LECTURE NOTES
ON
PARAMETRIC & NON-PARAMETRIC TESTS
FOR
SOCIAL SCIENTISTS/ PARTICIPANTS
OF
RESEARCH METODOLOGY WORKSHOP
BBAU, LUCKNOW

By

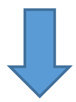
Dr. Rajeev Pandey
Professor in Statistics
University of Lucknow
Lucknow – 226 007

What is Statistics?

Statistics is neither really a science nor a branch of Mathematics. It is perhaps best considered as a meta-science (or meta-language) for dealing with data collection, analysis, and interpretation. As such its scope is enormous and it provides much guiding insight in many branches of science, business.

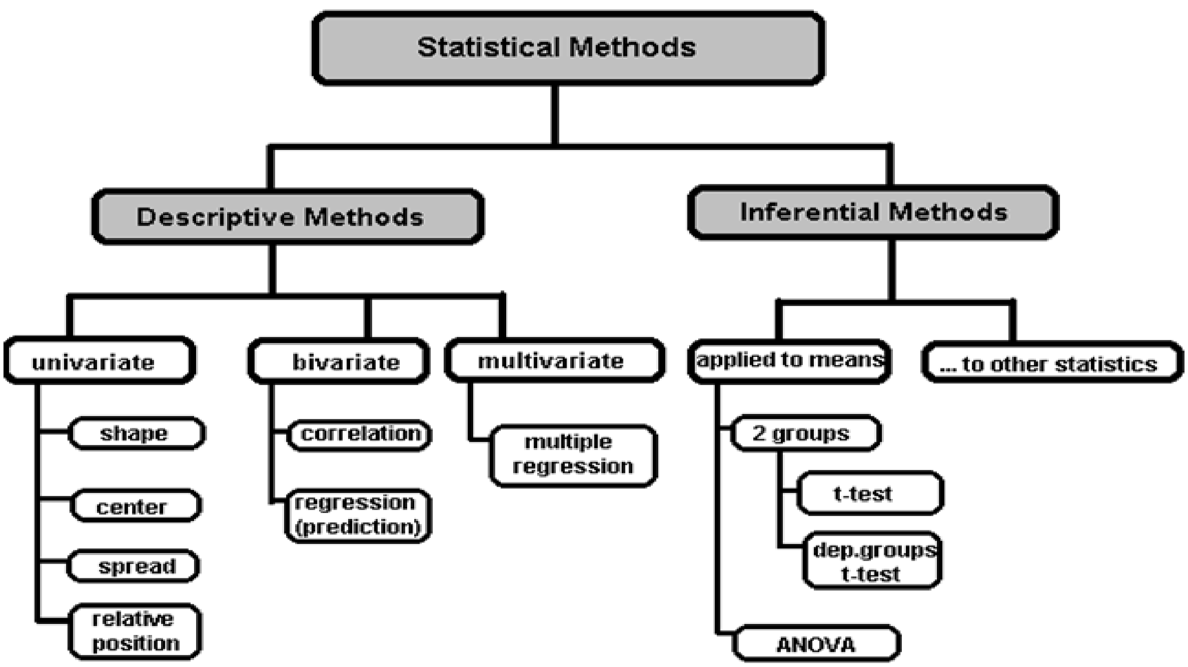
- *Statistics* is the science of collecting, organizing, analyzing, interpreting, and presenting data. (*Old Definition*).
- A *statistic* is a single measure (number) used to summarize a sample data set. For example, the average height of students in this class.
- A *statistician* is an expert with at least a master's degree in mathematics or statistics or a trained professional in a related field.

Statistics is a *tool* for creating *new understanding* from a set of numbers.



Statistics is a science of getting informed decisions.

A Taxonomy of Statistics



Statistical Inference: is to draw conclusions about the Population on the basis of information available in the sample which has been drawn from the population by a random sampling technique/ procedure. There are two branches Statistical Inference namely ESTIMATION & TESTING OF HYPOTHESIS.

In ESTIMATION, we try to find an estimate of any population characteristic while in TESTING OF HYPOTHESIS, we try to test the statement about any population characteristic. *Here, our main concern is with "TESTING OF HYPOTHESIS".*

Some Basic Definitions:

Population: Any collection of individuals under study is said to be Population (Universe). The individuals are often called the members or the units of the population may be may be physical objects or measurements expressed numerically or otherwise.

Sample: A part or small section selected from the population is called a sample and process of such selection is called sampling.

(The fundamental object of sampling is to get as much information as possible of the whole universe by examining only a part of it. An attempt is thus made through sampling to give the maximum information about parent universe with the minimum effort).

Parameters: Statistical measurements such as Mean, Variance etc. of the population are called *parameters.*

Statistic: It a statistical measure computed from sample observations alone. The theoretical distribution of a statistics is called its sampling distribution. *Standard deviation of the sampling distribution of a statistic is called Standard Error.*

Hypothesis: is a statement given by an individual. Usually it is required to make decisions about populations on the basis of sample information. Such decisions are called *Statistical Decisions*. In attempting to reach decisions it is often necessary to make assumption about population involved. Such assumptions, which are not necessarily true, are called *statistical hypothesis*.

Parametric Hypothesis: A statistical hypothesis which refers only to values of unknown parameters of population is usually called *a parametric hypothesis.*

Null Hypothesis and Alternative Hypothesis: A hypothesis which is tested under the assumption that it is true is called a *null hypothesis* and is denoted by H_0. Thus a hypothesis which is tested for possible rejection under the assumption that it is true is known as *Null Hypothesis*. The hypothesis which differs from the given Null Hypothesis H_0 and is accepted when H_0 is rejected is called an *alternative hypothesis and is denoted by* H_1 *(The hypothesis against which we test the null hypothesis, is an alternative hypothesis).*

Simple and Composite Hypothesis: A parametric hypothesis which describes a distribution completely is called a *simple hypothesis* otherwise it is called *composite hypothesis*. For example; In case of Normal Distribution N (μ, σ^2), $\mu = 5$, $\sigma = 3$ is simple hypothesis whereas $\mu = 5$ is a composite hypothesis as nothing have been said about σ.
Similarly, $\mu < 5$, $\sigma = 3$ is a composite hypothesis.
Let H_0: $\mu = 5$ be the null hypothesis, then

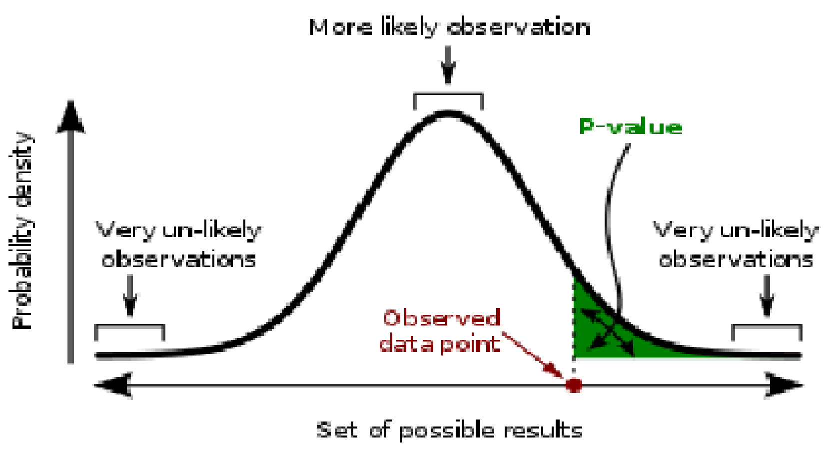
H_1: $\mu \neq 5$ is two sided composite alternative hypothesis.
H_1: $\mu < 5$ is one sided (Left) composite alternative hypothesis.
H_1: $\mu > 5$ is one sided (Right) composite alternative hypothesis.

Test: *Test is a rule through we test the null hypothesis against the given alternative hypothesis.*

Tests of Significance: Procedure which enables us to decide, on the basis of sample information whether to accept or reject the hypothesis or to determine whether observed sampling results differ significantly from expected results are called *tests of significance, rules of decisions or tests of hypothesis.*

Level of Significance: The probability level below which we reject the hypothesis is called *level of significance.* The levels of significance usually employed in testing of hypothesis are 5% and 1%.

P-Value: The **p-value** is the level of marginal significance within a **statistical** hypothesis test representing the probability of the occurrence of a given event. The **p-value** is used as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected.



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

A p value is used in hypothesis testing to help you support or reject the null hypothesis. The p value is the evidence **against** a null hypothesis. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

P values are expressed as decimals although it may be easier to understand what they are if you convert them to a percentage. For example, a p value of 0.0254 is 2.54%. This means there is a 2.54% chance your results could be random (i.e. happened by chance). That's pretty tiny. On the other hand, a large p-value of .9(90%) means your results have a 90% probability of being completely random and *not* due to anything in your experiment. Therefore, the smaller the p-value, the more important ("significant") your results.

Critical Region and Acceptance Region: A region (corresponding to a statistic t) is called the sample space. The part of sample space which amounts to rejection of null hypothesis H_0, is called *critical region or region of rejection*.

If $X = (x_1, x_2, \dots, x_n)$ is the random vector observed and W_c is the critical region (which corresponds the rejection of the hypothesis according to a prescribed test procedure) of the sample space W, then $W_a = W - W_c$ of the sample space is called the *acceptance region*.

Two Types of Errors in Testing of a Hypothesis: While testing a hypothesis H_0, the following four situations may arise:

(a) The test statistic may fall in the critical region even if H_0 is true then we shall be led to reject H_0 when it is true.

(b) The test statistic may fall in the acceptance region when H_0 is true we shall be led to accept H_0.

(c) The test statistic may fall in the critical region when H_0 is not true i.e. H_1 is true, then we shall be led to reject H_0 when H_1 is true.

(d) The test statistic may fall in the acceptance region even if H_0 is not true, then we shall be led to accept H_0 when it is not true.

It is quite obvious that the decisions taken in (b) and (c) are correct while the decisions taken in (a) and (d) are incorrect.

The wrong decision of rejecting a null hypothesis H_0, when it is true is called the ***Type I Error*** i.e. we reject H_0 when it is true. Similarly, the wrong decision of accepting the null hypothesis H_0 when it is not true is called the ***Type II Error*** i.e. we accept H_0 when H_1 is true.

Probability Forms:

$$P(reject\ H_0\ when\ it\ is\ true) = P(reject\ \ H_0/H_0) = \ \alpha$$

and

$$P(accept\ H_0\ when\ it\ is\ wrong) = P(reject\ \ H_0/H_1) = \ \beta$$

The α and β are called the size of Type I Error and size of Type II Error respectively.

