Statistics for Quality

Dr. Rajiv Nandan Rai
SCSQR
IIT Kharagpur

Population and Sample

- A population is the set of all items that possess a certain characteristic of interest.

- A sample is a subset of a population. Realistically, in many manufacturing or service industries, it is not feasible to obtain data on every element in the population. Measurement, storage, and retrieval of large volumes of data are impractical, and the costs of obtaining such information

Parameter and Statistic

- A parameter is a characteristic of a population, something that describes it.
- A statistic is a characteristic of a sample. It is used to make inferences on the population parameters that are typically unknown.

Subir Chowdhury School of Quality and Reliability

DESCRIPTIVE STATISTICS

Measures of Central Tendency (Location)

- Measures of central tendency tell us something about the location of the observations and the value about which they cluster and thus help us decide whether the settings of process variables should be changed.

- Mean : The mean is the simple average of the observations in a data set. In quality control, the mean is one of the most commonly used measures. It is used to determine whether , on average, the process is operating around a desirable target value. The sample mean, or average (denoted by X) , is found by adding all observations in a sample and dividing by the number of observations(n) in that sample. If the ith observation is denoted by Xi, the sample mean is calculated as:

Mean

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

The population mean(μ) is found by adding all the data values in the Population and dividing by the size of the population(N). It is calculated as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

The population mean is sometimes denoted as E(X), the expected value of the random variable X. It is also called the mean of the Probability distribution of X.

Mean

A random sample of five observations of the waiting time of customers in a Bank is taken. The times (in minutes) are 3, 2, 4, 1, and 2. The sample average .X., or Mean waiting time is

$$\overline{X} = \frac{3+2+4+1+2}{5} = \frac{12}{5} = 2.4 \text{ minutes}$$

The bank can use this information to determine whether the Waiting time needs to be improved by increasing the number of tellers.

Median

Given that the observations in a sample are $x_1, x_2, \ldots, x_n$, arranged in increasing order of magnitude, the sample median is

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

As an example, suppose the data set is the following: 1.7, 2.2, 3.9, 3.11, and 14.7. The sample mean and median are, respectively,

$$\bar{x} = 5.12, \quad \tilde{x} = 3.9.$$

Median

The median is the value in the middle when the observations are ranked. If there are an even number of observations, the simple average of the two middle numbers is chosen as the median. The median has the property that 50% of the values are less than or equal to it.

A random sample of 10 observations of piston ring diameters (in millimeters) yields the following values: 52.3, 51.9, 52.6, 52.4, 52.4, 52.1, 52.3, 52.0, 52.5, and 52.5. We first rank the observations:

51.9	52.0	52.1	52.3	52.3
52.4	52.4	52.5	52.5	52.6

The observations in the middle are 52.3 and 52.4. The median is $(52.3 + 52.4)/2$, or 52.35.

The median is less influenced by the extreme values in the data set; thus, it is said to be more "robust" than the mean.

Median

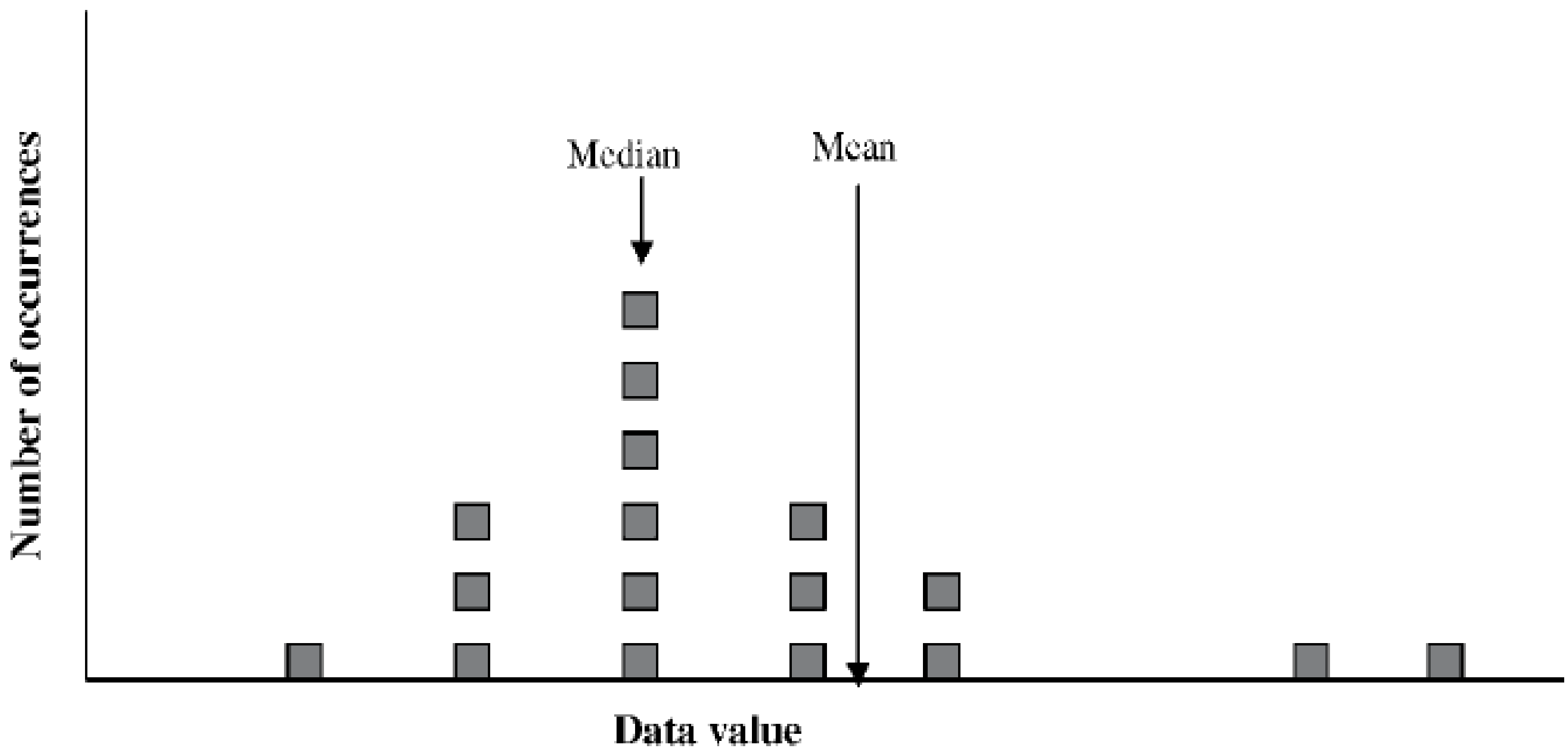
A department store is interested in expanding its facilities and wants to do a preliminary analysis of the number of customers it serves. Five weeks are chosen at random, and the number of customers served during those weeks were as follows:

$$3000, \quad 3500, \quad 500, \quad 3300, \quad 3800$$

The median number of customers is 3300, while the mean is 2820. On further investigation of the week with 500 customers, it is found that a major university whose students frequently shop at the store was closed for spring break. In this case, the median (3300) is a better measure of central tendency than the mean (2820) because it gives a better idea of the variable of interest. In fact, had the data value been 100 instead of 500, the median would still be 3300, although the mean would decrease further, another demonstration of the robustness of the median.

Outliers (values that are very large or very small compared to the majority of the data points) can have a significant influence on the mean, which is pulled toward the outliers. Figure(next slide) demonstrates the effect of outliers.

Median



Effect of outliers on a mean.

Example

- Often the nature of the scientific study will dictate the role that probability and deductive reasoning play in statistical inference. A study was conducted at the Virginia Polytechnic Institute and State University on the development of a relationship between the roots of trees and the action of a fungus. Minerals are transferred from the fungus to the trees and sugars from the trees to the fungus.

- Two samples of 10 northern red oak seedlings were planted in a greenhouse, one containing seedlings treated with nitrogen and the other containing seedlings with no nitrogen. All other environmental conditions were held constant. The stem weights in grams were recorded after the end of 140 days. The data are given in Table (Next Slide)

Example

Table 1: Data Set for Tutorial

No Nitrogen	Nitrogen
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52
0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

In this example there are two samples from two **separate populations**. The purpose of the experiment is to determine if the use of nitrogen has an influence on the growth of the roots.

Example

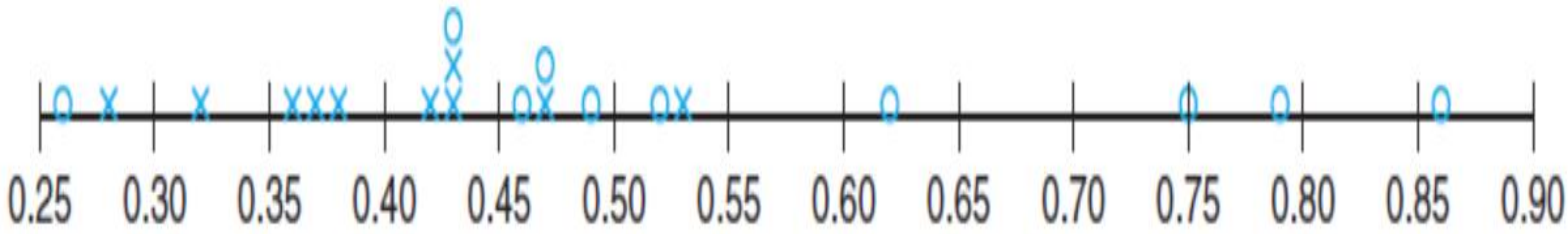


Figure 1: A dot plot of stem weight data.

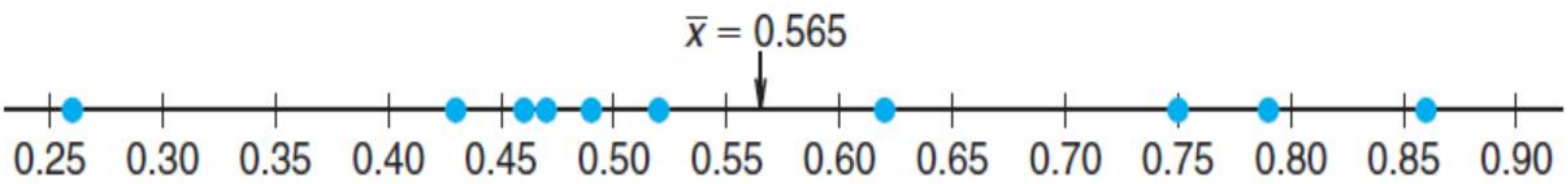
- The study is a comparative study (i.e., we seek to compare the two populations with regard to a certain important characteristic).
- It is instructive to plot the data as shown in the dot plot of Figure 1.
- The ◦ values represent the "nitrogen" data and the × values represent the "no-nitrogen" data.
- The general appearance of the data might suggest to the reader that, on average, the use of nitrogen increases the stem weight. Four nitrogen observations are considerably larger than any of the no-nitrogen observations. Most of the no-nitrogen observations appear to be below the center of the data. The appearance of the data set would seem to indicate that nitrogen is effective.

Example

The two measures of central tendency for the individual samples are:

$$\begin{aligned}\bar{x}\ (\text{no nitrogen}) &= 0.399 \text{ gram},\\ \tilde{x}\ (\text{no nitrogen}) &= \frac{0.38+0.42}{2} = 0.400 \text{ gram},\\ \bar{x}\ (\text{nitrogen}) &= 0.565 \text{ gram},\\ \tilde{x}\ (\text{nitrogen}) &= \frac{0.49+0.52}{2} = 0.505 \text{ gram}.\end{aligned}$$

Clearly there is a difference in concept between the mean and median. The sample mean is the centroid of the data in a sample. In a sense, it is the point at which a fulcrum can be placed to balance a system of "weights" which are the locations of the individual data.



Sample mean as a centroid of the with-nitrogen stem weight.

Other Measures of Locations (Trimmed Means)

A trimmed mean is computed by "trimming away" a certain percent of both the largest and the smallest set of values. For example, the 10% trimmed mean is found by eliminating the largest 10% and smallest 10% and computing the average of the remaining values.

For example, in the case of the stem weight data, we would eliminate the largest and smallest since the sample size is 10 for each sample. So for the without-nitrogen group the 10% trimmed mean is given by:

$$\bar{x}_{\mathrm{tr}(10)} = \frac{0.32 + 0.37 + 0.47 + 0.43 + 0.36 + 0.42 + 0.38 + 0.43}{8} = 0.39750,$$

And for the 10% trimmed mean for the with-nitrogen group we have:

$$\bar{x}_{\mathrm{tr}(10)} = \frac{0.43 + 0.47 + 0.49 + 0.52 + 0.75 + 0.79 + 0.62 + 0.46}{8} = 0.56625.$$

Other Measures of Locations (Trimmed Means)

- Note that in this case, as expected, the trimmed means are close to both the mean and the median for the individual samples.
- The trimmed mean is, of course, more insensitive to outliers than the sample mean but not as insensitive as the median.
- On the other hand, the trimmed mean approach makes use of more information than the sample median.
- Note that the sample median is, indeed, a special case of the trimmed mean in which all of the sample data are eliminated apart from the middle one or two observations.

Tutorial 1

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

Assume that the measurements are a simple random sample.

(a) What is the sample size for the above sample?
(b) Calculate the sample mean for these data.
(c) Calculate the sample median.
(d) Plot the data by way of a dot plot.
(e) Compute the 20% trimmed mean for the above data set.
(f) Is the sample mean for these data more or less descriptive as a center of location than the trimmed mean?

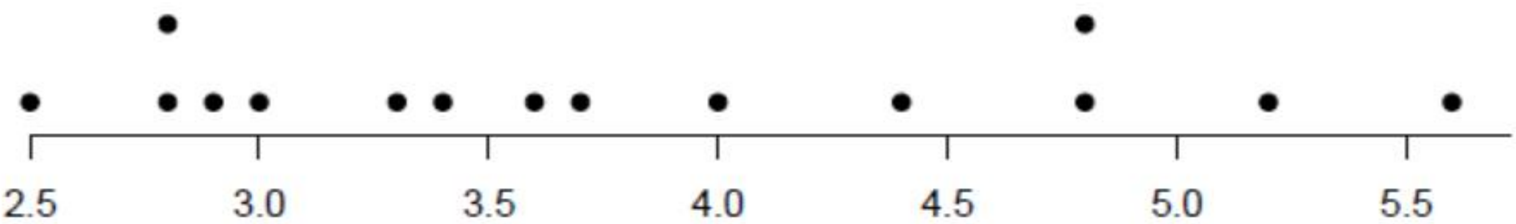
Tutorial 1

(a) 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + \cdots + 4.8) = 3.787.$

(c) Sample median is the 8th value, after the data is sorted from smallest to largest: 3.6.

(d) A dot plot is shown below.



(e) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:

2.9	3.0	3.3	3.4	3.6
3.7	4.0	4.4	4.8	

So. the trimmed mean is

$$\bar{x}_{\text{tr20}} = \frac{1}{9}(2.9 + 3.0 + \cdots + 4.8) = 3.678.$$

Tutorial 2

A certain polymer is used for evacuation systems for aircraft. It is important that the polymer be resistant to the aging process. Twenty specimens of the polymer were used in an experiment. Ten were assigned randomly to be exposed to an accelerated batch aging process that involved exposure to high temperatures for 10 days. Measurements of tensile strength of the specimens were made, and the following data were recorded on tensile strength in psi:

No aging:	227	222	218	217	225
	218	216	229	228	221
Aging:	219	214	215	211	209
	218	203	204	201	205

(a) Do a dot plot of the data.

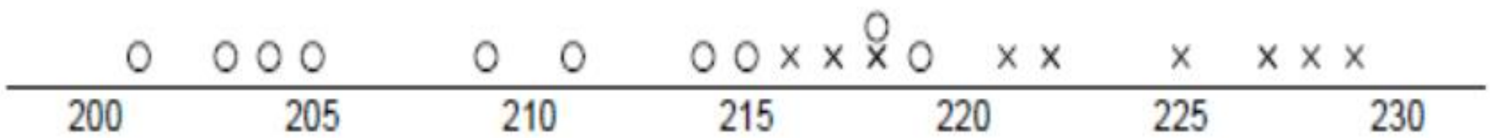
(b) From your plot, does it appear as if the aging process has had an effect on the tensile strength of this polymer? Explain.

(c) Calculate the sample mean tensile strength of the two samples.

(d) Calculate the median for both. Discuss the similarity or lack of similarity between the mean and median of each group.

Tutorial 2

(a) A dot plot is shown below.



In the figure, "×" represents the "No aging" group and "∘" represents the "Aging" group.

(b) Yes; tensile strength is greatly reduced due to the aging process.

(c) $\text{Mean}_{\text{Aging}} = 209.90$, and $\text{Mean}_{\text{No aging}} = 222.10$.

(d) $\text{Median}_{\text{Aging}} = 210.00$, and $\text{Median}_{\text{No aging}} = 221.50$. The means and medians for each group are similar to each other.

Mode

The mode is the value that occurs most frequently in the data set. It denotes a“typical”value from the process.

A hardware store wants to determine what size of circular saws it should stock. From past sales data, a random sample of 30 shows the following sizes(in millimeters):

80	120	100	100	150	120	80	150	120	80
120	100	120	120	150	80	120	100	120	80
100	120	120	150	120	100	120	120	100	100

Note that the mode has the highest frequency. In this case, the mode is 120 (13 is the largest number of occurrences). So, the manager may decide to stock more size 120 saws. A data set can have more than one mode, in which case it is said to be *multimodal.*

Measures of Dispersion

- An important function of quality control and improvement is to analyze and reduce the variability of a process.

- The numerical measures of location we have described give us indications of the central tendency, or middle, of a data set. They do not tell us much about the variability of the observations.

- Consequently, sound analysis requires an understanding of measures of dispersion, which provide information on the variability, or scatter, of the observations around a given value (usually, the mean).

Range

A widely used measure of dispersion in quality control is the range, which is the difference between the largest and smallest values in a data set. Notationally, the range R is defined as

$$R = X_L - X_S$$

where X_L is the largest observation and X_S is the smallest observation.

The following 10 observations of the time to receive baggage after landing are randomly taken in an airport. The data values (in minutes) are as follows:

15, 12, 20, 13, 22, 18, 19, 21, 17, 20

The range $R = 22 - 12 = 10$ minutes. This value gives us an idea of the variability in the observations. Management can now decide whether this spread is acceptable.

Variance

The variance measures the fluctuation of the observations around the mean. The larger the value, the greater the fluctuation. The population variance σ^2 is given by

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Where μ is the population mean and N represents the size of the population. The sample variance s^2 is given by

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

where $\overline{X}$ is the sample mean and n is the number of observations in the sample. In most applications, the sample variance is calculated rather than the population variance because calculation of the latter is possible only when every value in the population is known.

Variance

A modified version for calculating the sample variance is

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n}{n-1}$$

Standard Deviation

Like the variance, the standard deviation measures the variability of the observations around the mean. It is equal to the positive square root of the variance. A standard deviation has the same units as the observations and is thus easier to interpret. It is probably the most widely used measure of dispersion in quality control.

