое обеспечение дисциплины

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: компьютерный класс, проектор, экран.

Используемое лицензионное программное обеспечение:

- Операционные системы семейства Microsoft Windows, лицензия по подписке Microsoft Imagine.
- Любая свободно распространяемая реализация языка R (лицензия GNU).

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду

Программа составлена в соответствии с требованиями ОС ННГУ по направлению 02.04.02 Фундаментальная информатика и информационные технологии.

Автор д.ф.-м.н., доцент _____ А.В.Зорин

Рецензент (ы) _____

Заведующий кафедрой _____ А.В. Зорин

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики
от 30.11.2022 года, протокол № 3.