МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет управления

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Теория вероятностей и математическая статистика
(наименование дисциплины)
Кафедра бизнес информатики и высшей математики
факультета управления

Образовательная программа
38.03.05 Бизнес-информатика

Направленность (профиль) программы:
Корпоративные информационные системы

Форма обучения:
очная

Статус дисциплины:
Обязательная, математический модуль

Махачкала, 2022

Рабочая программа дисциплины **Теория вероятностей и математическая статистика** составлена в 2022 году в соответствии с требованиями ФГОС ВО – бакалавриат по направлению подготовки 38.03.05 Бизнес-информатика от «29» июля 2020г. №838.

Разработчик(и): кафедра бизнес-информатики и высшей математики ст. преподаватель Иванова Елена Владимировна.

Рабочая программа дисциплины одобрена:
на заседании кафедры БИиВМ от «16» 03 2022г., протокол № 7

Зав. кафедрой ________________ профессор Омарова Н.О.
(подпись) (Ф.И.О)

на заседании учебно-методической комиссии факультета управления от «16» 03 2022г., протокол № 6.

Председатель ________________ Гашимова Л.Г.
(подпись) (Ф.И.О)

Рабочая программа дисциплины согласована с учебно-методическим управлением «31» 03 2022г.

Начальник УМУ ________________ Гасангаджиева А.Г.
(подпись)

Аннотация рабочей программы дисциплины

Дисциплина Теория вероятностей и математическая статистика входит в обязательную часть ОПОП бакалавриата по направлению подготовки 38.03.05 Бизнес-информатика .

Дисциплина реализуется на факультете управления кафедрой бизнес-информатики и высшей математики.

Содержание дисциплины охватывает круг вопросов, связанных с получением базовых знаний и формированием основных навыков по теории вероятностей и математической статистике, необходимых для решения задач, возникающих в практическойдеятельности; развитием понятийной теоретико-вероятностной базы и формированием уровня алгебраической подготовки, необходимых для понимания основ экономической статистики и её применения.

Дисциплина нацелена на формирование следующих компетенций выпускника: универсальных - _УК-1__, общепрофессиональных – ОПК-1

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: *лекции, семинарские занятия, самостоятельная.*

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости в форме устного и письменного опросов ипромежуточный контроль в формезачета.

Объем дисциплины 108 ч. 3зачетных единицы, в том числе в академических часах по видам учебных занятий:

Очная форма обучения

Семестр	всего	Учебные занятия в том числе: Контактная работа обучающихся с преподавателем						СРС, в том числе зачет, дифференцированный зачет, экзамен	Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)
		всего	из них						
			Лекции	Лабораторные занятия	Практические занятия	…	..		
3	108	108	16	24	16			50	экзамен

1. Цели освоения дисциплины

Целями освоения дисциплины . Теория вероятностей и математическая статистика являются

В области воспитания целью является: развитие у студентов социально-личностных качеств, способствующих их творческой активности, общекультурному росту, социальной мобильности, целеустремленности, организованности, трудолюбия, ответственности, самостоятельности, гражданственности, приверженности этическим ценностям, коммуникативности, толерантности, настойчивости в достижении цели.

В области обучения целями являются: подготовка в области основ ИКТ, получение знаний, позволяющих проводить ориентированные на производство разработки и научные исследования, оформлять результаты научных исследований в виде публикаций в научных изданиях, излагать результаты в виде презентаций перед различными аудиториями.

В области обучения целью является формирование : универсальных - _УК-1__, общепрофессиональных – ОПК-1, компетенций, позволяющих выпускнику успешно работать в сфере информационных систем и технологий и быть устойчивым на рынке труда.

2.Место дисциплины в структуре ОПОП бакалавриата

Дисциплина Теория вероятностей и математическая статистикавходит в Обязательную часть базовый модуль ОПОП бакалавриата по направлению подготовки 38.03.05 Бизнес-информатика .

- перечень дисциплин (или их разделов), необходимых для изучения данной дисциплин: школьные курсы математики и информатики ;
- перечень дисциплин, использующих результаты изучения данной дисциплины:Системы искусственного интеллекта; вычислительные системы, сети и телекоммуникации; Архитектура предприятия;Управление жизненным циклом информационных систем; электронный бизнес…

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения и процедура освоения).

Код и наименование компетенции из ОПОП	**Код и наименование индикатора достижения компетенций**	**Результаты обучения**	**Процедура освоения**
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.И-1. Осуществляет поиск необходимой информации, опираясь на результаты анализа поставленной задачи. УК-1.И-2. Разрабатывает варианты решения проблемной ситуации на основе критического анализа	**Знает** принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач **Умеет** анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности **Владеет**	Устный, письменный опросы, защиты лабораторных работ. Экзамен

	доступных источников информации. УК-1.И-3. Выбирает оптимальный вариант решения задачи, аргументируя свой выбор.	навыками научного поиска и практической работы с информационными источниками; методами принятия решений; строит логические умозаключения на основе поступающих информации и данных для решения управленческих задач; выдвигает и обосновывает новые идеи, обосновывает альтернативы решения задач с учетом организационных, ресурсных, финансовых и иных ограничений и т.п	
ОПК-1. Способен проводить моделирование, анализ и совершенствование бизнес-процессов и информационно-технологической инфраструктуры предприятия в интересах достижения его стратегических целей с использованием современных методов и программного инструментария.	ОПК-1.И-1. Выявляет возможности для достижения предприятием своих стратегических целей за счет использования информационных систем и информационных технологий.. ОПК-1.И-2. Совершенствует процессы организации за счет использования информационных систем и информационных технологий. … ОПК-1.И-3. Применяет инструментальные средства для моделирования текущего и целевого состояний архитектуры предприятия.…	**Знает** методы интегрированного представления целей предприятия, процессов, информационных систем и ИТ-инфраструктуры в рамках архитектурного подхода; основные понятия и методы работы с вычислительным оборудованием, системами хранения данных, центрами обработки данных, с сетями передачи данных. **Умеет** выявлять и реализовывать возможности для совершенствования предприятия за счет использования информационных систем и информационных технологий; совершенствовать процессы организации за счет использования информационных систем и информационных технологий;	Устный, письменный опросы, защиты лабораторных работ. Экзамен

Владеет
способами применения облачных вычислений в области инфраструктурных решений;
навыками моделирования текущего и целевого состояние архитектуры предприятия с использованием

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет __3___ зачетных единиц, __108 академических часов.

4.2. Структура дисциплины.

4.2.1. Структура дисциплины в очной форме

№ п/п	***Разделы и темы дисциплины***	***Семестр***	***Неделя семестра***	***Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часа)***				***Самостоятельная работа***	***Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)***
				Лекции	Практические занятия	Лабораторные занятия	Контроль самостоятельной		
	Модуль1. Раздел 1. *Теория вероятностей*								
1	*Случайные события и их вероятностиЭлементы комбинаторики*	3		2	2			5	Устный, письменный опросы, защиты лабораторных работ.
2	*Теоремы сложения и умножения вероятностей.*	3		2	2	4		5	Устный, письменный опросы, защиты лабораторных работ.
3	*Повторение испытаний.*	3				4		5	Устный, письменный опросы, защиты лабораторных работ.

4	*Понятие и закон распределения СВ. Функция распределения случайной величины. Основные числовые характеристики ДСВ*	3		2	2	4		3	Устный и письменный опросы. Проверка домашнего задания
5	*Непрерывные случайные величины (НСВ)*	3		2	2			3	Устный и письменный опросы. Проверка домашнего задания
	Итого за 1 модуль:			**8**	**8**	**12**		**8**	
	Модуль 2. Раздел 3. *Математическая статистика**.***								
6	*Выборка и ее представление. Полигон , гистограмма*	3		2	2	4		5	Устный и письменный опросы. Проверка домашнего задания
7	*Точечные и интервальные оценки статистического ряда..*	3		2	2	4		5	Устный и письменный опросы. Проверка домашнего задания
8	*Статистическая гипотеза и статистический критерий. Проверка статистических гипотез.*	3		2	2	4			Устный и письменный опросы. Проверка домашнего задания
9	*Корреляционный и регрессионный анализы*	3		2	2				Устный и письменный опросы. Проверка домашнего задания
	Итого за 2 модуль:			**8**	**8**	**12**		**8**	
	Модуль 3: Подготовка к экзамену							**36**	
	Итого			16	16	12		52	экзамен

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине.

Модуль1. Раздел 1. *Теория вероятностей.*

Тема 1. Случайные события и их вероятности Элементы комбинаторики.

Случайные события, частота и вероятность. Классический способ подсчетавероятностей. Геометрические вероятности. Пространство элементарных событий.Случайное событие как подмножество в пространстве элементарных событий. Алгебрасобытий. Аксиомы вероятности и вероятностное пространство. Следствия из аксиом.Статистическое определение вероятности.

Основные понятия кормбинаторики: Комбинаторное правило умножения..Перестановки, сочетания из n по k, размещения из n по k, сочетания с повторениями..Бином Ньютона и свойства биномиальных коэффициентов.

Тема 2. Теоремы сложения и умножения вероятностей.

Основные формулы для вычисления вероятностей. Условные вероятности.

Теоремаумножения вероятностей. Формула полной вероятности и формула Байесавероятностей гипотез. Независимые события.

Тема 5. Случайные величины

Закон и функция распределения СВ.

Основные числовые характеристики ДСВ: математическое ожидание, дисперсия, среднее квадратическое отклонение.

Тема 8. Непрерывные случайные величины (НСВ).

Непрерывные и абсолютно непрерывные случайные величины. Свойства функции плотности. Математическое ожидание и дисперсия абсолютно непрерывной случайной величины. Математическое ожидание функции от абсолютно непрерывной случайной величины.Равномерное распределение на отрезке, показательное (экспоненциальное) распределение, нормальное распределения, их числовые характеристики. Мода, медиана, и квантили непрерывного распределения.

Модуль2. Математическая статистика

Тема 9.. Статистические методы обработки экспериментальных данных.

Генеральная и выборочная совокупности. Статистическое распределение выборки.

Эмпирическая функция распределения и вариационный ряд. Гистограмма. Мода и медиана.

Тема 10. Основные характеристики и показатели вариационного ряда.

Генеральные среднее, дисперсия, моменты высших порядков (симметрия, эксцесс).

Эмпирическая ковариация. Повторные и бесповторные выборки. Математическое ожидание и дисперсия выборочного среднего для повторной и бесповторной выборки.

Тема 11. Статистические оценки параметров распределения.

Несмещенность, состоятельность и эффективность точечных оценок. Оценка неизвестной вероятности по частоте. Точечные оценки для математического ожидания и дисперсии. Метод моментов. Метод максимального правдоподобия.

Тема 12. Интервальные оценки.

Доверительные вероятности и интервалы.

Приближенный доверительный интервал для оценки генерального доли признака.

Приближенный доверительный интервал для оценки генерального среднего.

Тема 13. Статистическая гипотеза и статистический критерий.

Статистическая проверка гипотез. Ошибки I и II рода. Уровень значимости и мощность критерия. Проверка гипотез о равенстве средних и дисперсий двух нормально распределенных генеральных совокупностей. Простые и сложные гипотезы. Хи-квадрат критерий Пирсона.

Тема 14 Проверка гипотез.

Проверка гипотезы о соответствии наблюдаемых значений предполагаемому распределению вероятностей (дискретному или непрерывному). Сравнение параметров двух нормальных распределений.

4.3.2. Содержание практических занятий по дисциплине.

Модуль1. Раздел 1. *Теория вероятностей.*

Занятие 1. Случайные события и их вероятности Элементы комбинаторики.

1. Случайные событие. Определения вероятностей.
2. Правила комбинаторики:Перестановки, сочетания из n по k, размещения из n по k,
3. Теорема сложения вероятностей. Теорема умножения вероятностей.

Занятие 2. Формула полной вероятности.

1. Вероятность наступления события хотя-бы одного испытании.
2. Формула полной вероятности.
3. Формула Байеса вероятностей гипотез.

Занятие 3. Случайные величины

1. Закон и функция распределения случайной величины.
2. . Математическое ожидание ДСВ.
3. Дисперсия ДСВ.
4. Среднеквадратическое отклонение.

Занятие4. Непрерывные случайные величины (НСВ).

1. Функция плотности распределения вероятностей НСВ.
2. Математическое ожидание, дисперсия НСВ.. Мода, медиана.
3. Нормальное, равномерное, показательные распределения.

Модуль2. Математическая статистика

Занятие 5. Статистические методы обработки экспериментальных данных.

1. Генеральная и выборочная совокупности.
2. Эмпирическая функция распределения и вариационный ряд.
3. Гистограмма. Полигон.

Занятие 6. Числовые и интервальные характеристики выборки .

1. Выборочная средняя.
2. Выборочная дисперсия.
3. Несмещенность, состоятельность и эффективность точечных оценок.
4. Метод моментов. Метод максимального правдоподобия.
5. Доверительные интервал математического ожидания и дисперсии.

Занятие7. Статистическая гипотеза и статистический критерий.

1. Схема проверки статистических гипотез.
2. Проверка гипотез о равенстве средних и дисперсий двух нормально распределенных генеральных совокупностей

Занятие 8.Корреляционный и регрессионный анализ

1. Коэффициент корреляции
2. Линейная регрессия

4.3.3. Содержаниелабораторных занятий по дисциплине.

Лабораторная работа №1.Свойства статистического распределения. Построение гистограмм.

Цель работы: приобретение навыков группирования и обработки первичной статистической информации в интерактивной среде Excel.

Задание. Проранжировать первичный ряд данных, определить частоты и частости нового ряда, найти абсолютную и относительные плотности распределения, графически изобразить данные в виде эмпирической функции распределения, гистограмм частот.

Теоретическая часть

Математическая статистика – наука, изучающая методы исследования закономерностей в массовых случайных явлениях и процессах по данным, полученным из конечного числа наблюдений за ними.

Построенные на основании этих методов закономерности относятся не к отдельным испытаниям, из повторения которых складывается данное массовое явление, а представляют собой утверждения об общих вероятностных характеристиках данного процесса. Такими характеристиками могут быть вероятности, плотности распределения вероятностей, математические ожидания, дисперсии и т.п.

Найденные характеристики позволяют построить *вероятностную модель* изучаемого явления. Применяя к этой модели методы теории вероятностей, исследователь может решать технико–экономические задачи, например, определять вероятность безотказной работы агрегата в течение заданного отрезка времени. Таким образом, теория вероятностей по вероятностной модели процесса предсказывает его поведение, а математическая статистика по результатам наблюдений за процессом строит его вероятностную модель. В этом состоит тесная взаимосвязь между данными науками.

Очевидно, что для обнаружения закономерностей случайного массового явления необходимо провести сбор статистических сведений, т.е. сведений, характеризующих отдельные единицы каких–либо массовых явлений. Пусть, например, мы располагаем материалом о числе дефектных изделий в изготовленной в определенных условиях партии продукции. Проблемы возникают тогда, когда на основании этой информации мы захотим сделать выводы относительно качества производства продукции, выпускаемой предприятием. Нас может интересовать вероятность производства дефектного изделия, средняя долговечность всех выпускаемых изделий и т.д. Собранный материал рассматривается лишь как некоторая пробная группа, одна из многих возможных пробных групп. Конечно, выводы, сделанные на основании этого ограниченного числа наблюдений, отражают данное массовое явление лишь приближенно. Математическая статистика указывает, как наилучшим способом использовать имеющуюся информацию для получения по возможности более точных характеристик массового явления.

Генеральная и выборочная совокупности

Для обнаружения закономерностей, описывающих исследуемое массовое явление, необходимо иметь опытные данные, полученные в результате обследования соответствующих объектов, отображающих изучаемое явление. Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений.

Определим основные понятия математической статистики.

Генеральная совокупность– все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности *N* и объем выборки *n* – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Замечание. Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была **репрезентативной** (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Пусть интересующая нас случайная величина X принимает в выборке значение x_1 – n_1 раз, x_2 – n_2 раз, …, $x_к$ – $n_к$ раз, причем $\sum_{i=1}^{k} n_k = n,$ где n – объем выборки. Тогда наблюдаемые значения случайной величины x_1, x_2,…, $x_к$ называют **вариантами**, а n_1,

$n_2,\ldots,\ n_к$ – **частотами**. Если разделить каждую частоту на объем выборки, то получим **относительные частоты** $w_i = \dfrac{n_i}{n}$. Последовательность вариант, записанных в порядке возрастания, называют **вариационным** рядом, а перечень вариант и соответствующих им частот или относительных частот – **статистическим рядом**:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Пример.

При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1.

Составим вариационный ряд: 0,1,2,3,4,5.

Статистический ряд для абсолютных и относительных частот имеет вид:

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
w_i	0,15	0,3	0,25	0,15	0,1	0,05

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае *удобнее* использовать **группированную выборку**. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал.

