МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИРОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В.Г.ШУХОВА»

(БГТУ им. В.Г. Шухова)

УТВЕРЖДАЮ
Директор института
экономики и менеджмента
_____ Дорошенко Ю.А.
«26» сентября 2016 г.

РАБОЧАЯ ПРОГРАММА

дисциплины
Анализ данных

направление подготовки
38.03.05 «Бизнес-информатика»

профиль подготовки
Технологическое предпринимательство

Степень
Бакалавр

Форма обучения
Очная

Срок обучения
4 года

Институт: экономики и менеджмента

Кафедра: экономики и организации производства

Белгород – 2016

Рабочая программа составлена на основании требований:

- Федерального государственного образовательного стандарта высшего профессионального образования по направлению подготовки бакалавриата 38.03.05 «Бизнес-информатика» утвержденного Приказом Министерства образования и науки РФ от 11августа 2016г №1002
- плана учебного процесса БГТУ им. В.Г. Шухова по направлению подготовки38.03.05 «Бизнес-информатика», введенного в 2016г.

Составитель (составители): к.э.н., профессор О.В.Доможирова

Рабочая программа согласована с выпускающей кафедрой Экономики и организации производства

Заведующий кафедрой: д.э.н., профессор Ю.И. Селиверстов

« 31 » августа 2016 г.

Рабочая программа обсуждена на заседании кафедры

« 31 » августа 2016 г., протокол № 1

Заведующий кафедрой: д.э.н., профессор Ю.И. Селиверстов

Рабочая программа одобрена методической комиссией института

« 23 » 09 2016 г., протокол № 1

Председатель к.э.н., профессор В.В.Выборнова

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Формируемые компетенции			Требования к результатам обучения
№	Код компетенции	Компетенция	
Общепрофессиональные			
1	ОПК-1	Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.	В результате освоения дисциплины обучающийся должен **Знать:** Теоретические основы, место и роль изучаемой дисциплины «Анализ данных» среди других наук; содержание ключевых понятий в области статистической обработки и анализа данных («программные средства», «статистический пакет», «статистическая обработка данных», «математическая статистика», «модульный принцип», «визуализация данных»). **Уметь:** Пользоваться методическим инструментарием анализа дан- ных;применятьматематическиеметодыиинструментальн ыесредства дляисследованияобъектовпрофессиональнойдеятельности. Строитьматематическиемоделиобъектовпрофессиональнойдеятельности;работать в современных пакетах программ статистической обработки информации; представлять статистическую информацию в виде графиков, диаграмм, гистограмм. **Владеть:** Современными методами статистической обработки данных рассчитывать доверительный интервал, определять объем выборки, проверять статистические гипотезы, определять коэффициенты детерминации, строить регрессионные зависимости, рассчитывать коэффициенты корреляции, измерять силу связи между явлениями, строить таблицы сопряженности, проводить кластер-анализ и уметь классифицировать объекты, проводить факторный анализ. Навыками применения математических инструментальных средствадляобработки, анализаисистематизациинформации; методикой построения, анализа и применения анализа данных для оценки состояния и прогноза развития экономических явлений и процессов

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Содержание дисциплины основывается и является логическим продолжением следующих дисциплин:

№	Наименование дисциплины (модуля)
1.	Теоретические основы информатики
2.	Программирование
3.	Вычислительные системы, сети и телекоммуникации
4.	Экономика фирмы

Содержание дисциплины служит основой для изучения следующих дисциплин:

№	Наименование дисциплины (модуля)
1.	Исследование операций
2.	Базы данных
3.	Объектно-ориентированный анализ и программирование

3. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет __4__ зач. единиц, _144_часов.

Вид учебной работы	Всего часов	Семестр № 3
Общая трудоемкость дисциплины, час	144	144
Контактная работа (аудиторные занятия), в т.ч.:	**51**	**51**
лекции	17	17
лабораторные	34	34
практические		
Самостоятельная работа студентов, в том числе:	**93**	**93**
Курсовой проект		
Курсовая работа		
Расчетно-графическое задания	18	18
Индивидуальное домашнее задание		
Другие виды самостоятельной работы	75	75
Форма промежуточная аттестация (зачет, экзамен)	диф.зачет	диф.зачет

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Наименование тем, их содержание и объем

Курс 2 Семестр № 3

№ п/п	Наименование раздела (краткое содержание)	Объем на тематический раздел по видам учебной нагрузки, час			
		Лекции	Практические занятия	Лабораторные занятия	Самостоятельная работа
1. Статистическая обработка и анализ данных: предмет изучения					
	“Математическая статистика” и “анализ данных”: способы получения данных и подходы к интерпретации результатов, сходства и особенности. Объекты статистического анализа в документоведении и архивном деле. Количественные и качественные признаки: измерение показателей. Номинальная шкала качественных признаков. Ранговые признаки.	2		4	8
2. Методы математической статистики					
	Теория статистического оценивания. Генеральная совокупность и выборочные данные, объем выборки. Оценка признака, величина ошибки. Доверительный интервал. Теория статистической проверки гипотез. Статистический критерий. Уровень значимости. Допустимое значение и критическая область. Гипотезы о типе закона распределения признака или критерии согласия; гипотезы о числовых значениях параметров совокупности; гипотезы о типе зависимости признаков.	2		4	8
3. Статистические функции MS Excel					
	Общие сведения о надстройке “Пакет анализа”. Технология работы в режиме “Анализ данных”. Работа с мастером функций. Статистические функции, связанные с режимом “Гистограмма”. Генерация случайных чисел. Статистические функции непрерывных и дискретных распределений. Выборка: технология работы. Статистические функции, связанные с режимом “Описательная статистика”. Статистические функции, связанные с режимом “Ранг и персентиль”. Выборочная дисперсия: правила расчета дисперсии, теоретическая дисперсия выборочного среднего. Основные свойства дисперсии. Методы проверки статистических гипотез. Дисперсионный анализ. Статистические методы изучения взаимосвязей явлений и процессов: ковариация и корреляция, регрессия. Статистические методы изучения динамики процессов: скользящее среднее и экспоненциальное сглаживание, трендовые модели, анализ Фурье.	2		4	9
4. Статистический пакет SPSS					

	Общие сведения о программе. Частотный анализ (частотные таблицы, графическое представление). Отбор данных: выбор наблюдения, извлечение случайной выборки, сортировка наблюдений, разделение наблюдений на группы. Модификация данных: вычисление новых переменных, подсчет частоты появлений определенных значений, перекодирование значений, агрегирование данных, ранговые преобразования, примеры вычисления новых переменных. Исследование данных. Таблицы сопряженности: создание, графическое представление, статистические критерии для таблиц. Анализ множественных ответов: дихотомный и категориальный методы. Сравнение средних: зависимые и независимые выборки. Непараметрические тесты. Корреляции. Регрессионный, дисперсионный, дискриминантный, факторный, кластерный анализ. Стандартные и интерактивные графики. Экспортирование выходных данных.	2		4	9
5. Статистическое описание данных в статистическом пакете STATISTICA					
	Возможности организации данных в пакете STATISTICA. Модульная структура пакета. Принципы работы в модулях. Ввод данных непосредственно в таблицу; экспорт данных из пакетов Lotus/Quattro, Excel, SPSS, dBASE. Чтение обычных ASCII-файлов. Добавление, удаление, перемещение, копирование и переименование строк и столбцов таблицы (объектов и признаков). Создание новых признаков на основе исходных (подсчет процентов и долей, относительных и суммарных показателей и т.п.). Дескриптивная статистика. Основные статистические характеристики: меры среднего уровня, меры рассеяния. Частотные распределения количественных и качественных признаков. Визуализация данных. Категоризованные распределения. Выборочный метод. Нормальное распределение. Основные понятия выборочного метода. Ошибки выборки. Точность и надежность. Доверительный интервал. Определение объема выборки. Статистическое оценивание доли качественного признака. Статистическая проверка гипотез. Основные понятия: статистическая гипотеза, статистический критерий и статистическая характеристика, уровень значимости и критическая область, ошибки, односторонняя и двусторонняя проверка. Критерии для средних. Критерии согласия.	4		8	15
6. Статистический анализ взаимосвязей в статистическом пакете STATISTICA					
	Корреляционный и регрессионный анализ. Анализ парных взаимосвязей. Построение диаграмм рассеяния. Построение уравнения линейной регрессии. Коэффициент корреляции. Проверка гипотезы о значимости коэффициента корреляции. Коэффициент детерминации. Множественная корреляция и регрессия. Визуализация множественной зависимости в пространстве трех переменных. Уравнение множественной регрессии. Проверка значимости в регрессионном анализе. Анализ взаимосвязей качественных данных. Типы качественных данных. Взаимосвязь ранговых качественных данных. Взаимосвязь номинальных качественных данных. Таблицы сопряженности. Критерий значимости связи качественных признаков. Коэффициенты взаимосвязи качественных признаков. Бинарные признаки. Четырехклеточные таблицы.	2		4	9
7. Многомерный статистический анализ в статистическом пакете STATISTICA					

