

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»

ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Кафедра теории вероятностей и математической статистики



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Прикладной многомерный статистический анализ»

| | |
|----------------------------|--|
| Направление подготовки: | 01.03.02 Прикладная математика и информатика |
| Профиль подготовки: | Статистика |
| Образовательная программа: | бакалавриат |
| Квалификация: | Академический бакалавр |
| Форма обучения: | <u>очная</u> , очно-заочная, заочная нужное подчеркнуть |

Донецк 2020

УТВЕРЖДАЮ:

Декан факультета математики
и информационных технологий
И. А. Моисеенко

«16» апреля 2020 г.


МП

Программа учебной дисциплины «Прикладной многомерный статистический анализ» составлена на основании Государственного образовательного стандарта высшего профессионального образования (ГОС ВПО) Донецкой Народной Республики (ДНР) по направлению подготовки 01.03.02 Прикладная математика и информатика, утвержденного приказом Министерства образования и науки ДНР от 04 апреля 2016 г. № 280; Порядка организации учебного процесса в образовательных организациях высшего профессионального образования Донецкой Народной Республики, утвержденного приказом Министерства образования и науки ДНР № 1171 от «10» ноября 2017 г.; учебного плана и основной образовательной программы высшего профессионального образования направления подготовки 01.03.02 Прикладная математика и информатика (профиль: Статистика), разработанных в ГОУ ВПО «Донецкий национальный университет».

Разработчик:
доцент кафедры теории вероятностей
и математической статистики

 А.В. Золотая

Программа учебной дисциплины утверждена на заседании кафедры теории вероятностей и математической статистики
Протокол № 14 от « 02 » апреля 2020 г.
Зам. заведующего кафедрой

 И.Л. Шурко

Программа учебной дисциплины одобрена учебно-методической комиссией факультета математики и информационных технологий
Протокол № 8 от «15» апреля 2020 г.

Председатель учебно-методической
комиссии факультета

 Л.И. Селякова

1. ОБЛАСТЬ ПРИМЕНЕНИЯ И МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ

Учебная дисциплина «Прикладной многомерный статистический анализ» относится к циклу вариативной части, по выбору студента, профессионального блока. Основывается на базе дисциплин: «Алгебра и геометрия», «Математический анализ», «Теория вероятностей и математическая статистика»

Является основой для изучения следующих дисциплин: «Дополнительные главы математической статистики», «Производственная практика (преддипломная, подготовка ВКР: дипломной работы)», «Государственная итоговая аттестация».

2. СТРУКТУРА ДИСЦИПЛИНЫ

| <i>Характеристика учебной дисциплины</i> | | |
|--|--|------------------------|
| Направление подготовки | 01.03.02 Прикладная математика и информатика | |
| Профиль | Статистика | |
| Образовательная программа | бакалавриат | |
| Квалификация | Академический бакалавр | |
| Количество содержательных модулей | 1 | |
| Дисциплина базовой / вариативной части образовательной программы | Профессиональный блок, вариативная часть | |
| Формы контроля (МК, экзамен, зачет) | 1 модульный контроль, 1 экзамен в 7 семестре | |
| Показатели | очная форма обучения | заочная форма обучения |
| Количество зачетных единиц (кредитов) | 4 | |
| Год подготовки | 4 | |
| Семестр | 7 | |
| Количество часов | 144 | |
| - лекционных | 32 | |
| - практических, семинарских | 32 | |
| - лабораторных | - | |
| - самостоятельной работы | 80 | |
| в т.ч. индивидуальное задание | - | |
| Недельное количество часов, | 9 | |
| в т.ч. аудиторных | 4 | |

3. ОПИСАНИЕ ДИСЦИПЛИНЫ

Цели и задачи

Целью изучения дисциплины является формирование у студентов базовых компетенций, позволяющих выработать обоснованные стратегические решения, сочетающие интуицию специалиста с тщательным анализом имеющейся информации. Дисциплина «Прикладной многомерный статистический анализ» является важным инструментом, обеспечивающим теоретическую и методологическую подготовку бакалавров данного направления, служит основой для разработки и совершенствования методов многомерного анализа, помогает определить взаимосвязь многомерных явлений и процессов и их закономерности.

Задачи изучения дисциплины: поиск, сбор, анализ и систематизация многомерных данных в экономике и управлении; применение статистического инструментария в исследовании многомерных совокупностей социально-экономических явлений и процессов; умение использовать в профессиональной деятельности основные многомерные статистические методы обработки и анализа данных наблюдений.