Population standard deviation is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Standard Deviation

Similarly, the sample standard deviation is

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n}{n-1}}$$

Standard Deviation

A random sample of 10 observations of the output voltage of transformers is taken. The values (in volts, V)are as follows:

9.2, 8.9, 8.7, 9.5, 9.0, 9.3, 9.4, 9.5, 9.0, 9.1

Find the sample mean and sample variance

Standard Deviation

the sample mean $\overline{X}$ is 9.16 V.

Table (next slide) shows the calculations.

From the table, $\sum (X_i - \overline{X})^2 = 0.644.$

The sample variance is given by

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} = \frac{0.644}{9} = 0.0716\,\mathrm{V}^2$$

Standard Deviation

Calculation of Sample Variance

X_i	X_i^2	Deviation from Mean, $X_i - \overline{X}$	Squared Deviation, $(X_i - \overline{X})^2$
9.2	84.64	0.04	0.0016
8.9	79.21	−0.26	0.0676
8.7	75.69	−0.46	0.2116
9.5	90.25	0.34	0.1156
9.0	81.00	−0.16	0.0256
9.3	86.49	0.14	0.0196
9.4	88.36	0.24	0.0576
9.5	90.25	0.34	0.1156
9.0	81.00	−0.16	0.0256
9.1	81.81	−0.06	0.0036
$\sum X_i = 91.60$	$\sum X_i^2 = 839.70$	$\sum (X_i - \overline{X}) = 0$	$\sum (X_i - \overline{X})^2 = 0.644$

Standard Deviation

The sample standard deviation is

$$s = \sqrt{0.0716} = 0.2675\ \text{V}$$

Next the calculations are shown in Table (previous slide) ,the sample variance is given by

$$s^2 = \frac{\sum X_i^2 - (\sum X_i)^2/n}{n-1}$$

$$= \frac{839.70 - (91.60)^2/10}{9}$$

$$= 0.0716\ \text{V}^2$$

The sample standard deviation is 0.2675V,as before.

Example

An engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH = 7.0). A sample of size 10 is taken, with results given by:

$$7.07 \quad 7.00 \quad 7.10 \quad 6.97 \quad 7.00 \quad 7.03 \quad 7.01 \quad 7.01 \quad 6.98 \quad 7.08.$$

Estimate the following:

Sample mean
Sample range
Sample variance
Sample Standard Deviation

Example

The sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{7.07 + 7.00 + 7.10 + \cdots + 7.08}{10} = 7.0250.$$

The sample variance s^2 is given by

$$s^2 = \frac{1}{9}[(7.07 - 7.025)^2 + (7.00 - 7.025)^2 + (7.10 - 7.025)^2 + \cdots + (7.08 - 7.025)^2] = 0.001939.$$

As a result, the sample standard deviation is given by

$$s = \sqrt{0.001939} = 0.044.$$

So the sample standard deviation is 0.0440 with $n - 1 = 9$ degrees of freedom.

Tutorial

The tensile strength of silicone rubber is thought to be a function of curing temperature. A study was carried out in which samples of 12 specimens of the rubber were prepared using curing temperatures of 20 Degree C and 45 degree C. The data below show the tensile strength values in megapascals.

20°C:	2.07	2.14	2.22	2.03	2.21	2.03
	2.05	2.18	2.09	2.14	2.11	2.02
45°C:	2.52	2.15	2.49	2.03	2.37	2.05
	1.99	2.42	2.08	2.42	2.29	2.01

(a) Show a dot plot of the data with both low and high temperature tensile strength values.

(b) Compute sample mean tensile strength for both samples.

(c) Does it appear as if curing temperature has an influence on tensile strength, based on the plot? Comment further.

(d) Does anything else appear to be influenced by an increase in curing temperature? Explain.

(e) Compute the sample standard deviation in tensile strength for the samples separately for the two temperatures. Does it appear as if an increase in temperature influences the variability in tensile strength? Explain.

Tutorial

For the cure temperature at 20°C: $s^2_{20°C} = 0.005$ and $s_{20°C} = 0.071$.
For the cure temperature at 45°C: $s^2_{45°C} = 0.0413$ and $s_{45°C} = 0.2032$.
The variation of the tensile strength is influenced by the increase of cure temperature.

Interquartile Range

Interquartile Range The lower quartile, Q_1, is the value such that one-fourth of the observations fall below it and three-fourths fall above it. The middle quartile is the median—half the observations fall below it and half above it. The third quartile, Q_3, is the value such that three-fourths of the observations fall below it and one-fourth above it.

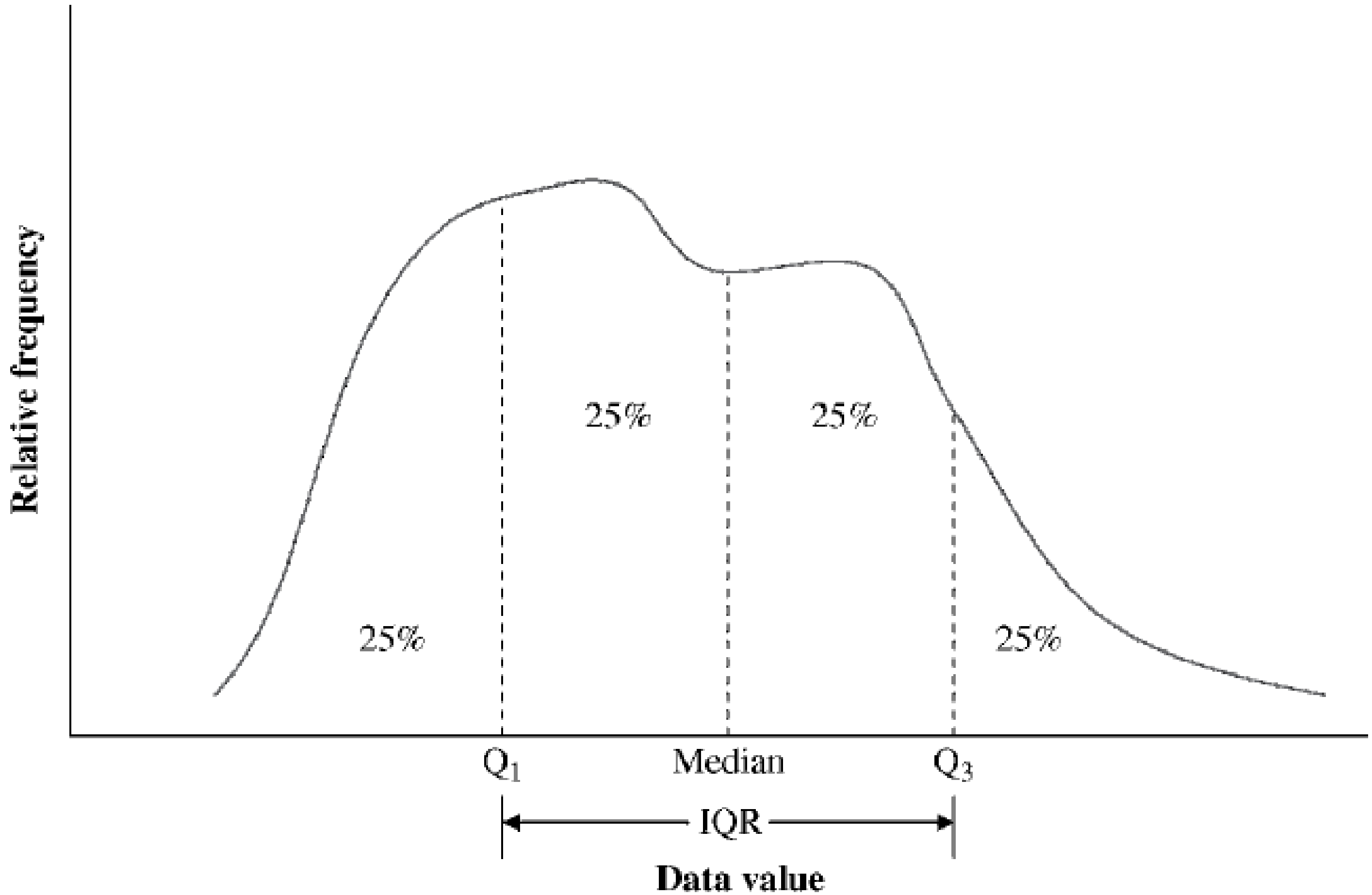
The **interquartile range** (IQR) is the difference between the third quartile and the first quartile. Thus,

$$\text{IQR} = Q_3 - Q_1$$

Note from Figure (next slide) that the IQR contains 50% of the observations. The larger the IQR value,

the greater the spread of the data. To find the IQR, the data are ranked in ascending order; Q_1 is located at rank $0.25(n+1)$, where n is the number of data points in the sample, and Q_3 is located at rank $0.75(n+1)$.

Interquartile Range



Inter quartile range for a distribution.

Interquartile Range

A random sample of 20 observations on the welding time (in minutes) of an operation gives the following values:

2.2	2.5	1.8	2.0	2.1	1.7	1.9	2.6	1.8	2.3
2.0	2.1	2.6	1.9	2.0	1.8	1.7	2.2	2.4	2.2

Interquartile Range

First let's find the locations of Q_1 and Q_3:

$$\text{location of } Q_1 = 0.25(n+1) = (0.25)(21) = 5.25$$
$$\text{location of } Q_3 = 0.75(n+1) = (0.75)(21) = 15.75$$

Now let's rank the data values:

Q_1's location $= 5.25$ Q_3's location $= 15.75$

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Data Value	1.7	1.7	1.8	1.8	1.8	1.9	1.9	2.0	2.0	2.0	2.1	2.1	2.2	2.2	2.2	2.3	2.4	2.5	2.6	2.6

$Q_1 = 1.825$ $Q_3 = 2.275$

Interquartile Range

Thus, linear interpolation yields a Q_1 of 1.825 and a Q_3 of 2.275. The interquartile range is then

$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ &= 2.275 - 1.825 = 0.45 \text{ minute} \end{aligned}$$

Measures of Skewness and Kurtosis

The skewness coefficient describes the asymmetry of the data set about the mean. The skewness coefficient is calculated as follows:

$$\gamma_1 = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{s^3}$$

•In Figure (next slide) , part(a) is a negatively skewed distribution (skewed to the left), part (b) is positively skewed (skewed to the right) , and part (c) is symmetric about the mean.

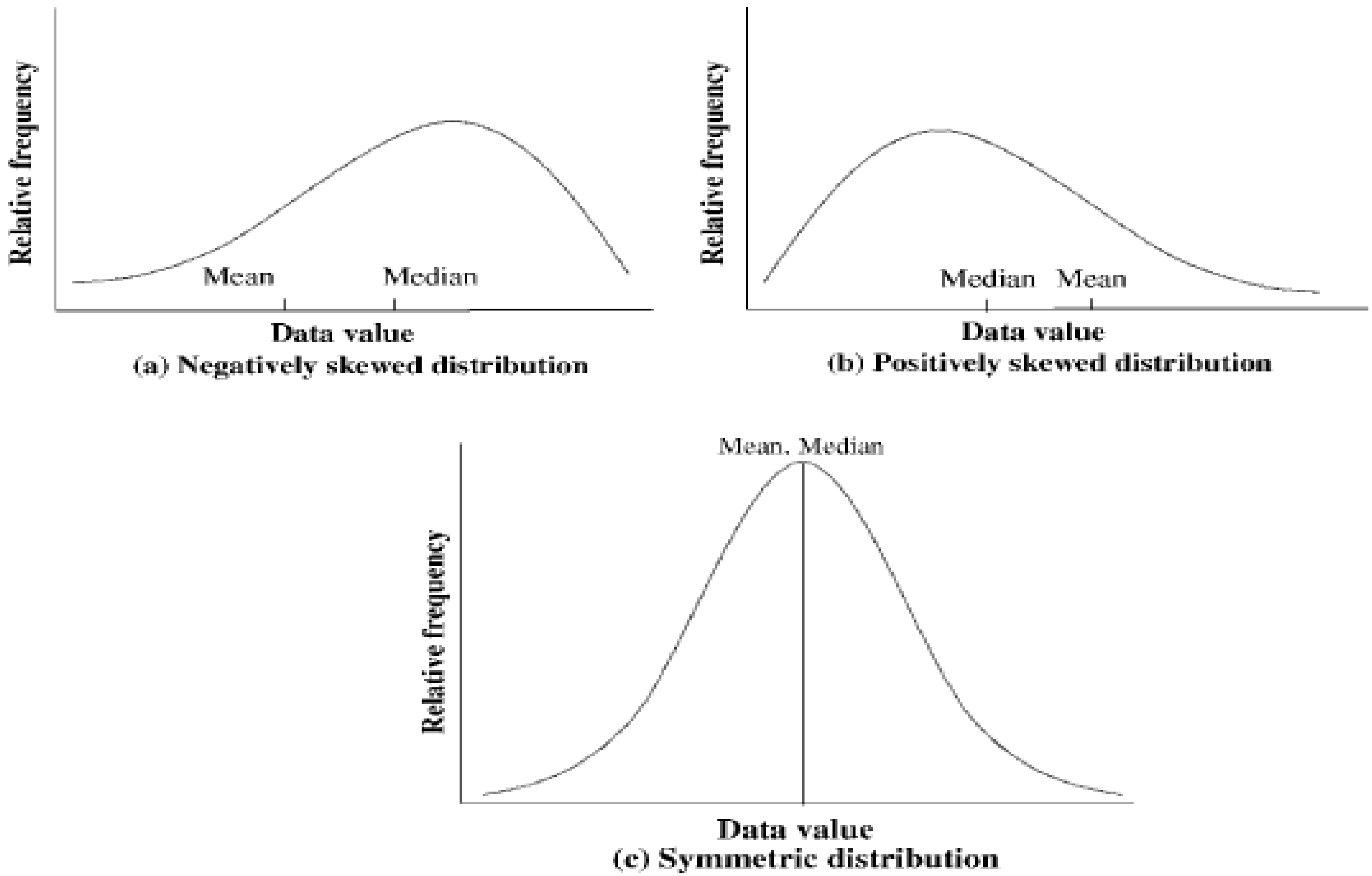
• The skewness coefficient is zero for asymmetric distribution, because [as shown in part(c)] the mean and the median are equal.

Measures of Skewness and Kurtosis

- For a positively skewed distribution, the mean is greater than the median because a few values are large compared to the others; the skewness coefficient will be a positive number.

- If a distribution is negatively skewed, the mean is less than the median because the outliers are very small compared to the other values, and the skewness coefficient will be negative. The skewness coefficient indicates the degree to which a distribution deviates from symmetry.

Measures of Skewness and Kurtosis



Symmetric and skewed distributions.

Measures of Skewness and Kurtosis

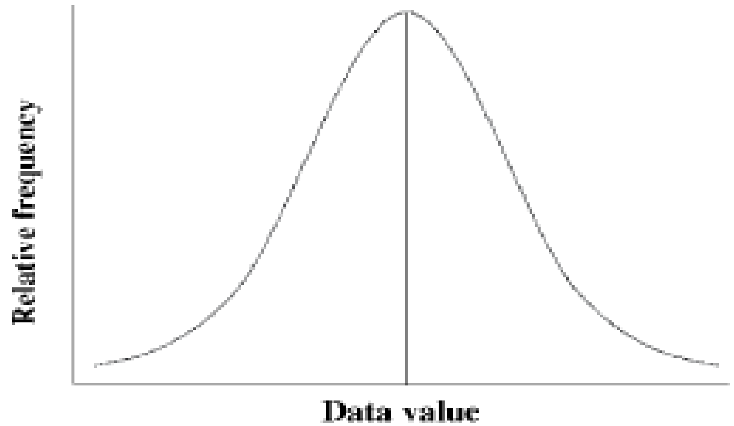
Kurtosis Coefficient: Kurtosis is a measure of the peakedness of the data set. It is also viewed as a measure of the “heaviness” of the tails of a distribution. The kurtosis coefficient is given by

$$\gamma_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \quad \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

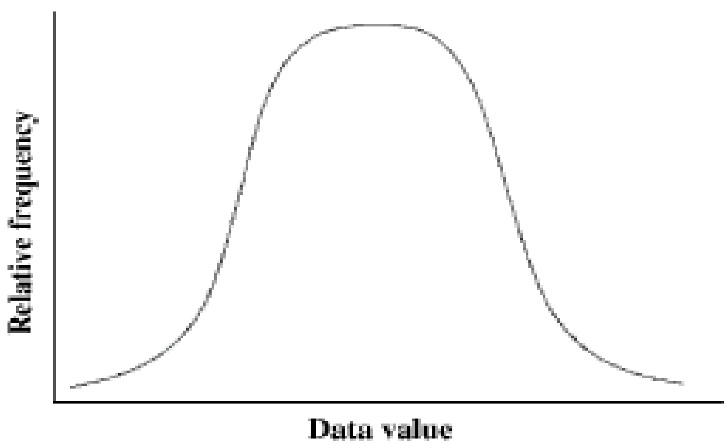
Measures of Skewness and Kurtosis

- The kurtosis coefficient is a relative measure. For normal distributions ,the kurtosis coefficient is zero.
- Figure (next slide) shows a normal distribution (mesokurtic), a distribution that is more peaked than the normal (leptokurtic), and one that is less peaked than the normal (platykurtic).
- For a leptokurtic distribution, the kurtosis coefficient is greater than zero. The more pronounced the peakedness, the larger the value of the kurtosis coefficient. For platykurtic distributions, the kurtosis coefficient is less than zero.

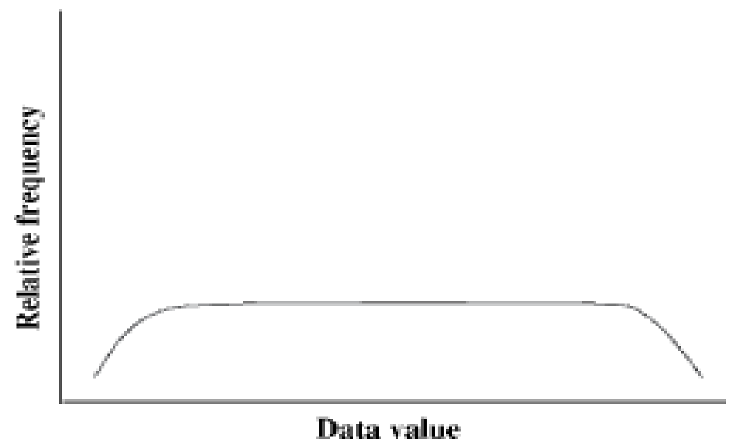
Measures of Skewness and Kurtosis



(a) Mesokurtic distribution (normal)



(b) Leptokurtic distribution



(c) Platykurtic distribution

Distributions with different degrees of peakedness.

Measures of Association

Measures of association indicate how two or more variables are related to each other. For instance, as one variable increases, how does it influence another variable? Small values of the measures of association indicate a nonexistent or weak relationship between the variables, and large values indicate a strong relationship.

- Correlation Coefficient : A correlation coefficient is a measure of the strength of the linear relationship between two variables in bivariate data.
- If two variables are denoted by X and Y, the correlation coefficient of a sample of observations is found from

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

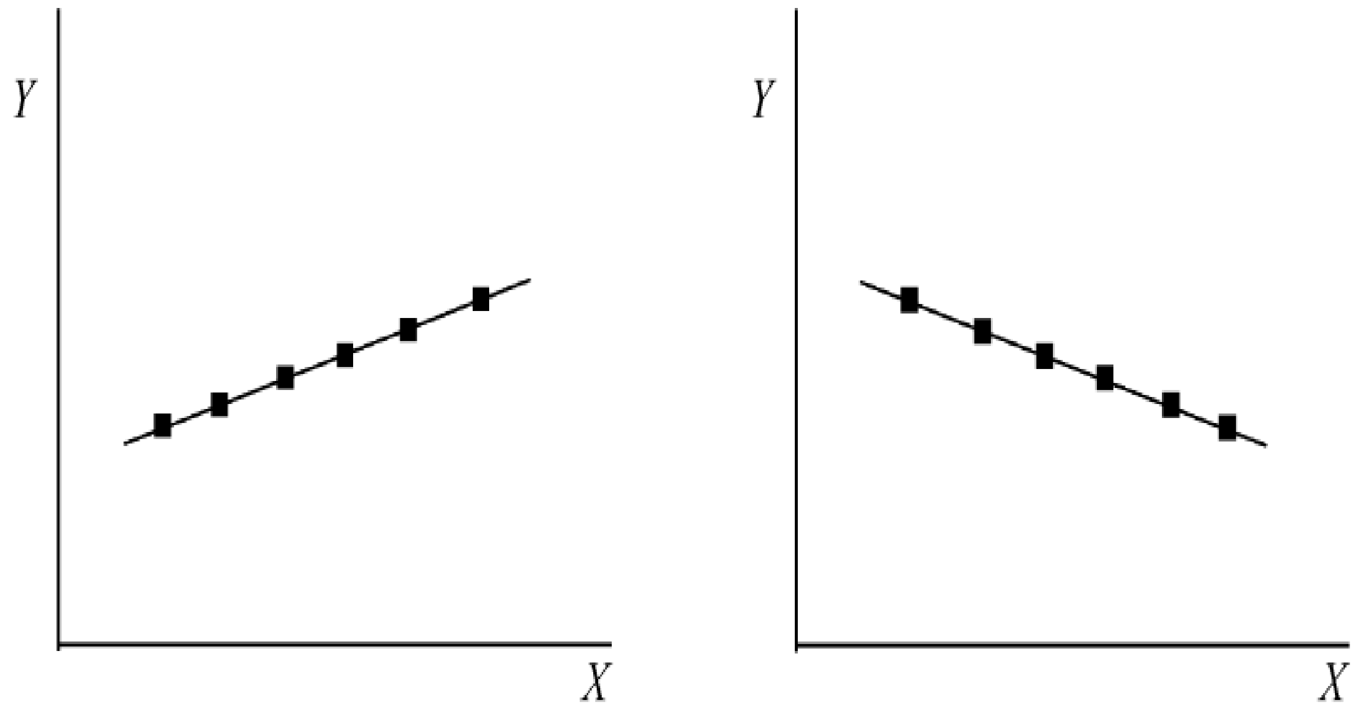
Measures of Association

where X_i and Y_i denote the coordinates of the ith observation, $\overline{X}$ is the sample mean of the X_i-values, $\overline{Y}$ is the sample mean of the Y_i-values, and n is the sample size. An alternative version for calculating the sample correlation coefficient is

$$r = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i)/n}{\sqrt{[\sum X_i^2 - (\sum X_i)^2/n][\sum Y_i^2 - (\sum Y_i)^2/n]}}$$

The sample correlation coefficient r is always between -1 and 1. An r-value of 1 denotes a perfect positive linear relationship between X and Y. This means that, as X increases, Y increases linearly and, as X decreases, Y decreases linearly. Similarly, an r-value of -1 indicates a perfect negative linear relationship between X and Y. If the value of r is zero, the two variables X and Y are uncorrelated, which implies that if X increases, we cannot really say how Y would change. A value of r that is close to zero thus indicates that the relationship between the variables is weak.

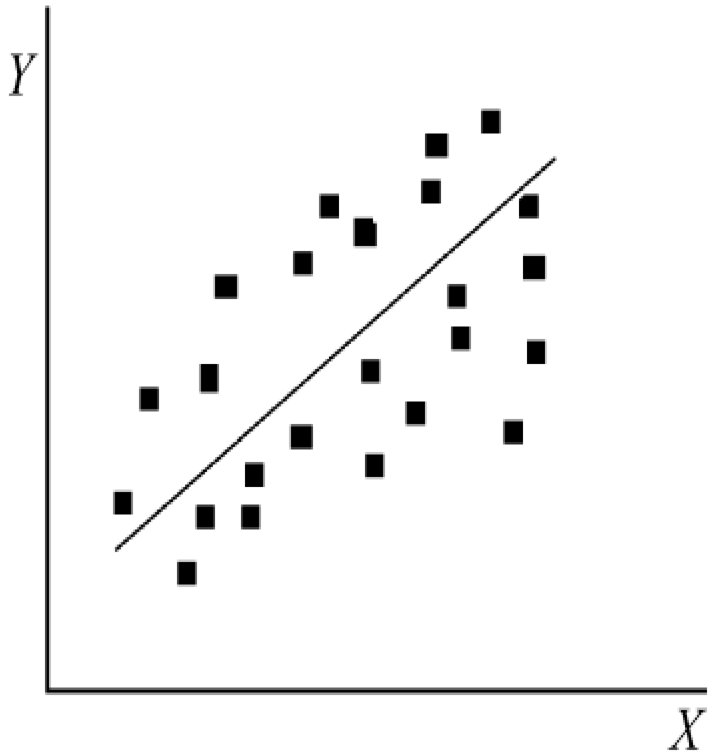
Measures of Association



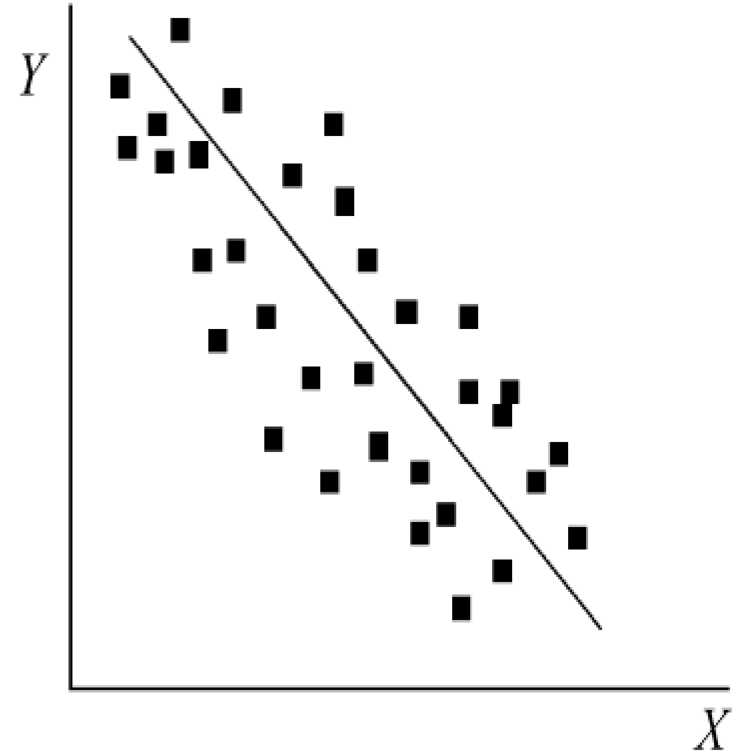
(a) Perfect positive linear relationship, $r = 1$ (b) Perfect negative linear relationship, $r = -1$

Scatter plots indicating different degrees of correlation.

Measures of Association



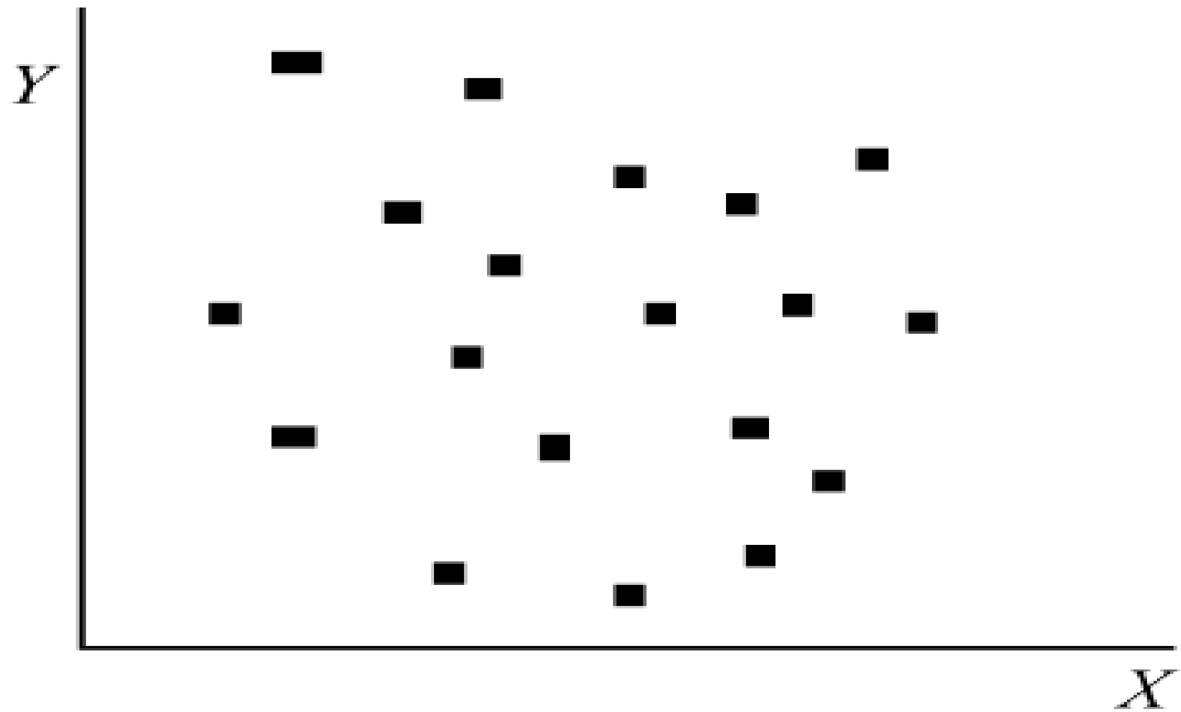
(c) Positive correlation of X and Y



(d) Negative correlation of X and Y

Scatter plots indicating different degrees of correlation.

Measures of Association



(e) No correlation between X and Y

Scatter plots indicating different degrees of correlation.

Random Variable

- Discrete Variables can assume a finite or countably infinite number of values are said to be discrete. These variables are counts of an event.

- For discrete random variables, a probability distribution shows the values that the random variable can assume and their corresponding probabilities. Some examples of discrete random variables are the number of defects in an assembly, the number of customers served over a period of time, and the number of acceptable compressors.

Random Variable

- Continuous random variables can take on an infinite number of values, so the probability distribution is usually expressed as a mathematical function of the random variable.

- This function can be used to find the probability that the random variable will be between certain bounds. Almost all variables for which numerical measurements can be obtained are continuous in nature: for example, the waiting time in a bank, the diameter of a bolt, the tensile strength of a cable, or the specific gravity of a liquid.

Probability Distributions

For a discrete random variable X, which takes on the values x_1, x_2, and so on, a **probability distribution function** $p(x)$ has the following properties:

1. $p(x_i) \geq 0$ for all i, where $p(x_i) = P(X = x_i), i = 1, 2, \ldots$
2. $\sum_{\text{all } i} p(x_i) = 1$

When X is a continuous random variable, the **probability density function** is represented by $f(x)$, which has the following properties:

1. $f(x) \geq 0$ for all x, where $P(a \leq x \leq b) = \int_a^b f(x)\, dx$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$

Note the similarity of these two properties to those for discrete random variables.

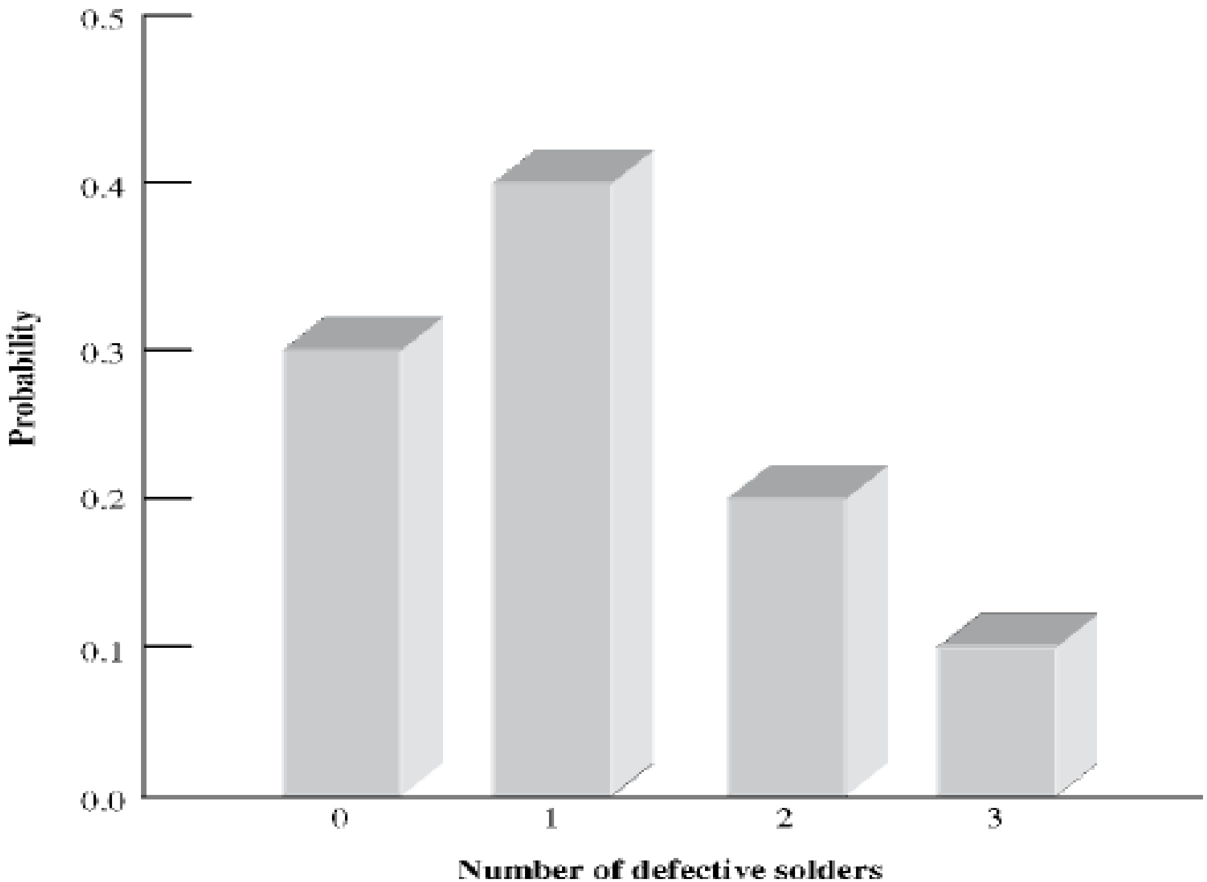
Probability Distributions

Let X denote a random variable that represents the number of defective solders in a printed circuit board. The probability distribution of the discrete random variable X may be given by

x	0	1	2	3
$p(x)$	0.3	0.4	0.2	0.1

This table gives the values taken on by the random variable and their corresponding probabilities. For instance, P(X.1).0.4; that is, there is a 40% chance of finding one defective solder. A graph of the probability distribution of this discrete random variable is shown in Figure(next slide).

Probability Distributions



Probability distribution of a discrete random variable.

Probability Distributions

Consider a continuous random variable X representing the time taken to assemble a part. The variable X is known to be between 0 and 2 minutes, and its probability density function (pdf) *,f(x),is given by*

$$f(x) = \frac{x}{2}, \quad 0 < x \leq 2$$

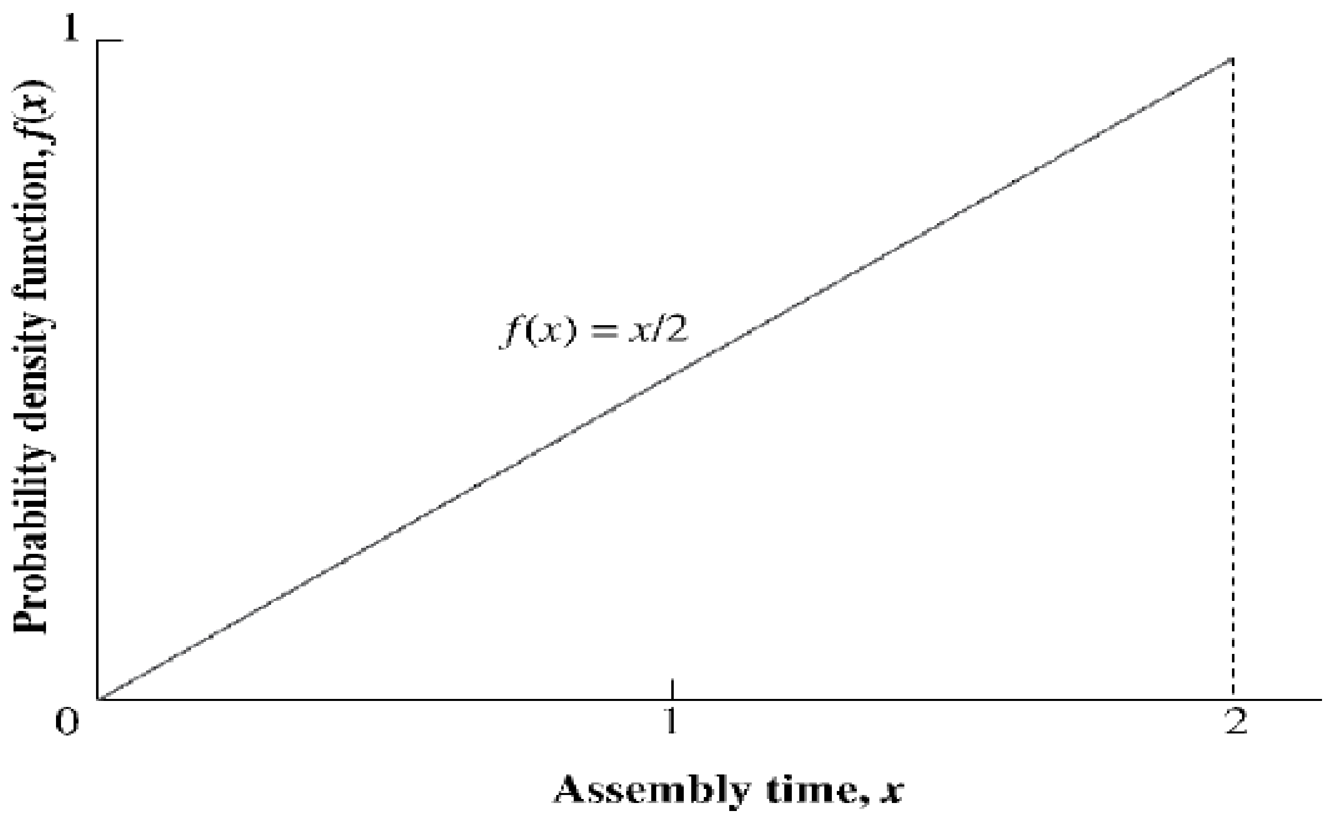
The graph of this probability density function is shown in Figure .Note that

$$\int_0^2 f(x)dx = \int_0^2 \frac{x}{2}dx = 1$$

The probability that X is between 1 and 2 is

$$P(1 \leq X \leq 2) = \int_1^2 \frac{x}{2}dx = \frac{3}{4}$$

Probability Distributions



Probability density function, $0 < x \le 2$.

Probability Distributions

Cumulative Distribution Function

The **cumulative distribution function (cdf)** is usually denoted by $F(x)$ and represents the probability of the random variable X taking on a value less than or equal to x, that is,

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

For a discrete random variable,

$$F(x) = \sum_{\text{all } i} p(x_i) \quad \text{for} \quad x_i \le x$$

If X is a continuous random variable,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Note that $F(x)$ is a nondecreasing function of x such that

$$\lim_{x \to \infty} F(x) = 1 \quad \text{and} \quad \lim_{x \to -\infty} F(x) = 0$$

Example

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and 2. Now,

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}, \quad f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190},$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}.$$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Example

Suppose that the error in the reaction temperature, in ◦C, for a controlled laboratory experiment is a continuous random variable *X* having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.
(b) Find $P(0 < X \leq 1)$.
(c) Find $F(x)$, and use it to evaluate
(d)$P(0 < X \leq 1)$.

Example

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \left.\frac{x^3}{9}\right|_0^1 = \frac{1}{9}.$$

For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} \frac{t^2}{3} dt = \left.\frac{t^3}{9}\right|_{-1}^{x} = \frac{x^3+1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

Tutorial

Classify the following random variables as discrete or continuous:

X: the number of automobile accidents per year in Virginia.

Y : the length of time to play 18 holes of golf.

M: the amount of milk produced yearly by a particular cow.

N: the number of eggs laid each month by a hen.

P: the number of building permits issued each month in a certain city.

Q: the weight of grain produced per acre.

Tutorial

Discrete;

continuous;

continuous;

discrete;

discrete;

continuous.

Tutorial

The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours;

(b) between 50 and 100 hours.

Tutorial

(a) $P(X < 1.2) = \int_0^1 x\, dx + \int_1^{1.2} (2-x)\, dx = \left.\frac{x^2}{2}\right|_0^1 + \left.\left(2x - \frac{x^2}{2}\right)\right|_1^{1.2} = 0.68.$

(b) $P(0.5 < X < 1) = \int_{0.5}^1 x\, dx = \left.\frac{x^2}{2}\right|_{0.5}^1 = 0.375.$

Tutorial

The probability distribution of *X*, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution function of *X*.

Tutorial

The c.d.f. of X is

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ 0.41, & \text{for } 0 \leq x < 1, \\ 0.78, & \text{for } 1 \leq x < 2, \\ 0.94, & \text{for } 2 \leq x < 3, \\ 0.99, & \text{for } 3 \leq x < 4, \\ 1, & \text{for } x \geq 4. \end{cases}$$

Tutorial

The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of *X*;
(b) using the probability density function of *X*.

Tutorial

(a) $P(X < 0.2) = F(0.2) = 1 - e^{-1.6} = 0.7981$;

(b) $f(x) = F'(x) = 8e^{-8x}$. Therefore, $P(X < 0.2) = 8\int_0^{0.2} e^{-8x}\,dx = -e^{-8x}\big|_0^{0.2} = 0.7981$.

Probability Distributions

Expected Value

The **expected value** or mean of a random variable is given by

$$\mu = E(X) = \sum_{\text{all } i} x_i p(x_i) \quad \text{if } X \text{ is discrete}$$

and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{if } X \text{ is continuous}$$

The variance of a random variable X is given by

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Probability Distributions

For the probability distribution of Example on the defective solders, the mean μ or expected value E(X) is given by

$$\mu = E(X) = \sum_{\text{all } i} x_i p(x_i)$$
$$= (0)(0.3) + (1)(0.4) + (2)(0.2) + (3)(0.1) = 1.1$$

The variance of X is

First, $E(X^2)$ is calculated as follows:

$$E(X^2) = \sum_{\text{all } i} x_i^2 p(x_i)$$
$$= (0)^2(0.3) + (1)^2(0.4) + (2)^2(0.2) + (3)^2(0.1) = 2.1$$

So

$$\text{Var}(X) = 2.1 - (1.1)^2 = 0.89$$

Hence, the standard deviation of X is $\sigma = \sqrt{0.89} = 0.943$.

Probability Distributions

For the probability distribution function in Example regarding a part's assembly time, the mean μ,or expected value E(X), is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{2} x \frac{x}{2} dx$$

$$= \frac{2^3}{6} = 1.333 \text{ minutes}$$

Thus, the mean assembly time for this part is 1.333 minutes.

Example

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Let *X* represent the number of good components in the sample. The probability distribution of *X* is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

Example

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

Example

Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $g(X) = 4X + 3$.

$$E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2} (4x^3 + 3x^2)\, dx = 8.$$

Tutorial

The probability distribution of *X*, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is as given below:

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the average number of imperfections per 10 meters of this fabric.

Tutorial

$$\mu = E(X) = (0)(0.41) + (1)(0.37) + (2)(0.16) + (3)(0.05) + (4)(0.01) = 0.88.$$

Tutorial

The density function of coded measurements of the pitch diameter of threads of a fitting is:

$$f(x) = \begin{cases} \frac{4}{\pi(1+x^2)}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of *X*.

Tutorial

$$E(X) = \frac{4}{\pi}\int_0^1 \frac{x}{1+x^2}\,dx = \frac{\ln 4}{\pi}.$$

Discrete Probability Distributions

•Now one-by-one, we will consider the important Probability Mass Functions (pmf) for discrete random variables.

•In each case we will provide the mean and variance of the random variable.

•Generally, in Quality studies, the random variable will be number of Defects/ successes in a given situation.

Binomial Distribution

Extremely useful and used extensively in Quality/Reliability Engineering.

If a Bernoulli trial resulting a success with probability *p* and a failure with probability *q = 1- p,* then the probability distribution of the binomial random variable X, the number of successes in *n* independent trials with replacements , is:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

Binomial Distribution

Binomial distribution is applicable when:

- Identical independent trials are performed
- Probability of success at each trial is constant
- Sampling is done with replacement
- Population size, N is larger than 50

Note: The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N. If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a **hypergeometric distribution** and not binomial. However, when n<<N (typically 0.1N) the binomial distribution is a good approximation, and widely used.

Binomial Distribution

For example, when we toss a coin, p (head) is 0.5.

If we toss a coin 10 times (n=10), then chance that we get head 6 times (r=6) is

$$\frac{n!}{r!(n-r)!}p^{r}(1-p)^{(n-r)}$$

$$= \frac{10!}{6!(10-6)!}0.5^{6}0.5^{4}$$

$$= 0.205$$

Complete tables are available in the set of statistical tables. The table for n=10 is reproduced on the next slide for reference.

Binomial Distribution

Binomial Distribution
Probability of r or fewer occurrence in n trials

		p (probability of occurrence on each trial)											
n	r	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	0.0003	0.0001
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107	0.0045	0.0017
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547	0.0274	0.0123
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719	0.1020	0.0548
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770	0.2616	0.1662
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230	0.4956	0.3669
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281	0.7340	0.6177
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453	0.9004	0.8327
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893	0.9767	0.9536
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9975	0.9940

Binomial Distribution

•From the table, probability of getting 6 or fewer heads is 0.8281. This is the probability of getting 6 or 5 or 4 or 3 or 2 or 1 or 0 heads.

•We want to find out probability of exactly 6 heads. We should therefore subtract probability of getting 5 or fewer heads. This is 0.6230. Thus probability of getting 6 heads is 0.8281-0.6231=0.205. This matches with the value calculated by the formula.

Binomial Distribution

Application Example

A transformer coil has historical failure rate of 10% within warranty period of 5000 hours. What is the probability that a batch of 10 transformers will all survive till warranty?

$$\frac{n!}{r!(n-r)!}p^r(1-p)^{(n-r)}$$

n=10, p=0.1, r=0. Thus probability that all will survive can be calculated as:

$$P(0)=\frac{n!}{r!(n-r)!}p^r(1-p)^{(n-r)}$$

$$=\frac{10!}{0!(10-0)!}0.1^0(1-0.1)^{(10-0)}$$

$$=0.9^{10}$$

$$=0.3486$$

Binomial Distribution

Using Table of Binomial distribution

		p (probability of occurrence on each trial)											
n	r	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	0.0003	0.0001
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107	0.0045	0.0017
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547	0.0274	0.0123
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719	0.1020	0.0548
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770	0.2616	0.1662
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230	0.4956	0.3669
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281	0.7340	0.6177
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453	0.9004	0.8327
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893	0.9767	0.9536
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9975	0.9940

Binomial Distribution

Binomial Average

$$\mu = np$$

Binomial Standard Deviation

$$\sigma = \sqrt{np(1-p)}$$

A manufacturing process is estimated to produce 5% nonconforming items. If a random sample of five items is chosen, find the probability of getting two nonconforming items.

Binomial Distribution

Solution Here, $n = 5$, $p = 0.05$ (if success is defined as getting a nonconforming item), and $x = 2$.

$$P(X = 2) = \binom{5}{2}(0.05)^2(0.95)^3 = 0.021$$

The expected number of nonconforming items in the sample is

$$\mu = E(X) = (5)(0.05) = 0.25 \text{ item}$$

while the variance is

$$\sigma^2 = (5)(0.05)(0.95) = 0.2375 \text{ item}^2$$

Binomial Distribution (Tutorial 1)

Problem 1 : A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Binomial Distribution (Tutorial 1)

(a) Denote by X the number of defective devices among the 20. Then X follows a $b(x; 20, 0.03)$ distribution. Hence,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03)$$
$$= 1 - (0.03)^0(1 - 0.03)^{20-0} = 0.4562.$$

(b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore,

$$P(Y = 3) = \binom{10}{3} 0.4562^3(1 - 0.4562)^7 = 0.1602.$$

Binomial Distribution (Tutorial 2)

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that

(a) at least 10 survive,

(b) from 3 to 8 survive,

(c) and (c) exactly 5 survive?

Binomial Distribution (Tutorial 2)

Let X be the number of people who survive.

(a) $$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662$$
$$= 0.0338$$

(b) $$P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4) = \sum_{x=0}^{8} b(x; 15, 0.4) - \sum_{x=0}^{2} b(x; 15, 0.4)$$
$$= 0.9050 - 0.0271 = 0.8779$$

(c) $$P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x; 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4)$$
$$= 0.4032 - 0.2173 = 0.1859$$

Poisson Distribution

Siméon Denis Poisson
(1781-1840)

Poisson Distribution is given by :

$$P(x) = (e^{-\mu} \mu^{x}) / (x\,!)$$

Where:

- P(X) is probability of finding X successes (often defects)
- μ is the process mean (μ=np) and
- X is the number of defects

Applicability of Poisson Distribution:

(1) Poisson distribution can be used as approximation instead of Binomial when sample size is large and probability is small (typically <0.1)

(2) When probability is small but sample size is large so that the product (np) is finite.

(3) When we want to find out probability of given number of events in a specified time interval. For example, what is the probability of getting 1 failure in a month if average number of failures is 3?

Poisson Distribution

Example

Probability of an item failing is 0.001. What is the probability that exactly 3 items will fail in 2000?

N=2000, p=0.001, μ=np=2000x0.001=2.

$$p(3) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-2}2^{3}}{3!} = 0.1804$$

Note: Poisson distribution is a good approximation to Binomial Distribution when sample size is large and probability is small so that the product np is finite. For example there are 10,000,000 vehicles in a city but probability of accident is 0.000001 so that np=10.

Poisson Distribution

Using Tables of Poisson Distribution

p=0.001, μ = np = 2000x0.001 = 2, X=3

Tables are for Cumulative Distribution. Observe that for np=2, cumulative P(3)=0.8571 and P(2)=0.6767. Thus Probability that exactly 3 items will fail=0.8571-0.6767 =0.1804

Institute of Quality & Reliability

1000xProbability of r or fewer occurrences of an event

np \ r	0	1	2	3	4	5	6	7	8	9
1.10	332.9	699.0	900.4	974.3	994.6	999.0	999.9			
1.20	301.2	662.6	879.5	966.2	992.3	998.5	999.7			
1.30	272.5	626.8	857.1	956.9	989.3	997.8	999.6			
1.40	246.6	591.8	833.5	946.3	985.7	996.8	999.4	999.9		
1.50	223.1	557.8	808.8	934.4	981.4	995.5	999.1	999.8		
1.60	201.9	524.9	783.4	921.2	976.3	994.0	998.7	999.7		
1.70	182.7	493.2	757.2	906.8	970.4	992.0	998.1	999.6		
1.80	165.3	462.8	730.6	891.3	963.6	989.6	997.4	999.4	999.9	
1.90	149.6	433.7	703.7	874.7	955.9	986.8	996.6	999.2	999.8	
2.00	135.3	406.0	676.7	857.1	947.3	983.4	995.5	998.9	999.8	

Poisson Distribution

$$\mu = np$$

$$Variance\,\sigma^2 = \mu = np$$

$$SD = \sigma = \sqrt{np}$$

Note that variance of Poisson distribution is same as its mean!

Example: It is estimated that the average number of surface defects in 20m^2 of paper produced by a process is 3. What is the probability of finding no more than two defects in 40m^2 of paper through random selection?

Poisson Distribution

Solution Here, one unit is 40 m^2 of paper. So, λ is 6 because the average number of surface defects per 40 m^2 is 6. The probability is

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= \frac{e^{-6}(6^0)}{0!} + \frac{e^{-6}(6^1)}{1!} + \frac{e^{-6}(6^2)}{2!} = 0.062$$

Poisson Distribution (Tutorial 1)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles.

What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

Poisson Distribution (Tutorial 1)

This is essentially a binomial experiment with $n = 8000$ and $p = 0.001$. Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = (8000)(0.001) = 8.$$

Hence, if X represents the number of bubbles, we have

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$

Poisson Distribution (Tutorial 2)

In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

Poisson Distribution (Tutorial 2)

Let X be a binomial random variable with $n = 400$ and $p = 0.005$. Thus, $np = 2$. Using the Poisson approximation,

(a) $P(X = 1) = e^{-2}2^1 = 0.271$ and

(b) $P(X \leq 3) = \sum_{x=0}^{3} e^{-2}2^x/x! = 0.857$.

Hypergeometric Distribution

Discrete probability distribution that describes the probability of x successes (random draws for which the object drawn has a specified feature) in n, *without* replacement, from a finite population of size N that contains exactly k objects with that feature, wherein each draw is either a success or a failure.

In contrast, the binomial distribution describes the probability of k successes in n draws *with* replacement.

The classical application of the hypergeometric distribution is **sampling without replacement**.

Hypergeometric Distribution

If we are interested in the probability of selecting ***x*** successes from ***k*** items labelled success and ***n-x*** failures from ***N-k*** items labelled as failures when a random sample of size ***n*** is selected from ***N*** items. The following conditions must be satisfied:

- •sampling is done without replacement
- •Sample size (n) is large (n>0.1N)

k of the ***N*** items may be designated as successes and ***N-k*** are classified as failures.

Let ***X*** be the random variable denoting the successes in Hypergeometric experiment,

Then ***X*** will have a probability mass function (with N, n and k parameters) given by:

Hypergeometric Distribution

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \; for \; x = 0, 1, 2, \ldots, n$$

$$with \; mean = \frac{nk}{N} \; and \; \mathrm{var} = \frac{N-n}{N-1}\, n \;\; \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

Hypergeometric Distribution

Example

If we receive 50 pieces from a supplier containing 4 defective pieces, and sample 8 pieces (without replacement), probability of finding 1 defective in this sample of 8 is:

N=50, x =1, k =4, n= 8
N-k = 50-4 =46
N-x= 7
Ans= 0.3987

Hypergeometric Distribution (Tutorial 1)

Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found.

What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot? Find the mean and variance of the random variable.

Hypergeometric Distribution (Tutorial 1)

Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Once again, this plan is not desirable since it detects a bad lot (3 defectives) only about 30% of the time.

Hypergeometric Distribution (Tutorial 2)

From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that

(a) all 4 will fire?

(b) at most 2 will not fire?

Hypergeometric Distribution (Tutorial 2)

(a) Probability that all 4 fire $= h(4; 10, 4, 7) = \frac{1}{6}$.

(b) Probability that at most 2 will not fire $= \sum_{x=0}^{2} h(x; 10, 4, 3) = \frac{29}{30}$.

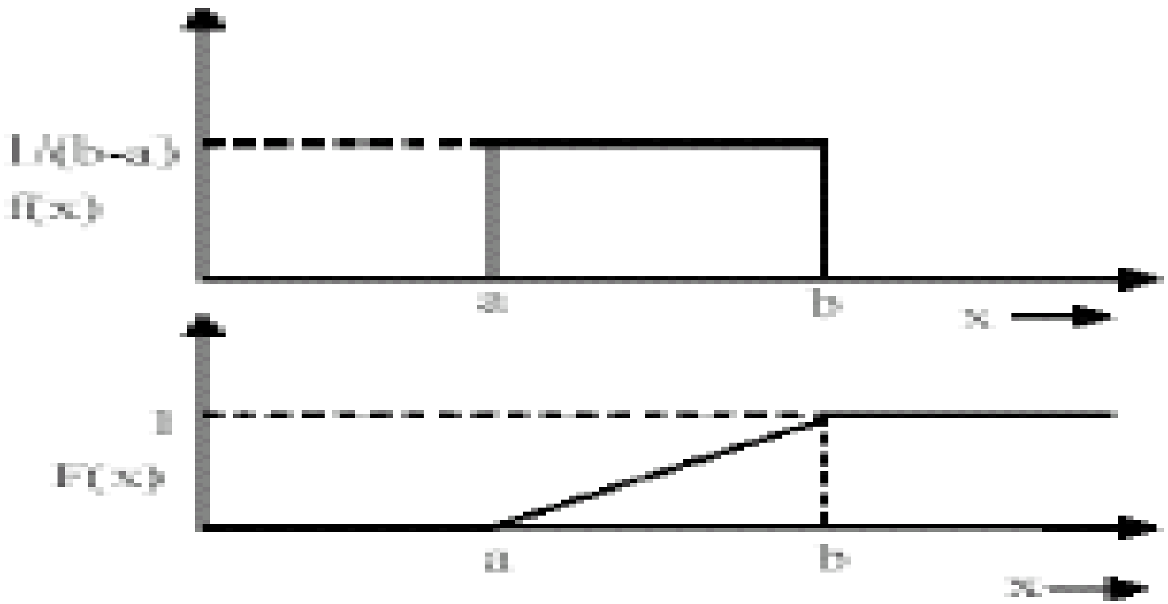
Distributions for continuous Random Variables
Distribution

Now we will consider the important Probability Density Functions for continuous random variables. In each case we will provide the mean and variance of the random variable.

Continuous Uniform Distribution

The probability density function is given by:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & for\ a \le x \le b \\ 0 & elsewhere \end{cases}$$



Continuous Uniform Distribution on Interval [a,b]

Continuous Uniform Distribution

The mean and variance of this distribution would be given by:

$$mean = \frac{a+b}{2}$$

$$Var = \frac{(b-a)^2}{12}$$

Continuous Uniform Distribution

Example: Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval [0, 4].

(a) What is the probability density function?

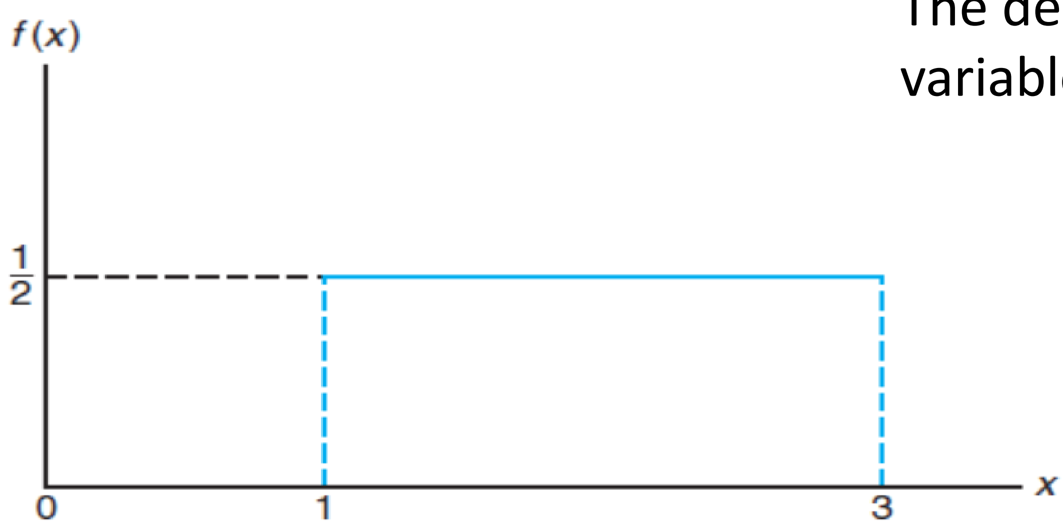
(b) What is the probability that any given conference lasts at least 3 hours?

Continuous Uniform Distribution

(a) The appropriate density function for the uniformly distributed random variable X in this situation is

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \le x \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(b) $P[X \ge 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}$.

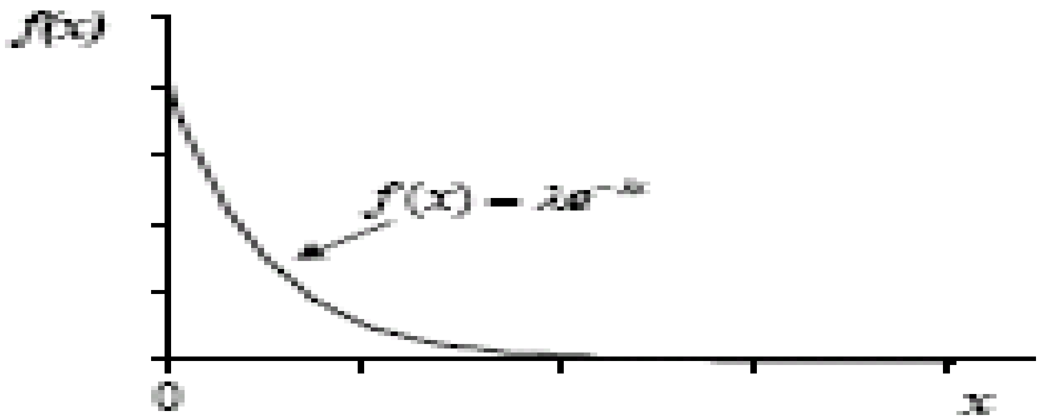


The density function for a random variable on the interval [1, 3].

Exponential Distribution

This is the most widely used single parameter distribution in reliability engineering. Its pdf is given by:

$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & for\, x \geq 0 \\ 0, & for\, x < 0. \end{cases}$$



pdf of Exponential Distribution

Exponential Distribution

•The exponential distribution has a memory-less property, i.e., the probability of occurrence of an event during a specified time interval has the same value, irrespective of the time elapsed before the interval under consideration.

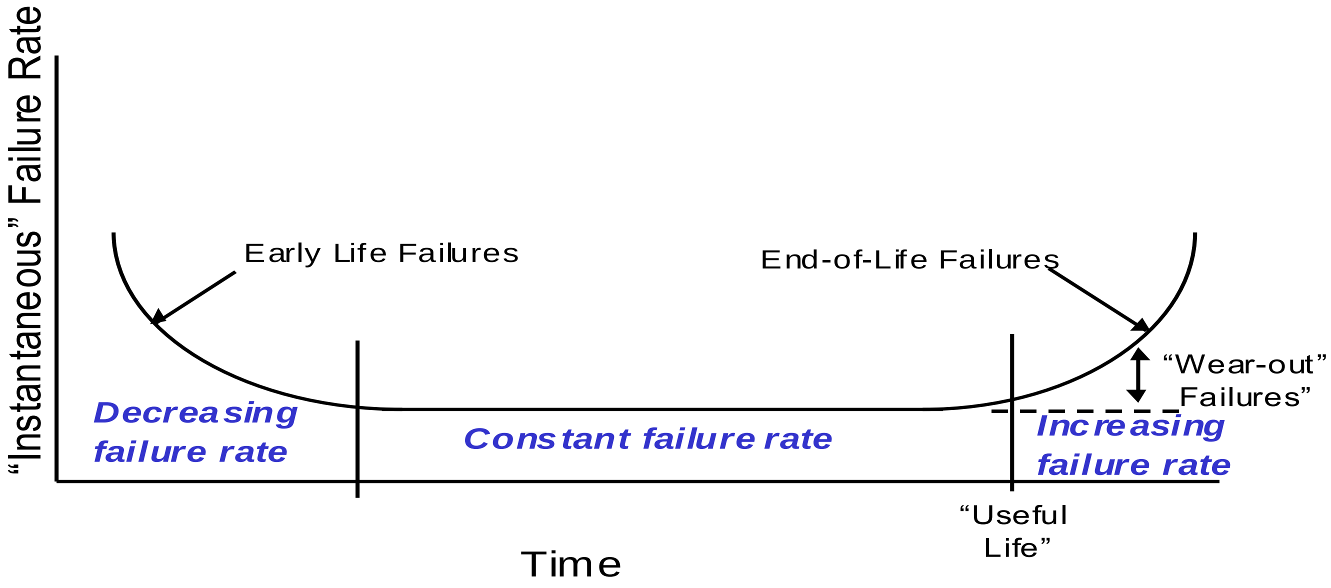
•Memorylessness occurs when failure is caused by external causes only and failure of a device can take place ***not*** due to ***deterioration*** or ***wear but*** due to ***random*** or ***sudden shocks***.

•Therefore it imparts an item a property of being ***as good as new***.

•Electronic and electrical components do have their failure times obeying this distribution.

Exponential Distribution

Exponential distribution is applicable when the <u>hazard rate is constant</u>:



As the hazard rate is constant, probability of failure does not depend on what happened earlier. Thus it is sometimes called 'Memoryless'.

Exponential Distribution

$$\lambda(t) = \lambda$$

$$R(t) = exp\left[-\int_0^t \lambda(t')dt'\right] = e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}$$

$$f(t) = \frac{dF(t)}{dt} = -\frac{d\,R(t)}{dt} = \lambda e^{-\lambda t}$$

$$MTTF = E(T) = \int_0^\infty R(t)\,dt = \int_0^\infty e^{-\lambda t}\,dt = \frac{1}{\lambda}$$

$$\sigma^2 = \int_0^\infty (t - MTTF)^2 f(t)dt = \int_0^\infty \left(t - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda t} dt = \frac{1}{\lambda^2}$$

$$S.D = \sigma = \frac{1}{\lambda} = MTTF$$

Exponential Distribution

Important Observations:

MTTF is the reciprocal of failure rate

Variability of failure times increases as reliability (MTTF) increases

A component having a CFR has a slightly better than one-third chance of surviving to its MTTF.

$$R(t = MTTF) = e^{-\frac{\lambda}{\lambda}} = e^{-1} = 0.368$$

Exponential Distribution

Important Observations:

The design life of a component having an exponentially distributed failure times may be obtained by

$$R(t_R) = e^{-\lambda t_R} = R$$

$$-\lambda t_R = \ln R$$

$$t_R = -\frac{1}{\lambda} \ln R$$

When R=0.5 , the median of the distribution is obtained from

$$t_{med} = -\frac{1}{\lambda} \ln 0.5 = \frac{0.69315}{\lambda} = 0.69315\, MTTF$$

Exponential Distribution

Memorylessness

- Lack of memory not shared by other failure distributions.

- Time to failure of a component is not dependent on how long the component has been operating. There is no aging or wear out effect.

- The probability that the component will operate for the next 1000 hours is the same regardless of whether the component is brand new, has been operating for several hundred hours, or has been operating for several thousand hours.

- This property is consistent with the completely random and independent nature of the failure process.

Exponential Distribution

Memorylessness

For example , when external, random environmental stresses are the primary causes of failures, the failure or operating history of the component will not be relevant.

A burn-in period T_0 has no subsequent effect on reliability and will not improve the component's reliability.

$$R(t|T_0) = \frac{R(T_0 + t)}{R(T_0)} = \frac{e^{-\lambda(t+T_0)}}{e^{-\lambda T_0}} = e^{-\lambda t} = R(t)$$

Exponential Distribution

Example 1: A component has constant failure rate with mean time to failure of 500 hours. What is its reliability at 375 hours?

Example 2: A component has constant failure rate of 0.0025 per hour. What is its reliability at 400 hours?

Example 3: A component has constant failure rate with mean time to failure of 1000 hours. If it has survived till 100 hours, what is the probability that it will survive till 150 hours?

Exponential Distribution

Solution 1

$$R(t) = e^{-\lambda t} = e^{-\frac{t}{\theta}} = e^{-\frac{375}{500}} = 0.4724$$

Solution 2

$$R(t) = e^{-\lambda t} = e^{-0.0025 \times 400} = 0.3679$$

Note that failure rate of 0.0025 corresponds to MTTF of 400 hours. Exponential reliability at t=MTTF is 0.3679. We can also say that probability of failure at MTTF is (1-0.3679) =0.6321 or 63.2% parts will fail up to MTTF. This is a property of exponential distribution.

Exponential Distribution

Solution 3

This problem requires estimation of reliability at 100 hours and at 150 hours. Probability of survival till 150 hours given that it has survived till 100 hours can be calculated as conditional probability as R(150)/R(100).

$$R(100) = e^{-\frac{t}{\theta}} = e^{-\frac{100}{1000}} = 0.9048$$

$$R(150) = e^{-\frac{t}{\theta}} = e^{-\frac{150}{1000}} = 0.8607$$

Thus reliability at 150 hrs given that it has survived till 100 hours =R(250)/R(100)=0.7788/0.9048=0.8607

Exponential Distribution

Example 4.

It is known that a battery for a video game has an average life of 500hours(h). The failures of batteries are known to be random and independent and may be described by an exponential distribution.

Exponential Distribution

(a) Find the probability that a battery will last at least 600 hours.

Solution Since the average life, or mean life, of a battery is given to be 500 hours, the failure rate is $\lambda = 1/500$.

If the life of a battery is denoted by X, we wish to find $P(X > 600)$:

$$P(X > 600) = 1 - P(X \leq 600) = 1 - [1 - e^{-(1/500)(600)}] = e^{-1.2} = 0.301$$

(b) Find the probability of a battery failing within 200 hours.

Solution

$$P(X \leq 200) = 1 - e^{-(1/500)(200)} = 1 - e^{-0.4} = 0.330$$

(c) Find the probability of a battery lasting between 300 and 600 hours.

Solution

$$\begin{aligned} P(300 \leq X \leq 600) &= F(600) - F(300) = e^{-(1/500)(300)} - e^{-(1/500)(600)} \\ &= e^{-0.6} - e^{-1.2} = 0.248 \end{aligned}$$

Exponential Distribution Tutorials

Problem 1

A component experiences chance (CFR) failures with an MTTF of 1100 hr. Find the following:

(a) The reliability for a 200-hr mission
(b) The design life for a 0.90 reliability
(c) The median time to failure
(d) The reliability for a 200-hr mission if a second, redundant (and independent) component is added

Problem 2:

A CFR system with $\lambda = 0.0004$ has been operating for 1000 hr. What is the probability that it will fail in the next 100 hr? The next 1000 hr?

Weibull Distribution

May be used to model both increasing and decreasing failure rates.

Hazard rate:

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} \quad \theta > 0, \beta > 0, t \geq 0$$

Reliability:

$$R(t) = e^{-\left(\frac{t}{\theta}\right)^{\beta}}$$

PDF

$$f(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}}$$

CDF

$$F(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}}$$

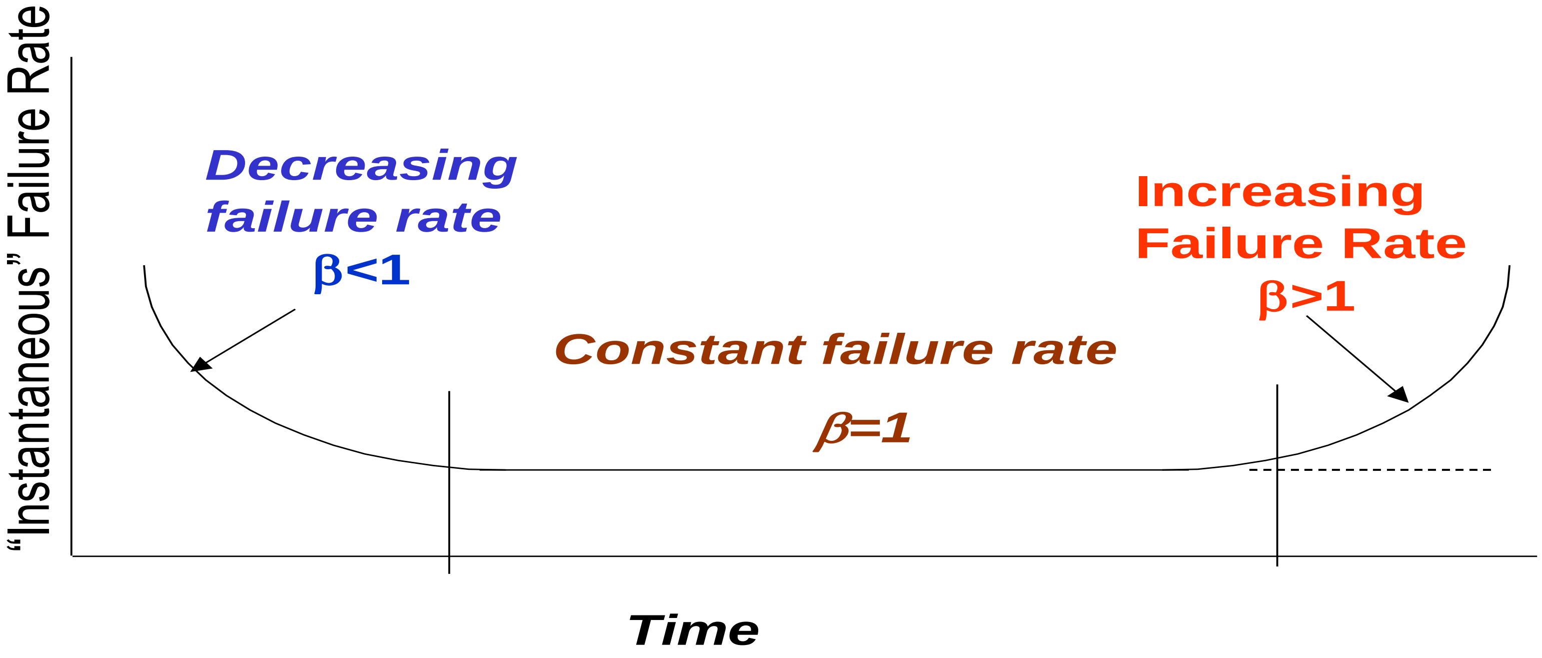
Weibull Distribution

The value of β provides insight into the behavior of the failure process.

Weibull Shape Parameter

Value	Property
$0 < \beta < 1$	Decreasing failure rate (DFR)
$\beta = 1$	Exponential Distribution (CFR)
$1 < \beta < 2$	IFR , Concave
$\beta = 2$	Rayleigh distribution (LFR)
$\beta > 2$	IFR, Convex
$3 \le \beta \le 4$	IFR, Approaches normal distribution, symmetrical

Weibull Distribution



Depending on the shape parameter, Weibull can be used to model life data for all 3 phases

Weibull Distribution

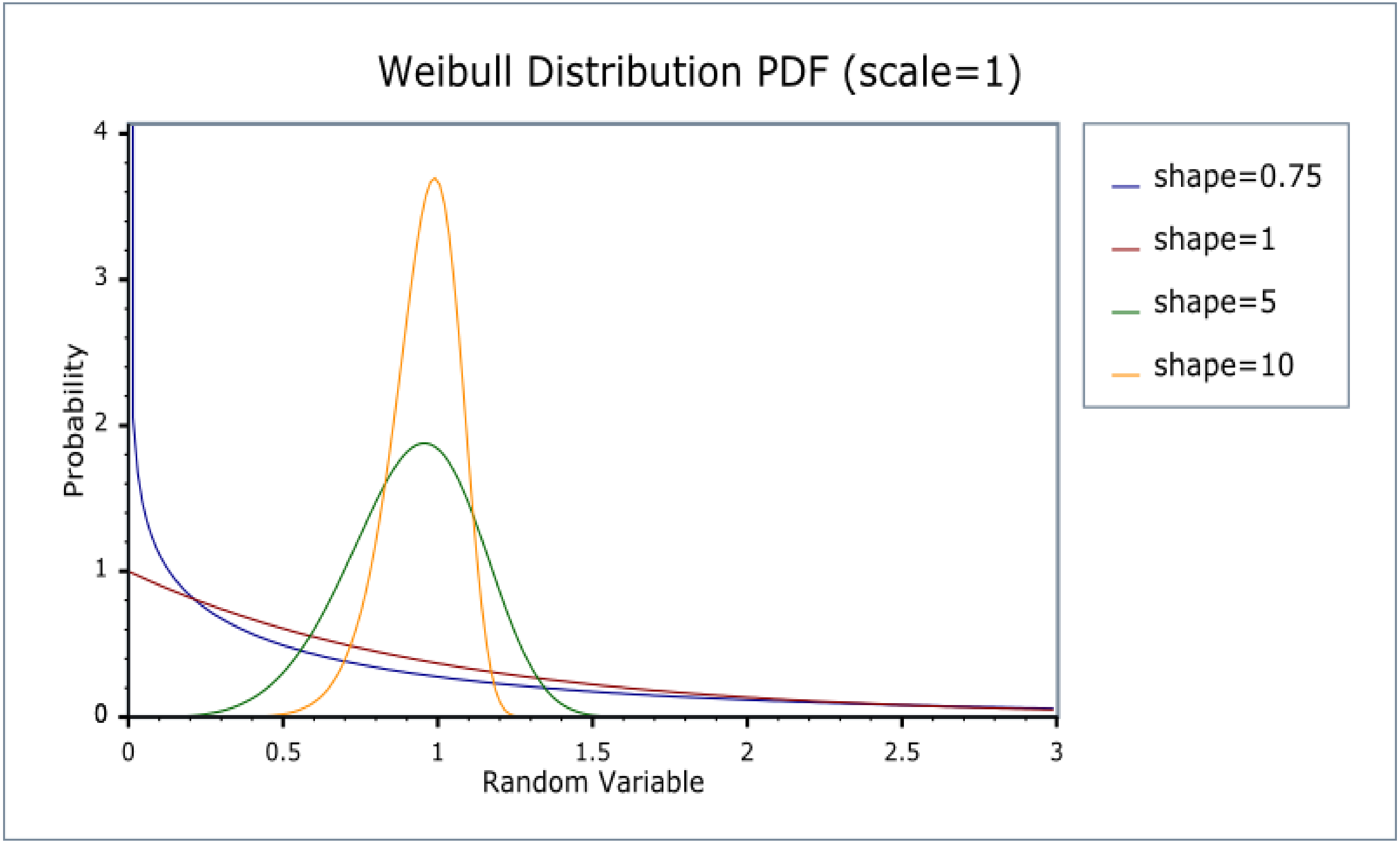
- At t=θ

$$R(\theta) = e^{-\left(\frac{\theta}{\theta}\right)^{\beta}} = exp(-1) = 0.368$$

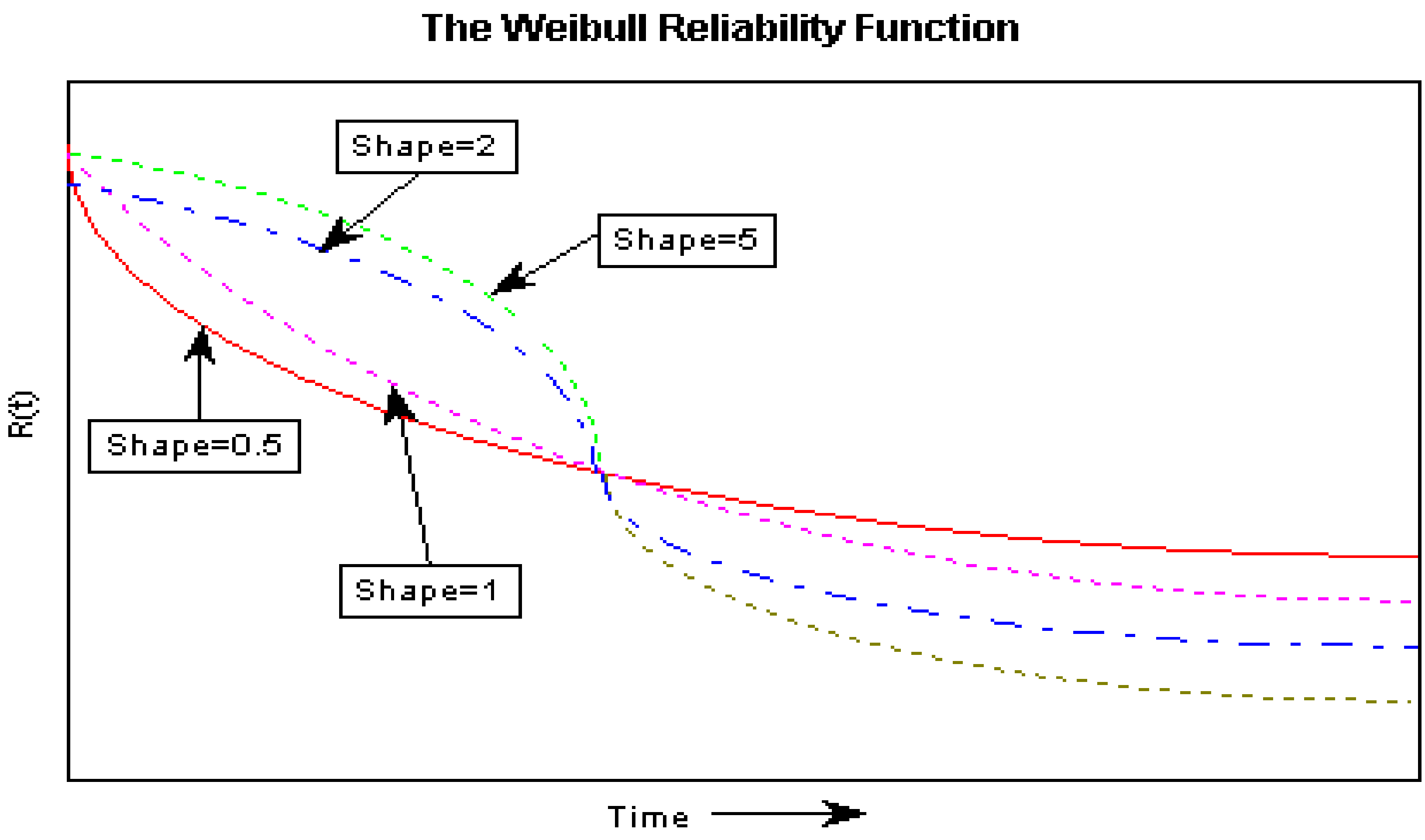
- Therefore , 63.2% of all Weibull failures will occur by time t=θ regardless of the value of the shape parameter.

- Theta (θ) is a scale parameter that influences both the mean and dispersion of the distribution. It is also called the *characteristic life* and it has units identical to those of the failure time , T.

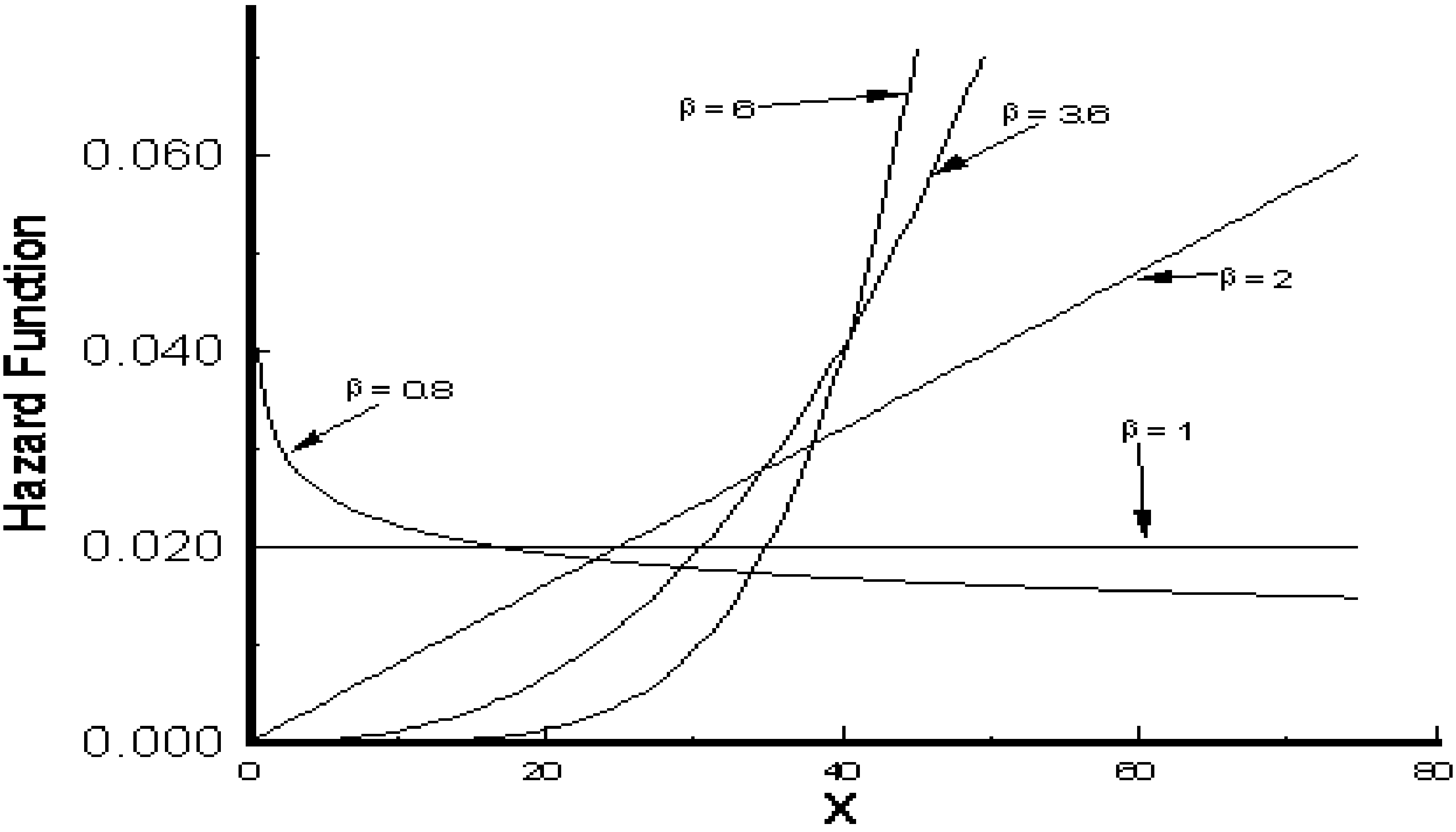
Weibull Distribution



Weibull Distribution



Weibull Distribution



Weibull Distribution

- MTTF: $$MTTF = \theta\Gamma\left(1 + \frac{1}{\beta}\right)$$

- Variance: $$\sigma^2 = \theta^2\left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right)^2\right]\right\}$$

- Where $\Gamma(x)$ is the gamma function

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$$

- Design Life : $R(t) = e^{-(t/\theta)^\beta} = R$ $\qquad t_R = \theta(-lnR)^{\frac{1}{\beta}}$

Weibull Distribution

- When R=0.99, $t_{0.99}$ is referred to as the B1 life, i.e., the time at which 1 percent of the population will have failed .

- Median: $t_{0.50} = t_{med} = \theta(-ln0.5)^{\frac{1}{\beta}} = \theta(0.69315)^{\frac{1}{\beta}}$

Burn-In Screening for Weibull :

Using the conditional reliability:

$$R(t|T_0) = \frac{exp\{-[(t + T_0/\theta)]^{\beta}\}}{exp[-(T_0/\theta)^{\beta}]} = exp\left[-\left(\frac{t+T_0}{\theta}\right)^{\beta} + \left(\frac{T_0}{\theta}\right)^{\beta}\right]$$

Weibull Distribution

- The Three-Parameter Weibull

  Whenever there is a minimum life t_0 such that $T > t_0$ the three parameter Weibull may be appropriate. The distribution assumes that no failures will take place prior to time t_0 .For this distribution,

$$R(t) = exp\left[-\left(\frac{t-t_0}{\theta}\right)^{\beta}\right] t \geq t_0$$

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t-t_0}{\theta}\right)^{\beta-1} t \geq t_0$$

$$MTTF = t_0 + \theta\Gamma\left(1 + \frac{1}{\beta}\right)$$

$$t_R = t_0 + \theta(-lnR)^{\frac{1}{\beta}}$$

Weibull Distribution

Example 1 : A compressor experiences wear out with a linear Hazard rate function

$$\lambda(t) = \frac{2}{1000}\left(\frac{t}{1000}\right)$$

Find:

(a) Design life for a desired reliability of 0.99

(b) MTTF

(c) Variance

(d) S.D

ANSWER

(a) 100.25 hour

(b) 886.23 hour

(c) 214,601.7

(d) 463.25 hour

Weibull Distribution

Example 2 .Given a Weibull failure distribution with a shape parameter of 1/3 and a scale parameter of 16,000. completely characterize the failure process

(a) Reliability function

(b) β

(c) MTTF

(d) Variance

(e) S.D

(f) Characteristic life

(g) Design life if the desired reliability is 0.90

Weibull Distribution

Answer

Reliability function

$$R(t) = exp\left[-\left(\frac{t}{16000}\right)^{\frac{1}{3}}\right]$$

β =1/3

MTTF:

$$MTTF = 16{,}000\,\Gamma\left(1+\frac{1}{1/3}\right) = 16{,}000\times 3! = 96{,}000\ hr.$$

Variance

$$\sigma^2 = (16000)^2\{\Gamma(7) - [\Gamma(4)]^2\} = 175{,}104\times 10^6$$

S.D.

$$\sigma = 418.4\times 10^3\ hr$$

Design Life :

$$t_d = 16{,}000(-\ln 0.90)^3 = 18.71\ hr$$

Weibull Distribution (Tutorial 1)

Example 3 *:* Leakage at front oil seal in a system exhibits Weibull distribution with characteristic life of 1500 hours. The shape parameter is 0.75.

What is its reliability at 500 hours? What % of seals should we expect to leak by 1000 hours?

Weibull Distribution (Tutorial 1)

$$\theta = 1500, \beta = 0.75$$

$$For\ t = 500\ hours,$$

$$R(500) = e^{-\left(\frac{t}{\theta}\right)^{\beta}} = e^{-\left(\frac{500}{1500}\right)^{0.75}} = 0.6449$$

$$For\ t = 1000\ hours, F(1000) = 1 - R(1000) = 1 - e^{-\left(\frac{1000}{1500}\right)^{0.75}} = 0.5218$$

Weibull Distribution (Tutorials)

Problem : The time to failure for a suspension of a motor cycle can be modeled using a Weibull distribution with parameters

$t_0 = 0$ β=1/3 ϴ=200 hours

(a) Find the mean time to failure and its standard deviation.

(b) What is the probability of suspension to operate for at least 800 hours?

Problem : The transmission system used in a motorcycle experiences failures that seem to be well approximated by a two parameter Weibull distribution with

$$\theta = 18,000 \text{ km and } \beta = 2.7$$

(1) What is the 10,000 km reliability? (2) What is the 24,000 km reliability?

Normal Distribution

The Gaussian or Normal distribution is given by:

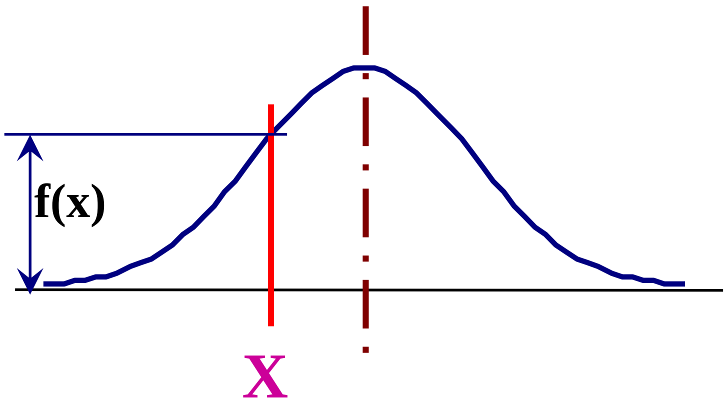
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

where μ is the *mean* and σ is the standard deviation. Mode and Median are coincident with the mean.

Normal Distribution

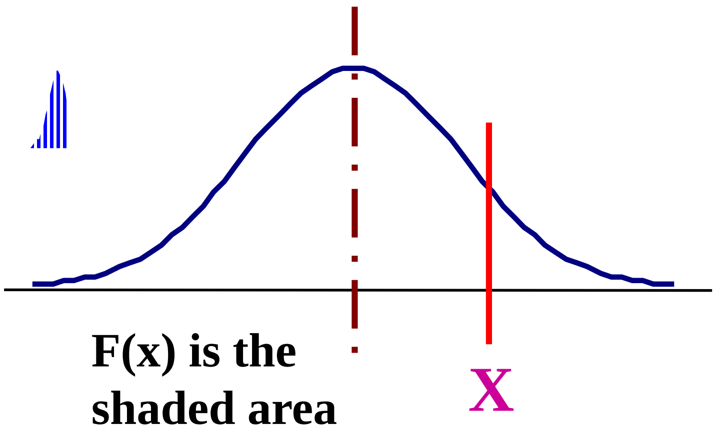
Probability Distribution Function f(x) for normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

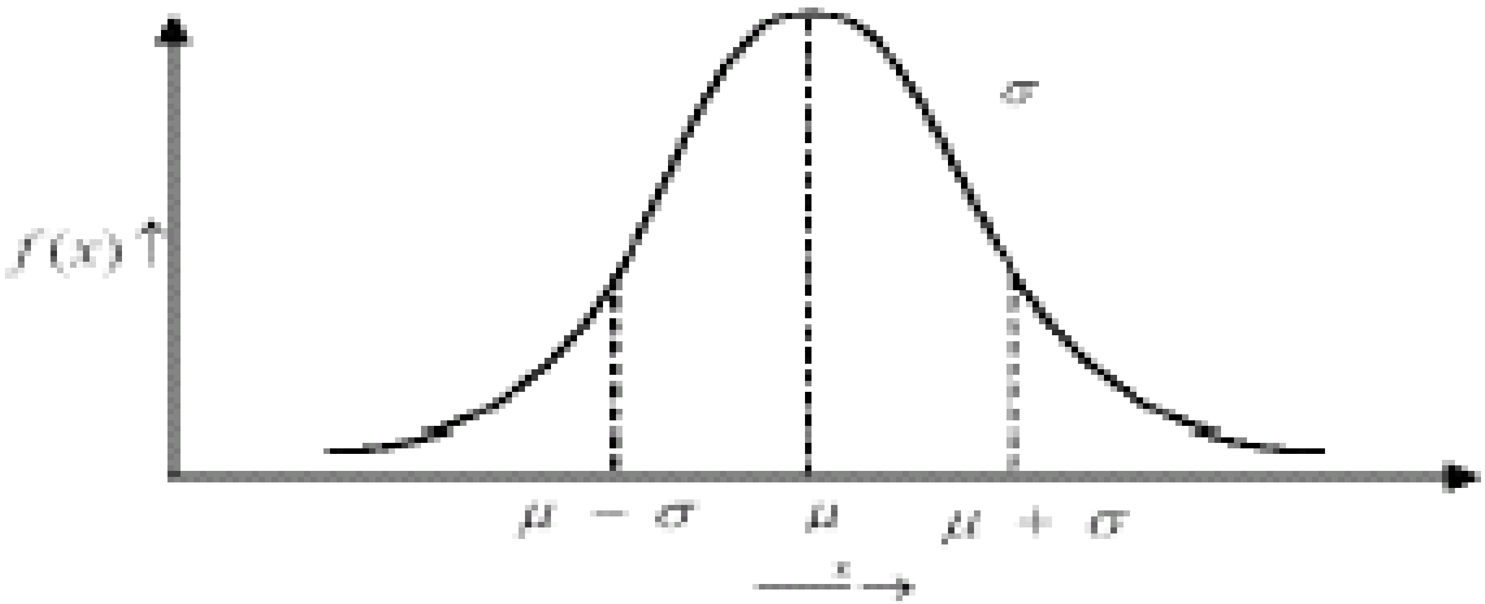


F(x) is Cumulative Distribution Function (CDF) for normal distribution and is given by:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{(2\pi)}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]} dx$$

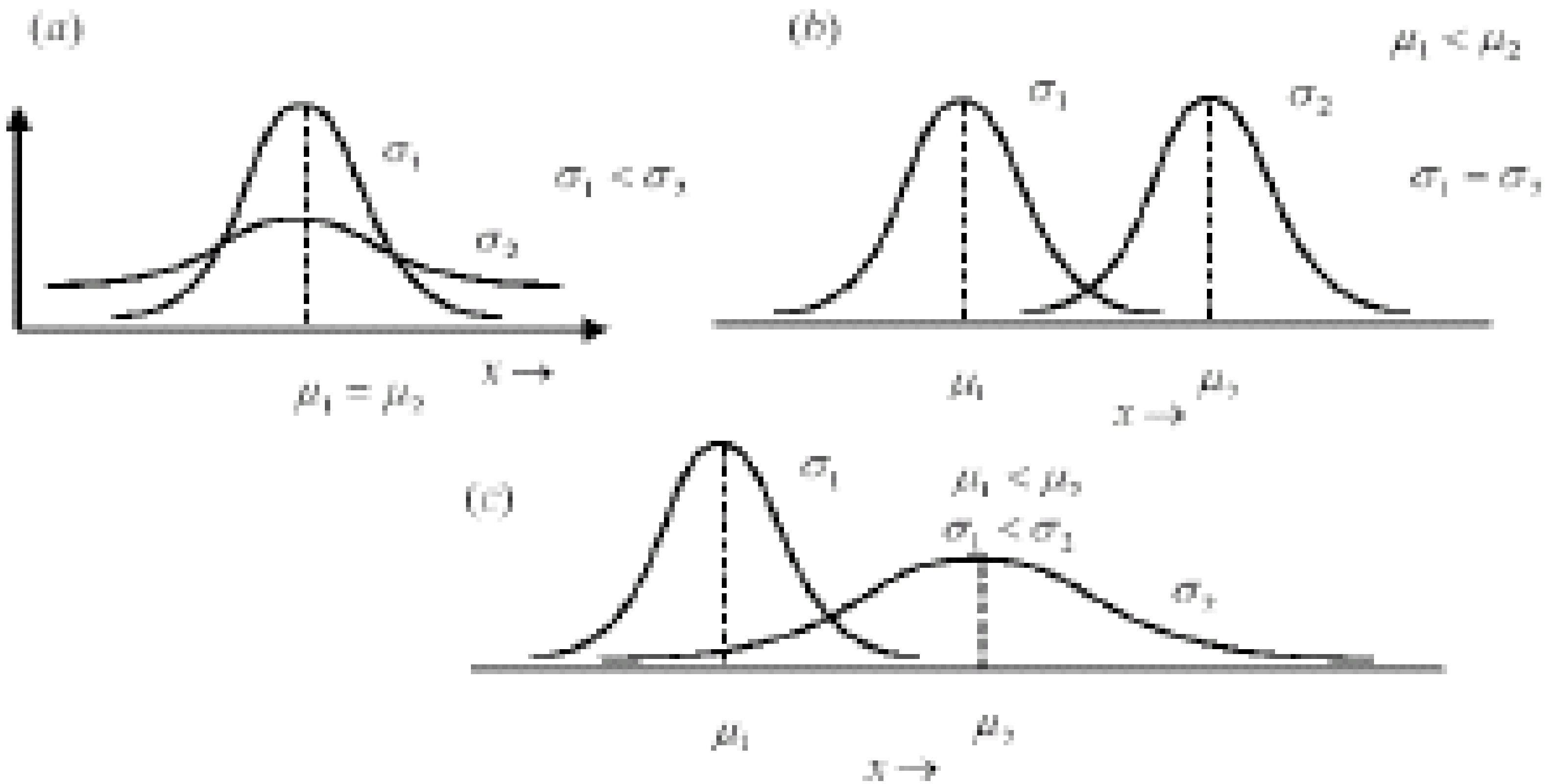


Normal Distribution



Probability Density Function of a Normal Distribution

Normal Distribution



Normal Distributions with Different μ and σ

Normal Distribution

Standard Normal Distribution

The distribution of a normal variate with mean 0 and variance 1 is called a standard normal distribution. The standard normal variable z is obtained through the transformation of random variable X by:

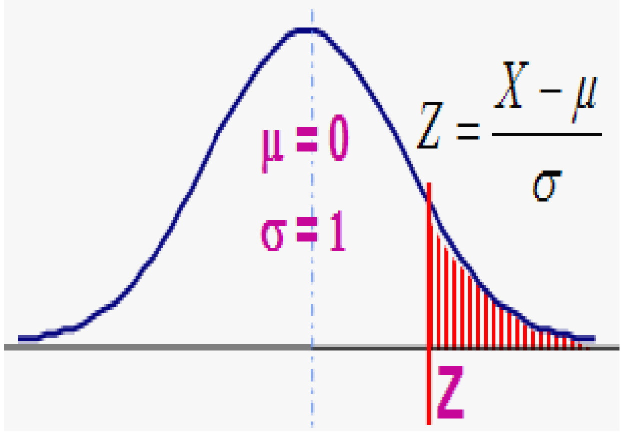
$$z = \frac{x - \mu}{\sigma}$$

And this curve has an area of 0.5 from -∞ up to z = 0 and the other 0.5 area extends from z = 0 to + ∞. The standard normal cdf tables are usually available.

Normal Distribution

Z Score

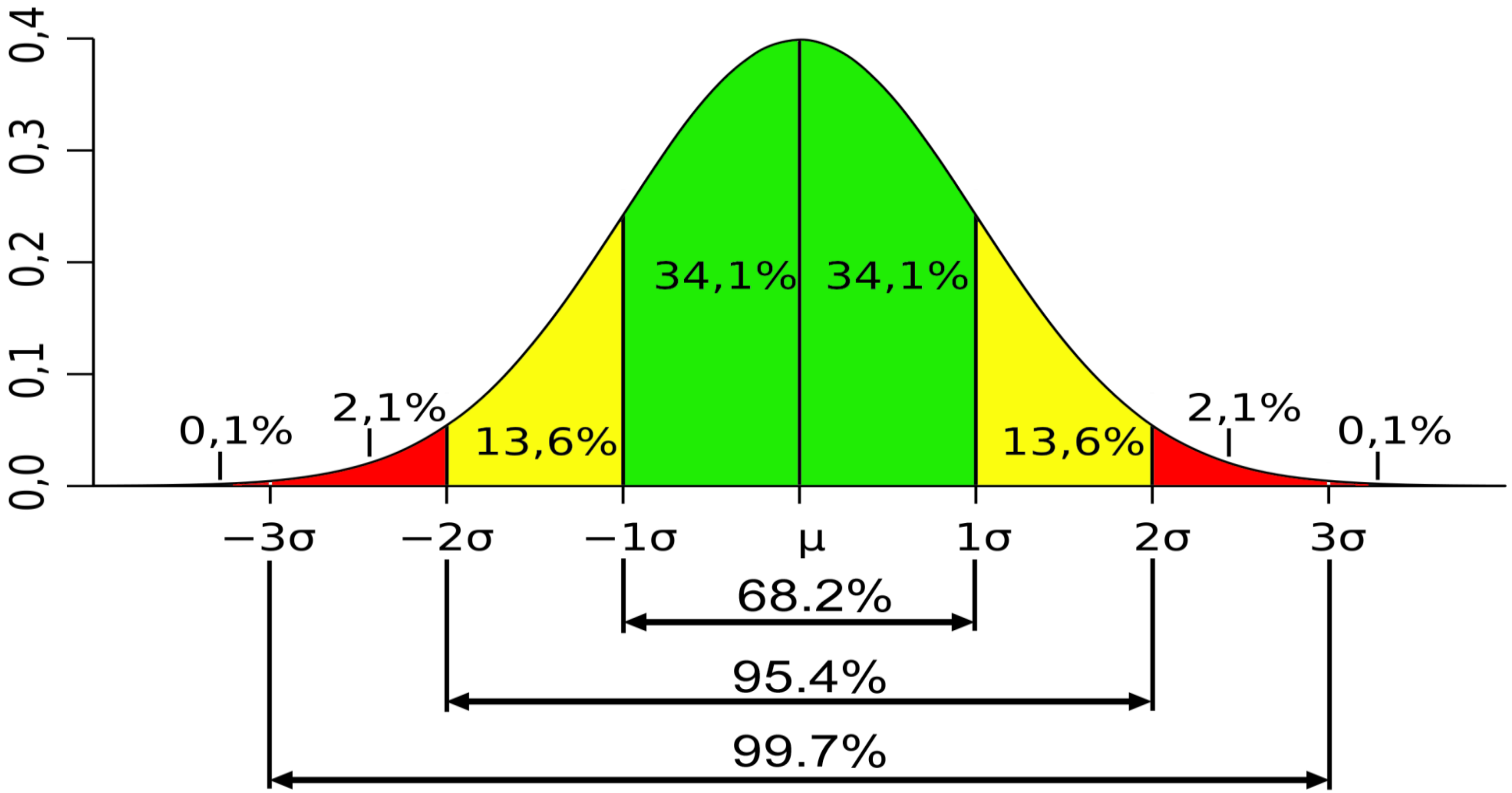
Z Score ‘converts’ the ‘X’ value in to ‘standard’ value. This is a very useful concept that helps us in ‘standardizing’ the X value.



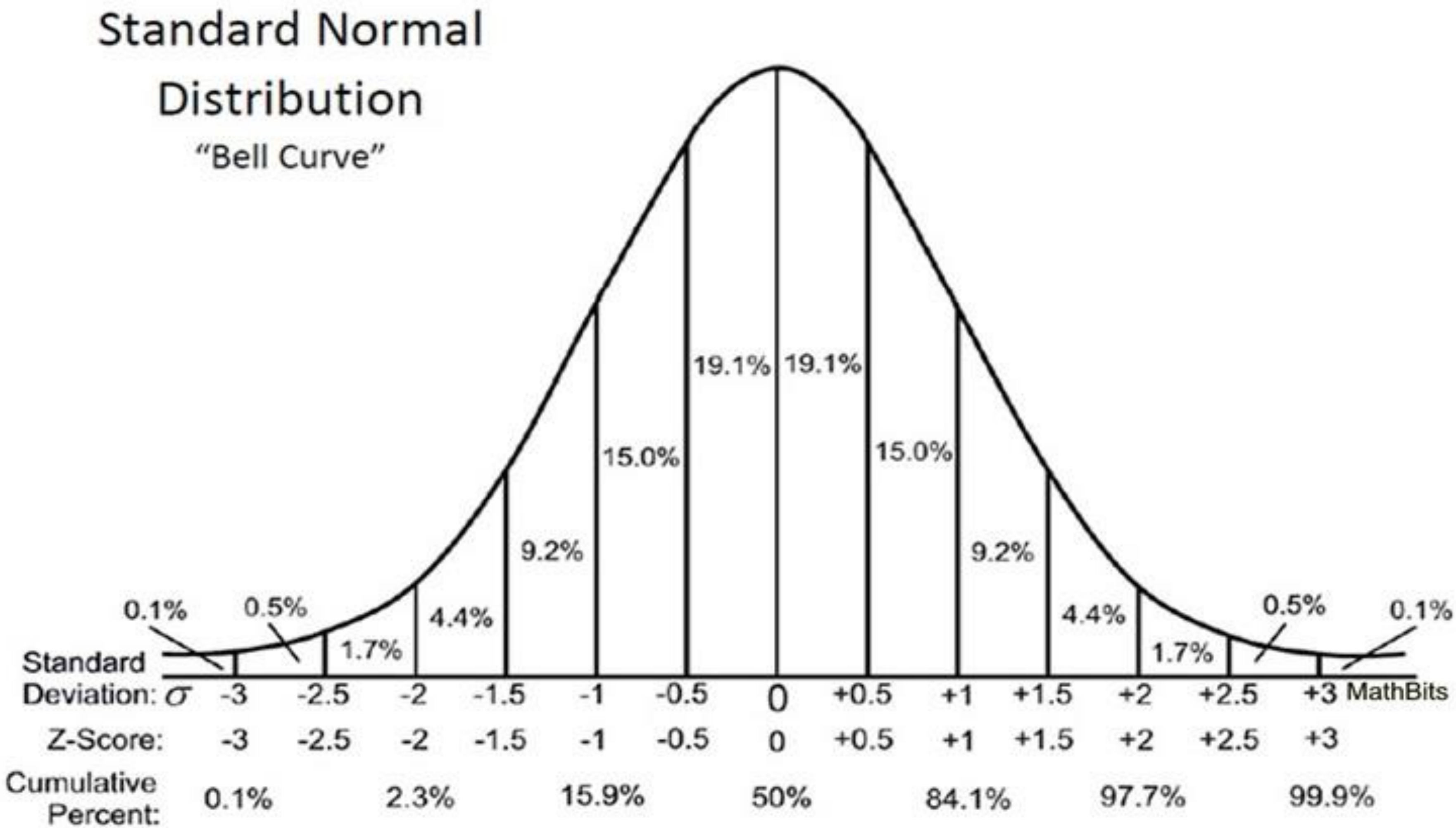
Z score helps us in finding area under the normal curve beyond the X value under consideration. This has important application in estimating defect rates when mean, standard deviation and tolerances are known. Standard Normal Tables are available.

Normal Distribution

Standard Normal Distribution (contd.)



Normal Distribution



Normal Distribution

Standard Normal Distribution (contd.)

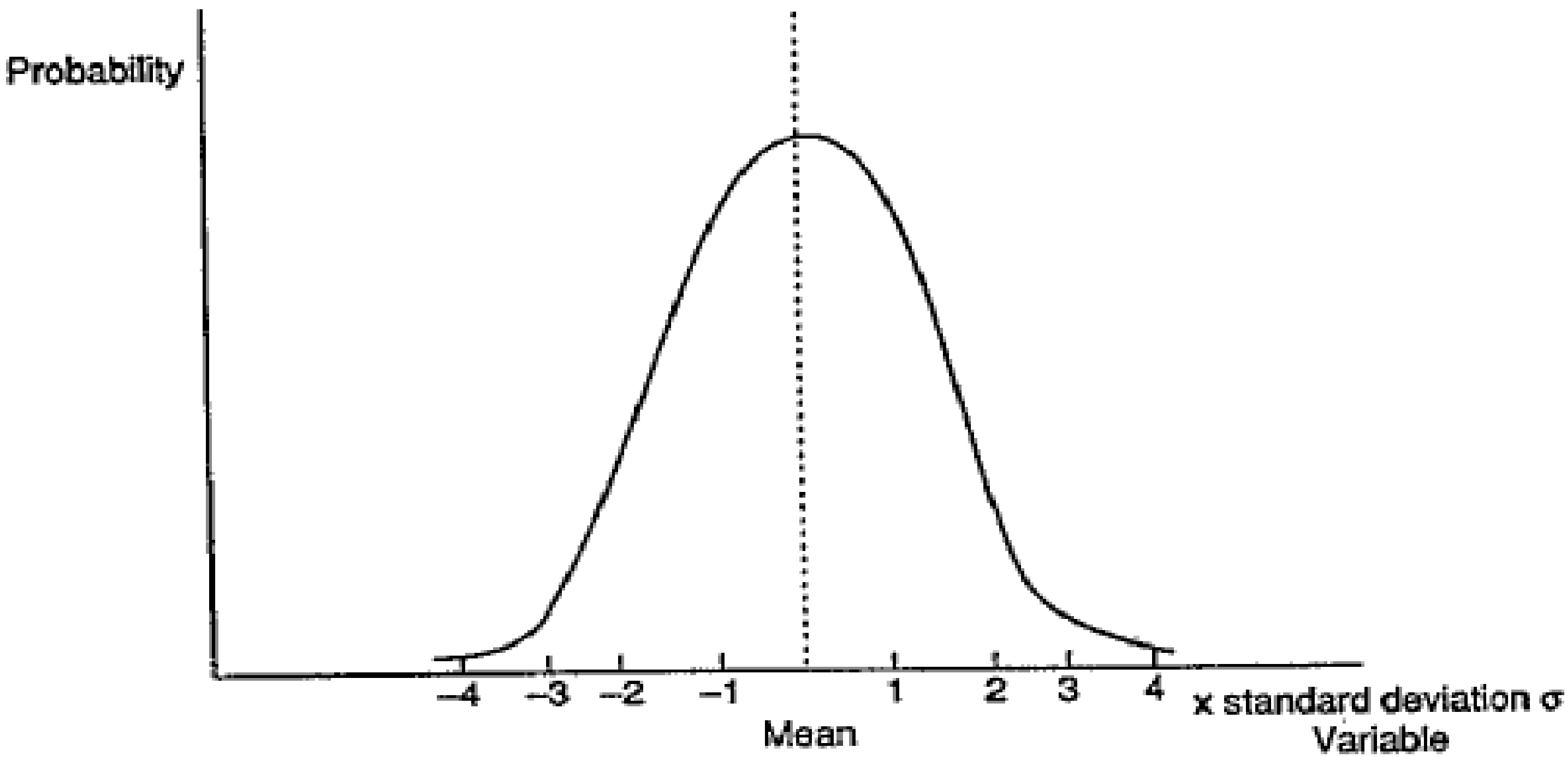
The standard normal distribution has the mean $\mu = 0$ and the variance $\sigma^2 = 1$

The probability density function and the cumulative distribution denoted by $\phi(z)$, and $\Phi(z)$ respectively, and are given by:

$$\phi(z; 0,1) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad for -\infty < z < \infty$$

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt, \quad , \quad for -\infty < z < \infty$$

Normal Distribution



Normal Distribution

Standard Normal Distribution (contd.)

The probability of $X<x_2$ would equivalently be given by:

$$P(X < x_2) = \int_{-\infty}^{z_2} \frac{1}{2\pi} e^{-\frac{1}{2}z^2} dz$$

$$= \int_{-\infty}^{z_2} \phi(z;\, 0,1) dz \qquad = P(Z < z_2)$$

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2)$$

Normal Distribution

Example 1

A microwave transmitting tube has been observed to follow a normal distribution with $\mu = 5000$ hours and $\sigma = 1500$ hours. Find the reliability of such a tube for a mission time of 4100 hours. (Source MIL Hdbk 338B)

At t=4100 hours,

$Z=(t-\mu)/\sigma = (4100-5000)/1500 = -0.6$

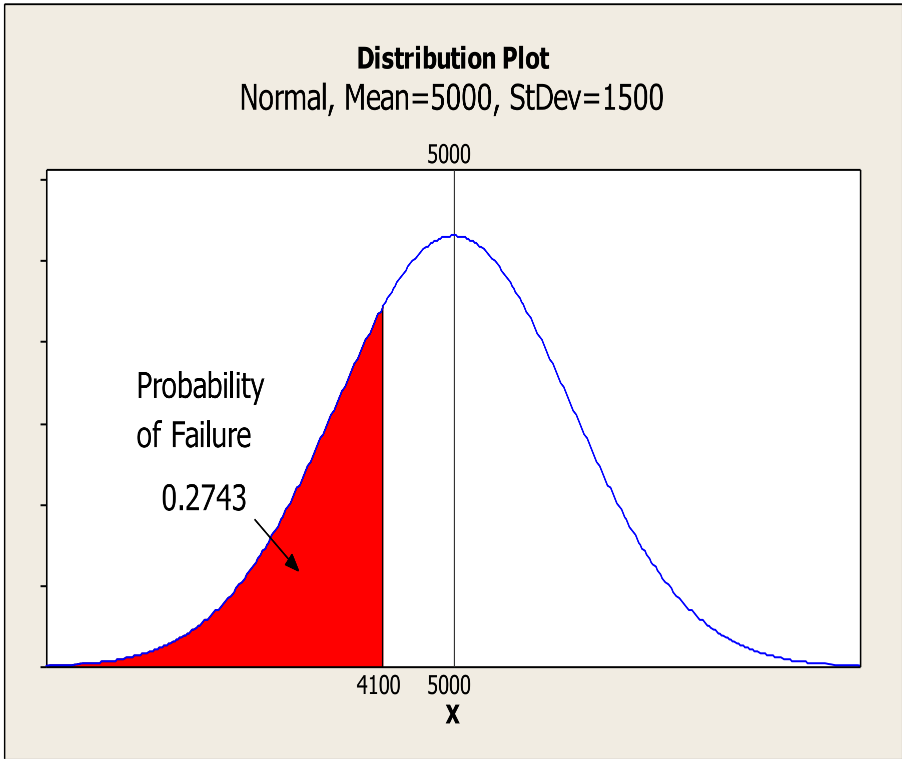
For Z=-0.6, shaded area is 0.2743. (next slide)
Note that this is area to the left as Z score is negative. This is therefore probability of failure.

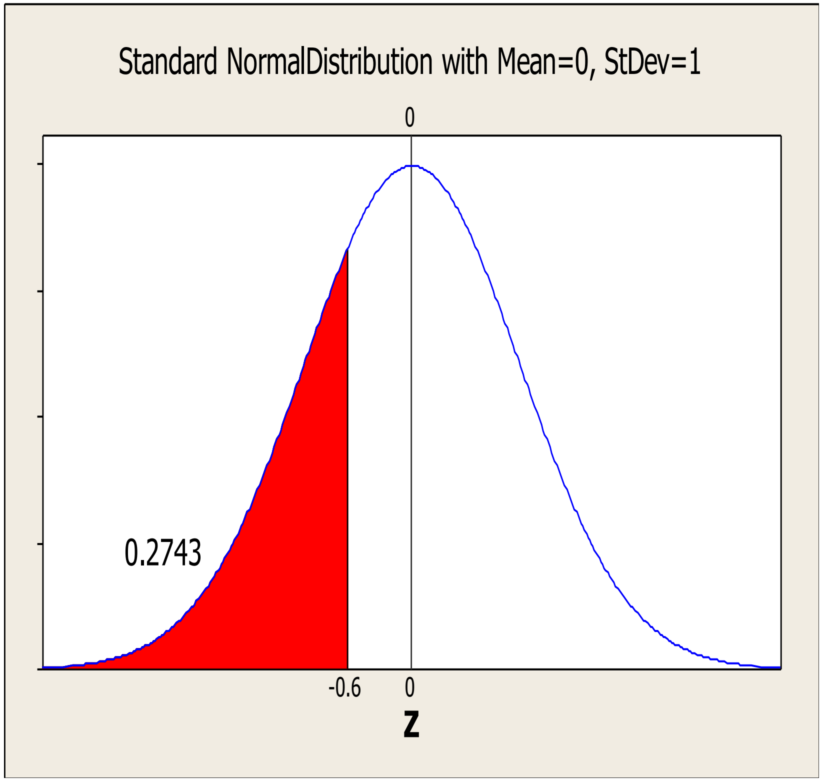
$F(4100)=0.2743$, $R(4100)=1-F(4100)=1-0.2743 = 0.7257$

Reliability at 4100 hours is 0.7257.

Normal Distribution

Example 1





Normal Distribution

Example 2

Wearout (failure) of an oil-drilling bit is normally distributed with a mean of 120 drilling hours and a standard deviation of 14 drilling hours. Drilling occurs for 12 hr each day. How many days should drilling continue before the operation is stopped in order to replace the drill bit ? A 95 percent reliability is desired.

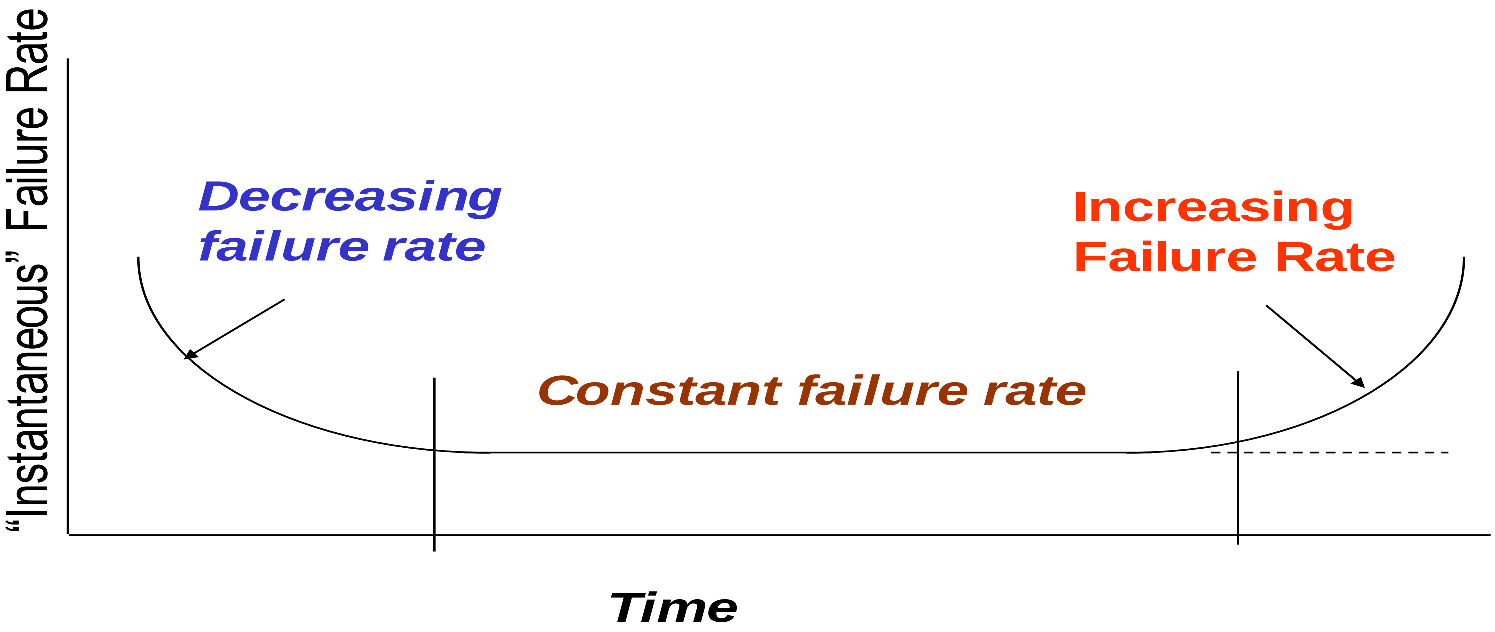
Find $t_{0.95}$ such that $\Pr\{T \geq t_{0.95}\} = 0.95$. Standardizing

$$\Pr\left\{z \geq \frac{t_{0.95} - 120}{14}\right\} = 1 - \Phi\left(\frac{t_{0.95} - 120}{14}\right) = 0.95$$

$$\left(\frac{t_{0.95} - 120}{14}\right) = -1.645 \qquad\qquad t_{0.95} = 96.97\, hr \;\approx 8\,(12 - hr)$$

Normal Distribution

Normal distribution can be a good fit for increasing failure rates

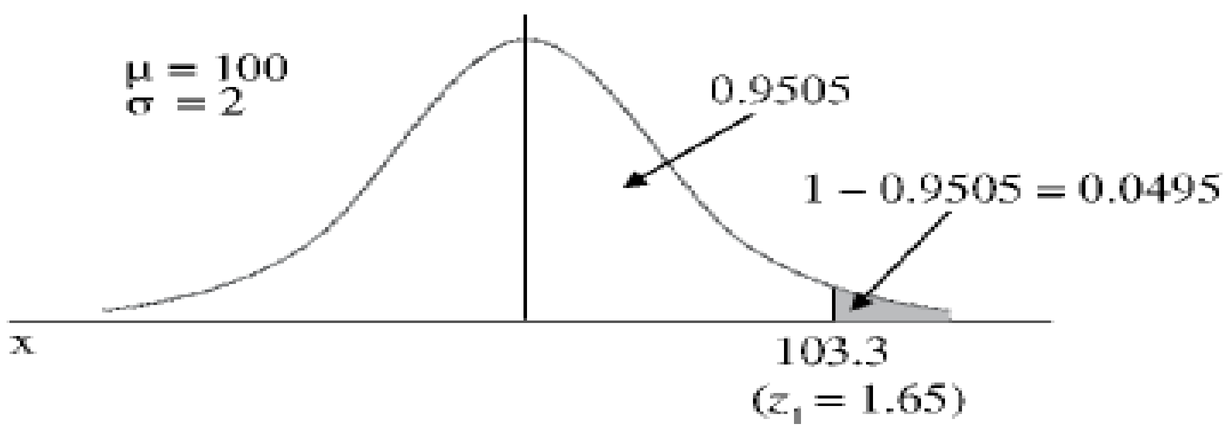


Normal Distribution

Example 3 :The length of a machined part is known to have a normal distribution with a mean of 100 mm and a standard deviation of 2mm.

(a)What proportion of the parts will be above 103.3mm?

Solution :Let X denote the length of the part. The parameter values for the normal distribution are μ.100 and σ.2. The probability required is shown in Figure a.



(a) $P(X > 103.3)$

Normal Distribution

The standardized value of 103.3 corresponds to

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{103.3 - 100}{2} = 1.65$$

Thus, $P(X > 103.3) = P(Z > 1.65)$. From Appendix A-3, $P(Z \leq 1.65) = 0.9505$, which also equals $P(X \leq 103.3)$. So

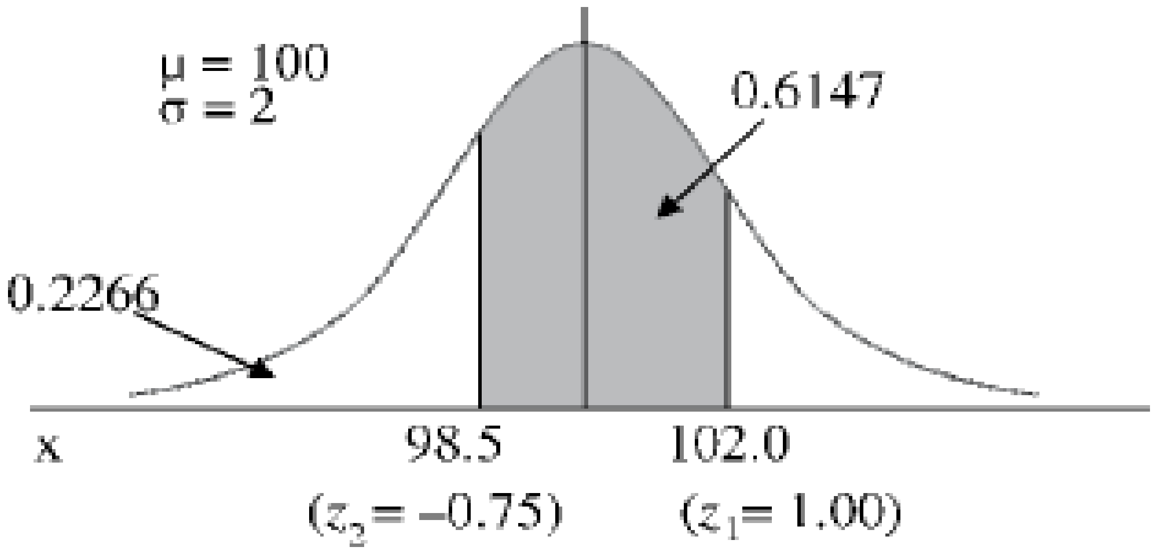
$$\begin{aligned} P(Z > 1.65) &= 1 - P(Z \leq 1.65) \\ &= 1 - 0.9505 = 0.0495 \end{aligned}$$

The desired probability $P(X > 103.3)$ is 0.0495, or 4.95%.

Normal Distribution

(**b)What proportion of the output will be between 98.5 and 102.0 mm?**

Solution :We wish to find P(98.5 ≤X ≤102.0), which is shown in Figure b.



(b) $P(98.5 \le X \le 102.0)$

Normal Distribution

The standardized values are computed as

$$z_1 = \frac{102.0 - 100}{2} = 1.00$$

$$z_2 = \frac{98.5 - 100}{2} = -0.75$$

From Appendix A-3, we have $P(Z \leq 1.00) = 0.8413$ and $P(Z \leq -0.75) = 0.2266$. The required probability equals $0.8413 - 0.2266 = 0.6147$. Thus, 61.47% of the output is expected to be between 98.5 and 102.0 mm

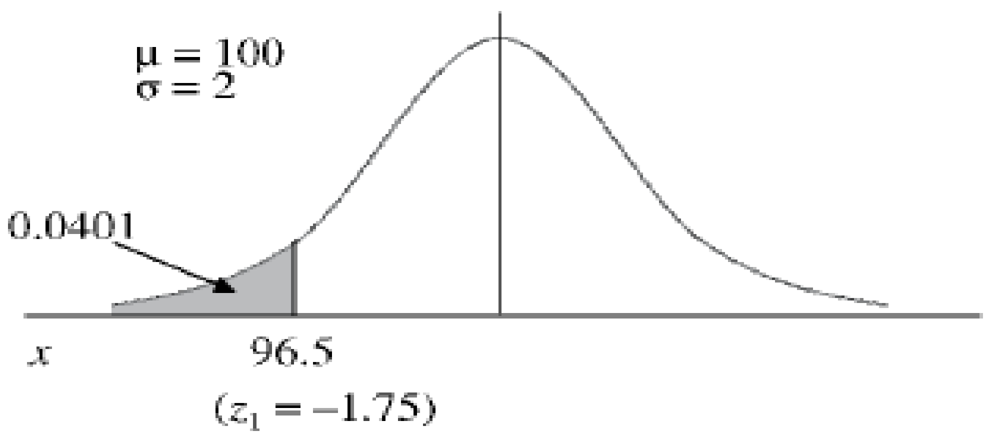
Normal Distribution

(c) What proportion of the parts will be shorter than 96.5 mm?

Solution We want $P(X < 96.5)$, which is equivalent to $P(X \leq 96.5)$, since for a continuous random variable the probability that the variable equals a particular value is zero. The standardized value is

$$z_1 = \frac{96.5 - 100}{2} = -1.75$$

The required proportion is shown in Figure c. Using Normal table, P(Z =-1.75)= 0.0401.Thus ,4.01% of the parts will have a length less than 96.5 mm.



(c) $P(X < 96.5)$

Normal Distribution

(d) It is important that not many of the parts exceed the desired length. If a manager stipulates that no more than 5% of the parts should be oversized, what specification limit should be recommended?

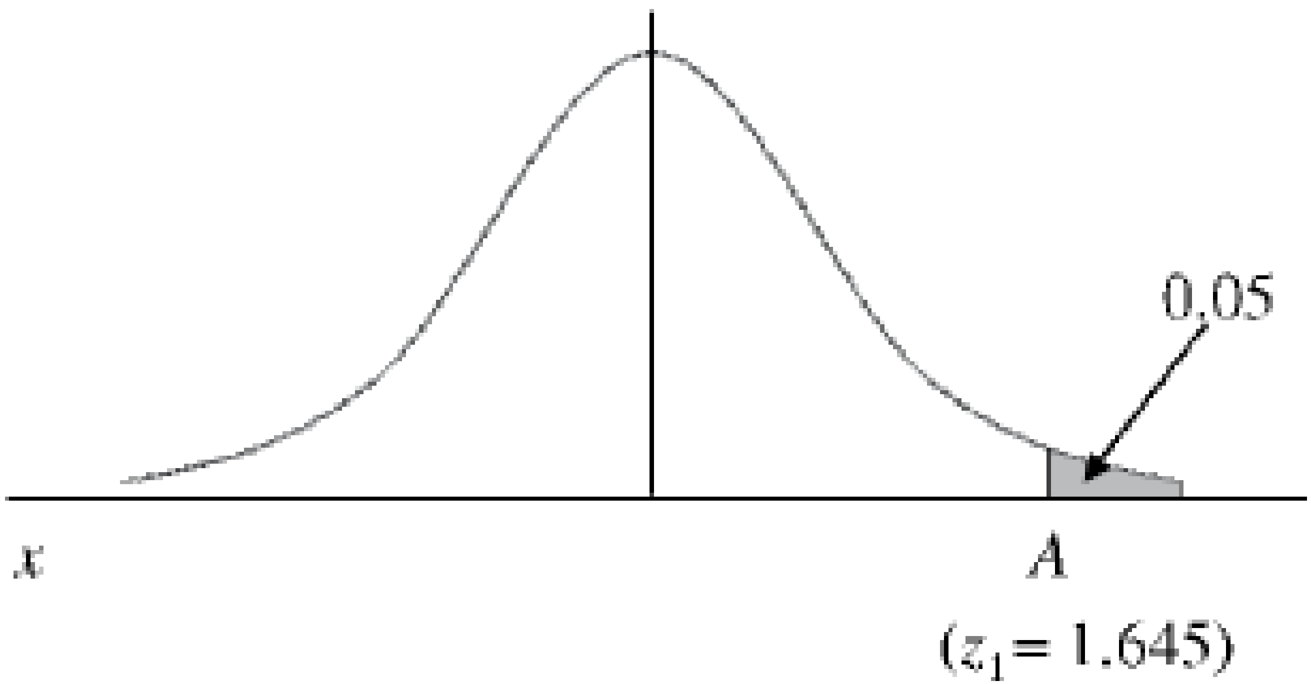
Solution Let the specification limit be A. From the problem information, $P(X \geq \text{A}) = 0.05$. To find A, we first find the standardized value at the point where the raw value is A. Here, the approach will be the reverse of what was done for the previous three parts of this example. That is, we are given an area, and we want to find the z-value. Here, $P(X \leq A) = 1–0.05 = 0.95$. We look for an area of 0.95 in Appendix A-3 and find that the linearly interpolated z-value is 1.645. Finally, we unstandardize this value to determine the limit A:

$$1.645 = \frac{x_1 - 100}{2}$$

$$x_1 = 103.29 \text{ mm}$$

Thus, A should be set at 103.29 mm to achieve the desired stipulation.

Normal Distribution



(d) $P(X \geq A) = 0.05$

Normal Distribution Tutorials

Problem 1: A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

Problem 2: An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

Problem 3: In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped?