

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Первый проректор

А.В. Толстиков

23 марта 2022 г.



ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ

**Компьютерные технологии и статистические методы в геозкологических  
исследованиях**  
по научной специальности  
1.6.21. Геозкология

## 1. Паспорт оценочных материалов по дисциплине

№ п/п	Темы дисциплины в ходе текущего контроля, вид промежуточной аттестации	Код и содержание компетенции (или ее части)	Оценочные материалы (виды и количество)
1.	Введение в анализ данных	<p>- <b>ПК-17</b> - знать современные междисциплинарные проблемы геоэкологии и использовать фундаментальные представления о составе, строении, свойствах, процессах, структуре и функционировании геосфер Земли как среды обитания человека и других организмов в сфере профессиональной деятельности</p> <p>- <b>ПК-18</b> - уметь использовать основные теории, концепции и принципы в области геоэкологической деятельности, быть способным к системному мышлению, демонстрировать знание истории и методологии геоэкологии, расширяющие общепрофессиональную, фундаментальную подготовку</p> <p>- <b>ПК-19</b> - владеть методами, самостоятельно анализировать необходимую информацию с использованием современных информационных технологий, выявлять фундаментальные проблемы и выполнять разномасштабные научные и прикладные исследования в профессиональной области</p>	<p>Выполнение практических заданий. Основные формы контроля – собеседование, интерпретация полученных результатов, выводы. Проектирование и разработка собственного проекта, применение статистических методов анализа и прогнозирования для реализации в среде программирования R.</p>
2.	Проверка статистических гипотез		
3.	Корреляционно - регрессионный анализ в экологическом мониторинге		
4.	Обзор ПО для анализа данных. R и статистика		
5.	Анализ экологических данных в R.		
6.	Дисперсионный анализ экологических данных.		
7.	Дифференцированный зачет (4 семестр)		Вопросы к зачету

## 2. Виды и характеристика оценочных средств

### 1. Задания к выполнению

На каждом занятии аспирант получает задание по варианту, выполнение которого заканчивается ответами на контрольные вопросы к работе.

### 2. Устный опрос обучающихся в ходе практических занятий.

Данное оценочное средство используется на каждом лабораторном занятии. Оцениваются фактические знания аспирантов, их использование в ходе выполнения лабораторных работ.

3. Промежуточная аттестация – дифференцированный зачет.  
Итоговое собеседование проводится по списку вопросов, составленных к зачету.

### 3. Оценочные средства

#### Примеры заданий

##### Задание I.

1. Инсталлируйте пакет: `ggplot2`.

Создайте новый проект и добавьте в него файл скрипта, текстовый файл и файл R Markdown.

2. Проведите линейный дискриминантный анализ для данных из файла “seeds2.txt”.

Сравните число верных предсказаний и число ошибок с результатами других методов классификации. Сделайте выводы.

3. Загрузите в объект `seeds` данные из файла “seeds2.txt”.

Постройте несколько диаграмм. Зёрна разного сорта должны отображаться на диаграммах разным цветом. Используйте функции из пакета `ggplot2`.

Постройте линейный классификатор для сорта зерна от всех семи параметров.

Используя полученную модель сделайте предсказания по сортам зёрен для данных из объекта `seeds`. Подсчитайте число верных и ошибочных предсказаний. Сравните с результатами предсказаний других моделей.

##### Задание II.

В целях оценки воздействия окружающей среды на здоровье населения обследованы две территории. На территории «А» с повышенным загрязнением атмосферного воздуха в течение 1 года проверено 13 населенных пунктов, на контрольной территории «В», расположенной в зелёной зоне, - 9 пунктов. Данные обследования населенных пунктов приведены в таблице, где  $x$  и  $y$  – выборочные значения случайных величин  $X$  и  $Y$ , выражающих соответственно количество заболеваний органов дыхания на 1000 человек на территориях «А» и «В». Предполагая, что  $X$  и  $Y$  имеют нормальный закон распределения, проверить существенность различий этих районов по уровню заболеваний при  $\alpha = 0,05$ .

$x$	109	175	158	148	154	165	141	159	164	173	162	153
$y$	104	132	114	127	119	122	117	106	113	127	141	113

Указание. Факт влияния окружающей среды может быть обоснован опровержением гипотезы о равенстве средних количеств заболеваний на данных территориях.

Использовать статистику Стьюдента. Предварительно необходимо проверить гипотезу о равенстве дисперсий случайных величин  $X$  и  $Y$  с помощью критерия Фишера при  $\alpha_1 = 0,05$ . Реализовать в R.

#### Примерный перечень вопросов к дифференцированному зачету

1. Классификация случайных событий. Классическое определение вероятности. Свойства вероятности события. Примеры.
2. Несовместные и совместные события. Сумма событий. Теоремы сложения вероятностей.
3. Зависимые и независимые события. Произведение событий. Понятие условной вероятности. Теорема умножения вероятностей.
4. Формула полной вероятности. Формула Байеса.
5. Повторные независимые испытания. Формула Бернулли. Наивероятнейшее число наступлений события.
6. Понятие случайной величины и ее описание. Дискретная случайная величина и закон ее распределения. Независимые случайные величины.



7. Математическое ожидание дискретной случайной величины и его свойства. Дисперсия дискретной случайной величины и ее свойства.
8. Функция распределения случайной величины, ее определение, свойства и график.
9. Непрерывная случайная величина (НСВ). Вероятность отдельно взятого значения НСВ. Плотность вероятности НСВ. Математическое ожидание и дисперсия НСВ.
10. Случайная величина, распределенная по биномиальному закону, ее математическое ожидание и дисперсия. Геометрическое распределение.
11. Закон распределения Пуассона. Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона.
12. Равномерный закон распределения.
13. Показательный (экспоненциальный) закон распределения.
14. Нормальный закон распределения. Функция распределения нормально распределенной случайной величины и ее выражение через функцию Лапласа.
15. Формулы для определения вероятности: а) попадания нормально распределенной случайной величины в заданный интервал; б) ее отклонения от математического ожидания. «Правило трех сигм».
16. Закон больших чисел. Теорема Бернулли и ее значение. Центральная предельная теорема.
17. Основы статистики. Проверка статистических гипотез. Выборочное оценивание. Несмещенные, эффективные и состоятельные оценки. Интервальное оценивание
18. Доверительные интервалы для выборочного среднего и выборочной дисперсии. Распределения Стьюдента и «хи-квадрат». Ошибки первого и второго рода.
19. Среда R: сведения из истории создания, источники получения, меню и структура пакета, настройка пользовательской конфигурации, редактор данных, визуализация данных.
20. Типы данных в среде R. Основные операторы.
21. Предварительная обработка данных в задачах анализа: нормализация, стандартизация, обработка пропущенных значений.
22. Способы получения репрезентативных выборок. Реализация в среде R.
23. Корреляционный анализ. Коэффициент корреляции. Критерий Стьюдента. Построение доверительного интервала.
24. Линейный регрессионный анализ: постановка задачи, метод наименьших квадратов, свойства решения, приложения. Реализация в среде R.
25. Однофакторная линейная регрессионная модель: графическая интерпретация решения задачи, исследование признаков на независимость. Реализация в среде R.
26. Множественная линейная регрессия: постановка и решение задачи, примеры применения модели. Несмещённость оценок. Реализация в среде R.
27. Задачи классификации: общая постановка, виды, обзор методов решения, возможные приложения.
28. Метод k-ближайших соседей: постановка задачи, идея и алгоритм метода, графическая иллюстрация, вычислительные аспекты, примеры применения. Реализация в среде R.
29. Понятие деревьев решений и смежные понятия. Решение задачи классификации методом построения деревьев решений. Реализация в среде R.
30. Линейный дискриминантный анализ: постановка задачи, идея и алгоритм метода, графическая иллюстрация (для однофакторной задачи), примеры применения. Реализация в среде R.

### **Критерии оценивания текущей и промежуточной аттестации**

Критерии оценки ответов:

· **оценка «отлично»** выставляется обучающемуся, который дает полный, развернутый ответ на поставленный вопрос, показывает совокупность осознанных знаний о статистических методах анализа данных об объекте или явлении, может привести пример. Свободно владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы.

· **оценка «хорошо»** » выставляется обучающемуся, который дает полный, развернутый ответ на поставленный вопрос, показывает совокупность осознанных знаний о статистических методах анализа данных об объекте или явлении, может привести пример. Владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы, но допускает не большие неточности.

· **оценка «удовлетворительно»** выставляется обучающемуся, который дает ответ на поставленный вопрос, имеет знания о статистических методах анализа данных об объекте или явлении. Владеет компьютерными технологиями для анализа данных, создает модели, правильно интерпретирует результаты моделирования, умеет делать прогнозы, но допускает неточности

**оценка «неудовлетворительно»** выставляется обучающемуся за отсутствие ответа.