

Минобрнауки России

Бузулукский гуманитарно-технологический институт (филиал)
федерального государственного бюджетного образовательного учреждения
высшего образования
«Оренбургский государственный университет»

Кафедра педагогического образования

Фонд
оценочных средств
по дисциплине *«Анализ данных»*

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки
21.03.01 Нефтегазовое дело
(код и наименование направления подготовки)

Эксплуатация и обслуживание объектов добычи нефти и газа
(наименование направленности (профиля) образовательной программы)

Квалификация
Бакалавр

Форма обучения
Очно-заочная

Год набора 2025

Фонд оценочных средств предназначен для контроля знаний обучающихся по направлению подготовки 21.03.01 Нефтегазовое дело по дисциплине «Анализ данных»

Фонд оценочных средств рассмотрен и утвержден на заседании кафедры
педагогического образования

наименование кафедры

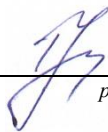
протокол № 6 от "17" января 2025 г.

Декан факультета

экономики и права

наименование факультета

подпись



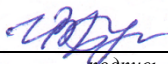
О. Н. Григорьева

расшифровка подписи

Исполнители:

ст. преподаватель

должность



подпись

И.В. Балан

расшифровка подписи

Раздел 1. Перечень компетенций, с указанием этапов их формирования в процессе освоения дисциплины

Формируемые компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций	Виды оценочных средств/ шифр раздела в данном документе
ОПК-6: Способен принимать обоснованные технические решения в профессиональной деятельности, выбирать эффективные и безопасные технические средства и технологии	ОПК-6-В-1 Знает принципы информационно-коммуникационных технологий и основные требования информационной безопасности	<u>Знать:</u> - основные понятия и положения, связанные со сбором, систематизацией, обработкой и анализом статистической данных; - основные подходы к анализу данных с использованием описательных и вероятностно-статистических методов; - основные требования техники безопасности	Блок А – задания репродуктивного уровня Тестовые задания Вопросы для собеседования
		<u>Уметь:</u> - определять методы анализа, необходимые для оценки степени и вида зависимостей между переменными, снижения размерности признакового пространства и многомерной классификации данных; - использовать методы устойчивого, робастного оценивания параметров и непараметрического моделирования; - анализировать временные данные и прогнозировать	Блок В – задания реконструктивного уровня Типовые задачи
		<u>Владеть:</u> - описательной статистики, табличного и графического представления данных, их содержательной интерпретации; - применения многомерных статистических методов исследования зависимостей, снижения размерностей и классификации при анализе социально-экономических данных с использованием аналитического программного обеспечения; - анализа динамики временных данных и прогнозирования социально-экономических процессов с использованием статистических методов и аналитического программного обеспечения	Блок С – задания практико-ориентированного уровня Задачи

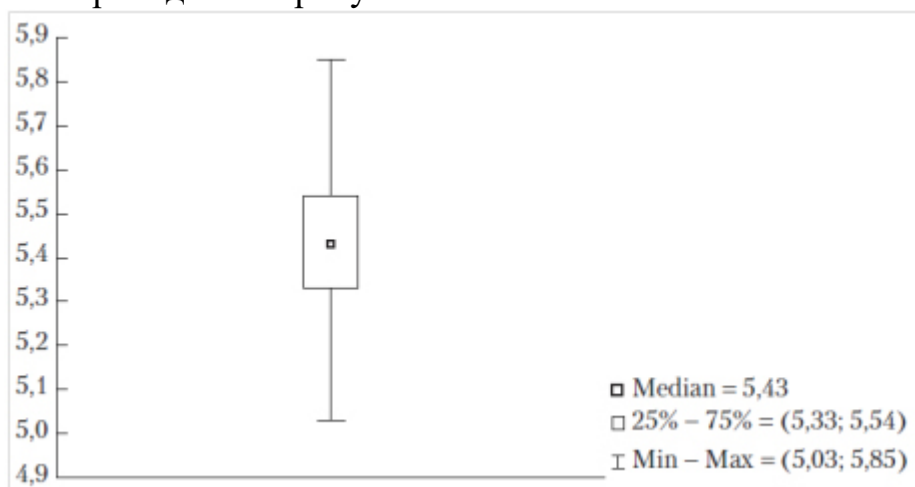
Раздел 2. Типовые контрольные задания и иные материалы, необходимые для оценки планируемых результатов обучения по дисциплине (оценочные средства). Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Блок А

А.0 Фонд тестовых заданий по дисциплине, разработанный и утвержденный в соответствии с Положением «О формировании фонда тестовых заданий по дисциплине»

Вопрос 1 (выбор одного правильного ответа)

Что приведено на рисунке?



- 1) Пример диаграммы «ящик с ногами»
- 2) Пример диаграммы «ящик с ушами»
- 3) Пример диаграммы «ящик с руками»
- 4) Пример диаграммы «ящик с усами»

Ответ: 4)

Вопрос 2 (выбор нескольких правильных ответов)

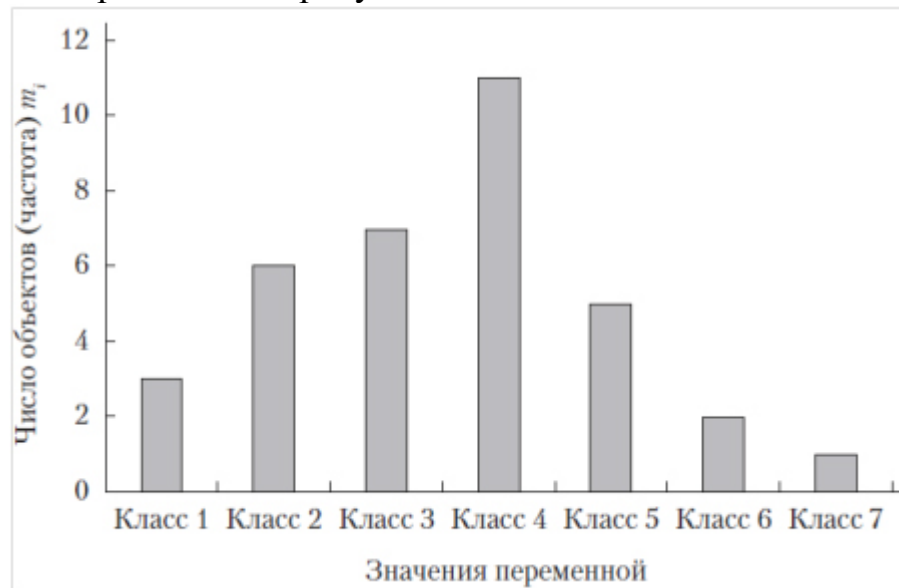
В многомерных (двумерных, трехмерных и т.д.) наборах данных у каждого наблюдения регистрируется несколько признаков. Статистические методы в этом случае используются для решения задач:

- 1) графическое представление данных с целью их визуализации и анализа
- 2) анализа наличия и степени зависимости между этими признаками
- 3) группировка данных и построение вариационных рядов
- 4) классификации наблюдений с целью получения однородных групп (кластеров) и выявления аномальных наблюдений
- 5) расчет средних значений и показателей вариации, размаха признака
- 6) исследования вида зависимости одной переменной от остальных
- 7) построения обобщающих, интегральных показателей с целью снижения размерности исходного признакового пространства
- 8) определения основных характеристик по каждому одномерному признаку
- 9) анализа рядов и прогнозирования (для временных данных)

Ответ: 2, 4, 6, 7, 8, 9

Вопрос 3 (выбор одного правильного ответа)

Что приведено на рисунке?



