Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 23.10.2023 18:02:48
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Декан факультета информационных технологий

/Д. Г. Демидов/

«28» 04 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Анализ данных»

Направление подготовки

09.03.03 Прикладная информатика

Образовательная программа (профиль подготовки)

«Большие и открытые данные»

Квалификация (степень) выпускника

бакалавр

Форма обучения

очная

Москва 2022

Программа дисциплины «Анализ данных» составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению **09.03.03 «Прикладная информатика»** и профилю подготовки **«Большие и открытые данные».**

Программу составил



_____/ Н.И. Царькова /

Программа дисциплины утверждена на заседании кафедры «Прикладная информатика»
«____» августа 2022 г. протокол № _____

Заведующий кафедрой
доцент, к.э.н.



/С.В.Суворов/

Программа согласована с руководителем образовательной программы по направлению подготовки **09.03.03 «Прикладная информатика»** по профилю подготовки **«Большие и открытые данные».**



_____ /С.В.Суворов/
«____» августа 2022 г.

Программа утверждена на заседании учебно-методической комиссии факультета Информационных технологий

Председатель комиссии  / Д. Г. Демидов/

«____»_____2022 г. Протокол:

Цели освоения дисциплины.

К **основным целям** освоения дисциплины «Анализ данных» следует отнести:

– формирование теоретических знаний о современных принципах, методах и средствах анализа данных, практических умений и навыков по применению современных методов анализа данных в различных сферах человеческой деятельности.

– подготовка студентов к деятельности в соответствии с квалификационной характеристикой бакалавра по направлению, в том числе формирование умений по выявлению необходимых усовершенствований и разработке новых, более эффективных средств анализа данных.

К **основным задачам**освоения дисциплины «Анализ данных» следует отнести:

– изучение существующих технологий подготовки данных к анализу;

- изучение основных методов поиска закономерностей, связей, правил в табулированных массивах данных большого объема; иллюстрированного их применения в различных областях деятельности;

-изучение методов статистического анализа данных;

- овладение практическими умениями и навыками реализации технологий анализа данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями;

- изучение принципов организации и проведения аналитического исследования; решение практических задач с целью прогнозирование и выработка рекомендаций.

- формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

Для успешного освоения дисциплины необходимы знания по следующим
дисциплинам и разделам ОП:
Теоретические основы информатика;
Математический анализ;
Линейная алгебра;
Теория вероятностей и математическая статистика.
Знания, умения и приобретенные компетенции будут
использованы при изучении
следующих дисциплин ОП:
Финансовые вычисления;
Моделирование бизнес-процессов;
Проектирование и разработка баз данных;
Интеллектуальный анализ данных.

1. Место дисциплины в структуре ООП бакалавриата.

Дисциплина «Анализ данных» относится к числуучебных дисциплин базовой части базового цикла (Б1) основной образовательной программы бакалавриата.

«Анализ данных» взаимосвязан логически и содержательно - методически со следующими дисциплинами и практиками ООП:

В базовой части базового цикла (Б1):

-Теория вероятностей
-Математическая статистика;

- Теоретические основы информатики.

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы.

В результате освоения дисциплины (модуля) у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующих компетенций:

Код компетенции	**В результате освоения образовательной программы обучающийся должен обладать**	**Перечень планируемых результатов обучения по дисциплине**
ПК-2	способностью понимать, совершенствовать и применять современный математический аппарат	**знать:** -основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных; - возможности отечественных и зарубежных универсальных программных средств и аналитических платформ, применяемых для анализа данных; **уметь:** - создавать хранилища данных, выполнять их загрузку, извлекать данные из хранилищ; - применять технологии интеллектуального анализа электронных массивов данных для решения конкретных практических проблем; **владеть:** - методами работы с пакетом анализа данных, предназначенных для решения сложных статистических и социологических задач, подробно рассмотрев весь набор инструментов, входящих в пакет анализа данных

4. Структура и содержание дисциплины.

Общая трудоемкость дисциплины составляет 4 зачетных единицы, т.е. 144 академических часов (из них 72 часов – самостоятельная работа студентов).

Разделы дисциплины «Анализа данных» изучаются на втором курсе в третьем семестре.

Третий семестр: лекции– 2 часа в неделю (36 часов), лабораторные работы – 2 час в неделю (36 часов), форма контроля - зачет.

Структура и содержание дисциплины «Анализ данных» по срокам и видам работы отражены в приложении.

Содержание разделов дисциплины

Третий семестр

1. **Введение**

   Предмет, задачи и содержание дисциплины.

   Содержание: Роль анализа данных в теории познания. Многообразие анализа данных. Основные этапы развития методов и средств анализа данных. Структура курса, его место и роль в подготовке бакалавра, связь с другими дисциплинами и ознакомление с компетенциями, которыми должны обладать студенты после изучения данной дисциплины. (ПК-2).

2. **Постановка задачи в анализе данных.**
   Содержание:Формула для вероятностей сумм независимых, одинаково распределенных, решетчатых случайных величин выраженная через сумму вероятностей полиномиального распределения. (ПК-2)

3. **Основы проверки статистических гипотез.**
   Содержание: Статистические модели. Проверка статистических гипотез. Схема испытаний Бернулли. Критерий знаков для одной выборки. Критерий Манна-Уитни. Критерий Уилкоксона. Анализ повторных парных наблюдений с помощью знаковых рангов. (ПК-2)

4. **Анализ одной и двух нормальных выборок.**
   Содержание: Об исследовании нормальных выборок. Одна выборка. Две выборки. Парные данные. (ПК-2)

5. **Однофакторный анализ.**
   Содержание: Постановка задачи. Непараметрические критерии проверки однородности. Критерий Краскела-Уоллиса. Критерий Джонкхиера. Практический пример. (ПК-2)

6. **Дисперсионный анализ.**
   Содержание: Оценивание эффектов обработки. Оценивание эффектов обработки в нормальной модели. Доверительные интервалы. Метод Шеффе множественных сравнений. (ПК-2)

7. **Двухфакторный анализ.**
   Содержание: Связь задач двухфакторного и однофакторного анализа.Таблица двухфакторного анализа. Непараметрические критерии проверки гипотезы об отсутствии эффектов обработки. Аддитивная модель данных двухфакторного эксперимента при независимом действии факторов. (ПК-2.)

8. **Линейный и нелинейный регрессионный анализ.**
   Содержание: Модель линейного регрессионного анализа. О стратегии, методах и проблемах регрессионного анализа. Простая линейная регрессия. (ПК-2)

9. **О проверке предпосылок в задаче регрессионного анализа.**
   Содержание: Непараметрическая линейная регрессия. Практический пример. Регрессионный анализ в пакетах Excel, STADIA и SPSS.(ПК-2)

10. **Независимость признаков.**
Содержание: Инструменты и стратегия исследования связи признаков. Связь признаков в количественных шкалах. Нормальная корреляция. Коэффициент корреляции. (ПК-2)

11. **Критерий согласия.**
Содержание: Критерии согласия Колмогорова и омега-квадратв случае простой гипотезы. Критерий согласия хи-квадрат К.Пирсона для простой гипотезы. Критерии согласия для сложной гипотезы. Критерий согласия хи-квадрат Фишера для сложной гипотезы. (ПК-2)

12. **Временные ряды.**
Содержание: Анализ временных рядов и его разделы. Цели, этапы и методы анализа временных рядов. Порядок анализа временных рядов.
Графические методы анализа временных рядов. . Методы исследования структуры стационарного временного ряда. Цели и методы анализа. (ПК-2)

13. **Анализ временных рядов на компьютере.**
Содержание: Анализ временных рядов в Excel и SPSS. Обзор возможностей. Подбор тренда и прогнозирование. Устранение сезонной компоненты. (ПК-2)

14. **Многомерный анализ и другие статистические методы.**
Содержание: Многомерный статистический анализ. Факторный анализ. Дискриминантный анализ. (ПК-2)

15. **Кластерный анализ.**
Содержание: Многомерное шкалирование. Методы контроля качества. Использование статистических пакетов. (ПК-2)

5. Образовательные технологии.

Методика преподавания дисциплины «Анализ данных» и реализация компетентностного подхода в изложении и восприятии материала предусматривает использование следующих активных и интерактивных форм проведения групповых, индивидуальных, аудиторных занятийв сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся:

– подготовка к выполнению лабораторных работ в лабораториях вуза;

– защита и индивидуальное обсуждение выполняемых лабораторных работ;

– организация и проведение текущего контроля знаний студентов в форме бланкового тестирования;

– проведение интерактивных занятий по процедуре подготовки к интернет-тестированию на сайтах: *i-exam.ru, fepo.ru;*

– использование интерактивных форм текущего контроля в форме аудиторного и внеаудиторного интернет-тестирования.

Удельный вес занятий, проводимых в интерактивных формах, определен главной целью образовательной программы, особенностью контингента обучающихся и содержанием дисциплины «Анализ данных» и в целом по дисциплине составляет 50% аудиторных занятий. Занятия лекционного типа составляют33% от объема аудиторных занятий.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов.

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

В третьем семестре

- подготовка к выполнению лабораторных работ и их защита.

Оценочные средства текущего контроля успеваемости включают контрольные вопросы и задания в форме бланкового тестирования, для контроля освоения обучающимися разделов дисциплины.

Образцы тестовых заданий, контрольных вопросов и лабораторных заданий для проведения текущего контроля, вопросов к зачету, приведены в приложении.

6.1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (модулю).

6.1.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.

В результате освоения дисциплины (модуля) формируются следующие компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
ПК-2	способностью понимать, совершенствовать и применять современный математический аппарат

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплины, в том числе лекций илабораторных работ в соответствии с учебным планом и календарным графиком учебного процесса.

6.1.2. Описание показателей и критериев оценивания компетенций, формируемых по итогам освоения дисциплины (модуля), описание шкал оценивания

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине (модулю).

ПК-2 - способностью понимать, совершенствовать и применять современный математический аппарат				
Показатель	**Критерии оценивания**			
	2	**3**	**4**	**5**
знать: статистические и машинные методы классификации и регрессии; -методики анализа и прогнозирования временных рядов; -технологию построения ансамблей и сравнения моделей;	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие следующих знаний: Основных методов регрессионного анализа и временных рядов	Обучающийся демонстрирует неполное соответствие следующих знаний: Основных методов регрессионного анализа и временных рядов. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие следующих знаний: Основных методов регрессионного анализа и временных рядов, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие следующих знаний: Основных методов регрессионного анализа и временных рядов, свободно оперирует приобретенными знаниями.
уметь: использовать возможности отечественных и зарубежных универсальных программных средств и аналитических платформ для поиска закономерностей, связей, правил, знаний в электронных массивах данных;	Обучающийся не умеет или в недостаточной степени умеет выполнять работы в пакете прикладных программах по анализу данных	Обучающийся демонстрирует неполное соответствие следующих умений: выполнять работы в пакете прикладных программах по анализу данных. Допускаются значительные ошибки, проявляется недостаточность умений, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании	Обучающийся демонстрирует частичное соответствие следующих умений: выполнять работы в пакете прикладных программах по анализу данных. Умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.	Обучающийся демонстрирует полное соответствие следующих умений: выполнять работы в пакете прикладных программах по анализу данных. Свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.

-свободно ориентироваться на современном динамичном рынке аналитических программных продуктов.		умениями при их переносе на новые ситуации.		
владеть: методами работы с современными технологиями на компьютере с целью реализации методов математической статистики, общей теории статистики и эконометрики на компьютере	Обучающийся не владеет или в недостаточной степени владеет методами и методиками регрессионно-дисперсионного анализа	Обучающийся владеет методами и методиками регрессионно-дисперсионного анализа в неполном объеме, допускаются значительные ошибки, проявляется недостаточность владения навыками по ряду показателей, Обучающийся испытывает значительные затруднения при применении навыков в новых ситуациях.	Обучающийся частично владеет методами и методиками регрессионно-дисперсионного анализа, навыки освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.	Обучающийся в полном объеме владеет методами и методиками регрессионно-дисперсионного анализа, свободно применяет полученные навыки в ситуациях повышенной сложности.

Шкалы оценивания результатов промежуточной аттестации и их описание:

Форма промежуточной аттестации: зачет.

Промежуточная аттестация обучающихся в форме зачёта проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «зачтено» или «не зачтено».

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине «Анализ данных» (указывается что именно – прошли промежуточный контроль, выполнили лабораторные работы, выступили с докладом и т.д.)

Шкала оценивания	Описание
Зачтено	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Не зачтено	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

Фонды оценочных средств представлены в приложении 1 к рабочей программе.

Приложение 1 к
рабочей программе

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Направление подготовки: 09.03.03 «Прикладная информатика»

Форма обучения: очная

Вид профессиональной деятельности: (В соответствии с ФГОС ВО)

Кафедра: Прикладная информатика

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

ПО ДИСЦИПЛИНЕ

Анализ данных

Состав: 1. Паспорт фонда оценочных средств
2. Описание оценочных средств:

Тест (Т)

Устный опрос, собеседование (УО)

Москва, 2022 год

Таблица 1

ПОКАЗАТЕЛЬ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Анализ данных					
ФГОС ВО 09.03.03 «Прикладная информатика»					
В процессе освоения данной дисциплины студент формирует и демонстрирует следующие **Общепрофессиональные и профессиональныекомпетенции**:					
КОМПЕТЕНЦИИ		**Перечень компонентов**	**Технология формирования компетенц**	**Форма оценочного средства****	**Степени уровней освоения компетенций**
ИН-ДЕКС	**ФОРМУЛИРОВ-КА**				

ПК-2	Способностью работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных компьютерных сетях	**знать:** -статистические и машинные мет классификации и регрессии; -методики анализа и прогнозиров временных рядов; -технологию построения ансамбл сравнения моделей; **уметь:** **-** использовать возможности отечественных и зарубежных универсальных программных ср и аналитических платформ для п электронных массивах данных; -свободно ориентироваться на современном динамичном рынке аналитических программных про **владеть:** методами работы с современным технологиями на компьютере с целью реализации методов мате эконометрики на компьютере	лекция, самостоятель ная работа, лабораторны е работы	УО, Т	**Базовый уровень** - способен проводить анализ данных используя эконометрический инструментарий **Повышенный уровень** **-** способен проводить анализ данных различных социально-экономических процессов, различные пакеты прикладных программ с целью дальнейшего прогнозирования развития

**- Сокращения форм оценочных средств см. в приложении 2 к РП.

Приложение 2
к рабочей программе

Перечень оценочных средств по дисциплине Анализ данных

№ ОС	Наименование оценочногосредства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Устный опрос собеседование, (УО	Средство контроля, организованное как специальная беседа педагогического работника с обучающимся на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п.	Вопросы по темам/разделам дисциплины
2	Тест (Т)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умениййобучающегося.	Фонд тестовых заданий

Примерные тесты
Тестовые задания

Парная регрессия и корреляция

1. Наиболее наглядным видом выбора уравнения парной регрессии является:

а) аналитический;

б) графический;

в) экспериментальный (табличный).

2. Рассчитывать параметры парной линейной регрессии можно, если у нас есть:

а) не менее 5 наблюдений;

б) не менее 7 наблюдений;

в) не менее 10 наблюдений.

3. Суть метода наименьших квадратов состоит в:

а) минимизации суммы остаточных величин;

б) минимизации дисперсии результативного признака;

в) минимизации суммы квадратов остаточных величин.

4. Коэффициент линейного парного уравнения регрессии:

а) показывает среднее изменение результата с изменением фактора на одну единицу;

б) оценивает статистическую значимость уравнения регрессии;

в) показывает, на сколько процентов изменится в среднем результат, если фактор изменится на 1%.

5. На основании наблюдений за 50 семьями построено уравнение регрессии $y = 284{,}56 + 0{,}672x$ ***, где*** y ***– потребление,*** x ***– доход. Соответствуют ли знаки и значения коэффициентов регрессии теоретическим представлениям?***

а) да;

б) нет;

в) ничего определенного сказать нельзя.

6. Суть коэффициента детерминации r_{xy}^2 ***состоит в следующем:***

а) оценивает качество модели из относительных отклонений по каждому наблюдению;

б) характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака;

в) характеризует долю дисперсии y , вызванную влиянием не учтенных в модели факторов.

7. Качество модели из относительных отклонений по каждому наблюдению оценивает:

а) коэффициент детерминации r_{xy}^2 ;

б) F -критерий Фишера;

в) средняя ошибка аппроксимации $\overline{A}$.

8. Значимость уравнения регрессии в целом оценивает:

а) F -критерий Фишера;

б) t -критерий Стьюдента;

в) коэффициент детерминации r_{xy}^2 .

9. Классический метод к оцениванию параметров регрессии основан на:

а) методе наименьших квадратов:

б) методе максимального правдоподобия:

в) шаговом регрессионном анализе.

10. Остаточная сумма квадратов равна нулю:

а) когда правильно подобрана регрессионная модель;

б) когда между признаками существует точная функциональная связь;

в) никогда.

11. Объясненная (факторная) сумма квадратов отклонений в линейной парной модели имеет число степеней свободы, равное:

а) $n-1$;

б) 1;

в) $n-2$.

12. Остаточная сумма квадратов отклонений в линейной парной модели имеет число степеней свободы, равное:

а) $n-1$;

б) 1;

в) $n-2$.

13. Общая сумма квадратов отклонений в линейной парной модели имеет число степеней свободы, равное:

а) $n-1$;

б) 1;

в) $n-2$.

14. Для оценки значимости коэффициентов регрессии рассчитывают:

а) F -критерий Фишера;

б) t -критерий Стьюдента;

в) коэффициент детерминации r_{xy}^2 .

15. Какое уравнение регрессии нельзя свести к линейному виду:

а) $y_x = a + b \cdot \ln x$;

б) $y_x = a \cdot x^b$:

в) $y_x = a + b \cdot x^c$.

16. Какое из уравнений является степенным:

а) $y_x = a + b \cdot \ln x$;

б) $y_x = a \cdot x^b$:

в) $y_x = a + b \cdot x^c$.

17. Параметр b в степенной модели является:

а) коэффициентом детерминации;

б) коэффициентом эластичности;

в) коэффициентом корреляции.

18. Коэффициент корреляции r_{xy} может принимать значения:

а) от –1 до 1;

б) от 0 до 1;

в) любые.

19. Для функции $y = a + \frac{b}{x} + \varepsilon$ ***средний коэффициент эластичности имеет вид:***

а) $\bar{Э} = \frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$;

б) $\bar{Э} = -\frac{b}{a \cdot \bar{x} + b}$;

в) $\bar{Э} = -\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$.

20. Какое из следующих уравнений нелинейно по оцениваемым параметрам:

а) $y = a + b \cdot x + \varepsilon$;

б) $y = a + b \cdot \ln x + \varepsilon$;

в) $y = a \cdot x^b \cdot \varepsilon$.

Множественная регрессия и корреляция

1. Добавление в уравнение множественной регрессии новой объясняющей переменной:

а) уменьшает значение коэффициента детерминации;

б) увеличивает значение коэффициента детерминации;

в) не оказывает никакого влияние на коэффициент детерминации.

2. Скорректированный коэффициент детерминации:

а) меньше обычного коэффициента детерминации;

б) больше обычного коэффициента детерминации;

в) меньше или равен обычному коэффициенту детерминации;

3. С увеличением числа объясняющих переменных скорректированный коэффициент детерминации:

а) увеличивается;

б) уменьшается;

в) не изменяется.

4. Число степеней свободы для остаточной суммы квадратов в линейной модели множественной регрессии равно:

а) $n-1$;

б) m;

в) $n-m-1$.

5. Число степеней свободы для общей суммы квадратов в линейной модели множественной регрессии равно:

а) $n-1$;

б) m;

в) $n-m-1$.

6. Число степеней свободы для факторной суммы квадратов в линейной модели множественной регрессии равно:

а) $n-1$;

б) m;

в) $n-m-1$.

7. Множественный коэффициент корреляции $R_{yx_1x_2} = 0{,}9$ ***. Определите, какой процент дисперсии зависимой переменной*** y ***объясняется влиянием факторов*** x_1 ***и*** x_2 ***:***

а) 90%;

б) 81%;

в) 19%.

8. Для построения модели линейной множественной регрессии вида $y = a + b_1x_1 + b_2x_2$ ***необходимое количество наблюдений должно быть не менее:***

а) 2;

б) 7;

в) 14.

9. Стандартизованные коэффициенты регрессии β_i ***:***

а) позволяют ранжировать факторы по силе их влияния на результат;

б) оценивают статистическую значимость факторов;

в) являются коэффициентами эластичности.

10. Частные коэффициенты корреляции:

а) характеризуют тесноту связи рассматриваемого набора факторов с исследуемым признаком;

б) содержат поправку на число степеней свободы и не допускают преувеличения тесноты связи;

в) характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании других факторов, включенных в уравнение регрессии.

11. Частный F-критерий:

а) оценивает значимость уравнения регрессии в целом;

б) служит мерой для оценки включения фактора в модель;

в) ранжирует факторы по силе их влияния на результат.

12. Несмещенность оценки параметра регрессии, полученной по МНК, означает:

а) что она характеризуется наименьшей дисперсией;

б) что математическое ожидание остатков равно нулю;

в) увеличение ее точности с увеличением объема выборки.

13. Эффективность оценки параметра регрессии, полученной по МНК, означает:

а) что она характеризуется наименьшей дисперсией;

б) что математическое ожидание остатков равно нулю;

в) увеличение ее точности с увеличением объема выборки.

14. Состоятельность оценки параметра регрессии, полученной по МНК, означает:

а) что она характеризуется наименьшей дисперсией;

б) что математическое ожидание остатков равно нулю;

в) увеличение ее точности с увеличением объема выборки.

15. Укажите истинное утверждение:

а) скорректированный и обычный коэффициенты множественной детерминации совпадают только в тех случаях, когда обычный коэффициент множественной детерминации равен нулю;

б) стандартные ошибки коэффициентов регрессии определяются значениями всех параметров регрессии;

в) при наличии гетероскедастичности оценки параметров регрессии становятся смещенными.

16. При наличии гетероскедастичности следует применять:

а) обычный МНК;

б) обобщенный МНК;

в) метод максимального правдоподобия.

17. Фиктивные переменные – это:

а) атрибутивные признаки (например, как профессия, пол, образование), которым придали цифровые метки;

б) экономические переменные, принимающие количественные значения в некотором интервале;

в) значения зависимой переменной за предшествующий период времени.

18. Если качественный фактор имеет три градации, то необходимое число фиктивных переменных:

а) 4;

б) 3;

в) 2.

Системы эконометрических уравнений

1. Наибольшее распространение в эконометрических исследованиях получили:

а) системы независимых уравнений;

б) системы рекурсивных уравнений;

в) системы взаимозависимых уравнений.

2. Эндогенные переменные – это:

а) предопределенные переменные, влияющие на зависимые переменные, но не зависящие от них, обозначаются через x .;

б) зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через y ;

в) значения зависимых переменных за предшествующий период времени.

3. Экзогенные переменные – это:

а) предопределенные переменные, влияющие на зависимые переменные, но не зависящие от них, обозначаются через x ;

б) зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через y ;

в) значения зависимых переменных за предшествующий период времени.

4. Лаговые переменные – это:

а) предопределенные переменные, влияющие на зависимые переменные, но не зависящие от них, обозначаются через x .;

б) зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через y ;

в) значения зависимых переменных за предшествующий период времени.

5. Для определения параметров структурную форму модели необходимо преобразовать в:

а) приведенную форму модели;

б) рекурсивную форму модели;

в) независимую форму модели.

6. Модель идентифицируема, если:

а) число приведенных коэффициентов меньше числа структурных коэффициентов;

б) если число приведенных коэффициентов больше числа структурных коэффициентов;

в) если число параметров структурной модели равно числу параметров приведенной формы модели.

7. Модель неидентифицируема, если:

а) число приведенных коэффициентов меньше числа структурных коэффициентов;

б) если число приведенных коэффициентов больше числа структурных коэффициентов;

в) если число параметров структурной модели равно числу параметров приведенной формы модели.

8. Модель сверхидентифицируема, если:

а) число приведенных коэффициентов меньше числа структурных коэффициентов;

б) если число приведенных коэффициентов больше числа структурных коэффициентов;

в) если число параметров структурной модели равно числу параметров приведенной формы модели.

9. Уравнение идентифицируемо, если:

а) $D + 1 < H$;

б) $D + 1 = H$;

в) $D + 1 > H$.

10. Уравнение неидентифицируемо, если:

а) $D + 1 < H$;

б) $D + 1 = H$;

в) $D + 1 > H$.

11. Уравнение сверхидентифицируемо, если:

а) $D + 1 < H$;

б) $D + 1 = H$;

в) $D + 1 > H$.

12. Для определения параметров точно идентифицируемой модели:

а) применяется двушаговый МНК;

б) применяется косвенный МНК;

б) ни один из существующих методов применить нельзя.

13. Для определения параметров сверхидентифицируемой модели:

а) применяется двушаговый МНК;

б) применяется косвенный МНК;

б) ни один из существующих методов применить нельзя.

14. Для определения параметров неидентифицируемой модели:

а) применяется двушаговый МНК;

б) применяется косвенный МНК;

б) ни один из существующих методов применить нельзя.

Временные ряды

1. Аддитивная модель временного ряда имеет вид:

а) $Y = T \cdot S \cdot E$;

б) $Y = T + S + E$;

в) $Y = T \cdot S + E$.

2. Мультипликативная модель временного ряда имеет вид:

а) $Y = T \cdot S \cdot E$;

б) $Y = T + S + E$;

в) $Y = T \cdot S + E$.

3. Коэффициент автокорреляции:

а) характеризует тесноту линейной связи текущего и предыдущего уровней ряда;

б) характеризует тесноту нелинейной связи текущего и предыдущего уровней ряда;

в) характеризует наличие или отсутствие тенденции.

4. Аддитивная модель временного ряда строится, если:

а) значения сезонной компоненты предполагаются постоянными для различных циклов;

б) амплитуда сезонных колебаний возрастает или уменьшается;

в) отсутствует тенденция.

5. Мультипликативная модель временного ряда строится, если:

а) значения сезонной компоненты предполагаются постоянными для различных циклов;

б) амплитуда сезонных колебаний возрастает или уменьшается;

в) отсутствует тенденция.

6. На основе поквартальных данных построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за первые три квартала равны: 7 – I квартал, 9 – II квартал и –11 – III квартал. Значение сезонной компоненты за IV квартал есть:

а) 5;

б) –4;

в) –5.

7. На основе поквартальных данных построена мультипликативная модель временного ряда. Скорректированные значения сезонной компоненты за первые три квартала равны: 0,8 – I квартал, 1,2 – II квартал и 1,3 – III квартал. Значение сезонной компоненты за IV квартал есть:

а) 0,7;

б) 1,7;

в) 0,9.

8. Критерий Дарбина-Уотсона применяется для:

а) определения автокорреляции в остатках;

б) определения наличия сезонных колебаний;

в) для оценки существенности построенной модели.

ПРИМЕРНЫЕ ВОПРОСЫ ДЛЯ УСТНЫХ ОПРОСОВ

1. Понятие технологии обработки информации. Этапы технологии обработки.
2. Классификация источников информации.
3. Данные как вид информации. Формы данных. Типы данных.
4. Структурированные данные.
5. Структуры данных.
6. Основные компоненты технологии обработки информации. Информационные процессы обработки информации.
7. Основные компоненты информационной технологии обработки данных. Стандартизация объектных баз данных.
8. Системы сбора и обработки информации. Виды систем сбора информации.
9. Общие визуализаторы: графики, диаграммы, гистограммы, статистика.
10. Регрессионный анализ, визуализаторы связей.
11. Многомерное представление данных.
12. Классификация программного обеспечения для создания аналитических решений.
13. Характеристики аналитических платформ.
14. Группировка/перегруппировка данных, объединение данных, преобразование даты
15. и времени.
16. Оценка качества данных. Визуальная оценка качества данных.
17. Набор инструментов предобработки данных в аналитическом приложении. Фильтрация данных.
18. Происхождение пропуска данных, восстановление пропущенных значений.
19. Информационная технология обработки данных: автоматизированная обработка графической информации.
20. Информационная технология анализа данных: системы статистической обработки
21. данных.
22. Системы сбора информации.
23. Системы анализа данных.
24. Системы статистических наблюдений.
25. Основные цели и задачи анализа данных.
26. Феноменологические и концептуальные модели и их характеристики.
27. Средства и методы анализа данных.
28. Формы представления, типы и виды анализируемых данных.
29. Общая схема анализа данных. Требования к алгоритмам анализа данных.
30. Назначение, основные этапы развития и виды систем анализа данных.
31. Применение систем анализа данных для информационно-аналитического обеспечения принятия решений (пример).
32. Основные формы представления и способы объяснения (интерпретации) результатов анализа данных.
33. Основные этапы анализа и интерпретации данных
34. Стратегии ввода, представления и организация исходных данных в информационных системах с анализом данных.
35. Измерительные шкалы, представление переменных, ввод и редактирование данных.

36. Трансформация данных и файлов в методах анализа данных (на выбранном методе, примере решения задачи).
37. Объединение (агрегирование и консолидация) данных в информационно-аналитических системах с применением средств анализа данных.
38. Базовые разведочные методы анализа данных.
39. Способы визуализации информации в системах анализа данных.
40. Графические способы представления результатов анализа данных.
41. Табличные способы представления результатов анализа данных.
42. Подготовка отчетов и экспорт результатов анализа из систем анализа данных.
43. Классические методы многомерного статистического анализа.
44. Роль и место методов классического статистического анализа для решения основных задач анализа данных.
45. Возможности программного обеспечения и практическое применение инструментов классического статистического анализа данных для решения задач анализа данных (примере системы или пакета прикладных программ).
46. Задачи выявления и восстановление зависимостей ванализа данных.
47. Простая регрессионная модель.
48. Оценка соответствия простой линейной регрессии реальным данным.
49. Множественная линейная регрессия.
50. Методы отбора переменных в регрессионные модели.
51. Ограничения применимости регрессионных моделей.
52. Логистическая регрессия. Интерпретация модели логистической регрессии.
53. Множественная логистическая регрессия.
54. Нелинейная регрессия
55. Задачи и методы таксономии (классификации) ванализа данных.
56. Применение методов классификации и кластеризации для сегментации данных.
57. Дискриминационный анализ как способ классификации явлений и объектов.
58. Цели, задачи и основное содержание кластерного анализа. Классификация методов кластеризации.
59. Принципы и общая характеристика методов кластерного анализа.
60. Способы определения меры расстояния между кластерами.
61. Применение кластерного анализа для сокращения количества переменных при моделировании (пример).
62. Использование деревьев решений в задачах классификации (пример).
63. Общая характеристика и методы решения задач прогнозирования
64. Построение прогнозов на основе различных моделей.
65. Оценивание качества прогноза и адекватности модели прогнозирования.
66. Анализ и прогнозирование временных рядов: цели, задачи, методы (временной и частотный подходы к анализу временных рядов).
67. Использование моделей временных рядов для анализа данных и прогнозирования (пример).
68. Способы декомпозиции временных рядов: выявления тренда, сезонной, циклической и случайных составляющих (пример).

Примерный перечень вопросов к зачету

1.Предмет, задачи и содержание дисциплины. (ПК-2).
2.Содержание: Роль анализа данных в теории познания. (ПК-2).
3.Многообразие анализа данных. (ПК-2).
4.Основные этапы развития методов и средств анализа данных. (ПК-2).
5.Структура курса, его место и роль в подготовке бакалавра, связь с другими дисциплинами и ознакомление с компетенциями, которыми должны обладать студенты после изучения данной дисциплины. (ПК-2).
6. Постановка задачи в анализе данных. (ПК-2)
7. Формула для вероятностей сумм независимых, одинаково распределенных, решетчатых случайных величин выраженная через сумму вероятностей полиномиального распределения. (ПК-2)
8. Основы проверки статистических гипотез. (ПК-2)
9. Статистические модели. (ПК-2)
10. Проверка статистических гипотез. (ПК-2)
11. Схема испытаний Бернулли. (ПК-2)
12. Критерий знаков для одной выборки. (ПК-2)
13. Критерий Манна-Уитни. (ПК-2)
14. Критерий Уилкоксона. (ПК-2)
15. Анализ повторных парных наблюдений с помощью знаковых рангов. (ОПК-3, ПК-18)
16. Анализ одной и двух нормальных выборок. (ПК-2)
17. Об исследовании нормальных выборок. (ПК-2)
18. Одна выборка. Две выборки. (ПК-2)
19. Парные данные. (ПК-2)
20. Однофакторный анализ. (ПК-2)
21. Постановка задачиоднофакторного анализа. (ПК-2)
22. Непараметрические критерии проверки однородности. (ПК-2)
23. Критерий Краскела-Уоллиса. (ПК-2)
24. Критерий Джонкхиера. (ПК-2)
25.Дисперсионный анализ. (ПК-2)
26. Оценивание эффектов обработки. (ПК-2)
27. Доверительные интервалы. (ПК-2)
28. Метод Шеффе множественных сравнений. (ПК-2)
29.Двухфакторный анализ. (ПК-2)
30. Связь задач двухфакторного и однофакторного анализа. (ПК-2.)
31. Таблица двухфакторного анализа. (ПК-2.)
32. Непараметрические критерии проверки гипотезы об отсутствии эффектов обработки. (ПК-2.)
33. Аддитивная модель данных двухфакторного эксперимента при независимом действии факторов. (ПК-2.)
34.Линейный и нелинейный регрессионный анализ**.** (ПК-2)
35. Модель линейного регрессионного анализа. (ПК-2)
36. О стратегии, методах и проблемах регрессионного анализа. (ПК-2)

37. Простая линейная регрессия. (ПК-2)
38. О проверке предпосылок в задаче регрессионного анализа. (ПК-2)
39. Непараметрическая линейная регрессия. (ПК-2)
40. Регрессионный анализ в пакетах Excel, STADIA и SPSS.
41. Независимость признаков. (ПК-2)
42. Инструменты и стратегия исследования связи признаков. (ПК-2)
43. Связь признаков в количественных шкалах. (ПК-2)
44. Нормальная корреляция. (ПК-2)
45. Коэффициент корреляции. (ПК-2)

46. Критерий согласия. (ПК-2)
47. Критерии согласия Колмогорова и омега-квадратв в случае простой гипотезы. (ПК-2)
48. Критерий согласия хи-квадрат К.Пирсона для простой гипотезы. (ПК-2)
49. Критерии согласия для сложной гипотезы. (ПК-2)
50.Критерий согласия хи-квадрат Фишера для сложной гипотезы. (ПК-2)

51. Двухфакторный анализ. (ПК-2)
52. Связь задач двухфакторного и однофакторного анализа. (ПК-2)
53. Таблица двухфакторного анализа. (ПК-2)
54. Непараметрические критерии проверки гипотезы об отсутствии эффектов обработки. (ПК-2)
55. Аддитивная модель данных двухфакторного эксперимента при независимом действии факторов. (ПК-2)
56. Временные ряды. (ПК-2)
57. Анализ временных рядов и его разделы. (ПК-2)
58. Цели, этапы и методы анализа временных рядов. (ПК-2)
59. Порядок анализа временных рядов. (ПК-2)
60. Графические методы анализа временных рядов. (ПК-2)
61. Методы исследования структуры стационарного временного ряда. Цели и методы анализа. (ПК-2)
62. Анализ временных рядов на компьютере. (ПК-2)
63. Анализ временных рядов в Excel и SPSS. (ПК-2)
64. Обзор возможностей. (ПК-2)
65. Подбор тренда и прогнозирование. (ПК-2)
66. Устранение сезонной компоненты. (ПК-2)
67. Многомерный анализ и другие статистические методы. (ПК-2)
68. Многомерный статистический анализ. (ПК-2)
69. Факторный анализ. (ПК-2)
70. Дискриминантный анализ. (ПК-2)
71. Кластерный анализ. (ПК-2)
72. Многомерное шкалирование. (ПК-2)
73 Методы контроля качества. (ПК-2)
74. Использование статистических пакетов. (ПК-2)

7. Учебно-методическое и информационное обеспечение дисциплины.

а) основная литература:

1. Методы и модели эконометрики, Ч. 2. Эконометрика пространственных данных
Седова Е. Н., Бантикова О. И., Васянина В. И., Жемчужникова Ю. А., Реннер А. Г.

Оренбургский государственный университет 2015 г. 435 страниц http://www.knigafund.ru/books/181870,

2.Сурина, Е.Е. Методы анализа данных. [Электронный ресурс] — Электрон.дан. — М. : ФЛИНТА, 2015. — 130 с. — Режим доступа: http://e.lanbook.com/book/72701 — Загл. с экрана.

б) дополнительная литература:

1.Новиков А.И. Эконометрика: учебное пособие: [Электронный ресурс].- М.: Дашков и К, 2017.- 224 с.- URL: http://www.knigafund.ru/books/199322

в) программное обеспечение и интернет-ресурсы:

1. Каталог бесплатных курсов Интернет университета Интуит. Курс Data Mining: http://www.intuit.ru/studies/courses/6/6/info
2. Библиотека BaseGroupLabs. Каталог статей об анализе данных:
3. http://basegroup.ru/community/articles
4. Портал знаний StastSoft: http://statistica.ru/local-portals/data-mining/
5. Журнал IEEE Computer Magazinehttp://www.computer.org/.

8. Материально-техническое обеспечение дисциплины.

-Лекционные аудитории с компьютерным и видеопроекционным оборудованием для презентаций с выходом в Интернет, средствами звуковоспроизведения

Компьютерные классы с соответствующим программным обеспечением и видеопроекционным оборудованием для презентаций с выходом в Интернет, средствами звуковоспроизведения

- (ауд.4809, 4810, 4811 и 4805).

9. Методические рекомендации для самостоятельной работы студентов

Запланированные часы учебной программы по самостоятельной работе предусмотрены для приобретения студентами навыков работы со специальной литературой, развития творческого мышления, применения теоретических знаний в конкретных ситуациях, а так же закрепления знаний, полученных в процессе изучения дисциплины на аудиторных занятиях. Это достигается за счет выполнения студентами лабораторных и контрольных работ, подготовки к тестам и итоговым аттестационным мероприятиям. Содержание аттестационных мероприятий приведено в 6 разделе учебно-методического комплекса, контрольных работ – в 4 разделе, материалов для подготовки – 1 разделе.

Табл. – Содержание самостоятельной работы студента

№	Наименование	Содержание
1	Подготовка к устному опросу	По определенной теме готовятся ответы на вопросы
2	Подготовка к тестам	По лекционным материалам курса повторяются и закрепляются вопросы, рассмотренные на аудиторных занятиях, самостоятельно прорабатываются вопросы, не освещенные на аудиторных занятиях, выполняется тест на самопроверку.
3	Подготовка к итоговым аттестационным мероприятиям	По лекционным материалам курса повторяются и закрепляются вопросы, рассмотренные на аудиторных занятиях, самостоятельно прорабатываются вопросы, не освещенные на аудиторных занятиях, выполняется тест на самопроверку.

Приложение

Структура и содержание дисциплины «Анализ данных» по направлению подготовки
09.03.03 «Прикладная информатика» **(бакалавр)**

n/n	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов, и трудоемкость в часах					Виды самостоятельной работы студентов					Формы аттестации	
				Л	П/С	Лаб	СРС	КСР	УО	Т	РГР	Реферат	К/р	Э	З
	Первый семестр														
1.1	Введение Предмет, задачи и содержание дисциплины. Постановка задачи анализа данных.	3	1	4			**2**								
1.2	*Вводное занятие по лабораторному практикуму*	3	1			2	2								
1.3	*Лабораторная работа Предварительный анализ данных с использованием пакета программ Microsoft Office Excel*	3	2			2	2		+						
1.4	Основы проверки статистических гипотез. Анализ одной и двух нормальных выборок.	3	3	4			**2**			+					
1.5	*Лабораторная работа*	3	3			2	2		+						

	Предварительный анализ данных с использованием пакета программ Microsoft Office Excel														
1.6	*Лабораторная работа «Изучение метода наименьших квадратов с использованием пакета программ Microsoft Office Excel»*	3	4			2	2								
1.7	Однофакторный анализ. Дисперсионный анализ. Двухфакторный анализ.	3	5	4			**2**			+					
1.8	*Лабораторная работа Предварительный анализ данных с использованием пакета программ Microsoft Office Excel*	3	5			2	2		+						
1.9	*Лабораторная работа Предварительный анализ данных с использованием пакета программ Microsoft Office Excel*	3	6			2	2		+						
1.10	Линейный и нелинейный регрессионный анализ.	3	7	4			2								
1.11	*Лабораторная работаИзучение методов дискриминантного анализа с использованием пакета программ Microsoft Office Excel*	3	**7**			2	**2**		**+**						
1.12	*Лабораторная работа Изучение методов кластер-анализа с*	3	8			2	2		+						

	использованием пакета программ Microsoft Office Excel														
1.13	О проверке предпосылок в задаче регрессионного анализа.	3	9	4			**2**			+					
1.14	*Лабораторная работа Изучение методов кластер-анализа с использованием пакета программ Microsoft Office Excel*	3	9			2	2		+						
1.15	*Лабораторная работа Изучение методов факторного-анализа с использованием пакета программ Microsoft Office Excel*	3	10			2	2		+						
1.16	Независимость признаков. Критерий согласия. .	3	11	4			**2**								
1.17	*Лабораторная работа Изучение методов факторного-анализа с использованием пакета программ Microsoft Office Excel*	3	11			**2**	**2**		+						
1.18	*Лабораторная работа*	3	12			2	4								

	Классификация данных и изучение методов снижения размерности данных с использованием пакета программ Microsoft Office Excel														
1.19	Временные ряды. Анализ временных рядов на компьютере	3	13	4			4			+					
1.20	*Лабораторная работа Классификация данных и изучение методов снижения размерности данных с использованием пакета программ Microsoft Office Excel*	3	13			2	4		+						
1.21	*Лабораторная работа Классификация данных и изучение методов снижения размерности данных с использованием пакета программ Microsoft Office Excel*	3	14			2	4		+						
1.22	Многомерный анализ и другие статистические методы.	3	15	4			**4**								
1.23	*Лабораторная работа Построение уравнения множественной регрессии для моделирования и прогнозирования экономических задач сиспользованием пакета программ Microsoft Office Excel*	3	15			2	**4**		+						
1.24	*Лабораторная работа Изучение методов прогнозирования временных рядов*	3	16			2	**4**		+						

	с использованием пакета программ Microsoft Office Excel														
1.25	Кластерный анализ.	3	17	4			**4**								
1.26	*Лабораторная работа Итоговое занятие*	3	17			2	**4**			+					
1.26	*Лабораторная работа Итоговое занятие*	3	18			2	**4**								
1.28	***Форма аттестации***	3	18												**З**
1.29	Всего часов по дисциплине в третьем семестре	3		36		36	72								