

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет математики и компьютерных наук
Кафедра прикладной математики

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
**«Статистический анализ и обработка данных с
использованием компьютерных технологий»**

Кафедра прикладной математики
факультета математики и компьютерных наук

Образовательная программа бакалавриата
01.03.05 – Статистика

Направленность (профиль) программы
Анализ больших данных

Форма обучения
Очная

Статус дисциплины: ***Входит в часть, формируемая участниками
образовательных отношений, дисциплина по выбору***

Махачкала, 2023

Фонд оценочных средств по дисциплине «Статистический анализ и обработка данных с использованием компьютерных технологий» составлена в 2023 году в соответствии с требованиями ФГОС ВО бакалавриата по направлению подготовки 01.03.05 - статистика от 14.08.2020 г. № 1032

Разработчики:

1. кафедра прикладной математики, Бейбалаев В.Д. к.ф.-м. н., доцент;

Фонд оценочных средств по дисциплине «Информационные системы и технологии в статистике» одобрен: на заседании кафедры _____ от «20» 01 2023г., протокол № 5

Зав. кафедрой К _____ Кадиев Р.И.

на заседании Методической комиссии № 11 факультета от «25» января 20__ г., протокол № 4.

Председатель М.К. Ризаев М.К.

Фонд оценочных средств «Информационные системы и технологии в статистике» согласован с учебно-методическим управлением

«20» февраля 2023г. _____

Рецензент (эксперт):
доц. Исхр. Ду и ФА
Исхр. Ду и ФА
Исхр. Ду и ФА
(полное наименование организации
и должности руководителя)
М.П.

И.О.
(подпись)

Фамилия И.О.

Активация W
Чтобы активировать

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине

«Информационные системы и технологии в статистике»

1.1. Основные сведения о дисциплине

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов		
	6 семестр	7 семестр	всего
Общая трудоёмкость	108		108
Контактная работа:	36		36
Лекции (Л)	16		16
Лабораторные занятия (ЛЗ)	16		16
Консультации			
Промежуточная аттестация (зачет, экзамен)	ЭКЗАМЕН		
Самостоятельная работа			
1. работа с лекционным материалом, с учебной литературой	10		10
2. опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	10		10
3. подготовка к лабораторным работам	20		20
4. подготовка к контрольным работам, коллоквиумам, экзамену	36		36

1.2. Требования к результатам обучения по дисциплине, формы их контроля и виды оценочных средств

*ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Информационные системы и технологии в статистике»*

№ п/п	Контролируемые модули, разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Оценочные средства		Способ контроля
			наименование	№№ заданий	
1	Модуль 1. Основы анализа данных с помощью компьютерных технологий	ПК-4 ПК-8	Вопросы для собеседования	1-30	устно
		ПК-4 ПК-8	Контрольные работы	1-2	письменно
		ПК-4 ПК-8	Лабораторные	1-4	письменно

			работы		
2	Модуль 2. Статистическая классификация и анализ временных рядов	ПК-4 ПК-8	Вопросы для собеседования	31-50	устно
ПК-4 ПК-8		Контрольные работы	3-4	письменно	
ПК-4 ПК-8		Лабораторные работы	5-7	письменно	

1.3. Показатели и критерии определения уровня сформированности компетенций

№ п/п	Код компетенции	Уровни сформированности компетенции			
		Недостаточный	Удовлетворительный (достаточный)	Базовый	Повышенный
		Отсутствие признаков удовлетворительного уровня	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:	Знать: Уметь: Владеть:
1	ОПК-4	Не знает на достаточном уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Не умеет на достаточном уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Не владеет на достаточном уровне навыками планирования и проведения аналитических работ, в том числе с применением технологий больших данных.	Знает на достаточном уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Умеет на достаточном уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Владеет на достаточном уровне навыками планирования и проведения аналитических работ, в том числе с применением технологий больших данных.	Знает на хорошем уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Умеет на хорошем уровне планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Владеет на хорошем уровне навыками планирования и проведения аналитических работ, в том числе с применением технологий больших данных.	Знает в совершенстве планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Умеет в совершенстве планировать и проводить аналитические работы, в том числе с применением технологий больших данных. Владеет в совершенстве навыками планирования и проведения аналитических работ, в том числе с применением технологий больших данных.
2	ПК-2	Не знает на	Знает на	Знает на	Знает в