Rules or Procedure for Testing of Hypothesis:
A test is a statistical procedure or a rule for deciding whether to accept or reject the hypothesis on the basis of sample values obtained.
Following is the Procedure for testing of Hypothesis:-

(a) Mention the null hypothesis H_0 to be tested along with an alternative hypothesis H_1.
(b) Make some assumptions such as the sample is random, the population is normal, the variance of two different populations are equal or unknown etc.
(c) Then find the most *appropriate test statistic* together with its sampling distribution. A statistic whose primary role is that of providing a test of some hypothesis is called a test statistic.
(d) On the basis of the sampling distribution make a decision to either accept or reject the null hypothesis H_0.
(e) Take a random sample and compute the test statistic. If the calculated value of the test statistic falls in the acceptance region, then accept the null hypothesis H_0. If it falls in the region of rejection (or Critical Region), reject the null hypothesis H_0.

Power Function of a Test: The Power Function of a test of a statistical hypothesis H_0: $\theta = \theta_0$, say, against alternative hypothesis H_1: $\theta > \theta_0$, $\theta < \theta_0$, $\theta \neq \theta_0$ is a function of the parameter, under consideration, which gives the probability that the test statistic will fall in the critical region when θ is the true value of the parameter.

Deduction: $P(\theta) = P$ (rejecting H_0 when H_1 is true)
$= P$ (W belong to W_c / H_1) $= 1 - P$ (accept H_0/ H_1)
$= 1 - \beta(\theta)$

The value of the power function at a particular value of the parameter is called the power of the test

Best Critical Region: In testing the hypothesis H_0: $\theta = \theta_0$ against the given alternative H_1: $\theta = \theta_1$, the critical region is best if the type II error is minimum or the power is maximum when compared to every other possible critical region of size α. *A test defined by this critical region is called* ***most powerful test.***

PARAMETRIC and NON-PARAMETRIC TESTS

In the literal meaning of the terms, a **parametric** statistical **test** is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn, while a non-**parametric test** is one that makes no such assumptions.

PARAMETRIC TESTS: Most of the statistical tests we perform are based on a set of assumptions. When these assumptions are violated the results of the analysis can be misleading or completely erroneous.

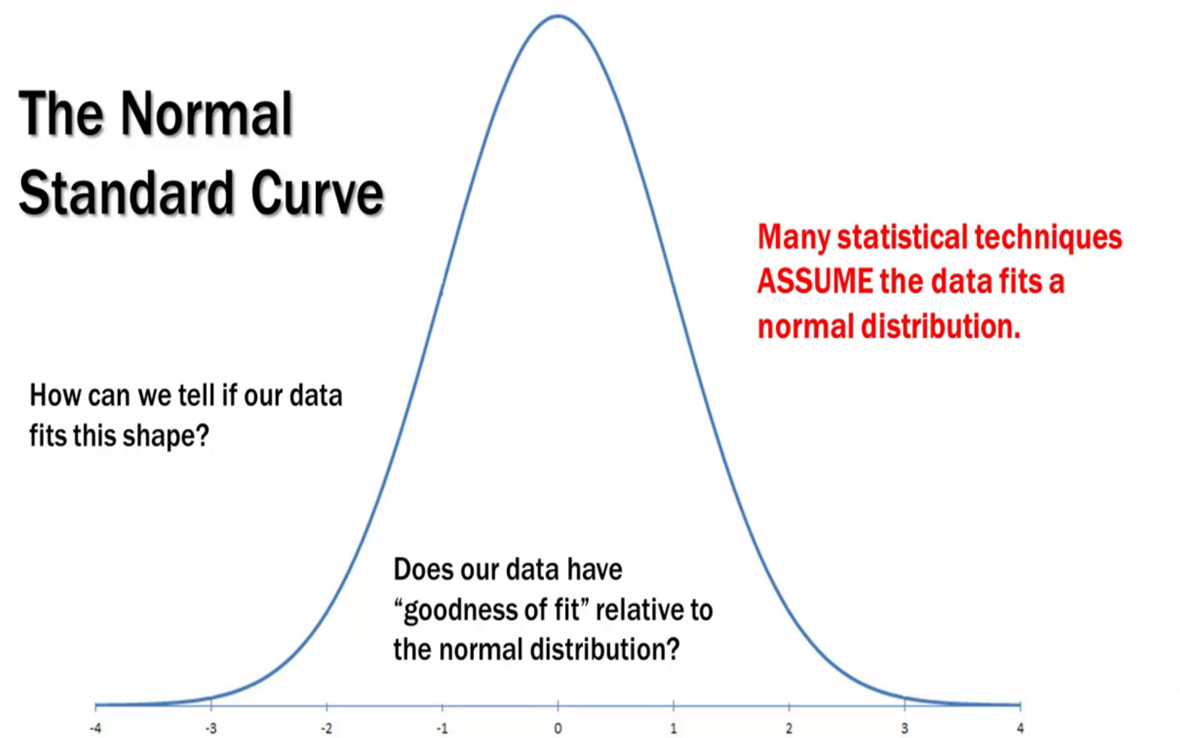
Typical assumptions are:

- **Normality**: Data have a normal distribution (or at least is symmetric)
- **Homogeneity of variances**: Data from multiple groups have the same variance
- **Linearity**: Data have a linear relationship
- **Independence**: Data are independent

We explore in detail what it means for data to be *normally distributed in Normal Distribution*, but in general it means that the graph of the data has the *shape of a bell curve*. Such data is symmetric around its mean and has kurtosis equal to zero. In Testing for Normality and Symmetry we provide tests to determine whether data meet this assumption.

Some tests (e.g. ANOVA) require that the groups of data being studied have the same variance. In Homogeneity of Variances we provide some tests for determining whether groups of data have the same variance.

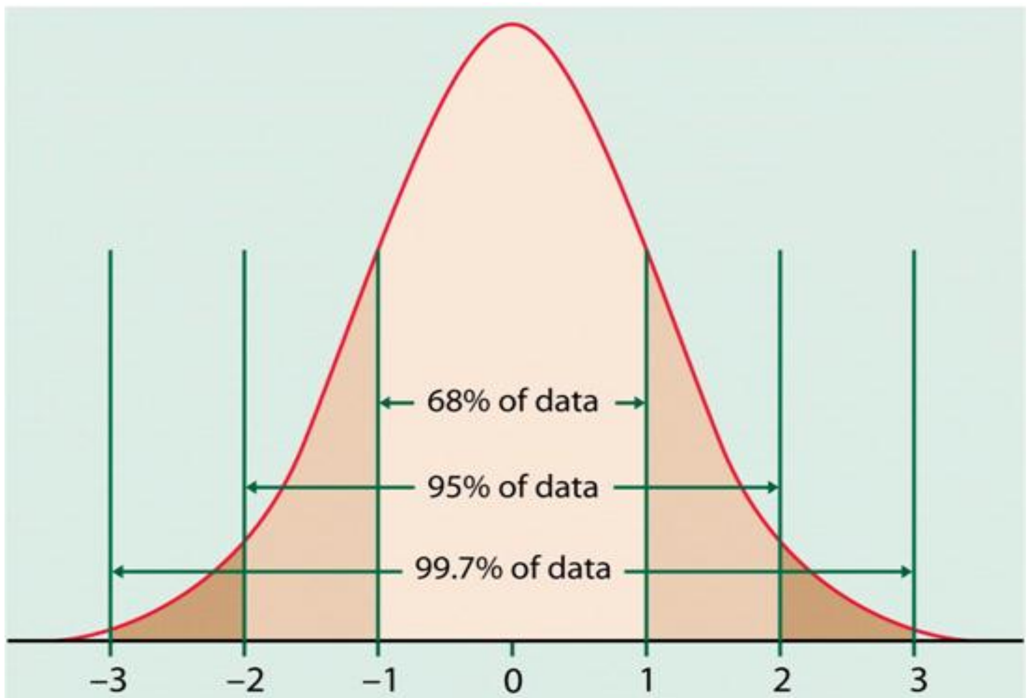
Some tests (e.g. Regression) require that there be a linear correlation between the dependent and independent variables. Generally linearity can be tested graphically using scatter diagrams or via other techniques explored in Correlation, Regression and Multiple Regression.

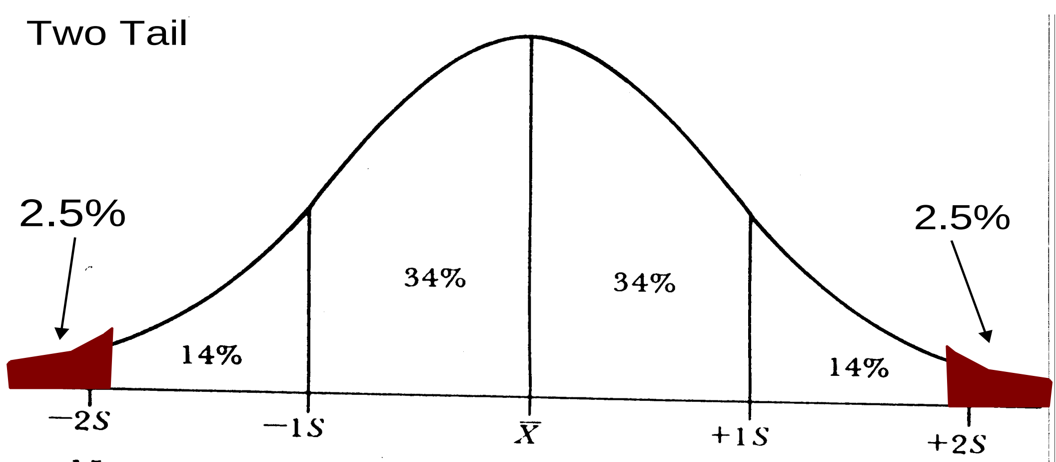


- *Many statistical methods require that the numeric variables we are working with have an approximate **normal distribution**.*

- *Standardized normal distribution with empirical rule percentages.*

- *For example, t-tests, F-tests, and regression analyses all require in some sense that the numeric variables are approximately normally distributed.*





5% region of rejection of null hypothesis
Non directional

t – test

Introduction:

- The t-test is a basic test that is limited to two groups. For multiple groups, you would have to compare each pair of groups. For example with three groups there would be three tests (AB, AC, BC) whilst with seven groups there would be need of 21 tests.
- The basic principle is to test the null hypothesis that means of the two groups are equal.

