

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ (МОДУЛЯ)/ПРАКТИКИ

АНАЛИЗ ДАННЫХ

наименование дисциплин (модуля)/практики

Автор: Вепрев С. Б.

Код и наименование направления подготовки, профиля:

38.03.05 Бизнес-информатика, профиль Бизнес-аналитика

Квалификация (степень) выпускника: Бакалавр

Форма обучения: Очная

Цель освоения дисциплины:

Сформировать компетенции:

проведение анализа инноваций в экономике, управлении и информационно-коммуникативных технологиях (ПК-4)

консультировать заказчиков по вопросам создания и развития электронных предприятий и их компонент (ПК-23)

способность использовать лучшие практики продвижения инновационных программно-информационных продуктов и услуг (ПК-27)

План курса:

Модуль 1. Статистический анализ данных

Тема 1. Параметрическая и непараметрическая проверка гипотез.

Нормальные параметрические критерии для проверки гипотез: гипотезы о положении, гипотезы о рассеивании. Систематизация критериев. Гипотеза о равенстве средних: критерий Стьюдента для одной и двух выборок, связанные выборки, гипотеза сдвига, метод множественных сравнений Шеффе, метод LSD. Гипотеза о равенстве дисперсий: критерий Фишера, критерий Кокрена, критерий Бартлетта. Критерии нормальности: Критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий омега-квадрат фон Мизеса, критерий хи-квадрат (Пирсона). Упрощённые проверки нормальности по асимметрии и эксцессу. Непараметрические ранговые критерии для проверки гипотез: критерий Уилкоксона-Манна-Уитни, гипотезы о положении, гипотезы о рассеивании. Элементы теории измерений: номинальные, порядковые и количественные переменные; инварианты. Разновидности средних: по Коши, по Колмогорову, мода, медиана. Среднее в порядковой шкале. Вариационный ряд, ранги и связки. Ранговые критерии: критерий Уилкоксона-Манна-Уитни, критерий знаков, критерий Уилкоксона двухвыборочный, критерий Уилкоксона для связанных выборок, критерий Краскела-Уоллиса, критерий Зигеля-Тьюки, медианный критерий: одновыборочный и двухвыборочный. Доверительные интервалы для медианы (Уилкоксона-Мозеса) и сдвига (Уилкоксона-Тьюки). Множественные сравнения на основе рангов Фридмана.

Тема 2. Дисперсионный анализ (ANOVA).

Модели факторного эксперимента. Примеры: факторы, влияющие на успешность решения математических задач; факторы, влияющие на объёмы продаж. Однофакторная параметрическая модель: метод Шеффе. Однофакторная непараметрическая модель:

критерий Краскела-Уоллиса, критерий Джонкхиера. Общий случай модели с постоянными факторами, теорема Кокрена. Двухфакторная непараметрическая модель: критерий Фридмана, критерий Пейджа. Двухфакторный нормальный анализ. Ковариационный анализ (постановка задачи).

Тема 3. Корреляционный и линейный регрессионный анализ.

Корреляция Пирсона, значимость коэффициента корреляции (критерий Стьюдента). Частная корреляция, значимость коэффициента частной корреляции. Множественная корреляция, значимость коэффициента множественной корреляции. Ранговая корреляция, коэффициент корреляции Спирмена, коэффициент корреляции Кенделла. Конкордация Кенделла. Многомерная линейная регрессия. Метод наименьших квадратов. МНК-решение и его запись через сингулярное разложение. Остаточная сумма квадратов (RSS). Основные предположения многомерной линейной регрессии. Статистические свойства МНК-оценок без предположения нормальности. Статистические свойства МНК-оценок при предположении нормальности. Доверительные интервалы для дисперсии шума, коэффициентов регрессии, прогнозного значения отклика.

Тема 4. Анализ регрессионных моделей и регрессионных остатков.

Анализ структуры линейной регрессионной модели. Значимость коэффициентов линейной регрессии (проверка равенства коэффициентов нулю), вложенные модели линейной регрессии, критерий Фишера. Шаговая регрессия. Проверка адекватности модели. Выборочный коэффициент детерминации. Дисперсия остатков. Проблема мультиколлинеарности. Методы понижения размерности: ридж-регрессия, лассо Тибширани, параметр селективности. Анализ регрессионных остатков: визуальный анализ. Непараметрические тесты: критерий Уилкоксона-Манна-Уитни, критерий Зигеля-Тьюки, критерий знаков, критерий серий, критерий экстремумов. Проверка нормальности остатков: Критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий омега-квадрат фон Мизеса, критерий хи-квадрат Пирсона, критерии асимметрии и эксцесса. Тест на корреляцию остатков, статистика Дарбина-Уотсона.

Тема 5. Анализ временных рядов.

Идентификация модели временных рядов. Систематическая составляющая и случайный шум. Два общих типа компонент временных рядов. Подгонка функции. Анализ тренда. Анализ сезонности. Автокорреляционная диаграмма. Исследование коррелограмм. Удаление периодической зависимости. Модель Бокса-Дженкинса и автокорреляции. Процесс авторегрессии и скользящего среднего. Требование стационарности. Обратимость. Модель авторегрессии и скользящего среднего. Идентификация. Оценивание и прогноз. Идентификация: число оцениваемых параметров, сезонные модели. Оценивание параметров: сравнение методов, стандартные ошибки оценок, штраф. Оценивание модели: оценки параметров, сравнение прогноза, анализ остатков, ограничения. Прерванные временные ряды. Простое экспоненциальное сглаживание. Выбор и оценивание лучшего значения параметра альфа. Индексы качества подгонки: средняя, средняя абсолютная и средняя относительная ошибка, среднеквадратичная ошибка. Сезонная и несезонная модели с трендом или без тренда. Сезонная декомпозиция. Общая модель. Аддитивная и мультипликативная сезонность. Аддитивный и мультипликативный тренд-цикл. Сезонная корректировка. Общая модель. Аддитивная и мультипликативная сезонность. Аддитивный и мультипликативный тренд-цикл. Выбросы. Последовательные уточнения. Критерии и итоговые статистики. Этапы процедуры корректировки. Анализ распределенных лагов. Общая модель. Распределенный лаг Алмона. Неправильная спецификация.

Тема 6. Метод главных компонент. Факторный анализ.

Переход к системе ортонормированных линейных комбинаций в исходном пространстве признаков. Собственные векторы корреляционной матрицы. Эллипсоид рассеивания. Собственные значения главных компонент. Процент дисперсии, приходящийся на каждую главную компоненту. Накопленный процент дисперсии.

Латентные переменные. Факторные нагрузки. Максимально возможное количество факторов. Основное соотношение факторного анализа. Неоднозначность решения задачи факторного анализа. Методы факторного анализа.

Модуль 2. Глубинный анализ данных (Data mining).

Тема 7. Основные системы интеллектуального анализа данных.

Этапы проекта интеллектуального анализа данных. Межотраслевой стандартный процесс для интеллектуального анализа данных (CRISP). Ступени процесса CRISP: понимание бизнеса, понимание данных, подготовка данных, моделирование, оценка, развертывание. Процесс SEMMA. Ступени процесса SEMMA: исследование данных, изменение данных, моделирование, оценка полезности и надежности.

Тема 8. Методы интеллектуального анализа данных.

Статистический анализ: дисперсионный анализ, кластерный анализ, регрессионный анализ, дискриминантный анализ. Методы, основанные на деревьях решений. Генетические алгоритмы. Нейронные сети. Нечеткая логика. Временные ряды. Методы добычи данных: ассоциация, классификация, кластеризация, прогноз временных рядов.

Тема 9. Топологический анализ данных.

Основные задачи. Основные методы топологического анализа данных. Симплициальные комплексы. Параметр близости. Алгебраические топологии. Основы теории устойчивых гомологий. Числа Бетти. Облако точек в Евклидовом пространстве. Преобразование облака точек в метрическом пространстве. Устойчивая гомология.

Формы текущего контроля и промежуточной аттестации:

Форма промежуточной аттестации – зачет.

В результате освоения дисциплины обучающийся знает, умеет, владеет:

Код этапа освоения компетенции	Результаты обучения
12.3	на уровне знаний: знать: - основы теории параметрической и непараметрической проверки гипотез; - основы теории дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов; - основы теории корреляционного и линейного регрессионного анализа; - методологию анализа регрессионных моделей и регрессионных остатков; - структуру систем интеллектуального и топологического анализа данных.
	на уровне умений: уметь применить основные методы статистического и глубинного анализа в решении экономических задач.
	на уровне навыков: владеть: - навыками применения методов параметрической и непараметрической проверки гипотез, дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов, корреляционного и регрессионного анализа для решения экономических задач; - методикой построения, анализа и применения регрессионных моделей для оценки состояния и прогноза развития экономических явлений и процессов; - навыками применения методов интеллектуального и топологического анализа данных для решения экономических задач.
25.1	на уровне знаний: знать - основы теории параметрической и непараметрической проверки гипотез; - основы теории дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов; - основы теории корреляционного и линейного регрессионного анализа; - методологию анализа регрессионных моделей и регрессионных остатков;

Код этапа освоения компетенции	Результаты обучения
	<p>структуру систем интеллектуального и топологического анализа данных</p> <p>на уровне умений: умеет применить основные методы статистического и глубинного анализа в решении экономических задач.</p> <p>на уровне навыков: владеет навыками применения методов параметрической и непараметрической проверки гипотез, дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов, корреляционного и регрессионного анализа для решения экономических задач; - методикой построения, анализа и применения регрессионных моделей для оценки состояния и прогноза развития экономических явлений и процессов; - навыками применения методов интеллектуального и топологического анализа данных для решения экономических задач.</p>
27.2	<p>на уровне знаний: знать:</p> <ul style="list-style-type: none"> - основы теории параметрической и непараметрической проверки гипотез; - основы теории дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов; - основы теории корреляционного и линейного регрессионного анализа; - методологию анализа регрессионных моделей и регрессионных остатков; <p>структуру систем интеллектуального и топологического анализа данных</p> <p>на уровне умений: умеет применить основные методы статистического и глубинного анализа в решении экономических задач.</p> <p>на уровне навыков: владеет навыками применения методов параметрической и непараметрической проверки гипотез, дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов, корреляционного и регрессионного анализа для решения экономических задач; - методикой построения, анализа и применения регрессионных моделей для оценки состояния и прогноза развития экономических явлений и процессов; - навыками применения методов интеллектуального и топологического анализа данных для решения экономических задач. - изучение основ математической теории и методов анализа данных, как статистического, так и глубинного анализа данных. Связать теорию и практику, научить студентов «видеть» статистические задачи в различных предметных областях. Правильно применять методы прикладной статистики, показать на практических примерах возможности и ограничения статистических методов.</p>

Информационные технологии, программное обеспечение, материально-техническая база, оценочные средства, необходимые для освоения дисциплины, адаптированы для обучающихся инвалидов и обучающихся с ограниченными возможностями здоровья.

Основная литература:

1. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2012