	Методы многомерной классификации. Кластерный анализ. Агломеративно-иерархический метод. Метод k-средних. Гибкая классификация: использование нечетких множеств. Факторный анализ. Факторные нагрузки. Факторные веса. Метод главных компонент. Факторный анализ как способ классификации.	2		4	8
8. Анализ динамики в статистическом пакете STATISTICA					
	Анализ временных рядов. Характеристики скорости и интенсивности изменения временного ряда. Средние характеристики временного ряда. Составляющие временного ряда. Временной тренд.	1		2	9
	ВСЕГО	17		34	75

4.2.Содержание практических (семинарских) занятий

Практические (семинарские) работы не предусмотрены учебным планом по данной дисциплине.

4.3.Содержание лабораторных занятий

Курс 2 Семестр № 3

№ п/п	**Наименование раздела учебной дисциплины (модуля)**	**Наименование лабораторных (практических) работ**	**Всего часов**
1.	Случайные величины, их статистические характеристики	Лабораторная работа №1	**4**
2.	Разведочный анализ данных	Лабораторная работа №2	**4**
3.	Ввод и редактирование формул	Лабораторная работа №3	**4**
4.	Работа с формулами массива	Лабораторная работа №4	**4**
5.	Зависимые случайные величины	Лабораторная работа №5	**8**
6.	Парная линейная регрессия	Лабораторная работа №6	**4**
7.	Методы вычисления параметров парной линейной регрессии	Лабораторная работа №7	**4**
8.	Автокорреляция остатков. Статистика Дарбина-Уотсона	Лабораторная работа №8	**2**
	ИТОГО:		**34**

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

№ п/п	Наименование раздела дисциплины	Содержание вопросов (типовых заданий)
1	Случайные величины. Их статистические характеристики	1. Что такое случайная величина? 2. Что такое непрерывная и дискретная случайные величины? 3. Каковы статистические характеристики случайных величин? Что такое дисперсия?
2	Разведочный анализ данных	1. Что такое гистограмма? 2. Что такое функция, плотность распределения случайной величины? 3. Что такое эмпирическая функция, плотность распределения случайной величины? 4. Что такое мода, медиана распределения случайной величины? 5. Что такое нормальное распределения случайной величины?
3	Ввод и редактирование формул	1. Что понимают под формулой в Excel? 2. Порядок формирования арифметической и геометрической прогрессий в Excel. 3. Объяснить расстановку ссылок в документах Excel.
4	Работа с формулами массива	1. Как создается формула массива? 2. Как изменить формулу массива? 3. Методика ввода данных для формулы массива в документах Excel.
5	Зависимые случайные величины	1. Какие зависимые величины являются зависимыми, независимыми? 2. Какие случайные величины являются коррелированными, некоррелированными? 3. Что такое коэффициент корреляции? 4. Что означает коэффициент корреляции, близкий к 0, 1, -1?
6	Парная линейная регрессия	1.Что такое условное математическое ожидание $M_x(Y)$? 2. Что такое корреляционная и регрессионная зависимости Y от X? 3.Что такое модельное уравнение регрессии? 4. Что такое спецификация модели регрессии, объясняемая и объясняющая переменные, параметры модели? 5. Почему невозможно получить модельное уравнение регрессии? 6. Что такое выборочное уравнение регрессии? 7. Что такое выборочное условное среднее? 8. Каковы задачи регрессионного анализа? 9. Какие модели наблюдения соответствуют модельному и выборочному уравнению регрессии? 10. Что такое парная линейная регрессия, для чего она используется?

