

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Поверинов Игорь Егорович
Должность: Проректор по учебной работе
Дата подписания: 27.09.2022 21:37:35
Уникальный программный идентификатор:
6d465b936eef331cede482bde6d12ab98216652f016465d53b72a2eab0de1b2

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«Чувашский государственный университет имени И. Н. Ульянова»
(ФГБОУ ВО «ЧГУ им. И.Н. Ульянова»)**

Факультет экономический

Кафедра бухгалтерского учета и электронного бизнеса

Утверждено
на заседании кафедры бухгалтерского
учета и электронного бизнеса
Заведующий кафедрой Львова М. В.

04.04.2022

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
(ФОНД ОЦЕНОЧНЫХ СРЕДСТВ)**

«Анализ многомерных данных»

Направление подготовки / специальность 38.04.01 Экономика

Квалификация выпускника Магистр

Направленность (профиль) / специализация « Искусственный интеллект и анализ
больших данных в банковской сфере»

Год начала подготовки - 2022

Паспорт
оценочных материалов для проведения текущего контроля и
промежуточной аттестации обучающихся по дисциплине (модулю)
Анализ многомерных данных

Перечень оценочных материалов и индикаторов достижения компетенций,
сформированность которых они контролируют

Наименование оценочного средства	Коды индикаторов достижения формируемых компетенции	Номер приложения
Тесты	ИД-1 ук-6, ИД-2 ук-6, ИД-3 ук-6, ИД-1 пк-1, ИД-2 пк-1, ИД-1 пк-7, ИД-2 пк-7, ИД-3 пк-7, ИД-1 пк-8	1
Выполнение и собеседование по лабораторным работам	ИД-1 ук-6, ИД-2 ук-6, ИД-3 ук-6, ИД-1 пк-1, ИД-2 пк-1, ИД-1 пк-7, ИД-2 пк-7, ИД-3 пк-7, ИД-1 пк-8	2
Курсовая работа	ИД-1 ук-6, ИД-2 ук-6, ИД-3 ук-6, ИД-1 пк-1, ИД-2 пк-1, ИД-1 пк-7, ИД-2 пк-7, ИД-3 пк-7, ИД-1 пк-8	3
Экзамен	ИД-1 ук-6, ИД-2 ук-6, ИД-3 ук-6, ИД-1 пк-1, ИД-2 пк-1, ИД-1 пк-7, ИД-2 пк-7, ИД-3 пк-7, ИД-1 пк-8	4

I. Текущий контроль

Приложение 1

Тесты

1. Процедура проведения тестирования

Количество проводимых тестов в течение всего периода освоения дисциплины	6 тестов
Общее количество тестовых вопросов в банке тестов	65 вопросов
Количество задаваемых тестовых вопросов в одном тесте	5 вопросов
Формат проведения тестирования	Электронный
Сроки / Периодичность проведения тестирования	Тест должен быть решен в течение 2 недель с момента начала изучения раздела дисциплины
Методические рекомендации (при необходимости)	Размещаются в электронной образовательной среде ВУЗа, случайный выбор вопроса из банка заданий по соответствующей теме

2. Шкала оценивания с учетом срока сдачи

Количество правильных ответов	Балл
5	Отлично
4	Хорошо
3	Удовлетворительно
0-2	Неудовлетворительно

3. Тестовые задания

Представляется полный перечень тестовых заданий:

Тест №1

1.1 Переменная «число дождливых дней в году в октябре» относится к типу ...

- а) категориальная номинальная;
- б) категориальная порядковая;
- в) количественная непрерывная;
- г) количественная дискретная.

1.2 Значения переменной «число преступлений в течение года в Москве» за 1990-2015 гг.» относятся к типу данных ...

- а) интервальный временной ряд;
- б) моментный временной ряд;
- в) пространственные данные;
- г) панельные (перекрестно-временные) данные.

1.3 Модальное значение определяется...

- а) только для количественных данных;
- б) только для категориальных данных;
- в) для количественных и категориальных данных.

1.4 Величине признака, которая делит упорядоченную последовательность его значений на две равные по численности части, соответствует следующая числовая характеристика...

- а) мода;
- б) медиана;
- в) среднее;
- г) дисперсия.

1.5 Возраст студентов курса колеблется в пределах от 20 до 26 лет:

Возраст, лет	20	21	22	23	24	25	26	Итого
Количество студентов	15	25	50	40	35	25	10	200

Определите медиану.

- а) 22;
- б) 23;
- в) 50;
- г) 100.

1.6 Определить, является ли распределение значений переменной, скошенным в какую-либо сторону, можно с помощью...

- а) дисперсии;
- б) коэффициента вариации;
- в) коэффициента асимметрии;
- г) коэффициента эксцесса.

1.7 Дисперсию можно рассчитать по формуле ... Укажите не менее двух вариантов ответа.

- а) $\frac{\sum(x_i - \bar{x}) \cdot f_i}{\sum f_i}$;
- б) $\frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$;
- в) $\sqrt{\frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$;
- г) $\overline{x^2} - (\bar{x})^2$.

1.8 Интерквартильный размах содержит следующий процент значений признака;

- а) 25;
- б) 50;
- в) 75;
- г) 100.

1.9 На главной диагонали ковариационной матрицы вектора X находятся следующие моменты элементов вектора X :

- а) центральные моменты первого порядка;

- б) центральные моменты второго порядка;
- в) начальные моменты первого порядка;
- г) начальные моменты второго порядка.

1.10 Уровень значимости статистического критерия выбирают в интервале...

- а) от 0 до 0,5;
- б) от -1 до 1;
- в) от 0,5 до 1;
- г) от -1 до 0.

1.11 При построении интервальной оценки для дисперсии σ^2 при числе наблюдений $n \leq 30$ используются выборочные характеристики, имеющие...

- а) χ^2 -распределение;
- б) t -распределение;
- в) F -распределение.

1.12 При построении интервальной оценки для математического ожидания μ , если число наблюдений $n \leq 30$ и дисперсия σ^2 не известна, используются выборочные характеристики, имеющие...

- а) χ^2 -распределение;
- б) t -распределение;
- в) F -распределение.

Тест №2

2.1 Зависимость, при которой каждому значению величины X соответствует единственное значение величины Y и наоборот, является...

- а) статистической;
- б) корреляционной;
- в) функциональной.

2.2 Зависимость, при которой каждому фиксированному значению величины X соответствует не одно, а множество значений величины Y , называется...

- а) статистической;
- б) корреляционной;
- в) функциональной.

2.3 Статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой, называется ...

- а) регрессия;
- б) корреляция;
- в) дисперсия;
- г) функция.

2.4 Анализ тесноты и направления связи двух признаков осуществляется на основе ...

- а) линейного коэффициента корреляции;
- б) частного коэффициента корреляции;
- в) множественного коэффициента корреляции;
- г) коэффициента детерминации.

2.5 Графическое изображение данных наблюдений в виде точек в декартовой системе координат называется...

- а) полем корреляции;
- б) гистограммой;
- в) круговой диаграммой;
- г) диаграммой рассеивания.

2.6 Парный коэффициент корреляции изменяется в пределах:

- а) $0 \leq r_{xy} \leq 1$;
- б) $-1 \leq r_{xy} \leq 1$;
- в) $0 \leq r_{xy} \leq \infty$.

2.7 Установите соответствие между показателями и их содержанием:

- | | |
|---|--|
| а) парный коэффициент корреляции | 1) показатель связи между двумя количественными признаками |
| б) множественный коэффициент корреляции | 2) показатель связи между одним и множеством других количественных признаков |
| в) коэффициент конкордации | 3) показатель связи между произвольным числом ранжированных признаков |
| г) частный коэффициент корреляции | 4) показатель связи между количественными результативным и факторным признаками при элиминированном влиянии других факторных признаков |

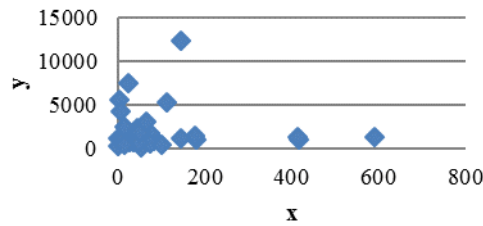
2.8 Установите соответствие между классификационными признаками и видами корреляционной связи:

- | | |
|-------------------------------------|--|
| а) теснота связи | 1) практически отсутствует, слабая, умеренная, сильная |
| б) направление связи | 2) прямая, обратная |
| в) аналитическое выражение связи | 3) линейная, нелинейная |
| г) число взаимосвязанных переменных | 4) парная, множественная |
| | 5) количественная, качественная |

2.9 Установите соответствие между значениями коэффициента корреляции и их смысловой интерпретацией:

- | | |
|---------|--|
| а) -0,8 | 1) связь обратная сильная |
| б) 0,9 | 2) связь прямая сильная |
| в) 0,1 | 3) связь прямая, практически отсутствует |
| г) -5 | 4) показатель не имеет смысла |
| | 5) связь обратная слабая |

2.10 Результаты наблюдений за совместным изменением значений изучаемых переменных x и y представлены на диаграмме рассеивания:



Сформулируйте вывод о характере связи между изучаемыми переменными:

- а) между переменными отсутствует какая-либо зависимость;
- б) между переменными имеется прямая линейная зависимость;
- в) между переменными существует обратная линейная зависимость.

Тест №3

3.1 Согласно методу наименьших квадратов минимизируется...

- а) $\sum_{i=1}^n (y_i - \bar{y})^2$
- б) $\sum_{i=1}^n |y_i - \hat{y}_i|$
- в) $\sum_{i=1}^n (y_i - \hat{y}_i)^2$
- г) $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

3.2 Для вектора $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ ковариационная матрица $S_b = \begin{pmatrix} 5,52 & -0,08 & -3,45 \\ -0,08 & 0,01 & 0,04 \\ -3,45 & 0,04 & 2,21 \end{pmatrix}$. Тогда

оценка дисперсии элемента b_2 равна:

- а) 5,52;
- б) 0,04;
- в) 0,01;
- г) 2,21.

3.3 Укажите свойства, характеризующие распределение случайных ошибок ε_i модели регрессии, полученной методом наименьших квадратов:

- а) $M(\varepsilon_i) = 1; Var(\varepsilon_i) = \sigma^2$
- б) $M(\varepsilon_i) = 0; Var(\varepsilon_i) = \sigma^2$
- в) $M(\varepsilon_i) = 1; Var(\varepsilon_i) = 0$
- г) $M(\varepsilon_i) = 0; Var(\varepsilon_i) = 0$

3.4 Критическое значение критерия Стьюдента определяет ...

- а) максимально возможную величину, допускающую принятие гипотезы о несущественности параметра;
- б) минимально возможную величину, допускающую принятие гипотезы о несущественности параметра;
- в) максимально возможную величину, допускающую принятие гипотезы о существенности параметра;
- г) минимально возможную величину, допускающую принятие гипотезы о равенстве нулю значения параметра.

3.5 Параметр уравнения регрессии является существенным, если...

- а) доверительный интервал не содержит ноль;

- б) расчетное значение критерия Стьюдента меньше критического значения;
- в) стандартная ошибка превышает половину значения самого параметра;
- г) доверительный интервал содержит ноль.

3.6 В регрессионном анализе критическое значение F -статистики служит для проверки статистической гипотезы о...

- а) незначимом отличии объясненной и необъясненной дисперсий зависимой переменной;
- б) о равенстве двух математических ожиданий;
- в) о равенстве дисперсии некоторой гипотетической величине;
- г) о равенстве математического ожидания некоторой гипотетической величине.

3.7 По результатам 50 статистических наблюдений построено уравнение множественной регрессии $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$. Число степеней свободы остаточной суммы квадратов отклонений для этого уравнения равно ...

- а) 46;
- б) 47;
- в) 49;
- г) 48.

3.8 По данным наблюдений оценены параметры модели множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon,$$

и записано уравнение (в скобках указаны значения стандартных ошибок параметров):

$$y = 2 - 3x_1 + 0,5x_2 + 4x_3 + \varepsilon$$

$$(2,4) \quad (-0,7) \quad (0,09) \quad (2,1)$$

Известны критические значения Стьюдента при различных уровнях значимости:

$$t_{(\alpha=0,1)} = 1,679; t_{(\alpha=0,05)} = 2,013; t_{(\alpha=0,01)} = 2,687.$$

Для данного уравнения при уровне значимости 0,01 значимыми являются...

- а) b_1, b_2 ;
- б) a, b_1, b_2, b_3 ;
- в) a, b_1, b_3 ;
- г) a, b_2, b_3 .

3.9 В процессе множественного регрессионного анализа получена матрица парных коэффициентов корреляции (y – зависимая переменная; $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}$ – независимые переменные):

	y	x_1	x_2	x_3	x_4
y	1				
x_1	0,75	1			
x_2	0,6	0,45	1		
x_3	0,89	0,82	0,3	1	
x_4	0,31	0,94	0,7	0,12	1

Тесно связанными независимыми переменными не являются ...

- а) $x^{(2)}$ и $x^{(3)}$;
- б) $x^{(1)}$ и $x^{(3)}$;
- в) $x^{(1)}$ и $x^{(4)}$;
- г) $x^{(2)}$ и $x^{(4)}$.

3.10 Фиктивные переменные используются в моделях множественной регрессии для учета влияния факторов, ...

- а) не имеющих количественных значений;
- б) имеющих количественные значения;
- в) имеющих вероятностную природу.

3.11 При анализе деятельности промышленных предприятий в трех регионах были построены три частных уравнения регрессии:

$$\hat{y} = 110 + 100 \cdot x \text{ для Псковской области;}$$

$$\hat{y} = 240 + 100 \cdot x \text{ для Новгородской области;}$$

$$\hat{y} = 500 + 100 \cdot x \text{ для Ленинградской области.}$$

Укажите вид фиктивных переменных и уравнение с фиктивными переменными, обобщающее три частных уравнения регрессии:

- | | | |
|----|--|--|
| а) | $\hat{y} = 110 + 100x + 130z_1 + 390z_2$ | $z_1 = \begin{cases} 1, \text{ если Новгородская область} \\ 0 - \text{ в остальных случаях} \end{cases}$ |
| б) | $\hat{y} = 110 + 100x + 240z_1 + 500z_2$ | $z_2 = \begin{cases} 1, \text{ если Ленинградская область} \\ 0 - \text{ в остальных случаях} \end{cases}$ |
| | | $z_1 = \begin{cases} 1, \text{ если Псковская область} \\ 0 - \text{ в остальных случаях} \end{cases}$ |
| | | $z_2 = \begin{cases} 1, \text{ если Новгородская область} \\ 0 - \text{ в остальных случаях} \end{cases}$ |

Тест №4

4.1 Нормирование признаков производят с целью...

- а) устранения влияния различных единиц измерения;
- б) уменьшения размерности признакового пространства;
- в) упрощения расчетов.

4.2 Относительный вклад (в %) первой главной компоненты в суммарную дисперсию для

матрицы $A = \begin{pmatrix} -0,90 & -0,30 & a_{13} \\ -0,54 & a_{22} & 0,03 \\ a_{31} & 0,20 & 0,30 \end{pmatrix}$ равен:

- а) 53;
- б) 59;
- в) 72;
- г) 66.

4.3 Получена матрица факторных нагрузок $A = \begin{pmatrix} 0,94 & 0,35 \\ -0,94 & 0,35 \end{pmatrix}$. Тогда матрица собственных значений имеет вид:

- | | |
|---|---|
| а) $\begin{pmatrix} 1,35 & 0 \\ 0 & 0,65 \end{pmatrix}$; | в) $\begin{pmatrix} 1,76 & 0 \\ 0 & 0,24 \end{pmatrix}$; |
| б) $\begin{pmatrix} 1,94 & 0 \\ 0 & 0,06 \end{pmatrix}$; | г) $\begin{pmatrix} 1,29 & 0 \\ 0 & 0,71 \end{pmatrix}$. |

4.4 Получена матрица факторных нагрузок $A = \begin{pmatrix} 0,94 & 0,35 \\ -0,94 & 0,35 \end{pmatrix}$. Тогда теснота линейной связи между исходными переменными x_1 и x_2 составляет:

- а) -0,76;
- б) 0,59;

- в) -0,59;
- г) 0,76.

