

«Обработка результатов измерений: общая последовательность выполнения обработки результатов измерений».

Вопросы:

- 1. Определение точечных оценок закона распределения результатов наблюдений.**
- 2. Определение координаты центра распределения.**
- 3. Определение выборочного среднего арифметического, среднее арифметическое 90 %-ной выборки.**
- 4. Медиана наблюдений, срединный размах вариационного ряда, центр размаха.**
- 5. Определение оценок среднеквадратического отклонения.**

● Порядок обработки результатов наблюдений при многократных измерениях состоит из ряда последовательно выполняемых этапов:

1) определение точечных оценок параметров законов распределения результатов измерений. На этом этапе после ранжирования значений выборки (X) в порядке возрастания и представления ее в виде вариационного ряда ($x_1 \leq x_2 \leq \dots \leq x_n$) определяются:

а) оценка центра распределения ($\overline{X}_{ц.р.}$);

б) оценка среднеквадратического отклонения СКО отдельных результатов наблюдений (S);

в) оценка СКО среднего арифметического значения ($S_{\bar{X}}$).

В соответствии с критериями исключаются грубые погрешности (промахи) и вводятся поправки на систематические погрешности. После их исключения проводится повторный расчет оценок среднего арифметического значения и оценок СКО наблюдений и измерений;

2) определение оценок параметров закона распределения результатов измерений или случайных погрешностей измерений.

В этом случае от выборки результатов измерений x_1, x_2, \dots, x_n переходят к выборке отклонений от среднего арифметического $\Delta x_1, \Delta x_2, \dots, \Delta x_n$

$$\Delta x_i = x_i - \overline{X}_{ц.р.} \quad (1)$$

Для оценки параметров закона распределения проводится построение по исправленным результатам измерений x_i , где $i=1,2, \dots, n$ - члены вариационного ряда (упорядоченной выборки) y_i , где $y_i = \min(x_i)$ и $y_n = \max(x_i)$. По виду статистических функций распределений (представленных в виде гистограмм или полигонов – для дифференциальной формы или в виде кумулятивной кривой – для интегральной формы) может быть оценен закон распределения результатов наблюдений;

3) оценка закона распределения по статистическим критериям согласия. Для проверки гипотез о виде функции распределения экспериментальных данных используют следующие критерии согласия: Пирсона, Мизеса-Смирнова, составной критерий d . При числе наблюдений $n > 50$ для идентификации закона распределения используется критерий Пирсона (хи-квадрат) или критерий Мизеса-Смирнова (ω^2). При $15 < n < 50$ для проверки нормальности закона распределения применяется составной критерий (d -критерий), приведенный в ГОСТ 8.207-76. При $n < 15$ принадлежность экспериментального распределения к нормальному не проверяется. При этом нахождение доверительных границ случайной погрешности результата измерения по методике, предусмотренной ГОСТ 8.207-76 и описываемый далее, возможна в том случае, если заранее известно, что результаты наблюдений принадлежат нормальному распределению.

● В случае если гипотеза о принадлежности полученных результатов к нормальному закону распределения не подтверждается, то проводится приближенная оценка параметров законов распределения (идентификацией формы и вида закона распределения соответствующему теоретическому);

4) определение доверительных интервалов случайной погрешности. Если удалось идентифицировать закон распределения результатов измерений, то с его использованием находят квантильный множитель z_p при заданном значении доверительной вероятности P . В этом случае доверительные границы случайной погрешности $\Delta = \pm z_p \cdot S_{\bar{X}}$;

5) определение границ не исключенной систематической погрешности результата измерений. Под этими границами понимают найденные нестатистическими методами границы интервала, внутри которого находится неисключенная систематическая погрешность. Она образуется из ряда составляющих, как правило, погрешностей метода и средств измерений, а также субъективной погрешности;

6) определение доверительной границы погрешности результата измерения Δ_p . Данная операция осуществляется путем суммирования СКО случайной составляющей $S_{\bar{X}}$ и границ не исключенной систематической составляющей ϑ в зависимости от соотношения этих погрешностей между собой;

7) формирование результата измерений.

1. Определение точечных оценок закона распределения результатов наблюдений.

В математической статистике под оценками понимают приближенные значения искомой (истинной) величины, полученные на основании результатов выборочного исследования и обеспечивающие возможность принятия обоснованных решений о неизвестных параметрах генеральной совокупности.

Чтобы оценки истинного значения измеряемой величины были надёжными, представительными, к ним предъявляется ряд требований. При этом следует помнить о том, что, производя оценку истинного значения измеряемой величины по результатам измерений, мы пользуемся методами теории вероятностей, применяемыми для оценки неизвестных параметров функции распределения случайной величины, т. е. оценки являются случайными величинами. Так, для нормального закона числовыми параметрами распределения являются математическое ожидание и дисперсия.

- **Функция нормального распределения случайных величин.**

Плотность нормального распределения для любой случайной величины описывается уравнением

$$f(x) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2S^2}}, \quad (2)$$

где $f(x)$ – плотность вероятности;

x – значение случайной величины, для которой определяется $f(x)$;

$e = 2,7183$ – основание натуральных логарифмов;

$a = M(x)$ – математическое ожидание;

S – среднее квадратическое отклонение.

- **Математическое ожидание случайной величины.**

Это такое значение случайной величины, вокруг которого группируются результаты отдельных наблюдений. Математическое ожидание дискретной случайной величины $M(x)$ определяется как сумма произведений всех возможных значений случайной величины на вероятность этих значений

$$M(x) = \sum x_i p_i. \quad (3)$$

Для непрерывных случайных величин математическое ожидание равно сумме бесконечно большого числа произведений всех возможных значений случайной величины x на бесконечно малые площади $\rho(x)dx$

$$M(x) = \int_{-\infty}^{+\infty} x \rho(x) dx. \quad (4)$$

Дисперсия.

● Мерой рассеяния значений случайной величины служит дисперсия $D(x)$, равная S^2 . Для дискретных распределений дисперсия отклонений от математического ожидания определяется выражением

$$D(x) = \sum_{i=1}^n [x_i - a]^2 p_i = \sum_{i=1}^n \delta_i^2 p_i = S^2. \quad (5)$$

И соответственно для непрерывных

$$D(x) = \int_{-\infty}^{+\infty} [x_i - a]^2 \rho(x) dx = S^2. \quad (6)$$

Среднее квадратическое отклонение.

Для характеристики рассеяния пользуются средним квадратическим отклонением, определенным как положительный квадратный корень из дисперсии

$$S = +\sqrt{Dx}. \quad (7)$$

Обычно при обработке результатов измерений оценку математического ожидания в виде среднего арифметического значения сопоставляют с оценкой истинного значения измеряемой величины. Но истинное значение является неизвестной нам величиной, имеющей единственное значение. Поэтому только с формальной точки зрения можно признать адекватными оценки математического ожидания случайной величины и истинного значения физической величины. Более того, некоторые результаты измерений находится ближе к истинному значению, чем среднее арифметическое значение результатов измерений. Формальным обоснованием указанной адекватности является то, что оценка параметров закона распределения случайной величины и оценка единственного истинного значения измеряемой величины выполняются по некоторому числу наблюдений, каждое из которых может рассматриваться как случайное событие.

Обработка результатов наблюдений предполагает вычисление математических оценок истинного значения измеряемой величины.

При многократных измерениях за оценку истинного значения измеряемой величины принимается координата центра опытного распределения. Статистическую обработку результатов наблюдений следует начинать с вычисления центра распределения, так как погрешность его нахождения влечет за собой неправильную оценку других характеристик (среднеквадратического отклонения – СКО, вида опытного распределения, оценки погрешностей результата измерений и др.).

Следует напомнить, что прямыми называются измерения, результат которых позволяет непосредственно получить искомые значения физических величин.

В условиях отсутствия сведений о виде и форме закона распределения результатов наблюдений (ограниченное число результатов, грубые СИ) среднее арифметическое значение не всегда может быть принято за оценку координаты центра распределения (центр кривой эмпирического распределения совпадает с оценкой математического ожидания только для нормального распределения).

В математической статистике известны несколько оценок координаты центра распределения: среднее арифметическое, медиана, мода, срединный размах, центр размаха.

Поскольку все перечисленные оценки являются точечными и выбор их неоднозначен, они должны, во-первых, сходиться к оцениваемому значению при $n \rightarrow \infty$ (состоятельные оценки), во-вторых, их математическое ожидание должно быть равно оцениваемому значению (несмещенные оценки), в-третьих, их выборочное распределение должно иметь наименьшую дисперсию (эффективные оценки). Остановимся на перечисленных свойствах оценок более подробно.

При увеличении числа независимых измерений n оценка должна сходиться по вероятности к математическому ожиданию случайной величины.

Такая оценка называется состоятельной. Требование “состоятельности” предъявляется к статистическим оценкам при рассмотрении выборок большого объема (т. е. число измерений должно быть велико!).

Одним из условий получения надёжных оценок является требование к их несмещенности, которое заключается в том, чтобы при замене оценкой (m_x^*) истинного значения X_n не допускалась систематическая погрешность (в сторону увеличения или уменьшения относительно X_n). Это требование приводит к необходимости выполнения условия: математическое ожидание оценки должно при любом числе измерений совпадать с истинным значением величины.

Если выбранная несмещенная оценка по сравнению с другими возможными оценками имеет наименьшую дисперсию, то такая оценка является эффективной, например, $D[m_x] = \min$. Оценка D_x^* не является эффективной. В случае нормального распределения результатов наблюдений статистическая дисперсия является асимптотической несмещенной, так как при увеличении числа измерений n отношение ее дисперсии к минимально возможной измеряемой величине стремится к единице.

Числовые характеристики случайных величин, полученных по результатам выборочных наблюдений (т. е. оценки истинных значений величин), подразделяются на три вида:

- 1) характеристики положения;
- 2) характеристики рассеяния;
- 3) характеристики формы распределения.

К характеристикам положения относятся:

- а) среднее арифметическое значение \overline{X} ;
- б) медиана $\overline{X_M}$;
- в) мода $\overline{X_0}$;
- г) среднее геометрическое значение $\overline{X_q}$;
- д) среднее гармоническое значение $\overline{X_h}$.

Все перечисленные числовые характеристики определяют координату центра распределения упорядоченной совокупности. Формулы для их определения приведены в следующем подразделе. Следует отметить, что только в случае нормально распределенных результатов наблюдений выборочное среднее арифметическое, медиана и мода совпадают между собой и могут быть приняты за центр распределения статистической совокупности физической величины, полученной при измерениях.

- К характеристикам рассеяния значений переменной относятся:
 - а) минимальное (X_{min}) и максимальное (X_{max}) значение;
 - б) размах вариационного ряда ($\overline{X}_R; \overline{R}$);
 - в) дисперсия (S^2);
 - г) среднее квадратическое (стандартное) отклонение (S);
 - д) 25 %-й (LQ) и 75 % (UQ) квантили и межквантильный размах ($RQ - UQ - LQ$);
 - е) среднее квадратичное отклонение среднего значения ($S_{\bar{x}}$);
 - ж) 95 %-й доверительный интервал истинного среднего значения.

2. Определение координаты центра распределения.

Координата центра распределения определяет положение случайной величины на числовой оси и может быть найдена несколькими способами. Наиболее фундаментальным является отыскание центра по принципу симметрии, т. е. такой точки X_M на оси x , слева и справа от которой вероятности появления различных значений случайной величины одинаковы и равны 0,5. В этом случае для интегральной функции распределения вероятностей должно выполняться условие:

$$F(X_M) = \int_{-\infty}^{X_M} P(X) dX = \int_{X_M}^{-\infty} P(X) dX = 0,5. \quad (8)$$

При этом точку X_M называют медианой или 50 %-ной квантилью. Для его нахождения у распределения случайной величины должен существовать только нулевой начальный момент. Нулевым моментом в математической статистике называют некоторое среднее значение, отсчитываемое от начала координат. Нулевой начальный момент равен единице. Он используется для задания условия нормирования плотности распределения и определяется по формуле:

$$\int_{-\infty}^{+\infty} X^0 P(x) dX = 1. \quad (9)$$

Первым начальным моментом является математическое ожидание случайной величины.

В качестве оценки центра распределения может выбираться одна из следующих оценок (в зависимости от типа распределения): выборочное среднее арифметическое, медиана, центр размаха, срединный размах.

При выборе оценок центра распределения следует учитывать, что они имеют различную чувствительность к наличию промахов в обрабатываемой совокупности исходных данных.

3. Определение выборочного среднего арифметического (\bar{X}), среднее арифметическое 90 %-ной выборки ($\overline{X_{0,9}}$).

● *Выборочное среднее арифметическое* определяют по формуле:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i, \quad (10)$$

где X_i – отдельные результаты наблюдений;
 n – общее количество результатов.

Выборочное среднее арифметическое для упорядоченной совокупности (вариационного ряда) вычисляется по формуле:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^k x_i \cdot m_i = \sum_{i=1}^k x_i \cdot \hat{p}_i, \quad (11)$$

где m – частота повтора отдельных результатов наблюдений;
 $p_i = \frac{m_i}{n}$ - частость (статистическая вероятность) попадания i -го наблюдения в определенный k -й интервал.

● Выборочное среднее арифметическое \bar{X} является несмещенной оценкой любого закона распределения, кроме этого – состоятельной, эффективной и достаточной (характеристика полноты использования всей содержащейся в выборке информации).

Однако оценка в виде среднего арифметического слабо защищена от влияния промахов. Она ослабляется лишь в \sqrt{n} раз, где n – число наблюдений, в то время как его возможный размер не ограничен.

Среднее арифметическое 90 %-ной выборки.

Среднее арифметическое (по ограниченному числу наблюдений) находится по формуле:

$$\overline{X_{(\varepsilon)}} = \frac{1}{n-2t} \cdot \sum_{i=t+1}^{n-t} x_i , \quad (12)$$

где $\varepsilon n \leq t \leq \varepsilon n + 1$ для случая, когда с каждого конца вариационного ряда исключают по t значений для получения более устойчивой оценки центра распределения. Обычно используют значения $\varepsilon = 0,05$ и $\varepsilon = 0,1$ (это означает, что следует отбрасывать по 5 или 10 % результатов наблюдений).

● В метрологии чаще находит применение среднее арифметическое 90 %-ной выборки (обозначаемой символами $\overline{X}_{0,9}$ или $\overline{X}_{0,1}$,). Она определяется по формуле:

$$\overline{X}_{0,9} = \frac{1}{n-2r} \cdot \sum_{i=r+1}^{n-r} x_i , \quad (13)$$

где $2r$ – число не учитываемых результатов.

Среднее арифметическое $\overline{X}_{0,9}$ также может быть определено:

$$\overline{X}_{0,9} = \frac{1}{n-2r} \cdot \sum_{i=1}^k x_i \cdot m_i , \quad (14)$$

где m_i – частота попадания i -го значения в k -й интервал (при интервальном представлении вариационного ряда).

Оценка $\overline{X}_{0,9}$ менее чувствительна к результатам с грубыми погрешностями, чем выборочное среднее арифметическое \overline{X} ; поскольку при обработке 90 % объема выборки отбрасываются из концов вариационного ряда $x_1 < x_2 < x_3 < \dots \leq x_n$ по 5 % наиболее удаленных результатов, в которых могут содержаться грубые погрешности.

4. Медиана наблюдений, срединный размах вариационного ряда, центр размаха.

Медиана наблюдений.

● Медианой \overline{X}_M называют наблюдаемое значение x_i (так называемую варианту), которая делит вариационный ряд на две части, равные по числу вариант.

Медиана \overline{X}_M вычисляется по нижеприведённым формулам.

Если n – четное, то медиана рассчитывается по формуле:

$$\overline{X}_M = \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), \quad (15)$$

Если n – нечетное, то по формуле:

$$\overline{X}_M = \frac{1}{2} \cdot x_{\frac{(n+1)}{2}}. \quad (16)$$

● Следует иметь в виду, что медиана \overline{X}_M является наиболее эффективной оценкой для симметричных экспоненциальных распределений, в которых контрэксцесс принадлежит интервалу $0 < \chi < 0,45$. Для класса распределений, близких к нормальному, с $0,45 < \chi < 0,67$ эффективными оценками являются среднее арифметическое \overline{X} , $\overline{X}_{(0,05)}$ и $\overline{X}_{(0,1)}$, занимающие медианное положение.

Для распределений, близких к равномерному и арксинусоидальному, с $0,67 < \chi < 1$ целесообразно использовать центр размаха \overline{X}_R . Для двухмодальных распределений с $0,67 < \chi < 1$ центр срединного размаха \overline{X}_C .

Медиана \overline{X}_M является эффективной оценкой центра экспоненциальных пологоспадающих одномодальных распределений Лапласа с эксцессом $\varepsilon > 3,8$.

Срединный размах вариационного ряда.

Центр срединного размаха \overline{X}_{R_2} определяют в зависимости от кратности членов ряда по нижеприведенным формулам.

При n , кратном 4 находится по формуле:

$$\overline{X}_{R_2} = \frac{1}{2} \cdot \left(x_{\frac{n}{(4+1)}} + x_{\frac{3n}{4}} \right). \quad (17)$$

- При четном n находится по формуле:

$$\overline{X_{R_2}} = \frac{1}{2} \cdot \left(x_{\frac{(n+2)}{4}} + x_{\frac{(3n+2)}{4}} \right). \quad (18)$$

При $(n-1)$, кратном 4 находится по формуле:

$$\overline{X_{R_2}} = \frac{1}{2} \cdot \left(x_{\frac{(n-1)}{4}+1} + x_{n-\frac{(n-1)}{4}} \right). \quad (19)$$

При $(n+1)$, кратном 4 находится по формуле:

$$\overline{X_{R_2}} = \frac{1}{2} \cdot \left(x_{\frac{(n+1)}{4}+1} + x_{n-\frac{(n+1)}{4}} \right). \quad (20)$$

- Центр срединного размаха вариационного ряда (\overline{X}_C) может быть определен также по формуле:

$$\overline{X}_C = X_{R_2} = \frac{1}{2} \cdot (x_{0,25} + x_{0,75}), \quad (21)$$

где $x_{0,25}$, $x_{0,75}$ – 25 % и 75 %-ные квантили опытного распределения (представляют собой усредненные значения конкретных результатов наблюдений).

Для симметричных двумодальных распределений с эксцессом $1 < \varepsilon < 2,4$ (например, композиция двух экспоненциальных распределений) эта оценка является эффективной.

Следует отметить, что обе квантильные оценки \overline{X}_{R_2} , \overline{X}_M являются защищенными от влияния промахов, поскольку они не зависят от координат промахов.

Центр размаха.

● Для ограниченных распределений (равномерных, треугольных, трапецеидальных и др.) эффективной оценкой центра распределения может служить центр размаха вариационного ряда, вычисляемый по формуле:

$$\overline{X}_R = \frac{1}{2} \cdot (x_1 + x_n), \quad (22)$$

где x_1, x_n – крайние значения вариационного ряда.

Однако эта оценка очень чувствительна к результатам с грубой погрешностью, так как она определяется по наиболее удаленным от центра распределения результатам наблюдений, каковыми и являются промахи.

В условиях, когда отсутствуют сведения о законе и виде распределения за оценку центра $\overline{X}_{ц.р.}$ рекомендуется принимать медиану оценок $\overline{X}, \overline{X}_{0,9}, \overline{X}_M, \overline{X}_C(\overline{X}_{R_2}), \overline{X}_p$, расположенных в вариационный ряд.

5. Определение оценок среднеквадратического отклонения.

● Оценка S^2 генеральной дисперсии σ^2 любого закона распределения может быть вычислена (при неизвестном математическом ожидании генерального среднего) по формуле:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{X}_{\text{ц.р.}})^2}. \quad (23)$$

Эта оценка является несмещенной и состоятельной, а для нормального распределения – еще и эффективной.

Для нормального закона распределения оценка генерального среднеквадратического отклонения (СКО) S результатов наблюдений определяется:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{X})^2}. \quad (24)$$

● Оценка является несмещенной, состоятельной и асимптотически эффективной только для нормального закона. В случае представления результатов вариационным рядом следует пользоваться следующей формулой.

$$\bar{X} = \sum_{i=1}^k x_{i0} \cdot \hat{p}_i , \quad (25)$$

где x_{i0} – значение измеряемой величины в середине k -го интервала;

k – количество интервалов, на которые разбит вариационный ряд.

Несмещенная оценка СКО для нормальных распределений так же определяется по формуле:

$$S = M_k \cdot S^* ; \quad (26)$$

$$S^* = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{X})^2} . \quad (27)$$

Значения коэффициента M_k приведены в таблице 1.

Таблица 1 – Значения коэффициента M_k в зависимости от количества наблюдений n .

n	M_k	n	M_k	n	M_k
1	1,253	10	1,025	19	1,013
2	1,128	11	1,023	20	1,012
3	1,085	12	1,021	25	1,010
4	1,064	13	1,019	30	1,008
5	1,051	14	1,018	35	1,007
6	1,042	15	1,017	40	1,006
7	1,036	16	1,016	45	1,006
8	1,032	17	1,015	50	1,005
9	1,028	18	1,014	60	1,004

● СКО случайной погрешности оценки центра распределения (СКО результата измерений) убывает по сравнению с СКО результата наблюдений в \sqrt{n} , как показано по формуле:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}. \quad (28)$$

Определение оценок третьего центрального момента μ_3 , коэффициента асимметрии γ_a , СКО коэффициента асимметрии $\sigma(\gamma_a)$ проводится по формулам:

$$\mu_3^* = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{X})^3; \quad (29)$$

$$\gamma_a = \frac{\mu_3^*}{\sigma_3^*}; \quad (30)$$

$$\sigma^*(\gamma_a) = \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}}. \quad (31)$$

Формулы для вычисления начальных a_k и центральных μ_k моментов и соотношения между ними приведены в таблице 2.

Таблица 2 – Перечень формул.

Выбор формул, приведенных во второй и третьей графах таблицы и дающих практически одинаковые результаты, осуществляют в зависимости от особенностей используемых средств вычислительной техники, алгоритмов и программ обработки информации.