- 1) Вид типичной круговой диаграммы для количественных данных
- 2) Вид типичной гистограммы для количественных данных
- 3) Вид типичной круговой диаграммы для категориальных данных
- 4) Вид типичной гистограммы для категориальных данных

Ответ: 4

Вопрос 4 (выбор одного правильного ответа)

Средние представляют собой обобщающие показатели, характеризующие центр группирования данных. Обычно рассматривают средние:

- 1) геометрическую
- 2) системную
- 3) арифметическую
- 4) гармоническую
- 5) взвешенную

Ответ: 1), 3), 4)

Вопрос 5 (выбор одного правильного ответа)

Коэффициент детерминации между x и y характеризует:

- 1) долю дисперсии y , обусловленную влиянием не входящих в модель факторов
- 2) долю дисперсии y , обусловленную влиянием x
- 3) направление зависимости между x и y
- 4) долю дисперсии x , обусловленную влиянием не входящих в модель факторов

Ответ: 2

Вопрос 6 (на сопоставление вариантов)

Сопоставьте методы с их математическим описанием

А) $\sum_{i=1}^n (y_i - f(x_i, \beta))^2 \rightarrow \min_{\beta}$

Б) $\max_{1 \leq i \leq n} |y_i - f(x_i, \beta)| \rightarrow \min_{\beta}$

В)
$$\sum_{i=1}^n |y_i - f(x_i, \beta)| \rightarrow \min_{\beta}$$

- 1) метод минимакса
- 2) метод наименьших квадратов
- 3) метод наименьших модулей

Ответ: А-2, Б-1, В-3

Вопрос 7 (выбор одного правильного ответа)

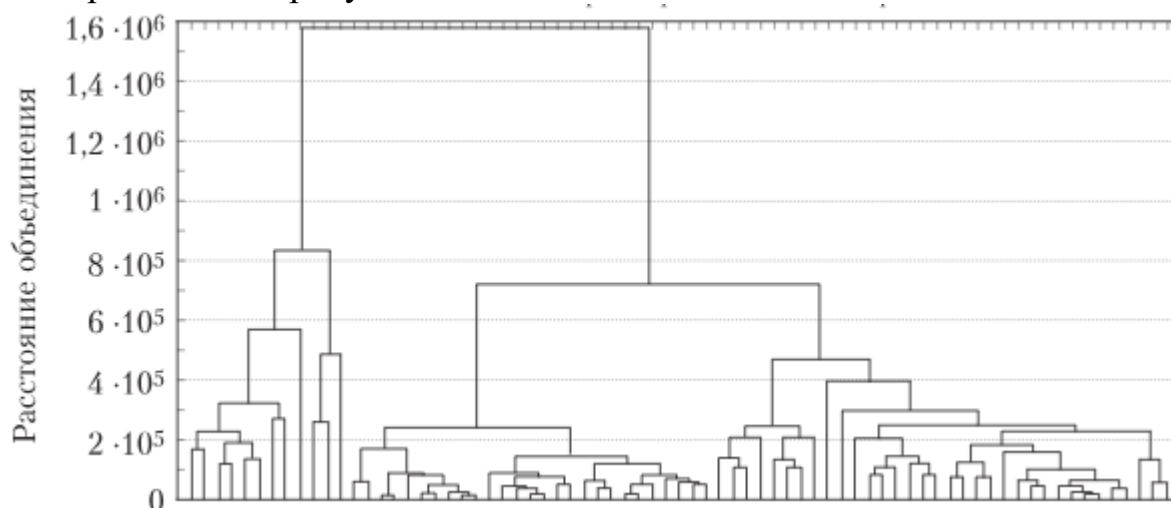
Для оценки коэффициентов регрессии, на которые оказывают влияние сопутствующие переменные, в регрессионную модель вводят так называемые фиктивные переменные, принимающие только два значения:

- 1) -1 или 0
- 2) 0 или 1
- 3) -1 или 1
- 4) 1 или 2

Ответ: 2

Вопрос 8 (выбор одного правильного ответа)

Что приведено на рисунке?

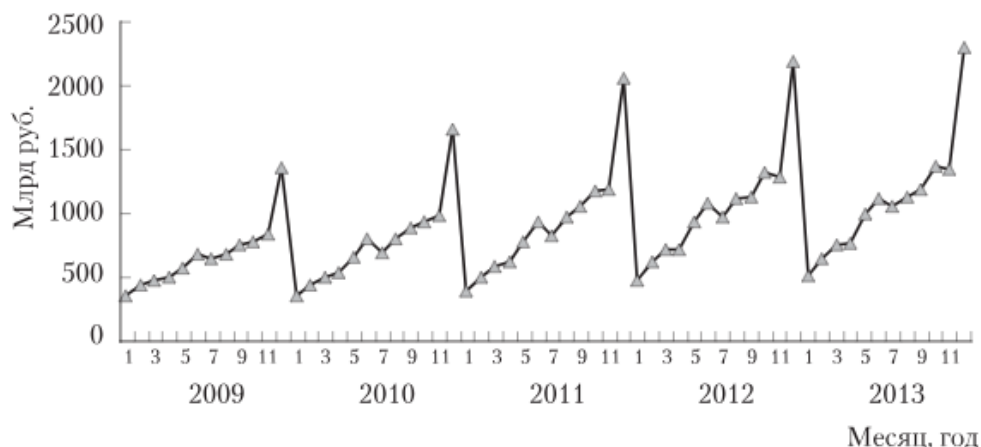


- 1) Пример Графика
- 2) Пример Гистограммы
- 3) Пример Диаграммы
- 4) Пример Дендрограммы

Ответ: 4

Вопрос 9 (выбор одного правильного ответа)

Что приведено на рисунке?



- 1) Пример возрастающего тренда, который сочетается с устойчивыми сезонными колебаниями
- 2) Пример убывающего тренда, который сочетается с устойчивыми сезонными колебаниями
- 3) Сглаживание временного ряда объемов продаж с помощью скользящей средней
- 4) Пример, демонстрирующий сезонные эффекты при наличии возрастающего тренда

Ответ: 1

Вопрос 10

Для мультипликативной модели временного ряда $Y = T \cdot S \cdot E$ сумма скорректированных сезонных компонент равна ...

- 1) средней оценке сезонной компоненты
- 2) скорректированной сезонной компоненте
- 3) лагу
- 4) константе

Ответ: 3

A.1 Вопросы для опроса открытого типа (ответ словом, числом):

1 В каких рядах динамики уровни характеризуют значения показателя по состоянию на определенные моменты времени.

Ответ: моментных

2 Анализ данных может проводиться с целью:

анализа и отображения конкретной собранной информации – в этом случае говорят о какой статистике;

описания всего класса явлений по имеющимся выборочным данным, характеризующим только часть этого класса. Эти задачи относятся к какой статистике.

Ответ: описательной, аналитической

3 Зависимость, при которой каждому значению величины X соответствует единственное значение величины Y и наоборот

Ответ: функциональная

4 Что представляет собой диаграмму, на которой изображается совокупность значений двух признаков. Каждая точка этой диаграммы имеет координаты (x_i, y_i) , соответствующие размерам признаков в i -м наблюдении.

Ответ: корреляционное поле

5 Значение критерия Дарбина – Уотсона можно приблизительно рассчитать по формуле $d \approx 2 \cdot (1 - r_\varepsilon)$, где r_ε – значение коэффициента автокорреляции остатков модели. Максимальная величина значения d будет наблюдаться при _____ автокорреляции остатков

Ответ: отрицательной

6 Из несмещенности оценки параметра следует, что среднее значение остатков равно

Ответ: 0

7 Если общая сумма квадратов отклонений $\sum(y - \bar{y})^2 = 120$, и остаточная сумма квадратов отклонений $\sum(y - \hat{y})^2 = 30$, то сумма квадратов отклонений, объясненная регрессией, равна

Ответ: 90

8 Процедура, при которой начальным является разбиение, состоящее из одного класса, а конечное — из n одноэлементных классов, называется

Ответ: агломеративной

9 Уровни временного ряда y_t , $t = 1, 2, \dots, n$ представлены в виде $y_t = u_t + s_t + v_t + e_t$, где u_t — трендовая составляющая; s_t — сезонная компонента; v_t — циклическая компонента; e_t — случайная компонента, что соответствует _____ форме модели:

Ответ: аддитивной

10 Значение коэффициента автокорреляции второго порядка равно $(-0,6)$, следовательно, ряд содержит ...

Ответ: тенденцию

Блок В

В.1 Типовые задачи

1 На основе интервального вариационного ряда объемов производства предприятия, представленного в таблице, требуется вычислить среднюю арифметическую этих объемов

Объем производства, усл.ед.	102 – 104	104 – 106	106 – 108	108 – 110	110 – 112
Количество предприятий	7	14	13	9	7

Ответ: 106,8 усл.ед.

2 При ранжировании оценок на вступительных экзаменах и средних баллов за экзаменационную сессию одних и тех же студентов получены следующие ранги.

Параметр	Студент									
	А	Б	В	Г	Д	Е	Ж	З	И	К
Вступительный экзамен	2	5	6	1	4	10	7	8	3	9
Экзаменационная сессия	3	6	4	1	2	7	8	10	5	9
d	-1	-1	2	0	2	3	-1	-2	-2	0
d^2	1	1	4	0	4	9	1	4	4	0

Требуется рассчитать значение коэффициента корреляции Спирмена

Ответ: 0,83

3 Дана матрица парных коэффициентов корреляции.

	y	x_1	x_2	x_3
y	1	–	–	–
x_1	0,987	1	–	–
x_2	0,754	0,451	1	–
x_3	0,857	0,789	0,154	1

Какие факторы являются коллинеарными?

Ответ: x_1, x_3

4 Если известно уравнение множественной регрессии $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$ построенное по результатам 50 наблюдений, для которого общая сумма квадратов отклонений равна 153, и остаточная сумма квадратов отклонений равна 3, то значение F -статистики равно

Ответ: 766,67

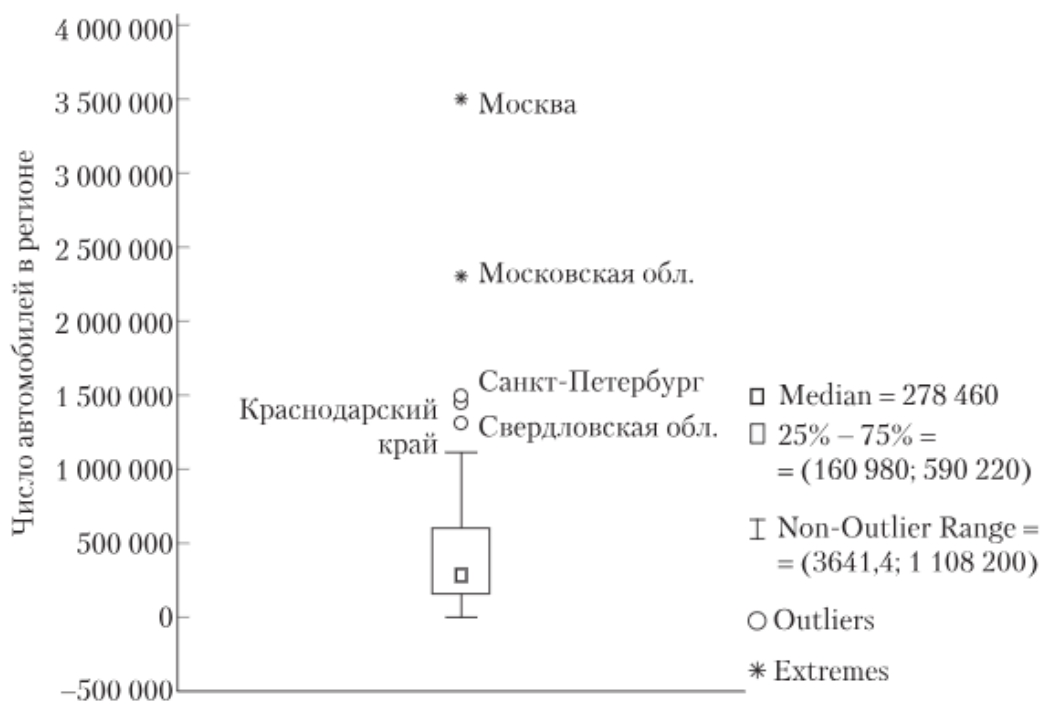
5 Уравнение регрессии потребления материалов от объема производства, построенное по 15 наблюдениям, имеет вид: $y = 5 + 5x + \varepsilon$, $t_b = 4,0$. Определить коэффициент детерминации для этого уравнения.

