

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

Филологический факультет

УТВЕРЖДЕНО:  
Декан  
И.В. Тубалова

Рабочая программа дисциплины  
**Статистические методы в гуманитарных исследованиях**  
по направлению подготовки

**45.04.03 Фундаментальная и прикладная лингвистика**

Направленность (профиль) подготовки:  
**Компьютерная и когнитивная лингвистика**

Форма обучения  
**Очная**

Квалификация  
**Магистр**

Год приема  
**2024**

СОГЛАСОВАНО:  
Руководитель ОП  
З.И. Резанова

Председатель УМК  
Ю.А. Тихомирова

Томск – 2025

## **1. Цель и планируемые результаты освоения дисциплины**

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен анализировать, сопоставлять и критически оценивать различные лингвистические направления, теории и гипотезы при решении задач профессиональной деятельности.

ОПК-6 Способен осуществлять эффективное управление разработкой программных средств информационных проектов в сфере своей профессиональной деятельности.

ПК-3 Способен разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты электронных ресурсов и интеллектуальных электронных систем (лингвистические корпуса, словари, онтологии, базы данных).

УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.2 Формулирует гипотезы в рамках предложенного лингвистического направления и теории при решении задач профессиональной деятельности

ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования

ИОПК-6.2 Разрабатывает алгоритмы и программы для решения лингвистических и междисциплинарных задач в том числе с применением высокопроизводительных вычислительных технологий

ИПК-3.3 Разрабатывает лингвистические компоненты интеллектуальных информационных систем (онтологии, базы данных)

ИУК-1.1 Выявляет проблемную ситуацию, на основе системного подхода осуществляет её многофакторный анализ и диагностику

## **2. Задачи освоения дисциплины**

- Освоить аппарат статистического анализа данных.
- Научиться применять понятийный аппарат и основные методы статистического анализа для решения практических задач профессиональной деятельности.

## **3. Место дисциплины в структуре образовательной программы**

Дисциплина относится к Блоку 1 «Дисциплина (модули)».

Дисциплина относится к обязательной части образовательной программы.

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

Второй семестр, экзамен

## **5. Входные требования для освоения дисциплины**

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

## **6. Язык реализации**

Русский

## **7. Объем дисциплины**

Общая трудоемкость дисциплины составляет 4 з.е., 144 часов, из которых:

-лекции: 6 ч.

-практические занятия: 20 ч.

в том числе практическая подготовка: 20 ч.

Объем самостоятельной работы студента определен учебным планом.

## **8. Содержание дисциплины, структурированное по темам**

Тема 1. Введение в статистический анализ данных.

Типы данных. Этапы статистического анализа. Статистические методы машинного обучения. Задачи обучения с учителем и без учителя. Предварительная обработка данных.

Тема 2. Выявление взаимосвязей.

Критерии сравнения групп. Дисперсионный анализ. Корреляционный анализ.

Тема 3. Задачи обучения с учителем.

Регрессионный анализ. Постановка задачи регрессии. Оценка параметров. Проверка адекватности. Задачи классификации. Построение модели и оценка параметров. Проверка качества классификатора.

Тема 4. Задачи обучения без учителя.

Кластерный анализ. Задачи снижения размерности.

## **9. Текущий контроль по дисциплине**

Текущий контроль по дисциплине проводится путем контроля посещаемости, тестов по лекционному материалу, выполнения итогового проекта и фиксируется в форме контрольной точки не менее одного раза в семестре.

## **10. Порядок проведения и критерии оценивания промежуточной аттестации**

Экзамен во втором семестре проводится в форме защиты исследовательского проекта и теста, включающего в себя как вопросы по теории, так и решение небольших практических задач.

Тест состоит из 15 вопросов разной сложности, за каждый из которых можно набрать от 1 до 3 баллов. Максимум за тест 30 баллов.

Баллы	Оценка
[26,30]	Отлично
[21,26)	Хорошо
[16,21)	Удовлетворительно
[0,16)	Неудовлетворительно

### **Примерные тестовые задания:**

- Для двух порядковых переменных при расчете коэффициента Спирмена были получены следующие результаты :  $r=0,75698236$        $p=0,0681236985$ .      Какой вывод можно сделать при уровне значимости 0.05?
- Для зависимых выборок по критерию Вилкоксона были получены следующие результаты:  $T=36,5$   $p=0,076522$ . Какой вывод можно сделать?

Исследовательский проект проверяет ИУК-1.1, ИОПК-2.2, ИОПК-6.1, ИОПК-6.2, ИПК-3.3.

## **11. Учебно-методическое обеспечение**

а) Электронный учебный курс по дисциплине в электронном университете «LMS IDO» - <https://lms.tsu.ru/course/view.php?id=29358>

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

в) Методические указания по организации самостоятельной работы студентов.

## **12. Перечень учебной литературы и ресурсов сети Интернет**

а) основная литература:

– Джеймс Г., Уиттон Д., Хасти Е., Тибширани Р., Введение в статистическое обучение с примерами на языке R. М.: ДМК Пресс, 2016 г., 450 с.

– Кабанова Т. В. Применение пакета R для решения задач прикладной статистики : учебное пособие : [для студентов и аспирантов университетов] / Т. В. Кабанова ; М-во образования и науки РФ, Нац. исслед. Том. гос. ун-т. - Томск : Издательский Дом Томского государственного университета, 2019. - 123 с.: ил., табл..

URL:<http://vital.lib.tsu.ru/vital/access/manager/Repository/vt1s:000668036>

б) дополнительная литература:

– Кендалл М. Д. Статистические выводы и связи / М. Кендалл, А. Стьюарт; Пер. с англ. Л. И. Гальчука, А. Т. Терехина; Под ред. А. Н. Колмогорова. - М. : Наука. Физматлит, 1973. - 899, [1] с.: ил..

URL: <http://sun.tsu.ru/limit/2016/000074332/000074332.djvu>

в) ресурсы сети Интернет:

– открытые онлайн-курсы  
– machine learning repository <https://archive.ics.uci.edu/ml/index.php>;  
– <https://www.kaggle.com/>.

## **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>

## **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.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

## **15. Информация о разработчиках**

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