

«УТВЕРЖДАЮ»
 Декан биологического факультета МГУ
 Академик М.П.Киришчинов
 _____ 2015 г.



Рабочая программа дисциплины (модуля)

1. Код и наименование дисциплины (модуля): **«АНАЛИЗ ДАННЫХ И КОМПЬЮТЕРНЫЕ МЕТОДЫ В БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ»**
2. Уровень высшего образования – подготовка научно-педагогических кадров в аспирантуре.
3. Направление подготовки – **06.06.01 Биологические науки**. Направленность (профиль) программы – **Общая биология и экология**.
4. Место дисциплины (модуля) в структуре ООП: вариативная часть ООП (весенний семестр), спецкурс по выбору (читается на кафедре зоологии позвоночных)
5. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции (код компетенции)	Планируемые результаты обучения по дисциплине (модулю)
<i>УК-1: Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях</i>	Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях Код В1 (УК-1) Владеть:

	<p>Владеть: навыками критического анализа и оценки современных научных достижений и результатов деятельности по решению исследовательских и практических задач, в том числе в междисциплинарных областях Код В2 (УК-1)</p>
<p>УК-2 Способность проектировать и осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения с использованием знаний в области истории и философии науки.</p>	<p>Знать: методы научно-исследовательской деятельности Код З1 (УК-2)</p>
<p>УК-3: Готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач</p>	<p>Владеть: технологиями оценки результатов коллективной деятельности по решению научных и научно-образовательных задач, в том числе ведущейся на иностранном языке Код В2 (УК-3)</p>
<p>УК-4: Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языке</p>	<p>Владеть: навыками анализа научных текстов на государственном и иностранном языках Код В1 (УК-4) Знать: стилистические особенности представления результатов научной деятельности в устной и письменной форме на государственном и иностранном языках Код З2 (УК-4)</p>
<p>ОПК-1 Способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-</p>	<p>Уметь: собирать, отбирать и использовать необходимые данные и эффективно применять количественные методы их анализа</p>

Оценочные средства для промежуточной аттестации по дисциплине (модулю) приведены в Приложении.

6. Объем дисциплины (модуля) составляет 2 зачетных единицы, всего 72 академических часа, из которых 48 часов составляет контактная работа аспиранта с преподавателем (48 часов занятий лекционного типа) и 24 часа составляет самостоятельная работа аспиранта (выполнение домашних заданий).

7. Входные требования для освоения дисциплины (модуля), предварительные условия:

ЗНАТЬ: теоретические и методологические основы биологических научных исследований, общие представления о статистике

УМЕТЬ: собирать и проводить первичную обработку фактических данных, вырабатывать на основе рационального анализа экспериментальных результатов свою точку зрения и отстаивать ее во время дискуссии со специалистами и неспециалистами; читать и реферировать научную литературу, в том числе на иностранных языках.

ВЛАДЕТЬ: современными информационно-коммуникационными технологиями, иностранным языком.

8. Образовательные технологии: лекция с элементами дискуссии с аудиторией, игровые задания, практические упражнения с применением компьютерных программ.

9. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы						Самостоятельная работа обучающегося, часы		
		из них						из них		
Индивидуальные	Групповые	Занятия	Занятия	Учебные занятия, направленные на проведение	Всего	Выполнение домашних	Подготовка рефератов и	Всего		
Вводная часть Роль математики в биологии, два взгляда на проблему. Трудности применения классической статистики к биологическим данным. Статистика как общепринятые правила для унификации и сравнения результатов. Относительность суждений, логика описания результатов по типу «опыт-контроль». Статистика и планирование сбора материала. Статистика и биологический смысл, валидность исследования. Современный уровень образования биологов в области статистического анализа данных.	3	1	0,5			0,5	2	1		1
Формирование выборки и первичная обработка данных. Принципы работы с базами данных Этапы проведения исследования: планирование работы, сбор данных, формирование выборки и первичная	3	1	0,5			0,5	2	1		1

