

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

**Кафедра прикладной математики факультета математики и
компьютерных наук**

Образовательная программа

01.04.02-Прикладная математика и информатика

Профиль подготовки
Математическое моделирование и вычислительная математика

Уровень высшего образования
магистратура

Форма обучения
Очная

Статус дисциплины: ***Вариативный***

Махачкала, 2018

Рабочая программа дисциплины «Статистические методы обработки данных» составлена в 2018 году в соответствии с требованиями ФГОС ВО по направлению подготовки **01.04.02-Прикладная математика и информатика** (уровень магистратура) утверждённого приказом Министерства образования и науки Российской Федерации от 28 августа 2015 года № 911

Разработчик: кафедра прикладной математики:
Лугуева А.С, к.ф-м.н., доцент,

Рабочая программа дисциплины одобрена:
на заседании кафедры прикладной математики
от « 14 » июня 2018г., протокол № 10
Зав. кафедрой Кадиев Р.И. (подпись) Кадиев Р.И.

на заседании Учебно-методической комиссии ФМиКН
от « 27 » июня 2018г., протокол № 6
Председатель УМС Бейбалаев В.Д. (подпись) доц. Бейбалаев В.Д.

Рабочая программа дисциплины согласована с учебно-методическим управлением « 28 » 06 2018 г. _____ (подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Статистические методы обработки данных» входит в вариативную часть образовательной программы подготовки магистров по направлению 01.04.02-Прикладная математика и информатика.

Дисциплина реализуется на факультете М и КН кафедрой ПМ.

Данный курс лекций ориентирован на изучение простых, но наиболее часто используемых методов статистической обработки данных.

Курс лекций носит сугубо прикладной характер. Особое внимание при разборе материала уделяется анализу ошибок, которые обычно делают начинающие исследователи при применении того или иного метода статистической обработки данных измерений.

Курс лекций преподается с использованием компьютерной сети. Лекционный материал содержит множество конкретных примеров, которые разбираются в интерактивном режиме.

При изучении курса «Статистические методы обработки данных» студенты должны иметь теоретическую подготовку по информатике и основным разделам математического анализа, дифференциального и интегрального исчисления. Студенты также должны обладать начальными практическими навыками работы на компьютере.

Дисциплина нацелена на формирование следующих компетенций выпускника: общекультурных – ОК-1, общепрофессиональных - ОПК-3, профессиональных – ПК-1.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: *лекции, лабораторные занятия и самостоятельная работа.*

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости в форме коллоквиума и промежуточный контроль в форме экзамена.

Объем дисциплины 3 зачетные единицы, в том числе в академических часах по видам учебных занятий

Семе стр	Учебные занятия							Форма промежуточной аттестации (зачет, дифференцирован ный зачет, экзамен
	в том числе							
	всего	Контактная работа обучающихся с преподавателем				СРС, в том числе экза мен		
		Всего	из них					
			Лекци и	Лабораторны е занятия	КСР	консульта ции		
В	108	36	6	30			72	экзамен

1. Цели освоения дисциплины

Цель изучения курса «Статистические методы обработки данных» является введение студентов в проблематику прикладной математики и информатики; расширение и углубление знаний студентов по вопросам статистической обработки данных, закрепление студентами ряд понятий изученных в курсах, позволяющего выпускнику успешно работать в избранной сфере деятельности, обладать универсальными и предметно специализированными компетенциями, способствующими его социальной мобильности, востребованности на рынке труда и успешной профессиональной карьере.

2. Место дисциплины в структуре ОПОП магистратуры

Дисциплина «Статистические методы обработки данных» входит в вариативную часть образовательной программы подготовки магистров по направлению 01.04.02 - Прикладная математика и информатика.

Курс «Статистические методы обработки данных» вводится после изучения дисциплин алгебра, информатика, математический анализ, теория вероятностей и математическая статистика, дифференциальные уравнения, так как для успешного усвоения этого курса студентам необходимы знания по указанным дисциплинам.

Данный курс лекций ориентирован на изучение простых, но наиболее часто используемых методов статистической обработки данных.

Курс лекций носит сугубо прикладной характер. Особое внимание при разборе материала уделяется анализу ошибок, которые обычно делают начинающие исследователи при применении того или иного метода статистической обработки данных биометрических измерений.

Лекционный материал содержит множество конкретных примеров.

При изучении курса «Статистические методы обработки данных» студенты должны иметь теоретическую подготовку по информатике и основным разделам математического анализа, дифференциального и интегрального исчисления. Студенты также должны обладать начальными практическими навыками работы на компьютере.

3. Компетенции обучающегося, формируемые в результате

освоения дисциплины (перечень планируемых результатов обучения).

Код компетенции из ФГОС ВО	Наименование компетенции из ФГОС ВО	Планируемые результаты обучения
ОК -1	Способность к абстрактному мышлению, анализу, синтезу.	Знает: основные формулы основы математической статистики, сбора, обработки и анализа статистических данных. Умеет: на основе применения аппарата математической статистики принимать нужные решения. Строить модели различных прикладных задач и перекладывать их на ЭВМ. Владеет: навыками дискуссии по профессиональной тематике
ОПК-3	Способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение	Знает: основные методы обработки статданных. Особое внимание при разборе материала уделяется анализу ошибок, которые обычно делают начинающие исследователи при применении того или иного метода статистической обработки Умеет: строить математические модели различных прикладных задач; программировать их для ЭВМ, проводить численное моделирование и по результатам этого находить новые закономерности и, если нужно, менять прежние модели. Владеет: статистическими пакетами для обработки и анализа экспериментальных данных
ПК-1	Способность проводить научные исследования и получать новые прикладные результаты самостоятельно и в составе научного коллектива.	Знает: - современные тенденции развития научных и прикладных достижений и их использование в прикладном исследовании; - подходы использования современных методов для решения научных и практических задач. Умеет: самостоятельно разбираться в современных методах и способах представления данных Владеет: навыками поиска информации, методами сбора информации и навыками применения набора стандартных методов статистической обработки данных.

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 3 зачетные единицы, 108 академических часов.

4.2. Структура дисциплины.

№	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Самостоятельная работа	Формы текущего контроля успеваемости (по неделям семестра)
				Лекц	Лаб	Прак	Контроль		Форма промежуточной аттестации (по семестрам)
Модуль 1. Обработка данных в рамках линейной регрессионной модели									
1	Массивы данных в прикладных задачах. Статистический и детерминированный подходы. Постановка задачи обработки данных.	В	1-2	2	6			6	Формы текущего контроля: устные опросы, реферат, выполнение лабораторной работы
2	Понятие функции распределения случайной величины и ее числовых характеристик. Точечные и интервальные оценки массива данных	В	3-4	2	6			4	
3	Регрессии, аппроксимирующие функции.	В	5-6		4			6	
	Итого по 1 модулю			4	16			16	
Модуль 2. Факторный и кластерный анализ									
4.	Выявление влияющих факторов. Элементы многофакторного анализа.	В	7-8	2	4			8	Формы текущего контроля: устные опросы, реферат, доклады ---
5.	Кластерный анализ. Методы	В	9-10		4			6	

	иерархической классификации.								
6.	Многомерное шкалирование.	В	11-12		6			6	
	Итого по 2 модулю			2	14			20	
Модуль 3 Подготовка к экзамену									
	Подготовка к экзамену	В						36	
	Итого по 3 модулю	В						36	
	ИТОГО:			6	30			72	экзамен

4.3. Содержание дисциплины, структурированное по темам (разделам).

Лекции

Модуль 1 Обработка данных в рамках линейной регрессионной модели

Тема 1. Массивы данных в прикладных задачах. Статистический и детерминированный подходы. Постановка задачи обработки данных.

Массивы данных как основной метод получения количественно информации о технических объектах с изменяемыми характеристиками. Философские обоснования статистической и детерминированной природы физических объектов. Задачи обработки массивов данных в зависимости от их природы (одномерные массивы, массивы функций двух и более переменных, интегральные зависимости связи между влияющими факторами, выявление различия/согласия различных массивов. Примеры массивов данных в задачах техногенной безопасности (загрязняющие вещества, воздействия ЭМП на население, многомерные задачи расчетов рисков и т.п.).

Тема 2. Понятие ФР случайной величины и ее числовых характеристик. Точечные и интервальные оценки массива данных

Функции распределения случайной величины (СВ) как основная форма ее представления. Математическое ожидание и дисперсия распределения СВ. Основные законы распределения: нормальный ЗР, бинормальный ЗР, равномерной плотности, распределение Пуассона и редких событий. Применение пакетов прикладных программ.

Используемые инструменты Mathcad. Машинные формы массивов

данных (выборки). Способы вычисления оценок математического ожидания и дисперсии. Понятия доверительной вероятности и доверительного интервала. Способы вычислений доверительных интервалов для оценок математического ожидания и дисперсии (стандарта отклонения). Применения инструментов пакетов MS Office Mathcad.

Тема 3. Регрессии, аппроксимирующие функции, МНК

Линейная регрессия для массива двух переменных. Нелинейная регрессия. Метод наименьших квадратов и его машинная реализация. Выбор аппроксимирующих функций. Применения инструментов пакетов Mathcad, Statistics.

Модуль 2. Факторный и кластерный анализ

Тема 4 Выявление влияющих факторов. Элементы многофакторного анализа.

Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ. Применения инструментов пакетов MS Office, Mathcad.

Понятие о критериях согласия. Проверка гипотезы о принадлежности двух выборок одной генеральной совокупности. Проверка гипотезы о функции распределения выборки. Проверка гипотезы о равенстве математических ожиданий двух выборок при равных неизвестных дисперсиях. Применения инструментов пакетов Mathcad.

Тема 5. Кластерный анализ. Методы иерархической классификации.

Общие принципы кластерного анализа. Виды кластерного анализа: двухэтапный кластерный анализ, иерархический кластерный анализ, кластерный анализ методом К-средних. Иерархический кластерный анализ. Задание метода иерархического кластерного анализа (Метод кластеризации, мера, преобразование значений, преобразование меры). Статистики для процедуры Иерархический кластерный анализ (порядок агломерации, матрица близостей, принадлежность к кластерам). Графики для процедуры Иерархический кластерный анализ (дендограмма, сосульчатый график). Сохранение новых переменных в процедуре. Кластеры на факторах. Интерпретация результатов кластерного анализа.

Тема 6. Многомерное шкалирование.

Суть и логика многомерного шкалирования. Взаимосвязь факторного анализа и многомерного шкалирования. Возможности визуализации представления данных. Задание размерности и осей координат. Интерпретация результатов многомерного шкалирования.

Лабораторные занятия

Модуль 1. Обработка данных в рамках линейной регрессионной модели

Лабораторная работа №1

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ОПЫТНЫХ ДАННЫХ. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ВБОРКИ

Цель: Научиться основным методам обработки данных, представленных выборкой. Изучить графические представления данных. Овладеет навыками расчета с помощью ЭВМ основных числовых характеристик выборки.

Основным объектом исследования в эконометрике является выборка. Выборкой объема n называются числа $x_1, x_2, \dots, x_n$ получаемые на практике при n – кратком повторении эксперимента в неизменных условиях. На практике выборку чаще всего представляют статистическим рядом. Для этого вся числовая ось, на которой лежат значения выборки, разбивается на k интервалов (это число выбирается произвольно от 5 до 10), которые обычно равны, вычисляются середины интервалов z_n и считается число элементов выборки, попадающих в каждый интервал n_1 . статистическим рядом называется последовательность пар (z_1, n_1) . Рассмотрим решение задачи на ЭВМ и ППП EXCEL на следующей примере.

ПРИМЕР. Дана выборка чисел выручки магазина за 30 дней:

2	4	9	1	3	8	3	7	6	7	6	6	6	4	5
5	0	5	1	9	2	9	8	2	7	2	1	5	2	9

Построим статистический ряд, полигон, гистограмму и кумулятивную кривую.

Откроем книгу программы EXCEL. Введем в первый столбец (ячейки A1-A30) исходные данные. Определим область чисел, на какой лежат данные. Для этого найдем максимальный и минимальный элементы выборки. Введем в B1 «Максимум», а в B2 «Минимум», а в соседних ячейках C1 и C2 определим функции «MAX» и «MIN», в качестве аргументов которых (в графе «число») обведем область данных (ячейки A1-A30). Результатом будут 64 и 81. видно, что все данные укладываются на отрезке [64;81]. Разделим его на 9 (выбирается произвольно от 5 до 10) интервалов:

64-66; 66-68; 68-70; 70-72; 72-74, 74-76, 76-78, 78-80, 80-82. в ячейке D1-D10 вводим верхние границы интегралов группировки – числа 66, 68, 70, 72, 74, 76, 78, 80, 82. Для вычисления частот n_1 используют функцию ЧАСТОТА, находящуюся в категории «Статистические». Введем ее в ячейку E1. в строке «Массив данных» введем диапазон выборки (ячейки A1-A30). В

строке «Двоичный массив» введем диапазон верхних границ интервалов группировки (ячейки D1-D9). Результат функции является массивом и выводится в ячейках E1-E9. для полного выбора (не только первого числа в E1) нужно выделить ячейки E1-E9, обведя их мышью, и нажать F2, а далее одновременно CTRL+SHIFT+ENTER. Результат – частоты интервалов 2,2,5,7,3,7,3,0,1.

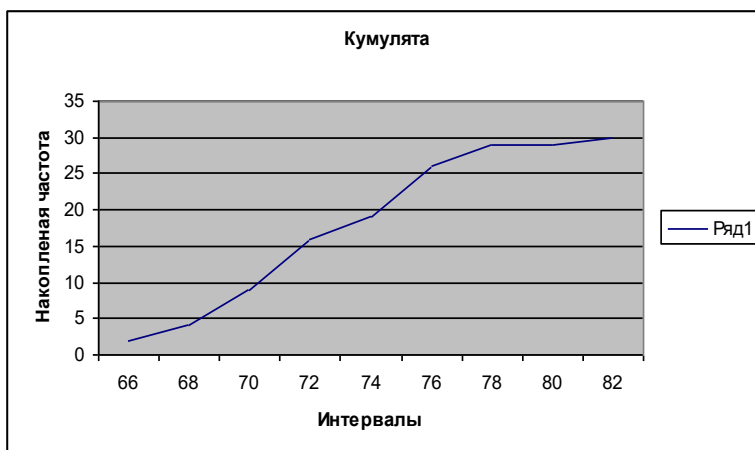
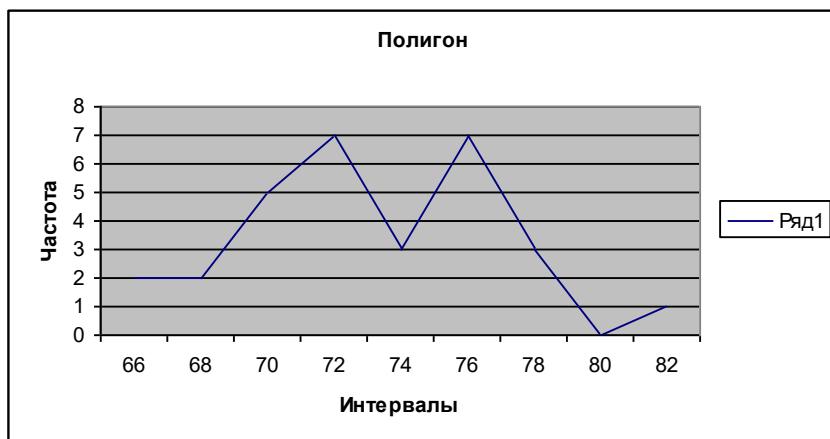
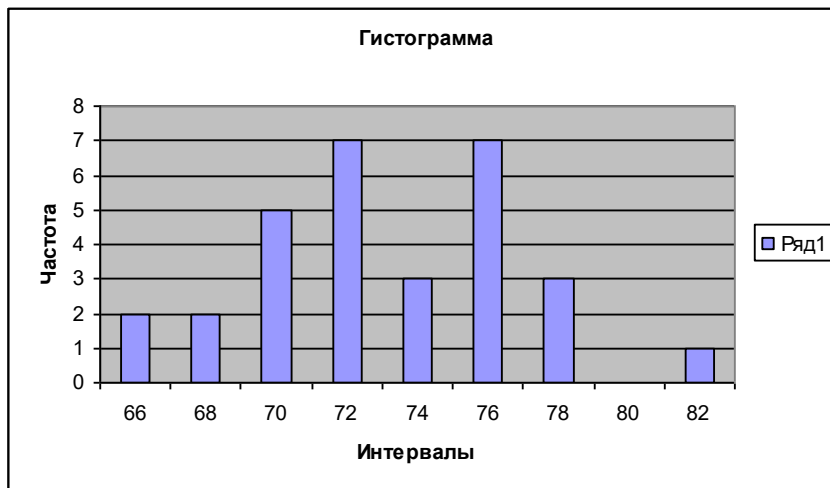
Для построения гистограммы нужно выбрать ВСТАВКА/ДИАГРАММА или нажать на соответствующий значок на основной панели (при этом курсор должен стоять в свободной ячейке) далее выбрать тип: ГИСТОГРАММА, вид по выборке, нажать «ДАЛЕЕ», в строке «ПОДПИСИ ОСИ X» ввести интервалы ячейках D1-D5, нажать «ДАЛЕЕ» ввести название «ГИСТОГРАММА», подписи осей «ИНТЕРВАЛЫ» и «ЧАСТОТА», нажать «ГОТОВО». Для создания полигона сделать то же самое, только вместо типа диаграммы «ГИСТОГРАММА», выбрать «ГРАФИК». Для построения кумулятивной кривой нужно посчитать накопленные частоты. Для этого в ячейку F1 вводим «=E1», в F2 – вводим «=F1+E2» и автозаполнением перетаскиваем эту ячейку до F9. далее строим график как и в случае полигона, но в строке «ДИАПАЗОН» вводим накопленные частоты, ссылаясь на F1- F9, а на вкладке «РЯД», в строке «ПОДПИСИ ОСИ X» вводим интервалы в ячейках D1-D9.

Находим основные числовые характеристики выборки. Для их ввода выделяем два столбца, например G и H, в первом вводим название характеристики, во втором – функцию, в которой в качестве массива данных (строка) «ЧИСЛО1»), указать ссылку на A1-A30

Характеристика	Функция
Объем выборки	30
Выборочное среднее	72,46666667
Дисперсия	15,63678161
Стандартное отклонение	3,954337063
Медиана	72
Мода	72
Коэффициент эксцесса	-0,214617804
Коэффициент асимметрии	-0,154098799
Персентиль 40%	72
Персентиль 80%	76

Существует другой способ вычисления числовых характеристик выборки. Для этого ставим курсор в свободную ячейку (например D11). Затем вызываем в меню «Сервис» подменю «Анализ данных». Если в меню «Сервис» отсутствует этот пункт, то в меню «Сервис» нужно выбрать пункт «Надстройки» и в нем поставить флажок напротив пункта «Пакет анализа». В окне «Анализ данных» нужно выбрать пункт «Описательная статистика».

В появившемся окне в поле «Входной интервал» делаем ссылку на выборку A1-A23. Оставляем группирование «По столбцам» в разделе «Параметры вывода» ставим флажок на «Выходной интервал» и в соседнем поле создаем ссылку на верхнюю левую ячейку области вывода (например D11), ставим флажок напротив «Описательная статистика», нажимаем «ОК». результат – основные характеристики выборки (сделайте шире столбцов D, переместив его границу в заголовок).



Лабораторная работа № 2

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Цель: Ознакомиться с методом проверки основных статистических гипотез, используемых в экономике, с помощью ЭВМ.

1. ПРОВЕРКА ГИПОТЕЗЫ О СООТВЕТСТВИИ (КРИТЕРИЙ СОГЛАСИЯ)

Используется для проверки предположения о том, что полученные в результате наблюдений данные соответствуют нормам. Рассматривается гипотеза о том, что отклонения от норм невелики, и ими можно пренебречь. При этом задается доверительная вероятность p которая имеет смысл вероятности не ошибиться при принятии гипотезы. Рассмотрим проверку на примере.

ПРИМЕР: 1. при производстве микросхем процессоров используются кристаллы кварца. Стандартом предусмотрено, чтобы 50% образцов не было обнаружено ни одного дефекта кристаллической структуры, у 15% - один дефект, у 13% - 2 дефекта, у 12% - 3 дефекта, у 10% более 3 дефектов. При анализе выборочной партии оказалось, что из 100 экземпляров распределение по дефектам партии оказалось, что из 1000 экземпляров распределение по дефектам следующего (вариант соответствует ЭВМ): Можно ли с вероятностью 0,99 считать, что партия соответствует стандарту?

Введем в A1 заголовок «НОРМА» и ниже в A2-A6 показатели – числа 500, 150, 130, 120, 100. в ячейку B1 введем заголовок «НАБЛЮДЕНИЯ» и ниже в B2-B6 наблюдаемые показатели 516, 148, 131, 110, 95. в третьем столбце вводятся формулы для критерия: C1 заголовок «КРИТЕРИЙ», в C2 формулу « $=(A2-B2)*(A2-B2)/A2$ ». Автозаполнением размножим эту формулу на C3-C6. в ячейку C7 запишем общее значение критерия – сумму столбца C2-C6. для этого поставим курсор в C6 и вызвав функцию в категории «Математический» найдем СУММ и в аргументе «Число 1» укажем ссылку на C2-C6. получится результат критерия $Z = 1,629692308$. Для ответа на вопрос, соответствуют ли опытные показатели нормам, Z сравнивают с критическим значением $Z_{кр}$. Вводим в D1 текст «критическое значение» в E1 вводим функцию ХИ2ОБР (категория «Статистические») у которой два аргумента: «Вероятность» - вводим уровень значимости $\alpha = 1-p$ и «Степени свободы» - вводят число $n-1$, где n – число норм). Результат 13,27670414. видно, что критическое значение больше критерия, следовательно опытные данные соответствуют стандартным и партия с заданной вероятностью можно отнести как соответствующую стандарту.

Норма	Наблюдения	Критерий	Критическое значение	13,27670414
500	516	0,512		
150	148	0,026666667		
130	131	0,007692308		
120	110	0,833333333		
100	95	0,25		
	1000	1,629692308		

2. ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДИСПЕРСИЙ

Используется в случае, если нужно проверить различается ли разброс данных (дисперсии) у двух выборок. Это может использоваться при сравнении точностей обработки деталей на двух станках, равномерности продаж товара в течении некоторого периода в двух городах и т.д. Для проверки статистической гипотезы, о равенстве дисперсий служит F – критерий Фишера. Основной характеристикой критерия является уровень значимости α , которой имеет смысла вероятности ошибиться, предполагая, что дисперсии и, следовательно, точность, различаются. Вместо α в задачах так же иногда задают доверительную вероятность $p=1-\alpha$, имеющую смысл вероятности того, что дисперсии и в самом деле равны. Обычно выбирают критическое значение уровня значимости, например 0,05 или 0,1, и если α больше критического значения, то дисперсии считаются равными, в противном случае, различны. При этом критерий может быть односторонним, когда нужно проверить, что дисперсия конкретной выделенной выборки больше, чем у другой, и двусторонним, когда просто нужно показать, что дисперсии не равны. Существует два способа проверки таких гипотез. Рассмотрим их на примерах.

ПРИМЕР 2. четыре станка в цеху обрабатывают детали. Для проверки точности обработки, взяли выборку размеров деталей у каждого станка. Необходимо сравнить с помощью F -теста попарно точности обработки всех станков (рассмотреть пары 1-2, 1-3, 1-4, 2-3, 2-4, 3-4) и сделать вывод, для каких станков точности обработки (дисперсии) равны, для каких нет. Взять уровень значимости $\alpha=0,02$.

1 станок	29,1	26,2	30,7	33,8	33,6	35,2	23,4	29,3	33,3	26,7
2 станок	29,0	28,9	34,0	29,7	39,4	28,5	35,9	32,6	37,1	28,0
3 станок	25,7	27,5	25,4	28,9	29,9	30,1	29,0	36,6	24,8	27,8
4 станок	32,1	31,0	27,2	29,3	30,4	31,7	30,4	27,3	35,7	31,5

Уровень значимости $\alpha=0,02$. вводим данные выборки (без подписей) в 4 строчки в ячейки A1-J1 и A2-J2 и т.д. соответственно. Для вычисления

ФТЕСТ (массив1; массив2). Вводим A5 подпись A5 «Уровень значимости», а в B5 функцию, ФТЕСТ, аргументами которой должны быть ссылки на ячейку A1-J1 и A2-J2 соответственно. Результат 0,873340161 говорит о том, что вероятность ошибиться, приняв гипотезу о различии дисперсий, около 0,9, что больше критического значения, заданного в условии задачи 0,02. следовательно, можно говорить что опытные данные с большей вероятностью подтверждают предположения о том, что дисперсии одинаковы и точность обработки станков одинакова, такие же результаты показало сравнение остальных пар. Следует отметить, что функции ФТЕСТ выводит уровень значимости двустороннего критерия и если нужно использовать односторонний, то результат необходимо уменьшить вдвое.

29,1	26,2	30,7	33,8	33,6	35,2	23,4	29,3	33,3	26,7
29	28,9	34	29,7	39,4	28,5	35,9	32,6	37,1	28
25,7	27,5	25,4	28,9	29,9	30,1	29	36,6	24,8	27,8
32,1	31	27,2	29,3	30,4	31,7	30,4	27,3	35,7	31,5
Уровень значимости									
1 - 2	0,873340161								
1 - 3	0,688084317								
1 - 4	0,190932274								
2 - 3	0,575576041								
2 - 4	0,144572063								
3 - 4	0,357739717								

3. ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ СРЕДНИХ

Используется для проверки предложения о том, что среднее значения двух показателей, представленных выборками, значимо различаются. Существует три разновидности критерия: один – для связанных выборок, и два для несвязных выборок (с одинаковыми и разными дисперсиями). Если выборки не связаны, то предварительно нужно проверить гипотезу о равенстве дисперсий, чтобы определить, какой из критериев использовать. Так же как и в случае сравнения дисперсий имеются 2 способа решения задачи, которые рассмотрим на примере.

ПРИМЕР 3. имеются данные о количестве продаж товара в двух городах. Проверить на уровне значимости 0,01 статистическую гипотезу о том, что среднее число продаж товара в городах различно.

23	25	23	22	23	24	28	16	18	23	29	26	31	19
22	28	26	26	35	20	27	28	28	26	22	29		

Используем пакет «Анализ данных». В зависимости от типа критерия выбирается один из трех: «Парный двухвыборочный t-тест для средних» - для связанных выборок, и «Двухвыборочных t-тест с одинаковыми дисперсиями» или «Двухвыборочных t-тест с разными дисперсиями» - для несвязанных выборок. Вызовите тест с одинаковыми дисперсиями, в открывшемся окне в полях «Интервал переменной 1» и «Интервал переменной 2» вводят ссылки на данные (A1-N1 и A2-L2, соответственно), если имеются подписи данных, то ставят флажок у надписи «Метки» (у нас их нет, поэтому флажок не ставится). Далее вводят уровень значимости в поле «Альфа» - 0,01. Поле «Гипотетическая средняя разность» оставляют пустыми. В разделе «Параметры вывода» ставят метку около «Выходной интервал» и поместив курсор в появившемся поле напротив надписи, щелкают левой кнопкой в ячейке B7. вывод результата будет осуществляться начиная с этой ячейки. Нажав на «ОК» появляется таблица результата. Сдвиньте границу между столбцами B и C, C и D, D и E увеличив ширину столбцов B, C и D так, чтобы уместались все надписи. Процедура выводит основные характеристики выборки, t-статистику, критические значения этих статистик и критические уровни значимости «P(T<=t) одностороннее» и «P(T<=t) двухстороннее». Если по модулю t-статистика меньше критического, то средние показатели с заданной вероятностью равны. В нашем случае $|-1,784242592| < 2,492159469$, следовательно, среднее число продаж значимо не отличается. Следует отметить, что если взять уровень значимости $\alpha=0,05$, то результаты исследования будут совсем иными.