The t-test assumes:

- A normal distribution (parametric data)
- Underlying variances are equal (if not, use welch's test)
- It is used when there is random assignment and only two sets of measurement to compare.

There are two main types of t-test:

- **Independent – measures – t- test: when samples are not matched.**
- **Match – pair – t-test: when samples appear in pairs (eg. before and after)**

- **A single – sample t-test** compares a sample against a known figure. For example when measures of a manufactured item are compared against the required standard.

APPLICATIONS:

- To compare the mean of a sample with population mean. (Simple t-test)
- To compare the mean of one sample with the independent sample. (Independent Sample t-test)
- To compare between the values (readings) of one sample but in two occasions. (Paired sample t-test)

Independent Samples t-Test (or 2-Sample t-Test)

The independent samples t-test is probably the single most widely used test in statistics. It is used to compare differences between separate groups. In Psychology, these groups are often composed by randomly assigning research participants to conditions. However, this test can also be used to explore differences in naturally occurring groups. For example, we may be interested in differences of emotional intelligence between males and females.

Any differences between groups can be explored with the independent t-test, as long as the tested members of each group are reasonably representative of the population.

There are some technical requirements as well. PRINCIPALLY, EACH VARIABLE MUST COME FROM A NORMAL (OR NEARLY NORMAL) DISTRIBUTION.

Example: Suppose we put people on 2 diets: *the pizza di*et and *the beer diet.*

Participants are randomly assigned to either 1-week of eating exclusively pizza or 1-week of exclusively drinking beer. Of course, this would be unethical, because pizza and beer should always be consumed together, but this is just an example.

At the end of the week, we measure weight gain by each participant. Which diet causes more weight gain?

In other words, the null hypothesis is: Ho: wt. gain pizza diet = wt. gain beer diet.

(The null hypothesis is the opposite of what we hope to find. In this case, our research hypothesis is that there ARE differences between the 2 diets. Therefore, our null hypothesis is that there are NO differences between these 2 diets.)

		Column 3	***Column 4***
X_1 :Pizza	**X_2 : Beer**	$(X_1 - \overline{X}_1)^2$	$(X_2 - \overline{X}_2)^2$
1	**3**	**1**	**1**
2	**4**	**0**	**0**
2	**4**	**0**	**0**
2	**4**	**0**	**0**
3	**5**	**1**	**1**
2	**4**	**0.4**	**0.4**

$$s_x^2 = \frac{\sum (X - \overline{X})^2}{n} = 0.4$$

The formula for the independent samples t-test is:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_{x_1}^2}{n_1 - 1} + \frac{S_{x_2}^2}{n_2 - 1}}}$$

$$t = \frac{2-4}{\sqrt{\frac{.4}{4}+\frac{.4}{4}}} \approx -4.47$$

df = (n_1-1) + (n_2-1) = (5-1) + (5-1) = 8

After calculating the "t" value, we need to know if it is large enough to reject the null hypothesis. The "t" is calculated under the assumption, called the null hypothesis, *that there are no differences between the pizza and beer diet. If this were true, when we repeatedly sample 10 people from the population and put them in our 2 diets, most often we would calculate a "t" of "0."*

The calculated t-value is 4.47 (notice, I've eliminated the unnecessary "-" sign), and the degrees of freedom are 8. In the research question we did not specify which diet should cause more weight gain, therefore this t-test is a so-called "2-tailed t."

In the last step, we need to find the critical value for a 2-tailed "t" with 8 degrees of freedom. *(This is available from tables that are in the back of any Statistics textbook).*

Look in the back for **"Critical Values of the t-distribution,"** or something similar. The value you should find is: **C.V.** $_{t(8),\ 2\text{-tailed}}$ **= 2.31.**

The calculated t-value of 4.47 is larger in magnitude than the C.V. of 2.31, therefore we can reject the null hypothesis. Even for a results section of journal article, this language is a bit too formal and general. It is more important to state the research result, namely:
Participants on the Beer diet (*M* = 4.00) gained significantly more weight than those on the Pizza diet (*M* = 2.00), t (8) = 4.47, p < .05 (two-tailed).

Chi-Square as a Statistical Test

- ***Chi-square test:*** an **inferential statistics** technique designed to test for **significant relationships** between two variables organized in a bivariate table.
- **Chi-square requires no assumptions** about the shape of the population distribution from which a sample is drawn.
- A statistical method used to determine **goodness of fit**
  - Goodness of fit refers to how close the observed data are to those predicted from a hypothesis
- Note:
  - **The chi square test does not prove that a hypothesis is correct**
  - It evaluates to what extent the data and the hypothesis have a good fit.

Limitations of the Chi-Square Test:

- **The chi-square test** does **not** give us much information about the ***strength*** of the relationship or its ***substantive significance*** in the population.
- **The chi-square test** is **sensitive** to ***sample size***. The size of the calculated chi-square is **directly proportional** to the size of the sample, independent of the strength of the relationship between the variables.
- **The chi-square test** is also **sensitive** to **small expected frequencies** in one or more of the cells in the table.

Statistical Independence:

Independence (statistical): the **absence of association** between two cross-tabulated variables. The percentage distributions of the dependent variable within each category of the independent variable are **identical**.

Hypothesis Testing with Chi-Square:

<u>Chi-square follows five steps</u>:

1. Making assumptions (**random sampling**)
2. Stating the research and null hypotheses
3. Selecting the sampling distribution and specifying the test statistic
4. Computing the test statistic
5. Making a decision and interpreting the results

The Assumptions:

- The chi-square test requires **no assumptions** about the **shape of the population distribution** from which the sample was drawn.
- However, like all inferential techniques it assumes **random sampling**.

H_1: The two variables are **related** in the population. Gender and fear of walking alone at night are ***statistically dependent***.

Afraid	**Men**	**Women**	**Total**
No	83.3%	57.2%	**71.1%**
Yes	16.7%	42.8%	**28.9%**
Total	**100%**	**100%**	**100%**

H_0: There is **no association** between the two variables. Gender and fear of walking alone at night are ***statistically independent***.

The Concept of Expected Frequencies:

Expected frequencies f_e ***:*** the cell frequencies that would be **expected** in a bivariate table **if** the two tables were **statistically independent**.

Observed frequencies f_o***:*** the cell frequencies **<u>actually observed</u>** in a bivariate table.

Calculating Expected Frequencies:

f_e **= <u>(column marginal)(row marginal) / N</u>**

To obtain the expected frequencies for any cell in any cross-tabulation in which the two variables are assumed independent, multiply the row and column totals for that cell and divide the product by the total number of cases in the table.

Chi-Square (obtained):

The test statistic that summarizes the differences between the observed (*fo*) and the expected (*fe*) frequencies in a bivariate table.

Calculating the Obtained Chi-Square:

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

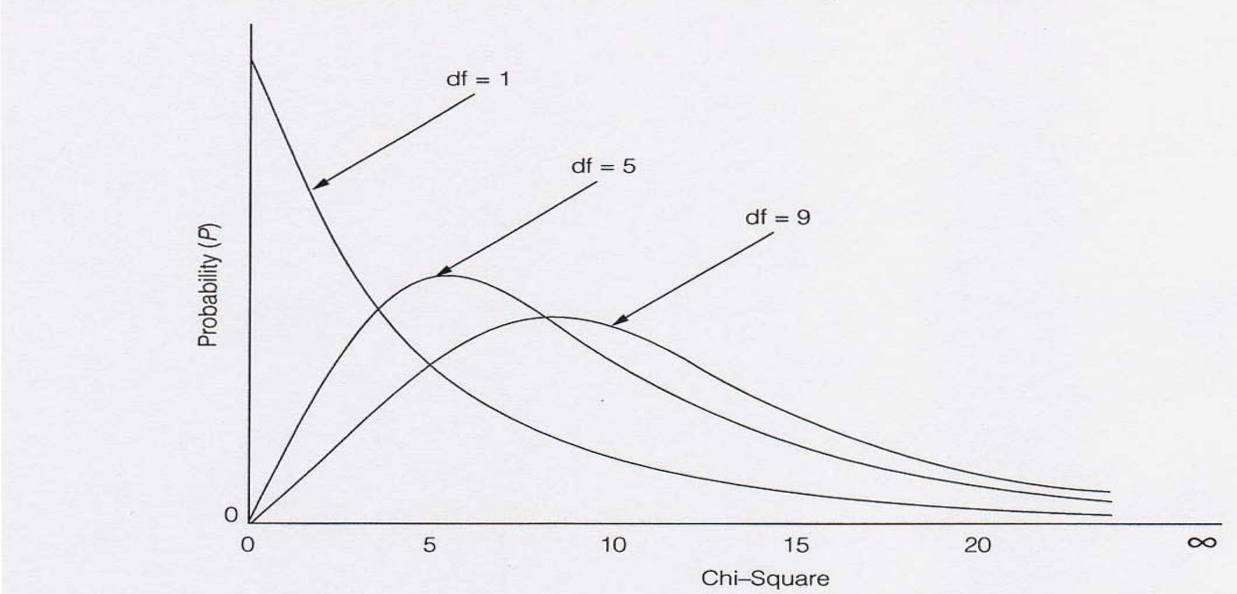
f_e = expected frequencies

f_o = observed frequencies

The Sampling Distribution of Chi-Square:

- The distributions are **positively skewed**. The research hypothesis for the chi-square is **<u>always</u>** a **one-tailed test**.
- Chi-square values are **<u>always</u> positive**. The minimum possible value is zero, with **no upper limit** to its maximum value.
- As the number of degrees of freedom increases, the **c^2** distribution becomes **more symmetrical**.

Figure 14.1 **Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom**



df = (r – 1)(c – 1)

where r = the number of rows ; c = the number of columns

(3 – 1)(2 – 1) = **2 degrees of freedom**

NON-PARAMETRIC STASTISTICS

The term non-parametric was first used by Wolfowitz, 1942. To understand the idea of nonparametric statistics it is required to have a basic understanding of parametric statistics which we have already discussed. A parametric test requires a sample to be normally distributed. A nonparametric test does not rely on parametric assumptions like normality.

Nonparametric test create flexible demands of the data. To make standard parametric legitimate, some provisions need to fulfilled, especially for minor sample sizes. For example, the requirement of the one sample t-test is that the observation must be made from ordinarily distributed population. In case, the provision is defined, then the resultants may not be credible. However, in case of Wilcoxon Signed rank test to illustrate valid inference, normality is not required.

We can assume that the sampling distribution is normal even if we are not sure that the distribution of the variable in the population is normal, as long as our sample is large enough, (for example, 100 or more observations). However, if our selected sample is too large, then those teats can only be utilized if we are assured that the variable is disseminated normally.

The applications of tests that are based on the normality assumptions are restricted by the deficiency of accurate measurement. For example, a study measures Grade Point Average (GPA) in place of percentage Marks. This measurement scale does not measure the exact distance between the marks of two students. GPA allows us only to rank the students from “good” to “poor” students. This measurement is called the ordinal scale. Statistical techniques such as Analysis of Variance, t-test etc. assume that the data are measured either on interval or ratio scale. In such situations where data is measured on nominal or ordinal scale nonparametric tests are more useful.

Thus, nonparametric tests are used when either:

- Sample is not normally distributed.
- Sample size is small.
- The variables are measured on nominal or ordinal scale.

There is at least one nonparametric equivalent for each parametric general type of test. Broadly, these tests fall into the following categories:

- Test of differences between groups (independent samples)
- Test of differences (dependent samples)
- Test of relationships between variables.

The concepts and procedure to undertake Run Test, Chi-Square Test, Wilcoxon Signed Rank Test, Mann-Whitney Test and Kruskal-Wallis Test are discussed here under:

Run Test: Run Test is used to examine the randomness of data. Many statistical procedures require data to be randomly selected. Run Test can be explained with the example of tossing of a fair coin, where the probability of getting a head or tail is equal which is 0.5. Suppose we denote Head by "1" and tail by "0" and record the outcomes as shown below, we conduct Run Test to see whether the sample is randomly chosen or not. The null and alternative hypotheses are:

H_0: The sample is randomly selected.

H_1: The sample is not randomly selected.

The Run length is calculated by computing the number of 0's or 1's in sequence. For example the run length for 0 and 1 in the following sequence is 5 and 4 respectively and number of runs is 2.

000001111

We shall conduct Run test on our data set to examine whether the students belonging to different faculties (graduation) are randomly selected or not.

Chi-Square Test: Chi-Square Test is used to examine the association between two or more variables measured on categorical scales. Chi-Square is used most frequently to test the statistical significance of result reported in bivariate tables, and interpreting bivariate tables is integral to interpreting the results of a chi-square test.

Bivariate tabular (Cross tabulation) analysis is used when trying to summarize the intersections of independent and dependent variables and to examine the relationship (if any) between those variables. For example, to know if there is any association between the **Gender** and their **Location,** Chi-Square test can be applied. In this case our dependent variable is **Location**. We control the independent variable **Gender** and elicit as well as measure the dependent variable Location to test the hypothesis, whether there is some association between these two variables.

The Chi-Square test is a statistical technique to examine the association or statistical independence between the row and column variables in a two – way table. The null and alternative hypotheses for Chi-Square Test are:

H_0: There is no association between the row (Gender) and column (Location) variables.

H_1: There is association between the row (Gender) and column (Location) variables.

Many researchers often get confused between statistical significance and strength of the relationship. People tend to think that more significant (the lower the P-value) relationship means stronger relationship. The significance level is influenced by the strength of the relationship and sample size. We require different measure to capture the strength of the relationship, or the effect size.

Mann- Whitney U Test: Generally, the t-test for independent samples is used, if two samples are compared over their mean value for some variable interest. Nonparametric alternatives for the test are the Wald-Wolfowitz Run test, the Mann Whitney U test, and the Kolmogorov-Smirnov two sample test.

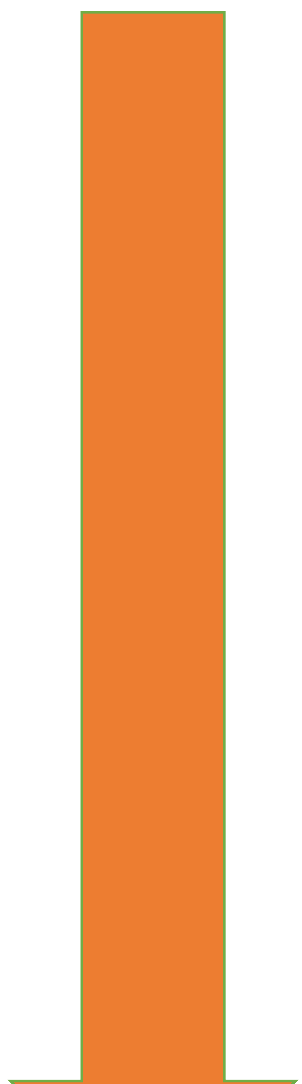
Mann Whitney U test compares the sums of ranks of two independent groups.

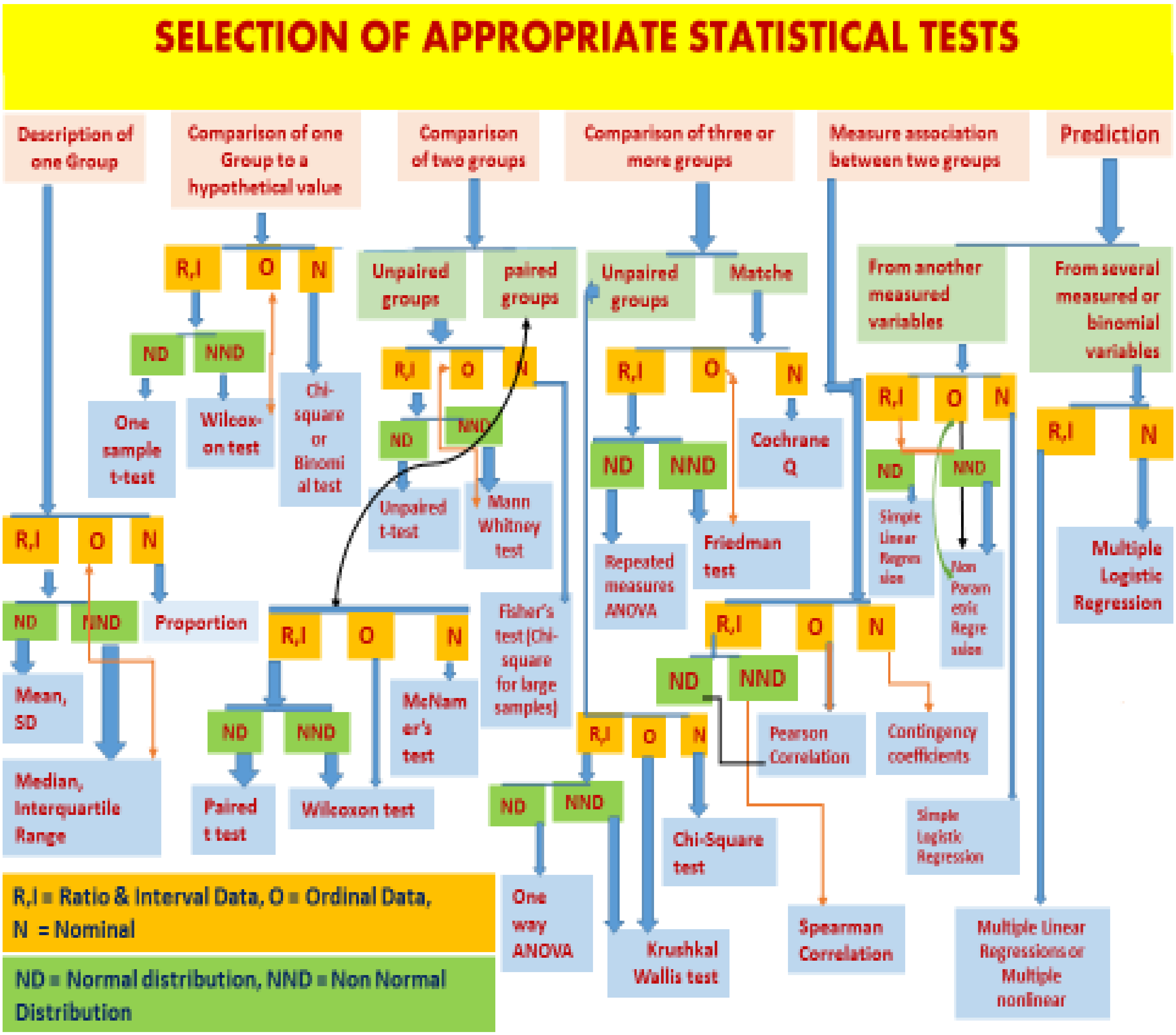
Wilcoxon Signed Rank Test: Wilcoxon Signed Rank Test (also known as Wilcoxon Matched Pair Test) is the non-parametric version of dependent sample t-test or paired sample t-test. Sign test is the other nonparametric alternative to the paired sample t-test. If the variables of interest are dichotomous in nature (Male and Female or Yes and No) then McNemar's Chi-Square test is used.

Wilcoxon Signed Rank Test is also a nonparametric version for one sample t-test. Wilcoxon Signed Rank Test compares the medians of the groups under two situations (paired samples) or it compares the median of the group with hypothesized median (one sample).

Kruskal –Wallis Test: Kruskal –Wallis Test is used with multiple groups. It is the non-parametric version of one-Way ANOVA. Median test is another nonparametric alternative to one-Way ANOVA. Kruskal –Wallis Test compares medians of more than two independent groups.

Below are the complete selection of Appropriate Statistical tests for Social Scientists/ Participants of the Workshop




SELECTION OF APPROPRIATE STATISTICAL TESTS
Description of one Group
Comparison of one Group to a hypothetical value
Comparison of two groups
Comparison of three or more groups
Measure association between two groups
Prediction
R,I
O
N
ND
NND
One sample t-test
Wilcox-on test
Chi-square or Binomial test
Unpaired groups
paired groups
Unpaired t-test
Mann Whitney test
Unpaired groups
Matche
Cochrane Q
Repeated measures ANOVA
Friedman test
From another measured variables
From several measured or binomial variables
Simple Linear Regression
Non Parametric Regression
Multiple Logistic Regression
Proportion
Mean, SD
Median, Interquartile Range
Paired t test
Wilcoxon test
McNamer's test
Fisher's test (Chi-square for large samples)
One way ANOVA
Krushkal Wallis test
Chi-Square test
Pearson Correlation
Contingency coefficients
Spearman Correlation
Simple Logistic Regression
Multiple Linear Regressions or Multiple nonlinear
R,I = Ratio & Interval Data, O = Ordinal Data, N = Nominal
ND = Normal distribution, NND = Non Normal Distribution

The flow chart of commonly used statistical tests: