**Федеральное государственное бюджетное образовательное учреждение высшего образования**
**«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Факультет информационных технологий и анализа данных
Кафедра эконометрики и математической экономики

УТВЕРЖДЕНА

решением кафедры системного анализа и информатики

Протокол от «03» сентября 2019 г.

№1

# РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

## Б1.Б.20 ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ

*(индекс и наименование дисциплины (модуля), в соответствии с учебным планом)*

по направлению подготовки
38.03.05 Бизнес-информатика
*(код и наименование направления подготовки)*

Цифровой бизнес
*направленность (профиль)*

Бакалавр
*квалификация выпускника*

Очная
*форма обучения*

*набор 2020 г.*

Москва, 2019 г.

**Автор–составитель:**

д.т.н. Вепрев С. Б.

*(ученое звание, ученая степень, должность)* *(Ф.И.О.)*

Заведующий кафедрой эконометрики и математической экономики

*(наименование кафедры)*

к.ф.-м.н., доцент Носко В. П.

*(ученая степень и(или) ученое звание )* *(Ф.И.О.)*

## СОДЕРЖАНИЕ


1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы…… 4
2. Объем и место дисциплины (модуля) в структуре образовательной программы ……... 6
3. Содержание и структура дисциплины (модуля)……... 6
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине (модулю)……..……... 10
5. Методические указания для обучающихся по освоению дисциплины (модуля)……... 15
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)……...……. 16
   6.1. Основная литература……..……...…… 16
   6.2. Дополнительная литература……...……...…...…… 16
   6.3. Учебно-методическое обеспечение самостоятельной работы…… 16
   6.4. Нормативные правовые документы…….……..…….. 17
   6.5. Интернет-ресурсы…….……..……... 17
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы …...……...…..…….. 17

## 1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения программы

Дисциплина «Введение в анализ данных» обеспечивает овладение следующими компетенциями:

| Код компетенции | Наименование компетенции | Код этапа освоения компетенции | Наименование этапа освоения компетенции |
|---|---|---|---|
| ОПК-ОС-4 | способность управлять ресурсами и сервисами информационных технологий (ИТ) и информационной средой (ИС) | ОПК-4.4 | Способен управлять информационной средой |

**Менеджер продуктов в области информационных технологий.** Приказ Минтруда России от 20.11.2014 N 915н (Зарегистрировано в Минюсте России 18.12.2014 N 35273)

В результате освоения дисциплины у студентов должны быть сформированы:

| ОТФ/ТФ<br><br>(при наличии профстандарта) | Код этапа освоения компетенции | Результаты обучения |
|---|---|---|
| С - Управление серией продуктов и группой их менеджеров.<br><br>Заказ и анализ результатов технологических исследований в интересах серии продуктов (С/01.6)<br><br>Разработка бизнес-планов, ценовой политики и стратегии развития серии продуктов (С/02.6)<br><br>Заключение партнерских соглашений и развитие отношений с партнерами (С/01.3)<br><br>Управление бюджетом серии продуктов (С/01.4)<br><br>Управление группой менеджеров продуктов (С/01.5)<br><br>Продвижение продуктов(С/016)<br><br>Заказ и контроль выполнения программы проектов по созданию, развитию, выводу на рынок и продажам продуктов серии (С/01.7)<br><br>Управление патентами на технологии, создаваемые в рамках продуктов (С/01.8)<br><br>Разработка предложений по приобретению и продаже технологических, продуктовых и прочих интеллектуальных активов и организаций (С/01.9) | ОПК-ОС-4.4 | на уровне знаний: знать:<br>- основы теории параметрической и непараметрической проверки гипотез;<br>- основы теории дисперсионного и факторного анализа, метода главных компонент, анализа временных рядов;<br>- основы теории корреляционного и линейного регрессионного анализа;<br>- методологию анализа регрессионных моделей и регрессионных остатков;<br>- структуру систем интеллектуального и топологического анализа данных. |
| | | на уровне умений: уметь<br>применить основные методы статистического и глубинного анализа в решении экономических задач. |
| | | на уровне навыков: владеть:<br>- 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;<br>- методикой построения, анализа и применения регрессионных моделей для оценки состояния и прогноза развития экономических явлений и процессов;<br>- навыками применения методов интеллектуального и топологического анализа данных для решения экономических задач. |

## 2. Объем и место дисциплины (модуля) в структуре ОП ВО

Дисциплина «Введение в анализ данных» имеет индекс Б1.Б.20, объем 72 .академических часа, 2 з.е., изучается на 2 курсе в 4 семестре в соответствии с учебным планом. Количество академических часов, выделенных на контактную работу с преподавателем - 20 часов, на самостоятельную работу обучающихся – 32 часа; форма промежуточной аттестации – зачет.

Дисциплина «Введение в анализ данных» включена в базовую часть математического и естественнонаучного цикла.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами: «Математический анализ», «Линейная алгебра и аналитическая геометрия», «Теоретические основы информатики», «Программирование», «Операционные среды, системы и оболочки».

Наименования последующих учебных дисциплин: «Архитектура предприятия», «Электронный бизнес», «Управление разработкой ИС».

## 3. Содержание и структура дисциплины (модуля)

| № п/п | Наименование тем (разделов) | Объем дисциплины, час. | | | | | | Форма текущего контроля успеваемости[4], промежуточной аттестации |
|---|---|---|---|---|---|---|---|---|
| | | Всего | Контактная работа обучающихся с преподавателем по видам учебных занятий | | | | СР | |
| | | | Л | ЛР | ПЗ | КСР | | |
| | | | | | | | | |
| | Модуль 1. Статистический анализ данных | | | | | | | |
| Тема 1 | Параметрическая и непараметрическая проверка гипотез Тема 8 | 8 | 2 | | 2 | | 4 | ДЗ |
| Тема 2 | Дисперсионный анализ (ANOVA). | 8 | 2 | | 2 | | 4 | ДЗ |
| Тема 3 | Корреляционный и линейный регрессионный анализ | 8 | 2 | | 2 | | 4 | ДЗ |
| Тема 4 | Анализ регрессионных моделей и регрессионных остатков. | 8 | 2 | | 2 | | 4 | ДЗ |
| Тема 5 | Анализ временных рядов | 8 | 2 | | 2 | | 4 | ДЗ |
| Тема 6 | Метод главных компонент. Факторный анализ. | 7 | 2 | | 2 | | 3 | ДЗ |
| | Модуль 2. Глубинный анализ данных (Data mining) | | | | | | | |
| Тема 7 | Основные системы интеллектуального анализа данных. | 7 | 2 | | 2 | | 3 | |
| Тема 8 | Методы интеллектуального анализа данных. | 7 | 2 | | 2 | | 3 | |

| № п/п | Наименование тем (разделов) | Объем дисциплины, час. | | | | | | Форма текущего контроля успеваемости[4], промежуточной аттестации |
|---|---|---|---|---|---|---|---|---|
| | | Всего | Контактная работа обучающихся с преподавателем по видам учебных занятий | | | | СР | |
| | | | Л | ЛР | ПЗ | КСР | | |
| | | | | | | | | |
| Тема 9 | Топологический анализ данных. | 11 | 4 | | 4 | | 3 | |
| | Промежуточная аттестация | | | | | | | Зачет |
| **Всего по дисциплине** | | **72** | **20** | | **20** | | **32** | |

** – при применении электронного обучения, дистанционных образовательных технологий в соответствии с учебным планом;*

*** – формы текущего контроля успеваемости: опрос (О), тестирование (Т), контрольная работа (КР), коллоквиум (К), эссе (Э), реферат (Р), диспут (Д) и др.*

## Содержание дисциплины (модуля)

### 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 Бартлета. 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 связных 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. Индентификация. 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 компонет. 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.

3.3.2. Темы практических занятий
1. Параметрическая и непараметрическая проверка гипотез. (2 час.)
2. Дисперсионный анализ. (2 час)
3. Корреляционный и регрессионный анализ. (2 час.)
4. Анализ регрессионных моделей и регрессионных остатков. (2 час.)
5. Анализ временных рядов. (2 час.)
6. Метод главных компонент. Факторный анализ. (2 час.)
7. Межотраслевой стандартный процесс для интеллектуального анализа данных. (2 час.)
8. Методы интеллектуального анализа данных, основанный на деревьях решений. (2 час.)
9. Топологический анализ данных. (2 час.)

**4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине (модулю)**

4.1. Формы и методы текущего контроля успеваемости и промежуточной аттестации.

В ходе реализации дисциплины используются следующие методы текущего контроля успеваемости обучающихся:
- при проведении занятий лекционного типа: беседа (диалог) с обучающимися,
- при проведении занятий семинарского типа: домашние работы по темам практических заданий.

4.2. Материалы текущего контроля успеваемости.

## Темы практических занятий

1. Параметрическая и непараметрическая проверка гипотез. (2 час.)
2. Дисперсионный анализ. (2 час)
3. Корреляционный и регрессионный анализ. (2 час.)
4. Анализ регрессионных моделей и регрессионных остатков. (2 час.)
5. Анализ временных рядов. (2 час.)
6. Метод главных компонент. Факторный анализ. (2 час.)
7. Межотраслевой стандартный процесс для интеллектуального анализа данных. (2 час.)
8. Методы интеллектуального анализа данных, основанный на деревьях решений. (2 час.)
9. Топологический анализ данных. (2 час.)

При проведении текущей аттестации студенты выполняют практические задания (домашние работы).

Примеры практических заданий.

Задача 1. Имеются поквартальные очищенные от сезонности данные о ВВП США в миллиардах долларов.

Моделируя данные временным рядом, провести его анализ. Определить ряд первых разностей, ряд разностей логарифмов ряда, автокорреляции и частные автокорреляции ряда прироста. Построить модели AR(1), MA(2), ARMA(1,2). Провести диагностику и сравнение моделей.

Задача 2. Имеются выборочные данные по доходу и потреблению.

Провести корреляционный анализ данных. Определить корреляцию Пирсона, значимость коэффициента корреляции, частную корреляцию, значимость коэффициента частной корреляции, множественную корреляцию, значимость коэффициента множественной корреляции, ранговую корреляцию, коэффициент корреляции Спирмена, коэффициент корреляции Кенделла.

Провести линейный регрессионный анализ данных. Получить МНК-решение и его запись через сингулярное разложение. Определить остаточную сумму квадратов, доверительные интервалы для дисперсии шума, коэффициентов регрессии, прогнозного значения отклика.

Задача 3. По данным задачи 2:

Провести анализ структуры линейной регрессионной модели. Оценить значимость коэффициентов линейной регрессии, критерий Фишера, выборочный коэффициент детерминации, дисперсию остатков.

Провести анализ регрессионных остатков. Провести непараметрические тесты: критерий Уилкоксона-Манна-Уитни, критерий Зигеля-Тьюки, критерий знаков, критерий серий, критерий экстремумов. Проверить нормальность остатков (критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий омега-квадрат фон Мизеса, критерий хи-квадрат Пирсона, критерии асимметрии и эксцесса). Провести тест на корреляцию остатков по статистике Дарбина-Уотсона.

В течение семестра выполняются домашние работы по темам практических занятий. Результаты выполнения этих работ являются основанием для выставления оценок текущего контроля. Выполнение всех работ является обязательным для всех студентов. Учитываются также результаты работы на практических занятиях.

## Шкала оценивания текущего контроля

| **10-бальная шкала** | **Традиционная шкала** | **«Зачтено»/ «Не зачтено»** | **Определение** |
|---|---|---|---|

| 10-бальная шкала | Традиционная шкала | «Зачтено»/ «Не зачтено» | Определение |
|---|---|---|---|
| 10 | Отлично | Зачтено | Полные, глубокие и систематические знания, знакомство с дополнительной литературой, полный и правильный ответ, творческий подход в понимании и изложении учебного материала, полное выполнение мероприятий текущего контроля. |
| 9 | Отлично | Зачтено | Полные, глубокие и систематические знания, полный и правильный ответ, полное выполнение мероприятий текущего контроля. |
| 8 | Отлично | Зачтено | Полные и систематические знания, отсутствие существенных неточностей в ответе, полное выполнение мероприятий текущего контроля. |
| 7 | Хорошо | Зачтено | Достаточно полные и систематические знания, отсутствие существенных неточностей в ответе, имеются погрешности при выполнении мероприятий текущего контроля. |
| 6 | Хорошо | Зачтено | Достаточно полные и систематические знания, отсутствие существенных неточностей в ответе, имеются погрешности при выполнении мероприятий текущего контроля. |
| 5 | Удовлетворительно | Зачтено | Знание основного учебного материала в объеме, необходимом для дальнейшей учебы и работы, имеются погрешности при выполнении мероприятий текущего контроля и при ответе. |
| 4 | Удовлетворительно | Зачтено | Знание основного учебного материала в минимальном объеме, необходимом для дальнейшей учебы и работы, имеются погрешности при выполнении мероприятий промежуточного контроля и при ответе. |
| 3 | Неудовлетворительно | Не зачтено | Имеются существенные погрешности при выполнении мероприятий текущего контроля, допущены существенные ошибки при ответе, необходима некоторая дополнительная работа. |
| 2 | Неудовлетворительно | Не зачтено | Имеются пробелы в знаниях по значительной части учебного материала, допущены существенные ошибки при ответе, необходима значительная дополнительная учебная работа. |
| 1 | Неудовлетворительно | Не зачтено | Не выполнены предусмотренные программой задания, не отработаны практические или лабораторные занятия, необходимы дополнительные занятия по соответствующей дисциплине. |
| 0 | Неудовлетворительно | Не зачтено | Нарушение академических норм (плагиат и т.п.) |

**4.3. Формы, методы (средства) промежуточной аттестации.**

4.3.1. Для контроля усвоения данной дисциплины учебным планом предусмотрены зачет (в соответствии с учебным планом), который проводится в устной форме. Задания

содержат вопросы, в которых необходимо использовать теоретические знания и практическое задание, демонстрирующие способность проведения анализа инноваций в экономике, управлении и информационно – коммуникативных технологиях. На зачет выносятся основные вопросы, рассматриваемые в рамках всего курса. Основой для определения оценки служит объем и уровень усвоения студентами материала, предусмотренного программой данного курса и подведения итогов по результатам выполнения заданий текущего контроля успеваемости.

Вопросы для подготовки.

1. Параметрическая проверка гипотез: гипотезы о положении, гипотезы о рассеивании. Систематизация критериев.

2. Параметрическая проверка гипотез: критерий Стьюдента для одной и двух выборок, связанные выборки, гипотеза сдвига, метод множественных сравнений Шеффе, метод LSD.

3. Параметрическая проверка гипотез: критерий Фишера, критерий Кокрена, критерий Бартлета. Критерии нормальности: Критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий омега-квадрат фон Мизеса, критерий хи-квадрат (Пирсона).

4. Параметрическая проверка гипотез: упрощённые проверки нормальности по асимметрии и эксцессу.

5. Непараметрическая проверка гипотез: критерий Уилкоксона-Манна-Уитни, гипотезы о положении, гипотезы о рассеивании.

6. Непараметрическая проверка гипотез: элементы теории измерений: номинальные, порядковые и количественные переменные; инварианты.

7. Непараметрическая проверка гипотез: среднее по Коши, по Колмогорову, мода, медиана. Среднее в порядковой шкале.

8. Непараметрическая проверка гипотез: критерий Уилкоксона-Манна-Уитни, критерий знаков, критерий Уилкоксона двухвыборочный, критерий Уилкоксона для связных выборок, критерий Краскела-Уоллиса, критерий Зигеля-Тьюки, медианный критерий: одновыборочный и двухвыборочный.

9. Дисперсионный анализ. Однофакторная параметрическая модель. метод Шеффе.

10. Дисперсионный анализ. Однофакторная непараметрическая модель: критерий Краскела-Уоллиса, критерий Джонкхиера.

11. Дисперсионный анализ. Двухфакторная непараметрическая модель: критерий Фридмана, критерий Пейджа. Двухфакторный нормальный анализ.

12. Корреляционный анализ. Корреляция Пирсона, значимость коэффициента корреляции (критерий Стьюдента). Частная корреляция, значимость коэффициента частной корреляции.

13. Корреляционный анализ. Множественная корреляция, значимость коэффициента множественной корреляции.

14. Корреляционный анализ. Ранговая корреляция, коэффициент корреляции Спирмена, коэффициент корреляции Кенделла. Конкордация Кенделла.

15. Линейный регрессионный анализ. Метод наименьших квадратов. МНК-решение и его запись через сингулярное разложение. Остаточная сумма квадратов (RSS).

16. Основные предположения многомерной линейной регрессии. Статистические свойства МНК-оценок без предположения нормальности. Статистические свойства МНК-оценок при предположении нормальности.

17. Линейный регрессионный анализ. Доверительные интервалы для дисперсии шума, коэффициентов регрессии, прогнозного значения отклика.

18. Анализ структуры линейной регрессионной модели. Значимость коэффициентов линейной регрессии (проверка равенства коэффициентов нулю), вложенные модели линейной регрессии, критерий Фишера. Шаговая регрессия.

19. Линейный регрессионный анализ. Проверка адекватности модели. Выборочный коэффициент детерминации. Дисперсия остатков.

20. Линейный регрессионный анализ. Проблема мультиколлинеарности. Методы понижения размерности: ридж-регрессия, лассо Тибширани, параметр селективности.

21. Анализ регрессионных остатков. Визуальный анализ.

22. Анализ регрессионных остатков. Непараметрические тесты: критерий Уилкоксона-Манна-Уитни, критерий Зигеля-Тьюки, критерий знаков, критерий серий, критерий экстремумов.

23. Анализ регрессионных остатков. Проверка нормальности остатков: Критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий омега-квадрат фон Мизеса, критерий хи-квадрат Пирсона, критерии асимметрии и эксцесса.

24. Анализ временных рядов. Идентификация модели временных рядов. Систематическая составляющая и случайный шум.

25. Анализ временных рядов. Два общих типа компонент временных рядов. Подгонка функции. Анализ тренда. Анализ сезонности. Автокорреляционная диаграмма. Исследование коррелограмм. Удаление периодической зависимости.

26. Анализ временных рядов. Модель Бокса-Дженкинса и автокорреляции. Процесс авторегрессии и скользящего среднего. Требование стационарности. Обратимость.

27. Анализ временных рядов. Оценивание и прогноз. Идентификация: число оцениваемых параметров, сезонные модели.

28. Анализ временных рядов. Оценивание параметров: сравнение методов, стандартные ошибки оценок, штраф. Оценивание модели: оценки параметров, сравнение прогноза, анализ остатков, ограничения.

29. 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.

30. Метод главных компонент. 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.

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

32. Межотраслевой стандартный процесс для интеллектуального анализа данных (CRISP). Ступени процесса CRISP: понимание бизнеса, понимание данных, подготовка данных, моделирование, оценка, развертывание.

33. Процесс SEMMA. Ступени процесса SEMMA: исследование данных, изменение данных, моделирование, оценка полезности и надежности.

34. Методы интеллектуального анализа данных. Статистический анализ. Методы, основанные на деревьях решений. Генетические алгоритмы. Нейронные сети. Нечеткая логика. Временные ряды.

35. Методы добычи данных: ассоциация, классификация, кластеризация, прогноз временных рядов.

36. Топологический анализ данных. Основные задачи. Основные методы топологического анализа данных. Симплициальные комплексы. Параметр близости.

37. Топологический анализ данных. Алгебраические топологии. Основы теории устойчивых гомологий. Числа Бетти.

38. Топологический анализ данных. Облако точек в Евклидовом пространстве. Преобразование облака точек в метрическом пространстве. Устойчивая гомология.

4.4. Методические материалы по проведению промежуточной аттестации

Зачет проводится в соответствии с графиком учебного процесса учетом проведения мониторинга уровня освоения компетенции по результатам выполнения самостоятельных заданий. Оценивание осуществляется в соответствии со шкалой оценивания. Студентам, не выполнившим домашние задания и (или) контрольные задания по уважительным причинам, предоставляется возможность их выполнения и сдачи.

**Критерии оценивания уровня формирования компетенций**

| **Этап освоения компетенции** | **Показатель оценивания**<br>*Что делает обучающийся (какие действия способен выполнить), подтверждая этап освоения компетенции* | **Критерий оценивания**<br>*Как (с каким качеством) выполняется действие. Соответствует оценке «отлично» в шкале оценивания в РПД.* | **Оценка (баллы)** |
|---|---|---|---|
| 4 этап (код этапа: ОПК ОС-4.4) Способен управлять информационной средой | Способен использовать навыки применения методов интеллектуального и топологического анализа данных и программирования для решения экономических задач и управления информационной средой. | Определено оптимальное количество необходимых для разработки информационной среды ресурсов.<br>Определены все возможные ограничения, существующие в рамках реализации информационной среды. | Промежуточная аттестация тест |

## 5. Методические указания для обучающихся по освоению дисциплины

Дисциплина включает 9 тем. На освоение каждой темы отводится от 9 до 18 часов аудиторной работы и от 1 до 6 самостоятельной. Самостоятельная работа студента по дисциплине «Введение в анализ данных» предусмотрена учебным планом и составляет 36 часов по очной форме обучения. Студенты выполняют следующие виды заданий: подготовка домашнего задания (в том числе индивидуального); подготовка доклада (сообщения).

Подготовка домашнего задания является основной формой самостоятельной работы студентов. Студенты получают задание на каждом практическом занятии и с той же регулярностью отчитываются о его выполнении. Бóльшая часть заданий имеет письменный характер. В связи с этим студенту рекомендуется завести отдельную тетрадь для самостоятельной работы, которая предъявляется преподавателю по его требованию. Индивидуальное задание назначается студенту в целях совершенствования актуальных для него навыков или для наилучшего усвоения отдельных тем дисциплины.

Подготовка докладов и сообщений может широко использоваться студентами при подготовке к практическим занятиям. Данный вид самостоятельной работы рассматривается как вспомогательный. В то же время темы выступлений на занятиях могут быть развернуты в темы студенческих научных исследований и стать основой для участия в студенческих научно-практических конференциях, олимпиадах, конкурсах студенческих научных работ.

**1. Советы по планированию и организации времени, необходимого для изучения дисциплины.** Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины: Изучение конспекта лекции в тот же день после лекции – 10-15 минут. Повторение лекции за день перед следующей лекцией – 10-15

минут. Изучение теоретического материала по учебнику и конспекту – 1 час в неделю. Подготовка к практическому занятию – 1 час. Тогда общие затраты времени на освоение курса студентами составят около 2,5 часа в неделю.

**2. Описание последовательности действий студента («сценарий изучения дисциплины»).** Следует внимательно слушать и конспектировать материал, излагаемый на аудиторных занятиях. Для его понимания и качественного усвоения рекомендуется следующая последовательность действий: 1. После окончания учебных занятий для закрепления материала просмотреть и обдумать текст лекции, прослушанной сегодня, разобрать рассмотренные примеры (10-15 минут). 2. При подготовке к лекции следующего дня повторить текст предыдущей лекции, подумать о том, какая может быть следующая тема (10-15 минут). 3. В течение недели выбрать время для работы с литературой в библиотеке и для решения задач (по 1 часу). 4. При подготовке к практическим занятиям повторить основные понятия и разобрать примеры на компьютере. Решая упражнение или задачу, – предварительно понять, какой теоретический материал нужно использовать. Наметить план решения, попробовать на его основе решить 1-2 аналогичные задачи.

**4. Рекомендации по работе с литературой.** Рекомендуется, кроме «заучивания» материала, добиться понимания изучаемой темы дисциплины. С этой целью после прочтения очередной главы желательно выполнить несколько простых упражнений на соответствующую тему. Кроме того, очень полезно мысленно задать себе и попробовать ответить на следующие вопросы: о чем эта глава, какие новые понятия в ней введены, каков их смысл.

**5. Советы по подготовке к зачету.** Дополнительно к изучению конспектов лекций необходимо пользоваться учебниками по дисциплине. Вместо «заучивания» материала важно добиться понимания изучаемых тем дисциплины. При подготовке к нужно освоить теорию: разобрать определения всех понятий структурного программирования, рассмотреть примеры и самостоятельно решить несколько типовых задач из каждой темы. При решении задач всегда необходимо комментировать свои действия и не забывать о содержательной интерпретации.

**6. Указания по организации работы с контрольно-измерительными материалами.** При выполнении домашних заданий и подготовке к контрольной работе необходимо сначала прочитать теорию и изучить примеры по каждой теме. Решая конкретную задачу, предварительно следует понять, что требуется от Вас в данном случае, какой теоретический материал нужно использовать, наметить общую схему решения. Если задача решается «по образцу» рассмотренного на практическом занятии или в методическом пособии примера, то желательно после этого обдумать процесс решения и попробовать решить аналогичную задачу самостоятельно.

## 6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1 Основная литература.

1. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2012
2. Шведов, А.С. Теория вероятностей и математическая статистика: промежуточный уровень [Электронный ресурс] : учебное пособие / А.С. Шведов. — Электрон. дан. — Москва : Издательский дом Высшей школы экономики, 2016. — 280 с. — Режим доступа: https://e.lanbook.com/book/100140.
3. Березинец, И.В. Практикум по теории вероятностей и математической статистике [Электронный ресурс] : учебное пособие / И.В. Березинец. — Электрон. дан. — Санкт-Петербург : СПбГУ, 2013. — 163 с. — Режим доступа: https://e.lanbook.com/book/47493.

6.2. Дополнительная литература.

1. Кулаичев А.П. Методы и средства комплексного анализа данных. - М.: Форум – Инфра-М, 2006.

6.3. Учебно-методическое обеспечение самостоятельной работы.

1. Ласковец С.В. Методология научного творчества [Электронный ресурс]: Учебное пособие. – Москва : Евразийский открытый институт, 2010. – 32 с. – URL: http://www.biblioclub.ru/90384_Metodologiya_nauchnogo_tvorchestva_Uchebnoe_posobie.html
2. Радаев В.В. Как организовать и представить исследовательский проект. 75 простых правил. – Москва : ГУ-ВШЭ : Инфра-М, 2001. – 203 с.
3. Панкратов В.Н. Искусство управлять собой: Практическое руководство. – Москва : Издательство института психотерапии, 2001. – 256 с.
4. ПОЛОЖЕНИЕ об организации самостоятельной работы студентов федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (в ред. приказа РАНХиГС от 11.05.2016 г. № 01-2211)
5. ПОРЯДОК освоения в РАНХиГС факультативных и элективных дисциплин (модулей) образовательных программ высшего образования - программ бакалавриата, программ специалитета, программ магистратуры. Приложение к приказу от 26 июля 2016 г. № 02-417.

