

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное учреждение
высшего профессионального образования
"Казанский (Приволжский) федеральный университет"
Институт фундаментальной медицины и биологии



подписано электронно-цифровой подписью

Программа дисциплины
Статистические методы в биологии М0.В.2

Направление подготовки: 020400.68 - Биология

Профиль подготовки: Биоинформатика

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Акберова Н.И.

Рецензент(ы):

Тарасов Д.С.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Алимова Ф. К.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института фундаментальной медицины и биологии:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No 849419214

Казань
2014

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Акберова Н.И. Кафедра биохимии и биотехнологии отделение биологии и биотехнологии, Natasha.Akberova@kpfu.ru

1. Цели освоения дисциплины

Ознакомление с основами биомедицинской статистики

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " М0.В.2 Гуманитарный, социальный и экономический" основной образовательной программы 020400.68 Биология и относится к вариативной части. Осваивается на 2 курсе, 3 семестр.

Дисциплина "Статистические методы в биологии" связана с курсами "Алгоритмы в геномике и протеомике", "Сравнительная геномика", "Вычислительная геномика и протеомика"

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-10 (профессиональные компетенции)	глубоко понимает и творчески использует в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов специальных дисциплин магистерской программы
ПК-16 (профессиональные компетенции)	имеет навыки формирования учебного материала, чтения лекций, готов к преподаванию в высшей школе и руководству научно-исследовательскими работами (НИР) студентов, умеет представлять учебный материал в устной, письменной и графической форме для различных контингентов слушателей
ОК-4 (общекультурные компетенции)	понимает пути развития и перспективы сохранения цивилизации, связь геополитических и биосферных процессов, проявляет активную жизненную позицию, используя профессиональные знания
ОК-5 (общекультурные компетенции)	проявляет инициативу, в том числе в ситуациях риска, способен брать на себя всю полноту ответственности, способен к поиску решений в нестандартных ситуациях
ПК-2 (профессиональные компетенции)	знает и использует основные теории, концепции и принципы в избранной области деятельности, способен к системному мышлению

В результате освоения дисциплины студент:

1. должен знать:

основные методы и подходы теории вероятности и статистики

2. должен уметь:

использовать статистические критерии

3. должен владеть:

методами статистической обработки и анализа данных

4. должен демонстрировать способность и готовность:

использовать методы статистики для анализа результатов собственных экспериментов

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 72 часа(ов).

Форма промежуточного контроля дисциплины зачет в 3 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Основные модели теории вероятности	3		2	0	0	коллоквиум
2.	Тема 2. Оценивание параметров модели. Пример использования метода максимального правдоподобия	3		2	3	0	контрольная работа
3.	Тема 3. Метод опорных векторов и его применения в вычислительной биологии	3		2	3	0	коллоквиум
4.	Тема 4. методы разведочного анализа	3		0	3	0	устный опрос
5.	Тема 5. Распознавание образов - базовые понятия	3		0	3	0	коллоквиум
6.	Тема 6. Выявление неоднородностей в данных	3		0	3	0	контрольная работа
7.	Тема 7. Методы кластеризации	3		0	3	0	контрольная работа
8.	Тема 8. Оценка вероятности события по частоте его появления	3		0	3	0	контрольная работа

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
9.	Тема 9. Расстояние Пирсона Хи-квадрат	3		0	3	0	контрольная работа
10.	Тема 10. Использование формулы Байеса для статистических выводов	3		0	3	0	контрольная работа
11.	Тема 11. Статистика в протеомике	3		0	3	0	коллоквиум
.	Тема . Итоговая форма контроля	3		0	0	0	зачет
	Итого			6	30	0	

4.2 Содержание дисциплины

Тема 1. Основные модели теории вероятности

лекционное занятие (2 часа(ов)):

Схема независимых испытаний Бернулли. Случайное распределение n точек по интервалу времени или пространства. Нормальное распределение (Гаусса-Лапласа).

Тема 2. Оценивание параметров модели. Пример использования метода максимального правдоподобия

лекционное занятие (2 часа(ов)):

Постановка задачи. Наивный подход. Метод максимального правдоподобия

практическое занятие (3 часа(ов)):

Работа с программным пакетом MATLAB (Octave)

Тема 3. Метод опорных векторов и его применения в вычислительной биологии

лекционное занятие (2 часа(ов)):

Обучение с учителем. Обобщение. Линейно разделимый случай. Лагранжиан. Решение оптимизационной задачи. Линейно неразделимый случай. Наиболее часто используемые ядра. Нелинейные SVM. Свойства SVM. Многоклассовая SVM. Применения SVM в геномике и протеомике

практическое занятие (3 часа(ов)):

Работа с программным пакетом MATLAB (Octave)

Тема 4. методы разведочного анализа

практическое занятие (3 часа(ов)):

Pipeline. Трансформации данных. Dimensionality reduction ? linear methods. Анализ главных компонент. Факторный анализ. Связь PCA и факторного анализа. Independent Component Analysis. Dimensionality reduction ? nonlinear methods. Multidimensional scaling. Metric MDS. Nonmetric MDS. Определение наличия линейных зависимостей в данных.

Тема 5. Распознавание образов - базовые понятия

практическое занятие (3 часа(ов)):

Выделение признаков. Два типа задач. Обучение без учителя. Обучение с учителем. Обучающая и тестовая выборки. Проблема: обобщение и переобучение. Кросс-валидация: проверка классификации. Уменьшение числа переменных (отбор признаков) . Этапы решения практических задач. Распознавание образов в биоинформатике.

Тема 6. Выявление неоднородностей в данных

практическое занятие (3 часа(ов)):

Различия в методике проверки гипотез сформулированных до и после опыта. Сравнение N нормальных величин с помощью хи-квадрат критерия. Множественные сравнения Шеффе. Разделение на два кластера в упорядоченной последовательности данных ("задача о разладке"). Критерий пустых ящиков

Тема 7. Методы кластеризации

практическое занятие (3 часа(ов)):

Общие вопросы. Расстояния: между бинарными переменными, между ordinal переменными (теми, которые могут быть упорядочены), между вещественными переменными, Махаланобиса, между смешанными переменными. Dissimilarity matrices. Пропущенные значения. Критерии для разбиений. Меры гетерогенности. Меры изолированности. Методы кластеризации. Иерархические аггломеративные. Single Linkage. Complete Linkage. Average Linkage. Centroid Linkage. Median. Ward. Иерархические дивизивные. Diana Density-based Partitioning. K-means. Mona. Кластеризация с использованием графов. Kohonen network. SOM. SOTA. Biclustering. Fuzzy clustering. Model-based clustering

Тема 8. Оценка вероятности события по частоте его появления

практическое занятие (3 часа(ов)):

Оценка вероятности события по частоте его появления. Оценка доли объектов в генеральной совокупности по их доле в выборке. Оценка параметра биномиального распределения. Построение доверительного интервала. Метод максимального правдоподобия. Построение интервала рассеяния и доверительного интервала на основе нормальной аппроксимации биномиального распределения. Вычислительные формулы для построения доверительного интервала. Простейшая задача планирования эксперимента.

Тема 9. Расстояние Пирсона ХИ-квадрат

практическое занятие (3 часа(ов)):

Применение хи-квадрат критерия для проверки простых гипотез. Применение хи-квадрат критерия для проверки сложных гипотез Важные моменты

Тема 10. Использование формулы Байеса для статистических выводов

практическое занятие (3 часа(ов)):

Использование формулы Байеса в теории вероятности и статистике. История вопроса. Основные сферы применения формулы Байеса.

Тема 11. Статистика в протеомике

практическое занятие (3 часа(ов)):

ПРИМЕНЕНИЕ СТАТИСТИКИ В ПРОТЕОМИКЕ. Разведочный анализ Нормировка Примеры Проблема пропущенных значений. Проблема выбросов. Классификация и снижение размерности. Отбор переменных. Классификация. ROC ? кривая. Сравнительный анализ классификаторов.

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Основные модели теории вероятности	3		подготовка к коллоквиуму	3	коллоквиум
2.	Тема 2. Оценивание параметров модели. Пример использования метода максимального правдоподобия	3		подготовка к контрольной работе	3	контрольная работа

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
3.	Тема 3. Метод опорных векторов и его применения в вычислительной биологии	3		подготовка к коллоквиуму	3	коллоквиум
4.	Тема 4. методы разведочного анализа	3		подготовка к устному опросу	3	устный опрос
5.	Тема 5. Распознавание образов - базовые понятия	3		подготовка к коллоквиуму	3	коллоквиум
6.	Тема 6. Выявление неоднородностей в данных	3		подготовка к контрольной работе	4	контрольная работа
7.	Тема 7. Методы кластеризации	3		подготовка к контрольной работе	4	контрольная работа
8.	Тема 8. Оценка вероятности события по частоте его появления	3		подготовка к контрольной работе	4	контрольная работа
9.	Тема 9. Расстояние Пирсона Хи-квадрат	3		подготовка к контрольной работе	3	контрольная работа
10.	Тема 10. Использование формулы Байеса для статистических выводов	3		подготовка к контрольной работе	3	контрольная работа
11.	Тема 11. Статистика в протеомике	3		подготовка к коллоквиуму	3	коллоквиум
	Итого				36	

5. Образовательные технологии, включая интерактивные формы обучения

Активные и интерактивные формы проведения занятий занимают большую часть аудиторной нагрузки в преподавании дисциплины "Статистические методы в биологии".
практические занятия проводятся на компьютерах с использованием биоинформационных баз данных и порталов в он-лайн режиме для решения конкретных биологических задач

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Основные модели теории вероятности

коллоквиум, примерные вопросы:

На коллоквиуме обсуждаются вопросы по теме "Основные модели теории вероятности"
Вопросы: -Схема независимых испытаний Бернулли -Случайное распределение n точек по интервалу времени или пространства -Нормальное распределение (Гаусса-Лапласа)
-Марковские цепи

Тема 2. Оценивание параметров модели. Пример использования метода максимального правдоподобия

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R -Задачи по регрессионному анализу

Тема 3. Метод опорных векторов и его применения в вычислительной биологии

коллоквиум , примерные вопросы:

На коллоквиуме обсуждаются вопросы по теме "Метод опорных векторов и его применения в вычислительной биологии " Вопросы: - Принципы обучения "с учителем" и "без учителя" -Свойства SVM - Применения SVM в геномике и протеомике

Тема 4. методы разведочного анализа

устный опрос , примерные вопросы:

Устный опрос проводится в виде обсуждения методов разведочного анализа -Методы Data Mining -Понятие разведочного анализа данных -Методы графического разведочного анализа данных и R -Аналитические методы первичного разведочного анализа данных - статистические характеристики количественных переменных -

Тема 5. Распознавание образов - базовые понятия

коллоквиум , примерные вопросы:

На коллоквиуме обсуждаются вопросы по теме "Распознавание образов " - Распознавание образов в биоинформатике, экспериментальные технологии - Вычислительные технологии: методы анализа последовательностей, методы смыслового анализа публикаций, молекулярное моделирование и докинг, методы анализа количественного соотношения структура-активность

Тема 6. Выявление неоднородностей в данных

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R - Задачи тестирование гипотез - Задачи на построение доверительных интервалов

Тема 7. Методы кластеризации

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R Задачи на классификацию - работа с пакетом lattice Визуализация многомерных данных с помощью пакета RGL

Тема 8. Оценка вероятности события по частоте его появления

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R Задачи на оценку доли объектов в генеральной совокупности по их доле в выборке. Задачи на оценку параметров биномиального распределения

Тема 9. Расстояние Пирсона ХИ-квадрат

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R Задачи на критерий ХИ-квадрат: общий случай, таблица сопряженности 2x2, критерий Мак-Нимара

Тема 10. Использование формулы Байеса для статистических выводов

контрольная работа , примерные вопросы:

Контрольная работа выполняется в среде R Задачи на оценку априорного распределения вероятностей

Тема 11. Статистика в протеомике

коллоквиум , примерные вопросы:

На коллоквиуме обсуждаются вопросы по теме "Статистика в протеомике " Вопросы: - Разведочный анализ : нормировка - Проблема пропущенных значений -Классификация и снижение размерности -ROC ? кривая

Тема . Итоговая форма контроля

Примерные вопросы к зачету:

Примерный план семинарских занятий

1. Комбинаторика. Классическое определение вероятностей
2. Геометрические вероятности
3. Независимые события. Условные вероятности. Схема Бернулли
4. Формула полной вероятности. Формула Байеса
5. Распределения случайных величин
6. Числовые характеристики случайных величин .
7. Предельные теоремы
8. Точечное оценивание. Метод моментов. Метод максимального правдоподобия. Свойства оценок. Сравнение оценок
9. Интервальное оценивание
10. Проверка статистических гипотез

Примеры задач по обработке числовых данных

1. По числовой выборке из нормальной совокупности с параметрами α , σ
2 построить доверительные интервалы для:
а) α , если σ
2 известно,
б) α , если σ
2 неизвестно,
в) σ
2
, если α известно,
г) σ
2
, если α неизвестно.
2. По данным числовым наблюдениям проверить основную гипотезу о равномерности распределения
с помощью: а) критерия Колмогорова; б) критерия хи-квадрат.
3. По данным двум выборкам из нормальных совокупностей проверить гипотезу
а) о совпадении дисперсий,
б) о совпадении средних, если известно, что дисперсии совпадают.

7.1. Основная литература:

Основная литература:

- Братусь А.С., Новожилов А.С., Платонов А.П. ? Динамические системы и модели биологии. - М.: Физматлит, 2010. - Режим доступа: <http://e.lanbook.com/view/book/2119>
- Иода Е. В. Статистика: Учебное пособие / Е.В. Иода. - М.: Вузовский учебник: НИЦ Инфра-М, 2012. - 303 с.: 60x90 1/16. (переплет) ISBN 978-5-9558-0144-5. - Режим доступа: <http://www.znaniyum.com/bookread.php?book=260143>
- Математические методы анализа и распознавания генетической информации: Монография / В.М. Гупал. - М.: ИЦ РИОР: НИЦ Инфра-М, 2012. - 154 с.: - Режим доступа: <http://znaniyum.com/bookread.php?book=309338>

7.2. Дополнительная литература:

Дополнительная литература:

Статистическая обработка результатов измерений. Компьютерные методы : учебное пособие [для студентов высших учебных заведений, обучающихся по специальностям 200503 "Стандартизация и сертификация" и 220501 - "Управление качеством"] / А. И. Сойко, Р. Н. Каратаев ; М-во образования и науки Рос. Федерации, Федер. агентство по образованию, Гос. образоват. учреждение высш. проф. образования "Казан. гос. техн. ун-т" .? Казань : [Изд-во Казанского государственного технического университета], 2010 .? 149, [1] с.

Статистика: Учебное пособие / О.А. Шумак, А.В. Гераськин. - М.: ИЦ РИОР: НИЦ Инфра-М, 2012. - 311 с. - Режим доступа: <http://znanium.com/catalog.php?bookinfo=261152>

7.3. Интернет-ресурсы:

Rosalind - обучение биоинформатике с помощью решения задач - <http://habrahabr.ru/post/149632/>

Журнал Biostatistics - <http://biostatistics.oxfordjournals.org/>

Обзор статистических программ - <http://www.sciencefiles.ru/section/46/>

Основы биостатистики - <http://mmb.bme.wisc.edu/stuff/GeneralInfo/website/Biostatisticsreview.pdf>

Программы статистического анализа данных - <http://www.sorashn.ru/index.php?id=2677>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Статистические методы в биологии" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

занятия по дисциплине проводятся в компьютерном классе, оборудованном мультимедийной техникой и выходом в Интернет

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 020400.68 "Биология" и магистерской программе Биоинформатика .

Автор(ы):

Акберова Н.И. _____

"__" _____ 201__ г.

Рецензент(ы):

Тарасов Д.С. _____

"__" _____ 201__ г.