

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Физический факультет

УТВЕРЖДАЮ:

Декан физического факультета

 С.Н. Филимонов

« 01 »  2022 г.

Рабочая программа дисциплины

Анализ данных в биомедицине

по направлению подготовки

03.04.02 Физика

Направленность (профиль) подготовки
«Физические методы и информационные технологии в биомедицине»

Форма обучения

Очная

Квалификация

Магистр

Год приема

2022

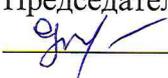
Код дисциплины в учебном плане: Б1.В.08

СОГЛАСОВАНО:

Руководитель ОП

 В.П. Демкин

Председатель УМК

 О.М. Сюсина

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

- УК-1 – способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий;
- ПК-2 – способен использовать свободное владение компьютерными программами анализа многомерных биомедицинских данных в задачах оценки состояния биосистем.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

- ИУК-1.1. Выявляет проблемную ситуацию, на основе системного подхода осуществляет ее многофакторный анализ и диагностику.
- ИУК-1.2. Осуществляет поиск, отбор и систематизацию информации для определения альтернативных вариантов стратегических решений в проблемной ситуации
- ИУК-1.3. Предлагает и обосновывает стратегию действий с учетом ограничений, рисков и возможных последствий.
- ИПК-2.1. Знает принципы и методы сбора, обработки и наглядного представления медико-биологической информации.
- ИПК-2.2. Умеет планировать и разрабатывать дизайн медико-биологических исследований с использованием современных компьютерных технологий и программных средств.
- ИПК-2.3. Владеет навыками визуализации, моделирования, анализа результатов биомедицинских исследований.

2. Задачи освоения дисциплины

- Освоение теоретических основ доказательной медицины и принципов планирования медико-биологических исследований;
- освоение методов и средств статистического анализа результатов медико-биологических исследований;
- приобретение навыков использования современного программного обеспечения для сбора, обработки и наглядного представления статистической информации.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений образовательной программы.

Дисциплина освещает вопросы об общих принципах анализа биомедицинских данных, проведения статистического анализа и представления его результатов с использованием современных программных средств.

Полученные в рамках дисциплины компетенции необходимы для:

- приобретения знаний научных принципов планирования медико-биологических исследований, основ доказательной медицины;
- усвоения общих принципов обобщения данных и проверки статистических гипотез, наглядного представления статистической информации;
- формирования умений разрабатывать дизайн медико-биологического исследования, осуществлять выбор статистических методов, соответствующих типам данных и виду дизайна исследования; обрабатывать данные и проводить статистические расчеты при помощи современных программных средств.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 2, экзамен.

5. Входные требования для освоения дисциплины

Входными требованиями для успешного освоения дисциплины являются:

- знание основ теории вероятности и математической статистики;
- базовые навыки работы с персональным компьютером.

Специальные компетенции для освоения дисциплины не предусмотрены.

6. Язык реализации

Английский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

- лекции: 12 ч.;
- практические занятия: 12 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Основы статистического анализа медико-биологических данных.

Общая характеристика биомедицинских данных. Типы данных. Шкалы измерений. Таблица экспериментальных данных. Пропуски в данных. Числовые характеристики случайных величин. Точечные и интервальные оценки. Случайные и систематические ошибки. Ошибки, специфичные для биомедицинских исследований.

Принципы доказательной медицины. Виды и дизайн клинических исследований. Критерии включения и исключения. Суррогатные и истинные оценки исхода. Рандомизированные контролируемые испытания – стандарт качества медицинских научных исследований.

Организация пакетов прикладных программ по статистическому анализу данных: интерфейс, настройка параметров, импорт и экспорт данных. Работа с данными: сортировка, графическое представление данных, расчёт описательных статистик.

Логика проверки статистических гипотез. Уровень статистической значимости и доверительная вероятность. Виды статистических критериев. Анализ повторных измерений. Виды и формы связей. Методы изучения корреляционной связи. Критерий Хи-квадрат. Цели и типовые задачи многомерного статистического анализа биомедицинских данных.

Тема 2. Методы принятия решений в биомедицине.

Постановка задачи принятия решений в биомедицине.

Геометрический подход к решению задачи принятия решений. Расстояние и мера сходства. Классификация по принципу минимума расстояния. Линейные дискриминантные функции. Алгоритмы кластеризации. Оценка результатов кластеризации и критерии качества распознавания. Методы обучения распознаванию. Нейронные сети. Модель и алгоритм перцептрона.

Статистический подход к решению задачи принятия решений Матрица потерь и Байесовский классификатор. Применение методов распознавания к решению задач медицинской диагностики. Последовательный анализ Вальда. Оценка информативности признаков. Информационная мера Кульбака.

Тема 3. Представление и публикация результатов медико-биологических исследований.

Источники научно-обоснованной информации. Виды нарушений практики добросовестных научных исследований. Авторство и плагиат. Конфликт интересов. Общая схема анализа научной публикации.

Требования Международного комитета редакторов медицинских журналов к рукописям, представляемым в биомедицинские журналы. Структура публикации

результатов биомедицинских исследований: резюме и введение, описание методов исследования и статистического анализа, графическое и табличное представление данных и результатов, обсуждение и выводы.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, оценки практических заданий, предполагающих самостоятельную работу по поиску, анализу, обработке информации, подготовке и оформлению результатов в форме отчетов.

Балльная оценка текущего контроля успеваемости студента по данной дисциплине составляет максимум **60 баллов**.

Таблица 9.1

№ п/п	Вид контроля	Количество	Количество баллов за 1 ед. контроля	Сумма
1.	Посещение лекций	6	1	6
2.	Выполнение практических заданий	3	18	54
	ИТОГО			60

Основным критерием балльной оценки текущего контроля успеваемости является **оценка качества выполнения практического задания** (содержание ответа, полнота ответа, владение профессиональным языком).

Индикаторы балльной оценки практического задания:

– 15-18 баллов – ответ не содержит ошибочных расчетов, элементов и утверждений, максимально полно раскрывает суть каждого вопроса, составлен профессиональным языком, содержит выводы;

– 10-14 баллов – в ответе допущены принципиальные ошибки и неточности в расчетах, ответ содержит упущения, составлен профессиональным языком, содержит выводы;

– 5-9 баллов – ответ содержит несколько ошибок в расчетах, упущения, содержание ответов не полное; составлен профессиональным языком, в выводах допущены неточности;

– 0-4 баллов – ответ содержит многочисленные ошибки в расчетах, упущения, содержание ответов не полное; выводы отсутствуют.

Текущий контроль фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен во втором семестре проводится в письменной форме по билетам. Каждый экзаменационный билет состоит из двух теоретических вопросов по одной из тем дисциплины. Продолжительность экзамена 1,5 часа.

Примерный перечень теоретических вопросов

1. Типы биомедицинских данных и шкалы измерений.
2. Таблица экспериментальных данных. Анализ пропусков в данных.
3. Случайные и систематические ошибки в биомедицинских исследованиях.
4. Числовые характеристики случайных величин.
5. Логика проверки статистических гипотез.
6. Виды статистических критериев.
7. Методы изучения взаимосвязи признаков.

8. Расстояние и мера сходства.
9. Линейные дискриминантные функции. Понятие обобщенных ЛДФ.
10. Случаи линейной разделимости объектов на классы.
11. Классификация объектов по принципу минимума расстояния.
12. Модель перцептрона обучения распознаванию.
13. Кластеризация. Оценка результатов кластеризации.
14. Алгоритмы максимального расстояния и M - групповых средних.
15. Статистический подход к распознаванию образов. Байесовский классификатор.
16. Последовательный анализ Вальда в задачах медицинской диагностики.
17. Оценка информативности признаков. Информационная мера Кульбака.
18. Источники научно-обоснованной медицинской информации.
19. Общая схема анализа медицинской публикации.
20. Типичные ошибки представления результатов биомедицинских исследований.

К экзамену допускаются только те студенты, кто удовлетворительно выполнил все практические задания.

Балльная оценка промежуточной аттестации (в форме экзамена) составляет максимум **40 баллов**.

Индикаторы балльной оценки ответа на экзамене:

– 32-40 баллов – ответы на вопросы билета не содержат ошибочных элементов и утверждений, ответы на дополнительные устные вопросы экзаменатора содержательны и убедительны;

– 22-32 баллов – в ответах на вопросы билета допущены не принципиальные ошибки и неточности, ответы на дополнительные устные вопросы экзаменатора содержат упущения;

– 11-21 баллов – в ответах на вопросы билета допущены несколько принципиальных ошибок, ответы на дополнительные устные вопросы экзаменатора содержат упущения;

– 0-10 баллов – ответы на вопросы билета имеют многочисленные ошибки, упущения или содержание ответов не имеет отношения к поставленному вопросу; ответы на дополнительные устные вопросы экзаменатора содержат ошибки.

Баллы, полученные на экзамене, суммируются с баллами, полученными по итогам текущего контроля. На основе итогового количества баллов выставляется оценка.

Соответствие 100-балльной шкалы оценок 4-альтернативной шкале оценок:

- 0-22 баллов – «неудовлетворительно»,
- 26-48 баллов – «удовлетворительно»,
- 52-74 баллов – «хорошо»,
- 77-94 баллов – «отлично».

В общем количестве баллов не учтены баллы за посещение лекций.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» – <https://moodle.tsu.ru/course/view.php?id=2950>;

б) Fokin V. Data analysis in biomedicine. Lecture course: trans. by Aleksandra Nabiullina/ Fokin V. – Tomsk: Publishing house of Tomsk State University. – 2016;

в) оценочные материалы текущего контроля и промежуточной аттестации по дисциплине;

г) примерные темы практических занятий:

1. Практическое занятие №1 «Знакомство со статистическими пакетами обработки данных».

2. Практическое занятие №2 «Расчет описательных статистик в статистических пакетах обработки данных».

3. Практическое занятие №3 «Анализ взаимосвязей количественных и качественных данных».
4. Практическое занятие №4 «Дискриминантный анализ как метод прогноза в медицине».
5. Практическое занятие №5 «Оценка информативности признаков для принятия решений».
6. Практическое занятие №6 «Анализ обоснованности статистических выводов в научных публикациях».

1. Практическое занятие №1 «Знакомство со статистическими пакетами обработки данных»

Примеры заданий

1. Ознакомиться со структурой ППП по статистической обработке данных Statistica.
2. Ознакомиться с основными видами документов, используемых в ППП Statistica: таблица данных, рабочая книга, график, отчет.
3. Создать простейшую таблицу с данными и ознакомиться с основными методами преобразования данных: сортировка, категоризация, вычисление по формулам, создание подвыборок.

2. Практическое занятие №2 «Расчет описательных статистик в статистических пакетах обработки данных»

Примеры заданий

Задание 1

Провести анализ данных на соответствие количественных признаков нормальному закону распределения с помощью изученных критериев. Сделайте выводы.

Задание 2

Провести расчет значений следующих описательных статистик всех количественных показателей для всей выборки, а также отдельно для каждой группы пациентов: среднее, медиана, стандартное отклонение, нижний и верхний квартиль, 95%-ый доверительный интервал.

3. Практическое занятие №3 «Анализ взаимосвязей количественных и качественных данных».

Примеры заданий

Задание 1

Провести для всей выборки, а также отдельно для каждой группы пациентов, корреляционный анализ данных и определите статистически значимые взаимосвязи между признаками. Сделайте выводы.

Задание 2

Оценить межгрупповые различия всех показателей из представленной базы данных. Сделайте выводы.

Задание 3

Оценить внутригрупповые различия показателей эхокардиографии до и после оперативного вмешательства в каждой группе. Сделайте выводы.

Задание 4

Проведено исследование, в котором исследовалась связь между курением и развитием ишемической болезни сердца (ИБС). С исследованием вошло 200 человек. У 60 курящих и 35 некурящих людей развилась ИБС, а у 40 курящих и 65 некурящих не развилась ИБС.

Оценить, имеется ли связь между курением и развитием ИБС?

Сделайте выводы.

4. Практическое занятие №4 «Дискриминантный анализ как метод прогноза в медицине».

Примеры заданий

По результатам обследования показателей красной крови получены следующие данные:

№ пациента	Нб, Гемоглобин (усл.ед.)	СП Цветной показатель (усл.ед.)	СОЭ Скорость оседания эритроцитов (мм/мин)	Анемия
1	1,3	0,9	5,3	Есть
2	1,4	0,8	5,6	Есть
3	1,6	1,0	4,2	Нет
4	1,2	0,9	5,8	Есть
5	1,5	1,3	5,1	Нет
6	1,7	1,1	5,0	Нет
7	1,4	1,3	5,5	Нет
8	1,1	1,0	4,7	Есть
9	0,8	1,0	5,1	Есть
10	1,0	1,1	5,7	
11	1,5	0,8	4,9	

– Построить правило классификации (ЛДФ) для диагностики анемии по принципу минимума расстояния, построить диагностическую таблицу, рассчитать чувствительность и специфичность теста, качество распознавания (общий % ошибок классификации).

– Установить, есть или нет анемия у 10 и 11 пациента.

5. Практическое занятие №5 «Оценка информативности признаков для принятия решений».

Примеры заданий

Оценить информативность признаков по статистике наблюдений над группами детей с ОРВ: (А1 - тяжелая степень, А2 - средняя степень), постройте диагностическую таблицу и рассчитайте пороги принятия решений на уровне ошибок первого и второго рода 0,05 и 0,1 соответственно.

Признаки	A1	A2	Гипер-термия	A1	A2	Возраст	A1	A2
X11	8	176	X21	2	63	X31	1	7
						X32	1	56
			X22	3	96	X31	2	2
						X32	1	94
						X31	1	3
X23	3	17	X32	2	14			
			X31	1	3			
X12	10	146	X21	3	30	X31	1	7
						X32	2	23
			X22	2	92	X31	1	9
						X32	1	83
						X31	2	2
X23	5	24	X32	3	22			
X13	43	48	X21	5	11	X31	3	2
						X32	2	9
			X22	7	23	X31	2	6
						X32	5	17
						X31	10	3
X23	31	14	X32	21	11			
X14	30	7	X21	3	2	X31	2	1
						X32	1	1
			X22	7	2	X31	4	1
						X32	3	1
						X31	9	2
X23	20	3	X32	11	1			

X1 – окраска кожи: X11 – нормальная, X12 – бледная однотонная, X13 – бледная мраморная, X14 – бледная с сыпью.

X2 – длительность гипертермии $>39^{\circ}$, в часах: X21 – менее часа, X22 – более часа и менее 6 часов, X23 – более 6 часов.

X3 – возраст: X31 – менее трех месяцев, X32 – более трех месяцев.

6. Практическое занятие №6 «Анализ обоснованности статистических выводов в научных публикациях».

Примеры заданий

Проанализировать предложенную преподавателем статью и ответьте вопросы:

- Отражает ли резюме название работы и ее содержание?
- Достаточно ли ясно описаны методы планирования исследования и методы статистического анализа?
- Обосновать адекватность методов статистического анализа заявленным целям исследования.
- Какой план исследования использован автором.
- Имеется ли контрольная группа.
- Оценить достаточность объема выборок.
- Проанализировать обоснованность выводов статьи.

- Оценить качество представления графических материалов.

Характерными показателями развития самостоятельности у студента в результате освоения дисциплины являются: теоретическое осмысление изучаемого материала, накопление необходимых умений и навыков, интерес к процессу создания продукта собственной самостоятельной деятельности, умение провести презентацию созданного продукта, умение отстаивать собственную точку зрения или предложенный вариант решения проблемы, рефлексия своей деятельности и результата.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Fokin V. Data analysis in biomedicine. Lecture course: trans. by Aleksandra Nabiullina/ Fokin V. – Tomsk: Publishing house of Tomsk State University. – 2016.

б) дополнительная литература:

1. Uniform requirements for manuscripts submitted to biomedical journals [Electronic resource] / International committee of medical journal editors // *Annals of internal medicine*. – 1997. – Vol. 126, is. 1. – P. 36–47. – The electronic version of the printing publication. – URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1488638/pdf/cmaj00295-0079.pdf> (access date: 20.02.2022).

2. Isaak C. A. EBM: evidence to practice and practice to evidence [Electronic resource] / C. A. Isaak, A. Franceschi // *Journal of evaluation in clinical practice*. – 2008. – Vol. 14, is. 5. – P. 656–659. – The electronic version of the printing publication. – URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2753.2008.01043.x/full> (access date: 20.02.2022).

3. Marewski J. N. Heuristic decision making in medicine [Electronic resource] / J. N. Marewski, G. Gigerenzer // *Dialogues in Clinical Neuroscience*. – 2012. – Vol. 14, is. 1. – P. 77–89. – The electronic version of the printing publication. – URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341653/> (access date: 20.02.2022).

4. Prasad K. Fundamentals of evidence based medicine / K. Prasad. – Springer India, 2013. – 154 p. – The electronic version of the printing publication. – URL: <http://link.springer.com/book/10.1007/978-81-322-0831-0> (access date: 20.02.2022).

5. Sidhu A. S. Biomedical data and applications [Electronic resource] / A. S. Sidhu, T. S. Dillon. – Berlin : Heidelberg : Springer-Verlag , 2009. – 342 p. – (Studies in computational intelligence, vol. 224). – The electronic version of the printing publication. – URL: <http://link.springer.com/book/10.1007/978-3-642-02193-0> (access date: 20.02.2022).

6. Zeger A. L. On time series analysis of public health and biomedical data [Electronic resource] / S. L. Zeger, R. Irizarry, R. D. Peng // *Annual Review of public health*. – 2006. – Vol. 27. – P. 57–79. – The electronic version of the printing publication. – URL: <http://www.annualreviews.org/doi/pdf/10.1146/annurev.publhealth.26.021304.144517> (access date: 20.02.2022).

7. Evidence-based medicine: how to practice and teach it [Electronic resource] / S. E. Straus. – 3rd ed. – Edinburgh: Elsevier, 2005. – 315 p. – The electronic version of the printing publication. – URL: <http://www.amazon.com/Evidence-Based-Medicine-Practice-Teach-Straus/dp/0702031275> (access date: 20.02.2022).

в) ресурсы сети Интернет:

1. Электронный учебник по пакету Statistika – <http://statsoft.ru/home/textbook/default.htm> (русскаяязычная версия); <http://www.statsoft.com/Textbook> (англыязычная версия).

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Professional Plus 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office Access, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.);
- программа Statistica 10 (академическая лицензия).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ –
<http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ –
<http://vital.lib.tsu.ru/vital/access/manager/Index>

- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных (при наличии):

14. Материально-техническое обеспечение

Для проведения лекционных и семинарских занятий используется лаборатория моделирования физических процессов в биологии и медицине (аудитория № 442 второго учебного корпуса ТГУ), оснащенная интерактивной доской, звуковым и видеооборудованием, мультимедийным оборудованием для демонстрации презентаций, ресурсов сети Интернет, других учебных материалов. Имеются персональные компьютеры студентов, с доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Для проведения практических занятий может использоваться материально-техническая база кафедры медицинской и биологической кибернетики Сибирского государственного медицинского университета.

Организация занятий возможна с использованием дистанционных образовательных технологий.

5. Информация о разработчиках

Фокин Василий Александрович, доктор технических наук, профессор кафедры медицинской и биологической кибернетики Сибирского государственного медицинского университета.