6.4. Нормативные правовые документы.

1. Федеральный закон от 29.12.2012 N 273-ФЗ (ред. от 13.07.2015) "Об образовании в Российской Федерации" (с изм. и доп.).
2. Федеральный государственный образовательный стандарт высшего профессионального образования по направлению подготовки 38.03.05 Бизнес-информатика (квалификация (степень) «бакалавр»), утвержденный приказом Министерства образования и науки Российской Федерации от «11» августа 2016 г. № 1002.
3. Нормативно-методические документы Минобрнауки России.
4. Устав Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российская академия народного хозяйства и государственной службы при Президенте РФ».

6.5. Интернет-ресурсы, справочные системы.

1. AnnualReviews [Электронный ресурс]. – URL: http://arjournals.annualreviews.org/.
2. EbscoHost [Электронный ресурс]. – URL: http://www.ebscohost.com/.
3. e-Library.ru [Электронный ресурс]: Научная электронная библиотека. – URL: http://elibrary.ru/.
4. Научная электронная библиотека «КиберЛенинка» [Электронный ресурс]. – URL: http://cyberleninka.ru//
5. ProQuest [Электронный ресурс]. – URL: http://search.proquest.com/index.
6. http://www.hr-training.net
7. http://www.lseptember.ru
8. http://www.tolerance.ru
9. Словари и энциклопедии на Академике [Электронный ресурс] // Академик. – URL: http://dic.academic.ru.
10. Университетская библиотека online [Электронный ресурс]. – URL: http://biblioclub.ru/.
11. Консультант Плюс http://www.consultant.ru/law/hotdocs/t7/

## 7. Материально-техническое и программное обеспечение дисциплины (модуля)

7.1. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Для реализации данной дисциплины (модуля), необходимы специализированные компьютерные аудитории для проведения всех видов контактной и самостоятельной работы. Аудитории должны быть оборудованы компьютерами в соответствии с минимальными техническими требованиями. Количество рабочих мест обучаемых должно быть не менее количества студентов в учебной группе. При использовании виртуальных машин должен быть единый защищенный сетевой ресурс, на котором обучаемые смогут сохранять результаты своей работы. В обязательном порядке в аудитории должна присутствовать проекционная аппаратура, обеспечивающая как показ презентаций по теме занятий, так и демонстрацию работы преподавателя в среде разработки в реальном режиме времени. Оборудование класса должно обеспечивать выход преподавателя и обучаемых в глобальную сеть Интернет для выполнения учебных занятий. К обязательному программному обеспечению для поддержки образовательного процесса необходимо отнести: MS Excel

7.2. Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)

При осуществлении образовательного процесса применяются информационные технологии, необходимые для подготовки презентационных материалов и материалов к занятиям (компьютеры с программным обеспечением для создания и показа презентаций, с доступом в сеть «Интернет», поисковые системы и справочные, профессиональные ресурсы в сети «Интернет»).

Содержание дисциплины размещено на сайте информационно-коммуникационной сети Интернет: ranepa.ru./.

7.3. Необходимое программное обеспечение

Для подготовки презентаций и их демонстрации необходима программа Impress из свободного пакета офисных приложений OpenOffice (или иной аналог с коммерческой или свободной лицензией).

Для контактной и самостоятельной работы используются мультимедийные комплексы, электронные учебники и учебные пособия, адаптированные к ограничениям здоровья обучающихся. Информационные средства обучения: электронные учебники, учебные фильмы по тематике дисциплины, презентации, интерактивные учебные и наглядные пособия, технические средства предъявления информации (многофункциональный мультимедийный комплекс) и контроля знаний (тестовые системы).