इन्दिरा गाँधी राष्ट्रीय मुक्त विश्वविद्यालय
INDIRA GANDHI NATIONAL OPEN UNIVERSITY

सत्रीय कार्य का प्रथम पृष्ठ

अध्ययन केन्द्र : इग्नू अध्ययन केन्द्र (27.....)
Study Centre : IGNOU Study Centre, (27..195..)

कार्यक्रम : MAPC : MA IN PSYCHOLOGY
Programme :

अनुक्रमांक संख्या : 2 0 0 1 4 8 5 9 0 7
Enrolment Number :

नाम : Mrs ASHUMITA KUMAR
Name :

दूरभाष : 8989753646, 6392839120
Mobile :

पाठ्यक्रम कोड : MPC-006
Course Code :

पाठ्यक्रम शीर्षक : STATISTICS IN PSYCHOLOGY
Course Title :

सत्र : जनवरी – 2020 / जुलाई – .....
Session :

Ashu kumar
(हस्ताक्षर)

MPC-006 : STATISTICS IN PSYCHOLOGY
Tutor Marked Assignment (TMA)

Course Code: MPC-006
Assignment Code: MPC-006/AST/TMA/2019-2020
Marks: 100

NOTE: All Questions Are Compulsory

SECTION A

Answer the following question in about 1000 words (wherever applicable) each — **15 *x* 3 = 45 Marks**

1. Discuss in detail organization of data.
2. Compute ANOVA for the following data: that indicates the scores obtained by three group on employees on Job Satisfaction scale:

Group A	56	66	55	66	45	34	33	23	34	33
Group B	23	34	33	44	55	76	43	35	57	34
Group C	34	56	54	33	56	34	23	24	26	34

3. Discuss divergence in normality with the help of suitable diagram and describe the factors causing divergence in the normal distribution. Discuss how divergence in normality is measured.

SECTION B

Answer the following questions in about 400 words (wherever applicable) each — **5 *x* 5 = 25 Marks**

4. Explain scales of measurement and discuss assumption of parametric statistics.
5. Using Spearman's rho for the following data:

Data 1	45	44	34	36	33	31	56	54	39	33
Data 2	12	15	22	13	14	11	16	10	19	20

MPC-006 : STATISTICS IN PSYCHOLOGY

SECTION - A

Q.1 Discuss in detail organization of data.

Ans. **INTRODUCTION**

The word statistics has different meaning to different persons. For some, it is a one number description of a set a data. Some consider statistics in terms of numbers used as measurements or counts. Mathematicians use statistics to describe data in one word. In behavioural sciences the word 'statistics' means something different, that is its prime function is to draw statistical inference about population on the basis of available quantitative and qualitative information.

Here statistics described as a branch of science which deals with the collection of data, their classification, analysis and interpretations of statistical data.

The science of statistics may be broadly studied under two headings i) Descriptive statistics and ii) Inferential statistics.

Descriptive statistics is a branch of statistics which deals with descriptions of obtained data. It include classification, tabulation, diagrammatic and graphical presentation of data, measures of central tendency and variability.

ORGANISATION OF DATA

These are four major statistical techniques for organising the data which enable the researchers to know about the tendency of data or the scores, and the ease in description of the phenomena. These are:

i) CLASSIFICATION

The arrangement of data in groups according to similarities is known as classification. A classification is a summary of the frequency of individual scores or ranges of scores for a variable. Once data are collected, it should be arranged in a format from which they would be able to draw some conclusions. Thus by classifying data, the investigators move a step ahead in regard to making a decision.
A much clear picture of the information of score emerges when the raw data are organised as a frequency distribution. Frequency distribution shows the number of cases following within a given class interval or range of scores.

Frequency Distribution can be with Ungrouped Data and Grouped Data
An ungrouped frequency distribution may be constructed by listing all score values either from highest to lowest or lowest to highest and placing a tally mark (|) besides each scores every times it occurs.

Grouped Frequency Distribution: A group frequency distribution is a table that organises data into classes. It shows the number of observations from the data set that fall into each of the class.

Construction of Frequency Distribution

To prepare a frequency distribution it is essential to determine the following:

i) The range of the given data : Difference between highest and lowest scores.

ii) The number of class intervals : The number of classes should be between 5 to 30

iii) Limits of each class Interval : Number of classes is the size/ width or range of the class

There are three methods for describing the **class limits for distribution** :

i) Exclusive Method : In this method the upper limit of one class become the lower limit of the next class.

ii) Inclusive method : This classification includes scores which are equal to the upper limit of the class.

iii) True or Actual class method : A score is an internal when it extends from 0.5 units below to 0.5 units above the face value of the score on a continuum. These class limits are known as true or actual class limits. (29.5 to 39.5, 39.5 to 49.5) etc.

Types of Frequency Distribution

There are various ways to arrange frequencies of a data array based on the requirement of the statistical analysis Some are here:

i) Relative Frequency Distribution: This distribution indicates the propotion of the total numbers of cases observed at each score value.

ii) Cumulative frequency Distribution: A cumulative frequency corresponding to a class-interval is the sum of frequencies for that class and of all classes prior to that class.

iii) Cumulative relative frequen: In this distribution the entry of any score of class interval expresses that score's cumulative frequency as a proportion of the total numbers of cases.

ii) TABULATION

Frequency distribution can be either in the form of a table or it can be in the form of graph. Tabulation is the process of presenting the classification data in the form of a table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. A table is a systematic arrangement of classified data in row and columns with appropriate headings and sub-headings. The main components of a table are:

i) **Table Number**: When there is more than one table in a particular analysis a table should be marked with a number for their reference and identification.

ii) **Title of the table**: Every table should have an appropriate title, which describe the content of the table.

iii) **Caption**: Caption are brief and self-explanatory headings for columns.

iv) **Stub**: stubs stand for brief and self-explanatory headings for rows.

v) **Body of the table**: This is the real table and contains numerical information or data in different cells.

vi) **Head Note**: This is written at the extreme right hand below the title and explains the unit of the measurements used in the body of tables.

vii) **Foot Note**: This is a qualifying statement which is to be written below the table explaining certain points related to the data which have not been covered in title, caption and stubs.

viii) **Source of data**: The source from which data have been taken is to be mentioned at the end of the table.