Требования к результатам освоения дисциплины. Процесс изучения дисциплины «Прикладной многомерный статистический анализ» направлен на формирование элементов следующих компетенций в соответствии с ГОС ВПО ДНР по направлению подготовки 01.03.02 Прикладная математика и информатика и основной образовательной программы высшего профессионального образования направления подготовки 01.03.02 Прикладная математика и информатики (Профиль: Статистика):

а) общекультурных (ОК): способностью к коммуникации в устной и письменной формах на русском и иностранном языках для решения задач межличностного и межкультурного взаимодействия (ОК-5); способностью работать в команде, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия (ОК-6); способностью к самоорганизации и самообразованию (ОК-7);

б) общепрофессиональных (ОПК): способностью использовать базовые знания естественных наук, математики и информатики, основные факты, концепции, принципы теорий, связанных с прикладной математикой и информатикой (ОПК-1); способностью приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии (ОПК-2); способностью к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям (ОПК-3);

в) профессиональных (ПК):

научно-исследовательская деятельность:

способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям (ПК-1); способностью понимать, совершенствовать и применять современный математический аппарат (ПК-2);

проектная и производственно-технологическая деятельность:

способностью работать в составе научно-исследовательского и производственного коллектива и решать задачи профессиональной деятельности (ПК-4); способностью к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения (ПК-7);

организационно-управленческая деятельность: способностью составлять и контролировать план выполняемой работы, планировать необходимые для выполнения работы ресурсы, оценивать результаты собственной работы (ПК-9).

В результате изучения учебной дисциплины студент должен:

знать: способы сбора и обработки многомерных данных; методики расчёта социально-экономических показателей; методы многомерного статистического анализа данных, необходимые для решения поставленных экономических задач;

уметь: собирать и обрабатывать многомерные данные с помощью различных статистических методов; выбирать инструментальные средства для обработки многомерных данных в соответствии с поставленной задачей; собирать, анализировать и интерпретировать необходимую информацию, содержащуюся в различных формах отчётности и прочих отечественных и зарубежных источниках.

владеть: навыками сбора и обработки необходимых многомерных данных; навыками

выбора и применения инструментальных средств для обработки многомерных данных; навыками многомерного статистического анализа и интерпретации информации, содержащейся в различных отечественных и зарубежных источниках; навыками выявления тенденций в развитии многомерных социально-экономических процессов; навыками интерпретации полученных в процессе многомерного статистического анализа результатов и формулирования выводов и рекомендаций.

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ И ФОРМЫ ОРГАНИЗАЦИИ УЧЕБНОГО ПРОЦЕССА

| Порядковый номер и тема | Краткое содержание темы |
|--|--|
| <i>Содержательный модуль 1</i> | |
| Тема 1. Предмет, цель, задачи и основные проблемы многомерного статистического анализа. Основы компонентного анализа. | Цель, задачи и основные проблемы многомерного статистического анализа. Предмет, метод и задачи дисциплины. Общая теория многомерных распределений. Понятие о множестве. Основные условия применения многомерного статистического анализа в социально-экономических исследованиях. Основные этапы многомерного статистического моделирования. Виды данных. Шкалы данных. Виды шкал. Количественные и качественные данные. Особенности анализа количественных и качественных показателей. Методы шкалирования при обработке качественных признаков. Экспертные оценки. |
| Тема 2. Методы снижения размерности. Основы факторного анализа | Линейная модель факторного анализа. Различие предпосылок компонентного и факторного анализа. Основные проблемы факторного анализа. Факторное отображение и факторная структура. Компоненты дисперсии в факторном анализе. Преобразование корреляционной матрицы в факторном анализе. Получение и использование матрицы нагрузок и матрицы индивидуальных значений. Метод главных факторов. Получение первого главного фактора. Использование методов вращения. Формирование названия общего фактора. Экономическая интерпретация результатов. |
| Тема 3. Классификация многомерных наблюдений. Кластерный анализ | Задача многомерной классификации объектов исследования. Классификация без обучения. Кластерный анализ. Методы кластерного анализа. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Классификация признаков на основе матриц коэффициентов статистической связи между ними. Иерархические кластер-процедуры. |
| Тема 4. Основные методы кластерного анализа | Метод К-средних. Классификация больших совокупностей объектов методами параллельных процедур. Классификация в пространстве главных компонент и общих факторов. Функционалы качества разбиения на классы. Зависимость выбора метода классификации от цели исследования. Классификация объектов (наблюдений) в социальных и экономических исследованиях. |

