

Федеральное бюджетное учреждение науки
«ГОСУДАРСТВЕННЫЙ НАУЧНЫЙ ЦЕНТР ВИРУСОЛОГИИ И
БИОТЕХНОЛОГИИ «ВЕКТОР»
Федеральной службы по надзору в сфере защиты прав потребителей и
благополучия человека
(ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора)



УТВЕРЖДАЮ
Генеральный директор
ФБУН ГНЦ ВБ «Вектор»
Роспотребнадзора
Р.А. Максютов

«08» сентября 2017 г.

**Рабочая программа дисциплины
Статистика и анализ данных**

Направление подготовки:
06.06.01 – Биологические науки

Направленность (профиль)
03.02.02 – Вирусология

Квалификация выпускника
Исследователь. Преподаватель-исследователь

Форма обучения: очная

Кольцово 2017

СОДЕРЖАНИЕ

1.	Цели освоения дисциплины.....	3
2.	Место дисциплины в структуре образовательной программы.....	3
3.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.....	3
4.	Объем дисциплины и виды учебной работы.....	3
5.	Содержание дисциплины и трудоемкость по видам учебных занятий.....	4
6.	Самостоятельная работа обучающихся.....	5
7.	Формы проведения занятий.....	6
8.	Фонд оценочных средств.....	7
	8.1. Паспорт фонда оценочных средств	7
	8.2. Аттестация (зачет).....	8
9.	Учебно-методическое и информационное обеспечение дисциплины.....	8
	9.1. Основная литература.....	8
10.	Материально-техническое обеспечение дисциплины.....	9

1. Цель освоения дисциплины

Цель дисциплины – обучить основным принципам математической статистики для анализа данных биологических исследований, научить правильно выбирать подходящий метод анализа в зависимости от имеющихся данных и использовать его.

2. Место дисциплины в структуре образовательной программы

Дисциплина отнесена к вариативной части Блока 1 «Дисциплины (модули)» Б1.В.ДВ.1 программы подготовки научно-педагогических кадров в аспирантуре по направлению подготовки 06.06.01 Биологические науки.

Для освоения данной дисциплины обучаемый должен:

знать: основы комбинаторики, логики, арифметики в рамках школьного курса математики; генетику; основные методы исследования в молекулярной биологии, микробиологии и вирусологии;

уметь: решать математические задачи, требующие вычислений с помощью калькулятора, проверять ответ; читать и понимать научные статьи и документацию к программам, в том числе на английском языке;

владеть: основами компьютерной грамотности, основами работы в программе Microsoft Excel; критическим мышлением; научным методом.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы аспирантуры обучающийся должен овладеть следующими компетенциями:

- универсальные компетенции: УК-1, УК-2;
- общепрофессиональные компетенции: ОПК-1.

4. Объем дисциплины и виды учебной работы

Объем дисциплины – 3 зачетных единицы (ЗЕ) или 108 академических часов.

Вид учебной работы	Всего часов
Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	48
Аудиторные занятия (всего)	48
в том числе:	
лекции (Л)	24
практические занятия (ПЗ), семинары (С)	24
лабораторные работы (ЛР)	
практикумы (ПР)	
Внеаудиторная работа (всего)	
в том числе:	
индивидуальная работа обучающихся с преподавателем	

Вид учебной работы		Всего часов
Консультации		
Самостоятельная работа обучающихся (СР) (всего)		24
в том числе: реферат		
Вид промежуточной аттестации зачет (З), экзамен (Э)		зачет
Общая трудоемкость	Часов	72
	зачётных единиц	2