		<p>достаточном уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Не умеет на достаточном уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Не владеет на достаточном уровне навыками поиска статистической информации, ее первичной обработки и подготовки для проведения аналитических исследований, в том числе с использованием больших данных.</p>	<p>достаточном уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Умеет на достаточном уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Владеет на достаточном уровне навыками поиска статистической информации, ее первичной обработки и подготовки для проведения аналитических исследований, в том числе с использованием больших данных.</p>	<p>хорошем уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Умеет на хорошем уровне осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Владеет на хорошем уровне навыками поиска статистической информации, ее первичной обработки и подготовки для проведения аналитических исследований, в том числе с использованием больших данных.</p>	<p>совершенстве осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Умеет в совершенстве осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием больших данных. Владеет в совершенстве навыками поиска статистической информации, ее первичной обработки и подготовки для проведения аналитических исследований, в том числе с использованием больших данных.</p>
--	--	--	--	--	--

2. КОНТРОЛЬНЫЕ ЗАДАНИЯ И ИНЫЕ МАТЕРИАЛЫ

**ОЦЕНКИ знаний, умений, навыков и (или) опыта
деятельности,
характеризующие этапы формирования компетенций в процессе
освоения дисциплины «Информационные системы и технологии в
статистике»**

Контрольная работа 1

1. Заполнить анкету «Изучение структуры свободного времени студентов» по организации сво-бодного времени (приложение 1).
2. Скопировать файл с базой данных Access и файл с таблицей “Эксперт”
3. Переименовать файл с базой данных, указав в названии файла свою фамилию.
4. Ввести данные анкеты в базу данных (приложение 2).
5. Экспортировать данные из базы данных Access в файл EXCEL. Указать в названии файла EXCEL свою фамилию.
6. Заполнить таблицу экспертных оценок “Эксперт”. Изменить название файла “Эксперт” на “Эксперт - ФАМИЛИЯ”.

	А	В	С
1	Таблица "ЭКСПЕРТ"		
2		Сколько времени в часах, на ваш взгляд, в среднем уходит на	Ф.И.О.
3	1	на одно посещение кинотеатра	
4	2	на одно посещение кафе в вечернее время	
5	3	на одно посещение ресторана/ночного клуба	
6	4	на одно посещение концерта	
7	5	на одно посещение музея, выставки, театра	
8	6	на одно посещение спортивного состязания	
9	7	на одно посещение молодежной компании (вечеринки)	
10	8	на одну поездку к родным (вопрос только для иногородних студентов)	
11	9	на одну поездку на отдых за городом	
12	10	на одно посещение парикмахерской	
13	11	на одно посещение салонов красоты (вопрос только для студенток)	

5. Разместить три файла в единую папку, отведенную для студентов всей группы.

Контрольная работа № 2

Предварительный анализ - одномерных признаков

1. Смоделировать данные по схеме:
 X_1, X_2 - случайные числа (функция СЛЧИС)
 X_3 - равномерное распределение с параметрами (a,b) - получить путём преобразования случайных чисел (a=5,b=15).

X_4, X_5 - стандартное нормальное распределение получить путём преобразования случайных чисел.

X_6, X_7 - нормальное распределение с параметрами $(m_1, s_1), (m_2, s_2)$.

Среднее признака X_6 $m_1=5$, стандартное отклонение признака X_6 $s_1=2$.

Среднее признака X_7 $m_2=15$, стандартное отклонение признака X_7 $s_2=4$

X_8 - нормальное распределение с параметрами (m_3, s_3) получить с помощью генератора EXCEL.

Среднее признака X_8 $m_3=10$, стандартное отклонение признака X_8 $s_3=3$.

2. По таблице данных рассчитать параметры признаков при пяти интервалах:

- среднее значение;
- дисперсию;
- среднеквадратичное отклонение;
- минимальное значение;
- максимальное значение;
- нижнюю границу диапазона;
- верхнюю границу диапазона;
- ширину диапазона значений признака;
- шаг по интервалу;
- дельта.

3. Рассчитать границы интервалов диапазона значений признаков.

4. Рассчитать абсолютные частотные ряды признаков.

5. Рассчитать относительные частотные ряды признаков

6. Построить гистограммы всех признаков, используя мастер диаграмм.

