

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)**

Отделение интеллектуальных систем в гуманитарной сфере
Кафедра математики, логики и интеллектуальных систем в гуманитарной сфере

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.03.03 Фундаментальная и прикладная лингвистика

Код и наименование направления подготовки/специальности

Фундаментальная и прикладная лингвистика

Наименование направленности (профиля)/ специализации

Уровень высшего образования: *бакалавриат*

Форма обучения: *Очная*

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2022

Математическая статистика
Рабочая программа дисциплины

Составитель:

д. ф-м. н., профессор МЛиИС М.Р. Пентус

УТВЕРЖДЕНО

Протокол заседания кафедры математики, логики и интеллектуальных систем в гуманитарной сфере

№_1__ от_____31.03.2022_____

ОГЛАВЛЕНИЕ

1. Пояснительная записка	4
1.1. Цель и задачи дисциплины	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	4
1.3. Место дисциплины в структуре образовательной программы	5
2. Структура дисциплины.....	5
3. Содержание дисциплины.....	6
4. Образовательные технологии	6
5. Оценка планируемых результатов обучения.....	7
5.1 Система оценивания	7
5.2 Критерии выставления оценки по дисциплине	8
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	9
6. Учебно-методическое и информационное обеспечение дисциплины.....	12
6.1 Список источников и литературы	12
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет» ... 12	
7. Материально-техническое обеспечение дисциплины	13
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	13
9. Методические материалы	14
9.1 Планы семинарских/ практических/ лабораторных занятий	14
9.2 Иные материалы	16
Приложение 1. Аннотация рабочей программы дисциплины	18

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель дисциплины: ознакомить студентов с математическими понятиями и средствами математической статистики, которые могут использоваться, в частности, при статистической обработке данных. Целью курса является также обучение слушателей стилю математического моделирования с использованием современных понятий и методов математической статистики.

Задачи:

- изучение теории и практики решения задач по математической статистике;
- приобретение навыков анализа данных методами математической статистики;
- развитие навыков применения изученного математического аппарата к решению практических задач.

В результате освоения дисциплины обучающийся должен:

Знать:

- основные понятия математической статистики;
- основные критерии проверки гипотез;
- основные методы статистического моделирования.

Уметь:

- использовать основные методы математической статистики;
- пользоваться программными средствами статистической обработки данных.

Владеть:

- методами описательной статистики;
- методами проверки статистических гипотез;
- начальными навыками корреляционного анализ и регрессионного анализа.

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

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ОПК-2. Владеет основами математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур	ОПК-2.1	Знает: основные определения и базовые факты теории множеств, комбинаторики, математической логики, теории алгоритмов, теории вероятностей, математической статистики и теории информации; наиболее подходящие для использования в лингвистике вероятностные модели и статистические методы; основные типы данных,

		операторы, стандартные функции одного из алгоритмических языков, имеющих практическое применение для обработки языковых данных.
	ОПК-2.2	Умеет: применять полученные знания при решении математических и лингвистических проблем в рамках теоретических и прикладных задач лингвистики; структурировать собственные рассуждения, анализировать логическую структуру рассуждений; доказывать основные теоремы изученных разделов математики; применять вероятностные модели для вычисления вероятности различных событий, определения степени достоверности выводов на основе ограниченных статистических данных.
	ОПК-2.3	Владеет: основными методами решения типичных задач теории множеств, комбинаторики, математической логики, теории алгоритмов, теории вероятностей, математической статистики и теории информации; навыками планирования, написания и отладки простых программ для обработки языковых данных на изученном алгоритмическом языке, использования основных функций соответствующей среды программирования.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Математическая статистика» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Понятийный аппарат математики», «Математическая логика», «Вероятностные модели».

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: «Компьютерная и корпусная лингвистика», «Программирование в лингвистике».

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часов.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
	Лекции	14
	Семинары	28
Всего:		42

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 66 академических часов.

3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1.	Описание статистических данных	Выборочный метод. Оценки математического ожидания и дисперсии. Свойства оценок. Понятие о доверительном интервале и доверительной вероятности. Доверительный интервал для математического ожидания нормального распределения.
2.	Критерии проверки гипотез	Проверка статистических гипотез. Ошибки первого и второго рода, уровень значимости. Критерии согласия.
3.	Статистическое моделирование	Корреляционный анализ. Линейная регрессия.

