Course content

Pre-requisites: BBM 112 and BBM 212

Purpose: To introduce students to business statistics necessary for data summarization and presentation

Expected Learning Outcomes of the Course:

By the end of the course unit the learners should be able to:-

i) Organize and present data using various methods
ii) Interpret summarized data

Course Content:

Data collection; Organization and Presentation of Data; Random variables: Discrete and continuous random variables, Their distribution such as binomial, Poisson, normal, and their business applications

Teaching / Learning Methodologies: Lectures and tutorials; group discussion; demonstration; Individual assignment; Case studies

Instructional Materials and Equipment: Projector; test books, design catalogues; computer laboratory; design software; simulators

Course Assessment

Examination - 70%; Continuous Assessment Test (CATS) - 20%; Assignments - 10%; Total - 100%

Recommended Text Books

i) Azel (2006); *Complete Business Statistics;* Tata Mcgraw Hill
ii) Beri (2008); *Business Statistics;* PVT Publishers New Delhi
iii) Chandra J. S. (2003); *Statistical for Business and Economics;* Tata McGraw –Hall, New Delhi

Text Books for further Reading

i) Mansfield Edwin; *Statistical for Business Economics Methods and Applications;* New York W.W Norton and Company
ii) Enns Phillip G, (1985); *Business Statistical Methods ad Application;* Homewood Richard D Irwin Inc

Module compiled by Charles Karuga

LECTURE 1

CHAPTER 1: INTRODUCTION

Purpose

To introduce the student to the world of statistics and to acquaint them with the role of statistics in Business.

Objectives

a) ***Define statistics and explain its uses.***
b) ***Define Business Statistics***
c) ***State limitations of statistics.***
d) ***Explain why statistics is distrusted.***
e) ***Distinguish between descriptive and inferential statistics.***
f) ***Explain the types of variables.***
g) ***State the levels and scales of measurement.***

1.1 What is Statistics?

The word 'statistics' is defined by Croxton and Cowden as follows:-

"The collection, presentation, analysis and interpretation of the numerical data."

This definition clearly points out four stages in a statistical investigation, namely:

1) Collection of data 2) Presentation of data
3) Analysis of data 4) Interpretation of data

In addition to this, one more stage i.e. organization of data is suggested

Definition:

Business statistics is the science of good decision making in the face of uncertainty and is used in many disciplines such as financial analysis, econometrics, auditing, production and operations including services improvement, and marketing research..

1.2 Uses of Statistics

a) To present the data in a concise and definite form : Statistics helps in classifying and tabulating raw data for processing and further tabulation for end users.
b) To make it easy to understand complex and large data : This is done by presenting the data in the form of tables, graphs, diagrams etc., or by condensing the data with the help of means, dispersion etc.
c) For comparison: Tables, measures of means and dispersion can help in comparing different sets of data..
d) In forming policies: It helps in forming policies like a production schedule, based on the relevant sales figures. It is used in forecasting future demands.
e) Enlarging individual experiences: Complex problems can be well understood by statistics, as the conclusions drawn by an individual are more definite and precise than mere statements on facts.

f) In measuring the magnitude of a phenomenon:- Statistics has made it possible to count the population of a country, the industrial growth, the agricultural growth, the educational level (of course in numbers)

1.3 Limitations of Statistics

1. Statistics does not deal with individual measurements. Since statistics deals with aggregates of facts, it cannot be used to study the changes that have taken place in individual cases. For example, the wages earned by a single industry worker at any time, taken by itself is not a statistical datum. But the wages of workers of that industry can be used statistically. Similarly the marks obtained by Kamau of your class or the height of Atieno (also of your class) are not the subject matter of statistical study. But the average marks or the average height of your class has statistical relevance.

- If one interviewer used, uniformity of approach.
- Used to pilot other methods.

Disadvantages:

- Need to set up interviews.
- Time consuming.
- Geographic limitations.
- Can be expensive.
- Normally need a set of questions.
- Respondent bias – tendency to please or impress, create false personal image, or end interview quickly.
- Embarrassment possible if personal questions.
- Transcription and analysis can present problems – subjectivity.
- If many interviewers, training required.

2.2.3 Case-studies

The term case-study usually refers to a fairly intensive examination of a single unit such as a person, a small group of people, or a single company. Case-studies involve measuring what is there and how it got there. In this sense, it is historical. It can enable the researcher to explore, unravel and understand problems, issues and relationships. It cannot, however, allow the researcher to generalize, that is, to argue that from one case-study the results, findings or theory developed apply to other similar case-studies. The case looked at may be unique and, therefore not representative of other instances. It is, of course, possible to look at several case-studies to represent certain features of management that we are interested in studying. The case-study approach is often done to make practical improvements. Contributions to general knowledge are incidental.

The case-study method has four steps:

1. Determine the present situation.
2. Gather background information about the past and key variables.
3. Test hypotheses. The background information collected will have been analysed for possible hypotheses. In this step, specific evidence about each hypothesis can be gathered. This step aims to eliminate possibilities which conflict with the evidence collected and to gain confidence for the important hypotheses. The culmination of this step might be the development of an experimental design to test out more rigorously the hypotheses developed, or it might be to take action to remedy the problem.
4. Take remedial action. The aim is to check that the hypotheses tested actually work out in practice. Some action, correction or improvement is made and a re-check carried out on the situation to see what effect the change has brought about.

The case-study enables rich information to be gathered from which potentially useful hypotheses can be generated. It can be a time-consuming process. It is also inefficient in researching situations which are already well structured and where the important variables have been identified. They lack utility when attempting to reach rigorous conclusions or determining precise relationships between variables.

2.2.4 Diaries

A diary is a way of gathering information about the way individuals spend their time on professional activities. They are not about records of engagements or personal journals of thought! Diaries can record either quantitative or qualitative data, and in management research can provide information about work patterns and activities.

Advantages:

ways but are usually based on having some information about population members. This information is usually in the form of an alphabetical list – called the sampling frame.

Three types of random sample can be drawn – a simple random sample (SRS), a stratified sample and a systematic sample.

2.3.1 Simple random sampling

Simple random sampling can be carried out in two ways – the lottery method and using random numbers.

The **lottery method** involves:

- transferring each person's name from the list and putting it on a piece of paper
- the pieces of paper are placed in a container and thoroughly mixed
- the required number are selected by someone without looking
- the names selected are the simple random sample.

This is basically similar to a game of bingo or the national lottery. This procedure is easy to carry out especially if both population and sample are small, but can be tedious and time consuming for large populations or large samples.

Alternatively **random numbers** can be used. Random numbers are strings of digits that have been generated by the lottery method and can be found in books of statistical tables. An example of these is:

03	47	43	73	86	36	96	47	36	61
97	74	24	67	62	42	81	14	57	20
16	76	62	27	66	56	50	26	71	07
12	56	85	99	26	96	96	68	27	31

This procedure is repeated until the nine people have been identified.

d) Any number occurring for second time is ignored as is any two-digit number over 89.

Simple random number sampling is used as the basis for many other sampling methods, but has two disadvantages:

- A sampling frame is required. This may not be available, exist or be incomplete.
- The procedure is unbiased but the sample may be biased. For instance, if the 90 people are a mixture of men and women and all men were selected this would be a biased sample.

2.3.2 Stratified Sampling

To overcome the second problem above, a stratified sample can be taken. In this the population structure is reflected in the sample structure, with respect to some criterion.

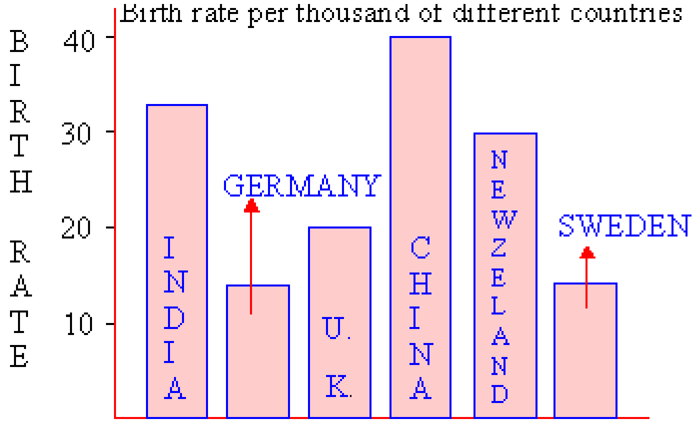
For example, suppose the 90 people consist of 30 men and 60 women. If gender is the criterion for stratification then:

$\frac{30}{90}$ of the sample should be men

i.e. $\frac{30}{90} \times 9 = 3men$

$\frac{60}{90}$ of the sample should be women

Represent the above data by a suitable diagram.



Comparing the size of bars, you can easily see that China's birth rate is the highest while Germany and Sweden equal in the lowest positions. Such diagrams are also known as component bar diagrams.

3.3.2 Sub - divided Bar Diagram

While constructing such a diagram, the various components in each bar should be kept in the same order. A common and helpful arrangement is that of presenting each bar in the order of magnitude with the largest component at the bottom and the smallest at the top. The components are shown with different shades or colors with a proper index.

Illustration:- During 1968 - 71, the number of students in University ' X ' are as follows. Represent the data by a similar diagram.

Year	Arts	Science	Law	Total
1968-69	20,000	10,000	5,000	35,000
1969-70	26,000	9,000	7,000	42,000
1970-71	31,000	9,500	7,500	48,000

3) Join the points by a smooth curve. Note that Ogives start at (i) zero on the vertical axis, and (ii) outside class limit of the last class. In most of the cases it looks like 'S'.

Note that cumulative frequencies are plotted against the 'limits' of the classes to which they refer.

(A) Less than Ogive:- To plot a less than ogive, the data is arranged in ascending order of magnitude and the frequencies are cumulated starting from the top. It starts from zero on the y-axis and the lower limit of the lowest class interval on the x-axis.

(B) Greater than Ogive:- To plot this ogive, the data are arranged in the ascending order of magnitude and frequencies are cumulated from the bottom. This curve ends at zero on the the y-axis and the upper limit of the highest class interval on the x-axis.

Illustrations:- On a graph paper, draw the two ogives for the data given below of the I.Q. of 160 students.

Class -intervals:	60 - 70	70 - 80	80 – 90	90 - 100	100 - 110
No. of students:	2	7	12	28	42

110 - 120	120 - 130	130 - 140	140 - 150	150 - 160
36	18	10	4	1

So we finish up with:

Stem	Leaf
13	6, 9, 9
14	2, 3, 3, 3, 3, 4
14	6, 7, 7, 8, 9
15	1, 3, 4
15	6, 7
16	2, 4

Key:
13 | 6 means 136

3.6.1 Back-to-back stem and leaf diagram

Back-to-back stem plots are used to compare two distributions side-by-side. This type of double stem plot contains three columns, each separated by a vertical line. The center column contains the stems. The first and third columns each contain the leaves of a different distribution. The numbers for the leaves of the distribution in the leftmost column are aligned to the right and are listed in increasing order from right to left. Here is an example of a back-to-back stem plot comparing the distribution o marks obtained in an exam by a sample of 25 boys and 25 girls.

					BOYS		GIRLS					
				3	4	40	5	4	1	2	8	5
		3	5	5	0	50	2	3	5	8	9	4
2	2	3	3	4	5	60	3	5	6	4	5	
5	5	2	8	0	2	70	0	3	3			
		3	1	3	4	80	3	6	4			
			4	4	9	90	3	4				

LECTURE4

CHAPTER 4: ANALYSIS AND INTERPRETATION OF DATA

Purpose

To examine various measures of central tendency.

Objectives

Compute numerical quantities that measure centrality in a set of data such as:

i. ***Arithmetic mean***
ii. ***Median***
iii. ***Mode***
iv. ***Geometric mean***
v. ***Weighted mean***
vi. ***Harmonic mean***

Measures of Central Tendency

4.1 Introduction

In the previous chapter, we have studied how to collect raw data, its classification and tabulation in a useful form, which contributes in solving many problems of statistical concern. Yet, this is not sufficient, for in practical purposes, there is need for further condensation, particularly when we want to compare two or more different distributions. We may reduce the entire distribution to one number which represents the distribution.

A single value which can be considered as typical or representative of a set of observations and around which the observations can be considered as Centered is called an 'Average' (or average value) or a Centre of location. Since such typical values tend to lie centrally within a set of observations when arranged according to magnitudes, averages are called measures of central tendency.

Families	Expenditure ($)	Deviation from as. mean
(x_i)		$(u_i = x_i - A)$
A	300	-200
B	700	200
C	100	- 400
D	750	250
E	500	0
F	80	-420
G	120	-380
H	250	-250
I	100	-400
J	370	-130
$n = 10$		$\Sigma u_i = -2180 + 450$ $= -1730$

Calculations :

$$\bar{u} = \frac{\Sigma u_i}{n} = \frac{-1730}{10} = -173$$

$$\bar{x} = A + \bar{u}$$

$$\therefore \bar{x} = 500 + (-173) = 327$$

4.2.2 Grouped data

There is a difference in the methods for finding the arithmetic means of the individual series and a discrete series. In the discrete series, every term (i.e. value of x) is multiplied by its corresponding frequency $f_i x_i$ and then their total (sum) is found $\sum f_i x_i$. The arithmetic mean is then obtained by dividing the total frequency $\sum f$ by the above sum so obtained $\sum f_i x_i$.

x	14	14x
125	8	1000
128	6	768
130	2	260
Total	$\sum f_i = 100$	$\sum f_i x_i = 9906 + 14x$

Now the arithmetic mean $\left(\bar{x}\right) = \dfrac{\sum f_i x_i}{\sum f_i}$

Therefore, $115.86 = \dfrac{9906 + 14x}{100}$

or $115.86 \times 100 = 9906 + 14x$

$11586 = 9906 + 14x$

$11586 - 9906 = 14x$

$1680 = 14x$

or $x = \dfrac{1680}{14}$

$x = 120$

Therefore the missing item is 120.

4.2.3 Properties of Arithmetic Mean

1. The sum of the deviations, of all the values of x, from their arithmetic mean, is zero.

   Justification

   $: \sum f_i \left(x_i - \bar{x}\right) = \sum f_i x_i - \bar{x} \sum f_i = 0$

therefore,

$$\sum x_i - 7 = -3$$

Subtracting the two equations we get,

$$\sum x_i - 4n = 72$$

$$\sum x_i - 7n = -3$$

(-) (+) (+)

3n=75

n = 25

Putting n = 25 in $\sum x_i - 4n = 72$, we get

$$\sum x_i - 4(25) = 72$$

$$\sum x_i = 72 + 100$$

$$\sum x_i = 172$$

Now Mean is given by $\bar{x} = \frac{\sum x_i}{n} - \frac{172}{25} = 688$

Example The mean weight of 98 students is found to be 50 kg. It is later discovered that the frequency of the class interval (30- 40) was wrongly taken as 8 instead of 10. Calculate the correct mean.

Solution:

Incorrect mean $(\bar{x}) = 50\,\text{Kg}$ and $\sum f_i = 98$

Incorrect $\bar{x} = \dfrac{incorrect \sum f_i x_i}{\sum f_i}$

$$50 = \frac{incorrect \sum f_i x_i}{98}$$

Therefore, Incorrect $\sum f_i x_i = 98 \times 50 = 4900$

Now correct

$\sum f_i x_i$ = Incorrect $\sum f_i x_i - (8 \times 35) + (10 \times 35)$

Note that the class-mark of class interval (30 - 40) is 35 and for the calculation of the mean we consider class marks.

\The correct $\sum f_i x_i = 4900 - 280 + 350 = 4970$

Also the correct $\sum f_i = 98 + 2 = 100$

Therefore, the correct mean

$$\bar{x} = \frac{correct \sum f_i x_i}{correct \sum f_i} = \frac{4970}{100}$$
$$= 49.7kg$$

Example The sum of the deviations of 'n' observation values of a variate from a

constant 'a', is S. Show that the arithmetic mean is $a + \dfrac{S}{n}$.

4.3 Median

It is the value of the size of the central item of the arranged data (data arranged in the ascending or the descending order). Thus, it is the value of the middle item and divides the series in to equal parts.

In Connor's words - "The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other all values lesser than the median." For example, the daily wages of 7 workers are 5, 7, 9, 11, 12, 14 and 15 dollars. This series contains 7 terms. The fourth term i.e. $11 is the median.

4.3.1 Median In Individual Series (ungrouped Data)

1. Set the individual series either in the ascending (increasing) or in the descending (decreasing) order, of the size of its items or observations.
2. If the total number of observations be 'n' then
   A. If 'n' is odd,

   $$\text{The median} = \text{size of} \left(\frac{n+1}{2}\right)^{th} \text{observation}$$

   B. If 'n' is even, the median

   $$= \frac{1}{2}\left[\begin{array}{l} \text{size of} \left(\frac{n}{2}\right)^{\text{th}} \text{observations} \\ + \text{size of} \left(\frac{n+2}{2}\right)^{\text{th}} \text{observations} \end{array}\right]$$

Example The following figures represent the number of books issued at the counter of a Statistics library on 11 different days. 96, 180, 98, 75, 270, 80, 102, 100, 94, 75 and 200. Calculate the median.

LECTURE 6

4.3.4 Merits of Median

1. It is rigidly defined.
2. It is easy to calculate and understand.
3. It is not affected by extreme values like the arithmetic mean. For example, 5 persons have their incomes \$2000, \$2500, \$2600, \$3000, \$5000. The median would be \$2600 while the arithmetic mean would be \$3020.
4. It can be found by mere inspection.
5. It is fully representative and can be computed easily.
6. It can be used for qualitative studies.
7. Even if the extreme values are unknown, median can be calculated if one knows the number of items.
8. It can be obtained graphically.

4.3.5 Demerits of Median

1. It may not be representative if the distribution is irregular and abnormal.
2. It is not capable of further algebraic treatment.
3. It is not based on all observations.
4. It is affected by sample fluctuations.
5. The arrangement of the data in the order of magnitude is absolutely necessary.

4.4 Mode

It is the size of that item which possesses the maximum frequency. According to Professor Kenney and Keeping, the value of the variable which occurs most frequently in a distribution is called the mode.

It is the most common value. It is the point of maximum density.

$$H.M = \overline{X} = \frac{n}{\sum\left(\frac{1}{x}\right)}$$

$$H.M = \overline{X} = \frac{5}{0.3417} = 14.63$$

Example:

Given the following frequency distribution of first year students of a particular college. Calculate the Harmonic Mean.

Age (Years)	13	14	15	16	17
Number of Students	2	5	13	7	3

Solution:

The given distribution belongs to a grouped data and the variable involved is ages of first year students. While the number of students Represent frequencies.

Ages (Years) x	**Number of Students** f	$\frac{1}{x}$
13	2	0.1538
14	5	0.3571
15	13	0.8667
16	7	0.4375
17	3	0.1765
Total	$\sum f = 30$	$\sum \frac{1}{x} = 1.9916$

Now we will find the Harmonic Mean as

4.9 Characteristics of a Good Measure of Central Tendency

1. It should be rigidly defined
2. It should be easy to understand and calculate
3. It should be based on all observations
4. It should be amenable to further algebraic manipulation.
5. It should not be affected much by extreme values
6. It should be least affected by fluctuations in sampling.

Chapter Review Questions

1. ***The mean of the ten numbers listed below is 5.5.***

    4, 3, a, 8, 7, 3, 9, 5, 8, 3

    (a) ***Find the value of a.***

    (b) ***Find the median of these numbers.***

2. ***In the following ordered data, the mean is 6 and the median is 5.***

    2, b, 3, a, 6, 9, 10, 12

    Find each of the following

    (a) ***the value of a;***

    (b) ***the value of b.***

3. ***For the set of {8, 4, 2, 10, 2, 5, 9, 12, 2, 6}***

    (a) ***calculate the mean;***

    (b) ***find the mode;***

    (c) ***find the median.***

Solution

$$\text{Arithmetic mean} = \bar{x} = \frac{12+6+7+3+15+10+18+5}{8}$$

$$= \frac{76}{8} = 9.5$$

x_i	$\lvert x_i - \bar{x} \rvert$
12	2.5
6	3.5
7	2.5
3	6.5
15	5.5
10	0.5
18	8.5
5	4.5
	34

$$\text{Mean deviation from the mean} = \frac{34}{8} = 4.25$$

Example (Continuous series) calculate the mean deviation and the coefficient of mean deviation from the following data using the mean.

Difference in ages between boys and girls of a class.

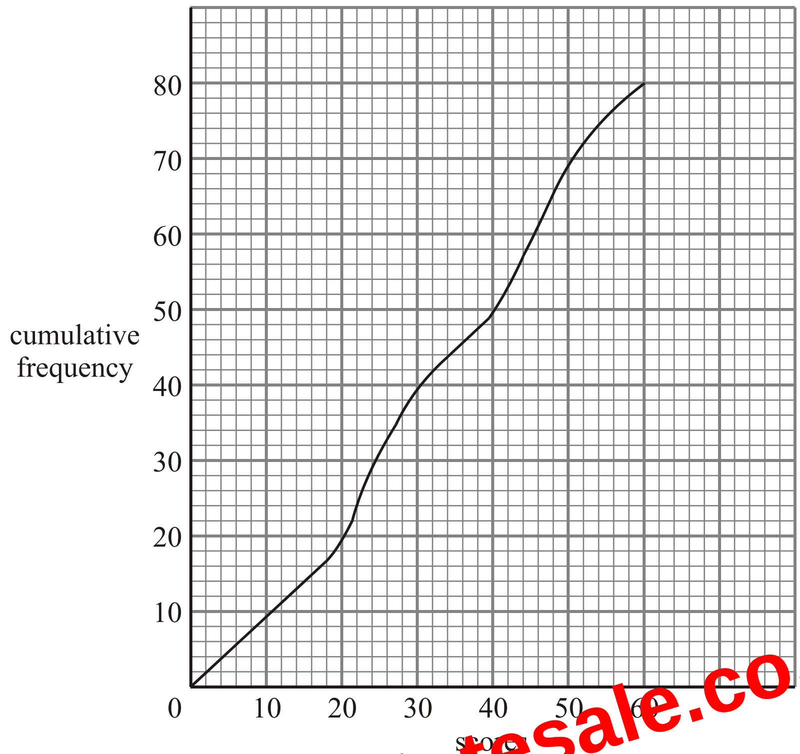
Diff. in years:	No. of students:
0 – 5	449

Solution:

Income group	f	c.f.
Below 50	1	1
50 – 70	16	17
70 – 90	39	56
90 – 110	58	114
110 – 130	60	174
130 – 150	46	220
150 – 170	22	242
170 – 190	15	257
190 – 210	15	252
210 – 230	9	281
230 & above	10	291
$n = \sum f = 291$		

Calculations :

1) Median = Size of $n+1/2^{th}$ item



From the graph find

(a) ***the median value;***

(b) ***the interquartile range;***

(c) ***the 35th percentile.***

6. The table below shows the number and weight (w) of fish delivered to a local fish market one morning.

weight (kg)	***frequency***	***cumulative frequency***
$0.50 \leq w < 0.70$	***16***	***16***
$0.70 \leq w < 0.90$	***37***	***53***
$0.90 \leq w < 1.10$	***44***	***c***

time it takes a number of students to complete a computer game.

(a) *From the graph find*

(i) *the median time*

(ii) *the interquartile range.*

The graph has been drawn from the data given in the table below.

Time in minutes	*Number of students*
$0 < x \leq 5$	*20*
$5 < x \leq 15$	*20*
$15 < x \leq 20$	p
$20 < x \leq 25$	*40*
$25 < x \leq 35$	*60*
$35 < x \leq 50$	q

35	36	1260	1225	1296
232	**231**	**7672**	**7710**	**7651**

$$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{232}{7} = 33 \text{ and } \bar{y} = \frac{\Sigma y}{n} = \frac{231}{7} = 33$$

Now $$r = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)}\sqrt{\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}}$$

$$r = \frac{7672 - \frac{232 \times 231}{7}}{\sqrt{\left(7710 - \frac{(232)^2}{7}\right)}\sqrt{\left(7651 - \frac{231^2}{7}\right)}}$$

$$\frac{7612 - 7656}{\sqrt{(7710 - 7689)}\sqrt{(7651 - 7623)}} = \frac{16}{\sqrt{21} \times \sqrt{28}}$$

Therefore r=0.65

b) The equation of the line of regression of y on x

$$y - \bar{y} = \frac{s_{xy}}{s_x^{\ 2}}\left(x - \bar{x}\right)$$

$$s_{xy} = 16 \qquad s_x^{\ 2} = 21$$

$$y - 33 = \frac{16}{21}\left(x - 33\right)$$

$$\Rightarrow \text{y} = 0.76\text{x}+7.92$$

c) Inserting x = 38, we get

y = 0.76 (38) +7.92

y = 36.8 = 37 (approximately)

Therefore, the Judge B would have given 37 marks to 8th performance

Alternative Formula for Calculating Regression

It is expressed as y = a + bx where a and b are two unknown constants which determine the position of the line completely.

If the values of a and b are completely determined the equation of the regression line of y on x is obtained.

The two basic equations which can be solved simultaneously to find a and b are:-

$\sum y = na + b\sum x$ …………(i)

$\sum xy = a\sum x + b\sum x^2$…..(ii)

Example 4

From 10 observations of price x and supply y of a commodity the results obtained

$\sum x = 130, \ \sum y = 220, \ \sum x^2 = 2288, \ \sum xy = 3467$

Compute the regression of y on x and interpret the result. Estimate the supply when the price of 16 units.

7.2.1 Features of a Binomial Distribution

The **Binomial** experiment consists of:

- A fixed number of trials, n
- Two possible outcomes, p and q (or 1 – p). p is called 'success' and 1 – p (the complement) is called 'failure'. 'Success' is what we are interested in. For example, the proportion of defective items produced by a factory in a day may be 1%. p = 0.01.
- Independent trials (the outcome of one trial does not affect the outcomes of any other trials).

7.2.2 Using the Binomial Tables:

Different probabilities can be read from the binomial tables.

Each page focuses on a given value of 'n'. Across the top are various values of 'p'. In the left margin are the possible values of 'k'. The body of the table gives **$P(X \leq k)$.**

Example 1

Find the probability of at least two defective items in a batch of 10 items with a defective rate of 10%.

Here, we would need to find the following:-

$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) +$
$P(X = 8) + P(X = 9) + P(X = 10)$

If we used $P(X \leq k)$ we would only need to find:

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

Your binomial tables are at the back of this Module.

where

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

X represents the name of the random variable. x represents the value of the random variable**.** The factorial sign (!) is best explained by example. 6! means 6 x 5 x 4 x 3 x 2 x 1 = 720. You will find the ! button on your calculator.

7.3 The Poisson Distribution

A particular event occurs 'on average' μ times within a certain period or situation. What is the probability that it will occur k times in that period?
A Poisson probability distribution is the number of occurrences per interval of **time** or **space**.
Time? Example: A programmer makes, on average, 2 mistakes per hour. Space?
Example: A programmer makes, on average, 2 mistakes per program.
Example : The average number of arrivals at a doctor's clinic is 3 per hour. What is the probability that in a given hour there will be 5 arrivals? Here we have μ=3 and k=5.

Note:
1) There is no maximum possible value for 'k'.

7.3.1 Features of the Poisson Experiment

- The number of successes that occur in a period of time or an interval of space is independent of the number of successes that occur in any other interval

<u>QUESTION TWO</u> (20 marks)

The manager of a fast food restaurant is concerned that the customers are waiting for too long for their food. She decides to gather some statistics on customer waiting times and the following times (in minutes are recorded).

1.25	2.5	8.5	4.6	10.5	3.4	3.7	6.25	7.7	4.1	5.15	5.95	7.35	5.8
2.9	3.4	6.6	8.8	2.7	10.2	4.5	5.2	4.1	2.5	3.8	2.1	5.5	6.25
4.3	1.8	3.7	4.4	6.2	3.3	7.2	8.6	3.45	6.55	2.85	9.4	4.25	5.6
11.9	6.4	4.8	5.8	2.5	4.1	8.1	6.1						

(a) Group this data into classes of 0 - 1.9, 2 - 3.9 etc and construct a frequency table which also shows cumulative frequencies. (4marks)

(b) Construct a histogram for the data. (4marks)

(c) Construct a less than ogive for the data and answer the following questions (4marks)

i) How many customers have to wait for less than 4 minutes to be served? (2marks)

ii) What percentage of customers has to wait for less than 5 minutes for their food? (2marks)

iii) If the restaurant's goal is for 90% of the customers to be given their food within 8 minutes, are they achieving this goal? (2marks)

iv) What is the mean waiting time? 2marks)

<u>QUESTION FOUR</u> **(20 marks)**

a) A reaction time experiment was performed first with 21 girls, and then with 24 boys. The results are shown on the stem and leaf diagram below

Key(Girls) 6|1 means0.16 sec

Key(Boys) 1|8 means 0.18 sec

Girls		Boys
4	2	4
3 3 2 2	2	2 2
1 0 0	2	0 0 0 1 1
9 9 9 8 8	1	8 8 8
7 7 6	1	6 6 7 7
5 5 4 4 4	1	4 5 5
	1	2 3
	1	0 1 1
	0	9

i) Find the median and inter quartile range for both sets of reaction times 6marks)

ii) Comment on your answers. (2marks)

42	93	46	52	72	77	53	41	48	86
62	54	85	60	58	43	58	43	58	74
52	82	78	86	94	63	72	63	72	44
78	56	80	44	52	74	68	82	57	47

e) Construct a stem and leaf diagram to represent these data. (4marks)

f) Find the median and the quartiles of this distribution. (4marks)

g) Draw a box plot to represent these data. (4marks)

h) Give one advantage of using

iii) a stem and leaf diagram (2marks)

iv) a box plot (2marks)

to illustrate data such as the one given above.

d) At the end of a statistics course, Diana sits for two written papers, P1 and P2 and hands in a piece of course work. Her marks out 100 were 76 for P1 and 67 for P2 and she gained 81 marks for her course work. Her overall percentage is to be weighted so that the two written papers account for 40% while the course work accounts for 20%.Calculate Diana's overall percentage mark. (4 marks)

QUESTION FIVE (20 Marks)

a) A shopkeeper wanted to investigate whether or not there was a correlation between the prices of food 10 years ago in 1992, with their prices today. He chose 8 everyday items and the prices are given in the table below.

The incorrect measurements of 44.5 m and 43.2 m must be removed from the data.

(b) Calculate the new value of *x* after removing the two unwanted values.

(3 marks)

<u>QUESTION FIVE</u> **(20marks)**

The following data relates to daily bill on consumption of a certain commodity for 60 households

Daily bills(KSh)	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of households	6	7	11	10	6	5	9	3	3

i) Calculate the mean (4 marks

ii) Calculate the median (3marks)

iii) Calculate the standard deviation (3marks)

iv) Calculate the coefficient of skewness (4 marks)

v) Comment on the skewness of this distribution (2 marks)

vi) Calculate the coefficient of variation (4 marks)

1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

The values in the table are the areas between zero and the z-score. That is, P(0<Z<z-score)

K \ P	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	0.99873	0.98304	0.92953	0.82080	0.65625	0.45568	0.25569	0.09888	0.01585
4	0.99994	0.99840	0.98906	0.95904	0.89062	0.76672	0.57982	0.34464	0.11426
5	1.00000	0.99994	0.99927	0.99590	0.98438	0.95334	0.88235	0.73786	0.46856
6	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 7

K \ P=.1	.2	.3	.4	.5	.6	.7	.8	.9
0 | 0.47830	0.20972	0.08235	0.02799	0.00781	0.00164	0.00022	0.00001	0.00000
1 | 0.85031	0.57672	0.32942	0.15863	0.06250	0.01884	0.00379	0.00037	0.00001
2 | 0.97431	0.85197	0.64707	0.41990	0.22656	0.09626	0.02880	0.00467	0.00018
3 | 0.99727	0.96666	0.87396	0.71021	0.50000	0.28979	0.12604	0.03334	0.00273
4 | 0.99982	0.99533	0.97120	0.90374	0.77344	0.58010	0.35293	0.14803	0.02569
5 | 0.99999	0.99963	0.99621	0.98116	0.93750	0.84137	0.67058	0.42328	0.14969
6 | 1.00000	0.99999	0.99978	0.99836	0.99219	0.97201	0.91765	0.79028	0.52170
7 | 1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 8

K \ P=.1	.2	.3	.4	.5	.6	.7	.8	.9
0 | 0.43047	0.16777	0.05765	0.01680	0.00391	0.00066	0.00007	0.00000	0.00000
1 | 0.81310	0.50332	0.25530	0.10638	0.03516	0.00852	0.00129	0.00008	0.00000
2 | 0.96191	0.79692	0.55177	0.31539	0.14453	0.04981	0.01129	0.00123	0.00002
3 | 0.99498	0.94372	0.80590	0.59409	0.36328	0.17367	0.05797	0.01041	0.00043
4 | 0.99957	0.98959	0.94203	0.82633	0.63672	0.40591	0.19410	0.05628	0.00502
5 | 0.99998	0.99877	0.98871	0.95019	0.85547	0.68461	0.44823	0.20308	0.03809
6 | 1.00000	0.99992	0.99871	0.99148	0.96484	0.89362	0.74470	0.49668	0.18690
7 | 1.00000	1.00000	0.99993	0.99934	0.99609	0.98320	0.94235	0.83223	0.56953
8 | 1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

N = 9

K \ P=.1	.2	.3	.4	.5	.6	.7	.8	.9
0 | 0.38742	0.13422	0.04035	0.01008	0.00195	0.00026	0.00002	0.00000	0.00000
1 | 0.77484	0.43621	0.19600	0.07054	0.01953	0.00380	0.00043	0.00002	0.00000
2 | 0.94703	0.73820	0.46283	0.23179	0.08984	0.02503	0.00429	0.00031	0.00000
3 | 0.99167	0.91436	0.72966	0.48261	0.25391	0.09935	0.02529	0.00307	0.00006
4 | 0.99911	0.98042	0.90119	0.73343	0.50000	0.26657	0.09881	0.01958	0.00089
5 | 0.99994	0.99693	0.97471	0.90065	0.74609	0.51739	0.27034	0.08564	0.00833
6 | 1.00000	0.99969	0.99571	0.97497	0.91016	0.76821	0.53717	0.26180	0.05297
7 | 1.00000	0.99998	0.99957	0.99620	0.98047	0.92946	0.80400	0.56379	0.22516
8 | 1.00000	1.00000	0.99998	0.99974	0.99805	0.98992	0.95965	0.86578	0.61258
9 | 1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

2	0.0005	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0085	0.0058	0.0038	0.0022	0.0012	0.0005	0.0002	0.0000	0.0000
4	0.0773	0.0608	0.0459	0.0328	0.0216	0.0125	0.0057	0.0015	0.0000
5	0.3936	0.3530	0.3101	0.2649	0.2172	0.1670	0.1142	0.0585	0.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=7

x	**p = 0.01**	**p = 0.02**	**p = 0.03**	**p = 0.04**	**p = 0.05**	**p = 0.06**	**p = 0.07**	**p = 0.08**	**p = 0.09**
0	0.9321	0.8681	0.8080	0.7514	0.6983	0.6485	0.6017	0.5578	0.5168
1	0.9980	0.9921	0.9829	0.9706	0.9556	0.9382	0.9187	0.8974	0.8745
2	1.0000	0.9997	0.9991	0.9980	0.9962	0.9937	0.9903	0.9860	0.9807
3	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9988	0.9982
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	**p = 0.10**	**p = 0.15**	**p = 0.20**	**p = 0.25**	**p = 0.30**	**p = 0.35**	**p = 0.40**	**p = 0.45**	**p = 0.50**
0	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
1	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
2	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
3	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
4	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
5	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
6	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922

5	0.4956	0.3669	0.2485	0.1503	0.0781	0.0328	0.0099	0.0016	0.0010
6	0.7340	0.6177	0.4862	0.3504	0.2241	0.1209	0.0500	0.0128	0.0088
7	0.9004	0.8327	0.7384	0.6172	0.4744	0.3222	0.1798	0.0702	0.0540
8	0.9767	0.9536	0.9140	0.8507	0.7560	0.6242	0.4557	0.2639	0.2254
9	0.9975	0.9940	0.9865	0.9718	0.9437	0.8926	0.8031	0.6513	0.6106
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	**p = 0.92**	**p = 0.93**	**p = 0.94**	**p = 0.95**	**p = 0.96**	**p = 0.97**	**p = 0.98**	**p = 0.99**	**p = 1.00**
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0006	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0058	0.0036	0.0020	0.0010	0.0004	0.0001	0.0000	0.0000	0.0000
7	0.0401	0.0283	0.0188	0.0115	0.0062	0.0028	0.0009	0.0001	0.0000
8	0.1879	0.1517	0.1176	0.0861	0.0582	0.0345	0.0162	0.0043	0.0000
9	0.5656	0.5160	0.4614	0.4013	0.3352	0.2626	0.1829	0.0956	0.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n=11

x	**p = 0.01**	**p = 0.02**	**p = 0.03**	**p = 0.04**	**p = 0.05**	**p = 0.06**	**p = 0.07**	**p = 0.08**	**p = 0.09**
0	0.8953	0.8007	0.7153	0.6382	0.5688	0.5063	0.4501	0.3996	0.3544
1	0.9948	0.9805	0.9587	0.9308	0.8981	0.8618	0.8228	0.7819	0.7399
2	0.9998	0.9988	0.9963	0.9917	0.9848	0.9752	0.9630	0.9481	0.9305

n=14

x	p = 0.01	p = 0.02	p = 0.03	p = 0.04	p = 0.05	p = 0.06	p = 0.07	p = 0.08	p = 0.09
0	0.8687	0.7536	0.6528	0.5647	0.4877	0.4205	0.3620	0.3112	0.2670
1	0.9916	0.9690	0.9355	0.8941	0.8470	0.7963	0.7436	0.6900	0.6368
2	0.9997	0.9975	0.9923	0.9833	0.9699	0.9522	0.9302	0.9042	0.8745
3	1.0000	0.9999	0.9994	0.9981	0.9958	0.9920	0.9864	0.9786	0.9685
4	1.0000	1.0000	1.0000	0.9998	0.9996	0.9990	0.9980	0.9965	0.9941
5	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	p = 0.10	p = 0.15	p = 0.20	p = 0.25	p = 0.30	p = 0.35	p = 0.40	p = 0.45	p = 0.50
0	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
1	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
2	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
3	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
4	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
5	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120