Министерство просвещения Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Ульяновский государственный педагогический университет
имени И.Н. Ульянова»
(ФГБОУ ВО «УлГПУ им. И.Н. Ульянова»)

Факультет физико-математического и технологического образования
Кафедра высшей математики

УТВЕРЖДАЮ
Проректор по учебно-методической работе
_____ С.Н. Титов
«21» июня 2021 г.

РЕГРЕССИОННЫЙ, ДИСПЕРСИОННЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ

Программа учебной дисциплины модуля
«Математические методы обработки данных»

основной профессиональной образовательной программы высшего образования – программы магистратуры по направлению подготовки
44.04.01 Педагогическое образование,

направленность (профиль) образовательной программы
Компьютерное моделирование и дизайн информационной образовательной среды
(заочная форма обучения)

Составители: Цыганов А.В., доцент кафедры высшей математики
Кувшинова А.Н., старший преподаватель кафедры высшей математики
Голубков А.В., старший преподаватель кафедры высшей математики

Рассмотрено и одобрено на заседании ученого совета факультета физико-математического и технологического образования, протокол от «21» июня 2021г. №7

Ульяновск,2021

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 01D77C644FC1E54000000006381D0002
Владелец: Титов Сергей Николаевич
Действителен: с 19.07.2021 до 19.07.2022

Место дисциплины в структуре образовательной программы

Дисциплина «Регрессионный, дисперсионный и кластерный анализ данных» относится к дисциплинам вариативной части Блока 1. Дисциплины (модули) модуля «Математические методы обработки данных» учебного плана основной профессиональной образовательной программы высшего образования – программы магистратуры по направлению подготовки 44.04.01 Педагогическое образование, направленность (профиль) образовательной программы «Компьютерное моделирование и дизайн информационной образовательной среды», заочной формы обучения.

Дисциплина опирается на результаты обучения, сформированные в рамках курса магистратуры следующих дисциплин: математические методы обработки результатов экспериментальных исследований и проверки статистических гипотез, методы математического моделирования в предметных областях.

Результаты изучения дисциплины являются основой для изучения дисциплин для профессиональной деятельности и прохождения государственной итоговой аттестации.

1. Перечень планируемых результатов обучения (образовательных результатов) по дисциплине

Целью освоения дисциплины «Регрессионный, дисперсионный и кластерный анализ данных» является способствование развитию профессиональной компетентности будущего педагога через овладение методологией и технологией анализа статистических данных, регрессионного, дисперсионного и кластерного анализа.

Задачами освоения дисциплины является изучение базовых понятий и методов анализа статистических данных, теоретических основ регрессионного, дисперсионного и кластерного анализа; формирование практических навыков анализа статистических данных для профессиональной деятельности; расширение кругозора и формирование вероятностно-статистического мышления.

В результате освоения программы обучающийся должен овладеть следующими результатами обучения (в таблице представлено соотнесение образовательных результатов обучения по дисциплине с индикаторами достижения компетенций):

Компетенция и индикаторы её достижения в дисциплине	Образовательные результаты дисциплины (этапы формирования дисциплины)		
	знает	умеет	владеет
ПК-5 - Способен осуществлять поиск, анализ и обработку научной информации в целях исследования проблем образования в предметной области направленности (профиля) магистратуры «Компьютерное моделирование и дизайн информационной образовательной среды»			
ИПК 5.1. Знает источники научной информации, необходимой для обновления содержания образования по дисциплинам (курсам) предметной математической области направленности (профиля) магистратуры и трансформации процесса обучения; методы работы с научной информацией; приемы	ОР-1. источники научной информации. ОР-2. методы работы с научной информацией, приемы дидактической обработки научной информации в		

дидактической обработки научной информации в целях ее трансформации в учебное содержание.	целях ее трансформации в учебное содержание.		

2. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Номер семестра	Учебные занятия						Форма промежуточной аттестации
	Всего		Лекции, час	Практические занятия, час	Лабораторные занятия, час	Самостоятельная работа, час	
	Трудоемк.						
	Зач. ед.	Часы					
3	3	108	4	10	-	94	зачет
Итого:	3	108	4	10	-	94	зачет

3. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

3.1.Указание тем (разделов) и отведенного на них количества академических часов и видов учебных занятий

Наименование раздела и тем	Количество часов по формам организации обучения			
	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа
3 семестр				
Обзор методов анализа статистических данных. Дисперсионный анализ	1	3	-	20
Корреляционные зависимости. Понятие о корреляционном анализе	1	2	-	20
Регрессионный анализ	1	3	-	30
Кластерный анализ	1	2	-	24
ИТОГО:	4	10	-	94

3.2. Краткое описание содержания тем (разделов) дисциплины

Краткое содержание курса (3 семестр)

1. ОБЗОР МЕТОДОВ АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ. ДИСПЕРСИОННЫЙ АНАЛИЗ.

Статистические данные. Независимые и зависимые переменные. Математические методы, применяемые для анализа статистических данных: корреляционный анализ, дисперсионный анализ, регрессионный анализ, канонический анализ, кластерный анализ, факторный анализ, дискриминантный анализ.
Однофакторная дисперсионная модель (модель I, модель II). Проверка существенности влияния факторного признака (F-критерий). Двухфакторная дисперсионная модель. Проверка существенности каждого фактора и их взаимодействия на результативный признак
Интерактивная форма: работа в микрогруппах

2. КОРРЕЛЯЦИОННЫЕ ЗАВИСИМОСТИ. ПОНЯТИЕ О КОРРЕЛЯЦИОННОМ АНАЛИЗЕ.
Функциональная и статистическая зависимость. Корреляционная зависимость как вид статистической зависимости. Линейная парная регрессия. Коэффициент корреляции, его свойства. Двумерная корреляционная модель. Корреляционное отношение, индекс корреляции. Проверка значимости коэффициента корреляции. Интервальная оценка параметров связи. Частный коэффициент корреляции. Ранговая корреляция.
Интерактивная форма: работа в микрогруппах

3. РЕГРЕССИОННЫЙ АНАЛИЗ.
Задачи регрессионного анализа. Парная регрессионная модель. Основные предпосылки регрессионного анализа. Оценка функции регрессии. Интервальная оценка параметров парной модели. Критерий Фишера-Снедекора. Нелинейная регрессия. Ковариационная матрица. Доверительный интервал для коэффициентов и функции регрессии. Проверка значимости уравнения регрессии.
Интерактивная форма: работа в микрогруппах

4. КЛАСТЕРНЫЙ АНАЛИЗ
Постановка задачи кластеризации. Алгоритм кластеризации. Гипотеза «компактности». Типы входных данных. Методы объединения кластеров. Метрики для количественных и качественных шкал. Типы кластеров: типа сгущения, с центром и другие.
Интерактивная форма: работа в микрогруппах

4. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

Самостоятельная работа студентов является особой формой организации учебного процесса, представляющая собой планируемую, познавательно, организационно и методически направляемую деятельность студентов, ориентированную на достижение конкретного результата, осуществляемую без прямой помощи преподавателя. Самостоятельная работа студентов является составной частью учебной работы и имеет целью закрепление и углубление полученных знаний и навыков, поиск и приобретение новых знаний, а также выполнение учебных заданий, подготовку к предстоящим занятиям и экзамену. Она предусматривает, как правило, разработку рефератов, написание докладов, выполнение творческих, индивидуальных заданий в соответствии с учебной программой (тематическим планом изучения дисциплины). Тема для такого выступления может быть предложена преподавателем или избрана самим студентом, но материал выступления не должен дублировать лекционный материал. Реферативный материал служит дополнительной информацией для работы на практических занятиях. Основная цель данного вида работы состоит в обучении студентов методам самостоятельной работы с учебным материалом. Для полноты усвоения тем, вынесенных в практические занятия, требуется работа с первоисточниками. Курс предусматривает самостоятельную работу студентов со специальной литературой. Следует отметить, что самостоятельная работа студентов результативна лишь тогда, когда она выполняется систематически, планомерно и целенаправленно.

Задания для самостоятельной работы предусматривают использование необходимых терминов и понятий по проблематике курса. Они нацеливают на практическую работу по применению изучаемого материала, поиск библиографического материала и электронных

источников информации, иллюстративных материалов. Задания по самостоятельной работе даются по темам, которые требуют дополнительной проработки.

Общий объем самостоятельной работы студентов по дисциплине включает аудиторную и внеаудиторную самостоятельную работу студентов в течение семестра.

Аудиторная самостоятельная работа осуществляется в форме выполнения тестовых заданий, кейс-задач, письменных проверочных работ по дисциплине. Аудиторная самостоятельная работа обеспечена базой тестовых материалов, кейс-задач по разделам дисциплины.

Внеаудиторная самостоятельная работа осуществляется в формах:

- подготовки к устным докладам (мини-выступлениям);
- подготовка к защите реферата;
- подготовка к групповому обсуждению по темам;
- подготовка стендовых докладов и постеров;
- разработка проектов.

Материалы, используемые для текущего контроля успеваемости обучающихся по дисциплине

ОС-1. Самостоятельная работа
Примерные темы рефератов

1. Меры качества прогнозов.
2. Прогнозирование иерархических совокупностей рядов.
3. Устойчивая оценка дисперсии Уайта, ее модификации.
4. Неразрешимость парадокса Симпсона в рамках классической статистики.
5. Интервенции. Оценка эффекта по обсервационным данным.
6. Регрессия счетного признака. Пуассоновская модель.
7. Байесовская классификация.
8. Логистическая регрессия.
9. Регрессионный анализ в среде R.
10. Методы Data Mining.
11. Линейные регрессионные модели педагогических исследований.
12. Исследование степени рассеяния вариационных рядов.
13. Методы составления приближенного распределения признака, имеющего непрерывное распределение.
14. Установления связи между случайными величинами при большом числе наблюдений.
15. Определение параметров выборочного уравнения прямой линии регрессии по сгруппированным данным.

5. Примерные оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

Организация и проведение аттестации студента

ФГОС ВО в соответствии с принципами Болонского процесса ориентированы преимущественно не на сообщение обучающемуся комплекса теоретических знаний, но на выработку у бакалавра компетенций – динамического набора знаний, умений, навыков и личностных качеств, которые позволят выпускнику стать конкурентоспособным на рынке труда и успешно профессионально реализовываться.

В процессе оценки бакалавров необходимо используются как традиционные, так и инновационные типы, виды и формы контроля. При этом постепенно традиционные средства совершенствуются в русле компетентностного подхода, а инновационные средства адаптированы для повсеместного применения в российской вузовской практике.

Цель проведения аттестации – проверка освоения образовательной программы дисциплины-практикума через сформированность образовательных результатов.

Промежуточная аттестация осуществляется в конце семестра и завершает изучение дисциплины; помогает оценить крупные совокупности знаний и умений, формирование определенных компетенций.

Оценочными средствами текущего оценивания являются: доклад, тесты по теоретическим вопросам дисциплины, защита практических работ и т.п. Контроль усвоения материала ведется регулярно в течение всего семестра на практических (семинарских, лабораторных) занятиях.

№ п/п	**СРЕДСТВА ОЦЕНИВАНИЯ,** используемые для текущего оценивания показателя формирования компетенции	Образовательные результаты дисциплины
	Оценочные средства для текущей аттестации ОС-1. Самостоятельная работа	ОР-1. Знает источники научной информации. ОР-2. Знает методы работы с научной информацией, приемы дидактической обработки научной информации в целях ее трансформации в учебное содержание
	Оценочные средства для промежуточной аттестации **зачет (экзамен)** ОС-2. Зачет в форме устного собеседования	

Материалы, используемые для текущего контроля успеваемости обучающихся по дисциплине

Материалы для организации текущей аттестации представлены в п.5 программы.

Материалы, используемые для промежуточного контроля успеваемости обучающихся по дисциплине

Примерный перечень вопросов к зачету

1. Понятие о дисперсионном анализе. Однофакторная дисперсионная модель. Двухфакторная дисперсионная модель.
2. Линейная парная регрессия. Коэффициент регрессии.
3. Коэффициент корреляции и его свойства.
4. Понятие о корреляционном анализе.
5. Интервальная оценка параметров связи. Проверка значимости параметров связи.
6. Корреляционное отношение и индекс корреляции. Их свойства.
7. Множественная корреляция. Множественный и частный коэффициент корреляции.
8. Понятие о ранговой корреляции. Коэффициент ранговой корреляции Спирмена.
9. Коэффициент ранговой корреляции Кендалла и его свойства.
10. Коэффициент конкордации (согласованности) рангов Кендалла.
11. Понятие о регрессионном анализе. Парная регрессионная модель.
12. Интервальная оценка функции регрессии.
13. Проверка значимости уравнения регрессии. Интервальная оценка параметров парной регрессии.
14. Нелинейная регрессия.
15. Множественный регрессионный анализ.
16. Выборочная оценка ковариационной матрицы.
17. Доверительный интервал для коэффициентов и функции регрессии.
18. Проверка значимости уравнения множественной регрессии.
19. Понятие о кластерном анализе. Цели и задачи кластерного анализа. Основные принципы кластеризации.
20. Основные понятия кластерного анализа.
21. Метрики для количественных и качественных шкал.
22. Методы кластеризации.

Критерии оценивания знаний обучающихся по дисциплине

Формирование балльно-рейтинговой оценки работы обучающихся

		Посещение лекций	Посещение практических занятий	Работа на практических занятиях и текущий контроль	Зачет
3 семестр	Разбалловка по видам работ	2х1=2 баллов	5х1=5 баллов	229 баллов	64 балла
	Суммарный макс. балл	2 баллов max	7 баллов max	236 баллов Max	300 баллов max

Критерии оценивания работы обучающегося по итогам 3 семестра

Оценка	**Баллы (3 ЗЕ)**
«зачтено»	151-300
«не зачтено»	150 и менее

6. Методические указания для обучающихся по освоению дисциплины

Успешное изучение курса требует от обучающихся посещения лекций, активной работы на практических занятиях, выполнения всех учебных заданий преподавателя, ознакомления с основной и дополнительной литературой.

Запись **лекции** – одна из форм активной самостоятельной работы обучающихся, требующая навыков и умения кратко, схематично, последовательно и логично фиксировать основные положения, выводы, обобщения, формулировки. В конце лекции преподаватель оставляет время (5 минут) для того, чтобы обучающиеся имели возможность задать уточняющие вопросы по изучаемому материалу. Из-за недостаточного количества аудиторных часов некоторые темы не удается осветить в полном объеме, поэтому преподаватель, по своему усмотрению, некоторые вопросы выносит на самостоятельную работу студентов, рекомендуя ту или иную литературу. Кроме этого, для лучшего освоения материала и систематизации знаний по дисциплине, необходимо постоянно разбирать материалы лекций по конспектам и учебным пособиям. В случае необходимости обращаться к преподавателю за консультацией.

Подготовка к практическим занятиям.

При подготовке к практическим занятиям студент должен изучить теоретический материал по теме занятия (использовать конспект лекций, изучить основную литературу, ознакомиться с дополнительной литературой, при необходимости дополнить конспект, делая в нем соответствующие записи из литературных источников). В случае затруднений, возникающих при освоении теоретического материала, студенту следует обращаться за консультацией к преподавателю. Идя на консультацию, необходимо хорошо продумать вопросы, которые требуют разъяснения.

В начале практического занятия преподаватель знакомит студентов с темой, оглашает план проведения занятия, выдает задания. В течение отведенного времени на выполнение работы студент может обратиться к преподавателю за консультацией или разъяснениями. В конце занятия проводится прием выполненных заданий, собеседование со студентом.

Результаты выполнения практических зданий оцениваются в баллах, в соответствии с балльно-рейтинговой системой университета.

Планы практических занятий (3 семестр)

Практическое занятие №1.

1. Статистические данные.
2. Независимые и зависимые переменные.

3. Математические методы, применяемые для анализа статистических данных
4. Обсуждение тем, вынесенных на лекции

Практическое занятие №2.

1. Однофакторная дисперсионная модель (модель I, модель II).
2. Проверка существенности влияния факторного признака (F-критерий).
3. Двухфакторная дисперсионная модель.
4. Проверка существенности каждого фактора и их взаимодействия на результативный признак
5. Работа с моделями

Практическое занятие №3.

1. Функциональная и статистическая зависимость.
2. Корреляционная зависимость как вид статистической зависимости.
3. Линейная парная регрессия.
4. Коэффициент корреляции, его свойства.
5. Двумерная корреляционная модель.

Практическое занятие №4.

1. Корреляционное отношение, индекс корреляции.
2. Проверка значимости коэффициента корреляции.
3. Интервальная оценка параметров связи.
4. Частный коэффициент корреляции.
5. Ранговая корреляция.
6. Практическая работа решение задач

Практическое занятие №5.

1. Задачи регрессионного анализа.
2. Парная регрессионная модель.
3. Основные предпосылки регрессионного анализа.
4. Оценка функции регрессии.
5. Обсуждение тем, вынесенных на лекции

Практическое занятие №6.

1. Интервальная оценка параметров парной модели.
2. Критерий Фишера-Снедекора.
3. Нелинейная регрессия.

Практическое занятие №7.

1. Ковариационная матрица.
2. Доверительный интервал для коэффициентов и функции регрессии.
3. Проверка значимости уравнения регрессии.
4. Работа над задачами

Практическое занятие №8.

1. Постановка задачи кластеризации.
2. Алгоритм кластеризации.
3. Гипотеза «компактности».
4. Защита докладов

Практическое занятие №9.

1. Типы входных данных.
2. Методы объединения кластеров.
3. Метрики для количественных и качественных шкал.
4. Типы кластеров: типа сгущения, с центром и другие.
5. Защита докладов

Практическое занятие №10.

1. Защита докладов
2. Проведение зачета

7. Перечень основной и дополнительной учебной литературы, Интернет-ресурсов, необходимых для освоения дисциплины

Основная литература

1. Лисицин, Д.В. Методы построения регрессионных моделей : учебное пособие / Д.В. Лисицин ; Министерство образования и науки Российской Федерации, Новосибирский государственный технический университет. - Новосибирск : НГТУ, 2011. - 77 с. - ISBN 978-5-7782-1621-1 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=228876
2. Кибзун, А. И. Теория вероятностей и математическая статистика: Базовый курс с примерами и задачами : учебное пособие / А. И. Кибзун, Е. Р. Горяинова, А. В. Наумов ; ред. А. И. Кибзун. – 3-е изд., перераб. и доп. – Москва : Физматлит, 2007. – 232 с. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=69320 (дата обращения: 05.12.2021). – ISBN 978-5-9221-0836-2
3. Дорофеев, В.А. Основы регрессионного моделирования для психологов: учебное пособие по дисциплине «Математическая статистика и математические методы в психологии» / В.А. Дорофеев, Ю.А. Мочалова ; Министерство науки и высшего образования РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет». - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2018. - 130 с. : ил. - Библиогр.: с. 120-121. - ISBN 978-5-9275-2549-2 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=499592

Дополнительная литература

1. Эконометрика : учебник / К.В. Балдин, В.Н. Башлыков, Н.А. Брызгалов и др. ; под ред. В.Б. Уткина. - 2-е изд. - Москва : Издательско-торговая корпорация «Дашков и К°», 2017. - 562 с. : ил. - Библиогр.: с. 473-477. - ISBN 978-5-394-02145-9 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=452991.
2. Комарова, Е.С. Парный регрессионный анализ : учебное пособие / Е.С. Комарова. - Москва ; Берлин : Директ-Медиа, 2015. - 59 с. : ил. - Библиогр. в кн. - ISBN 978-5-4475-4493-5 ; То же [Электронный ресурс]. - URL: http://biblioclub.ru/index.php?page=book&id=278250

Интернет-ресурсы

1. https://www.intuit.ru/ - Национальный Открытый Университете «ИНТУИТ»