Lecture 8

Random Variables and Probability Distributions

Definition

- A random variable is a **variable** that takes on **numerical values** determined by the **outcome of a random experiment**.

Notation

- It is important to distinguish between a random variable and the possible values that it can take.
- Capital letters, such as X is used to denote the random variable and the corresponding lowercase letter, x, is used to denote a possible value.
- **Example,** we can use the random variable X to denote the outcome of throwing a die. This random variable can take the specific values x = 1, x = 2, ….., x = 6, each with probability P(X = 1)=……………P(X = 6) = $\frac{1}{6}$

Types of Random Variable

There are two types of random variable:

(a) Discrete (b) Continuous

Discrete Random Variable

A *random variable* that assumes countable values is called a *discrete random variable*.

Example:

1. The number of cars sold at a dealership during a given month
2. The number of houses in a certain village
3. The number of customers who visit a bank during any given hour
4. The number of heads obtained in two tosses of a coin

Continuous Random Variable

A random variable is a continuous random variable if it can take any value in an interval.

The following are some examples of continuous random variables:

1. The time taken to commute from home to work
2. The weight of a fish

Probability Distribution of a Discrete Random Variable

- Let *X be* a discrete random variable. The **probability distribution** of *X* describes how the probabilities are distributed over all the possible values of X.
- The *probability distribution of a discrete random variable* lists all the possible values that the random variable can assume and their corresponding probabilities.

EXAMPLE

Suppose the following Table gives the frequency of the number of vehicles owned by all 2000 families living in a small town. **Let *X* be the number of vehicles owned by a randomly selected family. Write the probability** distribution of X.

Table: Probability Distribution of the Number of Vehicles Owned by Families

Number of Vehicles Owned X	Frequency	Relative Frequency/ Probability, P(x)
0	30	.015
1	470	.235
2	850	.425
3	490	.245
4	160	.080
Total	N=2000	Sum=1.00

Two Characteristics of a Probability Distribution

The probability distribution of a discrete random variable possesses the following two characteristics.

1. $0 \le P(x) \le 1$ for each value of *x*
2. $\sum P(x) = 1$

Presentation of a Probability Distribution

1. By using table
2. Graphically
3. Mathematical formula

Example

The following table lists the probability of the number of breakdowns per week for a machine based on past data.

Breakdowns per week	0	1	2	3
Probability, P(x)	0.15	0.20	0.35	0.30

a. Show that above table is a probability distribution table.
b. Present this probability distribution graphically.
c. Find the probability that the number of breakdowns for this machine during a given week is
   i. exactly 2 **ii.** 0 to 2 **iii.** more than 1 **iv.** at most 1

Solution

a. Let *X denote* the number of breakdowns for this machine during a given week.

The following Table is a probability distribution of *X* because

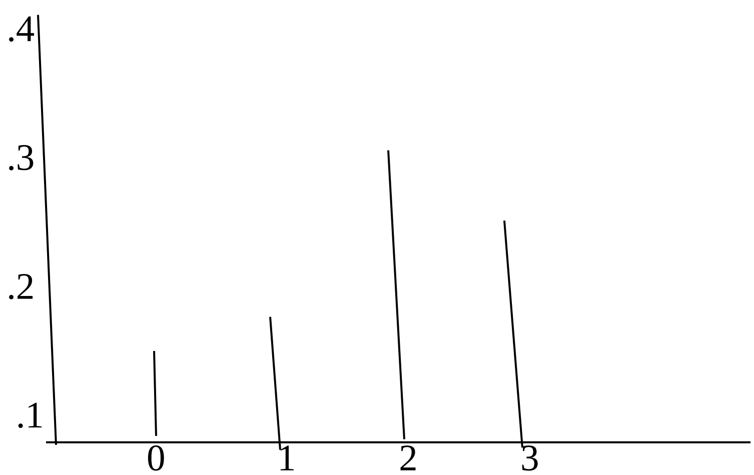
$0 \leq P(x) \leq 1$ for each value of x

$\sum P(x) = 1.$

Table: Probability Distribution of the Number of Breakdowns

X	P(x)
0	.15
1	.20
2	.35
3	.30
	$\sum P(x)$ =1.0

(b) The following Figure shows the graph of the probability distribution of the above Table.



(b) We can calculate the required probabilities as follows:

i. The probability of exactly two breakdowns is

$$P(\text{exactly 2 breakdowns}) = P(X = 2) = \mathbf{.35}$$

ii. The probability of 0 to 2 breakdowns is given by the sum of the probabilities of 0, 1, and 2 breakdowns:

$$\begin{aligned} P(\text{0 to 2 breakdowns}) &= P(0 \le X \le 2) \\ &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.15 + 0.20 + 0.35 \end{aligned}$$

iii. The probability of more than 1 breakdown is obtained by adding the probabilities of 2 and 3 breakdowns:

$$\begin{aligned} P(\text{more than 1 breakdown}) &= P(X > 1) \\ &= P(X = 2) + P(X = 3) \\ &= .35 + .30 \\ &= .65 \end{aligned}$$

iv. The probability of at most 1 breakdown is given by the sum of the probabilities of 0 and 1 breakdown:

$$\begin{aligned} P(\text{at most 1 breakdown}) &= P(X \le 1) \\ &= P(X = 0) + P(X = 1) \\ &= .15 + .20 \\ &= .35 \end{aligned}$$

Mean of a Discrete Random Variable

The *mean of a discrete random variable X* is denoted by μ and is calculated as

$$\mu = E(X) = \sum_x xP(X = x)$$

Variance and Standard Deviation of Discrete Random Variable

Let X be a discrete random variable. The variance, denoted σ^2 is given by

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 P(X = x)$$

The standard deviation, σ is the positive square root of the variance.

Example: An automobile dealer calculates the probability of new cars sold that have been returned various numbers of times for the correction of defects during the warranty period. The results are shown in the following table.

Number of returns	0	1	2	3	4
Probability P(X=x)	0.28	0.36	0.23	0.09	0.04

(a) Find the mean number of returns of an automobile for the corrections for defects during the warranty period.

(b) Find the variance of the number of returns of an automobile for corrections for defects during the warranty period.

Solution: The expected value is

$$\mu = E(X) = \sum_x xP(x) = (0)(.28) + 1(0.36) + 2(0.23) + 3(0.09) + 4(0.04) = 1.25$$

The variance is $\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 P(x)$

$$= (0-1.25)^2(0.28) + (1-1.25)^2(0.36) + (2-1.25)^2(0.23) +$$
$$(3-1.25)^2(0.09) + (4-1.25)^2(0.04) = 1.695$$

Probability Distributions

Probability Distributions for Discrete Random Variable

The Binomial Probability Distribution

An experiment that satisfies the following four conditions is called a **binomial experiment**.

1. There are *n* identical trials.
2. Each trial has two and only two outcomes. These outcomes are usually called a *success* and a *failure,* respectively.
3. The probability of success is denoted by *p* and that of failure by *q*, and *p* + *q* = 1. The probabilities *p* and *q* remain constant for each trial.
4. The trials are independent. In other words, the outcome of one trial does not affect the outcome of another trial.

Binomial Formula

For a binomial experiment, the probability of exactly *x* successes in *n* trials is given by the binomial formula

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where,

n = total number of trials
p = probability of success
q = 1 - *p* = probability of failure
x = number of successes in *n* trials
n - *x* = number of failures in *n* trials

Mean and Variance

Let X be the number of successes in n independent trials, each with probability of success p, then X follows a binomial distribution with mean

$$\mu = E(X) = np$$

and variance

$$\sigma^2 = E[(X-\mu)^2] = np(1-p)$$

Example: An insurance broker believes that for a particular contract the probability of making a sale is 0.4. Suppose that the broker has five contracts.

a. Find the probability that she makes at most one sale.
b. Find the probability that she makes between two and four sales (inclusive).
c. Graph the probability distribution function.

Solution:

a. P(At most 1 sale) $= P(X \le 1) = P(X=0) + P(X=1)$.

$$P(X=0) = \binom{5}{0}(0.4)^0(0.6)^5 = 0.078$$

$$P(X=1) = \binom{5}{1}(0.4)^1(0.6)^4 = 5(0.4)(0.6)^4 = 0.259$$

Hence $P(X \le 1) = 0.078 + 0.259 = 0.337$

b. $P(2 \le X \le 4) = P(X=2) + P(X=3) + P(X=4)$

$$P(X=2) = \binom{5}{2}(0.4)^2(0.6)^3 = 10(0.4)^2(0.6)^3 = 0.346$$

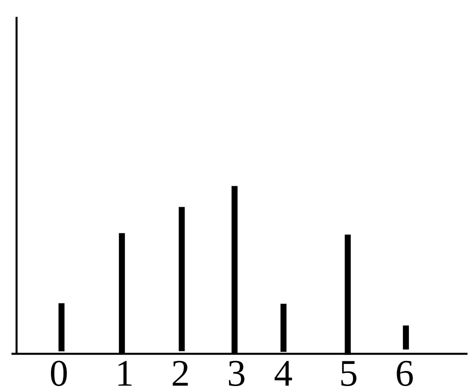
$$P(X=3) = \binom{5}{3}(0.4)^3(0.6)^2 = 10(0.4)^3(0.6)^2 = 0.230$$

$$P(X=4) = \binom{5}{4}(0.4)^4(0.6)^1 = 5(0.4)^4(0.6)^1 = 0.077$$

Hence $P(2 \le X \le 4) = 0.346 + 0.230 + 0.077 = 0.653$

c. The probability distribution function is shown in Figure 1.

X	P(X=x)
0	0.078
1	0.259
2	0.346
3	0.230
4	0.077
5	0.010



Probability of Success and the Shape of the Binomial Distribution

For any number of trials *n*:

1. The binomial probability distribution is symmetric if $p = 0.50$.
2. The binomial probability distribution is skewed to the right if $p < 0.50$.
3. The binomial probability distribution is skewed to the left if $p > 0.05$.

The Poisson Probability Distribution

- Assume that an interval is divided into a very large number of subintervals so that the probability of the occurrence of an event in any subinterval is very small.
- The assumptions of a Poisson probability distribution are as follows:
    1. The probability of the occurrence of an event is constant for all subintervals.
    2. There can be no more than one occurrence in each subinterval.
    3. Occurrences are independent; that is, the occurrences in non-overlapping intervals are independent of one another.

Examples of Poisson Distribution

1. Number of Network Failures per Week
2. Number of Bankruptcies Failed per Month
3. Number of Website Visitors per Hour
4. Number of Arrivals at a Restaurant
5. Number of Calls per Hour at a Call Center
6. Number of Books Sold per Week
7. Average Number of Storms in a City
8. Number of Emergency Calls Received by a Hospital Every Minute

Poisson Probability Distribution Formula

According to the *Poisson probability distribution*, the probability of *x* occurrences in an interval is

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \text{for x = 0, 1, 2,…………..}$$

where λ (pronounced *lambda*) is the mean number of occurrences in that interval and the value of *e* is approximately 2.71828.

- The λ is called the *parameter of the Poisson probability distribution* or the **Poisson parameter**.
- As is obvious from the Poisson probability distribution formula, we need to know only the value of λ to compute the probability of any given value of *x*.

The mean and variance of the Poisson probability distribution are

$$\mu = E(X) = \lambda \qquad \text{and} \quad \sigma^2 = E[(X-\mu)^2] = \lambda$$

Example: A professor receives, on average, 4.2 telephone calls from students the day before a final examination. If the distribution of calls is Poisson, what is the probability of receiving at least three of these calls on such a day?

Solution: The distribution of X, number of telephone calls, is Poisson with $\lambda = 4.2$. Thus the probability distribution function is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-4.2}(4.2)^x}{x!}$$

The probability of receiving at least three calls is

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - [P(X=0) + P(X=1) + P(X=2)] \end{aligned}$$

Now $P(X = 0) = \frac{e^{-4.2}(4.2)^0}{0!} = 0.015$

$$P(X = 0) = \frac{e^{-4.2}(4.2)^1}{1!} = 0.063$$

$$P(X = 0) = \frac{e^{-4.2}(4.2)^2}{2!} = 0.132$$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - [0.015 + 0.063 + 0.132] \\ &= 1 - .021 = 0.79 \end{aligned}$$

Example:

A washing machine breaks down an average of three times per month. Using the Poisson probability distribution formula, find the probability that during the next month this machine will have

(a) exactly two breakdowns

(b) at most one breakdown

Solution Let λ be the mean number of breakdowns per month, and let *x* be the actual number of breakdowns observed during the next month for this machine. Then,

$$\lambda = 3$$

(a) The probability that exactly two breakdowns will be observed during the next month

$$P(X=2)= \frac{e^{-3}3^{x}}{2!} = .2240$$

(b) $$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= \frac{e^{-3}3^{0}}{0!} + \frac{e^{-3}3^{1}}{1!} = 0.0498+0.1494=.1992$$

Poisson Approximation to the Binomial Distribution

- The Poisson probability distribution is obtained by starting with the Binomial probability distribution with p approaching 0 and n becoming very large.
- Thus, it follows that the Poisson distribution can be used to approximate the binomial probabilities when the number of trials, n is large and at the same time the probability, p, is small (generally such that $\lambda = np \leq 7$)

Example: An analysts predicted that 3.5% of all small corporations would file for bankruptcy in the coming year. For a random sample of 100 small corporations, estimate the probability that at least 3 will file for bankruptcy in the next year, assuming that the analyst's prediction is correct.

Solution: The distribution of X, the number of fillings for bankruptcy, is binomial with n = 100 and p = 0.035, so that the mean of the distribution is $\mu_x = np = 3.5$. Using the Poisson distribution to approximate the probability of at least 3 bankruptcies, we find

$$P(X \geq 3) = 1 - P(X \leq 2)$$

$$P(X = 0) = \frac{e^{-3.5}(3.5)^{0}}{0!} = e^{-3.5} = 0.030197$$

$$P(X = 1) = \frac{e^{-3.5}(3.5)^{1}}{1!} = (3.5)(0.030197) = 0.1056895$$

$$P(X = 2) = \frac{e^{-3.5}(3.5)^{2}}{2!} = (6.125)(0.030197) = 0.1849566$$

Thus, $P(X \leq 2) = P(0) + P(1) + P(2)$

$$= 0.030197 + 0.1056895 + 0.1849566$$

$$= 0.3208431$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.3208431 = 0.6791569$$

The binomial probability of $X \geq 3$ is

$P(X \geq 3) = 0.68$

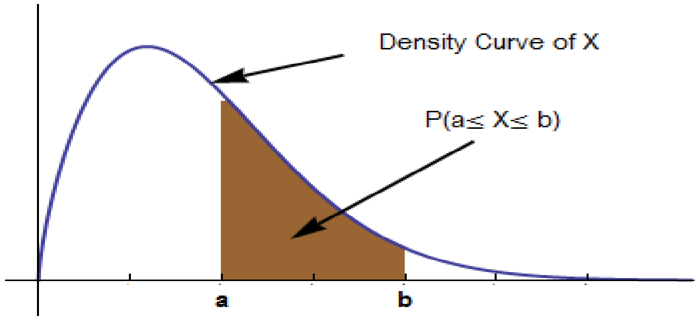
Continuous Random Variables and Probability Distributions

- We defined a **continuous random variable** as a random variable whose values are not countable.
- A continuous random variable can assume any value over an interval
- The probability that a continuous random variable X assumes a single value is always zero.

Probability Density Function:

Let X be a continuous random variable, and let x be any number lying in the range of values this random variable can take. The probability density function, ***f(x)***, of the random variable is a function with the following properties:

1. f(x) > 0 for all values of x.
2. The area under the probability density function, f(x), over all values of the random variable, X, is equal to 1.0.
3. Suppose that this density function is graphed. Let a and b be two possible values of random variable X, with a < b. Then the probability that X lies between a and b is the area under the density function between these points.



4. The cumulative distribution function, F(x_0), is the area under the probability density function , f(x), up to x_0:

$$F(x_0) = \int_{-x_m}^{x_0} f(x)dx$$

where x_m is the minimum value of the random variable x.

The Normal Distribution

The probability density function for a normally distributed random variable X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{for} \quad -\infty < x < \infty$$

where μ and σ^2 are any numbers such that $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$ and e and π are physical constants, e = 2.71828…. and $\pi = 3.14159.......$

Properties of the Normal Distribution

Suppose that the random variable X follows a normal distribution with parameters μ and σ^2. Then the following properties hold:

1. The mean of the random variable is μ:
$$E(X) = \mu$$
2. The variance of the random variable is σ^2:
$$\text{Var(X)} = E[(X-\mu)^2] = \sigma^2 E[(X-\mu)^2] = \sigma^2$$
3. The shape of the probability density function is a symmetrical bell-shaped curve centered on the mean, μ, as shown in the following Figure.
4. If we know the mean and variance, we can define the normal distribution by using the notation
$$X \sim N(\mu, \sigma^2)$$

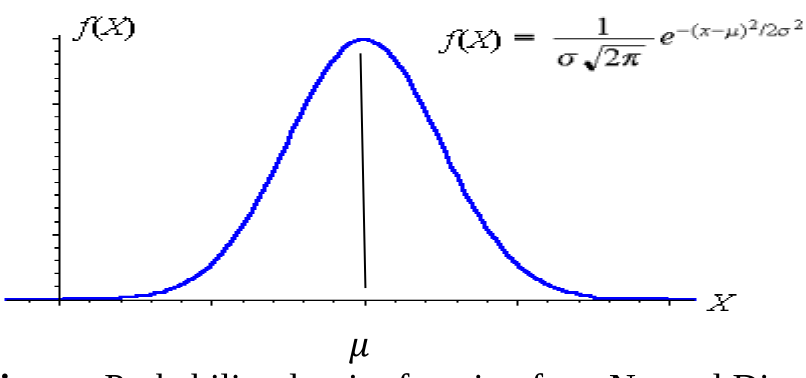


Figure: Probability density function for a Normal Distribution.

- The mean, μ , and the standard deviation, σ^2, are the *parameters* of the normal distribution.

- Given the values of these two parameters, we can find the area under a normal distribution curve for any interval. Each different set of values of μ and σ^2 gives a different normal distribution.
- The value of μ determines the center of a normal distribution curve on the horizontal axis, and the value of σ^2 gives the spread of the normal distribution curve.

Figure: Effects of μ and σ^2 on the probability density function of a Normal random variable.

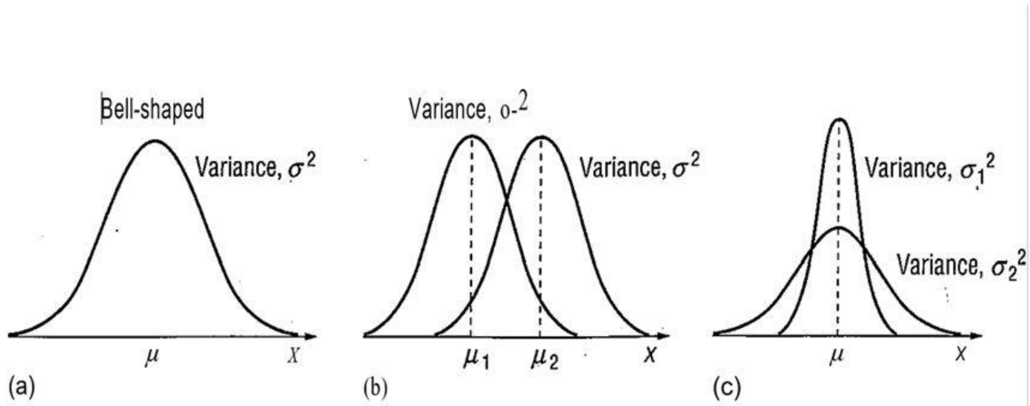


Figure 7.2 The probability density function of the Normal distribution of the variable x.

(a) Symmetrical about mean μ: variance = σ^2

(b) Effect of changing mean (μ2 > μ1)

(c) Effect of changing variance ($\sigma1^2 < \sigma2^2$)

Cumulative Distribution Function (CDF) of the Normal Distribution

Suppose that X is a normal random variable with mean μ and variance σ^2 ; that is, $X \sim N(\mu, \sigma^2)$. Then the cumulative distribution function is

$$F(x_0) = P(X \le x_0)$$

This is the area under the normal probability density function to the left of x_0. As for any proper density function, the total area under the curve is 1; that is,

$$F(\infty) = 1$$

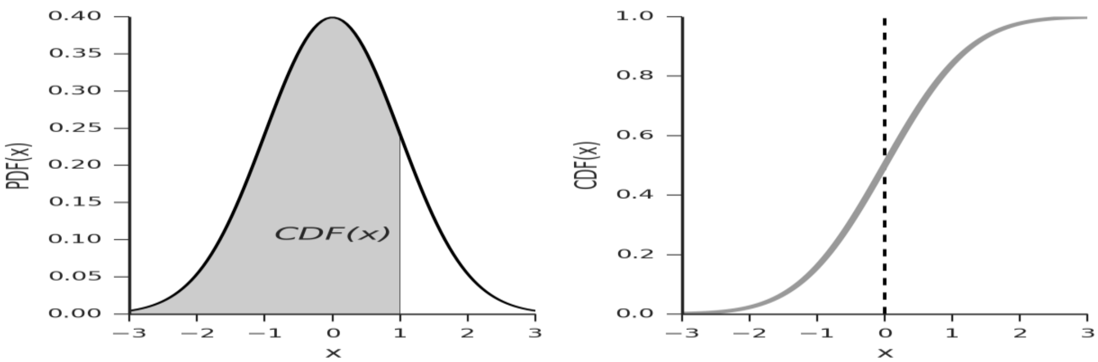


Figure: The shaded area is the probability that X does not exceed x_0 for a normal random variable.

Range Probabilities for Normal Random Variables

Let X be a normal random variable with cumulative distribution function F(x), and let a and b be two possible values of X, with a < b. Then

$$P(a < X < b) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$
$$= F(b) - F(a)$$

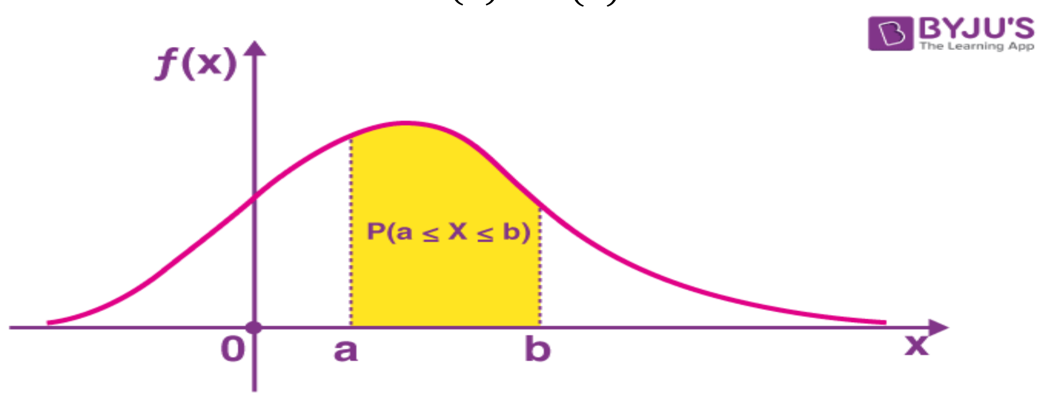


Figure: Normal density function with the shaded area indicating the probability that X is between a and b.

The Standard Normal Distribution

- The standard normal distribution, also called the z-distribution, is a special normal distribution where the mean is 0 and the standard deviation is 1.

- Any normal distribution can be standardized by converting its values into z scores. Z scores tell you how many standard deviations from the mean each value lies.

How to calculate a *z* score

Let X is a normally distributed random variable, i.e., $X \sim N(\mu, \sigma^2)$

To standardize a value from a normal distribution, convert the individual value into a *z*-score:

1. Subtract the mean from your individual value.
2. Divide the difference by the standard deviation.

***Z*-score formula**	**Explanation**
$z = \dfrac{x - \mu}{\sigma}$	• x = individual value • μ = mean • σ = standard deviation

Z be a normal random variable with mean 0 and variance 1; that is,

$$Z \sim N(0, 1)$$

Denote the cumulative distribution function as F(z) and a and b as two numbers with a < b; then

$$P(a < Z < b) = F(b) - F(a)$$

We can obtain probabilities for any normally distributed random variable by first converting the random variable to the standard normally distributed random variable, Z. There is always a direct relationship between any normally distributed random variable $X \sim N(\mu, \sigma^2)$ and Z. That relationship uses the transformation

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

This important result allows us to use the standard normal table to compute probabilities associated with any normally distributed random variable.

Example

Find the following areas under the standard normal curve.

(a) Area to the left of z = - 1.54
(b) Area to the right of z = 2.32

Solution

(a) Area to the left of z = - 1.54 = P(z < -1.54) = **.0618**
(b) Area to the right of z = 2.32 = P(z > 2.32) = 1.0 – P(z < 2.32)
= 1- .9898 = **.0102**

Sampling Distributions of Sample Means

Sample Mean
Let the random variables X_1, X_2,……….,X_n denote a random sample from a population. The sample mean value of these random variables is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Consider the sampling distribution of the random variable $\overline{X}$. We can determine the mean and variance of the sampling distribution.

$$E(\overline{X}) = \mu$$

The variance of the sample mean is denoted by $\sigma_{\bar{x}}^2$ and the corresponding standard deviation, called the standard error of $\overline{X}$, is given by

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Standard Normal Distribution for the Sample Means
Whenever the sampling distribution of the sample means is a normal distribution, we can compute a standardized normal random variable, Z, that has mean 0 and variance 1:

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Example: Suppose that the annual percentage salary increases for the chief executive officers of all midsize corporations are normally distributed with mean 12.2% and standard deviation 3.6%. A random sample of nine observations is obtained from this population and the sample mean computed. What is the probability that the sample mean will be less than 10%?

Solution: We know that

$$\mu = 12.2 \qquad \sigma = 3.6 \qquad n = 9$$

Let $\bar{x}$ denote the sample mean, and compute the standard error of the sample mean

$$\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{9}} = 1.2}$$

Then we compute

$$P(\bar{x} < 10) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{10-12.2}{1.2}\right) = P(Z < -1.83) = 0.0336$$

From this analysis we conclude that the probability that the sample mean will be less than 10% is only 0.0336.

Example: A spark plug manufacturer claims that the lives of its plugs are normally distributed with mean 36,000 miles and standard deviation 4,000 miles. A random sample of 16 plugs had an average life of 34,500 miles. If the manufacture's claim is correct, what is the probability of finding a sample mean of 34,500 or less?

Solution: Here

$$\mu = 36000, \qquad \sigma = 4000, \qquad n = 16$$

The desired probability is

$$P(\bar{x} < 34{,}500) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{34{,}500-36{,}000}{4000/\sqrt{16}}\right)$$
$$= P\left(\frac{-1500}{1000}\right) P(Z < -1.50) = 0.0668$$

We find that the probability that X is less than 34,500 is 0.0668. This probability suggests that, if the manufacturer's claims-- $\mu = 36{,}000$ and $\sigma = 4{,}000$ --are true, then a sample mean of 34,500 or less has a small probability. As a result, we are skeptical about the manufacturer's claims.