

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Антипова Наталья Викторовна
Должность: и.о. директора филиала
Дата подписания: 07.10.2024 10:20:28
Уникальный программный ключ:
fae5412acb1bf810dc69e6bc004ac45622b84b3a

Приложение 6
к основной профессиональной образовательной
программе по направлению подготовки
38.03.01 Экономика направленность
(профиль) программы Бизнес-статистика и аналитика

**Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования**

**«Российский экономический университет имени Г.В. Плеханова»
Улан-Баторский филиал РЭУ им. Г.В. Плеханова**



Одобрено
на заседании Совета Улан-Баторского
филиала РЭУ им. Г.В. Плеханова
протокол № 08 от «25» апреля 2024г.
Председатель совета
Н.В. Антипова

ОЦЕНОЧНЫЕ СРЕДСТВА

по учебной дисциплине **Б1.О. 08 МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ
МЕТОДЫ**

(наименование дисциплины в соответствии с учебным планом)

Направление подготовки 38.03.01 Экономика
(код) *(наименование направления)*

Направленность (профиль) программы "Бизнес-статистика и аналитика"

Уровень высшего образования Бакалавриат

Год начала подготовки 2024

Улан-Батор – 2024 г.

Оценочные материалы одобрены на заседании междисциплинарной кафедры
10.04.2024 г. протокол № 9

ОЦЕНОЧНЫЕ СРЕДСТВА

по учебной дисциплине **Многомерные статистические методы**

ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ И ЭТАПОВ ИХ ФОРМИРОВАНИЯ ПО ДИСЦИПЛИНЕ

| Формируемые компетенции (код и наименование компетенции) | Индикаторы достижения компетенций (код и наименование индикатора) | Результаты обучения (знания, умения) | Наименование контролируемых разделов и тем |
|--|--|---|---|
| <p>УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений;</p> | <p>УК-2.1. Понимает базовые принципы постановки задач и выработки решений;</p> | <p><i>УК-2.1. 3-1. Знает основные принципы и концепции в области целеполагания и принятия решений</i> <i>УК-2.1. 3-2. Знает методы генерирования альтернатив решений и приведения их к сопоставимому виду для выбора оптимального решения</i> <i>УК-2.1. 3-3. Знает природу данных, необходимых для решения поставленных задач</i> <i>УК-2.1. У-1. Умеет системно анализировать поставленные цели, формулировать задачи и предлагать обоснованные решения</i> <i>УК-2.1. У-2. Умеет критически оценивать информацию о предметной области принятия решений</i> <i>УК-2.1. У-3. Умеет использовать инструментальные средства для разработки и принятия решений</i></p> | <p>Тема 1. Введение. Теоретические основы многомерных статистических методов и их место в социально-экономических исследованиях.</p> <p>Тема 2. Методы первичной записи и представления многомерных данных.</p> <p>Тема 3. Многомерное признаковое пространство. Элементы математики, теории вероятностей и математической статистики в многомерных статистических методах.</p> <p>Тема 4. Робастное статистическое оценивание.</p> <p>Тема 5. Корреляционно-дисперсионный анализ (КДА) многомерных разнотипных данных.</p> |

| | | | |
|---|--|--|--|
| <p>ПК-1. Способен анализировать, обосновывать и выбирать решения.</p> | <p>ПК-1.2. Осуществляет оценку ресурсов, необходимых для реализации решений.</p> | <p><i>ПК-1.2. 3-1. Знает языки визуального моделирования</i> <i>ПК-1.2. 3-2. Знает теорию систем</i> <i>ПК-1.2. 3-3. Знает предметную область и специфику деятельности организации в объеме, достаточном для решения задач бизнес-анализа</i> <i>ПК-1.2. У-1. Умеет использовать техники эффективных коммуникаций</i> <i>ПК-1.2. У-2. Умеет определять связи и зависимости между элементами информации бизнес-анализа</i> <i>ПК-1.2. У-3. Умеет применять информационные технологии в объеме, необходимом для целей бизнес-анализа</i> <i>ПК-1.2. У-4. Умеет анализировать внутренние (внешние) факторы и условия, влияющие на деятельность организации</i></p> | <p>Тема 6. Методы и модели факторного анализа (ФА).</p> <p>Тема 7. Кластерный анализ (КА).</p> |
|---|--|--|--|

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ХАРАКТЕРИЗУЮЩИЕ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ

Полный перечень вопросов для проведения устного опроса студентов по дисциплине «Многомерные статистические методы»

Индикаторы достижения: УК-2.1, ПК-1.2

Тема 1. Введение. Теоретические основы многомерных статистических методов и их место в социально-экономических исследованиях

Вопросы для проведения опроса:

1. Что является предметом МСМ?
2. В чем заключаются особенности МСМ и их отличие от методов классической статистики?
3. Назовите основные этапы развития МСМ.
4. Перечислите методы МСМ и задачи, решаемые при помощи этих методов.
5. Что является методологической и теоретической основой МСМ?

Тема 2. Методы первичной записи и представления многомерных данных

Вопросы для проведения опроса:

1. Какие математические операции можно применять в интервальной шкале, порядковой шкале, номинальной шкале?
2. В каких шкалах измерены: номера телефонов, студенческих билетов, оценка знаний на экзаменах, цвет автомобилей?
3. Докажите, что результаты качественных измерений не являются элементами линейного (векторного) пространства.
4. Дайте по одному экономическому примеру: одномерной шкалы; двумерной шкалы; трехмерной шкалы.
5. Какие методы первичной записи многомерных данных вы знаете?

Тема 3. Многомерное признаковое пространство. Элементы математики, теории вероятностей и статистики в многомерных статистических методах

Вопросы для проведения опроса:

1. Что понимают под многомерной статистической гипотезой и какие характерные признаки для нее существуют?
2. Приведите примеры многомерных статистических гипотез.
3. Какого рода ошибки могут допускаться и чем определяется достоверность выводов при проверке многомерных статистических гипотез?
4. Будет ли статистической гипотеза:
 - а) о равенстве двух ковариационных матриц, представляющих связи признаков здоровья людей в двух группах: не имеющих хронической заболеваемости и с заболеваемостью?
 - б) о несущественности различий характеров двух человек? И если решать этот же вопрос относительно макроэкономического положения двух стран?
 - в) о том, что многомерная случайная величина $X^T=(X_1, X_2, X_3)$, представляющая совокупность предприятий, подчиняется нормальному закону распределения?

5. Какими критериями может быть проверена гипотеза о статистическом соответствии двух многомерных выборок?

Тема 4. Робастное статистическое оценивание

Вопросы для проведения опроса:

1. Что понимается под грубыми ошибками, и каковы причины их появления в статистической совокупности?
2. Какие существуют подходы при обработке грубых ошибок?
3. Назовите основные методы устойчивого оценивания параметров выборочной совокупности.
4. Приведите пример, в котором резко выделяющее наблюдение обязано своим появлением длинным хвостам распределения, иной модели для части данных, иной модели для всех данных.
5. Приведите примеры распознавания резко выделяющихся наблюдений, опираясь на специальные знания в конкретной области.

Тема 5. Корреляционно-дисперсионный анализ многомерных разнотипных данных

Вопросы для проведения опроса:

1. Приведите известные вам меры связи количественных и нечисловых многомерных данных.
2. В каких случаях целесообразно использовать вероятностные коэффициенты связи?
3. Приведите отличия дисперсионного и энтропийного коэффициента конкордации. В каких границах они изменяются, и что это означает?
4. Какая гипотеза применяется при проверке статистической значимости выборочных коэффициентов ранговой корреляции Спирмена и Кендалла для многомерных наблюдений?
5. Поясните критерий независимости для таблиц сопряженности.

Тема 6. Методы и модели факторного анализа (ФА)

Вопросы для проведения опроса:

1. Какие особенности должны быть учтены, если для проведения анализа выбирается метод главных компонент или метод главных факторов?
2. Чем отличаются методы ФА от метода главных компонент?
3. Чем объясняется возможность факторного отображения множества элементарных признаков?
4. Как определяется информативность главных компонент? Приведите примеры.
5. Поясните алгоритмическую схему метода Хотеллинга и его преимущества по сравнению с другими методами ФА.

Темы для проведения групповой дискуссии:

1. Что означает «простая структура» модели ФА? Зачем и какими средствами добиваются «простой структуры»?
2. Что подразумевается под ортогональным и косоугольным факторными решениями?
3. Назовите методы уточнения моделей ФА.
4. Какие статистические критерии проверки надежности решений вы знаете?
5. В чем заключается использование коэффициента конгруэнтности и подхода Хармана в оценке качества факторных решений?

Тема 7. Кластерный анализ (КА)

Вопросы для проведения опроса:

1. В чем состоит принципиальное отличие методов многомерных классификаций от комбинационных группировок?
2. Перечислите задачи, решаемые при помощи методов КА.
3. Назовите две основные группы методов КА и укажите их сходство и различие.
4. Какие меры сходства используются при проведении многомерной классификации?
5. Как оценивается качество полученного разбиения многомерной совокупности на кластеры?

Задания для текущего контроля

Индикаторы достижения: УК-2.1, ПК-1.2

Тема 2. Методы первичной записи и представления многомерных данных

Задание 1. В таблице приведены данные о количестве студентов курса, отсутствующих на занятиях, за 20 учебных дней.

| | | | | | | | | | | | | | | | | | | | | |
|--------|---|---|----|---|----|---|---|----|---|----|----|----|----|----|----|----|----|----|----|----|
| День | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Кол-во | 5 | 0 | 15 | 1 | 23 | 6 | 5 | 18 | 8 | 10 | 2 | 10 | 6 | 0 | 0 | 11 | 2 | 13 | 6 | 3 |

Представьте информацию в виде таблицы частот, принимая следующую шкалу: хорошая посещаемость, если отсутствовало на занятиях меньше 4 студентов; средняя посещаемость – при числе отсутствующих от 4 до 9; плохая посещаемость – при числе отсутствующих от 10 до 15; очень плохая посещаемость – при числе отсутствующих больше 15;

Задание 2. В таблице частот приведены данные о посещаемости студентами курса за 20 учебных дней.

| Событие | Кол-во отсутствующих | Число дней (эмпирическая частота) |
|----------------------------|----------------------|-----------------------------------|
| Очень хорошая посещаемость | 0–3 | 7 |
| Хорошая | 4–8 | 5 |
| Удовлетворительная | 9–13 | 4 |
| Плохая | 14–20 | 3 |
| Очень плохая | Более 21 | 1 |

Определите центр группирования и вариабельность показателя посещаемости.

Задание 3. В таблице частот приведены данные о количестве бракованных изделий, произведенных учеником художественной мастерской, за 20 дней.

| Событие | Кол-во бракованных изделий | Число дней (эмпирическая частота) |
|------------------------|----------------------------|-----------------------------------|
| Очень хорошее качество | 0–1 | 7 |
| Хорошее качество | 2–4 | 5 |
| Удовлетворительное | 5–7 | 4 |
| Плохое | 8–10 | 3 |
| Очень плохое | Более 10 | 1 |

Определите центр группирования и вариабельность показателя качества продукции.

Задание 4. Представьте данные анкеты в виде таблицы логического описания, если первый респондент по первому вопросу выбрал ответ №6, по второму вопросу – ответ №3; второй респондент по первому вопросу выбрал ответ №2, по второму вопросу – ответ №5; третий респондент по первому вопросу выбрал ответ №1, по второму вопросу – ответ №3.

Анкета:

1. Какой цвет автомобиля вы предпочитаете (выберите один из предложенных вариантов)?:

- (1) черный (2) белый
 (3) красный (4) синий
 (5) желтый (6) оранжевый

2. Какая цена для вас приемлема при покупке автомобиля (тыс. \$)?

- (1) 3,5 – 4 (2) 4,5 – 5 (3) 5,5 – 6,5
 (4) 6,5–10 (5) 10–15 (6) более 15

Задание 5. Принимая следующее правило оцифровки экспертных мнений: если объект $O_i \succ O_j$, то $a_{ij} = 1$, $a_{ji} = -1$ и если $O_i \propto O_j$, то $a_{ji} = a_{ij} = 0$, была составлена таблица парных сравнений оценки эксперта пяти объектов.

| | O_1 | O_2 | O_3 | O_4 | O_5 |
|-------|-------|-------|-------|-------|-------|
| O_1 | | 1 | 0 | 1 | 1 |
| O_2 | -1 | | 0 | -1 | 0 |
| O_3 | 0 | 0 | | 1 | -1 |
| O_4 | -1 | 1 | -1 | | 0 |
| O_5 | -1 | 0 | 1 | 0 | |

Какой объект является наилучшим по мнению эксперта?

Тема 3. Многомерное признаковое пространство. Элементы математики, теории вероятностей и статистики в многомерных статистических методах

Задание 1. В таблице приводятся данные о фонде заработной платы работников централизованной бухгалтерии x_1 и товарообороте обслуживаемых аптек x_2 :

| | | | | | | | | | |
|----------------------------------|-------|-----|-----|-----|-----|------|-----|-----|-----|
| Фонд заработной платы, тыс. руб. | x_1 | 20 | 36 | 28 | 51 | 70 | 45 | 30 | 56 |
| Товарооборот, млн. руб. | x_2 | 3,5 | 5,4 | 2,7 | 9,8 | 10,1 | 6,2 | 2,4 | 9,5 |

Найдите оценки математического ожидания, дисперсии и коэффициента корреляции двумерной случайной величины.

Задание 2. В таблице приводятся данные о фонде заработной платы работников централизованной бухгалтерии x_1 и товарообороте обслуживаемых аптек x_2 :

| | | | | | | | | | |
|----------------------------------|-------|-----|-----|-----|-----|------|-----|-----|-----|
| Фонд заработной платы, тыс. руб. | x_1 | 20 | 36 | 28 | 51 | 70 | 45 | 30 | 56 |
| Товарооборот, млн. руб. | x_2 | 3,5 | 5,4 | 2,7 | 9,8 | 10,1 | 6,2 | 2,4 | 9,5 |

Найдите доверительную область для вектора математических ожиданий с надежностью $\gamma = 0,95$.

Задание 3. По данным годовых отчетов $n=10$ промышленных предприятий с уровнем значимости $\alpha = 0,05$ проверить гипотезу о соответствии средних уровней экономических показателей работы группы предприятий (x_1 млн. руб. – объем валовой продукции, x_2 тыс.руб./чел. – производительность труда) контрольным значениям при известных значениях указанных в таблице параметров генеральной совокупности, которая является нормально распределенной. Подберите наиболее мощный критерий для проверки гипотезы.

| №, i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | σ_j^2 | Парная корреляция | Контрольные значения μ_{0j} |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------------|----------------------------|---------------------------------|
| x_{1i} | 3 | 4 | 5 | 5 | 5 | 5 | 6 | 7 | 10 | 10 | 4 | $\rho_{12} = 0,14\sqrt{5}$ | 5 |
| x_{2i} | 1,2 | 1,2 | 1,4 | 1,2 | 1,2 | 1,5 | 1,5 | 1,3 | 1,7 | 1,6 | 0,2 | | 1,5 |

Задание 4. Проверьте существенность различий уровня эффективности работы предприятий двух отраслей «А» и «Б», если в первой отрасли по 6 предприятиям рентабельность производства составила 14% годовых, а среднедневная выработка на одного работника – 4,2 тыс. руб., во второй отрасли по 8 предприятиям соответственно: 9% и 3.5 тыс. руб. Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид:

$$S_* = \begin{pmatrix} 5,4 & 0,6 \\ 0,6 & 2,0 \end{pmatrix}.$$

Задание 5. Проверьте предположение о равенстве двух ковариационных матриц, представляющих связи экономических показателей по 30 предприятиям:

$$S_1 = \begin{pmatrix} 60 & 18 \\ 18 & 36 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 84 & 20 \\ 20 & 41 \end{pmatrix}.$$

Тема 4. Робастное статистическое оценивание

Задание 1. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Определите наличие грубых ошибок критерием Смирнова-Граббса.

Задание 2. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Определите наличие грубых ошибок критерием Граббса.

Задание 3. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Определите наличие грубых ошибок критерием Титъена-Мура.

Задание 4. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Рассчитайте устойчивые средние по методу Пуанкаре.

Задание 5. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Рассчитайте устойчивые средние по методу Винзора.

Тема 5. Корреляционно-дисперсионный анализ многомерных разнотипных данных

Задание 1. При статистическом анализе двумерной случайной величины $\xi = (\xi^1, \xi^2)$ удается наблюдать лишь ее «зашумленные» значения, т.е. в действительности мы имеем дело со случайной величиной $\tilde{\xi} = (\tilde{\xi}^1, \tilde{\xi}^2)$, где $\tilde{\xi}^1 = \xi^1 + \delta^1$ и $\tilde{\xi}^2 = \xi^2 + \delta^2$, случайные ошибки δ^1 и δ^2

имеют нулевые средние, конечные дисперсии $D\delta^1 = \sigma_1^2$, $D\delta^2 = \sigma_2^2$, взаимно независимы и не зависят от ξ^1 и ξ^2 . Докажите, что $|r(\xi^1, \xi^2)| > |r(\tilde{\xi}^1, \tilde{\xi}^2)|$, т.е., что теснота парной корреляционной связи между случайными величинами ослабевает при аддитивном наложении на них случайных ошибок.

Задание 2. Получена корреляционная матрица взаимосвязи следующих экономических показателей:

- производительность труда, x_1 ;
- премии и вознаграждения на одного работника, x_2 ;
- среднегодовая численность ППП, x_3 ;
- среднегодовая стоимость ОПФ, x_4 ;
- среднегодовой фонд заработной платы ППП, x_5 ;
- оборачиваемость нормируемых оборотных средств, x_6 .

$$\hat{R} = \begin{pmatrix} 1,00 & 0,23 & 0,48 & 0,44 & 0,43 & -0,09 \\ & 1,00 & 0,32 & 0,38 & 0,34 & -0,26 \\ & & 1,00 & 0,79 & 0,92 & -0,21 \\ & & & 1,00 & 0,63 & -0,29 \\ & & & & 1,00 & -0,19 \\ & & & & & 1,00 \end{pmatrix}.$$

- а) сравните матрицы парных и частных коэффициентов корреляции;
- б) объясните разницу двух корреляционных матриц;
- в) для всех признаков вычислите выборочные значения множественного коэффициента корреляции $\hat{R}_{i,1..i-1i+1..6}$;
- г) проверьте значимость вычисленных множественных коэффициентов корреляции;
- д) определите признак, значение которого в наибольшей степени объясняется значением оставшихся признаков.

Задание 3. Группа из 8 предварительно отобранных абитуриентов проходила оценочные тесты. Первый тест на знание основ проведения экономико-плановых расчетов и вычислений, второй тест на умение логически излагать свои мысли. По результатам проведенного тестирования абитуриентам были присвоены следующие баллы:

| | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|
| 1й тест | 12 | 14 | 15 | 15 | 18 | 19 | 20 | 23 |
| 2й тест | 16 | 19 | 21 | 15 | 22 | 17 | 13 | 21 |

Определите наличие связи между знаниями и логикой у абитуриентов используя коэффициент ранговой корреляции Спирмена.

Задание 4. Группа из 8 предварительно отобранных абитуриентов проходила оценочные тесты. Первый тест на знание основ проведения экономико-плановых расчетов и вычислений, второй тест на умение логически излагать свои мысли. По результатам проведенного тестирования абитуриентам были присвоены следующие баллы:

| | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|
| 1й тест | 12 | 14 | 15 | 15 | 18 | 19 | 20 | 23 |
| 2й тест | 16 | 19 | 21 | 15 | 22 | 17 | 13 | 21 |

Определите наличие связи между знаниями и логикой у абитуриентов используя коэффициент ранговой корреляции Кендалла.

Задание 5. Предположим, что X и Y – независимые выборки. Чему равен, рассчитанный для них коэффициент корреляции Спирмена?

$$\hat{\rho}^S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i^x - R_i^y)^2$$

Тема 6. Методы и модели факторного анализа (ФА)

Задание 1. По десяти моделям автомобилей получены данные о мощности двигателя и средней рыночной цене.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Мощность (л.с.) | 129 | 152 | 190 | 86 | 107 | 136 | 105 | 102 | 122 | 90 |
| Цена (тыс. р.) | 647 | 790 | 890 | 424 | 423 | 812 | 712 | 529 | 699 | 461 |

Вычислите главные компоненты.

Задание 2. На основе наблюдений получена следующая корреляционная матрица:

$$\hat{R} = \begin{pmatrix} 1,00 & 0,38 & 0,44 & 0,52 \\ 0,38 & 1,00 & 0,91 & 0,67 \\ 0,44 & 0,91 & 1,00 & 0,02 \\ 0,52 & 0,67 & 0,02 & 1,00 \end{pmatrix}$$

Вычислите общности:

- методом наибольшей корреляции;
- методом Барта;
- методом триад.

Задание 3. По данным опроса практиков-экономистов построена матрица корреляционной зависимости R характерных признаков: x_1 – уровень оплаты труда, x_2 – возраст, x_3 – трудовой стаж

| Признак | x_1 | x_2 | x_3 |
|---------|--------|--------|-------|
| x_1 | 1 | -0,388 | 0,665 |
| x_2 | -0,388 | 1 | 0,740 |
| x_3 | 0,665 | 0,740 | 1 |

а) Проведите анализ данных матрицы парных корреляций R методом главных компонент, определите уровень информативности каждой из главных компонент и ее признаковый состав. Покажите распределение элементарных признаков в пространстве двух первых главных компонент (F_1 и F_2).

б) Найдите факторное решение методом главных факторов (для получения собственных чисел достаточно возведение матрицы парных корреляций в четвертую степень).

Задание 4. Имеется матрица значений общих факторов по шести типам промышленных товаров:

| | | |
|-----------|--------|--------|
| | F_1 | F_2 |
| n_1 | -2,245 | -0,711 |
| n_2 | 0,499 | 1,680 |
| $F = n_3$ | -0,402 | 0,060 |
| n_4 | 1,155 | 0,524 |
| n_5 | 1,048 | 0,675 |
| n_6 | -0,055 | -2,228 |

Известны названия главных факторов: F_1 – качество товара, F_2 – уровень обслуживания покупателя.

а) Покажите распределение товаров в пространстве двух общих факторов. Можно ли выделить однородные по F_1 и F_2 группы товаров?

б) Постройте линейную регрессионную модель, в которой y – это удельный вес продаж каждого вида товара из всей партии в течение первых 15 дней реализации:

$$y = (0,10 \ 0,42 \ 0,13 \ 0,82 \ 0,63 \ 0,21),$$

факторные признаки регрессионной модели: F_1 и F_2 .

Задание 5. По матрице факторных нагрузок A воспроизведите матрицу парных корреляций R^+ :

$$A = \begin{pmatrix} 0,76 & 0,42 \\ 0,45 & 0,21 \\ 0,65 & 0,37 \\ 0,38 & 0,19 \end{pmatrix}$$

Тема 7. Кластерный анализ (КА)

Задание 1. С использованием стандартного пакета StatGraphics на основании комплекса показателей, приведенных в таблице, проведите классификацию предприятий по методу поиска сгущений. Проанализируйте полученные результаты и оцените степень однородности кластеров на основании внутрикластерных дисперсий.

| Номер предприятия | Удельный вес продукции, изготовленной по новым технологиям, % | Коэффициент использования сырья и материалов, % | Удельный вес продукции высшего качества |
|-------------------|---|---|---|
| 1 | 45,0 | 60,3 | 15,7 |
| 2 | 38,5 | 64,1 | 11,8 |
| 3 | 55,3 | 88,9 | 14,0 |
| 4 | 24,0 | 44,3 | 16,1 |
| 5 | 37,8 | 45,7 | 12,5 |
| 6 | 30,5 | 71,0 | 11,4 |
| 7 | 50,0 | 90,6 | 12,8 |
| 8 | 25,5 | 68,9 | 15,7 |
| 9 | 36,1 | 72,3 | 13,0 |
| 10 | 41,7 | 66,4 | 16,5 |

Задание 2. Проведите кластерный анализ объектов по одному признаку методом ближнего соседа (метрика – Евклидово расстояние):

| Объект | Значение признака |
|--------|-------------------|
| 1 | 2 |
| 2 | 1 |
| 3 | 5 |
| 4 | 8 |

Задание 3. Имеются данные по пяти предприятиям, каждое из которых характеризуется тремя переменными:

x_1 – средняя стоимость основных производственных фондов, млрд. руб.;

x_2 – материальные затраты на 1 руб. произведенной продукции, коп.;

x_3 – объем произведенной продукции, млрд. руб.

| № п/п | x_1 | x_2 | x_3 |
|-------|-------|-------|-------|
| 1 | 12,0 | 9,4 | 16,0 |
| 2 | 8,5 | 7,5 | 9,2 |
| 3 | 14,0 | 8,1 | 12,0 |
| 4 | 7,8 | 7,5 | 8,6 |
| 5 | 7,0 | 7,6 | 10,4 |

Определите матрицу расстояний используя метрику «Сити-блок».

Задание 4. Имеются данные по пяти предприятиям, каждое из которых характеризуется двумя переменными:

x_1 – средняя стоимость основных производственных фондов, млрд. руб.;

x_2 – объем произведенной продукции, млрд. руб.

| № п/п | x_1 | x_2 |
|-------|-------|-------|
| 1 | 12,0 | 16,0 |
| 2 | 8,5 | 9,2 |
| 3 | 14,0 | 12,0 |
| 4 | 7,8 | 8,6 |
| 5 | 7,0 | 10,4 |

Определите матрицу расстояний используя метрику «Евклидово расстояние» предварительно проведя нормировку данных методом стандартизации.

Задание 5. Имеются данные по пяти предприятиям, каждое из которых характеризуется двумя переменными:

x_1 – средняя стоимость основных производственных фондов, млрд. руб.;

x_2 – объем произведенной продукции, млрд. руб.

| № п/п | x_1 | x_2 |
|-------|-------|-------|
| 1 | 12,0 | 16,0 |
| 2 | 8,5 | 9,2 |

| | | |
|---|------|------|
| 3 | 14,0 | 12,0 |
| 4 | 7,8 | 8,6 |
| 5 | 7,0 | 10,4 |

Определите матрицу расстояний используя метрику «Сити-блок» предварительно проведя нормировку данных методом стандартизации.

Комплект заданий для контрольной работы

Индикаторы достижения: УК-2.1, ПК-1.2

Тема: Многомерное признаковое пространство. Элементы математики, теории вероятностей и математической статистики в многомерных статистических методах

Вариант 1.

Задание 1. Раскройте понятия: линейная зависимость, базис, размерность в векторном пространстве.

Задание 2. Чем отличается квадратная матрица от прямоугольной, симметрическая от диагональной. Произвести операции сложения и умножения в матрицах $A \in R^{n,m}$, $B \in R^{n,m}$.

Задание 3. Как определить ранг матрицы, произвести транспонирование, обращение матрицы?

Задание 4. Запишите в матричном виде квадратичную форму:

$$q(x) = 2x_1^2 + 3x_2^2 - x_3^2 + 4x_1x_2 - 6x_1x_3 + 10x_2x_3$$

Задание 5. Пусть функция плотности $f(x)$ случайной величины равна

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2), & 0 \leq x \leq 1 \\ 0, & x > 1 \text{ и } x < 0 \end{cases}$$

Найдите функцию распределения.

Задание 6. Что понимают под статистической гипотезой и какие характерные признаки для нее существуют?

Задание 7. Будет ли статистической гипотеза:

а) о равенстве двух ковариационных матриц, представляющих связи признаков здоровья людей в двух группах: не имеющих хронической заболеваемости и с заболеваемостью?

б) о несущественности различий характеров двух человек? И если решать этот же вопрос относительно макроэкономического положения двух стран?

в) о том, что многомерная случайная величина $X^T = (x_1, x_2, x_3)$, представляющая совокупность предприятий, подчиняется нормальному закону распределения?

Приведите собственные примеры многомерных статистических гипотез.

Задание 8. Покажите графически доверительную область для трехмерной случайной величины

$X^T = (x_1, x_2, x_3)$, если x_1 – уровень оплатности продукции принимает параметрические значения $\bar{x}_1 = 0,20$ с предельно допустимой колеблемостью $(\bar{x}_1 - x_{i1}) \leq 0,5$, а x_2 – уровень энергоёмкости продукции при $\bar{x}_2 = 0,80$ и $(\bar{x}_2 - x_{i2}) \leq 0,1$, x_3 – удельный вес производственных налогов в стоимости продукции: $\bar{x}_3 = 0,45$ при $(\bar{x}_3 - x_{i3}) \leq 0,25$.

Задание 9. а). Определите функцию совместной плотности вероятностей $f(x_1, x_2, x_3)$, если X – трехмерная нормально распределенная случайная величины, $X \sim N(\mathbf{0}, \mathbf{I}_3)$, причем x_1, x_2, x_3 –

взаимно некоррелированы; б). Для данной случайной величины определите функции условных вероятностей.

Задание 10. В таблице приводятся данные о фонде заработной платы работников централизованной бухгалтерии x_1 и товарообороте обслуживаемых аптек x_2 :

| | | | | | | | | | |
|----------------------------------|-------|-----|-----|-----|-----|------|-----|-----|-----|
| Фонд заработной платы, тыс. руб. | x_1 | 20 | 36 | 28 | 51 | 70 | 45 | 30 | 56 |
| Товарооборот, млн. руб. | x_2 | 3,5 | 5,4 | 2,7 | 9,8 | 10,1 | 6,2 | 2,4 | 9,5 |

Найдите оценки математических ожиданий, дисперсии и коэффициента корреляции, доверительную область для вектора математических ожиданий с надежностью $\gamma = 0,95$.

Задание 11. По данным годовых отчетов $n=10$ промышленных предприятий с уровнем значимости $\alpha = 0,05$ проверить гипотезу о соответствии средних уровней экономических показателей работы группы предприятий (x_1 млн. руб. – объем валовой продукции, x_2 тыс.руб./чел. – производительность труда, x_3 млн. руб. – себестоимость товарной продукции) контрольным значениям при известных значениях указанных в таблице параметров генеральной совокупности, которая является нормально распределенной. Подберите наиболее мощный критерий для проверки гипотезы.

| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | σ_j^2 | Парная корреляция | Контрольные значения μ_{0j} |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------------|----------------------------|---------------------------------|
| x_1 | 3 | 4 | 5 | 5 | 5 | 5 | 6 | 7 | 10 | 10 | 4 | $\rho_{12} = 0,14\sqrt{5}$ | 5 |
| x_2 | 1,2 | 1,2 | 1,4 | 1,2 | 1,2 | 1,5 | 1,5 | 1,3 | 1,7 | 1,6 | 0,2 | $\rho_{23} = 0,12\sqrt{5}$ | 1,5 |
| x_3 | 2,1 | 2,8 | 3,2 | 4,5 | 4,8 | 4,9 | 5,5 | 6,5 | 8,5 | 8,2 | 4 | $\rho_{31} = 0,9$ | 5 |

Задание 12. Проверить влияние способа организации производства на экономические показатели группы предприятий на основе следующих выборочных данных.

| 1-й способ (X) | | 2-й способ (Y) | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| Фондовооруженность x_1 | Производительность x_2 | Фондовооруженность y_1 | Производительность y_2 |
| 56 | 52 | 58 | 74 |
| 51 | 64 | 56 | 75 |
| 52 | 63 | 62 | 71 |
| 72 | 68 | 69 | 83 |
| 67 | 59 | 62 | 78 |
| 51 | 62 | 65 | 72 |
| 52 | 64 | 59 | 80 |
| 50 | 68 | 71 | 70 |
| 59 | 70 | 38 | 60 |
| 60 | 70 | 47 | 65 |
| - | - | 74 | 60 |
| - | - | 59 | 76 |

Задание 13. Проверьте существенность различий уровня эффективности работы предприятий двух отраслей «А» и «Б» по следующим данным ($\alpha = 0,1$):

| Предприятия № | Рентабельность производства (%) | Среднегодовая выработка на одного работника, тыс. дол. США |
|---------------|---------------------------------|--|
| Отрасль «А» | | |
| 1 | 14 | 3,6 |
| 2 | 18 | 4,4 |
| 3 | 12 | 4,2 |
| 4 | 16 | 3,9 |
| 5 | 11 | 3,4 |
| 6 | 9 | 2,8 |
| Отрасль «Б» | | |
| 1 | 4 | 2,8 |
| 2 | 7 | 2,6 |
| 3 | 12 | 4,1 |
| 4 | 6 | 2,3 |
| 5 | 8 | 3,5 |
| 6 | 11 | 3,8 |
| 7 | 5 | 2,2 |
| 8 | 11 | 3,7 |

Задание 14. Проверьте предположение о равенстве двух ковариационных матриц, представляющих связи экономических показателей за два различных периода времени:

$$S_1 = \begin{pmatrix} 6050 & 1820 \\ 1820 & 3690 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 8400 & 2060 \\ 2060 & 4170 \end{pmatrix}.$$

Вариант 2.

Задание 1. Проверьте:

а) являются ли векторы a и b линейно независимыми: $a^T = (4, 12, 1, 7)$, $b^T = (4, -2, 1, 5)$;

б) являются ли векторы $a^T = (1, 3, 1, 3)$, $b^T = (2, 1, 1, 2)$, $c^T = (3, -1, 1, 1)$ линейно зависимыми.

Задание 2. Определите матрицу показателей стоимости продукции по ее видам и сортам, если имеется матрица A – объемов выпуска продукции по видам и сортам, и матрица B – уровней цен на каждый вид и сорт продукции:

$$A = \begin{pmatrix} 50 & 26 & 120 \\ 70 & 9 & 45 \\ 84 & 32 & 225 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 7 & 12 \\ 25 & 30 & 38 \\ 20 & 22 & 26 \end{pmatrix}.$$

Задание 3. Докажите, что приведенная система уравнений не имеет решения:

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ 4x_1 + 5x_2 + 6x_3 = 2 \\ 7x_1 + 8x_2 + 9x_3 = 3 \\ 5x_1 + 7x_2 + 9x_3 = 4 \end{cases}$$

Задание 4. Пусть совместное распределение двух случайных величин X и Y задано следующей таблицей:

| Случайные величины | | Y | | | |
|--------------------|---|-----|-----|-----|-----|
| | | -2 | -1 | 1 | 2 |
| X | 1 | 0 | 1/4 | 1/4 | 0 |
| | 4 | 1/4 | 0 | 0 | 1/4 |

Найдите математические ожидания, дисперсии и ковариацию величин. Проверьте, являются ли данные случайные величины независимыми?

Задание 5. Что означает: простая и сложная, одномерная и многомерная статистические гипотезы?

Задание 6. Какого рода ошибки могут допускаться и чем определяется достоверность выводов при проверке статистических гипотез?

Задание 7. Покажите графически доверительную область для двумерной случайной величины $X^T = (x_1, x_2)$, если x_1 – уровень олатоёмкости продукции принимает параметрические значения $\bar{x}_1 = 0,30$ с предельно допустимой колеблемостью $(\bar{x}_1 - x_{i1}) \leq 0,45$, а x_2 – уровень энергоёмкости продукции при $\bar{x}_2 = 0,70$ и $(\bar{x}_2 - x_{i2}) \leq 0,15$.

Задание 8. В связи с исследованием влияния прививок на заболеваемость гепатитом были получены следующие данные:

| | Не заболели | Заболели |
|------------|-------------|----------|
| Привитые | 276 | 3 |
| Непривитые | 473 | 66 |

Проанализируйте эффективность прививок.

Задание 9. Запишите плотность двумерной нормально распределенной случайной величины, в качестве параметра используя коэффициент корреляции.

Задание 10. По данным задачи 1 с помощью линейных комбинаций найти с надежностью $\gamma = 0,95$ интервальные оценки генеральных средних μ_1 и μ_2 .

Задание 11. Чтобы оценить производственную эффективность предложенной к внедрению технологии, проведена проверка качества продукции, выпущенной на старой и новой автоматических линиях, при этом получены следующие данные об удельном весе продукции высшего качества, %:

| Партия № | Старая линия | Партия № | Новая линия |
|----------|--------------|----------|-------------|
|----------|--------------|----------|-------------|

| | | | |
|---|----|---|----|
| 1 | 58 | 1 | 74 |
| 2 | 62 | 2 | 59 |
| 3 | 51 | 3 | 69 |
| 4 | 67 | 4 | 78 |
| 5 | 41 | 5 | 82 |
| 6 | 53 | 6 | 75 |
| | | 7 | 86 |
| | | 8 | 63 |

При $\alpha = 0,01$ следует установить, действительно ли новая линия, налаженная на передовую технологию, позволяет получить более высокий уровень качества продукции.

Задание 12. На частном предприятии с численностью работников 70 чел. средний уровень выработки одного рабочего составляет 500 деталей в день при среднеквадратическом отклонении 16 деталей. С целью корректировки среднего нормативного уровня выработки в сторону повышения администрацией была отобрана группа наиболее подготовленных рабочих и по ним установлен средний уровень выработки, равный 580 деталей. Можно ли вводить новый нормативный уровень выработки на предприятии, не опасаясь конфликтов с рабочими? Решите задачу при заданном уровне значимости $\alpha = 0,05$.

Задание 13. Оцените существенность различий двух рынков сбыта легковых автомобилей, если на первом рынке средний уровень реализационной цены автомобиля составляет 15 тыс. долларов, а экспертная оценка качества обслуживания (по 5-бальной шкале) – 3,4, на втором рынке соответственно: 18 тыс. долларов и 4,2 балла. Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид:

$$S_* = \begin{pmatrix} 9,4 & 0,28 \\ 0,28 & 2,0 \end{pmatrix}.$$

Задание 14. Проверьте предположение о равенстве двух ковариационных матриц, представляющих связи социально-экономических показателей за два различных периода времени:

$$S_1 = \begin{pmatrix} 7060 & 1940 \\ 1940 & 4690 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 9300 & 3050 \\ 3050 & 5270 \end{pmatrix}.$$

Тема: Корреляционно-дисперсионный анализ многомерных разнотипных данных

Задание 1. Пусть $\Sigma = (\sigma_{jk}); j, k = 1, \dots, p$ – ковариационная матрица случайного вектора $\xi = (\xi^1, \dots, \xi^p)$, т.е. $\sigma_{jk} = E[(\xi^j - E\xi^j)(\xi^k - E\xi^k)]$. Требуется доказать, что:

а) матрица Σ всегда симметрична и неотрицательно определена;

б) если компоненты ξ^1, \dots, ξ^p вектора ξ не связаны между собой линейным соотношением, то матрица Σ положительно определена.

Задание 2. При статистическом анализе двумерной случайной величины $\xi = (\xi^1, \xi^2)$ удается наблюдать лишь ее «зашумленные» значения, т.е. в действительности мы имеем дело со случайной величиной $\tilde{\xi} = (\tilde{\xi}^1, \tilde{\xi}^2)$, где $\tilde{\xi}^1 = \xi^1 + \delta^1$ и $\tilde{\xi}^2 = \xi^2 + \delta^2$, случайные ошибки δ^1 и δ^2 имеют нулевые средние, конечные дисперсии $D\delta^1 = \sigma_1^2$, $D\delta^2 = \sigma_2^2$, взаимно независимы и не зависят от ξ^1 и ξ^2 . Докажите, что $\left| r(\xi^1, \xi^2) \right| > \left| r(\tilde{\xi}^1, \tilde{\xi}^2) \right|$, т.е., что теснота парной корреляционной связи между случайными величинами ослабевает при аддитивном наложении на них случайных ошибок.

Задание 3. С целью исследования влияния погодных условий на урожайность кормовых трав рассмотрены данные Министерства земледелия Англии за 20 лет, характеризующие урожайность $x_i^{(0)}$ (в ц/акр), весеннее количество осадков $x_i^{(1)}$ (в дюймах) и накопленных за весну сумму «активных» (т.е. выше $+5^\circ\text{C}$) температур $x_i^{(2)}$ (в градусах по Фаренгейту) однородной в метеорологическом отношении области Англии, включающей в себя группу восточных графств. По выборке $(x_i^{(0)}, x_i^{(1)}, x_i^{(2)})$, $i = 1, \dots, 20$ были подсчитаны выборочные значения средних ($a^{(j)}$), дисперсий (s_j^2) и парных коэффициентов корреляции (r^{jk}) изучаемой трехмерной величины.

$$a^{(0)} = 28,02; \quad a^{(1)} = 4,91; \quad a^{(2)} = 594,0;$$

$$\sigma_0^2 = 19,54; \quad \sigma_1^2 = 1,21; \quad \sigma_2^2 = 7225;$$

$$r_{01} = 0,80; \quad r_{02} = -0,40; \quad r_{12} = -0,56.$$

Требуется:

- определить, действительно ли высокая температура в период созревания трав отрицательно влияет на их урожайность (ведь $r_{02} = -0,40$);
- построить интервальные оценки для $r_{01.2}$ и $r_{02.1}$ с уровнем доверия $P = 0,95$;
- вычислить значение выборочного множественного коэффициента корреляции $R_{0.12}$.

Задание 4. Пусть многомерная случайная величина $(\eta, \xi^1, \dots, \xi^p)$ подчиняется $(p+1)$ -мерному нормальному закону. Доказать, что $R^2_{\eta(\xi^1, \dots, \xi^p)}$ не может быть меньше парного или частного коэффициента корреляции между η и любой из объясняющих переменных ξ^j .

Задание 5. Группа из 8 предварительно отобранных абитуриентов проходила оценочные тесты. Первый тест на знание основ проведения экономико-плановых расчетов и вычислений, второй тест на умение логически излагать свои мысли. По результатам проведенного тестирования абитуриентам были присвоены следующие баллы:

| | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|
| 1й тест | 12 | 14 | 15 | 15 | 18 | 19 | 20 | 23 |
|---------|----|----|----|----|----|----|----|----|

| | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|
| 2й тест | 16 | 19 | 21 | 15 | 22 | 17 | 13 | 21 |
|---------|----|----|----|----|----|----|----|----|

Какое Ваше мнение о качестве тестирования.

Задание 6. Введем меру отклонения от статистической независимости категоризированных признаков $x^{(j)}$ и $x^{(k)}$ в виде:

$$D_{jk}^2 = n \sum_{i=1}^{m_j} \sum_{l=1}^{m_k} \frac{(\hat{p}_{il} - \hat{p}_{i\cdot} \hat{p}_{\cdot l})^2}{\hat{p}_{i\cdot} \hat{p}_{\cdot l}},$$

где $\hat{p}_{il} = \frac{n_{il}}{n}$, $\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}$, $\hat{p}_{\cdot l} = \frac{n_{\cdot l}}{n}$ (эмпирические частоты). Доказать, что $D_{jk}^2 = \hat{X}_{jk}^2$, где \hat{X}_{jk}^2 – коэффициент квадратической сопряженности признаков $x^{(j)}$ и $x^{(k)}$.

Задание 7. Докажите, что коэффициент ранговой корреляции Спирмена $-1 \leq \rho^S \leq 1$.

Задание 8. Предположим, что X и Y – независимые выборки. Определите математическое ожидание коэффициента корреляции Спирмена

$$\hat{\rho}^S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (1 - R_i)^2.$$

Задание 9. Покажите, что энтропийный коэффициент конкордации W удовлетворяет условию $0 \leq W \leq 1$.

ТЕСТЫ

по дисциплине «Многомерные статистические методы»

Индикаторы достижения: УК-2.1, ПК-1.2

Тестовый вопрос 1: Что Вы понимаете под многомерными статистическими методами?

- 1) методы классической статистики;
- 2) большое число вероятностно-статистических методов и приемов для обработки и анализа статистических многомерных данных;
- 3) методы визуализации данных;
- 4) множественный регрессионный анализ.

Тестовый вопрос 2: Объектом исследования в многомерных статистических методах является

- 1) многомерное признаковое пространство;
- 2) корреляционных анализ;
- 3) методы анализа данных;
- 4) всевозможные объекты.

Тестовый вопрос 3: Сформулируйте понятие многомерного признакового пространства.

- 5) пространство, не содержащее признаков или объектов;
- 6) пространство, размерность которого меньше двух;
- 7) пространство, объекты которого отражаются значениями одного какого-либо признака;
- 8) наблюдаемые объекты представлены значениями двух и более, до некоторого числа m признаков.

Тестовый вопрос 4: Какие данные можно использовать в многомерных статистических методах?

- 1) только количественные;
- 2) измеренные в интервальной шкале;
- 3) разнотипные данные;
- 4) как количественные, так и качественные.

Тестовый вопрос 5: Панельными данными являются:

- 1) данные, изменяющиеся в пространстве;
- 2) данные, изменяющиеся в пространстве и времени;
- 3) динамические данные;
- 4) временные ряды.

Тестовый вопрос 6: Укажите раздел мультивариативных методов, в котором проводится анализ выборочной совокупности на наличие грубых ошибок:

- 1) корреляционно-регрессионный анализ;
- 2) проектный анализ;
- 3) робастные статистические методы;
- 4) компонентный анализ;
- 5) дисперсионный анализ.

Тестовый вопрос 7: В чем принципиальное отличие параметрических и непараметрических многомерных методов? Выберите правильный ответ

- 1) параметрические методы основаны на оценках многомерных распределений и числовых характеристик многомерной статистической совокупности;

- 2) параметрические методы служат для анализа и обработки данных, несущественно отличающиеся от значений, которые наиболее часто встречаются в изучаемой совокупности;
- 3) параметрические методы служат для обработки количественных данных;
- 4) нет принципиальных отличий.

Тестовый вопрос 8: Методы факторного анализа предназначены для

- 1) оцифровки, представления, проверки статистических гипотез о характере распределения многомерных данных и т.д.;
- 2) оценивания статистических связей и закономерностей многомерных данных;
- 3) выявления и измерения скрытых причинно-следственных связей, сокращения размерности;
- 4) решения задач классификации многомерных данных.

Тестовый вопрос 9: Какие методы послужили первоосновой для развития многомерных статистических методов

- 1) корреляционного анализа;
- 2) регрессионного анализа;
- 3) факторного анализа;
- 4) дискриминантного анализа.

Тестовый вопрос 10: Методы корреляционно-дисперсионного и регрессионного анализа предназначены для

- 1) оцифровки, представления, проверки статистических гипотез о характере распределения многомерных данных и т.д.;
- 2) оценивания статистических связей и закономерностей многомерных данных;
- 3) выявления и измерения скрытых причинно-следственных связей, сокращения размерности;
- 4) решения задач классификации многомерных данных.

Тестовый вопрос 11: Ковариационно-дисперсионная матрица является

- 1) прямоугольной;
- 2) квадратной;
- 3) диагональной;
- 4) треугольной.

Тестовый вопрос 12: Определитель ковариационно-дисперсионной матрицы должен быть

- 1) положительным;
- 2) нулевым;
- 3) отрицательным;
- 4) любым.

Тестовый вопрос 13: Чему равен след ковариационно-дисперсионной матрицы?

- 1) числу признаков;
- 2) математическому ожиданию многомерной совокупности;
- 3) дисперсии признаков;
- 4) числу рассматриваемых объектов;

Тестовый вопрос 14: Какую числовую характеристику степени тесноты связи можно определить для таблицы сопряжённости? Выберите правильный ответ:

- 1) индекс корреляции;
- 2) парный коэффициент корреляции;
- 3) статистику Пирсона;
- 4) коэффициент детерминации.

Тестовый вопрос 15: Укажите статистику, которая используется при проверке гипотезы о равенстве ковариационных матриц различных выборочных совокупностей.

- 1) коэффициент множественной корреляции;
- 2) коэффициент детерминации;
- 3) χ^2 – статистика;
- 4) статистика Пирсона.

Тестовый вопрос 16: Какой показатель позволяет измерить степень согласованности экспертов?

- 1) парный коэффициент корреляции;
- 2) корреляционное отношение;
- 3) коэффициент ранговой корреляции Кендалла;
- 4) коэффициент конкордации.

Тестовый вопрос 17: Основным инструментом при измерении тесноты связи между двумя порядковыми переменными является.

- 1) парный коэффициент корреляции;
- 2) индекс корреляции;
- 3) коэффициент ранговой корреляции Кендалла;
- 4) коэффициент конкордации.

Тестовый вопрос 18: Основным инструментом при измерении нелинейной связи между количественными переменными является

- 1) парный коэффициент корреляции;
- 2) корреляционное отношение;
- 3) коэффициент ранговой корреляции Спирмена
- 4) коэффициент конкордации.

Тестовый вопрос 19: Какой показатель позволяет измерить степень взаимосвязи признаков, измеренных в номинальной шкале?

- 1) коэффициент конгруэнтности;
- 2) статистика квадратичной сопряженности;
- 3) коэффициент ранговой корреляции Кендалла;
- 4) энтропийный коэффициент согласия.

Тестовый вопрос 20: Укажите статистику, которая используется при проверке гипотезы о равенстве ковариационных матриц различных выборочных совокупностей.

- 5) коэффициент множественной корреляции;
- 6) коэффициент детерминации;
- 7) χ^2 – статистика;
- 8) статистика Пирсона.

Тестовый вопрос 21: В чем состоит различие понятий «общий фактор» и «элементарный признак»? Выберите правильный ответ:

- 1) эти понятия идентичны;
- 2) «общий фактор» есть независимая переменная, а «элементарный фактор» – зависимая переменная в уравнении регрессии;
- 3) «общий фактор» есть скрытый, латентный признак совокупности объектов, который устанавливается в результате обобщения элементарных, явных признаков;
- 4) «общий фактор» есть зависимая переменная, а «элементарный фактор» – независимая переменная в факторной модели.

Тестовый вопрос 22: На диагонали редуцированной корреляционной матрицы находится:

- 1) единичный вектор;

- 2) вектор средних значений элементарных признаков;
- 3) вектор характеристик;
- 4) вектор дисперсий общих факторов;
- 5) вектор специфичностей факторной модели.

Тестовый вопрос 23: Какая характеристика свидетельствует о достаточном числе факторов для изучаемого явления или процесса?

- 1) матрица факторного отображения;
- 2) матрица ковариаций;
- 3) коэффициент детерминации;
- 4) статистика Фишера.

Тестовый вопрос 24: С какой целью осуществляют вращение факторного пространства? Выберите правильный ответ:

- 1) для увеличения общих факторов;
- 2) для сокращения признакового пространства;
- 3) для лучшей интерпретации полученной факторной модели;
- 4) для получения главного фактора.

Тестовый вопрос 25: Что понимается под косоугольным вращением в факторном анализе?

- 1) поворот факторных осей на 90^0 ;
- 2) поворот координат исходных признаков на угол, меньше 90^0 ;
- 3) вращение пространства общих факторов, с углом между факторными осями, не равным 90^0 ;
- 4) поворот координат исходных признаков на угол, больше 90^0 .

Тестовый вопрос 26: Метод главных компонент лучше применять:

- 1) для слабокоррелированных признаков;
- 2) для коррелированных признаков;
- 3) если матрицы парных и частных корреляций статистически однородны;
- 4) когда определитель матрицы парных корреляций $\det R \rightarrow 0$.

Тестовый вопрос 27: Веса главных компонент показывают:

- 1) эластичность результативного признака;
- 2) тесноту линейно связи элементарных признаков;
- 3) долю дисперсии результативного признака, объясненную предикторным признаком;
- 4) тесноту линейной связи главной компоненты и элементарного признака.

Тестовый вопрос 28: Когда при построении регрессионных моделей целесообразно использовать метод главных компонент? Выберите правильный ответ:

- 1) при гетероскедастичности остатков модели;
- 2) при слабой зависимости регрессоров;
- 3) при наличии проблемы мультиколлинеарности;
- 4) если коэффициент детерминации меньше 0,8.

Тестовый вопрос 29: В каком виде может быть представлена модель в факторном анализе? Выберите правильный ответ:

- 1) $Y = AX + E$;
- 2) $F = AX$;
- 3) $X = AF + U$;
- 4) $F = AX + E$.

Тестовый вопрос 30: Что Вы понимаете под главной компонентой? Выберите правильный ответ:

- 1) главный фактор;
- 2) общий фактор;
- 3) линейную комбинацию первоначальных признаков;
- 4) взвешенную среднюю первоначальных признаков.

Критерии оценки (в баллах):

- 20 баллов выставляется обучающемуся, если по выполняемым заданиям в полном объеме достигнуты индикаторы УК-2.1, ПК-1.2;
- 10 баллов выставляется обучающемуся, если по выполняемым заданиям индикаторы УК-2.1, ПК-1.2 достигнуты с некоторыми замечаниями;
- 5 баллов выставляется обучающемуся, если по выполняемым заданиям индикаторы УК-2.1, ПК-1.2 достигнуты частично с ошибками и замечаниями;
- 0 баллов выставляется обучающемуся, если по выполняемым заданиям не достигнуты индикаторы УК-2.1, ПК-1.2.