Ответ: 0,552

Блок С

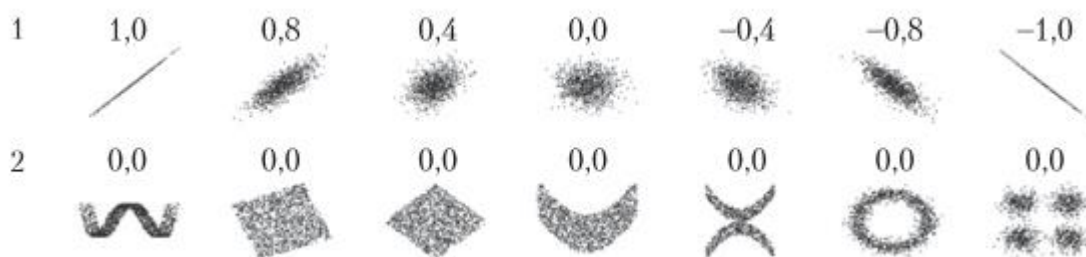
С.1 Индивидуальные творческие задания (вопрос типа эссе)

1 На рисунке представлены данные по регионам России, отражающие количество автомобилей в регионе. Опишите характеристики представленной совокупности.



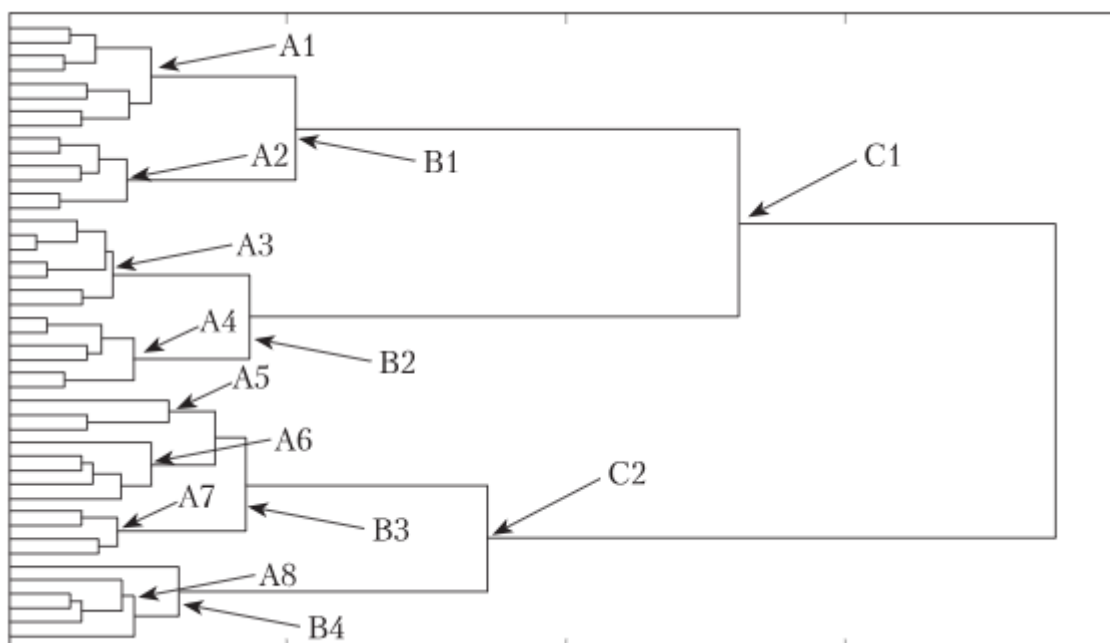
Ответ: медиана распределения (половина регионов имеет количество автомобилей, не превышающих 278 460 шт.), интерквартильный размах (25% регионов имеет менее 160980 автомобилей, а 25% - более 590220 автомобилей), аномальные наблюдения (Свердловская область), экстремальные наблюдения (Санкт-Петербург, Краснодарский край, Московская область, Москва), распределение обладает существенной правосторонней асимметрией.

2 На рисунке приведены поля корреляции между переменными x и y . Сверху каждого рисунка представлены соответствующие коэффициенты корреляции. Определите, в каком случае для исследования степени тесноты взаимосвязи между переменными могут быть использованы линейные коэффициенты корреляции.



Ответ: для анализа взаимосвязи между переменными в первой строке могут быть использованы линейные коэффициенты корреляции, во второй строке линейные коэффициенты корреляции не могут использоваться, так как связь носит сложный, нелинейный характер. В центре рисунка значение коэффициента корреляции не определено, так как дисперсия y равна нулю.

3 На рисунке представлена дендрограмма объединения объектов в кластеры. Сколько кластеров необходимо выделить?



Ответ: число выделяемых кластеров может быть равно 8 (A1-A8), 4 (B1-B4), 2 (C1-C2). Целесообразно выделить 4 кластера.

4 На основе помесечных данных за последние 5 лет была построена аддитивная временная модель потребления тепла в районе. Скорректированные значения сезонной компоненты приведены в таблице

Январь	+ 27	Май	- 20	Сентябрь	- 10
Февраль	+ 22	Июнь	- 34	Октябрь	+ 12
Март	+ 15	Июль	- 42	Ноябрь	+20
Апрель	- 2	Август	- 18	Декабрь	?

Уравнение тренда выглядит так: $T = 300 + 1,1 \cdot t$.

Определить значение сезонной компоненты за декабрь, а также точечный прогноз потребления тепла на 2-й квартал следующего года.

Ответ: значение сезонной компоненты за декабрь:

$$S_{12} = - \sum_{i=1, (i \neq 12)}^{12} S_i = (27 + 22 + 15 - 2 - 20 - 34 - 42 - 18 - 10 + 12 + 20) = -30.$$

Прогноз за весь второй квартал складывается из прогнозов за апрель, май и июнь.

$$\hat{y}(\text{апрель}) = (300 + 1,1 \cdot 64) - 2 = 368,4;$$

$$\hat{y}(\text{май}) = (300 + 1,1 \cdot 65) - 20 = 351,5;$$

$$\hat{y}(\text{июнь}) = (300 + 1,1 \cdot 66) - 34 = 338,6;$$

$$\hat{y}(2 - \text{й квартал}) = 368,4 + 351,5 + 338,6 = 1058,5.$$

5 Опишите порядок построения и анализа уравнения парной регрессии.

Ответ: 1) вычисление коэффициентов парной регрессии на основе метода наименьших квадратов; 2) исследование уравнения парной регрессии: оценка тесноты связи (линейный коэффициент корреляции, коэффициент детерминации, средняя ошибка аппроксимации); оценка значимости уравнения линейной регрессии и существенности параметров линейной регрессии

Блок D

Экзаменационные вопросы

- 1 Классификация статистических данных
- 2 Статистическое оценивание параметров генеральных совокупностей
- 3 Уравнение парной линейной регрессии. Поле корреляции.
- 4 Метод наименьших квадратов.
- 5 Нелинейные модели регрессии и их линеаризация.
- 6 Оценка тесноты связи. Коэффициент корреляции, индекс корреляции.
- 7 Множественная регрессия. Выбор вида уравнения.
- 8 Включение в уравнения качественных признаков. Фиктивные переменные.
- 9 Кластерный анализ, непараметрическая классификация без обучения.
- 10 Классификация с обучением.
- 11 Дискриминантный анализ
- 12 Ряды динамики. Компоненты ряда.
- 13 Мультипликативные и аддитивные модели.
- 14 Выбор уравнения тренда.
- 15 Моделирование тенденции временного ряда при наличии структурных изменений

Описание показателей и критериев оценивания компетенций, описание шкал оценивания

4-балльная шкала	Отлично	Хорошо	Удовлетворительно	Неудовлетворительно
100 балльная шкала	85-100	70-84	50-69	0-49
Бинарная шкала	Зачтено			Не зачтено

Оценивание выполнения практических заданий

4-балльная шкала	Показатели	Критерии
Отлично	1. Полнота выполнения практического задания; 2. Своевременность выполнения задания; 3. Последовательность и рациональность выполнения задания;	Задание решено самостоятельно. При этом составлен правильный алгоритм решения задания, в логических рассуждениях, в выборе формул и решении нет ошибок, получен верный ответ, задание решено рациональным способом.
Хорошо	4. Самостоятельность решения; 5. и т.д.	Задание решено с помощью преподавателя. При этом составлен правильный алгоритм решения задания, в логическом рассуждении и решении нет существенных ошибок; правильно сделан выбор формул для решения; есть объяснение решения, но задание решено нерациональным способом или допущено не более двух несущественных ошибок, получен верный ответ.
Удовлетворительно		Задание решено с подсказками преподавателя. При этом задание понято правильно, в логическом рассуждении нет существенных ошибок, но допущены существенные ошибки в выборе формул или в математических расчетах; задание решено не полностью или в общем виде.
Неудовлетвори-		Задание не решено.

4-балльная шкала	Показатели	Критерии
тельно		

Оценивание выполнения тестов

4-балльная шкала	Показатели	Критерии
Отлично	1. Полнота выполнения тестовых заданий; 2. Своевременность выполнения; 3. Правильность ответов на вопросы; 4. Самостоятельность тестирования;	Выполнено 85-100 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос.
Хорошо		Выполнено 70-84 % заданий предложенного теста, в заданиях открытого типа дан полный, развернутый ответ на поставленный вопрос; однако были допущены неточности в определении понятий, терминов и др.
Удовлетворительно		Выполнено 50-69 % заданий предложенного теста, в заданиях открытого типа дан неполный ответ на поставленный вопрос, в ответе не присутствуют доказательные примеры, текст со стилистическими и орфографическими ошибками.
Неудовлетворительно		Выполнено 49 % заданий предложенного теста, на поставленные вопросы ответ отсутствует или неполный, допущены существенные ошибки в теоретическом материале (терминах, понятиях).

Раздел 3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Основными этапами формирования компетенций по дисциплине при изучении студентами дисциплины являются последовательное изучение содержательно связанных между собой разделов.

Оценка знаний студентов на промежуточной аттестации производится по следующим критериям:

оценка «зачтено» выставляется студенту, если он усвоил программный материал курса, последовательно, четко и логически его излагает, умеет увязывать теорию с практикой, справляется с задачами и вопросами, обосновывает принятые решения, владеет навыками и приемами выполнения практических задач;

оценка «незачтено» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями решает практические задачи или не справляется с ними самостоятельно.

При оценивании результатов обучения: знания, умения, навыки и/или опыта деятельности (владения) в процессе формирования заявленных компетенций используются различные формы оценочных средств текущего и итогового контроля (промежуточной аттестации).