# 4 ПОСТРОЕНИЕ ВЫБОРОЧНОГО РАСПРЕДЕЛЕНИЯ И ОЦЕНИВАНИЕ ВАРИАЦИИ ПАРАМЕТРА

Выборочное распределение строят для получения информации о закономерностях вариации изучаемой случайной величины на основании выборки. При этом последовательно решаются две частные задачи – сначала составляется *вариационный ряд* а затем он отображается в виде специфических графиков (гистограмма, кумулята).

## 4.1.Построение интервального вариационного ряда

В общем виде интервальный вариационный ряд представляется, как показано в таблице 4.1. Его элементы имеют следующий смысл.

Карман – часть диапазона варьирования случайной величины, ограниченная ее значениями $u_{0j}$ и $u_{1j} = u_{0j} + l$, где $l$ - длина кармана.

Длины карманов могут быть различными, однако рекомендуют, чтобы они были одинаковыми. В этом случае:

$$l = R/(k-1) = (x_{max} - x_{min})/(k-1), \tag{4.1}$$

где $k$ - число карманов:

$$k \approx 1 + 3{,}322\, lg\, n, \tag{4.2}$$

Значение, найденное по формуле (4.2), рекомендуется округлять до ближайшего меньшего целого.

Структура интервального вариационного ряда представлена в табл. 4.1.

Началом интервального ряда, т. е. левой границей первого кармана ( $j$ =1), принимают значение случайной величины:

$$x = u_{01} = x_{min} - l/2. \tag{4.3}$$

Тогда правой границей первого кармана будет значение

$$x = u_{11} = u_{01} + l. \tag{4.4}$$

Таблица 4.1

Структура интервального вариационного ряда

| Карман | | | Вари-анта | Часто-та | Частость | |
|---|---|---|---|---|---|---|
| Номер | Левая граница | Правая граница | | | Диффе ренци-альная | Куму-лятив-ная |
| $j$ | $u_0$ | $u_1$ | $x^*$ | $m_j$ | $f_j$ | $F_j$ |
| 1 | $u_{01}$ | $u_{11}$ | $x_1^*$ | $m_1$ | $f_1$ | $F_1$ |
| 2 | $u_{02}$ | $u_{12}$ | $x_2^*$ | $m_2$ | $f_2$ | $F_2$ |
| … | … | … | … | … | … | … |
| $j$ | $u_{0j}$ | $u_{1j}$ | $x_j^*$ | $m_j$ | $f_j$ | $F_j$ |
| … | … | … | … | … | … | … |
| $k$ | $u_{0k}$ | $u_{1k}$ | $x_k^*$ | $m_k$ | $f_k$ | 1 |

Для второго и последующих карманов ( $j = 2, 3, \ldots, k$ ):

$$u_{0j} = u_{1(j-1)}; \tag{4.5}$$

$$u_{1j} = u_{0j} + l\ . \tag{4.6}$$

Варианта $x_j^*$ – значение случайной величины, которое считают характерным для $j$ -го кармана. Для непрерывной случайной величины:

$$x_j^* = \frac{u_{0j} + u_{1j}}{2}\ . \tag{4.7}$$

В вариационных рядах значения вариант $x_j^*$ случайной величины обязательно упорядочены (располагаются в возрастающей последовательности). Однако частоты и дифференциальные частости не будут упорядоченными, поскольку одни значения случайной величины неизбежно будут встречаться в исходной выборке реже, а другие - чаще.

Частота $m_j$ - число значений случайной величины, которые могут быть отнесены к данному карману. Возможные условия классификации значений параметра по карманам:

$$u_{0j} \le x < u_{1j}; \tag{4.8}$$

$$u_{0j} < x \le u_{1j}. \tag{4.9}$$

Сумма частот всех членов вариационного ряда, равна объему исходной выборки:

$$\sum_{j=1}^{k} m_j = n. \tag{4.10}$$

Дифференциальная частость $f_j$ - отношение частоты некоторого члена вариационного ряда к общему количеству наблюдений за случайной величиной:

$$f_j = \frac{m_j}{n}. \tag{4.11}$$

Значения частости - действительные положительные числа и при этом меньше 1. Сумма частостей всех членов вариационного ряда равна единице:

$$\sum_{j=1}^{k} f_j = 1. \tag{4.12}$$

По смыслу дифференциальная частость представляет собой оценку плотности вероятности для значения параметра $x = x_j^*$.

