

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
математического анализа

С.А. Шабров

25.05.2023

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.В.07 Статистический анализ данных массовых источников

Код и наименование дисциплины в соответствии с учебным планом

1. Код и наименование направления подготовки/специальности:

02.03.01 Математика и компьютерные науки

2. Профиль подготовки/специализация:

*Математические методы и компьютерные технологии в естествознании, экономике
и управлении*

3. Квалификация выпускника: бакалавр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра математического анализа

6. Составители программы: Бахтина Жанна Игоревна, канд. физ.-мат. наук, доцент, доцент кафедры математического анализа

7. Рекомендована: Научно-методическим Советом математического факультета, протокол от 25.05.2023 №0500-06

8. Учебный год: 2026/2027

Семестр(ы): 7

9. Цели и задачи учебной дисциплины

Обучающиеся должны знать виды массовых источников; основы математической статистики; различные методы статистического анализа данных; понятие массовых источников; виды анализа: качественный и количественный; этапы, формы и виды статистического наблюдения и методы обработки данных массовых источников.

Обучающиеся должны уметь работать с различными видами источников, уметь их классифицировать и обрабатывать.

Обучающиеся должны владеть навыками обработки массовых источников методами математической статистики.

В результате изучения дисциплины студенты должны обладать информационной компетентностью: должны знать понятие источника, математические основы статистических методов, уметь обрабатывать различные виды информации, в том числе и посредством вычислительной техники, владеть информационными технологиями, участвовать в информационных процессах; коммуникативной компетенцией: способностью вступать в контакт, высказывать свою точку зрения, вести дискуссию, быть способным решать коммуникативные задачи; учебно-познавательной компетентностью: быть способным самостоятельно планировать свою деятельность, обладать способностью самореализации, обладать способностью к самообразованию, участвовать в проектно-исследовательской деятельности; социальной и политической компетентностью: быть готовым жить и работать в обществе, понимать права как важнейший механизм регулирования социальных и межличностных отношений.

Целями освоения учебной дисциплины являются:

- изучение основных видов массовых источников и методов статистического анализа;
- развитие способности к самостоятельному поиску информации, а также методов ее обработки в соответствии с поставленной задачей;
- формирование у будущих специалистов системы компетенций в области теоретических основ практических методов анализа качественных и количественных данных.

Задачи учебной дисциплины:

- создание базовой теоретической основы и элементарных навыков, необходимых для поиска оптимальных методов обработки информации;
- овладение начальными навыками статистического анализа данных массовых источников;
- научить решать научно-технические задачи;
- научить уметь применять знания фундаментальных и прикладных дисциплин, решать задачи с помощью вычислительной техники, обладать информационной компетентностью, способностью к анализу и синтезу, поиску методов обработки и реализации этих методов при анализе данных.

10. Место учебной дисциплины в структуре ООП:

Дисциплина «Статистический анализ данных массовых источников» относится к учебным дисциплинам вариативной части блока Б1 основной образовательной программы направления подготовки 02.03.01 Математика и компьютерные науки – Бакалавр.

Быстрое развитие общества требуют от студента не знания конкретных методов и средств обработки информации, а базовых знаний и умения адаптироваться их к предложенным ему методам решения задачи. Поэтому главная цель учебного процесса – заложить базовые знания и научить осваивать предложенное программное обеспечение.

Студенты должны знать основы математической статистики, различные методы статистического анализа данных, понятие массовых источников, виды анализа: качественный и количественный, этапы, формы и виды статистического наблюдения и методы обработки данных массовых источников: вариационный, корреляционный анализ данных, анализ динамики и моделирование ряда распределения, методы многомерного анализа массовых источников, методы обработки массовых источников посредством вычислительной техники.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ПК-2	Способен оформлять результаты научно-исследовательских работ	ПК-2.1	Знает основные стандарты, нормы и правила оформления результатов научно-исследовательских работ	Знать: - основные понятия и определения курса Уметь: - находить схожесть и отличие в понятиях Владеть: - навыком обобщения результатов нескольких утверждений.
		ПК-2.2	Умеет четко ставить задачи и грамотно формулировать выводы по результатам исследования	Знать: - определения и утверждения, предшествующие данной теореме Уметь: - выбрать ранее изученные факты, на которых необходимо строить доказательство рассматриваемой теоремы Владеть: - навыком строгого обоснования шагов доказательства через ранее доказанные факты и определения
		ПК-2.3	Имеет практический опыт в оформлении результатов научно-исследовательской деятельности в математике и информатике	Знать: - основные стандарты, нормы и правила оформления результатов научно-исследовательских работ Уметь: - четко ставить задачи и грамотно формулировать выводы по результатам исследования Владеть:

				- навыками оформления результатов научно-исследовательской деятельности в математике и информатике
ПК-3	Способен решать задачи аналитического характера, предполагающие выбор и многообразие способов решения в области естествознания, экономики и управления	ПК-3.1	Знает современные методы разработки и реализации математических моделей	Знать: - принципы и этапы построения научной работы, способы научной аргументации Уметь: - выбирать подходящие методы решения задач и представлять научные результаты в различных форматах, исследования устойчивости моделей, описывающих реальные процессы Владеть: - навыками дискуссии, способностью публично представлять собственные и известные научные результаты
		ПК-3.2	Умеет разрабатывать математические модели в области естествознания, экономики и управления, а также реализовывать алгоритмы математических моделей на базе пакетов прикладных программ моделирования	

12. Объем дисциплины в зачетных единицах/час. (в соответствии с учебным планом) — 4/144.

Форма промежуточной аттестации (зачет/экзамен) экзамен.

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость		
		Всего	По семестрам	
			7 семестр	№ семестра
Аудиторные занятия		68	68	
в том числе:	лекции	34	34	
	практические	34	34	
	лабораторные			
Самостоятельная работа		40	40	
в том числе: курсовая работа (проект)				
Форма промежуточной аттестации (экзамен – 7 час.)		36	36	
Итого:		144	144	

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
1. Лекции		
1.1	Массовые источники	Рассматриваются основные понятия, подходы к изучению. Классификация и методология изучения массовых источников. Массовые источники в экономике.
1.2	Основы регрессионного анализа	Рассматриваются основы регрессионного анализа, применительно к массовым источникам. Массовые источники во времени.
1.3	Многомерный статистический анализ	Методы многомерного анализа, методы классификации данных. Кластерный анализ, факторный анализ.
1.4	Экономические индексы	Методы обработки статистической информации, понятие статистических индексов
1.5	Методы анализа текста	Частотный анализ, контент анализ. Методы анализа текста.
2. Практические занятия		
2.1	Основы статистического анализа	Рассматривается понятие статистического наблюдения, этапы, методы, программно-методологические вопросы, методы дескриптивной статистики.
2.2	Основы регрессионного анализа	Рассматриваются основы регрессионного анализа, применительно к массовым источникам. Массовые источники во времени.
2.3	Многомерный статистический анализ	Методы многомерного анализа, методы классификации данных. Кластерный анализ, факторный анализ.
2.4	Экономические индексы	Методы обработки статистической информации, понятие статистических индексов.

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)					
		Лекции	Практические	Лабораторные	Самостоятельная работа	Контроль	Всего
1	Массовые источники	6			2	6	14
2	Основы статистического анализа	2	6		6	6	20
3	Основы регрессионного анализа	6	8		8	6	28
4	Многомерный статистический анализ	8	12		8	6	34
5	Экономические индексы	6	8		8	6	28
6	Методы анализа текста	6			8	6	20
	Итого:	34	34		40	36	144

14. Методические указания для обучающихся по освоению дисциплины

При изучении дисциплины на лекционных занятиях рассматриваются основные понятия массовых источников. Студенты должны знать основы математической статистики, различные методы статистического анализа данных, понятие массовых источников, виды анализа: качественный и количественный, этапы, формы и виды статистического наблюдения и методы обработки данных массовых источников: вариационный, корреляционный анализ данных, анализ динамики и моделирование ряда распределения, методы многомерного анализа массовых источников, методы обработки массовых источников посредством вычислительной техники.

При решении задач необходимо использовать программные средства с целью выбора оптимального подхода при решении поставленных задач.

По темам № 2,3,5 обучающимся необходимо выполнить лабораторную работу, на основании которой формируется отчет. Данные работы являются формами контроля уровня усвоения материала по теме и уровня сформированности компетенций.

Лабораторные работы рассчитаны, что большую часть обучающиеся выполняют самостоятельно, таким образом, самостоятельная работа является важной составной частью процесса обучения.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Крылов, В. Е. Статистика : учебник / В. Е. Крылов, И. Б. Тесленко, Н. В. Муравьева. — Москва : Прометей, 2022. — 438 с. — ISBN 978-5-00172-312-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/265013
2	Годин, А. М. Статистика : учебник / А. М. Годин. — 13-е изд. — Москва : Дашков и К, 2022. — 412 с. — ISBN 978-5-394-04491-5. — Текст : электрон-

ный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/277529

б) дополнительная литература:

№ п/п	Источник
3	Буре, В. М. Теория вероятностей и математическая статистика : Учебник. / В. М. Буре, Е. М. Парилина. — СПб. : Издательство «Лань», 2013. — 416 с. — <URL : http://e.lanbook.com/view/book/10249/
4	Плотников, А. Н. Элементарная теория анализа и статистическое моделирование временных рядов : Учебное пособие / А. Н. Плотников. — СПб. : Издательство «Лань», 2015. — 224 с. — <URL : http://e.lanbook.com/view/book/65051/
5	Прикладные методы теории вероятностей : Учебник / А. А. Свешников [и др.] ; под ред. О. И. Зайца. — СПб. : Издательство «Лань», 2012. — 480 с. — <URL : http://e.lanbook.com/view/book/3184/
6	Туганбаев, А. А. Теория вероятностей и математическая статистика : учебное пособие / А. А. Туганбаев, В. Г. Крупин. — СПб. : Издательство «Лань», 2011. — 224 с. — <URL : http://e.lanbook.com/view/book/652/
7	Боровиков, В. П. Прогнозирование в системе STATISTICA в среде Windows : Основы теории и интенсивная практика на компьютере: Учебное пособие / В.П. Боровиков, Г.И. Ивченко. — М. : Финансы и статистика, 1999. — 382 с. — ISBN 5-279-01980-1 : 42.25.
8	Владимирова, Л. П. Прогнозирование и планирование в условиях рынка : учебное пособие / Л.П. Владимирова. — Изд. 5-е, перераб. и доп. — М. : Дашков и Ко, 2005. — 398 с. — ISBN 5-94798-613-2
9	Шепитько, Г. Е. Социальное прогнозирование и моделирование : учебное пособие для студ., обуч. по специальности "Финансы и кредит" / Г. Е. Шепитько. — М. : АЭБ МВД России, 2005. — 270 с. — ISBN 5-93479-019-6.

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
10	http://www.lib.vsu.ru – официальный сайт библиотеки ВГУ
11	http://www.math.vsu.ru – официальный сайт математического факультета ВГУ
12	Поисковые системы www.google.ru , www.yandex.ru
13	www.intuit.ru
14	http://lanbook.com/ebs.php Электронно-библиотечная система "Издательство "Лань"
15	http://ibooks.ru/home.php

16. Перечень учебно-методического обеспечения для самостоятельной работы

Целью самостоятельной работы обучаемых является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности. Самостоятельная работа учащихся способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня.

№ п/п	Источник
-------	----------

1	Буре, В. М. Теория вероятностей и математическая статистика : / Буре В. М., Парилина Е. М. — Москва : Лань, 2013 .— Допущено УМО по классическому университетскому образованию в качестве учебника для студентов вузов, обучающихся по направлениям ВПО 010400 — «Прикладная математика и информатика» и 010300 — «Фундаментальная информатика и информационные технологии» .— ISBN 978-5-8114-1508-3 .— <URL: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=10249 >.
---	---

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

Осуществляется интерактивная связь с преподавателем через сеть интернет, проводятся индивидуальные онлайн консультации. Практические занятия ведутся с привлечением мультимедийных технологий.

Перечень необходимого программного обеспечения : Microsoft Windows Server 2008, Microsoft Windows 10 Enterprise 64 bit, Android, программный пакет LibreOffice 6 (*Calc (электронные таблицы)*).

18. Материально-техническое обеспечение дисциплины:

Для проведения практических занятий требуется аудитория-компьютерный класс на группу студентов, оборудованная маркерной и интерактивной доской, мультимедийным проектором с экраном, компьютер преподавателя и персональные компьютеры слушателей с подключением к Internet, с установленным на них программными пакетами LibreOffice.Calc.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1.	Массовые источники	ПК-3	ПК - 3.1 ПК - 3.2	Промежуточная аттестация – экзамен, Устный опрос, контрольно-измерительные материалы к экзамену
2.	Основы статистического анализа.	ПК-2	ПК – 2.1 ПК – 2.2 ПК – 2.3	Промежуточная аттестация – экзамен, Лабораторная работа 1, контрольно-измерительные материалы к экзамену
3.	Основы регрессионного анализа.	ПК-2	ПК – 2.1 ПК – 2.2 ПК – 2.3	Промежуточная аттестация – экзамен, Лабораторная работа 2, контрольно-измерительные материалы к экзамену
4.	Многомерный статистический анализ.	ПК-2	ПК – 2.1 ПК – 2.2 ПК – 2.3	Промежуточная аттестация – экзамен, Лабораторная работа 3, контрольно-измерительные материалы к экзамену
5.	Экономические	ПК-2	ПК – 2.1	Промежуточная аттестация –

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
	индексы.		ПК – 2.2 ПК – 2.3	экзамен, Лабораторная работа 4, контрольно-измерительные материалы к экзамену
6.	Методы анализа текста.	ПК-3	ПК-3.1 ПК-3.2	Промежуточная аттестация – экзамен, Устный опрос, контрольно-измерительные материалы к экзамену
Промежуточная аттестация форма контроля - экзамен				Перечень вопросов к экзамену

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

Оценка знаний, умений и навыков, характеризующая этапы формирования компетенций в рамках изучения дисциплины осуществляется в ходе текущей и промежуточной аттестаций.

Текущая аттестация проводится в форме выполнения индивидуальных лабораторных работ.

ЛАБОРАТОРНАЯ РАБОТА:

РЕШЕНИЕ ЗАДАЧ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ СРЕДСТВАМИ LIBREOFFICE.CALC

LI-

Задачи работы

- Применение стандартных функций LibreOffice.Calc для решения задач описательной статистики. Ввод исходных данных. Вычисления размаха (вариации), оценки среднего, среднеквадратичного отклонения и дисперсии, асимметрии и эксцесса, квартилей. Построение таблицы частот и гистограммы.
- Применение пакета анализа для решения задач описательной статистики. Генерация случайных чисел. Сглаживание данных. Описательная статистика, построение гистограмм.

Задание 1

Даны значения прибыли, млн. руб., для 75 предприятий:

24,8 33,4 29,3 32,7 30,4 25,9 30,0 30,2 26,7 32,3 29,6 34,0 30,7 29,4 32,0 30,7 28,0 31,5 29,6
33,5 28,0 26,4 38,4 30,5 28,0 29,9 34,2 36,0 30,7 28,0 28,0 28,6 26,1 30,7 33,1 37,3 32,3 30,0
34,1 25,5 25,5 31,5 31,8 36,4 31,8 25,5 33,1 25,5 28,2 31,6 33,8 31,9 32,2 28,0 26,4 30,2 31,2
33,4 32,2 34,4 30,5 32,3 30,8 30,0 37,0 24,2 33,5 35,3 33,7 30,6 31,8 35,4 29,5 32,8 27,4.

Рассчитайте основные характеристики: количество выборок, среднее, среднеквадратичное отклонение, дисперсия, медиана, мода, асимметрия, эксцесс, наименьшее, наибольшее, размах, квартиль.

Охарактеризуйте выборку.

Постройте гистограмму (постройте таблицу частот выполнив группировку данных). Для этого рассчитайте минимальное количество интервалов, ширину интервала, и построьте таблицу частот. На основании получившейся таблицы строится гистограмма.

Проанализируйте влияние числа интервалов группирования на гистограмму для трех значений числа интервалов: минимального, в полтора раза и вдвое больше минимального. Проведите аппроксимацию и сглаживание построенных гистограмм, сравнение полиномиальных трендов для трех различных значений числа интервалов группирования проведите для степени аппроксимирующего полинома 2.

ЛАБОРАТОРНАЯ РАБОТА: МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

- Применение стандартных функций LibreOffice.Calc для решения задач прогнозирования.
- Применение пакета анализа для решения задач парной линейной регрессии.

Задача 1

Приводятся данные за 2000 год по территориям Северо-Западного федерального округа

Таблица.

Территории Северо-Западного федерального округа	Оборот розничной торговли за год, млрд. руб.	Общая сумма доходов населения за год, млрд. руб.
А	Y	X
1.Респ. Карелия	9,4	19,1
2.Респ. Коми	16,7	37,3
3.Архангельская обл.	16,3	30,0
4.Вологодская обл.	12,1	27,5
5.Калининградская обл.	14,0	19,0
6.Ленинградская обл.	15,6	26,2
7.Мурманская обл.	20,5	39,5
8.Новгородская обл.	9,3	14,8
9.Псковская обл.	7,3	11,6
10.г. Санкт-Петербург	83,1	133,6

Задание:

1. Расположите территории по возрастанию фактора X. Сформулируйте рабочую гипотезу о возможной связи Y и X. Постройте поле корреляции и сформулируйте гипотезу о возможной форме и направлении связи.
2. Рассчитайте параметры b_1 и b_0 парной линейной функции $y_x = b_0 + b_1 * x$ при помощи метода наименьших квадратов.
3. Оцените тесноту связи с помощью показателей корреляции (r) и детерминации (r^2), проанализируйте их значения.
4. Используя критерий Стьюдента оцените статистическую значимость коэффициентов регрессии и корреляции. Надёжность уравнений в целом оцените через F-критерий Фишера для уровня значимости $\alpha=0,05$.
5. Постройте интервальные оценки параметров регрессии. Проверьте, согласуются ли полученные результаты с выводами, полученными в предыдущем пункте.

6. Оцените полученные результаты, проинтерпретируйте полученное уравнение регрессии. Рассчитайте коэффициент эластичности.
7. По уравнению регрессии рассчитайте теоретические значения результата (\hat{Y}), по ним постройте теоретическую линию регрессии и определите скорректированную среднюю ошибку аппроксимации - ϵ'_{cp} , оцените её величину.

Задача 2

В течении 30 недель проводился анализ продаж. В таблице приведены данные на конец недели.

1. Постройте график, определить наличие линии тренда и определите скользящее среднее.
2. Выполните экспоненциальное сглаживание с константой сглаживания $\alpha = 0,1$
3. Найдите оптимальную константу экспоненциального сглаживания.
4. Выполните экспоненциальное сглаживание с полученной константой сглаживания α .
5. Какой из построенных наборов сглаженных данных подходит наилучшим образом для определения асимптотического ряда?

вари- ант неделя	1	2	3	4	5	6	7	8	9	10	11
1	33	43	24	272	1795	1523	1630	622	666	53	72
2	35	44	260	281	1738	1549	1659	620	670	33	73
3	37	45	265	289	1934	1576	1689	621	676	30	77
4	40	42	267	291	1835	1602	1720	630	684	29	81
5	38	47	269	296	2024	1630	1749	636	696	55	78
6	43	48	268	299	2083	1659	1778	650	705	44	79
7	44	49	272	302	1341	1689	1807	666	707	41	87
8	45	53	281	306	987	1720	1837	670	718	43	94
9	42	56	289	310	1650	1749	1865	676	731	68	93
10	47	60	291	315	2074	1778	1892	684	745	55	84

ЛАБОРАТОРНАЯ РАБОТА: ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ

Задание

Имеются данные по ценам в базисный и текущий период.

Товары	Базисный период		Текущий период	
	Цена (руб.)	Количество (шт.)	Цена (руб.)	Количество (шт.)
1	24	816	24	800
2	38	123	38	120
3	51	456	51	450

Рассчитайте следующие показатели.

1. Рассчитайте индивидуальные и общие индексы цены, физического объема и стоимости. Сделайте выводы о произошедших в текущем периоде изменениях.
2. Рассчитайте абсолютные изменения стоимости, возникшие в результате изменения

цен и физического объема.

3. Рассчитайте индексы переменного, фиксированного составов, индекс структурных сдвигов. Сделайте выводы о причинах, приведших к изменению средней цены совокупности товаров.

4. Рассчитайте абсолютные изменения средней цены, объясните их.

Примерные задачи:

Задача

Имеются данные по ценам в базисный и текущий период.

Товары	Базисный период		Текущий период	
	Цена (руб.)	Количество (шт.)	Цена (руб.)	Количество (шт.)
1	24	816	24	800
2	38	123	38	120
3	51	456	51	450

Рассчитайте следующие показатели.

1. Рассчитайте индивидуальные и общие индексы цены, физического объема и стоимости. Сделайте выводы о произошедших в текущем периоде изменениях.

2. Рассчитайте абсолютные изменения стоимости, возникшие в результате изменения цен и физического объема.

Задача

Провести корреляционно регрессионный анализ в зависимости выплаты труда от производительности.

№	1	2	3	4	5	6	7	8
y	23,6	31,6	35,2	27,1	23,6	25,7	38,2	17,8
x	70	79	91	94	72	81	96	45

Задача

Провести прогнозирование методом экспоненциального сглаживания для $\alpha=0$

T	1	2	3	4	5	6	7	8
X	11	7	6	12	10	9	8	-
f_{k+1}								

Задача

По данным о средних доходах на конечное потребление за десять лет, которые представлены в таблице, оцените наличие тренда и в случае положительного ответа постройте трендовую модель.

Год (t)	1-й	2-й	3-й	4-й	5-й	6-й	7-й	8-й	9-й	10-й
Расходы (y_t)	7	8	8	10	11	12	14	16	17	19

Задача

Рассмотрим объемы продаж в течение недели:

День недели	Понедельник	Вторник	Среда	Четверг	Пятница	Суббота	Воскресенье
Количество проданной продукции	11	7	6	12	10	9	8

Провести сглаживание методом скользящей средней.

Задача

Даны значения прибыли, млн. руб., для 75 предприятий:

24,8 33,4 29,3 32,7 30,4 25,9 30,0 30,2 26,7 32,3 29,6 34,0 30,7 29,4 32,0 30,7 28,0 31,5 29,6
 33,5 28,0 26,4 38,4 30,5 28,0 29,9 34,2 36,0 30,7 28,0 28,0 28,6 26,1 30,7 33,1 37,3 32,3 30,0
 34,1 25,5 25,5 31,5 31,8 36,4 31,8 25,5 33,1 25,5 28,2 31,6 33,8 31,9 32,2 28,0 26,4 30,2 31,2
 33,4 32,2 34,4 30,5 32,3 30,8 30,0 37,0 24,2 33,5 35,3 33,7 30,6 31,8 35,4 29,5 32,8 27,4.

Рассчитайте основные характеристики: количество выборок, среднее, среднеквадратичное отклонение, дисперсия, медиана, мода, асимметрия, эксцесс, наименьшее, наибольшее, размах, квартиль.

Охарактеризуйте выборку. Постройте гистограмму.

Задача

Найти уравнение линейной регрессии, описывающее корреляционную связь между размерами и доходом помещичьего имения

	1	2	3	4	5	6	7	8	9	10
x_i	240	255	265	270	285	295	310	320	325	330
y_i	1,5	1,25	1,55	1,4	1,45	1,6	1,8	1,8	1,85	1,9

Задача

Определить степень корреляционной зависимости между доходом и размерами помещичьего хозяйства в России на рубеже XIX-XX вв. по сведениям о размерах (в десятинах) и доходах (в тыс. руб.) десяти помещичьих имений.

	1	2	3	4	5	6	7	8	9	10
x_i	240	255	265	270	285	295	310	320	325	330
y_i	1,5	1,25	1,55	1,4	1,45	1,6	1,8	1,8	1,85	1,9

Задача

По урожайности зерновых культур 10 колхозов определить среднюю и предельную ошибки выборки и оценить пределы для генеральной средней.

Исходные данные (x_i , $i = 1, \dots, 10$ - урожайность зерновых в центнерах с гектара) и промежуточные вычисления можно записать в таблице:

1	2	3	4	5	6	7	8	9	10
6,5	6,2	5,4	9,3	7,2	8,4	4,3	6	6,3	7,4

Задача

Даны две группы людей, возраст которых (в годах):

1 группа: 27; 29; 30; 31; 31; 32;

2 группа: 13; 14; 14; 15; 61; 63.

Рассчитать средние характеристики и меры вариации признаков для обоих рядов. Провести сравнительный анализ. Характеризует ли среднее ряд?

При оценивании используются качественные шкалы оценок. Критерии оценивания приведены ниже.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств:

Промежуточная аттестация включает в себя теоретические вопросы, позволяющие оценить уровень полученных знаний и решение задачи в LibreOffice.Calc, позволяющую оценить степень сформированности умений и навыков.

Собеседование по экзаменационным билетам:

1. Понятие массового источника. Различные подходы к определению массовых источников.
2. Классификация массовых источников.
3. Массовые источники в экономике.
4. Методология изучения массовых источников. Общие методы анализа массовых источников.
5. Виды информации. Виды экономической информации.
6. Статистическая сводка. Группировка. Представление статистических данных: статистические графики и таблицы.
7. Генеральная совокупность и выборка. Определение объема выборки.
8. Средние величины. Показатели вариации. Характеристики закономерности рядов распределения.
9. Ошибки выборки.
10. Уравнение линейной регрессии. Метод наименьших квадратов.
11. Корреляционный анализ.
12. Прогнозирование.
13. Понятие временного ряда. Его характеристики.
14. Компоненты тендсезонного временного ряда.
15. Метод Четверикова.
16. Методы сглаживания ряда. Методы выявления временного тренда.
17. Автокорреляция и прогнозирование.
18. Методы многомерного анализа.
19. Кластерный анализ.
20. Метод к-средних.
21. Методы многомерной группировки данных.
22. Контент анализ.
23. Факторный анализ.
24. Методы анализа текста.

Для оценивания результатов обучения на экзамене используются следующие показатели:

- 1) Знание видов массовых источников,
- 2) Знание основ математической статистики, различные методы статистического анализа данных,
- 3) Знание понятия массовых источников, виды анализа: качественный и количественный, этапы, формы и виды статистического наблюдения и методы обработки данных массовых источников,
- 4) Умение работать с различными видами источников, умение их классифицировать и обрабатывать,
- 5) Владеть навыками обработки массовых источников методами математической статистики.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Обучающийся в полной мере владеет понятийным аппаратом данного курса (теоретическими основами дисциплины), полностью выполняет практические задания.	Повышенный уровень	отлично
Обучающийся владеет основными понятиями данного курса (теоретическими основами дисциплины), способен выполнять практические задания, допускает ошибки в ответе, которые исправляются по наводящим вопросам преподавателя.	Базовый уровень	хорошо

Обучающийся владеет частично теоретическими основами дисциплины, фрагментарно способен выполнять практические задания	Пороговый уровень	удовлетворительно
Ответ на контрольно-измерительный материал не соответствует перечисленным выше показателям. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки	Ниже порогового уровня	неудовлетворительно

20.3 Фонд оценочных средств сформированности компетенций студентов, рекомендуемый для проведения диагностических работ

1) закрытые задания (тестовые, средний уровень сложности):

1. По 10 предприятиям имеются следующие данные

Численность	800	950	1100	1400	1750	520	1000	400	2300	2000
Выпуск продукции	5	4,2	6,5	8	10	1,8	6	2	9,5	11

Для вычисления коэффициента рангов составлена вспомогательная таблица

№ п/п	Выпуск продукции x	Численность продуктов y	Ранги		Разность рангов d	Квадрат разности рангов d ²
			по x	по y		
1	5	800	4	3	1	1
2	4,2	950	3	4	-1	1
3	6,5	1100	6	6	0	0
4	8	1400	7	7	0	0
5	10	1750	9	8	1	1
6	1,8	520	1	2	-1	1
7	6	1000	5	5	0	0
8	2	400	2	1	1	1
9	9,5	2300	8	10	-2	4
10	11	2000	10	9	1	1
Итого						10

По этим данным вычислен коэффициент рангов и определен тип связи между выпуском продукции и численностью работников.

Ответы

- $\rho = 0,64$, связь заметная,
- (правильный ответ) $\rho = 0,94$, связь весьма тесная,
- $\rho = 0,85$, связь высокая,
- $\rho = 0,15$, связь слабая.

2. Охарактеризуйте вид ряда распределений коммерческих фирм по величине уставного капитала

Группы фирм по величине уставного капитала, млн.руб	Число фирм	Удельный вес фирм в % по итогу
До 9,0	4	13,3
9,0-14,0	5	16,7
14,0-19,0	10	33,3

19,0-24,0	6	20,0
24,0 и более	5	16,7

Ответы

- а) атрибутивный;
 б) (правильный ответ) интервальный вариационный;
 в) вариационный дискретный.

3. В таблице представлена информация о заработной плате рабочих цеха за месяц в тыс.руб

Заработная плата (x)	100	200	300	400	500	600	700
Кол-во (f)	2	5	20	30	15	10	5

Среднюю арифметическую можно вычислить по формуле:

Ответы

а) $\bar{x} = \frac{\sum_{i=1}^n x_i}{f}$;

б) (правильный ответ) $\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}$;

в) $\bar{x} = \sum_{i=1}^n \frac{x_i}{f_i}$.

4. Вычислено следующее уравнение множественной регрессии по сельскохозяйственным предприятиям области, характеризующее связь между размерами предприятия по посевной площади x га, числу коров z шт., доходами предприятия y млн. руб.:
 $y = 290 + 3,12x + 0,65z$.

Верно ли сделан вывод о параметрах уравнения регрессии: Увеличение посевной площади на один гектар приводит к повышению доходов в среднем на 3,12 млн. руб.; увеличение числа коров на 1 голову приводит к повышению доходов на 0,65 млн. руб.

Ответы

- а) (правильный ответ) верно;
 б) неверно.

5. Изучалось влияние на влажность вафельного листа y времени выдержки листа в печи x_1 , температуры печи x_2 и влажности теста x_3 . Проведено 20 наблюдений:

№	y	x ₁	x ₂	x ₃	№	y	x ₁	x ₂	x ₃
1	3,1	2,5	180	63	11	2,9	3	180	63
2	3,4	2,5	180	64	12	3,0	3	180	64
3	3,5	2,5	180	65	13	3,1	3	180	65
4	3,2	2,5	180	63	14	2,8	3	180	63
5	3,3	2,5	180	64	15	2,9	3	180	64
6	3,4	2,5	200	65	16	2,9	3	200	65
7	3,2	2,5	200	63	17	2,7	3	200	63
8	3,3	2,5	200	64	18	2,8	3	200	64
9	3,4	2,5	200	65	19	2,9	3	200	65
10	3,2	2,5	200	63	20	2,8	3	200	63

Построена модель множественной линейной регрессии, предполагая наличие линейной связи между влажностью вафельного листа и тремя указанными факторами: $y = -1,0506 - 0,84x_1 - 0,0041x_2 + 0,1132x_3$. Проведен регрессионный анализ с помощью инструмента Регрессия из пакета Анализ данных.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	№	y	x ₁	x ₂	x ₃	ВЫВОД ИТОГОВ							
2	1	3,1	2,5	180	63	Регрессионная статистика							
3	2	3,4	2,5	180	64	Множественный R 0,971647							
4	3	3,5	2,5	180	65	R-квадрат 0,944097							
5	4	3,2	2,5	180	63	Нормированный R-квадрат 0,933615							
6	5	3,3	2,5	180	64	Стандартная ошибка 0,063056							
7	6	3,4	2,5	200	65	Наблюдения в выборке 20							
8	7	3,2	2,5	200	63	Дисперсионный анализ							
9	8	3,3	2,5	200	64		df	SS	MS	F	Значимость F		
10	9	3,4	2,5	200	65	Регрессия	3	1,074382	0,358127	90,06996	3,1E-10		
11	10	3,2	2,5	200	63	Остаток	16	0,063618	0,003976				
12	11	2,9	3	180	63	Итого	19	1,138					
13	12	3	3	180	64	Коэффициенты стандартной ошибки					P-Значения	Вероятности 95%	
14	13	3,1	3	180	65	Y-пересечение	-1,05059	1,104471	-0,95121	0,355645	-3,39196	1,29078	
15	14	2,8	3	180	63	Коэффициент при X 1	-0,84	0,056399	-14,8938	8,5E-11	-0,95956	-0,7204	
16	15	2,9	3	180	64	Коэффициент при X 2	-0,00413	0,00142	-2,90947	0,010238	-0,00714	-0,0011	
17	16	2,9	3	200	65	Коэффициент при X 3	0,113235	0,017099	6,622507	5,86E-06	0,076988	0,14948	
18	17	2,7	3	200	63								
19	18	2,8	3	200	64								
20	19	2,9	3	200	65								
21	20	2,8	3	200	63								
22													

Определить, значима ли модель и значимы ли все факторы?

Ответы

- а) (правильный ответ) модель значима, все факторы значимы;
- б) модель не значима, все факторы значимы;
- в) модель значима, не все факторы значимы.

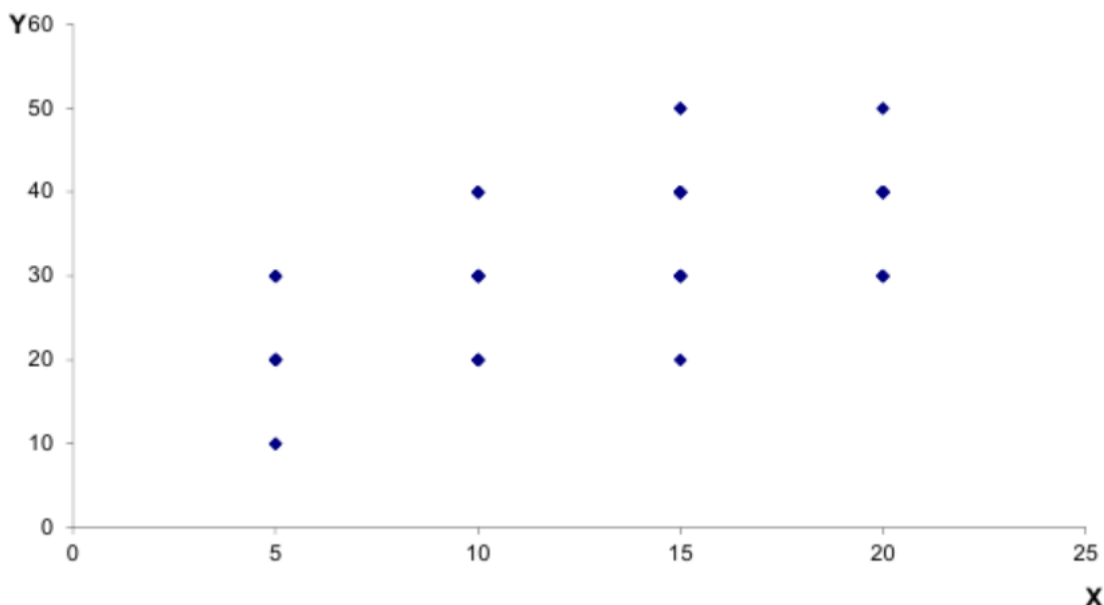
Решение:

Модель значима (см. проверку значимости по F-критерию: $3,14E-10 \ll 0,05$), все факторы также значимы: это следует из того, что все р-значения для переменных меньше, чем 0,05, коэффициент детерминации 0,944, что свидетельствует о хорошем качестве модели. Если бы некоторые из факторов (регрессоров) оказались незначимы, можно было бы попытаться построить новую модель, удалив их из нее.

6. Таблица значений признака Y при данных значениях признака X имеет вид:

Y	X				n_y
	5	10	15	20	
10	2	-	-	-	2
20	5	4	1	-	10
30	3	8	6	3	20
40	-	3	6	6	15
50	-	-	2	1	3
n_x	10	15	15	10	$n = 50$

Корреляционное поле данной двумерной выборки приведено на рисунке



Какой вывод можно сделать о зависимости между величинами по виду поля корреляции?

Ответы

- а) (правильный ответ) зависимость между величинами существует;
- б) зависимость между величинами отсутствует;
- в) нельзя установить зависимость между величинами.

7. Пусть случайная величина X имеет нормальное распределение с математическим ожиданием a и дисперсией σ^2 . Рассмотрена основная гипотеза: $H_0: a = 0, \sigma^2 = 1$. Какие из перечисленных гипотез могут быть выдвинуты в качестве альтернативных?

Ответы

- а) (правильный ответ) $H_1: a = 0, \sigma^2 = 2$;
- б) (правильный ответ) $H_1: a \neq 0, \sigma^2 = 1$
- в) $H_1: a = 2, \sigma^2 = 3$;
- г) $H_1: a \neq 0, \sigma^2 \neq 1$.

8. Участковый терапевт хочет проверить, есть ли различие между частотой биения сердца (кол-во ударов в минуту) курящих и некурящих студентов. Отобраны две группы: курящие $n_1 = 26$ чел., и некурящие $n_2 = 18$ чел. По данным этих выборок найдены исправленные выборочные дисперсии $S_1^2 = 36$ и $S_2^2 = 10$. По выборке вычислено значение

статистики $F_{\text{набл}} = 3,6$ и вычислена критическая точка $F_{\text{кр}} = 2,19$. Проверить при уровне значимости $\alpha = 0,05$, прав ли участковый терапевт.

Ответы

а) (правильный ответ) гипотезу о равенстве генеральных дисперсий необходимо отвергнуть;

б) гипотезу о равенстве генеральных дисперсий необходимо принять.

Решение:

$$H_0 : \sigma_1^2 = \sigma_2^2 ; H_1 : \sigma_1^2 \neq \sigma_2^2 .$$

По выборке вычисляем значение статистики: $F_{\text{набл}} = \frac{S_1^2}{S_2^2} = \frac{36}{10} = 3,6$.

По таблице, по уровню значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n_1 - 1 = 25$; $k_2 = n_2 - 1 = 17$, находим критическую точку $F_{\text{кр}}(0,05; 25; 17) = 2,19$.

Так как $F_{\text{набл}} > F_{\text{кр}}$, то необходимо отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются значимо.

9. Урожайность сахарной свеклы в сельхозпредприятиях Воронежской области характеризуется следующими показателями: $\bar{x} = 300$ ц/га, $\sigma = 120$ ц/га, $M_0 = 240$ ц/га, а урожайность зерновых — $\bar{x} = 30$ ц/га; $\sigma = 10$ ц/га, $M_0 = 27$ ц/га. Показатели асимметрии и направление асимметрии?

Ответы

1. (правильный ответ) $\Pi_{A_1} = 0,5$, $\Pi_{A_1} = 0,3$, асимметрия правосторонняя

2. $\Pi_{A_1} = -0,5$, $\Pi_{A_1} = -0,3$, асимметрия левосторонняя

3. $\Pi_{A_1} = 0,4$, $\Pi_{A_1} = 0,6$, асимметрия правосторонняя

4. $\Pi_{A_1} = 0,2$, $\Pi_{A_1} = -0,4$, асимметрия правосторонняя

11. Для номинального признака А, имеющего 5 градаций, и номинального признака В, имеющего 4 градации, составлена таблица сопряженности и вычислено значение статистики хи-квадрат. Значение статистики оказалось равным 23,13. Согласно таблицам, квантили распределения хи-квадрат $\chi_{12,095}^2 = 21,026$, $\chi_{12,099}^2 = 26,217$. Какой (какие) выводы можно сделать из полученных результатов:

Ответы

1. на уровне значимости 0,05 гипотезу о независимости признаков А и В следует отвергнуть,

2. (правильный ответ) на уровне значимости 0,05 гипотезу о независимости признаков А и В следует принять,

3. на уровне значимости 0,01 нет оснований для отклонения гипотезы о независимости признаков А и В,

4. (правильный ответ) на уровне значимости 0,01 гипотезу о независимости признаков А и В следует отвергнуть.

12. Гистограмма применяется для графического изображения:

Ответы

а) дискретных рядов распределения;

- б) (правильный ответ) интервальных рядов распределения;
- в) ряда накопленных частот;
- г) прерывного ряда распределения.

13. Размах вариации исчисляется как

Ответы

- а) (правильный вариант) разность между максимальным и минимальным элементами;
- б) разность между первым и последним членом ряда распределения;
- в) сумма первого и последнего членов ряда.

14. Для проверки значимости коэффициентов уравнения регрессии используют критерий

Ответы

- а) (правильный ответ) Стьюдента;
- б) Чупрова;
- в) Колмогорова;
- г) Фишера;
- д) Пирсона.

15. Знак «+» или «-» у коэффициента корреляции указывает на

Ответы

- а) (правильный ответ) направление связи;
- б) тесноту связи;
- в) наличие связи;
- г) форму связи.

16. Оценка значимости множественного уравнения регрессии осуществляется с помощью

Ответы

- а) коэффициента корреляции;
- б) (правильный ответ) F-критерия Фишера;
- в) коэффициента детерминации;
- г) коэффициента регрессии.

17. Известно, что коэффициент корреляции случайных величин X и Y равен нулю. Это означает, что

Ответы

- а) (правильный ответ) линейная связь между случайными величинами X и Y отсутствует;
- б) случайные величины X и Y линейно зависимы;
- в) случайные величины X и Y зависимы;
- г) случайные величины X и Y независимы;
- д) (правильный ответ) если случайный вектор X и Y является гауссовским, то случайные величины X и Y независимы.

2) открытые задания (тестовые, повышенный уровень сложности):

1. В таблице представлено распределение студентов заочного отделения по возрасту

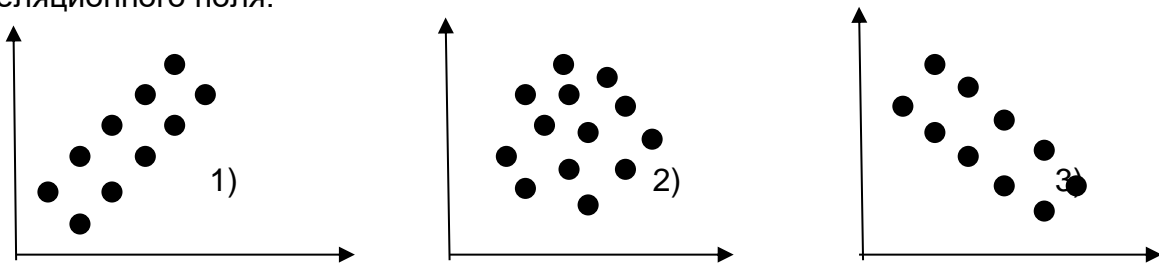
Возрастные группы	Число студентов	Накопленные частоты
-------------------	-----------------	---------------------

До 20 лет	345	346
20-25	872	1218
25-30	1054	2272
30-35	781	3053
35-40	212	3265
40-45	121	3386
45 лет и выше	76	3462
Итого	3462	

Определить медиану. Ответ округлите до целого числа лет

Ответ 27

2. Наличие корреляции приблизительно может быть определено с помощью корреляционного поля.



Ответы

Соотнесите рисунок и тип корреляционной зависимости.

1)

2)

3)

а) отсутствие корреляции;

б) сильная положительная корреляция;

в) сильная отрицательная корреляция.

Решение:

1-б; 2-в; 3-а.

3. По данным о месячной заработной плате 10 рабочих трех разных профессий (токарь, слесарь и кузнец) вычислены: общая дисперсия заработной платы $\sigma_0^2 = 1636$ и средняя из внутригрупповых дисперсий $\overline{\sigma^2} = 1140$. Вычислить корреляционное отношение. Ответ округлите до тысячных.

Ответ 0,303

Решение:

Корреляционное отношение вычисляется по формуле

$$\eta^2 = \frac{\delta^2}{\sigma_0^2}.$$

Следовательно, сначала необходимо найти межгрупповую дисперсию

$$\delta^2 = \sigma_0^2 - \overline{\sigma^2} = 1636 - 1140 = 496.$$

Подставляя это значение в вышеприведенную формулу, получим:

$$\eta^2 = \frac{\delta^2}{\sigma_0^2} = \frac{496}{1636} = 0,303.$$

Критерии и шкалы оценивания заданий ФОС:

1) Задания закрытого типа (выбор одного варианта ответа, верно/неверно):

- 1 балл – указан верный ответ;
- 0 баллов – указан неверный ответ.

2) Задания закрытого типа (множественный выбор):

- 2 балла – указаны все верные ответы;
- 0 баллов — указан хотя бы один неверный ответ.

3) Задания закрытого типа (на соответствие):

- 2 балла – все соответствия определены верно;
- 0 баллов – хотя бы одно сопоставление определено неверно.

4) Задания открытого типа (короткий текст):

- 2 балла – указан верный ответ;
- 0 баллов – указан неверный ответ.

5) Задания открытого типа (число):

- 2 балла – указан верный ответ;
- 0 баллов – указан неверный ответ.

Задания раздела 20.3 рекомендуются к использованию при проведении диагностических работ с целью оценки остаточных результатов освоения данной дисциплины (знаний, умений, навыков).