7. Построить одну из гистограмм с помощью программы "Гистограмма" пакета "Анализ данных".

Предварительный анализ - многомерных данных

1. Рассчитать коэффициенты ковариации и корреляции признаков X5,X6, используя расчетные формулы.
2. Рассчитать коэффициенты ковариации и корреляции признаков X7,X8, используя расчетные формулы.
3. Рассчитать ковариационные матрицы признаков X1-X8, используя программу пакета анализа.
4. Рассчитать корреляционные матрицы признаков X1-X8, используя программу пакета анализа.
5. Построить диаграмму рассеивания признаков X1,X2.
6. Построить диаграмму рассеивания признаков X4,X5.
7. Построить диаграмму рассеивания признаков X2,X6.
8. Построить диаграмму рассеивания признаков по классифицированным данным:
 - первый класс составляют признаки X5,X6;
 - второй класс составляют признаки X7,X8.

Контрольная работа 4

Задача 1. По промышленным предприятиям города имеются следующие данные за отчетный год:

№	Объем	Фонд	№	Объем	Фонд
пред-	продук-	заработной	пред-	продук-	заработной
приятия	ции,	платы,	приятия	ции,	платы,
	млн руб.	млн руб;		млн руб.	млн руб.
1	124,8	19,8	9	110,0	17,7
2	256,0	38,4	10	256,3	40,9
3	190,7	31,3	11	187,5	30,7
4	185,0	31,4	12	140,8	23,2
5	403,2	56,4	13	167,3	27,0
6	115,0	19,6	14	208,2	32,2
7	106,5	17,2	15	135,4	21,9
8	350,0	49,7	16	370,2	51,8

Требуется:

- a) сгруппировать предприятия по объему выработанной продукции, выделив три группы (интервалы группировки разработать самостоятельно);
- b) определить по каждой группе число предприятий, объем продукции, фонд заработной платы, размер заработной платы (тыс. руб.) на 1 млн руб. объема продукции;
- c) оформить решение в виде статистической таблицы. Сформулировать вывод.

Задача 2. Имеются следующие данные о распределении продовольственных магазинов региона по размеру товарооборота за месяц:

Группы магазинов по товарообороту, млн руб.	4 0- 5 0	5 0- 6 0	6 0- 7 0	7 0- 8 0	8 0- 9 0	90- 100	100- 110
Число магазинов	2	4	7	10	15	20	22

Рассчитайте:

- a) средний объем инвестиций;
- b) модальное и медианное значение объема инвестиций.

Критерии оценки:

- оценка «отлично» выставляется студенту, если верно и правильно выполнено 90%-100% заданий;
- оценка «хорошо» выставляется студенту, если верно и правильно выполнено 70%-80% заданий;
- оценка «удовлетворительно» выставляется студенту, если верно и правильно решено 50%-60% заданий, возможны некоторые исправления при решении;
- оценка «неудовлетворительно» выставляется студенту, если верно выполнено менее 50% заданий;

Вопросы для коллоквиумов, собеседования

Модуль 1: Основы анализа данных с помощью компьютерных технологий

1. Предмет и задачи дисциплины. Состав современного программного обеспечения задач статистического анализа, характеристика основного профессионального статистического программного обеспечения.
2. Основные направления и проблемы статистического анализа массовых

данных.

3. Категории статистики: совокупность, статистические показатели, признаки; выборочные и генеральные статистические характеристики.
4. Подготовка исходных массивов данных. Переменные (признаки) и наблюдения.
5. Шкалы измерения признаков. Логический и арифметический контроль. Длинные имена переменных и формулы. Генерация переменных.
6. Проведение математических операций с переменными. Генераторы случайных чисел.
7. Ряды распределения. Дисперсионный анализ.
8. Стандартизация данных. Описательные статистики.
9. Графический метод в статистике.
10. Современные приемы визуализации данных.
11. Элементы, виды и правила построения статистических графиков.
12. Проверка нормальности распределения первичных данных, нормализация первичных данных. Подгонка разных распределений.
13. Обработка корреляционных таблиц и рядов распределения.
14. Основы корреляционного анализа.
15. Постановка задачи и конкретные ситуации анализа корреляционных таблиц.
16. Анализ парной и множественной корреляции. Анализ частной корреляции.
17. Построение парной линейной регрессии методом наименьших квадратов.
18. Оценка статистической значимости параметров парной линейной регрессии.
19. Проверка нормальности распределения признаков-регрессоров, зависимые и независимые статистические величины.
20. Терминология оценки статистической значимости параметров регрессии и регрессионной модели в целом.

21. Частная и полураспределенная корреляция переменных регрессионного анализа. Анализ поведения отклонений.
 22. Построение таблиц сопряженности. Определение силы связи номинальных переменных.
 23. Оценка значимости связи в таблицах сопряженности. Оценка связи ранговых переменных.
 24. Назначение и основы теории факторного анализа и компонентного анализа, основные понятия и показатели.
 25. Методы выделения факторов и компонентов. Оценка общностей в факторном анализе.
 26. Методики оценки числа факторов и компонентов. Методика вращения факторов и компонентов.
 27. Содержательная интерпретация матриц факторных нагрузок, факторов и компонентов.
 28. Оценка уровней факторов и компонентов. Регрессионные уравнения факторов через исходные переменные (признаки).
 29. Многомерное шкалирование. Матрицы расстояний.
 30. Оценка осей в многомерном шкалировании. Уровень стресса и отчуждения. Диаграмма Шепарда.
- Модуль 2. Статистическая классификация и анализ временных рядов**
31. Постановка задачи группировки объектов (наблюдений).
 32. Постановка задачи группировки признаков.
 33. Формирование обучающих выборок.
 34. Дискриминантный анализ. Обобщенный дискриминантный анализ.
 35. Деревья классификации. Задача классификации при наличии обучающих выборок.
 36. Линейный дискриминантный анализ. Решающее правило и дискриминантная функция.
 37. Дискриминантный анализ при нормальном законе распределения показателей, критерий отношения правдоподобия.

38. Статистическое оценивание результатов дискриминантного анализа.
39. Графическое отображение временных рядов и выявление тренда, возможности выбора трендовой функции.
40. Техника сглаживания временных рядов скользящими средними (центрированными и простыми).
41. Автокорреляционная и частная корреляционная функции. Приведение временного ряда к стационарной форме.
42. Семейство моделей авторегрессии и скользящего среднего (AR, MA, ARMA, ARIMA, SARMA, SARIMA).
43. Идентификация временного ряда, определение порядков авторегрессии и скользящего среднего. Оценка отклонений от тренда.
44. Экспоненциальное сглаживание и прогнозирование. Анализ сезонности.
45. Основы методики сглаживания и прогнозирования временных последовательностей экспоненциальными средними величинами.
46. Графическое представление исходного ряда динамики. Модели Брауна, Холта, Винтерса.
47. Методы сезонной декомпозиции временного ряда.
48. Кодирование значений переменных.
49. Обработка вопросов с множественными вариантами ответа.
50. Проверка статистических гипотез. Анализ надежности.

Критерии оценки:

- оценка «отлично» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются единичные несущественные ошибки, самостоятельно исправляемые учащимися;
- оценка «хорошо» выставляется студенту, если изложение полученных знаний в устной форме полное, в системе, в соответствии с требованиями учебной программы; допускаются, отдельные несущественные ошибки, исправляемые учащимися после указания преподавателя на них;
- оценка «удовлетворительно» выставляется студенту, если изложение полученных знаний неполное, однако это не препятствует усвоению последующего программного материала; допускаются отдельные существенные ошибки, исправляемые с помощью преподавателя;
- оценка «неудовлетворительно» выставляется студенту, если изложение учебного материала неполное, бессистемное, что препятствует усвоению

последующей учебной информации; существенные ошибки, не исправляемые даже с помощью преподавателя;

Темы лабораторных работ Лабораторная работа №1.

Обработка и данных с помощью компьютерных технологий

Цель работы – формирование теоретических знаний и практических навыков использования электронных таблиц табличного процессора Excel для обработки и анализа данных.

Задачи работы:

- научиться создавать электронные таблицы для расчета процентов по вкладам;
- научиться создавать электронные таблицы для определения влияния инфляции на стоимость денег;
- научиться создавать электронные таблицы для определения процентов по остаткам на расчетном счете.