Кумулятивная частость $F_j$ является выборочной оценкой функции распределения вероятности для значения параметра $x = x_j^*$:

$$F_j = \sum_{i=1}^{j} f_i, \; (i = 1,2,\dots,j). \tag{4.13}$$

Для последнего кармана

$$F_k = \sum_{j=1}^{k} f_j = 1. \tag{4.14}$$

## 4.2.Графическое отображение интервального вариационного ряда

Графическое изображение вариационного ряда позволяет представить закономерности, присущие распределению случайной величины, в наглядной форме. Наиболее широко используются гистограмма,и кумулятивная кривая.

**Гистограмма** (рис. 4.1). В прямоугольной системе координат по оси абсцисс откладывают отрезки, изображающие карманы, а на этих отрезках, как на основании, строят прямоугольники с высотами, равными частотам $m_j$ или частостям $f_j$ соответствующего интервала. В результате получают ступенчатую фигуру, состоящую из прямоугольников, которую и принято называть гистограммой.

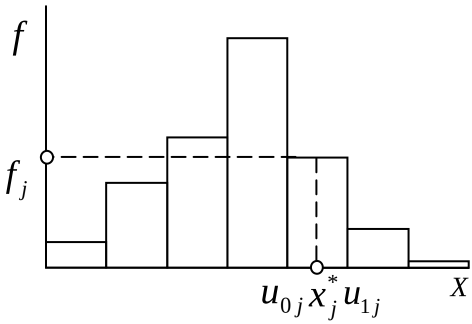

Рис. 4.1. Гистограмма

**Кумулятивная кривая** (**кумулята**) (рис. 4.2) отображает накопленные частоты, либо накопленные частости. При построении кумуляты по интервальному ряду в качестве абсцисс точек кумулятивной кривой принимают соответствующие верхние (правые) границы интервалов. Для нижней (левой) границы первого интервала значение ординаты (накопленной частоты или накопленной частости) принимается равным нулю. В соответствии с (11) для $x = u_{1k}$ накопленная частость $F_k$ =1. Полученные точки соединяют отрезками.

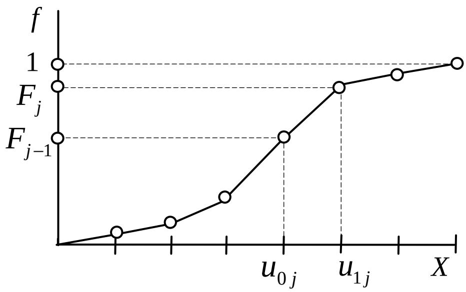

Рис. 4.2. Кумулятивная кривая

### 4.3.Построение теоретической кривой

Поскольку нормальное распределение является наиболее распространенным, изначально строится теоретическая кривая нормального распределения. С этой целью применяется критерий $\chi^2$, который определяется из свойств стандартного нормального распределения:

$$f(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right) \approx 0{,}4\, exp\left(-\frac{z^2}{2}\right). \qquad (4.15)$$

где $z = (x - \mu)/\sigma$.

Оценку нормальности выборочного распределения по $\chi^2$ производят после построения вариационного ряда. Для каждого кармана вычисляют нормированное отклонение

$$z_j = \frac{\left(x_j^* - \bar{x}\right)}{s} \qquad (4.16)$$

и теоретическую частоту

$$m_{\text{т}j} = k' \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z_j^2}{2}\right), \qquad (4.17)$$

где $k'$ – параметр, определяемый по выражению

$$k' = nl/s. \qquad (4.18)$$

Критерий $\chi^2$ характеризует степень несоответствия между теоретическими $m_j'$ и выборочными $m_j$ частотами:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(m_j - m_{\text{т}j}\right)^2}{m_{\text{т}j}}. \qquad (4.19)$$

Значение $\chi^2$, рассчитанное по формуле (4.19), сравнивается с табличным $\chi^2[\alpha; n-3]$. Для нахождения табличного значения $\chi^2$ в *MS Excel* имеется статистическая функция ХИ2ОБР($\alpha$; $n-3$). Гипотезу о соответствии выборочного распределения нормальному принимают, если выполняется условие:

$$\chi^2 > \chi^2[\alpha; n-3]. \quad (4.20)$$

В литературе отмечается, что проверка по критерию $\chi^2$ будет корректна, если для каждого из карманов $m_j$ ≥ 5. В противном случае некоторые карманы рекомендуют объединять.

## 4.4. Пример построения выборочного распределения и оценивания вариации параметра в *MS Excel*

Работа выполняется на отдельном рабочем листе книги *MS Excel.* Исходными данными должна быть однородная выборка, полученная, например, в результате выполнения работы «Обработка и анализ выборки». Пример оформления листа с результатами построения интервального вариационного ряда приведен на рис. 4.3.

В данном примере анализируется распределение толщины холоднокатаной полосы шириной 1550 мм при настройке стана на номинальную толщину 2,10 мм с допусками на повышенную точность по ГОСТ 19904 - 74.

### 4.4.1. Построение интервального вариационного ряда

Построение интервального вариационного ряда рекомендуется выполнять в следующей последовательности.

**Ввести исходные данные.** Исходную выборку (в примере – ячейки B2:B51) ввести с клавиатуры либо копированием через буфер обмена из рабочего листа, на котором выполнялась проверка ее однородности.

**Определить число, длину и границы карманов.** Чтобы выполнить указанные ранее требования к границам карманов, следует предусмотреть возможность ручной корректировки величины $l$ с одновременным расчетом границ $u_{0j}$ и $u_{1j}$. Для расчетов и окончательного выбора длины кармана на рабочем листе необходимо создать таблицу «Число и длина карманов». Здесь $k^*$

