# 第三节 随机误差的正态分布

## 一、 频率分布

在相同条件下对某样品中镍的质量分数（%）进行重复测定，得到90个测定值如下：

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1.60 | 1.67 | 1.67 | 1.64 | 1.58 | 1.64 | 1.67 | 1.62 | 1.57 | 1.60 |
| 1.59 | 1.64 | 1.74 | 1.65 | 1.64 | 1.61 | 1.65 | 1.69 | 1.64 | 1.63 |
| 1.65 | 1.70 | 1.63 | 1.62 | 1.70 | 1.65 | 1.68 | 1.66 | 1.69 | 1.70 |
| 1.70 | 1.63 | 1.67 | 1.70 | 1.70 | 1.63 | 1.57 | 1.59 | 1.62 | 1.60 |
| 1.53 | 1.56 | 1.58 | 1.60 | 1.58 | 1.59 | 1.61 | 1.62 | 1.55 | 1.52 |
| 1.49 | 1.56 | 1.57 | 1.61 | 1.61 | 1.61 | 1.50 | 1.53 | 1.53 | 1.59 |
| 1.66 | 1.63 | 1.54 | 1.66 | 1.64 | 1.64 | 1.64 | 1.62 | 1.62 | 1.65 |
| 1.60 | 1.63 | 1.62 | 1.61 | 1.65 | 1.61 | 1.64 | 1.63 | 1.54 | 1.61 |
| 1.60 | 1.64 | 1.65 | 1.59 | 1.58 | 1.59 | 1.60 | 1.67 | 1.68 | 1.69 |

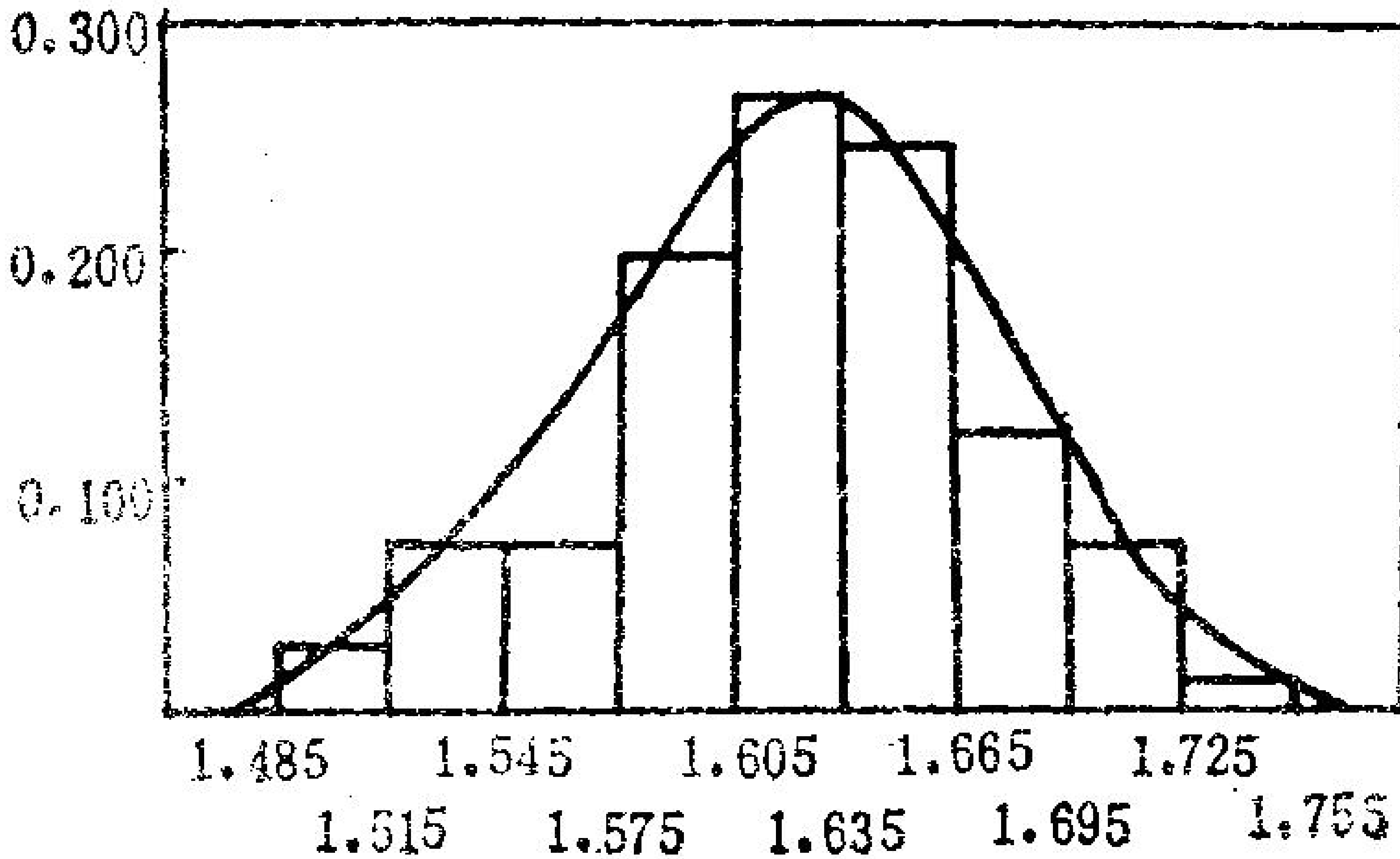
首先视样本容量的大小将所有数据分成若干组：容量大时分为10-20组，容量小时（n<50）分为5-7组，本例分为9组。