Теоретические положения

Электронная таблица – это двумерный или n-мерный массив данных, размещенный на машинных носителях информации. Электронные таблицы широко используются в следующих областях: бухгалтерский и банковский учет, планирование распределения ресурсов, проектно-сметные работы, инженерно-технические расчеты, обработка больших массивов информации, исследование динамических процессов и др.

К основным возможностям электронных таблиц относятся:

- анализ и моделирование на основе выполнения вычислений и обработки данных;
- оформление таблиц и отчетов;
- форматирование содержащихся в таблице данных;
- построение диаграмм;
- создание и ведение баз данных с возможностью выбора записей по заданному критерию и сортировки по любому параметру;
- перенесение (вставка) в таблицу информации из документов, созданных в других приложениях, работающих в среде Windows;
- печать итогового документа целиком или частично.

Создание электронной таблицы для расчета процентов по вкладу

Предположим, что процент по депозиту сложный и начисляется в конце года (то есть период капитализации равен одному году). В каждом

следующем году расчет процентов производится для суммы, положенной на депозит, плюс проценты, начисленные за предыдущий год.

Сумма вклада на конец периода рассчитывается по следующей формуле:

$$p_1 = p_0 \times (1 + r)^n,$$

(1)

где p_0 – сумма, размещенная на депозите;

r – ставка по депозиту; n – число периодов (лет).

Задание. В банк на депозит положена сумма в размере 1000,00 тыс. руб. Необходимо рассчитать, какая сумма с учетом процентов будет получена через определенное время.

Исходные данные для решения задачи представлены в области ввода таблицы (рисунок 1), которая содержит следующие управляющие параметры:

- первоначальную сумму вклада;
- начальное значение процентной ставки по депозиту и шаг ее изменения;
- начальное значение периода времени и шаг его изменения.

Максимальное значение процентной ставки по депозиту 7,5%, а период расчета процентов по вкладу составляет 10 лет.

Создайте электронную таблицу для автоматического расчета суммы с учетом процентов.

Лабораторная работа №2

Первичная обработка статистических данных

Целью лабораторной работы является получение практических навыков по проведению первичной обработки статистических данных.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Статистическая группировка.
2. Методы визуализации данных.
3. Проверка гипотез на соответствие теоретическим законам распределения вероятностей.
4. Дисперсионный анализ.

Примеры заданий для защиты лабораторной работы

Используя официальные ресурсы ФСГС (URL: <http://www.gks.ru>) сформируйте массив данных по признакам для всех субъектов РФ за 2022 г.:

	Признак
	ВРП на душу населения, тыс. руб.
	Число малых предприятий, ед.
	Стоимость основных фондов, тыс. руб.

	Среднедушевые денежные доходы населения, руб.
	Выпуск бакалавров, специалистов, магистров, чел.
	Удельный вес сельского населения, %
	Объем инвестиций в основной капитал, млрд. руб.

а) Постройте диаграмму рассеяния между v_1 и v_3 .

Обозначьте на диаграмме рассеяния линию регрессии.

Прокомментируйте результат.

б) Постройте ряды распределения субъектов РФ и гистограммы по признакам v_2 , v_4 , v_6 .

в) Постройте матричный график по всем переменным.

Прокомментируйте результат. Определите взаимосвязанные признаки.

г) Разделив наблюдения по переменной v_3 на 5 равных интервала, проведите дисперсионный анализ для оценки влияния на переменную v_1 .

Лабораторная работа №3

Анализ взаимосвязей

Целью лабораторной работы является получение практических навыков

по исследованию взаимосвязей переменных различных шкал.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Парный линейный корреляционно-регрессионный анализ.
2. Парный нелинейный корреляционно-регрессионный анализ.
3. Множественный линейный корреляционно-регрессионный анализ.
4. Анализ связей ранговых переменных.
5. Анализ связи номинальных переменных.

Примеры заданий для защиты лабораторной работы

По массиву, сформированному в теме 2 проведите исследование. Оформите отчет по исследованию в документ MS Word:

1. Построить матрицу парных линейных коэффициентов корреляции. Определите один наиболее сильно связанный с v_1 показатель. Сделайте выводы о силе и направлении связи.
2. Постройте корреляционное поле между v_1 и отобранным показателем в п.1. Отобразите на графике линию регрессии.
3. Самостоятельно определите зависимую и независимую переменную. Постройте уравнения, оцените на адекватность, параметры на значимость при $p=(1-\alpha)=0,95$.
4. По скорректированному коэффициенту детерминации выберите лучшую модель и сделайте выводы.

5. Построить матрицу парных линейных коэффициентов корреляции. Определите наиболее сильно связанные с v_1 показатели.
6. Постройте матричный график между v_1 и отобранными показателями.
7. Постройте множественное линейное уравнение регрессии.

Лабораторная работа 4

Снижение размерности признакового пространства

Целью лабораторной работы является получение практических навыков по снижению размерности признакового пространства.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Общности.
2. Факторные нагрузки.
3. Вращение факторов.
4. Определение числа факторов.
5. Компонентный анализ.
6. Многомерное шкалирование.

Примеры заданий для защиты лабораторной работы

Имеется матрица исходных данных по предприятиям.

№	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
1	0,23	0,78	0,4	1,37	1,23	0,23	1,45	26006	167,69	47750	6,4	166,
2	0,24	0,75	0,26	1,4	1,04	0,39	1,3	23935	186,1	50391	7,8	92,
3	0,19	0,68	0,4	1,44	1,8	0,43	1,37	22589	220,45	43149	9,78	158,
4	0,17	0,7	0,5	1,42	0,43	0,18	1,65	21220	169,3	41089	7,9	93,
5	0,23	0,62	0,4	1,35	0,88	0,15	1,91	7394	39,53	14257	5,35	173,
6	0,43	0,76	0,19	1,39	0,57	0,34	1,68	11586	40,41	22661	3,9	162,
7	0,31	0,73	0,25	1,16	1,72	0,38	1,94	26609	102,96	52509	4,5	88,
8	0,26	0,71	0,44	1,27	1,7	0,09	1,89	7801	37,02	14903	4,88	101,
9	0,49	0,69	0,17	1,16	0,84	0,14	1,94	11587	45,74	25587	3,46	166,
10	0,36	0,73	0,39	1,25	0,6	0,21	2,06	9475	40,07	16821	3,6	140,
11	0,37	0,68	0,33	1,13	0,82	0,42	1,96	10811	45,44	19459	3,56	128,
12	0,43	0,74	0,25	1,1	0,84	0,05	1,02	6371	41,08	12973	5,65	177,
13	0,35	0,66	0,32	1,15	0,67	0,29	1,85	26761	136,14	50907	4,28	114,
14	0,38	0,72	0,02	1,23	1,04	0,48	0,88	4210	42,39	6920	8,85	93,
15	0,42	0,68	0,06	1,39	0,66	0,41	0,62	3557	37,39	5736	8,52	126,
16	0,3	0,77	0,15	1,38	0,86	0,62	1,09	14148	101,78	26705	7,19	91,
17	0,32	0,78	0,08	1,35	0,79	0,56	1,6	9872	47,55	20068	4,82	69,
18	0,25	0,78	0,2	1,42	0,34	1,76	1,53	5975	32,61	11487	5,46	66,
19	0,31	0,81	0,2	1,37	1,6	1,31	1,4	16662	103,25	32029	6,2	67,
20	0,26	0,79	0,3	1,41	1,46	0,45	2,22	9166	38,95	18946	4,25	50,

где v_1 Трудоемкость единицы продукции

v_2 Доля рабочих в общей численности работников

- v3 Доля затрат на покупку изделия в общих затратах на производство
- v4 Коэффициент сменности работы оборудования
- v5 Премия на работника, тыс. руб.
- v6 Удельный вес потерь от брака в общих затратах на производство, %
- v7 Уровень фондоотдачи
- v8 Численность работающих, чел.
- v9 Среднегодовая стоимость ОФ, тыс. руб.
- v10 Среднегодовой ФОТ, тыс. руб.
- v11 Уровень фондовооруженности на 1 рабочего, млн.руб.
- v12 Коэффициент оборачиваемости оборотных средств

Проведите факторный анализ методами главных факторов. Количество факторов определите методом «каменистой осыпи». Сформируйте матрицу факторных нагрузок. Сделайте выводы

Лабораторная работа 5

Статистическая классификация

Целью лабораторной работы является получение практических навыков

по проведению статистической классификации переменных и наблюдений.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Многомерная группировка.
2. Меры сходства и различий, метрики расстояний
3. Методы иерархического кластерного анализа.
4. Методы неиерархического кластерного анализа.
5. Методика проведения дискриминантного анализа.

Примеры заданий для защиты лабораторной работы

По массиву, сформированному в теме 4 проведите исследование.

Оформите отчет по исследованию в документ MS Word:

1. Проведите кластерный анализ методом дендритов всеми возможными способами по метрике Евклида.
2. Отобразите на графиках характеристики кластеров.
3. Дайте смысловую характеристику кластеров.
4. Проведите кластерный анализ предприятий методом k-средних, если известно, что количество кластеров = {3; 4; 5}.
5. Отобразите на графиках характеристики кластеров.
6. Дайте смысловую характеристику кластеров.

Лабораторная работа 6

Анализ и прогнозирование временных рядов

Целью лабораторной работы является получение практических навыков по проведению обработки и анализа временных рядов.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Графическое отображение временных рядов и выявление тренда.
2. Методы сезонной декомпозиции временных рядов.
3. Приведение временного ряда к стационарной форме.
4. Семейство ARIMA-моделей.
5. Экспоненциальное сглаживание.
6. Спектральный анализ.

Примеры заданий для защиты лабораторной работы

Используя официальные ресурсы ФСГС (URL: <http://www.gks.ru>) сформируйте массив данных о ежемесячном количестве родившихся детей в РФ за период с января 2012 по декабрь 2018.

- а) Постройте линейное уравнение тренда. Оцените адекватность получившегося уравнения.
- б) Выделите тренд с помощью скользящей средней. Интервал сглаживания определите самостоятельно.
- в) Исключите тренд.
- г) Выделите сезонную составляющую. Сделайте выводы о сезонном изменении рождаемости в РФ.
- д) Сделайте графики исходных данных, сглаженных значений, сезонной компоненты, случайных отклонений.

Лабораторная работа 7

Компьютерная обработка статистических формуляров

Целью лабораторной работы является получение практических навыков по проведению обработке статистических формуляров.

Содержание вопросов, рассматриваемых на лабораторных занятиях:

1. Формирование массивов данных по формулярам анкеты
2. Кодирование значений переменных.
3. Обработка вопросов с множественными вариантами ответа.
4. Проверка статистических гипотез.
5. Анализ надежности

Примеры заданий для защиты лабораторной работы

Имеются результаты проведения исследования «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ (RLMS HSE)», проводимый Национальным исследовательским университетом "Высшая школа экономики" и ООО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии Федерального научно-исследовательского социологического центра РАН. (Сайты обследования RLMS HSE: <http://www.cpc.unc.edu/projects/rlms> и <http://www.hse.ru/rlms>).

1. Экспортируйте данные и сформируйте массив.
2. Проверьте наличие взаимосвязи между адресными вопросами анкеты и ответом на вопрос раздела М «4.(3) Скажите, пожалуйста, как Вы оцениваете Ваше здоровье? Оно у Вас...».
3. Выдвинете различные гипотезы и осуществите их проверку.

...Критерии оценки:

- оценка «отлично» выставляется студенту, если ответ соответствует заданию, вычисления произведены правильно, используемые формулы представлены в общем виде, их использование обосновано;
- оценка «хорошо» выставляется студенту, если ответ в целом соответствует заданию, отсутствуют ошибки в употреблении методов расчетов, возможны незначительные арифметические ошибки, используемые формулы представлены в общем виде, выбор метода расчета не обоснован
- оценка «удовлетворительно» выставляется студенту, если ответ в целом соответствует заданию, есть вычислительные ошибки (25–30%), продемонстрировано достаточное владение методиками расчета, есть ошибки в употреблении и трактовке формул, используемые формулы в общем виде представлены с ошибками.
- оценка «неудовлетворительно» выставляется студенту, если ответ не соответствует заданию, серьезные ошибки в методике расчета, либо выбор не подходящей методики, используемые для расчета формулы, не представлены в общем виде.

Вопросы к экзамену

1. Предмет и задачи дисциплины. Состав современного программного обеспечения задач статистического анализа, характеристика основного профессионального статистического программного обеспечения.
2. Основные направления и проблемы статистического анализа массовых данных.
3. Категории статистики: совокупность, статистические показатели, признаки; выборочные и генеральные статистические характеристики.
4. Подготовка исходных массивов данных. Переменные (признаки) и наблюдения.
5. Шкалы измерения признаков. Логический и арифметический контроль. Длинные имена переменных и формулы. Генерация переменных.
6. Проведение математических операций с переменными. Генераторы случайных чисел.
7. Ряды распределения. Дисперсионный анализ.
8. Стандартизация данных. Описательные статистики.
9. Графический метод в статистике.
10. Современные приемы визуализации данных.

11. Элементы, виды и правила построения статистических графиков.
12. Проверка нормальности распределения первичных данных, нормализация первичных данных. Подгонка разных распределений.
13. Обработка корреляционных таблиц и рядов распределения.
14. Основы корреляционного анализа.
15. Постановка задачи и конкретные ситуации анализа корреляционных таблиц.
16. Анализ парной и множественной корреляции. Анализ частной корреляции.
17. Построение парной линейной регрессии методом наименьших квадратов.
18. Оценка статистической значимости параметров парной линейной регрессии.
19. Проверка нормальности распределения признаков-регрессоров, зависимые и независимые статистические величины.
20. Терминология оценки статистической значимости параметров регрессии и регрессионной модели в целом.
21. Частная и полураспределенная корреляция переменных регрессионного анализа. Анализ поведения отклонений.
22. Построение таблиц сопряженности. Определение силы связи номинальных переменных.
23. Оценка значимости связи в таблицах сопряженности. Оценка связи ранговых переменных.
24. Назначение и основы теории факторного анализа и компонентного анализа, основные понятия и показатели.
25. Методы выделения факторов и компонентов. Оценка общностей в факторном анализе.
26. Методики оценки числа факторов и компонентов. Методика вращения факторов и компонентов.
27. Содержательная интерпретация матриц факторных нагрузок, факторов

и компонентов.

28. Оценка уровней факторов и компонентов. Регрессионные уравнения факторов через исходные переменные (признаки).
29. Многомерное шкалирование. Матрицы расстояний.
30. Оценка осей в многомерном шкалировании. Уровень стресса и отчуждения. Диаграмма Шепарда.
31. Постановка задачи группировки объектов (наблюдений).
32. Постановка задачи группировки признаков.
33. Формирование обучающих выборок.
34. Дискриминантный анализ. Обобщенный дискриминантный анализ.
35. Деревья классификации. Задача классификации при наличии обучающих выборок.
36. Линейный дискриминантный анализ. Решающее правило и дискриминантная функция.
37. Дискриминантный анализ при нормальном законе распределения показателей, критерий отношения правдоподобия.
38. Статистическое оценивание результатов дискриминантного анализа.
39. Графическое отображение временных рядов и выявление тренда, возможности выбора трендовой функции.
40. Техника сглаживания временных рядов скользящими средними (центрированными и простыми).
41. Автокорреляционная и частная корреляционная функции. Приведение временного ряда к стационарной форме.
42. Семейство моделей авторегрессии и скользящего среднего (AR, MA, ARMA, ARIMA, SARMA, SARIMA).
43. Идентификация временного ряда, определение порядков авторегрессии и скользящего среднего. Оценка отклонений от тренда.
44. Экспоненциальное сглаживание и прогнозирование. Анализ сезонности.
45. Основы методики сглаживания и прогнозирования временных последовательностей экспоненциальными средними величинами.

- 46.Графическое представление исходного ряда динамики. Модели Брауна, Холта, Винтерса.
- 47.Методы сезонной декомпозиции временного ряда.
- 48.Кодирование значений переменных.
- 49.Обработка вопросов с множественными вариантами ответа.
- 50.Проверка статистических гипотез. Анализ надежности.

Критерии оценки:

Структура билета для проведения экзамена и распределение количества баллов по структурным элементам билета

Раздел билета	Виды заданий раздела билета	Количество баллов
1	Теоретический вопрос	40
2	Задача	60
Всего баллов по билету		100

Итоговая оценка результатов зачета с оценкой

Количество баллов	0-50	51-66	66-85	86-100
Оценка	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично

Порядок проведения экзамена:

Экзамен проводится в комбинированной форме: студент отвечает на теоретический вопрос устно, задача выполняется на ПК.

Время на подготовку ответов – 60 минут.