

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Бианкина Алена Олеговна
Должность: Ректор
Дата подписания: 14.07.2023 15:32:47
Уникальный программный ключ:
b2aeadef209e4ec32d89f812db7eed614bb00b0c

**Автономная некоммерческая организация высшего образования
« Институт социальных наук »**



Протокол от 14.07.23 № 07

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.02 Математические методы статистической обработки и анализа данных
(краткое наименование дисциплины)

38.04.05 Бизнес-информатика
(код, наименование направления подготовки)

«Бизнес-аналитика»
(направленность)

магистр
(квалификация)

очная
(форма обучения)

Москва 2022

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине
 - 4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации
 - 4.2. Материалы текущего контроля успеваемости обучающихся
 - 4.3. Оценочные средства для промежуточной аттестации
 - 4.4. Методические материалы
5. Методические указания для обучающихся по освоению дисциплины
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине
 - 6.1. Основная литература
 - 6.2. Дополнительная литература
 - 6.3. Учебно-методическое обеспечение самостоятельной работы
 - 6.4. Нормативные правовые документы
 - 6.5. Интернет-ресурсы
 - 6.6. Иные источники
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

Дисциплина Б1.В.02 Математические методы статистической обработки и анализа данных обеспечивает овладение следующими компетенциями:

Таблица 1

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-2	Способен обосновывать подходы, используемые в бизнес-анализе, руководить и управлять бизнес-анализом с использованием информационно-коммуникационных технологий	ПКс-2.1	Способен использовать современные методы, информационные технологии, программный инструментарий в объеме, необходимом для решения задач бизнес-аналитики, использовать англоязычную документацию и справочные системы

В результате освоения дисциплины у магистрантов должны быть сформированы компетенции:

Таблица 2

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Управление проектами в области ИТ малого и среднего уровня сложности в условиях неопределенностей, порождаемых запросами на изменения, с применением формальных инструментов управления рисками и проблемами проекта, В/ Планирование в проектах малого и среднего уровня сложности в области ИТ, В/31.7	ПКс-2.1	на уровне знаний: Знать: <ul style="list-style-type: none"> – основные понятия и принципы планирования и организации эксперимента; – особенности экспериментов в области информатики и вычислительной техники и структуру получаемых экспериментальных данных; – порядок планирования экспериментов и обработки данных для получения достоверных результатов исследования; – основы математического моделирования информационных процессов
		на уровне умения: Уметь: <ul style="list-style-type: none"> - правильно поставить задачу эмпирического исследования, проанализировать полученные результаты, подтвердить или опровергнуть выдвинутые гипотезы; – самостоятельно подбирать и использовать статистические методы, адекватные задачам исследования; – применять пакеты программ для решения задач обработки эмпирических данных и визуализации

		<p>результатов;</p> <ul style="list-style-type: none"> – эффективно применять методы исследования к получению, анализу и интерпретации экспериментальных данных; – использовать вероятностные модели конкретных процессов; – создавать математические модели информационных процессов
		<p>на уровне владения:</p> <p>Владеть:</p> <ul style="list-style-type: none"> - статистическими методами обработки экспериментальных данных для анализа информационных процессов; – навыками представления и интерпретации результатов математической обработки данных; – технологией обработки экспериментальных данных на ЭВМ с применением современных пакетов прикладных программ; – методологией планирования эксперимента; – навыками построения моделей объектов исследования

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы /144 часа.

Дисциплина может реализоваться с применением дистанционных образовательных технологий (*далее - ДОТ*).

Таблица 3

Очная форма

Вид работы	Трудоемкость (акад/астр. часы)
Общая трудоемкость	144/108
Контактная работа с преподавателем	50/37,5
Лекции	20/15
Практические занятия	28/21
Самостоятельная работа	58/43,5
Контроль	36/27
Формы текущего контроля	Реферат
Форма промежуточной аттестации	Экзамен

Место дисциплины в структуре образовательной программы

Дисциплина Б1.В.02 Математические методы статистической обработки и анализа данных относится к части, формируемой участниками образовательных отношений, образовательной программы подготовки магистранта федерального государственного образовательного стандарта высшего образования.

Дисциплина изучается в 1 семестре на первом курсе.

Преподавание дисциплины Б1.В.02 «Математические методы статистической обработки и анализа данных» основано на знаниях, полученных при изучении дисциплин «Теория вероятностей и математическая статистика», «Эконометрика» (уровень подготовки – бакалавриат). В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как Б1.О.07 «Аналитическая поддержка принятия

решений», Б1.В.07 «Моделирование бизнес-процессов. Process mining», Б1.В.05 «Методы бизнес-аналитики», а также для выполнения выпускной квалификационной работы, сдачи государственного экзамена и выполнения задач практик, в том числе научно-исследовательской работы..

Формой промежуточной аттестации в соответствии с учебным планом является экзамен.

Доступ к системе дистанционных образовательных технологий осуществляется каждым обучающимся самостоятельно с любого устройства . Пароль и логин к личному кабинету / профилю предоставляется студенту в учебной части.

3. Содержание и структура дисциплины

Очная форма обучения

Таблица 4

№ п/п	Наименование тем (разделов),	Объем дисциплины (модуля), час.					Форма текущего контроля успеваемости*, промежуточной аттестации***	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					СР
			Л	ЛР *	ПЗ	КСР		
Тема 1	ОСНОВЫ МОДЕЛИРОВАНИЯ И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ	30	4		8		18	УО/Э
Тема 2	ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ	38	8		10		20	УО/Э
Тема 3	КЛАССИФИКАЦИЯ И СНИЖЕНИЕ РАЗМЕРНОСТИ	38	8		10		20	УО/Э/Т
Контроль		36						
Промежуточная аттестация						2		Экзамен
Всего (акад./астр. часы):		144/108	20/15		28/21	2/1,5	58/43,5	

УО – устный опрос

Т – тест

Э - экзамен

Содержание дисциплины

Тема 1. ОСНОВЫ МОДЕЛИРОВАНИЯ И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ

Статистики эмпирического ряда: Классификация признаков по шкалам измерений. Описательная статистика: среднее значение, математическое ожидание; медиана; мода; дисперсия, среднее квадратичное отклонение; коэффициент вариации; показатель точности опыта; минимум, максимум, размах выборки; моменты распределения. Вариационная статистика: параметры классовых интервалов, группировка, функции эмпирического распределения. Ранжирование. Проверка случайности выборки из нормальной совокупности. Репрезентативность выборки.

Проверка гипотез: Общая методика. Сравнение методик. Односторонняя и

двусторонняя гипотезы. Независимые и сопряженные выборки. Параметрические тесты: t-критерий Стьюдента, F-критерий Фишера, G-критерий различных средних, параметрические множественные сравнения. Непараметрические тесты. Проверка типа распределения эмпирических данных: простые и сложные гипотезы, простейшие методы, критерии согласия, критерии отклонения распределения от нормальности.

Дисперсионный анализ: Однофакторный анализ: однофакторный дисперсионный анализ, ранговый однофакторный анализ Краскела-Уоллиса, G-критерий Кокрена, критерии Шеффе, Дункана, Тьюки. Многофакторный анализ: двухфакторный дисперсионный анализ, ранговый критерий Фридмана, критерий Пейджа, критерий Шеффе для связанных выборок.

Тема 2. ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ

Теория распределений: Общая методика. Функции распределения и обратные функции распределения. Одномерные распределения: непрерывные, дискретные, генерация одномерных распределений. Многомерные распределения: нормальное распределение, генерация многомерных распределений. Теоретические и эмпирические распределения.

Корреляционный анализ: Корреляция количественных признаков: коэффициент корреляционного отношения Пирсона, коэффициент корреляции Фехнера, ковариация. Корреляция порядковых признаков: показатель ранговой корреляции Спирмена, коэффициент ранговой корреляции Кендалла. Корреляция номинальных признаков. Корреляция признаков, измеренных в различных шкалах. Множественные корреляции: коэффициент множественной корреляции, канонический корреляционный анализ, коэффициент конкордации. Критерии некоррелированности.

Методы планирования эксперимента. Аппроксимация зависимостей: Математическое моделирование и регрессионный анализ. Общая методика. Полиномиальная аппроксимация, интерполяционный полином Лагранжа. Экспоненциально-степенная аппроксимация. Логарифмическая функция. Нелинейная функция общего вида. Линейный множественный регрессионный анализ.

Тема 3. КЛАССИФИКАЦИЯ И СНИЖЕНИЕ РАЗМЕРНОСТИ

Методы снижения размерности признакового пространства. Факторный анализ: Метод главных компонент. Метод минимизации энтропии. Преобразование Карунена–Лоэва. Метод главных факторов: проблема общности, проблема факторов, измерение факторов. Метод максимума правдоподобия. Критерии максимального числа факторов. Визуализация результатов факторного анализа.

Кластерный анализ. Дискриминантный анализ: Меры различия и меры сходства. Кластерный анализ: метод ближайшей связи, метод средней связи Кинга, метод Уорда, метод средних Мак-Куина. Визуализация результатов кластерного анализа. Выявление информативных параметров. Метод Байеса. Линейный дискриминантный анализ.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

Промежуточная аттестация может проводиться с использованием ДОТ.

4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации.

В ходе реализации дисциплины Б1.В.02 Математические методы статистической обработки и анализа данных используются следующие методы текущего контроля успеваемости обучающихся:

Таблица 4.1

Тема (раздел)	Формы текущего контроля успеваемости
---------------	--------------------------------------

Тема 1. ОСНОВЫ МОДЕЛИРОВАНИЯ И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ	УО/Э
Тема 2. ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ	УО/Э
Тема 3. КЛАССИФИКАЦИЯ И СНИЖЕНИЕ РАЗМЕРНОСТИ	УО/Э/Т

4.1.2. Экзамен проводится с применением следующих методов (средств):

Экзамен проводится в компьютерном классе в устной форме. Во время экзамена проверяется уровень знаний по дисциплине Б1.В.02 Математические методы статистической обработки и анализа данных, а также уровень умений решать учебные задачи анализа данных с использованием программных приложений. К экзамену должны решить задания по всем темам учебной дисциплины. Результаты решения задач могут быть использованы при решении практической задачи в соответствии с имеемым перечнем задач. Пример задач приведен в программе. При ответе на вопросы магистрант показывает умение решать практические задачи с помощью средств бизнес-аналитики, которые изучаются в других дисциплинах образовательной программы, изучаемых параллельно.

4. 2. Материалы текущего контроля успеваемости обучающихся.

4.2.1.Задание

Примеры заданий

Задание по корреляционному, регрессионному, компонентному и кластерному анализу даны в таблице 1, а значения показателей производственно-хозяйственной деятельности предприятий машиностроения приведены в таблице 2. Рассматриваются следующие показатели:

- Y_1 - производительность труда;
- Y_2 - индекс снижения себестоимости продукции;
- Y_3 - рентабельность;
- X_4 - трудоемкость единицы продукции;
- X_5 - удельный вес рабочих в составе ППП;
- X_6 - удельный вес покупных изделий;
- X_7 - коэффициент сменности оборудования;
- X_8 - премии и вознаграждения на одного работника;
- X_9 - удельный вес потерь от брака;
- X_{10} - фондоотдача;
- X_{11} - среднегодовая численность ППП;
- X_{12} - среднегодовая стоимость ОПФ;
- X_{13} - среднегодовой фонд заработной платы ППП;
- X_{14} - фондовооруженность труда;
- X_{15} - оборачиваемость нормируемых оборотных средств;
- X_{16} - оборачиваемость ненормируемых оборотных средств;
- X_{17} - непроизводственные расходы.

Таблица 1

№ варианта	Результативный признак, Y	Номера факторных признаков, X
1	1	6,8, 11, 12, 17
2	1	6,8, 11, 13,17
3	1	8,11,12,13,17

4	1	6, 8, 13, 14, 17
5	1	8, 11, 13, 14, 17
6	1	6,8, 12, 13, 17
7	1	7, 11, 12, 13, 17
8	1	7,9, 12, 13, 17
9	1	8,11, 12, 13,17
10	1	8,9,13,14,17
11	1	5, 6, 7, 9, 17
12	1	5,7,9, 11,17
13	1	5,6,12,13,17
14	1	5,7,10,14,17
15	1	5,6,10,14,17
16	3	8, 10, 15, 16, 17
17	3	5,6, 10,15, 17
18	3	5,6,7,11,12
19	3	8,9,10,11,17
20	3	8,9,10,12,17
21	2	4, 5, 6, 8, 9
22	2	4, 5, 6, 7, 9
23	2	4,5,6,8,9
24	2	4, 5, 8, 9, 17
25	2	4, 5, 7, 9, 17

Таблица исходных данных

Таблица 2

№ пред- приятия	V;	V;	V.	X,	X ₅	x ₆	X ₇	X ₈	X ₉	X ₁₀
1	9,26	204,2	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45
2	9,38	209,6	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30
3	12,11	222,6	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37
4	10,81	236,7	12,85							
5	9,35	62,0	10,63							
6	9,87	53,1	9,12							
7	8,17	172,1	25,83							
8	9,12	56,5	23,39							

9	5,88	52,6	14,68							
10	6,30	46,6	10,05							
11	6,22	53,2	13,99							
12	5,49	30,1	9,68							
13	6,50	146,4	10,03							
14	6,61	18,1	9,13							
15	4,32	13,6	5,37							
16	7,37	89,8	9,86							
17	7,02	62,5	12,62	0,32	0,78	0,08	35	0,79	0,56	1,60
18	8,25	46,3	5,02	0,25	0,78	0,20	,42	0,34	1,76	1,53
19	8,15	103,5	21,18	0,31	0,81	0,20	1,37	1,60	1,31	1,40
20	8,72	73,3	25,17	0,26	0,79	0,30	1,41	1,46	0,45	2,22
21	6,64	76,6	19,40	0,37	0,77	0,24	1,35	1,27	0,50	1,32
22	8,10	73,01	21,0	0,29	0,78	0,10	1,48	1,58	0,77	1,48
23	5,52	32,3	6,57	0,34	0,72	0,11	1,24	0,68	1,20	0,68
24	9,37	199,6	14,19	0,23	0,79	0,47	1,40	0,86	0,21	2,30
25	13,17	598,1	15,81	0,17	0,77	0,53	1,45	1,98	0,25	1,37
26	6,67	71,2	5,23	0,29	0,80	0,34	1,40	0,33	0,15	1,51
27	5,68	90,8	7,99	0,41	0,71	0,20	1,28	0,45	0,66	1,43
28	5,22	82,1	17,50	0,41	0,79	0,24	1,33	0,74	0,74	1,82
29	10,02	76,2	17,16	0,22	0,76	0,54	1,22	0,03	0,32	2,62
30	8,16	119,5	14,54	0,29	0,78	0,40	1,28	0,99	0,89	1,75
31	3,78	21,9	6,24	0,51	0,62	0,20	1,47	0,24	0,23	1,54
32	6,48	48,4	12,08	0,36	0,75	0,64	1,27	0,57	0,32	2,25
33	10,44	173,5	9,49	0,23	0,71	0,42	1,51	1,22	0,54	1,07
34	7,65	74,1	9,28	0,26	0,74	0,27	1,46	0,68	0,75	1,44
35	8,77	68,6	11,42	0,27	0,65	0,37	1,27	1,00	0,16	1,40
36	7,00	60,8	10,31	0,29	0,66	0,38	1,43	0,81	0,24	1,31
37	11,06	355,6	8,65	0,01	0,84	0,35	1,50	1,27	0,59	1,12
38	9,02	264,8	10,94	0,02	0,74	0,42	1,35	1,14	0,56	1,16
39	13,28	526,6	9,87	0,18	0,75	0,32	1,41	1,89	0,63	0,88
40	9,27	118,6	6,14	0,25	0,75	0,33	1,47	0,67	1,10	1,07
41	6,70	37,1	12,93	0,31	0,79	0,29	1,35	0,96	0,39	1,24
42	6,69	57,7	9,78	0,38	0,72	0,30	1,40	0,67	0,73	1,49
43	9,42	51,6	13,22	0,24	0,70	0,56	1,20	0,98	0,28	2,03
44	7,24	64,7	17,29	0,31	0,66	0,42	1,15	1,16	0,10	1,84
45	5,39	48,3	7,11	0,42	0,69	0,26	1,09	0,54	0,68	1,22
46	5,61	15,0	22,49	0,51	0,71	0,16	1,26	1,23	0,87	1,72
47	5,59	87,5	12,14	0,31	0,73	0,45	1,36	0,78	0,49	1,75
48	6,57	108,4	15,25	0,37	0,65	0,31	1,15	1,16	0,16	1,46
49	6,54	267,3	31,34	0,16	0,82	0,08	1,87	4,44	0,85	1,60
50	4,23	34,2	11,56	0,18	0,80	0,68	1,17	1,06	0,13	1,47
51	5,22	26,8	30,14	0,43	0,83	0,03	1,61	2,13	0,49	1,38
52	18,00	43,6	19,71	0,40	0,70	0,02	1,34	1,21	0,09	1,41
53	11,03	72,0	23,56	0,31	0,74	0,22	1,22	2,20	0,79	1,39

Продолжение таблицы 2

№ пред- приятия	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇
1	26006	167,69	47750	6,40	166,32	10,08	17,72
2	23935	186,10	50391	7,80	92,88	14,76	18,39
3	22589	220,45	43149	9,76	158,04	6,48	26,46
4	21220	169,30	41089	7,90	93,96	21,96	22,37
5	7394	39,53	14257	5,35	173,88	11,88	28,13
6	11586	40,41	22661	9,90	162,30	12,60	17,55
7	26609	102,96	52509	4,50	88,56	11,52	21,92
8	7801	37,02	14903	4,88	101,16	8,28	19,52
9	11587	45,74	25587	3,46	166,32	11,52	23,99
10	9475	40,07	16821	3,60	140,76	32,40	21,76
11	10811	45,44	19459	3,56	128,52	11,52	25,68
12	6371	41,08	12973	5,65	177,84	17,28	18,13
13	26761	136,14	50907	4,28	114,48	16,20	25,74
14	4210	42,39	6920	8,85	93,24	13,32	21,21
15	3557	37,39	5736	8,52	126,72	17,28	22,97
16	14148	101,78	26705	7,19	91,80	9,72	16,38
17	9872	47,55	20068	4,82	69,12	16,20	13,21
18	5975	32,61	11487	5,46	66,24	24,84	14,48
19	16662	103,25	32029	6,20	67,68	14,76	13,38
20	9166	38,95	18946	4,25	50,40	7,56	13,69
21	15118	81,32	28025	5,38	70,56	8,64	16,66
22	11429	67,26	20968	5,88	72,00	8,64	15,06
23	6462	59,92	11049	9,27	97,20	9,00	20,09
24	24628	107,34	45893	4,36	80,28	14,76	15,98
25	49727	512,60	99400	10,31	51,48	10,08	18,27
26	11470	53,81	20719	4,69	105,12	14,76	14,42
27	19448	80,83	36813	4,16	128,52	10,44	22,76
28	18963	59,42	33956	3,13	94,68	14,76	15,41
29	9185	36,96	17016	4,02	85,32	20,52	19,35
30	17478	91,43	34873	5,23	76,32	14,40	16,83
31	6265	17,16	11237	2,74	153,00	24,84	30,53
32	8810	27,29	17306	3,10	107,64	11,16	17,98
33	17659	184,33	39250	10,44	90,72	6,48	22,09
34	10342	58,42	19074	5,65	82,44	9,72	18,29
35	8901	59,40	18452	6,67	79,92	3,24	26,05
36	8402	49,63	17500	5,91	120,96	6,48	26,20
37	32625	391,27	7888	11,99	84,60	5,40	17,26
38	31160	258,62	58947	8,30	85,32	6,12	18,83
39	46461	75,66	94697	1,63	101,52	8,64	19,70
40	13833	123,68	29626	8,94	107,64	11,88	16,87
41	6391	37,21	11688	5,82	85,32	7,92	14,63
42	11115	53,37	21955	4,80	131,76	10,08	22,17
43	6555	32,87	12243	5,01	116,64	18,72	22,62
44	11085	45,63	20193	4,12	138,24	13,68	26,44
45	9484	48,41	20122	5,10	156,96	16,56	22,26
46	3967	13,58	7612	3,49	137,52	14,76	19,13
47	15283	63,99	27404	4,19	135,72	7,92	18,28
48	20874	104,55	39648	5,01	155,52	18,36	28,23

49	19418	222,11	43799	11,44	48,60	8,28	12,39
50	3351	25,76	6235	7,67	42,84	14,04	11,64
51	6338	29,52	11524	4,66	142,20	16,92	8,62
52	9756	41,99	17309	4,30	145,80	11,16	20,10
53	11795	78,11	22225	6,62	120,52	14,76	19,41

Задание 4. Построить линейные модели с помощью регрессионного и факторного анализа для данных классического набора данных, предложенного Р.Фишером по ирисам.

4.2.2. Рефераты

1. Примеры практического использования методов многомерного статистического анализа.
2. Для непараметрической модели метода наименьших квадратов в случае линейной функции одной переменной разработайте алгоритмы
 - расчета доверительных границ для коэффициентов модели;
 - проверки гипотез относительно этих коэффициентов.
3. Докажите, что сумма исходных значений зависимой переменной должна быть равна сумме восстановленных значений.
4. Критерии качества регрессионной модели.
5. Использование непараметрических оценок плотности для восстановления зависимости.
6. Теоремы умножения и сложения для индекса инфляции.
7. Учет инфляции при проведении анализа финансово-хозяйственной деятельности предприятия.
8. Задача снижения размерности: измерение латентных переменных. Модель и метод главных компонент. Ограничения на применение метода.
9. Матрица нагрузок. Интерпретация матрицы нагрузок.
10. Задача снижения размерности: измерение латентных переменных. Модель и метод главных компонент. Ограничения на применение метода.
11. Модель и методы факторного анализа.
12. Графическое представление результатов факторного анализа и метода главных компонент.
13. Понятие о больших данных
14. Простейшие методы обработки больших данных
15. Многомерные статистические методы в экономике, управлении и финансах
16. Реализация статистических методов в пакетах прикладных программ
17. Математические основы многомерных статистических методов
18. Методы множественного корреляционно-регрессионного анализа.
19. Постановка задач классификации
20. Кластерный анализ
21. Использование кластерного анализа
22. Постановка задач снижения размерности
23. Использование компонентного анализа
24. Введение в теорию качественных признаков и нечисловой информации
25. Методы оцифровки 16. Введение в методы Data Mining
26. Базы данных
27. Новые технологии обработки и хранения больших данных
28. Технологии поиска
29. Интеграция данных из различных источников.

4.2.3. Тест

- 1) Какой метод используется для оценки параметров уравнения используется, в соответствии с которым $\hat{\beta} = (X'X)^{-1} X'Y$, где $\hat{\beta}$ – вектор оценок параметров уравнения, X – матрица независимых переменных, Y – вектор независимых переменных.
 1. Метод к-средних;
 2. Метод наименьших квадратов;
 3. Индексный метод;
 4. Иерархические кластер-процедуры.
- 2) По какой формуле определяется наблюдаемое значение F -критерия Фишера (R^2 – множественный коэффициент детерминации, n – число наблюдений, m – число независимых переменных)?
 1. $F_{набл} = \frac{R}{1-R^2} \times \frac{n-m}{m}$
 2. $F_{набл} = \frac{R^2}{1-R^2} \times \frac{n}{m}$
 3. $F_{набл} = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m}$
 4. $F_{набл} = \frac{R^2}{1-R^2} \times \frac{n-1}{m}$
- 3) Для оценки надежности регрессионного уравнения наблюдаемое значение F -критерия Фишера должно быть:
 1. Больше расчетного значения F -критерия Фишера
 2. Меньше расчетного значения F -критерия Фишера
 3. Равно расчетному значению F -критерия Фишера
 4. F -критерий Фишера не оценивает надежность уравнения.
- 4) Для определения количества значений, находящихся в выборке ниже или выше медианного значения, используется:
 1. Коэффициент конкордации;
 2. Коэффициент корреляции рангов Спирмена;
 3. Коэффициент корреляции рангов Кэндалла;
 4. Критерий знаков.
- 5) Какая математическая модель приведена ниже?
$$x_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \dots + a_{im}f_{mj} + d_j v_{ij}$$
, где $i = \overline{1, n}$; $j = \overline{1, m}$; $r = \overline{1, m}$, где n – число объектов наблюдения; m – число показателей, характеризующих объект; r – число значимых общих факторов; x_{ij} – центрированное значение j -го показателя (переменной) у i -го объекта исследования; f_r – r -й общий фактор; v_j – j -й характерный (индивидуальный) фактор, присущий только данной j -й переменной; a_{jr} – весовой коэффициент j -й переменной на r -м общем факторе; d_j – весовой коэффициент j -й переменной на j -м характерном факторе.
 1. Модель факторного анализа;
 2. Математическая модель метода главных компонент;
 3. Мультипликативная модель;
 4. Аддитивная модель.
- 6) Мультиколлинеарность факторных переменных – это:
 1. Отсутствие связи между факторными переменными;
 2. Тесная связь между факторными переменными;
 3. Многомерная связь между факторными переменными;
 4. Множественная регрессионная модель.
- 7) Элементами матрицы факторного отображения A по формуле $A = V \times \Lambda^{1/2}$ являются:
 1. Линейные коэффициенты корреляции между исходными переменными и

главными компонентами

3. Множественные коэффициенты корреляции между исходными переменными и главными компонентами

8) *Формула расчета расстояния Махаланобиса* имеет следующий вид::

$$1. d(X_i, X_j) = \sqrt{(X_i - X_j)^T \Delta^T \Sigma^{-1} \Delta (X_i - X_j)};$$

$$2. d(X_i, X_j) = \sqrt{\sum_{k=1}^p (x_i^{(k)} - x_j^{(k)})^2};$$

$$3. d(X_i, X_j) = \sum_{s=1}^p |x_i^{(s)} - x_j^{(s)}|;$$

$$4. p_{\min}(S_l, S_m) = \min_{X_i \in S_l, X_j \in S_m} d(X_i, X_j)$$

9) При использовании метода *k*-средних для классификации многомерных объектов в состав кластера включаются новые объекты таким образом, чтобы внутриклассовая дисперсия:

1. Стремилась к минимуму;
2. Стремилась к максимуму;
3. Оставалась постоянной.

10) Дискриминантный анализ – совокупность статистических методов многомерной классификации объектов при наличии:

1. Средних значений;
2. «Обучающих» выборок;
3. «Обычных» выборок;

4.2.4. Вопросы для самоподготовки

1. Показатели и объекты (измерения). Интервальные данные.
2. Нечисловые данные.
3. Простейшие статистические характеристики.
4. Приведение к нормальной форме.
5. Оцифровка нечисловых данных.
6. Роль и сущность многомерных статистических методов в экономике, управлении, финансах, социальных науках: постановка основных задач, примеры практического использования в социально-экономических исследованиях.
7. Многомерное нормальное распределение как основная модель современных многомерных статистических методов.
8. Практическое применение многомерных методов в финансовых, экономических и социальных исследованиях.
9. Методы статистического оценивания многомерных параметров и проверки гипотез.
10. Особенности анализа количественных и качественных признаков. Методы шкалирования.
11. Постановка основных прикладных задач классификации многомерных наблюдений.
12. Классификация с обучением и без обучения. Сущность методов классификации.
13. Кластерный анализ. Меры однородности объектов. Расстояния между объектами. Расстояния между кластерами.
14. Реализация методов кластерного анализа в современных пакетах прикладных программ.
15. Кластерный анализ финансовой деятельности предприятий.
16. Кластерный анализ мировой демографической статистики.
17. Кластерный анализ социологических опросов.
18. Кластерный анализ результатов аттестации персонала компании.
19. Зависимость выбора метода классификации от цели исследования.
20. Компонентный анализ. Математическая модель главных компонент. Геометрическая интерпретация главных компонент. Формирование названий главных компонент.
21. Экономическая интерпретация главных компонент.
22. Реализация методов компонентного анализа в современных пакетах прикладных программ.

23. Использование компонентного анализа в экономических и социальных исследованиях.
24. Линейная модель факторного анализа. Различие предпосылок компонентного и факторного анализа. Экономическая интерпретация
25. Реализация методов факторного анализа в современных пакетах прикладных программ. Использование факторного анализа в экономических и социальных исследованиях.
26. Введение в теорию качественных признаков и нечисловой информации
27. Роль и сущность статистики нечисловой информации в экономике, управлении, финансах, социальных науках.
28. Числовые (интервальная, отношений и абсолютная) и нечисловые (номинальная и порядковая) шкалы измерений. Дихотомическая шкала.

4.3. Оценочные средства для промежуточной аттестации.

Таблица 5

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-2	Способен обосновывать подходы, используемые в бизнес-анализе, руководить и управлять бизнес-анализом с использованием информационно-коммуникационных технологий	ПКс-2.1	Способен использовать современные методы, информационные технологии, программный инструментарий в объеме, необходимом для решения задач бизнес-аналитики, использовать англоязычную документацию и справочные системы

Таблица 6

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ПКс-2.1	Демонстрирует способность использовать современные методы статистической обработки и анализа данных, информационные технологии и программный инструментарий для решения задач бизнес-аналитики.	Полнота и качество решение поставленных задач, выполнения заданий, кейсов. Полнота ответа на вопросы на экзамене

Для оценки сформированности компетенций, знаний и умений, соответствующих данным компетенциям, используются контрольные вопросы, а также задачи, при решении которых необходимо продемонстрировать навыки владения современными информационными системами.

Комплект практических заданий в электронном виде размещен в локальной сети вуза и на портале дистанционного обучения .

Типовые вопросы, выносимые на экзамен

1. Предмет, метод и задачи курса «Математические методы статистической обработки и анализа данных».

2. Основные условия применения многомерного статистического анализа в социальноэкономических исследованиях.
3. Общая теория многомерных распределений.
4. Виды шкал. Количественные и качественные данные.
5. Робастность в многомерном статистическом анализе.
6. Одномерная группировка и одномерноецензурирование.
7. Таблицы сопряженности.
8. Многомерное цензурирование.
9. Выявление аномальности в многомерных совокупностях
10. Многомерное нормальное распределение.
11. Свойства многомерного нормального распределения.
12. Устойчивость многомерного нормального распределение относительно линейных преобразований.
13. Проблема размерности в многомерных исследованиях.
14. Многомерные методы оценивания и статистического сравнения.
15. Множественный корреляционно-регрессионный анализ.
16. Линейная множественная регрессионная модель.
17. Выбор адекватного уравнения регрессии.
18. Понятие о нелинейной регрессии.
19. Корреляционный анализ количественных связей и порядковых переменных.
20. Оценки частных и множественных коэффициентов корреляции.
21. Непараметрические показатели связи.
22. Сущность и алгоритм дисперсионного анализа.
23. Расчет внутригрупповой и межгрупповой вариации.
24. Множественный дисперсионный анализ.
25. Методы снижения размерности.
26. Модель, математическое обоснование и алгоритм метода главных компонент.
27. Получение и использование матриц факторов, индивидуальных значений главных компонент.
28. Экономическая интерпретация главных компонент
29. Метод главных факторов.
30. Регрессия на главные компоненты.
31. Линейная модель факторного анализа.
32. Экономическая интерпретация общих факторов.
33. Методы классификации без обучения.
34. Иерархические кластер-процедуры.
35. Функционалы качества разбиения на классы
36. Метод К-средних.
37. Классификация в пространстве главных компонент и общих факторов.
38. Методы классификации с обучением.
39. Линейный дискриминантный анализ при известных параметрах многомерного нормального закона распределения.
40. Вероятность ошибочной классификации с помощью дискриминантной функции.
41. Оценка качества дискриминантной функции и информативности отдельных признаков.
42. Пошаговый дискриминантный анализ.
43. Математическое описание метода дискриминантного анализа.
44. Модель метода канонических корреляций.
45. Применение многомерного статистического анализа в экономических исследованиях.
46. Интерпретация канонического коэффициента корреляции и соответствующих канонических величин.
47. Модель ковариационного анализа с одним фактором и одной сопутствующей переменной
48. Применение множественного ковариационного анализа в экономических

исследованиях

Шкала оценивания.

Оценка результатов производится на основе Положения о текущем контроле успеваемости обучающихся и промежуточной аттестации обучающихся по образовательным программам высшего образования в АНОВО ИСН.

Схема расчетов сформирована в соответствии с учебным планом направления, согласована с руководителем научно-образовательного направления, утверждена деканом факультета. Схема расчетов доводится до сведения студентов на первом занятии по данной дисциплине и является составной частью рабочей программы дисциплины и содержит информацию по изучению дисциплины для обучающихся в АНОВО ИСН

4.4. Методические материалы

При изучении дисциплины используются файлы, находящиеся в ресурсах сети Института.

5. Методические указания для обучающихся по освоению дисциплины

Рабочей программой дисциплины предусмотрены следующие виды аудиторных занятий: лекции, практические занятия. На лекциях рассматривается наиболее сложный материал дисциплины. Лекция сопровождается презентациями, компьютерными текстами лекции, что позволяет магистранту самостоятельно работать над повторением и закреплением лекционного материала. Для этого магистранту должно быть предоставлено право самостоятельно работать в компьютерных классах в сети Интернет.

Практические занятия предназначены для самостоятельной работы магистрантов по решению конкретных задач проектирования и использования информационных систем. Каждое практическое занятие сопровождается домашними заданиями, выдаваемыми магистрантам для решения во внеаудиторное время. Для оказания помощи в решении задач имеются тексты практических заданий с условиями задач и вариантами их решения.

С целью активизации самостоятельной работы магистрантов на портале дистанционного обучения разработан учебный курс «Математические методы статистической обработки и анализа данных», включающий набор файлов с текстами лекций, практикума, примерами задач, а также набором тестов для организации электронного тестирования.

Для активизации работы магистрантов во время контактной работы с преподавателем отдельные занятия проводятся в интерактивной форме. В основном, интерактивная форма занятий обеспечивается при проведении занятий в компьютерном классе. Интерактивная форма обеспечивается наличием разработанных файлов с заданиями, наличием контрольных вопросов, возможностью доступа к системе дистанционного обучения, а также к тестеру.

Для работы с печатными и электронными ресурсами имеется возможность доступа к электронным ресурсам. Организация работы студентов с электронной библиотекой указана на сайте института.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Основная литература

1. Буре В.М., Парилина Е.М. Теория вероятностей и математическая статистика. – СПб.: Лань, 2016. – 416 с. Электронный ресурс «Электронно-библиотечная система издательства «Лань» <http://e.lanbook.com/>. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=10249
2. Голубева Н.В. Математическое моделирование систем и процессов. – СПб.: Лань, 2017. – 192 с. Электронный ресурс «Электронно-библиотечная система издательства «Лань» <http://e.lanbook.com/>. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=4862.
3. Туганбаев А.А., Крупин В.Г. Теория вероятностей и математическая статистика. – СПб.: Лань, 2011. – 320 с. Электронный ресурс «Электронно-библиотечная система издательства «Лань» <http://e.lanbook.com/>. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=652.

6.2. Дополнительная литература

1. Боровиков, Владимир Павлович. Прогнозирование в системе Statistica в среде Windows : основы теории и интенсивная практика на компьютере : учеб. пособие для вузов, рек. М-вом образования Рос. Федерации / В. П. Боровиков, Г. И. Ивченко. - М. : Финансы и статистика, 2000. - 378 с.
2. Кацко И. А., Паклин Н. Б. Практикум по анализу данных на компьютере. – М.: КолосС, 2009. - 278 с.
3. Ларсон Б. Разработка Бизнес-аналитики в Microsoft SQL Server 2005. – М.: Питер, 2008.
4. Матвеев М.Г., Свиридов А.С., Алейникова Н.А. Модели и методы искусственного интеллекта. Применение в экономике.: [учебное пособие]. М.:Финансы и статистика, Изд.дом ИНФРА, 2014.
5. Наумов, Владимир Николаевич. Средства бизнес-аналитики : учеб. пособие / В. Н. Наумов ; Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2016. - 107 с.
6. Тихомиров, Николай Петрович. Методы эконометрики и многомерного статистического анализа : учебник, рек. М-вом образования и науки Рос. Федерации / Н. П. Тихомиров, Т. М. Тихомирова, О. С. Ушмаев. - М. : Экономика, 2011. - 637 с.
7. Халафян А.А. STATISTICA 6. Статистический анализ данных. – М.: ООО «Бином-Пресс», 2007.
8. Андерсон К. Аналитическая культура. От сбора данных до бизнес-результатов. – М.: Манн, Иванов и Фербер, 2017.

6.3. Учебно-методическое обеспечение самостоятельной работы

Приказ Минобрнауки России от 19.11.2013 N 1259 (ред. от 05.04.2016) "Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам подготовки научно-педагогических кадров в магистрантуре (адъюнктуре)" (Зарегистрировано в Минюсте России 28.01.2014 N 31137)

6.4. Нормативные правовые документы

Не используются

6.5. Интернет-ресурсы

Электронно-образовательные ресурсы на сайте научной библиотеки

Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс» http://www.nwapa.spb.ru/index.php?page_id=76

1. Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»

- http://www.nwapa.spb.ru/index.php?page_id=76
2. Электронные учебники электронно-библиотечной системы (ЭБС) «IPRbooks»
http://www.nwapa.spb.ru/index.php?page_id=76
 3. Электронные учебники электронно-библиотечной системы (ЭБС) «Юрайт»
http://www.nwapa.spb.ru/index.php?page_id=76
 4. Научно-практические статьи по экономике и финансам Электронной библиотеки ИД «Гребенников» http://www.nwapa.spb.ru/index.php?page_id=76
 5. Статьи из журналов и статистических изданий Ист-Вью
http://www.nwapa.spb.ru/index.php?page_id=76
 6. Англоязычные ресурсы EBSCO Publishing: доступ к мультидисциплинарным полнотекстовым базам данных различных мировых издательств по бизнесу, экономике, финансам, бухгалтерскому учету, гуманитарным и естественным областям знаний, рефератам и полным текстам публикаций из научных и научно-популярных журналов.
 7. Emerald eJournals Premier - крупнейшее мировое издательство, специализирующееся на электронных журналах и базах данных по экономике и менеджменту.

Возможно использование, кроме вышеперечисленных ресурсов, и других электронных ресурсов сети Интернет.

1. www.finexpert.ru
2. <http://www.itnews.ru/>
3. <http://www.cnews.ru/>
4. <http://www.prj-exp.ru/>
5. <http://piter-consult.ru/>
6. [http://www.gartner.com /](http://www.gartner.com/)
7. <http://www.idc.com>
8. <http://bpms.ru / BPMS.ru>
9. <http://www.betec.ru />
10. <http://www.cfin.ru / Интернет-проект «Корпоративный менеджмент»>
11. <http://www.osp.ru / Открытые системы>
12. <http://www.citforum.ru / CIT forum>
13. <http://www.iteam.ru / Портал iTeam – Технологии корпоративного управления>
14. <http://www.idef.com / Методологии IDEF>
15. <http://www.interface.ru/home.asp?artId=4449 / Электронная версия книги Дэвид А. Марка, Клемент МакГоуэн Методология структурного анализа и проектирования SADT.>
16. <http://www.fa.ru/dep/cko/msq/Pages/default.aspx / Международные стандарты качества.>
17. <http://office.microsoft.com/ru-ru/support/FX100996114.aspx / Microsoft Visio>
18. <http://wf.runa.ru/rus / СУБП RunaWFE>
19. <http://www.bizagi.com / Bizagi>
20. <http://www.businessstudio.ru / Business Studio>
21. http://www.casewise.com/ru/products/corporate_modeler_suite.html / Casewise Corporate Modeler Suite
22. <http://www.interface.ru/home.asp?artId=106 / Process Modeler>
23. <http://www.oracle.com/technetwork/ru/middleware/bpa/index.html / Oracle Business Process Analysis Suite>
24. <http://www.softwareag.com/ru / Software AG>
25. <http://www-03.ibm.com/software/products/ru/ru/wbi / IBM WebSphere Business>

Modeler

26. <http://www.consultant.ru/> СПС Консультант Плюс

27. [http://www.garant.ru /](http://www.garant.ru/) СПС Гарант

28. <http://www.consultant.ru/> СПС Консультант Плюс

29. [http://www.garant.ru /](http://www.garant.ru/) СПС Гарант

6.6. Иные источники.

Не используются.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Все практические занятия проводятся в компьютерном классе. Учебная дисциплина включает использование программного обеспечения Microsoft Excel, Microsoft Word, Microsoft Power Point для подготовки текстового и табличного материала, графических иллюстраций. Для формирования практических навыков - MathCad, Matlab, SPSS.

Методы обучения с использованием информационных технологий (компьютерное тестирование, демонстрация мультимедийных материалов).

Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).