4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	Описание статистических данных	Лекция 1 Семинары 1-2 Лекция 2 Семинары 3-4 Лекция 3 Семинары 5-6	Вводная лекция-беседа. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Работа с электронным конспектом, электронным задачником и интернет-ресурсами. Консультирование и приём

		Самостоятельная работа	домашних заданий посредством электронной почты
2	Критерии проверки гипотез	Лекция 4 Семинары 7-8 Лекция 5 Семинары 9-10 Лекция 6 Семинары 11-12 Лекция 7 Семинары 13-14 Самостоятельная работа	Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Работа с электронным конспектом, электронным задачником и интернет-ресурсами. Консультирование и приём домашних заданий посредством электронной почты
3	Статистическое моделирование	Лекция 8 Семинары 15-16 Лекция 9 Семинары 17-18 Лекция 10 Семинары 19-20 Семинары 21-22 Самостоятельная работа	Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Теоретическая лекция. Практикум по решению задач. Работа с электронным конспектом, электронным задачником и интернет-ресурсами. Консультирование и приём домашних заданий посредством электронной почты

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

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- контрольная работа (раздел 1)	30 баллов	30 баллов
- контрольная работа (разделы 2-3)	30 баллов	30 баллов
Промежуточная аттестация – экзамен		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу 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

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>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.</p> <p>Свободно 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 – «высокий».</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
82-68/ С	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	удовлетво- рительно/ 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	неудовлет- ворительно/ 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 литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

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

Текущий контроль

При оценивании контрольной работы учитывается:

- полнота выполненной работы – задание выполнено не полностью и/или допущены две и более ошибки или три и более неточности (1-4 балла);
- обоснованность содержания и выводов работы – задание выполнено полностью, но обоснование содержания и выводов недостаточны, но рассуждения верны (5-8 баллов);
- работа выполнена полностью, в рассуждениях и обосновании нет пробелов или ошибок, возможна одна неточность (9-10 баллов).

Промежуточная аттестация (экзамен)

При проведении промежуточной аттестации студент должен ответить на 5 вопросов (два вопроса теоретического характера и три вопроса практического характера).

При оценивании ответа на вопрос теоретического характера учитывается:

- теоретическое содержание не освоено, знание материала носит фрагментарный характер, наличие грубых ошибок в ответе (1-2 балла);
- теоретическое содержание освоено частично, допущено не более двух-трех недочетов (3-6 баллов);
- теоретическое содержание освоено почти полностью, допущено не более одного-двух недочетов, но обучающийся смог бы их исправить самостоятельно (7-8 баллов);
- теоретическое содержание освоено полностью, ответ построен по собственному плану (9-10 баллов).

При оценивании ответа на вопрос практического характера учитывается:

- ответ содержит менее 20% правильного решения (1-2 балла);
- ответ содержит 21-89 % правильного решения (3-8 баллов);
- ответ содержит 90% и более правильного решения (9-10 баллов).

Образцы заданий для самостоятельного выполнения

1. Найти медиану и квартили выборки (90,2, 91,6, 90,8, 93,6, 106,2, 92,2).
2. Найти выбросы в выборке (91,2, 91,6, 90,8, 93,6, 106,2, 91,2).
3. Найти таблицу контингентности случайных величин, заданных данной таблицей.
4. Найти коэффициент корреляции Пирсона по данной таблице контингентности.
5. Найти выборочную дисперсию выборки (2, 0, 1, 2, 8, 1).
6. С помощью одностороннего критерия Стьюдента выяснить, разнятся ли скорости чтения на родном и иностранном языках. Для каждого из 10 дикторов даны средняя скорость чтения на родном языке и средняя скорость чтения на иностранном языке.
7. Двенадцать раз подбрасывается пять игральных кубиков. Какова вероятность того, что сумма очков, равная 9, выпадет не менее десяти раз?
8. По одностороннему критерию знаков вычислили значение $p = 0,06$.
Какое значение p может получиться при применении двустороннего критерия знаков?
(А. Меньше чем 0,06. Б. Ровно 0,06. В. Больше чем 0,06. Г. Возможны все вышеперчисленные случаи.)

Образцы заданий для контрольных работ

Контрольная работа № 1

Вариант 1

1. (5 баллов) Найти ранжировку выборки (18, 16, 14, 12, 14, 18, 18).
2. (10 баллов) Для шести предложений даны количество слов в предложении и средняя длина слова в предложении. Найти коэффициенты корреляции Пирсона и Спирмена.
3. (5 баллов) Даны количества однословных предлогов в нескольких языках. Найти квартили и выбросы.

4. (5 баллов) Найти выборочное среднее и выборочную дисперсию данной выборки.

Контрольная работа № 2

Вариант 1

1. (5 баллов) Используя непараметрический критерий, выяснить (с уровнем значимости 5%), есть ли в третьем классе разница между мальчиками и девочками в уровне знания английского языка. Даны средние баллы мальчиков и девочек третьих классов в шести школах.

2. (10 баллов) Используя параметрический критерий, выяснить (с уровнем значимости 1%), есть ли в пятом классе разница между мальчиками и девочками в уровне знания английского языка. Даны средние баллы мальчиков и девочек пятых классов в шести школах.

3. (5 баллов) Студент получает на экзамене 5 с вероятностью 20%; 4 – с вероятностью 40%; 3 – с вероятностью 30%; 2 – с вероятностью 10%. За время обучения он сдаёт 100 экзаменов. Найти вероятность того, что сумма набранных баллов будет больше 352 баллов, но меньше 397. Найти симметричный относительно среднего значения интервал, в который с вероятностью 0,95 попадёт сумма набранных баллов.

4. (5 баллов) Выяснить (с уровнем значимости 1%), одинакова ли доля слов с иноязычными буквами среди односложных и двухсложных слов эстонского языка. В корпусе 10000 вхождений слов, из них 20% односложные и 26% двухсложные. Хотя бы одну иноязычную букву содержат 30 вхождений односложных слов и 60 вхождений двухсложных слов..

Список теоретических вопросов, выносимых на промежуточную аттестацию (экзамен)

1. Генеральная совокупность, выборка (однородная выборка), объём выборки, признак, выборка признака, варианта выборки.

2. Зависимые выборки (связанные выборки, выборки парных измерений), статистика выборки.

3. Номинальная шкала (номинативная шкала, классификационная шкала), порядковая шкала (ранговая шкала), интервальная шкала, абсолютная шкала, ранг (средний ранг, срединный ранг), ранжировка (дробная ранжировка, система рангов).

4. Частота реализаций эксперимента, относительная частота реализаций эксперимента (частость), таблица распределения частот, таблица контингентности (таблица сопряжённости, факторная таблица), гистограмма частот, разряд гистограммы (интервал гистограммы).

5. Выборочная квантиль, выборочная медиана, нижний выборочный квартиль, верхний выборочный квартиль, межквартильное расстояние, выброс.

6. Выброс, робастный метод, диаграмма размаха (ящик с усами).

7. Выборочная оценка математического ожидания, несмещённая выборочная оценка дисперсии, "исправленное" выборочное среднеквадратическое отклонение.

8. Выборочная ковариация, выборочный коэффициент корреляции Пирсона, коэффициент ранговой корреляции Спирмена.

9. Нормальное распределение, распределение хи-квадрат.

10. Распределение Стьюдента, распределение Фишера.

11. Статистическая гипотеза, нулевая гипотеза (основная гипотеза), альтернативная гипотеза.

12. Статистический критерий, критерий значимости, ошибка первого рода, ошибка второго рода.

13. Уровень значимости статистического критерия, односторонний статистический критерий, двусторонний статистический критерий.
14. Параметрические критерии проверки гипотез, z-критерий Фишера для проверки равенства средних значений, t-критерий Стьюдента для проверки равенства средних значений.
15. Критерий хи-квадрат для проверки независимости по таблице контингентности, поправка Йейтса.
16. Непараметрические критерии проверки гипотез, точный критерий Фишера для проверки независимости по таблице контингентности (точный тест Фишера).
17. U-критерий Манна--Уитни для проверки равенства медиан (критерий Уилкоксона--Манна--Уитни, критерий числа инверсий).
18. Критерий знаков для проверки равенства медиан.
19. Линейная регрессия, независимая переменная (регрессор, предиктор), зависимая переменная (критериальная переменная, отклик).
20. Метод наименьших квадратов, предсказанное значение, остаток регрессии, среднеквадратическая ошибка предсказания.
21. Дисперсионный анализ (ANOVA).

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Литература

Основная

Кремер, Н. Ш. Математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — Москва : Издательство Юрайт, 2022. — 259 с. — (Высшее образование). — ISBN 978-5-534-01654-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536959>

Малугин, В. А. Математическая статистика : учебное пособие для вузов / В. А. Малугин. — Москва : Издательство Юрайт, 2022. — 218 с. — (Высшее образование). — ISBN 978-5-534-06965-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/540131>

Дополнительная

Хуснутдинов, Р. Ш. Математическая статистика: Учебное пособие / Хуснутдинов Р.Ш. - М.:НИЦ ИНФРА-М, 2019. - 205 с. (Высшее образование: Бакалавриат) (Обложка. КБС)ISBN 978-5-16-009520-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1002159>

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

<http://www.sfs.uni-tuebingen.de/hbaayen/publications/baayenCUPstats.pdf>

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером и проектором для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. SPSS Statistics 22

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

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.

9. Методические материалы

9.1 Планы семинарских/ практических/ лабораторных занятий

Тема 1. Описание статистических данных

Цель занятия: ознакомиться с основными понятиями и определениями математической статистики, научиться использовать эти понятия при решении задач.

Форма проведения – решение задач.

1. Найти выборочную дисперсию выборки (12, 0, 1, 4, 8, 11).
2. Найти ранжировку выборки (16, 16, 14, 12, 14, 18, 14).
3. Даны количества односложных наречий в нескольких языках. Найти квартили и выбросы.
4. Найти выборочное среднее и выборочную дисперсию выборки (89,2, 91,6, 90,8, 93,6, 106,2, 89,2).
5. Для шести предложений даны количество слов в предложении и средняя длина слова в предложении. Найти коэффициенты корреляции Пирсона и Спирмена.

Контрольные вопросы:

1. Генеральная совокупность, выборка (однородная выборка), объем выборки, признак, выборка признака, варианта выборки.
2. Зависимые выборки (связанные выборки, выборки парных измерений), статистика выборки.
3. Номинальная шкала (номинативная шкала, классификационная шкала), порядковая шкала (ранговая шкала), интервальная шкала, абсолютная шкала, ранг (средний ранг, срединный ранг), ранжировка (дробная ранжировка, система рангов).

4. Частота реализаций эксперимента, относительная частота реализаций эксперимента (частость), таблица распределения частот, таблица контингентности (таблица сопряжённости, факторная таблица), гистограмма частот, разряд гистограммы (интервал гистограммы).

5. Выборочная квантиль, выборочная медиана, нижний выборочный квартиль, верхний выборочный квартиль, межквартильное расстояние, выброс.

6. Выброс, робастный метод, диаграмма размаха (ящик с усами).

7. Выборочная оценка математического ожидания, несмещённая выборочная оценка дисперсии, "исправленное" выборочное среднеквадратическое отклонение.

8. Выборочная ковариация, выборочный коэффициент корреляции Пирсона, коэффициент ранговой корреляции Спирмена.

9. Нормальное распределение, распределение хи-квадрат.

Тема 2. Критерии проверки гипотез

Цель занятия: ознакомиться с основными статистическими критериями, научиться использовать их при решении практических задач.

Форма проведения – решение задач.

1. Монету кинули 100 раз, она упала одной стороной 60 раз. Можно ли на уровне доверия 0,9 говорить о ее нечестности?

2. Для проверки эффективности нового лекарства были отобраны две случайные группы по 15 человек, страдающих гриппом. При применении старого лекарства средний срок выздоровления составлял 11 дней с выборочной дисперсией $S_{01}^2 = 3$, при применении нового – срок выздоровления составил 8 дней с выборочной дисперсией $S_{02}^2 = 4$. Проверить на уровне 0,99 гипотезу о преимуществе нового лекарства.

3. Каждого из 100 студентов просили назвать любимый вид спорта. Результаты представлены в таблице:

Пол/спорт	Футбол	Баскетбол	Плавание	Бег	Теннис	Всего
Мужской	21	5	9	12	13	60
Женский	9	3	1	15	12	40
Всего	30	8	10	27	25	100

Требуется проверить гипотезу о том, зависят ли предпочтения тех или иных видов спорта от пола опрашиваемых.

Контрольные вопросы:

1. Распределение Стьюдента, распределение Фишера.

2. Статистическая гипотеза, нулевая гипотеза (основная гипотеза), альтернативная гипотеза.

3. Статистический критерий, критерий значимости, ошибка первого рода, ошибка второго рода.

4. Уровень значимости статистического критерия, односторонний статистический критерий, двусторонний статистический критерий.

5. Параметрические критерии проверки гипотез, z-критерий Фишера для проверки равенства средних значений, t-критерий Стьюдента для проверки равенства средних значений.

6. Критерий хи-квадрат для проверки независимости по таблице контингентности, поправка Йейтса.

7. Непараметрические критерии проверки гипотез, точный критерий Фишера для проверки независимости по таблице контингентности (точный тест Фишера).

8. U-критерий Манна--Уитни для проверки равенства медиан (критерий Уилкоксона--Манна--Уитни, критерий числа инверсий).

9. Критерий знаков для проверки равенства медиан.

Тема 3. Статистическое моделирование

Цель занятия: приобретение навыков анализа данных методами математической статистики.

Форма проведения – решение задач.

1. В таблице приведены данные о зависимости стоимости эксплуатации самолета Y (в млн руб.) от времени его эксплуатации X (лет). Найти коэффициент корреляции, проверить гипотезу о наличии/отсутствии связи, найти уравнение линейной регрессии.

X	1	2	3	4	5	6	7	8
Y	3	3,5	3,5	4	4	6	9	10

Контрольные вопросы:

1. Линейная регрессия, независимая переменная (регрессор, предиктор), зависимая переменная (критериальная переменная, отклик).

2. Метод наименьших квадратов, предсказанное значение, остаток регрессии, среднеквадратическая ошибка предсказания.

3. Дисперсионный анализ (ANOVA).

9.2 Иные материалы

Рекомендуемая литература для более глубокого освоения программы.

Основная

1. Гмурман, В. Е. «Теория вероятностей и математическая статистика»: Учеб. пособие — 12-е изд., перераб.- М.: Высшее образование, 2006.-479 с.

2. Гмурман, В. Е. «Руководство к решению задач по теории вероятностей и математической статистике»: Учеб. пособие — 11-е изд., перераб. — М.: Высшее образование, 2006.-404 с

3. Ивашев-Мусатов О.С. Теория вероятностей и математическая статистика. М.: ФИМА, 2003. 224с.

4. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. М.: ИНФРА-М, 2003. С. 3-190, 236-329.

Дополнительная

1. Кибзун А. И., Горяинова Е. Н., Наумов А. В. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. 3-е изд., перераб. и доп. М.: Физматлит, 2013. 232 с.
2. Кремер Н. Ш. Теория вероятностей и математическая статистика: Учебник для ВУЗов. — 2- изд., перераб. и доп.-М:ЮНИТИ-ДАНА, 2004. — 573 с.
3. Baayen R. H. Analyzing Linguistic Data: A Practical Introduction to Statistics using R. Cambridge University Press, 2008. 368 p.
4. Gries S. T. Quantitative Corpus Linguistics with R: A Practical Introduction. Routledge, 2009. 256 p.
5. Levshina N. How to do Linguistics with R: Data exploration and statistical analysis. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015. 454 p.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина реализуется кафедрой математики, логики и интеллектуальных систем в гуманитарной сфере.

Цель дисциплины: ознакомить студентов с математическими понятиями и средствами математической статистики, которые могут использоваться, в частности, при статистической обработке данных. Целью курса является также обучение слушателей стилю математического моделирования с использованием современных понятий и методов математической статистики.

Задачи:

- изучение теории и практики решения задач по математической статистике;
- приобретение навыков анализа данных методами математической статистики;
- развитие навыков применения изученного математического аппарата к решению практических задач.

Дисциплина направлена на 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.

В результате освоения дисциплины обучающийся должен:

Знать:

- основные понятия математической статистики;
- основные критерии проверки гипотез;
- основные методы статистического моделирования.

Уметь:

- использовать основные методы математической статистики;
- пользоваться программными средствами статистической обработки данных.

Владеть:

- методами описательной статистики;
- методами проверки статистических гипотез;
- начальными навыками корреляционного анализ и регрессионного анализа.

По дисциплине предусмотрена промежуточная аттестация в форме *экзамена*.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы.