Properties of a Normal Distribution

Recall from the previous chapter that continuous random variables can take on any value within an interval of values of a certain set. Some examples of continuous random variables are height in centimeters, mass in kilograms, time in minutes, and temperature in degrees Celsius. Continuous random variables follow continuous probability distributions. One special continuous distribution called the normal distribution will be discussed in this lesson. This distribution is important because most significant theories and applications in statistical inference are based on the existence of a normal distribution.

The normal distribution is a probability density function for a continuous random variable. This distribution is also known as the *Gaussian distribution* in honor of the German mathematician Johann Carl Friedrich Gauss (1777–1855), who derived its equation.

For a continuous random variable X that follows a normal distribution denoted by $X \sim N(\mu, \sigma)$, the probability density function is

$$f(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & , -\infty < x < \infty, -\infty, < \mu < \infty, \sigma > 0 \\ 0 & , \text{otherwise} \end{cases}$$

where μ is the mean and σ is the standard deviation.

Note that $e \approx 2.71828$ and $\pi \approx 3.1416$.

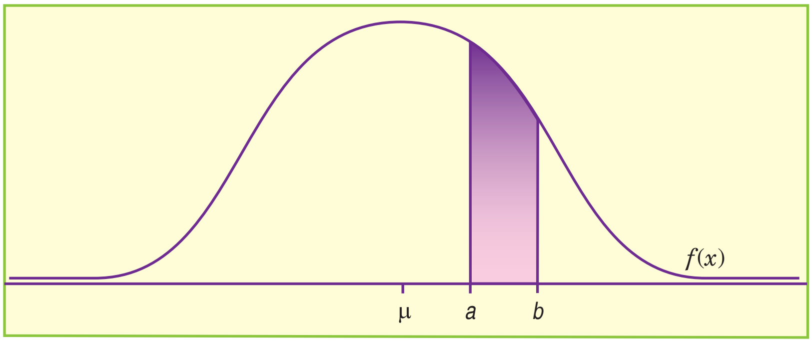

Figure 2.2 The probability $P(a < X < b)$ in a normal distribution with mean μ and standard deviation σ

Figure 2.2 shows a normal distribution with mean μ and standard deviation σ. The shaded region is the region bounded by the function $f(x)$, x-axis, $x = a$, and $x = b$. The area of this region is the probability that X lies between $x = a$ and $x = b$, denoted by $P(a < X < b)$.

For a normal distribution, about 68% lie in the interval $[\mu - \sigma, \mu + \sigma]$. This means that 68% are about 1 standard deviation away from the mean. About 95% lie in the interval $[\mu - 2\sigma, \mu + 2\sigma]$ while about 99.7% lie in the interval $[\mu - 3\sigma, \mu + 3\sigma]$. This is known as the *empirical rule*. Figure 2.3 shows the area in each region.

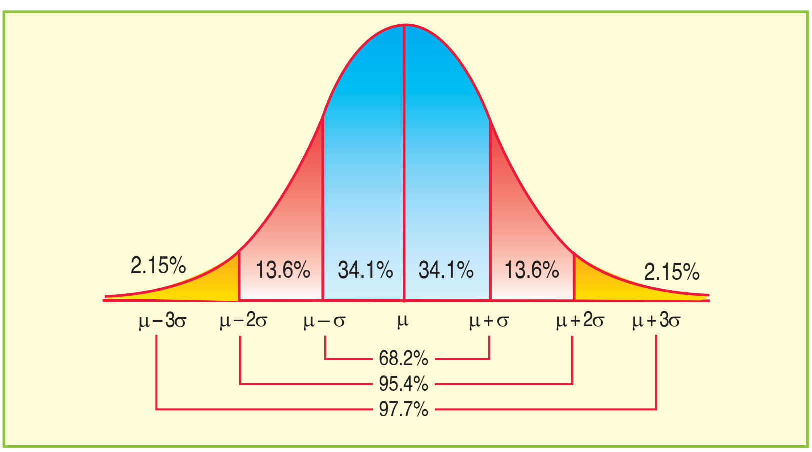

Figure 2.3 A normal distribution showing the regions as stated in the empirical rule