Example of Tabulation

TITLE

Stub Head	Caption			
Stub Entries	Column Head I		Column Head II	
	Sub Head	Sub Head	Sub Head	Sub Head
	MAIN BODY	OF	THE TABLE	
Total				

Foot note(s):

Source :

iii) GRAPHICAL PRESENTATION OF DATA

The purpose of preparing a frequency distribution is to provide a systematic way of "looking at" and understanding data. To extend this understanding, the information contained in a frequency distribution often is displayed in graphic and/or diagrammatic forms. In graphical presentation of frequency distribution, frequencies are plotted on a pictorial platform formed of horizontal and vertical lines known as graph.

A graph is created on two mutually perpendicular lines called the X and Y-axes on which appropriate scales are indicated. The horizontal line is called the abscissa and vertical the ordinate.

Some graphical patterns used in statistics which enhance the scientific understanding of the reader. These are:

i) Histogram: The histogram consists of series of rectangles, with its width equal to the class interval of the variable on horizontal axis and the corresponding frequency on the vertical axis as its heights.

ii) Frequency Polygon: In this graph mark each frequency against its concerned class on the height of its respective ordinate.

iii) Frequency curve: It is a smooth free hand curve drawn through frequency polygon.

iv) Cumulative Frequency Curve or Ogive: There are two types of cumulative frequency distribution "less than" and "More than" cumulative frequency.

iv) DIAGRAMMATIC PRESENTATION OF DATA

A diagram is a visual form for the presentation of statistical data. They present the data in simple, readily comprehensible form. Diagrammatic presentation is used only for presentation of the data in visual form, whereas graphic presentation of the data can be used for further analysis. There are different forms of diagram e.g. Bar diagram, Sub-divided bar diagram, Multiple bar diagram, Pie diagram and Pictogram.

Bar Diagram : Bar diagram is most useful for categorical data. It is drawn from the frequency distribution table representing the variable on the horizontal axis the frequency on the vertical axis.

Sub-divided bar diagram: Study of sub classification of a phenomenon can be done by using sub-divided bar diagram corresponding to each sub-category of the data the bar is divided and shaded.

Multiple Bar diagram : This diagram is used when comparisons are to be shown between two or more sets of interrelated phenomena or variables.

Pie diagram : It is also known as angular diagram. A pie chart or diagram is a circle divided into component sectors corresponding to the frequencies of the variables in the distrib

Each sector corresponding to the frequencies of the variables in the group. A circle represents 360°. So 360° angles is divided in propotion to percentages.

CONCLUSION :

Descriptive statistics are used to describe the basic features of the data in investigation. Such statistics provide summaries about the sample and measures. Data description comprises two operations: organising data and describing data. Organising data includes: classification, tabulation, graphical and diagrammatic presentation of raw scores. These measures enable the researchers to know about the tendency of data or the scores.

Q. 2. Compute ANOVA for the following data: that indicates the scores obtained by three group on employees on Job Satisfaction Scale:

GROUP-A	56	66	55	66	45	34	33	23	34	33
GROUP-B	23	34	33	44	55	76	43	35	51	34
GROUP-C	34	56	54	33	56	34	23	24	26	34

Ans. SOLUTION:

In the question there are three groups $n = 10$ each, Total number of scores $N = 30$

Null Hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$

Three group on employees on job satisfaction do not differ in their job satisfaction level. Thus

	GROUP-A		GROUP-B		GROUP-C	
S.No	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
1	56	3136	23	529	34	1156
2	66	4356	34	1156	56	3136
3	55	3025	33	1089	54	2916
4	66	4356	44	1936	33	1089
5	45	2025	55	3025	56	3136
6	34	1156	76	5776	34	1156
7	33	1089	43	1849	23	529
8	23	529	35	1225	24	576
9	34	1156	57	3249	26	676
10	33	1089	34	1156	34	1156
Sum	$\Sigma X_1 = 445$	$\Sigma X_1^2 = 21917$	$\Sigma X_2 = 434$	$\Sigma X_2^2 = 20990$	$\Sigma X_3 = 374$	$\Sigma X_3^2 = 1552$

$$\text{Mean} = \frac{\Sigma X_1}{n} = \frac{445}{10} = 44.5$$

$n = 10$, $N = 30$

$$\frac{\Sigma X_2}{n} = \frac{434}{10} = 43.4$$

$$\frac{\Sigma X_3}{n} = \frac{374}{10} = 37.4$$

Step-1 Correction term (C_x)

Formula = $C_x = \frac{\Sigma(X)^2}{N} = \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{n_1 + n_2 + n_3}$

$$C_x = \frac{(445 + 434 + 374)^2}{10 + 10 + 10}$$

$$= \frac{(1253)^2}{30}$$

$$= \frac{1570009}{30}$$

$$C_x = 52333.63$$

Step-2 SS_T (Sum of Squares of total)

Formula - $SS_T = \Sigma X^2 - C_x$ or $(\Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2) - \frac{(\Sigma X)^2}{N}$

$$SS_T = (21917 + 20990 + 15526) - 52333.63$$

$$= 58433 - 52333.63$$

$$= 6099.37$$

$$SS_T = 6099.37$$

Step-3 SS_A (Sum of Squares among the groups)

Formula - $SS_A = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - C_x$

$$SS_A = \frac{(445)^2}{10} + \frac{(434)^2}{10} + \frac{(374)^2}{10} - 52333.63$$

$$= \frac{198025}{10} + \frac{188356}{10} + \frac{139876}{10} - 52333.63$$

$$= \frac{526257}{10} - 52333.63$$

$$= 52625.7 - 52333.63$$

$$= 292.07$$

$$SS_A = 292.07$$

Step-4 SS_W (Sum of Squares within the groups)

Formula - $SS_W = SS_T - SS_A$

$$SS_W = 6099.37 - 292.07$$

$$SS_W = 5807.3$$

Step-5 MSS_A (Mean Sum of Squares among the groups)

