

Министерство науки и высшего образования Российской Федерации
Департамент координации деятельности организаций
в сфере сельскохозяйственных наук
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Волгоградский государственный аграрный университет»
Эколого-мелиоративный факультет
наименование факультета

УТВЕРЖДАЮ
Декан эколого-мелиоративного
наименование факультета

_____ О.А. Корчагина _____
подпись *инициалы фамилия*

_____ Г.
дата



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.ДВ.5.2 Анализ данных
индекс и наименование дисциплины

Кафедра Информационные системы и технологии
наименование кафедры
Уровень высшего образования бакалавриат
бакалавриат / специалитет / магистратура
Направление подготовки (специальность) 09.03.03 Прикладная информатика
шифр и наименование направления подготовки
Направленность (профиль) Прикладная информатика в экономике
наименование направленности (профиля) программы
Форма обучения очная/заочная
очная / заочная
Год начала реализации образовательной программы 2017

Волгоград
2022

Автор(ы):

доцент

должность

подпись

Стрижакова Е.А.

инициалы фамилия

Рабочая программа дисциплины согласована с руководителем основной профессиональной образовательной программы высшего образования по направлению подготовки (специальности)

09.03.03 Прикладная информатика

шифр и наименование направления подготовки (специальности)

Прикладная информатика в экономике

наименование направленности (профиля) программы

Заведующий кафедрой

должность

подпись

О.В. Кочеткова

инициалы фамилия

Рабочая программа дисциплины обсуждена и одобрена на заседании кафедры

Информационные системы и технологии

наименование кафедры

Протокол № 2 от 20 октября 2022 г.

Заведующий кафедрой

подпись

О.В. Кочеткова

инициалы фамилия

Рабочая программа дисциплины обсуждена и одобрена на заседании методической комиссии эколого-мелиоративного факультета

наименование факультета

Протокол № 2 от 25 октября 2022 г.

дата

Председатель

методической комиссии факультета

подпись

А.К. Васильев

инициалы фамилия

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

Целью изучения дисциплины является ознакомление студентов с методами количественного выражения взаимосвязей экономических процессов и явлений и освоение методов анализа информации и прогнозирования развития экономических процессов.

Изучение дисциплины направлено на решение следующих задач:

- овладение совокупностью математико-статистических методов, используемых для количественной оценки экономических явлений и процессов;
- обучение построению эконометрических моделей, параметры которых оцениваются средствами математической статистики;
- овладение математическим аппаратом, помогающим анализировать, моделировать и решать прикладные экономические задачи.

Изучение дисциплины направлено на формирование общепрофессиональных компетенций, а также знаний, умений и навыков, необходимых для решений профессиональных задач в научно-исследовательской деятельности.

Индекс компетенции	Содержание компетенции	Планируемые результаты
ОПК-2	способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования	Знать основные распределения вероятностей и области их применения; основы математико-статистических методов анализа данных
		Уметь применять вероятностные методы анализа; вычислять описательные статистические показатели; осуществлять выбор инструментальных средств анализа данных в соответствии с поставленной задачей
		Владеть методами вычисления вероятностей; навыками математико-статистического исследования эмпирических данных; техникой проверки статистических гипотез; навыками использования пакетов прикладных программ анализа данных
ПК-23	способностью применять системный подход и математические методы в формализации решения прикладных задач	Знать методы анализа зависимости между переменными
		Уметь исследовать зависимости и строить регрессионные модели; осуществлять выбор инструментальных средств статистического анализа данных
		Владеть основными приемами регрессионного анализа; навыками использования пакетов прикладных программ статистической направленности

2 Место дисциплины в структуре образовательной программы

Дисциплина «Анализ данных» (Б1.В.ДВ.5.2) входит в вариативную часть блока 1 «Дисциплины (модули)» ОПОП ВО подготовки бакалавров по направлению 09.03.03 «Прикладная информатика» (профиль «Прикладная информатика в экономике»).

Условием изучения дисциплины является освоение таких дисциплин программы подготовки бакалавров, как «Дискретная математика» (Б1.Б.6), «Математика» (Б1.Б.5), «Теория вероятностей и математическая статистика» (Б1.Б.9). Минимальные требования к «входным» знаниям, необходимым для успешного освоения данной дисциплины: удовлетворительное усвоение программ по указанным выше дисциплинам. Дисциплина «Эконометрика», в свою очередь, дает знания и умения, которые являются необходимыми для успешного освоения дисциплин «Математическое и имитационное моделирование» (Б1.В.ОД.6), «Интеллектуальные информационные системы» (Б1.В.ОД.12), выполнения научно-исследовательской работы (Б2.П.2).

3 Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Очная форма обучения

Вид учебной работы		Всего часов	Распределение часов по семестрам			
			4	...		
Контактная работа обучающихся с преподавателем в части аудиторных занятий, всего		36	36			
Лекции (Л)		18	18			
Практические занятия (ПЗ) / Семинары (С)		18	18			
Лабораторные работы (ЛР)						
Самостоятельная работа обучающихся, всего		72	72			
Курсовой проект (КП)						
Курсовая работа (КР)						
Расчетно-графическая работа (РГР)						
Реферат (Реф)						
Самостоятельное изучение разделов и тем		72	72			
Вид промежуточной аттестации*	зачет	0	0			
	зачет с оценкой					
	экзамен					
Общая трудоемкость	часов	108	108			
	зачетных единиц	3	3			

Заочная форма обучения

Вид учебной работы		Всего часов	Распределение часов по курсам			
			2	...		
Контактная работа обучающихся с преподавателем в части аудиторных занятий, всего		12	12			
Лекции (Л)		4	4			
Практические занятия (ПЗ) / Семинары (С)		8	8			
Лабораторные работы (ЛР)						
Самостоятельная работа обучающихся, всего		92	92			
Курсовой проект (КП)						
Курсовая работа (КР)						
Расчетно-графическая работа (РГР)						
Реферат (Реф)						
Контрольная работа (КРЗ)		14	14			
Самостоятельное изучение разделов и тем		78	78			
Вид промежуточной аттестации*	зачет	4	4			
	зачет с оценкой					
	экзамен					
Общая трудоемкость	часов	108	108			
	зачетных единиц	3	3			

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

4.1 Содержание лекций

III Содержание лекций			
№ п/п	Тема лекции	Объем, ч	
		Форма обучения	
		Очная	Заочная
Раздел 1. Основы математико-статистических методов анализа данных			
1	Законы распределения вероятностей.	2	2
2	Числовые характеристики случайных величин.	4	
3	Основы проверки статистических гипотез.	2	
Раздел 2. Корреляционный и регрессионный анализ.			
4	Корреляционный анализ.	2	2
5	Регрессионный анализ.	4	
6	Дисперсионный анализ.	4	
ВСЕГО		18	4

4.2 Практические (семинарские) занятия

№ п/п	Тема практического (семинарского) занятия	Объем, ч	
		Форма обучения	
		Очная	Заочная
Раздел 1. Основы математико-статистических методов анализа данных			
1	Числовые характеристики случайных величин.	6	4
2	Основы проверки статистических гипотез.	2	
Раздел 2. Корреляционный и регрессионный анализ.			
3	Оценка параметров уравнения парной линейной регрессии. Проверка качества уравнения регрессии.	6	4
4	Сравнение нескольких средних методом дисперсионного анализа.	4	
ВСЕГО		18	8

4.3 Лабораторные работы

Не предусмотрены

4.4 Перечень тем для самостоятельного изучения

№ п/п	Тема для самостоятельного изучения	Объем, ч	
		Форма обучения	
		Очная	Заочная
Раздел 1. Основы математико-статистических методов анализа данных			
1	Законы распределения вероятностей.	10	10
2	Числовые характеристики случайных величин.	14	16
3	Основы проверки статистических гипотез.	10	10
Раздел 2. Корреляционный и регрессионный анализ.			
4	Корреляционный анализ.	10	10
5	Регрессионный анализ.	16	18
6	Дисперсионный анализ.	12	14
Всего		72	78

4.5 Другие виды самостоятельной работы

№ п/п	Содержание самостоятельной работы	Объем, ч	
		Форма обучения	
		Очная	Заочная
1	Подготовка и написание контрольной работы	-	14
ВСЕГО		-	14

5 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

Для самостоятельной работы обучающихся по дисциплине рекомендуется следующая учебно-методическая литература:

1. Вуколов, Э.А. Основы статистического анализа. Практ. по стат. мет. и исслед. операций с исп. пакетов STATISTICA и EXCEL [Электронный ресурс]: учебное пособие / Э.А.Вуколов. – Электрон. текстовые дан. - 2 изд., испр. и доп. - М.: Форум: НИЦ Инфра-М, 2013. - 464 с. – Режим доступа: <http://znanium.com/bookread.php?book=369689>

2. Дайитбеков, Д.М. Компьютерные технологии анализа данных в эконометрике [Электронный ресурс]: монография / Д.М. Дайитбеков. – Электрон. текстовые дан. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с. - Режим доступа: <http://znanium.com/bookread2.php?book=365692>

1. Количественный анализ в экономике и менеджменте [Электронный ресурс]: учебник / В.А. Малугин, Л.Н. Фадеева. – Электрон. текстовые дан. - М.: НИЦ ИНФРА-М, 2014. – 615 с. - Режим доступа: <http://znanium.com/bookread.php?book=363305#none>

3. Мхитарян, В.С. Теория вероятностей и математическая статистика [Электронный ресурс]: учеб. пособие / В. С. Мхитарян, Е. В. Астафьева, Ю. Н. Миронкина, Л. И. Трошин; под ред. В. С. Мхитаряна. – Электрон. текстовые дан. - 2-е изд., перераб. и доп. - М.: Московский финансово-промышленный университет «Синергия», 2013. – Режим доступа: <http://znanium.com/bookread.php?book=451329>

4. Статистический анализ данных в MS Excel [Электронный ресурс]: учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. – Электрон. текстовые дан. - М.: ИНФРА-М, 2014. - 320 с. Режим доступа: <http://znanium.com/bookread.php?book=429722>

5. Шапкин, А. С. Задачи с решениями по высшей математике, теории вероятностей, математической статистике, математическому программированию [Электронный ресурс]: учебное пособие для бакалавров / А. С. Шапкин, В. А. Шапкин. – Электрон. текстовые дан. - 8-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2013. - 432 с. – Режим доступа: <http://znanium.com/bookread.php?book=430613>

6. Оценочные материалы для проведения промежуточной аттестации обучающихся по дисциплине (фонд оценочных средств)

6.1 Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Перечень компетенций, на освоение которых направлена дисциплина

Индекс компетенции	Содержание компетенции
ОПК-2	способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования
ПК-23	способностью применять системный подход и математические методы в формализации решения прикладных задач

Этапы формирования компетенций в результате изучения дисциплины в процессе освоения образовательной программы

Участвующие в формировании компетенций дисциплины, модули, практики		Форма обучения	Курсы обучения				
Индекс	Наименование		1 курс	2 курс	3 курс	4 курс	5 курс
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования							
Б1.Б.4	Экономическая теория	Очная	+				
		Заочная		+	+		
Б1.Б.5	Математика	Очная	+				
		Заочная	+				
Б1.Б.6	Дискретная математика	Очная	+				
		Заочная	+				
Б1.Б.7	Теория систем и системный анализ	Очная		+			
		Заочная		+			

