

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ



**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «РГГУ»)**

Институт лингвистики

**УНЦ компьютерной лингвистики**

Рабочая программа дисциплины

**«Статистические методы в лингвистике»**

**Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика**

**Магистерская программа: Фундаментальная и компьютерная лингвистика**

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

**Москва 2021**

**Применение статистических пакетов при решении лингвистических задач**  
**Рабочая программа дисциплины**

**Составитель:**

**И. П. Рыгаев**

**Ответственный редактор:**

**д. филол. н., профессор В.И.Подлеская**

**УТВЕРЖДЕНО**

Протокол заседания УНЦ компьютерной  
лингвистики

**№ 6 от «9» апреля 2021г.**

## **ОГЛАВЛЕНИЕ**

### **1. Пояснительная записка**

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

1.3. Место дисциплины в структуре образовательной программы

### **2. Структура дисциплины**

### **3. Содержание дисциплины**

### **4. Образовательные технологии**

### **5. Оценка планируемых результатов обучения**

5.1. Система оценивания

5.2. Критерии выставления оценок

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

### **6. Учебно-методическое и информационное обеспечение дисциплины**

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

### **7. Материально-техническое обеспечение дисциплины**

### **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

## **Приложения**

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

## Пояснительная записка

### 1.1 Предмет

Предметом дисциплины «Статистические методы в лингвистике» является ознакомление слушателей с основами теории вероятностей и её применением в лингвистике.

Курс направлен на решение следующих задач:

- освоение магистрантами базовых знаний в области теории вероятности;
- выработка у магистрантов умения оперировать случайными величинами, находить их параметры, вычислять вероятности и т. д.
- приобретение навыков применения вероятностных методов обработки и моделирования лингвистических данных.

### 1.2 Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
ПК-2 Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	ПК-2.1	Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных

### 1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Статистические методы в лингвистике» является дисциплиной по выбору цикла дисциплин ООП ВПО (магистратуры) и относится к части, формируемой участниками образовательных отношений.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: Введение в компьютерную лингвистику, Математические методы в лингвистике, Вероятностные методы в лингвистике.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 76ч., в том числе контактная работа обучающихся с преподавателем 20ч., промежуточная аттестация 18ч., самостоятельная работа обучающихся 38ч.

№ п/п	Раздел Дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				лекции	семинары	самостоятельная работа	промежуточная аттестация	
1.	Введение. Основные понятия теории вероятностей.	3	1		2	4		
2.	Теоремы сложения и умножения вероятностей.	3	2		2	4		
3.	Совместные и несовместные события. Формулы Байеса.	3	3		2	4		
4.	Виды случайных величин – дискретные и непрерывные.	3	4		2	4		
5.	Математическое ожидание и дисперсия.	3	5		2	4		
6.	Функция распределения вероятностей случайной величины.	3	6		2	4		
7.	Плотность распределения вероятностей случайной величины.	3	7		2	4		
8.	Нормальное распределение.	3	8		2	4		
9.	Показательное распределение.	3	9		2	3		
10.	Система двух	3	10		2	3		

	случайных величин.							
11.	Экзамен	3	11				18	Зачет по контрольным вопросам
12.	Итого:				20	38	18	

### 3. Содержание дисциплины

#### 1.1. Введение. Основные понятия теории вероятностей.

Испытания и события. Классическое определение вероятности. Основные формулы комбинаторики. Относительная частота. Статистическая вероятность. Геометрические вероятности.

#### 1.2. Теоремы сложения и умножения вероятностей.

Теорема сложения вероятностей несовместных событий. Полная группа событий. Противоположные события. Произведение событий. Условная вероятность. Теорема умножения вероятностей.

#### 1.3. Совместные и несовместные события. Формулы Байеса.

Теорема сложения вероятностей совместных событий. Формула полной вероятности. Вероятность гипотез. Формулы Байеса.

#### 1.4. Виды случайных величин – дискретные и непрерывные.

Случайная величина. Дискретные и непрерывные случайные величины. Закон распределения вероятностей дискретной случайной величины. Биномиальное распределение. Распределение Пуассона.

#### 1.5. Математическое ожидание и дисперсия.

Числовые характеристики случайных величин. Математическое ожидание случайной величины. Вероятностный смысл математического ожидания. Свойства математического ожидания. Дисперсия дискретной случайной величины. Формула для вычисления дисперсии. Свойства дисперсии. Среднеквадратичное отклонение.

#### 1.6. Функция распределения вероятностей случайной величины.

Определение функции распределения. Свойства функции распределения. График функции распределения.

#### 1.7. Плотность распределения вероятностей случайной величины.

Определение плотности распределения. Вероятность попадания непрерывной случайной величины в заданный интервал. Нахождение функции распределения по известной плотности распределения. Свойства плотности распределения.

#### 1.8. Нормальное распределение.

Кривая нормального распределения. Влияние параметров нормального распределения на форму нормальной кривой. Вычисление вероятности заданного отклонения. Правило трех сигм. Центральная предельная теорема. Распределение «хи квадрат». Распределение Стьюдента. Распределение Фишера – Снедекора.

#### 1.9. Показательное распределение.

Определение показательного распределения. Вероятность попадания в заданный интервал показательного распределенной случайной величины. Числовые характеристики показательного распределения. Функция надежности. Показательный закон надежности.

1.10. Система двух случайных величин.

Система случайных величин. Функция распределения двумерной случайной величины. Плотность совместного распределения. Коэффициент корреляции. Линейная регрессия. Линейная корреляция. Нормальная корреляция.

#### 4. Образовательные технологии

Интерактивные формы обучения в данном курсе предполагают:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с критическим разбором реферируемого научного сочинения);
2. он-лайн демонстрации работы с лингвистическими базами данных и энциклопедическими интернет-ресурсами (информационно-справочный ресурс по языкам мира Etnolog <http://www.ethnologue.com> (частично-платный ресурс, социолингвистическая информация, ареал распространения, карты); типологическая база данных WALS, The World Atlas of Language Structures <http://wals.info>; информационно-справочный ресурс по языкам мира <http://glottolog.org> (генеалогическая аффилиация и обширная библиография); многоязычная платформа для составления конкордансов и исследования грамматической и лексической дистрибуции WordSketchEngine <https://www.sketchengine.co.uk/>; лексикологическая база данных WordNet (<http://wordnet.princeton.edu/>; электронная энциклопедия по лингвистике серии Oxford Research Encyclopedia <http://linguistics.oxfordre.com/>; и др.;
3. использование открытых он-лайн аудио- и видео обучающих ресурсов (для академического английского: [www.cambridge.org/elt/english-for-academics](http://www.cambridge.org/elt/english-for-academics));
4. практическую работу с электронными анализаторами речи и платформами лингвистического документирования (платформы E-Language Archiving Technology, ELAN <http://www.lat-mpi.eu/tools/elan>; коллекция электронных ресурсов на сайте Summer Institute of Linguistics ([www.sil.org](http://www.sil.org)), в том числе, программа для акустического анализа речи Speech analyzer, программа IPA Help для обучения и пользования Международной фонетической транскрипцией, комплекс программ для полевой работы; программа для акустического анализа речи PRAAT (<http://www.fon.hum.uva.nl/praat>); открытое он-лайн издание по документированию и консервации языков Language Documentation & Conservation, LD&C <http://www.nflrc.hawaii.edu/ldc>).
5. В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:
  - видео-лекции;
  - онлайн-лекции в режиме реального времени;
  - электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
  - системы для электронного тестирования;
  - консультации с использованием телекоммуникационных средств.

## 5. Оценка планируемых результатов обучения

### 5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,
- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдает зачет по всему курсу и предьявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

### 5.2. Критерии выставления оценок

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»/ «зачтено (отлично)»/ «зачтено»	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе.  Оценка по дисциплине выставляется обучающемуся с



Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>учёт результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

### ***5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации***

Зачёт ориентирован на следующие контрольные вопросы

- Вычислить вероятность случайного события с использованием формул комбинаторики.
- Вычислить вероятность, что произойдёт хотя бы одно из нескольких событий и что все они произойдут совместно.
- Вычислить вероятность события при условии другого события.
- Написать закон распределения случайной величины в серии из нескольких испытаний.
- Найти математическое ожидание и дисперсию случайной величины, зная закон распределения.
- Зная функцию или плотность распределения, найти вероятность, что результат испытания окажется в определённом интервале.
- Найти функцию распределения по плотности распределения случайной величины.
- Найти математическое ожидание и дисперсию нормально распределённой случайной величины.
- Найти вероятность безотказной работы элемента в течение заданного времени, зная показательный закон распределения вероятности отказа.

## **6. Учебно-методическое и информационное обеспечение дисциплины**

### ***6.1. Список литературы***

Основная литература

1. Гмурман, В. Е. (2007). Теория вероятности и математическая статистика: учебное пособие для высших учебных заведений. М.: Высшее образование.
2. Кочетков Е.С. Теория вероятностей и математическая статистика: Учебник. - 2; испр. и перераб. - Москва; Москва: Издательство "ФОРУМ": ООО "Научно-издательский центр ИНФРА-М", 2018. - 240 с.
3. Белько И.В. Теория вероятностей, математическая статистика, математическое программирование: Учебное пособие. - 1. - Москва; Минск: ООО "Научно издательский центр ИНФРА-М": ООО "Новое знание", 2016. - 299 с.

Рекомендованная литература

1. Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика.

2. Тутубалин В. Н. Теория вероятностей и случайных процессов. М.: Издательство МГУ, 1992.

### **6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

### **7. Материально-техническое обеспечение дисциплины**

Занятия по курсу можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимые программы:

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

### **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья**

В ходе реализации дисциплины используются следующие дополнительные

методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
  - письменные задания выполняются на компьютере в письменной форме;
  - экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;
  - в форме электронного документа;
  - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - передвижными, регулируемые эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.

## АННОТАЦИЯ ДИСЦИПЛИНЫ

Дисциплина реализуется в Институте лингвистики УНЦ компьютерной лингвистики.

Цель дисциплины:

Предметом дисциплины «Статистические методы в лингвистике» является ознакомление слушателей с основами теории вероятностей и её применением в лингвистике.

Задачи дисциплины:

- освоение магистрантами базовых знаний в области теории вероятности;
- выработка у магистрантов умения оперировать случайными величинами, находить их параметры, вычислять вероятности и т. д.
- приобретение навыков применения вероятностных методов обработки и моделирования лингвистических данных.

Дисциплина направлена на формирование следующих компетенций:

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
ПК-2 Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	ПК-2.1	Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных

По дисциплине предусмотрена промежуточная аттестация в форме экзамен.  
Общая трудоемкость освоения дисциплины составляет 2 зачетных единиц

**ЛИСТ ИЗМЕНЕНИЙ**

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №		