A. Merlin Thanga Joy

NOORUL ISLAM COLLEGE OF ENGINEERING

Department of Management Studies

First Semester

BA 1601 – STATISTICS FOR MANAGEMENT

1. What do you meant by probability?

   The word "probability" that denotes that there is an uncertainty about the happening of event. It can also be defined as an expression of likelihood or chance of occurrence of an event. The probability is a number, which ranges from 0 to 1.

2. Define Event?

   An outcome or a combination of outcomes of a random experiment is called an event. Events are generally denoted by capital letters A, B, C, etc.

   For Example: If a coin is tossed, getting a head or tail is an event.

3. Define the term random experiment?

   The term experiment is used to describe an act, which can be repeated under some conditions. Random experiment is one which result depends on chance. That is the results can never be predicted.

   For example: Tossing a coin, throwing a die.

4. Define Mutually Exclusive events:

   Two events if the occurrence completely excludes the occurrence of the other. That is the two events cannot occur simultaneously.

   For Example: Tossing a coin, either head or tail will turn up.

5. Define Equally likely events:

   Events are said to be equally likely when one does not occur more often than others.

   For Example: If an unbiased coin or die is thrown, each face may be expected to be observed approximately the same number of times in the long run.

6. Define compound Events:

   The joint occurrence of two or more events is called compound events.

7. Define sample space:

   The set of all possible outcomes of a given experiment is called its sample space. Each element of the sample space is called a sample point.

   For example: When a die is thrown, the sample space S=$\{1, 2, 3, 4, 5, 6\}$

8. State the statistical definition of probability?

   If the probability of an event can be determined only after the actual happening of an event, is called statistical probability. .

   Statistical probability is also called Empirical Probability or A posterior probability. That is $P(A) = \lim_{N \to \alpha} \frac{M}{N}$

9. Define Independent Events?

   Two events are said to be independent if the occurrence of one does not affect the occurrence or non-occurrence of the other.

10. State the three axioms of probability?

    1. The probability of any event ranges from zero to one. That is $O \leq P(A) \leq 1$
    2. The probability of the entire space is 1 That is $P(S) = 1$
    3. If A and B are mutually exclusive events, then the probability of occurrence of either A or B denoted by $P(A \cup B)$ That is $P(A \cup B) = P(A) + P(B)$

11. Define non-independent events?

    If the occurrence of one event is affected by the occurrence of the other event, then the two events are said to be non-independent Events.

12. State the mathematical definition of probability?

    If the probability of an event can be calculated event before the actual happening of an event That is even before conducting the experiment, it is called mathematical probability. $P(A) = \frac{M}{N} = \frac{\text{Number of Favourable Cases}}{\text{Total Number of Exhaustive Events}}$

13. What do you meant by exhaustive events?

    Events are said to be exhaustive when their totality includes all the possible outcomes of a random experiment.

14. What do you meant by conditional probability?

Two events A and B are said to be dependent when B can occur only when A is known to have occurred. The probability attached to such an event is called conditional probability (A/B) =P (AB)/P(B)

For three events A, B and C

P (ABC) = P (A) x P (B/A) x p (C/AB).

15. Write short notes on Baye's Theorem?

Baye's Theorem is based on the concept that probabilities should be revised when some new information is available and it reduce the element of risk involved in decision making.

It is given by British mathematician Thomas bayes (1702-61). His theorem known as Baye's Theorem or Rule.

16. What is meant by prior probability and posterior probability?

Prior:

Probabilities before revision, by Baye's rule, are called prior or simply Priori because they are determined before the sample information is taken into account.

Posterior:

A probability, which has been revised in the light of sample information, is called posterior probability or revised probability.

17. Define Theoretical discrete distribution?

Distributions, which are not obtained by actual observation or experiment but are mathematically deducted on certain assumptions, are called theoretical Distribution.

18. Explain Binomial distribution?

If 'P' be the probability of success for an event and 'q' be the probability of its failure in a single trial, then the probability of exactly 'X' success and (n-x) failures in a series of 'n' independent trials is

$P(x) = NC_x\, p^r q^{n-r}$

20. State the assumptions of Binomial Distribution?

i. Trials are repeated under identical conditions for a fixed number of times.

ii. There are only two mutually exclusive outcomes (ie) success or failure.

iii. The probability of success in each trial remains constant and does not change from trial to trial.

21. What are the characteristics of a Binomial distribution?

1. It is a discrete distribution.

2. Mean = np, Variance = npq

   S.D = $\sqrt{npq}$

3. The mode of binomial distribution is that value of the variable, which occurs with the largest probability. It may be either one or two modes.
4. Skew ness = p-q / $\sqrt{npq}$

22. Explain Poisson distribution?

If the number of trials 'n' is very large and the probability of success P is very small so that the product np always exists and is finite. It relates to rare events.