Formula - $MSS_A = \frac{SS_A}{K-1}$

$$= \frac{292.07}{3-1}$$

$$MSS_A = \frac{292.07}{2}$$

$$= 146.035$$

$$MSS_A = 146.035$$

Step-6 MSS_w (Mean Sum of Squares within the groups)

$$\text{Formula} - MSS_w = \frac{SS_w}{N-K}$$

$$= \frac{5807.3}{30-3} = \frac{5807.3}{27} = 215.085$$

$$MSS_w = 215.085$$

Step-7 F-Ratio = $\text{Formula} = \frac{MSS_A}{MSS_w}$

$$= \frac{146.035}{215.085} = 0.680$$

$$F = 0.680$$

Step-8 Summary of ANOVA

Source of Variance	df	SS	MSS	F Ratio
Among the Groups	(K-1) 3-1=2	292.07	146.035	0.680
within the Groups	(N-K) 30-3=27	5807.3	215.085	
Total	29			

From F table for 2 and 27 df at .05 level, the F value is 19.48. Our calculated F value is 0.680, which is much lower than the critical value at 5% and 1% level. Therefore the obtained F ratio is not significant at .05 level and 1% level of significance at 2 and 27 df. Thus the Null hypothesis (Ho) is accepted.

Because null hypothesis is accepted at .05 and .01 level of Significance therefore with 95% confidence it can be said that the job satisfaction of employees do differ.

Q.3 Discuss divergence in normality with the help of suitable diagram and describe the factors causing divergence in the normal distribution. Discuss how divergence in normality is measured.

Ans. INTRODUCTION

In the normal curve model, the mean, the median and the mode all coincide and there is perfect balance between the right and left halves of the curve. Generally two types of divergence occur in the normal curve.
- Skewness
- kurtosis

SKEWNES

A distribution is said to be "skewed" when the mean and median fall at different points in the distribution and the balance i.e. the point of center of gravity is shifted to one side or the other to left or right. In a normal distribution the mean equals, the median exactly and the skewness is of course zero ($S_K = 0$).

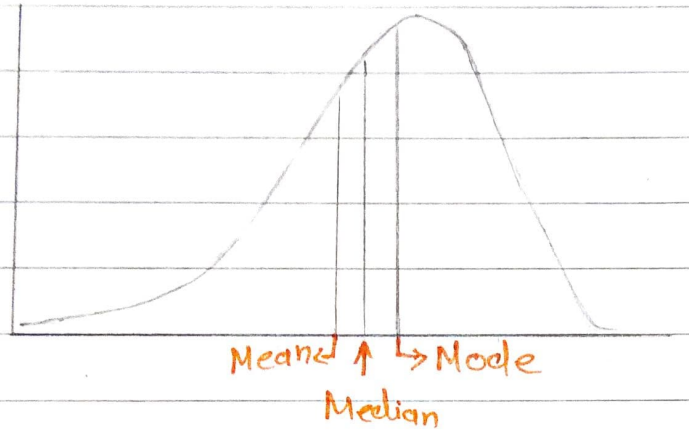
There are two types of skewness which appear in the normal curve.

a) Negative skewness

Distribution said to be skewed Negatively or to the left when scores are massed at the high end of the scale, i.e.

the left side of the curve. In negatively skewed distribution the value of median will be higher than that of the value of the mean.

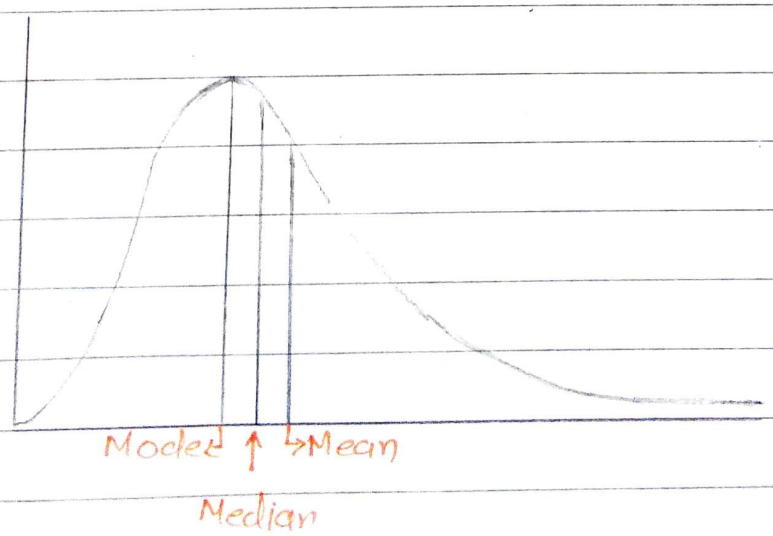
skewed to left



b) Positive skewness

Distributions are skewed positively or to the right, when scores are massed at the low, i.e. the left end of the scale, and are spread out gradually toward the high or right end as shown below.

skewed to right



KURTOSIS

The term kurtosis refers to the divergence in the height of the curve, specially in the peakness. There are two types of divergence in the peakness of the curve.

a) Leptokurtosis
b) Platy kurtosis

a) Lepto kurtosis

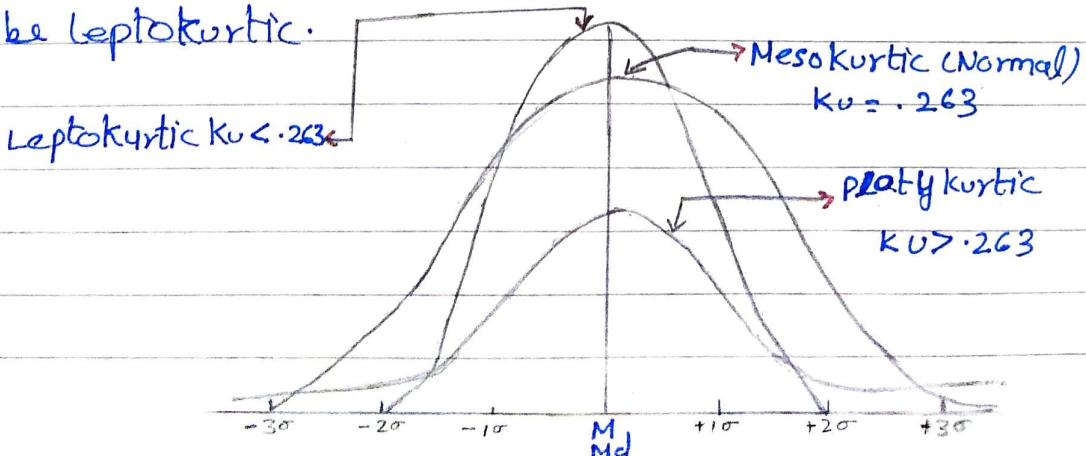
Lepto kurtosis distribution. the frequency distribution curve is more peaked than to the normal distribution curve.

b) Platy kurtosis

Platy kurtosis distribution of flatter peak than to the normal is known Platy kurtosis distribution.

