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

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0233C9D60038B01A9E48C1A3FBB4A9D9F6
Владелец: Титов Сергей Николаевич
Действителен: с 07.07.2023 до 28.04.2038

УТВЕРЖДАЮ
Проректор по учебно-методической
работе С.Н. Титов

РЕГРЕССИОННЫЙ, ДИСПЕРСИОННЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ ДАННЫХ

Программа учебной дисциплины вариативного 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
Методология математического образования

(очная форма обучения)

Составители:
Сибирева А.Р., к. ф.-м. н., доцент;
Макеева О.В., к.ф.-м.н.,
доцент кафедры высшей математики

Рассмотрено и одобрено на заседании ученого совета факультета физико-математического и технологического образования, протокол от 26 мая 2023 г. № 5.

Ульяновск, 2023

Место дисциплины в структуре образовательной программы

Дисциплина «Регрессионный, дисперсионный и кластерный анализ данных» относится к дисциплинам вариативного модуля «Математические методы обработки данных» Блока 1. *Дисциплины (модули)* части, формируемой участниками образовательных отношений, учебного плана основной профессиональной образовательной программы высшего образования – программы магистратуры по направлению подготовки 44.04.01 Педагогическое образование, направленность (профиль) образовательной программы «Методология математического образования», очной формы обучения.

Дисциплина опирается на результаты обучения, сформированные в рамках базовых математических дисциплин образовательных программ бакалавриата.

Результаты освоения дисциплины являются основой для научно-исследовательской работы магистрантов.

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

Актуальность включения дисциплины в ООП определяется необходимостью знакомства магистрантов с приложениями математики, в частности, математической статистики, в области психолого-педагогических и иных экспериментальных исследований, их подготовки к преподаванию линии анализа данных на различных уровнях образования.

Цель освоения дисциплины «*Регрессионный, дисперсионный и кластерный анализ данных*» - способствовать развитию профессиональной компетентности будущего педагога через овладение методологией и технологией анализа статистических данных, регрессионного, дисперсионного и кластерного анализа.

В соответствии с этим при преподавании дисциплины ставятся следующие **задачи**:

- 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;
- формирование математико-статистической грамотности и навыков анализа данных с помощью компьютерных сред.

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

Компетенция	Индикаторы достижения компетенции	Образовательные результаты дисциплины (этапы формирования компетенции)		
		теоретический	модельный	практический
ПК-5	**Способен осуществлять поиск, анализ и обработку научной информации в целях исследования проблем образования в предметной области направленности (профиля) магистратуры**			
	ПК.5.1	**Знает: источники научной информации, необходимой для обновления содержания образования по дисциплинам (курсам) предметной области направленности (профиля) магистратуры и трансформации процесса обучения; методы работы с научной информацией; приемы дидактической обработки научной информации в целях ее трансформации в учебное содержание**		
		Магистрант знает	Магистрант умеет	Магистрант владеет
		ОР-1: роль анализа статистических данных в научном исследовании; постановки задачи при различных методах анализа данных	ОР-3: проектировать структуру данных в исследовании, планировать способы получения и обработки данных, оценивать надежность данных, искать закономерности и связи между признаками	ОР-4: навыками изучения стохастических зависимостей, навыками статистической оценки влияния факторов на результат, навыками кластеризации
		ОР-2: основные методы анализа данных наблюдений и экспериментов; принципы регрессионного, корреляционного, дисперсионного анализа данных		
	ПК.5.2	**Умеет вести поиск и анализ научной информации; осуществлять дидактическую обработку и адаптацию научных текстов в целях их перевода в учебные материалы**		
		Магистрант знает	Магистрант умеет	Магистрант владеет
			ОР-5: осуществлять анализ данных, полученных в ходе исследования, изученными методами, включая методы выявления статистических зависимостей;	ОР-7: некоторыми компьютерными инструментами обработки статистических данных
			ОР-6: осваивать новые для себя методы анализа данных	

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	
	Трудоемкость						
	Зач. ед.	Часы					
3	3	108	4	20	-	84	Зачет
Итого:	3	108	4	20	-	84	

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

Отбор содержания дисциплины определяется необходимостью знакомства магистрантов с основными классическими и современными подходами к анализу данных. Основное внимание предлагается уделить базовым принципам статистической обработки информации, а также способам постановки проблем для статистического анализа и основам применения соответствующих инструментов, в том числе компьютерных сред.

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

Наименование разделов и тем	Количество часов по формам организации обучения			
	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа
3 семестр				
1. Введение. Обзор методов анализа статистических данных		4	-	12
2. Дисперсионный анализ	1	4	-	18
3. Корреляционные зависимости. Понятие о корреляционном анализе	1	4	-	18
4. Регрессионный анализ	1	4	-	18
5. Кластерный анализ	1	4	-	18
Всего по дисциплине:	4	20	-	84

3.2. Краткое описание содержания тем (разделов) дисциплины

1. Введение. Обзор методов анализа статистических данных.

Объект исследования. 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.

2. Дисперсионный 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.

3. **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.

4. **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.

5. **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.

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

Общий объем самостоятельной работы магистрантов по дисциплине включает аудиторную и внеаудиторную (78 часов) самостоятельную работу в течение семестра.
Аудиторная самостоятельная работа осуществляется преимущественно в малых группах, в рамках участия в интерактивных занятиях различных форм, предусмотренных программой (см. выше раздел 3.2). Внеаудиторная самостоятельная работа организуется как продолжение аудиторной и включает доработку консультационных материалов по отдельным темам в виде коллективной презентации; полное оформление мини-исследования, предполагающего анализ данных; подготовку к дискуссии и конференции; - подготовку проектов; подготовку к зачету (подробнее см. раздел 5.2).

Материалы, используемые для текущего контроля успеваемости обучающихся по дисциплине «Регрессионный, дисперсионный и кластерный анализ данных »

ОС-1. Домашняя контрольная работа

ОС-2.Примерные темы докладов и рефератов

ОС-3. Тест

ОС-4 Вопросы к зачету

Для самостоятельной подготовки к занятиям по дисциплине рекомендуется использовать учебно-методические материалы:

1. Владова Е.В. Теория вероятностей и математическая статистика [Текст]: / Е.В. Владова // Учебно-методическое пособие. – Ульяновск: УлГПУ, 2010. – 137 с. (Библиотека УЛГПУ).
2. Владова Е.В. Теория вероятностей и математическая статистика // Учебно-методическое пособие для бакалавров. – Ульяновск: УлГПУ им. И.Н. Ульянова, 2017. – 56 с.
3. Владова Е.В. Статистика: в 3-х частях. Часть 1: учебно-методическое пособие для бакалавров направления подготовки «Педагогическое образование». / Владова Е.В. – Ульяновск: ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», 2017. – 60 с.
4. Владова Е.В. Статистика: в 3-х частях. Часть 2: учебно-методическое пособие для бакалавров направления подготовки «Педагогическое образование»./ Владова Е.В.– Ульяновск: ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», 2017. – 59 с.
5. Владова Е.В. Статистика: в 3-х частях. Часть 3: учебно-методическое пособие для бакалавров направления подготовки «Педагогическое образование». / Владова Е.В. – Ульяновск: ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», 2017. – 45 с.
6. Стрюкова Г.А. Методы математической статистики в психолого-педагогических исследованиях: Учебно-методическое пособие / Г.А. Стрюкова. – Ульяновск: ФГБОУ ВО «УлГПУ им. И.Н. Ульянова», 2017. – 91 с.

7. Теория вероятностей и математическая статистика: учебно-методическое пособие для бакалавров и магистрантов направления подготовки «Педагогическое образование». Владова Е.В. – Ульяновск.УлГПУ им. И.Н. Ульянова. 2017 – 56 с.
8. Элементы теории вероятностей, математической статистики и анализа систем массового обслуживания. Часть 1. Введение в теорию вероятностей. Краткий исторический экскурс: учебное пособие для подготовки бакалавров и магистров нематематических направлений / сост. Н.А. Волкова, Н.В. Глухова. – Ульяновск: УлГПУ имени И.Н. Ульянова, 2017. – 96 с.

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

Цель проведения аттестации – диагностика сформированности запланированных образовательных результатов. Для проверки освоения программы дисциплины «Регрессионный, дисперсионный и кластерный анализ данных » используются технологии формирующего оценивания (текущая аттестация) и суммативного оценивания (промежуточная аттестация в форме зачёта, завершающая изучение дисциплины).

№ п/п	**СРЕДСТВА ОЦЕНИВАНИЯ,** используемые для текущего оценивания показателя формирования компетенции	Образовательные результаты дисциплины
	Оценочные средства для текущей аттестации ОС-1. Домашняя контрольная работа ОС-2.Примерные темы докладов и рефератов ОС-3. Тест	ОР-1: знает роль анализа статистических данных в научном исследовании; постановки задачи при различных методах анализа данных ОР-2: знает основные методы анализа данных наблюдений и экспериментов; принципы регрессионного, корреляционного, дисперсионного анализа данных ОР-3: умеет проектировать структуру данных в исследовании, планировать способы получения и обработки данных, оценивать надежность данных, искать закономерности и связи между признаками ОР-4: владеет навыками изучения стохастических зависимостей, навыками статистической оценки влияния факторов на результат, навыками кластеризации ОР-5: умеет осуществлять анализ данных, полученных в ходе исследования, изученными методами, включая методы выявления статистических зависимостей; ОР-6:умеет осваивать новые для себя методы анализа данных
	Оценочные средства для промежуточной аттестации **зачет (экзамен)** ОС-4 Зачет в форме устного собеседования	

		ОР-7: владеет некоторыми компьютерными инструментами обработки статистических данных

Описание оценочных средств и необходимого оборудования (демонстрационного материала), а также процедуры и критерии оценивания индикаторов достижения компетенций на различных этапах их формирования в процессе освоения образовательной программы представлены в Фонде оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине.

Материалы, используемые для промежуточного контроля успеваемости обучающихся по дисциплине

Программа зачёта по дисциплине
«*Регрессионный, дисперсионный и кластерный анализ данных*»

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.

ФОРМИРОВАНИЕ БАЛЛЬНО-РЕЙТИНГОВОЙ ОЦЕНКИ РАБОТЫ МАГИСТРАНТА

№ п/п	**Вид деятельности** (3 семестр)	**Максимальное количество баллов**
1.	Посещение лекционных занятий	2
2.	Посещение практических занятий	10

3.	Работа на занятиях и самостоятельная внеаудиторная работа, всего	224
4.	Зачёт	64
Итого	3 зачетных единицы	300

Критерии оценивания работы обучающегося по итогам семестра

По итогам освоения дисциплины «Регрессионный, дисперсионный и кластерный анализ данных», которая изучается в 3 семестре и трудоёмкость которой составляет 3 ЗЕ, магистрант набирает определённое количество баллов, которое соответствует результату «зачтено» или «не зачтено» согласно следующей таблице:

Итоговое количество баллов	Итоговая отметка
0-150	не зачтено
151-300	зачтено

6. Методические указания для обучающихся по освоению дисциплины

Дисциплина «*Регрессионный, дисперсионный и кластерный анализ данных*» нацелена на формирование представлений о характере и способах применения математико-статистических методов проектирования и обработки результатов педагогических наблюдений и экспериментов. Успешное освоение курса требует активной деятельности на практических занятиях и во внеаудиторное время, систематического выполнения домашних заданий, самостоятельной работы с математической и психолого-педагогической литературой, как учебной, так и научной.

Предполагается ознакомление студентов с методами математического исследования с учётом того, что эти методы невозможно изучать отдельно от соответствующих объектов. Основное время, отведенное на аудиторные занятия, предлагается посвятить именно формированию у будущих магистров основ умений, связанных с чётким выделением проблемы исследования и формулировкой гипотезы, которые допускали бы постановку математической задачи анализа результатов эксперимента, с построением математической модели наблюдения или эксперимента, в естественной связи с актуализацией и расширением понятийного аппарата, необходимого для такой деятельности.

Общие вопросы математической и социальной статистики (раздел 1), несомненно важные для формирования будущего психолога-исследователя, предлагается вынести в большой степени на самостоятельное изучение (и обдумывание). Обсуждение этих вопросов на аудиторных (практических/ семинарских) занятиях не должно сводиться к докладам студентов на общие темы. Предлагается сосредоточить внимание на проблемных точках, актуальных для каждого конкретного исследования, и на отражении содержательной специфики исследования в его математической модели.

На лекциях предполагается краткое изложение основных идей проверки статистических гипотез (лекция 1) и анализа статистических зависимостей между признаками (лекция 2). Могут быть указаны несколько содержательных проблемных полей, сформулированы соответствующие гипотезы/ задачи корреляционного, регрессионного, дисперсионного анализа и продемонстрированы типичные элементы дальнейшего исследования. С другой стороны, желательно, чтобы эти проблемы были связаны с интересами конкретного магистранта. При этом лекция не должна быть изложением «рецепта» обработки данных.

На практических занятиях предполагается активная работа студентов с теоретическим материалом и практическая реализация полного цикла обработки данных на модельных примерах. Одной из сверхзадач этих занятий является именно выравнивание уровней владения

студентами языком математики и анализа данных, а также компьютерными средами, реализующими алгоритмы обработки данных. На практических занятиях могут быть продемонстрированы и примеры «ручного» решения задач. При этом акцент должен делаться на исследовательских (учебных) задачах. Часть задач, поставленных на занятии, может после начального этапа совместного обсуждения быть предложена как домашнее задание, по некоторым из них целесообразно затем заслушивать краткие сообщения студентов. Как преподавателю, так и студентам во время их сообщений рекомендуется применять проблемный стиль изложения материала, ставить вопросы перед аудиторией, в том числе для последующего обдумывания, предусматривать время для ответа на вопросы слушателей. Желательно проработать глоссарий дисциплины и убедиться в четком понимании смысла содержащихся в нем терминов.

Контрольное задание по дисциплине «Регрессионный, дисперсионный и кластерный анализ данных » предлагается организовать как внеаудиторное, состоящее из двух отдельных частей (см. раздел 4). Формой рубежного контроля по дисциплине является зачёт (см. программу в разделе 5).

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 – 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

- Глобальный интеллектуальный ресурс «Портал знаний». http://statistica.ru .
- Портал MachineLearning. http://www.machinelearning.ru .
- Сайт Московского центра непрерывного математического образования. Лаборатория теории вероятностей и математической статистики https://www.mccme.ru.
- Общероссийский математический портал. http://www.mathnet.ru

Лист согласования рабочей программы
учебной дисциплины (практики)

Направление подготовки: 44.04.01 Педагогическое образование
Профиль: Компьютерное моделирование и дизайн информационной образовательной среды, МПИТОз, Методология математического образования
Рабочая программа Регрессионный, дисперсионный и кластерный анализ данных
Составители: А.Р. Сибирева – Ульяновск: УлГПУ, 2023.

Программа составлена с учетом федерального государственного образовательного стандарта высшего образования по направлению подготовки 44.04.01 Педагогическое образование, утверждённого Министерством образования и науки Российской Федерации, и в соответствии с учебным планом.

Составители _____ А.Р. Сибирева

Рабочая программа учебной дисциплины (практики) одобрена на заседании кафедры высшей математики «23» мая 2023 г., протокол № 10.

Заведующий кафедрой

_____ И.В. Столярова 23.05.23
личная подпись *расшифровка подписи* *дата*

Рабочая программа учебной дисциплины (практики) согласована с библиотекой

Сотрудник библиотеки

_____ Матвеева О.Н. 11.05.23
личная подпись *расшифровка подписи* *дата*

Программа рассмотрена и одобрена на заседании ученого совета факультета физико-математического и технологического образования «26» мая 2023г., протокол № 5.

Председатель ученого совета факультета физико-математического и технологического образования

_____ Е.М. Громова 26.05.23
личная подпись *расшифровка подписи* *дата*