

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

Филологический факультет

УТВЕРЖДАЮ:

Декан филологического факультета


И.В. Тубалова

« 31 » августа 2022 г.

Рабочая программа дисциплины

Математическая статистика

по направлению подготовки

45.03.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) подготовки:

Фундаментальная и прикладная лингвистика

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2022

Код дисциплины в учебном плане: Б1.В.06

СОГЛАСОВАНО:

Руководитель ОПОП


А.В. Васильева

Председатель УМК


Ю.А. Тихомирова

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 – Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур;

ПК-4 – Способен разрабатывать программный код при решении задач автоматической обработки текстов.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.3 Применяет математические методы в процедурах анализа и синтеза лингвистических структур.

ИОПК-2.2 Соотносит задачи формализации лингвистических знаний с основами математических знаний.

ИПК-4.1 Применяет способы формализации и алгоритмизации поставленных задач в сфере автоматической обработки текстов.

2. Задачи освоения дисциплины

- Изучить основные понятия и методы математической статистики;
- Освоить аппарат статистического анализа данных;
- Научиться применять изученные методы статистического анализа для решения практических задач с использованием пакетов программ для анализа данных

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения.

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

Пятый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: понятийный аппарат математики, информационные технологии и основы информационной культуры в лингвистике, информатика и основы программирования.

6. Язык реализации

Русский

7. Объем дисциплины

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

-лекции: 10 ч.

-практические занятия: 22 ч.

в том числе практическая подготовка: 22 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение.

Измерительные шкалы. Генеральная и выборочная совокупности. Способы отбора данных. Способы представления выборок (табличные, графические и др.)

Тема 2. Оценка параметров.

Точечные оценки и их свойства. Числовые характеристики выборки. Интервальное оценивание.

Тема 3. Проверка статистических гипотез.

Определение статистической гипотезы. Нулевая и альтернативная гипотезы. Ошибки первого и второго рода и их вероятности. Статистические критерии. Мощность критерия. Критические области и критические точки. Алгоритм проверки гипотез. Критерий согласия Пирсона. P-value. Проверка нормальности.

Тема 4. Критерии сравнения групп.

Параметрические критерии сравнения групп (t-критерий Стьюдента, критерий Фишера, дисперсионный анализ). Непараметрические критерии (критерии Манна-Уитни, Вилкоксона, Краскала-Уолиса, Фридмана).

Тема 5. Корреляционный анализ.

Количественная корреляция. Парный коэффициент корреляции Пирсона. Ранговая корреляция. Коэффициент Спирмена. Коэффициент конкордации. Анализ таблиц сопряженности.

Тема 6. Регрессионный анализ.

Постановка задачи регрессии. Парная регрессия. Оценка параметров уравнения регрессии. Проверка качества уравнения регрессии. Множественная регрессия.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения тестов по лекционному материалу и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в пятом семестре проводится в форме итогового теста по всему материалу курса. Итоговый тест проверяет ИОПК-2.2.

В тесте 15 вопросов от 1 до 3 баллов за каждый. Максимум за тест – 30 баллов.

0-15 неудовлетворительно,

16-20 удовлетворительно.

21-25 хорошо,

26-30 отлично.

Список теоретических вопросов для подготовки к тесту.

1. Типы статистических данных. Измерительные шкалы.
2. Понятие вариационного ряда.
3. Полигон частот и гистограмма.
4. Эмпирическая функция распределения и ее свойства.
5. Точечное оценивание.
6. Интервальное оценивание.
7. Средние показатели.
8. Показатели вариации.
9. Правило сложения дисперсий.

10. Выборочные моменты.
11. Выборочные коэффициенты асимметрии и эксцесса.
12. Выборочные мода и медиана.
13. Понятие выборочного квантиля.
14. Нулевая и альтернативная гипотезы.
15. Ошибки первого и второго рода. Мощность критерия.
16. Критические области.
17. Алгоритм проверки гипотез.
18. Критерий согласия Пирсона.
19. t-критерий Стьюдента.
20. F-критерий Фишера.
21. Непараметрические критерии сравнения двух выборок.
22. Непараметрические критерии сравнения нескольких выборок.
23. Парный коэффициент корреляции Пирсона.
24. Ранговый коэффициент корреляции Спирмена.
25. Коэффициент конкордации.
26. Общая постановка задачи корреляционного анализа категоризованных данных.
Таблица сопряженности.
27. Общая постановка задачи регрессионного анализа.
28. Парная регрессия. Постановка задачи. Оценка параметров.
29. Проверка качества уравнения парной регрессии.
30. Множественная регрессия. Постановка задачи.

Примерные вопросы теста:

1. Среднее геометрическое определяется по формуле:
 - а) $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$;
 - б) $\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i}$;
 - в) $\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$;
 - г) $\bar{x} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$.
2. Несмещенная оценка дисперсии выборки 8, 14, 14, 17, 23, 26, 29, 29, 32, 38 равна
 - а) 10
 - б) 30
 - в) 81
 - г) 90
 - д) 180
 - е) 810
3. Вероятностью ошибки второго рода является
 - а) вероятность отклонения верной нулевой гипотезы;

- б) вероятность принятия ложной нулевой гипотезы.
4. Для оценки связи между двумя категориальными переменными можно применить
- а) коэффициент корреляции Пирсона;
 - б) коэффициент корреляции Спирмена;
 - в) коэффициент квадратичной сопряженности.
5. Для выявления статистически значимых отличий между тремя зависимыми выборками можно применить критерий
- а) Вилкоксона
 - б) Краскала-Уолиса
 - в) Фридмана.

и т.п.

Практическая часть оценивается по результатам выполнения домашнего итогового исследования и проверяет ИОПК-2.3, ИПК-4.1. Необходимо сформулировать цель исследования, собрать данные, выбрать метод анализа, реализовать его и сделать выводы из полученных результатов. Оценивается на «зачтено» или «незачтено».

Зачтено – цель исследования поставлена корректно, выбран адекватный метод анализа, метод реализован верно, сделаны правильные выводы из полученных результатов.

Незачтено – цель сформулирована некорректно, выбран неадекватный метод анализа, не соответствующий типу данных и цели исследования, метод выбран правильно, но реализован некорректно, сделаны неверные выводы из полученных результатов.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=8908>;

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Гмурман, В. Е. Теория вероятностей и математическая статистика : учебник для прикладного бакалавриата / В. Е. Гмурман. — 12-е изд. — М. : Издательство Юрайт, 2016. — 479 с. — (Бакалавр. Прикладной курс). — URL: <https://www.biblio-online.ru/book/69836F34-AEF2-49FD-B438-3C1EC3996F17>.
2. Ивашев-Мусатов, О. С. Теория вероятностей и математическая статистика : учебник и практикум для академического бакалавриата / О. С. Ивашев-Мусатов. — 3-е изд., испр. и доп. — М. : Издательство Юрайт, 2016. — 224 с. — URL: <https://www.biblio-online.ru/book/BB281C15-FCBA-4362-B6C4-7A861A37F8A8>.
3. Кремер, Н. Ш. Теория вероятностей и математическая статистика в 2 ч. Часть 1. Теория вероятностей : учебник и практикум для академического бакалавриата / Н. Ш. Кремер. — 4-е изд., перераб. и доп. — М. : Издательство Юрайт, 2016. — 264 с. — (Бакалавр. Академический курс). — URL: <https://www.biblio-online.ru/book/3BC02C6C-E0AE-4E81-A340-00EC8442906A>.

б) дополнительная литература:

1. Статистика : учебник для прикладного бакалавриата / И. И. Елисеева, М. В. Боченина, Н. В. Бутова, Б. А. Михайлов ; под ред. И. И. Елисеевой. — 2-е изд., перераб. и доп. —

- М. : Издательство Юрайт, 2016. — 446 с. — (Бакалавр. Прикладной курс). — URL: <https://www.biblio-online.ru/book/E322DC6F-0297-4108-BD0E-77ED0CE55682>
2. Харченко Н. М. Статистика : учебник / Н. М. Харченко. - 2-е изд., перераб. и доп. - Москва : Дашков и К°, 2011. - 366, [1] с.: ил., табл.
 3. Сидоренко Е. В. Методы математической обработки в психологии / Е. В. Сидоренко. - СПб. : Речь, 2001. - 349,[3] с.: ил.
- Ермолаев-Томин, О. Ю. Математические методы в психологии в 2 ч. Часть 1. : учебник для академического бакалавриата / О. Ю. Ермолаев-Томин. — 5-е изд., испр. и доп. — М. : Издательство Юрайт, 2016. — 280 с. — URL: <https://www.biblio-online.ru/book/4676E3A0-BC8B-4569-86DA-0CED43842D94>.

13. Перечень информационных технологий

- а) лицензионное и свободно распространяемое программное обеспечение:
- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, 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.);
 - R – <https://www.r-project.org/>;
 - R Studio – <https://www.rstudio.com/>.
 - JASP - <https://jasp-stats.org/>.

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/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Лаборатории, оборудованные компьютерами для практических работ с соответствующим программным обеспечением и интернетом, проектором.

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, кандидат физико-математических наук, доцент кафедры теории вероятностей и математической статистики ИПМКН ТГУ