4.5 При исследовании взаимосвязи двух переменных x_1 и x_2 получен коэффициент корреляции $r_{x_1x_2} = 0,9$. Тогда максимальное собственное число, соответствующее первой главной компоненте, равно:

- а) 0,1;
- б) 1,8;
- в) 0,2;
- г) -0,9.

4.6 В двумерной совокупности коэффициент корреляции между величинами x_1 и x_2 равен 0,95. Тогда собственные значения λ_1 и λ_2 равны:

- а) 1,05 и 0,95;
- б) 1,40 и 0,22;
- в) 1,95 и 0,05;
- г) 1,45 и 0,55.

4.7 Финансовая устойчивость фирмы характеризуется восемью показателями. В результате расчетов получены собственные значения трех первых главных компонент: $\lambda_1 = 4,0$; $\lambda_2 = 1,6$ и $\lambda_3 = 0,8$. Тогда относительный вклад двух первых главных компонент равен (в %):

- а) 30;
- б) 70;
- в) 60;
- г) 80.

4.8 Получена матрица факторных нагрузок $A = \begin{pmatrix} -0,72 & 0,69 & -0,08 \\ 0,88 & 0,44 & 0,19 \\ 0,96 & 0,12 & -0,24 \end{pmatrix}$. Тогда

относительный вклад второй главной компоненты f_2 в суммарную дисперсию равен (в %):

- а) 74;
- б) 37;
- в) 4;
- г) 23.

4.9 Для матрицы факторных нагрузок $A = \begin{pmatrix} -0,90 & -0,30 & a_{13} \\ -0,54 & a_{22} & 0,03 \\ a_{31} & 0,20 & 0,30 \end{pmatrix}$ отсутствуют значения

элементов a_{13} , a_{22} , a_{31} . Собственное значение λ_3 , соответствующее третьей главной компоненте f_3 , равно:

- а) 1,95;
- б) 0,63;
- в) 0,21;
- г) 0,84.

4.10 Элементы матрицы факторных нагрузок варьируют в пределах:

- а) (-1;0);
- б) (0;1);
- в) (-1;1);
- г) (0;2).

Тест №5

5.1 В кластер S_I входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2, 5, 6, 7. Расстояние от пятого объекта до кластера S_I , если исходить из принципа «дальнего соседа», равно:

- а) 2;
- б) 5;
- в) 6;
- г) 7.

5.2 В кластер S_I входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2, 5, 6, 7. Расстояние от пятого объекта до кластера S_I , если исходить из принципа «ближнего соседа», равно:

- а) 2;
- б) 5;
- в) 6;
- г) 7.

5.3 В кластер S_I входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2, 5, 6, 7. Расстояние от пятого объекта до кластера S_I , если исходить из принципа «средней связи», равно:

- а) 2;
- б) 5;
- в) 6;
- г) 7.

5.4 Расстояние между пятью объектами характеризуется матрицей расстояний

$$R = \begin{pmatrix} 0 & 2,2 & 3,0 & 5,1 & 5,8 \\ 2,2 & 0 & 1,4 & 5,0 & 6,4 \\ 3,0 & 1,4 & 0 & 6,4 & 7,8 \\ 5,1 & 5,0 & 6,4 & 0 & 2,0 \\ 5,8 & 6,4 & 7,8 & 2,0 & 0 \end{pmatrix}.$$

Расстояние между кластерами $S_{1,2}$ и $S_{3,4,5}$, в которые входят соответственно первый-второй и третий-четвертый-пятый объекты, если исходить из метода «ближнего соседа», равно:

- а) 2,2;
- б) 3,0;
- в) 1,4;
- г) 2,0.

5.5 Расстояние между пятью объектами характеризуется матрицей расстояний

$$R = \begin{pmatrix} 0 & 2,2 & 3,0 & 5,1 & 5,8 \\ 2,2 & 0 & 1,4 & 5,0 & 6,4 \\ 3,0 & 1,4 & 0 & 6,4 & 7,8 \\ 5,1 & 5,0 & 6,4 & 0 & 2,0 \\ 5,8 & 6,4 & 7,8 & 2,0 & 0 \end{pmatrix}.$$

Расстояние между кластерами $S_{1,2}$ и $S_{3,4,5}$, в которые входят соответственно первый-второй и третий-четвертый-пятый объекты, если исходить из принципа «дальнего соседа», равно:

- а) 5,8;
- б) 5,1;

- в) 7,8;
г) 6,4.

5.6 Расстояние между пятью объектами характеризуется матрицей расстояний

$$R = \begin{pmatrix} 0 & 2,2 & 3,0 & 5,1 & 5,8 \\ 2,2 & 0 & 1,4 & 5,0 & 6,4 \\ 3,0 & 1,4 & 0 & 6,4 & 7,8 \\ 5,1 & 5,0 & 6,4 & 0 & 2,0 \\ 5,8 & 6,4 & 7,8 & 2,0 & 0 \end{pmatrix}.$$

Расстояние между кластерами $S_{1,2}$ и $S_{3,4,5}$, в которые входят соответственно первый-второй и третий-четвертый-пятый объекты, если исходить из принципа «средней связи», равно:

- а) 4,45;
б) 3,37;
в) 4,89;
г) 2,86.

5.7 Данные о четырех фирмах, деятельность которых характеризуется переменными $x^{(1)}$ и $x^{(2)}$, представлены в таблице:

i	1	2	3	4
$x_i^{(1)}$	1	7	1	9
$x_i^{(2)}$	5	9	3	7

Расстояние $\rho_E(1,2)$ между первым и вторым объектами, если в качестве метрики принять обычное евклидово расстояние, равно:

- а) 3,78;
б) 9,34;
в) 7,21;
г) 5,19.

5.8 Данные о четырех фирмах, деятельность которых характеризуется переменными $x^{(1)}$ и $x^{(2)}$, представлены в таблице:

i	1	2	3	4
$x_i^{(1)}$	1	7	1	9
$x_i^{(2)}$	5	9	3	7

Расстояние $\rho_{BE}(3,4)$ между третьим и четвертым объектами, если в качестве метрики принять взвешенное евклидово расстояние с весами $w_1 = 0,7$ и $w_2 = 0,3$, равно:

- а) 5,35;
б) 7,04;
в) 6,66;
г) 9,38.

5.9 Данные о четырех фирмах, деятельность которых характеризуется переменными $x^{(1)}$ и $x^{(2)}$, представлены в таблице:

i	1	2	3	4
$x_i^{(1)}$	1	7	1	9
$x_i^{(2)}$	5	9	3	7

Расстояние $\rho_{2,(1,3)}$ между вторым объектом и кластером $S_{1,3}$, включающим первый и третий объекты, если его измерять с помощью обычной евклидовой метрики $\rho_E(1,2)$ и принципа «центра тяжести», равно:

- а) 6,37;
- б) 7,81;
- в) 8,93;
- г) 5,72.

5.10 Данные о четырех фирмах, деятельность которых характеризуется переменными $x^{(1)}$ и $x^{(2)}$, представлены в таблице:

i	1	2	3	4
$x_i^{(1)}$	1	7	1	9
$x_i^{(2)}$	5	9	3	7

Расстояние между кластерами $S_{1,3}$ и $S_{2,4}$, если исходить и принципа «центра тяжести» и взвешенной евклидовой метрики с весами $w_1 = 0,2$ и $w_2 = 0,8$, равно:

- а) 7,45;
- б) 9,73;
- в) 5,39;
- г) 4,75.

5.11 Установите соответствие между характеристиками и формулами расчета:

- | | |
|--|--|
| а) обычное евклидово расстояние; | 1) $d(x_i, x_j) = \sum_{l=1}^k x_{il} - x_{jl} $ |
| б) взвешенное евклидово расстояние; | 2) $d(S_l, S_m) = d(\bar{X}_l, \bar{X}_m)$ |
| в) Хеммингово расстояние; | 3) $d(x_i, x_j) = \sqrt{\sum_{l=1}^k (x_{il} - x_{jl})^2}$ |
| г) расстояние Махаланобиса; | 4) $d(S_l, S_m) = \frac{1}{n_l n_m} \sum_{X_i \in S_l} \sum_{X_j \in S_m} d(X_i, X_j)$ |
| д) расстояние, измеряемое по принципу «дальнего соседа»; | 5) $d(X_l, X_j) = \sqrt{\sum_{l=1}^k w_l (x_{il} - x_{jl})^2}$ |
| е) расстояние, измеряемое по принципу «ближнего соседа»; | 6) $d_{\min}(S_l, S_m) = \min_{X_i \in S_l, X_j \in S_m} d(X_i, X_j)$ |
| ж) расстояние, измеряемое по принципу «средней связи»; | 7) $d_{\max}(S_l, S_m) = \max_{X_i \in S_l, X_j \in S_m} d(X_i, X_j)$ |
| з) расстояние, измеряемое по центрам тяжести групп. | 8) $d(X_i, X_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)}$ |

5.12 Процедура, при которой начальным является разбиение, состоящее из n одноэлементных классов, называется...

- а) дивизимной;
- б) корреляционной;
- в) агломеративной;
- г) линейной.

5.13 Кластерный анализ позволяет проводить...

- а) группировку объектов;

- б) группировку объектов и группировку признаков;
- в) группировку коэффициентов корреляции;
- г) группировку признаков.

5.14 Нормирование признаков производится с целью ...

- а) устранения влияния различных единиц измерения;
- б) уменьшения признакового пространства;
- в) упрощения расчетов;
- г) выделения латентных факторов.

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

- а) дивизимной;
- б) корреляционной;
- в) агломеративной;
- г) линейной.

5.16 Большинство программ с реализованным алгоритмом иерархической классификации, имею графическое представление результатов классификации в виде...

- а) дендрограммы;
- б) блок-схемы;
- в) дерева решений.

5.17 В задачах многомерной классификации объектов при $\alpha = \beta = -\delta = 1/2$ и $\gamma = 0$ расстояние между классами определяется по принципу...

- а) «дальнего соседа»;
- б) «средней связи»;
- в) «ближнего соседа».

Тест №6

6.1 Выявление аномальности k наибольших наблюдений предусматривает проверку на типичность ...