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | h, мм | | ВЫБОРОЧНЫЕ ХАРАКТЕРИСТИКИ | | | | | | ЧИСЛО И ДЛИНА КАРМАНОВ | | | | k' | | | |
| 2 | | 2,11 | | n | Xcp | S | Xmin | Xmax | R | k* | k | l* | l | | | | |
| 3 | | 2,09 | | 50 | 2,11 | 0,043 | 2,02 | 2,22 | 0,200 | 6,644 | 6 | 0,033 | 0,040 | 45,99 | | | |
| 4 | | 2,08 | | | | | | | | | | | | | | | |
| 5 | | 2,07 | | j | ВЫБОРОЧНОЕ РАСПРЕДЕЛЕНИЕ | | | | | | ПОСТРОЕНИЕ ТЕОРЕТИЧЕСКИХ КРИВЫХ | | | | | | |
| 6 | | 2,06 | | | $U0_j$ | $U1_j$ | X* | $m_j$ | $f_j$ | $F_j$ | $z_j$ | f(z) | $mт_j$ | $\chi_j^2$ | $fт_j$ | $Fт_j$ | |
| 7 | | 2,12 | | 1 | 2,00 | 2,04 | 2,02 | 3 | 0,060 | 0,060 | -1,996 | 0,054 | 3 | 0,098 | 0,050 | 0,050 | |
| 8 | | 2,08 | | 2 | 2,04 | 2,08 | 2,06 | 11 | 0,220 | 0,280 | -1,076 | 0,224 | 10 | 0,050 | 0,207 | 0,257 | |
| 9 | | 2,14 | | 3 | 2,08 | 2,12 | 2,10 | 21 | 0,420 | 0,700 | -0,156 | 0,394 | 18 | 0,456 | 0,365 | 0,622 | |
| 10 | | 2,06 | | 4 | 2,12 | 2,16 | 2,14 | 10 | 0,200 | 0,900 | 0,763 | 0,298 | 14 | 1,004 | 0,276 | 0,898 | |
| 11 | | 2,10 | | 5 | 2,16 | 2,20 | 2,18 | 3 | 0,060 | 0,960 | 1,683 | 0,097 | 4 | 0,473 | 0,090 | 0,988 | |
| 12 | | 2,17 | | 6 | 2,20 | 2,24 | 2,22 | 2 | 0,040 | 1,000 | 2,603 | 0,013 | 1 | 3,070 | 0,012 | 1,000 | |
| 13 | | 2,08 | | | | | | 50 | 1,000 | | | | 50 | | 1,000 | | |
| 14 | | 2,09 | | | ГРАФИЧЕСКОЕ ОТОБРАЖЕНИЕ | | | | | | | | | | | | |
| 15 | | 2,09 | | | РАСПРЕДЕЛЕНИЯ | | | | | | | | | | | | |
| 16 | | 2,10 | | | X* | Выборочное | | Теоретическое | | | | | | | | | |
| 17 | | 2,22 | | | | f(x) | F(x) | fт(x) | Fт(X) | | | | $\chi^2$ | 5,151 | | | |
| 18 | | 2,12 | | | 2,00 | | | | | | | | p | 0,95 | | | |
| 19 | | 2,03 | | | 2,02 | 0,060 | 0,060 | 0,050 | 0,050 | | | | $\chi^2[\alpha;k-3]$ | 7,815 | | | |
| 20 | | 2,05 | | | 2,06 | 0,220 | 0,280 | 0,207 | 0,257 | | | Распределение | | | | | |
| 21 | | 2,02 | | | 2,10 | 0,420 | 0,700 | 0,365 | 0,622 | | | можно считать | | | | | |
| 22 | | 2,09 | | | 2,14 | 0,200 | 0,900 | 0,276 | 0,898 | | | нормальным | | | | | |
| 23 | | 2,15 | | | 2,18 | 0,060 | 0,960 | 0,090 | 0,988 | | | | | | | | |
| 24 | | 2,11 | | | 2,22 | 0,040 | 1,000 | 0,012 | 1,000 | | | | | | | | |
| 25 | | 2,11 | | | 2,24 | | | | | | | | | | | | |

Рис. 4.3. Пример построения выборочного распределения и оценивания вариации параметра

и $l^*$ - рассчитываемые количество и длина карманов, $k$ и $l$ - задаваемые значения (вводятся с клавиатуры). Расчетные значения определяются программированием соответствующих формул в ячейках рабочего листа. Для примера на рис. 4.3:

| *Параметр* | *Ячейка* | *Формула для расчета в MS Excel* |
|---|---|---|
| $n$ | D3 | =СЧЁТ(B2:B51) |
| $\bar{x}$ | E3 | =СРЗНАЧ(B2:B51) |
| $s$ | F3 | =СТАНДОТКЛОН(B2:B51) |
| $x_{min}$ | G3 | =МИН(B2:B51) |
| $x_{max}$ | H3 | =МАКС(B2:B51) |
| $R$ | I3 | =H3-G3 |
| $k^*$ | J3 | =1+3,322*LOG10(D3) |
| $k$ | K3 | Ввод с клавиатуры |
| $l^*$ | L3 | =I3/F3 |
| $l$ | M3 | Ввод с клавиатуры |
| $k'$ | N3 | =D3*M3/F3 |

Расчет границ интервалов выполнять в таблице «Выборочное распределение». Здесь номера интервалов j вводятся с клавиатуры. Границы интервалов $u_{0j}$ и $u_{1j}$ определяются расчетом по формулам (4.5) и (4.6). В частности, для первых двух карманов из примера на рис. 4.3:

=G3-M3/2 (в ячейке E7);
=E7+$M$3 (в ячейке F7);
=F7 (в ячейке E8);
=E8+$M$3 (в ячейке F8)

Формула (4.5) обеспечивает необходимое значение нижней границы первого интервала при любом $l > l^*$. Поэтому определение длины интервала будет заключаться в подборе такого значения $l$, при котором верхняя граница последнего интервала превысит максимальное значение в выборке на величину $l/2$.

**Определить значения вариант интервального ряда.** Для определения значения вариант в соответствующих ячейках таблицы «Выборочное распределение» необходимо запрограммировать формулу (4.7). Например, для первого кармана:

=(E7+F7)/2 .

**Определить частоты.** Для определения частот $m_j$ применить статистическую функцию ЧАСТОТА(). Синтаксис функции:

ЧАСТОТА(<*Массив_данных*>;<*Массив_карманов*>),

где <*Массив_данных*> - это массив или ссылка на множество данных, для которых вычисляются частоты;

<*Массив_карманов*> - это массив или ссылка на множество карманов, в которые группируются значения аргумента <*Массив_данных*ю>.

При выполнении данной работы <*Массив_данных*> - ссылка на ячейки, в которых расположены элементы исходной выборки, а <*Массив_карманов*> - ссылка на ячейки, в которых расположены значения правых границ карманов U1j. **Внимание!** Ячейка с границей последнего кармана в <*Массив _карманов*> не включается.

Функция ЧАСТОТА() относится к классу функций массива. Поэтому при ее программировании следует соблюдать следующий порядок действий:

1. Выделить все смежные ячейки, в которых будут расположены значения частот (в рассматриваемом примере - от H7 до H12 включительно).
2. В первую из выделенных ячеек записать функцию ЧАСТОТА(). Для рассматриваемого примера:
   =ЧАСТОТА(B2:B51;F7:F11).
3. Завершить программирование нажатием комбинации клавиш <Ctrl>+<Shift>+<Enter>.

**Внимание!** Комбинация клавиш <Ctrl>+<Shift>+<Enter> должна применяться независимо от того, как программируется функция ЧАСТОТА() - с использованием Мастера Функций или с клавиатуры.

После определения частот следует убедиться, что их сумма равна объему обрабатываемой выборки. В рассматриваемом примере проверка выполняется в ячейке H13:

=СУММ(H7:H12).

**Рассчитать частости.** В каждой из ячеек столбца частостей расчет выполняется по формуле (9). Например, для первого интервала в рассматриваемом примере (ячейка I7):

=H7/$D$3.

Для проверки правильности расчета частостей необходимо убедиться, что их сумма равна 1. В рассматриваемом примере проверка выполнена в ячейке I13:

=СУММ(I7:I12).

**Рассчитать накопленные частости.** Накопленная частость для первого кармана принимается равной относительной частоте в этом кармане. Для рассматриваемого примера, в ячейке J8 запрограммировано:

=I7.

В каждом из остальных карманов накопленная частость определяется суммированием накопленной частости из предыдущего кармана и относительной частоты в данном кармане. Например, для второго кармана в рассматриваемом примере (ячейка J9):

=J7+I8.

Если расчеты запрограммированы верно, накопленная частость для последнего кармана должна быть равна 1.

### 4.4.2.Построение теоретической кривой и проверка нормальности выборочного распределения

В примере на рис. 4.3 соответствующая область рабочего листа названа «Построение теоретических кривых». Здесь рассчитываются не только теоретические частоты $m_{\text{т}j}$, но также теоретические частости - как дифференциальные $f_{\text{т}j}$, так и кумулятивные $F_{\text{т}j}$. В последствие $f_{\text{т}j}$ и $F_{\text{т}j}$ будут использованы для графического отображения теоретического распределения вероятности анализируемого параметра.

Особенности расчетов поясним для первого кармана:

| *Параметр* | *Ячейка* | *Формула для расчета в MS Excel* |
|---|---|---|
| $z_j$ | K7 | =(G7-$E$3)/$F$3 |
| $f(z_j)$ | L7 | =1/КОРЕНЬ(2*ПИ())*EXP(-K7*K7/2) |
| $m_{\text{т}j}$ | M7 | =$N$3*L7 |
| $\chi^2_j$ | N7 | =(H7-M7)^2/M7 |
| $f_{\text{т}j}$ | O7 | =M7/$M$13 |

Теоретические кумулятивные частости рассчитываются так же, как и выборочные. Для первого кармана (в ячейке P7):

=O7.

В каждом из остальных карманов $F_{\text{т}j} = F_{\text{т}j-1} + f_{\text{т}j}$. Например, для второго кармана (ячейка P8):

=P7+O8.

Для проверки гипотезы о соответствии выборочного распределения нормальному закону в ячейке N17 суммируются $\chi_j^2$:

=СУММ(N7:N15),

а в ячейке N19, с помощью статистической функции ХИ2ОБР() определяется табличное значение $\chi^2[\alpha; k-3]$:

=ХИ2ОБР(1-N18;K3-3).

При этом уровень значимости $\alpha$ рассчитывается через доверительную вероятность $p$ ($\alpha = 1 - p$), значение которой вводится с клавиатуры в ячейку N18.

Вывод относительно нормальности распределения генерируется в ячейке L20 с помощью функции ЕСЛИ():

=ЕСЛИ(N17<N19;"Распределение можно считать нормальным";"Распределение нельзя считать нормальным")

### 4.4.3.Создание графических отображений выборочного распределения с применением Мастера Диаграмм

Предварительно создается таблица «Графическое отображение распределения» (в примере на рис. 4.3 - область E14:I25).

Ряд значений *x* анализируемого параметра (ячейки E18:E25) формируется следующим образом. Первое значение (в ячейке E18) меньше варианты первого кармана на величину длины кармана:

=G7-M3.

Последнее значение (в ячейке E25) больше варианты последнего кармана на величину длины кармана:

=G12+M3.

Остальные значения принимаются равными значениям вариант. Например, второе значение *x* (в ячейке E19)

=G7,

а предпоследнее (в ячейке E24)

=G12.

Значения плотности f(x) и функции F(x) выборочного распределения заносятся в таблицу путем ссылки на соответствующие значения fj  и Fj из таблицы «Выборочное распределение». Например, в ячейках F19 и G19 соответственно записано:

=I7 и =J7.

Значения плотности fт(x) и функции Fт(x) теоретического распределения заносятся в таблицу путем ссылки на соответствующие значения fтj  и Fтj из таблицы «Построение теоретических кривых». Например, в ячейках H19 и I19 соответственно записано:

=O7 и =P7.

Плотность распределения анализируемого параметра для рассматриваемого примера представлена на рис. 4.4.

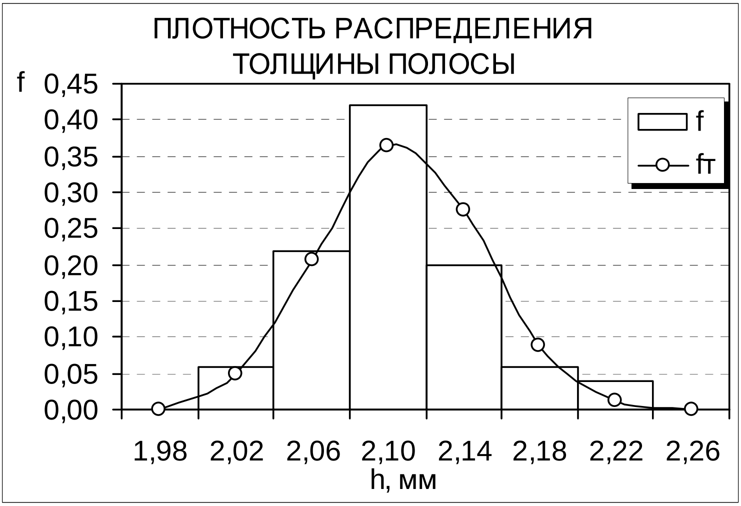


Рис.4.4. Иллюстрация плотности распределения анализируемого параметра:
f и fт – выборочная и теоретическая плотности

Используется диаграмма типа «График-гистограмма» из семейства нестандартных диаграмм. Выборочное распределение отображено гистограммой. Имя ряда – «f». Подписи по оси X для него заданы ссылкой на ячейки E18:E25, а значения – ссылкой на ячейки F18:F25. Теоретическое распределение отображено сглаженной линией. Имя ряда – «fт». Подписи по оси X для данного ряда также берутся из ячеек  E18:E25, но значения – из ячеек H18:H25.

Функция распределения анализируемого параметра для рассматриваемого примера представлена на рис. 4.5. Используется диаграмма типа «График» из семейства стандартных диаграмм. Выборочное распределение отображено рядом «F» в виде ломаной линии. Подписи по оси X для него заданы ссылкой на ячейки E18:E25, а значения – ссылкой на ячейки G18:G25. Теоретическое распределение отображено рядом «Fт» в виде сглаженной линии. Подписи по оси X для данного ряда также берутся из ячеек E18:E25, но значения – из ячеек I18:I25.

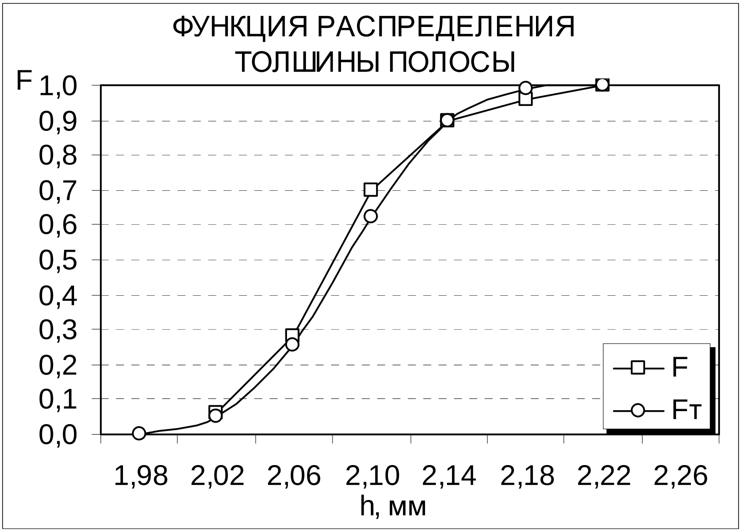


Рис. 4.5. Иллюстрация функции распределения анализируемого параметра:
F и Fт – выборочная и теоретическая функции

Результаты, полученные с применением Мастера Диаграмм, должны быть отредактированы путем форматирования различных элементов диаграммы. В частности, минимальное и максимальное значения, а также шаг шкалы по каждой из осей устанавливается форматированием соответствующей оси.

### 4.4.5.Анализ результатов оценивания вариации параметра

При оценивании вариации параметра представляют интерес ответы на следующие вопросы:

1. Распределение какого параметра было построено?
2. Можно ли считать распределение анализируемого параметра нормальным?
3. Чему равны допустимые отклонения исследуемого параметра?
4. Чему равны границы поля допуска исследуемого параметра?
5. Чему равны фактический и допустимый разброс исследуемого параметра?
6. Что следует из наблюдаемого соотношения допустимого и фактического разброса?

Для рассмотренного примера можно сделать следующие выводы.

1.По ГОСТ 19904 Прокат листовой холоднокатаный. Сортамент. предельные отклонения для листовой стали толщиной 2,5 мм равны $\delta_h^+$= 0,1 и $\delta_h^-$= 0,10 . Тогда минимальное допустимое значение $[h_{min}]$=2,5 -0,10=2,4 мм и максимально допустимое значение $[h_{max}]$ = 2,5 +0,10=2,6 мм.

3.По кумулятивной кривой процент годной продукции составляет 100 %.

### 4.5.Контрольные вопросы

1. Табличное представление вариационного ряда. Выборочное отображение плотности распределения
2. Табличное представление вариационного ряда. Выборочное отображение функции распределения
3. Поясните метод проверки нормальности распределения по критерию $\chi^2$ .
4. Применение выборочного распределения для анализа качества процесса
5. Применение выборочного распределения для оценки выхода годной продукции