**Вопросы**

1. Предметная область математической статистики. Особенности, задачи. Сравнительный анализ. Примеры использования методов.
2. Классификация погрешностей измерения.
3. Метод наименьших квадратов: постановка задачи, сущность метода, область применения.
4. Метод наименьших квадратов и его матричное представление.
5. Случайные величины. Дискретные случайные величины.
6. Случайные величины. Непрерывные случайные величины.
7. Генеральная и выборочная совокупности
8. Вариационный ряд: определение, графическое представление.
9. Алгоритм построения интервального вариационного ряда
10. Алгоритм построения дискретного вариационного ряда
11. Числовые характеристики вариационных рядов.
12. Законы распределения случайных величин.
13. Проверка статистических гипотез.
14. Анализ временных рядов.
15. Корреляция. Корреляционный анализ.
16. Дисперсия. Эксцесс. Показатель ассиметрии.
17. Числовые характеристики случайных величин.
18. Алгоритм вычисления числовых характеристик выборочного распределения.
19. Вычисление оценок по методу моментов.
20. Квантили, квартили, перцентили.

**Решения**

# 1. Предметная область математической статистики. Особенности, задачи. Сравнительный анализ. Примеры использования методов.

1)Решать задачи принятия решений в условиях вероятностной неопределенности

2)Принятие решений обычно преследует одну из целей: прогнозирование

будущего состояния процесса (объекта); управление (т. е. как следует

изменять одни параметры объекта (процесса), чтобы другие параметры

приняли желаемое значение); объяснение внутренней структуры объекта

(процесса).

3)Задачи:

1) организация наблюдений;

2) нахождение по результатам выборочных наблюдений оценок числовых

характеристик всей совокупности и исследование точности их

приближения (выборочный метод);

3) решение вопроса согласования результатов оценивания с опытными

данными (проверка статистических гипотез);

4) оценка существенности влияния факторных признаков на

результативный (дисперсионный анализ);

5) выявление аналитической зависимости между наблюдениями

факторных и результативных признаков (корреляционно-

регрессионный анализ).

4)Особенности?:По существу математическая статистика дает единственный, математически обоснованный аппарат для решения задач управления и прогнозирования при отсутствии явных закономерностей (наличии случайностей) в изучаемых процессах.

5)Методы математической статистики можно разделить на описательные

(дескриптивные) и аналитические

Описательные методы позволяют описать

реальные наблюдения с помощью таблиц, графиков, характеристик

положения (среднее арифметическое, мода, медиана), характеристик

рассеяния (среднее линейное отклонение, среднее квадратическое

отклонение, дисперсия, коэффициент вариации) и т. д.

Аналитические методы позволяют на основании выборочных наблюдений

сделать статистически значимые выводы о наличии закономерностей для

всей совокупности. Аналитические методы обычно основываются на

соответствующих вероятностных моделях, предполагающих нормальное

(или другое известное) распределение совокупности изучаемого признака –

методы параметрической статистики.

Основная цель математической статистики – это получение и обработка

данных для статистически значимой поддержки процесса принятия решения,

например, при решении задачи планирования, управления, прогнозирования.

.

Примеры:

1)рейтинг политиков

2)обоснованность рейтинга популярности

3)зависит ли производительность труда рабочего от стажа

4)какая зависимость существует между спросом на продукцию и курсом доллара

5)количество пятен на Солнце

6)контроль качества выпускаемой продукции.

Ссылка на лекцию: <https://drive.google.com/file/d/0B2aH_ii1oppXQnBrSlFqTzQxc0E/view>

# 2. Классификация погрешностей измерения.

Существуют различные виды классификации ошибок измерения.

1. по источникам возникновения погрешности измерений подразделяются на:

инструментальные;

внешние или погрешности среды;

личные.

Эта классификация имеет большое значение для специальных дисциплин, изучающих измерительные приборы и методы измерения.

2. По точности результаты измерений разделяются на:

равноточные и неравноточные.

Под равноточными измерениями понимают однородные результаты,

полученные при измерениях одним и тем же инструментом. То есть в этом случае принимается, что та группа основных факторов, которая поддаѐтся учѐту и оказывает существенное влияние на результаты измерения, сохраняет, в определѐнном смысле(более или менее), одинаковые значения.

Неравноточными называются результаты измерений, если

указанные условия не соблюдаются.

По закономерностям появления погрешности измерений подразделяются на:

грубые погрешности;

систематические;

случайные.

1)Грубой погрешностью называют погрешность, существенно превышающую ожидаемую при данных условиях измерения. Грубые погрешности или промахи возникают, как правило, из-за нарушения основных условий проведения экспериментов

2)Систематической погрешностью называют составляющую погрешности измерения, остающуюся постоянной или закономерно изменяющуюся при повторных измерениях одной и той же величины. Систематические погрешности обусловлены такими факторами, которые остаются постоянными в процессе измерения.

3) Случайной погрешностью называют составляющую погрешность измерения, изменяющуюся случайным образом при повторных измерениях одной и той же величины. Случайные погрешности вызываются большим количеством случайных факторов, однако эффект действия таких факторов бывает столь незначительным,

что его нельзя выделить и учесть в отдельности, поэтому случайную погрешность следует рассмотреть как суммарный эффект действия некоторых факторов.

Ссылка на лекцию: <https://drive.google.com/file/d/0B2aH_ii1oppXSThpRmRCc3l5TEk/view>

# 3. Метод наименьших квадратов: постановка задачи, сущность метода, область применения

М.Н.К - математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. Он может использоваться для «решения» переопределенных систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределенных) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции.

Пусть x - набор n неизвестных переменных (параметров), fi(x), i = 1, …, m, m > n, — совокупность функций от этого набора переменных. Задача заключается в подборе таких значений x, чтобы значения этих функций были максимально близки к некоторым значениям yi. По существу речь идет о «решении» переопределенной системы уравнений

fi(x) = yi, i=1, .., m, в указанном смысле максимальной близости левой и правой частей системы. Сущность МНК заключается в выборе в качестве «меры близости» суммы квадратов отклонений левых и правых частей abs(fi(x)-yi). Таким образом, сущность МНК может быть выражена следующим образом:



В случае, если система уравнений имеет решение, то наименьшее значение суммы квадратов будет равно нулю и могут быть найдены точные решения системы уравнений аналитически или, например, различными численными методами оптимизации. Если система переопределена, то есть, говоря нестрого, количество независимых уравнений больше количества искомых переменных, то система не имеет точного решения и метод наименьших квадратов позволяет найти некоторый «оптимальный» вектор x в смысле максимальной близости векторов y и f(x) или максимальной близости вектора отклонений e к нулю (близость понимается в смысле евклидова расстояния).

Ссылка( в лекциях не нашел): [Вики](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%BD%D0%B0%D0%B8%D0%BC%D0%B5%D0%BD%D1%8C%D1%88%D0%B8%D1%85_%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%BE%D0%B2)

# 4. Метод наименьших квадратов и его матричное представление.

Не найдено

# 5.Случайные величины. Дискретные случайные величины.

Случайная величина (СВ) – это величина, которая в результате опыта может принимать те или иные значения, причем до опыта мы не можем сказать, какое именно значение она примет.

СВ обозначаются буквами латинского алфавита X, Y, Z.

СВ могут быть трех типов:

-дискретные

-непрерывные

-смешанные

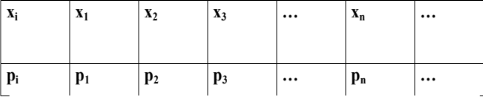
Дискретная случайная величина (ДСВ) может принимать конечное или бесконечное счетное число значений.

Например, подбрасываем монету 5 раз. Случайная величина X- число появлений герба: 0,1,2,3,4,5.

Пусть X - дискретная случайная величина, которая принимает значения: x1,

x2, ..., хn... с некоторой вероятностью pi ,где i = 1,2,..., n,...

Значения хi, и соответствующие рi, представляют в виде таблицы:



Эта таблица является одной из форм задания ДСВ. Обычно случайные величины располагаются в возрастающем порядке. Основное свойство таблицы заключено в том, что сумма вероятностей равна 1:

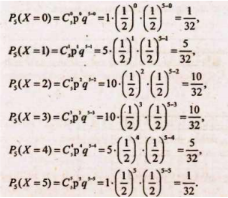
Задача-пример:

---------------------

Монета бросается 5 раз. Представим закон распределения ДСВ Х- числа появлений герба, в виде таблицы.

ДСВ X может принимать значения: 0, 1, 2, 3, 4, 5. Вероятность появления герба в одном опыте р=1/2, непоявления q=1/2, n = 5. Таким образом, выполняются условия применения формулы Бернулли.

Имеем:



С – это сочетания, проходили когда-то, формула-напоминание:



Полученные данные можно представить в виде таблицы распределения:



---------------------

Дискретная случайная величина может быть представлена в виде многоугольника распределения - фигуры, состоящей из точек (xi, pi), соединенных отрезками (рис. 8).

Над случайными величинами устанавливаются операции сло-

жения и умножения.

1. *Суммой* двух случайных величин X и Y называется

случайная величина, которая получается в результате

сложения всех значений случайной величины X и всех

значений случайной величины Y, соответствующие

вероятности перемножаются.

2. *Произведением* двух случайных величин Х и Y

называется случайная величина, которая получается в

результате перемножения всех значений случайной

величины X и всех значений случайной величины Y,

соответствующие вероятности перемножаются.

Задача-пример:

---------------------



Z = X + Y.



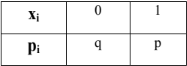
Одинаковые значения СВ можно записать один раз, предварительно сложив соответствующие вероятности:



---------------------

*Распределения дискретных случайных величин:*

1. Закон распределения Бернулли. Случайная величина X, распределенная по закону Бернулли (индикаторная случайная величина) принимает значения 1 - успех или 0 - неудача, с вероятностями р и q соответственно (p+q=1).



2. Биномиальный закон распределения. Случайная величина X принимает значения: 0, 1, 2, 3, 4, 5,..., п, с вероятностью, определяемой по формуле Бернулли :



*Числовые характеристики дискретных случайных величин:*

I. Характеристики положения ряда распределения:

1. Математическое ожидание.

2. Медиана.

3. Мода.

II. Характеристики рассеяния:

1. Дисперсия.

2. Среднее квадратическое отклонение.

*Математическим ожиданием* М(Х) ДСВ X называется среднее значение случайной величины:



Свойства математического ожидания:

1. Математическое ожидание постоянной величины равно ей самой: M[C]=C, C – постоянная;

2. M[C•X]=C•M[X]

3. Математическое ожидание суммы случайных величин равно сумме их математических ожиданий: M[X+Y]=M[X]+M[Y]

4. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий: M[X•Y]=M[X]•M[Y], если X и Y независимы.

*Медиана* Mе (X) - это значение случайной величины, которое делит таблицу распределения на две части таким образом, что вероятность попадания в одну из них равна 0,5.

*Мода* Мо (Х) распределения - это значение СВ, имеющее наиболее вероятное значение.

*Дисперсией* ДСВ X называется математическое ожидание квадрата отклонения СВ от ее математического ожидания:



Свойства дисперсии:

1. Дисперсия постоянной величины равна нулю: D(c)=0.

2. Постоянный множитель можно вынести из-под знака дисперсии, возведя его в квадрат: D(k\*X)= k2D(X).

3. Если случайные величины X и Y независимы, то дисперсия суммы равна сумме дисперсий: D(X+Y)=D(X)+D(Y).

4. Если случайные величины X и Y зависимы: D(X+Y)=DX+DY+2(X-M[X])(Y-M[Y])

5. Для дисперсии справедлива вычислительная формула:

D(X)=M(X2)-(M(X))2

Дисперсия характеризует средний квадрат отклонения ДСВ, поэтому на практике часто используют в качестве характеристики разброса среднее квадратическое отклонение:



# 6. Случайные величины. Непрерывные случайные величины.

Случайная величина (СВ) – это величина, которая в результате опыта может принимать те или иные значения, причем до опыта мы не можем сказать, какое именно значение она примет.

СВ обозначаются буквами латинского алфавита X, Y, Z.

СВ могут быть трех типов:

-дискретные

-непрерывные

-смешанные

Непрерывная случайная величина (НСВ) в отличие от ДСВ принимает бесконечное несчетное число значений.

Например, мишень имеет форму круга радиуса R. По этой мишени произвели

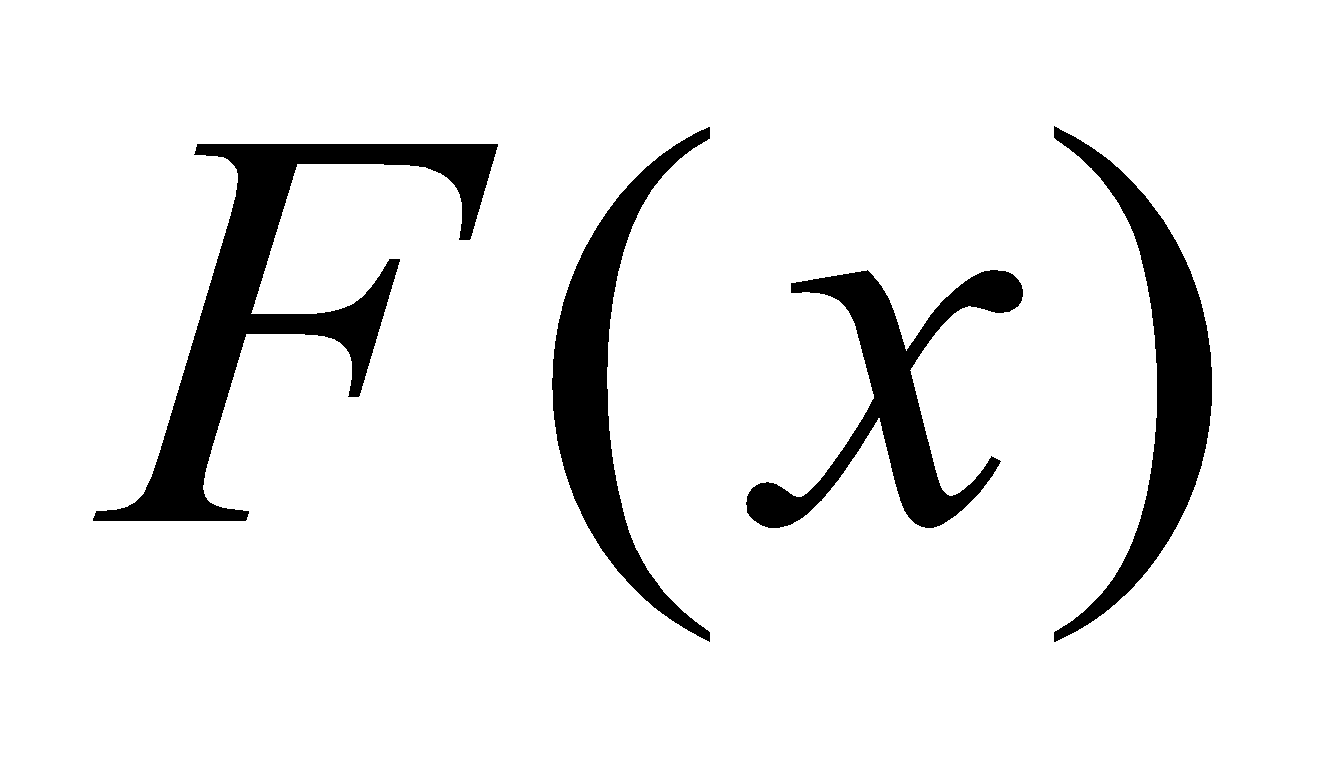
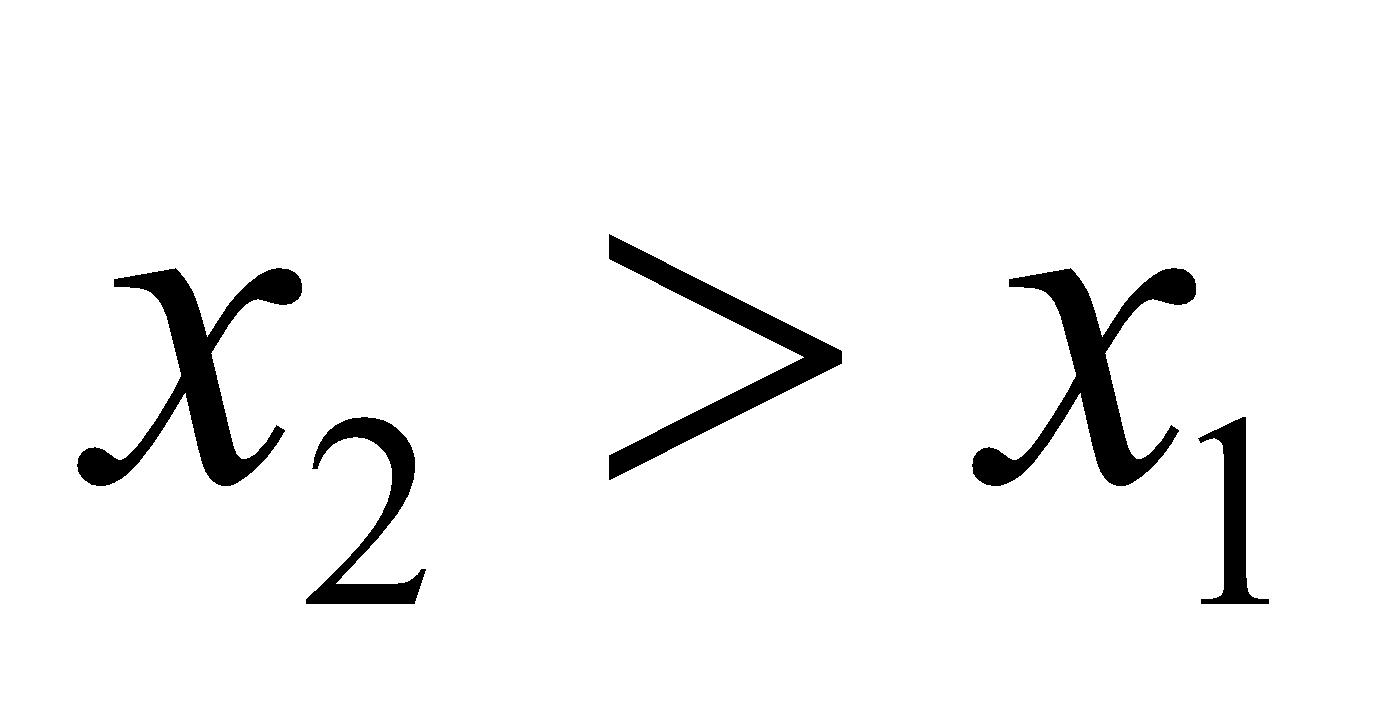
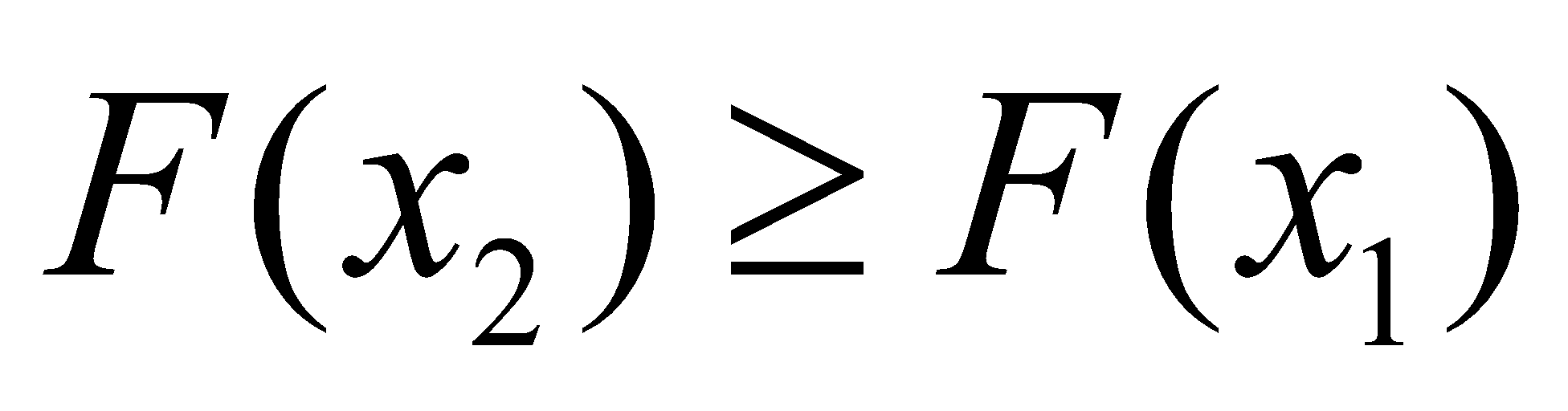
выстрел с обязательным попаданием. Обозначим через Y расстояние от центра до точки попадания в мишень, Y є [0; R].

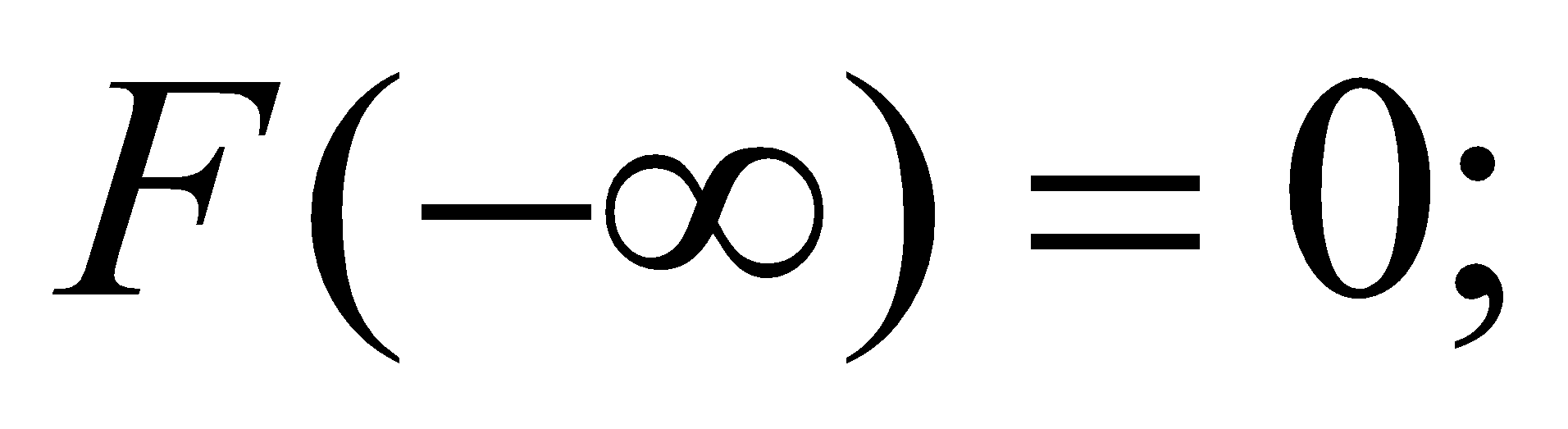
Y – непрерывная случайная величина, так как она принимает бесконечное

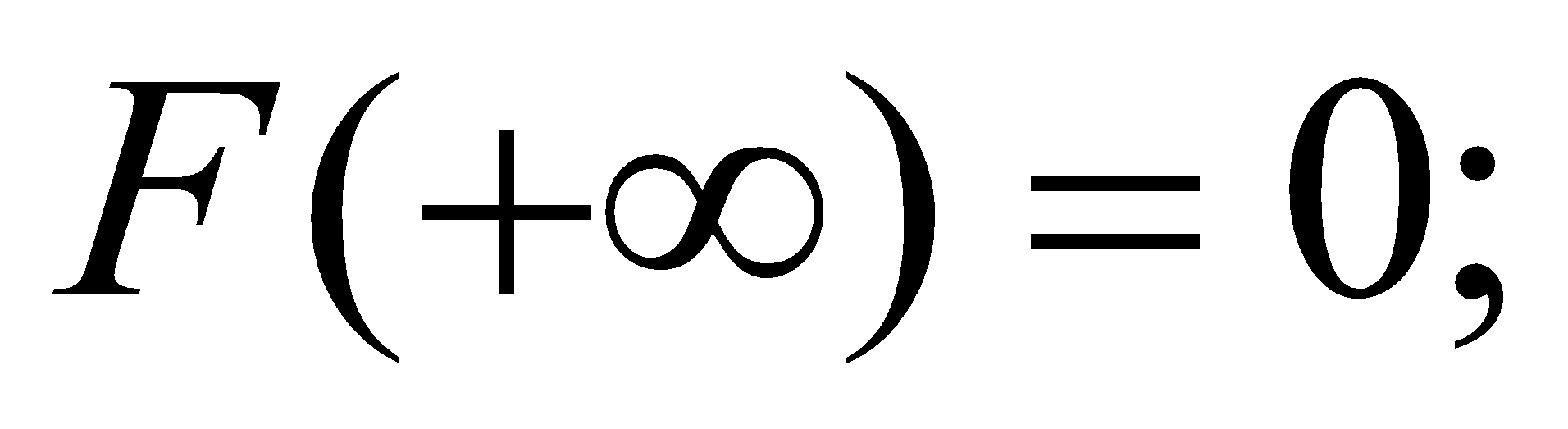
несчетное число значений.

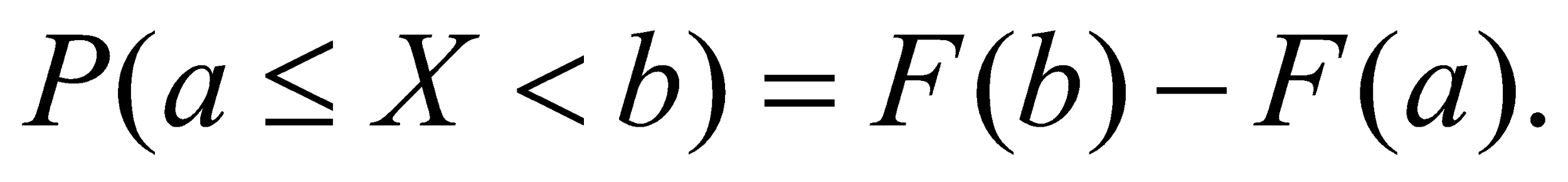
В общем случае случайная величина X задается функцией распределения F(x), которая выражает вероятность того, что X принимает значение, меньшее, чем x: F(x) = P (X < x).

Функция распределения обладает свойствами:

1. не убывает (если , то );

2. 

3. 

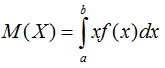
4. Вероятность попадания СВ X в интервал a<x<b: 

*Числовые характеристики непрерывных случайных величин:*

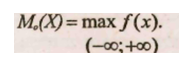
1)Математическое ожидание НСВ X определяется по формуле:



Если НСВ X определена на интервале (a;b), то:



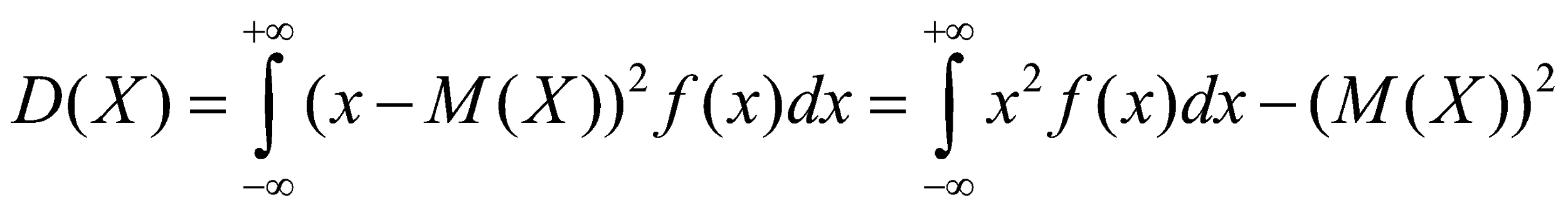
2)Мода НСВ X будет определяться как максимум ее дифференциальной функции:



3)Медиана определяется как значение случайной величины, которое делит площадь под дифференциальной функцией на две равные части:



4) Дисперсия НСВ:



Все свойства дисперсии и математического ожидания, установленные для ДСВ, сохраняются для НСВ.

5) Моменты случайных величин (начальные и центральные).

Кроме характеристик положения и рассеяния существует ряд других числовых характеристик распределения, например, моменты.

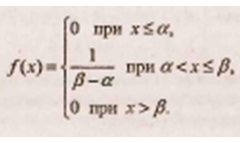
*Основные законы распределения непрерывных случайных величин:*

1.Равномерный закон распределения:

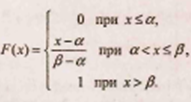
СВ X распределена по равномерному (прямоугольному) закону, если все значения СВ лежат внутри некоторого интервала и все они равновероятны (точнее, обладают одной плотностью вероятности).

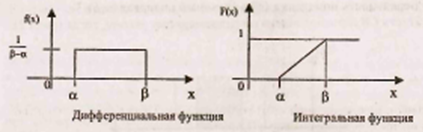
Например, если весы имеют точность 1г и полученное значение округляется до ближайшего целого числа k, то точный вес можно считать равномерно распределенной СВ на интервале (k-0,5; k+0,5).

Дифференциальная функция равномерного закона на интервале (α, β):



Интегральная функция равномерного закона на интервале (α, β):





Основные числовые характеристики равномерного закона:

1)Математическое ожидание:

M(X) совпадает, в силу симметрии распределения, с медианой.

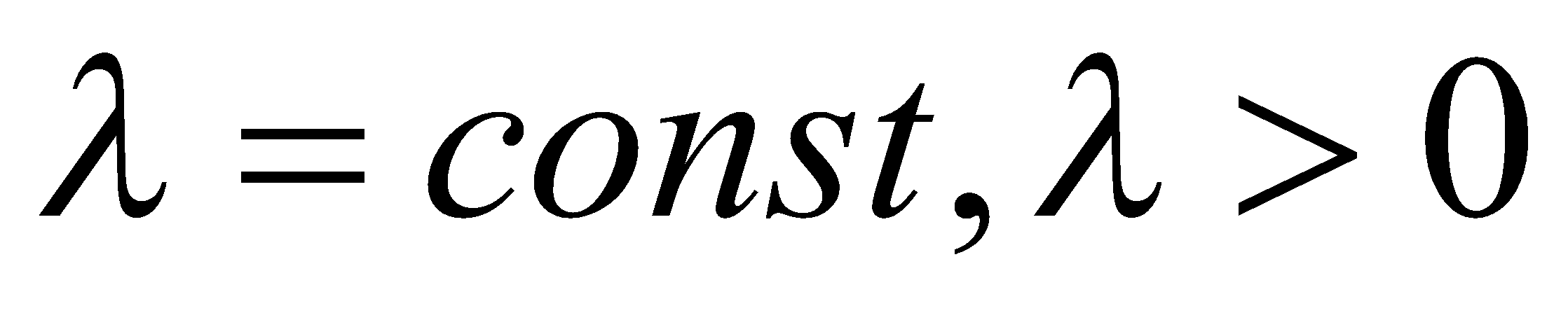
2)Моды равномерное распределение не имеет.

3)Дисперсия:

4)Вероятность попадания СВ в заданный интервал (а;b)

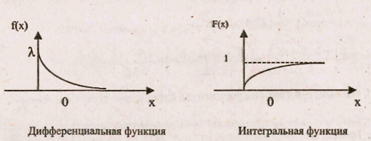
2. Показательное распределение:

НСВ X, принимающая неотрицательные значения, имеет показательное распределение, если ее дифференциальная функция имеет вид:

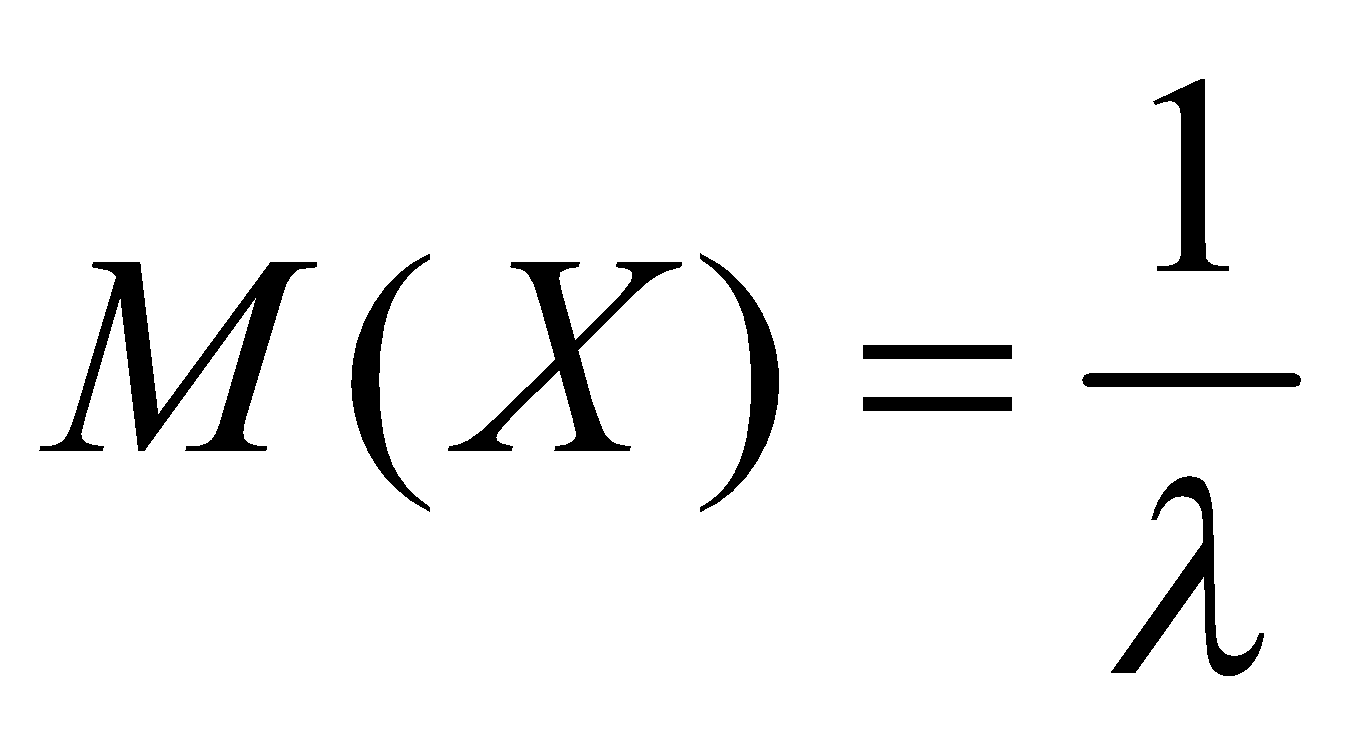
, где 

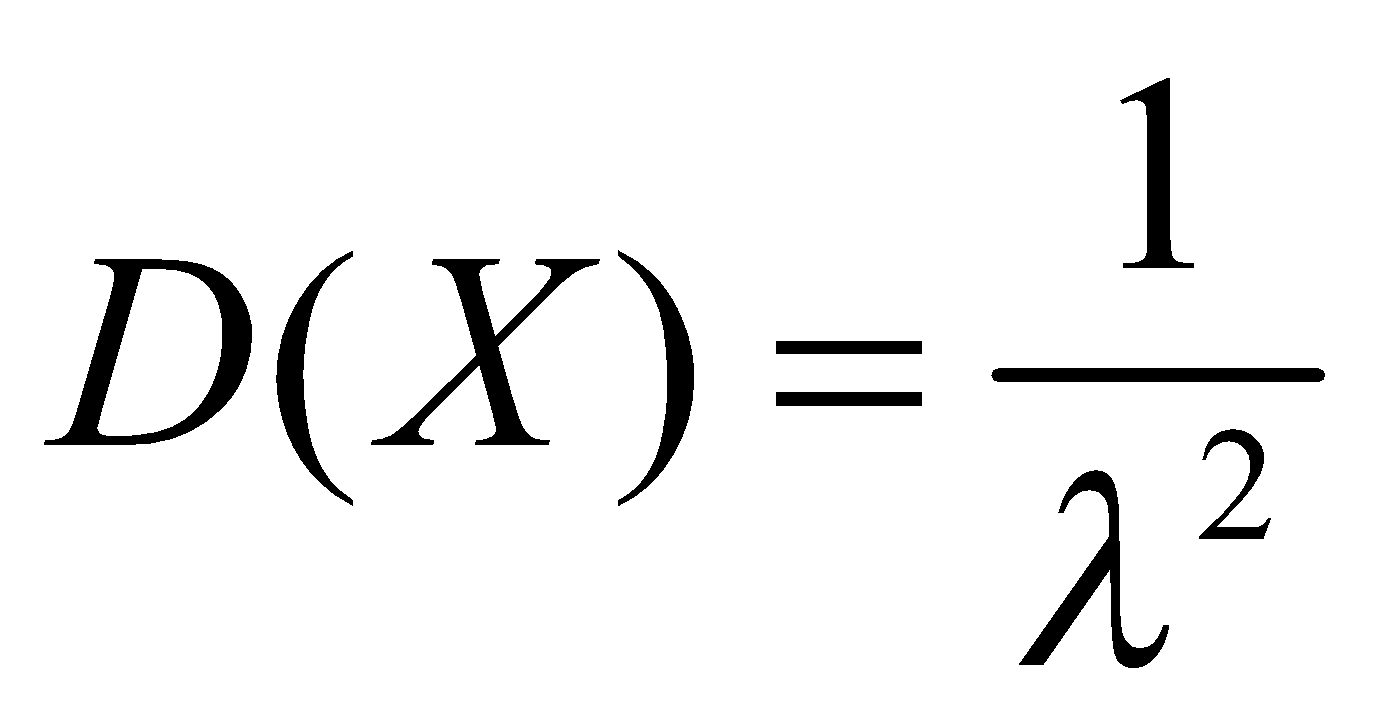
Интегральная функция показательного закона:

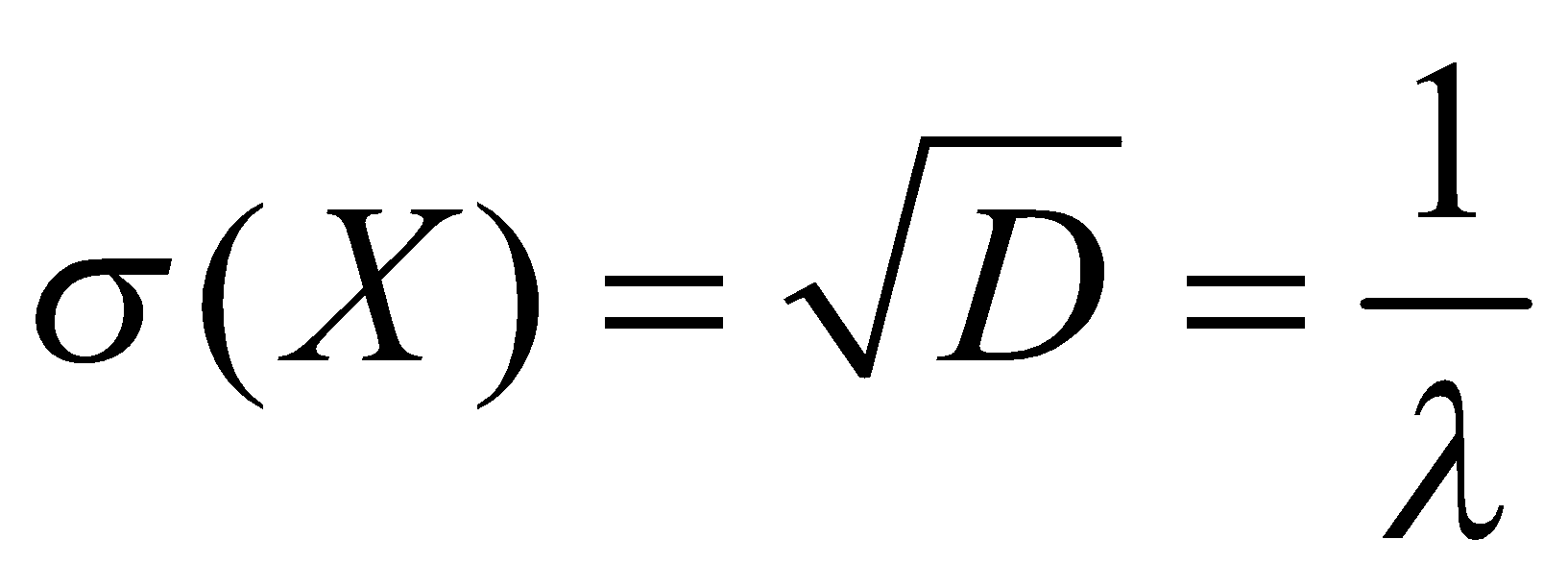


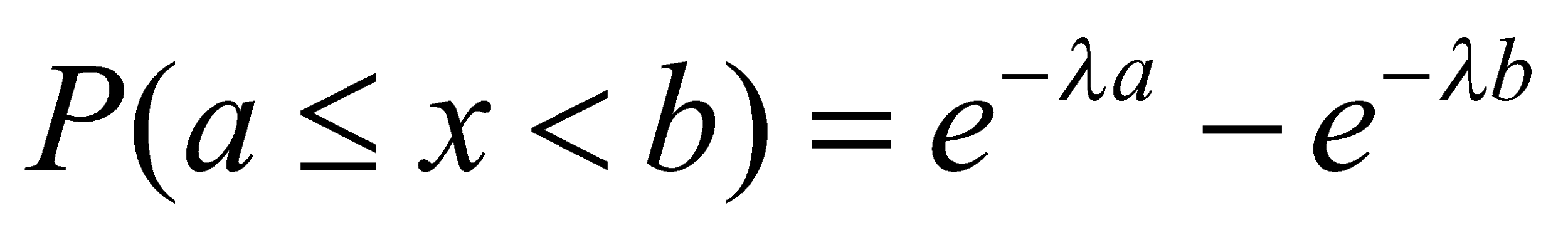


Числовые характеристики показательного закона:

1)Математическое ожидание: 

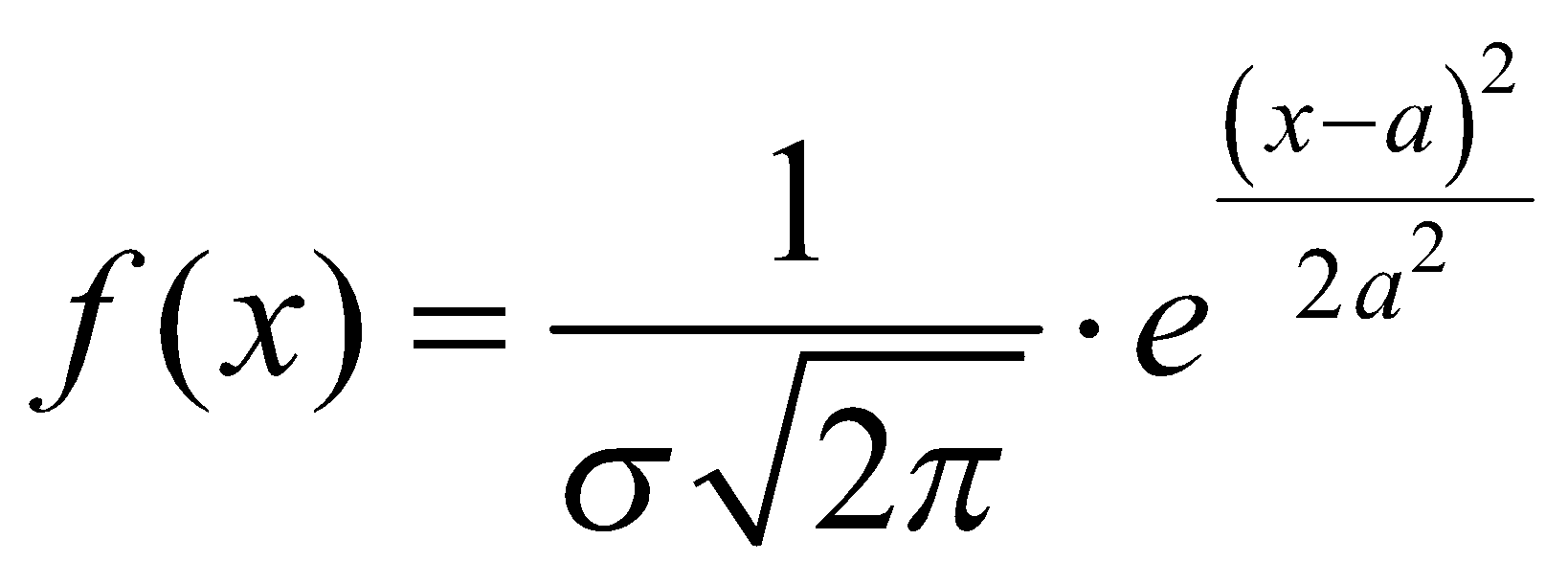
2)Дисперсия: ,

среднеквадратическое отклонение: 

3)Вероятность попадания СВ X в заданный интервал: 

3. Нормальный закон распределения играет исключительную роль в теории вероятностей. Это наиболее часто встречающийся закон распределения, главной особенностью которого то, что он является предельным законом, к которому, при определенных условиях, приближаются другие законы распределения.

Дифференциальная функция нормального закона имеет вид:



Числовые характеристики нормального закона:

1. Математическое ожидание характеризует центр распределения:



2. Дисперсия характеризует форму распределения:



# 7. Генеральная и выборочная совокупности

*Генеральной совокупностью* называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе.

В математической статистике генеральная совокупность часто понимается как совокупность всех мыслимых наблюдений, которые могли быть произведены при выполнении данного комплекса условий. Понятие генеральной совокупности аналогично понятию случайной величины (закону распределения вероятностей), так как обе они полностью определяются заданным комплексом условий. Так как понятия генеральной совокупности и совокупности всех значений случайной величины связаны с испытаниями (наблюдениями) в неизменных условиях, то в дальнейшем эти понятия не будут различаться.

Понятие генеральной совокупности несколько шире понятия случайной величины, так как случайная величина может быть результатом нескольких испытаний.

Генеральная совокупность может быть конечной или бесконечной.

Число объектов (наблюдений) в генеральной совокупности называется ее объемом.

Изучение всего набора элементов генеральной совокупности часто оказывается невозможным, в таких случаях рассматривают некоторую часть объема.

Часть объектов генеральной совокупности, используемая для исследования, называется *выборочной совокупностью* или выборкой.

*Пример.* Число единиц товара N, произведенного фирмой в течение года, есть конечная генеральная совокупность. Для исследования качества продукции на практике рассматривается выборка, состоящая из п единиц товара. Признаком, или случайной величиной, может быть число единиц товара, удовлетворяющих сертификатным требованиям.

Сущность выборочного метода в математической статистике заключается в том, чтобы по определенной части генеральной совокупности (выборке) судить о ее свойствах в целом.

Выборочный метод является единственно возможным в случае бесконечной генеральной совокупности или когда исследование связано с уничтожением (гибелью) наблюдаемых объектов (например, исследование предельных режимов приборов, исследование действия вирусов на подопытных животных и т.д.). Для того чтобы по выборке можно было адекватно судить о случайной величине, она должна быть представительной (репрезентативной).

Репрезентативность выборки обеспечивается случайностью отбора ее элементов, так как все элементы генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.

Имеются два способа образования выборки:

1) повторная выборка, когда каждый элемент, случайно отобранный и исследованный, возвращается в общую совокупность и может быть отобран повторно;

2) беповторная выборка, когда отобранный элемент не возвращается в общую совокупность.

# 8. Вариационный ряд: определение, графическое представление.

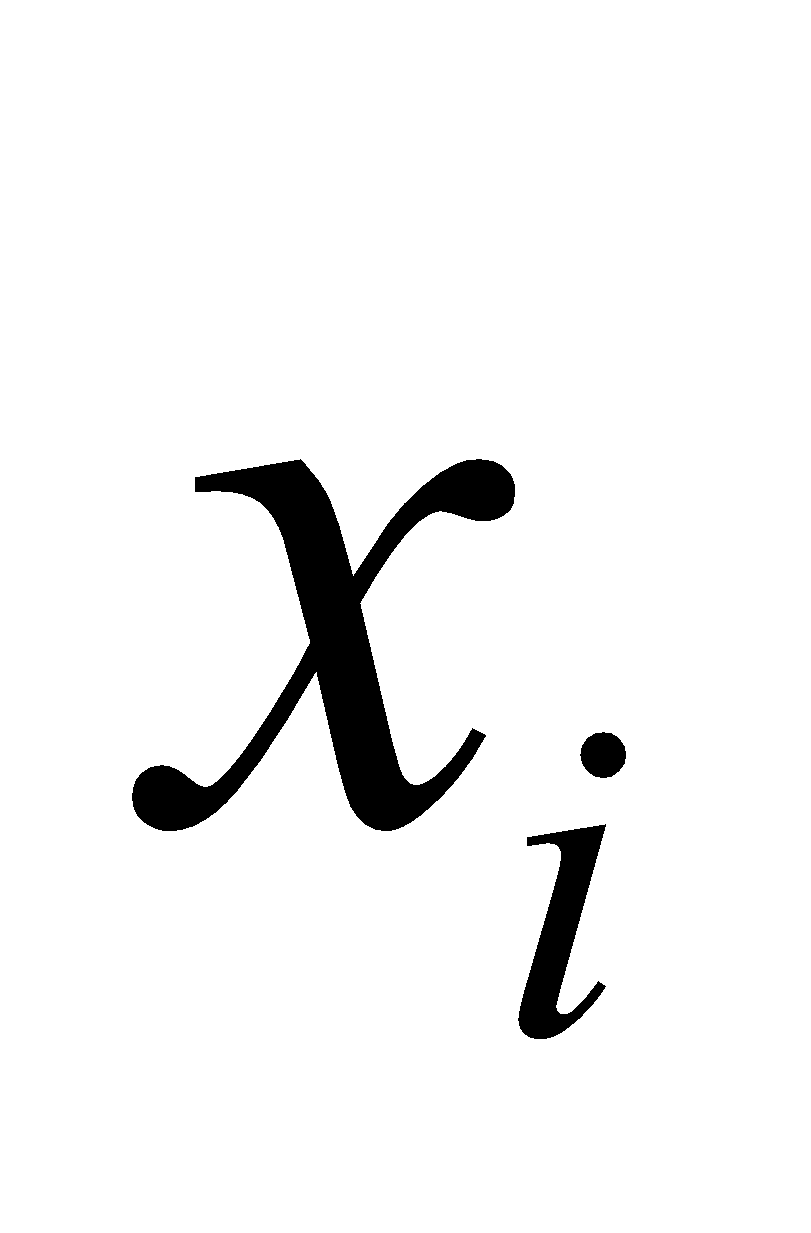
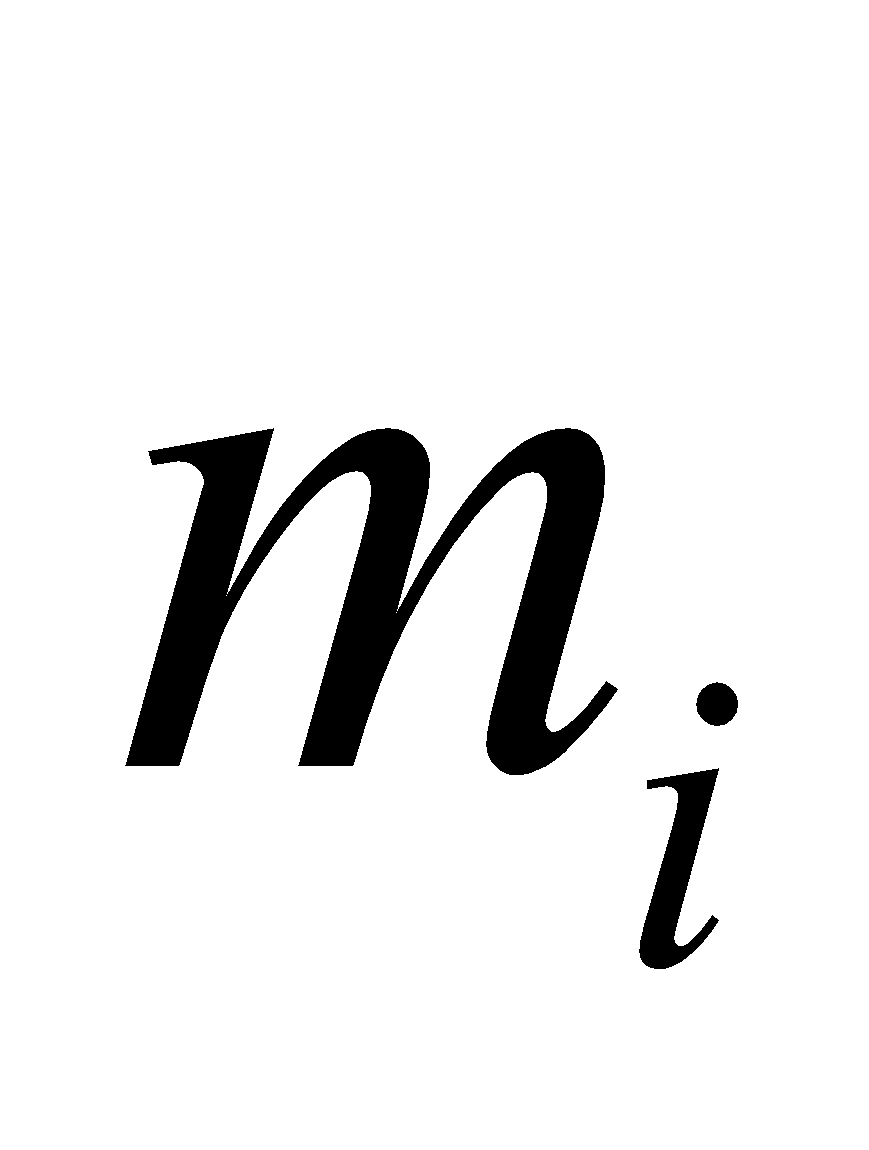
*Определение*

Пусть некоторый признак генеральной совокупности описывается случайной величиной *X.*

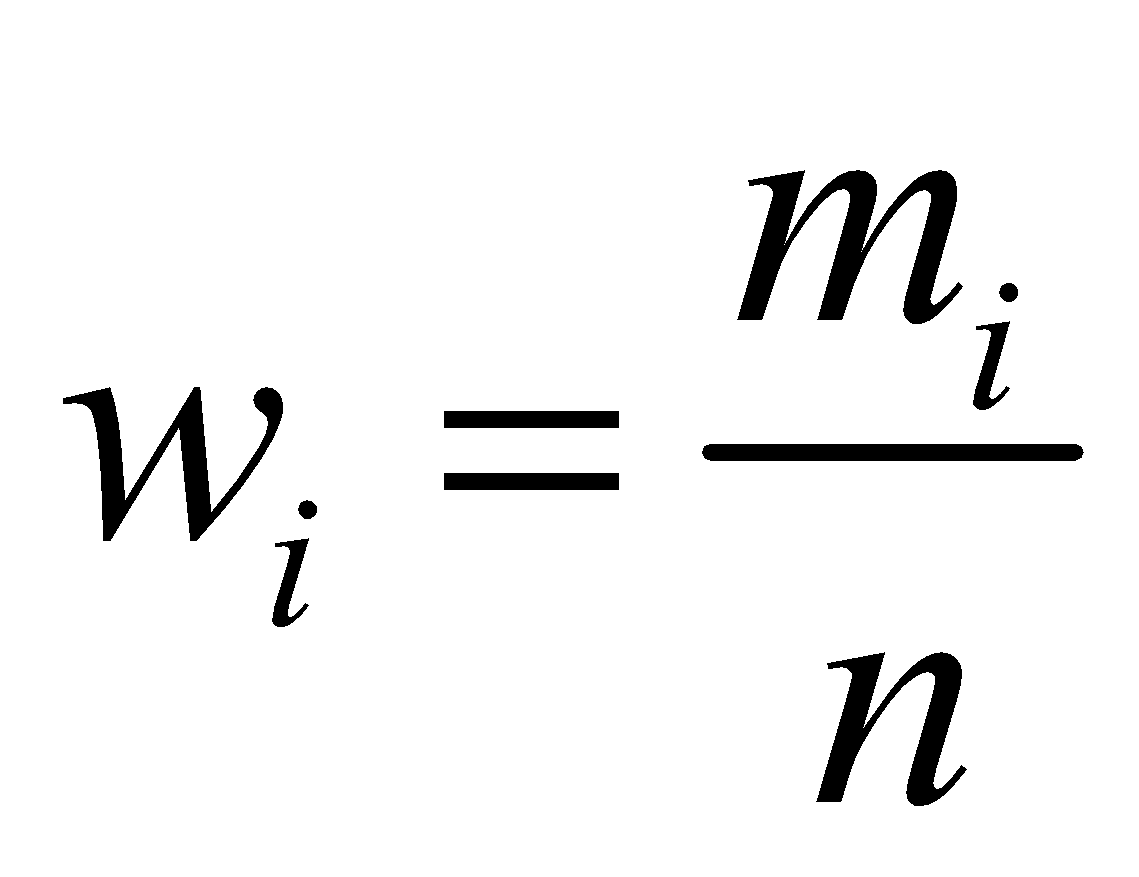
Рассмотрим выборку *{х1,х2,...,хп}* объема *п* из генеральной совокупности. Элементы этой выборки представляют собой значения случайной величины *X.*

На первом этапе статистической обработки производят *ранжирование* выборки, т.е. упорядочивание чисел *х1,х2,...,хп* по возрастанию.

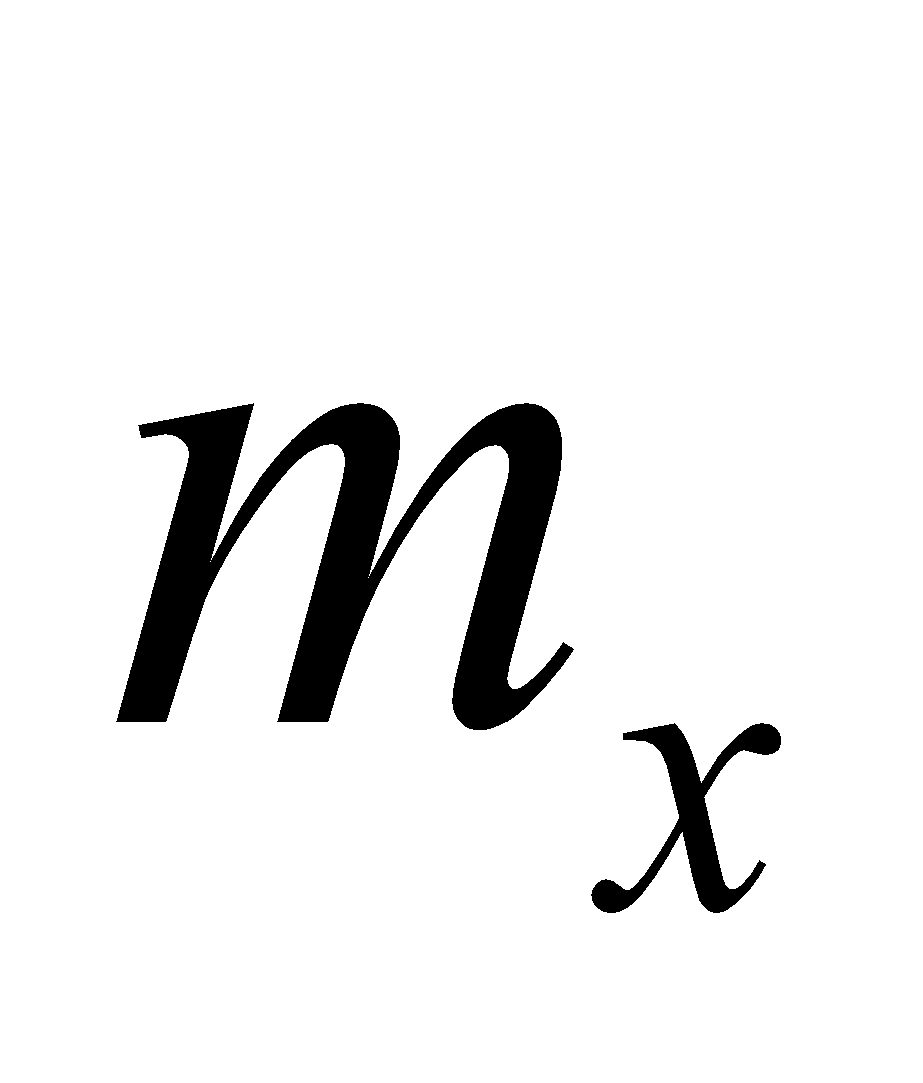
Различные элементы выборки называются **вариантами.**

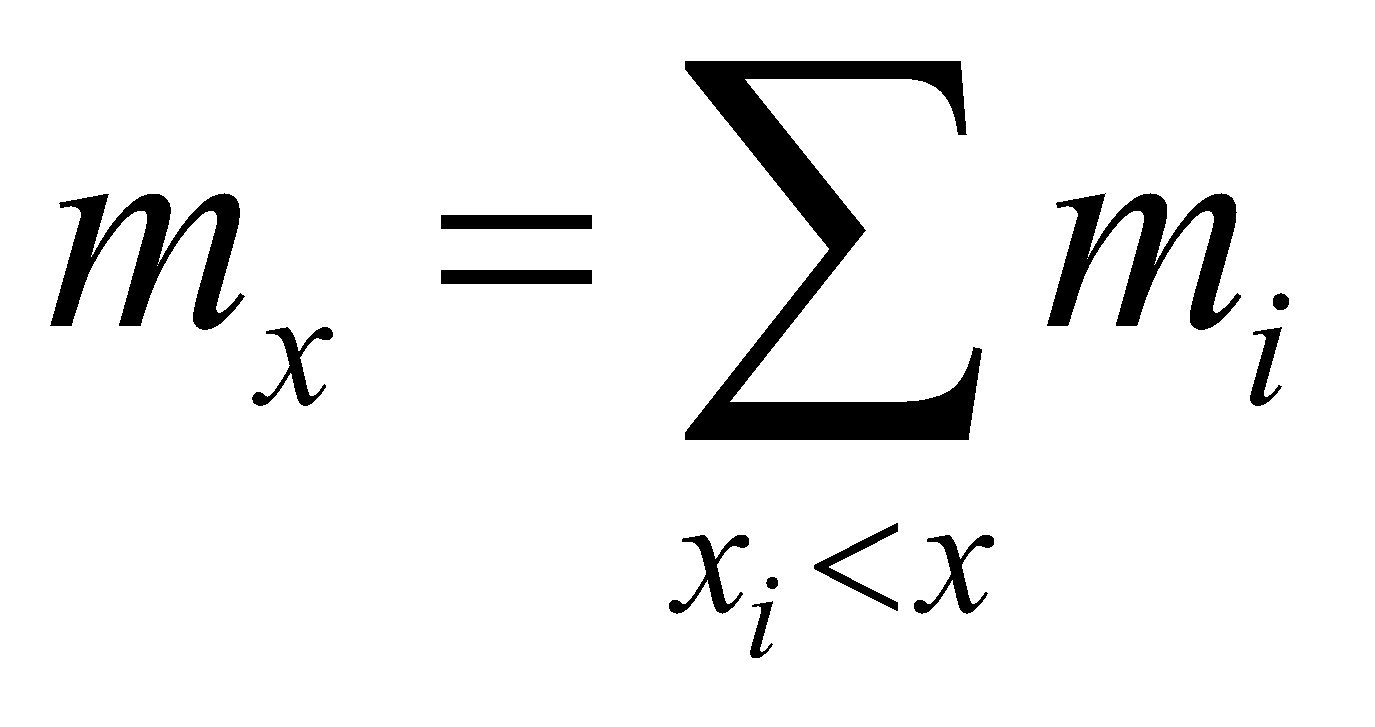
**Частотой варианты** называется число *,* показывающее, сколько раз эта варианта встречается в выборке.

**Частостью, относительной частотой**или **долей**варианты называется число

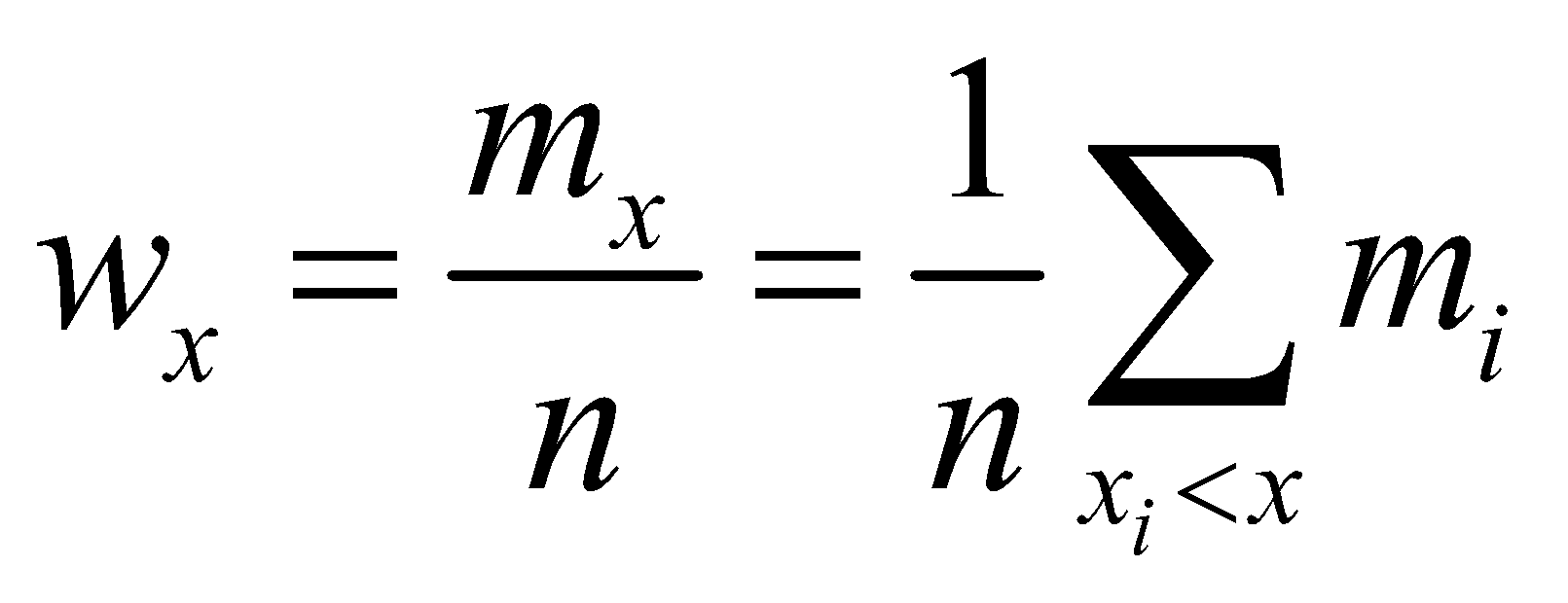
 (1.1)

Частоты и частости называются **весами**.

Пусть *х* некоторое число. Тогда количество вариант , значения которых меньше *х,* называется накопленной частотой, т.е.

 (1.2)

Отношение накопленной частоты к общему числу наблюдений *п* называется **накопленной частостью:**



Ряд вариант, расположенных в порядке возрастания их значений, с соответствующими им весами называется **вариационным рядом.**

Вариационные ряды бывают:

- дискретные;

- интервальные.

Вариационный ряд называется **дискретным,**если он представляет собой выборку значений дискретной случайной величины.

Ряд называется **непрерывным (интервальным)***,* если он представляет выборку непрерывной случайной величины.

Общий вид дискретного вариационного ряда показан

в таблице:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Варианты |  |  | … |  |
| Частоты |  |  | … |  |

*Графические изображения вариационных рядов*

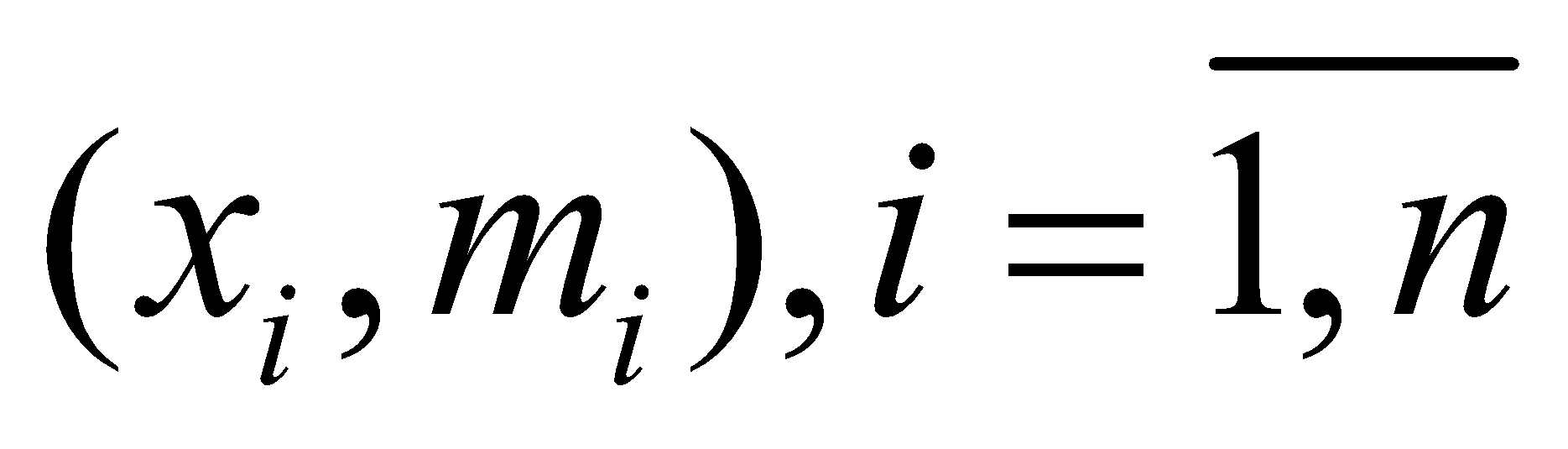
Для наглядности представления используют графические изображения вариационных рядов в виде:

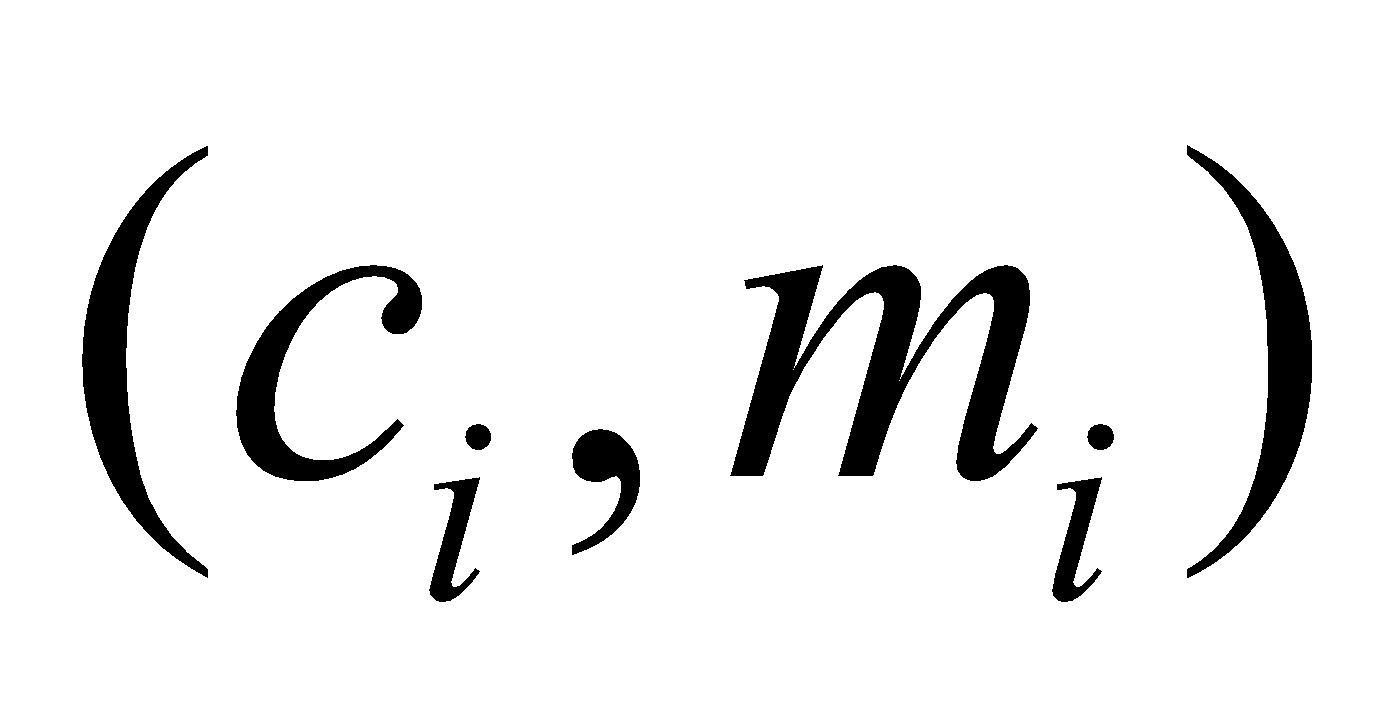
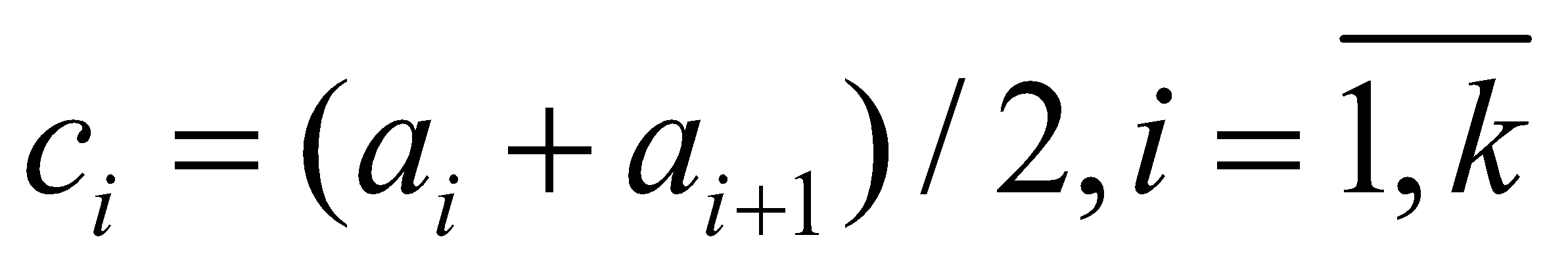
- полигона;

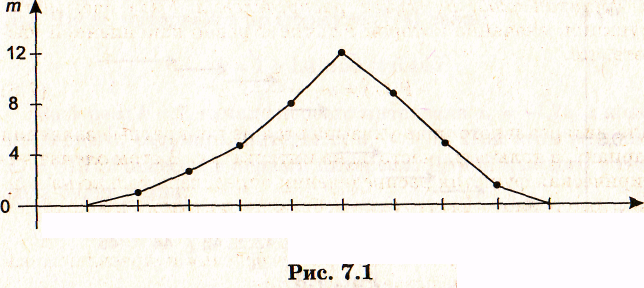
- гистограммы;

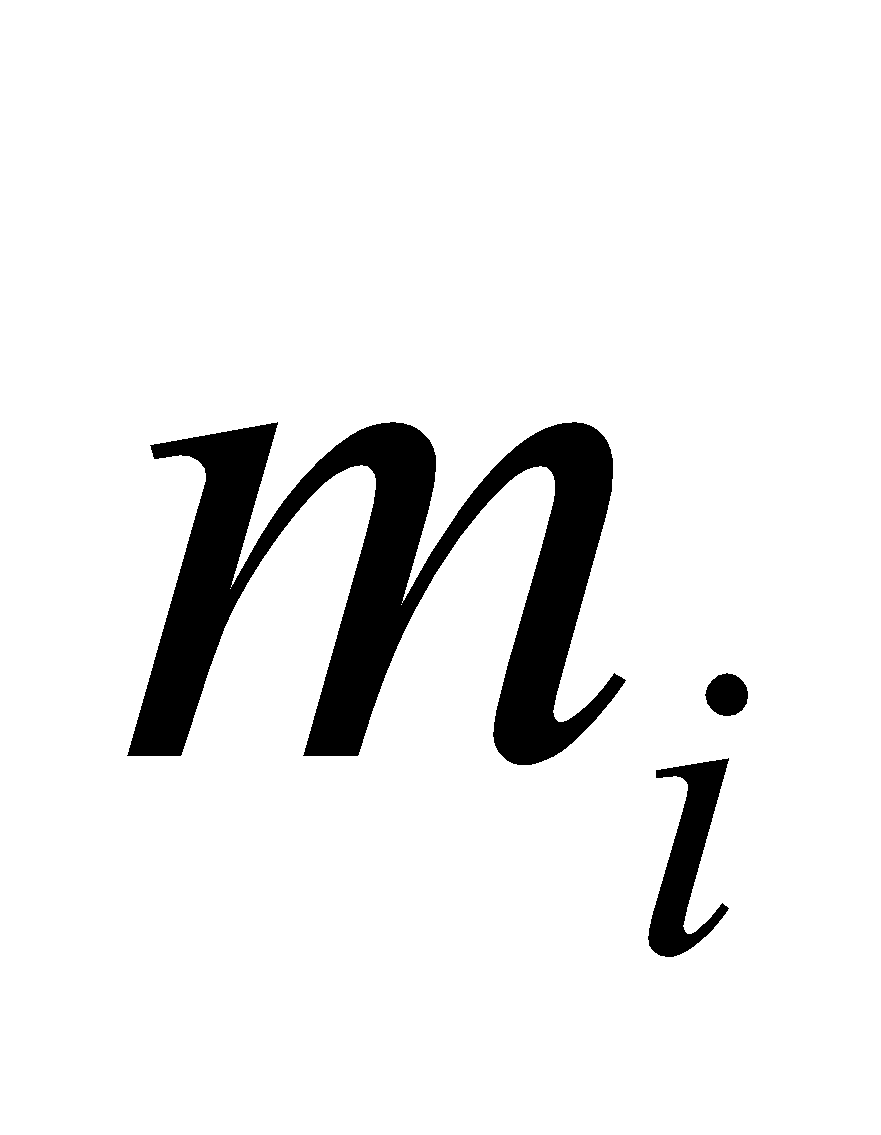
- кумулянты.

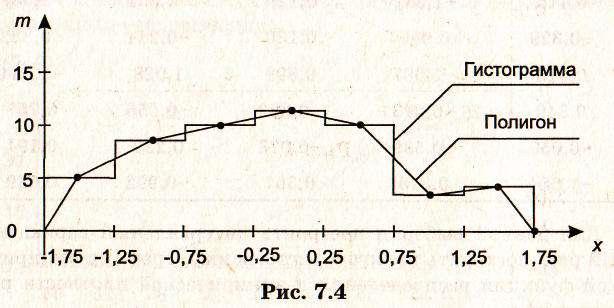
**Полигон***,* как правило, служит для изображения дискретного вариационного ряда.

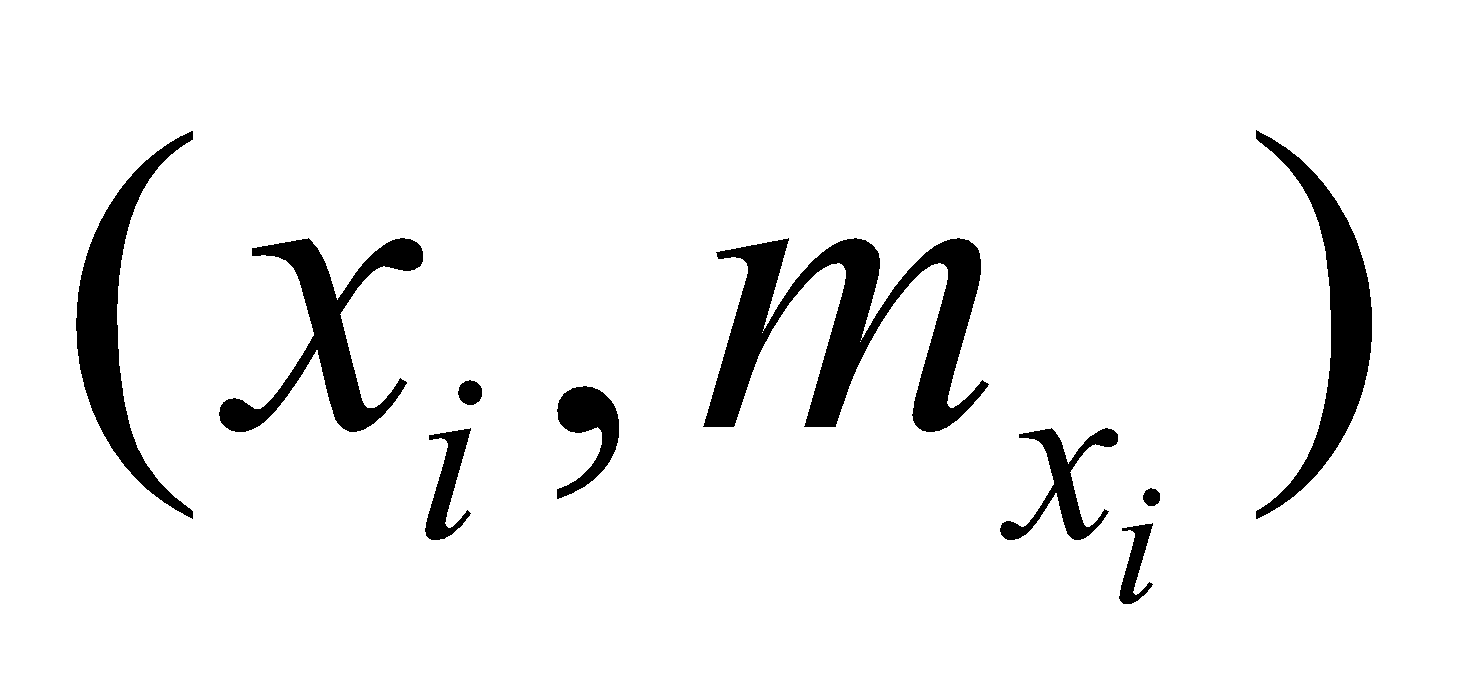
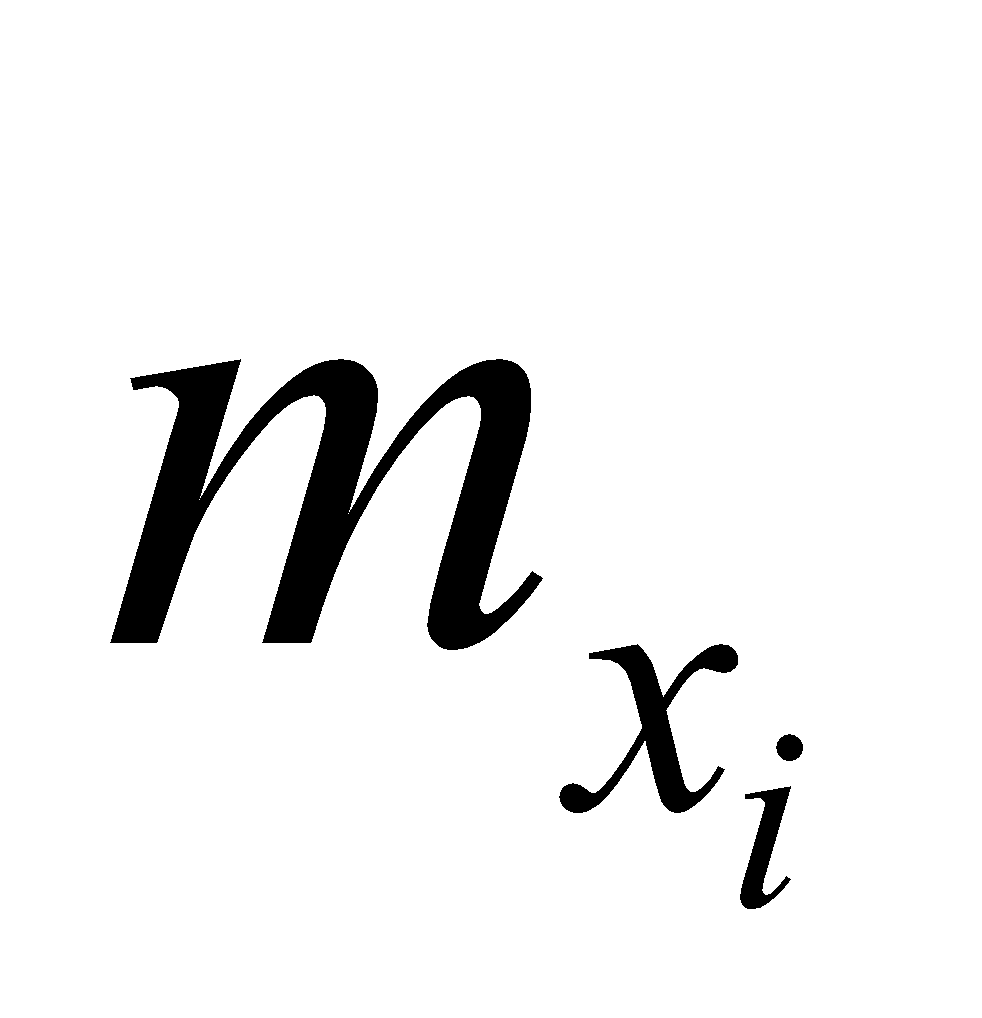
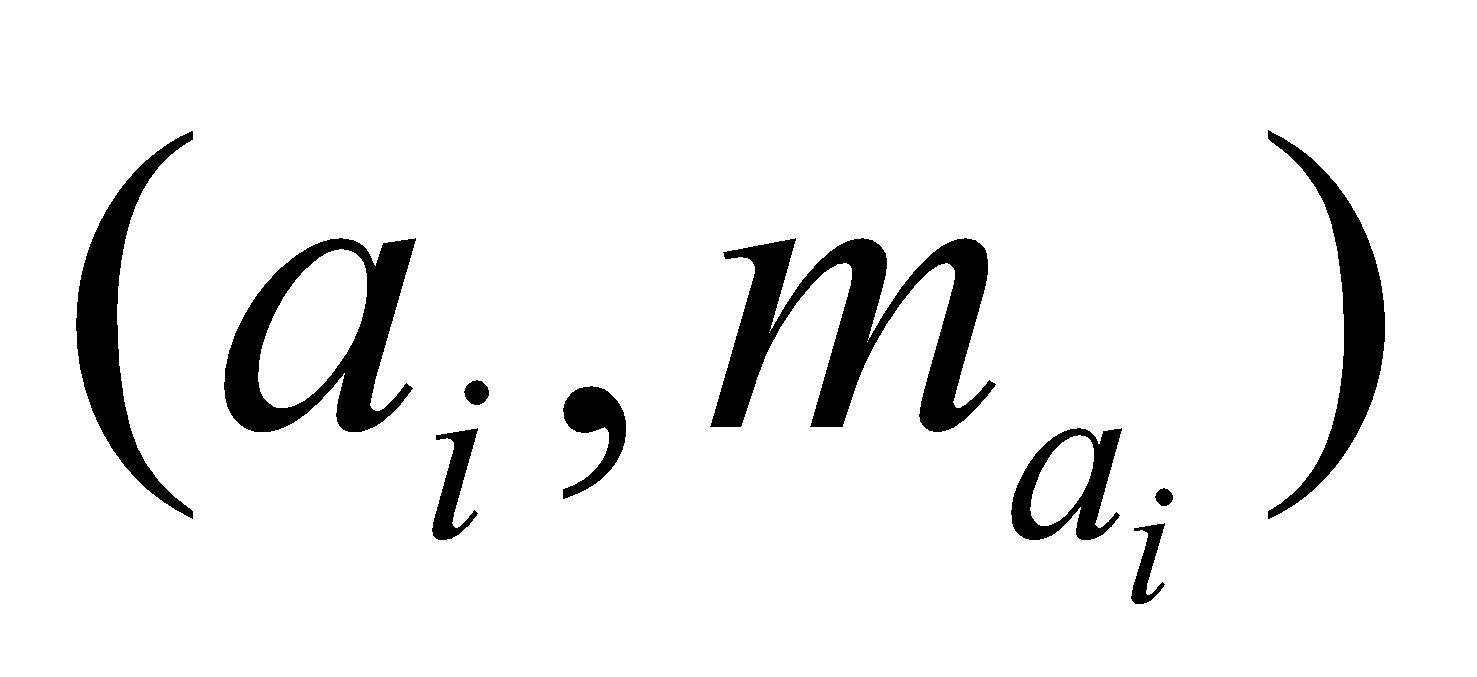
Представляет собой ломаную, соединяющую точки плоскости с координатами .

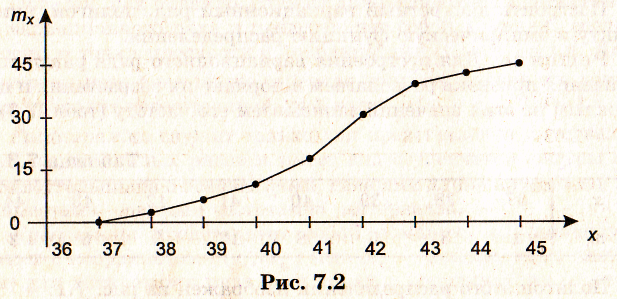
Для интервального ряда также строится полигон, только его ломаная проходит через точки , где .



**Гистограмма**служит только для представления интервальных вариационных рядов и имеет вид ступенчатой фигуры из прямоугольников с основаниями, равными длине интервалов Δ, и высотами, равными частотам  интервалов.

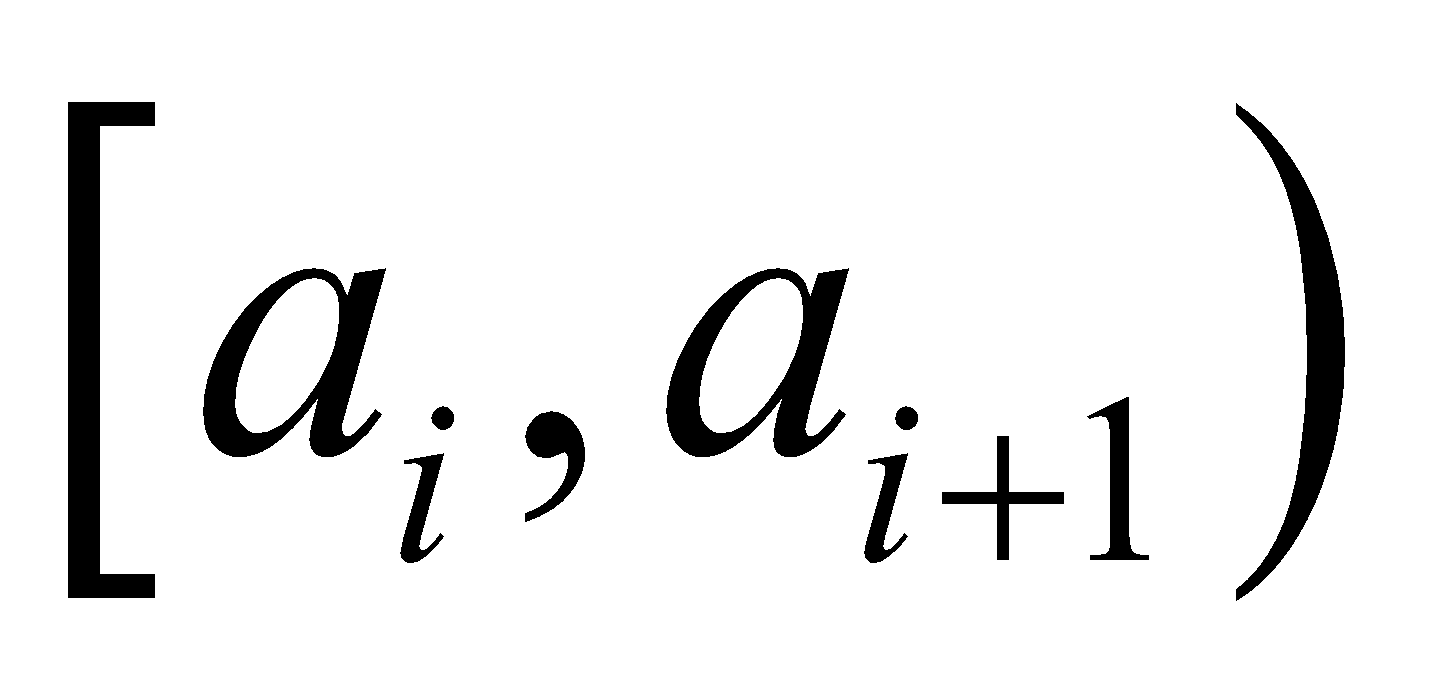


**Кумулянта**представляет собой ломаную, соединяющую точки с координатами (где— накопленные частоты) для дискретного ряда, или точки с координатами для интервального ряда.

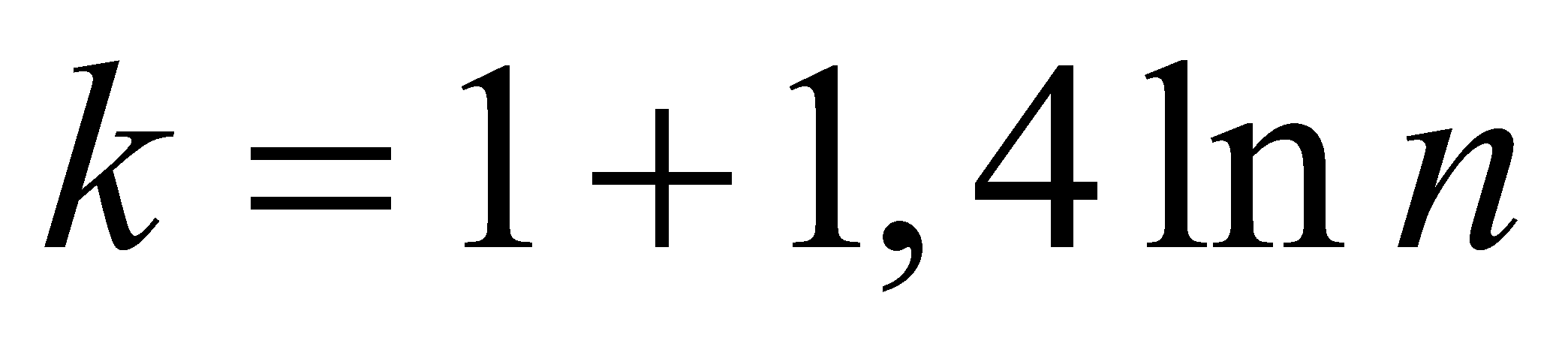


# 9. Алгоритм построения интервального вариационного ряда

Построение интервального вариационного ряда

1. Разбивают множество значений вариант на полуинтервалы  т.е. производят их *группировку.*

Рекомендуется количество интервалов *k* выбирать по формуле Стерджерса

 (1.4)

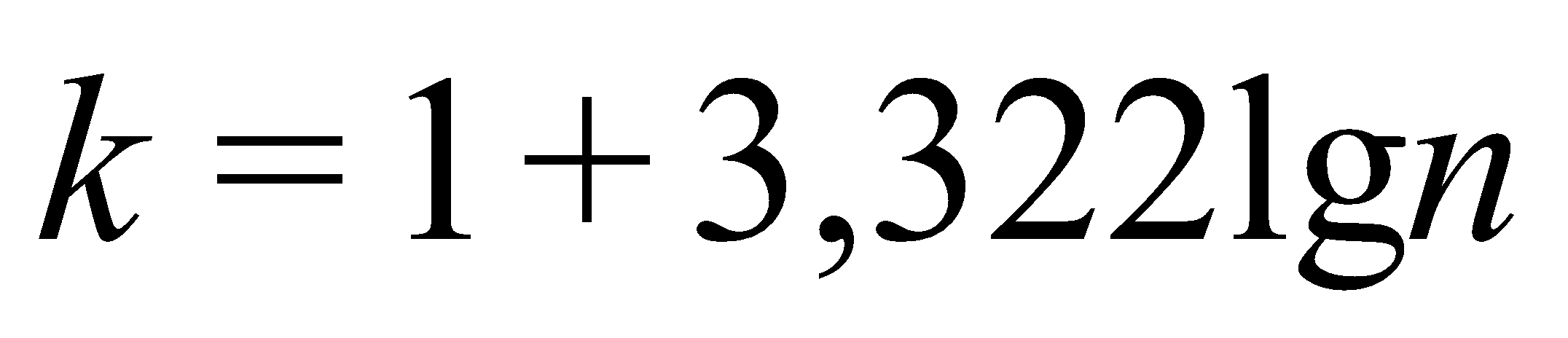
Длина интервала равна

Δ = xmax – xmin/ k

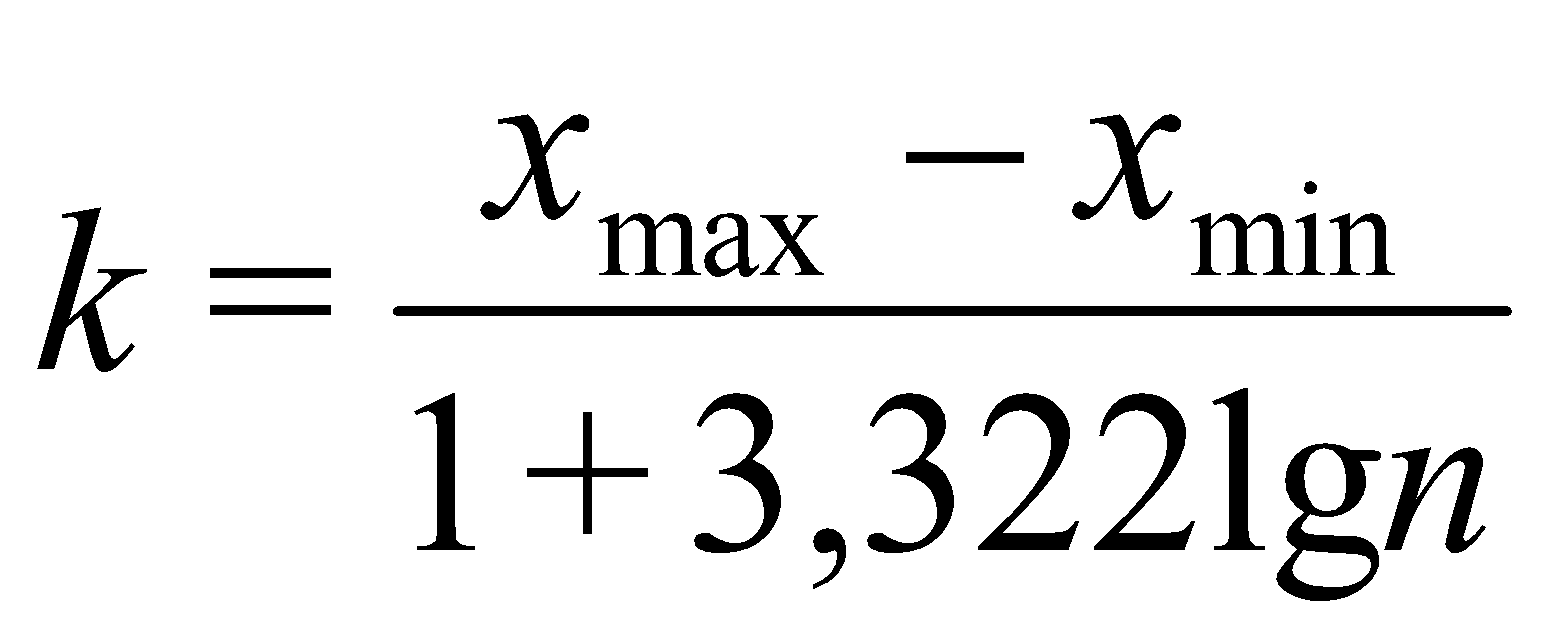
*Замечание 1.*

В литературе предлагается и такая форма записи формулы Стерджерса

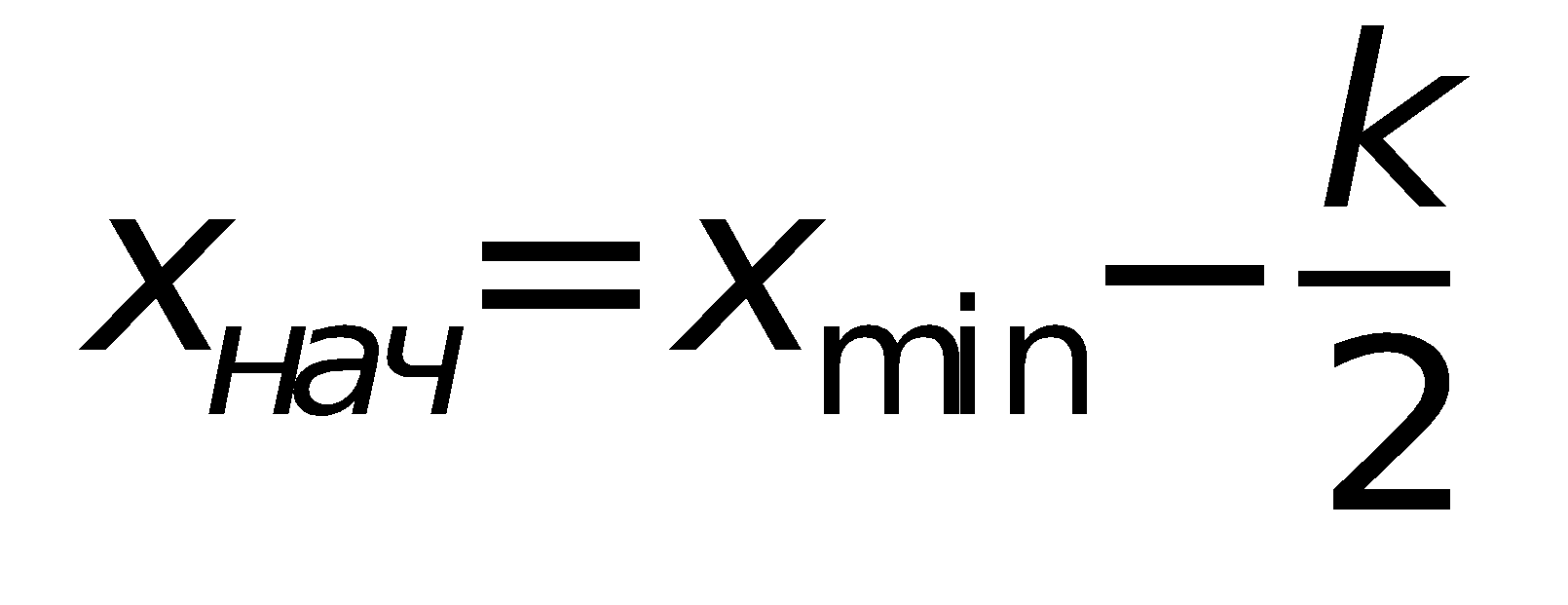
1. Рекомендуемое число интервалов

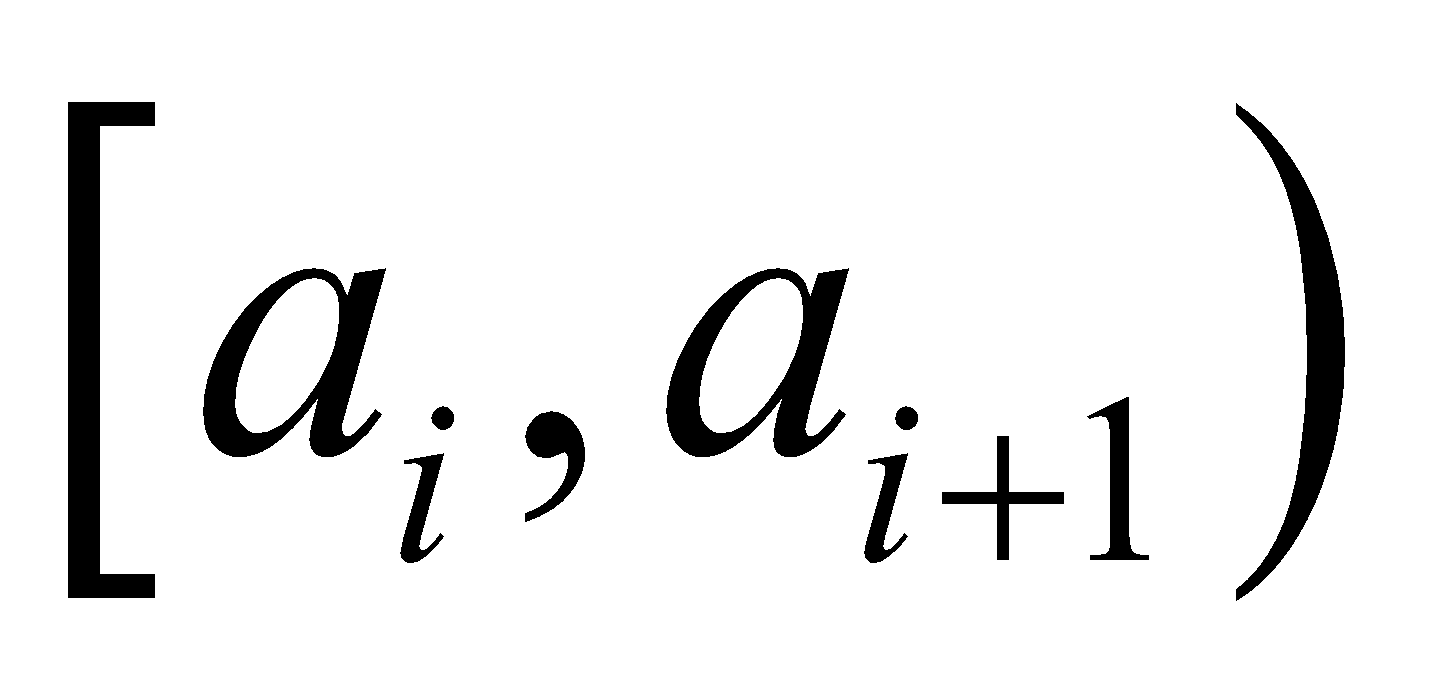


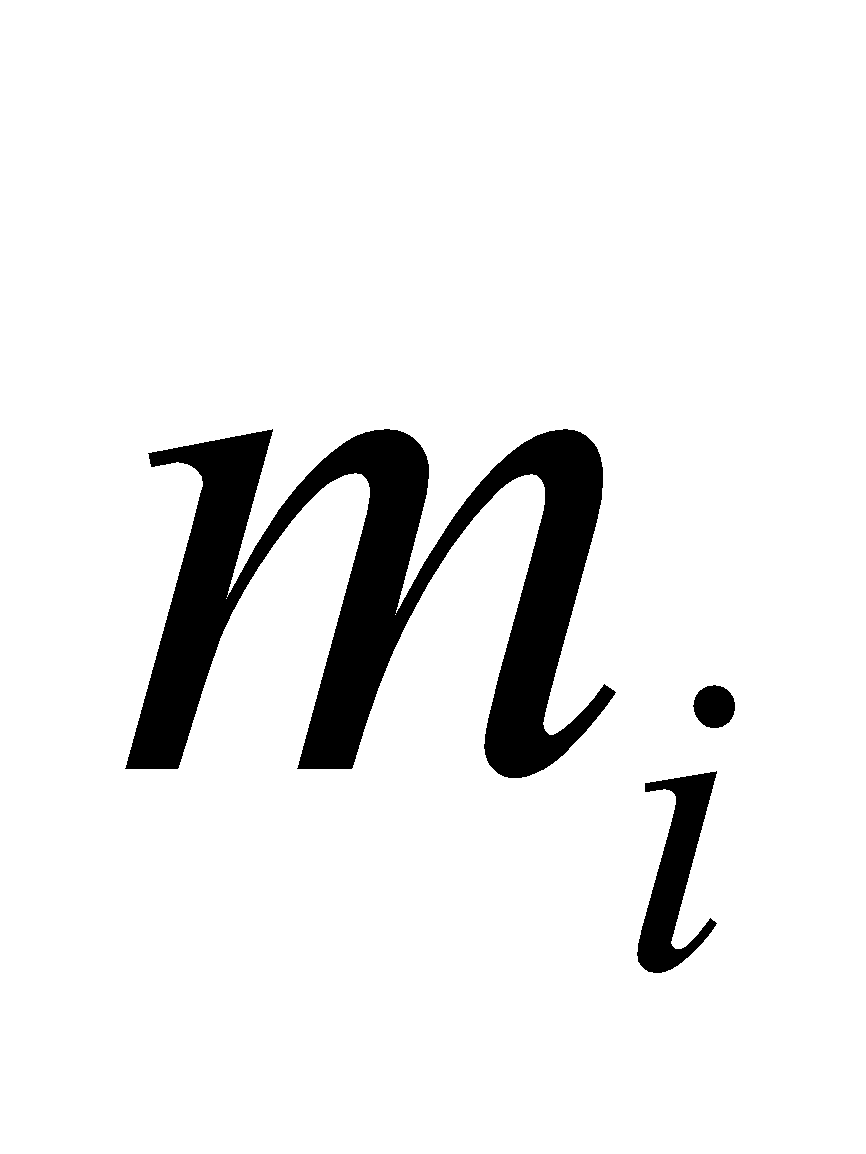
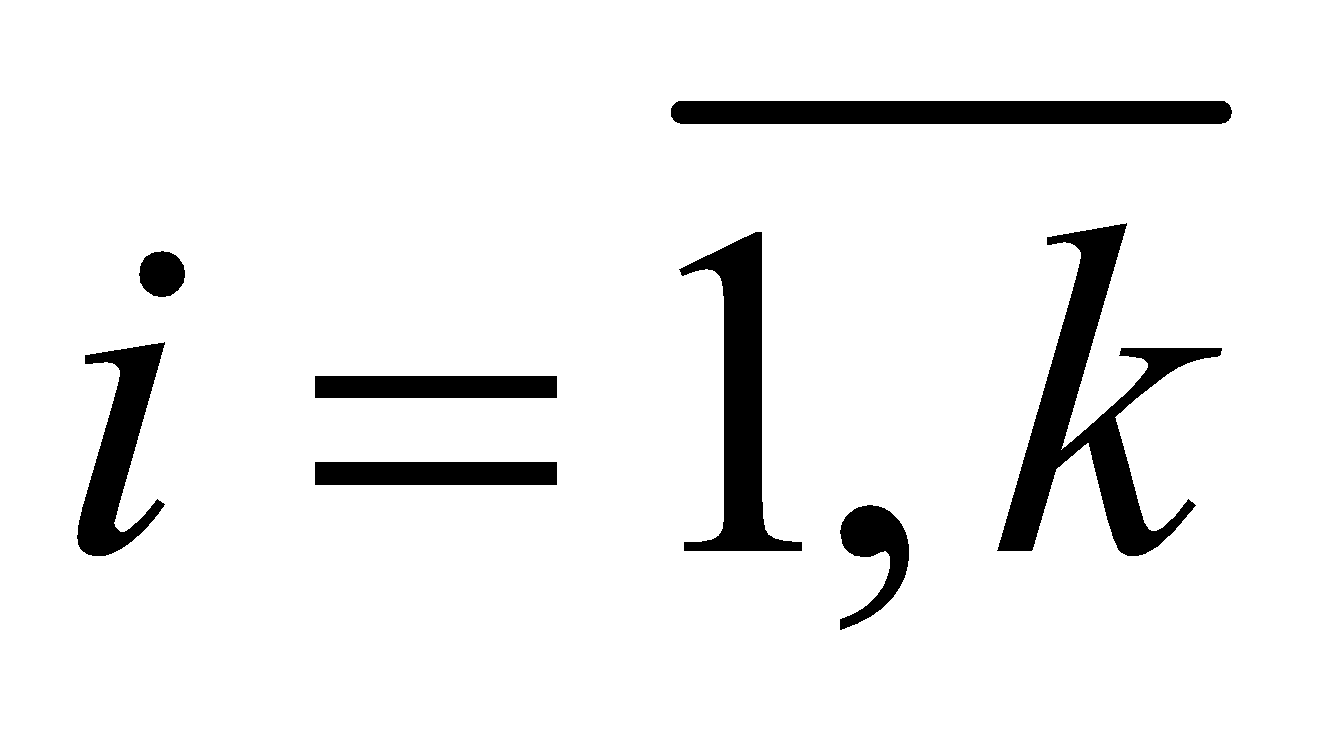
1. Величина интервала:



1. Строим интервал: за начало 1-го интервала берут:



2. Считают число вариант, попавших в полуинтервал .

Получают значения частот ,.

3. Интервальный ряд можно представить таблицей (табл. 1.2):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Варианты |  |  | … |  |
| Частоты |  |  | … |  |

*Замечание 2.*

Если варианта находится на границе интервала, то ее присоединяют к правому интервалу.

# 10. Алгоритм построения дискретного вариационного ряда

Дискретный вариационный ряд – это таблица, в которой указаны встречающиеся значения изучаемого признака как отдельные значения по возрастанию и их частоты

Для того, чтобы построить дискретный вариационный ряд нужно выполнить следующие действия:

1) упорядочить единицы наблюдения по возрастанию изучаемого значения признака,

2) определить все возможные значения признака xi, упорядочить их по возрастанию,

3) подсчитать сколько раз встречается каждое значение признака в изучаемой совокупности, т.е. определить частоту каждого значения признака ni.

4) записать полученные данные в таблицу из двух строк (столбцов) - xi и ni

Значение случайной величины, соответствующее отдельной группе наблюдаемых данных, называют значением признака, вариантом (вариантой) и обозначают xi.

Число, которое показывает, сколько раз встречается соответствующее значение признака в ряде наблюдений называют частота значения признака и обозначают n. Сумма всех частот ряда равна количеству элементов в изучаемой совокупности.

Пример дискретного вариационного ряда:

|  |  |
| --- | --- |
| xi (оценка) | ni (кол-во студентов с такой оценкой) |
| 2 | 8 |
| 3 | 12 |
| 4 | 23 |
| 5 | 17 |
| Всего | 60 |

# 11. Числовые характеристики вариационных рядов.

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.

где xi - 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:

где xi - i-е значение признака;

mi - частота 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.

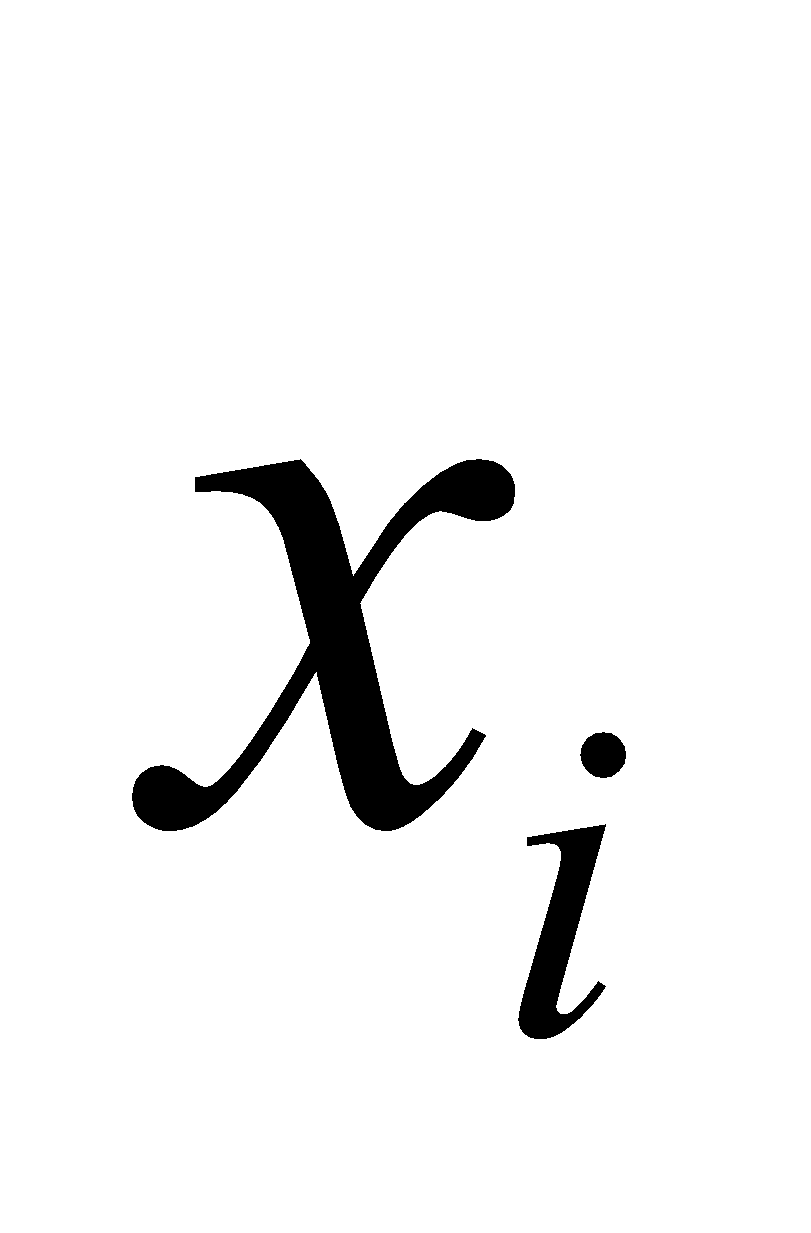
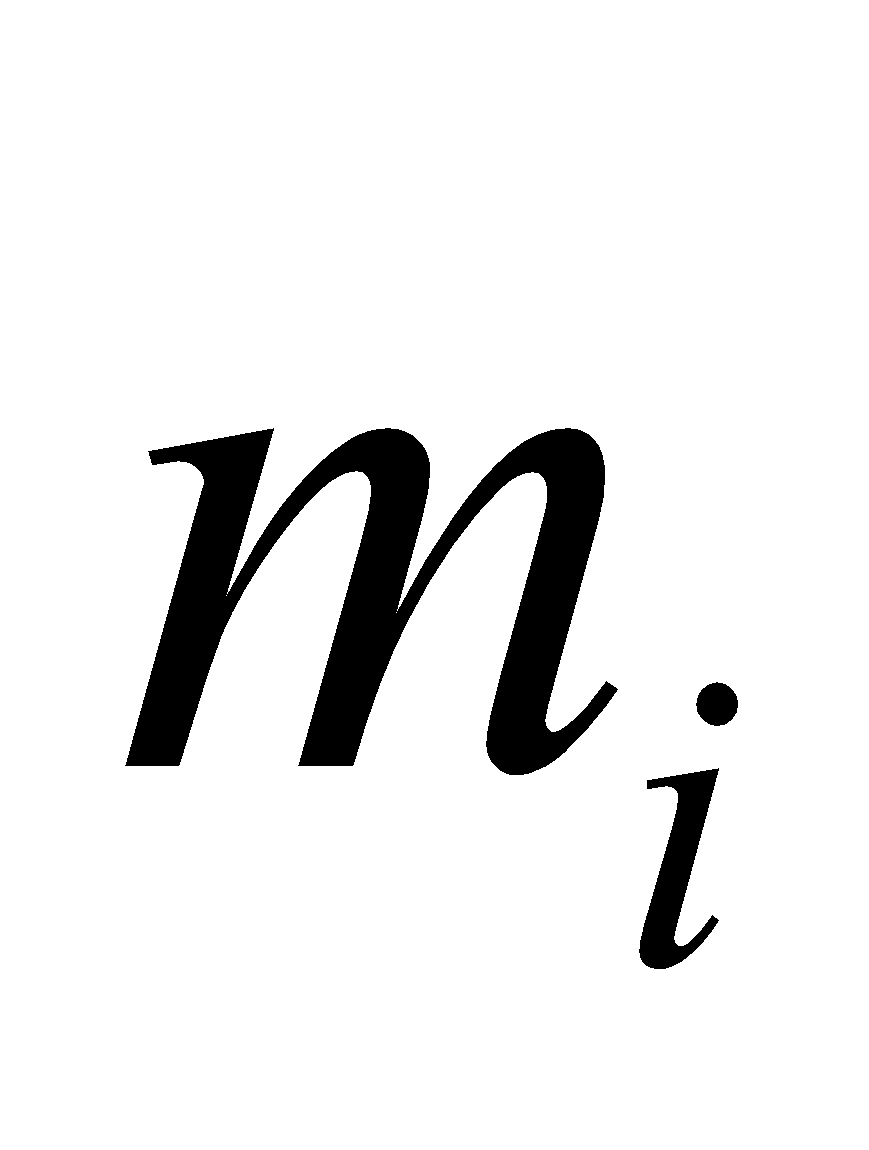
Дисперсия характеризует рассеяние значений (как сильно случайная величина отклоняется от среднего значения). Дисперсию можно рассчитать по простой и взвешенной формуле.

Простая:

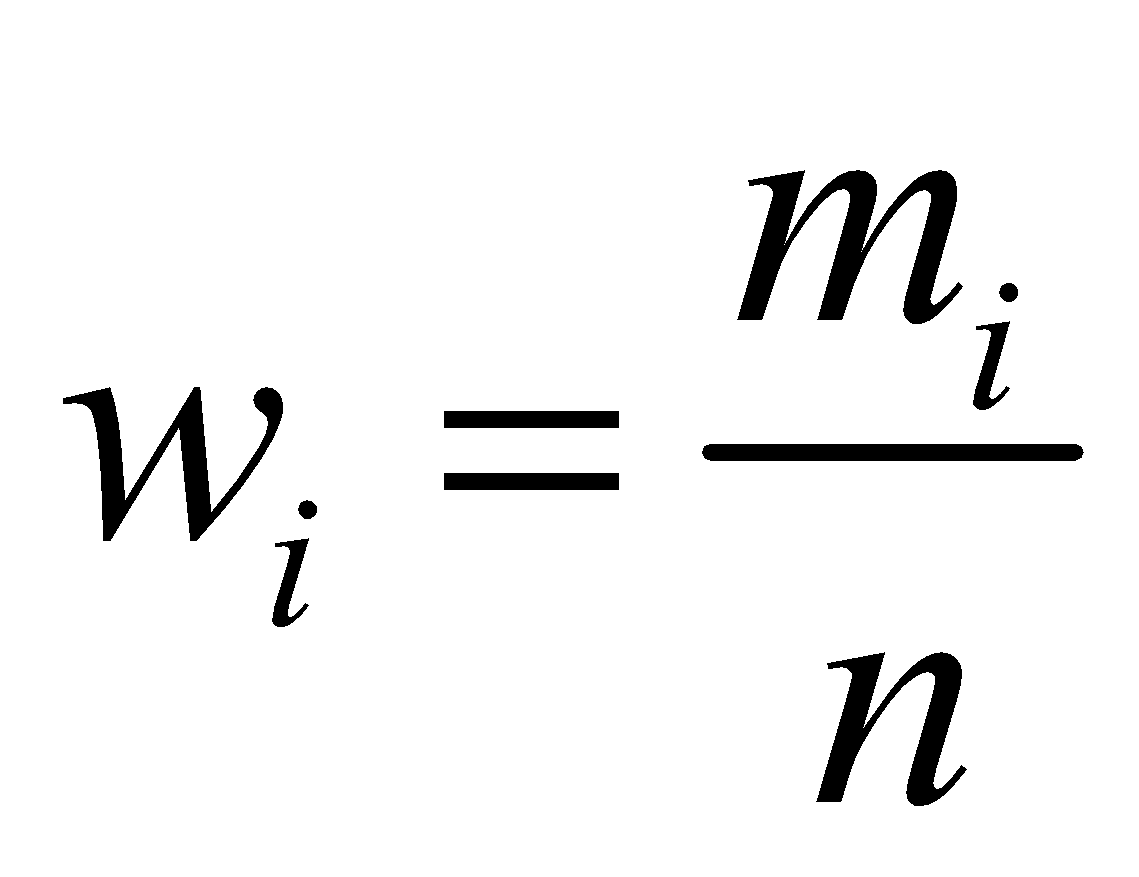
Взвешенная:

Среднее квадратическое отклонение показывает, на сколько в среднем отклоняются конкретные варианты от их среднего значения. Среднее квадратическое отклонение рассчитывается по формуле:

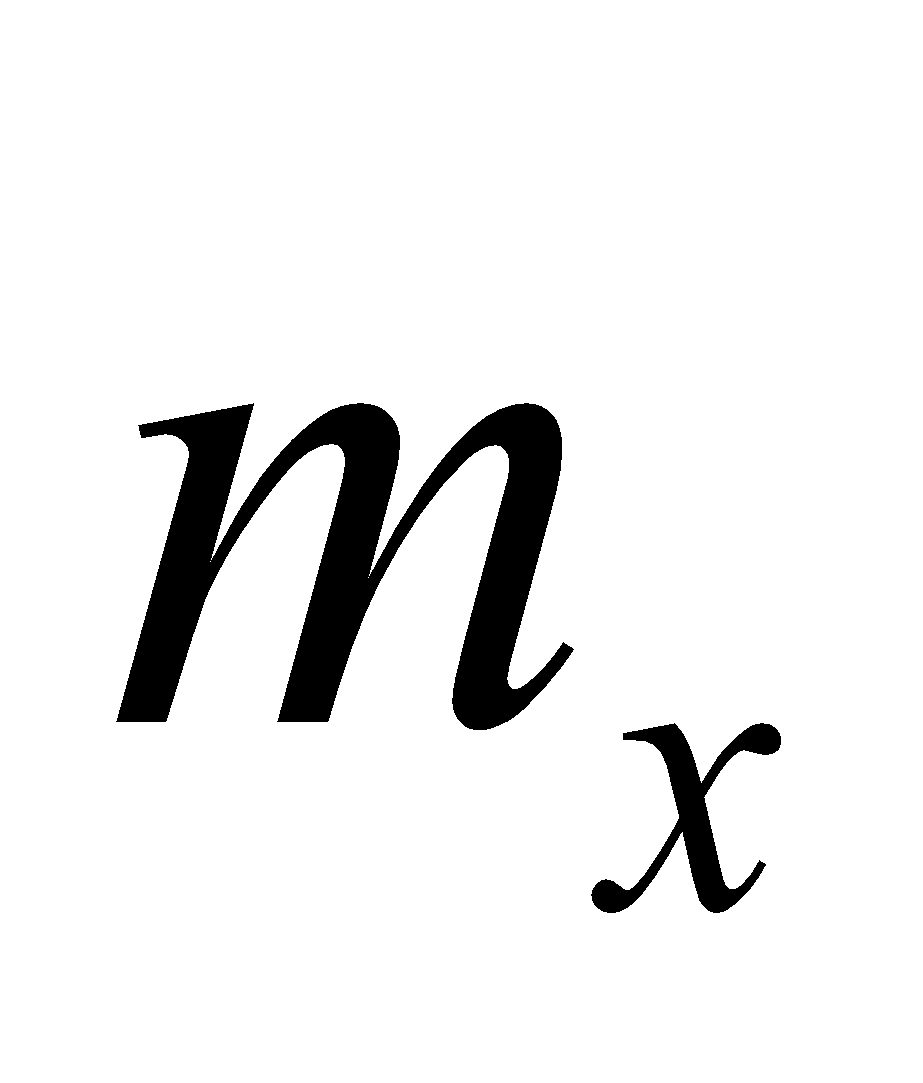
Важным показателем вариационного ряда является также коэффициент вариации, который показывает однородность исследуемого признака. Коэффициент вариации рассчитывается по формуле:

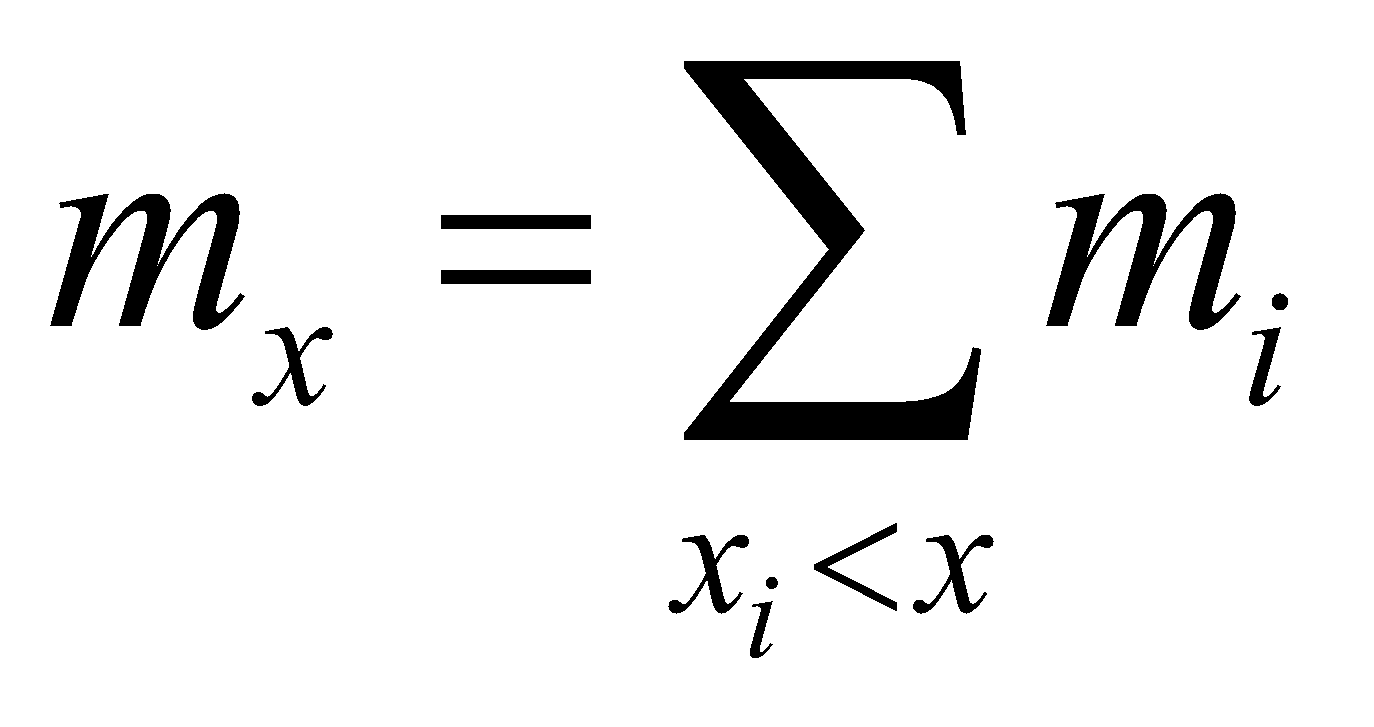
Частотой варианты  называется число , показывающее, сколько раз эта варианта встречается в выборке.

Частостью, относительной частотой или долей варианты называется число

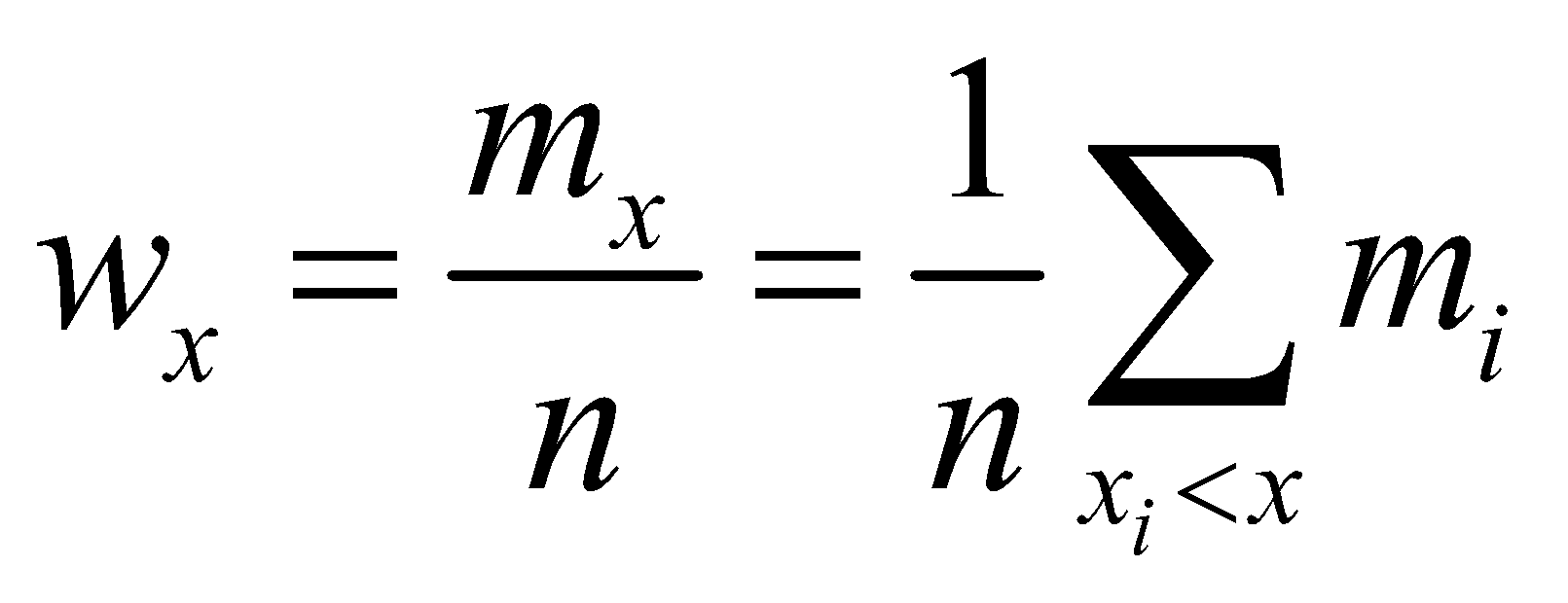


Частоты и частости называются весами.

Пусть х некоторое число. Тогда количество вариант , значения которых меньше х, называется накопленной частотой, т.е.



Отношение накопленной частоты к общему числу наблюдений п называется накопленной частостью:



# 12. Законы распределения случайных величин.

Законы распределения дискретных случайных величин:

1. Закон распределения Бернулли.

Случайная величина X, распределенная по закону Бернулли (индикаторная случайная величина) принимает значения - успех или 0 - неудача, с вероятностями р и q соответственно.

|  |  |  |
| --- | --- | --- |
| xi | 0 | 1 |
| pi | q | p |

1. Биномиальный закон распределения.

Случайная величина X принимает значения: 0, 1, 2, 3, 4, 5,..., n, с вероятностью, определяемой по формуле Бернулли:

1. Закон распределения Пуассона.

Случайная величина *X* принимает бесконечное счетное число значений: 0, 1, 2, 3, 4, *5,..., к,...,* с вероятностью, определяющейся по формуле Пуассона:

1. Геометрический закон распределения.

Вероятность появления m-неудач до первого наступления события *А*

Случайная величина *X,* распределенная по геометрическому закону, принимает значения: 0, 1, 2,..., m *,...,* с вероятностью, определяемой по формуле

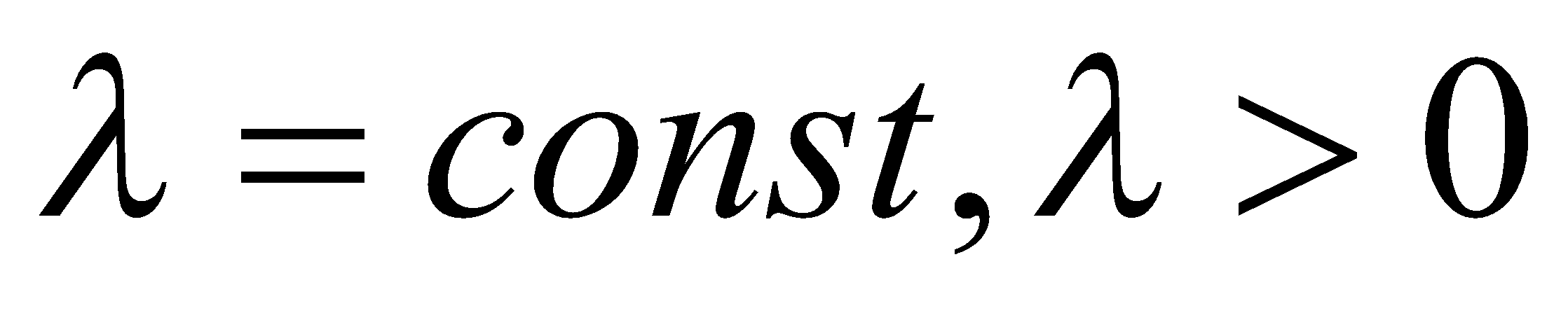
Законы распределения непрерывных случайных величин:

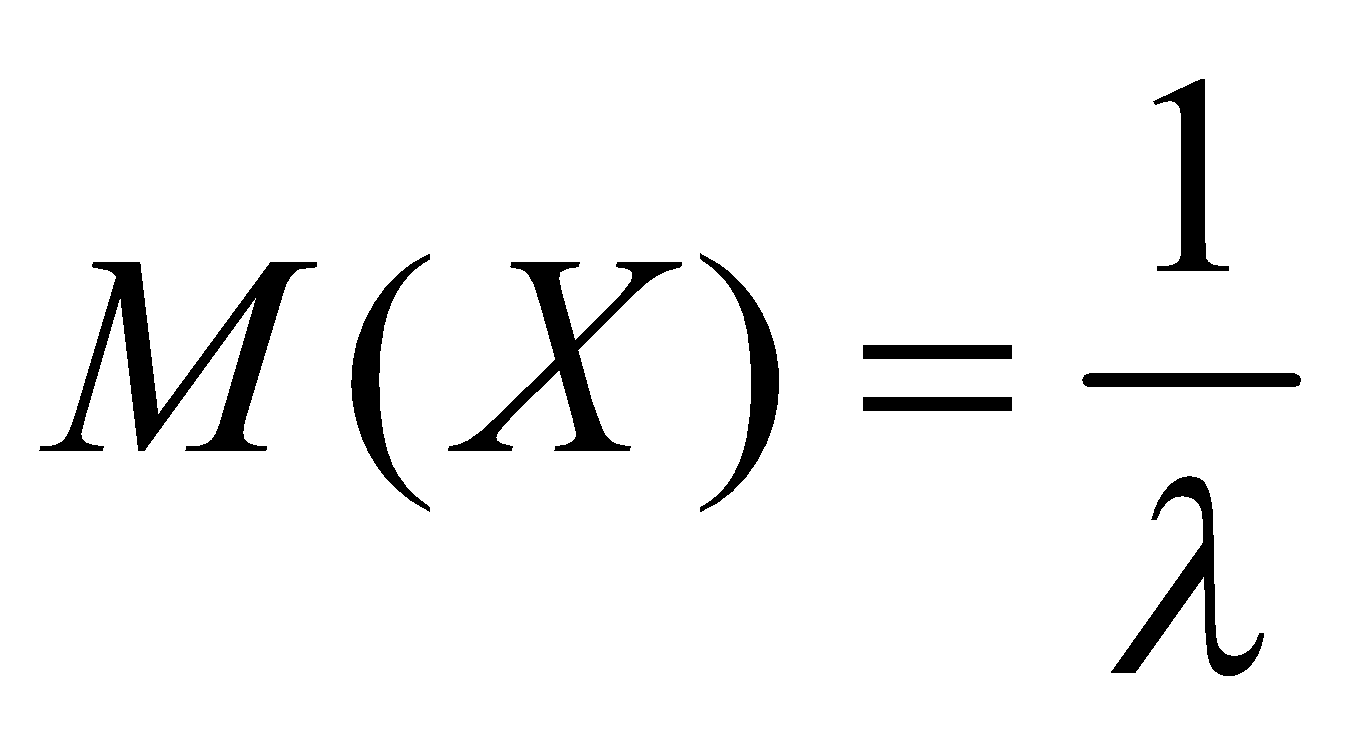
1.Равномерный закон распределения

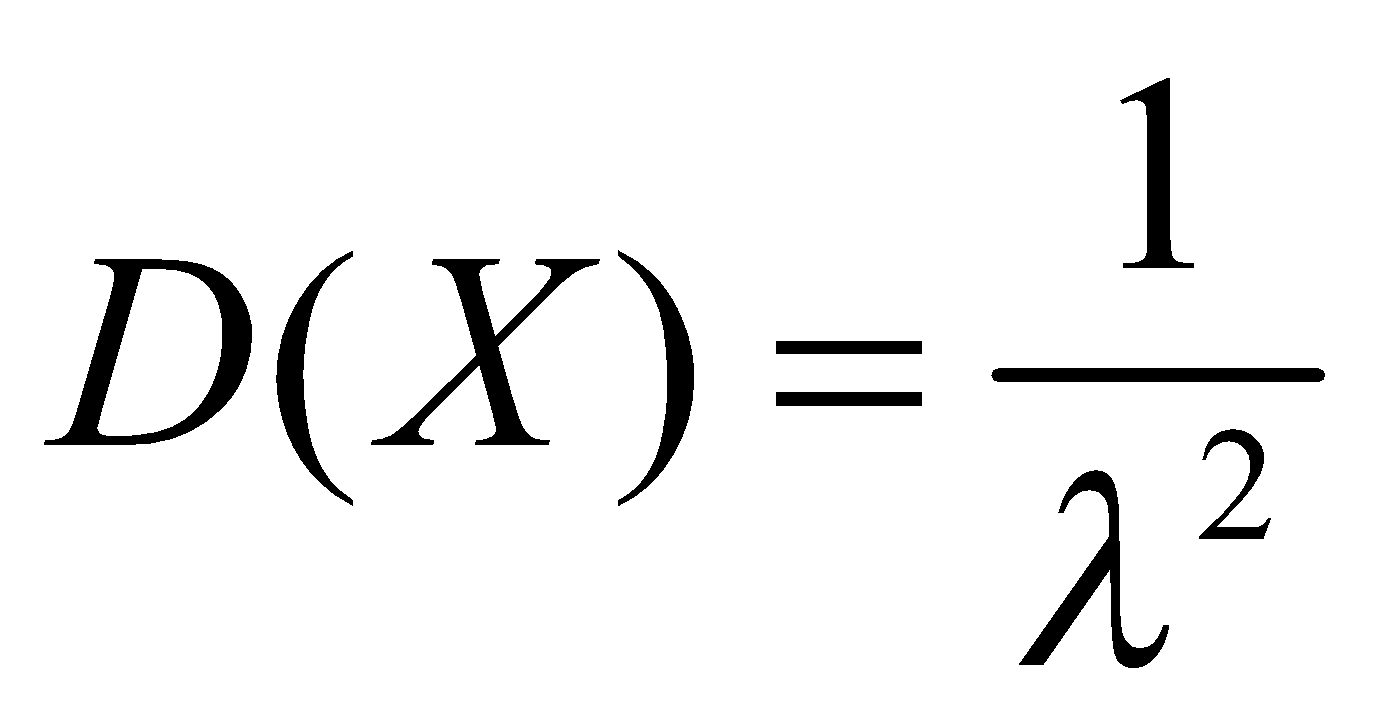
СВ X распределена по равномерному (прямоугольному) закону, если все значения СВ лежат внутри некоторого интервала и все они равновероятны (точнее, обладают одной плотностью вероятности).

2. Показательное распределение.

НСВ X, принимающая неотрицательные значения, имеет показательное распределение, если ее дифференциальная функция имеет вид



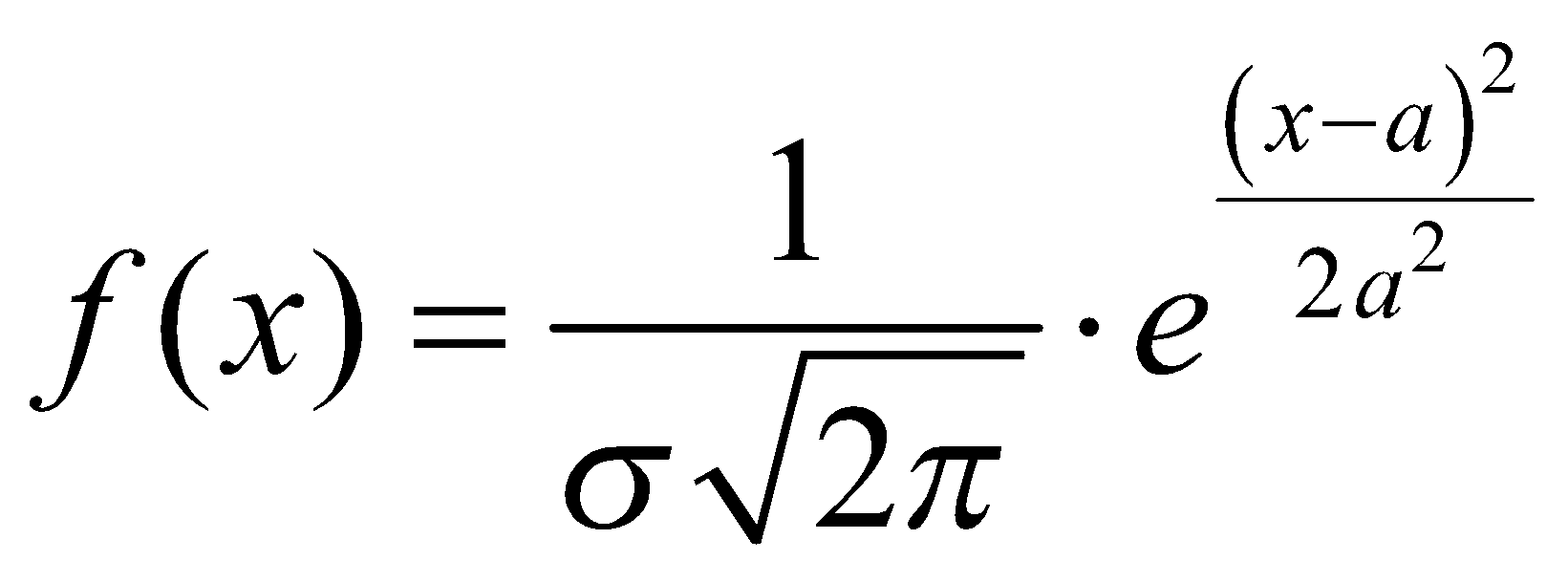




3. Нормальный закон распределения

Играет исключительную роль в теории вероятностей. Это наиболее часто встречающийся закон распределения, главной особенностью которого - то, что он является предельным законом, к которому, при определенных условиях, приближаются другие законы распределения.

Дифференциальная функция нормального закона имеет вид



# 13. Проверка статистических гипотез.

Статистической гипотезой называется любое предположение:

* о законе распределения случайной величины, т.е. предположение о типе закона распределения;
* о значениях его параметров;
* о равенстве параметров нескольких распределений;
* о независимости приведённых выборок и т.п.

Перед проверкой выдвигаемой гипотезы весь огромный массив наблюдения предварительно подвергают обработке, которую на практике обычно называют первичной (сортировка и объединение данных, представление их в удобном для дальнейшей обработки виде, отбраковка грубых и аномальных результатов, определение необходимого объёма выборки) и вторичной (проведение непосредственных вычислений по выдвигаемой гипотезе).

Проверка статистических гипотез осуществляется с помощью различных статистических критериев, значения которых вычисляются на основе имеющихся данных. Причём эти критерии также являются случайными величинами. Из множества возможных значений критерия выбирается подмножество, именуемое критической областью.

Если вычисленное значение критерия принадлежит критической области, то проверяемая (нулевая/основная) гипотеза отвергается. Критическая область выбирается таким образом, чтобы при этом выборе вероятность совершить ошибку первого рода (если мы отвергаем основную гипотезу, тогда как она в действительности верна) не превосходила некоторого заранее определённого положительного числа - уровня значимости и поэтому говорят: “нулевая гипотеза отвергается на уровне значимости ”. Обычно колеблется в пределах , но чаще всего равно .

Вероятность совершить ошибку второго рода (нулевая гипотеза принимается верной, тогда как она в действительности неверна) обозначают буквой . Наилучшей критической областью при данном будет та, для которой — наименьшее.

Вероятность отвергнуть неверную гипотезу равна:

и называется мощностью критерия, т.е. это вероятность правильного принятия решения.

Наиболее распространёнными являются критерии, связанные с использованием известных распределений - квадрат, Стьюдента, Фишера и т.п. Для этих критериев составлены таблицы, в которых приведены критические точки. Соответствующие определённой доверительной вероятности (или уровню значимости ) и определённому числу степеней свободы .

Исключение промахов из наблюдений.

Если серия небольшого числа измерений содержит грубую погрешность-промах, то наличие этого промаха может сильно исказить как среднее значение измеряемой величины, так и границы доверительного интервала. Поэтому из окончательного результата необходимо этот промах исключить.

Для выяснения возникших аномальных результатов наблюдения, прежде всего, необходимо тщательно проанализировать условия измерений и убедиться в полном их соблюдении на протяжении всего опыта. Если было обнаружено нарушение условий, то сомнительный результат не включается в выборку.

Однако вычислителю чаще приходится иметь дело с готовыми данными выборки и поэтому для выявления грубых результатов измерения он обычно использует статистические методы.

# 14. Анализ временных рядов.

Большое количество данных по экономике, коммерции, технике

и т.д. может рассматриваться как в пространстве, так и во времени,

путем построения и анализа одного или нескольких временных

рядов.

Дискретный временной ряд — это последовательность

измерений значений переменной (процесса) за определенный период

через одинаковые промежутки времени:

Z1, Z2, Z3, ..... , Zt, ....... Zn (1)

Последовательные наблюдения в (1) обычно зависимы.

Зависимость можно представить в виде

Zt = f(t) + **Ɛ**i

где t = 1, 2, ...., n;

f - гладкая (непрерывная и дифференцируемая) функция,

характеризующая долгосрочное движение в зависимости от времени -

тренд;

**Ɛ**i - случайный ряд возмущений, наложенный на систематическую

часть.

При наличии во временном ряду тенденции и циклических

колебаний значения каждого последующего уровня ряда зависят от

предыдущих. Корреляционную зависимость между

последовательными уровнями временного ряда называют

автокорреляцией уровней ряда.

Количественно ее можно измерить с помощью линейного

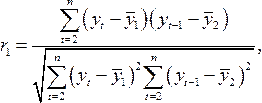
коэффициента корреляции между уровнями исходного временного

ряда и уровнями этого ряда, сдвинутыми на один или несколько шагов

во времени, называемого **коэффициентом автокорреляции**.

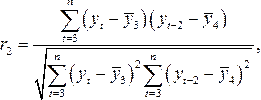
Коэффициент автокорреляции уровней ряда первого порядка,

смещенных на одну единицу времени, определяется по формуле:



http://ok-t.ru/studopediaru/baza9/458757026205.files/image075.gif

Коэффициент автокорреляции уровней ряда второго порядка:



http://ok-t.ru/studopediaru/baza9/458757026205.files/image087.gif

Аналогично можно определить коэффициент автокорреляции

более высоких порядков.

Коэффициент автокорреляции строится по аналогии с линейным

коэффициентом корреляции. Поэтому по нему можно удить о наличии линейной или близкой к линейной тенденции. Чем ближе

коэффициент автокорреляции уровней ряда первого порядка к

единице. Тем более выражена линейная тенденция. Для некоторых

временных рядов, имеющих сильную нелинейную тенденцию,

коэффициент автокорреляции уровней исходного ряда может

приближаться к нулю.

Последовательность коэффициентов автокорреляции уровней

первого, второго и т.д. порядков называют автокорреляционной функцией временного ряда.

Если наиболее высоким оказался коэффициент автокорреляции

порядка t , то ряд содержит циклические, или сезонные колебания с

периодичностью в t моментов времени. Если ни один коэффициент

не является значимым, то делаю вывод, что либо ряд не содержит

тенденции и циклических колебаний, либо содержит сильную

нелинейную тенденцию.

Число периодов или моментов времени, по которым

рассчитывает коэффициент автокорреляции называют ***лагом***.

Построение аналитической функции для моделирования

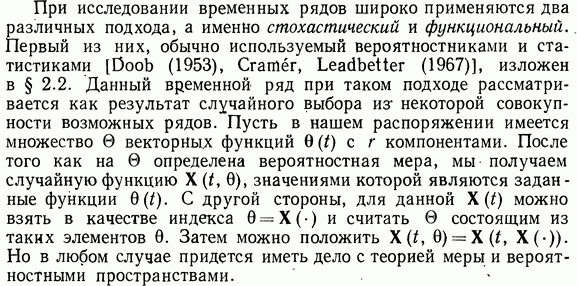
тенденции (тренда) врременного ряда называют аналитическим

выравниванием временного ряда. Тенденция во времени может

принимать разные формы, для ее формирования используют функции, рассмотренные в МНК (метод наименьших квадратов).

**Второй подход. Стохастический.**

[Стохастический процесс](https://economic_mathematics.academic.ru/4374) — [stochastic process] процесс называется стохастическим, если он состоит изслучайных переменных, значения которых меняются во времени. Заложил Эдни Юл в 1927 г.



**Третий подход к анализу временных рядов. Спектральный**

**анализ в частотной области.**

В этом случае можно получить выравнивание по ряду Фурье

(при этом обычно рассматривается не более 5 гармоник (j = 1, 2, 3, 4, 5)). Параметры aj и bj находятся с помощью МНК.

Анализ временных рядов преследует несколько целей:

1. Описание поведения ряда.

2. Построение модели для объяснения наблюдений.

3. Пункты 1) и 2) используют для прогноза, исходя из

предположения о сохранении тенденции развития в будущем.

Для достижения поставленных целей используют модели, основанные

на перечисленных подходах: детерминистском, стохастическом,

спектральном.

В общем случае можно предположить в модели наличие следующих

компонент:

1. тренд или долгосрочное колебание.

2. Регулярное движение относительного тренда.

3. Сезонная компонента.

4. Остаток.

Модель, в которой временной ряд представлен как сумма

перечисленных компонент, называется ***аддитивной моделью***

временного ряда. Если временной ряд представлен как произведение перечисленных компонент, то она называется ***мультипликативной*** ***моделью*** временного ряда.

Отделить тренд и сезонность в общем случае невозможно. т.к.

Они взаимно проникают друг в друга. При выделении тренда и

сезонности остается колеблющийся ряд. Удаление тренда

(сглаживание временного ряда) можно осуществить с помощью скользящей средней (СС). Скользящая средняя в отличие от простой средней для всей выбрки, содержит сведения о тенденциях изменения данных. Отделить тренд и сезонность в общем случае невозможно. Удаление тренда (сглаживание временного ряда) можно осуществить с помощью скользящей средней (СС). Для этого к первым (2m + 1) точкам ряда подбирают степеннйо полином и минимизируют. Затем подбирают полином того же порядка для второго, третьего, ...... (2m + 1) наблюдения.Эта процедура продолжается вдоль всего ряда до последней группы из (2m + 1) точек.

# 15. Корреляция. Корреляционный анализ.

**Корреляционный анализ** — это метод обработки статистических данных, заключающийся в изучении коэффициентов корреляции между переменными.

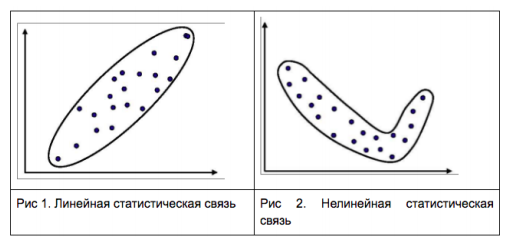
**Проще говоря:** связь между результативными и факторными переменными (урожайностью какой-либо культуры и количеством осадков, ростом и весом человека в однородных группах по полу и возрасту, частотой пульса и температурой тела и т.д.).

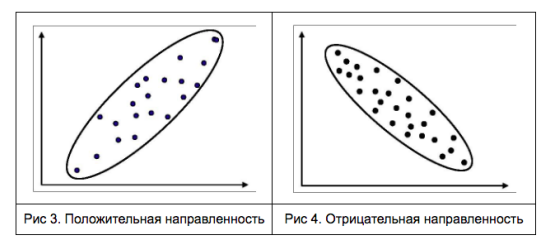
**Ложная корреляция —** корреляция по отношению к признакам абсурдных по отношению друг к другу (например: уровень IQ и размер ноги).

**Задачи корреляционного анализа:**

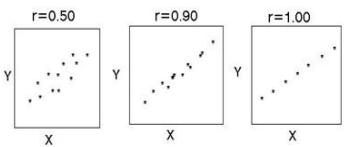
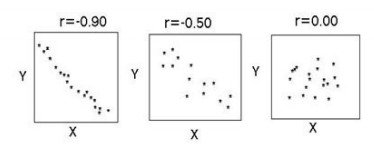
* получение информации об одной из искомых переменных с помощью другой
* определение тесноты связи между исследуемыми переменными
* выявление факторов, оказывающих наибольшее влияние на результативный признак
* выявление неизученных ранее причин связей
* построение корреляционной модели с ее параметрическим анализом
* исследование значимости параметров связи и их интервальная оценка.

**Корреляционное поле** (или диаграмма рассеяния) является графической зависимостью между результатами измерений двух признаков. (Ось х и у. Например: х - курение, у - смерность)





r - коэффициент корреляции

****

*Обратная корреляция Прямая корреляция*

Корреляции нет - точки расположены хаотично

Средняя степень связи - точки более-менее равномерно удалены от медианы.

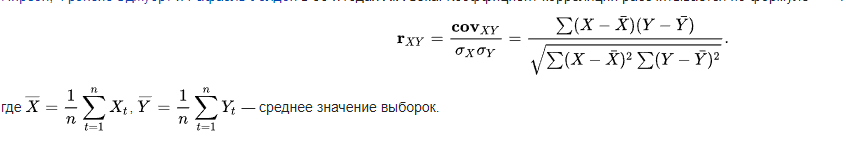
Сильная связь - стремится к прямой и при r=1, ровная линия

Знак коэффициента корреляции определяет направленность взаимосвязи: минус – отрицательная, плюс – положительная

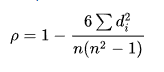


Если распределение является нормальным законом — результатом корреляционного анализа выступают коэффициенты корреляции Пирсона, либо, в случае, если признаки не подчиняются этому закону — коэффициент ранговой корреляции Спирмена.

**Коэффициент корреляции Бравэ-Пирсона** —линейный коэффициент корреляции



**Коэффициент корреляции рангов Спирмена (**rs**)** — ранговый коэффициент корреляции.



Значение коэффициента меняется от −1 (последовательности рангов полностью противоположны) до +1 (последовательности рангов полностью совпадают). Нулевое значение показывает, что признаки независимы.

**Ограничения использования коэффициента корреляции:**

* Нелинейность связи. Решения:

1. Найти точку перегиба и разделить на две части

2. Отказаться от использования коэффициента корреляции.

3. Если выявленная связь является монотонной, то целесообразно использовать ранговые коэффициенты корреляции.

* Дисперсионные выбросы, асимметричность распределения (значения r разбросаны от -1 до 1).Решения:

1. Исключать дисперсионные выбросы из выборки

2. Использовать ранговые коэффициенты корреляции

* Влияние третьей переменной.

**В Exel:**

1. КОРРЕЛ [CORREL] (массив1; массив2) - Коэффициент корреляции. ПИРСОН (PEARSON) с теми же массивами - Линейный коэффициент корреляции

2. СТЬЮДРАСПОБР (вероятность; степени\_свободы)

Оценка значимости коэффициента парной корреляции с использованием t критерия Стьюдента. Рассчитанное значение t-критерия сравнивается с табличной (критической) величиной данного показателя из соответствующей таблицы значений рассматриваемого параметра с учетом заданного уровня значимости и числа степеней свободы.

3. Матрица коэффициентов парной корреляции. Анализ осуществляется с помощью средства «Анализ данных», в котором выбирается «Корреляция».

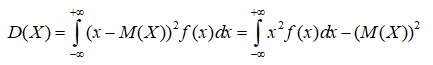
# 16. Дисперсия. Эксцесс. Показатель асимметрии.

Дисперсия служит для характеристики рассеяния случайных величин относительно ее математического ожидания и характеризует форму кривой распределения. Она является более полной оценкой дискретных случайных величин.

D(X) = (X- M(X))2,

где M(X) – математическое ожидание (M(X) =

Дисперсия *НСВ (непрерывных СВ):*



Свойства дисперсии:

1. D(X) = M(X2) – (M(X))2

2. D(C) = 0

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

4. D(X-Y) = D(X) + D(Y)

Дисперсия характеризует средний квадрат отклонения ДСВ, поэтому на практике часто используют в качестве характеристики разброса среднее квадратическое отклонение.

Для распределения Бернулли: D(X)=pq;

Для биномиального закона: D(X) = npq,

Для геометрического закона и для геометрического закона +1: D(X) = q/p2

Для отрицательного биномиального распределения: kq/p2

Для гипергеометрического:

Для распределения Пуассона: D(X) = α

(Теория: обобщенными числовыми характеристиками для случайных величин в теории вероятностей, а также математической статистике являются начальные и центральные моменты. Начальным моментом k-го порядка случайной величины Х называют математическое ожидание от величины в k-ой степени Xk. Центральным моментом k-го порядка называют математическое ожидание от величины (X-M(X)k)).

Третий центральный момент относительно среднего значения характеризует *ассиметрию распределения* относительно математического ожидания и является, следовательно, мерой ассиметрии (скошенности).

Для симметричного распределения M3 = 0. Если M3 < 0, то график плотности ассиметричен и скошен отрицательно — пик смещён вправо, а при M3 > 0 — пик смещён влево. Для удобства вычислений показателю ассиметрии придают безразмерную величину, для этого M3 делят на S3 (куб среднего квадратического отклонения)

Четвёртый центральный момент характеризует свойство островершинности или пологости кривой плотности вероятностей. За характеристику этого свойства принимается безразмерная величина E, называемая *эксцессом* и равная отношению M4 к M2, которое для нормального распределения равно 3.

Если S = 1 и M4 > 3, то график плотности распределения имеет эксцесс, превышающий нормальный (кривая заострена, т.е. E > 0).

Если M4 = 3 — то нормальный эксцесс — средняя заострённость, т.е. E = 0;

Если M4 < 3, то кривая плоская и имеет эксцесс менее нормального, т.е. E < 0.

Значение эксцесса по результатам наблюдений вычисляется по формуле

# 17. Числовые характеристики случайных величин.

- это отдельные числовые параметры распределения.

Это:

I. Характеристики положения ряда распределения:

1. *Математическое ожидание* - это сумма парных произведений случайной величины на соответствующую вероятность. M(X) =

2. *Медиана* - это значение случайной величины, которое делит таблицу распределения на две части таким образом, что вероятность попадания в одну из них равна 0,5. Медиана обычно не определяется для ДСВ.

3. *Мода* - это значение СВ, имеющее наиболее вероятное значение.

II. Характеристики рассеяния:

1. *Дисперсия* - математическое ожидание квадрата отклонения СВ от ее математического ожидания. D(X) = (X- M(X))2

2. *Среднее квадратическое отклонение* - наиболее распространённый показатель рассеивания значений случайной величины относительно её математического ожидания. σ(X)= √(D(X))

# 18. Алгоритм вычисления числовых характеристик выборочного распределения.

Используются для изучения и описания закономерностей распределения вероятностей случайных величин в математической статистике. Основными из них являются средние значения и меры рассеивания. Средними значениями являются: среднее арифметическое (математическое ожидание), медиана и моды, а в качестве мер рассеивания используются дисперсия, среднеквадратическое отклонение, размах, коэффициент вариации.

1. *Среднее арифметическое* является приближённой оценкой теоретического среднего (математического ожидания - M) генеральной совокупности величин и определяется по формуле

2. *Дисперсия*. Степень рассеивания элементов массива вокруг среднего значения характеризуется дисперсией.

3. *Средним квадратическим отклонением* (или стандартным отклонением) величин xi, x1, x2, …, xn от их среднего значения называется корень квадратный из дисперсии

Для эмпирического (S) распределения среднеквадратическое отклонение определяется по формулам:

при n ≥ 25

при n < 25

4. *Коэффициент вариации.*

Среднее квадратическое отклонение имеет ту же размерность, что и элементы исходного массива. Это не даёт возможности сравнить между собой степень рассеяния (изменчивости) разнородных величин. Для этой цели введён коэффициент вариации (коэффициент изменчивости)

# 19. Вычисление оценок по методу моментов.

В теории вероятностей под начальным моментом n-го порядка случайной величины понимается математическое ожидание n-й степени, которое для непрерывного и дискретного распределений записывается в виде:

*Центральные моменты:*

*1. Среднее арифметическое* (или математическое ожидание) называется первым моментом. Математическое ожидание — это точка на оси абсцисс, обладающая тем свойством, что момент левой ветви функции плотности вероятностей относительно этой точки равен моменту её правой ветви.

Первый момент характеризует как бы центр тяжести функции плотности вероятностей P(x).

*2. Дисперсия* является математическим ожиданием квадрата отклонения случайной величины от её среднего значения — математического ожидания и называется вторым центральным моментом или моментом инерции.

*3.* Третий центральный момент относительно среднего значения характеризует *ассиметрию распределения* относительно математического ожидания и является, следовательно, мерой ассиметрии (скошенности).

Для симметричного распределения M3 = 0. Если M3 < 0, то график плотности ассиметричен и скошен отрицательно — пик смещён вправо, а при M3 > 0 — пик смещён влево. Для удобства вычислений показателю ассиметрии придают безразмерную величину, для этого M3 делят на S3 (куб среднего квадратического отклонения)

4. Четвёртый центральный момент характеризует свойство островершинности или пологости кривой плотности вероятностей. За характеристику этого свойства принимается безразмерная величина E, называемая *эксцессом* и равная отношению M4 к M2, которое для нормального распределения равно 3.

Если S = 1 и M4 > 3, то график плотности распределения имеет эксцесс, превышающий нормальный (кривая заострена, т.е. E > 0).

Если M4 = 3 — то нормальный эксцесс — средняя заострённость, т.е. E = 0;

Если M4 < 3, то кривая плоская и имеет эксцесс менее нормального, т.е. E < 0.

Значение эксцесса по результатам наблюдений вычисляется по формуле

Дополнительными характеристиками функции плотности распределения вероятностей являются мода и медиана.

*Модой* называется число на оси абсцисс, соответствующее наибольшей плотности вероятности. Для определения оценки моды на основе опытных данных находится то результат наблюдений, который встречается наиболее часто.

Если данные сгруппированы по интервалам частот, то в качестве моды берётся середина интервала, содержащего наибольшее число наблюдений (наибольшую вероятность). Мода является второй характеристикой центра распределения случайной величины.

Третьей характеристикой центра распределения является срединная точка или *медиана*. Медианой M8 непрерывной случайной величины называется такое значение, для которого функция распределения равна 0,5

Геометрически медиана является абсциссой такой точки кривой плотности вероятности f(x), ордината которой делит площадь под кривой на две равные части.

# 20. Квантили, квартили, перцентили

**Квантили** вариационного ряда – это варианты, занимающие определенное место в ранжированной совокупности.

***P -ый перцентиль*** вариационного ряда – это значение признака, слева от которого лежит P% вариантов ряда. Позиция P -го перцентиля задается как

(n + 1) P /100, где n – число вариантов ряда.

**Перцентиль**– это значение признака в определенной позиции ранжированного ряда, мера относительной позиции варианта в ряду.

**Квартили** – квантили, которые делят вариационный ряд на четыре равные части первый квартиль, второй квартиль, третий квартиль четвертый квартиль (обозначаются Q1, Q2, Q3, Q4).

Первый квартиль (25-й перцентиль) = это значение признака в вариационном ряду, слева от которого лежит ¼ (или 25%) всех вариантов.

Второй квартиль (50-й перцентиль). Он называется *медианой* и обозначается *Ме*.

**Медиана** – значение признака ряда, относительно которого вариационный ряд делится на две равные по числу вариантов части. Это – 50-й перцентиль.

Третий квартиль – это точка, слева от которой находится ¾ или 75% вариантов ряда.

25-й перцентиль называют нижним квартилем, 50-й перцентиль (медиану) – средним квартилем, 75-й перцентиль – верхним квартилем.

**Децили** – квантили, которые делят вариационной ряд на десять равный частей (10, 20, … , 90 перцентили.)

****

***Мода*** – это значение признака, наиболее часто встречающееся в вариационном ряду. Обозначается Мо.

Для интервального вариационного ряда:

1. сначала определить модальный интервал (по максимальной частоте),
2. затем — значение модальной величины признака по формуле:



где:

* — значение моды
* — нижняя граница модального интервала
* *i* — величина интервала
* — частота модального интервала
* — частота интервала, предшествующего модальному
* — частота интервала, следующего за модальным

**Общие формулы для квантилей:**

Рассчитать номер квантиля: 

Если ряд дискретный, то рассчитать значение квантиля по формуле:

+ (

Если ряд интервальный, то рассчитать значение квантиля по формуле:



где:

* – квантиль,
* – номер квантиля,
* j– порядковый номер квантиля,
* - сумма всех частот (количество элементов в совокупности),
* – размерность квантиля (на сколько частей эти квантили делят совокупность),
* - нижняя граница квантильного интервала,
* – ширина квантильного интервала,
* - накопленная частота предквантильного интервала,
* - частота квантильного интервала.

**Медиана для интервального вариационного ряда:**

1. Определяется номер медианы по формуле: , полученное значение округляется до целого большего числа.
2. Затем по накопленной частоте определяется интервал, в который входит элемент с таким номером,
3. Затем — значение медианы по формуле:



где:

* — искомая медиана
* — нижняя граница интервала, который содержит медиану
* *i* — ширина интервала
* — сумма частот или число членов ряда
* - накопленная частота интервала, предшествующего медианному
* — частота медианного интервала

**Медиана для дискретного вариационного ряда:**

1. Сначала *порядковый номер медианы* по формуле: 
2. Затем определяют, какое значение признака обладает накопленной частотой, равной номеру медианы:
3. Если ряд содержит **четное** число элементов, то медиана будет равна средней из двух значений признака, находящихся в середине.
4. **Если ряд содержит нечетное** число элементов, определяют, какое значение варианта обладает накопленной частотой, равной номеру медианы.

**Примеры:**

1. Определить 25-й, 50-й и 90-й перцентили вариационном ряду.

\*тут должен быть какой-то ряд, но в лекции его нет, а решение есть :)\*

**? ? ? ? 13 14 ? ? ? ? ? 15 15 ? ? ? ? ? ? ? ? ? 19 21 ? ?**

Всего 26 значений (n = 26)

Определяем 25-й перцентиль.

1) Найти его позицию в вариационном ряду:

(n + 1) P /100 = (26 + 1) 25 /100 = (27) (0,25) =6,75.

Эта позиция находится между шестым и седьмым вариантами.

Шестой по порядку вариант в ранжированном ряду равен 13, седьмой – 14.

Значение перцентиля находится в точке, которая делит расстояние между 13

и 14 в соотношении 0,75 к 1, расстояние от 13 до 25-го перцентиля

составляет 0,75 от длины отрезка между 13 и 14.

Итак, 25-й перцентиль равен 13,75.

Найдем 50 перцентиль.

1) Найти значение варианта, соответствующего позиции:

(n + 1) P /100 = (26 + 1) 50 /100 = (27) (0,5) =13,5.

Просмотрим ранжированные варианты.

Видим, что значение 13-го по порядку варианта равно 15.

Значение 14-го по порядку варианта также равно 15.

Отсюда 50-й перцентиль равен 15.

Определяем 90-й перцентиль.

Определяем 90-й перцентиль как значение варианта, соответствующего

позиции

(n + 1) P /100 = (26 + 1) 90 /100 = (27) (0,9) =24,3.

Значение 24-го варианта равно 19, а 25-го равно 21, следовательно,

расстояние от 19 до 90-го перцентиля составляет 0,3 от длины отрезка между

19 и 21 (длина отрезка равна 2). Поэтому 90-й перцентиль равен 19,6.

Верхний квартиль – это точка, соответствующая позиции

(26 + 1) 75 /100 = (27) (0,75) =20,65. Его значение равно 16,65.

Шесть значений равно 14. Значение признака, равное 14

встречается наиболее часто. Следовательно мода равна 14.

**2.** По данным таблицы вычислить медиану.

*Сумма денег, израсходованные на покупки товаров в отделе верхней одежды*

Таблица

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Интервалы расходов | 100-300 | 300-500 | 500-700 | 700-900 | 900-1100 | 1100-1300 |
| Число покупателей mi | 30 | 38 | 50 | 31 | 22 | 13 |
| Доля покупателей νi | 0.163 | 0.207 | 0.272 | 0.168 | 0.120 | 0.070 |

Решение:

Ме = 500 + 200•((0,5•184 -68)/50) = 596.

**3.** По данным таблицы найти моду.

Решение:

Мо = 500 + 200•((50 - 38)/((50 -38) + (50 - 32)) = 577,42.

Вместо частот при вычислении квантилей и моды можно использовать частости.