Probability and Statistics for Engineers using MATLAB and Octave

Osama Hosameldeen
Associate Professor
Taibah University

Jan 2020
First Edition

Probability and Statistics for Engineers using MATLAB and Octave

ISBN (XXXXXXXXXXXXX)

كلمة شكر

أقدم الشكر لجميع أعضاء هيئة التدريس بقسم علوم الحاسبات بكلية علوم وهندسة الحاسبات بينبع. كما أتقدم بالشكر الخاص للطلاب الآتي أسماؤهم لما قدموه من مساعدة فعالة في كتابة محتوى المنهج على الحاسب الآلي

- محمد كعكي
- خالد العنزي
- عبد المحسن العنزي
- وليد بدر

تحياتي
أسامة حسام الدين
أكتوبر 2017م

Contents

Preface

Replace this wording with your own.

Phone: 054321000
Email:mohandesosama@yahoo.com
Website: go to my website, osama-hosam.blogspot.com

Syllabus

STAT 301 - Probability & Statistics for Engineers

Course Code	Course Title	Credit Hours	Lecture Hours	Labs Hours	Pre-requisite
STAT301	**Probability & Statistics for Engineers**	3	2	2	

Course Objectives

Having successfully completed this course, the student will be able to:

- Examine statistical data and methods of collecting, representing and organizing it.
- Identify measures of location and variability.
- Determine the basic concepts of probabilities.
- Apply the random variables and probability distributions and its properties
- Apply different types of distributions.
- Compute probabilities from some discrete and continuous distributions .
- Carry out confidence intervals calculations.
- Perform statistical hypothesis tests.
- Recognize methods of how to compute relationships between variables.
- Use the available statistical software to improve knowledge about probability.
- Improve problem solving skill, criticism thinking and group working.

Course Description

This course aims to make the students aware of the basic concepts of descriptive and inferential statistics and discusses the following topics: data types and data sources, presenting data in charts and tables, measures of location and variability, basic concept of probabilities, random variables and probability distributions, mathematical expectation, some discrete and continuous distributions, simple linear regression mode and correlation. The course also gives a general overview and foundations for statistical inference. The practical session aims to make students able to solve mathematical exercises using mathematical software such as SPSS, Minitab and R.

Course Outline

- Introduction to Statistics & data analysis
- Probability
- Random Variables and Probability Distributions.
- Sampling distribution of Means and the Central Limit Theorem,
- One and two Sample Estimation Problems
- Basic concepts, Confidence interval for the mean and the proportion, Confidence interval for the difference of means and the, Difference of proportions, Estimating the variance.
- Test of Hypotheses
- Basic concepts in testing statistical hypotheses

Lab

The practical session aims to make students able to solve mathematical exercises using mathematical software such as SPSS, Minitab and R

Textbooks:

- Ronald E. Walpole & others, "Probability and Statistics for Engineers and Scientists ", 9th ed., Pearson Prentice Hall, 2012.
- David Stirzaker, "Elementary Probability", 2nd ed., Cambridge University Press,2003 .
- David W. Stockburger, "Introductory statistics: concepts, models and applications", 3rd ed., Atomic Dog Publishing, 2007.

Chapter 1
Data representation

Statistics is defined as a collection of methods for collecting, organizing and representing data There two types of statistics:

1- Descriptive statistics : which describe data or represent it, data can be described using frequency count, average, modes, median and standard deviation
2- Inferential statistics: trying to guess the population parameters using the given sample . it performs hypothesis testing and mares prediction .

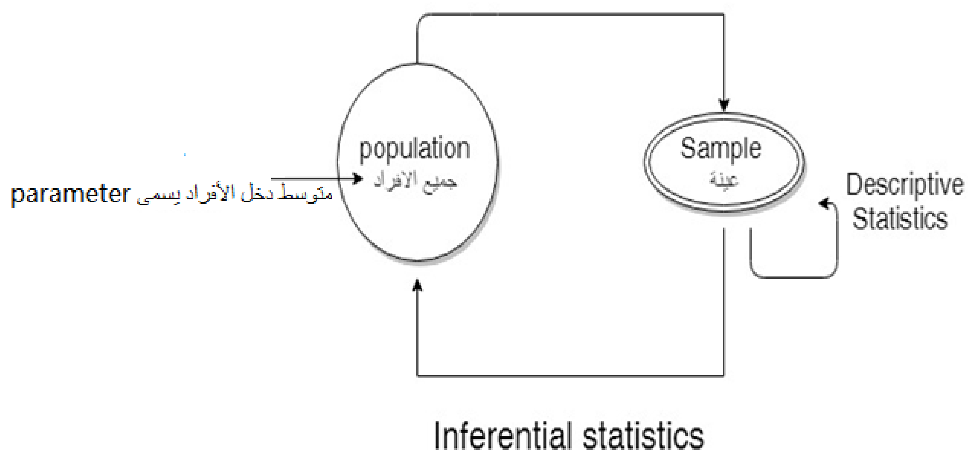


Figure 1.1: Scope of Statistics

Figure 1.1 shows the scope of statistics, which includes population, sample, parameters and statistic.

The population : the collection all elements (scores, persons, measurements, …. Etc.) to be studied

Sample : subset of the population

Parameter : characteristic or measure obtained from the population .

Statistic : is a characteristic or measure obtained from the sample .

In the field of statistics we have a collection or raw data, the data is collected from measurements, experiment results. As shown in Figure 1.2, The experiment study started by collecting data either from sensors (weather, salaries) or a result from another experiment. The data collected at this stage is called raw data. If the data is huge to commit study we have to take a sample of the data.

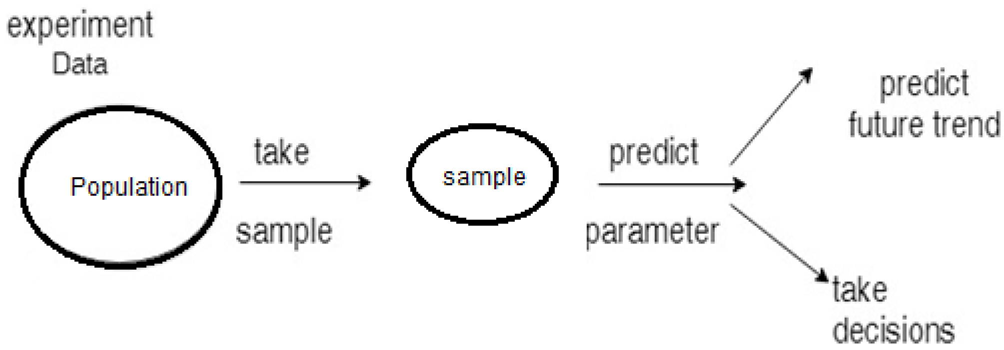


Figure 1.2: stages of making statistical experiments

1.1 Types of Data

The data can be qualitative such as “female”, “male”, or “Egyptian”, “Saudi” etc. generally qualitative data can’t be represented with numbers. Quantitative data is the data that can be represented numerically such as 0, 5, 21, 300, 0.12, etc. Quantitative data can be discrete or continuous (Bluman 2004)

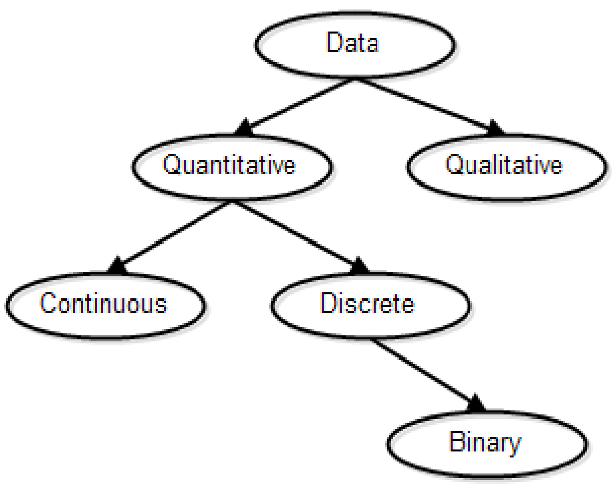


Figure 1.3: Types of data

Discrete Data : The data that its value is undividable such as

- Number of children in a family
- Number of students
- Number of crimes
- Number of cars sold in a day

Generally , we wouldn't expect to find 2.5 children or $\frac{1}{2}$ of a car sold in aday.

Binary date is the data that can have one of two values, such as of 1 or 0, Binary data is discrete data type.

Continuous Data **:** the observation that can take any value within finite or infinite interval real numbers, in other words it contains fractions.

Examples are :

- Weight
- Height
- Time to run 500 meters
- Age

1.2 Organization of Data

The statistician must organize data into meaningful way and present it so readers Can understand the distribution of data .

1.2.1 – Frequency table

Counting the number of repetition of specific value in the data.

Example : draw frequency table for the.

1 2 6 7 2 6 5 7 1 5 5 6 5 5 2

Solution :

Date	1	6	5	2	7
frequency	2	3	5	3	2

1.2.2 – Grouped Frequency table

Counting the frequency of repeating a value in specific range

Example : Draw the grouped frequency table of the following data with 5 classes .

12,22,18,9,25,31,28,19,22,27,32,14

Solution :

- Find the min and max values min 9, max 32 .
- wide the range over the class number
- $\frac{32-9}{5} = \frac{range}{\#\, classes} = \frac{23}{5} = 4.6$
- get the ceiling of the value width of class = $ciel\left(\frac{range}{\#\, classes}\right)$
- set the classes aB show in in the table each step increase 5

Class	**Frequency**
9 – 13	2
14 – 18	2
19 – 23	3
24 – 28	3

29 – 33	2

1.2.3 – Histogram

Displaying the frequency table with touching and non-overlapping bars .

Example : Use histogram to represent the above grouped frequency table .

Solution :

# Class	Frequency
9 – 13	2
14 – 18	2
19 – 23	3
24 – 28	3
29 – 33	2

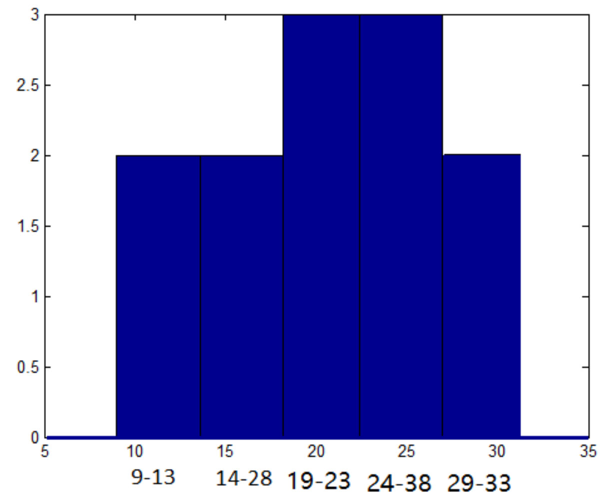


Figure 1.4 Histogram

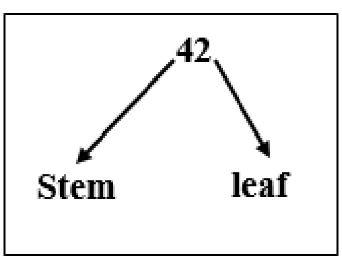
Frequency histogram with MATLAB use the following commands

```
>> a=[12,22,18,9,25,31,28,19,22,27,32,14]
>> hist ( a,5 )
```

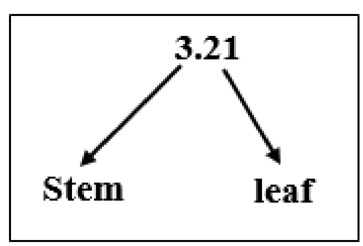
1.2.4 – Stem and leaf plot

Divide the data into two parts, STEM OF leaf list the Stems in Column and leaf's in rows .

Discrete



Continuous



<u>Example</u> : Draw Stem of leaf for the following Date .

90 70 70 70 75 70 65 68 60

74 70 95 75 70 68 65 40 65

Answer :

4	0
5	
6	055588
7	000000455
8	
9	05

Notice : 9 05 means there are data value, 90 and 95 in the data .

Example: Let's consider a random sample of 20 concentrations of calcium carbonate (CaCO3) in milligrams per liter.

130.8	129.9	131.5	131.2	129.5	132.7	131.5	127.8	133.7
132.2	134.8	131.7	133.9	129.8	131.4	128.8	132.7	132.8
131.4	131.3							

Create a stem-and-leaf plot of the data.

Solution: The following figure shows the resulting stem and leaf plot for the above continuous data

127	8
128	8
129	589
130	8
131	2344557
132	2778
133	79
134	8

1.3 Central tendency

The data by its nature if often tend to be concentrated around the center of data, the measure of this type is called measure of location such as the following measures.

- Mean
- Median
- Mode

1.3.1 : Arithmetic mean

The mean of measurements is the sum of all measurements divided by number of measurement

$mean\ \overline{x} = \frac{\sum xi}{n}$

n : is the number of measurements

xi : is the ith measurement

Note: notice that μ is the population mean while $\overline{x}$ is the sample mean.

Example: find the mean of following sample data, 2, 9, 11, 5, 6

Solution: $\overline{x} = \frac{\sum xi}{n} = \frac{2+9+11+5+6}{5} = \frac{33}{5} = 6.6$.

The following MATLAB command can be used to get the arithmetic mean

```
>> mean(a)
```

1.3.2 : Median

The median of a set of measurements is the middle measurement after sorting the data

The position of median value is $[\frac{n+1}{2}]$

Example: find the median of 2,4,9,8,6,5,3

Solution:

1- Sort the numbers 2,3,4,5,6,8,9
2- Position of median $= \frac{7+1}{2} = 4$

So the median is the fourth number which is 5. The median is obtained directly if the number of elements in the data (n) is odd.

Example: Find the median of 2,4,9,8,6,5

This an even problem, $n = 6$ so the median position will be between the third and fourth positions.

1- Sort the numbers 2,4,5,6,8,9
2- Position $= \frac{n+1}{2} = \frac{6+1}{2} = 3.5$

$$2,4,5, position, 6 , 8 , 6$$

The median is $= \frac{5+6}{2} = \frac{11}{2} = 5.5$

1.3.3 mode:

Mode measures the data which occurs most frequently

Example : the set $2, 4, 9, 8, 8 , 5, 3$ has mode of 8

Example : the set $2, 2, 9, 8, 8, 5, 3$ has mode of 2, 8 (bimodal) .

Example : the set $2, 9, 3, 6$ has mode of 2 the first number .

1.3.4 Removing Outliers (إزالة القيم الشاذة)

The median filter: the median filter collects 3 by 3 samples of the data and finds their median.

Example: filter the following data with median filter 2 , 4 , 10 , 5 , 3 , notice that 10 is an outlier

Solution:

find the median of (2 4 10)

$$sort \Rightarrow 2, 4, 10 \Rightarrow median = \underline{4}$$

find the median of (4 , 10 , 5)

$$sort \Rightarrow 4, 5, 10 \Rightarrow median = \underline{5}$$

find the median of (10 , 5 , 3)

$$sort \Rightarrow 3\,,5\,,10 \Rightarrow median = \underline{5}$$

The resulting signal is $2, 4, 5, 5, 3$, notice that the outlier value (10) is filtered out.

Note that median filter removes sharp notches (outliers)

Trimming: it is obtained by removing the highest and the lowest part of the data. For this purpose the data must be sorted first. For example 10% trimmed data is obtained by removing 10% of the data from both highest and lowest values.

Example: The given data is measurements taken as an output from experiment.

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

Assuming the measurements are a random sample, compute the 20% trimmed data for the data set:

Solution: sample size is n=15, 20% of 15= 0.20 * 15=3, So we remove the first and last 3 values after sorting them,

The sorted data is : 2.5, 2.8, 2.8, 2.9, 3.0, 3.3, 3.4, 3.6, 3.7, 4.0, 4.4, 4.8, 4.8, 5.2, 5.6

The trimmed data is: 2.9, 3.0, 3.3, 3.4, 3.6, 3.7, 4.0, 4.4, 4.8

MATLAB (A)

A.1: Basic operations .

Basic arithmetic and logic operations can be open done with MATLAB Such as

$\gg \mathbf{16/10}$ $\gg \mathbf{30 * 5}$ $\gg \mathbf{30 - 300}$ $\gg \mathbf{500 + 7}$ $\gg \mathbf{30 * (500 + 7)}$

In addition, a variable Can be defined. Then used such as

$\gg \boldsymbol{a} = \mathbf{10}$;

$\gg \boldsymbol{z} = \boldsymbol{a} * \mathbf{30}$

$\gg \boldsymbol{q} = \boldsymbol{z} * (\boldsymbol{a} + \mathbf{5})$

You can ignore the result (not display it) by adding ";" to the end of the command such as

$\gg \mathbf{16/10}$;

A.2:Vector representation :

Vector can hold 1D data such as voice signal or data transferred on a network, etc.

Vector can be represented as

$\gg \boldsymbol{a} = [\,\mathbf{5\ \ 10\ \ 15\ \ 30\ \ 55\ \ 16\ \ 27\ \ 300\ \ 5}\,]$;

Or

$\gg \boldsymbol{a} = [\,\mathbf{5\ \ 10\ \ 15\ \ 30\ \ 55\ \ 16\ \ 27\ \ 300\ \ 5}\,]$;

Regular vector Can be created as

$\gg \boldsymbol{s} = [\mathbf{1 : 10}]$

The result must be **1 2 3 4 5 6 7 8 9 10**
The vector is increased by 1, you can change .
The default incremental step to 2 such that
$\gg s = [\mathbf{1} : \mathbf{2} : \mathbf{10}\,]$
The result should be **1 3 5 7 9**
Random vector can be created with the command **rand.** it creates random values between 0 and 1 this default can be changed to create random values between 0 and 100 as follow
$\gg \mathbf{ceil}\,(\,\mathbf{rand}\,(\,\mathbf{1}, \mathbf{10}\,) * \mathbf{100}\,)$
Notice that **rand** command creates single random number
$\gg$ **rand**
While a vector of random numbers between 0 and 1 and the length of the vector is 10 Can be created with
$\gg \mathbf{rand}\,(\,\mathbf{1}\,, \mathbf{10}\,)$
The floating numbers can be removed with the function **ceil** for example *ceil (0.11) = 1* and *ceil (9.2) = 10*
the vector can be plotted and represent as follow:
$\gg \mathbf{a} = \mathbf{ceil}\,(\mathbf{rand}\,(\,\mathbf{1}, \mathbf{10}\,) * \mathbf{100}\,)$
$\gg \mathbf{plot}\,(\mathbf{a})$
or $\gg \mathbf{figure}, \mathbf{plot}\,(\mathbf{a})$

The following command also creates random numbers with uniform distribution, notice we used **fix** instead of **ceil.**
$\gg \mathbf{a} = \mathbf{fix}(\mathbf{rand}(\mathbf{1}, \mathbf{10}) * \mathbf{100})$
>> hist(a,6)
The following command can create random numbers with normal distribution.
>> a=fix(normrnd(70,10,1,20)
>>hist(a,6)
Where 70 is the center of data or the mean, 10 is the fluctuation range or the standard deviation, and 20 is the number of data points. Notice that the second figure displays the bell shaped distribution or histogram of heights frequencies.

A.3:Mean
The mean is the center of data which data values fluctuates around it, mean can be easily found with the **mean** MATLAB command

$\gg$ **a = fix (rand (1,10) * 100)**

$\gg$ **m = mean (a)**

For a signal a , it can be filtered using the following MATLAB command.

$\gg$ **sig 1 = conv(a, [0.333 0.333 0.333],′ same′)**

$\gg$ **plot (1 : 10, a , 1 : 10 , sig1)**

A.4:median

The median is the number located at the center of data, in the same manner as mean, median can be obtained as :

$\gg$ **a = fix(rand (1 , 10) * 100)**

$\gg$ **med = median (a)**

A signal can be filtered with median filter using the following Command.

$\gg$ **sig2 = medfilt1(a, 3)**

$\gg$ **plot (1 : 10 , a , 1: 10, sig2)**

A.5 mode:

Mode can be obtained with the **mode** MATLAB Command.

$\gg$ **a = fix (rand (1 , 10) * 100)**

$\gg$ **mode (a)**

1.4: Measures of Dispersion (مقاييس التشتت في البيانات)

Mean and median measure tells useful information about central tendency, but they lack the ability to describe data variance. for example, data may have the same mean but totally different variance.

Example 1 :

Calculate mean for the following two sets of data .

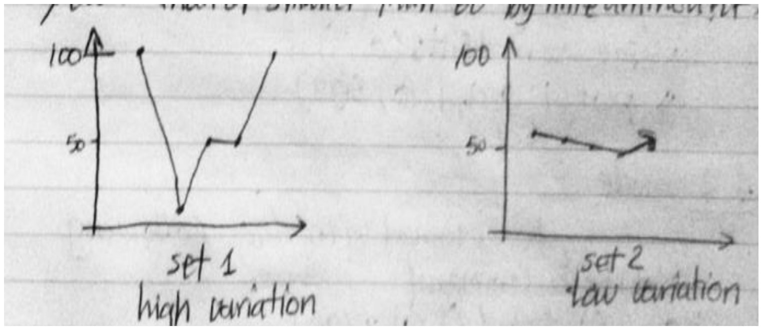
Set 1 : 100 10 50 50 90

Set 2: 62 61 54 58 60

Mean of set 1 = $\frac{100+10+50+90}{5} = \frac{300}{5} = 60$

Mean of set 2 = $\frac{62+61+59+58+60}{5} = \frac{300}{5} = 60$

Set 1 & set 2 have the same mean which is 60. But they have different variance . set 1 has very large values such as 100 & 90 . and very small values such as 10 . set2 has little variance since the values is greater than or smaller than 60 by little amount .



Set 1 gets far away from the central or mean value, but set 2 fluctuate around the center value. There are many methods that can be used to measure the variations, each method has its own advantages and disadvantages.

One of these methods is the range.

1. 4.1: Range

range is simply the difference between the largest and smallest values in the dataset .

$$\textbf{Range}\ (\,\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4}\,) = \textbf{max}(\,\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4}\,) - \textbf{min}(\,\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4}\,)$$

<u>Example</u> : find the ranges of the data in example 1

Set 1 : 100 10 50 50 90

Set 2: 62 61 59 58 60

Solation :

$$\min(100\,,10\,,50\,,50\,,90 = 10$$

$$\max(\,100\,,10\,,50\,,50\,,90\,) = 100$$

Set1 range = 100 – 10 = 90

$$\min(62\,,61\,,59\,,58\,,60 = 58$$

$$\max(\,62\,,61\,,59\,,58\,,60\,) = 62$$

Set2 range = 62 – 58 = 4

Set1 has high rang range which represents high variance and set 2 had low range (4) which represents low fluctuations.

Hint Range is a good measure for variability, but it is very weak if the data had outliers (قيم شاذة) the outliers affects the range but range in this case , will not express real variability.

القيم الشاذة تؤثر في المدى. على سبيل المثال إذا اردنا معرفة مدى الرواتب في مصنع من المصانع، وكان حوالي 100 عامل في المصنع يتقاضون رواتب بين 3 آلاف ريال وأربعة آلاف ريال. بذلك سيكون المدى بسيط وهو ألف ريال وهو مدى منطقي. أما إذا وجد مدير للمصنع يتقاضى 50 الف ريال. وهو الوحيد في المصنع الذي يتقاضى هذا الراتب. فإذا حسبنا المدى سيتأثر بالقيمة الشاذة وهي قيمة الراتب 50 الف. وسيصبح 47 ألف ريال. وهو غير منطقي حيث أن التذبذب الخاص بالبيانات ليس مداه 50 ألف.

1 . 4 . 2 : Variance and Standard Deviation

if (x1 , x2 , x3) is the data, then the variance an be calculated by

$$S^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2}{n-1}$$

Where $\overline{x} = \frac{x_1 + x_2 + x_3}{2}$ which is the data average or mean. and n is the number of data values which is 3 in this case .

The sample variance is S^2

The population variance is σ^2

The variance is simply the average value of the distances between each value in the dataset and data-set central value. or the average value or the fluctuations variance is expressed mathematically as:

$$S^2 = \frac{1}{n-1} \sum_{1=1}^{n} (x_i - \overline{x})^2$$

The variance is fine measure for describing variability, but the resulting value is squared, for example if the data set is heights of persons, the resulting values are not reflecting heights, it represent the square of the height.

Example: find the variance marks of the following sets which represents the degrees of STAT class students.

Set 1 : 100 10 50 50 90

Set 2 : 62 61 59 58 60

Set1 variance ($\overline{x} = 60$)

$$s_1^2 = \frac{1}{5-1} [(100-60)^2 + (10-60)^2 + (50-60)^2 + (50-60)^2 + (90-60)^2]$$

$$s_1^2 = 1300$$

Set2 variance ($\overline{x} = 60$)

$$s_1^2 = \frac{1}{5-1} [(62-60)^2 + (61-60)^2 + (59-60)^2 + (58-60)^2 + (60-60)^2]$$

$$s_2^1 = \frac{4+1+1+4+0}{4} = \frac{10}{4} = 2.5$$

So, set1 has higher variances (1300) them set2 (2.5)

Drawback: The problem is that 1300 is not a student mark, it doesn't reflect the same range of the data set , no one can get 1300 in the STAT Class !!

To solve the above problem, the standard deviation is used, the standard divination is simply the square root of the variance .

$$s = \sqrt{s^2}$$

For set 1 $s = \sqrt{1300} = 36.056$

For set 2 $s = \sqrt{2.5} = 1.5811$

Now we can easily say that the deviation of date from the center is 36 marks .

Another formula of variance

$$S^2 = \frac{\sum_{1=1}^{n} x_i{}^2 - n\bar{x}^2}{n-1}$$

So set1 variance $(\overline{x} = 60)$

$$s^2 = \frac{((100)^2 + (10)^2 + (50)^2 + (50)^2 + (90)^2 - (5 \times (60)^2)}{5-1}$$

$$= \frac{10000 + 100 + 2500 + 2500 + 8100 - (5 \times 3600)}{4}$$

$$= \frac{5200}{4} = 1300$$

And set2 variance

$$s^2 = \frac{((62)^2 + (61)^2 + (59)^2 + (58)^2 + (60)^2 - (5 \times (60)^2)}{5-41}$$

$$= \frac{10}{4} = 2.5$$

Exercise: Compute the variance and standard deviation of the following datasets

Set 1 10 21 33 53 54

Set 2 34 36 35 33 34

Explain the difference between the results, justify your answer.

1. 4. 3: Box Plot

To draw box plot, the inter – quartile range must be extracted from the data. The range of the middle of 50% score is the interquartile range. or

$$IRQ = Q3 - Q1$$

Where Q3 is the median of the right half of the data. and Q1 is the median of the left half of the data.

Example 1: Given the following dataset calculate the interquartile-range.

$$2, 3, 5, 6, 7, 8, 9, 10$$

Ans : the median = $\frac{6+7}{2} = 6.5$

Q1 is the median of 2 , 3 , 5 , 6 which is $\frac{3+5}{2} = 4$

Q3 is the median of 7 , 9 , 9 , 10 which is $\frac{9+9}{2} = 2$

Example 2: Calculate the interquartile range of $1, 3, 5, 6, 7, 8, 8$

Ans **:** median = 6

Q1 is the median of 1 , 3 , 5 which is 3

Q3 is the median of 7 , 8 , 8 which is 8

Drawing the box plot: the median, Q1 and Q3 can be used for drawing the box plot

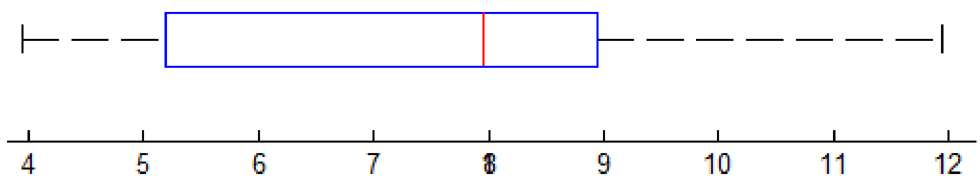
Example: Draw the box plot for the following data

$$4, 4, 5, 6, 8, 8, 9, 9, 9, 10, 12$$

Median = 8

$$Q1 = 5.5 \quad and \quad Q3 = 9$$

The box plot is shown in the figure .



MATLAB (B)

Try the following - exercises on MATLAB and Octave

```
>> a = [100  10  50  50  90]
>> b = [62  61  59  58  60]
>> range (a)
>> range (b)
>> var (a)
>> var (b)
>> std (a)
>> std (b)
>> boxplot(a)
```

range is function for getting range of data.

var is function to get variance of data.

std is the standard deviation.

boxplot is function for drawing the boxplot figure.

Exercises1

Question 1: for the following sample points 9,2,7,11,14,7,2,7,2,7,9

- Draw the frequency table
- Draw the histogram, number of classes = 5
- Calculate sample mean
- Calculate sample median
- Get sample mode
- Calculate sample Range

Question 2: A sample of 20 measurements is shown here:

2.6	3.4	2.1	3.2	4.2	3.6	2.8	3.8	1.7	3.9
2.2	1.2	5.6	3.9	2.5	4.1	3.0	2.3	2.7	1.9

- Use the data in the table to make a stem-and-leaf plot.
- Drive the frequency table from the data assuming number of classes = 5
- Draw the Histogram for the data.

Question 3: The ages of 5 randomly selected members of a club are as follows: 42, 52, 57, 63, 51

- The sample mean is (……….)
- The sample median is (……..)
- The sample Range is (…….)
- The sample Standard Deviation is (……)

Question 4: The following table represents the degrees out of 100 for Statistics and Math exams

STAT 301	74	68	62	69	67
MATH 333	58	98	38	78	68

(a) Calculate the sample Mean for the marks of each subject;
Can you differentiate between the two sets using Mean? Why?

(b) Calculate the sample Standard Deviation for the marks of each subject,
Can you differentiate between the two sets using Standard Deviation? Why? in what situation will the standard deviation be zero?

Question 5: Connect the following statements with the correct expression.

(a)	Eye color : blue, brown, hazel, green, etc.	(......)	Discrete
(b)	Number of children, Number of students,	(......)	Binary Data
(c)	Very unhappy, unhappy, neutral, happy,	(......)	Nominal Data
(d)	Gender: Male, Female	(......)	Continuous
(e)	Height, Temperature, Age	(......)	Ordinal Data

Question6: The following are the blood group of sample patients who attend clinic A.

A,	B,	O,	AB,	B,	A,	O,	O,	AB,	B
B,	B,	A,	O,	O,	AB,	B,	O,	B,	A
AB,	A,	O,	A,	A,	B,	O,	A,	A,	B

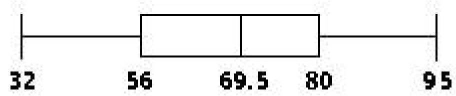
a. Construct the frequency table of the above data
b. Draw the histogram of the above data.

Question7. Draw a box plot for the following data set:

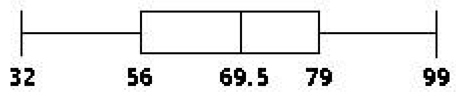
4.3, 5.1, 3.9, 4.5, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6, 4.4, 4.3, 4.8, 4.4, 4.2, 4.5, 4.4

Multiple Choices Questions: Choose the correct answer for the following questions.

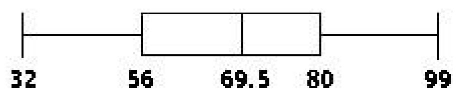
1) The number of students entering college in a certain year is 621.
a) Continuous b) Discrete
2) Calculate the Standard Deviation for the following data which was sampled from a large population: 4,10,12,2,15,5
a) 5.099 b) 26 c) 0 d) 15.12
3) The ages (in years) of the eight passengers on a bus are listed below.
10 7 26 16 21 43 40 30 Find the median age.
a)23.5 yr b)21 yr c)26 yr d)24.5 yr
4)The distances (in miles) in the past week by each of a company's sales representatives are listed below. 107 114 214 230 436 445 Find the mean.
a)214 miles b)220.50 miles c)257.67 miles d)230 miles
5)Find the mode for the given sample data. -20 -43 -46 -43 -49 -43 -49
a)-49 b)-46 c)-41.9 d)-43
6)The test scores of 32 students are listed below. Construct a box plot for the data set.
32 37 41 44 46 48 53 55 57 57 59 63 65 66 68 69 70 71 74 74 75 77 78 79 81 82 83 86 89 92 95 99
A)



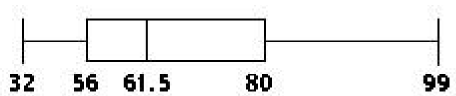
B)



C)



D)



7) Suppose S={1,2,3,4,5,6} is the space for an output of tossing a die, which of the following is a true event

a) E={-1,2,1} b) E={1,2,7} c) E={1,2,6} d) E={1,0,7}

8)The temperatures in 7 different cities on New Year's Day are listed below.

26 29 33 59 67 68 78

Find the median temperature.

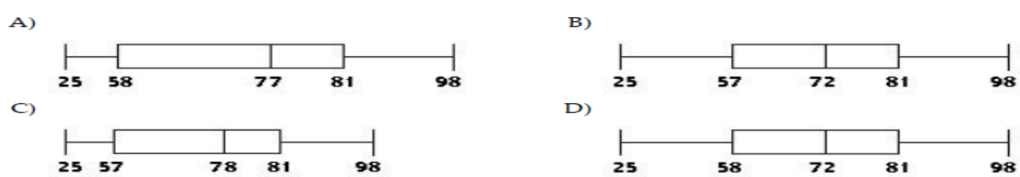
A)33°F B)59°F C)51°F D)67°F

9) Find the variance for the given data. 6.6 8.5 4.6 1.7 2.4

A)8.08 B)7.98 C)6.46 D)12.80

10) The test scores of 40 students are listed below. Construct a boxplot for the data set. D

25	35	43	44	47	48	54	55	56	57
59	62	63	65	66	68	69	69	71	72
72	73	74	76	77	77	78	79	80	81
81	82	83	85	89	92	93	94	97	98



11) If the sample size is 9 and the standard deviation is 7 then the variance is:

a) 49 b) 2.6 c) 2 d) 81

Solution Key

7 (c) 8 (b) 9 (a) 10 (d) 11 (a)

Question 4 (6 Marks)- Find the

a) Standard deviation and

b) Median for

60 62 62 61 61 60 62 61 61 60

mean = 61

a- $S^2 = \frac{1}{10-1} \sum (60-61)^2 + (62-61)^2 + (62-61)^2 + (61-61)^2 + (61-61)^2 + (60-61)^2 + (62-61)^2 + (61-61)^2 + (61-61)^2 + (60-6)^2$

$S^2 = \frac{6}{9} = 0.666$

$S = \sqrt{0.666} = 0.81$

6

b- sort 60 60 60 61 61 61 61 62 62 62

position $= \frac{10+1}{2} = 5.5$

median $= \frac{61+61}{2} = 61$

Chapter 2
Counting Sample Events

2.1 Sample Space and Events

The experiment is done on specific population to extract the samples and get the possible outcomes.

Random experiment: It is an experiment with a predictable outcome.

Example: to select randomly two students and check if they are smoker or non-smoker we have the following possibilities.

- The first student is non – smoker the second is non-smoker too.
- The first student is smoker the second is non – smoker.
- The first student is non – smoker the second is smoker .
- The first student is smoker, the second is smoker too .

If we define smoker by S and Non-smoker by N, We have the following sample points NN, NS , SN, SS.

The set of all possible outcomes is called sample space S.

$$S = \{NN , NS, SN, SS\}$$

Example: Get the sample space of tossing coin

Ans: We have two sample points for a coin either image / writing also called head / tail

$$S = \{H , T \}$$

Example: find the sample space of tossing a die.

Ans: we have six faces of the dice, so the sample space is

$$S = \{1 , 2 , 3 , 4 , 5 , 6\}$$

2.1.1 Event

The event is a subset of S which is defined with a condition on S. For tossing a die, we have the event of getting an even number E_1

$$E_1 = \{ 2 , 4 , 6 \}$$

Here $E_1 \subseteq S$

Example**:** what is the sample space of selecting 3 items from a manufacturing line. given that each item can be classified as either Defective D or Non – Defective N.

What the event of:

- Getting at least two defective items.
- Getting at most one defective item.
- Getting 3 defectives.

Answer: 3 items with 2 possibilities each, we get the following sample Space:

$$S = \{DDD , DDN , DND , DNN , NDD , NDN , NND , NNN\}$$

We call the first event A, the second event B, the third event C.

$A = \{DDD , DDN , DND , NDD\}$

$B = \{DNN , NDN , NND , NNN\}$

$C = \{DDD \}$

$A, B, C \ are \ \subseteq S$

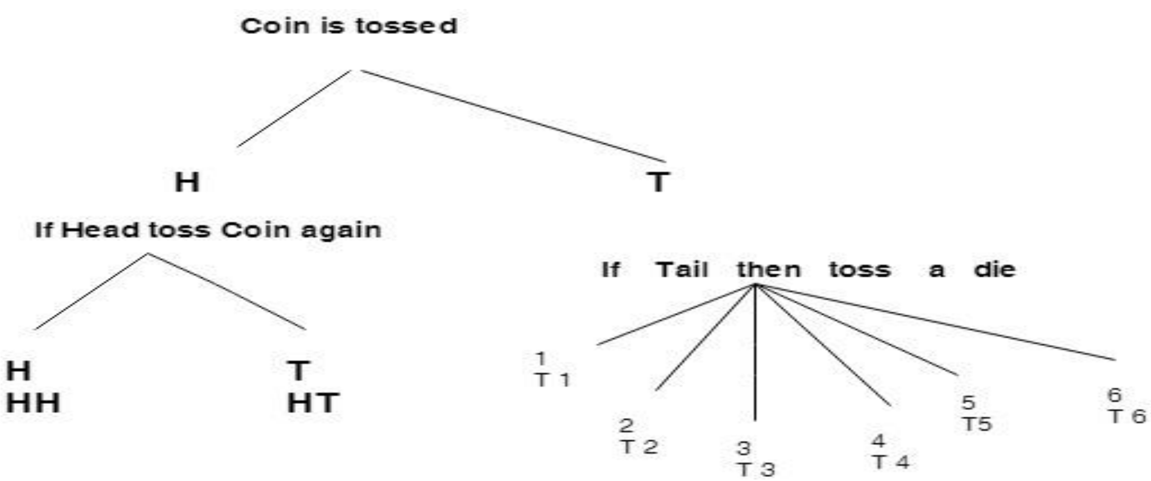
Example**:** An exponent consists of tossing a coin and then throw Adie if tail occurs in the coin otherwise the coin is tossed another time find the sample space and the event of getting at least one tail .

Ans : The sample space $S = \{HH , HT , T1 , T2 , T3 , T4 , T5 , T6\}$

The event of getting at least one tail

$$Et = \{, HT , T1 , T2 , T3 , T4 , T5 , T6\}$$

$Et = \emptyset$ If the event is not possible such as getting – 1 is case of tossing die



2.1.2 Rule Method: Events can be expressed with the use of rule method. It is difficult to write all the prime numbers from 1 to 1000 in a set. If the number of items is huge we have to use the rule method.

Example**:** Use rule method to write the set of all cities in the world .

$$E = \{x \ \ | x \ is \ a \ city \ in \ the \ world \}$$

All non – smoker students in Taibah University.

$$E = \{B : B \ is \ non - smooking \ student \ in \ taibah \ university\}$$

2 . 2 – Operations on Events

2.2.1 Complement

The complement of an event A is the elements found in S but not found in A .

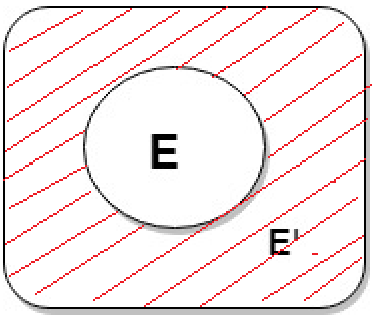
Example 1**:** $S = \{1 , 2 , 3 , 4 , 5 , \ 6\}$

E is the event getting and odd number

$$E = \{1\,,3\,,5\,\}$$

Then E complement is $E' = \{2\,,4\,,6\}$, Note that E and E' has no elements in common .

The following Venn – Diagram shows the complement of event E .



$S = \{\,1,2,3,4,5,6\}$

$E = \{\,1,3,5,\}$

$E' = \{\,2,4,6,\}$

2.2.2 Intersection

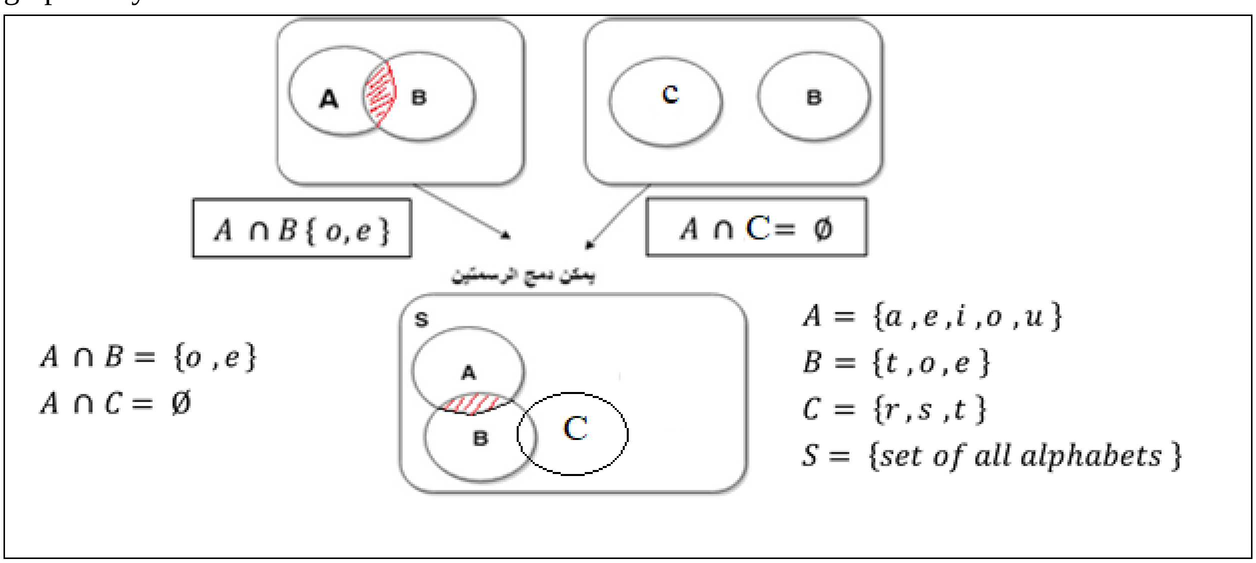
The interaction of two events A, B denoted by $A\ \cap B$ contains the elements which are common to both A and B.

<u>Example 2</u> **:** Suppose $A = \{a\,,e\,,i\,,o\,,u\,\}$, $B = \{t\,,o\,,e\,\}$, $C = \{r,s,t\,\}$

$find\ \ A\ \cap B\,,A\ \cap C$

<u>Ans</u>**:** $A\ \cap B = \{o\,,e\,\}$

$A\ \cap C = \{\emptyset\}$, Using the following Venn Diagram, we can see how intersection is represented graphically



2.2.3 Union

The union of two events A, B denoted as $A\ \cup B$ is the set of all elements in A, B without repetition.

Example 3: Suppose $A = \{a, r, s, t, u\}, R = \{L, m, n\}$ and $B = \{s, t, w, z, k\}, C = \{s, t, u\}$

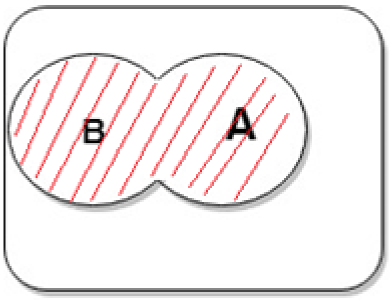
Find $A \cup B, A \cup C, A \cup R$

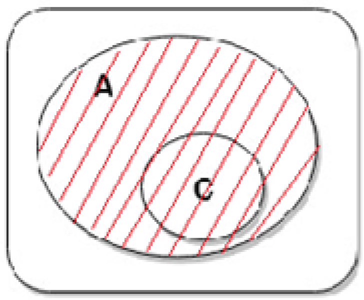
Ans:

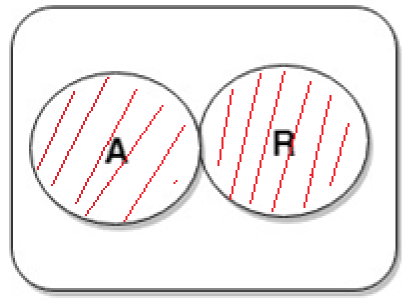
$A \cup B = \{a, r, r\ s, t, u, w, z, k\}$

$A \cup C = \{a, r, s, t, u\} = A$

$A \cup R = \{a, r, s, t, u, l, m, n\}$







$A \cup B$	$A \cup C$	$A \cup R$
يوجد قاسم مشترك	A تقع داخل C	لا يوجد قاسم

Example: if $m = \{x \mid 2 < x < g\}$

$$n = \{y \mid 5 < y < 12\}$$

$$m \cup n = \{z \mid 2 < z \mid 2\}$$

MATLAB (A)

To create random toss of coin we use

$\gg$ **a =' Ht'**

$\gg$ **a ((rand (1.1) < 0.5) + 1)**

The output should be randomly H or T. If we want to create sequence of heads and tails

$\gg$ **a ((rand (1.4) < 0.5) + 1)**

The output get randomly four consecutive H and T such as THTT.
To get randomly 3 items each item is defective or non – defective (N or D)

$\gg$ **a =' ND'**

$\gg$ **a = ((rand (1.3) < 0.5) + 1)**

The output should be randomly 3 items defective or non-defective Such as NND or NNN or DND etc.

To create toss of die use the following command

$\gg$ **unidrnd (b)**

The output should be random value from 1 to 6 to simulate many tosses for example 10 tosses

$\gg$ **unidrnd (6, 1, 10)**

The output should be something like 6 6 3 1 2 3 4 2 4 5

Operations on Events: For two sets a , b , the intersection is found as

$\gg$ **a = [100 10 50 50 90]**

$\gg$ **b = [2 100 10]**

$\gg$ **intersect (a , b)**

The output should be 10, 100

For the union of two sets use

$\gg$ **union (s , b)**

The output is 2 10 50 90 100

another example.

$\gg$ **A = {'cat','dog,'pig','bird'}**

$\gg$ **B = {'cat','fish','horse','bird'}**

$\gg$ **Interset (A , B)**

$\gg$ **union (A, B)**

For the complement of two sets use

$\gg$ **setdiff (a, b)**

The output should be 50, 90

The elements found in a and not found in b However.

$\gg$ **setdiff (b , a)**

The output should be 2

The elements found in b and not in a .

2.3 – Counting Techniques

Counting techniques are used to estimate the possible number of sample points in a random experiment. Some experiments produce thousands of cases which can't be listed one by one .

we talk about the following counting techniques.

- Multiplication rule .
- Permutations . تباديل
- Combinations . توافيق

2.3.1 – Multiplication rule

If experiment produces n1 possible cases and another experiment produces n2 possible case then we have a total of n1 x n2 possible cases.

Example: Tossing two dice produces 36 ways. Because n1 = 6 ways and n2 = 6 ways , and hence the two dice produce $6\ x\ 6 = 36$ points or ways .

Example: Car license plate has two digits and three English positions, how many plates can be produced?

Ans**:**

2	3	A: C: R

P1 (Plate position #1) Can be done with 10 possible numbers.

P2 Can be done with 10 possible numbers.

P3 Can be done with 26 possible letters.

P4 and p5 also can be done with 26 possible letters, a total of $= 10\ \ x\ 10\ x\ 26x\ 26x26$ license plates.

Example: An ATM machine is hacked by using a card with 4 digits secret number, how many trials the hacker can do to guess the secure number if he knows the following .

- The numbers have repeated digits.
- The numbers has no repeated digits.
- The numbers starts with 19 and the last digits are not similar .

Ans**:**

- The number has repeated digits the four digits are guessed among the 10 digits (0 ,1,2,3,4,5,6,7,8,9,10). Guesses = 10x10x10x10 = 10000 guesses .
- No repeated digits, So if the first digit can be tried in 10 ways, the second digit on be tried in 10 – 1 or 9 ways and so on. Guesses = 10 x9x8x7 = 5040
- Starts with 19 and no the last digits are not similar but including 1 and 9. guesses = 1 X 1 X 10 X 9 = 90 guesses .

Example 5 : How many 3 digits even numbers can be formed from the digits 3,4,6, 7. if it is known that no digit repetitions .

Ans **:**

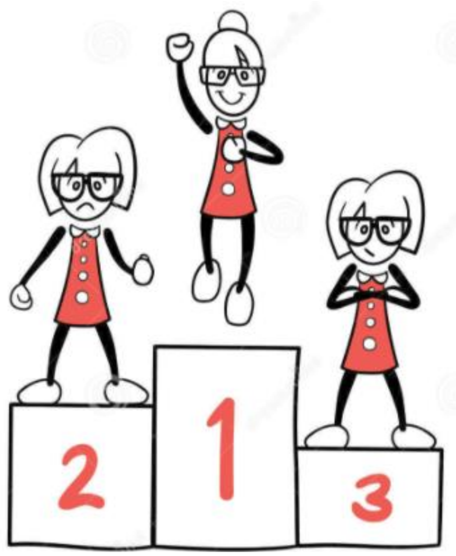
		4 or 6
3	**2**	**2**

A total of 3x2x2 = 4x3=12 possible even numbers can be formed from 3,4,6,4 without repetition.

2 .3 . 2 : permutation

The possible arrangements of all or part of a set of objects. It is applied when objects are "Ordinal". In other words, when order matters. For example three persons Mona, Hoda and Amira in a winning race .

التباديل: تستخدم في حالة وجود مجموعة من العناصر، عندما يتم تغيير ترتيبهم يظهر شكل جديد. ففي حالة السباق إذا كان لدينا ثلاث متسابقين فإن تغيير الترتيب يعني تغيير الفائز. وهذا شكل مختلف تماما.



We have 6 possible assignments of the persons. So if we use the first letters of the names MHA, we have 6 possible winning arrangements. MHA, MAH, AMH, AHM, HAM, HMA. The order matters.

Example : how many possible arrangements of permuting the letters x,y,z .

Ans :

We have $(x, y, z), (x, z, y), (y, z, x), (y, x, z), (z, x, y), (z. y. x\ )$ 6 possible cases .

The same result can be obtained by using the multiplication rule or 3x2x1 = 6 possible cases. Mathematically, permutation can be calculated by using factorial

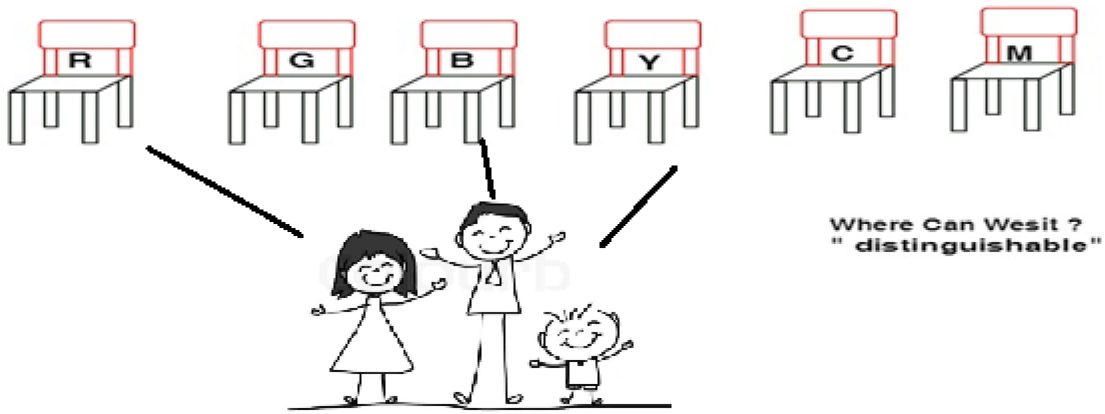
$$\boxed{\# \, permutations = n!}$$

the permutation of n distinct objects taken r at a time is (كمثال إذا كان عدد المتسابقين 6 وعدد والميداليات تكون ثلاثة ميداليات فقط)

$$\boxed{nPr = \frac{n!}{(\, n - r\,)!}}$$

Example 2 : How many possible ways of assigning mother, father and son on 6 possible colored chairs .

Ans :



Father , mother and the son need only 3 chairs out of 6 , they can set in the following possible ways

$$6P3 = \frac{6!}{(6-3)} = \frac{6x5x4x\cancel{3!}}{\cancel{3!}}$$

$$= 6\ x\ 5\ x4 = 20\ x\ 6 = 120\ ways\ .$$

Example 3 **:** How many ways a president and vice president can be chosen out of 5 students (V,W,X,Y,Z) .

- If there is no restrictions.
- If V will serve only if and only if he is a president.
- If W and X will be together or not at all.
- If Y and Z will not serve together.

Ans :

- No restriction means $\frac{5!}{(5-2)!} = \frac{5\ x\ 4x\ \cancel{3!}}{\cancel{3!}} = 20\ cases$
- V want to be president, So we have the possible 4 selections VW, VX, VY, VZ or (n – 1) selection . in addition to any other two except V or $\frac{4!}{(4-2)!} = \frac{4x3x2!}{2!} = 12$ a total of 4 + 12 = 16 cases.

- We have two cases that W, X appears together which are WX and XW, so we have 2 cases in addition to selecting 2 out of the remaining (n – 2) students (v, y, z) or $\frac{3!}{(3-2)!} = 3! = 6\ ways$. A total of 2 + 6 = Cases .
- Y can serve (n – 1) times if he is president. but in only another case when Z is Vice – President he will not serve as Vice in (n – 2) times. Y will not work if Z is the president. a total of $2\ x\ (n-2) = 2\ x\ (5-2) = 6$, Z can serve the same times of Y which is 6 times and also the other cases are possible if Y and Z are not selected (N – 2) a total of, $6 + 6 + \frac{3!}{(3-2)!} = 6 + 6 + 6 = 18\ times$

In Case Some of the things to be permuted are similar. for example if axx is given and we are asked to make three letter words using them we get axx, xax , xxa which are 3 generally .

$\left(\frac{n}{n1\ ,n\ 2\ ,n\ 3 \ldots .nr} = \frac{n\ !}{n1\ !\times\ n2\ !\times\ n3!\times \ldots ..nr\ !}\right)$, Where $n1 + n2 + n3 + \cdots \ldots \ldots \ldots + nr = n$

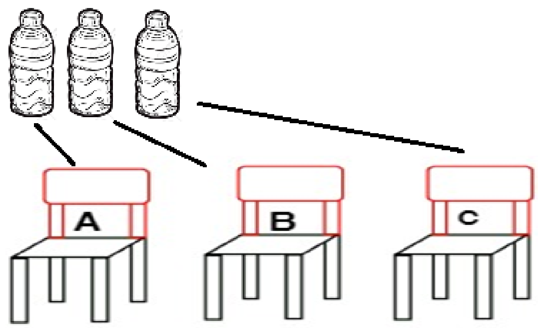
In the 'a x x' example we get , $\left(\frac{3}{1.2}\right) = \frac{3!}{1!x\ 2\ !} = 3$

Example: 10 students are to be assigned to two triple and two double rooms How many ways they can be selected .

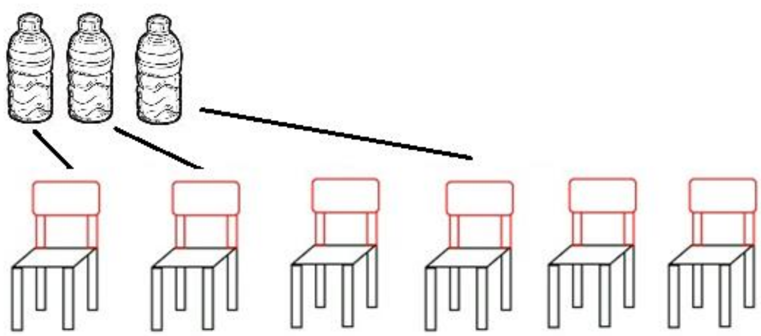
$$\left(\frac{10}{3\ ,3\ ,2\ ,2}\right) = \frac{10\ !}{3!\ \times 3!\ \times 2!\ \times 2!} = \frac{10\ \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3!}{3!\ \times 3! \times 2 \times 2} = 90\ \times 56$$
$$= 25200\ ways$$

2.3.3: Combinations

In many problems we are interested to find how many ways of selecting r objects out of n objects without regard to order. for example if we have 3 indistinguishable water bottles to be put on the three chairs we would have only single permutation .



If we changed the position of one bottle to another chair, that will not affect the permutation it will remain single combination .



If there are six chairs and 3 bottles, the resetting permutation would be.

$\frac{6\ !}{3!(\ 6-3\ )!} = \frac{6\ \times 5\ \times 4\ \times 3!}{3!\ \times 3\ !} = 20\ ways$

Compared to permutation, the combination ways are less than permutations. Generally.

$\binom{n}{r} = \frac{n!}{r!(\,n-r\,)!} \quad r = 0\,,1\,,2\,,00000\,,n$

<u>Example:</u> we have 10 operation rooms, 4 patients. How many ways can we assign the 4 patients to the rooms?

<u>Ans</u>: $n = \ 10 \qquad r \ = \ 4$

$$\binom{10}{4} = \frac{10!}{4!\,(\,10-4\,)!} = \frac{10\,!}{4! \ \times 6!} = 210\,Ways$$

MATLAB (B)

Permutations can be done with.

$\gg$ **a = [10 5 6]**

$\gg$ **perms (a)**

These will result in permuting all the numbers 3 at a time, to get the number of permutation We use the factorial function.

$\gg$ **factorial (3)**

To select 3 objects out of six we used the following formula

$\gg$ **n = 6; r = 3;**

$\gg$ **npr = factorial(n)/factorial (n − r)**

In combinations, we use similar formula

$\gg$ **n = 10; r = 4 ;**

$\gg$ **nCr = factorial(n)/foctorial(r) ∗ factorial (n − r)**

Exercises2

Question 1: Suppose S={1, 2, 3, 4, 5, 6} A={1, 2, 3, 4, 5 } B={3, 4, 5, 6 }, C={6}

1. Find A' or the complement of A.
2. A ∪ B
3. A ∩ B, draw the Venn diagram
4. A ∩ C, draw the Venn diagram

Question 2: Count the possible ways for the following cases

1. A pair of dice is thrown once.
2. Three dices and two coins are thrown once.
3. Permutation of a, b, c three letters at a time.
4. Permutation of a, b, c two letters at a time.
5. Combinations of a,b,c,d,e,f three letters at a time.

Question 3: Suppose S = {0,1,2,3,4,5,6,7,8,9} and A ={0,2,4,6,8}, B = {1,3,5,7,9}, C = {2,3,4,5}, and D = {1,6, 7} Find

i. A ∪ C, draw the Venn diagram
ii. A ∩ B, draw the Venn diagram
iii. Find C' or the complement of C.
iv. C' ∩ D, draw the Venn diagram

Question 4: How many even three-digit numbers can be formed from the digits 0, 1, 2 and 4 if each digit can be used only once.

Question 5: How many different letter arrangements can be made from the letters of the word GOOGLE

Question 6: if S={0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, and A= {0, 2, 4, 6, 8}, B={ 1, 3, 5, 7, 9}, C={2, 3, 4, 5} and D= {1, 6, 7}, list the elements of the sets corresponding to the following events, draw Venn diagram for each event:

a) A U C b) A ∩B c) C^c d) (C^c ∩ D) U B e) $(S \cap C)^c$

Question 7: Suppose that the checking of errors in transmission of CPU binary numbers is done with even parity. It is needed to transfer 101101 binary string.

a) How many permutations are expected to get Negative binary string (i.e. binary string without errors).
b) How many permutations are expected to get False Negative binary string (i.e. binary string with correct parity but has errors).

Multiple Choice Questions:

1) The difference between a permutation and a combination is:

a) In a permutation order is important and in a combination it is not.
b) In a permutation order is not important and in a combination it is important.

c) In both the order is important.
d) Non of above

2) ${}_5P_3 =$

a) 5 b)20 c)10 d)60

3) How many words consisting of 3 letters that can be construct from a x x ?
a) 1 b) 2 c) 3 d) none

4) If we have 7 equal–priority operations and only 3 operating rooms are available, in how many ways can we choose the 3 patients to be operated on first?
a) 15 b) 35 c) 25 d) 21

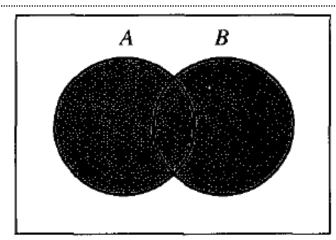
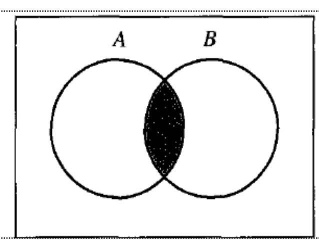
5) If an experiment consists of throwing a coin and then drawing a letter at random from the English alphabet, how many points are there in the sample space?
a) 52 b) 56 c) 156 d) 28

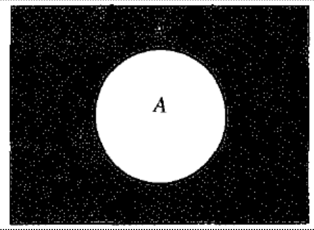
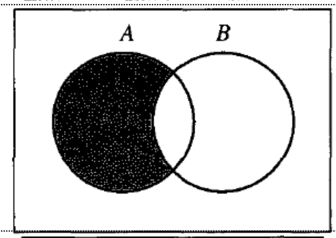
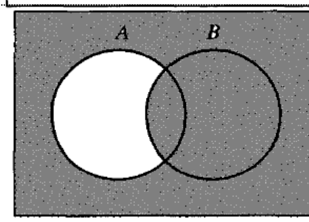
6) In 3 tosses of a coin which of following equals the event "getting two heads"?
[1] = {THH,HTH,HHT,HHH}
[2] = {THH,HTH,HHT}
[3] = {HTH,THH}
[4]={HHH,HHT,HTH,HHT,THH,THT,TTH,TTT}

7) In how many ways can a true-false test consisting of 9 questions be answered:
[1] 9 [2] 2 [3]18 [4] 2^9

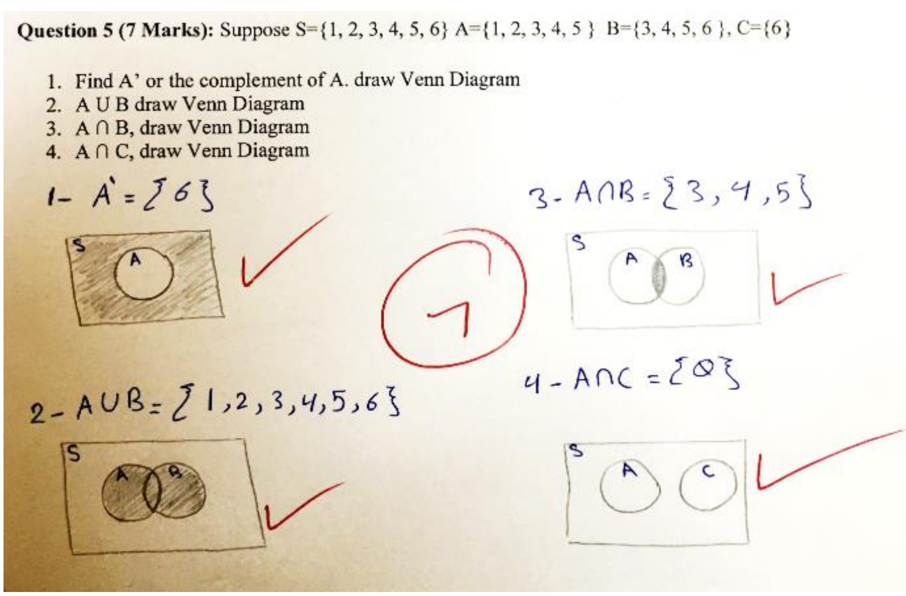
8) A president and treasure are to be chosen from student club consisting of 50 people. How many different choices of officer are possible if A will serve only if he president:
[1] 2450 [2] 2352 [3] 2401 [4] 49

Connection Question: Connect each set with the corresponding Venn diagram

	Answer		*Venn Diagram*
(........)	$\bar{A} \cup B$	(a)	
(........)	$A \cup B$	(b)	

(........)	$A \cap \overline{B}$	(c)	
(........)	$A \cap B$	(d)	
(........)	$\overline{A}$	(e)	

Answers to selected questions



Question 5 (7 Marks): Suppose S={1, 2, 3, 4, 5, 6} A={1, 2, 3, 4, 5 } B={3, 4, 5, 6 }, C={6}

1. Find A' or the complement of A. draw Venn Diagram
2. A U B draw Venn Diagram
3. A ∩ B, draw Venn Diagram
4. A ∩ C, draw Venn Diagram

1- $A' = \{6\}$

3- $A \cap B = \{3, 4, 5\}$

7

2- $A \cup B = \{1, 2, 3, 4, 5, 6\}$

4- $A \cap C = \{\emptyset\}$

Question 7: Suppose that the checking of errors in transmission of CPU binary numbers is done with even parity. It is needed to transfer 101101 binary string.

a) How many permutations are expected to get Negative binary string (i.e. binary string without errors).

b) How many permutations are expected to get False Negative binary string (i.e. binary string with correct parity but has errors).

Answer:

a) There is only one possible combination which is the exact string 101101, so the answer is 1.

b) Here, the string is of correct even parity if it has two 1’s or four 1’s or six 1’

for two 1's ~~we can~~ they can appear at any position in the string, selecting 2 out of 6

$$\binom{6}{2} = \frac{6!}{4! \times 2!} = \frac{{}^{3}\cancel{6} \times 5 \times \cancel{4!}}{\cancel{4!} \times \cancel{2}} = 15$$

selecting 4 out of 6

$$\binom{6}{4} = \frac{6!}{4! \times 1!} = \frac{6 \times 5 \times \cancel{4!}}{\cancel{4!} \times 1} = 30$$

One combination out of the 30 combinations is True Negative, so we have a total of

$$15 + 30 - 1 = 15 + 29 = 44$$

another combination $\binom{6}{6} = 1$

so the total is $44 + 1 = 45$

Chapter 3
Probability of an Event

Suppose we have the Sample space throwing 2 coins, $S = \{ HH , HT , TH , TT \}$, The event E is the Event of receiving at last one head. so event E can be expressed ay as .

$$E = \{ HH , HT , TH \}$$

The probability of the event E is expressed by the possibility of receiving at least one head . which occurs 3 times out 4 times (sample space points) .

The probability of 3 is then = $^3/_4 = 0.75$.

The probability is expressed by number between and including 0 and 1 . The probability of the event E has the following properties .

- $0 \le P (E) \le 1$
- $P (S) = 1$
- $P (\emptyset) = 0$

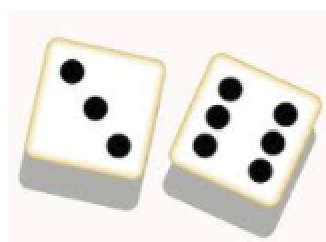
3 .1 Calculating Probability

The probability of the event E is calculated by

$$P(E) = \frac{n\,(E)}{n\,(\,s\,)} = \frac{n(E)}{N} = \frac{no.of\ sample\ points\ in\ E}{no\ .of\ sample\ Points\ in\ S}$$

Example: In tossing two dice experiment what is the probability of getting 6 in both dices

Sol : The sample space for two dice is 6 x 6 = 36 sample points only single sample point has 6 in both dice .



$$S = \{r, y | 1 \le x \le 6, 1 \le y \le 6\}$$

$$E = \{(6,6)\}$$

$$p(E) = \frac{n\,(E)}{N\,(S)} = \frac{1}{36} = 0.027$$

بمعني أنه في حالة رمي زهرتين فإن الحصول على 6 في كلا الأوجه يحتمل بنسبة 0.027 مما تعني انه بمعدل الف رمية يوجد احتمال وجود 27 مرة بالشكل (6 . 6)

Example: A jar of 13 candies (حلوى) , 6 of which are mint taste, 4 are toffee, and 3 are chocolate if a person selects two candies at a time. what is the probability of

$$\boxed{\begin{array}{ccc} M & T & C \\ \hline 6 & 4 & 3 \end{array}}$$

- Getting 2 mints. *m* event .
- Getting 2 toffee or 2 chocolate .

Solution: Selecting any two out of has the following no . of selections .

$$\binom{13}{2} = \frac{13\,!}{(\,13 - 2\,)!\ \times 2!} = \frac{13 \times \cancel{2} * 6 \times \cancel{11}!}{\cancel{11}!\ \times \cancel{2!}} = 13\ \times 6 = 78$$

- Selecting 2 mints has the following Ways .

$$\binom{6}{2} = \frac{6!}{(\,6-2)!\ \times 2!} = \frac{\cancel{36}\ \times 5 \times \cancel{4!}}{\cancel{4!} \times \cancel{2}!} = 15$$

$$P(m) = \frac{n(m)}{n(S)} = \frac{15}{78} = 0.19$$

- Call getting toffee event T and getting chocolate C, Selecting 2 toffee has the following ways .

$$\binom{4}{2} = \frac{4!}{(\,4-2)!\ \times 2!} = \frac{\cancel{34}\ \times 3 \times \cancel{2!}}{\cancel{2!} \times \cancel{2}!} = 6$$

selecting two chocolate out of 3 has the following ways.

$$\binom{3}{2} = \frac{3!}{2!\ \times (\,3-2\,)} = \frac{\cancel{3}\ \times 2 \times \cancel{2!}}{\cancel{2!} \times \cancel{1}!} = 3$$

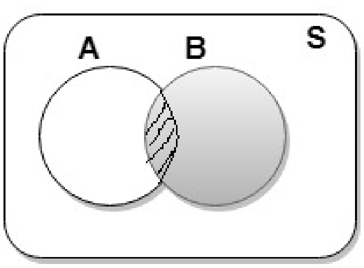
number of ways of selecting toffee or chocolate = number of ways select toffee + number of ways of selecting choc .

= 6+3 = 9

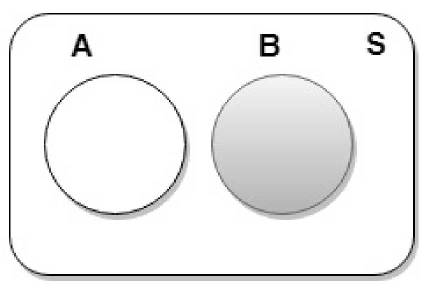
$$P\ (\,T\ U\ C) = \frac{n\ (\,T\ U\ C)}{n\ (S)} = \frac{9}{78}$$

3 .2 Additive Rule

An event probability can be calculated considering its sub. events. suppose there are two events A , B as show in figure .



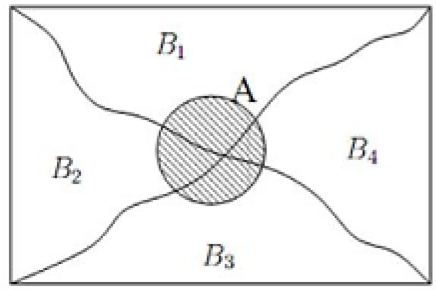
$$P\ (\,A\ \cup B\,) = P\ (A) + P(B) - P(A \cap B\,)$$



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

A,B disjoint

If the two events intersected, they have $A \cap B$ part repeated twice . So, part of them is removed when calculating $P\ (A \cup B\ )$.



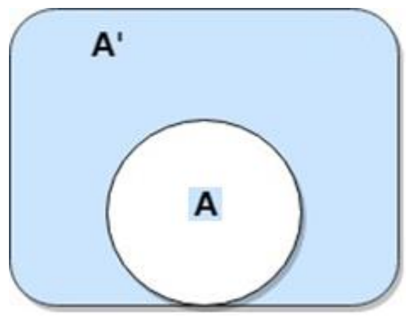
If the two events are mutually exclusive, i.e. $A \cap B = \emptyset$ then the probability of $A \cup B$ is simply the probability of the addition of the probability of sub events A , B

If the Sample space can be divided to B1, B2, B3, B4 events then.

$P\ (B1 \cup B2\ \cup B3\ \cup B4\ ) = P\ (B1) + P(B2) + P(B3) + P(B4) = P(S) = 1$

This is called partitioning. We also can conclude that if the event A and its complement A' are only found in the space then

$P(A) + P(A') = 1$
$P(A) = 1 - P(A')$



De Morgan theory:

$\overline{(A \cup B\ )} = A' \cap B'$

$\overline{(A \cap B\ )} = A' \cup B'$

So, $P(C \cup D\ ) = 1 - \overline{P\ (C \cup D\ )}$

$= 1 - P\ (C' \cap D'\ )$

$= 1 - P\ (C')\ P\ (D'\ )$

<u>Example:</u> The probability that Ahmed passes mathematics is $2/3$, the probability that he passes English is $4/9$. If the probability that he passes both courses is $1/4$. What is the probability that he will pass at least one course

<u>Solution :</u>

$P(M) = 2/3$, $P\ (E) = 4/9$

$p\ (M \cap E\ ) = 1/4$

$P(M \cup E\ ) = P(M) + P(E) - P(M \cap E)$

$= 2/3 + 4/4 - 1/4 = \dfrac{31}{36}$

Example: If the probabilities that a car service shop receives 3,4,5,6,7,8 or more cars in a work day is 0.12, 0.19, 0.28,0.24,0.10 and 0.07 respectively. what is the probability that the shop receives at least 5 cars in the next work day ?

solution **:** let E be the event that at least 5 cars are received now $P(E) = 1 - P(E')$, where E' is the event that less than 5 cars are received

$P(E') = P\ (\ getting\ 3\ cars\ ) + P\ (\ getting\ 4\ cars\ )$

$P(E') = 0.12 + 0.19 = 0.31$

So $P(E) = 1 - 0.31 = 0.69$

Example: A manufacturer is producing computer cables with length 2000 mm +- 10mm which is acceptable. Cables greater than 2010 and less than 1990 are not acceptable . and the cable is called defective, if the probability if getting acceptable cable is 0.99 .

- What is the probability of getting too large cable, i.e. cables with length more than 2010.
- What is the probability that a randomly selected cable is larger than 1990 .

Solution: Let us define the events and name them accordingly, the events are as follow

M is the event of producing acceptable cables.

L is the event that the cable is large.

T is the event that the cable is small.

- $P(M) = 0.99, P(M) + P(T) + P(L) = 1$

  $P(T) = P(L) = \frac{1-p(M)}{2} = \frac{1-0.99}{2} = 0.005$
- We want to calculate.

  $P(M) + P(L) = 1 - P(T) = 1 - 0.005 = 0.995$

3.3: Conditional Probability

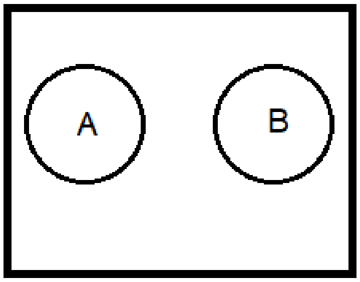
Any system is complex with multiple events. to calculate the probability of an event We shall consider the corresponding events in the same environments. There are two important relations between events.

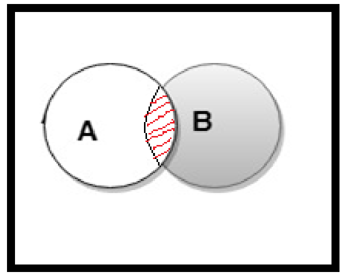
- Dependent events.
- Independent events.

Dependent events such as speed of the car and no# of accidents. Independent events such as person's height and the number of family members or the # of accidents of cars and # of Aeroplan accidents, they are not related but still they can occur at the same time. Disjoint events are the events that never occur together. Here, we consider only the dependent/independent events.

The probability that B occurs given that A occurs is called conditional probability and denoted by P (B | A), pronounced as “probability of B occurs given that A occurs.

We have two cases shown in the figure.





$P\ (\ B|A\ ) = 0$

A, B disjoint

$In\ case, P\ (\ B|A\ ) = P\ (\ B\ )$

A, B Independent

$$P\ (\ B|A\ ) = \frac{P\ (A \cap\ B\ )}{P\ (A)}$$

A, B dependent

<u>Note</u>**:** if it is proved that P (B | A) = P (B) then we are sure that A,B are Independent .

<u>Example:</u> In basket analysis the buyer history is shown in the table, Tid refers to the receipt id.

Tid	Items bought
10	Beer, Nuts, Diaper
20	Beer, Coffee, Diaper
30	Beer, Diaper, Eggs
40	Nuts, Eggs, Milk
50	Nuts, Coffee, Diaper, Eggs, Milk

If B is event of buying a Beer and D is the event of buying diaper, find P(D), P(B ∩ D), P(B|D) and P(D|B).

<u>Solution</u>: P(D) means the probability of buying Diaper. The byer bought Diaper 4 times out of 5 transactions. So, P(D)=4/5=0.8

P(B ∩ D) means getting the probability of buying both Beer and Diaper in the same time. Bear and Diaper appeared 3 times out of total of 5 times (5 is the number of receipts). So P(B ∩ D)=3/5=0.6.

P(B|D) is the probability of buying Beer given that the buyer already bought a Diaper. The buyer bought Diaper 4 times out of which she bought Beer 3 times. So, P(B|D)=3/4=0.75.

P(D|B) is the probability of buying Diaper given that the buyer already bought a Beer. The buyer bought Beer 3 times out of which she bought Diaper 3 times. So, P(D|B)=3/3=1.0

We can easily verify that $P\ (\ B|D\ ) = \frac{P\ (D \cap B\ )}{P\ (D)} = \frac{3/5}{4/5} = \frac{3}{4} = 0.75$

<u>Hint</u>: P(B|D) will be equal to P(B) only if the Diaper appeared in all transactions. In this case we can say that B and D are independent events. In other words, buying Diaper has no effect on buying Beer. In addition, buying Milk (M) and buying Beer B are two disjoint events.

Example**:** The probability that a flight arrives in time is P(A) = 0.82. The probability that it departed in time is P (D) = 0.83. The probability that it departed in time and arrives in time is $P\ (\ D\ \cap A\ ) = 0.78$ Find,

- The prob. that it arrives in time given that it departed in time.
- The prob. that it departed on time given that it had arrived in time.

Solution: this an example of dependent events.

- $P\ \ (A|D) = \frac{P\ (A \cap D\ )}{P(D)} = \frac{0.78}{0.83} = 0.94$
- $P\ \ (D|A) = \frac{P\ (A \cap D\ )}{P(A)} = \frac{0.78}{0.82} = 0.95$

Which is different in both cases, $P\ (D|A)$ is higher simply because if it arrived in time there is higher prob. that it departed in time.

Example: Consider an experiment in which 2 cards are drawn in succession from an ordinary deck with replacement the events are defined as .

- The first card is an ace.
- The second card is spade .

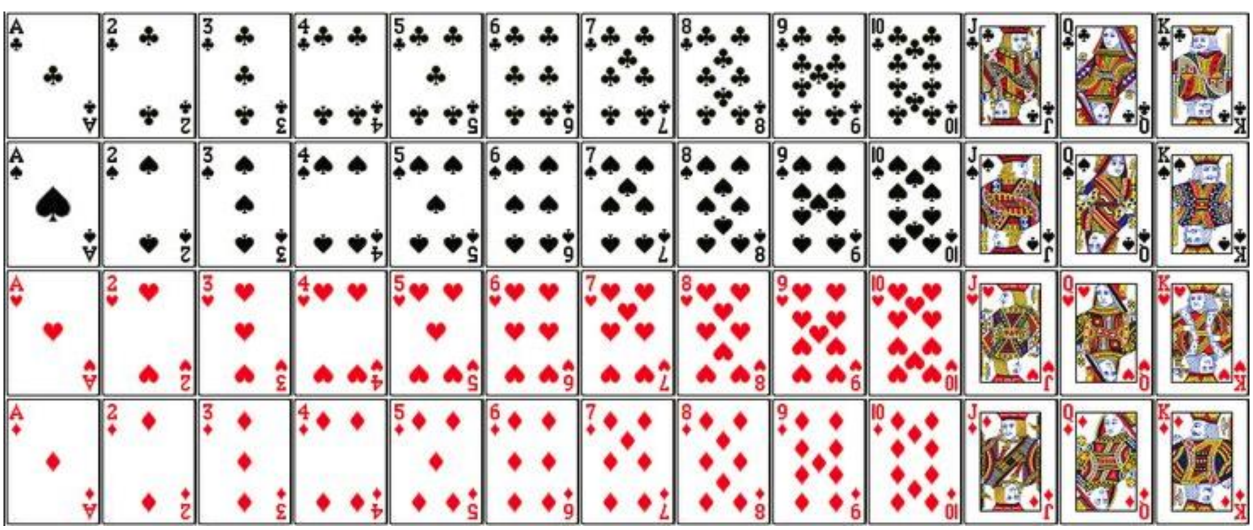
Find $P\ (B|A)\quad and \qquad P\ (A|B)$

Solution**:** this is an example of independent events since replacement forces the first event to be independent from the second event.

There are 4 aces in 52 cards and 13 spades.

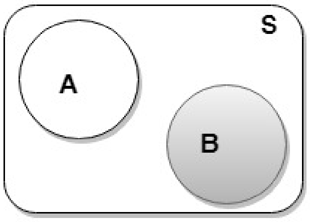
$P\ (B|A) = P\ (\ B\ ) \rightarrow independant\ ,P\ (B|A) = \ \frac{13}{52} = \ \frac{1}{4}$

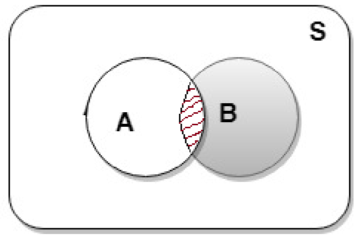
$P\ (A|B) = P\ (A) \rightarrow mdependant.\ ,P\ (A|B) = \ \frac{4}{52} = \ \frac{1}{13}$



3.4 Multiplicative Rules

There are three cases for many events either they are disjoint, dependent or independent.





$P\ (A \cap B\ ) = 0$

A, B disjoint

$P\ (B|A) = P\ (B),$ (1)

For A,B independent

substitute 1 In 3

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

A, B independent

$$P\ (B|A) = \frac{P(A \cap B)}{P(A)}, \qquad (2)$$

$$P(A \cap B) = P(B|A)\ P\ (A), \qquad (3)$$

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

A, B dependent

Example: Suppose that we have a fuse box containing 20 fuses, of which 5 are defective. if 2 fuses are taken from the box in succession without replacing the first, what is the probability that both fuses are defective?

Solution: Let A be the event that the first fuse is defective and B the second fuse is defective.

$P\ (\ A\ \cap B\ )$ is the event that A ocoures, then B occurs after A occurred.

The probability of the first fuse is defective is

$P\ (A) = \frac{5}{20} = \frac{1}{4}\ ,\ P\ (B) = \frac{4}{19}$

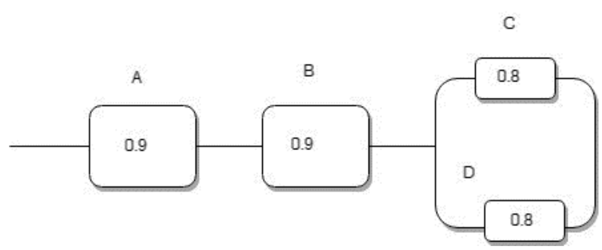
The probability of the second is defective given that the first fuse was defective is

$P(A \cap B) = P(A)P(B) \rightarrow independant$

$= \left(\frac{1}{4}\right)\left(\frac{4}{19}\right) = \frac{1}{19}$

Example: The entire system works, if A , B work together with either C o r D . give the probability of the following cases.

- The entire system works.
- The component C doesn't work given that the entire system works



A, B , C and D work independently.

Solution: A , B Serial C, D parallel

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

$$= P(A)P(B)\big(1 - P\,(C' \cap D')\big)$$
$$= P(A)P(B)\big(1 - P\,(C')\;P\,(D')\big)$$
$$= 0.9\;\times 0.9 \times (\,1 - 0.2\;\times 0.2\,) = 0.7776$$

- $P = \frac{P\,(\,system\;works\;but\;C\;not\;work)}{P\,(\,system\;Works\,)}$

$$= \frac{P\,(\,A\;\cap B \cap C' \cap D)}{P(\,sys.works)} = \frac{0.9 \times 0.9 \times 0.2 \times 0.8}{0.7776} = 0.1667\;\#$$

Notice that**:**

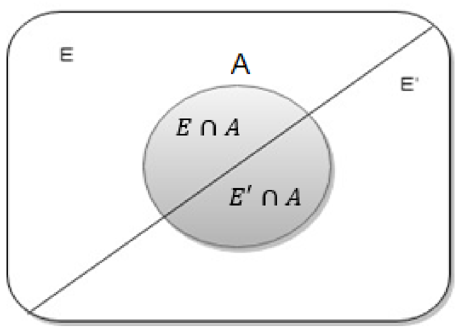
$P\,(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$P\,(A'|B) = \frac{P(A' \cap B)}{P(B)}$

$P\,(C'|SYS) = \frac{P(C' \cap SYS\,)}{P\,(SYS)}$, Where SYS is the event that the entire system works

3 .5 Bayes Rule

We can write A as a union of two independent events $E\;\cap A$, $E' \cap A$. Which is depicted in the following figure. Hence, A = $(E\;\cap A\,)\;\cup (\,E' \cap A\,)$

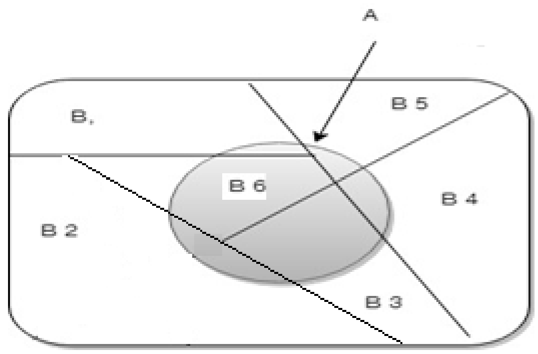


P (A) $= P[\,(\,E \cap A\,) \cup (E' \cap A\,)]$

$\quad = P(\,E\;\cap A\,) + P\,(\,E' \cap A\,)$

$\quad = P\,(E)P\,(A|E) + P\,(\,E')P\,(A|E')$, (1)

An extension of (1) is called the “Theorem of total probability” as shown in the figure



P (A) $= P\ [(B_1 \cap A) \cup (B_2 \cap A) \cup \ldots \quad (B_k \cap A)]$

$= P\ (B_1)\ P(A|B_1) + P(B_2)\ P(A|B_2) + \cdots \quad P(B_k)\ P(A|B_k)$

$= \sum_{i=1}^{k} P(Bi)\ P(A|Bi)$, and that is the "Rule of elimination", , (2)

Example: In a certain assembly plant, machines B1, B2 and B3 make 30%, 45% and 25%, respectively of products. it is known that 2%, 3% and 2% respectively for each machine are defective. What is the probability of that a selected random product is defective.

Solution: Consider the following events:

A: The product is defective.

B1: The product is made by machine 1

B2: The product is made by machine 2

B3: The product is made by machine 3

Using the total probability theorem,

$$P\ (A) = P\ (B_1)(\ P\ (A|B_1) + P(B_2)P\ (A|B_2) + P\ (B_3)P\ (A|B_3)$$

$P\ (B_1)P\ (A|B_1) = 0.3\ \times 0.02 = 0.006$

$P\ (B_2)P\ (A|B_2) = 0.45\ \times 0.03 = 0.0135$

$P\ (B_3)P\ (A|B_3) = 0.25\ \times 0.02 = 0.005$

Hence, $P(A) = 0.0006 + 0.0135 + 0.005 = 0.0245$

Suppose a random product is selected, what is the probability that it is made by machine B_1? We are not asking for the probability of getting defective product as solved by rule of domination, instead we ask for the source machine that produced that defective product. Bayes' rule solves such type of problems it is known that,

$P(B_r|A) = \frac{P(B_r \cap A)}{P(A)}$, substituting the P(A) from equation (2)

$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)}$, we know that $P(B \cap A) = P(B)\ P(A|B)$

$= \frac{P\ (B_r)P\ (A|B_r)}{\sum_{i=1}^{k} P\ (B_i)P\ (A|B_i)}$, Bayes' Rule

Example: Referring to example 1, if a product was chosen randomly and found to defective; what is the probability that it was made by machine B3?

Answer: Using Bayes' rule

$$P\ (B_3|A) = \frac{P\ (B_3)P\ (A|B_3)}{P\ (\ B_1)P\ (A|B_1) + P\ (\ B_2\ )P\ (A|B_2) + P\ (\ B_3)P\ (A|B_3)}$$

$$= \frac{0.005}{0.006 + 0.0135 + 0.005} = \frac{0.005}{0.0245} = \frac{10}{49}\ \#$$

Exercises3

Question1: Let $S = \{E1, E2, . . ., E6\}$; $A = \{E1, E3, E5\}$; $B = \{E1, E3\}$; $C = \{E1, E2, E3\}$; $D = \{E1, E6\}$. Suppose that all elementary events are equally likely.

(i) What does it mean that all elementary events are equally likely?

(ii) Use the complementation rule to find $P(A^c)$.

(iii) Find $P(A|B)$ and $P(B|A)$

(iv) Find $P(D)$ and $P(D|C)$

(v) Are A and B independent? Are C and D independent? *(Hint, check to see if $P(A|B)=P(A)$, and also check to see if $P(D|C) = P(D)$)*

Question 2: Suppose S={1, 2, 3, 4, 5, 6} A={1, 2, 3, 4, 5 } B={3, 4, 5, 6 }

Find Probability of A ∩ B assuming each sample point has equal probability of 1/6

Question 3: A coin is tossed twice. What is the probability that at least 1 head occurs?

Question 4: A spinner with five equally likely outcomes is spun. The outcomes are 1, 2, 3, 4 and 5.

b) What is the probability of getting a two?
c) What is the probability of getting an even number?

Question5: Khalid passes Math with probability ¼ and passes English with probability 2/3, if he passes both courses with probability ¾ what is the probability that he

a) pass at least one course
b) pass Math and fail English
c) fail both courses

Question6: complete the following formulas

a) P(B| A) = ………………, P(A) > 0, where A, B are events, P is the probability
b) P(S) = ……………… where S is the sample space and P is the probability
c) if P(A|B) = P(B|A) then both events A, B are said to be ………………
d) the distribution of random variable X has the following property $\sum_{all\ x} f(x) = \ldots\ldots\ldots\ldots$

Question7: Use event relationships to fill in the blanks in table below. Show your answers under the table.

P(A)	*P(B)*	*Conditions for Events for A and B*	*P(A ∩ B)*	*P(A ∪ B)*	*P(A | B)*
0.3	0.4	……….	0.12	……….	……….
0.3	0.4	……….	……….	0.7	……….
0.1	0.5	Mutually exclusive	……….	……….	……….
0.2	0.5	Independent	……….	……….	……….

Question8: If each coded item in a catalog begins with 3 distinct letters followed by 4 distinct nonzero digits, find the probability of randomly selecting one of these coded items with the first letter a vowel and the last digit odd.

Question9: It is common in many industrial areas to use a filling machine to fill boxes full of product. This occurs in the food industry as well as other areas in which the product is used in the home, for example, detergent. These machines are not perfect, and indeed they may A, fill to specification, B, underfill, and C, overfill. Generally, the practice of underfilling is that which one hopes to avoid. Let P(B) = 0.002 while P(A) = 0.980**.**

(a) Give P(C).

(b) What is the probability that the machine does not underfill?

Question10: Complete the following list of equations

- In case of two dependent events, P(M∪E)= P(M) + P(E) – …………
- The conditional probability of two dependent events *P(B|A) = ………………*
- P(A) + P(A') = ………………
- In case of 3 independent events A, B, and C the probability P(M∪E∪F)= ……………………………………………..=1
- The multiplicative rule in case of dependent events P(A ∩ B)= **………………………………..**

Multiple Choice Questions:

1. Two events, A and B, are said to be independent if:

a. P(A U B) = P(A).P(B)

b. P(A U B) = P(A) + P(B)

c. P(A | B) = P(B)

d. P(B | A) = P(A)

2. If P(A) = 0.8, P(B) = 0.3 and P(A|B) = 0.6, what is P(A ∩ B)?

a. 0.18

b. 0.24

c. 0.03

d. 0.30

3) P(A∩B) = Ø represents:

a) Independent events.

b) Mutually exclusive events.

c) Conditional events.

d) Dependent events.

4) The following formula: P(A∪B) = P(A) + P(B) - P(A∩B) represents:

a. the conditional probability.

b. the additive rule.
c. independence.
d. the multiplication rule.

5) Suppose that a sample space S consists of four simple events: A, B, C, and D. That is S = {A, B, C, D}. If P(A) = .4, P(B) = .1, P(C) = .2, what is P(D)?

a. 0.7
b. 0.1
c. 0.3
d. 1

6) An event A will occur with probability 0.5. An event B will occur with probability 0.6. The probability that both A and B will occur is 0.1. The conditional probability of A given B is

A) Cannot be determined from the information given.
B) is 0.167.
C) is 0.200.
D) is 0.833.

7) If three fair coins are tossed, what is the probability of getting at least two heads?
[1] 2/3 [2] 1/2 [3] 3/8 [4] 1/8

8) If two events (both with probability greater than 0) are mutually exclusive, then:
A. They also must be independent.
B. They also could be independent.
C. They cannot be independent.

9) suppose that the probability of event A is 0.2 and the probability of event B is 0.4. Also, suppose that the two events are independent. Then P(A|B) is:
A. P(A)=0.2
B. P(A)/P(B)=0.2/0.4=1⁄2
C. P(A)×P(B)=(0.2)(0.4)=0.08
D. None of the above.

10) Consider the two events A, B with: $P(A) = 0.5, P(B) = 0.7$ and $P(A \cap B) = 0.3$, find the probability $P(A \cup B)$
a- 0.14 b- 0.76 c- 0.9 d- none

11) Two events A, B with $P(A) = 0.3$, $P(B) = 0.6$ and $P(B|A) = 0.6$, The conditional probability $P(A|B)$ is equal to:
a- 0.3 b- 0.6 c- 0.2 d- 0.12

12) Suppose that the probability of event A is 0.4 and the probability of event B is 0.3, suppose A, B are independent, then $P(A \cap B)$ = …..
a- 0.4 b-0.3 c- 0.7 d- 0.12

Answers:

Question1: Let $S = \{E1, E2, \ldots, E6\}$; $A = \{E1, E3, E5\}$; $B = \{E1, E2, E3\}$; $C = \{E2, E4, E6\}$; $D = \{E6\}$. Suppose that all elementary events are equally likely.

(i) What does it mean that all elementary events are equally likely?

(ii) Use the complementation rule to find $P(A^c)$.

(iii) Find $P(A|B)$ and $P(B|A)$

(iv) Find $P(D)$ and $P(D|C)$

(v) Are A and B independent? Are C and D independent?

(vi) Find $P(A \cap B)$ and $P(A \cup B)$.

Question 1

i) this means every event has equal probability

so $P(E_1) = P(E_2) = \cdots = P(E6) = \frac{1}{6}$

ii) $P(A') = 1 - P(A)$

$= 1 - P(E_1) + P(E_3) + P(E_5)$

$= 1 - \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1 - \frac{3}{6} = \frac{6}{6} - \frac{3}{6}$

$= \frac{1}{2} = 0.5$

iii) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(E_1) + P(E_3)}{P(E_1) + P(E_2) + P(E_3)}$

$= \dfrac{2/6}{3/6} = \frac{2}{6} \times \frac{6}{3} = \frac{2}{3}$

$P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(E_1) + P(E_3)}{P(E_1) + P(E_3) + P(E_5)} = \dfrac{2/6}{3/6}$

$= \frac{2}{3}$

iv) $P(D) = P(E_6) = \frac{1}{6}$

$P(D|C) = \dfrac{P(D \cap C)}{P(C)} = \dfrac{P(E_6)}{P(E_3) + P(E_4) + P(E_6)}$

$= \dfrac{1/6}{3/6} = \frac{6}{3} \times \frac{1}{6} = \frac{1}{3} = 0.33$

v) $P(A|B) \overset{?}{=} P(A)$

$P(A|B) = \frac{2}{3}$ $\qquad$ $P(A) = P(A') = \frac{1}{2}$

$A, B \xrightarrow{\text{dependant}}$

$P(D|C) \overset{?}{=} P(D)$

$C, D \xrightarrow{\text{dependant}}$ $\frac{1}{5}$ $\qquad$ $\frac{1}{6}$

vi) ~~Fin~~ $P(A \cap B)$

$= P(E_1) + P(E_3) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

$P(A \cup B) = P(E_1) + P(E_2) + P(E_3) + P(E_5)$

$= \frac{4}{6} = \frac{2}{3}$

Question6: complete the following formulas

a) P(B| A) = ..., P(A) > 0, where A, B are events, P is the probability

b) P(S) = ... where S is the sample space and P is the probability

c) if P(A|B) = P(B|A) then both events A, B are said to be ....

d) the distribution of random variable X has the following property $\sum_{all\ x} f(x) = \ldots\ldots\ldots$

Question 6

a) $P(B|A) = \frac{P(A \cap B)}{P(A)}$ $\quad P(A) > 0$

b) $P(S) = 1$

c) If $P(A|B) = P(B|A)$, A, B is said equal

or $P(A) = P(B)$

d) $\sum_{x} f(x) = 1$ for all values of x

Question8: If each coded item in a catalog begins with 3 distinct letters followed by 4 distinct nonzero digits, find the probability of randomly selecting one of these coded items with the first letter a vowel and the last digit odd.

letters four digits.

Question 8

odd

1, 3, 5, 7, 9

Vowels: a, e, i, o, u

$5 \times 25 \times 24 \times 8 \times 7 \times 6 \times 5$

$$P(E) = \frac{5 \times 25 \times 24 \times 8 \times 7 \times 6 \times 5}{26 \times 25 \times 24 \times 9 \times 8 \times 7 \times 6}$$

$= 0.0855$

Question10: Complete the following list of equations

- In case of two dependent events, $P(M \cup E) = P(M) + P(E) -$ ....
- The conditional probability of two dependent events $P(B/A) =$ ....
- $P(A) + P(A') =$ ....
- In case of 3 independent events A, B, and C the probability $P(M \cup E \cup F) =$ ....=1
- The multiplicative rule in case of dependent events $P(A \cap B) =$ ....

Question 10

- $P(M \cap E)$
- $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$
- $P(A) + P(A') = 1$
- $P(S) = 1$
- $P(A \cap B) = P(B|A)\, P(A)$

The following is an example showing that two events can be independent and at the same time they are intersecting

$S = \{1, 2, 3, 4, 5, 6\}$

$C = \{1, 2, 6\} \rightarrow \} = S$

$D = \{1, 5\}$

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{1/6}{2/6} = \frac{1}{6} \times \frac{6}{2} = \frac{1}{2}$$

$$P(C) = \frac{1}{2} \implies P(C|D) = P(C)$$

C, D Independant

$$P(D|C) = \frac{P(C \cap D)}{P(C)} = \frac{1/6}{1/2} = \frac{1}{6} \times \frac{2}{1} = \frac{2}{6} = \frac{1}{3}$$

$P(D) = \frac{2}{6} = \frac{1}{3}$ $\qquad P(D|C) = P(D)$

C, D Independant.

Chapter 4
Random Variables and their Expected Values

4.1 Random Variable

When selecting three items from factory production line, in some cases the quality engineer is interested to know how many devices may be selected. The sample space should be:

$$S = \{ NNN\ , NND\ , NDN\ , NDD, DNN, DND\ , DDN\ , DDD\}$$

The engineer tracks the number as D's. which represents the number of defective devices. This number is called the random variable. is denoted by X. its corresponding value is x

Example: Basket contains 4 red balls, 3 black balls, two balls are drawn randomly, count the possibility of getting red ball.

Solution: Call R : Red   B : Black , We have the following possibilist.

Sample space	X
RR	2
RB	1
BR	1
BB	0

There are two types of random variables

- Discrete.
- Continuous

Discrete random variable is that with countable state such as the number of defective devices. *Continues random variable* takes values on continuous scale such as tracking height and body temperature.

4.2 Discrete Probability Distribution

Here we study the probability distribution of the random variable. The student may be interested in calculating the probability of getting for example 2 Defective devices.

Example: Unbiased coin is tossed twice. The probability of getting a tail is twice the probability of getting head. Calculate the probability distribution of the number of heads.

Solution**:** given $P\ (\ H\ ) = \frac{1}{2}\ P\ (T)$, we know from probability that *P(H) + P(T) = 1*

Then $P(H) = {}^{1}/_{3}\ , P\ (T) = {}^{2}/_{3}$

Suppose that X is the number of heads then, $S = \{\ HH\ , HT\ , TH\ , TT\ \}$

$P\ (HH) = P(H) \times P\ (H) = {}^{1}/_{3} \times {}^{1}/_{3} = \frac{1}{9}$

$\Rightarrow P\ (\ x\ = 2\ ) = \frac{1}{9}$

$P\ (HT) = P(TH) = P(x\ = 1\ ) = P(H)P(T) = {}^{1}/_{3} \times {}^{2}/_{3} = \frac{2}{9}$

$P\ (TT) = P(x\ = 0\ ) = P(T) \times P(T) = {}^{2}/_{3} \times {}^{2}/_{3} = \frac{4}{9}$

$\boldsymbol{X}$	**0**	**1**	**2**	**Total**
$\boldsymbol{P\ (X)}$	${}^{4}/_{9}$	${}^{2}/_{9} + {}^{2}/_{9}$	${}^{1}/_{9}$	1

The set of ordered pairs $\big(\ x, F\ (x)\big)$ is called the probability distribution of the random variable, is called PDF (Probability Distribution Function). PDF has the following properties

- $f\ (x) \geq 0$
- $\sum f\ (x) = 1$
- $P(X\ =\ x) = f(x)$

Another distribution called Commutative distillation function CDF is defined as

$$F(x) = P\ (X \leq x\ ) = \sum\nolimits_{t\ \leq x} f\ (t);\ -\infty < x < \infty$$

Example 2 : Calculate CDF of the above example (1)

$for\ x < 0\ , f\ (\ x) = 0$

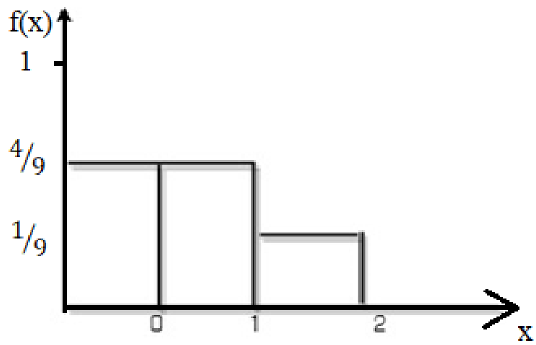
$for\ 0\ \leq x\ < 1\ , \qquad f\ (x) = \frac{4}{9}$

$for\ \ 1 \leq x\ < 2\ ,\ \ f\ (x) = P\ (\ x = 0\ ) + P\ (\ x = 1\ )$

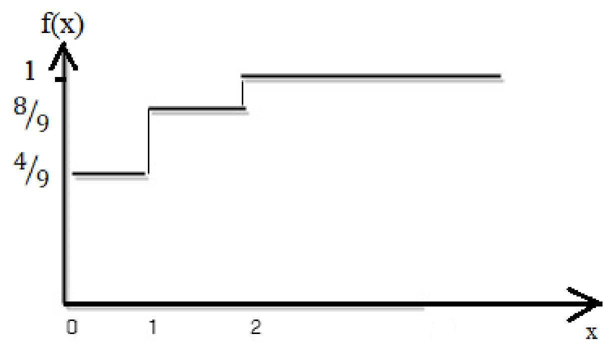
$$= \frac{4}{9} + \frac{4}{9} = \frac{8}{9}$$

$for\ x \geq 2\ \ \ F\ (x) = P\ (x = 0) + P\ (x = 1) + P\ (x = 2\ )$

$$= \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = \frac{9}{9} = 1$$



PDF



CDF

4.3 Continuous Probability Distribution

Continuous probability works with continuous variables. For example, if it is required to register patient's body temperature every day. it is found that in a month the degree was between 35 and 38 for 20 days and in 5 days it was less than 35 and for another 5 days it was greater than 38. The resulting measurements are shown in the table

35.183	32.712	38.561	35.509	35.565	35.633	37.587	35.962
35.850	39.291	36.696	35.768	37.293	36.434	36.204	36.559
33.911	34.883	36.780	38.081	38.867	34.332	36.092	38.169
39.542	36.557	38.414	36.146	35.762	34.237		

The measurements in the table can be created with MATLAB command (http://octave-online.net/)

```
>> a=normrnd(36,2,1,30)
```

The frequency table is shown in Figure 2 (Top) with the corresponding histogram (Left), the histogram can be obtained with MATLAB command

```
>> hist(a,6)
```

The curve equation in the figure for PDF must be defined in order to be able to calculate the probability for any period. A question such as, what is the probability that the patients temperature will fall between 36.214 and 37.369 ? The curve equation in this case is called bell curve. The curve follows Normal Distribution (one of the common continuous distribution), its PDF or the curve equation can be defined as

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

To draw the curve we don’t need the entire data (since we know that it is normally distributed), we need to know only the data’s mean and standard deviation. The mean of the patients data is 36.419 and standard deviation is 1.635. The following MATLAB code can be used to plot the curve in Figure 2 (right).

```
>> m=mean(a)
```

```
      m =  36.419
>> sigma=std(a)
      sigma =  1.6357
>> x = (-5 * sigma:0.01:5 * sigma) + m;  %// Plotting range
>> y = exp(- 0.5 * ((x - m) / sigma) .^ 2) / (sigma * sqrt(2 * pi));
>> plot(x,y,'b','linewidth',3)
```

Class center	33.281	34.419	35.558	36.696	37.835	38.973
frequency	1	4	9	7	4	5
Probability	1/30=0.03	4/30=0.13	9/30=0.30	7/30=0.23	4/30=0.13	5/30=0.16

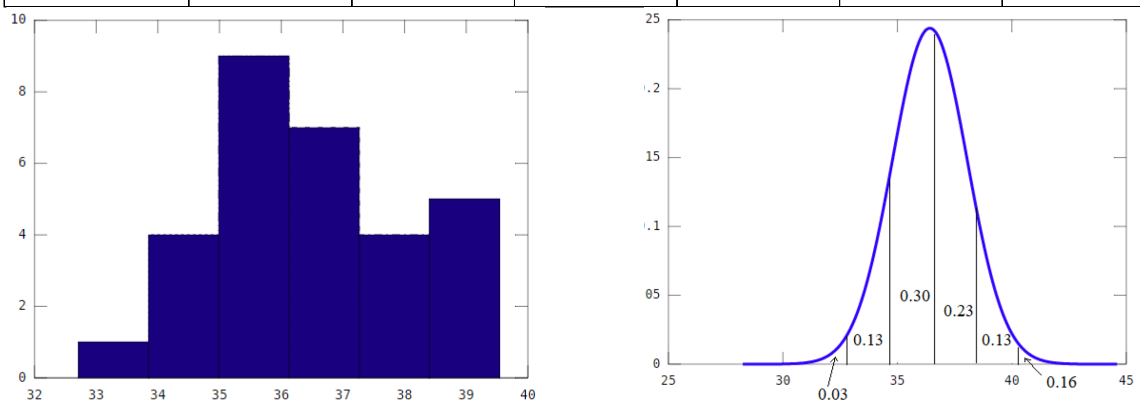


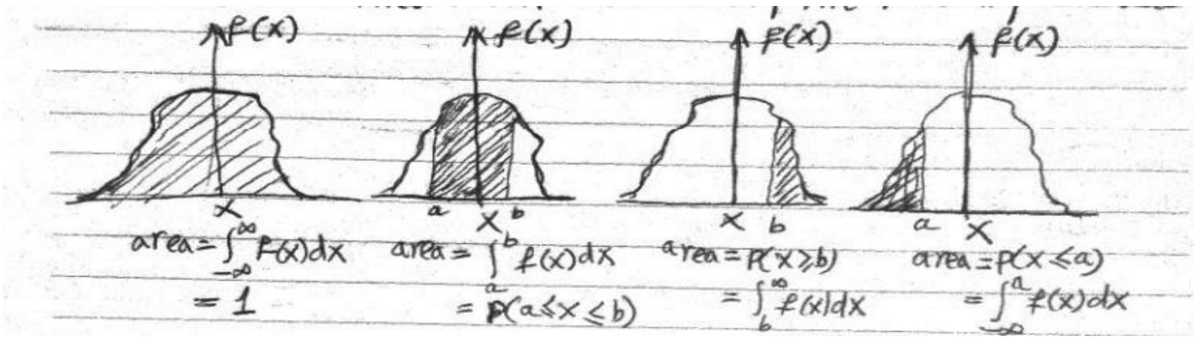
Figure 2: (Top) Frequency table of the patients data, (Left) The histogram of the patients data, (Right) The standard normal curve of the patients data

If for example, you are asked to calculate the probability of getting patient's temperature between a , b . this is an integration problem and must be solved using integration. So area under the curve must be calculated.

$$P\,(a\ <\ x\ < b) = \int_{a}^{b} f\,(x)dx$$

It now clear that *f(x)* will be defined by equation (1) only if the data follows Normal Distribution. The probability distribution or probability density function (PDF) for random variable x is defined as a set of real numbers if

1. $f(x) \geq 0\ \forall\ x \in R\ (R\ is\ real\ numbers\ set)$
2. $\int_{-\infty}^{\infty} f\,(x)dx = 1$
3. $P(\,a < \ x\ < b\,) = \int_{a}^{b} f\ (x)dx\ \forall\ a, b\ \in R\ ; a\ \leq b$



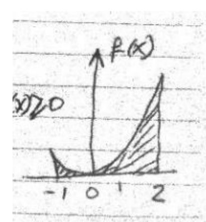
Example : Suppose the error in temperature C° for lab experiment is continues random variable x having the following probability density function .

$$f\ (x) = \begin{cases} 1/_3\ x^2\ , & -1\ < x\ < 2 \\ 0, & elsewhere \end{cases}$$

- Verify that $f(x) \geq 0$, And $\int_{\infty}^{\infty} f\ \ (x)\ dx = 1$
- Find $P\ (\ 0\ < x\ \leq 1\ )$

Solution :

- From the drawing , F (x) is quadratic so f (x) > 0



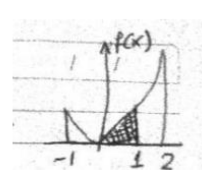
$$\int_{-\infty}^{\infty} f\ \ (x)dx = \int_{-\infty}^{-1} 0\ dx + \int_{-1}^{2} \frac{1}{3}\ x^2\ dx + \int_{2}^{\infty} 0\ dx$$

$$= \int_{-1}^{2} \frac{1}{3}\ x^2\ dx = \frac{1}{9}\ x^3\ |_{-1}^{2}$$

$$= \frac{1}{9}\ (\ 8 - (\ -1\ )) = 1$$

- $P\ (0 < x \leq 1\ ) = \int_0^1 f\ (x)dx = \int_0^1 \frac{1}{3}\ x^2\ dx$

$$= \frac{1}{9}\ x^3 \quad |_0^1 = \frac{1}{9}\ \big(\ 1 - (0)\big) = \frac{1}{9}$$



Another distribution is CDF for continuous random variable, CDF is calculated by

$$F\ (x) = p\ (X\ \ \leq x\ ) = \int_{-\infty}^{x} f\ (t)dt\ ;\ \ for - \infty\ < x\ < \infty$$

Then

$$p\ (\ a <\ \ X < b\ ) = P\ (\ X < b\ ) - P\ (X\ \ < a\ ) = f\ (b) - f(a)$$

Example:

- Find CDF for example 1
- Using CDF, find $P(\ 0 < x\ \ < 1)$

Solution:

- $f(x) = \int_0^{(1/3)x2} \quad \begin{matrix} -1<x<2 \\ elsewhere \end{matrix}$

$$\Rightarrow for\ x < -1, f\ (x) = \int_{-\infty}^{x} f\ (t)dt = \int_{-\infty}^{x} 0dt = 0$$

$$\Rightarrow for - 1\ < x\ < 2$$

$$f(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{-1} 0dt + \int_{-1}^{x} \frac{1}{3}\ t^2 dt$$

$$= \frac{1}{9}\ t^3\ \Big|_{-1}^{x} = \frac{1}{9}\ \big(\ x^3 - (\ -1\ )\big) = \frac{1}{9}\ (x^3+1)$$

$$\Rightarrow for\ x\ > 2$$

$$f(x) = \int_{-\infty}^{x} f\ (t)dt = \int_{-\infty}^{-1} 0dt\ + \int_{-1}^{2} \frac{1}{3}\ t^2\ dt + \int_{2}^{x} 0dt$$

$$= \int_{-1}^{2} \frac{1}{3}\ t^2\ dt = \frac{1}{9}\ t^3\ \Big|_{-1}^{2} = \frac{1}{9}\ \big(\ 8 - (-1\ )\big) = 1$$

And then

$$f(x) = \begin{cases} 0 & , \quad x < -1 \\ \frac{1}{9}(x^3+1\ ) & , \quad -1\ \le x\ < 2 \\ 1 & , \quad x \ge 2 \end{cases}$$

- $P\ (\ 0 < x\ \le 1\ ) = F\ (1) - F\ (0)$

$$= \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

4.4 Joint Probability Distribution

We studied single random variable, however there will be situations where we may find it desirable to record the simultaneous outcomes of several random variables. For example, we might measure the amount of precipitate *P* and volume *V* of gas released from a controlled chemical experiment, giving rise to a two-dimensional sample space.

In the discrete case, The function *f*(*x*, *y*) is a *joint probability distribution* or *probability mass function* of the discrete random variables *X* and *Y* if

1. $f(x,y) \ge 0$ for all (x,y),
2. $\sum_x \sum_y f(x,y) = 1$,
3. $P(X=x, Y=y) = f(x,y)$.

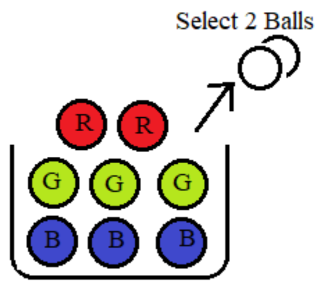
Example: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If *X* is the number of blue pens selected and *Y* is the number of red pens selected, find

(a) the joint probability function $f(x, y)$,
(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y)|x + y \leq 1\}$.
<u>Solution</u> **:** The possible pairs of values (x, y) are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0).

(a) Now, $f(0, 1)$, for example, represents the probability that a red and a green pens are selected. The total number of equally likely ways of selecting any 2 pens from the 8 is $\binom{8}{2}= 28$. The number of ways of selecting 1 red from 2 red pens and 1 green from 3 green pens is $\binom{2}{1}\binom{3}{1}= 6$. Hence, $f(0, 1) = 6/28 = 3/14$. Similar calculations yield the probabilities for the other cases, which are presented in Table 3.1. Note that the probabilities sum to 1.

Table : Joint Probability Distribution

$f(x,y)$		$x=0$	$x=1$	$x=2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1



(b) The probability that (X, Y) fall in the region A is

$$P[(X, Y) \in A] = P(X + Y \leq 1) = f(0, 0) + f(0, 1) + f(1, 0)$$
$$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}.$$

When X and Y are continuous random variables The function $f(x, y)$ is a <u>*joint density function*</u> of the continuous random variables X and Y if

1. $f(x,y) \geq 0$, for all (x,y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\ dx\ dy = 1$,
3. $P[(X,Y) \in A] = \int \int_A f(x,y)\ dx\ dy$, for any region A in the xy plane.

<u>Example:</u> A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of the joint density function definition.
(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < 0.5 , 0.25 < y < 0.5\}$.
<u>Solution</u>:
(a) The integration of $f(x, y)$ over the whole region is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dx\ dy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)\ dx\ dy$$
$$= \int_0^1 \left(\frac{2x^2}{5}+\frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1} dy$$
$$= \int_0^1 \left(\frac{2}{5}+\frac{6y}{5}\right) dy = \left(\frac{2y}{5}+\frac{3y^2}{5}\right)\Bigg|_0^1 = \frac{2}{5}+\frac{3}{5} = 1.$$

(b) To calculate the probability, we use

$$P[(X,Y)\in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$
$$= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)\ dx\ dy$$
$$= \int_{1/4}^{1/2} \left(\frac{2x^2}{5}+\frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2}\left(\frac{1}{10}+\frac{3y}{5}\right) dy$$
$$= \left(\frac{y}{10}+\frac{3y^2}{10}\right)\Bigg|_{1/4}^{1/2}$$
$$= \frac{1}{10}\left[\left(\frac{1}{2}+\frac{3}{4}\right)-\left(\frac{1}{4}+\frac{3}{16}\right)\right] = \frac{13}{160}.$$

4.5 Mean of Random Variable

Let X be random variable with probability distribution f(x). the mean or expected value of x is denoted by μ_x or E(x) is defined by:

$$E(x) = \mu_x = \begin{cases} \sum_{all\ x} x\, f(x) & if\ x\ is\ diserete \\ \int_{-\infty}^{\infty} x\, f(x)dx\ ; & if\ x\ is\ contions \end{cases}$$

If x defined by function g(x) then

$$E(g(x)) = \mu_{g(x)} = \begin{cases} \sum_{all\ x} g(x)f(x); & if\ x\ is\ discrete \\ \int_{-\infty}^{\infty} g(x)f(x)\ ; & if\ x\ is\ continous \end{cases}$$

Example 1:

8 computers with 3 defective, for random selection of 2 computers the probability distribution of X where X is the number of defectives is given by .

X	**0**	**1**	**2**	**Total**
$f(x)$	$^{10}/_{28}$	$^{15}/_{28}$	$^{3}/_{28}$	1

Find the expected value of the number of defectives X

Solution : $E(x) = \mu_x = \sum_{x=0}^{2} x\ f\ (x)$

$$= (0)\times\frac{10}{28}+(1)\times\frac{15}{28}+(2)\times\frac{3}{28}$$

$$= \frac{15}{28} + \frac{6}{28}$$

$$= \frac{21}{28} = 0.75\ computers$$

Example2: let X represents life in hours of a lamp, the PDF of X is given by

$$f\ (x) = \begin{cases} \dfrac{20{,}000}{x^3} & ;\ x\ > 100 \\ 0, & elsewhere \end{cases}$$

Find the expected life of the lamps

Solution: $E(x) = \ \mu_x = \ \int_{-\infty}^{\infty} x\ f\ (x)dx$

$$= \int_{100}^{\infty} x\ \ \frac{20{,}000}{x^3}\ dx = 20{,}000\ \int_{100}^{\infty} \frac{1}{x^2}\ dx$$

$$=\ 20{,}000\ \left[-\frac{1}{x}\ |_{x=100}^{x=\infty}\ \right] = -20{,}000\left[0 - \frac{1}{100}\right] = 200\ hours$$

We expect the lamp to last for 200 hours .

Example3 : for example 2 suppose X is a function g(x) defined by $g(x) = \frac{1}{x}$, find the expected value .

Solution : $E\ \left(g(X)\right) = E\ \left(\frac{1}{x}\right) = \ \int_{-\infty}^{\infty} g(x)f\ (x)dx = \ \int_{-\infty}^{\infty} \frac{1}{x}\ f(x)dx$

$$= \int_{100}^{\infty} \frac{1}{x} * \ \frac{20{,}000}{x^3}\ dx = \ \int_{100}^{\infty} \frac{1}{x^4} *\ 20{,}000\ dx = \ \frac{20{,}000}{-3}\ [\frac{1}{x^3}\ |_{x=100}^{x=00}]$$

$$= \ \frac{-20{,}000}{3}\ \left[\ 0 - \ \frac{1}{1{,}000{,}000}\right] = 0.0067\ (\ hours\ )$$

Example 4 **:** for example 1 , Suppose X is represented by function $g\ (x) = (\ x - 1)^2$

Solution : $E\ \left(\ g(x)\right) = \ \mu_{g(x)} = \ \sum_{x=0}^{2} g\ (x)f(x) = \ \sum_{x=0}^{2}(x - 1)^2\ f\ (x)$

$$= (0 - 1)^2\ f(0) + (1 - 1)^2\ f\ (1) + (2 - 1)^2\ f(2)$$

$$= (-1)^2\ \frac{10}{28} + (0)^2\ \frac{15}{28} + (1)^2\ \frac{3}{28}$$

$$= \ \frac{10}{28} + 0 + \ \frac{3}{28} = \ \frac{13}{28}\ \#\ computers$$

4.6 Variance of Random Variable

For two different datasets, the mean or expected value maybe equal, but the dispersion of data may differ, so we use another metric called variance of random variable is defined by :

$$Var\,(x) = \partial_x^2 = E\,(x-\mu)^2 = \begin{cases} \sum_x (x-\mu)^2 f(x)\ ;\ x \text{ is discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx\ ;\ x\ \text{continous} \end{cases}$$

The standard divination $\partial_x = \sqrt{\partial_x^2}$

Example1: Two companies A , B are using random number of cars in a given working day. suppose X is random variable of the number of cars . the distributions of x for company A and B is shown

Company A

x	**1**	**2**	**3**
F(x)	0.3	0.4	0.3

Company B

x	**0**	**1**	**2**	**3**	**4**
F(x)	0.2	0.1	0.3	0.3	0.1

Show that variance of company A data is greater than variance of company B data

Solution: For company A**:**

$$\mu A = E\,(x) = \sum_{x=1}^{3} x\,f\,(x) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0$$

$$\partial_A^2 = Var\,(x) = E\,[(x-\mu A)^2] = \sum_{x=1}^{3} (x-2\,)^2 f(x)$$

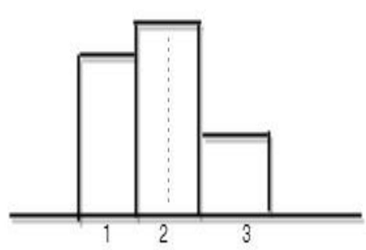
$$= (1-2\,)^2\,(0.3) + (2-2)\,^2(0.4) + (3-2)^2(0.3) = 0.6$$

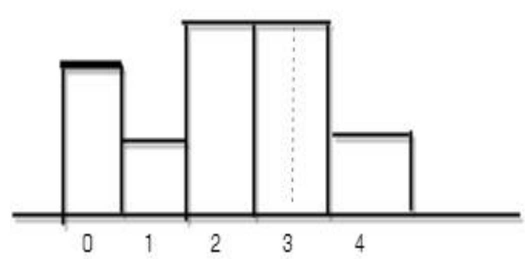
For company B:

$$\mu B = E\,(x) = \sum_{x=0}^{4} x\,f(x) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2$$

$$\partial_B^2 = Var\,(x) = E\,[(x-\mu B)^2]$$

$$= \sum_{x=0}^{4} (x-2\,)^2 f(x) = (0-2)^2\ (0.2) + (1-2)^2\,(0.1) + (2-2)^2\,(0.3) + (3-2)^2\,(0.3) + (\,4-2)^2\,(0.1) = 1.6$$





Example 2**:** The weekly demand for Pepsi in thousands of litters is continues random variable X has the following PDF.

$$f(x) = \begin{cases} 2(x-1) & ;\ 1< x <2 \\ 0 & ;elsewhere \end{cases}$$

Find the mean and variance of X

Solution:

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

$= \int_{1}^{2} x\, 2(x-1)dx = 2\int_{1}^{2} x^2 - x\, dx = 2\left[\frac{x^3}{3} - \frac{x^2}{2}\ \Big|_{x=1}^{x=2}\right] = \frac{5}{3}$

To calculate variance the following compact formula is used

$$\boxed{Var\,(x) = \partial_x^2 = E\,(\,x^2\,) - \mu^2}$$

$E(x^2) = \int_{-\infty}^{\infty} x^2\, f\,(x)dx = 2\int_{1}^{2} x^2\,(x-1)dx = 2\left[\frac{x^4}{1} - \frac{x^3}{3}\ \Big|_{x=1}^{x=2}\right] = \frac{17}{6}$

$$Var\,(x) = \partial_x^2 = E\,(\,x^2\,) - \mu^2 = \frac{17}{6} - (\frac{5}{3})^2 = \frac{1}{8}$$

4.7 Mean & Var of linear Combination of Random Variables

$if\ x_1, x_2, x_3, \dots\dots\dots . x_n$ are n random variables and a_1, a_2 , …. , a_n are constants , then the random variable .

$$X = \sum_{i=1}^{n} a_i\, x_i = a_1\, x_1 + a_2\, x_2 \dots\dots\dots . + a_n\, x_n$$

Is called *linear combination*, it has the following properties.

1. $E\,(\,a\,X \pm b\,) \quad = \quad aE\,(X) \pm b$
2. $E\,(\,X_1\ \pm\ X_2\,) = E\,(\,X_1\,) + E\,(\,X_2)$
3. $Var\,(aX\ \pm b\,) = a^2\, Var\,(X)$
4. $Var\,(\ aX_1\ \pm\ bX_2) = a^2\, Var\,(X_1\,) + b^2\, Var\,(\,X_2\,)$

Independent X_1, X_2

Example 1 **:** Let X be a random variable with the following PDF

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

Find E (4 X + 3)

Solution: $\mu = E(X) = \int_{-\infty}^{\infty} x\, f\,(x)dx = \int_{-1}^{2} x\ \left[\frac{1}{3}\, x^2\right] dx$

$$= \frac{1}{3}\int_{-1}^{2} x^3\, dx$$

$$= \frac{1}{3}\left[\frac{1}{4}\, x^4\ \Big|_{x=-1}^{x=2}\right] = {}^{5}/_{4}$$

E(4X + 3) = 4 E(X) + 3 = 4(5/4) + 3 = 8

Another Solution**:** $E[g\,(x)] = \int_{-\infty}^{\infty} g(x) f\,(x)dx \ \ ; g(x) = 4x + 3$

$E(\,4X + 3) = \int_{-\infty}^{\infty} (\,4\,x + 3\,) f\,(x)dx$

$$= \int_{-1}^{2} (4x+3)\left[\frac{1}{3}x^2\right] dx$$

$$= \int_{-1}^{2} \frac{4}{3}x^3 + x^2\, dx = \left[\frac{x^4}{3} + \frac{x^3}{3}\right]\Big|_{-1}^{2}$$

$$= \frac{16}{3} + \frac{8}{3} - \left(\frac{1}{3} - \frac{1}{3}\right) = \frac{24}{3} - 0 = 8$$

Example 2**:** if X, Y are independent random variables with Var(X) = 2 Var (Y) = 4

$$Z = 3X - 4Y + 8$$

Find Var(Z).

Solution: $Var\,(z) = Var\,(3X - 4Y + 8)$

$$= (3)^2\, Var\,(X) - (4)^2\, Var\,(Y) + 0$$

$$= 9 \times 2 + 16 \times 4 = 82$$

Example 3**:** if X, Y are independent random variables with

$$E\,(X) = 2 \quad Var\,(X) = 4 \quad Var\,(Y) = 1 \quad E(Y) = 7$$

Find

- $E\,(3X + 7)$,$Var\,(3X + 7)$
- $E\,(5X + 2Y - 2)$, $Var\,(5X + 2Y - 2)$

Solution:

- $E\,(3X + 7) = 3E(X) + 7 = 3 \times 2 + 7 = 13$

  $Var(3X + 7) = (3)^2\, Var\,(X) = 9 \times 4 = 36$
- $E(5X + 2Y - 2) = 5\,E\,(X) + 2\,E\,(Y) - 2$

  $= 5 \times 2 + 2 \times 7 - 2 = 22$

  $Var\,(5 \times +2Y - 2) = (5)^2\; Var\,(X) + (2)^2 Var\,(Y)$

  $= 25 \times 4 + 4 \times 1 = 104$ #

4.8 Mean and Covariance of Two Random Variables

Let X and Y be random variables with *joint probability distribution* $f(x, y)$. The **mean**, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y) f(x,y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\; dx\; dy$$

Dependent X , Y

if X and Y are continuous.

Example: Let X and Y be the random variables with joint probability distribution indicated in The Table. Find the expected value of $g(X, Y) = XY$.

	$f(x, y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

***Solution* :** By the Definition we write

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x, y)$$

$= (0)(0)f(0,0) + (0)(1)f(0, 1)$
$+ (1)(0)f(1, 0) + (1)(1)f(1,1) + (2)(0)f(2, 0)$

$$= f(1, 1) = \frac{3}{14}.$$

Let X and Y be random variables with joint probability distribution $f(x, y)$. The **covariance** of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_y) f(x, y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_y) f(x, y)\, dx\, dy$$

if X and Y are continuous.
The covariance between two random variables is a measure of the nature of the association between the two. The *sign* of the covariance indicates whether the relationship between two dependent random variables is positive or negative. When X and Y are statistically independent, it can be shown that the covariance is zero.

- *Positive sign* : Large values of X result in large values of Y and vice versa
- *Negative sign*: Large values of X result in small values of Y and vice versa
- *Zero*: X, Y are statistically independent.

The alternative and preferred formula for σXY The covariance of two random variables X and Y with means μX and μY , respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$

Example prev. example describes a situation involving the number of blue refills X and the number of red refills Y. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

	$f(x,y)$	x = 0	1	2	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y .

Solution **:** From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

And

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Exercises4

Question1: A shipment of 11 similar microcomputers to a retail outlet contains 4 that are defective. If a school makes a random purchase of 3 of these computers, find

a. the probability distribution of number of defectives
b. The expected value of number of defectives

Question2: For the following distribution of a random variable (5 Marks)

x	0	1	2
$f(x)$	$\frac{12}{25}$	$\frac{10}{25}$	$\frac{3}{25}$

a) Find the cumulative distribution function (CDF)
b) Find the expected value of the random variable E[x]

Question3: If x and y are independent random variables with: $\mu_x = 3$, $\mu_y = 6$, $\sigma^2_x = 1$, $\sigma^2_y = 4$:

- Find E(3X+7), Var(5X+2y-2).

Question4: The probability density function of X, the lifetime of a certain type of electronic device (measured in hours), is given by f(x) as shown. *Find E(1/X).*

$$F(x) = \begin{cases} \frac{10}{X^2}. & ;\ X > 10 \\ 0 \ ; & \end{cases}$$

Question5: The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{200}{(x+100)^2}, & x > 0, \\ 0, & elsewhere. \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of
(a) anywhere from 80 to 120 days.

Question6: The proportion of people who respond to a certain mail-order solicitation is a random variable X having the density function given in

$$f(x) = \begin{cases} \frac{2(x+3)}{7}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of X.

Question7: Complete the following sentences

1. The mean of discrete random variable can be calculated with ……………………….
2. The Variance of discrete random variable can be calculated with ……………………
3. The expected value of discrete random variable is calculated with ……………………
4. The standard deviation of the discrete random variable can be calculated with ………………………..

Question8: If x is a random variable with pdf $f(x) = x \quad , 1 < x < 2$ Find Var (x)

Question9: If X is a random variable with the following distribution

X	1	2	3	4	5
$f(X)$	c	2c	3c	c	3c

1- Find the value of c
2- Find E(X)

Multiple Choices Questions:

1) You are given the following probability distribution:

x	0	1	2	3	4
f(x)	0.2	0.1	0.3	0.3	0.1

What is the variance of X?

a. 0.60
b. 2.03
c. 1.59
d. 0.94
e. 0.48

2) random variable is said to be discrete if:

a) its outcomes are countable
b) it can assume any real number within an interval
c) its outcomes are uncountable
d) None of above

3) Which of the following cannot be a probability?

e) a) -1 b) 0 c) 1 d) .05

Answers

Question1: A shipment of 11 similar microcomputers to a retail outlet contains 4 that are defective. If a school makes a random purchase of 3 of these computers, find

a. the probability distribution of number of defectives
b. The expected value of number of defectives

(a) اجمالى 11 → 4D → 7ND

الحل اختيار 3 من 11 بشكل عام يعطى الأشكال الآتية

$$\binom{11}{3} = \frac{11!}{8! \times 3!} = \frac{11 \times \cancel{10}^{5} \times \cancel{9}^{3} \times \cancel{8!}}{\cancel{8!} \times \cancel{6}} = 11 \times 5 \times 3 = 15 \times 11 = 165$$

for $\underline{x=0}$ سنختار كل المنتجات سليمة من 7 السليمة

$$\binom{7}{3}\binom{4}{0} = \frac{7!}{3! \times 4!} = \frac{7 \times \cancel{6} \times 5 \times \cancel{4!}}{\cancel{6} \times \cancel{4!}} = 35$$

for $\underline{x=1}$ سنختار واحد من العاطبة و الباقى من السليم

$$\binom{7}{2}\binom{4}{1} = \frac{7!}{5! \times 2!} \times 4 = \frac{7 \times \cancel{6}^{3} \times \cancel{5!}}{\cancel{5!} \times \cancel{2}} \times 4 = 21 \times 4 = 84$$

for $\underline{x=2}$

$$\binom{7}{1}\binom{4}{2} = 7 \times \frac{4!}{2! \times 2!} = \frac{4 \times 3!}{4} = 3! \times 7 = 3! \times 7 = 6 \times 7 = 42$$

for $\underline{x=3}$

$$\binom{7}{0}\binom{4}{3} = 1 \times \frac{4!}{1! \times 3!} = \frac{4 \times \cancel{3!}}{\cancel{3!}} = \underline{4}$$

then distribution for x

X	0	1	2	3
$F(x)$	$\frac{35}{165}$	$\frac{84}{165}$	$\frac{42}{165}$	$\frac{4}{165}$

(b)

$$E(x) = 0 \times \frac{35}{165} + 1 \times \frac{84}{165} + 2 \times \frac{42}{165} + 3 \times \frac{4}{165}$$

$$= \frac{84}{165} + \frac{84}{165} + \frac{12}{165} = \frac{180}{165}$$

1- If X is a random variable with pdf $f(x) = x \quad 1 \le x \le 3$. Find $E(X)$

$f(x) = x$

$$E(x) = \int_1^3 x f(x)\, dx = \int_1^3 x \cdot x\, dx$$

$$= \int_1^3 x^2\, dx$$

$$= \left.\frac{x^3}{3}\right|_1^3 = \frac{(3)^3}{3} - \frac{(1)^3}{3}$$

$$= \frac{3 \times 9}{3} - \frac{1}{3} = \frac{3 \times 9 - 1}{3} = \frac{27-1}{3} = \frac{26}{3}$$

Chapter 5
Discrete Distributions

5.1 Discrete Distributions

In this section, we study the following discrete probability distributions.

- Discrete uniform distribution.
- Binomial distribution.
- Hypergeometric Distribution
- Poisson Distribution

5.1.1 Discrete Uniform Distribution

Assume random variable X with values $x1, x2, x3, x4$ ... ... with equal probabilities, then X has the discrete uniform distribution given by .

$$f\ (x) = p(X = x) = \begin{cases} 1/_k \ , x_1\ , x_2 \ldots\ldots\ldots x_k \\ 0\ , elsewhere \end{cases}$$

k is the distribution parameter, We can easily get the mean and variance of this distribution this distribution can be written in the following table.

x	$\boldsymbol{x_1} \quad \boldsymbol{x_2} \quad \boldsymbol{x_3} \quad \ldots \quad \ldots \ldots \boldsymbol{x_k}$
F(x)	$1/_k \quad 1/_k \quad 1/_k \ldots \ldots \ldots .\ 1/_k$

$$\mu = \sum_x x\ f\ (x) = \sum_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k} \sum_{i=1}^{k} x_i$$

$$\partial^2 = \sum_{i=1}^{k} (x_i - \mu)^2\ f\ (x) = \sum_{i=1}^{k} (x_i - \mu)^2\ \frac{1}{k} = \frac{1}{k} \sum_{i=1}^{k} (x_i - \mu)^2$$

Example1 : For tossing a balanced die the sample space is $S = \{\ 1,2,3,4,5,6\}$

$$p(x_1) = p(1) = \frac{1}{6}, p(x_2) = p\ (x_3) = p(x_4) = p\ (x_5) = p\ (x_6) = \frac{1}{6}$$

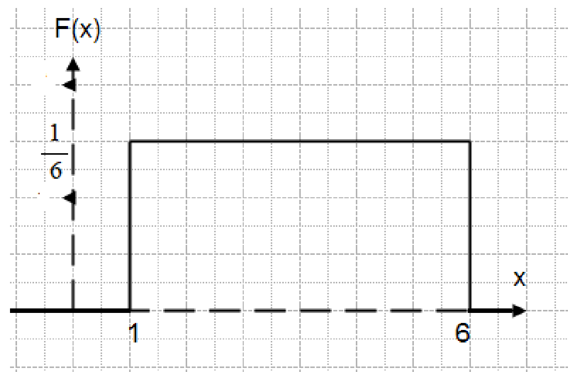
$$\mu = \frac{1}{k} \sum_{i=1}^{k} x_i = \frac{1}{6} \times (\ 1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

$$\partial = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

$$= \frac{1}{6}\,[\,(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2] = \frac{35}{12}$$

The resulting distribution is as follow

x	1	2	3	4	5	6
F(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



5.1.2 Binomial Distribution

Bernoulli trial: is an experiment with only two possible outcomes, namely Success, & Failure

P (Success)= p, P (Failure) = 1 – p = q

As an example, tossing a coin is a Bernoulli experiment with p (H) = 0.5 and P (T) = 1-0.5 = 0.5, repeating the above trials n times is called Bernoulli Process.

Binomial Random Variable: Consider the random variable x: is the number of successes in the n trials in Bernoulli process.

The random variable x has a binomial distribution with parameter n (number of trials) and P (probability of success as

$$f(x) = b\,(x; n, p) = \begin{cases} \binom{n}{x} P^x\,(1-p)^{n-x} & ,x=0\,,1\,,2\ldots n \\ 0\,, & otherwise \end{cases}$$

In a table it can be written as

x	0	1	……..	n – 1	n	Total
F(x)	$\binom{n}{0} P^0\,(1-p)^{n-0} = (1-p)^n$	$\binom{n}{1}\; P^1\,(1-p)^{n-1}$	……..	$\binom{n}{n-1} P^{n-1}\,(1-p)^0$	p^n	1

The mean & variance can be calculated as

$$\mu = np$$

$$\partial^2 = n\,p\,(1\,-p\,)$$

<u>Example 1:</u> Suppose that a chips box has 25% of the bags are winner bags, the child's mom lets him select three bags of the box. The child want to know the probability of getting no winner bags, one winner bag, two winner bags and three winner bags (i.e find the distribution)	

<u>Solution</u>: x is the number of winner bags. We need to find the probability distribution of x.

Probability of success = $= p = \frac{1}{4} = 0.25$

Probability of failure $= 1 \ - p = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

Number of trials n = 3, Then $f(x) = b\ (x; 3, \frac{1}{4}) = \begin{cases} \binom{3}{x}\left(\frac{1}{4}\right)^{x}\left(1-\frac{1}{4}\right)^{3-x} & x=0,1,2,3 \\ 0 & ,otherwise \end{cases}$

$$f(0) = b\left(0; 3, \frac{1}{4}\right) = \binom{3}{0}\left(\frac{1}{4}\right)^{0}\left(\frac{3}{4}\right)^{3} = \frac{27}{64}$$

$$f(1) = b\left(1; 3, \frac{1}{4}\right) = \binom{3}{1}\left(\frac{1}{4}\right)^{1}\left(\frac{3}{4}\right)^{2} = 3 \times \frac{1}{4} \times \frac{9}{16} = \frac{27}{64}$$

$$f(2) = b\left(2; 3, \frac{1}{4}\right) = \binom{3}{2}\left(\frac{1}{4}\right)^{2}\left(\frac{3}{4}\right)^{1} = 3 \times \frac{1}{16} \times \frac{3}{4} = \frac{9}{64}$$

$$f(3) = b\left(3; 3, \frac{1}{4}\right) = \binom{3}{3}\left(\frac{1}{4}\right)^{3}\left(\frac{3}{4}\right)^{0} = \frac{3!}{3!\ \times 0!} \times \frac{1}{64} \times 1 = 1 \times \frac{1}{64} \times 1 = \frac{1}{64}$$

$$\mu = np = 3\ \times \frac{1}{4} = \frac{3}{4}$$

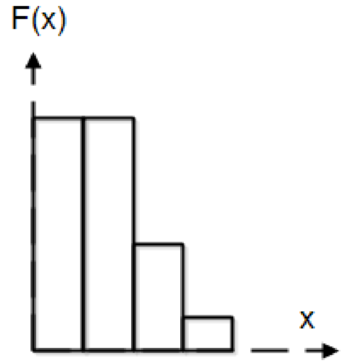
$$\delta^{2} = np\ (1 - p\ ) = 3 \times \frac{1}{4} \times \frac{3}{4} = \frac{9}{16}$$

<u>Example 2</u>**:** For the previous example find,

- Probability of getting at least two winner bags.
- Probability of getting at most two winner bags.

Solution**:** The previous example distributions is .

x	0	1	2	3
F(x)	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$



- $p(x \geq 2) = P\ (x\ = 2\ ) + P\ (\ x\ = 3\ ) = \frac{9}{64} + \frac{1}{64} = \frac{10}{64}$
- $p(x \leq 2) = P\ (x\ = 0\ ) + P\ (\ x\ = 1\ ) + P(x\ = 2\ ) = \frac{27}{64} + \frac{27}{64} + \frac{9}{64} = \frac{63}{64}$

5.1.3 Hypergeometric Distribution

The hypergeometric distribution is used to calculate probabilities when sampling is done without replacement. For example, suppose selecting a card from deck of 52 cards, then selecting a second card without replacing the first and so on. suppose there is a population with 2 types of elements success & failure.

N: is the population size

K: number of elements of the (success) type in the population.

N – K: number of elements of the failure type in the population. A Sample of size n elements is selected at random from the population, X is the probability of getting x elements in the sample of size n

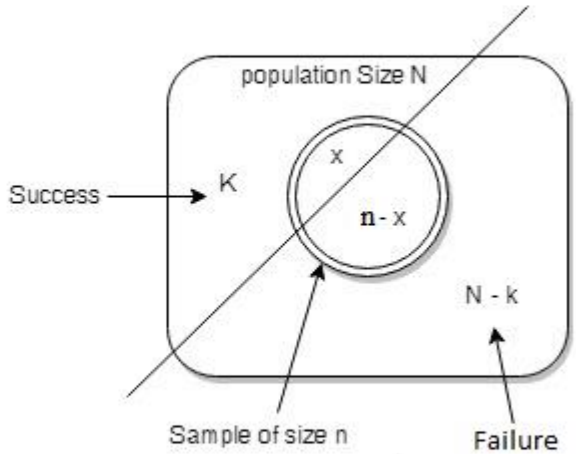
the probability distribution of X is hypergeometric and is obtained by;

$$f(x) = h\ (x, N, n\ , k\ ) = \begin{cases} \frac{\binom{k}{x} \times \binom{N-k}{n-x}}{\binom{N}{n}}, & x=0,1,2,\ldots\ldots n \\ 0 & otherwise \end{cases}$$

The hypergeometric distribution's mean and variance is respectively calculated by;

$$\mu = \frac{nk}{N}$$

$$\partial^2 = \sqrt{\frac{nk\ (\ N - k\ )(\ N - n)}{N^2\ (\ N - 1)}}$$



<u>Example 1</u>**:** Suppose you sampled 3 cards of a deck of 52 cards without replacement. what is the probability that exactly two of the drawn cards will be aces (4 of 52 are aces in the deck).

<u>Solution</u>: In this example, K = 4 because there are four aces in the deck, x = 2 because the problem asks about the probabilities of getting two aces , N = 52 is the entire cards . n=3 because 3 Cards are Sampled.

$$f(x) = \frac{\binom{4}{2}\binom{52-4}{3-2}}{\binom{52}{3}} = \frac{\frac{41}{2!\ 2!}\cdot\frac{48!}{47!\,1\,!}}{\frac{52!}{49!\,3!}} = 0.013$$

Use MATLAB to find the remaining values of x, https://octave-online.net/. The following function is used :

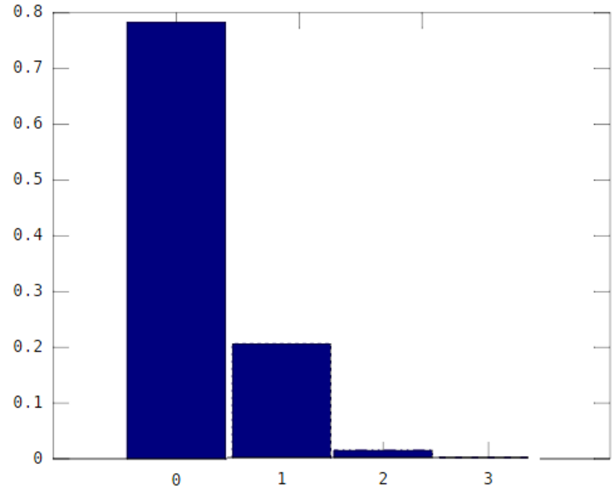
```
p = hygepdf(x values,N,K,n)
```

For the above example the following is used

```
>> x=0:3;
>> p = hygepdf(x,52,4,3)
>> bar(x,p)
```

The resulting distribution is shown in the figure

x	0	1	2	3
F(x)	0.782	0.204	0.013	0.001



Example 2: A chips box contains 40 chips bags, the child selects 5 bags randomly (without replacement) what is the probability of getting exactly one winning bag if the child knows that the entire box contains only 3 winning bags.

Solution : let x the number of winning bags in the sample .

$$N = 40,\;\; K = 3, \qquad n = 5\,,\;\; x = 1$$

$$f(x) = P\,(X = x\,) = \begin{cases} \frac{\binom{3}{1} \times \binom{37}{5-1}}{\binom{40}{5}} \\ 0 \quad otherwie \end{cases}$$

$$f(x) = \frac{\frac{3!}{1! \times 2!} \; \frac{37!}{4!\,.33!}}{\frac{40!}{5! \times 35!}} = 0.3011$$

5.1.4 Poisson Distribution

The Poisson random variable satisfies the following conditions:

- The number of successes in two disjoint time intervals is independent.
- The probability of a success during a small time interval is proportional to the entire length of the time interval.

Apart from disjoint time intervals, the Poisson random variable also applies to **disjoint regions of space**.

Application examples

- car accidents
- traffic flow and ideal gap distance
- number of typing errors on a page

The **probability distribution of a Poisson random variable** X representing the number of successes occurring in a given time interval or a specified region of space is given by the formula:

$$\boxed{P(X) = \frac{e^{-\mu}\mu^{x}}{x!}}$$

x=0,1,2,3…
e=2.71828 (but use your calculator's e button)
μ= mean number of successes in the given time interval or region of space
If μ is the average number of successes occurring in a given time interval or region in the Poisson distribution, then the mean and the variance of the Poisson distribution are both equal to μ.

$$E(X) = \mu$$

and

$$Var(X) = \sigma^2 = \mu$$

Note: In a Poisson distribution, only **one** parameter, μ is needed to determine the probability of an event.

Example 1: A life insurance salesman sells on the average 3 life insurance policies per week. Use Poisson's law to calculate the probability that in a given week he will sell

a) Some policies
b) 2 or more policies but less than 5 policies.
c) Assuming that there are 5 working days per week, what is the probability that in a given day he will sell one policy?

Solution: Here, μ = 3

(a) "Some policies" means "1 or more policies". We can work this out by finding 1 minus the "zero policies" probability: $P(X > 0) = 1 - P(x_0)$

$$\text{Now } P(X) = \frac{e^{-\mu}\mu^x}{x!} \text{ so } P(x_0) = \frac{e^{-3}3^0}{0!} = 4.9787 \times 10^{-2}$$

Therefore the probability of 1 or more policies is given by:

$$\begin{aligned}\text{Probability} &= P(X \geq 0)\\ &= 1 - P(x_0)\\ &= 1 - 4.9787 \times 10^{-2}\\ &= 0.95021\end{aligned}$$

(b) The probability of selling 2 or more, but less than 5 policies is:

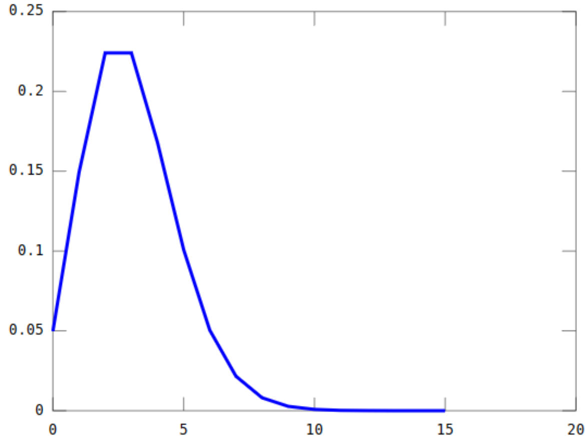
$$\begin{aligned}&P(2 \leq X < 5)\\ &= P(x_2) + P(x_3) + P(x_4)\\ &= \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} + \frac{e^{-3}3^4}{4!}\\ &= 0.61611\end{aligned}$$

(c) Average number of policies sold per day: 3/5=0.6

$$\text{So on a given day, } P(X) = \frac{e^{-0.6}(0.6)^1}{1!} = 0.32929$$

The distribution for selling policies/ week, when the average is 3 polices/week is obtained with MATLAB as follow

```
x = 0:15;
y = poisspdf(x,3);
plot(x,y,'linewidth',3)
```



Part (a) of the above example can be obtained with

```
y = poisscdf(0,3)
y =  0.049787
```

Part(b) of the example can be obtained with

```
y=poisscdf(4,3)-poisscdf(1,3)
y =  0.61611
```

Part(c) of the example can be obtained with

```
y=poisscdf(1,0.6)-poisscdf(0,0.6)
y =  0.32929
```

Example 2: If electricity power failures occur according to a Poisson distribution with an average of 3 failures every twenty weeks, calculate the probability that there will not be more than one failure during a particular week.

Solution: The average number of failures per week is: μ=3/20=0.15

"Not more than one failure" means we need to include the probabilities for " 0 failures" plus "1 failure".

$$P(x_0) + P(x_1) = \frac{e^{-0.15}0.15^0}{0!} + \frac{e^{-0.15}0.15^1}{1!} = 0.98981$$

Exercises:

1. A chips box contains 20 chips bags, the child selects 3 bags randomly (without replacement) what is the probability of getting exactly one winning bag if the child knows that the entire box contains only 2 winning bags.
2. Suppose that a chips box has 75% of the bags are winner bags. Child selects three bags from the box. Find the probability of getting at least two winner bags.
3. If X is random variable with Possisson distribution such that $P(X = 1) = P(X = 2)$Find $P(X = 4)$

Chapter 6
Continuous Distributions

5.2 Continuous Distributions

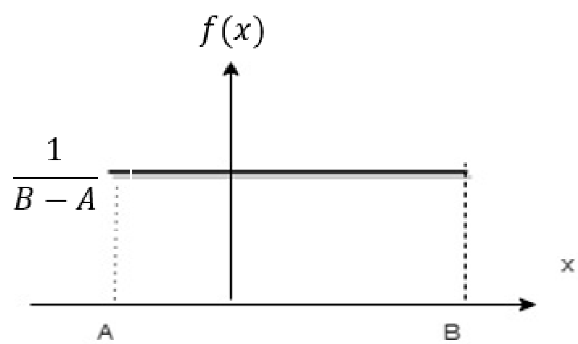
We study the following continuous probability distributions

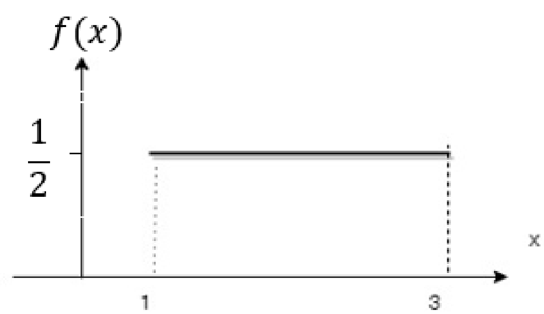
- Continuous uniform distribution.
- Normal distribution.

5.2.1 Continuous Uniform Distribution

The probability density function (PDF) of the continuous uniform random variable X on the interval [A, B] is given by :

$$f(x) = f\,(x, A\,, B\,) = \begin{cases} \frac{1}{B-A} & A \le x \le B \\ 0\,, & elsewhere \end{cases}$$





The mean and variance of the uniform continuous distribution is calculated by

$\mu = \frac{A+B}{2}$ is the distribution mean.

$\partial^2 = \frac{(B-A)^2}{12}$ is the distribution variance.

<u>Example1</u>: suppose that for a certain company, the conference time, X, has a uniform distribution on the interval [0, 4] hours .

- What is the probability density function of X?
- What is the probability that any conference. lasts at least 2 hours.

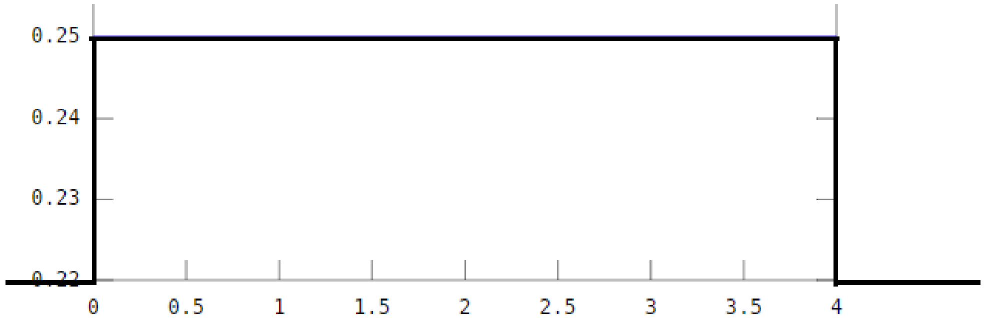
Solution**:**

- $f(x) = \begin{cases} \frac{1}{4-0} = \frac{1}{4} & 0 \le x \le 4 \\ 0\,, & elsewhere \end{cases}$
- $P\,(\,x \ge 2\,) = \int_2^4 f\,(x)dx = \int_2^4 \frac{1}{4}\,dx$

$$= \frac{x}{4} \Big|_{x=2}^{x=4} = \frac{1}{4}\,(\,4 - 2\,) = \frac{1}{4} \times 2 = \frac{2}{4} = \frac{1}{2}$$

MATLAB is used to plot the Probability Density Function (PDF), go to https://octave-online.net/

```
>> Y = unifpdf(0:4,0,4)
>> figure;
>> stairs(x,Y,'r','LineWidth',2);
>> hold on;
>> ylim([0 1]);
>> hold off;
```



5.2.2 Normal Distribution

Normal Distribution can describe data with central value μ and a fluctuation value called standard deviation. most of the data came around the mean for example the student height has μ = 174 cm and ∂ = 20 most of the height are near (above or below) 174 cm. the distribution shape is bell shaped. The probability density function (PDF) of the random variable X with normal distribution is given by.

$$f\,(x\,) = \frac{1}{\partial\sqrt{2\,\pi}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\partial}\right)^2}, \text{ where } \pi = 3.14 \text{ and } e = 2.71$$

The mean of normal distribution, E (x) = μ .

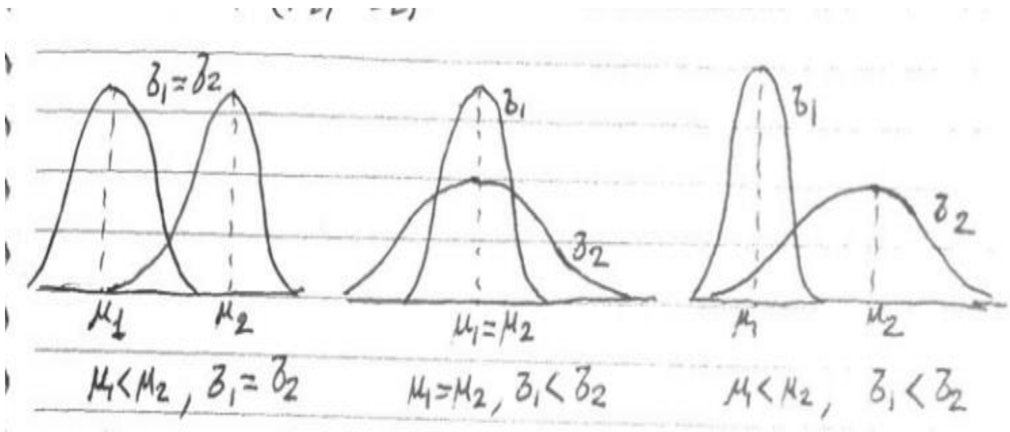
The variance of the normal distribution Var (x) = ∂^2

Two variables μ and = ∂^2 control the shape of the normal curve, suppose there are two distributions.

- $N\,(\mu_1, \partial_1)$
- $N\,(\mu_2, \partial_2)$

The properties of the normal curve are:

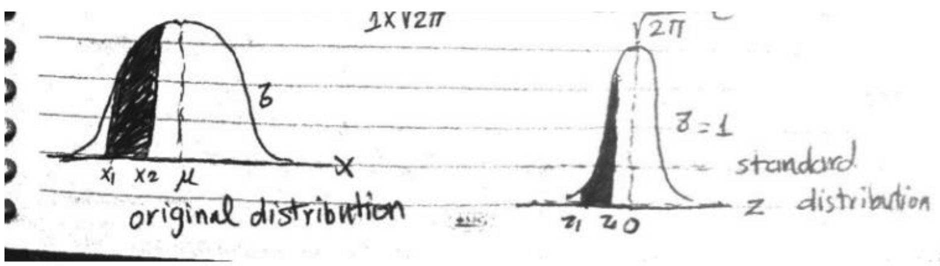
1. f(x) is symmetric about the mean μ .
2. f (x) has two points of inflection at $x = \mu \pm \partial$.
3. The total area under the curve of f (x) = 1 .
4. The highest point of the curve or f (x) at mean μ .



Standard Normal Distribution: Normal distribution with $\mu = 0$ and $\partial^2 = 1$ is called standard normal distribution and is denoted by N (0,1), We can get the standard Normal distribution by substituting $x = z$, $\mu = 0$, $\partial^2 = 1$, we then get

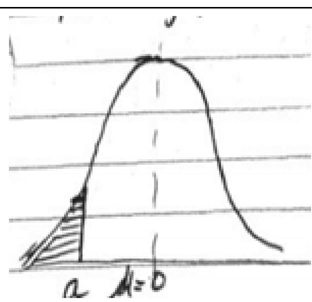
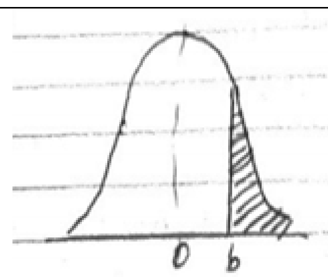
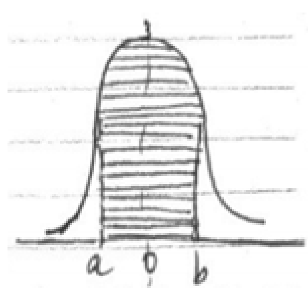
$$f\ (x\ ) = \frac{1}{\partial\sqrt{2\ \pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\partial}\right)^2}$$

$$f\ (x\ ) = \frac{1}{1\times\sqrt{2\ \pi}}\ e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\ \pi}}\ e^{-\frac{1}{2}z^2}$$



We can use the formula $z = \frac{x-\mu}{\partial}$, to convert between the original and standard distributions.

The standard Normal distribution Cumulative Probability Distribution (CDF) has one of the following forms.

		
$P(\ z \leq a\ )$ $\int_{-\infty}^{a} \frac{1}{\sqrt{2\ \pi}}\ e^{-\frac{1}{2}z^2}\ dz$	$P(z \geq b\ ) = 1 - p(z \leq b\ )$ $1 - \int_{-\infty}^{b} \frac{1}{\sqrt{2\ \pi}}\ e^{-\frac{1}{2}z^2}\ dz$	$P\ (a \leq z \leq b) = P(z \leq b) - P(z \leq a\ )$

		$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz$

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

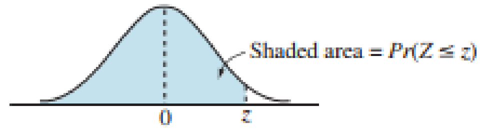


TABLE 1

Standard normal curve areas

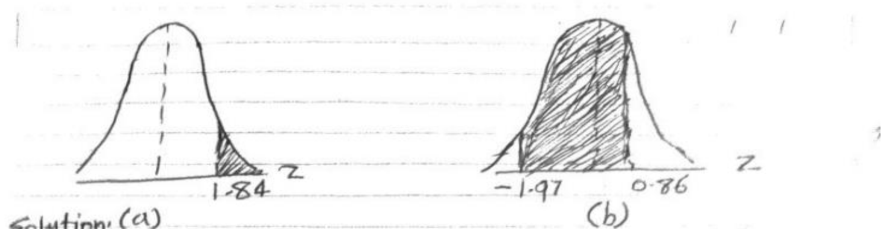
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Example 1**:** Given a standard normal distribution find the area under the curve that lies:

- To the right of z = 1.84.
- Between z = -1.97 and z = 0.86 as shown in the figure.

Solution**:**

- Normally, we must calculate $\int_{-\infty}^{1.84} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz$ but instead we can use the table A.3 to get the resulting value of the integration.
  Because the area to the right is needed, we have to get area to the left which is the CDF from table A.3 in figure (a). Then we subtract the CDF from 1 to get the right area. The area under curve to the right of Z = 1.84 is equal 1 – 0.9671 = 0.0329.
- The area in figure (b) between z = - 1.97 and z = 0.86 is equal to the area to the left of z 0.86 minus the area to the left of z = - 1.97. From table A.3 We find the desired area to be 0.8051 – 0.0244 = 0.7807.

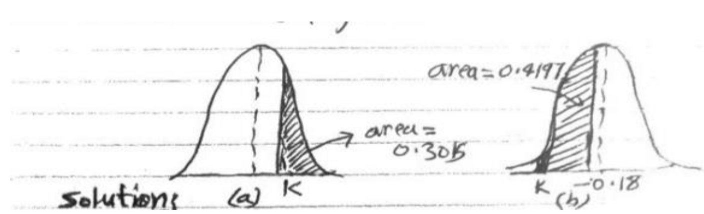


MATLAB can be used to solve the example above

```
>> left_area=normcdf(1.84)
 left_area =  0.96712
>> right_area=1-left_area
 right_area =  0.032884
>> a_left=normcdf(-1.97)
 a_left =  0.024419
>> b_left=normcdf(0.86)
 b_left =  0.80511
>> area=b_left-a_left
area =  0.78069
```

Example 2: given a standard normal distribution find the value of k such that:

- $P\,(z\ > k) = 0.3015$
- $P\,(\,k\ \leq z\ \leq\ -0.18\,) = 0.4197$, as shown in the figures.



Solution:

- In figure (a) we see that the value leaving an area of 0.3015 to the right must then leave an area of 0.6985 to the left. Then from table A.3 it follows that k= 0.52

- From table A.3 we note the total area to the left of – 0.18 is equal to 0.4286 , in figure (b) we see that the area between k and - 0.18 is 0.4197 so that the area to the left of K must be 0.4286 – 0.4197 = 0.0089 . Hence, from table A.3 We have k = - 2.37.

Example 3**:** Given a random variable X having a normal distribution with μ = 50 and ∂ = 10. Find the probability that X assumes a value between 45 and 62.

Solution**:**

The Z values corresponds to $x_1 = 45$, $x_2 = 62$ are

$$z_1 = \frac{45-50}{10} = -0.5 \qquad , z_2 = \frac{62-50}{10} = 1.2$$

Therefore

$$p\,(45 < x < 62) = p\,(-0.5 < z < 1.2) = P(z < 1.2) - P\,(z < -0.5)$$

$$= 0.8849 - 0.3085 = 0.5764$$

Example 4**:** Given normal distribution with μ = 40 and ∂ = 6 find the value of x that has .

- 45% of the area to the left.
- 14% of the area to the right.

Solution **:**

- We require a z value that leaves an area of 0.45 to the left , from table A . 3 We find $p\,(z < -0.13) = 0.45$ so that the desired z value is 0.13 Hence; $x = z\partial + \mu = -0.13 \times 6 + 40 = 39.22$
- We require z value that leaves an area of 1-0.14 = 0.86 to the left . form table A.3 we find that $P(z < 1.08) = 0.86$ so that the desired z value is 1.08 and $x = z\partial + \mu = 6 * 1.08 + 40 = 46.78$

Exercises 6

Question1: Given a standard normal dist., find the area under the curve that lies:(*Use the attached A.3 Table*)

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

(b) Between $z = 1.97$ and $z = 0.86$

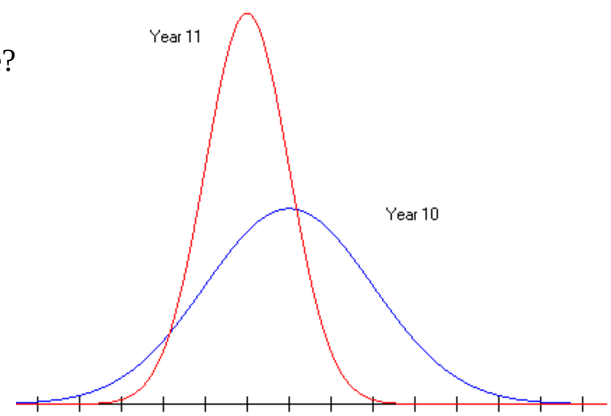
Question2: Suppose that, for a certain company, the conference time, X, has a uniform distribution on the interval [0,6] (hours).

(a) What is the probability density function of X?

(b)]What is the probability that any conference lasts at least 2 hours?

Multiple Choices Questions

1) The diagram shows two normal distribution curves, the scores achieved on an assignment by a group of Year 11 students, and the scores achieved on the same assignment by a group of Year 10 students.



Which one of the following sets of statements is true?

A) The mean score for the Year 11 students is higher than the mean score for the Year 10 students, but the variance of Year 11 more than the Year 10 marks.

B) The mean score for the Year 11 students is higher than the mean score for the Year 10 students, and the variance of Year 11 marks and Year 10 marks are equally variable.

C) The mean score for the Year 11 students is higher than the mean score for the Year 10 students, but the variance of Year 11 less than the Year 10 marks.

D) The mean score for the Year 11 students is lower than the mean score for the Year 10 students, but the variance of Year 11 less than the Year 10 marks.

2) Let x be a normal random variable with a mean of 50 and a standard deviation of 3.

A z score was calculated for x, and the z score is -1.2. What is the value of x?

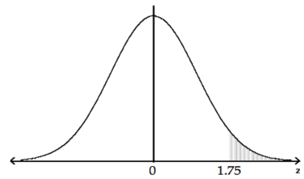
A. 53.6 B. 0.8849 C. 0.1151 D. 46.4

3) If X is a normally distributed random variable with a mean of 80 and a standard deviation of 12, then the $P(X \leq 68)$ is

A. .1587 B. .0000 C. .6587 D. .8413

4) The area of the following shaded area is:

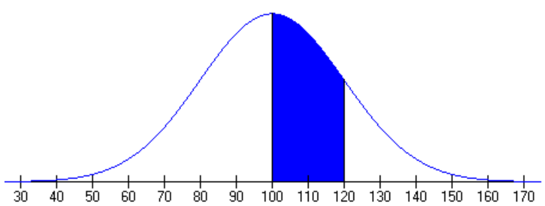
(a) 0.0228 (b) 0.9599 (c) 0.0401 (d) 0.0668



5) If the continuous random variable X is uniformly distributed over the interval [15,20] then the mean of X is:

a. 17.5
b. 15
c. 25.3
d. 35

6) If *X* is normally distributed random variable with mean 100 and standard deviation 20, and *Z* is the standard normal random variable, then the interval shaded in the diagram below can be written as:



A) $P(Z < 1)$
B) $P(Z > 100)$
C) $P(Z < 120)$
D) $P(100 < Z < 120)$

Chapter 7
Parameter Estimation

6.1 Central Limit Theorem

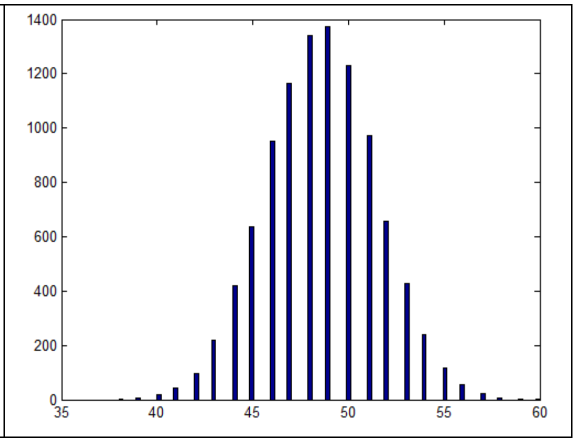
We have dataset of with unknown distribution. <u>Sampling:</u> means selecting sample from the dataset and study it to estimate the population (dataset) parameters. Parameters such as mean and standard deviation.

Sample Mean Distribution (Ignore Variance): Central limit theorem states that if you take random sample of size n and calculate the mean of the sample, and repeated this process many times you will find the calculated means for the sample take normal distribution shape (bell curve).

Example 1: MATLAB Based: Create random data of size 1000 for percentage of marks of 1000 students in MATH class. Take random sample of size n = 200. then calculate the mean. Take another sample then calculate the mean, repeat this process 100 times, then draw the histogram.

Answer: The code and the displayed histogram are shown in the figure

```
n=100; %sample size
N=10000; %Popoulation size
n_samples=10000; % number of samples
Data=fix(rand(1,N)*100);
X_=[     ]  ;
for i=1:n_samples
 Index=fix(rand(1,n)*N); %get rand index
 Index(Index==0)=1; % remove zero indies
 Sample=Data(Index); %get random Samples
 X_ = [ X_  fix(mean(Sample))];
end
hist( X_,100)
```



From the figure, we find that the histogram of the sample means is normally distributed.

Theory: The sample means distribution tend to be more normally distributed when sample size n becomes larger and larger.

- The mean of the sample means is the mean μ of the population shown in the figure.
- The standard deviation of the sampling distribution of means $\partial_x = \frac{\partial}{\sqrt{n}}$, and error percentage $z = \frac{\bar{x} - \mu}{\partial_x}$, generally the limiting form of the distribution $z = \frac{\bar{x} - \mu}{\frac{\partial}{\sqrt{n}}}$ with high value of n is a standard normal distribution.

Example 2: A certain brand of tires has mean life μ = 25,000 miles with standard deviation of 1600 miles. What is the probability that the mean life of 64 tires is less than 24,600 miles?

Answer: The population mean μ = 2500 mi, The population standard deviation ∂ = 1600 miles.

Sample standard deviation $\partial_x = \frac{\partial}{\sqrt{n}} = \frac{1600}{\sqrt{64}} = \frac{1600}{8}$ =200, so $\partial_x = 200$

Then $z = \frac{\overline{X} - \mu}{\partial_x} = \frac{24600-25000}{200} = -2$

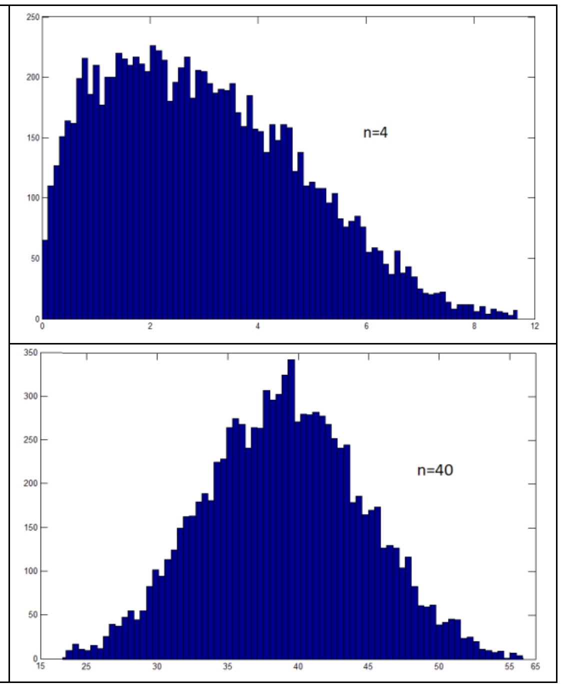
From table $A.3$ $P\ (z < -2) = 0.0228$, or 2.28 % of the sample means will be less than 24,600 mi.

Chi-Squared Distribution (Ignore Mean): If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom, The random variable χ^2 is called Chi-squared statistic, it is similar to mean and variance statistic. It has a sample distribution called Chi-squared distribution. The curve skewing is different according to the size of the sample n or degrees of freedom. Small n makes the curve skewed left, while large n makes the curve similar to normal distribution. The Chi-squared distribution is depicted in the figure

```
n=4; %sample size// n-1 degrees of freedom
N=10000; %Popoulation size
n_samples=10000;
Data=fix(rand(1,N)*100);
sigma2=var(Data);
X_=[     ]  ;
for i=1:n_samples
 Index=fix(rand(1,n)*N);  %get random index
 Index(Index==0)=1; % remove zero indies
 Sample=Data(Index); %get random Samples
 S2=var(Sample);
 chi_sqred=((n-1)*S2)/sigma2;
 X_ = [ X_  chi_sqred]; % collect chi-
square statistic
end
hist( X_,100)
```



6 .2 Types of Estimation:

There are two types of inference, Estimation and Hypothesis testing, hypothesis testing is studied in the upcoming chapter.

Estimation: The objective of estimation is to determine the approximate value of the population parameter based on sample statistic. for example, sample mean $\overline{X}$ can be employed to determine the population mean μ . There are two types of estimators.

- Point Estimator: use only single point to estimate the population parameter (take only single sample). the estimation is more accurate with greater sample size.
- Interval Estimator: We estimate that the population parameter is inside an interval with lower and upper limits. We also say we are confident % that the population parameter is a value in the estimated interval.

Example 1: We can estimate mean number of voters that said " YES" in each meeting with;

- *Point estimate*: calculate mean number of voters in specific meeting. the mean $\overline{X}$ for the sample is 200 persons (as an example).
- *Interval estimate*: take random meeting with size n and calculate $\overline{X}$ more than one time. We find that $\overline{X}$ always between 180, 210.

6.2.1 Confidence Interval Estimator for μ

If we want to estimate the parameter interval with 95% confidence, we say that $(1-\propto) = 0.95$. and the confidence level $\propto = 0.05$

Then the limits are.

$$UCL = \bar{x} + Z_{\propto/2}\frac{\partial}{\sqrt{n}} \rightarrow upper\ limit$$

$$LCL = \bar{x} - Z_{\propto/2}\frac{\partial}{\sqrt{n}} \rightarrow lower\ limit$$

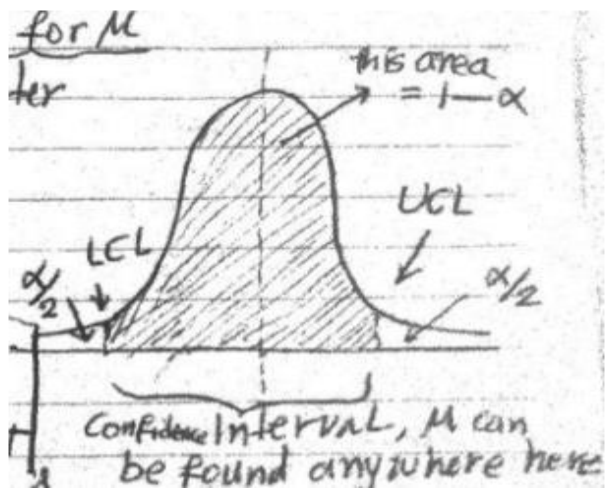


Table A.4: confidence level table (keep it for use) , this table is part from Table A.3

$1-\propto$	$\propto$	$\propto / 2$	$Z_{\propto/2}$
.90	.10	0.05	$Z_{.05} = 1.645$
.95	.05	0.025	$Z_{.25} = 1.96$
.98	.02	0.01	$Z_{.01} = 2.33$
.99	.01	0.005	$Z_{.005} = 2.575$

In MATLAB

```
>> norminv(0.05)
-1.645
```

Example1: A computer company samples demand over 25 times periods is found to be;

235	374	309	499	253
421	361	514	462	369
394	439	348	344	330
261	374	302	466	535
386	316	296	332	334

It is known that standard deviation for the entire population is 75 computers. We want to estimate the mean demand with 95% confidence in order to set inventory levels.

Solution: Sample mean: $\bar{x} = \frac{\sum xi}{n} = \frac{(235+374+309+499+253+421+\cdots+334)}{25} = 375.16$

$$1-\propto = 0.95\,, \propto/2 = 0.25\ from\ table\ A.4\ Z_{0.25} = 1.96$$

The population standard deviation = 75, the sample size = 25.

$$UCL = \bar{x} + Z_{\propto/2}\frac{\partial}{\sqrt{n}} = 370.16 + 1.96 * \frac{75}{\sqrt{25}} = 399.56$$

$$LCL = \bar{x} - Z_{\propto/2}\frac{\partial}{\sqrt{n}} = 370.16 - 1.96 * \frac{75}{\sqrt{25}} = 340.76$$

From example 1 we notice that wide interval provides little information. for example, if we are confident 95% that μ is between 399 and 390 that gives more accurate estimate.

Now the question is how to get narrow intervals? Look at the equation for UCL, LCL

$$\bar{x} \pm Z_{\propto/2}\frac{\partial}{\sqrt{n}}$$

- Solution 1: decrease the confidence level $Z_{\propto/2}$. i.e. calculate for 90% is narrower than calculate for 99 %.
- Solution 2: increase the sample size n so $\pm Z_{\propto/2}\frac{\partial}{\sqrt{n}}$ will be smaller.

From Solution 2 We can get the best sample size n that produce narrow intervals.

Example 2: Suppose that we want the interval of confidence to be within ±5 in example1, notice that confidence interval $\pm Z_{\propto/2}\frac{\partial}{\sqrt{n}} = \pm29.40$ in Example1; in Example 2 we want ±29 to be ±5.

Solution:

We want $Z_{\propto/2}\frac{\partial}{\sqrt{n}} = 5$

$$\Rightarrow \sqrt{n} = \frac{Z_{\propto/2}\partial}{5} \Rightarrow n = (\frac{Z_{\propto/2}\partial}{5})^2$$

$$\Rightarrow n = (\frac{(1.96)(75)}{5})^2 = 865$$

So we have to get sample size of 865 instead of 25 that was done in Example 1.

From Example 2 We conclude that the sample size of the required interval would be .

$$n = (\frac{Z_{\propto/2}\partial}{W})^2$$

Where W is the required interval width.

Example 3: A lumber company must estimate the mean diameter of trees to determine whether there is sufficient lumber to harvest an area of forest. they need to estimate this to within 1 inch at confidence level of 99%. the tree diameters are normally distributed with a standard deviation of 6 inches. How many trees need to be sampled?

Solution: Confidence level = 99%, therefore $\propto = 0.01, {}^{\propto}/_{2} = 0.005$ from table A.4, $Z_{\propto/2} = 2.575$

We want $\bar{x} \pm 1$ then W = 1, we are also given the population standard deviation ∂ = 6

We Compute.

$$n = (\frac{Z_{\propto/2}\partial}{W})^2 = (\frac{(2.575)(6)}{1})^2 = 239$$

That's, we need to sample at least 239 trees to have 99% confidence interval of $\bar{x} \pm 1$

Summery:

Sampling distribution and confidence intervals for the mean have been discussed

Exercises 7

1. A computer company samples demand over 12 times periods is found to be;

235	374	309	253
421	361	514	369
344	439	348	330

It is known that standard deviation for the entire population is 75 computers. We want to estimate the mean demand with 95% confidence in order to set inventory levels.

2. A lumber company must estimate the mean diameter of trees to determine whether or not there is sufficient lumber to harvest an area of forest. they need to estimate this to within **2** inches at confidence level of 99%. the tree diameters are normally distributed with a standard deviation of 6 inches. How many trees need to be sampled?

Chapter 8
Hypothesis Testing

8.1 Steps of Statistical Hypothesis

There are six steps to test any statistical hypothesis ([6] Douglas)

1. Define the population distribution, the population distribution must be known
2. Stating the hypothesis – H_0 : is null hypothesis (no difference), H_1 is the alternative hypothesis (there is a difference)
3. Define the level of significance
4. Calculate the testing statistic.
5. Define the critical region.
6. Take the final decision, accept or reject the hypothesis.

Step 1: There are two types of hypothesis tests

- Parametric tests. In parametric tests, knowing the population distribution is important.
- Non-parametric tests, In non-parametric tests, knowing the population distribution is not needed.

Step 2: There are two types of hypothesis

- H0: Null hypothesis, (No difference hypothesis)
- H1: Alternative hypothesis, it has three cases
    - Two tailed test: the population parameter is different from specific value, so we don't know which direction to calculate. For example, H1: $\mu \neq 10000$
    - Right test: H1: $\mu > 10000$
    - Left test: H1: $\mu < 10000$

Step 3: Define the level of significance (α), the level of significance is the probability of rejecting null hypothesis when it is correct. It is called type I error and shown in the following table.

Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

In one-side test the level of significance is α, In two-tailed test the level of significance is $\frac{\alpha}{2}$

Step 4: Calculate the testing statistic:

d

References:

[1] Ronald E. Walpole, Raymond H. Myers, Sharon.L. Myers and Keying Ye; Probability and Statistics for Engineers and Scientists, 8thed. 2007, Prentice Hall.
[2] Dr D JWilkinson: Introduction to Probability and Statistics
[3] W. W. Hines, D.C. Montgomery, D.M. Golsman, C.M. Borror; ; Probability and Statistics in Engineering, John Wiley & Sons, Inc., 2003.
[4] Windy L. Martinez, Angel R. Martinez- Computational Statistics Handbook with MATLAB
[5] Eric W . Hanson, Using MATLAB for Stochastic Simulation, Engineering Sciences 27 — Discrete and Probabilistic Systems, 2010
[6] Douglas A. Lind, William G. Marchal and Samuel A. Wathen “Statistical Techniques in business &Economics” , Fifteenth Edition
[7] Bluman, Allan G. Allan G. *Elementary statistics: A step by step approach*. No. QA 276.12. B58 2004. 2004.

Websites for similar courses

[1] 2DI90 - Probability and Statistics,
http://www.win.tue.nl/~rmcastro/2DI90/index.php?page=lectures
[2] King Abdulaziz University, Alghamdy, website contains questions, exams and solutions for statistics course.
http://saalghamdy.kau.edu.sa
http://www.kau.edu.sa/content.aspx?Site_ID=0003610&lng=AR&cid=26299