

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ им. В.Ф. УТКИНА**

Кафедра «Автоматики и информационных технологий в управлении»

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДИСЦИПЛИНЫ**

**Б1.В.08 «МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ»**

Направление 01.03.02  
«Прикладная математика и информатика»

ОПОП  
«Программирование и анализ данных»

Квалификация выпускника – бакалавр

Формы обучения – очная

Рязань 2021 г.

Оценочные материалы – это совокупность учебно-методических материалов (контрольных заданий, описаний форм и процедур), предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной профессиональной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретенных компетенций обучающихся целям и требованиям основной профессиональной образовательной программы в ходе проведения текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности общекультурных, общепрофессиональных и профессиональных компетенций, приобретаемых обучающимся в соответствии с этими требованиями.

Контроль знаний проводится в форме текущего контроля и промежуточной аттестации.

Текущий контроль успеваемости проводится с целью определения степени усвоения учебного материала, своевременного выявления и устранения недостатков в подготовке обучающихся и принятия необходимых мер по совершенствованию методики преподавания учебной дисциплины (модуля), организации работы обучающихся в ходе учебных занятий и оказания им индивидуальной помощи.

К контролю текущей успеваемости относятся проверка знаний, умений и навыков, приобретенных обучающимися в ходе выполнения индивидуальных заданий на практических занятиях и лабораторных работах. При оценивании результатов освоения практических занятий и лабораторных работ применяется шкала оценки «зачтено – не зачтено». Количество лабораторных и практических работ и их тематика определена рабочей программой дисциплины, утвержденной заведующим кафедрой.

Результат выполнения каждого индивидуального задания должен соответствовать всем критериям оценки в соответствии с компетенциями, установленными для заданного раздела дисциплины.

Промежуточный контроль по дисциплине осуществляется проведением экзамена.

Форма проведения экзамена – письменный ответ по утвержденным экзаменационным билетам, сформулированным с учетом содержания учебной дисциплины. После выполнения письменной работы обучаемого производится ее оценка преподавателем и, при необходимости, проводится теоретическая беседа с обучаемым для уточнения экзаменационной оценки.

## Паспорт оценочных материалов по дисциплине

| № п/п | Контролируемые разделы (темы) дисциплины<br>(результаты по темам)                              | Код контролируемой компетенции (или её части) | Вид, метод, форма оценочного мероприятия |
|-------|--|---|--|
| 1     | 2  | 3   | 4  |
| 1     | <i>Тема 1</i><br>Объект, предмет и задачи МСА  | ПК-2.1, ПК-2.2,<br>ПК-3.1, ПК-3.2             | Экзамен                                  |
| 2     | <i>Тема 2</i><br>Характеристики многомерных случайных величин и основы корреляционного анализа | ПК-2.1, ПК-2.2,<br>ПК-3.1, ПК-3.2             | Экзамен                                  |
| 3     | <i>Тема 3</i><br>Основы компонентного и факторного анализа                                     | ПК-2.1, ПК-2.2,<br>ПК-3.1, ПК-3.2             | Экзамен                                  |
| 4     | <i>Тема 4</i><br>Основы классификационного анализа   | ПК-2.1, ПК-2.2,<br>ПК-3.1, ПК-3.2             | Экзамен                                  |
| 5     | <i>Тема 5</i><br>Основы кластерного анализа  | ПК-2.1, ПК-2.2,<br>ПК-3.1, ПК-3.2             | Экзамен                                  |

### Критерии оценивания компетенций (результатов)

- 1). Уровень усвоения материала, предусмотренного программой.
- 2). Умение анализировать материал, устанавливать причинно-следственные связи.
- 3). Ответы на вопросы: полнота, аргументированность, убежденность, умение
- 4). Качество ответа (его общая композиция, логичность, убежденность, общая эрудиция)
- 5). Использование дополнительной литературы при подготовке ответов.