<p>обработка (группировка, выделение первостепенных данных), статистический анализ, интерпретация результатов. Логика работы с базами данных. Отличия СУБД (Систем Управления Базами Данных) на примере Microsoft Access от электронных таблиц Microsoft Excel. Основные компоненты базы данных (таблица, запрос, форма, отчет), их структура и назначение. Создание таблиц. Работа с записями. Контроль введенных данных. Изменение внешнего вида таблицы.</p>										
<p>Создание простых запросов Сортировка данных в таблице, использование функции поиска. Фильтрация данных. Создание запроса на основе фильтра. Запросы на выборку: отбор записей (данных) по условию и сортировка по нескольким полям таблицы, защита от введения новых записей в запрос, блокировка изменения уже существующих записей. Создание сложных условий отбора данных, использование основных операторов и функций.</p>	3	1	0,5			0,5	2	1		1
<p>Создание сложных запросов Использование языка SQL для создания запросов. Использование запросов на создание таблицы, запросов на удаление, обновление, добавление записей (данных). Схема данных, определение связей между таблицами. Использование группировки в</p>	3	1	0,5			0,5	2	1		1

запросах, групповые операции. Создание многотабличных запросов. Типы соединения таблиц в запросах. Перекрестные запросы. Работа с SQL при создании сложных запросов. Запросы на объединение данных.										
Создание связей баз данных с другими программами Общая логика связывания баз данных с другими компьютерными программами. Использование интерфейса ODBC для определения доступа программ к базе данных. Настройка подключений. Работа с запросами СУБД в программе статистического анализа Statistica. Настройка подключения СУБД к ГИС (Геоинформационной Системе) MapInfo. Создание карт на основе СУБД Microsoft Access.	3	1	0,5			0,5	2	1		1
Создание форм для ввода данных Назначение элемента «формы». Создание простых форм для ввода данных с помощью «мастера форм» в ACCESS. Настройка формы. Форматирование полей формы, условное форматирование. Создание полей со списком. Внедрение элементов управления в форму. Создание многотабличных форм. Общие сведения о создании отчетов.	3	1	0,5			0,5	2	1		1
Выбор метода анализа Основной принцип работы статистических критериев: критерий и модель, ошибка, мощность критерия. Статистическая	6	2	1			1	4	2		2

<p>значимость и понятие «статистически значимый». Одно- и двусторонние гипотезы. Вопросы, которые необходимо поставить перед тем, как выбрать метод. Единицы измерения и типы числовых данных: непрерывная шкала, частоты, ранги. Зависимые и независимые переменные. Взаимоисключаемость событий. Размер выборки. Описательная статистика и первичный анализ данных. Отображение материала на графиках, основные типы графиков (диаграммы рассеяния, гистограммы, «коробочки с усами», круговые диаграммы) и области их применения. Этапы выбора статистического критерия для анализа данных. Выбор между параметрическими и непараметрическими критериями. Проверка выборки на нормальность распределения: <i>критерий Колмогорова-Смирнова, критерий Шапиро-Вилка</i>. Трансформация данных. Проверка равенства дисперсий: <i>F-критерий Фишера</i>. Выбор группы критериев на основании решаемой задачи. Обзор основных групп критериев. Статистические ограничения конкретных критериев внутри выбранной группы.</p>										
<p>Сравнение частот и долей. Частотные критерии Частоты как тип данных. Сравнение частот 2</p>	6	2	1			1	4	2		2

<p>взаимоисключающих событий: <i>биномиальный критерий</i>. Сравнение долей проявления признака в двух выборках: <i>критерий двух пропорций</i>. Сравнение частот на интервальной шкале: <i>критерий хи-квадрат</i> (сравнение практических частот с теоретическими, сравнение двух рядов практических частот, объединение интервалов внутри ряда). Методы анализа четырехпольных таблиц: принципы построения таблицы, <i>критерий хи-квадрат</i>, <i>точный метод Фишера</i>, <i>критерий МакНемара</i>, поправка Ийтса.</p>										
<p>Критерии парного сравнения выборок</p> <p>- Параметрические критерии парного сравнения по средним значениям. Сравнение независимых выборок: <i>t-критерий для независимых выборок</i>. Сравнение зависимых (связанных) выборок: <i>t-критерий для зависимых выборок</i>. Принцип построения таблиц исходных данных для зависимых и независимых переменных в программе Statistica.</p> <p>- Параметрические критерии парного сравнения по разбросу данных. Сравнение дисперсий двух выборок: <i>F-критерий Фишера</i>.</p> <p>- Непараметрические критерии парного сравнения по средним тенденциям. Принцип ранжирования и</p>	6	2	1			1	4	2		2

<p>мощность непараметрических критериев. Сравнение независимых выборок: <i>критерий Манна-Уитни</i>. Сравнение зависимых (попарно связанных) выборок: <i>критерий знаков, критерий Вилкоксона для сопряженных пар</i>.</p> <p>- Непараметрические критерии парного сравнения по форме распределения. Сравнение независимых выборок: <i>критерий Колмогорова-Смирнова, критерий серий Вальда-Вольфовица</i>.</p>										
<p>Оценка взаимосвязи. Прямолинейная корреляция</p> <p>Различия между взаимосвязью (корреляцией) и зависимостью (регрессией). Оценка взаимосвязи двух выборок: коэффициент корреляции. Значение коэффициента корреляции и его достоверность. Параметрический <i>коэффициент корреляции Пирсона</i>. Непараметрический коэффициент корреляции <i>Спирмена, коэффициент корреляции тау Кендалла, коэффициент корреляции гамма</i>. Сравнение двух коэффициентов корреляции. Оценка взаимосвязи между тремя и более выборками: непараметрический <i>коэффициент согласованности Кендалла</i>, сравнение экспертных оценок. Оценка взаимосвязи между двумя бинарными признаками: <i>коэффициент сопряженности</i></p>	3	1	0,5		0,5	2	1			1

Φ_i , оценка согласованности простых поведенческих последовательностей.										
Множественное сравнение и оценка влияния факторов - Параметрический дисперсионный анализ ANOVA. Принцип работы и возможности применения метода: множественное сравнение, оценка влияния факторов. <i>Однофакторный дисперсионный анализ.</i> Апостериорное сравнение средних значений (критерий Тьюки). Поправка Бонферрони. <i>Двухфакторный дисперсионный анализ.</i> <i>Многофакторный дисперсионный анализ.</i> Введение ковариат. Задание способов взаимодействия факторов. <i>Дисперсионный анализ планов с повторными измерениями.</i> - Непараметрический дисперсионный анализ. Сравнение независимых выборок: <i>критерий Крускала-Уоллиса, критерий медианы.</i> Сравнение зависимых выборок (повторных наблюдений): <i>критерий Фридмана.</i> Особенности построения таблиц исходных данных в программе Statistica. Навыки работы в программной среде R. Дисперсионный анализ в среде R.	6	2	1			1	4	2		2
Выделение групп значимых факторов. Факторный анализ Основные предпосылки к использованию метода. Интерпретация результатов анализа. Оценка долей дисперсии, определяемых факторами; влияние отдельных переменных	3	1	0,5			0,5	2	1		1

на главные компоненты. Построение факторных диаграмм. Факторный анализ в среде R.										
Оценка зависимости. Регрессионный анализ Отбор значимых факторов методом дисперсионного анализа. Модель линейной зависимости от факторов и ограничения применения регрессионного метода. Анализ остатков, дистанция Кука. Информационный критерий Акаики. Линейная (для непрерывных значений) и логистическая (для бинарных признаков) регрессия. Регрессионный анализ в среде R. Оценка нелинейной регрессии в программе Curve Expert.	6	2	1			1	4	2		2
Оценка иерархии и классификация объектов. Кластерный анализ Значение отбора переменных и структуры выборки на построение классификации. Неоднозначность результатов кластерного анализа. Меры сходства: коэффициент корреляции Пирсона, меры расстояния (евклидово расстояние, квадратичное евклидово расстояние, Манхетенское расстояние), коэффициенты ассоциативности. Иерархические агломеративные методы группировки: метод одиночной связи, метод полных связей, метод средней связи, метод Уорда. Итеративные методы группировки: итерация по принципу k-средних.	3	1	0,5			0,5	2	1		1

Дендрограммы и их интерпретация.										
Общие проблемы анализа данных Использование среднего значения и медианы. Округление данных. Регулярность и пропорциональность наблюдений. Размер выборки, соотношение внутри- и межиндивидуальной изменчивости. Использование процентных соотношений. Разрешающая способность метода: выбор методики наблюдений и соответствующего метода статистической обработки. Проблема множественности сравнений, поправка Бонферрони. Проблема псевдорепликации: некорректная работа с повторными наблюдениями и при объединении данных. Методы обработки повторных наблюдений. Общие принципы моделирования случайных процессов, метод Монте-Карло.	7	2	1			1	4	3		3
Круговая статистика Особенности циклической шкалы данных. Операции с векторами. Варианты круговых диаграмм. Статистическая обработка угловых и временных данных в программе Ogiانا. Проверка выборки на равномерность распределения: <i>критерий Рэйли</i> , <i>критерий Rao</i> . Унимодальное распределение ван Мисес как аналог нормального распределения. Оценка отклонения среднего вектора от ожидаемого направления. Сравнение двух выборок: <i>критерий Хи-квадрат</i> для данных интервальной шкалы, параметрический <i>F</i> -	3	1	0,5			0,5	2	1		1

<i>критерий Уотсона-Уильямса для сравнения средних векторов, непараметрический серийный критерий Мардиа-Уотсона-Уиллера, непараметрический критерий U^2 Уотсона для сравнения дисперсий.</i>										
Решение предзачётных упражнений	3		1			1	2	1		1
Зачёт Решение зачётных упражнений	2					2	2			
Промежуточная аттестация - зачет										
Итого:	72						48			24

10. Учебно-методические материалы для самостоятельной работы аспирантов.

Конспекты лекций, аудио- и видеозаписи лекций, файлы презентаций лекций, основная и дополнительная учебная литература (см. п.11)

11. Ресурсное обеспечение:

Основная литература

Работа с базами данных

Бекаревич Ю.Б., Пушкина И.В., 2007. Самоучитель Access 2007. СПб.: БХВ-Петербург. 720с.

Кошелев В.Е., 2008. Access 2007. М.: Бином-пресс. 592с.

Статистический анализ

Книги:

Сидоренко Е.В., 2006. Методы математической обработки в психологии. СПб.: ООО "Речь". 350 с.

Реброва О.Ю., 2006. Статистический анализ медицинских данных. Применение пакета прикладных программ Statistica. М.:Медиасфера. 312 с.

Lehner P.N., 1996. Handbook of Ethological Methods. Cambridge University Press. 672p.

Zuur A. F., Ieno E. N., Smith G. M., 2007. Analysing Ecological Data. Springer Science + Business Media, LLC. 648 p.

- Zuur A.F., Ieno E.N., Meesters E.H.W.G., 2009. A Beginner's Guide to R. Springer Science + Business Media, LLC. 218 p.
- Огурцов С.В., 2015. Методическое пособие по обработке результатов и оформлению курсовых и дипломных работ, выпускных квалификационных работ бакалавров и магистерских диссертаций. М.:Биологический ф-т МГУ. 30 с.
- Ким Дж.-О., Мьюллер Ч.У., Клекка УР., Олдендерфер М.С., Блэшфилд Р.К., 1989. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика. 215 с.

Статьи:

- Fukuda H., Ohashi Y., 1997. A Guideline for Reporting Results of Statistical Analysis in Japanese Journal of Clinical Oncology // Japanese Journal of Clinical Oncology, V.27, №3, p.121–127
- Ланг Т., 2005. Двадцать ошибок статистического анализа, которые Вы сами можете обнаружить в биомедицинских статьях // Международный журнал медицинской практики, 2005, №1, с. 21-31
- Hurlbert S.H., 1984. Pseudoreplication and the design of ecological field experiments // Ecological Monographs, V.5, №2, p.187-211
- Козлов М. В., 2003. Мнимые повторности (pseudoreplications) в экологических исследованиях: проблема, не замеченная российскими учеными // Журнал общей биологии, Т.64, № 4, с.292-307
- Hargrove W.W., Pickering J., 1992. Pseudoreplication: a sine qua non for regional ecology // Landscape Ecology, V. 6, №4, p. 251-258

Дополнительная литература

Работа с базами данных

- Харитонов И.А., Рудикова Д.И., 2008. Microsoft Office Access 2007. СПб.: БХВ-Петербург. 1280с.

Статистический анализ

Книги:

- Малета Ю.С., Тарасов В.В., 1982. Непараметрические методы статистического анализа в биологии и медицине. М.: Изд-во Московского университета. 178с.
- Урбах В.Ю., 1964. Биометрические методы. М.: Наука. 415с.

Статьи:

- Орлов А.И., 2000. Прикладная статистика XXI в. // Экономика XXI века, 2000, №9, с.3-27
- Леонов В.П., Ижевский П.В., 1997. Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям // Бюллетень ВАК, 1997, №5, с.56-61
- Леонов В.П., Ижевский П.В., 1998. Применение статистики в статьях и диссертациях по медицине и биологии. Часть I. Описание методов статистического анализа в статьях и диссертациях // Международный Журнал Медицинской Практики, 1998, №4, с.7-12
- Леонов В.П., 1999. Применение статистики в статьях и диссертациях по медицине и биологии. Часть II. История биометрии и её применения в России // Международный Журнал Медицинской Практики, 1999, №4, с.7-19
- Леонов В.П., 2002. Применение статистики в статьях и диссертациях по медицине и биологии. Часть IV. Наукометрия статистической парадигмы экспериментальной биомедицины // Международный Журнал Медицинской Практики, 2002, №3, с.6-10

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Сайт поддержки пользователей Microsoft Office Access

<http://office.microsoft.com/ru-ru/access>

Сайт по базам данных и SQL

<http://www.sql.ru>

Сайт библиотеки знаний для разработчиков, использующих продукты и технологии Microsoft

<http://msdn.microsoft.com/ru-ru/library>

Сайт компании StatSoft, разработчика программы Statistica 6.0:

<http://www.statsoft.ru/>

Электронный журнал Биометрика:

<http://www.biometrica.tomsk.ru/index.htm>

Сайт программы CurveExpert (разработчик Daniel G. Hyams)

<http://curveexpert.webhop.biz/>

Сайт программы Oriana (Kovach Computing Services)

<http://www.kovcomp.co.uk/oriana/index.html>

Сайт программы Oriana (Kovach Computing Services)

<http://www.kovcomp.co.uk/oriana/index.html>

Сайт разработчиков программной среды R

<https://www.r-project.org/>

Перечень используемых информационных технологий, используемых при осуществлении образовательного процесса, включая программное обеспечение, информационные справочные системы (при необходимости):

СУБД Microsoft Access, Microsoft Excel, Statistica, Oriana, Curve Expert, программная среда R.

Описание материально-технической базы.

Кафедра зоологии позвоночных биологического факультета МГУ располагает необходимым аудиторным фондом, компьютерами, проекторами и экранами.

12. Язык преподавания: русский

13. Преподаватель (преподаватели):

доцент кафедры зоологии позвоночных С.В.Огурцов

младший научный сотрудник кафедры зоологии позвоночных Э.Н. Рахимбердиев



**Оценочные средства для промежуточной аттестации по дисциплине (модулю) «АНАЛИЗ ДАННЫХ И КОМПЬЮТЕРНЫЕ МЕТОДЫ В БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ»
на основе карт компетенций выпускников**

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модулю)	КРИТЕРИИ и ПОКАЗАТЕЛИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ по дисциплине (модулю), баллы БРС					ОЦЕНОЧНЫЕ СРЕДСТВА
	1, 0	2 1-29	3 30-59	4 60-89	5 90-100	
Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях Код В1 (УК-1)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, зачет
Владеть: навыками критического анализа и оценки современных научных достижений и результатов деятельности по решению исследовательских и практических задач, в том числе в междисциплинарных областях Код В2 (УК-1)	0	1-29	30-59	60-89	90-100	- - индивидуальное собеседование, реферат, зачет
Знать:	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат,

методы научно-исследовательской деятельности Код 31(УК-2)						<i>зачет</i>
Владеть: технологиями оценки результатов коллективной деятельности по решению научных и научно-образовательных задач, в том числе ведущейся на иностранном языке Код В2(УК-3)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>
Знать: стилистические особенности представления результатов научной деятельности в устной и письменной форме на государственном и иностранном языках Код 32(УК-4)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>
Владеть: навыками анализа научных текстов на государственном и иностранном языках Код В1(УК-4)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>
Уметь: собирать, отбирать и использовать необходимые данные и эффективно применять количественные методы их анализа	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>

Фонды оценочных средств, необходимые для оценки результатов обучения

Примеры вопросов к промежуточному контролю (темы рефератов, вопросы для индивидуального собеседования):

1. Практическое задание по работе с базой данных. Создание сложного запроса на конкретном примере.
2. Текстовый тест-вопрос на критический анализ методики, обработки данных и интерпертации результата на конкретном примере научной работы.
3. Задача на «простые» методы статистического анализа (частотные критерии, парное сравнение, корреляция) на конкретном численном примере.
4. Задача на многофакторные методы статистического анализа (дисперсионный, факторный, регрессионный анализ) на конкретном численном примере.

ПРОГРАММА

зачета по спецкурсу «АНАЛИЗ ДАННЫХ И КОМПЬЮТЕРНЫЕ МЕТОДЫ В БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ»

Структура зачётного билета:

1. Практическое задание по работе с базой данных. Создание сложного запроса на конкретном примере.
2. Текстовый тест-вопрос на критический анализ методики, обработки данных и интерпертации результата на конкретном примере научной работы.
3. Задача на «простые» методы статистического анализа (частотные критерии, парное сравнение, корреляция) на конкретном численном примере.
4. Задача на многофакторные методы статистического анализа (дисперсионный, факторный, регрессионный анализ) на конкретном численном примере.