Al-Muthanna University – Engineering collage

Chemical Department – 2nd class

ENGIEERING STATISTICS

Undergraduate study

Partial List of Symbols

- α alpha: Probability of a Type I error
- β beta: Probability of a Type II error
- β_1 Slope of a regression line
- β_0 Intercept of a regression line
- δ delta: A measure of effect size
- ϵ epsilon: The residual or error term in ANOVA and regression
- θ theta: The population median or the odds ratio
- μ mu: The population mean
- μ_t The population trimmed mean
- ν nu: Degrees of freedom
- Ω omega: The odds ratio
- ρ rho: The population correlation coefficient
- σ sigma: The population standard deviation
- ϕ phi: A measure of association
- χ chi: χ^2 is a type of distribution
- Δ delta: A measure of effect size
- $\sum$ Summation
- τ tau: Kendall's tau

Introduction to Statistics

Definitions:

Statistics: is the branch of scientific inquiry that provides methods for organizing and summarizing data, and for using information in the data to draw various conclusions.

Descriptive Statistics: The part of statistics that deals with methods for organization and summarization of data. Descriptive methods can be used with list of all population members (a census), or when the data consists of a samples.

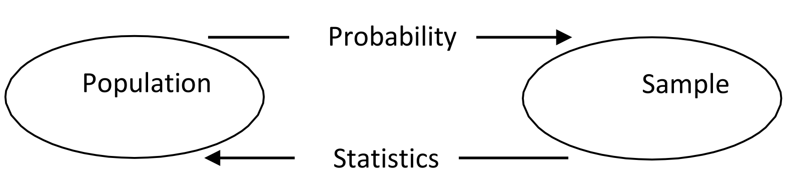
Inferential Statistics: When the data is a sample and the objective is to go beyond the sample to draw conclusions about the population based on sample information.

Population: A population of participants or objects consists of all those participants or objects that are relevant in a particular study.

Sample: A *sample* is any subset of the population of individuals or things under study.

Probability function: is a rule, denoted by $p(x)$ that assigns numbers to elements of the sample space

Link between statistics and Probability



Three fundamental components of statistics

Statistical techniques consist of a wide range of goals, techniques and strategies. Three fundamental components worth stressing are:

1. ***Design***, meaning the planning and carrying out of a study.

2. ***Description***, which refers to methods for summarizing data.

3. ***Inference***, which refers to making predictions or generalizations about a Population of individuals or things based on a sample of observations available to us.

Numerical Summaries of Data

1.0 Summation notation

In symbols, adding the numbers $X_1, X_2, \ldots, X_n$ is denoted by

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n,$$

where $\sum$ is an upper case Greek sigma. The subscript i is the index of summation and the 1 and n that appear respectively below and above the symbol $\sum$ designate the range of the summation.

Example 1:

$$1.2, 2.2, 6.4, 3.8, 0.9.$$

Then

$$\sum_{i=2}^{4} X_i = 2.2 + 6.4 + 3.8 = 12.4$$

and

$$\sum X_i = 1.2 + 2.2 + 6.4 + 3.8 + 0.9 = 14.5.$$

$$\sum X_i^2 = 1.2^2 + 2.2^2 + 6.4^2 + 3.8^2 + 0.9^2 = 62.49$$

and

$$\left(\sum X_i\right)^2 = (1.2 + 2.2 + 6.4 + 3.8 + 0.9)^2 = 14.5^2 = 210.25.$$

Problems

1. Given that

   $X_1 = 1 \quad X_2 = 3 \quad X_3 = 0$

   $X_4 = -2 \quad X_5 = 4 \quad X_6 = -1$

   $X_7 = 5 \quad X_8 = 2 \quad X_9 = 10$

   Find

   (a) $\sum X_i$, (b) $\sum_{i=3}^{5} X_i$, (c) $\sum_{i=1}^{4} X_i^3$, (d) $(\sum X_i)^2$, (e) $\sum 3$, (f) $\sum (X_i - 7)$

   (g) $3\sum_{i=1}^{5} X_i - \sum_{i=6}^{9} X_i$, (h) $\sum 10X_i$, (i) $\sum_{i=2}^{6} iX_i$, (j) $\sum 6$

2. Express the following in summation notation. (a) $X_1 + \frac{X_2}{2} + \frac{X_3}{3} + \frac{X_4}{4}$,
   (b) $U_1 + U_2^2 + U_3^3 + U_4^4$, (c) $(Y_1 + Y_2 + Y_3)^4$

3. Show by numerical example that $\sum X_i^2$ is not necessarily equal to $(\sum X_i)^2$.

Measures of location:

The sample mean:

The first measure of location, called the *sample mean*, is just the average of the values and is generally labeled $\bar{X}$. The notation $\bar{X}$ is read as X bar. In summation notation,

$$\bar{X} = \frac{1}{n}\sum X_i.$$

Example 1:

You sample ten married couples and determine the number of children they have. The results are 0, 4, 3, 2, 2, 3, 2, 1, 0, 8.

The sample mean is: $\bar{X}$ = (0+4+3+2+2+3+2+1+0+8)/10 = 2.5.

Of course, nobody has 2.5 children. The intention is to provide a number that is centrally located among the 10 observations with the goal of conveying what is typical.

Example 2

The salaries (in thousands Iraqi D) of the 11 individuals currently working at the company are:

300,250,320,280,350,310,300,360,290,2000,5000,

where the two largest salaries correspond to the vice president and president,

The average is 887, but it gives a distorted sense of what is typical!

Outliers are values that are unusually large or small.

2.0 The median

Another important measure of location is called the sample median. The basic idea is easily described using the example based on the weight of trout. The observed weights were
1.1,2.3,1.7,0.9,3.1.

Putting the values in ascending order yields

0.9,1.1,1.7,2.3,3.1.

Notice that the value 1.7 divides the observations in the middle in the sense that half of the remaining observations are less than 1.7 and half are larger.

If instead we have an even number of observations, there is no

middle value, 0.8, 1.3, 1.8, **2.6, 2.7**, 2.7, 3.1, 4.5

The sample median in this case is taken to be the average of 2.6 and 2.7, namely (2.6 + 2.7)/2 = 2.65.

<u>Problems</u>

4. Find the mean and median of the following sets of numbers. (a) −1, 03, 0, 2, −5. (b) 2, 2, 3, 10, 100, 1,000.

5. The final exam scores for 15 students are 73, 74, 92, 98, 100, 72, 74, 85, 76, 94,
89, 73, 76, 99. Compute the mean and median.

6. The average of 23 numbers is 14.7. What is the sum of these numbers?

7. Consider the ten values 3, 6, 8, 12, 23, 26, 37, 42, 49, 63. The mean is $\bar{X}$ = 26.9.

(a) What is the value of the mean if the largest value, 63, is increased to 100?
(b) What is the mean if 63 is increased to 1,000? (c) What is the mean if 63 is increased to 10,000?

8. Repeat the previous problem, only compute the median instead.

Measures of variation:

1.0The range

The *range* is just the difference between the largest and smallest observations. In symbols, it is $X(n) - X(1)$.

2.0 The variance and standard deviation

The following data written in ascending order:

7.5,8.0,8.0,8.5,9.0,11.0,19.5,19.5,28.5,31.0,36.0.

The data mean is $\bar{X} = 17$, so the deviation scores are

−9.5,−9.0,−9.0,−8.5,−8.0,−6.0,2.5,2.5,11.5,14.0,19.0.

Deviation scores reflect how far each observation is from the mean, but often it is best to find a single numerical quantity that summarizes the amount of variation in our data

The average difference is always ***zero***, so this approach is unsatisfactory
The average squared difference from the mean is called the ***sample variance***, which is:

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

The ***sample standard deviation*** is the (positive) square root of the variance, S.

<u>Example 1</u>

The following data are the sample test results

3,9,10,4,7,8,9,5,7,8.

The sample mean is $\bar{X} = 7$,

i	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	3	−4	16
2	9	2	4
3	10	3	9
4	4	−3	9
5	7	0	0
6	8	1	1
7	9	2	4
8	5	−2	4
9	7	0	0
10	8	1	1
$\sum$		0	48

The sum of the observations in the last column is

$$\sum(X_i - \bar{X})^2 = 48.$$

So,

$$S^2 = 48/9 = 5.33.$$

Problems

15. The height of 10 plants is measured in inches and found to be 12, 6, 15, 3, 12, 6, 21, 15, 18 and 12. Verify that $\sum(X_i - \bar{X}) = 0$.

16. For the data in the previous problem, compute the range, variance and standard deviation.

17. Use the rules of summation notation to show that it is always the case that $\sum(X_i - \bar{X}) = 0$.

18. Seven different thermometers were used to measure the temperature of a substance. The readings in degrees Celsius are −4.10, −4.13, −5.09, −4.08, −4.10, −4.09 and −4.12. Find the variance and standard deviation.

19. A weightlifter's maximum bench press (in pounds) in each of six successive weeks was 280, 295, 275, 305, 300, 290. Find the standard deviation.

GRAPHICAL SUMMARIES OF DATA:

1.0 Relative frequencies

The notation f_x is used to denote the frequency or number of times the value x occurs.

Plots of relative frequencies help add perspective on the sample variance, mean and median.

$$n = \sum f_x,$$

Table 1: One hundred results

2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 8 8 8 8

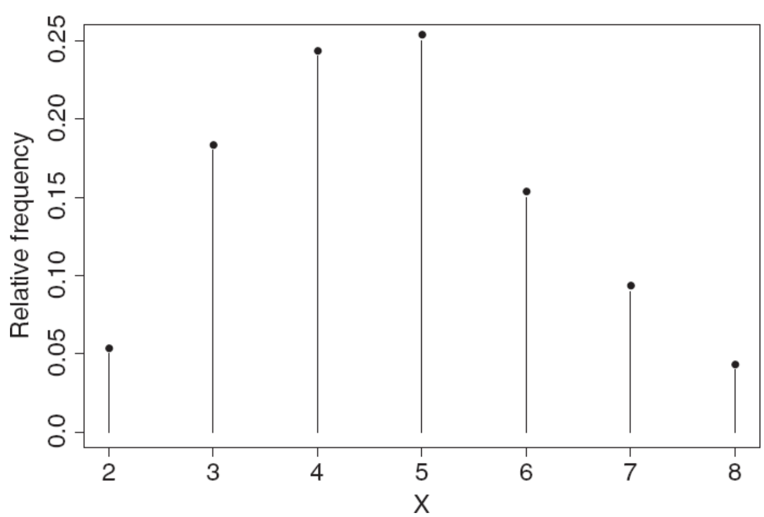


Figure 1: Relative frequencies for the data in table 1.

$$\sum f_x = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 + f_9 + f_{10}$$
$$= 0 + 5 + 18 + 24 + 25 + 15 + 9 + 40 + 0 + 0 = 100.$$

The sample mean is

$$\bar{X} = \frac{1}{n} \sum x f_x = \sum x \frac{f_x}{n}.$$

The sample variance is

$$s^2 = \frac{n}{n-1} \sum \frac{f_x}{n} (x - \bar{X})^2.$$

The ***cumulative relative frequency distribution*** $F(x)$ refers to the proportion of observations less than or equal to a given value.

Problems

1. Based on a sample of 100 individuals, the values 1, 2, 3, 4, 5 are observed with relative frequencies 0.2, 0.3, 0.1, 0.25, 0.15. Compute the mean, variance and standard deviation.

2. Fifty individuals are rated on how open minded they are. The ratings have the values 1, 2, 3, 4 and the corresponding relative frequencies are 0.2, 0.24, 0.4, 0.16, respectively. Compute the mean, variance and standard deviation.

3. For the values 0, 1, 2, 3, 4, 5, 6 the corresponding relative frequencies based on a sample of 10,000 observations are 0.015625, 0.093750, 0.234375, 0.312500, 0.234375, 0.093750, 0.015625, respectively. Determine the mean, median, variance, standard deviation and mode.

4. For a local charity, the donations in dollars received during the last month were 5, 10, 15, 20, 25, 50 having the frequencies 20, 30, 10, 40, 50, 5. Compute the mean, variance and standard deviation.

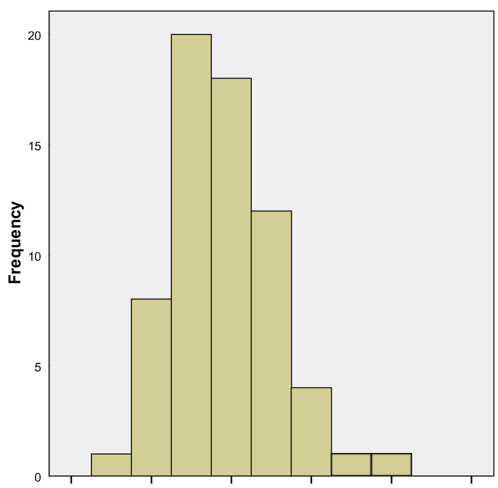
5. The values 1, 5, 10, 20 have the frequencies 10, 20, 40, 30. Compute the mean, variance and standard deviation.

2.0 **Histograms**: is an excellent graphical representation of the data.

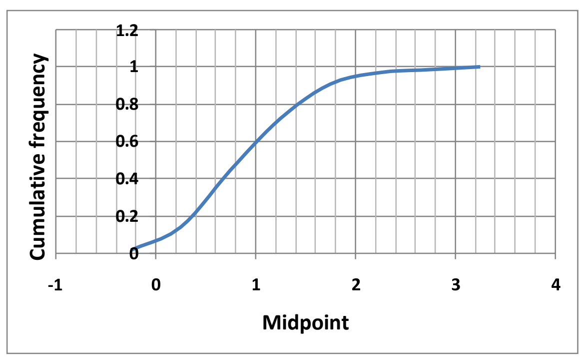
Table 2:

0.00	0.12	0.16	0.19	0.33	0.36	0.38	0.46	0.47	0.60	0.61	0.61	0.66	0.67	0.68
0.69	0.75	0.77	0.81	0.81	0.82	0.87	0.87	0.87	0.91	0.96	0.97	0.98	0.98	1.02
1.06	1.08	1.08	1.11	1.12	1.12	1.13	1.20	1.20	1.32	1.33	1.35	1.38	1.38	1.41
1.44	1.46	1.51	1.58	1.62	1.66	1.68	1.68	1.70	1.78	1.82	1.89	1.93	1.94	2.05
2.09	2.16	2.25	2.76	3.05										

class interval	**midpoint**	**frequency**	**Frequency Relative**	**Cumulative frequency**
–0.5–0.0	-0.25	1	1/65 = .0153	0.015385
>0.0–0.5	0.25	8	8/65 = .123	0.138462
>0.5–1.0	0.75	20	20/65 = .308	0.446154
>1.0–1.5	1.25	18	18/65 = .277	0.723077
>1.5–2.0	1.75	12	12/65 = .185	0.907692
>2.0–2.5	2.25	4	4/65 = .0625	0.969231
>2.5–3.0	2.75	1	1/65 = .0153	0.984615
>3.0–3.5	3.25	1	1/65 = .0153	1



Classes



Homework #1:

The frequency table below shows the compressive strength of concrete cubes results.

a) Construct a histogram, frequency table, frequency polygon, and cumulative frequency diagram?
b) Calculate mean value?
c) Calculate the percentage of the compressive strength results < 39.5 N/mm^2?
d) Calculate the percentage of the compressive strength results between a value of 36.5 and 39.5 N/mm^2?

Class interval	34 – <35	35 – <36	36 – <37	37 – <38	38-<39	39 –<40
Frequency	2	5	10	14	9	2

Homework #2:

The rainfall measurements data are 16, 22, 17, 18, 21, 14, 15, 23, 16, 19.

a) Arrange the data in ascending rank order?
b) Construct a histogram, frequency table, frequency polygon, and cumulative frequency diagram?
c) What is the probability of $(X \geq 13.5)$, (i.e compute $p(X \geq 13.5)$)?
d) compute $p(13.5 \leq X \geq 18.5)$?
e) compute $p(13.5 \leq X \geq 15.5)$?

Probability Theory

A ***random variable*** refers to a measurement or observation that cannot be known in advance.

An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a ***random experiment.***

Roman letter is used to represent a random variable, the most common letter being *X*.

A lower case x is used to represent an observed value corresponding to the random variable X. So the notation X =x means that the observed value of X is x.

The set of all possible outcomes or values of X we might observe is called the ***sample space***.

The set of all possible outcomes of a random experiment is called the ***sample space*** of the experiment. The sample space is denoted as *S*.

EXAMPLE 1:

Consider an experiment in which you select a plastic pipe, and measure its thickness.

Sample space as simply the positive real line because a negative value for thickness cannot occur

$S = R^{+} = \{ x | x>0 \}$

If it is known that all connectors will be between 10 and 11 millimeters thick, the sample space could be

$S = \{ x \mid 10 < x < 11 \}$

If the objective of the analysis is to consider only whether a particular part is ***low, medium, or high*** for thickness, the sample space might be taken to be the set of three outcomes:

S = { low, medium, high }

If the objective of the analysis is to consider only whether or not a particular part conforms to the manufacturing specifications, the sample space might be simplified to the set of two outcomes,

S = { yes, no }

that indicate whether or not the part conforms.

A ***discrete random variable*** meaning that there are gaps between any value and the next possible value.

A ***continuous random variable*** meaning that for any two outcomes, any value between these two values is possible.

Examples of Random Variables

Examples of **continuous** random variables:
electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:
number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error.

EXAMPLE 2:

If two connectors are selected and measured, the sample space is depending on the objective of the study.

If the objective of the analysis is to consider only whether or not the parts conform to the manufacturing specifications, either part may or may not conform. The sample space can be represented by the four outcomes:

$$S = \{ yy, yn, ny, nn \}$$

If we are only interested in the number of conforming parts in the sample, we might summarize the sample space as

$$S = \{ 0, 1, 2 \}$$

In random experiments in which items are selected from a batch, we will indicate whether or not a selected item is replaced before the next one is selected. For example, if the batch consists of three items **{*a, b, c*}** and our experiment is to select two items ***without replacement,*** the sample space can be represented as

$$S_{\text{without}} = \{ ab, ac, ba, bc, ca, cb \}$$

$$S_{\text{with}} = \{ aa, ab, ac, ba, bb, bc, ca, cb, cc \}$$

Events:

Often we are interested in a collection of related outcomes from a random experiment.

An **event** is a subset of the sample space of a random experiment.

Some of the basic set operations are summarized below in terms of events:

- The **union** of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as $E_1 \cup E_2$.
- The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.
- The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the component of the event *E as É*.

EXAMPLE 3:

Consider the sample space S {*yy*, *yn*, *ny*, *nn*} in Example 2. Suppose that the set of all outcomes for which at least one part conforms is denoted as E_1. Then,

$$E_1 = \{ yy, yn, ny \}$$

The event in which both parts do not conform, denoted as E_2, contains only the single outcome, $E_2\{nn\}$. Other examples of events are $E_3 = \emptyset$, the null set, and $E_4=S$, the sample space. If $E_5=\{yn, ny, nn\}$,

$$E_1 \text{ U } E_5 = S \qquad E_1 \cap E_5 = \{ yn , ny \} \qquad \acute{E}_1 = \{ nn \}$$

EXAMPLE 4:

Measurements of the time needed to complete a chemical reaction might be modeled with the sample space $S= R^+$, the set of positive real numbers. Let

$$E_1 = \{ x \mid 1 \leq x < 10\} \quad \text{and} \quad E_2 = \{ x \mid 1 < x < 118\}$$

Then,

$$E_1 \text{ U } E_2 = \{ x \mid 1 \leq x < 118\} \quad \text{and} \quad E_1 \cap E_2 = \{ x \mid 3 < x < 10\}$$

Also,

$$\acute{E}_1 = \{ x \mid x \geq 10\} \quad \text{and} \quad \acute{E}_1 \cap E_2 = \{ x \mid 10 \geq x < 118\}$$

EXAMPLE 5:

Samples of concrete surface are analyzed for abrasion resistance and impact strength. The results from 50 samples are summarized as follows:

		impact strength	
		High	Low
abrasion resistance	High	40	4
	Low	1	5

Let **A** denote the event that a sample has high impact strength,
Let **B** denote the event that a sample has high abrasion resistance.

Determine the number of samples in **A ∩ B**, **Á**, and **A U B**

The event **A ∩ B** consists of the 40 samples for which abrasion resistance and impact strength are high. The event **Á** consists of the 9 samples in which the impact strength is low. The event **A U B** consists of the 45 samples in which the abrasion resistance, impact strength, or both are high.

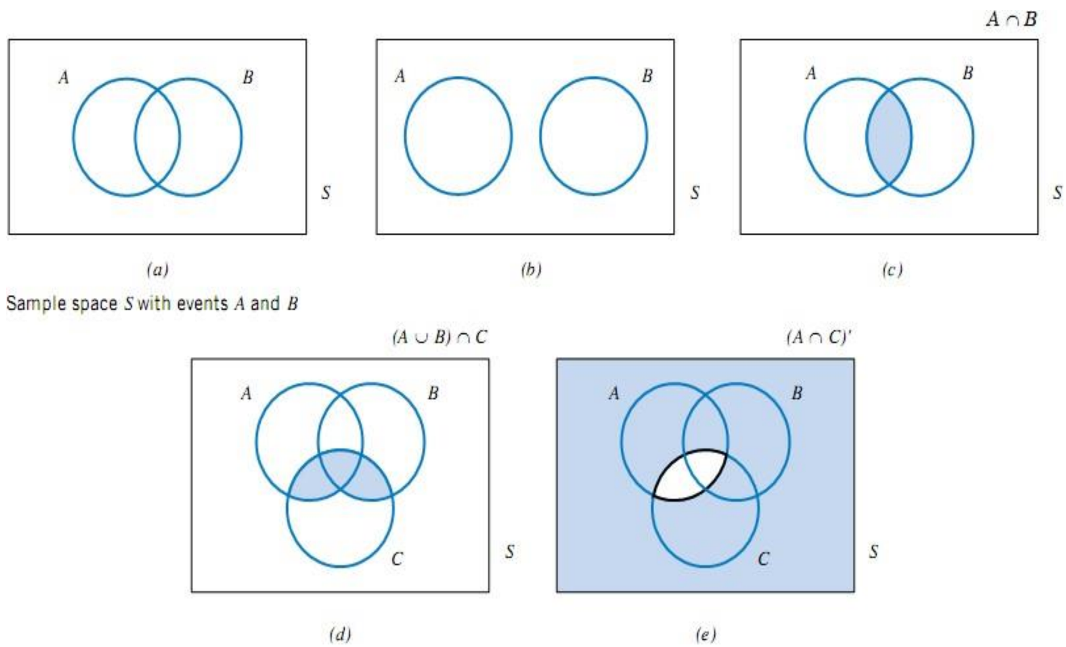


Figure 1: Venn diagrams

Venn diagrams are often used to describe relationships between events and sets.

Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \varnothing$$

are said to be ***mutually exclusive***.

The two events in Fig. 1(b) are mutually exclusive, whereas the two events in Fig. 1(a) are not. Additional results involving events are summarized below. The definition of the complement of an event implies that
1 *E*¿ 2 ¿ *E*
The distributive law for set operations implies that

Table 1: Corresponding statements in set theory and probability
Set theory Probability theory

Set theory	Probability theory
Space, S	Sample space, sure event
Empty set, $\emptyset$	Impossible event
Elements $a, b, \ldots$	Sample points a, $b, \ldots$ (or simple events)
Sets $A, B, \ldots$	Events A, $B, \ldots$
A	Event A occurs
$\bar{A}$	Event A does not occur
$A \cup B$	At least one of A and B occurs
AB	Both A and B occur
$A \subset B$	A is a subevent of B (i.e. the occurrence of A necessarily implies the occurrence of B)
$AB = \emptyset$	A and B are mutually exclusive (i.e. they cannot occur simultaneously)

Probability is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur. "The chance of rain today is 30%'' is a statement that quantifies our feeling about the possibility of rain.

A 0 probability indicates an outcome will not occur. A probability of 1 indicates an outcome will occur with certainty.

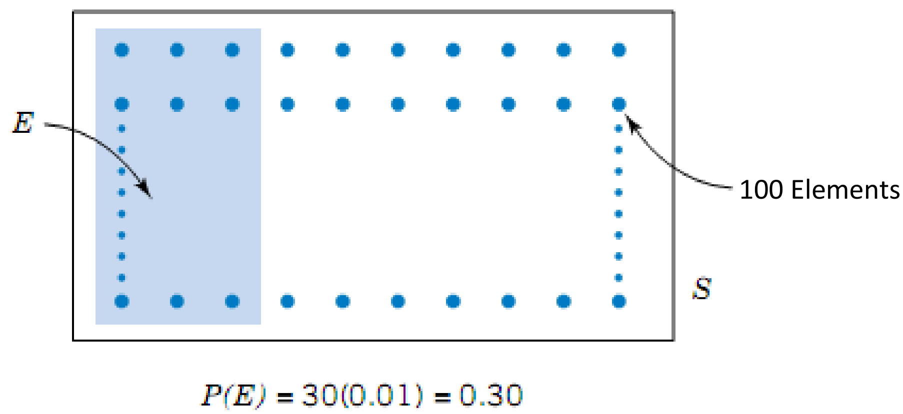


Fig. 2: Probability of the event E is the sum of the probabilities of the outcomes in E.

For a discrete sample space, the *probability of an event E*, denoted as P(*E*), equals the sum of the probabilities of the outcomes in *E*.

EXAMPLE 6:

A random experiment can result in one of the outcomes {*a*, *b*, *c*, *d*} with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let *A* denote the event {*a*, *b*}, *B* the event {*b*, *c*, *d*}, and *C* the event {*d*}.Then,

$P(A) = 0.1 + 0.3 = 0.4$
$P(B) = 0.3 + 0.5 + 0.1 = 0.9$
$P(C) = 0.1$

Also: $P(A') = 0.6$, $P(B') = 0.1$, $P(C') = 0.9$
$P(A \cap B) = 0.3$

P (A U B)= 1
P (A ∩ C)= 0

EXAMPLE 7:
A visual inspection of a defects location on concrete element manufacturing process resulted in the following table:

Number of defects	Proportion of concrete element
0	0.4
1	0.2
2	0.15
3	0.1
4	0.05
5 or more	0.1

a) If one element is selected randomly from this process to inspected, what is the probability that it contains no defects?

The event that there is no defect in the inspected concrete elements, denoted as E_1, can be considered to be comprised of the single outcome,

$$E_1 = \{0\}.$$

Therefore, $P(E_1) = 0.4$

b) What is the probability that it contains 3 or more defects?

Let the event that it contains 3 or more defects, denoted as E_2

$$P(E_2) = 0.1+0.05+0.1= 0.25$$

EXAMPLE 8:
Suppose that a batch contains six parts with part numbers {*a*, *b*, *c*, *d*, *e*, *f*}. Suppose that two parts are selected without replacement. Let *E* denote the event that the part number of the first part selected is *a*. Then *E* can be written as *E* {*ab*, *ac*, *ad*, *ae*, *af*}. The sample space can be counted. It has 30 outcomes. If each outcome is equally likely,

P(E) = 5/30 = 1/6

ADDITION RULES

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

EXAMPLE 8:

The defects such as those described in Example 7 were further classified as either in the “center’’ or at the “edge’’ of the concrete elements, and by the degree of damage. The following table shows the proportion of defects in each category. What is the probability that a defect was either at the edge or that it contains four or more defects?

Location in Concrete Element Surface			
Defects	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

Let E_1 denote the event that a defect contains four or more defects, and let E_2 denote the event that a defect is at the edge.

Defects Classified by Location and Degree

Number of defects	Center	Edge	Totals
0	0.30	0.10	0.40
1	0.15	0.05	0.20
2	0.10	0.05	0.15
3	0.06	0.04	0.10
4	0.04	0.01	0.05
5 or more	0.07	0.03	0.10
Totals	0.72	0.28	1.00

The requested probability is P (E_1 U E_2). Now, P (E_1) = 0.15 and P (E_2) = 0.28. Also, from the table above, P ($E_1 \cap E_2$) = 0.04

Therefore, P (E_1 U E_2) = 0.15 + 0.28 – 0.04 = 0.39

What is the probability that concrete surface contains less than two defects (denoted as E_3) or that it is both at the edge and contains more than four defects (denoted as E_4)?

The requested probability is P (E_3 U E_4). Now P (E_3) = 0.6, and P (E_4) = 0.03. Also, E_3 and E_4 are mutually exclusive.

Therefore, P ($E_3 \cap E_4$) = Ø

and P (E_3 U E_4) = 0.6 + 0.03 = 0.63

for the case of three events:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

A collection of events, $E_1, E_2, \ldots, E_k$, is said to be **mutually exclusive** if for all pairs,

$$E_i \cap E_j = \varnothing$$

For a collection of mutually exclusive events,

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = P(E_1) + P(E_2) + \ldots P(E_k)$$

EXAMPLE 9:

Let *X* denote the pH of a sample. Consider the event that *X* is greater than 6.5 but less than or equal to 7.8. This probability is the sum of any collection of mutually exclusive events with union equal to the same range for X. One example is:

$$P(6.5 < X \leq 7.8) = P(6.5 < X \leq 7.0) + P(7.0 < X \leq 7.5) + P(7.5 < X \leq 7.8)$$

Another example is

$$P(6.5 < X \leq 7.8) = P(6.5 < X \leq 6.6) + P(6.6 < X \leq 7.1) + P(7.1 < X \leq 7.4) + P(7.4 < X \leq 7.8)$$

The best choice depends on the particular probabilities available.

ENGIEERING STATISTICS

Lecture 4

2-49. If $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cap B) = 0.1$, determine the following probabilities:

(a) $P(A')$ (b) $P(A \cup B)$

(c) $P(A' \cap B)$ (d) $P(A \cap B')$

(e) $P[(A \cup B)']$ (f) $P(A' \cup B)$

2-50. If A, B, and C are mutually exclusive events with $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:

(a) $P(A \cup B \cup C)$ (b) $P(A \cap B \cap C)$

(c) $P(A \cap B)$ (d) $P[(A \cup B) \cap C]$

(e) $P(A' \cap B' \cap C')$

2-51. If A, B, and C are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

2-52. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		shock resistance	
		high	low
scratch	high	70	9
resistance	low	16	5

(a) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?

(b) If a disk is selected at random, what is the probability that its scratch resistance is high or its shock resistance is high?

(c) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?

2-53. The analysis of shafts for a compressor is summarized by conformance to specifications.

		roundness conforms	
		yes	no
surface finish	yes	345	5
conforms	no	12	8

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements?

(b) What is the probability that the selected shaft conforms to surface finish requirements or to roundness requirements?

(c) What is the probability that the selected shaft either conforms to surface finish requirements or does not conform to roundness requirements?

(d) What is the probability that the selected shaft conforms to both surface finish and roundness requirements?

2-54. Cooking oil is produced in two main varieties: mono- and polyunsaturated. Two common sources of cooking oil are corn and canola. The following table shows the number of bottles of these oils at a supermarket:

		type of oil	
		canola	corn
type of	mono	7	13
unsaturation	poly	93	77

(a) If a bottle of oil is selected at random, what is the probability that it belongs to the polyunsaturated category?

(b) What is the probability that the chosen bottle is monounsaturated canola oil?

2-55. A manufacturer of front lights for automobiles tests lamps under a high humidity, high temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

		useful life	
		satisfactory	unsatisfactory
intensity	satisfactory	117	3
	unsatisfactory	8	2

(a) Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.

(b) The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand?

2-56. The shafts in Exercise 2-53 are further classified in terms of the machine tool that was used for manufacturing the shaft.

Tool 1

		roundness conforms	
		yes	no
surface finish	yes	200	1
conforms	no	4	2

Tool 2

		roundness conforms	
		yes	no
surface finish	yes	145	4
conforms	no	8	6

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or to roundness requirements or is from Tool 1?

(b) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or does not conform to roundness requirements or is from Tool 2?

(c) If a shaft is selected at random, what is the probability that the shaft conforms to both surface finish and roundness requirements or the shaft is from Tool 2?

(d) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or the shaft is from Tool 2?

ENGIEERING STATISTICS

Lecture 4

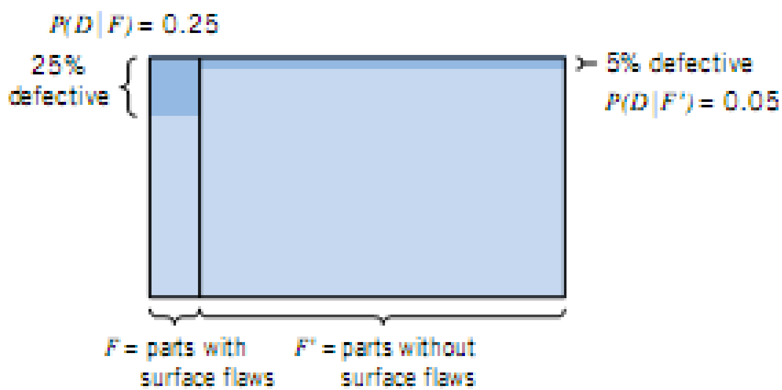
CONDITIONAL PROBABILITY

The **conditional probability** of an event B given an event A, denoted as $P(B|A)$, is

$$P(B|A) = P(A \cap B)/P(A)$$

for $P(A) > 0$.

In a manufacturing process, 10% of the parts contain visible surface flaws and 25% of the parts with surface flaws are (functionally) defective parts. However, only 5% of parts without surface flaws are defective parts. The probability of a defective part depends on our knowledge of the presence or absence of a surface flaw.



Let D denote the event that a part is defective
and let F denote the event that a part has a surface flaw.

Then, the probability of D given, or assuming, that a part has a surface flaw as $P(D|F)$. This notation is read as the **conditional probability** of D given F, and it is interpreted as the probability that a part is defective, given that the part has a surface flaw.

EXAMPLE 1:

Table 1 below provides an example of 400 parts classified by surface flaws and as (functionally) defective. For this table the conditional probabilities match those discussed previously in this section. For example, of the parts with surface flaws (40 parts) the number defective is 10.

Table 1: Parts Classified

		Surface Flaws		
		Yes (event F)	No	Total
Defective	Yes (event D)	10	18	38
	No	30	342	362
	Total	40	360	400

Therefore,

$$P(D|F) = 10/40 = 0.25$$

and of the parts without surface flaws (360 parts) the number defective is 18. Therefore,

$$P(D|F') = 18/360 = 0.05$$

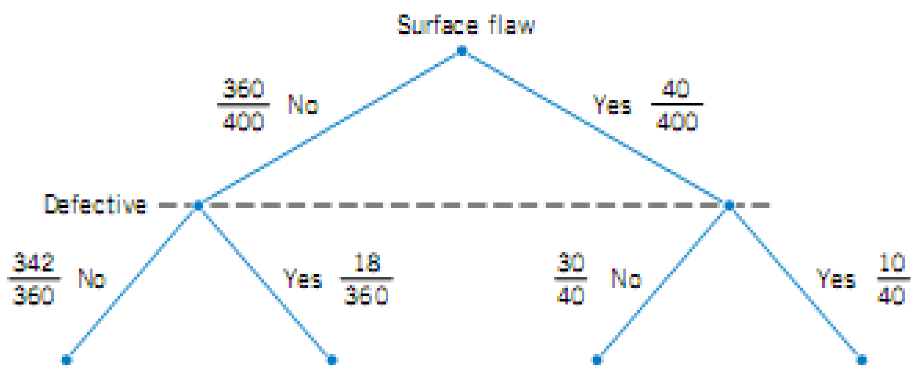


Figure 1: Tree diagram for parts classified

Therefore, P (B | A) can be interpreted as the relative frequency of event *B* among the trials that produce an outcome in event *A*.

EXAMPLE 2:

Again consider the 400 parts in Table 1 above (example 1). From this table

$$P(D|F) = P(D \cap F)/P(F) = \frac{10}{400}\Big/\frac{40}{400} = \frac{10}{40}$$

Note that in this example all four of the following probabilities are different:

$$P(F) = 40/400 \qquad P(F|D) = 10/28$$

$$P(D) = 28/400 \qquad P(D|F) = 10/40$$

Here, P (D) and P (D | F) are probabilities of the same event, but they are computed under two different states of knowledge.

Similarly, P (F) and P (F | D),

The tree diagram in Fig. 1 can also be used to display conditional probabilities.

$$P(D|F) = 10/40 \qquad \text{and} \qquad P(D'|F) = 30/40$$

Multiplication Rule (for counting techniques)

> If an operation can be described as a sequence of k steps, and
>
> if the number of ways of completing step 1 is n_1, and
>
> if the number of ways of completing step 2 is n_2 for each way of completing step 1, and
>
> if the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth,
>
> the total number of ways of completing the operation is
>
> $$n_1 \times n_2 \times \cdots \times n_k$$

Permutations

Another useful calculation is the number of ordered sequences of the elements of a set. Consider a set of elements, such as S {a, b, c}. A **permutation** of the elements is an ordered sequence of the elements. For example, *abc*, *acb*, *bac*, *bca*, *cab*, and *cba* are all of the permutations of the elements of S.

The number of **permutations** of n different elements is $n!$ where

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

In some situations, we are interested in the number of arrangements of only some of the elements of a set. The following result also follows from the multiplication rule.

The number of permutations of a subset of r elements selected from a set of n different elements is

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

<u>EXAMPLE 3:</u>

A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are possible?

Each design consists of selecting a location from the eight locations for the first component, a location from the remaining seven for the second component, a location from the remaining six for the third component, and a location from the remaining five for the fourth component. Therefore,

$$P_4^8 = 8 \times 7 \times 6 \times 5 = \frac{8!}{4!} = 1680 \text{ different designs are possible.}$$

Combinations

Another counting problem of interest is the number of subsets of *r* elements that can be selected from a set of *n* elements. Here, order is not important.

> The number of subsets of size r that can be selected from a set of n elements is denoted as $\binom{n}{r}$ or C_r^n and
>
> $$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

EXAMPLE 4:

A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible? Each design is a subset of the eight locations that are to contain the components. From the Equation above, the number of possible designs is

$$\frac{8!}{5!\,3!} = 56$$

The following example uses the multiplication rule in combination with the above equation to answer a more difficult, but common, question.

EXAMPLE 5:

A bin of 50 manufactured parts contains three defective parts and 47 non-defective parts. A sample of six parts is selected from the 50 parts. Selected parts are not replaced. That is, each part can only be selected once and the sample is a subset of the 50 parts. How many different samples are there of size six that contain exactly two defective parts?

A subset containing exactly two defective parts can be formed by first choosing the two defective parts from the three defective parts.

$$\binom{3}{2} = \frac{3!}{2!\ 1!} = 3 \text{ different ways}$$

Then, the second step is to select the remaining four parts from the 47 acceptable parts in the bin. The second step can be completed in

$$\binom{47}{4} = \frac{47!}{4!\ 43!} = 178{,}365 \text{ different ways}$$

Therefore, from the multiplication rule, the number of subsets of size six that contain exactly two defective items is

$$3 * 178{,}365 = 535{,}095$$

As an additional computation, the total number of different subsets of size six is found to be

$$\binom{50}{6} = \frac{50!}{6!\ 44!} = 15{,}890{,}700$$

Therefore, the probability that a sample contains exactly two defective parts is

$$\frac{535{,}095}{15{,}890{,}700} = 0.034$$

S2-1. An order for a personal digital assistant can specify any one of five memory sizes, any one of three types of displays, any one of four sizes of a hard disk, and can either include or not include a pen tablet. How many different systems can be ordered?

S2-2. In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by polishing, and followed by painting) for a part are possible?

S2-3. New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three locations for input valves, and four locations for output valves. How many different product designs are possible?

S2-4. A manufacturing process consists of 10 operations that can be completed in any order. How many different production sequences are possible?

S2-5. A manufacturing operations consists of 10 operations. However, five machining operations must be completed before any of the remaining five assembly operations can begin. Within each set of five, operations can be completed in any order. How many different production sequences are possible?

S2-6. In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

S2-7. A lot of 140 semiconductor chips is inspected by choosing a sample of five chips. Assume 10 of the chips do not conform to customer requirements.

(a) How many different samples are possible?

(b) How many samples of five contain exactly one nonconforming chip?

(c) How many samples of five contain at least one nonconforming chip?

S2-8. In the layout of a printed circuit board for an electronic product, there are 12 different locations that can accommodate chips.

(a) If five different types of chips are to be placed on the board, how many different layouts are possible?

(b) If the five chips that are placed on the board are of the same type, how many different layouts are possible?

S2-9. In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed two times each day to check the calibration of the laboratory instruments.

(a) How many different sequences of process and control samples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.

(b) How many different sequences of process and control samples are possible if we consider the five process samples to be different and the two control samples to be identical.

(c) For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

S2-10. In the design of an electromechanical product, seven different components are to be stacked into a cylindrical casing that holds 12 components in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top.

(a) How many different designs are possible?

(b) If the seven components are all identical, how many different designs are possible?

(c) If the seven components consist of three of one type of component and four of another type, how many different designs are possible? (more difficult)

S2-11. The design of a communication system considered the following questions:

(a) How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?

(b) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?

(c) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

S2-12. A byte is a sequence of eight bits and each bit is either 0 or 1.

(a) How many different bytes are possible?

(b) If the first bit of a byte is a parity check, that is, the first byte is determined from the other seven bits, how many different bytes are possible?

S2-13. In a chemical plant, 24 holding tanks are used for final product storage. Four tanks are selected at random and without replacement. Suppose that six of the tanks contain material in which the viscosity exceeds the customer requirements.

(a) What is the probability that exactly one tank in the sample contains high viscosity material?

(b) What is the probability that at least one tank in the sample contains high viscosity material?

(c) In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample contains high viscosity material and exactly one tank in the sample contains material with high impurities?

S2-14. Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts fall into a conveyor when the press opens. An inspector chooses 3 parts from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.

(a) What is the probability that the inspector finds exactly one nonconforming part?

(b) What is the probability that the inspector finds at least one nonconforming part?

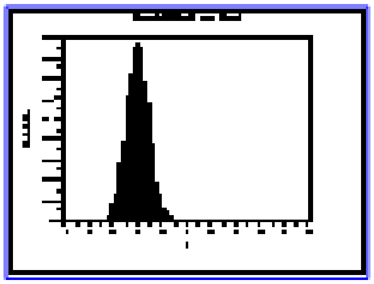
S2-15. A bin of 50 parts contains five that are defective. A sample of two is selected at random, without replacement.

(a) Determine the probability that both parts in the sample are defective by computing a conditional probability.

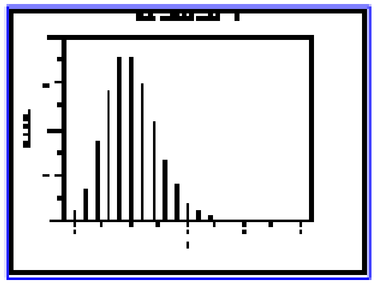
(b) Determine the answer to part (a) by using the subset approach that was described in this section.

Distributions

Discrete Distributions:

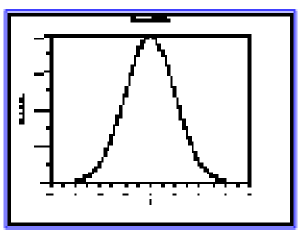


Binomial Distribution

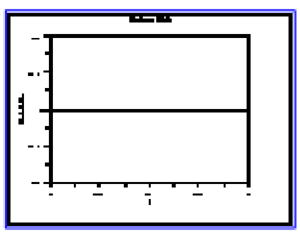


Poisson Distribution

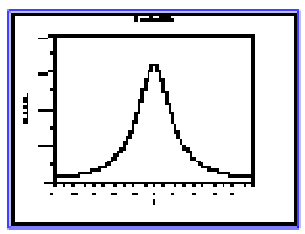
Continuous Distributions:



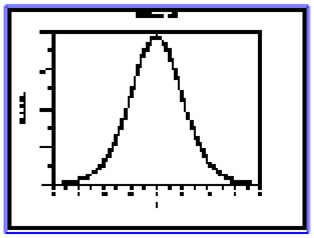
Normal Distribution



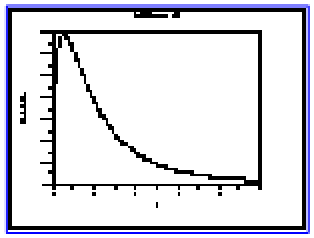
Uniform Distribution



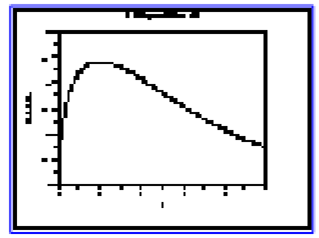
Cauchy Distribution



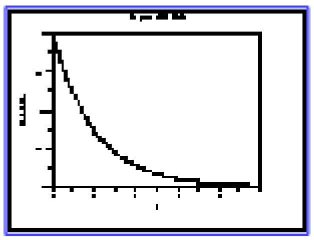
t Distribution



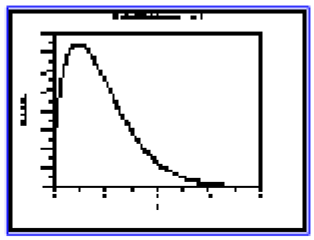
F Distribution



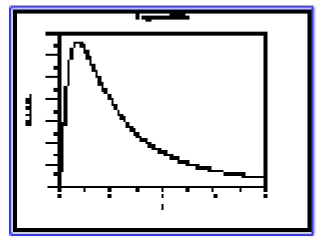
Chi-Square Distribution



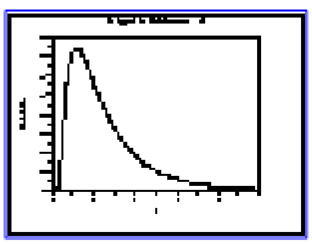
Exponential Distribution



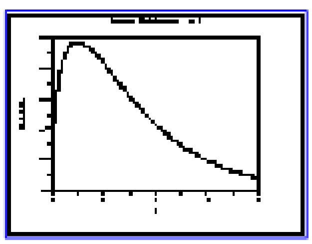
Weibull Distribution



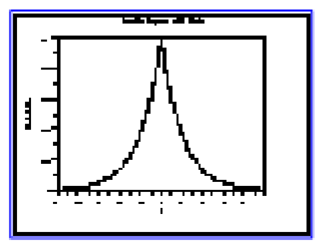
Lognormal Distribution



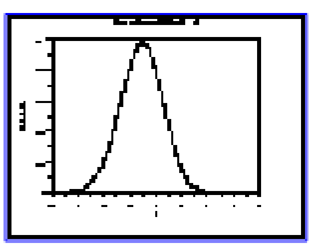
Fatigue Life Distribution



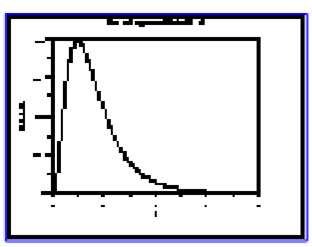
Gamma Distribution



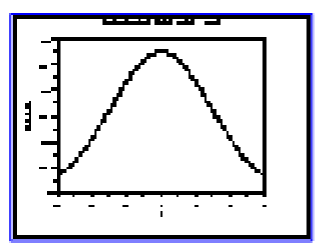
Double Exponential Distribution



Power Normal Distribution



Power Lognormal Distribution



Tukey-Lambda Distribution

Definition:

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

BINOMIAL DISTRIBUTION:

Definition:

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent

(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"

(3) The probability of a success in each trial, denoted as p, remains constant

The random variable X that equals the number of trials that result in a success has a **binomial random variable** with parameters $0 < p < 1$ and $n = 1, 2, \ldots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

EXAMPLE 1:

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, exactly 2 contain the pollutant. Let X the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with $p= 0.1$ and $n= 18$. Therefore,

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16}$$

Now $\binom{18}{2} = 18!/[2!\ 16!] = 18(17)/2 = 153$. Therefore,

$$P(X = 2) = 153(0.1)^2(0.9)^{16} = 0.284$$

Determine the probability that at least four samples contain the pollutant?

The requested probability is

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$

However, it is easier to use the complementary event,

$$P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{x=0}^{3} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$= 1 - [0.150 + 0.300 + 0.284 + 0.168] = 0.098$$

Determine the probability that $3 \leq X < 7$. Now

$$P(3 \leq X < 7) = \sum_{x=3}^{6} \binom{18}{x}(0.1)^x(0.9)^{18-x}$$
$$= 0.168 + 0.070 + 0.022 + 0.005$$
$$= 0.265$$

The mean and variance of a binomial random variable depend only on the parameters *p* and *n*.

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p)$$

EXERCISES:

1. For each scenario described below, state whether or not the binomial distribution is a reasonable model for the random variable and why. State any assumptions you make.

(a) A production process produces thousands of temperature transducers. Let *X* denote the number of nonconforming transducers in a sample of size 30 selected at random from the process.

(b) From a batch of 50 temperature transducers, a sample of size 30 is selected without replacement. Let *X* denote the number of nonconforming transducers in the sample.

(c) Four identical electronic components are wired to a controller that can switch from a failed component to one of the remaining spares. Let *X* denote the number of components that have failed after a specified period of operation.

(d) Defects occur randomly over the surface of a semiconductor chip. However, only 80% of defects can be found by testing. A sample of 40 chips with one defect each is tested. Let *X* denote the number of chips in which the test finds a defect.

2. The random variable X has a binomial distribution with n=10 and p=0.5. Determine the following probabilities:
(a) $P(X = 5)$ (b) $P(X \leq 2)$ (c) $P(X \geq 9)$ (d) $P(3 \leq X < 5)$

3. Sketch the probability mass function of a binomial distribution with n =10 and p = 0.01 and comment on the shape of the distribution.
(a) What value of X is most likely? (b) What value of X is least likely?

4. Batches that consist of 50 concrete blocks from a production process are checked for conformance to building requirements. The mean number of nonconforming concrete blocks in a batch is 5. Assume that the number of nonconforming concrete blocks in a batch, denoted as X, is a binomial random variable.
(a) What are n and p? (b) What is $P(X \leq 2)$? (c) What is $P(X \geq 49)$?

5. A manufacturing process has 100 customer orders to fill. Each order requires one component part that is purchased from a supplier. However, typically, 2% of the components are identified as defective, and the components can be assumed to be independent.

a) If the manufacturer stocks 100 components, what is the probability that the 100 orders can be filled without reordering components?
b) If the manufacturer stocks 102 components, what is the probability that the 100 orders can be filled without reordering components?
c) If the manufacturer stocks 105 components, what is the probability that the 100 orders can be filled without reordering components?

(This exercise illustrates that poor quality can affect schedules and costs).

POISSON DISTRIBUTION:

The random variable X that equals the number of counts in the interval is a **Poisson random variable** with parameter $0 < \lambda$, and the probability mass function of X is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots$$

If X is a Poisson random variable with parameter λ, then

$$\mu = E(X) = \lambda \qquad \text{and} \qquad \sigma^2 = V(X) = \lambda$$

EXAMPLE 2:

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter. Determine the probability of exactly 2 flaws in 1 millimeter of wire. Let X denote the number of flaws in 1 millimeter of wire. Then, $E(X) = 2.3$ flaws and

$$P(X = 2) = \frac{e^{-2.3}2.3^2}{2!} = 0.265$$

Determine the probability of 10 flaws in 5 millimeters of wire. Let X denote the number of flaws in 5 millimeters of wire. Then, X has a Poisson distribution with

$$E(X) = 5 \text{ mm} \times 2.3 \text{ flaws/mm} = 11.5 \text{ flaws}$$

Therefore,

$$P(X = 10) = e^{-11.5} \frac{11.5^{10}}{10!} = 0.113$$

Determine the probability of at least 1 flaw in 2 millimeters of wire. Let X denote the number of flaws in 2 millimeters of wire. Then, X has a Poisson distribution with

E(X) = 2 mm × 2.3 flaws/mm = 4.6 flaws

Therefore,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6} = 0.9899$$

EXERCISES:

3-97. Suppose X has a Poisson distribution with a mean of 4. Determine the following probabilities:
(a) $P(X = 0)$ (b) $P(X \leq 2)$
(c) $P(X = 4)$ (d) $P(X = 8)$

3-98. Suppose X has a Poisson distribution with a mean of 0.4. Determine the following probabilities:
(a) $P(X = 0)$ (b) $P(X \leq 2)$
(c) $P(X = 4)$ (d) $P(X = 8)$

3-99. Suppose that the number of customers that enter a bank in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$. Determine the mean and variance of X.

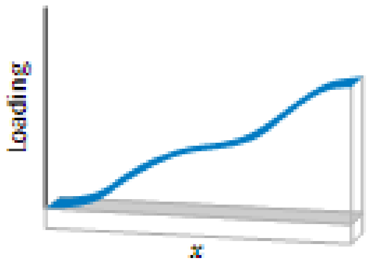
3-100. The number of telephone calls that arrive at a phone exchange is often modeled as a Poisson random variable. Assume that on the average there are 10 calls per hour.
(a) What is the probability that there are exactly 5 calls in one hour?
(b) What is the probability that there are 3 or less calls in one hour?
(c) What is the probability that there are exactly 15 calls in two hours?
(d) What is the probability that there are exactly 5 calls in 30 minutes?

3-101. The number of flaws in bolts of cloth in textile manufacturing is assumed to be Poisson distributed with a mean of 0.1 flaw per square meter.
(a) What is the probability that there are two flaws in 1 square meter of cloth?
(b) What is the probability that there is one flaw in 10 square meters of cloth?
(c) What is the probability that there are no flaws in 20 square meters of cloth?
(d) What is the probability that there are at least two flaws in 10 square meters of cloth?

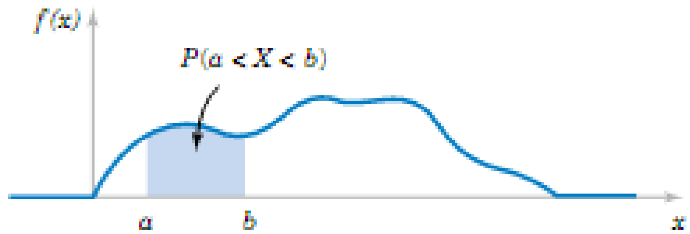
3-102. When a computer disk manufacturer tests a disk, it writes to the disk and then tests it using a certifier. The certifier counts the number of missing pulses or errors. The number of errors on a test area on a disk has a Poisson distribution with $\lambda = 0.2$.
(a) What is the expected number of errors per test area?
(b) What percentage of test areas have two or fewer errors?

3-103. The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.
(a) What is the probability that there are no cracks that require repair in 5 miles of highway?
(b) What is the probability that at least one crack requires repair in 1/2 mile of highway?
(c) If the number of cracks is related to the vehicle load on the highway and some sections of the highway have a heavy load of vehicles whereas other sections carry a light load, how do you feel about the assumption of a Poisson distribution for the number of cracks that require repair?

3-104. The number of failures for a cytogenics machine from contamination in biological samples is a Poisson random variable with a mean of 0.01 per 100 samples.
(a) If the lab usually processes 500 samples per day, what is the expected number of failures per day?



Density of a loading on a long, thin beam



Probability determined from the area under f(x)

Definition:

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx =$ area under $f(x)$ from a to b

for any a and b

For the density function of a loading on a long thin beam, because every point has zero width, the loading at any point is zero. Similarly, for a continuous random variable X and *any* value x.

$$P(X = x) = 0$$

If X is a **continuous random variable,** for any x_1 and x_2,

$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$$

EXAMPLE:

Let the continuous random variable *X* denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 mm. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of *X* can be modeled by a probability density function $f(x) = 20\,e^{-20(x-12.5)}$, $x \geq 12.5$.

If a part with a diameter larger than 12.60 millimeters is scrapped, what proportion of parts is scrapped? The density function and the requested probability are shown in Fig. 2. A part is scrapped if $X \geq 12.60$. Now,

$$P(X > 12.60) = \int_{12.6}^{\infty} f(x)\,dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}\,dx = -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = 0.135$$

`

What proportion of parts is between 12.5 and 12.6 millimeters? Now,

$$P(12.5 < X < 12.6) = \int_{12.5}^{12.6} f(x)\,dx = -e^{-20(x-12.5)}\Big|_{12.5}^{12.6} = 0.865$$

Because the total area under *f* (*x*) equals 1, we can also calculate

P (12.5< X <12.62) = 1 – P(X > 12.62) = 1- 0.135= 0.865.

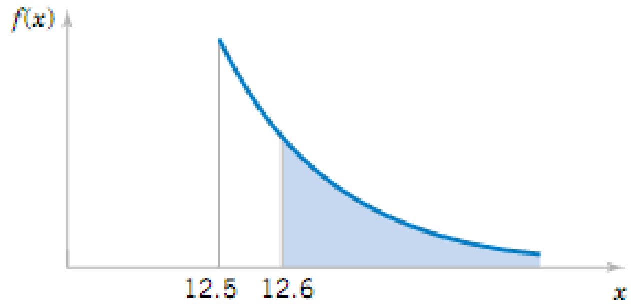


Figure 2: Probability density function

ENGIEERING STATISTICS

Lecture 7

EXERCISES:

4-1. Suppose that $f(x) = e^{-x}$ for $0 < x$. Determine the following probabilities:

(a) $P(1 < X)$ (b) $P(1 < X < 2.5)$

(c) $P(X = 3)$ (d) $P(X < 4)$

(e) $P(3 \le X)$

4-2. Suppose that $f(x) = e^{-x}$ for $0 < x$.

(a) Determine x such that $P(x < X) = 0.10$.

(b) Determine x such that $P(X \le x) = 0.10$.

4-3. Suppose that $f(x) = x/8$ for $3 < x < 5$. Determine the following probabilities:

(a) $P(X < 4)$ (b) $P(X > 3.5)$

(c) $P(4 < X < 5)$ (d) $P(X < 4.5)$

(e) $P(X < 3.5 \text{ or } X > 4.5)$

4-4. Suppose that $f(x) = e^{-(x-4)}$ for $4 < x$. Determine the following probabilities:

(a) $P(1 < X)$ (b) $P(2 \le X < 5)$

(c) $P(5 < X)$ (d) $P(8 < X < 12)$

(e) Determine x such that $P(X < x) = 0.90$.

4-5. Suppose that $f(x) = 1.5x^2$ for $-1 < x < 1$. Determine the following probabilities:

(a) $P(0 < X)$ (b) $P(0.5 < X)$

(c) $P(-0.5 \le X \le 0.5)$ (d) $P(X < -2)$

(e) $P(X < 0 \text{ or } X > -0.5)$

(f) Determine x such that $P(x < X) = 0.05$.

4-6. The probability density function of the time to failure of an electronic component in a copier (in hours) is $f(x) = \frac{e^{-x/1000}}{1000}$ for $x > 0$. Determine the probability that

(a) A component lasts more than 3000 hours before failure.

(b) A component fails in the interval from 1000 to 2000 hours.

(c) A component fails before 1000 hours.

(d) Determine the number of hours at which 10% of all components have failed.

4-7. The probability density function of the net weight in pounds of a packaged chemical herbicide is $f(x) = 2.0$ for $49.75 < x < 50.25$ pounds.

(a) Determine the probability that a package weighs more than 50 pounds.

(b) How much chemical is contained in 90% of all packages?

4-8. The probability density function of the length of a hinge for fastening a door is $f(x) = 1.25$ for $74.6 < x < 75.4$ millimeters. Determine the following:

(a) $P(X < 74.8)$

(b) $P(X < 74.8 \quad \text{or} \quad X > 75.2)$

(c) If the specifications for this process are from 74.7 to 75.3 millimeters, what proportion of hinges meets specifications?

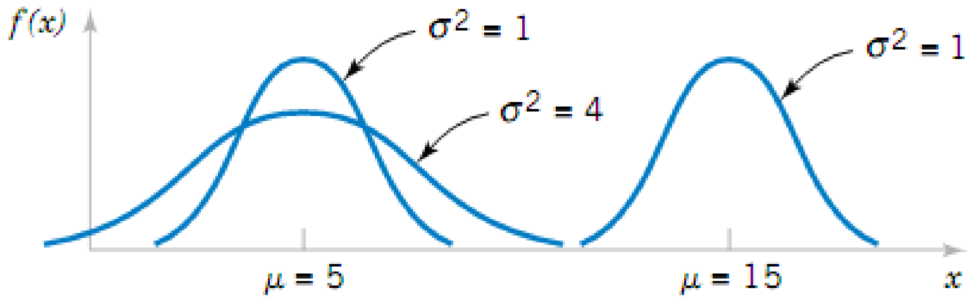
4-9. The probability density function of the length of a metal rod is $f(x) = 2$ for $2.3 < x < 2.8$ meters.

(a) If the specifications for this process are from 2.25 to 2.75 meters, what proportion of the bars fail to meet the specifications?

(b) Assume that the probability density function is $f(x) = 2$ for an interval of length 0.5 meters. Over what value should the density be centered to achieve the greatest proportion of bars within specifications?

4-10. If X is a continuous random variable, argue that $P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$.

<u>NORMAL DISTRIBUTION:</u>



Normal probability density functions for selected values of the parameters μ and σ^2

Definition:

A random variable X with probability density function

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

is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of X are shown to equal μ and σ^2, respectively, at the end of this Section 5-6.

<u>EXAMPLE 4:</u>

Assume that the current measurements in a strip of wire follow a normal distribution with a mean of 10 mA and a variance of 4 $(mA)^2$. What is the probability that a measurement exceeds 13 mA?

Let X denote the current in mA. The requested probability can be represented as:

$$P(X > 13)$$

This probability is shown as the shaded area under the normal probability density function in Fig. 3.

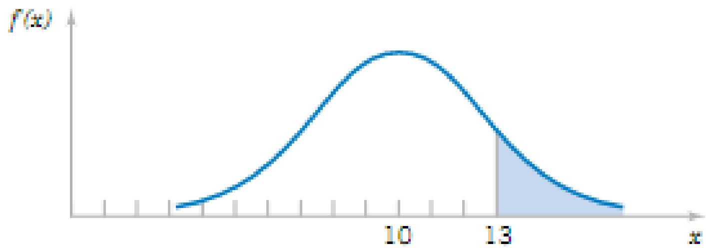


Figure 3 Probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$.

Some useful results concerning a normal distribution are summarized below and in Fig. 4. For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

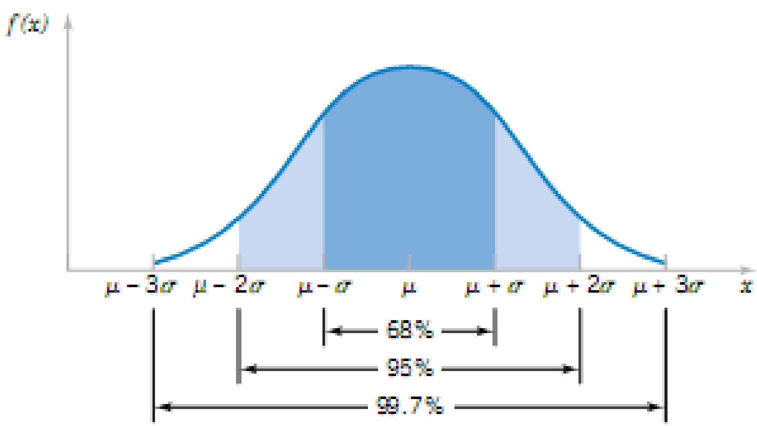


Figure 4 Probabilities associated with a normal distribution.

Definition:

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z.

Summary of Common Probability Distributions

Name	Probability Distribution	Mean	Variance
Discrete			
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$, $x = 0, 1, \ldots, n, 0 \le p \le 1$	np	$np(1-p)$
Geometric	$(1-p)^{x-1}p$, $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$, $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$
Hypergeometric	$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$, $x = \max(0, n-N+K), 1, \ldots \min(K, n), K \le N, n \le N$	np, where $p = \frac{K}{N}$	$np(1-p)\left(\frac{N-n}{N-1}\right)$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$	λ	λ
Continuous			
Uniform	$\frac{1}{b-a}, a \le x \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma$	μ	σ^2
Exponential	$\lambda e^{-\lambda x}, 0 \le x, 0 < \lambda$	$1/\lambda$	$1/\lambda^2$
Erlang	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, 0 < x, r = 1, 2, \ldots$	r/λ	r/λ^2
Gamma	$\frac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 < x, 0 < r, 0 < \lambda$	r/λ	r/λ^2
Weibull	$\frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^\beta}$, $0 < x, 0 < \beta, 0 < \delta$	$\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2$
Lognormal	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(\frac{-[\ln(x)-\theta]^2}{2\omega^2}\right)$	$e^{\theta+\omega^2/2}$	$e^{2\theta+\omega^2}(e^{\omega^2}-1)$

ENGIEERING STATISTICS

Lecture 8

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

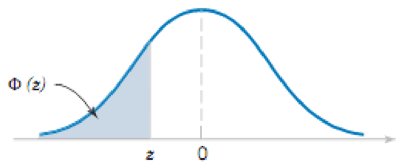


Table II Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968
−3.0	0.001001	0.001035	0.001070	0.001107	0.001144	0.001183	0.001223	0.001264	0.001306	0.001350
−2.9	0.001395	0.001441	0.001489	0.001538	0.001589	0.001641	0.001695	0.001750	0.001807	0.001866
−2.8	0.001926	0.001988	0.002052	0.002118	0.002186	0.002256	0.002327	0.002401	0.002477	0.002555
−2.7	0.002635	0.002718	0.002803	0.002890	0.002980	0.003072	0.003167	0.003264	0.003364	0.003467
−2.6	0.003573	0.003681	0.003793	0.003907	0.004025	0.004145	0.004269	0.004396	0.004527	0.004661
−2.5	0.004799	0.004940	0.005085	0.005234	0.005386	0.005543	0.005703	0.005868	0.006037	0.006210
−2.4	0.006387	0.006569	0.006756	0.006947	0.007143	0.007344	0.007549	0.007760	0.007976	0.008198
−2.3	0.008424	0.008656	0.008894	0.009137	0.009387	0.009642	0.009903	0.010170	0.010444	0.010724
−2.2	0.011011	0.011304	0.011604	0.011911	0.012224	0.012545	0.012874	0.013209	0.013553	0.013903
−2.1	0.014262	0.014629	0.015003	0.015386	0.015778	0.016177	0.016586	0.017003	0.017429	0.017864
−2.0	0.018309	0.018763	0.019226	0.019699	0.020182	0.020675	0.021178	0.021692	0.022216	0.022750
−1.9	0.023295	0.023852	0.024419	0.024998	0.025588	0.026190	0.026803	0.027429	0.028067	0.028717
−1.8	0.029379	0.030054	0.030742	0.031443	0.032157	0.032884	0.033625	0.034379	0.035148	0.035930
−1.7	0.036727	0.037538	0.038364	0.039204	0.040059	0.040929	0.041815	0.042716	0.043633	0.044565
−1.6	0.045514	0.046479	0.047460	0.048457	0.049471	0.050503	0.051551	0.052616	0.053699	0.054799
−1.5	0.055917	0.057053	0.058208	0.059380	0.060571	0.061780	0.063008	0.064256	0.065522	0.066807
−1.4	0.068112	0.069437	0.070781	0.072145	0.073529	0.074934	0.076359	0.077804	0.079270	0.080757
−1.3	0.082264	0.083793	0.085343	0.086915	0.088508	0.090123	0.091759	0.093418	0.095098	0.096801
−1.2	0.098525	0.100273	0.102042	0.103835	0.105650	0.107488	0.109349	0.111233	0.113140	0.115070
−1.1	0.117023	0.119000	0.121001	0.123024	0.125072	0.127143	0.129238	0.131357	0.133500	0.135666
−1.0	0.137857	0.140071	0.142310	0.144572	0.146859	0.149170	0.151505	0.153864	0.156248	0.158655
−0.9	0.161087	0.163543	0.166023	0.168528	0.171056	0.173609	0.176185	0.178786	0.181411	0.184060
−0.8	0.186733	0.189430	0.192150	0.194894	0.197662	0.200454	0.203269	0.206108	0.208970	0.211855
−0.7	0.214764	0.217695	0.220650	0.223627	0.226627	0.229650	0.232695	0.235762	0.238852	0.241964
−0.6	0.245097	0.248252	0.251429	0.254627	0.257846	0.261086	0.264347	0.267629	0.270931	0.274253
−0.5	0.277595	0.280957	0.284339	0.287740	0.291160	0.294599	0.298056	0.301532	0.305026	0.308538
−0.4	0.312067	0.315614	0.319178	0.322758	0.326355	0.329969	0.333598	0.337243	0.340903	0.344578
−0.3	0.348268	0.351973	0.355691	0.359424	0.363169	0.366928	0.370700	0.374484	0.378281	0.382089
−0.2	0.385908	0.389739	0.393580	0.397432	0.401294	0.405165	0.409046	0.412936	0.416834	0.420740
−0.1	0.424655	0.428576	0.432505	0.436441	0.440382	0.444330	0.448283	0.452242	0.456205	0.460172
0.0	0.464144	0.468119	0.472097	0.476078	0.480061	0.484047	0.488033	0.492022	0.496011	0.500000

ENGIEERING STATISTICS

Lecture 8

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

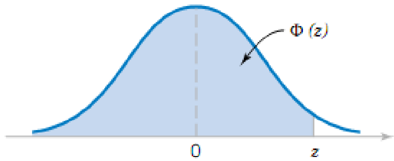


Table II Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

ENGIEERING STATISTICS

Lecture 8

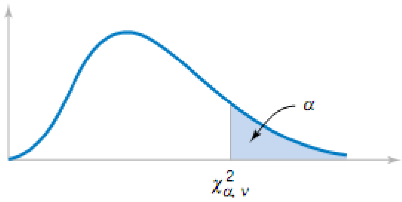


Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

= degrees of freedom.

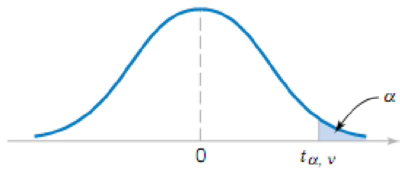


Table IV Percentage Points $t_{\alpha,\nu}$ of the t-Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

EXAMPLE 5:

The following calculations are shown pictorially in Fig. 5.

(1) $P(Z > 1.26) = 1 - P(Z \leq 1.26) = 1 - 0.89616 = 0.10384$

(2) $P(Z < -0.86) = 0.19490$.

(3) $P(Z > -1.37) = P(Z < 1.37) = 0.91465$

(4) $P(-1.25 < Z < 0.37)$. This probability can be found from the difference of two areas, $P(Z < 0.37) - P(Z < -1.25)$. Now,

$$P(Z < 0.37) = 0.64431 \quad \text{and} \quad P(Z < -1.25) = 0.10565$$

Therefore,

$$P(-1.25 < Z < 0.37) = 0.64431 - 0.10565 = 0.53866$$

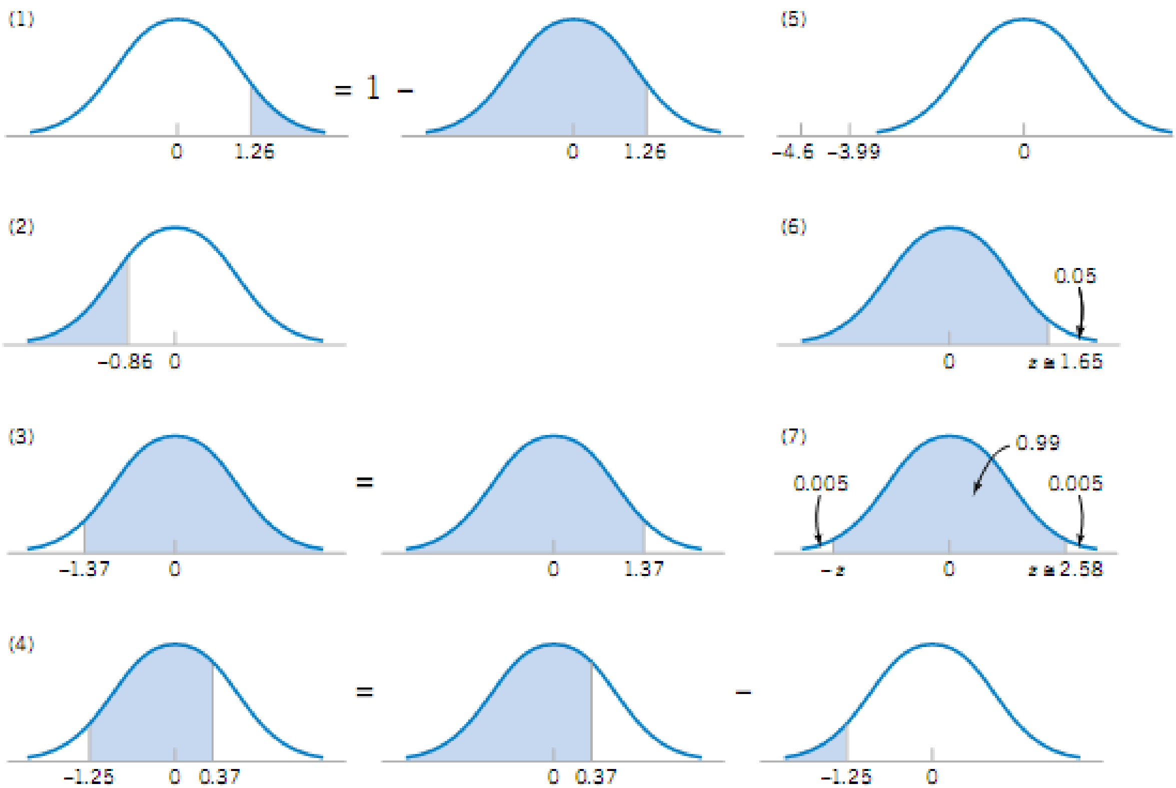


Figure 5: Graphical displays for standard normal distributions.

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

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

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

EXAMPLE 6:

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 mA and a variance of 4 $(mA)^2$. What is the probability that a measurement will exceed 13 mA?

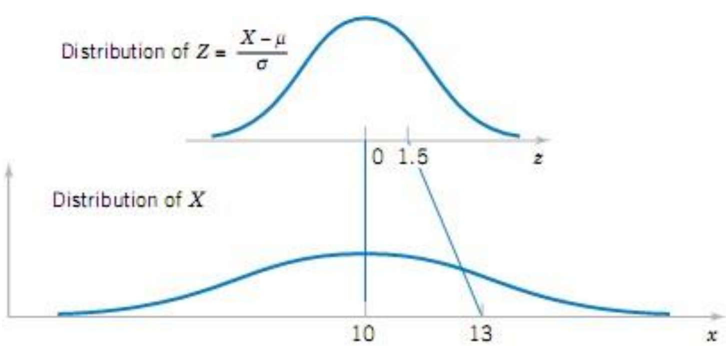
Let X denote the current in mA.

The requested probability can be represented as $P(X > 13)$.

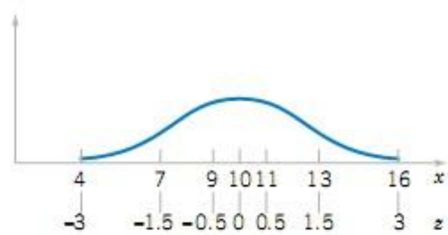
Let $Z = (X - 10)/2$.

We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table II,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$$



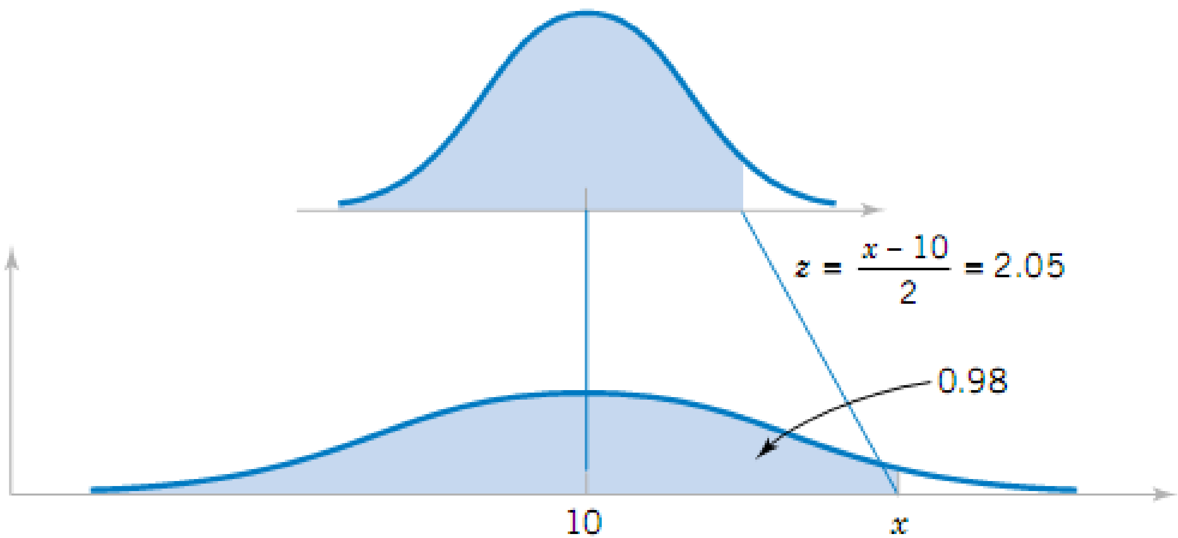
Standardizing a normal random variable.



EXAMPLE 7: Continuing the previous example, what is the probability that a current measurement is between 9 and 11 mA?

$$\begin{aligned} P(9 < X < 11) &= P((9 - 10)/2 < (X - 10)/2 < (11 - 10)/2) \\ &= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z < -0.5) \\ &= 0.69146 - 0.30854 = 0.38292 \end{aligned}$$

Determine the value for which the probability that a current measurement is below this value is 0.98. The requested value is shown graphically in the figure below. We need the value of x such that P(X < x) = 0.98. By standardizing, this probability expression can be written as



Appendix Table II is used to find the z-value such that P (Z < z) = 0.98. The nearest probability from Table II results in

P (Z< 2.05) = 0.97982

Therefore, (x - 10)/ 2= 2.05, and the standardizing transformation is used in reverse to solve for x. The result is

x = 2(2.05)/10 = 14.1 mA

EXAMPLE 8: The diameter of a shaft in an optical storage drive is normally distributed with `mean 0.2508 inch and standard deviation 0.0005 inch. The specifications on the shaft are 0.2500 ± 0.0015 inch. What proportion of shafts conforms to specifications?

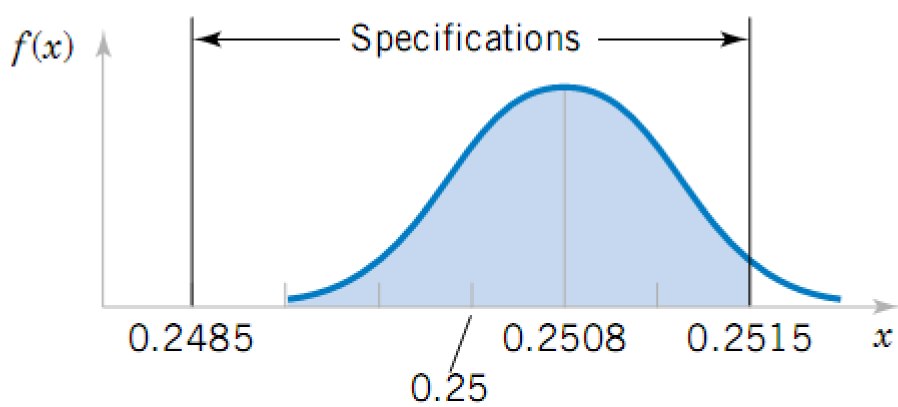
Let X denote the shaft diameter in inches. The requested probability is shown in the figure below and

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485 - 0.2508}{0.0005} < Z < \frac{0.2515 - 0.2508}{0.0005}\right)$$
$$= P(-4.6 < Z < 1.4) = P(Z < 1.4) - P(Z < -4.6)$$
$$= 0.91924 - 0.0000 = 0.91924$$

Most of the nonconforming shafts are too large, because the process mean is located very near to the upper specification limit. If the process is centered so that the process mean is equal to the target value of 0.2500,

$$P(0.2485 < X < 0.2515) = P\left(\frac{0.2485 - 0.2500}{0.0005} < Z < \frac{0.2515 - 0.2500}{0.0005}\right)$$
$$= P(-3 < Z < 3)$$
$$= P(Z < 3) - P(Z < -3)$$
$$= 0.99865 - 0.00135$$
$$= 0.9973$$

By recentering the process, the yield is increased to approximately 99.73%.



EXERCISES:

4-39. Use Appendix Table II to determine the following probabilities for the standard normal random variable Z:
(a) $P(Z < 1.32)$ (b) $P(Z < 3.0)$
(c) $P(Z > 1.45)$ (d) $P(Z > -2.15)$
(e) $P(-2.34 < Z < 1.76)$

4-40. Use Appendix Table II to determine the following probabilities for the standard normal random variable Z:
(a) $P(-1 < Z < 1)$ (b) $P(-2 < Z < 2)$
(c) $P(-3 < Z < 3)$ (d) $P(Z > 3)$
(e) $P(0 < Z < 1)$

4-41. Assume Z has a standard normal distribution. Use Appendix Table II to determine the value for z that solves each of the following:
(a) $P(Z < z) = 0.9$ (b) $P(Z < z) = 0.5$
(c) $P(Z > z) = 0.1$ (d) $P(Z > z) = 0.9$
(e) $P(-1.24 < Z < z) = 0.8$

4-42. Assume Z has a standard normal distribution. Use Appendix Table II to determine the value for z that solves each of the following:
(a) $P(-z < Z < z) = 0.95$ (b) $P(-z < Z < z) = 0.99$
(c) $P(-z < Z < z) = 0.68$ (d) $P(-z < Z < z) = 0.9973$

4-43. Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the following:
(a) $P(X < 13)$ (b) $P(X > 9)$
(c) $P(6 < X < 14)$ (d) $P(2 < X < 4)$
(e) $P(-2 < X < 8)$

4-44. Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the value for x that solves each of the following:
(a) $P(X > x) = 0.5$
(b) $P(X > x) = 0.95$
(c) $P(x < X < 10) = 0.2$
(d) $P(-x < X - 10 < x) = 0.95$
(e) $P(-x < X - 10 < x) = 0.99$

4-45. Assume X is normally distributed with a mean of 5 and a standard deviation of 4. Determine the following:
(a) $P(X < 11)$ (b) $P(X > 0)$
(c) $P(3 < X < 7)$ (d) $P(-2 < X < 9)$
(e) $P(2 < X < 8)$

4-46. Assume X is normally distributed with a mean of 5 and a standard deviation of 4. Determine the value for x that solves each of the following:
(a) $P(X > x) = 0.5$ (b) $P(X > x) = 0.95$
(c) $P(x < X < 9) = 0.2$ (d) $P(3 < X < x) = 0.95$
(e) $P(-x < X < x) = 0.99$

4-47. The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.
(a) What is the probability that a sample's strength is less than 6250 Kg/cm^2?
(b) What is the probability that a sample's strength is between 5800 and 5900 Kg/cm^2?
(c) What strength is exceeded by 95% of the samples?

4-48. The tensile strength of paper is modeled by a normal distribution with a mean of 35 pounds per square inch and a standard deviation of 2 pounds per square inch.
(a) What is the probability that the strength of a sample is less than 40 lb/in^2?
(b) If the specifications require the tensile strength to exceed 30 lb/in^2, what proportion of the samples is scrapped?

4-49. The line width of for semiconductor manufacturing is assumed to be normally distributed with a mean of 0.5 micrometer and a standard deviation of 0.05 micrometer.
(a) What is the probability that a line width is greater than 0.62 micrometer?
(b) What is the probability that a line width is between 0.47 and 0.63 micrometer?
(c) The line width of 90% of samples is below what value?

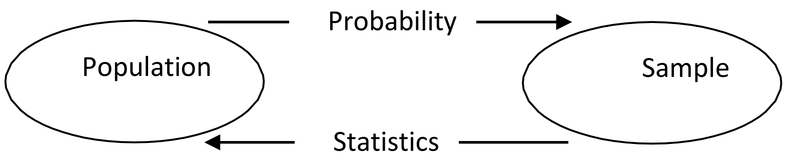
4-50. The fill volume of an automated filling machine used for filling cans of carbonated beverage is normally distributed with a mean of 12.4 fluid ounces and a standard deviation of 0.1 fluid ounce.
(a) What is the probability a fill volume is less than 12 fluid ounces?
(b) If all cans less than 12.1 or greater than 12.6 ounces are scrapped, what proportion of cans is scrapped?
(c) Determine specifications that are symmetric about the mean that include 99% of all cans.

4-51. The time it takes a cell to divide (called mitosis) is normally distributed with an average time of one hour and a standard deviation of 5 minutes.
(a) What is the probability that a cell divides in less than 45 minutes?
(b) What is the probability that it takes a cell more than 65 minutes to divide?
(c) What is the time that it takes approximately 99% of all cells to complete mitosis?

4-52. In the previous exercise, suppose that the mean of the filling operation can be adjusted easily, but the standard deviation remains at 0.1 ounce.
(a) At what value should the mean be set so that 99.9% of all cans exceed 12 ounces?
(b) At what value should the mean be set so that 99.9% of all cans exceed 12 ounces if the standard deviation can be reduced to 0.05 fluid ounce?

SAMPLING THEORY

Link between Population and Sampling:



1.0 SAMPLING DISTRIBUTIONS

Statistical inference is concerned with making **decisions** about a population based on the information contained in a random sample from that population.

For instance, the mean fill volume of a can (population) is required to be 300 mm.

An engineer takes a random sample of 25 cans and computes the sample average fill volume to be

$$\bar{x} = 298 \text{ mm}$$

The engineer will probably decide that the population mean is μ=300 mm, even though the sample mean was 298 mm because he or she knows that the sample mean is a reasonable estimate of μ and that a sample mean of 298 mm is very likely to occur, even if the true population mean is μ=300 mm.

Test values of $\bar{x}$ vary both above and below μ=300 mm.

Definition

The probability distribution of a statistic is called a **sampling distribution.**

The sampling distribution of a statistic depends on:

- The distribution of the population,
- The size of the sample, and
- The method of sample selection.

2.0 SAMPLING METHODS:

1. Random sampling
2. Systematic sampling
3. Stratified sampling
4. Multi-stage sampling

3.0 SAMPLING DISTRIBUTIONS OF MEANS

Suppose that a random sample of size n is taken from a normal population with mean μ and variance σ^2.

Now each observation in this sample, say, X_1, X_2, X_3... X_n, is a normally and independently distributed random variable with mean μ and variance σ^2

The sample mean:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

has a normal distribution with mean:

$$\mu_{\overline{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

and variance:

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n} \qquad \text{(For large N)}$$

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} \; \frac{N-n}{N-1} \qquad \text{(For small N)}$$

Theorem:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

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

EXAMPLE 1:

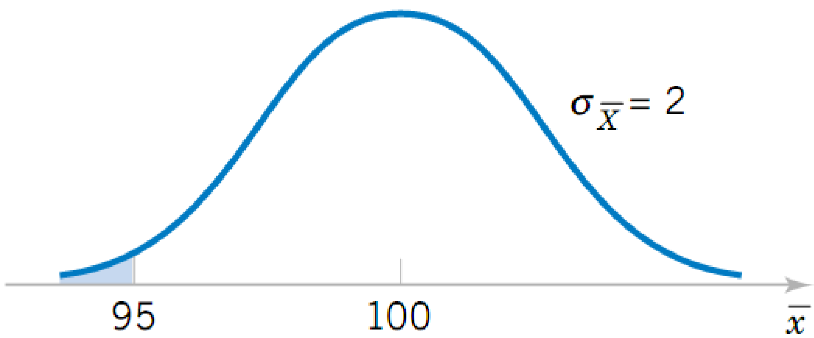
An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal.

Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\bar{x}$ is normal, with mean $\mu_{\bar{x}} = 100$ ohms and a standard deviation of:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability (shaded area) is shown in the figure below:



Standardizing the point $\bar{x}$ =95 in the Figure. We find that:

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$

3.0 SAMPLING DISTRIBUTIONS OF DIFFERENCES & SUM:

For two independent populations,

Let the first population has mean μ_1 and variance σ_1^2 and the second population has mean μ_2 and variance σ_2^2. Suppose that both populations are normally distributed. Then, we can say that the sampling distribution of ($\bar{x}_1$ - $\bar{x}_2$) is normal with mean:

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$$

And variance

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\bar{x}_1$ and $\bar{x}_2$ are the sample means of two independent random samples of sizes $n1$ and n_2 from these populations, then the sampling distribution is:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

with condition $n1, n2 \geq 30$

EXAMPLE 2:

The effective life of a component used in an engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution.

The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of n_1= 16 components is selected from the "old" process and a random sample of n_2=25 components is selected from the "improved" process.

What is the probability that the difference in the two sample means $\bar{x}_2 - \bar{x}_1$ is at least 25 hours? *Assume that the old and improved processes can be regarded as independent populations.*

the distribution of $\bar{x}_1$ is normal with mean μ_1= 5000 hours and standard deviation

$$\sigma_1/\sqrt{n_1} = 40/\sqrt{16} = 10 \text{ hours},$$

and the distribution of $\bar{x}_2$ is normal with mean μ_2= 5050 hours and standard deviation

$$\sigma_2/\sqrt{n_2} = 30/\sqrt{25} = 6 \text{ hours},$$

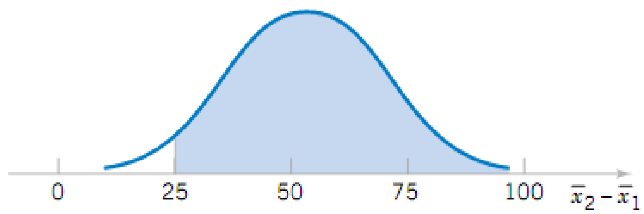
Now the distribution of $\bar{x}_2 - \bar{x}_1$ is normal with mean

$$\mu_2 - \mu_1 = 5050 – 5000 = 50 \text{ hours}$$

and variance

$\sigma_2^2/n_2^2 + \sigma_1^2/n_1^2 = 6^2 + 10^2 = 136$ hours^2.

This sampling distribution is shown in the Figure below:



The sampling distribution of in Example 2

The probability that $\bar{x}_2 - \bar{x}_1 \geq 25$ hours is the shaded portion of the normal distribution in this figure.

So,

$$z = \frac{25 - 50}{\sqrt{136}} = -2.14$$

and we find that:

$$P(\bar{X}_2 - \bar{X}_1 \geq 25) = P(Z \geq -2.14)$$
$$= 0.9838$$

EXERCISES:

1. PVC pipe is manufactured with a mean diameter of 1.01 inch and a standard deviation of 0.003 inch. Find the probability that a random sample of n = 9 sections of pipe will have a sample mean diameter greater than 1.009 inch and less than 1.012 inch.

2. A synthetic fiber used in manufacturing carpet has tensile strength that is normally distributed with mean 75.5 psi and standard deviation 3.5 psi. Find the probability that a random sample of n= 6 fiber specimens will have sample mean tensile strength that exceeds 75.75 psi.

3. A random sample of size n_1= 16 is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size n_2= 9 is taken from another normal population with mean 70 and standard deviation 12. Let $\bar{x}_1$ and $\bar{x}_2$ be the two sample means. Find

a) The probability that $\bar{x}_1 - \bar{x}_2$ exceeds 4
a) (b) The probability that $3.5 \leq \bar{x}_1 - \bar{x}_2 \leq 5.5$

4. The elasticity of a polymer is affected by the concentration of a reactant. When low concentration is used, the true mean elasticity is 55, and when high concentration is used the mean elasticity is 60. The standard deviation of elasticity is 4, regardless of concentration. If two random samples of size 16 are taken, find the probability that $\bar{x}_{high} - \bar{x}_{low} \geq 2$.

REGRESSION & CORRELATION

Many problems in engineering and science involve exploring the relationships between two or more variables. **Regression analysis** is a statistical technique that is very useful for these types of problems.

For example, in a chemical process, suppose that the yield of the product is related to the process-operating temperature. Regression analysis can be used to build a model to predict yield at a given temperature level. This model can also be used for process optimization, such as finding the level of temperature that maximizes yield, or for process control purposes.

Table ·1 Oxygen and Hydrocarbon Levels

Observation Number	Hydrocarbon Level x(%)	Purity y(%)
1	0.99	90.01
2	1.02	89.05
3	1.15	91.43
4	1.29	93.74
5	1.46	96.73
6	1.36	94.45
7	0.87	87.59
8	1.23	91.77
9	1.55	99.42
10	1.40	93.65
11	1.19	93.54
12	1.15	92.52
13	0.98	90.56
14	1.01	89.54
15	1.11	89.85
16	1.20	90.39
17	1.26	93.25
18	1.32	93.41
19	1.43	94.98
20	0.95	87.33

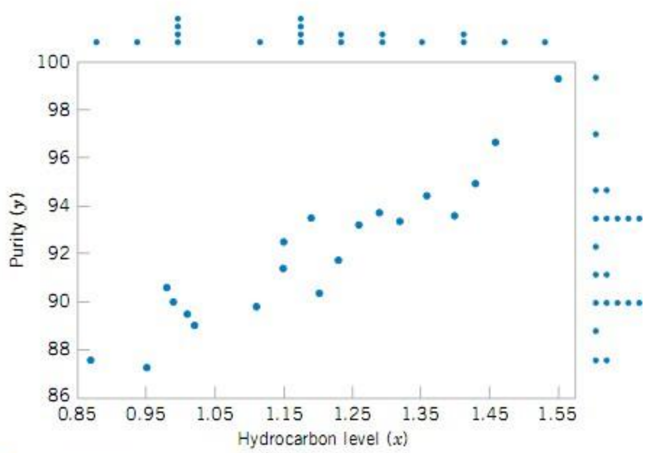


Figure ·1 Scatter diagram of oxygen purity versus hydrocarbon level from Table 11-1.

1.0 SIMPLE LINEAR REGRESSION

The case of **simple linear regression** considers a single **predictor** independent variable x and a dependent or **response variable** Y. Suppose that the true relationship between Y and x is a straight line and that the observation Y at each level of x is a random variable.

The expected value of Y, can be described by the model:

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where the intercept β_0 and the slope β_1 are unknown regression coefficients.
ε is a random error with mean zero

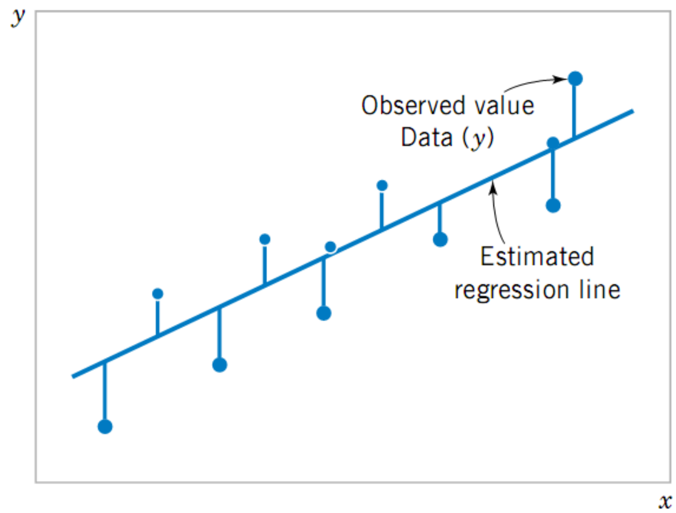


Figure 2: Deviation of data from the estimated regression model

We call this criterion for estimating the regression coefficients the **method of least squares. W**e may express the n observations in the sample as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

and the sum of the squares of the deviations of the observations from the true regression line is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

The least squares estimators of β_0 and β_1, must satisfy

$$\left.\frac{\partial L}{\partial \beta_0}\right| = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\left.\frac{\partial L}{\partial \beta_1}\right| = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Simplifying these two equations yields:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$

The solution to the normal equations results in the least squares estimators β_0 and β_1:

$$\beta_0 = \frac{\sum y_i \sum x_i^{\,2} - \sum x_i \sum x_i\ y_i}{n \sum x_i^{\,2} - (\sum x_i)^2}$$

$$\beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^{\,2} - (\sum x_i)^2}$$

Note that each pair of observations satisfies the relationship:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad i = 1, 2, \ldots\ldots\ldots, n$$

where $e_i = y_i - \hat{y}_i$ is called the **residual.** The residual describes the error in the fit of the model to the *i*th observation y_i.

Let:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

and

$$S_{xy} = \sum_{i=1}^{n} y_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

EXAMPLE 1: We will fit a simple linear regression model to the oxygen purity data in Table 1. The following quantities may be computed:

$$n = 20 \quad \sum_{i=1}^{20} x_i = 23.92 \quad \sum_{i=1}^{20} y_i = 1{,}843.21 \quad \bar{x} = 1.1960 \quad \bar{y} = 92.1605$$

$$\sum_{i=1}^{20} y_i^2 = 170{,}044.5321 \quad \sum_{i=1}^{20} x_i^2 = 29.2892 \quad \sum_{i=1}^{20} x_i y_i = 2{,}214.6566$$

$$\beta_0 = \frac{\sum y_i \sum x_i^{\,2} - \sum x_i \sum x_i\ y_i}{n \sum x_i^{\,2} - (\sum x_i)^2}$$

$$\beta_1 = \frac{n \sum x_i\, y_i - \sum x_i \sum y_i}{n \sum x_i^{\,2} - (\sum x_i)^2}$$

$$\beta_0 = \frac{1843.21 * 29.2892 - 23.92 * 2214.6566}{20 * 29.2892 - (23.92)^2}$$

$$\beta_1 = \frac{20 * 2214.6566 - 23.92 * 1843.21}{20 * 29.2892 - (23.92)^2}$$

$\beta_0 = 74.283$

$\beta_1 = 14.947$

As a double check:

$$\bar{y} =? \beta_0 + \beta_1 \bar{x}$$

So,

92.160 =? 74.283 + 14.947 * 1.196 if yes then continue

If not then re-check your calculations

The fitted simple linear regression model (with the coefficients reported to three decimal places) is:

$$\hat{y} = 74.283 + 14.947x$$

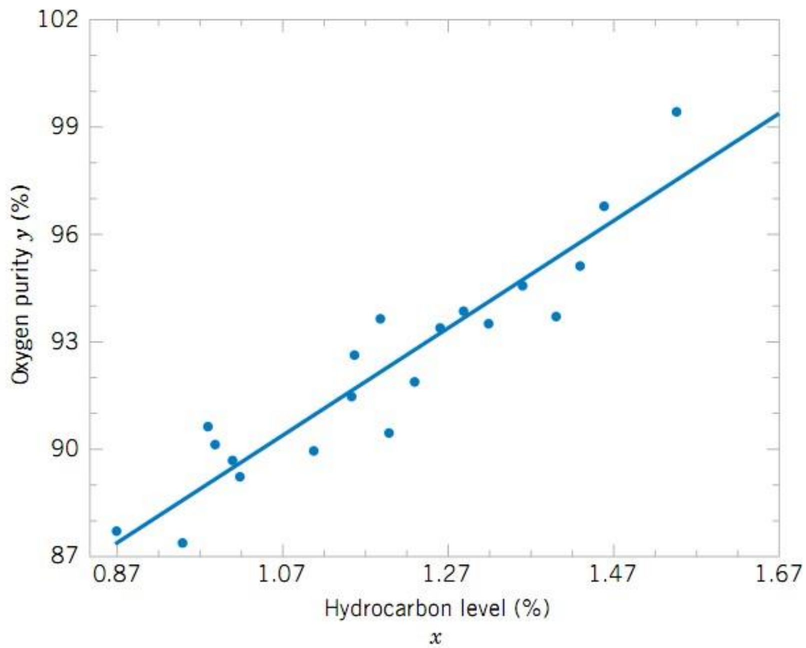


Figure 3 Scatter plot of oxygen purity y versus hydrocarbon level x and regression model $\hat{y} = 74.20 + 14.97x$.

Using the regression model of Example 1, we would predict oxygen purity of $\hat{y}$ = 89.23% when the hydrocarbon level is x = 1.00%.

The purity 89.23% may be interpreted as an estimate of the true population mean purity when x=1.00%, or as an estimate of a new observation when x = 1.00%. These estimates are, of course, subject to error; that is, it is unlikely that a future observation on purity would be exactly 89.23% when the hydrocarbon level is 1.00%. In subsequent sections we will see how to use confidence intervals and prediction intervals to describe the error in estimation from a regression model.

Estimating σ^2

There is actually another unknown parameter in our regression model, σ^2 (the variance of the error term ϵ). The residuals $e_i = y_i - \hat{y}_i$ are used to obtain an estimate of σ^2. The sum of squares of the residuals, often called the **error sum of squares,** is

$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{\sigma}^2 = \frac{SS_E}{n-2}$$

2.0 Correlation

A measure of the linear relationship between two numerical variables is provided by the correlation coefficient. A correlation coefficient takes a value between -1 (perfect negative correlation) to +1 (perfect positive correlation) with zero representing no correlation.

H.W No. 1:

The accompanying data was taken from published paper. The independent variable is SO_2 deposition rate ($mg/m^2/day$) and the dependent variable is steel weight loss (gm/m^2).

x: 14, 18, 40, 43, 45, 112

y: 280, 350, 470, 500, 560, 1200

a) Construct a scatter plot. Dose the simple linear regression model appear to be reasonable in this situation?
b) Calculate the equation of the estimated regression line?
c) Estimate the standard deviation of observation about the true regression line.

H.W No. 2:

The accompanying data resulted from a study carried out to examine the relationship between a measure of the corrosion of reinforcement (y) and the concentration of the corrosion inhibitor solution in concrete pores (x, in ppm):

x: 2.5, 5.03, 7.6, 11.6, 13, 19.6, 26.2, 33, 40, 50, 55

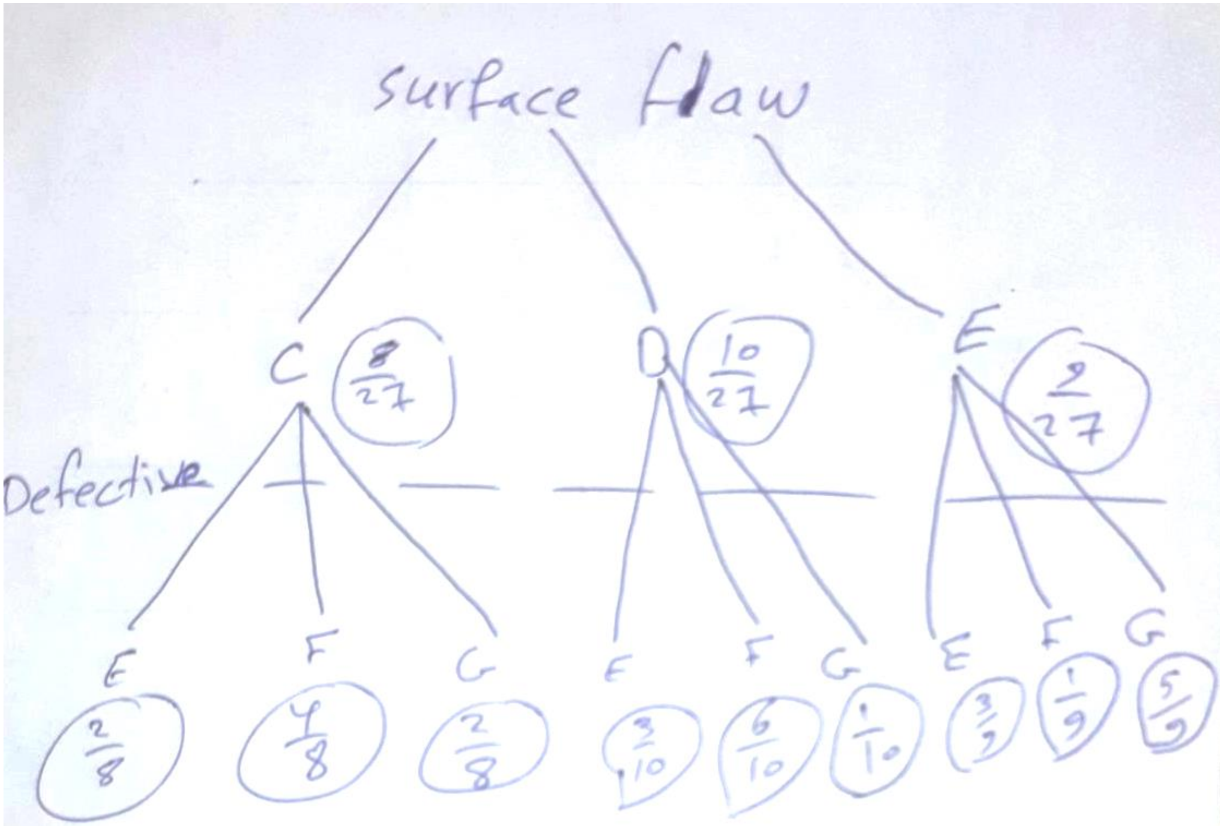
y: 7.68, 6.95, 6.3, 5.75, 5.01, 1.43, 0.93, 0.72, 0.68, 0.65, 0.56

a. Construct a scatter plot of the data. Dose the simple linear regression appear to be logical?
b. Calculate the equation of the estimated regression line, use it to predict the value of the corrosion rate that would be observed for a concentration of 33 ppm, and calculate corresponding residual.
c. Estimate the standard deviation of observation about the true regression line.

Surface flaw

event B \ event A	C	D	E
E	2	3	3
F	4	6	1
G	2	1	5

event B \ event A	C	D	E	Total
E	2	3	3	8
F	4	6	1	11
G	2	1	5	8
Total	8	10	9	27



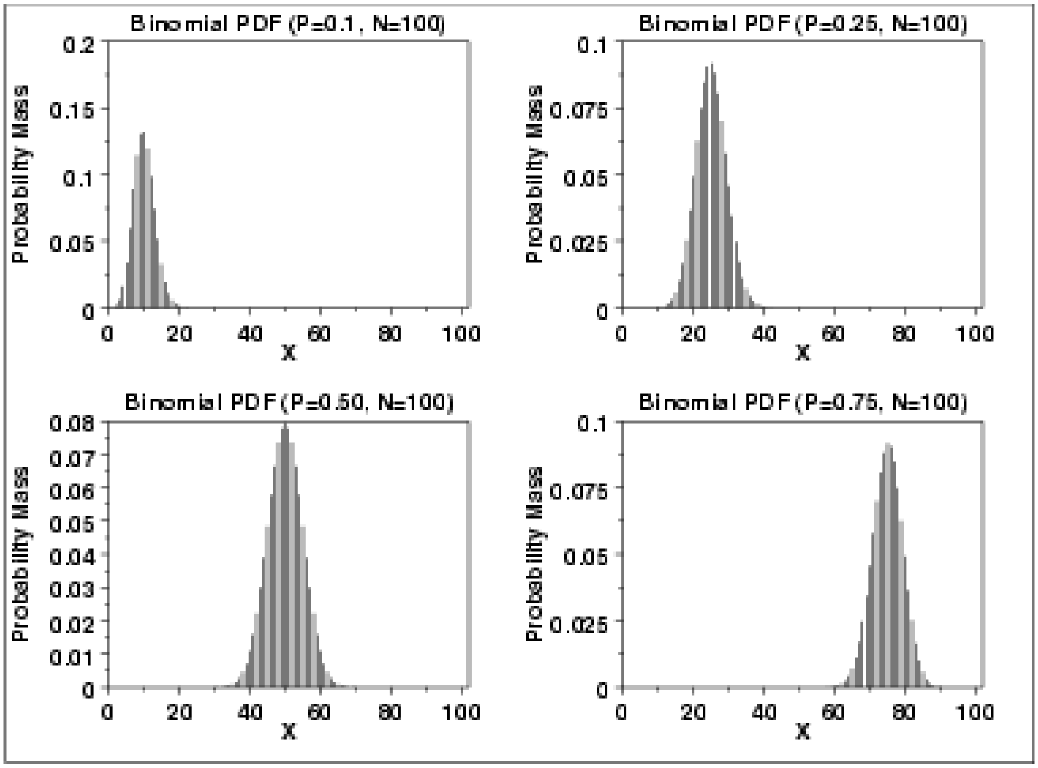
Probability mass function

In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n independent Bernoulli trials is given by the probability mass function:

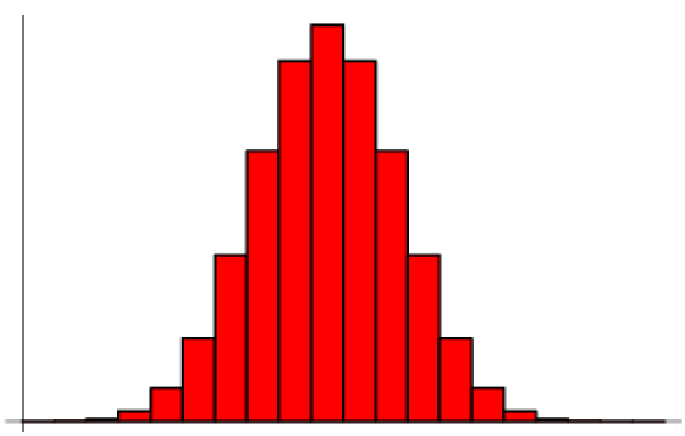
$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, \ldots, n$, where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



The formula can be understood as follows: k successes occur with probability p^k and $n - k$ failures occur with probability $(1 - p)^{n-k}$. However, the k successes can occur anywhere among the n trials, and there are $\binom{n}{k}$ different ways of distributing k successes in a sequence of n trials.



The binomial distribution is implemented in the Wolfram Language as Binomial Distribution [n, p].

The probability of obtaining more successes than the n observed in a binomial distribution is

$$P = \sum_{k=n+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} = I_p (n+1, N-n),$$

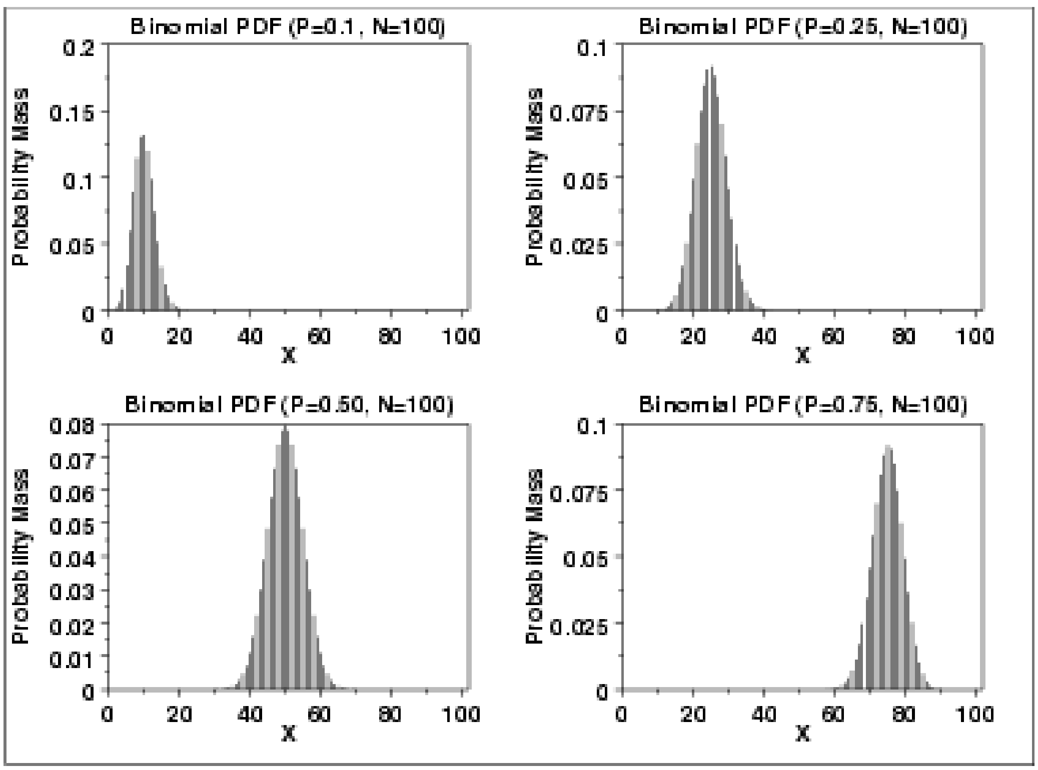
Probability mass function

In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n independent Bernoulli trials is given by the probability mass function:

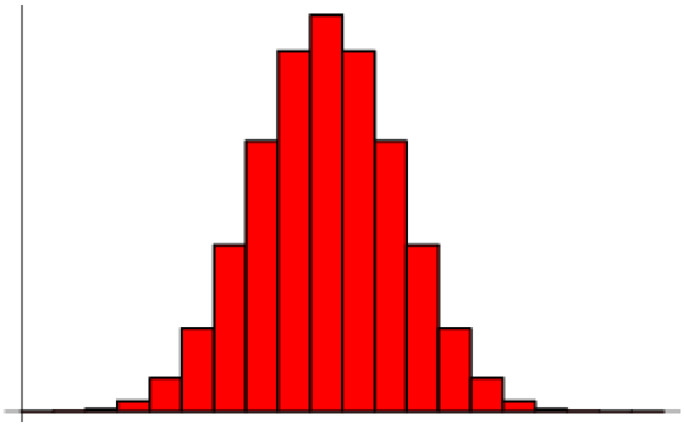
$$f(k,n,p) = \Pr(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, ..., n$, where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



The formula can be understood as follows: k successes occur with probability p^k and $n - k$ failures occur with probability $(1 - p)^{n - k}$. However, the k successes can occur anywhere among the n trials, and there are $\binom{n}{k}$ different ways of distributing k successes in a sequence of n trials.



The binomial distribution is implemented in the Wolfram Language as Binomial Distribution [n, p].

The probability of obtaining more successes than the n observed in a binomial distribution is

$$P = \sum_{k=n+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} = I_p (n+1, N-n),$$