

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ
Директор института прикладной
математики и компьютерных наук
А.В. Замятин
« 18 » мая 2022 г.



Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине
(Оценочные средства по дисциплине)

Статистический анализ данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Обработка данных, управление и исследование сложных систем

ОС составил(и):

канд. физ.-мат. наук, доцент

доцент кафедры теории вероятностей и математической статистики

Т.В. Кабанова

Рецензент:

канд. физ.-мат. наук, доцент

доцент кафедры теоретической кибернетики

Ж.Н. Зенкова

Оценочные средства одобрены на заседании учебно-методической комиссии института прикладной математики и компьютерных наук (УМК ИПМКН)

Протокол от 12.05. 2022г. № 4

Председатель УМК ИПМКН,
д-р техн. наук, профессор

С.П. Сущенко

Оценочные средства (ОС) являются элементом системы оценивания сформированности компетенций у обучающихся в целом или на определенном этапе ее формирования.

ОС разрабатывается в соответствии с рабочей программой (РП) дисциплины.

1. Компетенции и результаты обучения, формируемые в результате освоения дисциплины

Компетенция	Индикатор компетенции	Код и наименование результатов обучения (планируемые результаты обучения, характеризующие этапы формирования компетенций)	Критерии оценивания результатов обучения			
			Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
ОПК-1 Способен решать актуальные задачи фундаментальной и прикладной математики	ИОПК-1.3 Решает актуальные задачи фундаментальной и прикладной математики	ОР-1.1.1. обучающийся сможет: - выбирать адекватные метод для решения поставленной задачи; - реализовывать выбранный метод в программе анализа данных; - делать выводы и интерпретировать полученные результаты.	Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.	В целом успешные, но содержащие отдельные пробелы знания о математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.	Фрагментарное, неполное знание без грубых ошибок математически базовых понятий, которые необходимы для понимания статистических методов анализа данных.	Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.

<p>ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач.</p>	<p>ИОПК-2.1 Использует результаты прикладной математики для освоения, адаптации новых методов решения задач в области своих профессиональных интересов</p>	<p>ОР-2.1.1. обучающийся сможет: адаптировать модели для описания процессов реальной предметной области</p>	<p>Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>В целом успешные, но содержащие отдельные пробелы знания о математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Фрагментарное, неполное знание без грубых ошибок математически х основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.</p>
	<p>ИОПК-2.2 Реализует и совершенствует новые методы, решения прикладных задач в области профессиональной деятельности</p>	<p>ОР-2.2.1. обучающийся сможет: реализует и интерпретирует построенные модели для описания процессов реальной предметной области</p>	<p>Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>В целом успешные, но содержащие отдельные пробелы знания о математически х основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Фрагментарное, неполное знание без грубых ошибок математически х основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.</p>

	<p>ИОПК-2.3 Проводит качественный и количественный анализ полученного решения с целью построения оптимального варианта</p>	<p>ОР-2.3.1. Обучающийся сможет проводить качественный и количественный анализ построенных моделей и полученных на их основании прогнозов и выбирать наиболее оптимальный в соответствии с выбранной метрикой</p>	<p>Демонстрация высокого уровня знаний математических основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>В целом успешные, но содержащие отдельные пробелы знания о математически х основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Фрагментарное, неполное знание без грубых ошибок математически х основ и базовых понятий, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Не знает математические основы и базовые понятия, которые необходимы для понимания статистических методов анализа данных.</p>
--	--	---	--	---	--	--

2. Этапы формирования компетенций и виды оценочных средств

№	Этапы формирования компетенций (разделы дисциплины)	Код и наименование результатов обучения	Вид оценочного средства (тесты, задания, кейсы, вопросы и др.)
1.	Введение в статистический анализ.	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
2.	Критерии сравнения групп.	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
3.	Корреляционный анализ	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
4.	Парная регрессия	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
5.	Множественная регрессия	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
6.	Задачи классификации	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
7.	Кластерный анализ	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
8.	Анализ временных рядов	ОР-1.1.1, ОР-2.1.1, ОР-2.2.1, ОР-2.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста

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

3.1. Типовые задания для проведения текущего контроля успеваемости по дисциплине.

Текущий контроль осуществляется на основании посещаемости и проверки лабораторных работ.

Примеры заданий для лабораторных работ

Лабораторная работа. Генерация и анализ выборок из непрерывных и дискретных распределений

1. Сформировать выборку из генеральной совокупности с дискретным законом распределения (Пуассона, геометрическое или биномиальное). Объем от 100 до 200 наблюдений. Параметры задать самостоятельно.
2. Построить вариационный ряд абсолютных и относительных частот, полигон частот (см. рис.1), эмпирическую функцию распределения (см. рис.2), найти оценки числовых характеристик (выборочные среднее, дисперсию, СКО, моду, медиану, коэффициенты асимметрии и эксцесса.)
3. Найти теоретические мат ожидание и дисперсию при заданных параметрах. Сравнить найденные точечные оценки с теоретическими характеристиками.

4. Найти оценки параметров соответствующего распределения. Сравнить полученные оценки с заданными теоретическими значениями.
5. Проверить гипотезу о виде распределения.

Лабораторная работа. Предварительная обработка данных

Задание.

1. Импортировать заданный набор данных.
2. Проверить на наличие пропусков и выбросов.
3. Для количественных показателей построить гистограммы.
4. Найти оценки числовых характеристик.
5. Проверить гипотезу о нормальности.
6. Построить диаграммы размаха по группам на основании разбиения количественных показателей по уровням категориальных признаков.

Лабораторная работа. Множественная регрессия. Фиктивные переменные

Выполняется в R.

Задание.

Импортировать таблицу с данными в R.

1. Построить графики для визуализации данных и их взаимосвязей.
2. Проверить связи факторов друг с другом и их влияние на зависимую целевую переменную.
3. Построить и провести анализ множественной модели регрессии целевой переменной от всех представленных количественных и порядковых факторов.
4. Провести обработку и кодирование категориальных факторов.
5. Построить и провести анализ множественной модели регрессии с учетом всех предложенных факторов.
6. Удалить незначимые факторы. Построить окончательную модель.
7. Проверить остатки модели на нормальность.
8. Задать новое наблюдение со своими значениями признаков и построить прогноз целевого показателя для него.

Лабораторная работа. Линейные и нелинейные модели парной регрессии. Построение и анализ

Выполняется в R.

Пусть регрессионная модель описывается одним из уравнений:

- | | |
|---------------------|--|
| 1. Линейная | $y = a + bx + \varepsilon$ |
| 2. Степенная | $y = a \cdot x^b \cdot \varepsilon$ |
| 3. Экспоненциальная | $y = a \cdot e^{bx} \cdot \varepsilon$ |
| 4. Логарифмическая | $y = a + b \cdot \ln(x) + \varepsilon$ |
| 5. Гиперболическая | $y = a + \frac{b}{x} + \varepsilon$ |

Задание.

1. Сгенерировать выборки по n наблюдений по каждой из выше предложенных моделей по примеру линейной модели из учебно-методического пособия. Все необходимые параметры задать самостоятельно.
2. Построить диаграммы рассеяния для исходной модели.
3. Для нелинейных моделей провести линеаризацию и построить диаграммы рассеяния линеаризованных моделей.
4. Найти МНК-оценки параметров модели.
5. Найти дисперсии наблюдений и оценок параметров.
6. Построить доверительные интервалы для неизвестных параметров.
7. Проверить гипотезы о значимости коэффициентов регрессии.
8. Найти коэффициент детерминации модели.
Проверить гипотезу об адекватности модели.

3.2. Типовые задания для проведения промежуточной аттестации по дисциплине.

Примерный перечень теоретических вопросов и тем для подготовки к экзамену:

1. Типы данных и способы их представления.
2. Параметрические критерии сравнения групп.
3. Непараметрические критерии сравнения групп.
4. Корреляционный анализ количественных данных.
5. Ранговая корреляция.
6. Корреляционный анализ категоризованных данных.
7. Числовые характеристики оценок параметров парной регрессии.
8. Теорема Гаусса-Маркова для случая парной регрессии.
9. Проверка качества уравнения парной регрессии.
10. Скалярная и матричная записи уравнения множественной регрессии. МНК-оценки параметров. Условия Гаусса-Маркова.
11. Теорема Гаусса-Маркова для множественной регрессии.
12. Оценка дисперсии шума в матричном виде.
13. Проверка гипотез о значениях и значимости параметров множественной регрессии.
14. Доверительные интервалы для параметров и функции множественной регрессии.
15. Случай коррелированных гомоскедастических наблюдений.
16. Случай некоррелированных гетероскедастических наблюдений.

17. Мультиколлинеарность.
18. Фиктивные переменные.
19. Постановка задачи классификации. Логистическая регрессия.
20. Метрики качества бинарного классификатора.
21. ROC-анализ.
22. Типы и примеры алгоритмов кластеризации
23. Расстояния между объектами и между классами.
24. Структура временного ряда.
25. Методы сглаживания временного ряда.

4. Методические материалы, определяющие процедуры оценивания образовательных результатов обучения

4.1. Методические материалы для оценки текущего контроля успеваемости по дисциплине.

Для текущей аттестации необходимо иметь посещаемость не менее 75% от всех проведенных на момент аттестации занятий и сдать все данный ан момент аттестации лабораторные работы. Лабораторные работы оцениваются на зачтено/незачтено.

4.2. Методические материалы для проведения промежуточной аттестации по дисциплине.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Для письменного экзамена:

отлично	Ответ на вопрос билета дан в полном объеме, достаточно точно, возможны незначительные, несущественные неточности
хорошо	Ответ дан в неполном объеме, но на достаточно хорошем уровне, имеется пара не очень грубых ошибок.
удовлетворительно	Раскрыта основная суть ответа на вопрос, приведены основные результаты, но ответ недостаточно аргументирован, имеются не очень грубые ошибки.
неудовлетворительно	Основная суть ответа не раскрыта, ответ дан в недостаточном объеме, имеются грубые ошибки.

Для теста из 15 вопросов. За каждый вопрос в зависимости от его сложности можно получить от 1 до 3 баллов. Максимально 30.

отлично	От 26 до 30 баллов
хорошо	От 21 до 25 баллов
удовлетворительно	От 16 до 20 баллов
неудовлетворительно	От 0 до 15 баллов