

Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования
Московский государственный университет имени М.В. Ломоносова
Факультет биоинженерии и биоинформатики

УТВЕРЖДАЮ
Декан
факультета биоинженерии
и биоинформатики,
академик
_____/В.П. Скулачев /
« ____ » _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Базовые методы биологической статистики

Уровень высшего образования:
специалитет

Направление подготовки (специальность):

06.05.01 Биоинженерия и биоинформатика

Форма обучения:
очная

Рабочая программа рассмотрена и одобрена

Ученым советом факультета
(протокол № _____, _____)

Москва 20__

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по специальности 06.05.01 «Биоинженерия и биоинформатика» (программы специалитета) в редакции приказа МГУ от 30 декабря 2016 г.

Год (годы) приема на обучение – 2016, 2017, 2018, 2019.

© Факультет биоинженерии и биоинформатики МГУ имени М.В. Ломоносова

Программа не может быть использована другими подразделениями университета и другими вузами без разрешения факультета.

Цель и задачи дисциплины

Цель дисциплины - достичь качественного понимания базовых принципов статистического анализа.

Задачи:

- восстановить в памяти основные положения математической статистики, изученные ранее
- научиться выбирать подходящий статистический метод для основных исследовательских задач
- узнать о наиболее распространенных ошибках, встречающихся при применении статистики в биологических исследованиях

1. Место дисциплины в структуре ОПОП ВО: факультатив, курс IV – семестр 8.

2. Входные требования для освоения дисциплины, предварительные условия (если есть):
прохождение какого-либо курса статистики.

3. Планируемые результаты обучения по дисциплине:

Знать:

Знать стандартные методы статистической обработки

Знать основные ограничения на применение этих методов

Уметь

Выбирать необходимый критерий в типовых биологических задачах

Обнаруживать наиболее распространенные ошибки в применении статистики

4. Формат обучения – лекционные занятия.

5. Объем дисциплины составляет 1 з.е., в том числе 32 академических часа, отведенных на контактную работу обучающихся с преподавателем, 4 академических часа на самостоятельную работу обучающихся.

6. Краткое содержание дисциплины (аннотация):

Одна из основных проблем практического применения статистики в биологии – разрыв в понимании между математиками, придумывающими методы, и биологами, применяющими их. Цель данного курса – понимание, а не простое запоминание базовых принципов статистического анализа. Значительное место в курсе уделено тому, как в реальности происходит применение статистики – с распространенными ошибками и некоторыми традициями, не всегда имеющими строгое математическое обоснование.

Наименование и краткое содержание разделов и тем дисциплины, Форма промежуточной аттестации по дисциплине	Всего (часы)	В том числе			Самостоятельная работа обучающегося, часы (виды самостоятельной работы – эссе, реферат, контрольная работа и пр. – указываются при необходимости)
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, часы			
		Занятия лекционного типа	Занятия семинарского типа	Всего	
Тема 1. Типы переменных и описательные статистики	4	4	0	4	

Тема 2. Логика проверки статистических гипотез.	4	4	0	4	1
Тема 3. Сравнение групп. Проблема множественных сравнений..	6	6	0	6	1
Тема 4. Статистическая связь. Соотношение ее с причинно-следственной	6	6	0	6	0
Тема 5.Обобщенные линейные модели – объединение регрессионного и дисперсионного анализа	6	6	0	6	0
Тема 6. Анализ мощности. Как определить необходимую выборку?	6	6	0	6	1
Промежуточная аттестация – зачет					<i>1 (количество часов,отведенных на промежуточную аттестацию)</i>
Итого	36	32		4	

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

1) Исследователь закодировал формы листьев у некоторого растения: 1 - широкие, 2 - более-менее округлые, 3 - удлинённые и рассчитал медиану для характеристики центральной тенденции. Был ли он прав?

2) Длина некоторого вида жуков составляет (среднее±станд. откл.) $5,3 \pm 3$ г. Как вы думаете, нормально ли распределение длины в этой популяции?

3) Как оценить моду количественного признака в случае, если все значения в выборке различны?

4) Какова размерность дисперсии? Почему она выбрана в качестве основной характеристики разброса?

5) Про массу жуков в статье сказано, что она 10.5 ± 0.1 г. Как вы думаете, что стоит после плюса-минуса - стандартная ошибка или стандартное отклонение?

6) Что является нулевой гипотезой в тесте Хартиганов на мультимодальность : мультимодальность или унимодальность? Почему вы так решили? (гугление про тест портит карму)

7) Какой доверительный интервал шире: 90% или 95%? Почему? Что представляет собой 0% доверительный интервал?

8) Исследователь применил тест Тьюки для множественных сравнений и не нашел значимых различий. Какой тест ему стоит применить?

9) Когда острее стоит проблема множественных сравнений: при сравнении ЧЕТЫРЕХ групп между собой или при сравнении ПЯТИ экспериментальных групп с контрольной?

10) Коэффициент корреляции Спирмена между двумя переменными равен -1 (минус единица), а к.к. Пирсона равен -0.9. Что вы можете сказать о связи между этими переменными?

11) У некоторых животных коэффициент корреляции Пирсона между массой и длиной составляет 0.8, а между шириной и длиной - 0.7. Соответственно длина объясняет $0.8^2=0.64$, а ширина - $0.7^2=0.49$ вариации массы, что в сумме дает $0.64+0.49=1.13$. Где ошибка?

12) Приведите пример сильной связи между двумя переменными, которую не обнаружат коэффициенты корреляции ни Спирмена, ни Пирсона.

13) В каких пределах может изменяться мощность?

14) Как можно увеличить шансы на обнаружение эффекта некоторого интересующего вас размера, не увеличивая объем выборки?

15) Даже с маленькими выборками можно обнаружить слабые эффекты (только очень редко). Какая ошибка может возникнуть в тех исследованиях с малой выборкой, в которой статистической значимости достиг РЕАЛЬНО существующий слабый эффект?

16) Приведите пример статистического взаимодействия факторов из жизни.

17) В чем разница между generalized и general линейными моделями?

18) В чем биологический смысл свободного члена в линейных моделях?

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

Какой статистический метод можно применить в следующих ситуациях?

1) Выявляли, влияет ли освещенность на интенсивность окраски у мух. Личинок и куколок выдерживали в одном из четырех вариантов (темный, темноватый, светловатый и светлый) и смотрели окраску выплывших мух (бледные, желтенькие и желтые).

2) Выявляли влияние диет на ЖКТ. У тысячи опрошенных выяснили, диеты какой из пяти типов они придерживались, и какие заболевания ЖКТ у них возникли после перехода на эту диету (или не возникло никаких).

3) Выявляли связь срока выздоровления со способом лечения: стационарным и амбулаторным. Для этого собрали данные по длительности лечения (в днях) у сотни пациентов с одним и сотни - с другим способом лечения.

4) Выявляли, влияет ли курение на аварийность вождения. У пятисот водителей в ходе анкетирования выясняли, попадали ли они в аварию и являются ли они курильщиками.

5) Выявляли, влияет ли пол на предрасположенность к шизофрении. Материал исследования представлял данные о количестве шизофреников и здоровых среди группы наугад выбранных людей, мужчин и женщин.

6) Выявляли, зависит ли от пола время просмотра телевизора: сотню испытуемых (50 мужчин и 50 женщин) опросили, сколько времени они проводят, смотря телевизор, в неделю.

7) Выявляли, зависит ли доход людей от уровня образования: у 150 испытуемых (по 50 с низшим, средним и высшим образованием) выяснили среднемесячный доход.

8) Выясняли, различаются ли пять видов одного рода птиц по песенному репертуару. Для сотни отловленных птиц каждого вида смотрели, к какому из трех ранее выделенных типов относилась их песня.

9) Выявляли, влияет ли пол человека на длину волос: у сотни испытуемых записали пол и длину волос, отнесенную к одному из трех типов: короткие, средние и длинные.

Шкала и критерии оценивания результатов обучения по дисциплине.

Результаты обучения	«Неудовлетворительно»	«Удовлетворительно»	«Хорошо»	«Отлично»
Знания: - стандартные методы статистической обработки; - основные ограничения на применение этих методов	Знания отсутствуют	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения: - выбирать необходимый критерий в типовых биологических задачах; - обнаруживать наиболее распространенные ошибки в	Умения отсутствуют	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение обнаруживать наиболее распространенные ошибки	Успешное и систематическое умение

применении статистики				
Владения:	Навыки владения отсутствуют	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированн ые навыки (владения), применяемые при решении задач

8. Ресурсное обеспечение:

Перечень основной и дополнительной литературы:

1. Van Emden, H. F. (2015). *Statistics for terrified biologists*. John Wiley & Sons. Good, P. I., & Hardin, J. W. (2012). *Common Errors in Statistics (and How to Avoid Them)*. John Wiley & Sons.

Перечень профессиональных баз данных и информационных справочных систем:
<https://scholar.google.com/>

- Описание материально-технического обеспечения: маркерная доска, разноцветные маркеры