7	Методы вычисления параметров парной линейной регрессии	1. Что такое зависимая и независимая переменная, регрессия, коэффициент регрессии? 2. Что такое значимость уравнения регрессии? Что такое ошибка первого или второго рода? Нулевая гипотеза. 3. Что такое точечный прогноз и доверительный интервал? Какие прогнозы можно построить для линейной регрессии?
8	Автокорреляция остатков. Статистика Дарбина-Уотсона	1. Дать сравнительные характеристики исходных данных двух разделов, диаграмм остатков и знаков остатков. 2. Как проявляется зависимость остатков относительно линии регрессии? 3. Что такое статистика Дарбина-Уотсона, для чего она предназначена? 4. Как влияет на значение статистики Дарбина-Уотсона зависимость остатков? 5. Как влияет на значение статистики Дарбина-Уотсона независимость остатков? 6. При каких значениях статистики Дарбина-Уотсона можно сделать вывод о зависимости или независимости остатков? 7. Какой вывод следует сделать о характере зависимости между объясняемой и объясняющей переменной, если выяснилось, что остатки зависимы?

5.2.Перечень тем курсовых проектов, курсовых работ, их краткое содержание и объем

Курсовые проекты, работы не предусмотрены учебным планом по данной дисциплине.

5.3.Переченьиндивидуальных домашних заданий, расчетно-графических заданий

Цель расчетно-графического задания – углубление теоретических знаний и укрепление у студентов практических навыков в области анализа данных, его основных терминов и понятий, методов и приемов для осуществления аналитических расчетов, оценить полученные результаты и принимать адекватные управленческие решения на основе имеющейся информации.

На выполнение расчетно-графического задания предусмотрено 18 часов самостоятельной работы студента. При выполнении заданий студенты используют знания, полученные в процессе изучения дисциплины.

Расчетно-графическое задание включает теоретический вопрос и практическое задание. Студент имеет право самостоятельно выбрать тему теоретического вопроса из приведенного перечня. Практическое задание выполняется по вариантам, который определяется по номеру зачетной книжки.

РГЗ предоставляется преподавателю для проверки в виде работы на бумажных листах формата А4. Отчет расчетно-графического задания должен иметь следующую структуру: титульный лист, теоретическая часть, практическая часть.

Темы теоретической части расчетно-графического задания:

1. Этапы решения задачи анализа данных и их взаимосвязи
2. Случайные переменные и случайные выборки данных
3. Применение непараметрического теста для парных данных
4. Сводные таблицы. Вычисление ожидаемого количества наблюдений.
5. Распределение вероятностей.
6. Анализ табличных данных.
7. Статистика хи-квадрат Пирсона
8. Простая линейная взаимосвязь
9. Уравнение регрессии. Интерпретация параметров регрессии
10. Проверка модели регрессии
11. Проверка допущений регрессии
12. Понятие «корреляции». Матрица корреляции.
13. Матрица точечных диаграмм корреляций
14. Построение и графическое изображение вариационных рядов.
15. Статистические характеристики рядов распределения.
16. Проверка гипотезы о законе нормального распределения.
17. Статистические оценки параметров распределения.
18. Уравнение множественной регрессии
19. Дисперсионный анализ. Однофакторный дисперсионный анализ
20. Двухфакторный дисперсионный анализ

Типовое задание практической части РГЗ.

1. Определить для произвольных X и Y значения выражений:

$$\sqrt{\frac{|x+4|+\sqrt[3]{|y+16|}}{2-3!}}; \qquad \frac{\lg\left(5|x|-4\sqrt[3]{|x+2y|}\right)}{e^{|y|^3}}$$

$$\frac{\sin^3\left(\frac{\pi}{3}+1\right)\cdot 2tg\left(\left|\frac{\pi}{4}+1\right|\right)}{arctg\left(\frac{x}{5}\right)} \qquad \frac{cos^4\left(arcsin(\sqrt[3]{2y})\right)}{4x!+y};$$

2. Вычислить значения выражений:

$$\prod_{j=1}^{5}\sum_{i=1}^{6}\frac{6a_i}{i!+j!} \qquad \sum_{j=1}^{8}\prod_{i=1}^{7}\frac{i^2+6j^2}{5b_j}$$

где a_i– арифметическая прогрессия a_1=0.1, d=2
b_i– геометрическая прогрессия b_1=0.5, g=0.7

3. Табулировать функцию

$$F=\frac{\sqrt[5]{|3x|+|y|}}{a+4b^2}$$

для 9 значений x: x_1 = 1,3; шаг h = 0,6; используя метод таблиц подстановки (a, b, y – произвольные)

4. Решить задачу 2 следующими способами:

   - подготовка и копирование формул с различными типами ссылок,
   - использование формулы массива

   при следующих условиях: P_0=5000, N=8, p=5%, начисление идет по схеме простых процентов.

5. Решить задачу 6 – вариант Б – способом построения специфических вычислительных схем при следующих условиях: начисление идет по схеме сложных процентов, p=5%, поток денежных средств задан следующей таблицей.

Поток денежных средств							
Срок	0	1	2	3	4	5	6
Размер вклада	2000	50	60	70	80	90	100

6. Решить задачу 8 с использованием встроенных функций при следующих условиях P=8000, G=1000, N=6, метод уменьшающегося остатка. На основе полученного решения построить диаграмму типа График в двухмерной и трехмерной плоскостях.

5.4.Перечень контрольных работ

Контрольные работы не предусмотрены учебным планом по направлению.

6. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

6.1. Перечень основной литературы

1. Александровская Ю.П. Многомерный статистический анализ в экономике [Электронный ресурс]: учебное пособие / Ю.П. Александровская. — Электрон.текстовые данные. — Казань: Казанский национальный исследовательский технологический университет, 2017. — 96 с. — 978-5-7882-2191-5. — Режим доступа: http://www.iprbookshop.ru/79330.html
2. Жуковский, О.И. Информационные технологии и анализ данных: учебное пособие / О.И. Жуковский; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск: Эль Контент, 2014. - 130 с. : схем., ил. - Библиогр.: с. 126. - ISBN 978-5-4332-0158-3; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=480500
3. Мельниченко А.С. Математическая статистика и анализ данных [Электронный ресурс] : учебное пособие / А.С. Мельниченко. — Электрон.текстовые данные. — М.: Издательский Дом МИСиС, 2018. — 45 с. — 978-5-906953-62-9. — Режим доступа: http://www.iprbookshop.ru/78563.html

6.2. Перечень дополнительной литературы

1. Минашкин В.Г. Бизнес-статистика и прогнозирование [Электронный ресурс]: учебное пособие/ Минашкин В.Г., Садовникова Н.А., Шмойлова Р.А.— Электрон.текстовые данные.— М.: Евразийский открытый институт, 2010.— 256 с.— Режим доступа: http://www.iprbookshop.ru/10624.
2. Крутиков, В.Н. Анализданных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=278426

6.3. Перечень интернет ресурсов

1. Информационно-аналитический ресурс http://machinelearning.ru
2. Программа анализа данных AtteStat 12.5 - http://attestatsoft.narod.ru
3. Кузнецов С.Д. Основы современных баз данных - http://cdokp.tstu.tver.ru/moodle/mod/resource/view.php?id=779
4. Кричевский М.Л. Интеллектуальный анализ данных в менеджменте - http://cdokp.tstu.tver.ru/moodle/mod/resource/view.php?id=781

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

Лекционные и практические занятия – аудитория, оснащенная презентационной техникой, комплект электронных презентаций. Аудитория оборудована специализированной мебелью, мультимедийным проектор, переносным экраном, ноутбуком.

Читальный зал библиотеки для самостоятельной работы.

Лицензионное ПО: MicrosoftOfficeProfessional 2013.GoogleChrome свободно распространяемое ПО согласно условиям лицензионного соглашения.KasperskyEndpointCenter 10.

8. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Утверждение рабочей программы без изменений

Рабочая программа без изменений утверждена на 2017 /2018 учебный год.

Протокол № 11 заседания кафедры от «06» 06 2017г.

Заведующий кафедрой _____ Ю.И. Селиверстов

подпись, ФИО

Директор института _____ Ю.А.Дорошенко

подпись, ФИО

8. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Утверждение рабочей программы с изменениями, дополнениями

Рабочая программа с изменениями, дополнениями утверждена на 2018/2019 учебный год.

Протокол № 9 заседания кафедры от «21» 05 2018г.

6.1. Перечень основной литературы

1. Александровская Ю.П. Многомерный статистический анализ в экономике [Электронный ресурс]: учебное пособие / Ю.П. Александровская. — Электрон. текстовые данные. — Казань: Казанский национальный исследовательский технологический университет, 2017. — 96 с. — 978-5-7882-2191-5. — Режим доступа: http://www.iprbookshop.ru/79330.html
2. Анализ данных: методические указания к выполнению лабораторных работ для студентов направления подготовки 38.03.05 - Бизнес-информатика / сост.: О. В. Доможирова, В. Ю. Сорокина. - Белгород: Издательство БГТУ им. В. Г. Шухова, 2017. - 52 с.
3. Доможирова О. В., Сорокина В. Ю. Анализ данных: методические указания к выполнению лабораторных работ для студентов направления 38.03.05 – Бизнес-информатика Изд-во БГТУ им. В. Г. Шухова, 2017, 53с. Режим доступа: https://elib.bstu.ru/Reader/Book/2018022111151158300000658993
4. Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск: Эль Контент, 2014. - 130 с. : схем., ил. - Библиогр.: с. 126. - ISBN 978-5-4332-0158-3; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=480500
5. Мельниченко А.С. Математическая статистика и анализ данных [Электронный ресурс] : учебное пособие / А.С. Мельниченко. — Электрон. текстовые данные. — М.: Издательский Дом МИСиС, 2018. — 45 с. — 978-5-906953-62-9. — Режим доступа: http://www.iprbookshop.ru/78563.html

Заведующий кафедрой _____ Ю.И. Селиверстов
подпись, ФИО

Директор института _____ Ю.А.Дорошенко
подпись, ФИО

8. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Утверждение рабочей программы без изменений

Рабочая программа без изменений утверждена на 2019 /2020 учебный год.

Протокол № 9/1 заседания кафедры от «13» 06 2019.

Заведующий кафедрой _____ Ю.И. Селиверстов
подпись, ФИО

Директор института _____ Ю.А. Дорошенко
подпись, ФИО

8. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Утверждение рабочей программы без изменений

Рабочая программа без изменений утверждена на 2020/2021 учебный год.

Протокол № 8 заседания кафедры от «22» 05 2020г.

Заведующий кафедрой _____ Ю.И. Селиверстов
подпись, ФИО

Директор института _____ Ю.А. Дорошенко
подпись, ФИО

8. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Утверждение рабочей программы без изменений

Рабочая программа без изменений утверждена на 2021/2022 учебный год.

Протокол № 8 заседания кафедры от «13» 05 2021г.

Заведующий кафедрой _____ Ю.И. Селиверстов
подпись, ФИО

Директор института _____ Ю.А. Дорошенко
подпись, ФИО

ПРИЛОЖЕНИЯ

Приложение №1

Методические указания для обучающегося по освоению дисциплины (включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине).

Дисциплина «Анализ данных» представляет собой неотъемлемую составную часть подготовки студентов по направлению подготовки бакалавриата 38.03.05 «Бизнес-информатика».

Целью освоения дисциплины является овладение студентами базовыми знаниями об основных принципах, моделях, методах, алгоритмах решения задач в области анализа данных.

Задачами изучения дисциплины являются получение знаний об основных принципах, моделях и задачах анализа данных и распознавания образов, а также методах и алгоритмах их решения; приобретение практических навыков по построению моделей, анализу комбинаторной сложности и решению математических задач, к которым сводятся типовые проблемы в области анализа данных и распознавания образов.

Занятия проводятся в виде лекций и лабораторных занятий. Важное значение для изучения курса имеет самостоятельная работа студентов.

Формы контроля знаний студентов предполагают текущий и промежуточный контроль. Текущий контроль знаний проводится в форме защиты лабораторных работ. Формой промежуточного контроля является дифференцированный зачет.

Распределение материала дисциплины по темам и требования к ее освоению содержатся в рабочей программе дисциплины, которая определяет содержание и особенности изучения курса.

Изучение отдельных тем курса необходимо осуществлять в соответствии с поставленными в них целями, их значимостью, основываясь на содержании и вопросах, поставленных в лекции преподавателя и приведенных в планах и заданиях к практическим занятиям, а также методических указаниях для студентов заочного обучения.

В учебниках и учебных пособиях, представленных в *списке рекомендуемой литературы* содержатся возможные ответы на поставленные вопросы. Инструментами освоения учебного материала являются основные *термины и понятия*, составляющие категориальный аппарат дисциплины. Их осмысление, запоминание и практическое использование являются обязательным условием овладения курсом.

Для более глубокого изучения проблем курса при подготовке к лабораторным занятиям необходимо ознакомиться с публикациями в периодических экономических изданиях и статистическими материалами. Поиск и подбор таких изданий, статей, материалов и монографий осуществляется на основе библиографических указаний и предметных каталогов.

Изучение следует завершать выполнением лабораторных заданий, ответами на тесты, решением задач, содержащихся в соответствующих разделах учебников

и методических пособий по курсу «Анализ данных». Для обеспечения систематического контроля над процессом усвоения тем курса следует пользоваться перечнем контрольных вопросов для проверки знаний по дисциплине, содержащихся в планах и заданиях к практическим занятиям и методическим указаниях для студентов заочного отделения. Если при ответах на сформулированные в перечне вопросы возникнут затруднения, необходимо очередной раз вернуться к изучению соответствующей темы, либо обратиться за консультацией к преподавателю.

Успешное освоение курса дисциплины возможно лишь при систематической работе, требующей глубокого осмысления и повторения пройденного материала, поэтому необходимо делать соответствующие записи по каждой теме.

Основными формами текущего контроля успеваемости являются наблюдение и проверка заданий, выполняемых на компьютере в статистических пакетах MS Excel.

Степень освоения студентами системы необходимых компетенций проверяется на зачете в конце семестра. Зачет проводится в компьютерном классе и подразумевает выполнение практических заданий из предложенного перечня.

Тема 1. Статистическая обработка и анализ данных: предмет изучения

“Математическая статистика” и “анализ данных”: способы получения данных и подходы к интерпретации результатов, сходства и особенности.

Количественные и качественные признаки: измерение показателей. Номинальная шкала качественных признаков. Ранговые признаки. Объекты статистического анализа в документоведении и архивном деле.

Тема 2. Методы математической статистики

Теория статистического оценивания. Генеральная совокупность и выборочные данные, объем выборки. Оценка признака, величина ошибки. Доверительный интервал.

Теория статистической проверки гипотез. Статистический критерий. Уровень значимости. Допустимое значение и критическая область. Гипотезы о типе закона распределения признака или критерии согласия; гипотезы о числовых значениях параметров совокупности; гипотезы о типе зависимости признаков.

Тема 3. Статистические функции MS Excel

Общие сведения о надстройке “Пакет анализа”. Технология работы в режиме “Анализ данных”. Работа с мастером функций. Статистические функции, связанные с режимом “Гистограмма”. Выборка: технология работы. Статистические функции, связанные с режимом “Описательная статистика”. Статистические функции, связанные с режимом “Ранг и персентиль”. Генерация случайных чисел. Статистические функции непрерывных и дискретных распределений. Методы проверки статистических гипотез. Дисперсионный анализ. Статистические методы изучения взаимосвязей явлений и процессов: ковариация и корреляция, регрессия. Статистические методы изучения динамики процессов: скользящее среднее и экспоненциальное сглаживание, трендовые модели, анализ Фурье.

Тема 4. Статистический пакет SPSS

Общие сведения о программе. Частотный анализ (частотные таблицы, графическое представление). Отбор данных: выбор наблюдения, извлечение случайной выборки, сортировка наблюдений, разделение наблюдений на группы. Моди-

фикация данных: вычисление новых переменных, подсчет частоты появлений определенных значений, перекодирование значений, агрегирование данных, ранговые преобразования, примеры вычисления новых переменных. Исследование данных. Таблицы сопряженности: создание, графическое представление, статистические критерии для таблиц. Анализ множественных ответов: дихотомный и категориальный методы. Сравнение средних: зависимые и независимые выборки. Непараметрические тесты. Корреляции. Регрессионный, дисперсионный, дискриминантный, факторный, кластерный анализ. Стандартные и интерактивные графики. Экспортирование выходных данных.

Тема 5. Статистическое описание данных в статистическом пакете STATISTICA

Возможности организации данных в пакете STATISTICA. Модульная структура пакета. Принципы работы в модулях. Ввод данных непосредственно в таблицу; экспорт данных из пакетов Lotus/Quattro, Excel, SPSS, dBASE. Чтение обычных ASCII-файлов. Добавление, удаление, перемещение, копирование и переименование строк и столбцов таблицы (объектов и признаков). Создание новых признаков на основе исходных (подсчет процентов и долей, относительных и суммарных показателей и т.п.).

Дескриптивная статистика. Основные статистические характеристики: меры среднего уровня, меры рассеяния. Частотные распределения количественных и качественных признаков. Визуализация данных. Категоризованные распределения.

Выборочный метод. Нормальное распределение. Основные понятия выборочного метода. Ошибки выборки. Точность и надежность. Доверительный интервал. Определение объема выборки. Статистическое оценивание доли качественного признака.

Статистическая проверка гипотез. Основные понятия: статистическая гипотеза, статистический критерий и статистическая характеристика, уровень значимости и критическая область, ошибки, односторонняя и двусторонняя проверка. Критерии для средних. Критерии согласия.

Тема 6. Статистический анализ взаимосвязей в статистическом пакете STATISTICA

Корреляционный и регрессионный анализ. Анализ парных взаимосвязей. Построение диаграмм рассеяния. Построение уравнения линейной регрессии. Коэффициент корреляции. Проверка гипотезы о значимости коэффициента корреляции. Коэффициент детерминации.

Множественная корреляция и регрессия. Визуализация множественной зависимости в пространстве трех переменных. Уравнение множественной регрессии. Проверка значимости в регрессионном анализе.

Анализ взаимосвязей качественных данных. Типы качественных данных. Взаимосвязь ранговых качественных данных. Взаимосвязь номинальных качественных данных. Таблицы сопряженности. Критерий значимости связи качественных признаков. Коэффициенты взаимосвязи качественных признаков. Бинарные признаки. Четырехклеточные таблицы.

Тема 7. Многомерный статистический анализ в статистическом пакете STATISTICA

Методы многомерной классификации.

Кластерный анализ. Агломеративно-иерархический метод. Метод k-средних. Гибкая классификация: использование нечетких множеств.

Факторный анализ. Факторные нагрузки. Факторные веса. Метод главных компонент. Факторный анализ как способ классификации.

Тема 8. Анализ динамики в статистическом пакете STATISTICA

Анализ временных рядов. Характеристики скорости и интенсивности изменения временного ряда. Средние характеристики временного ряда. Составляющие временного ряда. Временной тренд.

В дисциплине «Анализ данных» одно из центральных мест занимает обработка данных (наблюдений, событий, примеров, результатов экспериментов и т.п.). С ее помощью осуществляется преобразование и приведение исходного фактического материала к наиболее удобной для восприятия форме с целью принятия управленческих решений и извлечения из данных необходимой информации.

Рекомендуется дать студентам информацию о литературе, которая необходима для более детального и углубленного изучения темы.

На лекциях рекомендуется использовать средства ТСО, рассмотреть вопросы, которые рассматривают состав и структура автоматизированных систем бухгалтерского учета, анализа и аудита, состав и структура современных систем управления ресурсами предприятия, методику компьютерного решения экономических задач.