Б1.Б.9	Теория вероятностей и математическая статистика	Очная	+				
		Заочная		+			
Б1.В.ОД.5	Исследование операций и методы оптимизации	Очная		+	+		
		Заочная		+	+		
Б1.В.ОД.6	Математическое и имитационное моделирование	Очная			+		
		Заочная				+	
Б1.В.ОД.12	Интеллектуальные информационные системы	Очная		+	+		
		Заочная			+	+	
Б1.В.ДВ.4.1	Численные методы	Очная		+			
		Заочная			+		
Б1.В.ДВ.4.2	Математические методы в инженерных и экономических расчетах	Очная		+			
		Заочная			+		
Б1.В.ДВ.5.1	Модели рискованных ситуаций в экономике	Очная		+			
		Заочная		+			
Б1.В.ДВ.5.2	Анализ данных	Очная		+			
		Заочная		+			
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач							
Б1.Б.6	Дискретная математика	Очная	+				
		Заочная	+				
Б1.Б.7	Теория систем и системный анализ	Очная		+			
		Заочная		+			
Б1.Б.9	Теория вероятностей и математическая статистика	Очная	+				
		Заочная		+			
Б1.В.ОД.5	Исследование операций и методы оптимизации	Очная		+	+		
		Заочная		+	+		
Б1.В.ОД.6	Математическое и имитационное моделирование	Очная			+		
		Заочная				+	
Б1.В.ОД.12	Интеллектуальные информационные системы	Очная		+	+		
		Заочная			+	+	
Б1.В.ДВ.1.1	Эконометрика	Очная		+			
		Заочная	+				
Б1.В.ДВ.1.2	Математическая экономика	Очная		+			
		Заочная	+				
Б1.В.ДВ.4.1	Численные методы	Очная		+			
		Заочная			+		
Б1.В.ДВ.4.2	Математические методы в инженерных и экономических расчетах	Очная		+			
		Заочная			+		
Б1.В.ДВ.5.1	Модели рискованных ситуаций в экономике	Очная		+			
		Заочная		+			
Б1.В.ДВ.5.2	Анализ данных	Очная		+			
		Заочная		+			
Б2.У.1	Практика по получению первичных профессиональных умений и навыков	Очная	+				
		Заочная	+				
Б2.П.1	Практика по получению профессиональных умений и опыта профессиональной деятельности	Очная			+		
		Заочная			+		
Б2.П.2	Научно-исследовательская работа	Очная				+	
		Заочная				+	
Б2.П.4	Преддипломная практика	Очная				+	
		Заочная					+

Основными этапами формирования указанных компетенций при освоении дисциплины является последовательное изучение содержательно связанных между собой модулей (разделов, тем). Изучение каждого модуля (раздела, темы) предполагает овладение обучающимися необходимыми компетенциями. Результат аттестации на различных этапах формирования компетенций показывает уровень освоения их обучающимися.

Этапы формирования компетенций в процессе изучения дисциплины

Контролируемые модули / разделы / темы дисциплины	Оценочные средства по этапам формирования компетенций	
	Текущий контроль	Промежуточная аттестация
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования		
Раздел 1. Основы математико-статистических методов анализа данных	Контрольная работа	Зачет
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач		
Раздел 2. Корреляционный и регрессионный анализ.	Контрольная работа, индивидуальное домашнее задание	Зачет

6.2 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

6.2.1 Текущий контроль

Показатели оценивания компетенций на различных этапах их формирования в процессе изучения дисциплины

Контролируемые модули / разделы / темы дисциплины	Показатели оценивания компетенций	
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования		
Раздел 1. Основы математико-статистических методов анализа данных	Знает	Основные распределения вероятностей и области их применения; основы математико-статистических методов анализа данных
	Умеет	Применять вероятностные методы анализа; вычислять описательные статистические показатели; осуществлять выбор инструментальных средств анализа данных в соответствии с поставленной задачей
	Владеет	Методами вычисления вероятностей; навыками математико-статистического исследования эмпирических данных; техникой проверки статистических гипотез; навыками использования пакетов прикладных программ анализа данных
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач		
Раздел 2. Корреляционный и регрессионный анализ.	Знает	Методы анализа зависимости между переменными
	Умеет	Исследовать зависимости и строить регрессионные модели; осуществлять выбор инструментальных средств статистического анализа данных
	Владеет	Основными приемами регрессионного анализа; навыками использования пакетов прикладных программ статистической направленности

Шкала и критерии оценивания формируемых компетенций в процессе изучения дисциплины, соотнесенные с этапами их формирования

Контролируемые модули / разделы / темы дисциплины	Форма оценочного средства	Шкала оценивания	Критерии оценки
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования			
Раздел 1. Основы математико-статистических методов анализа данных	Контрольная работа	«Отлично» (17-20 баллов)	Студент в полном объеме выполнил все задания (или ответил на все поставленные вопросы), проявив самостоятельность и знания межпредметного характера.
		«Хорошо» (12-16 баллов)	Студент выполнил задания, и в них содержатся недочеты или одна не грубая ошибка; при ответе на поставленные вопросы имел незначительные замечания и поправки со стороны преподавателя.
		«Удовлетворительно» (6-11 баллов)	Студент выполнил задания более чем на 50 % и работа содержит недочеты или две-три негрубые ошибки или две грубые ошибки; при ответе на поставленные вопросы имел значительные замечания и поправки со стороны преподавателя.
		«Неудовлетворительно» (0-5 баллов)	Студент выполнил работу менее чем на 50 % или работа содержит более двух грубых ошибок; при ответе на поставленные вопросы преподаватель оказывал ему постоянную помощь.
			Студент показал полное незнание вопроса, отказался отвечать или не приступил к выполнению работы.
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач			
Раздел 2. Корреляционный и регрессионный анализ.	Контрольная работа	«Отлично» (17-20 баллов)	Студент в полном объеме выполнил все задания (или ответил на все поставленные вопросы), проявив самостоятельность и знания межпредметного характера.
		«Хорошо» (12-16 баллов)	Студент выполнил задания, и в них содержатся недочеты или одна не грубая ошибка; при ответе на поставленные вопросы имел незначительные замечания и поправки со стороны преподавателя.
		«Удовлетворительно» (6-11 баллов)	Студент выполнил задания более чем на 50 % и работа содержит недочеты или две-три негрубые ошибки или две грубые ошибки; при ответе на поставленные вопросы имел значительные замечания и поправки со стороны преподавателя.
		«Неудовлетворительно» (0-5 баллов)	Студент выполнил работу менее чем на 50 % или работа содержит более двух грубых ошибок; при ответе на поставленные вопросы преподаватель оказывал ему постоянную помощь.
			Студент показал полное незнание вопроса, отказался отвечать или не приступил к выполнению работы.

	Индивидуальное домашнее задание	«Отлично» (8-10 баллов)	Глубокое и полное овладение содержанием учебного материала, понятийным аппаратом, умение связывать теорию с практикой, решать практические задачи, высказывать и обосновывать свои суждения. Отличная отметка предполагает грамотное, логическое изложение ответа, качественное внешнее оформление.
		«Хорошо» (5-7 баллов)	Студент полно освоил учебный материал, владеет понятийным аппаратом, ориентируется в изученном материале, осознанно применяет знания для решения практических задач, грамотно излагает ответ, но содержание, форма ответа имеют отдельные неточности.
		«Удовлетворительно» (2-4 балла)	Студент обнаруживает знание и понимание основных положений учебного материала, но излагает его неполно, непоследовательно, допускает ошибки в определении понятий, в применении знаний для решения практических задач, не умеет доказательно обосновать свои суждения.
		«Неудовлетворительно» (0-1 балл)	Студент демонстрирует разрозненные, бессистемные знания или полное незнание и непонимание учебного материала, беспорядочно и неуверенно излагает материал, не может применять знания для решения практических задач.

6.2.2 Промежуточная аттестация

Показатели оценивания компетенций в результате изучения дисциплины
в процессе освоения образовательной программы

Показатели оценивания компетенций	
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования	
Знает	Основные распределения вероятностей и области их применения; основы математико-статистических методов анализа данных
Умеет	Применять вероятностные методы анализа; вычислять описательные статистические показатели; осуществлять выбор инструментальных средств анализа данных в соответствии с поставленной задачей
Владеет	Методами вычисления вероятностей; навыками математико-статистического исследования эмпирических данных; техникой проверки статистических гипотез; навыками использования пакетов прикладных программ анализа данных
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач	
Знает	Методы анализа зависимости между переменными
Умеет	Исследовать зависимости и строить регрессионные модели; осуществлять выбор инструментальных средств статистического анализа данных
Владеет	Основными приемами регрессионного анализа; навыками использования пакетов прикладных программ статистической направленности

**Шкала и критерии оценивания формируемых компетенций
в результате изучения дисциплины в процессе освоения образовательной программы**

Шкала оценивания	Критерии оценки
На зачете	
«Зачтено» (61-100 баллов)	Выставляется студенту, если он определяет рассматриваемые понятия четко и полно, допускает отдельные погрешности в ответе или обнаруживает пробелы в знаниях основного учебно-программного материала; правильно решает практические задачи полностью или не в полном объеме. В результате следует считать, что компетенция сформирована. Присутствие сформированной компетенции свидетельствует о положительных результатах освоения дисциплины
«Не зачтено» (менее 61 балла)	Выставляется студенту, если он обнаруживает существенные пробелы в знаниях основных положений учебной дисциплины, неумение с помощью преподавателя получить правильное решение конкретной практической задачи из числа предусмотренных рабочей программой учебной дисциплины. В результате следует считать, что компетенция не сформирована. Отсутствие подтверждения наличия сформированности компетенции свидетельствует об отрицательных результатах освоения дисциплины

6.3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

6.3.1 Текущий контроль

Типовые контрольные задания
для оценки сформированности компетенций в процессе изучения
дисциплины, соотнесенные с этапами их формирования

Контролируемые модули / разделы / темы дисциплины	Форма оценочного средства	№ задания
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования		
Раздел 1. Основы математико-статистических методов анализа данных	Контрольная работа	Контрольная работа №1
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач		
Раздел 2. Корреляционный и регрессионный анализ.	Контрольная работа, индивидуальное домашнее задание	Контрольная работа №2, индивидуальная работа №1

Задания для контрольной работы

Контрольная работа № 1. На основе совокупности данных об урожайности ... за ... гг. Волгоградской области:

Задание 1. Определить параметры статистического распределения (построить вариационный интервальный ряд распределения; построить гистограмму частот; рассчитать числовые характеристики ряда).

Задание 2. Проверить при уровне значимости $\alpha=0,05$ статистическую гипотезу согласованности эмпирического распределения с нормальным теоретическим, применяя критерий Пирсона.

Вариант	Культура	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
1	Зерновые культуры	10,4	10	14,3	6,1	14	9,2	12,5	11,2	12,9	6,4	17,5	10,8	4,5	17,2	16
2	Пшеница	11,1	9,3	14,9	5,4	12,5	8,8	12,3	10,7	12,3	5,8	18,1	11,6	4,7	15,7	15,6
3	Пшеница озимая	14,5	13,7	19	5	12	11,3	17,5	13,1	17,9	9,6	25,4	16	12,4	23,8	20
4	Пшеница яровая	10,1	6,9	12,2	5,6	12,7	8	11,5	9,8	10,9	5,7	15,4	8,8	2,8	14	12,4
5	Рожь	9,3	12	11	2,9	10,6	8,1	8,6	8,6	11,1	7,1	16,7	10,9	7,9	17	14,5
6	Ячмень	11	11	16,3	7,9	16,9	11,1	13,9	13,5	14,9	6,3	18,7	10,2	4,1	18,6	17,8
7	Кукуруза на зерно	12,5	9,3	11,3	7,4	16,2	4,9	13,5	13	8,7	10,9	22,5	10,5	7,6	6,3	13,6
8	Овес	11,9	9,6	13,1	7,6	13,1	11,3	11,6	11,4	12,7	7,3	16,4	11	2,5	19,7	16,6
9	Просо	6,4	6,3	10,2	2	12,5	4	9,1	6,8	9,3	11,5	6,2	7,8	3,2	20,3	8,7
10	Гречиха	5,2	2,8	11	2,5	9,7	2,3	8	8,2	9,2	9,8	6,3	5,2	0,6	16,6	7,6
11	Сорго	1,2	2,9	4,6	2,1	3,8	0,6	3,7	1,2	0,7	1,1	1,4	1,2	0,5	4,6	1,9
12	Зернобобовые	6,9	4,3	15,7	7,8	10,7	7,8	9,4	11,3	12,1	5,7	15,7	7,9	4,4	14,1	15,9
13	Горох	2,7	3,4	20	8,2	10,7	8	9,4	11,4	12,2	5,8	15,8	7,9	4,5	14,2	15,9
Вариант	Культура	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
14	Зерновые культуры	3,8	19,2	10,2	19,7	6,8	11,9	8,2	10,2	13,1	4,1	14,1	10,6	13,2	18,5	18,4
15	Пшеница	5,2	18,9	10,6	22,1	7,5	12,8	11,4	10,7	18,8	7,3	17,2	15,5	12,7	25,7	24
16	Пшеница озимая	9,4	26,8	17	26,7	11,3	18,7	15	13,3	25,1	9,6	19,4	18,4	15	29,5	26,5
17	Пшеница яровая	1,7	15,2	7	16,6	6,1	8,5	4,8	5,6	8,9	1,2	12,8	6,8	9,8	7,6	9,3
18	Рожь	4	15,8	8,6	19,4	5,8	10,4	9,6	10	14,9	4,7	14,3	12,8	11,9	21,6	20
19	Ячмень	2,8	20,5	10,3	19	6,7	11,9	5,4	10,2	10,7	2,5	12,9	7,3	13,4	10,7	
20	Кукуруза на зерно	12,6	11,6	14	13,7	9,7	20,1	12	10	11,9	7,7	14,2	15,1	14,5	13,6	21,2
21	Овес	1,6	22,6	12,5	18,2	5,6	12,3	4,9	12,1	11,1	1,7	17,7	8,7	13,9	15,9	15,5
22	Просо	1,3	14,8	8,1	7	4,5	5,3	7,3	7,2	6,3	1,8	9,6	4,5	15	13,7	16
23	Гречиха	3,5	13,3	6,7	5,3	6,1	5,3	3,1	8,8	7,8	3,3	9,7	2,4	14,8	5,4	7,2
24	Сорго	0,6	4,7	1,8	1,2	5,2	1,5	3,9	3,1	4,8	1,6	2,6	2,9	7,6	7,6	10,9
25	Зернобобовые	1,7	15,8	10,4	15,1	4,3	12,6	4,8	8,9	11,5	1,6	10,9	4,7	10,6	8,8	10
26	Горох	1,7	15,8	10,4	15,1	4,3	12,6	4,8	8,9	11,5	1,5	10,9	4,7	10,7	8,9	10,1

Контрольная работа № 2. На основе приведенных исходных данных:

Задание 1. Оценить тесноту связи изучаемых признаков.

Задание 2. Построить уравнение регрессии.

Задание 3. Оценить статистическую значимость уравнения регрессии по F -критерию Фишера.Задание 4. Оценить статистическую значимость параметров регрессии с помощью t -критерия Стьюдента.

Задание 5. Дать точечный и интервальный прогноз расходов с вероятностью 0,95, принимая уровень дохода семьи равным 110% от среднего уровня.

№ варианта	Признак	Значения признака									
1	x	3,9	2,2	3,9	4,2	4,1	1,8	4,9	2,3	5,0	3,2
	y	22,8	15,4	20,5	28,2	25,0	12,6	27,8	14,8	29,1	20,4
2	x	3,1	2,9	3,0	5,0	2,3	4,0	2,4	5,0	5,1	4,0
	y	25,9	23,4	21,4	37,0	18,6	27,7	19,1	34,5	38,4	31,6
3	x	4,0	2,1	4,2	4,0	4,1	2,1	4,9	2,0	4,9	3,3
	y	28,8	18,4	26,5	34,0	31,2	15,4	35,5	18,0	33,0	25,0
4	x	2,9	4,8	3,2	3,2	2,2	5,2	4,0	3,1	3,6	4,8
	y	22,5	28,0	24,6	27,0	19,9	36,5	33,3	21,6	22,5	40,1
5	x	1,5	1,0	1,3	2,9	3,8	2,3	3,4	2,9	2,7	2,5
	y	23,6	17,7	20,1	24,6	33,9	28,8	40,5	29,7	32,6	28,9
6	x	3,9	0,7	4,1	4,1	1,0	3,8	1,2	3,9	1,3	0,8
	y	36,4	23,8	38,2	31,9	23,6	35,2	23,6	34,0	19,7	17,3
7	x	3,0	1,2	2,1	1,1	3,0	2,9	0,8	1,5	3,0	3,2
	y	38,5	15,8	29,6	18,5	37,6	29,1	18,8	20,6	33,4	36,8
8	x	2,9	4,0	3,0	1,2	1,0	4,3	1,0	1,7	1,7	3,1
	y	31,0	42,4	31,1	24,5	15,3	36,0	18,7	18,0	24,6	30,0
9	x	1,2	1,4	1,3	2,7	4,2	2,4	3,3	3,0	2,6	2,5
	y	23,2	18,0	19,5	24,4	34,2	29,1	40,3	31,1	32,4	28,5
10	x	2,5	2,9	2,4	2,0	4,1	5,0	4,1	3,4	3,9	5,0
	y	17,6	19,5	18,1	15,0	27,2	33,0	24,8	24,0	28,7	30,5
№	Признак	Значения признака									
11	y	90	120	180	220	260	290	330	380		
	x	120	310	530	740	960	1180	1450	1870		
12	y	95	125	185	225	260	270	325	350		
	x	125	300	510	720	950	1150	1450	1861		
13	y	90	120	180	220	260	290	330	380		
	x	110	300	515	730	940	1100	1450	1850		
14	y	85	110	155	210	245	285	325	360		
	x	120	310	530	740	960	1180	1450	1870		
15	y	80	100	130	165	200	255	300	345		
	x	150	280	330	500	880	1050	1350	1800		
16	y	100	150	220	300	330	350	380	400		
	x	200	280	350	600	750	1200	1400	1900		
17	y	85	110	155	210	245	285	325	360		
	x	150	280	330	500	880	1050	1350	1800		
18	y	87	95	115	135	150	200	250	335		
	x	180	200	250	310	650	980	1450	1750		
19	y	90	120	180	220	260	290	330	380		
	x	200	280	350	600	750	1200	1400	1900		
20	y	87	95	115	135	150	200	250	335		
	x	120	310	530	740	960	1180	1450	1870		

Задания для индивидуальной домашней работы

Индивидуальная работа №1. По имеющимся данным провести дисперсионный анализ.

№ 1			№ 2			№ 3		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
7	3,9	10	7	3,9	10	6	3,6	10
7	3,9	14	7	3,7	14	6	3,9	12
7	3,7	15	7	3,7	15	7	3,7	15
7	4,0	16	7	4,0	16	7	4,0	15
7	3,8	17	8	4,0	16	7	3,5	17
7	4,8	19	8	4,8	18	7	4,4	18
8	5,4	19	8	5,1	19	8	5,4	18
8	4,4	20	8	4,4	19	8	4,4	19
8	5,3	20	8	5,3	20	9	5,3	20
10	6,8	20	10	5,5	20	10	6,0	20
№ 4			№ 5			№ 6		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
6	3,5	12	6	3,6	12	7	3,9	10
7	3,7	14	6	3,9	14	7	3,7	12
7	3,7	15	7	3,7	15	7	3,7	15
7	4,0	16	7	4,0	16	7	4,0	15
8	4,2	16	7	3,5	16	8	4,0	17
8	4,8	17	8	4,4	17	9	4,8	18
8	5,0	18	8	5,4	18	9	5,1	18
9	4,4	19	9	4,4	19	9	4,4	19
9	5,8	20	9	5,3	20	9	5,3	20
10	5,5	20	10	6,0	20	10	5,5	20
№ 7			№ 8			№ 9		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
6	3,2	12	6	3,5	10	6	3,2	11
6	3,7	13	7	3,5	12	6	3,5	13
7	3,7	15	7	3,7	13	6	3,7	13
7	4,0	15	7	4,0	15	7	4,0	15
7	3,8	16	8	4,4	16	7	3,8	16
7	4,4	17	8	4,8	17	7	4,1	16
8	5,0	18	9	5,1	18	8	5,0	18
9	4,6	19	9	4,7	19	8	4,6	19
9	5,4	20	9	5,0	20	9	5,1	20
10	6,4	20	10	5,5	20	10	6,4	20
№ 10			№ 11			№ 12		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
6	3,5	11	6	3,9	10	7	3,5	10
7	3,5	11	7	3,9	12	7	3,7	14
7	3,7	14	7	3,7	15	7	3,7	15
8	4,0	15	7	4,0	15	7	4,0	16
8	3,8	16	8	3,8	17	7	4,2	16
8	4,8	17	8	4,8	18	7	4,8	18
9	5,1	18	8	5,4	18	8	5,0	19
9	4,7	19	9	4,4	19	8	4,4	19
9	5,4	19	9	5,3	20	8	5,8	20
10	5,5	20	10	6,8	20	10	5,5	20

№ 13			№ 14			№ 15		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
6	3,9	10	6	3,5	10	6	3,6	11
6	3,9	12	7	3,7	14	6	3,9	13
7	3,7	15	7	3,7	15	7	3,7	13
7	4,0	15	7	4,0	16	7	4,0	15
7	3,8	17	8	4,2	16	7	3,5	16
7	4,8	18	8	4,8	18	8	4,4	16
8	5,4	18	8	5,0	19	8	5,4	18
8	4,4	19	9	4,4	19	9	4,4	19
9	5,3	20	9	5,8	20	9	5,3	20
10	6,8	20	10	5,5	20	10	6,0	20
№ 16			№ 17			№ 18		
y	x_1	x_2	y	x_1	x_2	y	x_1	x_2
7	3,9	12	6	3,2	11	6	3,5	11
7	3,7	13	6	3,7	13	7	3,5	13
7	3,7	15	7	3,7	13	7	3,7	13
7	4,0	15	7	4,0	15	7	4,0	15
8	4,0	16	7	3,8	16	8	4,4	16
9	4,8	17	7	4,4	16	8	4,8	16
9	5,1	18	8	5,0	18	9	5,1	18
9	4,4	19	9	4,6	19	9	4,7	19
9	5,3	20	9	5,4	20	9	5,0	20
10	5,5	20	10	6,4	20	10	5,5	20
№ 19			№ 20					
y	x_1	x_2	y	x_1	x_2			
6	3,2	11	6	3,6	11			
7	3,5	13	7	3,9	11			
7	3,7	13	7	3,7	14			
7	4,0	15	8	4,0	15			
8	3,8	16	8	3,5	16			
8	4,1	16	8	4,4	17			
9	5,0	18	9	5,4	18			
9	4,6	19	9	4,4	19			
9	5,1	20	9	5,3	19			
10	6,4	20	10	6,0	20			

Типовые контрольные задания
для оценки сформированности компетенций в результате изучения
дисциплины в процессе освоения образовательной программы,
соотнесенные с этапами их формирования

Контролируемые модули / разделы / темы дисциплины	№ вопроса / задания для проверки уровня обученности		
	Знать	Уметь	Владеть
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования			
Раздел 1. Основы математико-статистических методов анализа данных	Вопросы 1-12	Задание 1-10	Задание 1-10
ПК-23 – способностью применять системный подход и математические методы в формализации решения прикладных задач			
Раздел 2. Корреляционный и регрессионный анализ.	Вопросы 13-24	Задание 11-20	Задание 11-20

Вопросы для проверки уровня обученности ЗНАТЬ (ответьте на теоретические вопросы)

1. Предмет и задачи математико-статистических методов обработки информации.
2. Генеральная и выборочная совокупности.
3. Вариационные и статистические ряды распределения, их графическое изображение.
4. Построение интервального статистического ряда распределения.
5. Законы распределения выборочных характеристик.
6. Эмпирическая функция распределения.
7. Статистические характеристики положения (средняя, мода, медиана).
8. Показатели вариации (рассеяния).
9. Моменты, коэффициенты асимметрии и эксцесса.
10. Точность оценки. Доверительная вероятность. Доверительный интервал.
11. Статистические гипотезы.
12. Проверка гипотезы о нормально распределении. Критерий согласия Пирсона.
13. Спецификация эконометрических моделей.
14. Дисперсионный анализ. Общая, факторная и остаточная дисперсии.
15. Корреляционная и функциональная зависимости.
16. Уравнение регрессии. Коэффициент регрессии.
17. Метод наименьших квадратов.
18. Коэффициент корреляции и коэффициент детерминации.
19. Оценка значимости уравнения регрессии с помощью F-критерия Фишера.
20. Оценка значимости параметров уравнения регрессии по t-критерию Стьюдента.
21. Средняя ошибка аппроксимации.
22. Предпосылки метода наименьших квадратов (МНК). Проверка предпосылок МНК.
23. Информационные технологии статистических исследований.
24. Интервальный прогноз на основе уравнения регрессии.

Вопросы / Задания для проверки уровня обученности УМЕТЬ (выполните предложенное задание)

Задание 1. Ниже приведены выборочные данные о производительности труда на 76 предприятиях мясной промышленности за месяц. За производительность труда на предприятии принято частное от деления валовой продукции за год, выраженной в рублях, на число всех видов работников предприятия. Приведенные данные округлены, и каждое число выражает в среднем долю выработанной валовой продукции в тысячах рублей на одного работника предприятия в месяц:

24, 32, 36, 34, 26, 35, 29, 28, 40, 37, 18, 27, 22,
26, 34, 28, 32, 33, 39, 41, 35, 32, 35, 41, 30, 42,
31, 23, 25, 39, 27, 25, 32, 18, 36, 33, 31, 22, 41.

Построить интервальный статистический ряд, изобразив его геометрически в виде гистограммы частот.

Задание 2. В приведенной ниже таблице представлены данные о валовой продукции за год по 76 предприятиям мясной промышленности. Все данные округлены и выражены в миллионах рублей (под валовой продукцией понимается вся выработанная за год продукция)

Валовая продукция, млн руб.	2-8	8-14	14-20	20-26	26-32	32-38	38-44	44-50	50-56
Число предприятий	27	17	8	6	5	3	3	3	4

Построить гистограмму частот и кумуляту.

Задание 3. Для дискретного статистического ряда, заданного таблицей, построить полигоны относительных частот, а также эмпирическую функцию распределения.

x_i	13	14	15	16	17	19	21
n_i	3	6	3	3	2	1	1

Задание 4. Для интервального статистического ряда, заданного таблицей, построить гистограммы частот и относительных частот, а также эмпирическую функцию распределения.

Интервалы	7,5 - 10,5	10,5 - 13,5	13,5 - 16,5	13,5 - 16,5	19,5 - 22,5
n_i	1	5	14	3	1

Задание 5. В течение недели регистрировались пропуски занятий студентами одной группы. В результате регистрации получили следующие статистические данные:

2, 1, 3, 1, 2, 1, 2, 4, 3, 5, 3, 2, 2, 2, 1, 2, 3, 1, 0, 0, 0, 2, 3, 1, 4.

Составить статистический дискретный ряд распределение числа пропущенных занятий и построить их полигон частот. Найти эмпирическую функцию распределения.

Задание 6. Получены следующие данные о количестве производственных подразделений на каждом из 60 сельскохозяйственных предприятий:

2, 4, 5, 3, 4, 6, 7, 4, 5, 3, 3, 4, 2, 6, 5, 4, 7, 2, 3, 4,
4, 5, 4, 3, 4, 6, 6, 5, 2, 3, 4, 3, 5, 6, 7, 2, 4, 3, 4, 5,
4, 6, 7, 2, 5, 3, 5, 4, 3, 7, 2, 4, 3, 4, 5, 4, 3, 2, 6, 7.

Составить дискретный статистический ряд распределения сельскохозяйственных предприятий по числу производственных подразделений на одно хозяйство. Найти накопленные частоты и относительные частоты.

Задание 7. Получены данные о числе цветных телевизоров, продаваемых ежедневно в магазине электроники в течение 26 дней:

16; 12; 15; 15; 23; 9; 15; 13; 14; 14; 21; 15; 14;
17; 27; 15; 16; 12; 16; 19; 14; 16; 17; 13; 14; 14.

Требуется: составить вариационный ряд; составить дискретный статистический ряд распределения частот и относительных частот; составить интервальный статистический ряд частот.

Задание 8. Выборочная совокупность задана таблицей распределения.

x_i	1	2	5	8	9
n_i	3	4	6	4	3

Найти эмпирическую функцию распределения и построить ее график.

Задание 9. Найти и изобразить графически эмпирическую функцию распределения, соответствующую выборке, представленной в следующей таблице.

x_i	1	2	4	8	10
n_i	5	15	12	13	5

Вычислить выборочное среднее.

Задание 10. Имеются данные о числе цветных телевизоров, продаваемых ежедневно в магазине электроники в течение некоторого месяца:

5; 16; 18; 19; 14; 12; 22; 23; 25; 20; 32; 17; 34; 25; 14; 14; 17; 8; 5; 11; 13; 6; 7; 9; 14; 7; 21; 28; 23.

Построить интервальный статистический ряд с корректировкой границ первого и последнего интервалов. Построить полигон.

Задание 11. Построить уравнение парной линейной регрессии и пояснить смысл его параметров:

$$n = 10, \sum_{i=1}^n x_i = 35,3; \sum_{i=1}^n y_i = 238,4; \sum_{i=1}^n x_i^2 = 134,81; \sum_{i=1}^n (y_i \cdot x_i) = 898,32.$$

Задание 12. Построить уравнение показательной регрессии $\hat{y} = b_0 \cdot b_1^x$, если $n = 10$,

$$\sum_{i=1}^n x_i = 36,8; \sum_{i=1}^n z_i = 14,3; \sum_{i=1}^n x_i^2 = 146,1; \sum_{i=1}^n (z_i \cdot x_i) = 53,7, \text{ где } z = \lg y.$$

Задание 13. Построить уравнение степенной регрессии $\hat{y} = b_0 \cdot x^{b_1}$, если $n = 10$,

$$\sum_{i=1}^n v_i = 5,543; \sum_{i=1}^n z_i = 14,295; \sum_{i=1}^n v_i^2 = 3,196; \sum_{i=1}^n (z_i \cdot v_i) = 8,017,$$

где $z = \lg y$, $v = \lg x$.

Задание 14. С вероятностью 0,99 определить точечный и интервальный прогноз результативного признака y при значении фактора x , составляющем 120 % от среднего уровня:

$$\hat{y} = 71,24 + 0,138 \cdot x; \quad n = 8; \quad \bar{y} = 170,88; \quad \overline{x^2} = 847363.$$

Задание 15. Определить числовые характеристики (среднее значение, среднее квадратическое отклонение, дисперсию) результативного признака y и факторных признаков x_1 и x_2 . Рассчитать множественный коэффициент корреляции. Сделать вывод.

y	6	6	7	7	7	7	8	8	9	10
x_1	3,6	3,9	3,7	4	3,5	4,4	5,4	4,4	5,3	6
x_2	10	12	15	15	17	18	18	19	20	20

Задание 16. Определить коэффициент парной корреляции и оценить тесноту связи между признаками y и x : $\hat{y} = 58,62 + 0,033 \cdot x$; $\bar{x} = 64,9$; $\overline{y^2} = 3713,1$; $\overline{x^2} = 4246,3$.

Задание 17. Оценить качество уравнения регрессии по F -критерию Фишера:

$$\hat{y} = 17,2 + 4,51 \cdot x; \quad n = 10; \quad \sigma_x = 1,49; \quad \sigma_y = 7,16; \quad \alpha = 0,1.$$

Задание 18. Оценить статистическую значимость параметра регрессии b_1 через t -критерий Стьюдента: $\hat{y} = 3,606 + 6,564 \cdot x$; $\overline{x^2} = 14,61$; $\bar{x} = 3,68$; $t_{кр} = 2,306$.

y	25,9	23,4	21,4	37,0	18,6	27,7	19,1	34,5	38,4	31,6
$\hat{y}$	24,0	22,6	23,3	36,4	18,7	29,9	19,4	36,4	37,1	29,9

Задание 19. С вероятностью 0,9 оценить статистическую значимость параметра регрессии b_0 через t -критерий Стьюдента:

$$\hat{y} = 90,88 - 0,403 \cdot x; \quad S_{ост}^2 = 35,44; \quad n = 7; \quad \sum x^2 = 26166,4; \quad \bar{x} = 60,61.$$

Задание 20. Оценить с помощью средней ошибки аппроксимации качество уравнения регрессии: $\hat{y} = 82,36 + 0,17 \cdot x$.

x	120	310	530	740	960	1180	1450	1870
y	90	120	180	220	260	290	330	380

Задания для проверки уровня обученности ВЛАДЕТЬ (решите практическую (ситуационную) задачу)

Задание 1. В штате предприятия числится 60 рабочих, которые имеют следующие разряды:

1, 5, 2, 4, 3, 4, 6, 4, 5, 1, 2, 2, 3, 4, 5, 3, 4, 5, 2, 1,
4, 5, 5, 4, 3, 4, 6, 1, 2, 4, 4, 3, 5, 6, 4, 3, 3, 1, 3, 4,
3, 1, 2, 4, 4, 5, 6, 1, 3, 4, 5, 3, 4, 4, 3, 2, 6, 1, 2, 4.

Составить дискретный статистический ряд распределения рабочих по разрядам. Определить моду и медиану.

Задание 2. Путем устного опроса изучалось качество продукции, выпускаемой фирмой и реализуемой в магазине этой фирмы. Посетители давали оценку качества по десятибалльной шкале. В результате были получены сводные данные, представленные в таблице.

Оценка качества продукции, балл	1-2	3-4	5-6	7-8	9-10
Число случаев	3	8	36	89	45

Определить средний балл качества продукции, среднее квадратическое отклонение, выборочный момент третьего порядка.

Задание 3. Администрацию универсама интересует оптимальный уровень запасов продуктов в торговом зале, а также среднемесячный объем покупок товаров, которые не являются предметом ежедневного потребления в семье (например, сода). Для выяснения этого вопроса менеджер универсама в течение месяца регистрировал частоту покупок 100-граммовых пакетов с содой и собрал следующие данные:

4, 4, 9, 3, 3, 1, 2, 0, 4, 2, 3, 5, 7, 10, 6,
5, 7, 3, 2, 9, 8, 1, 4, 6, 5, 4, 2, 1, 0, 8.

Построить вариационный и дискретный статистический ряды, определить выборочное среднее, среднее квадратическое отклонение, коэффициент вариации.

Задание 4. В табл. представлены данные об урожайности ржи на различных участках поля.

Урожайность ржи, ц/га	[9-12)	[12-15)	[15-18)	[18-21)	[21-24)	[24-27]
Доля участка в общей площади, %	6	12	33	22	19	8

Найти выборочное среднее, выборочную дисперсию, коэффициент вариации и размах вариации признака X - урожайности ржи.

Задание 5. По статистическому распределению, заданному таблицей,

x_i	1	3	5	7	9
n_i	10	15	30	33	12

вычислить числовые характеристики: выборочное среднее, среднее квадратическое отклонение, моду, медиану и коэффициент вариации.

Задание 6. Выборочным путем были получены следующие данные об урожайности ржи в районе

Урожайность, ц/га	18	20	21
Число гектаров	10	20	20

Определить выборочное среднее, коэффициенты асимметрии и эксцесса.

Задание 7. В результате первичной обработки статистических данных о затратах времени (в минутах) на разработку и оформление планового приказа директора фирмы получили интервальную таблицу распределения.

Интервалы, мин	18-22	22-26	26-30	30-34	34-38	38-42
Число наблюдений на интервале	7	13	33	31	12	3

Найти выборочное среднее, моду и медиану.

Задание 8. В одном из банков в течение дня измеряли время обслуживания клиентов. Получено следующее распределение.

Время обслуживания, мин	6-8	8-10	10-12	12-14	14-16	16-18	18-20
Число клиентов	1	3	16	15	14	6	2

Найти среднее арифметическое, среднее квадратическое отклонение, коэффициент вариации.

Задание 9. Следующие данные представляют собой годовой процент продаж в общем объеме продаж автомобилей импортного производства в России с 2002 по 2014 г. (данные условные):

6,7; 9,5; 9,3; 12,3; 12,0; 16,6; 21,3; 21,8; 22,6; 20,9; 18,3; 20,1; 22,8.

Найти выборочное среднее, медиану и моду, стандартное отклонение построенного статистического ряда.

Задание 10. На некотором предприятии собраны данные о числе дней, пропущенных работниками по болезни.

Число дней, пропущенных в текущем месяце	0	1	2	3	4	5
Число работников	10	17	25	28	30	27

Найдите среднее число пропущенных дней, стандартное отклонение, медиану и моду распределения. Является ли распределение симметричным?

Задание 11. На основе приведенных исходных данных требуется:

1. Построить уравнение парной линейной регрессии и пояснить экономический смысл его параметров.

2. Найти прогноз y , принимая уровень фактора x равным 120 % от среднего уровня.

x	1,5	1,0	1,3	2,9	3,8	2,3	3,4	2,9	2,7	2,5
y	23,6	17,7	20,1	24,6	33,9	28,8	40,5	29,7	32,6	28,9

Задание 12. На основе приведенных исходных данных требуется:

1. Построить уравнение парной линейной регрессии.

2. Определить линейный коэффициент корреляции и коэффициент детерминации.

x	2,9	4,8	3,2	3,2	2,2	5,2	4,0	3,1	3,6	4,8
y	22,5	28,0	24,6	27,0	19,9	36,5	33,3	21,6	22,5	40,1

Задание 13. На основе данных по предприятиям требуется:

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.

2. Определить множественный коэффициент корреляции. Сделать вывод.

3. Определить прогнозное значение результативного признака $\hat{y}_p$, если фактор x_1 увеличится на 15% от среднего уровня, а фактор x_2 - на 10% от среднего уровня.

y	6	6	7	7	7	8	8	9	9	10
x_1	3,6	3,9	3,7	4	3,5	4,4	5,4	4,4	5,3	6
x_2	12	14	15	16	16	17	18	19	20	20

Задание 14. На основе данных по предприятиям требуется:

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.
2. Определить парные и частные коэффициенты корреляции. Сделать выводы.

y	7	7	7	7	7	7	8	8	8	10
x_1	3,9	3,9	3,7	4	3,8	4,8	5,4	4,4	5,3	6,8
x_2	10	14	15	16	17	19	19	20	20	20

Задание 15. На основе приведенных исходных данных требуется:

1. Построить уравнение парной линейной регрессии.
2. Оценить статистическую значимость уравнения регрессии по F-критерию Фишера.

x	3,1	2,9	3,0	5,0	2,3	4,0	2,4	5,0	5,1	4,0
y	25,9	23,4	21,4	37,0	18,6	27,7	19,1	34,5	38,4	31,6

Задание 16. На основе приведенных исходных данных требуется:

1. Построить уравнение парной линейной регрессии.
2. Оценить статистическую значимость параметров уравнения регрессии по t-критерию Стьюдента.

x	4,0	2,1	4,2	4,0	4,1	2,1	4,9	2,0	4,9	3,3
y	28,8	18,4	26,5	34,0	31,2	15,4	35,5	18,0	33,0	25,0

Задание 17. На основе приведенных исходных данных требуется:

1. Построить уравнение парной линейной регрессии.
2. Оценить качество уравнения регрессии через среднюю ошибку аппроксимации.

x	3,9	2,2	3,9	4,2	4,1	1,8	4,9	2,3	5,0	3,2
y	22,8	15,4	20,5	28,2	25,0	12,6	27,8	14,8	29,1	20,4

Задание 18. На основе данных по предприятиям требуется:

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.
2. Оценить статистическую значимость уравнения регрессии с помощью F-критерия Фишера.

y	7	7	7	7	8	8	8	8	8	10
x_1	3,9	3,7	3,7	4	4	4,8	5,1	4,4	5,3	5,5
x_2	10	14	15	16	16	18	19	19	20	20

Задание 19. На основе данных по предприятиям требуется:

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.
2. Оценить качество уравнения регрессии через среднюю ошибку аппроксимации

y	6	6	7	7	7	7	8	8	9	10
x_1	3,6	3,9	3,7	4	3,5	4,4	5,4	4,4	5,3	6
x_2	10	12	15	15	17	18	18	19	20	20

Задание 20. На основе данных по предприятиям требуется:

1. Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.
2. Оценить с помощью t-критерия Стьюдента статистическую значимость коэффициентов регрессии.

y	6	7	7	7	8	8	8	9	9	10
x_1	3,5	3,7	3,7	4	4,2	4,8	5	4,4	5,8	5,5
x_2	12	14	15	16	16	17	18	19	20	20

6.4 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Методические материалы, определяющие процедуры оценивания сформированности компетенций, соотнесенные с этапами их формирования

Контролируемые модули / разделы / темы дисциплины	Форма оценочного средства	Методические материалы
ОПК-2 - способностью анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования		
Раздел 1. Основы математико-статистических методов анализа данных	Контрольная работа	Методические указания по подготовке к контрольной работе
ПК-23 - способностью применять системный подход и математические методы в формализации решения прикладных задач		
Раздел 2. Корреляционный и регрессионный анализ.	Контрольная работа	Методические указания по подготовке к контрольной работе
	Индивидуальное домашнее задание	Методические указания по выполнению домашней работы

Методические указания по подготовке к контрольной работе

Контрольная работа - Средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу.

Контрольная работа №1 (пример). На основе совокупности данных об урожайности зерновых культур за 1950-1964 гг. Волгоградской области (таблица 1).

Требуется:

1. Определить параметры статистического распределения (построить вариационный интервальный ряд распределения; построить гистограмму частот; рассчитать числовые характеристики ряда).

2. Проверить при уровне значимости $\alpha=0,05$ статистическую гипотезу согласованности эмпирического распределения с нормальным теоретическим, применяя критерий Пирсона.

Таблица 1 - Урожайность зерновых культур за 1950-1964 гг., ц/га

Год	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
Урожайность, ц/га	4,6	4,8	7,5	4,8	3,6	8,4	4,8	3,9	13,7	5,5	10,4	10	14,3	6,1	14

Решение:

1. Построим интервальный ряд распределения урожайностей по наименьшему и наибольшему значению исследуемого признака.

По таблице экспериментальных данных (таблица 1) находим наименьшее $x_{\min}=3,6$ и наибольшее $x_{\max}=14,3$ значения. Определяем шаг интервала: $i=(x_{\max}-x_{\min})/(1+3,32*\lg(n))$;
 $i=(14,3-3,6)/(1+3,32*\lg(15))\approx 2,2$.

Таблица 2 - Интервальный ряд распределения урожайности зерновых культур

$x_{i-1}-x_i$	3,5–5,7	5,7–7,9	7,9–10,1	10,1–12,3	12,3–14,5
n_i	7	2	2	1	3

Графическое изображение интервального распределения называется гистограммой – ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат интервалы длины i и высотами служат частоты n_i . Площадь гистограммы частот равна сумме всех частот, т.е. объему выборки n .

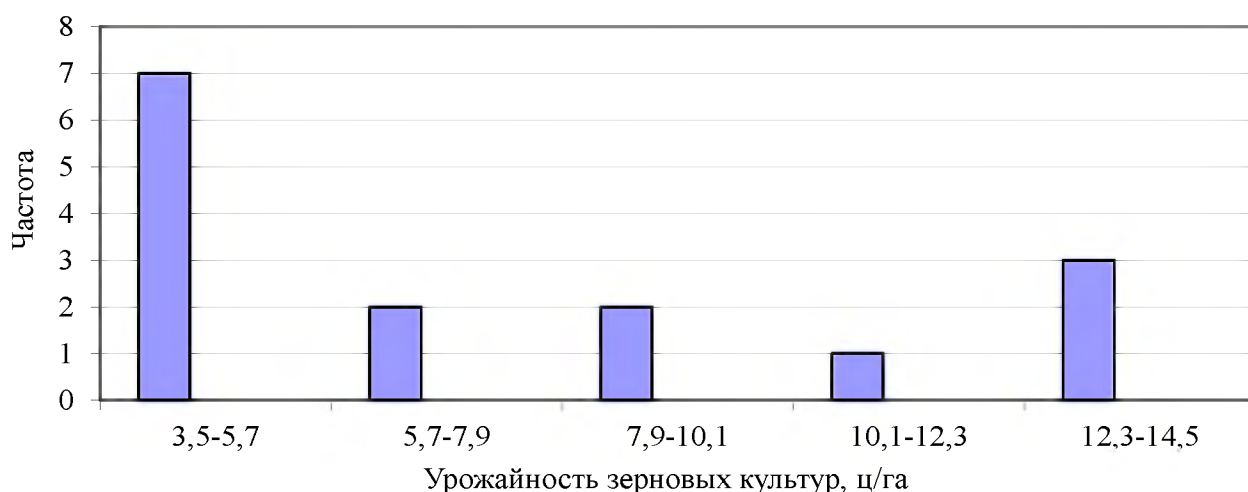


Рис. 1 – Гистограмма распределения урожайности

Чтобы охарактеризовать наиболее существенные свойства распределения, определим средние показатели, которые называются выборочными числовыми характеристиками.

Выборочную среднюю рассчитаем по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i n_i}{n}$$

При расчете средней арифметической для интервального ряда нужно сначала определить середины интервалов как полусуммы значений верхней и нижней границ интервала (табл. 3).

Таблица 3 - Дискретный ряд распределения

x_i	4,6	6,8	9,0	11,2	13,4
n_i	7	2	2	1	3

$$\bar{x} = \frac{4,6 \cdot 7 + 6,8 \cdot 2 + 9 \cdot 2 + 11,2 \cdot 1 + 13,4 \cdot 3}{15} = \frac{115,2}{15} \approx 7,7 \quad \text{ц/га}$$

Средняя урожайность зерновых культур составляет 7,7 ц/га.

Мода – варианта, которая наиболее часто встречается в данной совокупности, т.е. варианта с наибольшей частотой ($n_i = \max$).

Для интервального ряда (с равными интервалами) мода определяется по следующей формуле:

$$M_o = x_{M_o} + i \frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)},$$

где i – длина интервала; n_2 – частота модального, n_1 – домодального, n_3 – замодального интервалов; x_{M_o} – начало модального интервала.

$$M_o = 3,5 + 2,2 \frac{7 - 0}{(7 - 0) + (7 - 2)} \approx 4,8 \quad \text{ц/га}$$

Наиболее часто в наблюдаемой совокупности встречается урожайность 4,8 ц/га.

Медиана – варианта, относительно которой статистическая совокупность делится на две равные по объему части, причем в одной из них содержатся члены, у которых значения признака не больше M_e , а в другой – члены со значениями не меньше M_e .

Для интервального распределения сначала находят медианный интервал. При нечетном числе членов ряда ($n = 2k + 1$) медианным будет являться интервал, в котором накопленная частота содержит значение $k + 1$, а при четном числе членов ряда ($n = 2k$) – значение k .

Накопленной частотой в точке x называют суммарную частоту членов статистической совокупности со значениями признака меньшими x (табл. 4).

Таблица 4

$x_{i-1}-x_i$	3,5–5,7	5,7–7,9	7,9–10,1	10,1–12,3	12,3–14,5
Накопленная частота	7	9	11	12	15

Медиану вычисляют по формуле:

$$M_e = x_{M_e} + i \frac{n/2 - n_n}{n_{M_e}},$$

где n_n – накопленная частота интервала, предшествующего медианному; n_{M_e} – локальная частота медианного интервала; x_{M_e} – начало медианного интервала.

$$n = 15 = 2 \cdot 7 + 1, \text{ следовательно } k+1 = 7+1 = 8.$$

Медианным является интервал (5,7 – 7,9), так как его накопленная частота содержит $k+1=8$.

$$M_e = 5,7 + 2,2 \cdot \frac{15/2 - 7}{2}.$$

50% значений признака составляют не более 6,3 ц/га, а остальные 50% составляют 6,3 ц/га и более.

Размах вариации – разность между наибольшей и наименьшей вариантой:

$$R = x_{\max} - x_{\min} = 14,3 - 3,6 = 10,7 \text{ ц/га}.$$

Разность между наибольшей и наименьшей вариантами равна 10,7 ц/га.

Среднее квадратическое отклонение – показатель вариации, измеряющий величину, на которую все варианты в среднем отклоняются от средней арифметической.

Применяется для оценки диапазона значений случайной величины. При этом используется так называемое правило трех сигм, состоящим в том, что диапазон практически возможных значений случайной величины X не выходит за пределы $\bar{x} \pm 3\sigma_x$.

Среднее квадратическое отклонение рассчитывается по формуле

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot n_i}{\sum n_i}}.$$

$$\sigma = \sqrt{\frac{(4,6 - 7,7)^2 \cdot 7 + (6,8 - 7,7)^2 \cdot 2 + \dots + (13,4 - 7,7)^2 \cdot 3}{7 + 2 + \dots + 3}} = \sqrt{\frac{182}{15}} \approx 3,48.$$

В среднем величина урожайности в исследуемой совокупности отклоняется от средней на 3,48 ц/га.

Дисперсия – средний квадрат отклонений значений признака от средней.

$$D = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{\sum n_i} = \sigma^2 = 3,48^2 \approx 12,1.$$

Средний квадрат отклонений урожайности от среднего значения равен 12,1 (D).

Коэффициент вариации показывает меру колебимости признака относительно его среднего значения и позволяет сравнивать изменчивость признаков, выраженных в разных единицах измерения.

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Для нормального распределения изменчивость считается значительной, если $V > 0,2$; средней, если $0,1 < V < 0,2$, и незначительной, если $V < 0,1$.

$$V = \frac{3,48}{7,7} \cdot 100\% = 45\%.$$

Так как вариация урожайности зерновых культур составляет 45%, изменчивость признака значительная ($V > 20\%$), а рассматриваемая совокупность неоднородна ($V > 33\%$).

2. Случайная величина X имеет нормальное распределение, если ее плотность вероятности составляет

$$f(x) = \frac{1}{\sigma_x} \cdot e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}$$

где m_x – математическое ожидание случайной величины X ;

σ_x – среднее квадратическое отклонение случайной величины X .

Нормальное распределение полностью определяется величинами $\bar{x}$ и σ_x .

Вероятность попадания случайной величины X , имеющей нормальное распределение с параметрами $\bar{x}$ и σ_x , на участке от α до β выражается формулой:

$$P = \{X \in (\alpha, \beta)\} = \Phi\left(\frac{\beta - m_x}{\sigma_x}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma_x}\right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$.

Для проверки гипотезы о нормальном распределении сформулируем нулевую гипотезу (H_0) как утверждение, что случайная величина X – урожайность зерновых культур в генеральной совокупности – имеет нормальное распределение с параметрами $X = 7,7$ и $\sigma = 3,48$.

Необходимо рассчитать вероятности p_i попадания случайной величины X в i -й интервал (x_{i-1}, x_i) , а затем определить теоретические частоты, как пропорциональные указанным вероятностям:

$$n_i^* = p_i n \quad (n=15).$$

Если случайная величина X имеет нормальное распределение, то нормированная случайная величина $Z = \frac{X - \bar{X}}{\sigma}$ имеет стандартное нормальное распределение, и для расчета вероятностей можно использовать табличную функцию Лапласа:

$$p_i = P(x_{i-1} \leq X < x_i) = P\left(\frac{x_{i-1} - \bar{X}}{\sigma} \leq \frac{X - \bar{X}}{\sigma} < \frac{x_i - \bar{X}}{\sigma}\right) = P(z_{i-1} \leq Z < z_i) = \Phi(z_i) - \Phi(z_{i-1}) = \Delta\Phi.$$

Значения функции Лапласа $\Phi(z_i)$ приводятся в таблице интегральной функции Лапласа.

Для нахождения значений функции Лапласа для отрицательных значений z воспользуемся свойством нечетности этой функции, т.е.

$$\Phi(-z) = -\Phi(z).$$

При расчете вероятностей p_i и, соответственно, теоретических частот n_i^* мы принимаем нижнюю границу первого интервала равной $-\infty$, а верхнюю границу последнего интервала $+\infty$. Значения функции Лапласа принимаются:

$$\Phi(-\infty) = -0,5; \quad \Phi(+\infty) = 0,5.$$

Если нулевая гипотеза справедлива, то эмпирические частоты n_i должны быть близкими к теоретическим частотам n_i^* . Для оценки расхождения эмпирических и теоретических частот используем критерий χ^2 Пирсона:

$$\chi^2 = \sum_{i=1}^q \frac{(n_i - n_i^*)^2}{n_i^*}.$$

Если справедлива нулевая гипотеза, то статистика χ^2 имеет так называемое χ^2 -распределение с числом степеней свободы

$$v = q - 1 - k,$$

где q – число интервалов, на которые разбит весь диапазон изменения варьирующего признака ($q=5$);

k – число параметров генерального распределения, оцененных по данным выборки ($k=2$).

– по данным выборки оценены два параметра: средняя $\bar{X}$ и среднее квадратическое отклонение σ).

Таблица 5

Расчеты для проверки гипотезы нормального распределения урожайности зерновых культур

Урожайность, зерновых культур, ц/га $x_{i-1} - x_i$	Эмпирические частоты n_i	Нормированные верхние границы интервалов $z_i = \frac{x_i - \bar{X}}{\sigma}$	Значения функции Лапласа $\Phi(z_i)$	Вероятности $p_i = \Delta\Phi$	Теоретические частоты $n_i^* = p_i n$	$\frac{(n_i - n_i^*)^2}{n_i}$
1	2	3	4	5	6	7
$-\infty$	-	$-\infty$	-0,5	-	-	-
3,5–5,7	7	-0,57	-0,2157	0,2843	4,2	1,8667
5,7–7,9	2	0,06	0,0239	0,2396	3,6	0,7111
7,9–10,1	2	0,69	0,2549	0,231	3,5	0,6429
10,1–12,3	1	1,32	0,4066	0,1517	2,3	0,7348
12,3–14,5	3	$+\infty$	0,5	0,0934	1,4	1,8286
Итого	15	-	-	1,0000	15,0	5,7831

По данным выборки рассчитываем конкретное значение статистики $\chi^2_{\text{набл}}$. При полном совпадении эмпирических и теоретических частот $\chi^2_{\text{набл}} = 0$, в противном случае $\chi^2_{\text{набл}} > 0$, однако при справедливости нулевой гипотезы большие расхождения маловероятны.

$\chi^2_{\text{набл}}$ рассчитывается в 7-й колонке таблицы 5 ($\chi^2_{\text{набл}} = 5,7831$).

По заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $v = 5 - 1 - 2 = 2$ находим табличное (критическое) значение $\chi^2_{\alpha} = 5,99$.

Если $\chi^2_{\text{набл}} \leq \chi^2_{\alpha}$, то гипотеза H_0 не отклоняется, если же $\chi^2_{\text{набл}} > \chi^2_{\alpha}$, то гипотеза H_0 отклоняется.

Так как $\chi^2_{\text{набл}} = 5,7831 < \chi^2_{\alpha} = 5,99$, с вероятностью $p = 1 - \alpha = 1 - 0,05 = 0,95$ нет оснований отвергать гипотезу о нормальном распределении урожайности зерновых культур в генеральной совокупности.

Контрольная работа №2 (пример). По данным проведенного опроса восьми семей известны данные связи расходов населения на продукты питания с уровнем доходов семьи.

Таблица 1– Исходные данные для построения модели

Расходы на продукты питания, Y , тыс. руб.	5,5	4,4	4,9	4,4	5,2	5,8	6,2	6,0
Доходы семьи, X , тыс. руб.	20,4	15,2	16,6	14,8	19,2	23,6	25,1	26,4

Требуется:

1. Построить уравнение парной линейной регрессии y от x .
2. Оценить тесноту связи изучаемых признаков.
3. Оценить качество уравнения регрессии через среднюю ошибку аппроксимации и F-критерий Фишера.
4. Оценить статистическую значимость параметров регрессии с помощью t-критерия Стьюдента.
5. Дать точечный и интервальный прогноз расходов с вероятностью 0,95, принимая уровень дохода семьи равным 110% от среднего уровня.

Решение:

1. Предположим, что связь между доходами семьи и расходами на продукты питания линейная. Для подтверждения нашего предположения построим поле корреляции (Рис. 1).

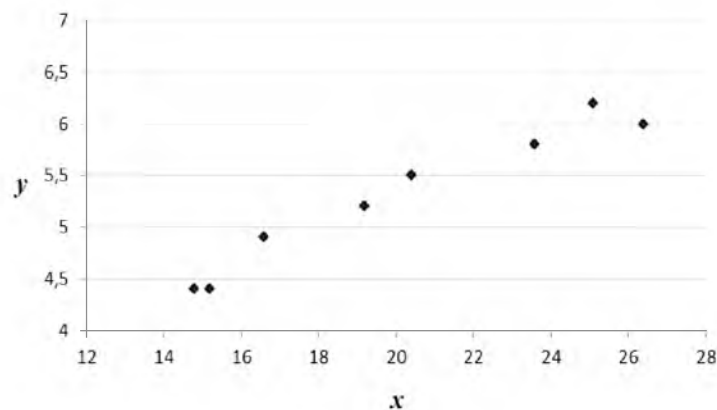


Рис. 1 – Поле корреляции

По графику видно, что точки выстраиваются в некоторую прямую линию.

Расчет параметров b_0 и b_1 линейной регрессии $\hat{y} = b_0 + b_1x$ выполним двумя способами. Для удобства дальнейших вычислений составим таблицу (табл. 2).

1 способ. Решим систему нормальных уравнений относительно b_0 и b_1 :

$$\begin{cases} b_0 \cdot n + b_1 \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n y_i; \\ b_0 \cdot \sum_{i=1}^n x_i + b_1 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i. \end{cases}$$

Подставим суммы из табл. 2 в систему:

$$\begin{cases} b_0 \cdot 8 + b_1 \cdot 161,3 = 42,4 \\ b_0 \cdot 161,3 + b_1 \cdot 3394,4 = 876,28 \end{cases}$$

Решив систему уравнений, получим следующие значения параметров:

$$b_1 = 0,15, b_0 = 2,27.$$

Таблица 2 – Определение параметров модели и оценка её качества

№	x	y	xy	x ²	y ²	$\bar{y}$	$y - \bar{y} = \varepsilon$	A	$(y - \bar{y})^2$
1	20,4	5,5	112,2	416,16	30,25	5,34	0,16	3,0	0,03
2	15,2	4,4	66,88	231,04	19,36	4,55	-0,15	3,5	0,02
3	16,6	4,9	81,34	275,56	24,01	4,76	0,14	2,8	0,02
4	14,8	4,4	65,12	219,04	19,36	4,49	-0,09	2,1	0,01
5	19,2	5,2	99,84	368,64	27,04	5,16	0,04	0,9	0,00
6	23,6	5,8	136,88	556,96	33,64	5,82	-0,02	0,3	0,00
7	25,1	6,2	155,62	630,01	38,44	6,04	0,16	2,5	0,02
8	26,4	6	158,4	696,96	36	6,24	-0,24	4,0	0,06
Итого	161,3	42,4	876,28	3394,4	228,1	42,4	0,0	19,0	0,16
Ср. знач.	20,16	5,30	109,54	424,3	28,51	-	-	2,4	-

2 способ. Рассчитаем параметры регрессии с помощью формул:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 424,3 - 20,16^2 = 17,8;$$

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} = \frac{109,54 - 20,16 \cdot 5,3}{17,8} \approx 0,15,$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 5,3 - 0,15 \cdot 20,16 \approx 2,27.$$

Уравнение парной линейной регрессии имеет вид:

$$\hat{y} = 2,27 + 0,15 \cdot x. \quad (*)$$

С увеличением дохода семьи на 1 тыс. руб. расходы на продукты питания увеличиваются в среднем на 0,15 тыс. руб.

2. Рассчитаем линейный коэффициент парной корреляции r_{yx} :

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{17,8} = 4,22$ – в среднем доход семьи отклоняется от среднего дохода ($\bar{x} = 20,16$) на 4,22 тыс. руб.;

$$V_x = \frac{\sigma_x}{\bar{x}} \cdot 100 = \frac{4,22}{20,16} \cdot 100 = 20,9\% \text{ - вариация признака } x \text{ умеренная;}$$

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2 = 28,51 - 5,3^2 = 0,42;$$

$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{0,42} = 0,65$ – в среднем расходы на питания отклоняются от средних расходов ($\bar{y} = 5,3$) на 0,65 тыс. руб.;

$$V_y = \frac{\sigma_y}{\bar{y}} \cdot 100 = \frac{0,65}{5,3} \cdot 100 = 12,3\% \text{ - вариация признака } y \text{ умеренная;}$$

$$r_{yx} = b_1 \cdot \frac{\sigma_x}{\sigma_y} = 0,15 \cdot \frac{4,22}{0,65} = 0,976.$$

Линейная связь между рассматриваемыми признаками тесная, прямая. Определим коэффициент детерминации:

$$R^2 = r_{yx}^2 = 0,976^2 = 0,952.$$

Вариация результата на 95,2% объясняется вариацией фактора x . На долю прочих факторов, не учитываемых в уравнении регрессии, приходится 4,8%.

3. Расчетные значения результативного признака $\hat{y}_i, i = \overline{1, n}$ определим подставляя в уравнение регрессии (*) значения факторного признака x :

$$\hat{y}_1 = 2,27 + 0,15 \cdot 20,4 = 5,34 \text{ и т. д. (гр. 7 табл. 2)}$$

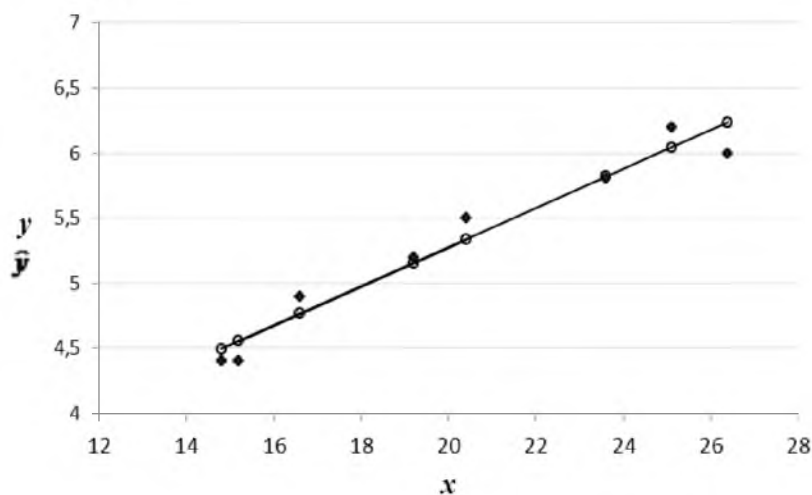


Рис. 2

Рассчитаем индивидуальные ошибки аппроксимации: $A_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100$;

$$A_1 = \left| \frac{y_1 - \hat{y}_1}{y_1} \right| \cdot 100 = \left| \frac{\varepsilon_1}{y_1} \right| \cdot 100 = \left| \frac{0,16}{5,5} \right| \cdot 100 = 3,0\% \text{ и т.д. (гр. 9 табл. 2)}$$

Найдем величину средней ошибки аппроксимации: $\bar{A} = \frac{\sum_{i=1}^n A_i}{n} = \frac{19,0}{8} = 2,4\%.$

Ошибка аппроксимации показывает отличное соответствие расчетных и фактических данных: среднее отклонение составляет 2,4 %.

Для оценки статистической значимости уравнения регрессии в целом определим наблюдаемое значение F-критерия Фишера:

$$F_{набл} = \frac{R^2}{1-R^2} \cdot (n-2) = \frac{0,952}{1-0,952} \cdot (8-2) = 120.$$

и сравним его с критическим, которое находим по таблице значений F-критерия при уровне значимости α и степенях свободы $k_1=m$, $k_2=n-m-1$.

Примем вероятность $P=0,95$, тогда $\alpha=1-P=1-0,95=0,05$. Для уравнения парной линейной регрессии $m=1$, следовательно $k_1=1$, $k_2=8-1-1=6$. По таблицам находим: $F_{кр}(0,05; 1; 6) = 5,99$.

Так как $F_{набл} > F_{кр}$ ($120 > 5,99$), с вероятностью 0,95 уравнение регрессии (*) в целом признается статистически значимым и надежным.

4. Оценку статистической значимости параметров регрессии проведем с помощью t -критерия Стьюдента и путем расчета доверительного интервала каждого из показателей.

Определим стандартные ошибки параметров S_{b_0} , S_{b_1} :

$$S_{осм}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-m-1} = \frac{0,16}{8-1-1} = 0,027; \quad S_{осм} = \sqrt{S_{осм}^2} = \sqrt{0,027} = 0,16;$$

$$S_{b_0} = S_{осм} \cdot \frac{\sqrt{\sum_{i=1}^n x_i^2}}{n \cdot \sigma_x} = 0,16 \cdot \frac{\sqrt{3394}}{8 \cdot 4,22} = 0,283;$$

$$S_{b_1} = \frac{S_{осм}}{\sqrt{n \cdot \sigma_x}} = \frac{0,16}{\sqrt{8 \cdot 4,22}} = 0,014.$$

Наблюдаемые значения t -критерия Стьюдента:

$$t_{b_0} = \frac{b_0}{S_{b_0}} = \frac{2,27}{0,283} = 8,0; \quad t_{b_1} = \frac{b_1}{S_{b_1}} = \frac{0,15}{0,014} = 10,9.$$

Критическое значение t -критерия определим по таблицам для уровня значимости $\alpha=0,05$ и числа степеней свободы $df=n-m-1=8-1-1=6$: $t_{кр} = 2,45$.

Так как $t_{b_1} = 10,9 > t_{кр} = 2,45$, коэффициент b_1 статистически значим.

Так как $t_{b_0} = 8,0 > t_{кр} = 2,45$, параметр b_0 статистически значим.

Рассчитаем доверительные интервалы для b_0 и b_1 :

$$\beta_0 = [b_0 \pm t_{кр} \cdot S_{b_0}] = [2,27 \pm 2,45 \cdot 0,283]; \quad \beta_{0_{\min}} = 1,57; \quad \beta_{0_{\max}} = 2,96;$$

$$\beta_1 = [b_1 \pm t_{кр} \cdot S_{b_1}] = [0,15 \pm 2,45 \cdot 0,014]; \quad \beta_{1_{\min}} = 0,12; \quad \beta_{1_{\max}} = 0,18.$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $P=1-\alpha=0,95$ параметры b_0 и b_1 , находясь в указанных границах, не принимают нулевых значений, т.е. являются статистически значимыми.

5. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Точечный прогноз расходов на продукты питания при прогнозном значении дохода семьи $x_p=1,1 \cdot 20,16=22,2$ тыс. руб. составит:

$$\hat{y}_p = 2,27 + 0,15 \cdot 22,2 = 5,6 \text{ тыс. руб.}$$

Чтобы получить интервальный прогноз, найдем среднюю ошибку прогноза:

$$S_{\hat{y}_p} = S_{осм} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}} = 0,16 \cdot \sqrt{1 + \frac{1}{8} + \frac{(22,2 - 20,16)^2}{8 \cdot 17,8}} = 0,18.$$

Доверительный интервал прогнозируемых расходов составит:

$$y_p = [5,6 \pm 2,45 \cdot 0,18],$$

т.е. при доходе семьи, равном 22,2 тыс. руб., расходы на питание с вероятностью 0,95 находятся в пределах от 5,2 до 6,0 тыс. руб.

Индивидуальная работа - средство проверки умений применять полученные знания для решения задач определенного типа по теме.

Индивидуальная домашняя работа (пример). Зависимость потребления продукта А от среднедушевого дохода по данным 20 семей характеризуется следующим образом: уравнение регрессии $\hat{y}_x = 2 \cdot x^{0,3}$; индекс корреляции $\rho_{xy}=0,9$; остаточная дисперсия $\sigma_{ост}^2=0,06$.

Требуется: провести дисперсионный анализ полученных результатов.

Решение:

Общая сумма квадратов отклонений индивидуальных значений результативного признака y от среднего значения вызвана влиянием множества причин. Условно разделим всю совокупность причин на две группы: изучаемый фактор x и прочие факторы. Если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо и фактор x оказывает существенное воздействие на результат y (R^2 близок к 1).

Любая сумма квадратов связана с числом степеней свободы (df), т.е. с числом свободы независимого варьирования признака. Число степеней свободы связано с числом единиц совокупности n и с числом определяемых по ней констант. Применительно к исследуемой проблеме число степеней свободы должно показать, сколько независимых отклонений из n возможных $[(y_1 - \bar{y}), (y_2 - \bar{y}), \dots, (y_n - \bar{y})]$ требуется для образования данной суммы квадратов.

Имеем два равенства:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2, \\ n-1 = 1 + (n-2).$$

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим средний квадрат отклонений, или, что то же самое, дисперсию на одну степень свободы D .

$$D_{общ} = \frac{\sum (y - \bar{y})^2}{n-1}; \quad D_{факт} = \frac{\sum (\hat{y}_x - \bar{y})^2}{1}; \quad D_{ост} = \frac{\sum (y - \hat{y}_x)^2}{n-2}.$$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F -критерий для проверки гипотезы H_0 :

$$D_{факт} = D_{ост} (F = D_{факт} / D_{ост}).$$

Результаты дисперсионного анализа приведены в таблице 1.

Таблица 1

Вариация результата y	Число степеней свободы	Сумма квадратов отклонений, S	Дисперсия на одну степень свободы, D	$F_{факт}$	$F_{табл}$ $\alpha=0,05$, $k_1=1, k_2=18$
Общая	$df=n-1=19$	6,316	-	-	-
Факторная	$k_1=m=1$	5,116	5,116	76,7	4,41
Остаточная	$k_2=n-m-1=18$	1,200	0,0667	-	-

$$S_{ост} = \sigma_{ост}^2 * n = 0,06 * 20 = 1,2; \\ S_{общ} = S_{ост} (1 - \rho_{xy}^2) = 1,2 : (1 - 0,81) = 6,316; \\ S_{факт} = 6,316 - 1,2 = 5,116; \\ F_{факт} = 0,9^2 / (1 - 0,9^2) * 18 / 1 = 76,7.$$

В силу того, что $F_{факт} > F_{табл}$, гипотеза о случайности различий факторной и остаточной дисперсий отклоняется. Эти различия существенны, статистически значимы, уравнение надежно, значимо, показатель тесноты связи надежен и отражает устойчивую зависимость потребления продукта А от среднедушевого дохода.

7 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

7.1. Основная литература:

1. Количественный анализ в экономике и менеджменте [Электронный ресурс]: учебник / В.А. Малугин, Л.Н. Фадеева. – Электрон. текстовые дан. - М.: НИЦ ИНФРА-М, 2014. – 615 с. - Режим доступа: <http://znanium.com/bookread.php?book=363305#none>

7.2. Дополнительная литература:

1. Вуколов, Э.А. Основы статистического анализа. Практ. по стат. мет. и исслед. операций с исп. пакетов STATISTICA и EXCEL [Электронный ресурс]: учебное пособие / Э.А. Вуколов. – Электрон. текстовые дан. - 2 изд., испр. и доп. - М.: Форум: НИЦ Инфра-М, 2013. - 464 с. – Режим доступа: <http://znanium.com/bookread.php?book=369689>

2. Дайитбегов, Д.М. Компьютерные технологии анализа данных в эконометрике [Электронный ресурс]: монография / Д.М. Дайитбегов. – Электрон. текстовые дан. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с. - Режим доступа: <http://znanium.com/bookread2.php?book=365692>

3. Мхитарян, В.С. Теория вероятностей и математическая статистика [Электронный ресурс]: учеб. пособие / В. С. Мхитарян, Е. В. Астафьева, Ю. Н. Миронкина, Л. И. Трошин; под ред. В. С. Мхитаряна. – Электрон. текстовые дан. - 2-е изд., перераб. и доп. - М.: Московский финансово-промышленный университет «Синергия», 2013. – Режим доступа: <http://znanium.com/bookread.php?book=451329>

4. Статистический анализ данных в MS Excel [Электронный ресурс]: учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. – Электрон. текстовые дан. - М.: ИНФРА-М, 2014. - 320 с. Режим доступа: <http://znanium.com/bookread.php?book=429722>

5. Шапкин, А. С. Задачи с решениями по высшей математике, теории вероятностей, математической статистике, математическому программированию [Электронный ресурс]: учебное пособие для бакалавров / А. С. Шапкин, В. А. Шапкин. – Электрон. текстовые дан. - 8-е изд. - М.: Издательско-торговая корпорация «Дашков и К°», 2013. - 432 с. – Режим доступа: <http://znanium.com/bookread.php?book=430613>

8 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

- Библиотека экономической, бухгалтерской и управленческой литературы <http://eup.ru/>
- Международный журнал «Квантиль» <http://www.quantile.ru/>
- Федеральная служба государственной статистики РФ <http://www.gks.ru/>
- Федеральный образовательный портал ЭСМ <http://ecsocman.hse.ru/>
- Экономический портал <http://institutiones.com/download/books.html>
- Электронный учебник http://www.statsoft.ru/resources/statistica_text_book.php

9 Методические указания для обучающихся по освоению дисциплины

В ходе лекционных занятий обучающимся рекомендуется: 1) вести конспектирование основных положений излагаемого материала; 2) обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации по их применению; 3) задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций; 4) желательно оставить в рабочих конспектах поля, на которых во внеаудиторное время можно сделать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений.

На практических занятиях в зависимости от темы занятия выполняется поиск информации по решению соответствующих содержанию дисциплины проблем, выработка индивидуальных решений, итоговое обсуждение с обменом знаниями, разбор и описание конкретных ситуаций.

Самостоятельная работа обучающихся осуществляется в виде изучения литературных источников и эмпирических данных по публикациям, выполнения индивидуальных заданий, работы с лекционным материалом, самостоятельного изучения отдельных тем дисциплины.

Подготовка к контрольным мероприятиям требует от обучающегося не только повторения пройденного материала на аудиторных занятиях, но поиска и анализа материала, выданного на самостоятельное изучение.

Оценка знаний, умений, навыков, характеризующая этапы формирования компетенций в процессе изучения дисциплины «Анализ данных», проводится в форме текущего контроля и промежуточной аттестации. Текущий контроль успеваемости осуществляется в течение семестра в ходе повседневной учебной работы, обеспечивая оценивание хода освоения дисциплины. В частности, текущий контроль успеваемости проводится с целью определения уровня усвоения обучающимися знаний, оценки формирования у них умений и навыков, своевременного выявления преподавателем недостатков в подготовке обучающихся и принятия необходимых мер по ее корректировке, совершенствованию методики обучения, организации учебной работы и оказания обучающимся индивидуальной помощи. Данный вид контроля стимулирует у обучающихся стремление к систематической самостоятельной работе по изучению дисциплины. Текущий контроль успеваемости осуществляется на практических занятиях, а также в ходе индивидуальных консультаций с преподавателем. К оценочным средствам для проведения текущего контроля успеваемости по дисциплине относятся выполнение индивидуальных домашних заданий и контрольных работ.

Промежуточная аттестация обучающихся позволяет определить степень достижения запланированных результатов обучения по дисциплине и проводится в форме зачета. Данная форма контроля включает в себя теоретические вопросы, позволяющие оценить уровень освоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и навыков. Форма проведения зачета (устная, письменная) определяется преподавателем. По результатам зачета выставляется оценка: «зачтено», «не зачтено».

10 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем

При осуществлении образовательного процесса по дисциплине используется следующее программное обеспечение и информационные справочные системы:

1. Desktop Education ALNG LicSAPk OLVS E 1Y Academic Edition Enterprise.
2. СДО «Прометей».
3. IBM / D0FRELL / IBM SPSS Statistics Base Academic Authorized User License на условиях Академической лицензии.

11 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

№ п/п	Наименование оборудованных учебных аудиторий (помещений)	Перечень основного оборудования, приборов и материалов
1.	Учебная аудитория для проведения занятий лекционного типа: аудитория 507 «Инновационно - образовательный центр компьютерных технологий»	Оснащена специализированной мебелью, мультимедийная система, акустическая система, компьютеры.
2.	Учебная аудитория для проведения занятий семинарского типа: 505 «Лаборатория телекоммуникационных технологий и сетевого администрирования» (компьютерный класс)	Компьютеры, аудиторная доска - (мультимедийная)
3.	Учебная аудитория для самостоятельной работы и проведения групповых и индивидуальных консультаций: аудитория 505 «Лаборатория телекоммуникационных технологий и сетевого администрирования» (компьютерный класс)	
4.	Учебная аудитория для текущего контроля и промежуточной аттестации: аудитория 505 «Лаборатория телекоммуникационных технологий и сетевого администрирования» (компьютерный класс)	

12 Иные сведения и (или) материалы

12.1 Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине

При изучении дисциплины «Анализ данных» используется сочетание отдельных видов учебной работы с методами и формами активизации познавательной деятельности обучающихся с целью достижения запланированных результатов обучения и формирования соответствующих компетенций.

Методы активного и интерактивного обучения
при разных видах учебных занятий

№	Методы активного и интерактивного обучения	Лекции	Практические (семинарские) занятия	Лабораторные работы	Самостоятельная работа
1	Проблемная лекция	+			
2	Решение ситуационных задач		+		+