Уровень освоения сформированности знаний, умений и навыков по дисциплине оценивается в форме бальной отметки:

**«Отлично»** заслуживает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные программой, усвоивший основную и знакомый с дополнительной литературой, рекомендованной программой. Как правило, оценка «отлично» выставляется студентам, усвоившим взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявившим творческие способности в понимании, изложении и использовании учебно-программного материала.

**«Хорошо»** заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в програм-

ме задания, усвоивший основную литературу, рекомендованную в программе. Как правило, оценка «хорошо» выставляется студентам, показавшим систематический характер знаний по дисциплине и способным к их самостоятельному пополнению и обновлению в ходе дальнейшей учебной работы и профессиональной деятельности.

«Удовлетворительно» заслуживает студент, обнаруживший знания основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по специальности, справляющийся с выполнением заданий, предусмотренных программой, знакомый с основной литературой, рекомендованной программой. Как правило, оценка «удовлетворительно» выставляется студентам, допустившим погрешности в ответе на экзамене и при выполнении экзаменационных заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя.

«Неудовлетворительно» выставляется студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных программой заданий. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение или приступить к профессиональной деятельности по окончании вуза без дополнительных занятий по соответствующей дисциплине.

## Типовые контрольные задания или иные материалы

### Вопросы к экзамену по дисциплине

1. Объект и предмет МСА. Матрица данных и задачи МСА.
2. Скалярные непрерывные случайные величины.
3. Несмещенные, состоятельные и эффективные оценки.
4. Доверительный интервал для  $m_x$  при известной дисперсии  $D_x$  гауссовой СВ.
5. Доверительный интервал для  $m_x$  при неизвестной дисперсии  $D_x$  гауссовой случайной величины. Доверительный интервал для  $D_x$  гауссовой случайной величины.
6. Двумерные случайные величины. Распределение вероятностей. Плотность вероятностей. Условная плотность вероятностей.
7. Двумерные случайные величины. Моменты случайной величины. Коэффициенты ковариации и корреляции случайной величины. Ковариационная и корреляционная матрицы.
8. Двумерная гауссова случайная величина.
9. Многомерные случайные величины. Распределение вероятностей. Плотность вероятностей.

10. Основные характеристики многомерной случайной величины.
  11. Выборочные оценки числовых характеристик многомерных случайных величин.
  12. Проверка значимости выборочных характеристик.
  13. Метод главных компонент. Модель представления и основное уравнение компонентного анализа. Сингулярное разложение прямоугольной вещественной матрицы.
  14. Метод главных компонент. Сингулярное разложение корреляционной матрицы. Матрица нагрузок и декоррелирующее преобразование МД.
  15. Метод главных компонент. Влад главных компонент в суммарную дисперсию.
  16. Метод главных компонент. Примеры.
  17. Факторный анализ. Модель и основное уравнение. Усечение матрицы нагрузок на главные компоненты.
  18. Факторный анализ. Вращение факторного пространства. Принципы Терстоуна.
  19. Факторный анализ. Методы кватримакс и варимакс.
  20. Факторный анализ. Примеры.
  21. Постановка задачи классификации данных. Расстояние между строками МД.
  22. Решение задачи классификации. Примеры на плоскости.
  23. Метод 1-ближайших соседей (k-Nearest Neighbors, KNN).
  24. Метод опорных векторов (support vector machine, SVM). Жесткий зазор.
  25. Метод опорных векторов (support vector machine, SVM). Мягкий зазор.
- Расширение пространства признаков.
26. Наивный байесовский метод.
  27. Использование деревьев решений.
  28. Иерархический кластерный анализ.
  29. Метод кластеризации «k-средних».
  30. Метод кластеризации ФорЭл.

## Лабораторный практикум

| № п/п | № темы дисциплины | Наименование лабораторной работы  | Трудоемкость, час |
|-------|-------------------|-----------------------------------|-------------------|
| 1     | 2                 | Многомерный корреляционный анализ | 6                 |
| 2     | 3                 | Компонентный анализ               | 5                 |
| 3     | 5                 | Кластерный анализ                 | 5                 |

### Типовые задачи для проверки знаний

#### Задачи компонентного анализа

**Задача 3.1.** Имеются следующие факторные нагрузки (табл. 3.17).

Таблица 3.17

| $x_i$ | $a_{i1}$ | $a_{i2}$ | $a_{i3}$ | $a_{i4}$ | $a_{i5}$ |
|-------|----------|----------|----------|----------|----------|
| 1     | 0,72     | 0,80     | 0,13     | -0,08    | -0,02    |
| 2     | 0,67     | -0,46    | -0,22    | 0,13     | 0,09     |
| 3     | 0,92     | 0,68     | 0,19     | 0,02     | 0,00     |
| 4     | 0,85     | -0,21    | -0,34    | 0,21     | 0,01     |
| 5     | 0,58     | -0,47    | -0,06    | -0,15    | -0,10    |

Требуется:

1) определить вклад каждой компоненты в суммарную дисперсию; 2) выделить главные компоненты; 3) объяснить экономический смысл проведенных расчетов.

**Задача 3.2.** Имеются следующие собственные значения семи компонент (табл. 3.18).

Таблица 3.18

| $G_j$       | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
|-------------|------|------|------|------|------|------|------|
| $\lambda_j$ | 3,21 | 2,60 | 1,96 | 1,12 | 0,85 | 0,44 | 0,17 |

Требуется:

1) определить долю вариации каждой компоненты, а также накопленную долю вариации;

2) используя сочетание критериев, ответить на вопрос, сколько компонент и какие следует выделить для нахождения существенных оценок?

**Задача 3.3.** Имеются семь компонент, собственные значения двух из которых больше единицы. Ниже приведены факторные нагрузки этих компонент (табл. 3.19).

Таблица 3.19

| Показатель                                       | $a_{i1}$ | $a_{i2}$ |
|--|----------|----------|
| Доля городского населения                        | 0,815    | 0,092    |
| Доля населения трудоспособного возраста          | 0,508    | 0,684    |
| Плотность населения                              | 0,463    | 0,142    |
| Денежные вклады населения в сберегательные кассы | 0,689    | - 0,304  |
| Денежные доходы населения                        | 0,874    | - 0,215  |
| Средний размер семьи                             | - 0,316  | 0,828    |
| Доля женщин                                      | - 0,387  | 0,792    |

Требуется определить вклад каждой компоненты в суммарную дисперсию и объяснить его экономический смысл.

**Задача 3.4.** В табл. 3.20 приведены следующие коэффициенты корреляции четырех показателей бытовой инфраструктуры.

Таблица 3.20

| Показатель | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|------------|-------|-------|-------|-------|
| $x_1$      | 1     | 0,92  | 0,86  | 0,78  |
| $x_2$      |       | 1     | 0,73  | 0,95  |
| $x_3$      |       |       | 1     | 0,84  |
| $x_4$      |       |       |       | 1     |

$x_1$  – обеспеченность городского населения жилой площадью;  
 $x_2$  – доля комплексно благоустроенного жилья;  
 $x_3$  – объем бытовых услуг на 1000 жителей;  
 $x_4$  – число телефонных аппаратов на 1000 жителей.

Требуется определить собственные значения вектора корреляционной матрицы и сделать необходимые экономические выводы.

**Задача 3.5.** По приведенным в задаче 3.4 данным требуется выделить главные компоненты и определить их факторные нагрузки.