Тематический план

| Содержательный модуль 1 | | | | | | | | | | | |
|--|----------------------|--------|--------------|--------------|------------------------|-----------------------|------------------------|--------|--------------|--------------|------------------------|
| Названия содержательных модулей и тем | Количество часов | | | | | | | | | | |
| | Очная форма обучения | | | | | | Заочная форма обучения | | | | |
| | всего | В Т.Ч. | | | | | всего | В Т.Ч. | | | |
| | | лекции | практические | лабораторные | самостоятельная работа | индивидуальная работа | | лекции | практические | лабораторные | самостоятельная работа |
| Тема 1. Предмет, цель, задачи и основные проблемы многомерного статистического анализа. Основы компонентного анализа. | 45 | 10 | 10 | | 25 | | | | | | |
| Тема 2. Методы снижения размерности. Основы факторного анализа | 27 | 6 | 6 | | 15 | | | | | | |
| Тема 3. Классификация многомерных наблюдений. Кластерный анализ | 18 | 4 | 4 | | 10 | | | | | | |
| Тема 4. Основные метода кластерного анализа | 54 | 12 | 12 | | 30 | | | | | | |
| Итого по содержательному модулю 1 | 144 | 32 | 32 | | 80 | | | | | | |
| Всего по дисциплине | 144 | 32 | 32 | | 80 | | | | | | |

5. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРОВЕДЕНИЯ ЛЕКЦИОННЫХ, ПРАКТИЧЕСКИХ И ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Темы лекционных занятий

| № п/п | Название темы | Количество часов |
|----------|---|------------------|
| 1 | Введение в многомерный статистический анализ | 2 |
| 2 | Основы компонентного анализа. Нахождение главных компонент | 2 |
| 3 | Свойства главных компонент. Проблема количества главных компонент | 2 |
| 4 | Геометрическая интерпретация метода главных компонент. Выборочные главные компоненты | 2 |
| 5 | Проверка статистических компонентном анализе | 2 |
| 6 | Введение в факторный анализ. Каноничекая модель факторного анализа. | 2 |
| 7 | Сравнение компонентного и факторного анализа. Свойства факторного анализа. Решение факторной модели | 2 |

| | | |
|----|---|-----------|
| 8 | Проблема общности в факторном анализе | 2 |
| 9 | Общая постановка задачи кластерного анализа | 2 |
| 10 | Расстояния между отдельными объектами. Меры близости объектов | 2 |
| 11 | Общая характеристика методов кластерного анализа. Иерархический кластерный анализ | 2 |
| 12 | Параллельные и последовательные кластерные процедуры. Функционалы разбиения на кластеры | 2 |
| 13 | Оптимизационные процедуры. | 2 |
| 14 | Метод k-средних | 2 |
| 15 | Алгоритмы разрезания графа в кластерном анализе | 2 |
| 16 | Этапы проведения кластерного анализа | 2 |
| | ВСЕГО | 32 |

Темы (практических, лабораторных, семинарских) занятий

| № п/п | Название темы | Количество часов |
|------------------|---|-----------------------------|
| 1 | Введение в многомерный статистический анализ | 2 |
| 2 | Основы компонентного анализа. Нахождение главных компонент | 2 |
| 3 | Свойства главных компонент. Проблема количества главных компонент | 2 |
| 4 | Геометрическая интерпретация метода главных компонент. Выборочные главные компоненты | 2 |
| 5 | Проверка статистических компонентном анализе | 2 |
| 6 | Введение в факторный анализ. Каноническая модель факторного анализа. | 2 |
| 7 | Сравнение компонентного и факторного анализа. Свойства факторного анализа. Решение факторной модели | 2 |
| 8 | Проблема общности в факторном анализе | 2 |
| 9 | Общая постановка задачи кластерного анализа | 2 |
| 10 | Расстояния между отдельными объектами. Меры близости объектов | 2 |
| 11 | Общая характеристика методов кластерного анализа. Иерархический кластерный анализ | 2 |
| 12 | Параллельные и последовательные кластерные процедуры. Функционалы разбиения на кластеры | 2 |
| 13 | Оптимизационные процедуры. | 2 |
| 14 | Метод k-средних | 2 |
| 15 | Алгоритмы разрезания графа в кластерном анализе | 2 |
| 16 | Этапы проведения кластерного анализа | 2 |
| | ВСЕГО | 32 |

6. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Организация самостоятельной работы студентов

| <i>№ п/п</i> | <i>Название темы</i> | <i>Количество часов</i> |
|------------------|---|-----------------------------|
| 1 | Введение в многомерный статистический анализ | 5 |
| 2 | Основы компонентного анализа. Нахождение главных компонент | 5 |
| 3 | Свойства главных компонент. Проблема количества главных компонент | 5 |
| 4 | Геометрическая интерпретация метода главных компонент. Выборочные главные компоненты | 5 |
| 5 | Проверка статистических компонентном анализе | 5 |
| 6 | Введение в факторный анализ. Каноническая модель факторного анализа. | 5 |
| 7 | Сравнение компонентного и факторного анализа. Свойства факторного анализа. Решение факторной модели | 5 |
| 8 | Проблема общности в факторном анализе | 5 |
| 9 | Общая постановка задачи кластерного анализа | 5 |
| 10 | Расстояния между отдельными объектами. Меры близости объектов | 5 |
| 11 | Общая характеристика методов кластерного анализа. Иерархический кластерный анализ | 5 |
| 12 | Параллельные и последовательные кластерные процедуры. Функционалы разбиения на кластеры | 5 |
| 13 | Оптимизационные процедуры. | 5 |
| 14 | Метод k-средних | 5 |
| 15 | Алгоритмы разрезания графа в кластерном анализе | 5 |
| 16 | Этапы проведения кластерного анализа | 5 |
| | ВСЕГО | 80 |

Темы для выполнения самостоятельной домашней работы

1. Статистический анализ инновационной активности в России.
2. Статистический анализ патентной деятельности в России
3. Статистический анализ кооперационной активности в России
4. Статистический анализ инвестиционной деятельности в России
5. Статистический анализ миграционных процессов.
6. Статистический анализ структуры доходов домохозяйств.
7. Статистический анализ употребления спиртных напитков среди населения (женщин, мужчин, молодежи).
8. Статистический анализ использования населением информационных технологий (Интернета).
9. Статистический анализ структуры расходов домохозяйства.
10. Статистический анализ уровня образования населения.
11. Статистический анализ уровня рождаемости.
12. Статистический анализ уровня смертности.
13. Статистический анализ политических взглядов населения в Европе и России.
14. Статистический анализ состояния здоровья населения.
15. Статистический анализ уровня жизни населения.

16. Статистический анализ отношения населения к институту брака в России (в Европе).
17. Статистический анализ уровня брачности.
18. Статистический анализ уровня разводимости.
19. Статистический анализ демографической структуры семьи.
20. Статистический анализ жилищных условий домохозяйств в России.
21. Статистический анализ сберегательного поведения домохозяйств.
22. Анализ макроэкономических детерминант инновационного развития.
23. По согласованию любая тема по социально-экономическому развитию России или стран мира

7. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ

Образец индивидуального задания

1. Запишите модель факторного анализа и систему уравнений при количестве признаков $k = 4$ и количестве общих факторов $m = 2$. Сколько уравнений и неизвестных в этой системе?

2. Приведите пример ортогональной матрицы размерности $2 \cdot 2$.

При проведении факторного анализа исходные признаки были предварительно центрированы и нормированы. Чему равны вклады общих факторов и специфического фактора в дисперсию исходного признака $x^{(3)} = 0,6f^{(1)} + 0,7f^{(2)} + 0,3f^{(3)} + e^{(3)}$? Какова общность признака $x^{(3)}$?

3. Постройте матрицы расстояний и сходства для пяти точек $(1; 3)$, $(7; 5)$, $(-1; -1)$, $(-1; 0)$, $(-5; 5)$ приняв за меру такие расстояния:

- а) Евклида;
- б) Махаланобиса;
- в) Хэмминга;
- г) Чебышева.

4. Пусть задана матрица расстояний между объектами множества, состоящего из четырех элементов:

| Объекты | a | b | c | d |
|---------|-----|-----|-----|-----|
| a | 0 | 2,0 | 4,0 | 6,3 |
| b | 2,0 | 0 | 2,5 | 4,1 |
| c | 4,0 | 2,5 | 0 | 2,2 |
| d | 6,3 | 4,1 | 2,2 | 0 |

Проведите кластеризацию заданного множества объектов таким методом:

- а) одиночной связи;
- б) полной связи.

Постройте дендограмму для каждого метода. Сравните их.

5. Пусть задана матрица расстояний между объектами множества, состоящего из четырех элементов:

| Объекты | a | b | c | d | e |
|---------|-----|-----|-----|-----|-----|
| a | 0 | 2,0 | 4,0 | 6,3 | 2,1 |
| b | 2,0 | 0 | 2,5 | 4,1 | 5,4 |
| c | 4,0 | 2,5 | 0 | 2,2 | 3,6 |
| d | 6,3 | 4,1 | 2,2 | 0 | 4,8 |
| e | 2,1 | 5,4 | 3,6 | 4,8 | 0 |

6. Проведите кластеризацию заданного множества объектов таким методом: а) невзвешенного попарного среднего; б) взвешенного попарного среднего.