23. What are conditions of Poisson distribution?

a) The number of trials 'n' is indefinitely large $n \to \alpha$

b) np = m (finite)

c) The prob. of success p for each trial is very small, ie $P \to O$

24. State the characteristics of Poisson distribution?

a) It is a discrete distribution, where random variable assumes a countable infinite number of values 0, 1, 2. . . . . .

b) Mean = m= variance, S.D =$\sqrt{}$ m

   np = m remains constant

25. Write down the density function of normal distribution?

It is defined by the probability density function.

$$Y = 1/\sigma\sqrt{2\pi} \exp [-(x-\mu)^2 / 2\sigma^2]$$

π = 3.14; e = 2.71828

Ranges from $(-\alpha + \alpha)$ $\sigma \to$ S.D

26. State the properties of N.D?

1. The normal curve is perfectly symmetrical about the mean (μ) and is bell shaped
2. It has only one mode.
3. The mean and median coincide with mode
4. The points of inflexion are at $x = \mu \pm \sigma$
5. The maximum ordinate is at $x = \mu$ Its value is $1/\sigma\sqrt{2\pi}$

27. Define population?

The word 'population' or Universe denotes aggregate or group of individual objects of any nature whose general characteristics are studied by a statistical investigation. The population may finite or infinite.

28. Define sample?

Sample is a finite sub set of the population and the number of items in a sample is called size of a sample. It may be large or small sample.

29. Define standard Error?

The standard deviation of sampling distribution of statistic is known as standard error.

30. What is a point estimate?

An estimate of a population parameter given by a single number is called a point estimate of a parameter.

31. What is interval estimates?

An interval estimate would always be specified by two values. That is the lower one and the upper one. In simple words, interval estimation refers to the estimation of a parameter by a random interval is called confidential interval.

32. Who is an estimator and estimate?

The numerical value of the sample mean is said to be an estimate of the population.

In statistical measure used that is the method of estimation is referred to as an estimate.

33. Define parameters and statistic?

Statistical constants of population namely mean (μ) and variance (σ^2) etc, which are usually referred as parameter

The statistical measures from sample observation are known as mean (x) and S.D (S), variable (S^2).

34. Explain the central limit theorem?

According to this theorem, if we selected larger number of simple random samples, say from any population and determine the mean of each sample, the distribution of these sample means will tend to be described by the normal probability distribution with a mean μ and variance σ^2/n

35. What is hypothesis?

"A hypothesis in statistics is simply a quantitative statement about a population". It is based an assumptions.

36. What is Null hypothesis?

Null hypothesis is the hypothesis, which is tested for possible rejection under the assumption that it is true and is denoted as Ho.

37. What is Alternative hypothesis?

It is the statement about the population, which gives an alternative to the null hypothesis and is denoted by H_1.

38. Define Type I and Type II error?

Rejection of the hypothesis when it should be accepted is known as Type I error.

Acceptance of a hypothesis when it should be rejected is known as Type II error.

	Accept Ho	Reject Ho
Ho is true	Correct decision	Type I error
Ho is false	Type II error	Correct decision

39. What do you meant by level of significance?

In testing a given hypothesis, the maximum probability with which we could be willing to risk is called level of significance of the test.

40. What is critical value?

The value of the test statistic, which separates the sample space into rejection region and the acceptance region, is called the critical value.

41. Explain the procedure for testing of hypothesis?

- Set up the Null hypothesis: H_o
- Set up the alternative hypothesis: H_1
- Choose an appropriate level of significance
- Calculate the test statistic

  $Z = t - \varepsilon(t) / s.\varepsilon(t)$
- Compare the computed value with the table value.

  if Z > table value → Reject the Null

  Z < table value → Accept the Null

42. Explain one-tailed test and two-tailed test?

<u>One-tailed test:</u>

In any test, the critical region is represented by a portion of the area under the probability curve of the sampling distribution of the test statistic. A test of any statistical

hypothesis where the alternative hypothesis is one tailed (right or left tailed) is called a one-tailed test.

Two-tailed test:

A test of statistical hypothesis where the alternative hypothesis is two tail.

Ho: $\mu = \mu o$ against the alternative hypothesis H_1: $\mu > \mu o$ and H_1: $\mu < \mu o$ is known as two tailed test in such case the critical region is given by the portion of the area lie in both the tails of the probability curve of the test statistic.

43. Define sampling Distribution?

"The distribution of all possible values which can be assumed by some statistic computed from samples of same size rand only drawn from the same population is called the sampling distribution of the statistic".

44. What is a non-parametric test?

The tests, which do not depend upon the population parameters such as mean and the variance, they are called non-parametric tests.

45. What do you meant by non-parametric statistics?

Non-parametric statistics is a collection of tools for data analysis that offers a different approach to many of the decision problems.

46. What are the merits of non-parametric tests?

- Non-parametric tests are distribution free. That is they do not require any assumption to be made about population.
- They are simple to understand and easy to apply when the sample sizes are small.
- Non-parametric test make fewer and less stringent assumptions than do the classical procedures.
- It is less time consuming.

47. What are the limitations of non-parametric test?

- As the sample size gets larger, data manipulations required for non-parametric procedures are sometimes laborious unless appropriate computer software is applicable.
- A collection of tabulated critical values for a variety of non-parametric tests under situations dealing with small and large 'n' is not readily available.

48. What are the methods used in non-parametric tests?

The following are the methods used in non-parametric tests. They are:

1. The sign test

2. A Rank sum test
3. The one sample Runs Test
4. The kruskal wallis or H test
5. The spearman's Rank correlation procedure

49. What do you meant by the sign test?

The sign test is the simplest of the non-parametric tests. It is based on the direction of a pair of observations and not on their numerical magnitude.

It can be used as: we count:

Number of + signs

Number of – signs

Number of O's That is which cannot be included either as positive or negative.

50. What do you meant by a rank sum test?

It is also known as Mann-Whitney U test. It is used only when two population are involved. It is also based on Ranks (R_1, R_2)

51. What do you meant by one sample runs test?

The Run test is based on order in which the sample observations are obtained. It is the useful technique for testing the null hypothesis that the observations have been indeed to be dream.

52. Define the team "Run"?

Run is defined as, "set of identical or related symbols contained between two different symbols are no symbol".

53. What is Kruskal Wallis test?

It is also known as H – test, It is the extension of Mann Whitney U – test. It is used when more than two populations are involved.

54. Define correlation?

Correlation analysis deals with the association between two or more variables.

55. What is the significance of correlation?

The following are the significances of correlation:

- There are some kinds of relationship between variables. For example relationship between price and supply, income and expenditure etc.
- The two variables are closely related. That is the estimate the value of one variable given the value of another.
- The effect of correlation is to reduce the range of uncertainty.
-

56. What is positive correlation?

If two variables tend to move together in the same direction. That is an increase in the value of one variable is accompanied by an increase in the value of other variable.

57. What is negative correlation?

If two variables tend to move together in opposite directions so that an increase or decrease in the value of one variable is accompanied by a decrease or increase in the value of other variable then the correlation is called negative or inverse correlation.

58. What do you meant by linear correlation?

If the amount of change in one variable bears constant ration to the amount of change in the other variable then the correlation is said to be linear.

59. What do you meant by non-linear correlation?

If the amount of change in one variable does not bear constant ratio to the amount of change in other variable. It is also known as curvilinear correlation.

60. What are the methods used in correlation?

The following are the various methods used in correlation.

- ❖ Scatter diagram method.
- ❖ Graphic Method.
- ❖ karl pearson's coefficient of correlation.
- ❖ Concurrent Deviation Method.
- ❖ Method of least squares.

61. What do you meant by scatter diagram?

This is a simple method of diagrammatic representation of a bivariate distribution for ascertaining the nature of correlation between two variables.

The 'n' pairs of values (x_1, y_1), (x_2, y_2) and ……. (x_n, y_n) of two variables x and y can be plotted as dots in the xy – plane. This diagram of dots so obtained is called as scatter diagram.

62. What do you meant by Karl Pearson's coefficient of correlation?

Karl Pearson, a great Biometrician and statistician suggested a mathematical method for measuring the magnitude of linear relationship between two variables. This method is widely used and is known as pearsonian coefficient of correlation. It is denoted by the symbol 'r'.

63. State the assumptions of the pearsonian coefficient?

- ❖ There is linear relationship between the variables. That is when the two variables are plotted on a scatter diagram a straight line will be formed by the points so plotted.
- ❖ There is a cause and effect relationship between the forces affecting the distribution of the items in the two series.
- ❖ Variables like height, such forces affect weight, price, demand, supply etc that a normal distribution is formed.

64. What are the limitations of Pearson's coefficient?

- ❖ The correlation coefficient always assumes linear relationship regardless of the fact whether that assumption is correct or not.
- ❖ This method takes more time to computer the value of correlation coefficient.
- ❖ The value of the coefficient is unduly affected by the extreme items
- ❖ Great care must be exercised in interpreting the value of this coefficient as very often the coefficient is misinterpreted.

65. State the properties of correlation coefficient?

1. The coefficient of correlation of correlation lies between – 1 and + 1.
2. The coefficient of correlation is independent of change of scale and origin of the variable x and y.
3. The coefficient of correlation is the geometric mean of two regression coefficients.
4. The degree of relationship between the two variables is symmetrical.

66. What do you meant by Rank correlation coefficient?

In 1904, Charles Edwin Spearman a British psychologist found out the method by determining the coefficient of correlation by ranks. This measure is useful in dealing with qualitative characteristics such as intelligence, beauty, morality, character etc.

67. State the features of spearman's correlation coefficient?

1. The sum of the difference of ranks between two variables shall be zero
   That is $\Sigma d = o$
2. Spearman's correlation coefficient is distribution free.

68. What are the Merits of Rank Method?

1. It is simple to understand and easy when compared to Kart pearson's Method.
2. Where the data are of a qualitative nature like honesty, efficiency, intelligence etc.

69. State the limitations of Rank Method?

1. This method cannot be used for finding out correlation in a grouped frequency distribution.
2. Where the numbers of items exceeds 30, the calculations become quite tedioces and require a lot of time.

70. Define the term regression?

"Regression is the measure of the average relationship between two or more variables in terms of the original units of data".

71. State the uses of regression analysis?

1. Regression analysis provides estimates of value of the dependent variable from values of the independent variable.
2. With the help of regression coefficients, we can calculate the correlation coefficient (r) and the coefficient of determination (r^2).
3. The regression analysis is highly useful and the regression line equation helps to estimate the value of dependent variable, when the values of independent variables are used in the equation.

72. State the properties of regression lines?

1. The two regression lines are $x - x = r\sigma x/\sigma y\ (y-y)$

   $$y-y = r\sigma y/\sigma x\ (x-x)$$

   These two lines pass through the point of Arithmetic Means (x,y)
2. In each regression line, one variable is dependent and the other variable is independent. In case of x on y, the variable y is independent and in case of y on x, the variable x is independent.
3. The angle between the regression lines indicates the degrees of dependence between the variables.
4. In regression equation, we find the high values of one variable are associated with the high values of the other then we conclude that correlation is positive.

   On the other hand, if it is determined from the regression equation that the high values of one variable are associated with the low values of the other, then the correlation between the two is negative.

73. Difference between correlation and Regression?

No	Correlation	Regression
1.	It is the relationship between two or more variables.	It is a mathematical showing the average relationship between two variables.
2.	Both the variables x and y are random variables.	Here x is a random variables and y is a fixed variable. Sometimes both the variables may be random variables.
3.	It finds out the degree of relationship between two variables and not the cause and effect of the variables.	It indicates the cause and effect relationship between the variables and establishes a functional relationship.
4.	It is used for testing and verifying the relation between two variables and gives limited information.	Besides verification it is used for the prediction of one value in relationship to the other given value.
5.	The coefficient of correlation is a relative measures. The range of relationship lies between -1 and +1.	Regression coefficient is an absolute figure That is If we know the value of the independent variable, we can find the value of the dependent variable.
6.	It has limited application.	It has under application.
7.	It is not very useful for further mathematical measure.	It is widely used for further mathematical measure.
8.	Correlation coefficient is independent of the origin and scale.	Regression coefficient is independent in origin but not have scale.
9.	If the coefficient of correlation is positive, then the two variables are positively correlated and vice-versa.	The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable.

74. Define time series?

A time series may be defined as a collection of readings belonging to different periods of some economic variable or composite of variables.

75. What are the importances of time series?

1. The data related to past years helps to understand the changes took place in those periods.
2. It facilitates for forecasting and planning.
3. It facilitates comparison.

76. State the various components of time series?

The following are the various components of time series.

1. Trend
2. Seasonal charges
3. Cyclical charges
4. Irregular or Random fluctuations.

The changes in the value of variable in different periods of time are due to so many factors. These factors are called the components of a time series.

77. What do you meant by trend?

The trend may be defined as the changes over a long period of time. The longer period depends upon the nature of the study or nature of the data. Trend also called as secular trend or long – term trend.

78. What do you meant by seasonal variations?

Seasonal variations are those periodic movements in business activity, which occur regularly every year and have their origin in the nature of the year itself.

79. What are the factors that cause seasonal variations?

The factors that cause seasonal variations are:

(a) Climate and weather condition.

(b) Customs, traditions and habits.

80. What do you meant by cyclical variations?

The term 'cycle' refers to the recurrent variations in time series that usually last longer than a year and are regular neither in amplitude nor in length.

Cyclical fluctuations are long term movements that represent consistently rises and declines in activity.

81. What do you meant by irregular variations?

Irregular variations are also called as erratic, accidental, and random; refer to such variations in business activities that do not repeat in definite pattern.

Irregular movements are like changes in technology, war, famine, flood, strike etc.

82. State the reasons for causing cyclical variations?

    1. Business cycle does not show regular periodicity. That is they differ widely in timing, amplitude and pattern, which makes their study very tough and tedious.
    2. The cyclical variations are mixed with erratic, random or irregular forces which make it impracticable to isolate separately the effect of cyclical and irregular forces.

83. Mention the reasons for irregular movements?

    1. To suggest that on occasions it may be possible to explain certain movements in the data due to specific causes and to simplify further analysis.
    2. To emphasize the fact that predictions of economic conditions are always subject to degree of error owing to the unpredictable erratic influences.

84. State the reasons for measuring trend?

    - ❖ To find out trend characteristics in and of themselves.
    - ❖ To enable us to eliminate trend in order to study other elements.

85. What are the various methods used for determining trend?

    1. Freehand or graphic method
    2. Semi – average method
    3. Moving average method
    4. Method of least squares

86. What are the methods for measuring seasonal variations?

    1. Method of simple averages.
    2. Ratio – to – Trend method.
    3. Ration – to – moving average method.
    4. Link Relative method.

87. What are methods used for measuring cyclical variations?

    1. Residual method.
    2. Reference cycle analysis method.
    3. Direct method
    4. Harmonic analysis method.

88. What do you meant by de-seasonalization?

    De-seasonalisation of data means elimination of the seasonal effects from the given values. It can be done for the multiplication model as well as Additional Model.

89. If the probability of the horse A winning the race is 1/5 and the probability of the horse B winning the same race is 1/6, what is the probability that one of the horses will win the race?

Probability of winning of the horse A=1/5

Probability of winning of the horse B=1/6

P(A+B)= P(A) + P(B) = 1/5+1/6 = 11/30

90. If five coins are tossed, what is the probability that all will show a head?

Let P(A) denote the probability of showing a head.

P(A) = ½

Probability of showing a head when 5 coins are tossed

= P(A)X P(A)X P(A)X P(A)X P(A) = 1/2X 1/2X 1/2X 1/2X ½ = 1/32

91. A random sample of 900 items is taken from a normal population whose mean and the variance are 4. can the sample with mean 4.5 be regarded as truly random one at 1% level of significance?

Null Hypotheses: H0 = m=4

Alternative Hypothesis: H1: m≠ 4

It is a case of two tailed test

Calculation of test statistic:

$$Z = \frac{X - m}{\sigma/\sqrt{n}} = \frac{4.5 - 4}{2/30} = 7.5$$

The modulus value of calculated Z is greater than 1% value of tabulated which is given to be 2.58. hence the null hypothesis is rejected and it is concluded that the sample cannot be regarded as a sample from a normal population with mean 4.

92. For 10 observations on price (x) and supply (y), the following data were obtained (in appropriate units)

$\sum x = 130, \sum y = 220, \sum x^2 = 2288, \sum y^2 = 5506 = 3467$

Obtain the line of regression of y on x and estimate the supply when the price is 16 units

The regression line y on x is y = a+bx, where a and b are given by the normal equation,

220 = 10a + 130 b ……. (1)

3467 = 130a + 2288 b …….(2)

Solving (1) and (2) we get, a = 8.8 and b= 1.015

Y = 8.8+ 1.015x

X = 16 to get y = 25.04

93. Ten students are selected at random from a college and their heights are found to be 100, 104, 108, 110, 118, 120, 122, 124, 126 and 128 cms. In the light of these data, discuss the suggestion that the mean height of the students of the college is 110 cms. Use 5 % level of significance.

X= 1160/10 = 116 cm

$S = \sqrt{\sum (x-x)^2} = \sqrt{864/9} = 9.798$

$n-1$

Test Statistic:

$T = \frac{x - \mu}{S/\sqrt{n}} = \frac{(116-110)}{9.798} \times \sqrt{10} = 1.937$

94. Find Mean and Variance for all possible sequence consisting of three a's and two b's and give the number of runs, V, corresponding to each sequence.

$\mu v = 2n1\ n2/(n1+ n2)+ 1 = 2x3x2/(3+2) + 1 = 17/5$

$\sigma v^2 = 2n1n2(2n1n2-n1-n2)/ (n1+n2)^2 (n1+n2-1) = 2x3x2(2x3x2-3-2)/(3+2)^2 (3+2-1)$

$= 21/25$

95. A random variable X is normally distributed with meanμ = 12 and S.D = 2 find P(9.6<X<13.8) Given that the area: A = 0.3159 for Z = 0.9 and a = 0.3849 for Z = 1.2 where Z is a standard normal variate

Standardised variate $Z = x - \mu / \sigma$

μ= 12, σ = 2

X = 9.6, then Z = 9.6-12 / 2 = -1.2

When X = 13.8, then Z = 13.8 -12 / 2 = 0.9

P(9.6<X<13.8) = P(-1.2<Z<0.9)

= 0.3849 + 0.3159 = 0.7008

96. A factory is producing 50,000 pairs of shoes dialy. From a sample of 500 pairs 2% were found to be of substandard quality. Estimate the number of pairs that can be reasonably expected to be spoiled in the daily production and assign limits at 95% level of confidence.

N=500, sample percentage p = 2%

S.E. (p) = $\sqrt{}$ p(100-p) / n

= $\sqrt{}$ 2(100 – 2)/ 500 = $\sqrt{}$ 2x 98/500 = 0.63

(1- α) % = 95 % Zα = 1.96

(2- 1.23 and 2 + 1.23 = 0.77 and 3.23

Number of pair of shoes that can be expected to be spoiled in daily production between 0.77/100 X 50,000 = 385 and 3.23/100 X 50,000 = 1615

97. In a survey of 200 boys of which 75 were intelligent, 40 had educated fathers, while 85 of the unintelligent boys had uneducated fathers. Do these figures support the hypothesis that educated fathers have intelligent boys?

Let the null hypothesis be that the education of fathers has no effect on the intelligence of boys.

	Intelligent boys	Unintelligent boys	Total
Educated fathers	40(30)	40(50)	80
Un-educated fathers	35(45)	85(75)	120
Total	75	125	200

Expected frequencies are given in brackets and are calculated by multiplying row total by column total and then dividing it by gross total.

$\chi^2 = \sum (O-E)^2 = \frac{(40-30)^2}{30} + \frac{(40-50)^2}{50} + \frac{(35-45)^2}{45} + \frac{(85-75)^2}{75} = 8.88$

Since the caluculated value is higher than the tabular value of 3.841, so we reject the null hypothesis ie. The education of fathers has an effect on the intelligence of boys.

98. For a certain bivariate data, the equation of line of regression of Y on X is 2.5, Y=X+35 and X= 15, given Y = 20. the equation of line of regression X on Y is 10X = Y + 70 and Y = 20 in this equation gives X = 9 and not the original value of X = 15. Explain why this is so. For what values of X and Y does this process give consistent results?

Solving the equations 2.5Y = X + 35 ……. (i)

Y = -10X + 70 …….(ii)

X = 35/4, Y = 35/2

Hence equation will give consistent results of the both X and Y.

99. It is known that the population standard deviation in waiting time for L.P.G. gas cylinder in Delhi is 15 days. How large a sample should be chosen to be 95% confident, the waiting time is within 7 days of true average.

The required sample size

$n = (\sigma Z/E)^2 = (15 \text{ X } 196/7)^2 = 17.64$

Hence the size of the sample is 18.

100. A firm wishes to estimate with an error of not more than 0.03 and a level of confidence of 98% the proportion of consumers that prefers its brand of household detergent. Sales report indicates that about 0.20 of all consumers prefer the firm's brand. What is the requisite sample size?

$N = Z^2 PQ/E^2 = (2.33)^2 \text{X } 0.2 \text{ X } 0.8/ (0.03)^2$
$= 965.14$ [Z = 2.33 from the Table]
Hence the sample size n = 965.

Part – B

1. Between 2 and 4 P.M the average number of phone calls per minute coming into the switch board of a company is 2.5. Find the probability that during one particular minute there will be (i) no phone call at all (ii) exactly 3 calls

(Given $e^{-2.5} = 0.13534$ $e^{-0.5} = 0.13534 = 0.060650$)

$P(0) = \frac{e^{-2.5}\ (2..5)0}{0!} = e^{-2.5} \text{ X } e^{-0.5} = 0.13534 \text{ X } 0.60650 = 0.0821$

$P(3) = \frac{e^{-2.5}(2.5)^3}{3!} = 0.214$

2. In a certain factory manufacturing razor blades, there is a small chance, 1/50 for any blade to be defective. The blades are placed in packets, each containing 10 blades. Using the Poisson distribution, calculate the approximate number of packets containing not more than 2 defective blades in a consignment of 10,000 packets.

$M = np = 10 \text{ x } 1/50 = 0.2$

$P(0) = e^{-m} = e^{-0.2} = 0.8187$

$NX\ P(0) = P(0) \text{ X } 10{,}000 = 0.8187 \text{ X } 10{,}000 = 8187$

$NXP(1) = mXNX\ P(0) = 1637.4 \text{ X } P(2) = m/2 \text{ X } N \text{ X } P(1) = 163.74$
Required number = [8187 + 1637.4 + 163.74] = 9988.

3. The two regression line obtained from certain data were Y = X + 5 and 16X = 9Y – 94. Find the variance of X, if the variance of Y is 16. Also find the covariance between X and Y.

The two regression lines are given by
$Y = X + 5$ …….. (1)
$16X = 6Y - 94$
$X = 9/16\ Y - 47/8$……..(2)

$B_{yx} = 1$, $b_{xy} = 9/16$

$r^2 = B_{yx} + b_{xy}$
$r = \sqrt{1 X 9/16} = ¾$
$b_{xy} = r \sigma x/\sigma y$
$16 = 6/4 \sigma x/4\sqrt{16}$
$\sigma x = 3$
$Cov (X,Y)/ \sigma x.\sigma y = r$

$Cov (X,Y) = ¾ \times 3 \times 4 = 9$

4.Find the coefficient of concurrent deviations for the following data:

Year	1984	1985	1986	1987	1988	1989	1990
Supply	150	154	160	172	160	165	180
Price	200	180	170	160	190	180	172

N= Total number of observation years = 7
n= N-1 = 7 – 1 = 6

$C = \sum C \times Cy = 0$
$r_c = \pm \sqrt{} \pm (2x0 – 6)/ 6 = - \sqrt{} - (-1)$
$= \sqrt{} + 1 = - 1$, which is negative, we shall multiply it by (-1) to get $\sqrt{} +1 = 1$. the value of r_c Will be negative as $2C – n / n$ is negative.

5. The Standard deviation calculated from two random samples of size 9 and 13 are 2 and 1.9 respectively. May the sample be regarded as drawn from the normal population with the same standard deviation.

H0 : $\sigma 1^2 = \sigma 2^2$
H1: $\sigma 1^2 \neq \sigma 2^2$
$\sigma 1^2 = ni\ S1^2/ n1 – 1 = 9/8 x4 = 4.5$

$\sigma 2^2 = 13/12x3.61 = 3.91$

$F = \sigma 1^2/ \sigma 2^2 = 4.5 / 3.91 = 1.15 < F_{0.05} = 3.51.$

6. In the accounting department of a bank 100 accounts are selected at random and examined for errors. The following results have been obtained.

No. of errors:	0	1	2	3	4	5	6
No. of Accounts:	36	40	19	2	0	2	1

Does this information verify that the errors are distributed according to Poisson probability law?

Soln.

1. Null Hypothesis Ho: The errors are distributed according to Poisson Probability law

Alternative Hypothesis H1: The errors are not distributed according to Poison Probability law.

2. Computation of Test Statistics:

Mean of the given data = m = $\frac{\Sigma f X x}{N}$

$\frac{0x36+ 1x40+2x19+3x2+4x0+6x1}{36+40+19+2+0+2+1} = 100/100 = 1$

$e^{-1} = 0.367$

Poisson Probability distribution = $P(x) = \frac{e^{-m} m^{x}}{x!}$, x = 0,1,2,3…….

Expected Frequencies = $Nxp(x) = N X \frac{e^{-m} m^{x}}{x!}$, x = 0,1,2,3…….

$= \frac{0.637 \times 100 \times 1^{x}}{x!}$, x = 0,1,2,3…….

The expected probabilities for the various errors are

$\frac{0.367 X 100 X 1^{0}}{0!}, \frac{0.367 X 100 X 1^{1}}{1!}, \frac{0.367 X 100 X 1^{2}}{2!}, \frac{0.367 X 100 X 1^{3}}{3!},$

$\frac{0.367 X 100 X 1^{4}}{4!}, \frac{0.367 X 100 X 1^{5}}{5!}, \frac{0.367 X 100 X 1^{6}}{6!},$

7.Calculate the correlation co-efficient between the height of sisters and height of brothers from the given data

Height of sisters (in cm)	64	65	66	67	68	69	70
Height of brothers (in cm)	66	67	65	68	70	68	72

$\bar{x}$ 67, y = 68

$\sum x = 0$, $\sum Y = 0$

$\sum x^2 = 28$, $\sum y^2 = 34$

$\sum xy = 25$

$R = \sum xy / \sqrt{x^2} X \sqrt{y^2} = 25/ \sqrt{28x 34} = 0.81$

8. Given r = 0.8, $\sum xy$, $\sum Y = 2.5$ and $\sum x^2 = 90$. find the number of items (x and y are deviations from their respective A. M.)

$\sigma x^2 = 1/n \sum x^2 = 90/n$

$r = \sum (x-X) (y-Y)/n\sigma x\sigma y = \sum xy / n\sigma x \sigma y$

$r^2 = 0.64$
$n = 10$

9. If $N = 50$, $\sum x = 75$, $\sum y = 80$ $\sum x^2 = 150$ $\sum y^2 = 120$. Find the value of r.

$$\frac{120 - 75 \times 80/50}{\sqrt{130 - (75)^2/5}\ \sqrt{140 - (80)^2/50}}$$

$= 0$

10. For the following data, calculate the coefficient of Rank correlation

X:	80	91	99	71	61	81	70	59
Y:	123	135	154	110	105	134	121	106

$$R = 1 - \frac{6 \sum D^2}{N(n^2-1)} = 1 - \frac{6 X4}{8(64-1)}$$

$1-3/63 = 60/64 = 20/21$

11. The coefficient of rank correlation of marks obtained by 10 students, in English and Economics was found to be 0.5. It was later discovered that the difference in ranks in the two subjects obtained by one student was wrongly taken as 3 instead of 7. Find the correct co-efficient of rank correlation.

$R = 0.5$, $N = 10$

$6 \sum D^2 = 82.5$

Corrected $R \approx 0.26$

12. Find the coefficient of correlation r when its probable error is 0.2 and the number of pairs of items is 9.

P.E. = 0.2

$N = 9$

$R = \sqrt{0.1105} = 0.332$

13. State the reasons for measuring trend? What are the various methods used for determining trend?

a) To find out trend characteristics in and of themselves.
b) To enable us to eliminate trend in order to study other elements.

c) Freehand or graphic method
d) Semi – average method
e) Moving average method
f) Method of least squares

14. What is a non-parametric test? What are the merits and demerits of non-parametric tests?

a) The tests, which do not depend upon the population parameters such as mean and the variance, they are called non-parametric tests.
b) Non-parametric statistics is a collection of tools for data analysis that offers a different approach to many of the decision problems.
c) Non-parametric tests are distribution free. That is they do not require any assumption to be made about population.
d) They are simple to understand and easy to apply when the sample sizes are small.
e) Non-parametric test make fewer and less stringent assumptions than do the classical procedures.
f) It is less time consuming.
g) As the sample size gets larger, data manipulations required for non-parametric procedures are sometimes laborious unless appropriate computer software is applicable.
h) A collection of tabulated critical values for a variety of non-parametric tests under situations dealing with small and large 'n' is not readily available.

15. What are the methods used in non-parametric tests?

The following are the methods used in non-parametric tests. They are:

1. The sign test
2. A Rank sum test
3. The one sample Runs Test
4. The kruskal wallis or H test
5. The spearman's Rank correlation procedure

16. Explain Binomial distribution? . State the assumptions of Binomial Distribution?

If 'P' be the probability of success for an event and 'q' be the probability of its failure in a single trial, then the probability of exactly 'X' success and (n-x) failures in a series of 'n' independent trials is

$P(x) = NC_x\, p^r q^{n-r}$

Trials are repeated under identical conditions for a fixed number of times.

There are only two mutually exclusive outcomes (ie) success or failure.

iv. The probability of success in each trial remains constant and does not change from trial to trial.

17. What are the characteristics of a Binomial distribution?

1. It is a discrete distribution.
2. Mean = np, Variance = npq

   S.D = $\sqrt{npq}$
3. The mode of binomial distribution is that value of the variable, which occurs with the largest probability. It may be either one or two modes.
4. Skew ness = p-q / $\sqrt{npq}$

18. Explain the following

1. hypothesis 2.Null hypothesis 3. Alternative hypothesis?

"A hypothesis in statistics is simply a quantitative statement about a population". It is based an assumptions. Null hypothesis is the hypothesis, which is tested for possible rejection under the assumption that it is true and is denoted as Ho.It is the statement about the population, which gives an alternative to the null hypothesis and is denoted by H_1.

19.Define Type I and Type II error?

Rejection of the hypothesis when it should be accepted is known as Type I error.

Acceptance of a hypothesis when it should be rejected is known as Type II error.

20.What do you meant by level of significance and two tailed test

In testing a given hypothesis, the maximum probability with which we could be willing to risk is called level of significance of the test.

<u>Two-tailed test:</u>

A test of statistical hypothesis where the alternative hypothesis is two tail.

Ho: $\mu = \mu o$ against the alternative hypothesis H_1: $\mu > \mu o$ and H_1: $\mu < \mu o$ is known as two tailed test in such case the critical region is given by the portion of the area lie in both the tails of the probability curve of the test statistic.