Example:

When the distribution and related curve is normal, the vain of kurtosis is 0.263 (KU = 0.263). If the values of the KU is greater than 0.263, the distribution and related curve obtained will be platykurtic. When the values of KU is less than 0.263, the distribution and related curve obtained will be leptokurtic.



FACTORS CAUSING DIVERGENCE IN THE NORMAL DISTRIBUTIO

The reasons on why distribution exhibit skewness and kurtosis are numerous and often complex, but a careful analysis of the data will often permit the common causes of asymmetry. Some of common causes are –

i) **SELECTION OF THE SAMPLE**

Selection of the subjects (individuals) produce skewness and kurto in the distribution. If the sample size is small or sample is biased ou skewness is possible in the distribution of scores obtained on the basi of selected sample or group of individuals.

If the scores made by small and homogeneous group are likely yield narrow and leptokurtic distribution. Scores from small and highly hetrogeneous groups yield platykurtic distribution.

ii) **UNSUITABLE OR POORLY MADE TESTS**

If the measuring tool or test is unappropriate, or poorly made, th asymmetry is possible in the distribution of scores. If a test is to easy, scores will pile up at the high end of the scale, whereas test is too heard, scores will pile up at the low end of the scale.

iii) **THE TRAIT BEING MEASURED IS NON-NORMAL**

skewness or kurtosis or both will appear when there is a rea lack of normality in the trait being measured, e.g. interest, attitude, suggestibility, deaths in a old age or early childhood

due to certain degenerative diseases etc.

iv) ERRORS IN THE CONSTRUCTION AND ADMINISTRATION OF TEST

The unstandardised with poor item-analysis test may cause asymmetry in the distribution of the scores. Similarly, while administrating the test, the nuclear instructions - Error in timings, Errors in the scoring, practice and motivation to complete the test all the these factors may cause skewness in the distribution.

MEASURING DIVERGENCE IN THE NORMAL DISTRIBUTION

Measuring Skewness

There are two methods to study the skewness in a distribution.

i) OBSERVATION METHOD

There is a simple method of detecting the directions of skewness by the inspection of frequency polygon prepared on the basis of the scores obtained regarding a trait of the population or a sample drawn from a population.

Looking at the tails of the frequency polygon of the distribution obtained, if longer tail of the curve is towards the higher value or upper side or right side to the centre or mean, the skewness is positive. If the longer tail is towards the lower values or lower side or left to the mean, the skewness is negative.

ii) STATISTICAL METHOD

To know the skewness in the distribution we may also use the statistical method. For the purpose we use measures of central tendency, specifically mean and median values and use the following formula.

$$S_K = \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

Another measure of skewness based on percentile values is as under

$$S_K = \frac{P_{90} - P_{10}}{2} - P_{50}$$

Here, it is to be kept in mind that the above two measures ar not mathematically equivalent. A normal curve has the value of $S_K = 0$. Deviations from normality can be negative and positively direction leading to negatively skewed and positive skewed distributions respectively.

Measuring kurtosis

For judging whether a distribution lacks normal symmetry o peakedness; it may detected by inspection of the frequency poly obtained. If a peak of curve is thin and sides are narrow to th centre, the distribution is leptokurtic and if the peak of frequency distribution is too flat and sides of the curve are d viating from the centre towards $\pm 4\sigma$ or $\pm 5\sigma$ than the distri -tion is platykurtic.

Kurtosis can be measured by following formula using percentile values.

$$K_U = \frac{Q}{P_{90} - P_{10}}$$

where Q = quartile deviation i.e.
P_{10} = 10th percentile
P_{90} = 90th percentile

A normal distribution has $K_U = 0.263$. If the value of K_U is less than 0.263 ($K_U < 0.263$), the distribution is lepto-kutic and if K_U is greater than 0.263 ($K_U > 0.263$), the distribution is platykurtic.

CONCLUSION

There are two types of divergence, skewness and kurtosis. factors causing divergence in the normal distribution are selection of sample, unsuitable or poorly made test, the trait being measured is not normal and errors of construction and administration of test. Divergence can be measured by observation and statistical method.

SECTION - B

Q.4 Explain scales of measurement and discuss a assumption of parametric statistics.

Ans- SCALES OF MEASUREMENT

The scales of measurement of the dependent variable helps us to choose the broad category of statistical procedures appropriate for our hypothesis (nonparametric or parametric) The Scale of measurement of the independent variable helps us to determine which statistical procedure within the broad category is appropriate.

These are four types of scales used in measurement viz., Nominal scale, Ordinal scale, Interval scale, and Ratio scale.

NOMINAL SCALE :

Nominal scale deals with nominal data or classified data such as for example the population divided into males and females. There is no ordering of the data in that it has no meaning when we say Male > female. These data are also given arbitrary labels such as m/f and 1/0.

ORDINAL SCALE :

Ordinal scale deals with interval data. These are in certain orde but the differences between values are not important. For example

degree of satisfaction ranging in a 5 point scale of 1 to 5, with 1 indicating least satisfaction and 5 indicating high satisfaction.

INTERVAL SCALE

Interval scale deals with ordered data with interval. This is a constant scale but has no natural zero. Differences do make sense. Example of this kind of data includes for instance temperature in Centigrade or Fahrenheit. Interval scale possesses two out of three important requirements of a good measurement scale, that is, magnitude and equal intervals but lacks the real or absolute zero point.

RATIO SCALE :

Ratio scale deals with ordered, constant scale with a natural zero. Example of this type of data include for instance, height, weight, age, length etc.

Once we have identified the independent and dependent variables, our next step in choosing a statistical test is to identify the scale of measurement of the variables. All of the parametric tests require and Interval or Ratio Scale of measurement for the dependent variable. Many psychologists also apply paramatric tests to variables with an approximately interval scale of measurement.

ASSUMPTIONS OF PARAMETRIC STATISTICS

Parametric tests normally involve data expressed in absolute numbers or values rather than ranks; an example is the Student's t-test.
The parametric statistical test operates under certain conditions. Since these conditions are not ordinarily tested, they are assumed to hold valid. The meaningfulness of the results of a parametric test depends on the validity of the assumption. proper interpretation of parametric test based on normal distribution also assumes that the scene being analysed results from measurement in at least an interval scale.

Parametric test like, t and f tests may be used for analysing the data which satisfy the following conditions:

i) The population from which the sample have been drawn should be normally distributed.

ii) Normal Distributions refer to frequency distribution following a normal curve, which is infinite at both the ends.

iii) The variables involved must have been measured interval or ratio scale.

iv) variable and its types that can have different values.

v) The observation must be independent. The inclusion or exclusion of any case in the sample should not unduly affect the result of study.

vi) These population must have the same variance or, in special cases, must have a known ratio of variance. This we call homosedasticty.

vii) The samples have equal or nearly equal variance. This condition is known as equality or homogeneity of variances and is particularly important to determine when the sample are small.

viii) The observations are independent. The selection of one case in the sample is not dependent upon the selection of any other case.

2·5 Using Spearman's rho for the following data :

Data - 1	45	44	34	36	33	31	56	54	39	33
Data - 2	12	15	22	13	14	11	16	10	19	20

Ans. Calculation of Spearman's rho aforsaid data :

Data-1 (X)	Data-2 (Y)	Rank X	Rank Y	(Rank X)^2	(Rank Y)^2	(Rank X)(Rank Y)
45	12	8	3	64	9	24
44	15	7	6	49	36	42
34	22	4	10	16	100	40
36	13	5	4	25	16	20
33	14	2·5	5	6·25	25	12·5
31	11	1	2	1	4	2
56	16	10	7	100	49	70
54	10	9	1	81	1	9
39	19	6	8	36	64	48
33	20	2·5	9	6·25	81	22·5
Sum:		$\Sigma X = 55$	$\Sigma X = 55$	$\Sigma X^2 = 384·5$	$\Sigma Y^2 = 385$	$\Sigma XY = 290$

Formula

$$r_s = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right]}}$$

$\Sigma XY = 290$
$\Sigma X = 55$
$\Sigma Y = 55$
$\Sigma X^2 = 384·5$
$\Sigma Y^2 = 385$
$n = 10$

$$r_s = \frac{290 - \frac{(55)(55)}{n}}{\sqrt{\left[384.5 - \frac{(55)^2}{n}\right]\left[385 - \frac{(55)^2}{10}\right]}}$$

$$= \frac{290 - \frac{3025}{10}}{\sqrt{\left[384.5 - 302.5\right]\left[385 - 302.5\right]}}$$

$$= \frac{290 - 302.5}{\sqrt{[82]\,[82.5]}}$$

$$= \frac{12.5}{\sqrt{6765}}$$

$$= \frac{12.5}{82.2}$$

$$= 0.15$$

$$r_s = 0.15$$

Q.6 With the help of t test find if significant difference exists between the scores obtained on Emotional Intelligence by male and female teachers.

	Scores on Emotional Intelligence Scale
Male teachers	45, 32, 25, 57, 36, 42, 35, 55, 66, 65, 30, 35, 22, 27, 26
Female teachers	36, 53, 64, 55, 52, 34, 62, 73, 61, 34, 45, 38, 36, 25, 45

Ans. SOLUTION :

First Sample: (X_1)

MALE TEACHERS (X_1)

X_1	M_1	$(X_1 - M_1)$ x_1	x_1^2
45	39.86	5.14	26.41
32	39.86	-7.86	61.77
25	39.86	-14.86	220.81
57	39.86	17.14	293.77
36	39.86	-3.86	14.89
42	39.86	2.14	4.57
35	39.86	-4.86	23.61
55	39.86	15.14	229.21
66	39.86	26.14	683.29
65	39.86	25.14	632.01
30	39.86	-9.86	97.21
35	39.86	-4.86	23.61
22	39.86	-17.86	318.97
27	39.86	-12.86	165.37
26	39.86	-13.86	192.09

$\Sigma X_1 = 598$ $\Sigma x_1^2 = 2987.59$

Second Sample : (X_2)

FEMALE TEACHERS (X_2)

X_2	M_2	x_2	x_2^2
36	47.53	-11.53	132.94
53	47.53	5.47	29.92
64	47.53	16.47	271.26
55	47.53	7.47	55.80
52	47.53	4.47	19.98
34	47.53	-13.53	183.06
62	47.53	14.47	209.38
73	47.53	25.47	648.72
61	47.53	13.47	181.44
34	47.53	-13.53	183.06
45	47.53	-2.53	6.40
38	47.53	-9.53	90.82
36	47.53	-11.53	132.94
25	47.53	-22.53	507.60
45	47.53	-2.53	6.40

$\Sigma X_2 = 713$ $\Sigma x_2^2 = 2659.7$

$\Sigma X_1 = 598$ $\quad\quad$ $\Sigma X_2 = 713$

$N_1 = 15$ $\quad\quad$ $N_2 = 15$

$M_1 = \Sigma X_1 / N$ $\quad\quad$ $M_2 = \Sigma X_2 / N$

$M_1 = 598/15$ $\quad\quad$ $M_2 = 713/15$

$M_1 = 39.86$ $\quad\quad$ $M_2 = 47.53$

$\Sigma X_1^2 = 2987.59$ $\quad\quad$ $\Sigma X_2^2 = 2659.72$

Pooled SD or σ Formula:

$$= \sqrt{\frac{\Sigma X_1^2 + \Sigma X_2^2}{(N_1 - 1) + (N_2 - 1)}}$$

$$SD \text{ or } \sigma = \sqrt{\frac{2987.59 + 2659.72}{(15-1) + (15-1)}}$$

$$SD \text{ or } \sigma = \sqrt{\frac{5647.31}{28}}$$

$$SD \text{ or } \sigma = \sqrt{201.689}$$

$$SD \text{ or } \sigma = 14.201$$

SED or σ_D Formula:

$$= \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

$$= 14.20 \sqrt{\frac{1}{15} + \frac{1}{15}}$$

$$SE_D \text{ or } \sigma_D = 14.201\sqrt{\frac{2}{15}}$$

$$SE_D \text{ or } \sigma_D = 14.201\sqrt{0.133}$$

$$SE_D \text{ or } \sigma_D = 14.201 \times 0.364$$

$$SE_D \text{ or } \sigma_D = 5.169$$

t (value) Formula :

$$t = \frac{M_1 - M_2}{\sigma_D}$$

$$t = \frac{39.86 - 47.53}{5.169}$$

$$t = \frac{-7.67}{5.169}$$

$$t = -1.48$$

df = Formula :- $N_1 + N_2 - 2$ OR $(N_1 - 1) + (N_2 - 1)$

$$df = 15 + 15 - 2$$

$$= 28$$

In 't' distribution table then we find critical values of t with degree of freedom (df) 28 i.e. at .05 level is 2.05.

Since calculated value of 't' is less than critical value hence it can be said that there exist No significant difference between the scores. Hence Null hypothesis is accepted.

2.1 Describe Point-biserial correlation and Phi coefficient.

Ans. **POINT BISERIAL CORRELATION (r_{PB})**

INTRODUCTION

In educational or psychological studies, we often come across situations where both the variables correlated are continuously measurable, while one of them is artificially reduced to Dichotomy. In such a situation, when we try to compute correlation between a continuous variable and a variable reduced to artificial dichotomy, we always compute the coefficient of biserial correlation.

The Dichotomous variable is the one that can be divided into two sharply distinguished or mutually exclusive categories. This reduction into two categories may be the consequence of the nature of the data obtained. For example, in study to find out whether or not a student passes or fails a certain standard, we place the crucial point dividing pass and fail student anywhere we please. Hence measurement in the variable is reduced to two categories (pass and fail). This reduction into two categories, however, is not natural as we can have the crucial or dividing point according to our convenience.

Such a reduction of the variable into two artificial categories (artificial Dichotomy) may be seen in the following classification:

- Socially adjusted and socially maladjusted
- Athletic and non-athletic
- Radical and conservative

- Poor and not poor
- Social minded and mechanical minded
- Drop-outs and Stay-ins
- Successful and unsuccessful
- Moral and immoral

If we try to analyze the nature of distributions involving these dichotomized variables i.e. adjustment in the topmost classification, we can come to the conclusion that artificial dichotomy is based on a clear assumption that the variable underlying the dichotomy should be continous and normal.

In the two-fold division of socially adjusted and socially maladjusted, the division is quite artificial. If sufficient data were available, we could have found the trait 'adjustment' normally distributed among the studied population and it could have been distributed equally, instead of being discrete or limited (restricted to two-fold division).

In conclusion, we may term a dichotomy an artificial dichotomy, when we do not have any clear-cut crucial point or criteria for such a division. We fix the dividing point according to our own convenience.

In case sufficient data were available, the continuity as well as the normality of the distribution involving this variable can be easily established. Hence the basic assumption in using Biserial correlation as an estimate of the relationship between a continuous variable and a dichotomous variable

is that the variable underlying the dichotomy is continuous and normal. This implies that it should be an artificial dichotomous variable rather than a natural dichotomous variable.

In contrast to artificial dichotomy, the variable can be reduced to two categories - Natural or Genuine Dichotomy. Here we do not apply an artificial crucial point for the division as we do in artificial dichotomy. The examples of such a division of the related variables into natural categories are:

- Scored as 1 and scored as 0
- Right and Wrong
- Male and female
- Owning a home and not owning a home
- Living in Delhi and not living in Delhi
- Being alcho alcoholic and Non alcholic

In these categories, the division of the relevant variable into two categories is quite clear. In such cases, even if sufficient data were available, we could not have more than two categories. The answer scored as one and zero or right and wrong, cannot have more than two categories.

Thus, before deciding the type of measure of correlation between a continuous variable and a variable reduced to dichotomy, we should first try to find out what type of dichotomy - artificial or natural is involved in the categorization of the second variable. If it is artificial, then we should try to compute the coefficient of biserial correlation (r_{bis}). But if there is a genuine or a natural dichotomy, then we should try to compute coefficient of Point Bise

($r_{p,bis}$) instead of r_{bis}.

THE POINT BISERIAL CORRELATION (r_{pB})

We resorted to the computation of Point biserial correlation coefficient ($r_{p.bis}$) for estimating the relationship between two variables when one variable is in a continuous state and the other is in the state of a natural or genuine dichotomy. If we are sure that the dichotomized variable does not belong to the category of artificial dichotomy, then we should try to compute Point Biserial Correlation Coefficient ($r_{p.bis}$)

Computation of Point Biserial Correlation Coefficient (r_{pB})

FORMULA: $r_{p,bis} = \frac{M_p - M_q}{\sigma_t}\sqrt{pq}$ (The general formula)

Alternative formula for this: $r_{p,bis}$

$$r_{p,bis} = \frac{M_p - M_t}{\sigma_t}\sqrt{p/q}$$

Where: **p** = Propotion of cases in one of the categories (higher group) of dichotomous variable

q = Poropotion of cases in the lower group = $1-p$

M_p = Mean of the higher group, the first category of the dichotomous variable

M_q = Mean of the values of lower group

M_t = Mean of the entire group

σ_t = Standard deviation (SD) of the entire group

THE PHI COEFFICIENT (ϕ)

Where we have to compute correlation between two such variables which are genuinely dichotomous, it is the ϕ coefficient that is computed. Generally, its computation may involve the following situations:

- When the classification of the variables into two categories is entirely and truly discrete, we are not allowed to have more than two categories i.e. living vs. dead, employed vs. not employed etc.

- When we have test items which are scored as pass-fail, True-False, or opinion and attitude responses, which are available in the form of yes-no, like-dislike etc. no other intermediate type of responses is allowed

- With such dichotomized variables which may be continuous and may even be normally distributed, but are treated in practical spre operation or if they were genuine dichotomics, e.g. test items that are scored as either right or wrong, 1 and 0 and the like.

Features and characteristics of ϕ coefficient

- The phi coefficient is used for measuring the correlation between two variables when both are expressed in the form of genuine or natural dichotomies.
- The phi coefficient has the same relation with tetrachoric correlation (r_{t}) as point biserial ($r_{p.bis}$) has with the biserial coefficient (r_{bis}).

- It can be checked against pearson 'r' obtained from the same table.
- It is most useful in item analysis when we want to know the item to item correlation.
- It bears a relationship with χ^2 to be expressed as $\chi^2 = N\phi^2$
- The values of phi coefficient range between -1 and $+1$, but these are influenced by marginal totals.
- It makes no assumptions regarding the form of distribution in dichotomized variables like r_t which needs the assumptions of large N and continuity and normality of the distributions.
- Standard error of ϕ can be easily computed and ϕ can be easily tested against the null hypothesis by means of its relationship to χ^2.
- When there is any doubt regarding the exact nature of the dichotomized variable, it is always safe to compute ϕ. Also, its computation is much easier and regarded as a better and a more dependable statistics than r_t.

CONCLUSION

For dichotomous variable point biserial correlation is used whereas when variables are genuine or natural dichotomous.

Q.8 Compute Chi-Square for the following data :

Job Position	Work Motivation Scores		
	High	Average	Low
Junior Managers	10	15	15
Senior Managers	10	10	10

Ans. CALCULATION OF chi-Square AFOERSAID DATA :

Job Position	High	Average	Low	Total
Junior Managers	10	15	15	40
Senior Managers	10	10	10	30
Total	20	25	25	70

Computation of Expected Frequencies (fe / E)

$\frac{40 \times 20}{70} = 11.428$

$\frac{40 \times 25}{70} = 14.285$

$\frac{40 \times 25}{70} = 14.285$

$\frac{30 \times 20}{70} = 8.571$

$\frac{30 \times 25}{70} = 10.714$

$\frac{30 \times 25}{70} = 10.714$

Job Position Observed Frequencies Expected Frequencies

Job Position		High	Average	Low
Junior Managers	OF	10	15	15
	EF	11.428	14.285	14.285
Senior Managers	OF	8.571	10	10
	EF	8.571	10.714	10.714

[OF - Observed Frequencies
EF - Expected Frequencies]

Computation of value of χ^2 (chi - Square)

Formula:

Chi - Square = $\Sigma [(O-E)^2 / E]$ OR $\chi^2 = \Sigma [(O-E)^2 / E]$

O	E	O-E	$(O-E)^2$	$(O-E)^2 / E$
10	11.428	-1.428	2.039	0.178
15	14.285	0.715	0.511	0.035
15	14.285	0.715	0.511	0.035
10	8.571	1.429	2.082	0.242
10	10.714	-0.714	0.509	0.047
10	10.714	-0.714	0.509	0.047
Total - 70	69.997			0.584

Chi - Square = 0.584.

SECTION - C

Q.9 The sign test

Ans. THE SIGN TEST

The sign test is the simplest test of significance in the category of non-parametric tests. It makes use of plus and minus signs rather than quantitative measures as its data. It is particularly useful in situations in which quantitative measurement is impossible or inconvenient, but on the basis of superior or inferior performance it is possible to rank with respect to each other the two members of each pair.

The sign test is used either in the case of single sample from which observations are obtained under two experimental conditions. The researcher wants to establish that the two conditions are different.

The use of this test does not make any assumption about the form of the distribution of differences. The only assumption underlying this test is that the variable under investigation has a continuous distribution.

Q.10 Point Estimation

Ans. **POINT ESTIMATION**

We have Indicated that $\bar{x}$ obtained from a sample is an unbiased and consistent estimator of the population mean (μ). Thus, if an Investigator obtains Adjustment score from 100 students and wanted to estimate the value of (μ) for the popula-tion from which these scores were selected, researcher would use the value of $\bar{x}$ as an estimate of population mean (μ). If the obtained value of $\bar{x}$ were 45.0 then this value would be used as estimate of population mean(μ).

This form of estimate of population parameters from sample statistic is called point estimation. point estimation is estimating the value of a parameter as a single point, for example, population mean (μ) = 45.0 from the value of the statistic $\bar{x}$ = 45.0.

Q. 11 Decision Errors

Ans

DECISION ERRORS

The crucial topic for making sense of statistical significance is the kind of errors that are possible in the hypothesis-testing process. The kind of errors we consider here are about how, in spite of doing all your figuring correctly, your conclusions from hypothesis-testing can still be incorrect. It is not about making mistakes in calculations or even about using the wrong procedures. That is, decision errors are situations in which the right procedures lead to the wrong decisions.

Decision errors are possible in hypothesis testing because you are making decisions about populations based on information in samples. The whole hypothesis testing process is based on probabilities. The hypothesis-testing process is set up to make the prop probability of decision errors as small as possible. For example we only decide to reject the null hypothesis if a sample's mean is so extreme that there is a very small probability (say, less than 5%.) that we could have gotten such an extreme sample if the null hypothesis is true. But a very small probability is not the same as a zero probability! Thus in spite of your best intention, decision errors are always a ll errors.

Q.12 Direction of correlation

Ans. DIRECTION OF CORRELATION

The direction of the relationship is an important aspect of the description of relationship. If the two variables are correlated then the relationship is either positive or negative. The absence of relationship indicates "zero correlation". Let's look at the positive, negative and zero correlation.

Positive Correlation

The positive correlation indicates that as the values of one variable increases the values of other variable also increase. Consequently, as the values of one variable decreases, the values of other variable also decrease. This means that both the variables move in the same direction. For ex. As the intelligence (IC increases the marks obtained increases.

Negative correlation

The Negative correlation indicates that as the values of one variable increases, the values of the other variable decrease. This means that two variables move in the opposite direction. For example As the intelligence (IQ) increases the errors on reasoning task decreases.

Zero Correlation / No correlation

Sometimes also possible that there is no relationship between X and Y. When they do not share any relationship then the direction of the correlation is neither positive nor negative and it is called zero or NO correlation.

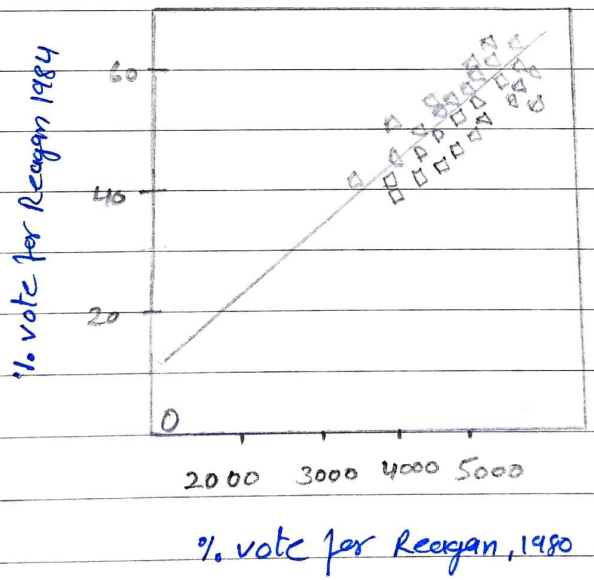
Q.13 Linear Regression

Ans. LINEAR REGRESSION

Regression goes one step beyond correlation in identifying the relationship between two variables. It creates an equation so that values can be predicted within the range framed by the data. That is if you know X you can predict Y and if you know Y you can predict X. This is done by an equation called regression equation.

When we have a scatter plot you have learnt that the correlation between X and Y are scattered in the graph and we can draw a straight line covering the entire data. This line is called the regression line.

Here is the line and the regression equation superimposed on the scatter plot: Linear Regression



From this line, you can predict X from Y that is % votes in 1984 if kno... You can find out the % of vote in 198... Similary if you know % of votes in 198... you can know % of votes in 1984.

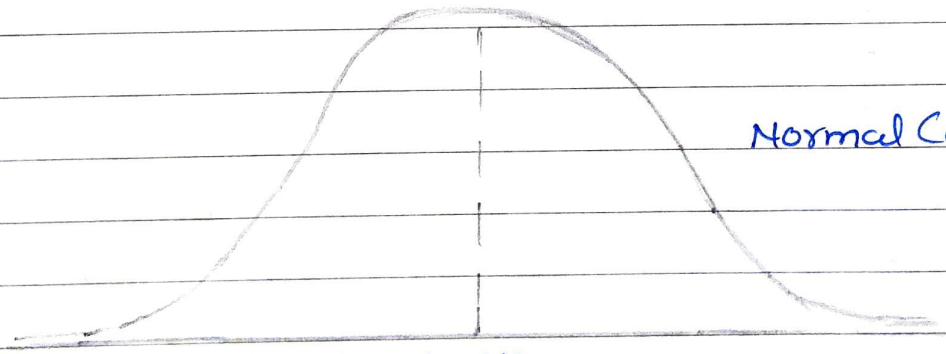
Q.14 Normal curve

Ans. NORMAL CURVE

This Bell shaped curve technically known as Normal Probability Curve or Simply Normal curve and the corresponding frequency distribution of sour scores, having just the same values of all three measures of Central tendency (Mean, Median and Mode) is known as Normal Distribution.

The Normal distribution is a continous probability distribution that is ~~syst~~ symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side.

The area under the normal distribution curve represents probability and the total area under the curve sums to one.



Q.15 Sampling and standard errors

Ans. The statistics error is of two types: Sampling Error and Standard Error of statistics.

SAMPLING ERROR

Sampling error refers to the difference between the mean of the entire population and the mean obtained of the sample taken from the population.

Thus sampling Error = $M_{pop} \sim M$ or $\overline{M} - M$

As the difference is low the mean obtained on the basis of sample is near to the population mean and sample mean is considered to be representing the population mean (or M_{opp})

STANDARD ERRORS

The standard error is nothing but the intra differences in the sample measurements of number of samples taken from a single population.

Q. 16 Scatter Plot

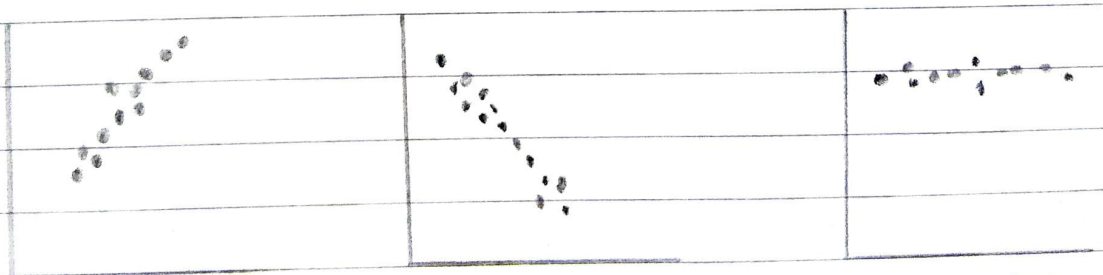
Ans. SCATTER PLOT

A scatter plot is a type of plot or mathematical diagram using cartesian coordinates to display values for typically two variables for a set of data. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

The first step is creating a scatter plot of the data. "There is no excuse for failing to plot and look. In general, scatter plots may reveal a :

- Positive correlation (high values of X associated with high values of Y)
- Negative correlation (high value of X associated with low values of Y)
- No correlation (values of x are not at all predictive of values of Y).

These patterns are demonstrated in the figure below



A) Positive correlation B) Negative Correlation C) No correlation

Q.17 Goodness of fit

Ans. GOODNESS OF FIT

Goodness of fit is one function of Chi Square. It describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing eg. to test for normality of residuals, to test whether two samples are drawn from identical distributions, or whether outcome frequencies follow a specified distribution.

A common use is to assess whether a measured/observed set of measures follows an expected pattern. The expected frequency may be determined from prior knowledge or by calculation of an average from the given data.

The null hypothesis, Ho is that the two sets of measures are not significantly different.