

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
радиотехники и компьютерных
технологий**

А.В. Дворкович

	Рабочая программа дисциплины (модуля)
по дисциплине:	Статистический анализ данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра инфокоммуникационных систем и сетей
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет

8 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 60 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: Е.Н. Платонов, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры инфокоммуникационных систем и сетей 09.04.2020

Аннотация

В данном курсе изучаются основные понятия математической статистики и методы статистического анализа эмпирических данных. Для того чтобы начать изучение прикладных методов анализа статистических данных прежде всего обучающиеся знакомятся с основными методами математической статистики. Описывается постановка и формулируются подходы к решению проблемы первичного анализа статистических данных. Рассматриваются параметрические и непараметрические методы проверки гипотез. Изучается проблема обнаружения зависимости статистических данных, измеряемых в различных шкалах. Также рассмотрены некоторые современные методы анализа данных, например, подробно изучаются непараметрические робастные методы, которые можно использовать для обработки информации в условиях априорной статистической неопределенности, свойственной реальным статистическим экспериментам. Для успешного прохождения курса необходимо посещение и конспектирование лекций, выполнение практических заданий и самостоятельная работа с дополнительными литературными источниками.

1. Цели и задачи

Цель дисциплины

изучение основ статистического анализа данных – направления математики, относящегося к разработке методов получения научно обоснованных выводов о массовых явлениях и процессах по данным наблюдений или экспериментов.

Задачи дисциплины

- формирование базовых знаний в области математической статистики как части теории вероятности;
- обучение студентов понятиям выборочной теории, точечным и интервальным оценкам, проверки гипотез, основам корреляционного и, регрессионного и дисперсионного анализов, непараметрическим методам статистики;
- формирование подходов к обработке массивов статистических данных при изучении разнообразных задач: природных, в том числе биологических, инфокоммуникационных, экономических, демографических, социологических и пр.;
- выполнение студентами научных исследований;
- оказание консультаций и помощи студентам в проведении теоретических и экспериментальных работ.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-4 Способен осуществлять сбор и обработку данных, технической и (или) иной информации	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности

обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- ☐ место и роль общих вопросов науки в научных исследованиях;
- ☐ фундаментальные понятия, законы, теории классической и современной физики;
- ☐ современные проблемы физики, химии, математики;
- ☐ теоретические модели фундаментальных процессов и явлений в физике и ее приложениях;
- ☐ новейшие открытия естествознания в биологии мозга;
- ☐ постановку проблем моделирования в сложных биологических системах.

уметь:

- ☐ эффективно использовать на практике теоретические компоненты науки: понятия, суждения, умозаключения, законы;
- ☐ представлять панораму универсальных методов и законов современного естествознания;
- ☐ пользоваться своими знаниями для решения фундаментальных и прикладных задач и технологических задач;
- ☐ делать правильные выводы из сопоставления результатов теории и эксперимента;
- ☐ производить численные оценки по порядку величины;
- ☐ делать качественные выводы при переходе к предельным условиям в изучаемых проблемах;
- ☐ видеть в технических задачах физическое содержание;
- ☐ осваивать новые предметные области, теоретические подходы и экспериментальные методики;
- ☐ работать на современном экспериментальном оборудовании;
- ☐ абстрагироваться от несущественных влияний при моделировании реальных физических ситуаций;
- ☐ эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

владеть:

- ☐ навыками освоения большого объема информации;
- ☐ навыками самостоятельной работы в лаборатории и Интернете;
- ☐ культурой постановки и моделирования физических задач;
- ☐ навыками грамотной обработки результатов опыта и сопоставления с теоретическими данными;
- ☐ практикой исследования и решения теоретических и прикладных задач;
- ☐ навыками теоретического анализа реальных задач, связанных со свойствами систем, обладающих как дискретным, так и непрерывным спектрами
- ☐ планированием, постановкой и обработкой результатов физического эксперимента;
- ☐ научной картиной мира;
- ☐ математическим моделированием физических задач.

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

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост.