Творческие задания по дисциплине «Многомерные статистические методы»

Индикаторы достижения: УК-2.1, ПК-1.2

Задания для написания аналитических записок:

Тема 1. Корреляционно-дисперсионный анализ (КДА) многомерных разнотипных данных

1. По данным индивидуального проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть I подготовьте аналитическую записку с подробным ответом на вопросы:
 - 1). Какие характеристики статистической связи могут быть использованы для анализа структуры полученных анкетных данных?
 - 2). Сравните оценки различных характеристик связи исследуемых многомерных данных.
2. По данным индивидуального проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть II подготовьте аналитическую записку с подробным ответом на вопросы:
 - 3). Какие характеристики статистической связи могут быть использованы для анализа структуры полученных порядковых данных?
 - 4). Сравните оценки различных характеристик связи исследуемых многомерных данных.

Критерии оценки (в баллах):

- 3 балла выставляется обучающемуся, если по выполняемым заданиям в полном объеме достигнуты индикаторы УК-2.1, ПК-1.2;
- 2 балла выставляется обучающемуся, если по выполняемым заданиям индикаторы УК-2.1, ПК-1.2 достигнуты с некоторыми замечаниями;
- 2 балла выставляется обучающемуся, если по выполняемым заданиям индикаторы УК-2.1, ПК-1.2 достигнуты частично с ошибками и замечаниями;
- 0 баллов выставляется обучающемуся, если по выполняемым заданиям не достигнуты индикаторы УК-2.1, ПК-1.2.

Темы индивидуальных и/или групповых проектов

Индикаторы достижения: УК-2.1, ПК-1.2

1. Многомерный статистический анализ кредитоспособности заемщика коммерческого банка на панельных данных.
2. Моделирование динамики ставки рефинансирования Банка России в зависимости от макроэкономических показателей методом главных компонент.
3. Многомерный статистический анализ зависимости удельных затрат на 1 кв. м. промышленного строительства от производственно-экономических характеристик.
4. Многомерный статистический анализ населения по структуре расходов на панельных данных.
5. Многомерный статистический анализ населения по структуре потребления продуктов питания на панельных данных.
6. Многомерный статистический анализ регионов РФ по макроэкономическим показателям.

7. Многомерный статистический анализ регионов РФ по социально-экономическим показателям.
8. Многомерный статистический анализ регионов РФ по демографическим показателям.
9. Многомерный статистический анализ регионов РФ по заболеваемости населения на панельных данных.
10. Многомерный статистический анализ регионов РФ по структуре регионального бюджета на панельных данных.
11. Многомерный статистический анализ стран мирового сообщества по макроэкономическим показателям на панельных данных.
12. Многомерный статистический анализ стран мирового сообщества по социально-экономическим показателям.
13. Многомерный статистический анализ стран мирового сообщества по демографическим показателям на панельных данных.
14. Многомерный статистический анализ государств по структуре бюджета на панельных данных.
15. Многомерный статистический анализ изменения ставки по федеральным фондам ФРС США от макроэкономических показателей.
16. Многомерное статистическое моделирование заболеваемости населения от социально-экономических факторов.
17. Многомерное статистическое моделирование рождаемости населения от социально-экономических факторов на панельных данных.
18. Многомерный статистический анализ зависимости среднедушевых расходов на питание от макроэкономических показателей.
19. Многомерный статистический анализ зависимости среднедушевых расходов на жилье от макроэкономических показателей.
20. Многомерный статистический анализ зависимости среднедушевых расходов на отдых от макроэкономических показателей.
21. Многомерный статистический анализ зависимости среднедушевых расходов на личный автотранспорт от макроэкономических показателей.
22. Многомерный статистический анализ зависимости среднедушевых расходов на питание от социально-экономических показателей.
23. Многомерный статистический анализ зависимости среднедушевых расходов на жилье от социально-экономических показателей.
24. Многомерный статистический анализ зависимости среднедушевых расходов на отдых от социально-экономических показателей.
25. Многомерный статистический анализ занятости населения в зависимости от макроэкономических показателей на панельных данных.
26. Моделирование цены на недвижимость в зависимости от макроэкономических показателей с учетом робастного оценивания.
27. Моделирование цены на загородную недвижимость в зависимости от макроэкономических показателей с учетом робастного оценивания.
28. Моделирование цены на недвижимость в зависимости от ее качественных характеристик с учетом робастного оценивания.
29. Моделирование средней цены на недвижимость в зависимости от социальной инфраструктуры с учетом робастного оценивания.
30. Моделирование кредитоспособности заемщика коммерческого банка с учетом робастного оценивания.

Задания для индивидуальных проектов:

Тема. Методы первичной записи и представления многомерных данных

Проект: «Оцифровка качественных признаков. Многомерное ранжирование»

Часть I. Статистическая обработка разнотипных данных

- 1) Используя произвольные анкетные данные сформулируйте цель и задачи эмпирического социологического исследования.
- 2) Проведите опрос по репрезентативной выборке.
- 3) Оцифруйте исходную информацию (преобразуйте качественные признаки в количественные данные).
- 4) Проведите статистическую обработку полученных данных.
- 5) Оформите полученные результаты в виде аналитического отчета.

Часть II. Многомерная группировка

- 1) Определите перечень характеристик многомерного статистического исследования в соответствии с выбранной темой индивидуального проекта.
- 2) Используя статистические ежегодники РФ (статистические ежегодники стран мирового сообщества, данные министерств и ведомств и т.п.) за период с 2000 по 2019 гг. соберите информацию по уровню социально-экономического развития (финансовой сферы, конъюнктуры рынка и т.п.) регионов РФ (стран мирового сообщества и других объектов) в динамике в соответствии с поставленной целью исследования.
- 3) Предложите обоснованный метод свертки многомерных данных по времени.
- 4) Проведите группировку регионов (стран мирового сообщества и других объектов), используя методы многомерного ранжирования.
- 5) Проведите группировку регионов (стран мирового сообщества и других объектов), используя альтернативные статистические методы.
- 6) Результаты оформите в виде аналитического отчета.

Тема. Робастное статистическое оценивание

Проект: «Факторный анализ многомерных данных»

Используя количественные данные проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть II по регионам РФ (странам мирового сообщества и т.п.) проведите робастное статистическое оценивание многомерной совокупности.

- 1) Для всех признаков найдите оценку матрицы R парных коэффициентов корреляции.
- 2) На основе матрицы парных корреляций определите наиболее информативные признаки.
- 3) Используя результаты многомерного ранжирования проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть II визуальным образом выявить нетипичные объекты выборки и проверьте статистически свои выводы.
- 4) Проведите статистический анализ нетипичных объектов.
- 5) Результаты оформите в виде аналитического отчета.

Тема. Модели и методы факторного анализа

Часть I. Используя количественные данные проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть II по регионам РФ (странам мирового сообщества и т.п.) проведите факторный анализ.

- 1) Для всех признаков оцените матрицу R парных коэффициентов корреляции.
- 2) На основании матрицы собственных значений матрицы R определите вклад компонент в суммарную дисперсию. Обоснуйте вклад значимых главных компонент ($k < m$).
- 3) Используя матрицу весов главных компонент, дайте экономическую интерпретацию полученным главным компонентам.
- 4) Проведите визуализацию и классификацию объектов по первым двум-трем (в зависимости от анализа п.3) главным компонентам. Дайте интерпретацию полученным результатам.

5) Визуально выявите нетипичные объекты выборки и проверьте статистически свои выводы.

Часть II.

- 1) Для факторных признаков x оцените матрицу R парных коэффициентов корреляции.
- 2) На основании матрицы собственных значений матрицы R определите вклад компонент в суммарную дисперсию. Обоснуйте вклад существенных главных компонент ($k < m$).
- 3) Используя матрицу весов главных компонент, дайте экономическую интерпретацию полученным главным компонентам.
- 4) Используя вектор значений результативного признака y и значения главных компонент, постройте уравнение регрессии на главных компонентах с максимальным числом значимых коэффициентов регрессии.
- 5) Прокомментируйте результаты.

Часть III.

- 1) Повторите расчеты части II для типичных (не содержащих выбросов) данных.
- 2) С учетом результатов выполнения части I и части II постройте регрессию на главных компонентах для типичных данных.
- 3) Проведите сопоставительный анализ полученных регрессионных моделей.
- 4) Результаты оформите в виде аналитического отчета.

Тема. Кластерный анализ (КА)

Проект: «Построение устойчивой многомерной классификации без обучающей выборки»

Часть I. Используя количественные данные проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть II по регионам РФ (странам мирового сообщества и т.п.), проведите структурный анализ связей в признаковом пространстве (для признаков).

- 1) Отберите информативные признаки для описания ваших объектов.
- 2) Проведите классификацию объектов в полном и сокращенном признаковом пространстве с учетом отобранных информативных признаков. Проанализируйте результат. Выберите экономически обоснованную классификацию регионов (стран, предприятий и т.д.).
- 3) Проверьте качество полученной классификации.
- 4) Сравните полученную группировку с результатами индивидуального проекта №1.
- 5) Дайте описание полученных результатов.

Часть II. Используя анкетные данные проекта «Оцифровка качественных признаков. Многомерное ранжирование» часть I по респондентам, проведите структурный анализ связей в признаковом пространстве (для признаков).

- 1) Отберите информативные признаки для описания ваших респондентов.
- 2) Проведите типологическую группировку респондентов в полном и сокращенном признаковом пространстве с учетом отобранных информативных признаков. Проанализируйте результат. Выберите обоснованную группировку респондентов.
- 3) Проверьте качество полученной группировки.
- 4) Дайте описание полученных результатов.
- 5) Результаты оформите в виде аналитического отчета.

Критерии оценки (в баллах):

- 17 баллов выставляется обучающемуся, если по выполняемым индивидуальным проектам в полном объеме достигнуты индикаторы УК-2.1, ПК-1.2;
- 12 баллов выставляется обучающемуся, если по выполняемым индивидуальным проектам индикаторы УК-2.1, ПК-1.2 достигнуты с некоторыми замечаниями;

- 5 баллов выставляется обучающемуся, если по выполняемым индивидуальным проектам индикаторы УК-2.1, ПК-1.2 достигнуты частично с ошибками и замечаниями;
- 0 баллов выставляется обучающемуся, если по выполняемым индивидуальным проектам не достигнуты индикаторы УК-2.1, ПК-1.2.

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ХАРАКТЕРИЗУЮЩИЕ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ ВО ВРЕМЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Структура экзаменационного билета/зачетного задания

| <i>Наименование оценочного средства</i> | <i>Максимальное количество баллов</i> |
|---|---------------------------------------|
| Вопрос 1. Особенности социально-экономической информации. Измерение различных данных. | 8 |
| Вопрос 2. Выборочный коэффициент ранговой корреляции Спирмена для многомерных данных и проверка гипотезы о его значимости. | 12 |
| Практическое задание 1. Покажите графически доверительную область для трехмерной случайной величины $x' = (x_1, x_2, x_3)$, если x_1 – уровень платоемкости продукции принимает параметрические значения $\bar{x}_1 = 0,30$ с предельно допустимой колеблемостью $(\bar{x}_1 - x_{i1}) \leq 0,4$, а x_2 – уровень энергоемкости продукции при $\bar{x}_2 = 0,70$ и $(\bar{x}_2 - x_{i2}) \leq 0,1$, x_3 – удельный вес производственных налогов в стоимости продукции: $\bar{x}_3 = 0,45$ при $(\bar{x}_3 - x_{i3}) \leq 0,25$. | 10 |
| Практическое задание 2. Оцените существенность различий двух сетевых магазинов, если у первого магазина средний чек составляет 500 рублей, а экспертная оценка качества обслуживания (по 5-бальной шкале) – 3, у второго – цена чека в среднем 380 рублей и 4,2 балла. Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид: | 10 |

Задания, включаемые в экзаменационный билет/зачетное задание

Типовой перечень вопросов к экзамену:

1. МСМ как самостоятельная область научных знаний. Их особенности и отличия от методов классической статистики.
2. Основные исторические этапы развития МСМ. Классификация МСМ. Параметрические и непараметрические методы, их различия.
3. Особенности социально-экономической информации. Измерение различных данных.
4. Оцифровка неколичественной информации. Основные способы оцифровки.
5. Понятие признакового пространства. Примеры одномерного, двумерного и многомерного признакового пространства. Геометрическая интерпретация.
6. Многомерные случайные величины и их распределения. Многомерные статистические гипотезы. Примеры.

7. Проверка гипотез и доверительные области для вектора математических ожиданий.
8. Критерий Хотеллинга для двух многомерных выборок.
9. Критерий Бартлетта и проверка гипотезы об однородности дисперсии.
10. Цели и задачи робастного статистического оценивания. Многомерные методы выявления ошибок в данных. Примеры.
11. Грубые ошибки. Причины их появления в статистической совокупности. Статистические методы и критерии их выявления.
12. Основные методы устойчивого оценивания параметров выборочной совокупности.
13. Цели и задачи многомерного корреляционно-дисперсионного анализа. Примеры использования для количественных и качественных данных.
14. Меры связи количественных многомерных данных и критериальная проверка их значимости.
15. Меры связи многомерных данных, измеренных в номинальной шкале.
16. Вероятностные коэффициенты связи.
17. Критерий независимости для таблиц сопряженности.
18. Многомерные методы анализа экспертных оценок.
19. Особенности дисперсии данных, измеренных в порядковой шкале.
20. Дисперсионный и энтропийный коэффициенты конкордации. В каких границах они изменяются, и что это означает?
21. Выборочный коэффициент ранговой корреляции Спирмена для многомерных данных и проверка гипотезы о его значимости.
22. Выборочный коэффициент ранговой корреляции Кендалла для многомерных данных и проверка гипотезы о его значимости.
23. Энтропийный метод получения интегрального решения в экспертном оценивании.
24. Цели и задачи метода многомерного ранжирования в социально-экономических исследованиях. Примера использования.
25. Многомерные статистические методы снижения размерности признакового пространства, измеренного в различных шкалах. Примеры использования.
26. Цели и задачи метода главных компонент. Вычислительная процедура метода главных компонент. Геометрическая интерпретация.
27. Информативность и весовые коэффициенты главных компонент. Примеры.
28. Взаимосвязи весовых коэффициентов главных компонент с дисперсией признаков и с корреляцией признаков.
29. Алгоритмическая схема использования главных компонент в многомерном регрессионном анализе.
30. Основная концепция факторного анализа (ФА) и существование модели ФА. Состав дисперсии элементарного признака в ФА.
31. Алгоритмическая схема реализации методов ФА.
32. Структурные параметры модели ФА.
33. Вычислительная процедура метода главных компонент.
34. Информативность и весовые коэффициенты главных компонент. Примеры
35. Метод главных факторов. Отличия от метода главных компонент.
36. Методы вычисления редуцированной корреляционной матрицы в ФА.
37. Итеративный метод Хотеллинга получения главных факторов. Его преимущества по сравнению с другими методами ФА.
38. Итеративный метод максимального правдоподобия улучшения факторных нагрузок и характеристик.
39. Применение методов ФА в корреляционно-регрессионном анализе.
40. Методы многомерной группировки объектов в экономике.
41. Понятие однородности объектов в кластерном анализе. Способы измерения однородности объектов, измеренных в разнотипных шкалах.
42. Понятие стратегии классификации объектов в кластерном анализе. Алгоритмы стратегий.

43. Оценка информативности признаков в кластерном анализе.
 44. Методы иерархического кластерного анализа. Алгоритмическая схема. Геометрическая интерпретация результатов.
 45. Итеративные методы кластерного анализа. Их отличия от иерархических методов.
 46. Оценка качества классификации в кластерном анализе.

Типовые расчетно-аналитические задания/задачи:

1. В связи с исследованием влияния занятий фитнесом на уровень здоровья сотрудников компании, наблюдалась заболеваемость в течение года. Были получены следующие данные:

| Событие | Занятие фитнесом | |
|-----------------------|------------------|-----|
| | Да | Нет |
| Не болели | 47 | 102 |
| Болели 1 раз | 9 | 54 |
| Болели 2 и более раза | 1 | 6 |

Проанализируйте эффективность вложений в оплату фитнеса с точки зрения работодателя.

2. За два различных периода времени на одном предприятии осуществлялся сбор данных относительно объема выпуска продукции и затрат. Располагая двумя ковариационными матрицами, проверьте изменились ли экономические показатели работы предприятия:

$$S_1 = \begin{pmatrix} 405 & 182 \\ 182 & 169 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 640 & 206 \\ 206 & 217 \end{pmatrix}.$$

3. Оцените существенность различий эффективности двух предприятий, если на первом предприятии производительность труда составляет 15 ед. на чел., а экспертная оценка качества продукции (по 5-бальной шкале) – 3,4, на втором – соответственно: 18 ед. на чел. и 4,2 балла.

Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид:

$$S_* = \begin{pmatrix} 10,4 & 0,4 \\ 0,4 & 3,0 \end{pmatrix}.$$

4. Пусть $\Sigma = (\sigma_{jk}); j, k = 1, \dots, p$ – ковариационная матрица случайного вектора $\xi = (\xi^1, \dots, \xi^p)$,

т.е. $\sigma_{jk} = E \left[\left(\xi^j - E \xi^j \right) \left(\xi^k - E \xi^k \right) \right]$. Требуется доказать, что матрица Σ всегда симметрична и неотрицательно определена.

5. По 10 исследуемым объектам недвижимости имеются данные о средней стоимости аренды квадратного метра площади и минимальном сроке аренды.

| Объект | Стоимость аренды | Минимальный срок аренды | Объект | Стоимость аренды | Минимальный срок аренды |
|--------|------------------|-------------------------|--------|------------------|-------------------------|
| 1 | 600 | 6 | 6 | 1350 | 14 |
| 2 | 500 | 8 | 7 | 900 | 5 |
| 3 | 800 | 9 | 8 | 500 | 7 |
| 4 | 700 | 8 | 9 | 800 | 7 |
| 5 | 1000 | 15 | 10 | 1050 | 14 |

Определите наличие грубых и рассчитайте устойчивые средние по методам Пуанкаре, Винзора и сравните их. Получите винзорированные данные. Прокомментируйте результаты.

6. По матрице факторных нагрузок A воспроизведите матрицу парных корреляций R :

$$A = \begin{pmatrix} 0,78 & 0,62 \\ 0,78 & 0,62 \end{pmatrix}.$$

7. По 10 наблюдаемым объектам имеются данные о средней концентрации загрязняющих веществ в воздухе:

| Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ | Объект | Пыль, мг/м ³ | Окись углерода, мг/м ³ |
|--------|-------------------------|-----------------------------------|--------|-------------------------|-----------------------------------|
| 1 | 0,12 | 1,2 | 6 | 0,27 | 2,8 |
| 2 | 0,10 | 1,6 | 7 | 0,18 | 1,1 |
| 3 | 0,16 | 1,8 | 8 | 0,10 | 1,3 |
| 4 | 0,14 | 1,7 | 9 | 0,16 | 1,4 |
| 5 | 0,20 | 3,0 | 10 | 0,21 | 2,9 |

Определите наличие грубых ошибок, рассчитайте устойчивые средние по методам Пуанкаре и Винзора.

8. В таблице приводятся данные о выручке предприятия x_1 и затратах x_2 :

| | | | | | | | | | |
|--------------------|-------|-----|-----|-----|-----|------|-----|-----|-----|
| Выручка, тыс. руб. | x_1 | 20 | 36 | 28 | 51 | 70 | 45 | 30 | 56 |
| Затраты, тыс. руб. | x_2 | 3,5 | 5,4 | 2,7 | 9,8 | 10,1 | 6,2 | 2,4 | 9,5 |

Найдите оценки векторного математического ожидания, дисперсии и доверительную область для вектора математических ожиданий с надежностью $\gamma = 0,95$.

9. В таблице приводятся данные предприятия по объему продаж x_1 и расходах на рекламу x_2 :

| | | | | | | | | | |
|-------------------------------|-------|-----|-----|-----|-----|------|-----|-----|-----|
| Объем продаж, млн. руб. | x_1 | 25 | 41 | 33 | 56 | 75 | 50 | 35 | 61 |
| Расходы на рекламу, тыс. руб. | x_2 | 3,5 | 5,4 | 2,7 | 9,8 | 10,1 | 6,2 | 2,4 | 9,5 |

Найдите оценки математических ожиданий, дисперсии доверительную область для вектора математических ожиданий с надежностью $\gamma = 0,95$.

10. При статистическом анализе двумерной случайной величины $\xi = (\xi^1, \xi^2)$ удается наблюдать лишь ее «зашумленные» значения, т.е. в действительности мы имеем дело со

случайной величиной $\tilde{\xi} = (\tilde{\xi}^1, \tilde{\xi}^2)$, где $\tilde{\xi}^1 = \xi^1 + \delta^1$ и $\tilde{\xi}^2 = \xi^2 + \delta^2$, случайные ошибки δ^1 и δ^2

имеют нулевые средние, конечные дисперсии $D\delta^1 = \sigma_1^2$, $D\delta^2 = \sigma_2^2$, взаимно независимы и не

зависят от ξ^1 и ξ^2 . Докажите, что $\left| r(\xi^1, \xi^2) \right| > \left| r(\tilde{\xi}^1, \tilde{\xi}^2) \right|$, т.е., что теснота парной корреляционной связи между случайными величинами ослабевает при аддитивном наложении на них случайных ошибок.

11. Пусть многомерная случайная величина $(\eta, \xi^1, \dots, \xi^p)$ подчиняется $(p+1)$ – мерному

нормальному закону. Доказать, что $R^2_{\eta(\xi^1, \dots, \xi^p)}$ не может быть меньше парного или частного коэффициента корреляции между η и любой из объясняющих переменных ξ^j .

12. В связи с исследованием влияния занятий фитнесом на уровень здоровья сотрудников компании, наблюдалась заболеваемость в течение года. Были получены следующие данные:

| Событие | Занятие фитнесом | |
|-----------------------|------------------|-----|
| | Да | Нет |
| Не болели | 47 | 102 |
| Болели 1 раз | 9 | 54 |
| Болели 2 и более раза | 1 | 6 |

Критериально подтвердите эффективность вложений в оплату фитнеса.

13. Введем меру отклонения от статистической независимости категоризированных признаков $x^{(j)}$ и $x^{(k)}$ в виде:

$$D_{jk}^2 = n \sum_{i=1}^{m_j} \sum_{l=1}^{m_k} \frac{(\hat{p}_{il} - \hat{p}_{i\cdot} \hat{p}_{\cdot l})^2}{\hat{p}_{i\cdot} \hat{p}_{\cdot l}},$$

где $\hat{p}_{il} = \frac{n_{il}}{n}$, $\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}$, $\hat{p}_{\cdot l} = \frac{n_{\cdot l}}{n}$ (эмпирические частоты). Доказать, что $D_{jk}^2 = \hat{X}_{jk}^2$, где \hat{X}_{jk}^2 – коэффициент квадратической сопряженности признаков $x^{(j)}$ и $x^{(k)}$.

14. Оцените существенность различий тарифов ЖКХ в двух регионах, если в первом регионе тариф на отопление составляет 80 руб. на кв. м. площади помещения, а тариф на электричество – 4 руб. за кВт, во втором – соответственно: 65 и 5 руб. Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид:

$$S_* = \begin{pmatrix} 105 & 40 \\ 40 & 30 \end{pmatrix}.$$

15. Докажите, что коэффициент ранговой корреляции Спирмена $-1 \leq \rho^S \leq 1$.

16. В страховой компании исследовалось влияние пола застрахованного клиента на частоту выплат по КАСКО за календарный год. Были получены следующие данные:

| Событие | Пол клиента | |
|------------------------|-------------|---------|
| | Женский | Мужской |
| Нет выплат | 47 | 102 |
| Выплаты 1 раз | 44 | 54 |
| Выплаты 2 и более раза | 1 | 6 |

Проанализируйте наличие зависимости.

17. Предположим, что X и Y – независимые выборки. Определите математическое ожидание коэффициента корреляции Спирмена

$$\hat{\rho}^S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (i - R_i)^2.$$

18. Покажите, что энтропийный коэффициент конкордации W удовлетворяет условию $0 \leq W \leq 1$.

19. Используя матрицу парных корреляций, получите главные компоненты.

$$R = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}.$$

20. По матрице факторных нагрузок A определите дисперсию главных компонент:

$$A = \begin{pmatrix} 0,78 & 0,62 \\ 0,78 & 0,62 \end{pmatrix}.$$

21. За два различных периода времени на одном предприятии осуществлялся сбор данных относительно объема выпуска продукции и затрат. Располагая двумя ковариационными матрицами, проверьте изменились ли экономические показатели работы предприятия.

$$S_1 = \begin{pmatrix} 305 & 82 \\ 82 & 69 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 540 & 106 \\ 106 & 117 \end{pmatrix}.$$

22. Запишите плотность двумерной нормально распределенной случайной величины, в качестве параметра используя коэффициент корреляции.

23. По матрице факторных нагрузок A определите дисперсию первоначальных признаков:

$$A = \begin{pmatrix} 10,09 & 1,98 \\ 2,98 & 6,70 \end{pmatrix}.$$

24. Определите функцию совместной плотности вероятностей $f(x_1, x_2, x_3)$ для трехмерной нормально распределенной случайной величины $X \sim N(0, I_3)$, причем x_1, x_2, x_3 – взаимно некоррелированные величины.

25. Используя матрицу факторных нагрузок A , получите уравнения для главных компонент:

$$A = \begin{pmatrix} 0,78 & 0,62 \\ 0,78 & 0,62 \end{pmatrix}.$$

26. Дана X – трехмерная нормально распределенная случайная величины, $X \sim N(0, I_3)$, причем x_1, x_2, x_3 – взаимно некоррелированы. Для данной случайной величины определите функции условных вероятностей.

27. По матрице факторных нагрузок A определите вклад главных компонент в общую дисперсию:

$$A = \begin{pmatrix} 10,09 & 1,98 \\ 2,98 & 6,70 \end{pmatrix}.$$

28. Используя матрицу парных корреляций, получите главные компоненты.

$$R = \begin{pmatrix} 1 & 0,6 \\ 0,6 & 1 \end{pmatrix}.$$

29. Покажите графически доверительную область для двумерной случайной величины $x'=(x_1, x_2)$, если x_1 – уровень олатоемкости продукции принимает параметрические значения $\bar{x}_1 = 0,10$ с предельно допустимой колеблемостью $(\bar{x}_1 - x_{i1}) \leq 0,5$, а x_2 – уровень энергоемкости продукции при $\bar{x}_2 = 0,90$ и $(\bar{x}_2 - x_{i2}) \leq 0,1$.

30. Оцените существенность различий эффективности двух предприятий, если на первом предприятии производительность труда составляет 20 ед. на чел., а экспертная оценка качества продукции (по 5-бальной шкале) – 4,2, на втором – соответственно: 18 ед. на чел. и 3,2 балла.

Пусть $\alpha = 0,05$, объединенная ковариационная матрица имеет вид:

$$S_* = \begin{pmatrix} 12 & 0,5 \\ 0,5 & 3,0 \end{pmatrix}.$$

Показатели и критерии оценивания планируемых результатов освоения компетенций и результатов обучения, шкала оценивания

| Шкала оценивания | | Формируемые компетенции | Индикатор достижения компетенции | Критерии оценивания | Уровень освоения компетенций |
|------------------|-------------------------|-------------------------|----------------------------------|---|------------------------------|
| 85 – 100 баллов | «отлично»/ «зачтено» | УК-2, | УК-2.1, | <p>Знает верно и в полном объеме: основные принципы и концепции в области целеполагания и принятия решений, методы генерирования альтернатив решений и приведения их к сопоставимому виду для выбора оптимального решения, природу данных, необходимых для решения поставленных задач</p> <p>Умеет верно и в полном объеме: системно анализировать поставленные цели, формулировать задачи и предлагать обоснованные решения, критически оценивать информацию о предметной области принятия решений, использовать инструментальные средства для разработки и принятия решений</p> | Продвинутый |
| | | ПК-1 | ПК-1.2 | <p>Знает верно и в полном объеме: языки визуального моделирования, теорию систем, предметную область и специфику деятельности организации в объеме, достаточном для решения задач бизнес-анализа</p> <p>Умеет верно и в полном объеме: использовать техники эффективных коммуникаций, определять связи и зависимости между элементами информации бизнес-анализа, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, анализировать внутренние (внешние) факторы и</p> | |

| | | | | | |
|-----------------------|----------------------------|-------|---------|---|-------------------|
| | | | | <i>условия, влияющие на деятельность организации</i> | |
| 70 – 84 баллов | «хорошо»/ «зачтено» | УК-2, | УК-2.1, | <p>Знает с незначительными замечаниями: <i>основные принципы и концепции в области целеполагания и принятия решений, методы генерирования альтернатив решений и приведения их к сопоставимому виду для выбора оптимального решения, природу данных, необходимых для решения поставленных задач</i></p> <p>Умеет с незначительными замечаниями: <i>системно анализировать поставленные цели, формулировать задачи и предлагать обоснованные решения, критически оценивать информацию о предметной области принятия решений, использовать инструментальные средства для разработки и принятия решений</i></p> | Повышенный |
| | | ПК-1 | ПК-1.2 | Знает с незначительными замечаниями: <i>языки визуального моделирования, теорию систем, предметную область и специфику деятельности организации в</i> | |

| | | | | | |
|----------------|-----------------------------------|-------|---------|---|---------|
| | | | | <p><i>объеме, достаточном для решения задач бизнес-анализа</i></p> <p>Умеет с незначительными замечаниями: использовать техники эффективных коммуникаций, определять связи и зависимости между элементами информации бизнес-анализа, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, анализировать внутренние (внешние) факторы и условия, влияющие на деятельность организации</p> | |
| 50 – 69 баллов | «удовлетворительно»/ «зачтено» | УК-2, | УК-2.1, | <p>Знает на базовом уровне, с ошибками: основные принципы и концепции в области целеполагания и принятия решений, методы генерирования альтернатив решений и приведения их к сопоставимому виду для выбора оптимального решения, природу данных, необходимых для решения поставленных задач</p> <p>Умеет на базовом уровне, с ошибками: системно анализировать поставленные цели, формулировать задачи и предлагать обоснованные решения, критически оценивать информацию о предметной области принятия решений, использовать инструментальные средства для разработки и принятия решений</p> | Базовый |
| | | ПК-1 | ПК-1.2 | <p>Знает на базовом уровне, с ошибками: языки визуального моделирования, теорию систем, предметную область и специфику деятельности организации в объеме,</p> | |

| | | | | | |
|-----------------|--|-------|---------|---|-----------------------------|
| | | | | <p><i>достаточном для решения задач бизнес-анализа</i></p> <p>Умеет на базовом уровне, с ошибками: <i>использовать техники эффективных коммуникаций, определять связи и зависимости между элементами информации бизнес-анализа, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, анализировать внутренние (внешние) факторы и условия, влияющие на деятельность организации</i></p> | |
| менее 50 баллов | «неудовлетворительно»/ «не зачтено» | УК-2, | УК-2.1, | <p>Не знает на базовом уровне: <i>основные принципы и концепции в области целеполагания и принятия решений, методы генерирования альтернатив решений и приведения их к сопоставимому виду для выбора оптимального решения, природу данных, необходимых для решения поставленных задач</i></p> <p>Не умеет на базовом уровне: <i>системно анализировать поставленные цели, формулировать задачи и предлагать обоснованные решения, критически оценивать информацию о предметной области принятия решений, использовать инструментальные средства для разработки и принятия решений</i></p> | Компетенции не сформированы |
| | | ПК-1 | ПК-1.2 | <p>Не знает на базовом уровне: <i>языки визуального моделирования, теорию систем, предметную область и специфику деятельности организации в объеме,</i></p> | |

| | | | | | |
|--|--|--|--|--|--|
| | | | | <p><i>достаточном для решения задач бизнес-анализа</i></p> <p>Не умеет на базовом уровне: <i>использовать техники эффективных коммуникаций, определять связи и зависимости между элементами информации бизнес-анализа, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, анализировать внутренние (внешние) факторы и условия, влияющие на деятельность организации</i></p> | |
|--|--|--|--|--|--|