

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

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:

Директор



А. В. Замятин

« 15 » июня 2023 г.

Рабочая программа дисциплины

Статистический анализ данных

по направлению подготовки

02.04.02 Фундаментальная информатика и информационные технологии

Направленность (профиль) подготовки:

Математика беспроводных сетей связи и интернета вещей

Форма обучения

Очная

Квалификация

Магистр

Год приема

2023

Код дисциплины в учебном плане: Б1.О.02.04

СОГЛАСОВАНО:

Руководитель ОП

С.П. Моисеева

Председатель УМК

С.П. Сущенко

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

– ОПК-1 – Способен решать актуальные задачи фундаментальной и прикладной математики.

– ОПК-3 – Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.3 Решает актуальные задачи фундаментальной и прикладной математики.

ИОПК-3.1 Проводит анализ математических моделей и систем.

ИОПК-3.2 Применяет математические модели, методы для решения прикладных задач профессиональной деятельности.

ИОПК-3.3 Разрабатывает новые алгоритмы и методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

2. Задачи освоения дисциплины

- Научить студентов решать задачи статистического анализа данных, начиная от их формулирования исходных задач соответствующей предметной области на языке прикладной статистики, выбора методов решения и критериев качества полученных решений и заканчивая формулировкой полученных выводов на языке предметной области.
- Изучить основные методы статистического анализа данных.
- Сформировать навыки работы в программах статистической обработки данных.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательной части образовательной программы. Дисциплина входит в модуль «Общепрофессиональные дисциплины».

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Первый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 16 ч.

-лабораторные: 16 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в статистический анализ.

Типы данных. Графические и табличные способы представления данных. Предварительная обработка данных.

Тема 2. Критерии сравнения групп.

Параметрические критерии. t-критерий Стьюдента. Критерий Фишера. Дисперсионный анализ. Непараметрические критерии. Критерии Манна-Уитни, Вилкоксона, Краскала-Уолиса, Фридмана.

Тема 3. Корреляционный анализ.

Парный коэффициент корреляции Пирсона. Z-преобразование Фишера. Корреляционный анализ, Ранговая корреляция. Коэффициент Спирмена, Кендалла, конкордации Кендалла. Корреляционный анализ категоризованных данных.

Тема 4. Парная регрессия.

Определение простой регрессии. Метод наименьших квадратов оценки параметров простой регрессии. Условия Гаусса-Маркова. Теорема Гаусса-Маркова. Оценки дисперсий. Проверка качества модели регрессии, Коэффициент детерминации, его интерпретация, общая адекватность модели. Нелинейные модели и линеаризация.

Тема 5. Множественная регрессия.

Основные понятия и задачи регрессионного анализа, Общая постановка задачи множественной регрессии. Метод наименьших квадратов оценки параметров регрессии. Теорема Гаусса-Маркова. Оценка дисперсий. Проверка качества модели множественной регрессии. Фиктивные переменные. Случай коррелированных наблюдений Гетероскедастичность. Мультиколлинеарность.

Тема 6. Задача классификации.

Основные понятия и задачи классификации. Бинарная классификация и логистическая регрессия. Метрики качества. ROC-анализ.

Тема 7. Кластерный анализ.

Основные подходы в задачах кластеризации. Итерационные, плотностные, иерархические алгоритмы. Расстояния между объектами Расстояния между классами. Проверка качества кластеризации.

Тема 8. Анализ временных рядов.

Понятие временного ряда, основные модели временных рядов, задачи анализа временных рядов. Декомпозиция временных рядов. Прогнозирование во временных рядах.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, выполнения лабораторных работ, и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в первом семестре проводится в письменной форме по билетам или итогового тестирования. Экзаменационный билет состоит из одного вопроса, требующего развернутого ответа. Тест состоит из 15-20 вопросов. Продолжительность экзамена 1 академический час (45 минут).

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.

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.

1. Импортировать таблицу с данными в R.
2. Построить графики для визуализации данных и их взаимосвязей.
3. Проверить связи факторов друг с другом и их влияние на зависимую целевую переменную.
4. Построить и провести анализ множественной модели регрессии целевой переменной от всех представленных количественных и порядковых факторов.
5. Провести обработку и кодирование категориальных факторов.
6. Построить и провести анализ множественной модели регрессии с учетом всех предложенных факторов.
7. Удалить незначимые факторы. Построить окончательную модель.
8. Проверить остатки модели на нормальность.
9. 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.
9. Проверить гипотезу об адекватности модели.

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

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=00000>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Методические указания по проведению лабораторных работ.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Кабанова Т.В. Применение пакета R для решения задач прикладной статистики: учебное пособие: [для студентов и аспирантов университетов]. – Томск: Издательский Дом Томского государственного университета. 2019. 124 с.

2. Мыльников Л.А. Статистические методы интеллектуального анализа данных. – СПб.: БХВ-Петербург, 2021. – 240 с.

3. Дж. Д. Лонг, Пол Титор. R. Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 510 с.

б) дополнительная литература:

1. Кендалл М., Стьюарт А. Статистические выводы и связи. Наука. Физматлит. 1973. 432 с.

2. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. Финансы и статистика. 1989. 608 с.

3. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 1. ЮНИТИ-ДАНА. 2001, 270 с.

4. Айвазян С.А. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 2. ЮНИТИ-ДАНА. 2001, 432 с.

5. Марголис Н.Ю., Кабанова Т.В. Прикладная статистика: учебно-методическое пособие. Ч. 1. Том. гос. ун-т. 2007. 46 с.

6. Марголис Н.Ю., Кабанова Т.В. Прикладная статистика: учебно-методическое пособие. Ч. 2. Том. гос. ун-т. 2007. 58 с.

7. Джеймс Г., Уиттон Д., Хастис Е., Тибширани Р. Введение в статистическое обучение с примерами на языке R. – М.: ДМК Пресс, 2016. 450 с.

в) ресурсы сети Интернет:

– Статистика в Data Science – исчерпывающий гид для амбициозных практиков ML <https://habr.com/ru/company/skillfactory/blog/526972/>

– Введение в Data Science и машинное обучение <https://stepik.org/course/4852>

- 10 примеров использования статистических методов в проекте машинного обучения <https://www.machinelearningmastery.ru/statistical-methods-in-an-applied-machine-learning-project/>
- Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных www.machinelearning.ru/

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.)
- R The R Foundation, США свободно распространяемое.
- RStudio RStudio, PBC, США свободно распространяемое.
- JASP Амстердамский университет, Нидерланды свободно распространяемое.

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных (*при наличии*):

- Искусственный интеллект и сферы его применения. Новости разработки квантовых компьютеров. Исследования искусственных нейронных сетей. <https://ai-news.ru>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Лаборатории, оборудованные персональными компьютерами, соответствующим необходимым программным обеспечением, выходом в интернет.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешанном формате («Актру»).

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, кандидат физ.-мат. наук, доцент, кафедра ТВиМС ИПМКН ТГУ.

МИНОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ

Директор института прикладной
математики и компьютерных наук

_____ А.В. Замятин

« _____ » _____ 2022 г.

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

Статистический анализ данных

по направлению подготовки

02.04.02 Фундаментальная информатика и информационные системы

Направленность (профиль) подготовки:

Математика беспроводных сетей связи и интернета вещей

Томск–2022

ОС составил(и):

канд. физ.-мат. наук, доцент

доцент кафедры теории вероятностей и математической статистики

Т.В. Кабанова

Рецензент:

д-р техн. наук, профессор,

профессор кафедры прикладной математики

И.О. Фамилия

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

Протокол от _____ 20xx г. № _____

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 математических основ и 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-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области.</p>	<p>ИОПК-3.1 Проводит анализ математических моделей и систем</p>	<p>ОР-3.1.1. 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 математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Фрагментарное, неполное знание без грубых ошибок математических 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-3.2 Применяет математические модели, методы для решения прикладных задач профессиональной деятельности</p>	<p>ОР-3.2.1. 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 математических 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-3.3 Разрабатывает новые алгоритмы и методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования</p>	<p>ОР-3.3.1. Обучающийся 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 математических основах и базовых понятиях, которые необходимы для понимания статистических методов анализа данных.</p>	<p>Фрагментарное, неполное знание без грубых ошибок математических 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-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
2.	Критерии сравнения групп.	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
3.	Корреляционный анализ	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
4.	Парная регрессия	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
5.	Множественная регрессия	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
6.	Задачи классификации	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
7.	Кластерный анализ	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	Лабораторные работы; ответы на вопросы экзамена или теста
8.	Анализ временных рядов	ОР-1.1.1, ОР-3.1.1, ОР-3.2.1, ОР-3.3.1.	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

6. Сформировать выборку из генеральной совокупности с дискретным законом распределения (Пуассона, геометрическое или биномиальное). Объем от 100 до 200 наблюдений. Параметры задать самостоятельно.
7. 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.)
8. Найти теоретические мат ожидание и дисперсию при заданных параметрах. Сравнить найденные точечные оценки с теоретическими характеристиками.
9. Найти оценки параметров соответствующего распределения. Сравнить полученные оценки с заданными теоретическими значениями.
10. Проверить гипотезу о виде распределения.

Лабораторная работа. Предварительная обработка данных

Задание.

7. Импортировать заданный набор данных.
8. Проверить на наличие пропусков и выбросов.
9. Для количественных показателей построить гистограммы.
10. Найти оценки числовых характеристик.
11. Проверить гипотезу о нормальности.
12. 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.

10. Построить графики для визуализации данных и их взаимосвязей.
11. Проверить связи факторов друг с другом и их влияние на зависимую целевую переменную.
12. Построить и провести анализ множественной модели регрессии целевой переменной от всех представленных количественных и порядковых факторов.
13. Провести обработку и кодирование категориальных факторов.
14. Построить и провести анализ множественной модели регрессии с учетом всех предложенных факторов.
15. Удалить незначимые факторы. Построить окончательную модель.
16. Проверить остатки модели на нормальность.
17. 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:

- | | |
|---------------------|--|
| 6. Линейная | $y = a + bx + \varepsilon$ |
| 7. Степенная | $y = a \cdot x^b \cdot \varepsilon$ |
| 8. Экспоненциальная | $y = a \cdot e^{bx} \cdot \varepsilon$ |

9. Логарифмическая $y = a + b \cdot \ln(x) + \varepsilon$
10. Гиперболическая $y = a + \frac{b}{x} + \varepsilon$

Задание.

10. Сгенерировать выборки по n наблюдений по каждой из выше предложенных моделей по примеру линейной модели из учебно-методического пособия. Все необходимые параметры задать самостоятельно.
11. Построить диаграммы рассеяния для исходной модели.
12. Для нелинейных моделей провести линеаризацию и построить диаграммы рассеяния линеаризованных моделей.
13. Найти МНК-оценки параметров модели.
14. Найти дисперсии наблюдений и оценок параметров.
15. Построить доверительные интервалы для неизвестных параметров.
16. Проверить гипотезы о значимости коэффициентов регрессии.
17. 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:

26. Типы данных и способы их представления.
27. Параметрические критерии сравнения групп.
28. Непараметрические критерии сравнения групп.
29. Корреляционный анализ количественных данных.
30. Ранговая корреляция.
31. Корреляционный анализ категоризованных данных.
32. Числовые характеристики оценок параметров парной регрессии.
33. Теорема Гаусса-Маркова для случая парной регрессии.
34. Проверка качества уравнения парной регрессии.
35. Скалярная и матричная записи уравнения множественной регрессии. МНК-оценки параметров. Условия Гаусса-Маркова.
36. Теорема Гаусса-Маркова для множественной регрессии.
37. Оценка дисперсии шума в матричном виде.
38. Проверка гипотез о значениях и значимости параметров множественной регрессии.
39. Доверительные интервалы для параметров и функции множественной регрессии.
40. Случай коррелированных гомоскедастических наблюдений.
41. Случай некоррелированных гетероскедастических наблюдений.
42. Мультиколлинеарность.
43. Фиктивные переменные.
44. Постановка задачи классификации. Логистическая регрессия.
45. Метрики качества бинарного классификатора.
46. ROC-анализ.
47. Типы и примеры алгоритмов кластеризации
48. Расстояния между объектами и между классами.
49. Структура временного ряда.
50. 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