- а) k наблюдений последовательно в порядке уменьшения, начиная с максимального, с помощью теста Диксона;
- б) $(N - k)$ -го наблюдения в ряду упорядоченных по возрастанию наблюдений;
- в) одновременно всех k наибольших наблюдений с помощью теста Титъена-Мура;
- г) k наблюдений последовательно в порядке уменьшения, начиная с максимального, с помощью теста Смирнова-Граббса.

6.2 Определение аномальности одновременно k наибольших и наименьших наблюдений проводится путем проверки гипотезы о типичности ...

- а) упорядоченных по возрастанию модулей отклонений от среднего значения по тому же критерию, что и для k наибольших отклонений;
- б) упорядоченных по возрастанию модулей отклонений от среднего по тому же критерию, что и для k наименьших отклонений;
- в) упорядоченных по возрастанию модулей отклонений от среднего значения по тому же критерию, что и для k наибольших значений признака.

6.3 Возможным следствием маскирующего эффекта аномальных наблюдений является ...

- а) изменение результата проверки на типичность других «подозрительных» наблюдений;
- б) изменение результата проверки на типичность одновременно данного и других «подозрительных» наблюдений;
- в) увеличение числа «подозрительных» наблюдений;
- г) снижение числа «подозрительных» наблюдений.

6.4 Последствиями игнорирования гетероскедастичности при оценивании параметров регрессионной модели методом наименьших квадратов являются ...

- а) завышенные значения оценок коэффициентов при регрессорах;
- б) заниженные значения оценок коэффициентов при регрессорах;
- в) завышенные значения оценок дисперсий коэффициентов при регрессорах;
- г) заниженные значения оценок дисперсий коэффициентов при регрессорах.

6.5 Робастные МНК-оценки стандартных ошибок коэффициентов линейной модели регрессии при автокоррелированности случайных ошибок — это ...

- а) ошибки в форме Уайта;
- б) стандартные ошибки коэффициентов для классической линейной регрессионной модели;
- в) ошибки в форме Ньюи-Веста;
- г) стандартные ошибки коэффициентов, полученных методом максимума правдоподобия.

Выполнение лабораторных работ

1. Процедура выполнения лабораторных работ

Количество проводимых лабораторных работ в течение всего периода освоения дисциплины	8 работ
Формат проведения результатов	Бумажный / Электронный
Методические рекомендации (при необходимости)	Размещаются в электронной образовательной среде ВУЗа

2. Шкала оценивания с учетом срока сдачи

Критерии оценивания	Балл
Студент правильно выполнил задание работы, продемонстрировал знания теоретического и практического материала	Отлично
Студент правильно выполнил задание работы, но продемонстрировал неполноту знания теоретического и практического материала	Хорошо
Студент выполнил задание работы, но допустил значительные неточности при выполнении, продемонстрировал неполноту знания теоретического и практического материала	Удовлетворительно
Студент неправильно выполнил задание работы, не продемонстрировал знания теоретического и практического материала	Неудовлетворительно

3. Перечень лабораторных работ и вопросов при собеседовании

1. Предварительный анализ данных с использованием MS Excel: табличное и графическое представление данных, группировка, описательная статистика.
2. Проверка статистических гипотез с использованием MS Excel.
3. Корреляционный анализ в MS Excel.
4. Многофакторный регрессионный анализ. Оценка параметров модели линейной регрессии и ее качества средствами MS Excel.
5. Работа в программном пакете Gretl. Множественный регрессионный анализ в Gretl.
6. Модели с переменной структурой.
7. Снижение размерности и преобразование признакового пространства при построении регрессионной модели.
8. Факторный анализ методом главных компонент.

Примерный перечень вопросов для проведения собеседования

- 1) Назовите основные критерии классификации данных.
- 2) В чем состоит различие пространственных, временных и пространственно-временных данных?

- 3) Назовите основные типы переменных в зависимости от шкалы измерения.
- 4) Назовите основные числовые характеристики центра группирования количественных данных.
- 5) Какие показатели вариации количественных данных Вы можете назвать? В чем состоят различия между ними?
- 6) Что понимается под точечной и интервальной оценками?
- 7) Какие законы распределения выборочных характеристик используются для получения интервальных оценок?
- 8) Какая зависимость называется корреляционной?
- 9) Какие коэффициенты корреляции Вы знаете? В чем заключается различие между ними?
- 10) Какова цель расчета матрицы коэффициентов парной корреляции?
- 11) Как интерпретируется множественный коэффициент детерминации?
- 12) Какую информацию можно получить, анализируя корреляционное поле?
- 13) Что представляет собой каноническая корреляция и в чем состоит ее отличие от линейных коэффициентов корреляции?
- 14) На какой идее основан метод наименьших квадратов и для какой цели он используется?
- 15) Как интерпретируется значение коэффициента регрессии?
- 16) Перечислите свойства оценок случайных возмущений.
- 17) Каковы свойства МНК-оценок в классической модели линейной регрессии?
- 18) Что характеризуют элементы ковариационной матрицы?
- 19) Какая идея лежит в основе расчета коэффициента множественной детерминации в регрессионном анализе?
- 20) Для решения каких задач регрессионного анализа требуется условие нормальности регрессионных остатков?
- 21) Какова цель проведения компонентного анализа?
- 22) Опишите модель метода главных компонент.
- 23) Что представляют собой собственные векторы и собственные значения корреляционной матрицы, и как они могут быть использованы для получения матрицы весовых коэффициентов?
- 24) Дайте определение квадратичных форм и главных компонент. Укажите главные компоненты для двумерного, трехмерного и конечномерного пространства.
- 25) Как получают и для чего используют матрицы индивидуальных значений главных компонент?
- 26) Каковы свойства ортогональной матрицы собственных векторов в модели метода главных компонент?
- 27) В чем сущность регрессии на главные компоненты?
- 28) В чем состоит основная проблема применимости метода регрессии на главные компоненты?
- 29) Как определить относительный вклад m первых главных компонент в суммарную дисперсию?
- 30) Сколько главных компонент (факторов) следует выделять при снижении размерности признакового пространства?
- 31) Как проинтерпретировать выделенные главные компоненты (факторы)?
- 32) С какой целью проводится кластерный анализ?
- 33) Назовите способы представления информации для проведения кластерного анализа.
- 34) Как можно рассчитать расстояние между объектами (признаками)?
- 35) В каких случаях может быть использовано обычное евклидово расстояние?
- 36) Как можно рассчитать расстояние между объектами, если они представлены дихотомическими признаками?

- 37) Что представляет собой функционал качества разбиения, какова цель его использования?
- 38) Какие иерархические кластер-процедуры Вы знаете?
- 39) В чем состоит различие агломеративных и дивизимных кластер-процедур?
- 40) Что представляет собой дендрограмма, с какой целью она используется?
- 41) В каких случаях требуется нахождение робастных оценок параметров?
- 42) В чем состоит основной недостаток непосредственного применения правила «трех сигм» для выявления аномальных наблюдений?
- 43) Как выявить аномальные наблюдения с помощью критерия Диксона?
- 44) В каком случае критерий Титъена — Мура эквивалентен критерию Граббса при выявлении аномальных наблюдений?
- 45) Для чего производят цензурирование выборки?
- 46) Как получить робастные оценки параметров линейной регрессионной модели?

Курсовая работа

1. Процедура проведения

Этапы выполнения курсовой работы с указанием сроков выполнения:

1. Выдача задания (3 неделя семестра).
2. Ознакомление с темой работы и заданием (4 неделя семестра).
3. Изучение теории (5-6 неделя семестра).
4. Составление плана решения поставленной задачи (7-8 неделя семестра).
5. Составление и отладка программы, реализующей заданный алгоритм решения задачи (9-11 неделя семестра).
6. Проведение расчетов, получение выводов (12-13 неделя семестра).
7. Оформление пояснительной записки (14-15 неделя семестра).

2. Шкала оценивания с учетом срока сдачи

Критерии оценки уровня сформированности компетенций	Балл
Выставляется при выполнении работы в полном объеме; оформлении с соблюдением установленных правил; студент свободно владеет теоретическим материалом, безошибочно применяет его при практическом исследовании; применении адекватных методов и алгоритмов для выполнения задания; на все вопросы дает правильные и обоснованные ответы, убедительно защищает свою точку зрения.	Отлично
Выставляется при выполнении работы в полном объеме; оформлении с соблюдением установленных правил; студент твердо владеет теоретическим материалом, может применять его при практическом исследовании; использованы правильные методы и алгоритмы для выполнения задания работы с несущественными неточностями; на большинство вопросов даны правильные ответы, защищает свою точку зрения достаточно обоснованно.	Хорошо
Выставляется при выполнении работы в полном объеме, работа оформлена с соблюдением установленных правил; при выполнении работы без достаточно глубокой проработки вопросов применены правильные методы и алгоритмы для выполнения задания с существенными неточностями; на вопросы отвечает неуверенно или допускает ошибки, неуверенно защищает свою точку зрения.	Удовлетворительно
Выставляется, когда студент не может защитить свои решения, допускает грубые ошибки и искажение фактов при ответах на поставленные вопросы или не отвечает на них.	Неудовлетворительно

3. Варианты (темы) курсовой работы

Тематика курсовых работ по дисциплине определяется преподавателем кафедры. При этом выбор основывается как на государственном стандарте, так и на направлениях научно-исследовательской и учебно-методической работы, актуальных направлениях работы других организаций, деятельность которых связана с разработкой информационного и программного обеспечения.

При разработке тематики необходимо предложить и реализовать решение следующих задач:

- 1) Изучить теоретические основы анализа многомерных данных и составить реферат.
- 2) Произвести сбор статистических данных по актуальной проблеме, определить адекватные методы их анализа (включая методы, изученные в курсе) и применить их.
- 3) По всем пунктам сделать выводы и дать практические рекомендации.

Экзамен

1. Процедура проведения

Общее количество вопросов к экзамену (зачету с оценкой)	16 вопросов
Количество вопросов в билете	2 вопросов
Наличие задач в билете	1 задача
Формат проведения	Устно и письменно

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Обучающийся показал глубокие знания теоретического материала по поставленному вопросу, грамотно, логично и стройно его изложил, правильно решена задача	Отлично
Обучающийся твердо знает материал, грамотно его излагает, но допустил несущественные неточности в ответе на вопрос / при решении задачи	Хорошо
Обучающийся показал знание только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допустил отдельные неточности в ответе на вопрос / при решении задачи	Удовлетворительно
Обучающийся допустил грубые ошибки в ответе на поставленный вопрос / при решении задачи	Неудовлетворительно

3. Вопросы к экзамену

1. Основные критерии классификации наборов данных и виды классификации.
2. Основные числовые характеристики количественных данных.
3. Корреляционный анализ взаимосвязи количественных и качественных признаков.
4. Методы оценивания параметров регрессии. Оценивание параметров двумерного и множественного уравнения регрессии методом наименьших квадратов.
5. Свойства оценок параметров регрессии, полученных методом наименьших квадратов.
6. Регрессионные модели с переменной структурой.
7. Основные понятия и задачи снижения размерности признакового пространства.
8. Метод главных компонент.
9. Факторный анализ.
10. Многомерное шкалирование.
11. Кластерный анализ.
12. Дискриминантный анализ.
13. Метод канонических корреляций.
14. Методы выявления аномальных значений в выборочных данных.

