Министерство образования Республики Беларусь
Белорусский государственный университет
Механико-математический факультет
Кафедра общей математики и информатики

СОГЛАСОВАНО	СОГЛАСОВАНО
Заведующий кафедрой	Декан факультета
______________ Самаль С.А.	_____________ Босяков С.М.
«13» декабря 2023 г.	«26» декабря 2023 г.

Основы математической статистики

Электронный учебно-методический комплекс для специальностей
1-21 03 02 «Регионоведение»,
6-05-0222- 02 «Регионоведение»

Регистрационный № 2.4.2-24/394

Авторы:
Матейко О. М., кандидат физико-математических наук, доцент;
Велько О. А., старший преподаватель.

Рассмотрено и утверждено на заседании Научно-методического совета БГУ 18.01.2024г., протокол № 5.

Минск 2023

УДК 519.22(075.8)
М 34

Утверждено на заседании Научно-методического совета БГУ
Протокол № 5 от 18.01.2024 г.

Решение о депонировании вынес:
Совет механико-математического факультета
Протокол № 6 от 26.12.2023 г.

А в т о р ы:

Матейко Олег Михайлович, доцент кафедры общей математики и информатики механико-математического факультета БГУ, кандидат физико-математических наук, доцент,

Велько Оксана Александровна, старший преподаватель кафедры общей математики и информатики механико-математического факультета БГУ.

Рецензенты:

кафедра экономики и переводческого дела Института предпринимательской деятельности (заведующий кафедрой экономики и переводческого дела, кандидат философских наук, доцент Г.В. Бороздина);

Сташулёнок С.П., доцент кафедры функционального анализа и аналитической экономики механико-математического факультета БГУ, кандидат физико-математических наук, доцент.

Матейко, О.М. Основы математической статистики : электронный учебно-методический комплекс для специальностей: 1-21 03 02, 6-05-0222-02 «Регионоведение» / О.М. Матейко, О.А. Велько ; БГУ, Механико-математический фак., Каф. общей математики и информатики. – Минск : БГУ, 2023. – 147 с. : ил. – Библиогр.: 145–147 с.

Электронный учебно-методический комплекс (ЭУМК) по учебной дисциплине «Основы математической статистики» предназначен для студентов специальностей 1-21 03 02 «Регионоведение», 6-05-0222- 02 «Регионоведение». В ЭУМК содержатся лекционный материал, задания для практических занятий, примерные задания для управляемой самостоятельной работы студентов, тестовые задания, контрольные работы, вопросы для подготовки к зачёту, примерный тематический план, содержание учебного материала, список литературы.

СОДЕРЖАНИЕ

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Электронный учебно-методический комплекс (ЭУМК) по учебной дисциплине «Основы математической статистики» предназначен для студентов 1 курса специальности 6-05-0222-02 Регионоведение.

Комплекс подготовлен в соответствии с требованиями Положения об учебно-методическом комплексе на уровне высшего образования, утвержденного Постановлением министерства образования Республики Беларусь от 26.07.2011 № 167.

Содержание разделов ЭУМК соответствует образовательным стандартам по дисциплине «Основы математической статистики».

Главные цели ЭУМК: помощь студентам в организации самостоятельной работы, повышение качества подготовки и усиление практико-ориентированности учебного процесса по дисциплине «Основы математической статистики».

ЭУМК состоит из следующих разделов.

Теоретический. Включает аннотацию учебного пособия, написанного в соответствии с программой дисциплины. Материал данного пособия, наряду с конспектом лекций, может быть использован для самостоятельной подготовки студентов к практическим занятиям, контрольным работам и зачёту.

Практический. Содержит задания практических занятий. Данные материалы используются для подготовки к практическим занятиям и их организации, для самостоятельной работы над курсом.

Раздел контроля знаний представлен вопросами к зачету, тренировочными тестами и промежуточными контрольными работами. Описаны формы диагностики.

Вспомогательный раздел включает рекомендуемую литературу и содержание учебной программы курса по отдельным темам, на основе которой построено изучение дисциплины и контроль знаний.

Место учебной дисциплины в системе подготовки специалиста. Учебная дисциплина «Основы математической статистики» относится к модулю «Естественнонаучные дисциплины и информационные технологии» компонента учреждения образования.

Дисциплина «Основы математической статистики» основана на школьной учебной дисциплине «Математика» и необходима для изучения следующих учебных дисциплин: «Экспертно-аналитическая деятельность» и «Историко-социологические методы в регионоведении», формирующих навыки работы с профессиональной информацией. Кроме того, практические навыки, полученные при изучении дисциплины, будут полезны студентам при написании курсовых и дипломной работ, проведении исследовательских проектов, а также в самообразовании.

Цель учебной дисциплины – повышение уровня математической подготовки студентов и ориентация их на использование методов математической статистики в профессиональной деятельности, подготовка к использованию современного аппарата математической статистики в качестве эффективного инструмента для решения научных и практических задач в области регионоведения.

Задачи учебной дисциплины:

– сформировать умение корректной математической постановки прикладных задач, способствовать дальнейшему развитию у студентов способностей к логическому и критическому мышлению;

– обучить студентов основным понятиям и методам математической статистики;

– подготовить будущего специалиста-регионоведа к самостоятельному изучению тех разделов современной математики, теории вероятностей и математической статистики, которые могут потребоваться дополнительно в его практической и научно-исследовательской работе.

В результате изучения учебной дисциплины студент должен *знать*:

– природу математических абстракций, элементы теории множеств и возможности их использования в социально-экономической и общественно-политической сферах;

– основные методы теории вероятностей и математической статистики, которые используются для решения задач в профессиональной деятельности;

Студенты должны *уметь*:

– использовать математический язык и аппарат для описания явлений и закономерностей при изучении стран и регионов;

– пользоваться математическими методами для формализации отдельных информационных задач и процессов в регионоведении;

– делать оценки правдоподобности информации, основанной на количественных параметрах и соотношениях;

Студенты должны *владеть*:

– терминологией дисциплины «Основы математической статистики»;

– навыками вычисления вероятностей событий при решении прикладных задач;

– навыками использования элементов математической статистики при обработке и систематизации регионоведческой информации.

Освоение учебной дисциплины «Основы математической статистики» должно обеспечить формирование следующих универсальных и специализированных компетенций:

Универсальные компетенции:

УК-2: решать стандартные задачи профессиональной деятельности на основе информационно-коммуникационных технологий.

Специализированные компетенции:

СК-2: подбирать и систематизировать страновую и регионоведческую информацию, подготавливать аналитические материалы и рекомендацим для государственных, коммерческих и иных организаций по различным аспектам социально-экономического, общественно-политического и духовно-культурного развития.

Дисциплина изучается в 1 семестре. Всего на изучение учебной дисциплины «Основы математической статистики» отведено:

– для очной формы получения высшего образования – 100 часов, в том числе 50 аудиторных часов, из них: лекции – 18 часов, практические занятия– 24 часа, управляемая самостоятельная работа 8 часов (из них – 2ч/ДОТ).

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – зачет.

1. ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

1.1. ЭЛЕМЕНТЫ ТЕОРИИ МНОЖЕСТВ

1.1.1. Основные понятия теории множеств. Диаграммы Эйлера – Венна

Начиная что-то рассматривать или изучать, возникает необходимость кратко описать те объекты, с которыми будет происходить работа. Например, мы говорим: «граждане Республики Беларусь», «европейские страны», «планеты Солнечной системы», «факультеты Белорусского государственного университета», «буквы алфавита русского языка», «натуральные числа». При этом каждый раз рассматривается некоторая совокупность объектов, обладающих общими свойствами. Математиков интересуют множества чисел, функций, геометрических фигур и т.д. Однако следует заметить, что элементами множеств могут быть объекты различной природы, поэтому любое утверждение, сформулированное в теории множеств, верно также для произвольного множества.

Построение любой теории в математике начинается с введения основных начальных неопределяемых понятий. Такие понятия называются *неопределяемыми* потому, что любая попытка определить их через другие термины приводит к появлению других понятий, которые также нуждаются в определении. Примерами таких понятий в геометрии являются *точка* – то, что, как говорили древнегреческие философы, не имеет частей, и *линия* – длина без ширины. Понятие *множества* является в математике первичным, несводимым к более простым понятиям.

Теория множеств была основана в конце 19 века немецким математиком и философом **Георгом Кантором.** Он давал такое описание понятия множества: «*Множество – это многое, мыслимое нами как единое*». Под множеством в математике понимается совокупность некоторых объектов, обладающих одинаковыми свойствами". Вместо слова «совокупность» можно использовать слова-синонимы: набор, коллекция, семейство, собрание, система, класс, группа, комплект.

Для *обозначения множеств* используются прописные буквы алфавита латинского языка $A, B, C, \ldots, X, Y, Z$, или какого-либо другого по соглашению.

Каждое множество состоит из отдельных объектов, которые называются его ***элементами***. Для *обозначения элементов множества* используются строчные буквы того же алфавита: $a, b, c, \ldots x, y, z$.

Например, если множество представляет собой совокупность факультетов Белорусского государственного университета, то исторический факультет (ИФ) является его элементом.

Тот факт, что «*a является элементом множества A*», записывается в виде $a \in A$. Это высказывание можно также прочесть следующим образом: «*a принадлежит множеству A*» или «*a содержится в множестве A*». Символ $\in$ называется ***символом принадлежности***.

Если «*a не является элементом множества A*», то пишут $a \notin A$ и читают как «*a не принадлежит множеству A*», «*a не содержится в множестве A*».

Множество, состоящее из конечного числа элементов, называется *конечным множеством*.

Пример. Множество всех стран мира – конечное множество.

Множество, не содержащее элементов, называется *пустым множеством* и обозначается символом $\varnothing$.

Замечание. *Символ для пустого множества только один, потому что пустое множество единственно*.

Возможны различные способы задания множества. Один из них состоит в том, что дается полный список элементов, входящих во множество. При этом элементы, принадлежащие множеству, записывают между двумя фигурными скобками и разделяют их запятыми.

Например, множество букв алфавита русского языка: {а, б, в, …, я}, множество арабских цифр десятичной системы счисления: {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.

Замечание. *В дальнейшем, для удобства, будем давать словесное описание множеств в кавычках, **например**, множество A – «множество всех стран мира»*.

В общем случае множество задается с помощью указания *характеристических свойств* его элементов. Так запись

{*x*: *x обладает свойством P*}

задает множество, содержащее только те объекты, которые имеют свойство *P*. Двоеточие в этой записи можно читать как «*такой, что*».

Пример. Множество двузначных натуральных чисел с помощью способа указания характеристического свойства его элементов можно задать так: $A = \{a$: a – двузначное натуральное число$\} = \{10, 11, 12, \ldots, 99\}$.

Замечание. *Множество, состоящее из одного элемента* $\{x\}$, *не следует отождествлять с самим элементом x.*

Пример. Множества $\varnothing$ и $\{\varnothing\}$ являются различными, так как $\varnothing$ – это пустое множество, не содержащее ни одного элемента, а множество $\{\varnothing\}$ – не пусто, его единственным элементом является само пустое множество, т. е. $\varnothing \in \{\varnothing\}$.

Диаграммы Эйлера – Венна

Определение подмножества. *Множество A называется **подмножеством** множества B, обозначается* $A \subset B$, *если каждый элемент множества A является элементом множества B*:

$$\boxed{A \subset B \;\Leftrightarrow\; (\forall\, x)\ (x \in A \Rightarrow x \in B)}.$$

В этой символической записи использованы логические обозначения: символ $\forall$ – *квантор всеобщности* («для всех», «для любого»), знак $\Rightarrow$ – *импликация* (т. е. $X \Rightarrow Y$ означает: «если *X*, то *Y*», «в случае *X* выполняется *Y*»), наконец, логическая связка $\Leftrightarrow$ – *эквивалентность* («если и только если», «тогда и только тогда, когда»), называемая иногда «двойной импликацией».

Пример. Пусть *A* – «множество стран Западной Европы», *B* – «множество европейских стран». Тогда множество *A* является подмножеством множества *B*, т. е. $A \subset B$.

Замечания. 1. *Всякое множество является подмножеством самого себя:* $A \subset A$. **2.** *Пустое множество является подмножеством любого другого множества*: $\varnothing \subset A$.

Действительно, по определению подмножества каждый элемент множества *A* является элементом множества *A*. Если бы пустое множество $\varnothing$ не было подмножеством множества *A*, то оно содержало бы элемент принадлежащий $\varnothing$, но не принадлежащий *A*, а поскольку пустое множество не содержит элементов, то это невозможно.

Множества $\varnothing$ и *A* называются *несобственными подмножествами* множества *A*. Остальные подмножества, если таковые есть, называются *собственными подмножествами* множества *A*.

Пример. Множество гласных букв является собственным подмножеством множества букв русского алфавита.

Для подмножеств верно следующее свойство: подмножество подмножества само является подмножеством, т. е. если $A \subset B$ и $B \subset C$, то $A \subset C$.

Пример. Пусть *A* – «множество студентов первого курса ИФ», *B* – «множество студентов ИФ», *C* – «множество студентов БГУ». Тогда $A \subset B \subset C$.

Если в множестве *A* находится хотя бы один элемент, не содержащийся в множестве *B*, то множество *A не является подмножеством* множества *B, обозначается* $A \not\subset B$.

Пример. Множество стран Содружества независимых государств (СНГ) не является подмножеством множества европейских стран.

Определим, сколько подмножеств имеет конечное множество, состоящее из *n* элементов (*n* – натуральное число).

Пример. Выпишем все подмножества следующих конечных множеств: **а)** $A = \{1\}$, **б)** $B = \{1, 2\}$, **в)** $C = \{1, 2, 3\}$.

Решение. **а)** Подмножествами одноэлементного множества $A = \{1\}$ являются два множества: $\varnothing$, $\{1\}$; **б)** подмножествами двухэлементного множества $B = \{1, 2\}$ являются четыре множества: $\varnothing$, $\{1\}$, $\{2\}$, $\{1, 2\}$; **в)** подмножествами трехэлементного множества $C = \{1, 2, 3\}$ являются восемь множеств: $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$. Заметим, что количество подмножеств каждый раз удваивается.

Конечное множество, состоящее из n элементов (*n – натуральное число*), *имеет* 2^n *подмножеств, включая пустое множество и само множество*.

Обычно все множества, рассматриваемые в рамках какого-либо исследования, являются подмножествами некоторого фиксированного множества, называемого ***универсальным множеством***. Будем обозначать его через ***U***.

Примеры. Пусть универсальное множество ***U*** – «множество всех стран мира». Тогда его подмножествами являются: «множество европейских стран», «множество азиатских стран», «множество стран-членов блока НАТО (организации Североатлантического договора)», «множество стран Европейского союза».

Существует удобный прием наглядного изображения взаимоотношений между множествами – это ***диаграммы Эйлера – Венна***. Множества на этих диаграммах изображаются кругами, а прямоугольник изображает *универсальное множество* ***U***. В диаграммах Эйлера – Венна не имеет значения относительный размер кругов, важно только их взаимное расположение.

Пример. Изобразить на диаграмме Эйлера – Венна взаимное расположение следующих множеств: ***U*** – «множество всех студентов БГУ», *A* – «множество студентов исторического факультета БГУ», *B* – «множество девушек-студенток исторического факультета БГУ».

<u>Решение</u>. Соответствующая диаграмма изображена на рисунке 1, причем видно, что множество *B* включено в множество *A*, т. е. $B \subset A$.

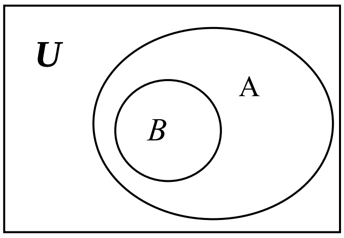


Рис.1

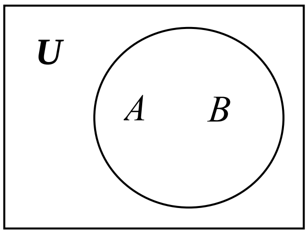


Рис.2

Определение равенства множеств. *Множества А и В* ***равны****, обозначается А = В, если все элементы множества А принадлежат также множеству В, и наоборот, все элементы множества В принадлежат также множеству А:*

$$\boxed{A = B \Leftrightarrow (A \subset B \text{ и } B \subset A)}.$$

Согласно этому определению, *А* = *В*, если каждое из двух множеств есть подмножество другого множества, поэтому можно говорить, что *множества А и В состоят из одних и тех же элементов* (см. рис. 2).

Пример. Пусть *A* – «множество нечетных однозначных натуральных чисел», $B = \{1, 3, 5, 7, 9\}$. Тогда множества *A* и *B* равны, т.е. *A* = *B*.

Определение неравенства множеств. *Множества А и В* ***не равны****, обозначается* $A \neq B$*, если в одном из этих множеств есть хотя бы один элемент, которого нет в другом множестве.*

Пример. Если $A = \{1, 2, 4\}$ и $B = \{1, 2, 3\}$, то множества *A* и *B* не равны, то есть $A \neq B$.

Замечание. *Для неравных множеств не выполняется хотя бы одно из включений* $A \subset B$, $B \subset A$.

Пример. Если $A = \{1, 2\}$ и $B = \{1, 2, 3\}$, то $A \neq B$, так как множество *B* не является подмножеством множества *A*, то есть $B \not\subset A$.

1.1.2. Операции над множествами

Из множеств с помощью определенных операций можно образовывать новые множества. Основными операциями над множествами являются: пересечение множеств, объединение множеств, разность множеств, дополнение множества.

Определение пересечения множеств. Пересечением *двух множеств А и В, обозначается* $A \cap B$*, называется множество, которое состоит из всех* *элементов, принадлежащих каждому из множеств A и B:*

$$\boxed{A \cap B = \{x \colon\ x \in A \ \boldsymbol{и} \ x \in B\}}.$$

Примеры. Если *A* – «множество студентов ИФ», а *B* – «множество девушек-студенток БГУ», тогда $A \cap B$ – «множество девушек-студенток ИФ».

Если, например *A* – «множество нечетных натуральных чисел», а *B* – «множество двузначных натуральных чисел», тогда $A \cap B$ – «множество нечетных двузначных натуральных чисел».

Если множества *A* и *B* не имеют общих элементов, то их пересечение пусто, $A \cap B = \varnothing$, и в таком случае говорят, что множества *A* и *B* ***не пересекаются***. Если *A* – подмножество множества *B*, т. е. $A \subset B$, то $A \cap B = A$, поскольку общими для множеств *A* и *B* будут все элементы множества *A*, и только они.

На рисунках 3 и 4 приведены диаграммы Эйлера – Венна для двух множеств *A* и *B* в случаях, когда, соответственно, $A \cap B \neq \varnothing$ и $A \subset B$. Множеству $A \cap B$ на этих рисунках соответствуют заштрихованные части диаграмм.

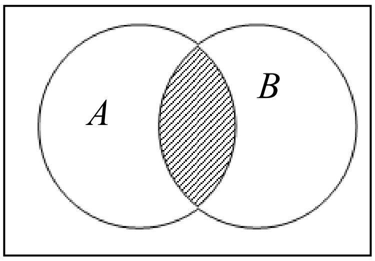


Рис. 3

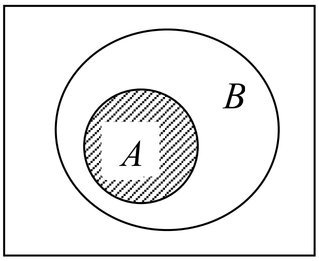


Рис. 4

Замечание. *Отметим свойства операции пересечения, справедливые для любых множеств A и B*: $A \cap B \subset A$, $A \cap B \subset B$.

Для любого множества A имеют место равенства:

$$A \cap \varnothing = \varnothing, \quad A \cap \boldsymbol{U} = A, \quad A \cap A = A.$$

Пример. Пусть *A* – «множество, состоящее из различных букв русского алфавита, входящих в слово МНОЖЕСТВО», *B* – «множество, состоящее из различных букв русского алфавита, входящих в слово ГОСУДАРСТВО». Найдем пересечение этих множеств $A \cap B$.

<u>Решение</u>. Множество *A* состоит из 8 различных букв: *A* = {М, Н, О, Ж, Е, С, Т, В}, а множество *B* состоит из 9 букв: *B* = {Г, О, С, У, Д, А, Р, Т, В}. Пересечением этих множеств является следующий набор из 4 букв: $A \cap B$ = {О, С, Т, В}, который содержится как во множестве *A*, так и во множестве *B*.

Операцию пересечения множеств можно обобщить для пересечения трех и более множеств. **Например**, на рис. 5 заштрихованная часть диаграммы Эйлера – Венна соответствует множеству $A \cap B \cap C$.

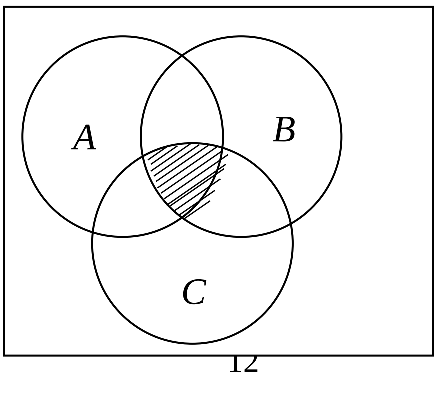


Рис. 5

***Определение объединения множеств. Объединением** двух множеств A и B, обозначается $A \cup B$, называется множество, которое состоит из всех элементов, принадлежащих хотя бы одному из множеств A и B:*

$$\boxed{A \cup B = \{x\colon\ x \in A \ \textbf{\textit{или}}\ x \in B\}}.$$

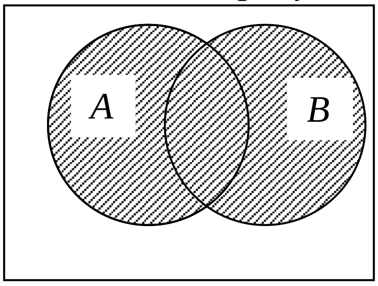
Примеры. Если A – «множество всех девушек-студенток БГУ», а B – «множество всех юношей-студентов БГУ», тогда $A \cup B$ – «множество всех студентов БГУ».

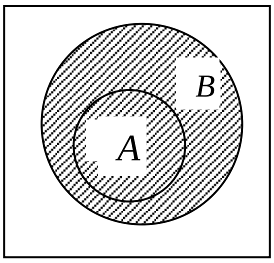
Если A – «множество всех нечетных натуральных чисел», а B – «множество всех четных натуральных чисел», то $A \cup B$ – «множество всех натуральных чисел».

Если A – «множество нечетных чисел», а B – «множество двузначных чисел», тогда $A \cup B$ – «множество, состоящее из нечетных чисел и четных двузначных чисел».

Если A является подмножеством множества B, т. е. $A \subset B$, то $A \cup B = B$, так как элементы из множества A принадлежат множеству B и второй раз включать их в объединение не надо.

На рис. 6 – 8 приведены диаграммы Эйлера – Венна для двух множеств A и B в случаях, когда, соответственно, $A \cap B \neq \varnothing$, $A \subset B$ и $A \cap B = \varnothing$. Множеству $A \cup B$ на этих рисунках соответствуют заштрихованные части диаграмм.





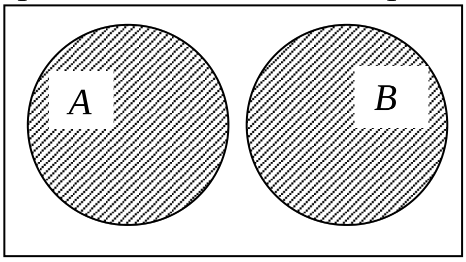


Рис. 6 Рис. 7 Рис. 8

Замечание. *Отметим свойства операции объединения, справедливые для любых множеств A и B*:

$$A \subset A \cup B, \quad B \subset A \cup B.$$

Для любого множества A имеют место равенства:

$$A \cup \varnothing = A, \quad A \cup \boldsymbol{U} = \boldsymbol{U}, \quad A \cup A = A.$$

Пример. Пусть A – «множество, состоящее из различных букв русского алфавита, входящих в слово МНОЖЕСТВО», B – «множество, состоящее из различных букв русского алфавита, входящих в слово ГОСУДАРСТВО». Найдем объединение этих множеств $A \cup B$.

<u>Решение</u>. Множество A состоит из 8 различных букв: A = {М, Н, О, Ж, Е, С, Т, В}, а множество B состоит из 9 букв: B = {Г, О, С, У, Д, А, Р, Т, В}. Объединением этих множеств является следующий набор из 13 букв:

$A \cup B$ = {Г, О, С, У, Д, А, Р, Т, В, М, Н, Ж, Е}.

Операцию объединения множеств можно обобщить в случае объединения трех и более множеств. **Например**, на рис. 9 заштрихованная часть диаграммы Эйлера – Венна соответствует множеству $A \cup B \cup C$.

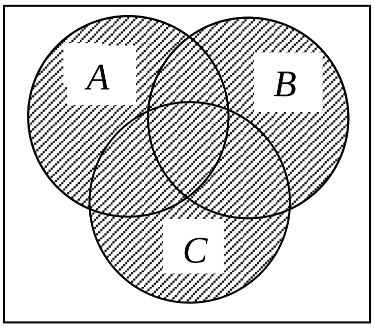


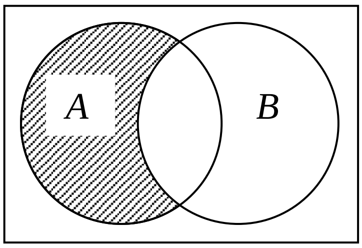
Рис. 9

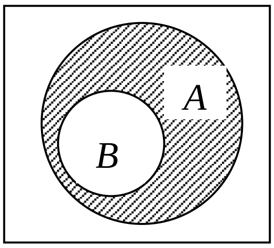
Определение разности множеств. Разностью *двух множеств A и B, обозначается* $A \setminus B$, *называется множество, которое состоит из всех элементов, принадлежащих множеству A, но не принадлежащих множеству B:*

$$\boxed{A \setminus B = \{x\colon\ x \in A \ \textbf{\textit{и}}\ x \notin B\}}\ .$$

Пример. Пусть *A* – «множество студентов группы, изучающих английский язык», *B* – «множество студентов группы, изучающих немецкий язык», а также $A \cap B$ не пусто. Тогда $A \setminus B$ – «множество студентов группы, которые изучают английский язык, но не изучают немецкий язык», $B \setminus A$ – «множество студентов группы, которые изучают немецкий язык, но не изучают английский язык».

На рис. 10 – 12 приведены диаграммы Эйлера – Венна для двух множеств *A* и *B* в случаях, когда, соответственно, $A \cap B \neq \varnothing$, $B \subset A$ и $A \cap B = \varnothing$. Множеству $A \setminus B$ на этих рисунках соответствуют заштрихованные части диаграмм.





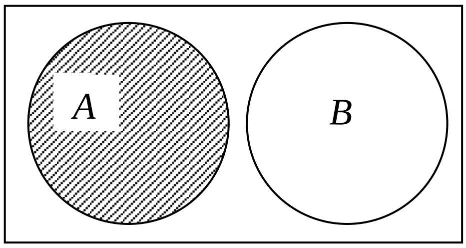


Рис. 10 Рис. 11 Рис. 12

Заметим, что если *A* – подмножество множества *B*, т. е. $A \subset B$, то тогда их разность $A \setminus B = \varnothing$ (см. рис. 7).

Замечание. *Отметим свойства операции разности, справедливые для любых множеств A и B*:

$$A \setminus B \subset A, \quad (A \setminus B) \cap (B \setminus A) = \varnothing, \quad A \setminus B = A \setminus (A \cap B).$$

Для любого множества A имеют место равенства:

$$A \setminus A = \varnothing, \quad A \setminus \varnothing = A, \quad \varnothing \setminus A = \varnothing\ .$$

Пример. Пусть *A* – «множество, состоящее из различных букв русского алфавита, входящих в слово МНОЖЕСТВО», *B* – «множество, состоящее из различных букв русского алфавита, входящих в слово ГОСУДАРСТВО». Найдем разность этих множеств вида $A \setminus B$ и $B \setminus A$.

Решение. Множество *A* состоит из 8 различных букв: *A* = {М, Н, О, Ж, Е, С, Т, В}, а множество B состоит из 9 букв: B = {Г, О, С, У, Д, А, Р, Т, В}. Разностью этих множеств вида $A \setminus B$ является набор букв $A \setminus B$ = {М, Н, Ж, Е}, который принадлежит множеству A, но не содержится в множестве B. Другая разность множеств $B \setminus A$ = {Г, У, Д, А, Р}, т.е. $A \setminus B \neq B \setminus A$.

Определение дополнения множества. *Если* $\boldsymbol{U}$ *– универсальное множество, содержащее множество А, то разность* $\boldsymbol{U} \setminus A$ *называется* ***дополнением*** *множества А и обозначается* $\overline{A}$*:*

$$\boxed{\overline{A} = \boldsymbol{U} \setminus A = \{x\colon\ x \in \boldsymbol{U} \ \ \text{и}\ \ x \notin A\}}.$$

Заметим, что дополнение $\overline{A}$ множества A – это множество элементов фиксированного универсального множества $\boldsymbol{U}$, не входящих в множество A.

Пример. Если $\boldsymbol{U}$ – множество всех действительных чисел, то дополнением множества всех рациональных чисел будет множество всех иррациональных чисел.

Пусть на диаграмме Эйлера – Венна, изображенной на рисунке 13, универсальное множество представлено прямоугольником на плоскости,

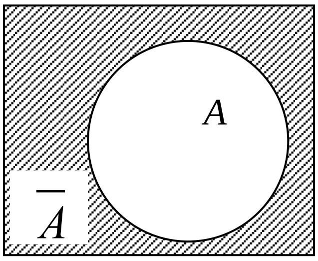


Рис. 13

а множество A – круг, лежащий внутри этого прямоугольника. Тогда дополнению множества A, т. е. множеству $\overline{A}$, соответствует заштрихованная часть этого прямоугольника.

Замечание. *Отметим следующие свойства операции дополнения, справедливые для любого множества А и содержащего его универсального множества* $\boldsymbol{U}$:

$$A \cup \overline{A} = \boldsymbol{U}, \quad A \cap \overline{A} = \varnothing;$$
$$\overline{\overline{A}} = A, \quad \overline{\varnothing} = \boldsymbol{U}, \quad \overline{\boldsymbol{U}} = \varnothing.$$

Замечание. *Если А – подмножество множества В, т. е.* $A \subset B$*, то* $\overline{B} \subset \overline{A}$ *или* $\boldsymbol{U} \setminus B \subset \boldsymbol{U} \setminus A$.

Замечание. *Разность множеств А и В можно выразить через пересечение множеств А и В, а именно,*

$$A \setminus B = A \cap \overline{B},$$

т. е. разность множеств $A \setminus B$ *есть пересечение множества А и дополнения множества В.*

В заключение рассмотрим на следующем примере все четыре операции над множествами: пересечение, объединение, разность, дополнение.

Пример. Пусть $\boldsymbol{U}$ – «множество всех стран мира», A – «множество стран, с которыми Беларусь имеет дипломатические отношения», а B – «множество стран, с которыми Россия имеет дипломатические отношения». Опишем словами множества: $A \cap B$, $A \cup B$, $A \setminus B$, $B \setminus A$, $\overline{A}$.

Решение. Получим: $A \cap B$ – «множество стран, с которыми одновременно и Беларусь и Россия имеют дипломатические отношения», $A \cup B$ – «множество стран, с которыми имеют дипломатические отношения Беларусь или Россия», $A \setminus B$ – «множество стран, с которыми только Беларусь имеет дипломатические отношения», $B \setminus A$ – «множество стран, с которыми только Россия имеет дипломатические отношения», $\overline{A}$ – «множество стран, с которыми Беларусь не имеет дипломатических отношений».

Сформулируем основные свойства операций над множествами.

1. *Законы коммутативности*. Для любых двух множеств A и B выполняются **свойства коммутативности** операций пересечения и объединения:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

2. *Законы ассоциативности*. Для любых трех множеств A, B и C выполняются **свойства ассоциативности** операций пересечения и объединения:

$$(A \cap B) \cap C = A \cap (B \cap C), \quad (A \cup B) \cup C = A \cup (B \cup C).$$

3. *Законы дистрибутивности*. При чередовании операций объединения и пересечения для любых трех множеств A, B и C выполняются **свойства дистрибутивности** одной операции относительно другой:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

4. *Закон двойного дополнения*. Для любого множества A верно равенство:

$$\overline{\overline{A}} = A.$$

5. *Законы де Моргана* (*законы двойственности*). Для любых двух множеств A и B верны следующие равенства:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}, \quad \overline{A \cup B} = \overline{A} \cap \overline{B}.$$

Замечание. *Операция разности множеств свойствами коммутативности и ассоциативности не обладает, т.е. верны неравенства*:

$$A \setminus B \neq B \setminus A, \ (A \setminus B) \setminus C \neq A \setminus (B \setminus C).$$

На рисунке 14 приведена диаграмма Эйлера – Венна для двух множеств A и B, на которой множествам $A \setminus B$ и $B \setminus A$ соответствуют различные заштрихованные части диаграммы.

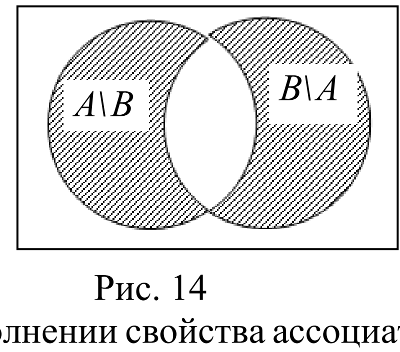
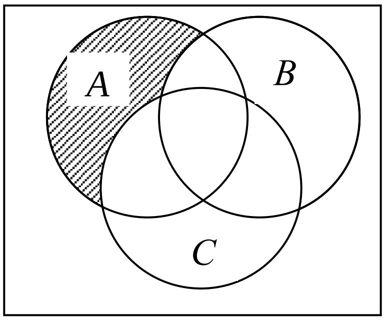


Рис. 14

Утверждение о невыполнении свойства ассоциативности операции разности множеств можно проверить на конкретных примерах. На рисунках 15 и 16 приведены диаграммы Эйлера – Венна для трех множеств *A*, *B*, и *C*, на которых множествам $(A \setminus B) \setminus C$ и $A \setminus (B \setminus C)$ соответствуют различные заштрихованные части диаграмм.



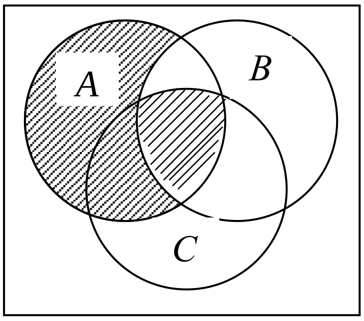


Рис. 15 Рис.16

Пример. Множество всех студентов факультета международных отношений БГУ является объединением трех множеств *A*, *B*, и *C*, где *A* – «множество всех успевающих студентов», *B* – «множество всех девушек-студенток», *C* – «множество всех неуспевающих юношей-студентов». Опишем множества, входящие в равенства законов дистрибутивности:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Решение. Ясно, что каждый студент ФМО БГУ принадлежит хотя бы одному из указанных множеств. Заметим, что множества *A* и *B* имеют общие элементы – успевающие девушки входят и в первое, и во второе множество.

Поскольку $B \cup C$ – «множество всех девушек-студенток и всех неуспевающих юношей-студентов», то $A \cap (B \cup C)$ – «множество всех успевающих девушек-студенток», с другой стороны, так как $A \cap C = \varnothing$, то $(A \cap B) \cup (A \cap C) = A \cap B$ – «множество всех успевающих девушек-студенток». Значит, для множеств *A*, *B* и *C* справедливы равенства:

$$A \cap (B \cup C) = A \cap B = (A \cap B) \cup (A \cap C).$$

Поскольку $B \cap C = \varnothing$, то $A \cup (B \cap C) = A \cup \varnothing = A$. С другой стороны, $A \cup B$ – «множество всех успевающих студентов и всех девушек-студенток», $A \cup C$ – «множество всех успевающих студентов и всех неуспевающих юношей-студентов», следовательно, $(A \cup B) \cap (A \cup C)$ – «множество всех успевающих студентов», т. е. это множество *A*. Значит, для множеств *A*, *B* и *C* справедливы равенства:

$$A \cup (B \cap C) = A = (A \cup B) \cap (A \cup C).$$

1.1.3. Число элементов в объединении конечного множества

Обозначим через $n(A)$ – *число элементов конечного множества A*.

Посчитаем число элементов объединения двух множеств *A* и *B* в случае, когда их пересечение не пусто, т. е. если $A \cap B \neq \varnothing$.

Утверждение. *Для произвольных конечных множеств A и B число* *элементов объединения этих множеств* $n(A \cup B)$ *равно:*

$$n(A \cup B) = n(A) + n(B) - n(A \cap B).$$

Эта формула следует из того факта, что когда число элементов множества A суммируется с числом элементов множества B, то элементы, принадлежащие множеству $A \cap B$, учитываются дважды (см. рис. 3).

Пример. В группе из 40 студентов 25 человек изучают английский язык, 10 человек – немецкий язык, а 5 человек – одновременно английский и немецкий языки. Сколько студентов изучают английский или немецкий язык? Сколько студентов не изучают ни английский, ни немецкий язык?

<u>Решение</u>. Пусть A – «множество студентов, изучающих английский язык», B – «множество студентов, изучающих немецкий язык», а универсальное множество $\boldsymbol{U}$ – «множество всех студентов группы». Тогда, в силу предыдущего утверждения, количество студентов, изучающих английский или немецкий языки, равно

$n(A \cup B) = n(A) + n(B) - n(A \cap B) = 25 + 10 - 5 = 30.$

Так как 40 – 30 = 10, то всего 10 студентов группы не изучают эти иностранные языки.

Посчитаем число элементов объединения трех множеств A, B и C в случае, когда их пересечение не пусто, т. е. если $A \cap B \cap C \neq \varnothing$ (см. рис. 5). Если просуммировать количество элементов в каждом множестве A, B и C, то некоторые подмножества при подсчете будут учтены дважды. Если вычесть число элементов множества $A \cap B$, множества $A \cap C$ и множества $B \cap C$, т. е. $n(A \cap B)$, $n(A \cap C)$ и $n(B \cap C)$, соответственно, из числа элементов $n(A \cup B \cup C)$, то, как видно из диаграммы Эйлера – Венна (см. рис. 5), элементы множества $A \cap B \cap C$ совсем не будут учтены. Поэтому к указанной разности нужно добавить $n(A \cap B \cap C)$, тогда каждый элемент множества $A \cup B \cup C$ будет учтен ровно один раз. Таким образом, верно следующее утверждение.

Утверждение. *Для произвольных конечных множеств* A, B *и* C *число элементов объединения этих множеств* $n(A \cup B \cup C)$ *равно:*

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C). \quad (*)$$

Пример (*планета мифов*)[1]**.** Однажды во время беседы за чашкой кофе в клубе Межгалактических путешественников знаменитый член этого клуба, Мюнхгаузен космической эры, Йон Тихий рассказал: «*Высадка на планету Гесиод была очень трудна. Но когда я оказался на поверхности, то пожалел, что решил опуститься: на ней жили чудовища, более страшные, чем описанные в древних мифах греков. Навстречу мне вышла делегация из* 1000 *жителей планеты. У* 811 *из них был один глаз, как у циклопа Полифема, у* 752 *– вместо волос были змеи, как у Медузы Горгоны, а* 418 *имели рыбий хвост, как нереиды. При этом* 570 *чудовищ были одновременно одноглазы и змееволосы,* 356 *– одноглазы и при этом имели рыбий хвост,* 348 *– змееволосы и кроме того с*

---

[1] Виленкин, Н.Я. Рассказы о множествах / Н.Я. Виленкин. – 3-е изд. – М.: МЦНМО, 2004. – С. 41.

рыбьим хвостом, а 297 – *были одновременно одноглазы, змееволосы и с рыбьим хвостом. Старший из них обратился ко мне и сказал...*»

Но члены клуба так и не узнали, что услышал Йон Тихий на планете чудовищ. Слушавший рассказ путешественника профессор Тарантога мгновенно произвел в уме какие-то выкладки и воскликнул: «*Дорогой Йон! Я готов поверить, что на этой планете жили существа с одним глазом, со змеями вместо волос и с рыбьими хвостами. Тебе приходилось встречать еще более страшных чудовищ – вспомни о курдлях. Но я надеюсь, что законы математики на этой планете не превратились в мифы*».

<u>Вопрос</u>. *Почему профессор Тарантога сделал такое заключение*?

<u>Решение</u>. Обозначим через $\boldsymbol{U}$ – «множество всех членов делегации», через A – «множество одноглазых делегатов», через B – «множество змееволосых делегатов», а через C – «множество делегатов, имеющих рыбьи хвосты». По условию число элементов этих конечных множеств равно:

$n(\boldsymbol{U}) = 1000$, $n(A) = 811$, $n(B) = 752$, $n(C) = 418$.

В силу принятых обозначений $A \cap B$ – «множество одноглазых со змеями вместо волос делегатов», $A \cap C$ – «множество одноглазых, имеющих рыбий хвост делегатов», $B \cap C$ – «множество змееволосых с рыбьим хвостом делегатов», и, наконец, $A \cap B \cap C$ – «множество делегатов, которые одновременно одноглазы, змееволосы и с рыбьими хвостами». По условию число элементов этих множеств равно

$n(A \cap B) = 570$, $n(A \cap C) = 356$, $n(B \cap C) = 348$, $n(A \cap B \cap C) = 297$.

Согласно предыдущему утверждению для $A \cup B \cup C$ – «множества делегатов, которые одноглазы или змееволосы или с рыбьими хвостами» справедливо равенство (∗). Поэтому число элементов множества $A \cup B \cup C$ равно

$n(A \cup B \cup C) = 811 + 752 + 418 - 570 - 356 - 348 + 297 = 1004$.

Подсчитаем теперь, сколько членов делегации жителей планеты Гесиод не были ни одноглазыми, ни змееволосыми и не имели рыбьих хвостов, то есть сколько элементов содержит множество $\boldsymbol{U} \setminus (A \cup B \cup C)$, где $A \cup B \cup C$ – подмножество множества $\boldsymbol{U}$. Так как $n(\boldsymbol{U}) = 1000$, а $n(A \cup B \cup C) = 1004$, то разность множеств $\boldsymbol{U} \setminus (A \cup B \cup C)$ содержит $n\big(\boldsymbol{U} \setminus (A \cup B \cup C)\big) = 1000 - 1004 = -4$ планетных существа.

<u>Заключение</u>. Поэтому профессор Тарантога имел все основания сказать: «*Но согласись сам, дорогой Йон, даже на планете мифов ни одно множество не может иметь отрицательной численности. Даже для тебя такие выдумки слишком невероятны*».

1.2. ЭЛЕМЕНТЫ КОМБИНАТОРИКИ И ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.2.1. Элементы комбинаторики

Комбинаторикой называется раздел математики, посвященный решению задач расположения элементов в соответствии с заданными правилами и выяснению, сколькими способами такие расположения могут быть осуществлены. Методы комбинаторики помогают в теории вероятностей осуществить подсчет числа возможных исходов и числа благоприятствующих исходов в конкретных случаях.

При решении многих комбинаторных задач оказываются полезными следующие утверждения, которые часто называют основными правилами комбинаторики.

Правило суммы. Если множество A содержит m различных элементов, а множество B – n различных элементов и $A \cap B = \varnothing$, то множество $A \cup B$ содержит $m+n$ элементов.

Замечание. Если некоторый объект A можно выбрать m способами, а другой объект B, отличный от A, – n способами, то согласно правилу суммы выбрать либо A, либо B можно $m+n$ способами.

Правило произведения. Если множество A содержит m различных элементов, т. е. $A = \{a_i : i = 1, 2, \dots, m\}$, а множество B – n различных элементов, т. е. $B = \{b_j : j = 1, 2, \dots, n\}$, то множество C, составленное из всех возможных пар элементов множеств A и B, т. е. $C = \{(a_i, b_j) : i = 1, 2, \dots, m; \ j = 1, 2, \dots, n\}$, содержит $m \cdot n$ элементов.

Поясним указанные правила на примерах.

Пример. Сколькими способами студент исторического факультета БГУ может выбрать одну книгу, если на полке находятся 15 книг по всемирной истории и 10 книг по истории Беларуси?

<u>Решение.</u> Заметим, что книгу по всемирной истории можно выбрать 15 способами, а книгу по истории Беларуси – 10 способами. Поэтому согласно правилу суммы студент может выбрать одну книгу на полке $15 + 10 = 25$ способами.

Пример. В группе из 20 человек должен быть выбран староста и его заместитель. Сколькими способами это можно сделать, если каждый студент может быть избран на одну из этих должностей?

<u>Решение.</u> Всего возможно 20 вариантов выбора старосты группы, а после того как староста будет выбран, существует 19 способов выбора его заместителя из оставшихся студентов. Поэтому согласно правилу произведения есть $20 \cdot 19 = 380$ способов выбора старосты и его заместителя.

Замечание. Правила суммы и произведения могут быть распространены на любое конечное число k подмножеств множества U, обладающих требуемыми свойствами.

Размещения. Перестановки. Сочетания

Пусть есть множество X, содержащее n элементов. Множество X называется *упорядоченным*, если задан порядок расположения его элементов, т. е. указано, какой элемент является первым, какой – вторым, какой – третьим и т. д. Например, множества $\{a, b, c\}$ и $\{b, a, c\}$ как упорядоченные множества являются различными.

Определение. **Упорядоченное подмножество множества X, содержащее m различных элементов $(m \le n)$, называется размещением из n элементов по m.**

В соответствии с данным определением два размещения из n элементов по m элементов считаются различными, если они отличаются друг от друга либо хотя бы одним элементом, либо порядком следования элементов. Число размещений из n элементов по m обозначается символом A_n^m (A – первая буква французского слова *arrangement* – размещение). Читается: «а из эн по эм».

Теорема 1. Число размещений из n элементов по m определяется формулой

$$A_n^m = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+1), \tag{1}$$

т. е. равно произведению m последовательных натуральных чисел от $n-m+1$ до n включительно.

Доказательство. Справедливость сформулированного утверждения вытекает из правила умножения. Действительно, первый из элементов размещения может быть выбран n способами; второй элемент может быть выбран $n-1$ способами, поскольку использование выбранного элемента недопустимо; третий элемент может быть выбран $n-2$ способами и т. д. Последний, m-й элемент размещения может быть выбран $n-m+1$ способами. Применяя правило умножения, получаем формулу (1).

Понятие размещения применяется при решении комбинаторных задач следующего типа: *сколькими способами можно разместить по m различным местам m из n разных объектов*?

Пример. Сколькими способами можно выбрать три человека на три различные должности из 10 кандидатов?

Решение. Искомое число способов равно числу размещений из 10 элементов по 3, т. е.

$$A_{10}^3 = 10 \cdot 9 \cdot 8 = 720.$$

Пример. В седьмом классе изучается 14 предметов. Сколькими способами можно составить расписание на субботу, если в этот день недели должно быть 5 различных уроков?

Решение. Искомое число способов равно числу размещений из 14 элементов по 5, т. е.

$$A_{14}^5 = 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 = 240\,240.$$

Среди всех размещений особая роль отводится случаю размещений из n элементов по n.

Определение. **Размещение из n элементов по n называется перестановкой из n элементов.**

В соответствии с этим определением две перестановки из n элементов являются различными, если они отличаются друг от друга порядком следования хотя бы двух элементов. Число перестановок из n элементов обозначается символом P_n (P – первая буква французского слова *permutation* – перестановка). Читается: «пэ из эн».

Теорема 2. Число перестановок из n элементов определяется формулой

$$P_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1. \tag{2}$$

Доказательство. Формула (2) непосредственно следует из формулы (1) при $m = n$.

Замечание. Для произведения вида (2) введен специальный символ $n!$ (читается: «эн факториал»; например, $10!$ – десять факториал). Таким образом, $P_n = n!,$ а формула для числа размещений из n элементов по m может быть записана в виде $A_n^m = \dfrac{n!}{(n-m)!}.$

Отметим также, что если в равенство $(n+1)! = n!\cdot(n+1)$ подставить $n = 0,$ то получится $1! = 0!\cdot 1.$ Поэтому принято считать, что $0! = 1.$

Пример. Число перестановок из трех элементов $a,$ $b,$ c равно $P_3 = 3! = 3\cdot 2\cdot 1 = 6.$ Ими являются следующие множества:

$$\{a, b, c\},\quad \{a, c, b\},\quad \{b, a, c\},\quad \{b, c, a\},\quad \{c, a, b\},\quad \{c, b, a\}.$$

Понятие перестановки применяется при решении комбинаторных задач следующего типа: *скольким способами можно разместить n разных объектов по n разным местам*?

Пример. Сколькими способами могут разместиться на скамейке пять человек?

Решение. Число способов размещения на скамейке пяти человек равно числу перестановок из пяти элементов, т. е. $P_5 = 5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120.$

В некоторых задачах по комбинаторике порядок расположения объектов в той или иной совокупности не имеет значения. Важно лишь то, какие именно элементы ее составляют. В связи с этим вводят понятие сочетания.

Определение. **Неупорядоченное подмножество множества X, содержащее m различных элементов $(m \le n)$, называется сочетанием из n элементов по m.**

В соответствии с этим определением два сочетания из n элементов по m элементов считаются различными, если они отличаются друг от друга хотя бы одним элементом (порядок следования элементов при этом не учитывается). Число сочетаний из n элементов по m обозначается символом C_n^m (C – первая буква французского слова *combinaison* – сочетание). Читается: «цэ из эн по эм».

Теорема 3. Число сочетаний из n элементов по m определяется формулой

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m!\cdot(n-m)!}. \qquad (3)$$

Доказательство. Все возможные сочетания из n элементов по m могут быть получены следующим образом. Рассмотрим сначала все размещения из n элементов по m. Их число будет равно A_n^m. Но, как было указано ранее, размещения представляют собой упорядоченные подмножества множества X, т. е. подмножества, которые состоят из одних и тех же элементов, но следующих в разном порядке, считаются различными. Однако такие подмножества являются совпадающими, если рассматривать их как сочетания. Это означает, что все размещения, отличающиеся друг от друга порядком следования элементов, должны отождествляться. Число таких отождествляемых размещений при фиксированном наборе m элементов равно $P_m = m!$. Отсюда и следует справедливость формулы (3).

Отметим, что $C_n^0 = 1$, $C_n^1 = n$, $C_n^n = 1$. Обратим также внимание на то, что если в формуле (3) заменить m на $n - m$, то получится то же самое выражение, только факториалы в знаменателе правой части поменяются местами:

$$C_n^m = C_n^{n-m}.$$

Понятие сочетания применяется при решении комбинаторных задач следующего типа: *сколькими способами можно выбрать m из n разных объектов*?

Пример. Сколькими способами можно выбрать три человека на три одинаковые должности из 10 кандидатов?

<u>Решение.</u> Число способов выбора трех человек на три одинаковые должности из 10 кандидатов равно числу сочетаний из 10 по 3, т. е.

$$C_{10}^3 = \frac{10!}{3! \cdot 7!} = \frac{10 \cdot 9 \cdot 8}{3!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120.$$

Пример. Сколько экзаменационных комиссий, состоящих из пяти человек, можно образовать из 12 преподавателей?

<u>Решение.</u> Нужное число экзаменационных комиссий равно

$$C_{12}^5 = \frac{12!}{5! \cdot 7!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 11 \cdot 9 \cdot 8 = 792.$$

Рассмотрим еще несколько задач, которые решаются с применением указанных выше формул.

1.2.2. Случайные события и вероятности

Теория вероятностей зародилась в XVI в. благодаря попыткам создать теорию распространенных в ту эпоху азартных игр. Н. Тарталья, Дж. Кардано и другие производили теоретический подсчет вероятности, с которой может появиться та или иная комбинация очков при игре в кости. На примерах отдельных задач, разбиравшихся этими учеными, складывались первые понятия и простейшие теоремы теории вероятностей.

Возникновение теории вероятностей как науки относится к середине XVII в. и связано с именами Б. Паскаля, П. Ферма, Х. Гюйгенса. *Теория вероятностей* – это наука, изучающая закономерности случайных явлений. Если в начале своего существования теория вероятностей состояла из объяснений различных казусов, к которым приводили игры и лотереи, то с течением времени области, охватываемые теорией вероятностей, становились все шире и серьезнее. Довольно быстро теория вероятностей стала применяться в страховом деле, а в конце XVIII в. П. Лаплас развил на основе теории вероятностей теорию ошибок наблюдений, имевшую важное значение для астрономии. Теория вероятностей используется в теории стрельбы, телефонии, страховом деле, машиностроении в виде теории допусков и т. д.

Основой всех точных исследований является наблюдение за поведением и признаками изучаемых объектов и явлений, которое может осуществляться с помощью соответствующего опыта или эксперимента. Под *опытом* будем подразумевать однократное или многократное повторение некоторого действия, а под *испытанием* – отдельное такое действие. Результат испытания будем называть *исходом* испытания. Каждое *испытание* должно удовлетворять следующим условиям:

а) исход (результат) испытания точно неизвестен, т. е. испытание может иметь более одного исхода;

б) конечное множество всех исходов может быть определено до начала испытания;

в) испытание можно повторить неограниченное число раз при тех же условиях.

Игровые модели очень удобны для первоначального рассмотрения элементов теории вероятностей. Напомним, что *игральная кость* – это кубик с округленными углами, на гранях которого нанесены точки, изображающие числа от 1 до 6.

Пусть опыт состоит в многократном бросании игральной кости на стол. В этом случае исходом каждого испытания является выпадение одной из шести граней кубика, соответствующей числам: 1, 2, 3, 4, 5, 6. При подбрасывании монеты возможно два исхода: выпадение «орла» («герба») и выпадение «решки» («числа»).

Множество всех исходов испытания называется *множеством (пространством) элементарных событий,* и обозначается Ω*, а случайным событием* – подмножество множества элементарных событий.

В дальнейшем случайные события будем называть просто событиями. События обозначаются прописными буквами латинского алфавита *A*, *B*, *C*, ...

Рассмотрим следующий пример испытания. Если из группы студентов, в которой обучаются 25 человек, выбираются наудачу пять студентов для участия в университетской конференции, то тогда множество элементарных событий Ω состоит из C_{25}^{5} элементов. Его подмножествами являются, например, такие события: A – «выбраны пять девушек-студенток», B – «выбраны пять иностранных студентов», C – «выбраны пять студентов-отличников».

Замечание. В теории множеств аналогом пространства элементарных событий Ω является универсальное множество *U*, а аналогом событий – множества, которые являются подмножествами *U*.

Каждое случайное событие, в частности выпадение «орла», есть следствие действия очень многих случайных причин (сила, с которой брошена монета, форма монеты и др.). Невозможно учесть влияние всех этих причин на результат испытания. Поэтому теория вероятностей не ставит перед собой задачу предсказать, произойдет единичное событие или нет. В теории вероятностей рассматриваются только *массовые* случайные события, т. е. предполагается, что возможно создать много раз одни и те же условия, при каждом из которых данное случайное событие может произойти либо не произойти.

Оказывается, что массовые случайные события подчиняются некоторым закономерностям, которые называют вероятностными закономерностями. Так, нельзя наперед определить результат одного бросания монеты, но можно предсказать, причем с небольшой погрешностью, промежуток, в который попадет число появлений «орла», если монета будет брошена достаточно большое количество раз.

Исходя из реального смысла понятия события, можно определить следующие частные случаи события и операции над событиями.

Событие называется *достоверным*, если в данном опыте оно заведомо происходит. Оно совпадает со всем пространством элементарных событий, поэтому обозначается символом Ω. Событие называется *невозможным*, если в данном опыте оно заведомо не может произойти ни при одном испытании (обозначается символом $\varnothing$).

Например, если в урне находятся только белые шары, а испытание состоит в том, что из урны вынимают один шар, то событие «достали белый шар» является достоверным, а событие «достали черный шар» является невозможным.

События, которые не могут произойти одновременно в рассматриваемом опыте, называются *несовместными.*

Замечание. В терминах теории множеств события A и B являются несовместными в данном опыте, если $A \cap B = \varnothing$.

Например, при бросании игральной кости события A – «выпадение грани с числом 1» и B – «выпадение грани с числом 2» являются несовместными событиями.

Два события называются *противоположными*, если в данном опыте появление одного из них равносильно непоявлению другого.

Если некоторое событие обозначено буквой A, то противоположное ему событие обозначают $\overline{A}$.

Замечание. В терминах теории множеств для события A и противоположного ему события $\overline{A}$ верно равенство: $A \cup \overline{A} = \Omega$.

Например, если событие A – «попадание», то противоположное ему событие $\overline{A}$ – «промах» при одном выстреле по мишени.

Суммой двух событий A и B называется событие, состоящее в наступлении хотя бы одного из событий A, B и обозначаемое через $A + B$ или $A \cup B$.

Произведением двух событий A и B называется событие, обозначаемое через AB или $A \cap B$, которое происходит тогда и только тогда, когда происходят и A, и B вместе.

Разностью событий A и B называется событие $A \setminus B$, которое происходит тогда и только тогда, когда происходит A, но не происходит B.

Пример. Пусть бросается игральная кость. В процессе данного испытания может выпасть число очков, равное какому-либо числу из множества $\{1, 2, 3, 4, 5, 6\}$. Событиями в этом случае являются, например, события

$A = \{\text{выпало четное число очков}\}$, т. е. $A = \{2, 4, 6\}$,

$B = \{\text{выпавшее число очков не меньше трех}\}$, т. е. $B = \{3, 4, 5, 6\}$.

Тогда

$A + B = \{\text{выпадает число очков, отличное от одного}\} = \{2, 3, 4, 5, 6\}$,

$AB = \{$выпадает число очков, равное 4 или 6$\} = \{4, 6\}$,

$A \setminus B = \{$выпадает число очков, равное 2$\} = \{2\}$,

$\overline{A} = \{$выпадает нечетное число очков$\} = \{1, 3, 5\}$,

$\overline{B} = \{$выпадает число очков, меньшее трех$\} = \{1, 2\}$.

Пример. В опыте с бросанием игральной кости пространство элементарных событий имеет вид $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, где ω_k – элементарное событие, состоящее в том, что на верхней грани игральной кости выпала цифра k ($k = 1, 2, 3, 4, 5, 6$).

Пусть событие A состоит в появлении нечетного числа очков на верхней грани. Этому событию благоприятствуют элементарные события из подмножества $\{\omega_1, \omega_3, \omega_5\}$ множества Ω.

Классическое определение вероятности

Для количественной оценки возможности появления случайного события A вводится понятие вероятности.

Каждому событию A ставится в соответствие характеризующее возможность его появления число $P(A)$ (от первой буквы французского слова *probabilite* – вероятность), которое называется *вероятностью* события A.

Для определения классической вероятности нам потребуется понятие *равновозможных* исходов испытания (или *равновозможных* элементарных событий). Элементарные события (исходы испытания), шансы появления которых одинаковы в данном испытании, будем называть *равновозможными*.

Обычно понятие классической вероятности иллюстрируется на игровых моделях, поскольку здесь равновозможность исходов испытания прямо задана внешней или геометрической симметрией объекта – монеты, игральной кости, колоды карт и т. д.

Если монета ровная, неизогнутая, то можно ожидать, что при ее многократном подбрасывании «орел» и «решка» будут выпадать одинаково часто, т. е. примерно в половине случаев будет выпадать «орел», а в половине случаев – «решка». Поэтому в условиях этого эксперимента принято считать, что для такой монеты вероятность выпадения «орла» или «решки» равна $\frac{1}{2}$.

При многократном бросании идеально правильной игральной кости естественно ожидать, что на долю каждой грани с числами 1, 2, 3, 4, 5, 6 будет приходиться примерно шестая часть общего числа исходов испытаний, т. е. бросаний кости. Поэтому считают, что вероятность выпадения каждой грани, соответствующей числам 1, 2, 3, 4, 5, 6, равна $\frac{1}{6}$.

Какова в таком случае вероятность события A – «выпадение нечетного числа очков»? Поскольку существует 6 одинаково возможных исходов каждого испытания, причем три из них (выпадение граней с числами 1,3,5) благоприятствуют событию A, то вероятность $P(A)$ можно считать равной $\frac{3}{6} = \frac{1}{2}$.

Пусть пространство Ω состоит из конечного числа равновозможных элементарных событий: $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, $A \subseteq \Omega$.

Определение. **Под вероятностью $P(A)$ события A понимается отношение числа равновозможных элементарных событий, благоприятствующих событию A, к общему числу всех равновозможных элементарных событий.**

Если m – число элементарных событий, благоприятствующих событию A, n – общее число всех равновозможных элементарных событий опыта, то по определению

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

Это определение вероятности называется *классическим*.

При использовании формулы классической вероятности в решении конкретных задач числовые значения величин m и n, входящих в эту формулу, не всегда очевидны. Их определение требует применения основных правил и формул комбинаторики.

Из определения вероятности события следуют ее простейшие свойства.

1. $P(\Omega) = 1$, так как все элементарные события являются благоприятствующими Ω, т. е. $m = n$.

2. $P(\varnothing) = 0$, так как для невозможного события нет ни одного элементарного события, ему благоприятствующего, т. е. $m = 0$.

3. Для всякого события $A \subseteq \Omega$ справедливы соотношения

$$0 \le P(A) \le 1,$$

так как $0 \le m \le n$.

Пример. Найти вероятность того, что при бросании игральной кости выпадет четное число очков.

Решение. Для данного примера имеем: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, $A = \{\omega_2, \omega_4, \omega_6\}$, где ω_k – элементарное событие, состоящее в том, что на верхней грани игральной кости выпала цифра k ($k = 1, 2, 3, 4, 5, 6$). Поэтому $m = 3$, $n = 6$ и $P(A) = \frac{3}{6} = \frac{1}{2}$.

Пример. В ящике три белых, четыре черных и пять красных шаров. Какова вероятность вынуть из ящика черный шар?

Решение. В этом примере число всех равновозможных элементарных событий опыта $n = 3 + 4 + 5 = 12$, т. е. это количество шаров в ящике. Число элементарных событий, благоприятствующих событию

$$A = \{\text{из ящика вынут черный шар}\},$$

равно количеству черных шаров, т. е. $m = 4$. Отсюда $P(A) = \frac{4}{12} = \frac{1}{3}$.

В теории вероятностей классическим является эксперимент с урной, из которой надо, не глядя, извлекать одинаковые шары разных окрасок.

Рассмотрим задачу о выборке, которую называют также «*урновой схемой*», имеющую практическое применение в различных областях знания.

Пример. В урне всего K шаров, из них L белых и $K-L$ черных. Из урны наудачу вынимаются k шаров. Какова вероятность того, что среди вынутых k шаров окажется l белых?

Решение. Исход этого испытания состоит в появлении любых k шаров из общего числа K шаров, находящихся в урне. В данном примере событие A – «среди вынутых k шаров окажется l белых». Поскольку нас не интересует порядок появления шаров, то число всех исходов данного испытания равно числу сочетаний C_K^k. Заметим, что мы извлекаем не только l белых шаров, но и оставшиеся $k-l$ черных. Очевидно, что $0 \le l \le L$ и $0 \le k-l \le K-L,$ в противном случае вероятность появления интересующего нас события равна 0. Извлечь l белых шаров, из имеющихся в урне L белых шаров, можно C_L^l способами. Соответственно извлечь $k-l$ черных шаров из $K-L$ черных шаров в урне можно C_{K-L}^{k-l} способами. Согласно правилу произведения число благоприятствующих исходов для события A равно $C_L^l \cdot C_{K-L}^{k-l}$. Следовательно, искомая вероятность по формуле классической вероятности равна

$$P(A) = \frac{C_L^l \cdot C_{K-L}^{k-l}}{C_K^k}.$$

1.2.3. Основные теоремы теории вероятностей

Теорема сложения вероятностей

Теорема .4. Вероятность суммы двух событий равна сумме вероятностей этих событий без вероятности их одновременного осуществления, т. е.

$$P(A+B) = P(A) + P(B) - P(AB).$$

Доказательство. Используем классическое определение вероятности. Пусть $|\Omega| = n$ – общее число элементарных событий; m_1 – число элементарных событий, благоприятствующих событию A; m_2 – число элементарных событий, благоприятствующих событию B; l – число элементарных событий, благоприятствующих событию AB. Тогда

$$P(A) = \frac{m_1}{n}, \quad P(B) = \frac{m_2}{n}, \quad P(AB) = \frac{l}{n}.$$

Событию $A+B$ будут благоприятствовать $m_1 + m_2 - l$ элементарных событий. Действительно, складывая числа m_1 и m_2 элементарных событий, благоприятствующих событиям A и B соответственно, мы дважды считаем элементарные события, благоприятствующие событию AB.

Поэтому при подсчете числа элементарных событий, благоприятствующих событию $A+B,$ значение l необходимо исключить. Таким образом,

$$P(A+B) = \frac{m_1 + m_2 - l}{n} = \frac{m_1}{n} + \frac{m_2}{n} - \frac{l}{n} = P(A) + P(B) - P(AB).$$

Теорема доказана.

Следствие 1. Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий:

$$P(A+B)=P(A)+P(B). \qquad (4)$$

Действительно, так как $AB=\varnothing,$ то $P(AB)=0.$

Формула (4) допускает распространение на случай произвольного конечного числа слагаемых.

Следствие 2. Если $A_1, A_2, \dots, A_n$ – попарно несовместные события (т. е. никакие два из них не могут произойти одновременно), то

$$P(A_1+A_2+\dots+A_n)=P(A_1)+P(A_2)+\dots+P(A_n). \qquad (5)$$

Действительно, при $n=2$ равенство (5) превращается в равенство (4). Предположим, что формула (5) верна при $n=k-1.$ Тогда при $n=k$ имеем

$$(A_1+A_2+\dots+A_{k-1})A_k=A_1A_k+A_2A_k+\dots+A_{k-1}A_k=\varnothing,$$

откуда

$$P(A_1+A_2+\dots+A_{k-1}+A_k)=P((A_1+A_2+\dots+A_{k-1})+A_k)=$$
$$=P(A_1+A_2+\dots+A_{k-1})+P(A_k)=P(A_1)+P(A_2)+\dots+P(A_{k-1})+P(A_k).$$

Следовательно, по индукции формула (5) верна для любого конечного числа слагаемых.

Говорят, что события $A_1, A_2, \dots, A_n$ образуют *полную группу событий*, если они попарно несовместны и их сумма является достоверным событием:

1) $A_iA_j=\varnothing \quad \forall i,\ j=1,2,\dots,n;\ i\neq j;$

2) $A_1+A_2+\dots+A_n=\Omega.$

Следствие 3. Сумма вероятностей событий $A_1, A_2, \dots, A_n,$ образующих полную группу, равна 1.

Действительно, по формуле (5)

$$P(A_1)+P(A_2)+\dots+P(A_n)=P(A_1+A_2+\dots+A_n)=P(\Omega)=1.$$

Следствие 4. Сумма вероятностей противоположных событий равна 1, т. е.

$$P(A)+P(\overline{A})=1. \qquad (6)$$

Действительно, события A и $\overline{A}$ образуют полную группу событий, откуда в силу следствия 3 и вытекает равенство (6).

Пример. В ящике находится два белых, три красных и пять синих шаров. Какова вероятность, что шар, случайным образом извлеченный из ящика, будет цветным (т. е. не белым)?

Решение. Пусть событие A – «из ящика извлечен красный шар», событие B – «из ящика извлечен синий шар». Тогда событие $A+B$ – «из ящика извлечен цветной шар». Поскольку события A и B несовместны, то

$$P(A+B)=P(A)+P(B)=0{,}3+0{,}5=0{,}8.$$

Пример. Бросают две игральные кости. Какова вероятность выпадения хотя бы одной шестерки?

Решение. Пусть событие A – «выпадение шестерки на первой кости», а событие B – «выпадение шестерки на второй кости». Тогда событие $A+B$ – «выпадение хотя бы одной шестерки при бросании двух игральных костей». В данном примере не исключен случай, что события A и B могут произойти одновременно. Очевидно, что $P(A)=\frac{1}{6}$, $P(B)=\frac{1}{6}$, $P(AB)=\frac{1}{36}$, откуда

$$P(A+B)=P(A)+P(B)-P(AB)=\frac{1}{6}+\frac{1}{6}-\frac{1}{36}=\frac{11}{36}.$$

Пример. Студент первого курса исторического факультета выучил 10 вопросов из 30 по курсу «Высшая математика». Каждый зачетный билет состоит из трех тестовых вопросов, распределенных случайным образом. Найти вероятность того, что студент ответит хотя бы на один вопрос из вытянутого наугад билета.

Решение. Рассмотрим событие A – «студент ответит хотя бы на один вопрос из билета». Тогда противоположное ему событие $\overline{A}$ –«студент не ответит ни на один вопрос из билета», и выполняется равенство $P(A)+P(\overline{A})=1.$ Так как порядок вопросов в билетах несущественен, то общее количество исходов такого испытания равно количеству способов выбора трех вопросов из всех 30 вопросов, т. е. $n=C_{30}^{3}$. Поскольку студент выучил 10 вопросов по курсу «Высшая математика», то $30-10=20$ вопросов остались невыученными. Значит, количество благоприятствующих исходов для события $\overline{A}$ равно числу способов выбора трех вопросов из 20 вопросов, невыученных студентом, т. е. $m=C_{20}^{3}$. Поэтому по формуле классической вероятности имеем

$$P(\overline{A})=\frac{m}{n}=\frac{C_{20}^{3}}{C_{30}^{3}}=\frac{20!}{3!\cdot 17!}\cdot\frac{3!\cdot 27!}{30!}=\frac{20\cdot 19\cdot 18}{30\cdot 29\cdot 28}=\frac{19\cdot 3}{29\cdot 7}=\frac{57}{203}.$$

Следовательно, вероятность того, что студент ответит хотя бы на один вопрос из вытянутого наугад билета, равна

$$P(A)=1-P(\overline{A})=1-\frac{57}{203}=\frac{146}{203}.$$

Условная вероятность. Теоремы умножения вероятностей

В некоторых случаях вероятность наступления события зависит от того, предшествовало или нет ему другое событие. Такую вероятность называют условной. Прежде чем дать точное определение условной вероятности, рассмотрим следующий пример.

Пример. Пусть в ящике находится 10 шаров: шесть белых и четыре черных. Из ящика вынимаются наудачу один за другим два шара. Рассмотрим следующие события: B – «первый вынутый шар белый», A – «второй вынутый шар белый».

Очевидно, что $P(B)=0,6.$ Если первый извлеченный шар был белым, т. е. событие B произошло, то среди оставшихся девяти шаров будет пять белых и поэтому вероятность события A будет равна $\frac{5}{9}$. Если же первый извлеченный шар был черным, т. е. событие B не произошло, то среди оставшихся девяти шаров окажется шесть белых и поэтому вероятность события A будет равна $\frac{6}{9}$

или $\frac{2}{3}$. Таким образом, вероятность события A зависит от того, произошло или не произошло событие B.

Определение. **Пусть вероятность события B – положительная величина, т. е. $P(B) > 0$. Вероятность события A при условии, что произошло событие B, называется условной вероятностью события A и обозначается $P(A \mid B)$ или $P_B(A)$.**

Рассмотрим, как вычисляется условная вероятность, исходя из классического определения вероятности.

Пусть $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, событию A благоприятствуют m элементарных событий, событию B – k элементарных событий, а событию AB – l элементарных событий ($l \leq m, l \leq k, m \neq 0, k \neq 0$). Если событие B произошло, то в этой ситуации возможны лишь те k элементарных событий, которые благоприятствовали событию B, причем l из них, очевидно, благоприятствуют событию A. Таким образом,

$$P(A \mid B) = \frac{l}{k} = \frac{\frac{l}{n}}{\frac{k}{n}} = \frac{P(AB)}{P(B)}.$$

Аналогично,

$$P(B \mid A) = \frac{l}{m} = \frac{\frac{l}{n}}{\frac{m}{n}} = \frac{P(AB)}{P(A)}.$$

Дадим теперь общее определение условной вероятности события, применимое не только для классической вероятности.

Определение. **Пусть вероятность события B – положительная величина, т. е. $P(B) > 0$. Условной вероятностью события A, при условии, что произошло событие B, называют число**

$$\boldsymbol{P(A \mid B) = \frac{P(AB)}{P(B)}}.$$

Аналогично, если $P(A) > 0$, то условной вероятностью события B, при условии, что произошло событие A, называют число

$$P(B \mid A) = \frac{P(AB)}{P(A)}.$$

Из определения условной вероятности следует *теорема умножения вероятностей*:

Теорема 5. Пусть $P(A) > 0$, $P(B) > 0$. Тогда

$$P(AB) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B),$$

т. е. вероятность произведения (пересечения) двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое событие произошло.

Пример. В ящике находится шесть белых и восемь черных шаров. Из ящика вынули два шара (не возвращая первый вынутый шар в ящик). Найти вероятность того, что оба вынутых шара белые.

Решение. Пусть событие A – «первый раз вынут белый шар», событие B – «второй раз вынут белый шар». Тогда $P(A)=\frac{6}{14}=\frac{3}{7}$, $P(B\,|\,A)=\frac{5}{13}$ и по теореме 5 вероятность события AB, заключающегося в том, что оба вынутых шара белые, равна

$$P(AB)=P(A)\cdot P(B\,|\,A)=\frac{3}{7}\cdot\frac{5}{13}=\frac{15}{91}.$$

Пример. Известно, что число жарких дней июля на протяжении десяти лет составило 73 дня, а число влажных дней для этого же месяца – 50 дней, причем только 14 влажных дней были жаркими. Найти вероятность того, что какой-либо день июля будет одновременно жарким и влажным. При условии, что выбранный день июля оказался жарким, найти вероятность того, что данный день является влажным.

Решение. Обозначим через A событие, состоящее в том, что наугад выбранный день июля является жарким, через B – событие, состоящее в том, что наугад выбранный день июля является влажным. Задача сводится к нахождению вероятностей $P(AB)$ и $P(B\,|\,A)$.

Согласно условию задачи $P(A)=\frac{73}{310}\approx 0,24,$ $P(B)=\frac{50}{310}\approx 0{,}16$ и $P(A\,|\,B)=\frac{14}{50}=0,28.$ По теореме 9.5 имеем

$$P(AB)=P(B)\cdot P(A\,|\,B)=\frac{50}{310}\cdot\frac{14}{50}=\frac{14}{310}\approx 0,045.$$

С другой стороны, $P(AB)=P(A)P(B\,|\,A),$ откуда

$$P(B\,|\,A)=\frac{P(AB)}{P(A)}=\frac{0,045}{0,24}\approx 0{,}188.$$

Теорему 5 можно распространить на случай n сомножителей.

Теорема 6. Пусть $P(A_1A_2\cdots A_{n-1})>0.$ Тогда

$$P(A_1A_2\cdots A_n)=P(A_1)\cdot P(A_2\,|\,A_1)\cdot P(A_3\,|\,A_1A_2)\cdot\ldots\cdot P(A_n\,|\,A_1A_2\cdots A_{n-1}),$$

т. е. вероятность произведения нескольких событий равна произведению вероятностей этих событий, причем вероятность каждого следующего по порядку события вычисляется при условии, что все предыдущие события имели место.

Доказательство теоремы проводится методом математической индукции с применением теоремы 5.

Пример. Студент с практики отправил на кафедру 15 образцов пород, 10 из которых составляет руда. В отсутствие руководителя практики четыре образца случайным образом поместили на полку. Найти вероятность того, что все помещенные образцы оказались рудой.

Решение. Введем обозначения событий: A_1 – «первый помещенный на полку образец является рудой», A_2 – «второй помещенный на полку образец является рудой», A_3 – «третий помещенный на полку образец является рудой»,

A_4 – «четвертый помещенный на полку образец является рудой». По теореме 6 искомая вероятность равна

$$P(A_1A_2A_3A_4) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)P(A_4 \mid A_1A_2A_3) =$$

$$= \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} = \frac{2}{13} \approx 0{,}154.$$

Независимость событий

Определение. **Событие *A* называется независимым от события *B*, если**

$$\boldsymbol{P(A \mid B) = P(A),} \qquad (7)$$

т. е. вероятность события *A* не зависит от того, произошло или нет событие *B*.

В силу теоремы 5 $P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$, откуда с учетом (7) получаем $P(A) \cdot P(B \mid A) = P(B) \cdot P(A)$ или, сокращая на $P(A) > 0$, $P(B \mid A) = P(B)$. Последнее равенство означает, что событие *B* не зависит от события *A*. Таким образом, свойство независимости событий является взаимным, поэтому можно говорить просто о *независимых событиях* *A* и *B*.

Для независимых событий *A* и *B* имеет место следующая теорема умножения вероятностей, которая является следствием теоремы 5.

Теорема 7. Вероятность произведения двух независимых событий равна произведению вероятностей этих событий:

$$P(AB) = P(A) \cdot P(B). \qquad (8)$$

Замечание 1. Равенство (8) иногда принимают за определение независимости событий *A* и *B*. При таком определении вероятности событий *A* и *B* могут быть равными нулю.

Независимость более чем двух событий является понятием более сложным, чем двух событий. Введем понятие независимости *n* событий.

События $A_1, A_2, \dots, A_n$ называются *независимыми в совокупности* или *независимыми*, если для любого подмножества индексов $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$ выполняется равенство

$$P(A_{i_1}A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}).$$

Так, для трех событий *A*, *B* и *C* независимость в совокупности означает выполнение четырех равенств:

$$P(AB) = P(A) \cdot P(B), \quad P(AC) = P(A) \cdot P(C), \quad P(BC) = P(B) \cdot P(C),$$

$$P(ABC) = P(A) \cdot P(B) \cdot P(C).$$

Очевидно, что независимость в совокупности сильнее попарной независимости событий, т. е. из нее вытекает попарная независимость. Обратное, вообще говоря, неверно.

Замечание 2. Если события *A* и *B* независимы, то независимы также события $\overline{A}$ и *B*, *A* и $\overline{B}$, $\overline{A}$ и $\overline{B}$ (доказательство предлагается читателю в качестве упражнения).

Замечание 3. Если события $A_1, A_2, \dots, A_n$ независимы в совокупности, то противоположные им события $\overline{A}_1, \overline{A}_2, \dots, \overline{A}_n$ также независимы в совокупности.

Пример. Два стрелка стреляют по мишени с вероятностями попадания 0,7 и 0,8. Найти вероятность того, что цель будет поражена.

Решение. Для поражения цели необходимо попадание хотя бы одного стрелка. Пусть событие A – «первый стрелок попал в цель», событие B – «второй стрелок попал в цель». Тогда событие $A+B$ – «цель поражена». События A и B являются независимыми, поэтому $P(AB)=P(A)\cdot P(B)=0,7\cdot 0,8=0,56$ и по теореме сложения вероятностей

$$P(A+B)=P(A)+P(B)-P(AB)=0,7+0,8-0,56=0,94.$$

Пример. Наблюдениями установлено, что в некоторой местности в сентябре бывает 12 дождливых дней. Найти вероятность того, что из случайно зафиксированных в этом месяце семи дней ни один не окажется дождливым.

Решение. Наблюдения в условиях данной задачи являются независимыми. Вероятность выпадения дождя в любой день сентября равна $p=\dfrac{12}{30}=0,4,$ а вероятность того, что в любой день сентября дождя не будет, равна $q=1-p=1-0,4=0,6.$ Пусть A_i – событие, состоящее в том, что в i-й день сентября дождя не будет; $i=1,2,\dots,7.$ Тогда вероятность события B – «из случайно зафиксированных семи дней ни один не окажется дождливым» – равна

$$P(B)=P(A_1A_2A_3A_4A_5A_6A_7)=P(A_1)P(A_2)P(A_3)P(A_4)P(A_5)P(A_6)P(A_7)=$$
$$=q^7=0,6^7\approx 0,028.$$

Формула полной вероятности. Формулы Байеса

Пусть события $H_1, H_2, \dots, H_n$ образуют полную группу событий и $A\subseteq\Omega$. Вероятности событий H_i известны, причем $P(H_i)>0$ для всех $i=1,2,\dots,n$; известны также условные вероятности $P(A\,|\,H_i)$. Требуется найти вероятность события A.

Теорема 8. Пусть события $H_1, H_2, \dots, H_n$ образуют полную группу событий и $P(H_i)>0$ для всех $i=1,2,\dots,n.$ Тогда вероятность произвольного события A может быть найдена по формуле

$$P(A)=\sum_{i=1}^{n}P(H_i)\cdot P(A\,|\,H_i). \tag{9}$$

Доказательство. Представим событие A в виде

$$A=A\Omega=A(H_1+H_2+\dots+H_n)=AH_1+AH_2+\dots+AH_n.$$

Поскольку события $H_1, H_2, \dots, H_n$ попарно несовместны, то события $AH_1, AH_2, \dots, AH_n$ также попарно несовместны. Пользуясь равенством (5) и теоремой 5, получим

$$P(A)=P(AH_1+AH_2+\dots+AH_n)=P(AH_1)+P(AH_2)+\dots+P(AH_n)=$$
$$=P(H_1)\cdot P(A\,|\,H_1)+P(H_2)\cdot P(A\,|\,H_2)+\dots+P(H_n)\cdot P(A\,|\,H_n)=$$
$$=\sum_{i=1}^{n}P(H_i)\cdot P(A\,|\,H_i).$$

Теорема доказана.

Формула (9) называется *формулой полной вероятности*. События $H_1, H_2, \dots, H_n$ иногда называют *гипотезами*.

Пример. Имеются три ящика с шарами: в первом – три белых и четыре черных шара, во втором – десять белых шаров, в третьем – два белых и пять черных шаров. Наудачу выбирается один из ящиков и из него извлекают один шар. Какова вероятность того, что извлеченный шар окажется черным?

Решение. Введем в рассмотрение события: H_1 – «шар извлечен из первого ящика», H_2 – «шар извлечен из второго ящика», H_3 – «шар извлечен из третьего ящика», A – «извлеченный шар оказался черным». Тогда согласно условию задачи

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3},$$

$$P(A\,|\,H_1) = \frac{4}{7}, \quad P(A\,|\,H_2) = 0, \quad P(A\,|\,H_3) = \frac{5}{7},$$

откуда по формуле полной вероятности

$$P(A) = P(H_1)\cdot P(A\,|\,H_1) + P(H_2)\cdot P(A\,|\,H_2) + P(H_3)\cdot P(A\,|\,H_3) =$$

$$= \frac{1}{3}\cdot\frac{4}{7} + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot\frac{5}{7} = \frac{3}{7}.$$

Пример. На город примерно 100 дней в году дует ветер с севера и 200 дней в году – с запада (считается, что год невисокосный и с других направлений ветер дуть не может). Промышленные предприятия, расположенные на севере, производят выброс вредных веществ каждый третий день, а расположенные на западе – в последний день каждой недели. Определить, как часто город подвергается воздействию вредных выбросов. Иными словами, какова вероятность того, что в наугад выбранный день город будет накрыт промышленным смогом?

Решение. Пусть событие H_1 – «ветер дует с севера», событие H_2 – «ветер дует с запада», событие A – «город подвергается воздействию вредных выбросов». Тогда согласно условию задачи

$$P(H_1) = \frac{100}{365} = \frac{20}{73}, \quad P(H_2) = \frac{200}{365} = \frac{40}{73}, \quad P(A\,|\,H_1) = \frac{1}{3}, \quad P(A\,|\,H_2) = \frac{1}{7}.$$

Отсюда по формуле полной вероятности

$$P(A) = P(H_1)P(A\,|\,H_1) + P(H_2)P(A\,|\,H_2) = \frac{20}{73}\cdot\frac{1}{3} + \frac{40}{73}\cdot\frac{1}{7} \approx 0{,}17.$$

Таким образом, около двух месяцев в году город накрыт промышленным смогом.

Пусть события $H_1, H_2, \dots, H_n$ образуют полную группу и их вероятности $P(H_i)$, $i = 1, 2, \dots, n$, известны до проведения опыта (так называемые *априорные вероятности*). Производится опыт, в результате которого происходит событие A. Каковы будут вероятности событий H_i после опыта, т. е. после того как событие A уже наступило?

Искомые вероятности $P(H_i \mid A)$ носят название *апостериорных* и находятся путем использования теоремы умножения вероятностей и формулы полной вероятности. Имеем

$$P(AH_i) = P(A) \cdot P(H_i \mid A) = P(H_i) \cdot P(A \mid H_i)$$

(здесь предполагается, что $P(A) > 0$ и $P(H_i) > 0$ для всех $i = 1, 2, \ldots, n$), откуда

$$P(H_i \mid A) = \frac{P(H_i) \cdot P(A \mid H_i)}{P(A)}.$$

Подставляя в последнее равенство выражение для $P(A)$ из формулы полной вероятности (9), получим

$$P(H_i \mid A) = \frac{P(H_i) \cdot P(A \mid H_i)}{\sum_{k=1}^{n} P(H_k) \cdot P(A \mid H_k)}, \quad i = 1, 2, \ldots, n. \qquad (10)$$

Формулы (10) называют *формулами Байеса*. Они отвечают на поставленный вопрос.

Пример. Имеются три ящика с шарами: в первом – три белых и четыре черных шара, во втором – десять белых шаров, в третьем – два белых и пять черных шаров. Из наугад взятого ящика наугад вынимается шар, который оказывается белым. Найти вероятность того, что шар взят из первого ящика.

Решение. Введем в рассмотрение события: H_1 – «шар извлечен из первого ящика», H_2 – «шар извлечен из второго ящика», H_3 – «шар извлечен из третьего ящика», A – «извлеченный шар оказался белым». Тогда согласно условию задачи

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3},$$

$$P(A \mid H_1) = \frac{3}{7}, \quad P(A \mid H_2) = 1, \quad P(A \mid H_3) = \frac{2}{7},$$

откуда по формуле полной вероятности

$$P(A) = P(H_1) \cdot P(A \mid H_1) + P(H_2) \cdot P(A \mid H_2) + P(H_3) \cdot P(A \mid H_3) =$$

$$= \frac{1}{3} \cdot \frac{3}{7} + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{2}{7} = \frac{4}{7}.$$

Искомая вероятность равна

$$P(H_1 \mid A) = \frac{P(H_1) \cdot P(A \mid H_1)}{P(A)} = \frac{\frac{1}{3} \cdot \frac{3}{7}}{\frac{4}{7}} = \frac{1}{4}.$$

Пример. На складе находятся детали, изготовленные на двух заводах, причем деталей первого завода 80 %, а второго – 20 % от общего количества. Вероятность брака на первом заводе равна 0,05, на втором заводе – 0,01. Наудачу взятая деталь оказалась бракованной. Какова вероятность того, что эта деталь изготовлена первым заводом?

Решение. Пусть событие H_1 – «взятая деталь изготовлена первым заводом», H_2 – «взятая деталь изготовлена вторым заводом», A – «взятая деталь оказалась бракованной». Тогда согласно условию задачи

$$P(H_1)=0{,}8,\quad P(H_2)=0{,}2,\quad P(A\,|\,H_1)=0{,}05,\quad P(A\,|\,H_2)=0{,}01.$$

Искомая вероятность равна

$$P(H_1\,|\,A)=\frac{P(H_1)\cdot P(A\,|\,H_1)}{P(H_1)\cdot P(A\,|\,H_1)+P(H_2)\cdot P(A\,|\,H_2)}=$$

$$=\frac{0{,}8\cdot 0{,}05}{0{,}8\cdot 0{,}05+0{,}2\cdot 0{,}01}=\frac{20}{21}\approx 0{,}952.$$

Повторение испытаний. Схема Бернулли

Пусть независимо друг от друга проводятся n испытаний, в каждом из которых возможны лишь два исхода: либо происходит событие A («успех»), либо противоположное ему событие $\overline{A}$ («неудача»). Будем считать, что испытания проводятся в одинаковых условиях и вероятность появления события A в каждом испытании одна и та же. Обозначим эту вероятность через p, а вероятность появления противоположного события $\overline{A}$ через q $(q=1-p)$. Данная схема носит название *схемы Бернулли*.

Поставим вопрос о вероятности того, что в данной серии из n испытаний событие A наступило ровно m раз $(0\le m\le n)$. Искомую вероятность обозначим через $P_n(m)$.

Теорема 9. Вероятность $P_n(m)$ наступления события A ровно m раз в последовательности из n испытаний схемы Бернулли равна

$$P_n(m)=C_n^m p^m q^{n-m}. \qquad (11)$$

Доказательство. Событие A в проведенных n испытаниях может наступить m раз в разных вариантах, например в варианте

$$\underbrace{A\ldots A}_{m}\underbrace{\overline{A}\ldots\overline{A}}_{n-m},$$

где событие A занимает m первых мест (произошло в первых m испытаниях), а событие $\overline{A}$ занимает последующие $n-m$ мест (произошло в последующих $n-m$ испытаниях). Всего таких вариантов существует C_n^m, причем все составленные таким образом сложные события являются несовместными. Поскольку все события в каждом из вариантов независимы, то по теореме умножения вероятностей независимых событий вероятность каждого варианта равна $p^m q^{n-m}$. Так как все варианты-события несовместны, то по теореме сложения вероятностей несовместных событий искомая вероятность равна сумме вероятностей всех возможных событий, т. е. $C_n^m p^m q^{n-m}$, что и доказывает теорему.

Вероятности $P_n(m)$ называются *биномиальными вероятностями*, а формула (11) – *формулой Бернулли*.

Пример. Найти вероятность того, что при десятикратном бросании монеты «орел» выпадет ровно 5 раз.

Решение. В данном случае $p = q = \frac{1}{2}$, $n = 10$, $m = 5$. Отсюда находим

$$P_{10}(5) = C_{10}^5 \cdot \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right)^5 = \frac{10!}{5! \cdot 5!} \cdot \left(\frac{1}{2}\right)^{10} = \frac{252}{1024} \approx 0,246.$$

Пример. Вероятность попадания в мишень при одном выстреле для данного стрелка равна 0,7 и не зависит от номера выстрела. Найти вероятность того, что при пяти выстрелах произойдет ровно два попадания в мишень.

Решение. В данном случае $p = 0,7$, $q = 0,3$, $n = 5$, $m = 2$. Отсюда находим

$$P_5(2) = C_5^2 \cdot 0,7^2 \cdot 0,3^3 = 10 \cdot 0,49 \cdot 0,027 = 0,1323.$$

Вероятность того, что в n испытаниях схемы Бернулли событие A наступит: а) менее m раз; б) более m раз; в) не менее m раз; г) не более m раз, находят соответственно по формулам:

$$P_n(0) + P_n(1) + \ldots + P_n(m-1);$$

$$P_n(m+1) + P_n(m+2) + \ldots + P_n(n);$$

$$P_n(m) + P_n(m+1) + \ldots + P_n(n);$$

$$P_n(0) + P_n(1) + \ldots + P_n(m).$$

1.2.4 Случайные величины

Часто встречаются опыты, в результате которых случайным образом получаются числа. Например, при бросании игральной кости мы случайным образом получаем одно из чисел: $1, 2, 3, 4, 5, 6$ или при измерениях получаются случайные ошибки и т. п. В таких случаях говорят, что мы имеем дело со случайной величиной.

Случайной величиной называют величину, которая в результате испытания принимает одно и только одно возможное значение, наперед неизвестное и зависящее от случайных причин, которые заранее не могут быть учтены. На примере с бросанием игральной кости мы видим, что каждому исходу опыта ставится в соответствие единственное число $k \in \{1, 2, 3, 4, 5, 6\}$ – значение случайной величины. Поэтому естественно рассматривать случайную величину как функцию, заданную на множестве исходов данного опыта.

Определение. **Пусть Ω – пространство элементарных событий. Числовую функцию от элементарного события $\omega \in \Omega$ назовем случайной величиной.**

Таким образом, если каждому элементарному событию ω можно поставить в соответствие некоторое число, то говорят, что задана случайная величина. Случайные величины будем обозначать прописными латинскими буквами X, Y, Z и т. д., а их возможные значения – соответствующими строчными латинскими буквами x, y, z и т. д. Например, если случайная величина X имеет три возможных значения, то они будут обозначены x_1, x_2, x_3.

Случайная величина, принимающая значения, которые можно записать в виде конечного набора или счетной последовательности чисел, называется *дискретной*, т. е. дискретная случайная величина принимает отдельные, изолированные возможные значения, число которых конечно или счетно.

Случайная величина, которая может принимать все значения из некоторого числового промежутка, называется *непрерывной* (например, ошибка измерения, время безотказной работы прибора и т. д.).

Каждому значению x_n случайной величины дискретного типа отвечает определенная вероятность p_n; каждому промежутку (a, b) из области значений случайной величины непрерывного типа также отвечает определенная вероятность p того, что значение x, принятое случайной величиной, попадет в этот промежуток. Соотношение, устанавливающее тем или иным способом связь между возможными значениями случайной величины и их вероятностями, называется *законом распределения случайной величины*.

Закон распределения дискретной случайной величины обычно задается в виде таблицы:

X	x_1	x_2	x_3	$\ldots$	x_{n-1}	x_n	$\ldots$
p	p_1	p_2	p_3	$\ldots$	p_{n-1}	p_n	$\ldots$

В верхней строке записывают возможные значения x_i случайной величины X, а в нижней – их вероятности $p_i = P(X = x_i)$. Так как события $A_i = \{X = x_i\}$, $i = 1, 2, \ldots,$ образуют полную группу событий, то $\sum_{i=1}^{\infty} p_i = p_1 + p_2 + \ldots = 1$. В случае конечного числа значений случайной величины, равного n, справедливо равенство $\sum_{i=1}^{n} p_i = p_1 + p_2 + \cdots + p_n = 1$.

В целях наглядности закон распределения дискретной случайной величины можно изобразить графически: построить точки (x_i, p_i) в декартовой прямоугольной системе координат и соединить их отрезками прямых. Полученная фигура называется *многоугольником распределения*.

Пример. Даны вероятности значений случайной величины X: значение 10 имеет вероятность 0,3; значение 2 – вероятность 0,4; значение 8 – вероятность 0,1 и значение 4 – вероятность 0,2. Запишем эти значения в таблицу и построим многоугольник распределения (рисунок 17):

X	2	4	8	10
p	0,4	0,2	0,1	0,3

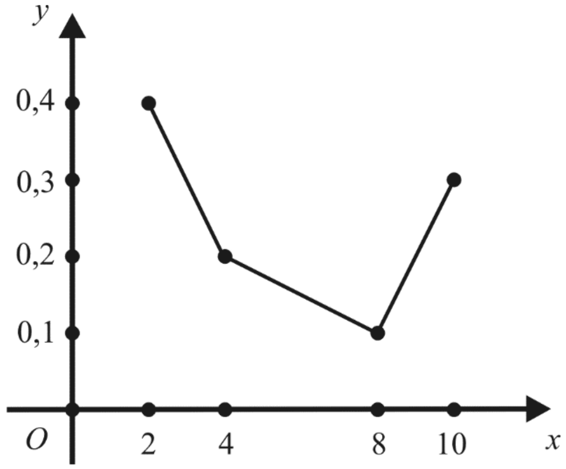


Рис.17

Построенная таблица задает закон распределения случайной величины, так как

$$\sum_{i=1}^{4} p_i = p_1 + p_2 + p_3 + p_4 = 0,4 + 0,2 + 0,1 + 0,3 = 1.$$

Пример. Подбрасываются две монеты и подсчитывается количество выпавших «орлов» на обеих верхних сторонах монет. Рассматривается дискретная случайная величина X – число выпадения «орлов» на обеих монетах. Записать закон распределения этой случайной величины.

Решение. В данном испытании $\Omega = \{(o, o), (o, p), (p, o), (p, p)\}$, где o означает, что выпал «орел», p – выпала «решка». «Орел» может выпасть 1 раз, 2 раза или не появиться ни разу. Следовательно, случайная величина X может принимать только три значения: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$. Вероятности этих значений равны соответственно

$$p_1 = P(X = 0) = \frac{1}{4}, \quad p_2 = P(X = 1) = \frac{2}{4} = \frac{1}{2}, \quad p_3 = P(X = 2) = \frac{1}{4}.$$

Таким образом, закон распределения случайной величины X имеет вид

X	0	1	2
p	1/4	1/2	1/4

Пример. Стрелок производит по мишени три выстрела. Вероятность попадания в мишень при каждом выстреле равна 0,3. Случайная величина X – число попаданий. Записать закон ее распределения.

Решение. Воспользуемся формулой Бернулли

$$P_n(m) = C_n^m p^m q^{n-m}.$$

В данном случае $p = 0,3$, $q = 0,7$, $n = 3$, откуда

$$P_3(0) = C_3^0 \cdot 0,3^0 \cdot 0,7^3 = 0,343,$$

$$P_3(1) = C_3^1 \cdot 0,3^1 \cdot 0,7^2 = 3 \cdot 0,3 \cdot 0,49 = 0,441,$$

$$P_3(2) = C_3^2 \cdot 0,3^2 \cdot 0,7^1 = 3 \cdot 0,09 \cdot 0,7 = 0,189,$$

$$P_3(3) = C_3^3 \cdot 0,3^3 \cdot 0,7^0 = 0,027.$$

Закон распределения случайной величины X имеет вид

X	0	1	2	3
p	0,343	0,441	0,189	0,027

Числовые характеристики дискретных случайных величин

Закон распределения полностью задает дискретную случайную величину. Однако часто встречаются случаи, когда закон распределения случайной величины неизвестен. В таких ситуациях случайную величину изучают по ее числовым характеристикам, основными из которых являются математическое ожидание, дисперсия и среднее квадратическое отклонение.

Математическое ожидание дискретной случайной величины

Определение. **Математическим ожиданием $M(X)$ дискретной случайной величины X называют сумму произведений всех ее возможных значений на соответствующие вероятности.**

Если дискретная случайная величина X принимает конечное число значений $x_1, x_2, \dots, x_n$ с вероятностями $p_1, p_2, \dots, p_n$ соответственно, то по определению

$$M(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_n p_n.$$

Пример. Подбрасывается игральная кость. Найти математическое ожидание дискретной случайной величины X, равной числу выпавших очков.

Решение. Закон распределения случайной величины X имеет вид

X	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Следовательно, по определению

$$M(X) = (1+2+3+4+5+6)\cdot\frac{1}{6} = 3,5.$$

Отметим, что постоянную величину C можно рассматривать как дискретную случайную величину, принимающую лишь одно значение $x = C$ с вероятностью $p = 1$. Поэтому $M(C) = C \cdot 1 = C,$ т. е. математическое ожидание постоянной величины равно самой этой величине.

Приведем без доказательства важнейшие свойства математического ожидания для дискретных случайных величин.

1. Постоянный множитель можно выносить за знак математического ожидания, т. е.

$$M(CX) = C\,M(X).$$

2. Математическое ожидание суммы двух случайных величин X и Y равно сумме их математических ожиданий:

$$M(X+Y) = M(X) + M(Y).$$

Прежде чем сформулировать следующее свойство, укажем, что две случайные величины X и Y называют *независимыми*, если закон распределения каждой из них не зависит от того, какие возможные значения приняла другая величина. Несколько случайных величин независимы, если закон распределения любой из них не зависит от того, какие возможные значения приняли остальные случайные величины.

3. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY) = M(X)M(Y).$$

Замечание. Свойства 2 и 3 имеют место для любого конечного числа случайных величин.

Пример. Найти математическое ожидание случайной величины $Z = 2X + 3Y + 7$, если $M(X) = 4$, $M(Y) = 1$.

Решение. Используя свойства математического ожидания, имеем

$$M(Z) = M(2X + 3Y + 7) = M(2X) + M(3Y) + M(7) =$$
$$= 2M(X) + 3M(Y) + 7 = 2 \cdot 4 + 3 \cdot 1 + 7 = 18.$$

Пример. Независимые случайные величины заданы законами распределения:

X	1	2
p	0,6	0,4

Y	4	6
p	0,2	0,8

Найти математическое ожидание случайной величины XY.

Решение. Найдем математические ожидания каждой из данных случайных величин:

$$M(X) = 1 \cdot 0,6 + 2 \cdot 0,4 = 1,4, \quad M(Y) = 4 \cdot 0,2 + 6 \cdot 0,8 = 5,6.$$

В силу независимости случайных величин X и Y искомое математическое ожидание

$$M(XY) = M(X)M(Y) = 1,4 \cdot 5,6 = 7,84.$$

Можно показать что при достаточно большом числе испытаний математическое ожидание дискретной случайной величины приближенно равно среднему арифметическому всех ее значений. В связи с этим математическое ожидание случайной величины называют также ее *средним значением*.

Дисперсия и среднее квадратическое отклонение дискретной случайной величины

Математическое ожидание характеризует случайную величину не полностью. Зная математическое ожидание, нельзя сказать, какие значения принимает случайная величина и как они отклоняются от среднего значения. Например, при одинаковой средней величине годовых осадков одна местность может быть засушливой и неблагоприятной для сельскохозяйственных работ (нет дождей весной и летом), а другая – благоприятной для ведения сельского хозяйства. Чтобы знать, как рассеяны значения случайной величины вокруг ее математического ожидания, вводят такую числовую характеристику как дисперсия.

Пусть X – дискретная случайная величина, $M(X)$ – ее математическое ожидание.

Величина $X-M(X)$ называется *отклонением* случайной величины *X* от ее математического ожидания. Отклонение является случайной величиной и его математическое ожидание

$$M(X-M(X))=M(X)-M(M(X))=M(X)-M(X)=0.$$

Из полученного равенства следует, что с помощью отклонения невозможно определить среднее отклонение возможных значений случайной величины *X* от ее математического ожидания, т. е. степень рассеяния случайной величины *X*. Это объясняется взаимным погашением возможных положительных и отрицательных значений отклонения. Данного недостатка можно избежать, если рассматривать квадрат отклонения случайной величины *X*.

Определение. **Дисперсией $D(X)$ дискретной случайной величины X называют математическое ожидание квадрата ее отклонения:**

$$\boldsymbol{D(X)=M((X-M(X))^2).}$$

Дисперсия случайной величины постоянна, т. е. является числовой характеристикой этой величины.

Если дискретная случайная величина *X* принимает конечное число значений $x_1, x_2, \dots, x_n$ с вероятностями $p_1, p_2, \dots, p_n$ соответственно, то закон распределения случайной величины $(X-M(X))^2$ имеет вид:

$(X-M(X))^2$	$(x_1-M(X))^2$	$(x_2-M(X))^2$	…	$(x_n-M(X))^2$
p	p_1	p_2	…	p_n

Исходя из определения математического ожидания, получаем

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

Дисперсия дискретной случайной величины обладает следующими свойствами.

1. Дисперсия – величина неотрицательная, т. е. $D(X)\geq 0$.

2. Дисперсия случайной величины *X* равна разности между математическим ожиданием квадрата случайной величины *X* и квадратом ее математического ожидания:

$$D(X)=M(X^2)-(M(X))^2.$$

Действительно, используя свойства математического ожидания, имеем

$$D(X)=M((X-M(X))^2)=M(X^2-2XM(X)+(M(X))^2)=$$
$$=M(X^2)-2M(X)M(X)+(M(X))^2=$$
$$=M(X^2)-2(M(X))^2+(M(X))^2=M(X^2)-(M(X))^2.$$

3. Дисперсия постоянной величины *C* равна нулю:

$$D(C)=0.$$

Действительно, $D(C)=M((C-M(C))^2)=M((C-C)^2)=M(0)=0.$

4. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(CX)=C^2D(X).$$

Из определения дисперсии и свойств математического ожидания следует, что

$$D(CX)=M((CX-M(CX))^2)=M((CX-CM(X))^2)=$$
$$=M(C^2(X-M(X))^2)=C^2M((X-M(X))^2)=C^2D(X).$$

5. Дисперсия суммы двух независимых случайных величин X и Y равна сумме дисперсий этих величин:

$$D(X+Y)=D(X)+D(Y).$$

Применяя свойство 2 дисперсии и свойства математического ожидания, имеем

$$D(X+Y)=M((X+Y)^2)-(M(X+Y))^2=$$
$$=M(X^2+2XY+Y^2)-(M(X)+M(Y))^2=$$
$$=M(X^2)+2M(XY)+M(Y^2)-((M(X))^2+2M(X)M(Y)+(M(Y))^2)=$$
$$=M(X^2)+2M(X)M(Y)+M(Y^2)-(M(X))^2-2M(X)M(Y)-(M(Y))^2=$$
$$=M(X^2)-(M(X))^2+M(Y^2)-(M(Y))^2=D(X)+D(Y).$$

Замечание. Свойство 5 распространяется на случай любого конечного числа случайных величин.

6. Если C – постоянная, то

$$D(X+C)=D(X).$$

Действительно,

$$D(X+C)=M(((X+C)-M(X+C))^2)=$$
$$=M((X+C-(M(X)+C))^2)=M((X-M(X))^2)=D(X).$$

Определение. **Квадратный корень из дисперсии случайной величины X называется ее средним квадратическим отклонением и обозначается $\sigma(X)$:**

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

Введение среднего квадратического отклонения объясняется тем, что дисперсия измеряется в квадратных единицах относительно размерности самой случайной величины. Например, если возможные значения некоторой случайной величины измеряются в метрах, то ее дисперсия – в квадратных метрах. В тех случаях, когда нужно иметь числовую характеристику рассеяния возможных значений той же размерности, что и сама случайная величина, используется среднее квадратическое отклонение.

Пример . Дисперсия случайной величины X равна 2. Найти дисперсию случайной величины $Y=5X+3$.

Решение. Согласно свойствам дисперсии имеем

$$D(Y)=D(5X+3)=D(5X)=5^2\cdot D(X)=25\cdot 2=50.$$

Пример . Найти дисперсию и среднее квадратическое отклонение случайной величины X, заданной законом распределения

X	0	1	2
p	0,3	0,5	0,2

Решение. Находим $M(X) = 0 \cdot 0,3 + 1 \cdot 0,5 + 2 \cdot 0,2 = 0,9$. Запишем закон распределения квадрата отклонения случайной величины X, т. е. величины $(X - M(X))^2$:

$(X - M(X))^2$	$(0-0,9)^2$	$(1-0,9)^2$	$(2-0,9)^2$
p	0,3	0,5	0,2

По формуле для дисперсии дискретной случайной величины, принимающей конечное число значений, находим

$$D(X) = (0-0,9)^2 \cdot 0,3 + (1-0,9)^2 \cdot 0,5 + (2-0,9)^2 \cdot 0,2 =$$
$$= 0,81 \cdot 0,3 + 0,01 \cdot 0,5 + 1,21 \cdot 0,2 = 0,49.$$

Отсюда $\sigma(X) = \sqrt{D(X)} = \sqrt{0,49} = 0,7$.

Отметим, что дисперсию случайной величины X можно было найти и по формуле $D(X) = M(X^2) - (M(X))^2$. Запишем закон распределения случайной величины X^2:

X^2	0	1	4
p	0,3	0,5	0,2

Находим $M(X^2) = 0 \cdot 0,3 + 1 \cdot 0,5 + 4 \cdot 0,2 = 1,3$, откуда

$$D(X) = M(X^2) - (M(X))^2 = 1,3 - 0,9^2 = 1,3 - 0,81 = 0,49.$$

Функция распределения вероятностей случайной величины

Дискретная случайная величина может быть задана законом распределения, представляющим собой перечень всех возможных значений этой случайной величины и их вероятностей. Однако такой способ задания не является общим: он неприменим, например, для непрерывных случайных величин. Введение понятия функции распределения вероятностей случайной величины устраняет этот недостаток.

Введем следующие обозначения. Пусть x – действительное число, т.е. $x \in \boldsymbol{R}$. Обозначим через $F(x)$ – вероятность события A, состоящего в том, что случайная величина X примет значение, меньшее x, т. е. $A = \{X < x\}$. Очевидно, что мы получили функцию $F(x)$ от переменной x.

Определение функции распределения. ***Функцией распределения*** *вероятностей случайной величины X называют функцию $F(x)$, определяющую вероятность того, что X примет значение, меньшее x:* $F(x) = P(X < x)$.

Функция распределения содержит в себе всю информацию, заложенную в случайной величине. Поэтому считается, что случайная величина (дискретная либо непрерывная) задана, если задана ее функция распределения.

Свойства функции распределения любой случайной величины:

1. $0 \le F(x) \le 1$.

Это свойство следует из того факта, что функция $F(x)$ есть вероятность.

2. Функция распределения $F(x)$ – неубывающая функция, т. е. если $x_1<x_2$, то $F(x_1)\le F(x_2)$.

Покажем это:

Действительно, пусть $x_1 < x_2$. Событие «случайная величина X примет значение, меньшее x_2» можно представить в виде суммы двух несовместных событий: «X примет значение, меньшее x_1» и «X примет значение, удовлетворяющее неравенствам $x_1<X<x_2$». Обозначим вероятности последних двух событий через $P(X < x_1)$ и $P(x_1 \le X< x_2)$ соответственно. По теореме о вероятности суммы двух несовместных событий имеем:

$$P(X < x_2) = P(X < x_1) + P(x_1 \le X < x_2),$$

откуда

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1).$$

Поскольку вероятность любого события – число неотрицательное, то $P(x_1\le X<x_2)\ge 0$ и, следовательно, $F(x_1) \le F(x_2)$.

3. Вероятность попадания значений случайной величины X в полуинтервал $[a,b)$ равна разности между значениями функции распределения в правом и левом концах этого интервала:

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

4. Если все возможные значения случайной величины X принадлежат интервалу (a, b), то $F(x) = \begin{cases} 0, & при\ x \le a, \\ 1, & при\ \ x \ge b. \end{cases}$

Пример. Случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0 & \text{при } \ x \le -1, \\ \frac{1}{3}x + \frac{1}{3} & \text{при } \ -1 < x \le 2, \\ 1 & \text{при } \ x > 2. \end{cases}$$

Найдите вероятность того, что в результате испытания случайная величина X принимает значения из полуинтервала $[0,1)$.

<u>Решение.</u> Так как на полуинтервале $[0,1)$ функция распределения задается формулой $F(x) = \frac{1}{3}x + \frac{1}{3}$, то

$$P(0 \le X < 1) = F(1) - F(0) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

Определение функции распределения. Функция распределения дискретной случайной величины *X имеет вид*:

$$F(x) = \sum_{x_k < x} P(X = x_k),$$

где суммируются вероятности тех значений случайной величины X, которые меньше x. График функции распределения дискретной случайной величины имеет ступенчатый вид.

Пример. Закон распределения дискретной случайной величины X задан таблицей:

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

Найдите функцию распределения случайной величины X и построить график функции распределения.

Решение. Если $x \le 0$, то событие $A = \{X < x\}$ является невозможным (случайная величина не принимает значений, строго меньших нуля) и, следовательно, $F(x) = 0$.

Если $0 < x \le 1$, то $F(x) = P(X) = 0{,}1$. Действительно, в данной ситуации случайная величина X может принять только одно значение, находящееся левее 1, – значение 0 с вероятностью 0,1.

Если $1 < x \le 2$, то $F(x) = P(X=0) + P(X=1) =0{,}1+0{,}5=0{,}5$. Действительно, $F(x)$ равно вероятности события $A = \{X < x\}$, которое может быть осуществлено, когда случайная величина X примет значение 0 или значение 1. Поскольку два этих события несовместны, то по теореме сложения вероятность события $A = \{X < x\}$ равна сумме вероятностей событий $A_1 = \{X =0\}$ и $A_2 = \{X =1\}$.

Если $x > 2$, то $F(x) = 1$, так как событие $A = \{X < x\}$ является достоверным.

Таким образом, получаем функцию распределения вида:

$$F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ 0{,}1 & \text{при } 0 < x \le 1, \\ 0{,}5 & \text{при } 1 < x \le 2, \\ 1 & \text{при } x > 2. \end{cases}$$

График функции $F(x)$ приведен на рисунке 18.

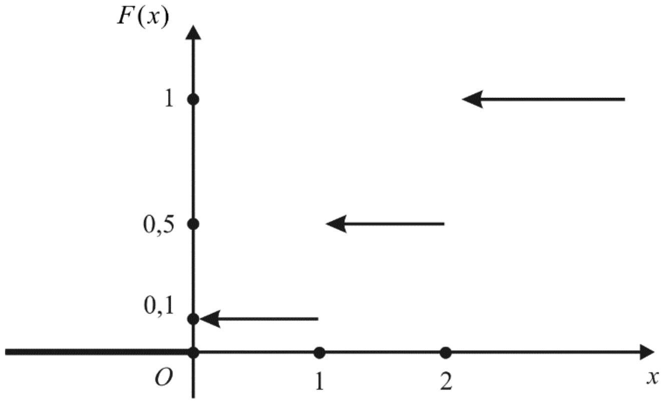


Рис.18.

Как видно из рисунка, функция $F(x)$ является разрывной, причем точками разрыва являются значения x_k, принимаемые случайной величиной X. Величины скачков функции равны вероятностям $p_k = P(X = x_k)$.

1.2.5. Некоторые законы распределения случайных величин

Биномиальное распределение

Рассмотрим n независимых испытаний, в каждом из которых наступает либо событие A с вероятностью p, либо противоположное ему событие $\bar{A}$ с вероятностью $q = 1 - p$ (схема Бернулли). Пусть X – случайная величина, равная числу появлений события A в n испытаниях.

Понятно, что событие A может не наступить вообще, наступить один раз, два раза, $\ldots$, n раз. Следовательно, возможными значениями случайной величины X будут числа $0, 1, 2, \ldots, n$ (дискретная случайная величина с конечным числом значений). По формуле Бернулли можно найти вероятности этих значений:

$$P_n(0) = C_n^0 p^0 q^n = q^n, \quad P_n(1) = C_n^1 p q^{n-1},$$

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

Запишем эти данные в виде следующей таблицы:

X	0	1	2	…	m	…	n
p	q^n	$C_n^1 pq^{n-1}$	$C_n^2 p^2 q^{n-2}$	…	$C_n^m p^m q^{n-m}$	…	p^n

Построенный закон распределения носит название *биномиального распределения*.

Для случайной величины X, имеющей биномиальное распределение, имеем

$$M(X) = np, \quad D(X) = npq.$$

Пример. Вероятность пересыхания реки в засушливый сезон каждого года составляет 0,2. Найти ожидаемое число случаев пересыхания реки за двадцатилетний период.

Решение. Пусть X – число случаев пересыхания реки за двадцатилетний период. Эта случайная величина имеет биномиальное распределение с $n = 20$ и $p = 0,2$. Ожидаемое число случаев пересыхания реки равно $M(X) = np = 20 \cdot 0,2 = 4$.

Распределение Пуассона

Пусть в одинаковых условиях производится серия n независимых испытаний, в каждом из которых может появиться либо событие A с вероятностью p, либо противоположное ему событие $\bar{A}$ с вероятностью $q = 1 - p$. Если n является достаточно большим, а p – достаточно малым, причем $np = a$, где a – некоторое число, то

$$P_n(m) \approx \frac{a^m}{m!} e^{-a}.$$

Говорят, что случайная величина X имеет *распределение Пуассона*, если ее закон распределения задается следующей таблицей:

X	0	1	2	…	m	…
p	e^{-a}	ae^{-a}	$\frac{a^2}{2!}e^{-a}$	…	$\frac{a^m}{m!}e^{-a}$	…

Для случайной величины, распределенной по закону Пуассона, имеют место равенства $M(X) = D(X) = a$.

Распределение Пуассона часто встречается в задачах из области географии. Оно представляет собой грубую модель частоты встречаемости катастрофических наводнений при довольно длительном периоде наблюдений. Распределение микроэлементов в образце почвы также может приближаться к пуассоновскому.

Пример . Вероятность нахождения микроэлемента в случайным образом выбранном образце почвы равна 0,001. Найти вероятность того, что среди 10 000 выбранных образцов почвы пять образцов будут содержать данный микроэлемент.

Решение. Пусть X – число образцов почвы, содержащих нужный микроэлемент. В данной задаче можно считать, что случайная величина X имеет распределение Пуассона с $n = 10\,000$, $p = 0{,}001$. Тогда $a = 10\,000 \cdot 0{,}001 = 10$ и искомая вероятность равна

$$P(X = 5) = \frac{10^5}{5!} e^{-10} \approx 0{,}0378.$$

Равномерное распределение

Непрерывная случайная величина X называется *равномерно распределенной на отрезке* $[a, b]$, если ее плотность распределения вероятностей постоянна на данном отрезке и равна нулю вне этого отрезка:

$$p(x) = \begin{cases} c, & \text{если } a \le x \le b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases}$$

Поскольку в силу свойства плотности распределения

$$\int_{-\infty}^{+\infty} p(x)\,dx = \int_a^b c\,dx = c(b - a) = 1,$$

то $c = \frac{1}{b - a}$. Следовательно,

$$p(x) = \begin{cases} \frac{1}{b - a}, & \text{если } a \le x \le b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases}$$

Вероятность попадания значений случайной величины X в интервал (α, β), принадлежащий отрезку $[a, b]$, пропорциональна длине этого интервала:

$$P(\alpha < X < \beta) = \int_\alpha^\beta p(x)\,dx = \int_\alpha^\beta \frac{dx}{b - a} = \frac{\beta - \alpha}{b - a}.$$

Непосредственно из определений находятся функция распределения вероятностей, математическое ожидание и дисперсия равномерно распределенной случайной величины X (вывод предлагается читателю в качестве упражнения):

$$F(x) = \begin{cases} 0 & \text{при } x \le a, \\ \frac{x - a}{b - a} & \text{при } a < x < b, \\ 1 & \text{при } x \ge b; \end{cases}$$

$$M(X)=\frac{a+b}{2},\quad D(X)=\frac{(b-a)^2}{12}.$$

Пример. Известно, что температура водоема в течение месяца является равномерно распределенной случайной величиной на отрезке $[6,10]$. Найти среднюю температуру водоема в данном месяце.

Решение. Пусть X – температура водоема в течение месяца. Тогда ее среднее значение равно $M(X)=\frac{6+10}{2}=8.$

Показательное распределение

Показательным (*экспоненциальным*) называют распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

$$p(x)=\begin{cases}0 & \text{при } x<0,\\ \lambda e^{-\lambda x} & \text{при } x\geq 0,\end{cases}$$

где λ – постоянная положительная величина.

Найдем функцию распределения вероятностей данной случайной величины. При $x<0$ имеем

$$F(x)=\int\limits_{-\infty}^{x}p(t)\,dt=\int\limits_{-\infty}^{x}0\,dt=0.$$

Если $x\geq 0,$ то

$$F(x)=\int\limits_{-\infty}^{x}p(t)\,dt=\int\limits_{-\infty}^{0}p(t)\,dt+\int\limits_{0}^{x}p(t)\,dt=\int\limits_{-\infty}^{0}0\,dt+\int\limits_{0}^{x}\lambda e^{-\lambda t}\,dt=$$

$$=-e^{-\lambda t}\Big|_{0}^{x}=-(e^{-\lambda x}-e^{0})=1-e^{-\lambda x}.$$

Следовательно, функция распределения имеет вид

$$F(x)=\begin{cases}0 & \text{при } x<0,\\ 1-e^{-\lambda x} & \text{при } x\geq 0.\end{cases}$$

С помощью метода интегрирования по частям можно показать, что

$$M(X)=\frac{1}{\lambda},\quad D(X)=\frac{1}{\lambda^2}.$$

Пример. Продолжительность жизни (в днях) растений данного вида в определенной среде представляет собой непрерывную случайную величину X, имеющую показательное распределение с параметром $\lambda=\frac{1}{140}$. Определить, какая доля растений данного вида погибает за период 100 дней.

Решение. Находим

$$P(0\leq X\leq 100)=\int\limits_{0}^{100}\frac{1}{140}e^{-\frac{x}{140}}dx=-e^{-\frac{x}{140}}\Bigg|_{0}^{100}=1-e^{-\frac{5}{7}}\approx 0{,}51.$$

Нормальное распределение

Нормальным распределением или *распределением Гаусса*, называют распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (a,\sigma\in \boldsymbol{R},\ \sigma>0).$$

Постоянные a и σ называют параметрами нормального распределения. Параметр a совпадает с математическим ожиданием случайной величины X: $a=M(X)$, а параметр σ является средним квадратическим отклонением случайной величины X: $\sigma=\sigma(X)=\sqrt{D(X)}$

График функции $p(x)$ называют *нормальной кривой* или *кривой Гаусса*. Он симметричен относительно прямой $x=a$, имеет асимптоту – ось Ox и схематически изображен на рисунке 19.

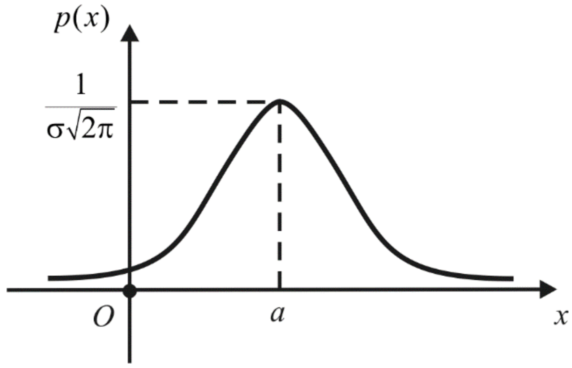


Рис.19

Нормальное распределение с параметрами $a=0$ и $\sigma=1$ называется *нормированным* или *стандартным*.

Вероятность того, что нормально распределенная случайная величина X примет значение, принадлежащее интервалу (α,β), равна

$$P(\alpha<X<\beta)=\frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right), \qquad (16)$$

где $\Phi(x)$ – *функция Лапласа*, определяемая формулой

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt.$$

Функция Лапласа является нечетной:

$$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{-x}e^{-\frac{t^2}{2}}dt=-\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{z^2}{2}}dz=-\Phi(x) \quad (t=-z,\, dt=-dz).$$

Для значений функции Лапласа составлены специальные таблицы, которые приводятся в приложениях к учебникам по теории вероятностей (см., например, [4, 6]).

Частным случаем формулы (16) является формула для определения вероятности того, что отклонение нормально распределенной случайной величины X от ее математического ожидания по абсолютной величине меньше заданного положительного числа δ:

$$P(|X-a|<\delta)=P(a-\delta<X<a+\delta)=\Phi\left(\frac{\delta}{\sigma}\right)-\Phi\left(-\frac{\delta}{\sigma}\right)=2\Phi\left(\frac{\delta}{\sigma}\right)$$

(здесь использовано свойство нечетности функции $\Phi(x)$).

Пример. Вес найденного образца породы подчиняется нормальному закону распределения с параметрами $a=400$ г и $\sigma=25$ г. Найти вероятность того, что вес образца породы будет заключен в пределах от 325 до 450 г.

Решение. Пусть случайная величина X – вес образца породы. Пользуясь таблицей значений функции Лапласа, находим

$$P(325<X<450)=\Phi\left(\frac{450-400}{25}\right)-\Phi\left(\frac{325-400}{25}\right)=\Phi(2)-\Phi(-3)=$$
$$=\Phi(2)+\Phi(3)=0,4772+0,4987=0,9759.$$

1.3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.3.1. Выборочный метод. Статистическое распределение

Математическая статистика

Еще в древнем Риме и Китае до нашей эры статистике уделялось большое внимание. Статистика определялась как один из разделов науки об управлении государством, сборе, классификации и обсуждении сведений о состоянии общества и государства.

Элементы статистики математической можно найти уже в трудах создателей теории вероятностей Я. Бернулли, П. Лапласа, С. Пуассона, К. Гаусса, Ж. Бюффона. В России методы математической статистики в применении к демографии и страховому делу развивал на основе теории вероятностей В. Я. Буняковский. Большое значение для всего дальнейшего развития математической статистики имели работы русской классической школы теории вероятностей второй половины XIX – начала XX веков, основателем которой стал П. Л. Чебышёв и его ученики А. А. Марков и А.М. Ляпунов. В начале XX века заметной в статистике была лондонская школа под руководством К. Пирсона прежде всего с точки зрения применения статистики в биологии. Этой школой велась интенсивная работа по составлению таблиц функций, необходимых для применения статистических методов. Важные результаты в теории математической статистики были получены У. Госсетом (Student), Р. Фишером, Э. Пирсоном, Е. Нейманом, В. И. Романовским, А. Н. Колмогоровым, Е. Е. Слуцким, Н. В. Смирновым, Ю. В. Линником и др.

В XXI веке основные типы статистических расчетов выполняются с помощью программного обеспечения. Но для грамотного применения компьютерных программ и встроенных функций необходимо понимание математической сути задачи.

Предмет и задачи математической статистики

Статистические данные – это числовые характеристики того или иного признака исследуемого процесса или явления.

Метод исследования, опирающийся на рассмотрение статистических данных о различных совокупностях однородных объектов, называется ***статистическим***.

При обработке результатов эксперимента статистическими методами основные понятия и положения теории вероятностей выступают как некоторые математические модели реальных явлений.

Статистический метод применяется в самых различных областях знания: социально-экономическая статистика, таможенная статистика, физическая статистика, звездная статистика, биологическая статистика и т. д. Общие черты статистических методов в различных областях знания составляют формальную математическую сторону статистических методов исследования, уже безразличную к специфической природе изучаемых объектов.

Математическая статистика – раздел математики, который изучает методы обработки результатов, а также методы планирования и применения статистических исследований. Предметом математической статистики являются статистические методы.

Основные задачи математической статистики состоят в разработке методов:

1) организации и планирования статистических наблюдений;
2) сбора статистических данных;
3) группировки и сокращения статистических данных с целью сведения большого числа таких данных к небольшому числу параметров, которые в сжатом виде характеризуют всю исследуемую совокупность;
4) анализа статистических данных;
5) принятия решений на основе анализа;
6) прогнозирования случайных явлений.

Выборочный метод. Статистическое распределение и его числовые характеристики

Одним из основных методов статистического исследования является *выборочный метод*.

Введем основные понятия выборочного метода.

Определение. *Статистической совокупностью* (в дальнейшем – *совокупность*) называется множество однородных объектов, объединенных по некоторому общему отличительному признаку.

Сущность выборочного метода заключается в том, что обследованию подвергаются не все объекты совокупности, а только их часть, случайно выбранная из этой совокупности; выводы, полученные при изучении этой части, распространяются на всю совокупность объектов.

Определение. ***Генеральной совокупностью*** **называется совокупность всех однородных объектов, подлежащих изучению по определенному признаку при данном комплексе условий *S***.

Например, перепись населения дает полную информацию о ряде признаков граждан данного государства. Результаты переписи населения можно считать генеральной совокупностью.

Определение. ***Выборочной совокупностью*** **или** ***выборкой*** **называется конечная совокупность объектов, случайным образом извлеченная из генеральной совокупности при постоянном комплексе условий *S*.**

Определение. ***Объемом*** **совокупности называется количество ее объектов. Объем генеральной совокупности обычно обозначается буквой *N*, а объем выборочной совокупности – буквой *n*.**

Выборка называется *повторной*, если отобранный объект перед отбором следующего объекта возвращается в генеральную совокупность.

Выборка называется *бесповторной*, если отобранный объект перед отбором следующего объекта не возвращается в генеральную совокупность.

Замечание. Если объем генеральной совокупности достаточно велик по сравнению с объемом выборки, то различия между повторной и бесповторной выборками можно считать незначительными.

Поставим в соответствие генеральной совокупности (изучаемому количественному признаку) случайную величину X. Тогда выборка понимается следующим образом. Случайная выборка – это результат n последовательных и независимых наблюдений над случайной величиной X. Пусть случайная величина X наблюдается в эксперименте. Этот эксперимент повторяется n раз в предположении, что условия проведения эксперимента и распределение X не изменяются от эксперимента к эксперименту. Математической моделью этого нового составного эксперимента является n-мерная случайная величина $\left(X_1, X_2, \dots, X_n\right)$, где X_i – случайная величина, соответствующая i-му эксперименту. Логично предположить, что X_i – независимые в совокупности случайные величины, каждая из которых имеет ту же функцию распределения $F\left(x\right)$, что и случайная величина X, соответствующая генеральной совокупности. Эта формулировка является основополагающей гипотезой математической статистики.

Статистическое распределение. Статистическая (эмпирическая) функция распределения

Пусть из генеральной совокупности объема N извлечена выборка объема n. Расположим все элементы выборки в неубывающем порядке и пронумеруем их:

$$x_1, x_2, \dots, x_n \quad (x_1 \le x_2 \le \dots \le x_n) . \tag{17}$$

Члены упорядоченной выборки (17) называются *вариантами*, а сама упорядоченная выборка (17) называется *дискретным вариационным рядом*.

Определение. **Число m_i наблюдений варианты x_i в дискретном вариационном ряде называется *абсолютной частотой* (*частотой*) этой варианты.**

Если значение x_1 встречается m_1 раз, значение x_2 встречается m_2 раз, ... , значение x_k – m_k раз, и $x_1, x_2, \dots, x_k$ – это все попарно различные варианты вариационного ряда, упорядоченные по возрастанию, то табл. 1 представляет *статистический ряд* выборки.

Таблица 1

Варианты x_i	x_1	x_2	...	x_k
Частоты m_i	m_1	m_2	...	m_k

В этой таблице сумма всех частот равна объему выборки:

$$\sum_{i=1}^{k} m_i = n. \tag{18}$$

Определение. ***Относительной частотой* или *частостью* варианты x_i называется отношение m_i к объему выборки, которое обозначается w_i:**

$$w_i = \frac{m_i}{n}. \tag{19}$$

Определение. ***Статистическим распределением выборки* называется соответствие между всеми различными вариантами и их относительными частотами (частостями).**

Иллюстрацией такого распределения является таблица, аналогичная табл. 1, где во второй строке указываются относительные частоты всех различных вариант, упорядоченных по возрастанию. Сумма относительных частот во второй строке этой таблицы равна единице:

$$\sum_{i=1}^{k} w_i = 1.$$

Определение. ***Статистической (эмпирической) функцией распределения* случайной величины X называют функцию $F^*(x)$, определяющую для каждого значения $x \in \boldsymbol{R}$ относительную частоту события $\{X < x\}$:**

$$F^*(x) = \sum_{i\,:\,x_i < x} w_i, \quad x \in \boldsymbol{R}. \tag{20}$$

$F^*(x)$ называют еще *функцией распределения выборки*.

Для того, чтобы найти значение статистической функции распределения при данном действительном значении x, достаточно подсчитать число опытов, в которых случайная величина X приняла значение, меньшее чем x, и разделить на общее число n произведенных опытов. Из определения следует, что статистическая функция распределения обладает следующими свойствами.

Свойства статистической функции распределения

1. $0 \le F^*(x) \le 1$, $x \in \boldsymbol{R}$.

2. $F^*(x)$ – не убывающая функция на множестве $\boldsymbol{R}$.

3. Если $x_{\min}$ и $x_{\max}$ – соответственно наименьшее и наибольшее значения наблюдаемых вариант выборки, то $F^*(x) = 0$, $x \le x_{\min}$ и $F^*(x) = 1$, $x > x_{\max}$.

При достаточно большом объеме n выборки значения функций $F(x)$ и $F^*(x)$ при одинаковых значениях аргумента мало отличаются одно от другого, поэтому статистическую функцию $F^*(x)$ берут в качестве оценки функции $F(x)$. Теоретической основой для такого оценивания служит теорема, суть которой состоит в следующем. Пусть $F^*(x)$ – эмпирическая функция распределения, построенная по выборке объема n из генеральной совокупности, соответствующей случайной величине X с функцией распределения $F(x) = P(X < x)$, $x \in \boldsymbol{R}$. Тогда для любого $x \in (-\infty, +\infty)$ и любого $\varepsilon > 0$ справедливо предельное равенство

$$\lim_{n\to\infty}\left(\left|F^*(x) - F(x)\right| < \varepsilon\right) = 1.$$

Пример. Дан статистический ряд выборки.

Варианты x_i	1	2	4
Частоты m_i	12	18	30

Построить эмпирическую функцию распределения и ее график.

Решение. Найдем объем выборки: n = 12 + 18 + 30 = 60. Наименьшее значение варианты равно 1, следовательно, $F^*(x) = 0$ при $x \le 1$.

Значение $x_1 = 1 < 2$, наблюдалось 12 раз. Поэтому согласно формуле (20)

$$F^*(x) = \frac{12}{60} = 0,2 \qquad \text{при } x \in (1;\ 2].$$

Значения $x < 4$, а именно, $x_1 = 1$, $x_2 = 2$, наблюдались 12 + 18 = 30 раз. Поэтому

$$F^*(x) = \frac{30}{60} = 0,5 \quad \text{при} \quad x \in (2;\ 4].$$

Так как $x = 4$ – наибольшее значение варианты, то при $x > 4$ значения $F^*(x) = 1$. Таким образом, эмпирическая функция распределения имеет график, изображенный на рис. 20.

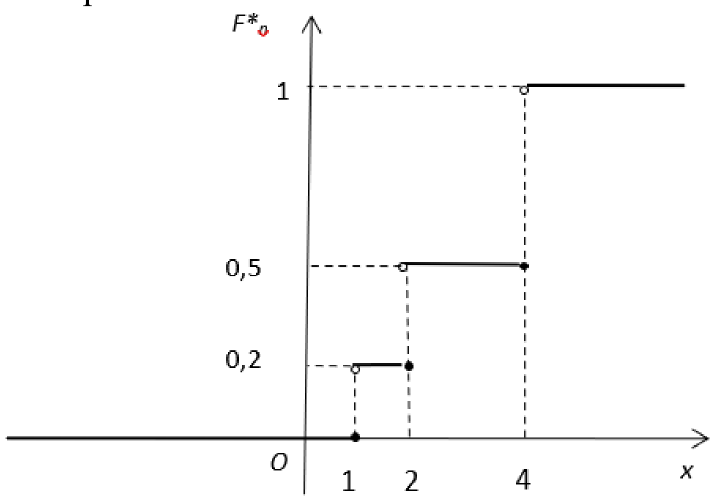


Рис. 20

Эта функция задается формулой :

$$F^*(x) = \begin{cases} 0 & \text{при } x \le 1; \\ 0,2 & \text{при } x \in (1;\ 2]; \\ 0,5 & \text{при } x \in (2;\ 4]; \\ 1 & \text{при } x > 4. \end{cases}$$

Для наглядного изображения статистического распределения строят полигон частот или относительных частот выборки.

Определение. ***Полигоном абсолютных частот*** **(*полигоном частот*) называется ломаная линия, последовательно соединяющая точки с координатами** $(x_i, m_i), i = \overline{1,k},$ **(рис. 21), а *полигоном относительных частот* – ломаная, соединяющая точки с координатами** $\left(x_i, \dfrac{m_i}{n}\right), i = \overline{1,k}$.

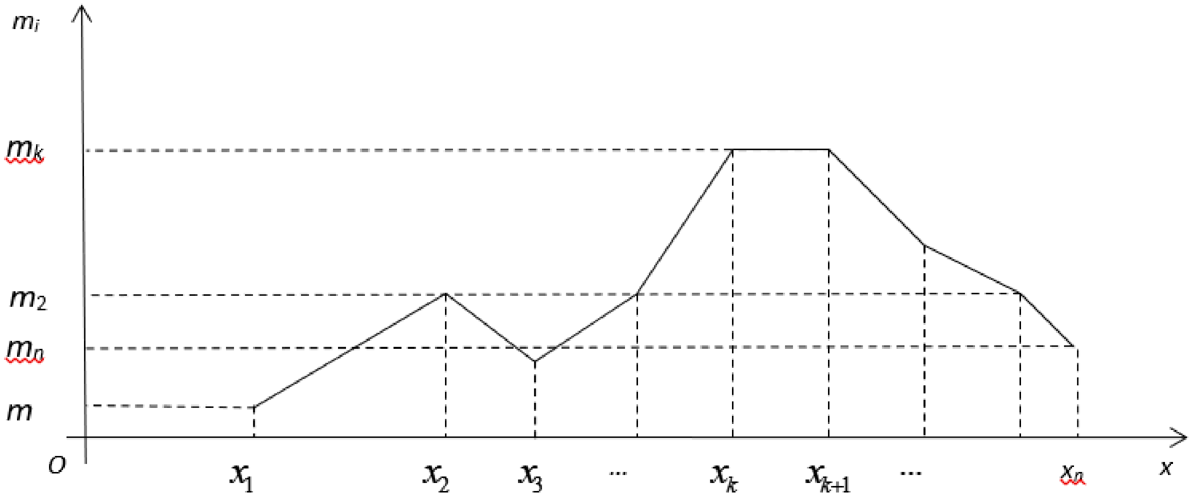


Рис. 21

В случае *интервального статистического ряда* делят весь диапазон наблюдаемых значений x_i на k интервалов и подсчитывают количество m_i значений вариант, попавших в каждый i-тый интервал, $i = \overline{1,k}$. Далее строят таблицу, в первой строке которой приводят интервалы в порядке их расположения вдоль оси абсцисс, а во второй – соответствующие им частоты или относительные частоты (таблица 2).

Таблица 2

Интервалы	$[x_1, x_2)$	$[x_2, x_3)$	. . .	$[x_k, x_{k+1}]$
Частоты m_i	m_1	m_2	. . .	m_k

Сумма всех чисел во второй строке таблицы равна объему выборки n.

В дальнейшем будем полагать длины интервалов равными

$$h = x_{i+1} - x_i, \quad i = \overline{1,\ k} \ .$$

При обработке выборок больших объемов весь диапазон значений выборки делится на k интервалов, число которых в случае нормально распределенной случайной величины X можно вычислять по формуле Стерджесса

$$k = 1 + 3{,}322 \cdot \lg n$$

с округлением до ближайшего целого (часто здесь округляют в большую сторону).

Для определения эмпирической функции распределения $F^*(x)$ интервального вариационного ряда вводятся *вспомогательные варианты*

$$x_i^* = \frac{x_{i+1} + x_i}{2}, \quad i = \overline{1,k} \ , \tag{21}$$

которые представляют свойства всего i-го интервала. Тогда функция $F^*(x)$ определяется для каждого действительного значения x равенством

$$F^*(x) = \sum_{i:\, x_i^* < x} w_i \ , \ x \in \boldsymbol{R} . \tag{22}$$

Для графического представления интервальных статистических рядов строятся полигон и гистограмма частот или относительных частот.

Определение. ***Гистограммой абсолютных частот*** **(*частот*) интервального статистического ряда называется ступенчатая фигура, состоящая из прямоугольников с основанием длины h и высотой $\frac{m_i}{h}$, построенных на интервалах группировки.**

Определение. ***Гистограммой относительных частот*** **(*частостей*) интервального статистического ряда называется ступенчатая фигура, состоящая из прямоугольников с основанием длины h и высотой $\frac{w_i}{h}$, построенных на интервалах группировки** (рис. 22).

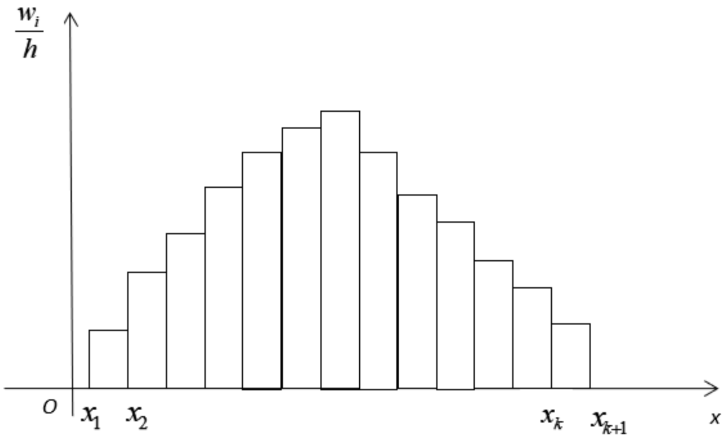


Рис. 22

Из способа построения следует, что площадь гистограммы относительных частот равна единице, так как $\sum_{i=1}^{k} w_i = 1$.

Пример. Проводились опыты по взвешиванию определенного вещества, входящего в состав одной таблетки лекарства сильного действия (провели 122 независимых однотипных эксперимента). Результаты экспериментов показали, что масса взвешиваемого вещества в одной таблетке − 1 мг и соответствующее номеру опыта количество микрограммов (мкг), указанное в выборке:

0,20	0,30	0,40	0,47	0,74	0,44	0,50	0,49	0,44	0,46	0,68
0,46	0,42	0,41	0,50	0,32	0,80	0,46	0,37	0,45	0,39	0,55
0,38	0,50	0,28	0,52	0,49	0,54	0,53	0,47	0,54	0,82	0,44
0,57	0,50	0,55	0,38	0,64	0,32	0,30	0,50	0,43	0,49	0,88
0,50	0,39	0,62	0,57	0,45	0,49	0,57	0,56	0,60	0,48	0,55
0,47	0,63	0,27	0,47	0,60	0,50	0,52	0,25	0,68	0,54	0,53
0,58	0,53	0,65	0,58	0,38	0,45	0,42	0,60	0,43	0,47	0,29
0,39	0,43	0,46	0,61	0,60	0,51	0,65	0,59	0,53	0,50	0,28
0,38	0,70	0,67	0,42	0,48	0,37	0,12	0,47	0,58	0,66	0,53
0,60	0,50	0,38	0,59	0,54	0,62	0,67	0,72	0,45	0,27	0,29
0,48	0,39	0,62	0,44	0,76	0,46	0,51	0,38	0,47	0,55	0,43 0,48

Построить интервальный статистический ряд, гистограмму относительных частот, эмпирическую функцию распределения для этой выборки.

Решение. Объем данной выборки $n = 122$. Построим интервальный статистический ряд. Для определения количества интервалов используем формулу Стерджесса:

$$k = 1 + 3{,}322 \cdot \lg n = 1 + 3{,}322 \cdot \lg 122 \approx 8.$$

Будем считать, что все интервалы имеют одинаковую длину. Длину одного интервала вычислим по формуле

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{0{,}88 - 0{,}12}{8} \approx 0{,}1.$$

Интервальный статистический ряд имеет вид:

Интервалы	m_i
[0,1; 0,2)	1
[0,2; 0,3)	8
[0,3; 0,4)	16
[0,4; 0,5)	37
[0,5; 0,6)	35
[0,6; 0,7)	18
[0,7; 0,8)	4
[0,8; 0,9)	3

Здесь числа m_i $(i = \overline{1, 8})$ показывают, сколько выборочных значений принадлежат соответствующему интервалу.

Дополним предыдущую таблицу еще тремя столбцами:

Интервалы	x_i^*	m_i	$w_i = m_i / n$	$m_i / (nh)$
[0,1; 0,2)	0,15	1	1/122	$\approx 0{,}08$
[0,2; 0,3)	0,25	8	8/122	$\approx 0{,}66$
[0,3; 0,4)	0,35	16	16/122	$\approx 1{,}31$
[0,4; 0,5)	0,45	37	37/122	$\approx 3{,}03$
[0,5; 0,6)	0,55	35	35/122	$\approx 2{,}87$
[0,6; 0,7)	0,65	18	18/122	$\approx 1{,}48$
[0,7; 0,8)	0,75	4	4/122	$\approx 0{,}33$
[0,8; 0,9)	0,85	3	3/122	$\approx 0{,}25$
		$\sum_{i=1}^{8} m_i = 122$	$\sum_{i=1}^{8} w_i = 1$	

Построим с помощью этой таблицы гистограмму относительных частот интервального статистического ряда (рис. 23).

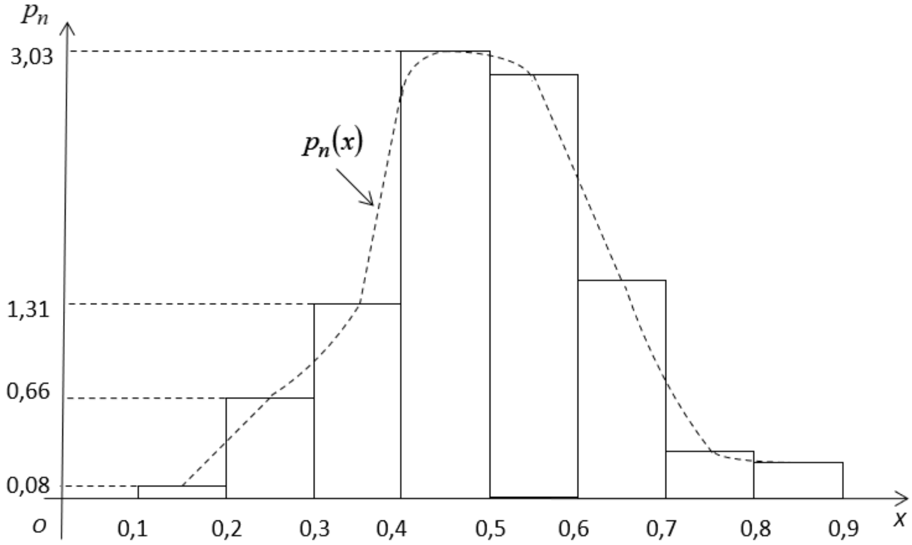


Рис. 23

График функции $p_n(x)$, изображенный на рис. 23 плавной штрихованной линией, проходит через точки $\left(x_i^*, \dfrac{m_i}{nh}\right)$, $i = \overline{1, 8}$. Своим видом он дает приближенное представление о графике функции плотности распределения вероятностей случайной величины *X*, соответствующей генеральной совокупности. В данном случае можно высказать гипотезу о нормальном законе распределения случайной величины *X*.

Эмпирическая функция распределения имеет вид:

$$F^*(x)=\begin{cases} 0, & x \le 0,15, \\ 1/122, & 0,15 < x \le 0,25, \\ 9/122, & 0,25 < x \le 0,35, \\ 25/122, & 0,35 < x \le 0,45, \\ 31/61, & 0,45 < x \le 0,55, \\ 97/122, & 0,55 < x \le 0,65, \\ 115/122, & 0,65 < x \le 0,75, \\ 119/122, & 0,75 < x \le 0,85, \\ 1, & x > 0,85. \end{cases}$$

1.3.2. Числовые характеристики статистического распределения

Конкретная выборка $(x_1, x_2, \ldots, x_n)$ с одной стороны понимается как реализация многомерной случайной величины $(X_1, X_2, \ldots, X_n)$ с независимыми в совокупности компонентами, каждая из которых имеет одну и ту же функцию распределения $F(x)$; с другой стороны саму конкретную выборку можно рассматривать как конечную генеральную совокупность с равновероятными исходами $\left(P(X = x_i) = \frac{1}{n}\right)$ и определять числовые характеристики этой выборочной случайной величины *X*. Определения таких выборочных характеристик аналогичны соответствующим определениям характеристик в теории вероятностей при условии, что в формулировках термин «вероятность» заменяется на «относительная частота».

Размахом выборки называется разность между наибольшим и наименьшим значениями вариант

$$R = x_{\max} - x_{\min} \ . \tag{23}$$

Пусть задан дискретный статистический ряд таблицей 13.1 (см. 13.1), для которого $\sum_{i=1}^{k} m_i = n$.

Выборочное математическое ожидание или *выборочное среднее значение* (*выборочное среднее*) определяется по формуле:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} m_i x_i \,. \tag{24}$$

В случае, если вариационный ряд не группированный, либо все варианты различны, то

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \tag{25}$$

Выборочный начальный момент s-го порядка вычисляется по формуле:

$$\alpha_s^* = \overline{X^s} = \frac{1}{n}\sum_{i=1}^{k} m_i x_i^s \ . \tag{26}$$

В случае, если вариационный ряд не группированный, либо все варианты различны, то

$$\alpha_s^* = \overline{X^s} = \frac{1}{n}\sum_{i=1}^{n} x_i^s \quad . \tag{27}$$

Выборочная дисперсия – неотрицательная числовая характеристика статистического распределения, определяемая формулой:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} m_i \left(x_i - \bar{X} \right)^2 \quad . \tag{28}$$

В случае, если вариационный ряд не группированный, либо все варианты различны, то

$$D_B = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \bar{X} \right)^2 \quad . \tag{29}$$

Выборочным средним квадратическим (*квадратичным*) *отклонением* называется квадратный корень из выборочной дисперсии

$$\sigma_B = \sqrt{D_B} \, . \tag{30}$$

Эта числовая характеристика является мерой вариации *X*, а также показателем однородности статистической совокупности.

Выборочным центральным моментом s-го порядка называется числовая характеристика статистического распределения, определяемая формулой

$$\mu_s^* = \frac{1}{n}\sum_{i=1}^{k} m_i \left(x_i - \bar{X} \right)^s . \tag{31}$$

В случае, если вариационный ряд не группированный, либо все варианты различны, то

$$\mu_s^* = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \bar{X} \right)^s \quad . \tag{32}$$

Выборочной модой $d_X^* = Mo$ называется варианта, имеющая наибольшую частоту (унимодальный ряд).

Если несколько соседних элементов дискретного вариационного ряда имеют наибольшую частоту, то выборочная мода равна среднему арифметическому этих значений. Если несколько несмежных вариант имеют наибольшие частоты, то говорят, что вариационный ряд является полимодальным. Если все варианты имеют одинаковые частоты, то ряд моды не имеет.

Выборочной медианой вариационного ряда называется число $h_X^* = Me$, которое делит ряд на две части, содержащие равное количество элементов:

если объем выборки $n = 2l + 1$ – нечетное число, то $Me = x_{l+1}$;

если объем выборки $n = 2l$ – четное число, то $Me = \frac{1}{2}\left(x_l + x_{l+1} \right)$.

Для определения выборочных моментов интервального статистического ряда, вычисляют средние арифметические значения концов элементарных интервалов

$$x_i^* = \frac{x_i + x_{i+1}}{2}, \quad i = \overline{1,k} \, .$$

Выборочное математическое ожидание или выборочное среднее интервального статистического ряда определяется по формуле

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} m_i x_i^* . \tag{33}$$

Выборочная дисперсия для интервального статистического ряда определяется по формуле

$$D_B = \frac{1}{n}\sum_{i=1}^{k} m_i \left(x_i^* - \overline{X} \right)^2 . \tag{34}$$

Аналогично определяются начальные и центральные моменты, т. е. в формулах (26), (31) значения вариант x_i следует заменить на x_i^*.

Для интервального статистического ряда выборочная мода вычисляется следующим образом:

$$d_X^* = Mo = x_i + h \cdot \frac{m_i - m_{i-1}}{\left(m_i - m_{i-1} \right) + \left(m_i - m_{i+1} \right)} , \tag{35}$$

где h – длина элементарных интервалов. Сначала определяется модальный интервал $\left[x_i \, ; \, x_{i+1} \right)$, т. е. интервал с наибольшей частотой m_i, $i = \overline{2, k-1}$, затем определяются m_{i-1}, m_{i+1} – частоты предмодального и послемодального интервалов соответственно.

В случае интервального статистического ряда объема *n* выборочная медиана вычисляется по формуле:

$$h_X^* = Me = x_\ell + h \cdot \frac{\frac{n}{2} - \sum_{i=1}^{\ell-1} m_i}{m_\ell} , \tag{36}$$

где h – длина медианного интервала, m_ℓ – частота медианного интервала. Для определения медианного интервала $\left[x_\ell \, ; \, x_{\ell+1} \right)$ подсчитываются накопленные частоты и сравниваются с $\frac{n}{2}$. Медианный интервал – интервал с наименьшим номером, накопленная частота которого не меньше $\frac{n}{2}$:

$$\sum_{i=1}^{l-1} m_i < \frac{n}{2} \quad \text{и} \quad \sum_{i=1}^{\ell} m_i \ge \frac{n}{2} .$$

Замечание. Выборочную медиану можно вычислять также по формуле

$$h_X^* = Me = x_{l+1} + h \cdot \frac{\sum_{i=1}^{l-1} m_i - \frac{n}{2}}{m_\ell} .$$

Важную роль при исследовании статистических совокупностей играют *асимметрия* и *эксцесс*, которые характеризуют форму статистического распределения. Выборочный коэффициент асимметрии вычисляется по формуле:

$$a_X^* = \frac{\mu_3^*}{\sqrt{D_B^3}} \;. \qquad (37)$$

Для симметричного статистического распределения $a_X^* \approx 0$. При $a_X^* > 0$ вершина кривой эмпирического распределения сдвинута относительно ординаты среднего выборочного влево. При $a_X^* < 0$ вершина кривой сдвинута вправо. Выборочный коэффициент эксцесса вычисляется по формуле

$$e_X^* = \frac{\mu_4^*}{D_B^2} - 3 \,. \qquad (38)$$

Эта характеристика применяется в основном для исследования симметричных распределений. Если $e_X^* \approx 0$, то распределение близко к нормальному. Если $e_X^* > 0$, то статистическое распределение называют островершинным. Если $e_X^* < 0$, то распределение называют пологим. Заметим, что понятие эксцесса используется только для унимодальных распределений.

Относительными мерами вариации являются *коэффициенты вариации* и *осцилляции*. Для сравнения средних квадратических отклонений различных вариаций вариационных рядов вычисляется процентное отношение $\sigma_В$ к среднему выборочному $\overline{X}$

$$V = \frac{\sigma_B}{\overline{X}} \cdot 100 \;\% \;, \qquad (39)$$

где V называется *коэффициентом вариации*. *Коэффициент осцилляции* определяется по формуле

$$K_R = \frac{R}{\overline{X}} \cdot 100 \;\% \;, \qquad (40)$$

где R – размах выборки.

В заключение отметим, что все приведенные выше формулы для дискретного статистического ряда используются и для интервального статистического ряда, если по нему построить дискретный статистический ряд, заменив частичные интервалы их серединами.

Пример. При статистическом исследовании доходности предприятий региона получили следующий интервальный статистический ряд:

Доходность	[87,93)	[93,99)	[99,105)	[105,111)	[111,117)	[117,123]
w_i	4/30	6/30	6/30	4/30	7/30	3/30

На основе этих данных построить гистограмму относительных частот.

<u>Решение.</u> В данном случае длина частичного интервала $h = 6$. Поэтому высоты $h_i = \frac{w_i}{h}$ прямоугольников равны: $h_1 = h_4 = \frac{4}{6 \cdot 30} \approx 0{,}022$, $h_2 = h_3 = \frac{6}{6 \cdot 30} \approx 0{,}033$, $h_5 = \frac{7}{6 \cdot 30} \approx 0{,}039$, $h_6 = \frac{3}{6 \cdot 30} \approx 0{,}017$.

Гистограмма относительных частот изображена на рис. 24.

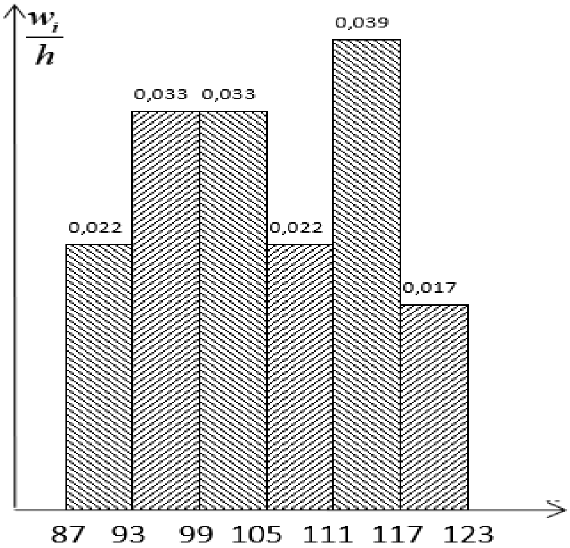


Рис. 24

Пример. За отчетный период на пункте «Дубровка» выборка таможенных платежей физических лиц представлена в виде таблицы:

75	191	76	138	113	157	129	100	182	144
100	155	200	111	115	95	120	180	147	146
145	154	100	117	96	137	168	189	204	149
145	200	111	83	139	89	172	125	140	183
122	124	110	123	138	125	129	70	142	169
210	150	136	100	102	107	154	108	135	150
90	173	184	131	154	160	187	151	159	137

Требуется:

1. определить объем выборки;
2. найти размах выборки;
3. разбив выборку на $k = 7$ интервалов, построить интервальный статистический ряд;
4. построить полигон абсолютных частот;
5. построить гистограмму относительных частот;
6. построить эмпирическую функцию распределения;
7. найти выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение;
8. найти выборочные моду и медиану;
9. вычислить выборочные коэффициенты асимметрии и эксцесса;
10. вычислить коэффициенты вариации и осцилляции.

Решение. Объем выборки $n = 70$. Размах выборки согласно (23): $R = 210 - 70 = 140$, $h = \frac{R}{k} = \frac{140}{7} = 20$. Интервальный статистический ряд для $k = 7$ имеет вид:

Таблица 3

№ интервала, i	Интервал, $x_i - x_{i+1}$	Частота, m_i
1	70 – 90	5
2	90 – 110	10
3	110 – 130	14
4	130 – 150	16
5	150 – 170	12
6	170 – 190	8
7	190 – 210	5

Для построения полигона частот и гистограммы относительных частот применим табличный процессор Microsoft Excel. Результаты изображены на рис. 25 и 26 соответственно.

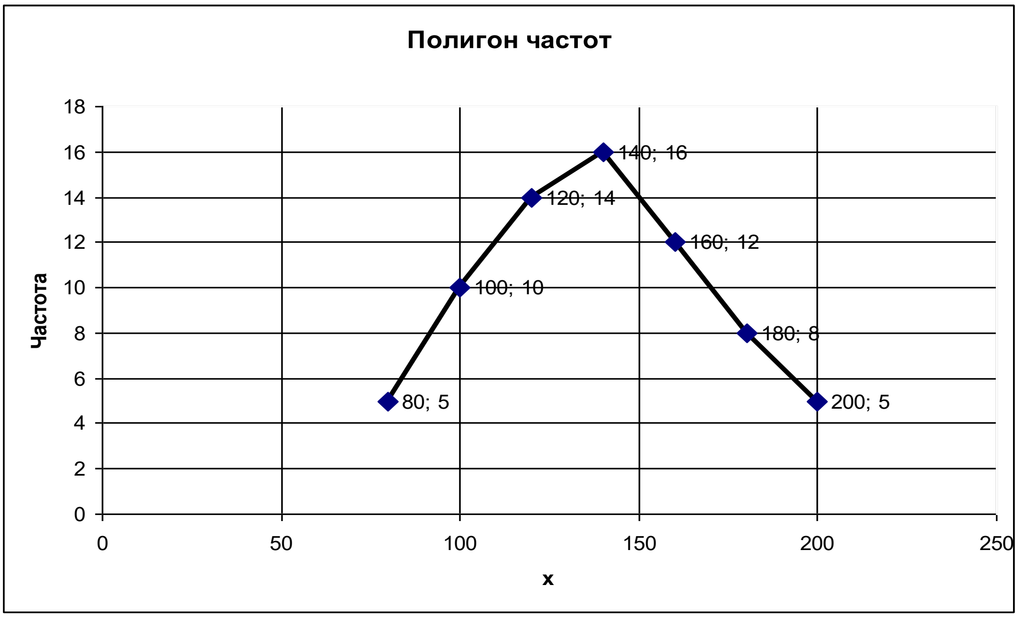


Рис. 25

Гистограмма относительных частот имеет вид

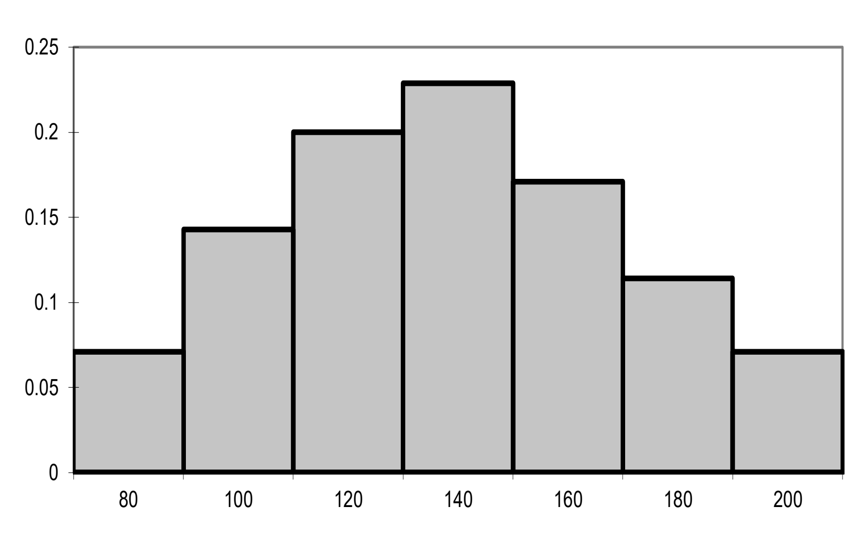


Рис. 26

Определим значения вспомогательных вариант $x_i^* = \dfrac{x_i + x_{i+1}}{2}$, $i = \overline{1,7}$ (середин соответствующих интервалов), накопленных частот и занесем результаты в следующую таблицу:

Таблица 4

№ интервала, i	x_i^*	$\sum m_i$
1	80	5
2	100	15
3	120	29
4	140	45
5	160	57
6	180	65
7	200	70

Эмпирическая функция распределения имеет вид:

$$F^*(x) = \begin{cases} 0, & x \le 80, \\ 1/14, & 80 < x \le 100, \\ 3/14, & 100 < x \le 120, \\ 29/70, & 120 < x \le 140, \\ 9/14, & 140 < x \le 160, \\ 57/70, & 160 < x \le 180, \\ 13/14, & 180 < x \le 200, \\ 1, & x > 200. \end{cases}$$

Вычислим выборочное среднее значение $\overline{X}$ по формуле (33).

$$\overline{X} = \frac{1}{70} \cdot (5 \cdot 80 + 10 \cdot 100 + 14 \cdot 120 + 16 \cdot 140 + 12 \cdot 160 + 8 \cdot 180 + 5 \cdot 200) =$$

$$= \frac{9680}{70} = 138{,}286 \approx 138{,}3.$$

Найдем выборочную дисперсию

$$D_B = \frac{1}{70} \cdot \Big[5 \cdot (80 - 138{,}3)^2 + 10 \cdot (100 - 138{,}3)^2 + 14 \cdot (120 - 138{,}3)^2 +$$

$$+ 16 \cdot (140 - 138{,}3)^2 + 12 \cdot (160 - 138{,}3)^2 + 8 \cdot (180 - 138{,}3)^2 +$$

$$+ 5 \cdot (200 - 138{,}3)^2 \Big] = \frac{69604{,}02}{70} \approx 994{,}34.$$

Выборочное среднее квадратическое отклонение вычислим по формуле (30)

$$\sigma_{\text{В}} = \sqrt{993{,}34} \approx 31{,}53.$$

Для нахождения медианного интервала используем накопленные частоты $\left(\sum m_i\right)$ (см. табл. 4). Интервал с номером $\ell = 4$ будет медианным, так как имеет накопленную частоту $\sum_{i=1}^{4} m_i = 45 > \frac{n}{2} = 35$ и $\sum_{i=1}^{3} m_i = 29 < 35$, тогда по формуле (36)

$$h_X^* = x_\ell + h \cdot \frac{\frac{n}{2} - \sum_{i=1}^{\ell-1} m_i}{m_\ell} = 130 + 20 \cdot \frac{35-29}{16} = 137{,}5 .$$

Интервалу с номером $i = 4$ соответствует максимальная частота $m_4 = 16$. Такой интервал единственный, следовательно, распределение – унимодальное. Выборочную моду определим по формуле (35):

$$d_X^* = x_i + h \cdot \frac{m_i - m_{i-1}}{(m_i - m_{i-1}) + (m_i - m_{i+1})} = 130 + 20 \cdot \frac{16-14}{(16-14)+(16-12)} \approx 136{,}67 .$$

По формуле (37) находим выборочный коэффициент асимметрии

$$a_X^* = \frac{\mu_3^*}{\sqrt{D_B^3}} \approx 0{,}11 > 0 ,$$

следовательно, распределение выборки имеет отклонение влево.

Выборочный коэффициент эксцесса вычисляем по формуле (38)

$$e_X^* = \frac{\mu_4^*}{D_B^2} - 3 = 2{,}2997 - 3 \approx -0{,}7 < 0 ,$$

следовательно, распределение выборки – пологое.

Согласно формуле (39) вычислим коэффициент вариации:

$$V = \frac{\sigma_B}{\bar{X}} \cdot 100\,\% = \frac{31{,}53}{138{,}3} \cdot 100\% \approx 22{,}8\% \ .$$

Коэффициент осцилляции определим по формуле (40)

$$K_R = \frac{R}{\bar{X}} \cdot 100\,\% = \frac{140}{138{,}3} \cdot 100\% \approx 101{,}23\,\% .$$

Ответ. 1. $n = 70$, 2. $R = 140$, 3. Табл. 3, 4. Рис. 25, 5. Рис. 26, 6. $F^*(x)$ (см. вид в тексте решения), 7. $\bar{X} \approx 138{,}3,\ D_B \approx 994{,}34, \sigma_B \approx 31{,}53$, 8. $h_X^* = 137{,}5$, $d_X^* \approx 136{,}67$, 9. $a_X^* \approx 0{,}11,\ e_X^* \approx -0{,}7$, 10. $V \approx 22{,}8\%$, $K_R \approx 101{,}23\,\%$.

1.3.3. Статистическое оценивание

Точечное оценивание

Пусть изучается случайная величина X с законом распределения, зависящим от одного или нескольких параметров. Требуется по выборке $X_1, X_2, \dots, X_n$, полученной в результате наблюдений, оценить неизвестный параметр θ.

Определение. ***Статистической оценкой*** θ **(или просто –** ***оценкой*** θ**) параметра** θ **теоретического распределения называют его приближенное значение, зависящее от данных выбора.**

Понятно, что оценка θ представляет собой значение некоторой функции наблюдений над случайной величиной, т. е.

$$\theta = \theta(X_1, X_2, \cdots, X_n).$$

Функцию результатов наблюдений (т. е. функцию выборки) называют *статистикой*.

Можно сказать, что оценка θ параметра θ есть статистическая характеристика, которая в определенном смысле близка к истинному значению θ. Так, например, $F^*(x)$ есть оценка $F(x)$, гистограмма частостей – это оценка плотности распределения вероятностей $f(x)$ случайной величины *X*.

Оценка θ является случайной величиной, поскольку представляет собой функцию независимых в совокупности случайных величин $X_1, X_2, \cdots, X_n$. Если реализовать разные конкретные выборки, то, вообще говоря, функция может принять разные числовые значения.

Статистика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется *точечной оценкой* этого параметра. Таким образом, точечная оценка в случае конкретной выборки – это число, определяемое по выборке с помощью формулы, зависящей от выборочных данных.

Определение. **Статистическая оценка $\theta = \theta(X_1, X_2, \cdots, X_n)$ неизвестного параметра θ называется *несмещенной*, если**

$$M(\theta) = \theta \tag{41}$$

при любом объеме *n* выборки.

Определение. ***Смещенной* называют оценку, математическое ожидание которой не равно оцениваемому параметру, т. е. не выполняется равенство (41).**

Однако несмещенность не является достаточным условием хорошего приближения к истинному значению оцениваемого параметра. Если при этом возможные значения θ могут значительно отклоняться от среднего значения, то есть дисперсия θ велика, то значение, найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра. Следовательно, требуется наложить ограничения на дисперсию.

Определение. **Несмещенная оценка называется *эффективной*, если она при заданном объеме выборки *n* имеет наименьшую возможную дисперсию.**

При рассмотрении выборок большого объема обязательным требованием к статистическим оценкам является требование состоятельности.

Определение. ***Состоятельной* называется статистическая оценка $\theta = \theta(X_1, X_2, \cdots, X_n)$ неизвестного параметра θ, которая для любого $\varepsilon > 0$ удовлетворяет предельному равенству**

$$\lim_{n\to\infty} P\left(\left|\theta(X_1, X_2, \cdots, X_n) - \theta\right| < \varepsilon\right) = 1.$$

Теорема 10. Если оценка $\theta = \theta(X_1, X_2, \cdots, X_n)$ несмещенная и при $n\to\infty$ ее дисперсия стремится к 0, то θ будет состоятельной.

Пример. Доказать, что $\overline{X}$ представляет собой несмещенную оценку математического ожидания $M(X)$ случайной величины X.

Решение. Будем рассматривать выборочные значения как независимые в совокупности случайные величины $X_1, X_2, \ldots, X_n$ с одним и тем же законом распределения и с математическим ожиданием a, и, следовательно, выборочное среднее значение $\overline{X}$ – как случайную величину. Из свойств математического ожидания следует, что

$$M(\overline{X}) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n} M\left(X_1 + X_2 + \ldots + X_n\right).$$

Поскольку каждая из величин $X_1, X_2, \ldots, X_n$ имеет такое же распределение, что и генеральная совокупность с $M(X) = a$, то $M(\overline{X}) = \frac{1}{n} na = a$.

Что и требовалось доказать.

Выборочное среднее значение является состоятельной оценкой математического ожидания. Если предположить, что $X_1, X_2, \ldots, X_n$ имеют ограниченные дисперсии, то из теоремы Чебышева следует, что их среднее арифметическое, то есть $\overline{X}$, при неограниченном увеличении n стремится по вероятности к математическому ожиданию a каждой из величин, то есть к $M(X)$. Следовательно, выборочное среднее есть состоятельная оценка математического ожидания.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_B) = \frac{n-1}{n} D_\Gamma, \tag{42}$$

где D_Γ – истинное значение дисперсии генеральной совокупности.

В качестве несмещенной оценки дисперсии используют *исправленную дисперсию* s^2, вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}. \tag{43}$$

Такая оценка является несмещенной в силу свойств математического ожидания и формулы (42). Ей соответствует *исправленное среднее квадратическое отклонение*

$$s = \sqrt{s^2}. \tag{44}$$

Оценка θ некоторого неизвестного параметра θ теоретического распределения называется *асимптотически несмещенной*, если для случайной выборки $X_1, X_2, \ldots, X_n$ любого объема n она содержит систематическую ошибку ($M(\theta) \neq \theta$), но обладает свойством

$$\lim_{n \to \infty} M(\hat{\theta}) = \theta. \tag{45}$$

Методы нахождения точечных оценок

Далее рассмотрим два наиболее распространенных метода получения точечных оценок параметров распределения: *метод моментов* и *метод максимального правдоподобия*.

Метод моментов для нахождения точечных оценок неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов распределения соответствующим эмпирическим моментам, найденным по реализации выборки (по выборке).

Если распределение зависит от одного параметра θ, то для нахождения его оценки необходимо решить относительно θ уравнение

$$M(X) = \overline{X}.$$

Если распределение зависит от двух параметров θ_1 и θ_2, то для нахождения их оценок необходимо решить относительно θ_1 и θ_2 одну из систем уравнений:

$$\begin{cases} M(X) = \overline{X}, \\ D(X) = D_{\text{В}}, \end{cases} \quad \text{или} \quad \begin{cases} M(X) = \overline{X}, \\ M(X^2) = \overline{X^2}. \end{cases}$$

Если же нужно оценить l параметров $\theta_1, \theta_2, \cdots, \theta_l$, то нужно решить относительно $\theta_1, \theta_2, \cdots, \theta_l$ одну из систем уравнений:

$$\begin{cases} M(X) = \overline{X}, \\ D(X) = D_{\text{В}}, \\ \ldots \\ M(X - M(X))^l = \sum\limits_{i=1}^{k} (x_i - \overline{X})^l p_i^*, \end{cases} \quad \text{или} \quad \begin{cases} M(X) = \overline{X}, \\ M(X^2) = \overline{X^2}, \\ \ldots \\ M(X^l) = \sum\limits_{i=1}^{k} x_i^l p_i^*. \end{cases}$$

Пример. Методом моментов найти по реализации выборки $x_1, x_2, \cdots, x_n$ оценку параметра λ показательного распределения.

Решение. Поскольку нужно оценить один параметр $\theta = \lambda$, решим уравнение $M(X) = \overline{X}$. Математическое ожидание показательного распределения с параметром λ, как известно, равно $\frac{1}{\lambda}$, поэтому $\frac{1}{\lambda} = \overline{X}$.

Ответ. $\lambda = \dfrac{1}{\overline{X}} = \dfrac{n}{\sum\limits_{i=1}^{n} x_i}$.

Пример. Методом моментов найти по реализации выборки $x_1, x_2, \cdots, x_n$ оценки параметров a и b равномерного на отрезке $[a, b]$ распределения.

Решение. В данной задаче нужно оценить два параметра $\theta_1 = a$ и $\theta_2 = b$. Так как для случайной величины $X \sim R[a, b]$ верны равенства $M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$, то

$$\begin{cases} \dfrac{a+b}{2} = \overline{X}, \\ \dfrac{(b-a)^2}{12} = D_{\text{в}} \end{cases} \qquad \text{или} \qquad \begin{cases} a+b = 2\overline{X}, \\ b-a = 2\sqrt{3}\sigma_{\text{в}}, \end{cases}$$

таким образом

$$\begin{cases} a = \overline{X} - \sqrt{3}\sigma_{\text{в}}, \\ b = \overline{X} + \sqrt{3}\sigma_{\text{в}}. \end{cases}$$

Ответет. $a = \overline{X} - \sqrt{3}\sigma_{\text{в}}$, $b = \overline{X} + \sqrt{3}\sigma_{\text{в}}$.

Метод моментов является наиболее простым методом оценки параметров, однако далеко не всегда приводит к наилучшим оценкам. Далее рассмотрим более эффективный (и более трудоемкий) метод нахождение точечных оценок – *метод максимального правдоподобия*.

Пусть $x_1, x_2, \cdots, x_n$ – реализация выборки, полученная в результате проведения n независимых наблюдений за случайной величиной X. Пусть вид закона распределения случайной величины X известен (в частности, известен вид плотности распределения вероятностей $f(x,\theta)$), но неизвестен параметр θ, влияющий на этот вид. Нужно по выборке оценить параметр θ.

Ключевым понятием метода максимального правдоподобия является понятие *функции правдоподобия*:

$$L(x_1, x_2, \cdots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta),$$

где $f(x,\theta)$ – плотность распределения вероятностей непрерывной случайной величины X.

Если же случайная величина X – дискретная, то функция правдоподобия имеет вид

$$L(x_1, x_2, \cdots, x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta),$$

где $p(x,\theta) = P(X = x)$, причем эта вероятность зависит от параметра θ.

Из определения функции правдоподобия вытекает, что чем больше ее значение, тем более вероятно (правдоподобно) появление при фиксированном θ в результате наблюдений числовых значений $x_1, x_2, \cdots, x_n$.

За точечную оценку параметра θ берут такое его значение θ, при котором функция правдоподобия достигает максимума. Эта оценка, называемая *оценкой максимального правдоподобия*, является решением уравнения

$$\left. \frac{dL(x_1, x_2, \cdots, x_n, \theta)}{d\theta} \right|_{\theta=\theta} = 0.$$

Зачастую вместо отыскания максимума функции $L(x_1, x_2, \cdots, x_n, \theta)$ ищут максимум функции $\ln L(x_1, x_2, \cdots, x_n, \theta)$, поскольку обе эти функции достигают максимума при одном и том же значении θ, но, как правило, проще искать максимум последней функции.

Итак, для нахождения оценки методом максимального правдоподобия необходимо:

1) решить уравнение правдоподобия

$$\left.\frac{d\left(\ln L(x_1, x_2, \cdots, x_n, \theta)\right)}{d\theta}\right|_{\theta=\theta} = 0;$$

2) отобрать то решение, которое обращает функцию в максимум (как правило, удобно использовать вторую производную: если $\left.\frac{d^2\left(\ln L(x_1, x_2, \cdots, x_n, \theta)\right)}{d\theta^2}\right|_{\theta=\theta} < 0$, то в точке $\theta = \theta$ достигается максимум).

Пример. Методом максимального правдоподобия найти по реализации выборки $x_1, x_2, \cdots, x_n$ оценку параметра λ показательного распределения.

<u>Решение.</u> В данной задаче речь идет о законе распределения непрерывной случайной величины с плотностью вероятности $f(x,\lambda) = \lambda e^{-\lambda x}$, где $x \geq 0$, а λ – неизвестный параметр распределения. Составим функцию правдоподобия:

$$L(x_1, x_2, \cdots, x_n, \lambda) = f(x_1, \lambda) \cdot f(x_2, \lambda) \cdot \ldots \cdot f(x_n, \lambda) =$$

$$= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdot \ldots \cdot \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

Тогда

$$\ln L(x_1, x_2, \cdots, x_n, \lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i, \qquad \frac{d\left(\ln L(x_1, x_2, \cdots, x_n, \lambda)\right)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

Приравнивая к нулю последний результат, находим искомый параметр

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{X}}.$$

Поскольку $\left.\frac{d^2\left(\ln L(x_1, x_2, \cdots, x_n, \lambda)\right)}{d\lambda^2}\right|_{\lambda=\frac{1}{\overline{X}}} = -\frac{n}{\left(1/\overline{X}\right)^2} < 0$,

то формула $\lambda = \frac{1}{\overline{X}}$ представляет собой искомую оценку.

<u>Ответ</u>**.** $\lambda = \frac{1}{\overline{X}}$ является искомой оценкой.

Точечные оценки неизвестного параметра θ хороши для первоначальных результатов обработки наблюдений. Их основной недостаток в том, что точечная оценка не несет информации о точности процедуры оценивания, а потому точечная оценка может значительно отличаться от истинного значения оцениваемого параметра.

Интервальное оценивание неизвестных параметров. Доверительные интервалы

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому часто целесообразнее пользоваться *интервальными оценками*, то есть указывать интервал, который покрывает истинное значение оцениваемого параметра с заданной вероятностью. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки $\hat{\theta}$ некоторого параметра θ справедливо неравенство $|\hat{\theta} - \theta| < \delta$, то число $\delta > 0$ характеризует *точность оценки* (чем меньше δ, тем точнее оценка). Однако статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью или с некоторой степенью надежности.

Задача интервального оценивания может быть сформулирована следующим образом: по данным выборки построить числовой интервал (θ_1, θ_2), относительно которого с заранее заданной вероятностью γ можно утверждать, что он покроет точное значение оцениваемого параметра.

Определение. **Интервал (θ_1, θ_2), покрывающий с вероятностью γ истинное значение неизвестного параметра θ, называется *доверительным интервалом* (иногда $\gamma \cdot 100\%$ *доверительным интервалом*), а γ – *доверительной вероятностью* или *надежностью оценки*.**

Определение. ***Надежностью* (*доверительной вероятностью*) оценки θ неизвестного параметра θ называется вероятность γ события, состоящего в том, что выполняется неравенство $|\theta - \theta| < \delta$, где $\delta > 0$ характеризует точность оценки.**

Если заменить последнее неравенство двойным неравенством $-\delta < \theta - \theta < \delta$, то получим формулу $P(\hat{\theta} - \delta < \theta < \hat{\theta} + \delta) = \gamma$. Таким образом, γ есть вероятность того, интервал $(\theta - \delta,\ \theta + \delta)$ покроет истинное значение параметра θ. Получили доверительный интервал радиуса δ, симметричный относительно оценки θ.

Построение доверительных интервалов

Далее получим формулы доверительных интервалов для неизвестных параметров a и σ в случае нормально распределенной случайной величины, соответствующей генеральной совокупности.

1. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии.

Пусть исследуемая случайная величина X распределена по нормальному закону с известным средним квадратическим значением σ, и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание a. Будем рассматривать выборочное среднее как случайную величину $\bar{X}$, а значения вариант выборки $x_1, x_2, \ldots, x_n$ – как одинаково распределенные независимые в совокупности случайные величины $X_1, X_2, \ldots, X_n$, каждая из которых имеет математическое ожидание a и среднее квадратическое отклонение σ. При этом $M(\bar{X}) = a$, $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ (здесь использованы свойства математического ожидания и дисперсии суммы независимых в совокупности случайных величин). Оценим вероятность выполнения неравенства $|\bar{X}-a|<\delta$, обозначив эту вероятность γ. Применим формулу для отыскания вероятности попадания нормально распределенной случайной величины $\bar{X}$ в заданный интервал:

$P(|\overline{X}-a|<\delta)=2\Phi\left(\dfrac{\delta}{\sigma\left(\overline{X}\right)}\right)$. Тогда, с учетом того, что $\sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$, получаем формулу $P(|\bar{X}-a|<\delta)=2\Phi\left(\dfrac{\delta\sqrt{n}}{\sigma}\right)$=2Φ($t$), где $t=\dfrac{\delta\sqrt{n}}{\sigma}$. Отсюда $\delta=\dfrac{t\cdot\sigma}{\sqrt{n}}$, и предыдущее равенство можно переписать так:

$$P\left(\overline{X}-\frac{t\cdot\sigma}{\sqrt{n}}<a<\overline{X}+\frac{t\cdot\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma\,.$$

Итак, интервал $\left(\overline{X}-\dfrac{t\cdot\sigma}{\sqrt{n}};\overline{X}+\dfrac{t\cdot\sigma}{\sqrt{n}}\right)$ покрывает значение неизвестного математического ожидания a с вероятностью (надежностью) γ, где значение t определяется с помощью таблиц для интегральной функции Лапласа (см. [4], Приложение 3) из условия 2Φ(t) = γ.

Пример. Найти доверительный интервал для математического ожидания нормально распределенной случайной величины, если объем выборки n = 49, $\bar{x}_в=2{,}8$, σ = 1,4, а доверительная вероятность γ = 0,9.

Решение. Определим значение t, при котором Φ(t) = 0,9:2 = 0,45. По таблице Приложения 3 находим, что $t\approx$ 1,645. Тогда искомый доверительный интервал имеет вид

$$2{,}8-\frac{1{,}645\cdot 1{,}4}{\sqrt{49}}<a<2{,}8+\frac{1{,}645\cdot 1{,}4}{\sqrt{14}}$$ или 2,471< a <3,129.

Ответ. Доверительный интервал **(**2,471;3,129) покрывает неизвестное математическое ожидание a с надежностью 0,9.

2. Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания a используем новую случайную величину

$$T=\frac{\overline{X}-a}{\dfrac{s}{\sqrt{n}}},$$

где $\overline{X}$ – выборочное среднее, s – исправленная дисперсия, n – объем выборки. Случайная величина T, возможные значения которой будем обозначать t, имеет распределение Стьюдента с числом степеней свободы равным $\nu=n-1$.

Плотность распределения Стьюдента $s(t,n)=B_n\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}$, где $B_n=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}$, явным образом не зависит от a и σ. Поэтому можно задать вероятность попадания значений случайной величины T в некоторый интервал $(-t_\gamma, t_\gamma)$, учитывая четность плотности данного распределения, следующим образом: $P\left(\left|\frac{\overline{X}-a}{\frac{s}{\sqrt{n}}}\right|<t_\gamma\right)=2\int_0^{t_\gamma}s(t,n)dt=\gamma$. В данной формуле γ – это заданная надежность. Отсюда следует формула

$$P\left(\overline{X}-\frac{t_\gamma\cdot s}{\sqrt{n}}<a<\overline{X}+\frac{t_\gamma\cdot s}{\sqrt{n}}\right)=\gamma.$$

Таким образом, получен симметричный относительно выборочного среднего $\overline{X}$ доверительный интервал для неизвестного математического ожидания a, где t_γ можно найти при заданных n и γ по таблице распределения Стьюдента, составленной для отыскания двусторонней критической области. Заметим, что число $\alpha=1-\gamma$ называется *уровнем значимости*. Значение искомого параметра t_γ находят по таблице, зная уровень значимости $\alpha=1-\gamma$ и число степеней свободы $\nu=n-1$.

Замечание. Иногда в учебниках приводят таблицу распределения Стьюдента, составленную для отыскания правосторонней критической области (см. Приложение 4). Чтобы по такой таблице определить число t_γ, используемое в формуле доверительного интервала, нужно уровень значимости α разделить пополам и найти значение параметра t_γ по таблице для односторонней критической области, принимая в качестве уровня значимости число $\frac{\alpha}{2}$, а в качестве числа степеней свободы $\nu=n-1$.

Пример. Пусть объем выборки n = 25, $\overline{x}_B=3$, s = 1,5. Найти доверительный интервал для математического ожидания a при γ = 0,99.

<u>Решение.</u> Из таблицы ([4], Приложение 4) находим t_γ (n = 25, γ = 0,99, $\alpha=1-\gamma$=0,01, $\nu=n-1=24$): t_γ ≈ 2,797. Тогда формула доверительного интервала для математического ожидания, покрывающего a с надежностью 0,99, имеет вид

$$3-\frac{2{,}797\cdot 1{,}5}{\sqrt{25}}<a<3+\frac{2{,}797\cdot 1{,}5}{\sqrt{25}}$$ или 2,161< a < 3,839.

Ответ. Доверительный интервал (2,161;3,839) покрывает неизвестное математическое ожидание a с надежностью 0,99.

3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал в виде $(s-\delta\,;\,s+\delta)$, где s – исправленное выборочное среднее квадратическое отклонение, а точность оценки δ удовлетворяет условию

$P(|\sigma-s|<\delta)=\gamma$ (γ – это заданная надежность). Запишем неравенство

$|\sigma-s|<\delta$ в виде $s\left(1-\dfrac{\delta}{s}\right)<\sigma<s\left(1+\dfrac{\delta}{s}\right)$ или

$$s(1-q)<\sigma<s(1+q),\ \text{где}\ q=\frac{\delta}{s}.$$

Рассмотрим случайную величину χ, определяемую по формуле

$$\chi=\frac{s}{\sigma}\sqrt{n-1},$$

которая распределена по закону «хи-квадрат» с числом степеней свободы, равным n–1. Плотность ее распределения

$$R(\chi,n)=\frac{\chi^{n-2}e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\dfrac{n-1}{2}\right)}$$

не зависит от оцениваемого параметра σ, а зависит только от объема выборки n. Вероятность того, что значение случайной величины χ удовлетворяет неравенству $\chi_1<\chi<\chi_2$, равна доверительной вероятности γ, следовательно, $\int\limits_{\chi_1}^{\chi_2}R(\chi,n)d\chi=\gamma.$ Предположим, что $q<1$, тогда неравенство $s(1-q)<\sigma<s(1+q)$ можно записать в виде

$$\frac{1}{s(1+q)}<\frac{1}{\sigma}<\frac{1}{s(1-q)}$$

или, после умножения на число $s\sqrt{n-1}$, в виде

$$\frac{\sqrt{n-1}}{1+q}<\frac{s\sqrt{n-1}}{\sigma}<\frac{\sqrt{n-1}}{1-q}.$$

Следовательно, $\dfrac{\sqrt{n-1}}{1+q}<\chi<\dfrac{\sqrt{n-1}}{1-q}$. Тогда $\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}}R(\chi,n)d\chi=\gamma.$ Существуют таблицы, составленные для распределения «хи-квадрат», из которых можно найти q по заданным n и γ, не решая это уравнение (см. [4], Приложение 6). Таким образом, вычислив по выборке значение s и определив по таблице значение q, можно найти доверительный интервал, который покрывает неизвестное значение σ с заданной вероятностью γ.

Замечание. Если $q > 1$, то с учетом условия $\sigma > 0$ доверительный интервал для σ будет иметь границы

$$0 < \sigma < s(1+q).$$

Пример. Пусть $n = 20$, $s = 1{,}3$. Найти доверительный интервал для неизвестного параметра σ при заданной надежности γ = 0,95.

<u>Решение.</u> Из таблицы (см. [4], Приложение 6) находим значение параметра q: $q(n = 20, \gamma = 0{,}95) \approx 0{,}37$. Следовательно, границы доверительного интервала будут такими: 1,3(1– 0,37) = 0,819 и 1,3(1+0,37) = 1,781. Итак, интервал 0,819 < σ < 1,781 покрывает параметр σ с вероятностью 0,95.

<u>Ответ.</u> Доверительный нтервал (0,819; 1,781) покрывает неизвестный параметр σ с вероятностью 0,95.

1.3.4. Проверка статистических гипотез

Статистические гипотезы. Статистические критерии

Очень часто при решении практических задач, связанных с применением методов математической статистики, возникает вопрос: может ли на основании данных некоторой выборки быть принято или отвергнуто некоторое предположение (гипотеза) относительно генеральной совокупности (или, иными словами, наблюдаемой случайной величины). Например, испытана партия новой модели автомобилей. Можно ли по результатам испытания сделать обоснованный вывод о том, что новая модель по сравнению с предыдущей более безопасна (имеет лучшие технические характеристики, меньше загрязняет окружающую среду). Аналогичный вопрос возникает и при апробации новых лекарств, новых методик обучения, при внедрении новых технологии и т. п.

Определение. ***Статистическая гипотеза*** **– это некоторое предположение относительно генеральной совокупности, проверяемое по выборочным данным.**

Гипотезу называют *параметрической*, если в ней содержится некоторое утверждение о значении неизвестного параметра (значениях неизвестных параметров) известного вида распределения случайной величины, соответствующей генеральной совокупности. Гипотезу называют *непараметрической*, если в ней содержится некоторое утверждение относительно неизвестного вида распределения случайной величины, соответствующей генеральной совокупности.

Обычно выделяют некоторую *основную* или *нулевую* гипотезу H_0 и *альтернативную* ей (*конкурирующую*) гипотезу H_1, которая противоречит H_0. Правило, по которому принимают решение о принятии или отклонении гипотезы H_0 (соответственно, об отклонении или принятии гипотезы H_1), называют *критерием*.

Определение. ***Статистический критерий*** **– это некоторый статистический показатель, являющийся числовой мерой отклонения, различия в том или ином смысле между выборочным и теоретическим**

распределениями, где теоретическое распределение получено в предположении, что подлежащая проверке гипотеза H_0 верна.

Статистический критерий зависит от выборочных значений $(x_1, x_2, ..., x_n)$ и является случайной величиной. Множество всех возможных значений такой случайной величины можно представить как объединение двух непересекающихся подмножеств:

— множество всех тех возможных значений критерия, при которых принимают гипотезу H_0 и отклоняют гипотезу H_1;

— множество всех тех возможных значений критерия, при которых принимают гипотезу H_1 и отклоняют гипотезу H_0 – так называемая *критическая область* значений статистического критерия.

При принятии решения могут иметь место ошибки двух родов:

— если принята гипотеза H_1, тогда как на самом деле верна гипотеза H_0, то говорят, что имеет место *ошибка первого рода*; вероятность такой ошибки называют *уровнем значимости* и обозначают α;

— если принята гипотеза H_0, тогда как на самом деле верна гипотеза H_1, то говорят, что имеет место *ошибка второго рода*; вероятность такой ошибки обозначают β.

Соответственно правильные решения тоже бывают двух родов:

— принята гипотеза H_0, тогда как она верна; согласно определению ошибки первого рода, вероятность такого решения равна $(1-\alpha)$;

— принята гипотеза H_1, тогда как она верна; согласно определению ошибки второго рода, вероятность такого решения равна $(1-\beta)$.

Вероятность ошибки первого рода определяют выбранным изначально уровнем значимости α. Если, например, принять $\alpha = 0{,}05$, то это означает, что в среднем в пяти случаях из 100 мы рискуем отвергнуть верную гипотезу в ходе процедуры статистической проверки. Может показаться, что эффективность статистической процедуры можно увеличить, понизив уровень значимости, т. е. уменьшив вероятность совершения ошибки первого рода. Однако следующий пример помогает понять ошибочность такого мнения. Пусть проводят тестирование нового лекарства, чтобы убедиться в его безопасности, т. е. отсутствии серьезных побочных эффектов при массовом использовании. При принятии лекарства наблюдают опасные реакции у малой части взрослых пациентов. Нулевая гипотеза H_0 сводится в данном случае к утверждению: лекарство безопасно для массового использования. Альтернативная гипотеза H_1 сводится к утверждению: лекарство опасно. Таким образом, уменьшая уровень значимости α, мы уменьшаем вероятность в ходе статистической проверки сделать вывод об опасности лекарства для пациентов. Подчеркнем также, что последствия ошибок первого и второго родов могут оказаться различными. В данном случае, если будет принята гипотеза H_1: «лекарство опасно» при условии, что на самом деле верна гипотеза H_0: «лекарство безопасно», то эта ошибка первого рода может повлечь материальный ущерб. Если же будет принята гипотеза H_0: «лекарство безопасно», тогда как на самом деле верна гипотеза H_1: «лекарство опасно», то эта ошибка второго рода может повлечь не только материальный ущерб, но и ухудшение здоровья, а, возможно, и гибель людей.

В контексте того или иного статистического исследования можно устанавливать определенный уровень значимости α – вероятность ошибки первого рода, но исследователь не обладает подобной свободой в отношении ошибок второго рода. Наиболее распространенная причина ошибки второго рода – это слишком малый объем выборки, поэтому в случаях относительно простых исследований снижение риска ошибки второго рода должно достигаться за счет увеличения объема выборки. Число $(1-\beta)$ называют *мощностью критерия*. Для определения наилучшего критерия проверки основной гипотезы среди всех рассматриваемых критериев выбирают тот, у которого мощность является наибольшей при одном и том же выбранном уровне значимости α.

Замечание. Принятие основной гипотезы H_0 в результате применения того или иного статистического критерия не означает, что она – единственно подходящая для этого случая. Просто нулевая гипотеза не противоречит выборочным данным, но таким же свойством наряду с ней могут обладать и другие гипотезы.

Алгоритм проверки статистических гипотез

1. Выдвигают основную гипотезу H_0 и альтернативную ей гипотезу H_1. Задают число α – вероятность ошибки первого рода (т. е. вероятность в результате проведения статистической процедуры принять гипотезу H_1 и отклонить гипотезу H_0, в то время как на самом деле верна гипотеза H_0).

2. Выбирают статистический критерий K для проверки гипотезы H_0. Критерий K – это случайная величина, т. к. его значения определяются по выборочным значениям согласно конкретной формуле. Распределение случайной величины K известно. Как правило, составлены таблицы значений для критерия K, в которых приведено решение $K_{\text{кр.}}$ уравнения $P\left(|K| > K_{\text{кр.}}\right) = \alpha$ или $P\left(K > K_{\text{кр.}}\right) = \alpha$. В случае первого уравнения в скобках записано событие, состоящее в попадании возможного значения критерия K в *двустороннюю критическую область* значений критерия, а в случае второго уравнения в скобках кратко записано событие, состоящее в попадании возможного значения критерия K в *правостороннюю критическую область* значений критерия.

3. Вычисляют по таблице критическое значение $K_{\text{кр.}}$ по заданным уровню значимости α и числу степеней свободы; определяют критическую область значений критерия.

4. Вычисляют фактическое значение $K_{\text{набл.}}$ статистического критерия K по конкретной случайной выборке, предполагая при этом, что гипотеза H_0 верна.

5. Принимают статистическое решение. Если $K_{\text{набл.}}$ принадлежит критической области значений критерия, то гипотезу H_0 отклоняют при заданном α и принимают гипотезу H_1. Такой вывод делают потому, что число

α в исследованиях мало, и тогда событие, состоящее в попадании наблюдаемого значения критерия в критическую область, в случае конкретной случайной выборки, крайне маловероятно. Если рассмотрев одну случайную выборку, мы сталкиваемся с наступлением крайне маловероятного события, то это может лишь означать, что гипотеза H_0 опровергается опытными данными и, следовательно, принимается гипотеза H_1. Если же $K_{\text{набл.}}$ не принадлежит критической области значений критерия, то гипотезу H_0 принимают при заданном α и отклоняют гипотезу H_1.

В данном алгоритме вероятность отклонения гипотезы H_0 при первоначальном предположении, что она верна, есть вероятность α. Это соответствует определению вероятности ошибки первого рода или уровня значимости статистической процедуры проверки гипотезы. При реализации алгоритма желательно, чтобы статистический критерий K не зависел от конкретной функции распределения случайной величины, соответствующей генеральной совокупности; имел достаточно простой вид и был чувствителен к отклонению данных, полученных опытным путем, от предполагаемой гипотезы.

Примеры параметрических гипотез и их проверки

Пусть генеральная совокупность изучается по выборочной совокупности объема *n*. Пусть изучаемая по выборке случайная величина X, соответствующая генеральной совокупности, распределена по нормальному закону и ее параметры $M(X)=a$ и $\sigma(X)=\sigma$ неизвестны, но по выборке построены их точечные оценки $\overline{X}$ и $s=\sqrt{s^2}$. Выдвигают основную статистическую гипотезу $H_0:\ \ a=a_0$, где a_0 – некоторое конкретное предполагаемое число. Альтернативную гипотезу формулируют так: $H_1:\ \ a\neq a_0$. Задают уровень значимости α статистической процедуры проверки гипотезы – вероятность ошибки первого рода.

В качестве статистического критерия проверки гипотезы выбирают *T*-критерий: $T=\frac{\overline{X}-a_0}{s}\sqrt{n}$. Известно, что для случайной выборки объема *n* из нормально распределенной генеральной совокупности этот критерий имеет распределение Стьюдента с числом степеней свободы $\nu=n-1$. По заданным в задаче значениям α и $\nu=n-1$ находят $T_{\text{кр.}}$ по таблице значений распределения Стьюдента для двусторонней критической области. Речь идет о двусторонней критической области, т. к. согласно альтернативной гипотезе отклонение $M(X)=a$ от предполагаемого значения a_0 допускается как в большую, так и в меньшую сторону, а значит, значение $T=\frac{\overline{X}-a_0}{s}\sqrt{n}$ может быть как положительным, так и отрицательным. Затем вычисляют $T_{\text{набл.}}$, делая расчеты по указанной формуле по конкретным выборочным данным. Если $T_{\text{набл.}}$ попало в критическую область $(-\infty;-T_{\text{кр.}})\cup(T_{\text{кр.}};+\infty)$, то гипотеза H_0 опровергается опытными данными, т. к. в случае рассматриваемой выборки произошло практически невероятное событие (таблица для двустороннего распределения составлена так, что в ней дается решение $T_{\text{кр.}}$ уравнения $P(|T| > T_{\text{кр.}}) = \alpha$, а α задается как достаточно малое число). Если $T_{\text{набл.}}$ не попало в критическую область $(-\infty; -T_{\text{кр.}}) \cup (T_{\text{кр.}}; +\infty)$, то гипотеза H_0 не противоречит опытным данным и, следовательно, принимается при заданном уровне значимости α.

Аналогично рассматриваются случаи:

— $H_0: \ a = a_0$, $H_1: \ a > a_0$. Отличие от предыдущих рассуждений состоит в том, что согласно альтернативной гипотезе отклонение $M(X) = a$ от предполагаемого значения a_0 допускается лишь в большую сторону, а значит, значение критерия $T = \frac{\overline{X} - a_0}{s}\sqrt{n}$ рассматривается как положительное, что приводит к построению правосторонней критической области $(T_{\text{кр.}}; +\infty)$ значений статистического критерия.

— $H_0: \ a = a_0$, $H_1: \ a < a_0$. Теперь согласно альтернативной гипотезе отклонение $M(X) = a$ от предполагаемого значения a_0 допускается лишь в меньшую сторону, а значит, значение критерия $T = \frac{\overline{X} - a_0}{s}\sqrt{n}$ рассматривается как отрицательное, что приводит к построению левосторонней критической области $(-\infty; -T_{\text{кр.}})$ значений статистического критерия.

Замечание. Таблица распределения Стьюдента, составленная для отыскания правосторонней критической области, приведена в [4], Приложении 4. По ней можно определить значение $T_{\text{кр.}}$ для правосторонней критической области по заданному уровню значимости α и числу степеней свободы $\nu = n - 1$. Левосторонняя критическая область симметрична правосторонней относительно начала координат. Чтобы по таблице Приложения 4 определить критическое значение T–критерия, используемое для отыскания двусторонней критической области, нужно сначала уровень значимости α разделить пополам. Затем найти значение параметра $T_{\text{кр.}}$ по таблице для правосторонней критической области, принимая в качестве уровня значимости число $\frac{\alpha}{2}$, а в качестве числа степеней свободы $\nu = n - 1$.

Пример. Проводят эксперимент с целью исследовать выживаемость однотипных патогенных микроорганизмов при определенном комплексе неблагоприятных воздействий. В таблице приведена информация о времени (с) активной жизнедеятельности тридцати микроорганизмов в данном эксперименте после комплекса воздействий.

Время (с)	28	29	30	31	32	33	34
m_i (количество микроорганизмов)	3	4	6	7	6	2	2

Можно ли при уровне значимости $\alpha = 0{,}01$ принять предположение, что среднее время активной жизнедеятельности микроорганизмов в данном эксперименте равно 30 с при альтернативном предположении, что оно превышает 30 с?

Решение. $H_0: \quad a = 30$, $H_1: \quad a > 30$, $n = 30$.

$$\overline{X} = \frac{1}{30}\left(28 \cdot 3 + 29 \cdot 4 + 30 \cdot 6 + 31 \cdot 7 + 32 \cdot 6 + 33 \cdot 2 + 34 \cdot 2\right) = 30{,}767;$$

$$s^2 = \frac{1}{29}\Big(\left(28 - 30{,}767\right)^2 \cdot 3 + \left(29 - 30{,}767\right)^2 \cdot 4 + \ldots + \left(34 - 30{,}767\right)^2 \cdot 2\Big) \approx 2{,}737;$$

$$s \approx 1{,}654, \quad T_{\text{набл.}} = \frac{\overline{X} - a_0}{s}\sqrt{n} \approx \frac{30{,}767 - 30}{1{,}654}\sqrt{30} \approx 2{,}54.$$

Наша задача соответствует рассмотренной ранее математической модели с использованием правосторонней критической области T–критерия, значение a_0 в этой модели в силу условия равно 30 с. Используем таблицу (см. [4], Приложение 4). Правосторонняя критическая область значений T–критерия в нашем случае имеет вид $\left(T_{\text{кр.}};+\infty\right) = \left(2{,}462;+\infty\right)$. Так как наблюдаемое по выборке значение $T_{\text{набл.}}$ принадлежит критической области, то делаем вывод, что при уровне значимости $\alpha = 0{,}01$ принимается альтернативная гипотеза H_1 и отклоняется основная гипотеза H_0. Заметим, что при уменьшении α, например, при $\alpha = 0{,}005$ или $\alpha = 0{,}001$ в этой же задаче будет приниматься гипотеза H_0 и отклоняться гипотеза H_1, т. к. наблюдаемое значение критерия не будет попадать в критическую область.

Ответ. При уровне значимости $\alpha = 0{,}01$ гипотеза о том, что среднее время активной жизнедеятельности микроорганизмов в данном эксперименте равно 30 с, отклоняется, и принимается гипотеза о том, что оно больше 30 с.

Пусть сняты показания для одной и той же случайной величины X, соответствующей генеральной совокупности, с использованием двух различных методов (например, методов химического анализа), либо метод один и тот же, но показания сняты различными лабораториями. Получены выборки одного объема n, математическими моделями которых являются совокупности $X_1, X_2, \ldots X_n$; $Y_1, Y_2, \ldots Y_n$ случайных величин. Пусть X распределена по нормальному закону, параметры $M(X) = a$ и $\sigma(X) = \sigma$ неизвестны. Требуется определить, значимо или незначимо различаются снятые в двух случаях показания. Математически это можно формализовать следующим образом:

$$H_0: \quad M(\overline{X}) = M(\overline{Y}), \quad H_1: \quad M(\overline{X}) \neq M(\overline{Y}).$$

Задача о сравнении двух средних для двух зависимых выборок сводится к задаче сравнения одного выборочного среднего с гипотетическим значением генерального среднего. Поступают так: формируют новую выборку $D_i = X_i - Y_i, \quad i = \overline{1,n}$. Для нее:

$$\overline{D}=\frac{\sum_{i=1}^{n}D_i}{n}=\frac{\sum_{i=1}^{n}X_i}{n}-\frac{\sum_{i=1}^{n}Y_i}{n}=\overline{X}-\overline{Y},\quad M\left(\overline{D}\right)=M\left(\overline{X}\right)-M\left(\overline{Y}\right).$$

Тогда задачу можно переформулировать следующим образом:

$$H_0:\quad M(\overline{D})=0,\quad H_1:\quad M(\overline{D})\neq 0.$$

Так как случайные величины X и Y распределены по нормальному закону, то и случайные величины $\overline{X},\overline{Y}$ распределены по нормальному закону, следовательно, случайная величина $\overline{D}$ также распределена по нормальному закону, причем, ее параметры неизвестны. Для этого случая применяют T–критерий, изложенный выше, где в качестве числа a_0 берут нуль (модель, в которой используется двусторонняя критическая область).

Пример. В соответствии с изложенной выше моделью получены выборки

X: 6, 7, 7, 7, 6, 7, 8, 8, 6, 5, 7, 5, 8, 6, 5, 7, 7, 7, 6, 5, 5, 6, 7, 7, 6, 6, 7, 8, 6, 6,

Y: 5, 6, 7, 7, 5, 8, 6, 7, 7 ,7, 6, 8, 5, 6, 6, 5, 7, 7, 7, 8, 6, 6, 7, 8, 7, 7, 7, 6, 5, 7.

Проверить, значимо ли различаются результаты выборочных исследований при уровне значимости $\alpha=0{,}05$.

Решение. Формируем выборку объема n=30 для случайной величины D:

1,1,0,0,1,-1,2,1,-1,-2,1,-3,3,0,-1,2,0,0,-1,-3,-1,0,0,-1,-1,-1,0,2,1,-1.

Аналогично, как и в примере 13.33, рассчитываем по этой выборке числовые показатели:

$$\overline{D}=\frac{\sum_{i=1}^{n}D_i}{n}\approx-0,067,\quad s\approx1,413,\ T_{\text{набл.}}=\frac{\overline{D}-0}{s}\sqrt{n}\approx\frac{-0,067}{1,413}\sqrt{30}\approx-0,259.$$

Критическая область в данном случае двусторонняя, т. к. согласно альтернативной гипотезе отклонение $\overline{D}$ от нуля может быть как в положительную, так и в отрицательную сторону. Находим ее с использованием таблицы распределения Стьюдента (см. [4], Приложение 4) по данным: $\alpha=0,05,\ \nu=n-1=29$. Область имеет вид $\left(-\infty;-2{,}045\right)\cup\left(2{,}045;+\infty\right)$. Поскольку наблюдаемое значение по выборке не принадлежит критической области, делаем вывод о том, что основная гипотеза согласуется с опытными данными, следовательно, расхождение результатов двух исходных выборочных исследований при $\alpha=0{,}05$ не является значимым.

Ответ. При уровне значимости $\alpha=0{,}05$ расхождение результатов двух исходных выборочных исследований не является значимым.

Одними из часто применяемых статистических критериев или критериев согласия для проверки непараметрических гипотез о неизвестном виде распределения случайной величины, соответствующей генеральной совокупности, являются *критерий согласия Пирсона* (*критерий* χ^2) и *критерий согласия Колмогорова.* Кратко изложим описание применения каждого из этих критериев.

Критерий согласия Пирсона (критерий согласия χ^2)

Пусть дана выборка $(x_1, x_2, ..., x_n)$ наблюдаемых значений случайной величины X, соответствующей генеральной совокупности. Проверяют непараметрическую гипотезу H_0, утверждающую, что закон распределения случайной величины X имеет функцию распределения $F(x)$. В качестве альтернативной гипотезы принимают противоположное утверждение.

Алгоритм применения критерия согласия Пирсона (критерия согласия χ^2)

1. Находят по выборке точечные оценки всех неизвестных числовых параметров предполагаемого гипотезой H_0 закона распределения случайной величины X. Количество таких неизвестных параметров обозначают l.

2. Если X – дискретная случайная величина, то определяют абсолютные частоты $m_i \left(i = \overline{1, k}\right)$, с которыми соответственно все попарно различные варианты $x_i \left(i = \overline{1, k}\right)$ встречаются в выборке.

Если X – непрерывная случайная величина, то разбивают область ее наблюдаемых значений на k интервалов и определяют количество m_i $\left(i = \overline{1, k}\right)$ элементов выборки, принадлежащих соответственно i-му интервалу.

3. Если X – дискретная случайная величина, то, используя предполагаемый гипотезой H_0 закон распределения этой случайной величины (ее функцию распределения $F(x)$), вычисляют вероятности $p_i \left(i = \overline{1, k}\right)$, с которыми данная случайная величина принимает соответственно значения $x_i \left(i = \overline{1, k}\right)$.

Если X – непрерывная случайная величина, то определяют вероятности $p_i \left(i = \overline{1, k}\right)$ попадания возможных значений этой случайной величины соответственно в i-й интервал, считая, что гипотеза H_0 верна. При этом за начало первого интервала принимают $(-\infty)$, а за конец последнего интервала – $(+\infty)$.

4. Вычисляют фактическое (наблюдаемое по выборке) значение статистического критерия χ^2 по формуле

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

5. Принимают статистическое решение: гипотеза H_0 не противоречит выборочным значениям при заданном уровне значимости α, если выполняется неравенство

$$\chi^2_{\text{набл}} \le \chi^2_{\alpha,\, k-l-1},$$

где значение $\chi^2_{\alpha,\, k-l-1}$ определяют при помощи Приложения 5, составленного для χ^2–распределения (число степеней свободы χ^2–распределения определяют формулой $\nu = k - l - 1$). Если же указанное выше неравенство не выполняется, то гипотезу H_0 отклоняют при заданном α и принимают альтернативную гипотезу.

Алгоритм критерия согласия Пирсона следует из теорем, суть которых сводится к следующему: *при описанных в алгоритме условиях распределение суммы вида* $\sum_{i=1}^{k}\frac{(m_i - np_i)^2}{np_i}$ *при неограниченном возрастании объема выборки* n *стремится к* χ^2*–распределению с заданным уровнем значимости* α *и с числом степеней свободы, равным* $\nu = k - l - 1$*, причем независимо от того, по какому закону* $F(x)$ *распределена случайная величина X, соответствующая генеральной совокупности*.

Таблица для χ^2–распределения (см. [4], Приложение 5) составлена таким образом, что позволяет определить по числу степеней свободы ν и уровню значимости α решение $\chi^2_{\alpha,\nu}$ уравнения

$$P\left(\chi^2 > \chi^2_{\alpha,\nu}\right) = \alpha,$$

т. е. позволяет найти так называемое *критическое значение* $\chi^2_{\alpha,\nu}$. Последнее уравнение означает, что число α – это вероятность того, что случайная величина, имеющая распределение χ^2 с числом ν степеней свободы, превзойдет критическое значение. Поскольку в задачах α задается достаточно малым, то вероятность $P\left(\chi^2 > \chi^2_{\alpha,\nu}\right)$ мала, т. е. можно считать практически несомненным, что событие $\left(\chi^2 > \chi^2_{\alpha,\nu}\right)$ не произойдет в случайном испытании. Поэтому, если для конкретной выборки $(x_1, x_2, \ldots, x_n)$ наблюдаемых значений случайной величины *X*, соответствующей генеральной совокупности, предположили, что нулевая гипотеза H_0 верна, и оказалось, что $\chi^2_{\text{набл}} = \sum_{i=1}^{k}\frac{(m_i - np_i)^2}{np_i}$ больше, чем $\chi^2_{\alpha,\nu} = \chi^2_{\alpha,\,k-l-1}$, то приходим к выводу, что гипотеза H_0 опровергнута опытом, т. к. в случайном испытании произошло практически невозможное событие. Если же $\chi^2_{\text{набл}} \le \chi^2_{\alpha,k-l-1}$, то гипотеза H_0 согласуется с опытными данными.

Пример. Проверить гипотезу о нормальном законе распределения случайной величины, соответствующей генеральной совокупности, используя критерий согласия Пирсона при уровне значимости $\alpha = 0{,}05$. Интервальный статистический ряд выборки имеет вид

Интервалы	m_i
[0,1; 0,2)	1
[0,2; 0,3)	8
[0,3; 0,4)	16
[0,4; 0,5)	37
[0,5; 0,6)	35

[0,6; 0,7)	18
[0,7; 0,8)	4
[0,8; 0,9)	3

Решение. Воспользуемся алгоритмом применения критерия согласия Пирсона.

1. Согласно сформулированной в условии гипотезе H_0, считаем, что случайная величина распределена по нормальному закону, т. е. ее плотность распределения вероятностей выражается формулой

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x\in \boldsymbol{R},$$

где неизвестными параметрами являются числа $a=M(X)$, $\sigma=\sqrt{D(X)}$. Их количество – $l=2$. Точечные оценки этих параметров вычисляются по интервальному вариационному ряду по формулам $\overline{X}=\frac{1}{n}\sum_{i=1}^{k}m_i x_i^*$, $s^2=\frac{\sum_{i=1}^{k}m_i\left(x_i^*-\overline{X}\right)^2}{n-1}$, $s=\sqrt{s^2}$, где $x^*_i\ \left(i=\overline{1,8}\right)$ – ‘это середины соответствующих интервалов. Количество интервалов в нашей задаче – $k=8$. Приведем результаты расчетов по указанным формулам: $\overline{X}\approx 0,50;\ \ s\approx 0,13$. Для их получения необходимо знать объем выборки, который легко находится сложением всех абсолютных частот из исходной таблицы (условие) и составляет $n=122$.

2. Определим вероятности $p_i\ \left(i=\overline{1,8}\right)$ попадания возможных значений случайной величины X соответственно в i-й интервал, считая, что гипотеза H_0 верна. При этом за начало первого интервала примем $(-\infty)$, а за конец последнего интервала – $(+\infty)$. Согласно нулевой гипотезе вероятность попадания возможных значений случайной величины X в некоторый интервал $[A;B)$ будем вычислять по формуле

$$P\left(X\in[A;B)\right)=\Phi\left(\frac{B-\overline{X}}{s}\right)-\Phi\left(\frac{A-\overline{X}}{s}\right).$$

Используя эту формулу, вычислим вероятности $p_i\ \left(i=\overline{1,8}\right)$, а также числа $np_i\ \left(i=\overline{1,8}\right)$, поскольку при использовании критерия Пирсона следует учитывать следующее *эмпирическое правило*: *если число степеней свободы* $\nu=k-l-1$ *больше единицы, то для каждого* i*-го интервала должно выполняться условие* $np_i\geq 5$*; если же число степеней свободы* $\nu=1$*, то более предпочтительным является условие* $np_i\geq 10$. Если это правило для некоторого интервала не выполняется, то его следует объединить с соседним интервалом, чтобы полученный в результате объединения интервал удовлетворял данному эмпирическому правилу.

Проведем необходимые вычисления: $\nu = k - l - 1 = 8 - 2 - 1 = 5$;

$$p_1 = P\left(X \in (-\infty; 0,2) \right) \approx \Phi\left(\frac{0,2 - 0,5}{0,13} \right) - \Phi(-\infty) \approx \Phi(-2,31) + \Phi(+\infty) \approx$$

$$\approx -0,4895 + 0,5 = 0,0105; \quad np_1 \approx 122 \cdot 0,0105 \approx 1,28 < 5.$$

Объединим первые два интервала статистического ряда и получим новый первый интервал $(-\infty; 0,3)$. Пересчитаем для него вероятность p_1:

$$p_1 = P\left(X \in (-\infty; 0,3) \right) \approx \Phi\left(\frac{0,3 - 0,5}{0,13} \right) - \Phi(-\infty) \approx \Phi(-1,54) + \Phi(+\infty) \approx$$

$$\approx -0,4382 + 0,5 = 0,0618; \quad np_1 \approx 122 \cdot 0,0618 \approx 7,54 > 5;$$

$$p_2 = P\left(X \in (0,3; 0,4) \right) \approx \Phi\left(\frac{0,4 - 0,5}{0,13} \right) - \Phi\left(\frac{0,3 - 0,5}{0,13} \right) \approx \Phi(-0,77) + 0,4382 \approx$$

$$\approx -0,2794 + 0,4382 = 0,1588; \quad np_2 \approx 122 \cdot 0,1588 \approx 19,37 > 5;$$

$$p_3 = P\left(X \in (0,4; 0,5) \right) \approx \Phi\left(\frac{0,5 - 0,5}{0,13} \right) - \Phi\left(\frac{0,4 - 0,5}{0,13} \right) \approx \Phi(0) + 0,2794 \approx$$

$$\approx 0,2794; \quad np_3 \approx 122 \cdot 0,2794 \approx 34,09 > 5;$$

$$p_4 = P\left(X \in (0,5; 0,6) \right) \approx \Phi\left(\frac{0,6 - 0,5}{0,13} \right) - \Phi\left(\frac{0,5 - 0,5}{0,13} \right) \approx \Phi(0,77) \approx$$

$$\approx 0,2794; \quad np_4 \approx 122 \cdot 0,2794 \approx 34,09 > 5;$$

$$p_5 = P\left(X \in (0,6; 0,7) \right) \approx \Phi\left(\frac{0,7 - 0,5}{0,13} \right) - \Phi\left(\frac{0,6 - 0,5}{0,13} \right) \approx \Phi(1,54) - 0,2794 \approx$$

$$\approx 0,4382 - 0,2794 = 0,1588; \quad np_5 \approx 122 \cdot 0,1588 \approx 19,37 > 5;$$

$$p_6 = P\left(X \in (0,7; 0,8) \right) \approx \Phi\left(\frac{0,8 - 0,5}{0,13} \right) - \Phi\left(\frac{0,7 - 0,5}{0,13} \right) \approx \Phi(2,31) - 0,4382 \approx$$

$$\approx 0,4895 - 0,4382 = 0,0513; \quad np_6 \approx 122 \cdot 0,0513 \approx 6,26 > 5;$$

$$p_7 = P\left(X \in (0,8; +\infty) \right) \approx \Phi(+\infty) - \Phi\left(\frac{0,8 - 0,5}{0,13} \right) \approx 0,5 - 0,4895 \approx 0,0105;$$

$$np_7 \approx 122 \cdot 0,0105 \approx 1,28 < 5.$$

Объединим последние два интервала и получим новый интервал $(0,7; +\infty)$. Пересчитаем для него вероятность p_6:

$$p_6 = P\left(X \in (0,7; +\infty) \right) \approx \Phi(+\infty) - \Phi\left(\frac{0,7 - 0,5}{0,13} \right) \approx 0,5 - 0,4382 \approx 0,0618;$$

$$np_6 \approx 122 \cdot 0,0618 \approx 7,54 > 5.$$

Сведем вычисления в таблицу:

Таблица 13.5

Интервалы	m_i	p_i	np_i	$(m_i - np_i)^2$	$(m_i - np_i)^2 / (np_i)$
$(-\infty; 0,3)$	9	0,0618	7,54	2,13	0,28
[0,3; 0,4)	16	0,1588	19,37	11,36	0,59

[0,4; 0,5)	37	0,2794	34,09	8,47	0,25
[0,5; 0,6)	35	0,2794	34,09	0,83	0,02
[0,6; 0,7)	18	0,1588	19,37	1,88	0,10
[0,7; $+\infty$)	7	0,0618	7,54	0,29	0,04

$\sum m_i = 122$ $\sum p_i = 1$ $\chi^2_{\text{набл}} \approx 1{,}28$

3. Сложив все элементы последнего столбца, получаем наблюдаемое по выборке значение статистического критерия $\chi^2_{\text{набл}} \approx 1{,}28$.

4. Вычисляем число степеней свободы по формуле $\nu = k - l - 1 = 6 - 2 - 1 = 3$ и значение $\chi^2_{\alpha,\, k-l-1} = \chi^2_{0{,}05,\, 3} \approx 7{,}81$ (см. [4], Приложение 5). Замечаем, что $1{,}28 < 7{,}81$. Принимаем статистическое решение: гипотеза H_0 не противоречит выборочным значениям при уровне значимости $\alpha = 0{,}05$, т. к. $\chi^2_{\text{набл}} < \chi^2_{0{,}05,3}$. Итак, при уровне значимости $\alpha = 0{,}05$ нулевая гипотеза принимается.

Ответ. При уровне значимости $\alpha = 0{,}05$ гипотеза о нормальном законе распределения исследуемой случайной величины принимается.

Критерий согласия Колмогорова

Если гипотетический закон распределения непрерывной случайной величины *X*, соответствующей генеральной совокупности, известен полностью, т. е. вид функции распределения $F(x)$ и числовые значения всех параметров распределения известны, то можно применять критерий согласия Колмогорова, который в вычислительном плане проще критерия согласия χ^2. Пусть дана выборка $(x_1, x_2, \ldots, x_n)$ наблюдаемых значений непрерывной случайной величины *X*. Проверяют непараметрическую гипотезу H_0, утверждающую, что закон распределения случайной величины *X* имеет функцию распределения $F(x)$. В качестве альтернативной гипотезы принимают противоположное утверждение.

Алгоритм применения критерия согласия Колмогорова

1. Разбивают область наблюдаемых значений непрерывной случайной величины *X* на *k* интервалов вида $[x_{i-1}; x_i), i = \overline{1, k}$, и определяют количество m_i $\left(i = \overline{1, k}\right)$ элементов выборки, принадлежащих соответственно *i*-му интервалу.

2. Определяют *накопленную частоту* ν_i $\left(i = \overline{1, k}\right)$ для каждого из интервалов $[x_{i-1}; x_i), i = \overline{1, k},$ соответственно по формуле:

$$\nu_i = m_1 + m_2 + \cdots + m_i .$$

Находят значения эмпирической функции распределения $F^*(x)$ интервального статистического ряда в точках x_i $\left(i = \overline{1, k}\right)$ по формуле

$$F^*(x_i) = \frac{\nu_i}{n} .$$

3. Находят значения теоретической функции распределения $F(x)$ наблюдаемой случайной величины в точках x_i $\left(i=\overline{1,k}\right)$, считая, что гипотеза H_0 верна. При этом за конец x_k последнего интервала принимают $(+\infty)$, а за начало x_0 первого – $(-\infty)$.

4. Вычисляют значение величины $D_n = \max\limits_{i=\overline{1,k}}\left|F^*(x_i) - F(x_i)\right|$ и фактическое (наблюдаемое по выборке) значение $\lambda_{\text{набл}} = D_n\sqrt{n}$ – статистический показатель критерия Колмогорова.

5. Принимают статистическое решение: гипотеза H_0 не противоречит выборочным значениям при заданном уровне значимости α, если выполняется неравенство

$$\lambda_{\text{набл}} \le \lambda_\alpha,$$

где значение λ_α определяют по таблице критических значений распределения Колмогорова ([4], Приложение 7). Если же указанное выше неравенство не выполняется, то гипотеза H_0 отклоняется при заданном α и принимается альтернативная гипотеза.

Алгоритм критерия согласия Колмогорова следует из теорем, суть которых сводится к следующему: *при неограниченном возрастании объема выборки n функция распределения случайной величины $\lambda_n = D_n\sqrt{n}$ стремится к функции Колмогорова независимо от того, по какому закону $F(x)$ распределена непрерывная случайная величина X, соответствующая генеральной совокупности*.

Таблица критических значений распределения Колмогорова (см. ([4], Приложение 7) составлена таким образом, что позволяет определить по заданному уровню значимости α решение λ_α уравнения

$$P(\lambda > \lambda_\alpha) = \alpha,$$

т. е. позволяет найти так называемое *критическое значение* λ_α. Последнее уравнение означает, что число α – это вероятность того, что случайная величина, имеющая распределение Колмогорова, превзойдет критическое значение. Поскольку в задачах число α задается достаточно малым, то вероятность $P(\lambda > \lambda_\alpha)$ мала, т. е. можно считать практически несомненным, что событие $(\lambda > \lambda_\alpha)$ не произойдет в случайном испытании. Поэтому, если для конкретной выборки $(x_1, x_2, \ldots, x_n)$ наблюдаемых значений непрерывной случайной величины *X*, соответствующей генеральной совокупности, предположили, что нулевая гипотеза H_0 верна, и оказалось, что $\lambda_{\text{набл}} = D_n\sqrt{n}$ больше, чем λ_α, то приходим к выводу, что гипотеза H_0 опровергнута опытом, т. к. в случайном испытании произошло практически невозможное событие. Если же $\lambda_{\text{набл}} \le \lambda_\alpha$, то гипотеза H_0 согласуется с опытными данными.

Замечание. Если имеются несколько гипотетических распределений, к которым применяют критерий согласия Колмогорова, то из них лучше согласуется с эмпирическим распределением то, для которого $\lambda_{\text{набл}}$ принимает наименьшее значение.

Критерий Колмогорова значительно проще в вычислительном плане критерия Пирсона. Однако для его применения необходимо, чтобы предполагаемый теоретический закон непрерывного распределения был полностью задан, т. е. вид функции распределения $F(x)$ и числовые значения всех параметров распределения должны быть известны. Если при неизвестных числовых параметрах закона распределения применить критерий Колмогорова, взяв за значения параметров их оценки, то возможна ситуация, когда по этому критерию гипотеза будет признана допустимой, а на самом деле она неверна.

Пример. Используя выборочные данные из примера 13.35, проверить при уровне значимости $\alpha = 0{,}05$ гипотезу H_0, утверждающую, что случайная величина X, соответствующая генеральной совокупности, распределена по нормальному закону с известными числовыми параметрами $a = 0{,}5; \sigma = 0{,}1$.

Решение. Воспользуемся алгоритмом применения критерия согласия Колмогорова. Согласно сформулированной в условии гипотезе H_0 считаем, что случайная величина распределена по нормальному закону с числовыми параметрами $a = 0{,}5; \sigma = 0{,}1$, т. е. ее плотность распределения вероятностей выражается формулой

$$p(x) = \frac{1}{0{,}1\cdot\sqrt{2\pi}} e^{-\frac{(x-0{,}5)^2}{2\cdot 0{,}1^2}}, \quad x \in \boldsymbol{R}.$$

1. Интервальный статистический ряд выборки приведен в условии примера 13.35.

2–3. Накопленную частоту ν_i для каждого из интервалов $[x_{i-1}; x_i)$, $i = \overline{1,8}$, соответственно вычислим по формуле

$$\nu_i = m_1 + m_2 + \cdots + m_i.$$

При этом за конец x_8 последнего интервала примем $(+\infty)$, а за начало x_0 первого – $(-\infty)$. Значение эмпирической функции распределения $F^*(x)$ интервального статистического ряда в каждой точке $x_i \left(i = \overline{1,8}\right)$ вычислим по формуле $F^*(x_i) = \frac{\nu_i}{n} = \frac{\nu_i}{122}$. Согласно нулевой гипотезе, значения теоретической функции распределения $F(x)$ случайной величины X в точках x_i будем вычислять таким образом:

$$F(x_i) = \frac{1}{2} + \Phi\left(\frac{x_i - a}{\sigma}\right), i = \overline{1,8}.$$

Результаты выполнения пунктов 2, 3 алгоритма применения критерия согласия Колмогорова оформим в виде таблицы.

Интервалы	m_i	ν_i	$F^*(x_i)$	$(x_i - a)/\sigma$	$F(x_i)$	$\lvert F^*(x_i) - F(x_i)\rvert$

$[x_{i-1}; x_i)$						
(-∞; 0,2)	1	1	0,0082	-3	0,0014	0,0068
[0,2; 0,3)	8	9	0,0738	-2	0,0228	0,0510
[0,3; 0,4)	16	25	0,2049	-1	0,1587	0,0462
[0,4; 0,5)	37	62	0,5082	0	0,5000	0,0082
[0,5; 0,6)	35	97	0,7951	1	0,8413	0,0462
[0,6; 0,7)	18	115	0,9426	2	0,9772	0,0346
[0,7; 0,8)	4	119	0,9754	3	0,9987	0,0233
[0,8; +∞)	3	122	1,0000	+∞	1,0000	0,0000

4. Из последнего столбца находим наибольший модуль разности между соответствующими значениями эмпирической и гипотетической функций распределения: $D_n = \max\limits_{i=\overline{1,k}} \left| F^*(x_i) - F(x_i) \right| = 0,0510$. Тогда статистический показатель критерия Колмогорова, вычисляемый по выборке, равен $\lambda_{\text{набл}} = D_n \sqrt{n} = 0,0510 \cdot \sqrt{122} \approx 0,563$.

5. По таблице критических значений распределения Колмогорова (см. ([4], Приложение 7) вычисляем $\lambda_\alpha = \lambda_{0,05} = 1,358$ и замечаем, что 0,563<1,358. Принимаем статистическое решение: гипотеза H_0 не противоречит выборочным данным при заданном уровне значимости α=0,05, т. к. выполняется неравенство $\lambda_{\text{набл}} < \lambda_{0,05}$. Итак, при уровне значимости $\alpha = 0,05$ основная гипотеза H_0 принимается.

Ответ. При уровне значимости $\alpha = 0,05$ гипотеза о нормальном законе распределения исследуемой случайной величины с числовыми параметрами $a = 0,5; \sigma = 0,1$. принимается.

2. ПРАКТИЧЕСКИЙ РАЗДЕЛ
ПРИМЕРНАЯ ТЕМАТИКА ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Занятие № 1. Основные понятия теории множеств.

Примеры решения задач

1. *Задайте множество способом перечисления его элементов:*

а) *Множество А – «множество дней недели».*

Решение. Имеем, множество A = {понедельник, вторник, среда, четверг, пятница, суббота, воскресенье}.

б) *Множество В – «множество основных арифметических действий».*

Решение. Следуя школьной традиции, будем считать основными четыре арифметических действия, значит, множество B = {сложение, вычитание, умножение, деление}.

2. *Задайте множества из предыдущей задачи* 1, *указывая только характеристические свойства их элементов.*

Указанные множества A и B можно задать следующим образом:

а) $A = \{a$: a – день недели$\}$.

б) $B = \{b$: b – основное арифметическое действие$\}$.

3. *Задайте множество стран Содружества независимых государств* (*СНГ*) *с помощью указания характеристического свойства его элементов и способом перечисления элементов, составляющих это множество.*

Решение. Пусть A – «множество стран СНГ». Зададим множество A с помощью указания характеристического свойства его элементов: $A = \{a$: a – страна СНГ$\}$. Зададим множество A способом перечисления всех 11 элементов (стран), составляющих это множество: A = {Азербайджан, Армения, Беларусь, Казахстан, Киргизия, Молдова, Россия, Таджикистан, Туркменистан, Узбекистан, Украина}.

4. *Перечислите элементы множеств, заданных с помощью характеристических свойств*:

а) *Множество решений квадратного уравнения* $x^2-2x-15=0$, *т. е.* $X=\{x: x^2-2x-15=0\}$.

Решение. Имеем, множество $X = \{-3, 5\}$.

б) *Множество* $X = \{x$ – *натуральное число*: $-4 < x \le 3\}$.

Решение. Имеем, множество $X = \{1, 2, 3\}$.

5. *Пусть* ***U*** *– «множество языков мира». Приведите примеры подмножеств данного множества.*

Подмножествами множества ***U*** являются, например, «множество европейских языков», «множество романских языков», «множество восточных языков».

6. *Пусть А – «множество стран СНГ», В – «множество стран таможенного союза». Проверить справедливость включений*: $A \subset B$, $B \subset A$.

Решение. Зададим множество A способом перечисления его одиннадцати элементов-государств: A = {Азербайджан, Армения, Беларусь, Казахстан, Киргизия, Молдова, Россия, Таджикистан, Туркменистан, Узбекистан, Украина}. Зададим множество B способом перечисления его трех элементов-государств: B = {Беларусь, Казахстан, Россия}. Поскольку множество A содержит элементы, которые не входят в множество B, то включение $A \subset B$ неверно. Так как каждый элемент множества B является также элементом множества A, то множество B является подмножеством множества A, значит, справедливо включение $B \subset A$.

7. *Выписать все подмножества данного множества:*

а) *Множество* $A = \{\{0\}, 2\}$.

Решение. Множество A имеет четыре подмножества: $\varnothing$, $\{\{0\}\}$, $\{2\}$, $\{\{0\}, 2\}$.

б) *Множество стран таможенного союза* B = {Беларусь, Казахстан, Россия}.

Решение. Так как множество B состоит из трех элементов, то множество B имеет восемь подмножеств. Перечислим их: $\varnothing$, {Беларусь}, {Казахстан}, {Россия}, {Беларусь, Казахстан}, {Беларусь, Россия}, {Казахстан, Россия}, {Беларусь, Казахстан, Россия}.

8. *Совпадают ли множества A и B, если* $A = \{\{1, 2\}, 3\}$ *и* $B = \{1, 2, 3\}$?

Решение. Нет, не совпадают, так как множество A состоит из двух элементов $\{1, 2\}$ и 3, а множество B – из трех элементов: 1, 2, 3.

9. *Верны, ли следующие включения и вложения:*

а) $\{1, 2\} \in \{\{1, 2, 3\}, \{1, 3\}, 1, 2\}$.

Решение. Нет, не верно, так как в множестве $\{\{1, 2, 3\}, \{1, 3\}, 1, 2\}$ нет элемента $\{1, 2\}$. Это множество содержит 4 элемента: $\{1, 2, 3\}$, $\{1, 3\}$, 1 и 2.

б) $\{2, 5\} \subset \{1, 2, 5, 7\}$.

Решение. Да, верно, так как множество $\{2, 5\}$ – это подмножество множества $\{1, 2, 5, 7\}$.

Задачи для самостоятельного решения

1. *Задайте множество X – «множество решений квадратного уравнения* $x^2 - 5x + 6 = 0$*» способом перечисления его элементов.*

2. *Задайте множество X – «множество решений квадратного уравнения* $x^2 - 5x + 6 = 0$*» с помощью указания характеристических свойств его элементов.*

3. *Составьте список элементов множества, заданного с помощью характеристического свойства элементов:*

а) *Множество* $X = \{x$ *– натуральное число:* $x < 7\}$.

б) *Множество* $X = \{x$ *– натуральное число:* $|x| < 4\}$.

4. *Пусть* $\boldsymbol{U}$ *– «множество всех студентов БГУ». Являются ли «множество студентов ИФ», «множество иностранных студентов БГУ» и «множество девушек-студенток БГУ» подмножествами множества* $\boldsymbol{U}$*?*

5. *Выписать все подмножества данных множеств*:

а) $A = \{1, 3\}$; **б)** $B = \{\{1, 2\}, 3\}$; **в)** $C = \{\{1, 2, 3\}\}$.

6. *Выяснить, справедливы ли включения*:

а) $\{1, 2\} \subset \{\{1, 2, 3\}, \{1, 3\}, 1, 2\}$;
б) $\{2, 5\} \subset \{1, 2, 4, \{5\}, 7\}$;
в) $\{-1, 1\} \subset \{x\colon\ x^2 = 1\}$.
7. *Проверить, совпадают ли множества A и B:*
а) $A = \{\{1, 3\}, 2\}$ *и* $B = \{1, 2, 3\}$.
б) $A = \{2, 3\}$ *и* $B = \{x\colon x^2 - 5 \cdot x + 6 = 0\}$.

Занятие № 2. Операции над множествами, их свойства. Число элементов в объединении конечных множеств.

Примеры решения задач

Для решения задач определим следующие множества стран.

Множество стран Европейского союза (*Евросоюз, ЕС*): A = {Австрия, Бельгия, Болгария, Великобритания, Венгрия, Германия, Греция, Дания, Ирландия, Испания, Италия, Кипр, Латвия, Литва, Люксембург, Мальта, Нидерланды, Польша, Португалия, Румыния, Словакия, Словения, Финляндия, Франция, Чехия, Швеция, Эстония}. Множество A содержит 27 элементов (стран).

Множество стран организации Североатлантического договора (*блок НАТО*): B = {Албания, Бельгия, Болгария, Великобритания, Венгрия, Германия, Греция, Дания, Исландия, Испания, Италия, Канада, Латвия, Литва, Люксембург, Нидерланды, Норвегия, Польша, Португалия, Румыния, Словакия, Словения, США, Турция, Франция, Хорватия, Чехия, Эстония}. Множество B содержит 28 элементов (стран).

Множество стран Шенгенской зоны: C = {Австрия, Бельгия, Венгрия, Германия, Греция, Дания, Исландия, Испания, Италия, Латвия, Литва, Лихтенштейн, Люксембург, Мальта, Нидерланды, Норвегия, Польша, Португалия, Словакия, Словения, Финляндия, Франция, Чехия, Швейцария, Швеция, Эстония}. Множество C содержит 26 элементов (стран).

1. *Пусть A – «множество стран, которые входят в Европейский союз», B – «множество стран-членов блока НАТО». Описать словами множества*: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.

Решение. Дадим описания заданных множеств: $A \cup B$ – «множество стран, которые входят в Евросоюз или в НАТО» (34 государства), $A \cap B$ – «множество стран, которые одновременно входят и в Евросоюз, и в НАТО» (21 государство), $A \setminus B$ – «множество стран, которые входят в Евросоюз, но не являются членами блока НАТО» (Австрия, Ирландия, Кипр, Мальта, Финляндия и Швеция), $B \setminus A$ – «множество стран, которые являются членами блока НАТО, но не входят в Евросоюз» (Албания, Исландия, Канада, Норвегия, США, Турция, Хорватия).

2. *Пусть A – «множество студентов факультета международных отношений* (*ФМО*) *Белорусского государственного университета* (*БГУ*)», *B – «множество людей, которые имеют гражданство стран СНГ». Описать словами множества*: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.

Решение. Дадим описания заданных множеств: $A \cup B$ – «множество, которое состоит из студентов ФМО и людей, не являющихся студентами ФМО, но имеющих гражданство стран СНГ», $A \cap B$ – «множество студентов ФМО, у которых есть гражданство стран СНГ», $A \setminus B$ – «множество студентов ФМО, у которых нет гражданства стран СНГ», $B \setminus A$ – «множество людей, которые имеют гражданство стран СНГ, но не являющихся студентами ФМО».

3. *Пусть A – «множество всех студентов, сдавших сессию на баллы не ниже* 7 (*т. е. на* 7, 8, 9, 10)*», и пусть B – «множество всех студентов, сдавших сессию на баллы не выше* 8 (*т. е. на* 4, 5, 6, 7, 8)*». Описать словами множества*: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.

Решение. Имеем, $A \cup B$ – «множество студентов, сдавших сессию, т. е. получившие оценки от 4 до 10», $A \cap B$ – «множество студентов, сдавших сессию на баллы 7 и 8», $A \setminus B$ – «множество студентов, сдавших сессию на баллы 9 и 10», $B \setminus A$ – «множество студентов, сдавших сессию на баллы 4, 5 и 6».

4. *Пусть A – «множество иностранных студентов ФМО», B – «множество студентов* 1*-го курса ФМО», **U** – «множество всех студентов ФМО». Описать словами множества*: $\overline{A} \cup B$, $\overline{A} \cap B$, $\overline{A} \setminus B$, $B \setminus \overline{A}$.

Решение. Дадим описания заданных множеств: $\overline{A} = \boldsymbol{U} \setminus A$ – «множество студентов ФМО, которые являются гражданами Республики Беларусь (РБ)», тогда $\overline{A} \cup B$ – «множество, которое состоит из студентов ФМО, являющихся гражданами РБ, и иностранных студентов 1-го курса ФМО», $\overline{A} \cap B$ – «множество студентов 1-го курса ФМО, являющихся гражданами Республики Беларусь», $\overline{A} \setminus B$ – «множество студентов ФМО, являющихся гражданами Республики Беларусь, кроме студентов 1-го курса ФМО», $B \setminus \overline{A} = B \cap A$ – «множество иностранных студентов 1-го курса ФМО».

5. *Пусть A – «множество студентов-отличников ФМО», B – «множество студентов* 1*-го курса ФМО», **U** – «множество всех студентов ФМО». Описать словами множества*: $A \cup \overline{B}$, $A \cap \overline{B}$, $A \setminus \overline{B}$, $\overline{B} \setminus A$.

Решение. Имеем, $\overline{B} = \boldsymbol{U} \setminus B$ – «множество студентов со 2-го по 5-й курсы ФМО», тогда $A \cup \overline{B}$ – «множество, которое состоит из студентов со 2-го по 5-й курсы ФМО и студентов-отличников 1-го курса ФМО», $A \cap \overline{B}$ – «множество студентов-отличников со 2-го по 5-й курсы ФМО», $A \setminus \overline{B}$ – «множество студентов-отличников 1-го курса ФМО», $\overline{B} \setminus A$ – «множество студентов со 2-го по 5-й курсы ФМО, не являющихся отличниками».

6. *Пусть A – «множество стран, которые входят в Европейский союз», B – «множество стран-членов блока НАТО», C – «множество стран, которые входят в Шенгенскую зону». Описать словами множества*: $A \cup B \cup C$ и $A \cap B \cap C$.

Решение. Множество $A \cup B \cup C$ – «множество стран, которые входят в Европейский союз или в блок НАТО, или в Шенгенскую зону». Данное множество содержит 36 элементов (государств), ему соответствует заштрихованная часть диаграммы на рис. 9. Другое множество $A \cap B \cap C$ –

«множество стран, которые одновременно входят и в Европейский союз, и в НАТО, и в Шенгенскую зону». Данному множеству соответствует заштрихованная часть диаграммы на рис. 5. Это множество содержит 18 элементов-государств: Бельгия, Венгрия, Германия, Греция, Дания, Испания, Италия, Латвия, Литва, Люксембург, Нидерланды, Польша, Португалия, Словакия, Словения, Франция, Чехия, Эстония.

7. *Пусть A – «множество стран, которые входят в Европейский союз (ЕС)», B – «множество стран-членов блока НАТО», C – «множество стран, которые входят в Шенгенскую зону», **U** – «множество всех стран мира». Изобразить на диаграмме Эйлера – Венна и описать словами множества*: **а)** $A \cup \left(B \cap \overline{C}\right)$, **б)** $A \setminus (B \cap C)$, **в)** $\overline{A} \cap \left(B \setminus \overline{C}\right)$.

Решение. **а)** Множеству $A \cup \left(B \cap \overline{C}\right)$ соответствует заштрихованная часть диаграммы на рис. 27.

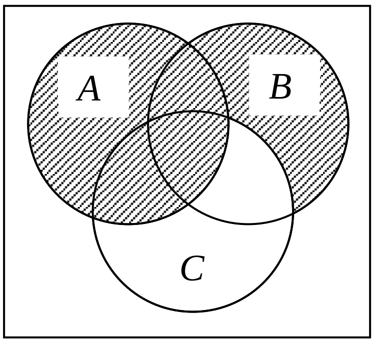


Рис. 27

Тогда $A \cup \left(B \cap \overline{C}\right)$ – «множество, которое состоит из всех стран Европейского союза, а также тех стран-членов блока НАТО, которые не входят в Шенгенскую зону». Данное множество состоит из 32 государств.

б) Множеству $A \setminus (B \cap C)$ соответствует заштрихованная часть диаграммы на рис. 28.

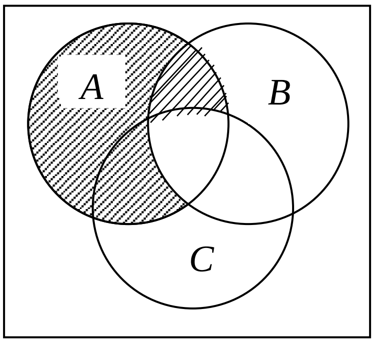


Рис. 28

Тогда $A \setminus (B \cap C)$ – «множество таких стран Европейского союза, которые не являются одновременно членами НАТО и Шенгенской зоны». Это следующие государства: Австрия, Болгария, Великобритания, Ирландия, Кипр, Мальта, Румыния, Финляндия, Швеция.

в) Множеству $\overline{A} \cap \left(B \setminus \overline{C}\right)$ соответствует заштрихованная часть диаграммы на рис. 29.

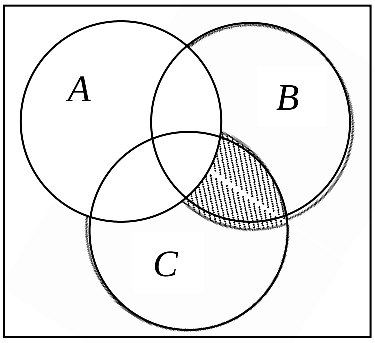


Рис. 29

Тогда $\overline{A} \cap \left(B \setminus \overline{C}\right)$ – «множество стран, которые одновременно являются членами НАТО и Шенгенской зоны, но не входят в Европейский союз». Это Исландия и Норвегия.

8. *В туристической группе из* 30 *человек* 15 *человек владеют английским языком,* 11 *– немецким, трое владеют обоими языками. Сколько человек в группе не владеют ни одним из этих языков?*

Решение. Пусть $\boldsymbol{U}$ – «множество туристов группы», A – «множество туристов, владеющих английским языком», B – «множество туристов, владеющих немецким языком». Найдем количество туристов, владеющих английским или немецким языком:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B) = 15 + 11 - 3 = 23.$$

Тогда количество человек в группе, не владеющих ни одним из этих языков, равно $n\left(\overline{A \cup B}\right) = 30 - 23 = 7.$

9. *В группе студентов из* 35 *человек* 18 *студентов посещали спецкурс «ООН и современный мир», а* 13 *человек – спецкурс «ЕС в современном мире». Сколько студентов посещали оба эти спецкурса, если* 7 *человек группы посещали другие спецкурсы?*

Решение. Пусть A – «множество студентов, которые посещали спецкурс «ООН и современный мир», B – «множество студентов, которые посещали спецкурс «ЕС в современном мире». Тогда количество студентов, которые посещали хотя бы один из этих спецкурсов $n(A \cup B) = 35 - 7 = 28.$

Далее получаем:

$$n(A \cup B) = 28 = n(A) + n(B) - n(A \cap B) = 18 + 13 - n(A \cap B).$$

Отсюда, $n(A \cap B) = 3,$ т. е. 3 студента посещали оба спецкурса.

10. *Министерство образования направило в гуманитарный лицей инспектора для проверки качества преподавания иностранных языков. Он представил следующий отчет. «В лицее* 100 *учащихся, каждый из которых изучает, по крайней мере, один из трех иностранных языков: английский, немецкий или французский. Причем все три языка изучают* 5 *человек, английский и немецкий –* 20*, английский и французский –* 8*, немецкий и французский –* 10*,* *английский* – 50 *человек, немецкий* – 33*, французский* – 30*». Инспектор, представивший этот отчет, был уволен. Почему?*

Решение. Пусть A – «множество учащихся лицея, изучающих английский язык», B – «множество учащихся лицея, изучающих немецкий язык», C – «множество учащихся лицея, изучающих французский язык», а универсальное множество $\boldsymbol{U}$ – «множество всех учащихся лицея». По условию задачи имеем: $n(\boldsymbol{U}) = 100$, $n(A) = 50$, $n(B) = 33$, $n(C) = 30$ $n(A \cap B) = 20$, $n(A \cap C) = 8$, $n(B \cap C) = 10$, $n(A \cap B \cap C) = 5$. Тогда по формуле для числа элементов $n(A \cup B \cup C)$ получим:

$n(A \cup B \cup C) = 50 + 33 + 30 - 20 - 8 - 10 + 5 = 80$.

Поскольку в отчете инспектора сказано, что *каждый* из 100 учащихся изучает хотя бы один из трех иностранных языков, то должно выполняться равенство: $n(\boldsymbol{U}) = n(A \cup B \cup C)$. Получили противоречие, так как $100 \neq 80$. Значит, инспектор был уволен, так как предоставленные им сведения оказались противоречивы.

11. *В студенческой группе* 40 *человек, и каждый изучает два иностранных языка из следующих трех языков: немецкий, английский итальянский.* 25 *студентов посещают занятия по английскому языку, и столько же – по немецкому языку. Сколько человек посещают занятия по итальянскому языку?*

Решение. 1-*й способ*. Пусть A – «множество студентов, изучающих немецкий язык», B – «множество студентов, изучающих английский язык», C – «множество студентов, изучающих итальянский язык». Здесь рисунок с тремя кругами не отражает условия задачи, так как нет элементов, принадлежащих всем трем множествам, и нет элементов, принадлежащих только одному множеству. Каждое из множеств A, B и C представим как две трети круга (рис. 30). Три таких множества будут накрывать круг дважды.

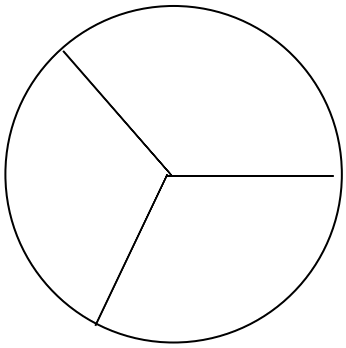


Рис. 30

Значит, верно равенство:

$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cup B \cup C)$,

из которого получаем, что $2 \cdot n(A \cup B \cup C) = n(A) + n(B) + n(C)$. Тогда $2 \cdot 40 = 25 + 25 + n(C)$, откуда $n(C) = 30$. Таким образом, занятия по итальянскому языку посещают 30 человек.

2-*й способ*. Так как каждый из 40 студентов изучает ровно два иностранных языка, то $2 \cdot 40 = 80$ – общее число посещений занятий по иностранным языкам. Тогда $80 - (25 + 25) = 30$ человек изучают итальянский язык.

Задачи для самостоятельного решения

1. *Пусть А – «множество иностранных студентов первого курса ИФ», В – «множество студентов 1-й группы первого курса ИФ». Описать словами множества:* $A \cup B$, $A \cap B$, $A \setminus B$.

2. *Пусть А – «множество студентов-отличников ИФ», В – «множество студентов 1-го курса ИФ»,* **U** *– «множество всех студентов ИФ». Описать словами множества:* $A \cup B$, $A \cap B$ $A \setminus B$, $\overline{A}$, $\overline{B}$.

3. *Пусть А – «множество всех студентов, сдавших сессию на баллы не ниже 8 (т. е. на 8, 9, 10)», и пусть В – «множество всех студентов, сдавших сессию на баллы не выше 9 (т. е. на 4, 5, 6, 7, 8, 9)». Описать словами множества*: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.

4. *Пусть А – «множество стран, которые входят в Европейский союз», В – «множество стран-членов блока НАТО», С – «множество стран, которые входят в Шенгенскую зону»,* **U** *– «множество всех стран мира». Изобразить на диаграмме Эйлера – Венна и описать словами множества:* **а)** $A \cap \left(\overline{B} \cap \overline{C}\right)$, **б)** $\overline{A} \cap (B \cup C)$.

5. *В студенческой группе* 20 *человек изучают английский,* 18 *человек – французский язык. При этом* 5 *человек изучают оба эти языка, а* 7 *человек вообще не изучают иностранный язык. Сколько студентов в группе?*

6. *В туристической группе из* 27 *человек* 9 *туристов владеют английским языком,* 10 *– французским, двое владеют обоими языками. Сколько человек в группе не владеют ни одним из этих языков? Сколько туристов владеют только английским языком?*

7. *В туристической группе из* 40 *человек* 26 *туристов владеют английским языком,* 25 *– немецким,* 27 *– французским. При этом* 15 *человек знают немецкий и английский,* 16 *человек – английский и французский,* 18 *– немецкий и французский. Один человек не владеет ни одним из этих языков. Сколько человек в группе владеют всеми тремя этими языками? Сколько человек владеют только одним иностранным языком?*

8. *В группе из* 30 *студентов* 15 *любят петь,* 17 *– танцевать,* 11 *– заниматься спортом. Известно, кроме того, что* 5 *спортсменов любят петь,* 7 *– танцевать,* 8 *певцов любят танцевать, а трое – все эти три вида досуга. Сколько студентов не любят ни один из указанных способов проведения свободного времени? Сколько студентов танцуют и поют, но не занимаются спортом?*

Занятие № 3. Основные принципы комбинаторики. Перестановки, размещения и сочетания.

Примеры решения задач

1. *Пусть в магазине имеются 7 различных видов коробок конфет и 5 различных коробок печенья. Сколькими способами можно выбрать в подарок коробку конфет или коробку печенья? Сколькими способами можно составить набор, состоящий из коробки конфет и коробки печенья?*

Решение. Коробку конфет или коробки печенья можно выбрать согласно правилу суммы 7+5=12 способами. Составить набор из коробки конфет и коробки печенья можно согласно правилу произведения $7 \cdot 5=35$ способами.

2. *Сколькими способами можно выбрать гласную и согласную буквы из слова МОСКВА?*

Решение. Гласную букву можно выбрать двумя способами (О или А), а согласную – существует четыре варианта (М, С, К, В). Следовательно, согласно *правилу произведения* гласную и согласную буквы можно выбрать $2 \cdot 4=8$ способами.

3. *Сколькими способами 6 студенток могут встать в столовой в очередь?*

Решение. Необходимо пересчитать способы расположения семи девушек в последовательность, т.е. речь идет о всех перестановках из n элементов, где n=6. Число таких перестановок равно $P_n = n! = 6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$.

4. *Из группы студентов, состоящей из 25 человек, играющих в шахматы, надо выбрать шахматную команду из четырех человек играющих на I, II, III и IV доске. Сколькими способами это можно сделать?*

Решение. Так как из 25 человек выбираются четверо и порядок важен при распределении их по доскам, то число способов есть число размещений из 25 по 4, т.е.:

$$A_{25}^{4} = 25!/(25-4)!=25!/21!=22 \cdot 23 \cdot 24 \cdot 25=303600 \text{ способов.}$$

5. *Сколько трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5, 6, 7, если цифры в числах не повторяются?*

Решение. Выбираем 3 цифры из 7, порядок важен: $A_7^3 = 7!/4! = 5 \cdot 6 \cdot 7= 210$ способов.

6. *Сколькими способами можно составить трехцветный флаг с полосами одинаковой ширины, если есть материал трех различных цветов и возможно как вертикальное, так и горизонтальное расположение полос?*

Решение. При вертикальном расположении полос их можно расположить $P_3 = 3! = 6$ способами (число перестановок из трех элементов), и такое же количество способов – для горизонтального расположения полос. Так как на флагах возможно вертикальное или горизонтальное расположение полос, то согласно правилу суммы общее число таких флагов будет равно $6+6=12$.

7. *Для занятий по информатике группа студентов исторического факультета БГУ из 25 человек разбивается на две подгруппы А и Б по 13 и 12 человек соответственно. Сколькими способами это можно сделать?*

Решение. Выбираем сначала подгруппу А. Выбор 13 человек из 25 можно осуществить C_{25}^{13} способами. Оставшиеся 12 человек составят подгруппу Б (ее можно выбрать $C_{12}^{12}=1$ способом). Таким образом, данное разбиение можно осуществить $C_{25}^{13} = \dfrac{25!}{13! \cdot 12!} = 5\,200\,300$ способами.

Задачи для самостоятельного решения

1. *Сколькими способами 10 студентов могут встать в очередь друг за другом в университетской библиотеке?*

2. *Студенты изучают в каждом семестре восемь различных дисциплин. Расписание занятий на понедельник состоит из 4 различных дисциплин. Сколько различных расписаний на понедельник может составить методист факультета?*

3. *Сколько поединков по борьбе должны быть проведены между 15 спортсменами, если каждый из них должен встретиться с каждым?*

4. *В книжный магазин привезли новых 20 книг, из них по философии − 5, по социологии – 8, по высшей математике − 4 и по психологии − 3. Сколькими способами можно составить набор для университетской библиотеке, чтобы в него входило 3 книги по философии, 5 по социологии, 3 по высшей математике и 2 по психологии?*

5. *Сколькими способами 3 награды (за I, II, III места) могут быть распределены между 10 участниками математической олимпиады?*

6. *В конкурсе участвовало 8 фирм, трем из которых жюри должно присудить 1-е, 2-е и 3-е места. Сколько вариантов решения жюри существует?*

7. *Перед выпуском группа студентов, состоящая из 19 человек, обменялась фотографиями. Сколько всего фотографий было роздано?*

8. *Сколько трёхзначных чисел можно составить из цифр 1,2,3, если каждая цифра входит в запись числа только один раз?*

9. *Собрание из 35 человек выбирает председателя, секретаря и помощника. Сколько существует способов это сделать?*

10. *Вычислить* C_{11}^{3} , A_{7}^{4}.

11. *Предположим, что в записи 7-значного номера телефона используются только две цифры: 2 и 5. Сколько может быть таких различных номеров телефонов?*

12. *Из девяти различных задач для контрольной работы выбирается пять. Сколько различных вариантов контрольной работы можно составить?*

13. *Сколько пятизначных чисел можно составить из цифр 1, 2, 3, 4, 5, 6 так, чтобы каждое из них начиналось с комбинации «45»?*

14. *Сколько среди трёхзначных чисел встречается таких, в записи которых не участвует число 6?*

15. *5 групп занимаются в 5-ти расположенных подряд аудиториях. Сколько существует вариантов расписания, при которых группы №1 и №2 находились бы в соседних аудиториях?*

16. *В стране 20 городов, каждые 2 из которых соединены авиалиниями. Сколько авиалиний в этой стране?*

17. *Сколькими способами можно разбить 10 человек на 2 баскетбольные команды по 5 человек в каждой?*

Занятие № 4. Операции над событиями и их свойства. Вероятности случайных событий. Условные вероятности. Независимость событий. Теоремы сложения и умножения вероятностей.

Примеры решения задач

1. *Из 30 экзаменационных билетов студент исторического факультета знает материал 20 билетов. Какова вероятность того, что студент вытянет билет, который знает, если он отвечает первым?*

Решение. Общее число исходов данного испытания $n = 30,$ так как студент отвечает первым и может вытянуть любой из 30 билетов, а исходов, благоприятствующих событию *A* – «студент вытянул билет, материал которого знает», $m = 20.$ Тогда по формуле классической вероятности получим

$$P(A) = \frac{m}{n} = \frac{20}{30} = \frac{2}{3}.$$

2. *На Евразийском саммите присутствовали восемь представителей стран Восточной Европы и 10 представителей стран Западной Европы. Найти вероятность того, что среди шести случайно отобранных делегатов саммита трое окажутся представителями стран Восточной Европы.*

Решение. Пусть событие *A* – «среди шести отобранных делегатов саммита трое окажутся представителями стран Восточной Европы». Исход этого испытания состоит в выборе любых шести делегатов из 18 делегатов саммита. Число всех исходов $n = C_{18}^{6}.$ Заметим, что из шести выбранных делегатов трое должны представлять Восточную Европу и трое – Западную Европу. Выбрать трех из восьми стран Восточной Европы можно C_8^3 способами, а трех из десяти стран Западной Европы – C_{10}^3 способами. Согласно правилу произведения число благоприятствующих исходов для события *A* равно $m = C_8^3 \cdot C_{10}^3.$ Следовательно, по формуле классической вероятности

$$P(A) = \frac{m}{n} = \frac{C_8^3 \cdot C_{10}^3}{C_{18}^6} = \frac{8!}{3!\cdot 5!} \cdot \frac{10!}{3!\cdot 7!} \cdot \frac{6!\cdot 12!}{18!} =$$

$$= \frac{8\cdot 7\cdot 6}{2\cdot 3} \cdot \frac{10\cdot 9\cdot 8}{2\cdot 3} = \frac{2\cdot 3\cdot 4\cdot 5\cdot 6}{18\cdot 17\cdot 16\cdot 15\cdot 14\cdot 13} = \frac{10\cdot 8}{13\cdot 17} = \frac{80}{221}.$$

3. *В ЕС входят 27 государств, из них три страны Балтии (Литва, Латвия, Эстония). Найти вероятность того, что в случайным образом сформированную комиссию из пяти государств войдет одна из стран Балтии.*

Решение. Пусть событие *A* – «в комиссию из пяти государств войдет одна из стран Балтии». Общее число исходов испытания равно количеству способов выбора пяти стран из 27 стран ЕС, т. е. $n = C_{27}^5,$ а $m = C_3^1 \cdot C_{24}^4$ – число благоприятствующих исходов для события *A*, так как в комиссии должна быть одна из стран Балтии и четыре страны, не являющиеся балтийскими. Следовательно, по формуле классической вероятности получим

$$P(A)=\frac{m}{n}=\frac{C_3^1\cdot C_{24}^4}{C_{27}^5}=\frac{3!}{1!\cdot 2!}\cdot\frac{24!}{4!\cdot 20!}\cdot\frac{5!\cdot 22!}{27!}=$$

$$=3\cdot\frac{24\cdot 23\cdot 22\cdot 21}{2\cdot 3\cdot 4}\cdot\frac{2\cdot 3\cdot 4\cdot 5}{17\cdot 26\cdot 25\cdot 24\cdot 23}=\frac{11\cdot 7}{3\cdot 5\cdot 13}=\frac{77}{195}.$$

4. *Брошены 2 игральные кости. Найдите вероятность следующих событий: а) сумма выпавших очков равна 7; б) сумма выпавших очков равна 8, а разность 4.*

Решение. Для случая **а)** n = 36, так как число равновозможных исходов испытания состоит из числа различных комбинаций очков на одной и второй игральных костях. Для первой кости возможны 6 вариантов от 1 до 6, для второй аналогично 6 вариантов, следовательно, для первой и второй костей, согласно комбинаторному принципу умножения будет 36 исходов. Число благоприятных исходов m, т. е. сумма выпавших очков равна 7, состоит из следующих вариантов: (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3), т. е. m = 6, и поэтому

$$P(A)=\frac{m}{n}=\frac{6}{36}=\frac{1}{6}.$$

Для случая **б)** как и в случае **а)** n = 36. Число благоприятных исходов m, т. е. сумма выпавших очков равна 8, а разность 4, состоит из следующих двух вариантов: (2, 6), (6, 2), т. е. m = 2, и поэтому

$$P(A)=\frac{m}{n}=\frac{2}{36}=\frac{1}{18}.$$

5. *В урне 15 голубых, 10 зеленых и 25 белых шаров. Найдите вероятность того, что из урны наугад будет извлечен цветной шар.*

Решение. Извлечение цветного шара означает появление либо голубого, либо зеленого шара. Пусть событие A означает «появление голубого шара», B – «появление зеленого шара». Тогда

$$P(A)=\frac{15}{15+10+25}=\frac{15}{50}=\frac{3}{10},\quad P(B)=\frac{10}{15+10+25}=\frac{10}{50}=\frac{1}{5}.$$

Так как события A и B несовместны, то по теореме сложения вероятностей двух несовместных событий имеем:

$$P(A+B)=P(A)+P(B)=\frac{3}{10}+\frac{1}{5}=\frac{1}{2}.$$

6. *Тридцать экзаменационных билетов по курсу «Основы высшей математики» пронумерованы числами от 1 до 30. Билеты тщательно перемешаны. Какова вероятность вытянуть наудачу студенту билет с номером, кратным 2 или 3?*

Решение. Обозначим события: A – «вытянут билет с четным номером», событие B – «вытянут билет с номером, кратным 3», событие AB – «вытянут билет с четным номером, кратным 3». Найдем вероятность искомого события $A+B$. Поскольку события A и B – это совместные события, то вероятность события $A+B$ находим по теореме сложения вероятностей двух событий $P(A+B) = P(A) + P(B) - P(AB)$.

Событию A благоприятствуют 15 исходов, событию B – всего 10 исходов, а событию AB – только 5 исходов. Поэтому искомая вероятность равна

$$P(A+B)=P(A)+P(B)-P(AB) = \frac{15}{30}+\frac{10}{30}-\frac{5}{30}=\frac{20}{30}=\frac{2}{3}=0{,}667.$$

7. *Из урны, в которой 10 черных и 5 белых шаров вынимают 2 шара. Чему равна вероятность того, что оба шара черные?*

Решение. Пусть событие A_1 – «первый шар черный», A_2 – «второй шар черный»

$$P(A_1)=\frac{10}{10+5}=\frac{2}{3}.$$

Найдем условную вероятность $P(A_2|A_1)=\frac{10-1}{15-1}=\frac{9}{14}$. (второй шар был черным, если первый был черным). Отсюда находим искомую вероятность:

$P(A_1A_2)=P(A_1)P(A_2|A_1)=\frac{9}{14}\cdot\frac{2}{3}=\frac{3}{7}$;

8. *Вероятность попадания каждого из трех стрелков соответственно равна: 0,8; 0,7; 0,9. Стрелки произвели один залп. Найдите вероятность: а) только одного попадания; б) ровно двух попаданий; в) хотя бы одного попадания.*

Решение. Пусть A_i – «попадание в мишень при i-м выстреле», $\overline{A_i}$ – «непопадание в мишень при i-м выстреле» $i=1,2,3$.

Т.к. $P(A_1)=0{,}8$; $P(A_2)=0{,}7$; $P(A_3)=0{,}9$, то $P(\overline{A}_1)=1-0{,}8=0{,}2$; $P(\overline{A}_2)=1-0{,}7=0{,}3$; $P(\overline{A}_3)=1-0{,}9=0{,}1$.

а) Обозначим событие A – «ровно одно попадание в мишень». Оно распадается на три несовместных события: может быть попадание при первом выстреле, промахи при втором и третьем; попадание при втором, промахи при первом и третьем; попадание при третьем, промахи при первом и втором. Тогда

$$A=A_1\overline{A}_2\overline{A}_3+\overline{A}_1A_2\overline{A}_3+\overline{A}_1\overline{A}_2A_3.$$

Учитывая, что события $\overline{A}_1A_2\overline{A}_3$; $\overline{A}_1\overline{A}_2A_3$ – несовместные события (стрелки производят выстрелы независимо друг от друга), то из теорем сложения и умножения вероятностей следует:

$$P(A)=P(A_1)P(\overline{A}_2)P(\overline{A}_3)+P(\overline{A}_1)P(A_2)P(\overline{A}_3)+P(\overline{A}_1)P(\overline{A}_2)P(A_3)=$$
$$=0{,}8\cdot0{,}3\cdot0{,}1+0{,}2\cdot0{,}7\cdot0{,}1+0{,}2\cdot0{,}3\cdot0{,}9=0{,}092.$$

б) Обозначим событие B – «ровно два попадания в мишень». Оно распадается на три несовместных события: может быть попадание при первом и втором выстрелах, промах при третьем; попадание при первом и третьем, промах при втором; попадание при втором и третьем, промах при первом. Тогда

$$B=A_1A_2\overline{A_3}+A_1\overline{A_2}A_3+\overline{A_1}A_2A_3$$

Учитывая, что $A_1A_2\overline{A_3}$, $A_1\overline{A_2}A_3$, $\overline{A_1}A_2A_3$ – несовместные события (стрелки производят выстрелы независимо друг от друга), то из теорем сложения и умножения вероятностей следует

$$P(B)=P(A_1A_2\overline{A_3}+A_1\overline{A_2}A_3+\overline{A_1}A_2A_3)$$
$$=P(A_1)P(A_2)P(\overline{A}_3)+P(A_1)P(\overline{A}_2)P(A_3)+P(\overline{A}_1)P(A_2)P(A_3)=$$
$$=0{,}8\cdot 0{,}7\cdot 0{,}1+0{,}8\cdot 0{,}3\cdot 0{,}9+0{,}2\cdot 0{,}7\cdot 0{,}9=0{,}398.$$

в) Обозначим событие C – «ровно три попадания в мишень»; событие D – «хотя бы одно попадание в мишень», тогда $D=A+B+C$.

Найдем вероятность события $C=A_1A_2A_3$, $P(C)=P(A_1)P(A_2)P(A_3)=0{,}8\cdot 0{,}7\cdot 0{,}9=0{,}504$.

Получаем
$$P(D)=P(A+B+C)=P(A)+P(B)+P(C)=0{,}092+0{,}398+0{,}504=0{,}994.$$

Так как попадания каждого стрелка независимые в совокупности события, то вероятность события D можем найти:
$$P(D)=1-P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})=1-0{,}2\cdot 0{,}3\cdot 0{,}1=0{,}994.$$

9. *Студент–заочник знает ответы на 10 тестовых вопросов из 15 вопросов. Пусть они для него будут «счастливые». Предположим, что четыре вопроса задаются лектором последовательно один за другим. Найдите вероятность того, что четыре подряд заданных вопроса – «счастливые».*

Решение. Введем обозначения событий: A_1 – «первый вопрос счастливый», A_2 – «второй вопрос счастливый», A_3 – «третий вопрос счастливый», A_4 – «четвертый вопрос счастливый». По теореме умножения вероятностей n событий искомая вероятность равна
$$P(A_1A_2A_3A_4)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3)=$$
$$=\frac{10}{15}\cdot\frac{9}{14}\cdot\frac{8}{13}\cdot\frac{7}{12}=\frac{2}{13}\approx 0{,}154.$$

Задачи для самостоятельного решения

1. *Набирая номер телефона, абонент забыл последние две цифры, помня лишь, что эти цифры различны, набрал их наудачу. Найдите вероятность того, что набраны нужные цифры.*

2. *Игральную кость бросают один раз. Найти вероятность того, что число выпавших очков не меньше пяти.*

3. *Подбрасываются два игральных кубика, подсчитывается сумма очков на верхних гранях. Что вероятнее – получить в сумме 7 или 8 очков?*

4. *Подбрасываются 3 симметричные монеты. Чему равна вероятность того, что на верхних сторонах монет выпало 2 герба?*

5. *В урне имеется 7 черных и несколько белых шаров. Какова вероятность вытащить белый шар, если вероятность вытащить черный шар равна 1/6. Сколько белых шаров в урне?*

6. *Из 30 экзаменационных билетов по дисциплине «Социология» студент может ответить на 24 билета. Какова вероятность его успешного ответа на экзамене при однократном извлечении билета?*

7. *Из 20 изделий первого сорта и 10 второго сорта, имеющихся на складе наугад взято 2 изделия. Найдите вероятность того, что оба эти изделия:* **а)** первого сорта; **б)** второго сорта.

8. *На стеллажах библиотеки в случайном порядке расставлено 15 учебников, причем 5 из них в переплете. Библиотекарь наудачу берет три учебника. Найдите вероятность того, что все они будут в переплете.*

9. *Вероятность того, что потребитель увидит рекламу определенного продукта по каждому из 3 центральных телеканалов, равна 0,05. Предполагается, что эти события – независимы в совокупности. Чему равна вероятность того, что потребитель увидит рекламу:* **а)** по всем трем каналам; **б)** по одному каналу?

10. *Покупатель может приобрести акции 2 компаний А и В. Надежность 1-й оценивается экспертами на уровне 90%, а второй – 80%. Чему равна вероятность того, что обе компании в течении года не станут банкротами.*

11. *В студенческой группе из 20 человек 5 отличников. Наудачу по списку выбирают двух человек. Чему равна вероятность того, что:*

a) оба студента отличники;

b) оба студента не отличники.

12. *В городе имеется 3 коммерческих банка. Вероятности того, что банки обанкротятся в течение года соответственно равны 0,1; 0,2; 0,05. Найдите вероятности того, что в течение года обанкротятся*

a) ровно два банка;

b) ровно один банк;

c) все три банка;

d) ни одного банка;

e) хотя бы один банк.

13. *Симметричную игральную кость бросают 3 раза. Найти вероятность того, что хотя бы один раз появится шестерка.*

14. *Из группы студентов 20 человек выбираются наугад трое. Найти вероятность, что среди них обязательно будет староста группы.*

15. *Студент знает только 10 из 50 билетов, выносимых на экзамен. Когда шансы получить знакомый билет выше: когда он пойдет 1-й, 2-й или 3-й?*

16. *Из колоды карт (36 шт.) последовательно вынимаются 4 карты. Найти вероятность того, что это будут карты* ***а*****)** *четырех разных мастей;* ***б*****)** *одной и той же масти.*

17. *В лотерее разыгрывается 100 билетов, из которых 10 выигрышные. студент покупает 4 билета. Какова вероятность того, что хотя бы 1 из них выиграет?*

18. *На карточках написаны буквы слова «Миссисипи». Карточки перемешали и выложили в ряд одну за другой. Какова вероятность, что получим то же самое слово?*

19. *В продаже имеется 27 белых роз и 12 красных. Продавец наугад вынул 9 цветов. Какова вероятность, что в полученном букете будет пять красных роз?*

20. *Имеется хорошо перетасованная колода из 36 карт. Найти вероятность того, что первые 4 карты в колоде – тузы.*

Занятие № 5 Формула полной вероятности. Повторение испытаний. Схема Бернулли.

Примеры решения задач

1. *На фабрике на машинах a, b, c производят соответственно 25, 35 и 40% всех изделий. В их продукции брак составляет соответственно 5%, 4% и 2%. Какова вероятность того, что случайно выбранное изделие фабрики дефектно?*

Решение. Пусть событие A состоит в том, что случайно выбранное изделие дефектно, и пусть H_1, H_2, H_3 – события, состоящие в том, что изделие произведено на машинах a, b, c соответственно. Очевидно, события H_1, H_2, H_3 образуют полную группу событий.

По условию $P(H_1)=0{,}25$; $P(H_2)=0{,}35$; $P(H_3)=0{,}40$. Аналогично, $P(A|H_1)=0{,}05$; $P(A|H_2)=0{,}04$; $P(A|H_3)=0{,}02$ являются условными вероятностями события A при выполнении гипотез H_1, H_2, H_3 соответственно.

Применив формулу полной вероятности, найдем:

$$P(A)=P(H_1)P(A|H_1)+P(H_2)P(A|H_2)+P(H_3)P(A|H_3)=0{,}25\cdot 0{,}05+0{,}35\cdot 0{,}04+0{,}40\cdot 0{,}02=0{,}0345.$$

2. *Пусть выполнены условия задачи 1 и пусть известно, что случайно выбранное изделие оказалось дефектным. Какова вероятность того, что оно было сделано на машинах a, b, c соответственно?*

Решение. Пусть A, H_1, H_2, H_3 означает то же, что и в задаче **1**. Тогда задача состоит в нахождении условных вероятностей гипотез H_1, H_2, H_3 при условии, что событие A уже произошло. По условию $P(H_1)=0{,}25$; $P(H_2)=0{,}35$; $P(H_3)=0{,}40$. Аналогично, $P(A|H_1)=0{,}05$; $P(A|H_2)=0{,}04$; $P(A|H_3)=0{,}02$.

В задаче **1** найдено:

$P(H_1)P(A|H_1)+P(H_2)P(A|H_2)+P(H_3)P(A|H_3)=0{,}0345$.

Применяя формулу Байеса, имеем

$$P(H_1|A)=\frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1)+P(H_2)P(A|H_2)+P(H_3)P(A|H_3)}=\frac{0{,}25\cdot 0{,}05}{0{,}0345}=\frac{25}{69}.$$

Аналогично получаем $P(H_2|A)=\frac{28}{69}$, $P(H_3|A)=\frac{16}{69}$.

3. *Социолог проводил исследование психологического климата в разных отделах фирмы. При этом было установлено, что мужчины и женщины по-разному реагируют на некоторые жизненные обстоятельства. Результаты исследования показали, что 68 % женщин позитивно реагируют на эти ситуации, в то время как 37 % мужчин реагируют на них негативно. 15 женщин и 5 мужчин заполнили анкету, в которой отразили свое отношение к предлагаемым ситуациям. Случайно извлеченная анкета содержит негативную реакцию. Чему равна вероятность, что ее заполнял мужчина?*

Решение. Обозначим через А событие, состоящее в том, что «случайно извлеченная анкета будет содержать негативную реакцию», а через Н1, – событие, состоящее в том, что «анкету заполнял мужчина», через Н2, – событие, состоящее в том, что «анкету заполняла женщина». Из условия задачи следует, что P(A|H1)=0,37; P(A|H2)=1–0,68=0,32.

Далее, $P(H_1)=\frac{5}{15+5}=\frac{5}{20}=\frac{1}{4}$, $P(H_2)=\frac{15}{15+5}=\frac{15}{20}=\frac{3}{4}$.

По формуле полной вероятности вероятность того, что случайно извлеченная анкета будет содержать негативную реакцию

$$P(A)=P(H_1)P(A|H_1)+P(H_2)P(A|H_2)=0{,}37\cdot\frac{1}{4}+0{,}32\cdot\frac{3}{4}=0{,}093+0{,}24=0{,}333.$$

Случайно извлеченная анкета содержит негативную реакцию. Вероятность того, что ее заполнял мужчина, найдём по формуле Байеса

$$P(H_1|A)=\frac{P(H_1)P(A\,|\,H_1)}{P(H_1)P(A\,|\,H_1)+P(H_2)P(A\,|\,H_2)}=0{,}278.$$

4. *В результате наблюдений на некоторой реке установлено, что в течение 20 лет имело место четыре случая пересыхания реки. Найти вероятность того, что за двадцатилетний период будет наблюдаться:*

а) от трех до пяти случаев пересыхания реки (включительно);

б) не более пяти случаев пересыхания реки;

в) более пяти случаев пересыхания реки.

Решение. а) Вероятность того, что за 20 лет произойдет m случаев пересыхания реки $(0\le m\le 20)$ равна $P_{20}(m)=C_{20}^m p^m q^{20-m}$, где $p=\frac{4}{20}=0{,}2$, $q=1-p=1-0{,}2=0{,}8$. Тогда искомая вероятность для а) составит

$$P_{20}(3)+P_{20}(4)+P_{20}(5)=$$
$$=C_{20}^3\cdot 0{,}2^3\cdot 0{,}8^{17}+C_{20}^4\cdot 0{,}2^4\cdot 0{,}8^{16}+C_{20}^5\cdot 0{,}2^5\cdot 0{,}8^{15}\approx$$
$$\approx 0{,}2054+0{,}2182+0{,}1746=0{,}5982.$$

б)

$$P(m\le 5)=\sum_{m=0}^{5}P_{20}(m)=C_{20}^0\cdot 0{,}2^0\cdot 0{,}8^{20}+C_{20}^1\cdot 0{,}2^1\cdot 0{,}8^{19}+$$
$$+C_{20}^2\cdot 0{,}2^2\cdot 0{,}8^{18}+C_{20}^3\cdot 0{,}2^3\cdot 0{,}8^{17}+C_{20}^4\cdot 0{,}2^4\cdot 0{,}8^{16}+C_{20}^5\cdot 0{,}2^5\cdot 0{,}8^{15}\approx$$
$$\approx 0{,}0115+0{,}0576+0{,}1369+0{,}2054+0{,}2182+0{,}1746=0{,}8042.$$

в)

$$P(m>5)=1-P(m\le 5)=1-0{,}8042=0{,}1958.$$

Задачи для самостоятельного решения

1. *Судоходная компания организует средиземноморские круизы в течение летнего времени и проводит несколько круизов в сезон. Эксперт по туризму, нанятый компанией, предсказывает, что вероятность того, что корабль будет полон в течение сезона, будет равна 0,92, если доллар не подорожает по отношению к рублю, и с вероятностью – 0,75, если доллар подорожает. По оценкам экономистов, вероятность того, что в течение сезона доллар* *подорожает по отношению к рублю, равна 0,23. Чему равна вероятность того, что билеты на все круизы будут проданы?*

2. *Сотрудники отдела маркетинга полагают, что в ближайшее время ожидается рост спроса на продукцию фирмы. Вероятность этого они оценивают в 0,8. Консультационная фирма, занимающаяся прогнозом рыночной ситуации, подтвердила предположение о росте спроса. Положительные прогнозы консультационной фирмы сбываются с вероятностью 0,95, а отрицательные – с вероятностью 0,99. Какова вероятность того, что рост спроса действительно произойдет?*

3. *Среди студентов института – 30% первокурсники, 35% студентов учатся на 2-м курсе, на 3-м и 4-м курсе их 20% и 15% соответственно. По данным деканата известно, что на первом курсе 20% студентов сдали сессию только на отличные оценки, на 2-м курсе – 30%, на 3-м курсе – 35%, на 4-м курсе – 40% отличников. Наудачу вызванный студент оказался отличником. Чему равна вероятность того, что он (или она) – третьекурсник?*

4. *Из числа авиалиний некоторого аэропорта 60% – местные авиалинии, 30% – авиалинии по СНГ и 10% – международные авиалинии. Среди пассажиров местных авиалиний 50% путешествуют по делам, связанным с бизнесом, на линиях СНГ таких пассажиров 60%, на международных – 90%. Из прибывших в аэропорт пассажиров случайно выбирается один пассажир. Чему равна вероятность того, что он: а) бизнесмен; б) прибыл из стран СНГ по делам бизнеса; в) прилетел местным рейсом по делам бизнеса; г) прибывший международным рейсом бизнесмен?*

5. *По статистике турист, прибывший накануне на автобусе, поезде или самолете, заказывает утренний экскурсионный тур с вероятностями 0,2, 0,4 и 0,7 соответственно. Найдите вероятность того, что прибывший в отель турист закажет тур, если автобусом приезжает 55%, поездом – 30%, а самолетом – 15% постояльцев.*

6. *На горнолыжном курорте фуникулер оборудован тремя независимыми предохранительными системами с надежностью срабатывания 0,95, 0,93 и 0,9 соответственно. Известно, что в результате перепада напряжения в сети, сработала одна из них. Найдите вероятность того, что не сработала третья система.*

7. *В санатории 30% пациентов – мужчины и 70% – женщины. Сердечные болезни среди мужчин встречаются в два раза чаще, чем среди женщин. Какова вероятность, что наугад выбранный пациент сердечник?*

8. *Вероятность встретить реку, загрязняемую постоянным фактором, равна 0,4, временным фактором – 0,1 и обоими факторами – 0,05. Найти:*

а) *вероятность того, что река, загрязняемая временным фактором, будет к тому же загрязнена и постоянным фактором;*

б) *вероятность того, что река, загрязняемая постоянным фактором, будет еще загрязнена и временным фактором.*

9. *Вероятность отказа каждого прибора при испытании не зависит от отказов остальных приборов и равна 0,2. Испытано 9 приборов. Какова* *вероятность того, что откажут: а) два прибора; б) один прибор; в) менее трех приборов?*

10. *В семье пять детей. Найти вероятность того, что в этой семье одна девочка и четыре мальчика, если вероятность рождения мальчика равна 0,5.*

11. *Многолетний опыт показал, что в сентябре в данном районе 10 дней бывают дождливыми. Экспериментальное хозяйство должно в течение первых трех дней сентября выполнить некоторые работы. Определить вероятность того, что ни один из этих дней не будет дождливым.*

12. *Применяемый метод лечения приводит к выздоровлению в 80% случаев. Какова вероятность того, что из пяти больных выздоровеет не менее четырех?*

13. *Имеются 2 одинаковые с виду коробки, в первой из которых 7 белых и 3 красных шара, во второй 1 белый и 4 красных шара. Из первой коробки наугад выбранный шар перекладывается во вторую, шары в ней перемешиваются, и из нее наугад вынимается 1 шар. Определите вероятность того, что этот шар будет 1) белым; 2) красным.*

14. *Имеются 2 одинаковые с виду коробки, в первой из которых 7 белых и 3 красных шара, во второй 1 белый и 4 красных шара. Из первой коробки наугад выбранных 3 шара перекладываются во вторую, шары в ней перемешиваются, и из нее наугад вынимается 1 шар. Определите вероятность того, что этот шар будет 1) белым; 2) красным.*

Занятие № 6. Случайные величины: дискретные и непрерывные. Закон распределения дискретной случайной величины. Математическое ожидание, дисперсия, их свойства.

Примеры решения задач

1. *В ящике находятся шесть белых и четыре черных шара. Из него пять раз подряд извлекают шар, причем каждый раз вынутый шар возвращают в ящик и шары перемешивают. Приняв за случайную величину X число извлеченных белых шаров, составить закон распределения этой величины, определить ее математическое ожидание и дисперсию.*

Решение. Согласно условию задачи случайная величина X может принять значения $0, 1, 2, 3, 4, 5$. Найдем вероятности этих значений по формуле Бернулли (9.11). В данном случае $p = 0,6,\ q = 0,4,\ n = 5,$ откуда

$$P_5(0) = C_5^0 \cdot 0,6^0 \cdot 0,4^5 = 0,01024,$$

$$P_5(1) = C_5^1 \cdot 0,6^1 \cdot 0,4^4 = 5 \cdot 0,6 \cdot 0,0256 = 0,0768,$$

$$P_5(2) = C_5^2 \cdot 0,6^2 \cdot 0,4^3 = 10 \cdot 0,36 \cdot 0,064 = 0,2304,$$

$$P_5(3) = C_5^3 \cdot 0,6^3 \cdot 0,4^2 = 10 \cdot 0,216 \cdot 0,16 = 0,3456,$$

$$P_5(4) = C_5^4 \cdot 0,6^4 \cdot 0,4^1 = 5 \cdot 0,1296 \cdot 0,4 = 0,2592,$$

$$P_5(5) = C_5^5 \cdot 0,6^5 \cdot 0,4^0 = 0,07776.$$

Закон распределения случайной величины X имеет вид

X	0	1	2	3	4	5
p	0,01024	0,0768	0,2304	0,3456	0,2592	0,07776

Отсюда

$$M(X) = 0 \cdot 0,01024 + 1 \cdot 0,0768 + 2 \cdot 0,2304 + 3 \cdot 0,3456 +$$
$$+ 4 \cdot 0,2592 + 5 \cdot 0,07776 = 3,$$

$$D(X) = (0-3)^2 \cdot 0,01024 + (1-3)^2 \cdot 0,0768 + (2-3)^2 \cdot 0,2304 +$$
$$+ (3-3)^2 \cdot 0,3456 + (4-3)^2 \cdot 0,2592 + (5-3)^2 \cdot 0,07776 = 1,2.$$

2. *Подбрасываются две монеты и подсчитывается количество выпавших «орлов» на обеих верхних сторонах монет. Рассматривается дискретная случайная величина X – число выпадения «орлов» на обеих монетах. Записать закон распределения этой случайной величины.*

Решение. В данном испытании пространство элементарных событий равно Ω = {(О, О), (Р, Р), (О, Р), (Р, О)}, где О означает, что выпал «орел», Р – выпала «решка». «Орел» может выпасть 1 раз, 2 раза или не появиться ни разу. Следовательно, случайная величина X может принимать только три значения: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$. Вероятности этих значений равны соответственно

$$p_1 = P(X = 0) = \frac{1}{4}, \quad p_2 = P(X = 1) = \frac{2}{4} = \frac{1}{2}, \quad p_3 = P(X = 2) = \frac{1}{4}.$$

Таким образом, закон распределения случайной величины X имеет вид:

X	0	1	2
P	1/4	1/2	1/4

3. *Двое студентов играют в интересную игру: первый студент загадывает число от 1 до 6, второй студент должен угадать это число. Пусть случайная величина X – число попыток, сделанных вторым игроком при угадывании числа.*

a) Составить закон распределения случайной величины X.

b) Найдите математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

Решение. а) Дискретная случайная величина X может принимать значения 1,2,3,4,5,6 (студент угадает либо с первой, либо со второй и т.д. до шестой попытки) $P(x=1)=1/6$ (с первой попытки угадано нужное число, тогда число благоприятных исходов равно 1, а число всех исходов равно 6). Если $x=2$, то число угадано со второй попытки $P(x=2)=\frac{5}{6}\cdot\frac{1}{5}=\frac{1}{6}$. Если событие состоит в том, что первые две попытки были неудачные и только третья оказалась удачной, то вероятность такого события $P(x=3)=\frac{5}{6}\cdot\frac{4}{5}\cdot\frac{1}{4}=\frac{1}{6}$. Далее $P(x=4)=\frac{5}{6}\cdot\frac{4}{5}\cdot\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{6}$. Аналогично $P(x=5)=\frac{1}{6}$. $P(x=6)=\frac{1}{6}$.

Закон распределения числа сделанных попыток будет иметь вид:

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

б) $M(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}$.

Для нахождения дисперсии воспользуемся формулой

$D(X)=M(X^2)-(M(X))^2=\frac{91}{6}-\left(\frac{21}{6}\right)^2=15{,}17-12{,}25\approx 2{,}92$, где

$M(X^2)=1^2\cdot\frac{1}{6}+2^2\cdot\frac{1}{6}+3^2\cdot\frac{1}{6}+4^2\cdot\frac{1}{6}+5^2\cdot\frac{1}{6}+6^2\cdot\frac{1}{6}=\frac{1}{6}(1+4+9+16+25+36)=\frac{91}{6}$.

Среднеквадратическое отклонение $\sigma(X)=\sqrt{D(X)}=\sqrt{2{,}92}\approx 1{,}7$.

4. *Дискретная случайная величина X задана законом распределения:*

X	-1	0	2
P	p_1	p_2	p_3

Найдите p_1, p_2, p_3, если известны величины – математическое ожидание этой случайной величины $M(X)=0{,}1$ и математическое ожидание ее квадрата $M(X)^2=0{,}9$.

Решение. Пользуясь тем, что сумма вероятностей всех возможных значений случайной величины равна 1, имеем $p_1+p_2+p_3=1$.

Принимая во внимание, что $M(X)=0{,}1$, получаем соотношение

$-1p_1+0p_2+2p_3=0{,}1$.

Составим ряд распределения для случайной величины X^2:

X^2	1	0	4
P	p_1	p_2	p_3

и запишем выражение

$M(X^2)=1p_1+0p_2+4p_3$.

Используя условие задачи $M(X)^2=0{,}9$, имеем

$1p_1+4p_3=0{,}9$,

откуда получается система трех линейных уравнений с тремя неизвестными

$p_1+p_2+p_3=1$,

$-1p_1+0p_2+2p_3=0{,}1$,

$1p_1+4p_3=0{,}9$.

Решив эту систему, найдем искомые вероятности:

$p_1=\frac{7}{30}$, $p_2=\frac{3}{5}$, $p_3=\frac{1}{6}$.

Задачи для самостоятельного решения

1. *В урне 4 белых и 3 черных шара. Из нее последовательно вынимают шары до первого появления белого шара. Построить закон распределения случайной величины X – числа извлеченных шаров.*

2. *В урне 5 белых и 3 черных шара. Из нее наудачу извлекли три шара. Построить закон распределения случайной величины X – числа извлеченных белых шаров.*

3. *Дан ряд распределения случайной величины X. Найдите M(X).*

а)

X	–1	5	6
P	0,2	0,4	0,4

б)

X	1,2	2,3	4,1
P	0,35	p_2	0,24

4. *Найдите дисперсию и среднее квадратическое отклонение случайной величины X, заданной законом распределения:*

а)

X	4,3	5,1	10,6
P	0,2	0,3	0,5

б)

X	131	140	160	180
P	0,05	0,1	0,25	0,6

5. *Случайная величина X задана своим законом распределения:*

X	–1	0	x_3	5
P	0,3	0,2	0,1	0,4

Найдите x_3, если известно, что $M(X) = 1{,}9$.

6. *Два стрелка делают по одному выстрелу в одну мишень. Вероятность попадания для первого стрелка при одном выстреле 0,6; для второго 0,8. Случайная величина X – количество попаданий в мишень. Найти закон ее распределения и вычислить математическое ожидание X.*

7. *Стрелок стреляет по мишени 3 раза. Вероятность попадания равна 0,7 и не зависит от номера выстрела. Случайная величина X – количество попаданий. Найти закон ее распределения и вычислить математическое ожидание M(X).*

8. *Из колоды (36 карт) вытягиваем одну карту и выигрываем 15$, если это – картинка (Туз, К, Д, В) и теряем B$ если это цифра (6,7,8,9,10). При каких B нам играть выгодно?*

9. *Из колоды (36 карт) вытягиваем одну карту и выигрываем 8$, если это – Туз и теряем A$ если нет. При каких A нам играть невыгодно?*

10. *Закон распределения случайной величины X задан таблицей:*

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

Найти ее математическое ожидание и дисперсию.

11. *В коробке 5 синих и 4 красных шара. Из коробки случайным образом извлекают 3 шара. Случайная величина X – число извлеченных синих шаров. Найти закон ее распределения и вычислить математическое ожидание M(X).*

12. *В коробке 6 синих и 4 красных шара. Из коробки случайным образом извлекают 2 шара. Случайная величина X – число извлеченных красных шаров. Найти закон ее распределения, вычислить математическое ожидание M(X) и дисперсию D(X).*

Занятие № 7. Некоторые законы распределения случайных величин.

Примеры решения задач

1. *Проверкой качества установлено, что из каждых 100 деталей не имеют дефектов 75 штук в среднем. Составить биномиальный закон распределения вероятностей числа пригодных деталей из взятых наугад 6 деталей.*

Решение. Из условия задачи следует, что $p = 0{,}75$, $q = 1 - 0{,}75 = 0{,}25$, $n = 6$. В соответствии с формулой Бернулли находим:

$$P(X = 0) = 1 \cdot (0{,}25)^6 \approx 0{,}0002;$$

$$P(X) = 6 \cdot (0{,}75) \cdot (0{,}25)^5 \approx 0{,}0044;$$

$$P(X) = 15 \cdot (0{,}75)^2 \cdot (0{,}25)^4 \approx 0{,}033;$$

$$P(X = 3) = 20 \cdot (0{,}75)^3 \cdot (0{,}25)^3 \approx 0{,}132;$$

$$P(X = 4) = 15 \cdot (0{,}75)^4 \cdot (0{,}25)^2 \approx 0{,}297;$$

$$P(X = 5) = 6 \cdot (0{,}75)^5 \cdot (0{,}25) \approx 0{,}356;$$

$$P(X = 6) = 1 \cdot (0{,}75)^6 \approx 0{,}178.$$

Закон распределения данной случайной величины X – «числа стандартных деталей из 6 взятых наудачу» запишем в виде таблицы:

X	0	1	2	3	4	5	6
P	0,0002	0,0044	0,033	0,132	0,297	0,356	0,178

Проверка: $0{,}0002 + 0{,}0044 + 0{,}033 + 0{,}132 + 0{,}297 + 0{,}356 + 0{,}178 = 1.$

2. *Телевизионный канал рекламирует новый вид детского питания. Вероятность того, что телезритель увидит рекламу, оценивается в 0,002. В случайном порядке выбраны 500 телезрителей. Найдите вероятность того, что рекламу увидят:* **а)** ровно три телезрителя; **б)** менее трех телезрителей.

Решение. По условию $n = 500,\ p = 0{,}002,\ k = 3$. Найдем $\lambda = np = 500 \cdot 0{,}02 = 1$. Имеет место формула Пуассона:

$$P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

а) Найдем вероятность того, что рекламу увидят ровно три телезрителя:

$$P_{500}(3) = \frac{e^{-1}}{3!} = \frac{1}{6e} = 0{,}0613.$$

б) Найдем вероятность того, что рекламу увидят менее трех телезрителей:

$$P_{500}(0) + P_{500}(1) + P_{500}(2) = e^{-1} + e^{-1} + \frac{e^{-1}}{2} = \frac{5}{2} e^{-1} = \frac{5}{2e} = 0{,}9197.$$

3. *Цена товара X может быть в равной степени любой в пределах от 15 до 25 тыс. у.е. Найдите числовые характеристики данной случайной величины.*

Решение. Случайная величина X распределена равномерно, следовательно $M(X) = \dfrac{a+b}{2},\ D(X) = \dfrac{(b-a)^2}{12}.$

$$M(X) = \frac{15 + 25}{2} = 20,\quad D(X) = \frac{(25 - 15)^2}{12} = \frac{100}{12} = 8{,}33,\quad \sigma(X) = \sqrt{8{,}33} = 2{,}89.$$

4. *Среднее время обслуживания покупателя 10 минут. Чему равна вероятность простоя в очереди от 10 до 20 минут?*

<u>Решение.</u> По условию $M(X)=\frac{1}{\lambda}=10$, тогда $\lambda=\frac{1}{10}$. Если X – время обслуживания, то $P(\alpha<X<\beta)=\int_{\alpha}^{\beta}p(x)dx=F(\beta)-F(\alpha)=e^{-\alpha\lambda}-e^{-\beta\lambda}$.

$$P(10\le X\le 20)=F(20)-F(10)=e^{-\frac{1}{10}\cdot 10}-e^{-\frac{1}{10}\cdot 20}=e^{-1}-e^{-2}=0{,}2326.$$

5. *Предположим, что в течение года цена на акции некоторой компании есть случайная величина, распределенная по нормальному закону с математическим ожиданием, равным 30 у.е., и средним квадратическим отклонением, равным 10. Определить вероятность того, что в случайно выбранный день обслуживаемого периода цена за акцию была между 10 и 50 у.е.*

<u>Решение.</u> Воспользуемся формулой $P(\alpha\le X<\beta)=\Phi(\frac{\beta-a}{\sigma})-\Phi(\frac{\alpha-a}{\sigma})$. По условию $\alpha=10,\ \beta=50,\ a=30,\ \sigma=10$, тогда

$$P(10\le X<50)=\Phi(\frac{50-30}{10})-\Phi(\frac{10-30}{10})=\Phi(2)+\Phi(2)=2\Phi(2).$$

По таблице приложения 2 находим $\Phi(2)=0{,}4772$. Отсюда, искомая вероятность равна:

$$P(10<X<50)=2\cdot 0{,}4772=0{,}9544.$$

6. *Вес тропического грейпфрута, выращенного в Краснодарском крае – случайная величина, подчиненного нормальному закону с математическим ожиданием 0,5 кг и средним квадратическим отклонением 0,09 кг. Установить:* ***а)*** *вероятность того, что наудачу взятый грейпфрут имеет вес в пределах от 0,4 до 0,7 кг;* ***б)*** *вероятность того, что вес наудачу взятого грейпфрута отличается от математического ожидания не более, чем на 0,2 кг.*

<u>Решение.</u>

а) $P(0{,}4<X<0{,}7)=\Phi\left(\frac{0{,}7-0{,}5}{0{,}09}\right)-\Phi\left(\frac{0{,}4-0{,}5}{0{,}09}\right)=\Phi(2{,}22)+\Phi(1{,}11)=0{,}4867+0{,}3664=0{,}8531,$

б) $P(|X-0{,}5|<0{,}2)=2\Phi\left(\frac{\varepsilon}{\sigma}\right)=2\Phi\left(\frac{0{,}2}{0{,}09}\right)=2\Phi(2{,}22)=2\cdot 0{,}4867=0{,}9734,$

Задачи для самостоятельного решения

1. *Вероятность пересыхания реки в засушливый сезон каждого года составляет 0,3. Пусть случайная величина X – число случаев пересыхания реки за пятилетний период. Найти:*

а) закон распределения случайной величины X;

б) функцию распределения вероятностей случайной величины X;

в) числовые характеристики случайной величины X: математическое ожидание, дисперсию и среднее квадратическое отклонение.

2. *Телефонная станция обслуживает 400 абонентов. Для каждого абонента вероятность того, что в течение часа он позвонит на станцию,* *равна 0,01. Найдите вероятность следующих событий: «в течение часа 5 абонентов позвонят на станцию»; «в течение часа не более 4 абонентов позвонят на станцию»; «в течение часа не менее 3 абонентов позвонят на станцию».*

3. *Непрерывная случайная величина X распределена по показательному закону с функцией плотности* $p(x) = \begin{cases} 0, & x \le 0, \\ 7e^{-7x}, & x \ge 0. \end{cases}$.

Найдите вероятность того, что в результате испытаний случайная величина *X* попадет в интервал (0,15; 0,6), математическое ожидание, дисперсию случайной величины *X*.

4. *Известно, что дневная температура воздуха является равномерно распределенной случайной величиной X на отрезке* [*18; 20*]. *Найти*

1) плотность распределения вероятностей случайной величины X;
2) функцию распределения вероятностей случайной величины X;
3) вероятность попадания значений случайной величины X в интервал
4) среднюю дневную температуру воздуха.

5. *Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 мин. Найдите вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее 3 мин.*

6. *Случайная величина X распределена нормально с математическим ожиданием, равным 12,5. Вероятность попадания случайной величины X в интервал (10, 15) равна 0,2. Чему равна вероятность попадания случайной величины X в интервал (35, 40)?*

7. *Вес пойманной рыбы подчиняется нормальному закону распределения с параметрами* а=*375 г,* σ=*25 г. Найдите вероятность того, что вес одной рыбы будет: а) от 300 до 425 г; б) не более 450 г; в) больше 300 г.*

8. *Нефтеразведывательная компания получила финансирование для проведения 6 нефтеразработок. Вероятность успешной нефтеразведки 0,5. Предположим, что нефтеразведку осуществляют независимые друг от друга разведывательные партии. Составить закон распределения случайной величины X – числа успешных нефтеразведок. Найдите математическое ожидание и дисперсию случайной величины.*

Занятие № 8. Выборочный метод. Основные понятия, связанные с выборочным методом: генеральная и выборочная совокупности, дискретный и интервальный вариационные ряды, частоты, репрезентативность выборки.

Примеры решения задач

1. *Сто тысяч абитуриентов проходят централизованное тестирование по математике. Каждый из них может набрать от 0 до 100 баллов включительно. Для проведения статистического анализа отобрали группу из 15 абитуриентов. Они набрали следующие количества баллов: 30, 37, 78, 21, 92, 14, 67, 69, 37, 80, 30, 30, 10, 45, 52. Укажите генеральную и выборочную совокупности этого* *эксперимента. Что представляет собой реализация выборки? Оформите ее в виде вариационного и дискретного статистического ряда.*

Решение. Пусть X_i – число баллов, набранных i-м ($i=\overline{1,100000}$) абитуриентом. Тогда $X_1, X_2, \dots, X_{100000}$ набор случайных величин и есть генеральная совокупность данного эксперимента. Выборочная совокупность, это, например, набор случайных величин $X_1, X_2, \dots, X_{15}$.

Набор чисел

$$x_1=30, x_2=37, x_3=78, x_4=21, x_5=92, x_6=14, x_7=67, x_8=69, x_9=37, x_{10}=80, x_{11}=30,$$
$$x_{12}=30, x_{13}=10, x_{14}=45, x_{15}=52$$

и есть ее реализация. Обычно ее записывают кратко: 30, 37, 78, 21, 92, 14, 67, 69, 37, 80, 30, 30, 10, 45, 52.

Проранжировав статистические данные, получим вариационный ряд $x_{i_1}, x_{i_2}, \dots, x_{i_{15}}$, т. е. ряд $x_{13}=10, x_6=14, x_4=21, x_1=30, x_{11}=30, x_{12}=30, x_2=37,$ $x_9=37, x_{14}=45, x_{15}=52, x_7=67, x_8=69, x_3=78, x_{10}=80, x_5=92$.

Обычно вариационный ряд записывают без буквенных обозначений: 10, 14, 21, 30, 30, 30, 37, 37, 45, 52, 67, 69, 78, 80, 92. Отсюда видим, что в данной задаче всего 12 вариант (различных значений реализации выборки). А именно: 10, 14, 21, 30, 37, 45, 52, 67, 69, 78, 80, 92.

Подсчитав частоту и частость вариант, получим дискретный статистический ряд:

x_i	10	14	21	30	37	45	52	67	69	78	80	92
n_i	1	1	1	3	2	1	1	1	1	1	1	1

или

x_i	10	14	21	30	37	45	52	67	69	78	80	92
w_i	1/15	1/15	1/15	3/15	2/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15

2. *Для статистического исследования доходности предприятий региона выбрали группу из 30 фирм. Результаты исследований таковы: 90, 93, 123, 92, 111, 115, 119, 100, 99, 95, 94, 92, 98, 98, 103, 93, 102, 110, 115, 120, 117, 111, 113, 91, 99, 102, 112, 105, 108, 107. Постройте интервальный статистический ряд.*

Решение. Для удобства проранжируем данные: 90, 91, 92, 92, 93, 93, 94, 95, 98, 98, 99,99, 100, 102, 102, 103, 105, 107, 108, 110, 111, 111, 112, 113, 115, 115, 117, 119, 120, 123. По формуле Стерджеса найдем длину частичного интервала: $h=\dfrac{123-90}{m}$, где $m=1+log_2 30\approx 6$. Итак, $h=\dfrac{33}{6}=5.5$. Примем $h=6$, а в качестве начала первого интервала возьмем точку $90-\dfrac{6}{2}=87$. Тогда исходные данные будут разбиты на интервалы $[87,93)$, $[93,99)$, $[99,105)$, $[105,111)$, $[111,117)$, $[117,123]$, благодаря чему получим следующий интервальный статистический ряд:

Доходность	$[87,93)$	$[93,99)$	$[99,105)$	$[105,111)$	$[111,117)$	$[117,123]$

n_i	4	6	6	4	7	3

Или

Доходность	$[87,93)$	$[93,99)$	$[99,105)$	$[105,111)$	$[111,117)$	$[117,123]$
w_i	4/30	6/30	6/30	4/30	7/30	3/30

Задачи для самостоятельного решения

1. *В универсаме работает 120 человек. Изучается случайная величина X – число продаж каждого из 30 случайно выбранных продавцов универсама: 16, 12, 13, 15, 14, 21, 23, 9, 12, 14, 16,15, 22, 10, 16, 13, 15, 11, 19, 20, 24, 17, 16, 14, 14, 14, 12, 19, 20, 13.*

1) Что в данном исследовании представляет собой генеральная совокупность?

2) Что представляет собой выборка? Приведите два примера ее реализаций.

3) Постройте вариационный ряд, дискретный статистический ряд, интервальный статистический ряд.

2. *Имеются данные о еженедельном количестве проданных компьютеров одной из фирм: 332, 405, 456, 543, 345, 765, 567, 600, 613, 457, 504, 432, 457, 711, 249, 587, 459, 599,600, 499, 585, 600, 332, 459, 523, 428, 368, 444, 399, 548, 457, 598. Постройте вариационный ряд, дискретный статистический ряд, интервальный статистический ряд.*

3. *Число пассажиров одного из рейсов за 30 дней составило: 39, 42, 39, 46, 30,43, 32, 32, 48, 36, 44, 44, 43, 38, 42, 44, 44, 31, 39, 35, 42, 47, 31, 30, 43, 32, 41, 47, 46, 38. Постройте вариационный ряд, дискретный статистический ряд, интервальный статистический ряд.*

Занятие № 9. Статистическое распределение выборки. Полигон частот и гистограмма частот. Эмпирическая (статистическая) функция распределения.

Примеры решения задач

1. *Составить вариационный ряд оценок, полученных студентами одной группы по высшей математике: 3, 4, 5, 4, 3, 2, 5, 5, 3, 3, 4, 3, 4, 2, 3, 3, 4, 4, 5, 3.*

Решение. Составим вариационный ряд и запишем его в виде таблицы

Варианты (x_i)	2	3	4	5
Частоты (n_i)	2	8	6	4

Общая сумма частот вариационного ряда равна объему выборки, т.е.

$$n = \sum_{i=1}^{4} n_i = 2 + 8 + 6 + 4 = 20 .$$

Построим полигон частот по данному распределению. Для этого отложим на оси абсцисс варианты x_i, а на оси ординат – соответствующие им частоты n_i, соединив точки (x_i, n_i) отрезками прямых, получим искомый полигон частот (рис. 31).

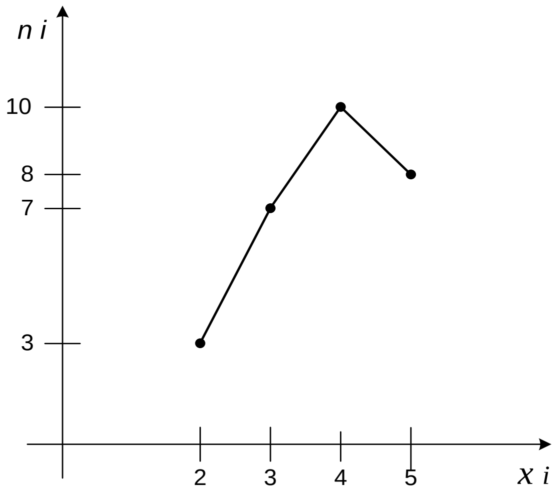


Рис. 31.

Частоты могут накапливаться. Накопленные частоты получаются последовательным суммированием значений частот от первой до последней.

Варианты (x_i)	2	3	4	5
Частоты (n_i)	2	8	6	4
Кумуляты частот	2	10	16	20

2. *Приведены результаты контрольной работы по высшей математике у студентов одной группы: 2, 4, 4, 2, 3, 3, 5, 3, 5, 5, 5, 2, 3, 3, 4, 4, 4, 5, 4, 3, 5, 5, 3, 4, 4, 4, 5, 4. Найти распределение относительных частот. Построить полигон частот и полигон относительных частот.*

Решение. Построим соответствующий вариационный ряд 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5.

Выборка содержит 4 различные значения, опишем ее статистическим рядом

x_i	2	3	4	5
n_i	3	7	10	8

Объем выборки: n=3+7+10+8=28. Найдем относительные частоты по формуле $w_i = \frac{m_i}{n}$: $w_1 = \frac{3}{28} = 0{,}11$; $w_2 = \frac{7}{28} = 0{,}25$; $w_3 = \frac{10}{28} = 0{,}36$; $w_4 = \frac{8}{28} = 0{,}28$. Напишем искомое распределение относительных частот:

x_i	2	3	4	5
w_i	0,11	0,25	0,36	0,28

Построим полигон частот по данному распределению. Для этого отложим на оси абсцисс варианты x_i, а на оси ординат – соответствующие им частоты n_i соединив точки (x_i, n_i) отрезками прямых, получим искомый полигон частот рис. 32.

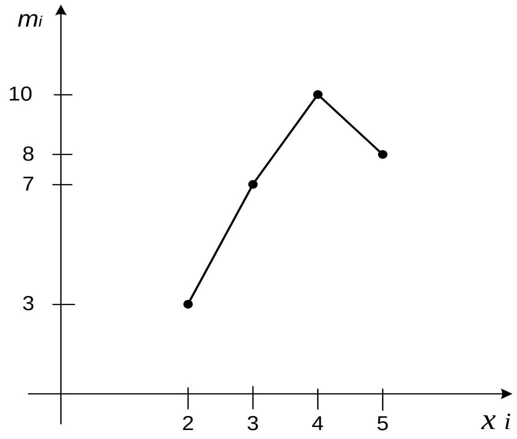


Рис. 32

Чтобы построить полигон относительных частот, воспользуемся распределением относительных частот. Отложим на оси абсцисс варианты x_i, а на оси ординат w_i – соответствующие относительные частоты, соединим точки (x_i, w_i) отрезками прямых и получим искомый полигон относительных частот на рис. 33.

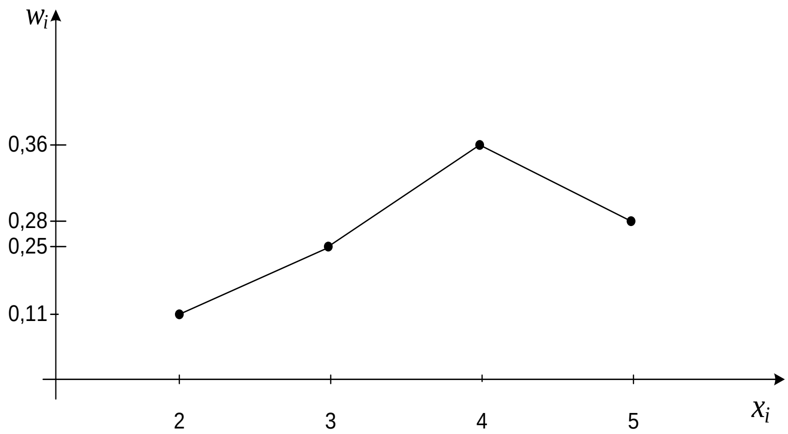


Рис. 33

3. *Был проведен следующий эксперимент: книгу «Основы высшей математики и теории вероятностей» открывали на случайной странице и искали слово «вероятность». При этом фиксировалось, сколько раз встречается искомое слово. В результате 20 опытов получена следующая выборка 4, 1, 4, 5, 1, 13, 4, 10, 2, 4, 7, 2, 2, 4, 6, 4, 5, 6, 2, 4. Найти эмпирическую функцию распределения и построить ее график.*

Решение. Соответствующий вариационный ряд 1, 1, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 5, 5, 6, 6, 7, 10, 13. Выборка содержит 8 различных значений и описывается следующим статистическим рядом:

x_i	1	2	4	5	6	7	10	13
m_i	2	4	7	2	2	1	1	1

Объем выборки n=2+4+7+2+2+1+1+1=20. Наименьшая варианта равна 1, тогда при $x \le 1$ $F_n(x)$=0, значение x<2, а именно x_1=1 наблюдалось 2 раза, тогда, $F_n(x)$=2/20=0,1 при $1 < x \le 2$. Значения x<4, а именно x_1=1 и x_2=2 принимались

2+4=6 раз, следовательно $F_n(x)$=6/20=0,3 при $2 < x \le 4$. Значения x<5 принимались 2+4+7=13 раз, тогда $F_n(x)$=13/20=0,65 при $4 < x \le 5$. Значения x<6 принимались 2+4+7+2=15 раз, тогда $F_n(x)$=15/20=0,75 при $5 < x \le 6$. Значения x<7 принимались 2+4+7+2+2=17 раз, тогда $F_n(x)$=17/20=0,85 при $6 < x \le 7$. Значения x<10 наблюдались 17+1=18 раз, тогда $F_n(x)$=18/20=0,9 при $7 < x \le 10$. Значения x<13 принимались 18+1=19 раз, тогда $F_n(x)$=19/20=0,95 при $10 < x \le 13$. Так как x=13 – наибольшая варианта, то $F_n(x)$=1, при x>13.Аналитическое выражение искомой эмпирической функции распределения:

$$F_n(x) = \begin{cases} 0, & x \le 1 \\ 0{,}1, & 1 < x \le 2 \\ 0{,}3, & 2 < x \le 4 \\ 0{,}65, & 4 < x \le 5 \\ 0{,}75, & 5 < x \le 7 \\ 0{,}85, & 6 < x \le 7 \\ 0{,}9, & 7 < x \le 10 \\ 0{,}95, & 10 < x \le 13 \\ 1, & x > 13 \end{cases}$$

Построим график этой функции (рис. 34)

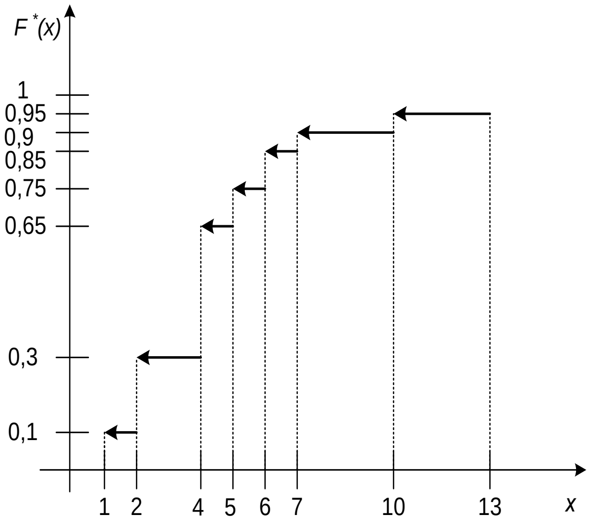


Рис. 34

Занятие № 10. Выборочная средняя, мода, медиана, выборочная статистическая дисперсия и выборочное среднее квадратическое отклонение.

Примеры решения задач

1. *Рабочая бригада из 11 человек имеет следующие тарифные разряды: 9,5,9,6,6,8,7,6,7,8,6. Найти моду выборки.*

Решение. Упорядочим статистический ряд по возрастанию 5,6,6,6,6,7,7,8,8,9,9. Мода отражает наиболее распространенную варианту рассматриваемого признака (т.е. варианту, которая имеет наибольшую частоту). В данном случае M_o=6, т.е. это будет рабочий шестого разряда.

2. *По данным социологического исследования был получен ряд распределения рабочих цеха по числу изготавливаемых за смену деталей.*

x_i	8	10	12	14
n_i	1	4	10	6

Вычислить выборочное среднее, выборочную дисперсию, выборочное среднеквадратическое отклонение, коэффициент вариации и размах вариации.

Решение. Для нахождения выборочного среднего воспользуемся формулой:

$$\overline{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + x_3 \cdot n_3 + x_4 \cdot n_4}{n}, \quad \overline{x} = \frac{8 \cdot 1 + 10 \cdot 4 + 12 \cdot 10 + 14 \cdot 6}{1 + 4 + 10 + 6} = \frac{8 + 40 + 120 + 84}{21} = 12 \cdot$$

Найдем выборочную дисперсию

$$s^2 = \frac{(x_1 - \overline{x})^2 n_1 + (x_2 - \overline{x})^2 n_2 + (x_3 - \overline{x})^2 n_3 + (x_4 - \overline{x})^2 n_4}{n} =$$

$$= \frac{(8-12)^2 \cdot 1 + (10-12)^2 \cdot 4 + (12-12)^2 \cdot 10 + (14-12)^2 \cdot 6}{21} = \frac{4^2 \cdot 1 + 2^2 \cdot 4 + 0^2 \cdot 10 + 2^2 \cdot 6}{21} =$$

$$= \frac{16 + 16 + 0 + 24}{21} = \frac{56}{21}.$$

Выборочное среднеквадратическое отклонение $s = \sqrt{s^2} = \sqrt{\frac{56}{21}} \approx 1,63$. По формуле $v = \frac{s}{x} \cdot 100\%$ вычисляем коэффициент вариации: $v = \frac{1,63}{12} \cdot 100\% = 13,58$ %.

Размах вариации: $R = x_{\max} - x_{\min}$ =14–8=6.

3. *Обследование качества пряжи на прочность дало следующие результаты*

Прочность нити	Число случаев (частота)	Накопленная частота
120-140	1	1
140-160	6	7
160-180	19	26
180-200	58	84
200-220	53	137
220-240	24	161
240-260	16	177
260-280	3	180

Найти моду, медиану.

Решение. Так как наибольшая частота $n_i = 58$ отвечает интервалу 180-200, то $x_i = 180$, $n_{i-1} = 19$, $n_{i+1} = 53$, $h = 20$. По формуле

$$Mo = x_i + h \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

получаем:

$$Mo = 180 + 20 \cdot \frac{58 - 19}{(58 - 19) + (58 - 52)} = 197,73.$$

Номер медианного интервала определяется на основании неравенства:

$$\sum_{i=1}^{l} m_i \le \frac{n}{2} < \sum_{i=1}^{l+1} m_i .$$

$$\sum_{i=1}^{l} m_i \le \frac{180}{2} < \sum_{i=1}^{l+1} m_i \Rightarrow \sum_{i=1}^{4} m_i \le 90 < \sum_{i=1}^{5} m_i \Rightarrow 84 \le 90 < 137.$$

Следовательно, номер медианного интервала $m_{Me} = 5$, а сам интервал 200-220. По формуле:

$$Me = x_l + h \cdot \frac{\frac{n}{2} - \sum_{i=1}^{l-1} m_i}{m_{Me}} .$$

получаем

$$Me = 200 + 20 \cdot \frac{90 - 84}{53} = 202{,}26 .$$

4. *Требуется найти выборочную среднюю и выборочную дисперсию для интервального ряда заданного таблицей:*

Рост	150-154	154-158	158-162	162-166	166-170	170-174	174-178	178-182	182-186
Число студентов, n_i	10	30	105	230	250	220	115	30	10

Решение. Для расчета средней переходим от интервального ряда к дискретному, заменяя каждый интервал его серединой. Далее перейдем к условным вариантам по формуле: $u_i = \frac{x_i - c}{k}$, где $c = 168$, а $k = 4$, т.е. $u_i = \frac{x_i - 168}{4}$. Все необходимые вычисления приведем в таблице:

Рост, (см)	Число студентов, n_i	Середина интервала, x_i	$u_i = \frac{x_i - c}{k}$	$u_i \cdot n_i$	$u_i^2 \cdot n_i$
150-154	10	152	-4	-40	160
154-158	30	156	-3	-90	270
158-162	105	160	-2	-210	420
162-166	230	164	-1	-230	230
166-170	250	168	0	0	0
170-174	220	172	1	220	220
174-178	115	176	2	230	460
178-182	30	180	3	90	270
182-186	10	184	4	40	160
Сумма	1000	-	-	10	2190

Тогда $\bar{u} = \frac{\sum u_i \cdot n_i}{n} = \frac{10}{1000}$.

Отсюда найдем выборочную среднюю $\bar{x} = \bar{u} \cdot k + c = \frac{10}{1000} \cdot 4 + 168 = 168{,}04$.

Найдем выборочную дисперсию по формуле

$$s_x^2 = \frac{\sum_{i=1}^{m} u_i^2 n_i}{n} \cdot k^2 - \left(\overline{x} - c\right)^2 = \frac{2190}{1000} \cdot 4^2 - \left(168,04 - 168\right)^2 = 35,0384 \approx 35,04.$$

Задачи для самостоятельного решения

Для указанных статистических данных выполнить задания требуется:

1. определить объем выборки;
2. найти размах выборки;
3. разбив выборку на $k = 7$ интервалов, построить интервальный статистический ряд;
4. построить полигон абсолютных частот;
5. построить гистограмму относительных частот;
6. построить эмпирическую функцию распределения;
7. найти выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение;
8. найти выборочные моду и медиану;
9. вычислить выборочные коэффициенты асимметрии и эксцесса;
10. вычислить коэффициенты вариации и осцилляции.

1. $k = 9$

17	21	15	19	22	20	17	18	21	16
19	20	14	16	17	18	19	15	18	17
16	17	22	19	21	15	17	19	14	20
17	19	18	15	14	18	21	16	19	17
18	14	17	19	16	20	17	23	17	15
15	17	21	22	20	14	19	18	16	18
16	18	14	18	19	15	21	17	22	20
22	17	20	15	21	18	20	14	18	16
20	16	18	21	17	23	21	19	15	18
18	14	19	20	15	20	21	18	21	20

2. $k = 7$

16	17	2	14	19	18	15	9	20	16
19	18	22	18	16	18	19	21	14	16
21	14	18	2	15	18	17	22	20	19
17	19	1	18	14	18	19	21	15	17
22	17	22	16	21	15	21	14	20	16
20	15	19	17	20	23	18	15	18	20
17	21	14	21	16	15	17	19	22	19
22	17	6	20	15	18	16	18	15	17
19	4	7	17	16	17	18	14	21	16
15	20	17	20	2	18	15	21	20	15

3. $k = 8$

183	207	213	208	186	210	198	219	231	227

202	211	220	236	227	220	210	183	213	190
197	227	187	226	213	191	209	196	202	235
211	214	220	195	182	228	202	207	192	226
193	203	232	202	215	195	220	233	214	185
234	215	196	220	203	236	225	221	193	215
204	184	217	193	216	205	197	203	229	204
225	216	233	223	208	204	207	182	216	191
210	190	207	205	232	222	198	217	211	201
185	217	225	201	208	211	189	205	207	199

4. $k = 7$

20	26	32	34	26	28	22	30	17	24
30	28	18	22	24	26	34	28	22	20
34	24	28	20	32	17	22	24	26	30
30	22	26	35	28	24	30	32	28	18
20	30	17	24	32	28	22	26	24	30
34	26	24	28	22	30	35	32	20	17
28	22	36	30	20	26	28	23	24	32
20	26	30	24	32	17	22	28	35	26
28	35	32	22	26	24	26	24	30	24
18	24	26	28	35	30	26	22	26	28

5. $k = 8$

57	46	33	49	29	50	38	41	27	34
37	49	51	26	55	42	59	43	46	30
31	43	58	41	35	47	33	45	49	37
47	34	54	39	60	49	25	50	31	53
38	41	30	51	37	55	47	43	35	42
35	46	27	45	41	34	50	29	51	39
42	59	43	31	38	58	54	37	26	43
29	42	33	41	24	39	53	45	33	51
45	21	54	50	37	30	41	61	42	46
38	53	34	47	35	49	57	39	55	31

6. $k = 10$

20	49	43	31	44	38	40	31	28	43
32	44	47	29	50	25	43	38	41	32
38	24	49	40	32	34	31	28	37	46
41	35	43	25	37	46	38	24	41	50
38	29	41	32	34	49	44	37	31	47
50	34	25	37	40	32	35	28	44	43
46	37	41	35	29	43	38	31	26	34
49	32	46	26	38	35	40	50	37	46
37	25	40	34	24	44	32	28	34	38
44	34	29	47	37	49	43	35	47	50

7. $k = 9$

70	95	75	85	60	77	55	63	80	67
90	78	57	76	84	82	75	68	73	62
62	81	77	72	97	68	85	56	92	71
73	78	98	63	83	85	70	90	66	91

86	68	55	93	71	96	77	81	86	72
82	62	70	78	67	87	91	99	78	87
91	58	81	97	75	83	71	66	61	76
73	85	65	90	86	61	54	75	78	93
87	58	72	92	66	98	65	81	76	63
95	83	65	57	80	87	61	92	56	71

8. $k=10$

181	141	162	103	136	124	41	117	69	153
101	24	67	154	172	110	62	59	197	121
135	58	199	159	81	39	142	87	179	85
171	107	125	192	163	200	133	150	178	98
148	56	113	169	73	138	104	31	90	109
127	116	190	20	111	94	157	119	53	76
66	132	166	91	44	115	72	26	128	149
46	75	105	137	82	64	186	96	176	97
156	33	188	58	112	139	86	174	106	77
152	130	43	108	119	129	37	71	96	114

9. $k=7$

32	105	48	80	144	128	64	112	18	81
66	129	113	17	94	78	90	51	104	34
110	149	36	103	82	53	93	130	68	150
114	84	55	131	70	38	102	77	16	135
41	19	142	61	85	159	115	57	72	101
56	100	86	146	73	40	141	25	87	126
151	71	94	15	125	76	54	99	39	140
17	124	52	98	139	37	147	88	69	109
35	158	67	30	93	123	50	138	21	97
96	121	49	137	89	145	91	65	92	33

10. $k=10$

88	72	100	60	116	74	36	143	114	70
56	75	30	76	89	53	117	90	135	103
35	128	71	86	43	76	61	113	34	83
62	84	50	69	120	91	102	47	119	99
33	76	91	37	85	17	85	63	121	74
46	85	63	104	77	92	54	78	42	105
85	79	49	80	93	32	106	81	64	79
73	19	80	65	107	123	51	94	80	108
52	83	124	81	96	82	109	20	95	68
66	41	82	98	111	67	125	97	112	58

Занятие № 11. Точечные и интервальные оценки числовых характеристик случайной величины.

Примеры решения задач

1. *Из партии, содержащей 2000 деталей, для проверки по схеме бесповторной выборки было отобрано 200 деталей, среди которых оказалось*

184 стандартных. Найти: а) вероятность того, что доля нестандартных деталей во всей партии отличается от полученной доли в выборке не более чем на 0,02 (по абсолютной величине); б) границы, в которых с надежностью 0,95 заключена доля нестандартных деталей во всей партии.

Решение. Имеем $N = 2000$, $n = 200$, $n = 200 - 184 = 16$ нестандартных деталей. Выборочная доля нестандартных деталей $w = \frac{m}{n} = \frac{16}{200} = 0,08$.

а) Находим среднюю квадратическую ошибку бесповторной выборки для доли:

$$\sigma'_w = \sqrt{\frac{0,08 \cdot 0,92\left(1 - \frac{200}{2000}\right)}{200}} = 0,0182.$$

Тогда искомая доверительная вероятность**:**

$$P(|w - p| \le 0,02) = 2\Phi\left(\frac{0,02}{0,0182}\right) = 2\Phi(1,10) = 0,729,$$

т.е. вероятность того, что выборочная доля нестандартных деталей будет отличаться от генеральной доли не более чем на 0,02, равна 0,729.

б) Учитывая, что $\gamma = 2\Phi(t) = 0,95$ и $t = 1,96$, найдем предельную ошибку выборки для доли**:** $\Delta = 1,96 \cdot 0,0182 = 0,0357$**.** Теперь искомый доверительный интервал:

$$0,08-0,0357 \le p \le 0,08+0,0357 \text{ или } 0,044 \le p \le 0,116.$$

Итак, с надежностью 0,95 доля нестандартных деталей во всей партии заключена от 0,044 до 0,116.

2. *Найти с надежностью 0,95 доверительный интервал для неизвестного математического ожидания a, нормально распределенного признака X, если генеральная дисперсия $\sigma^2 = 7$, выборочная средняя $\bar{x} = 4,5$ и объем выборки n=20.*

Решение. Требуется найти доверительный интервал по формуле

$$\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}.$$

По условию $\bar{x} = 4,5$; n=20; $\sigma = \sqrt{7}$; т.е. все величины кроме t_γ известны. Найдем t_γ из соотношения $\Phi(t) = \frac{\gamma}{2}$, где $\gamma = 0,95$, $\Phi(t) = \frac{0,95}{2} = 0,475$. По таблице (Приложение 2) находим t_γ =1,96.

Подставим t_γ =1,96 в формулу $\bar{x} - t_\gamma \frac{\sigma}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{\sigma}{\sqrt{n}}$ и получим $4,5 - 1,96 \cdot \frac{\sqrt{7}}{\sqrt{20}} < a < 4,5 + 1,96 \cdot \frac{\sqrt{7}}{\sqrt{20}}$. Окончательно получим 95% доверительный интервал для математического ожидания

$$(4,5-1,16;\ 4,5+1,16)=(3,34;\ 5,66).$$

3. *В некотором районе в личном владении находится 3500 коров. Выборочно обследовали 800 коров и установили, что среднегодовой удой у этой* *группы коров 4200 кг. Оценить вероятность того, что среднегодовой удой у всех коров отличается от среднего выборочного не более чем на 10 кг, если генеральное среднеквадратичное отклонение равно 250 кг.*

Решение. Рассмотрим случаи повторного и бесповторного отбора. Из условия имеем n= 800; $\Delta = 10$; $\sigma = 250$; $a = 4200$; $N = 3500$.

Для повторного отбора

$$P(|\bar{x} - a| < \Delta) = P(|\bar{x} - 4200| < 10) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{10 \cdot \sqrt{800}}{250}\right) = 2\Phi(1{,}13) = 2 \cdot 0{,}3708 = 0{,}7416 \approx 74\%.$$

Для бесповторного отбора

$$P(|\bar{x} - a| < \Delta) = P(|\bar{x} - 4200| < 10) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sigma\sqrt{1 - \frac{n}{N}}}\right) = 2\Phi\left(\frac{10 \cdot \sqrt{800}}{250\sqrt{1 - \frac{800}{3500}}}\right) = 2\Phi(1{,}288) \approx 2\Phi(1{,}29) =$$

$$= 2 \cdot 0{,}4015 = 0{,}803 = 80\%$$

Мы видим, что бесповторная выборка дает более точный результат при одном и том же объеме.

4. *Определить объем повторной и бесповторной выборок для определения средней продолжительности горения лампочки в партии из 5000 лампочек, чтобы с вероятностью 0,99 предельная ошибка выборки не превосходила 25 часов. Предельная ошибка распределена нормально со среднеквадратическим отклонением равным 150.*

Решение. По условию $\sigma = 150$; Δ =25; γ =0,99; N=5000.

Из соотношения $\Phi(t) = \frac{\gamma}{2}$ найдем t: $\Phi(t) = \frac{0{,}99}{2} = 0{,}495$. По таблице (Приложение 2) находим t=2,58.

Следовательно, необходимый объем выборки в случае повтора

$$n = \frac{t^2\sigma^2}{\Delta^2} = \frac{(150)^2 \cdot (2{,}58)^2}{(25)^2} = 239{,}6 \approx 240.$$

В случае бесповторного отбора

$$n = \frac{Nt^2\sigma^2}{t^2\sigma^2 + \Delta^2 N} = \frac{5000 \cdot (150)^2 \cdot (2{,}58)^2}{(2{,}58)^2 \cdot (25)^2 + (150)^2 \cdot (2{,}58)^2} \approx 229.$$

5. *Из 2000 деталей было отобрано 400, распределение которых по размеру задается следующей таблицей.*

Размер детали, мм	7,975	8,025	8,075	8,125	8,175	8,225
Количество деталей	12	28	132	150	62	16

Найти среднюю ошибку выборки при повторном и бесповторном отборе.

Решение. Выборочную среднюю и “исправленное” среднеквадратическое отклонение найдем соответственно по формулам

$n = \sum m_i$ =12+28+132+150+62+16,

$$\bar{x} = \frac{\sum x_i m_i}{n} = \frac{1}{400}\left(12 \cdot 7{,}975 + 28 \cdot 8{,}025 + 132 \cdot 8{,}075 + 150 \cdot 8{,}125 + 62 \cdot 8{,}175 + 16 \cdot 8225\right) \approx 8{,}11,$$

$$s=\sqrt{\frac{\sum m_i(x_i-\bar{x})^2}{n-1}}=$$

$$=\sqrt{\frac{1}{399}\cdot(12\cdot(7{,}975-8{,}11)^2+28\cdot(8{,}025-8{,}11)^2+132\cdot(8{,}075-8{,}11)^2+150\cdot(8{,}125-8{,}11)^2+62\cdot(8{,}175-8{,}11)^2+16\cdot(8{,}225-8{,}11)^2)}\approx 0{,}052.$$

Таким образом, средняя ошибка для повторной выборки

$$\mu=\frac{s}{\sqrt{n}}=\frac{0{,}052}{\sqrt{400}}=0{,}0026,$$

а для бесповторной выборки $\mu=\frac{s}{\sqrt{n}}\cdot\sqrt{1-\frac{n}{N}}=\frac{0{,}052}{\sqrt{400}}\cdot\sqrt{1-\frac{400}{2000}}=\frac{0{,}052}{20}\cdot\frac{2}{\sqrt{5}}\approx 0{,}0023$.

6. *Из генеральной совокупности, где признак распределен по нормальному закону, извлечена выборка объема* n=12:

x_i	-0,5	-0,4	-0,2	0	0,2	0,6	0,8	1	1,2	1,5
n_i	1	2	1	1	1	1	1	1	2	1

Найти доверительный интервал для математического ожидания, который отвечает доверительной вероятности $\gamma=0{,}95$.

Решение. В данном примере воспользуемся формулой

$$\bar{x}-\frac{st_\alpha}{\sqrt{n}}<a<\bar{x}+\frac{st_\alpha}{\sqrt{n}}$$

Сначала найдем

$$\bar{x}=\frac{\sum x_i n_i}{n}=\frac{1}{12}\left(-0,5\cdot1-0,4\cdot2-0,2\cdot1+0+0,2\cdot1+0,6\cdot1+0,8\cdot1+1+1,2\cdot2+1,5\cdot1\right)\approx 0,4$$

$$s=\sqrt{\frac{\sum n_i(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{1}{11}((-0,5-0,417)^2\cdot1+(-0,4-0,417)^2\cdot2+(-0,2-0,417)^2\cdot1+(0-0,417)^2\cdot1+(0,2-0,417)^2\cdot1+(0,6-0,417)^2\cdot1+(0,8-0,417)^2\cdot1+(1-0,417)^2\cdot1+(1,2-0,417)^2\cdot2+(1,5-0,417)^2\cdot1)}\approx 0,720.$$

В случае малой выборки (n<30) вместо нормального распределения пользуются распределением Стьюдента, которое зависит от одного параметра – числа степеней свободы этого распределения. В нашем случае $\nu=n-1=12-1=11$. По соответствующей таблице распределения Стьюдента находим $t_{кр}$(11;0,05)=2,2, где $\alpha=1-0{,}95=0{,}05$. Тогда искомый интервал $\left(0{,}417-\frac{0{,}720}{\sqrt{12}}\cdot2{,}2;0{,}417+\frac{0{,}720}{\sqrt{12}}\cdot2{,}2\right)=(-0{,}04;0{,}88)$. Это означает, что если будет произведено достаточно большое число выборок данного объема, то 95% из них найденный доверительный интервал накроет математическое ожидание и только в 5% случаев оцениваемое математическое ожидание может выйти за границы доверительного интервала.

7. *Перед выборами было опрошено* n = 650 *человек. Из них* k = 350 *человек отдали предпочтение демократам. На сколько голосов могут рассчитывать* *демократы, если число избирателей в городе равно* $N = 65000$. *Вычисления произвести с доверительной вероятностью* 0,95.

Решение. Обозначим через m число избирателей, которые проголосовали «за». Применим формулу для бесповторного отбора:

$$P\left(\left|w-p\right|<\varepsilon\right)\approx 2\Phi\left(\frac{\varepsilon}{\mu}\right),$$

где

$$\mu=\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}.$$

Откуда $p-\varepsilon \le W \le p+\varepsilon$.

Так как $W=\frac{m}{n}$, то $\left(p-\varepsilon\right)N \le m \le \left(p+\varepsilon\right)N$. Найдем p и ε. По условию $N = 65000$; $p=\frac{k}{n}=\frac{350}{650}=\frac{7}{13}$; $q=1-\frac{7}{13}=\frac{6}{13}$.

Доверительная вероятность по условию $P = 0{,}95$, а значит $2\Phi\left(t\right)=0{,}95$ или $\Phi\left(t\right)=0{,}475$. Из приложения 2 находим, что $t = 1{,}96$. Тогда

$$\frac{\varepsilon}{\mu}=t=1{,}96 \quad\Rightarrow\quad \varepsilon=1{,}96\cdot\mu=1{,}96\cdot\sqrt{\frac{\frac{7}{13}\cdot\frac{6}{13}}{650}\cdot\left(1-\frac{650}{65000}\right)}\approx 0{,}038.$$

Таким образом, $\frac{7}{13}-\varepsilon \le W \le \frac{7}{13}+\varepsilon \quad\Rightarrow\quad 65000\left(\frac{7}{13}-\varepsilon\right)\le m\le 65000\left(\frac{7}{13}+\varepsilon\right)$

.Следовательно, $65000\cdot 0{,}5 \le m \le 65000\cdot 0{,}576$, т.е. $32500 \le m \le 37440$.

Итак, демократы смогут рассчитывать на m голосов, где $32500 \le m \le 37440$

Задачи для самостоятельного решения

1. *Методом моментов и методом максимального правдоподобия найти по реализации выборки* $x_1, x_2, \dots, x_n$ *оценку неизвестной вероятности успеха p в схеме Бернулли.*

2. *Пользуясь методом моментов и методом максимального правдоподобия, оценить вероятность появления решки, если при 10 бросаниях монеты она появилась 6 раз.*

3. *Методом моментов и методом максимального правдоподобия найти по реализации выборки* $x_1, x_2, \dots, x_n$ *оценку параметра a распределения Пуассона.*

4. *Случайная величина* X *распределена по закону Пуассона. Для нахождения параметра a этого распределения была произведена выборка, объема 1000:*

x_i	0	1	2	3	4	5	6
m_i	405	366	175	40	8	4	2

Методом моментов найти оценку параметра a.

5. *Методом максимального правдоподобия найти по реализации выборки* $x_1, x_2, \dots, x_n$ *оценку параметра b равномерного на отрезке* $[0, b]$ *распределения.*

6. *Произведено 10 независимых наблюдений значений нормально распределенной случайной величины* $X \sim N(a, \sigma)$. *В результате наблюдений было посчитано, что* $\bar{x}_{\text{в}} = 49$. *Построить 90%, 95% и 99% доверительные интервалы для математического ожидания случайной величины* X. *Как изменятся доверительные интервалы, если число наблюдений увеличить до 100 и полагать, что* $\bar{x}_{\text{в}}$ *по-прежнему равно 49?*

7. *Измерительный прибор не имеет систематической ошибки, а случайные ошибки распределены по нормальному закону с* $\sigma = 20$. *Сколько нужно сделать независимых измерений, чтобы определить истинное значение измеряемой величины с ошибкой, не превосходящей 10 при надежности* $\gamma = 0,9$, $\gamma = 0,95$ *и* $\gamma = 0,99$?

8. *Аналитик оценивает среднюю доходность определенных акций. Случайная выборка 15 дней показала, что средняя доходность* $\bar{x}_{\text{в}}$ *составляет 10,5% и имеет среднее квадратическое отклонение* $\sigma_{\text{в}} = 3,4\%$. *Считая, что доходность акций подчиняется нормальному закону распределения, построить 95% и 99% доверительные интервалы для средней доходности.*

9. *Для оценки параметра* σ *нормально распределенной случайной величины была сделана выборка объема 50 и вычислено* $s = 15$. *Найти 90% доверительный интервал для параметра* σ, *считая, что*

а) математическое ожидание $a = 20$,

б) математическое ожидание неизвестно.

10. *Результат социологического опроса среди курильщиков о числе выкуренных сигарет в день представляет собой нормально распределенную случайную величину* $X \sim N(a, \sigma)$. *Случайная выборка объема 15 представляет следующие результаты: 7, 10, 12, 15, 16, 16, 16, 8, 19, 10, 9, 20,25, 29, 17. Построить 95% доверительный интервал для:*

а) математического ожидания a, полагая, что $\sigma = 4$;

б) математического ожидания a, полагая, что величина σ неизвестна;

в) среднеквадратического отклонения σ, считая, что $a = 16$;

г) среднеквадратического отклонения σ, считая, что величина a неизвестна.

11. *Известно, что доходность группы акций представляет собой нормально распределенную случайную величину* $X \sim N(a, \sigma)$. *Случайная выборка объема 10 представляет следующие результаты: 24, 26, 28, 23, 25, 25, 30, 29, 27, 27. Построить 95% доверительный интервал для:*

а) математического ожидания a, полагая, что $\sigma = 2$;

б) математического ожидания a, полагая, что величина σ неизвестна;

в) среднеквадратического отклонения σ, считая, что $a = 26$;

г) среднеквадратического отклонения σ, считая, что величина a неизвестна.

Занятие № 12. Понятие о статистической гипотезе, основная и альтернативная гипотезы. Статистический критерий, уровень значимости. Ошибки первого и второго рода. Некоторые критерии согласия.

Примеры решения залач

1. *На основании сделанного прогноза средняя дебиторская задолженность однотипных предприятий региона должна составить* $a_0 = 120$ *ден. ед. Выборочная проверка* 10 *предприятий дала среднюю задолженность* $\bar{x} = 135$ *ден. ед., а среднее квадратическое отклонение задолженности* $s = 20$ *ден. ед. На уровне значимости* 0,05*: а) выяснить, можно ли принять данный прогноз; б) найти мощность критерия, если в действительности средняя дебиторская задолженность всех предприятий региона равна* 130 *ден. ед*.

<u>Решение.</u> а) Проверяемая гипотеза $H_0 : \bar{x}_0 = a = 120$. В качестве альтернативной возьмем гипотезу $H_1 : a > 120$. Так как генеральная дисперсия σ^2 неизвестна, то используем t-критерий Стьюдента. Статистика критерия в соответствии с табл. 5.5 равна $t = \dfrac{\bar{x} - a_0}{s/\sqrt{n-1}} = \dfrac{135-120}{20/\sqrt{10-1}} = 2{,}25.$ Критическое значение статистики

$$t_{1-2\cdot 0{,}05;10-1} = t_{0{,}9;9} = 1{,}83.$$

Так как $|t| > t_{0{,}9;9} (2{,}25 > 1{,}83)$,то гипотеза H_0 отвергается, т.е. на 5%-ном уровне значимости сделанный прогноз должен быть отвергнут.

б) Альтернативная гипотеза $H_1 : \bar{x}_0 = a_1 = 120.$ Так как $a_1 = 130 > a_0 = 120$, то критическая область правосторонняя и критическое значение выборочной средней

$$\bar{x}_{кр} = \bar{x}_0 + t_{1-2\alpha,n-1}\frac{s}{\sqrt{n-1}} = a + t_{0{,}9;9}\frac{s}{\sqrt{n-1}} = 120 + 1{,}83\frac{20}{\sqrt{10-1}} = 132{,}2(ден.ед.)$$

т.е. критическая область значений для $\bar{x}$ есть интервал $(132{,}2;+\infty)$. Мощность критерия равна вероятности P отвергнуть гипотезу H_0, когда верна гипотеза H_1 т.е.

$$P = P(132{,}2 < \bar{x} < +\infty) = \frac{1}{2} + \frac{1}{2}\theta(t, n-1),$$

где $t = \dfrac{\bar{x} - a_1}{s/\sqrt{n-1}} = \dfrac{132{,}2-130}{20/\sqrt{10-1}} = 0{,}33.$ Следовательно, $\theta(0{,}33;9) \approx 0{,}25.$

Итак, $P \approx \dfrac{1}{2}(1 + 0{,}25) = 0{,}62$.

2. *Сравниваются средние доходы двух фирм X и Y в двух однотипных отраслях, имеющих нормальное распределение с дисперсиями 1 млн. долларов и 4 млн. долларов соответственно. В первой отрасли по выборке из 20 фирм получен средний доход 1 млн. долларов, а во второй по выборке из 25 фирм получен средний доход 0,9 млн. долларов. Определить, есть ли основание отклонить гипотезу H*$_0$*: MX=MY с уровнем значимости а =0,05.*

Решение. В качестве критерия проверки H_0 примем СВ $U_{набл}=\dfrac{\overline{x}-\overline{y}}{\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{k}}}$.

По условию $\overline{x}=1; \overline{y}=0{,}9; \sigma_x^2=1; \sigma_y^2=4; n=20; k=25$, тогда

$$U_{набл}=\frac{1-0{,}9}{\sqrt{\frac{1}{20}+\frac{4}{25}}}=\frac{7}{\sqrt{21}}\approx 0{,}22.$$

При конкурирующей гипотезе H_1: $MX \neq MY$ определяем две критические точки $U_{1-\frac{\alpha}{2}}$ и $U_{\frac{\alpha}{2}}$ из условий $\Phi\left(U_{\frac{\alpha}{2}}\right)=\dfrac{1-\alpha}{2}=\dfrac{1-0{,}05}{2}=0{,}475; U_{1-\frac{\alpha}{2}}=U_{\frac{\alpha}{2}}$. Из приложения 2 находим $U_{\frac{\alpha}{2}}=U_{1-\frac{\alpha}{2}}=1{,}95$. Так как $U_{набл.}\approx 0{,}22<1{,}95$, то нет основания для отклонения гипотезы H_0: $MX=MY$.

3. *Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема* n=200:

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение. Сначала найдем выборочное среднее $\overline{x}_в$ и выборочное среднеквадратическое отклонение $\sigma_в$. Перейдем к условным вариантам по формуле $v_i=\dfrac{x_i-x_0}{h}$, где h=2, x_0=13, тогда $v_i=\dfrac{x_i-13}{2}$. Все необходимые вычисления приведем в таблице:

x_i	n_i	v_i	$n_i\cdot v_i$	$n_i\cdot v_i^2$
5	15	-4	-60	240
7	26	-3	-78	234
9	25	-2	-50	100
11	30	-1	-30	30
13	26	0	0	0
15	21	1	21	21
17	24	2	48	96
19	20	3	60	180
21	13	4	52	208
Сумма	200		-37	1109

Следовательно, $\overline{v}=\dfrac{\sum\limits_i m_i\cdot v_i}{n}=\dfrac{-37}{200}=-0{,}185$, тогда

$\overline{x}=\overline{v}_i\cdot 2+13=-0{,}185\cdot 2+13=12{,}63$; $\sigma_v^2=\dfrac{\sum\limits_i m_i\cdot v_i^2}{n}-\left(\dfrac{\sum\limits_i m_i\cdot v_i}{n}\right)^2=\dfrac{1109}{200}-\left(\dfrac{-37}{200}\right)^2=5{,}511$;

$\sigma_в^2 = \sigma_v{}^2 \cdot h^2 = 5{,}511 \cdot 2^2 = 22{,}043$, тогда $\sigma_в = \sqrt{22{,}044} = 4{,}695$. Далее вычислим теоретические частоты по формуле $m_i' = \frac{n \cdot h}{\sigma_в} \cdot \varphi(u_i) = \frac{200 \cdot 2}{4{,}695} \cdot \varphi(u_i) = 85{,}2 \cdot \varphi(u_i)$, где $u_i = \frac{(x_i - \bar{x})}{\sigma_в}$. Значение функции *φ(u)* найдем из приложения 1.

Все необходимые вычисления приведены в таблице:

x_i	$u_i = \frac{(x_i - \bar{x})}{\sigma_в}$	*φ(u)*	$n_i' = 85{,}2 \cdot \phi(u_i)$
5	-1,6251	0,1074	9,0056
7	-1,1991	0,1942	16,5458
9	-0,7731	0,2966	25,2703
11	-0,3471	0,3752	31,9670
13	0,0788	0,3977	33,8840
15	0,5047	0,3503	29,9989
17	0,9307	0,2589	22,0582
19	1,3567	0,1582	13,4786
21	1,7827	0,0818	6,9693

Так как $\chi^2_{набл} = \sum \frac{\left(n_i - n_i'\right)^2}{n_i'}$, то производя соответствующие вычисления в таблице

n_i	n_i'	$n_i - n_i'$	$\left(n_i - n_i'\right)^2$	$\frac{\left(n_i - n_i'\right)^2}{n_i'}$
15	9,0056	5,9944	35,9328	3,9900
26	16,5458	9,4542	89,3819	5,4020
25	25,2703	-0,2703	0,07306	0,0028
30	31,9670	-1,967	3,8690	0,1210
26	33,8840	-7,884	62,1574	1,8344
21	29,9989	-8,9989	80,9802	2,6994
24	22,0582	1,9418	3,7705	0,1709
20	13,4786	6,5214	42,5286	3,1552
13	6,9693	6,0307	36,3693	5,2185
Сумма				22,59

получаем $\chi^2_{набл} = 22{,}59$. Найдем $\chi^2_{кр}$ по уровню значимости α =0,05 и числу степеней свободы $k=s-r-1=9-2-1=6$ из приложения 4. $\chi^2_{кр}(0{,}05;6) = 12{,}6$. Так как $\chi^2_{набл} > \chi^2_{кр}$, то гипотезу о нормальном распределении генеральной совокупности отвергаем, т.е. эмпирические и теоретические частоты различаются значимо.

4. *Распределение СВ X задается следующим вариационным рядом*:

Интервалы СВ X	m_i
5 - 10	2
10 - 15	14
15 - 20	11
20 - 25	9
25 - 30	4

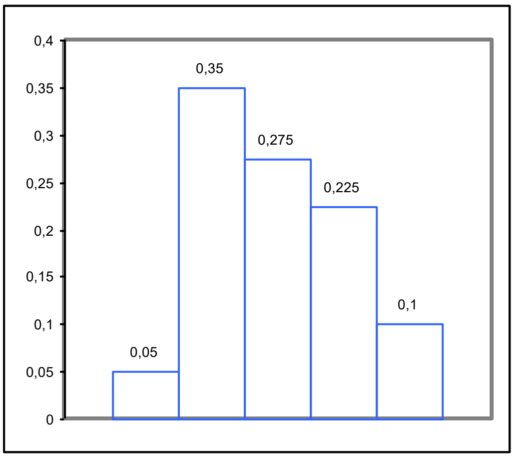
Выдвинуть гипотезу о виде распределения этой случайной величины. Используя критерий Пирсона, при уровне значимости 0,05 *проверить, согласуется ли выдвинутая гипотеза с эмпирическим распределением выборки.*

Решение. Найдем объем выборки n: $\sum_{i=1}^{l} m_i = n \quad \Rightarrow \quad n = 2 + 14 + 11 + 9 + 4 = 40$.

Составим следующую таблицу:

x	m_i	$p_i = m_i / n$	Середины интервалов
5-10	2	0,05	7,50
10-15	14	0,35	12,5
15-20	11	0,275	17,5
20-25	9	0,225	22,5
25-30	4	0,1	27,5
	40		

Построим полигон полученного распределения:



Используя таблицу, найдем выборочное среднее значение дисперсию и среднее квадратическое отклонение интервального вариационного ряда.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{5} x_i m_i = \frac{695}{40} = 17{,}375;$$
$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{5} x_i^2 m_i = \frac{13250}{40} = 331{,}25$$

$$\sigma^2 = \overline{X^2} - \left(\overline{X}\right)^2 = 331{,}25 - 17{,}375^2 = 29{,}36;$$
$$\sigma = \sqrt{29{,}36} = 5{,}42$$

По виду полигона частот выдвинем гипотезу, что случайная величина распределена по нормальному закону. Проверим эту гипотезу с помощью критерия χ^2 Пирсона при уровне значимости $\alpha = 0{,}05$. Будем считать, что случайная величина X распределена по нормальному закону распределения с функцией распределения $F(x) = \frac{1}{2}\left(1 + \Phi\left(\frac{x - 17{,}375}{5{,}42}\right)\right)$. По формуле $P(x_1 < X < x_2) = \frac{1}{2}\left(\Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right)\right)$ вычислим вероятность p_i $\left(i = \overline{1,5}\right)$ того, что случайная величина содержится в интервале (x_i; x_{i+1}), а затем по формуле $m_i' \approx n \cdot p_i$ найдем соответствующую теоретическую частоту случайной величины X в этом интервале.

	$P(x_1 < X < x_2)$	$m_i' \approx n \cdot p_i$	p_i'
$P(5<X<10)=$	0,0755952	$m_1' \approx 3$	3,0238073
$P(10<X<15)=$	0,2438195	$m_2' \approx 10$	9,7527793
$P(15<X<20)=$	0,355296	$m_3' \approx 14$	14,211842
$P(20<X<25)=$	0,2343409	$m_4' \approx 9$	9,373635
$P(25<X<30)=$	0,0698188	$m_5' \approx 3$	2,7927537

Проверим степень согласия теоретического и эмпирического распределений с помощью критерия Пирсона. Для этого заполним таблицу:

m_i	m_i'	$m_i - m_i'$	$\left(m_i - m_i'\right)^2$	$\frac{\left(m_i - m_i'\right)^2}{m_i'}$
2	3	1	1	0,3333333
14	10	-4	16	1,6
11	14	3	9	0,6428571
9	9	0	0	0
4	3	-1	1	0,3333333
			СУММА	2,9095238

$$\chi_{набл}^{2} = \sum \frac{\left(m_i - m_i'\right)^2}{m_i'} = 2{,}9095238$$

Итак, $\chi_{набл}^{2} = 2{,}91$. Найдем число степеней свободы: $5 - 2 = 3$. Находим области при уровне значимости 0,05 $\chi^2_{прак} = 7{,}82$. Так как $\chi_{набл}^{2} < \chi_{кр}^{2}$, то нет оснований отвергать гипотезу о нормальном распределении СВ X.

Задачи для самостоятельного решения

1. *Событие А в 10000 независимых испытаний произошло 6825 раз. Пусть случайная величина X – число наступлений события А в одном испытании, а гипотеза* H_0 *заключается в том, что случайная величина X распределена по биномиальному закону с параметром р=0,7. При уровне значимости* $\alpha = 0{,}05$ *проверить гипотезу* H_0, *используя*

а) критерий согласия Пирсона,

б) критерий согласия Колмогорова.

2. *Событие А в 5000 независимых испытаний произошло 3809 раз. Пусть случайная величина X – число наступлений события А в одном испытании, а гипотеза* H_0 *заключается в том, что случайная величина X распределена по биномиальному закону. При уровне значимости* $\alpha = 0{,}01$ *проверить гипотезу* H_0 *, используя*

а) критерий согласия Пирсона,

б) критерий согласия Колмогорова.

3. *С целью определения загруженности телефонной станции в течение 100 случайно выбранных 10-секундных интервалов регистрировалось число вызовов:*

Число вызовов	0	1	2	3	4	5
Число 10-секундных интервалов (m_i)	18	32	33	26	16	10

Используя критерий согласия Пирсона, при уровне значимости $\alpha = 0{,}05$ *проверить гипотезу о том, что случайная величина X – число вызовов – имеет распределение Пуассона.*

4. *Отдел технического контроля проверил 500 партий однотипных изделий и установил, что число X нестандартных деталей в одной партии имеет эмпирическое распределение, приведенное в таблице.*

Число нестандартных изделий в одной партии (x_i)	0	1	2	3	4	5
Количество партий, содержащих x_i нестандартных изделий (m_i)	193	188	80	24	13	2

Используя критерий согласия Пирсона, при уровне значимости $\alpha = 0{,}01$ *проверить гипотезу о том, что случайная величина X – число нестандартных изделий в одной партии – распределена по закону Пуассона.*

5. *Результаты наблюдений числа появлений автобуса на остановке, имеющего 4-минутный интервал движения, приведены в следующей таблице.*

Интервал в минутах	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$
Число появлений автобуса (m_i)	45	47	48	50	45

При уровне значимости $\alpha = 0{,}01$ *проверить гипотезу о равномерном законе распределения, используя*

а) критерий согласия Пирсона,

б) критерий согласия Колмогорова.

6. *Владелец фирмы считает, что добиться более высокой прибыли ему помешала неравномерность поставок по месяцам года. Поставщик утверждает, что поставки были достаточно равномерными. Распределение поставок имеет следующий вид:*

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Объем поставок (m_i)	20	22	27	17	20	20	20	30	29	29	33	40

С помощью критерия согласия Пирсона при уровне значимости $\alpha = 0{,}05$ *определить, кто прав владелец фирмы или поставщик?*

7. *Данные о числе посетителей гипермаркета в первые 60 минут после его открытия приведены в таблице:*

Временные интервалы (в минутах)	$[0,10)$	$[10,20)$	$[20,30)$	$[30,40)$	$[40,50)$	$[50,60)$
Число посетителей (m_i)	148	72	31	20	19	10

Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ *проверить гипотезу о том, что число посетителей распределено по показательному закону.*

8. *Компания по производству электроприборов тестирует новый образец ламп. Число вышедших из строя ламп за 1000 часов работы имеет следующее распределение*

Временные интервалы	$[0,200)$	$[200,400)$	$[400,600)$	$[600,800)$	$[800,1000)$
Число вышедших из строя ламп (m_i)	541	134	79	32	14

Аналитик считает, что число вышедших из строя ламп распределено по показательному закону с параметром $\lambda = 0{,}005$. *Используя критерий Пирсона, при уровне значимости* $\alpha = 0{,}025$ *выяснить, прав ли аналитик.*

9. *При уровне значимости* $\alpha = 0{,}025$ *проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические (*m_i*) и теоретические частоты (*$m_i^{\text{теор}}$*):*

m_i	5	10	20	25	14	3
$m_i^{\text{теор}}$	6	14	27	19	11	3

10. *Данные о числе высокооплачиваемых работников некоторого предприятия, приведены в таблице:*

Заработная плата в тыс. у. е.	$[1, 1.5)$	$[1.5, 2)$	$[2, 2.5)$	$[2.5, 3)$	$[3, 3.5)$	$[3.5, 4)$
Число работников (m_i)	30	97	203	217	103	50

Используя критерий Пирсона, при уровне значимости $\alpha = 0,01$ *проверить гипотезу* H_0 *о том, что случайная величина X – число высокооплачиваемых работников – имеет нормальный закон распределения.*

3. РАЗДЕЛ КОНТРОЛЯ ЗНАНИЙ

3.1. Примерные вопросы к зачёту

1. Основные понятия теории множеств, примеры множеств.

2. Операции над множествами (пересечение, объединение, разность), их свойства.

3. Число элементов в объединении конечных множеств.

4. Основные принципы комбинаторики. Правило умножения.

5. Перестановки.

6. Размещения.

7. сочетания.

8. Классификация событий. Операции над событиями и их свойства. Вероятности случайных событий. Классическое и статистическое определения вероятности события.

9. Условные вероятности. Независимость событий.

10. Теоремы сложения и умножения вероятностей.

11. Формула полной вероятности.

12. Повторение испытаний. Схема Бернулли.

13. Случайные величины: дискретные и непрерывные. Закон распределения дискретной случайной величины.

14. Математическое ожидание, дисперсия, их свойства.

15. Функция распределения случайной величины. Некоторые законы распределения случайных величин (биномиальное, нормальное).

16. Выборочный метод. Основные понятия, связанные с выборочным методом: генеральная и выборочная совокупности, дискретный и интервальный вариационные ряды, частоты, репрезентативность выборки.

17. Статистическое распределение выборки. Полигон частот и гистограмма частот. Эмпирическая (статистическая) функция распределения.

18. Выборочная средняя, мода, медиана, выборочная статистическая дисперсия и выборочное среднее квадратическое отклонение.

19. Точечные и интервальные оценки числовых характеристик случайной величины.

20. Статистическая гипотеза, основная и альтернативная гипотезы. Статистический критерий, уровень значимости. Ошибки первого и второго рода. Некоторые критерии согласия.

3.2. Средства диагностики

Оценка и определение уровня знаний и практических профессиональных Перечень рекомендуемых средств диагностики:

1. Компьютерное тестирование.
2. Контрольные работы.
3. Устный опрос

Оценка за ответы на практических занятиях включает в себя полноту ответа, наличие аргументов, примеров из практики, правильности решения практических примеров и задач и т.д.

Формой промежуточной аттестации по дисциплине «Основы математической статистики» учебным планом предусмотрен зачет.

3.3. Примерные промежуточные контрольные работы

Контрольная работа 1. Элементы теории множеств (2 ч.)

1. В туристической группе из 25 человек 8 человек владеет английским языком, 11 – немецким, 3-е владеют обоими языками. Сколько человек в группе не владеет ни одним из этих языков?

2. A – множество иностранных студентов 1-го курса исторического факультета БГУ, B – множество студентов 1-го курса специальности «регионоведение». Описать множества $A \cap B$, $A \cup B$, $A \setminus B$.

3. Изобразить на диаграмме Эйлера-Венна множество $A \setminus (B \cap \overline{C})$. Описать словами это множество, если A – множество стран, которые входят в Европейский союз (ЕС), B – множество стран членов блока NATO, C – множество стран, которые входят в Шенгенскую зону. Принадлежат ли этому множеству страны США, Литва?

4. Доказать равенство множеств $A \cup B = A \cup (B \setminus A)$.

Контрольная работа 2. Элементы комбинаторики и теории вероятностей (2 ч.)

1. В коробке 6 синих и 4 красных шара. Из коробки случайным образом извлекают 2 шара. Случайная величина X – число извлеченных синих шаров. Найти закон ее распределения, вычислить математическое ожидание $\mathrm{M}(X)$ и дисперсию $\mathrm{D}(X)$. Найти функцию распределения вероятностей случайной величины X.

2. Монета брошена 2 раза. Найти вероятность того, что, хотя бы 1 раз появится «орел».

3. Из букв слова КАРАНДАШ, составленного с помощью разрезной азбуки, наудачу последовательно извлекаются 4 буквы и выкладываются в ряд. Какова вероятность того, что получится слово ДАША?

4. Сколько среди трёхзначных чисел встречается таких, в записи которых не участвует число 6?

5. В коробке 15 книг, среди которых 9 детективов. Наудачу берем 4 книги. Найти вероятность того, что среди них окажется 3 детектива.

6. В первой урне находится 3 белых шара и 5 черных, а во второй урне -7 белых и 2 черных шара. Извлекаем один шар из первой урны, перекладываем его во вторую урну и, кроме того, мы добавляем во вторую урну еще два шара того же цвета, что извлеченный шар. Затем мы извлекаем два шара из второй урны. Найти вероятность того, что оба шара, извлеченные из второй урны, будут черными.

Контрольная работа 3. Выборочный метод. Статистическое распределение. Статистические оценки параметров распределения (2 ч.)

1. По результатам измерений уровня загрязненности атмосферы в центре города получены следующие значения индекса загрязнения:

27; 29; 23; 30; 31; 25; 30; 29; 24; 29; 31; 28; 28; 24; 29; 26; 30; 29; 28; 25; 28; 30; 29; 27; 25; 28; 27; 32; 31; 28.

Построить вариационный и статистический ряды, статистическое распределение, полигон относительных частот, эмпирическую функцию распределения для этой выборки

2. Используя полученные в задаче 1 результаты, вычислить следующие числовые характеристики выборки из указанного примера: выборочное среднее значение, моду, медиану, выборочную статистическую дисперсию, исправленную статистическую дисперсию, исправленное среднее квадратичное отклонение выборки.

3. За отчетный период на пункте «Дубровка» выборка таможенных платежей физических лиц представлена в виде таблицы:

75	191	76	138	113	157	129	100	182	144
100	155	200	111	115	95	120	180	147	146
145	154	100	117	96	137	168	189	204	149
145	200	111	83	139	89	172	125	140	183
122	124	110	123	138	125	129	70	142	169
210	150	136	100	102	107	154	108	135	150
90	173	184	131	154	160	187	151	159	137

Требуется:

1. определить объем выборки;
2. найти размах выборки;
3. разбив выборку на $k = 7$ интервалов, построить интервальный статистический ряд;
4. построить полигон абсолютных частот;
5. построить гистограмму относительных частот;
6. построить эмпирическую функцию распределения;
7. найти выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение;
8. найти выборочные моду и медиану.

3.4. Примерные тестовые задания

Элементы комбинаторики. Случайные события и вероятности (2 ч/ДОТ).

1. Сколько существует способов расставить на полке 6 книг?

1) 720. 2) 30. 3) 36. 4) 120. 5) 6. 6) 600.

2. Сколько различных трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5 при условии, что в каждом числе нет одинаковых цифр?

1) 60. 2) 120. 3) 15. 4) 125. 5) 50. 6) 27.

3. Сколькими способами можно сделать трехцветный флаг с горизонтальными полосами одинаковой ширины, если есть материя шести различных цветов?

1) 120 2) 36 3) 30 4) 18 5) 216 6) 60

4. Из десяти различных задач для контрольной работы выбирается пять. Сколько различных вариантов контрольной работы можно составить?

1) 252. 2) 200. 3) 10. 4) 1200. 5) 720. 6) 600.

5. В классе 20 учеников. Сколько имеется способов выбрать из класса двух дежурных?

1) 190. 2) 20. 3) 40. 4) 380. 5) 200. 6) 400

6. В чемпионате страны по футболу принимает участие 16 команд. Сколькими способами может определиться тройка призёров этих состязаний?

1) 3360. 2) 560. 3) 48. 4) 240. 5) 200. 6) 400

7. Вероятность попадания в цель первым стрелком равна 0,7; вторым стрелком — 0,8. Стрелки выстрелили одновременно. Какова вероятность того, что, хотя бы один из них попал в цель?

1) 0,94. 2) 1,5. 3) 0,56. 4) 0,83. 5) 0,97. 6)0,5.

8. Симметричная монета подброшена 3 раза. Какова вероятность того, что цифра выпадет не менее двух раз?

1) 1/2. 2) 1/8. 3) 2/3. 4) 3/8. 5) 1/3. 6) 3/4.

9. В коробке 7 белых и 4 чёрных шара. Случайным образом вынули 2 шара. Вероятность того, что они оба белые равна:

1) 21/55. 2) 1/3 3) 2/7 4) 19/51 5) 7/11. 6) 1/2.

10. Бросаются два игральных кубика. Вероятность того, что сумма выпавших очков окажется 7, равна:

1) 1/6. 2) 0,7. 3) 7/36. 4) 1/7. 5) 2/7. 6) 1/18.

11. Монета бросается 3 раза. Тогда вероятность того, что герб выпадет ровно 2 раза равна:

1) 3/8. 2) 2/3. 3) 3/4. 4) 1/2. 5) 1/4. 6) 1/3.

12. В коробке 5 белых и 4 чёрных шара. Случайным образом вынули 2 шара. Вероятность того, что они оба черные равна:

1) 1/6. 2) 1/3 3) 2/9 4) 4/9. 5) 4/27. 6) 1/2.

4. ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ

4.1. Содержание учебного материала

Раздел 1. ЭЛЕМЕНТЫ ТЕОРИИ МНОЖЕСТВ

Тема 1.1. Основные понятия теории множеств

Понятие множества, подмножества. Способы задания множеств. Примеры множеств. Пустое и универсальное множество. Диаграммы Эйлера–Венна.

Тема 1.2. Операции над множествами

Пересечение, объединения и разность множеств. Свойства операций над множествами. Число элементов в объединении конечных множеств.

Раздел 2. ЭЛЕМЕНТЫ КОМБИНАТОРИКИ И ТЕОРИИ ВЕРОЯТНОСТЕЙ

Тема 2.1. Элементы комбинаторики

Правило суммы, правило произведения. Перестановки, размещения и сочетания, их число.

Тема 2.2. Случайные события и вероятности

Основы теории вероятностей. Классификация событий. Операции над событиями и их свойства. Вероятности случайных событий. Условные вероятности. Независимость событий. Теоремы сложения и умножения вероятностей. Формула полной вероятности. Схема независимых испытаний, формула Бернулли.

Тема 2.3. Случайные величины

Случайные величины: дискретные и непрерывные. Закон распределения дискретной случайной величины. Математическое ожидание случайной величины, дисперсия случайной величины, их свойства. Функция распределения случайной величины. Некоторые законы распределения случайных величин.

Раздел 3. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Тема 3.1. Выборочный метод. Статистическое распределение

Выборочный метод. Основные понятия, связанные с выборочным методом: генеральная и выборочная совокупности, дискретный и интервальный вариационные ряды, частоты, репрезентативность выборки. Статистическое распределение выборки. Полигон частот и гистограмма частот. Эмпирическая (статистическая) функция распределения.

Тема 3.2. Статистические оценки параметров распределения

Выборочная средняя, мода, медиана, выборочная статистическая дисперсия и выборочное среднее квадратическое отклонение. Точечные и интервальные оценки числовых характеристик случайной величины.

Тема 3.3. Проверка статистических гипотез

Понятие о статистической гипотезе, основная и альтернативная гипотезы. Статистический критерий, уровень значимости. Ошибки первого и второго рода. Некоторые критерии согласия.

4.2. Рекомендуемая литература

Основная учебная литература

1. Бондаренко, Н. Н. Теория вероятностей. Математическая статистика. Практикум: учебное пособие для студентов учреждений высшего образования по специальностям "Менеджмент (по направлениям)", "Бизнес-администрирование (по направлениям)", "Маркетинг" / Н. Н. Бондаренко, Л. Г. Третьякова, М. Л. Зеленкевич; [под ред. М. Л. Зеленкевич]; М-во образования

Республики Беларусь, БГУ, Институт бизнеса БГУ. – Минск: Ин-т бизнеса БГУ, 2021. – 231 с. - URL: https://elib.bsu.by/handle/123456789/259178.

2. Велько, О.А. Основы высшей математики для социологов: Учебно-методическое пособие / О.А. Велько, М.В. Мартон, Н.А. Моисеева. – Минск: БГУ, 2020. – 303 с. - URL: https://elib.bsu.by/handle/123456789/254094.

3. Велько, О.А. Основы высшей математики и теории вероятностей: Учебно-методическое пособие / О.А. Велько, М.В. Мартон, Н.А. Моисеева. – Минск: БГУ, 2022. – 399 с. - URL: https://elib.bsu.by/handle/123456789/290012.

4. Высшая математика. Практикум : учебное пособие в 2 ч. / О. М. Матейко [и др.]; под ред. С. А. Самаля. – Минск : РИВШ, 2022. – Ч. 2. – 360 с.

5. Коган, Е. А. Теория вероятностей и математическая статистика : учебник / Е. А. Коган, А. А. Юрченко. — Москва : ИНФРА-М, 2024. — 250 с. — (Среднее профессиональное образование). - ISBN 978-5-16-015649-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/2078388 (дата обращения: 17.08.2023). – Режим доступа: по подписке.

6. Матейко, О.М. Высшая математика : учеб.-метод. пособие. В 2 ч. Ч. 2 / О. М. Матейко, С. А. Самаль, Н. Б. Яблонская. – Минск : БГУ, 2022. – 176 с.

7. Пирогова, И. Н. Элементы теории вероятностей и математической статистики : учебно-методическое пособие / И. Н. Пирогова, Е. Г. Филиппова. — Екатеринбург : , 2018. — 81 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/121341 (дата обращения: 17.08.2023). — Режим доступа: для авториз. пользователей.

Дополнительная учебная литература

1. Велько, О.А. Основы математической статистики и их применение в социологических исследованиях : учебно-методическое пособие / О. А. Велько, М. В. Мартон, Н. А. Моисеева ; БГУ, Механико-математический фак., Каф. общей математики и информатики. – Минск : БГУ, 2023. – 110 с. : ил., табл.– Библиогр.: с. 102–104. № 002729032023, Деп. в БГУ 29.03.2023. https://elib.bsu.by/handle/123456789/295986.

2. Велько, О.А. Элементы теории вероятностей : учебно-методическое пособие / О. А. Велько, М. В. Мартон, Н. А. Моисеева ; БГУ, Механико-математический фак., Каф. общей математики и информатики. – Минск : БГУ, 2023. – 104 с. : ил., табл. – Библиогр.: с. 99–101. № 002429032023, Деп. в БГУ 29.03.2023. https://elib.bsu.by/handle/123456789/295981.

3. Гайшун, Л.Н. Теория вероятностей: Учебное пособие для студентов экономических специальностей / Л.Н. Гайшун, Г.К. Игнатьева, О.А. Велько. – Минск: МИУ, 2002. – 167 с.

4. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебник для вузов / В. Е. Гмурман. – 12-е изд. – М.: Юрайт, 2020. – 479 с.

5. Грес, П.В. Математика для гуманитариев: учеб. пособие / П.В. Грес. - М.: Логос, 2003. – 120 с.

6. Дубатовская, М. В. Математическая статистика : учеб.-метод. пособие для студ. учреждений высш. образования, обуч. по спец. 1 ступени высш.

образования: 1-25 01 01 "Экономическая теория", 1-25 01 02 "Экономика" / М. В. Дубатовская, С. В. Рогозин, Е. И. Васенкова ; БГУ. - Минск : БГУ, 2015. - 143 с.

7. Математическая статистика в психологии : учеб.-метод. пособие / М-во образования РБ, УО "БГПУ им. М. Танка" ; [сост.: Н. А. Литвинова, Н. П. Радчикова]. –Минск : БГПУ, 2016. – 87 с.

8. Петров, В.А. Теория вероятностей и математическая статистика: Учебно-методический комплекс / В.А. Петров, Г.К. Игнатьева, О.А. Велько. – 2-е изд. – Минск: МИУ, 2009. – 268 с.

9. Письменный, Д. Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д. Письменный. - 10-е изд. - Москва : Айрис пресс, 2022. - 287 с.

10. Толстова, Ю. Н. Математическая статистика для социологов : учебник и практикум для вузов, для студентов, обучающихся по гуманитарным направлениям / Ю. Н. Толстова. – 2-е изд., испр. и доп. – Москва:Юрайт, 2020. – 258 с.

4.3. Электронные ресурсы

1. Образовательный портал БГУ [Электронный ресурс]. – Режим доступа: https://eduhist.bsu.by/course/view.php?id=842– Дата доступа: 16.09.2023.

2. Велько, О.А. Основы высшей математики: УМК для специальности 1-23 01 05 «Социология» [Электронный ресурс] / О. А. Велько, Н.А. Моисеева // БГУ. Минск, 05.03.2020. – Режим доступа: https://elib.bsu.by/handle/123456789/241078. – Дата доступа: 16.09.2023.