Министерство науки и высшего образования Российской Федерации
Набережночелнинский институт (филиал)
федерального государственного автономного
образовательного учреждения высшего образования
«Казанский (Приволжский) федеральный университет»

кафедра математики

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие
для самостоятельной работы студентов заочной формы обучения
по инженерно-техническим направлениям подготовки бакалавров

г. Набережные Челны
2025

Теория вероятностей и математическая статистика. *Учебно-методическое пособие для самостоятельной работы студентов заочной формы обучения по инженерно-техническим направлениям подготовки бакалавров.* /Составители: ***Зайцева Ж.И., Углов А.Н.*** - Набережные Челны: Изд-во: НЧИ К(П)ФУ, 2025, 71 с.

1.Цель и задачи дисциплины, её место в учебном процессе.

Цель преподавания дисциплины «Теория вероятностей и математическая статистика» - формирование системы базовых знаний по данной дисциплине, которая позволит будущим специалистам решать в своей повседневной деятельности актуальные задачи практики, понимать написанные на современном научном уровне результаты других исследований и тем самым совершенствовать свои профессиональные навыки.

Основными **задачами дисциплины** являются**:**
- ознакомление студентов с ролью математики в современной жизни, с характерными чертами математического метода изучения реальных задач;
- обучение студентов теоретическим основам курса;
- привитие практических навыков математического моделирования реальных социально-экономических задач с использованием математического аппарата данного курса;
- развитие у студентов навыков творческого и логического мышления, повышение общего уровня математической культуры.

Данная дисциплина является основой при изучении дисциплин, использующих современные математические методы. В свою очередь, для изучения данной дисциплины необходимо знание элементарной математики, линейной алгебры и математического анализа.

В результате изучения данной дисциплины студент должен:
- знать теоретические основы теории вероятностей и математической статистики;
- **уметь** использовать полученные знания для решения практических задач.

Изучение дисциплины предусматривает проведение лекционных, практических занятий и самостоятельную работу студентов. В лекциях излагается содержание тем программы с учётом требований, установленных для специалиста в квалификационной характеристике. Практические занятия проводятся с целью закрепления теоретических основ курса, получения практических навыков решения математических задач. Контроль знаний осуществляется с помощью контрольной работы и итогового экзамена (зачёта) в конце семестра обучения.

2. Содержание и структура дисциплины.

2.1 Содержание дисциплины (наименование тем).

Раздел. Теория вероятностей.

Тема. Случайные события и их вероятности.

Предмет теории вероятностей. Случайный эксперимент. Случайные события, действия над ними. Классическое, геометрическое, статистическое определения вероятности. Свойства вероятности. Правила и формулы комбинаторики, вычисление вероятностей с их помощью. Условная вероятность события. Формулы сложения и умножения вероятностей, полной вероятности и Байеса. Повторные испытания. Схема и формула Бернулли.

Тема. Случайные величины. Системы случайных величин.

Понятие случайной величины (СВ). Функция распределения вероятностей. Дискретная и непрерывная СВ, способы их задания. Функция плотности распределения вероятностей. Числовые характеристики СВ. Основные законы распределения СВ: биномиальный, Пуассона, равномерный, показательный и нормальный. Понятие многомерной СВ. Дискретная двумерная СВ: таблица распределения вероятностей, законы распределения составляющих, числовые характеристики. Коэффициент корреляции. Коррелированность и зависимость СВ. Функция регрессии.

Тема 3. Предельные теоремы теории вероятностей.

Неравенство Чебышева. Законы больших чисел в форме Чебышева и Бернулли. Понятие о центральной предельной теореме.

Раздел. Математическая статистика.

Тема 4. Основные понятия и задачи математической статистики. Предварительная обработка экспериментальных данных.

Предмет и основные задачи математической статистики. Генеральная совокупность и выборка из неё. Выборочный метод. Повторная и бесповторная выборки. Понятия случайной выборки, выборочного пространства, выборочной характеристики (статистики). Основные способы записи выборки: вариационный ряд; статистический дискретный и интервальный ряды. Числовые характеристики выборки и её графическое представление.

Тема 5. Статистические методы оценивания параметров распределений, проверки гипотез и исследования зависимостей.

Понятие точечной и интервальной оценок параметров распределения. Основные методы построения оценок. Точечные и интервальные оценки параметров нормального и биномиального распределений. Принцип практической уверенности. Понятие статистической гипотезы. Статистический критерий и критическое множество. Статистика критерия. Ошибки 1-го и 2-го рода, допускаемые при принятии гипотез. Общая схема критерия проверки гипотез.

Проверка гипотез о параметрах нормального и биномиального распределений. Критерий «хи-квадрат» и его применение для проверки гипотезы о виде распределения генеральной совокупности. Понятия функциональной, стохастической и корреляционной зависимостей. Парный линейный корреляционно-регрессионный анализ, его задачи и проведение. Корреляционное поле. Коэффициент корреляции, его оценивание по выборке и проверка значимости. Уравнение парной линейной регрессии, оценивание по выборке его параметров методом наименьших квадратов и проверка значимости.

2.2. Практические занятия, их содержание.

Тема. Случайные события и их вероятности.

Вычисление вероятностей случайных событий с помощью классического определения, формул сложения и умножения, формулы Бернулли, формул полной вероятности и Байеса.

Тема. Случайные величины.

Нахождение законов распределения дискретных и непрерывных случайных величин и их числовых характеристик.

Тема. Предварительная обработка статистических данных. Оценка параметров, проверка гипотез, исследования взаимосвязей.

Построение статистических рядов, вычисление их числовых характеристик и графическое представление. Построение точечных и интервальных оценок. Проведение парного линейного корреляционно-регрессионного анализа.

2.3. Виды самостоятельной работы студентов.

Самостоятельная работа студентов предполагает изучение теоретического материала и выполнение контрольной работы.

3. Рекомендуемая литература.

Основная литература:

1. Бочаров П.П., Печинкин А.В. Теория вероятностей. Математическая статистика. –М.: ФИЗМАТЛИТ, 2005. -296с. –ISBN: 978-5-9221-0633-3. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=2115.
2. Белько И.В., Свирид Г.П. Теория вероятностей и математическая статистика: примеры и задачи. Учеб. пособие для вузов. –Минск: Новое знание, 2004. -251с.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. Учебник для вузов. –М.: Высшая школа, 2005. -479с.
4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Учебник для вузов. –М.: Высшая школа, 2005. -400с.
5. Кремер Н.Ш. Теория вероятностей и математическая статистика. Учебник для вузов. –М.: ЮНИТИ-ДАНА, 2004. -543с.

Дополнительная литература:

1. Горлач Б.А. Теория вероятностей и математическая статистика: Учеб. пособие. –СПб.: Изд-во «Лань», 2013. -320с. ISBN 978-5-8114-1429-1. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=4864.
2. Емельянов Г.В., Скитович В.П. Задачник по теории вероятностей и математической статистике: Учеб. пособие. –СПб.: Изд-во «Лань», 2007. -336с. ISBN 978-5-8114-0743-9. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=141.
3. Мхитарян В.С., Астафьева Е.В., Миронкина Ю.Н., Трошин Л.И. Теория вероятностей и математическая статистика: Учеб. пособие /Под ред. В.С. Мхитаряна. –М.: Московский финансово-промышленный университет «Синергия», 2013. -336с. ISBN 978-5-4257-0106-0. Режим доступа: http://znanium.com/catalog.php?item=bookinfo&book=451329.
4. Сборник задач по математике для вузов. Учеб. пособие для студентов вузов. /Абрамова В.В., Бикчурина Л.Ж., Валеева М.И. и др.; под ред. Котляра Л.М., Углова А.Н.; 5-е изд., перераб. и доп. -Наб. Челны: ИНЭКА, 2006. – 472с. (Гриф Министерства образования и науки РФ)
5. Шапкин А.С., Шапкин В.А. Задачи с решениями по высшей математике, теории вероятностей, математической статистике, математическому программированию: Учебное пособие для бакалавров. –М.: Изд-во «Дашков и К», 2013. -432с. ISBN 978-5-394-01943-2. Режим доступа: http://znanium.com/catalog.php?item=bookinfo&book=430613.

4. Методические указания по изучению дисциплины.

В процессе изучения данной дисциплины студенты должны сначала изучить теоретический материал и выработать навыки решения типовых задач, используя рекомендованную литературу, а затем выполнить по одной контрольной работе в каждом из семестров обучения (задания для контрольной работы приведены в разделе **5.1**).

При выполнении контрольной работы необходимо придерживаться указанных ниже правил:

1. Контрольная работа должна быть выполнена студентом в отдельной ученической тетради с полями не менее 3 см для замечаний преподавателя.

2. На обложке тетради указываются: название дисциплины; номер варианта и номера решаемых задач; Ф.И.О. студента, выполнившего работу, его номер группы и номер зачетной книжки; Ф.И.О. преподавателя, проверяющего работу (образец оформления обложки приведён в *Приложении 6.4*).

3. Номер варианта соответствует номеру студента в списке группы.

4. Номера решаемых задач выбираются из *ТАБЛИЦЫ НОМЕРОВ ВЫПОЛНЯЕМЫХ ЗАДАНИЙ (Приложение 6.5).*

5. Условия задач переписываются полностью, без сокращения слов, после чего приводится их подробное решение (чертежи можно выполнять аккуратно от руки). В конце решения приводится ответ.

6. В работу должны быть включены все задачи, указанные в задании, строго по порядку номеров. Контрольные работы, содержащие не все задания, а также задачи не своего варианта, не зачитываются.

7. Если в работе имеются ошибки, студент должен выполнить все требования преподавателя, изложенные в рецензии, и сдать работу с исправлениями на повторную проверку.

8. Никакие исправления в тексте уже проверенной работы не допускаются, все исправления записываются после рецензии преподавателя с указанием номера задачи, к которой относятся дополнения и исправления.

9. Работа может быть выполнена заново в случае выявления серьёзных замечаний и ошибок.

10. В конце тетради рекомендуется оставлять несколько чистых страниц для дополнений и исправлений.

После проверки контрольная работа предъявляется к защите. На защите студент должен показать свое умение решать задачи, подобные тем, что имеются в его контрольной работе.

Образец решения типового варианта контрольной работы приведён в *Приложении 6.1*.

5. Материалы для контроля знаний студентов.

Итоговой формой контроля знаний является экзамен (зачёт) в конце первого семестра обучения. На экзамене (зачёте) студент должен показать знание теоретических основ курса в объёме вопросов, приведённых в разделе **5.2** и умение решать задачи, подобные тем, что имеются в его контрольной работе.

5.1. Задания для контрольной работы.

Раздел. Теория вероятностей.

1-10. Требуется найти вероятности указанных событий, используя ***классическое определение*** вероятности:

1. Бросаются два игральных кубика. Найти вероятность того, что число выпавших очков на одном из кубиков в два раза больше, чем на другом.
2. В первом ящике находятся шары с номерами от 1 до 5, а во втором – с номерами от 6 до 10. Из каждого ящика вынимают по одному шару. Найти вероятность того, что сумма номеров вынутых шаров равна десяти.
3. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 35. Найти вероятность того, что номер первого наудачу извлеченного жетона содержит цифру 4.
4. Бросаются два игральных кубика. Найти вероятность того, что сумма числа очков на выпавших гранях равна восьми.
5. В первом ящике находятся шары с номерами от 1 до 5, а во втором – с номерами от 6 до 10. Из каждого ящика вынимают по одному шару. Найти вероятность того, что произведение номеров вынутых шаров больше пятнадцати.
6. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 35. Найти вероятность того, что номер первого наудачу извлеченного жетона не содержит цифру 3.
7. Бросаются два игральных кубика. Найти вероятность того, что сумма числа очков на выпавших гранях больше четырёх, но меньше семи.
8. В первом ящике находятся шары с номерами от 1 до 5, а во втором – с номерами от 6 до 10. Из каждого ящика вынимают по одному шару. Найти вероятность того, что абсолютная величина разности номеров вынутых шаров равна двум.
9. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 35. Найти вероятность того, что номер первого наудачу извлеченного жетона является числом, кратным 5 (делится на 5 без остатка).
10. Бросаются два игральных кубика. Найти вероятность того, что сумма числа очков на выпавших гранях меньше их произведения.

11-20. Требуется найти вероятности указанных событий, используя ***<u>классическое определение</u>*** вероятности:

11. В магазине выставлены для продажи 20 изделий, среди которых 6 изделий некачественные. Найти вероятность того, что взятые случайным образом 2 изделия будут некачественными.

12. В партии из 20 изделий 4 изделия имеют скрытый дефект. Найти вероятность того, что из взятых наугад 5 изделий дефектными являются 2 изделия.

13. На ТЭЦ 15 сменных инженеров, из них 3 женщины. В смену занято 3 человека. Найти вероятность того, что в случайно выбранную смену окажется две женщины.

14. Из партии, содержащей 10 изделий, среди которых 3 бракованные, наудачу извлекают 3 изделия для контроля. Найти вероятность того, что в выборке содержится одно бракованное изделие.

15. В магазине имеются 20 холодильников, причем 15 из них – импортные. Найти вероятность того, что среди 5 проданных в течение дня холодильников окажется три импортных холодильника.

16. Среди 25 студентов, из которых 15 девушек, разыгрываются 4 билета в театр, причем каждый может выиграть только один билет. Найти вероятность того, что среди обладателей билетов окажутся два юноши и две девушки.

17. В магазине выставлены для продажи 18 изделий, среди которых 8 изделий некачественные. Найти вероятность того, что взятые случайным образом 3 изделия будут некачественными.

18. В партии из 30 изделий 5 изделий имеют скрытый дефект. Найти вероятность того, что из взятых наугад 5 изделий дефектными являются 3 изделия.

19. На ТЭЦ 15 сменных инженеров, из них 3 женщины. В смену занято 3 человека. Найти вероятность того, что в случайно выбранную смену окажется хотя бы один мужчина.

20. Среди 25 студентов, из которых 15 девушек, разыгрываются 4 билета в театр, причем каждый может выиграть только один билет. Какова вероятность того, что среди обладателей билетов окажется хотя бы одна девушка.

21-30. Требуется найти вероятности указанных событий, используя ***<u>формулы сложения и умножения вероятностей.</u>***

21. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом справочнике 0.8, во втором – 0.7, в третьем – 0.85. Найти вероятность того, что формула содержится хотя бы в одном справочнике.

22. В первом ящике 5 белых и 7 чёрных шаров. Во втором 3 белых и 12 чёрных шаров. Из каждого ящика вынули по шару. Какова вероятность, что оба шара разного цвета.
23. Рабочий обслуживает три станка. Вероятность того, что в течение часа станок не потребует внимания рабочего, равна для первого станка 0.9, для второго – 0.8 и для третьего – 0.85. Найти вероятность того, что в течение некоторого часа два станка из трёх потребуют внимания рабочего.
24. Три стрелка стреляют по разу в одну мишень независимо друг от друга. Вероятность попадания в цель первым стрелком равна 0.7, вторым – 0.8, третьим – 0.9. Найти вероятность того, что мишень будет поражена два раза.
25. Экзаменационный билет по теории вероятностей содержит три вопроса. Вероятности того, что студент ответит на первый и второй вопросы билета равны 0.9; на третий – 0.8. Найти вероятность того, что студент сдаст экзамен, если для этого необходимо ответить на все вопросы.
26. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом справочнике 0.8, во втором – 0.7, в третьем – 0.85. Найти вероятность того, что формула содержится только в одном справочнике.
27. В первом ящике 6 белых и 4 чёрных шара, во втором -7 белых и 3 чёрных. Из каждого ящика наугад вынимают по одному шару. Чему равна вероятность того, что вынутые шары одного цвета.
28. Рабочий обслуживает три станка. Вероятность того, что в течение часа станок не потребует внимания рабочего, равна для первого станка 0.9, для второго – 0.8 и для третьего – 0.85. Найти вероятность того, что в течение некоторого часа все три станка потребуют внимания рабочего.
29. Три стрелка стреляют по разу в одну мишень независимо друг от друга. Вероятность попадания в цель первым стрелком равна 0.7, вторым – 0.8, третьим – 0.9. Найти вероятность того, что мишень будет поражена хотя бы один раз.
30. Экзаменационный билет по теории вероятностей содержит три вопроса. Вероятности того, что студент ответит на первый и второй вопросы билета равны 0.9; на третий – 0.8. Найти вероятность того, что студент сдаст экзамен, если для этого необходимо ответить хотя бы на два вопроса из трёх.

31-40. Требуется найти вероятности указанных событий, используя ***формулы полной вероятности и Байеса.***
31. В магазин поступает продукция трёх фабрик. Причём продукция первой фабрики составляет 20%, второй -45% и третьей -35% изделий. Известно, что средний процент нестандартных изделий для первой фабрики равен 3%, для второй -2%, и для третьей -4%. Найти вероятность того, что оказавшееся нестандартным изделие произведено на первой фабрике.

32. На сборочное предприятие поступили однотипные комплектующие с трёх заводов в количестве: 25 - с первого завода, 35 – со второго, 40 - с третьего. Вероятность качественного изготовления изделий на первом заводе 0.9, на втором – 0.8, на третьем – 0.7. Найти вероятность того, что взятое случайным образом изделие будет качественным.
33. На двух станках производятся одинаковые детали. Вероятность того, что деталь стандартная, для первого станка равна 0.8, для второго – 0.9. Производительность второго станка вдвое больше, чем первого. Найти вероятность того, что взятая наудачу деталь окажется нестандартной.
34. На сборочное предприятие поступили однотипные комплектующие с трёх заводов в количестве: 40 - с первого завода, 35 – со второго, 25 - с третьего. Вероятность качественного изготовления изделий на первом заводе 0.9, на втором – 0.7, на третьем – 0.9. Найти вероятность того, что взятое случайным образом изделие будет качественным.
35. Известно, что 90% выпускаемой заводом продукции удовлетворяет стандарту. Упрощенная схема контроля признает пригодной стандартную продукцию с вероятностью 0.95 и признает пригодной нестандартную продукцию с вероятностью 0.1. Найти вероятность того, что изделие, не прошедшее упрощенный контроль, удовлетворяет стандарту.
36. Среди поступающих на сборку деталей с первого станка-автомата 1% нестандартных, со второго – 2%, с третьего – 2.5%, с четвертого – 5%. Производительности их относятся как 4:3:2:1. Взятая наудачу деталь оказалась стандартной. Найти вероятность того, что деталь изготовлена на третьем станке-автомате.
37. В магазин поступает продукция трёх фабрик. Причём продукция первой фабрики составляет 20%, второй -45% и третьей -35% изделий. Известно, что средний процент нестандартных изделий для первой фабрики равен 3%, для второй -2%, и для третьей -4%. Найти вероятность того, что оказавшееся нестандартным изделие произведено на второй фабрике.
38. На сборочное предприятие поступили однотипные комплектующие с трёх заводов в количестве: 25 - с первого завода, 10 – со второго, 15 - с третьего. Вероятность качественного изготовления изделий на первом заводе 0.7, на втором – 0.9, на третьем – 0.8. Найти вероятность того, что взятое случайным образом изделие будет качественным.
39. На двух станках производятся одинаковые детали. Вероятность того, что деталь стандартная, для первого станка равна 0.8, для второго – 0.9. Производительность второго станка втрое больше, чем первого. Найти вероятность того, что взятая наудачу деталь окажется стандартной.
40. Трое рабочих изготавливают однотипные изделия. Первый рабочий изготовил 40 изделий, второй – 35, третий -25. Вероятность брака у первого рабочего 0.03, у второго – 0.02, у третьего – 0.01. Взятое наугад изделие оказалось

бракованным. Найти вероятность того, что это изделие изготовил второй рабочий.

41-50. Требуется найти вероятности указанных событий, используя ***формулу Бернулли***.

41. Экзамен состоит из 6 вопросов. На каждый вопрос дано четыре возможных ответа, среди которых необходимо выбрать один правильный. Найти вероятность того, что методом простого угадывания удастся ответить не менее чем на 5 вопросов.

42. Покупатель приобрел шесть изделий, изготовленных на данном предприятии, 80% изделий которого составляет продукция высшего сорта. Найти вероятность того, что не менее пяти из них являются изделиями высшего сорта.

43. Для нормальной работы автобазы на линии должно быть не менее 8 машин из 10 имеющихся. Вероятность невыхода каждой машины на линию равна 0.1. Найти вероятность того, что автобаза будет работать нормально в ближайший день.

44. В урне 20 белых и 10 черных шаров. Из урны вынимают подряд 4 шара, причем каждый вынутый шар возвращают в урну перед извлечением следующего и шары в урне перемешивают. Найти вероятность того, что из четырех вынутых шаров окажется три белых.

45. Для стрелка, выполняющего упражнение в тире, вероятность попасть в «яблочко» при одном выстреле не зависит от результатов предшествующих выстрелов и равна 0.25. Спортсмен сделал 5 выстрелов. Найти вероятность того, что стрелок попал в «яблочко» не менее четырёх раз.

46. Найти вероятность того, что в семье, имеющей 6 детей, не менее двух девочек. Предполагается, что вероятности рождения мальчика и девочки одинаковые.

47. Экзамен состоит из 5 вопросов. На каждый вопрос дано три возможных ответа, среди которых необходимо выбрать один правильный. Найти вероятность того, что методом простого угадывания удастся ответить на все вопросы.

48. Покупатель приобрел пять изделий, изготовленных на данном предприятии, 80% изделий которого составляет продукция высшего сорта. Найти вероятность того, что четыре из них являются изделиями высшего сорта.

49. Для нормальной работы автобазы на линии должно быть не менее 9 машин из 10 имеющихся. Вероятность невыхода каждой машины на линию равна 0.2. Найти вероятность того, что автобаза будет работать нормально в ближайший день.

50. Вероятность выиграть по лотерейному билету равна $1/7$. Найти вероятность выиграть не менее чем по двум билетам из шести.

51-60. Требуется составить закон распределения дискретной случайной величины X; построить многоугольник полученного распределения; вычислить её математическое ожидание MX и дисперсию DX.

51. В партии из 10 деталей содержится 3 нестандартных. Наудачу отобраны две детали. Дискретная случайная величина X – число нестандартных деталей среди двух отобранных;
52. Стрелок производит по мишени два выстрела. Вероятность попадания в мишень при каждом выстреле равна 0.9. Дискретная случайная величина X – число попаданий в мишень.
53. В экзаменационном билете две задачи. Вероятность правильного решения первой задачи равна 0.9, второй – 0.6. Дискретная случайная величина X – число правильно решённых задач в билете.
54. В партии из 6 деталей имеется 4 стандартных. Наудачу отобраны 2 детали. Дискретная случайная величина X – число нестандартных деталей среди отобранных.
55. Стрелок производит по мишени два выстрела. Вероятность попадания в мишень при каждом выстреле равна 0.7. Дискретная случайная величина X – число промахов.
56. Из пяти купленных роз 2 красные. Для составления букета наудачу берут 3 розы. Дискретная случайная величина X – число красных роз среди отобранных.
57. В партии из 6 деталей имеется 4 стандартных. Наудачу отобраны 2 детали. Дискретная случайная величина X – число стандартных деталей среди отобранных.
58. Устройство состоит из двух независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0.1. Дискретная случайная величина X – число отказавших элементов в одном опыте.
59. В урне 6 белых и 4 чёрных шаров. Из неё извлекли два шара. Дискретная случайная величина X – число белых шаров среди отобранных.
60. В урне 6 белых и 4 чёрных шаров. Из неё извлекли два шара. Дискретная случайная величина X – число чёрных шаров среди отобранных.

61-70. Требуется найти функцию плотности вероятностей $f(x)$ непрерывной случайной величины X; вычислить её математическое ожидание MX, дисперсию DX и вероятность $P(X \in (\alpha, \beta))$.

61. Непрерывная случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 & x \le 0 \\ x^2/25 & 0 < x \le 5 \\ 1 & x > 5 \end{cases}, \quad \alpha = 1, \beta = 4.$$

62. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le 0\\ x^2 & 0<x\le 1\\ 1 & x>1\end{cases}$, $\alpha=\dfrac{1}{4}, \beta=\dfrac{3}{4}$.

63. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le -1\\ \dfrac{(x+1)^2}{9} & -1<x\le 2\\ 1 & x>2\end{cases}$, $\alpha=0, \beta=1$.

64. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le 1\\ (x-1)/5 & 1<x\le 6\\ 1 & x>6\end{cases}$, $\alpha=2, \beta=5$.

65. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le 0\\ x/4 & 0<x\le 4\\ 1 & x>4\end{cases}$, $\alpha=2, \beta=3$.

66. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le 0\\ x^2/4 & 0<x\le 2\\ 1 & x>2\end{cases}$, $\alpha=1, \beta=2$.

67. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le 0\\ \dfrac{8x^2-x^4}{16} & 0<x\le 2\\ 1 & x>2\end{cases}$, $\alpha=1, \beta=2$.

68. Непрерывная случайная величина X задана функцией распределения $F(x)=\begin{cases}0 & x\le -2\\ \dfrac{x}{4}+\dfrac{1}{2} & -2<x\le 2\\ 1 & x>2\end{cases}$, $\alpha=-1, \beta=1$.

69. Непрерывная случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 & x\le 3 \\ \dfrac{(x-3)^2}{4} & 3<x\le 5, \\ 1 & x>5 \end{cases}\quad \alpha=3,\ \beta=4.$$

70. Непрерывная случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 & x\le 2 \\ x^2-4x+4 & 2<x\le 3, \\ 1 & x>3 \end{cases}\quad \alpha=2,\ \beta=\frac{5}{2}.$$

Раздел. <u>Математическая статистика.</u>

71-80. Для приведённых выборок требуется: построить вариационный и дискретный статистический ряды; вычислить числовые характеристики выборки: $x_{\min}$, $x_{\max}$, $\widehat{R}$ (размах), $\overline{x}$ (среднее арифметическое), $\widehat{\sigma}^2$ (дисперсию); построить полигон частот.

71. Выборка объема $n=20$:

0	**4**	**2**	**0**	**5**	**1**	**1**	**3**	**0**	**2**	**2**	**4**	**3**	**2**	**3**	**3**	**0**	**4**	**5**	**1**

72. Выборка объема $n=15$:

4	**5**	**6**	**4**	**4**	**6**	**2**	**2**	**5**	**4**	**5**	**5**	**4**	**2**	**3**

73. Выборка объема $n=15$:

2	**2**	**1**	**3**	**4**	**2**	**1**	**1**	**3**	**3**	**4**	**3**	**2**	**4**	**5**

74. Выборка объема $n=15$:

9	**8**	**10**	**6**	**6**	**7**	**9**	**9**	**10**	**4**	**10**	**11**	**11**	**11**	**6**

75. Выборка объема $n=20$:

5	**5**	**4**	**2**	**6**	**2**	**1**	**5**	**3**	**3**	**1**	**5**	**6**	**4**	**3**	**3**	**4**	**1**	**5**	**5**

76. Выборка объема $n=20$:

2	**1**	**2**	**3**	**1**	**1**	**0**	**2**	**2**	**4**	**3**	**3**	**0**	**3**	**0**	**2**	**3**	**1**	**2**	**2**

77. Выборка объема $n=15$:

6	**8**	**9**	**9**	**4**	**9**	**6**	**8**	**9**	**4**	**6**	**6**	**8**	**6**	**4**

78. Выборка объема $n=15$:

5	**10**	**10**	**9**	**5**	**8**	**8**	**9**	**6**	**6**	**6**	**8**	**8**	**10**	**8**

79. Выборка объема $n=15$:

1	**0**	**2**	**6**	**5**	**4**	**1**	**4**	**5**	**1**	**2**	**4**	**2**	**2**	**2**

80. Выборка объема $n=20$:

1	**3**	**3**	**2**	**0**	**2**	**4**	**3**	**2**	**1**	**2**	**2**	**2**	**2**	**3**	**3**	**1**	**1**	**1**	**3**

81-90. Для приведённых выборок требуется: вычислить числовые характеристики группированной выборки: $x_{\min}$, $x_{\max}$, $\widehat{R}$ (размах), $\bar{x}$ (среднее арифметическое), $\widehat{\sigma}^2$ (дисперсию); построить гистограмму частот; найти 95 %-ный доверительный интервал для генеральной средней MX (предполагая нормальным закон распределения генеральной совокупности X, из которой получена выборка).

81. Выборочные данные о месячной заработной плате X по цеху (в тыс.руб.)

Зар.плата	**2-4**	**4-6**	**6-8**	**8-10**	**10-12**	**12-14**	**14-16**	**16-18**
Число рабочих	**1**	**3**	**10**	**15**	**20**	**12**	**7**	**2**

82. Выборочные данные о расходе X фирм, продающих компьютеры, на рекламу (в % к общему расходу фирмы):

Расход на рекламу	**2-4**	**4-6**	**6-8**	**8-10**	**10-12**
Число фирм	**5**	**8**	**16**	**12**	**9**

83. Выборочные данные о стаже работы X сотрудников банка:

Стаж, лет	**0-2**	**2-4**	**4-6**	**6-8**	**8-10**	**10-12**
Число сотрудников	**3**	**9**	**18**	**14**	**10**	**6**

84. Выборочные данные о годовом товарообороте X (млн.руб) продовольственных магазинов города

Товарооборот	**10-12**	**12-14**	**14-16**	**16-18**	**18-20**
Число магазинов	**17**	**40**	**32**	**8**	**3**

85. Выборочные данные о месячной заработной плате X по цеху (в тыс.руб.)

Зар.плата	**2-4**	**4-6**	**6-8**	**8-10**	**10-12**
Число рабочих	**5**	**8**	**16**	**12**	**9**

86. Данные измерений роста X (в см) студентов одного из вузов города:

Рост	**150-160**	**160-170**	**170-180**	**180-190**	**190-200**
Число студентов	**10**	**34**	**25**	**21**	**10**

87. Выборочные данные о годовом товарообороте X (млн.руб) мебельных магазинов города

Товарооборот	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Число магазинов	**12**	**17**	**46**	**12**	**13**

88. Данные измерений внутреннего диаметра X (в мкм) поршневых колец:

Диаметр	**28-32**	**32-36**	**36-40**	**40-44**	**44-48**
Число колец	**5**	**16**	**11**	**8**	**10**

89. Выборочные данные о расходе X фирм, продающих автомобили, на рекламу (в % к общему расходу фирмы):

Расход на рекламу	**3-5**	**5-7**	**7-9**	**9-11**	**11-13**
Число фирм	**1**	**6**	**14**	**7**	**2**

90. Данные о содержании X (в куб. м) деловой древесины в одном дереве:

Содержание деловой древесины	**0.2-0.6**	**0.6-1.0**	**1.0-1.4**	**1.4-1.8**
Число деревьев	**2**	**14**	**15**	**9**

91-100. Для приведенных выборок (предполагается, что выборки получены из двумерных нормально распределённых генеральных совокупностей):

а) Построить диаграмму рассеивания (корреляционное поле).

б) Вычислить выборочный коэффициент линейной корреляции r; проверить его значимость на уровне $\alpha = 0.05$, сделать выводы (для значимого r) о тесноте и направлении связи между величинами X и Y .

в) Найти выборочное уравнение прямой регрессии Y на X и построить ее график на одном чертеже с диаграммой рассеивания.

г) Вычислить, используя регрессионную зависимость Y на X , ожидаемое среднее значение $\bar{y}_{x_0}$ величины Y при $X = x_0$.

91. *Результаты измерений (в метрах) уровней X и Y воды в реке соответственно в пунктах A и B (пункт B находится на 50 км ниже по течению пункта A) в первые 10 дней апреля:*

Y	**12.1**	**11.2**	**9.8**	**10.4**	**9.2**	**8.5**	**8.8**	**7.4**	**6.6**	**7.0**
X	**10.5**	**9.3**	**8.3**	**9.6**	**8.6**	**7.1**	**6.9**	**5.8**	**5.2**	**5**

$x_0 = 12$.

92. *Результаты измерений роста X (в см) и веса Y (в кг) 10 случайно выбранных студентов-первокурсников:*

X	**164**	**168**	**170**	**172**	**174**	**180**	**182**	**183**	**183**	**184**
Y	**53**	**62**	**71**	**76**	**67**	**76**	**86**	**93**	**82**	**79**

$x_0 = 175$.

93. *Результаты опроса 10 студентов, проведенного с целью выявления зависимости между средним баллом Y по результатам предыдущей сессии и числом часов X (в неделю), затраченных студентом на самоподготовку:*

X	**25**	**22**	**9**	**15**	**15**	**30**	**20**	**30**	**10**	**17**
Y	**4.6**	**4.3**	**3.8**	**3.8**	**4.2**	**4.3**	**3.8**	**4.0**	**3.1**	**3.9**

$x_0 = 35$.

94. *Данные об уровне механизации работ X (в %) и производительности труда Y (в т/ч) для 10 промышленных предприятий города:*

X	30	32	36	40	41	47	54	60	69	76
Y	24	20	28	30	31	33	37	38	45	48

$x_0 = 50$.

95. *Результаты 10 измерений для исследования зависимости выхода продукта Y (в кг/ч) от температуры реакции X (в $°C$):*

X	30	35	40	45	50	55	60	65	70	75
Y	5	15	22	39	53	56	64	79	94	101

$x_0 = 80$.

96. *Данные, собранные ПАТП с целью выявления зависимости между пробегом автобусов X (в тыс. км.) и стоимостью их ежегодного технического обслуживания Y (в тыс. руб.) для 10 случайно отобранных автобусов:*

X	13	16	15	20	19	21	26	24	30	30
Y	6	7	8	9	10	12	12	13	14	16

$x_0 = 10$.

97. *Данные о розничном годовом товарообороте Y (в млн. руб.) и среднесписочном числе работников X (чел.) для 10 магазинов города:*

X	73	85	102	115	122	126	134	147	120	90
Y	50	70	90	110	140	140	170	190	130	100

$x_0 = 110$.

98. *Результаты опроса 10 студентов, проведенного с целью выявления зависимости между средним баллом Y по результатам предыдущей сессии и общим числом X пропущенных студентом за семестр занятий (в часах):*

X	2	8	30	25	12	20	18	18	40	26
Y	4.6	4.3	3.2	3.8	4.2	4.2	3.8	4.0	3.1	3.9

$x_0 = 10$.

99. *Данные о месячной выработке Y (в тыс. руб.) на одного работника торговли и величине товарооборота X (в млн.руб.) для 10 магазинов торговой сети «Эльдорадо»:*

X	80	55	75	110	85	70	40	100	50	95
Y	44	40	37	57	42	38	31	43	38	50

$x_0 = 90$.

100. *Данные о рентабельности Y (в %) производства и дневной производительности X труда рабочего (в тыс. руб.) для 10 предприятий города:*

X	9	7	10	6	5	9	5	6	7	6
Y	15	12	17	7	4	14	6	8	10	9

$x_0 = 8$.

5.2. Вопросы к экзамену (зачёту).

Раздел. <u>Теория вероятностей.</u>

1. Предмет теории вероятностей. Понятия случайного эксперимента, случайного события. Свойство статистической устойчивости исходов случайного эксперимента, пример такого эксперимента.
2. Элементарное событие. Пространство элементарных событий $\boldsymbol{\Omega}$. Случайное событие, как подмножество $\boldsymbol{\Omega}$. Достоверное и невозможное события. Представление событий в виде диаграмм Эйлера-Венна.
3. Действия над случайными событиями, их геометрическая иллюстрация с помощью диаграмм Эйлера-Венна. Совместные и несовместные, противоположные события.
4. Комбинаторика: правила суммы и произведения; сочетания, размещения и перестановки, подсчёт их числа.
5. Равновозможные события. Классическое определение вероятности. Частота, относительная частота появления события. Статистическое определение вероятности.
6. Основные свойства вероятности. Условная вероятность события. Зависимые и независимые события. Формулы сложения и умножения вероятностей (для двух событий).
7. Полная группа событий, гипотезы. Формулы полной вероятности,Байеса.
8. Повторные испытания. Схема Бернулли. Формула Бернулли.
9. Понятие случайной величины (СВ). Функция распределения случайной величины и её основные свойства.
10. Дискретная случайная величина (ДСВ). Ряд распределения, многоугольник распределения, функция распределения ДСВ, их построение.
11. Непрерывная случайная величина (НСВ). Функция плотности распределения, её основные свойства. Представление функции распределения НСВ через функцию плотности распределения.
12. Математическое ожидание дискретной и непрерывной случайной величин. Основные свойства математического ожидания.
13. Дисперсия и среднее квадратичное отклонение случайной величины. Основные свойства дисперсии. Вычисление дисперсии дискретной и непрерывной случайных величин.
14. Начальные и центральные моменты k -ого порядка, взаимосвязь между ними. Асимметрия и эксцесс.
15. Основные законы распределения ДСВ (биномиальный, закон Пуассона), их числовые характеристики (математическое ожидание и дисперсия).
16. Равномерный закон распределения НСВ, его числовые характеристики (математическое ожидание и дисперсия).

17. Показательный закон распределения НСВ, его числовые характеристики (математическое ожидание и дисперсия).
18. Нормальный закон распределения НСВ, его числовые характеристики (математическое ожидание, дисперсия, асимметрия, эксцесс). График функции плотности нормального распределения, его особенности.
19. Стандартный нормальный закон распределения. Интеграл Лапласа и его применение для вычисления вероятности попадания нормально распределённой СВ в заданный интервал. Правило «трёх сигм».
20. Понятие n-мерной случайной величины. Двумерная ДСВ, законы её распределения. Одномерные безусловные и условные законы распределения составляющих. Зависимость и независимость двух дискретных СВ. Ковариация случайных величин. Функция регрессии.
21. Неравенства Чебышева. Законы больших чисел в форме Чебышева и Бернулли, их значение для практики.
22. Центральная предельная теорема ТВ, её значение для практики. Особая роль нормального закона распределения. Интегральная формула Муавра-Лапласа, как следствие центральной предельной теоремы ТВ.

Раздел. <u>Математическая статистика.</u>

23. Предмет математической статистики. Основные задачи математической статистики. Взаимосвязь математической статистики и теории вероятностей.
24. Генеральная совокупность и выборка. Основные способы организации выборки (повторный и бесповторный отбор). Репрезентативность выборки. Случайная выборка. Выборочный метод, как основной метод математической статистики.
25. Вариационный ряд. Медиана и размах выборки, их нахождение для негруппированных и группированных данных.
26. Статистический ряд распределения выборки. Интервальный статистический ряд и его построение. Формула Стерджесса определения числа интервалов группировки. Графическое представление выборки: полигон, гистограмма, их построение.
27. Среднее арифметическое выборки, его свойства и вычисление для негруппированных и группированных данных.
28. Дисперсия $\widehat{\sigma}^2$ выборки, её свойства и вычисление для негруппированных и группированных данных. Исправленная дисперсия s^2 выборки. Взаимосвязь дисперсий $\widehat{\sigma}^2$ и s^2.
29. Понятие точечной оценки неизвестного параметра θ распределения генеральной совокупности. Свойства оценок (несмещённость, состоятельность, эффективность).

30. Точечные оценки основных числовых характеристик (математического ожидания, дисперсии, генеральной доли) генеральной совокупности, их свойства.
31. Метод моментов нахождения точечных оценок. Нахождение методом моментов точечных оценок параметров равномерного закона распределения.
32. Функция правдоподобия. Метод максимального правдоподобия нахождения точечных оценок. Свойства оценок максимального правдоподобия.
33. Нахождение методом максимального правдоподобия оценок параметров a и σ нормального закона распределения.
34. Понятие интервальной оценки (доверительного интервала) неизвестного параметра θ распределения генеральной совокупности. Доверительная вероятность и предельная ошибка выборки. Основные виды доверительных интервалов.
35. Доверительные интервалы: для среднего значения нормального распределения при известной и неизвестной дисперсиях; для генеральной доли (для параметра p биномиального распределения)
36. Определение необходимого объёма повторной выборки при оценке неизвестных параметров распределений генеральной совокупности.
37. Понятие статистической гипотезы. Виды гипотез: основная и альтернативная, простая и сложная. Критерий проверки гипотез и критическое множество. Статистика критерия, её критическое множество. Ошибки первого и второго рода, допускаемые при принятии гипотез.
38. Основные статистические гипотезы о параметрах генеральной совокупности, общая схема их проверки.
39. Критерий «хи-квадрат». Общая схема проверки статистической гипотезы о законе распределения генеральной совокупности.
40. Понятие функциональной, стохастической (*статистической*) и корреляционной зависимости. Основные задачи статистического исследования зависимостей между случайными величинами.
41. Парный корреляционный анализ и его проведение. Корреляционное поле и корреляционная таблица, их построение. Предварительный анализ корреляционной связи.
42. Коэффициент линейной корреляции, его свойства, оценивание по выборке, проверка значимости выборочного коэффициента корреляции.
43. Парный регрессионный анализ, его предположения и проведение. Парная линейная и нелинейная регрессия. Понятие о методе МНК оценивания неизвестных параметров уравнений регрессий.
44. Выборочное уравнение линейной регрессии, нахождение его параметров методом МНК.

6. Приложения.
6.1. Образец решения контрольных задач типового варианта.

Раздел. <u>Теория вероятностей.</u>

1-10. Требуется найти вероятность указанного события, используя ***классическое определение*** вероятности.
Бросаются два игральных кубика. Найти вероятность того, что сумма очков на выпавших гранях – четная, причем на грани хотя бы одного из кубиков появится шестерка.

При классическом определении вероятность случайного события A определяется равенством: $P(A)=\dfrac{m(A)}{n}$, где $m(A)$ - число элементарных (далее неделимых и взаимно исключающих друг друга) исходов эксперимента, благоприятных появлению события A; n - общее число равновозможных элементарных исходов эксперимента. Равновозможность элементарных исходов обеспечивается такими условиями проведения эксперимента (опыта, испытания), при выполнении которых можно считать, что ни один из исходов не является объективно более возможным, чем другие.

Если событие A определяется словами «хотя бы один…», то непосредственное нахождение $P(A)$ по формуле классического определения вероятности приводит обычно к громоздким вычислениям. Проще сначала найти вероятность события $\overline{A}$, противоположного событию A и определяемого словами «ни один…», а затем, используя формулу для вероятностей противоположных событий: $P(A)=1-P(\overline{A})$, вычислить вероятность искомого события.

Для нахождения вероятности события по формуле $P(A)=\dfrac{m(A)}{n}$ необходимо:

1) Рассмотреть событие A, вероятность которого следует найти.
2) Правильно определить, что является в данном испытании элементарным исходом.
3) Найти общее число n элементарных исходов, предварительно выписав их все непосредственно. Если выписать все элементарные исходы не представляется возможным из-за их чрезмерного количества, то при подсчете их числа используют правила и формулы комбинаторики.

4) Установить какое число $m(A)$ элементарных исходов данного испытания благоприятствуют появлению события A.

Решение.

Рассмотрим событие $A=\{$**сумма очков на выпавших гранях – четная, причем на грани хотя бы одного из кубиков появится шестерка**$\}$.

Элементарными исходами данного испытания (подбрасывание двух игральных кубиков) являются всевозможные комбинации очков: $1, 2, 3, 4, 5, 6$, которые могут появиться на верхних гранях двух кубиков.

Общее число элементарных исходов n данного испытания найдём, используя правило умножения комбинаторики.

Пусть α_1, α_2 – действия из некоторого конечного множества действий.

Правило умножения. Если действие α_1 можно выполнить n_1 способами и, после каждого такого выполнения, действие α_2 можно выполнить n_2 способами, то последовательное выполнение пары действий α_1 и α_2 можно осуществить $n=n_1 \cdot n_2$ способами.

На каждом игральном кубике 6 граней, поэтому возможны шесть исходов бросания каждого из них. Если испытание представить в виде последовательно выполняемых подбрасываний кубиков, то первое подбрасывание можно выполнить $n_1=6$ способами, второе подбрасывание - $n_2=6$ способами, тогда последовательно выполняемое подбрасывание двух кубиков можно осуществить $n=n_1 \cdot n_2=6 \cdot 6=36$ способами.

Общее число элементарных исходов n можно найти и, выписав непосредственно все возможные исходы испытания:

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

Теперь найдём число элементарных исходов $m(A)$ данного испытания, благоприятных событию A, выписав их непосредственно. Такими исходами, очевидно, являются: (2,6); (4,6); (6,2); (6,4); (6,6). Их число $m(A)=5$.

Тогда искомая вероятность $P(A)=\dfrac{5}{36} \approx 0.139$.

Ответ: $P(A)=\dfrac{5}{36} \approx 0.139$.

11-20. Требуется найти вероятности указанных событий, используя ***<u>классическое определение</u>*** вероятности:
В урне находятся 5 черных и 6 белых шаров. Случайным образом из урны вынимают 4 шара. Найти вероятности того, что среди вынутых шаров окажутся: «2 белых шара»; «не более одного белого шара»; «хотя бы один белый шар».

Решение.

Рассмотрим события: $A = \{$**среди четырёх вынутых шаров - 2 белых**$\}$, $B = \{$**среди четырёх вынутых шаров – не более одного белого шара**$\}$, $C = \{$**среди четырёх вынутых шаров - хотя бы один белый шар**$\}$.

Элементарными исходами данного испытания (случайное вынимание четырех шаров) являются всевозможные комбинации по 4 шара из находящихся в урне 11 шаров.

Для подсчёта общего числа элементарных исходов n данного испытания и чисел $m(A), m(B), m(C)$ элементарных исходов, благоприятных событиям A, B, C, используем правила и формулы комбинаторики.

Пусть α_1, α_2 – действия из некоторого конечного множества действий.

Правило сложения. Если действие α_1 можно выполнить n_1 способами, действие α_2 - другими n_2 способами, отличными от первых n_1, то выполнение одного из действий: или α_1, или α_2 (но не двух одновременно) можно осуществить $n = n_1 + n_2$ способами.

Сочетаниями из n ***элементов по*** m называются всевозможные комбинации элементов, отличающиеся друг от друга только составом элементов. Они рассматриваются как элементарные исходы эксперимента, состоящего в одновременном неупорядоченном выборе без возвращения m элементов из n различных элементов, а их общее число C_n^m определяется формулой:

$$C_n^m = \frac{n!}{m! \cdot (n-m)!}, \quad \text{где } n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n, \quad 0! = 1.$$

Общее число элементарных исходов n данного испытания, очевидно, равно числу всевозможных неупорядоченных комбинаций по 4 шара из находящихся в урне 11 шаров, т.е. числу сочетаний C_{11}^4. Тогда:

$$n = C_{11}^4 = \frac{11!}{4! \cdot (11-4)!} = \frac{11!}{4! \cdot 7!} = \frac{7! \cdot 8 \cdot 9 \cdot 10 \cdot 11}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 7!} = 330.$$

Подсчитаем теперь число элементарных исходов $m(A), m(B), m(C)$ благоприятных событиям A, B, C, соответственно.

Событие $A=\{$**среди четырёх вынутых шаров - 2 белых**$\}$ означает, что среди вынутых шаров – «2 белых и 2 черных шара». Следовательно, благоприятными событию A являются всевозможные комбинации по 4 шара (два белых и два черных шара) из находящихся в урне 11 шаров. Их число $m(A)$ найдём, используя правило умножения комбинаторики. Представим для этого выбор четырёх шаров в виде двух последовательно выполняемых действий: сначала выбор двух белых шаров из имеющихся в урне 6 белых шаров и затем выбор двух чёрных шаров из имеющихся в урне 5 чёрных шаров. Получим: $m(A)=C_6^2\cdot C_5^2=\dfrac{6!}{2!\cdot 4!}\cdot\dfrac{5!}{2!\cdot 3!}=15\cdot 10=150$. Тогда: $P(A)=\dfrac{150}{330}\approx 0.455$

Событие $B=\{$**среди четырёх вынутых шаров – не более одного белого шара**$\}$ означает, что среди вынутых шаров - или «один белый и три черных шара», или «четыре чёрных шара». Следовательно, благоприятствующими событию B являются всевозможные комбинации по 4 шара (один белый и три черных или четыре черных шара) из находящихся в урне 11 шаров. Их число $m(B)$ найдём, используя правила сложения и умножения комбинаторики. Сначала, используя правило умножения комбинаторики, найдём число способов выбрать один белый и три черных шара. Получим $C_6^1\cdot C_5^3=\dfrac{6!}{1!\cdot 5!}\cdot\dfrac{5!}{3!\cdot 2!}=6\cdot 10=60$. Затем, используя правило умножения комбинаторики, найдём число способов выбрать 4 чёрных шара. Получим $C_5^4=\dfrac{5!}{4!\cdot 1!}=5$. Теперь, используя правило сложения комбинаторики, найдём число $m(B)$ способов выбрать или один белый и три чёрных шара, или четыре чёрных шара. Получим $m(A)=C_6^1\cdot C_5^3+C_5^4=60+5=65$. Тогда $P(B)=\dfrac{65}{330}\approx 0.197$.

Событие $C=\{$**среди четырёх вынутых шаров-хотя бы один белый шар**$\}$ определяется словами «хотя бы один…». Прямое решение задачи, учитывая, что событие C означает, среди вынутых шаров: или «один белый и три черных шара», или «два белых и два черных шара», или «три белых и один черный шар», или «четыре белых шара», приводит к громоздким вычислениям. Поэтому сначала найдём вероятность противоположного события $\overline{C}=\{$**среди вынутых четырёх шаров нет ни одного белого шара, т.е. все шары – чер-**

ные$\}$. Получим $m(\overline{C}) = C_5^4 = 5$, тогда $P(\overline{C}) = \dfrac{m(\overline{C})}{n} = \dfrac{5}{330}$. Затем по формуле $P(C) = 1 - P(\overline{C})$ найдём вероятность искомого события: $P(C) = 1 - \dfrac{5}{330} = \dfrac{325}{330} \approx 0.985$.

Ответ: $P(A) = \dfrac{150}{330} \approx 0.455$; $P(B) = \dfrac{65}{330} \approx 0.197$; $P(C) = \dfrac{325}{330} \approx 0.985$.

21-30. Требуется найти вероятности указанных событий, используя: ***формулы сложения и умножения вероятностей.***
Экзаменационная сессия состоит из трёх экзаменов. Студент оценивает свои шансы успешно сдать экзамены следующим образом: вероятность сдать первый экзамен - 0.8, второй – 0.9, третий – 0.7. Найти вероятности того, что студентом будут успешно сданы: «все три экзамена», «по крайней мере два экзамена», «хотя бы один экзамен». Предполагается, что сдача экзаменов – независимые события.

Сложным называют событие, наблюдаемое в эксперименте и выражаемое через другие наблюдаемые в том же эксперименте события с помощью допустимых алгебраических операций над событиями.

Вероятность осуществления того или иного сложного события вычисляется с помощью *формул умножения вероятностей*:

1) $P(A \cdot B) = P(A) \cdot P(B \mid A)$, $P(A) > 0$;

2) $P(A \cdot B) = P(A) \cdot P(B)$ (для независимых событий)

и *формул сложения вероятностей*:

3) $P(A + B) = P(A) + P(B) - P(A \cdot B)$;

4) $P(A + B) = P(A) + P(B)$ (для несовместных событий).

События A и B называют *несовместными*, если $A \cdot B = \varnothing$. Несовместными событиями являются, например, элементарные исходы эксперимента.

События A и B, называются *независимыми*, если выполняется равенство $P(A \mid B) = P(A)$, в противном случае они называются *зависимыми*. Часто, независимость событий определяется условиями проведения эксперимента.

Для решения задач с использованием формул сложения и умножения вероятностей следует:

1) рассмотреть «сложное» событие, вероятность которого нужно вычислить;

2) выразить «сложное» событие, посредством допустимых алгебраических операций, через наблюдаемые в том же эксперименте «простые» события, вероятности которых известны или легко определяются из условий задачи, например, по формуле классического определения вероятности;
3) вычислить вероятность «сложного» события с помощью формул сложения и умножения вероятностей, учитывая зависимость или независимость, совместность или несовместность составляющих его событий.

Решение.

Рассмотрим «сложные» события: $A = \{$**студент успешно сдаст все три экзамена**$\}$, $B = \{$**студент успешно сдаст по крайней мере два экзамена из трёх**$\}$, $C = \{$**студент успешно сдаст хотя бы один экзамен из трех**$\}$.

Выразим сначала «сложные» события A, B, C через «простые» события: $D_1 = \{$студент успешно сдаст первый экзамен$\}$, $D_2 = \{$студент успешно сдаст второй экзамен$\}$, $D_3 = \{$студент успешно сдаст третий экзамен$\}$, вероятности которых известны и равны: $P(D_1) = 0.8$, $P(D_2) = 0.9$, $P(D_3) = 0.7$. Затем вычислим вероятности $P(A), P(B), P(C)$, используя формулы сложения и умножения вероятностей, учитывая при этом зависимость и независимость, совместность и несовместность составляющих событий.

Событие A представим в виде $A = D_1 \cdot D_2 \cdot D_3$. Тогда, учитывая независимость событий D_1, D_2, D_3, по формуле умножения вероятностей для независимых событий получим: $P(A) = P(D_1) \cdot P(D_2) \cdot P(D_3) = 0.8 \cdot 0.9 \cdot 0.7 = 0.504$.

Событие B означает, очевидно, что студент сдаст или все три экзамена, или только любые два экзамена из трёх. Следовательно:

$$B = D_1 \cdot D_2 \cdot D_3 + \overline{D}_1 \cdot D_2 \cdot D_3 + D_1 \cdot \overline{D}_2 \cdot D_3 + D_1 \cdot D_2 \cdot \overline{D}_3,$$

где $\overline{D}_1, \overline{D}_2, \overline{D}_3$ - события, противоположные к событиям D_1, D_2, D_3: $\overline{D}_1 = \{$студент не сдаст первый экзамен$\}$, $\overline{D}_2 = \{$студент не сдаст второй экзамен$\}$, $\overline{D}_3 = \{$студент не сдаст третий экзамен$\}$, вероятности которых:

$$P(\overline{D}_1) = 1 - P(D_1) = 0.2, \; P(\overline{D}_2) = 1 - P(D_2) = 0.1, \; P(\overline{D}_3) = 1 - P(D_3) = 0.3.$$

Тогда, учитывая несовместность событий $D_1 \cdot \overline{D}_2 \cdot \overline{D}_3$, $\overline{D}_1 \cdot D_2 \cdot \overline{D}_3$, $\overline{D}_1 \cdot \overline{D}_2 \cdot D_3$, являющихся элементарными исходами эксперимента (экзаменационной сессии), а также независимость событий D_1, D_2, D_3, $\overline{D}_1, \overline{D}_2, \overline{D}_3$, используя формулы сложения (для несовместных событий) и умножения вероятностей (для независимых событий), получим:

$$P(B) = P(D_1)\cdot P(D_2)\cdot P(D_3) + P(\overline{D}_1)\cdot P(D_2)\cdot P(D_3) + P(D_1)\cdot P(\overline{D}_2)\cdot P(D_3) +$$
$$+ P(D_1)\cdot P(D_2)\cdot P(\overline{D}_3) = 0.504 + 0.2\cdot 0.9\cdot 0.7 + 0.8\cdot 0.1\cdot 0.7 + 0.8\cdot 0.9\cdot 0.3 = 0.902$$

Событие C, определяемое словами «хотя бы один», означает, что студент сдаст или все три экзамена, или только любые два экзамена из трёх, или только любой один экзамен из трёх. Прямое вычисление вероятности данного события приводит к громоздким вычислениям. Поэтому, сначала найдём вероятность противоположного события $\overline{C}$ ={**студент не сдаст ни одного экзамена**}, представляемого в виде $\overline{C} = \overline{D}_1\cdot\overline{D}_2\cdot\overline{D}_3$. Учитывая независимость событий $\overline{D}_1, \overline{D}_2, \overline{D}_3$, по формуле умножения вероятностей для независимых событий получим: $P(\overline{C}) = P(\overline{D}_1)\cdot P(\overline{D}_2)\cdot P(\overline{D}_3) = 0.2\cdot 0.1\cdot 0.3 = 0.006$.

Тогда $P(C) = 1 - P(\overline{C}) = 1 - 0.006 = 0.994$.

Ответ: $P(A) = 0.504$, $P(B) = 0.902$, $P(C) = 0.994$.

31-40. Требуется найти вероятность указанного события, используя ***формулы полной вероятности и Байеса.***

В группе из 20 студентов, пришедших сдавать экзамен по «Теории вероятностей», 3 студента подготовлены на «5», 5 –на «4», 8 – на «3» и 4 – на «2». В экзаменационных билетах имеется 60 вопросов. Отлично подготовленный студент может ответить на все 60 вопросов, хорошо – на 45, удовлетворительно – на 30 и неудовлетворительно – на 20. Наудачу вызванный студент ответил на произвольно заданный преподавателем вопрос. Найти вероятность того, что студент был подготовлен неудовлетворительно.

Для любого наблюдаемого в эксперименте события A имеет *место формула полной вероятности:* $P(A) = \sum_{i=1}^{n} P(H_i)P(A \mid H_i)$, где $P(H_i) > 0$, $\sum_{i=1}^{n} P(H_i) = 1$. Здесь $H_1, H_2, \ldots, H_n$- наблюдаемые для данного эксперимента события, попарно несовместные и образующие полную группу, только с одним из которых событие A происходит. Такие события называют *гипотезами* по отношению к событию A.

Если стало известно, что в результате эксперимента событие A произошло, то можно переоценить *априорные* (*доопытные*) вероятности $P(H_i)$ гипотез по *формуле Байеса*:

$P(H_i \mid A) = \dfrac{P(H_i)\cdot P(A \mid H_i)}{P(A)}$, где $P(A) = \sum_{i=1}^{n} P(H_i)P(A \mid H_i)$. $P(H_i \mid A)$ называют тогда *апостериорными* (*послеопытными*) вероятностями гипотез H_i.

Для решения задач с использованием формул полной вероятности и Байеса следует:

1) рассмотреть событие A, вероятность которого нужно найти и набор гипотез $H_1, H_2, \ldots, H_n$ - наблюдаемых в том же эксперименте событий, с одним из которых событие A происходит;

2) вычислить вероятности гипотез $P(H_i)$ и условные вероятности $P(A|H_i)$ события A, в предположении, что имела место гипотеза H_i

3) по формуле полной вероятности найти $P(A)$ и, если это требуется по условиям задачи, найти по формуле Байеса апостериорные вероятности $P(H_i \mid A)$, в предположении, что событие A в эксперименте произошло.

Решение.

Рассмотрим событие $A=\{$**наудачу вызванный студент ответил на произвольно заданный вопрос**$\}$. Данное событие, очевидно, может произойти только с одним из следующих несовместных событий (гипотез): $H_1=\{$наудачу вызванный студент подготовлен на «5»$\}$, $H_2=\{$наудачу вызванный студент подготовлен на «4»$\}$, $H_3=\{$наудачу вызванный студент подготовлен на «3»$\}$, $H_4=\{$наудачу вызванный студент подготовлен на «2»$\}$. Из условия задачи следует, что в эксперименте (сдача экзамена) событие A произошло и необходимо переоценить вероятность гипотезы H_4, с учётом дополнительной информации относительно осуществления события A. Решение такой задачи, сводится к нахождению вероятности $P(H_4 \mid A)$ по формуле Байеса: $P(H_4 \mid A)=\dfrac{P(H_4)\cdot P(A|H_4)}{P(A)}$, где $P(A)=\sum\limits_{i=1}^{4} P(H_i)\cdot P(A|H_i)$.

Вычислим, сначала, используя формулу классического определения вероятности, вероятности $P(H_i)$ гипотез H_i и условные вероятности $P(A|H_i)$ события A, в предположении, что имела место каждая из гипотез:

$$P(H_1)=\frac{m(H_1)}{n}=\frac{3}{20}=0.15, \qquad P(H_2)=\frac{m(H_2)}{n}=\frac{5}{20}=0.25,$$

$$P(H_3)=\frac{m(H_3)}{n}=\frac{8}{20}=0.4, \qquad P(H_4)=\frac{m(H_4)}{n}=\frac{4}{20}=0.2, \quad \sum_{i=1}^{4} P(H_i)=1;$$

$$P(A|H_1)=\frac{m(A|H_1)}{n}=\frac{60}{60}=1, \qquad P(A|H_2)=\frac{m(A|H_2)}{n}=\frac{45}{60}=0.75,$$

$$P(A|H_3)=\frac{m(A|H_3)}{n}=\frac{30}{60}=0.5, \qquad P(A|H_4)=\frac{m(A|H_4)}{n}=\frac{20}{60}\approx 0.33.$$

Затем, по формуле полной вероятности, вычислим $P(A)$:

$$P(A)=\sum_{i=1}^{4}P(H_i)\cdot P(A|H_i)=\frac{3}{20}\cdot\frac{60}{60}+\frac{5}{20}\cdot\frac{45}{60}+\frac{8}{20}\cdot\frac{30}{60}+\frac{4}{20}\cdot\frac{20}{60}=\frac{725}{1200}\approx 0.60.$$

После чего, по формуле Байеса, найдём $P(H_4 \mid A)$:

$$P(H_4 \mid A)=\frac{P(H_4)\cdot P(A|H_4)}{P(A)}=\frac{\frac{4}{20}\cdot\frac{20}{60}}{\frac{725}{1200}}=\frac{80}{725}\approx 0.11.$$

Ответ: $P(H_4 \mid A)\approx 0.11$ - вероятность, того на заданный преподавателем вопрос ответил студент подготовленный неудовлетворительно.

41-50. Требуется найти вероятность указанного события, используя ***формулу Бернулли***.
В урне 15 белых и 10 черных шаров. Из урны вынимают подряд 5 шаров, причем каждый вынутый шар возвращают в урну перед извлечением следующего и шары в урне перемешивают. Найти вероятность того, что из пяти вынутых шаров окажется не более двух белых.

Схемой Бернулли называют последовательность испытаний, удовлетворяющую условиям: **1)** результатом каждого испытания является один из двух возможных исходов: «успех» (появление некоторого события A) и «неудача»; **2)** испытания являются независимыми, т.е. вероятность «успеха» в каждом следующем испытании не зависит от результатов предыдущих испытаний; **3)** вероятность «успеха» во всех испытаниях одинакова и равна $P(A)=p$.

Вероятность $P_n(k)$ того, что в n испытаниях по схеме Бернулли произойдёт ровно k «успехов», определяется *формулой Бернулли*:

$$P_n(k)=C_n^k p^k (1-p)^{n-k}, \; k=0,1,2,\ldots,n.$$

Следствием формулы Бернулли является формула: $P_n(k\geq 1)=1-(1-p)^n$ - вероятность того, что в n испытаниях по схеме Бернулли «успех» наступит хотя бы один раз.

Для решения задач с использованием формулы Бернулли следует:
1) установить, что эксперимент представляет собой схему Бернулли (вероятности событий, связанных с таким экспериментом, всегда можно выразить через вероятности $P_n(k)$, вычисляемые по формуле Бернулли);
2) рассмотреть событие A, которое может наступить или не наступить в каждом испытании и вычислить его вероятность $p = P(A)$;
3) рассмотреть событие B, вероятность которого нужно найти и которое состоит в том, что событие A в данном эксперименте появляется определённое число раз;
4) найти $P(B)$, выразив её предварительно, через вероятности $P_n(k)$, вычисляемые по формуле Бернулли.

Решение.

Эксперимент (последовательный выбор пяти шаров из урны с неизменным составом шаров) представляет собой, очевидно, схему Бернулли.

Рассмотрим событие $A = \{$вынутый из урны шар – белый$\}$. Это событие происходит или не происходит при каждом выборе шара из урны с одной и той же вероятностью $p = P(A) = \dfrac{m(A)}{n} = \dfrac{15}{25} = \dfrac{3}{5} = 0.6$.

Рассмотрим событие $B = \{$из пяти вынутых из урны шаров, белых - не более двух$\}$. Таким образом, событие B состоит в том, что в данном эксперименте событие A произойдёт $0, 1$ или 2 раза.

Выразим $P(B)$ через $P_n(k)$ -вероятности того, что событие A в n испытаниях по схеме Бернулли произойдёт ровно k раз: $P(B) = P_5(0) + P_5(1) + P_5(2)$.

Вычислим вероятности $P_n(k)$ по формуле Бернулли:

$$P_5(0) = C_5^0 \cdot \left(\frac{3}{5}\right)^0 \cdot \left(\frac{2}{5}\right)^5 = 1 \cdot 1 \cdot \frac{32}{3125} = 0.01024,$$

$$P_5(1) = C_5^1 \cdot \left(\frac{3}{5}\right)^1 \cdot \left(\frac{2}{5}\right)^4 = 5 \cdot \frac{3}{5} \cdot \frac{16}{625} = 0.0768,$$

$$P_5(2) = C_5^2 \cdot \left(\frac{3}{5}\right)^2 \cdot \left(\frac{2}{5}\right)^3 = 10 \cdot \frac{9}{25} \cdot \frac{8}{125} = 0.2304.$$

Тогда $P(B) = 0.01024 + 0.0768 + 0.2304 = 0.31744 \approx 0.32$.

Ответ: $P(B) \approx 0.32$ - вероятность того, что среди пяти вынутых шаров окажутся не более двух белых шаров.

51-60. Производятся последовательные независимые испытания пяти приборов на надежность. Каждый следующий прибор испытывается только в том случае, если предыдущий оказался надежным. Вероятность выдержать испытания для каждого из приборов равна 0.9. Требуется: составить закон распределения дискретной случайной величины X – числа испытанных приборов; построить многоугольник полученного распределения; вычислить её математическое ожидание MX и дисперсию DX.

Закон распределения ДСВ удобно задавать рядом распределения. *Рядом распределения* ДСВ называют таблицу, в которой перечислены все возможные значения $x_1, x_2, \ldots$ этой случайной величины и соответствующие им вероятности $p_1, p_2, \ldots$.

Решение.

Случайная величина X – число испытанных приборов, может, очевидно, принимать значения: $1, 2, 3, 4, 5$. Вычислим вероятности этих значений $p_i = P(X = x_i)$, используя формулы сложения и умножения вероятностей.

Для вычисления вероятностей p_i могут, в зависимости от условий задачи, использоваться также формулы классического определения вероятности и Бернулли.

Рассмотрим события $A_i = \{$ i **-ый испытанный прибор – надёжный**$\}$ ($i = 1,2,3,4,5$), вероятность которых одинакова и равна $P(A_i) = 0.9$. Противоположными к событиям A_i являются события $\overline{A_i} = \{$ i **-ый испытанный прибор–ненадёжный**$\}$, вероятность их одинакова и равна $P(\overline{A_i}) = 1 - P(A_i) = 0.1$.

Выразим события $X = x_i$, где $x_i \in \{1, 2, 3, 4, 5\}$, через события A_i и $\overline{A_i}$:

$\{X = 1\} = \overline{A_1} = \{$испытывался один прибор$\}$,

$\{X = 2\} = A_1 \cdot \overline{A_2} = \{$испытывались два прибора$\}$,

$\{X = 3\} = A_1 \cdot A_2 \cdot \overline{A_3} = \{$испытывались три прибора$\}$,

$\{X = 4\} = A_1 \cdot A_2 \cdot A_3 \cdot \overline{A_4} = \{$испытывались четыре прибора$\}$,

$\{X = 5\} = A_1 \cdot A_2 \cdot A_3 \cdot A_4 \cdot (\overline{A_5} + A_5) = \{$испытывались все пять приборов$\}$. Очевидно, все пять приборов будут испытаны только при условии, что первые четыре оказались надежными, причем они будут испытаны при любом исходе пятого испытания: $\overline{A_5}$ или A_5.

Вычислим вероятности $p_i = P(X = x_i)$, используя формулы умножения вероятностей для независимых, по условиям задачи, событий A_i и $\overline{A_i}$:

$p_1 = P(X=1) = P(\overline{A_1}) = 1 - P(A_1) = 0.1$,

$p_2 = P(X=2) = P(A_1 \cdot \overline{A_2}) = P(A_1)\cdot(1 - P(A_2)) = 0.9 \cdot 0.1 = 0.09$,

$p_3 = P(X=3) = P(A_1 \cdot A_2 \cdot \overline{A_3}) = P(A_1)\cdot P(A_2)\cdot(1 - P(A_3)) = 0.9 \cdot 0.9 \cdot 0.1 = 0.081$

$p_4 = P(X=4) = P(A_1 \cdot A_2 \cdot A_3 \cdot \overline{A_4}) =$

$$= P(A_1)\cdot P(A_2)\cdot P(A_3)\cdot(1 - P(A_4)) = 0.9^3 \cdot 0.1 = 0.0729.$$

$p_5 = P(X=5) = P(A_1 \cdot A_2 \cdot A_3 \cdot A_4 \cdot (A_5 + \overline{A_5})) =$

$$= P(A_1)P(A_2)P(A_3)P(A_4)P(A_5 + \overline{A_5}) = 0.9^4 \cdot 1 = 0.6561.$$

Если при вычислении вероятностей p_i производится округление их значений, то округление выполняется таким образом, чтобы $\sum_i p_i = 1$.

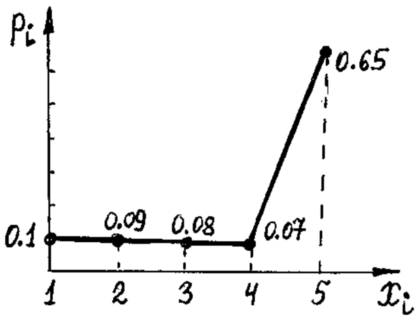
Тогда ряд распределения дискретной случайной величины X имеет вид:

x_i	1	2	3	4	5
p_i	0.1	0.09	0.081	0.0729	0.6561

.

Для наглядности закон распределения ДСВ изображают графически, для чего в прямоугольной системе координат строят точки $M_i(x_i, p_i)$ и соединяют их отрезками прямых. Полученную фигуру и называют *многоугольником распределения*.

Построим в прямоугольной системе координат, многоугольник полученного распределения:



Математическим ожиданием (средним значением) дискретной случайной величины X называется число $MX = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots x_n p_n$.

Вычислим математическое ожидание MX:

$$MX = \sum_{i=1}^{5} x_i p_i = 1 \cdot 0.1 + 2 \cdot 0.09 + 3 \cdot 0.081 + 4 \cdot 0.0729 + 5 \cdot 0.6561 = 4.0951 \approx 4.10.$$

Дисперсией случайной величины X называется неотрицательное число $DX = M(X - MX)^2$. Дисперсию дискретной случайной величины X вычисляют по формулам: $DX = \sum_i (x_i - MX)^2 p_i$ или $DX = \sum_i x_i^2 p_i - (MX)^2$.

Дисперсию DX вычислим по формуле $DX = M(X^2) - (MX)^2$, где

$$M(X^2) = \sum_{i=1}^{5} x_i^2 p_i = 1^2 \cdot 0.1 + 2^2 \cdot 0.09 +$$

$$+ 3^2 \cdot 0.081 + 4^2 \cdot 0.0729 + 5^2 \cdot 0.6561 = 18.7579 \approx 18.76.$$

Тогда $DX \approx 18.76 - (4.10)^2 = 1.95$.

Ответ:

x_i	1	2	3	4	5
p_i	0.1	0.09	0.081	0.0729	0.6561

, $MX \approx 4.10$, $DX \approx 1.95$.

61-70. Найти функцию плотности вероятностей $f(x)$ непрерывной случайной величины X, заданной функцией распределения $F(x) = \begin{cases} 0 & x \le 1 \\ \dfrac{(x^2 - x)}{2} & 1 < x \le 2 \\ 1 & x > 2 \end{cases}$; вычислить её математическое ожидание MX, дисперсию DX и вероятность $P(X \in (5/4, 3/2))$.

Закон распределения непрерывной случайной величины удобно задавать функцией плотности вероятностей $f(x)$ - неотрицательной и интегрируемой в бесконечных пределах функцией.
Функция плотности вероятностей $f(x)$ в точках, где $F(x)$ дифференцируема, определяется равенством: $f(x) = F'(x)$. В точках, где $F(x)$ недифференцируема, $f(x)$ определяется произвольным образом, чаще всего по непрерывности слева или справа.

Непрерывная функция, задаваемая в области своего определения несколькими аналитическими выражениями, может оказаться недифференцируемой в точках, в окрестности которых она задаётся разными аналитическими выражениями.

Решение.

Найдём функцию плотности вероятностей $f(x)$, как производную от функции распределения $F(x)=\begin{cases} 0 & x\le 1 \\ \dfrac{(x^2-x)}{2} & 1<x\le 2 \\ 1 & x>2 \end{cases}$. Учитывая, что: $(0)'=0$,

$\left(\dfrac{x^2-x}{2}\right)'=\dfrac{2x-1}{2}$, $(1)'=0$, получим $f(x)=\begin{cases} 0 & x<1 \\ \dfrac{2x-1}{2} & 1\le x\le 2 \\ 0 & x>2 \end{cases}$.

В точках $x=1$ и $x=2$, являющихся концами промежутка, где $f(x)>0$, и в которых функция $F(x)$ недифференцируема, функцию плотности вероятностей $f(x)$, определили таким образом, чтобы на концах промежутка она была непрерывной справа (в точке $x=1$) и слева (в точке $x=2$).

Математическим ожиданием непрерывной случайной величины X называется число $MX=\int\limits_{-\infty}^{+\infty} xf(x)dx$.

Вычислим математическое ожидание MX:

$$MX=\int\limits_{-\infty}^{+\infty} x\,f(x)dx=\int\limits_{-\infty}^{1} x\cdot 0dx+\int\limits_{1}^{2} x\cdot\frac{(2x-1)}{2}dx+\int\limits_{2}^{+\infty} x\cdot 0dx=0+\int\limits_{1}^{2}\left(x^2-\frac{x}{2}\right)dx+0=$$

$$=\left(\frac{x^3}{3}-\frac{x^2}{4}\right)\Bigg|_1^2=\left(\frac{8}{3}-1\right)-\left(\frac{1}{3}-\frac{1}{4}\right)=\frac{19}{12}\approx 1.58.$$

Дисперсию непрерывной случайной величины X вычисляют по формулам:

$$DX=\int\limits_{-\infty}^{+\infty}(x-MX)^2 f(x)dx \text{ или } DX=\int\limits_{-\infty}^{+\infty}x^2 f(x)dx-(MX)^2.$$

Дисперсию DX вычислим по формуле $DX = M(X^2) - (MX)^2$, где

$$M(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \int_{-\infty}^{1} x^2 \cdot 0dx + \int_{1}^{2} x^2 \cdot \frac{(2x-1)}{2}dx + \int_{2}^{+\infty} x^2 \cdot 0\,dx =$$

$$= 0 + \int_{1}^{2}\left(x^3 - \frac{x^2}{2}\right)dx + 0 = \left(\frac{x^4}{4} - \frac{x^3}{6}\right)\Bigg|_1^2 = \left(\frac{16}{4} - \frac{8}{6}\right) - \left(\frac{1}{4} - \frac{1}{6}\right) = \frac{31}{12} \approx 2.58\,.$$

Тогда $DX \approx 2.58 - (1.58)^2 \approx 0.08$.

Для непрерывной случайной величины X справедлива формула: $P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = P(a < X < b) =$ $= F(b) - F(a) = \int_a^b f(x)dx$.

Вероятность $P(X \in (5/4, 3/2))$ вычислим по формуле:

$$P(X \in (5/4, 3/2)) = = F\left(\frac{5}{4}\right) - F\left(\frac{3}{2}\right) = \left(\frac{x^2 - x}{2}\right)\Bigg|_{x=3/2} - \left(\frac{x^2 - x}{2}\right)\Bigg|_{x=5/4} =$$

$$= \frac{(3/2)^2 - (3/2)}{2} - \frac{(5/4)^2 - (5/4)}{2} = \frac{7}{32} \approx 0.22\,.$$

Ответ: $f(x) = \begin{cases} 0 & x < 1 \\ \dfrac{2x-1}{2} & 1 \le x \le 2 \\ 0 & x > 2 \end{cases}$, $MX \approx 1.58$, $DX \approx 0.08$, $P(X \in (5/4, 3/2)) \approx 0.22$.

Раздел. <u>Математическая статистика.</u>

71-80. Дана выборка объема $n = 15$:

23	**23**	**21**	**20**	**20**	**23**	**23**	**25**	**23**	**20**	**20**	**24**	**21**	**25**	**21**

Требуется: построить вариационный и дискретный статистический ряды; вычислить числовые характеристики выборки: $x_{\min}$, $x_{\max}$, $\widehat{R}$ (размах), $\bar{x}$ (среднее арифметическое), $\widehat{\sigma}^2$ (дисперсию); построить полигон частот.

Вариационным рядом выборки $x_1, x_2, \ldots, x_n$ называется такой способ её записи, при котором элементы выборки упорядочиваются по величине, т.е. записываются в виде последовательности $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, где $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$. Разность $x_{(n)} - x_{(1)} = \widehat{R}$ называется *размахом выборки.*

Различные значения x_i, $i = \overline{1,k}$ ($k \le n$), называются *вариантами.* Число n_i повторений варианты x_i в выборке называется её *частотой.*

Дискретным статистическим рядом называется упорядоченная в порядке возрастания значений вариант x_i последовательность пар (x_i, n_i), $i = \overline{1,k}$. Обычно его записывают в виде таблицы, первая стока которой содержит варианты x_i, а вторая их частоты.

Полигоном частот называется фигура, расположенная под ломаной линией с вершинами в точках $M_i(x_i, n_i)$.

Решение.

Построим вариационный ряд выборки, расположив элементы выборки в порядке возрастания их значений. Получим:

20	**20**	**20**	**20**	**21**	**21**	**21**	**23**	**23**	**23**	**23**	**23**	**24**	**25**	**25**

Построим дискретный статистический ряд и запишем его в виде таблицы, в первой строке которой расположим различные значения элементов выборки в порядке их возрастания, а во второй соответствующие им частоты. Получим:

x_i	20	21	23	24	25
n_i	4	3	5	1	2

.

Если выборка записана в виде дискретного статистического ряда

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

, где $\sum_{i=1}^{k} n_i = n$, то *среднее арифметическое выборки* $\bar{x}$ вычисляют по формуле $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$, а *дисперсию выборки* $\widehat{\sigma}^2$ - по формулам: $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i$ или $\widehat{\sigma}^2 = \overline{x^2} - (\bar{x})^2$, где $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i$.

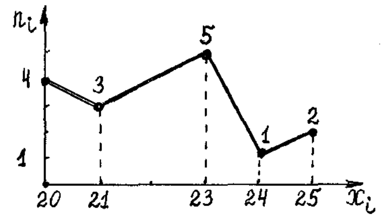
Вычислим числовые характеристики выборки: $x_{\min}$, $x_{\max}$, $\widehat{R}$, $\overline{x}$, $\widehat{\sigma}^2$. Получим: $x_{\min}=20$, $x_{\max}=25$, $\widehat{R}=x_{\max}-x_{\min}=25-20=5$,

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i=\frac{1}{15}\cdot(20\cdot4+21\cdot3+23\cdot5+24\cdot1+25\cdot2)=\frac{332}{15}\approx22.13,$$

$$\overline{x^2}=\frac{1}{n}\sum_{i=1}^{k}x_i^2 n_i=\frac{1}{15}\cdot\left(20^2\cdot4+21^2\cdot3+23^2\cdot5+24^2\cdot1+25^2\cdot2\right)=\frac{7394}{15}\approx492.93,$$

$$\widehat{\sigma}^2=\overline{x^2}-(\overline{x})^2\approx492.93-(22.13)^2\approx492.93-489.74=3.19.$$

Построим полигон частот. Для его построения в прямоугольной системе координат, в которой по оси абсцисс откладываются варианты x_i, по оси ординат – частоты n_i, начало системы координат совмещено с точкой $(x_{\min},0)$, изобразим точки $M_i(x_i,n_i)$ и соединим их отрезками.



Ответ: Вариационный ряд:

20	20	20	20	21	21	21	23	23	23	23	23	24	25	25

Дискретный статистический ряд:

x_i	20	21	23	24	25
n_i	4	3	5	1	2

.

$x_{\min}=20$, $x_{\max}=25$, $\widehat{R}=5$, $\overline{x}\approx22.13$, $\widehat{\sigma}^2\approx3.19$.

81-90. Получены данные о содержании меди (в %) в 60 образцах сплава:

Содержание меди	**52-56**	**56-60**	**60-64**	**64-68**	**68-72**
Число образцов сплава	**3**	**9**	**18**	**14**	**16**

Требуется: вычислить числовые характеристики группированной выборки: $x_{\min}$, $x_{\max}$, $\widehat{R}$ (размах), $\overline{x}$ (среднее арифметическое), $\widehat{\sigma}^2$ (дисперсию); построить гистограмму частот; найти $95\ \%$-ный доверительный интервал для генеральной средней MX (предполагая нормальным закон распределения генеральной совокупности X, из которой получена выборка).

Если выборка записана в виде интервального статистического ряда

J_i	J_1	J_2	$\cdots$	J_k
n_i	n_1	n_2	$\cdots$	n_k

, где $\sum_{i=1}^{k} n_i = n$, то *среднее арифметическое выборки* $\bar{x}$ вычисляют по формуле $\bar{x} = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i n_i$, а *дисперсию выборки* $\hat{\sigma}^2$ - по формулам: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}(\tilde{x}_i - \bar{x})^2 n_i$ или $\hat{\sigma}^2 = \overline{x^2} - (\bar{x})^2$, где $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i^2 n_i$, $\tilde{x}_i$ - середина интервала J_i.

Решение.

Вычислим числовые характеристики выборки: $x_{\min}$, $x_{\max}$, $\hat{R}$, $\bar{x}$, $\hat{\sigma}^2$. Получим: $x_{\min} = 52$, $x_{\max} = 72$, $\hat{R} = x_{\max} - x_{\min} = 72 - 52 = 20$,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i n_i = \frac{1}{60}\cdot(54\cdot 3 + 58\cdot 9 + 62\cdot 18 + 66\cdot 14 + 70\cdot 16) = \frac{3844}{60} \approx 64.07,$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i^2 n_i = \frac{1}{60}\cdot\left(54^2\cdot 3 + 58^2\cdot 9 + 62^2\cdot 18 + 66^2\cdot 14 + 70^2\cdot 16\right) =$$

$$= \frac{247600}{60} \approx 4126.67,$$

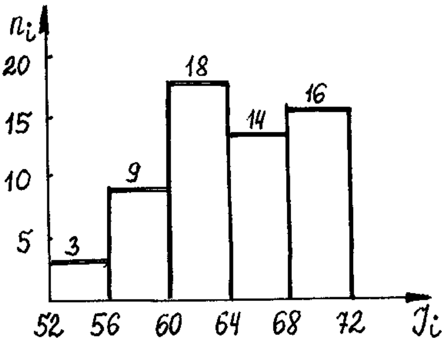
$$\hat{\sigma}^2 = \overline{x^2} - (\bar{x})^2 \approx 4126.67 - (64.07)^2 \approx 4126.67 - 4104.96 = 21.71.$$

Гистограммой частот называется ступенчатая фигура, составленная из прямоугольников, построенных на интервалах группировки так, что площадь каждого прямоугольника равна частоте n_i, $i = \overline{1,k}$. Если длины всех интервалов одинаковы и равны h, то высоты прямоугольников равны $\frac{n_i}{h}$.

Часто, при построении гистограмм частот по интервалам равной длины, высоту прямоугольников выбирают равной частоте.

Построим гистограмму частот. Для этого в прямоугольной системе координат, в которой по оси абсцисс откладываются интервалы J_i, по оси ординат

– частоты n_i, начало системы координат совмещено с точкой $(x_{\min}, 0)$, на интервалах J_i, как на основаниях, построим прямоугольники высоты n_i.



Если функция распределения $F_X(x, \theta)$ генеральной совокупности X известна с точностью до параметра θ, то его *доверительным интервалом* называется случайный интервал $(\hat{\theta}_1, \hat{\theta}_2)$, который накрывает неизвестное значение параметра θ с заданной вероятностью $0 < \gamma < 1$, т.е. $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma$. Число $\gamma = 1 - \alpha$ называется *доверительной вероятностью*, а число α - *уровнем значимости.* Обычно используются значения γ, равные $0.90, 0.95, 0.99$.

Если выборка получена из генеральной совокупности X, имеющей нормальное распределение $N(a, \sigma)$, с неизвестной дисперсией σ^2, то доверительный интервал для генеральной средней $MX = a$ находится по формуле:

$$\bar{x} - \frac{\hat{\sigma}}{\sqrt{n-1}} \cdot t_\alpha(n-1) < a < \bar{x} + \frac{\hat{\sigma}}{\sqrt{n-1}} \cdot t_\alpha(n-1),$$

где $t_\alpha(k)$ - коэффициент доверия, определяемый по заданным значениям $\alpha = 1 - \gamma$ и $k = n - 1$ с помощью следующей таблицы (для $\alpha = 0.05$):

k	1	2	3	4	5	6	7	8	9	10
$t_{0.05}(k)$	**12.71**	**4.30**	**3.18**	**2.78**	**2.57**	**2.45**	**2.36**	**2.31**	**2.26**	**2.23**
k	11	12	13	14	15	16	17	18	19	20
$t_{0.05}(k)$	**2.20**	**2.18**	**2.16**	**2.14**	**2.13**	**2.12**	**2.11**	**2.10**	**2.09**	**2.09**
k	21	22	23	24	25	26	27	28	29	30
$t_{0.05}(k)$	**2.08**	**2.07**	**2.07**	**2.06**	**2.06**	**2.06**	**2.05**	**2.05**	**2.05**	**2.04**

k	40	50	60	90	120	∞
$t_{0.05}(k)$	**2.02**	**2.01**	**2.00**	**1.99**	**1.98**	**1.96**

Если для данного k значения коэффициента доверия $t_{\alpha}(k)$ в таблице нет, то в качестве $t_{\alpha}(k)$ берут значение коэффициента доверия для ближайшего меньшего числа k.

Построим 95 %-ный доверительный интервал для генеральной средней MX (предполагая нормальным закон распределения генеральной совокупности X, из которой получена выборка). Учитывая, что: $n = 60$, $\gamma = 0.95$, $\alpha = 1 - \gamma = 0.05$, $\bar{x} \approx 64.07$, $\hat{\sigma} = \sqrt{\hat{\sigma}^2} \approx \sqrt{21.71} \approx 4.66$, $t_{\alpha}(k) = t_{0.05}(59) \approx 2.01$,

получим: $$\bar{x} - \frac{\hat{\sigma}}{\sqrt{n-1}} \cdot t_{\alpha}(n-1) < MX < \bar{x} + \frac{\hat{\sigma}}{\sqrt{n-1}} \cdot t_{\alpha}(n-1) \Rightarrow$$

$$64.07 - \frac{4.66}{\sqrt{60-1}} \cdot 2.01 < MX < 64.07 + \frac{4.66}{\sqrt{60-1}} \cdot 2.01 \Rightarrow$$

$$64.07 - 1.22 < MX < 64.07 + 1.22 \quad \Rightarrow \quad 62.85 < MX < 65.29.$$

Таким образом, интервал $(62.85, 65.29)$ накроет значение MX с вероятностью не меньшей, чем $\gamma = 0.95$.

Ответ: $x_{\min} = 52$, $x_{\max} = 72$, $\hat{R} = 20$, $\bar{x} \approx 64.07$, $\hat{\sigma}^2 \approx 21.71$, $(62.85, 65.29)$.

91-100.

Получены *данные о текучести кадров Y (% в год) и среднемесячной зарплате X (в тыс.руб) работников 10 торговых центров города:*

Y	**4.2**	**3.8**	**3.5**	**4.0**	**4.4**	**3.8**	**4.5**	**4.2**	**4.8**	**4.6**
X	**5.5**	**6.0**	**6.5**	**6.2**	**5.2**	**5.8**	**4.5**	**5.2**	**4.2**	**4.3**

Для приведённой выборки (предполагается, что выборка получена из двумерной нормально распределённой генеральной совокупности) требуется:

а) Построить диаграмму рассеивания (корреляционное поле).

б) Вычислить выборочный коэффициент линейной корреляции r; проверить его значимость на уровне $\alpha = 0.05$, сделать выводы (для значимого r) о тесноте и направлении связи между величинами X и Y.

в) Найти выборочное уравнение прямой регрессии Y на X и построить ее график на одном чертеже с диаграммой рассеивания.

г) Вычислить, используя регрессионную зависимость Y на X, ожидаемое среднее значение $\bar{y}_{x_0}$ величины Y при $X = x_0 = 5$.

Решение.

а) Построим диаграмму рассеивания, изобразим в прямоугольной декартовой системе координат точки (x_i, y_i).

Предварительное представление о зависимости между случайными величинами X и Y можно получить, изобразив в прямоугольной системе координат на плоскости точки (x_i, y_i). Такое графическое представление двумерной выборки называют *диаграммой рассеивания* (*корреляционным полем*).

Количественной характеристикой степени линейной зависимости между величинами X и Y является *коэффициент корреляции* $\rho = \dfrac{M(XY) - MX \cdot MY}{\sqrt{DX \cdot DY}}$.

Его оценкой служит выборочный коэффициент линейной корреляции r, значение которого вычисляется по формуле $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{\sigma}_x \cdot \hat{\sigma}_y}$, где $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$,

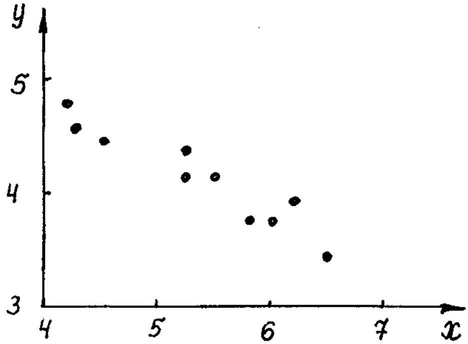
$\bar{y} = \dfrac{1}{n}\sum\limits_{i=1}^{n} y_i$, $\overline{xy} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i y_i$, $\hat{\sigma}_x^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$, $\hat{\sigma}_y^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(y_i - \bar{y})^2$

Проверка значимости выборочного коэффициента линейной корреляции r (значимого отличия от нуля найденного по выборке значения r) сводится к проверке статистической гипотезы $H_0 : \rho = 0$ (о равенстве нулю коэффициента линейной корреляции ρ в генеральной совокупности) при альтернативе $H_1 : \rho \neq 0$. Проверка осуществляется с помощью статистики $T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

При условии справедливости основной гипотезы H_0 статистика T имеет распределение Стьюдента $t(k)$ с числом степеней свободы $k = n-2$. Решение об отклонении основной гипотезы H_0 (признании значимости выборочного коэффициента линейной корреляции r) принимается так: если $|t_{набл}| > t_\alpha(k)$ на заданном уровне значимости α, то H_0 отклоняется. В противном случае делается вывод о согласии выборочных данных с основной гипотезой. Здесь $t_\alpha(k)$-критическая точка распределения Стьюдента, значения которой находятся (для $\alpha = 0.05$) по таблице на странице ***39-40***.

По статистически значимому выборочному коэффициенту корреляции r делают выводы о тесноте линейной корреляционной связи и её направлении. Если $|r| < 0.3$, то связь считается слабой; если $0.3 < |r| < 0.7$, то – умеренной; если $|r| > 0.7$ - то – тесной. Если $r < 0$, то связь является обратной; если $r > 0$, то - прямой.

Выборочной прямой регрессии Y на X является прямая: $\bar{y}_x = \bar{y} + r\dfrac{\hat{\sigma}_y}{\hat{\sigma}_x}(x - \bar{x})$, которая устанавливает линейную корреляционную зависимость между значениями x величины X и соответствующими им условными средними $\bar{y}_x$ величины Y.



б) Вычислим значение выборочного коэффициента линейной корреляции r.
Учитывая, что:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}\left(5.5+6.0+6.5+6.2+5.2+5.8+4.5+5.2+4.2+4.3\right) = \frac{53.4}{10} = 5.34,$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{10}\left(4.2+3.8+3.5+4.0+4.4+3.8+4.5+4.2+4.8+4.6\right) = \frac{41.8}{10} = 4.18,$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \ \frac{1}{10}(5.5\cdot 4.2+6.0\cdot 3.8+6.5\cdot 3.5+6.2\cdot 4.0+5.2\cdot 4.4+5.8\cdot 3.8+$$

$$+4.5\cdot 4.5+5.2\cdot 4.2+4.2\cdot 4.8+4.3\cdot 4.6) = \frac{220.4}{10} = 22.04,$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{10}(5.5^2+6.0^2+6.5^2+6.2^2+5.2^2+5.8^2+4.5^2+5.2^2+$$

$$+4.2^2+4.3^2) = \frac{291.04}{10} \approx 29.10,$$

$$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{10}(4.2^2+3.8^2+3.5^2+4.0^2+4.4^2+3.8^2+4.5^2+4.2^2+$$

$$+4.8^2+4.6^2)=\frac{176.22}{10}\approx 17.62,$$

$$\hat{\sigma}_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\overline{x^2}-(\bar{x})^2\approx 29.10-(5.34)^2\approx 0.58,$$

$$\hat{\sigma}_y^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2=\overline{y^2}-(\bar{y})^2\approx 17.62-(4.18)^2\approx 0.15,$$

получим: $$r=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\hat{\sigma}_x\cdot\hat{\sigma}_y}=\frac{22.04-5.34\cdot 4.18}{\sqrt{0.58}\cdot\sqrt{0.15}}\approx -0.95.$$

Проверим значимость выборочного коэффициента корреляции r. Для этого сначала вычислим по выборке наблюдаемое значение $t_{набл}$ статистики $T=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. Получим $t_{набл}\approx\dfrac{-0.95\sqrt{10-2}}{\sqrt{1-(-0.95)^2}}\approx 8.61$. Затем, с помощью таблицы критических точек распределения Стьюдента (страница **38-39**), найдём по заданному уровню значимости $\alpha=0.05$ и числу степеней свободы $k=n-2=10-2=8$ значение критической точки $t_\alpha(k)$. Получим $t_{0.05}(8)=2.31$. Так как $|t_{набл}|\approx 8.61>t_{0.05}(8)=2.31$, то делаем вывод о значимом отличии от нуля найденного по выборке коэффициента линейной корреляции r.

Поскольку $r\approx -0.95$ - статистически значим и выполнены условия: $|r|>0.7$ и $r<0$, то делаем вывод о тесной обратной линейной корреляционной зависимости между величинами X и Y.

Если выборочный коэффициент корреляции r незначим, то никаких выводов по найденному значению r о тесноте и направлении связи не делают.

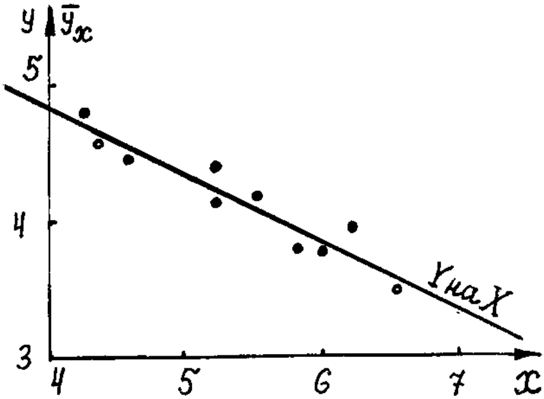
в) Найдём выборочное уравнение прямой регрессии Y на X в виде:

$$\bar{y}_x=\bar{y}+r\frac{\hat{\sigma}_y}{\hat{\sigma}_x}(x-\bar{x}).$$

Учитывая, что $\bar{x}=5.34$, $\bar{y}=4.18$, $\hat{\sigma}_x\approx\sqrt{0.58}\approx 0.76$, $\hat{\sigma}_y\approx\sqrt{0.15}\approx 0.39$, $r\approx -0.95$, получим: $\bar{y}_x=4.18+(-0.95)\cdot\dfrac{0.39}{0.76}(x-5.34)\Rightarrow$

$$\bar{y}_x=4.18-0.4875\cdot(x-5.34)\quad\Rightarrow\quad \bar{y}_x=6.78-0.49x$$

Построим на одном чертеже с диаграммой рассеивания график функции регрессии $\bar{y}_x=6.78-0.49x$, откладывая по оси ординат условные средние значения $\bar{y}_x$. Для построения прямой, найдём точки, через которые она проходит. Положим $x=4$ и $x=7$, тогда из уравнения прямой получим: $\bar{y}_{x=4}=4.82\approx 4.8$ и $\bar{y}_{x=7}=3.35$. Таким образом, прямая проходит через точки: $(4, 4.8)$ и $(7, 3.35)$ в плоскости переменных x и $\bar{y}_x$.



г) Вычислим, используя регрессионную зависимость Y на X, ожидаемое среднее значение $\bar{y}_{x_0}$ величины Y при $X=x_0=5$. Получим: $\bar{y}_{x=5}=6.78-0.49\cdot 5=4.33$, т.е. при среднемесячной зарплате в 5тыс руб. текучесть кадров (% в год) в торговых центрах города составит в среднем 4.33%.

Ответ: $r\approx -0.95$ - значим; связь между Y и X - обратная и тесная; $\bar{y}_x=6.78-0.49x$; $\bar{y}_{x=5}=4.33$.

6.2. Краткие теоретические сведения.

Тема. Случайные события и их вероятности.

1. Классическое и геометрическое определения вероятности.

При ***классическом определении вероятность*** $P(A)$ ***случайного события*** A определяется равенством $P(A)=\dfrac{m(A)}{n}$, где $m(A)$ - число элементарных исходов эксперимента (опыта, испытания), благоприятствующих появлению события A; n - общее число равновозможных элементарных исходов эксперимента. Каждый из исходов (далее неделимых и взаимно исключающих друг друга) эксперимента называется его ***элементарным исходом*** (элементарным событием) и обозначается ω. Элементарные исходы называются ***равновозможными***, если в силу условий проведения эксперимента можно считать, что ни один из них не является объективно более возможным, чем другие. Множество всех элементарных исходов эксперимента называется ***пространством элементарных исходов*** и обозначается Ω. Исход ω называется ***благоприятствующим*** данному событию, если его появление влечёт за собой наступление такого события.

Противоположным событию A называется событие $\overline{A}$, состоящее в том, что событие A не происходит. Например, противоположным событию, определяемому словами «хотя бы один…» является событие, определяемое словами «ни один…». Если вероятность $P(\overline{A})$ известна или легко может быть найдена, то вероятность $P(A)$ вычисляют по формуле: $P(A)=1-P(\overline{A})$.

Для вычисления общего числа n элементарных исходов и числа $m(A)$ элементарных исходов, благоприятствующих рассматриваемому событию, широко используются правила и формулы комбинаторики. Одной из основных задач комбинаторики является подсчёт числа комбинаторных конфигураций (комбинаций элементов), образованных из элементов некоторых конечных множеств в соответствии с заданными правилами. Примерами таких комбинаций являются перестановки, размещения и сочетания.

Сочетаниями из n ***элементов по*** m называются комбинации элементов, отличающиеся друг от друга только составом элементов. Они рассматриваются как элементарные исходы эксперимента, состоящего в одновременном выборе без возвращения любых m элементов из n различных элементов, а их общее число C_n^m определяется формулой:

$$C_n^m=\frac{n!}{m!\cdot(n-m)!},\ \text{где}\ n!=1\cdot2\cdot3\cdot\ldots\cdot(n-1)\cdot n,\ 0!=1.$$

Размещениями из n ***элементов по*** m называются комбинации элементов, отличающиеся друг от друга как составом элементов, так и порядком их следования. Они рассматриваются как элементарные исходы эксперимента, состоящего сначала в одновременном выборе без возвращения любых m элементов из n различных элементов, а затем в произвольном их упорядочивании. Общее число A_n^m размещений определяется формулой: $A_n^m = \dfrac{n!}{(n-m)!}$.

Перестановками из n ***элементов*** называются комбинации элементов, отличающиеся друг от друга только порядком их следования. Они рассматриваются как элементарные исходы эксперимента, состоящего в произвольном упорядочивании множества, состоящего из n различных элементов, а их общее число P_n определяется формулой $P_n = n!$.

Для подсчёта числа всевозможных комбинаторных конфигураций широко используются правила комбинаторики.

Пусть $\alpha_1, \alpha_2, \ldots, \alpha_k$ - элементы (действия) из некоторого конечного множества элементов (действий), которые можно выбрать (выполнить), соответственно, $n_1, n_2, \ldots, n_k$ способами. Тогда справедливы следующие правила.

Правило сложения. Осуществить выбор (выполнение) только одного из элементов (действий) можно $N = n_1 + n_2 + \ldots + n_k$ способами.

Правило умножения. Осуществить последовательный выбор (выполнение) всех элементов (действий) можно $N = n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Пусть эксперимент состоит в том, что наудачу бросается точка в некоторую область Ω. Слово «наудачу» означает, что в таком эксперименте все точки области Ω «равновозможны». В этом случае вероятность попадания точки в некоторую часть A области Ω равна отношению меры (длины, площади, объёма) этой части к мере всей области Ω: $P(A) = \dfrac{\mu(A)}{\mu(\Omega)}$, в предположении, что указанные меры определены, причём $\mu(\Omega) \neq 0$. Данное определение вероятности события называют ***геометрическим определением вероятности***.

2. Условная вероятность. Формулы сложения и умножения вероятностей. Формула полной вероятности. Формула Байеса.

Всякое случайное событие A можно рассматривать как подмножество Ω (обратное утверждение, вообще говоря, места не имеет), состоящее из всех тех $\omega \in \Omega$, которые благоприятствуют событию A ($\omega \in A$). Множество Ω называют ***достоверным событием***, а пустое множество $\varnothing$, являющееся по определению подмножеством Ω, называют ***невозможным событием***.

Если $A \subset B$, то говорят, что ***событие*** A ***влечёт событие*** B.

Произведением событий A и B называют событие $A \cdot B$, происходящее тогда и только тогда, когда происходят одновременно оба события A и B. События A и B называют ***несовместными***, если $A \cdot B = \varnothing$.

Суммой событий A и B называют событие $A + B$, происходящее тогда и только тогда, когда происходит хотя бы одно из событий A или B.

Разностью событий A и B называют событие $A \setminus B$, происходящее тогда и только тогда, когда происходит событие A, но не происходит событие B. Событие $\overline{A} = \Omega \setminus A$, происходящее тогда и только тогда, когда событие A не происходит, называют ***противоположным*** событию A. Разность событий $A \setminus B$ всегда можно представить в виде $A \setminus B = A \cdot \overline{B}$.

Из определений вероятности следуют следующие её свойства:

1) $P(\varnothing) = 0$ **; 2)** $P(\Omega) = 1$; **3)** $0 \le P(A) \le 1$; **4)** Если $A \subset B$, то $P(A) \le P(B)$;

5) $P(\overline{A}) = 1 - P(A)$; **6)** $P(A+B) = \begin{cases} P(A) + P(B), & \text{если } A \cdot B = \varnothing \\ P(A) + P(B) - P(A \cdot B), & \text{если } A \cdot B \ne \varnothing \end{cases}$.

Пусть A и B - наблюдаемые события в эксперименте, причём $P(A) > 0$. ***Условной вероятностью*** $P(B \mid A)$ осуществления ***события*** B при условии, что событие A произошло в результате данного эксперимента, называется величина, определяемая равенством: $P(B \mid A) = \dfrac{P(A \cdot B)}{P(A)}$.

События A и B, имеющие ненулевую вероятность, называются ***независимыми,*** если выполняется равенство $P(B \mid A) = P(B)$ или $P(A \mid B) = P(A)$, в противном случае события A и B называются ***зависимыми***.

Сложным называют событие, наблюдаемое в эксперименте и выраженное через другие наблюдаемые в том же эксперименте события с помощью допустимых алгебраических операций над событиями.

Вероятность осуществления того или иного сложного события вычисляется с помощью ***формул умножения вероятностей***:

1) $P(A \cdot B) = P(A) \cdot P(B \mid A)$, $P(A) > 0$;

2) $P(A \cdot B) = P(A) \cdot P(B)$ (для независимых событий)

и ***формул сложения вероятностей***:

3) $P(A+B) = P(A) + P(B) - P(A \cdot B)$;

4) $P(A+B) = P(A) + P(B)$ (для несовместных событий).

Пусть $H_1, H_2, \ldots, H_n$ - наблюдаемые события для данного эксперимента, попарно несовместные ($H_i \cdot H_j = \varnothing$ при $i \ne j$) и образующие полную группу событий ($H_1 + H_2 + \ldots + H_n = \Omega$). Такие события H_i принято называть

гипотезами по отношению к событию A. Тогда для любого наблюдаемого в эксперименте события A имеет место ***формула полной вероятности:***

$$P(A)=\sum_{i=1}^{n}P(H_i)P(A|H_i)\text{, где } P(H_i)>0.$$

Пусть $H_1,H_2,\ldots,H_n$ - совокупность гипотез по отношению к событию A, безусловные вероятности которых $P(H_i)>0$, называемые ***априорными*** (*до-опытными*), известны и пусть стало известно, что в результате эксперимента событие A произошло. Тогда ***апостериорные*** (*послеопытные*) вероятности $P(H_i \mid A)$ гипотез H_i при условии, что событие A имело место, вычисляются по ***формуле Байеса***:

$$P(H_i \mid A)=\frac{P(H_i)\cdot P(A|H_i)}{P(A)}\text{, где } P(A)=\sum_{i=1}^{n}P(H_i)P(A|H_i).$$

Формула Байеса позволяет переоценить вероятность каждой из гипотез после поступления дополнительной информации относительно осуществления тех или иных наблюдаемых событий.

3. Схема Бернулли. Формула Бернулли.

Схемой Бернулли называют последовательность испытаний, удовлетворяющую условиям: **1)** результатом каждого испытания является один из двух возможных исходов: «успех» (появление некоторого события A) и «неудача»; **2)** испытания являются независимыми, т.е. вероятность «успеха» в каждом следующем испытании не зависит от результатов предыдущих испытаний; **3)** вероятность «успеха» во всех испытаниях одинакова и равна $P(A)=p$.

Вероятность $P_n(k)$ того, что в n испытаниях по схеме Бернулли произойдёт ровно k «успехов», определяется ***формулой Бернулли***:

$$P_n(k)=C_n^k p^k (1-p)^{n-k},\ k=0,1,2,\ldots,n.$$

Следствиями формулы Бернулли являются формулы:

1) $P_n(k_1\le k\le k_2)=\sum\limits_{k=k_1}^{k_2}P_n(k)$ - вероятность того, что в n испытаниях по схеме Бернулли «успех» наступит не более k_1 раз и не менее k_2 раз;

2) $P_n(k\ge 1)=1-(1-p)^n$ - вероятность того, что в n испытаниях по схеме Бернулли «успех» наступит хотя бы один раз.

Тема. Случайные величины. Системы случайных величин.

1. Одномерные случайные величины.

Под ***случайной величиной*** X понимают величину, принимающую свои возможные значения x в зависимости от исхода ω эксперимента, с которым она связана.

Законом распределения *(вероятностей)* случайной величины называют любое правило, позволяющее найти вероятность того, что случайная величина примет значение из некоторого подмножества своих возможных значений. Общим законом распределения, присущим всем случайным величинам, является функция распределения.

Функцией распределения *(вероятностей)* случайной величины X называется функция $F(x)$ действительной переменной x, $-\infty < x < +\infty$, определяемая формулой $F(x) = P(X < x)$.

Каждая функция распределения $F(x)$ обладает следующими свойствами:

1) $0 \le F(x) \le 1$, $-\infty < x < +\infty$; **2)** $F(x)$ не убывает;

3) $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$, $F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$; **4)** $F(x)$ непрерывна слева.

Любая неубывающая непрерывная слева действительная функция $F(x)$, удовлетворяющая условиям $F(-\infty) = 0$ и $F(+\infty) = 1$, является функцией распределения некоторой случайной величины.

Вероятность события $a \le X < b$ определяется формулой:

$$P(a \le X < b) = F(b) - F(a).$$

Случайная величина X называется ***дискретной случайной величиной*** (ДСВ), если множество её возможных значений $\{x_1, x_2, \ldots\}$ конечно или счётно, причём $P(X = x_i) = p_i > 0$, $\sum\limits_i p_i = 1$, где суммирование распространяется на все возможные значения i. Функция распределения в этом случае имеет ступенчатый вид и задаётся формулой $F(x) = \sum\limits_{x_i < x} P(X = x_i)$, где суммирование распространяется на все значения индекса i, для которых $x_i < x$.

Закон распределения ДСВ удобно задавать рядом распределения. ***Рядом распределения*** ДСВ называют таблицу, в которой перечислены все возможные значения $x_1, x_2, \ldots$ этой случайной величины и соответствующие им вероятности $p_1, p_2, \ldots$. Для наглядности закон распределения ДСВ изображают графически, для чего в прямоугольной системе координат строят точки $M_i(x_i, p_i)$ и соединяют их отрезками прямых. Полученную фигуру называют ***многоугольником распределения***.

Математическим ожиданием дискретной случайной величины X называется число $MX=\sum_i x_i p_i = x_1 p_1 + x_2 p_2 + \ldots$, если ряд сходится абсолютно.

Дисперсией случайной величины X называется неотрицательное число $DX = M(X - MX)^2$. Число $\sigma(X)=\sqrt{DX}$ называется ***средним квадратичным отклонением.***

Дисперсию дискретной случайной величины X вычисляют по формулам:

$$DX=\sum_i (x_i - MX)^2 p_i \text{ или } DX=\sum_i x_i^2 p_i - (MX)^2.$$

Пусть C -постоянная величина. Математическое ожидание и дисперсия случайной величины обладают следующими свойствами:

Свойства математического ожидания: **1)** $MC=C$; **2)** $M(CX)=CMX$; **3)** $M(X \pm Y)=MX \pm MY$; **4)** $M(XY)=MX \cdot MY$, если X и Y независимы.

Свойства дисперсии: **1)** $DC=0$; **2)** $D(CX)=C^2 DX$; **3)** $D(X \pm C)=DX$; **4)** $DX=M(X^2)-(MX)^2$; **5)** $D(X \pm Y)=DX+DY$, если X и Y независимы.

Случайная величина X называется *(абсолютно)* ***непрерывной случайной величиной*** (НСВ), если её функция распределения представляется в виде $F(x)=\int_{-\infty}^{x} f(t)dt$, $-\infty < x < +\infty$, где $f(x)$ -неотрицательная и интегрируемая в бесконечных пределах функция, называемая ***функцией плотности*** *(распределения)* ***вероятностей***. Множество возможных значений непрерывной случайной величины несчётно и обычно представляет собой некоторый конечный или бесконечный промежуток числовой прямой.

Функция распределения $F(x)$ непрерывной случайной величины X является непрерывной неубывающей функцией на всей числовой прямой, причём вероятность попадания в любую фиксированную точку равна нулю: $P(X=x)=0$, $-\infty < x < +\infty$.

Функция $f(x)$ является плотностью вероятностей некоторой НСВ X, тогда и только тогда, когда: **1)** $f(x) \geq 0$**;** **2)** $\int_{-\infty}^{+\infty} f(x)dx = 1$**.**

Плотность вероятностей $f(x)$ в точках, где $F(x)$ дифференцируема, определяется равенством: $f(x)=F'(x)$. В точках, где $F(x)$ не дифференцируема, плотность вероятностей $f(x)$, определяется произвольным образом, чаще всего по непрерывности слева или справа.

Для непрерывной случайной величины X с плотностью вероятностей $f(x)$: $P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = P(a < X < b) =$

$$= F(b) - F(a) = \int_a^b f(x)dx .$$

Математическим ожиданием непрерывной случайной величины X называется число $MX = \int_{-\infty}^{+\infty} xf(x)dx$, если интеграл сходится абсолютно.

Дисперсию непрерывной случайной величины X вычисляют по формулам:

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 f(x)dx \text{ или } DX = \int_{-\infty}^{+\infty} x^2 f(x)dx - (MX)^2 .$$

Медианой *непрерывной случайной величины* X называется число $Me(X)$, удовлетворяющее условию $P(X < Me) = P(X > Me)$ или $F(Me) = 0.5$.

Начальным моментом k ***-го порядка*** ($k = 0,1,2,\dots$) распределения случайной величины X (если он существует) называется число $\nu_k = M(X^k)$.

Центральным моментом k ***-го порядка*** ($k = 0,1,2,\dots$) распределения случайной величины X (если он существует) называется число $\mu_k = M(X - MX)^k$.

Для непрерывной случайной величины X начальные и центральные моменты вычисляют по формулам: $\nu_k = \int_{-\infty}^{+\infty} x^k f(x)dx$, $\mu_k = \int_{-\infty}^{+\infty} (x - MX)^k f(x)dx$.

2. <u>Основные законы распределения одномерных случайных величин.</u>

Дискретная случайная величина X имеет ***биномиальное распределение*** $B(n, p)$, если: $P(X = k) = C_n^k p^k q^{n-k}$, $k = \overline{0, n}$. Если $X \sim B(n, p)$, то: $MX = np$, $DX = npq$.

Дискретная случайная величина X имеет ***распределение Пуассона*** $P(\lambda)$ если: $P(X = k) = \dfrac{\lambda^k \cdot e^{-\lambda}}{k!}$ $(\lambda > 0)$, $k = \overline{0, \infty}$. Если $X \sim P(\lambda)$, то: $MX = \lambda$, $DX = \lambda$.

Непрерывная случайная величина X имеет ***равномерное распределение*** $R(a,b)$, если: $f(x)=\begin{cases} 1/(b-a) & x\in(a,b) \\ 0 & x\notin(a,b)\end{cases}$. Если $X\sim R(a,b)$, то:

$MX=\dfrac{a+b}{2}$, $DX=\dfrac{(b-a)^2}{12}$, $P(\alpha\le X<\beta)=\dfrac{\beta-\alpha}{b-a}$.

Непрерывная случайная величина X имеет ***показательное распределение*** $E(\lambda)$, если: $f(x)=\begin{cases} 0 & x<0 \\ \lambda\cdot e^{-\lambda x} & x\ge 0\end{cases}$ $(\lambda>0)$. Если $X\sim E(\lambda)$, то:

$MX=1/\lambda$, $DX=1/\lambda^2$, $P(\alpha\le X<\beta)=e^{-\lambda\alpha}-e^{-\lambda\beta}$.

Непрерывная случайная величина имеет ***нормальное распределение*** $N(a,\sigma)$, если: $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2/2\sigma^2}$, $-\infty<x<+\infty$. Если $X\sim N(a,\sigma)$, то: $MX=a$, $DX=\sigma^2$, $P(|X-a|<\varepsilon)=2\Phi\left(\dfrac{\varepsilon}{\sigma}\right)$,

$P(\alpha\le X<\beta)=\Phi\left(\dfrac{\beta-a}{\sigma}\right)-\Phi\left(\dfrac{\alpha-a}{\sigma}\right)$, где $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-t^2/2}dt$ - функция Лапласа, значения которой находят с помощью специальных таблиц.

3. <u>Многомерные случайные величины</u>.

Под n***-мерной случайной величиной (случайным вектором)*** понимают совокупность случайных величин $X_1(\varpi), X_2(\omega),\ldots,X_n(\omega)$, принимающих свои возможные значения $x_1,x_2,\ldots,x_n$ в зависимости от исхода ω эксперимента, с которым они связаны. Ограничимся рассмотрением двумерных случайных величин $(X,Y)=(X(\omega),Y(\omega))$, $\omega\in\Omega$.

Функцией распределения случайного вектора (X,Y) называется функция $F(x,y)$ действительных переменных x и y, $(x,y)\in R^2$, определяемая формулой $F(x,y)=P(X<x,Y<y)$.

Зная функцию распределения (совместную) вектора (X,Y), можно найти функцию распределения (частную) каждой компоненты:

$$F_X(x)=F(x,+\infty)=\lim_{y\to+\infty}F(x,y),\quad F_Y(x)=F(+\infty,y)=\lim_{x\to+\infty}F(x,y).$$

Случайные величины X и Y называются ***независимыми***, если для всех $(x, y) \in R^2$: $F(x, y) = F_X(x) \cdot F_Y(y)$. В противном случае случайные величины называют ***зависимыми.***

Случайный вектор (X, Y) называется ***дискретным случайным вектором***, если каждая из его компонент является дискретной случайной величиной. Ограничимся рассмотрением дискретных случайных величин X и Y с конечным множеством возможных значений $x_1, x_2, \dots, x_n$ и $y_1, y_2, \dots, y_m$.

Функция распределения $F(x, y)$ дискретного случайного вектора (X, Y) задаётся формулой $F(x, y) = \sum\limits_{x_i < x} \sum\limits_{y_j < y} p_{ij}$, где $p_{ij} = P(X = x_i, Y = y_j) > 0$, $i = \overline{1, n}$, $j = \overline{1, m}$ и суммирование распространяется на все значения индексов i и j для которых $x_i < x$ и $y_j < y$.

Закон распределения дискретного случайного вектора (X, Y) удобно задавать ***таблицей распределения*** *(вероятностей)*, в которой перечислены все возможные пары значений (x_i, y_j), $i = \overline{1, n}$, $j = \overline{1, m}$ компонент вектора и соответствующие им вероятности p_{ij}, причём $\sum\limits_i \sum\limits_j p_{ij} = 1$.

Частные законы распределения $(x_i, p_{i\bullet})$, $i = \overline{1, n}$ и $(y_j, p_{j\bullet})$, $j = \overline{1, m}$ компонент X и Y, где $p_{i\bullet} = P(X = x_i) = \sum\limits_j p_{ij}$, $p_{\bullet j} = P(Y = y_j) = \sum\limits_i p_{ij}$, можно найти, производя в таблице суммирования p_{ij} в каждой строке по столбцам и в каждом столбце по строкам.

Дискретные случайные величины X и Y независимы тогда и только тогда, когда $p_{ij} = p_{i\bullet}\, p_{\bullet j}$, $\forall i, j$, $i = \overline{1, n}$, $j = \overline{1, m}$. В противном случае они зависимы.

Ковариацией (корреляционным моментом) случайных величин X и Y называют число $\operatorname{cov}(X, Y) = M((X - MX)(Y - MY))$. Очевидно, что $\operatorname{cov}(X, X) = DX$. Более удобной для вычисления $\operatorname{cov}(X, Y)$ является формула $\operatorname{cov}(X, Y) = M(XY) - MX \cdot MY$. Для независимых случайных величин X и Y: $\operatorname{cov}(X, Y) = 0$ (***необходимое условие независимости***).

Коэффициентом корреляции случайных величин X и Y называют число

$$\rho = \rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{DX \cdot DY}} = \frac{M(XY) - MX \cdot MY}{\sqrt{DX \cdot DY}}, \text{ где } DX > 0,\ DY > 0.$$

Коэффициент корреляции обладает свойствами: **1)** $|\rho|\le 1$; **2)** $|\rho|=1$ тогда и только тогда, когда X и Y связаны линейной зависимостью $Y=aX+b$, $a\ne 0$; **3)** если X и Y независимы, то $\rho=0$ (***необходимое условие независимости***). Если $\rho(X,Y)=0$, то случайные величины X и Y называют ***некоррелированными***.

Условные законы распределения компоненты X при $Y=y_j$, $j=\overline{1,m}$ (индекс j сохраняет одно и тоже значение при всех возможных значениях X) задают рядами распределения, указывая все возможные значения X и соответствующие им условные вероятности: $(x_i, P(X=x_i \mid Y=y_j))$, $i=\overline{1,n}$. Аналогично задают условные законы распределения компоненты Y при $X=x_i$, $i=\overline{1,n}$: $(y_j, P(Y=y_j \mid X=x_i))$, $j=\overline{1,m}$. Условные вероятности компонент X и Y вычисляют соответственно по формулам:

$$P(X=x_i \mid Y=y_j)=\frac{p_{ij}}{p_{\bullet j}},\quad P(Y=y_j \mid X=x_i)=\frac{p_{ij}}{p_{i\bullet}}.$$

Числовые характеристики MX, MY, $M(XY)$, DX, DY, вычисляют по формулам: $MX=\sum_{i=1}^{n} x_i p_{i\bullet}$, $MY=\sum_{j=1}^{m} y_j p_{\bullet j}$, $M(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij}$,

$$DX=\sum_{i=1}^{n} x_i^2 p_{i\bullet}-(MX)^2, \qquad DY=\sum_{j=1}^{m} y_j^2 p_{\bullet j}-(MY)^2.$$

Условные математические ожидания дискретных случайных величин X *и* Y ***при условиях*** $Y=y_j$ *и* $X=x_i$ определяются соответственно формулами: $M(X \mid Y=y_j)=\sum_{i=1}^{n}\left(x_i \frac{p_{ij}}{p_{\bullet j}}\right)$, $M(Y \mid X=x_i)=\sum_{i=1}^{n}\left(x_i \frac{p_{ij}}{p_{\bullet j}}\right)$.

Вероятность события $\{Z(X,Y)\le A\}$, где A - постоянная величина, вычисляется по формуле $P(Z(X,Y)\le A)=\sum_{(i,j):\, z(x_i,y_j)\le A} p_{ij}$, где суммирование распространяется на все значения индексов i и j для которых $z(x_i,y_j)\le A$.

4. <u>Функции случайных величин</u>.

Случайную величину Y, которая каждому исходу $\omega\in\Omega$ ставит в соответствие число $Y(\omega)=\varphi(X(\omega))$, называют ***функцией от скалярной случайной величины*** X и пишут $Y=\varphi(X)$.

Функция $Y=\varphi(X)$ от дискретной случайной величины X также является дискретной. Если X задана рядом распределения $(x_i, P(X=x_i))$, $i=\overline{1,n}$, то рядом распределения случайной величины $Y=\varphi(X)$ является ряд: $(y_j, P(Y=y_j))$, $j=\overline{1,m}$, $m\le n$, где $y_1, y_2, \ldots, y_m$ - различные числа среди чисел $\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)$, $P(Y=y_j)=\sum\limits_{i:\varphi(x_i)=y_j} P(X=x_i)$ (суммирование распространяется на все значения индекса i для которых $\varphi(x_i)=y_j$).

Функция $Y=\varphi(X)$ от непрерывной случайной величины X может быть как непрерывной, так и дискретной случайной величиной.

Для вычисления числовых характеристик неслучайной функции от случайной величины $Y=\varphi(X)$ можно не знать закон распределения зависящей от X случайной величины Y, а достаточно знать закон распределения случайного аргумента X. Математическое ожидание и дисперсия случайной величины $Y=\varphi(X)$, где дискретная случайная величина X задана рядом распределения (x_i, p_i), $i=\overline{1,n}$, $p_i=P(X=x_i)$, могут быть найдены по формулам:

$$MY=\sum_{i=1}^{n}\varphi(x_i)p_i\,, \qquad DY=\sum_{i=1}^{n}(\varphi(x_i)-MY)^2 p_i=\sum_{i=1}^{n}\varphi^2(x_i)p_i-(MY)^2\,.$$

Тема. Предельные теоремы теории вероятностей.

Если для неотрицательной случайной величины $X\ge 0$ существует математическое ожидание MX, то для всех $\varepsilon>0$ выполняется неравенство:

$$P(X\ge\varepsilon)\le\frac{MX}{\varepsilon}$$ (***первое неравенство Чебышева***).

Если для случайной величины X существует дисперсия DX, то для всех $\varepsilon>0$ выполняется ***второе неравенство Чебышева***:

$$P(|X-MX|\ge\varepsilon)\le\frac{DX}{\varepsilon^2} \qquad \text{или} \qquad P(|X|\ge\varepsilon)\le\frac{M(X^2)}{\varepsilon^2}$$

Второе неравенство Чебышева часто используют в виде:

$$P(|X-MX|<\varepsilon)\ge 1-\frac{DX}{\varepsilon^2}\,, \qquad P(|X|<\varepsilon)\ge 1-\frac{M(X^2)}{\varepsilon^2}\,.$$

Последовательность случайных величин $X_1, X_2, \ldots, X_n, \ldots$ называют ***сходящейся по вероятности*** к случайной величине X (кратко записывается $X_n \underset{n\to\infty}{\overset{P}{\to}} X$), если для всех $\varepsilon>0$: $\lim\limits_{n\to\infty} P(|X_n-X|\ge\varepsilon)=0$.

Говорят, что для последовательности случайных величин $X_1, X_2, \dots, X_n, \dots$, имеющих математические ожидания MX_i, $i = 1,2,\dots$, выполняется ***закон больших чисел***, если $\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{P}{\to}} \frac{1}{n}\sum_{i=1}^{n} MX_i$, т.е. для всех $\varepsilon > 0$ $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| \geq \varepsilon\right) = 0$.

Закон больших чисел в форме Чебышева. Если последовательность независимых случайных величин $X_1, X_2, \dots, X_n, \dots$ такова, что существуют MX_i и DX_i, причём дисперсии этих величин равномерно ограничены (не превышают постоянного числа C), то для неё выполняется закон больших чисел, т.е. $\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{P}{\to}} \frac{1}{n}\sum_{i=1}^{n} MX_i$. В частности, если случайные величины X_i, $i = 1,2,\dots$ являются также одинаково распределёнными (в этом случае $MX_i = a$, $DX_i = \sigma^2$), то $\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{P}{\to}} a$.

Закон больших чисел в форме Бернулли. Если m_n - число успехов в n испытаниях по схеме Бернулли с вероятностью успеха p в отдельном испытании, то $\frac{m_n}{n} \underset{n\to\infty}{\overset{P}{\to}} p$, т.е. для всех $\varepsilon > 0$ $\lim\limits_{n\to\infty} P\left(\left|\frac{m_n}{n} - p\right| \geq \varepsilon\right) = 0$.

Закон больших чисел в форме Бернулли является частным случаем закона больших чисел в форме Чебышева.

Центральная предельная теорема. Пусть $X_1, X_2, \dots, X_n, \dots$ - последовательность независимых одинаково распределённых случайных величин ($MX_i = a$, $DX_i = \sigma^2$, $i = 1,2,\dots$), тогда последовательность нормированных случайных величин $Z_n = \frac{\overline{X}_n - a}{(\sigma/\sqrt{n})}$, где $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, сходится по распределению при $n \to \infty$ к стандартной нормальной величине $Z \sim N(0,1)$, т.е. для всех $x \in R$: $F_{Z_n}(x) = P(Z_n < x) \underset{n\to\infty}{\to} F_Z(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$.

Тема. <u>Основные понятия и задачи математической статистики. Предварительная обработка экспериментальных данных.</u>

Выборкой объёма n из генеральной совокупности X называется совокупность $x_1, x_2, \ldots, x_n$ наблюдаемых значений случайной величины X, соответствующих n независимым повторениям случайного эксперимента с которым связана величина X. В математической статистике генеральную совокупность отождествляют со случайной величиной, совокупность всех возможных значений которой и называют ***генеральной совокупностью***.

Выборка может быть записана в виде вариационного и статистического (дискретного или интервального) рядов. Выборку, записанную в виде статистического ряда, называют ***группированной***.

Вариационным рядом выборки $x_1, x_2, \ldots, x_n$ называется такой способ её записи, при котором элементы выборки упорядочиваются по величине, т.е. записываются в виде последовательности $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, где $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$. Разность $x_{(n)} - x_{(1)} = \widehat{R}$ называется ***размахом выборки***. Всюду в дальнейшем выборочные характеристики будем, как правило, обозначать символом с «$\wedge$» наверху.

Различные значения x_i, $i = \overline{1,k}$ ($k \le n$), называются ***вариантами***. Число n_i повторений варианты x_i в выборке называется её ***частотой***, а отношение $w_i = n_i / n$ называется её ***относительной частотой***.

Дискретным статистическим рядом называется упорядоченная в порядке возрастания значений вариант x_i последовательность пар (x_i, n_i), $i = \overline{1,k}$. Обычно его записывают в виде таблицы, первая стока которой содержит варианты x_i, а вторая их частоты.

Полигоном частот называется фигура, расположенная под ломаной линией с вершинами в точках $M_i(x_i, n_i)$, построенных в прямоугольной системе координат.

Интервальным статистическим рядом называется последовательность пар (J_i, n_i), $i = \overline{1,k}$, где $J_1, J_2, \ldots, J_k$ - непересекающиеся интервалы, как правило, равной длины, объединением которых является отрезок J, содержащий все выборочные значения; n_i - частота интервала J_i, равная числу элементов выборки, значения которых попали в данный интервал. Обычно его записывают в виде таблицы, первая строка которой содержит границы интервалов или их середины $\tilde{x}_i$, а вторая – частоты интервалов.

Гистограммой частот называется ступенчатая фигура, составленная из прямоугольников, построенных на интервалах группировки так, что площадь каждого прямоугольника равна частоте n_i, $i=\overline{1,k}$. Если длины всех интервалов одинаковы и равны h, то высоты прямоугольников равны $\dfrac{n_i}{h}$.

Основные числовые характеристики выборки.

Негруппированная выборка	Группированная выборка
1.Среднее арифметическое выборки	
$\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$	$\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{k}\tilde{x}_i n_i$
2.Дисперсия выборки	
$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$	$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{k}(\tilde{x}_i-\bar{x})^2 n_i$
3.Исправленная дисперсия выборки: $s^2=n\hat{\sigma}^2/(n-1)$	
4. Размах выборки: $\hat{R}=x_{\max}-x_{\min}$	

Тема. <u>Статистические методы оценивания параметров распределений, проверки гипотез и исследования зависимостей.</u>

1. <u>Точечные оценки.</u>

Одной из основных задач математической статистики является ***оценка*** неизвестных параметров, характеризующих распределение генеральной совокупности X. Совокупность независимых случайных величин $X_1, X_2, \ldots, X_n$, каждая из которых имеет то же распределение, что и случайная величина X называют ***случайной выборкой*** объёма n из генеральной совокупности X и обозначают $\vec{X}_n$. Любую функцию $g(X_1, X_2, \ldots, X_n)$ случайной выборки называют ***статистикой.***

Если функция распределения $F_X(x,\theta)$ генеральной совокупности X известна с точностью до параметра θ, то его ***точечной оценкой*** называют статистику $\hat{\theta}(\vec{X}_n)$, значение которой $\hat{\theta}(\vec{x}_n)$ на данной выборке $(x_1,\ldots,x_n)=\vec{x}_n$ принимают за приближённое значение неизвестного параметра θ: $\theta \approx \hat{\theta}(\vec{x}_n)$.

Чтобы точечные оценки давали «хорошее» приближение оцениваемых параметров, они должны удовлетворять определённым требованиям. «Хорошей» считается оценка, обладающая свойствами состоятельности, несмещённости и эффективности. Оценка $\hat{\theta}(\vec{X}_n)$ называется: **1)** ***состоятельной*** оценкой параметра θ, если при неограниченном увеличении объёма выборки она сходится по вероятности к оцениваемому параметру, т.е. $\widehat{\theta}(\vec{X}_n) \underset{n\to\infty}{\overset{P}{\to}} \theta$; **2)** ***несмещённой*** (оценкой без систематических ошибок), если её математическое ожидание при любом n равно оцениваемому параметру, т.е. $M\widehat{\theta}(\vec{X}_n) = \theta$; **3)** ***эффективной*** (в некотором классе несмещённых оценок), если она имеет минимальную дисперсию в этом классе.

Пусть распределение генеральной совокупности X известно с точностью до вектора параметров $\vec{\theta} = (\theta_1, \ldots, \theta_r)$ и требуется найти значение его оценки по выборке $\vec{x}_n$.

Оценкой метода моментов вектора параметров $\vec{\theta} = (\theta_1, \ldots, \theta_r)$ называют статистику $\hat{\vec{\theta}}(\vec{X}_n)$ значение $\hat{\vec{\theta}}(\vec{x}_n) = (\widehat{\theta}_1(\vec{x}_n), ..., \widehat{\theta}_r(\vec{x}_n))$ которой для любой выборки $\vec{x}_n$ удовлетворяет системе уравнений:

$$\nu_k(\widehat{\theta}_1(\vec{x}_n), ..., \widehat{\theta}_r(\vec{x}_n)) = \widehat{\nu}_k(\vec{x}_n),\ k = 1,2, \ldots, r,$$

где $\nu_k(\theta_1, \ldots, \theta_r)$ - теоретические начальные моменты k-го порядка случайной величины X, $\widehat{\nu}_k(\vec{x}_n) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ - эмпирические начальные моменты k-го порядка выборки $\vec{x}_n$. В систему уравнений метода моментов могут входить и уравнения вида $\mu_k(\widehat{\theta}_1(\vec{x}_n), ..., \widehat{\theta}_r(\vec{x}_n)) = \widehat{\mu}_k(\vec{x}_n)$, где $\mu_k(\theta_1, \ldots, \theta_r)$ - теоретические центральные моменты k-го порядка случайной величины X, $\widehat{\mu}_k(\vec{x}_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$ эмпирические центральные моменты k-го порядка выборки $\vec{x}_n$. Часто для нахождения значения оценки одного параметра используют первый начальный момент. Для нахождения значений оценок двух параметров используют первый начальный и второй центральный моменты.

Оценкой метода максимального правдоподобия вектора параметров $\vec{\theta} = (\theta_1, \ldots, \theta_r)$ называют статистику $\hat{\vec{\theta}}(\vec{X}_n)$, значение $\hat{\vec{\theta}}(\vec{x}_n)$ которой для любой выборки $\vec{x}_n$ удовлетворяет условию: $L(\vec{x}_n, \hat{\vec{\theta}}) = \max_{\vec{\theta}\in\Theta} L(\vec{x}_n, \vec{\theta})$, где $L(\vec{x}_n, \vec{\theta})$ - функция правдоподобия выборки $\vec{x}_n$, Θ - множество всех возможных значений вектора параметров $\vec{\theta}$.

Функция правдоподобия имеет вид:

1) $L(\vec{x}_n,\vec{\theta})=\prod\limits_{i=1}^{n}P(X=x_i,\vec{\theta})$ - для дискретной случайной величины X;

2) $L(\vec{x}_n,\vec{\theta})=\prod\limits_{i=1}^{n}f(x_i,\vec{\theta})$ - для непрерывной случайной величины X.

Если функция $L(\vec{x}_n,\vec{\theta})$ дифференцируема как функция аргумента $\vec{\theta}=(\theta_1,\ldots,\theta_r)$ для любой выборки $\vec{x}_n$ и максимум $L(\vec{x}_n,\vec{\theta})$ достигается во внутренней точке Θ, то значение точечной оценки $\hat{\vec{\theta}}(\vec{x}_n)$ максимального правдоподобия находят, решая систему уравнений максимального правдоподобия: $\dfrac{\partial L(\vec{x}_n,\theta_1,\ldots,\theta_r)}{\partial\theta_k}=0$, $k=\overline{1,r}$. Нахождение $\hat{\vec{\theta}}(\vec{x}_n)$ упрощается, если максимизировать не саму функцию правдоподобия, а её логарифм $\ln L(\vec{x}_n,\vec{\theta})$, так как при логарифмировании точки экстремума остаются теми же, а уравнения, как правило, упрощаются и записываются в виде: $\dfrac{\partial\ln L(\vec{x}_n,\theta_1,\ldots,\theta_r)}{\partial\theta_k}=0$, $k=\overline{1,r}$.

2. <u>Интервальные оценки.</u>

Если функция распределения $F_X(x,\theta)$ генеральной совокупности X известна с точностью до параметра θ, то его ***интервальной оценкой*** или ***доверительным интервалом*** называется случайный интервал $(\hat{\theta}_1(\vec{X}_n),\hat{\theta}_2(\vec{X}_n))$, который накрывает неизвестное значение параметра θ с заданной вероятностью $0<\gamma<1$, т.е. $P(\hat{\theta}_1<\theta<\hat{\theta}_2)=\gamma$. Число $\gamma=1-\alpha$ называется ***доверительной вероятностью,*** а число α - ***уровнем значимости***. Обычно используются значения γ, равные 0.90, 0.95, 0.99.

Точность интервальной оценки характеризуется длиной $\hat{\theta}_2(\vec{X}_n)-\hat{\theta}_1(\vec{X}_n)$ доверительного интервала и зависит от объёма n выборки и доверительной вероятности γ. Очевидно, что, чем меньше длина доверительного интервала, тем точнее оценка. Доверительный интервал, симметричный относительно точечной оценки $\hat{\theta}(\vec{X}_n)$, определяется формулой $P(|\hat{\theta}(\vec{X}_n)-\theta|<\Delta)=\gamma$ и имеет вид $(\hat{\theta}-\Delta,\hat{\theta}+\Delta)$, где Δ характеризует отклонение выборочного значения параметра от его истинного значения и называется ***предельной ошибкой выборки.*** Доверительные интервалы часто строятся в предположении, что выборка получена из генеральной совокупности, имеющей нормальное распределение.

Доверительный интервал для параметра a нормально распределённой генеральной совокупности.

Параметр	Точечная оценка	Доверительный интервал
a (σ^2 неизвестна)	$\overline{X}$	$\overline{X} - \Delta < a < \overline{X} + \Delta$, где $\Delta = \dfrac{\widehat{\sigma} t_{\alpha}^{дв}(n-1)}{\sqrt{n-1}}$,

Здесь: $t_{\alpha}^{дв}(k)$, где $k = n-1$ - двусторонняя критическая точка распределения Стьюдента (находится с помощью специальных таблиц).

3. <u>Проверка статистических гипотез.</u>

Статистической гипотезой H называют любое предположение относительно параметров или вида распределения генеральной совокупности (случайной величины) X. Гипотезы относительно неизвестного значения параметра распределения генеральной совокупности (случайной величины) называются ***параметрическими*** и непараметрическими в иных случаях. Статистическая гипотеза называется ***простой***, если она однозначно определяет распределение X, в противном случае она называется ***сложной***. Проверяемая гипотеза называется ***основной*** и обозначается H_0. Наряду с гипотезой H_0 рассматривают одну из альтернативных гипотез H_1, противоречащих основной. Например, если проверяется гипотеза о равенстве параметра θ распределения X некоторому заданному значению θ_0, т.е. $H_0 : \theta = \theta_0$, то в качестве альтернативной гипотезы, как правило, рассматривается одна из следующих гипотез: $\theta > \theta_0$, $\theta < \theta_0$, $\theta \neq \theta_0$. Выбор альтернативы определяется конкретной постановкой задачи.

Правило, по которому принимается решение принять или отклонить основную гипотезу H_0, называется ***критерием*** K проверки гипотезы. Критерий K задают с помощью критического множества $W \subset \chi_n$, где $\chi_n \subset R^n$ - ***выборочное пространство*** (множество всех возможных значений случайной выборки $\vec{X}_n$). Решение принимают на основе выборки $\vec{x}_n = (x_1, x_2, \ldots, x_n)$ наблюдаемых значений случайной величины X, используя для этого подходящую статистику $U(\vec{X}_n)$, называемую ***статистикой критерия*** K. При проверке параметрической гипотезы $H_0 : \theta = \theta_0$ в качестве статистики критерия выбирают ту же статистику, что и при оценивании параметра θ.

Решение принимают следующим образом: **1)** если выборка $\vec{x}_n \notin W$, то принимают основную гипотезу H_0; **2)** если выборка $\vec{x}_n \in W$, то основную гипотезу H_0 отклоняют и принимают альтернативную гипотезу H_1.

При использовании любого критерия возможны ошибки двух видов:
1) отклонить верную основную гипотезу H_0 - ***ошибка первого рода***;
2) принять неверную основную гипотезу H_0 - ***ошибка второго рода***.

Вероятности совершения ошибок первого и второго рода обозначают α и β: $\alpha = P(\vec{X}_n \in W \mid H_0)$, $\beta = P(\vec{X}_n \notin W \mid H_1)$, где $P(A \mid H_j)$ - вероятность события A при условии, что справедлива гипотеза H_j, $j = 0,1$. Вероятность совершения ошибки первого рода α называют также ***уровнем значимости*** критерия K, а величину $(1-\beta)$, равную вероятности отклонить основную гипотезу H_0, когда она неверна, называют ***мощностью критерия***. Уровень значимости α определяет «размер» критического множества. Обычно используются значения α , равные $0.1, 0.05, 0.01$.

Проверка статистической гипотезы H_0 основывается на принципе, в соответствии с которым маловероятные события считаются невозможными, т.е. если выборка $\vec{x}_n$ попадает в критическое множество W с исключительно малой вероятностью, то естественно предположить, что утверждение, которое привело к этому маловероятному событию, не соответствует истине и отклонить его. Поступая так, мы будем отклонять в действительности верную основную гипотезу H_0 крайне редко – не более чем в $100\alpha\%$ случаев. Поэтому за основную гипотезу естественно принять утверждение, отклонение которого, когда оно в действительности является верным, приводит к более тяжёлым последствиям, чем его принятие при справедливости альтернативы.

Общая ***схема проверки параметрической гипотезы*** $H_0 : \theta = \theta_0$ состоит в следующем: **1)** формулируется альтернативная гипотеза H_1; **2)** задаётся уровень значимости α ; **3)** выбирается статистика $U(\vec{X}_n)$ критерия K проверки гипотезы H_0; **4)** определяется выборочное распределение статистики $U(\vec{X}_n)$ при условии, что гипотеза H_0 является верной; **5)** по заданным значениям α и n определяется критическое множество W критерия K в зависимости от формулировки альтернативной гипотезы H_1; **6)** по выборке $\vec{x}_n$ вычисляется наблюдаемое значение $u_{набл} = U(\vec{x}_n)$ статистики критерия; **7)** принимается статистическое решение: если $\vec{x}_n \in W$, то основная гипотеза H_0 отклоняется как не согласующаяся с данными выборки; если $\vec{x}_n \notin W$, то H_0 принимается, т.е. считается, что гипотеза H_0 не противоречит данным выборки.

Критерии, используемые для проверки гипотезы H_0 о виде распределения случайной величины (генеральной совокупности) X называют ***критериями согласия*** (с основной гипотезой), при этом альтернатива H_1, как правило, не формулируется, подразумевая под ней «всё остальное». Одним из наиболее широко применяемых на практике критериев согласия, является критерий согласия χ^2 («хи-квадрат»).

Критерий «хи-квадрат» в качестве меры расхождения эмпирического и теоретического законов распределения случайной величины X использует значения статистики $\chi^2(\vec{X}_n) = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}$, где n - объём выборки; r - число непересекающихся множеств Δ_i на которые разбита область возможных значений случайной величины X; n_i -эмпирическая частота попадания X в Δ_i; $p_i = P(X \in \Delta_i)$ -вероятность попадания X в Δ_i, вычисленная для теоретического закона распределения X. Закон распределения статистики $\chi^2(\vec{X}_n)$ при $n \to \infty$ независимо от вида закона распределения случайной величины X стремится к закону χ^2-распределения с $k = r - l - 1$ степенями свободы (l -число параметров теоретического закона распределения $F(x, \theta_1, \theta_2, \ldots)$, вычисляемых по выборке). Для его применения практически достаточно, чтобы $n \geq 50 \div 60$.

Общая схема проверки непараметрической гипотезы H_0, утверждающей, что случайная величина X имеет теоретический закон распределения $F(x, \theta_1, \theta_2, \ldots)$, состоит в следующем.

1) Задают уровень значимости α.

2) По выборке $\vec{x}_n$ находят значения оценок $\hat{\theta}_1(\vec{x}_n), \hat{\theta}_2(\vec{x}_n), \ldots$ неизвестных параметров предполагаемого закона распределения X.

3) Множество возможных значений случайной величины X разбивают на r непересекающихся множеств $\Delta_1, \Delta_2, \ldots, \Delta_r$: r интервалов, если X - непрерывная величина или r групп отдельных значений, если X - дискретная величина, и подсчитывают их частоты n_i, $i = 1, 2, \ldots, r$.

4) Используя предполагаемый закон распределения X вычисляют вероятности $p_i = P(X \in \Delta_i)$, $i = 1,2,\ldots,r$ - вероятности того, что наблюдаемое значение X принадлежит множеству Δ_i. <u>***Замечание.***</u> *Критерий «хи-квадрат» использует тот факт, что случайные величины* $\dfrac{n_i - np_i}{\sqrt{np_i}}$, $i = 1,2,\ldots,r$, *имеют распределения, близкие к нормальному* $N(0,1)$. *Чтобы это утверждение было достаточно точным, необходимо, чтобы для всех* Δ_i *выполнялось условие* $np_i \geq 5$. *Если для некоторых* Δ_i *это условие не выполняется, то их объединяют с соседними.*
5) По заданным значениям α и n определяют критическое множество W критерия «хи-квадрат»: $W = \{\vec{x}_n \mid \chi^2(\vec{x}_n) \geq \chi^2_\alpha(k)\}$, $k = r - l - 1$, где $\chi^2_\alpha(k)$ - критическая точка χ^2-распределения (находится с помощью специальных таблиц). <u>***Замечание.***</u> *Если проводилось объединение* Δ_i, *то* r *- число множеств* Δ_i, *оставшихся после их объединения.*

6) По выборке $\vec{x}_n$ вычисляют наблюдаемое значение $\chi^2_{набл} = \sum\limits_{i=1}^{r} \dfrac{(n_i - np_i)^2}{np_i}$ статистики критерия «хи-квадрат».
7) Принимают решение: если $\vec{x}_n \in W$, то основная гипотеза H_0 отклоняется как не согласующаяся с данными выборки; если $\vec{x}_n \notin W$, то H_0 принимается, т.е. считается, что гипотеза H_0 не противоречит данным выборки.

4. <u>Корреляционно-регрессионный анализ.</u>

На практике часто бывает важно знать, существует ли зависимость между некоторыми наблюдаемыми величинами, насколько тесно они связаны между собой, можно ли по значению одной величины сделать какие-либо выводы о предполагаемом значении другой величины и т.д. Для решения задач такого рода и применяется корреляционно-регрессионный анализ.

Пусть $(x_1, y_1), \ldots, (x_n, y_n)$ - выборка из двумерной генеральной совокупности (X, Y). Предварительное представление о зависимости между случайными величинами X и Y можно получить, изобразив в прямоугольной системе координат на плоскости точки (x_i, y_i). Такое графическое представление двумерной выборки называют ***диаграммой рассеивания*** (***корреляционным полем***). Количественной характеристикой степени линейной зависимости между величинами X и Y является ***коэффициент корреляции***

$\rho = \dfrac{M(XY) - MX \cdot MY}{\sqrt{DX \cdot DY}}$. Его состоятельной оценкой служит статистика $r = \dfrac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\hat{\sigma}_X \cdot \hat{\sigma}_Y}$, где $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \dfrac{1}{n}\sum_{i=1}^{n} Y_i$, $\overline{XY} = \dfrac{1}{n}\sum_{i=1}^{n} X_i Y_i$, $\hat{\sigma}_X^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $\hat{\sigma}_Y^2 = \dfrac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

Если $|\rho| = 1$, то все выборочные точки (x_i, y_i), $i = 1, 2, \ldots, n$ лежат на одной прямой. При $|\rho| < 1$ выборочные данные только имеют тенденцию сосредотачиваться около прямых: $y = a_Y + \rho \dfrac{\sigma_Y}{\sigma_X}(x - a_X)$, $x = a_X + \rho \dfrac{\sigma_X}{\sigma_Y}(y - a_Y)$, называемых (*теоретическими) **прямыми регрессии*** Y на X и X на Y, соответственно. Здесь $a_X = MX$, $a_Y = MY$. Первое уравнение даёт наилучший в среднем квадратичном прогноз ожидаемых значений Y по наблюдениям X, второе – прогноз значений X по наблюдениям Y.

Прямые $y = \bar{y} + r\dfrac{\hat{\sigma}_y}{\hat{\sigma}_x}(x - \bar{x})$, $x = \bar{x} + r\dfrac{\hat{\sigma}_x}{\hat{\sigma}_y}(y - \bar{y})$ называются ***эмпирическими прямыми регрессии*** Y на X и X на Y, соответственно. Здесь $\bar{x}$, $\bar{y}$, $\hat{\sigma}_x^2$, $\hat{\sigma}_y^2$, $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{\sigma}_x \cdot \hat{\sigma}_y}$ - найденные по выборке (x_i, y_i), $i = \overline{1, n}$, значения статистик $\bar{X}$, $\bar{Y}$, $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$, r, являющихся состоятельными оценками параметров a_X, a_Y, σ_X^2, σ_Y^2, ρ двумерной генеральной совокупности.

Проверка гипотезы о значимости выборочного коэффициента корреляции r.

Гипотеза H_0	Статистика критерия	Критическое множество
$\rho = 0$	$T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	$W_{\rho \neq 0} = \left\{ \lvert T \rvert > t_\alpha^{дв}(k) \right\}$, где $k = n - 2$

Здесь: $t_\alpha^{дв}(k)$ - двусторонняя критическая точка распределения Стьюдента (находится с помощью специальных таблиц), n .- объём выборки.

6.3 Основные математические формулы.

Формулы сокращённого умножения:

1. $(a \pm b)^2 = a^2 \pm 2ab + b^2$ **2.** $(a-b)(a+b) = a^2 - b^2$

3. $(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac$

4. $(a \pm b)^3 = a^3 \pm 3a^2 b + 3ab^2 \pm b^3$ **5.** $(a \pm b)(a^2 \mp ab + b^2) = a^3 \pm b^3$

Формулы тригонометрии:

1. $\sin^2\alpha + \cos^2\alpha = 1,$ **2.** $tg\alpha \cdot ctg\alpha = 1,$

3. $1 + tg^2\alpha = 1/\cos^2\alpha$, **4.** $1 + ctg^2\alpha = 1/\sin^2\alpha$ **.**

5. $\sin(\alpha \pm \beta) = \sin\alpha \cdot \cos\beta \pm \cos\alpha \cdot \sin\beta$

6. $\cos(\alpha \pm \beta) = \cos\alpha \cdot \cos\beta \mp \sin\alpha \cdot \sin\beta$

7. $\sin 2\alpha = 2\sin\alpha \cdot \cos\alpha$ **8.** $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$

9. $\cos\alpha + \cos\beta = 2\cos[(\alpha+\beta)/2] \cdot \cos[(\alpha-\beta)/2]$

10. $\cos\alpha - \cos\beta = -2\sin[(\alpha+\beta)/2] \cdot \sin[(\alpha-\beta)/2]$

11. $\sin\alpha + \sin\beta = 2\sin[(\alpha+\beta)/2] \cdot \cos[(\alpha-\beta)/2]$

12. $\sin\alpha - \sin\beta = 2\sin[(\alpha-\beta)/2] \cdot \cos[(\alpha+\beta)/2]$

13. $\cos^2\alpha = \frac{1}{2}(1 + \cos 2\alpha)$ **14.** $\sin^2\alpha = \frac{1}{2}(1 - \cos 2\alpha)$

Формулы приведения.

Функция	$\beta = \frac{\pi}{2} \pm \alpha$	$\beta = \pi \pm \alpha$	$\beta = \frac{3\pi}{2} \pm \alpha$	$\beta = 2\pi - \alpha$
$\sin\beta$	$+\cos\alpha$	$\mp\sin\alpha$	$-\cos\alpha$	$-\sin\alpha$
$\cos\beta$		$-\cos\alpha$	$\pm\sin\alpha$	$+\cos\alpha$
$tg\beta$	$\mp ctg\alpha$	$\pm tg\alpha$	$\mp ctg\alpha$	$-tg\alpha$
$-tg\beta$	$\mp tg\alpha$	$\pm ctg\alpha$	$\mp tg\alpha$	$-ctg\alpha$

Значения тригонометрических функций некоторых углов.

α	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	π	$3\pi/2$	2π
$\sin\alpha$	0	$1/2$	$\sqrt{2}/2$	$\sqrt{3}/2$	1	$\sqrt{3}/2$	0	-1	0
$\cos\alpha$	1	$\sqrt{3}/2$	$\sqrt{2}/2$	$1/2$	0	$-1/2$	-1	0	1
$tg\alpha$	0	$1/\sqrt{3}$	1	$\sqrt{3}$	∞	$-\sqrt{3}$	0	∞	0
$ctg\alpha$	∞	$\sqrt{3}$	1	$1/\sqrt{3}$	0	$-1/\sqrt{3}$	∞	0	∞

Таблица производных и дифференциалов основных элементарных функций.

№ п/п	$f(x)$	$f'(x)$	$df(x)$
1	$x^{\alpha} \quad (\alpha \neq 0)$	$\alpha x^{\alpha-1}$	$\alpha x^{\alpha-1}dx$
2	$a^{x} \quad (a>0, \neq 1)$	$a^{x}\ln a$	$a^{x}\ln a dx$
3	e^{x}	e^{x}	$e^{x}dx$
4	$\log_{a} x \quad (a>0, \neq 1)$	$\frac{1}{x\ln a}$	$\frac{dx}{x\ln a}$
5	$\ln x$	$\frac{1}{x}$	$\frac{dx}{x}$
6	$\sin x$	$\cos x$	$\cos xdx$
7	$\cos x$	$-\sin x$	$-\sin xdx$
8	tgx	$\frac{1}{\cos^{2} x}$	$\frac{dx}{\cos^{2} x}$
9	$ctgx$	$-\frac{1}{\sin^{2} x}$	$-\frac{dx}{\sin^{2} x}$
10	$\arcsin x$	$\frac{1}{\sqrt{1-x^{2}}}$	$\frac{dx}{\sqrt{1-x^{2}}}$
11	$\arccos x$	$-\frac{1}{\sqrt{1-x^{2}}}$	$-\frac{dx}{\sqrt{1-x^{2}}}$
12	$arctgx$	$\frac{1}{1+x^{2}}$	$\frac{dx}{1+x^{2}}$
13	$arcctgx$	$-\frac{1}{1+x^{2}}$	$-\frac{dx}{1+x^{2}}$
14	chx	shx	$shxdx$
15	shx	chx	$chxdx$
16	thx	$\frac{1}{ch^{2}x}$	$\frac{dx}{ch^{2}x}$
17	$cthx$	$-\frac{1}{sh^{2}x}$	$-\frac{dx}{sh^{2}x}$

Таблица основных неопределенных интегралов.

№ п/п	$\int f(x)dx$	№ п/п	$\int f(x)dx$
1	$\int x^k dx = \frac{x^{k+1}}{k+1} + C$ $(k \in R, k \neq -1)$	**2**	$\int (x+a)^k dx = \frac{(x+a)^{k+1}}{k+1} + C$ $(k \in R, k \neq -1)$
3	$\int \frac{dx}{x} = \ln\lvert x \rvert + C$	**4**	$\int \frac{dx}{x+a} = \ln\lvert x+a \rvert + C$
5	$\int a^x dx = \frac{a^x}{\ln a} + C$ $(a > 0, a \neq 1)$	**6**	$\int e^x dx = e^x + C$
7	$\int \sin x dx = -\cos x + C$	**8**	$\int \cos x dx = \sin x + C$
9	$\int \frac{dx}{\cos^2 x} = tgx + C$	**10**	$\int \frac{dx}{\sin^2 x} = -ctgx + C$
11	$\int \frac{dx}{\sin x} = \ln\left\lvert tg\frac{x}{2} \right\rvert + C$	**12**	$\int \frac{dx}{\cos x} = \ln\left\lvert tg\left(\frac{x}{2} + \frac{\pi}{4}\right) \right\rvert + C$
13	$\int \frac{dx}{x^2 + a^2} = \frac{1}{a} arctg\frac{x}{a} + C$	**14**	$\int \frac{dx}{x^2 - a^2} = \frac{1}{2a} \ln\left\lvert \frac{x-a}{x+a} \right\rvert + C$
15	$\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin\frac{x}{a} + C$	**16**	$\int \frac{dx}{\sqrt{x^2 \pm a^2}} = \ln\left\lvert x + \sqrt{x^2 \pm a^2} \right\rvert + C$
17	$\int shxdx = chx + C$	**18**	$\int chxdx = shx + C$
19	$\int \frac{dx}{ch^2 x} = thx + C$	**20**	$\int \frac{dx}{sh^2 x} = -cthx + C$
21	$\int \sqrt{a^2 - x^2}\, dx = \frac{a^2}{2} \cdot \arcsin\frac{x}{a} + \frac{x}{2} \cdot \sqrt{a^2 - x^2} + C$		
22	$\int \sqrt{x^2 \pm a^2}\, dx = \pm\frac{a^2}{2} \ln\lvert x + \sqrt{x^2 \pm a^2} \rvert + \frac{x}{2}\sqrt{x^2 \pm a^2} + C$		

6.4 Образец оформления обложки с контрольной работой.

Федеральное государственное автономное образовательное учреждение высшего профессионального образования

«Набережночелнинский институт
Казанского (Приволжского) федерального университета»

кафедра математики

Контрольная работа

по дисциплине «____________________»

Вариант № ____
(номера выполняемых заданий: __________________________)

Выполнил: студент группы № _______
***Ф.И.О. студента*__________**
зач. книжка - № _________
Проверил: преподаватель кафедры математики
***Ф.И.О. преподавателя*_____**

Набережные Челны
201…

6.5. Таблица номеров выполняемых заданий.

Номер варианта	*Номера выполняемых заданий*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	**1**	**11**	**21**	**31**	**41**	**51**	**61**	**71**	**81**	**91**
2	**2**	**12**	**22**	**32**	**42**	**52**	**62**	**72**	**82**	**92**
3	**3**	**13**	**23**	**33**	**43**	**53**	**63**	**73**	**83**	**93**
4	**4**	**14**	**24**	**34**	**44**	**54**	**64**	**74**	**84**	**94**
5	**5**	**15**	**25**	**35**	**45**	**55**	**65**	**75**	**85**	**95**
6	**6**	**16**	**26**	**36**	**46**	**56**	**66**	**76**	**86**	**96**
7	**7**	**17**	**27**	**37**	**47**	**57**	**67**	**77**	**87**	**97**
8	**8**	**18**	**28**	**38**	**48**	**58**	**68**	**78**	**88**	**98**
9	**9**	**19**	**29**	**39**	**49**	**59**	**69**	**79**	**89**	**99**
10	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
11	**3**	**12**	**21**	**32**	**43**	**52**	**61**	**72**	**83**	**92**
12	**4**	**13**	**22**	**33**	**44**	**53**	**62**	**73**	**84**	**93**
13	**5**	**16**	**27**	**38**	**49**	**60**	**69**	**78**	**87**	**96**
14	**6**	**17**	**28**	**39**	**50**	**59**	**68**	**77**	**86**	**95**
15	**7**	**18**	**29**	**40**	**49**	**58**	**67**	**76**	**85**	**94**
16	**8**	**19**	**30**	**39**	**48**	**57**	**66**	**75**	**84**	**93**
17	**9**	**20**	**29**	**38**	**47**	**56**	**65**	**74**	**83**	**92**
18	**1**	**12**	**23**	**34**	**45**	**56**	**67**	**78**	**89**	**100**
19	**2**	**13**	**24**	**35**	**46**	**57**	**68**	**79**	**90**	**99**
20	**3**	**14**	**25**	**36**	**47**	**58**	**69**	**80**	**89**	**98**
21	**4**	**15**	**26**	**37**	**48**	**59**	**70**	**79**	**88**	**97**
22	**5**	**14**	**23**	**32**	**41**	**52**	**63**	**74**	**85**	**96**
23	**6**	**15**	**24**	**33**	**42**	**51**	**62**	**73**	**84**	**95**
24	**7**	**16**	**25**	**34**	**43**	**52**	**61**	**72**	**83**	**94**
25	**8**	**17**	**26**	**35**	**44**	**53**	**62**	**71**	**82**	**93**
26	**9**	**18**	**27**	**36**	**45**	**54**	**63**	**72**	**81**	**92**
27	**10**	**19**	**28**	**37**	**46**	**55**	**64**	**73**	**82**	**91**
28	**2**	**11**	**22**	**33**	**44**	**55**	**66**	**77**	**88**	**99**
29	**3**	**12**	**21**	**32**	**43**	**54**	**65**	**76**	**87**	**98**
30	**4**	**13**	**22**	**31**	**42**	**53**	**64**	**75**	**86**	**97**

Номер варианта соответствует номеру студента в списке группы.

С О Д Е Р Ж А Н И Е