

## **Программа дисциплины «МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ»**

### **I. Название дисциплины - Математические методы в биологии.**

Данный курс посвящен статистическим и графическим методам анализа данных с использованием пакетов прикладных программ для персональных компьютеров и рассчитан для студентов второго курса биологического факультета МГУ.

Усвоению материала, излагаемого в данном разделе курса, способствует большое число примеров из различных областей биологии. Большое внимание в курсе уделяется, образно говоря, подводным рифам статистических методов анализа данных, с которыми часто сталкивается исследователь.

Отличительная особенность курса – широкое использование пакетов прикладных компьютерных программ при анализе экспериментальных данных. Такой подход позволяет опустить изложение утомительных вычислительных процедур и сосредоточить все внимание на главном – применении методов анализа данных и интерпретации результатов. Предполагается, что на практических занятиях при анализе экспериментальных данных в конкретных областях биологии студенты будут использовать персональные компьютеры. При этом студенты обучаются работе с очень популярным в настоящее время пакетом прикладных компьютерных программ анализа данных STATISTICA. Студенты также обучаются самостоятельному составлению компьютерных программ для анализа данных.

Конечная цель курса – освоение студентами теоретических подходов и практических методов анализа биологических данных, применяемых в современной мировой науке.

### **II. Шифр дисциплины:**

### **III. Цели и задачи дисциплины:**

А. Основная цель изучения дисциплины «Математические методы в биологии» - освоение студентами теоретических подходов и практических методов анализа биологических данных, применяемых в современной мировой науке и дать системное представление о статистических, дескриптивных и графических методах анализа данных, с которыми часто сталкивается исследователь-биолог.

Б. Задачи дисциплины:

- познакомить студентов с основными понятиями теории вероятностей и математической статистики.
- сформировать у студентов способность применять дескриптивные и графические методы анализа данных.
- сформировать у студентов способность применять современные методы анализа данных (статистическое оценивание и проверка гипотез, методы многомерного статистического анализа и исследования зависимостей) с использованием пакета прикладных компьютерных программ анализа данных STATISTICA.

### **IV. Место дисциплины в структуре ООП:**

А. Информация об образовательном стандарте и учебном плане:

- Тип образовательного стандарта и вид учебного плана: ИБ – интегрированный магистр МГУ, учебный план бакалавриата.
- Направление подготовки: Антропология, Ботаника, Зоология, Биохимия, Генетика, Микробиология, Физиология.
- Наименование учебного плана:

- Профиль подготовки:

Б. Информация о месте дисциплины в образовательном стандарте и учебном плане:

- Базовая часть.
- Блок дисциплин - математический и естественнонаучный цикл.
- Тип – обязательный.
- Курс – 2.
- Семестр – 3.

В. Перечень дисциплин, которые должны быть освоены до начала освоения данной дисциплины: курс «Высшей математики».

Г. Общая трудоемкость: 144 ак. часа и 4 зачетных единицы.

Д. Форма аттестации: экзамен.

#### V. Форма проведения:

Формы занятий:

- Лекции – 36 ак. часов.
- Практические занятия (семинары) – 36 ак. часов.
- Самостоятельная работа – 72 ак. часа.

Формы текущего контроля: домашние занятия, две контрольные и одно контрольное занятие на компьютере.

#### VI. Распределение трудоемкости по разделам и темам

№ п/п	Наименование разделов и тем дисциплины / Наименование разделов (этапов) практики	Трудоемкость (в ак. часах) по формам занятий (для дисциплин) и видам работ (для практик)				Формы контро ля
		Аудиторная работа (с разбивкой по формам занятий (для дисциплин) и видам работ (для практик)			Самостоятельн ая работа	
		Лекция	Практическ ие занятия (семинары)	Лабораторн ая работа		
1.	Вероятностные модели. Понятие случайного испытания. Пространство элементарных событий. Операции над событиями. Полная система событий.	2	2		4	
2.	Теорема Бернулли. Примеры введения вероятностей событий в биологических задачах. Вероятностное пространство. Модели вероятностного пространства.	2	2		4	
3.	Условная вероятность. Независимость событий, Формула полной вероятности. Формула Байеса.	2	2		4	
4.	Одномерная случайная величина. Функция распределения случайной величины и ее свойства. Дискретные случайные величины. Непрерывные случайные величины. Функция плотности вероятностей и ее свойства.	2	2		4	Контрольная работа на дом
5.	Нормально распределенная случайная величина. Стандартное нормальное распределение. Центральной предельной теоремы. Распределения случайных величин,	2	2		4	

	связанные с нормальным распределением.					
6.	Характеристики распределений. Математическое ожидание случайной величины и его свойства. Дисперсия случайной величины и ее свойства. Неравенство Чебышева.	2	2		4	
7.	Характеристики положения и разброса случайной величины. Примеры вычисления рассмотренных характеристик.	2	2		4	
8.	Многомерные случайные величины. Функция распределения многомерной случайной величины и ее свойства. Ковариация и коэффициент корреляции двух случайных величин. Свойства коэффициента корреляции. Двумерное нормальное распределение.	2	2		4	Контрольная работа в аудитории
9.	Генеральная совокупность. Простой случайный выбор. Случайная выборка. Эмпирическая функция распределения. Гистограмма выборки	2	2		4	
10.	Точечное статистическое оценивание. Выборочные оценки. Свойства оценок: несмещенность; состоятельность; Эффективность. Достаточное условие состоятельности. Выборочные характеристики. Методы получения выборочных оценок. Метод максимального правдоподобия.	2	2		4	
11.	Интервальное статистическое оценивание. Доверительный интервал. Доверительная вероятность. Центральные и нецентральные доверительные интервалы. Доверительный интервал для математического ожидания нормально распределенной случайной величины.	2	2		4	
12.	Доверительный интервал для дисперсии нормально распределенной случайной величины; параметра биномиального распределения; пуассоновского распределения; для коэффициента корреляции.	2	2		4	
13.	Статистическая проверка гипотез. Простые и сложные альтернативные гипотезы. Статистика критерия. Критическая область. Вероятности ошибок первого и второго родов. Уровень значимости критерия. Мощность критерия. Параметрические и непараметрические критерии проверки статистических гипотез	2	2		4	
14.	Одновыборочный t-критерий. Двухвыборочный t-критерий для независимых выборок и для связанных выборок. Одновыборочный хи-квадрат критерий. Двухвыборочный F-критерий. Сравнение параметров двух биномиальных распределений; двух пуассоновских распределений. Проверка гипотезы о равенстве заданному числу коэффициента корреляции	2	2		4	
15.	Критерии согласия: хи-квадрат и Колмогорова и Смирнова. Непараметрические критерии:	2	2		4	

	для проверки гипотезы об отсутствии сдвига; критерий Смирнова об однородности; критерий Спирмена для проверки гипотезы о независимости.					
16.	Классификация методов анализа многомерных данных. Агломеративно-иерархический кластерный анализ. Анализ главных компонент. Факторный анализ.	2	2		4	
17.	Регрессионный анализ. Модели линейного и нелинейного регрессионного анализа. Множественная линейная регрессия. Метод наименьших квадратов. Пошаговый регрессионный анализ.	2	2		4	
18.	Дисперсионный анализ. План эксперимента. Модели дисперсионного анализа: с постоянными факторами; со случайными факторами; смешанные модели. Задачи, решаемые дискриминантным анализом и анализом таблиц сопряженности.	2	2		4	Контрольное занятие на компьютерах

## VII. Содержание дисциплины

### Основные понятия теории вероятностей.

Пространство элементарных исходов и случайные события; операции над событиями; вероятности событий и их свойства; условная вероятность и независимость событий; теорема Бернулли; формула полной вероятности и формула Байеса; Формула Бернулли; случайные величины и функция распределения вероятностей; дискретные случайные величины; непрерывные случайные величины и функция плотности вероятностей; характеристики распределения случайной величины: математическое ожидание, дисперсия, медиана, мода и др.; примеры распределений случайных величин: распределение Бернулли, биномиальное распределение, распределение Пуассона, непрерывное и дискретное равномерные распределения, нормальное и логнормальное распределения; центральная предельная теорема; распределения, связанные с нормальным: распределение хи-квадрат, распределение Стьюдента и распределение Фишера. Характеристики распределения случайных величин: математическое ожидание, дисперсия, среднее квадратическое отклонение, коэффициент вариации, момент порядка  $\nu$ , центральный момент порядка  $\nu$ , коэффициент асимметрии, коэффициент эксцесса, медиана, мода, первая и третья квартили, интерквартильный размах, квантиль порядка  $p$ . Неравенство Чебышева. Статистические таблицы; многомерные случайные величины и их распределения; условные распределения и независимость случайных величин; коэффициент корреляции; двумерное нормальное распределение.

### Статистические данные.

Понятие случайной выборки; примеры реальных биологических экспериментов; многомерные статистические данные: матрица экспериментальных данных, переменные и наблюдения, количественные, номинальные и ранговые переменные.

### Дескриптивные и графические методы анализа данных.

Гистограмма; эмпирическая функция распределения; полигон частот; таблица частот; двумерные диаграммы рассеяния; множественные двумерные диаграммы рассеяния;

трехмерные диаграммы рассеяния; множественные трехмерные диаграммы рассеяния; столбчатые диаграммы; секторные диаграммы; составные линейные диаграммы; звездные графики; лучевые графики; оконные графики; разведочный анализ данных.

### **Статистическое оценивание.**

Понятие статистической оценки; свойства оценок: несмещенность, состоятельность, эффективность; метод максимального правдоподобия; точечное оценивание характеристик распределения (эмпирическая частота, выборочное среднее, выборочная дисперсия, выборочное среднееквадратическое отклонение, выборочный коэффициент вариации, выборочный коэффициент асимметрии, выборочный коэффициент эксцесса, выборочная медиана, выборочные квартили и интерквартильный размах, выборочная мода, выборочные квантили, выборочный коэффициент корреляции); интервальное оценивание; доверительный интервал для неизвестной вероятности; доверительные интервалы для математического ожидания и дисперсии нормального распределения; доверительный интервал для коэффициента корреляции.

### **Статистическая проверка гипотез.**

Логика проверки статистических гипотез; ошибки первого и второго рода, уровень значимости и мощность критерия; одновыборочные и двухвыборочные  $t$ -критерий и  $F$ -критерий; сравнение параметров двух биномиальных и двух пуассоновских распределений; проверка значимости отличия от нуля коэффициента корреляции; критерии согласия ( $\chi^2$ -квадрат критерий, одновыборочный критерий Колмогорова–Смирнова); непараметрические критерии: критерий Манна–Уитни для проверки гипотезы об отсутствии сдвига; критерий однородности – двухвыборочный критерий Колмогорова–Смирнова; критерий Спирмена для проверки гипотезы о независимости признаков.

### **Методы многомерного статистического анализа.**

Классификация методов многомерного статистического анализа; методы анализа связи между двумя системами переменных: корреляционный анализ, регрессионный анализ, дисперсионный анализ и дискриминантный анализ; методы анализа структуры многомерных данных: кластерный анализ, факторный анализ и анализ главных компонент.

### **Исследование зависимостей.**

Линейный регрессионный анализ; множественная линейная регрессия; метод наименьших квадратов; доверительные интервалы и проверка гипотез в линейном регрессионном анализе; множественный и частный коэффициенты корреляции; пошаговый регрессионный анализ; нелинейный регрессионный анализ; дисперсионный анализ: однофакторный и многофакторный дисперсионный анализ для моделей с постоянными уровнями факторов: проверка гипотез; методы множественного сравнения; модели со случайными уровнями факторов: проверка гипотез; коэффициент внутриклассовой корреляции; смешанные модели дисперсионного анализа: планы эксперимента со случайными блоками: планы с группировкой; непараметрические методы анализа связи между количественной и качественными переменными.

## **Лекция 1**

Описание методов и задач, решаемых теорией вероятностей и математической статистикой. Вероятностные модели. Понятие случайного испытания. Пространство элементарных событий. Примеры пространств элементарных событий в вопросах биологии. Случайные события. Операции над событиями (объединение, пересечение, дополнение). Достоверное и невозможное события. Несовместные события. Полная система событий. Обозначения, принятые в теории вероятности. Примеры введенных понятий.

## **Лекция 2**

Частота появления события в длинном ряду испытаний. Аксиоматическое введение вероятностей событий. Свойства вероятностей событий, вытекающие из аксиоматического введения вероятности. Теорема Бернулли. Естественность аксиоматического определения, вытекающая из теоремы Бернулли. Примеры введения вероятностей событий в биологических задачах. Вероятностное пространство. Модель равновероятного пространства элементарных событий (классическое введение вероятностей). Схемы выбора. Выбор без возвращения. Выбор с возвращением, Порядок выбора важен и не важен. Формулы комбинаторики для различных схем выбора. Геометрическая модель вероятностного пространства.

## **Лекция 3**

Определение условной вероятности. Естественность введенного определения на биологических примерах. Правило умножения. Независимость событий, Примеры независимых событий. Естественность введенного определения независимости событий на примере из биологии. Формула полной вероятности. Примеры задач на формулу полной вероятности. Формула Байеса. Примеры задач на формулу Байеса. Формула Бернулли. Примеры задач на формулу Бернулли.

## **Лекция 4**

Одномерная случайная величина. Примеры случайных величин в задачах биологии. Функция распределения случайной величины и ее свойства. Два типа случайных величин. Дискретные случайные величины. Табличная форма представления дискретной случайной величины. Функция распределения вероятностей дискретной случайной величины. Примеры дискретных случайных величин. Случайная величина, распределенная по дискретному равномерному закону. Случайная величина, распределенная по закону Бернулли. Биномиально распределенная случайная величина. Случайная величина, распределенная по закону Пуассона. Непрерывные случайные величины. Примеры непрерывных случайных величин в задачах биологии. Функция плотности вероятностей. Свойства функции плотности. Равномерно распределенная случайная величина на отрезке. Случайная величина, распределенная по экспоненциальному закону. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

## **Лекция 5**

Нормально распределенная случайная величина. Стандартное нормальное распределение. Формулировка центральной предельной теоремы. Графики функции плотности и функции распределения нормально распределенной случайной величины. Примеры применения рассмотренных распределений в задачах биологии. Распределения случайных величин, связанные с нормальным распределением. Случайная величина распределенная по закону хи-квадрат. Распределение Стьюдента. F- распределение. Таблицы нормального распределения, распределения хи-квадрат, распределения Стьюдента и F-распределения. Случайная величина, распределенная по логнормальному

закону. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 6**

Характеристики распределений. Математическое ожидание случайной величины. Свойства математического ожидания случайной величины. Примеры вычисления математического ожидания для ранее рассмотренных случайных величин. Дисперсия случайной величины. Свойства дисперсии случайной величины. Примеры вычисления дисперсии для ранее рассмотренных случайных величин. Неравенство Чебышева.

### **Лекция 7**

Характеристики положения и разброса случайной величины. Стандартное отклонение случайной величины. Коэффициент вариации случайной величины. Моменты и центральные моменты случайной величины. Коэффициент асимметрии и коэффициент эксцесса случайной величины. Медиана и мода случайной величины. Нижняя и верхняя квартили случайной величины. Квантили и процентиля распределения. Примеры вычисления рассмотренных характеристик.

### **Лекция 8**

Многомерные случайные величины. Функция распределения многомерной случайной величины и ее свойства. Примеры многомерных случайных величин в задачах биологии. Дискретные многомерные случайные величины. Непрерывные многомерные случайные величины. Характеристики связи случайных величин. Коэффициент корреляции и коэффициент ковариации двух случайных величин. Свойства коэффициента корреляции. Многомерное нормальное распределение на примере двумерного нормального распределения.

### **Лекция 9**

Генеральная совокупность. Простой случайный выбор. Случайная выборка. Объем выборки. Выборка как многомерная случайная величина. Эмпирическая функция распределения. Гистограмма выборки. Примеры построения эмпирической функции распределения и гистограммы. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 10**

Точечное статистическое оценивание. Выборочные оценки. Несмещенные оценки. Состоятельные оценки. Достаточное условие состоятельности. Асимптотически несмещенные оценки. Эффективные оценки. Выборочное среднее как несмещенная и состоятельная оценка математического ожидания случайной величины. Выборочная дисперсия. Выборочные моменты и центральные моменты. Выборочные коэффициенты асимметрии и эксцесса. Выборочная медиана. Выборочные нижняя и верхняя квартили. Выборочная квантиль. Выборочная мода. Методы получения выборочных оценок. Метод максимального правдоподобия. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 11**

Интервальное статистическое оценивание. Доверительный интервал. Доверительная вероятность. Центральные доверительные интервалы. Доверительный интервал для математического ожидания нормально распределенной случайной величины. Робастные доверительные интервалы. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 12**

Доверительный интервал для дисперсии нормально распределенной случайной величины. Доверительный интервал для параметра биномиального распределения. Доверительный интервал для параметра пуассоновского распределения. Доверительный интервал для коэффициента корреляции. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 13**

Статистическая проверка гипотез. Нулевая и альтернативная гипотезы. Простые и сложные альтернативные гипотезы. Статистика критерия. Критическая область. Вероятности ошибок первого и второго родов. Уровень значимости критерия. Мощность критерия. Наиболее мощные критерии. Взаимосвязь между вероятностями ошибок первого и второго родов. Примеры статистической проверки гипотез в задачах биологии. Параметрические и непараметрические критерии проверки статистических гипотез. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 14**

Одновыборочный t-критерий. Двухвыборочный t-критерий для независимых выборок и для связанных выборок. Одновыборочный хи-квадрат критерий. Двухвыборочный F-критерий. Сравнение параметров двух биномиальных распределений. Сравнение параметров двух пуассоновских распределений. Проверка гипотезы о равенстве заданному числу коэффициента корреляции. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 15**

Критерии согласия. Критерий согласия хи-квадрат. Одновыборочный критерий согласия Колмогорова-Смирнова. Непараметрические критерии. Ранги. Критерий знаков. Критерий знаковых рангов. Проверка гипотезы об отсутствии сдвига. Критерий Манна-Уитни. Критерий Вилкоксона. Критерий Ван-дер-Вардена. Критерии однородности. Двухвыборочный критерий Колмогорова-Смирнова. Проверка гипотезы о независимости. Коэффициент ранговой корреляции Спирмена. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 16**

Классификация методов анализа многомерных данных. Наблюдения. Переменные. Качественные, количественные и ранговые переменные. Методы анализа структуры многомерных данных. Методы анализа взаимосвязи между переменными. Агломеративно-иерархический кластерный анализ. Дендрограмма кластерного анализа. Кластеры. Проблема выбора меры различия. Проблема выбора расстояния между кластерами. Метод K-средних. Анализ главных компонент. Главные компоненты. Факторный анализ. Факторы. Факторные нагрузки. Метод варимаксного вращения. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 17**

Регрессионный анализ. Зависимые и независимые переменные. Модели линейного и нелинейного регрессионного анализа. Множественная линейная регрессия. Метод наименьших квадратов. Таблица дисперсионного анализа для модели множественной линейной регрессии. Доверительные интервалы для коэффициентов регрессии. Проверка гипотез о коэффициентах регрессии. Множественный и частный коэффициенты



корреляции. Пошаговый регрессионный анализ. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Лекция 18**

Дисперсионный анализ. План эксперимента. Модели дисперсионного анализа с постоянными факторами и со случайными факторами. Смешанные модели. Однофакторный дисперсионный анализ. Таблица дисперсионного анализа. Двухфакторный дисперсионный анализ. Задачи, решаемые дискриминантным анализом. Задачи, решаемые анализом таблиц сопряженности. Применение пакета статистических программ STATISTICA к рассматриваемым в лекции вопросам.

### **Семинар 1**

Пространство элементарных событий  $\Omega = \{\omega\}$ . Примеры пространства элементарных событий, которое может порождаться одной, двумя и более измеримыми характеристиками. Полное, достоверное и невозможное события. Операции над событиями: объединение событий  $A$  и  $B$ ; пересечение событий  $A$  и  $B$ ; дополнение события  $A$ . Наглядное представление событий в виде диаграммы Венна. Несовместные события. Полная система событий.

### **Семинар 2**

Аксиоматическое определение вероятности. Задание вероятностного пространства. Теорема Бернулли, связывающая аксиоматическое определение вероятности с ее частотной интерпретацией. Классическая модель вероятности случайного события. Четыре различные постановки задачи для определения числа способов,  $N$ , выбора  $m$  элементов из исходного множества  $n$  элементов: 1. Выбор без возвращения; порядок выбора важен; 2. Выбор без возвращения; порядок выбора не важен; 3. Выбор с возвращением; порядок выбора важен; 4. Выбор с возвращением; порядок выбора не важен. Повторение комбинаторных формул.

### **Семинар 3**

Геометрическая модель вероятности случайного события. Условная вероятность и независимость событий. Формула полной вероятности.

### **Семинар 4**

Формула Байеса: априорные вероятности; апостериорные вероятности. Испытания Бернулли. Формула Бернулли. Контрольная работа на дом.

### **Семинар 5**

Одномерные случайные величины. Функция распределения вероятностей одномерной случайной величины и ее свойства. Дискретные случайные величины. Примеры распределений дискретных случайных величин: дискретное равномерное распределение; распределение Бернулли; биномиальное распределение; распределение Пуассона.

### **Семинар 6**

Непрерывная случайная величина. Свойства функции плотности вероятностей. Примеры распределений непрерывных случайных величин: непрерывное равномерное

распределение; экспоненциальный закон распределения; нормальное распределение. Преобразование стандартизации и стандартное нормальное распределение. Распределения, связанные с нормальным: распределение хи-квадрат, распределение Стьюдента и распределение Фишера; логнормальное распределение.

### Семинар 7

Характеристики распределения случайных величин: математическое ожидание и его свойства, дисперсия и ее свойства. Вычисление этих характеристик для некоторых, рассмотренных ранее распределений. Характеристики распределения случайных величин: среднее квадратическое отклонение, коэффициент вариации, момент порядка  $\nu$ , центральный момент порядка  $\nu$ , коэффициент асимметрии, коэффициент эксцесса, медиана, мода, первая и третья квартили, интерквартильный размах, квантиль порядка  $p$ .

### Семинар 8

Контрольная работа.

### Семинар 9

Основы информатики. Пакет статистических программ Statistica. Рабочее окно системы Statistica: главное меню, панель инструментов. Настройка формы вывода результатов. Модульная структура системы Statistica. Модуль **Основные статистики/Таблицы**. Вероятностный калькулятор. Создать матрицы данных с вероятностями дискретных распределений: биномиального и пуассоновского с разными параметрами. Модуль **Подгонка распределения**. Построить графики функций распределения вероятностей для полученных распределений.

### Семинар 10

Модуль **Основные статистики/Таблицы**. **Вероятностный калькулятор**. Непрерывные распределения. Для нормального, логнормального, распределения хи-квадрат, распределения Стьюдента и распределения Фишера построить графики функций распределения и функций плотности с различными значениями параметров; задавая различные значения квантили, вычислить порядок квантили; по заданному значению порядка квантили вычислить значение квантили. Для каждого случая определить нижнюю и верхнюю квартили и интерквартильный размах. Исследовать свойства этих распределений в зависимости от параметров этих распределений.

### Семинар 11

Открытие файла данных. Организация файла данных. Примеры создания файла данных. Создание рабочего файла данных. Разделы главного меню **Графики, Файл**. Построение графиков **Усатые ящики**. Создание новых переменных. Редактирование созданного файла данных. Для переменных созданного файла данных построить графики эмпирической функции распределения, графики рассеяния, графики **Усатые ящики**, гистограммы. Редактировать и сохранить полученные результаты.

### Семинар 12

Модули **Описательные статистики и Корреляционные матрицы**. Вычисление значений выборочных характеристик. Построение двумерных и трехмерных графиков рассеяния. Проверка гипотез о коэффициентах корреляции.

### Семинар 13

Модули **Описательные статистики** и **Корреляционные матрицы**. Построение доверительного интервала для математического ожидания. Построение доверительного интервала для неизвестного параметра  $p$  биномиального распределения и для неизвестного параметра  $\lambda$  пуассоновского распределения. Построение двумерных и трехмерных графиков рассеяния. Построение доверительного интервала для коэффициента корреляции двумерного нормального распределения.

#### Семинар 14

Статистическая проверка гипотез. Модуль **Описательные статистики**. Одновыборочные и двухвыборочные  $t$ -критерий (для зависимых и независимых выборок) и  $F$ -критерий. Проверка гипотезы о равенстве нулю коэффициента корреляции двумерного нормального распределения. Критерии согласия. Одновыборочный критерий Колмогорова–Смирнова для случая простой нулевой гипотезы и для случая сложной нулевой гипотезы. Критерий  $\chi^2$ -квадрат. Для графической иллюстрации результатов использовать графики **Усатые ящики**.

#### Семинар 15

Непараметрические критерии. Двухвыборочные критерии: для проверки гипотезы об отсутствии сдвига для независимых выборок – критерий Манна-Уитни, и связанных выборок – критерий Вилкоксона. Для проверки гипотезы однородности - двухвыборочный критерий Колмогорова–Смирнова и критерий Вальда-Волфовитца. Для проверки гипотезы независимости признаков - критерий Спирмена. Эти критерии реализованы в Модуле **Непараметрические статистики**. Для графической иллюстрации результатов использовать графики **Усатые ящики**.

#### Семинар 16

Анализ многомерных данных. Переменные и наблюдения, количественные, качественные и ранговые переменные. Стандартизация переменных. Методы анализа структуры многомерных данных. Проблемы кластерного анализа: выбор меры различия – расстояния; выбор алгоритма последовательного объединения наблюдений (или признаков) в кластеры. Агломеративно- иерархический кластерный анализ. Модуль **Многомерные Исследовательские Методы/Кластерный анализ/Соединение**. Анализ дендрограммы. Метод  $K$ -средних. Факторный анализ. Метод главных компонент. Для графической иллюстрации результатов использовать графики рассеяния проекций наблюдений на пары первых главных компонент.

#### Семинар 17

Методы анализа зависимостей. Регрессионный анализ. Модуль – **Множественная регрессия**. Задать вид модели. Результаты анализа: мнк-оценки неизвестных параметров регрессии и предсказанные по модели значения зависимой переменной. Для проверки адекватности модели проводится анализ остатков. Для графической иллюстрации результатов используются графики зависимости предсказанных и наблюдаемых значений зависимой переменной от независимых переменных, а также графики остатков от наблюдаемых или предсказанных значений зависимой переменной. Дисперсионный анализ. Модуль **ANOVA**. Однофакторный и двухфакторный дисперсионный анализ для моделей с постоянными уровнями фактора. Для графической иллюстрации результатов использовать графики **Усатые ящики**. Однофакторный дисперсионный анализ для модели со случайными уровнями фактора. Результаты анализа - проверка гипотезы о незначимости влияния фактора (таблица дисперсионного анализа) и значения оценок компонент дисперсий и коэффициента внутриклассовой корреляции. Планы эксперимента

для смешанных моделей, которые наиболее часто используются при проведении биологических исследований: план со случайными блоками и план с группировкой.

### **Семинар 18**

Контрольное занятие. Каждому студенту предлагается выполнить десять заданий с использованием пакета статистических программ Statistica.

Целью первых восьми семинаров является усвоение теоретического материала курса, решение задач, проверка и разбор домашних заданий, и проведение контрольной работы. Первые восемь семинаров проводятся в аудитории. Целью последних десяти семинаров является подготовка грамотных пользователей персональных компьютеров; обучение использованию современных пакетов прикладных программ для освоения студентами теоретических методов анализа данных биологических исследований. Для каждого задания создаются и используются соответствующие файлы данных. Для иллюстрации результатов решения каждой задачи используются графические представления этих результатов. Все результаты сохраняются в специально созданной папке. Последние десять семинаров проводятся в дисплейном классе.

### **VIII. Перечень компетенций, формируемых в результате освоения дисциплины:**

ОНК – в результате освоения дисциплины студент должен знать принципы теоретических подходов и практических методов анализа биологических данных, применяемых в современной мировой науке.

ИК – в результате освоения дисциплины студент должен уметь использовать полученные знания для применения в использовании пакетов прикладных компьютерных программ при анализе экспериментальных данных

СК – в результате освоения дисциплины студент должен сформировать у себя системное представление о статистических, дескриптивных и графических методах анализа данных, с которыми часто сталкивается исследователь-биолог.

ПК – в результате освоения дисциплины у студента должна сформироваться способность применять современные методы анализа данных (статистическое оценивание и проверка гипотез, методы многомерного статистического анализа и исследования зависимостей) с использованием пакета прикладных компьютерных программ анализа данных STATISTICA.

### **IX. Используемые образовательные, научно-исследовательские и научно-производственные технологии:**

В преподавании дисциплины на лекциях используются демонстрационные компьютерные технологии. На семинарских занятиях студенты работают в дисплейном классе, при этом студенты обучаются работе с очень популярным в настоящее время пакетом прикладных компьютерных программ анализа данных STATISTICA.

### **X. Учебно-методическое обеспечение самостоятельной работы студентов, оценочные средства контроля успеваемости и промежуточной аттестации:**

Примерный список вопросов для проведения текущей, промежуточной аттестации и экзамена:

1. Пространство элементарных событий. Привести три примера.
2. Операции над событиями. Несовместные события. Привести примеры.

3. Задание вероятностного пространства. Привести примеры.
4. Свойства вероятности.
5. Условная вероятность. Привести примеры. Теорема умножения.
6. Формула полной вероятности.
7. Формула Байеса.
8. Независимые события. Привести примеры.
9. Формула Бернулли. Определить вероятность выпадения ровно 1 герба при 5-ти бросаниях правильной монеты.
10. Случайная величина. Привести три примера.
11. Функция распределения и ее свойства.
12. График функции распределения для какой-нибудь конкретной дискретной случайной величины.
13. Дискретная случайная величина. Привести примеры.
14. Распределение Бернулли. Математическое ожидание и дисперсия случайной величины, распределенной по Бернулли.
15. Биномиальное распределение. Математическое ожидание и дисперсия биномиально распределенной случайной величины.
16. Распределение Пуассона. Математическое ожидание и дисперсия случайной величины, распределенной по Пуассону.
17. Непрерывная случайная величина. Свойства функции плотности.
18. Непрерывное равномерное распределение. Математическое ожидание и дисперсия случайной величины, распределенной по равномерному закону на интервале (a,b).
19. Логнормальное распределение.
20. Нормальное распределение.
21.  $\chi^2$  –распределение, t- распределение, F- распределение.
22. Математическое ожидание случайной величины и его свойства.
23. Дисперсия случайной величины и ее свойства.
24. Средне-квадратическое отклонение и коэффициент вариации случайной величины.
25. Моменты, центральные моменты, абсолютные моменты случайной величины.
26. Коэффициенты асимметрии и эксцесса случайной величины.
27. Медиана, нижняя и верхняя квартили случайной величины.
28. Интерквартильный размах, мода случайной величины.
29. Квантили и процентиля распределения.
30. Многомерная случайная величина.
31. Коэффициент корреляции двух случайных величин и его свойства (без доказательства).
32. Случайная выборка. Выборочные значения. Объем выборки.
33. Гистограмма. График гистограммы для какого-нибудь примера.
34. Эмпирическая функция распределения. График этой функции для какого-нибудь примера.
35. Точечное оценивание. Несмещенные оценки.
36. Состоятельные оценки. Достаточное условие состоятельности.
37. Выборочное среднее, выборочная дисперсия; выборочное средне-квадратическое отклонение; выборочный коэффициент вариации.
38. Доказать несмещенность и состоятельность выборочного среднего, как оценки математического ожидания.
39. Выборочная квантиль, выборочная медиана, выборочные нижняя и верхняя квартили, выборочная мода.
40. Выборочные коэффициенты асимметрии и эксцесса.
41. Выборочный коэффициент корреляции.
42. Доверительный интервал для математического ожидания нормального распределения.
43. Доверительный интервал для дисперсии нормального распределения.

44. Доверительный интервал для коэффициента корреляции.
45. Доверительный интервал для параметра  $p$  биномиального распределения.
46. Доверительный интервал для параметра  $\lambda$  распределения Пуассона.
47. Проверка статистических гипотез. Вероятность ошибки 1-ого и 2-ого рода. Уровень значимости критерия и мощность критерия.
48. Одновыборочный  $t$ -критерий.
49. Двухвыборочный  $t$ -критерий (для независимых и связанных выборок).
50. Двухвыборочный  $F$ -критерий.
51. Проверка гипотезы о равенстве параметров биномиальных случайных величин.
52. Проверка гипотезы о равенстве параметров пуассоновских случайных величин.
53. Критерии согласия  $\chi^2$ , Колмогорова и Смирнова.
54. Критерии знаков и ранговых знаков.
55. Критерии для проверки гипотезы об отсутствии сдвига (для независимых и связанных выборок).
56. Проверка гипотез о независимости (для двумерного нормального и произвольных распределений).
57. Классификация методов многомерного статистического анализа.
58. Регрессионный анализ.
59. Дисперсионный анализ.
60. Кластерный анализ.

#### XI. Учебно-методическое и информационное обеспечение дисциплины:

Автор	Название книги/статьи	Отв. Редактор(для коллективных работ)	Место издания	Издательство	Год издания	Название журнала / сборника	Том	Номер

#### Основная литература

1. Мешалкин Л.Д. Сборник задач по теории вероятностей. М.: Изд-во МГУ, 1963, 154 с.
2. Мятлев В.Д., Панченко Л.А., Ризниченко Г.Ю., Терехин А.Т. Теория вероятностей и математическая статистика. Математические модели. Москва, Издательский центр «Академия», 2009. – 315 с.
3. В.Д.Мятлев, Л.А.Панченко, А.Т.Терехин «Основы теории вероятностей», Москва, МАКС Пресс, 2002 48 с.
4. В.Д.Мятлев, Л.А.Панченко, А.Т.Терехин «Основы математической статистики», Москва, МАКС Пресс, 2002, 57 с.
5. Мятлев В.Д., Панченко Л.А., Терехин А «Анализ многомерных данных», Москва, МАКС Пресс, 2007, 72 с.

#### Дополнительная литература

1. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. М.: Мир, 1982, 488 с.

2. Гмурман В.Е. Теория вероятностей и математическая статистика. 12-е изд. М.: Юрайт, 2005, 479 с.
3. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей. 1982. М.: Наука.
4. Колмогоров А.Н. Основные понятия теории вероятностей. М.: Наука, 1974, 119 с.
3. Крамер Г. Математические методы статистики. Москва: Мир, 1975, 648 с.
5. Тугубалин В.Н. Теория вероятностей и случайных процессов. М.: Из-во МГУ, 1992, 400 с.
8. Чжун К.Л., Айтсахлиа Ф. Элементарный курс теории вероятностей. М.: Бином, 2007, 455 с.
9. Шеффе Г. Дисперсионный анализ. М.: Наука, 1980, 512 с.
14. Sokal R.R., Rohlf F.J. 1995. Biometry. 3rd ed. N.Y.: W.H. Freeman & Co, 887 p.
15. Tereokhin A.T., de Meeus T., Guegan J,-F. On the Power of Some Binomial Modifications of the Bonferroni Multiple Test. Журнал общей биологии, 2007, Том 68.
16. Zar J.H.. Biostatistical Analysis. 4th ed. Upper Saddle River, N.J.: Prentice-Hall, Inc., 1999, 998 p.

## **ХII. Материально-техническое обеспечение дисциплины**

Лекции проводятся в аудитории, оснащённой проектором для показа презентаций. Студентам выдается презентация лекций в электронном виде. Большинство семинарских занятий проводятся в дисплейном классе (14 персональных компьютеров объединенных в сеть с возможностью выхода в интернет).

Авторы программы:

**д.б.н., профессор Терехин А.Т.**

к.ф.-м.н, доцент Мятлев В.Д. (Биологический ф-т МГУ)

к.т.н., доцент Панченко Л.А. (Биологический ф-т МГУ)