

*Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Северо-Осетинский государственный университет  
имени Коста Левановича Хетагурова»*



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
«Введение в анализ данных»**

Направление подготовки 01.03.02 Прикладная математика и информатика

Профиль: «Математическое моделирование и вычислительная математика»

**Форма обучения – очная**

Владикавказ, 2019

Программа составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 01.03.02 Прикладная математика и информатика, утвержденным приказом Министерства образования и науки Российской Федерации от 10.01.2018 г. № 9, учебным планом подготовки бакалавриата по направлению 01.03.02 Прикладная математика и информатика, профиль: «Математическое моделирование и вычислительная математика», утвержденным Ученым советом ФГБОУ ВО «СОГУ» от 28.05.2019 г. № 10.

Составитель: к.ф.-м.н. Басаева Е.К.

Рабочая программа

обсуждена и утверждена на заседании кафедры математического анализа (протокол № 7 от 27.03.2019г.);

одобрена советом факультета математики и информационных технологий (протокол № 5 от 29.03.2019 г.).

## 1. Структура и общая трудоемкость дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единиц (180 часов).

	Всего	1 семестр	2 семестр
Курс	3	3	3
Семестр	5, 6	5	6
Лекции	66	34	32
Практические занятия	–	–	–
Лабораторные занятия	32		32
Консультации			+
Итого аудиторных занятий	98	34	64
Самостоятельная работа	55	38	17
Курсовая работа	–		
Зачет		+	–
Экзамен	27	–	27
Общее количество часов	180	72	108

## 2. Цели освоения дисциплины

Анализ данных – предмет, порожденный компьютерной революцией, приведшей к накоплению больших объемов данных.

Цели освоения дисциплины – ознакомление с методами анализа одномерных, двумерных и многомерных данных. Рассмотрены вопросы анализа и интерпретации связей между двумя количественными признаками, двумя качественными, а также качественным и количественным признаками. Из многомерных методов рассмотрены наивный Байесовский классификатор, метод К-средних для кластерного анализа и методы снижения размерности признакового пространства.

## 3. Место дисциплины в структуре ОПОП:

Б1.О.23. Дисциплины (модули). Обязательная часть.

Для освоения данной дисциплины полезны знания, умения, навыки и компетенции, формируемые предшествующими дисциплинами «Алгебра» «Аналитическая геометрия», «Математический анализ», «Теория вероятностей и математическая статистика», «Алгоритмы и структуры данных», «Основы и методология программирования», «Вычислительная математика».

Дисциплина обеспечивает формирование знаний, навыков и умений применяемых при изучения следующих дисциплин: «Прикладные задачи анализа данных», «Методы и системы обработки больших данных», Проектный семинар "Машинное обучение и приложения", «Эконометрика».

## 4. Требования к результатам освоения дисциплины

Компетенции обучающегося, формируемые в результате освоения дисциплины:

УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач;

ОПК-2 – Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач;

ОПК-3 – Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности.

Взаимосвязь планируемых результатов обучения по дисциплине с формируемыми компетенциями ОПОП:

Компетенции		Планируемые результаты обучения, соответствующие формируемым компетенциям ОПОП		
Код	Формулировка	Знать:	Уметь	Владеть:
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	– основные положения, связанные со сбором, систематизацией, обработкой и анализом статистических данных.	– определять методы анализа, необходимые для оценки степени и вида зависимости между переменными, снижения размерности признакового пространства и многомерной классификации данных.	– опытом сбора статистических данных из доступных информационных источников, – опытом проведения анализа данных и опытом создания научных текстов.
ОПК-2	Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач	– основные понятия и методы анализа данных и смежных дисциплин.	– применять методы устойчивого оценивания параметров и непараметрического моделирования при решении прикладных задач.	– компьютерного представления и обработки реальных данных.
ОПК-3	Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности	– основные подходы к анализу данных с использованием описательных и вероятностно-статистических методов и моделей.	собирать и обрабатывать статистические и экспериментальные данные для построения математических моделей.	– навыками применения одномерных и многомерных статистических методов и моделей для решения прикладных задач.

## 5. Содержание и учебно-методическая карта дисциплины

Таблица 5.1

Номер недели	Наименование тем (вопросов), изучаемых по данной дисциплине	Занятия			Самостоятельная работа студентов		Формы контроля	Баллы		Литература
		л	пр	лаб	Содержание	Часы		min	max	
5 семестр										
1–2	Введение в анализ данных. Предварительный анализ данных. Описательная статистика. Классификация статистических данных.	4			Генеральная и выборочная совокупности	10	Устный опрос, доклад; решение задач	0	25	[1]–[4]
3–7	Одномерный анализ количественных, категориальных и бинарных признаков.	10			Нормирование (стандартизация) и унификация данных	4				
8–12	Двумерный анализ: суммаризация и корреляция двух признаков. Количественные признаки: линейная регрессия, нелинейная и линеаризованная регрессии.	10			Линейная регрессионная модель себестоимости продукции	6	Устный опрос, доклад; решение задач	0	25	[1]–[7]
13–15	Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки	6			Канонические корреляции и канонические величины генеральной совокупности	6				
16–17	Двумерный анализ. Случай двух номинальных признаков	4			Оценка канонических корреляций и канонических величин	12				
	Всего за 1 семестр:	34				38	зачет	0	50	
6 семестр										
1	Корреляция и суммаризация для многомерных данных: Классы решающих правил.	2		2	Аномальные значения. Методы обнаружения засорения выборки	10	Устный опрос, доклад; лабораторная работа	0	25	[1]–[4]
2–3	Байесовское решающее правило.	4		4						
4	Меры качества классификатора.	2		2						
5	Задача кластеризации.	4		4	Устойчивые параметрические методы оценивания	5				
6–7	Кластеризация методом К-средних	6					Устный опрос, доклад;	0	25	[1]–[7]
8–9	Модель множественной регрессии	4		4						
10	Нелинейные модели регрессии и их линеаризация	2		2						

11	Регрессионные модели с фиктивными переменными	2		2		лабораторная работа			
12–13	Снижение размерности признакового пространства. Компонентный анализ	4		4	Эвристические методы снижения размерности признакового пространства				
14–16	Снижение размерности признакового пространства. Факторный анализ.	6		6					
	<b>Всего за 1 семестр:</b>	<b>32</b>		<b>32</b>		<b>17</b>	<b>экзамен</b>	<b>0</b>	<b>50</b>
	<b>ИТОГО</b>	<b>66</b>		<b>32</b>		<b>55</b>			

**Примечания:**

- Все виды учебной работы могут проводиться дистанционно на основании локальных нормативных актов.
- В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по индивидуальной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины может осуществляться через индивидуальные консультации преподавателя очно, в часы консультаций, по электронной почте и с использованием платформ дистанционного обучения.

## 6. Образовательные технологии

При преподавании дисциплины в основном используются традиционные образовательные технологии: лекции, лабораторные занятия и самостоятельная работа студентов. Также при проведении занятий могут быть использованы современные интерактивные и информационно-коммуникационные образовательные технологии такие как:

- интерактивная лекция представляет собой выступление преподавателя перед студенческой аудиторией с применением следующих интерактивных форм обучения: управляемая дискуссия или беседа; демонстрация слайдов или учебных фильмов; мотивационная речь и др.;

- видеоконференция – технология интерактивного взаимодействия двух и более участников образовательного процесса для обмена информацией в реальном режиме времени;

- онлайн-семинар – разновидность веб-конференции, проведение онлайн-встреч или презентаций через Интернет в режиме реального времени;

- творческое задание требует от студента воспроизведения полученной ранее информации в форме, определяемой преподавателем и требующей творческого подхода: подбор материалов по заданной теме; подбор примеров из практики; самостоятельная постановка и решение нетиповых практических задач;

- презентация проекта – слайд-презентации позволяют эффектно и наглядно представить содержание, выделить и проиллюстрировать сообщение.

## 7. Учебно-методическое обеспечение самостоятельной работы

По дисциплине предусмотрены следующие виды самостоятельной работы студентов:

- самостоятельное повторение и изучение теоретического материала;
- подготовка доклада по темам, вынесенным на самостоятельную работу;
- подготовка к лабораторным занятиям, зачету, экзамену.

Для обеспечения самостоятельной работы студентов предусмотрены следующие методические материалы (см. разделы 8–9): перечень вопросов для подготовки к зачету и экзамену, перечень рекомендованной литературы. Учебно-методические материалы необходимые для обеспечения самостоятельной работы студентов размещаются на дистанционной площадке СОГУ в начале каждого модуля.

## 8. Оценочные средства для текущего контроля успеваемости, рубежной аттестации и промежуточной аттестации по итогам освоения дисциплины

*Основными формами текущего контроля* по дисциплине являются устный опрос на лекционном или практическом занятии, доклад по теме, вынесенной на самостоятельное изучение, решение задач на практических занятиях.

*Формы промежуточной аттестации:* зачет в конце 5 семестра и экзамен в конце 6 семестра.

Студенты, набравшие в ходе текущего контроля 56–100 баллов, автоматически получают «Зачет» или соответствующую шкале экзаменационную оценку. Результирующая оценка рассчитывается по согласно действующей балльно-рейтинговой системе.

### 8.1. Формы контроля и критерии оценивания

Этап	Форма контроля	Критерии оценивания (процент от максимального кол-ва баллов)			
		86-100 %	71–85%	56–70%	Менее 56%
		отлично / зачет	хорошо / зачет	удовлетвори- тельно / зачет	неудовлетвори- тельно / незачет
1. Текущий контроль (max 25 баллов за 1 модуль)					
		7–8 баллов	5–6 баллов	4–5 баллов	0–3 баллов

	Оценка текущей работы студента в течение модуля (max 106.)	Студент активно работает на занятиях, превосходно выполняет все задания преподавателя.	Студент активно работает на занятиях, хорошо выполняет задания преподавателя.	Студент недостаточно активно работает на занятиях, удовлетворительно выполняет задания преподавателя.	Студент недостаточно активно работает на занятиях, неудовлетворительно выполняет задания преподавателя.
		4 балла	3 балла	2 балла	0–1 балл
	Доклад по теме, вынесенной на самостоятельную работу	Тема полностью раскрыта. Превосходное владение материалом. Высокий уровень самостоятельности, логичности, аргументированности. Превосходный стиль изложения.	Тема в основном раскрыта. Хорошее владение материалом. Средний уровень самостоятельности, логичности, аргументированности. Хороший стиль изложения.	Тема частично раскрыта. Удовлетворительное владение материалом. Низкий уровень самостоятельности, логичности, аргументированности. Удовлетворительный стиль изложения.	Тема не раскрыта. Неудовлетворительное владение материалом. Недостаточный уровень самостоятельности, логичности, аргументированности. Неудовлетворительный стиль изложения.
		3 балла	2 балла	1 балл	0 баллов
	лабораторная работа	Работа выполнена полностью, самостоятельно и в точном соответствии с заданием. Студент четко и аргументированно отвечает на дополнительные вопросы.	Работа выполнена полностью и в точном соответствии с заданием. При выполнении работы потребовались подсказки преподавателя. При ответе на дополнительные вопросы делает незначительные ошибки.	Работа выполнена на 70%. При выполнении работы потребовались помощь преподавателя. При ответе на дополнительные вопросы делает ошибки.	Работа не выполнена или выполнена менее чем на 70%. Недостаточный уровень самостоятельности. На дополнительные вопросы не отвечает или отвечает неверно.
<b>2. Рубежный контроль (25баллов за один модуль)</b>					
		22–25 баллов	18–21 балл	14–17 баллов	0–13 баллов
	Контрольная работа	Правильно выполнены все задания. Продemonстрирован высокий уровень владения материалом. Проявлены превосходные способности применять знания и умения к выполнению конкретных заданий.	Правильно выполнено более 70% заданий. Присутствуют незначительные ошибки. Продemonстрирован хороший уровень владения материалом. Проявлены средние способности применять знания и умения к выполнению конкретных заданий.	Задания выполнены более чем наполовину. Присутствуют серьезные ошибки. Продemonстрирован удовлетворительный уровень владения материалом. Проявлены низкие способности применять знания и умения к выполнению конкретных заданий.	Задания выполнены менее чем наполовину. Продemonстрирован неудовлетворительный уровень владения материалом. Проявлены недостаточные способности применять знания и умения к выполнению конкретных заданий.
<b>3. Итоговый контроль по дисциплине (max 50 баллов)</b>					
		43–50 баллов	36–42 балла	28–35 баллов	0–27 баллов
	Устный экзамен/зачет	Дан полный, развернутый ответ на поставленные вопросы. Ответ формулируется в терминах науки, изложен литературным языком, логичен, доказателен, демонстрирует авторскую позицию студента.	Дан полный ответ на поставленные вопросы, показано умение выделить существенные и не существенные признаки, причинно-следственные связи. Но допущены незначительные ошибки, ис-	Дан недостаточно полный ответ. Студент не способен самостоятельно выделить существенные и не существенные признаки и причинно-следственные связи. Речевое оформление требует поправок, коррекции.	Не получены ответы по базовым вопросам дисциплины или дан неполный ответ и допущены грубые ошибки. Речь неграмотная. Уточняющие вопросы преподавателя не приводят к коррекции ответа студента не только



			правленные студентам с помощью «наводящих» вопросов преподавателя.		на поставленный вопрос, но и на другие вопросы дисциплины.
--	--	--	--	--	--

## 8.2. Примерные темы лабораторных работ

1. Оценка параметров генеральной совокупности по выборке. Описательные статистики. Визуализация данных.
2. Анализ взаимосвязи социально-экономических показателей группы стран.
3. Исследование методами корреляционного анализа взаимосвязи между следующими показателями машиностроительных предприятий: рентабельность (%), премии на одного работника (руб.), фондоотдача.
4. Наивный байесовский классификатор.
5. Кластеризация методом К-средних.
6. Множественная линейная регрессия.
7. Нелинейные модели регрессии и их линеаризация.
8. Регрессионные модели с фиктивными переменными
9. Анализ главных компонент.
10. Факторный анализ.

## 8.3. Вопросы для подготовки к текущим опросам, зачету и экзамену

### *1 семестр*

1. Предварительный анализ данных. Описательная статистика.
2. Классификация статистических данных.
3. Генеральная и выборочная совокупности.
4. Одномерный анализ количественных признаков.
5. Одномерный анализ категориальных и бинарных признаков.
6. Нормирование (стандартизация) и унификация данных.
7. Двумерный анализ. Количественные признаки: линейная регрессия,
8. Двумерный анализ. Количественные признаки: нелинейная и линеаризованная регрессии.
9. Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки. Целевой количественный признак.
10. Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки. Номинальный целевой признак.
11. Двумерный анализ. Случай двух номинальных признаков.
12. Канонические корреляции и канонические величины генеральной совокупности.
13. Оценка канонических корреляций и канонических величин.

### *2 семестр*

14. Аномальные значения. Методы обнаружения засорения выборки.
15. Корреляция и суммаризация для многомерных данных: Классы решающих правил.
16. Байесовское решающее правило.
17. Меры качества классификатора.
18. Задача кластеризации.
19. Кластеризация методом К-средних.
20. Модель множественной регрессии.
21. Нелинейные модели регрессии и их линеаризация.
22. Регрессионные модели с фиктивными переменными.
23. Снижение размерности признакового пространства. Компонентный анализ.

24. Снижение размерности признакового пространства. Факторный анализ.

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

### **а) основная литература:**

1. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — М.: Издательство Юрайт, 2018. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/413060>.

2. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — М.: Издательство Юрайт, 2018. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/412967>.

### **б) дополнительная литература:**

3. Крутиков, В.Н. Анализ данных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Кемеровский государственный университет. — Кемерово: Кемеровский государственный университет, 2014. — 138 с. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=278426>. — Текст: электронный.

4. Каган, Е.С. Прикладной статистический анализ данных: учебное пособие / Е.С. Каган; Кемеровский государственный университет. — Кемерово: Кемеровский государственный университет, 2018. — 235 с. : ил., табл. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=573550>. — Текст: электронный.

5. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. 456с. — URL: [https://scask.ru/h\\_book\\_lra.php?id=1](https://scask.ru/h_book_lra.php?id=1).

6. Кремер, Н. Ш. Эконометрика: учебник и практикум для академического бакалавриата / Н. Ш. Кремер, Б. А. Путко; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 354 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-02760-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/401922>.

7. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel: учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/400278>.

### **в) Профессиональные базы данных и другие интернет-ресурсы**

- ЭБС «Научная электронная библиотека eLibrary.ru» (<http://www.elibrary.ru>).
- ЭБС «Университетская библиотека online» (<http://www.biblioclub.ru>)
- ЭБС «Юрайт» (<http://biblio-online.ru>)
- ЭБС «Консультант студента» ([studentlibrary.ru](http://studentlibrary.ru))
- Федеральная служба государственной статистики. Центральная база статистических данных. <http://www.gks.ru/>
- Общероссийский математический портал Math-Net.Ru (<http://www.mathnet.ru/>)
- Math24.ru (сайт «Высшая математика»). <http://math24.ru/>

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

Для проведения занятий используются:

— учебные аудитории для проведения лекционных занятий, занятий семинарского типа, текущего контроля, промежуточной аттестации, оборудованные аудиторной мебелью, доской (меловой, маркерной или интерактивной), мультимедийным проектором, экраном, компьютером или ноутбуком с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду СОГУ.

– компьютерный класс (корпус 10, ауд. №505, 506, 605), оборудованный аудиторной мебелью, доской (меловой, маркерной или интерактивной), компьютерами или ноутбуками с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде СОГУ, мультимедийным проектором, экраном.

*Лицензионное программное обеспечение:*

1. Windows 10 Pro for Workstations, (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
2. Office Standard 2016 (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
3. Система поиска текстовых заимствований «Антиплагиат ВУЗ»;

*Перечень ПО в свободном доступе:* Kaspersky Free; WinRar; Google Chrome; Yandex Browser; Opera Browser; Acrobat Reader; LaTeX; системы компьютерной алгебры SciLab и Maxima, R, Python.

Помещение для самостоятельной работы студентов: Зал электронных ресурсов Научной библиотеки СОГУ (корпус 6, кабинет № 1.8), укомплектован специализированной мебелью (рабочие места студентов), необходимыми техническими средствами обучения: компьютеры, принтер, возможность подключения к сети «Интернет», доступ в электронную информационно-образовательную среду СОГУ.

## **11. Лист обновления/актуализации**

1. Рабочая программа  
пересмотрена и актуализирована на заседании кафедры математического анализа,  
протокол № 7 от 23.03.2020г.;
- одобрена на заседании совета факультета математики и информационных техноло-  
гий, протокол № 5 от 27.03.2020 г.