

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

Механико-математический факультет

УТВЕРЖДАЮ:

Декан



Л.В. Гензе

« 30 » 06 20 22 г.

Рабочая программа дисциплины

**Статистический анализ данных**  
по направлению подготовки

**01.03.01 Математика, 02.03.01 Математика и компьютерные науки**

Направленность (профиль) подготовки:

**Основы научно-исследовательской деятельности в области математики**  
**Основы научно-исследовательской деятельности в области математики и**  
**компьютерных наук**

Форма обучения

**Очная**

Квалификация

**Бакалавр**


Год приема

**2021**

Код дисциплины в учебном плане: Б1.В.3.ДВ.01.04

СОГЛАСОВАНО:

Руководитель ОП



Л.В. Гензе

Председатель УМК



Е.А.Тарасов

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

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 Способен проводить под научным руководством исследование на основе существующих методов в конкретной области профессиональной деятельности.

ПК-1 Способен проводить научно-исследовательские разработки по отдельным разделам выбранной темы.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК 4.1 Проводит поиск и обработку научной и научно-технической информации, необходимой для решения исследовательских задач

ИОПК 4.2 Оценивает полученные результаты и формулирует выводы по итогам проведенных исследований

ИПК 1.1 Проводит работы по обработке и анализу научно-технической информации и результатов исследований

ИПК 1.2 Подготавливает планы и программы проведения отдельных этапов научно-исследовательской работы

ИПК 1.3 Проводит отдельные этапы научно-исследовательской работы

## **2. Задачи освоения дисциплины**

– Освоить основные методы регрессионного анализа при обработке статистических данных;

– Сформировать практические навыки использования изученных методов для обработки данных с помощью Python, Matlab, R;

– Сформировать навыки интерпретации полученных на практике результатов для задач регрессионного анализа.

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

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

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

Седьмой семестр, зачет с оценкой

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

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: математический анализ, функциональный анализ, теория вероятностей, математическая статистика, дополнительные главы теории вероятностей и математической статистики, математическое моделирование.

## **6. Язык реализации**

Русский

## **7. Объем дисциплины**

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

-лекции: 36 ч, практики 36 ч.

Объем самостоятельной работы студента определен учебным планом.

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

Тема 1. Линейные регрессионные модели.

Модель парной линейной регрессии. Модель множественной линейной регрессии. Оценивание параметров модели МНК. Мультиколлинейность. Линейная модель для случая зависимых факторов. Модель ридж-регрессии и Лассо.

Тема 2. Робастные модели регрессии

М-оценки. Метод наименьшей медианы квадратов. Сглаживание данных при помощи метода скользящей медианы.

Тема 3. Бинарная регрессия.

Модель логистической регрессии. Пробит-модель. Оценивание параметров моделей ММП, проверка гипотез о значимости модели. Меры адекватности. ROC-кривая, AUC.

Тема 4. Байесовский подход к статистическому оцениванию параметров.

Априорные распределенные, сопряженные с генеральной совокупностью. САЗ – распределения. Переход от априорного распределения параметров к апостериорному с помощью формулы Байеса. Байесовский прогноз отклика для модели линейной регрессии в предположении нормальности остатков.

Тема 5. Бутстраппинг.

Рецентрирование. Бутстраповская корректировка смещения. Ресэмплинг в регрессии.

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

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольной работы, выполнения практических заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

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

Зачет проводится в письменной форме с учетом работы студента в семестре. Билет содержит один теоретический вопрос.

Результаты дифференцированного зачета определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Студенту ставится оценка

- «удовлетворительно» ставится, если студент выполнил все задания на минимальную оценку «удовлетворительно» и ответил на теоретический вопрос в билете, понимая суть вопроса, но со значительными пробелами в выводах;
- «хорошо», если все задачи выполнены в среднем на оценку «хорошо» и студент ответил на теоретический вопрос, но в ответе были мелкие неточности и не ответил на дополнительные вопросы, то есть ответ был неполным;
- «отлично» ставится, если студент выполнил все практические задания на оценку «отлично» и ответил на теоретический вопрос достаточно развернуто и без ошибок.

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

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=33784>

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

в) План лекционных и практических занятий по дисциплине.

г) Электронный файл с конспектами лекций.

## 12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Боровков А.А. Математическая статистика. СПб.: Лань, 2016, 703 с.

- Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику, М.: Изд-во ЛКИ, 2015, 599 с.

- Chibara L., Hesterberg T. Mathematical statistics with resampling and R. Singapor: Wiley, 2011, 434 p.

- Норман Р. Дрейпер Прикладной регрессионный анализ: Книга 1. М.: Книга по Требованию, 2021, 364 с.

б) дополнительная литература:

- Efron B., Tibshirani R.J. An introduction to the bootstrap. N.Y.: Chapman & Hall, 1993

- Amemiya T. Advanced Econometrics, Cambridge: Harvard University Press, 1985.

- Greene W. H. Econometric Analysis, 5th edition. New Jearsey: Pearson Education, 2003

- Long J. S. Regression models for categorial and limited dependent variables, Thousand Oaks: Sage Publ., 1997.

в) ресурсы сети Интернет:

– [http://www.machinelearning.ru/wiki/images/1/17/Psad\\_linreg.pdf](http://www.machinelearning.ru/wiki/images/1/17/Psad_linreg.pdf)

## 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;

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

– Mathlab, Python;

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/>

## 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.

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

1. Пчелинцев Евгений Анатольевич, доцент, кандидат физико-математических наук, доцент кафедры математического анализа и теории функций ММФ ТГУ
2. Губин Владимир Николаевич, кандидат физико-математических наук, доцент кафедры математического анализа и теории функций ММФ ТГУ