15. Методы устойчивого параметрического оценивания.
 16. Основные подходы к оцениванию параметров при отсутствии информации о законе распределения значений признаков.

4. Задачи к экзамену

1. По выборке из 10 наблюдений роста (в см) отцов (X) и их взрослых сыновей (Y), представленной в таблице, найти выборочный коэффициент корреляции и проверить его значимость при $\alpha = 0,05$ (распределение случайных величин X и Y предполагается нормальным):

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

2. Две группы экспертов проанализировали 12 проектов с точки зрения их эффективности:

Группы экспертов	Номера проектов и присвоенные им ранги											
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12
Первая	1	3	4	2	5	6	12	7	8	9	10	11
Вторая	2	3	1	4	6	5	9	7	8	10	12	11

Рассчитайте коэффициент Спирмена и сделайте вывод о согласованности мнений экспертов.

3. По данным 15 предприятий отрасли рассчитан коэффициент корреляции $r_{x_1, x_2} = -0,7$ между себестоимостью продукции $x^{(1)}$ и производительностью труда $x^{(2)}$. Требуется:
 а) построить матрицу собственных значений и матрицу факторных нагрузок;
 б) определить дисперсию и относительный вклад первой главной компоненты в обобщенную дисперсию исходных признаков.

4. По выборке из 10 наблюдений роста (в см) отцов (X) и их взрослых сыновей (Y), представленной в таблице, рассчитать оценки параметров линейной регрессии, описывающей зависимость роста сыновей от роста отцов:

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

Дать интерпретацию полученным результатам.

5. Деятельность 10 промышленных предприятий региона охарактеризована двумя показателями $x^{(1)}$ — общие затраты на руль товарной продукции и $x^{(2)}$ — фондоотдача (см. табл.):

№ предприятия / показатели	1	2	3	4	5	6	7	8	9	10
$x^{(1)}$	0,92	0,72	0,83	0,81	0,82	0,93	0,84	0,89	0,89	0,95
$x^{(2)}$	0,51	0,59	1,03	1,21	0,63	0,68	0,57	1,52	1,04	0,99

Требуется:

- а) проранжировать предприятия по первой главной компоненте;
 б) отобразить объекты (предприятия) в пространстве двух главных компонент, отложив по оси ординат значения первого главного фактора, а по оси абсцисс — второго;
 в) дать интерпретацию полученным результатам.
6. Дана информация по шести объектам, каждый из которых характеризуется двумя признаками:

Номер объекта (i)	1	2	3	4	5	6
x_{i1}	5	6	5	10	11	10
x_{i2}	10	12	13	9	9	7

Проведите классификацию:

- используя обычное евклидово расстояние и метод «ближнего соседа»;
- используя обычное евклидово расстояние и метод «дальнего соседа»;
- используя обычное евклидово расстояние и метод «средней связи».

7. По иерархическому агломеративному алгоритму проведите классификацию четырех хозяйств, работа которых характеризуется показателями объема реализованной продукции $x^{(1)}$ — растениеводства и $x^{(2)}$ — животноводства (млн руб./га):

Номер хозяйства (i)	1	2	3	4
x_{i1}	1	7	1	9
x_{i2}	5	9	3	7

Постройте дендрограмму. Для расчета расстояния между объектами используйте взвешенное евклидово расстояние с весами $w_1 = 0,1$; $w_2 = 0,9$, а расстояние между кластерами измеряйте по принципу «дальнего соседа».

8. Данные о деятельности четырех фирм по показателям $x^{(1)}$ и $x^{(2)}$ представлены в таблице:

Номер фирмы (i)	1	2	3	4
x_{i1}	2	2	8	10
x_{i2}	3	5	9	7

Рассчитайте:

- расстояние между вторым и четвертым объектами ($d_E(2,4)$), исходя из обычного евклидова расстояния;
- расстояние между первым и третьим объектами ($d_{EE}(1,3)$), исходя из взвешенного евклидова расстояния с весами $w_1 = 0,8$; $w_2 = 0,2$.

9. Расстояния между тремя объектами равны $d_{1,2} = 7,8$; $d_{1,3} = 6,2$; $d_{2,3} = 2,4$. Постройте матрицу расстояний.

10. Дана информация о шести регионах по двум показателям:

Номер региона (i)	1	2	3	4	5	6
x_{i1}	35	31	32	36	30	34
x_{i2}	126	112	123	128	115	123

При решении задачи классификации регионов использован метод k -средних. Координаты первого и второго регионов были взяты в качестве центров группирования. В результате проведения первой итерации алгоритма регионы были разделены на группы следующим образом: I группа — №1, №3, №4, №6; II группа — №2, №5. Проведите вторую и (при необходимости) последующие итерации алгоритма для получения окончательного ответа данной задачи.