

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП

А.В. Григорьев

«13» июня 2023г.

УТВЕРЖДАЮ
Заведующий кафедрой
информационных технологий

А.Н. Марьенков

«13» июня 2023г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«ВЕРОЯТНО-СТАТИСТИЧЕСКИЕ МЕТОДЫ В АНАЛИЗЕ ДАННЫХ»

Составитель	Головки Ю.А., к.т.н., доцент кафедры информационных технологий
Направление подготовки / специальность	09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА
Направленность (профиль) ОПОП	ПРИКЛАДНАЯ ИНФОРМАТИКА В СОЦИАЛЬНЫХ НАУКАХ
Квалификация (степень)	бакалавр
Форма обучения	очная
Год приёма	2023
Курс	2
Семестр	3

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Вероятно-статистические методы в анализе данных» являются: изучение статистическим методам, используемым при решении исследовательских задач, связанных с обработкой реальных данных, собираемых и обрабатываемых в процессе прикладного исследования.

1.2. Задачи освоения дисциплины (модуля):

- изучить методики построения дизайна исследования;
- научиться собирать и подготавливать данные;
- освоить методы статистической обработки данных.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина «Вероятно-статистические методы в анализе данных» относится к обязательной части и осваивается в 3 семестре.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения, навыки, формируемые предшествующими учебными дисциплинами (модулями):

- Математические основы информационных технологий и вычислительной техники;
- Введение в информационные технологии.

Знания:

- основных понятий и методов математического анализа, аналитической геометрии и теории алгебраических систем;
- основных понятий информатики.

Умения:

- использовать математические методы и модели для решения прикладных задач;
- использовать программные и аппаратные средства персонального компьютера.

Навыки:

- овладеть методами количественного анализа процессов обработки, поиска и передачи информации;
- поиска информации в глобальной информационной сети Интернет и работы с офисными приложениями (текстовыми процессорами, электронными таблицами, средствами подготовки презентационных материалов, СУБД и т.п.).

2.3. Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):

- Управление данными;
- Программная инженерия.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс освоения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки:

ОПК-7 Способен разрабатывать алгоритмы и программы, пригодные для практического применения.

ПК-6 Способен обрабатывать и анализировать данные для подготовки аналитически х

решений, экспертных заключений и рекомендаций.

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
ОПК-7 Способен разрабатывать алгоритмы и программы, пригодные для практического применения	ОИПК-7.1.1 методы алгоритмизации, языки и технологии программирования, пригодные для практического применения	ОИПК-7.2.1 применять методы алгоритмизации, языки и технологии программирования при решении профессиональных задач	ОИПК-7.3.1 программирования, отладки и тестирования прототипов программно- технических комплексов задач
ПК-6 Способен обрабатывать и анализировать данные для подготовки аналитически х решений, экспертных заключений и рекомендаций	ИПК-6.1.1 методы обработки данных для подготовки аналитических решений	ИПК-6.2.1 использовать соответствующие методы обработки данных для подготовки аналитических решений	ИПК-6.3.1 обработкой данных для подготовки аналитических решений

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объём дисциплины составляет 4 зачётных единицы, в том числе 54 часа, выделенных на контактную работу обучающихся с преподавателем (из них 18 часов – лекции, 36 часов – лабораторные работы), и 90 часов – на самостоятельную работу обучающихся.

Таблица 2 – Структура и содержание дисциплины (модуля)

Раздел, тема дисциплины (модуля)	Семестр	Контактная работа (в часах)			Самост. работа		Форма текущего контроля успеваемости, форма промежуточной аттестации
		Л	ПЗ	ЛР	КР	СР	
<i>Тема 1. Элементы комбинаторики</i>	3	2		4		10	Отчет по лабораторной работе
<i>Тема 2. Классическая формула вероятности</i>	3	2		4		10	Отчет по лабораторной работе
<i>Тема 3. Повторные независимые испытания</i>	3	2		4		10	Отчет по лабораторной работе

<i>Тема 4. Дискретная случайная величина</i>	3	2		4		10	Отчет по лабораторной работе
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда.</i>	3	2		4		10	Отчет по лабораторной работе
<i>Тема 6. Виды распределения. Критерии Пирсона.</i>	3	2		4		10	Отчет по лабораторной работе
<i>Тема 7. Парный коэффициент корреляции</i>	3	2		6		10	Отчет по лабораторной работе
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии</i>	3	4		6		20	Отчет по лабораторной работе
Итого		18		36		90	Диф.зачет

Таблица 3 – Матрица соотнесения разделов, тем учебной дисциплины (модуля) и формируемых компетенций

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции		Общее количество компетенций
		ОПК-7	ПК-6	
<i>Тема 1. Элементы комбинаторики</i>	16		+	1
<i>Тема 2. Классическая формула вероятности</i>	16	+	+	2
<i>Тема 3. Повторные независимые испытания</i>	16		+	1
<i>Тема 4. Дискретная случайная величина</i>	16		+	1
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда</i>	16	+	+	2
<i>Тема 6. Виды распределения. Критерии Пирсона</i>	16	+		1
<i>Тема 7. Парный коэффициент корреляции</i>	18	+		1
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии</i>	30	+	+	2
Итого	144			2

Краткое содержание каждой темы дисциплины (модуля)

Тема 1. Элементы комбинаторики

Понятия комбинаторики. Основные правила комбинаторики. Соединение без повторений. Бином Ньютона. Треугольник Паскаля. Непосредственные подсчеты.

Тема 2. Классическая формула вероятности

Элементы теории вероятности. Частота событий. Классическая теория вероятности. Геометрическая теория вероятности. Несовместные события. Теорема сложения вероятностей. Независимые события. Теорема умножения вероятностей. Зависимые события. Сложение и умножение вероятностей.

Тема 3. Повторные независимые испытания

Понятие независимых испытаний. Схема Бернулли. Наиболее вероятное число появлений событий. Приближение Пуассона. Локальная и интегральная теорема Лапласа. Отклонение частоты появления события от его вероятности.

Тема 4. Дискретная случайная величина

Понятие случайной величины. Определение дискретной случайной величины. Способы задания дискретной случайной величины. Закон распределения дискретной случайной величины. Числовые характеристики дискретных случайных величин.

Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда

Составление сводных таблиц. Преобразование формы информации. Проверка данных. Группировка данных. Повышение наглядности собранных данных. Оценка уровня связи между рассматриваемыми характеристиками.

Тема 6. Виды распределения. Критерии Пирсона

Основные понятия математической статистики: случайное событие, вариация, частота, вероятность, распределение вероятности, выборка, генеральная совокупность, вариационный ряд, полигон частот, гистограмма, кривая распределения. Характеристики статистических совокупностей: меры положения, меры изменчивости, меры связи меры возможной ошибки и достоверности основные этапы статистической обработки результатов исследований.

Тема 7. Парный коэффициент корреляции

Сбор и анализ данных, определение формы корреляционной связи (линейная, криволинейная). Метод случайной выборки объектов из однородной совокупности. Визуальный анализ данных. Вычисление показателя тесноты корреляционной связи. Оценка статистической значимости показателя тесноты корреляционной связи.

Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии

Проверка гипотезы о виде закона распределения. Вычисление оценок коэффициентов регрессии. Проверка адекватности математической модели. Переход к исходным физическим переменным. Дисперсия параметра оптимизации.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

При подготовке к лекционным и практическим занятиям необходимо пользоваться учебно-методической литературой из п.8. Лекции необходимо проводить в интерактивном формате с использованием презентаций, созданных в прикладном пакете Microsoft Office PowerPoint.

При подготовке к лабораторным занятиям необходимо воспользоваться учебно-методической литературой (основной и дополнительной) и информационно-справочными ресурсами из п.8.

По окончании курса необходимо провести итоговое тестирование с целью проверки сформированности указанных в рабочей программе компетенций (специализированных образовательных площадках Moodle).

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Таблица 4 – Содержание самостоятельной работы обучающихся

Вопросы, выносимые на самостоятельное изучение	Кол-во часов	Форма работы
<i>Тема 1. Элементы комбинаторики (Непосредственные подсчеты)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 2. Классическая формула вероятности (Условная и безусловная вероятность)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 3. Повторные независимые испытания (Решение задач)</i>	10	Решение задач по теме самостоятельного обучения
<i>Тема 4. Дискретная случайная величина (Функция распределения случайной величины)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда (Вычисление мер центральной тенденции)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 6. Виды распределения. Критерии Пирсона (Сравнение совокупностей)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 7. Парный коэффициент корреляции (Метод наименьших квадратов)</i>	10	Изучения материалов, конспект в электронном виде
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии (Эконометрика)</i>	20	Изучения материалов, конспект в электронном виде

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины (модуля), выполняемые обучающимися самостоятельно

В процессе изучения дисциплины предусмотрены следующие виды и формы письменных работ для самостоятельного выполнения:

- аудиторная контрольная работа;
- типовой расчет - внеаудиторная работа;
- домашнее задание, как теоретического, так и практического характера;

- лабораторная работа, полностью или частично выполняемая в математических пакетах;
- итоговое тестирование.

Контрольные работы и зачетная работа выполняется студентом в аудитории.

В установленный срок студент сдает типовой расчет и устно отчитывается преподавателю по выполненной работе. Зачетную работу студент сдает в том случае, если по результатам семестрового контроля он набрал рейтинг менее 60 баллов. Лабораторные работы выполняются в аудитории и оцениваются как зачет/незачет. Для получения зачета по дисциплине по результатам текущего семестрового контроля студенту необходимо иметь рейтинг по контрольным работам не менее 60 баллов и зачет по всем лабораторным работам.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6.1. Образовательные технологии

Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
<i>Тема 1. Элементы комбинаторики</i>	<i>Обзорная лекция</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 2. Классическая формула вероятности</i>	<i>Лекция-диалог</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 3. Повторные независимые испытания</i>	<i>Лекция-диалог</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 4. Дискретная случайная величина</i>	<i>Лекция</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда</i>	<i>Лекция</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>

<i>Тема 6. Виды распределения. Критерии Пирсона</i>	<i>Лекция</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 7. Парный коэффициент корреляции</i>	<i>Лекция-диалог</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии</i>	<i>Лекция</i>	<i>Не предусмотрено</i>	<i>выполнение контрольной работы, выполнение лабораторной работы</i>

6.2. Информационные технологии

При реализации различных видов учебной и внеучебной работы используются следующие информационные технологии:

- использование возможностей Интернета в учебном процессе (использование информационного сайта преподавателя (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление учащихся с оценками и т.д.));
- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источников информации;
- использование возможностей электронной почты преподавателя;
- использование средств представления учебной информации (электронных учебных пособий и практикумов, применение новых технологий для проведения очных (традиционных) лекций и семинаров с использованием презентаций и т.д.);
- использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е.
- информационные ресурсы (доступ к мировым информационным ресурсам, на базе которых строится учебный процесс);
- использование виртуальной обучающей среды (или системы управления обучением LMS Moodle «Электронное образование») или иных информационных систем, сервисов и мессенджеров.

6.3. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

6.3.1. Программное обеспечение

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда

Google Chrome	Браузер
Microsoft Office 2013, Microsoft Office Project 2013 , Microsoft Office Visio 2013	Пакет офисных программ
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
R	Программная среда вычислений

6.3.2. Современные профессиональные базы данных и информационные справочные системы

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем» <https://library.asu.edu.ru/catalog/>;
2. Справочная правовая система КонсультантПлюс. Содержится огромный массив справочной правовой информации, российское и региональное законодательство, судебную практику, финансовые и кадровые консультации, консультации для бюджетных организаций, комментарии законодательства, формы документов, проекты нормативных правовых актов, международные правовые акты, правовые акты, технические нормы и правила. <http://www.consultant.ru>;
3. Электронный каталог «Научные журналы АГУ» <https://journal.asu.edu.ru/>;
4. Электронно-библиотечная система BOOK.ru <https://book.ru>;
5. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента» Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей доступ через Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог содержит более 15 000 наименований изданий. www.studentlibrary.ru.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине «Вероятно-статистические методы в анализе данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6 – Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

Контролируемый раздел, тема дисциплины (модуля)	Код контролируемой компетенции	Наименование оценочного средства
<i>Тема 1. Элементы комбинаторики</i>	ПК-6	Лабораторная работа №1

Контролируемый раздел, тема дисциплины (модуля)	Код контролируемой компетенции	Наименование оценочного средства
<i>Тема 2. Классическая формула вероятности</i>	ОПК-7 ПК-6	Лабораторная работа №2
<i>Тема 3. Повторные независимые испытания</i>	ПК-6	Лабораторная работа №3
<i>Тема 4. Дискретная случайная величина</i>	ПК-6	Лабораторная работа №4
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда</i>	ОПК-7 ПК-6	Лабораторная работа №5, 6
<i>Тема 6. Виды распределения. Критерии Пирсона</i>	ОПК-7	Лабораторная работа №7
<i>Тема 7. Парный коэффициент корреляции</i>	ОПК-7	Лабораторная работа №8
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии</i>	ОПК-7 ПК-6	Лабораторная работа №9 Итоговое тестирование

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 8 – Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые

Шкала оценивания	Критерии оценивания
	выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание по подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задания

7.3. Контрольные задания и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

<i>Тема 1. Элементы комбинаторики</i>
<i>Тема 2. Классическая формула вероятности</i>
<i>Тема 3. Повторные независимые испытания</i>
<i>Тема 4. Дискретная случайная величина</i>
<i>Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда</i>
<i>Тема 6. Виды распределения. Критерии Пирсона</i>
<i>Тема 7. Парный коэффициент корреляции</i>
<i>Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии</i>

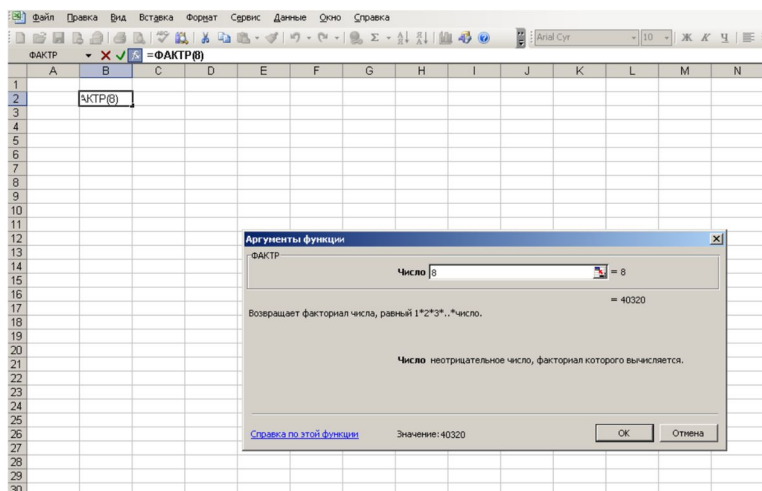
Тема 1. Элементы комбинаторики Лабораторная №1

При подсчете количества комбинаций элементов чаще всего используют перестановки, сочетания, размещения.

Перестановкой из n элементов (например чисел $1, 2, \dots, n$) называется всякий упорядоченный набор из этих элементов.

Для вычисления факториала в Excel используется функция ФАКТР(), относящаяся к математическим (Меню: Вставка — Функция — Математические). Аргументом функции является количество элементов n для которых нужно подсчитать перестановки. Сколькими способами могут восемь человек стать в очередь к театральной кассе.

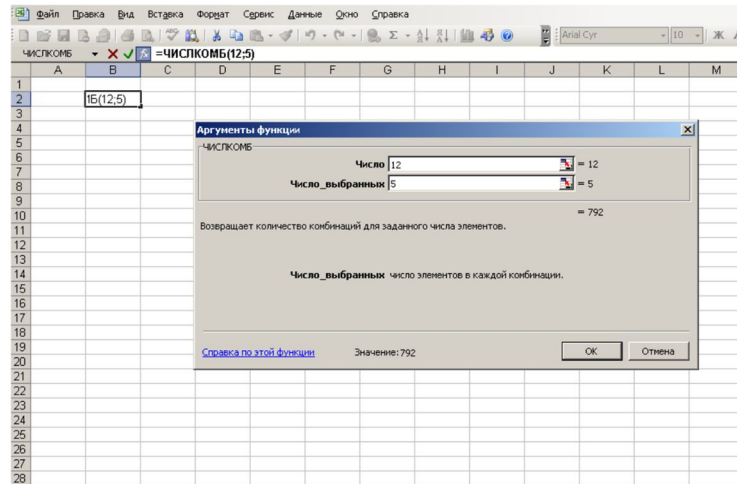
Решение: Нужно подсчитать сколько комбинаций можно составить из 8 элементов, отличающиеся друг от друга порядком следования элементов.



Количество сочетаний в Excel можно вычислить с помощью функции ЧИСЛОКОМБ(), также относящейся к математическим (Меню: Вставка — Функция — Математические). Аргументами функции является количество элементов n из которых выбирают наборы (Число) и количество элементов k , входящих в наборы (Число выбранных) для которых нужно подсчитать сочетания.

Пример типовой задачи на сочетания: Сколькими способами из двенадцати человек можно создать комиссию из пяти членов.

Решение: Нужно подсчитать сколько комбинаций можно составить по 5 элементов из 12, отличающиеся друг от друга составом.



Для вычисления размещений записывают формулу в ручную, как отношение факториалов, используя функцию ФАКТР().

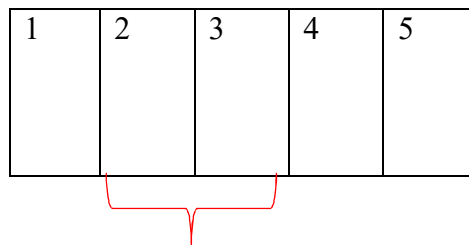
Пример типовой задачи на размещения. Алфавит некоторого языка содержит 30 букв. Сколько шестибуквенных «слов» (цепочка букв от пробела до пробела) можно составить из букв этого алфавита, если буквы в «словах» не повторяются.

Решение. Нужно подсчитать сколько комбинаций можно составить по 6 элементов из 30, отличающиеся друг от друга составом или порядком следования элементов.

$$A_{30}^6 = \frac{30!}{(30-6)!} = \frac{30!}{24!} = 427518000.$$

1. Сколькими способами можно расставить 5 книг на полке? Сколькими способами можно расставить 5 книг на полке таким образом, что две определенные книги оказались рядом?

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$



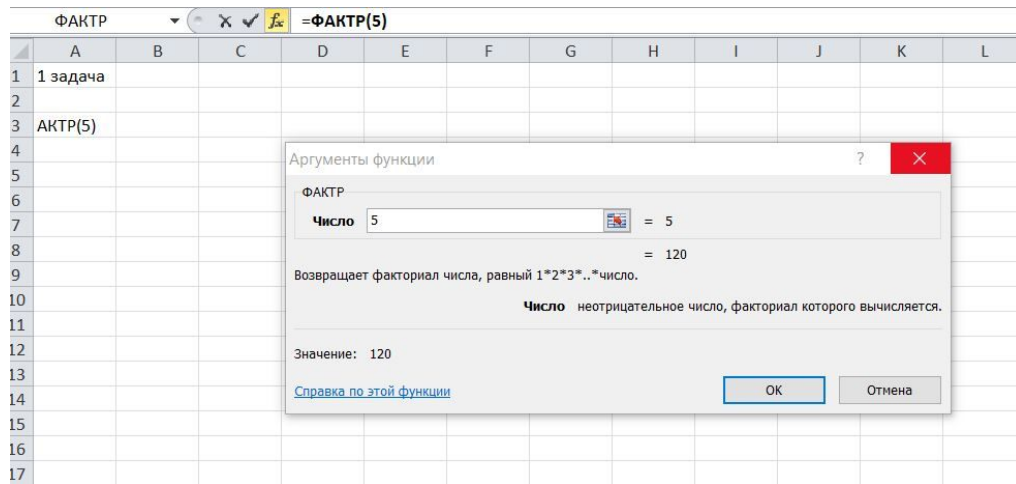
Объединим в один том

У нас есть всего 5 книг, 2 книги обязательно должны быть рядом (по условию). Свяжем две книги в один том, и тогда у нас получится 4 тома. Неизвестно где эти 2 книги могут оказаться среди остальных. Таких способов перестановки существует 4. Еще между

собой эти книги будут меняться (стоять в одном порядке или в другом). Таких способов: $4!$ и $2!$, поэтому работает правило произведения:

$$4! \cdot 2! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 2 = 48.$$

В ячейках вы пишете не готовый ответ, а формулу, по которой вы эту величину находите.



- 2. Предприятие может предоставить работу по одной специальности 4 женщинам, по другой – 6 мужчинам, по третьей – 3 работникам независимо от пола. Сколькими способами можно заполнить вакантные места, если имеются 14 претендентов: 6 женщин и 8 мужчин?**

1 спец. = 4 женщины

2 спец. = 6 мужчин

3 спец. = 3 работника

Необходимо посчитать количество способов, которыми можно заполнить 1 специальность:

$$C_6^4 = \frac{6!}{(6-4)! \cdot 4!} = \frac{6!}{2! \cdot 4!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 4} = \frac{30}{2} = 15$$

$$C_8^6 = \frac{8!}{(8-6)! \cdot 6!} = \frac{8!}{2! \cdot 6!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8}{1 \cdot 2 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = \frac{56}{2} = 28$$

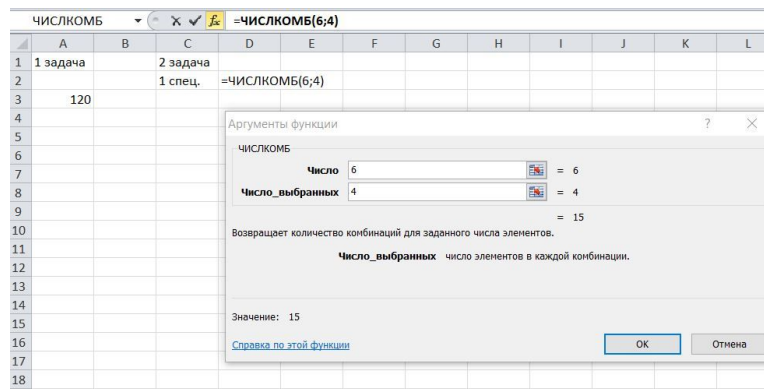
Не забываем, что уже были выбраны 4 женщины и их осталось 2, а также выбрали 6 мужчин и их осталось 2.

Теперь необходимо выбрать трех работников из оставшихся людей:

$$C_4^3 = \frac{4!}{(4-3)! \cdot 3!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 3} = 4$$

Далее применим правило умножения (произведения):

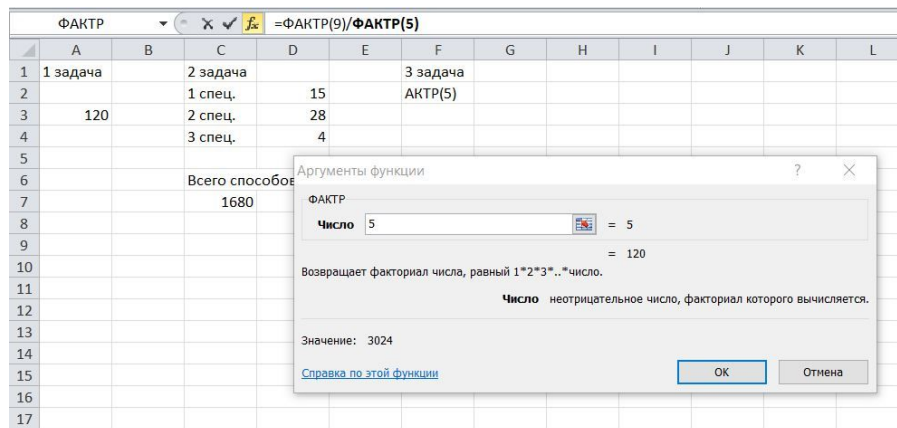
$$C_6^4 \cdot C_8^6 \cdot C_4^3 = 15 \cdot 28 \cdot 4 = 1680.$$



3. В пассажирском поезде 9 вагонов. Сколькими способами можно рассадить в поезде 4 человека, при условии, что все они должны ехать в различных вагонах?

Необходимо использовать размещение:

$$A_9^4 = \frac{9!}{(9-4)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 3024$$



Тема 2. Классическая формула вероятности

Вероятностью $P(A)$ события A называется отношение числа благоприятствующих этому событию случаев (исходов опыта) m к числу всех возможных случаев (исходов опыта) n , образующих полную группу несовместных равновероятных событий:

$$P(A) = \frac{m}{n}.$$

Иногда вероятность события A удобно вычислять через вероятность противоположного события \bar{A} :

$$P(A) + P(\bar{A}) = 1 \rightarrow P(A) = 1 - P(\bar{A})$$

Задачи:

- Задумано двузначное число. Найти вероятность того, что задуманным числом окажется:
 - случайно названное число;
 - случайно названное число, цифры которого различны.
- Монета брошена три раза. Найти вероятность того, что хотя бы один раз появится изображение герба.

3. В коробке семь одинаковых пронумерованных кубиков. Наудачу извлекают все кубики по очереди. Найти вероятность того, что номера кубиков появятся в убывающем порядке.
4. В пачке 30 пронумерованных карточек. Наудачу взяли 3 карточки. Какова вероятность того, что взяли карточки с номерами 12, 24, 30?
5. Среди 25 участников розыгрыша лотереи находятся 10 девушек. Разыгрывается 5 призов. Вычислить вероятность того, что обладателями двух призов окажутся девушки.
6. В коробке 4 белых и 5 чёрных футболок. Наугад вытаскивают две футболки. Найти вероятность того, что одна из футболок белая, другая чёрная.
7. При подготовке к зачёту студент выучил 60 из необходимых 90 вопросов. Какова вероятность того, он сдаст зачёт, если для этого нужно ответить не менее чем на два из трёх предложенных вопросов?
8. Из партии, состоящей из 20 игроков, для проверки произвольно отбирают три игрока. Партия содержит 2 игрока с дефектами. Какова вероятность того, что в число отобранных игроков попадут только два бракованных игрока?

Тема 3. Повторные независимые испытания

Пусть производится несколько испытаний, в каждом из которых может появиться событие A . Если вероятность события A в каждом испытании не зависит от того, появилось или не появилось это событие в других испытаниях, то такие испытания называются независимыми относительно события A .

Пусть производится серия из n независимых испытаний, в каждом из которых может появиться событие A с вероятностью $P(A) = p$. Вероятность того, что событие A не наступит, для каждого испытания равна

$$P(\bar{A}) = 1 - p = q.$$

Вероятность того, что при n независимых испытаниях событие A появится ровно k раз ($P_n(k)$), вычисляется по формуле, называемой формулой Бернулли:

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}$$

Если требуется вычислить вероятность того, событие A наступит не менее k_0 раз в n испытаниях, то используют формулу сложения вероятностей:

Формула Бернулли представляется функцией БИНОМРАСП ($k, n, p, \text{ЛОЖЬ}$), где k – количество появления события, n – число независимых испытаний; p – вероятность появления события; "ЛОЖЬ" – указание на то, что определяется вероятность появления ровно k событий. В случае, когда последний аргумент функции равен "ИСТИНА", функция возвращает вероятность того, что в n испытаниях событие наступит не менее k раз.

Если нужно вычислить вероятность того, что событие наступит менее k_0 раз или не более k_0 раз, то используют противоположную вероятность:

$$P_n(k < k_0) = 1 - P_n(k \geq k_0)$$

или

$$P_n(k \leq k_0) = 1 - P_n(k \geq k_0 + 1)$$

Пример. В освещении помещения фирмы используются 14 лампочек. Для каждой лампочки вероятность того, что она останется исправной в течение года равна. Какова вероятность того, что в течение года останутся исправными: а) половина лампочек; б) не менее половины лампочек.

Решение.

а)

	A	B	C	D	E	F
1						
2		Формула Бернулли				
3						
4		Количество появлений события	7	k		
5		Число независимых испытаний	14	n		
6		Вероятность появления события в одном испытании	7/8	p		
7		Событие наступит ровно k раз	ЛОЖЬ	локальная		
8		Событие наступит не менее k раз	ИСТИНА	интегральная		
9						
10		Вероятность	0,000643	"=БИНОМРАСП(C4;C5;C6;C7)"		
11						
12						
13						
14						
15						

б)

	A	B	C	D	E	F
1						
2		Формула Бернулли				
3						
4		Количество появлений события	7	k		
5		Число независимых испытаний	14	n		
6		Вероятность появления события в одном испытании	7/8	p		
7		Событие наступит ровно k раз	ЛОЖЬ	локальная		
8		Событие наступит не менее k раз	ИСТИНА	интегральная		
9						
10		Вероятность	0,000731	"=БИНОМРАСП(C4;C5;C6;C8)"		
11						
12						
13						
14						
15						

Для определения вероятности применяется функция ПУАССОН(). Аргументами функции являются количество появлений события k , среднее (математическое ожидание количества появлений события в n испытаниях), логическая переменная Интегральная, принимающая значение «ЛОЖЬ» если определяется вероятность появления ровно k событий или «ИСТИНА», если определяется вероятность того, что в n испытаниях событие наступит не менее k раз.

Пример. Вероятность выигрыша в лотерее равна 0,001. Какова вероятность того, что среди 1000 наугад купленных билетов: а) ровно 5 выигрышных; б) не менее 5 выигрышных?

Решение.

	A	B	C	D	E	F
1						
2		Формула Пуассона				
3						
4		Количество появлений события	5	k		
5		Число независимых испытаний	1000	n		
6		Вероятность появления события в одном испытании	0,001	p		
7		Среднее	1	"=C5*C6		
8		Событие наступит ровно k раз	ЛОЖЬ	локальная		
9		Событие наступит не менее k раз	ИСТИНА	интегральная		
10						
11		Вероятность	0,9994058	"=ПУАССОН(C4;C7;C9)		
12						
13						

	A	B	C	D	E	F
1						
2		Формула Пуассона				
3						
4		Количество появлений события	5	k		
5		Число независимых испытаний	1000	n		
6		Вероятность появления события в одном испытании	0,001	p		
7		Среднее	1	"=C5*C6		
8		Событие наступит ровно k раз	ЛОЖЬ	локальная		
9		Событие наступит не менее k раз	ИСТИНА	интегральная		
10						
11		Вероятность	0,0030657	"=ПУАССОН(C4;C7;C8)		
12						

Задачи:

1. Мастер обслуживает шесть однотипных станков. Вероятность того, что станок потребует внимания мастера в течение дня, равна 0,2. Найти вероятность того, что в течение дня мастеру придется вмешаться в работу станков: а) меньше одного раза; б) больше двух раз; в) от двух до пяти раз.
2. В освещении помещения фирмы используются 20 лампочек. Для каждой лампочки вероятность того, что она останется исправной в течение года, равна $\frac{3}{5}$. Какова вероятность того, что в течение года придется заменить не меньше половины всех лампочек?
3. Вероятность того, что студент забросит мяч в корзину, равна 0,4. Студент произвел 24 броска. Найти наиболее вероятное число попаданий и соответствующую вероятность.
4. Вероятность попадания в мишень при одном выстреле для данного охотника равна 0,9 и не зависит от номера выстрела. Найти наиболее вероятное число попаданий в мишень при 7 выстрелах и соответствующую этому числу вероятность.
5. Стрелок стреляет по цели до первого попадания. Найти вероятность того, что у стрелка останется хотя бы один неизрасходованный патрон, если он получил 7 патронов и вероятность попадания в цель при одиночном выстреле равна $\frac{1}{7}$.

Тема 4. Дискретная случайная величина

Дискретной случайной величиной (X) называется случайная величина, которая в результате испытания принимает отдельные значения (x_1, x_2, \dots) с определёнными вероятностями (p_1, p_2, \dots). Число возможных значений дискретной случайной величины может быть конечным и бесконечным.

Соотношение, устанавливающее связь между отдельными возможными значениями случайной величины и соответствующими им вероятностями, называется законом распределения дискретной случайной величины:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

$$\sum_{i=1}^n p_i = 1$$

Закон (ряд) распределения можно изобразить графически, в виде точек с координатами (x_i, p_i), соединённых отрезками. Получим многоугольник распределения вероятностей (полигон распределения).

Дискретная случайная величина может быть задана функцией распределения.

Функцией распределения случайной величины X называется функция $F(x)$, выражающая вероятность того, что X примет значение, меньшее чем x :

$$F(x) = P(X < x)$$

Пример. Закон распределения случайной величины X :

X	0	1	2	3
P	0,198	0,457	0,293	0,052

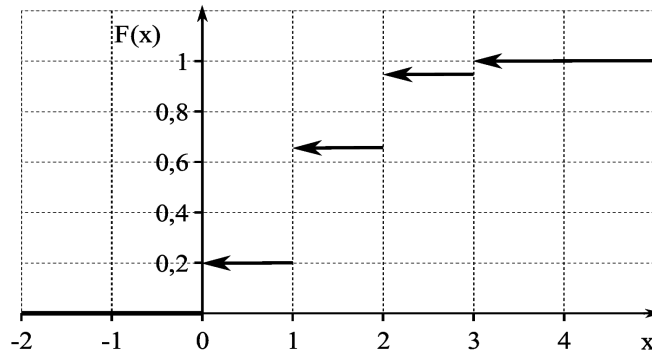
Проверка: $\sum_{i=1}^4 p_i = 0,198 + 0,457 + 0,293 + 0,052 = 1$

Функция распределения вероятностей случайной величины X :

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,198, & 0 < x \leq 1 \\ 0,198 + 0,457 = 0,655, & 1 < x \leq 2 \\ 0,198 + 0,457 + 0,293 = 0,948, & 2 < x \leq 3 \\ 0,198 + 0,457 + 0,293 + 0,052 = 1, & x > 3 \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,198, & 0 < x \leq 1 \\ 0,655, & 1 < x \leq 2 \\ 0,948, & 2 < x \leq 3 \\ 1, & x > 3 \end{cases}$$

График функции распределения:



Числовые характеристики дискретной случайной величины.

Математическое ожидание случайной величины X .

$$M(X) = \sum_{i=1}^n p_i \cdot x_i$$

Дисперсия случайной величины X .

$$D(X) = M((X - M(X))^2) = M(X^2) - (M(X))^2 = \sum_{i=1}^n p_i \cdot x_i^2 - (M(X))^2$$

Среднее квадратическое отклонение.

$$\sigma(X) = \sqrt{D(X)}$$

Обобщёнными числовыми характеристиками для случайных величин в теории вероятностей, а также математической статистике являются начальные и центральные моменты.

Начальным моментом k -го порядка случайной величины X называют математическое ожидание от величины в k -ой степени:

$$v_k = M(X^k) = \sum_{i=1}^n p_i \cdot x_i^k, \quad k = 1, 2, 3, \dots$$

$$v_1 = M(X) = \sum_{i=1}^n p_i \cdot x_i$$

Начальный момент первого порядка:

$$v_2 = M(X^2) = \sum_{i=1}^n p_i \cdot x_i^2$$

Начальный момент второго порядка:

$$v_3 = M(X^3) = \sum_{i=1}^n p_i \cdot x_i^3$$

Начальный момент третьего порядка:

Центральным моментом k -го порядка случайной величины X называют математическое ожидание от величины $(X - M(X))^k$:

$$\mu_k = M((X - M(X))^k) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^k, \quad k = 1, 2, 3, \dots$$

Центральный момент первого порядка: $\mu_1 = M(X - M(X)) = M(X) - M(X) = 0$

Центральный момент второго порядка:

$$\mu_2 = M((X - M(X))^2) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^2 = D(X) = M(X^2) - (M(X))^2 = v_2 - v_1^2$$

$$\mu_3 = M((X - M(X))^3) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^3$$

Центральный момент третьего порядка:

Задание.

1. Построить многоугольник распределения.
2. Составить функцию распределения и построить её график.
3. Найти начальные и центральные моменты первого, второго и третьего порядка.
4. Найти числовые характеристики случайной величины (математическое ожидание, дисперсию, среднее квадратическое отклонение).

Вариант № 1

X	23	28	34	45	47	52	56	67	69	73
P	0,01	0,03	0,04	0,13	0,15	0,28	0,16	0,08	0,06	0,06

Вариант № 2

X	35	40	46	57	59	64	68	79	81	85
P	0,05	0,07	0,14	0,31	0,18	0,11	0,05	0,04	0,03	0,02

Тема 5. Первичная обработка эмпирических данных. Числовые характеристики вариационного ряда

Имеющийся набор эмпирических данных является выборкой из генеральной совокупности.

Набор данных расположенный в порядке возрастания, называется вариационным рядом. Если набор данных достаточно большой, то удобнее всего представить его в виде интервального вариационного ряда.

Для построения интервального вариационного ряда необходимо выполнить следующие действия:

1. Имеющиеся данные располагают в порядке возрастания.

H33		f _x							
	A	B	C	D	E	F	G	H	I
1	104								
2	112								
3	114	xmin		104		R		260	
4	115	xmax		364					
5	118	n		80					
6	118	k		7,318259					
7	119			7 округляем до целого значения					
8	121								
9	125								
10	126	h		37,14286					
11	129			37					
12	132								
13	133	x0		85,42857					
14	135								
15	135	границы интервалов					частоты		
16	136	1	85,42857	122,5714				8	
17	140	2	122,5714	159,7143				17	много пассажиров
18	141	3	159,7143	196,8571				21	пик перевозок
19	145	4	196,8571	234				11	
20	146	5	234	271,1429				14	
21	148	6	271,1429	308,2857				5	
22	148	7	308,2857	345,4286				2	мало пассажиров
23	149	8	345,4286	364				2	
24	150							80	
25	158								

Тема 6. Виды распределения. Критерии Пирсона

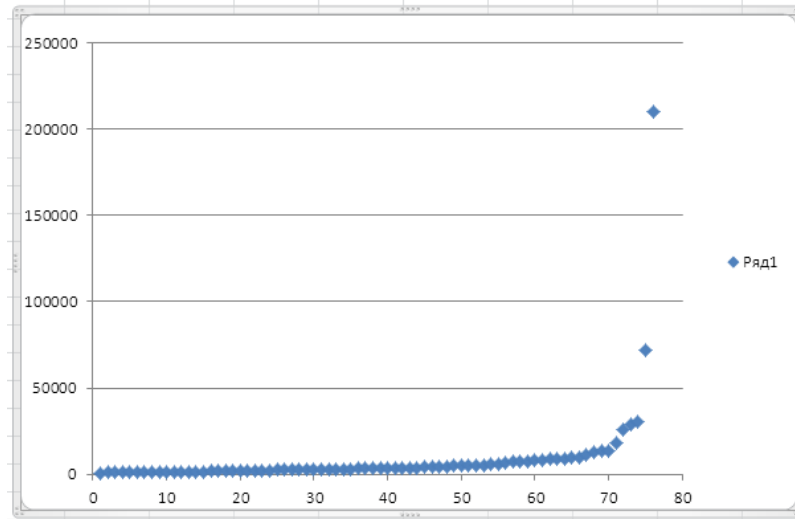
Для построенного интервального ряда проверить гипотезу о виде распределения. Для начала проверим гипотезу о виде распределения в Excel.

Для анализа возьмем данные о собственном капитале банков России за 2018 г. из файла «Данные для работы 2».

Построим интервальный статистический ряд стандартным способом:

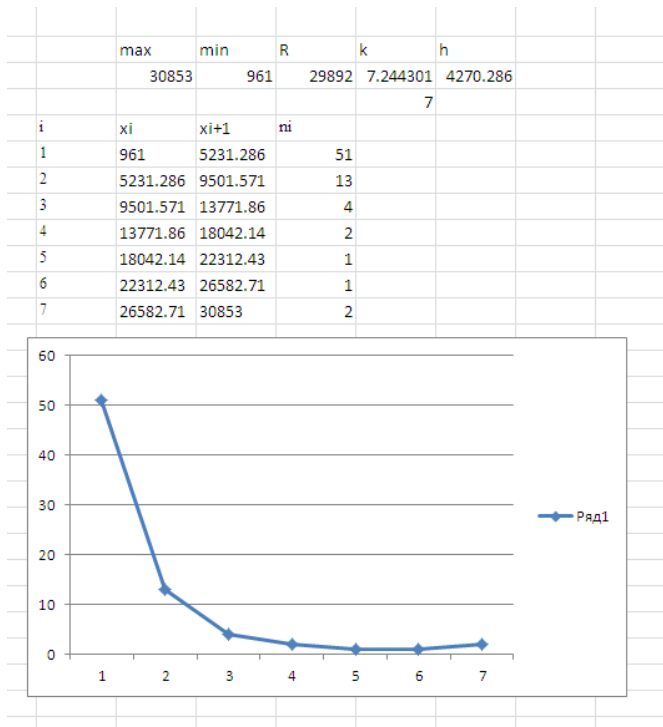
	max	min	R	k	h
	209933	961	208972	7.244301	29853.14
					7
i	x _i	x _{i+1}	n _i		
1	961	30814.14	73		
2	30814.14	60667.29	1		
3	60667.29	90520.43	1		
4	90520.43	120373.6	0		
5	120373.6	150226.7	0		
6	150226.7	180079.9	0		

Из распределения частот видны интервалы с нулевыми частотами, это произошло из-за аномальных выбросов в выборке. Убедимся в это, построив точечную диаграмму.



Мы видим два таких наблюдения на диаграмме, для достоверного анализа данных исключим эти наблюдения из выборки и продумаем алгоритм разбиения на интервалы еще раз.

Построим гистограмму:



Предварительный визуальный анализ данных показывает, что распределение значений выборки похоже на показательное. Начнем с проверки гипотезы о показательном распределении.

Найдем теоретические частоты при выдвинутой гипотезе о показательном законе распределения для i -го интервала по формуле:

$$n_i^* = n \cdot (F(x_i) - F(x_{i-1})) = n \cdot \left(e^{-x_{i-1} \cdot \lambda} - e^{-x_i \cdot \lambda} \right)$$

, где n — объем выборки; x_{i-1} , x_i — границы i -го интервала; теоретический параметр показательного распределения оценивается

$$\lambda \approx \frac{1}{x_6}$$

по выборке (x_6);

i	xi-1	xi	ni	$e^{-(xi-1*\lambda)}$	$e^{-(xi*\lambda)}$	ni*
1	961	5231.286	51	0.840559357	0.388492	34.35715
2	5231.286	9501.571	13	0.388491537	0.179554	15.87926
3	9501.571	13771.86	4	0.179553857	0.082987	7.339112
4	13771.86	18042.14	2	0.082986589	0.038355	3.392007
5	18042.14	22312.43	1	0.038354921	0.017727	1.567725
6	22312.43	26582.71	1	0.01772696	0.008193	0.724575
7	26582.71	30853	2	0.008193085	0.003787	0.334885

Затем вычислим наблюдаемое значение критерия Пирсона $\chi^2_{наб} = \sum_{i=1}^k \frac{(n_i - n_i^*)^2}{n_i^*}$:

i	xi-1	xi	ni	$e^{-(xi-1*\lambda)}$	$e^{-(xi*\lambda)}$	ni*	XI
1	961	5231.286	51	0.840559357	0.388492	34.35715	8.061911
2	5231.286	9501.571	13	0.388491537	0.179554	15.87926	0.522075
3	9501.571	13771.86	4	0.179553857	0.082987	7.339112	1.519213
4	13771.86	18042.14	2	0.082986589	0.038355	3.392007	0.57125
5	18042.14	22312.43	1	0.038354921	0.017727	1.567725	0.205592
6	22312.43	26582.71	1	0.01772696	0.008193	0.724575	0.104695
7	26582.71	30853	2	0.008193085	0.003787	0.334885	8.279269
						X2 набл	19.264

Найдем табличное (критическое) значение критерия Пирсона, используя функцию Х2.ОБР.ПХ. В данном случае оно зависит от уровня значимости 0,05 и числа степеней свободы $n-m-1$, где $n=76$ – объем выборки, $m=1$ – число оцениваемых параметров (в показательном распределении – один параметр $\lambda \approx \frac{1}{x_g}$).

X2 набл	19.264
X2 критич	=ХИ2.ОБР.ПХ(0.05;74)

Аргументы функции

ХИ2.ОБР.ПХ

Вероятность 0.05 = 0.05

Степени_свободы 74 = 74

= 95.08146667


Возвращает значение обратное к правосторонней вероятности распределения хи-квадрат.

Степени_свободы число степеней свободы - число от 1 до 10¹⁰, исключая 10¹⁰.

Таким образом, $\chi^2_{критич} = 95,08$. Поскольку табличное значение больше наблюдаемого, то нулевая гипотеза о показательном распределении принимается, т.е. отклонения экспериментальных частот от теоретических являются несущественными.

Гипотеза о показательном распределении подтвердилась. Проверим, однако, гипотезы о других видах распределения. Начнем с нормального, определим соответствующие

теоретические частоты по формуле: $n_i^* = n \cdot (F(x_i) - F(x_{i-1})) = n \cdot \left(\Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right) \right)$, где n — объем выборки; x_{i-1} , x_i - границы i -го интервала; теоретические параметры нормального распределения a и σ^2 оцениваются по выборке ($a \approx \bar{x}_e$, $\sigma \approx \sqrt{S^2}$; значение $\Phi(t)$ вычисляется с помощью функции НОРМРАСП()).

i	x_{i-1}	x_i	n_i	$\Phi\left(\frac{x_{i-1} - a}{\sigma}\right)$	$\Phi\left(\frac{x_i - a}{\sigma}\right)$	n_i^*	XI
1	961	5231.286	51	0.219891217	0.479675	19.74355	49.48277
2	5231.286	9501.571	13	0.479674779	0.748769	20.45116	2.71475
3	9501.571	13771.86	4	0.748768997	0.918071	12.86694	6.110439
4	13771.86	18042.14	2	0.918070867	0.982734	4.914393	1.728329
5	18042.14	22312.43	1	0.982733936	0.997712	1.138298	0.016802
6	22312.43	26582.71	1	0.997711536	0.999812	0.159668	4.422661
7	26582.71	30853	2	0.999812432	0.999991	0.01354	291.425
							
				X2 набл	355.9008		
				X2 критич	95.08147		

Поскольку табличное значение оказалось меньше наблюдаемого (что вполне ожидаемо), то в этом случае нулевая гипотеза отклоняется в пользу конкурирующей.

Теоретическая частота при выдвинутой гипотезе о равномерном законе распределения для i -го интервала определяется по формуле:

$$n_i^* = n \cdot (F(x_i) - F(x_{i-1})) = n \cdot \frac{x_i - x_{i-1}}{x_{max} - x_{min}}$$

i	x_{i-1}	x_i	n_i	n_i^*	XI
1	961	5231.286	51	10.85714286	148.4229
2	5231.286	9501.571	13	10.85714286	0.422932
3	9501.571	13771.86	4	10.85714286	4.330827
4	13771.86	18042.14	2	10.85714286	7.225564
5	18042.14	22312.43	1	10.85714286	8.949248
6	22312.43	26582.71	1	10.85714286	8.949248
7	26582.71	30853	2	10.85714286	7.225564
				X2 набл	185.5263
				X2 критич	95.08147

В данном случае мы также видим, что табличное значение оказалось меньше наблюдаемого, поэтому нулевая гипотеза о равномерном распределении отклоняется.

Рассмотрим, как критерий проверки гипотезы о виде распределения реализован в R.

Большая часть инструментов статистики работает в предположении о том, что выборка получена из нормальной совокупности. Основные классические критерии проверки на нормальность собраны в пакете nortest. Пакет можно установить с CRAN при помощи вызова функции `install.packages()`. Подключить установленный пакет можно при помощи функции `library()`:

```
> library(nortest).
```

Функция `shapiro.test(x)` выполняет тест Шапиро–Уилка. Нулевая гипотеза заключается в том, что случайная величина, выборка x которой известна,

распределена по нормальному закону. Объем выборки должен быть не меньше 3 и не больше 5000.

Объект, возвращаемый функцией `shapiro.test`, это список со следующими полями:

- `statistics` - значение статистики Шапиро–Уилка,
- `p.value` - вероятность того, что гипотеза будет принята,
- `method` - строка "Shapiro-Wilk normality test"
- `data.name` - строка, содержащее имя данных, подвергнутых тесту.

Для проверки гипотезы о нормальном распределении вызовем данную функцию:

```
> shapiro.test(x)
```

```
Shapiro-Wilk normality test
```

```
data: x
```

```
W = 0.67664, p-value = 1.757e-11
```

Проанализируем результаты теста: $p\text{-value}=1,757 \cdot 10^{-11} < 0.05$, поэтому гипотеза о нормальном распределении отклоняется.

Тема 7. Парный коэффициент корреляции

Найти в папке «Данные для работы в Excel» данные вашей группы (файл Word). Рассчитать парный коэффициент корреляции. Проверить гипотезу о том, что парный коэффициент корреляции равен нулю. Для выполнения задания использовать программу Excel.

Вариант № 1. В следующей выборке представлены данные по цене X некоторого товара и количеству (Y) данного товара, приобретаемому домохозяйством ежемесячно в течение года.

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
X	10	20	15	25	30	35	40	35	25	40	45	40
Y	110	75	100	80	60	55	40	80	60	30	40	30

Значение X для построения прогноза: 38.

Вариант № 2. Имеются следующие данные об уровне механизации работ $X(\%)$ и производительности труда $Y(\text{т/ч})$ для 14 однотипных предприятий:

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Значение X для построения прогноза: 58.

Вариант №3. В следующей выборке представлены данные по цене (X) некоторого товара и количеству (Y) данного товара, приобретаемому домохозяйством ежемесячно в течение года.

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
X	22	32	27	37	42	47	52	47	37	52	57	52
Y	118	83	108	88	68	63	48	88	68	38	48	38

Значение X для построения прогноза: 44.

Вариант № 4. Имеются следующие данные об уровне механизации работ X (%) и производительности труда Y(т/ч) для 14 однотипных предприятий:

X	38	36	42	46	47	53	62	60	66	61	67	73	75	82
Y	22	26	30	32	33	35	36	39	40	42	43	45	47	50

Значение X для построения прогноза: 62.

Вариант № 5. В следующей таблице приведены статистические данные по располагаемому доходу домохозяйств (X) и затратам домохозяйств на розничные покупки(Y) за 15 лет:

X	9,098	9,137	9,095	9,280	9,230	9,348	9,525	9,755
Y	5,490	5,540	5,305	5,505	5,420	5,320	5,540	5,690
X	10,280	10,665	11,020	11,305	11,430	11,450	11,697	
Y	5,870	6,157	6,342	5,905	6,125	6,185	6,225	

Значение X для построения прогноза: 9,852.

Вариант № 6 Известны данные в (у.е.) по доходам (X) и расходам (Y) на непродовольственные товары 20 домохозяйств:

X	26,2	33,1	42,5	47,0	48,5	49,0	49,1	50,9	52,4	53,2
Y	10,0	11,2	15,0	20,5	21,2	19,5	23,0	19,0	19,5	18,0
X	54,0	54,8	59,0	61,3	62,5	63,1	64,0	66,2	70,0	71,5
Y	24,5	21,5	35,4	25,0	17,3	21,6	15,3	32,6	34,0	23,8

Значение X для построения прогноза: 38.

Тема 8. Вычисление оценок коэффициентов регрессии. Статистические свойства коэффициентов регрессии

Если сразу по величине стандартной ошибки трудно судить о точности оценки, то необходимо протестировать коэффициент функции регрессии, т.е. проверить гипотезу о том, что этот коэффициент равен нулю. Особенно важно протестировать коэффициент при

объясняющей переменной, поскольку если он незначимо отличается от нуля, то зависимости нет.

Итак, мы собираемся проверить гипотезу

$$H_0: \beta_1 = 0.$$

Для проверки этой гипотезы используется следующая критическая статистика

$$t = \frac{b_1}{s.e.(b_1)}.$$

Если $|t_e| > t_{кр}$, то гипотеза отклоняется, точность оценки хорошая, зависимость есть.

Если $|t_e| \leq t_{кр}$, то гипотеза не отклоняется, точность оценки плохая, зависимости нет.

Проверка гипотезы для коэффициента β_0 выполняется точно также.

Приведем примеры тестирования коэффициентов регрессии.

Вернемся к модели зависимости прироста GDP от прироста занятости. Мы видели, что коэффициенты этой регрессии нуждаются в тестировании. Выборочные значения t-статистики приведены в таблице результатов справа от стандартных ошибок. По таблице находим критическое значение t-статистики для 23 степеней свободы

$$t_{кр}(0,01;23) = 2,807.$$

Мы видим, что оба выборочных значения статистики больше критического $6,67 > 2,807$; $5,75 > 2,807$.

Эти неравенства показывают, что оба коэффициента регрессии значимо отличаются от нуля при уровне значимости 1%. Таким образом, оценки коэффициентов точны, и зависимость есть.

Для проверки гипотезы можно было поступить и по-другому. В последнем столбце таблицы приведены вероятности получения значений t-статистики превосходящих выборочное значение.

Если вероятность выборочного значения t-статистики меньше выбранного уровня значимости, то гипотеза H_0 отклоняется, точность оценки хорошая, зависимость есть.

Если вероятность выборочного значения t-статистики больше выбранного уровня значимости, то гипотеза H_0 не отклоняется, точность оценки плохая, зависимости нет.

В рассмотренном примере вероятности получения значений t-статистики, превосходящих выборочное значение, менее 0,0001, т.е. меньше любого общепринятого уровня значимости. Следовательно, нулевая гипотеза отклоняется.

Рассмотрим другой пример. Пусть по 30 наблюдениям оценена регрессия

$$y = 34,56 - 2,53x.$$

(41,45) (1,67)

Под оценками коэффициентов подписаны стандартные ошибки. Мы видим, что стандартная ошибка свободного члена больше его оценки, следовательно, точность этой оценки плохая. Для коэффициента наклона выполняются неравенства

$$0,1 \cdot 2,53 < 1,67 < 2,53.$$

Следовательно, необходим тест. Выборочное значение t-статистики

$$t = \frac{2,53}{1,67} = 1,515.$$

Критическое значение t-статистики

$$t_{кр}(0,05;28) = 2,042.$$

Мы видим, что модуль выборочного значения меньше критического, гипотеза не отклоняется, точность оценки плохая, зависимости нет.

Используя t-статистику можно построить доверительный интервал для истинного значения коэффициента β_1 . Истинное значение и его оценка совместимы, если

$$-t_{кр} < \frac{b_1 - \beta_1}{s.e.(b_1)} < t_{кр}.$$

Выразив из этого двойного неравенства β_1 , получим

$$b_1 - t_{кр} s.e.(b_1) < \beta_1 < b_1 + t_{кр} s.e.(b_1).$$

Доверительный интервал для коэффициента β_1 .

Вернемся к модели зависимости прироста GDP от прироста занятости. В этой модели получаем следующий 99% доверительный интервал для коэффициента наклона

$$0,618 < \beta_1 < 1,792.$$

Мы видим, что в этот интервал не попало нулевое значение, т.е. с вероятностью 99% можно утверждать, что коэффициент наклона отличен от нуля, и зависимость существует.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	x	y	xy	x^2	y^2																
2	35	29	1015	1225	841																
3	35	31	1085	1225	961																
4	33	35	1155	1089	1225																
5	34	35	1190	1156	1225																
6	38	45	1710	1444	2025																
7	40	46	1840	1600	2116																
8	40	45	1800	1600	2025																
9	39	44	1716	1521	1936																
10	37	38	1406	1369	1444																
11	36	37	1332	1296	1369																
12																					
13			1424,9	1352,5	1516,7																
14																					
15	x ср	y ср	cov(x,y)	var x	var y	r(x,y)	-1				0										1
16	36,7	38,5	11,95	5,61	34,45	0,859591						0,85									
17													Необходим тест								
18	t выб	t крит																			
19	4,758068	2,306004		$\alpha=0,05$																	
20																					
21	t выб	>	t крит	Существует положительная зависимость																	
22																					
23																					
24	1. Взять любые данные из файла (последнее значение x брать не нужно).						b1	2,130125	$y^A = -39,67 + 2,13x$	Уравнение регрессии											
25	2. Посчитать: x ср, y ср, cov(x,y), var x, var y, r(x,y)						b0	-39,6756													
26	3. Проанализировать полученные данные и если будет необходимость провести тест.																				
27	4. Сделать вывод о существовании зависимости и ее характере.						Стоимость квартиры будет увеличиваться на 2,13 тыс. долларов при увеличении площади на 1 кв.м.														
28							b0 не имеет экономического смысла, так как площадь квартиры не может быть нулевой.														

Перечень вопросов выносимых на дифференцированный зачёт:

1. Введение и основы синтаксиса. Списки и циклы.
2. Операции с таблицами. Условия и функции.
3. Библиотека Pandas. Анализ данных и оформление результатов.
4. Веб-оболочка Jupyter Notebook.
5. Процесс анализа данных.
6. Предобработка данных.
7. Основные принципы предобработки. Работа с пропусками.
8. Изменение типов. Поиск дубликатов. Категоризация данных.
9. Введение в исследовательский анализ. Графики и выводы.
10. Изучение срезов данных.
11. Работа с несколькими источниками данных.
12. Взаимосвязь данных.
13. Валидация данных.
14. Введение в статистический анализ.

15. Описательная статистика.
16. Математическая статистика.
17. Выборочные методы математической статистики.
18. Основы теории оценивания.
19. Основы проверки статистических гипотез.
20. Основы регрессионного анализа.
21. Статистические оценки параметров распределения.
22. Основные понятия теории вероятностей.
23. Элементы комбинаторики.
24. Теорема сложения вероятностей. Теорема умножения вероятностей.
25. Следствия теорем сложения и умножения вероятностей.
26. Дискретные случайные величины (ДСВ).
27. Закон больших чисел.
28. Функция распределения вероятностей случайной величины.
29. Плотность распределения вероятностей непрерывной случайной величины.
30. Числовые характеристики непрерывной случайной величины.
31. Сетевое планирование. Транспортная задача.

Таблица 9 – Примеры оценочных средств с ключами правильных ответов

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
Код и наименование проверяемой компетенции				
...				
1.	Задание закрытого типа	На плоскости нарисованы две концентрические окружности, радиусы которых 6 и 12 см соответственно. Какова вероятность того, что точка брошенная наудачу в большой круг, попадет в кольцо, образованное указанными окружностями? 1) 0,5; 2) 0,65; 3) 0,12; 4) 0,75; 5) 0,60.	4	2
2.		На сборку попадают детали с двух автоматов: 80 % из первого и 20 % из второго. Первый автомат дает 10 % брака, второй – 5 % брака. Найти вероятность попадания на сборку доброкачественной детали. 1) 0,90; 2) 0,09; 3) 0,91; 4) 0,85; 5) 0,15	3	3
3.		Случайная величина распределена по нормальному закону, причем $M(X) = 15$. Найти $P(10 < X < 15)$, если известно, что $P(15 < X < 20) = 0,25$. 1) 0,10;	4	3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		2) 0,15; 3) 0,20; 4) 0,25; 5) 0,30.		
4.		По выборке объема $n = 51$ найдена смещенная оценка $Dv=3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности. 1) 3,05; 2) 3,06; 3) 3,51; 4) 3,60; 5) 0.	2	3
5.		Как называется число m_0 (наступления события в n независимых испытаниях, в каждом из которых вероятность появления события равна p), определяемое из неравенства: $np - q \leq m_0 \leq np + p$ 1) наибольшее; 2) оптимальное; 3) наимвероятнейшее; 4) невозможное; 5) минимальное	3	3
6.	Задание открытого типа	На пяти одинаковых карточках написаны буквы И, Л, О, С, Ч. Если перемешать их, и разложить наудачу в ряд две карточки, То вероятность p получить слово ИЛ равна В ответе запишите число $1/p$.	20	5
7.		Данное предприятие в среднем выпускает 20 % продукции высшего сорта и 70 % продукции первого сорта. Найти вероятность P того, что случайно взятое изделие этого предприятия будет высшего или первого сорта. В ответ записать число $30 P$.	27	6
8.		Случайная величина X распределена равномерно на интервале $(2; 6)$ и $p(x)$ – ее плотность вероятности. Найти $p(3)$. В ответ записать число $40 p(3)$.	10	8
9.		Студентам нужно сдать 4 экзамена за 6 дней. Сколькими способами можно составить расписание сдачи экзаменов?	360	8
10.		Вероятность появления события A в	64	8

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		каждом из 100 независимых испытаний равна 0,4. Найдите математическое ожидание и дисперсию случайной величины X – числа появлений события A . В ответ запишите их сумму.		

Полный комплект оценочных материалов по дисциплине (модулю) (фонд оценочных средств) хранится в электронном виде на кафедре, утверждающей рабочую программу дисциплины (модуля), и в Центре мониторинга и аудита качества обучения.

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

По дисциплине «Вероятностно-статистические методы в анализе данных» итоговой формой отчетности в 3 семестре является дифференцированный зачет. Согласно действующей в АГУ им. В.Н. Татищева в системе оценивания БАРС для дисциплин, итоговой формой отчетности для которых является зачет, отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины и распределяются по возможности равномерно по всему семестру.

Система накопления баллов, а также система штрафов, представлена в технологической карте дисциплины «Вероятностно-статистические методы в анализе данных».

Отчет по лабораторной работе

Отчет по лабораторной работе представляется в электронном виде. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя. В случае, если оформление отчета и поведение студента во время защиты соответствуют указанным требованиям, студент получает максимальное количество баллов.

Основаниями для снижения количества баллов в диапазоне от \max до \min являются:

- небрежное выполнение;
- отсутствие выводов;
- нарушение сроков предоставления отчета.

Отчет не может быть принят и подлежит доработке в случае:

- отсутствия необходимых разделов;
- отсутствия необходимого графического материала;
- неверных результатов расчета.

Контрольные работы

Контрольная работа состоит из 2-х заданий.

Основаниями для снижения оценки за задание являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ для теоретических заданий;
- небрежное выполнение;
- многократное переписывание контрольной работы.

Задание не может быть засчитано, если:

- даны два неверных ответа на теоретические вопросы.

В соответствии с балльно-рейтинговой системой БАРС по дисциплине отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины. Оценивание студентов на зачете осуществляется в соответствии с требованиями и критериями 100-балльной шкалы. Учитываются как результаты текущего контроля, так и знания, навыки и умения, непосредственно показанные студентами в ходе зачета. Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных и тематических контрольных работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку. На зачете осуществляется комплексная проверка знаний, навыков и умений студентов по всему теоретическому материалу дисциплины и с проверкой практических навыков и умений по разработке документов различных видов. Теоретические знания оцениваются путем компьютерного тестирования или на основании письменных ответов студентов по нескольким теоретическим вопросам.

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	<i>Выполнение лабораторной работы</i>	6/10	60	По расписанию
2.	<i>Выполнение контрольной работы</i>	1/15	15	
3.	<i>Тест</i>	1/15	15	
Всего			90	-
Блок бонусов				
4.	<i>Посещение занятий без пропусков</i>	1	3	
5.	<i>Своевременное выполнение всех заданий</i>	1	3	
6.	<i>Активная работа студента на занятии</i>	1	4	
Всего			10	-

Таблица 11 – Система штрафов (для одного занятия)

Показатель	Балл
<i>Опоздание на занятие</i>	-1
<i>Нарушение учебной дисциплины</i>	-1
<i>Неготовность к занятию</i>	-2
<i>Пропуск занятия без уважительной причины</i>	-5

Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)

Сумма баллов	Оценка по 4-балльной шкале	Зачтено
90–100	5 (отлично)	
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	

Сумма баллов	Оценка по 4-балльной шкале	
60–64		
Ниже 60	2 (неудовлетворительно)	Не зачтено

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Основная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL:<https://urait.ru/bcode/450262>;
2. Теория вероятностей и математическая статистика. Практикум [Электронный ресурс]: учеб. пособие / И.Ю. Мацкевич, Н.П. Петрова, Л.И. Тарусина -Минск : РИПО, 2017. - URL: <http://www.studentlibrary.ru/book/ISBN9789855037119.html> (ЭБС «Консультант студента»);
3. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие/ А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/90591>;
4. Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - М. : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. – URL : <https://www.studentlibrary.ru/book/ISBN9785970604090.html>;
5. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166>.

8.2. Дополнительная литература

1. Теория вероятностей. Примеры и задачи [Электронный ресурс]: учеб. пособие /Васильчик М.Ю. - Новосибирск : Изд-во НГТУ, 2014. - URL:<http://www.studentlibrary.ru/book/ISBN9785778224872.html>
2. (ЭБС «Консультант студента»);
3. Теория вероятностей и математическая статистика [Электронный ресурс] /Балдин К. В. - М. : Дашков и К, 2014. - URL:<http://www.studentlibrary.ru/book/ISBN9785394021084.html> (ЭБС «Консультант студента»);
4. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/26444.html>;
5. Билл, Фрэнкс Укрощение больших данных : как извлекать знания из массивов информации с помощью глубокой аналитики / Фрэнкс Билл ; перевод А. Баранов. — Москва : Манн, Иванов и Фербер, 2014. — 340 с. — ISBN 978-5-00057-146-0. — Текст

:электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/39433.html>;

6. Флах П., Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Флах П. - М. : ДМК Пресс, 2015. - 400 с. - ISBN 978-5-97060-273-7 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. –
7. URL : <https://www.studentlibrary.ru/book/ISBN9785970602737.html>.

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». www.studentlibrary.ru;
2. Электронно-библиотечная система IPR BOOKS <http://www.iprbookshop.ru/>.

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лабораторных занятий необходима компьютерная аудитория, в которой организован доступ к сети Интернет и установлено программное обеспечение. Для проведения публичной защиты проектов, необходима мультимедийная аудитория с проектором. Во время лекций используется ноутбук и проектор для презентаций.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).