

СОСТАВИТЕЛЬ:

Анна Анатольевна Жукова, доцент кафедры общей экологии и методики преподавания биологии Белорусского государственного университета, кандидат биологических наук

РЕЦЕНЗЕНТЫ:

Екатерина Эдуардовна Хейдорова, Ученый секретарь Государственного научно-производственного объединения «Научно-практический центр Национальной академии наук Беларуси по биоресурсам», кандидат биологических наук;

Сергей Владимирович Буга, заведующий кафедрой зоологии Белорусского государственного университета, доктор биологических наук, профессор

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой общей экологии и методики преподавания биологии Белорусского государственного университета (протокол № 17 от 20 марта 2015 г.);

Учебно-методической комиссией биологического факультета Белорусского государственного университета протокол № 8 от 25 марта 2015 г.)

Ответственный за редакцию: Анна Анатольевна Жукова

Ответственный за выпуск: Анна Анатольевна Жукова

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Современная биология давно перестала быть исключительно описательной наукой. Сегодня ее существование и развитие невозможно без использования методов и подходов такой области математики как статистика. Статистика позволяет компактно описать данные, понять их структуру, провести классификацию, увидеть закономерности в хаосе случайных явлений. Игнорирование и недооценка статистической обработки и математического анализа полученного исследователем материала может свести на нет результаты многих важных опытов, привести к необоснованным и даже ошибочным выводам. Умелое применение статистических методов позволяет объективно оценивать результаты массовых наблюдений, выявлять скрытые закономерности, правильно трактовать их, что в конечном итоге делает биологию точной наукой. В связи этим курс лекций «Биометрия» является обязательным при подготовке специалистов биологического профиля.

Цель курса – сформировать у студентов целостную систему знаний о современных подходах статистического анализа данных. В **задачи** дисциплины входит освоение методов, позволяющих выявлять количественные закономерности в биологических явлениях; ознакомление с принципами построения математических моделей биологических явлений и процессов; формирование навыков и умений компьютерной обработки экспериментальных данных; ознакомление с правилами корректного представления результатов исследований; формирование способности к критическому анализу представляемых в публикациях данных.

В результате изучения дисциплины обучаемый должен:

знать:

- классификацию основных методов статистического анализа биологических данных;
- способы описания центральной тенденции и разброса в совокупностях, подчиняющихся различным законам распределения;
- условия применения параметрических и непараметрических методов анализа данных;

уметь:

- распознавать разные типы биологических данных;
- строить графические изображения вариационных рядов;
- описывать наиболее выраженные свойства анализируемой совокупности по графическому изображению вариационного ряда;
- рассчитывать основные показатели описательной статистики при помощи компьютера;
- выполнять сравнение двух и более выборок при помощи компьютера;
- выполнять анализ частот при помощи компьютера;
- выполнять корреляционный и регрессионный анализы при помощи компьютера.

владеть:

- основными методами сравнения двух и более совокупностей;
- методами анализа частот;

– методами выявления связи между биологическими признаками и ограничениями по их применению.

Программа курса составлена с учетом межпредметных связей и программ по смежным учебным дисциплинам («Информационные технологии в биологических исследованиях», «Основы информационной биологии», «Популяционная экология», «Биоиндикация качества природной среды», «Общая экология» и др.).

Программа рассчитана на 88 часов, в том числе 52 часов аудиторных: 24 – лекционных и 28 – лабораторных занятий.

ПРИМЕРНЫЙ ТЕМАТИЧЕСКИЙ ПЛАН

№ п/п	Наименование разделов, тем	Количество часов				
		Аудиторные				Самост. работа
		Лек ции	Практич., семинар.	Лаб. занятия	УСР	
1.	Введение	2	–	–	–	–
2.	Данные в биологии	2	–	2	–	4
3.	Элементы теории планирования исследований	2	–	2	–	4
4.	Описательная статистика	4	–	4	–	4
5.	Статистическая гипотеза	2	–	–	–	4
6.	Основы дисперсионного анализа	2	–	8	–	6
7.	Анализ частот	2	–	–	–	4
8.	Корреляционный анализ	2	–	4	–	4
9.	Регрессионный анализ	4	–	4	–	4
10.	Элементы многомерной статистики	2	–	4	–	2
	ИТОГО:	24	–	28	–	36

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

I. ВВЕДЕНИЕ

Биометрия как наука. Значение биометрии в исследовательской работе и профессиональной подготовке специалистов-биологов. Роль работ У. Петти, Дж. Гранта, П.-С. де Лапласа, П. Пуассона, П. Л. Чебышева, А. Кетле, К. Ф. Гаусса, Ф. Гальтона, К. Пирсона, У. Госсета, Р. Фишера и других ученых в развитии биометрии.

Краткая история развития биостатистического направления и использования математических методов в биологии в СССР и Беларуси. Современная практика использования биометрии в фундаментальной и прикладной биологии.

II. ДАННЫЕ В БИОЛОГИИ

Понятие о наименьшей выборочной единице (единице наблюдения) и данных в биологии. Переменные (признаки). Генеральная совокупность и выборка. Количественные переменные: дискретные и непрерывные. Качественные переменные. Ранговая шкала измерений. Производные переменные: пропорции, индексы, интенсивности протекания процессов.

III. ЭЛЕМЕНТЫ ТЕОРИИ ПЛАНИРОВАНИЯ ИССЛЕДОВАНИЙ

Сплошное и выборочное обследование совокупностей. Важность случайного (рандомизированного) отбора единиц наблюдения при формировании выборок. Понятие о репрезентативной и смещенной выборках. Полностью случайный отбор и его реализация при помощи таблиц случайных чисел. Стратифицированный отбор. Систематический отбор.

IV. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Группировка данных в вариационный ряд. Способы графического изображения вариационного ряда: полигон (кривая) распределения, гистограмма. Теоретические распределения случайных величин и их свойства: биномиальное распределение, распределение Пуассона, нормальное распределение. Коэффициенты асимметрии и эксцесса.

Средние величины: средняя арифметическая, взвешенная средняя, геометрическая средняя. Меры разброса единиц совокупности: дисперсия и стандартное отклонение. Коэффициент вариации.

Мода. Медиана и процентиля. 25-й и 75-й процентиля (квартили).

Расчет параметров описательной статистики при качественной изменчивости.

Оценка репрезентативности выборочных показателей при помощи стандартной ошибки. Центральная предельная теорема. Закон больших чисел. Определение достаточного объема выборки. Доверительные интервалы для средней арифметической и для доли.

Выбросы и их устранение. Пропуски и их заполнение.

Способы представления средних величин, мер разброса, стандартных ошибок и доверительных интервалов в научных публикациях.

V. СТАТИСТИЧЕСКАЯ ГИПОТЕЗА

Понятие о статистической гипотезе. Нулевая и альтернативная гипотезы. Статистические критерии (тесты). Вероятность справедливости нулевой гипотезы (уровень значимости). Статистические ошибки I и II типа. Мощность критерия (теста). Понятие о параметрических и непараметрических критериях (тестах). Способы трансформации данных для приведения их к нормальному распределению: логарифмирование, извлечение квадратного корня, преобразование Бокса-Кокса, угловое преобразование.

VI. ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА

Назначение дисперсионного анализа (ANOVA). Нулевая гипотеза при дисперсионном анализе. Расчет внутри- и межгрупповой дисперсий при однофакторном анализе с равномерным дисперсионным комплексом. F -критерий Фишера. Определение внутри- и межгруппового числа степеней свободы. Однофакторный дисперсионный анализ повторных измерений. Понятие о многофакторном дисперсионном анализе.

Допущения дисперсионного анализа. Проверка нормальности распределения данных: визуальный анализ гистограммы распределения, использование нормальной вероятностной бумаги, тесты Колмогорова-Смирнова и Шапиро-Уилка. Проверка равенства групповых дисперсий: тесты Бартлетта, Левене, Кохрана, F -тест Хартли.

Эффект множественных сравнений. Апостериорный (*post-hoc*) анализ и его методы: тесты Тьюки, Ньюмена-Кейлса, Шеффе, Даннета.

Непараметрические аналоги однофакторного дисперсионного анализа: H -тест Крускала-Уоллиса и тест Фридмана.

Сравнение двух групп. Тест Стьюдента как частный случай дисперсионного анализа. t -распределение. Тест Стьюдента для парных измерений. Использование доверительных интервалов для проверки гипотезы о равенстве двух средних. Введение поправки Бонферрони для t -критерия при проведении множественных сравнений средних. Непараметрические аналоги критерия Стьюдента: U -тест Манна-Уитни, тест Уилкоксона, тест Уэлча.

VII. АНАЛИЗ ЧАСТОТ

z -критерий для сравнения двух выборочных долей и условие его применимости. Анализ таблиц сопряженности при помощи χ^2 -критерия. Поправка Йетса на непрерывность. Использование критерия χ^2 для определения нормальности распределения данных. Определение числа степеней свободы при анализе таблиц сопряженности. Точный критерий Фишера. Одностороннее и двустороннее значения точного критерия Фишера.

VIII. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Понятие о функциональной и корреляционной зависимостях. Степень и направление корреляционной зависимости. Коэффициент корреляции Пирсона и оценка его статистической значимости.

Непараметрические процедуры корреляционного анализа, области применения. Коэффициент ранговой корреляции Спирмена.

IX. РЕГРЕССИОННЫЙ АНАЛИЗ

Назначение регрессионного анализа. Общий вид регрессионного уравнения. Связь коэффициента регрессии с коэффициентом корреляции. Оценка параметров регрессионного уравнения по выборке с помощью метода наименьших квадратов. Статистическая значимость регрессии. Проверка нулевой гипотезы о равенстве коэффициента регрессии нулю. Стандартные ошибки параметров регрессионного уравнения. Коэффициент детерминации.

Анализ остатков. Оценка величины остаточной дисперсии с помощью F -критерия. Нахождение доверительной области для линии регрессии. Понятие о нелинейной и множественной регрессионной зависимости.

Х. ЭЛЕМЕНТЫ МНОГОМЕРНОЙ СТАТИСТИКИ

Понятие о многомерной совокупности и многомерном пространстве. Принцип «сворачивания» информации, заключенной в многомерных совокупностях.

Кластерный анализ и области его применения. Правила объединения объектов в кластеры. Графическое изображение результатов кластерного анализа. Выбор процедур кластеризации исходя из характера используемых данных. Биологическая интерпретация визуализируемых кластеров.

Дискриминантный анализ и области его применения. Дискриминантное уравнение и его параметры.

Анализ главных компонент и области его применения. Принцип ортогональности главных компонент. Кумулятивная объясненная дисперсия.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная и дополнительная литература

Основная:

1. *Боровиков В.П., Боровиков И.П.* STATISTICA: Статистический анализ и обработка данных в среде Windows. – М.: Информационно-издательский дом «Филинь», 1998. – 608 с.
2. *Вуколов Э.А.* Основы статистического анализа (практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL). – 2-е изд. – Москва, 2008. – 464 с.
3. *Гланц С.* Медико-биологическая статистика. – М., Практика, 1999. – 459 с.
4. *Ивантер Э.В., Коросов А.В.* Элементарная биометрия: учеб. пособие. – 2-е изд., Петрозаводск: Изд-во ПетрГУ, 2010. – 104 с.
5. *Ивантер Э.В., Коросов А.В.* Введение в количественную биологию: учеб. пособие. – Петрозаводск: Изд-во ПетрГУ, 2011. – 302 с.
6. *Лакин Г.Ф.* Биометрия: учеб. пособие для биологич. спец. вузов. – 4-е изд., перераб. доп. – М.: Высш. шк., 1990. – 352 с.
7. *Петри А., Сэбин К.* Наглядная медицинская статистика. – М.: Геотар-Мед, 2003. – 143 с.
8. *Рокицкий П.Ф.* Биологическая статистика. – Минск: Вышэйш. шк., 1973. – 320 с.

