МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЗЕМЛЕУСТРОЙСТВЕ

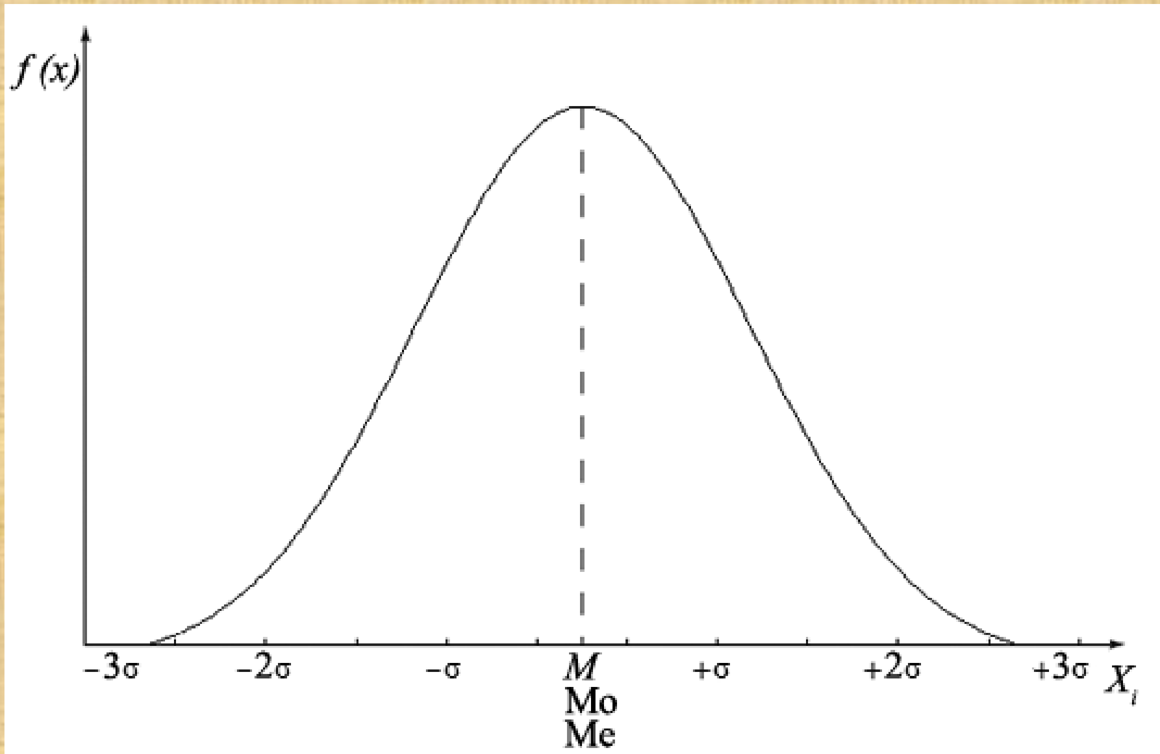
Карпиченко Александр Александрович

доцент кафедры почвоведения и геоинформационных систем

Всего 46 ч., в т.ч. лекций 24 ч., лабораторных 18 ч., УСР 4 ч.

1.2. Обработка вариационного ряда

После удаления артефакта приступают к вычислению показателей ***описательной статистики*** при условии, что тип распределения вариант соответствует **нормальному** или **логнормальному закону распределения**. *В иных случаях с выборкой работают как с непараметрической, на которые теория вероятности не распространяется*.



1.2. Обработка вариационного ряда

При установлении типа распределения принимается следующий порядок действий. Сначала определяется величина классового интервала *i*, которая зависит от принятого числа классов *k* и объема выборки *N*:

$$i = (x_{max} - x_{min}) / k.$$

Число классов в зависимости от объема выборки определяется по формуле:

$$k = 1 + 3{,}3 \lg N. \qquad (1.1)$$

Исходя из формулы (1.1), можно рекомендовать следующее число классов в зависимости от объема выборки:

N	30–50	51–10	101–400	401–1000	1001–2000
k	4–5	6–7	8–9	9–10	11–12

1.2. Обработка вариационного ряда

Величина классового интервала должна быть одинаковой на протяжении всего вариационного ряда. Границы классов выбираются такими, чтобы каждая варианта могла быть отнесена только к одному классу. Примеры правильной границы классов: 5–9, 10–14, 15–19 или 5,1–9,1, 9,2–13,2, 13,3–17,3, первый и последний классы могут быть неполными. Границы классов желательно выбирать так, чтобы крайние варианты ряда по возможности оказались ближе к середине интервала своего класса.

Пусть в выборке объемом N = 64 по количеству осадков за время наблюдения x_{max} = 179 мм, x_{min} = 103 мм. Согласно формуле (1.4), вариационный ряд разбиваем на 8 классов. Затем находим классовый интервал:

$$i = (179 – 103) / 8 = 9,5, \text{ или округленно } 10.$$

1.2. Обработка вариационного ряда

Исходя из величины классового интервала и минимального значения в выборке, за начало левой границы первого класса удобно принять величину 100. Прибавляя к 100 классовый интервал 10, получаем левые границы последующих классов: 110, 120, 130, 140, 150, 160, 170 мм. Правые границы классов должны отличаться на единицу точности наблюдения от левой границы следующего класса, чтобы граничные значения вариант были отнесены к определенному классу. В нашем примере точность измерения составляет 1,0 мм, поэтому правые границы классов будут следующими: 109, 119, 129, 139, 149, 159, 169, 179.

Срединное значение класса (x) вычисляем путем сложением границ классов и делением суммы на два. Для первого класса срединное значение равно: (100 + 109) / 2 = 104,5. Срединное значение последующих классов определяется путем последовательного прибавления классового интервала к срединному значению предыдущего класса: 104,5 + 10= 114,5.

1.2. Обработка вариационного ряда

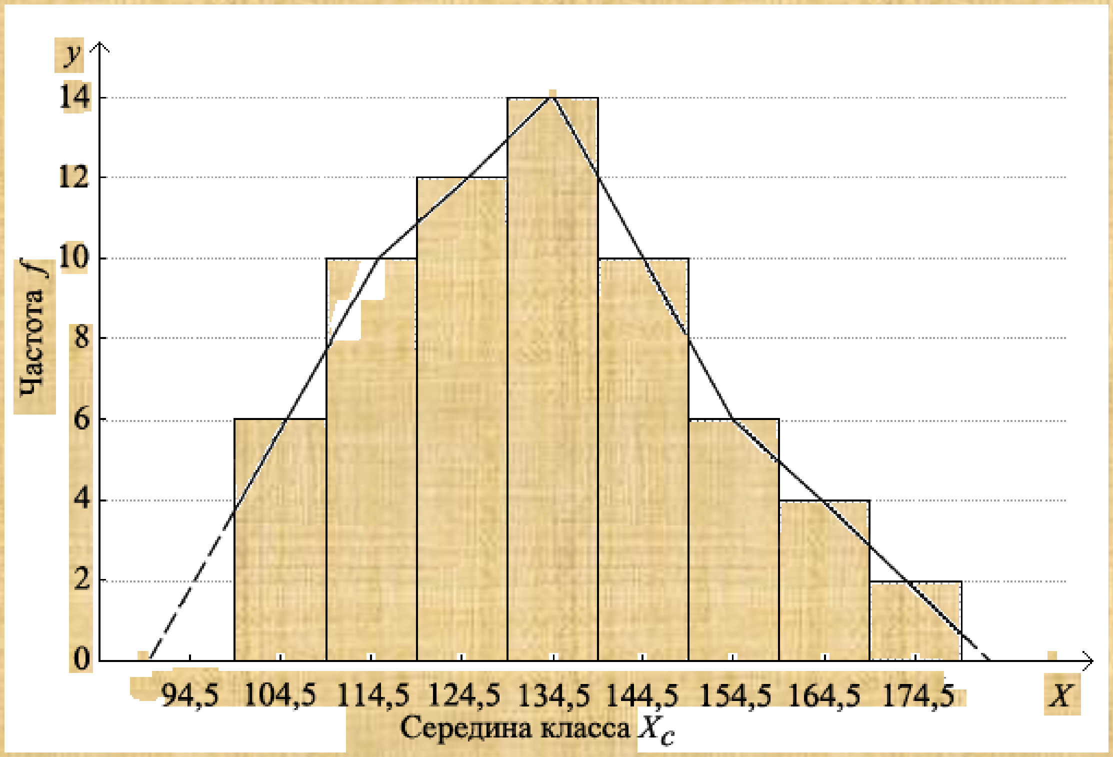
Затем производим разноску вариант по классам (подсчитываем количество вариант, вошедших в тот или иной класс в зависимости от их абсолютных величин). Получаем частоту (f) класса (см. табл. 1.1). Сумма частот должна соответствовать объему выборки (64), сумма частостей *fч'* (частота, выраженная в процентах) должна равняться 100 %.

Группировка вариант в классы при дискретной изменчивости признака

Границы класса	Середина класса, x	Частота, f	Частость, $f_ч$, %
100–109	104,5	6	9,37
110–119	114,5	10	15,62
120–129	124,5	12	18,75
130–139	134,5	14	21,87
140–149	144,5	10	15,62
150–159	155,5	6	9,37
160–169	165,5	4	6,25
170–179	175,5	2	3,12
$i = 10$	$k = 8$	$N = 64$	$\sum$ 100,00

1.2. Обработка вариационного ряда

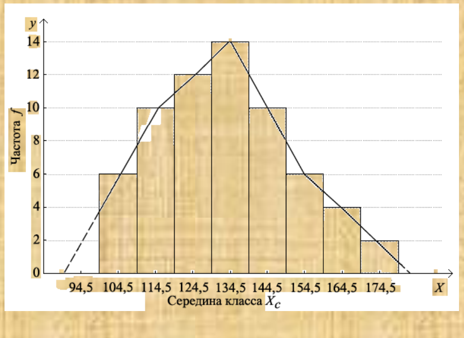
По частоте и середине класса представим вариационный ряд графически в виде полигона и кривой распределения частот.



Способы графического представления вариационного ряда: кривая распределения и гистограмма

1.2. Обработка вариационного ряда

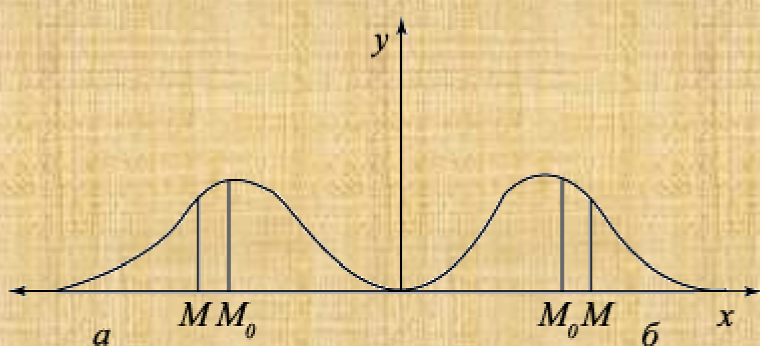
При построении вариационной кривой по оси абсцисс откладываются значения середины класса, по оси ординат – частоты. При построении гистограммы по оси абсцисс откладываются границы классов, а число вариант каждого класса обозначается высотой или площадью соответствующего прямоугольника. При сравнении изменчивости одинаковых условий или признаков полученные вариационные кривые распределения частот наносятся на один график. Группировка вариант в классы для сравниваемых выборок должна быть одинаковой. Если объем выборок не одинаков, все частоты должны быть выражены в процентах от объема выборки по каждой совокупности.

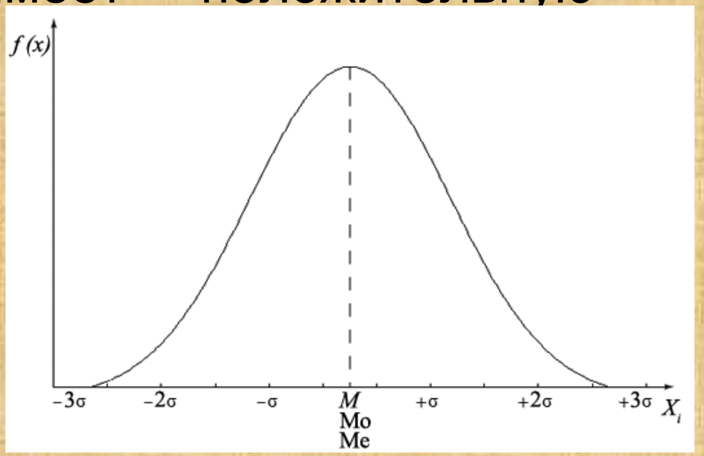


1.2. Обработка вариационного ряда

Показатели асимметрии и эксцесса. Распределение частот в изучаемом объекте не всегда подчиняется закону нормального распределения. Это особенно четко проявляется при выражении вариационного ряда в виде графика. Распределение частот может быть представлено *асимметричной*, *островершинной* или *туповершинной* кривой.

Асимметрия кривой распределения обусловлена неравномерным размещением вариант по обе стороны от модального значения признака. Если число вариант больше справа от моды, распределение имеет положительную асимметрию, если слева – отрицательную.





а – отрицательная асимметрия, *б* – положительная асимметрия

1.2. Обработка вариационного ряда

При получении асимметричной кривой следует проверить асимметричность распределения. Если асимметричность не будет доказана по критерию Стьюдента, то рассматриваемое распределение относят к симметричному. Для проверки асимметричности распределения вычисляют коэффициент асимметрии, его ошибку, затем на основании показателя достоверности устанавливают вид кривой распределения. Коэффициент асимметрии находят:

$$K_{as} = (M - Mo) / \sigma, \text{ или } K_{as} = (M - Me) / \sigma.$$

1.2. Обработка вариационного ряда

При изучении содержания подвижного бора в дерново-подзолистых почвах были получены следующие показатели: М = 0,25 мг/кг, Мо = 0,28, σ= 0,02, *N* = 20. Для получения представления о форме кривой распределения бора предварительно вычисляем коэффициент асимметрии:

$$K_{as} = (0{,}25 - 0{,}28) / 0{,}02 = -1{,}5.$$

Полученная величина указывает на наличие отрицательной асимметрии в распределении вариант содержания подвижного бора в дерново-подзолистых почвах. Затем находим ошибку коэффициента асимметрии:

$$m_{as} = I = \sqrt{6/(20+3)} = 0{,}51.$$

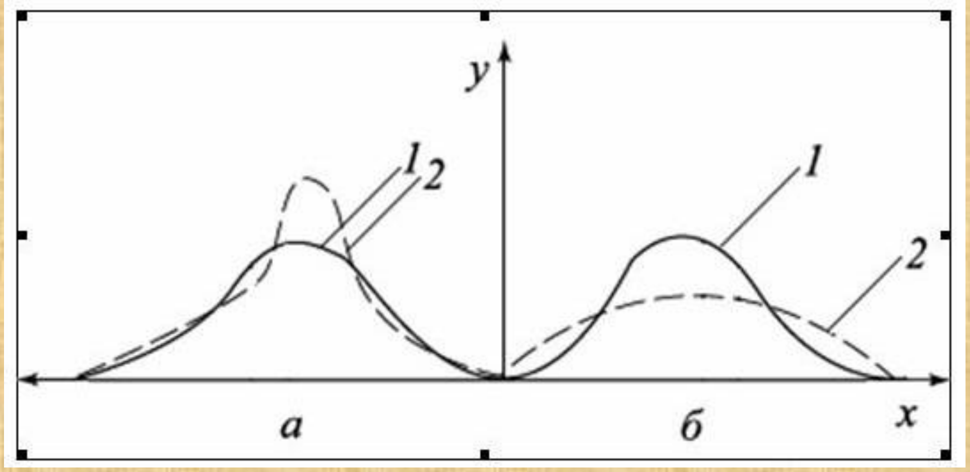
Достоверность коэффициента асимметрии определяется по критерию Стьюдента: $t = K_{as} / m_{as} = -1{,}5 / 0{,}51 = -2{,}94$.

Величина критерия Стьюдента (см. прил. 4) для $P_{0,99}$ при $v \to \infty$ составляет 2,58 (число степеней свободы принимается равным бесконечности). Рассчитанный критерий Стьюдента (2,94) больше табличного для $P_{0,99}$ (2,58), что указывает на асимметричность распределения подвижного бора. Если бы расчетная величина критерия Стьюдента была меньше табличной, то распределение отнесли бы к симметричному даже при наличии незначительной асимметрии.

1.2. Обработка вариационного ряда

Эксцесс кривой распределения *(Е)* имеет место в тех случаях, когда большинство вариантов совокупности сосредоточено около среднего арифметического. Тогда эмпирическая кривая распределения отклоняется от нормальной теоретической кривой у ее вершины и количественно выражается показателем эксцесса (рис. 1.3).

Положительный эксцесс представлен островершинной кривой (эксцессивной, или лептокуртичной) (см. рис. 1.3, *а*), отрицательный – плосковершинной (депрессивной, или платикуртичной) (см. рис. 1.3, *б*). При сильном отрицательном эксцессе кривая может приобрести вид двухвершинной:



1 – теоретическая линия распределения, *2* – эмпирическая линия распределения

1.2. Обработка вариационного ряда

Показатель эксцесса определяется по формуле:

$$E = [\sum (x - M)^4 / N \cdot \sigma^4] - 3.$$

Вычисляют ошибку коэффициента эксцесса: m_E = 2 Оценка достоверности показателя эксцесса производится аналогично оценке показателя асимметрии по критерию Стьюдента: $t = E / m_E$.

Оценить достоверность показателей эксцесса и асимметрии можно более простым способом. **Отклонение эмпирического ряда по асимметрии и эксцессу от нормального распределения считают существенным, если K_{as} и E более, чем в 3 раза превышают свои ошибки (m_{as}, m_E).** Если показатель эксцесса меньше –2, это указывает на наличие в выборке вариант, относящихся к разным совокупностям. Эксцесс считается незначительным, если $|E| < 0,4$. Чем меньше показатель эксцесса, тем ближе распределение к нормальному.

Асимметрия и эксцесс эмпирических кривых указывают иногда на важные особенности объекта исследования, например, на изменение признака в ходе усовершенствования технологии на предприятии при выпуске той же продукции. В таких случаях изучение степени и характера асимметрии и эксцесса вариационных кривых может быть самостоятельной задачей при проведении исследовательских работ.

1.3. Показатели описательной статистики

Одна из основных задач статистической обработки – *нахождение параметров, представляющих в обобщенном виде распределение данной статистической совокупности*. Для решения этих задач используются методы описательной статистики.

1.3. Показатели описательной статистики

Статистические показатели распределения

Показатели	Назначение показателей	Показатели распределения
Центра распределения (средние величины)	**Описывают положение середины распределения**	***Структурные (непараметрические) средние*: мода (Мо) и медиана (Ме).** ***Степенные средние*:** **среднее арифметическое (M),** **среднее гармоническое ($M_{гар}$),** **среднее геометрическое ($M_{г}$),** **среднее квадратическое ($M_{кв}$),** **среднее кубическое ($M_{куб}$),** **среднее взвешенное ($M_{взв}$)**
Рассеивания вариант	**Описывают степень разброса (вариабельности, изменчивости) вариант**	**Лимит (lim)** **Размах варьирования (*ampl*)** **Среднеквадратическое отклонение (σ)** **Дисперсия (σ^2)** **Коэффициент вариации (V)** **Квантили ($V_{0,25;\ 0,5;\ 0,75}$)**
Формы распределения	**Описывают симметрию и островершинность распределения вариант около центра**	**Коэффициент асимметрии (*As*)** **Эксцесс (*E*)** **Гистограмма** **Полигон распределения**

1.3. Показатели описательной статистики

Показатели центра распределения. Для обоснования представления о генеральной совокупности на основании выборки необходимо использовать наиболее характерные параметры признаков. К ним относятся показатели центра распределения, или среднего положения: мода, медиана, среднее арифметическое, гармоническое, геометрическое, квадратическое, кубическое, взвешенное. Средняя величина выражает характерную, типичную для данного ряда величину признака и является равнодействующей всех факторов, влияющих на признак. В ней погашаются индивидуальные различия вариант в ряду, обусловленные случайными обстоятельствами.

Мода (Мо) представляет собой наиболее часто встречающуюся варианту в вариационном ряду. На графике она соответствует максимальной ординате и находится на вершине вариационной кривой. Если вариационный ряд разбит на классы, то мода соответствует максимальной частоте класса, который называется *модальным*. При полимодальном (многовершинном) распределении вариационный ряд имеет несколько значений моды.

1.3. Показатели описательной статистики

Медиана **(Ме)** представляет собой среднюю варианту в ранжированном вариационном ряду, которая делит его на две равные части. При нечетном числе вариант середину ряда будет составлять одна варианта (медиана). При четном числе вариант середину ряда образуют две варианты, среднее арифметическое которых будет характеризовать медиану.

При наличии в вариационном ряду сильно отличающихся вариант медиана будет характеризовать середину ряда более точно, чем среднее арифметическое. Мода и медиана используются в тех случаях, когда о выборочных параметрах необходимо иметь ориентировочное представление.

1.3. Показатели описательной статистики

Среднее арифметическое (*M*, $\bar{x}$) представляет собой величину, сумма положительных и отрицательных отклонений от которой равна нулю. Оно является основной характеристикой статистической совокупности и вычисляется по формуле:

$$M = \sum x_i / N,$$

где $\sum x_i$ – сумма всех вариант совокупности. Среднее арифметической рассчитывается в тех случаях, когда противопоказано вычислять другие средние.

Пример. Определено следующее количество осадков, выпавших в трех пунктах наблюдений: 10, 15 и 20 мм (N = 3). Среднее арифметическое равно: M = (10+15+20) / 3 = 15 мм.

1.3. Показатели описательной статистики

Среднее гармоническое* (*$M_{гар}$*)** вычисляется при усреднении меняющихся скоростей процессов (скорость течения воды), показателей обратно пропорциональной зависимости между процессами и явлениями, сложных абсолютных величинах измерений (тонна/километр). Оно рассчитывается по формуле: ***$M_{гар} = N / \sum (1 / x_i)$. Его величина меньше, чем средней арифметической. Для вычисления сохраним те же количественные варианты, что и для определения среднего арифметического.

Пример. При измерении скорости воды в реке на трех створах русла получены следующие результаты: 10, 15 и 20 м/с:

$$M_{гар} = 3/ \sum [(1:10) + (1:15) + (1:20)] = 13{,}8 \text{ м/с}.$$

1.3. Показатели описательной статистики

***Среднее геометрическое* (*$M_г$*)** вычисляется в тех случаях, когда в вариационном ряду отдельные значения распределяются в геометрической прогрессии (резко различаются между собой, например, 4 и 16). В данном случае среднее геометрическое равно 8. Оно в два раза меньше 16 и в два раза больше 4. Среднее арифметическое из этих вариант 10, т. е. больше среднего геометрического. При наличии нулевой варианты рассчитывается приближенное среднее арифметическое. Если варианты представлены логарифмами чисел (pH и др.), то вычисляют **среднее логарифмическое**.

Пример. Строение стоит 100 тыс. у.е. Одним лицом оно оценивается в 10 млн, другим – в 1000 млн. С арифметической точки зрения в первом случае получаем ошибку в 90 млн у. е., во втором – в 900 млн у.е. Если оценивать, во сколько раз ошиблись покупатели, то получаем один ответ в обоих случаях – в 10 раз.

1.3. Показатели описательной статистики

Среднее квадратическое ($M_{кв}$) используется, когда необходима проверка результатов эксперимента на единство суммарного действия (средний радиус или диаметр объекта, площадь земельных участков, функциональных зон и т.д.).

Пример. Имеются данные по величине радиусов трех спилов дуба: 10, 15 и 20 см. Среднее квадратическое будет равно:

$$M_{кв} = \sqrt{\sum x_i^2 / N} = \sqrt{\sum (10^2 + 15^2 + 20^2) / 3} = 15{,}56 \text{ см.}$$

1.3. Показатели описательной статистики

Среднее кубическое **($M_{куб}$)** применяется при проверке на единство суммарного действия, например, при нахождении средней величины объема.

Пример. Кубатура древесины по трем ключевым участкам составляет 10, 15, и 20 м^3. Определяем среднее кубическое по формуле:

$$M_{куб} = \sqrt[3]{\sum x_i^3 / N} = \sqrt[3]{\sum (10^3 + 15^3 + 20^3)/3} = 16{,}03 \text{ м}^3.$$

1.3. Показатели описательной статистики

Средневзвешенная **($M_{взв}$).** Сгруппированный вариационный ряд по классам иногда называют взвешенным из-за той роли, которую выполняют частоты. Чем больше частота вариант в классе, тем *больший вес* она имеет в характере распределения числового ряда. Среднее арифметическое, рассчитанное в этом ряду, называют *взвешенным средним*:

$$M_{взв} = \sum[(x_1 \cdot f_1) + (x_2 \cdot f_2) + \ldots + (x_n \cdot f_n)] / \sum f_i,$$

где x_n – варианты; f_i – частоты по классам.

Если совокупность вариант разбита на несколько неравных по численности групп, то среднюю арифметическую вычисляют для каждой группы. Затем их объединяют, определяя *общее среднее* ($M_{общ}$):

$$M_{общ} = \sum M_j \cdot n_j / \sum n_j ,$$

где M_j – среднее по группам; n_j – число вариант в группе.