Двухвыборочный t-тест с одинаковыми дисперсиями		
	город 1	город 2
Среднее	23,57142857	26,41666667
Дисперсия	17,34065934	15,35606061
Наблюдения	14	12
Объединенная дисперсия	16,43105159	
Гипотетическая разность средних	0	
df	24	
t-статистика	-1,784242592	
P(T<=t) одностороннее	0,043516846	
t критическое одностороннее	2,492159469	
P(T<=t) двухстороннее	0,087033692	
t критическое двухстороннее	2,796939498	

Лабораторная работа №3

ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Цель: Освоить методы построения линейного уравнения парной регрессии с помощью ЭВМ, научиться получать и анализировать основные характеристики регрессионного уравнения.

Рассмотрим методику построения регрессионного уравнения на примере.

ПРИМЕР. Даны выборки факторов x_i и y_i . По этим выборкам найти уравнение линейной регрессии $\hat{y} = ax + b$. Найти коэффициент парной корреляции. Проверить на уровне значимости $\alpha = 0,05$ регрессионную модель на адекватность.

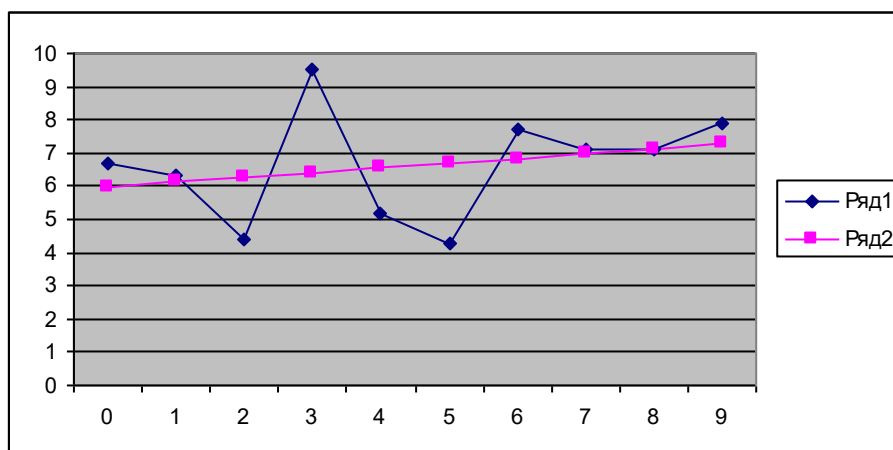
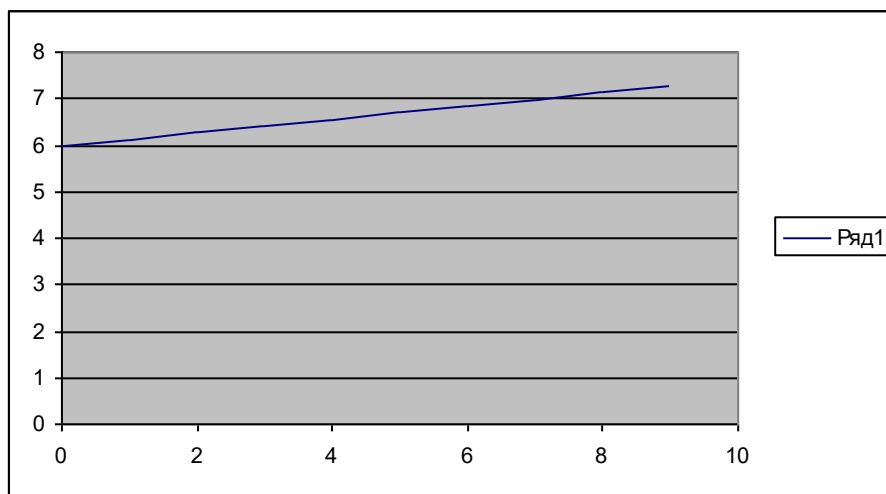
	,7	,3	,4	,5	,2	,3	,7	,1	,1	,9

Для нахождения коэффициентов a и b уравнения регрессии служат функции НАКЛОН и ОТРЕЗОК, категории «Статистические». Вводим в А5 подпись «a=» а в соседнюю ячейку В5 вводим функцию НАКЛОН, ставим курсор в поле «Изн_знач_y» задаем ссылку на ячейки В2-К2, обводя их мышью. Результат 0,14303. Найдем теперь коэффициент b . Вводим в А6 подпись «b=», а в В6 функцию ОТРЕЗОК с теми же параметрами, что и функции НАКЛОН. Результат 5,976364. следовательно, уравнение линейной регрессии есть $y=0,14303x+5,976364$.

Построим график уравнения регрессии. Для этого в третью строчку таблицы введем значения функции в заданных точках X (первая строка) – $y(x_1)$. Для получения этих значений используются функция ТЕНДЕНЦИЯ категории «Статистические». Вводим в А3 подпись «Y(X) и, поместив курсор в В3, вызываем функцию ТЕНДЕНЦИЯ. В полях «Изн_знач_y» и «Изн_знач_x» даем ссылку на В2-К2 и В1-К1. в поле «Нов_знач_x» вводим также ссылку на В1-К1. в поле «Константа» вводят 1, если уравнение регрессии имеет вид $y=ax+b$, и 0, если $y=ax$. В нашем случае вводим единицу. Функция ТЕНДЕНЦИЯ является массивом, поэтому для вывода всех ее значений выделяем область В3-К3 и нажимаем F2 и Ctrl+Shift+Enter. Результат – значения уравнения регрессии в заданных точках. Строим график. Ставим курсор в любую свободную клетку, вызываем мастер диаграмм, выбираем категорию «Точечная», вид графика – линия без точек (в нижнем правом углу), нажимаем «Далее», в поле «Диагноз» вводим ссылку на В3-К3. переходим на закладку «Ряд» и в поле «Значения X» вводим ссылку на В1-К1, нажимаем «Готово». Результат – прямая линия регрессии. Посмотрим, как различаются графики опытных данных и уравнения

регрессии. Для этого ставим курсор в любую свободную ячейку, вызываем мастер диаграмм, категория «График», вид графика – ломанная линия с точками (вторая сверху левая), нажимаем «Далее», в поле «Диапазон» вводим ссылку на вторую и третью строки В2-К3. переходим на закладку «Ряд» и в поле «Подписи оси Х» вводим ссылку на В1-К1, нажимаем «Готово». Результат – две линии (Синяя (ряд 1) – исходные, красная (ряд 2)– уравнение регрессии). Видно, что линии мало различаются между собой.

a=	0,14303
b=	5,976364



Для вычисления коэффициента корреляции r_{xy} служит функция ПИРСОН. Размещаем график так, чтобы они располагались выше 25 строки, и в А25 делаем подпись «Корреляция», в В25 вызываем функцию ПИРСОН, в полях которой «Массив 2» вводим ссылку на исходные данные В1-К1 и В2-К2. результат 0,993821. коэффициент детерминации R_{xy} – это квадрат коэффициента корреляции r_{xy} . В А26 делаем подпись «Детерминация», а в В26 – формулу «=В25*В25». Результат 0,265207.

Однако, в Excel существует одна функция, которая рассчитывает все

основные характеристики линейной регрессии. Это функция ЛИНЕЙН. Ставим курсор в B28 и вызываем функцию ЛИНЕЙН, категории «Статистические». В полях «Изм_знач_y» и «Изм_знач_x» даем ссылку на B2-K2 и B1-K1. поле «Константа» имеет тот же смысл, что и функции ТЕНДЕНЦИЯ, у нас она равна 1. поле «Стат» должно содержать 1, если нужно вывести полную статистику о регрессии. В нашем случае ставим туда единицу. Функция возвращает массив размеров 2 столбца и 5 строк. После ввода выделяем мышью ячейку B28-C32 и нажимаем F2 и Ctrl+Shift+Enter. Результат – таблица значений, числа в которой имеют следующий смысл:

Коэффициент a	Коэффициент b
Стандартная ошибка m_o	Стандартная ошибка m_b
Коэффициент детерминации R_{xy}	Среднеквадратическое отклонение y
F – статистика	Степени свободы $n-2$
Регрессионная сумма квадратов S_n^2	Остаточная сумма квадратов S_n^2

0,14303	5,976364
0,183849	0,981484
0,070335	1,669889
0,60525	8
1,687758	22,30824

Анализ результата: в первой строчке – коэффициенты уравнения регрессии, сравните их с рассчитанными функциями НАКЛОН и ОТРЕЗОК. Вторая строчка – стандартные ошибки коэффициентов. Если одна из них по модулю больше, чем сам коэффициент, то коэффициент считается нулевым. Коэффициент детерминации характеризует качество связи между факторами. Полученное значение 0,070335 говорит об очень хорошей связи факторов, F – статистика проверяет гипотезу о адекватности регрессионной модели. Данное число нужно сравнить с критическим значением, для его получения вводим в E33 подпись «F-критическое», а в F33 функцию ФРАСПОБР, аргументами которой вводим соответственно «0,05» (уровень значимости), «1» (число факторов X) и «8» (степени свободы).

F-критическое	5,317655
---------------	----------

Видно, что F-статистика меньше, чем F-критическое, значит, регрессионная модель не адекватна. В последней строке приведены регрессионная сумма квадратов $S_b^2 = \sum_{i=1}^n (\tilde{y}(x_i) - y)^2$ и остаточные суммы

квадратов $S_b^2 = \sum_{i=1}^n (\tilde{y}(x_i) - y)^2$. Важно, чтобы регрессионная сумма (объясненная регрессией) была намного больше остаточной (не объясненная регрессией, вызванная случайными факторами). В нашем случае это условие не выполняется, что говорит о плохой регрессии.

Вывод: В ходе работы я освоил методы построения линейного уравнения парной регрессии с помощью ЭВМ, научился получать и анализировать основные характеристики регрессионного уравнения.

Лабораторная работа № 4

НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Цель: освоить методы построения основных видов нелинейных уравнений парной регрессии с помощью с помощью ЭВМ (внутренне линейные модели), научиться получать и анализировать показатели качества регрессионных уравнений.

Рассмотрим случай, когда нелинейные модели с помощью преобразования данных можно свести к линейным (внутренне линейные модели).

ПРИМЕР. Построить уравнение регрессии $y = f(x)$ для выборки $x_n y_n$ ($f = 1, 2, \dots, 10$). В качестве $f(x)$ рассмотреть четыре типа функций – линейная, степенная, показательная и гиперболу:

$$y = Ax + B; \quad y = Ax^B; \quad y = Ae^{Bx}; \quad y = A/x + B.$$

Необходимо найти их коэффициенты A и B , и сравнив показатели качества, выбрать функцию, которая наилучшим образом описывает зависимость.

Прибыль Y	0,3	1,2	2,8	5,2	8,1	11,0	16,8	16,9	24,7	29,4
Прибыль X	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50

Введем данные в таблицу вместе с подписями (ячейки A1-K2). Оставим свободными три строчки ниже таблицы для ввода преобразованных данных, выделим первые пять строк, проведя по левой серой границе по числам от 1 до 5 и выбрать какой-либо цвет (светлый – желтый или розовый) раскрасить фон ячеек. Далее, начиная с A6, выводим параметры линейной регрессии. Для этого в ячейку A6 делаем подпись «Линейная» и в соседнюю ячейку B6 вводим функцию ЛИНЕЙН. В полях «Изн_знач_x» даем ссылку на B2-K2 и B1-K1, следующие два поля принимают значения по единице. Далее обводим область ниже в 5 строчек и левее в 2 строки и нажимаем F2 и

Ctrl+Shift+Enter. Результат - таблица с параметрами регрессии, из которых наибольший интерес представляет коэффициент детерминации в первом столбце третий сверху. В нашем случае он равен $R_1 = 0,951262$. Значение F-критерия, позволяющего проверить адекватность модели $F_1 = 156,1439$

(четвертая строка, первый столбец). Уравнение регрессии равно $y = 12,96x + 6,18$ (коэффициенты a и b приведены в ячейках B6 и C6).

Линейная	12,96	-6,18
	1,037152	1,60884
	0,951262	2,355101
	156,1439	8
	866,052	44,372

Определим аналогичные характеристики для других регрессий и в результате сравнения коэффициентов детерминации найдем лучшую регрессионную модель. Рассмотрим гиперболическую регрессию. Для ее получения преобразуем данные. В третьей строке в ячейку A3 введем подпись «1/x» а в ячейку B3 введем формулу «=1/B2». Растянем автозаполнением данную ячейку на область B3-K3. Получим характеристики регрессионной модели. В ячейку A12 введем подпись «Гипербола», а в соседнюю функцию ЛИНЕЙН. В полях «Изв_знач_у» и «Изв_знач_x2» даем ссылку на B1-K1 и преобразованные данные аргумента x – B3-K3, следующие два поля принимают значения по единице. Далее обводим область ниже 5 строчек и левее в 2 строки и нажимаем F2 и Ctrl+Shift+Enter. Получаем таблицу параметров регрессии. Коэффициент детерминации в данном случае равен $R_2 = 0,475661$, что намного хуже, чем в случае линейной регрессии. F-статистика равна $F_2 = 7,257293$. Уравнение регрессии равно $y = -6,25453x^{18,96772}$.

Гипербола	-6,25453	18,96772
	2,321705	3,655951
	0,475661	7,724727
	7,257293	8
	433,0528	477,3712

Рассмотрим экспоненциальную регрессию. Для ее линеаризации получаем уравнение $y = \tilde{a}x + \tilde{b}$, где $\tilde{y} = \ln y$, $\tilde{a} = b$, $\tilde{b} = \ln a$. Видно, что надо сделать преобразование данных – y заменить на $\ln y$. Ставим курсор в ячейку A4 и делаем заголовок «ln y». Ставим курсор в B4 и вводим формулу LN (категория «Математические»). В качестве аргумента делаем ссылку на B1. Автозаполнением распространяем формулу на четвертую строку на ячейки B4-K4. Далее в ячейке F6 задаем подпись «Экспонента» и в соседней G6 вводим функцию ЛИНЕЙН, аргументами которой будут преобразованные данные B4-K4 (в поле «Изв_знач_у»), а остальные поля такие же как и для случая линейной регрессии (B2-K2, 1, 1). Далее обводим ячейки G6-H10 и нажимаем F2 и Ctrl+Shift+Enter. Результат $R_3 = 0,89079$, $F_3 = 65,25304$, что говорит об очень хорошей регрессии. Для нахождения

коэффициентов уравнения регрессии $b = \tilde{a}$; $a = e^{\tilde{b}}$ ставим курсор в J6 и делаем заголовок «a=», а в соседней K6 формулу «=EXP(H6)», в J7 даем заголовок «b=», а в K7 формулу «=G6». Уравнение регрессии есть $y = 0,511707 \cdot e^{6,197909x}$.

Экспонента	1,824212	-0,67	a=	0,511707
	0,225827	0,350304	b=	6,197909
	0,89079	0,512793		
	65,25304	8		
	17,15871	2,103652		

Рассмотрим степенную регрессию. Для ее линеаризации получаем уравнение $\tilde{y} = \tilde{a}\tilde{x} + \tilde{b}$, где $\tilde{y} = \ln y$, $\tilde{x} = \ln x$, $\tilde{a} = b$, $\tilde{b} = \ln a$. Видно, что надо сделать преобразование данных – y заменить на $\ln y$ и x заменить на $\ln x$. Строчка с $\ln y$ у нас уже есть. Преобразуем переменные x . В ячейку A5 даем подпись « $\ln x$ », а в B5 и вводим формулу LN (категория «Математические»). В качестве аргумента делаем ссылку на B2. Автозаполнением распространяем формулу на пятую строку на ячейки B5-K5. Далее в ячейке F12 задаем подпись «Степенная» и в соседней G12 вводим функцию ЛИНЕЙН, аргументами которой будут преобразованные данные B4-K4 (в поле «Изн_знач_y»), и B5-K5 (в поле «Изн_знач_x»), остальные поля – единицы. Далее освободим ячейки G12-H16 и нажимаем F2 и Ctrl+Shift+Enter. Результат $R_4 = 0,997716$, $F_4 = 3494,117$, что говорит об хорошей регрессии. Для нахождения коэффициентов уравнения регрессии $b = \tilde{a}$; $a = e^{\tilde{b}}$ ставим курсор в J12 и делаем заголовок «a=», а в соседней K12 формулу «=EXP(H12)», в J13 даем заголовок «b=», а в K13 формулу «=G12». Уравнение регрессии есть $y = 4,90767/x + 7,341268$.

Степенная	1,993512	1,590799	a=	4,90767
	0,033725	0,023823	b=	7,341268
	0,997716	0,074163		
	3494,117	8		
	19,21836	0,044002		

Проверим, все ли уравнения адекватно описывают данные. Для этого нужно сравнить F-статистики каждого критерия с критическим значением. Для его получения вводим в A21 подпись «F-критическое», а в B21 функцию ФРАСПОБР, аргументами которой вводим соответственно «0,05» (уровень значимости), «1» (число факторов X в строке «Уровень значимости 1») и «8» (степень свободы $2 = n - 2$). Результат 5,317655. F – критическое больше F – статистики значит модель адекватна. Также адекватны и остальные регрессии. Для того, чтобы определить, какая модель наилучшим образом описывает данные, сравним индексы детерминации для каждой модели R_1 , R_2 , R_3 , R_4 . Наибольшим является $R_4 = 0,997716$. Значит опытные данные лучше описывать $y = 4,90767/x + 7,341268$.

Вывод: В ходе работы я освоил методы построения основных видов

нелинейных уравнений парной регрессии с помощью с помощью ЭВМ (внутренне линейные модели), научился получать и анализировать показатели качества регрессионных уравнений.

Y	0,3	1,2	2,8	5,2	8,1	11	16,8	16,9	24,7	29,4
X	0,25	0,5	0,75	1	1,25	1,5	1,75	2	2,25	2,5
1/x	4	2	1,333 333	1	0,8	0,666 667	0,571 429	0,5	0,444 444	0,4
ln y	- 1,203 97	0,1823 22	1,029 619	1,648 659	2,09186 41	2,397 895	2,821 379	2,827 314	3,206 803	3,380 995
ln x	- 1,386 29	- 0,6931 5	- 0,287 68	0	0,22314 36	0,405 465	0,559 616	0,693 147	0,810 93	0,916 291
Линей ная	12,96	-6,18			Экспон ента	1,824 212	-0,67		a=	0,511 707
	1,037 152	1,6088 4				0,225 827	0,350 304		b=	6,197 909
	0,951 262	2,3551 01				0,890 79	0,512 793			
	156,1 439	8				65,25 304	8			
	866,0 52	44,372				17,15 871	2,103 652			
Гипер бола	- 6,254 53	18,967 72			Степен ная	1,993 512	1,590 799		a=	4,907 67
	2,321 705	3,6559 51				0,033 725	0,023 823		b=	7,341 268
	0,475 661	7,7247 27				0,997 716	0,074 163			
	7,257 293	8				3494, 117	8			
	433,0 528	477,37 12				19,21 836	0,044 002			
F - критическое		5,317655								

Лабораторная работа № 5

ПОЛИНОМИНАЛЬНАЯ РЕГРЕССИЯ

Цель: По опытным данным построить уравнение регрессии вида $y = ax^2 + bx + c$.

ХОД РАБОТЫ:

Рассматривается зависимость урожайности некоторой культуры y_i от количества внесенных в почву минеральных удобрений x_i . Предполагается, что эта зависимость квадратичная. Необходимо найти уравнение регрессии вида $\tilde{y} = ax^2 + bx + c$.

x	0	1	2	3	4	5	6	7	8	9
y	29,8	58,8	72,2	101,5	141	135,1	156,6	181,7	216,6	208,2

Введем эти данные в электронную таблицу вместе с подписями в ячейки A1-K2. Построим график. Для этого обведем данные Y (ячейки B2-K2), вызываем мастер диаграмм, выбираем тип диаграммы «График», вид диаграммы – график с точками (второй сверху левый), нажимаем «Далее», переходим на закладку «Ряд» и в поле «Подписи оси X» делаем ссылку на B2-K2, нажимаем «Готово». График можно приблизить полиномом 2 степени $y = ax^2 + bx + c$. Для нахождения коэффициентов a , b , c нужно решить систему уравнений:

$$\begin{cases} a\sum x^4 + b\sum x^3 + c\sum x^2 = \sum x^2 y, \\ a\sum x^3 + b\sum x^2 + c\sum x = \sum xy, \\ a\sum x^2 + b\sum x + cn = \sum y \end{cases}$$

Рассчитаем суммы. Для этого в ячейку A3 вводим подпись «X^2», а в B3 вводим формулу «=B1*B1» и Автозаполнением переносим ее на всю строку B3-K3. В ячейку A4 вводим подпись «X^3», а в B4 формулу «=B1*B3» и Автозаполнением переносим ее на всю строку B4-K4. В ячейку A5 вводим «X^4», а в B5 формулу «=B4*B1», автозаполняем строку. В ячейку A6 вводим «X*Y», а в B8 формулу «=B2*B1», автозаполняем строку. В ячейку A7 вводим «X^2*Y», а в B9 формулу «=B3*B2», автозаполняем строку. Теперь считаем суммы. Выделяем другим цветом столбец L, щелкнув по заголовку и выбрав цвет. В ячейку L1 помещаем курсор и щелкнув по кнопке автосуммы со

значком Σ , вычисляем сумму первой строки. Автозаполнением переносим формулу на ячейки L1-710.

Решаем теперь систему уравнений. Для этого вводим основную матрицу системы. В ячейку A13 вводим подпись «A=», а в ячейки матрицы B13-D15 вводим ссылки, отраженные в таблице

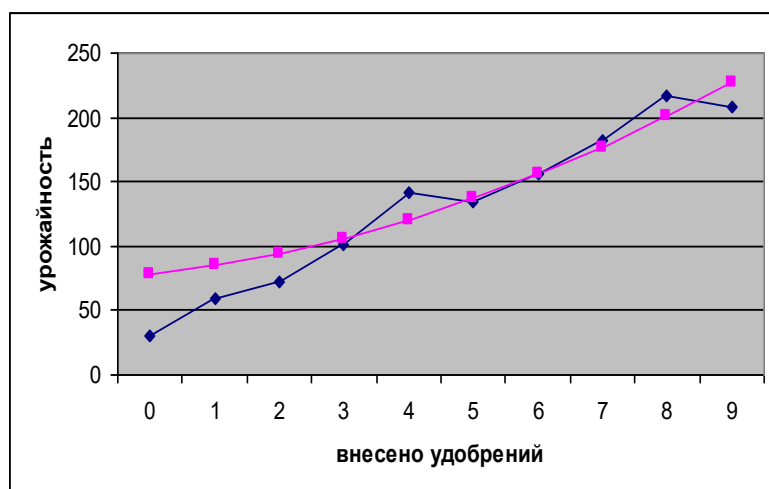
	B	C	D
13	=L5	=L4	=L3
14	=L3	=L2	=L1
15	=L2	=L1	=9

Вводим также правые части системы уравнений. В G13 вводим подпись «B=», а в H13-H15 вводим, соответственно ссылки на ячейки «=L7», «=L6», «=L2». Решаем систему матричным методом. Из высшей математики известно, что решение равно $A^{-1}B$. Находим обратную матрицу. Для этого в ячейку J13 вводим подпись «A обр.» и, поставив курсор в K13 задаем формулу МОБР (категория «Математические»). В качестве аргумента «Массив» даем ссылку на ячейки B13:D15. Результатом также должна быть матрица размером 4×4. Для ее получения обводим ячейки K13-M15 мышью, выделяя их и нажимаем F2 и Ctrl+Shift+Enter. Результат – матрица A^{-1} . Найдем теперь произведение этой матрицы на столбец B (ячейки H13-H15). Вводим в ячейку A18 подпись «Коэффициенты» и в B18 задаем функцию МУМНОЖ (категория «Математические»). Аргументами функции «Массив 1» служит ссылка на матрицу A^{-1} (ячейки K13-M15), а в поле «Массив 2» даем ссылку на столбец B (ячейки H13-H16). Далее выделяем B18-B20 и нажимаем F2 и Ctrl+Shift+Enter. Получившийся массив – коэффициенты уравнения регрессии a , b , c . В результате получаем уравнение регрессии вида: $y = 1,201082x^2 - 5,619177x + 78,48095$.

Построим графики исходных данных и полученных на основе уравнения регрессии. Для этого в ячейку A8 вводим подпись «Регрессия» и в B8 вводим формулу «=\$B\$18*B3+\$B\$19*B1+\$B\$20». Автозаполнением переносим формулу в ячейки B8-K8. Для построения графика выделяем ячейки B8-K8 и, удерживая клавишу Ctrl, выделяем также ячейки B2-M2. Вызываем мастера диаграмм, выбираем тип диаграммы «График», вид диаграммы – график с точками (второй сверху левый), нажимаем «Далее», переходим на закладку «Ряд» и в поле «Подписи оси X» делаем ссылку на B2-M2, нажимаем «Готово». Видно, что кривые почти совпадают.

ВЫВОД: в процессе работы я по опытным данным научился строить уравнение регрессии вида $y = ax^2 + bx + c$.

[illegible]



5. Образовательные технологии

В ходе изучения дисциплины предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий:

- во время лекционных занятий используется презентация с применением слайдов с графическим и табличным материалом, что повышает наглядность и информативность используемого теоретического материала;
- практические занятия предусматривают использование групповой формы обучения, которая позволяет студентам эффективно взаимодействовать в микрогруппах при обсуждении теоретического материала;
- подготовка рефератов и докладов по самостоятельной работе студентов и выступление с докладом перед аудиторией, что способствует формированию навыков устного выступления по изучаемой теме и активизирует познавательную активность студентов.

При проведении занятий кроме указанных средств используются также интернет ресурсы.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Содержание самостоятельной работы студентов

В процессе самостоятельной работы над каждой темой студентом должны осуществляться следующие виды деятельности:

- Проработка учебного материала по конспектам лекций, основной и рекомендуемой учебной литературе.
- Работа над домашними заданиями
- Работа над вопросами и заданиями для самоподготовки, представленными.
- Написание рефератов.
- Работа с тестами.
- Поиск и обзор научных публикаций и электронных источников информации.

Темы, виды и содержание самостоятельной работы по дисциплине

Темы	Виды и содержание самостоятельной работы	Форма контроля
Модуль 1. Обработка данных в рамках линейной регрессионной модели		
1.Массивы данных в прикладных задачах. Статистический и детерминированный подходы. Постановка задачи обработки данных.	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос
2.Понятие функции распределения случайной величины и ее числовых характеристик. Точечные и интервальные оценки массива данных	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Подготовка к	Устный опрос, тестирование

	лабораторному занятию по теме, составление конспекта.	
3.Регрессии, аппроксимирующие функции.	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 3. Аналитический разбор и конспектирование источников по данной теме.	Устный опрос, тестирование
Модуль 2. Факторный и кластерный анализ		
4.Выявление влияющих факторов. Элементы многофакторного анализа.	1. Проработка конспекта лекций. 2. Поиск и анализ дополнительной литературы.	Устный опрос
5.Кластерный анализ. Методы иерархической классификации.	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Подготовить научный доклад по теме.	Устный опрос
6.Многомерное шкалирование.	1. Проработка конспекта лекций, изучение учебной и научной литературы и интернет ресурсов; 2. Разработать электронную презентацию	Устный опрос

Целью подготовки реферата является приобретение навыков творческого обобщения и анализа имеющейся литературы по рассматриваемым вопросам, что обычно является первым этапом самостоятельной работы. По каждому

модулю предусмотрены написание и защита одного реферата. Всего по дисциплине студент может представить шесть рефератов. Тему реферата студент выбирает самостоятельно из предложенной тематики. При написании реферата надо составить краткий план, с указанием основных вопросов избранной темы. Реферат должен включать введение, несколько вопросов, посвященных рассмотрению темы, заключение и список использованной литературы. В вводной части реферата следует указать основания, послужившие причиной выбора данной темы, отметить актуальность рассматриваемых в реферате вопросов. В основном разделе излагаются наиболее существенные сведения по теме, производится их анализ, отмечаются отдельные недостатки или нерешенные еще вопросы и т.д. В заключении реферата на основании изучения литературных источников должны быть сформулированы краткие выводы и предложения. Список литературы оформляется в соответствии с требованиями ГОСТ 7.1-84 «Библиографическое описание документа». Перечень литературы составляется в алфавитном порядке фамилий первых авторов, со сквозной нумерацией. Примерный объем реферата 15-20 страниц.

Предусмотрено проведение индивидуальной работы (консультаций) со студентами в ходе изучения материала данной дисциплины.

Примерные вопросы для самостоятельного изучения (рефераты)

1. Классификация методов статистического анализа социологической информации.

2. Подготовка баз данных к статистическому анализу

Анализ различий: обоснование выбора метода оценки

3. Реализация t-теста SPSS

4. Реализация однофакторного дисперсионного анализа

5. Регрессионный анализ, общая характеристика метода

6. Реализация линейной регрессии в SPSS и Vortex

7. Факторный анализ – общая характеристика метода.

8. Реализация факторного анализа в SPSS

9. Кластерный анализ – общая характеристика метода.

10. Реализация кластерного анализа в SPSS и Vortex

11. Многомерное шкалирование – общая характеристика метода.

12. Реализация многомерного шкалирования в SPSS

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

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

Перечень компетенций с указанием этапов их формирования приведен в описании образовательной программы.

Код и наименование компетенции из ФГОС ВО	Код и наименование индикатора достижения компетенций (в соответствии с ПООП (при наличии))	Планируемые результаты обучения	Процедура освоения
ОК-1. Способность к абстрактному мышлению, анализу, синтезу.		Знает: основные формулы основы математической статистики, сбора, обработки и анализа статистических данных. Умеет: на основе применения аппарата математической статистики принимать нужные решения. Строить модели различных прикладных задач и перекладывать их на ЭВМ. Владеет: навыками дискуссии по профессиональной тематике	Устный опрос, написание рефератов. Выполнение лабораторной работы
ОПК-3 Способность самостоятельно приобретать с помощью информационных технологий и использовать в практической		Знает: основные методы обработки статданных. Особое внимание при разборе материала уделяется анализу ошибок, которые обычно делают начинающие исследователи при применении того или иного	Устный опрос, написание рефератов. Выполнение лабораторной работы

деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение		метода статистической обработки Умеет: строить математические модели различных прикладных задач; программировать их для ЭВМ, проводить численное моделирование и по результатам этого находить новые закономерности и, если нужно, менять прежние модели. Владеет: статистическими пакетами для обработки и анализа экспериментальных данных	
ПК-1. Способность проводить научные исследования и получать новые прикладные результаты самостоятельно и в составе научного коллектива.		Знает: - современные тенденции развития научных и прикладных достижений и их использование в прикладном исследовании; - подходы использования современных методов для решения научных и практических задач. Умеет: самостоятельно разбираться в современных методах и способах представления данных Владеет: навыками поиска информации, методами сбора информации и навыками применения набора стандартных методов статистической обработки данных.	Устный опрос, написание рефератов.

7.2. Типовые контрольные задания

Формы контроля при изучении дисциплины «Статистические методы обработки данных»: текущий контроль, промежуточный контроль по модулю, итоговая аттестация по дисциплине.

По результатам текущего и промежуточного контроля составляется академический рейтинг магистранта по каждому модулю и выводится средний рейтинг по всем трем модулям.

Текущий контроль – все виды аудиторной и внеаудиторной работы магистрантов по данному дисциплинарному модулю, результаты которой оцениваются до промежуточного контроля.

Промежуточный контроль – это проверка полноты знаний по освоенному материалу дисциплинарного модуля.

Итоговая аттестация – это подведение итогов текущей работы и промежуточных контролей по дисциплинарным модулям.

По результатам итоговой аттестации магистранту засчитывается трудоемкость дисциплины в дисциплинарных модулях, выставляется дифференцированная отметка в принятой системе баллов, характеризующая качество освоения студентом знаний, умений и навыков по данной дисциплине.

Дисциплинарный модуль (ДМ) – относительная часть учебной дисциплины, по окончании изучения которой осуществляется промежуточный контроль знаний магистрантов. Количество дисциплинарных модулей определяется в зависимости от содержания и трудоемкости дисциплины.

Текущий контроль включает оценку:

- посещения занятий
 - активного участия на семинарских занятиях
 - текущего контрольного тестирования
- написания, оформления и защиты рефератов (докладов)

Задания для контроля самостоятельной работы

1. Сравнение средних.
2. Рассчитать средний балл, в разных населенных пунктах (переменная «Населенный пункт»).
3. Обосновать выбор метода проверки различий средних: t-тесты или дисперсионный анализ? И почему?
4. Проверить наличие статистически значимых различий среднего уровня дохода в разных должностных группах. Интерпретировать полученные данные.
5. Проиллюстрируйте результат с помощью графика «Столбики ошибок».
6. Линейная регрессия.

7. Вычислите зависимость продолжительности жизни женщин в разных странах (переменная «Average male life expectancy») от калорийности питания (переменная Daily calorie intake).
8. Проинтерпретируйте полученные данные.
9. Каков объяснительный потенциал полученной регрессионной модели?
10. Запишите уравнение регрессии.
11. Проиллюстрируйте результат с помощью графика (на графике обязательно должна быть линия регрессии).
12. Кластерный анализ.
13. Осуществить кластеризацию массива на основании переменных (метод кластеризации – k-means, количество кластеров - 3). Не забудьте вариант «затрудняюсь ответить» закодировать как пропущенное значение!
14. Определить наполненность кластеров.
15. Проанализировать различия между кластерами на основании тех переменных, которые были использованы для кластеризации.
16. Придумать кластерам названия.
17. Задания выполняются индивидуально каждым студентом.
18. Результаты работы должны быть оформлены в формате Word.
19. Работа должна содержать все необходимые для аргументации таблицы с показателями и подробное объяснение того, что дает нам тот или иной показатель, какие выводы позволяет сделать относительно полученных данных.

Вопросы к экзамену

1. Выборочный метод в статистике. Эмпирическая функция распределения выборки

2. Выборочное среднее квадратическое отклонение. Выборочные начальные и центральные моменты. Асимметрия. Эксцесс
3. Понятие о точечной оценке числовой характеристике случайной величины и свойства точечной оценки. Методы получения точечных оценок
4. Доверительные интервалы для оценки числовых характеристик.
5. Метод произведения для вычисления выборочной средней и дисперсии.
6. Функциональная и корреляционная зависимости
7. Определение формы связи. Понятие регрессии
8. Выборочный коэффициент корреляции. Методика вычисления выборочного коэффициента корреляции
9. Проверка статистических гипотез.
10. Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны.
11. Массивы данных как основной метод получения количественно информации о технических объектах с изменяемыми характеристиками
12. Задачи обработки массивов данных в зависимости от их природы (одномерные массивы, массивы функций двух и более переменных, интегральные зависимости связи между влияющими факторами, выявление различия/согласия различных массивов).
13. Примеры массивов данных в задачах техногенной безопасности (загрязняющие вещества, воздействия ЭМП на население, многомерные задачи расчетов рисков и т.п.).
14. Функции распределения случайной величины как основная форма ее представления.
15. Математическое ожидание и дисперсия распределения случайной величины.
16. Используемые инструменты Mathcad. Машинные формы массивов данных (выборок). Способы вычисления оценок математического ожидания и дисперсии.
17. Линейная регрессия для массива двух переменных.
18. Метод наименьших квадратов и его машинная реализация.
19. Однофакторный дисперсионный анализ.
20. Двухфакторный дисперсионный анализ.

7.3. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля - 50% и промежуточного контроля - 50%.

Текущий контроль по дисциплине включает:

- посещение занятий - 10 баллов,
- участие на практических занятиях - 50 баллов,
- выполнение домашних (аудиторных) контрольных и лабораторных работ – 40 баллов.

Промежуточный контроль по дисциплине включает:

- зачет - 100 баллов.

Критерии оценки знаний студентов

100 баллов – студент показал глубокие и систематизированные знания учебного материала по теме; глубоко усвоил учебную литературу; хорошо знаком с научной литературой; активно использовал материалы из первоисточников; цитировал различных авторов; принимал активное участие в обсуждении узловых вопросов на всём протяжении семинарского занятия; умеет глубоко и всесторонне анализировать те или иные исторические события; в совершенстве владеет соответствующей терминологией; материал излагает чётко и лингвистически грамотно; отличается способностью давать собственные оценки, делать выводы, проводить параллели и самостоятельно рассуждать.

90 баллов – студент показал полные знания учебно-программного материала по теме; хорошо усвоил учебную литературу; знаком с научной литературой; использовал материалы из первоисточников; цитировал различных авторов; принимал активное участие в обсуждении узловых вопросов; проявил способность к научному анализу материала; хорошо владеет соответствующей терминологией; материал излагается последовательно и логично; отличается способностью давать собственные оценки, делать выводы, рассуждать; показал высокий уровень исполнения заданий, но допускает отдельные неточности общего характера.

80 баллов – студент показал достаточно полное знание учебно-программного материала; усвоил основную литературу, рекомендованную программой; владеет методом комплексного анализа; показал способность аргументировать свою точку зрения с использованием материала из первоисточников; правильно ответил практически на все вопросы преподавателя в рамках обсуждаемой темы; систематически участвовал в групповых обсуждениях; не допускал в ответе существенных неточностей.

70 баллов – студент показал достаточно полное знание учебного материала, не допускал в ответе существенных неточностей, активно работал на семинарском занятии, показал систематический характер знаний по дисциплине, цитирует первоисточники, но не может теоретически обосновать некоторые выводы.

60 баллов – студент обладает хорошими знаниями по всем вопросам темы семинарского занятия, не допускал в ответе существенных неточностей, самостоятельно выполнил основные предусмотренные программой задания, усвоил основную литературу, отличается достаточной активностью на семинарском занятии; умеет делать выводы без существенных ошибок, но при этом не дан анализ информации из первоисточников.

50 баллов – студент усвоил лишь часть программного материала, вместе с тем ответ его стилистически грамотный, умеет логически рассуждать; допустил одну существенную или несколько несущественных ошибок; знает терминологию; умеет делать выводы и проводить некоторые параллели.

40 баллов – студент знает лишь часть программного материала, не отличался активностью на семинарском занятии; усвоил не всю основную литературу,

рекомендованную программой; нет систематического и последовательного изложения материала; в ответах допустил достаточное количество несущественных ошибок в определении понятий и категорий, дат и т.п.; умеет делать выводы без существенных ошибок; наличие грамматических и стилистических ошибок и др.

30 баллов – студент имеет недостаточно полный объём знаний в рамках образовательного стандарта; знает лишь отдельные вопросы темы, кроме того допускает серьёзные ошибки и неточности; наличие в ответе стилистических и логических ошибок.

20 баллов – у студента лишь фрагментарные знания или отсутствие знаний по значительной части заданной темы; не знает основную литературу; не принимал участия в обсуждении вопросов по теме семинарского занятия; допускал существенные ошибки при ответе; студент не умеет использовать научную терминологию дисциплины; наличие в ответе стилистических и логических ошибок.

10 балл — отсутствие знаний по теме или отказ от ответа.

Шкала диапазона для перевода рейтингового балла по дисциплине с учётом итогового контроля в «5»- балльную систему.

0 – 50 баллов – «неудовлетворительно»;

51 – 65 баллов – «удовлетворительно»;

66 – 85 баллов – «хорошо»;

86 – 100 баллов – «отлично».

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.

Основная литература

1. Шорохова И.С. Статистические методы анализа [Электронный ресурс] : учебное пособие / И.С. Шорохова, И.В. Кисляк, О.С. Мариев. — Электрон. текстовые данные. — Екатеринбург: Уральский федеральный университет, ЭБС АСВ, 2015. — 300 с. — 978-5-7996-1633-5. — Режим доступа: <http://www.iprbookshop.ru/65987.html> (12.05.18)
2. Новикова, Е.Н. Компьютерная обработка результатов измерений : учебное пособие / Е.Н. Новикова, О.Л. Серветник ; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь : СКФУ, 2017. - 182 с. : ил. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=483751> (16.05.2018).
3. Горяинова, Е.Р. Прикладные методы анализа статистических данных : учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. - Москва : Издательский дом Высшей школы экономики, 2012. - 312 с. - ISBN 978-5-7598-0866-4 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=227280> (16.05.2018).

2. А.А.Боровков Математическая статистика. Учебник. 4-е издание Санкт-Петербург, Лань, 2010, 704 с
3. А.Н. Бородин Элементарный курс теории вероятностей и математической статистики: Учебное пособие. 7-е изд. Санкт-Петербург, Лань, 2010, 256с
4. В.Н. Калинина, В. Ф. Панкин. Математическая статистика. 2-е издание, стереотипное, М: Высшая школа, 1998, 336 с.

Дополнительная литература

1. Новиков, Д.А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д.А. Новиков, В.В. Новочадов. - Волгоград : ВолГМУ, 2005. - 84 с. - ISBN 5-9652-0011-0 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=82775> (16.05.2018).
3. Айвазян С.А. Многомерные статистические методы [Электронный ресурс] : учебное пособие / С.А. Айвазян, В.С. Мхитарян, В.А. Зехин. — Электрон. текстовые данные. — М. : Евразийский открытый институт, Московский государственный университет экономики, статистики и информатики, 2003. — 77 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/10803.html> (16.05.2018).
3. Крянев, А.В. Математические методы обработки неопределенных данных / А.В. Крянев, Г.В. Лукин. - Москва : Физматлит, 2006. - 281 с. - ISBN 978-5-9221-0724-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=68359> (16.05.2018).
4. Харин Ю.С., Малюгин В.И., Кирлица В.П. и др. Основы имитационного и статистического моделирования. – Минск: Дизайн ПРО, 1997. – 288 с.
5. Дж. Бендат, А. Пирсол. Прикладной анализ случайных данных. М.:Мир, 1989, 540 с.
6. Козлов, А.Ю. Статистический анализ данных в MS Excel : учеб. пособие для студентов вузов / А. Ю. Козлов, В. С. Мхитарян. - М. : ИНФРА-М, 2014. - 320 с. - (Высшее образование: серия основана в 1996 г.). - ISBN 978-5-16-004579-5 : 375-00.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

1. <http://elib.dgu.ru/?q=node/876> - Научная библиотека ДГУ
2. <http://www.book.ru> - Электронная система BOOK.RU
3. <http://www.iprbookshop.ru> - Электронно-библиотечная система IPRBOOKSHOP
4. <http://ibooks.ru> - Электронно-библиотечная система IBOOKS.RU
5. <http://www.biblio-online.ru> - Издательство «Юрайт»
6. <http://books.google.com> - Интернет каталог общемирового книжного фонда Google Books

10. Методические указания для обучающихся по освоению дисциплины.

Перечень учебно-методических изданий, рекомендуемых студентам, для подготовки к занятиям представлен в разделе «Учебно-методическое обеспечение. Литература».

Для успешного освоения курса студентам рекомендуется проводить самостоятельный разбор материалов семинарских занятий в течении семестра. В случае затруднений в понимании и освоении каких-либо тем решать дополнительные задания из учебных пособий, рекомендуемых к данному курсу.

Такие виды учебно-познавательной деятельности студента как лекции, семинарские, лабораторные занятия и самостоятельная работа составляют систему вузовского образования.

Лекция является главным звеном дидактического цикла обучения в отечественной высшей школе. Несмотря на развитие современных технологий и появление новых методик обучения лекция остаётся основной формой учебного процесса. Она представляет собой последовательное и систематическое изложение учебного материала, разбор какой-либо узловой проблемы. Вузовская лекция ориентирована на формирование у студентов информативной основы для последующего глубокого усвоения материала методом самостоятельной работы, призвана помочь студенту сформировать собственный взгляд на ту или иную проблему.

При изучении дисциплины рекомендуется рейтинговая технология обучения, которая позволяет реализовать комплексную систему оценивания учебных достижений студентов. Текущие оценки усредняются на протяжении семестра при изучении модулей. Комплексность означает учет всех форм учебной и творческой работы студента в течение семестра.

Рейтинговый балл студента на каждом занятии зависит от его инициативности, качества выполненной работы, аргументированности выступления, характера использованного материала и т.д. Уровень усвоения материала напрямую зависит от внеаудиторной самостоятельной работы, которая традиционно такие формы деятельности, как выполнение письменного домашнего задания, подготовка к разбору ранее прослушанного лекционного материала, подготовка доклада и выполнение реферата.

Для успешного освоения курса студентам рекомендуется проводить самостоятельный разбор материалов семинарских занятий в течении семестра. В случае затруднений в понимании и освоении каких-либо тем решать дополнительные задания из учебных пособий, рекомендуемых к данному курсу.

Рекомендуется самостоятельно повторять материал, пройденный во время лекций с подробным разбором.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Информационные средства обучения: электронные учебники, презентации, технические средства предъявления информации (многофункциональный мультимедийный комплекс) и контроля знаний (тестовые системы). Электронные ресурсы Научной библиотеки ДГУ. Электронно-образовательные ресурсы Дагестанского государственного университета.

Для успешного освоения дисциплины, обучающийся использует следующие программные средства: WINDOWS XP, пакет MS OFFICE.

Для успешного освоения дисциплины, обучающийся использует следующие программные средства: пакеты для решения задач Mathcad, MS Office.

12. Описание материально-технической базы, необходимой для

осуществления образовательного процесса по дисциплине.

Реализация учебной дисциплины требует наличия типовой учебной аудитории с возможностью подключения технических средств: аудиовизуальных, компьютерных и телекоммуникационных, компьютерные классы факультета и ИВЦ ДГУ. В университете имеется пакет необходимого лицензионного программного обеспечения.

При кафедре прикладной математики имеется студенческая научно-исследовательская лаборатория «Математическое моделирование».