Постройте дендограмму для каждого метода. Сравните их.

Проведите кластеризацию множества объектов из предыдущей задачи таким методом:

а) невзвешенным центроидным; б) взвешенным центроидным.

Постройте дендограмму для каждого метода. Сравните их.

8. КОНТРОЛЬНЫЕ ВОПРОСЫ К ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

1. В чем заключается необходимость снижения размерности признакового пространства?
2. Каковы предпосылки, обуславливающие возможность снижения размерности признакового пространства?
3. Что понимается под «мерой информативности» в методе главных компонент?
4. Сформулировать определение k -й главной компоненты.
5. Сформулировать оптимизационную задачу для построения первой главной компоненты.
6. В чем заключается алгоритм нахождения коэффициентов линейного преобразования исходных признаков?
7. Каковы основные числовые характеристики главных компонент?
8. Линейная модель факторного анализа.
9. В чем состоят различия компонентного и факторного анализа?
10. Чем отличаются исходные признаки от общих факторов?
11. В чем различие общих и специфических факторов?

12. Единственно ли решение задачи факторного анализа? Ответ поясните.
13. Что такое редуцированная корреляционная матрица?
14. В каких пределах лежат диагональные элементы редуцированной корреляционной матрицы?
15. Какой вид имеет ковариационная матрица специфических факторов?
16. Методы расчета общностей.
17. Метод максимального правдоподобия решения задачи факторного анализа.
18. В чем состоит статистический смысл факторных нагрузок?
19. Центроидный метод решения задачи факторного анализа.
20. Метод главных осей решения задачи факторного анализа.
21. Метод минимальных остатков решения задачи факторного анализа.
22. Зачем проводить вращения факторного пространства?
23. В чем суть ортогонального вращения?
24. Методы ортогонального вращения факторного пространства.
25. Назовите этапы проведения факторного анализа.
26. Модель, математическое обоснование и алгоритм метода главных компонент.
27. Получение и использование матриц факторов, индивидуальных значений главных компонент.
28. Экономическая интерпретация главных компонент.
29. Метод главных факторов.
30. Регрессия на главные компоненты.
31. Линейная модель факторного анализа.
32. Экономическая интерпретация общих факторов.
33. Методы классификации без обучения.
34. Иерархические кластер-продукты.
35. Функционалы качества разбиения на классы.
36. Метод К-средних.
37. Классификация в пространстве главных компонент и общих факторов.
38. Применение многомерного статистического анализа в экономических исследованиях.
39. Интерпретация канонического коэффициента корреляции и соответствующих канонических величин.
40. Какие задачи решает кластерный анализ?
41. В чем состоит общая постановка задачи кластерного анализа?
42. Что является решением задачи кластерного анализа?
43. Какие подходы к кластерному анализу существуют?
44. Какие используют шкалы измерения показателей?
45. Зачем проводят нормирование признаков, какие есть способы его проведения?
46. Что такое меры сходства и расстояния между объектами?
47. Какие имеются меры сравнения объектов между собой?
48. Какие существуют основные методы близости кластеров?
49. Для каких задач обработки экспериментальных данных используются методы иерархического кластерного анализа?
50. В чем состоит принципиальное отличие между иерархическими и неиерархическими алгоритмами классификации?
51. Что такое дендрограмма?
52. Что представляют собой иерархические агломеративные методы кластерного анализа?
53. Что представляют собой иерархические дивизимные методы кластерного анализа?
54. Что такое функционал качества разбиения?
55. В чем состоит суть оптимизационных кластер-процедур?

56. Какие можно выделить отличия параллельных и последовательных кластер-процедур?
57. В чем заключается принцип действия алгоритма метода k -средних?
58. В чем состоит принцип действия алгоритма Форель?
59. В чем суть алгоритмов типа разрезания графа?
60. Какие этапы включает в себя кластерный анализ?

9. ОБРАЗЕЦ МОДУЛЬНОГО КОНТРОЛЯ

ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ» Факультет математики и информационных технологий

Направление подготовки: **01.03.02 Прикладная математика и информатика**
 Профиль: **Статистика**
 Программа подготовки: **бакалавриат**
 Семестр: **7**
 Учебная дисциплина: **Прикладной многомерный статистический анализ**

МОДУЛЬНАЯ КОНТРОЛЬНАЯ РАБОТА ВАРИАНТ №1

1. Центроидный метод решения задачи факторного анализа.
2. Модель, математическое обоснование и алгоритм метода главных компонент.
3. Метод k -средних.

Утверждено на заседании кафедры теории вероятностей и математической статистики, протокол № ____ от «__» _____ 20__ г.

Зам.заведующего кафедрой _____ И.Л.Шурко
 Преподаватель _____ А.В.Золотая

Критерии оценивания модульного контроля

| <i>Номер задания</i> | <i>Количество баллов</i> |
|----------------------|--------------------------|
| 1 | 20 |
| 2 | 15 |
| 3 | 15 |
| Всего | 50 |

10. ОБРАЗЕЦ ЭКЗАМЕНАЦИОННОГО БИЛЕТА

Теоретические вопросы к экзамену

1. В чем заключается необходимость снижения размерности признакового пространства?
2. Каковы предпосылки, обуславливающие возможность снижения размерности признакового пространства?
3. Что понимается под «мерой информативности» в методе главных компонент?
4. Сформулировать определение k -й главной компоненты.
5. Сформулировать оптимизационную задачу для построения первой главной компоненты.

6. В чем заключается алгоритм нахождения коэффициентов линейного преобразования исходных признаков?
7. Каковы основные числовые характеристики главных компонент?
8. Линейная модель факторного анализа.
9. В чем состоят различия компонентного и факторного анализа?
10. Чем отличаются исходные признаки от общих факторов?
11. В чем различие общих и специфических факторов?
12. Единственно ли решение задачи факторного анализа? Ответ поясните.
13. Что такое редуцированная корреляционная матрица?
14. В каких пределах лежат диагональные элементы редуцированной корреляционной матрицы?
15. Какой вид имеет ковариационная матрица специфических факторов?
16. Методы расчета общностей.
17. Метод максимального правдоподобия решения задачи факторного анализа.
18. В чем состоит статистический смысл факторных нагрузок?
19. Центроидный метод решения задачи факторного анализа.
20. Метод главных осей решения задачи факторного анализа.
21. Метод минимальных остатков решения задачи факторного анализа.
22. Зачем проводить вращения факторного пространства?
23. В чем суть ортогонального вращения?
24. Методы ортогонального вращения факторного пространства.
25. Назовите этапы проведения факторного анализа.
26. Модель, математическое обоснование и алгоритм метода главных компонент.
27. Получение и использование матриц факторов, индивидуальных значений главных компонент.
28. Экономическая интерпретация главных компонент.
29. Метод главных факторов.
30. Регрессия на главные компоненты.
31. Линейная модель факторного анализа.
32. Экономическая интерпретация общих факторов.
33. Методы классификации без обучения.
34. Иерархические кластер-продукты.
35. Функционалы качества разбиения на классы.
36. Метод К-средних.
37. Классификация в пространстве главных компонент и общих факторов.
38. Применение многомерного статистического анализа в экономических исследованиях.
39. Интерпретация канонического коэффициента корреляции и соответствующих канонических величин.
40. Какие задачи решает кластерный анализ?
41. В чем состоит общая постановка задачи кластерного анализа?
42. Что является решением задачи кластерного анализа?
43. Какие подходы к кластерному анализу существуют?
44. Какие используют шкалы измерения показателей?
45. Зачем проводят нормирование признаков, какие есть способы его проведения?
46. Что такое меры сходства и расстояния между объектами?
47. Какие имеются меры сравнения объектов между собой?
48. Какие существуют основные методы близости кластеров?
49. Для каких задач обработки экспериментальных данных используются методы иерархического кластерного анализа?

50. В чем состоит принципиальное отличие между иерархическими и неиерархическими алгоритмами классификации?
51. Что такое дендрограмма?
52. Что представляют собой иерархические агломеративные методы кластерного анализа?
53. Что представляют собой иерархические дивизимные методы кластерного анализа?
54. Что такое функционал качества разбиения?
55. В чем состоит суть оптимизационных кластер-процедур?
56. Какие можно выделить отличия параллельных и последовательных кластер-процедур?
57. В чем заключается принцип действия алгоритма метода k -средних?
58. В чем состоит принцип действия алгоритма Форель?
59. В чем суть алгоритмов типа разрезания графа?
60. Какие этапы включает в себя кластерный анализ?

**ГОУ ВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»
Факультет математики и информационных технологий**

Направление подготовки: **01.03.02 Прикладная математика и информатика**
 Профиль: **Статистика**
 Программа подготовки: **бакалавриат**
 Семестр: **7**
 Учебная дисциплина: **Прикладной многомерный статистический анализ**

БИЛЕТ №1

1. Классификация в пространстве главных компонент и общих факторов.
2. Линейная модель факторного анализа.
3. Какие этапы включает в себя кластерный анализ?
4. Применение многомерного статистического анализа в экономических исследованиях.

Утверждено на заседании кафедры теории вероятностей и математической статистики, протокол № ____ от « ____ » _____ 20__ г.

Зам. заведующего кафедрой
Экзаменатор

И.Л.Шурко
А.В.Золотая

Критерии оценивания экзамена

| Номер задания | Количество баллов |
|----------------------|--------------------------|
| 1 | 25 |
| 2 | 25 |
| 3 | 25 |
| 4 | 25 |
| Всего | 100 баллов |

11. ОБРАЗЕЦ ТЕСТОВОГО ЗАДАНИЯ

1. Матрица наблюдений многомерных векторов X имеет размерность 10×5 . Какова размерность оценки корреляционной матрицы R ?
 - а) 10×10 ;
 - б) 10×5 ;
 - в) 5×5 ;
 - г) 5×10 .
2. Матрица наблюдений многомерных векторов X имеет размерность 10×5 . Сколько собственных чисел и собственных векторов имеет оценка корреляционной матрицы R ?
 - а) 5 собственных чисел и 10 собственных векторов;
 - б) 10 собственных чисел и 5 собственных векторов;
 - в) 10 собственных чисел и 10 собственных векторов;
 - г) 5 собственных чисел и 5 собственных векторов.
3. Матрица наблюдений многомерных векторов X имеет размерность 10×5 . Чему равна сумма собственных чисел оценки корреляционной матрицы R ?
 - а) 1; б) 5; в) 10; г) сумме оценок дисперсий главных компонент.
4. В каком случае существует корреляция главных компонент?
 - а) при нормальном законе распределения многомерных векторов;
 - б) при коррелированности многомерных векторов;
 - в) корреляция отсутствует во всех случаях;
 - г) при размерности векторов более 10.
5. Метод главных компонент предназначен для
 - а) увеличения размерности модели;
 - б) фиксирования исходной размерности модели;
 - в) уменьшения размерности модели;
 - г) определения значимых коэффициентов регрессии.
6. Первая главная компонента соответствует направлению, вдоль которого дисперсия векторов исходного набора
 - а) максимальна; б) постоянна; в) минимальна; г) любая.
7. Что является оценками дисперсий главных компонент, если стандартизация исходных данных не проводилась?
 - а) собственные числа матрицы оценок ковариаций;
 - б) квадраты собственных чисел матрицы оценок ковариаций;
 - в) корни квадратные собственных чисел матрицы оценок ковариаций;
 - г) зависит от обстоятельств.
 - д) сумме оценок дисперсий главных компонент.
8. В каком случае существует корреляция главных компонент?
 - а) при нормальном законе распределения многомерных векторов;
 - б) при коррелированности многомерных векторов;
 - в) корреляция отсутствует во всех случаях;
 - г) при размерности векторов более 10.
9. Метод главных компонент предназначен для
 - а) увеличения размерности модели;

- б) фиксирования исходной размерности модели;
- в) уменьшения размерности модели;
- г) определения значимых коэффициентов регрессии.

10. Первая главная компонента соответствует направлению, вдоль которого дисперсия векторов исходного набора

- а) максимальна; б) постоянна; в) минимальна; г) любая.

11. Что является оценками дисперсий главных компонент, если стандартизация исходных данных не проводилась?

- а) собственные числа матрицы оценок ковариаций;
- б) квадраты собственных чисел матрицы оценок ковариаций;
- в) корни квадратные собственных чисел матрицы оценок ковариаций; г) зависит от обстоятельств.

11. Методом k -средних необходимо разделить множество объектов на два кластера. Количество кластеров, которое будет образовано на начальном этапе кластеризации, будет равно такому значению:

- а) 1; б) 2;
- в) 3; г) 4.

12. Методом k -средних необходимо разделить множество объектов на три группы. Количество кластеров, которое будет образовано по завершении кластеризации, будет равно такому значению:

- а) 1; б) 2;
- в) 3; г) 4.

13. Обязательно ли при использовании метода поиска сгущений необходимо указывать количество кластеров:

- а) да;
- б) нет;
- в) количество кластеров всегда равно двум;
- г) в зависимости от алгоритма.

12. КРИТЕРИИ ОЦЕНИВАНИЯ

По курсу предполагается проведение промежуточной аттестации в виде модульного контроля, выполнения домашних работ и экзамена. Экзамен сдают студенты с целью повышения рейтинга.

*Распределение баллов, которые могут получить студенты
в процессе изучения дисциплины*

| Организационно-учебная работа студента | СРС | | | Всего |
|--|-----------------|--------------------|----------------------------------|------------|
| | Домашняя работа | Модульный контроль | Индивидуальная творческая работа | |
| Мах 20 баллов | тах 20 баллов | тах 50 баллов | тах 10 баллов | 100 баллов |

Шкала соответствия баллов национальной шкале

| Оценка по шкале ECTS | Оценка по 100-балльной шкале | Оценка по государственной шкале (экзамен, дифференцированный зачет) | Оценка по государственной шкале (зачет) |
|----------------------|------------------------------|--|---|
| A | 90-100 | 5 (отлично) | зачтено |
| B | 80-89 | 4 (хорошо) | зачтено |
| C | 75-79 | 4 (хорошо) | зачтено |
| D | 70-74 | 3 (удовлетворительно) | зачтено |
| E | 60-69 | 3 (удовлетворительно) | зачтено |
| FX | 35-59 | 2 (неудовлетворительно) с возможностью повторной сдачи | не зачтено |
| F | 0-34 | 2 (неудовлетворительно) с возможностью повторной сдачи при условии обязательного набора дополнительных баллов | не зачтено |

13. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОГО ПРОЦЕССА

Для проведения лекционных занятий требуется аудитория на группу, оборудованная меловой или интерактивной доской, мультимедийным проектором и экраном, для практических занятий - компьютерная лаборатория.

14. РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА

| № п/п | Наименование | Кол-во экземпляров в библиотеке ДонНУ | Наличие электронной версии в ЭБС |
|----------------------------------|--|---------------------------------------|----------------------------------|
| Основная литература | | | |
| 1. | Многомерный статистический анализ в экономике: Учеб. Пособие для студентов вузов / Л.А.Сошникова, В.Н. Тамашевич, Г. Уебе, М. Шефер; Под ред. В.Н. Тамашевича. – М.: ЮНИТИ, 1999. – 598 с. | 6 | - |
| Дополнительная литература | | | |
| 2. | Гирко, В.Л. Многомерный статистический анализ:[Учеб. Пособие для вузов по специальностям «Математика» и «Прикл. математика»] / В.Л.Гирко, - К.: Выща шк., 1988.- 318 с. | 6 | - |
| 3. | Ермаков С. М. Метод Монте-Карло в вычислительной математике: Вводный курс / С. М. Ермаков. – СПб.: Невский Диалект; М.: БИНОМ. Лаборатория знаний, 2009. – 192 с. | 4 | - |
| 4. | Ермаков С. М. Статистическое моделирование / С. М. Ермаков, Г. А. Михайлов. – М.: Главная редакция физико-математической литературы издательства «Наука», 1982. – 296 с. | 3 | - |
| 5. | Соболь И. М. Численные методы Монте-Карло / И. М. Соболь. – М.: Главная редакция физико-математической литературы издательства «Наука», 1973. – 312 с. | 3 | - |

15. ИНФОРМАЦИОННЫЕ РЕСУРСЫ

1. www.donnu.ru – ГОУ ВПО «Донецкий национальный университет»
2. www.newlibrary.ru - новая электронная библиотека;
3. www.edu.ru – федеральный портал российского образования;
4. www.mathnet.ru – общероссийский математический портал;
5. www.elibrary.ru – научная электронная библиотека;
6. www.nehudlit.ru - электронная библиотека учебных материалов

16. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

1. Windows 7 PRO (корпоративная лицензия ДОННУ № 46484614);
2. Microsoft Office (корпоративная лицензия ДОННУ лицензия № 46472919);
3. Microsoft Visual Studio (лицензия программы Dream Spark для высших учебных заведений);
4. Лицензии GPL, Apache, BSD для свободного программного обеспечения: Free Lab, Scilab, R Studio, Python, Eclipse, Free Pascal, Tries Mode, Prolog, Антивирус Касперского, Linux Fedora, Libre Office, Adobe Acrobat Reader, xPDF, Blender, КОМПАС-3D LT, Paint.NET, Gimp.

Рабочая программа рассмотрена и переутверждена на заседании теории вероятностей и математической статистики с изменениями (без изменений) на 20____ год.

Протокол № ____ от «____» _____ 20____ г.

Заведующий кафедрой _____