

Министерство науки и высшего образования РФ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Б1.Б.17 Анализ данных**

---

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

**38.03.05 БИЗНЕС-ИНФОРМАТИКА**

---

Направленность (профиль)

**38.03.05 БИЗНЕС-ИНФОРМАТИКА**

---

Форма обучения

**очная**

---

Год набора

**2020**

---

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили \_\_\_\_\_

к.ф.-м.н., доцент, Лукьянова Н. А.

\_\_\_\_\_  
должность, инициалы, фамилия

## 1 Цели и задачи изучения дисциплины

### 1.1 Цель преподавания дисциплины

Цель курса заключается в представлении студентам фундаментальных понятий и теоретических основ анализа многомерных данных, а также обучении практическим навыкам статистических вычислений в экономике. Курс посвящен основным методам анализа многомерных данных, которые используются в процессе бизнес-аналитики. Рассмотрены модели и технологии интеллектуального анализа данных, которые являются теоретической и методологической основой бизнес-аналитики.

### 1.2 Задачи изучения дисциплины

Задачи дисциплины:

1. Ознакомить студентов с фундаментальными понятиями и теоретическими основами анализа многомерных данных, основными методами анализа многомерных данных, моделями и технологиями интеллектуального анализа данных.
2. Дать представление об основных методах статистического анализа числовых величин, многомерного статистического анализа, анализа временных рядов, статистики нечисловых и интервальных данных.
3. Сформировать навыки использования математических методов при решении прикладных задач.
4. Научить владеть методами факторного, дискриминантного и кластерного анализа.
5. Сформировать основу для дальнейшего самообразования в процессе профессиональной деятельности.

### 1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
<b>ОПК-3: способностью работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных компьютерных сетях</b>	
ОПК-3: способностью работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных компьютерных сетях	основные источники информации по дисциплине, разделы дисциплины и их взаимосвязь, основные требования к выполнению заданий и освоению курса методы хранения, обработки и передачи информации принципы использования информационных технологий при решении прикладных задач пользоваться справочной литературой, поисковыми системами, необходимыми прикладными пакетами программ использовать информационные технологии на всех необходимых этапах решения прикладных задач использовать языки и системы программирования для решения профессиональных задач

	<p>навыками использования сети Internet для решения образовательных задач</p> <p>навыками самостоятельного изучения теоретического материала</p> <p>навыками использования вычислительной техники, в том числе стандартных пакетов прикладных компьютерных программ</p>
<p><b>ПК-17: способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования</b></p>	
<p>ПК-17: способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования</p>	<p>фундаментальные понятия и теоретические основы анализа многомерных данных</p> <p>основные методы анализа многомерных данных</p> <p>модели и технологии интеллектуального анализа данных</p> <p>применять методы статистического анализа числовых величин</p> <p>применять методы многомерного статистического анализа, анализа временных рядов, статистики нечисловых и интервальных данных</p> <p>применять методы факторного, дискриминантного и кластерного анализа</p> <p>методами количественного анализа и моделирования</p> <p>методами теоретического и экспериментального исследования</p> <p>методами интеллектуального анализа данных</p>
<p><b>ПК-18: способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования</b></p>	
<p>ПК-18: способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования</p>	<p>основные методики статистического анализа</p> <p>основные понятия теории корреляции</p> <p>основные методы факторного и кластерного анализа</p> <p>применять основные прикладные пакеты статистического анализа числовых величин</p> <p>применять основные прикладные пакеты многомерного статистического анализа, анализа временных рядов, статистики нечисловых и интервальных данных</p> <p>применять основные прикладные пакеты факторного, дискриминантного и кластерного анализа</p> <p>навыками использования MS Office, MathCad</p> <p>навыками использования математических методов при решении типовых задач</p> <p>навыками использования математических методов при решении прикладных задач</p>

#### 1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=13222>.

## 2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
<b>Контактная работа с преподавателем:</b>	<b>1,5 (54)</b>	
занятия лекционного типа	0,5 (18)	
практические занятия	1 (36)	
<b>Самостоятельная работа обучающихся:</b>	<b>0,5 (18)</b>	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	
<b>Промежуточная аттестация (Экзамен)</b>	<b>1 (36)</b>	

### 3 Содержание дисциплины (модуля)

#### 3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
<b>1. Введение в анализ данных</b>									

<p>1. Задачи анализа данных и знаний. Сходства и отличия. Закономерности. Теория измерений. Классификация задач анализа данных. Куб данных. Базовые гипотезы, лежащие в основе методов анализа данных. Гипотеза компактности. Задачи таксономии.</p> <p>Нечисловые данные. Нечеткие множества – частный случай нечисловых данных. Данные и расстояния в пространствах произвольной природы.</p> <p>Аксиоматическое введение расстояний. Статистическая устойчивость как основное условие применение вероятно-статистических методов. Законы больших чисел. Центральные предельные теоремы. Теоремы о наследовании сходимости. Метод линеаризации.</p> <p>Принцип инвариантности. Нечеткие множества как проекции случайных множеств. Устойчивость выводов и принцип уравнивания погрешностей. Описание данных, оценивание и проверка гипотез. Модели порождения данных. Таблицы и выборочные характеристики. Шкалы измерения, инвариантные алгоритмы и средние величины. Вероятностные модели порождения нечисловых данных.</p> <p>Анализ структурированной информации с помощью статистических и математических методов: отбор признаков, стратификация, кластеризация, ассоциации, визуализация, регрессия, прогнозирование временных рядов, последовательности. Анализ неструктурированной или слабоструктурированной информации: категоризация, разведка и семантическая обработка текстов, расширенный поиск информации.</p>	6							
---	---	--	--	--	--	--	--	--

<p>2. Эмпирическая и математическая системы.  Кодирование данных и кодировочная таблица.  Частотный анализ. Проверка репрезентативности.  Отбор данных. Модификация данных. Создание и анализ таблиц сопряженности. Подготовка данных к анализу. Корреляции. Проверка связанности переменных. Сжатие данных. Реализация расчетов при помощи стандартных пакетов прикладных компьютерных программ.</p>			12					
<p>3. Самостоятельное изучение теоретического материала, выполнение расчетных заданий, выполнение реферата, включая выполнение заданий на электронном курсе с использованием ЭО и ДОТ.</p>							6	
<p><b>2. Многомерный статистический анализ</b></p>								

<p>1. Статистический анализ числовых величин. Средние. Непараметрические оценки плотности. Методы оценивания параметров. Одношаговые оценки. Асимптотика решений экстремальных статистических задач. Робастность статистических процедур. Метод моментов проверки гипотез. Неустойчивость параметрических методов отбраковки выбросов. Предельная теория непараметрических критериев. Метод проверки гипотез по совокупности малых выборок. Проблема множественных проверок статистических гипотез.</p> <p>Коэффициент корреляции. Общие понятия и определения. Понятие корреляции. Количественные критерии оценки тесноты связи. Значимость результатов или статистическая достоверность. Меры связанности для переменных с порядковой шкалой. Ранговые шкалы. Коэффициент ранговой корреляции Спирмена или Кендала. Коэффициент конкордации оценивает согласованность ранжировок всей выборочной совокупности. Меры связанности для переменных с номинальной шкалой. Коэффициент сопряженности Пирсона. Коэффициент Крамера (V). Критерий Фишера, коэффициент Юла. Тау (г) Гудмена-Крускала. Коэффициент неопределенности.</p> <p>Задачи дискриминантного анализа. Построение дискриминантной функции. Классификация. Вывод групп и статистик.</p>	6							
--	---	--	--	--	--	--	--	--

<p>2. Дискриминантный анализ. Сравнение средних. Применение t-критерия. T-критерий для независимых выборок. T-критерий для зависимых выборок. T-критерий для одной выборки. Пошаговый алгоритм вычисления. Тестовая величина U, определенная с помощью теста Манна и Уитни. Наименьшее значение из обеих ранговых сумм (W-тест Уилкоксона). Тестовая величина Z, определенная по тесту Колмогорова-Смирнова. Тест Мозеса (Moses). Однофакторный дисперсионный анализ. Метод главных компонент. Факторный анализ. Двухэтапный кластер-анализ. Иерархический кластерный анализ. Кластерный анализ методом k- средних. Реализация расчетов при помощи стандартных пакетов прикладных компьютерных программ.</p>			12					
<p>3. Самостоятельное изучение теоретического материала, выполнение расчетных заданий, выполнение реферата, включая выполнение заданий на электронном курсе с использованием ЭО и ДОТ.</p>						6		
<p><b>3. Факторный и кластер-анализ</b></p>								

<p>1. Разделение на группы. Критерии сходства. Значимость критериев. Дисперсия. Критерий Тьюки. ANOVA- метод. Многофакторный дисперсионный анализ. Геометрическое описание метода главных компонент. Основные гипотезы и основания применения метода главных компонент. Извлечение. Вращение. Классификация. Интерпретация. Алгоритм. Модели факторного анализа. Основные гипотезы и основания применения методов и алгоритмов факторного анализа. Равенства Тэрстоуна. Ограничения факторного анализа. Алгоритм факторного анализа. Maximum likelihood (Метод максимального правдоподобия). Principal-axis factoring (Метод главных факторов). Alpha-factoring (Альфа-факторный анализ). Image factoring (Факторный анализ образов). Вывод результатов факторного анализа. Этапы кластерного анализа. Выбор переменных-критериев для кластеризации. Выбор способа измерения расстояния между объектами, или кластерами. Формирование кластеров. Модели парной регрессии. Метод наименьших квадратов. Модель множественной регрессии. Теорема Гаусса- Маркова. Проверка статистических гипотез. Доверительные интервалы и доверительные области.</p>	6							
<p>2. Самостоятельное изучение теоретического материала, выполнение расчетных заданий, выполнение реферата, включая выполнение заданий на электронном курсе с использованием ЭО и ДОТ.</p>							6	

3. Модели парной регрессии. Метод наименьших квадратов. Многомерная регрессия. Анализ временных рядов. Логистическая регрессия. Пробит-анализ. Метод нелинейной регрессии. Многомерная регрессия. Реализация расчетов при помощи стандартных пакетов прикладных компьютерных программ. Защита расчетных заданий и рефератов.			12					
Всего	18		36				18	

## 4 Учебно-методическое обеспечение дисциплины

### 4.1 Печатные и электронные издания:

1. Лейнартас Д. Е. Анализ данных. Математические методы анализа данных: учеб.-метод. пособие для самостоят. работы [для студентов напр. 080500.62 «Бизнес-информатика» очной формы обуч.] (Красноярск: СФУ).
2. Дайитбегов Д. М. Компьютерные технологии анализа данных в эконометрике: монография(Москва: Вузовский учебник).
3. Гобарева Я. Л., Городецкая О. Ю., Золотарюк А. В. Бизнес-аналитика средствами EXCEL: учебное пособие для студентов вузов, обучающихся по направлению "Экономика"(Москва: Вузовский учебник).
4. Воскобойников Ю. Е. Регрессионный анализ данных в пакете Mathcad: учеб. пособие(Санкт-Петербург: Лань).
5. Тюрин Ю.Н., Макаров А. А. Анализ данных на компьютере: учебное пособие по направлениям "Математика", "Математика. Прикладная математика"(Москва: Форум).
6. Якунин Ю. Ю. Системный анализ данных и технологий принятия решений: учеб.- метод. пособие(Красноярск: СФУ).
7. Кремер Н. Ш., Путко Б. А., Тришин И. М., Фридман М. Н., Кремер Н. Ш. Математика для экономистов: от Арифметики до Эконометрики: учебно-справочное пособие для студентов вузов (бакалавров) по специальностям 080116 (061800) "Математические методы в экономике" и другим экономическим специальностям(Москва: Юрайт).
8. Козлов А. Ю., Мхитарян В. С., Шишов В. Ф. Статистический анализ данных в MS Excel: учебное пособие(Москва: ИНФРА-М).
9. Лемешко Б. Ю., Постовалов С. Н., Лемешко С. Б., Чимитова Е. В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход(Москва: ООО "Научно-издательский центр ИНФРА-М").
10. Боровиков В. П. Популярное введение в современный анализ данных в системе STATISTICA: Учебное пособие для вузов(Москва: Горячая линия-Телеком).
11. Берк К., Кэйри П. Анализ данных с помощью Microsoft Excel(М.: Вильямс).
12. Якунин Ю. Ю. Системный анализ данных и технологий принятия решений: учебно-методическое пособие [для магистров очной формы обучения напр. 220100.68 «Системный анализ и управление»] (Красноярск: СФУ).
13. Лемешко Б. Ю., Постовалов С. Н., Лемешко С. Б., Чимитова Е. В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход(Новосибирск: Новосибирский государственный технический университет (НГТУ)).

**4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):**

1. Методика проведения занятий предполагает использование технических средств (проекторы, интерактивные доски), обеспеченных соответствующим программным обеспечением, для проведения практических занятий применяется вычислительная техника и стандартные пакеты прикладных программ (MS Office, MathCad, MathLab и др.). Для проведения практических занятий используется вычислительная техника (компьютерный класс на 10-15 мест).

**4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:**

1. Электронные учебные курсы в LMS Moodle, e.sfu-kras.ru на сайте СФУ.
2. Федеральный портал «Российское образование» <http://www.edu.ru/>
3. Информационно-образовательный портал <http://www.faito.ru/>
4. Математический портал <http://allmath.ru/>
5. <http://www.e-library.ru/defaultx.asp> E-Library
6. <http://ict.edu.ru/konkurs> «Образование в Рунете»
7. <http://www.gnpbu.ru/> - Научная педагогическая библиотека им. К.Д. Ушинского
8. <http://www.i-exam.ru/> - Интернет-тренажеры и тестовая база данных Росаккредагентства для проведения репетиционного тестирования (ФЭПО)
9. Учебная и научная литература по курсу. Компьютерные демонстрации, связанные с программой курса, компьютерные демонстрации, технические возможности для их просмотра и прослушивания. Свободный доступ в Интернет, наличие компьютерных программ общего назначения.
10. Операционные системы: семейства Windows (не ниже Windows XP).

**5 Фонд оценочных средств**

Оценочные средства находятся в приложении к рабочим программам дисциплин.

**6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)**

В течение четвертого семестра аудиторные занятия по дисциплине включают в себя: лекции – 1 раз в две недели, практические занятия – 1 раз в неделю. В конце семестра проводится зачет. Результирующая оценка выставляется по текущей работе в семестре и итоговой работе в конце семестра. Положительная оценка ставится с 50% от общей суммы баллов.

В течение семестра проводятся три контрольных работы, студенты самостоятельно выполняют индивидуальные расчетные задания и реферат.

Теоретическая подготовка студентов предполагает, наряду с чтением лекций, использование учебников и учебных пособий по приведенному списку литературы. Лекции дополняются практическими занятиями, на которых студенты учатся решать задачи и применять лекционный материал. В целом каждое занятие соответствует определенной лекции. Практические занятия проводятся с целью освоения теоретического материала и создания навыков решения задач по соответствующим разделам, в том числе с использованием современного программного обеспечения (в компьютерном классе). Каждое занятие заключается в решении комплекта задач по определенной теме, с теоретическим обоснованием (определения, теоремы). Для подготовки к занятиям студенты должны повторить пройденный теоретический материал, желательно иметь при себе конспект лекций.

Самостоятельная работа студентов по дисциплине включает: самостоятельное изучение теоретического материала, выполнение расчетных заданий, подготовку к контрольным работам, выполнение реферата. Расчетные задания выдаются преподавателем с указанием учебно-методической литературы либо в виде раздаточного материала по вариантам.

Типовые расчеты выполняются студентами в отдельной тетради и передаются для проверки преподавателю. Оценка выставляется в 100-балльной шкале в соответствии с долей выполненных заданий и допущенными ошибками. Проверенная работа возвращается студенту для исправления и доработки, по окончании которой оценка может быть скорректирована. Расчетно-графические задания предусматривают использование студентами численных методов и стандартного программного обеспечения (MS Excel, MathCad и др). РГЗ оформляются, как правило, в виде распечатки из использованной программы. Допускается самостоятельное программирование, расчеты, построение графиков от руки.

Для проведения всех практических занятий используется вычислительная техника (компьютерный класс на 10-15 мест).

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации в зависимости от нозологии:

Для лиц с нарушениями зрения:

- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.