再将全部数据由小至大排列成序，找出其中最大值和最小值，算出极差R。由极差除以组数算出组距。本例中的R=1.74%-1.49%=0.25%，组距=R/9=0.25%/9=0.03%。每组内两个数据相差0.03%即：1.48-1.51,1.51-1.54等等。为了使每一个数据只能进入某一组内，将组界值较测定值多取一位。即：

1.485-1.515,1.515-1.545,1.545-1.575等等。

统计测定值落在每组内的个数（称为频数），再计算出数据出现在各组内的频率（即相对频数）。

| 分组（%） | 频数 | 频率 |
| --- | --- | --- |
| 1.485-1.515 | 2 | 0.022 |
| 1.515-1.545 | 6 | 0.067 |
| 1.545-1.575 | 6 | 0.067 |
| 1.575-1.605 | 17 | 0.189 |
| 1.605-1.635 | 22 | 0.244 |
| 1.635-1.665 | 20 | 0.222 |
| 1.665-1.695 | 10 | 0.111 |
| 1.695-1.725 | 6 | 0.067 |
| 1.725-1.755 | 1 | 0.011 |
| $\sum$ | 90 | 1.00 |



0.300
0.200
0.100
1.485
1.515
1.545
1.575
1.605
1.635
1.665
1.695
1.725
1.755

由表中的数据和图可以看出，测定数据的分布并非杂乱无章，而是呈现出某些规律性。在全部数据中，平均值1.62%所在的组（第五组）具有最大的频率值，处于它两侧的数据组，其频率值仅次之。统计结果表明：测定值出现在平均值附近的频率相当高，具有明显的集中趋势；而与平均值相差越大的数据出现的频率越小。

## 二、正态分布

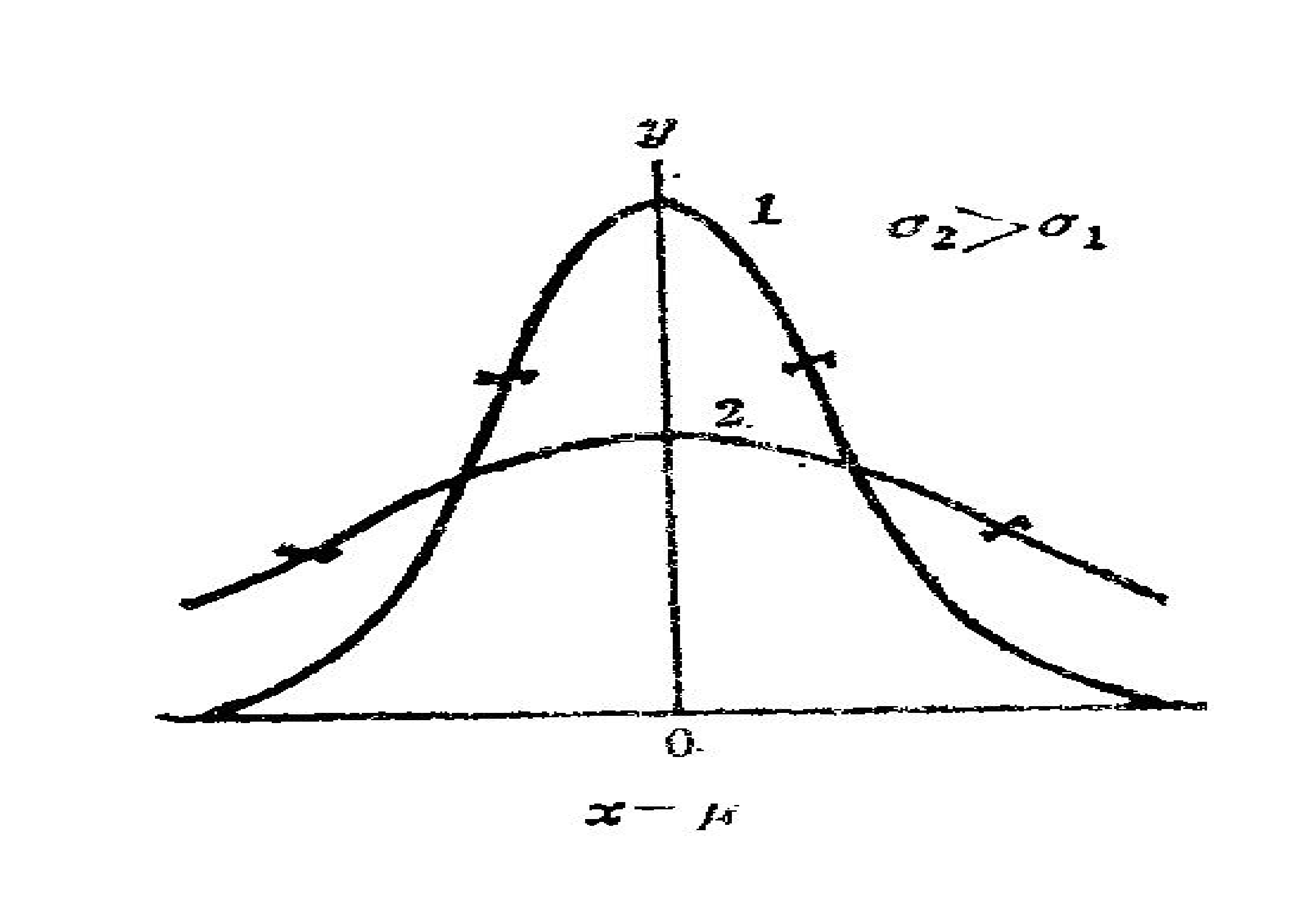
正态分布，又称高斯分布，它的数学表达式即正态分布函数式为：

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

说明了在等精密度的许多测定值中，平均值是出现概率最大的值。

式中的σ为总体标准偏差，是曲线两侧的拐点之一到直线x=μ的距离，它表征了测定值的分散程度。

标准偏差σ较小的曲线陡峭，表明测定值位于μ附近的概率较大，即测定的精密度高。与此相反，具有较大标准偏差较大的曲线平坦，表明测定值位于μ附近的概率较小，即测定的精密度低。




正态分布曲线

（μ相同，$\sigma_2 > \sigma_1$）

在μ和σ确定后，正态分布曲线的位置和形状也就确定，因此μ和σ是正态分布的两个基本参数，这种正态分布用N（μ，$\sigma^2$）表示。

正态分布曲线关于直线x=μ呈钟形对称，且具有以下特点：

1.对称性 绝对值大小相等的正负误差出现的概率相等，因此它们常可能部分或完全相互低消。

2.单峰性 峰形曲线最高点对应的横坐标x-μ值等于0，表明随机误差为0的测定值出现的概率密度最大。

3.有界性 一般认为，误差大于$|\pm 3\sigma|$的测定值并非是由随机误差所引起的。也就是说，随机误差的分布具有有限的范围，其值大小是界的。

# 三、标准正态分布

由于μ和σ不同时就有不同的正态分布，曲线的形状也随之而变化。为了使用方便，将正态分布曲线的横坐标改用u来表示（以σ为单位表示随机误差），并定义

$$u = \frac{x-\mu}{\sigma} \qquad (3\text{-}14)$$

代入得：

$$y = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

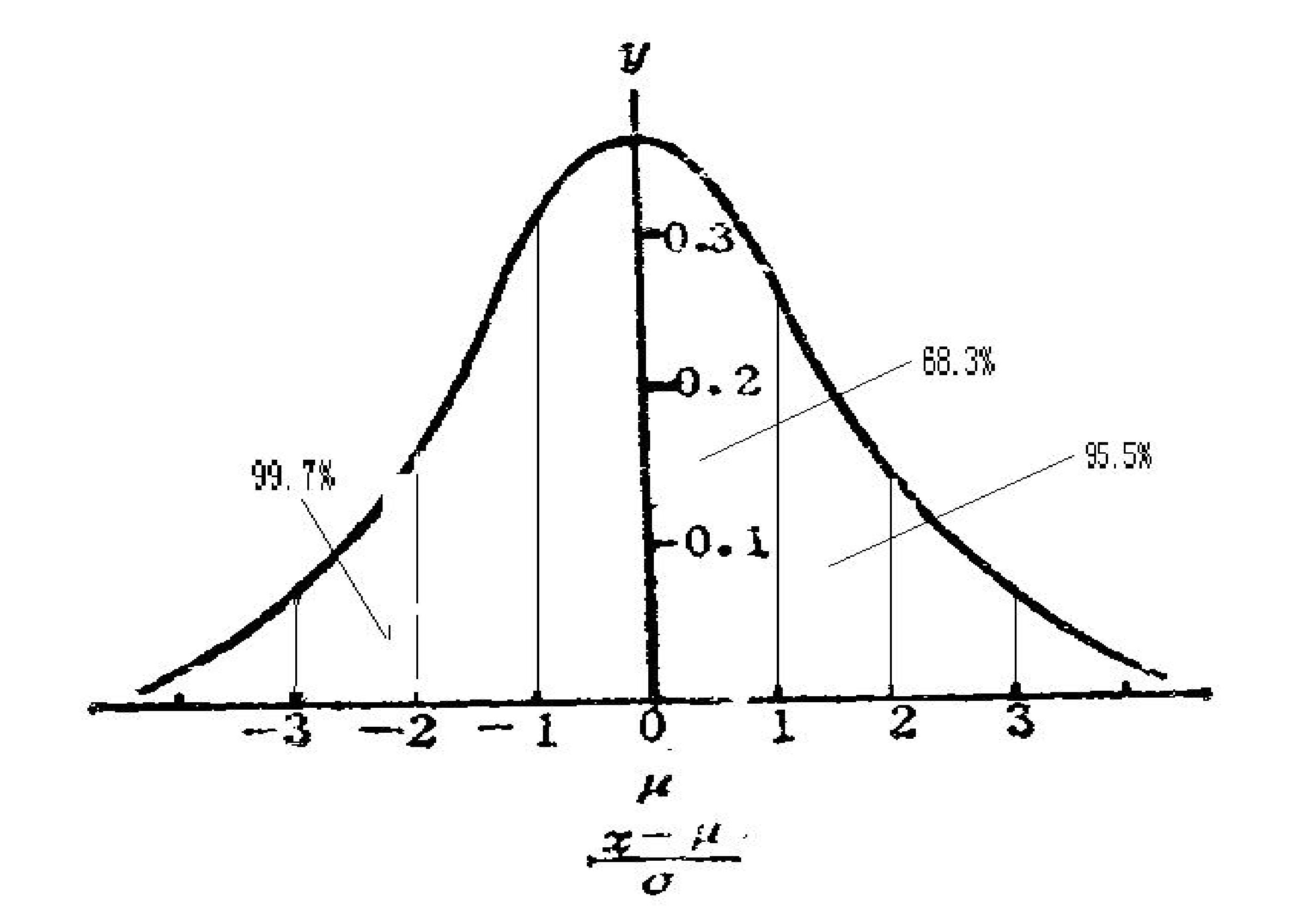
由于 $dx = \sigma du$

故 $$f(x)dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(u)du$$

u称为标准正态变量。此时式（3-13）就转化成只有变量u的函数表达式：

$$y = \Phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad \text{（3-15）}$$

经过上述变换，总体平均值为μ的任一正态分布均可化为μ=0，$\sigma^2$=1的标准正态分布，以N（0，1）表示。标准正态分布曲线如图3-5所示，曲线的形状与μ和σ的大小无关。




标准正态分布曲线

## 四、随机误差的区间概率

正态分布曲线与横坐标之间所夹的总面积，就等于概率密度函数从-∞至+∞的积分值。它表示来自同一总体的全部测定值或随机误差在上述区间出现概率的总和为100%，即为1。

$$\int_{-\infty}^{+\infty}\Phi(u)du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{u^2}{2}}du=1$$

欲求测定值或随机误差在某区间出现的概率P，可取不同的u值对式积分求面积而得到。例如随机误差在±σ区间（u=±1），即测定值在μ±σ区间出现的概率是：