

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Механико-математический факультет

УТВЕРЖДЕНО:

Декан

Л. В. Гензе

Оценочные материалы по дисциплине

Математическая статистика

по направлениям подготовки

01.03.01 Математика

02.03.01 Математика и компьютерные науки

01.03.03 Механика и математическое моделирование

Направленность (профиль) подготовки/ специализация:

Современная математика и математическое моделирование

Вычислительная математика и компьютерное моделирование

Теоретическая, вычислительная и экспериментальная механика

Форма обучения

Очная

Квалификация

Математик. Преподаватель / Математик. Аналитик / Математик. Исследователь

Математик. Преподаватель / Математик. Вычислитель /

Исследователь в области математики и компьютерных наук

Механик. / Механик. Исследователь

Год приема

2024, 2025

СОГЛАСОВАНО:

Руководитель ОП

Л.В. Гензе

Председатель УМК

Е.А. Тарасов

Томск – 2024

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен применять фундаментальные знания, полученные в области математических наук и механики в профессиональной деятельности.

ОПК-2 Способен разрабатывать, анализировать и внедрять математические модели в современной науке и технике, экономике и управлении.

ПК-3 Способен проводить эффективный статистический анализ данных и строить стохастические модели на основе оптимальных вероятностных методов.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

РООПК-1.1 Знает типовые постановки задач математики и механики, классические методы решения, теоретические основы методов и границы их применимости

РООПК-1.2 Способен адаптировать известные математические методы для решения поставленной задачи в области математики и механики

РООПК-1.3 Способен провести решение поставленной задачи в области математики и механики с использованием полученных фундаментальных знаний и получить результат

РООПК-2.1 Умеет обоснованно выбрать тип математической модели для формализации решаемой задачи

РООПК-2.2 Применяет стандартные и типовые действия при построении математической модели определенного типа

РООПК-2.3 Применяет подходы визуализации и представления результатов математического моделирования для апробации и демонстрации в виде отчетов, презентаций и научных текстов

РОПК-3.1 Знает:

- актуальные методы и передовые результаты в области стохастического анализа и его применений

- основные современные математические методы обработки статистических данных

- основные принципы стратегического управления развитием методологической и технологической инфраструктуры анализа больших данных

- методы получения, хранения, передачи и обработки информации для систем больших данных

РОПК-3.2 Будет владеть фундаментальными знаниями и базовыми навыками в области ТВ и МС для применения в части:

- разработки и совершенствования статистических и вероятностных методов анализа числовой и нечисловой информации

- осуществлять консультирование в области статистической и актуарной деятельности

- совершенствовать и разрабатывать новые методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными

РОПК-3.3 Будет владеть фундаментальными знаниями и базовыми навыками в области ТВ и МС для освоения:

- навыков подготовки аналитических отчетов, а также обзоров, докладов, рекомендаций по результатам проведенных статистических расчетов

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- контрольные работы;
- кейсы;
- коллоквиум.

Контрольная работа 1 (РОПК-3.1, 3.2, 3.3)

Контрольная работа состоит из 3 задач.

Примеры задач:

1. Пусть $X^{(n)} = (X_1, \dots, X_n)$ – выборка из генеральной совокупности независимых и одинаково $U[0, \theta]$ распределенных случайных величин. Является ли оценка $T = \frac{n+1}{n} X_{(n)}$,

$X_{(n)} = \max_{1 \leq i \leq n} X_i$, параметра θ несмещенной, состоятельной?

2. Пусть $X^{(n)} = (X_1, \dots, X_n)$ – выборка, элементы которой независимы и имеют дискретное распределение, задаваемое таблицей

X	1	2	3
$P(X=k)$	θ	$1-2\theta$	θ

Обладает ли такая статистическая структура достаточной статистикой для параметра θ ? Найти оценки по методу моментов и методу максимального правдоподобия. Сравнить оценки.

3. Пусть $X_1, \dots, X_n$ – независимы и имеют нормальное распределение $N(\theta, 1)$. Построить γ -доверительный интервал для θ , основанный на центральной статистике $\sqrt{n} \cdot (\bar{X} - \theta)$.

4. По выборке $X = (X_1, \dots, X_n)$ из распределения $N(\theta, \sigma^2)$ построить двусторонний γ -доверительный интервал для $\tau = -\theta$.

