

Направление подготовки 09.03.01. «Информатика и вычислительная техника»
Профиль «Вычислительные машины, комплексы, системы и сети»
РПД Б1.В.ДВ.05.02 «Методы анализа данных»



**Филиал федерального государственного бюджетного образовательного учреждения
высшего образования
«Национальный исследовательский университет «МЭИ»
в г. Смоленске**

УТВЕРЖДАЮ

Зам. директора
по учебно-методической работе
филиала ФГБОУ ВО
«НИУ «МЭИ» в г. Смоленске
В.В. Рожков
« 29 » 08 20 17 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МЕТОДЫ АНАЛИЗА ДАННЫХ

(НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ)

Направление подготовки: **09.03.01. «Информатика и вычислительная техника»**

Профиль **«Вычислительные машины, комплексы, системы и сети»**

Уровень высшего образования: **бакалавриат**

Нормативный срок обучения: **5 лет**

Форма обучения: **заочная**

Год набора: **2017**

Смоленск

Программа составлена с учетом ФГОС ВО – бакалавриат по направлению подготовки 09.03.01 «Информатика и вычислительная техника», утвержденного приказом Минобрнауки России от «12» января 2016 г. № 5.

Программу составил:

к.т.н., доцент

М.А. Свириденкова

«_26_» _____06_____2017 г.

Программа обсуждена и одобрена на заседании кафедры «Вычислительная техника»
«_28» _____06_____2017_ г., протокол №_10_

Заведующий кафедрой вычислительной техники

д.т.н., профессор

А.С. Федулов

«_02_» _____07_____2017 г.

РПД адаптирована для лиц с ограниченными возможностями здоровья и инвалидов

Ответственный в филиале по работе

с ЛОВЗ и инвалидами

Зам. нач. УУ

подпись

Е.В. Зуева

ФИО

«_02_» _____07_____2017 г.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины (модуля): изучение принципов и методов обработки данных, ознакомление с современными средствами обработки и анализа данных, выработка первоначальных навыков построения моделей для решения прикладных задач.

Задачи: ознакомление студентов с методами математического анализа и моделирования, теоретического и экспериментального исследования; приобретение навыков постановки и проведения экспериментальных исследований; умения обосновать правильность выбранной модели, сопоставляя результаты экспериментальных данных и полученных решений; способность использовать математические методы анализа данных для решения практических задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Методы анализа данных» относится к вариативной части программы

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами:

Физика

Теория вероятностей и математическая статистика

Информатика

Инженерная графика

Базы данных

Высшая математика

Вычислительная математика

Электротехника

Электроника

Схемотехника

Программирование

Дискретная математика

Теория алгоритмов

Операционные системы

Компьютерная графика

Технология программирования

Сети и телекоммуникации

Сетевые технологии

Микропроцессорные системы

Моделирование

Теория автоматов

Основы теории управления

Тестирование программного обеспечения

Сопровождение разработки программного обеспечения

Конструирование и технологии средств вычислительной техники

Инженерное проектирование и САПР

Введение в оптимизацию

Теория систем

Введение в цифровую обработку сигналов

Теория сигналов

Перечень последующих дисциплин, для которых необходимы знания, умения и навыки, формируемые данной дисциплиной

- Аппаратная реализация алгоритмов
- Технология проектирования устройств на ПЛИС
- Теория передачи информации
- Методы и средства цифровой связи
- Основы теории надежности
- Надежность и диагностика технических средств
- Проектирование информационных систем
- Информационные технологии
- Корпоративные и ведомственные сети
- Технологические сети для сбора данных и управления
- Интернет-технологии
- Проектирование WEB-приложений
- Практика по получению первичных профессиональных умений и навыков, в том числе первичных умений и навыков научно-исследовательской деятельности
- Исполнительская практика
- Практика по получению профессиональных умений и опыта профессиональной деятельности
- Технологическая практика
- Преддипломная практика
- Государственная итоговая аттестация

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины направлено на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки:

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция	Результаты обучения
ОПК-2 Способностью осваивать методики использования программных средств для решения практических задач	Знает: современный математический аппарат и инструментальные средства для обработки, анализа и систематизации данных. Умеет: использовать современный математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования. Владеет: навыками использования современного математического аппарата и инструментальных средств для решения практических задач анализа данных.
ПК-1 Способностью разрабатывать модели компонентов информационных систем, включая модели	Знает: теоретические основы разработки моделей компонентов информационных систем Умеет: разрабатывать модели компонентов информационных систем с уче-

<p>баз данных и модели интерфейсов "человек - электронно-вычислительная машина"</p>	<p>том их особенностей на основе методов анализа данных. Владеет: навыками формирования моделей компонентов информационных систем.</p>
<p>ПК-3 Способностью обосновывать принимаемые проектные решения, осуществлять постановку и выполнять эксперименты по проверке их корректности и эффективности</p>	<p>Знает: методы и технологии обработки и анализа экспериментальных данных, а также их применение на практике. Умеет проводить предварительную обработку данных, решать задачи регрессионного и дисперсионного анализа, выбирать эффективные модели и методы решения прикладных задач, выполнять эксперименты по проверке корректности моделей и их эффективности. Владеет: навыками построения моделей анализа данных, разработки и отладки программ реализации методов анализа данных, инструментальными средствами для проведения расчетов и обоснования принимаемых решений на основе построенных моделей</p>



4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Структура дисциплины:

Индекс	Наименование	Курс 4										Итого за курс												
		Контроль					Академических часов					Академических часов					3.е.							
		Всего	Кон такт.	Лек	Лаб	Пр	КРП	СР	Контроль	3.е.	Всего	Кон такт.	Лек	Лаб	Пр	КРП		СР	Контроль					
Б1.В.ДВ.05.02	Методы анализа данных	180	16	8	8							160	4	5	180	16	8	8					4	5

ОБОЗНАЧЕНИЯ:

Виды промежуточной аттестации (виды контроля):

Экз - экзамен;

ЗаО - зачет с оценкой;

За - зачет;

Виды работ:

Контакт. – контактная работа обучающихся с преподавателем;

Лек. – лекционные занятия;

Лаб.– лабораторные работы;

Пр. – практические занятия;

КРП – курсовая работа (курсовой проект);

РГР – расчетно-графическая работа (реферат);

СР – самостоятельная работа студентов;

3.е.– объем дисциплины в зачетных единицах.

Содержание дисциплины:

№	Наименование видов занятий и тематик, содержание
1	<p>Лекционные занятия 4 шт. по 2 часа:</p> <p>1.1. Работа с данными: Понятие данных. Измерения. Типы шкал измерений. Дискретные и непрерывные данные. Этапы решения задачи анализа данных и их взаимосвязи. Особенности анализа данных на качественном и количественном уровне. Использование математических методов обработки, анализа и синтеза результатов в профессиональных исследованиях. Обзор математического аппарата и инструментальных средств, используемых для решения практических задач анализа данных.</p> <p>1.2 Методы статистического анализа взаимосвязи признаков, их отличие от методов описательной статистики. Задачи и функции статистических методов. Измерение связи и статистической значимости. Парная и множественная корреляция. Коэффициенты связи. Проверка на статистическую значимость.</p> <p>1.3 Основы регрессионного анализа. Парная линейная регрессия: Понятие «регрессия». Простая линейная взаимосвязь. Построение модели парной линейной регрессии, проверка качества модели регрессии, интерпретация параметров регрессии. Множественная регрессия: Уравнение множественной регрессии: построение модели множественной регрессии, проверка качества модели регрессии.</p> <p>1.4 Моделирование временных рядов. Обобщенная модель динамического ряда: Понятие обобщенной модели динамического ряда. Классификация моделей динамического ряда. Порядок построения обобщенной модели динамического ряда. Оценка параметров модели и ее адекватности. Анализ модели.</p>
2	<p>Лабораторные работы 4 шт. по 2 часа</p> <p>2.1 Предварительная обработка данных. Построение дискретного вариационного ряда. Расчет и анализ его характеристик. Построение интервального вариационного ряда. Расчет и анализ его характеристик.</p> <p>2.2 Исследование корреляционных зависимостей между признаками. Отсев грубых погрешностей. Проверка требований к отбору исходных факторов для изучения корреляционной зависимости между ними. Поле корреляции. Коэффициенты парной линейной корреляции и их анализ. Построение корреляционной матрицы и ее анализ. Множественная корреляция.</p> <p>2.3 Исследование модели линейной регрессии. Парная линейная регрессия. Оценка качества модели. Множественная линейная регрессия. Оценка качества модели.</p> <p>2.4 Исследование модели регрессии. Использование фиктивных переменных в уравнении множественной линейной регрессии. Проверка предпосылок МНК.</p>
3	<p>3.1 Расчетно-графическая работа «Разработка модели объекта, системы, процесса с использованием методов анализа данных»</p>
4	<p>Самостоятельная работа студентов:</p> <p>4.1 Защита лабораторных работ 2.1 – 2.4.</p> <p>4.2 Выполнение РГР.</p> <p>4.3 Подготовка к зачету по дисциплине (оценочные материалы приведены в разделе 6 настоящей РПД).</p>

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Таблица - Образовательные технологии, используемые при реализации различных видов учебной занятости по дисциплине

№ п/п	Виды учебных занятий	Образовательные технологии
1	Лекции	Классическая (традиционная, информационная) лекция
2	Лабораторные работы	Технология выполнения лабораторных работ индивидуально.
3	Самостоятельная работа студентов (внеаудиторная)	Информационно-коммуникационные технологии (доступ к ЭИОС филиала, к ЭБС филиала, доступ к информационно-методическим материалам по дисциплине)
4	Контроль (промежуточная аттестация: зачет с оценкой)	Технология устного опроса

6. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ – ДЛЯ ОЦЕНКИ КАЧЕСТВА ОСВОЕНИЯ ДИСЦИПЛИНЫ

К промежуточной аттестации студентов по дисциплине могут привлекаться представители работодателей, преподаватели последующих дисциплин, заведующие кафедрами.

Оценка качества освоения дисциплины включает как текущий контроль успеваемости, так и промежуточную аттестацию.

Примерные вопросы к защите лабораторной работы №1:

1. Дайте определение случайной величины.
2. В чем отличие случайной переменной от неслучайной (детерминированной)? Какие виды случайных переменных Вы знаете? Приведите примеры.
3. Перечислите основные вероятностные характеристики дискретных случайных величин и дайте их определения.
4. Что представляет собой дискретный вариационный ряд? Какие характеристики можно рассчитать по данным вариационного дискретного ряда?
5. Что такое мода? Поясните этот показатель по данным своего варианта.
6. Что такое медиана? Поясните этот показатель по данным своего варианта.
7. Для каких целей рассчитываются показатели разброса?
8. Что такое «полигон»?
9. Что представляет собой интервальный вариационный ряд? Какие характеристики можно рассчитать по данным интервального вариационного ряда?
10. В каком случае целесообразен переход от дискретного вариационного ряда к интервальному?
11. Что такое полигон? Что такое гистограмма? Для каких целей они используются?
12. Дайте понятие «степенные средние». Для каких целей используются эти характеристики?

13. Дайте понятие позиционные (или структурные) средние. Для каких целей используются эти характеристики?
14. Перечислите основные характеристики разброса случайных величин?

Примерные вопросы к защите лабораторной работы №2:

1. Сформулируйте понятия функциональной и стохастической зависимостей.
2. Какая взаимосвязь случайных величин называется корреляционной?
3. Перечислите требования к отбору исходных факторов для изучения корреляционной зависимости между ними.
4. Поясните требование однородности исходных данных.
5. Поясните, что представляет собой коэффициент вариации и как он рассчитывается?
6. Что представляет собой дифференциальная функция нормального распределения?
7. Назовите основные свойства нормального распределения.
8. Назовите варианты проверки гипотезы о нормальном распределении исходных данных.
9. Поясните, каким образом проверяется гипотеза о нормальном распределении случайной величины на основе коэффициентов асимметрии и эксцесса?
10. Поясните, каким образом проверяется гипотеза о нормальном распределении случайной величины на основе критерия χ^2 (хи-квадрат).
11. Поясните, в чем заключается правило «трех сигм»? Каким образом его используют в практических исследованиях?
12. В каких случаях требуется корректировка исходной информации, предполагаемой для использования в корреляционном анализе?
13. Каким образом осуществляют корректировку исходной информации, используя правило «трех сигм»?
14. Каким образом можно осуществить отсев грубых погрешностей методом максимального относительного отклонения?
15. Каким образом можно осуществить отсев грубых погрешностей с использованием размаха вариации?
16. В чем заключается основная задача корреляционного анализа?
17. Для оценки какой корреляционной зависимости используется выборочный коэффициент корреляции? Каковы его свойства?
18. Что такое «поле корреляции»? Каким образом оно анализируется?
19. Как проверяется значимость коэффициента корреляции?
20. Поясните, что представляет собой матрица парных линейных коэффициентов корреляции?
21. Для совокупности трех случайных величин X, Y, Z получена матрица выборочных коэффициентов корреляции
$$\begin{bmatrix} 1 & 0,4 & -0,7 \\ 0,4 & 1 & 0,6 \\ -0,7 & 0,6 & 1 \end{bmatrix}.$$
 Укажите наиболее тесно связанные пары величины.
22. Каким образом выявляются мультиколлинеарные признаки?
23. Что оценивает выборочный коэффициент множественной корреляции?
24. Поясните практическое использование матрицы парных линейных коэффициентов корреляции.

Тест «Исследование корреляционных зависимостей между признаками» (лабораторная работа №2)

Выбрать правильный ответ.

1. Коэффициент корреляции, равный нулю, означает, что между переменными: а) линейная связь отсутствует; б) существует линейная связь; в) ситуация не определена.

2. Коэффициент корреляции, равный 1, означает, что между переменными: а) линейная связь отсутствует; б) существует линейная связь; в) функциональная зависимость; г) ситуация не определена.

3. В каких пределах изменяется множественный коэффициент корреляции: а) от 0 до 1; б) от -1 до 0; в) от -1 до 1; г) от 0 до 10.

4. В каких пределах изменяется коэффициент детерминации: а) от 0 до 1; б) от -1 до 0; в) от -1 до 1; г) от 0 до 10.

5. Коэффициент детерминации – это: а) квадрат парного коэффициента корреляции; б) квадрат частного коэффициента корреляции; в) квадрат множественного коэффициента корреляции.

6. Величина, рассчитанная по формуле $r = \frac{\overline{xy} - \bar{x} * \bar{y}}{s_x s_y}$, является оценкой: а) коэффициента детерминации; б) парного коэффициента корреляции; в) частного коэффициента корреляции; г) множественного коэффициента корреляции.

7. На практике о наличии мультиколлинеарности обычно судят по корреляционной матрице. Если один из элементов корреляционной матрицы больше _____, то считают, что имеет место мультиколлинеарность и в уравнение регрессии следует включать только один из факторов. Вставьте недостающее значение: а) 0,3; б) 0,5; в) 0,65; г) 0,8; д) 0,9; е) другое значение.

Примерные вопросы к защите лабораторной работы №3:

1. Каковы основные задачи регрессионного анализа?
2. В чем особенность классической модели парной линейной регрессии?
3. В каких случаях для описания результативного признака целесообразно построение модели парной линейной регрессии?
4. Перечислите требования к исходной информации при построении модели парной линейной регрессии.
5. Поясните смысл коэффициента регрессии, назовите способы его оценивания.
6. В чем суть метода наименьших квадратов (МНК)?
7. В каких случаях возможно использование МНК?
8. Поясните предпосылки МНК.
9. Что такое число степеней свободы?
10. Что определяет уровень значимости α ?
11. Какова концепция F-критерия Фишера?
12. Как определяется табличное значение критерия Фишера? Как определяется фактическое (расчетное) значение критерия Фишера?
13. В чем смысл средней ошибки аппроксимации и как она определяется?
14. Что характеризует эластичность?
15. Как строится доверительный интервал коэффициентов регрессии?
16. Как рассчитывается прогнозная оценка результативного признака?
17. Каковы основные задачи множественного регрессионного анализа?
18. Запишите модель множественной линейной регрессии. Поясните, каким образом осуществляется интерпретация модели регрессии.
19. Поясните смысл коэффициентов регрессии, назовите способы их оценивания.
20. Изложите сущность дисперсионного анализа и результаты его представления в Excel.
21. Изложите результаты представления регрессионного анализа в Excel.
22. Что такое число степеней свободы?
23. Что определяет уровень значимости α ?

24. Какова концепция F-критерия Фишера?
25. Как определяется табличное значение критерия Фишера? Как определяется фактическое (расчетное) значение критерия Фишера?
26. Как определяется значимость (существенность) коэффициентов регрессии?
27. Как определяется табличное значение критерия Стьюдента? Как определяется фактическое (расчетное) значение критерия Стьюдента?
28. Перечислите и поясните критерии, по которым сравниваются регрессионные модели, описывающие один и тот же результативный признак.
29. Назовите трудности при построении уравнения множественной линейной регрессии.
30. Изложите сущность явления мультиколлинеарности.
31. Назовите основные направления решения проблемы мультиколлинеарности признаков в регрессионном уравнении.
32. Каким образом строится прогноз по уравнению регрессии?
33. Каким образом строится доверительный интервал оценки прогноза по уравнению регрессии?

Примерные вопросы к защите лабораторной работы №4:

1. Дайте понятие адекватной модели регрессии. Приведите пример. Поясните графически.
2. Дайте понятие точной модели регрессии. Приведите пример. Поясните графически.
3. Поясните, каким образом, адекватность и точность модели регрессии зависят от регрессионных остатков.
4. Поясните сущность условий Гаусса-Маркова.
5. Поясните смысл понятий гомоскедастичность и гетероскедастичность.
6. Поясните последствия гетероскедастичности.
7. Поясните, что означает наличие автокорреляции остатков уравнения регрессии. Каковы последствия наличия этого явления.
8. Какими свойствами обладают оценки коэффициентов уравнения регрессии, рассчитанные по МНК?
9. Перечислите условия, которые помимо предпосылок МНК (условий Гаусса-Маркова) предполагаются при построении классических регрессионных моделей.
10. Как проверить выполнение 1-й предпосылки МНК $M(\mathbf{e}_i) = 0$?
11. Как проверить выполнение 2-й предпосылки МНК $D(\mathbf{e}_i) = \sigma^2$ для всех i ?
12. Как проверить выполнение 3-й предпосылки МНК - значения ε_i независимы между собой?
13. Как проверить выполнение 4-й предпосылки МНК - $\sigma_{\varepsilon_i x_i} = \text{cov}(\varepsilon_i, x_i) = 0$?
14. Каким образом проверяется случайность остатков уравнения регрессии?
15. Каким образом проверяется нормальность распределения остатков уравнения регрессии?
16. Каким образом можно изучить влияние качественных переменных на результативный признак в регрессионной модели?
17. Дайте понятие «фиктивная переменная».
18. Какой эффект будет достигнут включением фиктивных переменных в регрессионную модель – множественную регрессию?
19. Каким образом составляется выборка для построения множественной линейной регрессии с включением фиктивных переменных? Это временная или пространственная выборка?
20. Каким образом рассчитываются коэффициенты уравнения регрессии с фиктивными переменными?

21. Каким образом осуществляется прогнозирование результативного признака по уравнению регрессии с фиктивными переменными?

Примерные вопросы по теме «Моделирование временных рядов»:

1. Назовите основные концепции и предпосылки моделирования временных рядов.
2. Какова роль статистического прогнозирования в принятии научно-обоснованных управленческих решений.
3. Приведите примеры задач прогнозирования процессов.
4. Какие вы знаете виды временных рядов?
5. Перечислите требования, предъявляемые к временным рядам при прогнозировании.
6. Дать понятие компоненты временного ряда.
7. Перечислит этапы первичного анализа компонентного состава временного ряда.
8. Поясните, в чем состоят характерные отличия временных рядов от пространственных выборок?
9. Объясните назначение скользящих средних. Влияние каких компонент временного ряда устраняется с их помощью?
10. Поясните, когда целесообразно использовать простые скользящие средние, а для каких временных рядов предпочтительнее применение взвешенных.
11. Дать понятие аддитивных, мультипликативных, смешанных моделей временных рядов.
12. Каким образом определяется характер сезонности (аддитивный или мультипликативный) на основе графического анализа данных?
13. Что представляют собой трендовая, сезонная, циклическая и случайная компоненты, в чем их отличие?
14. Что представляет собой аддитивная модель временного ряда?
15. В чем отличие сезонной компоненты от циклической? Что у них общего?
16. Что представляет собой мультипликативная модель временного ряда?
17. Какие вы знаете методы проверки гипотезы о существовании тенденции?
18. Назовите методы, используемые при сглаживании временных рядов.
19. Как можно восстановить недостающие уровни временного ряда?
20. Какие вы знаете классы моделей кривых роста (трендовых моделей)?
21. Как можно оценить параметры полиномов?
22. Какие вы знаете характеристики точности моделей?
23. Как проводить анализ случайной компоненты для проверки адекватности выбранных моделей реальному процессу?
24. Как проверить наличие автокорреляции в остатках?
25. Какие требования предъявляются к временным рядам как к исходной информации при прогнозировании?
26. Что такое доверительный интервал прогноза?

Оценочные средства для промежуточной аттестации:

Примерные задания расчетно-графической работы:

Задание 1

Для каждого варианта представлены выборочные значения случайной величины X (выборка объемом $n=45$). X - это результаты регистрации значений затухания сигнала x_i на частоте 1000 Гц коммутируемого канала телефонной сети. Эти значения, измеренные в дБ, в виде вариационно-

го ряда представлены в таблице исходных данных. Необходимо построить вариационный ряд, представить его графически, проанализировать и рассчитать характеристики.

Необходимо:

1. Вычислить выборочное среднее по первым 25 значениям, вычислить выборочное среднее по всему объему выборки. Сравнить.
2. Вычислить выборочную дисперсию по первым 25 значениям, вычислить выборочную дисперсию по всему объему выборки. Сравнить.
3. Рассчитать моду и медиану вариационного ряда ($n=45$).
4. Получить интервальный вариационный ряд по всему объему выборки, разделив выборочные значения на 5 интервалов.
5. Построить гистограмму. Сделать вывод.
6. Рассчитать выборочное среднее, выборочную дисперсию интервального вариационного ряда, моду и медиану.
7. Рассчитать коэффициент вариации, коэффициенты асимметрии и эксцесса и сделать вывод о характере распределения случайной величины X . Пояснить результат.
8. Проверить, включает ли выборка аномальные значения (по правилу «трех сигм»). Сделать вывод по п.7 и 8.

Задание 2

Представлена выборка, включающая результативный признак y и несколько факторных признаков, влияющих на него.

Требуется:

А. Изучить основы корреляционного анализа (построить поля корреляции, получить корреляционную матрицу; проанализировать корреляционную матрицу и сделать вывод о возможности построения регрессии (указать конкретный вид модели).

Б. Изучить основы регрессионного анализа (получить линейное уравнение множественной регрессии, выбрав в качестве зависимой переменной – Y , в качестве независимых переменных X_i , соответствующие варианту (в определении уравнения опираться на результаты п.А); определить коэффициент множественной корреляции и коэффициент детерминации полученной модели; проверить значимость построенной модели, используя уровень значимости $\alpha = 0,05$; если модель значима, дать оценку коэффициентов множественной регрессии на основе t -критерия Стьюдента; пересчитать уравнение множественной регрессии, используя только значимые факторы; проверить адекватность полученной регрессионной модели; выполнить прогнозирование в соответствии с вариантом задания.

Примерные тесты по дисциплине «Методы анализа данных»

1. Для какой из следующих шкал измерений приведен правильный пример?
 - 1 Интервальная шкала – номера домов на улицах.
 - 2 Шкала отношений – температура в градусах Цельсия.
 - 3 Номинальная – время решения задачи студентом.
 - 4 Порядковая – студенты решили, что профессор X самый строгий, профессор Y второй после него, и т.д.
2. Генеральная совокупность - это:
 - 1 все общество;
 - 2 изучаемая социальная группа;
 - 3 множество всех объектов, которые имеют качества, свойства, интересующие исследователя;
 - 4 изучаемый социальный институт.
3. Чем более однородна генеральная совокупность, тем...
 - 1 больше величина возможной ошибки выборки;
 - 2 меньше величина возможной ошибки выборки;

- 3 ошибка выборки не зависит от однородности генеральной совокупности;
- 4 больший объем выборки потребуется исследователю.
4. Понятие «репрезентативность» означает:
 - 1 свойство выборки воспроизводить структуру генеральной совокупности;
 - 2 свойство объекта воспроизводить ситуацию исследования;
 - 3 свойство генеральной совокупности воспроизводить структуру объекта;
 - 4 свойство задач исследования воспроизводить его цель.
5. Построение случайной выборки основывается:
 - 1 на возможностях и желании исследователя;
 - 2 на учете случайных факторов, влияющих на состояние объекта;
 - 3 на вероятностных методах отбора респондентов;
 - 4 на экспертных методах отбора респондентов.
6. Центральную тенденцию характеризуют такие величины, как:
 - 1 среднеквадратичное отклонение;
 - 2 медиана;
 - 3 размах;
 - 4 размер выборочной совокупности.
7. Величина дисперсии (среднеквадратического отклонения) показывает:
 - 1 числовое значение средней величины переменной;
 - 2 силу связи между двумя и более переменными;
 - 3 направление связи между и более переменными;
 - 4 степень разброса всех зафиксированных значений переменной вокруг среднего.
8. Связь между двумя переменными проявляется в:
 - 1 устойчивости (неизменности) значений одной переменной при изменении значений другой;
 - 2 воздействии на характер ответа порядка, в котором вопросы размещены в вопроснике;
 - 3 изменении значений одной переменной при изменении значений другой переменной;
 - 4 наличии одинакового вербального выражения.
9. Корреляция может быть:
 - 1 положительной и отрицательной;
 - 2 линейной и нелинейной;
 - 3 парной и множественной;
 - 4 все перечисленное верно.
10. Нулевая гипотеза – это:
 - 1 гипотеза о равенстве параметра генеральной совокупности заранее заданному значению
 - 2 гипотеза об отсутствии различий;
 - 3 гипотеза о наличии различий;
 - 4 гипотеза, альтернативная той, которую мы хотим проверить.
11. p -уровень – это:
 - 1 рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы;
 - 2 пороговый уровень статистической значимости (популярными являются 10 %, 5 %, 1 %, и 0,1 %).
 - 3 вероятность ошибочного неотклонения нулевой гипотезы;
 - 4 вероятность принятия нулевой гипотезы;
12. Статистический критерий – это:
 - 1 метод сравнения выборочной совокупности с теоретическим распределением;
 - 2 критерий, по которому принимается решение о равенстве или неравенстве средних значений в двух выборках;
 - 3 решающее правило, по которому на основе результатов наблюдений принимается решение в задаче проверки статистических гипотез;
 - 4 правило, по которому определяется зависимость или независимость двух выборок.

13. Параметрические статистические критерии:
 - 1 предполагают независимость выборок;
 - 2 проверяют, что вычисляемое выборочное значение равно определенному параметру;
 - 3 проверяют наличие различий между выборками;
 - 4 Предполагают, что вид распределения или функция распределения выборки нам заданы.
14. Какой из перечисленных ниже методов анализа не является статистическим методом анализа взаимосвязи признаков?
 - 1 Корреляционный.
 - 2 Регрессионный.
 - 3 Контент-анализ.
 - 4 Факторный.
15. Что из перечисленного является целью кластерного анализа?
 - 1 Определение взаимосвязей между переменными.
 - 2 Сокращение числа переменных x для описания данных.
 - 3 Классификация многомерных данных наблюдений.
 - 4 Поиск зависимостей в экспериментальных данных.

Примерные вопросы к зачету по дисциплине «Методы анализа данных»

1. Анализ данных и его основные этапы. Цели и задачи анализа.
2. Сбор и подготовка данных.
3. Предварительная обработка результатов эксперимента.
4. Дискретный вариационный ряд и его характеристики.
5. Корреляционный анализ: основные понятия. Корреляционный анализ и его использование в анализе данных.
6. Корреляционный анализ. Требования к отбору факторов для расчета коэффициентов корреляции.
7. Проверка гипотез о нормальном распределении признаков.
8. Исследование линейной зависимости Y от единственной объясняющей переменной X : парный линейный коэффициент корреляции. Поле корреляции. Проверка гипотез о наличии линейной корреляционной связи.
9. Расчет парного линейного коэффициента корреляции. Построение корреляционной матрицы. Анализ корреляционной матрицы. Мультиколлинеарность и способы ее устранения.
10. Статистическое оценивание параметров. Статистические оценки и их основные свойства: состоятельность, несмещенность, эффективность.
11. Регрессионный анализ: основные понятия. Регрессионный анализ и его использование в анализе данных.
12. Регрессионный анализ. Требования к отбору признаков для построения модели регрессии.
13. Метод наименьших квадратов (МНК). Предпосылки МНК.
14. Построение и исследование модели парной линейной регрессии. Интерпретация модели парной линейной регрессии.
15. Расчет и интерпретация коэффициента корреляции для парной линейной регрессии. Коэффициент детерминации и его характеристика.
16. Дисперсионный анализ: сущность и методика проведения.
17. F- тест на значимость уравнения регрессии. Использование t-статистики для оценки значимости коэффициентов уравнения регрессии. Оценка точности уравнения регрессии.
18. Нелинейные регрессии и их характеристика. Основные подходы к построению модели нелинейной регрессии.
19. Множественный регрессионный анализ. Основные понятия.

20. Построение и исследование модели множественной линейной регрессии.
21. Отбор факторных признаков при построении множественной регрессии. Мультиколлинеарность. Трудности построения модели множественной регрессии.
22. Предпосылки метода наименьших квадратов. Гомоскедастичность и гетероскедастичность остатков. Обнаружение гетероскедастичности.
23. Автокорреляция остатков уравнения регрессии. Ее причины и обнаружение.
24. Линейные регрессионные модели с переменной структурой. Введение «манекенов» (фиктивных переменных) в линейную модель регрессии.
25. Понятие временного ряда. Требования, предъявляемые к исходной информации для построения модели временного ряда.
26. Обобщенная модель динамического ряда.
27. Моделирование тенденции временного ряда. Основные типы трендов и их распознавание.
28. Выявление сезонной компоненты во временном ряде. Сезонная корректировка.
29. Ряд Фурье и его применение в оценке динамического ряда.
30. Автокорреляция уровней ряда динамики и связанные с ней факторы. Обнаружение автокорреляции. Учет автокорреляции при моделировании динамических рядов.
31. Многомерный анализ данных. Цели и задачи анализа. Метод многомерной средней: постановка задачи и алгоритм реализации.
32. Однофакторный дисперсионный анализ: постановка задачи и алгоритм реализации.
33. Кластерный анализ. Постановка задачи.
34. Кластерный анализ. Подготовка данных для кластеризации.
35. Методы кластерного анализа. Графическое представление результатов кластеризации.
36. Факторный анализ. Цели и задачи факторного анализа. Основные понятия.
37. Факторный анализ. Постановка задачи. Линейная модель факторного анализа.
38. Факторный анализ. Алгоритмы решения задач факторного анализа. Метод главных компонент.

В филиале используется система с традиционной шкалой оценок – "отлично", "хорошо", "удовлетворительно", "неудовлетворительно", "зачтено", "не зачтено" (далее - пятибалльная система).

Форма промежуточной аттестации по настоящей дисциплине – **зачет с оценкой**.

Применяемые критерии оценивания по дисциплинам (в соответствии с инструктивным письмом НИУ МЭИ от 14 мая 2012 года № И-23):

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
«отлично»/ «зачтено (отлично)»/ «зачтено»	Выставляется обучающемуся, обнаружившему всестороннее, систематическое и глубокое знание материалов изученной дисциплины, умение свободно выполнять задания, предусмотренные программой, усвоивший основную и знакомый с дополнительной литературой, рекомендованной рабочей программой дисциплины; проявившему творческие способности в понимании, изложении и использовании материалов изученной дисциплины, безупречно ответившему не только на вопросы билета, но и на дополнительные вопросы в рамках рабочей программы дисциплины, правильно выполнившему практическое задание. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущего контроля.

Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
	Компетенции, закреплённые за дисциплиной, сформированы на уровне – «эталонный».
«хорошо»/ «зачтено (хорошо)»/ «зачтено»	Выставляется обучающемуся, обнаружившему полное знание материала изученной дисциплины, успешно выполняющему предусмотренные задания, усвоившему основную литературу, рекомендованную рабочей программой дисциплины; показавшему систематический характер знаний по дисциплине, ответившему на все вопросы билета, правильно выполнивший практическое задание, но допустивший при этом непринципиальные ошибки. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущего контроля. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «продвинутый».
«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	Выставляется обучающемуся, обнаружившему знание материала изученной дисциплины в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии, справляющемуся с выполнением заданий, знакомому с основной литературой, рекомендованной рабочей программой дисциплины; допустившему погрешность в ответе на теоретические вопросы и/или при выполнении практических заданий, но обладающему необходимыми знаниями для их устранения под руководством преподавателя, либо неправильно выполнившего практическое задание, но по указанию преподавателя выполнившего другие практические задания из того же раздела дисциплины.. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «пороговый».
«неудовлетворительно»/ не зачтено	Выставляется обучающемуся, обнаружившему серьезные пробелы в знаниях основного материала изученной дисциплины, допустившему принципиальные ошибки в выполнении заданий, не ответившему на все вопросы билета и дополнительные вопросы и неправильно выполнившего практическое задание (неправильное выполнение только практического задания не является однозначной причиной для выставления оценки «неудовлетворительно»). Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение по образовательной программе без дополнительных занятий по соответствующей дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущего контроля. Компетенции на уровне «пороговый», закреплённые за дисциплиной, не сформированы.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебное и учебно-лабораторное оборудование

Учебная аудитория для проведения занятий лекционного типа, оснащенная:

- специализированной мебелью; доской аудиторной.

Учебная аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная:

- специализированной мебелью; доской аудиторной.

Учебная аудитория для лабораторных работ, выполняемых в компьютерном классе, оснащенная:

- специализированной мебелью; доской аудиторной; персональными компьютерами с подключением к сети "Интернет" и доступом в ЭИОС филиала.

Для самостоятельной работы обучающихся по дисциплине используется помещение для самостоятельной работы обучающихся, оснащенное:

- специализированной мебелью; доской аудиторной; персональным компьютерами с подключением к сети "Интернет" и доступом в ЭИОС филиала.

8. ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ И ИНВАЛИДОВ

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

для слепых и слабовидящих:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

- письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;

- обеспечивается индивидуальное равномерное освещение не менее 300 люкс;

- для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;

- письменные задания оформляются увеличенным шрифтом;

- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

для глухих и слабослышащих:

- лекции оформляются в виде электронного документа;

- письменные задания выполняются на компьютере в письменной форме;

- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

для лиц с нарушениями опорно-двигательного аппарата:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

- письменные задания выполняются на компьютере со специализированным программным обеспечением;

- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере;

- используется специальная учебная аудитория для лиц с ЛОВЗ – ауд. 106 главного учебного корпуса по адресу 214013, г. Смоленск, Энергетический пр-д, д.1, здание энергетического института (основной корпус).

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены филиалом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

для слепых и слабовидящих:

- в печатной форме увеличенным шрифтом;
- в форме электронного документа;
- в форме аудиофайла.

для глухих и слабослышащих:

- в печатной форме;
- в форме электронного документа.

для обучающихся с нарушениями опорно-двигательного аппарата:

- в печатной форме;
- в форме электронного документа;
- в форме аудиофайла.

9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература.

1. Крутиков, В.Н. Анализ данных [Электронный ресурс]: учеб. пособие / В.Н. Крутиков, В.В. Мешечкин - Кемерово: Кемеровский государственный университет, 2014. - 138 с. Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=278426> – Библиогр. в кн. – ISBN 978-5-8353-1770-7. – Текст : электронный.
2. Прикладная математическая статистика [Электронный ресурс]: учеб. пособие / сост. А.А. Мицель - Томск: ТУСУР, 2016. - 113 с. Режим доступа - <http://biblioclub.ru/index.php?page=book&id=480889>.

Дополнительная литература.

1. Большаков А.А. Методы обработки многомерных данных и временных рядов: учеб. Пособие по напр. «Информатика и вычислительная техника»/ А.А. Большаков, Р.Н. Каримов. – 2-е изд., стер. – М.: Горячая линия – Телеком, 2014. – 520 с. : ил. – Библиогр. в конце глав. – ISBN 978-5-7782-3183-2. – Текст электронный. (1 ч/з)
2. Новикова, Е.Н. Компьютерная обработка результатов измерений [Электронный ресурс] : учеб. пособие / Е.Н. Новикова, О.Л. Серветник - Ставрополь : СКФУ, 2017. - 182 с. Режим доступа - <http://biblioclub.ru/index.php?page=book&id=483751>.
3. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python: [16+] / В.М. Волкова, М.А. Семенова, Е.С. Четвертакова, С.С. Вожов ; Новосибирский государственный технический университет. – Новосибирск: Новосибирский государственный технический университет, 2017. – 74 с.: ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576496>. – Библиогр.: с. 48. – ISBN 978-5-7782-3183-2. – Текст: электронный.



ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

Но- мер изме- мене- ния	Номера страниц				Всего стра- ниц в доку- менте	Наименование и № документа, вводящего изменения	Подпись, Ф.И.О. внесшего измене- ния в данный эк- земпляр	Дата внесения из- менения в данный эк- земпляр	Дата введения из- менения
	изме- нен- ных	заме- нен- ных	но- вых	анну- лиро- ванн- ых					
1	2	3	4	5	6	7	8	9	10