

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

Центр сопровождения образовательных инициативных проектов

УТВЕРЖДЕНО:  
Руководитель сетевой ОПОП

З.И. Резанова

Рабочая программа дисциплины

**Статистические методы в гуманитарных исследованиях**

по направлению подготовки

**45.04.03 Фундаментальная и прикладная лингвистика**

Направленность (профиль) подготовки :  
**Анализ естественного языка (NLP) в лингвистике и IT**

Форма обучения  
**Очная**

Квалификация  
**Магистр**

Год приема  
**2023**

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

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий и актуальные концепции в области лингвистики.

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

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

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.2 Решает профессиональные задачи применяя основные понятия, категории и положения лингвистических теорий.

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

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.
- Владеть библиотеками языка Python, а также другими пакетами программ для статистической обработки данных.

## **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**

Семестр 1, 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 часа, из которых:

-лекции: 18 ч.

-практические занятия: 20 ч.

Объем самостоятельной работы студента определен учебным планом.

## 8. Содержание дисциплины, структурированное по темам

### Тема 1. Статистика. Введение

- Введение в статистический анализ
- Типы данных
- Генеральная и выборочная совокупности. Способы представления выборок: табличные и графические
- Точечные оценки параметров. Основные числовые характеристики. Интервальное оценивание
- Проверка статистических гипотез. Алгоритм проверки. p-value
- Практика по числовым характеристикам

### Тема 2. Критерии сравнения групп

- Параметрические критерии сравнения групп. t-test, Fisher, ANOVA
- Непараметрические критерии сравнения групп
- Практика по критериям сравнения групп

### Тема 3. Корреляционный анализ

- Корреляционный анализ. Постановка задачи. Корреляционный анализ количественных данных
- Ранговая корреляция
- Корреляционный анализ категоризованных данных. Таблицы сопряженности
- Практика по корреляционному анализу

### Тема 4. Регрессионный анализ

- Парная регрессия. Постановка задачи. Оценка параметров. Проверка качества
- Множественная регрессия
- Практика по регрессионному анализу

### Тема 5. Задача классификации

- Постановка задачи классификации. Бинарный классификатор
- Построение и проверка качества бинарного классификатора
- Практика по классификации или теория по кластеризации

## 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 семестр.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» – <https://www.tsu.ru/sveden/education/eduop/>.

## 10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в первом семестре проводится в форме 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

**Примерные тестовые задания:**

1. 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?
2. Для зависимых выборок по критерию Вилкоксона были получены следующие результаты:  $T=36,5$   $p=0,076522$ . Какой вывод можно сделать?

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» – <https://www.tsu.ru/sveden/education/eduop/>.

**11. Учебно-методическое обеспечение**

- а) Электронный учебный курс по дисциплине на интерактивной платформе «Skillfactory CS».
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

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

- а) основная литература:
  - Джеймс Г., Уиттон Д., Хасты 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/vtls: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**

Каждый обучающийся обеспечен доступом к интерактивной платформе «Skillfactory CS» <https://apps.skillfactory.ru/learning/login>.

#### **15. Информация о разработчиках**

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