Module 6

Hypothesis

WHAT IS A HYPOTHESIS?

- hypothesis may be defined as a proposition
forth as an explanation for the occurrence o
phenomena either asserted merely as a prov
guide some investigation or accepted as high
established facts.

- Quite often a research hypothesis is a predic
of being tested by scientific methods, that r
variable to some dependent variable

Characteristics of hypothesis:

(i) Hypothesis should be clear and precise. If t
and precise, the inferences drawn on its b
reliable.

(ii) Hypothesis should be capable of being test

(iii) Hypothesis should state relationship
happens to be a relational hypothesis.

(iv) Hypothesis should be limited in scope
researcher must remember that narrower h
more testable and he should develop such hyp

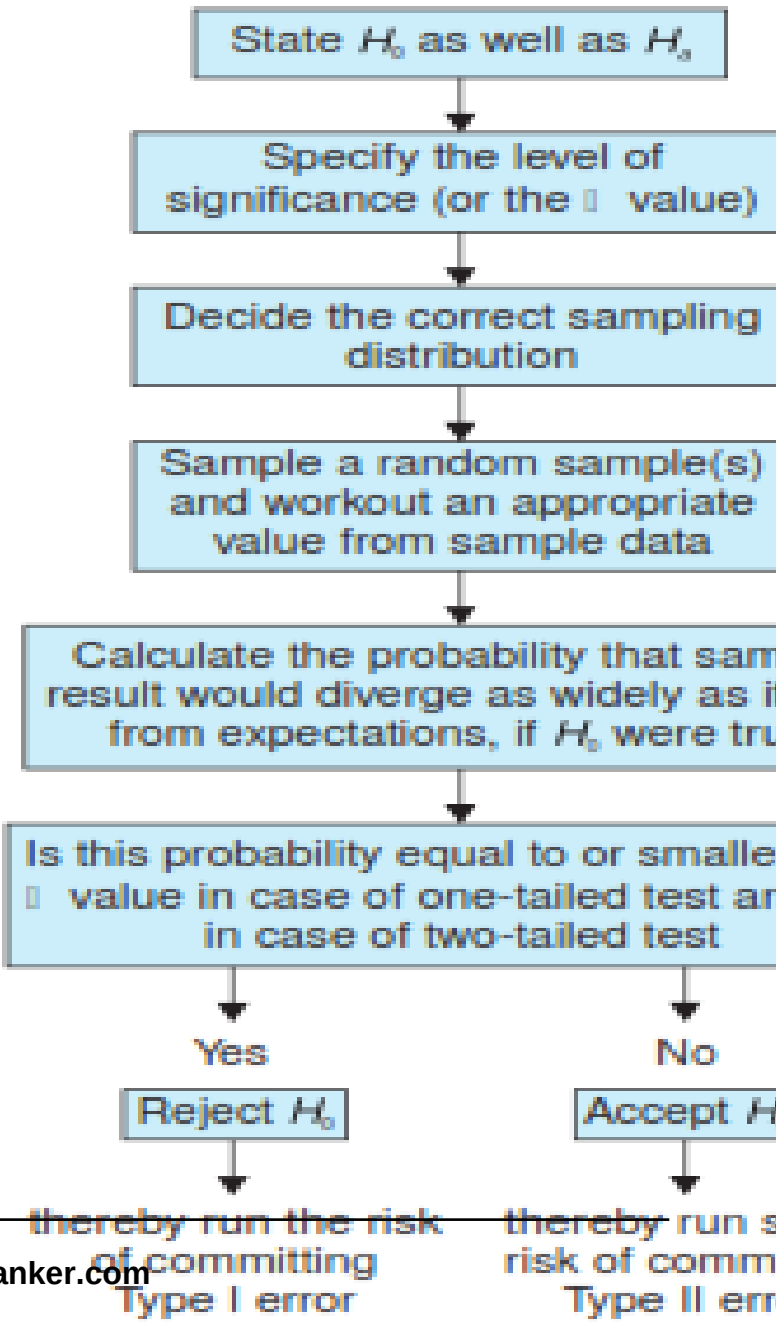
Characteristics of hypothesis:

(v) Hypothesis should be stated as far as possi
so that the same is easily understandable by
must remember that simplicity of hypothesis l
significance.

(vi) Hypothesis should be consistent with mos
be consistent with a substantial body of es
words, it should be one which judges accept a

PROCEDURE FOR HYPOTHESIS

FLOW DIAGRAM FOR HYPOTHESIS



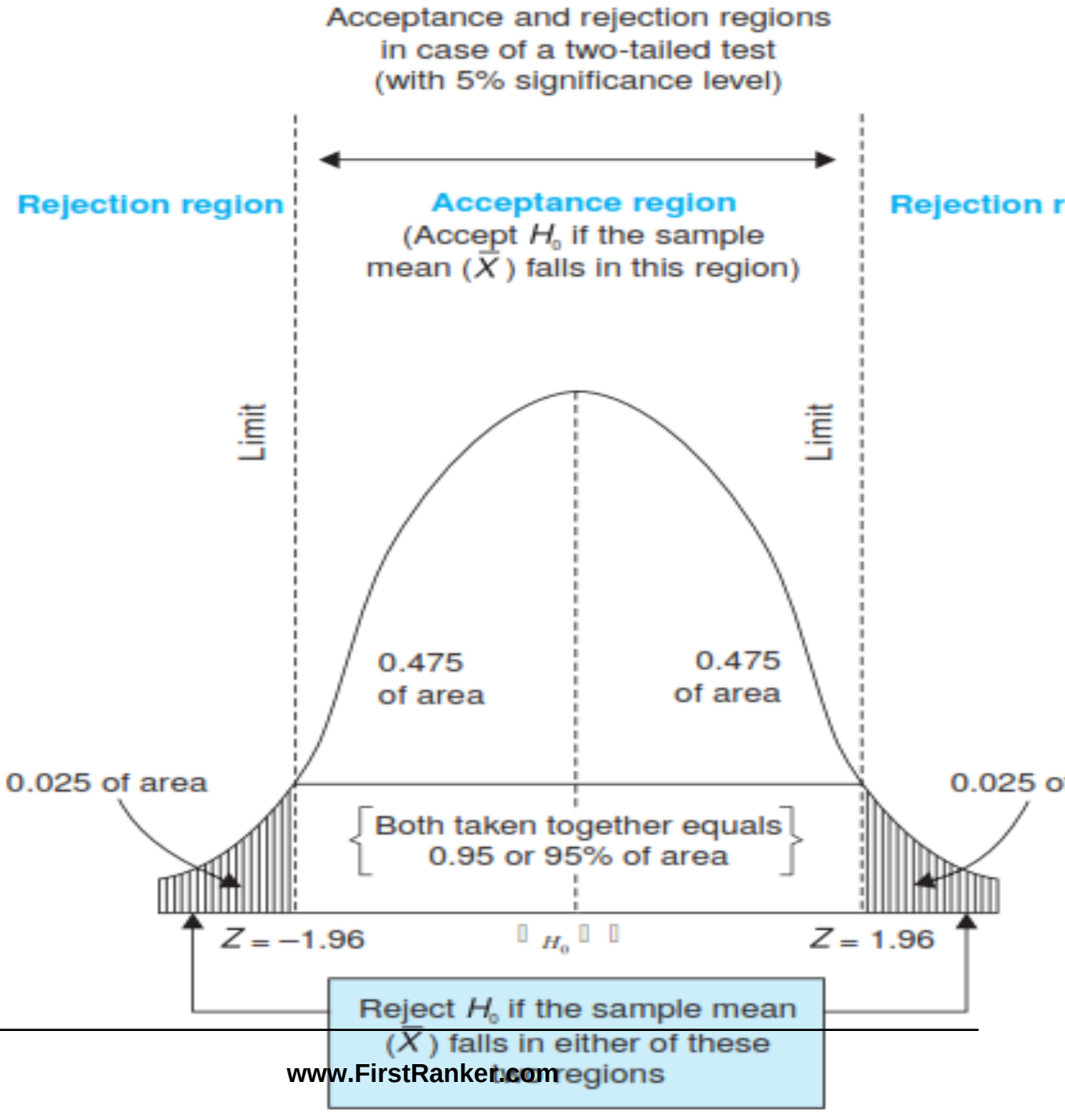
(i) State Ho and H1:

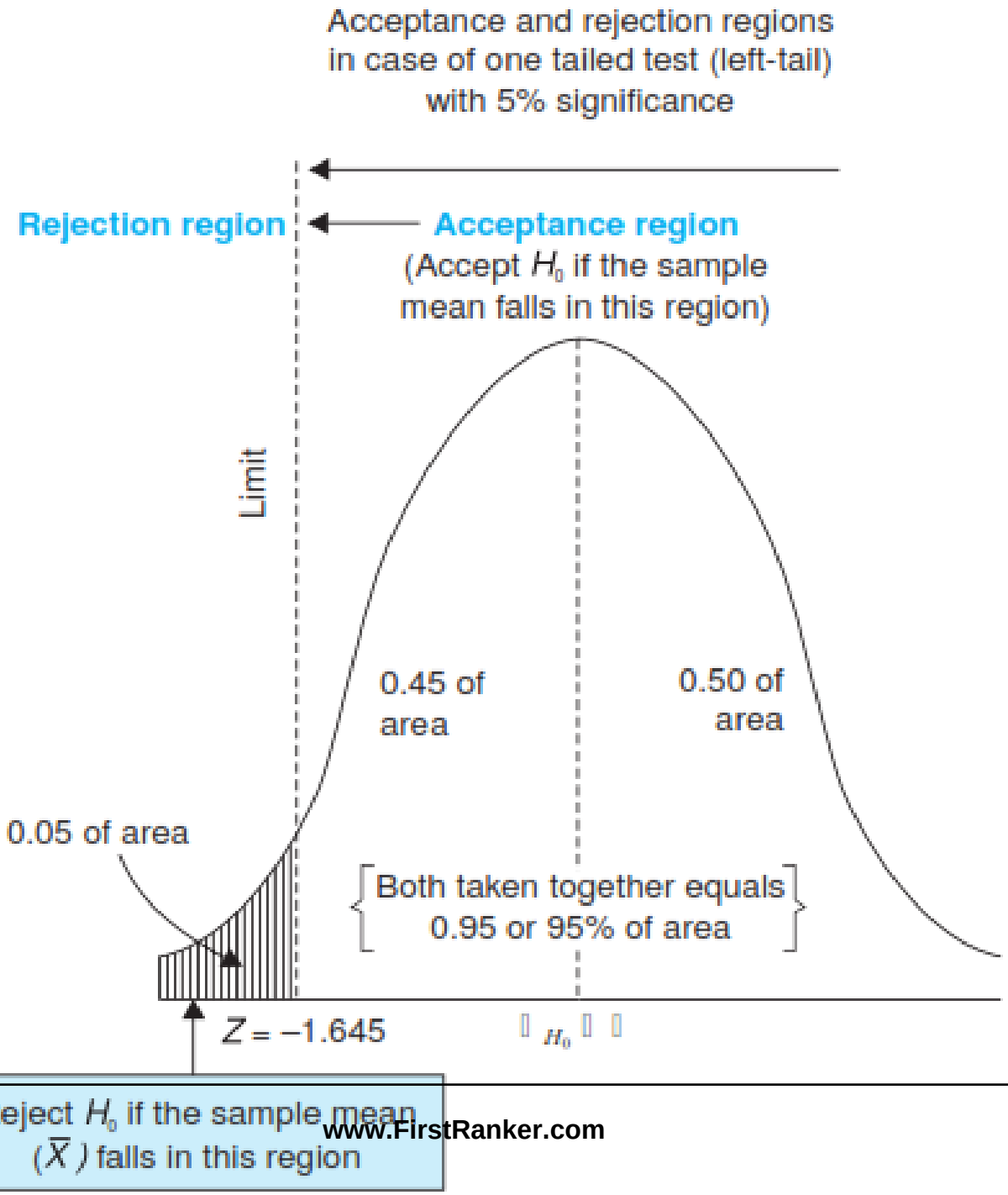
Null hypothesis $H_0 : \mu = 10$ tons

Alternative Hypothesis $H_a : \mu > 10$ tons

(ii) Selecting a Significance lev

- The hypotheses are tested on a pre-determ and as such the same should be specified.
- Generally, in practice, **either 5% level or 1%** purpose.
- The 5 per cent level of significance means th take as much as a 5 per cent risk of rejecting it (H) happens to be true.





(iii) Deciding the distribution t

After deciding the level of significance, the testing is to determine the appropriate sampli

The choice generally remains between normal t-distribution.

(iv) Selecting a random sampl computing an appropriate val

- Another step is to select a random sam appropriate value from the sample data cor utilizing the relevant distribution.
- In other words, draw a sample to furnish em

(v) Calculation of the probabil

One has then to calculate the probability that diverge as widely as it has from expectation were in fact true.

(vi) Comparing the probability

- Yet another step consists in comparing the p with the specified value for α , the significan
- If the calculated probability is equal to or sm case of one-tailed test, then reject the null h alternative hypothesis),
- but if the calculated probability is greater, th hypothesis.

Errors in hypothesis

- Type 1 error
  - Hypothesis is rejected when it is true
- Type 2 error
  - Hypothesis is not rejected when it is false

H_0 (true)		
H_0 (false)		

Types of tests

- Parametric test
- Non-parametric test

PARAMETRIC TESTS

- z- test – for large samples
- t- test- for small samples
- f- test- for significance of difference variance.

z-test

- A type of statistical analysis that considers th

The mean of the variable in a sample set and

The mean of the variable in a larger populatio

Circumstance where the Z tes

- A z-test is used for testing the mean of a popu comparing the means of two populations, with la you know the population standard deviation or n
- It is also used for testing the proportion of standard proportion, or comparing the proportion

Example: Comparing the average engineering sal

Example: Comparing the fraction defectives from 2

T-test

- A **T-test** is a statistical examination of two po
- A two-sample **t-test** examines

whether two samples are different and is com

- when the variances of two normal distributio

- when an experiment uses a small sample size

t-test – when to use

- A t-test is used for testing the mean of one popu
  comparing the means of two populations if you d
  standard deviation and when you have a limited sa
- If you know the populations' standard deviation, y

Example: Measuring the average diameter of shafts f
you have a small sample.

F-test

- **Definition**: **F-test** is a statistical **test** that is use
  populations having normal distribution hav
  standard deviation. This is an important pa
  (ANOVA).

WHEN?

- An F-test is used to compare 2 populations' var
  any size. It is the basis of ANOVA.

Example: Comparing the variability of bolt diame

Non Parametric te

U test

- (also called the **Mann–Whitney–Wilcoxon** (M
sum test, or **Wilcoxon–Mann–Whitney test**)

- is a nonparametric test of the null hypothesi
from the same population against an
especially that a particular population tend
than the other.

K-W test

- is a non-parametric method for testing whether
  same distribution.
- It is used for comparing two or more indepen
  different sample sizes. It extends the Mann–Wh
  more than two groups.
- The parametric equivalent of the Kruskal-Wallis
  of variance (ANOVA).

Bivariate Analysis

Introduction to bivaria

- When one measurement is made on eac analysis is applied.

  If more than one measurement is ma multivariate analysis is applied.

  In this section, we focus on bivariate an measurements are made on each obser

  The two measurements will be called X are obtained for each observation, the is the pair (X, Y).

- Bivariate data can be stored in a table with

	X	Y
Obs. 1	2	1
Obs. 2	4	4
Obs. 3	3	1
Obs. 4	7	5
Obs. 5	5	6
Obs. 6	2	1
Obs. 7	4	4
Obs. 8	3	1
Obs. 9	7	5
Obs. 10	5	6

- Some examples:
  - Height (X) and weight (Y) are meas
    ual in a sample.
  - Stock market valuation (X) and qua
    ings (Y) are recorded for each comp
  - A cell culture is treated with varying
    drug, and the growth rate (X) and
    (Y) are recorded for each trial.
  - Temperature (X) and precipitation
    a given day at a set of weather stat

CHI SQUARE T

INTRODUCTION

- The chi-square test is an important test amo
  tests of significance developed by statisticia
- Is was developed by Karl Pearson in1900.
- CHI SQUARE TEST is a non parametric test
  assumption or distribution of any variable.
- This statistical test follows a specific distrib
  square distribution.
- In general The test we use to measure the di
  what is observed and what is expected acco
  assumed hypothesis is called the chi-square

IMPORTANT CHARACTERISTI
SQUARE TEST

- This test (as a non-parametric test) is ba frequencies and not on the parameters standard deviation.
- The test is used for testing the hypothes useful for estimation.
- This test can also be applied to a compl table with several classes and as such i test in research work.
- This test is an important non-parametri assumptions are necessary in regard to population, no need of parameter valu less mathematical details are involved.

APPLICATIONS OF A CHI SQ

This test can be used in

1) Goodness of fit of distributions

2) test of independence of attributes

3) test of homogenity.

1) TEST OF GOODNESS OF FIT OF DISTRI

- This test enables us to see how well does t theoretical distribution (such as Binomia Poisson distribution or Normal distributi observed data.

- The $\chi 2$ test formula for goodness of fit is:

$$X^2 = \sum \frac{(o-e)^2}{e}$$

Where,
o = observed frequency
e = expected frequency

- If $\chi 2$ (calculated) > $\chi 2$ (tabulated), with (n- hypothesis is rejected otherwise accepted.

- ~~And if null hypothesis is accepted, then it~~ that the given distribution follows theoretic

2) TEST OF INDEPENDENCE OF ATTRIBUTES

- Test enables us to explain whether or not two associated.

- For instance, we may be interested in knowing medicine is effective in controlling fever or no useful.

- In such a situation, we proceed with the null h the two attributes (viz., new medicine and con independent which means that new medicine in controlling fever.

- χ^2 (calculated) > χ^2 (tabulated) at a certain leve significancfe for given degrees of freedom, the is rejected, i.e. two variables are dependent.(i. medicine is effective in controlling the fever) (calculated) < χ^2 (tabulated) ,the null hypothesi i.e. 2 variables are independent.(i.e., the new m effective in controlling the fever).

- when null hypothesis is rejected, it can be con a significant association between two attributes

3) TEST OF HOMOGENITY

- This test can also be used to test whether events follow uniformity or not e.g. the ac patients in government hospital in all day uniform or not can be tested with the hel test.

- χ^2 (calculated) < χ^2 (tabulated), then null h accepted, and it can be concluded that the in the occurance of the events. (uniformit of patients through out the week)

CONDITIONS FOR THE APPLI
TEST

The following conditions should be satisfied
be applied:

1) The data must be in the form of frequenc
2) The frequency data must have a precise
must be organised into categories or grou
3) Observations recorded and used are colle
basis.
4) All the itmes in the sample must be indep
5) No group should contain very few items,
case where the frequencies are less than
done by combining the frequencies of ad
that the new frequencies become greate
statisticians take this number as 5, but 1
better by most of the statisticians.)
6) The overall number of items must also be
It should normally be at least 50.

Multivariate analysis

- Multivariate analysis is essentially the simultaneously analysing multiple indep variables with multiple dependent (outcom using matrix algebra (most multivariate anal

Purpose.

- Behaviours, emotions, cognitions, and a described in terms of one or two variables.
- Furthermore, these traits cannot be measure speed, but must be inferred from constr measured by multiple factors or variables.

- Importance is usually based upon how much variance can be extracted from the data.
- Variance is a numerical representation of the (behaviour, emotion, cognition, etc.) in the p
- We assume it represents how much of that t individual.
- If two variables are associated or correlated they share some common underlying trait/fa equality in how they vary on the scores in th
- That underlying trait is causing them to co-va

Why the multivariate approac

With univariate analyses we have just one depend

Although any analysis of data involving more than as ‘multivariate’, we typically reserve the term variables

So MV analysis is an extension of UV ones, or c analyses are special cases of MV ones

Multivariate Pros and Cons Su

Advantages of using a multivariate statistic

- Richer realistic design
- Looks at phenomena in an overarching way (pro analysis)
- Each method differs in amount or type of Indep DVs
- **Can help control for Type I Error**

Disadvantages

- Larger Ns are often required
- More difficult to interpret
- Less known about the robustness of assumption

Introductio

- ANOVA is an abbreviation for the
method: Analysis Of Variance

Invented by R.A. Fis

- ANOVA is used to test the significa
difference between more than two
and to make inferences about whe
are drawn from population having

Theoretical Example: Sections At IMT

Sec: A

Sec: B

GK

CGPA

Indecipline

Continued..

- ANOVA is comparison of means. Ea
value of a factor or combination of
treatment.

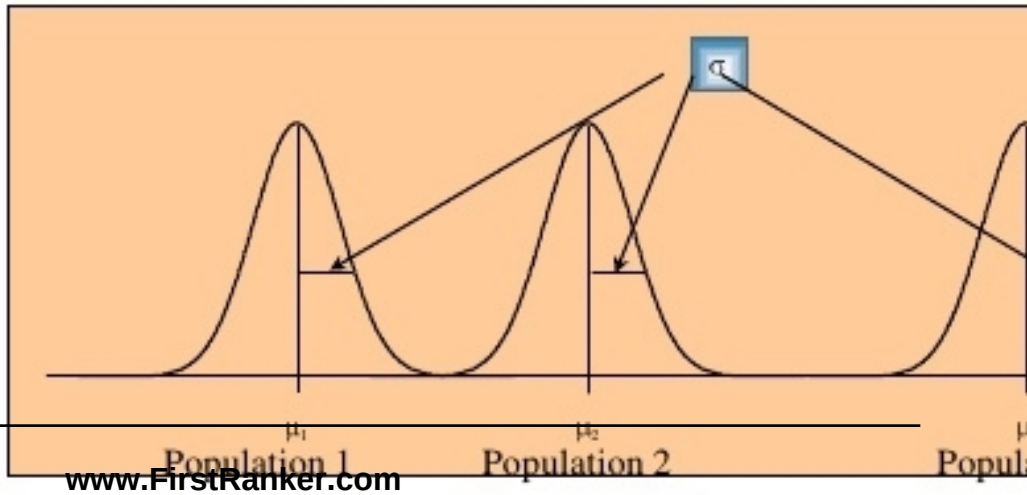
- The ANOVA is a powerful and com
procedure in the social sciences. It
variety of situations.

Why ANOVA instead of tests?

- If you are comparing means betwe two groups, why not just do severa tests to compare the mean from or mean from each of the other group
  - Before ANOVA, this was the only optic compare means between more than tw
- The problem with the multiple t-te that as the number of groups incre of two sample t-tests also increase
- As the number of tests increases th making a Type I error also increases

Analysis of Variance : Assu

- We assume *independent random sampli* the *r* populations
- We assume that the *r* populations unde
  - are *normally distributed*,
  - with means μ_i that may or may not be equal,
  - but with *equal variances*, σ_i^2.



ANOVA Hypoth

- The Null hypothesis for ANOV
means for all groups are equal

$$H_o : \mu_1 = \mu_2 = \mu_3 = \ldots =$$

- The Alternative hypothesis for
at least two of the means are n
- The test statistic for ANOVA is
F-statistic.

One & N way AN

- One way ANOVA
  Analysis of variance, so named
  can consider only one indepen
  variable at a time.

- N way ANOVA
  As its name suggests, this is a
  allows you to examine the effe
  independent variables concurr