5. Содержание дисциплины и трудоемкость по видам учебных занятий

Содержание раздела		
№ п/п	Наименование темы	Объем, час
1	2	3
Лекции		
1.	История статистики, в том числе в биологии. Важность статистики для научной работы. Что такое описательная статистика и сравнение гипотез. Случайные события, свойства вероятности, независимые и зависимые события. Основные понятия статистики.	2
2.	Распределение вероятностей. Биномиальное распределение. Нормальное распределение, его характеристики. Центральная предельная теорема.	2
3.	Программы для статистической обработки данных. Microsoft Excel. Программа Statistica. Среда R.	2
4.	Описательная статистика. Типы данных. Варианты представления данных различных типов. Квартет Энского. Представление многомерных данных. Корреляции. Снижение размерности данных.	2
5.	Распределения, отличающиеся от нормальных. Закон Бенфорда. Логнормальное распределение, распределение Пуассона. Выбросы. Преобразование данных.	2
6.	Основные понятия теории проверки гипотез. Определение гипотезы. Определение критерия. Ошибки первого и второго рода. Критерий Стьюдента.	2
7.	Статистические критерии для различных типов данных. Выбор статистических критериев для решения конкретных задач.	2
8.	Анализ выживаемости. Методы работы с цензурированными данными.	2
9.	Множественные сравнения, «кошмар Бонферрони». Проблемы с GWAS исследованиями. Рандомизация и бутстреп.	2
10.	Анализ многомерных данных. Метод главных компонент.	2

11.	Регрессия. Анализ зависимости доза-эффект.	2
12.	Машинное обучение. Подготовка данных для машинного обучения. Методология использования программ для машинного обучения. Оценка качества. Характеристическая кривая и площадь под ней.	2
	Итого	24
Практические занятия и семинары		
1.	Решение задач по теории вероятностей.	2
2.	Задачи на вычисление параметров нормального распределения, проверку нормальности, точный тест Фишера	2
3.	Практикум по основам работы в R. Основы работы в командной строке.	2
4.	Практикум по отображению данных с помощью Excel и R.	2
5.	Практикум по преобразованию данных, проверке нормальности, поиску выбросов с помощью R.	2
6.	Решение задач по проверке гипотез с помощью Excel и R.	2
7.	Решение задач по проверке гипотез с помощью R.	2
8.	Решение задач по анализу выживаемости с помощью R.	2
9.	Практикум по анализу многомерных данных с помощью R.	2
10.	Практикум по анализу многомерных данных с помощью R.	2
11.	Решение задач по регрессионному анализу с помощью Excel и R.	2
12.	Практикум по машинному обучению в среде R, обзор других инструментов для машинного обучения.	2
	Итого:	24
	Всего:	48

6. Самостоятельная работа обучающихся

Аспиранты могут выполнять необходимую при изучении дисциплины самостоятельную работу в читальном зале библиотеки ГНЦ ВБ «Вектор» Роспотребнадзора, в учебных кабинетах, на рабочих местах и на дополнительно оборудованных стационарных местах с выходом в Интернет, а также в домашних условиях. При составлении задач рекомендуется использовать реальные данные, полученные в своей работе или коллегами.

№ п/п	Наименование вида самостоятельной работы	Трудоемкость
		часы
1	2	3
1.	Поиск и изучение информации о когнитивных искажениях, связанных со статистикой и анализом данных.	2
2.	Самостоятельно составить и решить задачу с использованием точного теста Фишера.	2
3.	Повторить самостоятельно действия в среде R, которые изучались на семинаре.	2
4.	Самостоятельно изучать возможности отображения данных в среде R.	2
5.	Найти примеры данных, распределенных по логнормальному закону.	2
6.	Самостоятельно составить и решить задачи по проверке гипотез.	2
7.	Самостоятельно составить и решить задачи по проверке гипотез.	2
8.	Самостоятельно составить и решить задачи по анализу выживаемости.	2
9.	Повторить самостоятельно действия в среде R, которые изучались на семинаре.	2
10.	Повторить самостоятельно действия в среде R, которые изучались на семинаре.	2
11.	Самостоятельно составить и решить задачи по регрессионному анализу.	2
12.	Найти в научных публикациях примеры использования машинного обучения в биологии.	2
	ИТОГО:	24

Самостоятельная работа аспирантов проходит с использованием компьютеров в конференц-зале в заранее выделенное время, также возможно использование компьютеров в лабораториях, где аспиранты проходят практику. Самостоятельная работа выполняется аспирантами по заданию преподавателя.

7. Формы проведения занятий

Занятия проводятся в интерактивной форме с использованием наглядных пособий, мультимедийного оборудования (компьютер, проектор) и маркерной доски для записей. Для мониторинга качества усвоения материала используются опросы студентов, в том числе по самостоятельной работе. На семинарах аспиранты работают за компьютерами индивидуально или в парах. На занятиях используются технологии проблемного обучения, постановки задач формулируются с учетом тем квалификационных работ аспирантов. Решение задач на семинарах происходит как индивидуально, так и методом мозгового штурма.

8. Фонд оценочных средств

8.1. Паспорт фонда оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части)	Наименование оценочного средства
1	2	3	4
1	Основные понятия теории вероятностей. Основные понятия статистики.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
2	Формулировка центральной предельной теоремы. Распределение случайной величины. Параметры распределения. Нормальное и биномиальное распределение.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
3	Основные функции программы Microsoft Excel и среды R.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
4	Описательная статистика, виды диаграмм.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
5	Логнормальное распределение, распределение Пуассона. Выбросы. Виды преобразования данных.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
6	Основные понятия теории проверки гипотез. Определение гипотезы. Определение критерия. Ошибки первого и второго рода. Применимость критерия Стьюдента.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
7	Факторы, влияющие на выбор статистического критерия. Параметрические и непараметрические критерии. Мощность критерия. Состоятельность критерия.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
8	Основные понятия анализа выживаемости. Методы работы с цензурированными данными.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
9	Проблема множественных сравнения и пути решения этой проблемы.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
10	Методы анализа многомерных данных.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
11	Основные понятия регрессионного анализа.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос
12	Принципы машинного обучения. Методы машинного обучения, их ограничения. Возможные ошибки при использовании машинного обучения.	УК-1, УК-2, ОПК-1	Собеседование, устный опрос

8.2. Аттестация (зачет)

Зачет проходит в форме выполнения практического задания в среде R либо в программе Microsoft Excel, по выбору аспиранта. Во время зачета разрешено пользоваться конспектами и информацией в сети интернет. К зачету допускаются аспиранты, посетившие не менее 80% занятий. При пропуске большего количества занятий аспирант выполняет письменные работы по пропущенным темам и сдает их на проверку преподавателю.

При приеме зачета экзаменатор может задавать дополнительные вопросы по программе курса, чтобы убедиться, что аспирант правильно понимает свои действия.

Варианты практических заданий:

- Найти выбросы в наборе данных и обосновать свое решение.
- Определить среднюю эффективную дозу (ED50) по заданным данным.
- Сравнить две заданные выборки, самостоятельно выбрать статистический критерий.
- Сравнить несколько заданных выборок, самостоятельно выбрать статистический критерий.
- Наглядно представить заданный набор данных с использованием методов описательной статистики.

Оценка «зачтено» выставляется обучающемуся, показавшему систематизированные знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач.

Оценка «не зачтено» выставляется обучающемуся, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

9. Учебно-методическое и информационное обеспечение дисциплины

Аспирантам ГНЦ ВБ «Вектор» Роспотребнадзора обеспечен полный доступ к обслуживанию в библиотеке ГНЦ ВБ «Вектор» Роспотребнадзора, в т.ч. библиотечное обслуживание, обслуживание по межбиблиотечному абонементу, справочно-библиографическое и информационное обслуживание.

Учебно-методическое и информационное обеспечение дисциплины включает в себя основную и дополнительную учебную и научную литературу, ресурсы информационно-телекоммуникационной сети Интернет, конспекты лекций.

9.1. Основная литература:

- 1 А. Б. Шипунов, Е. М. Балдин, П. А. Волкова, А. И. Коробейников, С. А. Назарова, С. В. Петров, В. Г. Суфиянов. Наглядная статистика. Используем R!
- 2 В.К. Шитиков, Г.С. Розенберг. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R.

10. Материально-техническое обеспечение дисциплины




- Конференц-зал корпуса №12а ГНЦ ВБ «Вектор» Роспотребнадзора, оснащенный презентационной техникой;
- средства мультимедиа: проектор, экран, компьютер/ноутбук;
- доска учебная маркерная;
- рабочее место аспиранта с выходом в Интернет.

Составители:

д.б.н., зав. отделом

к.б.н., с.н.с.


к.б.н., с.н.с.

 С.И. Бажан
 Д.В. Антонев
 А.Ю. Бакулина

Согласовано:

Зав. аспирантурой ФБУН ГНЦ ВБ «Вектор»

Роспотребнадзора

 Т.А. Косогова

Рабочая программа рассмотрена и одобрена на заседании ученого совета ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора.

Протокол от «07» сентября 2017 г. № 8.