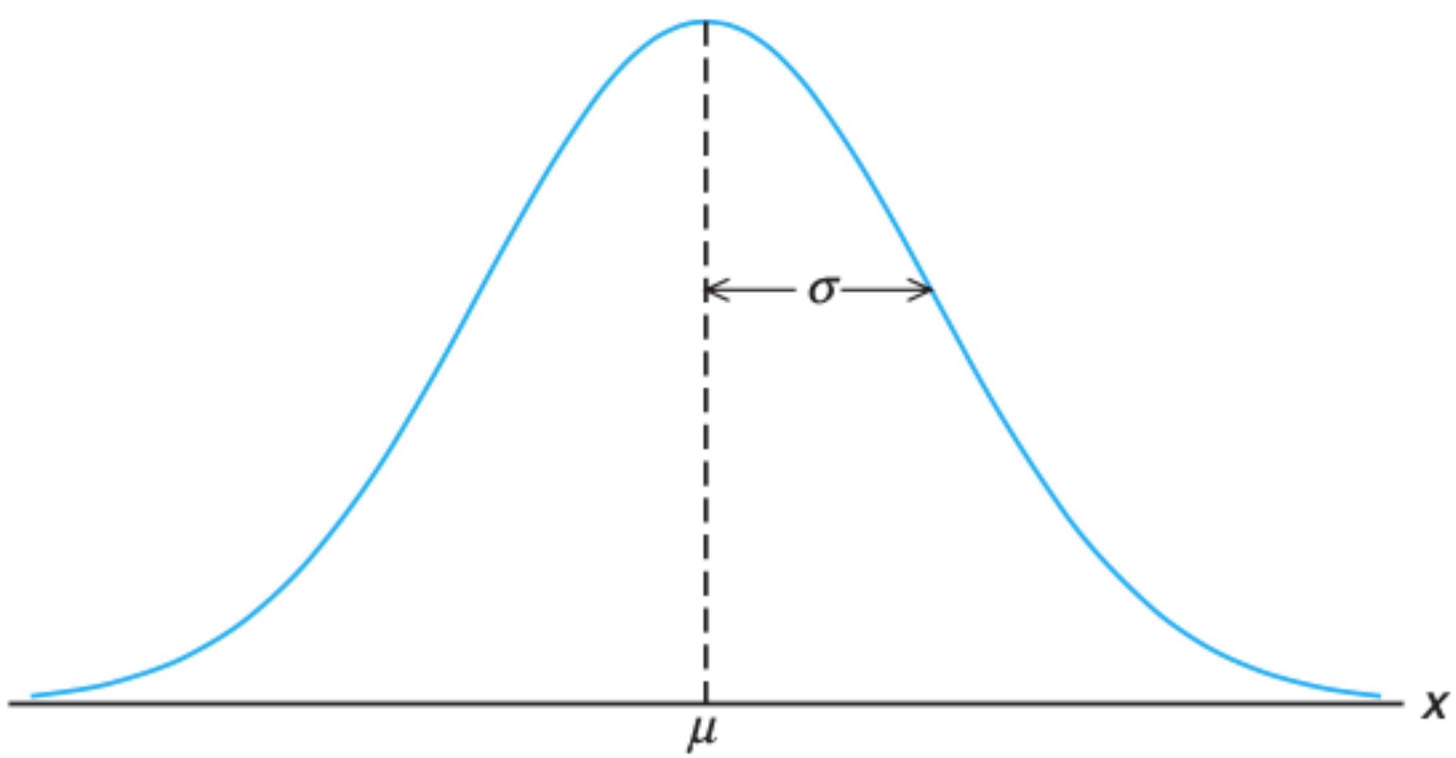
Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the normal distribution. Its graph, called the normal curve, is the bell-shaped curve which approximately describes many phenomena that occur in nature, industry, and research.



A continuous random variable X having the bell-shaped distribution is called a normal random variable. The mathematical equation for the probability distribution of the normal variable depends on the two parameters μ and σ, its mean and standard deviation, respectively. Hence, we denote the values of the density of X by n(x; μ, σ).

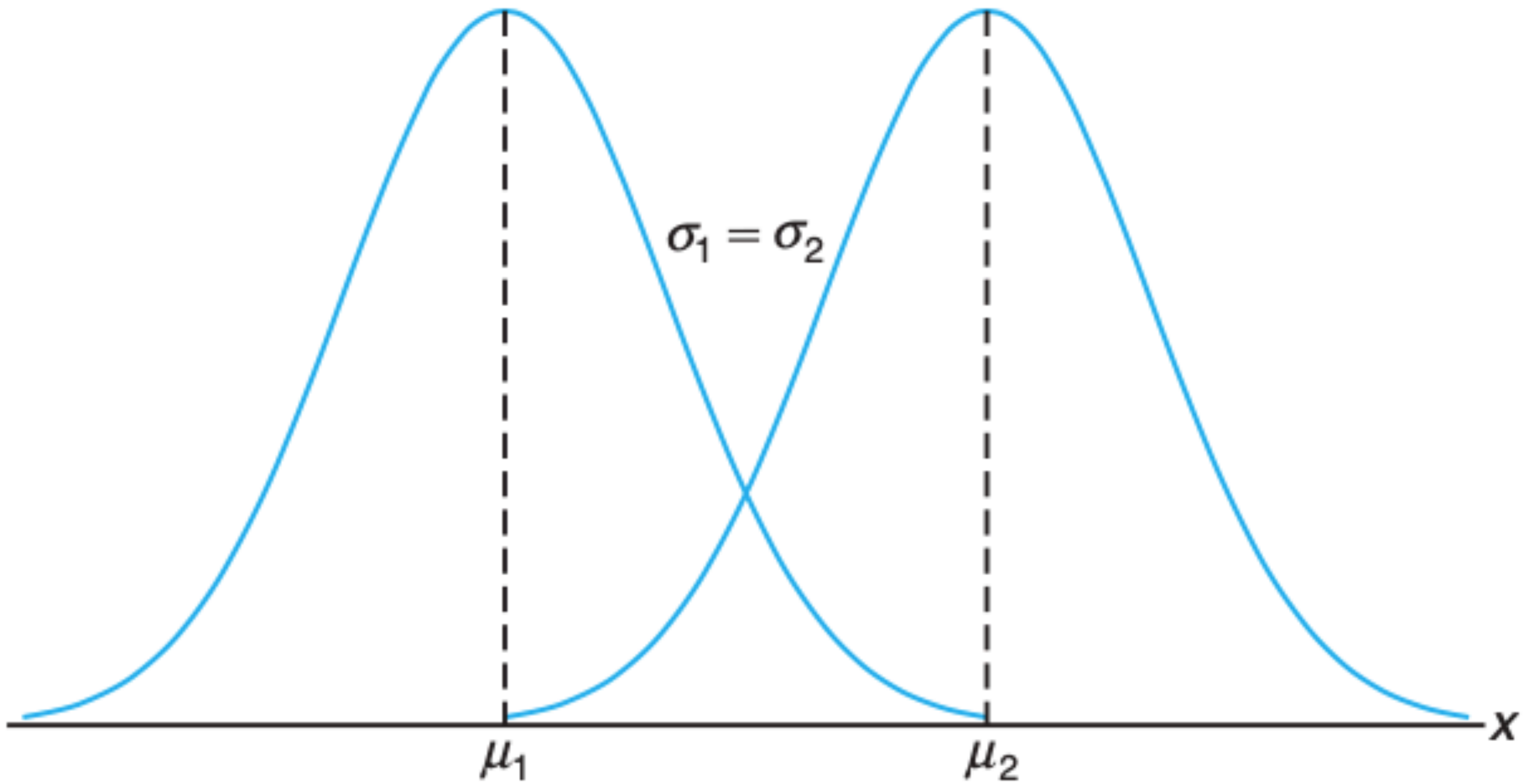
Definition

The density of the normal random variable X, with mean μ and variance σ^2, is

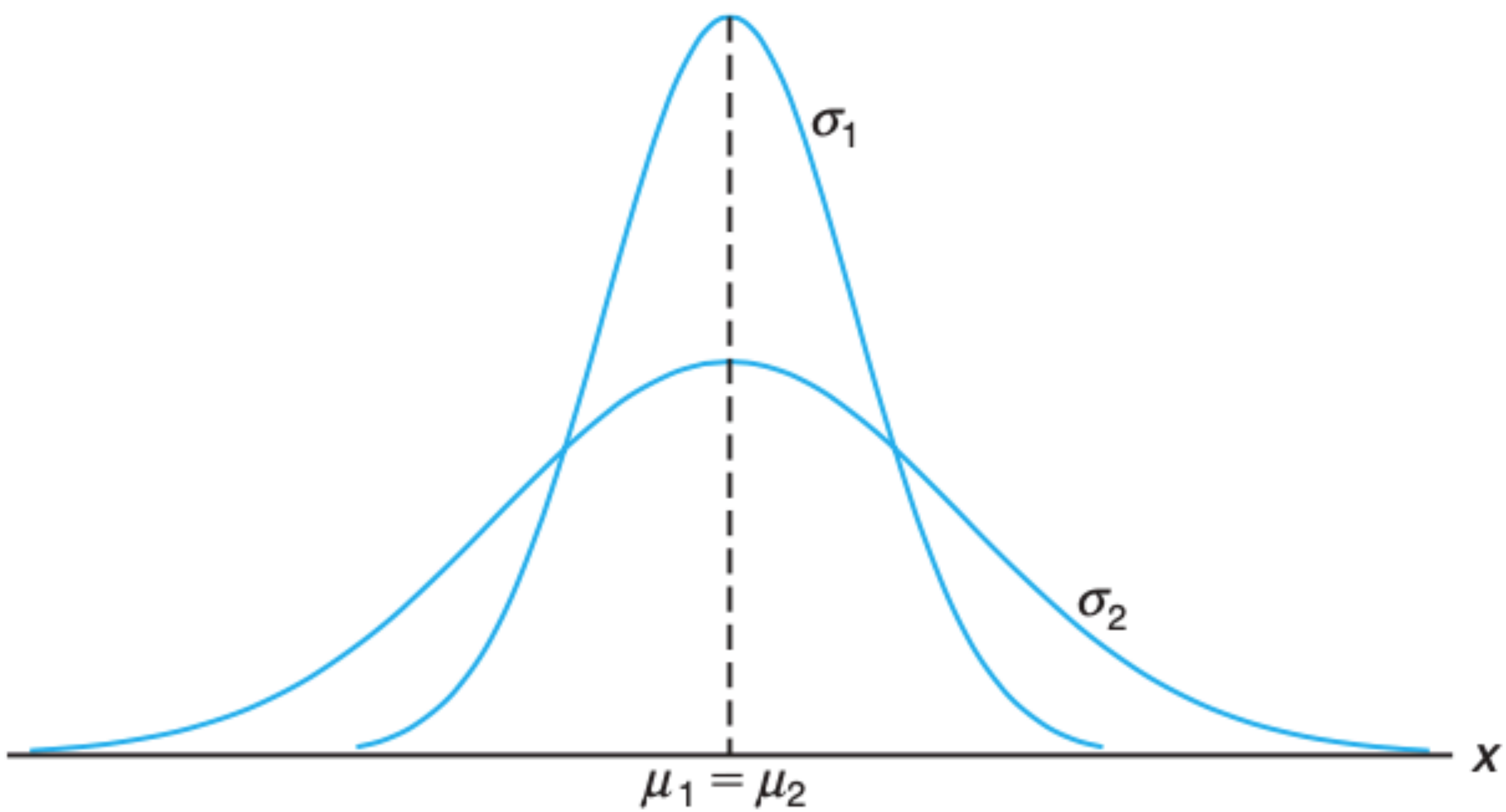
$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$

where $\pi = 3.14159\ldots$ and $e = 2.71828\ldots$.

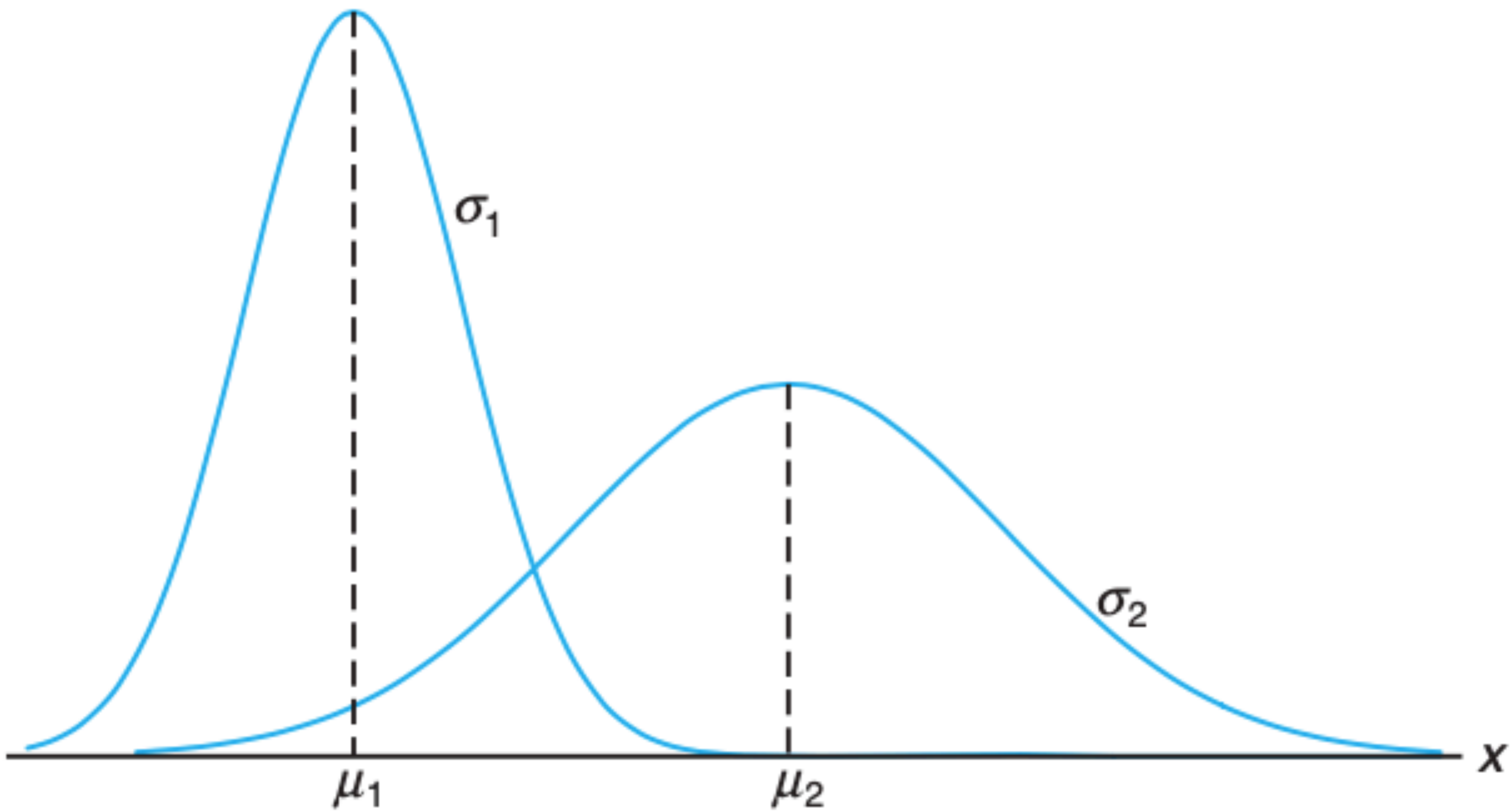
Once μ and σ are specified, the normal curve is completely determined.



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.



Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.



Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.

Properties of Normal Distribution

1. The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at $x = \mu$.
2. The curve is symmetric about a vertical axis through the mean μ.
3. The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.
4. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
5. The total area under the curve and above the horizontal axis is equal to 1.

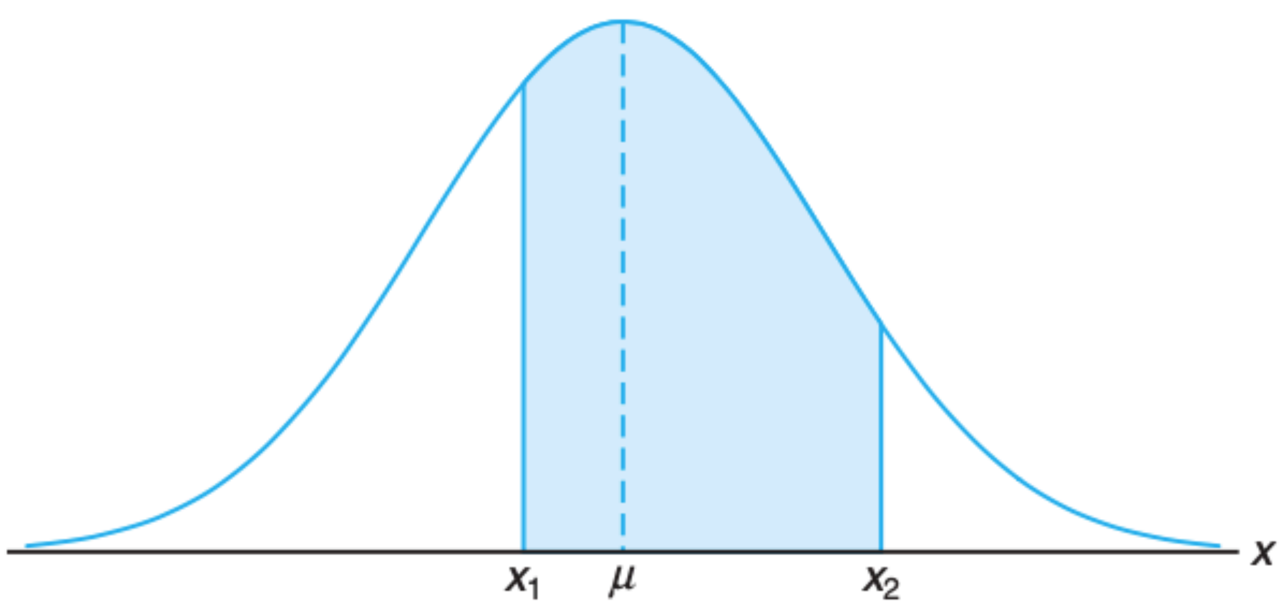
Note: The mean and variance of $n(x;\mu,\sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.

Areas under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates x = x1 and x = x2 equals the probability that the random variable X assumes a value between x = x1 and x = x2. Thus, for the normal curve

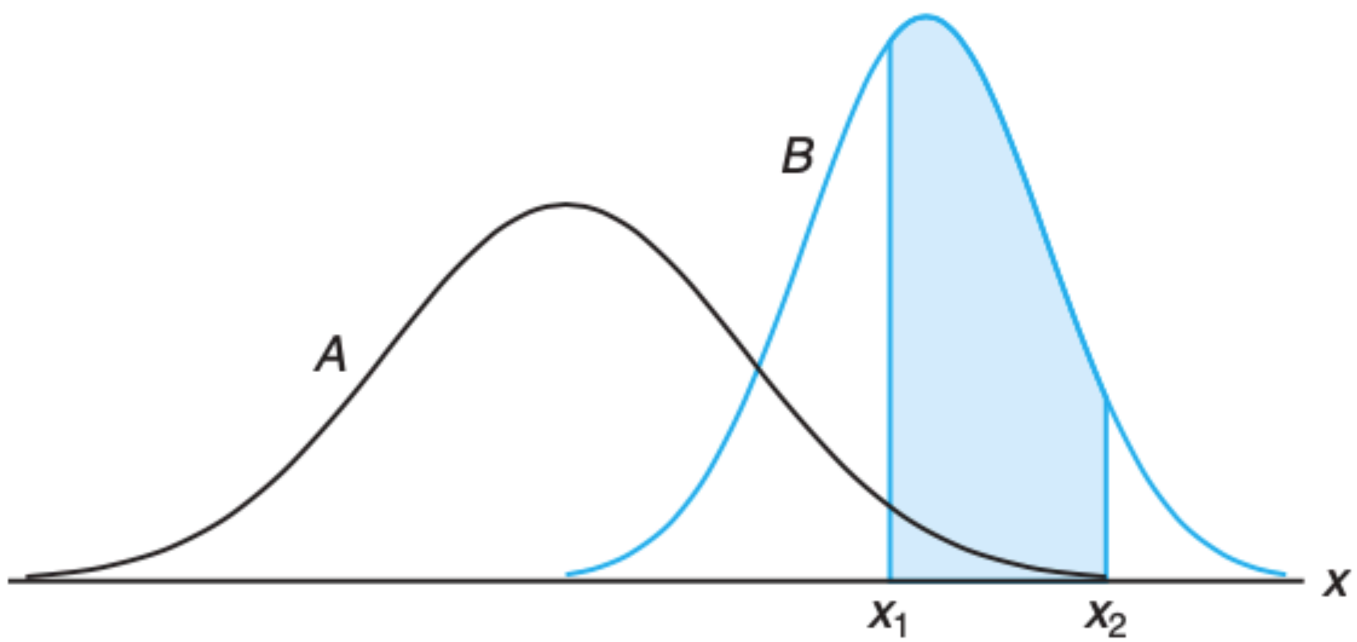
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma)\, dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

is represented by the area of the shaded region.



$P(x_1 < X < x_2) =$ area of the shaded region.

Note: The normal curve is dependent on the mean and the standard deviation of the distribution. The area under the curve between any two ordinates must then also depend on the values μ and σ.



$P(x_1 < X < x_2)$ for different normal curves.

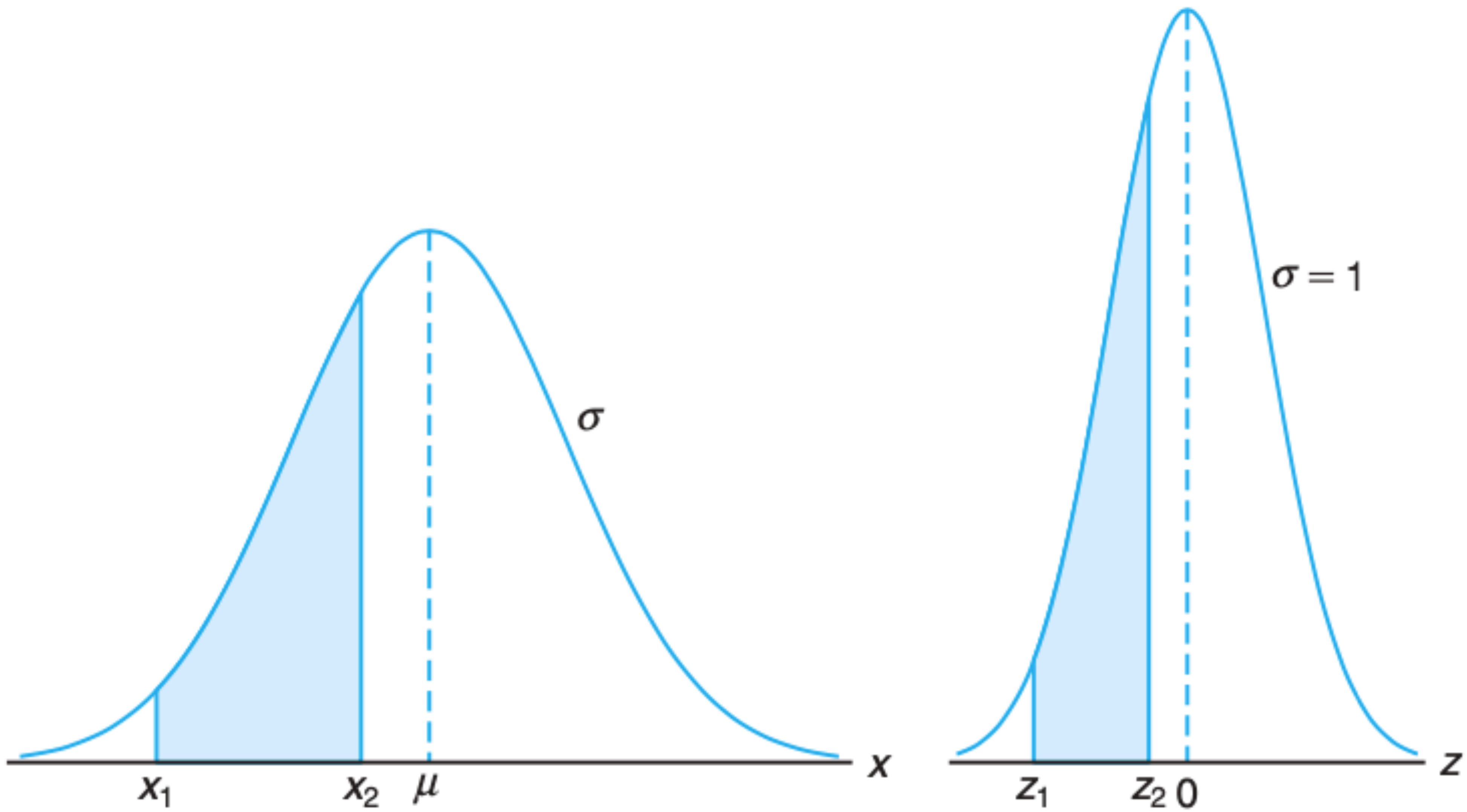
Standard Normal Distribution

There are many types of statistical software that can be used in calculating areas under the normal curve. The difficulty encountered in solving integrals of normal density functions necessitates the tabulation of normal curve areas for quick reference. However, it would be a hopeless task to attempt to set up separate tables for every conceivable value of μ and σ.

we are able to transform all the observations of any normal random variable X into a new set of observations of a normal random variable Z with mean 0 and variance 1. This can be done by means of the transformation

$$Z = \frac{x - \mu}{\sigma}$$

The distribution of a normal random variable with mean 0 and variance 1 is called a standard normal distribution.

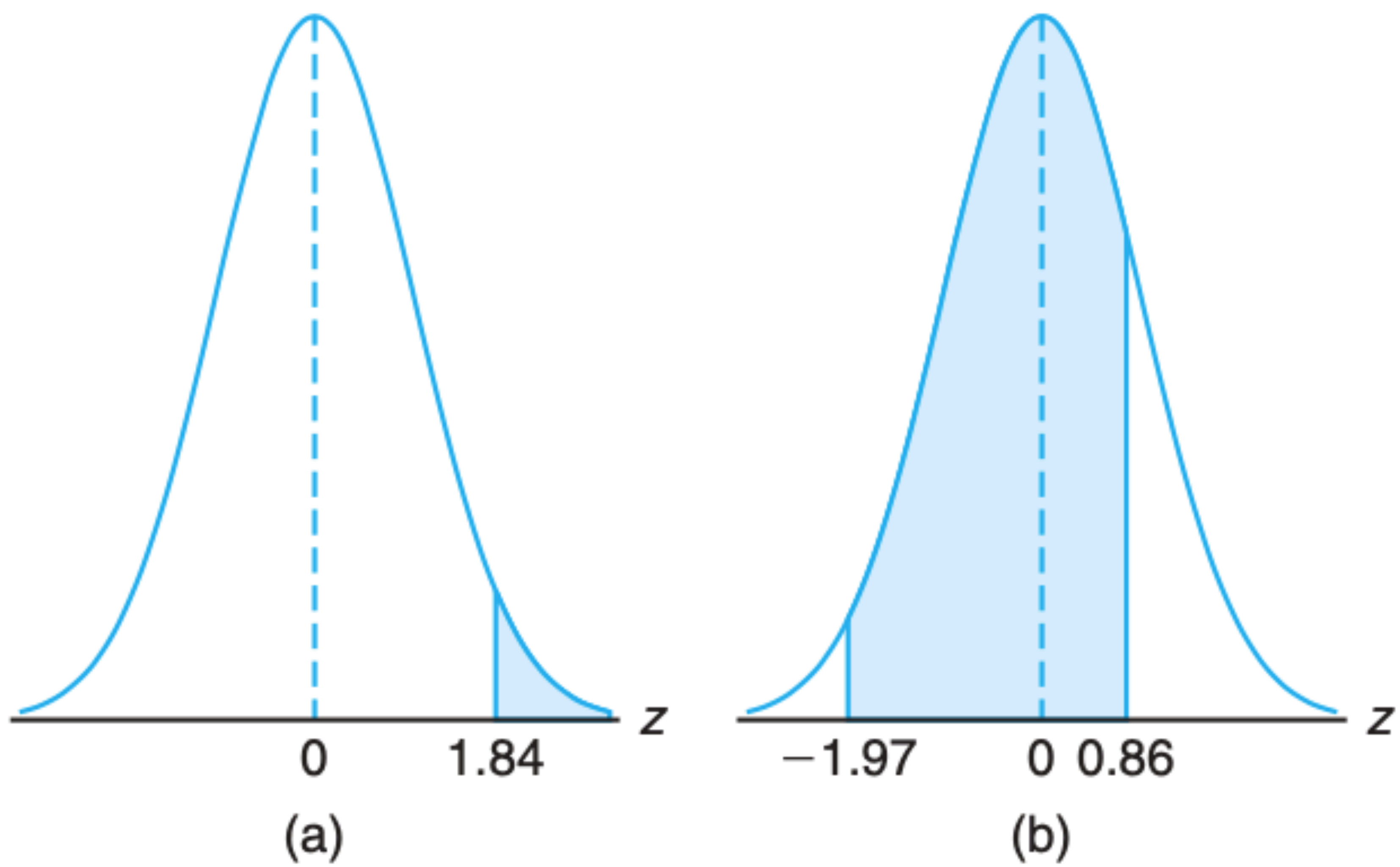


The original and transformed normal distributions.

Given a standard normal distribution, find the area under the curve that lies

(a) to the right of $z = 1.84$ and

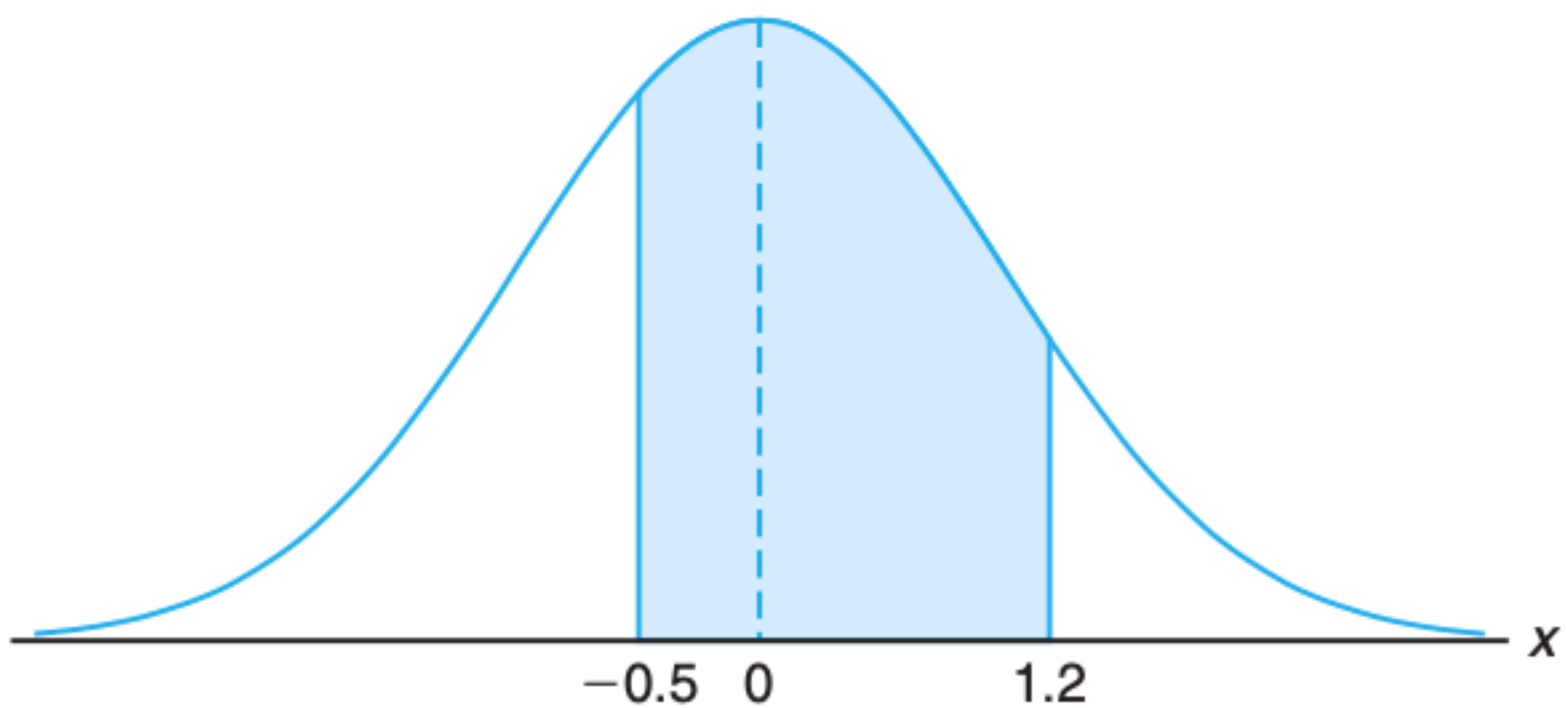
(b) between $z = -1.97$ and $z = 0.86$.



(a) The area in Figure (a) to the right of $z = 1.84$ is equal to 1 minus the area in Table A.3 to the left of $z = 1.84$, namely, $1 - 0.9671 = 0.0329$.

(b) The area in Figure (b) between $z = -1.97$ and $z = 0.86$ is equal to the area to the left of $z = 0.86$ minus the area to the left of $z = -1.97$. From Table we find the desired area to be $0.8051 - 0.0244 = 0.7807$.

Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.



The z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

$$z_1 = \frac{45 - 50}{10} = -0.5 \text{ and } z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$

$P(-0.5 < Z < 1.2)$ is shown by the area of the shaded region in Figure This area may be found by subtracting the area to the left of the ordinate $z = -0.5$ from the entire area to the left of $z = 1.2$. Using Table we have

$$\begin{aligned} P(45 < X < 62) &= P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5) \\ &= 0.8849 - 0.3085 = 0.5764. \end{aligned}$$

Using the Normal Curve in Reverse

6.6 Find the value of z if the area under a standard normal curve

(a) to the right of z is 0.3622;

(b) to the left of z is 0.1131;

(a)

Lets find the value of z, if the area under the normal curve to the right of z is 0.3622. In other words $P(Z > z) = 0.3622$.

$$P(Z > z) = 1 - P(Z \leq z)$$
$$P(Z \leq z) = 1 - 0.3622$$
$$P(Z \leq z) = 0.6378$$

$$z = 0.3526$$

(b)

Lets find the value of z, if the area under the normal curve to the left of z is 0.1131. In other words $P(Z < z) = 0.1131$.

$$z = -1.21$$

Applications of the Normal Distribution

A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) what fraction of the cups will contain more than 224 milliliters?

(b) what is the probability that a cup contains between 191 and 209 milliliters?

(c) how many cups will probably overflow if 230-milliliter cups are used for the next 1000 drinks?

(d) below what value do we get the smallest 25% of the drinks?

- Let X represent the amount of drink distributed.

-The mean is $\mu = 200$ milliliters per cup.

-The standard deviation $\sigma = 15$ milliliters per cup.

---

a)

Lets calculate what fraction of the cups will contain more than 224 milliters.

$$z = \frac{x - \mu}{\sigma} = \frac{224 - 200}{15} = 1.6$$

Therefore,

$$\begin{aligned} P(X > 224) &= P(Z > 1.6) \\ &= 1 - P(Z < 1.6) \\ &= 1 - 0.9452 \\ &= 0.0548 \end{aligned}$$

b)

Now, lets find the probability that a cup contains between 191 and 209 milliliters.

$$z = \frac{x_1 - \mu}{\sigma} = \frac{191 - 200}{15} = -0.6$$

$$z = \frac{x_2 - \mu}{\sigma} = \frac{209 - 200}{15} = 0.6$$

Therefore,

$$\begin{aligned} P(-0.6 < X < 0.6) &= P(-0.6 < Z < 0.6) \\ &= P(Z < 0.6) - P(Z < -0.6) \\ &= 0.7257 - 0.2743 \\ &= 0.4514 \end{aligned}$$

c)

Lets determine how many cups will probably overflow if 230 - milliliter cups are used for the next 1000 drinks.

$$z = \frac{x-\mu}{\sigma} = \frac{230-200}{15} = 2$$

Therefore,

$$\begin{aligned} P(X > 230) &= P(Z > 2) \\ &= 1 - P(Z < 2) \\ &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$

Hence, using the binomial property, we get:

$$\begin{aligned} E(X) &= n \cdot p \\ &= 1000 \cdot 0.0228 \\ &= 22.8 \approx 23 \end{aligned}$$

d)

Now, lets find below what value do we get the smallest 25% of the drinks.

$$
\begin{aligned}
P(X < x) &= 0.25 \\
P\left(Z < \frac{x-\mu}{\sigma}\right) &= 0.25 \\
P(Z < -0.68) &= 0.25 \\
\frac{x-\mu}{\sigma} &= -0.68 \\
x &= -0.68 \cdot \sigma + \mu \\
x &= -0.67 \cdot 15 + 200 \\
x &= 189.95
\end{aligned}
$$

A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.

(a) What is the probability that a trip will take at least 1/2 hour?

(b) If the office opens at 9:00 A.M. and the lawyer leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work?

(c) If he leaves the house at 8:35 A.M. and coffee is served at the office from 8:50 A.M. until 9:00 A.M., what is the probability that he misses coffee?

(d) Find the length of time above which we find the slowest 15% of the trips.

(e) Find the probability that 2 of the next 3 trips will take at least 1/2 hour.

- Let X represents the trip times from lawyer's suburban home to his midtown office normally distributed with the mean 24 minutes and a standard deviation of 3.8 minutes.

---

a)

Lets find the probability that a trip will take at least 1/2 hour.

$$
\begin{aligned}
P(X \geq 30) &= 1 - P(X < 30) \\
&= 1 - P\left(\frac{X-\mu}{\sigma} < \frac{30-24}{3.8}\right) \\
&= 1 - P(Z < 1.58) \\
&= 1 - 0.9429 \\
&= 0.0571
\end{aligned}
$$

b)

Lets find what percentage of the time is lawyer late for work if the office opens at 9: 00 A.M. and the he leaves his house at 8: 45 A.M. daily.

$$\begin{aligned}
P(X > 15) &= 1 - P(X \leq 15) \\
&= 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{15 - 24}{3.8}\right) \\
&= 1 - P(Z \leq -2.37) \\
&= 1 - 0.0089 \\
&= 0.9911 = 99.11\%
\end{aligned}$$

c)

Now lets calculate the probability that he misses coffee, If he leaves the house at 8: 35 A.M. and coffee is served at the office from 8: 50 A.M. until 9: 00 AM.

$$
\begin{aligned}
P(X > 25) &= 1 - P(X \leq 25) \\
&= 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{25-24}{3.8}\right) \\
&= 1 - P(Z \leq 0.23) \\
&= 1 - 0.6026 \\
&= 0.3974
\end{aligned}
$$

d)

Lets find the length of time above which we find the slowest 15% of the trips.

$$
\begin{aligned}
P(X > x) &= 0.15 \\
1 - P(X \leq x) &= 0.15 \\
P(X \leq x) &= 1 - 0.15 \\
P(X \leq x) &= 0.85 \\
P\left(Z \leq \frac{x - \mu}{\sigma}\right) &= 0.85 \\
P(Z < 1.04) &= 0.85 \\
\frac{x - \mu}{\sigma} &= 1.04 \\
x &= 1.04 \cdot \sigma + \mu \\
x &= 1.04 \cdot 3.8 + 24 \\
x &= 3.952 + 24 \\
x &= 27.952
\end{aligned}
$$

e)

Lets determine the probability that 2 of the next 3 trips will take at least 1/2 hour.

$$
\begin{aligned}
P(X > 30) &= 1 - P(X \leq 30) \\
&= 1 - P\left(\frac{X - \mu}{\sigma} < \frac{30 - 24}{3.8}\right) \\
&= 1 - P(Z < 1.58) \\
&= 1 - 0.9429 \\
&= 0.0571
\end{aligned}
$$

Therefore, X has binomial distribution with parameters $n = 3$ and $p = 0.0571$

The probability mass function of X is:

$$
\begin{aligned}
P(X=x) &= \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \\
&= \left(\begin{array}{c} 3 \\ x \end{array} \right) (0.0571)^x (1-0.0571)^{3-x} \qquad x = 0,1,2,3
\end{aligned}
$$

Therefore the probability that 2 of the next 3 trips will take at least 1/2 hour is:

$$
\begin{aligned}
P(X=2) &= \left(\begin{array}{c} 3 \\ 2 \end{array} \right) (0.0571)^2 (1-0.0571)^{3-2} \\
&= 3 \cdot 0.00326 \cdot 0.9429 \\
&= 0.0092
\end{aligned}
$$