		лекции	семинары	лаборат. работы	работа
1	Основные понятия выборочной теории	3			1
2	Точечные оценки	3			2
3	Интервальные оценки и доверительные интервалы	3			4
4	Проверка гипотез. Параметрические модели	3			2
5	Проверка непараметрических гипотез	3			2
6	Основы корреляционного анализа	3			2
7	Основы регрессионного анализа	6			1
8	Основы дисперсионного анализа	6			1
9	Непараметрические методы статистики	10			10
10	Гауссовские марковские случайные процессы	10			10
11	Оптимальные фильтрация и прогнозирование	10			10
Итого часов		60			45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 7 (Осенний)

1. Основные понятия выборочной теории

Генеральная совокупность. Выборка. Статистический ряд. Вариационный ряд. Эмпирическая функция распределения. Теорема Гливенко. Гистограмма. Полигон частот. Выборочные числовые моменты. Примеры и решения типовых задач.

2. Точечные оценки

Состоятельные, несмещенные и эффективные оценки. Достаточные статистики. Методы получения точечных оценок. Примеры и решения типовых задач.

3. Интервальные оценки и доверительные интервалы

Понятия интервальной оценки и доверительного интервала. Сведения о некоторых распределениях, связанных с нормальным распределением. Построение интервальных оценок. Примеры и решения типовых задач.

4. Проверка гипотез. Параметрические модели

Основные понятия. Проверка двух простых гипотез. Критерий Неймана-Пирсона. Сложные параметрические гипотезы. Последовательный критерий отношения правдоподобия. Примеры и решения типовых задач.

5. Проверка непараметрических гипотез

Критерии согласия (простая гипотеза). Критерии согласия (сложная гипотеза). Критерии независимости. Примеры и решения типовых задач.

6. Основы корреляционного анализа

Исходные понятия. Анализ парных связей. Анализ коэффициента корреляции. Анализ корреляционного отношения. Анализ множественных связей. Примеры и решения типовых задач.

7. Основы регрессионного анализа

Исходные предположения. Метод наименьших квадратов. Статистический анализ регрессионной модели. О выборе допустимой модели регрессии. Примеры и решения типовых задач.

8. Основы дисперсионного анализа

Однофакторный дисперсионный анализ. Линейные контрасты. Двухфакторный дисперсионный анализ. Примеры и решения типовых задач.

Семестр: 8 (Весенний)

9. Непараметрические методы статистики

Одновыборочная задача о сдвиге. Двухвыборочная задача о сдвиге. Критерий знаков. Критерий Вилкоксона, Манна и Уитни. Критерий серий. Ранговая корреляция. Примеры и решения типовых задач.

10. Гауссовские марковские случайные процессы

Гауссовские марковские случайные последовательности. Случайные процессы и марковское свойство. Гауссовские марковские случайные процессы. Аппроксимация гауссовского марковского случайного процесса с помощью гауссовской марковской последовательности. Координаты состояния системы и марковское свойство. Примеры и решения типовых задач.

11. Оптимальная фильтрация и прогнозирование

Оценивание параметров методом взвешенных наименьших квадратов. Оптимальная фильтрация для линейных одношаговых переходов. Оптимальная фильтрация и прогнозирование линейных многошаговых процессов. Оптимальная фильтрация непрерывных линейных динамических систем с непрерывными измерениями. Примеры и решения типовых задач.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Необходимое оборудование для лекций: компьютер и мульти-медийное оборудование (проектор, звуковая система).

Обеспечение самостоятельной работы - базы данных по журналам Physica Status Solidi b, Physical Review, J of Apply Physics.

6. Перечень рекомендуемой литературы

Основная литература

1. Введение в математическую статистику [Текст] : [учебник для вузов] / Г. И. Ивченко, Ю. И. Медведев .— М. : ЛКИ, 2010, 2014, 2015 .— 600 с.
2. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.

Дополнительная литература

1. Математическая статистика [Текст]/Б. Л. Ван дер Варден, -М., Иностранная литература, 1960
2. Линейная и нелинейная регрессии [Текст]/Е. З. Демиденко, -М., Финансы и статистика, 1981
3. Теория ранговых критериев [Текст]/Я. Гаек, З. Шидак, -М., Наука, 1971

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. <http://lib.mipt.ru> – электронная библиотека Физтеха.
2. <http://www.edu.ru> – федеральный портал «Российское образование».
3. <http://benran.ru> – библиотека по естественным наукам Российской академии наук.
4. <https://scikit-learn.org/> – анализ данных с помощью Python.
5. <https://r-analytics.blogspot.com/> – анализ и визуализация данных в программном среде R.
6. <https://www.kaggle.com/> – система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению.
7. Журналы по физике и математике (Physical Review B, АиТ, Теория вероятностей и ее применения и др.), доступные через Internet научные и научно-технические журналы: <http://scitation.aip.org/>, <http://jre.cplire.ru/jre/jan12/8/text.html>, <http://elibrary.ru/>.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций (Microsoft PowerPoint).

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения, понятия.

Успешное освоение курса требует напряжённой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам лекций, учебной и научной литературе);
- подготовку к дифференцированному зачету и экзамену.

Руководство и контроль за самостоятельной работой студента осуществляется в форме индивидуальных консультаций.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Радиотехника и компьютерные технологии Физтех-школа Радиотехники и Компьютерных Технологий кафедра инфокоммуникационных систем и сетей
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет

8 (весенний) - Экзамен

Разработчик: Е.Н. Платонов, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

2. Показатели оценивания компетенций

В результате изучения дисциплины «Статистический анализ данных» обучающийся должен:

знать:

- ☐ место и роль общих вопросов науки в научных исследованиях;
- ☐ фундаментальные понятия, законы, теории классической и современной физики;
- ☐ современные проблемы физики, химии, математики;
- ☐ теоретические модели фундаментальных процессов и явлений в физике и ее приложениях;
- ☐ новейшие открытия естествознания в биологии мозга;
- ☐ постановку проблем моделирования в сложных биологических системах.

уметь:

- ☐ эффективно использовать на практике теоретические компоненты науки: понятия, суждения, умозаключения, законы;
- ☐ представлять панораму универсальных методов и законов современного естествознания;
- ☐ пользоваться своими знаниями для решения фундаментальных и прикладных задач и технологических задач;
- ☐ делать правильные выводы из сопоставления результатов теории и эксперимента;
- ☐ производить численные оценки по порядку величины;
- ☐ делать качественные выводы при переходе к предельным условиям в изучаемых проблемах;
- ☐ видеть в технических задачах физическое содержание;
- ☐ осваивать новые предметные области, теоретические подходы и экспериментальные методики;
- ☐ работать на современном экспериментальном оборудовании;
- ☐ абстрагироваться от несущественных влияний при моделировании реальных физических ситуаций;
- ☐ эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

владеть:

- ☐ навыками освоения большого объема информации;
- ☐ навыками самостоятельной работы в лаборатории и Интернете;
- ☐ культурой постановки и моделирования физических задач;
- ☐ навыками грамотной обработки результатов опыта и сопоставления с теоретическими данными;
- ☐ практикой исследования и решения теоретических и прикладных задач;
- ☐ навыками теоретического анализа реальных задач, связанных со свойствами систем, обладающих как дискретным, так и непрерывным спектрами
- ☐ планированием, постановкой и обработкой результатов физического эксперимента;
- ☐ научной картиной мира;
- ☐ математическим моделированием физических задач.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

С целью текущего контроля в течение семестра проводятся контрольные работы.

Задания для контрольной работы в первом семестре.

Вариант 1

1. Найдите оценку параметра θ с помощью метода максимального правдоподобия по выборке $\{X_1, \dots, X_n\}$ соответствующей распределению $N(\theta, 1)$. Доказать, что выборочное среднее является несмещенной, состоятельной и эффективной оценкой математического ожидания.
2. Случайный вектор $Z = \{X, Y\}$ имеет плотность вероятности . Вычислить $P(X > Y)$.
3. На последних выборах в N-кой области партия X получила 30% голосов избирателей в городах и 70% в сельской местности. Сколько нужно опросить избирателей области, чтобы оценить поддержку партии X на ближайших выборах с точностью до 3% и доверительной вероятностью 0,95? Городское население в области составляет 60%.
4. Известно, что примерно 83% студентов факультета менеджмента защищают диплом на отличную оценку. На основе наблюдений этого года было выяснено, что из 100 случайно отобранных дипломников отличную оценку получил 91 студент. Можно ли с 95% уровнем доверия сказать, что наблюдения этого года не противоречат ранее сделанному статистическому выводу?

Вариант 2

1. Получены два наблюдения $k = 1, 2$ случайной величины $X \sim N(1; 4)$. Ошибки наблюдения не зависят друг от друга и от X и имеют распределение $N(0; 1)$. Найти с.к.-оптимальную оценку для X по наблюдениям y_1 и y_2 .
2. За кандидата на выборную должность по предварительным данным собирается голосовать от 30 до 40% избирателей пенсионного возраста и от 60 до 70% остальных избирателей. Как велико должно быть количество опрошенных жителей этого избирательного округа, чтобы можно было оценить долю голосов за данного кандидата с точностью не менее 0,01 на уровне доверия 0,95? Избиратели пенсионного возраста составляют 25% от общего числа избирателей.

3. Перед экспертами поставлена задача оценить сравнительную активность электората Москвы и Санкт-Петербурга при избрании депутатов Государственной Думы. С этой целью была осуществлена случайная выборка в двух этих городах из состава населения, которое имеет право голоса. Было выяснено, какая часть выборки реально пришла на избирательный участок для участия в выборах. Данные оказались следующими: в Москве из 1500 потенциальных случайно выбранных избирателей реально в выборах приняли участие 480 человек, а в Санкт-Петербурге из 1630 потенциальных избирателей на избирательные участки пришли 490 человек. На уровне значимости 0,05 проверить гипотезу о равенстве генеральных долей избирателей в двух этих городах, реально принявших участие в выборах.

Вариант 3

1. Найдите оценку параметра θ с помощью метода максимального правдоподобия по выборке $\{X_1, \dots, X_n\}$ соответствующей распределению $N(\theta, 1)$. Доказать, что выборочное среднее является несмещенной, состоятельной и эффективной оценкой математического ожидания.
2. Главный врач ветеринарной клиники утверждает, что не менее 70 % его пациентов после приёма новейших лекарств верно служат своим хозяевам не более на протяжении трех лет. Можно ли считать это утверждение верным, если из 100 пациентов ветеринарной клиники 59 остаются здоровыми после приема таблеток ещё три года? Принять уровень значимости 0,05.
3. В некотором районе, насчитывающем 4000 семей, нужно оценить долю семей, имеющих более двух телевизоров. По предварительным данным эта доля приблизительно равна 10%. Каково должно быть число обследованных семей, чтобы с вероятностью не менее 0,97 можно было оценить эту долю с точностью 0,03.
4. Тысячу человек классифицировали по признаку дальтонизма. Среди 480 мужчин оказалось 38 дальтоников, среди 520 женщин 6 дальтоников. Можно ли считать, что вероятность заболевания дальтонизмом у мужчин и женщин одинакова?

Вариант 4

1. Найдите оценку параметра θ с помощью метода максимального правдоподобия по выборке $\{X_1, \dots, X_n\}$ соответствующей распределению $N(0, \theta)$. Доказать, что оценка является эффективной.
2. В стране 5% трудоспособного населения имеют сверхвысокие доходы, 10% - высокие доходы и 85% - средние и низкие. Правительство решило ввести прогрессивную ставку подоходного налога. Сколько людей трудоспособного возраста нужно опросить, чтобы с доверительной вероятностью 0,98 получить оценку населения этого предложения с точностью до 0,01. При этом известно, что 80-90% людей с высокими доходами против этого предложения.
3. В течение месяца завод изготовил 200 корпусов, из которых 3 оказалось дефектными. В следующий месяц было изготовлено 350 корпусов, из которых 7 оказалось дефектными. Изменилась ли доля дефектных корпусов?
4. Пусть ковариационная матрица гауссовского вектора Z имеет вид $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, математическое ожидание $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Найти оценку первой компоненты вектора Z по наблюдению за двумя другими случайными величинами Z_2 и Z_3 .

Задания для контрольной работы во втором семестре.

Вариант 1

1. Проведено исследование по оценке роли чистой мотивации (знания цели работы) на выполнение скучных монотонных операций при вытачивании металлической заготовки. Восемнадцать рабочих одинаковой квалификации были случайным образом разделены на три группы. Группа А не имела информации о требуемой производительности труда, группа В получила лишь общие сведения, группа С имела точную информацию о задании.

Количество обработанных заготовок:

А: 40; 35; 38; 43; 44; 41; 42.

В: 38; 40; 47; 44; 40; 42.

С: 48; 40; 45; 43; 46; 44.

Можно ли на уровне значимости 0,05 считать, что производительность труда растет с осведомленностью?

2. У каждого из 6800 респондентов измерялось два признака. Признак А (цвет глаз) имеет три градации: карий, серый, зеленый. Признак В (цвет волос) имеет градации: брюнет, шатен, блондин.

	Брюнет	Шатен	Блондин
Карий	1768	806	236
Серый	946	1387	800
Зеленый	115	438	304

Имеется ли статистическая зависимость между признаками А и В на уровне значимости 0,05?

3. Согласно опросам 29 семей, проводившимся в 1968 г. в юго-западном регионе Англии, выборочное среднее еженедельной арендной платы за мебелированную квартиру составило 2,5 фунта, а выборочная дисперсия 0,67 фунта². В Уэльсе выборочное среднее арендной платы 16 семей составило 2,06 фунта, а выборочная дисперсия 0,42 фунта². Проверьте, является ли различие арендной платы в этих регионах статистически значимым, предполагая, что выборки порождены гауссовскими случайными величинами.

4. Выборочная корреляционная матрица трёх экономических показателей X_1 , X_2 , X_3 имеет вид

1	-0.66	-0.94
-0.66	1	0.71
-0.94	0.71	1

Вычислить оценку множественного коэффициента корреляции и про верить гипотезу о равенстве нулю множественного коэффициента корреляции.

Вариант 2

1. Три группы случайно отобранных людей обучались навыкам скорочтения тремя разными методами. В конце обучения проводился зачет, на котором оценивалась скорость чтения. Обучающиеся показали следующие результаты (страниц за десять минут).

Первая группа: 20, 23, 24, 24, 25, 26, 28, 30, 31, 32.

Вторая группа: 38, 42, 42, 44, 47, 48, 49, 50, 51.

Третья группа: 29, 32, 33, 35, 35, 37, 38, 39, 40, 42, 30

Можно ли считать на уровне значимости 0,05, что предлагаемые методы обучения имеют различную эффективность?

2. Спортсмены, ранги которых построении по росту были равны 1,2,...,10, показали в прыжке в длину следующие результаты (в м)

6,94; 7,12; 7,01; 6,98; 7,24; 7,42; 7,13; 6,95; 7,34; 7,82. Имеется ли зависимость между ростом спортсмена и длиной его прыжка?

3. Центром исследования гражданского общества и некоммерческого сектора НИУ ВШЭ была сформирована репрезентативная выборка. Респондентов попросили ответить, удовлетворены ли они своими доходами.

Из 1091 участвовавших в опросе женщин 120 затруднились ответить, 831 ответили отрицательно и 140 положительно.

Из 909 мужчин 19 затруднились ответить, 652 ответили отрицательно и 238 положительно.

Вычислить меры прогноза Гутмана.

4. Уровень гистамина в мокроте у семи курильщиков, склонных к аллергии, составил (в микрограммах):

102,4; 100,0; 67,5; 65,9; 64,7; 39,6; 31,2,

а у десяти курильщиков, не склонных к аллергии:

48,1; 45,5; 41,7; 35,4; 29,1; 18,9; 58,3; 66,8; 71,3; 94,3.

Верно ли предположение о том, что уровень гистамина у курильщиков, подверженных аллергии, выше, чем у неаллергиков?

Вариант 3

1. В таблице приведены 2663 случая, классифицированных по двум признакам: А — курит/не курит, В — профессия.

	A1	A2
Врач	68	6
Токарь	46	13

Учитель 15 38

Космонавт 16 67

Вычислить меры прогноза Гутмана.

2. Имеется 12 одинаковых по размеру дисков, окраска которых отличается тоном---- от светло-голубого до темно-синего. Для того чтобы оценить способность модельера одежды различать цветовые оттенки, ему предлагают расположить диски в порядке увеличения степени интенсивности цвета.

Модельер установил следующий порядок дисков: 1, 4, 7, 2, 3, 5, 8, 12, 10, 6, 11, 9.

Объективная оценка интенсивности цвета, полученная с помощью колориметрического испытания: 1, 2,...,12. Опираясь на эти данные, охарактеризуйте способность модельера различать оттенки цвета.

3. Судейская коллегия из шести человек оценивает выступления четырёх фигуристов, вышедших в финал соревнований.

A B C D

1 2 3 4

1 3 4 2

4 3 2 1

1 2 3 4

2 1 3 4

1 4 3 2

Вычислите коэффициент конкордации Кендалла.

4. По имеющимся данным об экономических показателях X_1 , X_2 и X_3 для 20 регионов России вычислены выборочные коэффициенты корреляции , , . Предполагается, что наблюдения имеют гауссовское распределение. Вычислите частные коэффициенты корреляции

Вариант 4

1. В таблице приведены 2663 случая, классифицированных по двум признакам: А — наличие прививки против холеры, В — отсутствие заболевания.

Проверьте гипотезу о независимости признаков А и В на уровне значимости 0,05.

1625 5

1022 11

2. Деятельность отделения банка характеризуется некоторым показателем X . Для проверки была случайным образом выбрана группа из 10 однотипных отделений банка. Показатель X у этих отделений составил:

258, 588, 477, 577, 619, 614, 641, 543, 517, 593.

После экономического кризиса показатель X у 9 случайным образом выбранных отделений составил:

537, 398, 256, 440, 376, 524, 527, 589, 479.

Можно ли считать, опираясь на эти данные, что экономический кризис привел к снижению показателя X ?

3. В конкурсе красоты участвовало 13 девушек. Разыгрывались призы жюри и зрителей. Места присуждённые девушкам жюри и зрителями:

Жюри : 12 13 10 7 8 2 5 1 6 3 9 4 11

Зрители: 13 11 8 2 9 6 4 5 3 7 10 1 12

Зависимы ли мнения жюри и зрителей?

4. Имеются данные о величине прожиточного минимума трудоспособного населения в федеральных округах (на душу населения руб. в месяц), установленной в субъектах РФ за IV квартал 2008 г.

Центральный Сибирский Дальневосточный

Белгородская область Республика Бурятия Чукотский автономный округ

4143 5039 10061

Брянская область Республика Тыва Камчатский край

4292 5005 10146

Ивановская область Кемеровская область Приморский край

4623 4407 6525

Курская область Красноярский край Амурская область

4424 5625 6069

Московская область Омская область Сахалинская область

5786 4836 7972

Рязанская область

4772

г. Москва

7510

Можно ли считать, что величина прожиточного минимума во всех представленных федеральных округах в среднем одинакова?

При выполнении контрольной работы студент получает оценку в соответствии со следующими критериями оценивания:

10 баллов (отлично) Все задания выполнены верно.

8-9 баллов (отлично) В ходе выполнения работы допущены арифметические и другие ошибки не влияющие на ход решения задачи.

7 баллов (хорошо) В ходе выполнения работы допущены ошибки, влияющие на ход решения задач, или одна из задач не решена целиком.

5-6 баллов

(хорошо) Одна задача не решена целиком, а при решении других задач допущены ошибки, влияющие на ход решения задачи.

3-4 балла (удовл) Не решено две задачи.

1-2 балла (неудовл) Выполнено не более одного задания из четырех.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Перечень контрольных вопросов для сдачи дифференцированного зачета в 7-ом семестре:

1. Основные понятия математической статистики. Предмет и задачи математической статистики.
2. Выборка и её характеристики. Генеральная совокупность. Распределение генеральной совокупности.
3. Выборочный вектор. Выборочное пространство.
4. Вариационный ряд. Размах выборки. Частота элемента выборки.
5. Группированный статистический ряд. Эмпирическая (выборочная) функция распределения.
6. Теорема Гливенко. Теорема Бернулли. Гистограмма.
7. Выборочное среднее, дисперсия, медиана.
8. Выборочные начальные и центральные моменты. Свойства выборочных характеристик.
9. Формулы для вычислений выборочных характеристик группированной выборки. Выборочная квантиль.
10. Точечные оценки и их свойства. Состоятельность. Несмещенная оценка, смещение. Метод подстановки.
11. Задачи на несмещенность и состоятельность оценок, полученных методом подстановки.
12. Метод максимального правдоподобия. Функция правдоподобия. Свойства оценок, полученных методом максимального правдоподобия. Принцип инвариантности
13. Эффективные оценки. Регулярные распределения. Информация Фишера. Вклад выборки.
14. Неравенство Рао-Крамера. Сверэффективные оценки.
15. Интервальные оценки. Определение интервальной оценки.
16. Центральная статистика и способ её построения.
17. Распределения Фишера, хи-квадрат и Стьюдента.
18. Асимптотический доверительный интервал.

19. Доверительные интервалы для параметров гауссовского распределения. Построение центральных доверительных интервалов для математического ожидания и дисперсии гауссовского распределения.

Перечень контрольных вопросов для сдачи экзамена в 8-ом семестре:

1. Проверка параметрических гипотез. Простая и сложная гипотеза. Статистический критерий. Уровень значимости. Ошибки первого и второго рода. Мощность критерия.
2. Теорема Неймана-Пирсона.
3. Проверка непараметрических гипотез. Статистика хи-квадрат Пирсона. Ассимптотическое распределение статистики Пирсона. Критерий согласия Пирсона.
4. Формула Стерджеса. Проверка гипотезы о независимости дискретных случайных величин.
5. Проверка гипотезы об однородности в двухвыборочной модели. Постановка задачи. Критерии Стьюдента и Фишера.
6. Критерий Колмогорова-Смирнова
7. Критерии свободные от распределения. Понятие рангов элементов выборки.
8. Критерии Вилкоксона и Ансари-Брэдли.
9. Коэффициенты корреляции. Выборочный коэффициент корреляции. Ранговые коэффициенты Кендалла и Спирмена. Частный и множественный коэффициент корреляции.
10. Однофакторный дисперсионный анализ. Постановка задачи. Основное тождество дисперсионного анализа.
11. Критерий Фишера и Краскела-Уоллиса. Линейные контрасты.
12. Линейная модель регрессии. Метод наименьших квадратов.
13. Свойства МНК-оценки. Теорема Гаусса-Маркова. Коэффициент детерминации. Критерий Фишера.
14. Различные типы моделей регрессии. Гетероскедастичность и автокорреляция.
15. Мультиколлиниарные модели. Метод наименьших модулей.
16. Нелинейная модель регрессии. Квантильная регрессия.
17. Анализ временных рядов. Примеры временных рядов. Стационарные и нестационарные временные ряды. Система уравнения Юла-Уолкера.

Критерии оценивания

Оценка отлично (10) выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9) выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8) выставляется студенту, показавшему систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7) выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6) выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5) выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4) выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3) выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2) выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1) выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Экзамен и дифференцированный зачет проводится в устной форме.

При проведении дифференцированного зачета обучающемуся предоставляется 20 минут на подготовку. Опрос обучающегося проводится в течение 30 минут.

При проведении устного экзамена обучающемуся предоставляется не менее 60 минут на подготовку. Опрос обучающегося на устном экзамене не должен превышать двух астрономических часов.

Во время проведения экзамена и дифференцированного зачета обучающиеся могут пользоваться программой дисциплины.