ISSN No:-2456-2165

The Application of Inferential Statistics in Testing Research Hypotheses in Library and Information Science

SANUSI MAMMAN (1), MADUAGWU, K. LILIAN (2) &
OMORODION, E. EMMANUEL (3)
Department of Lib. & Infor. Science, Waziri Umaru Federal Polytechnic, Birnin Kebbi
OLULEYE HELEN REMILEKUN (4)
College Library
Kogi State College of Education (technical), Kabba

Abstract:- There is growing concern on different type of analyzing research hypothesis that is carried out in whole spectrum of knowledge and many of the researchers fail to understand the right tool for testing the hypothesis and this leads to poor data analysis and wrong inferences. Testing hypothesis in research is the basic ingredient of all research activities and if done correctly, suitable inferences were made which lead to concrete and accurate findings that can be used for products or services development or improvement, or policy formulation. Hence, this paper discusses the different types of inferential statistics which include both parametric and non-parametric and also the condition of applying any type of inferential statistics by researchers. In the process, the paper divides parametric and non-parametric inferential statistics in to relational and differential. The paper also outlines the conditions for using each type inferential statistics and procedures for making inferences. The paper concludes with an emphasis for librarians and other researchers in the body of knowledge to up skill and retools themselves in order to enhance their skills, knowledge of research and attain better strategies on how to advance the frontiers of knowledge, working environments and other sectors in a 21st century.

Keywords:- *Application, Inferential Statistics, Testing Hypothesis, Correlation, Regression, Regression Analysis*

I. INTRODUCTION

Inferential Statistics: is the process of making inference or drawing conclusion and/or generalization about certain characteristics considering the nature or appearance of the sample. It is basically concerned in decision making or inference. It used mostly in making judgment based on the research hypothesis formulated in the research study. The hypothesis should be in relation or difference formed.

In short, inferential statistics involves:
- Drawn conclusion
- Making generalization
- Taking decision

Inferential statistics is basically divided in to 2:
1. Parametric Inferential Statistics
2. Non parametric Inferential Statistics

1. Parametric Inferential Statistics: These are powerful techniques used to detect the relationship or differences between two-or-more groups when the data is normally distributed. It can be bi-viarite and multi-viariate analysis. Bi-viarite is dealing with two variables of analysis and multi-viarite is dealing with more than two variables of analysis

The researcher may use parametric when the data to be analyze meet the following conditions:
1. The variable measured should be normally distributed
2. Homogeneity of the variables for some of the parametric test of analysis

Types of Parametric Inferential Statistics
1. Relational Parametric Inferential Statistics
2. Differential Parametric Inferential Statistics

1. Relational Parametric Inferential Statistics is called
Correlation is the study of relationship between two or more variables. Co-efficient of correlation can be used to compare two variables i.e. measure linear relationship between 2 variables X and Y. A numerical descriptive measure of correlation is provided by Pearson product-moment and Spearman's ranking coefficients of correlation [2]. The relationship is always within $-1 \leq X \leq 1$ (signifies X is greater or equal to -1, less than or equal to 1).

The correlation parametric statistics involves:
i. Pearson Product Moment Correlation (PPMC)
ii. Regression Analysis
In showing the relationship between or among the variables or data set in the research study can either be direct relationship or inverse relationship
- Direct Relationships: Signifies positive relationship such as +1, +0.7, +0.2
- Inverse Relationship: Signifies negative relationship such as -1, -0.7, -0.2

i. Pearson Product Moment Correlation (PPMC):
The value of coefficient (r) is interpreted as follows:-
When r = 0, it signifies that there is no relationship between the variables
When r = negative value, it means that the relationship between the variables is inverse.

When r = positive value, it means there is a direct relationship between the variables.
When r = 1 then it means there is a perfect positive relationship between the variables.
r is used to measure the strength of association between two variables and ranges between -1 (perfect negative correlation) to 1 (perfect positive correlation). Cohen has the following interpretation of the absolute value of the correlation:

Table 1:

Different Type of Relational Result & its Level		
Positive Correlation coefficient value	**Negative Correlation coefficient value**	**Association**
+0.2	-0.2	Mild or weak
0.2 to 0.5	-0.5 to -0.2	Moderate
0.5 to 0.9	-0.9 to -0.5	Strong
0.9 to 1.0	-1.0 to -0.9	Very strong or perfect

[1] The formula for calculating coefficient of correlation under Pearson product moment method the formula is:-

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N(\sum X^2) - (\sum X)^2][N(\sum Y^2)(\sum Y)]^2}}$$

Example 1

Find the relationship between use of information resources in reference section and reserve section in UMYUK library and the following data were recorded

Table 2:

Relationship Between Use of Information Resources in Reference Section and Reserve Section in UMYUK Library		
S/n	**Reference section**	**Reserve section**
1	10	11
2	21	12
3	25	12
4	34	15
5	40	22
6	42	25
7	53	20
8	52	20
9	60	32
10	63	31

$H_{0:}$ There is no significant relationship between use of information resources in reference section and reserve section in UMYUK library

Table 3:

Relationship on the Use of Information Resources Between Reference and Reserve Section					
S/n	**Reference section (X)**	**Reserve section (Y)**	X^2	Y^2	XY
1	10	11	100	121	110
2	21	12	441	144	252
3	25	12	625	144	300
4	34	15	1156	225	510
5	40	22	1600	484	880
6	42	25	1764	625	1050
7	53	20	2809	400	1060
8	52	20	2704	400	1040
9	60	32	3600	1024	1920
10	63	31	3969	961	1953
	∑X= 400	∑Y= 200	18768	4528	9075

N = number of frequency or number of students/respondents which 10

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N(\sum X^2) - (\sum X)^2][N(\sum Y^2)(\sum Y)]^2}}$$

$$r = \frac{(10)(9075) - (400)(200)}{\sqrt{(10)(18768) - (400)^2 (10)(4528)(200)^2}}$$

$$r = \frac{90750 - 80000}{\sqrt{(187680 - 160000)(45280 - 40000)}}$$

$$r = \frac{10750}{\sqrt{(27680)(5290)}}$$

$$r = \frac{10750}{\sqrt{146427400}}$$

$$r = \frac{10750}{12089.26797} = 0.889$$

Therefore, r value calculated is equals to +0.89

So there is a very high positive relationship between the use of information resources in reference section and reserve section in UMYUK library, therefore the null hypothesis formulated is rejected

ii. Regression: is a statistical method used to describe the nature of the relationship between variables, that is, positive or negative, linear or nonlinear [3]

Regression analysis is a statistical process for estimating the relationships among variables

Key assumptions for multiple regressions are similar to those for a simple regression:
• Variables are normally distributed.
• The standard deviations of dependent variables are the same for each value of the independent variable.
• There is a linear relationship between the outcome variable and the independent variables.
• The independent variables are not highly correlated with each other.

Regression Analysis can be in two ways:
➢ Linear Regression
➢ Multiple Regression

2. Differential Parametric Inferential Statistics: This is also another type of parametric statistic that is used in detecting the difference between two-or-more groups (i.e. difference between and/or among the selected sample). It involves z-test, t-test, ANOVA and ANCOVA.

Though, there are other types of differential parametric analyses that are not common, such include MANOVA, MANCOVA

i. T- Test

T-test statistically measures the difference between two groups' means, most likely reflects a real difference in the population from which the groups were sampled. One sample t-test, independent samples t-test, and dependent samples t-test are all parametric tests used at the bivariate level and all compare means between two groups [2]. **T**-test is used to measure the difference between two groups which can either one group with one (currently) and the previous one, two related or paired groups or two independent groups. It is bi-viarite analysis of differences [3].

In calculating t-test, it required the following values in making inferences, conclusion and/or decision making.

- **T- critical:** Signifies the number obtained from the table considering the degree of freedom and significant level
- **Degree of freedom (df)** which is mostly based on the types of analysis used by in researcher in making inferences.
- **Significant level: Signifies** 0.05 and 0.01 levels 0.05significance level means 95/100 times that you sample perfectly represents from the population. It is the most commonly used in scientific research.

Types of T-test

a. Independent T-test (Unrelated Sample): This type of test is used to detect the difference between two independents groups. For example, male or female. It is the difference observed between two means (average) to test its significant. Formula for calculating independent-samples T-test is:

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1^2 + SS_2^2}{N_1 + N_2 - 2}\right]\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Where: $\bar{X}_1$ = Mean of the first variable $\bar{X}_2$ = Mean of the second variable N_1 = Number of first variable N_2 = Number of second variable S_1^2 = Variance of the first sample S_2^2 = Variance of the second sample

Where: $SS_1^2 = EX^2{}_1 - (EX_1/N_1)^2$

$SS_2^2 = EX_2^2 - (EX_2/N_2)^{2`}$

For example, utilization of information resources in the library of two different groups were scored and the following data were recorded.

Table 4:

Utilization of Information Resources of two Different Groups in the Library										
Gender	**1**	**1**	**2**	**2**	**1**	**2**	**2**	**1**	**1**	**2**
Score	**3**	**4**	**2**	**3**	**5**	**3**	**4**	**6**	**7**	**3**

$\bar{X}_1$

Table 5:

Result on the Utilization of Information Resources of two Different Groups in the Library			
X_1	X_2	$X^2{}_1$	X_2^2
3	2	9	4
4	3	16	9
5	3	25	9
6	4	36	16
7	3	49	9
EX_1= 25	EX_2 =15	E $X^2{}_1$ = 135	E $X^2{}_2$ = 47

Let Male to be X_2 and female to be X_1

= EX_1/N_1 = 25/5 =5

$\bar{X}_2$ = EX_2/N_2 15/5 =3

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1^2 + SS_2^2}{N_1 + N_2 - 2}\right]\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Where $SS_1^2 = EX^2{}_1 - (EX_1/N_1)^2$

SS_1^2 = 135- $(25/5)^2$

= 135 – $(5)^2$

=135-25 =110

$SS_2^2 = EX_2^2 - (EX_2/N_2)^{2`}$

= 47- $(15/5)^2$

= 47- $(3)^2$

= 47-9 = 38

Substitute from the formulae

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1^2 + S\,S_2^2}{N_1 + N_2 - 2}\right]\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

=

$$\frac{5-3}{\sqrt{\left[\frac{110+38}{5+5-2}\right]\left(\frac{1}{5}+\frac{1}{5}\right)}}$$

=

$$= \frac{2}{\sqrt{\left[\frac{148}{8}\right]\left(\frac{2}{5}\right)}}$$

$$= \frac{2}{\sqrt{[18.5](0.4)}}$$

$$\frac{2}{2{,}72} = 0.735$$

level of significance, **t** critical = **2.896**

t_{cal} = **0.735**, **t** critical is found using **degree of freedom d/f=** $N_1 + N_2 - 2$

= **5+5−2=8**

Therefore, t_{cal} = **0.735 is less than $t_{critical}$ = 2.896** and the hypothesis is retained thereby saying there is no significant difference in the use of information resources by male and female in UMYUK library**.**

Note: In all the Parametric and Non-Parametric Tests, Null Hypothesis is retained when t_{cal} **is less than $t_{critical}$** except for U-test (Mann Whitney U- Test) where the reverse is the case With advent Statistical Software for Data analysis such as SPSS and among others, the researcher may decide to compare and contrast with t_{cal} and $t_{critical}$ and follow the above instruction or to compare probability and level of significance for example P-value vs level of significance in either retaining or rejecting the null hypothesis formulated by the researcher.

This is to say that; hypothesis can as well be tested using a calculated probability value which is compared with alpha value (assumed level of significance).

In this case, if P value is greater than α value (which is assumed level of significance like 0.05); the null hypothesis is retained or accepted otherwise is rejected.

Example 2

Consider the scores below rating the use public and national libraries in Katsina metropolis.

Table 6:

Use Public and National Libraries in Katsina Metropolis.		
S/N	**LIBRARIES SCORES**	**SCORES**
1	1	30
2	2	35
3	2	18
4	1	36
5	2	28
6	2	24
7	2	29
8	2	19
9	1	27
10	2	25
11	1	39
12	2	27
13	2	24
14	1	40

Using T-test, answer this research question "Is there any significant difference between the use public and national libraries by the users in Katsina metropolis? (Use 2 for public libraries and 1 for national libraries). Assuming the significant level is 0.05.

b. Dependent T- test, Related T-test or Paired t- test: This is also another type of parametric test used in detecting the difference between related samples. It is measured in a continuous scale.

The rationale of using Dependent t- test (Related t- test in research)

- **Normally distribution of data**
- **Related groups on a particular sample**
- **Continuous scale**

Formula use in dependent T-test calculations

$$T = \frac{\bar{d}}{\frac{Sd}{\sqrt{N-1}}}$$

Where:

d = difference between paired score

$\bar{d}$= Mean of the difference of the two variables

Sd = Variance of the differences of the two variables

N = Number of samples

$$\textbf{Sd} = \sqrt{\frac{Ed2}{N} - \left(\frac{Ed}{N}\right)2}$$

The score of Test and Retest of the research instrument were collected in a particular survey

Table 7:

Test and Retest of the Instruments on the Use of Library										
Test	56	43	50	44	40	55	48	60	58	62
Retest	63	71	52	57	51	52	54	63	55	58

Using t-test, test the hypothesis "There is significant difference between test and retest of the research instrument. (Significant level at 0.05)

Table 8:

Result on the Test and Retest on the Use of Library			
Test	**Retest**	**D**	**d^2**
56	63	-7	49
43	71	-28	784
50	52	-2	4
44	57	-13	169
40	51	-11	121
55	52	3	9
48	54	-6	36
60	63	3	9
58	55	3	9
62	58	4	16
		Ed = -54	Ed^2 = 1206

$\bar{d}$ = -54/10 = -5.4

$$\textbf{Sd} = \sqrt{\frac{1206}{10} - \left(\frac{-54}{10}\right)2}$$

$$= \sqrt{120.6 - (-5.4)2}$$

$$= \sqrt{120.6 - 29.16} \quad = \sqrt{91.44} \quad = 9.56$$

Hence, $\frac{\frac{\bar{d}}{Sd}}{\sqrt{N-1}} = \frac{\frac{-5.4}{9.56}}{\sqrt{10-1}} = \frac{\frac{-5.4}{9.56}}{\sqrt{9}} = \frac{-5.4}{3.19} = -1.6928$

Degree of Freedom = N - 1 = 10 – 1 = 9 @ 0.05 confidence level

$t_{critical}$ using T- table = 1.833 Therefore, t_{cal} = -1.6928 **is less than $t_{critical}$** = 1.833 and the hypothesis is retained thereby saying significant difference exist between the test and retest of the research instrument

c. One sample t- test: In this type of t- test, the mean (average) is compared with another specified group of population. For example in academic environment one sample t- test is used in comparing the result of the student for example result of this year in Umaru Musa Yar'adua University, Katsina and the one of previous year.

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{N-1}}}$$

$$S = \sqrt{\frac{EX2}{N} - \left(\frac{EX}{N}\right)2}$$

For example, a researcher wishes to test whether time to complete the research is significantly change. Scores of (10) researchers were obtained and recorded as

If the time requires completing the research was known to be 4 days, test whether the completion time has significantly changed or not. @ 95% c confidence level

Table 9:

Time Change on the Research Completion										
Scores	**3**	**8**	**9**	**12**	**7**	**4**	**8**	**5**	**6**	**8**

Table 10:

Result on the Time Changes at the Research Completion	
X	**X^2**
3	**9**
8	**64**
9	**81**
12	**144**
7	**49**
4	**16**
8	**64**
5	**25**
6	**36**
8	**64**
EX = 70	**EX^2 = 552**

Therefore,

$$= \frac{\overline{X} - \mu}{\frac{S}{\sqrt{N-1}}}$$

$$S = \sqrt{\frac{EX2}{N} - \left(\frac{EX}{N}\right)2} = S = \sqrt{\frac{552}{10} - \left(\frac{70}{10}\right)2} = \sqrt{\frac{552}{10} - \left(\frac{70}{10}\right)2} = \sqrt{55.2 - (7)2} = \sqrt{55.2 - 49}$$

$$= \sqrt{6.2} = 2.449$$

$$= \frac{7-4}{\frac{2.449}{\sqrt{10-1}}} = \frac{3}{\frac{2.449}{\sqrt{9}}} = \frac{3}{0.8} = 3.75$$

therefore, t_{cal} = 3.75 and t_{crit} = 2.262

df = N – 1 = 10 – 1 = 9 and 95% confidence level involves 0.05 level of significance

t_{cal} 3.75 is greater than t_{crit} 2.262, therefore the hypothesis is rejected by saying that the time of completion the research has not change

ii. ANOVA

Analysis of Variation (or Variance): This is also another type of parametric statistic used in detecting the difference among three groups. ANOVA is an extension to t-test that the researchers used to compare the differences that exist between more than two means (averages) obtained from a defined population. It is also called as F-ratio or F-value. It involves Variation within a sample and Variation between samples.

Assumption of using ANOVA in Data Analysis

- Normal distribution of Data
- Homogeneity of variables (X_2)
- More than two groups

Types of ANOVA

a. One-way b. Two- way

a. One-way ANOVA: It is used to detect the difference in means of 3 or more independent groups. It can be thought of as an extension of the t-test for 3 or more independent groups. ANOVA uses the ratio of the between group variance to the within group variance to decide whether there are statistically significant differences between the groups or not.

$$SS_T = SS_{between} + SS_{within\ groups}$$

$$SS_T = EX^2 - \frac{(EX)_2}{N}$$

$$SS_b = \frac{(EX)_2}{N1} + \frac{(EX2)_2}{N2} + \frac{(EX3)_2}{N3} -- \frac{(EX)_2}{N}$$

Where EX = $EX_1 + EX_2 + EX_3$

b. Two-way ANOVA:

Note: For the entire above parametric test can be done using Statistical software like SPSS (Software Package for Social Science).

ISSN No:-2456-2165

Normality Test before Parametric Test

One of the important and notable assumptions for parametric tests is the data is the normally distribution. The normal distribution peaks in the middle and is symmetrical about the mean. Data does not need to be perfectly normally distributed for the tests to be reliable. Checking normality of the data can be done using the followings:

1. **Histogram:** When the drawing of the data is going like the below curve, then the data is normally distributed;

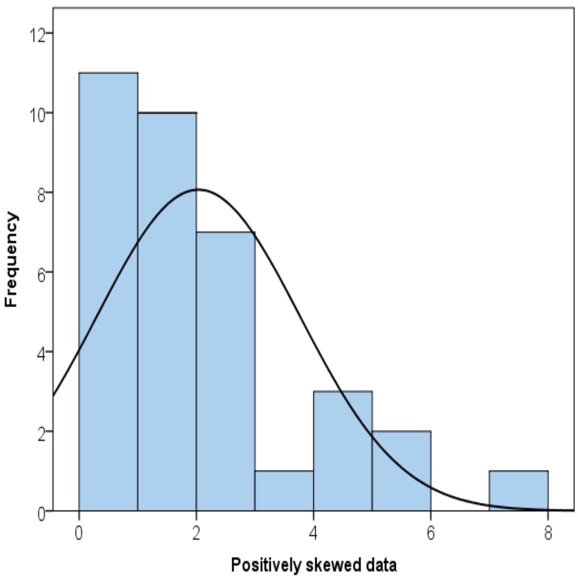


Fig 1: Sample of Normality Curve of the Distribution of Data using Histogram

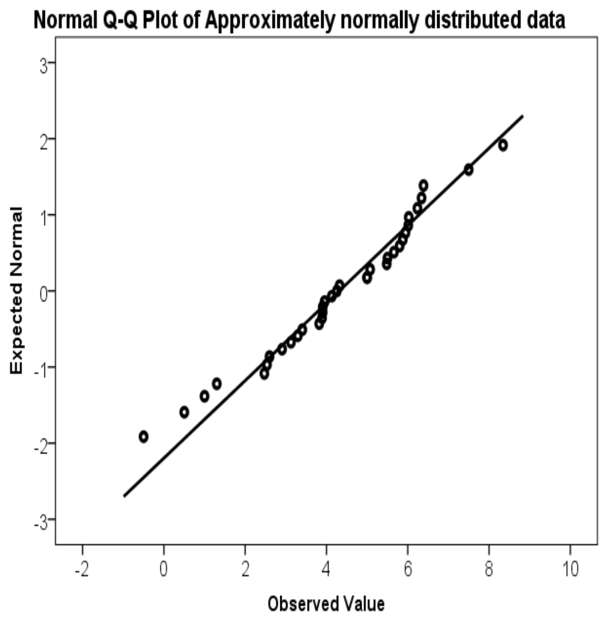
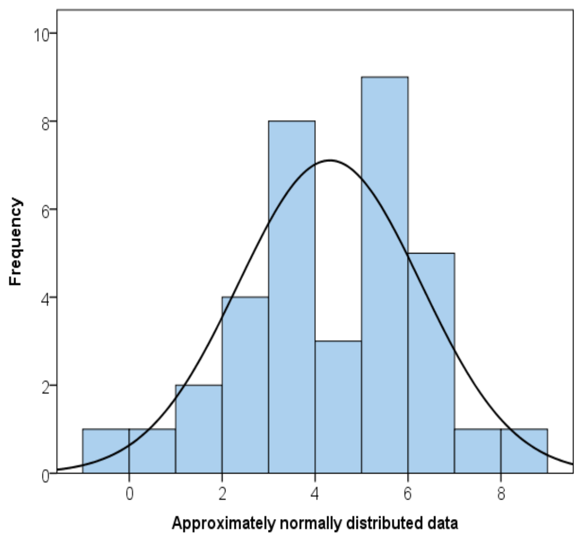


Fig 2: Sample of **Normality Curve of the Distribution of Data using Histogram**



Source [3]
Fig 3: Sample of Normality Curve of the Distribution of Data using Q-Q plots

2. Normal Q-Q (Quantile-Quantile) plot is an alternative graphical method of assessing normality to the histogram and is easier to use when there are small sample sizes. The scatter should lie as close to the line as possible with no obvious pattern coming away from the line for the data to be considered normally distributed. Below are the same examples of normally distributed and skewed data.

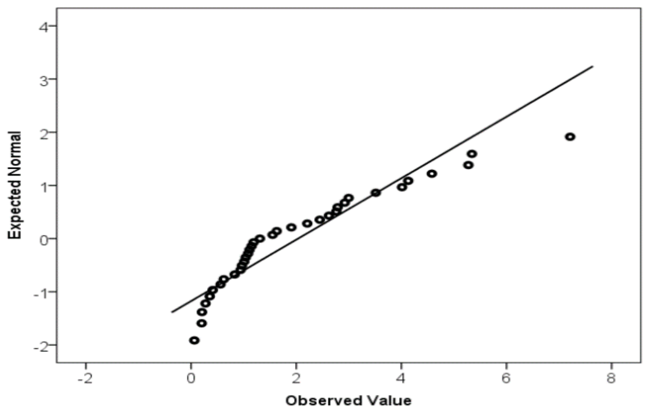


Fig 4: Sample of Normality Curve of the Dist
Source: (Samuel & Marshall, 2017)

2. Measures of **Skewness and Kurtosis** and their standard errors are also provided in the Explore output and can also be used in normality checking. In SPSS, the result which is called statistics is divided by standard error and compares the result with -1.96 to 1.96. When the result is within the above scale, the data is normally distributed but when it is otherwise, then the data is not normally distributed

4. Kolmogorov Sminorv and Shapiro Wilky Test: These are two different normality of two set of data. Kolmogorov Sminorv is used when the data is one hundred and above while Shapiro-wilky is used for the data that is less than one-hundred.

In all these tests the result of each variable is known as P-value which appears as (Sig. in SPSS) and is compared with level of significance. If p-value is less than level of significance (0.05), the data is set to be normally distributed.

II. NON PARAMETRIC INFERENTIAL STATISTICS

2. Non Parametric Inferential Statistics: Is another type of inferential statistics where researchers in different spectrum of knowledge used to make inferences or making effective and efficient decision making. This type of inferential is less power technique used in detecting the relationship or difference between two-or-more variables when the data failed to meet normality distribution [3]. It is important, for the researchers to know that; non-parametric statistic is used; when the data failed to meet the assumptions of parametric test especially ***Normality Distribution***.

Non Parametric Test is like parametric test where is divided in two:
1. Relational Non-Parametric Test
2. Differential Non-Parametric Test

1. Relational Non-Parametric Test: These are non-parametric tools used in ascertaining the relationship between two or more variables when the data failed to meet normality distribution. These tools are the substitutes of relational parametric test. This includes:

i. Spearman ranking order: Spearman ranking order is the substitutes of Pearson Product Moment Correlation (PPMC). It is only used, when the data is not normally distributed. The formula is as follows: -

$$r = 1 - \frac{6\sum d^2}{N(N^2-1)}$$

Where d is the difference between the rankings of the two variables.

N is the total number of observations in a given distribution

Example 1: Find the relationship between use of information resources in reference section and reserve section in UMYUK library and the following data were recorded

Table 11: Relationship on the Use of Information Resources in Reference & Reserve Sections using Spearman Ranking order

S/NO.	REFERENCE SECTION	RESERVE SECTION
1	81	77
2	65	52
3	86	70
4	48	40
5	79	66
6	75	60
7	58	63
8	61	50
9	51	43
10	78	63
11	45	41
12	88	80
13	74	68
14	63	55
15	65	58

H_0: There is no significant relationship between use of information resources in reference section and reserve section in UMYUK library

Assuming the above data is not normally distributed.

Solution
First, form a table with all elements required in the formula

Table 12: Result on the Relationship on the Use of Information Resources in Reference & Reserve Sections using Spearman Ranking order

S/N	Reference (X)	Reserve (Y)	Rank of X	Rank of Y	Difference (D)	D^2
1	81	77	3	2	1	1
2	67	52	8	11	-3	9
3	86	70	2	3	-1	1
4	48	40	14	15	-1	1
5	79	66	4	4	0	0
6	75	60	6	8	-2	4
7	58	63	12	6	6	36
8	61	50	11	12	-1	1
9	51	43	13	13	0	0
10	78	62	5	7	-2	4
11	45	41	15	14	1	1
12	88	80	1	1	0	0
13	74	68	7	5	2	4
14	63	55	10	10	0	0
15	65	58	9	9	0	0
						62

ISSN No:-2456-2165

$$r = 1 - \frac{6\sum \square^2}{\square(\square^2 - 1)}$$

$$r = 1 - \frac{6 \times 62}{15(15^2 - 1)}$$

$$r = 1 - \frac{372}{15(225 - 1)}$$

$$r = 1 - \frac{372}{15(224)}$$

$$r = 1 - \frac{372}{3360}$$

$$r = 1 - 0.1107$$

$$r = 0.8893$$

From the above, the answer shows that there is a high positive relationship between the two variables, so the null hypothesis is therefore rejected

ii. Chi square distribution is another type of relational non-parametric inferential statistics that is theoretical or mathematical distribution which has wide applicability in statistical work. Chi-square test is a particularly useful technique for testing whether observed data are representative of a particular distribution. It is widely used in Educational and Scientific research studies (Marshall & Samuel, 2017).

The formula of calculating Chi square test is

$$\chi = \sum \frac{(\square - \square)^2}{\square}$$

Type of Chi-square

One-variable chi square: It is used when one wants to analyze just one variable that is organized either as raw data or as a frequency table.

Two-variable chi square: It is used when one has two variables that are summarized in the two-way table.

Cross tabulation chi square: It is used when you have more than two variables that are organized either as raw data or as frequency data.

2. Differential Non-parametric Inferential Statistics: This is another type of non-parametric use to detect the differences between two-or-more variables when the data is not normally distributed. Differential Non-parametric inferential statistics involve the followings:

- **One-variable chi square**: It is used when one wants to analyze just one variable that is organized either as raw data or as a frequency table.
- **Two-variable chi square**: It is used when one has two variables that are summarized in the two-way table.
- **Cross tabulation chi square**: It is used when you have more than two variables that are organized either as raw data or as frequency data.

2. Differential Non-parametric Inferential Statistics: This is another type of non-parametric use to detect the differences between two-or-more variables when the data is not normally distributed. Differential Non-parametric inferential statistics involve the followings:

i. Mann-Whitney test (non-parametric equivalent to the independent t-test): It is used to compare whether two groups containing different people are the same or not. The Mann-Whitney test ranks all of the data and then compares the sum of the ranks for each group to determine whether the groups are the same or not. There are two types of Mann-Whitney U tests. If the distribution of scores for both groups has the same shape, the medians can be compared. If not, use the default test which compares the mean ranks (Marshall & Samuel, 2017).

The Mann-Whitney test is also known as the Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon–Mann–Whitney test.

ii. Wilcoxon Signed Rank test (non-parametric equivalent to the dependent or paired t-test): It is used to compare two related samples, matched samples or repeated measurements on a single sample to assess whether their population means ranks differ. It is a paired difference test and is the non-parametric alternative to the paired t-test. The absolute differences are ranked then the signs of the actual differences used to add the negative and positive ranks

iii. Kruskal-Wallis test (non-parametric equivalent to the one-way ANOVA): It compares the medians of two or more samples to determine if the samples have come from different populations. It is an extension of the Mann–Whitney U test to 3 or more groups. The distributions do not have to be normal and the variances do not have to be equal.

iv. Friedman test (non-parametric equivalent to the two-way ANOVA): Is non-parametric test used to detect differences in treatment across the multiple tests. It is substitute to repeated measures ANOVA.

In summary the tables below provide the short description of both parametric and nor nob-parametric inferential statistics:

Table 13:

Tests of Relationship or Association					
S/n	**Type of Relationship**	**Parametric test**	**Parametric Scale**	**Non parametric**	**Non-Para Scale**
1.	Relationship between 2 continuous variables	Pearson's Correlation Coefficient	Nominal scale	Spearman's Correlation Coefficient	Ordinal scale
2.	Predicting the value of one variable from the value of a predictor variable or looking for significant relationships	Regression Analysis (single or Multiple) depending the variables	Nominal Scale	Assessing the relationship between two categorical variables	Categorical scale

Table 14:

Test of Differences				
S/n	**Parametric Test**	**Scale for Parametric Test**	**Non Parametric Test**	**Scale for Non Parametric Test**
1	Independent t-test	Nominal Scale	Mann-Whitney test/ Wilcoxon rank sum	Ordinal Scale
2.	Paired/ Dependent-test/	Nominal Scale	Wilcoxon signed rank test	Ordinal Scale
	ANOVA (One-way)	Time/Condition variable	Kruskal-Wallis test	Ordinal scale and/or Time/ Condition variable
	ANOVA (Two-way)	Time/Condition variable	Friedman test	Ordinal scale and/or Time/ Condition variable

Note

Hypothesis testing is an **objective** method of making decisions or **inferences** from sample data (evidence). Sample data is used to choose between two choices i.e **hypotheses** or statements about a population. Typically, this is carried out by comparing what we have observed to what we expected if one of the statements (**Null Hypothesis**) was true.

Key terms:

- **Null Hypothesis (H_0)** is a statement about the population and sample data used to decide whether to reject that statement or not. Typically the statement is that there is no relationship or association for relational and there is no difference for differential hypothesis between or among variables.
- **Alternative Hypothesis (H_1)** is often the research question and varies depending on whether the test is one or two tailed. Typically, the statement is that there is relationship or association for relational and there is difference, for differential hypothesis between or among variables.
- **Significance Level (SL):** The probability of rejecting the null hypothesis when it is true, (also known as a type 1 error). This is decided by the individual but is normally set at 5% (0.05) which means that there is a 1 in 20 chance of rejecting the null hypothesis when it is true.
- **"Test Statistic** is a value calculated from a sample to decide whether to accept or reject the null (H0) and varies between tests. The test statistic compares differences between the samples or between observed and expected values when the null hypothesis is true."
- **P-value:** the probability of obtaining a test statistic at least as extreme as ours if the null is true and there really is no difference or association in the population of interest. P-values are calculated using different probability distributions depending on the test. A significant result is when the p-value is less than the chosen level of significance (usually 0.05).

III. CONCLUSION

The application of inferential statistic is necessary in any research study that involves research hypothesis. It has become a pre-requisite for making effective judgment. Selecting appropriate statistical is the key element of solving any experimental and hypothetical research study.

REFERENCES

[1]. Allan G. B., (2009). *Elementary Statistics: A Brief Version*, 7th edition, New York: McGraw-Hill.
[2]. [Olatoye, R. A., Olutola, T. A. & Suleiman, I. A. (2019). Understanding basic inferential statistics (correlation, chi-square, t-test and anova). Paper presented at Federal University Dustinma, Faculty of Education, Katsina State; Nigeria.
[3]. Marshall, E. & Samuel, P. (2017). Checking Normality for Parametric Tests: Statistical Hypothesis Testing worksheet and Normality Checking example solutions worksheet workshop 7: SPSS and Workshop 8: Parametric Testing. Retrieved from:https://www.researchgate.net/publication/3208 31642 (accessed on 8th September, 2021)