

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 25.05.2026 13:55:43
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»**

Инженерная академия

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

01.04.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

DATA SCIENCE И ЦИФРОВАЯ ТРАНСФОРМАЦИЯ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Статистические методы анализа данных» входит в программу магистратуры «Data Science и цифровая трансформация» по направлению 01.04.02 «Прикладная математика и информатика» и изучается в 1 семестре 1 курса. Дисциплину реализует Кафедра механики и процессов управления. Дисциплина состоит из 8 разделов и 21 тема и направлена на изучение фундаментальных основ аппарата дескриптивного статистического анализа и доверительного оценивания; разбор основных методов решения типовых задач и знакомство с областью их применения в профессиональной деятельности.

Целью освоения дисциплины является формирование фундаментальных знаний и навыков применения методов решения задач, необходимых для профессиональной деятельности, повышение общего уровня математической грамотности студентов, чтобы научить их вычислять точечные оценки неизвестных параметров распределений и строить доверительные интервалы, дать студентам представление об основах корреляционного анализа.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Статистические методы анализа данных» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики	ОПК-1.1 Знает основные законы, положения и методы в области решения актуальных задач фундаментальной и прикладной математики;; ОПК-1.2 Умеет выявлять сущность задач фундаментальной и прикладной математики;; ОПК-1.3 Владеет инструментами для решения задач фундаментальной и прикладной математики.;
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК-2.1 Знает существующие математические методы решения прикладных задач;; ОПК-2.2 Владеет инструментами реализации новых математических методов решения прикладных задач;; ОПК-2.3 Владеет необходимыми знаниями, позволяющими совершенствовать существующие методы решения прикладных задач.;
ОПК-3	Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.1 Знает методы и подходы для разработки математических моделей и анализа получаемых решений в области профессиональной деятельности;; ОПК-3.2 Умеет применять и выбирать наиболее приемлемые методы и подходы для разработки математических моделей и анализа получаемых решений;; ОПК-3.3 Владеет инструментами разработки и анализа математических моделей объектов профессиональной деятельности.;
ПК-1	Способен разрабатывать новые методики выполнения аналитических работ	ПК-1.1 Знает основные методы и подходы к анализу данных;; ПК-1.2 Умеет применять известные методы и подходы для проведения анализа данных;; ПК-1.3 Владеет алгоритмами по разработке методик проведения аналитических работ в профессиональной области.;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Статистические методы анализа данных» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Статистические методы анализа данных».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики		Преддипломная практика; Geoinformation Systems and Applications;
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач		Искусственные нейронные сети (Глубокое обучение); Преддипломная практика;
ОПК-3	Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности		Преддипломная практика; Научно-исследовательская работа; Проектирование автоматизированных систем управления;
ПК-1	Способен разрабатывать новые методики выполнения аналитических работ		Преддипломная практика; Научно-исследовательская работа; Системы искусственного интеллекта; Искусственные нейронные сети (Глубокое обучение); <i>Динамика и управление космическими системами**</i> ; <i>Dynamics and Control of Space Systems**</i> ; Geoinformation Systems and Applications; Регулирование сбора, хранения, обработки и использования данных;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Статистические методы анализа данных» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			1
<i>Контактная работа, ак.ч.</i>	34		34
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	17		17
Практические/семинарские занятия (СЗ)	0		0
<i>Самостоятельная работа обучающихся, ак.ч.</i>	83		83
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Вероятностно-статистическая модель. Статистики. Выборочное среднее и дисперсия. Медиана и мода. Эмпирическая функция распределения. Гистограмма	1.1	Понятие выборки. Способы представления выборочных данных	Выборка как часть генеральной совокупности, отобранная для изучения её свойств. Требования к репрезентативности выборки: случайность отбора и достаточный объём для надёжных выводов. Способы представления выборочных данных: упорядоченный ряд с сортировкой по возрастанию, вариационный ряд с группировкой одинаковых значений, статистический ряд распределения с частотами и относительными частотами.	ЛК, ЛР
		1.2	Основные выборочные характеристики: выборочное среднее, выборочная дисперсия, медиана, мода	Выборочное среднее как мера центральной тенденции, вычисляемая как сумма всех значений, делённая на объём выборки. Выборочная дисперсия как мера разброса данных относительно среднего, вычисляемая через сумму квадратов отклонений. Медиана как значение, делящее упорядоченную выборку пополам, устойчивое к выбросам. Мода как наиболее часто встречающееся значение в выборке, применимая для качественных данных.	ЛК, ЛР
		1.3	Эмпирическая функция распределения. Визуализация выборочных данных. Гистограмма. "Ящик с усами"	Эмпирическая функция распределения как ступенчатая функция, показывающая долю наблюдений, не превышающих заданное значение. Сходимость эмпирической функции распределения к теоретической по теореме Гливленко-Кантелли. Гистограмма как графическое представление распределения выборки с группировкой данных по интервалам и построением столбцов высотой, пропорциональной частоте или плотности частоты. Выбор числа интервалов по правилу Стёрджеса или формуле Фридмана-Диакониса. Ящик с усами как компактная визуализация пяти статистик: минимум, первый квартиль, медиана, третий квартиль, максимум. Выбросы на ящике с усами как точки за пределами полутора межквартильных размахов.	ЛК, ЛР
Раздел 2	Свойства статистик. Задача точечного оценивания. Метод максимального правдоподобия. Метод	2.1	Понятие дескриптивной статистики	Дескриптивная статистика как раздел статистики, занимающийся сбором, систематизацией и наглядным представлением данных без распространения выводов на генеральную совокупность. Отличие дескриптивной статистики	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
	моментов			от индуктивной: первая описывает имеющиеся данные, вторая делает выводы о всей совокупности. Основные задачи дескриптивной статистики: измерение центральной тенденции, измерение вариации, визуализация распределения.	
		2.2	Свойства выборочных оценок: состоятельность, несмещённость, эффективность	Состоятельность оценки как свойство сходиться по вероятности к истинному значению параметра при увеличении объёма выборки. Несмещённость оценки как равенство математического ожидания оценки истинному значению параметра при любом объёме выборки. Эффективность оценки как минимальная дисперсия среди всех несмещённых оценок параметра. Сравнение оценок по этим трём свойствам для выбора наилучшей.	ЛК, ЛР
		2.3	Методы нахождения точечных оценок неизвестных параметров распределений: метод максимального правдоподобия, метод моментов	Метод максимального правдоподобия как нахождение значений параметров, максимизирующих функцию правдоподобия выборки. Функция правдоподобия как совместная плотность вероятности наблюдений при заданных параметрах. Применение логарифмирования для упрощения максимизации. Метод моментов как приравнивание теоретических моментов распределения к соответствующим выборочным моментам. Решение полученной системы уравнений относительно неизвестных параметров.	ЛК, ЛР
Раздел 3	Точечное и доверительное оценивание математического ожидания нормального распределения при известной и неизвестной дисперсии	3.1	Определение понятия доверительного множества	Доверительное множество как интервал или область, покрывающая истинное значение параметра с заданной доверительной вероятностью. Доверительная вероятность как вероятность того, что построенный интервал накроет истинное значение параметра. Уровень значимости как дополнение доверительной вероятности до единицы, вероятность ошибки. Отличие доверительного интервала от точечной оценки: интервал даёт представление о точности оценивания.	ЛК, ЛР
		3.2	Вывод формул для границ доверительных интервалов для математического ожидания нормального распределения при известной и неизвестной дисперсии	Доверительный интервал для математического ожидания при известной дисперсии на основе нормального распределения выборочного среднего. Использование квантилей стандартного нормального распределения. Доверительный интервал для математического ожидания при неизвестной дисперсии на основе распределения Стьюдента. Замена неизвестного среднеквадратического отклонения его выборочной оценкой.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				Увеличение ширины доверительного интервала при неизвестной дисперсии по сравнению со случаем известной дисперсии.	
Раздел 4	Доверительный интервал для среднего квадратического отклонения нормального распределения при неизвестных параметрах. Распределение хи-квадрат. Распределение Стьюдента. Доверительный интервал для пропорции (доли) в случае больших выборок	4.1	Теоретическое обоснование формул для границ доверительного интервала для среднего квадратического отклонения нормального распределения (при неизвестных параметрах) и для пропорции (доли) в случае больших выборок	Доверительный интервал для среднего квадратического отклонения нормального распределения на основе распределения хи-квадрат. Использование квантилей распределения хи-квадрат с числом степеней свободы на единицу меньше объёма выборки. Доверительный интервал для пропорции доли в случае больших выборок на основе нормальной аппроксимации биномиального распределения. Условия применимости нормальной аппроксимации: достаточное число успехов и неудач.	ЛК, ЛР
		4.2	Свойства распределений Хи-квадрат и Стьюдента	Распределение хи-квадрат как распределение суммы квадратов независимых стандартных нормальных случайных величин. Параметр числа степеней свободы как количество суммируемых квадратов. Свойства распределения хи-квадрат: неотрицательность, асимметрия с убывающей к нулю при росте числа степеней свободы. Распределение Стьюдента как распределение отношения стандартной нормальной величины к квадратному корню из хи-квадрат, делённого на число степеней свободы. Симметричность распределения Стьюдента относительно нуля и сходимось к стандартному нормальному распределению при росте числа степеней свободы.	ЛК, ЛР
Раздел 5	Основные понятия теории проверки гипотез. Ошибки 1-го и 2-го рода. Р-значение. Проверка гипотез о параметрах нормального распределения.	5.1	Теоретические сведения о статистических гипотезах и релевантном понятийном аппарате.	Статистическая гипотеза как предположение о свойствах распределения генеральной совокупности. Нулевая гипотеза как основное проверяемое предположение, обычно об отсутствии эффекта. Альтернативная гипотеза как противостоящее нулевой предположение. Статистический критерий как правило принятия или отклонения нулевой гипотезы на основе выборочных данных. Ошибка первого рода как отклонение верной нулевой гипотезы. Ошибка второго рода как принятие неверной нулевой гипотезы. Уровень значимости как вероятность ошибки первого рода. Мощность критерия как вероятность верного отклонения неверной нулевой гипотезы. Р-значение как наименьший уровень значимости, при котором нулевая гипотеза может быть отклонена.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
		5.2	Критерии проверки гипотез о параметрах нормального распределения	Критерий для проверки гипотезы о равенстве математического ожидания заданному значению при известной дисперсии на основе нормального распределения статистики критерия. Критерий для проверки гипотезы о равенстве математического ожидания заданному значению при неизвестной дисперсии на основе распределения Стьюдента. Критерий для проверки гипотезы о равенстве дисперсии заданному значению на основе распределения хи-квадрат. Правила принятия решений через сравнение вычисленного значения статистики с критическим значением или через сравнение р-значения с уровнем значимости.	ЛК, ЛР
Раздел 6	Сравнение двух групп	6.1	Возможные постановки задачи о сравнении выборок в форме статистических гипотез	Постановка задачи сравнения двух выборок как проверка гипотезы о равенстве параметров распределений двух генеральных совокупностей. Сравнение независимых выборок от разных объектов и зависимых парных выборок от одних и тех же объектов до и после воздействия. Нулевая гипотеза об отсутствии различий между группами. Альтернативные гипотезы: двусторонняя о любом различии и односторонние о направленном различии.	ЛК, ЛР
		6.2	Критерии проверки гипотез о равенстве средних нормальных генеральных совокупностей при известных и неизвестных дисперсиях	Критерий для проверки гипотезы о равенстве средних двух нормальных совокупностей при известных дисперсиях на основе нормального распределения. Критерий при неизвестных, но равных дисперсиях на основе распределения Стьюдента с объединённой оценкой дисперсии. Критерий при неизвестных и неравных дисперсиях проблема Беренса-Фишера с приближённым распределением Стьюдента по формуле Сэттертвейта.	ЛК, ЛР
		6.3	Критерий Стьюдента	Критерий Стьюдента как основной инструмент сравнения средних двух групп. Формула вычисления t-статистики как отношения разности выборочных средних к стандартной ошибке этой разности. Число степеней свободы для независимых выборок как сумма объёмов выборок минус два. Условия применимости критерия Стьюдента: нормальность распределений и равенство дисперсий для двухвыборочного критерия. Парный критерий Стьюдента для зависимых выборок с переходом к разностям между парами наблюдений.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				Увеличение мощности парного критерия за счёт устранения межобъектной вариации.	
Раздел 7	Критерий Хи-квадрат. Проверка независимости признаков	7.1	Возможные постановки задачи о независимости количественных и качественных признаков (компонент двумерных случайных величин)	Задача о независимости двух признаков как проверка гипотезы о том, что распределение одного признака не зависит от значений другого. Количественные признаки с разбиением на интервальные категории для применения критерия хи-квадрат. Качественные номинальные признаки с произвольным числом категорий. Таблица сопряжённости как основная форма представления данных для анализа независимости.	ЛК, ЛР
		7.2	Теоретическое обоснование критерия Хи-квадрат	Критерий хи-квадрат как проверка гипотезы о независимости признаков на основе сравнения наблюдаемых и ожидаемых частот в таблице сопряжённости. Ожидаемые частоты при гипотезе независимости как произведение соответствующих маргинальных сумм, делённое на общий объём выборки. Статистика критерия как сумма квадратов отклонений наблюдаемых частот от ожидаемых, делённых на ожидаемые частоты. Распределение статистики как асимптотическое хи-квадрат с числом степеней свободы, равным произведению числа категорий одного признака минус один и числа категорий другого признака минус один.	ЛК, ЛР
		7.3	Мозаичная диаграмма как средство визуальной оценки правдоподобности гипотезы о независимости признаков (в случае качественной группировки)	Мозаичная диаграмма как графическое представление таблицы сопряжённости с площадями прямоугольников, пропорциональными частотам. Построение мозаичной диаграммы: последовательное деление области по уровням одного признака, затем по уровням другого. Визуальная оценка независимости признаков через сравнение долей категорий внутри каждого уровня. Отклонение от независимости проявляется как различие высоты или ширины прямоугольников для разных категорий.	ЛК, ЛР
Раздел 8	Непараметрические критерии проверки гипотез. Критерий Уилкоксона. Основы корреляционного анализа	8.1	Теоретическое обоснование применения непараметрических критериев проверки гипотез. Критерий Уилкоксона	Непараметрические критерии как методы проверки гипотез, не требующие предположения о виде распределения генеральной совокупности. Применение непараметрических критериев при малых выборках, отсутствии нормальности, наличии выбросов, порядковых данных. Критерий Уилкоксона для двух независимых выборок как ранговый аналог критерия Стьюдента. Алгоритм критерия Уилкоксона: объединение	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				выборки, ранжирование всех наблюдений, вычисление суммы рангов для меньшей выборки. Критерий Уилкоксона для парных выборок как критерий знаковых рангов на основе рангов абсолютных разностей.	
		8.2	Основные понятия корреляционного анализа: ковариация, коэффициент корреляции, их выборочные оценки	Ковариация как мера совместной изменчивости двух признаков, показывающая направление линейной связи. Выборочная ковариация как сумма произведений отклонений каждой пары от своих средних, делённая на объём выборки или объём минус один. Коэффициент корреляции как нормированная ковариация, приведённая к интервалу от минус единицы до плюс единицы. Выборочный коэффициент корреляции Пирсона как отношение выборочной ковариации к произведению выборочных среднеквадратических отклонений.	ЛК, ЛР
		8.3	Применение выборочного коэффициента корреляции для оценки степени зависимости признаков (компонент двумерной нормальной случайной величины)	Интерпретация коэффициента корреляции по абсолютной величине: близкая к единице означает сильную линейную связь, близкая к нулю – отсутствие линейной связи. Знак коэффициента указывает направление связи: положительный при прямой зависимости, отрицательный при обратной. Проверка гипотезы о значимости коэффициента корреляции на основе распределения Стьюдента. Отличие корреляции от причинно-следственной связи: корреляция не означает причинности. Ограничения коэффициента корреляции Пирсона: чувствительность к выбросам, измерение только линейной связи, требование нормальности распределений для проверки значимости.	ЛК, ЛР

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 3 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. - СПб.: Лань, 2011. - 256с
2. Ширяев А.Н., Эрлих И.Г., Яськов П.А. Вероятность в теоремах и задачах (с доказательствами и решениями). Книга 1. - М.: МЦНМО, 2013. - 648 с.
3. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов и др. - М.: НИЦ ИНФРА-М, 2015. - 890 с.

Дополнительная литература:

1. Боровков А.А. Математическая статистика. - СПб.: Лань, 2010. - 704 с
2. Свешников А.А. Прикладные методы теории вероятностей. - СПб.: Лань, 2012. - 480 с.
3. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистический анализ данных в MS Excel. - М.: ИНФРА-М, 2014. - 320 с.

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров
- Электронно-библиотечная система РУДН – ЭБС РУДН

<http://lib.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС Юрайт <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Троицкий мост»

2. Базы данных и поисковые системы

- электронный фонд правовой и нормативно-технической документации

<http://docs.cntd.ru/>

- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS

<http://www.elsevierscience.ru/products/scopus/>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Статистические методы анализа данных».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИКИ:

Доцент

Должность, БУП

Подпись

Салтыкова Ольга
Александровна

Фамилия И.О.

Доцент

Должность, БУП

Подпись

Демидов Александр
Сергеевич

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой

Должность БУП

Подпись

Разумный Юрий
Николаевич

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Профессор

Должность, БУП

Подпись

Разумный Юрий
Николаевич

Фамилия И.О.