

Статистическая обработка результатов измерений.

Вопросы:

- 1. Определение точечных оценок исправленных результатов измерений.**
- 2. Определение закона распределения результатов измерений.**

1. Определение точечных оценок исправленных результатов измерений.

● После исключения результатов с грубыми погрешностями и внесения поправок на систематическую погрешность проводят математико-статистическую обработку исправленных результатов измерений.

Для этого определяют точечные оценки координаты центра распределения и СКО результатов наблюдений и измерений. Точечные оценки могут отличаться от определенных ранее, так как может уменьшиться число результатов (если имеются грубые погрешности), а при наличии систематической составляющей погрешности, расчет точечных оценок должен быть выполнен по исправленным (после исключения систематической погрешности) результатам наблюдений.

Для расчета среднеквадратического отклонения среднего арифметического значения (результата измерений) может быть так же использована формула Питерса:

$$S_{\bar{x}} = \frac{5}{4} \cdot \frac{\sum_{i=1}^n |\Delta x_i|}{n \cdot (n - 1)^{\frac{1}{2}}}$$

● где Δx_i – отклонения отдельных, полученных при измерениях значений от среднего арифметического.

Если $n < 4$ вместо для оценки S_x используют приближенную формулу:

$$S_{\bar{x}} = \frac{1}{n} \cdot \sum_{i=1}^n |\Delta x_i|$$

Если количество измерений достаточно большое (не менее 40-50), то требуется систематизация исходных данных и разделение вариационного ряда на интервалы.

Для интервальных вариационных рядов используют формулы:

$$\bar{X} = \sum_{i=1}^k x_{io} \cdot \hat{p}_i$$

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

k – количество интервалов, на которые разбит вариационный ряд.

$$S^* = \sqrt{\sum_{i=1}^k \hat{p}_i \cdot (x_{io} - \overline{X_{ц.р.}})^2}$$

где \hat{p}_i – статистическая вероятность попадания i -го результата в данный интервал.

Она находится по формуле:

$$\hat{p}_i = \frac{m_i}{n}$$

где m_i – частота попадания результатов в каждый k -й интервал.

Причем, следует заметить, что S^* - смещенная оценка СКО результатов наблюдений. Для определения несмещенной оценки необходимо получить результат, используя формулу:

$$S^* = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^k (x_{io} - \overline{X_{ц.р.}})^2 \cdot m_i}$$

● Число интервалов определяют, пользуясь формулой Старджесса:

$$k = 1 + 3,31 \cdot \lg n$$

или

$$k = \sqrt{n}.$$

Можно использовать приведенные в таблице 1 рекомендуемые числа интервалов k , определяемые в зависимости от числа экспериментальных данных n .

Таблица 1 – Рекомендуемое число интервалов.

Число экспериментальных данных n	Рекомендуемое число интервалов k
40-100	7-9
100-500	8-12
500-1000	10-16
1000-10000	12-22

Рекомендуется выбирать нечетное число интервалов. При расчете числа интервалов также следует округлять значение до целого числа. Применение числа интервалов менее 3 не рекомендуется, т. к. это неинформативно.

- Затем вычисляют ширину интервала h по формуле:

$$h = \frac{(x_{max} - x_{min})}{k}$$

где x_{max} , x_{min} – наибольшее и наименьшее значения данного вариационного ряда (размах).

Определяют границы интервалов (например, x_{min} ; $x_{min} + h$ - границы 1-го интервала, $x_{min} + h - x_{min} + 2h$ – границы 2-го интервала и т. д.), затем определяют частоту попадания в интервалы и середины интервалов.

Упорядоченные значения рекомендуется оформлять таблицей по форме, представленной в виде таблицы 2.

Таблица 2 – Промежуточные значения интервального ряда.

1	2	3	4

2. Определение закона распределения результатов измерений.

Для определения закона распределения случайной величины необходимо, прежде всего, отнести ее либо к дискретной, либо к непрерывной.

Большинство измеряемых величин мы считаем непрерывными. В ряде случаев это связано с недостаточной чувствительностью имеющихся СИ, которые не дают возможности проводить измерения путем счета отдельных частиц.

Также следует заметить, что граница между дискретными и непрерывными величинами далеко не так определена, как это может показаться на первый взгляд. Например, некоторое количество воды, определяемое счетчиком расхода, может рассматриваться как величина непрерывная. Но вода состоит из отдельных молекул, и количество ее может отличаться одно от другого только на целое число молекул, т.е., если бы мы могли считать молекулы, количество воды нужно было бы рассматривать как прерывистую (дискретную) величину.

Случайная величина X может быть полностью охарактеризована с вероятностной точки зрения, если имеется возможность вычислить вероятность появления каждого ее значения. Этим устанавливается закон распределения случайной величины.

Дискретная случайная величина может быть задана перечнем всех ее возможных значений и их вероятностей, но это неприемлемо для непрерывных случайных величин, т. к. невозможно составить перечень всех возможных значений X , которые сплошь заполняет некоторый интервал (a, b) . С этой целью вводится универсальный способ задания любых типов случайных величин в виде функции распределения вероятностей.

Функцией распределения случайной величины называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания (измерения) примет значение, меньшее x , т. е.:

$$F(x) = P(X < x).$$

Иногда вместо термина “функция распределения” используют термин “интегральная функция”.

Можно дать более точное (с учетом рассмотренного понятия о $F(x)$) определение непрерывной случайной величины: случайную величину называют непрерывной, если ее функция распределения $F(x)$ есть непрерывная, кусочно-дифференцируемая функция с непрерывной производной.

● Непрерывную случайную величину можно также задать, используя другую функцию, которую называют плотностью распределения или плотностью вероятности (иногда ее называют дифференциальной функцией).

Плотностью распределения вероятностей непрерывной случайной величины X называют функцию $f(x)$ - первую производную от функции распределения $F(x)$:

$$f(x) = F'(x).$$

Из этого определения следует, что функция распределения является первообразной для плотности распределения.

Для описания распределения вероятности дискретной случайной величины плотность распределения неприменима.

В математической статистике (в том числе и метрологической практике) для описания статистического распределения частот количественного признака X пользуются эмпирической функцией распределения. Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F(x)$, определяющую для каждого значения x относительную частоту события $X < x$, т. е.

$$\hat{F}(x) = \frac{n_x}{n}$$

где n_x – число вариантов, меньших x ;

n – объем выборки.

- Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция $F(x)$ генеральной совокупности определяет вероятность события $X < x$, а эмпирическая $\hat{F}(x)$ определяет относительную частоту этого же события. Можно сказать, что эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Для наглядности используют графическое представление статистического распределения в виде полигона, гистограммы и многоугольника распределения. Напомним, что полигон представляет собой ломаную кривую, соединяющую середины верхних оснований каждого столбца гистограммы. Он более наглядно, чем гистограмма, отражает форму кривой распределения для непрерывной случайной величины. Полигон частот может быть использован и для дискретной случайной величины. В этом случае он представляет собой ломаную, отрезки которой соединяют точки $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$, т. е. показывают соответствие между наблюдаемыми значениями (результатами наблюдений) и соответствующими им частотами появления n_i .

Статистическое распределения, представленное многоугольником распределения или гистограммой имеет дифференциальную форму.

В качестве функции плотности распределения вероятностей погрешности измерений или ее составляющих следует принимать закон, близкий к нормальному усеченному, при соблюдении следующего условия: имеются основания полагать, что реальная (статистическая) функция плотности распределения — функция симметричная, одномодальная, отличная от нуля на конечном интервале значений аргумента, и другая информация о плотности распределения отсутствует согласно МИ 1317-2004.

В тех случаях, когда нет основания полагать, что указанное выше условие выполняется, следует принимать какую-либо другую аппроксимацию функции плотности распределения вероятностей погрешности измерений.

Принятая аппроксимация считается удовлетворительной при следующих условиях:

- а) она позволяет рассчитывать интервальные характеристики погрешности измерений по ее средним квадратическим отклонениям;
- б) возможные значения погрешности расчета, обусловленные отличием принятой аппроксимации от реальной функции плотности распределения, лежат в пределах, допустимых для решения данной конечной задачи (цепи) измерений.

При отсутствии сведений о подходящей аппроксимации функции плотности распределения вероятностей погрешности измерений, не могут быть рассчитаны интервальные характеристики погрешности измерений и погрешности испытаний, а также показатели достоверности контроля параметров образцов продукции.

Расчет характеристик погрешности измерений, при известных типах средств измерений, должен быть основан на использовании метрологических характеристик средств измерений, нормированных по ГОСТ 8.009-84.