Дополнительная:

9. *Афифи А., Эйзен С.* Статистический анализ: Подход с использованием ЭВМ. Под редакцией Башарина Г.П. – М.: Мир, 1982. – 488 с.
10. *Берк К.Н., Кэйри П.* Анализ данных с помощью Microsoft Excel. – Москва-Санкт-Петербург-Киев, 2005. – 560 с.
11. *Бейли Н.* Математика в биологии и медицине. – М.: Мир, 1970. – 325 с.
12. *Боровиков В.* STATISTICA. Искусство анализа данных на компьютере. – Питер, 2003. – 688 с.
13. *Бурева Н.Н.* Многомерный статистический анализ с использованием ППП «STATISTICA». – Нижний Новгород, 2007. – 112 с.
14. *Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др.* Факторный, дискриминантный и кластерный анализ / Под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
15. *Куприенко Н.В., Пономарева О.А., Тихонов Д.В.* Статистика. Методы анализа распределений. Выборочное наблюдение. – СПб.: Изд-во Политехн. ун-та, 2009. – 138 с.
16. *Любищев А.А.* Дисперсионный анализ в биологии. – М.: Изд-во Моск. ун-та, 1986. – 200 с.
17. *Мэгарран Э.* Экологическое разнообразие и его измерение. – Москва, 1992. – 89 с.
18. *Плохинский Н. А.* Биометрия. – 2-е изд. – М.: МГУ, 1970. – 367 с.
19. *Ростова Н.С.* Корреляции: структура и изменчивость. – СПб.: Изд-во С.-Петербур. ун-та, 2002. – 308 с.
20. *Рунион Р.* Справочник по непараметрической статистике: современный подход. – М.: Финансы и статистика, 1982. – 198 с.
21. *Халафян А.А.* STATISTICA 6.0. Статистический анализ данных. – М., 2007. – 512 с.
22. *Чайковский Ю.В.* О природе случайности. – 2-е изд., испр. и доп. – Вып. 27. «Ценологические исследования». – М.: Центр системных исследований, 2004. – 280 с.
23. *Sokal R.R., Rohlf F.J.* Introduction to Biostatistics. 2nd ed. – Dover edition, 2009. – 366 p.
24. *Zar J. H.* Biostatistical analysis. 5th edition. Prentice Hall, New York, 2009. – 960 p.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

В курсе подробно рассматриваются параметрические процедуры анализа данных. Наряду с этим большое внимание уделяется непараметрическим методам, использование которых в практике биологических исследований постоянно возрастает. На примере кластерного и дискриминантного анализов, а также метода главных компонент слушатели знакомятся с элементами многомерной статистики. Теоретические положения лекционного курса развиваются и закрепляются на лабораторных занятиях, при выполнении которых студенты приобретают навыки и умения статистической обработки данных при помощи персонального компьютера. Студенты изучают возможности использования программы STATISTICA для расчета параметров описательной статистики, построения кривых распределения и гистограмм, выполнения дисперсионного анализа и сравнения двух групп, расчета коэффициентов корреляции, анализа частот, выполнения регрессионного анализа.

Организация самостоятельной работы студентов по курсу предполагает размещение в сетевом доступе комплекса учебных и учебно-методических материалов (программа, курс лекций, мультимедийные презентации, список рекомендуемой литературы и информационных ресурсов, темы лабораторных занятий и методические и информационные материалы к ним и др.).

Эффективность самостоятельной работы студентов целесообразно проверять в ходе текущего и итогового контроля знаний в форме устного и письменного опросов, тестового компьютерного контроля по темам и разделам курса. Для общей оценки качества усвоения студентами учебного материала рекомендуется использование рейтинговой системы.

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ

Учебным планом специальностей 1-31 01 01 Биология (по направлениям); 1-33 01 01 Биоэкология; 1-31 01 02 Биохимия; 1-31 01 03 Микробиология в качестве формы итогового контроля по учебной дисциплине рекомендован зачет. Для текущего контроля качества усвоения знаний студентами можно использовать следующий диагностический инструментарий:

- устные опросы;
- защита индивидуальных заданий при выполнении лабораторных работ;
- письменные контрольные работы по отдельным темам курса;
- компьютерное тестирование.