

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Антипова Наталья Викторовна
Должность: и.о. директора филиала
Дата подписания: 01.09.2026 12:26:12
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

Кафедра статистики

Одобрено
на заседании Совета Улан-Баторского
филиала РЭУ им. Г.В. Плеханова
протокол №11 от 11 июня 2026 г.
Председатель совета
Н.В. Антипова



ОЦЕНОЧНЫЕ СРЕДСТВА

Б1.В.09

(шифр по УП)

Интеллектуальный анализ данных в экономике
и социальной сфере
(наименование дисциплины)

Направление подготовки

38.03.01

(код)

Экономика

(наименование направления)

Направленность (профиль) программы

"Бизнес-статистика и аналитика"

Уровень высшего образования Бакалавриат

Год начала подготовки 2026

Улан-Батор – 2026 г.

Составитель(и):

К.э.н., доц., доцент

А.В. Безруков

ОЦЕНОЧНЫЕ СРЕДСТВА

по учебной дисциплине «Интеллектуальный анализ данных в экономике и социальной сфере»

(наименование дисциплины в соответствии с учебным планом)

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 ¹
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2. Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации	УК-1.2. З-1. Знает критерии сопоставления различных вариантов решения поставленной задачи	Тема 1, 2
		УК-1.2. У-1. Умеет осуществлять критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи	
		УК-1.2. У-2. Умеет отличать факты от мнений, интерпретаций и оценок при анализе собранной информации	
		УК-1.2. У-3. Умеет сопоставлять и оценивать различные варианты решения поставленной задачи, определяя их достоинства и недостатки	Тема 3, 4
ПК-1. Способен анализировать, обосновывать и выбирать решения	ПК-1.3. Рассматривает эффективность вариантов решений как соотношения между ожидаемыми уровнями использования ресурсов и их ценностью	ПК-1.3. З-1. Знает языки визуального моделирования	Тема 5, 6, 7
		ПК-1.3. З-2. Знает методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации бизнес-анализа	
		ПК-1.3. З-3. Знает информационные технологии (программное обеспечение), применяемые в организации, в объеме, необходимом для целей	

¹ Наименования разделов и тем соответствует рабочей программе дисциплины.

		бизнес-анализа	
		ПК-1.3. У-1. Умеет оформлять результаты бизнес-анализа в соответствии с выбранными подходами	
		ПК-1.3. У-2. Умеет применять информационные технологии в объеме, необходимом для целей бизнес-анализа	
		ПК-1.3. У-3. Умеет проводить оценку эффективности решения с точки зрения выбранных критериев	
		ПК-1.3. У-4. Умеет оценивать бизнес-возможность реализации решения с точки зрения выбранных целевых показателей	

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ХАРАКТЕРИЗУЮЩИЕ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ

**Полный перечень вопросов для проведения устного опроса студентов
по дисциплине «Интеллектуальный анализ данных в экономике и социальной
сфере»**

Индикаторы достижения: УК-1.2, ПК-1.3

Тема 1: Основы интеллектуального анализа данных

1. Современные программные пакеты средств статистического анализа.
2. Средства описательной статистики в R, Python: возможности и ограничения.
3. Анализ вариационного ряда в R, Python.
4. Инструменты описательной статистики в R, Python: дисперсия, мода, медиана, стандартная ошибка.
5. Анализ формы распределения в R, Python.
6. Уровень надежности в R, Python. Интерпретация полученных результатов.

Тема 2: Первичная обработка данных для интеллектуального анализа

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 3: Методы классификации в интеллектуальном анализе

1. Сформулируйте основные направления и методы кластерного анализа.
2. Что представляет собой формальная постановка задачи и какие данные требуются для иерархической кластеризации?
3. Назовите необходимые исходные данные и постановку задачи для метода К средних.

Тема 4: Ассоциативное моделирование

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 5: Компонентное моделирование

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.

Тема 6: Кластерный анализ

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?

Тема 7: Нейросетевое моделирование

1. Приведите известные вам меры связи количественных и не количественных многомерных данных.
2. В каких случаях целесообразно использовать вероятностные коэффициенты связи?

Критерии оценки (в баллах):

Максимальный балл за правильный ответ на устный вопрос преподавателя 0,4 балла.

Неправильный или неполный ответ на вопрос преподавателя 0 баллов.

ТЕСТЫ

по дисциплине «Интеллектуальный анализ данных в экономике и социальной сфере»

Индикаторы достижения: УК-1.2, ПК-1.3

- 1) Какой метод используется для оценки параметров уравнения используется, в соответствии с которым $\hat{\beta} = (X'X)^{-1}X'Y$, где $\hat{\beta}$ – вектор оценок параметров уравнения, X – матрица независимых переменных, Y – вектор независимых переменных.
- а) Метод к-средних;
 б) Метод наименьших квадратов;
 в) Индексный метод;
 г) Иерархические кластер-процедуры.
- 2) По какой формуле определяется наблюдаемое значение F -критерия Фишера (R^2 – множественный коэффициент детерминации, n – число наблюдений, m – число независимых переменных)?

а) $F_{набл} = 1 - \frac{R^2}{2 \cdot m} \frac{n-m}{n}$

б) $F_{набл} = 1 - \frac{R^2}{2 \cdot m} \frac{n}{n-m}$

в) $F_{набл} = 1 - \frac{R^2}{2 \cdot m - 1} \frac{n}{n-m}$

г) $F_{набл} = 1 - \frac{R^2}{2 \cdot m - 1} \frac{n-m}{n}$

- 3) Для оценки надежности регрессионного уравнения наблюдаемое значение F -критерия Фишера должно быть:
- а) Больше расчетного значения F -критерия Фишера
 б) Менше расчетного значения F -критерия Фишера
 в) Равно расчетному значению F -критерия Фишера
 г) F -критерий Фишера не оценивает надежность уравнения.
- 4) Для определения количества значений, находящихся в выборке ниже или выше медианного значения используется:
- а) Коэффициент конкордации;
 б) Коэффициент корреляции рангов Спирмена;
 в) Коэффициент корреляции рангов Кэндалла;
 г) Критерий знаков.
- 5) Какая математическая модель приведена ниже?

$X_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \dots + a_{im}f_{mj} + d_j v_{ij}$, где $i = 1, n$; $j = 1, m$; $r = 1, m$, где n – число объектов наблюдения; m – число показателей, характеризующих объект; r – число значимых общих факторов; X_{ij} – центрированное значение j -го показателя (переменной) у i -го объекта исследования; f_r – r -й общий фактор; v_j – j -й характерный (индивидуальный) фактор, присущий только данной j -й переменной; a_{jr} – весовой коэффициент j -й переменной на r -м общем факторе; d_j – весовой коэффициент j -й переменной на j -м характерном факторе.

- а) Модель факторного анализа;
- б) Математическая модель метода главных компонент;
- в) Мультипликативная модель; Аддитивная модель.

6) Мультиколлинеарность факторных переменных - это:

- а) Отсутствие связи между факторными переменными;
- б) Тесная связь между факторными переменными;
- в) Многомерная связь между факторными переменными;
- г) Множественная регрессионная модель.

7) Элементами матрицы факторного отображения A по формуле $A = V \cdot \Lambda^{1/2}$ являются:

- а) Линейные коэффициенты корреляции между исходными переменными и главными компонентами
- б) Частные коэффициенты корреляции между исходными переменными и главными компонентами
- в) Множественные коэффициенты корреляции между исходными переменными и главными компонентами

8) Формула расчета расстояния Махаланобиса имеет следующий вид:

$$d(X_i, X_j) = \sqrt{(X_i - X_j) \cdot \Sigma^{-1} \cdot (X_i - X_j)};$$

$$а) d(X_i, X_j) = \sqrt{\sum_{k=1}^p (x_i^{(k)} - x_j^{(k)})^2};$$

$$б) d(X_i, X_j) = \sum_{s=1}^p |x_i^{(s)} - x_j^{(s)}|;$$

$$в) d(X_i, X_j) = \sum_{s=1}^p |x_i^{(s)} - x_j^{(s)}|;$$

$$г) p_{\min}(S_l, S_m) = \min_{X_i \in S_l, X_j \in S_m} d(X_i, X_j)$$

9) При использовании метода k-средних для классификации многомерных объектов в состав кластера включаются новые объекты таким образом, чтобы внутриклассовая дисперсия:

- а) Стремилась к минимуму;
- б) Стремилась к максимуму;
- в) Оставалась постоянной.

10) Дискриминантный анализ – совокупность статистических методов многомерной классификации объектов при наличии:

- а) Средних значений;
- б) «Обучающих» выборок;
- в) «Обычных» выборок;
- г) Коэффициентов корреляции.

11) Как интерпретируется DATA MINING?

- а) интеллектуальный анализ данных
- б) значение
- в) хранение
- г) перечисление данных

12) Если несколько событий связаны друг с другом, то это...

- а) Ассоциация
- б) Последовательность
- в) Классификация
- г) Кластеризация

13) Data Mining — это процесс обнаружения в сырых данных знаний, необходимых для:

- а) принятия решений в различных сферах человеческой деятельности
- б) замены аналитика в процессе принятия решений
- в) увеличения стоимости анализа данных
- г) объективных закономерностей

14) Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:

- а) Data Mining
- б) машинное обучение
- в) статистика
- г) Все вышеперечисленное

- 15) “Data Mining” представляет собой направление анализа, основной целью которого является
- а) поиск нетривиальных закономерностей внутри данных
 - б) разделение данных по определенному признаку
 - в) снижение размерности данных
 - г) исследование взаимосвязей между исходными переменными
- 16) Основой для каких систем не служит историческая информация, хранящаяся в БД в виде временных рядов?
- а) Классификации
 - б) Последовательности
 - в) Прогнозирования
 - г) Ассоциации
- 17) Data Mining — это процесс обнаружения в сырых данных
- а) ранее сформулированных гипотез
 - б) неочевидных закономерностей
 - в) практических закономерностей
 - г) большого количества закономерностей
- 18) Выберите характеристику, наиболее подходящую для Data Mining
- а) подходит для понимания ретроспективных данных
 - б) опирается на ретроспективные данные для получения ответов на вопросы о будущем
 - в) подходит для обобщения ретроспективных данных
 - г) подходит для классификации ретроспективных данных
- 19) Для данных небольшого объема методики Data Mining:
- а) неприменимы
 - б) применимы
 - в) применимы, но дают неверные результаты
 - г) применимы только для однородных данных
- 20) Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:
- а) быть очевидными
 - б) быть неочевидными
 - в) быть практически полезными
 - г) быть объективными
- 21) Исходные данные для Data Mining:
- а) представляют собой результат специально организованного статистического наблюдения б) носят спонтанный характер
 - в) засорены большим количеством грубых ошибок и аномальных наблюдений
 - г) как правило, имеют низкую однородность
- 22) Какое требование к переработке информации не верно?
- а) Данные имеют неограниченный объем
 - б) Данные являются разнородными
 - в) Результаты должны быть конкретны и понятны
 - г) Инструменты для обработки сырых данных должны быть сложны в использовании
- 23) Целесообразность применения методик Data Mining определяется:
- а) большим объемом исходной информации;
 - б) практической задачей выявления неочевидных закономерностей; наличием в данных неоднородности; все перечисленное
- 24) По логике вычисления средней в совокупности, заведомо засоренной «грубыми ошибками», вес (роль) аномальных наблюдений в формировании среднего значения:
- а) увеличивается
 - б) уменьшается
 - в) не влияет на среднюю
 - г) аппроксимируется
- 25) В статистике робастным оцениванием называют:
- а) способы статистического анализа, направленные на снижение влияния на результат аномальных значений и грубых ошибок
 - б) оценку распределений, имеющих отклонения от нормального;
 - в) оценку устойчивости совокупности к внешним и внутренним воздействиям;
 - г) оценку колеблемости данных

- 26) Требованием к проведению дисперсионного анализа является**
- а) равенство групповых дисперсий
 - б) равенство групповых средних
 - в) равные объемы выборки
 - г) все вышеперечисленное
- 27) Для нечисловых данных может применяться**
- а) факторный анализ
 - б) многомерное шкалирование
 - в) компонентное моделирование
 - г) регрессионный анализ
- 28) При проведении цензурирования большого объема данных для исключения подозрительных значений в случае отсутствия нормальности целесообразно применять** а) среднее значение
- б) t-критерий Стьюдента
 - в) правило Томпсона
 - г) подходят все варианты
- 29) При наличии в статистической совокупности аномальных наблюдений и грубых ошибок**
- а) такие наблюдения отбрасываются (исключаются из совокупности и анализ проводится без них)
 - б) такие наблюдения модифицируются (заменяются на значения, близкие к ним);
 - в) такие наблюдения заменяются на среднюю (математическое ожидание);
 - г) все вышеперечисленное;
- 30) Группировка является одним из хороших способов получения робастных оценок. По Вашему мнению, наименьшие потери информации (дисперсии) об исходной совокупности имеет:**
- а) группировка равными интервалами
 - б) группировка равными частотами
 - в) группировка равными вероятностями
 - г) асимптотическая группировка (по F-критерию Фишера)
- 31) Существенные расхождения между сгруппированными и несгруппированными данными могут быть вызваны**
- а) засорением выборки аномальными значениями;
 - б) ошибками в расчетах;
 - в) неверным предположением о законе распределения генеральной совокупности;
 - г) все вышеперечисленное
- 32) Дисперсионный анализ при отсутствии нормальности распределения** а) применим
- б) неприменим
 - в) применим с ограничениями
 - г) требует преобразования данных
- 33) Наличие нормального распределения требует**
- а) факторный анализ
 - б) многомерное шкалирование
 - в) метод главных компонент
 - г) дисперсионный анализ
- 34) Применение методов Data Mining для данных большого объема:**
- а) снимает проблему наличия выбросов в данных
 - б) снимает проблему содержательного анализа переменных
 - в) решает проблему мультиколлинеарности в данных
 - г) ничего из перечисленного
- 35) Для поиска нетривиальных закономерностей в данных статистические методы и модели**
- а) имеют ограниченное применение (применяются только в исключительных случаях)
 - б) не используются
 - в) применяются наравне со специфическими для Data Mining методами
 - г) не могут быть использованы для данных большого объема
- 36) Методики Data Mining находят широкое применение вследствие:**
- а) современных возможностей получения, хранения и обработки большого объема информации
 - б) поиск нетривиальных закономерностей позволяет существенно повысить показатели экономической эффективности
 - в) методики Data Mining являются более совершенными по отношению к традиционным статистическим методам
 - г) только а) и б), но не в)

- 37) Выбросом (аномальным наблюдением) принимается такой объект (элементарная единица совокупности), данные о котором:**
- а) являются выбросом хотя бы по одной переменной
 - б) являются выбросом сразу по всем переменным
 - в) содержат пропущенные значения
 - г) возможны а) и б)
- 38) Для выявления структуры и однородных групп в данных, применяется**
- а) иерархический кластерный анализ
 - б) метод К средних
 - в) дискриминантный анализ
 - г) метод главных компонент
- 39) Если необходимо максимизировать дисперсию между кластерами, для объединения объектов используется**
- а) метод одиночной связи
 - б) метод полных связей
 - в) метод средней связи
 - г) метод Уорда
- 40) Задачей дискриминантного анализа является**
- а) выявление существующих внутри совокупности классов, заведомо неизвестных
 - б) определение принадлежности объектов совокупности к той или иной группе
 - в) выявление целостности данных
 - г) снижение размерности данных
- 41) Для кластерного анализа большой объем совокупности**
- существенен несущественен
 - должен иметь нормальное распределение не должен
 - содержать выбросов
- 42) Кластерный анализ позволяет проводить**
- а) группировку объектов
 - б) группировку признаков
 - в) группировку объектов и группировку признаков
 - г) выделение главных факторов
- 43) Метод Уорда заключается в следующем**
- а) на основании матрицы расстояний выбираются ближайшие объекты, которые образуют начальные кластеры, затем в них включаются наиболее близкие объекты, и т. д.
 - б) включение объекта в кластер происходит, если расстояние между объектами не ниже заданного уровня
 - в) каждый кластер изначально состоит из одного объекта; вначале объединяются ближайшие кластеры; дальнейшее объединение кластеров происходит при минимальных приращениях вариации внутри кластеров
 - г) объекты включаются в кластер, если среднее значение близости не ниже заданного порогового уровня
- 44) Функционалом называют**
- а) критерий качества разбиения
 - б) формулу расстояния между объектами
 - в) условие оптимальности разбиения
 - г) количество итераций
- 45) В случае, если исследуемые признаки коррелированы, для кластерного анализа применяется**
- а) квадрат Евклидова расстояния
 - б) взвешенное Евклидово расстояние
 - в) расстояние «Городских кварталов»;
 - г) расстояние Махаланобиса
- 46) В случае, если необходимо получить наиболее однородные группы, для объединения объектов в кластеры применяется**
- а) метод одиночной связи
 - б) метод полных связей
 - в) метод средней связи
 - г) метод Уорда
- 47) Целевая функция в кластерном анализе представляет собой**
- а) количество итераций

- б) критерий качества разбиения
- в) расстояние между объектами
- г) условие оптимальности, согласно которому происходит разбиение **48) Метод**

группировки заключается:

- а) в формировании новых классов, если их количество и состав неизвестны;
- б) в разнесении наблюдений по существующим классам, для которых известны их параметры;
- в) в ликвидации грубых ошибок и аномальных наблюдений;
- г) в проверке данных на однородность **49) Метод**

классификации заключается:

- а) в формировании новых классов, если их количество и состав неизвестны;
- б) в разнесении наблюдений по существующим классам, для которых известны их параметры;
- в) в ликвидации грубых ошибок и аномальных наблюдений;
- г) в проверке данных на однородность

50) Задачей иерархического кластерного анализа является

- а) выявление существующих внутри совокупности классов, заведомо неизвестных
- б) определение принадлежности объектов совокупности к той или иной группе
- в) выявление однородности данных
- г) снижение размерности данных

51) В случае, если необходимо получить наиболее компактные группы, для объединения объектов в кластеры применяется

- а) метод одиночной связи
- б) метод полных связей
- в) метод средней связи
- г) метод Уорда

52) Главные факторы содержат информацию

- а) о самих наблюдениях
- б) о вариации или дисперсии наблюдений
- в) об исходных переменных
- г) о геометрии исходных данных

53) Отличие авторегрессионного процесса от процесса скользящей средней состоит в том, что процесс авторегрессии

- а) имеет лучшую память о предыдущем состоянии
- б) помнит компонент прошлого случайного шума
- в) более сглаженный
- г) все вышеперечисленное

54) Какую иерархическую структуру создают деревья решений?

- а) "ЕСЛИ... ТО..."
- б) "НИ... НИ..."
- в) "КОГДА... ТО..."
- г) "... НИКОГДА..."

55) При проверке значимости частного коэффициента корреляции, найденного при $n=16$, используется распределение

- а) Нормальное
- б) Стюдента
- в) Пирсона
- г) Пуассона **56) В каких пределах изменяется множественный коэффициент**

корреляции? а) $0 \leq r_{y/xz} \leq 1$

- б) $-1 \leq r_{y/xz} \leq 1$
- в) $-\sqrt{2} \leq r_{y/xz} \leq \sqrt{2}$
- г) $0 \leq r_{y/xz} \leq \sqrt{2}$

57) Задачей логистической регрессии является

- а) выявление существующих внутри совокупности классов, заведомо неизвестных
- б) определение принадлежности объектов совокупности к той или иной группе
- в) выявление целостности данных
- г) снижение размерности данных

58) Главные компоненты содержат информацию

- а) о самих наблюдениях
- б) о вариации или дисперсии наблюдений

- в) об исходных переменных
- г) о геометрии исходных данных

59) Парный коэффициент корреляции между факторами равен 1. Это означает

- а) наличие нелинейной функциональной связи
- б) отсутствие связи

- в) наличие функциональной связи
 - г) отрицательную линейную связь
- 60) Как называется класс систем, архитектура которых имеет аналогию с построением нервной ткани из нейронов?**
- а) Статистические пакеты
 - б) Деревья решений
 - в) Нейронные сети
 - г) Генетические алгоритмы
- 61) На основании 20 наблюдений выяснено, что выборочная доля дисперсии случайной величины y , вызванной вариацией x , составит 64%. Чему равен выборочный парный коэффициент корреляции**
- а) 0,64
 - б) 0,36
 - в) 0,8
 - г) 0,8 или -0,8
- 62) На основании 20 наблюдений выяснено, что выборочная доля дисперсии случайной величины y , вызванной вариацией x , составит 36%. Чему равен выборочный парный коэффициент корреляции:**
- а) 0,36
 - б) 0,16
 - в) 0,6
 - г) + 0,6 или -0,6
- 63) На основании 20 наблюдений выяснено, что выборочная доля дисперсии случайной величины y , вызванной вариацией x , составит 81%. Чему равен выборочный парный коэффициент корреляции:**
- а) 0,81
 - б) 0,36
 - в) 0,9
 - г) + 0,9 или -0,9
- 64) Если в данных отчетливо выделяются отдельные группы объектов, то это:** а) Ассоциация
- б) Последовательность
 - в) Классификация
 - г) Кластеризация
- 65) Задачей компонентного моделирования является**
- а) выявление существующих внутри совокупности классов, заведомо неизвестных
 - б) определение принадлежности объектов совокупности к той или иной группе
 - в) выявление целостности данных
 - г) снижение размерности данных
- 66) Главные факторы, полученные на основе факторного анализа любым некомпонентным методом**
- а) линейно неотделимы
 - б) имеют нулевую корреляцию
 - в) коррелированы
 - г) ортогональны
- 67) Имеется таблица сопряженности о количестве предприятий, разделенных по организационной форме и величине (крупные, средние, мелкие). Для анализа взаимосвязей между этими факторами в данном случае не применим**
- а) кластерный анализ
 - б) факторный анализ
 - в) дисперсионный анализ
 - г) метод главных компонент
- 68) Для дискриминантного анализа исходные данные**
- а) должны быть нормализованы
 - б) должны быть достаточно большого объема в зависимости от количества параметров
 - в) должны быть линейно зависимы
 - г) всё вышеперечисленное
- 69) Применение адаптивных моделей семейств Хольта, Брауна, Уинтерса подразумевает содержательное наличие тренда**
- а) верно
 - б) неверно
 - в) необязательно
 - г) верно, но с поправками при построении моделей

- 70) Модель авторегрессионного скользящего среднего (ARMA) отличается от модели интегрированного авторегрессионного скользящего среднего (ARIMA), тем, что ARIMA**
- а) имеет тенденцию оставаться вблизи долгосрочного среднего значения
 - б) описывает нестационарный процесс
 - в) обладает улучшенной краткосрочной памятью
 - г) все вышеперечисленное
- 71) Многомерное нормальное распределение могут составлять**
- а) все значения, которые может принимать случайный вектор
 - б) все значения, которые принимают несколько совместных случайных величин
 - в) . все значения, которые принимает любой набор взаимонезависимых случайных величин
 - г) . все значения, которые принимает любой набор совместных случайных величин
- 72) У исследователя имеются данные по показателям деятельности нескольких предприятий. Некоторые из этих показателей являются количественными (величина балансовой прибыли, фондоотдача, показатели устойчивости, ликвидности, и т.д.), некоторые – качественными (организационная форма и пр.). Для выявления сходств отдельных наблюдений и групп предприятий наиболее целесообразно применить а) дисперсионный анализ**
- б) факторный анализ
 - в) шкалирование
 - г) метод главных компонент
- 73) Задачей "дерева решений" является**
- а) выявление существующих внутри совокупности классов, заведомо неизвестных
 - б) определение принадлежности объектов совокупности к той или иной группе
 - в) выявление целостности данных
 - г) снижение размерности данных
- 74) Задачей "случайного леса" является**
- а) выявление существующих внутри совокупности классов, заведомо неизвестных
 - б) определение принадлежности объектов совокупности к той или иной группе
 - в) выявление целостности данных
 - г) снижение размерности данных
- 75) Задачей нейросетевого моделирования является**
- а) выявление существующих внутри совокупности классов, заведомо неизвестных
 - б) определение принадлежности объектов совокупности к той или иной группе
 - в) выявление целостности данных
 - г) снижение размерности данных
- 76) Для выделения групп наиболее похожих предприятий по организационной форме и величине лучше всего подходит**
- а) дисперсионный анализ
 - б) факторный анализ
 - в) шкалирование
 - г) метод главных компонент
- 77) Для нейросетевого моделирования проблема выбора переменных**
- а) не существует
 - б) основана на содержательном анализе
 - в) решается на основе подгонки модели
 - г) зависит от кластеризации данных
- 78) Можно ли интерпретировать нейросетевую модель с точки зрения факторов, влияющих на результативные параметры?**
- а) можно, если количество скрытых факторов совпадает с числом параметров, влияющих на совокупность
 - б) нельзя
 - в) зависит от ситуации
 - г) скрытые факторы интерпретируются так же, как главные компоненты
- 79) Метод классификации "случайный лес" склонен к "переобучению" в случае:**
- а) большого количества исходных переменных
 - б) малого количества исходных наблюдений
 - в) малого количества исходных переменных
 - г) большого количества исходных наблюдений
- 80) Выберите верное утверждение**
- а) Нейросетевое моделирование является аналогом работы человеческого мозга;
 - б) Нейросетевые модели позволяют не проводить содержательный анализ исходных факторных признаков;

- в) Нейросетевые модели не имеют содержательную интерпретацию;
г) Нейросетевые модели не могут быть использованы для подгонки "белого шума" 81) Какое значение может принять коэффициент детерминации? а) 0,68
б) -0,68;
в) 1,2
г) -1,2
- 82) Сходство задачи метода главных компонент и факторного анализа заключается в:
а) снижении размерности данных при минимальной потере информации
б) выявлении линейных комбинаций исходных признаков, объясняющих изменчивость совокупности
в) поиске подмножеств меньшей размерности
г) определении действительного числа главных факторов
- 83) Для выделения латентных компонент и определения их количества применяется а) факторный анализ
б) многомерное шкалирование
в) дисперсионный анализ
г) дерево решений
- 84) В ...процедурах начальным является разбиение, состоящее из n одноэлементных классов, а конечным - из одного класса;
а) агломеративных
б) дивизимных
в) дисконтированных
г) иерархических
- 85) Наиболее употребительными расстояниями и мерами близости между классами объектов являются: (выберите необходимый вариант)
а) расстояние, измеряемое по принципу "ближайшего соседа"
б) расстояние, измеряемого по принципу "дальнего соседа"
в) расстояние, измеряемое по принципу "родственной связи"
г) расстояние, измеряемое по принципу «значимой связи»
- 86) Оценка значимости моделей семейств Бокса-Дженкинса основывается
а) на количестве переменных авторегрессии и скользящей средней
б) на наличии автокорреляции в регрессионных остатках
в) на значимости переменных авторегрессии и скользящей средней
г) на подогнанности под поведение данных
- 87) Чему равен коэффициент корреляции показателя самого на себя а) 0
б) 1
в) -1
г) ∞
- 88) Многомерное шкалирование решает задачу сходств или различий объектов и представления их в меньшей размерности, основываясь на
а) максимизации дисперсии вокруг главных осей
б) сохранении геометрической структуры данных
в) извлечении латентных переменных
г) извлечении главных координатных осей
- 89) Коэффициент детерминации между x и y характеризует
а) - долю дисперсии y , обусловленную влиянием не входящих в модель факторов
б) долю дисперсии y , обусловленную влиянием x
в) долю дисперсии x , обусловленную влиянием не входящих в модель факторов
г) направление зависимости между x и y
- 90) Для какого из перечисленных видов анализа данные можно не нормировать? а) кластерный анализ
б) дискриминантный анализ
в) метод главных компонент
г) регрессионный анализ
- 91) Параметры авторегрессии в моделях ARMA/ARIMA содержат информацию:
а) о предыдущих уровнях ряда динамики
б) о предыдущих изменениях ряда динамики
в) о предыдущей дисперсии ряда динамики
г) о предыдущих значениях "белого шума" ряда динамики
- 92) Параметры скользящей средней в моделях ARMA/ARIMA содержат информацию:
а) о предыдущих уровнях ряда динамики

- б) о предыдущих изменениях ряда динамики
 - в) о предыдущей дисперсии ряда динамики
 - г) о предыдущих значениях "белого шума" ряда динамики
- 93) При декомпозиции ряда динамики с постоянной амплитудой колебаний целесообразно использовать:**
- а) мультипликативную модель
 - б) аддитивную модель
 - в) нелинейную модель
 - г) регрессионную модель
- 94) При декомпозиции ряда динамики с нарастающей/затухающей амплитудой колебаний целесообразно использовать:**
- а) мультипликативную модель
 - б) аддитивную модель
 - в) нелинейную модель
 - г) регрессионную модель
- 95) Модели семейства ARIMA подходят для моделирования:**
- а) стационарных рядов динамики
 - б) нестационарных рядов динамики
 - в) рядов распределения
 - г) ошибок прогнозирования
- 96) Большинство программ, реализующих алгоритм иерархической классификации, предусматривает графическое представление результатов классификации в виде а) дендрограммы**
- б) блок-схемы
 - в) главных компонент
 - г) многомерного распределения
- 97) При применении метода главных компонент**
- а) дисперсия исходного набора максимизируется вдоль главных осей
 - б) дисперсия минимизируется вдоль главных осей
 - в) дисперсия минимизируется вдоль регрессии
 - г) моделируется матрица корреляций
- 98) Случайный вектор можно рассматривать как**
- а) зафиксированные значения нескольких независимых параметров объекта
 - б) многомерную случайную величину
 - в) значения нескольких несовместных случайных величин
 - г) значения нескольких совместных случайных величин
- 99) Опорой случайного вектора называются**
- а) все допустимые исходы наблюдения
 - б) счетное количество исходов наблюдения
 - в) все ненулевые исходы наблюдения
 - г) несчетное количество исходов наблюдения
- 100) Многомерная случайная величина представляет собой**
- а) набор параметров исследуемого объекта, зафиксированных в определенный момент времени
 - б) точка в многомерном пространстве координат
 - в) набор совместных исходов наблюдения одного объекта
 - г) всё вышеперечисленное
- 101) Метод канонических корреляций целесообразно применять в случае**
- а) исходный набор признаков можно разделить на несколько линейных комбинаций
 - б) в исходном наборе признаков встречаются коллинеарные переменные
 - в) число исходных переменных достаточно большое
 - г) все перечисленное
- 102) Канонические переменные**
- а) взаимно не коррелированы
 - б) являются линейными комбинациями исходного набора факторов
 - в) выбираются таким образом, чтобы канонические корреляции были максимальны г) все перечисленное
- 103) Основное назначение метода главных компонент заключается в следующем**
- а) выявление факторов, в наибольшей степени влияющих на распределение
 - б) снижение размерности данных без потери информативности
 - в) построение многофакторной модели совокупности

г) всё вышеперечисленное

104) Кластерный анализ методом К-средних может применяться:

- а) для группировки данных
- б) для классификации данных
- в) для задач группировки и классификации
- г) для моделирования данных

105) В отличие от логистической регрессии, дискриминантный анализ

- а) может использоваться для небольшого объема исходной информации
- б) требует большего объема исходной информации
- в) применим только для классификации дихотомической переменной
- г) представляет собой нелинейную модель классификации

106) Для оценки значимости канонических корреляций наиболее целесообразно с точки зрения робастности использовать критерий:

- а) след Пиллаи
- б) критерий Хи-квадрат Пирсона
- в) F-критерий Фишера
- г) лямбда Уилкса

107) Какое значение могут принимать канонические корреляции? а) 0,68

- б) -0,68;
- в) 1,2
- г) -1,2

108) Факторные нагрузки в каноническом моделировании

- а) показывают процент объясненной исходной вариации
- б) показывают процент вариации одного подвектора, который может быть объяснен при наличии второго;
- в) показывают степень взаимосвязи с исходными переменными;
- г) показывают процент излишней объясненной вариации

109) Критерий чрезмерности в каноническом моделировании

- а) показывает процент объясненной исходной вариации
- б) показывает процент вариации одного подвектора, который может быть объяснен при наличии второго;
- в) показывает степень взаимосвязи с исходными переменными;
- г) показывает процент излишней объясненной вариации

110) "Случайный лес" позволяет проводить

- а) группировку объектов
- б) группировку признаков
- в) группировку объектов и группировку признаков
- г) выделение главных факторов

Критерии оценки (в баллах):

Рубежный контроль в виде комплексного теста, формируемого из тестовых заданий по темам курса

- если тест выполнен в полном объеме (85 и более%) – 4 балла;
- если тест выполнен по большей части (70-84%) – 3 балла;
- если тест выполнен наполовину (50-69%) – 2 балл;
- если тест выполнен в объеме меньше половины (до 50%) – 1 балла;
- если тест выполнен не выполнен вообще – 0 баллов;

Творческие задания
по дисциплине «Интеллектуальный анализ данных в экономике и
социальной сфере»

Индикаторы достижения: УК-1.2, ПК-1.3

Тема 1: Основы интеллектуального анализа данных

Задание 1. Соберите данные и исследуйте описательную статистику по одному из следующих статистических наблюдений, или самостоятельно выбранной области:

- 1) обследование работников промышленных предприятий;
- 2) перепись оборудования на промышленных предприятиях;
- 3) обследование семей рабочих и служащих;
- 4) обследование строительных организаций;
- 5) обследование торговых предприятий;
- 6) изучение спроса на некоторые товары;
- 7) изучение общественного мнения по отдельным вопросам.

По указанному наблюдению определите цель и задачи наблюдения; объекты и единицу наблюдения; основные признаки, подлежащие регистрации; вид, форму и способ наблюдения.

Задание 2. Соберите данные по нескольким (2-4) показателям в интересующей вас сфере, или выбранной из следующего перечня:

- 1) обследование работников промышленных предприятий;
- 2) перепись оборудования на промышленных предприятиях;
- 3) обследование семей рабочих и служащих;
- 4) обследование строительных организаций;
- 5) обследование торговых предприятий;
- 6) изучение спроса на некоторые товары;
- 7) изучение общественного мнения по отдельным вопросам.

По указанному наблюдению определите цель и задачи наблюдения; объекты и единицу наблюдения; основные признаки, подлежащие регистрации; вид, форму и способ наблюдения.

Для собранных данных укажите тип данных (временной ряд либо данные об одном временном срезе), по каждой переменной: вид распределения, тип переменных (количественная, качественная, дискретная, непрерывная, номинальная и т.д.)

Пользуясь средствами R, рассчитайте описательные статистики, интерпретируйте показатели и сделайте выводы.

Задание 3. На основе тщательного изучения выборки объемом 868, извлеченной из 11013 хранящихся на складе контейнеров, обнаружено, что 3,6% контейнеров не готовы к отгрузке.

а) определите стандартную ошибку, связанную с приведенной выше оценкой процента, и поясните полученный результат;

б) была бы неожиданной информация, что в действительности 4% из 11013 контейнеров не готовы к отгрузке? Почему? Если 10%? Почему?

Опрос клиентов показал, что из случайно отобранных 200 клиентов 42 недовольны постпродажной поддержкой и обслуживанием.

а) Вычислите основные статистики: размер выборки n , выборочный процент w , стандартную ошибку доли S_p .

б) Постройте двусторонний 95% доверительный интервал для процента недовольных среди *всех* клиентов;

в) Пусть генеральная совокупность состоит из 28209 клиентов. Рассчитайте количество людей, недовольных постпродажным сервисом, для граничных точек доверительного интервала п. б).

г) сформулируйте и обоснуйте результаты расчетов.

Задание 4. Часть сборочной линии необходимо регулировать, если консистенция подаваемого пластика становится слишком вязкой или недостаточно вязкой по сравнению со значением вязкости 56,0. Вы решаете производить настройку только тогда, когда есть уверенность, что система “вышла из-под контроля”. Среднее значение вязкости в последних 13 измерениях составило 51,22 со стандартным отклонением 3,18.

1. Сформулируйте нулевую и исследовательскую (альтернативную) гипотезы для двусторонней проверки.

2. Выполните двусторонний тест при уровне значимости 5% и опишите полученный результат.

3. Выполните двусторонний тест при уровне значимости 1% и опишите полученный результат.

4. Сформулируйте, если это возможно, односторонний вывод на основе результата двустороннего теста на уровне значимости 5%.

Задание 5. Пекарня изготавливает хлеб, на этикетках которого указан вес 500 грамм. Ниже приведены значения веса случайно отобранного хлеба (грамм): 510, 480, 490, 505, 500, 510, 515, 495, 480, 520.

1. Определите 95% доверительный интервал для среднего значения веса хлеба 2. Сформулируйте гипотезы H_0 и H_1 .

3. Выполните проверку гипотезы (двусторонняя, уровень 0,05) и поясните полученные результаты.

Тема 2. Первичная обработка данных для интеллектуального анализа

Задание 1. Задание должно быть выполнено на основе информации, приведенной в Приложении 1 «Основные технико-экономические показатели работы предприятия». Для этого:

1. По данным приложения 1 следует произвести методом случайного бесповторного отбора выборку 30 предприятий по одному из показателей работы. Показатели распределены между студентами в зависимости от начальной буквы их фамилий:

номер колонки прилож.1	табл.1 начальная буква фамилии студента	номер колонки табл.1прилож.1	начальная буква фамилии студента
1	А, Б	8	М
2	В, Г	9	Н, Р
3	Д, Е	10	П
4	Ж, З	11	Щ, Ю
5	И, Я	12	У, Ф, Х
6	К	13	Ц, Ч, Ш
7	Л, О	14	С, Т

Случайная выборка может быть проведена следующим образом. В новую таблицу Excel необходимо перенести столбец с данными. В соседней колонке затем генерируют, напротив каждого значения, ряд случайных чисел с помощью функции СЛЧИС(). Поскольку значения случайных чисел меняются каждый раз при редактировании ячеек, необходимо их зафиксировать: выделить столбец случайных чисел – «копировать» - и вставить вместо этих же значений следующим образом: «специальная вставка» - «значения». Функция СЛЧИС(), таким образом, заменяется выведенными случайными числами. Весь массив данных затем сортируют по возрастанию столбца со случайными числами. Первые n значений исходного показателя будут представлять собой случайную выборку (в данном случае 30).

2. По выборочной совокупности рассмотрите закономерность распределения исследуемого признака. Для этого:
 - а. пользуясь формулой Стерджесса, постройте интервальный ряд распределения и изобразите его графически в виде гистограммы, полигона и кумуляты.
 - б. рассчитайте описательные характеристики распределения: размах, среднюю арифметическую, моду, медиану, дисперсию, среднее квадратическое отклонение, коэффициент вариации и коэффициенты асимметрии и эксцесса. Проанализируйте исчисленные показатели и

сделайте выводы. Для асимметрии и эксцесса сделайте выводы о их наличии в генеральной совокупности, проведя проверку значимости.

3. Так как вы рассматриваете выборочную совокупность, то целесообразно вычислить ошибку выборки и определить границы генеральной средней с заданной вероятностью.

С вероятностью 0,954 (confidence probability) определите ошибку выборки средней и границы генеральной средней. Сделайте выводы.

Предельная ошибка выборки ($\Delta_{\bar{x}}$) для бесповторного случайного отбора определяется по формуле:

$$\Delta_{\bar{x}} = t \sqrt{\frac{\sigma^2 \cdot n}{N \cdot (1 - \frac{n}{N})}}$$

где t – коэффициент доверия (confidence coefficient), определенный таблично по доверительной вероятности (P), при этом для $P=0.997$ коэффициент доверия $t=3$, для $P=0.954$ $t=2$, для $P=0.683$ $t=1$; n – численность выборки (30);

N – численность генеральной совокупности (100).

Границы генеральной средней ($\bar{x}$) определяются по формуле:

$$\bar{x} - \Delta_{\bar{x}} \leq X \leq \bar{x} + \Delta_{\bar{x}}$$

Критерии оценки (в баллах):

Творческие задания по отдельным темам:

- если творческое задание выполнено в полном объеме (85 и более%) – 4 балла;
- если творческое задание выполнено по большей части (70-84%) – 3 балла;
- если творческое задание выполнено наполовину (50-69%) – 2 балл;
- если творческое задание выполнено в объеме меньше половины (до 50%) – 1 балла;
- если творческое задание не выполнено вообще – 0 баллов;

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

Структура экзаменационного билета

Наименование оценочного средства							Максимальное количество баллов
1. Вопрос 1. Общая теория многомерных распределений.							5
Вопрос 2. Оценки частных и множественных коэффициентов корреляции							8
Практическое задание (расчетно-аналитическое) В таблице представлены данные о количестве заключенных договоров в разрезе страховых продуктов каждым агентом. Используя статистический ППП, по представленным данным необходимо: 1) Построить интервальный вариационный ряд, гистограмму. 2) Вычислить выборочные характеристики: среднее значение, дисперсию, СКО, коэффициент вариации, асимметрию, эксцесс, моду, медиану. 3) Проверить гипотезу о нормальном законе распределения анализируемой случайной величины. 4) Построить 95% интервальные оценки математического ожидания, дисперсии и СКО.							12
вой №	Страхование имущества	Ипотечное страхование	Автострахование	Страхование туристов	ОСАГО	ДМС	
	26	10	6	4	11	7	
	38	18	17	3	10	2	
	62	4	5	5	9	1	
	56	24	26	3	4	8	
	19	20	21	1	6	6	
	85	8	7	3	21	7	
	44	9	2	1	14	8	
	63	10	9	0	5	5	
	52	7	7	4	8	6	
	47	5	1	1	9	10	
	57	10	9	5	10	1	
	28	13	12	1	12	2	
	20	4	5	1	8	7	
	70	18	18	3	16	3	
	46	17	17	0	10	5	
	30	13	13	1	13	3	
	38	15	14	2	9	14	
	38	17	17	1	10	12	
	53	15	15	1	15	10	
	48	15	14	1	9	9	

21	57	14	13	2	6	21	113														
22	62	26	29	1	13	2	134														
23	46	10	9	1	3	6	74														
24	59	27	30	4	10	11	142														
25	30	18	17	3	21	15	103														
26	68	10	9	2	7	4	100														
27	61	13	0	1	11	7	93														
28	56	5	6	1	1	12	82														
29	39	4	1	2	10	7	63														
30	53	11	10	3	7	7	91														
31	59	16	15	1	11	6	107														
32	39	0	3	2	6	14	64														
33	40	14	13	1	9	5	82														
34	59	11	10	1	12	3	96														
35	47	0	3	4	6	9	70														
36	22	31	38	5	13	7	115														
37	63	17	16	4	12	7	119														
38	53	1	3	2	18	10	88														
39	48	12	10	1	1	9	82														
40	57	23	24	0	11	12	127														
41	39	17	16	4	12	4	91														
42	64	8	0	4	10	2	88														
43	47	10	9	5	7	5	84														
44	40	5	2	4	12	6	69														
45	23	12	11	1	6	13	66														
46	27	3	4	0	2	8	44														
47	43	12	11	4	12	5	88														
48	72	19	19	2	19	1	132														
49	40	21	22	2	13	4	104														
50	39	16	16	2	16	14	103														
<i>Практическое задание (расчетно-аналитическое)</i> Для оценки удовлетворенности населения жизнью проведен социологический опрос в 80 странах мира. В таблице представлены агрегированные результаты проведенного опроса в разрезе стран, единицей измерения показателей является удельный вес респондентов, положительно оценивающих параметр. В качестве основных параметров удовлетворенности жизнью выбраны следующие: x_1 – работа, %; x_2 – здоровье, %; x_3 - материальное благополучие, %; x_4 - достижение поставленных целей, %; x_5 - социальный статус, %; x_6 - социальные контакты, %. Необходимо с помощью метода главных компонент с помощью ППП STATISTICA выявить факторы удовлетворенности населения жизнью и ранжировать страны по уровню удовлетворенности населения жизнью. <i>Результаты опроса населения стран об удовлетворенности жизнью</i> <table> <tr> <th rowspan="2">№</th><th rowspan="2">Страна</th><th colspan="6">Показатели</th></tr> <tr> <th>x_1</th><th>x_2</th><th>x_3</th><th>x_4</th><th>x_5</th><th>x_6</th></tr> </table>							№	Страна	Показатели						x_1	x_2	x_3	x_4	x_5	x_6	15
№	Страна	Показатели																			
		x_1	x_2	x_3	x_4	x_5	x_6														

1	Израиль	80	80	71	88	81	85
2	Греция	80	82	57	90	92	79
3	Словакия	76	72	47	85	78	93
4	Эстония	79	64	46	72	79	85
5	Венгрия	83	69	43	88	88	90
6	Португалия	90	80	47	92	93	87
7	Польша	82	72	67	87	91	89
8	Чили	81	73	68	90	93	83
9	Латвия	79	63	33	79	80	78
10	Хорватия	78	77	48	83	74	90
11	Болгария	73	67	29	77	77	81
12	Тринидад и Тобаго	76	82	40	97	93	85
13	Сербия	73	73	35	84	77	82
14	Белоруссия	66	55	34	70	71	88
15	Перу	74	72	54	96	89	79
16	Албания	72	75	43	78	68	79
17	Россия	74	56	36	79	83	88
18	Казахстан	82	68	51	88	81	88
19	Азербайджан	73	68	42	87	79	72
20	Босния и Герцеговина	76	75	39	80	67	74
21	Украина	71	55	23	74	78	81
22	Иран	71	82	55	87	81	62
23	Югославия	71	82	34	93	81	78
24	Грузия	63	50	22	86	83	54
25	Армения	61	53	31	93	89	67
26	Эквадор	80	76	57	98	93	78
27	Гайана	79	83	69	90	75	83
28	Ямайка	82	88	50	98	80	91
29	Турция	71	76	44	85	68	64
30	Доминиканская Республика	69	80	57	96	92	84
31	Сальвадор	82	80	60	97	89	72
32	Шри-Ланка	86	77	58	91	76	82
33	Таиланд	91	79	63	95	75	82
34	Боливия	83	79	67	94	90	82
35	Парагвай	85	84	63	93	96	89
36	Филиппины	83	77	68	96	94	77
37	Ботсвана	58	67	41	92	83	83
38	Молдова	68	60	39	79	73	83
39	Гана	84	87	64	95	77	84
40	Намибия	84	87	61	98	86	83
41	Гондурас	84	83	65	95	91	81
42	Индонезия	63	83	62	95	92	78
43	Киргизия	78	74	48	91	86	85
44	ЮАР	66	79	42	97	83	88
45	Таджикистан	78	75	69	91	76	65
46	Вьетнам	72	79	59	98	92	79
47	Марокко	69	88	71	90	89	85
48	Никарагуа	80	80	62	98	91	83
49	Индия	74	85	61	91	72	66
50	Камбоджи	80	69	51	81	87	82
51	Кения	57	70	25	98	78	79
52	Бангладеш	76	73	63	94	87	53
53	Гана	54	66	34	98	88	63
54	Камерун	63	69	40	93	85	73
55	Йемен	74	80	53	88	84	75
56	Мадагаскар	46	76	24	96	77	77
57	Мавритания	57	79	47	93	85	81
58	Нигерия	65	80	40	92	81	72
59	Уганда	53	64	35	96	79	85
60	Сенегал	39	68	27	89	85	81
61	Гаити	51	51	35	81	66	64
62	Ангола	72	67	54	90	83	58
63	Танзания	45	67	21	95	74	76
64	Замбия	48	78	34	93	83	62
65	Руанда	41	64	37	88	77	56
66	Малави	62	77	64	99	88	72
67	Судан	65	77	64	97	89	89
68	Афганистан	71	79	53	83	64	54
69	Гвинея	68	75	27	96	86	58
70	Эфиопия	50	79	33	89	74	76
71	Сьерра Леон	49	47	19	98	81	59
72	Центрально-Африканская республика	78	81	31	96	74	56
73	Мали	30	71	30	99	86	75

	74	Буркина-Фасо	46	70	27	94	83	73		
	75	Либерия	47	70	46	100	82	58		
	76	Чад	78	69	52	93	79	57		
	77	Мозамбик	74	82	46	93	89	75		
	78	Нигер	54	82	52	99	93	77		
	79	Конго	60	74	40	98	79	67		
	80	Зимбабве	49	72	27	91	81	81		

Задания, включаемые в зачетное задание
(в соответствии со структурой зачетного задания)

Полный перечень вопросов к зачету по дисциплине
«Первичная обработка данных для интеллектуального анализа»:

1. Современные программные пакеты средств статистического анализа.
2. Средства описательной статистики в R, Python: возможности и ограничения.
3. Анализ вариационного ряда в R, Python.
4. Инструменты описательной статистики в R, Python: дисперсия, мода, медиана, стандартная ошибка.
5. Анализ формы распределения в R, Python.
6. Уровень надежности в R, Python. Интерпретация полученных результатов.
7. Средства дисперсионного анализа в R, Python.
8. Оценка возможности проведения дисперсионного анализа: числовая и графическая. построение блочной диаграммы.
9. Оценка результатов проверки гипотез дисперсионного анализа, F-критерий.
10. Проведение двумерного дисперсионного анализа в R, Python.
11. Факторный анализ. Снижение размерности данных.
12. Методы снижения размерности и факторного анализа в R, PYTHON. Объединение и обобщение.
13. Выделение главных компонент по результатам анализа в R, PYTHON. Анализ главных компонент. Обзор результатов анализа.
14. Факторный анализ как метод классификации.
15. Подтверждающий факторный анализ (моделирование структурными соответствиями).
16. Анализ соответствий (изучение структуры качественных переменных).
17. Многомерный дисперсионный анализ.
18. Проведение основных этапов множественного регрессионного анализа в R, PYTHON. Корреляционная матрица. Выявление мультиколлинеарности.
19. Выбор числа переменных. Подгонка полиномиальных моделей. Определение степени и формы модели.
20. Анализ регрессионных остатков, выбросы. Критерии значимости. Критерий Дарбина-Уотсона.
21. Основное назначение кластерного анализа. Возможности кластерного и дискриминантного анализа в R, PYTHON.
22. Древовидная классификация (объединение). Меры расстояния. Правила объединения.
23. Метод К средних: общая логика и постановка задачи. Примеры анализа и интерпретации результатов.
24. Дискриминантный анализ: основная цель и вычислительный подход. Пошаговый дискриминантный анализ.

- 25. Дискриминация для двух групп.
- 26. Дискриминация для нескольких групп. Проведение классификации дискриминантным методом.
- 27. Авторегрессионное моделирование методом скользящих средних (ARMA, ARIMA) в R, PYTHON.
- 28. Проведение авторегрессионного моделирования динамики одномерных данных.
- 29. Выбор параметров временного лага и сезонных колебаний.
- 30. Проблема формальных критериев определения параметров модели.
- 31.** Многофакторное моделирование результативной переменной на основе авторегрессионных моделей.

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-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2. Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации	Знает верно и в полном объеме: критерии сопоставления различных вариантов решения поставленной задачи Умеет верно и в полном объеме: осуществлять критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи, отличать факты от мнений, интерпретаций и оценок при анализе собранной информации, сопоставлять и оценивать различные варианты решения поставленной задачи, определяя их достоинства и недостатки	Продвинутый
		ПК-1. Способен анализировать, обосновывать и выбирать решения	ПК-1.3. Рассматривает эффективность вариантов решений как соотношения между ожидаемыми уровнями использования ресурсов и их ценностью	Знает верно и в полном объеме: языки визуального моделирования, методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации бизнес-анализа, информационные технологии (программное обеспечение), применяемые в организации, в объеме, необходимом для целей бизнес-анализа Умеет верно и в полном объеме: оформлять результаты бизнес-анализа в соответствии с выбранными подходами, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, проводить оценку эффективности решения с точки зрения выбранных критериев, оценивать бизнес-возможность реализации решения с точки зрения выбранных целевых показателей	

Шкала оценивания		Формируемые компетенции	Индикатор достижения компетенции	Критерии оценивания	Уровень освоения компетенций
70 – 84 баллов	«хорошо»/ «зачтено»	УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.2. Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации	Знает с незначительными замечаниями: критерии сопоставления различных вариантов решения поставленной задачи Умеет с незначительными замечаниями: осуществлять критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи, отличать факты от мнений, интерпретаций и оценок при анализе собранной информации, сопоставлять и оценивать различные варианты решения поставленной задачи, определяя их достоинства и недостатки	Повышенный
		ПК-1. Способен анализировать, обосновывать и выбирать решения	ПК-1.3. Рассматривает эффективность вариантов решений как соотношения между ожидаемыми уровнями использования ресурсов и их ценностью	Знает с незначительными замечаниями: языки визуального моделирования, методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации бизнес-анализа, информационные технологии (программное обеспечение), применяемые в организации, в объеме, необходимом для целей бизнес-анализа Умеет с незначительными замечаниями: оформлять результаты бизнес-анализа в соответствии с выбранными подходами, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, проводить оценку эффективности решения с точки зрения выбранных критериев, оценивать бизнес-возможность реализации решения с точки зрения выбранных целевых показателей	
50 – 69	«удовлетворительно»/	УК-1.	УК-1.2.	Знает на базовом уровне,	Базовый

Шкала оценивания		Формируемые компетенции	Индикатор достижения компетенции	Критерии оценивания	Уровень освоения компетенций
баллов	«зачтено»	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации	с ошибками: критерии сопоставления различных вариантов решения поставленной задачи Умеет на базовом уровне, с ошибками: осуществлять критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи, отличать факты от мнений, интерпретаций и оценок при анализе собранной информации, сопоставлять и оценивать различные варианты решения поставленной задачи, определяя их достоинства и недостатки	
		ПК-1. Способен анализировать, обосновывать и выбирать решения	ПК-1.3. Рассматривает эффективность вариантов решений как соотношения между ожидаемыми уровнями использования ресурсов и их ценностью	Знает на базовом уровне, с ошибками: языки визуального моделирования, методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации бизнес-анализа, информационные технологии (программное обеспечение), применяемые в организации, в объеме, необходимом для целей бизнес-анализа Умеет на базовом уровне, с ошибками: оформлять результаты бизнес-анализа в соответствии с выбранными подходами, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, проводить оценку эффективности решения с точки зрения выбранных критериев, оценивать бизнес-возможность реализации решения с точки зрения выбранных целевых показателей	
менее 50 баллов	«неудовлетворительно»/ «не зачтено»	УК-1. Способен осуществлять	УК-1.2. Разрабатывает варианты	Не знает на базовом уровне: критерии сопоставления различных	Компетенции не сформированы

Шкала оценивания		Формируемые компетенции	Индикатор достижения компетенции	Критерии оценивания	Уровень освоения компетенций
		поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	решения проблемной ситуации на основе критического анализа доступных источников информации	вариантов решения поставленной задачи Не умеет на базовом уровне: осуществлять критический анализ собранной информации на соответствие ее условиям и критериям решения поставленной задачи, отличать факты от мнений, интерпретаций и оценок при анализе собранной информации, сопоставлять и оценивать различные варианты решения поставленной задачи, определяя их достоинства и недостатки	
		ПК-1. Способен анализировать, обосновывать и выбирать решения	ПК-1.3. Рассматривает эффективность вариантов решений как соотношения между ожидаемыми уровнями использования ресурсов и их ценностью	Не знает на базовом уровне: языки визуального моделирования, методы сбора, анализа, систематизации, хранения и поддержания в актуальном состоянии информации бизнес-анализа, информационные технологии (программное обеспечение), применяемые в организации, в объеме, необходимом для целей бизнес-анализа Не умеет на базовом уровне: оформлять результаты бизнес-анализа в соответствии с выбранными подходами, применять информационные технологии в объеме, необходимом для целей бизнес-анализа, проводить оценку эффективности решения с точки зрения выбранных критериев, оценивать бизнес-возможность реализации решения с точки зрения выбранных целевых показателей	