МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Рубцовский индустриальный институт (филиал)
федерального государственного бюджетного образовательного
учреждения высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»
(РИИ АлтГТУ)

О.В. Ефременкова

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ НАДЕЖНОСТИ

Учебное пособие
для студентов заочной формы обучения направления
«Строительство»

Рекомендовано Рубцовским индустриальным институтом (филиалом) ФГБОУ ВО «Алтайский государственный технический университет им. И.И. Ползунова» в качестве учебного пособия для студентов, обучающихся по направлению подготовки «Строительство»

Рубцовск 2017

УДК 517.8

Ефременкова О.В. Математические основы теории надежности: Учебное пособие для студентов заочной формы обучения направления «Строительство»./ Рубцовский индустриальный институт. – Рубцовск, 2017.- 131 с.

Данное учебное пособие предназначено для студентов всех форм обучения, направления «Строительство». В пособии представлен широкий спектр, как теоретического материала, так и задач прикладной математики. Содержание задач затрагивает разные отрасли практических знаний, с которыми будущий инженер-электрик может столкнуться в своей профессиональной деятельности.

Рассмотрено и одобрено на заседании
НМС Рубцовского индустриального
института
Протокол № 8 от 27.12.2017 г.

Рецензент: к.т.н. О.А. Михайленко

СОДЕРЖАНИЕ

1. ПОНЯТИЕ НАДЕЖНОСТИ И ЕЕ СВОЙСТВА

Надежностью называют свойство объекта сохранять во времени в установленных пределах значения всех параметров, характеризующих способность выполнять требуемые функции в заданных режимах и условиях применения, технического обслуживания, ремонтов, хранения и транспортировки.

Особенности надежности:

- **Надежность есть внутреннее свойство объекта**, заложенное в него при проектировании и изготовлении и проявляющееся во время эксплуатации. Надежность может быть оценена количественно.
- **Надежность проявляется во времени**. Она отражает устойчивость начального качества объекта во времени.
- **Надежность по - разному проявляется при различных условиях эксплуатации и различных режимах применения объекта**. При изменении режимов и условий эксплуатации изменяются и характеристики надежности. **Нельзя оценивать надежность объекта, не уточнив условия его эксплуатации и режимов применения.**

Расширение условий эксплуатации, повышение ответственности выполняемых **системами обработки информации и управления** функций, их усложнение приводит к повышению требований к надежности её элементов.

Надежность является сложным свойством и включает в себя единичные свойства:

- **Безотказность** - способность объекта непрерывно сохранять работоспособное состояние в течение времени. Потому наиболее важным в обеспечении надежности систем обработки информации и управления является повышение её безотказности.
- **Ремонтопригодность** - это свойство объекта, заключающееся в приспособленности к поддержанию и восстановлению работоспособного состояния путем технического обслуживания и ремонта. Не каждый ремонтнопригодный объект является восстанавливаемым, т.к. для этого дополнительно требуется подготовленный обслуживающий персонал, специальные организационно - технические мероприятия по обслуживанию и снабжению запасными частями, а также мероприятия по созданию приемлемых условий эксплуатации объекта.
- **Сохраняемость** - это свойство объекта сохранять в заданных пределах значения параметров, характеризующих способность объекта выполнять требуемые функции в течение и после хранения и/или транспортирования.
- **Долговечность** - это свойство объекта сохранять работоспособное состояние до наступления предельного состояния, т. е. до момента, когда дальнейшая эксплуатация объекта недопустима или нецелесообразна либо восстановление его работоспособного состояния невозможно или нецелесообразно.

Особенностью проблемы **надежности** является ее связь со всеми этапами жизненного цикла систем обработки информации и управления от зарождения идеи создания до её списания:

- при расчете и проектировании систем обработки информации и управления её надежность закладывается в проект,
- при изготовлении систем обработки информации и управления её надежность обеспечивается,
- при эксплуатации систем обработки информации и управления её надёжность реализуется.

Проблема надежности - комплексная проблема, и решать ее необходимо на всех этапах и разными средствами.

На этапе проектирования систем обработки информации и управления определяется её структура, производится выбор или разработка элементной базы. Поэтому здесь имеются наибольшие возможности обеспечения требуемого уровня надежности систем обработки информации и управления.

Основным методом решения этой задачи являются **расчеты надежности** (в первую очередь - **безотказности**), в зависимости от структуры **систем обработки информации и управления** и характеристик её составляющих частей, с последующей необходимой коррекцией проекта.

Безотказность (и другие составляющие свойства **надежности) систем обработки информации и управления** проявляется через случайные величины, например, **наработку до очередного отказа** и **количество отказов за заданное время**.

Для описания этих случайных величин используют стандартные **распределения случайных величин**, рассматриваемых в теории вероятностей. Для каждого распределения рассматриваются четыре **основные характеристики**:

- функция распределения **F (t)** ;
- плотность распределения ***f* (t)** ;
- математическое ожидание (средняя наработка до отказа) $\mathbf{T_0}$;
- дисперсия $\mathbf{DT_0}$.

Наработка до очередного отказа есть продолжительность или объем работы объекта.

Для систем обработки информации и управления естественно исчисление наработки в единицах времени, тогда как для других технических средств могут быть удобнее иные единицы измерения (например, наработка автомобиля - в километрах пробега).

Для невосстанавливаемых и восстанавливаемых изделий понятия наработки различаются:

- в первом случае подразумевается **наработка до первого отказа** (он же является и последним отказом),
- во втором случае – **наработка между двумя соседними во времени отказами** (после каждого отказа производится восстановление работоспособного состояния).

Вероятностные характеристики наработки до очередного отказа и являются показателями безотказности объекта.

Эти вероятностные характеристики определяются по результатам наблюдений за некоторым множеством экземпляров однотипных изделий, но используются в качестве показателя надежности каждого конкретного изделия.

Средняя наработка до отказа (между отказами) есть математическое ожидание наработки до очередного отказа (интегрирование осуществляется в пределах от 0 до ∞):

$$M[T] = \int t f(t)dt = T_0,$$

где через **t** обозначено текущее значение наработки, а $f(t)$ - плотность вероятности ее распределения.

Вероятностью безотказной работы P(t) называют вероятность того, что изделие будет работоспособно в течение заданной наработки при заданных условиях эксплуатации.

По статистическим данным об отказах, вероятность безотказной работы определяют по формуле:

$$P(t) = [N(0) - n(t)]/(N(0),$$

где **N(0)** – число изделий в начале наблюдений; n(t) – число изделий, отказавших за время t.

В начальный момент времени P(0) = 1. При увеличении времени вероятность P(t) монотонно уменьшается и для любых технических изделий асимптотически приближается к нулю.

Вероятность отказа F(t) есть вероятность того, что при заданных условиях эксплуатации в течение заданной наработки произойдет отказ.

Отказ и **безотказная работа** – противоположенные события. Поэтому **вероятность F(t) отказа** можно найти по формуле:

$$F(t) = 1 - P(t).$$

Частота отказов есть плотность распределения времени безотказной работы изделия:

$$f(t) = d\,F(t)/dt = d[1 - P(t)]\,/dt = -\,d\,P(t)/dt,$$

откуда очевидно, что **частота отказов** характеризует **скорость уменьшения вероятности безотказной работы во времени.**

Интенсивностью отказов λ(t) называют условную плотность вероятности возникновения отказа изделия при условии, что к моменту **t** отказ не возник:

$$\lambda(t) = f(t)/P(t) = -\,[1/P(t)]dP(t)/dt.$$

Функции **λ(t)** и **f(t)** измеряются в **$ч^{-1}$**.

Интегрируя **λ(t)** в пределах от 0 до t, можно получить:

$$P(t) = \exp[\,-\int\lambda(t)dt\,].$$

Это выражение, называемое **основным законом надежности**, позволяет установить временное изменение вероятности безотказной работы при любом характере изменения интенсивности отказов во времени.

В частном случае постоянства интенсивности отказов $\lambda(t) = \lambda = const$ равенство переходит в известное в теории вероятностей **экспоненциальное распределение:**

$P(t) = \exp(-\lambda t)$; $F(t) = 1 - \exp(-\lambda t)$; $f(t) = \lambda \exp(-\lambda t)$.

Экспоненциальное распределение однопараметрическое и обладает уникальными свойствами:

- **вероятность безотказной работы изделия $P(t) = \exp(-\lambda t)$** не зависит от того, сколько времени изделие проработало до рассматриваемого интервала времени;
- **средняя остаточная (предстоящая) наработка до отказа T_0** также не зависит от того, сколько времени проработало изделие ранее.

Эти закономерности являются проявлением свойства, называемого **отсутствием последствия**.

В этом случае показатели надежности изделия зависят только от состояния изделия в начале рассматриваемого интервала времени, но не зависят от наработки до этого интервала времени.

Существуют физические предпосылки, объясняющие это свойство. Одно из объяснений. Любое изделие работает в условиях определенной нагрузки (электрической, механической и пр.) и имеет ограниченную «прочность», поэтому существует некоторая предельная нагрузка, которую изделие способно выдержать без отказа. Если же нагрузка превосходит предельное значение, то наступает внезапный отказ. Пиковые значения нагрузок возникают случайным образом. В теории случайных процессов доказывается, что при определенных условиях время до первого пересечения случайным процессом некоторого порогового уровня имеет как раз экспоненциальное распределение.

В приведенном объяснении важным является то, что отказ возникает не вследствие постепенного изменения внутреннего состояния изделия, а вследствие внешнего воздействия, значение которого превышает допустимое.

Отсюда следует, что **при экспоненциальном распределении наработки до отказа профилактические работы, включающие в себя замену элементов или их периодический ремонт, теряют всякий смысл, так как не могут повлиять на причину отказа.**

Естественный путь повышения надежности изделия состоит либо в его конструктивном улучшении, либо в снижении действующих нагрузок.

Поток отказов при постоянстве интенсивности отказов **$\lambda(t)=const$** называется **простейшим потоком отказов,** и именно он реализуется для большинства **систем обработки информации и управления в течение периода нормальной эксплуатации от окончания приработки до начала старения и износа.**

Подставив выражение плотности вероятности $f(t) = \lambda \exp(-\lambda t)$ экспоненциального распределения, получим:

$$T_o = 1 / \lambda,$$

т.е. при простейшем потоке отказов средняя наработка T_0 обратно пропорциональна интенсивности отказов **λ**.

Можно показать, что за время средней наработки, $t = T_0$, вероятность безотказной работы изделия составляет 1/e**.**

2. ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

2.1. Определение вероятности и свойства, вытекающие из её определения. Классификация событий. Диаграммы Венна

Под вероятностью в широком смысле понимают количественную меру неопределенности. Это - число, которое выражает степень уверенности в наступлении того или иного *случайного события.* Например, нас может интересовать вероятность того, что объем продаж некоторого продукта не упадет, если цены вырастут, или вероятность того, что строительство нового дома завершится в срок.

Случайным называется событие, которое может произойти или не произойти в результате некоторого испытания. В дальнейшем для простоты мы будем опускать термин “случайный”.

Достоверное событие - это событие, которое обязательно произойдет в результате испытания. Например, если в урне содержатся только белые шары, то извлечение из нее белого шара есть событие достоверное; другой пример, если мы подбросим вверх камень, то он обязательно упадет на Землю в силу действия закона притяжения, то есть результат этого опыта заведомо известен. Достоверные события условимся обозначать символом Ω.

Невозможное событие - это событие, которое не может произойти в результате данного опыта (испытания). Извлечение черного шара из урны с белыми шарами есть событие невозможное; выпадение выигрыша на все номера облигаций в каком-либо тираже выигрышного займа также невозможное событие. Невозможное событие обозначим $\varnothing$.

Достоверные и невозможные события, вообще говоря, не являются случайными.

Совместные события. Несколько событий называются совместными, если в результате эксперимента наступление одного из них не исключает появления других. Например, при бросании трех монет выпадение цифры на одной не исключает появления цифр на других монетах.

В магазин вошел покупатель. События “в магазин вошел покупатель старше 60 лет” и “в магазин вошла женщина” - совместные, так как в магазин может войти женщина старше 60 лет.

Несовместные события. Несколько событий называются несовместными в данном опыте, если появление одного из них исключает появления других. Например, выигрыш, ничейный исход и проигрыш при игре в шахматы (одной партии) - три несовместных события.

Единственно возможные события. События называются единственно возможными, если в результате испытания хотя бы одно из них обязательно произойдет (или одно, или два, или ... или все события из рассматриваемой совокупности событий произойдут; одно точно произойдет). Например, некоторая фирма рекламирует свой товар по радио и в газете. Обязательно произойдет одно и только одно из следующих событий: “потребитель услышал о товаре по радио”, “потребитель прочитал о товаре в газете”, “потребитель получил информацию о товаре по радио и из газеты”, “потребитель не слышал о товаре по радио и не читал газеты”. Эти четыре события единственно возможные.

Равновозможные события. Несколько событий называются равновозможными, если в результате испытания ни одно из них не имеет объективно большую возможность появления, чем другие. При бросании игральной кости появление каждой из её граней - события равновозможные.

Противоположные события. Два единственно возможных и несовместных события называются противоположными. Купля и продажа определенного вида товара есть события противоположные.

Полная группа событий. Совокупность всех единственно возможных и несовместных событий называется полной группой событий.

Пример 2.1. Магазин в целях рекламы нового товара проводит лотерею, в которой 1 главный приз, 5 вторых призов, 100 третьих призов и 1000 четвертых призов. В конце рекламного дня выяснилось, что лотерейные билеты получили 10000 покупателей. По правилам розыгрыша, после извлечения выигрышного билета он не возвращается в урну, и покупатель не может получить более одного выигрыша. Чему равна вероятность того, что покупатель, который приобрел рекламируемый товар: а) выиграет первый приз; б) выиграет хотя бы один приз; в) не выиграет ни одного приза?

Решение. Определим событие *A*: «Покупатель выиграл первый приз». Согласно условию задачи в лотерее участвовало 10000 покупателей, отсюда общее число испытаний N = 10000, а число исходов, благоприятствующих событию *A,* M = 1. Все исходы являются равновозможными, единственно возможными и несовместными элементарными событиями. Следовательно, по формуле классической вероятности: P (A)=0,0001.

Соответственно, определим событие B: «Покупатель выиграл любой приз». Для этого события число благоприятствующих исходов M = =1+5+100+1000 = 1106.

$$P(B) = \frac{M}{N} = \frac{1106}{10000} = 0.1106 .$$

Событие «Покупатель не выиграет ни одного приза» - противоположное событию B: «Покупатель выиграет хотя бы один приз», поэтому обозначим его как $\overline{B}$. По формуле 2.3 найдем:

$$P(\overline{B}) = 1 - P(B) = 1 - 0{,}1106 = 0{,}8894 .$$

Ответ. Вероятность того, что покупатель выиграет первый приз, равна 0,0001; любой приз - 0,1106; ни одного приза - 0,8894.

Пример 2.2. Структура занятых в региональном отделении крупного банка имеет следующий вид:

	Женщины	Мужчины
Администрация	25	15
Операционисты	35	25

Если один из служащих выбран случайным образом, то какова вероятность, что он: 1. Мужчина-администратор? 2. Женщина-операционист? 3. Мужчина? 4. Операционист?

Решение.

1. В банке работают 100 человек, N = 100. Из них 15 – мужчины-администраторы, M = 15. Следовательно,

$$P(\text{мужчина - администратор}) = \frac{15}{100} = 0{,}15.$$

2. 35 служащих в банке – женщины-операционисты, следовательно,

$$P(\text{женщина - операционист}) = \frac{35}{100} = 0{,}35.$$

3. 40 служащих в банке – мужчины, следовательно,

$$P(\text{мужчина}) = \frac{40}{100} = 0{,}40.$$

4. Из общего числа служащих в банке 60 – операционисты, следовательно,

$$P(\text{операционист}) = \frac{60}{100} = 0{,}60.$$

Ответ. Вероятность того, что один из служащих:

1. мужчина – администратор – 0,15;
2. женщина – операционист – 0,35;
3. мужчина – 0,40;
4. операционист – 0,60.

Пример 2.3. Компания производит 40000 холодильников в год, которые реализуются в различных регионах России. Из них 10000 экспортируются в страны СНГ, 8000 продаются в регионах Европейской части России, 7000 продаются в страны дальнего зарубежья, 6000 в Западной Сибири, 5000 в Восточной Сибири, 4000 в Дальневосточном районе. Чему равна вероятность того, что определенный холодильник будет: 1. Произведен на экспорт? 2. Продан в России?

Решение. Обозначим события:

A – «Холодильник будет продан в странах СНГ»,

B – «Холодильник будет продан в Европейской части России»,

C – «Холодильник будет продан в страны дальнего зарубежья»,

D – «Холодильник будет продан в Западной Сибири»,

E – «Холодильник будет продан в Восточной Сибири»,

F – «Холодильник будет продан в Дальневосточном районе».

Соответственно:

вероятность того, что холодильник будет продан в странах СНГ:

$$P(A) = 10000/40000 = 0{,}25;$$

вероятность того, что холодильник будет продан в Европейской части России:

$$P(B) = 8000/40000 = 0{,}20;$$

вероятность того, что холодильник будет продан в страны дальнего зарубежья:

$$P(C) = 7000/40000 = 0{,}175;$$

вероятность того, что холодильник будет продан в Западной Сибири:

$$P(D) = 6000/40000 = 0{,}15;$$

вероятность того, что холодильник будет продан в Восточной Сибири:

$$P(E) = 5000/40000 = 0{,}125;$$

вероятность того, что холодильник будет продан на Дальнем Востоке:

$$P(F) = 4000/40000 = 0{,}10.$$

События A, B, C, D, E, F – несовместные.

1.Событие, состоящее в том, что холодильник произведен на экспорт, означает, что холодильник будет продан или в страны СНГ, или страны дальнего зарубежья. Отсюда находим его вероятность:

$$P(\text{холодильник произведен на экспорт})=$$
$$=P(A+B)=P(A)+P(B)=0{,}25+0{,}175=0{,}425.$$

2. Событие, состоящее в том, что холодильник будет продан в России, означает, что холодильник будет продан или в Европейской части России, или в Западной Сибири, или в Восточной Сибири, или на Дальнем Востоке. Отсюда находим его вероятность:

$$P(\text{холодильник будет продан в России})=$$
$$=P(A+D+E+F)=P(B)+P(D)+P(E)+P(F)=0{,}20+0{,}15+0{,}125+0{,}10=0{,}575.$$

Этот же результат можно было получить, рассуждая по-другому. События «Холодильник произведен на экспорт» и «Холодильник будет продан в России» – два взаимно противоположных события, отсюда:

$$P(\text{холодильник будет продан в России}) =$$
$$=1-P(\text{холодильник произведен на экспорт })=1–0{,}425=0{,}575.$$

Ответ: 1. P(холодильник произведен на экспорт) = 0,425,
2. P(холодильник будет продан в России) == 0,575.

Пример 2.4. Опыт состоит в случайном извлечении карты из колоды в 52 карты. Чему равна вероятность того, что это будет или туз, или карта масти треф?

Решение. Определим события: А - “извлечение туза”, В - “извлечение карты трефовой масти”. Вероятность извлечения туза из колоды карт $P(A)=4/52$; вероятность извлечения карты трефовой масти - $P(B) = 13/52$; вероятность их пересечения - извлечение трефового туза – $P(AB) = 1/52$.

События А и В - совместные, поскольку в колоде есть трефовый туз.

Согласно условию задачи, нас интересует вероятность суммы совместных событий А и В. Получим:

$$P(A + B) = P(A) + P(B) – P\ (AB) = 4/52 + 13/52 - 1/52 = 16/52 = 0{,}3077.$$

Ответ: Вероятность того, что случайно выбранная карта будет или туз, или масти трефа, равна 0,3077.

Пример 2.5. В урне два белых и три черных шара. Чему равна вероятность появления белого шара при первом извлечении из урны? При втором извлечении из урны?

Решение. Здесь возможны два случая.

Первый случай. Схема возвращенного шара, то есть шар после первого испытания возвращается в урну.

Пусть событие A - “появление белого шара при первом испытании”. Так как N = 5, а M = 2, то P(A) = 2/5.

Пусть событие B - “появление белого шара при втором испытании”. Так как шар после первого испытания возвратился в урну, то N = 5, а M = 2 и P(B)=2/5.

Таким образом, вероятность каждого из событий не зависит от того, произошло или не произошло другое событие. События A и B в этом случае называются **независимыми**.

Итак, события A и B называются независимыми, если вероятность каждого из них не зависит от того, произошло или нет другое событие.

Вероятности независимых событий называются **безусловными**.

Второй случай. Схема невозвращенного шара, то есть шар после первого испытания в урну не возвращается.

Вероятность появления белого шара при первом испытании P(A) = 2/5. Белый шар в урну не возвращается, следовательно, в урне остались один белый и три черных шара. Чему равна вероятность события B при условии, что событие A произошло? N = 4, M = 1.

Искомую вероятность обозначают P(B/A) или $P(B)_A$ или $P_A(B)$. Итак, P(B/A)=1/4 называют *условной вероятностью*, а события A, B называются зависимыми. В предыдущем примере с картами P(A)=4/52; P(A/B)=4/16.

Например, тот факт, что человек работает научным сотрудником, является зависимыми от наличия у него высшего образования; событие, состоящее в том, что станок может выйти из строя, не является независимым от срока его эксплуатации, событие, состоящее в том, что цена акций компании пошла вверх, не является независимым от того, с прибылью или с убытком сработала компания в прошлом периоде и т.д.

Итак, события A и B называются зависимыми, если вероятность каждого из них зависит от того, произошло или нет другое событие. Вероятность события B, вычисленная в предположении, что другое событие A уже произошло.

Пример 2.6. Консультационная фирма претендует на два заказа от двух крупных корпораций. Эксперты фирмы считают, что вероятность получения консультационной работы в корпорации A равна 0,45. Эксперты также полагают, что если фирма получит заказ у корпорации A, то вероятность того, что и корпорация B обратится к ним, равна 0,9. Какова вероятность того, что консультационная фирма получит оба заказа?

Решение. Обозначим события:

A - “получение консультационной работы в корпорации A”,

B - “получение консультационной работы в корпорации B”.

События A и B - зависимые, т.к. событие B зависит от того, произойдет или нет событие A.

По условию мы имеем: P(A) = 0,45, а также знаем, что P(B/A) = 0,9.

Необходимо найти вероятность того, что оба события (и событие A, и событие B) произойдут, т.е. P(AB). Для этого используем правило умножения вероятностей.

Отсюда получим:

$$P(A\ B) = P(A)\ P(B/A) = 0{,}45\ 0{,}9 = 0{,}405.$$

Ответ. Вероятность того, что фирма получит оба заказа, 0,405.

Пример 2.7. В большой рекламной фирме 21% работников получают высокую заработную плату. Известно также, что 40% работников фирмы - женщины, а 6,4% работников - женщины, получающие высокую заработную плату. Можем ли мы утверждать, что на фирме существует дискриминация женщин в оплате труда?

Решение. Сформулируем условие этой задачи в терминах теории вероятностей. Для ее решения необходимо ответить на вопрос: "Чему равняется вероятность того, что случайно выбранный работник будет женщиной, имеющей высокую заработную плату?" и сравнить ее с вероятностью того, что наудачу выбранный работник любого пола имеет высокую зарплату.

Обозначим события:

А - "случайно выбранный работник имеет высокую зарплату";

В - "случайно выбранный работник - женщина".

События А и В - зависимые.

По условию: $P(AB) = 0{,}064$; $P(B) = 0{,}4$; $P(A) = 0{,}21$.

Нас интересует вероятность того, что наудачу выбранный работник имеет высокую зарплату при условии, что это женщина, т.е. - условная вероятность события А.

Тогда, используя теорему умножения вероятностей, получим:

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} = \frac{0{,}064}{0{,}40} = 0{,}16.$$

Поскольку $P(A/B)=0{,}16$ меньше, чем $P(A)=0{,}21$, то мы можем заключить, что женщины, работающие в рекламной фирме, имеют меньше шансов получить высокую заработную плату по сравнению с мужчинами.

Ответ. На фирме существует дискриминация женщин в оплате труда.

Пример 2.8. Студент пришел на экзамен, изучив только 20 из 25 вопросов программы. Экзаменатор задал студенту три вопроса. Вычислить вероятность того, что студент ответит:

1. на все три вопроса;
2. хотя бы на один вопрос.

Решение. Обозначим события:

А -"студент знает все три вопроса";

A_1 - "студент знает первый вопрос";

A_2 - "студент знает второй вопрос";

A_3 - "студент знает третий вопрос".

По условию: $P(A_1) = 20/25$; $P(A_2/A_1) = 19/24$; $P(A_3/A_2 \cdot A_1) = 18/23$.

1. Искомое событие А состоит в совместном наступлении событий A_1, A_2, A_3.

События A_1, A_2, A_3 - зависимые.

Для решения задачи используем правило умножения вероятностей конечного числа *n* зависимых событий.

$$P(A)=\frac{20}{25}\cdot\frac{19}{24}\cdot\frac{18}{23}=\frac{57}{115}=0{,}496.$$

Вероятность того, что студент ответит на все три вопроса, равна 0,496.

2. Обозначим событие:

В -“студент ответит хотя бы на один вопрос”;

Событие В состоит в том, что произойдет или событие A_1, а события A_2 и A_3 - не произойдут, или произойдет событие A_2, а события A_1 и A_3 - не произойдут, или произойдет событие A_3, а события A_1 и A_2 - не произойдут, или произойдут события A_1 и A_2, а событие A_3 - не произойдет, или произойдут события A_1 и A_3, а событие A_2 - не произойдет, или произойдут события A_2 и A_3, а событие A_1 - не произойдет, или произойдут все три события A_1, A_2, A_3.

Для решения этой задачи можно было бы использовать правила сложения и умножения вероятностей. Однако здесь проще применить правило для вероятности наступления хотя бы одного из n зависимых событий:

Учитывая, что:

$$P(\overline{A}_1)=1-P(A_1)=1-\frac{20}{25}=\frac{5}{25};$$

$$P(\overline{A}_2/\overline{A}_1)=1-P(A_2/\overline{A}_1)=1-\frac{20}{24}=\frac{4}{24};$$

$$P(\overline{A}_3/\overline{A}_1\cdot\overline{A}_2)=1-P(A_3/\overline{A}_1\cdot\overline{A}_2)=1-\frac{20}{23}=\frac{3}{23},$$

получим:

$$P(B)=1-\frac{5}{25}\cdot\frac{4}{24}\cdot\frac{3}{23}=\frac{229}{230}\approx 0{,}9957.$$

Вероятность того, что студент ответит хотя бы на один вопрос, равна 0,9957.

Ответ. Вероятность того, что студент ответит на все три вопроса, равна 0,496.

Вероятность того, что студент ответит хотя бы на один вопрос, равна 0,9957.

Пример 2.9. Вероятность того, что потребитель увидит рекламу определенного продукта по телевидению, равна 0,04. Вероятность того, что потребитель увидит рекламу того же продукта на рекламном стенде, равна 0,06. Предполагается, что оба события - независимы. Чему равна вероятность того, что потребитель увидит: 1. обе рекламы; 2. хотя бы одну рекламу?

Решение. Обозначим события:

А -“ потребитель увидит рекламу по телевидению”;

В - “потребитель увидит рекламу на стенде”.

С - “потребитель увидит хотя бы одну рекламу”. Это значит, что потребитель увидит рекламу по телевидению, или на стенде, или по телевидению и на стенде.

По условию: P(A) = 0,04; P(B) = 0,06.

События А и В - совместные и независимые.

1. Поскольку вероятность искомого события есть вероятность совместного наступления независимых событий A и B (потребитель увидит рекламу и по телевидению, и на стенде), т.е. их пересечения, для решения задачи используем правило умножения вероятностей для независимых событий.

Отсюда:

$P(AB) = P(A) \cdot P(B) = 0{,}04 \cdot 0{,}06 = 0{,}0024.$

Вероятность того, что потребитель увидит обе рекламы, равна 0,0024.

2. Так как событие C состоит в совместном наступлении событий A и B, искомая вероятность может быть найдена с помощью правила сложения вероятностей.

$P(C) = P(A + B)= P(A) + P(B) - P(AB) = 0{,}04 + 0{,}06 - 0{,}0024 = 0{,}0976.$

Вместе с тем, при решении этой задачи может быть использовано правило о вероятности наступления хотя бы одного из n независимых событий:

Учитывая, что

$P(\overline{A}) = 1 - P(A) = 1 - 0{,}04 = 0{,}96$ и

$P(\overline{B}) = 1 - P(B) = 1 - 0{,}06 = 0{,}94$,

получим: $P(C) = 1 - P(\overline{A}) \cdot P(\overline{B}) = 1 - 0{,}96 \cdot 0{,}94 = 0{,}0976.$

Вычисление вероятностей событий такого типа характеризует эффективность рекламы, поскольку эта вероятность может означать долю (процент) населения, охватываемого, и отсюда следует оценка рекламных усилий.

Ответ. Вероятность того, что потребитель увидит обе рекламы, равна 0,024.

Вероятность того, что потребитель увидит хотя бы одну рекламу, равна 0,0976.

3. ФОРМУЛЫ ПОЛНОЙ ВЕРОЯТНОСТИ И БАЙЕСА

Часто мы начинаем анализ вероятностей, имея предварительные, **априорные** значения вероятностей интересующих нас событий. Затем из источников информации, таких как выборка, отчет, опыт и т.д., мы получаем дополнительную информацию об интересующем нас событии. Имея эту новую информацию, мы можем уточнить, пересчитать значения априорных вероятностей. Новые значения вероятностей для тех же интересующих нас событий будут уже **апостериорными** (послеопытными) вероятностями. Теорема Байеса дает нам правило для вычисления таких вероятностей.

Пусть событие A может осуществиться лишь вместе с одним из событий H_1, H_2, H_3, ..., H_n, образующих полную группу. Пусть известны вероятности $P(H_1)$, $P(H_2)$,…$P(H_i)$,…$P(H_n)$. Так как события H_i образуют полную группу, то $\sum_{i=1}^{n} P(H_i) = 1$. Также известны и условные вероятности события A: $P(A/H_1)$, $P(A/H_2)$, …$P(A/H_i)$…, $P(A/H_n)$, i=1, 2, …, n. Так как заранее неизвестно, с каким из событий H_i произойдет событие A, то события H_i называют гипотезами.

Необходимо определить вероятность события А и переоценить вероятности событий H_i с учетом полной информации о событии А.

Вероятность события А определяется как:

$$P(A)=\sum_{i=1}^{n}P(H_i)\cdot P(A/H_i). \tag{3.1}$$

Эта вероятность называется **полной вероятностью**.

Если событие А *может наступить только вместе с одним из событий* H_1, H_2, H_3, ..., H_n, *образующих полную группу несовместных событий и называемых гипотезами, то вероятность события* А *равна сумме произведений вероятностей каждого из событий* H_1, H_2, ..., H_n *на соответствующую условную вероятность события* А.

Условные вероятности гипотез вычисляются по формуле:

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)},$$

или (3.2)

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{\sum_{i=1}^{n}P(H_i)\cdot P(A/H_i)}.$$

Это - формулы Байеса (по имени английского математика Т.Байеса, опубликовавшего их в 1764 году), выражение в знаменателе - формула полной вероятности.

Пример 3.1. Предприятие, производящее компьютеры, получает одинаковые ЧИПы от двух поставщиков. Первый поставляет 65% ЧИПов, второй 35%. Известно, что качество поставляемых ЧИПов разное. Основываясь на предыдущих данных о рейтингах качества, составили следующую таблицу:

Поставщик	% качественной продукции	% брака
Поставщик 1	98	2
Поставщик 2	95	5

Предприятие осуществляет гарантийный ремонт компьютеров. Имея данные о числе компьютеров, поступающих на гарантийный ремонт в связи с неисправностью ЧИПов, переоцените вероятности того, что возвращенный для ремонта компьютер укомплектован ЧИПом от поставщика 1, поставщика 2.

Решение задач с использованием формул полной вероятности и Байеса удобнее оформлять в виде таблицы следующего вида:

Гипотезы H_i	Априорные вероятности $P(H_i)$	Условные вероятности $P(A/H_i)$	Совместные вероятности $P(H_i \cap A)$	Апостериорные вероятности $P(H_i/A)$
1	2	3	4	5

Шаг 1. В первой колонке перечисляем события, которые задают априорную информацию в контексте решаемой проблемы.

Соб. H_1 - ЧИП от первого поставщика;

Соб. H_2 - ЧИП от второго поставщика.

Это – гипотезы, и они образуют полную группу независимых и несовместных событий.

Во второй колонке записываем вероятности этих событий:

$P(H_1) = 0{,}65$, $P(H_2) = 0{,}35$.

В третьей колонке определим условные вероятности события А - «ЧИП бракованный» для каждой из гипотез.

Шаг 2. В колонке 4 находим вероятности для событий «ЧИП от первого поставщика и он бракованный» и «ЧИП от второго поставщика, и он бракованный». Они определяются по правилу умножения вероятностей путем перемножения значений колонок 2 и 3. Поскольку сформулированные события являются результатом пересечения двух событий: А и Hi, то их называют совместными вероятностями, то есть $P(H_i \cap A) = P(H_i) \cdot P(A/H_i)$.

Шаг 3. Суммируем вероятности в колонке 4 для того, чтобы найти вероятность события А. В нашем примере 0,0130 - вероятность поставки некачественного ЧИПа от поставщика 1, 0,0175 - вероятность поставки некачественного ЧИПа от поставщика 2. Поскольку, как мы уже сказали выше, ЧИПы поступают только от двух поставщиков, то сумма вероятностей 0,0130 и 0,0175 показывает, что 0,0305 есть вероятность бракованного ЧИПа в общей поставке, по формуле (3.1):

$$P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i) = 0{,}0130 + 0{,}0175 = 0{,}0305.$$

Шаг 4. В колонке 5 вычисляем апостериорные вероятности, используя формулу (3.2)

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0{,}65 \cdot 0{,}02}{0{,}0305} = 0{,}426,$$

$$P(H_2/A) = \frac{P(H_2)P(A/H_2)}{P(A)} = \frac{0{,}35 \cdot 0{,}05}{0{,}0305} = 0{,}574.$$

Заметим, что совместные вероятности находятся в строках колонки 4, а вероятность события А как сумма колонки 4.

Гипотезы H_i	Априорные вероятности $P(H_i)$	Условные вероятности $P(A/H_i)$	Совместные вероятности $P(H_i \cap A)$	Апостериорные вероятности $P(H_i/A)$
1	2	3	4	5
ЧИП от 1-го поставщика	0,65	0,02	0,0130	0,426
ЧИП от 2-го поставщика	0,35	0,05	0,0175	0,574
	1	-	P(A) = 0,305	1

Пример 3.2. Экономист полагает, что вероятность роста стоимости акций некоторой компании в следующем году равна 0,75, если экономика страны будет на подъёме; и эта же вероятность равна 0,30, если экономика страны не будет успешно развиваться. По его мнению, вероятность экономического подъёма в будущем году равна 0,80. Используя предположения экономиста,

оцените вероятность того, что *акции компании поднимутся в цене в следующем году*?

Решение. Определим события:

А - “акции компании поднимутся в цене в будущем году”.

Событие А - “акции компании поднимутся в цене в будущем году” - может произойти только вместе с одной из гипотез:

H_1 - экономика страны будет на подъёме;

H_2 - экономика страны не будет успешно развиваться.

По условию известны вероятности гипотез:

$P(H_1) = 0,8$; $P(H_2) = 0,2$

и условные вероятности события А:

$P(A/H_1) = 0,75$; $P(A/H_2) = 0,3$.

Гипотезы образуют полную группу, сумма их вероятностей равна 1. Рассмотрим событие А - это или H_1A, или H_2A. События H_1A и H_2A - несовместные попарно, так как события H_1 и H_2 - несовместны.

События H_1 и А, H_2 и А - зависимые.

Вышеизложенное позволяет применить для определения искомой вероятности события А формулу полной вероятности:

$P(A)=P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) = 0,8 \cdot 0,75 + 0,2 \cdot 0,3 = 0,66$.

Оформим решение в рабочей таблице:

Гипотезы H_i	$P(H_i)$	$P(A/H_i)$	$P(H_i) \cdot P(A/H_i)$
H_1 - “подъем экономики”	0,80	0,75	0,60
H_2- “спад экономики”	0,20	0,30	0,06
□	1,00	-	0,66

Ответ. Вероятность того, что акции компании поднимутся в цене в следующем году, составляет 0,66.

Пример 3.3. Экономист полагает, что в течение периода активного экономического роста американский доллар будет расти в цене с вероятностью 0,7, в период умеренного экономического роста доллар подорожает с вероятностью 0,4 и при низких темпах экономического роста доллар подорожает с вероятностью 0,2. В течение любого периода времени вероятность активного экономического роста 0,3, умеренного экономического роста 0,5 и низкого роста - 0,2. Предположим, что доллар дорожает в течение текущего периода. Чему равна вероятность того, что анализируемый период совпал с периодом активного экономического роста?

*Решение***.** Определим события:

А - “доллар дорожает”. Оно может произойти только вместе с одной из гипотез:

H_1 - “активный экономический рост”;

H_2 - “умеренный экономический рост”;

H_3 - “низкий экономический рост”.

По условию известны доопытные (априорные) вероятности гипотез и условные вероятности события А:

$P(H_1) = 0,3$; $P(H_2) = 0,5$; $P(H_3) = 0,2$;

$P(A/H_1) = 0,7$; $P(A/H_2) = 0,4$ и $P(A/H_3) = 0,2$.

Гипотезы образуют полную группу, сумма их вероятностей равна 1. Событие А - это или $H_1 \cdot A$, или $H_2 \cdot A$, или $H_3 \cdot A$. События $H_1 \cdot A$, $H_2 \cdot A$ и $H_3 \cdot A$ - несовместные попарно, так как события H_1 , H_2 и H_3 - несовместны.

События H_1 и А, H_2 и А, H_3 и А - зависимые.

По условию требуется найти уточненную (послеопытную, апостериорную) вероятность первой гипотезы, т.е. необходимо найти вероятность активного экономического роста, при условии, что доллар дорожает (событие А уже произошло), то есть $P(H_1/A)$ - ?

Используя формулу Байеса (3.2) и подставляя заданные значения вероятностей, имеем:

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + P(H_3) \cdot P(A/H_3)} =$$

$$= \frac{0,3 \cdot 0,7}{0,3 \cdot 0,7 + 0,5 \cdot 0,4 + 0,2 \cdot 0,2} = 0,467.$$

Мы можем получить тот же результат при помощи таблицы:

Гипотезы H_i	Априорные вероятности $P(H_i)$	Условные вероятности $P(A/H_i)$	Совместные вероятности $P(A \cap H_i)$	Апостериорные вероятности $P(H_i/A)$
H_1	0,30	0,70	0,21	0,21 / 0,45 = 0,467
H_2	0,50	0,40	0,20	0,20 / 0,45 = 0,444
H_3	0,20	0,20	0,04	0,04 / 0,45 = 0,089
	1,00	-	0,45	1

Ответ. Вероятность активного экономического роста, при условии, что доллар дорожает, составляет 0,467.

Пример 3.4. В каждой из двух урн содержится 6 черных и 4 белых шара. Из первой урны во вторую наудачу переложен один шар.

1. Найти вероятность того, что *шар, извлеченный из второй урны после перекладывания, окажется черным*?

2. Предположим, что шар, извлеченный из второй урны после перекладывания, оказался черным. Какова тогда вероятность того, что из *первой урны во вторую был переложен белый шар*?

Решение. Определим события:

А - “шар, извлеченный из второй урны, - черный”. Оно может произойти только вместе с одной из гипотез:

H_1 - “из 1-й урны во 2-ю урну переложили черный шар” и

H_2 - “из 1-й урны во 2-ю урну переложили белый шар”.

Используя классическое определение вероятностей, определим вероятности гипотез:

$$P(H_1) = 6/10;\ P(H_2) = 4/10$$

и условные вероятности события А.

После перекладывания во второй урне окажется 11 шаров. Если из первой урны во вторую переложили черный шар, то во второй урне окажется 7 черных и 4 белых шаров.

Тогда $P(A/H_1) = 7/11$.

Если из первой урны во вторую переложили белый шар, то во второй урне окажется 6 черных и 5 белых шаров.

Тогда $P(A/H_2) = 6/11$.

Гипотезы образуют полную группу, сумма их вероятностей равна 1. Рассмотрим событие А - это (или H_1A, или H_2A). События H_1A и H_2A - несовместные попарно, так как события H_1 и H_2 - несовместны.

События H_1 и А, H_2 и А - зависимые.

1. Вышеизложенное позволяет применить для определения вероятности события А и ответа на первый вопрос формулу полной вероятности (3.1):

$P(A)=P(H_1)\ P(A/H_1) + P(H_2)\ P(A/H_2)=6/10\quad 7/11 + 4/10\quad 6/11 = 0{,}6$.

Это же решение можно оформить в рабочей таблице:

Гипотезы H_i	$P(H_i)$	$P(A/H_i)$	$P(H_i)\ P(A/H_i)$
H_1 - "из 1-й урны во 2-ю урну переложили черный шар"	6/10	7/11	42/110
H_2- "из 1-й урны во 2-ю урну переложили белый шар"	4/10	6/11	24/110
	1,00	-	0,6

Ответ. Вероятность того, что шар, извлеченный из второй урны после перекладывания, окажется черным, составляет 0,6.

2. Во второй части задачи предполагается, что событие А уже произошло, т.е. шар, извлеченный из второй урны, оказался черным. Требуется найти уточненную (послеопытную, апостериорную) вероятность второй гипотезы, т.е. необходимо найти вероятность того, что из первой урны во вторую был переложен белый шар при условии, что шар, извлеченный из второй урны после перекладывания, оказался черным.

$P(H_2/A)$ - ?

Для определения искомой вероятности воспользуемся формулой Байеса (3.2):

$$P(H_2\ /\ A) = \frac{P(H_2) \cdot P(A\ /\ H_2)}{P(A)} = \frac{4\ /\ 10 \cdot 6\ /\ 11}{0{,}6} = 0{,}3636\,.$$

Мы можем получить тот же результат при помощи таблицы:

Гипотезы H_i	Априорные вероятности $P(H_i)$	Условные вероятности $P(A/H_i)$	Совместные вероятности $P(A \cap H_i)$	Апостериорные вероятности $P(H_i/A)$
H_1	6/10	7/11	42/110=0,3818	0,3818/0,6 = 0,6364
H_2	4/10	6/11	24/110=0,2182	0,2182/0,6 = 0,3636
	1,00	-	0,6	1

Ответ. Вероятность того, что из первой урны во вторую был переложен белый шар, при условии, что шар, извлеченный из второй урны после перекладывания, оказался черным, составляет 0,3636.

4. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

4.1. Определение дискретной случайной величины

Величина, которая в результате испытания может принять то или иное значение, заранее неизвестно, какое именно, называется *случайной величиной.*

Дискретной случайной величиной называется такая переменная величина, которая может принимать конечную или бесконечную совокупность значений, причем принятие ею каждого из значений есть случайное событие с определенной вероятностью.

Соотношение, устанавливающее связь между отдельными возможными значениями случайной величины и соответствующими им вероятностями, называется законом распределения дискретной случайной величины. Если обозначить возможные числовые значения случайной величины X через x_1, x_2, ... x_n..., а через $p_j = P(X = x_i)$ – вероятность появления значения x_i, то дискретная случайная величина полностью определяется следующей таблицей 4.1:

Таблица 4.1

x_i	x_1	x_2	...	x_n
p_j	p_1	p_2	...	p_n

где значения x_1, x_2,..., x_n записываются, как правило, в порядке возрастания. Таблица называется *законом* или *рядом распределения* дискретной случайной величины X. Поскольку в верхней строчке ряда распределения записаны все значения случайной величины X, то нижняя строчка обладает тем свойством, что

$$p_1 + p_2 + ... + p_n = \sum_{i=1}^{n} p_i = 1. \qquad (4.1)$$

Вероятность попадания случайной величины X в промежуток от α до β (включая α) выражается формулой:

$$P(\alpha \le X < \beta) = F(\beta) - F(\alpha). \qquad (4.2)$$

4.2.Числовые характеристики

Математическим ожиданием дискретной случайной величины называется:

$$M(X) = x_1p_1 + x_2p_2 + ... + x_np_n = \sum_{i=1}^{n} x_i * p_i \qquad (4.3)$$

В случае бесконечного множества значений x_i в правой части (4.4) находится ряд, и мы будем рассматривать только те значения X, для которых этот ряд абсолютно сходится.

M(X) представляет собой среднее ожидаемое значение случайной величины. Оно обладает следующими свойствами:

1) M(C)=C, где C=const;

2) M (CX)=CM (X); (4.4)

3) M (X+Y)=M(X)+M(Y), для любых X и Y;

4) M (XY)=M (X)M(Y), если X и Y независимы.

Для оценки степени рассеяния значений случайной величины около ее среднего значения M(X)=a вводятся понятия *дисперсии* D(X) и среднего квадратического (стандартного) отклонения $\sigma(x)$. *Дисперсией* называется математическое ожидание квадрата разности (X-$\boldsymbol{a}$), т.е. :

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

где $\boldsymbol{a}$=M(X)**;** если X и У независимы.

Размерность величин $M(X)$ и $\sigma(X)$ совпадает с размерностью самой случайной величины X, а размерность D(X) равна квадрату размерности случайной величины X.

4.3. Распределения Бернулли и Пуассона

Рассмотрим последовательность n идентичных повторных испытаний, удовлетворяющих следующим условиям:

1. Каждое испытание имеет два исхода, называемые успех и неуспех.

Эти два исхода - взаимно несовместные и противоположные события.

2. Вероятность успеха, обозначаемая p, остается постоянной от испытания к испытанию. Вероятность неуспеха обозначается q.

3. Все n испытаний - независимы. Это значит, что вероятность наступления события в любом из n повторных испытаний не зависит от результатов других испытаний.

Вероятность того, что в n независимых повторных испытаниях, в каждом из которых вероятность появления события равна $p(0<p<1)$, событие наступит ровно m раз (в любой последовательности), равна

$$P_{n,m}=C_n^m p^m q^{k-m}, \tag{4.5}$$

где q=1-p.

Выражение (4.5) называется формулой Бернулли. Вероятности того, что событие наступит:

а) менее m раз,

б) более m раз,

в) не менее m раз,

г) не более m раз - находятся соответственно по формулам:

а)$P_{n;0} + P_{n;1} + ... + P_{n;m-1}$,

б)$P_{n;m+1} + P_{n;m+2} + ... + P_{n;n}$,

в)$P_{n;m} + P_{n;m+1} + ... + P_{n;n}$,

г)$P_{n;0} + P_{n;1} + ... + P_{n;m}$.

Биномиальным называют **закон** распределения дискретной случайной величины X - числа появлений события в n независимых испытаниях, в каждом из которых вероятность наступления события равна p; вероятности возможных значений X = 0,1,2,..., m,...,n вычисляются по формуле Бернулли (таблица 4.2).

Таблица 4.2

Число успехов X=m	0	1	2	...	m	...	n
Вероятность $P_{n;m}$	$C_n^0 p^0 q^n =$ $= P_{n;0}$	$C_n^1 p^1 q^{n-1} =$ $= P_{n,1}$	$C_n^2 p^2 q^{n-2} =$ $= P_{n;2}$	...	$C_n^m p^m q^{n-m} =$ $= P_{n;m}$	...	$C_n^n p^n q^0 =$ $= P_{n;n}$

Так как правая часть формулы (4.5) представляет общий член биноминального разложения $(q+p)^n$, то этот закон распределения называют *биномиальным*. Для случайной величины X, распределенной по биноминальному закону, имеем:

$$M(X)=np, \tag{4.6}$$

$$D(X)=npq. \tag{4.7}$$

Если число испытаний велико, а вероятность появления события p в каждом испытании очень мала, то вместо формулы (4.5) пользуются приближенной формулой:

$$P_{n;m} \approx \frac{\lambda^m e^{-\lambda}}{m!}, \tag{4.8}$$

где m - число появлений события в n независимых испытаниях, $\lambda = np$ (среднее число появлений события в n испытаниях).

Выражение (4.8) называется формулой Пуассона. Придавая m целые неотрицательные значения m=0,1,2,...,n, можно записать ряд распределения вероятностей, вычисленных по формуле (4.8), который называется **законом распределения Пуассона** (таблица 4.3):

Таблица 4.3

M	0	1	2	...	m	...	n
$P_{n;m}$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	...	$\frac{\lambda^m e^{-\lambda}}{m!}$	...	$\frac{\lambda^n e^{-\lambda}}{n!}$

Распределение Пуассона часто используется, когда мы имеем дело с числом событий, появляющихся в промежутке времени или пространства. Например, число машин, прибывших на автомойку в течение часа, число дефектов на новом отрезке шоссе длиной в 10 километров, число мест утечки воды на 100 километров водопровода, число остановок станков в неделю, число дорожных происшествий.

Если распределение Пуассона применяется вместо биномиального распределения, то n должно иметь порядок не менее нескольких десятков, лучше нескольких сотен, а np< 10.

Математическое ожидание к дисперсии случайной величины, распределенной по закону Пуассона, совпадают и равны параметру λ, которая определяет этот закон, т.е.

$$M(X)=D(X)=n\cdot p=\lambda . \tag{4.9}$$

Пример 4.1. Известно, что в определенном городе 20% горожан предпочитают добираться на работу личным автотранспортом. Случайно выбраны 4 человека.

а) Составьте ряд распределения числа людей в выборке, предпочитающих добираться на работу личным автотранспортом, и постройте его график.

б) Найдите числовые характеристики этого распределения.

в) Напишите функцию распределения числа людей в выборке, предпочитающих добираться на работу личным автотранспортом, и постройте её график.

г) Чему равна вероятность того, что среди 4-х случайно отобранных человек *не будет ни одного, предпочитающего добираться на работу личным автотранспортом*?

д) Чему равна вероятность того, что среди 4-х случайно отобранных людей *окажется хотя бы один, предпочитающий добираться на работу личным автотранспортом*?

е) Чему равна вероятность того, что среди 4-х случайно отобранных человек *будет не больше двух, предпочитающих добираться на работу личным автотранспортом*?

Решение. В качестве случайной величины в данной задаче выступает число людей в выборке, предпочитающих добираться на работу личным автотранспортом. Обозначим ее через X.

Перечислим все возможные значения случайной величины X: 0, 1, 2, 3, 4.

Вероятность того, что каждый из отобранных людей предпочитает добираться на работу личным автотранспортом, - постоянна и равна 0,2 (p=0,2). Вероятность противоположного события, т.е. того, что каждый из отобранных людей предпочитает добираться на работу не личным автотранспортом, а как-то иначе - также постоянна и составляет 0,8 (q = 1 - p = 1 0,2 = 0,8).

Все 4 испытания - независимы, т.е. вероятность того, что каждый из отобранных людей предпочитает добираться на работу личным автотранспортом, не зависит от того, каким способом предпочитает добираться на работу любой другой человек из числа случайно отобранных.

Очевидно, что случайная величина X - подчиняется биномиальному закону распределения вероятностей с параметрами n=4 и p=0,2.

Итак, по условию задачи: n = 4; p = 0,2; q = 0,8; X = m.

а) Чтобы построить ряд распределения, необходимо вычислить вероятности того, что случайная величина примет каждое из своих возможных значений, и записать полученные результаты в таблицу.

Расчет искомых вероятностей осуществляется по формуле Бернулли.

$$P=(X=m)=P_{n,m}=C_n^m\cdot p^m\cdot q^{n-m}=\frac{n!}{m!(n-m)!}\cdot p^m\cdot q^{n-m}.$$

Подставим в эту формулу данные задачи:

$$P(X=0)=P_{4,0}=C_4^0\cdot 0{,}2^0\cdot 0{,}8^{4-0}=\frac{4!}{0!\cdot(4-0)!}\cdot 0{,}2^0\cdot 0{,}8^{4-0}=$$

$$=1\cdot 0{,}2^0\cdot 0{,}8^{4-0}=0{,}4096;$$

$$P(X=1)=P_{4,1}=C_4^1\cdot 0{,}2^1\cdot 0{,}8^{4-1}=\frac{4!}{1!\cdot(4-1)!}\cdot 0{,}2^1\cdot 0{,}8^{4-1}=$$

$$=4\cdot 0{,}2^1\cdot 0{,}8^{4-1}=0{,}4096;$$

$$P(X=2)=P_{4,2}=C_4^2\cdot 0{,}2^2\cdot 0{,}8^{4-2}=\frac{4!}{2!\cdot(4-2)!}\cdot 0{,}2^2\cdot 0{,}8^{4-2}=$$

$$=6\cdot 0{,}2^2\cdot 0{,}8^{4-2}=0{,}1536;$$

$$P(X=3)=P_{4,3}=C_4^3\cdot 0{,}2^3\cdot 0{,}8^{4-3}=\frac{4!}{3!\cdot(4-3)!}\cdot 0{,}2^3\cdot 0{,}8^{4-3}=$$

$$=4\cdot 0{,}2^3\cdot 0{,}8^{4-3}=0{,}0256;$$

$$P(X=4)=P_{4,4}=C_4^4\cdot 0{,}2^4\cdot 0{,}8^{4-4}=\frac{4!}{4!\cdot(4-4)!}\cdot 0{,}2^4\cdot 0{,}8^{4-4}=$$

$$=1\cdot 0{,}2^4\cdot 0{,}8^{4-4}=0{,}0016.$$

Получим ряд распределения числа людей в выборке, предпочитающих добираться на работу личным автотранспортом (таблица 4.4):

Таблица 4.4

X	0	1	2	3	4
P	0,4096	0,4096	0,1536	0,0256	0,0016

Так как все возможные значения случайной величины образуют полную группу событий, то сумма их вероятностей должна быть равна 1.

Проверка: 0,4096 + 0,4096 + 0,1536 + 0,0256 + 0,0016 = 1.

Вместо ряда распределения дискретная случайная величина может быть задана графически многоугольником (полигоном) распределения (рис. 4.1).

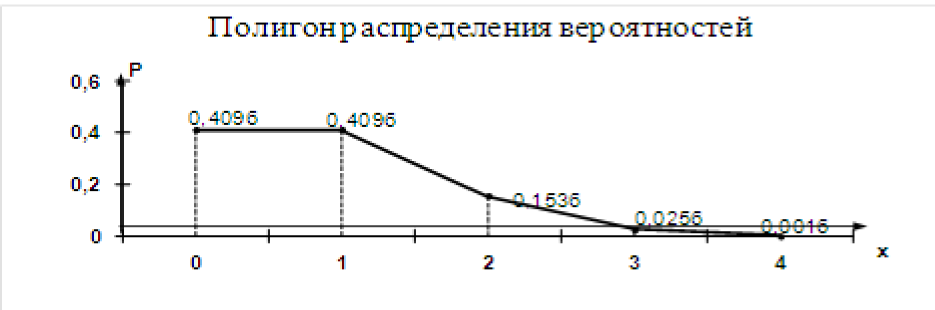


Рис. 4.1

б) Найдем основные числовые характеристики распределения данной случайной величины: математическое ожидание, дисперсию и среднее квадратическое (стандартное) отклонение.

Математическое ожидание любой дискретной случайной величины может быть рассчитано по формуле (4.4):

$$M(X)=\sum_{i=1}^{n} x_i \cdot p_i = 0 \cdot 0{,}4096 + 1 \cdot 0{,}4096 + 2 \cdot 0{,}1536 + 3 \cdot 0{,}0256 + 4 \cdot 0{,}0016 =$$

$$=0{,}8 \text{ (чел.)}.$$

Вместе с тем, ввиду того, что в данном случае речь идет о математическом ожидании частоты, для его расчета можно воспользоваться более простой формулой (4.10):

$$M(X = m) = np = 4 \cdot 0{,}2 = 0{,}8 \text{ (чел.)}.$$

Рассчитаем дисперсию числа человек, предпочитающих добираться на работу личным автотранспортом, среди 4-х отобранных. Дисперсия любой дискретной случайной величины может быть рассчитана по формуле :

$$D(X)=\sum_{i=1}^{n}[x_i - M(X)]^2 \cdot p_i = (0 - 0{,}8)^2 \cdot 0{,}4096 + (1 - 0{,}8)^2 \cdot 0{,}4096 + (2 - 0{,}8)^2 \times$$

$$\times 0{,}1536 + (3 - 0{,}8)^2 \cdot 0{,}0256 + (4 - 0{,}8)^2 \cdot 0{,}0016 = 0{,}64 \text{ (кв.ед)}.$$

В данном случае речь идет о дисперсии частоты, а её можно найти по формуле:

$$D(X = m) = npq = 4 \cdot 0{,}2 \cdot 0{,}8 = 0{,}64 \text{ (кв.ед)}.$$

Рассчитаем среднее квадратическое отклонение числа людей в выборке, предпочитающих добираться на работу личным автотранспортом. Среднее квадратическое отклонение рассчитывается по формуле:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}64} = 0{,}8 \ .$$

в) Дискретную случайную величину можно задать функцией распределения:

$$F(X) = P(X < x) = \sum_{\substack{i \\ x_i < x}} P(X = x_i)\,,$$

где для каждого значения x суммируются вероятности тех значений x_i , которые лежат левее точки x.

Зададим функцию распределения дискретной случайной величины применительно к условию данной задачи:

$$F(X = m) = P(X < x) = \sum_{\substack{i \\ x_i < x}} C_4^m \cdot 0{,}2^m \cdot 0{,}8^{4-m}\ .$$

Для построения графика функции распределения вероятностей дискретной случайной величины необходимо рассчитать кумулятивные (накопленные) вероятности, соответствующие значениям случайной величины. Алгоритм их расчета вытекает из смысла функции распределения:

$$F(X_i) = P(X_1) + P(X_2) + \ldots + P(X_{i-2}) + P(X_{i-1}).$$

Эта формула справедлива для всех F(Xi), кроме F(X0). Так как по определению функция распределения определяет вероятность того, что случайная величина примет значение, меньшее заданного, понятно, что вероятность того, что случайная величина примет значение, не более минимального, равна 0:

$$F(X_0) = 0.$$

Рассчитаем значения F(x):

$$F(x) = \begin{cases} 0 \text{ при } x \le 0 \\ 0{,}4096 \text{ при } 0 < x \le 1 \\ 0{,}8192 \text{ при } 1 < x \le 2 \\ 0{,}9728 \text{ при } 2 < x \le 3 \\ 0{,}9984 \text{ при } 3 < x \le 4 \\ 1 \text{ при } x > 4 \end{cases}$$

Эти данные можно представить и в виде таблицы 4.5:

Таблица 4.5

X	$x \le 0$	$0 < x \le 1$	$1 < x \le 2$	$2 < x \le 3$	$3 < x \le 4$	$x > 4$
F(x)	0	0.4096	0.8192	0.9728	0.9984	1

График функции распределения вероятностей дискретной случайной величины имеет ступенчатый вид. Скачки равны вероятностям, с которыми случайная величина принимает возможные значения (рис. 4.2).

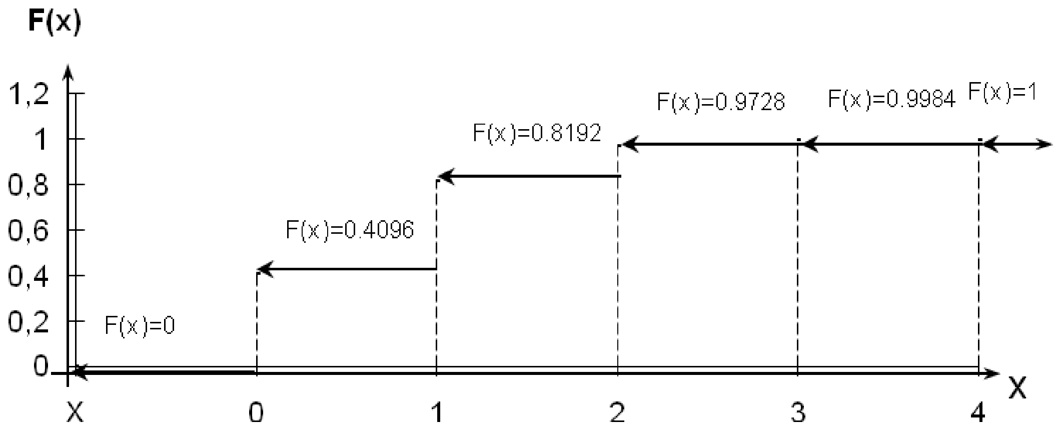


Рис. 4.2

График функции распределения вероятностей дискретной случайной величины

г) Определим вероятность того, что среди 4-х случайно отобранных человек *не будет ни одного человека, предпочитающего добираться на работу личным автотранспортом.*

$$P(X = 0) = 0{,}4096.$$

Вероятность того, что среди 4-х случайно отобранных человек *не будет ни одного, предпочитающего добираться на работу личным автотранспортом,* составляет 0,4096.

д) Определим вероятность того, что среди 4-х случайно отобранных человек *будет хотя бы один человек, предпочитающий добираться на работу личным автотранспортом.*

"Хотя бы один" - "как минимум один" - "один или больше". Другими словами, "хотя бы один" - это "или один, или два, или три, или четыре".

Исходя из этого, для определения вероятности того, что среди 4-х случайно отобранных человек будет хотя бы один, предпочитающий добираться на работу личным автотранспортом, можно использовать теорему сложения вероятностей несовместных событий:

$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$,

$P(X \geq 1) = 0{,}4096 + 0{,}1536 + 0{,}0256 + 0{,}0016 = 0{,}5904$.

С другой стороны, все возможные значения случайной величины образуют полную группу событий, а сумма их вероятностей равна 1. По отношению к событию ($X \geq 1$) до полной группы событий не хватает события ($X = 0$), которое является противоположным событию ($X \geq 1$). Поэтому искомую вероятность того, среди 4-х случайно отобранных человек будет хотя бы один, предпочитающий добираться на работу личным автотранспортом, проще найти следующим образом:

$P(X \geq 1) + P(X < 1) = 1$, откуда

$P(X \geq 1) = 1 - P(X = 0) = 1 - 0{,}4096 = 0{,}5904$.

Вероятность того, что среди 4-х случайно отобранных человек *будет хотя бы один человек, предпочитающий добираться на работу личным автотранспортом,* составляет 0,5904.

е) Определим вероятность того, что среди 4-х случайно отобранных человек *будет не больше двух, предпочитающих добираться на работу личным автотранспортом.*

"Не больше двух" - "два или меньше", т.е. "или ноль, или один, или два".

Используем теорему сложения вероятностей несовместных событий:

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$,

$P(X \leq 2) = 0{,}4096 + 0{,}4096 + 0{,}1536 = 0{,}9728$.

Вероятность того, что среди 4-х случайно отобранных человек будет не больше двух, предпочитающих добираться на работу личным автотранспортом, составляет 0,9728.

Пример 4.2. Среднее число инкассаторов, прибывающих утром на автомобиле в банк в 15-минутный интервал, равно 2. Прибытие инкассаторов происходит случайно и независимо друг от друга.

а) Составьте ряд распределения числа инкассаторов, прибывающих утром на автомобиле в банк в течение 15 минут;

б) Найдите числовые характеристики этого распределения;

в) Напишите функцию распределения числа инкассаторов, прибывающих утром на автомобиле в банк в течение 15 минут, и постройте её график;

г) Определите, чему равна вероятность того, что в течение 15 минут в банк прибудут на автомобиле *хотя бы два инкассатора;*

д) Определите вероятность того, что в течение 15 минут число прибывших инкассаторов окажется *меньше трех.*

Решение. Пусть случайная величина X - число инкассаторов, прибывающих утром на автомобиле в банк в течение 15 минут. Перечислим все возможные значения случайной величины X: 0, 1, 2, 3, 4, 5, ... , n.

Это - дискретная случайная величина, т.к. ее возможные значения отличаются друг от друга не менее чем на 1, и множество ее возможных значений является счетным.

По условию прибытие инкассаторов происходит случайно и независимо друг от друга. Следовательно, мы имеем дело с независимыми испытаниями.

Если мы предположим, что вероятность прибытия инкассаторов на автомобиле одинакова в любые два периода времени равной длины и что прибытие или неприбытие автомобиля в любой период времени не зависит от прибытия или неприбытия в любой другой период времени, то последовательность прибытия инкассаторов в банк может быть описана распределением Пуассона.

Итак, случайная величина X - число инкассаторов, прибывающих утром на автомобиле в течение 15 минут, подчиняется распределению Пуассона. По условию задачи: $\lambda = np = 2$; $X = m$.

а) Составим ряд распределения.

Вычислим вероятности того, что случайная величина примет каждое из своих возможных значений, и запишем полученные результаты в таблицу.

Так как данная случайная величина X подчинена распределению Пуассона, расчет искомых вероятностей осуществляется по формуле Пуассона (4.8).

Найдем по этой формуле вероятность того, что в течение 15 минут утром на автомобиле прибудет 0 инкассаторов:

$$P(X=0) = \frac{2^0}{0!} \cdot e^{-2} = 0{,}1353.$$

Однако расчет вероятностей распределения Пуассона легче осуществлять, пользуясь специальными таблицами вероятностей распределения Пуассона. В этих таблицах содержатся значения вероятностей при заданных m и (Приложение 6).

По условию $\lambda = 2$, а m изменяется от 0 до n.

Воспользовавшись таблицей распределения Пуассона, получим:

$P(X = 0) = 0{,}1353$; $P(X = 1) = 0{,}2707$;
$P(X = 2) = 0{,}2707$; $P(X = 3) = 0{,}1804$;
$P(X = 4) = 0{,}0902$; $P(X = 5) = 0{,}0361$;
$P(X = 6) = 0{,}0120$; $P(X = 7) = 0{,}0034$;
$P(X = 8) = 0{,}0009$; $P(X = 9) = 0{,}0002$.

Данных для $\lambda = 2$, и m > = 10 в таблице нет, что указывает на то, что эти вероятности составляют менее 0, 0001, т.е.

$P(X = 10) \approx 0$. Понятно, что $P(X = 11)$ еще меньше отличается от 0.

Занесем полученные результаты в таблицу 4.6:

Таблица 4.6

X	0	1	2	3	4	5	6	7	8	9	10
P(X)	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120	0,0034	0,0009	0,0002	0,0000

Так как все возможные значения случайной величины образуют полную группу событий, сумма их вероятностей должна быть равна 1.

Проверим: 0,1353 + 0,2707 + 0,2707 + 0,1804 + 0,0902 + 0,0361 + 0,0120 + + 0,0034 + 0,0009 + 0,0002 = 0,9999 ≈ 1.

График полученного ряда распределения дискретной случайной величины X – полигон распределения вероятностей:

Рис. 4.3

б) Найдем основные числовые характеристики полученного распределения случайной величины X.

Можно рассчитать математическое ожидание, дисперсию и среднее квадратическое отклонение по общим для любой дискретной случайной величины формулам.

Математическое ожидание случайной величины, подчиняющейся распределению Пуассона, может быть рассчитано и по формуле(4.13.):

$$M(X = m) = n \cdot p = \lambda.$$

M(X = m) = λ = 2 (инкассатора).

Для выполнения дисперсии случайной величины, подчиняющейся распределению Пуассона, можно применить формулу:

$$D(X = m) \approx \lambda.$$

Итак, дисперсия числа инкассаторов, прибывающих утром на автомобиле в течение 15 минут:

$$D(X = m) = \lambda = 2 \text{ (кв.ед.)}.$$

Среднее квадратическое отклонение числа инкассаторов, прибывающих утром на автомобиле в течение 15 минут:

$\sigma(X) = \sqrt{2} = 1{,}4142$ (инкассатора).

в) Зададим теперь дискретную случайную величину в виде функции распределения:

$$F(X = m) = P(X < x) = \sum_{\substack{i \\ x_i < x}} \frac{2^m}{m!} \cdot e^{-2}.$$

Рассчитаем значения F(x):

$$F(x) = \begin{cases} 0 & \text{при } x \le 0 \\ 0{,}1353 & \text{при } 0 \le x \le 1 \\ 0{,}4060 & \text{при } 1 < x \le 2 \\ 0{,}6767 & \text{при } 2 < x \le 3 \\ 0{,}8571 & \text{при } 3 < x \le 4 \\ 0{,}9473 & \text{при } 4 < x \le 5 \\ 0{,}9834 & \text{при } 5 < x \le 6 \\ 0{,}9954 & \text{при } 6 < x \le 7 \\ 0{,}9988 & \text{при } 7 < x \le 8 \\ 0{,}9997 & \text{при } 8 < x \le 9 \\ 1 & \text{при } x > 9 \end{cases}$$

Эти данные можно представить и в виде таблицы 4.7:

Таблица 4.7

X	x ≤0	0<x≤1	1<x≤2	2<x≤3	3<x≤4	4<x≤5	5<x≤6	6<x≤7	7<x≤8	8<x≤9	x > 9
P(X)	0	0,1353	0,4060	0,6767	0,8571	0,9473	0,9834	0,9954	0,9988	0,9997	1

г) Определим вероятность того, что в течение 15 минут в банк прибудут *хотя бы два инкассатора.*

“Хотя бы два” - “как минимум два” - “два или больше”. Другими словами, “хотя бы два” - это “или два, или три, или четыре, или ...”.

Исходя из этого, для определения вероятности того, что в течение 15 минут в банк прибудут хотя бы два инкассатора, можно использовать теорему сложения вероятностей несовместных событий:

$$P(X \ge 2) = P(X = 2) + P(X = 3) + P(X = 4) + ... + P(X = n).$$

График функции (вероятностная гистограмма)

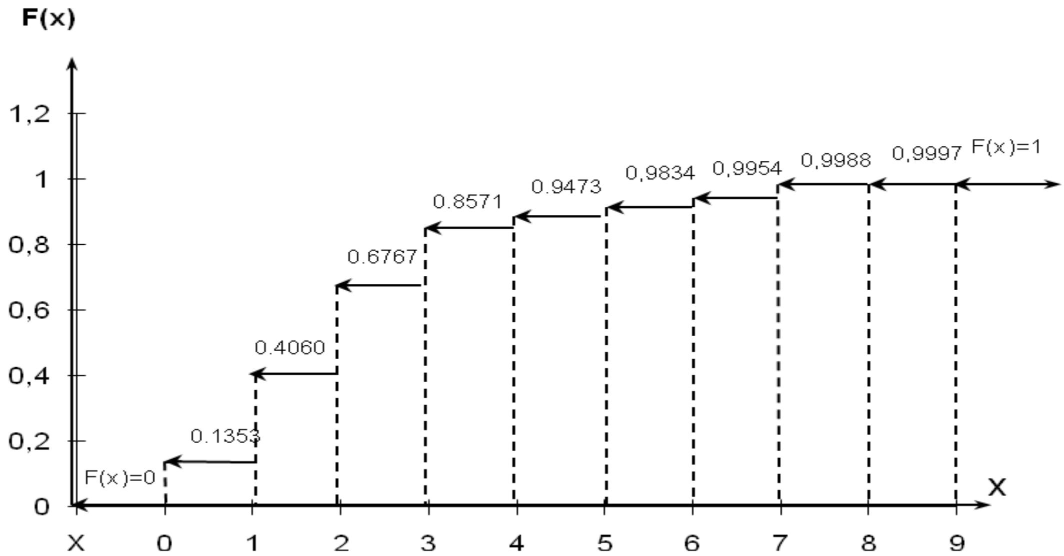


Рис. 4.4

С другой стороны, все возможные значения случайной величины образуют полную группу событий, а сумма их вероятностей равна 1. По отношению к событию ($X \geq 2$) до полной группы событий не хватает события ($X < 2$), т. е. ($x \leq 1$), которое является противоположным событию ($X \geq 2$). Поэтому искомую вероятность того, что в течение 15 минут в банк прибудут на автомобиле хотя бы два инкассатора, проще найти следующим образом:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X=1)) = 1 - (0{,}1353 + 0{,}2707) = 1 - 0{,}406 = 0{,}594.$$

Вероятность того, что в течение 15 минут в банк прибудут на автомобиле хотя бы два инкассатора, составляет 0,5904.

д) Определим вероятность того, что в течение 15 минут число прибывших инкассатор окажется *меньше трех*.

“Меньше трех” - это “или ноль, или один, или два”.

Из теоремы сложения вероятностей несовместных событий следует:

$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$.

$P(X < 3) = 0{,}1353 + 0{,}2707 + 0{,}2707 = 0{,}6767$.

Ответ. Вероятность того, что в течение 15 минут в банк прибудет меньше трех инкассаторов, составляет 0,6767.

Пример 4.3. Из 20 лотерейных билетов выигрышными являются 4. Наудачу извлекаются 4 билета.

а) Составьте ряд распределения числа выигрышных билетов среди отобранных;

б) Найдите числовые характеристики этого распределения;

в) Напишите функцию распределения числа выигрышных билетов среди отобранных и постройте ее график;

г) Определите вероятность того, что среди отобранных 4 билетов окажется *не меньше трех* выигрышных билетов;

д) Определите вероятность того, что среди отобранных 4 билетов окажется *не больше одного* выигрышного билета.

Решение. В качестве случайной величины в данной задаче выступает число выигрышных билетов среди отобранных. Обозначим ее через X.

Перечислим все возможные значения случайной величины X: 0, 1, 2, 3, 4.

Это - дискретная случайная величина, т.к. ее возможные значения отличаются друг от друга не менее чем на 1, и множество ее возможных значений является счетным.

Очевидно, что отбор лотерейных билетов - бесповторный. Следовательно, испытания - зависимые.

Вышеперечисленные признаки указывают на то, что рассматриваемая случайная величина - число выигрышных билетов среди отобранных - подчиняется гипергеометрическому закону распределения.

Случайная величина, интересующая нас, $X = m$ - число выигрышных билетов в выборке объемом в n билетов. Число всех возможных случаев отбора n билетов из общего числа N билетов равно числу сочетаний из N по n (C_N^n), а число случаев отбора m выигрышных билетов из общего числа M выигрышных билетов (*и значит, (n-m) проигрышных из общего числа (N - M) проигрышных*) равно произведению

$$C_M^m \cdot C_{N-M}^{n-m}$$

(отбор каждого из m выигрышных билетов может сочетаться с отбором любого из (n-m) проигрышных). Событие, вероятность которого мы хотим определить, состоит в том, что в выборке из n лотерейных билетов окажется ровно m выигрышных. По формуле для расчета вероятности события в классической модели вероятность получения в выборке m выигрышных билетов (то есть вероятность того, что случайная величина X примет значение m) равна:

$$P_{n,m} = \frac{C_M^m . C_{N-M}^{n-m}}{C_N^n},$$

где C_N^n - общее число всех единственно возможных, равновозможных и несовместных исходов,

$C_M^m \cdot C_{N-M}^{n-m}$ - число исходов, благоприятствующих наступлению интересующего нас события;

$m \le n$, если $n \le M$ и $m \le M$, если $M < n$.

Если по этой формуле вычислить вероятности для всех возможных значений m и поместить их в таблицу, то получим ряд распределения.

а) Составим ряд распределения.

Вычислим вероятности того, что случайная величина примет каждое из своих возможных значений, и запишем полученные результаты в таблицу.

По условию задачи N =20; M = 4; n = 4; m = 0, 1, 2, 3, 4.

$$P_{4,0} = \frac{C_4^0 \cdot C_{16}^4}{C_{20}^4} = \frac{1 \cdot \dfrac{16 \cdot 15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3 \cdot 4}}{\dfrac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}} = 0{,}37564;$$

$$P_{4,1} = \frac{C_4^1 \cdot C_{16}^3}{C_{20}^4} = \frac{4 \cdot \dfrac{16 \cdot 15 \cdot 14}{1 \cdot 2 \cdot 3}}{\dfrac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}} = 0{,}46233;$$

$$P_{4,2} = \frac{C_4^2 \cdot C_{16}^2}{C_{20}^4} = \frac{\dfrac{4 \cdot 3}{1 \cdot 2} \cdot \dfrac{16 \cdot 15}{1 \cdot 2}}{\dfrac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}} = 0{,}14861;$$

$$P_{4,3} = \frac{C_4^3 \cdot C_{16}^1}{C_{20}^4} = \frac{4 \cdot 16}{\dfrac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}} = 0{,}01321;$$

$$P_{4,4} = \frac{C_4^4 \cdot C_{16}^0}{C_{20}^4} = \frac{1 \cdot 1}{\dfrac{20 \cdot 19 \cdot 18 \cdot 17}{1 \cdot 2 \cdot 3 \cdot 4}} = 0{,}00021.$$

Занесем полученные результаты в таблицу 4.8:

Таблица 4.8

X	0	1	2	3	4
P(X)	0,37564	0,46233	0,14861	0,01321	0,00021

Произведем проверку. Так как все возможные значения случайной величины образуют полную группу событий, сумма их вероятностей должна быть равна 1.

Проверка: 0,37564 + 0,46233 + 0,14861 + 0,01321 + 0,00021 = 1.

График полученного распределения вероятностей дискретной случайной величины - полигон распределения вероятностей; изображенный на рис 4.5.

Рис. 4.5

б) Найдем основные числовые характеристики распределения данной случайной величины.

Можно рассчитать математическое ожидание, дисперсию и среднее квадратическое отклонение по общим для любой дискретной случайной величины формулам.

Но математическое ожидание случайной величины, подчиняющейся гипергеометрическому распределению, может быть рассчитано по более простой формуле:

$$M(X=m)=n\cdot\frac{M}{N}.$$

Рассчитаем математическое ожидание числа выигрышных билетов среди отобранных:

$$M(X=m)=n\cdot\frac{M}{N}=4\cdot\frac{4}{20}=0{,}8\,(\text{билета}).$$

Дисперсию случайной величины, подчиняющейся распределению, также можно рассчитать по более простой формуле:

$$D(X=m)=n\cdot\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)\cdot\left(1-\frac{n-1}{N-1}\right).$$

Вычислим дисперсию числа выигрышных билетов среди отобранных:

$$D(X=m)=n\cdot\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)\cdot\left(1-\frac{n-1}{N-1}\right)=4\cdot\frac{4}{20}\cdot\left(1-\frac{4}{20}\right)\cdot\left(1-\frac{4-1}{20-1}\right)=$$

$$=\frac{16}{20}\cdot\frac{16}{20}\cdot\frac{16}{19}=\frac{256}{475}=0{,}53895.$$

D(X = m) = 0,53895 (кв.ед.).

Рассчитаем среднее квадратическое отклонение числа выигрышных билетов среди отобранных:

$$\sigma(X) = \sqrt{0,5389} = 0,73413 \text{ (билета)}.$$

в) Зададим дискретную случайную величину в виде функции распределения:

$$F(X = m) = P(X < x) = \sum_{\substack{i \\ x_i < x}} \frac{C_4^m \cdot C_{16}^{4-m}}{C_{20}^4}.$$

Рассчитаем значения F(x):

$$F(x) = \begin{cases} 0 & \text{при } x \le 0 \\ 0,37564 & \text{при } 0 < x \le 1 \\ 0,83797 & \text{при } 1 < x \le 2 \\ 0,98658 & \text{при } 2 < x \le 3 \\ 0,99979 & \text{при } 3 < x \le 4 \\ 1 & \text{при } x > 4 \end{cases}$$

Эти данные можно представить и в виде таблицы 4.9:

Таблица 4.9

X	$x \le 0$	$0 < x \le 1$	$1 < x \le 2$	$2 < x \le 3$	$3 < x \le 4$	$x > 4$
F(x)	0	0,37564	0,83797	0,98658	0,99979	1

График функции распределения

Рис. 4.6

г) Определим вероятность того, что среди 4-х отобранных билетов окажется *не меньше трех* выигрышных.

"Не меньше трех" - "как минимум три" - "три или больше". Другими словами, "не меньше трех" - это "или три, или четыре".

Исходя из этого, для определения вероятности того, что среди отобранных 4-х билетов окажется не меньше трех выигрышных билетов, можно применить теорему сложения вероятностей несовместных событий:

$P(X \ge 3) = P(X = 3) + P(X = 4) = 0,01321 + 0,00021 = 0,01342.$

Вероятность того, что среди отобранных окажется не меньше трех выигрышных билетов, составляет 0,01342.

д) Определим теперь вероятность того, что среди отобранных 4-х билетов окажется *не больше одного* выигрышного билета.

"Не больше одного" - это "один или меньше" - "или ноль, или один".

Следовательно, для определения вероятности того, что среди отобранных окажется не больше одного выигрышного билета, также применяем теорему сложения вероятностей для несовместных событий:

P(X ≤ 1) = P(X = 0) + P(X = 1) = 0,37564 + 0,46233 = 0,83797.

Ответ. $P(X \geq 3) = 0{,}01342$; $P(X \leq 1) = 0{,}83797$.

5. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

5.1. Функция распределения и плотность распределения непрерывной случайной величины

Случайная величина X называется непрерывной, если ее функция распределения непрерывна и имеет производную.

Как уже было показано в разделе 4 (формула 4.2), функцией распределения случайной величины X называется функция F(X), выражающая вероятность выполнения условия $X < x$:

$$F(X) = P(X < x). \quad (5.1)$$

Функция распределения обладает следующими свойствами:

1. Вероятность попадания случайной величины в промежуток от α до β равна приращению функции распределения на концах этого промежутка:

$$P(\alpha < x < \beta) = F(\beta) - F(\alpha), \quad (5.2)$$

так как вероятность любого отдельного значения случайной величины равна нулю, если функция распределения непрерывна при этом значении, т. е.:

$P(X = x_1) = 0$, когда F(X) - непрерывна в точке $x = x_1$.

2.Функция распределения удовлетворяет условиям:

$$F(-\infty) = 0. \qquad F(+\infty) = 1. \quad (5.3)$$

Плотностью распределения (дифференциальной функцией) непрерывной случайной величины называется функция

$$f(x) = F'(x). \quad (5.4)$$

Плотность распределения любой случайной величины неотрицательна: $f(x) \geq 0$.

Несобственный интеграл от дифференциальной функции в пределах от $-\infty$ до $+\infty$ равен 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1. \quad (5.5)$$

График функции y = f(x) называется *кривой распределения* или графиком плотности распределения. Кривая y = f (x) располагается над осью абсцисс.

Вероятность попадания случайной величины в промежуток от α до β может быть вычислена по формуле:

$$P(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x)dx. \quad (5.6)$$

Подинтегральное выражение *f(x)dx* называется элементом вероятности. Оно выражает вероятность попадания случайной точки в промежуток между точками x и $x+\Delta x$, где Δx бесконечно малая величина.

Функция распределения F(x) выражается через плотность f(x) формулой:

$$F(x)=\int\limits_{-\infty}^{x} f(x)dx. \tag{5.7}$$

Математическое ожидание непрерывной случайной величины X вычисляется по формуле:

$$M(X)=\int\limits_{-\infty}^{+\infty} xf(x)dx, \tag{5.8}$$

дисперсия

$$D(X)=\int\limits_{-\infty}^{+\infty}(x-M(X))^2 f(x)dx. \tag{5.9}$$

5.2. Нормальное распределение

Если плотность распределения (дифференциальная функция) случайной переменной определяется выражением:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{5.10}$$

то говорят, что X имеет нормальное распределение с параметрами a и σ^2. Вероятностный смысл параметров: $a=M(X)$, а $\sigma^2=D(X)$.Обозначение:

$$X \sim N\left(a;\sigma^2\right).$$

Для расчета вероятности попадания нормально распределенной случайной величины X в промежуток от α до β используется формула:

$$P(\alpha < X < \beta)=\Phi_0\left(\frac{\beta-a}{\sigma}\right)-\Phi_0\left(\frac{\alpha-a}{\sigma}\right), \tag{5.11}$$

где $\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{\frac{z^2}{2}}dz$ - интеграл Лапласа.

Формула (5.11) иногда в литературе называется интегральной теоремой Лапласа.

Функция $\Phi_0(x)$ обладает свойствами:

1) $\Phi_0(0)=0,$

2) $\lim\limits_{x\to\pm\infty}\Phi_0(x)=\pm 0{,}5.$

3) $\Phi_0(-x)=-\Phi_0(x)$ (см. таблицу приложения 2).

Функция $\Phi_0(x)$ табулирована. В частности, для симметричного относительно a промежутка $(a-\Delta;a+\Delta)$ имеем:

$$P\left(\left|X-a\right|<\Delta\right)\approx 2\Phi_0\left(\frac{\Delta}{\sigma}\right). \tag{5.12}$$

Формула (5.12) применима и к частоте *m,* поскольку ее закон распределения при достаточно большом числе испытаний практически совпадает с нормальным. Применительно к случайной величине *m*, с учетом ее числовых характеристик

$$M(m) = np \text{ и } \quad \sigma^2(m) = npq, \tag{5.13}$$

формула (5.12) примет вид:

$$P(|m - np| < \Delta) \approx 2\Phi_0\left(\frac{\Delta}{\sqrt{npq}}\right). \tag{5.14}$$

Формула (5.12) может быть применена и к относительной частоте $\frac{m}{n}$ с числовыми характеристиками

$$M\left(\frac{m}{n}\right) = p \text{ и } \quad \sigma^2\left(\frac{m}{n}\right) = \frac{pq}{n}, \tag{5.15}$$

$$P\left(\left|\frac{m}{n} - p\right| < \Delta\right) \approx 2\Phi_0\left(\frac{\Delta}{\sqrt{\frac{pq}{n}}}\right) = 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right), \tag{5.16}$$

С вероятностью, очень близкой к единице (равной $2\Phi_0\int(3) = 0{,}9973$), нормально распределенная случайная величина X удовлетворяет неравенству:

$$a - 3\sigma < X < a + 3\sigma. \tag{5.17}$$

В этом состоит *правило трех сигм*: если случайная величина распределена по нормальному закону, то ее отклонение от математического ожидания практически не превышает $\pm 3\sigma$.

Локальная теорема Муавра-Лапласа. При $p \neq 0$ и $p \neq 1$ и достаточно большом n биноминальное распределение близко к нормальному закону (причем их математические ожидания и дисперсии совпадают), т.е. имеет место равенство:

$$\lim_{n\to\infty} P_{n;m} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \text{ где } \sigma = \sqrt{npq},\ a = np.$$

Тогда:

$$P_{n;m} \approx \frac{\varphi(x)}{\sqrt{npq}} \tag{5.18}$$

для достаточно больших n (здесь $\varphi(x)$ - плотность вероятностей стандартной нормальной случайной величины $t = \frac{x - a}{\sigma}$ и $x = \frac{m - np}{\sqrt{npq}}$).

Пример 5.1. На рынок поступила крупная партия говядины. Предполагается, что вес туш - случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием *a* = 950 кг и средним квадратическим отклонением σ = 150 кг. Определите вероятность того, что вес случайно отобранной туши:

а) окажется больше 1250 кг;

б) окажется меньше 850 кг;

в) будет находиться между 800 и 1300 кг;

г) отклонится от математического ожидания меньше, чем на 50 кг;

д) отклонится от математического ожидания больше, чем на 50 кг;

е) Найдите границы, в которых отклонение веса случайно отобранной туши от своего математического ожидания не превысит утроенного среднего квадратического отклонения (проиллюстрируйте правило трех сигм);

ж) С вероятностью 0,899 определите границы, в которых будет находиться вес случайно отобранной туши. Какова при этом условии максимальная величина отклонения веса случайно отобранной туши от своего математического ожидания?

Решение.

а) Вероятность того, что вес случайно отобранной туши окажется больше 1250 кг – можно понимать как вероятность того, что вес случайно отобранной туши окажется в интервале от 1250 кг до $+\infty$.

Формула расчета вероятности попадания в заданный интервал нормально распределенной случайной велечины X имеет вид:

$$P(\alpha < X < \beta) = \Phi_0\left(\frac{\beta - a}{\sigma}\right) - \Phi_0\left(\frac{\alpha - a}{\sigma}\right),$$

где $\Phi_0(z)$ - функция Лапласа:

$$\Phi_0(z) = \int_0^z \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}} dx.$$

Функция $\Phi_0(z)$ является нечетной функцией, т.е. $\Phi_0(-z) = -\Phi_0(z)$.

Найдем вероятность того, что вес случайно отобранной туши окажется больше 1250 кг. По условию: $\alpha = 1250$, $\beta = +\infty$, $a = 950$, $\sigma = 150$.

Используем формулу расчета вероятности попадания в заданный интервал нормально распределенной случайной величины X.

$$P(X > 1250) = P(1250 < X < +\infty) = \Phi_0\left(\frac{\infty - 950}{150}\right) - \Phi_0\left(\frac{1250 - 950}{150}\right) =$$

$$= \Phi_0(\infty) - \Phi_0(2).$$

Найдем по таблице функции Лапласа (см. приложение 2) значения $\Phi_0(z)$.

Значения $\Phi_0(+\infty)$ в таблице нет. Однако известно, что $\Phi_0(z) \to 0{,}5$ при $z \to +\infty$. Уже при $z = 5$ $\Phi_0(z = 5) = 0{,}49999997 \approx 0{,}5$. Очевидно, что $\Phi_0(+\infty)$ – величина, бесконечно близкая к 0,5. $\Phi_0(-\infty)$ – величина, бесконечно близкая к -0,5.

По таблице функции Лапласа $\Phi_0(2) = 0{,}47725$.

Отсюда: $P(X>1250) = 0{,}5 - 0{,}47725 = 0{,}02275$.

Итак, вероятность того, что вес случайно отобранной туши окажется больше 1250 кг, составляет 0,02275.

Проиллюстрируем решение задачи графически (рис. 5.1).

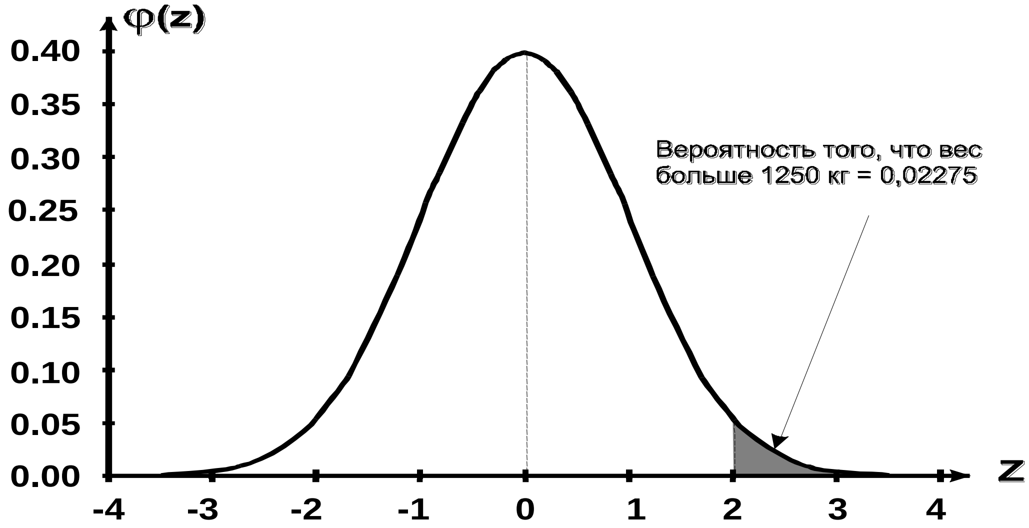


Рис. 5.1. Графическая интерпретация к примеру 5.1

Итак, нам задана нормально распределенная случайная величина X с математическим ожиданием a = 950 кг и средним квадратическим отклонением σ = 150 кг, то есть X ~N(950;150^2). Мы хотим найти вероятность того, что X больше 1250, то есть определить P(X > 1250). Преобразуем X в Z, и тогда искомая вероятность определится по таблице приложения 2 стандартного нормального распределения $Z = \frac{x-a}{\sigma}$.

$$P(X > 1250) = P\left(\frac{X-a}{\sigma} > \frac{1250-a}{\sigma}\right) = P\left(Z > \frac{1250-a}{\sigma}\right) =$$

$$= P\left(Z > \frac{1250-950}{150}\right) = P(Z > 2) = 0{,}5 - 0{,}47725 = 0{,}02275.$$

Точка z = 0 соответствует математическому ожиданию, то есть a = 950 кг.

б) Вероятность того, что вес случайно отобранной туши окажется меньше 850 кг - это то же самое, что вероятность того, что вес случайно отобранной туши окажется в интервале от -∞ до 850 кг.

По условию: $\alpha = -\infty$, $\beta = 850$, $a = 950$, $\sigma = 150$.

Для расчета искомой вероятности используем формулу расчета вероятности попадания в заданный интервал нормально распределенной случайной величины X.

$$\mathrm{P}(\mathrm{X} < 850) = \mathrm{P}(-\infty < \mathrm{X} < 850) \approx \Phi_0\left(\frac{850-950}{150}\right) - \Phi_0\left(\frac{-\infty-950}{150}\right) =$$

$$= \Phi_0(-2/3) - \Phi_0(-\infty) = \Phi_0(+\infty) - \Phi_0(0{,}67).$$

Согласно свойству функции Лапласа:

$-\Phi_0(-\infty) = \Phi_0(+\infty)$, а $\Phi_0(-0{,}67) = -\Phi_0(0{,}67)$.

Найдем по таблице функции Лапласа (приложение 2) значения $\Phi_0(z)$.

$\Phi_0(+\infty) \approx 0{,}5$;

$\Phi_0(0{,}67) = 0{,}24857$.

Отсюда: P(X < 850) = 0,5 - 0,24857 = 0,25143.

Вероятность того, что вес случайно отобранной туши окажется меньше 850 кг, составляет 0,25143.

Проиллюстрируем решение задачи графически (рис. 5.2).

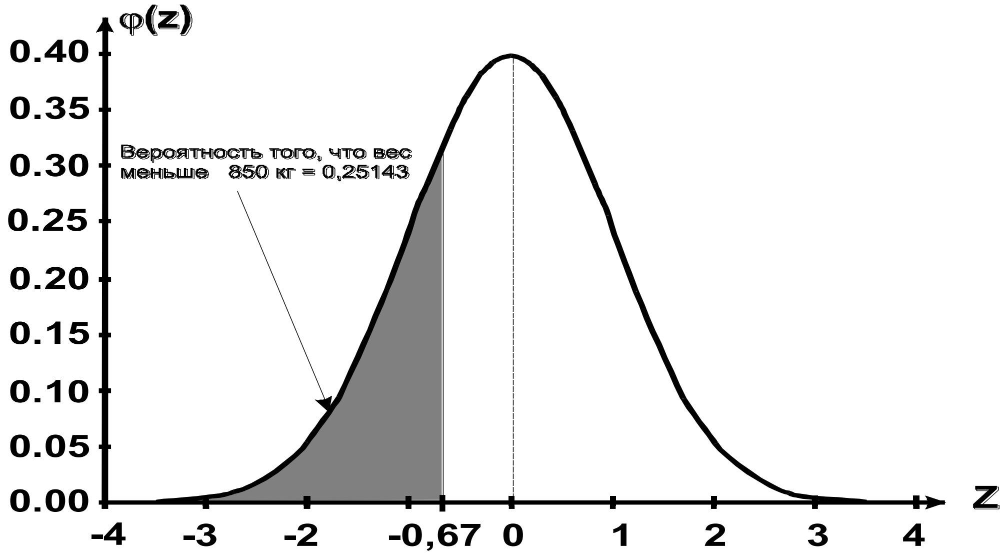


Рис. 5.2. Графическая интерпретация к примеру 5.1

По условию данной задачи точка на оси абсцисс (z = -0,67) соответствует x=850, т.е. весу, равному 850 кг. Заштрихованная на графике площадь представляет собой вероятность того, что вес наудачу выбранной туши окажется меньше 850 кг, т.е. в интервале от -∞ до 850 кг.

в) Найдем вероятность того, что вес случайно отобранной туши окажется в интервале от 800 до 1300 кг.

По условию: $\alpha = 800$, $\beta = 1300$, $a = 950$, $\sigma = 150$.

Для расчета искомой вероятности используем формулу расчета вероятности попадания в заданный интервал нормально распределенной случайной величины X.

$$P(800 < X < 1300) \approx \Phi_0\left(\frac{1300-950}{150}\right) - \Phi_0\left(\frac{800-950}{150}\right) = \Phi_0(2{,}33) - \Phi_0(-1) =$$

$$= \Phi_0(2{,}33) + \Phi_0(1).$$

Согласно свойству функции Лапласа:

$-\Phi_0(-1) = \Phi_0(1)$.

Найдем по таблице функции Лапласа (см. приложение 2) значения $\Phi_0(z)$.

$\Phi_0(2{,}33) = 0{,}49010$;

$\Phi_0(1) = 0{,}34134$.

Отсюда: $P(800 < X < 1300) = 0{,}49010 + 0{,}34134 = 0{,}83144$.

Вероятность того, что вес случайно отобранной туши окажется в интервале от 800 кг до 1300, составляет 0,83144.

Проиллюстрируем решение задачи графически (рис. 5.3).

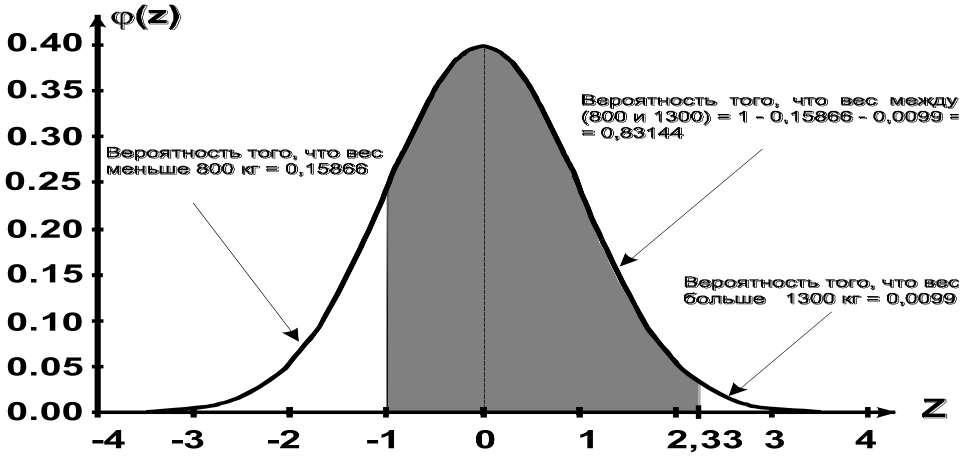


Рис. 5.3. Графическая интерпретация к примеру 5.1

По условию данной задачи точка на оси абсцисс (z = -1) соответствует x = 800, т.е. весу, равному 800 кг, а точка (z = 2,33) соответствует x = 1300, т.е. весу, равному 1300 кг. Заштрихованная на графике площадь представляет собой вероятность того, что вес наудачу выбранной туши окажется в интервале от 800 до 1300 кг.

На графике видно, что искомую вероятность, что вес наудачу выбранной туши окажется в интервале от 800 до 1300 кг, можно было найти другим способом. Для этого необходимо было найти вероятность того, что вес наудачу выбранной туши окажется меньше 800 кг, а также больше 1300 кг. Полученные вероятности – сложить и вычесть из единицы.

Так, вероятность того, что вес наудачу выбранной туши окажется меньше 800 кг, - это, другими словами, вероятность того, что вес случайно отобранной туши окажется в интервале от -∞ до 850 кг.

$$P(X < 800) = P(-\infty < X < 800) \approx \Phi_0\left(\frac{800-950}{150}\right) - \Phi_0\left(\frac{-\infty-950}{150}\right) =$$

$$= \Phi_0(-1) - \Phi_0(-\infty) = \Phi_0(+\infty) - \Phi_0(1) = 0{,}5 - 0{,}34134 = 0{,}15866.$$

Вероятность того, что вес случайно отобранной туши окажется больше 1300 кг, - это, другими словами, вероятность того, что вес случайно отобранной туши окажется в интервале от 1300 кг до +∞.

$$P(X > 1300) = P(1300 < X < +\infty) = \Phi_0\left(\frac{\infty-950}{150}\right) - \Phi_0\left(\frac{1300-950}{150}\right) =$$

$$= \Phi_0(\infty) - \Phi_0(2{,}33) = 0{,}5 - 0{,}49010 = 0{,}0099.$$

Отсюда, искомая вероятность того, что вес наудачу выбранной туши окажется в интервале от 800 до 1300 кг:

$P(800 < X < 1300) = 1 - (P(X<800) + P(X>1300)) = 1 - (0{,}15866 + 0{,}0099) =$
$=1 - 0{,}16856 = 0{,}83144.$

г) Найдем вероятность того, что вес случайно отобранной туши отклонится от математического ожидания меньше, чем на 50 кг, т.е.:

$P(|X - 950| < 50) = ?$

Что значит: $|X - 950| < 50$?

Это неравенство можно заменить двойным неравенством:

$$-50 < X - 950 < 50 \text{ или}$$
$$950 - 50 < X < 950 + 50,$$
$$900 < X < 1000.$$

Следовательно:

$P(|X - 950| < 50) = P(900 < X < 1000).$

А это вероятность попадания в заданный интервал нормально распределенной случайной величины X.

Отсюда:

$$P(|X-950|<50) = P(900 < X < 1000) = \Phi_0(\frac{1000-950}{150}) - \Phi_0(\frac{900-950}{150}) =$$
$$= \Phi_0(0{,}33) - \Phi_0(-0{,}33) = \Phi_0(0{,}33) + \Phi_0(0{,}33) = 2\Phi_0(0{,}33).$$

Согласно свойству функции Лапласа:

$-\Phi_0(-0{,}33) = \Phi_0(0{,}33).$

Найдем по таблице функции Лапласа (см. приложение 2) значения $\Phi_0(z)$.

$\Phi_0(0{,}33) = 0{,}1293.$

Следовательно:

$P(|X - 950| < 50) = P(900 < X < 1000) = 2 \cdot 0{,}1293 = 0{,}2586.$

Вероятность того, что вес случайно отобранной туши отклонится от математического ожидания меньше, чем на 50 кг, составляет 0,2586.

Эту задачу легче решить, используя формулу расчета вероятности заданного отклонения нормально распределенной случайной величины X от своего математического ожидания:

$$P(|X-a| < \Delta) = 2 \cdot \Phi_0(\frac{\Delta}{\sigma}),$$

где - Δ величина отклонения случайной величины X от математического ожидания.

По условию $\Delta = 50$; $a = 950$, $\sigma = 150$.

Используя эту формулу, сразу получим:

$P(|X - 950| < 50) = 2\Phi_0(50 / 150) = 2\Phi_0(0{,}33) = 2 \cdot 0{,}1293 = 0{,}2586.$

Проиллюстрируем решение задачи графически (рис. 5.4).

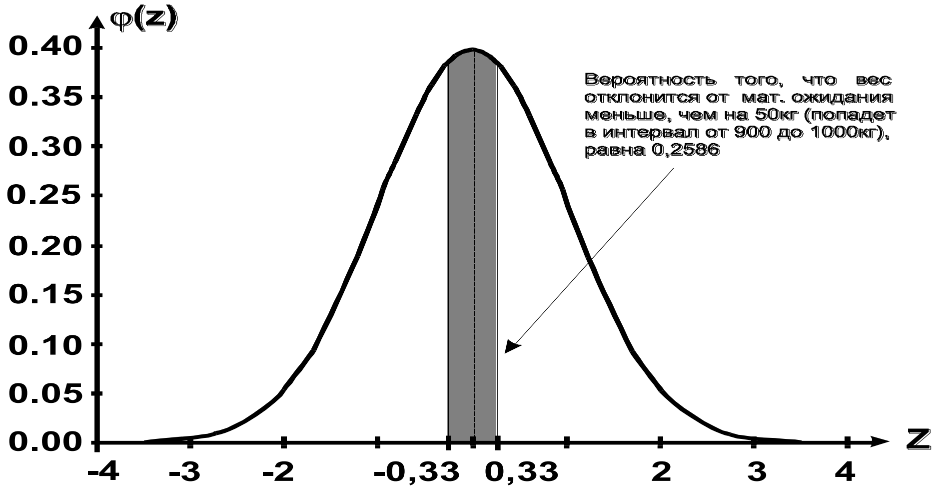


Рис.5.4. Графическая иллюстрация к задаче 5.1

По условию данной задачи точка на оси абсцисс ($z = -0,33$) соответствует $x = 900$, т.е. весу, равному 900 кг, а точка ($z = 0,33$) соответствует $x = 1000$, т.е. весу, равному 1000 кг. Заштрихованная на графике площадь представляет собой вероятность того, что вес наудачу выбранной туши окажется в интервале от 900 до 1000 кг, т.е. отклонится от математического ожидания меньше, чем на 50 кг.

д) Найдем вероятность того, что вес случайно отобранной туши отклонится от математического ожидания больше, чем на 50 кг, т.е.:

$P(|X - 950| > 50) = ?$

Это вероятность события, противоположного по отношению к событию: вес случайно отобранной туши отклонится от математического ожидания меньше, чем на 50 кг ($P(|X - 950| < 50)$).

Следовательно:

$P(|X - 950| > 50) = 1 - P(|X - 950| < 50) = 1 - 0,2586 = 0,7414.$

Вероятность того, что вес случайно отобранной туши отклонится от математического ожидания больше, чем на 50 кг, составляет 0,7414.

Можно использовать другой алгоритм решения.

Вероятность того, что вес случайно отобранной туши отклонится от математического ожидания больше, чем на 50 кг, - это вероятность того, что вес случайно отобранной туши будет или меньше (950 - 50 = 900) кг или больше (950 + 50 = 1000) кг.

По теореме сложения вероятностей несовместных событий имеем:

$P(|X - 950| > 50) = P(X < 900) + P(X > 1000).$

$$P(X < 900) = P(-\infty < X < 900) = \Phi_0\left(\frac{900-950}{150}\right) - \Phi_0\left(\frac{-\infty-950}{150}\right) =$$

$$= \Phi_0(-1/3) - \Phi_0(-\infty) = \Phi_0(+\infty) - \Phi_0(0{,}33) = 0{,}5 - 0{,}1293 = 0{,}3707.$$

$$P(X > 1000) = P(1000 < X < +\infty) = \Phi_0\left(\frac{\infty-950}{150}\right) - \Phi_0\left(\frac{1000-950}{150}\right) =$$

$$= \Phi_0(\infty) - \Phi_0(0{,}33) = 0{,}5 - 0{,}1293 = 0{,}3707.$$

Отсюда:

$P(|X - 950| > 50) = P(X < 900) + P(X > 1000) = 0,3707 + 0,3707 = 0,7414.$

е) Найдем границы, в которых отклонение веса случайно отобранной туши от своего математического ожидания не превысит утроенного среднего квадратического отклонения.

В этом задании студентам предлагается проиллюстрировать *правило трех сигм*, которое можно сформулировать следующим образом:

Если случайная величина распределена по нормальному закону, то ее отклонение от математического ожидания практически не превышает $\pm 3\sigma$.

$$P(|X - a| < 3\sigma) = 2\Phi_0(3) = 0{,}9973.$$

Вероятность того, что отклонение нормально распределенной случайной величины X от своего математического ожидания будет меньше 3σ или, другими словами, вероятность того, что нормально распределенная случайная величина X попадет в интервал (a - 3σ; a + 3σ), равна 0,9973.

Следовательно, вероятность того, что отклонение случайной величины от своего математического ожидания по абсолютной величине превысит утроенное среднее квадратическое отклонение, очень мала и равна 0,0027. Другими словами, лишь в 27 случаях из 10000 случайная величина X в результате испытания может оказаться вне интервала (a - 3σ; a + 3σ). Такие события считаются практически невозможными.

Формулу, описывающую правило трех сигм, несложно получить из формулы вероятности заданного отклонения нормально распределенной случайной величины X от своего математического ожидания:

$$P(|X-a|<\Delta)=2\cdot\Phi_0\left(\frac{\Delta}{\sigma}\right).$$

Если взять $\Delta = 3\sigma$, то получим $\Delta/\sigma = 3$.

Отсюда:

$$P(|X - a| < 3\sigma) = 2\Phi_0(3) = 0{,}9973.$$

По условию задачи: $a = 950$; $\sigma = 150$.

Правило трех сигм можно представить так:

$$P(a - 3\sigma < X < a + 3\sigma) = 2\Phi_0(3) = 0{,}9973.$$

Интересующие нас границы - это границы интервала (a - 3σ; a + 3σ), т.е.:

$$a - 3\sigma < X < a + 3\sigma,$$

$$950 - 3 \cdot 150 < X < 950 + 3 \cdot 150,$$

$$500 < X < 1400.$$

Учитывая, что вес отобранной туши - нормально распределенная случайная величина, можно быть практически уверенным, что вес случайно отобранной туши не выйдет за пределы от 500 до 1400 кг.

ж) Определим границы, в которых с вероятностью 0,899 будет находиться вес случайно отобранной туши.

Формулу вероятности заданного отклонения нормально распределенной случайной величины X от своего математического ожидания можно представить следующим образом:

$$P(|X-a|<\Delta)=2\cdot\Phi_0\left(\frac{\Delta}{\sigma}\right)=\gamma,$$

или

$$P(a-\Delta<X<a+\Delta)=2\cdot\Phi_0\left(\frac{\Delta}{\sigma}\right)=\gamma,$$

где γ - вероятность того, что отклонение нормально распределенной случайной величины X от своего математического ожидания не превысит заданной величины Δ.

По условию задачи: $a = 950$; $\sigma = 150$.

Используя последнюю формулу, получим:

$$P(950-\Delta<X<950+\Delta)=2\cdot\Phi_0\left(\frac{\Delta}{150}\right)=0{,}899.$$

Из соотношения $2\cdot\Phi_0\left(\frac{\Delta}{150}\right)=0{,}899$ найдем Δ.

$$2\cdot\Phi_0\left(\frac{\Delta}{150}\right)=0{,}899,$$

$$\Phi_0\left(\frac{\Delta}{150}\right)=\frac{0{,}899}{2}=0{,}4495.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком z $=\frac{\Delta}{150}$ $\Phi_0(z)=0{,}4495.$

z = 1,64, т.е. $\Phi_0(1{,}64)=0{,}4495.$

Отсюда: $\frac{\Delta}{150}=1{,}64.$

Δ= 1,64 · 150 = 246.

С вероятностью 0,899 можно ожидать, что отклонение веса случайно отобранной туши от своего математического ожидания не превысит 246 кг.

Найдем границы интересующего нас интервала:

$$a-\Delta < X < a+\Delta,$$

$$950-246 < X < 950+246,$$

$$704 < X < 1196.$$

С вероятностью 0,899 можно ожидать, что вес случайно отобранной туши будет находиться в пределах от 704 до 1196 кг.

Ответ а) 0,02275; б) 0,25143; в) 0,83144; г) 0,2586; д) 0,7414; е) (500; 1400); *ж)* 246; (704; 1196).

Пример 5.2. Изменим условие предыдущей задачи.

На рынок поступила крупная партия говядины. Предполагается, что вес туш – случайная величина, подчиняющаяся нормальному закону распределения с неизвестным математическим ожиданием и средним квадратическим отклонением σ =150 кг. Известно, что 37,07% туш имеют вес более 1000 кг. Определите ожидаемый вес случайно отобранной туши.

Решение. По условию задачи: σ = 150; α = 1000; β = +∞; P(X > 1000) = 0,3707.

Ожидаемый вес случайно отобранной туши - это среднеожидаемый вес, математическое ожидание, т.е. *a* =?

Используем формулу (5.10) расчета вероятности попадания в заданный интервал нормально распределенной случайной величины X:

$$P(1000 < X < +\infty)=\Phi_0\left(\frac{\infty-a}{150}\right)-\Phi_0\left(\frac{1000-a}{150}\right)=0{,}3707;$$

$$\Phi_0(\infty)-\Phi_0\left(\frac{1000-a}{150}\right)=0{,}3707;$$

$$0{,}5-\Phi_0\left(\frac{1000-a}{150}\right)=0{,}3707;$$

$$\Phi_0\left(\frac{1000-a}{150}\right)=0{,}5-0{,}3707=0{,}1293.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком $z = \frac{1000 - a}{150}$ $\Phi_0(z) = 0{,}1293.$

$z = 0{,}33$, т.е. $\Phi_0(0{,}33) = 0{,}1293.$

Отсюда: $\frac{1000 - a}{150} = 0{,}33.$

$1000 - a = 0{,}33 \cdot 150 = 50.$

$a = 1000 - 50 = 950.$

Ответ. Среднеожидаемый вес случайно отобранной туши составляет 950кг.

Пример 5.3. Вновь изменим условие задачи.

На рынок поступила крупная партия говядины. Предполагается, что вес туш - случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием a = 950 кг и неизвестным средним квадратическим отклонением. Известно, что 15,87% туш имеют вес менее 800 кг. Определите среднее квадратическое (стандартное) отклонение веса туш.

Решение. По условию задачи: $a = 950$; $\alpha = -\infty$ $\beta = 800$; $P(X < 800) = 0{,}1587$; $\sigma = ?$

Используем формулу (5.10) расчета вероятности попадания в заданный интервал нормально распределенной случайной величины X:

$$P(-\infty < X < 800) = \Phi_0(\frac{800 - 950}{\sigma}) - \Phi_0(\frac{-\infty - 950}{\sigma}) = 0{,}1587\,;$$

$$\Phi_0(\frac{800 - 950}{\sigma}) + \Phi_0(\infty) = 0{,}1587\,;$$

$$\Phi_0(\frac{800 - 950}{\sigma}) + 0{,}5 = 0{,}1587\,;$$

$$0{,}5 - \Phi_0(\frac{950 - 800}{\sigma}) = 0{,}1587\,;$$

$$\Phi_0(\frac{150}{\sigma}) = 0{,}5 - 0{,}1587 = 0{,}3413\,.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком $z = \frac{150}{\sigma}$ $\Phi_0(z) = 0{,}3413.$

$z = 1$, т.е. $\Phi_0(1) = 0{,}3413.$

Отсюда: $\frac{150}{\sigma} = 1$; $\sigma = 150$.

Ответ. Среднее квадратическое отклонение веса туш составляет 150 кг.

Пример 5.4. В очередной раз изменим условие задачи.

На рынок поступила крупная партия говядины. Предполагается, что вес туш - случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием a = 950 кг и неизвестным средним квадратическим отклонением. Каким должно быть среднее квадратическое (стандартное) отклонение, чтобы с вероятностью 0,81648 можно было утверждать, что абсолютное отклонение веса случайно отобранной туши от математического ожидания не превысит 200 кг?

Решение. По условию задачи: $a = 950$; $\Delta = 200$; $P(|X - 950| < 200) = 0{,}81648$; $\sigma = ?$

Используем формулу (5.11) расчета вероятности попадания заданного отклонения нормально распределенной случайной величины X от своего математического ожидания:

Тогда получим:

$P(|X - 950| < 200) = 2\Phi_0(200 / \sigma) = 0{,}81648$;

$2\Phi_0(200 / \sigma = 0{,}81648$;

$\Phi_0(200 / \sigma) = 0{,}81648 / 2$;

$\Phi_0(200 / \sigma) = 0{,}40824$.

По таблице функции Лапласа (см. приложение 2) найдем, при каком $z = \frac{200}{\sigma}$ $\Phi_0(z) = 0{,}40824$.

$z = 1{,}33$, т.е. $\Phi_0(1{,}33) = 0{,}40824$.

Отсюда: $\frac{200}{\sigma} = 1{,}33$; $\sigma = 200 / 1{,}33 = 150$.

Ответ. Чтобы с вероятностью 0,81648 можно было утверждать, что абсолютное отклонение веса случайно отобранной туши от математического ожидания не превысит 200 кг, среднее квадратическое отклонение веса туш должно составлять 150 кг.

Пример 5.5. Фирма собирается приобрести партию из 100 000 единиц некоторого товара. Из прошлого опыта известно, что 1% товаров данного типа имеют дефекты. Какова вероятность того, что в данной партии окажется от 950 до 1050 дефектных единиц товара?

Решение. В качестве случайной величины в данной задаче выступает число дефектных единиц товара в общей партии из 100000 единиц. Обозначим ее через X.

Перечислим все возможные значения случайной величины X: 0, 1, 2, ... , 99999, 100000.

Это - дискретная случайная величина, т.к. ее возможные значения отличаются друг от друга не менее чем на 1 и множество ее возможных значений является счетным.

По условию вероятность того, что единица товара окажется дефектной, - постоянна и составляет 0,01 ($p = 0{,}01$). Вероятность противоположного события, т.е. того, что единица товара не имеет дефекта - также постоянна и составляет 0,99 ($q = 1 - p = 1 - 0{,}01 = 0{,}99$).

Все 100000 испытаний - независимы, т.е. вероятность того, что каждая единица товара окажется дефектной, не зависит от того, окажется дефектной или нет любая другая единица товара.

Значения случайной величины X - это, в общем виде, число появлений интересующего нас события в 100000 независимых испытаниях.

Это позволяет сделать вывод о том, что случайная величина X - число дефектных единиц товара в общей партии из 100000 единиц - подчиняется биномиальному закону распределения вероятностей с параметрами n = 100000 и p = 0,01.

Итак, по условию задачи: n = 100000; p = 0,01; q = 0,99, X = m.

Необходимо найти вероятность того, что число дефектных единиц товара окажется в пределах от m_1 = 950 до m_2 =1050, т.е. - вероятность того, что случайная величина X = m попадет в интервал от 950 до 1050, т.е.:

$$P(m_1 < m < m_2) = ?$$

Так как мы имеем дело со случайной величиной, подчиняющейся биномиальному распределению, вероятность появления события m раз в n независимых испытаниях необходимо вычислять по формуле Бернулли (4.9).

В данном случае для определения искомой вероятности нам необходимо с помощью формулы Бернулли найти $P_{100000,\ 950}$; $P_{100000,\ 951}$; $P_{100000,\ 952}$; ... ; $P_{100000,\ 1049}$; $P_{100000,\ 1050}$, а затем - сложить их, используя теорему сложения вероятностей несовместных событий.

Очевидно, что такой способ определения искомой вероятности связан с громоздкими вычислениями. Так, например:

$$P_{100000,950} = C_{100000}^{950} \cdot 0{,}01^{950} \cdot 0{,}99^{100000-950}.$$

Можно значительно облегчить расчеты, если аппроксимировать биномиальное распределение нормальным, т.е. выразить функции биномиального распределения через функции нормального.

Когда n - число испытаний в биномиальном эксперименте - возрастает, дискретное биномиальное распределение стремится к непрерывному нормальному распределению. Это означает, что для больших n мы можем аппроксимировать биномиальные вероятности вероятностями, полученными для нормально распределенной случайной величины, имеющей такое же математическое ожидание и такое же среднее квадратическое отклонение.

Подставим параметры биномиального распределения (5.15) в формулу (5.10) и получим формулу для приближенного расчета вероятности появления события от m_1 и до m_2 раз в n независимых испытаниях $P(m_1 < m > m_2)$:

$$P_n(m_1 < m < m_2) \approx \Phi_0\left(\frac{m_2 - M(m)}{\sigma(m)}\right) - \Phi_0\left(\frac{m_1 - M(m)}{\sigma(m)}\right),$$

$$P_n(m_1 < m < m_2) \approx \Phi_0\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{m_1 - np}{\sqrt{npq}}\right), \qquad (5.19)$$

где $\Phi_0(z)$ - функция Лапласа: $\Phi_0(z) = \int\limits_0^z \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}} dx$.

Формулу для приближенного вероятности появления события не менее m_1 и не более m_2 раз в n независимых испытаниях $P_n(m_1 < m < m_2)$ называют *интегральной теоремой Лапласа.*

Использование локальной и интегральной теорем Лапласа дает приближенные значения искомых вероятностей. Погрешность будет невелика при условии, что npq > 9.

Для решения данной задачи воспользуемся интегральной теоремой Лапласа:

$$P_{100000}(950 \le m \le 1050) \approx \Phi_0(\frac{1050 - 100000 \cdot 0{,}01}{\sqrt{100000 \cdot 0{,}01 \cdot 0{,}99}}) - \Phi_0(\frac{950 - 100000 \cdot 0{,}01}{\sqrt{100000 \cdot 0{,}01 \cdot 0{,}99}});$$

$$P_{100000}(950 \le m \le 1050) \approx \Phi_0(\frac{1050 - 1000}{31{,}4643}) - \Phi_0(\frac{950 - 1000}{31{,}4643});$$

$$P_{100000}(950 \le m \le 1050) \approx \Phi_0(\frac{50}{31{,}4643}) - \Phi_0(\frac{-50}{31{,}4643});$$

$$P_{100000}(950 \le m \le 1050) \approx \Phi_0(\frac{50}{31{,}4643}) + \Phi_0(\frac{50}{31{,}4643});$$

$$P_{100000}(950 \le m \le 1050) \approx \Phi_0(1{,}59) + \Phi_0(1{,}59) = 2\Phi_0(1{,}59).$$

По таблице функции Лапласа (см. приложение 2) найдем $\Phi_0(1{,}59)$:

$\Phi_0(1{,}59) = 0{,}44408$.

$$P_{100000}(950 \le m \le 1050) \approx 2 \cdot 0{,}44408 = 0{,}88816.$$

Вероятность того, что в партии из 100000 единиц окажется от 950 до 1050 дефектных единиц товара, составляет 0,88816.

Данную конкретную задачу можно было решить еще более просто.

Математическое ожидание числа дефектных единиц товара равно 1000 единиц:

$$M(m) = n \cdot p = 100000 \cdot 0{,}01 = 1000.$$

Абсолютное отклонение нижней и верхней границ интервала $[m_1; m_2]$ от математического ожидания $M(m) = n \cdot p$ составляет 50 единиц:

$|m_1 - n \cdot p| = |950 - 100000 \cdot 0{,}01| = 50;$

$|m_2 - n \cdot p| = |1050 - 100000 \cdot 0{,}01| = 50.$

Следовательно, искомую вероятность можно рассматривать как вероятность заданного отклонения частоты от своего математического ожидания:

$$P(|m - np| < \Delta).$$

Подставив параметры биномиального распределения в формулу расчета вероятности заданного отклонения нормально распределенной случайной величины от своего математического ожидания, получим формулу для приближенного расчета вероятности заданного отклонения частоты от своего математического ожидания:

$$P(|(X = m) - M(m)| < \Delta) \approx 2 \cdot \Phi_0(\frac{\Delta}{\sigma(m)});$$

$$P(|m - np| < \Delta) \approx 2 \cdot \Phi_0(\frac{\Delta}{\sqrt{npq}}). \quad (5.20)$$

При использовании этой формулы для решения задачи сразу получим:

$$P(|m - 100000 \cdot 0{,}01| < 50) \approx 2 \cdot \Phi_0(\frac{50}{\sqrt{100000 \cdot 0{,}01 \cdot 0{,}99}});$$

$$P(|m - 1000| < 50) \approx 2 \cdot \Phi_0(1{,}59) = 0{,}88816.$$

Ответ. Вероятность того, что в партии из 100000 единиц окажется от 950 до 1050 дефектных единиц товара, составляет 0,88816.

Пример 5.6. Фирма собирается приобрести партию из 100 000 единиц некоторого товара. Из прошлого опыта известно, что 1% товаров данного типа имеют дефекты. Какова вероятность того, что в данной партии окажется 1020 дефектных единиц товара?

Решение. Так как $npq > 9$, применим локальную теорему Муавра-Лапласа:

$$P_{n;m} \approx \frac{\varphi(x)}{\sqrt{npq}}.$$

Найдем $x = \frac{m - np}{\sqrt{npq}} = \frac{1020 - 100000 \cdot 0{,}01}{\sqrt{100000 \cdot 0{,}01 \cdot 0{,}99}} = \frac{1020 - 1000}{31{,}4643} = 0{,}64$.

Найдем $\varphi(x)$ по таблице плотности вероятностей стандартной нормальной случайной величины (приложение 1) - $\varphi(0{,}64) = 0{,}3251$.

$$P_{100000;1020} \approx \frac{0{,}3251}{\sqrt{100000 \cdot 0{,}01 \cdot 0{,}99}} = 0{,}0103.$$

Ответ. Вероятность того, что в партии из 100000 единиц окажется 1020 дефектных единиц товара, составляет 0,0103.

Пример 5.7. Подлежат исследованию 400 проб руды. Вероятность промышленного содержания металла в каждой пробе для всех проб одинакова и равна 0,8. Найти вероятность того, что доля проб с промышленным содержанием металла отклонится от вероятности промышленного содержания металла в каждой пробе не более чем на 0,05.

Решение. В отличие от предыдущей задачи, в данном случае речь идет о расчете вероятности заданного отклонения частости (относительной частоты) появления события от вероятности его появления в отдельном независимом испытании, т.е.:

$$P\left(\left|\frac{m}{n} - M\left(\frac{m}{n}\right)\right| \le \Delta\right) = ?$$

При возрастании числа независимых испытаний распределение частости стремится к нормальному распределению точно так же, как и распределение частоты. Это означает, что при больших *n* мы можем аппроксимировать распределение частости нормальным распределением случайной величины, имеющей такое же математическое ожидание и такое же среднее квадратическое отклонение.

Подставив параметры распределения частости в формулу расчета вероятности заданного отклонения нормально распределенной случайной величины от своего математического ожидания, получим формулу для приближенного расчета вероятности заданного отклонения частости от своего математического ожидания (вероятности):

Параметры распределения частости:

$$M\left(\frac{m}{n}\right) = \frac{np}{n} = p; \qquad D\left(\frac{m}{n}\right) = \frac{npq}{n^2} = \frac{pq}{n}; \qquad \sigma\left(\frac{m}{n}\right) = \sqrt{\frac{pq}{n}}.$$

Используя эти формулы, получим:

$$P\left(\left|\left(X=\frac{m}{n}\right)-M\left(\frac{m}{n}\right)\right|<\Delta\right)\approx 2\Phi_0\left(\frac{\Delta}{\sigma\left(\frac{m}{n}\right)}\right);$$

$$P\left(\left|\frac{m}{n}-p\right|<\Delta\right)\approx 2\Phi_0\left(\frac{\Delta}{\sqrt{\frac{pq}{n}}}\right). \qquad (5.21)$$

Применим данную формулу для решения задачи.

По условию: n = 400; p = 0,8; q = 1- 0,8 = 0,2; Δ = 0,05.

$$P\left(\left|\frac{m}{400}-0{,}8\right|\le 0{,}05\right)\approx 2\Phi_0\left(\frac{0{,}05}{\sqrt{\frac{0{,}8\cdot 0{,}2}{400}}}\right);$$

$$P(\left|\frac{m}{400}-0{,}8\right|\le 0{,}05)\approx 2\Phi_0(2{,}5)=2\cdot 0{,}49379=0{,}98758.$$

Ответ. Вероятность того, что доля проб с промышленным содержанием металла отклонится от вероятности промышленного содержания металла в каждой пробе не более чем на 0,05, составляет 0,98758.

6. ВАРИАЦИОННЫЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ

6.1. Понятие вариационного ряда. Виды вариационных рядов

Совокупность предметов или явлений, объединенных каким-либо общим признаком или свойством качественного или количественного характера, называется ***объектом наблюдения***.

Всякий объект статистического наблюдения состоит из отдельных элементов - ***единиц наблюдения***.

Результаты статистического наблюдения представляют собой числовую информацию - ***данные***. ***Статистические данные*** - это сведения о том, какие значения принял интересующий исследователя признак в статистической совокупности.

Если значения признака выражаются числами, то признак называется ***количественным***.

Если признак характеризует некоторое свойство или состояние элементов совокупности, то признак называется ***качественным***.

Если исследованию подлежат все элементы совокупности (сплошное наблюдение), то статистическую совокупность называют ***генеральной.***

Если исследованию подлежит часть элементов генеральной совокупности, то статистическую совокупность называют ***выборочной (выборкой)***. Выборка из генеральной совокупности извлекается случайно, так, чтобы каждый из n элементов выборки имел равные шансы быть отобранным.

Значения признака при переходе от одного элемента совокупности к другому изменяются (варьируют), поэтому в статистике различные значения признака также называют ***вариантами***. Варианты обычно обозначаются малыми латинскими буквами x, y, z.

Порядковый номер варианта (значения признака) называется ***рангом***. x_1 - 1-й вариант (1-е значение признака), x_2 - 2-й вариант (2-е значение признака), x_i - i-й вариант (i-е значение признака).

Упорядоченный в порядке возрастания или убывания ряд значений признака (вариантов) с соответствующими им весами называется ***вариационным рядом (рядом распределения).***

В качестве ***весов*** выступают частоты или частости.

Частота (m_i) показывает сколько раз встречается тот или иной вариант (значение признака) в статистической совокупности.

Частость или относительная частота (w_i) показывает, какая часть единиц совокупности имеет тот или иной вариант. Частость рассчитывается как отношение частоты того или иного варианта к сумме всех частот ряда.

$$w_i = \frac{m_i}{\sum_{i=1}^{k} m_i}. \tag{6.1}$$

Сумма всех частостей равна 1.

$$\sum_{i=1}^{n} w_i = 1. \tag{6.2}$$

Вариационные ряды бывают дискретными и интервальными.

Дискретные вариационные ряды строят обычно в том случае, если значения изучаемого признака могут отличаться друг от друга не менее чем на некоторую конечную величину.

В дискретных вариационных рядах задаются точечные значения признака.

Общий вид дискретного вариационного ряда указан в таблице 6.1.

Таблица 6.1

Значения признака (x_i)	x_1	x_2	…	x_k
Частоты (m_i)	m_1	m_2	…	m_k

где i = 1, 2, …, k.

Интервальные вариационные ряды строят обычно в том случае, если значения изучаемого признака могут отличаться друг от друга на сколь угодно малую величину.

В интервальных вариационных рядах значения признака задаются в виде интервалов.

Общий вид интервального вариационного ряда показан в таблице 6.2.

Таблица 6.2

Значения признака	$a_1 - a_2$	$a_2 - a_3$	...	$a_{l-1} - a_l$
Частоты (m_i)	m_1	m_2	...	m_l

где i = 1, 2, … , *l*.

В интервальных вариационных рядах в каждом интервале выделяют верхнюю и нижнюю границы интервала.

Разность между верхней и нижней границами интервала называют ***интервальной разностью*** или ***длиной (величиной) интервала***.

Величина первого интервала k_1 определяется по формуле:

$k_1 = a_2 - a_1$;

второго: $k_2 = a_3 - a_2$;

последнего: $k_l = a_l - a_{l-1}$.

В общем виде *интервальная разность* k_i рассчитывается по формуле:

$$k_i = x_{i\,(max)} - x_{i\,(min)}. \qquad (6.3)$$

Если интервал имеет обе границы, то его называют ***закрытым***.

Первый и последний интервалы могут быть ***открытыми***, т.е. иметь только одну границу.

Например, первый интервал может быть задан как "до 100", второй - "100-110", … , предпоследний - "190-200", последний - "200 и более". Очевидно, что первый интервал не имеет нижней границы, а последний - верхней, оба они - открытые.

Часто открытые интервалы приходится условно закрывать. Для этого обычно величину первого интервала принимают равной величине второго, а величину последнего - величине предпоследнего. В нашем примере величина второго интервала равна 110-100=10, следовательно, нижняя граница первого интервала условно составит 100-10=90; величина предпоследнего интервала равна 200-190=10, следовательно, верхняя граница последнего интервала условно составит 200+10=210.

Кроме этого, в интервальном вариационном ряде могут встречаются интервалы разной длины. Если интервалы в вариационном ряде имеют одинаковую длину (интервальную разность), их называют ***равновеликими***, в противном случае - ***неравновеликими.***

При построении интервального вариационного ряда часто встает проблема выбора величины интервалов (интервальной разности).

Для определения оптимальной величины интервалов (в том случае, если строится ряд с равными интервалами) применяют *формулу Стэрджесса:*

$$k = \frac{x_{(\max)} - x_{(\min)}}{1 + 3{,}322 \lg n}, \qquad (6.4)$$

где n - число единиц совокупности,

$x_{(max)}$ и $x_{(min)}$ - наибольшее и наименьшее значения вариантов ряда.

Для характеристики вариационного ряда наряду с частотами и частостями используются накопленные частоты и частости.

Накопленные частоты (частости) показывают сколько единиц совокупности (какая их часть) не превышают заданного значения (варианта) x.

Накопленные частоты (v_i) по данным дискретного ряда можно рассчитать по следующей формуле:

$$v_i = m_i + m_{i-1} + ... + m_1. \qquad (6.5)$$

Для интервального вариационного ряда - это сумма частот (частостей) всех интервалов, не превышающих данный.

Дискретный вариационный ряд графически можно представить с помощью *полигона распределения частот или частостей*.

При построении полигона распределения по оси абсцисс откладываются значения признака (варианты), а по оси ординат - частоты или частости. На пересечении значений признака и соответствующих им частот (частостей) откладываются точки, которые, в свою очередь, соединяются отрезками. Получающаяся таким образом ломаная называется полигоном распределения частот (частостей).

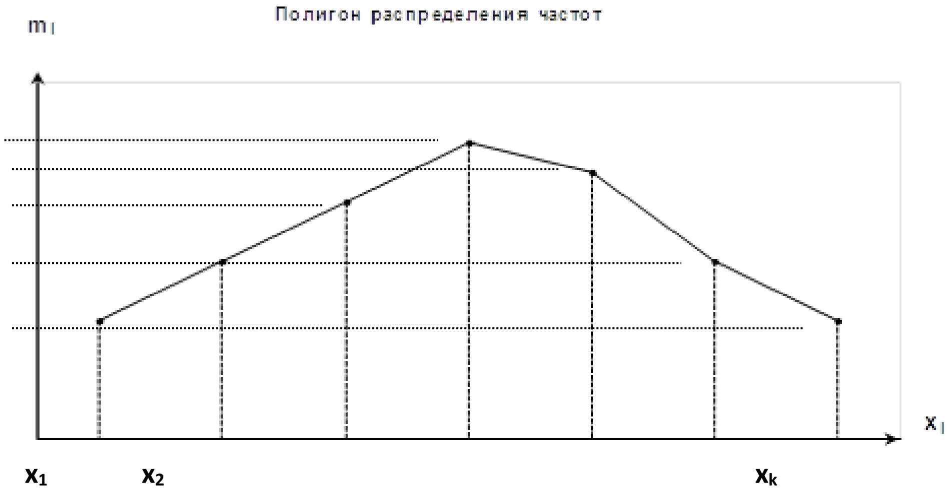


Рис. 6.1

Интервальные вариационные ряды графически можно представить с помощью *гистограммы*, т.е. столбчатой диаграммы.

При построении гистограммы по оси абсцисс откладываются значения изучаемого признака (границы интервалов).

В том случае, если интервалы - одинаковой величины, по оси ординат можно откладывать частоты или частости.

Если же интервалы имеют разную величину, по оси ординат необходимо откладывать значения абсолютной или относительной плотности распределения.

Абсолютная плотность - отношение частоты интервала к величине интервала:

$$f(a)_i = \frac{m_i}{k_i}, \tag{6.6}$$

где $f(a)_i$ - абсолютная плотность i-го интервала;

m_i - частота i-го интервала;

k_i - величина i-го интервала (интервальная разность).

Абсолютная плотность показывает, сколько единиц совокупности приходится на единицу интервала.

Относительная плотность - отношение частости интервала к величине интервала:

$$f(o)_i = \frac{w_i}{k_i}, \qquad (6.7)$$

где $f(o)_i$ - относительная плотность i-го интервала;

w_i - частость i-го интервала.

Относительная плотность показывает, какая часть единиц совокупности приходится на единицу интервала.

Рис. 6.2

И дискретные и интервальные вариационные ряды графически можно представить в виде кумуляты и огивы.

При построении *кумуляты* по данным дискретного ряда по оси абсцисс откладываются значения признака (варианты), а по оси ординат - накопленные частоты или частости. На пересечении значений признака (вариантов) и соответствующих им накопленных частот (частостей) строятся точки, которые, в свою очередь, соединяются отрезками или кривой. Получающаяся таким образом ломаная (кривая) называется кумулятой (кумулятивной кривой).

При построении кумуляты по данным интервального ряда по оси абсцисс откладываются границы интервалов. Абсциссами точек являются верхние границы интервалов. Ординаты образуют накопленные частоты (частости) соответствующих интервалов. Часто добавляют еще одну точку, абсциссой которой является нижняя граница первого интервала, а ордината равна нулю. Соединяя точки отрезками или кривой, получим кумуляту.

Огива строится аналогично кумуляте с той лишь разницей, что на оси абсцисс наносятся точки, соответствующие накопленным частотам (частостям), а по оси ординат - значения признака (варианты).

6.2. Числовые характеристики вариационного ряда

Одной из основных числовых характеристик ряда распределения (вариационного ряда) является ***средняя арифметическая.***

Существует две формулы расчета средней арифметической: простая и взвешенная.

Простую среднюю арифметическую обычно используют, когда данные наблюдения не сведены в вариационный ряд либо все частоты равны единице или одинаковы.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (6.8)$$

где x_i - i-е значение признака;

n - объем ряда (число наблюдений; число значений признака).

В том случае, если частоты отличны друг от друга, расчет производится по формуле ***средней арифметической взвешенной***:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot m_i}{\sum_{i=1}^{k} m_i}, \qquad (6.9)$$

где x_i - i-е значение признака;

m_i - частота i-го значения признака;

k - число значений признака (вариантов).

При расчете средней арифметической в качестве весов могут выступать и частости. Тогда формула расчета средней арифметической взвешенной примет следующий вид:

$$\bar{x} = \sum_{i=1}^{k} x_i \cdot w_i, \qquad (6.10)$$

где x_i - i-е значение признака;

w_i - частость i-го значения признака;

k - число значений признака (вариантов).

Колеблемость изучаемого признака можно охарактеризовать с помощью различных *показателей вариации*. К числу основных показателей вариации относятся: дисперсия, среднее квадратическое отклонение, коэффициент вариации.

Дисперсию можно рассчитать по простой и взвешенной формуле.

Простая имеет вид:

$$D(X) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}. \qquad (6.11)$$

А *взвешенная* -

$$D(X) = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot m_i}{\sum_{i=1}^{k} m_i}. \qquad (6.12)$$

Среднее квадратическое отклонение рассчитывается по формуле:

$$\sigma(X) = \sqrt{D(X)}. \qquad (6.13)$$

Коэффициент вариации рассчитывается по формуле:

$$V(X) = \frac{\sigma(X)}{\bar{x}} \cdot 100\%. \quad (6.14)$$

Пример 6.1. При обследовании 50 членов семей рабочих и служащих установлено следующее количество членов семьи: 5; 3; 2; 1; 4; 6; 3; 7; 9; 1; 3; 2; 5; 6; 8; 2; 5; 2; 3; 6; 8; 3; 4; 4; 5; 6; 5; 4; 7; 5; 6; 4; 8; 7; 4; 5; 7; 8; 6; 5; 7; 5; 6; 6; 7; 3; 4; 6; 5; 4.

а) Составьте вариационный ряд распределения частот.

б) Постройте полигон распределения частот, кумуляту.

в) Определите средний размер (среднее число членов) семьи.

г) Охарактеризуйте колеблемость размера семьи с помощью показателей вариации (дисперсии, среднего квадратического отклонения, коэффициента вариации).

Объясните полученные результаты, сделайте выводы.

Решение.

а) В данной задаче изучаемый признак является дискретно варьирующим, т.к. размер семей не может отличаться друг от друга менее чем на одного человека. Следовательно, необходимо построить дискретный вариационный ряд.

Чтобы построить вариационный ряд, необходимо подсчитать: сколько раз встречаются те или иные значения признака, и упорядочить их в порядке возрастания или убывания.

Значения изучаемого признака - размер семьи - обозначим x_i, частоты - m_i.

Произведем упомянутые расчеты и запишем полученные результаты в таблице 6.3:

Таблица 6.3

x_i	1	2	3	4	5	6	7	8	9
m_i	2	4	6	8	10	9	6	4	1

б) Дискретный вариационный ряд графически можно представить с помощью полигона распределения частот или частостей.

Построим полигон распределения частот (рис. 6.3):

Рис. 6.3

Для того чтобы построить кумуляту, необходимо рассчитать накопленные частоты или частости.

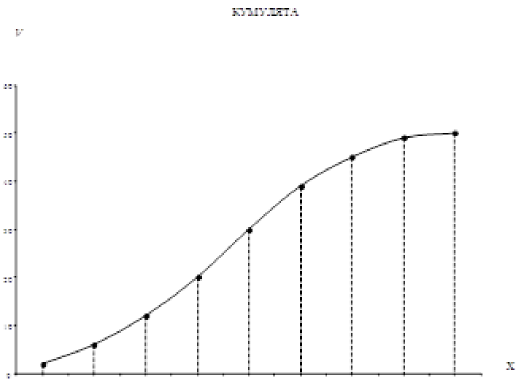
Накопленная частота первого варианта $x_1 = 1$ равна самой частоте этого варианта, т.е. двум: $v_1 = 2$.

Накопленная частота второго варианта $x_2 = 2$ равна сумме частот первого и второго вариантов, т.е. $v_2 = 2 + 4 = 6$.

Далее, аналогично:

$v_3 = 12$; $v_4 = 20$; $v_5 = 30$; $v_6 = 39$; $v_7 = 45$; $v_7 = 49$; $v_8 = 50$.

Построим кумуляту:



в) Рассчитаем средний размер (среднее число членов) семьи. Так как частоты отличны друг от друга, расчет средней арифметической произведем по формуле (6.9).

$$\bar{x} = \frac{1\cdot 2 + 2\cdot 4 + 3\cdot 6 + 4\cdot 8 + 5\cdot 10 + 6\cdot 9 + 7\cdot 6 + 8\cdot 4 + 9\cdot 1}{2+4+6+8+10+9+6+4+1} =$$

$$= \frac{2+8+24+32+50+54+42+32+9}{50} = \frac{253}{50} = 5{,}06.$$

Средний размер семьи - 5,06 человека.

г) Так как частоты - неодинаковы, для расчета дисперсии размера семьи используем формулу (6.12).

$$D(X) = \frac{(1-5{,}06)^2\cdot 2 + (2-5{,}06)^2\cdot 4 + (3-5{,}06)^2\cdot 6 + (4-5{,}06)^2\cdot 8}{2+4+6+8+10+9+6+4+1} +$$

$$+ \frac{(5-5{,}06)^2\cdot 10 + (6-5{,}06)^2\cdot 9 + (7-5{,}06)^2\cdot 6 + (8-5{,}06)^2\cdot 4}{50} +$$

$$+ \frac{(9-5{,}06)^2\cdot 1}{50} = 3{,}6964.$$

Дисперсия размера семьи - 3,6964 чел^2.

Найдем среднее квадратическое отклонение размера семьи по формуле (6.13).

$$\sigma(X) = \sqrt{3{,}6964} = 1{,}9226.$$

Среднее квадратическое отклонение размера семьи - 1,9226 чел.

Найдем коэффициент вариации размера семьи по формуле (6.14).

$$V(X) = \frac{1{,}9226}{5{,}06}\cdot 100\% = 38\%.$$

Коэффициент вариации составляет 38%. Так как коэффициент вариации больше 35%, можно сделать вывод о том, что изучаемая совокупность семей является неоднородной, чем и объясняется высокая колеблемость размера семьи в данной совокупности.

Ввиду неоднородности семей, попавших в выборку, использование средней арифметической для характеристики наиболее типичного уровня размера семьи не вполне оправданно - средняя арифметическая нетипична для изучаемой совокупности. В качестве характеристик наиболее типичного уровня размера семьи в данной совокупности лучше использовать моду или медиану.

Пример 6.2. Имеются данные о годовой мощности предприятий цементной промышленности:

Предприятия с годовой мощностью, тыс. тонн	Количество предприятий
до 500	27
500 – 1000	11
1000 – 2000	8
2000 – 3000	8
свыше 3000	2

а) Постройте гистограмму, кумуляту.

б) Рассчитайте среднюю мощность предприятий.

в) Найдите дисперсию, среднее квадратическое отклонение, коэффициент вариации.

Объясните полученные результаты, сделайте выводы.

Решение.

а) Данные о годовой мощности предприятий цементной промышленности представлены в виде интервального вариационного ряда - значения признака заданы в виде интервалов. При этом первый и последний интервалы - открытые: оба интервала не имеют одной из границ. Наконец, данный интервальный вариационный ряд - с неравными интервалами: интервальные разности (разность между верхней и нижней границами интервала) интервалов неодинаковы.

Условно закроем границы открытых интервалов.

Интервальная разность второго интервала равна: 1000 - 500 = 500. Следовательно, нижняя граница первого интервала составит: 500 - 500 = 0.

Интервальная разность предпоследнего интервала равна: 3000 - 2000 = 1000. Следовательно, верхняя граница последнего интервала составит: 3000 + 1000 = 4000.

В результате получим следующий вариационный ряд:

x_i	m_i
0 - 500	27
500 - 1000	11
1000 - 2000	8
2000 - 3000	8
3000 - 4000	2

Учитывая неодинаковую величину интервалов, для построения гистограммы рассчитаем абсолютные плотности распределения по формуле (6.6).

$$f(a)_1 = \frac{27}{500-0} = 0{,}054;$$

$$f(a)_2 = \frac{11}{1000-500} = 0{,}022;$$

$$f(a)_3 = \frac{8}{2000-1000} = 0{,}008;$$

$$f(a)_4 = \frac{8}{3000-2000} = 0{,}008;$$

$$f(a)_1 = \frac{2}{4000-3000} = 0{,}002.$$

Построим гистограмму.

Для того чтобы построить кумуляту, необходимо рассчитать накопленные частоты или частости.

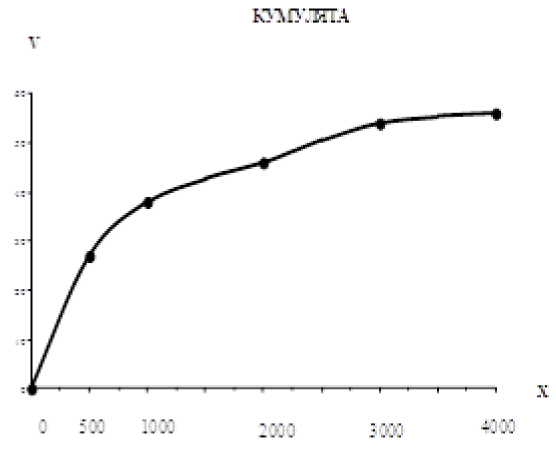
Накопленная частота нижней границы первого варианта x=0 равна нулю. Накопленная частота верхней границы первого интервала равна частоте этого интервала, т.е. 27.

Накопленная частота верхней границы второго интервала равна сумме частот первого и второго интервалов, т.е. 27 + 11 = 38.

Далее, аналогично:

38 + 8 = 46; 46 + 8 = 54; 54 + 2 = 56.

Построим кумуляту:



б) Рассчитаем среднюю мощность предприятий цементной промышленности.

Так как частоты интервалов - разные, используем для расчета средней арифметической формулу (6.9). При расчете числовых характеристик интервального вариационного ряда в качестве значений признака принимаются середины интервалов.

Рассчитаем середины интервалов:

$$x_1 = \frac{500+0}{2} = 250; \quad x_2 = \frac{1000+500}{2} = 750; \quad x_3 = \frac{2000+1000}{2} = 1500;$$

$$x_4 = \frac{3000+2000}{2} = 2500; \quad x_5 = \frac{4000+3000}{2} = 3500.$$

Теперь расчет средней арифметической примет вид:

$$\bar{x} = \frac{250 \cdot 27 + 750 \cdot 11 + 1500 \cdot 8 + 2500 \cdot 8 + 3500 \cdot 2}{27+11+8+8+2} =$$
$$= \frac{6750 + 8250 + 12000 + 20000 + 7000}{56} = \frac{54000}{56} = 964{,}2857.$$

Средняя мощность предприятий цементной промышленности составила 964,2857 тыс. тонн.

Следует отметить, что использование с той или иной целью средней арифметической, рассчитанной по данным интервального ряда с открытыми интервалами, может привести к серьезным ошибкам. Это связано с тем, что открытые интервалы закрываются условно, в действительности значения признака у объектов, попадающих в открытые интервалы, могут выходить далеко за их условные границы.

В связи с этим, для оценки наиболее типичного уровня изучаемого признака по данным интервального ряда с открытыми интервалами лучше использовать моду или медиану.

в) Оценим колеблемость мощности предприятий цементной промышленности.

Так как частоты - неодинаковы, для расчета дисперсии используем формулу (6.12)

$$D(X) = \frac{(250-964{,}2857)^2 \cdot 27 + (750-964{,}2857)^2 \cdot 11 + (1500-964{,}2857)^2 \cdot 8}{27+11+8+8+2} +$$
$$+ \frac{(2500-964{,}2857)^2 \cdot 8 + (3500-964{,}2857)^2 \cdot 2}{56} = 862563{,}7755.$$

Дисперсия мощности предприятий - 862563,7755 (тыс. тонн)^2.

Найдем среднее квадратическое отклонение мощности предприятий по формуле (6.13)

$$\sigma(X) = \sqrt{862563{,}7755} = 928{,}7431.$$

Среднее квадратическое отклонение мощности предприятий - 928,7431 тыс. тонн.

Найдем коэффициент вариации по формуле (6.14)

$$V(X) = \frac{928{,}7431}{964{,}28575} \cdot 100\% = 96{,}31\%.$$

Коэффициент вариации годовой мощности предприятий цементной промышленности составляет 96,31%. Так как коэффициент вариации больше 35%, можно сделать вывод о том, что изучаемая совокупность предприятий является неоднородной, в ее состав вошли и крупные, и мелкие предприятия, что и обусловило высокую колеблемость годовой мощности.

Следовательно, использование средней арифметической для характеристики наиболее типичного уровня годовой мощности предприятий цементной промышленности неверно - средняя арифметическая нетипична для изучаемой совокупности. Это еще раз подтверждает необходимость использования моды или медианы для характеристики наиболее типичного уровня годовой мощности данной совокупности предприятий цементной промышленности.

7. ВЫБОРОЧНЫЙ МЕТОД И СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

7.1. Основные понятия и определения выборочного метода

Одно из популярных определений статистики говорит, что это – ***наука, позволяющая распространять выводы, сделанные на основе изучения части совокупности (случайной выборки), на всю совокупность (генеральную совокупность).*** В этом определении заключена сущность выборочного метода и его ведущая роль в статистике.

Все единицы совокупности, обладающие интересующими исследователя признаками, составляют *генеральную совокупность*.

Часть совокупности, случайным образом отобранная из генеральной совокупности, – *выборочная совокупность* – выборка.[1]

Число единиц (элементов) статистической совокупности называется её объёмом. Объем генеральной совокупности обозначается N, а объем выборочной совокупности n. Если объем совокупности велик, то его полагают равным бесконечности.

Случайная выборка из n элементов - это такой отбор, при котором элементы извлекаются по одному из всей генеральной совокупности и каждый из них имеет равный шанс быть отобранным. Требование случайности обеспечивается отбором по таблицам случайных чисел или по жребию. Такая выборка называется ***собственно-случайной.*** Одним из примеров использования собственно-случайной выборки является проведение тиражей выигрышей денежно-вещевых лотерей, при которых обеспечивается равная возможность попадания в тираж любого номера лотерейного билета.

По способу отбора элементов различают два типа случайных выборок: ***собственно-случайная повторная*** выборка (схема возвращенного шара); ***собственно-случайная бесповторная*** выборка (схема невозвращенного шара).

Выбор схемы отбора зависит от характера изучаемого объекта. Напомним, что при повторном отборе единица наблюдения после извлечения из генеральной совокупности регистрируется и вновь возвращается в генеральную совокупность, откуда опять может быть извлечена случайным образом. При бесповторном отборе отобранный элемент в выборку обратно не возвращается. Необходимо заметить, что независимо от способа организации выборки она должна представлять собой уменьшенную копию генеральной совокупности, то есть быть *представительной (репрезентативной).*

---

[1] В учебниках по математической статистике вместо термина “статистическая совокупность” используется термин “набор данных”, а вместо термина “единица совокупности” используется термин “элемент выборки”.

7.2. Статистическое оценивание

Пусть из генеральной совокупности извлекается выборка объема *n*, причем значение признака x_1 наблюдается m_1 раз, x_2 m_2 раз,..., x_k наблюдается m_k раз, $\sum_{i=1}^{k} m_i = n$ - объем выборки.

Мы можем сопоставить каждому значению x_i относительную частоту m_i/n**.**

Статистическим распределением выборки называют перечень возможных значений признака x_i и соответствующих ему частот или относительных частот (частостей) m_i (w_i).

Числовые характеристики генеральной совокупности, как правило, неизвестные (средняя, дисперсия и др.), называют **параметрами генеральной совокупности** (обозначают, например, $\overline{X}$ или $\overline{\mathrm{X}}_{ген.}$, $\sigma^2{}_{ген.}$). Доля единиц, обладающих тем или иным признаком в генеральной совокупности, называется генеральной долей и обозначается *p***.**

По данным выборки рассчитывают числовые характеристики, которые называют **статистиками** (обозначают $\tilde{X}$, или $\tilde{X}_{выб.}$, $\sigma^2{}_{выб.}$, выборочная доля обозначается *w*). Статистики, получаемые по различным выборкам, как правило, отличаются друг от друга. Поэтому статистика, полученная из выборки, является только *оценкой* неизвестного параметра генеральной совокупности. *Оценка параметра - определенная числовая характеристика, полученная из выборки.* Когда оценка определяется одним числом, ее называют **точечной оценкой**.

В качестве точечных оценок параметров генеральной совокупности используются соответствующие выборочные характеристики. Теоретическое обоснование возможности использования этих выборочных оценок для суждений о характеристиках и свойствах генеральной совокупности дают закон больших чисел и центральная предельная теорема Ляпунова.

Выборочная средняя является точечной оценкой генеральной средней, т.е. $\tilde{\mathrm{X}}_{выб.} \approx \overline{\mathrm{X}}_{ген.}$.

Генеральная дисперсия имеет 2 точечные оценки: - $\sigma^2_{выб.}$ выборочная дисперсия; S^2 - исправленная выборочная дисперсия[2]. -$\sigma^2_{выб.}$ исчисляется при $\mathrm{n} \geq 30$, а S^2 - при $\mathrm{n} < 30$. Причем в математической статистике доказывается, что

$$S^2 = \frac{\sum_{i=1}^{k}(X_i - \tilde{X}_{выб.})^2 \cdot n_i}{n-1}, \quad \text{или} \quad S^2 = \frac{n}{n-1} \cdot \sigma^2{}_{выб.} \qquad (7.1)$$

При больших объемах выборки $\sigma^2_{выб.}$ и S^2 практически совпадают.

[2] Для того, чтобы любые статистики служили хорошими оценками параметров генеральной совокупности, они должны обладать рядом свойств: несмещённости, эффективности, состоятельности, достаточности. Всем указанным свойствам отвечает выборочная средняя. $\sigma^2_{выб.}$ -смещённая оценка. Для устранения смещения при малых выборках вводится поправка $n/n{-}1$ (см. 7.1).

Генеральное среднее квадратическое отклонение $\sigma_{ген.}$ также имеет 2 точечные оценки: $\sigma_{выб.}$ - выборочное среднее квадратическое отклонение и S - исправленное выборочное среднее квадратическое отклонение. $\sigma_{выб.}$ используется для оценивания $\sigma_{ген.}$ при $n \geq 30$, а **S** для оценивания $\sigma_{ген.}$, при $n < 30$; при этом $\sigma_{выб.} = \sqrt{\sigma^2_{выб.}}$, а $S = \sqrt{S^2}$.

7.3. Ошибки выборки

Поскольку выборочная совокупность представляет собой лишь часть генеральной совокупности, то вполне естественно, что выборочные характеристики не будут точно совпадать с соответствующими генеральными. Ошибка репрезентативности может быть представлена как разность между генеральными и выборочными характеристиками изучаемой совокупности: $\varepsilon = \tilde{X} - \overline{X}$, либо $\varepsilon = p - w$.

Применительно к выборочному методу из теоремы Чебышева следует, что с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности разность между выборочной средней и генеральной средней будет сколь угодно мала.

$$P\left(\left|\tilde{X} - \overline{X}\right| < \frac{t \cdot \sigma_{ген.}}{\sqrt{n}}\right) > 1 - \frac{1}{t^2}, \tag{7.2}$$

где $\tilde{X}$ - средняя по совокупности выбранных единиц,

$\overline{X}$ - средняя по генеральной совокупности,

$\sigma_{ген.}$ - среднее квадратическое отклонение в генеральной совокупности.

Запись показывает, что о величине расхождения между параметром и статистикой $\Delta = \frac{t \cdot \sigma}{\sqrt{n}} = t \cdot \mu$ можно судить лишь с определенной вероятностью, от которой зависит величина t.

Формула (7.2) устанавливает связь между пределом ошибки $\mathbf{\Delta}$, гарантируемым с некоторой вероятностью P, величиной t и средней ошибкой выборки $\mu = \frac{\sigma_{ген.}}{\sqrt{n}}$.

Согласно центральной предельной теореме Ляпунова выборочные распределения статистик (при n ≥ 30) будут иметь нормальное распределение независимо от того, какое распределение имеет генеральная совокупность. Следовательно:

$$P\left(\left|\tilde{X} - \overline{X}\right| < t \cdot \mu\right) \approx 2\Phi_0(t), \tag{7.3}$$

где Φ_0(t) - функция Лапласа.

Значения вероятностей, соответствующие различным t, содержатся в специальных таблицах: при n ≥ 30 - в таблице значений Φ_0(t), а при n < 30 в таблице распределения t-Стьюдента. Неизвестное значение $\sigma_{ген.}$ при расчете ошибки выборки заменяется $\sigma_{выб.}$.

В зависимости от способа отбора средняя ошибка выборки определяется по-разному:

Таблица 7.1

Формулы расчёта ошибки выборки для собственно-случайного отбора

μ	Собственно-случайный повторный отбор	Собственно-случайный бесповторный отбор
Для средней	$\sqrt{\frac{\sigma^2}{n}}$	$\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$
Для доли	$\sqrt{\frac{w(1-w)}{n}}$	$\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$

Здесь σ^2 - выборочная дисперсия значений признака;

$w(1-w)$ - выборочная дисперсия доли значений признака;

n - объем выборки;

N - объем генеральной совокупности;

$\frac{n}{N}$ - доля обследованной совокупности;

$\left(1-\frac{n}{N}\right)$ - поправка на конечность совокупности[3].

7.4. Определение численности (объема) выборки

Одной из важнейших проблем выборочного метода является определение необходимого объема выборки. От объема выборки зависит размер средней ошибки (μ) и экономичность проводимого выборочного наблюдения, т.к. чем больше объем выборки, тем больше затраты на изучение элементов выборки, но тем меньше при этом ошибка выборки.

Из формулы предельной ошибки $\Delta = t \cdot \mu$ и формул средних ошибок выборки определяются формулы необходимой численности выборки для различных способов отбора.

Таблица 7.2

Формулы расчёта необходимой численности выборки
для собственно-случайного отбора

n	Собственно-случайный повторный отбор	Собственно-случайный бесповторный отбор
Для средней	$\frac{t^2\sigma^2}{\Delta^2}$	$\frac{t^2\sigma^2 N}{N\Delta^2 + t^2\sigma^2}$
Для доли	$\frac{t^2 w(1-w)}{\Delta^2}$	$\frac{t^2 N w(1-w)}{N\Delta^2 + t^2 w(1-w)}$

[3] В литературе (1 – n /N) иногда называется "поправкой на бесповторность отбора".

7.5. Интервальное оценивание

Мы уже знаем, что $\varepsilon = \tilde{X} - \overline{X}$. Если Δ представляет собой предел, которым ограничена сверху абсолютная величина $|\varepsilon| < \Delta$, то $\left|\tilde{X} - \overline{X}\right| < \Delta$. Следовательно,

$$\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta. \tag{7.4}$$

Мы получили интервальную оценку генеральной средней. Из теоремы Чебышева следует, что

$$P(\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta) = 2\Phi_0(t) = \gamma. \tag{7.5}$$

Интервальной оценкой называют оценку, которая определяется двумя числами - концами интервала, который с определенной вероятностью накрывает неизвестный параметр генеральной совокупности. Интервал, содержащий оцениваемый параметр генеральной совокупности, называют ***доверительным интервалом.*** Для его определения вычисляется ***предельная ошибка выборки*** $\boldsymbol{\Delta}$ **,** позволяющая установить предельные границы, в которых с заданной вероятностью (надёжностью) должен находиться параметр генеральной совокупности.

Предельная ошибка выборки равна t-кратному числу средних ошибок выборки. Коэффициент t позволяет установить, насколько надежно высказывание о том, что заданный интервал содержит параметр генеральной совокупности. Если мы выберем коэффициент таким, что высказывание в 95% случаев окажется правильным и только в 5% - неправильным, то мы говорим: *со статистической надежностью в 95% доверительный интервал выборочной статистики содержит параметр генеральной совокупности.* Статистической надежности в 95% соответствует доверительная вероятность - 0,95. В 5% случаев утверждение "параметр принадлежит доверительному интервалу" будет неверным. То есть 5% задает *уровень значимости* (α) или 0,05 вероятность ошибки. Обычно в статистике уровень значимости выбирают таким, чтобы он не превысил 5% (α < 0,05). Доверительная вероятность и уровень значимости дополняют друг друга до 1 (или 100%) и определяют надежность статистического высказывания.

С помощью доверительного интервала можно оценить не только генеральную среднюю, но и другие неизвестные параметры генеральной совокупности.

Для оценки *математического ожидания a (генеральной средней)*[4] нормально распределенного количественного признака X по выборочной средней $\tilde{x}$ при известном среднем квадратическом отклонении $\boldsymbol{\sigma}$ генеральной совокупности (на практике - *при большом объеме выборки*, т.е. при n ≥ 30) и *собственно-случайном повторном отборе* формула (7.5) примет вид:

$$P\left(\tilde{X} - t \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} < \tilde{X} + t \cdot \frac{\sigma}{\sqrt{n}}\right) = 2\Phi_0(t) = \gamma, \tag{7.6}$$

---

[4] Для нормально распределенной случайной величины $M(\tilde{X}) = a \approx \overline{X}$. Поэтому справедливо: $P\left(\left|\tilde{X} - \mathrm{a}\right| < \Delta\right) \approx \overline{X}$.

где t определяется по таблицам функции Лапласа из соотношения $2\Phi_0(t) = \gamma$;

σ - среднее квадратическое отклонение;

n - объем выборки (число обследованных единиц).

Δ определяется по формуле:

$$\Delta = t \cdot \frac{\sigma}{\sqrt{n}}.$$

Для оценки *математического ожидания а (генеральной средней)* нормально распределенного количественного признака X по выборочной средней $\tilde{X}$ при известном среднем квадратическом отклонении **σ** генеральной совокупности (*при большом объеме выборки*, т.е. при $n \geq 30$) и *собственно-случайном бесповторном отборе* формула (7.6) примет вид:

$$P\left(\tilde{X} - t \cdot \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}} < \overline{X} < \tilde{X} + t \cdot \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}\right) = 2\Phi_0(t) = \gamma. \quad (7.7)$$

Δ определяется по формуле:

$$\Delta = t \cdot \frac{\sigma}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}.$$

Для оценки *математического ожидания а (генеральной средней)* нормально распределенного количественного признака X по выборочной средней $\tilde{X}$ при неизвестном среднем квадратическом отклонении σ генеральной совокупности (на практике - *при малом объеме выборки*, т.е. при n < 30) и *собственно-случайном повторном отборе* формула (7.6) примет вид:

$$P\left(\tilde{X} - t \cdot \frac{s}{\sqrt{n}} < \overline{X} < \tilde{X} + t \cdot \frac{s}{\sqrt{n}}\right) = 2S(t) = \gamma, \quad (7.8)$$

где t определяется по таблицам Стьюдента по уровню значимости $\alpha = 1 - \gamma$ и числу степеней свободы $k = n - 1$;

s - исправленное выборочное среднее квадратическое отклонение;

n - объем выборки.

Δ определяется по формуле:

$$\Delta = t \cdot \frac{s}{\sqrt{n}}.$$

Для оценки *математического ожидания а (генеральной средней)* нормально распределенного количественного признака X по выборочной средней $\tilde{X}$ при неизвестном среднем квадратическом отклонении σ генеральной совокупности (*при малом объеме выборки*, т.е. при n < 30) и *собственно-случайном бесповторном отборе* формула (7.8) примет вид:

$$P\left(\tilde{X} - t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}} < \overline{X} < \tilde{X} + t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}\right) = 2S(t) = \gamma. \quad (7.9)$$

Δ определяется по формуле:

$$\Delta = t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}.$$

Для оценки *генеральной доли p* нормально распределенного количественного признака по выборочной доле $w = \frac{m}{n}$ *при большом объеме выборки*, т.е. при n $\geq$ 30), и *собственно-случайном повторном отборе* формула (7.5) примет вид:

$$P\left(w - t \cdot \sqrt{\frac{w(1-w)}{n}} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n}} \right) = 2\Phi_0(t) = \gamma, \qquad (7.10)$$

где t определяется по таблицам функции Лапласа из соотношения $2\Phi_0(t) = \gamma$;
w - выборочная доля;
n - объем выборки (число обследованных единиц).

Δ определяется по формуле:

$$\Delta = t \cdot \sqrt{\frac{w(1-w)}{n}}.$$

Для оценки *генеральной доли p* нормально распределенного количественного признака по выборочной доле $w = \frac{m}{n}$ *при большом объеме выборки*, т.е. при n $\geq$ 30, и *собственно-случайном бесповторном отборе* формула (7.10) примет вид:

$$P\left(w - t \cdot \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} \right) = 2\Phi_0(t) = \gamma. \qquad (7.11)$$

Δ определяется по формуле:

$$\Delta = t \cdot \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)}.$$

Для оценки *генеральной доли p* нормально распределенного количественного признака по выборочной доле $w = \frac{m}{n}$ *при малом объеме выборки*, т.е. при n < 30, и *собственно-случайном повторном отборе* формула (7.10) примет вид:

$$P\left(w - t \cdot \sqrt{\frac{w(1-w)}{n-1}} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n-1}} \right) = 2S(t) = \gamma, \qquad (7.12)$$

где t определяется по таблицам Стьюдента по уровню значимости $\alpha = 1 - \gamma$ и числу степеней свободы k = n - 1.

Δ определяется по формуле:

$$\Delta = t \cdot \sqrt{\frac{w(1-w)}{n-1}}.$$

Для оценки *генеральной доли p* нормально распределенного количественного признака по выборочной доле $w = \frac{m}{n}$ *при малом объеме выборки*, т.е. при n < 30, и *собственно-случайном бесповторном отборе* формула (7.12) примет вид:

$$P\left(w - t \cdot \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n-1}\left(1 - \frac{n}{N}\right)} \right) = 2S(t) = \gamma. \qquad (7.13)$$

Δ определяется по формуле:

$$\Delta = t \cdot \sqrt{\frac{w(1-w)}{n-1}\left(1-\frac{n}{N}\right)}.$$

Пример 7.1. С помощью собственно-случайного повторного отбора руководство фирмы провело выборочное обследование 900 своих служащих. Средний стаж их работы в фирме равен 8,7 года, а среднее квадратическое (стандартное) отклонение - 2,7 года. Среди обследованных оказалось 270 женщин. Считая стаж работы служащих фирмы распределённым по нормальному закону, определите:

а) с вероятностью 0,95 доверительный интервал, в котором окажется средний стаж работы всех служащих фирмы;

б) с вероятностью 0,90 доверительный интервал, накрывающий неизвестную долю женщин во всем коллективе фирмы.

Решение. По условию задачи выборочное обследование проведено с помощью собственно-случайного повторного отбора. Объем выборки n = 900 единиц, т.е. выборка - большая.

а) Найдем границы доверительного интервала среднего стажа работы всего коллектива фирмы, т.е. границы доверительного интервала для генеральной средней.

По условию: $\tilde{X} = 8{,}7$; $\sigma = 2{,}7$; $n = 900$; $\gamma = 0{,}95$.

Используем формулу:

$$P(\tilde{X} - t \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} < \tilde{X} - t \cdot \frac{\sigma}{\sqrt{n}}) = 2\Phi_0(t) = \gamma .$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}95;$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t $\Phi_0(t)=0{,}475$.

$$\Phi_0(1{,}96)\ \Phi_0(t) = 0{,}95 / 2 = 0{,}475.$$

Следовательно, t = 1,96.

Найдем предельную ошибку выборки:

$$\Delta = t \cdot \frac{\sigma}{\sqrt{n}};$$

$$\Delta = 1{,}96 \cdot \frac{2{,}7}{\sqrt{900}} = 1{,}96 \cdot 0{,}09 = 0{,}1764 .$$

$$\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta ;$$

$$8{,}7 - 0{,}1764 < \overline{X} < 8{,}7 + 0{,}1764 ;$$

$$8{,}5236 < \overline{X} < 8{,}8764 .$$

С вероятностью 0,95 можно ожидать, что средний стаж работы всего коллектива фирмы находится в интервале от 8,5236 до 8,8764 года.

б) Теперь оценим истинное значение доли женщин во всем коллективе фирмы.

По условию: m = 270; n = 900; $\gamma = 0{,}9$.

Выборочная доля $w = \frac{270}{900} = 0{,}3.$

Рассмотрим формулу:

$$P\left(w - t \cdot \sqrt{\frac{w(1-w)}{n}} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n}} \right) = 2\Phi_0(t) = \gamma \,.$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}9;$$

$$\Phi_0(t) = 0{,}9 / 2 = 0{,}45.$$

По таблице функции Лапласа (см. приложение 2) определим, при каком t $\Phi_0(t) = 0{,}45$.

$\Phi_0(1{,}64) = 0{,}45$.

Следовательно, $t = 1{,}64$.

Предельная ошибка выборки определяется по формуле:

$$\Delta = t \cdot \sqrt{\frac{w(1-w)}{n}}\,;$$

$$\Delta = 1{,}64 \cdot \sqrt{\frac{0{,}3 \cdot (1-0{,}3)}{900}} = 1{,}64 \cdot \sqrt{\frac{0{,}3 \cdot 0{,}7}{900}} = 1{,}64 \cdot 0{,}0153 = 0{,}0251.$$

$$w - \Delta < p < w + \Delta;$$

$$0{,}3 - 0{,}0251 < p < 0{,}3 + 0{,}0251;$$

$$0{,}2749 < p < 0{,}3251\,.$$

Итак, с вероятностью 0,9 можно ожидать, что доля женщин во всем коллективе фирмы находится в интервале от 0,2749 до 0,3251.

Ответ. Можно ожидать, что с вероятностью 0,95, средний стаж работы всех служащих фирмы находится в интервале от 8,5236 до 8,8764 года. С вероятностью 0,90 можно гарантировать, что доля женщин во всем коллективе фирмы находится в интервале от 0,2749 до 0,3251.

Пример 7.2. Изменим условие примера 7.1.

а) С помощью собственно-случайного повторного отбора определяется средний стаж работы служащих фирмы. Предполагается, что он подчиняется нормальному закону. Каким должен быть объем выборки, чтобы с доверительной вероятностью 0,95 можно было утверждать, что (принимаем полученный средний стаж работы за истинный) совершается погрешность, не превышающая 0,5 года, если стандартное отклонение σ равно 2,7 года?

б) Каким должен быть объем собственно-случайной повторной выборки, чтобы с надежностью 0,90 можно было утверждать, что максимальное отклонение выборочной доли женщин в выборке от доли женщин во всем коллективе фирмы не превышало 0,05, если в прошлом аналогичном обследовании выборочная доля женщин оказалась равной 0,3?

Решение. В данной задаче нужно найти необходимую численность выборки. Расчет необходимой численности выборки дает ответ на вопрос: “Сколько нужно обследовать единиц совокупности, чтобы с заранее заданной вероятностью не превысить заранее заданную ошибку?”

а) Дано: $\Delta = 0{,}5$; $\sigma = 2{,}7$; $\gamma = 0{,}95$.

По условию задачи требуется найти необходимую численность выборки для средней при повторном отборе.

Воспользуемся формулой расчета необходимой численности выборки для средней для собственно-случайного повторного отбора:

$$n = \frac{t^2 \sigma^2}{\Delta^2}.$$

Неизвестное значение t найдем из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}95;$$

$$\Phi_0(t) = 0{,}95 / 2 = 0{,}475.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t

$\Phi_0(t) = 0{,}475$.

$\Phi_0(1{,}96) = 0{,}475$.

Следовательно, $t = 1{,}96$.

Рассчитаем необходимую численность выборки:

$$n = \frac{1{,}96^2 \cdot 2{,}7^2}{0{,}5^2} = 112{,}02.$$

Так как n - целое число, а также, учитывая необходимость не превысить заданную ошибку, округлим полученный результат до большего целого.

Следовательно, необходимо обследовать не менее 113 служащих.

Ответ. Чтобы с вероятностью 0,95 и $\Delta = 0{,}5$ года с помощью собственно-случайного повторного отбора определить средний стаж работы в фирме, необходимо обследовать не менее 113 служащих.

б) Дано: $\Delta = 0{,}05$; $w = 0{,}3$; $\gamma = 0{,}9$.

По условию задачи требуется найти необходимую численность выборки для доли для собственно-случайного повторного отбора.

Воспользуемся формулой расчета необходимой численности выборки для доли для собственно-случайного повторного отбора:

$$n = \frac{t^2 w(1-w)}{\Delta^2}.$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}9;$$

$$\Phi_0(t) = 0{,}9 / 2 = 0{,}45.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t

$\Phi_0(t) = 0{,}45$.

$\Phi_0(1{,}64) = 0{,}45$.

Следовательно, $t = 1{,}64$.

Рассчитаем необходимую численность выборки:

$$n = \frac{1{,}64^2 \cdot 0{,}3 \cdot (1-0{,}3)}{0{,}05^2} = 225{,}93.$$

Так как n - целое число, а также, учитывая необходимость не превысить заданную ошибку, округлим полученный результат до большего целого.

Следовательно, $n \approx 226$.

Ответ. Чтобы с вероятностью 0,9 и ошибкой $\Delta =$ 0,05 с помощью собственно-случайного повторного отбора определить долю женщин во всем коллективе фирмы, необходимо обследовать не менее 226 служащих.

Пример 7.3. Владелец автостоянки опасается обмана со стороны своих служащих (охраны автостоянки). В течение года (365 дней) владельцем автостоянки проведено 40 проверок. По данным проверок, среднее число автомобилей, оставляемых на ночь на охрану, составило 400 единиц, а среднее квадратическое (стандартное) отклонение их числа - 10 автомобилей.

Считая отбор собственно-случайным, с вероятностью 0,99 оцените с помощью доверительного интервала истинное среднее число автомобилей, оставляемых на ночь на охрану. Обоснованны ли опасения владельца автостоянки, если по отчетности охранников среднее число автомобилей, оставляемых на ночь на стоянку, составляет 395 автомобилей?

Решение. По условию задачи выборочное обследование проведено с помощью собственно-случайного отбора. Очевидно, что отбор - бесповторный, т.к. не имеет смысла производить проверку более 1 раза в сутки. Объем выборки n = 40, что больше 30 единиц, т.е. выборка - большая. Объём генеральной совокупности N==365.

Найдем границы доверительного интервала для оценки среднего числа автомобилей, оставляемых на ночь на охрану, т.е. границы доверительного интервала для генеральной средней.

По условию: $\tilde{X} = 400$; $\sigma = 10$; $n = 40$; $\gamma = 0,99$; $N=365$.

Используем формулу:

$$P\left(\tilde{X} - t\cdot\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}} < \overline{X} < \tilde{X} + t\cdot\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}\right) = 2\Phi_0(t) = \gamma\,.$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0,99;$$

$$\Phi_0(t) = 0,99 / 2 = 0,495.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t

$\Phi_0(t) = 0,495$.

$\Phi_0(2,58) = 0,495$.

Следовательно, $t = 2,58$.

Найдем предельную ошибку выборки:

$$\Delta = t\cdot\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}\,;$$

$$\Delta = 2,58\cdot\frac{10}{\sqrt{40}}\cdot\sqrt{1-\frac{40}{365}} = 3,8493\,.$$

$$\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta\,;$$

$$400 - 3,8493 < \overline{X} < 400 + 3,8493\,;$$

$$396,1507 < \overline{X} < 403,8493\,.$$

Ответ. С уверенностью в 99% можно ожидать, что среднее число автомобилей, оставляемых на ночь на охрану, находится в интервале от 396 до

404. Таким образом, можно утверждать, что служащие автостоянки обманывают ее владельца.

Пример 7.4. В 24 из 40 проверок число автомобилей на автостоянке не превышало 400 единиц.

С вероятностью 0,98 найдите доверительный интервал для оценки истинной доли дней в течение года, когда число оставляемых на стоянку автомобилей не превышало 400 единиц.

Решение. Определим границы доверительного интервала для доли дней в течение года, когда число оставляемых на стоянку автомобилей не превышало 400 единиц.

По условию: $m = 24$; $n = 40$; $\gamma = 0{,}98$.

Выборочная доля $w = \frac{24}{40} = 0{,}6.$

Так как $P\left(w - t\cdot\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} < p < w + t\cdot\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}\right) = 2\Phi_0(t) = \gamma$, то найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}98;$$

$$\Phi_0(t) = 0{,}98 / 2 = 0{,}49.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t $\Phi_0(t) = 0{,}49$.

$\Phi_0(2{,}33) = 0{,}49$.

Следовательно, $t = 2{,}33$.

Найдем предельную ошибку выборки:

$$\Delta = t\cdot\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)};$$

$$\Delta = 2{,}33\cdot\sqrt{\frac{0{,}6\cdot(1-0{,}6)}{40}\cdot\left(1-\frac{40}{365}\right)} = 2{,}33\cdot\sqrt{\frac{0{,}6\cdot 0{,}4}{40}\cdot\left(1-\frac{40}{365}\right)} = 0{,}1703.$$

$$w - \Delta < p < w + \Delta;$$

$$0{,}6 - 0{,}1703 < p < 0{,}6 + 0{,}1703;$$

$$0{,}4297 < p < 0{,}7703.$$

Ответ. С вероятностью 0,98 можно ожидать, что доля дней в течение года, когда число оставляемых на стоянку автомобилей не превышало 400 единиц, находится в интервале от 0,4297 до 0,7703.

Пример 7.5. Изменим условие примера 7.3.

С помощью собственно-случайного бесповторного отбора определяется среднее число автомобилей, оставляемых на ночь на охрану. Предполагается, что оно подчиняется нормальному закону. Каким должен быть объём выборки, чтобы с вероятностью 0,95 можно было утверждать, что (принимаем полученное среднее число автомобилей по выборке за истинное) совершается погрешность, не превышающая 3 автомобилей, если среднее квадратическое отклонение σ равно 10 автомобилям?

Решение. Дано: $\Delta = 3$; $\sigma = 10$; $\gamma = 0{,}95$; $N=365$.

Воспользуемся формулой расчета необходимой численности выборки для средней для собственно-случайного бесповторного отбора:

$$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta^2 \cdot N + t^2 \cdot \sigma^2}.$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}95;$$

$$\Phi_0(t) = 0{,}95 / 2 = 0{,}475;$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t

$\Phi_0(t) = 0{,}475$.

$\Phi_0(1{,}96) = 0{,}475$.

Следовательно, t = 1,96.

Рассчитаем объём выборки:

$$n = \frac{1{,}96^2 \cdot 10^2 \cdot 365}{3^2 \cdot 365 + 1{,}96^2 \cdot 10^2} = 38{,}22.$$

Так как n - целое число, а также, учитывая необходимость не превысить заданную ошибку, округлим полученный результат до большего целого.

Следовательно, необходимо провести не менее 39 проверок.

Ответ. Для определения среднего числа автомобилей, оставляемых на ночь на охрану с вероятностью 0,95 и Δ=3, необходимо провести не менее 39 проверок.

Пример 7.6. Изменим условие примера 7.4.

Каким должен быть объём собственно-случайной бесповторной выборки, чтобы с вероятностью 0,9 можно было утверждать, что максимальное отклонение выборочной доли дней от доли дней в течение года (когда среднее число оставляемых на охрану автомобилей не превышало 400 единиц) не превышало 0,1, если, по данным прошлых проверок, выборочная доля таких дней составляла 0,6?

Решение. Дано: $\Delta = 0{,}1$; $w = 0{,}6$; $\gamma = 0{,}9$; N=365.

Воспользуемся формулой расчета необходимой численности выборки для доли при собственно-случайном бесповторном отборе:

$$n = \frac{t^2 \cdot w \cdot (1-w) \cdot N}{\Delta^2 \cdot N + t^2 \cdot w \cdot (1-w)}.$$

Найдем t из соотношения $2\Phi_0(t) = \gamma$:

$$2\Phi_0(t) = 0{,}9;$$

$$\Phi_0(t) = 0{,}9 / 2 = 0{,}45.$$

По таблице функции Лапласа (см. приложение 2) найдем, при каком t

$\Phi_0(t) = 0{,}45$.

$\Phi_0(1{,}64) = 0{,}45$.

Следовательно, t = 1,64.

Рассчитаем необходимую численность выборки:

$$n = \frac{1{,}64^2 \cdot 0{,}6 \cdot (1-0{,}6) \cdot 365}{0{,}1^2 \cdot 365 + 1{,}64^2 \cdot 0{,}6 \cdot (1-0{,}6)} = 54{,}85.$$

Так как n - целое число, а также, учитывая необходимость не превысить заданную ошибку, округлим полученный результат до большего целого.

Следовательно, $n \approx 55$.

Ответ. Для того чтобы с вероятностью 0,9 и предельной ошибкой 0,1 с помощью собственно-случайного бесповторного отбора определить искомую долю дней в течение года, необходимо провести не менее 55 проверок.

Пример 7.7. Служба контроля Энергосбыта провела выборочную проверку расхода электроэнергии жителями одного из многоквартирных домов. С помощью собственно-случайного отбора выбрано 10 квартир и определен расход электроэнергии в течение одного из летних месяцев (кВт-час): 125; 78; 102; 140; 90; 45; 50; 125; 115; 112.

С надежностью 0,95 определите доверительный интервал для оценки среднего расхода электроэнергии на 1 квартиру во всем доме при условии, что в доме 70 квартир, а отбор был:

а) повторным;

б) бесповторным.

Решение. По условию задачи выборочное обследование проведено с помощью собственно-случайного отбора. Объем выборки n = 10 единиц, т.е. *выборка - малая*.

а) Считая отбор повторным, найдем доверительный интервал для оценки среднего расхода электроэнергии на 1 квартиру во всем доме, т.е. границы доверительного интервала для оценки генеральной средней.

Для этого используем формулы:

$$P\left(\tilde{X} - t \cdot \frac{s}{\sqrt{n}} < \overline{X} < \tilde{X} + t \cdot \frac{s}{\sqrt{n}}\right) = 2S(t) = \gamma;$$

$$\Delta = t \cdot \frac{s}{\sqrt{n}}.$$

Для определения границ доверительного интервала необходимо рассчитать выборочные среднюю и среднее квадратическое (стандартное) отклонение.

Рассчитаем выборочную среднюю арифметическую:

$$\tilde{X} = \frac{125 + 78 + 102 + 140 + 90 + 45 + 50 + 125 + 115 + 112}{10} = \frac{982}{10} = 98{,}2.$$

Найдем исправленную выборочную дисперсию:

$$s^2 = \frac{(125-98{,}2)^2 + (78-98{,}2)^2 + (102-98{,}2)^2 + (140-98{,}2)^2 + (90-98{,}2)^2}{10-1} +$$

$$+ \frac{(45-98{,}2)^2 + (50-98{,}2)^2 + (125-98{,}2)^2 + (115-98{,}2)^2 + (112-98{,}2)^2}{10-1} =$$

$$= 1033{,}2889.$$

Найдем исправленное выборочное среднее квадратическое отклонение:

$$s = \sqrt{s^2} = \sqrt{1033{,}2889} = 32{,}1448.$$

Итак, дано: $\tilde{X} == 98{,}2$; $s = 32{,}1448$; $n = 10$; $\gamma = 0{,}95$.

По таблице Стьюдента (приложение 5) найдем t по уровню значимости α и числу степеней свободы k.

$\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05$;

$k = n - 1 = 10 - 1 = 9$.

$t_{\alpha=0,05;\ k=9} = 2{,}26$.

Найдем предельную ошибку выборки:

$$\Delta = t \cdot \frac{s}{\sqrt{n}}.$$

$$\Delta = 2{,}26 \cdot \frac{32{,}1448}{\sqrt{10}} = 22{,}9731.$$

$$\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta;$$

$$98{,}2 - 22{,}9731 < \overline{X} < 98{,}2 + 22{,}9731;$$

$$75{,}2269 < \overline{X} < 121{,}1731.$$

Ответ. При условии, что отбор квартир был повторным, с вероятностью 0,95 можно ожидать, что средний расход электроэнергии на 1 квартиру во всем доме находится в интервале от 75,2269 до 121,1731 кВт-часа.

б) Найдем границы доверительного интервала для оценки среднего расхода электроэнергии на 1 квартиру во всем доме, считая отбор бесповторным.

Для этого используем формулы:

$$P\left(\tilde{X} - t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}} < \overline{X} < \tilde{X} + t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}\right) = 2S(t) = \gamma.$$

$$\Delta = t \cdot \frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}.$$

По условию: $\tilde{X} == 98{,}2$; $s = 32{,}1448$; $n = 10$; $\gamma = 0{,}95$; $t_{\alpha=0,05;\ k=9} = 2{,}26$; $N = 70$.

Найдем предельную ошибку выборки:

$$\Delta = 2{,}26 \cdot \frac{32{,}1448}{\sqrt{10}} \cdot \sqrt{(1 - \frac{10}{70})} = 21{,}2689.$$

$$\tilde{X} - \Delta < \overline{X} < \tilde{X} + \Delta;$$

$$98{,}2 - 21{,}2689 < \overline{X} < 98{,}2 + 21{,}2689;$$

$$76{,}9311 < \overline{X} < 119{,}4689.$$

Ответ. При условии, что отбор квартир был бесповторным, с вероятностью 0,95 можно ожидать, что средний расход электроэнергии на 1 квартиру во всем доме находится в интервале от 76,9311 до 119,4689 кВт-часа.

8. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

На разных стадиях статистического исследования и моделирования возникает необходимость в формулировке и экспериментальной проверке некоторых предположений (гипотез) относительно природы и величины неизвестных параметров анализируемой генеральной совокупности (совокупностей). Например, исследователь высказывает предположение: "выборка извлечена из нормальной генеральной совокупности" или "генеральная средняя анализируемой совокупности равна пяти". Такие предположения называются ***статистическими гипотезами.***

Сопоставление высказанной гипотезы относительно генеральной совокупности с имеющимися выборочными данными, сопровождаемое количественной оценкой степени достоверности получаемого вывода, осуществляется с помощью того или иного статистического критерия и называется ***проверкой статистических гипотез***.

Выдвинутая гипотеза называется ***нулевой (основной)***. Ее принято обозначать H_0.

По отношению к высказанной (основной) гипотезе всегда можно сформулировать ***альтернативную (конкурирующую)***, противоречащую ей. Альтернативную (конкурирующую) гипотезу принято обозначать H_1.

Цель статистической проверки гипотез состоит в том, чтобы на основании выборочных данных принять решение о справедливости основной гипотезы H_0.

Если выдвигаемая гипотеза сводится к утверждению о том, что значение некоторого неизвестного параметра генеральной совокупности *в точности равно* заданной величине, то эта гипотеза называется *простой*, например: "среднедушевой совокупный доход населения России составляет 650 рублей в месяц"; "уровень безработицы (доля безработных в численности экономически активного населения) в России равна 9%". В других случаях гипотеза называется *сложной*.

В качестве нулевой гипотезы H_0 принято выдвигать простую гипотезу, т.к. обычно бывает удобнее проверять более строгое утверждение.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов[5]:

- гипотезы о виде закона распределения исследуемой случайной величины;
- гипотезы о числовых значениях параметров исследуемой генеральной совокупности[6];
- гипотезы об однородности двух или нескольких выборок или некоторых характеристик анализируемых совокупностей;
- гипотезы об общем виде модели, описывающей статистическую зависимость между признаками и др.

Так как проверка статистических гипотез осуществляется на основании выборочных данных, т.е. ограниченного ряда наблюдений, решения относительно нулевой гипотезы H_0 имеют вероятностный характер. Другими словами, такое решение неизбежно сопровождается некоторой, хотя возможно и очень малой, вероятностью ошибочного заключения как в ту, так и в другую сторону.

Так, в какой-то небольшой доле случаев α нулевая гипотеза H_0 может оказаться отвергнутой, в то время как в действительности в генеральной совокупности она является справедливой. Такую ошибку называют ***ошибкой первого рода***. А ее вероятность принято называть ***уровнем значимости*** и обозначать α.

---

[5] В этой работе рассматриваются первые два типа гипотез.

[6] Эти гипотезы часто называют параметрическими, тогда как все остальные - непараметрическими.

Наоборот, в какой-то небольшой доле случаев β нулевая гипотеза H_0 принимается, в то время как на самом деле в генеральной совокупности она ошибочна, а справедлива альтернативная гипотеза H_1. Такую ошибку называют ***ошибкой второго рода***. Вероятность ошибки второго рода принято обозначать β. Вероятность $1 - \beta$ называют ***мощностью критерия***.

При фиксированном объеме выборки можно выбрать по своему усмотрению величину вероятности только одной из ошибок α или β. Увеличение вероятности одной из них приводит к снижению другой. Принято задавать вероятность ошибки первого рода α - уровень значимости. Как правило, пользуются некоторыми стандартными значениями уровня значимости α: 0,1; 0,05; 0,025; 0,01; 0,005; 0,001. Тогда, очевидно, из двух критериев, характеризующихся одной и той же вероятностью α отклонить правильную в действительности гипотезу H_0, следует принять тот, который сопровождается меньшей ошибкой второго рода β, т.е. большей мощностью. Снижения вероятностей обеих ошибок α и β можно добиться путем увеличения объема выборки.

Правильное решение относительно нулевой гипотезы H_0 также может быть двух видов:

- будет принята нулевая гипотеза H_0, тогда как и на самом деле в генеральной совокупности верна нулевая гипотеза H_0 ; вероятность такого решения $1 - \alpha$;

- нулевая гипотеза H_0 будет отклонена в пользу альтернативной H_1, тогда как и на самом деле в генеральной совокупности нулевая гипотеза H_0 отклоняется в пользу альтернативной H_1; вероятность такого решения $1 - \beta$ - *мощность критерия.*

Результаты решения относительно нулевой гипотезы можно проиллюстрировать с помощью таблицы 8.1.

Проверка статистических гипотез осуществляется с помощью *статистического критерия* (назовем его в общем виде K), являющего функцией от результатов наблюдения.

Статистический критерий - это правило (формула), по которому определяется мера расхождения результатов выборочного наблюдения с высказанной гипотезой H_0.

Таблица 8.1

Нулевая гипотеза H_0	Результаты решения относительно нулевой гипотезы H_0	
	отклонена	принята
верна	ошибка первого рода, ее вероятность $P(H_1/H_0) = \alpha$	правильное решение, его вероятность $P(H_0/H_0) = 1 - \alpha$
не верна	правильное решение, его вероятность $P(H_1/H_1) = 1 - \beta$	ошибка второго рода, ее вероятность $P(H_0/H_1) = \beta$

Статистический критерий, как и всякая функция от результатов наблюдения, является случайной величиной и в предположении справедливости нулевой гипотезы $\boldsymbol{H_0}$ подчинена некоторому хорошо изученному (и затабулированному) теоретическому закону распределения с плотностью распределения $f(k)$.

Выбор критерия для проверки статистических гипотез может быть осуществлен на основании различных принципов. Чаще всего для этого пользуются *принципом отношения правдоподобия*, который позволяет построить критерий наиболее мощный среди всех возможных критериев. Суть его сводится к выбору такого критерия K с известной функцией плотности $f(k)$ при условии справедливости гипотезы H_0, чтобы при заданном уровне значимости α можно было бы найти критическую точку $K_{кр}$. распределения $f(k)$, которая разделила бы область значений критерия на две части: область допустимых значений, в которой результаты выборочного наблюдения выглядят наиболее правдоподобными, и критическую область, в которой результаты выборочного наблюдения выглядят менее правдоподобными в отношении нулевой гипотезы H_0.

Если такой критерий K выбран и известна плотность его распределения, то задача проверки статистической гипотезы сводится к тому, чтобы при заданном уровне значимости α рассчитать по выборочным данным наблюдаемое значение критерия $K_{набл.}$ и определить, является ли оно наиболее или менее правдоподобным в отношении нулевой гипотезы H_0.

Проверка каждого типа статистических гипотез осуществляется с помощью соответствующего критерия, являющегося наиболее мощным в каждом конкретном случае. Например, проверка гипотезы о виде закона распределения случайной величины может быть осуществлена с помощью критерия согласия Пирсона χ^2; проверка гипотезы о равенстве неизвестных значений дисперсий двух генеральных совокупностей - с помощью критерия F - Фишера; ряд гипотез о неизвестных значениях параметров генеральных совокупностей проверяется с помощью критерия Z - нормальной распределенной случайной величины и критерия T- Стьюдента и т.д.

Значение критерия, рассчитываемое по специальным правилам на основании выборочных данных, называется *наблюдаемым значением критерия* ($K_{набл.}$).

Значения критерия, разделяющие совокупность значений критерия на *область допустимых значений* (наиболее правдоподобных в отношении нулевой гипотезы H_0) и *критическую область* (область значений, менее правдоподобных в отношении таблицам распределения случайной величины K, выбранной в качестве критерия), называются ***критическими точками($K_{кр.}$).***

Областью допустимых значений (областью принятия нулевой гипотезы H_0) называют совокупность значений критерия $\boldsymbol{K}$, при которых нулевая гипотеза $\boldsymbol{H_0}$ не отклоняется.

Критической областью называют совокупность значений критерия $\boldsymbol{K}$, при которых нулевая гипотеза $\boldsymbol{H_0}$ отклоняется в пользу конкурирующей $\boldsymbol{H_1}$.

Различают *одностороннюю* (правостороннюю или левостороннюю) и *двустороннюю критические области*.

Если конкурирующая гипотеза - правосторонняя, например, H_1: $a > a_0$, то и критическая область - *правосторонняя* (рис. 8.1). При правосторонней конкурирующей гипотезе критическая точка *($K_{кр.\ правосторонняя}$)* принимает положительные значения.

Если конкурирующая гипотеза - левосторонняя, например, H_1: $a < a_0$, то и критическая область - *левосторонняя* (рис. 8.2). При левосторонней конкурирующей гипотезе критическая точка принимает отрицательные значения *($K_{кр.\ левосторонняя}$)*.

Если конкурирующая гипотеза - двусторонняя, например, H_1: $a \neq a_0$, то и критическая область - *двусторонняя* (рис. 8.3). При двусторонней конкурирующей гипотезе определяются две критические точки *($K_{кр.\ левосторонняя}$* и *$K_{кр.\ правосторонняя}$)*.

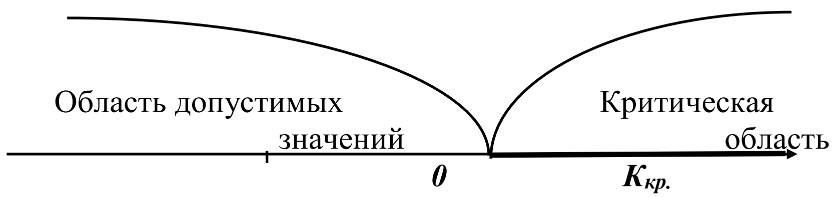


Рис. 8.1. Правосторонняя критическая область

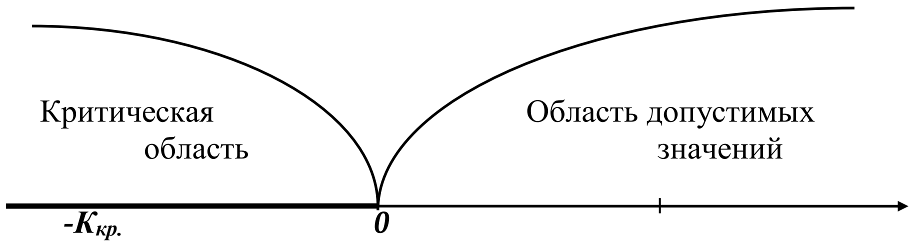


Рис. 8.2. Левосторонняя критическая область

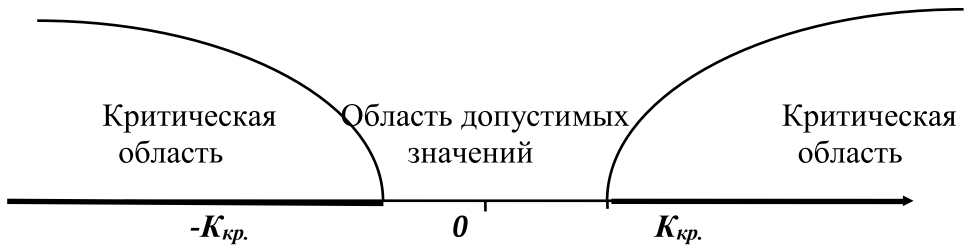


Рис. 8.3. Двусторонняя критическая область

Основной принцип проверки статистических гипотез состоит в следующем:

- *если наблюдаемое значение критерия ($K_{набл.}$) принадлежит критической области, то нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1;*

- *если наблюдаемое значение критерия ($K_{набл.}$) принадлежит области допустимых значений, то нулевую гипотезу H_0 нельзя отклонить.*

Можно принять решение относительно нулевой гипотезы H_0 путем сравнения наблюдаемого *($K_{набл.}$)* и критического значений критерия *($K_{кр.}$)*.

При правосторонней конкурирующей гипотезе:

Если $K_{набл.} \leq K_{кр.}$, то нулевую гипотезу H_0 *нельзя отклонить*;

если $K_{набл.} > K_{кр.}$, то нулевая гипотеза H_0 *отклоняется в пользу конкурирующей* H_1.

При левосторонней конкурирующей гипотезе:

Если $K_{набл.} \geq - K_{кр.}$, то нулевую гипотезу H_0 *нельзя отклонить*;

если $K_{набл.} < - K_{кр.}$, то нулевая гипотеза H_0 *отклоняется в пользу конкурирующей* H_1.

При двусторонней конкурирующей гипотезе:

Если $- K_{кр.} \leq K_{набл.} \leq K_{кр.}$, то нулевую гипотезу H_0 *нельзя отклонить*;

если $K_{набл.} > K_{кр.}$ *или* $K_{набл.} < - K_{кр.}$, то нулевая гипотеза H_0 *отклоняется в пользу конкурирующей* H_1.

Алгоритм проверки статистических гипотез сводится к следующему:

1. Сформулировать нулевую H_0 и альтернативную H_1 гипотезы;

2. Выбрать уровень значимости α;

3. В соответствии с видом выдвигаемой нулевой гипотезы H_0 выбрать статистический критерий для ее проверки, т.е. - специально подобранную случайную величину K, точное или приближенное распределение которой заранее известно;

4. По таблицам распределения случайной величины K, выбранной в качестве статистического критерия, найти его критическое значение $K_{кр.}$ (критическую точку или точки);

5. На основании выборочных данных по специальному алгоритму вычислить наблюдаемое значение критерия $K_{набл.}$;

6. По виду конкурирующей гипотезы H_1 определить тип критической области;

7. Определить, в какую область (допустимых значений или критическую) попадает наблюдаемое значение критерия $K_{набл.}$, и в зависимости от этого - принять решение относительно нулевой гипотезы H_0.

Следует заметить, что даже в том случае, если нулевую гипотезу H_0 нельзя отклонить, это не означает, что высказанное предположение о генеральной совокупности является единственно подходящим: просто ему не противоречат имеющиеся выборочные данные, однако таким же свойством наряду с высказанной могут обладать и другие гипотезы.

Можно интерпретировать результаты проверки нулевой гипотезы следующим образом:

- если в результате проверки нулевую гипотезу H_0 нельзя отклонить, то это означает, что имеющиеся выборочные данные не позволяют с достаточной уверенностью отклонить нулевую гипотезу H_0, вероятность нулевой гипотезы H_0 больше α, а конкурирующей H_1 - меньше $1 - \alpha$;

- если в результате проверки нулевая гипотеза H_0 отклоняется в пользу конкурирующей H_1, то это означает, что имеющиеся выборочные данные не позволяют с достаточной уверенностью принять нулевую гипотезу H_0, вероятность нулевой гипотезы H_0 меньше α, а конкурирующей H_1 - больше $1-\alpha$.

Пример 8.1. В семи случаях из десяти фирма-конкурент компании "А" действовала на рынке так, как будто ей заранее были известны решения, принимаемые фирмой "А". На уровне значимости 0,05 определите, случайно ли это, или в фирме "А" работает осведомитель фирмы-конкурента.

Решение. Для того чтобы ответить на вопрос данной задачи, необходимо проверить статистическую гипотезу о том, совпадает ли данное эмпирическое распределение числа действий фирмы-конкурента с равномерным теоретическим распределением.

Если ходы, предпринимаемые конкурентом, выбираются случайно, т.е. в фирме "А" - нет осведомителя (инсайдера), то число "правильных" и "неправильных" ее действий должно распределиться поровну, т.е. по 5 (10/2). А это и есть отличительная особенность равномерного распределения.

Этот вид статистических гипотез относится к гипотезам о виде закона распределения генеральной совокупности.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $X \sim R(a;\ b)$ - случайная величина X подчиняется равномерному распределению с параметрами (a; b) (в контексте задачи - "в фирме "А" - нет осведомителя (инсайдера)"; "распределение числа удачных ходов фирмы-конкурента - случайно").

H_1: Случайная величина X не подчиняется равномерному распределению (в контексте задачи - "в фирме "А" - есть осведомитель (инсайдер)"; "распределение числа удачных ходов фирмы-конкурента - не случайно").

В качестве критерия для проверки статистических гипотез о неизвестном законе распределения генеральной совокупности используется случайная величина χ^2. Этот критерий называют критерием Пирсона.

Его наблюдаемое значение ($\chi^2_{набл.}$) рассчитывается по формуле:

$$\chi^2_{набл.} = \sum_{i=1}^{n} \frac{(m_{(эмп.)i} - m_{(теор.)i})^2}{m_{(теор.)i}}, \qquad (8.1)$$

где $m_{(эмп.)i}$ - эмпирическая частота i-той группы выборки;

$m_{(теор.)i}$ - теоретическая частота i-той группы выборки.

Составим таблицу распределения эмпирических и теоретических частот:

$m_{(эмп.)i}$	7	3
$m_{(теор.)i}$	5	5

Найдем наблюдаемое значение $\chi^2_{набл.}$:

$$\chi^2_{набл.} = \frac{(7-5)^2}{5} + \frac{(3-5)^2}{5} = \frac{4}{5} + \frac{4}{5} = 1{,}6.$$

Критическое значение ($\chi^2_{кр.}$) следует определять по таблице распределения χ^2 (см. приложение 4) по уровню значимости α и числу степеней свободы k.

По условию α = 0,05, а число степеней свободы рассчитывается по формуле:

$k = n - l - 1,$

где k - число степеней свободы;

n - число групп выборки;

l - число неизвестных параметров предполагаемой модели, оцениваемых по данным выборки (если все параметры предполагаемого закона известны точно, то $l = 0$).

По условию задачи число групп выборки (n) равно 2, т.к. могут быть только два варианта действий фирмы-конкурента: "удачные" и "неудачные", а число неизвестных параметров равномерного распределения (l) равно 0.

Отсюда k = 2 - 0 - 1 = 1.

Найдем $\chi^2_{кр.}$ по уровню значимости α = 0,05 и числу степеней свободы k=1.

$$\chi^2_{кр.(\alpha=0,05;k=1)} = 3{,}8.$$

$\chi^2_{набл.} < \chi^2_{кр.}$, следовательно, на данном уровне значимости нулевую гипотезу нельзя отклонить, расхождения эмпирических и теоретических частот - незначимые. Данные наблюдений согласуются с гипотезой о равномерном распределении генеральной совокупности.

Это означает, что для утверждения о том, что действия фирмы-конкурента на рынке неслучайны; на уровне значимости α = 0,05 можно утверждать, что в фирме "А" нет платного осведомителя фирмы-конкурента.

Ответ. На уровне значимости α = 0,05 можно утверждать, что в фирме "А" нет платного осведомителя фирмы-конкурента.

Пример 8.2. На уровне значимости α = 0,025 проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

$m_{(эмп.)i}$	5	10	20	25	14	3
$m_{(теор.)i}$	6	14	28	18	8	3

Решение. Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: X~N(a; σ^2) - случайная величина X подчиняется нормальному закону распределения с параметрами a и σ^2.

H_1: случайная величина X не подчиняется нормальному закону распределения с параметрами a и σ^2.

В качестве критерия для проверки нулевой гипотезы используем критерий Пирсона χ^2.

Найдем наблюдаемое значение ($\chi^2_{набл.}$):

$$\chi^2_{набл.} = \frac{(5-6)^2}{6} + \frac{(10-14)^2}{14} + \frac{(20-28)^2}{28} + \frac{(25-18)^2}{18} + \frac{(14-8)^2}{8} + \frac{(3-3)^2}{3} \approx 10{,}8175.$$

Найдем критическое значение критерия ($\chi^2_{кр.}$) по таблице распределения χ^2 (см. приложение 4) по уровню значимости α и числу степеней свободы k.

По условию $\alpha = 0{,}025$; число степеней свободы найдем по формуле:

$k = n - l - 1$,

где k - число степеней свободы;

n - число групп выборки;

l - число неизвестных параметров предполагаемой модели, оцениваемых по данным выборки.

По условию задачи число групп выборки (n) равно 6, а число неизвестных параметров нормального распределения (l) равно 2.

Отсюда, k = 6 - 2 - 1 = 3.

Найдем $\chi^2_{кр.}$ по уровню значимости $\alpha = 0{,}025$ и числу степеней свободы k=3.

$\chi^2_{кр.(\alpha=0{,}025;k=3)} = 9{,}4.$

$\chi^2_{набл.} > \chi^2_{кр.}$, следовательно, на данном уровне значимости нулевая гипотеза отвергается в пользу конкурирующей, расхождения эмпирических и теоретических частот - значимые. Данные наблюдений не согласуются с гипотезой о нормальном распределении генеральной совокупности.

Ответ. На уровне значимости $\alpha = 0{,}025$ данные наблюдений не согласуются с гипотезой о нормальном распределении генеральной совокупности.

Пример 8.3. Техническая норма предусматривает в среднем 40 сек. на выполнение определенной технологической операции на конвейере по производству часов. От работниц, работающих на этой операции, поступили жалобы, что они в действительности затрачивают на эту операцию больше времени. Для проверки данной жалобы произведены хронометрические измерения времени выполнения этой технологической операции у 16 работниц, занятых на этой операции, и получено среднее время выполнения операции $\tilde{X}$ = 42 сек. Можно ли по имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, если:

а) исправленное выборочное среднее квадратическое отклонение s составило 3,5 сек.;

б) выборочное среднее квадратическое отклонение σ составило 3,5 сек.?

Решение.

а) Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная средняя нормальной совокупности точно равна определенному числу, когда *дисперсия генеральной совокупности неизвестна* (выборка мала, т.к. n = 16, меньше 30).

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $a = a_0 = 40$ - неизвестное математическое ожидание a (нормально распределенной генеральной совокупности с неизвестной дисперсией) равно гипотетическому предполагаемому числовому значению a_0 (применительно к условию данной задачи - время выполнения технологической операции соответствует норме).

H_1: $a > 40$ - неизвестное математическое ожидание a (нормально распределенной генеральной совокупности с неизвестной дисперсией) больше числовому значению a_0 (применительно к условию данной задачи - время выполнения технологической операции больше установленной нормы).

Так как конкурирующая гипотеза - правосторонняя, то и критическая область - правосторонняя.

В качестве критерия для сравнения неизвестного математического ожидания a (нормально распределенной генеральной совокупности с неизвестной дисперсией) с гипотетическим числовым значением a_0 используется случайная величина t - критерий Стьюдента:

Его наблюдаемое значение ($t_{набл.}$) рассчитывается по формуле:

$$t_{набл.} = \frac{\tilde{X} - a_0}{s}\sqrt{n}\,. \tag{8.2}$$

где $\tilde{X}$ - выборочная средняя;

a_0 - числовое значение генеральной средней;

s - исправленное среднее квадратическое отклонение;

n - объем выборки.

Найдем наблюдаемое значение $t_{набл.}$:

$$t_{набл.} = \frac{42-40}{3{,}5}\cdot\sqrt{16-1} \approx 2{,}2131.$$

Критическое значение ($t_{кр.}$) следует находить по таблице распределения Стьюдента (см. приложение 5) по уровню значимости α и числу степеней свободы k.

По условию $\alpha = 0{,}01$; число степеней свободы найдем по формуле:

$k = n - 1$,

где k - число степеней свободы;

n - объем выборки;

$k = 16 - 1 = 15$.

Найдем $t_{кр.}$ по уровню значимости $\alpha = 0{,}01$ (для односторонней критической области) и числу степеней свободы $k = 15$:

$$t_{кр.(\alpha=0{,}01;k=15)} = 2{,}6.$$

Заметим, что при левосторонней конкурирующей гипотезе H_1: $a < 40$ $t_{кр.}$ следует находить по таблицам распределения Стьюдента (см. приложение 5) по уровню значимости α (для односторонней критической области) и числу степеней свободы $k = n - 1$ и присваивать ему "минус".

При двусторонней конкурирующей гипотезе H_1: $a \neq 40$ $t_{кр.}$ следует находить по таблицам распределения Стьюдента (см. приложение 5) по уровню значимости α (для двусторонней критической области) и числу степеней свободы $k = n - 1$).

$t_{набл.} < t_{кр}$, следовательно, на данном уровне значимости нет оснований отклонить нулевую гипотезу.

Ответ. По имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ нельзя отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме. Следовательно, жалобы работниц - необоснованны.

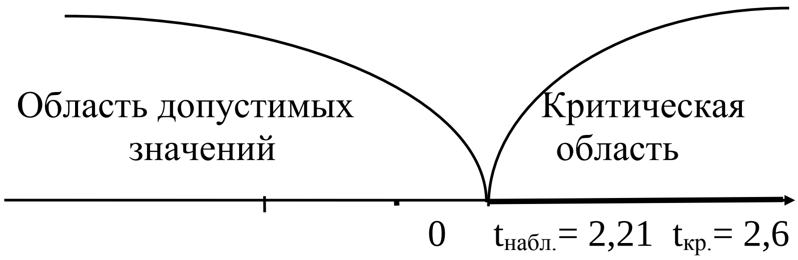


Рис. 8.4

Наблюдаемое значение критерия попадает в область допустимых значений, следовательно, нет оснований отклонить нулевую гипотезу.

б) Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная средняя нормальной совокупности точно равна определенному числу, когда *дисперсия генеральной совокупности неизвестна.*

Алгоритм решения задачи будет тот же, что и в первом случае. Однако наблюдаемое значение $t_{набл.}$ будет рассчитывается по формуле:

$$t_{набл.} = \frac{\tilde{X} - a_0}{\sigma_{выб.}} \sqrt{n}\,. \tag{8.3}$$

где $\tilde{X}$ - выборочная средняя;

a_0 - числовое значение генеральной средней;

$\sigma_{выб.}$ - выборочное среднее квадратическое отклонение;

n - объем выборки.

Найдем наблюдаемое значение ($t_{набл.}$):

$$t_{набл.} = \frac{42-40}{3{,}5} \cdot \sqrt{16} \approx 2{,}2857.$$

Критическое значение ($t_{кр.}$) следует находить по таблице распределения Стьюдента (см. приложение 5) по уровню значимости α и числу степеней свободы k.

$$t_{кр.(\alpha=0{,}01;k=15)} = 2{,}6.$$

$t_{набл.} < t_{кр}$, следовательно, на данном уровне значимости нет оснований отвергнуть нулевую гипотезу, жалобы работниц - необоснованны.

Ответ. По имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ нельзя отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, жалобы работниц - необоснованны.

Пример 8.4. Изменим условие предыдущей задачи. Техническая норма предусматривает в среднем 40 сек. на выполнение определенной технологической операции на конвейере по производству часов. От работниц, работающих на этой операции, поступили жалобы, что они в действительности затрачивают на эту операцию больше времени. Для проверки данной жалобы произведены хронометрические измерения времени выполнения этой технологической операции у 36 работниц, занятых на этой операции, и получено среднее время выполнения операции $\tilde{X} =$ 42 сек. Можно ли (предполагая время выполнения технологической операции случайной величиной, подчиняющейся нормальному закону) по имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, если известно, что среднее квадратическое отклонение генеральной совокупности σ составляет 3,5 сек.?

Решение. Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная средняя нормальной совокупности точно равна числовому значению, когда *дисперсия генеральной совокупности известна* (большая выборка, т.к. n = 36, больше 30).

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $a = a_0 = 40$ - неизвестная генеральная средняя нормально распределенной совокупности с известной дисперсией равна числовому значению (применительно к условию данной задачи - время выполнения технологической операции соответствует норме).

H_1: $a > 40$ - неизвестная генеральная средняя нормально распределенной совокупности с известной дисперсией больше числового значения (применительно к условию данной задачи - время выполнения технологической операции больше установленной нормы).

Так как конкурирующая гипотеза - правосторонняя, то и критическая область - правосторонняя.

В качестве критерия для сравнения выборочной средней с гипотетической генеральной средней нормальной совокупности, когда дисперсия генеральной совокупности известна, используется случайная величина U:

Его наблюдаемое значение ($u_{набл.}$) рассчитывается по формуле:

$$u_{набл.} = \frac{\tilde{X} - a_0}{\sigma_{ген.}} \sqrt{n} \,. \tag{8.4}$$

где $\tilde{X}$ - выборочная средняя;

a_0 - числовое значение генеральной средней;

$\sigma_{ген.}$ - выборочное среднее квадратическое отклонение;

n - объем выборки.

Найдем наблюдаемое значение ($u_{набл.}$):

$$u_{набл.} = \frac{42-40}{3{,}5} \cdot \sqrt{36} \approx 3{,}4286.$$

Так как конкурирующая гипотеза - правосторонняя, критическое значение $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства:

$\Phi_0(u_{кр}) = (1 - 2\alpha) / 2$.

По условию $\alpha = 0{,}01$.

Отсюда:

$\Phi_0(u_{кр}) = (1 - 2\cdot 0,01) / 2 = 0,49$.

По таблице функции Лапласа (см. приложение 2) найдем, при каком $u_{кр.}$ $\Phi_0(u_{кр}) = 0,49$.

$\Phi_0(2,33) = 0,49$.

Следовательно: $u_{кр.} = 2,33$.

Заметим, что при левосторонней конкурирующей гипотезе H_1: $a < 40$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = (1 - 2\alpha) / 2$ и присваивать ему "минус".

При двусторонней конкурирующей гипотезе H_1: $a \neq 40$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = (1 - \alpha) / 2)$.

$u_{набл.} > u_{кр}$, следовательно, на данном уровне значимости нулевая гипотеза отвергается в пользу конкурирующей. По имеющимся хронометрическим данным с более чем 99%-ной надежностью можно утверждать, что среднее время выполнения этой операции превышает норму. Следовательно, жалобы работниц - обоснованны.

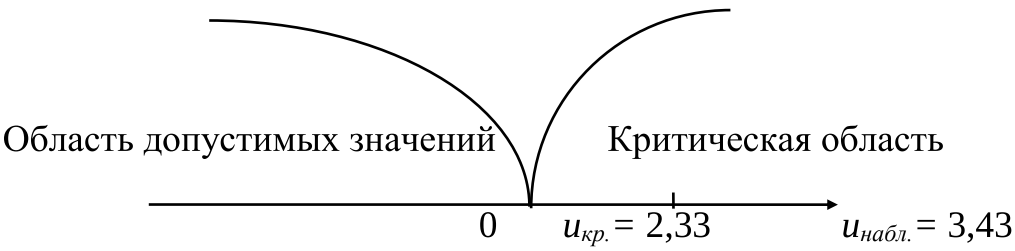


Рис. 8.5

Наблюдаемое значение критерия попадает в критическую область, следовательно, нулевая гипотеза отвергается в пользу конкурирующей.

Ответ. По имеющимся хронометрическим данным на уровне значимости $\alpha = 0,01$ можно утверждать, что среднее время выполнения этой операции превышает норму, жалобы работниц - обоснованны.

Пример 8.5. Экономический анализ производительности труда предприятий отрасли позволил выдвинуть гипотезу о наличии двух типов предприятий с различной средней величиной показателя производительности труда. Выборочное обследование 42 предприятий первой группы дало следующие результаты: средняя производительность труда $\tilde{X}$ составила 119 деталей. По данным выборочного обследования 35 предприятий второй группы средняя производительность труда $\tilde{Y}$ составила 107 деталей. Генеральные дисперсии известны: $D(X) = 126,91$ (дет.^{2}); $D(Y) = 136,1$ (дет.^{2}). Считая, что выборки извлечены из нормально распределенных генеральных совокупностей X и Y, на уровне значимости 0,05 проверьте, случайно ли полученное различие средних показателей производительности труда в группах или же имеются два типа предприятий с различной средней величиной производительности труда.

Решение. Для решения данной задачи необходимо сравнить две средние нормально распределенные генеральные совокупности, *генеральные дисперсии которых известны* (большие независимые выборки). В данной задаче речь идет о больших выборках, так как $n_x = 42$ и $n_y = 35$ больше 30. Выборки - независимые, так как из контекста задачи видно, что они извлечены из непересекающихся генеральных совокупностей.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $\overline{X} = \overline{Y}$ - генеральные средние двух нормально распределенных совокупностей с известными дисперсиями равны (применительно к условию данной задачи - предприятия двух групп относятся к одному типу предприятий, - средняя производительность труда в двух группах - одинакова).

H_1: $\overline{X} \neq \overline{Y}$ - генеральные средние двух нормально распределенных совокупностей с известными дисперсиями не равны (применительно к условию данной задачи - предприятия двух групп относятся к разному типу предприятий, - средняя производительность труда в двух группах - неодинакова).

Выдвигаем двустороннюю конкурирующую гипотезу, так как из условия задачи не следует, что необходимо выяснить, больше или меньше производительность труда в одной из групп предприятий по сравнению с другой.

Так как конкурирующая гипотеза - двусторонняя, то и критическая область - двусторонняя.

В качестве критерия для сравнения двух средних генеральных совокупностей, дисперсии которых известны (большие независимые выборки), используется случайная величина Z.

Его наблюдаемое значение ($z_{набл.}$) рассчитывается по формуле:

$$z_{набл.} = \frac{\tilde{X} - \tilde{Y}}{\sqrt{\frac{D(X)}{n_x} + \frac{D(Y)}{n_y}}}, \tag{8.5}$$

где $\tilde{X}$ - выборочная средняя для X;

$\tilde{Y}$ - выборочная средняя для Y;

D(X) - генеральная дисперсия для X;

D(Y) - генеральная дисперсия для Y;

n_x - объем выборки для X;

n_y - объем выборки для Y.

Найдем наблюдаемое значение ($z_{набл.}$):

$$z_{набл.} = \frac{119 - 107}{\sqrt{\frac{129{,}91}{42} + \frac{136{,}1}{35}}} \approx 4{,}5649.$$

Так как конкурирующая гипотеза - двусторонняя, критическое значение ($z_{кр.}$) следует находить по таблице функции Лапласа (см. приложение 2) из равенства:

$\Phi_0(z_{кр}) = (1 - \alpha) / 2.$

По условию α = 0,05.

Отсюда:

$\Phi_0(z_{кр}) = (1 - 0,05) / 2 = 0,475$.

По таблице функции Лапласа (см. приложение 2) найдем, при каком $z_{кр.}$ $\Phi_0(z_{кр}) = 0,475$.

$\Phi_0(1,96) = 0,475$.

Учитывая, что конкурирующая гипотеза - двусторонняя, находим две критические точки:

$z_{кр.(прав.)} = 1,96$; $z_{кр.(лев.)} = -1,96$.

Заметим, что при левосторонней конкурирующей гипотезе H_1: $\overline{X} < \overline{Y}$ $z_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(z_{кр}) = (1 - 2\alpha) / 2$ и присваивать ему "минус".

При правосторонней конкурирующей гипотезе H_1: $\overline{X} > \overline{Y}$ $z_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(z_{кр})=(1- -2\alpha) / 2)$.

$z_{набл.} > z_{кр}$, следовательно, на данном уровне значимости нулевая гипотеза отвергается в пользу конкурирующей. На уровне значимости α = 0,05 можно утверждать, что полученное различие средних показателей производительности труда в группах - неслучайно, имеются два типа предприятий с различной средней величиной производительности труда.

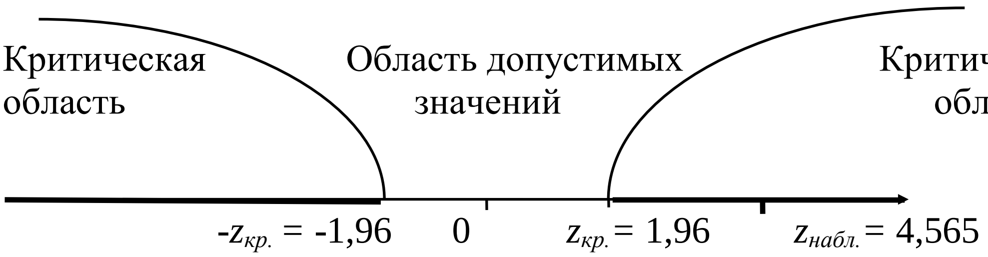


Рис. 8.6

Наблюдаемое значение критерия попадает в критическую область, следовательно, нулевая гипотеза отвергается в пользу конкурирующей.

Ответ. На уровне значимости α = 0,05 можно утверждать, что полученное различие средних показателей производительности труда в группах - неслучайно, имеются два типа предприятий с различной средней величиной производительности труда.

Пример 8.6. Предполагается, что применение нового типа резца сократит время обработки некоторой детали. Хронометраж времени обработки 9 деталей, обработанных старым типом резцов, дал следующие результаты: среднее время обработки детали $\tilde{X}$ составило 57 мин., исправленная выборочная дисперсия s_x^2 = 186,2 (мин.^{2}). Среднее время обработки 15 деталей, обработанных новым типом резца, $\tilde{Y}$ по данным хронометражных измерений составило 52 мин., а исправленная выборочная дисперсия s_y^2 = 166,4 (мин.^{2}). На уровне значимости α = 0,01 ответьте на вопрос, позволило ли использование нового типа резцов сократить время обработки детали.

Решение. Для решения данной задачи необходимо сравнить две средние нормально распределенные генеральные совокупности, *генеральные дисперсии которых неизвестны,* но предполагаются одинаковыми (малые независимые выборки). В данной задаче речь идет о малых выборках, так как $n_x = 9$ и $n_y = 15$ меньше 30. Выборки - независимые, так как из контекста задачи видно, что они извлечены из непересекающихся генеральных совокупностей.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $\overline{X} = \overline{Y}$ - генеральные средние двух нормально распределенных совокупностей с неизвестными дисперсиями (но предполагаемыми одинаковыми) равны (применительно к условию данной задачи - среднее время, затрачиваемое на обработку детали резцами нового и старого типа - одинаково, т.е. использование нового типа резца не позволяет снизить время на обработку детали).

H_1: $\overline{X} > \overline{Y}$ - генеральная средняя для X больше, чем генеральная средняя для Y (применительно к условию данной задачи - среднее время, затрачиваемое на обработку детали резцами старого типа, больше, чем - нового, т.е. использование нового типа резца позволяет снизить время на обработку детали).

Так как конкурирующая гипотеза - правосторонняя, то и критическая область - правосторонняя.

Приступать к проверке гипотезы о равенстве генеральных средних двух нормально распределенных совокупностей с неизвестными дисперсиями можно лишь в том случае, если генеральные дисперсии равны. В противном случае данная задача в теории неразрешима.

Поэтому, прежде чем проверять эту гипотезу, проверим гипотезу о равенстве генеральных дисперсий нормальных совокупностей.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: D(X) = D(Y) - генеральные дисперсии двух нормально распределенных совокупностей равны.

H_1: D(X) > D(Y) - генеральная дисперсия для X больше генеральной дисперсии для Y. Выдвигаем правостороннюю конкурирующую гипотезу, так как исправленная выборочная дисперсия для X значительно больше, чем исправленная выборочная дисперсия для Y.

Так как конкурирующая гипотеза - правосторонняя, то и критическая область - правосторонняя.

В качестве критерия для сравнения двух дисперсий нормальных генеральных совокупностей используется случайная величина F - критерий Фишера-Снедекора.

Его наблюдаемое значение ($f_{набл.}$) рассчитывается по формуле:

$$f_{набл.} = \frac{s^2_{б.}}{s^2_{м.}}, \tag{8.6}$$

где $s^2_б$ - большая (по величине) исправленная выборочная дисперсия;

$s_м^2$ - меньшая (по величине) исправленная выборочная дисперсия.

Найдем $f_{набл.}$:

$$f_{набл.} = \frac{186,2}{166,4} \approx 1,119.$$

Критическое значение ($f_{кр.}$) следует находить по таблице распределения Фишера-Снедекора (приложение 6) по уровню значимости α и числу степеней свободы k_1 и k_2.

По условию $\alpha = 0,01$; число степеней свободы найдем по формуле:

$k_1 = n_1 - 1$; $k_2 = n_2 - 1$,

где k_1 - число степеней свободы большей (по величине) исправленной дисперсии;

k_2 - число степеней свободы меньшей (по величине) исправленной дисперсии;

n_1 - объем выборки большей (по величине) исправленной дисперсии;

n_2 - объем выборки меньшей (по величине) исправленной дисперсии.

Найдем k_1 и k_2:

$k_1 = 10 - 1 = 9$;

$k_2 = 15 - 1 = 14$.

Определяем $f_{кр.}$ по уровню значимости $\alpha = 0,01$ и числу степеней свободы $k_1=9$ и $k_2=14$:

$$f_{кр.(\alpha=0,01;k_1=8;k_2=14)} = 4,14.$$

$f_{набл.} < f_{кр}$, следовательно, на данном уровне значимости нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий нормальных совокупностей.

Следовательно, можно приступить к проверке гипотезы о равенстве генеральных средних двух нормально распределенных совокупностей.

В качестве критерия для проверки этой гипотезы используется случайная величина t - критерий Стьюдента:

Его наблюдаемое значение ($t_{набл.}$) рассчитывается по формуле:

$$t_{набл.} = \frac{\tilde{X} - \tilde{Y}}{\sqrt{(n_x - 1)\cdot s_x^2 + (n_y - 1)\cdot s_y^2}} \cdot \sqrt{\frac{n_x \cdot n_y \cdot (n_x + n_y - 2)}{n_x + n_y}}, \qquad (8.7)$$

где $\tilde{X}$ - выборочная средняя для X;

$\tilde{Y}$ - выборочная средняя для Y;

D(X) - генеральная дисперсия для X;

D(Y) - генеральная дисперсия для Y;

n_x - объем выборки для X;

n_y - объем выборки для Y.

Найдем $t_{набл.}$:

$$t_{набл.} = \frac{57 - 52}{\sqrt{(9-1)\cdot 186,2 + (15-1)\cdot 166,4}} \cdot \sqrt{\frac{9\cdot 15\cdot (9+15-2)}{9+15}} \approx 0,9.$$

Критическое значение ($t_{кр.}$) следует находить по таблице распределения Стьюдента (см. приложение 5) по уровню значимости α и числу степеней свободы k.

По условию $\alpha = 0{,}01$; число степеней свободы найдем по формуле:

$$k = n_x + n_y - 2,$$

где k - число степеней свободы;

n_x - объем выборки для X;

n_y - объем выборки для Y;

$k = 9 + 15 - 2 = 22.$

Найдем $t_{кр.}$ по уровню значимости $\alpha = 0{,}01$ (для односторонней критической области) и числу степеней свободы $k = 22$:

$$t_{кр..(\alpha=0,01;k=22)} = 2{,}51.$$

Заметим, что при левосторонней конкурирующей гипотезе $\overline{X} < \overline{Y}$ $t_{кр.}$ следует находить по таблицам распределения Стьюдента (см. приложение 5) по уровню значимости α (для односторонней критической области) и числу степеней свободы $k = n_x + n_y - 2$ и присваивать ему "минус".

При двусторонней конкурирующей гипотезе $\overline{X} \neq \overline{Y}$ $t_{кр}$ следует находить по таблицам распределения Стьюдента (см. приложение 5) по уровню значимости α (для двусторонней критической области) и числу степеней свободы $k = n_x + n_y - 2$).

$t_{набл.} < t_{кр}$, следовательно, на данном уровне значимости нет оснований отвергнуть нулевую гипотезу. По имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ нельзя отклонить гипотезу о том, что генеральные средние равны, т.е. среднее время, затрачиваемое на обработку детали старым и новым типом резцов, отличается незначимо, расхождения между средними - случайны, использование нового типа резцов не позволяет снизить время обработки детали.

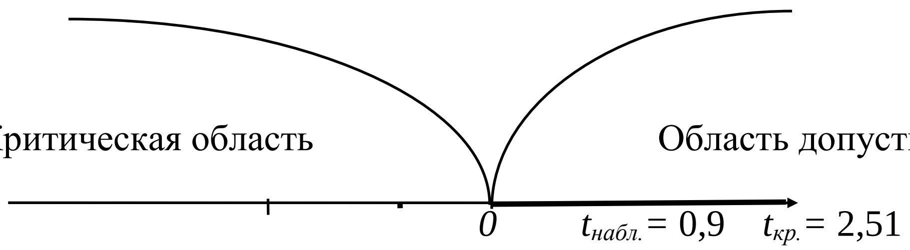


Рис. 8.7

Наблюдаемое значение критерия попадает в область допустимых значений, следовательно, нулевую гипотезу нельзя отвергнуть.

Ответ. На уровне значимости $\alpha = 0{,}01$ нельзя утверждать, что использование нового типа резцов позволило сократить время обработки детали.

Пример 8.7. Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, составляет не менее 0,97. Среди случайно отобранных 200 изделий проверяемой партии оказалось 193 соответствующих стандарту. Можно ли на уровне значимости $\alpha = 0{,}02$ принять партию?

Решение. Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная доля точно равна определенному числу.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $p = p_0 = 0,97$ - неизвестная генеральная доля p равна p_0 (применительно к условию данной задачи - вероятность того, что деталь из проверяемой партии окажется соответствующей стандарту, равна 0,97; то есть партию изделий можно принять).

H_1: $p < 0,97$ - неизвестная вероятность p меньше гипотетической вероятности p_0 (применительно к условию данной задачи - вероятность того, что деталь из проверяемой партии окажется соответствующей стандарту, меньше 0,97; то есть партию изделий нельзя принять).

Так как конкурирующая гипотеза - левосторонняя, то и критическая область - левосторонняя.

В качестве критерия для сравнения наблюдаемой относительной частоты с гипотетической вероятностью появления события используется случайная величина U:

Его наблюдаемое значение ($u_{набл.}$) рассчитывается по формуле:

$$u_{набл.} = \frac{\frac{m}{n} - p_0}{\sqrt{p_0 \cdot q_0}} \cdot \sqrt{n}\,, \tag{8.8}$$

где m / n - относительная частота (частость) появления события;
p_0 - гипотетическая вероятность появления события;
q_0 - гипотетическая вероятность непоявления события;
n - объем выборки.

По условию: m = 193; n = 200; $p_0 = 0,97$; $q_0 = 1 - p_0 = 0,03$; $\alpha = 0,02$.

Найдем наблюдаемое значение ($u_{набл.}$):

$$u_{набл.} = \frac{\frac{193}{200} - 0,97}{\sqrt{0,97 \cdot 0,03}} \cdot \sqrt{200} \approx -0,2771.$$

Так как конкурирующая гипотеза - левосторонняя, то критическое значение ($u_{кр.}$) следует находить по таблице функции Лапласа (см. приложение 2) из равенства:

$\Phi_0(u_{кр}) = (1 - 2\alpha) / 2$.

По условию $\alpha = 0,02$.

Отсюда:

$\Phi_0(u_{кр}) = (1 - 2 \cdot 0,02) / 2 = 0,48$.

По таблице функции Лапласа (см. приложение 2) найдем, при каком $u_{кр.}$ $\Phi_0(u_{кр}) = 0,48$.

$\Phi_0(2,05) = 0,48$.

Учитывая, что конкурирующая гипотеза - левосторонняя, критическому значению необходимо присвоить знак "минус".

Следовательно: $u_{кр.} = -2,05$.

Заметим, что при правосторонней конкурирующей гипотезе H_1: $p > 0{,}97$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = (1 - 2\alpha) / 2$.

При двусторонней конкурирующей гипотезе H_1: $p \neq 0{,}97$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = (1 - \alpha) / 2$).

$u_{набл.} > u_{кр}$, следовательно, на данном уровне значимости нет оснований отвергнуть нулевую гипотезу. По имеющимся данным на уровне значимости $\alpha=0{,}02$ нельзя отклонить гипотезу о том, что вероятность того, что изделие окажется соответствующим стандарту, составляет 0,97. Следовательно, партию изделий принять можно.

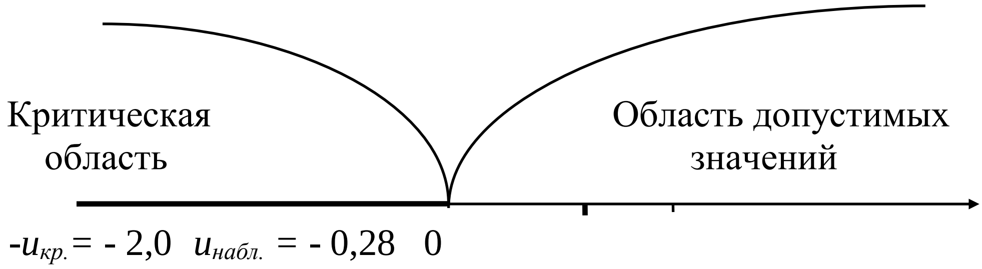


Рис. 8.8

Наблюдаемое значение критерия попадает в область допустимых значений, следовательно, нет оснований отклонить нулевую гипотезу.

Ответ. На уровне значимости $\alpha = 0{,}02$ партию изделий принять можно.

Пример 8.8. Два завода изготавливают однотипные детали. Для оценки их качества извлечены выборки из продукции этих заводов и получены следующие результаты:

	Завод №1	Завод №2
Объем выборки	n_1	n_2
Число бракованных деталей	m_1	m_2

На уровне значимости $\alpha = 0{,}025$ определите, имеется ли существенное различие в качестве изготавливаемых этими заводами деталей.

Решение. Для решения данной задачи необходимо сравнить две вероятности биномиальных распределений.

Сформулируем нулевую и конкурирующую гипотезы согласно условию задачи.

H_0: $p_1 = p_2$ - вероятности появления события в двух генеральных совокупностях, имеющих биномиальное распределение, равны (применительно к условию данной задачи - вероятность того, что деталь изготовленная на первом заводе, окажется бракованной, равна вероятности того, что деталь изготовленная на втором заводе, окажется бракованной).

H_1: $p_1 \neq p_2$ - вероятности появления события в двух генеральных совокупностях, имеющих биномиальное распределение, не равны (применительно к условию данной задачи - вероятность того, что деталь, изготовленная на первом заводе, окажется бракованной, не равна вероятности того, что деталь, изготовленная на втором заводе, окажется бракованной; заводы изготавливают детали разного качества). Так как по условию задачи не требуется проверить, на каком заводе качество изготавливаемых деталей выше, выдвигаем двустороннюю конкурирующую гипотезу.

Так как конкурирующая гипотеза - двусторонняя, то и критическая область - двусторонняя.

В качестве критерия для сравнения двух вероятностей биномиальных распределений используется случайная величина U:

Его наблюдаемое значение $u_{набл.}$ рассчитывается по формуле:

$$u_{набл.} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{\bar{p} \cdot \bar{q}}{n_1} + \frac{\bar{p} \cdot \bar{q}}{n_2}}}, \qquad (8.9)$$

где m_1 / n_1 - относительная частота (частость) появления события в первой выборке;

m_2 / n_2 - относительная частота (частость) появления события во второй выборке;

$\bar{p}$ - средняя частость появления события; $\bar{p} = \frac{m_1 + m_2}{n_1 + n_2}$;

$\bar{q}$ - средняя частость непоявления события; $\bar{q} = 1 - \bar{p}$;

n_1 - объем первой выборки;

n_2 - объем второй выборки.

По условию: $m_1 = 20$; $n_1 = 200$; $m_2 = 15$; $n_2 = 300$; $\alpha = 0,025$.

Найдем $\bar{p}$ - среднюю частость появления события:

$$\bar{p} = \frac{20 + 15}{200 + 300} = 0,07.$$

Найдем $\bar{q}$ - среднюю частость непоявления события:

$\bar{q} = 1 - \bar{p} = 1 - 0,07 = 0,93$.

Найдем $u_{набл.}$:

$$u_{набл.} = \frac{\frac{20}{200} - \frac{15}{300}}{\sqrt{\frac{0,07 \cdot 0,93}{200} + \frac{0,07 \cdot 0,93}{300}}} \approx 2,1467.$$

Так как конкурирующая гипотеза - двусторонняя, критическое значение ($u_{кр.}$) следует находить по таблице функции Лапласа (см. приложение 2) из равенства:

$\Phi_0(u_{кр}) = (1 - \alpha) / 2$.

По условию $\alpha = 0,025$.

Отсюда:

$\Phi_0(u_{кр}) = (1 - 0,025) / 2 = 0,4875$.

По таблице функции Лапласа (см. приложение 2) найдем, при каком $u_{кр.}$ $\Phi_0(u_{кр}) = 0,4875$.

$\Phi_0(2,24) = 0,4875$.

Учитывая, что конкурирующая гипотеза - двусторонняя, находим две критические точки:

$u_{кр.(прав.)} = 2,24$; $u_{кр.(лев.)} = -2,24$.

Заметим, что при правосторонней конкурирующей гипотезе H_1: $p_1 > p_2$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = (1 - 2\alpha) / 2$.

При левосторонней конкурирующей гипотезе H_1: $p_1 < p_2$ $u_{кр.}$ следует находить по таблице функции Лапласа (см. приложение 2) из равенства $\Phi_0(u_{кр}) = =(1 - 2\alpha) / 2$ и присваивать ему знак "минус").

$-u_{кр.} < u_{набл.} < u_{кр}$, следовательно, на данном уровне значимости нет оснований отвергнуть нулевую гипотезу. По имеющимся данным на уровне значимости $\alpha = 0,025$ нет оснований отклонить нулевую гипотезу. Следовательно, заводы изготавливают детали одинакового качества.

Наблюдаемое значение критерия попадает в область допустимых значений, следовательно, нет оснований отклонить нулевую гипотезу.

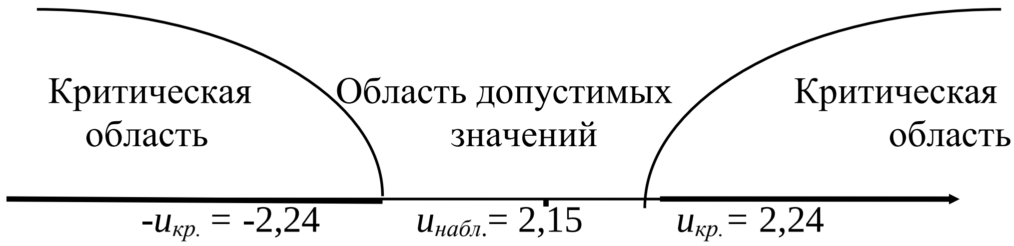


Рис. 8.9

Ответ. Нет оснований отклонить нулевую гипотезу, то есть имеющееся различие в качестве изготавливаемых этими заводами деталей - случайно, незначимо.

9. ПРАВИЛА ВЫПОЛНЕНИЯ И ОФОРМЛЕНИЯ ДОМАШНЕГО ЗАДАНИЯ (КОНТРОЛЬНОЙ РАБОТЫ)

При выполнении домашнего задания (контрольной работы) следует ***строго придерживаться*** следующих правил:

1. Работу следует выполнять ***в отдельной тетради*** чернилами синего или черного цвета, оставляя поля для замечаний.
2. Перед решением каждой задачи надо ***полностью выписывать*** ее условие.
3. Решать задачи необходимо по порядку. Решение задач нужно ***излагать подробно и аккуратно, объясняя все действия и указывая правила и формулы***, использованные при решении каждой задачи.
4. Все искомые величины при расчетах нужно вычислять ***с точностью до четырех цифр после запятой.***
5. Студент должен ***уметь решать задачи, аналогичные*** задачам, входящим в его домашнее задание.

6. Вариант выбирается ***по последней цифре зачетной книжки***. В случае если последняя цифра ноль, решается 10 вариант.

Вариант	***Тема 1***	***Тема 2***	***Тема 3***	***Тема 4***	***Тема 5***	***Тема 6***
Первый	1,11	1,11	1,11	1,11	1,11	1,11
Второй	2,12	2,12	2,12	2,12	2,12	2,12
Третий	3,13	3,13	3,13	3,13	3,13	3,13
Четвертый	4,14	4,14	4,14	4,14	4,14	4,14
Пятый	5,15	5,15	5,15	5,15	5,15	5,15
Шестой	6,16	6,16	6,16	6,16	6,16	6,16
Седьмой	7,17	7,17	7,17	7,17	7,17	7,17
Восьмой	8,18	8,18	8,18	8,18	8,18	8,18
Девятый	9,19	9,19	9,19	9,19	9,19	9,19
Десятый	10,20	10,20	10,20	10,20	10,20	10,20

7. Домашние задания (контрольные работы), ***выполненные не по своему варианту, не проверяются и не засчитываются.***

Задачи к теме 1

1. Руководство компании выяснило, что в среднем 85% сотрудников, отправленных на стажировку по применению новых информационных технологий, успешно завершают курс обучения. В дальнейшем из них 60% активно применяют в работе полученные знания. Среди тех сотрудников, которые не смогли успешно завершить обучение, новые информационные технологии успешно применяют лишь 10%. Если случайно выбранный сотрудник компании активно применяет новые информационные технологии, то какова вероятность того, что он успешно прошел стажировку?

2. Агент по недвижимости пытается продать участок земли под застройку. Он полагает, что участок будет продан в течение ближайших шести месяцев с вероятностью 0,85, если экономическая ситуация в регионе не будет ухудшаться. Если же экономическая ситуация будет ухудшаться, то вероятность продать участок составит 0,4. Экономист, консультирующий агента, полагает, что с вероятностью, равной 0,6, экономическая ситуация в регионе в течение следующих шести месяцев будет ухудшаться. Чему равна вероятность того, что участок будет продан в течение ближайших шести месяцев?

3. Судоходная компания организует средиземноморские круизы в течение летнего времени и проводит несколько круизов в сезон. Поскольку в этом виде бизнеса очень высокая конкуренция, то важно, чтобы все каюты зафрахтованного под круизы корабля были полностью заняты туристами, тогда компания получит прибыль. Эксперт по туризму, нанятый компанией, предсказывает, что вероятность того, что корабль будет полон в течение сезона, равна 0,87, если доллар не подорожает по отношению к рублю, и с вероятностью - 0,64, если доллар подорожает. По оценкам экономистов, вероятность того, что в течение сезона доллар подорожает по отношению к

рублю, равна 0,1. Чему равна вероятность того, что билеты на все круизы будут проданы?

4. Исследованиями маркетологов установлено, что мужчины и женщины по-разному реагируют на рекламу средств бытовой химии. Результаты исследований показали, что 64% женщин позитивно реагируют на такую рекламу, считая, что она дает полезную информацию о новинках в этой сфере, в то время как 48% мужчин реагируют на подобную рекламу негативно. 12 женщин и 8 мужчин заполнили анкету, в которой оценили новую рекламу средств бытовой химии. Случайно извлеченная анкета содержит негативную реакцию. Чему равна вероятность того, что её заполняла женщина?

5. Компьютерная фирма разработала программу автоматизации учета в кафе и ресторанах. Рекламные материалы были разосланы в крупнейшие кафе и рестораны города, которые составляют 70% от общего числа предприятий питания города. Закупили программу 40% кафе и ресторанов, которые получили рекламные материалы, и 15% не получавших ее. Какова вероятность того, что случайно выбранное кафе заказало новую программу автоматизации учета?

6. Экспортно-импортная фирма собирается заключить контракт на поставку сельскохозяйственного оборудования в одну из развивающихся стран. Если основной конкурент фирмы не станет одновременно претендовать на заключение контракта, то вероятность получения контракта оценивается в 0,55; в противном случае - в 0,35. По оценкам экспертов компании, вероятность того, что конкурент выдвинет свои предложения по заключению контракта, равна 0,30. Чему равна вероятность заключения контракта?

7. Сотрудники отдела маркетинга полагают, что в ближайшее время ожидается рост спроса на продукцию фирмы. Вероятность этого они оценивают в 0,72. Консультационная фирма, занимающаяся прогнозом рыночной ситуации, подтвердила предположение о росте спроса. Положительные прогнозы консультационной фирмы сбываются с вероятностью 0,93, а отрицательные - с вероятностью 0,96. Какова вероятность того, что рост спроса действительно произойдет?

8. Из числа авиалиний некоторого аэропорта 70% - местные, 20% - по СНГ и 10% - в дальнее зарубежье. Среди пассажиров местных авиалиний 60% путешествуют по делам, связанным с бизнесом, на линиях СНГ таких пассажиров 50%, на международных - 90%. Из прибывших в аэропорт пассажиров случайно выбирается один. Чему равна вероятность того, что он бизнесмен?

9. Аудитор осуществляет проверку фирмы. В ходе работы у него накопилось 2 стопы бухгалтерских документов. В первой стопе из 67 документов 7 содержат ошибки, а во второй стопе из 45 документов 4 документа с ошибками. Случайно был переложен один документ из первой стопы во вторую. Какова вероятность того, что документ, извлеченный из второй стопы, содержит ошибку?

10. Компьютерная фирма продает мониторы 4 марок. При этом известно, что мониторы Sony составляют 24% от продаж, Panasonic-28%, LG – 16%,

Samsung-32%. Вероятность неполадок в первый год работы для мониторов Sony составляет 0,01, Panasonic-0,02, LG – 0,03, Samsung-0,02. Какова вероятность неполадок в первый год работы случайно выбранного монитора?

11. При слиянии акционерного капитала двух фирм аналитики фирмы, получающей контрольный пакет акций, полагают, что сделка принесет успех с вероятностью, равной 0,65, если председатель совета директоров поглощаемой фирмы выйдет в отставку; если он откажется, то вероятность успеха равна 0,3. Предполагается, что вероятность ухода в отставку председателя составляет 0,7. Чему равна вероятность успеха сделки?

12. На АЭС установлена система аварийной сигнализации. Когда возникает аварийная ситуация, звуковой сигнал срабатывает с вероятностью 0,999. Звуковой сигнал может сработать случайно и без аварийной ситуации с вероятностью 0,002. Реальная вероятность аварийной ситуации равна 0,004. Предположим, что звуковой сигнал сработал. Чему равна вероятность того, что это случилось в условиях реальной аварийной ситуации?

13. Нефтеразведочная экспедиция проводит исследования для определения вероятности наличия нефти на месте предполагаемого бурения скважины. Исходя из результатов предыдущих исследований, нефтеразведчики считают, что вероятность наличия нефти на проверяемом участке равна 0,55. На завершающем этапе разведки проводится сейсмический тест, который имеет определенную степень надежности: если на проверяемом участке есть нефть, то тест укажет на нее в 92% случаев; если нефти нет, то в 14% случаев тест может ошибочно указать на ее наличие. Сейсмический тест указал на присутствие нефти. Чему равна вероятность того, что запасы нефти на этом участке существуют реально?

14. Вероятность того, что новый товар будет пользоваться спросом на рынке, если конкурент не выпустит в продажу аналогичный продукт, равна 0,58. Вероятность того, что товар будет пользоваться спросом при наличии на рынке конкурирующего товара, 0,32. Вероятность того, что конкурирующая фирма выпустит аналогичный товар на рынок в течение интересующего нас периода, 0,24. Чему равна вероятность того, что товар будет иметь успех?

15. Вероятность того, что клиент банка не вернет заем в период экономического роста, равна 0,06, а в период экономического кризиса - 0,23. Предположим, вероятность того, что начнется период экономического роста, равна 0,79. Чему равна вероятность того, что случайно выбранный клиент банка не вернет полученный кредит?

16. Экономист-аналитик условно подразделяет экономическую ситуацию в стране на “хорошую”, “посредственную” и “плохую” и оценивает их вероятности для данного момента времени в 0,25, 0,60 и 0,15 соответственно. Некоторый индекс экономического состояния возрастает с вероятностью 0,7, когда ситуация “хорошая”; с вероятностью 0,2, когда ситуация “посредственная”, и с вероятностью 0,1, когда ситуация “плохая”. Пусть в настоящий момент индекс экономического состояния возрос. Чему равна вероятность того, что экономика страны на подъеме?

17. Керамическая плитка одной марки, цвета и размера выпускается двумя цехами завода: первый цех выпускает 60% плитки, а второй 40%. Причем известно, что 8% продукции первого цеха имеют дефекты, тогда как этот же показатель для второго цеха равен 5%. Случайно взятая плитка имеет дефект. Чему равна вероятность того, что она выпущена первым цехом?

18. Опрос показал, что из 26 студентов, обучающихся в первой группе, 18 ростовчан, а остальные живут в других городах, во второй группе 17 студентов-ростовчан, а остальные 10 живут в других городах. Из второй группы в первую был переведен один студент. После перевода один студент первой группы был вызван в деканат, и оказалось, что это студент ростовчанин. Какова вероятность того, что из второй группы в первую был переведен студент-ростовчанин?

19. Страховая компания делит водителей, заключивших договор автокаско, на следующие группы риска: 1 группа – низкий риск; 2 группа - средний; 3 группа – высокий риск. Среди клиентов страховой компании 25% - первой группы; 65% - второй группы; 10% - третьей группы. Вероятность того, что страховое событие произойдет и страховая компания будет вынуждена выплатить страховое возмещение, для первой группы риска оценивается как 0,1; для второй группы – 0,2; для третьей – 0,3. Какова вероятность того, что случайно выбранный клиент, получивший страховое возмещение, относится к группе среднего риска?

20. Работа сотрудников торгового зала супермаркета организована в две смены. В первой смене работают 5 мужчин и 7 женщин, во второй смене – 9 мужчин и 10 женщин. Из второй смены в первую был переведен один сотрудник. Клиент супермаркета пригласил сотрудника торгового зала для консультации. Консультировал клиента сотрудник – мужчина. Какова вероятность того, что из второй смены в первую была переведена женщина?

Задачи к теме 2

1. Нефтеразведывательная компания получила финансирование для проведения 7 нефтеразработок. По оценкам специалистов, вероятность успешной нефтеразведки составляет 0,2. Предположим, что нефтеразведку осуществляют независимые друг от друга разведывательные партии.

а) Составьте ряд распределения возможного числа успешных нефтеразведок и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что как минимум три нефтеразведки принесут успех?

2. В салоне мобильной техники представлены 4 модели телефона Samsung, 5 моделей телефона Nokia и 6 моделей телефона Motorola. В течение дня было продано 3 телефона.

а) Составьте ряд распределения числа телефонов Samsung, среди 3 проданных телефонов, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что в течение дня было продано как минимум два телефона Samsung?

3. Некоторый ресторан славится хорошей кухней. Управляющий рестораном утверждает, что в субботний вечер в течение получаса подходит в среднем 5 групп посетителей.

а) Составьте ряд распределения возможного числа групп посетителей ресторана в течение получаса; постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что три или более групп посетителей прибудут в ресторан в течение 10-минутного промежутка времени?

4. В кредитном отделе банка работают 5 специалистов с высшим финансовым образованием и 3 специалиста с высшим юридическим образованием. Руководство банка решило направить 3 специалистов для повышения квалификации, выбирая их в случайном порядке.

а) Составьте ряд распределения числа специалистов с высшим юридическим образованием, которые могут быть направлены на повышение квалификации, и постройте его график;

б) Найдите числовые характеристики этого распределения.

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Какова вероятность того, что повышать квалификацию будут не более двух специалистов с высшим юридическим образованием?

5. Для экспертной оценки качества растворимого кофе было отобрано 9 образцов разных производителей: 6 образцов фирмы Nestle и 3 образца фирмы Kraft Food. В результате проверки выяснилось, что 4 случайно выбранных образца соответствуют стандартам качества.

а) Составьте ряд распределения числа образцов продукции фирмы Nestle среди отобранных и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что как минимум два образца фирмы Nestle соответствуют качеству?

6. В течение часов пик в общественном транспорте города происходит в среднем два дорожных происшествия в час. Утреннее время пик длится полтора часа, а вечернее - два часа.

а) Составьте ряды распределения числа дорожных происшествий в утренние и вечерние часы пик и постройте их графики;

б) Найдите числовые характеристики этих распределений;

в) Запишите функции распределений вероятностей и постройте их графики;

г) Чему равна вероятность того, что в определенный день в течение и утреннего, и вечернего времени не произойдет ни одного дорожного происшествия?

7. В городе 6 коммерческих банков. У каждого из них риск банкротства в течение года составляет 10%.

а) Составьте ряд распределения числа банков, которые могут обанкротиться в течение следующего года; постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что в течение года обанкротятся не более двух банков?

8. В течение семестра преподаватели проводят консультации по вопросам, которые остались неясными для студентов. Преподаватель, проводящий консультации по статистике, заметил, что в среднем 12 студентов посещают его за час консультационного времени, хотя число студентов, посещающих консультацию в определенный день, в назначенный час, - случайная величина.

а) Составьте ряд распределения числа студентов, посещающих консультации преподавателя по статистике в течение получаса, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что трое студентов придут на консультацию в течение определенных 15 минут?

9. Сеть кафе «Пить кофе» включает 7 кофеен, 3 из которых имеют круглосуточный режим работы. Для оценки качества обслуживания клиентов администрация кафе случайным образом отбирает 4 кофейни.

а) Составьте ряд распределения числа кофеен с круглосуточным режимом работы, отобранных для анализа, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что в исследовании будут участвовать не более двух круглосуточно работающих кофеен?

10. Туристическая фирма оценивает вероятность того, что клиент отменит уже оплаченное путешествие вследствие личных обстоятельств, как 0,1. Группа из 6 туристов оплатила тур в Индию.

а) Составьте ряд распределения числа туристов, отменивших поездку вследствие личных обстоятельств, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Определите вероятность того, что не более одного туриста отменят поездку.

11. В мастерскую по ремонту бытовой техники поступили 8 холодильников, из которых 3 подлежали гарантийному обслуживанию. Бригада специалистов, работающая в первую смену, получила наряд на ремонт 4 холодильников.

а) Составьте ряд распределения числа холодильников, отремонтированных по гарантии в первую смену; если холодильники для ремонта отбирались случайным образом, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Определите вероятность того, что по гарантии было отремонтировано не более двух холодильников.

12. Для того чтобы проверить правильность своих финансовых счетов, компания регулярно пользуется услугами аудиторов для проверки в бухгалтерских проводках счетов. Предположим, что служащие компании при обработке входящих счетов допускают примерно 5% ошибок. Предположим, аудитор случайно отбирает 5 входящих документов.

а) Составьте ряд распределения числа ошибок, выявленных аудитором, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Определите вероятность того, что аудитор обнаружит не менее двух ошибок.

13. В магазине имеется 11 автомобилей определенной марки. Среди них - 6 автомобилей черного цвета, 3 - серого и 2 - белого. Представители фирмы обратились в магазин с предложением о продаже им трех автомобилей этой марки, безразлично какого цвета.

а) Составьте ряд распределения числа проданных автомобилей черного цвета при условии, что автомобили отбирались случайно, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Напишите функцию распределения вероятностей и постройте ее график;

г) Какова вероятность того, что среди проданных фирме автомобилей окажется по крайней мере 2 автомобиля черного цвета?

14. В международном аэропорту время прибытия самолетов различных рейсов высвечивается на электронном табло. Появление информации о различных рейсах происходит случайно и независимо друг от друга. В среднем в аэропорт прибывает 6 рейсов в течение получаса.

а) Составьте ряд распределения числа сообщений о прибытии самолетов в течение получаса и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что в течение получаса появится информация о прибытии не менее трех рейсов?

д) Чему равна вероятность того, что в течение четверти часа не появится информация о прибытии ни одного самолета?

15. Телевизионный канал рекламирует новую марку автомобилей. Вероятность того, что телезритель увидит эту рекламу, оценивается в 0,4. В случайном порядке выбраны 5 телезрителей.

а) Составьте ряд распределения числа лиц, видевших рекламу, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что по крайней мере 2 телезрителя этого канала видели рекламу новой марки автомобиля?

16. Экзаменационный тест содержит 5 вопросов, каждый из которых имеет 4 варианта ответа и только 1 из них верный.

а) Составьте ряд распределения числа правильных ответов и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что по крайней мере 3 ответа будут правильными?

17. Менеджер ювелирного магазина утверждает, что в течение дня совершается в среднем 4 покупки.

а) Составьте ряд распределения числа покупок, совершенных в ювелиром магазине в течение дня, постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что за два дня в магазине будет совершено не более 2 покупок?

18. В подгруппе английского языка занимается 9 студентов, 4 из которых окончили школы с углубленным изучением языка. Для стажировки по бухгалтерскому учету в Англии случайным образом отбирают 3 студентов.

а) Составьте ряд распределения числа студентов, среди отобранных, углубленно изучавших английский язык, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что на стажировку будет отправлено не более двух студентов, окончивших ранее спецшколы?

19. По данным страховой компании, вероятность неурожая составляет 0,3. В случае неурожая страховая фирма обязуется выплатить страховое возмещение. Договор страхования был заключен с 5 фермерскими хозяйствами.

а) Составьте ряд распределения числа фермерских хозяйств, которые могут получить страховое возмещение, и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что страховое возмещение было выплачено не более трем фермерским хозяйствам?

20. На предприятии 2000 единиц оборудования определенного вида Вероятность отказа единицы оборудования в течение часа составляет 0,001.

а) Составьте ряд распределения числа отказов оборудования в течение часа и постройте его график;

б) Найдите числовые характеристики этого распределения;

в) Запишите в общем виде функцию распределения вероятностей и постройте ее график;

г) Чему равна вероятность того, что в течение часа откажут как минимум 3 единицы оборудования?

Задачи к теме 3

1. Компьютерная система содержит 50 одинаковых микрочипов. Вероятность того, что любой микрочип будет работать в заданное время, равна 0,9. Для выполнения некоторой операции требуется, чтобы, по крайней мере, 30 микрочипов было в рабочем состоянии.

а) Чему равна вероятность того, что операция будет выполнена успешно?

б) Чему равна вероятность того, что будут работать 47 микрочипов?

2. Почтовое отделение быстро оценивает объём переводов в рублях, взвешивая почтовые отправления, полученные в течение каждого текущего рабочего дня. Установлено, что если вес почтовых отправлений составляет N кг, то объём переводов в рублях есть случайная величина, распределенная по нормальному закону со средним значением 160N и стандартным отклонением 20N руб. Найти вероятность того, что в день, когда вес почтовых отправлений составит 150 кг, объём переводов в рублях будет находиться в пределах:

а) от 21000 до 27000 руб.; б) более 28500 руб.; в) менее 22000 руб.

3. Менеджер крупного ресторана по опыту знает, что только 80% людей, сделавших заказ на вечер, придут в ресторан поужинать. В один из вечеров менеджер решил принять 60 заказов, хотя в ресторане было лишь 55 свободных столиков. Чему равна вероятность того, что более 55 посетителей придут на заказанные места?

4. Экзамен по математической статистике успешно сдают 75% студентов дневного отделения. Если на втором курсе факультета обучается 250 студентов, то какова вероятность того, что 203 студента сдадут экзамен успешно?

5. В отделе продаж страховой компании работают 45 сотрудников. Вероятность того, что сотрудник выполнит план по числу заключенных договоров, оценивается начальником отдела как 0,7. Какова вероятность того, что:

а) план выполнят как минимум 35 сотрудников?

б) план выполнят не более 30 сотрудников?

в) план выполнят 37 сотрудников?

6. Отдел маркетинга фармацевтической компании утверждает, что новая модификация таблеток от головной боли используется 30% пациентов. Если среди пациентов было отобрано 80 человек, то какова вероятность того, что отобранная доля лиц, предпочитающих новую модификацию таблеток, не будет отличаться по абсолютной величине от истинной доли более чем на 0,1?

7. Дневная выручка супермаркета распределена по нормальному закону с математическим ожиданием 10000 у.е. и стандартным отклонением 1400 у.е. Найдите вероятность того, что:

а) выручка супермаркета окажется более 13000 у.е.;

б) выручка супермаркета окажется менее 8000 у.е.;

в) найдите границы, в которых будет находиться выручка супермаркета согласно правилу трех сигм.

8. По данным независимого исследования, хлеб определенного сорта составляет 15% от совокупной реализации хлебобулочных изделий. Если выборочному обследованию были подвергнуты 80 торговых предприятий, то какова вероятность того, что доля реализации хлеба определенного сорта в генеральной совокупности будет отличаться по абсолютной величине от истинной доли менее чем на 5%?

9. В течение месяца кредитным отделом банка было выдано 68 ипотечных кредитов. Менеджер банка оценивает вероятность просрочки оплаты таких кредитов как 0,2. Какова вероятность того, что в течение срока кредитования будут просрочены:

а) как минимум 15 кредитов?

б) не более 18 кредитов?

в) 16 кредитов?

10. Фирма, занимающаяся продажей товаров по каталогу, ежемесячно получает по почте заказы. Число этих заказов - нормально распределенная случайная величина со средним квадратическим отклонением σ = 560 и неизвестным математическим ожиданием. В 90% случаев число ежемесячных заказов превышает 12439. Найдите ожидаемое среднее число заказов, получаемых фирмой за месяц.

11. Дневная добыча угля в некоторой шахте распределена по нормальному закону с математическим ожиданием 870 тонн и стандартным отклонением 90 тонн.

а) Найдите вероятность того, что в определенный день будут добыты по крайней мере 900 тонн угля.

б) Определите долю рабочих дней, в которые будет добыто от 860 до 940 тонн угля.

в) Найдите вероятность того, что в данный день добыча угля окажется ниже 750 тонн.

12. Кандидат на выборах считает, что 20% избирателей в определенной области поддерживают его избирательную платформу. Если 72 избирателя случайно отобраны из числа избирателей данной области, найдите вероятность того, что выборочная доля избирателей, поддерживающих кандидата, не будет отличаться по абсолютной величине от истинной доли более чем на 0,09.

13. Еженедельный выпуск продукции на заводе приблизительно распределен по нормальному закону со средним значением, равным 150000 единиц продукции в неделю, и стандартным отклонением - 12000 ед. Найдите вероятность того, что еженедельный выпуск продукции:

а) превысит 170000 единиц;

б) окажется ниже 100000 единиц в данную неделю?

в) Предположим, что возникли трудовые споры, и недельный выпуск продукции стал ниже 90000 ед. Менеджеры обвиняют профсоюз в беспрецедентном падении выпуска продукции, а профсоюз утверждает, что выпуск продукции находится в пределах принятого уровня ($\pm 3\sigma$). Можно ли доверять профсоюзу?

14. Вес тропического грейпфрута, выращенного в Краснодарском крае, - нормально распределенная случайная величина с неизвестным математическим ожиданием и дисперсией, равной 0,09. Агрономы знают, что 75% фруктов весят меньше, чем 0,5 кг. Найдите ожидаемый вес случайно выбранного грейпфрута.

15. Один из методов, позволяющих добиться успешных экономических прогнозов, состоит в применении согласованных подходов к решению конкретной проблемы. Обычно прогнозом занимается большое число аналитиков. Средний результат таких индивидуальных прогнозов представляет собой общий согласованный прогноз. Пусть этот прогноз относительно величины банковской процентной ставки в текущем году подчиняется нормальному закону со средним значением a = 11% и стандартным отклонением σ = 3,6%. Из группы аналитиков случайным образом отбирается один человек. Найдите вероятность того, что, согласно прогнозу этого аналитика, уровень процентной ставки:

а) превысит 13%;

б) окажется менее 16%;

в) будет в пределах от 13 до 17%.

16. Предположим, что в течение года цена на акции некоторой компании есть случайная величина, распределенная по нормальному закону с математическим ожиданием, равным 50 условным денежным единицам, и стандартным отклонением, равным 10. Чему равна вероятность того, что в случайно выбранный день обсуждаемого периода цена за акцию будет:

а) более 70 условных денежных единиц?

б) ниже 50 за акцию?

в) между 45 и 58 условными денежными единицами за акцию?

17. По данным университета, лишь 45% абитуриентов получают положительные оценки на вступительных экзаменах. Предположим, что в приемную комиссию поступило 2120 заявлений. Чему равна вероятность того, что:

а) хотя бы 970 абитуриентов получат положительные оценки на вступительных экзаменах?

б) 950 абитуриентов получат положительные оценки на вступительных экзаменах?

18. Средний срок службы коробки передач до капитального ремонта у автомобиля определенной марки составляет 56 месяцев со стандартным отклонением σ = 16 мес. Привлекая покупателей, производитель хочет дать гарантию на этот узел, обещая сделать ремонт коробки передач нового автомобиля в случае ее поломки до определенного срока. Пусть срок службы коробки передач подчиняется нормальному закону. На сколько месяцев в таком

случае производитель должен дать гарантию для этой детали, чтобы число бесплатных ремонтов не превышало 2,275% проданных автомобилей?

19. При производстве безалкогольных напитков специальный аппарат разливает определенное число унций (1 унция = 28,3 г) напитка в стандартную ёмкость. Число разлитых унций подчиняется нормальному закону с математическим ожиданием, зависящим от настройки аппарата. Количество унций напитка, разлитых отдельным аппаратом, имеет стандартное отклонение σ = 0,4 унции. Пусть ёмкости объёмом в 8 унций наполняются кока-колой. Сколько унций напитка должен в среднем разливать аппарат, чтобы не более 3% ёмкостей оказались переполненными?

20. Налоговая инспекция утверждает, что нарушения налогового законодательства характерны для 35% предприятий города. Тщательной проверке были подвергнуты 59 предприятий. Чему равна вероятность того, что доля предприятий – нарушителей будет отличаться от истинной доли более чем на 0,12?

Задачи к теме 4

1. Имеются данные об уровне рентабельности ряда сталеплавильных компаний:

Компания, №	1	2	3	4	5
Рентабельность, (%)	12	17	20	14	16

Определите среднюю рентабельность компании. Охарактеризуйте колеблемость рентабельности с помощью различных показателей вариации. Сделайте выводы.

2. По данным Всероссийского союза страховщиков, статистика страховых выплат мошенникам (среди всех произведенных выплат) в 2014 году в различных странах имела следующий вид:

Страны	Франция	Германия	Великобритания	Канада	США	Россия
Доля страхо-вых выплат мошен-никам (%)	1,8	2,9	3,2	3,4	9,5	10,0

Определить средний процент страховых выплат мошенникам. Охарактеризовать колеблемость доли страховых выплат с помощью различных показателей вариации. Сделать выводы.

3. На основании данных о выпуске иностранных автомобилей различных марок в России в 2013 году определить средний объем производства иномарок, среднее квадратическое отклонение, коэффициент вариации. Объяснить полученные результаты.

Марка автомобиля	Kia	Renault	Hyundai	Ford	Chevrolet	Chery	Hummer
Произведено в 2005 году, (тыс.штук)	16,3	10,2	44,4	32,0	51,8	8,3	3,5

4. На основании данных о динамике импорта рыбных товаров Россией в 2008-2014 годах (в млн. долл.) определить среднегодовой объем импорта рыбных товаров, дисперсию, среднее квадратическое отклонение и коэффициент вариации. Объяснить полученные результаты.

Год	2008	2009	2010	2011	2012	2013	2014*
Рыба свежая и охлажденная	6,2	13,9	32,4	72,2	131,9	150,2	170,5

* Данные за 2014 год являются прогнозными.

5. Имеются данные о размерах чистой прибыли крупнейших российских нефтяных компаний в первом полугодии 2014 года:

Компания	«Лукойл»	«Роснефть»	«ТНК-ВР»	«Сургут-нефть»	«Газпром-нефть»	«Татнефть»
Чистая прибыль (млрд. руб.)	43,2	60,0	38,7	47,9	30,0	23,4

Определите средний размер чистой прибыли нефтяной компании, дисперсию, среднее квадратическое отклонение и коэффициент вариации. Объясните полученные результаты.

6. Имеются данные о среднемесячных суммах, которые россияне тратили на оплату электроэнергии в 2014 году:

Города и другие населенные пункты	Города - миллионники	Крупные города	Средние города	Малые города	Села
Размер оплаты (руб.)	303	240	250	269	177

Определить среднемесячный размер оплаты электроэнергии по всей совокупности. Найти и проанализировать дисперсию, среднее квадратическое отклонение, коэффициент вариации. Сделать выводы.

7. На основании данных о численности студентов учебных заведений среднего профессионального образования за период 2009-2013гг. определить среднегодовую численность студентов, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Объяснить полученные результаты.

Год	2009	2010	2011	2012	2013
Число студентов, (млн.чел.)	2,470	2,585	2,612	2,503	2,461

8. Имеются данные о распределении городского населения по затратам на ежемесячную оплату электроэнергии:

Размер оплаты (руб.)	Менее 100	100-200	200-300	300-400	400-500	500-600	Более 600
Удельный вес в общей численности населения (%)	12	29	25	15	11	6	2

Определить среднемесячные затраты городского населения на оплату электроэнергии. Найти и проанализировать дисперсию, среднее квадратическое отклонение, коэффициент вариации. Построить гистограмму распределения городского населения по затратам на ежемесячную оплату электроэнергии. Сделать выводы.

9.По данным поискового сайта Рамблер доля Интернет-пользователей в различных возрастных группах распределена следующим образом:

Возраст, лет	18-25	25-35	35-45	45 и более
Доля Интернет-пользователей (% от числа опрошенных)	36	31	20	13

На основании этих данных определить средний возраст Интернет-пользователей. Найти и проанализировать дисперсию, среднее квадратическое отклонение, коэффициент вариации. Построить гистограмму распределения доли Интернет-пользователей по различным возрастным группам. Сделать выводы.

10. Имеются данные о распределении объемов продаж мобильных телефонов в сетевых салонах связи по ценовым группам:

Цена, тыс. руб.	1-2	2-3	3-4	4-5	5-6	6-7
Доля в объеме продаж (%)	14	23	25	23	8	9

Определить среднюю цену мобильного телефона, продаваемого в сетевых салонах связи, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Построить гистограмму распределения объемов продаж мобильных телефонов по ценовым группам. Сделать выводы.

11.Для выяснения возрастных особенностей кадрового состава сотрудников фирмы было произведено выборочное обследование, в результате которого получены следующие данные:

Возраст сотрудников, лет	20-25	25-30	30-35	35-40	40-45	45-50	Старше 50
Число сотрудников	20	25	30	20	28	15	12

Определить средний возраст сотрудника фирмы, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Построить гистограмму распределения числа сотрудников по интервалам возраста. Сделать анализ полученных результатов.

12. Ниже приводятся данные о возрастном составе безработных города, зарегистрированных в службе занятости, в %:

Возраст (лет)	до 20	20-24	25-29	30-49	50-54	55-59	60 и старше
Мужчины	7,7	17,0	11,9	50,9	4,2	5,7	2,6
Женщины	11,2	18,5	11,7	49,5	4,0	3,8	1,3

Найдите средний возраст безработных мужчин и женщин, дисперсию, среднее квадратическое отклонение и коэффициент вариации. Оцените различия показателей возрастного состава безработных мужчин и женщин. Сделайте выводы.

13. Для оценки состояния деловой активности промышленных предприятий различных форм собственности были проведены выборочные бизнес-обследования и получены следующие результаты:

Интервалы значений показателя деловой активности (в баллах)	0 - 8	8 - 16	16 - 24	24 - 32
Число предприятий (акционерные общества открытого типа)	10	15	8	5

Постройте гистограмму распределения частот. Найдите среднее значение показателя деловой активности, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Объясните полученные результаты.

14. Предположим, что на некотором предприятии собраны данные о числе дней, пропущенных работниками по болезни.

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

Постройте полигон распределения частот. Найдите среднее число пропущенных дней, стандартное отклонение, коэффициент вариации. Является ли распределение симметричным?

15. Постройте гистограмму частот, найдите среднюю арифметическую, среднее квадратическое отклонение и коэффициент вариации для данных о дневной выручке в магазине электроники:

Выручка, у.е.	0-200	200-300	300-400	400-500	500-600	600-700
Число дней	3	5	9	14	8	3

16. Администрацию универсама интересует оптимальный уровень запасов продуктов в торговом зале, а также среднемесячный объем покупок товаров, которые не являются предметом ежедневного потребления в семье (например, таких как сода). Для выяснения этого вопроса менеджер универсама в течение января регистрировал частоту покупок стограммовых пакетиков с содой и собрал следующие данные (x_i): 8, 4, 4, 9, 3, 3, 1, 2, 0, 4, 2, 3, 5, 7, 10, 6, 5, 7, 3, 2, 9, 8, 1, 4, 6, 5, 4, 2, 1, 8.

Постройте вариационный ряд, определите его числовые характеристики. Какие рекомендации Вы дали бы администрации универсама?

17. Число пассажиров компании «Аэрофлот - Дон» рейса Ростов – Стамбул в мае текущего года составило: 125, 130, 121, 124, 128, 136, 125, 130, 124, 128, 125, 125, 130, 128, 125, 128.

Составьте вариационный ряд. Чему равно среднее число пассажиров в рейсе? Рассчитайте показатели вариации. Сделайте анализ полученных результатов.

18. Имеются данные об объемах экспорта российской нефти в Польшу по нефтепроводу «Дружба» за первый квартал 2014 года:

Компания - экспортер	«Лукойл»	«Роснефть»	«ТНК-ВР»	«Сургутнефть»	«Газпромнефть»	«Татнефть»
Объем экспорта (млн.т)	0,496	1,380	1,055	1,000	0,600	0,300

Определите средний объем экспорта нефти в Польшу в первом квартале 2014 года. Рассчитайте дисперсию, среднее квадратическое отклонение, коэффициент вариации. Проанализируйте полученные результаты.

19. Имеются данные о вредных выбросах в атмосферу в 2013 году по ряду крупных российских городов:

Город	Москва	Санкт – Петербург	Самара	Краснодар	Ростов-на-Дону	Новосибирск	Челябинск
Объем выбросов в атмосферу (тыс. тонн)	89,0	52,5	33,5	99,0	10,6	109,2	140,9

Определить средний объем выбросов в атмосферу, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Проанализировать полученные результаты.

20. Имеются данные об объемах загрязненных сточных вод по ряду крупных российских городов в 2014 году:

Город	Москва	Санкт-Петербург	Самара	Краснодар	Ростов-на-Дону	Новосибирск	Челябинск
Объем загрязненных сточных вод (тыс. тонн)	1922,0	753,0	238,0	74,0	104,0	4,1	234,0

Определить средний объем выбросов в атмосферу, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Проанализировать полученные результаты.

Задачи к теме 5

1. Результаты 10-дневного наблюдения в молочном отделе супермаркета показали, что в среднем в день реализуется 144 пачки творога с исправленным средним квадратическим отклонением в 23 пачки. Оцените потребность супермаркета в закупке творога, построив 99% доверительный интервал.

2. Фирма, торгующая автомобилями в небольшом городе, собирает информацию о состоянии местного автомобильного рынка в текущем году. С этой целью из 8500 горожан в возрасте 18 лет и старше отобрано 500 человек. Среди них оказалось 130 человек, планирующих приобрести новый автомобиль в текущем году. Оцените долю лиц в генеральной совокупности в возрасте 18 лет и старше, планирующих приобрести новый автомобиль в текущем году, если $\alpha = 0{,}01$.

3. При выборочном опросе 1200 телезрителей оказалось, что 456 из них регулярно смотрят программы телеканала НТВ. Постройте 99%-ный доверительный интервал, оценивающий долю всех телезрителей, предпочитающих программы телеканала НТВ.

4. Выборочные обследования показали, что доля покупателей, предпочитающих новую модификацию зубной пасты, составляет 60% от общего числа покупателей данного товара. Каким должен быть объём выборки, чтобы можно было получить оценку генеральной доли с точностью не менее 0,1 при доверительной вероятности 0,954?

5. Среднемесячные расходы на питание домохозяйств из трех человек оцениваются по случайной выборке. С вероятностью 0,997 определите объем выборки, необходимой для такой оценки, если ошибка выборки не должна превышать 500 рублей, а по результатам более ранних исследований среднее квадратическое отклонение составило 2000 рублей.

6. Менеджер компании, занимающейся прокатом автомобилей, хочет оценить среднюю величину пробега одного автомобиля в течение месяца. Из 280 автомобилей, принадлежащих компании, методом случайной бесповторной выборки отобрано 35. По данным этой выборки установлено, что средний пробег автомобиля в течение месяца составляет 1342 км со стандартным отклонением 227 км. Считая пробег автомобиля случайной величиной, распределённой по нормальному закону, найти 95%-ный доверительный интервал, оценивающий средний пробег автомобилей всего парка в течение месяца.

7. Выборочные маркетинговые исследования показали, что 68% потребителей предпочитают приобретать черный чай без вкусовых добавок. Определите границы 95%-ного доверительного интервала доли таких потребителей в генеральной совокупности, если объем выборки составил 500 человек.

8. Выборочное исследование деятельности коммерческих банков региона показало, что в среднем каждый банк имеет 14 филиалов в регионе (со стандартным отклонением, равным 8). Найти объем выборки, позволивший сделать такую оценку, если предельная ошибка оценки генеральной средней находится в пределах 20% от ее выборочного среднего значения, а доверительная вероятность составляет 0,95.

9. Выборочное обследование распределения населения города по среднедушевому денежному доходу показало, что 25% обследованных в выборке имеют доход ниже прожиточного минимума. В каких пределах с надежностью 0,954 находится доля населения, имеющего среднедушевой доход ниже прожиточного минимума, в генеральной совокупности, если в городе проживает 1 млн. чел. и выборочное обследование осуществляется с помощью собственно-случайного бесповторного отбора?

10. Аудиторская фирма хочет проконтролировать состояние счетов одного из коммерческих банков. Для этого случайно отбираются 55 счетов. По 21 счету из 55 отобранных имело место движение денежных средств в течение месяца. Построить 95%-ный доверительный интервал, оценивающий долю счетов в генеральной совокупности, по которым имело место движение денежных средств в течение месяца.

11. Выборочные обследования, проведенные администрацией строительных магазинов города, показали, что 45% горожан планируют ремонт квартиры или дома в течение следующих трех лет. Каким должен быть объем выборки, чтобы можно было получить оценку генеральной доли с точностью не менее 0,05 при доверительной вероятности 0,95, если в городе проживает 500000 человек ?

12. Предварительный опрос покупателей магазина рыболовных принадлежностей «Серебряный ручей» показал, что 25% из них планируют в дальнейшем делать покупки в этом магазине, если им будет предоставлена дисконтная карта. Каким должен быть объем выборки, необходимый для оценки генеральной доли постоянных покупателей, при заданной точности не менее 0,04 и доверительной вероятности 0,954?

13. Среднемесячный бюджет студентов в колледжах одного из штатов США оценивается по случайной выборке. Найдите наименьший объём выборки, необходимый для такой оценки с вероятностью 0,954, если среднее квадратическое отклонение предполагается равным 100 у.е., а предельная ошибка средней не должна превышать 25 у.е.

14. Коммерческий банк, изучая возможности предоставления долгосрочных кредитов населению, опрашивает своих клиентов для определения среднего размера такого кредита. Из 9700 клиентов банка опрошено 1000 человек. Среднее значение необходимого кредита в выборке составило 7750 у.е. со стандартным отклонением 1560 у.е. Найдите границы 95%-ного доверительного интервала для оценки неизвестного среднего значения кредита в генеральной совокупности.

15. Выборочное обследование показало, что 20% студентов университета нуждаются в общежитии. Каким должен быть объем случайной бесповторной

выборки, в результате которой будет оценена генеральная доля с точностью не менее 0,03 при доверительной вероятности 0,954, если в университете обучается 5000 студентов дневного отделения?

16. По предварительным данным коммунальных служб города, 10% потребителей имеют задолженности по оплате коммунальных услуг. Каким должен быть объем выборки, необходимой для оценки генеральной доли задолжников, если предельная ошибка выборки не должна превышать 0,05 при доверительной вероятности 0,954?

17. Строительная компания хочет оценить возможности успешного бизнеса на рынке ремонтно-строительных работ. Каким должен быть объем выборки среди 1200 клиентов строительной фирмы, если среднее квадратическое отклонение по результатам пробного обследования составило 850 у.е., а предельная ошибка выборки не должна превышать 200 у.е.?

18. По данным автосалона, услугами гарантийного ремонта в течение года гарантии воспользовались 28% покупателей автомобилей. Постройте 95% доверительный интервал доли покупателей, пользующихся гарантийным ремонтом, если автосалон продал за год 297 автомобилей.

19. Опрос 20 горожан показал, что среднемесячные расходы на покупку журналов и газет составляют 125 рублей с исправленным средним квадратическим отклонением 60 рублей. Постройте 99% доверительный интервал для оценки среднемесячных расходов на прессу горожан в генеральной совокупности.

20. Для определения среднего размера дневной выручки маршрутных такси города была произведена 10%-ная случайная бесповторная выборка из 1200 маршрутных такси. В результате были получены данные о средней дневной выручке, которая составила 5000 рублей. В каких пределах с доверительной вероятностью 0,95 может находиться средняя дневная выручка всех маршрутных такси города, если среднее квадратическое отклонение составило 650 рублей?

Задачи к теме 6

1. Компания, производящая средства для потери веса, утверждает, что прием таблеток в сочетании со специальной диетой позволяет сбросить в среднем в неделю 800 граммов веса. Случайным образом отобраны 25 человек, использующих эту терапию, и обнаружено, что в среднем еженедельная потеря в весе составила 830 граммов со средним квадратическим отклонением 250 граммов. Ответьте, правда ли, что потеря в весе составляет 800 граммов? Уровень значимости $\alpha = 0,05$.

2. Компания утверждает, что новый вид зубной пасты для детей лучше предохраняет зубы от кариеса, чем зубные пасты, производимые другими фирмами. Для проверки эффекта в случайном порядке была отобрана группа из 500 детей, которые пользовались новым видом зубной пасты. Другая группа из 600 детей, также случайно выбранных, в это же время пользовалась другими видами зубной пасты. После окончания эксперимента было выяснено, что у 30 детей, использующих новую пасту, и 35 детей из контрольной группы

появились новые признаки кариеса. Имеются ли у компании достаточные основания для утверждения о том, что новый сорт зубной пасты эффективнее предотвращает кариес, чем другие виды зубной пасты? Принять уровень значимости $\alpha = 0{,}05$.

3. По оценкам оператора сотовой связи, средняя длительность ежедневных звонков составляет 24 минуты на одного абонента. Выборочное обследование 100 абонентов показало, что среднедневная длительность звонков составляет 30 минут. На уровне значимости $\alpha = 0{,}05$ оцените статистическую значимость различий выборочного обследования, если известно, что стандартное отклонение длительности звонков в генеральной совокупности составляет 3 минуты.

4. По оценкам финансовых аналитиков, риск потери денежных средств для инвесторов арт-бизнеса составляет 17% в течение пяти лет. Среди 400 постоянных клиентов аукционного дома был проведен опрос, в ходе которого выяснилось, что 65 из них потеряли средства на вложениях в предметы искусства за последние пять лет. Можно ли утверждать, что оценки финансовых аналитиков совпадают с действительностью на уровне значимости $\alpha = 0{,}01$?

5. Крупный коммерческий банк заказал маркетинговое исследование по выявлению эффекта «премирования» (калькулятор, набор ручек и др.), как стимула для открытия счета в банке. Для проверки случайным образом было отобрано 230 «премированных» посетителей и 200 «непремированных». В результате выяснилось, что 80% посетителей, которым предлагалась премия, и 75% посетителей, которым не предлагалась премия, открыли счет в банке в течение 6 месяцев. Используя эти данные, проверьте гипотезу о том, что доля «премированных» посетителей, открывших счет в банке, статистически существенно отличается от удельного веса «непремированных» посетителей, открывших счет в банке. Принять уровень значимости $\alpha = 0{,}01$.

6. По данным российской аналитической компании, средняя розничная цена покупки мобильного телефона в 2006 году составила 5000 рублей. Выборочная оценка 25 случайно выбранных телефонов, купленных в одном из салонов города, показала, что средняя цена купленного телефона составляет 5200 рублей с исправленным средним квадратическим отклонением 250 рублей. На уровне значимости $\alpha = 0{,}01$ проверьте гипотезу о том, что средняя розничная цена мобильного телефона, купленного в 2006 году, равна 5200 рублей.

7. Компания, выпускающая в продажу новый сорт сока, проводит оценку вкусов покупателей по случайной выборке из 500 человек, и оказалось, что 310 из них предпочли новый сорт всем остальным. Проверьте на уровне значимости $\alpha = 0{,}01$ гипотезу о том, что новый сорт сока предпочитают 65 % потребителей.

8. Страховая компания изучает вероятность дорожных происшествий для подростков, имеющих мотоциклы. За прошедший год проведена случайная выборка 1000 страховых полисов подростков-мотоциклистов, и выявлено, что 11 из них попадали в дорожные происшествия и предъявили компании

требование о компенсации за ущерб. Может ли аналитик компании отклонить гипотезу, о том, что менее одного процента всех подростков-мотоциклистов, имеющих страховые полисы, попадали в дорожные происшествия в прошлом году? Принять уровень значимости $\alpha = 0{,}05$.

9. Новое лекарство, изобретенное для лечения атеросклероза, должно пройти экспериментальную проверку для выяснения возможных побочных эффектов. В ходе эксперимента лекарство принимали 7000 мужчин и 6000 женщин. Результаты выявили, что 100 мужчин и 100 женщин испытывали побочные эффекты при приеме нового медикамента. Можем ли мы на основании эксперимента утверждать, что побочные эффекты нового лекарства у женщин проявляются в большей степени, чем у мужчин? Принять уровень значимости $\alpha = 0{,}01$.

10. Руководство фирмы - провайдера полагает, что проведение рекламной акции приведет к увеличению числа новых клиентов. За 30 рабочих дней после проведения рекламной акции число новых клиентов составило 120 чел., тогда как до нее в среднем за день к услугам Internet впервые подключились 2 чел. Считая среднее квадратическое отклонение равным 3, на уровне значимости 0,01 определите, принесла ли успех рекламная акция.

11. Владелец фирмы считает, что добиться более высоких финансовых результатов ему помешала неравномерность поставок комплектующих по месяцам года, несмотря на то, что поставщик в полном объеме выполнил свои обязательства за год. Поставщик утверждает, что поставки были не так уж неравномерны. Распределение поставок по месяцам года имеет следующий вид:

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Объем поставок, единиц	19	23	26	18	20	20	20	20	32	27	35	40

На уровне значимости $\alpha = 0{,}05$ определите, кто прав: владелец фирмы или поставщик? Изменится ли ответ на поставленный вопрос, если уровень значимости принять равным 0,01? Объясните результаты.

12. Годовой оборот 8 супермаркетов некоторой федеральной сети в Ростовской области составил 16 млн. у.е. с исправленным средним квадратическим отклонением 0,25 млн. у.е., а годовой оборот 5 супермаркетов этой же сети в Краснодарском крае составил 9,5 млн. у.е. с исправленным средним квадратическим отклонением 0,4 млн. у.е. Можно ли на уровне значимости $\alpha = 0{,}05$ утверждать, что в Ростовской области сеть супермаркетов работает более эффективно?

13. Компания по производству безалкогольных напитков предполагает выпустить на рынок новую модификацию популярного напитка, в котором сахар заменен сукразитом. Компания хотела бы быть уверенной в том, что не менее 60% её потребителей предпочтут новую модификацию напитка. Новый напиток был предложен на пробу 1500 человек, и 850 из них сказали, что он вкуснее старого. Может ли компания отклонить предположение о том, что 60% всех её потребителей предпочтут новую модификацию напитка старой? Принять уровень значимости $\alpha = 0{,}01$.

14. Кондитерская компания решила выяснить, действительно ли новая упаковка увеличивает объем продаж дорогих конфет. Исследования были проведены в 35 магазинах и супермаркетах, продающих конфеты в старой упаковке и в 42 магазинах, в которых продавались конфеты в новой упаковке. Среднедневной объем продаж конфет в старой упаковке составил 27,4 коробки с дисперсией 6,8, а объем продаж конфет в новой упаковке составил 35,6 с дисперсией 4,2. Можно ли на уровне значимости $\alpha = 0{,}01$ утверждать, что новая упаковка увеличила объем продаж конфет?

15. Производители нового типа аспирина утверждают, что он снимает головную боль за 30 минут. Случайная выборка 100 человек, страдающих головными болями, показала, что новый тип аспирина снимает головную боль за 33,6 минуты при среднем квадратическом отклонении 4,2 минуты. Проверьте на уровне значимости $\alpha = 0{,}05$ справедливость утверждения производителей аспирина о том, что это лекарство излечивает головную боль за 30 минут.

16. Для определения среднего размера валютного вклада клиентов коммерческого банка осуществлена случайная выборка 200 вкладчиков банка. В результате были получены следующие данные:

Размер вклада (в долларах)	До 500	500-1000	1000-1500	1500-2000	2000-2500	2500-3000	Более 3000
Число вкладов	8	16	40	72	36	18	10
Теоретические частоты	6	18	36	76	39	18	7

На основании этих данных проверить на 5% уровне значимости гипотезу о нормальном законе распределения размера валютного вклада.

17. На двух станках с программным управлением обрабатываются одинаковые детали. Для оценки точности станков отобраны 10 деталей с первого станка и 12 деталей со второго станка. По этим выборкам найдены исправленные выборочные дисперсии, равные соответственно 30 кв.ед. и 10 кв.ед. Можно ли на основании этих данных утверждать на 5% уровне значимости, что точность станков существенно различается?

18. По данным Росстата, средний возраст безработного по РФ составляет 40 лет. Выборочное обследование демографических характеристик безработных в регионе выявило, что средний возраст безработного составил 38 лет, со стандартным отклонением 4 года. Выяснить, существенны ли результаты выборочного исследования, если в выборку попало 25 человек? Ответ дать на 5% уровне значимости

19. Главный бухгалтер большой корпорации провел обследование по данным прошедшего года с целью выяснения доли некорректных счетов. Из 2000 выбранных счетов в 25 оказались некорректные проводки. Для уменьшения доли ошибок он внедрил новую систему. Год спустя он решил проверить, как работает новая система, и выбрал для проверки в порядке случайного отбора 3000 счетов компании. Среди них оказалось 30 некорректных. Можно ли утверждать, что новая система позволила уменьшить долю некорректных проводок в счетах? Принять уровень значимости $\alpha = 0{,}05$.

20. На предприятии исследовалось изменение расхода сырья на производство продукции в условиях применения новой и старой технологий изготовления изделий. Дисперсия расхода сырья на изделие по новой технологии составила 124 кв.ед., а по старой – 189 кв.ед. Считая, что расход сырья на изделие по старой и новой технологии имеет нормальный закон распределения с одинаковыми дисперсиями, выяснить, существенны ли различия в вариации расхода сырья на изделие при использовании старой и новой технологий. Ответ дать на 1% уровне значимости, применив двухстороннюю альтернативную гипотезу.

СПИСОК ЛИТЕРАТУРЫ

1. Бережная Е.В., Бережной В.И. Математические методы моделирования экономических систем: Учеб. пособие / Е.В. Бережная, В.И.Бережной –М.: Финансы и статистика, 2002. – 368 с.

2. Введение в математическое моделирование: Учеб. пособие / под ред. П.В. Трусова. – М.: Университетская книга, Логос, 2007. – 440 с.

3. Вентцель Е.С. Исследование операций. Задачи, принципы, методология. Учеб. пособие для студентов втузов.- 2-е изд., стер. – М.: Высш. шк., 2001. – 208 с.

4. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2003. – 479 с.

5. Соболь И.М. Метод Монте-Карло / И.М. Соболь – 4-е изд. – М.: Наука. Главная редакция физико-математической литературы, 1985. – 80 с.

6.Соболь И.М. Численные методы Монте-Карло / И.М. Соболь – М.: Наука. Главная редакция физико-математической литературы, 1973. – 312 с.

7. Щипачев А.М. Системный анализ и математическое моделирование процессов в машиностроении: учебное пособие / А.М. Щипачев; Уфимск. гос. авиац. техн. ун-т.- Уфа, 2008. -173 с.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица функции $\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}}$

(кривая вероятностей)

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,2	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Приложение 2

Таблица функции $\Phi_0(z) = \int\limits_0^z \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}} dx$ (функция Лапласа)

z	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,499968									
4,5	0,499997									
5,0	0,4999997									

Приложение 3

Таблица значений функции Пуассона: $P(X=m)=P_{n,m}\approx\frac{\lambda^m}{m!}\cdot e^{-\lambda}$

m \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3596	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4	-	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5	-	-	-	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6	-	-	-	-	-	-	0,0001	0,0002	0,0003

m \ λ	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,3679	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0,3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0,1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050
3	0,0613	0,1805	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150
4	0,0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337
5	0,0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0,0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0,0001	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171
8	-	0,0009	0,0081	0,0298	0,0655	0,1033	0,1304	0,1396	0,1318
9	-	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318
10	-	-	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186
11	-	-	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970
12	-	-	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13	-	-	-	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14	-	-	-	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15	-	-	-	-	0,0002	0,0009	0,0033	0,0090	0,0194
16	-	-	-	-	-	0,0003	0,0015	0,0045	0,0109
17	-	-	-	-	-	0,0001	0,0006	0,0021	0,0058
18	-	-	-	-	-	-	0,0002	0,0009	0,0029
19	-	-	-	-	-	-	0,0001	0,0004	0,0014
20	-	-	-	-	-	-	-	0,0002	0,0006
21	-	-	-	-	-	-	-	0,0001	0,0003
22	-	-	-	-	-	-	-	-	0,0001

Приложение 4

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,99**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Приложение 5

Критические точки распределения Стьюдента

Число степеней свободы k	**Уровень значимости α(двусторонняя критическая область)**					
	0,10	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,00	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,70
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,28	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,07	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
Число степеней свободы k	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α(односторонняя критическая область)					

Приложение 6

Критические точки распределения Фишера-Снедекора
($К_1$ - число степеней свободы большей дисперсии,
$К_2$ - число степеней свободы меньшей дисперсии)

Уровень значимости α = 0,01												
$К_1$ / $К_2$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	38,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,96	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,72	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,44	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости α = 0,05												
K_1 / K_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Приложение 7

Параметрические критерии значимости

Нулевая гипотеза $\mathbf{H}_0$	Дополнитель-ные условия	Критерий проверки H_0 (критериальная статистика)	Используемое распределе-ние	Конкур гип. $\mathbf{H}_1$	Критическая область и формулы для нахождения её границ		Гипотеза $\mathbf{H}_0$ не отвергается, если
$H_0 : \tilde{x} = a_0$	$\sigma_г$ известно	$U_{набл} = \dfrac{\tilde{x} - a_0}{\sigma}\sqrt{n}$	Нормальный закон Функция Лапласа	$\tilde{x} > a_0$	ПКО	$\Phi_0(u_{кр}) = \dfrac{1-2\alpha}{2}$	$\lvert U_{íàáë}\rvert < u_{êð}$
				$\tilde{x} < a_0$	ЛКО		
				$\tilde{x} \neq a_0$	ДКО	$\Phi_0(u_{кр}) = \dfrac{1-\alpha}{2}$	
$H_0 : \tilde{x} = a_0$	$\sigma_г$ неизвестно	$T_{набл} = \dfrac{\tilde{x} - a_0}{S}\sqrt{n-1}$	Стьюдента с $k_{св} = n-1$ степенями свободы	$\tilde{x} > a_0$	ПКО	$t_{правкр}(\alpha;k)$	$T_{набл} < t_{правкр}$
				$\tilde{x} < a_0$	ЛКО	$t_{лев.кр} = -t_{правкр}$	$T_{набл} > -t_{правкр}$
				$\tilde{x} \neq a_0$	ДКО	$t_{двуст.кр}(\alpha;k)$	$\lvert T_{набл}\rvert < t_{двуст.кр}$
$H_0 : \bar{x}_1 = \bar{x}_2$	$n \geq 30$	$z_{набл} = \dfrac{\tilde{x}_1 - \tilde{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	Нормальный закон Функция Лапласа	$x_1 > x_2$	ПКО	$\Phi_0(z_{кр}) = \dfrac{1-2\alpha}{2}$	$z_{набл} < z_{кр}$
				$x_1 < x_2$	ЛКО		$z_{набл} > -z_{кр}$
				$x_1 \neq x_2$	ДКО	$\Phi_0(z_{кр}) = \dfrac{1-\alpha}{2}$	$\lvert z_{набл}\rvert < z_{кр}$
$H_0 : \bar{x}_1 = \bar{x}_2$	$x_1 \to N(a_1;\sigma_1^2)$ $x_2 \to N(a_2;\sigma_2^2)$ $\sigma_{1г}^2 = \sigma_{2г}^2$	$T_{набл} = \dfrac{\tilde{x}_1 - \tilde{x}_2}{\sqrt{\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	Стьюдента с $k_{св} = n_1 + n_2 - 2$ степенями свободы	$x_1 > x_2$	ПКО	$t_{правкр}(\alpha;k)$	$T_{набл} < t_{правкр}$
				$x_1 < x_2$	ЛКО	$t_{лев.кр} = -t_{правкр}$	$T_{набл} > -t_{правкр}$
				$x_1 \neq x_2$	ДКО	$t_{двуст.кр}(\alpha;k)$	$\lvert T_{набл}\rvert < t_{двуст.кр}$
$H_0 : \sigma_{ген}^2 = \sigma_0^2$	$x \to N(a;\sigma^2)$	$\chi_{набл}^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	χ^2 распред. с $k_{св} = n-1$ степенями свободы	$\sigma_{ген}^2 > \sigma_0^2$	ПКО	$\chi_{кр}^2(\alpha;k)$	$\chi_{набл}^2 < \chi_{кр}^2$
				$\sigma_{ген}^2 < \sigma_0^2$	ЛКО	$\chi_{кр}^2(1-\alpha;k)$	$\chi_{набл}^2 > \chi_{кр}^2$
				$\sigma_{ген}^2 \neq \sigma_0^2$	ДКО	$\chi_{правкр}^2(\alpha/2;k)$ $\chi_{лев.кр}^2(1-\alpha/2;k)$	$\chi_{л.кр}^2 < \chi_{набл}^2 < \chi_{пр.кр}^2$

Нулевая гипотеза $\mathbf{H}_0$	Дополнительные условия	Критерий проверки H_0 (критериальная статистика)	Используемое распределение	Конкур гип. H_1	Критическая область и формулы для нахождения её границ		Гипотеза H_0 не отвергается, если
$H_0 : \sigma_1^2 = \sigma_2^2$	$x_1 \to N(a_1;\sigma_1^2)$ $x_2 \to N(a_2;\sigma_2^2)$	$F_{набл} = \dfrac{S_1^2(наиб)}{S_2^2(наим)}$	F распред. с $k_1 = n_1 - 1$ $k_2 = n_2 - 1$ степенями свободы	$\sigma_1^2 > \sigma_2^2$	ПКО	$F_{кр}(\alpha;k_1;k_2)$	$F_{набл} < F_{кр}$
				$\sigma_1^2 \neq \sigma_2^2$	ДКО	$F_{кр}(\alpha/2;k_1;k_2)$	$\lvert F_{набл}\rvert < F_{кр}$
$H_0 : p = p_0$	$n \to \infty$ $(n > 30)$ p неизвестна	$U_{набл} = \dfrac{m/n - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$ $q_0 = 1 - p_0$	Нормальный закон Функция Лапласа	$p > p_0$	ПКО	$\varPhi_0(u_{кр}) = \dfrac{1-2\alpha}{2}$	$U_{набл} < u_{кр}$
				$p < p_0$	ЛКО		$U_{набл} > -u_{кр}$
				$p \neq p_0$	ДКО	$\varPhi_0(u_{кр}) = \dfrac{1-\alpha}{2}$	$\lvert U_{набл}\rvert < u_{кр}$
$H_0 : p_1 = p_2$	$n \to \infty$ $(n > 30)$	$U_{набл} = \dfrac{\dfrac{m_1}{n_1} - \dfrac{m_2}{n_2}}{\sqrt{\dfrac{\overline{p}\overline{q}}{n}}}$ $\overline{p} = \dfrac{m_1 + m_2}{n_1 + n_2}$ $\overline{q} = 1 - \overline{p}$ $n = \dfrac{n_1 n_2}{n_1 + n_2}$	Нормальный закон Функция Лапласа	$p > p_0$	ПКО	$\varPhi_0(u_{кр}) = \dfrac{1-2\alpha}{2}$	$U_{набл} < u_{кр}$
				$p < p_0$	ЛКО		$U_{набл} > -u_{кр}$
				$p \neq p_0$	ДКО	$\varPhi_0(u_{кр}) = \dfrac{1-\alpha}{2}$	$\lvert U_{набл}\rvert < u_{кр}$

Ефременкова Ольга Валентиновна

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ НАДЕЖНОСТИ

Учебное пособие для студентов заочной формы обучения направления «Строительство»

Редактор Е.Ф. Изотова

Подписано к печати 29.12.17. Формат 60х84/16.
Усл. печ. л. 8,18. Тираж 50 экз. Зак. 171639. Рег. № 23.

Отпечатано в ИТО Рубцовского индустриального института
658207, Рубцовск, ул. Тракторная, 2/6.