**Задача 3.6.** По данным задачи 3.4 требуется определить значения главных компонент со следующими стандартизованными значениями переменных  $x_i$  (табл. 3.21):

Таблица 3.21

| $x_i$ | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> |
|-------|----------|----------|----------|----------|
| $x_1$ | 1        | 0,92     | 0,86     | 0,78     |
| $x_2$ |          | 1        | 0,73     | 0,95     |
| $x_3$ |          |          | 1        | 0,84     |
| $x_4$ |          |          |          | 1        |

**Задача 3.7.** Имеются следующие данные по 17 объектам и 6 признакам (табл. 3.22).

Требуется построить компонентную модель с собственными значениями  $\lambda_1 = 2,8$ ;  $\lambda_2 = 1,3$  и дать экономическую интерпретацию выделенных компонент.

Таблица 3.22

| Объект | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|--------|-------|-------|-------|-------|-------|-------|
| 1      | 8,2   | 0,6   | 37,3  | 15,5  | 121,4 | 46,1  |
| 2      | 9,9   | 0,5   | 33,1  | 15,9  | 128,6 | 54,5  |
| 3      | 9,9   | 0,7   | 32,7  | 17,9  | 160,7 | 39,4  |
| 4      | 10,3  | 0,6   | 56,1  | 20,4  | 119,3 | 54,7  |
| 5      | 11,7  | 0,8   | 32,5  | 15,9  | 91,9  | 45,3  |
| 6      | 11,2  | 0,8   | 55,2  | 21,7  | 137,4 | 30,5  |
| 7      | 14,3  | 0,8   | 47,2  | 20,9  | 88,0  | 41,8  |
| 8      | 14,2  | 0,8   | 47,6  | 18,9  | 138,2 | 30,6  |
| 9      | 15,3  | 0,7   | 51,8  | 20,6  | 184,6 | 43,1  |
| 10     | 12,9  | 1,1   | 33,6  | 22,2  | 12,6  | 29,2  |
| 11     | 13,4  | 6,8   | 57,7  | 30,8  | 146,0 | 76,6  |
| 12     | 16,8  | 0,9   | 60,3  | 4,0   | 200,6 | 55,5  |
| 13     | 14,6  | 0,8   | 56,2  | 33,0  | 203,1 | 36,4  |
| 14     | 13,8  | 0,8   | 68,7  | 29,9  | 473,3 | 39,2  |
| 15     | 16,0  | 0,9   | 61,3  | 28,3  | 190,4 | 46,3  |
| 16     | 17,3  | 0,9   | 76,3  | 30,9  | 163,4 | 39,5  |
| 17     | 19,5  | 1,0   | 81,7  | 28,9  | 223,0 | 50,5  |

### Задачи факторного анализа

**Задача 3.8.** Определите вклады общих факторов в дисперсию  $i$ -переменной при  $r_1 = 0,8G_1 + 0,5G_2 + 0,33G$ . Прокомментируйте смысл полученных оценок.

**Задача 3.9.** На основе парных коэффициентов корреляции (табл. 3.23) постройте редуцированную корреляционную матрицу, проведите разложение вариации каждой переменной на: а) обусловленную общим фактором; б) специфическую (характерную).

Таблица 3.23

| $x_i$ | 1 | 2    | 3    | 4    |
|-------|---|------|------|------|
| 1     | 1 | 0,65 | 0,72 | 0,44 |
| 2     |   | 1    | 0,58 | 0,86 |
| 3     |   |      | 1    | 0,35 |
| 4     |   |      |      | 1    |

**Задача 3.10.** Получены факторные нагрузки для двух взаимно некоррелированных факторов (табл. 3.24).

Таблица 3.24

| $x_i$ | $a_{i1}$ | $a_{i2}$ |
|-------|----------|----------|
| $x_1$ | 0,60     | - 0,10   |
| $x_2$ | 0,40     | - 0,40   |
| $x_3$ | 0,70     | 0,30     |
| $x_4$ | 0,70     | 0,40     |
| $x_5$ | 0,50     | - 0,50   |

Требуется:

1) определить факторную нагрузку и специфичность признаков  $x_i$ ; 2) найти суммарную общность.

**Задача 3.11.** По данным задачи 3.3 требуется определить, какую долю общей суммарной дисперсии объясняет межгрупповая (факторная) дисперсия, вклад каждого фактора в общую дисперсию.

**Задача 3.12.** Имеются два некоррелируемых фактора со следующими нагрузками (табл. 3.25).

Таблица 3.25

| $x_i$ | $a_{i1}$ | $a_{i2}$ |
|-------|----------|----------|
| $x_1$ | 0,40     | 0,70     |
| $x_2$ | 0,60     | -0,50    |
| $x_3$ | -0,30    | 0,80     |
| $x_4$ | 0,50     | 0,40     |
| $x_5$ | +0,70    | 0,50     |

Требуется преобразовать факторные нагрузки путем вращения системы координат на  $30^\circ$  ( $\sin 30^\circ = 0,5000$ ,  $\cos 30^\circ = 0,8660$ ) против часовой стрелки.

**Задача 3.13.** Имеются следующие собственные векторы и собственные значения корреляционной матрицы (табл. 3.26).

Таблица 3.26

|                   | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|-------------------|-------|-------|-------|-------|
| $\lambda_1 = 2,5$ | 2,32  | 1,74  | 0,56  | 1     |
| $\lambda_2 = 1,7$ | -1,12 | -0,46 | -0,17 | 2     |

Требуется определить факторные нагрузки  $a_{ij}$  и провести преобразование факторного решения путем вращения системы координат на  $45^\circ$  ( $\sin 45^\circ = \cos 45^\circ = 0,71$ ) против часовой стрелки.

**Задача 3.14.** Имеются следующие коэффициенты корреляции (табл. 3.27).

Таблица 3.27

| $x_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|-------|-------|-------|-------|-------|
| $x_1$ | x     | 0,78  | 0,92  | 0,86  |
| $x_2$ | 0,78  | x     | 0,67  | 0,76  |
| $x_3$ | 0,92  | 0,67  | x     | 0,82  |
| $x_4$ | 0,86  | 0,76  | 0,82  | x     |

Требуется определить общности и с помощью критерия триад проверить гипотезу о наличии одного общего фактора.

## Задачи кластерного анализа

**Задача 3.15.** Имеются следующие исходные данные (табл. 3.28).

Таблица 3.28

| Объект | Признак |       |       |       |
|--------|---------|-------|-------|-------|
|        | $x_1$   | $x_2$ | $x_3$ | $x_4$ |
| 1      | 13,8    | 0,8   | 68,7  | 29,9  |
| 2      | 16,0    | 0,9   | 61,3  | 28,3  |
| 3      | 15,3    | 0,7   | 51,8  | 20,6  |
| 4      | 9,9     | 0,7   | 32,7  | 17,9  |
| 5      | 16,8    | 0,9   | 60,3  | 4,0   |

Требуется построить матрицу евклидовых расстояний между объектами, стандартизуя приведенные данные по значениям следующих общих средних и среднеквадратических отклонений (табл. 3.29).

Таблица 3.29

| Показатель | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|------------|-------|-------|-------|-------|
| $x_i$      | 13,49 | 0,79  | 52,31 | 22,10 |
| $\sigma_i$ | 3,02  | 0,15  | 15,19 | 7,50  |

**Задача 3.16.** Требуется определить группы однородных объектов по данным задачи 3.15. Расчеты провести по центрографическому алгоритму «ближайшего соседа».

**Задача 3.17.** Требуется определить группы однородных объектов (кластеры) по центрографическому алгоритму «дальнего соседа» по данным задачи 3.15.

**Задача 3.18.** Имеются данные о 36 объектах, каждый из которых попарно характеризуется признаками  $x_1$  и  $x_2$ , и  $x_3$  и  $x_4$  (табл. 3.30). Требуется определить центрографическим методом существующие кластеры.

| Объект | Признак |       |       |       |     | Объект | Признак |       |       |       |     |
|--------|---------|-------|-------|-------|-----|--------|---------|-------|-------|-------|-----|
|        | $x_1$   | $x_2$ | $x_3$ | $x_4$ | $S$ |        | $x_1$   | $x_2$ | $x_3$ | $x_4$ | $S$ |
| 1      | 1,75    | 5,25  | 8     | 16    | 1   | 19     | 5,28    | 5,0   | 13    | 7     | 3   |
| 2      | 2,65    | 5,50  | 11    | 15    | 1   | 20     | 5,21    | 4,54  | 10    | 5     | 3   |
| 3      | 1,80    | 4,47  | 10    | 6     | 1   | 21     | 6,01    | 5,00  | 12    | 9     | 3   |
| 4      | 2,50    | 4,75  | 7     | 9     | 1   | 22     | 5,23    | 4,23  | 15    | 8     | 3   |
| 5      | 3,00    | 5,00  | 12    | 7     | 1   | 23     | 6,00    | 4,37  | 14    | 8     | 3   |
| 6      | 3,54    | 4,71  | 8     | 6     | 1   | 24     | 5,76    | 3,92  | 15    | 7     | 3   |
| 7      | 2,82    | 4,22  | 10    | 5     | 1   | 25     | 3,50    | 2,00  | 8     | 4     | 4   |
| 8      | 2,53    | 4,07  | 8     | 5     | 1   | 26     | 2,77    | 1,54  | 4     | 4     | 4   |
| 9      | 2,25    | 4,04  | 12    | 16    | 1   | 27     | 3,51    | 1,40  | 3     | 6     | 4   |
| 10     | 2,06    | 3,95  | 9     | 8     | 1   | 28     | 4,00    | 1,75  | 7     | 2     | 4   |
| 11     | 2,75    | 3,75  | 9     | 7     | 1   | 29     | 3,03    | 1,00  | 8     | 2     | 4   |
| 12     | 3,24    | 3,93  | 10    | 8     | 1   | 30     | 3,25    | 0,88  | 6     | 3     | 4   |
| 13     | 4,14    | 3,52  | 12    | 11    | 2   | 31     | 3,69    | 0,49  | 5     | 2     | 4   |
| 14     | 4,62    | 3,34  | 13    | 7     | 2   | 32     | 4,20    | 1,00  | 7     | 2     | 4   |
| 15     | 4,27    | 3,11  | 11    | 8     | 2   | 33     | 2,30    | 6,11  | 10    | 14    | —   |
| 16     | 3,78    | 2,92  | 15    | 6     | 2   | 34     | 5,21    | 1,18  | 13    | 4     | —   |
| 17     | 4,25    | 2,76  | 14    | 5     | 2   | 35     | 4,75    | 3,25  | 11    | 8     | —   |
| 18     | 4,53    | 2,60  | 14    | 5     | 2   | 36     | 3,13    | 2,50  | 7     | 3     | —   |

**Задача 3.19.** По данным задачи 3.18 из совокупности первых 18 объектов выделить второй кластер, используя центрографический алгоритм поиска однородных объектов, приняв за начальную точку координаты 15-го объекта.

**Задача 3.20.** По данным задачи 3.18 из совокупности вторых 18 объектов (объекты 19–36) выделить третий кластер, приняв за начальную точку координаты 28-го объекта.

**Задача 3.21.** Имеются следующие данные по 10 объектам

Таблица 3.31

| Признак | Объект |     |     |     |     |     |     |     |     |     |
|---------|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|         | 1      | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
| $x_1$   | 1      | 2,3 | 1,2 | 1,1 | 1,5 | 2,4 | 1,6 | 1,5 | 1,6 | 1,7 |
| $x_2$   | 2      | 8   | 2   | 1   | 5   | 8   | 3   | 4   | 4   | 5   |

Требуется выделить множество кластеров и обосновать предпочтительный метод кластеризации.