Составленная по этим результатам таблица называется **группированным статистическим рядом**:

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a + h)$	$(a + h, a + 2h)$	…	$(b - h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

Полигон частот. Выборочная функция распределения и гистограмма.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами $(x_1,\ n_1)$, $(x_2,\ n_2),\ldots,\ (x_k,\ n_k)$, где x_iоткладываются на оси абсцисс, а n_i– на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные (w_i) частоты, то получим **полигон относительных частот.**

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события $X<x$.

Определение. Выборочной **(эмпирической) функцией распределения** называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X<x$. Таким образом,

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число вариант, меньших x, n – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*. $F(x)$ определяет вероятность события $X<x$, а $F^*(x)$ – его относительную частоту. При достаточно больших n, как следует из теоремы Бернулли, $F^*(x)$ стремится по вероятности к $F(x)$.

Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами $F(x)$, а именно:

1) $0 \le F^*(x) \le 1$.
2) $F^*(x)$ – неубывающая функция.
3) Если x_1 – наименьшая варианта, то $F^*(x) = 0$ при $x \le x_1$;

если $x_к$– наибольшая варианта, то $F^*(x) = 1$ при $x > x_к$.

Для непрерывного признака графической иллюстрацией служит **гистограмма**, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами – отрезки длиной n_i /h (гистограмма частот) или w_i/h(гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице .

Вычисление выборочных характеристик в Excel

Вычисление частот. Для вычисления частот n_i можно использовать ***функцию ЧАСТОТА***, обращение к которой имеет вид:

=ЧАСТОТА(*массив_данных;массив_границ*),

где *массив_данных* – адреса ячеек, для которых вычисляется частота n_i; *массив_границ* – адреса ячеек, в которых размещаются упорядоченные по возрастанию значения z_j, $j = 1, 2, ..., m+1$, где m – число интервалов.

При использовании этой функции необходимо помнить:

1. Функция ЧАСТОТА вводится как формула массива, т.е. предварительно выделяется интервал ячеек, в который будут помещены вычисленные частоты (число ячеек должно быть на 1 больше числа границ), затем вводится функция ЧАСТОТА с соответствующими аргументами, потом одновременно нажимаются клавиши [Ctrl] + [Shift] + [Enter].
2. Функция ЧАСТОТА игнорирует пустые ячейки и текстовые данные.
3. Если *массив_границ* не содержит возрастающих значений границ и интервалов, то осуществляется автоматическое вычисление границ интервалов равной ширины, причем число интервалов равно корню квадратному из числа элементов *массива_данных.*

Результатом ***работы*** является массив значений, определяемый по следующему правилу: первый элемент равен числу n_0 элементов *массива_данных* меньше z_1; последний элемент равен числу n_{m+1} элементов *массива_данных* больше z_{m+1}; остальные элементы определяются как числа n_j элементов x_i*массива_данных*, удовлетворяющих условию

$$z_j < x_i \le z_{j+1},\ j = 1, 2, ..., m.$$

Другими словами, кроме m значений частот n_j, $j = 1, 2, ..., m$, соответствующих m интервалам, вычисляются частоты n_0 (число значений x_i, лежащих левее z_1) и n_{m+1} (число значений x_i, лежащих правее z_{m+1}).

$$\sum_{j=0}^{m+1=9} n_j = 55,\ \sum_{j=0}^{m+1=9} \omega_j = 1.$$ ☻

Для подсчета количества элементов выборки (т.е. объема выборки n) используется ***функция СЧЁТ***, обращение к которой имеет вид:

СЧЁТ(*массив_данных*),

где *массив_данных* – адреса ячеек или числовые константы.

Функция МАКС вычисляет максимальное значение из заданных аргументов. Обращение к ней имеет вид:

=МАКС(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые величины.

Функция МИН вычисляет минимальное значение из заданных аргументов. Обращение к ней имеет вид:

=МИН(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые величины.

Пример решения задачи в Excel

Условие. Имеются разрозненные данные по рентабельности активов банков с доходами от 50 до 100 млн. долл.:

1,51; 0,85; 1,37; 1,62; 0,80; 2,0; 1,49; 1,58; 1,75; 1,24; 1,28; 1,04; 1,98; 1,15; 1,66; 1,33; 1,73; 1,13; 1,36; 1,28.

По этим данным получить сгруппированный статистический ряд распределения, найти и построить эмпирическую функцию распределения выборки, построить гистограммы частот.

Выполнение задания.

1. В программной среде Excel заполняется столбец исходных данных (рис. 1).
2. Первоначально, начиная с ячейки A5 (рис.1), введем в столбец A 20 элементов выборки. (диапазон A5:A24).Выполняется сортировка столбца ***A*** - первичного ряда в порядке возрастания. В результате получен новый интервальный ранжированный ряд. На рис.1 приведен уже результат сортировки по возрастанию.

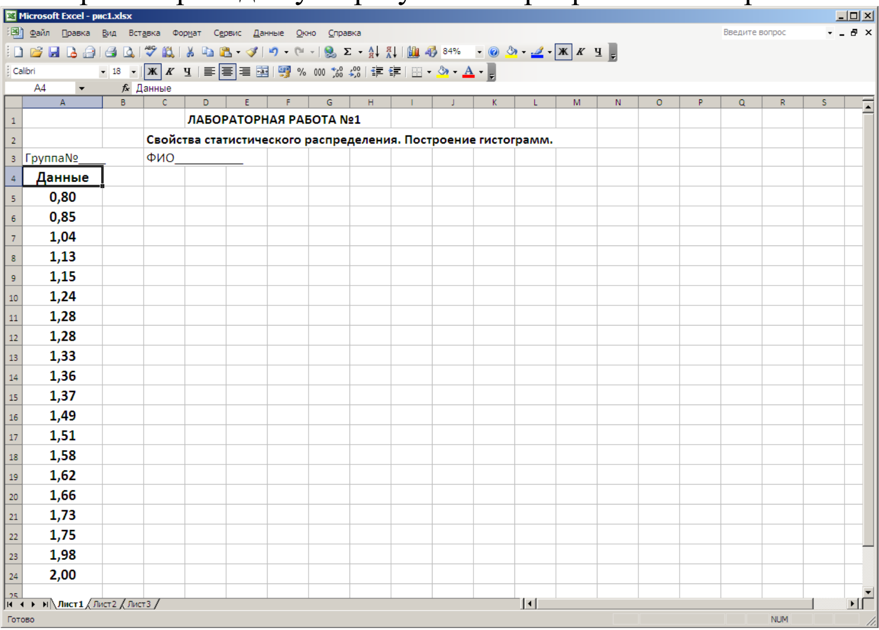


Рис. 1. Результат сортировки по возрастанию.

3.Определяются частоты нового ряда. Для этого используются данные об объеме совокупности исследуемых банков *n*. Найдем объем данной выборки n, используя встроенную функцию **СЧЁТ(*массив_данных*).**

Дискретный вариационный ряд разбивается на интервалы, число которых подсчитывается по формуле Стерджесса

$$k = \left[\ 1 + 3{,}322 \lg n\ \right], \tag{1}$$

в которой квадратные скобки означают округление числа 5,32 , тогда $k = 5$. Длина частичного интервала (интервальный шаг) определяется по формуле

$$h = \frac{x_{\max} - x_{\min}}{k} \tag{2}$$

Значения $x_{\max} = 2{,}0$, $x_{\min} = 0{,}8$ находим, используя соответствующие встроенные функции *МАКС* и *МИН*. Определяем интервальный шаг $h = 0{,}24$. Тогда границы интервалов будут такими:

$x0 = x_{\min} = 0{,}8$;

$x1 = x_{\min} + h = 1{,}04$;

$$x2 = x_{\min} + 2h = 1{,}28\ ;$$

$$x3 = x_{\min} + 3h = 1{,}52\ ;$$

$$x4 = x_{\min} + 4h = 1{,}76\ ;$$

$$x5 = x_{\min} + 5h = 2$$

Затем, начиная с ячейки B11, введем границы заданных интервалов от $x_{\min}$ до $x_{\max}$, соблюдая интервальный шаг h (см. рис. 2). Введем рядом эти же данные в диапазон C11:C16, как на рис.2.

После подготовки этих данных, для определения частот n_i, выделяем ячейки D11:D16 (число ячеек должно быть на 1 больше числа границ!), вводим выражение

=ЧАСТОТА (A5:A24;B11:B16)

и нажимаем одновременно клавиши [Ctrl] + [Shift] + [Enter]. В ячейках D11:D15 появляется результат выполнения функции (см. рис. 2.). Таблицу оформить согласно рис.2.

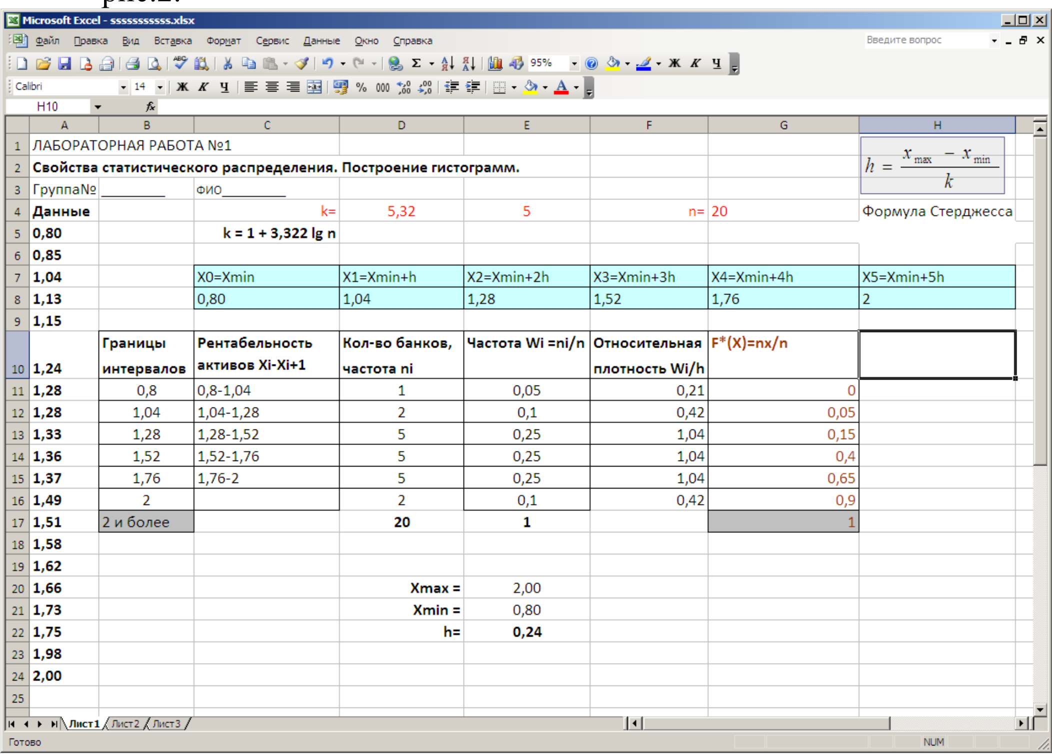
Определив количество банков, принадлежащих каждому из интервалов (частоты *ni)*, получим сгруппированный статистический ряд распределения.

Для вычисления относительных частот w_i (частостей) необходимо частоты поделить на число элементов выборки, т.е. $wi=ni/n$. Эти вычисления реализованы в ячейках E11:E15 (см. рис.2) . Здесь обратите внимание на абсолютную адресацию ячейки, содержащей значение n! Для контроля правильности вычисления частот n_i и частостей w_iв ячейках D17, E17 определить контрольные суммы по соответствующим столбцам.

3. Найдем эмпирическую функцию распределения

$$F^*(x) = \frac{n_x}{n}\ ,$$

где n_x – число вариант, меньших x, n – объем выборки. Результаты представлены на рис.2.



	A	B	C	D	E	F	G	H
1	ЛАБОРАТОРНАЯ РАБОТА №1							$h = \frac{x_{\max} - x_{\min}}{k}$
2	Свойства статистического распределения. Построение гистограмм.							
3	Группа№ ________		ФИО_________					
4	Данные		k=	5,32	5	n=	20	Формула Стерджесса
5	0,80		k = 1 + 3,322 lg n					
6	0,85							
7	1,04		X0=Xmin	X1=Xmin+h	X2=Xmin+2h	X3=Xmin+3h	X4=Xmin+4h	X5=Xmin+5h
8	1,13		0,80	1,04	1,28	1,52	1,76	2
9	1,15							
10	1,24	Границы интервалов	Рентабельность активов Xi-Xi+1	Кол-во банков, частота ni	Частота Wi =ni/n	Относительная плотность Wi/h	F*(X)=nx/n	
11	1,28	0,8	0,8-1,04	1	0,05	0,21	0	
12	1,28	1,04	1,04-1,28	2	0,1	0,42	0,05	
13	1,33	1,28	1,28-1,52	5	0,25	1,04	0,15	
14	1,36	1,52	1,52-1,76	5	0,25	1,04	0,4	
15	1,37	1,76	1,76-2	5	0,25	1,04	0,65	
16	1,49	2		2	0,1	0,42	0,9	
17	1,51	2 и более		20	1		1	
18	1,58							
19	1,62							
20	1,66			Xmax =	2,00			
21	1,73			Xmin =	0,80			
22	1,75			h=	0,24			
23	1,98							
24	2,00							
25								

Рис. 2.Примерный вид таблицы при расчете частостей и эмпирической функции распределения.

5. Построим графики эмпирической функции распределения и гистограмм частот. Для функции F*(x) предварительно выделим данные по столбцам *Границы интервалови*

$F^*(x)=\frac{n_x}{n}$. Затем, открывая вкладку меню *Вставка*, строим график, выбирая соответствующие опции меню. Названия диаграмм, осей вводим, открывая контекстное меню диаграммы.

Для гистограмм частот выделяем данные по столбцам: *Рентабельность активов* X_i-X_{i+1} и частота n_i или w_i.

Получаем следующее графическое представление данных для эмпирической функции распределения:

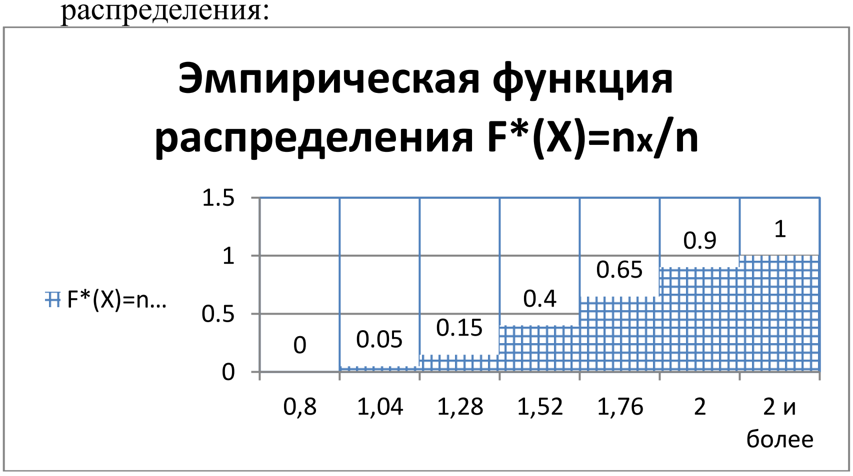


Рис.3.Эмпирическая функция распределения.

Для гистограмм частот и гистограмм относительных частот:

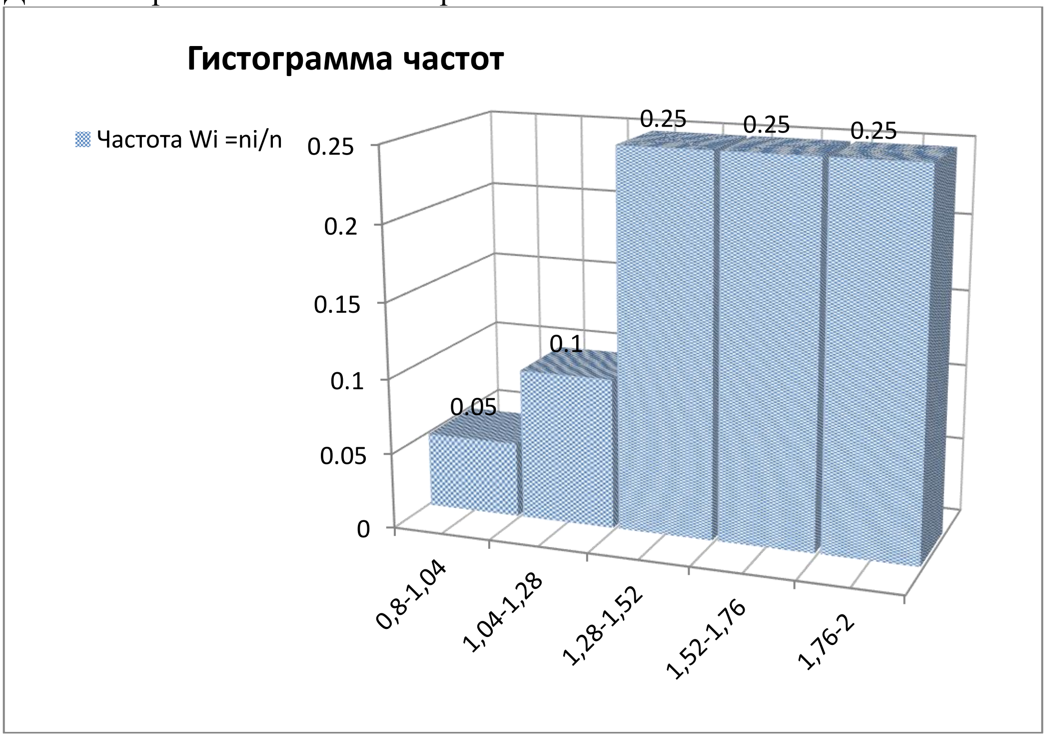


Рис.4 Графическое представление данных для гистограммы частот

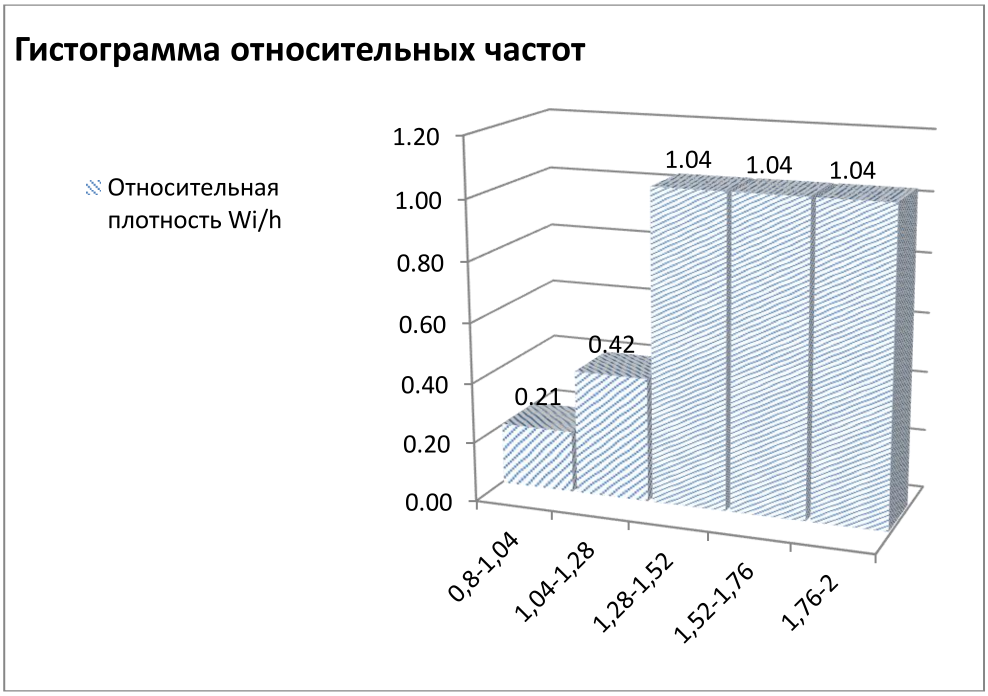


Рис.5. Графическое представление данных для гистограммы относительных частот

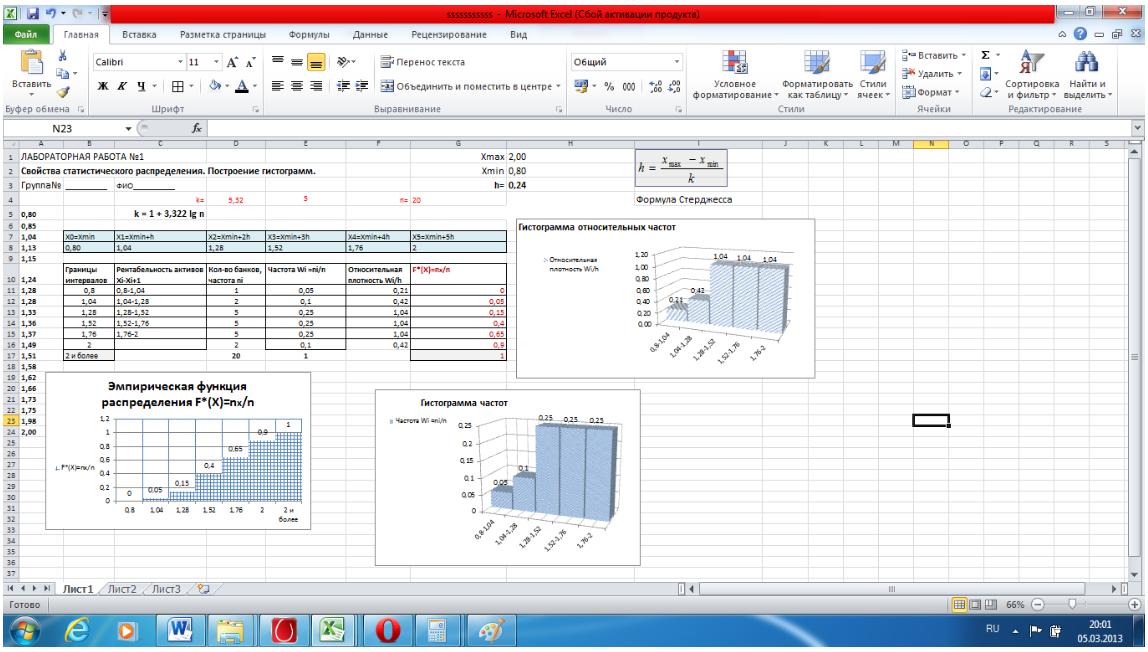


Рис.6. Результат выполнения типового примера Лабораторной работы №1.

Варианты заданий для самостоятельного решения.

Варианты указаны римскими цифрами.

Вар.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
1	0,52	0,65	0,89	1,21	1,25	1,69	1,45	1,85	0,35	0,68	0,48	1,25	1,79	1,96	0,86
2	1,89	1,63	1,06	1,69	1,85	1,52	0,42	1,64	1,05	1,78	1,51	1,22	1,11	1,54	1,88
3	1,22	0,53	1,29	0,62	1,84	0,94	1,21	1,68	1,43	1	0,49	1,22	1,95	1,05	1,22

Вар.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
4	1,43	1,45	1,06	0,41	1,98	0,78	1,78	1,28	1,21	1,43	0,65	0,74	1,01	1,43	0,63
5	0,87	1,34	1,06	0,43	0,89	1,21	1,54	1,78	0,73	1,11	0,85	0,65	1,54	1,73	1,64
6	1,55	1,68	0,69	1,75	0,95	1,54	1,66	1,37	1,25	0,74	0,53	1,01	1,55	1,51	1,01
7	1,14	1,88	1,73	0,43	1,96	0,50	1,10	1,37	1,43	1,05	0,63	1,21	0,79	1,83	1,43
8	0,65	0,99	1,56	1,48	1,43	0,65	1,12	0,89	1,12	1,24	0,25	1,01	0,74	1,78	1,33
9	0,65	1,75	0,94	0,54	0,97	1,05	1,25	1,51	1,14	1,22	0,68	1,54	1,43	1,42	1,21
10	1,89	0,59	1,75	1,67	1,99	1,25	1,48	1,88	1,64	0,78	1,11	0,75	0,95	0,63	1,69
11	1,14	2,10	0,92	1,52	1,54	2,01	1,03	1,56	0,75	0,89	0,72	1,25	0,83	1,64	0,65
12	0,91	1,87	0,89	0,65	1,05	0,94	1,66	1,11	0,63	1,92	1,43	2,14	1,37	0,89	1,54
13	1,37	1,43	0,92	1,68	1,47	1,14	0,65	2	1,07	1,64	0,98	1,01	0,64	1,78	0,89
14	1,43	1,36	1,25	1,49	1,03	1,62	1,96	1,43	1,08	0,72	0,35	0,89	0,63	1,47	1,61
15	1,78	1,37	1,45	1,37	1,21	1,78	1,62	1,22	0,74	1,23	1,78	0,67	1,25	1,63	1,37
16	0,96	0,89	1,51	0,63	1,07	0,59	1,43	1,01	1,51	1,21	0,74	0,76	1,05	1,22	1,51
17	1,25	1,65	1,65	1,76	1,42	1,45	1,51	1,23	1,11	1,01	1,21	1,51	0,89	1,92	2
18	1,11	1,21	1,78	1,25	1,65	1,29	1,81	0,91	0,65	1,37	0,86	0,32	0,99	1,88	1,65
19	0,58	1,43	1,08	1,56	2	1,21	1,11	0,91	1,88	0,63	1,05	0,63	0,69	0,74	1,51
20	1,56	1,52	1,04	1,37	1,08	1,11	1,70	1,25	1,54	1,09	0,39	0,67	1,06	1,25	0,65
21	1,09	1,34	1,23	0,89	1,84	1,44	1,37	0,63	0,85	0,78	1,89	1,54	1,56	1,75	1,11
22	2	1,37	1,42	1,66	1,54	1,11	1,67	0,63	0,68	1,25	0,45	1,05	1,15	1,32	1,56
23	1,21	1,87	1,05	0,74	1,89	1,22	1,74	1,54	1,55	0,46	1,01	0,64	2,12	1,62	0,74
24	0,99	1,25	1,67	1,66	1,84	1,98	1,42	1,21	0,89	1,43	1,64	1,11	1,24	1,84	1,05
25	1,45	1,21	0,65	1,64	1,42	1,32	1,83	0,95	1,22	1,23	0,89	0,85	1,88	0,65	1,78
26	0,89	1,47	1,58	1,57	1,37	0,89	1,06	0,97	1,24	0,67	1,05	0,81	1,25	1,74	1,37
27	2	1,14	1,02	1,58	1,65	1,64	1,43	1,25	1,21	1,54	0,81	1,43	1,13	1,71	1,64
28	1,87	1,35	1,67	1,78	1,24	1,56	1,59	1,05	0,71	1,51	1,22	1,85	1,22	1,21	1,25
29	0,62	1,11	1,79	1,52	1,06	1,42	1,04	0,74	1,78	1,65	1,54	0,93	1,51	1,01	1,35
30	1,23	1,02	1,21	1,85	1,06	1,56	0,89	1,28	1,05	0,65	0,41	0,89	0,69	1,65	1,74

Контрольные вопросы для самопроверки .

1.Генеральная и выборочная совокупности.
2. Статистическое распределение выборки.
3. Встроенные в Excel функции составления статистического ряда распределения
4. Принцип построения интервального статистического ряда при помощи формулы Стерджесса.
5.Эмпирическая функция распределения и вариационный ряд.
6. Гистограмма. Мода и медиана.

Лабораторная работа №2.

Вычисление выборочных характеристик в Excel

Цель работы:

Овладение различными способами отбора статистических данных. Нахождение точечных характеристик вариационного ряда. Овладение методами установления связи между случайными величинами. Приобретение навыка работы со встроенными функциями для расчета точечных статистических оценок в Excel.

Задание.

1.Все приведенные точечные оценки рассчитать для статистических измерений своего варианта;

2. Вычислить описательные статистики, используя пакет *Анализ данных*.

3. Сравнить значения точечных оценок, полученных с помощью пакета *Анализ данных*, со значениями аналогичных характеристик, вычисленных с помощью встроенных функций.

Теоретическая часть

Большинство случайных величин, рассмотренных в курсе теории вероятностей, имели распределения, зависящие от одного или нескольких параметров. Так, биномиальное распределение зависит от параметров p и n, нормальное – от параметров a и σ, распределение Пуассона – от параметра λ и т.п. Одной из основных задач математической статистики является оценивание этих параметров по наблюдаемым данным, т.е. по выборочной совокупности. Выборочные среднее и дисперсия интерпретируются как приближенные значения неизвестных значений математического ожидания и дисперсии изучаемой случайной величины X, т.е. являются оценками этих неизвестных характеристик.

Выборочная характеристика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется *точечной оценкой* этого параметра. В этом определении слово "точечная" означает, что значение оценки представляет собой число или точку на числовой оси.

Обозначим через θ некоторый неизвестный параметр генеральной совокупности, а через θ_n^* – точечную оценку этого параметра. Оценка θ_n^* есть функция $\varphi(X_1, X_2, ..., X_n)$ от n независимых экземпляров $X_1, X_2, ..., X_n$ генеральной совокупности, где n – объем выборки (см. п. 2.1). Поэтому оценка θ_n^*, как функция случайных величин, также является случайной, и свойства θ_n^* можно исследовать с использованием понятий теории вероятностей.

В общем случае точечная оценка θ_n^* не связана с оцениваемым параметром θ. Поэтому естественно потребовать, чтобы θ_n^* была близка к θ. Это требование формулируется в терминах несмещенности, состоятельности и эффективности.

Оценка θ_n^* параметра θ называется *несмещенной*, если для любого фиксированного объема выборки n математическое ожидание оценки равно оцениваемому параметру, т.е. $M(\theta_n^*) = \theta$.

Оценка θ_n^* называется *состоятельной*, если $\theta_n^* \xrightarrow{p} \theta$, т.е. для любого $\varepsilon > 0$ при $n \to \infty$ $P\left(\left|\theta_n^* - \theta\right| < \varepsilon \right) \to 1$. Несмещенная оценка θ_n^* параметра θ называется *несмещенной эффективной*, если она среди всех других несмещенных оценок того же параметра обладает *наименьшей дисперсией*. Как же выяснить, является ли несмещенная оценка эффективной? Очевидно, для этого необходимо сравнить дисперсию этой оценки с минимальной дисперсией.

Генеральной средней $\overline{X}r$ дискретной случайной величины называют среднее арифметическое всей генеральной совокупности.

$$\bar{x}_r = \frac{1}{n}\sum_{i=1}^{n} x_n$$

Если генеральную совокупность образует непрерывная случайная величина, то генеральная средняя определяется как ее математическое ожидание *a*: $\overline{X}r$ =M(X)= *a*

Для изучения генеральной совокупности обычно извлекается выборка объема n. Анализируя эту выборку, можно сформировать некоторое представление о свойствах генеральной совокупности, например, о числовых характеристиках ее закона распределения.

Выборочной средней $\overline{X}в$ называют среднее арифметическое значений элементов выборки.

Если все значения $x_1, x_2, \ldots, x_k$ элементов выборки различны, то

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i \qquad (1)$$

Если же значения $x_1, x_2, \ldots, x_k$ элементов выборки имеют частоты $n_1, n_2, \ldots, n_k$, причем

$\sum_{i=1}^{k} n_i = n$, то $$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \qquad (2)$$

В качестве оценки математического ожидания генеральной совокупности (генерального среднего) принимается среднее арифметическое полученных элементов выборки (выборочных значений), то есть выборочную среднюю (1) или (2). Таким образом, в общем случае M(X) = *a*.

Теорема. Выборочное среднее $\overline{X}в$ есть *состоятельная и несмещенная* оценка генеральной средней $\overline{X}_r$.

Теорема. Выборочное среднее $\overline{X}_в$ является *эффективной несмещенной* оценкой для $\overline{X}_r$.

Дисперсией случайной величины *x* называют математическое ожидание квадрата отклонения случайной величины от её математического ожидания:

$$D(x) = M[x - M(x)]^2.$$

Дисперсию удобно вычислять по формуле:

$$D(x)=M(x^2) - [M(x)]^2.$$

Средним квадратическим отклонением случайной величины называют квадратный корень из дисперсии:

$$\sigma(x) = \sqrt{D(x)}\,.$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их значений $\overline{X}в$: $D_в = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_в\right)^2$

Если значения признака $x_1, \ldots, x_n$ имеют соответственно частоты $n_1, \ldots, n_k$, причём $n_1 + \ldots + n_k = n$, то выборочную дисперсию можно найти по формуле: $D_в = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}_в\right)^2$,

т.е. выборочная дисперсия есть среднее взвешенное квадратов отклонений с весами, равными соответствующим частотам.

Теорема. Выборочная дисперсия $D_в$ равна среднему квадратов значений признака минус квадрат общей средней:

$$D_в = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (x_в)^2$$

.

Выборочная дисперсия $D_в$ в роли случайной величины является смещённой оценкой дисперсии:

$$M(D_в) = \frac{n-1}{n} D$$

Легко «исправить» выборочную дисперсию так, чтобы ее математическое ожидание было равно ***генеральной дисперсии*** D(x). Достаточно для этого $D_в$ умножить на дробь *n/(n-1).* Сделав это, получим ***исправленную выборочную дисперсию***:

$$S^2 = \frac{n}{n-1} D_в = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2$$

Исправленная дисперсия является несмещённой оценкой генеральной дисперсии:

$$M(S^2) = M\left(\frac{n}{n-1} D_в\right) = \frac{n}{n-1} M(D_в) = \frac{n}{n-1}\cdot\frac{n-1}{n} D = D$$

Итак, в качестве несмещённой оценки генеральной дисперсии принимают исправленную дисперсию:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_в)^2$$

Теорема. Исправленная дисперсия S^2 является состоятельной и несмещенной оценкой для генеральной дисперсии D.

Вычисление выборочных характеристик в Excel

Вычисление исправленной дисперсии. Ранее было показано, что оценка

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_в)^2 \qquad (3)$$

является несмещенной точечной оценкой для дисперсии случайной величины, и такую оценку называют *исправленной дисперсией.*

Для вычисления выборочного значения этой оценки можно использовать статистическую ***функцию***Excel***ДИСП***, обращение к которой имеет вид: =ДИСП(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числа или адреса ячеек, содержащих числовые величины.

Пример 1. При изменении диаметра изделия после корректировки на производстве была получена следующая выборка (объемом $n = 55$):

20.3	15.4	17.2	19.2	23.3	18.1	21.9
15.3	16.8	13.2	20.4	16.5	19.7	20.5
14.3	20.1	16.8	14.7	20.8	19.5	15.3
19.3	17.8	16.2	15.7	22.8	21.9	12.5
10.1	21.1	18.3	14.7	14.5	18.1	18.4
13.9	19.8	18.5	20.2	23.8	16.7	20.4
19.5	17.2	19.6	17.8	21.3	17.5	19.4
17.8	13.5	17.8	11.8	18.6	19.1	

вычислить оценку (3).

*Решение.*Первоначально, введем в ячейки A3:A57 55 элементов выборки (рис. 7). Затем, используя функции КВАДРОТКЛ, ДИСП (как показано на рис. 7), вычислим оценку (3). Видно ожидаемое совпадение двух вычисленных значений. (Сделать вывод).

	A	B	C	D	E
1					
2	Выборочные значения				
3	20,3	=КВАДРОТКЛ(A3:A57)/(55-1)			
4	15,3				
5	14,3		8,760		
6	19,3				
7	10,1		8,760		
8	13,9				
9	19,5		=ДИСП(A3:A57)		
10	17,8				

Рис. 7. Фрагмент вычисления исправленной дисперсии

Функции Excel для вычисления других точечных оценок:

Для вычисления среднеквадратичных отклонений можно использовать следующие функции Excel.

Функция СТАНДОТКЛОН вычисляет

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_в)^2}\ .$$

Обращение к ней имеет вид:

=СТАНДОТКЛОН(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Функция СТАНДОТКЛОНП вычисляет

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_в)^2}\ .$$

Обращение к ней имеет вид:

=СТАНДОТКЛОНП(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Функция ЭКСЦЕСС вычисляет оценку

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}_в}{d_в}\right)^2-\frac{3(n-1)^2}{(n-2)(n-3)}$$

для характеристики *эксцесс* $\frac{\mu_4}{\sigma^4}-3$, которая определяет островершинность или плосковершинность плотности распределения.

Обращение к функции имеет вид:

=ЭКСЦЕСС(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Функция МОДА вычисляет наиболее часто встречающееся значение в заданных аргументах функции, т.е. значение, встречающееся в выборке с максимальной частотой.

Обращение к функции имеет вид:

=МОДА(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Если в заданных значениях аргументов ***нет повторяющихся значений***, то функция возвращает признак ошибки #Н/Д.

Функция МЕДИАНА вычисляет значение выборки, приходящееся на середину упорядоченной выборочной совокупности. Если выборка имеет четное число элементов, то значение функции будет равно среднему двух значений, находящихся по середине упорядоченной выборочной совокупности. Например, медиана выборки (200, 236, 250, 305, 337, 220) будет равна (236 + 250) / 2 = 243.

Обращение к функции имеет вид:

=МЕДИАНА(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Функция СКОС вычисляет оценку

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\frac{\left(x_i-\overline{x}_в\right)^3}{d_в^{3/2}}$$

для характеристики асимметрии $\frac{\mu_3}{\sigma^3}$, которая для симметричной плотности распределения равна 0.

Обращение к функции имеет вид:

=СКОС(*арг1; арг2; …; арг30*),

где *арг1; арг2; …; арг30* – числовые константы или адреса ячеек, содержащих числовые данные.

Вычисление описательных статистик. Описательные статистики можно разделить на следующие группы:

- *характеристики положения* описывают положение данных на числовой оси (среднее, минимальное и максимальное значения, медиана и др.);
- *характеристики разброса* описывают степень разброса данных относительно своего центра (дисперсия, размах выборки, эксцесс, среднеквадратическое отклонение и др.);
- *характеристики асимметрии* определяют симметрию распределения данных относительно своего центра (коэффициент асимметрии, положение медианы относительно среднего и др.);
- *характеристики, описывающие закон распределения* (частоты, относительные частоты, гистограммы и др.).

Основные характеристики положения, разброса и асимметрии можно вычислить, используя режим ***Описательная статистика*** команды *Пакет анализа*.

Для вызова режима ***Описательная статистика*** необходимо обратиться к вкладке ***Данные***, выбрать опцию ***Пакетанализа***, выбрать в списке режимов ***Описательная статистика*** и щелкнуть на кнопке ОК. В появившемся диалоговом окне Описательная статистика задать следующие параметры (рис. 8):

Входной интервал: – адреса ячеек, содержащих элементы выборки.

Группирование: – задает способ расположения (по столбцам или по строкам) элементов выборки.

Метки в первой строке – включается, если первая строка (столбец) во входном интервале содержит заголовки.

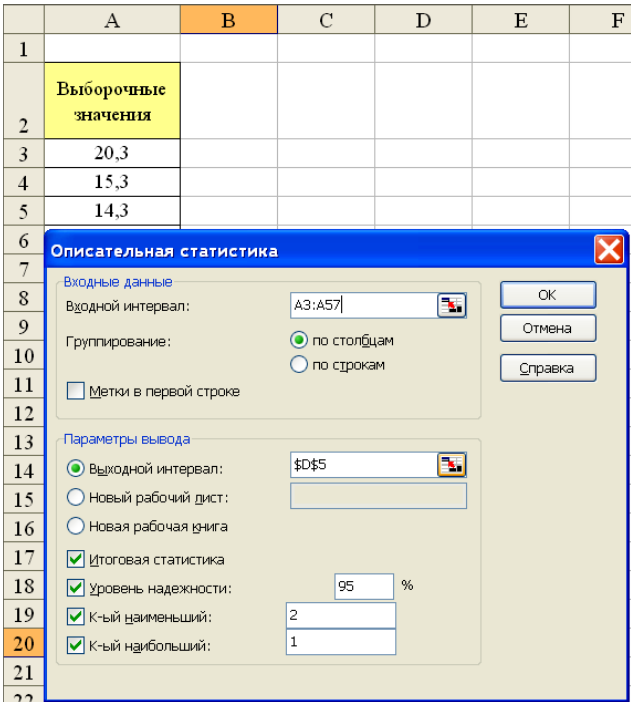


Рис. 8. Параметры режима ***Описательная статистика***

Выходной интервал: / Новый рабочий лист: / Новая рабочая книга – определяет место вывода результатов вычислений. При включении *Выходной интервал:* в поле вводится адрес ячейки, начиная с которой будут выводиться результаты.

Итоговая статистика: – включается, если необходимо вывести по одному полю для каждой из вычисленных характеристик.

Уровень надежности: – включается, если необходимо вычислить доверительный интервал для математического ожидания с задаваемым ($в$ %) уровнем надежности γ .

К-й наименьший: – включается, если необходимо вычислить к-й наименьший (начиная с $x_{\min}$) элемент выборки. При к = 1 вычисляется наименьшее значение.

К-й наибольший: – включается, если необходимо вычислить к-й наибольший (начиная с $x_{\max}$) элемент выборки. При к = 1 вычисляется наибольшее значение.

Пример задания параметров приведен на рис. 8.

Результаты работы режима ***Описательная статистика*** выводятся в виде таблицы, в левом столбце которой приводится название вычисленной характеристики (рис. 3.7), позволяющее однозначно трактовать характеристику. Тем не менее, поясним следующие названия характеристик:

- *Интервал* – определяет размах выборки $x_{\max} - x_{\min}$;
- *Сумма* – определяет сумму всех элементов выборки;
- *Счет*– определяет число обработанных элементов выборки;
- *Уровень надежности* – определяет величину $\Delta_{\bar{x}}$, от которой зависит доверительный интервал для математического ожидания, имеющий вид

$$[\bar{x}_в - \Delta_{\bar{x}}, \bar{x}_в + \Delta_{\bar{x}}],$$

где $\bar{x}_в$ – выборочное среднее.

Пример 2. По выборке Примера 1 вычислить описательные статистики, используя режим ***Описательная статистика.***

Решение. Первоначально, начиная с ячейки А3, введем в столбец А 55 элементов выборки. Последовательно обратимся к командам *Данные-Пакет анализа*. В списке режимов выберем ***Описательная статистика***. В появившемся диалоговом окне включим параметры, показанные на рис. 8, и щелкнем ОК. Вычисленные характеристики приведены на рис. 9.

Столбец1	
Среднее	***17,907***
Стандартная ошибка	***0,399***
Медиана	***18,100***
Мода	***17,800***
Стандартное отклонение	***2,960***
Дисперсия выборки	***8,760***
Эксцесс	***-0,078***
Асимметричность	***-0,386***
Интервал	***13,700***
Минимум	***10,100***
Максимум	***23,800***
Сумма	***984,900***
Счет	***55,000***
Наибольший(2)	***23,300***
Наименьший(1)	***10,100***
Уровень надежности(95,0%)	***0,800***

Рис. 9. Результаты работы *Описательная статистика*

Сравните значения полученных характеристик (см. рис. 9) со значениями аналогичных характеристик, вычисленных в предыдущих примерах.

Варианты заданий для самостоятельного решения:

№	Результаты статистических измерений											
1	11,7	5,49	7,43	9,92	3,41	6,83	8,22	8,30	8,14	9,29	9,27	7,43
	7,41	7,72	12,1	6,06	10,6	6,76	8,21	9,86	8,13	9,04	4,75	9,33
2	4,49	7,94	9,10	6,27	6,77	3,47	8,84	6,48	4,92	6,98	10,1	6,32
	6,36	7,92	12,0	7,46	7,01	13,0	7,34	6,71	5,48	9,95	11,9	8,89
3	6,13	9,77	9,17	8,89	6,19	7,70	6,96	6,72	6,08	4,41	5,52	9,59
	9,02	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98
4	6,52	7,91	5,77	8,02	3,07	2,22	5,76	11,67	6,62	7,07	12,5	1,65
	10,5	7,62	4,94	5,39	3,64	4,62	8,88	6,75	5,77	6,38	10,3	5,74
5	8,18	6,06	5,85	6,78	5,60	10,8	7,70	6,44	8,64	6,95	5,66	4,84
	4,96	5,57	6,47	5,97	8,02	3,66	9,24	4,13	6,58	7,51	5,67	7,89
6	10,2	8,77	10,48	9,44	9,09	6,30	9,42	6,12	9,69	8,59	8,68	7,97
	8,64	5,29	5,00	8,42	8,84	8,26	6,66	6,96	6,51	6,72	6,00	5,36
7	7,13	9,77	9,17	8,89	6,19	7,71	6,96	6,72	6,08	4,41	5,52	9,59
	8,06	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98
8	3,53	7,03	9,18	7,45	5,59	6,85	11,3	7,90	6,00	6,68	5,66	8,64
	8,87	11,34	5,02	4,33	9,31	10,3	5,99	6,98	5,23	8,75	7,73	9,16
9	3,38	4,04	8,21	4,08	3,46	4,37	6,66	1,46	5,59	3,78	8,73	5,57
	8,22	3,38	4,20	2,49	6,11	4,54	6,53	5,20	3,84	5,35	9,72	4,63
10	4,21	3,45	6,79	3,39	2,99	3,88	3,77	1,43	5,96	4,94	6,55	5,92
	4,20	5,64	5,58	5,87	5,05	3,55	7,95	4,45	5,85	6,68	1,24	7,09
11	9,38	7,61	7,52	8,42	8,96	9,07	6,98	8,07	10,89	9,95	9,95	9,04
	11,5	8,95	8,52	6,47	6,52	5,89	6,15	8,67	10,15	9,77	8,65	5,87
12	8,31	9,39	8,47	8,79	9,78	9,42	10,6	10,92	10,17	6,57	7,26	6,16
	8,09	9,81	12,53	10,2	8,65	9,22	9,26	7,98	10,08	10,7	12,6	6,36
13	6,44	5,30	4,43	5,32	7,03	5,14	3,49	3,97	8,77	7,92	8,59	4,99
	6,87	6,78	7,54	5,33	9,13	4,91	2,56	7,62	3,61	5,40	2,21	4,46
14	9,90	8,77	4,57	11,2	8,53	8,57	9,74	12,01	6,24	7,81	8,81	10,9
	9,89	9,32	8,10	9,80	8,68	8,40	9,29	10,93	9,02	9,33	11,8	9,94
15	11,0	6,73	10,90	11,6	13,3	10,7	10,3	11,38	12,04	10,1	8,89	9,05
	12,4	9,64	10,28	7,29	9,39	11,4	9,68	10,13	11,34	11,3	10,4	8,43

Контрольные вопросы для самопроверки .

1.Генеральные среднее, дисперсия, моменты высших порядков (симметрия, эксцесс).
2. Повторные и бесповторные выборки.
3.Математическое ожидание и дисперсия выборочного среднего.
4.Несмещенность, состоятельность и эффективность точечных оценок.
5. Точечные оценки для математического ожидания и дисперсии.
6. Назначение пакета Анализ данных в Excel
7. Встроенные функции для вычисления точечных характеристих в Excel.

Лабораторная работа № 3. Вычисление границ доверительных интервалов в Excel.

Цели и задачи работы.

Цель работы – научиться вычислять интервальные оценки параметров случайной величины: интервальные оценки математического ожидания и дисперсии нормального распределения, интервальная оценка вероятности события.

Задачи работы:

- уметь исследовать распределение случайной величины χ^2.заданной распределение К. Пирсона;
- уметь исследовать распределение случайной величины заданной распределением Стьюдента (t-распределение);
- уметь исследовать распределение случайной величины заданной распределением Фишера (F-распределение);
- иметь понятие интервальной оценки параметра случайной величины;
- уметь вычислять интервальные оценки математического ожидания нормального распределения;
- уметь вычислять интервальные оценки дисперсии нормального распределения;
- уметь вычислять интервальные оценки вероятности события;
- научиться вычислять величины x_γ, $t(\gamma, n)$, $\frac{x_\gamma \sigma}{\sqrt{n}}$,$\chi^2_{\text{лев.}\gamma}$, $\chi^2_{\text{пр.}\gamma}$ с помощью встроенных функций Excel.

Теоретическая часть.

1. Некоторые распределения выборочных характеристик.

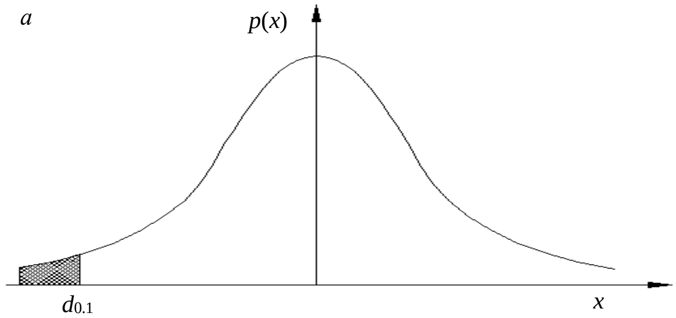
Генеральные совокупности часто имеют нормальный закон распределения. В этом случае многие выборочные характеристики, в том числе $\overline{X}_в, D_в, S^2$, выражаются через небольшое число распределений. В математической статистике, как правило, используются не плотности этих распределений, а некоторые характеристики, представленные таблицами. Чаще всего в качестве такой характеристики выступает квантиль распределения.

Квантилем уровня $p(0 < p < 1)$ или ***p-квантилем*** случайной величины X называется такое число d_p, что вероятность $P(X < d_p)$ равна заданной величине p.

Из определения следует, что если непрерывная случайная величина X имеет плотность распределения *p(x)*, то квантиль d_p определяется равенством

$$\int_{-\infty}^{d_p} p(x)dx = p \qquad (1)$$

Это означает, что площадь фигуры, ограниченной осью абсцисс, кривой *f(x)* и прямой $x=d_p$, равна величине p. На рис.1,а показан квантиль $d_{0.1}$, а на рис.1,б – квантиль $d_{0.9}$. Площади заштрихованных фигур равны 0,1 и 0,9 соответственно.



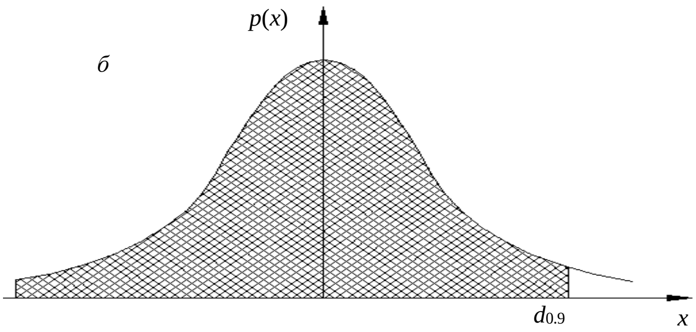


Рис.1 Определение квантилей случайной величины.

Рассмотрим несколько распределений, которым подчиняются выборочные характеристики и которые используются для построения интервальных оценок.

Распределение χ^2 (распределение К.Пирсона). Пусть $N_1 \dots N_n$ – независимые нормально распределенные случайные величины с параметрами (0,1), распределение случайной величины

$$\chi_n^2 = N_1^2 + N_2^2 + N_3^2 + \dots + N_n^2 \qquad (2)$$

Называется *распределением* χ^2 *с n степенями свободы*, а сама величина χ^2 - *случайной величиной* χ^2 *с n степенями свободы.*

Заметим, что количество степеней свободы n является единственным параметром χ^2 - распределения и значения χ^2 неотрицательны, т.е. $P(\chi_n^2 < 0) = 0$.

Определим математическое ожидание величины χ^2. По определению (2) имеем

$$M(\chi_n^2) = M\left(\sum_{i=1}^{n} N_i^2\right) = \sum_{i=1}^{n} M(N_i^2) = \sum_{i=1}^{n}\left[D(N_i) + M^2(N_i)\right]$$

так как $D(X) = M(X^2) - M^2(X)$,

Но $D(N_1) = 1$, $M(N_1) = 0$, а значит, $M(\chi_n^2) = n$.

Нетрудно вычислить и дисперсию случайной величины χ_n^2. Так как случайные величины$N_1^2, \dots, N_n^2$ независимы , то

$$D(\chi_n^2) = nD(N_1^2) = n[M(N_1^4) - M(N_1^2)]. \qquad (3)$$

Плотность распределения случайной величины N_1 равна $p(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = 3$.

Следовательно, $M(N_1^4) = \int_{-\infty}^{\infty} x^4 p(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} = 3$

.

Последний интеграл вычисляется методом интегрирования по частям. Далее, так как $M(N_1^2) = 1$, то $D(\chi_n^2) = n(3-1) = 2n$. Таким образом, χ_n^2 - распределение с n степенями свободы имеет следующие числовые характеристики: $M[\chi_n^2] = n; \quad D[\chi_n^2] = 2n$ *(4)*

Согласно центральной предельной теореме, если случайные величины $N_1^2, \dots, N_n^2$ независимы, одинаково распределены и имеют конечные дисперсии, то последовательность $\chi_n^2 = N_1^2 + N_2^2 + \dots + N_n^2$ асимптотически нормальна. Другими словами, при больших значениях n распределение случайной величины χ_n^2 близко к нормальному распределению с параметрами $a = n, \sigma^2 = 2n$. Однако при малых значениях n функция плотности случайной величины χ_n^2значительно отличается от кривой Гаусса.

На рис. 2 показаны плотности распределения *p(x)* случайной величины χ_n^2 при *n=2, n=6 и n=20*. Видно, что при увеличении *n* плотность *p(x)* «приближается» к плотности нормального распределения.

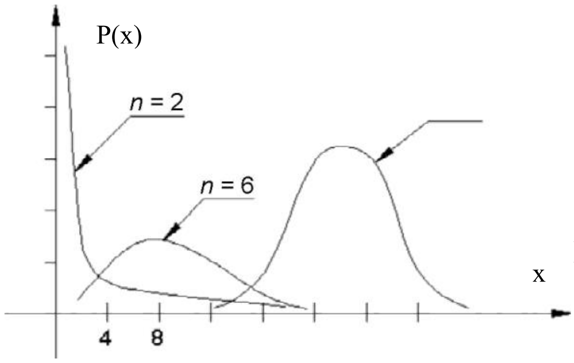


Рис. 2. Плотность распределенияχ_n^2

Обратим внимание на одно замечательное свойство распределения χ_n^2. Строго, говоря это свойство состоит в том, что сумма независимых случайных величин $\chi_n^2 + \chi_m^2$ также распределена по закону χ^2 с *(n+m)* степенями свободы. Объясняется это тем,что случайная величина $\chi_n^2 + \chi_m^2$ представляется в виде суммы *(n+m)* квадратов случайных величин, независимых и нормально распределенных с параметрами (0,1).

Распределение Стьюдента (t-распределение). Пусть *N(0,1)* – нормально распределенная величина с параметрами $a = 0, \sigma = 1$, а χ_n^2 - независимая от *N(0,1)* случайная величина, подчиняющаяся распределению χ^2 с n степенями свободы. Тогда

$$T_n = \frac{N(0,1)\sqrt{n}}{\sqrt{\chi_n^2}} \qquad (5)$$

называется *t-распределением или распределение Стьюдента*. Сама случайная величина (5) называется *t-величиной с n степенями свободы*. Плотность вероятности случайной величины T_n имеет вид

$p_n = B_n\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$, где B_n- некоторая константа, удовлетворяющая условию нормирования $\int_{-\infty}^{\infty} p_n(x)dx = 1$. При больших значениях n кривая $p_n(x)$ близка к кривой нормального распределения *N(0,1)*. Поэтому в практических расчетах при *n>30* часто считают, что

$$p_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Заметим, что функция плотности $p_n(x)$симметрична относительно оси ординат.

Распределение Фишера (F-распределение). Пусть χ_n^2 и χ_m^2 -независимые случайные величины, имеющие χ_n^2- распределение с n и m степенями свободы соответственно. Распределение случайной величины

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m} \qquad (6)$$

называется F- распределением или распределением Фишера с n и m степенями свободы, а сама величина (6) - $F_{n,m}$ величиной. Так как случайные величины$\chi_n^2 \geq 0$ и $\chi_m^2 \geq 0$, то $F_{n,m} \geq 0$.

В дальнейшем мы часто будем ссылаться на следующую теорему о распределении выборочных характеристик $\bar{X}, D_в$, доказанную Р.Фишером.

Теорема: (о распределении выборочных характеристик). Если генеральная совокупность X распределена по нормальному закону с параметрами a и *σ, то:*

а) случайная величина $\bar{X}_в$ распределена нормально с параметрами $\left(a, \frac{\sigma}{\sqrt{n}}\right)$;

б) $\frac{nD_в}{\sigma^2}$ имеет распределение $\chi_{n-1}^2$$

в) случайные величины $\bar{X}_в$ и $D_в$ независимы.

Мы не будем полностью доказывать эту теорему, а ограничимся доказательством утверждения а). Очевидно, что $\bar{X}_в$ есть линейная комбинация
$\bar{X}_в = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$ независимых, нормально распределенных случайных величин. Как отмечалось в курсе теории вероятностей, в этом случае случайная величина $\bar{X}_в$ распределена нормально.
Легко получить, что
$M(\bar{X}_в) = M\left[\frac{x_1+x_2+\cdots+x_n}{n}\right] = \frac{M(x_1)+M(x_2)+\cdots+M(x_n)}{n} = \frac{na}{n} = a,$
$D(\bar{X}_в) = D\left[\frac{x_1+x_2+\cdots+x_n}{n}\right] = \frac{D(x_1)+D(x_2)+\cdots+D(x_n)}{n^2} = \frac{n\sigma^2}{n^2} = n.$
Тем самым первое утверждение теоремы доказано.

Как следует из в), используя случайные величины $\bar{X}_в$ и $D_в$, можно составить случайную величину T_{n-1}.
Действительно, пронормировав $\bar{X}_в$, получим $\frac{(\bar{X}_в - a)\sqrt{n}}{\sigma} = N(0,1)$.
Так как $\bar{X}_в$ и $D_в$ независимы, то по (5)

$$T_{n-1} = \frac{(\bar{X}_в - \mathrm{a})\sqrt{\mathrm{n}}\sqrt{\mathrm{n}-1}}{\sigma} : \sqrt{\frac{nD_n}{\sigma^2}} = \frac{(\bar{X}_в - \mathrm{a})\sqrt{\mathrm{n}-1}}{\sqrt{D_в}}.$$

Итак, мы получим
Следствие. Если условия теоремы о распределении выборочных характеристик выполнены, то случайная величина

$$\frac{(\bar{X}_в - \mathrm{a})\sqrt{\mathrm{n}-1}}{\sqrt{D_в}}$$

имеет распределение Стьюдента с $(n-1)$ степенями свободы. Напомним, что исправленная дисперсия S^2 определяется как

$$S^2 = \frac{n}{n-1}D_в.$$

Тогда получаем новое
Следствие. Если условия теоремы о распределении выборочных характеристик выполнены, то случайная величина

$$\frac{(\bar{X}_в - \mathrm{a})\sqrt{\mathrm{n}-1}}{\sqrt{S^2}}$$

имеет распределение с *(n-1)* степенями свободы.

2. Понятие интервальной оценки параметра случайной величины.

Вычисляя на основе результатов наблюдений точечную оценку θ^* неизвестного параметра θ, мы понимаем, что величина θ^*является (в силу своей случайности) лишь приближенным значением параметра θ. При большом числе наблюдений точность приближения бывает достаточной для практических выводов в силу несмещенности, состоятельности и эффективности «хороших» оценок. Для выборок малого объема точечные оценки могут значительно отличаться от оцениваемого параметра и вопрос о точечности получаемых оценок становится очень важным. В математической статистике он решается введением интервальных оценок.

Интервальной оценкойдля параметра θ называется такой интервал $\left(\underline{\theta^*}, \overline{\theta^*}\right)$ со случайными границами, что

$$P\left(\underline{\theta^*} < \theta < \overline{\theta^*}\right) = \gamma.$$

Вероятность γ называется *надежностью интервальной оценки* или *доверительной вероятностью*, случайные величины $\underline{\theta^*}, \overline{\theta^*}$ - *доверительными границами*, а сам интервал $\left(\underline{\theta^*}, \overline{\theta^*}\right)$ иногда называют *доверительным интервалом*. Центром этого интервала является значение точечной оценки θ^*.

Надежность γ принято выбирать равной 0,95 . 0,99. Тогда событие, состоящее в том, что интервал $\left(\underline{\theta^*}, \overline{\theta^*}\right)$покроет параметр θ, будет практически достоверным.

Общая теория построения интервальных оценок заключается в определении *случайной величины, зависящей от оцениваемого параметра*. Зная распределение этой случайной величины, находят соответствующие доверительные границы и сам доверительный интервал с требуемой точностью. Посмотрим, как эта идея реализует для различных параметров.

3. Интервальные оценки математического ожидания нормального распределения.

Пусть генеральная совокупность X распределена по нормальному закону $N(a,\sigma)$, причем параметр σ известен, а параметр a требует оценить с надежностью γ. По теореме о распределении выборочных характеристик случайная величина $\frac{(\bar{X}_{\text{в}}-\mathrm{a})\sqrt{\mathrm{n}}}{\sigma}$ распределена по закону $N(0,1)$. На рис. 3 изображен график функции плотности этой случайной величины, т.е. кривая $y=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Выберем число x_γ, т.е.
$P\left(-x_\gamma<\frac{(\overline{X_{\text{в}}}-a)\sqrt{n}}{\sigma}<x_\gamma\right)=\gamma.$

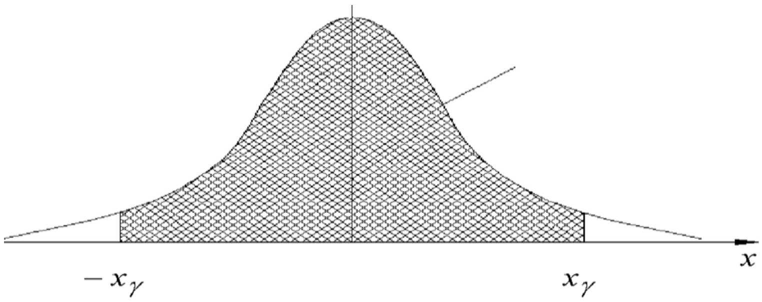


Рис. 3. Построение доверительных интервалов

Это значение легко находится с использованием интегральной функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt.$
Действительно,

$$P\left(-x_\gamma<N(0,1)<x_\gamma\right)=\Phi\left(x_\gamma\right)-\Phi\left(-x_\gamma\right)=\gamma(9)$$

Значение x_γ, удовлетворяющие нелинейному уравнению

$$\Phi\left(x_\gamma\right)=\frac{\gamma}{2} \qquad (10)$$

находится по таблице нормированной функции Лапласа (Приложение 1).

Так как $\sigma>0$, то события $-x_\gamma<\frac{(\overline{X_{\text{в}}}-a)\sqrt{n}}{\sigma}<x_\gamma$ и $\overline{X_{\text{в}}}-\frac{x_\gamma\sigma}{\sqrt{n}}<a<\overline{X_{\text{в}}}+\frac{x_\gamma\sigma}{\sqrt{n}}$ ‘эквивалентны, а значит, их вероятности равны:

$$P\left(\overline{X_{\text{в}}}-\frac{x_\gamma\sigma}{\sqrt{n}}<a<\overline{X_{\text{в}}}+\frac{x_\gamma\sigma}{\sqrt{n}}\right)=\gamma. \qquad (11)$$

Таким образом, для параметра a мы построили доверительный интервал (интервальную оценку), левая граница которого - $\overline{X_{\text{в}}}-\frac{x_\gamma\sigma}{\sqrt{n}}$, правая - $\overline{X_{\text{в}}}+\frac{x_\gamma\sigma}{\sqrt{n}}$, а точность – $\delta=\frac{x_\gamma\sigma}{\sqrt{n}}$. Центр этого интервала находится в точке с координатой $\overline{X_{\text{в}}}$, а длина интервала $2\frac{x_\gamma\sigma}{\sqrt{n}}$. Если объем выборки неограниченно возрастает, то интервал стягивается в одну точку $\overline{X_{\text{в}}}$, которая является состоятельной и несмещенной оценкой для параметра a.

Пример 1: По выборке объема n=9 найдено среднее значение $\overline{x_{\text{в}}}=1{,}5$. Считая, что генеральная совокупность распределена по нормальному закону с $\sigma=2$, определить интервальную оценку для математического ожидания с надежностью $\gamma=0{,}95$.
Решение. Используя таблицу Приложения 1, находим, что

$$\Phi\left(x_\gamma\right)=\frac{0{,}95}{2}=0{,}475$$

При $x_\gamma = 1{,}96$. Тогда $\delta = 1{,}96 \cdot \frac{2}{\sqrt{9}} = 1{,}31$ и доверительный интервал (11) имеет границы $(\overline{X_в} - 1{,}31; \overline{X_в} + 1{,}31)$. Таким образом, с вероятностью 0,95 можно быть уверенным в том, что интервал

$$(\overline{X_в} - 1{,}31; \overline{X_в} + 1{,}31) \qquad (12)$$

накроет параметр a или, другими словами, с вероятностью 0,95 значение $X_в$дает значение параметра a с точностью $\delta = 1{,}31$.

Заметим, что эта трактовка неверна, если вместо случайной величины $X_в$ использовать вычисленное по конкретной выборке значение $x_в = 1.5$. Тогда границы интервала (0,19;2,81) будут не случайными и возможны два случая:

- Точка a лежит внутри этого интервала, тогда $P(0{,}19 < a < 2.81) = 1$;
- Точка a не лежит внутри (0,19;2,81), тогда $P(0{,}19 < a < 2.81) = 0$;

Поэтому только для интервала (12) со случайными границами можно утверждать, что

$$P\left(\overline{X_в} - 1{,}31 < a < \overline{X_в} + 1{,}31\right) = 0{,}95$$

Определим теперь интервальную оценку для неизвестной генеральной средней $\overline{x_г}$ нормально распределенной генеральной совокупности X в том случае, когда генеральная дисперсия $D_г$ неизвестна, т.е. построим доверительный интервал для параметра a, если параметр σ неизвестен.

В отличие от предыдущего случая, вместо случайной величины $\frac{(\overline{X_в}-a)\sqrt{n}}{\sigma}$, распределенной по закону N(0,1), рассмотрим случайную величину $\frac{(\overline{X_в}-a)\sqrt{n-1}}{\sqrt{D_в}}$, которая согласно следствию из теоремы 1 распределена по закону Стьюдента T_{n-1}. При заданном значении γ, пользуясь таблицей Приложения 3 (Значения чисел q в зависимости от объема выборки n и надежности γ для определения доверительного интервала среднего квадратического отклонения σ_x), вычислим значение *t(γ,n)* из условия

$$P(-t(\gamma,n) < \frac{(\overline{X_в}-a)\sqrt{n}}{\sqrt{D_в}} < t(\gamma,n)) = \gamma, \qquad (13)$$

где γ – надежность интервальной оценки. Заметим, что в таблице Приложения 3 *n* означает не число степеней свободы, а объем выборки. Число степеней будет *n-1*.

Замена случайной величины $\frac{(\overline{X_в}-a)\sqrt{n}}{\sigma}$ на случайную величину $\frac{(\overline{X_в}-a)\sqrt{n-1}}{\sqrt{D_в}}$ вызвана тем, что закон распределения последней случайной величины известен и в ее запись не входит неизвестный в данном случае параметр σ. Из условия (13) получаем

$$P\left(\overline{X_в} - \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}} < a < \overline{X_в} + \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}}\right) = \gamma.$$

Таким образом, интервальная оценка надежности γ для неизвестной генеральной средней a имеет границы

$$\left(\overline{X_в} - \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}}; \overline{X_в} + \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}}\right).$$

Выразим границы интервала через исправленную дисперсию S^2. Так как $S^2 = \frac{n}{n-1}D_в$, то $\frac{\sqrt{D_в}}{n-1} = \frac{S}{\sqrt{n}}$. Поэтому

$$\frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}} = \frac{t(\gamma,n)S}{\sqrt{n}}$$

Значит, границы доверительного интервала можно записать как

$$\left(\overline{X_в} - \frac{t(\gamma,n)S}{\sqrt{n}}; \overline{X_в} + \frac{t(\gamma,n)S}{\sqrt{n}}\right), \qquad (14)$$

а точность интервальной оценки определить соотношением

$$\delta = \frac{t(\gamma,n)}{\sqrt{n}}S \qquad (15)$$

Как и в предыдущем случае, центр интервала находится в точке $X_в$, но длина интервала $2\frac{t(\gamma,n)}{\sqrt{n}}S$ является случайной величиной, принимающей тем меньшие значения, чем больше значение n. Это объясняется тем, что наличие большей информации $x_1, \ldots, x_n$ о генеральной совокупности X позволяет сузить интервал.

Пример 2: По выборке объема $n=9$ из нормально распределенной генеральной совокупности найдены значения $\overline{x_\text{в}} = 1{,}5$ и $s=2$. Построить интервальную оценку для математического ожидания с надежностью $\gamma = 0{,}95$.

Решение. Пользуясь таблицей Приложения 3, находим величину $t(0{,}95; 9) = 2{,}31$. Тогда точность δ определяется соотношением (15): $\delta = \frac{t(0{,}95;9)S}{\sqrt{n}} = \frac{2{,}31}{3}S = 0{,}77S$, а интервальная оценка имеет границы $(\overline{X_\text{в}} - 0{,}77 \cdot S; \overline{X_\text{в}} + 0{,}77 \cdot S)$, которые зависят от двух случайных величин: $\overline{X_\text{в}}$ и S . Подставляя вместо S ее вычисленное значение S=2, получаем интервал $(\overline{X_\text{в}} - 1{,}54; \overline{X_\text{в}} + 1{,}54)$.

Сравнивая эту оценку с интервальной оценкой примера 1 (см.(12)), видим, что замена неизвестной величины σ вычисляемой величиной S приводит к уменьшению точности интервальной оценки и увеличению длины доверительного интервала. Подставив вместо случайной величины $X_\text{в}$ ее конкретное значение $\overline{x_\text{в}} = 1{,}5$, получаем конкретное значение границ (0;3).

a. Интервальные оценки дисперсии нормального распределения.

Как и при построении интервальных оценок для математического ожидания, в данном случае также необходимо определить случайную величину, распределение которой было известно и включало оцениваемый параметр σ. В соответствии с теоремой 1 такой отправной точкой для построения доверительного интервала может быть случайная величина $\frac{nD_\text{в}}{\sigma^2}$, распределенная по закону χ^2 с *(n-1)* степенями свободы. Заметим, что доверительные интервалы, построенные для параметра a, вообще говоря, можно было выбрать несимметричными относительно $\overline{X_\text{в}}$ и это не противоречило бы определению интервальной оценки. Но такой выбор интервала, когда в его середине лежит состоятельная и несовмещенная оценка параметра, являлся предпочтительным. В данном случае целесообразно выбрать два предела$\chi^2_{\text{лев.}\gamma}\chi^2_{\text{пр.}\gamma}$ так, что

$$P\left(\chi^2_{n-1} < \chi^2_{\text{лев.}\gamma}\right) = P\left(\chi^2_{n-1} < \chi^2_{\text{пр.}\gamma}\right) = \frac{a}{2},$$

где $a = 1 - \gamma$, γ-надежность интервальной оценки.

Следовательно,$\chi^2_{\text{лев.}\gamma}$ –квантиль χ^2_{n-1}-распределения уровня $\frac{a}{2}$. $\chi^2_{\text{пр.}\gamma}$ - уровня $1 - \frac{a}{2}$.

Тогда имеет место равенство $P\left(\chi^2_{\text{лев.}\gamma} < \frac{nD_\text{в}}{\sigma^2} < \chi^2_{\text{пр.}\gamma}\right) = \gamma$, а интервал

$$\left(\frac{nD_\text{в}}{\chi^2_{\text{пр.}\gamma}}, \frac{nD_\text{в}}{\chi^2_{\text{лев.}\gamma}}\right) \qquad (16)$$

является интервальной оценкой σ^2 надежности γ.

Так как $D_\text{в} = {(n-1)S^2}/{n}$, то $nD_\text{в} = (n-1)S^2$ и интервал

$$\left(\frac{n-1}{\chi^2_{\text{пр.}\gamma}}S^2, \frac{n-1}{\chi^2_{\text{лев.}\gamma}}S^2\right) \quad , \qquad (17)$$

является интервальной оценкой для σ^2 надежности γ.

Заметим, что границы интервалов (16), (17) являются случайными величинами и с вероятностью γ можно утверждать, что интервалы (16), (17) накроют неизвестную дисперсию σ^2.

Пример 3: По выборке объема n=20 из нормально распределенной генеральной совокупности вычислено значение дисперсии выборки $d_\text{в} = 1{,}5$. Построить интервальную оценку для параметра σ^2 надежности γ=0,96.

Решение: Значения $\chi^2_{\text{лев.}\gamma}\chi^2_{\text{пр.}\gamma}$ находим из условий:

$$P\left(\chi^2_{19} < \chi^2_{\text{лев.}\gamma}\right) = 0{,}02;\ P\left(\chi^2_{19} < \chi^2_{\text{пр.}\gamma}\right) = 0{,}98.$$

Эти условия означают, что $\chi^2_{\text{лев.}\gamma}$есть квантиль χ^2-распределения с 19 степенями свободы уровня 0,02, а $\chi^2_{\text{пр.}\gamma}$ – квантиль уровня 0,98. По таблице Приложения 3 квантилей χ^2-распределения находим

$$\chi^2_{\text{лев.}\gamma} = 8{,}6;\ \chi^2_{\text{пр.}\gamma} = 33{,}7$$

Тогда интервальная оценка (16) примет вид

$$(0{,}59D_{\text{в}}; 2{,}33D_{\text{в}}).$$

Подставляя вычисленное значение $d_{\text{в}} = 1{,}5$случайной величины $D_{\text{в}}$, получаем

$$0{,}89 < \sigma^2 < 3{,}488.$$

b. Интервальная оценка вероятности события.

Выше было показано, что «хорошей» точечной оценкой вероятности p события является частность $p^* = \frac{m}{n}$ (см.(17)), где n – общее число независимых испытаний, в каждом из которых событие A может произойти с вероятностью p, а m – число испытаний , в которых произошло событие A.

Зададим надежностью интервальной оценки γ и найдем числа $p_{\text{лев.}\gamma}$, $p_{\text{пр.}\gamma}$такие, чтобы выполнялось соотношение

$$P(p_{\text{лев.}\gamma} < p < p_{\text{пр.}\gamma)=\gamma}.$$

Интервальную оценку построим для двух случаев: когда число испытаний n сравнительно велико $(np > 10, n > 30)$ и для малого числа испытаний.

Интервальная оценка вероятности при большом числе испытаний.

Если $(np > 10, n > 30)$, то распределение случайной величины $p^* = \frac{m}{n}$ можно аппроксимировать нормальным распределением $N(p, \sqrt{pq/n})$. Следовательно, при этих же условиях распределение величины $\frac{(p^*-p)}{\sqrt{pq/n}}$ близко к нормальному с нулевым математическим ожидание и единичной дисперсией, т.е.

$$\frac{(p^*-p)}{\sqrt{pq/n}} = N(0{,}1). \ (18)$$

По аналогии с (8) найдем такое число x_γ, для которого справедливо равенство

$$P\left(-x_\gamma < \frac{(p^*-p)}{\sqrt{pq/n}} < x_\gamma\right) = \gamma. \ (19)$$

Это число является корнем уравнения

$$\Phi(x_\gamma) = \frac{\gamma}{2},$$

где $\Phi(x)$- функция Лапласа, и корень может быть найден с помощью табл. П1.

Неравенство, стоящее в скобках выражения (19), разрешим относительно р. Для этого неравенство перепишем в виде эквивалентного неравенства $\left|\frac{(p^*-p)}{\sqrt{pq/n}}\right| < x_\gamma$. Возведем в квадрат, в результате получим $(p^* - p)^2 < \frac{p(1-p)}{n} x_\gamma^2$. Далее возведя в квадрат $(p^* - p)$ и перенеся все члены влево, получим

$$\left(1 + \frac{x_\gamma^2}{n}\right)p^2 - \left(2p^* + \frac{x_\gamma^2}{n}\right)p + p^{*2} < 0.$$

Корни p_1и p_2 квадратного трехчлена, состоящего в правой части неравенства, определяются выражениями

$$p_1 = \frac{p^* + x_\gamma^2/(2n) - x_\gamma\sqrt{p^*(1-p^*)/n + x_\gamma^2/(4n^2)}}{1 + x_\gamma^2/n}, \qquad (20)$$

$$p_2 = \frac{p^* + x_\gamma^2/(2n) + x_\gamma\sqrt{p^*(1-p^*)/n + x_\gamma^2/(4n^2)}}{1 + x_\gamma^2/n}. \qquad (21)$$

Корни этого уравнения и являются границами интервальной оценки (18)

$$p_{\text{лев.}\gamma} = p_1;\ p_{\text{пр.}\gamma} = p_2. \qquad (22)$$

Если $n \gg 100$, то для вычисления p_1и p_2 можно использовать приближенные формулы:

$$p_1 = p^* - x_\gamma\sqrt{p^*(1-p^*)/n};\ \ p_2 = p^* + x_\gamma\sqrt{p^*(1-p^*)/n}. \qquad (23)$$

Видно, что границы интервала (18) являются случайными величинами и конкретные значения границ получаются в результате подстановки наблюдаемого значения случайной величины p^*.

Пример 4: Событие A в серии из $n=100$ испытаний произошло $m=78$ раз. Построить интервальную оценку для вероятности p события с надежностью γ=0,9.

Решение. Значение точечной оценки вероятности p равно $p^* = \frac{78}{100} = 0{,}78$. По табл. П1 определяем $x_\gamma = 1{,}64$ и вычисляем по формулам (20), (21) значения p_1 и p_2 при $p^* = 0{,}78$: $p_1=0{,}705$, $p_2=0{,}848$. Таким образом, получили реализацию достоверного интервала (0,705;0,848) для вероятности p события A.

Интервальная оценка вероятности при малом числе испытаний. При малом числе испытаний n предположение о приближенном распределении случайной величины m необходимо использовать формулу Бернулли:

$$P(m = x) = C_n^x p^x (1-p)^{n-x},\ x = 1,1, \dots n.$$

Можно показать, что граничные точки интервальной оценки (18) являются решениями следующих нелинейных уравнений:

$$\sum_{x=0}^{m-1} C_n^x p_{\text{лев.}\gamma}^x \left(1 - p_{\text{лев.}\gamma}\right)^{n-x} = \frac{1+\gamma}{2}; \qquad (24)$$

$$\sum_{x=0}^{m-1} C_n^x p_{\text{пр.}\gamma}^x \left(1 - p_{\text{пр.}\gamma}\right)^{n-x} = \frac{1-\gamma}{2}, \qquad (25)$$

где γ – надежность интервальной оценки. Вновь заметим, что решения $p_{\text{лев.}\gamma,}, p_{\text{пр.}\gamma}$ этих уравнений являются случайными величинами и только при подстановке конкретного значения m (количество испытаний, в которых появилось событие A) будут получены конкретные значения граничных точек интервальной оценки (18).

Корни уравнений (24), (25) могут быть найдены одним из известных численных методов решения нелинейных уравнений. Кроме этого, существуют специальные таблицы для нахождения $p_{\text{лев.}\gamma,}, p_{\text{пр.}\gamma}$, удовлетворяющих уравнениям (24), (25) по заданным n, m-n, γ. Фрагмент этих таблиц представлении в приложении (табл.П4).

Пример 5: В пяти испытаниях событие A произошло три раза. Построить интервальную оценку для вероятности p события А с надежностью, $\gamma=0{,}95$.

Решение: Из условий примера , имеем $n=5$, $m=3$, $\gamma=0{,}95$. По таб.П4 находим $p_{\text{лев.}\gamma,} = 0{,}147$, $p_{\text{пр.}\gamma} = 0{,}947$, а интервальная оценка определяется как (0,147;0,947).

Сравнивая интервальные оценки примеров 4, 5, видим, что длина доверительного интервала для примера 5 (равна 0,8) существенно больше длины доверительного интервала примера 4 (0,143). Это является следствие разного объема выборок (n=5 и n=100) и различных дисперсий случайной величины $p^* = \frac{m}{n}$.

4. *Вычисление границ доверительных интервалов в Excel.*

Границы доверительных интервалов зависят от некоторой величины, которая зависит от распределения точечной оценки и доверительной вероятности. Эта величина находится по специальным таблицам. Поэтому часто возникает необходимость интерполяции или экстраполяции табличных данных и, следовательно, требуются дополнительные вычисления. В табличном процессоре Excel определены функции, позволяющие в интервальные оценки для различных числовых характеристик случайной величины.

Вычисление величины $\boldsymbol{x_\gamma}$, входящей в доверительный интервал (11):

$$\left| \overline{X_{\text{в}}} - \frac{x_\gamma \sigma}{\sqrt{n}}; \overline{X_{\text{в}}} + \frac{x_\gamma \sigma}{\sqrt{n}} \right|, \qquad (26)$$

Величина x_γ является корнем нелинейного уравнения (10) и вычисляется с помощью функции **НОРМСТОБР**:

$$x_\gamma = \text{НОРМСТОБР}((\gamma + 1)/2),$$

Где γ – надежность интервальной оценки (26).

Вычисление величины $\frac{x_\gamma \sigma}{\sqrt{n}}$ осуществляется с помощью функции **ДОВЕРИТ:**

$$\Delta \overline{X}_{\text{в}} = \frac{x_\gamma \sigma}{\sqrt{n}} = \text{ДОВЕРИТ } (a; \sigma; n) \ ,$$

где $a = 1 - \gamma$, σ – известное среднеквадратическое отклонение, n – объем выборки. Тогда интервальную оценку (26) можно записать в виде

$\left| X_{\text{в}} - \Delta \overline{X}_{\text{в}}; X_{\text{в}} + \Delta \overline{X}_{\text{в}} \right|$.

Вычисление величины $\boldsymbol{t(\gamma, n)}$**,** входящих в доверительный интервал

$$\left| \overline{X}_{\text{в}} - \frac{t(\gamma, n)\sqrt{D_{\text{в}}}}{\sqrt{n-1}}; \overline{X}_{\text{в}} + \frac{t(\gamma, n)\sqrt{D_{\text{в}}}}{\sqrt{n-1}} \right|$$

осуществляют с использованием функции **СТЬЮДРАСПОБР,** обращение к которой имеет вид:

$$t(\gamma, n) = \text{СТЬЮДРАСПОБР}(a; n),$$

где $a = 1 - \gamma$, n – число степеней свободы.

Вычисление величин $\boldsymbol{\chi^2_{\text{лев.}\gamma}}$**,** $\boldsymbol{\chi^2_{\text{пр.}\gamma}}$, входящих в доверительный интервал (17) для дисперсии σ^2:

$$\left| \frac{n-1}{\chi^2_{\text{пр.}\gamma}} S^2, \frac{n-1}{\chi^2_{\text{лев.}\gamma}} S^2 \right|,$$

гдеS^2 - исправленная дисперсия. Используется функция **ХИ2ОБР**:

$$\boldsymbol{\chi^2_{\text{лев.}\gamma}} = \text{ХИ2ОБР}(1 - \frac{\text{a}}{2} : \text{n});$$

$$\boldsymbol{\chi^2_{\text{пр.}\gamma}} = \text{ХИ2ОБР}(\frac{\text{a}}{2} : \text{n}).$$

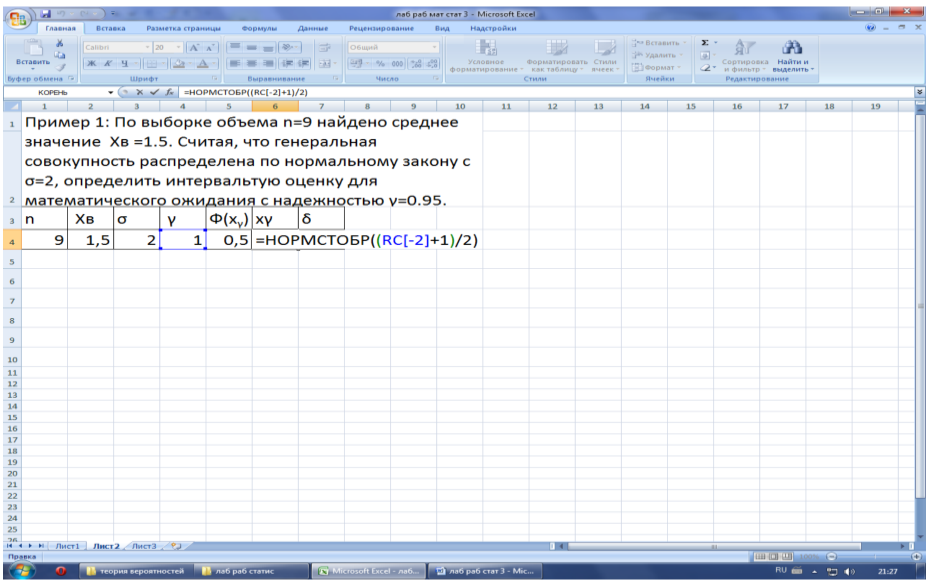
где $a = 1 - \gamma$, γ – надежность интервальной оценки.

Пример решения задач в Excel.

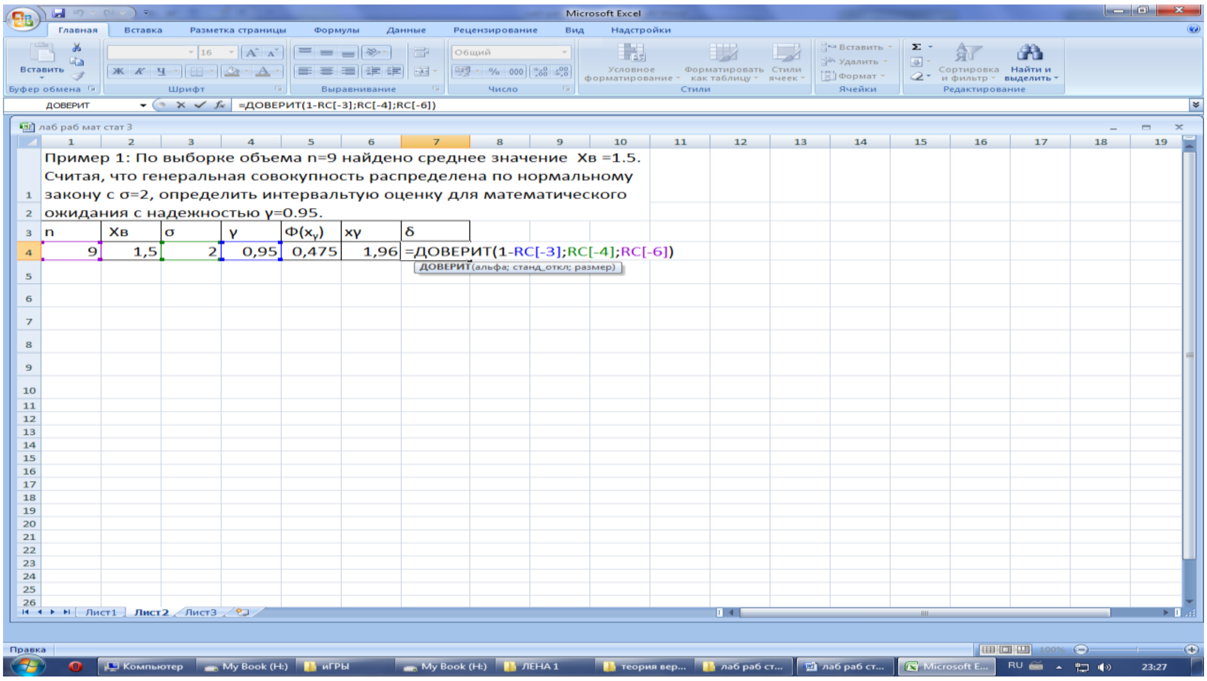
Запишите условие Примера 1 в среде Excel. Оформите исходные данные примера в виде таблицы. Значения n, $x_{\text{в}}$, σ, γ, Φ(x), x_γ.δ занесите в ячейки 1:4,2:4,3:4,4:4,5:4,6:4,7:4.

В ячейке5:4 вычислите значение интегральной Функции Лапласса.

В ячейке 6:4 вычислите значение x_γ с помощью встроенной функции **НОРМСТОБР**



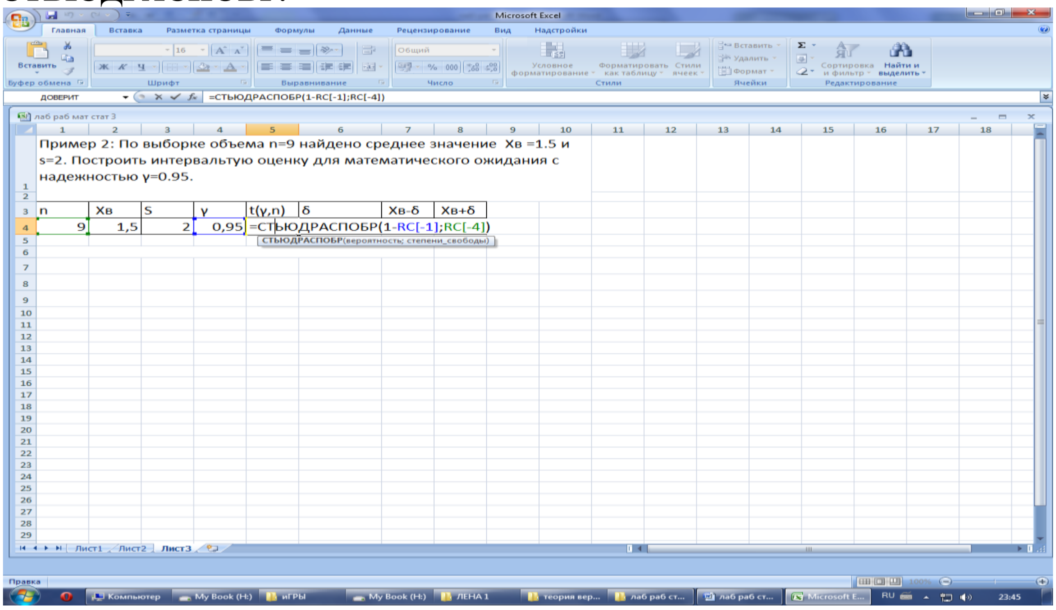
В ячейке 7:4 вычислите точность δ с помощью встроенной функции **ДОВЕРИТ**



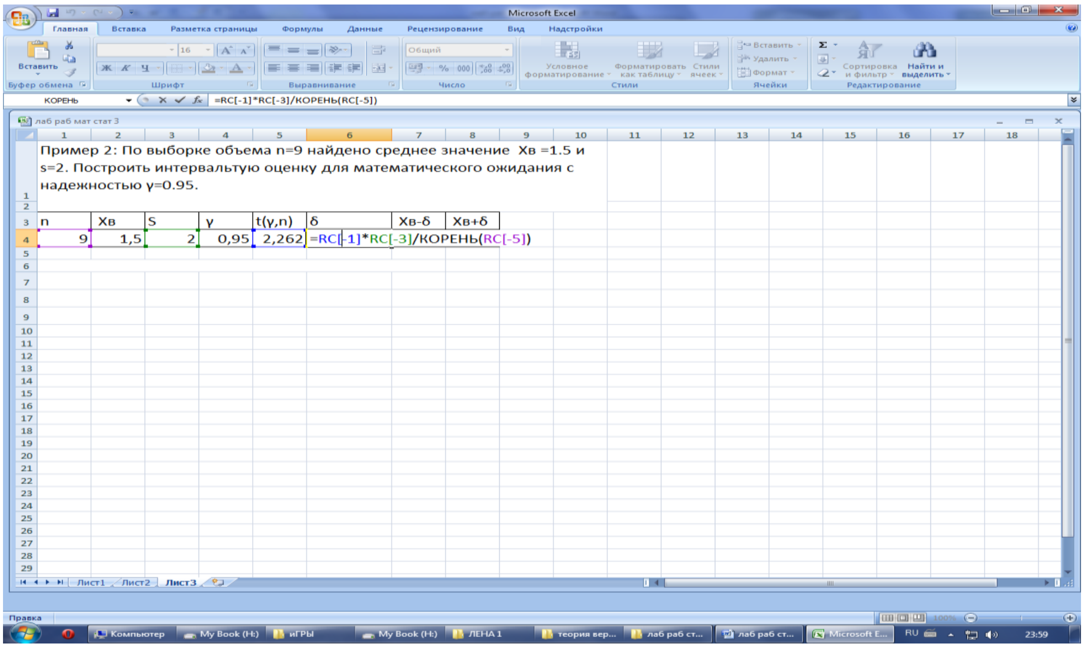
Запишите условие Примера 2 в среде Excel. Оформите исходные данные примера в виде таблицы. Значения n, $x_в$, s, γ, t(γ, n), δ занесите в ячейки 1:4,2:4,3:4,4:4,5:4,6:4,7:4.

Под таблицей запишите результат интервальной оценки математического ожидания случайной велисчины.

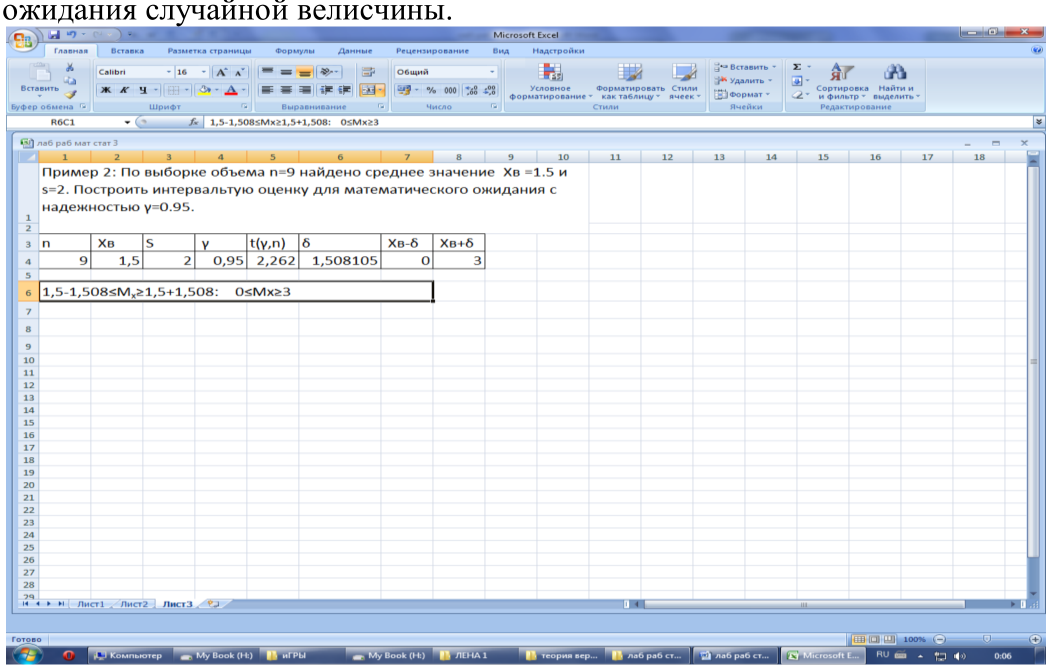
В ячейке 5:4 вычислите значение величины t(γ, n) с помощью встроенной функции **СТЬЮДРАСПОБР.**



В ячейке 6:4 вычислите значение точности δ

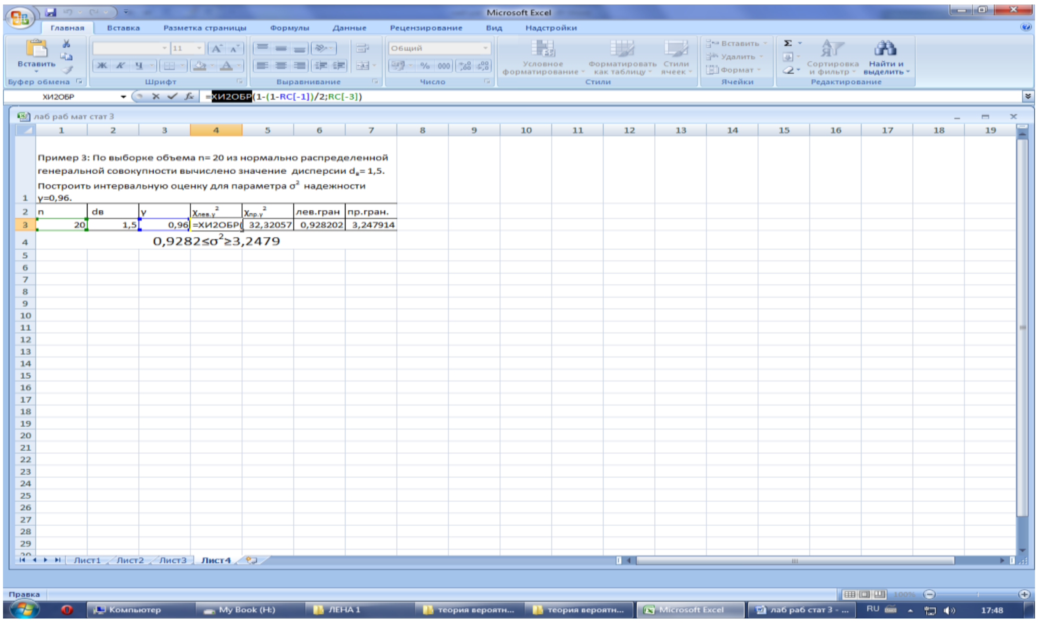


Под таблицей запишите результат интервальной оценки математического ожидания случайной велисчины.



Запишите условие Примера 3 в среде Excel. Оформите исходные данные примера в виде таблицы. Значения n, $d_в$, γ, $\chi^2_{лев.\gamma}$, $\chi^2_{пр.\gamma}$ занесите в ячейки 1:4,2:4,3:4,4:4,5:4.

С помощью функции **ХИ2ОБР** в ячейках 4:4;4:5 вычислите значения $\chi^2_{лев.\gamma}$ и $\chi^2_{пр.\gamma}$. В ячейках 6:4;7:4 вычислите значения левой и правой границ интервальной оценки дисперсии нормального распределения σ^2.



Варианты заданий для самостоятельного решения:

1. Найти доверительный интервал для оценки с надежностью γ неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если генеральное среднее квадратическое отклонение σ, выборочная средняя $\overline{x_в}$ и объем выборки n .

№ варианта	Надежность γ	Генеральное среднее квадратическое отклонение σ	Выборочная средняя $\overline{x_в}$	Объем выборки n
1	0,95	5	14	25
2	0,95	8	12	25
3	0,95	5	14	20
4	0,95	8	12	25
5	0,97	6	15	25
6	0,97	7	13	20
7	0,97	6	12	30
8	0,97	8	11	30
9	0,99	4	10,4	16
10	0,99	5	10,2	25
11	0,99	4	16,2	20
12	0,99	7	15	16
13	0,95	40	2000	5
14	0,95	10	100	5
15	0,95	30	2000	10

2. По выборке объема n из нормально распределенной генеральной совокупности найдены значения $\overline{x_в}$ и s. Построить интервальную оценку для математического ожидания с надежностью γ.

№ варианта	Надежность γ	«исправленное» среднее квадратическое отклонение s	Выборочная средняя $\overline{x_в}$	Объем выборки n
1	0,95	5	1,4	9

2	0,95	8	1,2	10
3	0,95	5	1,4	9
4	0,99	2	12	15
5	0,95	6	1,5	8
6	0,95	7	1,8	10
7	0,95	6	10	9
8	0,99	2	4,8	9
9	0,99	4	10,4	10
10	0,99	5	10,2	15
11	0,99	4	16,2	10
12	0,99	7	15	16
13	0,95	40	2000	5
14	0,95	10	100	5
15	0,95	30	2000	10

3. По выборке объема n из нормально распределенной генеральной совокупности вычислено значение дисперсии выборки $d_{в}$. Построить интервальную оценку для параметра σ^2 надежности γ.

№ варианта	Надежность γ	Дисперсия выборки $d_{в}$	Объем выборки n
1	0,95	2,5	9
2	0,95	1,8	10
3	0,95	5,7	30
4	0,99	2,9	15
5	0,95	6,6	8
6	0,95	7,1	10
7	0,95	6	9
8	0,99	2,7	30
9	0,99	4,2	10
10	0,99	5,3	15
11	0,99	4,4	10
12	0,99	7,2	16
13	0,95	4,8	20
14	0,95	1,7	5
15	0,95	3,0	10

Контрольные вопросы для самопроверки .

1. Квантилем случайной величины X называется?
2. Запишите распределение К.Пирсона.
3. Чему равны математическое ожидание и дисперсия распределения К Пирсона?
4. Запишите распределение Стьюдента.
5. Запишите распределение Фишера.
6. Чему равна исправленная дисперсия?

7. Чему равны интервальные оценки математического ожидания нормального распределения случайной величины?
8. Чему равны границы интервальной оценки дисперсии случайной нормального распределения?
9. С помощью какой статистической функции Excel можно вычислить величину $\boldsymbol{x_\gamma}$**?**
10. С помощью какой статистической функции Excel можно вычислить величину $\frac{x_\gamma \sigma}{\sqrt{n}}$**?**
11. С помощью какой статистической функции Excel можно вычислить величину $\boldsymbol{t(\gamma, n)}$?
12. С помощью какой статистической функции Excel можно вычислить величины $\boldsymbol{\chi^2_{\text{лев.}\gamma}, \chi^2_{\text{пр.}\gamma}}$?

5. **Образовательные технологии**

С целью формирования и развития профессиональных навыков обучающихся в соответствии с требованиями ФГОС ВО по направлению подготовки предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий:

- во время лекционных занятий используется презентация с применением слайдов с графическим и табличным материалом, что повышает наглядность и информативность используемого теоретического материала;
- Практические занятия предусматривают решение задач по соответствующим темам.
- Лабораторные работы проходят с применением пакета MSExcel, использованием его встроенных функций и графики.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Приводятся виды самостоятельной работы обучающегося, порядок их выполнения и контроля, дается учебно-методическое обеспечение (возможно в виде ссылок) самостоятельной работы по отдельным разделам дисциплины.

№п/п	**Раздел дисциплины**	**Трудоемкость (час)**	**Компетенции ОК,ПК**	**Контроль выполнения работы**
1.	**Теория вероятностей**	8	УК-1; ОПК-1;	Устный, письменный опросы. Решение задач. Защита лабораторных работ.
2	**Математическая статистика**	8		Устный, письменный опросы. Решение задач. Защита лабораторных работ.
	Итого	16		

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.
7.1. Типовые контрольные задания

Примерный вариант контрольной контрольной работы №1 ()

1. В банк поступило 20 авизо. Подозревают, что среди них 4 фальшивых. Тщательной проверке подвергают 15 авизо. Чему равна вероятность того, что в ходе проверки обнаружится ровно 2 фальшивых авизо?

2. Покупатель может приобрести акции трех компаний: А, В и С. Надежность первой компании оценивается экспертами на уровне 90%, второй – 95% и третьей – 85%. Чему равна вероятность того, что а) только одна компания станет банкротом

б) наступит хотя бы одно банкротство?

3. На сборку поступают однотипные детали с трёх предприятий, причём первое поставляет 40% , второе - 20% и третье - остальное количество. Вероятность появления брака для первого, второго и третьего поставщиков соответственно равны 0,2; 0,1 и 0,15. Выборочный контроль обнаружил брак. Какому предприятию вероятнее всего принадлежит брак?

Примерный вариант модульной контрольной работы №2

1.

	1	2	3	4	5
P	0,1	0,2	0,4	0,2	0,1

Случайная величина X задана законом распределения
Найти M(X) ,D(X), σ_x

2.Случайная величина X задана плотностью вероятности

$$f(x) = \begin{cases} 0, \text{при } x \leq 0 \\ 1, \text{при } x > 0 \end{cases}$$

Найти вероятность P(1<X<2).

Задания по промежуточному контролю

2. В банк поступило 20 авизо. Подозревают, что среди них 4 фальшивых. Тщательной проверке подвергают 15 авизо. Чему равна вероятность того, что в ходе проверки обнаружится ровно 2 фальшивых авизо?

2. Покупатель может приобрести акции трех компаний: А, В и С. Надежность первой компании оценивается экспертами на уровне 90%, второй – 95% и третьей – 85%. Чему равна вероятность того, что а) только одна компания станет банкротом

б) наступит хотя бы одно банкротство?

3. На сборку поступают однотипные детали с трёх предприятий, причём первое поставляет 40% , второе - 20% и третье - остальное количество. Вероятность появления брака для первого, второго и третьего поставщиков соответственно равны 0,2; 0,1 и 0,15. Выборочный контроль обнаружил брак. Какому предприятию вероятнее всего принадлежит брак?

4. В жилом доме имеется 100 ламп, вероятность включения каждой из них в вечернее время равна 0,8. Найти вероятность того, что число одновременно включённых ламп будет не меньше 60.

5.Дан ряд распределения случайной величины

X	1	4	5	7
P(X)	**0,4**	**0,1**	**0,3**	**0,2**

Найти функцию распределения F(x) и построить ее график

6. Сочетания, размещения, перестановки. Примеры.

Примерный вариант модульной контрольной работы №2

Случаная величина X задана законом распределения:

1.

X	1	2	3	4	5
P	0,1	0,2	0,4	0,2	0,1

Найти M(X) ,D(X), σ_x.

2. Случайная величина X задана плотностью вероятности
2.Случайная величина X задана плотностью вероятности

$$f(x) = \begin{cases} 0, \text{при } x \leq 0 \\ 1, \text{при } x > 0 \end{cases}$$

Найти вероятность P(1<X<2).

3. Случайная величина X задана плотностью распределения

$$f(x) = \begin{cases} 0, \text{при } x \leq 0 \\ 1, \text{при } x > 0 \end{cases}$$

Найти D(x)

4. По данной дифференциальной функции НСВ X ^(**x**) =0, *прих* <0 l-cos2*x*

ж 0, *прих* >***ж***, *приО* <***х*** < ***ж***

Найти ее интегральную функцию F(X)
5. Найти M(X) и D(X) случайной величины X, распределенной равномерно на интервале (2,8).
6. Случайная величина X имеет равномерное распределение с M(X) =7 и D(X)=3. Найти интегральную и дифференциальную функции распределения и вероятность того, что случайная величина попадет в интервал (4;6).

Вопросы подготовки к экзамену
1 модуль

1. Основные понятия теории вероятностей
2.Теорема сложения вероятностей
3.Теорема умножения вероятностей
4. Следствие теорем сложения и умножения вероятностей случайных событий
5.Повторение испытаний. Формула Бернулли. Интегральная и локальная теоремы Лапласа. Формула Пуассона.
1.Виды случайных величин. Закон распределения вероятностей дискретной случайной величины.
2. Функция распределения вероятностей случайной величины, свойства.
3. Плотность распределения вероятностей непрерывной случайной величиныВероятность попадания непрерывной случайной величины в заданный интервал.
4. Нахождение функции распределения по известной плотности распределения
5. Свойства плотности распределения и вероятностный смысл плотности распределения.
7.Математические ожидания дискретных и непрерывных случайных величин.

8. Вероятностный смысл математического ожидания свойства матем. ожидания.
9. Математическое ожидание числа появления события в независимых испытаниях
10.Дисперсия случайной величины. Формула для вычисления дисперсии. Свойства дисперсии. Дисперсия числа появлений события в независимых испытаниях 11. Среднее квадратическое отклонение. Среднее квадратическое отклонение суммы взаимно независимых случайных величин.

12. Биномиальные распределения. Распределение Пуассона.
13. Равномерное распределение. Вероятность попадания СВ в интервал. Числовые характеристики этого распределения.
14. Нормальное распределение. Вероятность параметров нормального распределения на форму нормальной кривой. Вероятность попадания в заданный интервал нормальной случайной величины.
15. Вычисления вероятности заданного отклонения, правило трех сигм.
16. Показательное распределение. Определение этого распределения. Вероятность попадания СВ в заданный интервал показательно распределенной случайной величины.
17. Числовые характеристики показательного распределения.

Модуль 2

1.Генеральная и выборочная совокупности. Статистическое распределение выборки.
2.Эмпирическая функция распределения и вариационный ряд. 3. Гистограмма. Мода и медиана.
4.Генеральные среднее, дисперсия, моменты высших порядков (симметрия, эксцесс). 5. Эмпирическая ковариация. Повторные и бесповторные выборки. 6.Математическое ожидание и дисперсия выборочного среднего для повторной и бесповторной выборки.
7.Несмещенность, состоятельность и эффективность точечных оценок.
8. Оценка неизвестной вероятности по частоте.
9. Точечные оценки для математического ожидания и дисперсии.
10. Метод моментов. Метод максимального правдоподобия.
11.Доверительные вероятности и интервалы.
12.Приближенный доверительный интервал для оценки генерального доли признака.
13. Приближенный доверительный интервал для оценки генерального среднего.
14. Статистическая проверка гипотез. Ошибки I и II рода. Уровень значимости и мощность критерия. Проверка гипотез о равенстве средних и дисперсий двух нормально распределенных генеральных совокупностей.
15. Простые и сложные гипотезы. Хи-квадрат критерий Пирсона.
16.Проверка гипотезы о соответствии наблюдаемых значений предполагаемому распределению вероятностей (дискретному или непрерывному).
17. Сравнение параметров двух нормальных распределений.

7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

1. Общий результат по модулю выводится как интегральная оценка, складывающая из текущей работы - __50__% и текущего контроля - _50____%.
Текущий работа по дисциплине включает:
- посещение занятий - ___ баллов,
- участие на практических занятиях - 100__ баллов,
- защита лабораторных работ -…
Текущий контроль по дисциплине включает:
- устный опрос - 100___ баллов,
- письменная контрольная работа - _100___ баллов,
…
2. Промежуточный контроль
Собеседование- 100___ баллов,
…

8. Учебно-методическое обеспечение дисциплины.

а) адрес сайта курса
Интернет-адрес сайта. В качестве сайта курса рекомендуется использовать сайт кафедры или факультета (института), специализированные учебные сайты (например, на платформе Moodle).
б) основная литература:

1. Бухтоярова В.И. Высшая математика. Часть III. Теория вероятностей. Математическая статистика [Электронный ресурс]: учебное пособие/ В.И. Бухтоярова [и др.].— Электрон. текстовые данные.— Кемерово: Кемеровская государственная медицинская академия, 2006.— 88 с.— Режим доступа: http://www.iprbookshop.ru/6113.html (01.03.2022)
2. **Гмурман, В.Е.** Теория вероятностей и математическая статистика : учеб. для бакалавров / В. Е. Гмурман. - 12-е изд., перераб. и доп. - М. : Юрайт, 2014. - 470-47. **Местонахождение:** Научная библиотека ДГУ **URL:**
3. Кацман Ю.Я. Теория вероятностей, математическая статистика и случайные процессы [Электронный ресурс]: учебник/ Кацман Ю.Я.— Электрон. текстовые данные.— Томск: Томский политехнический университет, 2013.— 131 с.— Режим доступа: http://www.iprbookshop.ru/34722.html (01.02.2022)
4. **Кремер, Наум Шевелевич.** Теория вероятностей и математическая статистика = Probability theory and mathematical statistics : [учеб. для вузов по экон. специальностям] / Кремер, Наум Шевелевич. - 2-е изд., перераб. и доп. - М. : ЮНИТИ-Дана, 2006. - 573 с. : граф. ; 21 см. - Библиогр.: с. 533-534. - Предм. указ.: с. 562-573. - Рекомендовано МО РФ. - ISBN 5-238-00573-3 : 320-00. **Местонахождение:** Научная библиотека ДГУ **URL:**

Дополнительная литература

1. **Бородин, Андрей Николаевич.** Элементарный курс теории вероятностей и математической статистики : [учеб. пособие] / Бородин, Андрей Николаевич. - 6-е изд. - СПб. : Лань, 2006. - 254 с. - (Учебники для вузов. Специальная литература). - ISBN 5-8114-0442-5 : 121-00. **Местонахождение:** Научная библиотека ДГУ **URL:**
2. **Браилов, А.В. .** Теория вероятностей и математическая статистика : учебник-практикум / А. В. Браилов. - Ижевск : Регулярная и хаотическая динамика, Институт компьютерных исследований, 2016. - 414 с. - ISBN 978-5-4344-0415-0 **Местонахождение:** PRbooks **URL:** http://www.iprbookshop.ru/69368.html
3. **Вентцель, Елена Сергеевна.** Теория вероятностей и её инженерные приложения : учеб. пособие / Вентцель, Елена Сергеевна, Л. А. Овчаров. - 4-е изд., **стер. - М. : Высш. шк., 2007. - 491 с. : ил. - Рекомендовано МО РФ. - ISBN 978-5-06-005714-0 : 330-00. Местонахождение: Научная библиотека ДГУ URL:**
4. **Калинина, Вера Николаевна.** Теория вероятностей и математическая статистика : учеб. для бакалавров (компьютерно-ориентированный курс) / Калинина, Вера Николаевна. - 2-е изд., перераб. и доп. - М. : Юрайт, 2013. - 450-01. **Местонахождение:** Научная библиотека ДГУ **URL:**
5. Маталыцкий М.А. Теория вероятностей, математическая статистика и случайные процессы [Электронный ресурс]: учебное пособие/ Маталыцкий М.А., Хацкевич Г. А.— Электрон. текстовые данные.— Минск: Вышэйшая школа, 2012.— 720 с.— Режим доступа: http://www.iprbookshop.ru/20289.html (01.09.2018)
6. **Михайлов, Геннадий Алексеевич.** Численное статистическое моделирование: методы Монте-Карло : учеб. пособие / Михайлов, Геннадий Алексеевич, А. В. Войтишек. - Допущено МО РФ. - М. : Академия, 2006. - 367

с. - (Университетский учебник. Прикладная математика и информатика). - ISBN 5-7695-2739-0 : 286-00. **Местонахождение:** Научная библиотека ДГУ **URL:**

7. **Семенчин, Евгений Андреевич.** Теория вероятностей в примерах и задачах : учеб. пособие для вузов / Семенчин, Евгений Андреевич. - СПб. : Лань, 2007. - 351 с. : ил. - (Учебники для вузов. Специальная литература). - Рекомендовано УМО РФ. - ISBN 978-5-8114-0648-7 : 242-00. **Местонахождение:** Научная библиотека ДГУ **URL**
8. Тарасов В.Н. Теория вероятностей, математическая статистика и случайные процессы [Электронный ресурс]: учебное пособие/ Тарасов В.Н., Бахарева Н.Ф.— Электрон. текстовые данные.— Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017.— 283 с.— Режим доступа: http://www.iprbookshop.ru/71890.html (01.02.2022)

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

1)Bibliclub [электронный ресурс]: электронная библиотека https://biblioclub.ru/index.php?page=main_ub_re

2)eLIBRARY.RU[Электронный ресурс]: электронная библиотека / Науч. электрон. б-ка. – – Москва, 1999 – . Режим доступа: http://elibrary.ru/defaultx.asp (дата обращения: ____).

3) Электронный каталог НБ ДГУ[Электронный ресурс]: база данных содержит сведения о всех видах лит, поступающих в фонд НБ ДГУ/Дагестанский гос. ун-т. – Махачкала, 2010 – Режим доступа: http://elib.dgu.ru, свободный (дата обращения: __________).

10. Методические указания для обучающихся по освоению дисциплины.

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

1. **MS Word, MS PowerPoint, MS Excel. Пакет офисных приложений OfficeStd 2016 RUSOLPNLAcdmc**, Контракт №219-ОА от 19.12.2016 г. с ООО «Фирма АС».

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

г. Махачкала, ул. Батырая 2/12, № 405 - учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	- количество посадочных мест - 64 ; - проектор BenqMP 730; - Экран для проектора Draper STAR; - меловая; - стол преподавателя – 1 шт.; - кафедра – 1шт.; - выход в интернет.
г. Махачкала, ул. Батырая 2/12, № 411 - учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	- количество посадочных мест - 30 ; - проектор BenQ MX661; - экран ScreenMedia 200*200; - меловая и маркерная доска; - стол преподавателя – 1 шт.; - выход в интернет.