Лекция 4.

Статистические методы обработки информации в нефтегазовом деле.

Составитель асс. каф. БНГС СамГТУ, магистр Никитин В.И.

3.1. Математическое ожидание случайной величины

Математическое ожидание - мера центральной тенденции в рассеянии случайной величины, одна из важнейших числовых характеристик распределения вероятностей случайной величины. Для *непрерывной* случайной величины математическое ожидание выражается формулой:

$$M_x = \int_{-\infty}^{+\infty} xp(x)dx \tag{3.1}$$

- это выражение является математическим выражением координаты центра тяжести, т. е. M_x можно представить себе как абсциссу центра тяжести массы, расположенной под кривой, являющейся *плотностью вероятности* $p(x)$.

Для *дискретной* случайной величины *X*, принимающей последовательность значений $x_1, x_2, ..., x_n$ с вероятностями, равными соответственно $p_1, p_2, ..., p_n$ математическое ожидание определяется формулой:

$$M_x = \sum_{i=1}^{\infty} x_i p_i \tag{3.2}$$

В научных и прикладных исследованиях математическое ожидание характеризует наиболее вероятное значение физической величины, получаемой в результате экспериментального определения, но оно отличается от моды, которая характеризует расположение максимума кривой $p(x)$.

Проблема любого наблюдения и эксперимента заключается в том, что значение какой-либо характеристики явления или процесса абсолютно точно определить невозможно. В результате многократных измерений физической величины получится множество значений, имеющих большее или меньшее рассеяние относительно среднего значения. Принято считать, что это среднее значение является оценкой неизвестного математического ожидания (истинного значения измеряемой величины).

Для физических величин определяемых в результате экспериментов математическое ожидание определить невозможно, его можно только *оценить*.

3.2. Множественность оценок математического ожидания

Оценка - количественная характеристика параметра, получаемая по результатам выборки. Проблема оценки неизвестного параметра является одной из центральных в теории обработки результатов наблюдений. К оценкам параметров предъявляется комплекс требований. Важнейшие среди них: несмещённость, состоятельность и эффективность.

Важно отметить, что в отличие от математического ожидания (некоторой неизвестной абстрактной величины) оценок математического ожидания множество, например, арифметическое взвешенное среднее, *арифметико-геометрическое среднее, арифметическое среднее, взвешенное степенное среднее, гармоническое среднее, геометрическое среднее, среднее квадратичное, среднее кубическое,* а также *мода, медиана* и *начальный момент первого порядка*:

Очевидно, что разные виды средних различаются; отсюда возникает проблема правильного выбора формы среднего значения. Решающую роль здесь играет физическая сущность объекта (процесса, явления), опыт и компетентность исследователя.

Существует несколько способов оценки математического ожидания по результатам выборки, но только некоторые из них используются практически. В большинстве случаев важно знать среднее значение выборки или совокупности, которое удовлетворяло бы некоторому критерию, соответствующему физической сущности задачи.

Максимуму кривой плотности вероятностей соответствует *мода,* x_0 это наиболее вероятный результат. Мода в статистике - то, что в обычной жизни называется массовым, типичным. Например, цена, по которой данный товар чаще всего реализуется на рынке.

Если распределение асимметрично (Рис.3.1), то иногда представляет интерес *медиана,* $x_{0.5}$- то значение случайной величины, которое делит распределение на две равные части. Другими словами, вероятности событий по обе стороны медианы одинаковы. Мода и медиана имеют больше теоретическое, чем практическое значение - для экспериментальной выборки моду и медиану вычислить непросто. Также Рис.3.1. демонстрирует тот факт, что для ассиметричного распределения оценки математического ожидания не равны друг другу.

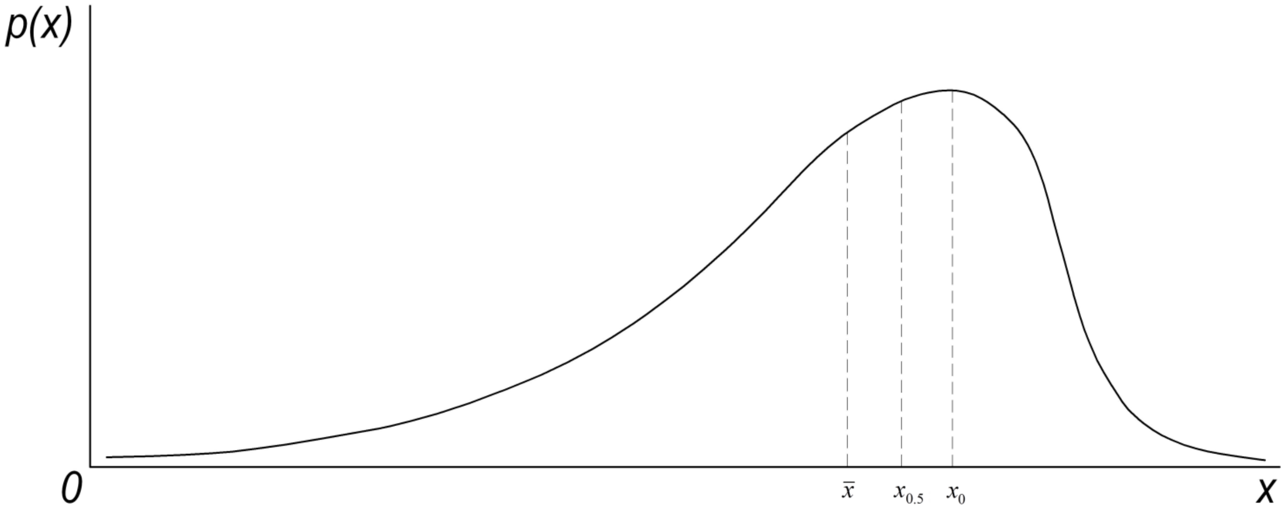

Рис.3.1. Три различные оценки математического ожидания для ассиметричного распределения.

Как правило, выбор оценки математического ожидания необходимо осуществлять на основании знаний о физической сущности процесса. К сожалению, не во всех задачах можно сразу выбрать лучшую из оценок. Поэтому, будет полезно ввести критерий эффективности оценки. *Эффективная статистическая оценка* - оценка, обладающая наименьшей дисперсией среди нескольких оценок одного и того же параметра.

3.3. Дисперсия распределения случайной величины

Дисперсия - в математической статистике и теории вероятностей теории - мера рассеяния значений случайной величины относительно центра (соответствующего математическому ожиданию), одна из характеристик распределения вероятностей случайной величины, т. е. отклонения её от среднего значения.

В теории вероятностей дисперсия D_x, случайной величины *X* определяется как математическое ожидание квадрата отклонения *X* от ее математического ожидания. Для случайной величины с дискретным распределением дисперсия определяется формулой:

$$D_x = \sum_{i=1}^{\infty} (x_i - M_x)^2 p_i. \tag{3.3}$$

Для случайной величины *X* с непрерывным распределением, имеющим плотность вероятностей *p(x),* - формулой:

$$D_x = \int_{-\infty}^{+\infty} (x_i - M_x)^2 p(x)dx. \tag{3.4}$$

Если $D_x = 0$, то случайная величина ***X*** принимает с вероятностью 1 единственное значение M_x. Дисперсия имеет важное значение в характеристике качества статистической оценки случайной величины. Наряду с дисперсией в качестве меры рассеяния (той же размерности, что и сама случайная величина) используется квадратный корень из дисперсии:

$$\sigma_x = \sqrt{D_x} \tag{3.5}$$

- *стандартное (квадратичное) отклонение **X**.*

На рис. 3.2 приведены графики плотности различных распределений. Очевидно, что математическое ожидание определяет центр распределения, а дисперсия и следовательно – квадратичное отклонение опрелеляют рассеяние случайной величины. Таким образом можно сделать вывод, что рис. 3.2 (1) соответствует меньшая дисперсия, чем на рис. 3.2 (2) и рис. 3.2 (3). Данная иллюстрация может отображать реальные измеренные физические процессы и тогда можно сделать вывод о погрешности эксперимента и, например, сказать, что на рис. 3.2 (2) погрешность измерений больше, чем на рис. рис. 3.2 (1), из-за разницы дисперсий. рис. 3.2 (3)- самая большая дисперсия

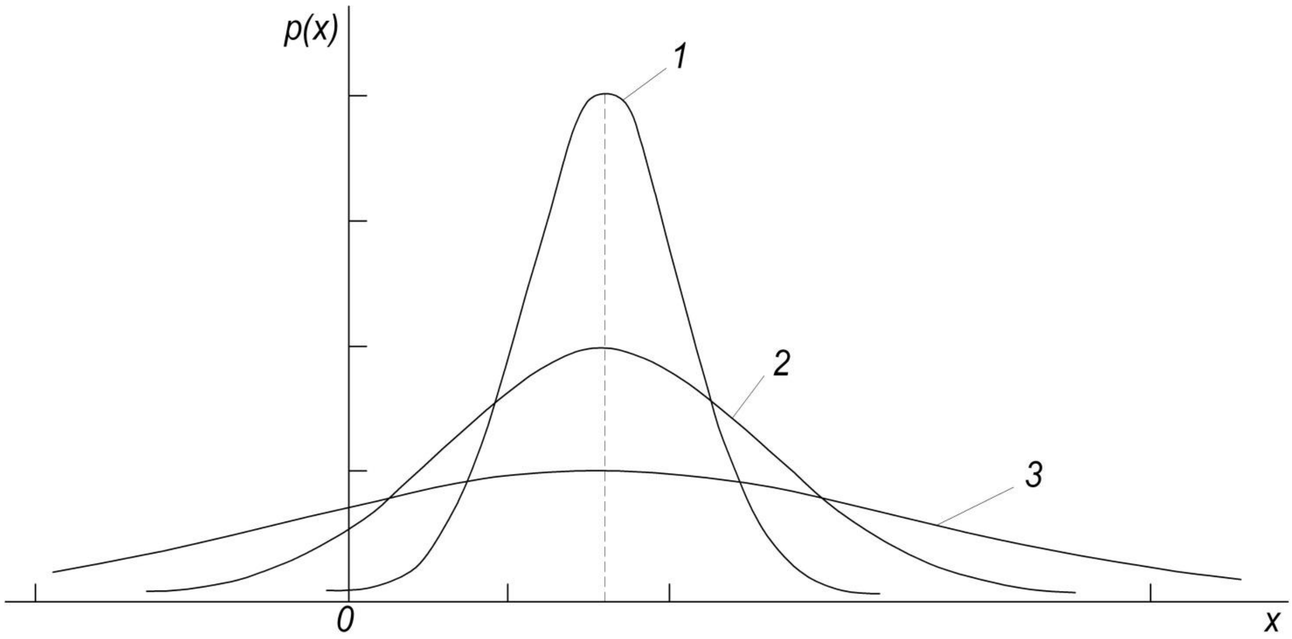

Рис. 3.2. Плотности распределений при различных значениях математического ожидания и дисперсии: (1) - малая дисперсия, (2) - большая дисперсия (3)- самая большая дисперсия

Поскольку в практике математическое ожидание – величина неизвестная, на практике определяют *среднее значение выборки* и *выборочную (опытную)* дисперсию:

$$S_{on}^2 = \frac{1}{n_{on}-1}\sum_{i=1}^{n_{on}}\left(x_i - \bar{x}\right)^2 = \frac{1}{n_{on}-1}\left(\sum_{i=1}^{n_{on}} x_i^2 - n_{on}\bar{x}^2\right), \qquad (3.6)$$

n_{on}- число опытов в выборке, $\upsilon_{on} = n_{on} - 1$- число степеней свободы выборочной дисперсии. $\bar{x}$ определенная оценка математического ожидание (среднее значение). Множитель $\frac{1}{n_{on}-1}$ объясняется тем, что если использовать формулу (3.6) то вычисляемая опытная дисперсия будет являться смещённой оценкой дисперсии генеральной совокупности, уменьшающей её.

В соответствии с (3.5) вычисляется квадратичное отклонение:

$$s = \pm\sqrt{S_{on}^2} \qquad (3.7)$$

- *квадратичное отклонение* или *стандартное отклонение*.

S_{on} ещё называют *дисперсией воспроизводимости. Дисперсия воспроизводимости* - количественная характеристика точности эксперимента или воспроизводимости результатов наблюдений в природе, науке, технике и технологии.

Так как часто выборка представлена в виде *дискретного и интервального вариационного* ряда, то представим формулы, для вычисления опытной дисперсии для них.

В случае дискретного вариационного ряда, среднее опытная дисперсия вычисляется как:

$$S_{on}^2 = \frac{1}{n_{on}-1}\sum_{i=1}^{K} n_i \left(x_i - \bar{x}\right)^2 = \frac{1}{n_{on}-1}\left(\sum_{i=1}^{K} n_i x_i^2 - n_{on}\bar{x}^2\right), \tag{3.8}$$

K – количество вариант, n_i - их частоты.

При интервальном вариационном ряде:

$$S_{on}^2 = \frac{1}{n_{on}-1}\sum_{i=1}^{K} n_i \left(x_i^* - \bar{x}\right)^2 = \frac{1}{n_{on}-1}\left(\sum_{i=1}^{K} n_i (x_i^*)^2 - n_{on}\bar{x}^2\right), \tag{3.9}$$

где K – количество интервалов, n_i - частоты им соответствующие, x_i^* - середины интервалов.

Следует различать дисперсию эмпирического распределения (дисперсию выборочного распределения) и дисперсию воспроизводимости (опытную дисперсию). Для определения дисперсии воспроизводимости осуществляют так называемые параллельные опыты и оценку математического ожидания как правило производят по формуле среднего арифметического. При этом предполагают (проверяют), что ошибки измерения физических величин подчиняются закону нормального распределения. В случае эмпирического распределения параллельные опыты имеют второстепенное значение и производятся для отладки методики конкретных наблюдений, экспериментов. Главная задача исследования эмпирического распределения - сделать представительную выборку из совокупности. Если эмпирическое распределение асимметрично. то возникает достаточно серьёзная проблема выбора центра распределения.

3.4. Коэффициент вариации

Вариация - изменение, видоизменение, различие, разнообразие, разновидность, некоторое отклонение от основной формы.

Коэффициент вариации - безразмерная мера рассеяния случайной величины относительно математического ожидания:

$$V_x = \frac{\sigma_x}{M_x}, \tag{3.10}$$

где σ_x - стандартное отклонение случайной величины, M_x - математическое ожидание. Коэффициент вариации был предложен в 1895 году К.Пирсоном (Pearson Karl; 1857-1936).

Выборочный коэффициент вариации вычисляется по формуле:

$$V_x = \frac{s_x}{\bar{x}}, \tag{3.11}$$

где s_x - квадратичное отклонение случайной величины, $\bar{x}$ - оценка математического ожидания. В практике экспериментальных исследований коэффициент вариации служит также оценкой точности (воспроизводимости) эксперимента; в отечественной литературе его называют относительной ошибкой и обычно выражают в процентах. Т.е. умножая (3.11) на 100% можно указать относительную погрешность эксперимента.

3.5. Мода и медиана, как оценки математического ожидания

Как обсуждалось в пункте 3.2. существует некоторое количество оценок математического ожидания.

Рассмотрим разные формы оценок математического ожидания для выборки $x_1, x_2, ..., x_n$ (<u>не путать с вариационным рядом</u>). Также неупорядоченные выборки, среди которых могут повторяться некоторые значения x_i,будем называть массивом значений или массивом данных.

Наибольшему значению вероятности соответствует *мода,* x_0 это наиболее вероятный результат. Следовательно, модой в дискретной выборке будет являться наиболее часто встречающаяся величина, т.е. случайная величина с наибольшей частотой.

При интервальной группировке выбирается интервал, которому соответствует наибольшая частота. Пусть это k-й интервал $(x_{k-1};x_k)$. его частота равна n_k, а ширина Δ. Тогда

$$x_0 = x_{k-1} + \Delta\frac{n_k - n_{k-1}}{2n_k - n_{k-1} - n_{k+1}} \qquad (3.12)$$

Медианой $x_{0.5}$ будем называть значение случайной величины, которая делит весь упорядоченный по неубыванию массив данных на две равные части. Так как из массива значений можно составить вариационный ряд, то $x_{0.5}$ будем называть ту варианту, которая делит вариационный ряд на равные по числу вариант части. Способы вычисления медианы:

пусть $n -$ объём выборки. Возможны два случая – объём выборки может быть чётным или нечетным. То есть существует такое число $k,$ что $n = 2k+1$ - для нечетного n. Или, в противном случае, $n = 2k$ - для четного n. Тогда медиана вычисляется следующим образом:

$$x_{0.5} = \begin{cases} \dfrac{x_k + x_{k+1}}{2}, & \text{если} \quad n = 2k, \Rightarrow n - \text{четное} \\ x_{k+1}, & \text{если} \quad n = 2k+1, \Rightarrow n - \text{нечетное} \end{cases} \qquad (3.13)$$

При работе с интервальным вариационным рядом, сначала находят медианный интервал $(x_{s-1};x_s)$, номер s которого определяется из равенств:

$$\sum_{i<s} n_i \le \frac{n}{2}, \quad \sum_{i\le s} n_i > \frac{n}{2}, \qquad (3.14)$$

где $\sum\limits_{i<s} n_i$ - сумма частот всех интервалов левее медианного, $\sum\limits_{i\le s} n_i$ - сумма частот, включающая частоту медианного интервала. Тогда медиану оценивают с помощью интерполяционной формулы:

$$x_{0.5} = x_{s-1} + \Delta\left(\frac{\frac{n}{2} - \sum\limits_{i<s} n_i}{n_s}\right), \qquad (3.15)$$

где $n_s -$ частота медианного интервала.

ПРИМЕР 3. (Вычисление моды и медианы для массива данных и для дискретного вариационного ряда)

Имеется массив данных об обводнённой нефти из насосных скважин (в %):

61,2 61,4 60,2 61,2 61,3 60,4 61,4 60,8 61,2 60,6
61,6 60,2 61,3 60,3 60,7 60,9 61,2 60,5 61,0 61,4

Задания:

1) По массиву данных вычислить моду и медиану.

2) Перейти от массива данных к дискретному вариационному ряду и вычислить моду и медиану.

Решение

1) Для нахождения моды и медианы по массиву данных отсортируем его по неубыванию:

60,2 60,2 60,3 60,4 60,5 60,6 60,7 60,8 60,9 61

61,2 61,2 61,2 61,2 61,3 61,3 61,4 61,4 61,4 61,6

Очевидно, что 61,2 повторяется наибольшее число раз, т.е. его частота $n_{61,2}=4$, следовательно $x_0=61{,}2$ - мода.

Для вычисления медианы, сначала определим объём выборки- $n=20$. Далее, чтобы воспользоваться формулой (3.11) определим число k:

$$n=20-\text{четное},\quad k=2k,\Rightarrow k=n/2=20/2=10$$

Согласно (3.13):

$$x_{0.5}=\frac{x_k+x_{k+1}}{2}=\frac{x_{10}+x_{11}}{2}$$

По упорядоченному массиву данных обводнённой находим, что $x_{10}=61$, $x_{11}=61{,}2$, тогда:

$$x_{0.5}=\frac{61+61{,}2}{2}=61{,}1$$

$x_{0.5}=61{,}1-$медиана.

2) Для того, чтобы перейти от массива данных к дискретному вариационному ряду, воспользуемся отсортированным массивом, записанным выше, и составим таблицу, соответствующую табл.1.1.

Объём выборки - $n = 20$, подсчитаем количество вариант, т.е. количество различных элементов выборки - $K = 13$, т.е. таблица для вариационного ряда будет иметь 13 столбцов+ столбец оглавлений.

Запишем все варианты в первую строку таблицы, а во второй строке перечислим частоты соответствующие им, т.е. сколько раз каждая из вариант встречается в выборке.

x_i	60,2	60,3	60,4	60,5	60,6	60,7	60,8	60,9	61	61,2	61,3	61,4	61,6
n_i	2	1	1	1	1	1	1	1	1	4	2	3	1

Для самоконтроля проверим, что $\sum_{i=1}^{K} n_i = n = 20$. Таким образом получен дискретный вариационный ряд.

Воспользуемся определением моды, для её нахождения. Очевидно, что $x_0 = 61{,}2$, так как соответствующая варианта имеет наибольшую частоту.

Определим теперь медиану.

Выборка четная, $n = 20 - \textit{четное}, \quad k = 2k, \Rightarrow k = n/2 = 20/2 = 10$, теперь необходимо найти случайные величины с индексами 10 и 11, т.е. x_{10} и x_{11}. Для этого воспользуемся записанным вариационным рядом, учтём, что частоты n_i указывают сколько раз повторяется одна и та же варианта. Например, в первом столбике записано, что $x_1 = 60{,}2$, $x_2 = 60{,}2$ следовательно второй столбик указывает, что $x_3 = 60{,}3$, следовательно $x_{10} = 61$, $x_{11} = x_{12} = x_{13} = x_{14} = 61{,}2$, следовательно, согласно (3.13):

$$x_{0.5} = \frac{61 + 61{,}2}{2} = 61{,}1$$

В заключении примера заметим, что результаты, 1) и 2) абсолютно идентичны, что является следствием того факта, что можно работать как с упорядоченным массивом данных, так и с вариационным рядом, т.к. они являются различными формами записи одних и тех же эмпирических данных.

ПРИМЕР 4 (Вычисление моды и медианы для интервального вариационного ряда)

Дана выборка в виде интервального вариационного ряда.

Интервалы	(5;7)	(7;9)	(9;11)	(11;13)	(13;15)	(15;17)	(17.19)
Частоты n_i	4	8	11	7	5	3	2

Найти оценки моды и медианы для данной выборки.

Решение

Определим объём выборки - $\sum_{i=1}^{k} n_i = n = 40$. $\Delta = x_i - x_{i-1} = 2$ - длина интервалов.

Согласно (3.12) определим моду по интервальному вариационному ряду. Для этого определим интервал, которому соответствует наибольшая частота. Третьему интервалу (9;11) соответствует наибольшая частота $n_3 = 11$. Тогда по формуле (3.12) вычисляется мода:

$$x_0 = x_{k-1} + \Delta \frac{n_k - n_{k-1}}{2n_k - n_{k-1} - n_{k+1}} = 9 + 2\frac{11-8}{2 \cdot 11 - 8 - 7} \approx 9{,}86$$

Определим оценку медианы. Для этого определим медианный интервал по правилу (3.14),для этого сначала определим $\frac{n}{2} = 20$. Далее подсчитаем сумму частот всех интервалов левее медианного, таким образм, чтобы она не превышала 20. В данном примере это будет сумма частот первых двух интервалов:

$$n_1 + n_2 = 4 + 8 = 12 < 20$$

Теперь вычислим сумму частот интервалов, включающий медианный, т.е. сумму частот первых трёх интервалов в данном примере:

$$n_1 + n_2 + n_3 = 4 + 8 + 11 = 23 > 20$$

Таким образом, обе вычисленные суммы удовлетворяют условию (3.14) и следовательно, третий интервал (9;11) – медианный, тогда по формуле (3.15) вычислим медиану:

$$x_{0.5} = x_{s-1} + \Delta\left(\frac{\frac{n}{2} - \sum_{i<s} n_i}{n_s}\right), = 9 + 2\left(\frac{20-12}{11}\right) \approx 10,45$$

3.6. Средние значения, как оценки математического ожидания

Среднее значение совокупности чисел - концептуальное число, заключённое между наименьшим и наибольшим из них и получаемое из элементов выборки при помощи некоторой процедуры, которая как раз и определяет специфический вид среднего значения.

Среднее значение может быть целью исследования, а может привлекаться для характеристики совокупности в целом как один из *моментов* или как оценка математического ожидания. Среднее значение совокупности выражает равнодействующую влияния множества факторов на вариацию признака независимо от вида распределения случайной величины.

Среднее значение - понятие математической статистики, эта величина зависит, зависящая от метода её вычисления. В общем виде среднее значение величины X будем обозначать $\bar{x}$.

В научных исследованиях технологических процессов в большинстве случаев особых проблем с исчислением среднего значения не возникает.

Перечислим применяемые на практике формулы для вычисления средних значений по массиву данных (выборке) объёмом n. Наиболее употребительным средним является

Арифметическое среднее:

$$x_{ap} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}. \qquad (3.16)$$

Гармоническое среднее:

$$x_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}. \qquad (3.17)$$

Геометрическое среднее

$$x_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}. \qquad (3.18)$$

Среднее квадратичное:

$$x_{кв} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} \qquad (3.19)$$

Среднее кубическое:

$$x_{куб} = \sqrt[3]{\frac{\sum_{i=1}^{n} x_i^3}{n}} = \sqrt[3]{\frac{x_1^3 + x_2^3 + \ldots + x_n^3}{n}} \qquad (3.20)$$

Арифметико-геометрическое среднее $x_{ар.-геом.}$ вычисляется по следующему правилу:

Арифметико-геометрическое среднее - общий предел последовательностей арифметического среднего $x_{ар}$ и геометрического среднего x_g, получаемых в результате следующих операций:

1) по заданной выборке определяется среднее арифметическое и среднее геометрическое;

2) По полученным значениям $x_{ар}$ и x_g вычисляются среднее арифметическое и среднее геометрическое;

3) Пункт 2) следует повторить пока $|x_{ар} - x_g| \geq \varepsilon$, где $\varepsilon -$ заданная точность вычислений.

Так как арифметическое среднее является наиболее часто встречающейся оценкой математического ожидания, укажем выражения для неё в случае *дискретного и интервального вариационного* ряда:

В случае дискретного вариационного ряда, среднее арифметическое вычисляется как:

$$x_{ар} = \frac{\sum_{i=1}^{K} n_i x_i}{n} = \sum_{i=1}^{K} h_i x_i, \qquad (3.21)$$

$K -$ количество вариант, n_i - их частоты, h_i -их относительные частоты

При интервальном вариационном ряде:

$$x_{ap} = \frac{\sum_{i=1}^{K} n_i x_i^*}{n} = \sum_{i=1}^{K} h_i x_i^*, \tag{3.22}$$

где K – количество интервалов, n_i - частоты им соответствующие, h_i - относительные частоты, x_i^* - середины интервалов.

ПРИМЕР 5 (Оценка эффективности средних значений, как оценки математического ожидания выборки)

Пусть дан массив проходки на долото за рейс, в метрах. Проходка на долото - пройденный долотом интервал ствола скважины с момента спуска долота и до его подъёма на поверхность (рейс).

3,34 3,36 3,43 3,45 3,45

3,46 3,47 3,47 3, 48 3,51

Массив получен при проведении приёмочных испытаний коронок КАП в породах IX-X категории по буримости с различной степенью трещиноватости и абразивности.

Требуется вычислить средние значения: x_{ap}, x_h, x_g, $x_{кв}$, $x_{куб}$ $x_{ap.-геом.}$. Определить наиболее эффективную из предложенных оценок математического ожидания.

Решение

Данную задачу удобно решать при помощи вспомогательных таблиц, работать с которыми можно как вручную, так и используя специализированные программы, такие как Microsoft Office Excel, или его аналоги.

Вычислим x_{ap}, x_h, x_g, $x_{кв}$, $x_{куб}$ $x_{ap.-геом.}$, их дисперсии, а далее дадим оценку . Наиболее эффективным средним значение будет то, у которого меньшая дисперсия.

Вся задача заключается в правильном использовании формул (3.16)-(3.20), т.е. в выполнении тривиальных арифметических операций. Поэтому обсудим только вычисление *арифметико-геометрического среднего,* по алгоритму представленному выше.

Вычислим по (3.16) среднее арифметическое

$$x_{ap} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3{,}34 + 3{,}36 + 3{,}43 + 3{,}45 + 3{,}45 + 3{,}46 + 3{,}47 + 3{,}47 + 3{,}\,48 + 3{,}51}{10} = 3{,}442$$

Вычислим по (3.16) геометрическое среднее:

$$x_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[10]{3{,}34 \cdot 3{,}36 \cdot 3{,}43 \cdot 3{,}45 \cdot 3{,}45 \cdot 3{,}46 \cdot 3{,}47 \cdot 3{,}47 \cdot 3{,}\,48 \cdot 3{,}51} = 3{,}441628158$$

Зададим точность, с которой будет вычисляться арифметико-геометрическое среднее. Пусть $\varepsilon = 0{,}001$, данной точности достаточно для многих тонических приложений. Следовательно по имеющимся $x_{ap},\ x_g$ будем вычислять $x_{ap},\ x_g$ заново, пока $\left|x_{ap} - x_g\right| \geq \varepsilon$.

Проделаем первую итерацию:

$$x_{ap} = \frac{3{,}442 + 3{,}441628158}{2} = 3{,}441814079;\ x_g = \sqrt{3{,}442 \cdot 3{,}441628158} = 3{,}441814074$$

Проверяем условие: $\left|x_{ap} - x_g\right| \geq 0{,}001$, очевидно, что оно не выполняется после первой же итерации, так как $x_{ap},\ x_g$ совпадают с точностью до девятого знака после запятой. Теперь мы можем записать искомое значение - $x_{ap.-геом.} = 3{,}44181407$.

Все вычисленные средние значения запишем в сводную таблицу. Также вычислим дисперсии. Дисперсии для каждого среднего значения вычисляются по формуле (3.6). В каждом случае в формулу поставляется соответствующее среднее значение на место $\bar{x}$. Все значения дисперсий занесем в сводную таблицу. В качестве примера разберем вычисление дисперсии для среднего арифметического. Пример находится на рисунке ниже.

	$\boxed{x}$	$\boxed{x-x_{cp}}$	$\boxed{(x-x_{cp})^2}$
	3,34	-0,102	0,010404
	3,36	-0,082	0,006724
	3,43	-0,012	0,000144
	3,45	0,008	6,4E-05
	3,45	0,008	6,4E-05
	3,46	0,018	0,000324
	3,47	0,028	0,000784
	3,47	0,028	0,000784
	3,48	0,038	0,001444
	3,51	0,068	0,004624
$\boxed{x_{cp}}$	3,442	Сумма	0,02536
$\boxed{n}$	10	S^2	0,002818

Пример расчёта дисперсии при помощи электронных таблиц

По результатам вычислений, представленным в сводной таблице очевидно, что разные виды средних различаются; отсюда возникает проблема правильного выбора формы среднего значения. Также очевидно, что у арифметического среднего минимальная дисперсия. Другими словами, арифметическое среднее является эффективной статистической оценкой.

Средние значения	$\bar{x}$	Дисперсии S^2
x_{ap}	3,442	0,002817778
x_h	3,441252854	0,002818398
x_g	3,441628158	0,002817931
$x_{кв}$	3,442368371	0,002817929
$x_{куб}$	3,442733265	0,002818375
$x_{ар.-геом.}$	3,44181407	0,002817816

Сводная таблица средних значений и их дисперсий