5. Пусть (X_1, X_2, X_3) – выборка из нормального распределения $N(0, \sigma^2)$. Определим статистику

$$p_T(x) = \frac{1}{2T} \mathbf{1}_{\{|x| \leq T\}},$$

где $T = \sqrt{X_1^2 + X_2^2 + X_3^2}$, которая как функция переменного x представляет собой плотность равномерного распределения $U[-T, T]$. Проверить, что $p_T(x)$ при всех x является несмещенной оценкой теоретической плотности нормального распределения $N(0, \sigma^2)$.

6. Пусть $(X_1, \dots, X_n)$ – выборка из гамма-распределения $\Gamma(\theta, \lambda)$. Построить оценку максимального правдоподобия для $\tau(\theta) = \theta^{-1}$. Убедиться в ее асимптотической несмещенности, состоятельности и асимптотической нормальности.

7. Пусть $(X_1, \dots, X_n)$ – выборка из нормального распределения $N(\theta, \sigma^2)$ с неизвестным средним θ и известной дисперсией σ^2 . Показать, что дисперсия любой несмещенной оценки $\hat{\theta}$ параметра θ не меньше, чем σ^2/n .

Контрольная работа 2 (РОПК-3.1, 3.2, 3.3)

Контрольная работа состоит из 3 задач.

Примеры задач:

1. Пусть $X_1, \dots, X_n$ - независимы и имеют распределение Пуассона $P(\theta)$. Построить равномерно наиболее мощный критерий размера α для проверки гипотезы $H_0: \theta = \theta_0$ при альтернативе $H_1: \theta > \theta_0$. Найти функцию мощности.

2. В течение эпидемии гриппа среди 2000 контролируемых индивидуумов одно заболевание наблюдалось у 181 человека, а дважды болели гриппом лишь 9 человек. Согласуются ли при уровне значимости 0,05 эти данные с гипотезой, согласно которой число заболеваний отдельного индивидуума – биномиальная $(2, p)$ случайная величина?

3. Из 300 абитуриентов, поступивших в вуз, 97 человек имели 5 в школе и 48 получили 5 на вступительных экзаменах по тому же предмету, причем только 18 человек имели 5 и в школе и на экзамене. Проверить гипотезу о независимости оценок 5 в школе и на экзамене при уровне значимости 0,1.

4. Изучая физические свойства нефтяного пласта, исследователь получил следующие значения пористости пласта в зонах расположения двух достаточно удаленных скважин на залежи:

Первая скважина	5,9	8,2	6,0	7,7	6,5	5,2			$n_1 = 6$
Вторая скважина	7,5	7,8	8,5	10,3	8,1	7,6	8,8	9,9	$n_2 = 8$

Можно ли считать, что коллектор однороден, т.е. его пористость в среднем меняется незначительно?

Критерии оценивания: результаты контрольной работы определяются оценками «зачтено», «не зачтено».

Оценка «зачтено» выставляется, если 2 задачи решены без ошибок.

Иначе выставляется оценка «не зачтено».

Пример кейса (РОПК 3.1 - 3.3)

1. Известны факторы, влияющие на цену выпускаемого продукта ООО «Двери», результаты наблюдений за которыми приведены в таблице. Провести первичную обработку данных. Определите зависимость цены от затрат удельных и постоянных, а также от рентабельности предприятия, то есть от объема выпускаемой продукции, используя модель множественной линейной регрессии. Исследуйте качество модели.

Цена	20	30	21	25	23	18	22	24	29	27
Удельные затраты	10	12	7	11	14	5	8	6	9	13
Постоянные затраты	100	120	90	94	91	80	93	95	103	101
Объем выпускаемой продукции	12	8	7,5	7,9	13	7	7,7	6	5	8,3

Написать отчет и защитить его.

2. Самостоятельно найти данные (скачать из официальных источников или воспользоваться датчиками случайных чисел). Провести первичную обработку данных. Определить зависимую переменную и факторы, которые влияют на зависимую переменную. Определить зависимость, используя модель множественной линейной/нелинейной регрессии, то есть

1) произвести идентификацию модели (определить тип зависимости и оценить неизвестные параметры модели);

2) вычислить стандартные ошибки (дисперсии) параметров регрессии;

3) проверить значимость каждого параметра и произвести интервальное оценивание параметров регрессионной модели;

- 4) рассчитать общую, факторную и остаточную дисперсии;
- 5) вычислить коэффициент детерминации;
- 6) вычислить стандартную ошибку регрессии;
- 7) проверить общее качество модели при уровне значимости, равном 0,05;
- 8) проверить выполнение предположений регрессионного анализа;
- 9) по результатам анализа качества модели (в случае некачества) произвести ее усовершенствование;
- 10) в случае парной/двухфакторной регрессионной модели, изобразить ее графически (построить линию/поверхность);
- 11) Написать отчет и защитить его.

Критерии оценивания: результат защиты отчета по выполнению кейса определяется оценками «зачтено», «не зачтено».

Оценка «зачтено» выставляется, если кейс выполнен верно не менее, чем на 70%.
Иначе выставляется оценка «не зачтено».

Коллоквиум (РООПК 1.1 - 2.3)

Билет на коллоквиум состоит из 1 теоретического вопроса.

Перечень теоретических вопросов:

1. Проблема статистического вывода. Задачи математической статистики. Схема статистического вывода.
2. Выборка. Статистическая структура. Статистика. Статистическое правило. Функции потерь и риска. Плотность распределения выборки.
3. Выборочные характеристики: порядковые статистики, эмпирическая функция распределения, выборочные моменты и квантили, гистограммная оценка плотности распределения.
4. Достаточные статистики. Теорема факторизации с доказательством.
5. Достаточные статистики. Теорема Рао-Блэкуэлла-Колмогорова с доказательством.
6. Примеры построения достаточных статистик для статистических структур Бернулли, равномерной, нормальной и Коши.
7. Примеры построения достаточных статистик для статистических структур Пуассона, показательной, гамма и биномиальной.
8. Оценка неизвестного параметра распределения: определение, свойства. Формулировка основной задачи статистической теории оценивания. Метод моментов.
9. Примеры построения оценок по методу моментов.
10. Теорема о состоятельности оценок по методу моментов с доказательством.
11. Теорема об асимптотической нормальности оценок по методу моментов с доказательством.
12. Принцип и метод максимального правдоподобия для оценивания неизвестных параметров распределений. Функция правдоподобия. Уравнения правдоподобия. Примеры построения оценок ММП.
13. Теорема о состоятельности оценок ММП с доказательством.
14. Информация по Фишеру: определение и свойства.
15. Теорема об асимптотической нормальности оценок ММП с доказательством.
16. Оптимальная оценка. Неравенство Рао-Крамера с доказательством.
17. Неравенство Рао-Крамера для несмещенных оценок. Метод построения оптимальных несмещенных оценок.
18. Эффективность несмещенной оценки. Теорема об эффективности оценки ММП с доказательством. Примеры эффективных несмещенных оценок неизвестных

параметров распределений.

19. Доверительные области (интервалы). Построение доверительных интервалов для среднего и доверительных верхних границ для дисперсии нормального распределения.
20. Лемма Фишера с доказательством.
21. Асимптотически доверительные области. Построение асимптотически доверительных интервалов для вероятности успеха в схеме Бернулли и параметра интенсивности распределения Пуассона.

Критерии оценивания коллоквиума и зачета

Не зачтено	Зачтено
Дан неправильный ответ, однозначно неправильная трактовка темы.	Дан правильный и развернутый ответ на вопрос. Знание основной и дополнительной литературы.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Зачет в шестом семестре проводится в письменной форме с последующим собеседованием по билетам. Билет содержит теоретический вопрос и задачу.

Перечень теоретических вопросов (РОПК 1.1, 1.2, 1.3):

1. Статистическая проверка гипотез: основные понятия и определения.
2. Критерии на основе состоятельных оценок параметров распределений. Проверка гипотезы о величине среднего значения нормального распределения при известной и неизвестной дисперсии.
3. Критерии на основе состоятельных оценок параметров распределений. Проверка гипотезы о величине дисперсии нормального распределения при известном и неизвестном среднем значении.
4. Двухвыборочные критерии Стьюдента (сравнение средних) и Фишера (сравнение дисперсий).
5. Подходы к сравнению критериев. Наиболее мощные критерии.
6. Критерий отношения правдоподобия. Лемма Неймана-Пирсона с доказательством.
7. Рандомизированные критерии. Проверка гипотезы о вероятности успеха в схеме Бернулли.
8. Проверка надежности при показательном распределении долговечности.
9. Принцип двойственности между задачами проверки гипотез и доверительного оценивания. Оптимальность доверительных границ, соответствующих равномерно наиболее мощным критериям.
10. Критерии согласия хи-квадрат и Колмогорова.
11. Теорема Пирсона с доказательством.
12. Критерий проверки независимости хи-квадрат.
13. Критерий проверки однородности хи-квадрат.
14. Критерий Смирнова. Критерий Манна-Уитни.
15. Исследование статистической зависимости. Математическая модель регрессии. Метод максимального правдоподобия и метод наименьших квадратов. Коэффициент корреляции. Модель нормальной регрессии.
16. Идентификация множественной линейной регрессии методом наименьших квадратов. Теорема Гаусса-Маркова с доказательством.
17. Статистический анализ качества множественной линейной регрессии (построение доверительных интервалов, проверка гипотез для параметров модели, коэффициент детерминации и F-критерий).

Примеры задач:

1. Ниже приведены результаты измерения роста (в см) случайно отобранных студентов.

Рост	154-158	158-162	162-166	166-170	170-174	174-178	179-182
Число студентов	10	14	26	28	12	8	2

Считая, что измерения нормально распределены с параметрами (a, σ^2) . Найти наиболее правдоподобную оценку вектора (a, σ^2) , оценить эффективность полученной оценки. Каково будет решение этой задачи, когда

- а) σ^2 - известно, a - неизвестно;
- б) σ^2 - неизвестно, a - известно.

2. Согласуются ли приведенные выше данные с нормальным распределением?

3. Пусть $(X_1, \dots, X_n)$ - выборка из равномерного распределения $U[0, \theta]$. Построить кратчайший γ -доверительный интервал для параметра θ , основанный на центральной

статистике $T(X^{(n)}) = \frac{\max_{1 \leq i \leq n} X_i}{\theta}$.

4. Пусть $(X_1, \dots, X_n)$ - выборка из биномиального $(1, p)$ распределения. Построить равномерно наиболее мощный критерий размера α для проверки гипотезы $H_0: p = p_0$ при альтернативе $H_1: p < p_0$. Найти мощность критерия.

5. Пусть $(X_1, \dots, X_n)$ - выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \exp\{-(x - \theta)\}, & x > \theta, \\ 0, & x \leq \theta. \end{cases}$$

Построить равномерно наиболее мощный критерий размера α для проверки гипотезы $H_0: \theta = \theta_0$ при альтернативе $H_1: \theta = \theta_1$, $\theta_1 < \theta_0$. Найти мощность критерия.

Зачет может быть оценен «Зачтено» и «Не зачтено». Итоговая оценка – «Зачтено» выставляется, если зачтены контрольные работы, коллоквиум и зачет, иначе выставляется - «Не зачтено». При ответе на теоретический вопрос оценивается полнота и точность ответа, логичность и аргументированность изложения материала, умения использовать в ответе фактический материал, знания основной и дополнительной литературы.

В случае если какая-либо часть текущего контроля не сдана студентом, она переносится на зачет в виде дополнительного теоретического вопроса и/или задачи по соответствующему разделу.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Задачи (РООПК 1.1-1.3)

1. Пусть $X_1, \dots, X_n$ независимы и имеют нормальное распределение $N(\theta, 1)$.

Исследовать несмещенность и состоятельность оценки $T(X) = \bar{X}$ параметра θ .

2. Дана среднегодовая температура в городе за два десятилетия.

1-ое десятилетие	7,5	10,6	9,7	6,8	6,9	10,5	9,5	8,9	11,2	8,1
------------------	-----	------	-----	-----	-----	------	-----	-----	------	-----

2-ое десятилетие	10,1	8,4	5,6	11,7	9,8	7,8	8,8	11,1	9,7	7,8
------------------	------	-----	-----	------	-----	-----	-----	------	-----	-----

Определить, является ли расхождение между среднегодовыми температурами существенным, т.е. меняется ли климат данного города?

Ответы:

Задача 1. Несмещенная и состоятельная.

Задача 2. Нет, не значимо расхождение.

Теоретические вопросы (РОПК 1.1, 1.2, 1.3):

1. Дать определение статистики. Что такое достаточная статистика? Сформулировать основную задачу математической статистики.
2. Что такое оценка по методу максимального правдоподобия? Перечислить ее свойства.
3. Сформулировать задачу построения доверительной области.
4. Сформулировать задачу проверки гипотезы о параметре распределения.
5. Что такое равномерно наиболее мощный критерий. Сформулировать теорему Неймана-Пирсона.
6. Что такое регрессия?

Информация о разработчиках

Пчелинцев Евгений Анатольевич, доктор физико-математических наук, доцент, заведующий кафедрой математического анализа и теории функций ММФ ТГУ.

Губин Владимир Николаевич, кандидат физико-математических наук, доцент кафедры математического анализа и теории функций ММФ ТГУ.

Емельянова Татьяна Вениаминовна, кандидат физико-математических наук, доцент кафедры математического анализа и теории функций ММФ ТГУ.

Мурзинцева Алена Андреевна, ст. преподаватель кафедры математического анализа и теории функций ММФ ТГУ.

Перелевский Святослав Сергеевич, ст. преподаватель кафедры математического анализа и теории функций ММФ ТГУ.