КОНТРОЛЬНАЯ РАБОТА

4.1. Общие методические указания

Контрольная работа имеет 11 заданий (5 по теории вероятностей и 6 по математической статистике) по 10 вариантов на каждое задание. Студент выполняет тот вариант, который совпадает с последней цифрой его учебного шифра. При оформлении контрольной работы необходимо выполнять следующие правила:

- работу нужно выполнять в отдельной тетради, на обложке которой должны быть указаны фамилия и инициалы студента, его полный шифр, номер контрольной работы, дата отправки в университет;
- задачи оформляются в порядке номеров, указанных в задании, при этом записывается условие задачи и к решению прилагаются пояснения, которые должны быть подробными и написаны без сокращений;
- если работа выполнена неудовлетворительно, то вместе с рецензией ее возвращают студенту. Студент должен в короткий срок сделать необходимые доработки и послать на повторное рецензирование. При необходимости студент должен давать пояснения по всем или некоторым задачам контрольной работы.

4.2. Контрольные задания по теории вероятностей

1. Определить вероятность случайных событий по классическому определению вероятности.
2. Определить вероятности сложных событий, используя формулы сложения и умножения вероятностей.
3. Установить, являются ли случайные события независимыми, используя понятие условной вероятности.
4. Определить вероятности случайных событий, используя формулу полной вероятности и формулу Байеса.
5. Определить закон распределения случайной величины (СВ), ее числовые характеристики и вероятность попадания в заданный промежуток.

4.2.1. Методические указания к выполнению контрольных заданий и варианты заданий

Классическое определение вероятности

Случайным событием называется событие, которое в результате эксперимента может произойти или не произойти. При этом рассматриваются только такие эксперименты, которые можно повторять, при неизменном комплексе условий, произвольное число раз.

Результатом эксперимента является **случайный исход** или **элементарное событие** $\mathbf{e_i}$. Все возможные элементарные исходы образуют выборочное пространство $\mathbf{E} = \{\mathbf{e}_i\}$, где $\mathbf{i = 1, \ldots, n}$. Случайное событие может совпадать со случайным исходом эксперимента или являться какой-либо комбинацией этих исходов. Если случайное событие обозначить **A**, а все возможные исходы эксперимента $\mathbf{e_i}$, то каждому случайному событию **A** будет благоприятствовать какое-то число элементарных событий (исходов) $\mathbf{e_i}$ (ими будут $\mathbf{e_i}$, через которые можно определить случайное событие **A**).

Классическое определение вероятности случайного события **A** используется для экспериментов со случайными исходами $\mathbf{e_i}$ (элементарными событиями), которые можно повторять (воспроизводить) при неизменном комплексе условий произвольное число раз и при этом каждый раз любой исход $\mathbf{e_i}$ имеет один и тот же шанс появиться. В этом случае вероятность любого события **A**, являющегося комбинацией $\mathbf{e_i}$, определяется по следующей формуле:

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

где **m** число элементарных исходов $\mathbf{e_i}$, которые благоприятствуют появлению **A**, и **n** – число всевозможных исходов.

Пример. Эксперимент состоит в подбрасывании один раз правильной шестигранной игральной кости. Пусть **x** – число очков, выпавших на верхней грани кости. Описать множество всех элементарных событий, и указать состав подмножеств, соответствующих событиям **A = {x кратно 3}, B = {x нечетно}, C** = {**x** > **3**}. Определить вероятности случайных событий **A, B, C.**

Решение. Множество всех возможных случайных исходов составляют элементарные события, состоящие в выпадении $\mathbf{e_1-1}$, $\mathbf{e_2-2}$, $\mathbf{e_3-3}$, $\mathbf{e_4-4}$, $\mathbf{e_5-5}$, $\mathbf{e_6-6}$. Событиям **A, B, C** соответствуют подмножества из $\{\mathbf{e_i}\}$, в которые входят следующие исходы:

$$\mathbf{A=\{e_3,e_6\}},\qquad \mathbf{B=\{e_1,e_3,e_5\}},\qquad \mathbf{C=\{e_4,e_5,e_6\}}.$$

$$\mathbf{P(A)=2/6},\qquad \mathbf{P(B)=3/6=1/2},\qquad \mathbf{P(C)=3/6=1/2}.$$

Варианты заданий

1. Игральная кость подбрасывается два раза. Найти вероятность события **A** – сумма очков равна **6**, **D** – сумма очков меньше **5**.
2. Подбрасывается две игральных кости. Отмечается число очков на верхних гранях. Что вероятнее: получить число очков, в сумме дающих 7 или получить одинаковое число очков на обеих костях?
3. Подбрасываются две игральные кости, подсчитывается сумма очков на верхних гранях. Что вероятнее: получить в сумме 7 или 8.
4. Монета подбрасывается два раза. Определить вероятность того, что появится не более двух гербов.

5. Монета подбрасывается два раза. Определить вероятность того, что появится хотя бы одна решка.

6. Среди 500 пассажиров метрополитена 240 мужчин. Определить вероятность того, что первым выбранным для интервью пассажиром окажется женщина.

7. В денежно-вещевой лотерее на каждые **1000** билетов **5** денежных и **25** вещевых выигрышей. Какова вероятность выигрыша на **1** билет?

8. В ящике **250** яиц, из которых **20** бракованных. Какова вероятность того, что первое взятое из ящика не окажется бракованным?

9. В группе **17** юношей и **8** девушек. Какова вероятность, что студент, фамилия которого в списке группы находится на 1-ом месте, окажется девушкой.

10. В урне **23** шара: **10** синих, **5** желтых и **8** белых. Что более вероятно извлечение желтого шара или появление **5** очков при бросании игральной кости.

Комбинаторный метод вычисления вероятностей в классической схеме

Предполагается, что производится выбор **m** элементов из **n** различных элементов всех возможных исходов $\mathbf{E = \{e_1, e_2, ..., e_n\}}$. При этом предполагается, что выбор осуществляется без возвращения. Это предполагает, что выбираются сразу все **m** элементов из **n** , или по одному, причем каждый выбранный исключается. В результате имеем различные схемы по выбору **m** элементов из общего числа исходов **E**.

1. Схема выбора, приводящая к сочетаниям

Если опыт состоит в выборе **m** элементов без возвращения и упорядочивания, то различными исходами будут **сочетания из n элементов по m**. Их общее число определяется по формуле

$$\mathbf{C_n^m = \frac{n!}{m!(n-m)!} = \frac{1\cdot 2\cdot 3\cdot \ldots \cdot n}{1\cdot 2\cdot \ldots \cdot m\cdot 1\cdot 2\cdot \ldots \cdot (n-m)}}.$$

Пример. В урне **10** белых и **5** черных шаров. Из урны вынимают **2** шара. Определить вероятность того, что оба шара будут белыми.

Решение. Событие **A** – оба шара будут белыми.

Общее число случаев $\mathbf{C_{10+5}^2 = \frac{15!}{2!13!} = \frac{1\cdot 2\cdot 3\cdot \ldots \cdot 14\cdot 15}{1\cdot 2\cdot 1\cdot 2\cdot \ldots \cdot 12\cdot 13} = 105}$. Число благоприятных случаев

$$\mathbf{C_{10}^2 = \frac{1\cdot 2\cdot \ldots \cdot 9\cdot 10}{1\cdot 2\cdot 1\cdot 2\cdot \ldots \cdot 8} = 45;\quad P(A) = 45/105 = 3/7}.$$

Пример. Ящик содержит **6** деталей, среди которых **2** детали неисправны, **4** детали исправны. **Три** из **шести** деталей выбираются из ящика.

а. Какое количество элементарных исходов мы имеем в данном случае?

б. Какова вероятность, что одна из 3-х деталей будет неисправна?

Решение.

$$\text{а. } n = C_6^3 = \frac{6!}{3!3!} = \frac{1\cdot2\cdot3\cdot4\cdot5\cdot6}{1\cdot2\cdot3\cdot1\cdot2\cdot3} = 20.$$

$$\text{б. } P(A) = \frac{C_2^1 \cdot C_4^2}{C_6^3} = \frac{2\cdot\frac{4!}{2!2!}}{20} = \frac{2\cdot6}{20} = 0.6.$$

Число благоприятных исходов в данном случае определяется как произведение числа комбинаций для выбора **одной** неисправной детали из **двух** и числа комбинаций для выбора **остальных двух** деталей из **трех** исправных.

2. Схема выбора, приводящая к размещениям

Если опыт состоит в выборе **m** элементов без возвращения, но с упорядочиванием их по мере выбора, в последовательную цепочку, то различными исходами будут **m** элементарных подмножеств множества **E,** отличающиеся либо набором элементов, либо порядком их следования. Получаемые при этом комбинации называются **размещениями из n элементов по m**, а их общее число определяется формулой

$$A_n^m = \frac{n!}{(n-m)!} = \frac{1\cdot2\cdot3\cdot\ldots\cdot n}{1\cdot2\cdot\ldots\cdot(n-m)}.$$

Пример. На пяти карточках написаны цифры от **1** до **5**. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятность события **A = {появится число 123}.**

Решение. Общее число случаев $n = A_n^m = A_5^3 = \frac{5!}{2!} = 3\cdot4\cdot5 = 60$.

Благоприятных исходов 1, т.е. **P(A)=1/60.**

В частном случае упорядочивания, когда $n = m$ *имеем* $A_n^m = n!$ или ***число перестановок*** P_n ***из n элементов****, где* $P_n = 1\cdot2\cdot\ldots\cdot n = n!$

Пример. Из урны, содержащей **n** перенумерованных шаров, наугад вынимаются один за другим все находящиеся в ней шары. Найти вероятность того, что номера вынутых шаров будут идти по порядку **1, 2, …, n.**

Решение. Общее число случаев $\mathbf{n = n!}$. Благоприятных случаев $\mathbf{m = 1}$, $\mathbf{P(A) = 1 / n!}$.

Примечание: $\mathbf{0! = 1}$.

Варианты заданий

1. Буквы **а, а, в, к, к, о, х** написаны на отдельных карточках. Какова вероятность того, что, извлекая эти карточки по одной наудачу (без возвращения обратно) мы получим в порядке их выхода слово **«каховка»**?
2. Из полной колоды карт (**52** карты) вынимаются наугад **3** карты (без возврата). Вычислить вероятность того, что среди вынутых карт будет точно **один туз.**
3. Из партии, содержащей **10** изделий, среди которых **3** бракованных наудачу извлекают **3** изделия для контроля. Определить вероятность того, что в полученной выборке нет **ни одного** бракованного.
4. В группе **25** студентов. Вызываются во время занятий **3** студента. Полагая, что вызов производится случайно, определить, какова вероятность того, что будут вызваны **3** студента **А, В, С** в определенном порядке.
5. Студент знает ответы на **15** из **25** вопросов. Какова вероятность того, что из 3 выпавших ему вопросов только два счастливых.
6. В коробке **15** папирос, одинаковых по внешнему виду, но отличающихся сортом табака, а именно: **10** папирос сорта **А** и **5** папирос сорта **Б**. Из коробки берут наудачу **шесть** папирос сразу. Какова вероятность того, что среди этих **шести** папирос **четыре** папиросы сорта **А** и **две** папиросы сорта **Б**.
7. Множество **Е** состоит из **10** первых букв русского алфавита. Опыт состоит в выборе без возвращения **четырех** букв в записи слова. Сколько слов (из **4-х** букв) может быть получено в данном опыте. Какова вероятность, что наудачу составленное слово будет оканчиваться буквой **а**?
8. Ребенок играет с буквами разрезной азбуки **Б, И, И, И, Л, С, Т.** Какова вероятность того, что раскладывая эти буквы в ряд, он получит слово **«ТБИЛИСИ»?**
9. Имеем **10** кандидатов на **3** различные должности. Какова вероятность того, что кандидаты **А, В, С** получат необходимые им должности.
10. Компания имеет **20** работников. **6** из них должны быть выбраны для интервью. Определить вероятность того, что среди них будет **2** женщины, если всего в компании работает **8** женщин.

Классификация событий

Сложным событием называется событие, выражаемое через другие наблюдаемые события в том же эксперименте с помощью допустимых алгебраических операций. Допустимыми алгебраическими операциями над случайными событиями являются сложение, умножение, дополнение, разность.

Суммой двух событий A и B является событие **C**, которое состоит в появлении хотя бы одного из **A** и **B**.

Произведением событий **A** и **B** является событие **C**, состоящее в совместном осуществлении **A** и **B.**

Дополнение $\overline{A} = \Omega - A$ - противоположное событие, где **Ω** - событие, которое происходит всегда, т.е. **P(Ω) = 1,** событие **Ω** называется **достоверным** событием.

Разность A – B - событие **C**, состоящее в том, что событие **A** происходит, а событие **B** не происходит.

Событие ∅ называется **невозможным** и его вероятность **P(∅) = 0.**

События **A** и **B** называются **несовместными**, если появление события **A** исключает появление события **B**.

Операции над событиями помогают упростить вычисление вероятностей событий.

Формула сложения вероятностей **P(A+B)=P(A)+P(B)-P(AB)**, если **A** и **B** совместны, **P(A+B)=P(A)+P(B)**, если **A** и **B** несовместны.

Пример. В урне **12 белых** и **7 черных** шаров. Из урны вынимается **2** шара. Какова вероятность получить **один черный** шар и **один белый** шар?

Решение. Пусть **A** событие – вытащили **один белый** шар и **один черный** шар. Событие **A** произойдет, если произойдут следующие события:

B – в **первый** раз извлечен **белый** шар, во **второй черный** или
C - в **первый** раз извлечен **черный** шар, во **второй белый**.

Событие **A** можно представить как сумму событий **B** и **C**, т.е. **A=B+C**. События **B** и **C** несовместны, значит

$$P(A) = P(B) + P(C);$$
$$P(B) = 12/19\cdot 7/18;$$
$$P(C) = 12/19\cdot 7/18;$$
$$P(A) = 12/19\cdot 7/18 + 12/19\cdot 7/18 = \frac{28}{57}.$$

Условная вероятность. Независимость событий

Условная вероятность это вероятность события **A** при условии, что появилось событие **B.** Условная вероятность события **A** при условии, что произошло событие **B** обозначается как **P(A/B).**

Условная вероятность определяется как отношение вероятности совместного появления событий **A** и **B** к вероятности события **B**: **P(A/B)=P(AB)/P(B).** Аналогично определяется условная вероятность события **B** относительно **A**. Определение условной вероятности выражает принцип умножения вероятностей: вероятность совместного появления двух событий равна вероятности одного из них, умноженной на условную вероятность другого относительно первого :

$$P(AB)=P(A)P(B/A)=P(B)P(A/B).$$

Событие **A** зависит от события **B, если** $P(A/B)\neq P(A)$, т.е, вероятность события **A** изменяется при появлении события **B**. Если **A** не зависит от **B,** то $P(A/B)=P(A)$.

Пример 1. В урне имеется **5 черных и 7 белых** шаров. Опыт состоит в том, что из урны вынимается **один** шар и регистрируется его цвет, после чего шар **возвращается** в урну. После перемешивания снова вынимается **один** шар. Найти вероятность того, что оба раза будет вынут **белый** шар.

Решение. Пусть событие **A** появление **белого** шара в **первый** раз, событие **B** появление **белого** шара во **второй** раз. Как в первый так и во второй раз благоприятных возможностей вынуть белый шар **7**, а общее число возможностей **12.** Следовательно, в данном случае $P(A) = P(B) = 7/12$; $P(A/B) = P(B/A) = 7/12$. Поэтому, события **A** и **B** в данном случае независимы.

Пример 2. В условиях предыдущего примера найти вероятность появления **двух белых** шаров, если **белый** вынутый шар в урну **не возвращается.**

Решение. В этом случае при вынимании **второго** шара имеется **11** случаев, из которых **6** благоприятных при условии, что **первый** шар **- белый.** Следовательно, $P(B/A)=6/11$, если **первый** шар был **черным.** то $P(B/A)=7/12$. Таким образом, появление белого шара во второй раз зависит от того, какой шар был вынут в первый раз, т.е. $P(B/A)\neq P(B)$, события **A** и **B** в данном случае зависимы. $P(AB)=7/12\cdot 6/11=7/22$, если первый шар белый. $P(AB)=7/12\cdot 7/11=49/132$, если первый шар черный.

Варианты заданий

Установить являются ли события зависимыми.

1. Две монеты последовательно бросаются. Рассматриваются события: **A** – выпадение герба на первой монете, **E** – выпадение хотя бы одной цифры. Определить являются ли эти события зависимыми.
2. Из колоды карт (**52** карты) вынимается одна карта. Событие **A** – появление туза, событие **B** – появление карты красной масти. Зависимы ли эти события?
3. Из колоды карт (**52** карты) вынимается одна. Событие **A** – появление туза, событие **C** – появление бубнового туза. Зависимы ли эти события?
4. При последовательном бросании двух монет определить условные и безусловные вероятности для следующих событий: **D** – выпадение хотя бы одного герба, **F** – выпадение герба на второй монете.
5. Из полной колоды карт вынимается одна карта. Рассматриваются события **C** – появление бубнового туза, **B** – появление карты красной масти. Зависимы ли эти события?
6. Брошены **две** игральные кости. **A** – число очков на первой делится на **3**, **B** - число очков на **второй** делится на **3**. Являются ли эти события зависимыми?
7. Брошены **две** игральные кости: **A** – число очков на первой кости делится на **2**, **C** – сумма очков на первой и второй кости делится на **2**. Являются ли эти события **A** и **C** зависимыми?
8. Игральная кость брошена два раза. X_1 и X_2 – числа очков, выпавших при

этих испытаниях. Рассматриваются события: $\mathbf{A_1}$ – $\mathbf{X_1}$ делится на **2**, $\mathbf{X_2}$ делится на **3**, $\mathbf{A_2}$ – $\mathbf{X_1}$ делится на **3**, $\mathbf{X_2}$ делится на **2**. Установить являются ли $\mathbf{A_1}$ и $\mathbf{A_2}$ независимыми.

9. Игральная кость брошена два раза. $\mathbf{X_1}$ и $\mathbf{X_2}$ – числа выпавших очков. Рассматриваются события $\mathbf{A_1}$**:** $\mathbf{X_1}$ делится на $\mathbf{X_2}$; $\mathbf{A_2}$**:** $\mathbf{X_2}$ делится на $\mathbf{X_1}$. Являются ли $\mathbf{A_1}$ и $\mathbf{A_2}$ зависимыми?

10. Игральная кость бросается два раза. $\mathbf{X_1}$ и $\mathbf{X_2}$ – числа выпавших очков. Рассматриваются события $\mathbf{A_1}$**:** $\mathbf{X_1}$ делится на 2; $\mathbf{X_2}$ делится на 3; $\mathbf{A_2}$**:** $\mathbf{X_1}$ делится на $\mathbf{X_2}$. Являются ли эти события зависимыми?

Варианты заданий

1. Вероятность попадания в мишень одного стрелка при одном выстреле для первого стрелка равна **0.8**, для второго стрелка – **0.85**. Стрелки произвели по одному выстрелу в мишень. Считая попадание в цель для отдельных стрелков событиями независимыми, найти вероятность события **A** – ни одного попадания в цель.
2. В условиях задачи варианта 1 определить вероятность **B** – ровно одно попадание в цель.
3. В условиях задачи варианта 1 определить вероятность события **C** – попал первый стрелок.
4. В условиях задачи варианта 1 определить вероятность события **D** – хотя бы одно попадания в цель.
5. В лаборатории работают два компьютера. Вероятность того, что первый компьютер потребует ремонта в течении месяца равна 0.15. Вероятность, что второй компьютер потребует внимания в течении месяца равна 0.2. Определить вероятность того, что оба компьютера не потребуют внимания в течении месяца.
6. В условиях задачи варианта 5 определить вероятность того, что хотя бы один компьютер потребует внимания в течении месяца.
7. Проводится **3** повторных независимых измерения некоторой физической величины. Вероятность того, что при одном измерении (любом) ошибка выйдет за пределы допуска равна **0,1**. Найти вероятность события **A** – во всех измерениях была достигнута точность.
8. В условиях задачи варианта 7 определить вероятность события **B** – не более, чем в одном измерении ошибка выйдет за пределы допуска.
9. В условиях задачи варианта 7 определить вероятность события **C** – по крайней мере в двух измерениях подряд была достигнута заданная точность.
10. Среди товаров, продаваемых супермаркетом **25 %** товаров первого сорта и **65 %** – высшего сорта. Какова вероятность, что первый из двух выбранных товаров окажется товаром первого сорта?

Формула полной вероятности и формула Байеса

Пусть случайное событие **A** может произойти только с одним из событий $\mathbf{H_1, \ldots, H_n}$, попарно несовместных. Тогда вероятность события **A** может быть определена по формуле $\mathbf{P(A)=P(H_1)P(A/H_1)+\ldots+P(H_n)P(A/H_n)}$. События $\mathbf{H_1, \ldots, H_n}$ называются **гипотезами.**

Пример. Имеются три одинаковые урны. В первой **a** белых шаров и **b** черных, во второй – **c** белых и **d** черных; в третьей только белые шары. Некто подходит наугад к одной из урн и вынимает из нее один шар. Найти вероятность того, что этот шар белый.

Решение. $\mathbf{H_1}$ **-** выбор первой урны; $\mathbf{H_2}$ **-** выбор второй урны; $\mathbf{H_3}$**-**выбор третьей урны.

$$\mathbf{P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+P(H_3)P(A/H_3);}$$

$$\mathbf{P(Hl)=P(H2)=P(H3)=l/3.}$$

$$P(A/H_1)=\frac{a}{a+b};\ P(A/H_2)=\frac{c}{c+b}\cdot P(A/H_3)=1;$$

$$P(A)=\frac{1}{3}\cdot\frac{a}{a+b}+\frac{1}{3}\cdot\frac{c}{c+d}+\frac{1}{3}\cdot 1=\frac{1}{3}\cdot\left(\frac{a}{u+b}+\frac{c}{c+d}+1\right).$$

Формула Байеса имеет вид :

$$P(H_i/A)=\frac{P(H_i)P(A/H_i)}{P(H_i)P(A/H_i)+\ldots+P(H_n)P(A/H_n)},\quad i=1,2,\ldots,n.$$

Эта формула для вычисления условной вероятности $\mathbf{P(H_i/A)}$ гипотезы $\mathbf{H_i}$ после испытания, при котором произошло событие **A**. Формула Байеса позволяет переоценить вероятности гипотез до испытания (априорные), по результатам уже проведенного опыта.

Пример. Пусть в предыдущем примере некто вынул шар и он оказался белым. Определить вероятность того, что он вынул его из первой урны.

Решение.

$$P(H_1/A)=\frac{\frac{1}{3}P(A/H_1)}{\frac{1}{3}\sum_{i=1}^{3}P(A/H_i)}=\frac{\frac{a}{a+b}}{\frac{a}{a+b}+\frac{c}{c+d}+1}.$$

Варианты заданий

1. **Три** станка выпускают одинаковые детали. Дневная выработка **первого** станка составляет **6000** изделий, **второго - 1000** изделий, **третьего - 3000** изделий. Детали проверяются с точки зрения одного определенного признака, причем первый станок выпускает **10%** деталей данного свойства, второй

– **8 %,** третий -**15 %.** На складе продукция трех станков смешивается. Какова вероятность выбора из этой суммарной партии детали с определенным свойством ?

2. На предприятии имеется **- три** станка одного типа. Один из них дает **20 %** общей продукции, второй – **30 %,** третий – **50 %,** При этом первый станок производит **5 %** брака, второй **4 %,** третий – **2 %.** Найти вероятность того, что случайно отобранное **негодное** изделие выпущено **первым** станком.

3. В группе **40** стрелков, из них **10** человек стреляют отлично, **20 -** хорошо, **6** -удовлетворительно, **4** - плохо. Вероятность попадания в цель при одном выстреле для отличного стрелка равна **0.9,** для хорошего **– 0.8,** для удовлетворительного **– 0.6,** для плохого **– 0.4.** Вызывают наугад одного из стрелков. Он производит **1** выстрел. Найти вероятность того, что он попал в цель.

4. Пусть при массовом производстве некоторого изделия вероятность того, что оно окажется стандартным равна **0.95.** Для контроля производится проверка стандартности изделия, которая дает положительный результат в **99 %** случаев для стандартных изделий и в **3 %** случаев для нестандартных изделий. Какова вероятность того, что изделие стандартнее, если оно выдержало упрощенную проверку ?

5. В пирамиде установлены **5** винтовок, из которых **3** снабжены оптическим прицелом. Вероятность того, что стрелок поразит мишень при выстреле из винтовки с прицелом равна **0.95,** для винтовки без оптического прицела эта вероятность равна **0.7.** Найти вероятность того, что мишень будет поражена, если стрелок произведет один выстрел из наудачу взятой винтовки.

6. В группе **30** спортсменов : **20** лыжников, **6** конькобежцев и **4** бегуна. Вероятность выполнить норму мастера спорта равна : для лыжника **0.9,** для конькобежца **0.8** и для бегуна **0.75.** Найти вероятность того, что спортсмен, выбранный наудачу, выполнит норму мастера спорта.

7. **Два** автомата производят одинаковые детали, которые сбрасываются на общий конвейер. Производительность **первого** автомата **вдвое** больше производительности **второго.** Первый автомат вырабатывает в среднем **60 %** деталей отличного качества, а второй – **84 %.** Наудачу взятая с конвейера деталь оказалась отличного качества. Найти вероятность того, что эта деталь произведена **первым** автоматом.

8. В группе из **10** студентов, пришедших на экзамен **3** подготовлены отлично, **4** -хорошо, **2** - посредственно и **1** - плохо. В экзаменационных билетах имеется **20** вопросов. Отлично подготовленный студент знает все **20** вопросов, хорошо подготовленный - может ответитъ на **16** вопросов, посредственный **- 10** и плохо подготовленный на **5**. Вызванный наугад студент ответил на **3** произвольно заданных вопроса. Найти вероятность того, что он подготовлен отлично.

9. В условиях задачи 8 определить вероятность, что студент подготовлен плохо.

10. Имеется 2 партии деталей по **12 и 10** штук, причем в каждой партии одно изделие бракованное. Изделие взятое из **первой** партии перекладывается во **вторую,** после чего выбирается изделие **из второй** партии. Определить вероятность извлечения бракованного изделия **из второй** партии.

Случайные величины

Случайной величиной (СВ) называется величина **X,** которая в результате опыта может принимать то или иное значение, причем заранее (до опыта) неизвестное. В теоретико вероятностной схеме случайной величиной называют числовую функцию от элементарных исходов $\mathbf{X = X(e_i)}$. Следовательно, можно говорить о числовой функции $\mathbf{X(e_i)}$, определенной на пространстве элементарных событий. Значения случайных величин обозначаются **x, y, z, ...**

Случайные величины могут быть дискретными и непрерывными.

Дискретной случайной величиной называется величина, которая принимает конечное или бесконечное счетное множество значений.

Непрерывные случайные величины – это величины, значения которых непрерывно заполняют некоторый числовой отрезок.

В качестве примеров можно рассмотреть следующие.

1. Предположим эксперимент состоит в том, что необходимо определить число автомобилей, проезжающих по магистрали за **30** сек. Возможные значения случайной величины **0, 1, 2, …, n**.

2. Предположим, что нужно измерить время завершения двух операций на конвейере. Возможные значения, которые принимает случайная величина от **0** секунд до **n** секунд.

В первом случае имеем дискретную случайную величину, во втором - непрерывную.

Случайная величина считается заданной, если задан ее закон распределения. **Законом распределения случайной величины** называется всякое соотношение между значениями, принимаемыми случайной величиной и вероятностями, с которыми она принимает эти значения. Закон распределения позволяет полностью описать случайную величину с вероятностной точки зрения.

Для **дискретных** случайных величин закон распределения задается в виде **ряда распределения** и **функции распределения**.

Для **непрерывных** случайных величин закон распределения задается в виде **функции распределения** и **плотности распределения**, которые будут рассмотрены ниже.

Для **дискретной** СВ закон распределения может быть задан рядом распределения, который представляется таблицей соответствия значений СВ и соответствующих им вероятностей. Ряд распределения может быть также задан графиком, когда по оси **x** откладываются значения случайной величины, а по **y** значения вероятностей; ломаная линия, соединяющая точки $(\mathbf{x_i}, \mathbf{p_i})$ называется **многоугольником распределения**. Кроме того закон распределения СВ может быть задан аналитически в виде функции распределения.

Функцией распределения называется вероятность того, что случайная величина **X** примет значение, меньшее произвольно выбранного значения **x**, то есть $\mathbf{X \in [-\infty; x]}$;

$$\mathbf{F(x) = P(X < x) = P}.$$

Для дискретных случайных величин функция распределения **F(x)** вычисляется по формуле $\mathbf{F(x) = \sum_{x_i < x} P(X < x_i)}$, где суммирование ведется по тем значениям **i,** для которых $\mathbf{x_i < x, i = 1,2,...}$. Свойства функции распределения приводятся далее (см. функция распределения для непрерывных случайных величин).

Наиболее часто встречающиеся законы распределения дискретных случайных величин

1. Гипергеометрическое распределение $P(x=k)=C_M^k C_{N-M}^{n-k} / C_N^n$, где **N, M, n** – натуральные числа: $M \le N, n \le N, k = 0,1,\ldots,\min(n,M)$.

Пример. Из партии, содержащей **M** белых и **N-M** черных, наудачу извлекаются **n** шаров. Какова вероятность того, что среди выбранных **n** шаров окажется ровно **k** белых?

В этом примере СВ, равная числу шаров, имеет гипергеометрическое распределение $P_n(k)=C_M^k C_{N-M}^{n-k} / C_N^n$, где $k=k_0, k_0+1,\ldots,k_1$, $k_0=\max(0, M-N+n)$, $k_1=\min(n,M)$.

2. Распределение СВ, равное числу успехов в **n** испытаниях, определяется по биноминальному закону распределения:

$$P(x=k)=P_n(k)=C_n^k p^k q^{n-k}, \qquad (1)$$

где **n** – число испытаний, **k** – число успехов, **p** – вероятность успеха в одном испытании, $q=1-p$ - вероятность неудачи.

Формула (1) носит название формулы Бернулли.

3. Для бесконечной последовательности испытаний Бернулли СВ **X**, равная числу испытаний до первого успеха включительно, имеет геометрическое распределение

$$P(x=k)=(1-p)^{k-1}p,$$

где **p** – вероятность успеха **k = 1, 2, …** .

4. Если вероятность появления события в последовательности испытаний Бернулли мала, а **n** велико, то применение формулы Бернулли затруднительно, в этом случае пользуются ее предельным значением $P_n(k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$, где $\lambda = np$. Это формула Пуассона и распределение называется Пуассоновским.

Пример 1. Из партии, содержащей **100** изделий, среди которых **10** дефектных, выбраны случайным образом **5** изделий для проверки их качества. Построить ряд распределения случайного числа X дефектных изделий, содержащихся в выборке.

Решение. Случайная величина **X** распределена по гипергеометрическому закону, то есть $P(X=k)=\dfrac{C_{10}^k C_9^{5-k}}{C_{100}^5}$, **k** – число бракованных изделий, **k = 0, 1, 2, 3, 4, 5**. Ряд распределения представляется следующей таблицей:

x_i	0	1	2	3	4	5
p_i	0.583	0.340	0.07	0.007	0	0

Пример 2. Вероятность того, что некоторая деталь окажется дефектной, равна **p**. В случае обнаружения дефекта линию останавливают и делают переналадку. Составить ряд распределения для числа **X** годных деталей между двумя переналадками.

Решение. Случайная величина **X** принимает значения **0, 1, 2, …, k, …** . **X = k**, когда **(k+1)**-ая деталь окажется дефектной. Следовательно

$$P_k = P(X=k) = q^k \cdot p \quad (k=0,1,2,\ldots)$$

и ряд распределения представится в виде

x_i	0	1	2	…	k	…
p_i	p	qp	q^2p	…	q^kp	…

Плотность распределения. Свойства функции распределения

Как уже указывалось, закон распределения для непрерывной СВ задается в виде функции распределения **F(x)** и плотности распределения **f(x)=F′(x).** График плотности распределения называется кривой распределения. Функция распределения может быть представлена

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Основные свойства функции распределения **F(x):**

1. $F(-\infty) = 0;$
2. $F(+\infty) = 1;$
3. **F(x) – неубывающая;**
4. $P(a \le x \le b) = F(b) - F(a).$

Основные свойства плотности распределения **f(x):**

1. $f(x) \ge 0;$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1;$
3. $P(a \le x \le b) = \int_{a}^{b} f(x)dx.$

Прежде, чем ввести наиболее встречающиеся законы распределения непрерывных СВ, рассмотрим основные числовые характеристики СВ.

Числовые характеристики случайных величин

Математическое ожидание СВ **X** есть число, обозначаемое **M(x)**, а также m_x и μ, и определяемое по формуле

$$M(x) = \begin{cases} \sum_i x_i p_i, & \text{для дискретных СВ X,} \\ \int_{-\infty}^{+\infty} xf(x)dx, & \text{для непрерывных СВ X,} \end{cases}$$

где $p_i = P(X = x_i)$, **f(x)** – плотность распределения.

Предполагается, что $\sum_i x_i p_i$ и $\int_{-\infty}^{+\infty} xf(x)dx$ абсолютно сходятся, в противном случае считают, что **M(x)** не существует. **M(x)** является характеристикой положения СВ **X**.

Дисперсией D(x) СВ X называют математическое ожидание квадрата отклонения СВ **X** от ее математического ожидания

$$D(x) = M(x - m_x)^2,$$

$$D(x) = \begin{cases} \sum_i (x - m_x)^2 p_i, \text{ для дискретных СВ X}, \\ \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx, \text{ для непрерывных СВ X}. \end{cases}$$

Дисперсия является характеристикой рассеяния СВ X. Кроме нее характеристикой рассеяния является $\sigma = \sqrt{D(x)}$, называемая **среднеквадратическим отклонением**.

Пример. Производится ряд независимых испытаний, в каждом из которых может появиться некоторое событие **A**. Вероятность **A** в каждом опыте равна **p**. Опыты производятся до первого появления **A**, после чего они прекращаются. Случайная величина **X** – число произвольных опытов. Построить ряд распределений этой случайной величины и найти математическое ожидание и дисперсию.

Решение. СВ **X** имеет геометрическое распределение $P(x = k) = (1 - p)^{k-1} p$, **q = 1-p, k** – число испытаний до первого успеха.

x_i	0	1	2	3	4	…	i	…
p_i	p	qp	q2p	q3p	q4p	…	qip	…

$$M(x) = \sum_{i=1}^{\infty} x_i p_i = 0 \cdot p + 1qp + 2q^2p + 3q^3p + \ldots + iq^ip + nq^np = p\sum_{i=1}^{\infty} iq^{i-1},$$

$$D(x) = \sum_{i=1}^{\infty} (x_i - m_x)^2 p_i.$$

На практике удобней бывает использование формулы $D(x) = M[x^2] - M^2(x)$, в нашем случае

$$M[x^2] = \sum_{i=1}^{\infty} x_i^2 p_i = \sum_{i=1}^{\infty} i^2 q^{i-1} p_i = p\sum_{i=1}^{\infty} i^2 q^{i-1},$$

$$D[x] = p\sum_{i=1}^{\infty} i^2 q^{i-1} - p(\sum_{i=1}^{\infty} iq^{i-1})^2.$$

Основные законы распределения непрерывных СВ

Среди различных распределений особое место занимает нормальное или гауссовское распределение, которое задается плотностью распределения вида

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma^2}} \qquad (-\infty < x < +\infty)$$

и функцией распределения вида

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m_x)^2}{2\sigma^2}} dx \qquad (-\infty < x < +\infty),$$

где m_x – математическое ожидание, σ - среднеквадратическое отклонение.

m_x и σ являются параметрами нормального закона распределения, которое обозначается $N(m_x, \sigma)$. Для каждого конкретного значения **M(x)** и **σ** имеем свое нормальное распределение.

Кривая нормального распределения (кривая Гаусса) симметрична относительно прямой $x = m_x$ и имеет единственный максимум в точке $x = m_x$. Кривая нормального распределения приведена на рис. 1. Ось абсцисс является асимптотой для этой кривой.

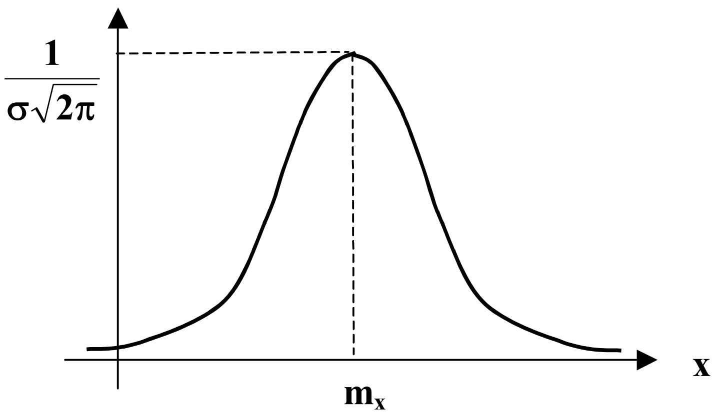


Рис. 1. График нормального распределения **N(m$_x$,σ)**

Нормальное распределение с параметрами $m_x = 0$ и $\sigma = 1$ называется стандартным или нормированным и обозначается **N(0,1)**, его плотность распределения определяется формулой

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (-\infty < x < +\infty).$$

Функция распределения стандартного (нормированного) нормального закона распределения имеет вид:

$$\Phi_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

Значения этой функции протабулированы и приводятся в таблицах (см. приложение).

Функцию распределения нормального закона с параметрами m_x и σ можно выразить через $\Phi_{N(0,1)}(x)$ следующим образом

$$F(x) = \Phi_{N(0,1)}\left(\frac{x - m_x}{\sigma}\right).$$

Вероятность попадания случайной величины на отрезок $[x_1, x_2]$ определяется в виде:

$$P(x_1 < x \le x_2) = F(x_2) - F(x_1) = \Phi_{N(0,1)}\left(\frac{x_2 - m_x}{\sigma}\right) - \Phi_{N(0,1)}\left(\frac{x_1 - m_x}{\sigma}\right).$$

Функция $\Phi_{N(0,1)}(x)$ имеет свойство

$$\Phi_{N(0,1)}(-x) = 1 - \Phi_{N(0,1)}(x).$$

Часто на практике необходимо определить значение x_p из уравнения $\Phi(x_p) = p$. Т.е. необходимо определить значение аргумента по заданному значению функции $\Phi(x_p)$, которое задается в таблице стандартного нормального распределения (смотри таблицу II приложения). Как видно из таблицы, значения функции $\Phi(x_p)$ задаются начиная со значения равного 0.5. В случае, когда необходимо определить x_p из уравнения $\Phi(x_p) = p$, при **p < 0.5,** то можно воспользоваться указанным выше свойством. Значение x_p называется **квантилью уровня р.**

Пример 1. Пусть необходимо определить x_p, из условия $\Phi_{N(0,1)}(x_p) = 0.05$.

Решение. По определению функции распределения нужно найти x_p из условия $P(x < x_p) = 0.05$. Так как значение **0.05** < **0.5**, переходим к $P(x > x_p) = 1 - P(x < x_p) = 1 - 0.05 = 0.95$.
$1 - P(x < x_p) = \Phi_{N(0,1)}(-x_p) = 0.95$, из таблицы $x_p = -1.64$. Квантили нормального закона связаны $-x_p = -x_{1-p}$.

Часто в таблицах приложений приводится не значение функции стандартного нормального закона распределения **N(0,1),** а интеграл вероятностей или функция Лапласа, определяемая в виде $\Phi(x) = \frac{2}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ или $\Phi_1(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$, где $\Phi_1(x) = 1/2\Phi(x)$. Для функции Лапласа справедливо $\Phi(-x) = -\Phi(x)$ и $\Phi_1(-x) = -\Phi_1(x)$.

Функция распределения нормального стандартного закона распределения связана с функцией Лапласа следующими соотношениями:

$$\Phi_{N(0,1)}(x) = 1/2(\Phi(x)+1) \quad \text{или}$$

$$\Phi_{N(0,1)}(x) = 1/2(2\Phi_1(x)+1).$$

Пример 2. Определить вероятность того, что случайная величина **X** примет значение в интервале **(158, 162),** если известно, что она распределена по нормальному закону с параметрами $m_x = 168,\ \sigma = 5.92$.

Решение. Так как $P(\alpha < x < \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - m_x}{\sigma}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma}\right)\right]$. Вычисляем $t_1 = (158-168)/5.92 = -1.69$; $t_2 = (162-168)/5.92 = -1.014$. По таблице функции Лапласа $\Phi(x)$ имеем

$$\Phi(t_2) = \Phi(-1.014) = -\Phi(1.014) = -0.6894;$$

$$\Phi(t_1) = -\Phi(1.690) = -0.9090.$$

Следовательно $P(158 \le x \le 162) = (1/2)[-0.6894 + 0.9090] = 0.1106$. По таблице 2 приложения функции распределения стандартного нормального закона распределения имеем

$$\Phi_{N(0,1)}(t_2) = \Phi_{N(0,1)}(-1.014) = 1 - \Phi_{N(0,1)}(1.014) = 1 - 0.84375 = 0.15625,$$

$$\Phi_{N(0,1)}(t_1) = \Phi_{N(0,1)}(-1.69) = 1 - \Phi_{N(0,1)}(1.69) = 1 - 0.95449 = 0.0455,$$

$$\Phi_{N(0,1)}(t_2) - \Phi_{N(0,1)}(t_1) = 0.15625 - 0.0455 = 0.11075.$$

Вероятность того, что абсолютная величина отклонения СВ от своего математического ожидания меньше любого $\varepsilon > 0$ определяется формулой $P(|x - m_x| < \varepsilon) = 2\Phi_{N(0,1)}(\varepsilon/\sigma) - 1$.

Варианты заданий

1. Производится стрельба по цели, вероятность попадания при каждом выстреле равна **0,6.** Найти закон распределения СВ **X,** равный числу попаданий по цели при двух выстрелах, если число попаданий при втором выстреле не зависит от числа попаданий при первом. Определить математическое ожидание числа попаданий.
2. В денежной лотерее выпущено **100** билетов. Разыгрывается **один** выигрыш в **10** т.руб., **четыре** в **5** т.руб., **пять** в **4** т.руб. и **десять** в **1** т.руб, Найти закон распределения, математическое ожидание и дисперсию стоимости выигрыша.
3. Случайная величина **X** число попаданий мячом в корзину при броске, причем вероятность попадания равна **0,3.** Построить ряд распределения, функцию распределения и определить математическое ожидание **X**.
4. Функция распределения СВ **X** имеет вид :

$$F(x) = \begin{cases} 0, & \text{при } x < 0 \\ x^2/26, & \text{при } 0 \le x \le 5 \\ 1, & \text{при } x > 5. \end{cases}$$

Найти вероятность того, что случайная величина окажется в интервале **(3,6).** (см. свойства F(x)).

5. Случайная величина **X** имеет закон распределения

$X = x_i$	-1	0	1
$P(x = x_i)$	¼	½	¼

Написать выражение для функции распределения СВ **X.**

6. Из урны, в которой лежат **2** белых и **8** черных шаров, вынимают **3** шара. Найти распределение вероятностей числа **X** вынутых белых шаров.

7. Случайная величина **X** распределена по нормальному закону с **M(x)=3** и $\sigma = 0.5$. Определить вероятность того, что ее значения будут отклоняться от **M(x)** по абсолютной величине не более, чем на **1.3.**

8. На автомате изготовляются заклепки. Диаметр их головок представляет собой случайную величину, распределенную по нормальному закону и имеет **M(x)** = **2** мм и дисперсию равную **0.01** мм^2. Какие размеры диаметра головок заклепки можно гарантировать с вероятностью **0.95.**

9. Производится взвешивание некоторого вещества без систематических ошибок. Случайные ошибки взвешивания подчинены нормальному закону со средним квадратическим отклонением **σ = 20** г. Найти вероятность того, что взвешивание будет произведено с ошибкой, не превосходящей по абсолютной величине **10 г.**

10. Производятся многократные испытания некоторого элемента на надежность до тех пор, пока элемент не откажет. Найти: математическое ожидание дискретной случайной величины **X -** числа опытов, которые надо произвести. Вероятность отказа элемента в каждом опыте равна **0.1.**

Для случайной величины **X**, распределенной по нормальному закону с заданными m_x и σ определить вероятность попадания в интервал **[a,b].** Значения m_x, σ и границы интервала **a** и **b** см. в вариантах заданий.

Варианты заданий

1. $m_x = 5.5$, σ = 1.2, **a** = 2, **b** = 7;
2. $m_x = 144$, σ = 4.8, **a** = 138, **b** = 150;
3. $m_x = 12$, σ = 3.1, **a** = 9, **b** = 14;
4. $m_x = 17.1$, σ = 2.4, **a** = 16, **b** = 19;
5. $m_x = 138.1$, σ = 5.9, **a** = 135, **b** = 140;
6. $m_x = 14.2$, σ = 1.5, **a** = 11, **b** = 15;

7. $\mathbf{m_x} = 125.1$, $\sigma = 4.9$, $\mathbf{a} = 120$, $\mathbf{b} = 127$;

8. $\mathbf{m_x} = 14.4$, $\sigma = 5.8$, $\mathbf{a} = 13$, $\mathbf{b} = 16$;

9. $\mathbf{m_x} = 126.1$, $\sigma = 14.1$, $\mathbf{a} = 110$, $\mathbf{b} = 130$;

10. $\mathbf{m_x} = 10.5$, $\sigma = 3.1$, $\mathbf{a} = 9$, $\mathbf{b} = 12$.

4.3. Контрольные задания по математической статистике

1. Представить исходную выборку в виде статистического ряда и изобразить его графически. Привести график эмпирической функции распределения.
2. Определить моду и медиану.
3. Определить точечные оценки для среднего арифметического, дисперсии, среднеквадратического отклонения.
4. Определить квартили Q_1, Q_2, Q_3.
5. Установить, является ли распределение симметричным, используя коэффициент асимметрии и графический способ Box and Whisker Plot (смотри [8], стр 90 – 91.)
6. Определить интервальные оценки для математического ожидания с уровнями значимости $\alpha = 0,05$ и $\alpha = 0,01$.

Варианты заданий

Построить интервальный ряд и выполнить пункты задания по математической статистике (см. стр. 10, раздел 4.2.2).

№ 1	50.0	61.7	72.1	80.9	90.9	51.2	61.9	73.4
	81.3	91.3	52.7	62.8	74.5	82.4	92.5	63.7
	76.7	82.9	64.0	77.7	83.7	66.1	64.1	78.1
	65.8	65.2	79.8	66.1	68.5	66.8	67.4	70.1
№ 2	30.0	41.7	52.1	60.9	70.9	31.2	41.9	53.4
	45.8	45.2	59.8	46.1	48.5	46.8	47.4	50.1
	61.3	71.3	32.7	42.8	54.5	62.4	72.5	43.7
	56.7	62.9	44.0	57.7	63.7	46.1	44.1	58.1
№ 3	7.34	6.19	5.12	9.09	8.09	7.21	6.17	5.00
	8.13	9.13	6.37	9.25	8.24	7.45	6.28	5.27
	6.58	6.52	7.98	6.61	6.85	6.68	6.74	7.01
	7.67	8.29	6.40	7.77	8.37	6.61	6.41	7.81
№ 4	153.4	141.9	131.2	170.9	160.9	152.1	141.7	130.0
	145.8	150.1	147.4	146.8	148.5	146.1	159.8	145.2
	161.3	171.3	132.7	142.8	154.5	162.4	172.7	143.7

	156.7	158.1	144.1	146.1	163.7	157.7	144.0	162.9			
№ 5	123.4	121.9	101.2	140.9	130.9	122.1	111.7	100.0			
	126.7	128.1	114.1	116.1	133.7	127.7	114.0	132.9			
	115.8	120.1	117.4	116.8	118.5	116.1	129.8	115.2			
	131.3	141.3	102.7	112.8	124.5	132.4	142.5	113.7			
№ 6	8.8	7.5	4.3	5.4	6.7	3.8	4.6	6.7	5.2	6.5	5.5
	5.8	6.5	4.7	7.4	6.2	4.2	5.8	7.2	4.7		
№ 7	75.41	38.97	108.54	37.18	62.75	118.32	88.84	127.74			
	55.42	39.03	29.41	47.99	28.73	84.05	66.12	88.22			
	47.81	89.01	84.92	97.36							
№ 8	15.62	5.29	16.25	10.92	11.46	21.62	8.45	8.58	5.41	11.42	
	11.62	7.29	17.50	17.96	14.42	10.5	9.29	7.54	18.92		
№ 9	9.54	11.46	16.62	12.62	25.75	15.41	14.29	13.13	13.71	10.04	
	5.75	12.46	19.17	13.21	16.00	12.23	14.25	15.37	16.25	19.71	
№ 10	7.55	3.75	0.1	1.1	0.6	0.52	3.30	2.10	0.58	4.02	3.75
	0.65	1.92	0.6	1.53	4.23	0.08	1.48	1.65	0.72		

4.3.1. Методические указания к выполнению контрольных заданий

Основы выборочного метода. Представление результатов наблюдений

Математическая статистика (МС) для своих выводов использует наблюдения за случайным явлением с целью получения заключения об источнике наблюдений. При этом выбирается интересующее исследователя свойство (например, производительность труда, здоровье людей и т.д.) и определяются переменные **x, y, …,** характеризующие это свойство (например, число деталей за смену, показатели сердечно-сосудистой системы и т.д.). Затем случайным образом выбираются объекты из источника наблюдений, на которых измеряются выбранные показатели. Результаты наблюдений обозначаются малыми латинскими буквами **x, y, z, …** в отличие от переменных, характеризующих свойство, которые обозначаются большими латинскими буквами **X, Y, Z, …** .

Таким образом, полный набор всех значений **N**, которое принимает случайная величина **X**, называется **генеральной совокупностью**, где число **N** может быть конечным и бесконечным. Часть генеральной совокупности из **n** элементов, отобранных случайным образом, называется **выборкой**. Число **n** называется **объемом выборки**. Различают выборки большого и малого объема. Выборки с $n \le 20$ называются **малыми**, выборки $n > 20$ называют **большими.** Это является условным и зависит от решаемой задачи. Выборка называется **репрезентативной**, если она дает достаточное представление об особенностях генеральной совокупности.

Статистические методы представляют собой методы для получения выводов о генеральной совокупности через выборку.

Методы, касающиеся сбора данных называются **описательной статистикой**.

Методы, касающиеся заключений о генеральной совокупности называются **статистическими выводами**.

Поскольку конечная цель статистических исследований - получение заключений о генеральной совокупности по выборке с определенной вероятностью, то методы описательной статистики могут рассматриваться как методы первичной обработки результатов наблюдений и представления их в виде удобном для дальнейшего анализа.

Чтобы правильно обработать исходный ряд наблюдений, необходимо выяснить какого типа количественную случайную переменную предстоит исследовать.

В зависимости от типа множества значений, которые может принимать случайная переменная, выделяют два типа: **дискретный** и **непрерывный**.

Случайная переменная **дискретного** типа может принимать лишь изолированные значения. Случайная переменная **непрерывного** типа – любое значение из некоторого интервала.

К **непрерывным** относятся переменные, значения которых получены в результате измерений (иногда они могут быть округлены и представлены в виде целых чисел). Такие переменные как вес, рост, температура, расстояние, время и т.д. являются непрерывными.

Дискретными являются переменные, значения которых определяются в виде некоторых подсчетов.

Первоначально выборки представляют в виде таблицы, состоящей из двух строк, в первой даны номера измерений, во второй – их результаты.

Таблица 1

i – номер измерений	**1**	**2**	**…**	**N**
результат измерений	$\mathbf{x_1}$	$\mathbf{x_2}$	**…**	$\mathbf{x_n}$

Таблица такого вида называется **простым статистическим рядом**. Далее этот ряд преобразуют в **вариационный ряд**, где все наблюдения представляются в порядке возрастания, т.е. в виде:

$$\mathbf{x_{1\,min}, x_2, \ldots, x_{n\,max}},$$

где $\mathbf{x_1 \le x_2 \le \ldots \le x_{nmax}}$.

На следующем этапе данный вариационный ряд представляется в виде статистического ряда. **Статистический ряд** для дискретной переменной это сгруппированный или частотный ряд вида:

Таблица 2

Значение переменной x	**Частота**	**Частость (относит.**	**Накопленная частота**	**Накопленная частость**

		Частота)		
	m_i	m_i / n	$\sum\limits_{x_i<x} m_i$	$\sum\limits_{x_i<x} m_i / n$
	$\sum m_i = n$	$\sum \frac{m_i}{n} = 1$		

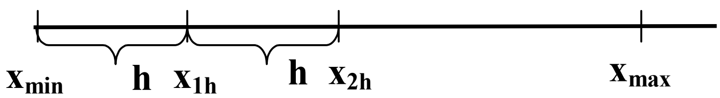
Для **непрерывных** вариационных рядов результаты наблюдений представляются в виде **интервального** ряда.

Таблица 3

Класс границ	Частота m_i	Средняя точка класса x_i	Частости (относит. частота)	Накопленные частоты	Накопленные частости
$x_i,\ x_{i+k}$	m_i	$\bar{x}_{i_k}$	m_i / n	$\sum\limits_{x_i<x} m_i$	$\sum\limits_{x_i<x} m_i / n$
	$\sum\limits_i m_i = n$		$\sum\limits_i m_i / n = 1$		

Табл. 1 и 2 дают полный вид статистических рядов (они могут представляться и в усеченном виде (в зависимости от решаемой задачи).

Класс границ для интервального ряда можно изобразить на числовой оси.



Из рисунка видно, что граничные точки классов отстоят друг от друга на одну и ту же величину, равную шагу **h**, который равен $h = (x_{max} - x_{min}) / k$, где **k** - число классов. Чтобы определить величину этого шага, необходимо установить на какое количество классов **k** разбивается данный ряд наблюдений. Это рекомендуется сделать в соответствие со следующими формулами: $k = 1 + 3 \cdot 322 \lg n$ или $k \le 5 \lg n$, где **n** – число наблюдений (объем выборки). В литературе предлагается также выбирать число классов в зависимости от объема выборки. Для малых выборок $k = 5 \div 7$, для больших $k = 10 \div 20$. Выбор числа классов является важным моментом. При слишком малом **k** гистограмма не будет отражать особенностей распределения, при слишком большом **k** гистограмма будет излишне изрезанной. Значения **h** и **k** обычно округляются до ближайшего целого. За начало первого интервала берется точка x_{min} или $x_{min} - h/2$.

Средняя точка класса определяется в виде $\bar{x}_{i_k} = x_{i_h} + h/2$.

Частота m_i представляет собой количество наблюдений, соответствующих данному наблюдению для дискретной переменной для сгруппированного ряда или число наблюдений, попавших в данный интервал для интервального ряда. Значение соответствующей частоты, деленной на объем выборки характеризует **частость** попадания $\mathbf{x_i}$ в частичные интервалы.

Закон больших чисел в форме Бернулли утверждает, что если эксперимент повторяется **n** раз при одинаковых условиях, то частость $\mathbf{m_i / n}$ сходится по вероятности к $\mathbf{p_i}$, то есть $\frac{\mathbf{m_i}}{\mathbf{n}} \xrightarrow{p} \mathbf{p_i}$. Следовательно, $\mathbf{m_i / n}$ является приближенными значениями вероятности $\mathbf{p_i}$. В отличие от теоретического закона распределения СВ **X**, для выборки определяется эмпирический закон распределения.

Сгруппированный и интервальный ряды представляют эмпирическое распределение частот. Для наглядного представления эмпирических распределений **сгруппированного** ряда строится график, где по оси X откладываются значения переменной, а по оси У значения частот (частостей). Полученный точки соединяют ломаной линией. Этот график называется **полигоном**.

Интервальный ряд графически представляется в виде **гистограммы**. Гистограмма представляет собой ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы $[\mathbf{x_i}, \mathbf{x_{i+k}}]$, а их высоты равны $\mathbf{m_i / n}$ или $\mathbf{m_i}$, **h** – длина интервала $[\mathbf{x_i}, \mathbf{x_{i+k}}]$ или шаг. В первом случае площадь гистограммы равна **1**, во втором случае -объему выборки **n**. Результаты в столбцах «накопленные частоты» и «накопленные частости» используются для определения эмпирической функции распределения, с помощью которой также задается частотное распределение.

Эмпирическая функция распределения определяется по значениям **накопленных частостей** из следующих соотношений:

$$\mathbf{F_n^*(x) = 1/n \sum_{x_i < x} m_i},$$

где суммируются частости тех элементов выборки, для которых выполняется неравенство $\mathbf{x_i < x}$ (**x** – некоторое значение). Из приведенных формул следует, что

$$\mathbf{F_n^*(x) = 0}, \quad \text{если} \quad \mathbf{x \le x_1},$$

$$\mathbf{F_n^*(x) = 1}, \quad \text{если} \quad \mathbf{x > x_n}.$$

На промежутке $\mathbf{(x_1, x_n)}$ $\mathbf{F_n^*(x)}$ представляет собой неубывающую кусочно-постоянную функцию. Согласно теореме Гливенко эмпирическая функция распределения $\mathbf{F_n^*(x)}$ является хорошей оценкой генеральной функции распределения **F(x)** при $\mathbf{n \to \infty}$.

Пример. Получить статистические ряды для следующих наблюдений и представить графически их частотные распределения.

а) производится наблюдение над величиной урожая пшеницы на 10 опытных участках, результаты представлены в табл. 4.

Таблица 4

№ уч	1	2	3	4	5	6	7	8	9	10
X	15.2	19.1	17.1	20.8	18.4	16.4	22.0	20.4	17.6	18.2

в) имеем выборку 21 пар обуви

39, 41, 40, 41, 44, 40, 42, 41, 43, 39, 42, 41, 42, 38, 42, 41, 43, 41, 39, 40, 40.

Данные выборки представляют собой **непрерывный ряд** наблюдений - случай а) и **дискретный ряд** наблюдений – случай в). Поэтому для случая а) нужно построить статистический ряд в виде интервального ряда, в случае в) в виде сгруппированного ряда.

Решение.

Случай **а.**

Представляем данный простой статистический ряд в виде вариационного ряда. Имеем:

15.2, 16.4, 17.1, 17.6, 18.2, 18.4, 19.1, 20.4, 20.8, 22.0.

Объем выборки **n = 10**.

Для представления вариационного ряда в виде интервального ряда необходимо определить ширину интервала **h.**

$$h = \frac{x_{max} - x_{min}}{k}, \quad k \text{ – число классов.}$$

Число классов **k** определяем по формуле

$$k = 1 + 3.322\lg n = 1 + 3.322\lg 10 = 4.322.$$

В результате округления получаем **k = 4.**

$$h = \frac{22 - 15.2}{4} = 1.75 \approx 2.$$

Интервальный ряд представлен табл. 5.

Таблица 5

Границы классов	Частоты m_i	Средние точки классов	Частости m_i / n	Накоплен. частоты	Накоплен. частости
15-17	2	16	0.2	2	0.2
17-19	4	18	0.4	6	0.6
19-21	3	20	0.3	9	0.9
21-23	1	22	0.1	10	1
	$\sum m_i = 10$		$\sum m_i / n = 1$		

Накопленные частоты определяются по формуле $\sum\limits_{x_i < x} m_i$, то есть суммируются, те значения частот, которые соответствуют соотношению $x_i < x$. При заполнении данного столбца на первое место записывается частота, соответствующая первому интервалу. В данном случае это $m_1 = 2$. Далее суммируются последовательно частоты первого и второго интервалов и т.д. В результате имеем значение **6**, далее **9** и **10**.

Столбец “накопленные частости” определяется делением соответствующих накопленных частот на **10**, то есть на значение, равное общему объему выборки.

Графическое представление полученного интервального ряда (гистограмма) представлена на рис. 2.

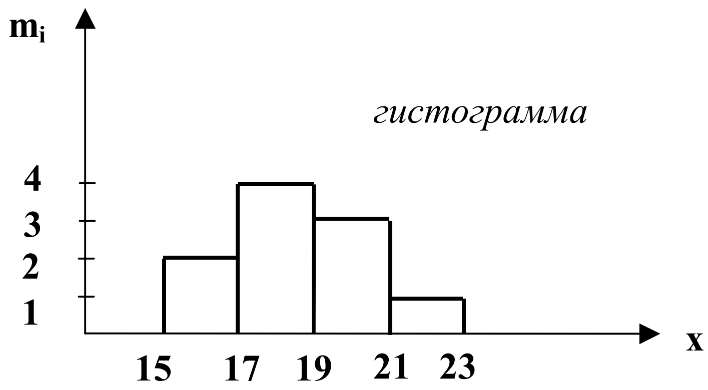


Рис. 2. Графическое представление интервального ряда

На рис. 3 представлено графическое изображение эмпирической функции распределения (кумулята). При построении кумуляты по оси X откладываются начальная точка и верхние точки классов, а по оси У значение накопленных частостей. Полученные точки соединяются кривой.

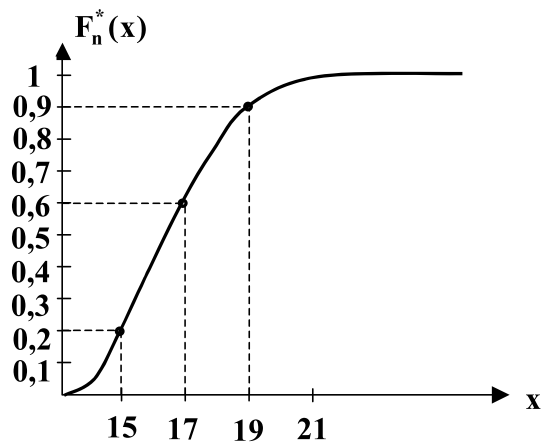


Рис. 3. Графическое представление эмпирической функции распределения $\mathbf{F_n^*(x)}$ для случайных переменных непрерывного типа

Гистограмма и кумулята представляют собой графики эмпирических частотных распределений.

Случай **б.** Вариационный ряд имеет вид:

38, 39, 39, 39, 40, 40, 40, 40, 40, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 43, 43, 44.

Сгруппированный ряд представлен табл. 6.

Таблица 6

Размеры	**Частоты**	**Частости**	**Накоплен.**	**Накоплен.**

обуви			**частоты**	**частости**
38	**1**	**1/21**	**1**	**1/21**
39	**3**	**3/21**	**4**	**4/21**
40	**4**	**4/21**	**8**	**8/21**
41	**6**	**6/21**	**14**	**14/21**
42	**4**	**4/21**	**18**	**18/21**
43	**2**	**2/21**	**20**	**20/21**
44	**1**	**1/21**	**21**	**1**
	$\sum_i m_i = 21$	$\sum_i \frac{m_i}{n} = 1$		

Для графического изображения частотных распределений необходимо построить полигон и кумуляту.

Для построения полигона по оси X откладываются возможные значения переменной, по оси У частоты (частости).

В данном случае полигон имеет вид

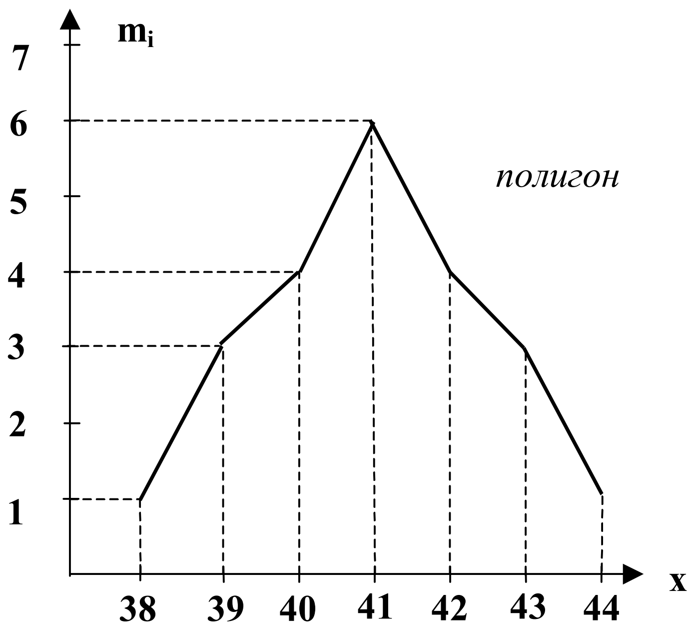


Рис. 4. Графическое представление сгруппированного ряда

Эмпирическая функция распределения изображается **кумулятой,** которая представляет собой неубывающую ступенчатую функцию, скачки которой происходят в точках, соответствующих возможным значениям переменной и равны значению частости в этой точке. По оси X откладываются значения переменной, по оси У соответствующие накопленные частости.

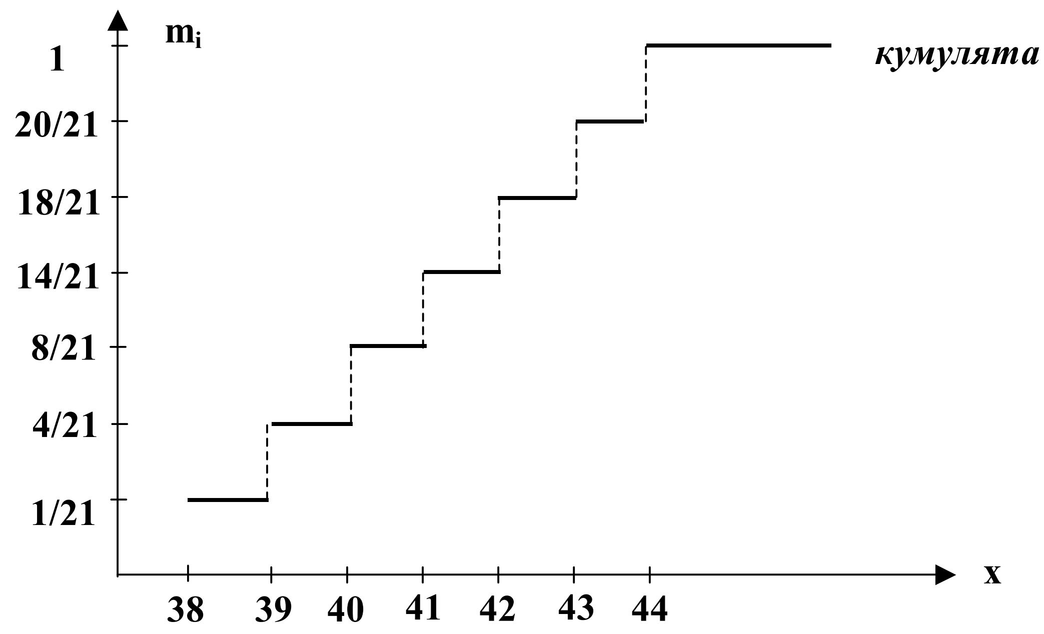


Рис. 5. Графическое представление эмпирической функции распределения $F_n^*(x)$ для случайной величины дискретного типа

На практике эмпирическая функция распределения используется, когда при принятии решения необходимо знать текущий итог.

Числовые характеристики выборочных распределений

Эмпирические распределения, представляемые для дискретных вариационных рядов в виде полигона и кумуляты, а для непрерывных – в виде гистограммы и кумуляты, позволяют предположить вид генерального распределения.. Кроме этого на практике часто используются числовые характеристики для описания ряда наблюдений. Данные характеристики позволяют проанализировать ряд наблюдений другим способом. Они включают меры расположения, меры рассеивания и меры, характеризующие форму распределения.

Вычисление этих характеристик зависит от того как представлены результаты наблюдений: в виде статистических рядов или в виде простого статистического ряда (исходной выборки). Как было уже сказано, статистический ряд является такой формой представления, когда данные группируются и определяются частоты. В дальнейшем будем называть статистический ряд сгруппированным рядом. Простой статистический ряд не является сгруппированным.

Меры расположения

Меры расположения дают информацию о среднем значении, наиболее часто встречающемся значении и о значении, являющимся серединой данных измерений и т.д. Эти меры не дают информации о том, как разбросаны наблюдения относительно среднего числа. К ним относятся **мода, медиана, среднее арифметическое, процентили и квартили.**

Меры расположения в несгруппированном ряду наблюдений

Мода – наиболее часто встречающееся значение в ряду наблюдений. Для определения моды ряд наблюдений необходимо представить в виде вариационного ряда и определить наиболее часто встречающееся наблюдение. Ряд наблюдений может иметь два и более таких значений. В первом случае ряд называется **бимодальным**, во втором - **мультимодальным**.

Медиана – величина, соответствующая середине в вариационном ряду (то есть число наблюдений меньше медианы равно числу наблюдений больше медианы). Чтобы определить медиану необходимо выполнить следующее:

- представить исходный ряд наблюдений в виде вариационного;
- если объем выборки число нечетное, то медиана это серединный элемент вариационного ряда, номер которого определяется в виде $(\mathbf{n}+\mathbf{1})/\mathbf{2}$;
- если объем выборки число четное, то медиана равна среднему значению двух серединных элементов вариационного ряда.

Среднее арифметическое значение, обозначаемое $\overline{\mathbf{X}}$ (вместо **X** может быть любая буква латинского алфавита) определяется в виде:

$$\overline{\mathbf{X}}=\frac{\mathbf{1}}{\mathbf{n}}(\mathbf{x}_1+\mathbf{x}_2+\ldots+\mathbf{x}_n)=\frac{\mathbf{1}}{\mathbf{n}}\sum_{i=1}^{n}\mathbf{x}_i\ .$$

Мода и медиана называются **структурными средними**, так как они связаны с рядом частот, то есть со структурой распределения. Следует отметить, что на значение среднего арифметического оказывают влияние все значения выборки. Медиана не подвержена такому влиянию.

Пример 1. Для данного вариационного ряда, представляющего собой объем акций 10 наиболее действующих предприятий (в тысячах) определить моду, медиану и среднее арифметическое

651 681 789 827 903 903 904 922 979 1514.

Решение.

Мода: наиболее часто встречающееся значение **903** миллиона.

Медиана: так как объем выборки n = 10, то значение $\frac{\mathbf{903}+\mathbf{903}}{\mathbf{2}}=\mathbf{903}$ является медианой.

Среднее арифметическое:

$$\overline{\mathbf{X}}=\frac{\mathbf{1}}{\mathbf{10}}(\mathbf{651}+\mathbf{681}+\mathbf{789}+\mathbf{827}+\mathbf{903}+\mathbf{903}+\mathbf{904}+\mathbf{922}+\mathbf{979}+\mathbf{1514})=\mathbf{907.3}$$

Процентили являются мерами расположения, которые делят наблюдения на 100 частей. Можно определить **99** процентилей. **Р-ый** процентиль – это такая величина, что **Р %** данных являются меньше этой величины и **(100-Р) %** являются больше.

Процентили широко используются в различного рода отчетах. Для того, чтобы определить **Р-ый** процентиль необходимо выполнить следующее:

- представить результаты наблюдений в виде вариационного ряда;
- вычислить номер Р-го процентиля в вариационном ряду

$$\mathbf{i}=\frac{\mathbf{P}}{\mathbf{100}}(\mathbf{n}),$$

где **P** – значения процентиля,
n – объем выборки,
i – номер процентиля в ряду наблюдений.

- определить значение Р-го процентиля:

а) если **i** – целое, то **Р-ый** процентиль является средней величиной **i – го** и **(i+1) – го** наблюдений в вариационном ряду.

в) если **i** не является целым числом, то номер **Р-го** процентиля определяется как целая часть от значения **(i+1)**.

Пример 2. Определить **30-ый** процентиль для следующего ряда наблюдений:

14 12 19 23 5 13 28 17.

Решение. Вариационный ряд:

5 12 13 14 17 19 23 28,

$$\mathbf{i} = \frac{30}{100} \cdot 8 = 2.4\,.$$

Так как **i** не является целым числом, то номер **30-го** процентиля в данном вариационном ряду определяется как целая часть от значения **2.4+1=3.4**, то есть **3.** Следовательно **30-ым** процентилем является значение $\mathbf{x_3 = 13}$.

Квартили это меры расположения, которые делят ряд наблюдений на **4** части. Для этого необходимо **3** квартиля, которые обозначаются $\mathbf{Q_1, Q_2, Q_3}$.

$\mathbf{Q_1}$ является **25-ым** процентилем, то есть $\mathbf{Q_1 = P_{25}}$.

$\mathbf{Q_2}$ является **50-ым** процентилем, то есть $\mathbf{Q_2 = P_{50}}$.

$\mathbf{Q_3}$ - это **75-ый** процентиль, то есть $\mathbf{Q_3 = P_{75}}$.

Пример 3. Определить $\mathbf{Q_1, Q_2, Q_3}$ для следующей выборки:

109 121 122 129 106 116 125 114.

Решение. Вариационный ряд:

106 109 114 116 121 122 125 129.

Так как $\mathbf{Q_1 = P_{25}}$, то определяем 25-ый процентиль. Для **n = 8**

$$\mathbf{i} = \frac{25}{100} \cdot 8 = 2\,.$$

Так как **i** – целое число, то $\mathbf{P_{25}}$ определяется как среднее второго и третьего значений в вариационном ряду.

$$\mathbf{P_{25}} = \frac{109+114}{2} = 111.5\,.$$

$\mathbf{Q_1 = 111.5}$, $\mathbf{Q_2 = P_{50}}$ и является медианой. Так как **n** – четное, то

$$\mathbf{Q_2} = \frac{116+121}{2} = 118.5\,,\ \mathbf{Q_3 = P_{75}},$$

$$\mathbf{i} = \frac{75}{100} \cdot 8 = 6\,,\ \mathbf{P_{75}} = \frac{122+125}{2} = 123.5\,,\ \mathbf{Q_3 = 123.5}.$$

Меры рассеивания

Меры расположения дают информацию об особых точках в ряду наблюдений.

Меры рассеивания характеризуют разброс или дисперсию в выборке.

Меры расположения вместе с мерами рассеивания дают более полное числовое описание наблюдений.

Одними из основных показателей, характеризующих рассеивание в выборке, являются **выборочная дисперсия** и **среднеквадратическое отклонение.** Для несгруппированных данных выборочная дисперсия определяется в виде:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2,$$

где **n** – объем выборки,

$\overline{X}$ - среднее арифметическое,

x_i - значение **i – го** наблюдения.

Среднеквадратическое отклонение

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

Следует отметить, что среднее арифметическое $\overline{X}$, дисперсия S^2 и среднее квадратическое отклонение **S**, определяемые по выборке, являются оценками соответственно для матожидания **M(x),** дисперсии **D(x)** и среднеквадратического отклонения σ, являющихся параметрами генеральной совокупности. При этом оценки должны удовлетворять определенным свойствам (эффективности, состоятельности и несмещенности см. [1, 2]), поэтому в формуле для выборочной дисперсии (для выполнения условия несмещенности) вводится сомножитель $\frac{1}{n-1}$.

Вычислительные формулы

В практических задачах кроме представленных выше могут использоваться следующие вычислительные формулы для выборочной дисперсии и стандартного отклонения.

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}\right), \qquad S = \sqrt{S^2}.$$

Если среднее арифметическое задано, то

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n(\overline{X})^2\right).$$

Смысл среднеквадратического отклонения (стандартного отклонения) может быть установлен по его применению. Два наиболее используемых способа этого применения: это **эмпирическое правило** и **теорема Чебышева**.

Эмпирическое правило справедливо для наблюдений с нормальным законом распределения и утверждает, что

68% данных лежит внутри интервала $\overline{X} \pm 1S$.

95% данных лежит внутри интервала $\overline{X} \pm 2S$.

99.7% данных лежит внутри интервала $\overline{X} \pm 3S$.

Теорема Чебышева применяется для всех распределений. Она утверждает, что внутри **k** стандартных отклонений от среднего находится $(1 - 1/k^2)$ значений выборки.

Пример. Определить внутри какого интервала лежит по крайней мере 85 % всех величин, если известно $\overline{X} = 16.85; \quad S = 4.7$.

Решение.

$$0.85 = 1 - 1/k^2; \quad 0.15 = 1/k^2; \quad k^2 = 6.667; \quad k = 2.58.$$

Полученный интервал определяется как $\overline{X} \pm 2.58 \cdot 4.7$, то есть границами являются **4.72** и **28.98**.

Формулы для среднего арифметического дисперсии и среднеквадратического отклонения

Для **сгруппированных данных** указанные числовые характеристики определяются в виде:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{X})^2 m_i, \qquad S = \sqrt{S^2},$$

где x_i – средние точки классов; **k** – число классов.

Вычислительные формулы для сгруппированных данных

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} m_i x_i^2 - \frac{\left(\sum m_i x_i\right)^2}{n}\right), \qquad S = \sqrt{S^2},$$

где **n** – объем выборки,
k – число классов,
m_i - частоты,
x_i - средние точки в классе.

Структурные средние для сгруппированных данных

Для определения моды M_0 и медианы M_e рассмотрим фрагмент гистограммы, соответствующий интервалу с наибольшей частотой и двумя соседними интервалами.

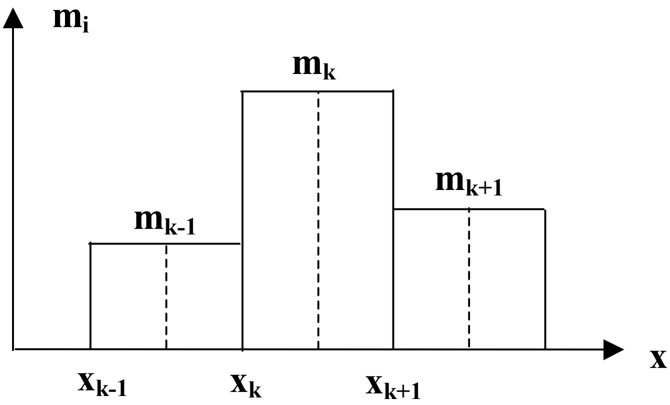


Рис. 6. Фрагмент гистограммы для определения моды $\mathbf{M_0}$

$$\mathbf{M_0 = x_k + \frac{m_k - m_{k-1}}{2m_k - (m_{k-1} + m_{k+1})} \cdot \Delta x},$$

где $\mathbf{\Delta x = x_{k+1} - x_k = h}$.

Медиана определяется по формуле

$$\mathbf{M_e = x_{l+1} + \frac{n/2 - F_l(x)}{m_{l+1}} \Delta x},$$

где $\mathbf{F_l(x) = \sum_{x_i < x} m_i}$ выбирается из условия $\mathbf{F_l(x) \le n/2}$ и $\mathbf{F_{l+1}(x) > n/2}$,

где $\mathbf{n}$ – объем выборки, $\mathbf{\Delta x = h}$,

$\mathbf{m_{l+1}}$ - частота интервала, $\mathbf{x_{l+1}}$ - нижняя граница интервала, для которого выполняется $\mathbf{F_l(x) \le n/2}$ и $\mathbf{F_{l+1}(x) > n/2}$ (см. интервальную таблицу).

Пример. Вычислить среднее арифметическое, медиану, моду и среднеквадратическое (стандартное) отклонение для следующего интервального ряда, представленного табл. 7.

Таблица 7

Класс границ	Частоты	Средние точки класса	Накопленные частоты
10-15	6	12.5	6
15-20	22	17.5	28
20-25	35	22.5	63
25-30	29	27.5	92
30-35	16	32.5	108
35-40	8	37.5	116
40-45	4	42.5	120
45-50	2	47.5	122

$$n = \sum m_i = 122 .$$

Решение.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \frac{1}{122}\sum_{i=1}^{8} x_i m_i = \frac{3130}{122} = 25.66.$$

Медиана

$$M_e = x_{l+1} + \frac{n/2 - F_l^*(x)}{m_{l+1}}\Delta x = 20 + \frac{122/2 - 28}{35}\cdot 5 =$$

$$= 20 + 0.943\cdot 5 = 20 + 4.714 = 24.714.$$

Мода

$$M_0 = 20 + \frac{(35-22)\cdot 5}{70-(22+29)} = 20 + \frac{13\cdot 5}{70-51} = 23.42.$$

Дисперсия и стандартное отклонение могут быть определены как будет показано далее. Вначале для определения дисперсии используем исходную формулу $S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{X})^2 m_i$. Вычисления представлены в табл. 8.

Таблица 8

Класс границ	Частоты	Средняя точка класса x_i	$x_i - \overline{X}$	$(x_i - \overline{X})^2$	$m_i(x_i - \overline{X})^2$
10-15	6	12.5	-13.16	173.19	1039.14
15-20	22	17.5	-8.16	66.59	1464.98
20-25	35	22.5	-3.16	9.99	349.65
25-30	29	27.5	1.84	3.39	98.31
30-35	16	32.5	6.84	46.79	748.64
35-40	8	37.5	11.84	140.19	1121.52
40-45	4	42.5	16.84	283.59	1134.36
45-50	2	47.5	21.84	476.99	953.98

$$\sum m_i = 122$$

$$\sum m_i(x_i - \overline{X})^2 = 6910.58$$

$$S^2 = \frac{\sum (x_i - \overline{X})^2 m_i}{n-1} = \frac{6910.58}{121} = 57.11, \quad S = \sqrt{57.11} = 7.56.$$

Определим дисперсию, используя вычислительную формулу. Вычисления представлены в табл. 9.

Таблица 9

Класс границ	Частоты	Средняя точка класса x_i	$m_i x_i$	mx_i^2
10-15	6	12.5	75.0	937.50
15-20	22	17.5	385.0	6737.50
20-25	35	22.5	787.0	17718.75

25-30	29	27.5	797.0	21931.25
30-35	16	32.5	520.0	16900.00
35-40	8	37.5	300.0	11250.00
40-45	4	42.5	170.0	7225.00
45-50	2	47.5	95.0	4512.50

$$\sum m_i = 122 \qquad \sum m_i x_i = 3130.0$$

$$\sum m_i x_i^2 = 87212.50.$$

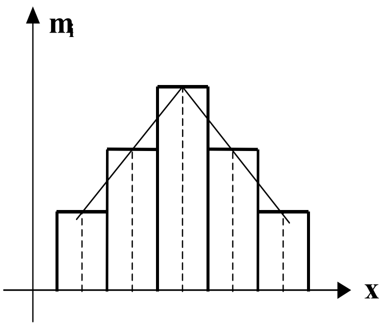
$$S^2 = \frac{\sum m_i x_i^2 - \dfrac{\left(\sum m_i x_i\right)^2}{n}}{n-1} = \frac{87212.50 - (3130.0)^2 / 122}{121} = \frac{6910.04}{121} = 57.11,$$

$$S = \sqrt{57.11} = 7.56.$$

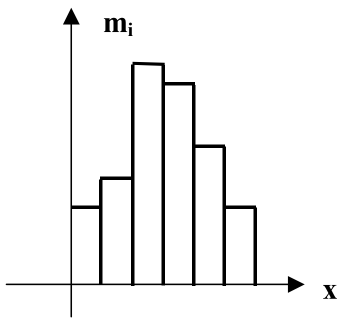
Меры формы

К мерам формы относятся числовые характеристики, позволяющие сделать вывод о форме частотных распределений выборочных данных.

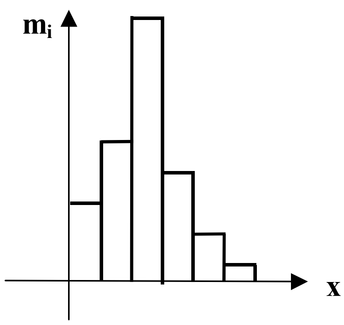
Рассмотрим числовые характеристики, которые позволяют оценить симметричность распределения. Если относительно наибольшей частоты остальные частоты для соответствующих интервалов приблизительно равны, то такое распределение частот будет изображено гистограммой или полигоном, имеющими **симметричную** форму (случай а) на рис. 6. Случаи в) и с) приводятся соответственно левосторонняя и правосторонняя эмпирические асимметричные распределения.



а) Симметричное распределение частот



в) Левая асимметрия



с) Правая асимметрия

Рис. 7

Коэффициент асимметрии

В данном разделе приводится одна из формул для определения коэффициента асимметрии, который используется для определения симметричности распределения.

$$\mathbf{A_x} = \frac{\mathbf{3(\overline{X} - M_e)}}{\mathbf{S}},$$

где $\mathbf{A_x}$ – коэффициент асимметрии,

$\mathbf{S}$ – среднеквадратическое отклонение,

$\mathbf{M_e}$ – медиана,

$\mathbf{\overline{X}}$ - среднее арифметическое.

Если значение этого коэффициента имеет положительное значение, то распределение имеет правостороннюю (положительную) асимметрию. Если значение $\mathbf{A_x}$ отрицательно, то распределение имеет левостороннюю (отрицательную асимметрию. Если $\mathbf{A_x} \approx 0$, то распределение симметрично. Для симметричных распределений выполняется условие: $\mathbf{\overline{X} \approx M_0 \approx M_e}$.

Методы статистических выводов

Рассмотрим один из методов статистических выводов, часто используемый на практике, метод оценки параметров генеральной совокупности. Другие методы статистических выводов такие как методы проверки гипотез и оценки взаимосвязи показателей можно найти в [1,2].

Методы оценки параметров генеральной совокупности делятся на точечные и интервальные.

Точечными оценками параметров генеральной совокупности являются выборочные оценки этих параметров, которые рассматривались нами ранее. Точечные оценки параметров распределения, полученных при выборке, не дают информацию о степени близости их значений теоретическому значению оцениваемого параметра. В этом случае нельзя ответить на вопрос, какую ошибку мы совершаем, принимая вместо теоретического значения параметра его оценку. В связи с этим во многих случаях следует определять не точечную оценку теоретического параметра, а **интервальную оценку**, которая основана на определении некоторого интервала, внутри которого с определенной вероятностью находится истинное значение параметра. то есть следует определить интервал, который с вероятностью **P** охватывает оцениваемое значение параметра.

Одним из наиболее важных для практики генеральных параметров является математическое ожидание или генеральное среднее. Например, управляющий крупной компанией хочет установить, какое количество дней в среднем сотрудники его фирмы пропустили по болезни. Если в фирме тысячи работников, то напрямую получить это число сложно. Используя статистические методы можно определить по выборке среднее арифметическое пропущенных дней и взять эту среднюю арифметическую в качестве оценки генерального среднего и сделать вывод о всей генеральной совокупности по полученному среднему значению. Среднее арифметическое будет являться точечной оценкой генерального среднего, но как точно оценивает оно это генеральное среднее, мы не знаем.

Интервальные оценки дают ответ на этот вопрос.

Если среднее значение, найденное по результатам выборки, является точечной оценкой математического ожидания **M(x),** то вероятность того, что действительное значение оцениваемой величины $\mathbf{M(x)}$ лежит в пределах $(\overline{\mathbf{X}} - \Delta < \mathbf{M(x)} < \overline{\mathbf{X}} + \Delta)$ определяется в виде:

$$\mathbf{P}(\overline{\mathbf{X}} - \Delta < \mathbf{M(x)} < \overline{\mathbf{X}} + \Delta) = \mathbf{1} - \alpha,$$

где $\mathbf{P} = \mathbf{1} - \alpha$ - доверительная вероятность,

α - уровень значимости,

Δ - точность оценки.

Выбор доверительной вероятности определяется конкретными условиями задачи. Обычно используются значения **P = 0.9, 0.95, 0.99.**

Для данных значений доверительной вероятности уровень значимости соответственно равен **0.1, 0.05, 0.01.** Понимать доверительную вероятность следует следующим образом. Если из генеральной совокупности делать выборки, а по ним определять оценки истинного параметра, то в α **%** случаев эти оценки не будут принадлежать указанному интервалу, который называется доверительным.

На практике важную роль играет длина доверительного интервала. Следует отметить, что чем меньше длина доверительного интервала, тем точнее оценка. Если длина доверительного интервала велика, то оценка мало пригодна для практики. Длина доверительного интервала связана с объемом выборки и доверительной вероятностью, что видно из формул, приводимых ниже.

Интервальное оценивание математического ожидания случайной величины

1. Исходим из следующих предположений:

- генеральная выборка распределена по нормальному закону;
- выборка бесповторная;
- значение генеральной дисперсии **D(x)** известно.

В этом случае

$$\mathbf{P}\left(\overline{\mathbf{X}} - \mathbf{Z}_{\alpha/2}\frac{\sigma}{\sqrt{\mathbf{n}}} < \mathbf{M(x)} < \overline{\mathbf{X}} + \mathbf{Z}_{\alpha/2}\frac{\sigma}{\sqrt{\mathbf{n}}}\right) = \mathbf{1} - \alpha,$$

где $\overline{\mathbf{X}}$ - выборочное среднее,

$\mathbf{Z}_{\alpha/2}$ - квантиль нормального распределения или такое значение (см. таблицу функции распределения $\mathbf{\Phi(x)}$ стандартизированного нормального распределения **N(0,1))** при котором $\Phi(\mathbf{Z}_{\alpha/2}) = \mathbf{1} - \frac{\alpha}{\mathbf{2}}$, $\varepsilon = \frac{\sigma}{\sqrt{\mathbf{n}}}$ - средняя квадратическая ошибка, $\Delta = \mathbf{Z}_{\alpha/2}\frac{\sigma}{\sqrt{\mathbf{n}}}$ - точность оценивания.

Для наиболее употребительных значений доверительной вероятности **1-**α **= 0.9, 0.95, 0.99, 0.9973** квантили стандартизованного нормального распределения $\mathbf{Z}_{\alpha}$ соответственно равны : $\mathbf{Z}_{\alpha/2}$ **= 1.64, 1.96, 2.58, 3, 3.37.**

2. Исходим из следующих предположений :

- генеральная совокупность распределена по нормальному закону;
- выборка бесповоротная;
- значение D(x) неизвестно, а вместо него используется ее оценка, определяемая по выборке в виде:

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2}\,.$$

В этом случае

$$P\left(\overline{X}-t_{\alpha,k}\frac{S}{\sqrt{n}}<M(x)<\overline{X}+t_{\alpha,k}\frac{S}{\sqrt{n}}\right)=1-2\alpha,$$

где $t_{\alpha,k}$ - значение квантиля статистики t, распределенной по закону Стьюдента;

k – число степеней свободы или число элементов выборки минус число связей, которые наложены на элементы выборки **k = n – 1**;

α - уровень значимости.

Пример 1. Необходимо установить в течение скольких лет компании данной страны ведут торговлю с Индией. Случайно выбранные **44** фирмы показали, что ведут торговлю в среднем **10.455** лет. Предположим, что известно генеральное среднеквадратическое отклонение и оно равно 7.7 лет. Построить **90 %** доверительный интервал для генерального среднего.

Решение. Имеем **n** = **44**; $\overline{X}=10.55$; **σ** = **7.7**. Так как известно генеральное среднеквадратическое отклонение **σ**, то доверительный интервал определяется по формуле

$$\overline{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le M(x)\le\overline{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

$P=1-\alpha=0.9;\quad \alpha/2=0.05$. По таблице функции распределения стандартизованного нормального распределения определяем $Z_{\alpha/2}=-1.645$. Доверительный интервал вычисляется как

$$10.455-1.645\cdot\frac{7.7}{\sqrt{44}}\le M(x)\le 10.455+1.645\frac{7.7}{\sqrt{44}},$$

$$10.455-1.91\le M(x)\le 10.455+1.91,\ 8.545\le M(x)\le 12.365,$$

$$P(8.545\le M(x)\le 12.365)=0.9\,.$$

Точечная оценка генерального среднего составляет **10.455** лет. Интервальная оценка указывает, что генеральное среднее находится между **8.545** и **12.365**. Точность оценивания **1.91.**

Пример 2. На основании обработки данных выборки из 145 значений получено $\overline{X}=31.40$; **S** = **3.26**. Определить доверительный интервал для генеральной средней с **P = 0.99**.

Решение. Доверительный интервал вычисляется по формуле

$$\overline{X}-t_{\alpha,k}\frac{S}{\sqrt{n}}\le M(x)\le\overline{X}+t_{\alpha,k}\frac{S}{\sqrt{n}}\,.$$

Определяем $\mathbf{t_{\alpha,k} = t_{0.01,144}}$ по таблице распределения Стьюдента (см. таблицу в приложении). $\mathbf{t_{0.01,144} = 2.58}$. Границы доверительного интервала

$$\mathbf{31.40 - 2.58\frac{3.26}{\sqrt{145}} \le M(x) \le 31.40 + 2.58\frac{3.26}{\sqrt{145}}},$$

$$\mathbf{31.40 - 0.70 \le M(x) \le 31.40 + 0.70}, \quad \mathbf{30.70 \le M(x) \le 32.10}.$$

Генеральное среднее будет находиться в интервале **(30.70, 32.10).**