Module 5

Hypothesis

WHAT IS A HYPOTHESIS?

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

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

Characteristics of hypothesis:

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

(ii) **Hypothesis should be capable of being tes**

(iii) Hypothesis should state relationship b
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 a

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

Characteristics of hypothesis:

- (vii) Hypothesis should be amenable to tes
  time.

One should not use even an excellent hypothe
tested in reasonable time for one cannot spe
data to test it.

Characteristics of hypothesis:

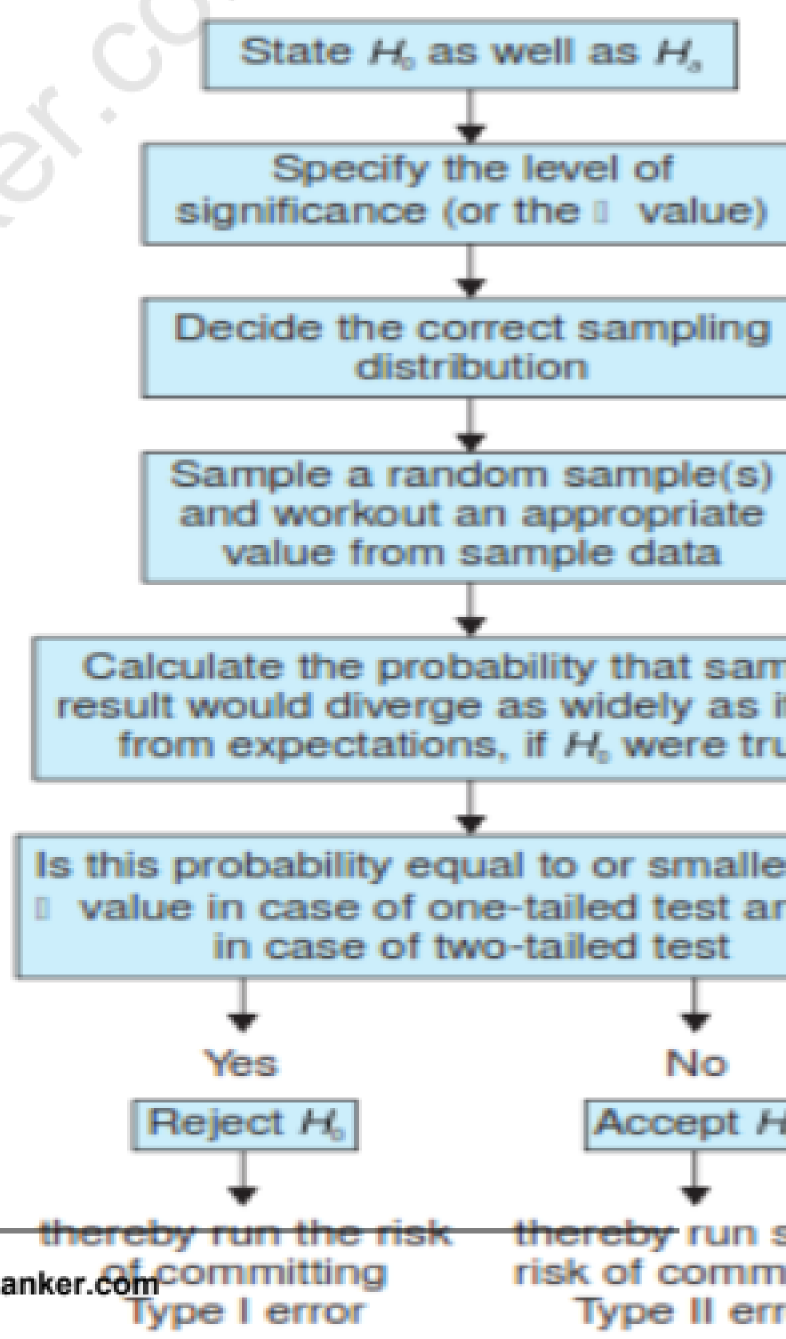
(viii) Hypothesis must explain the facts that g

explanation.

- Thus hypothesis must actually explain what i should have empirical reference

PROCEDURE FOR HYPOTHESIS

FLOW DIAGRAM FOR HYPOTHESIS



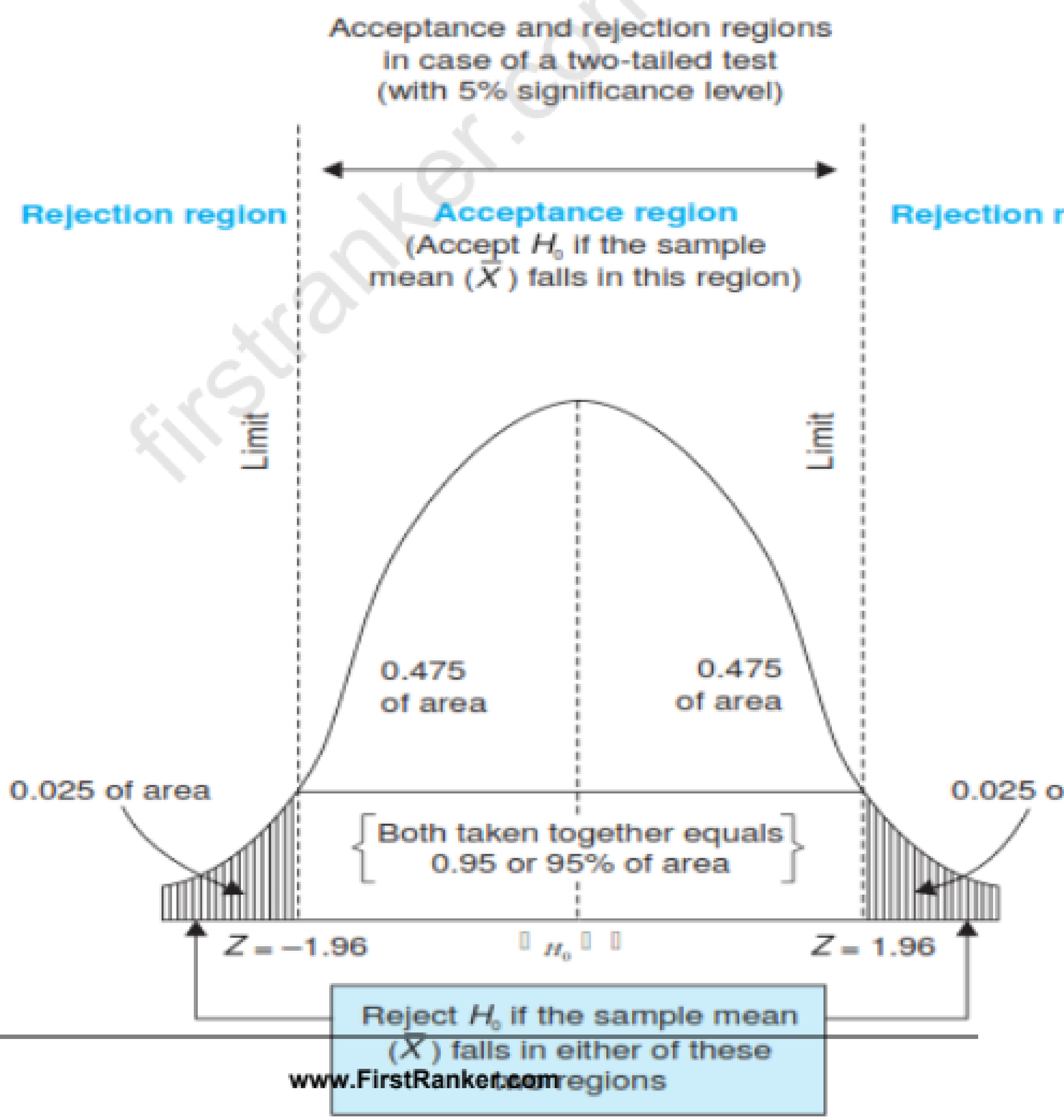
(i) State Ho and H1:

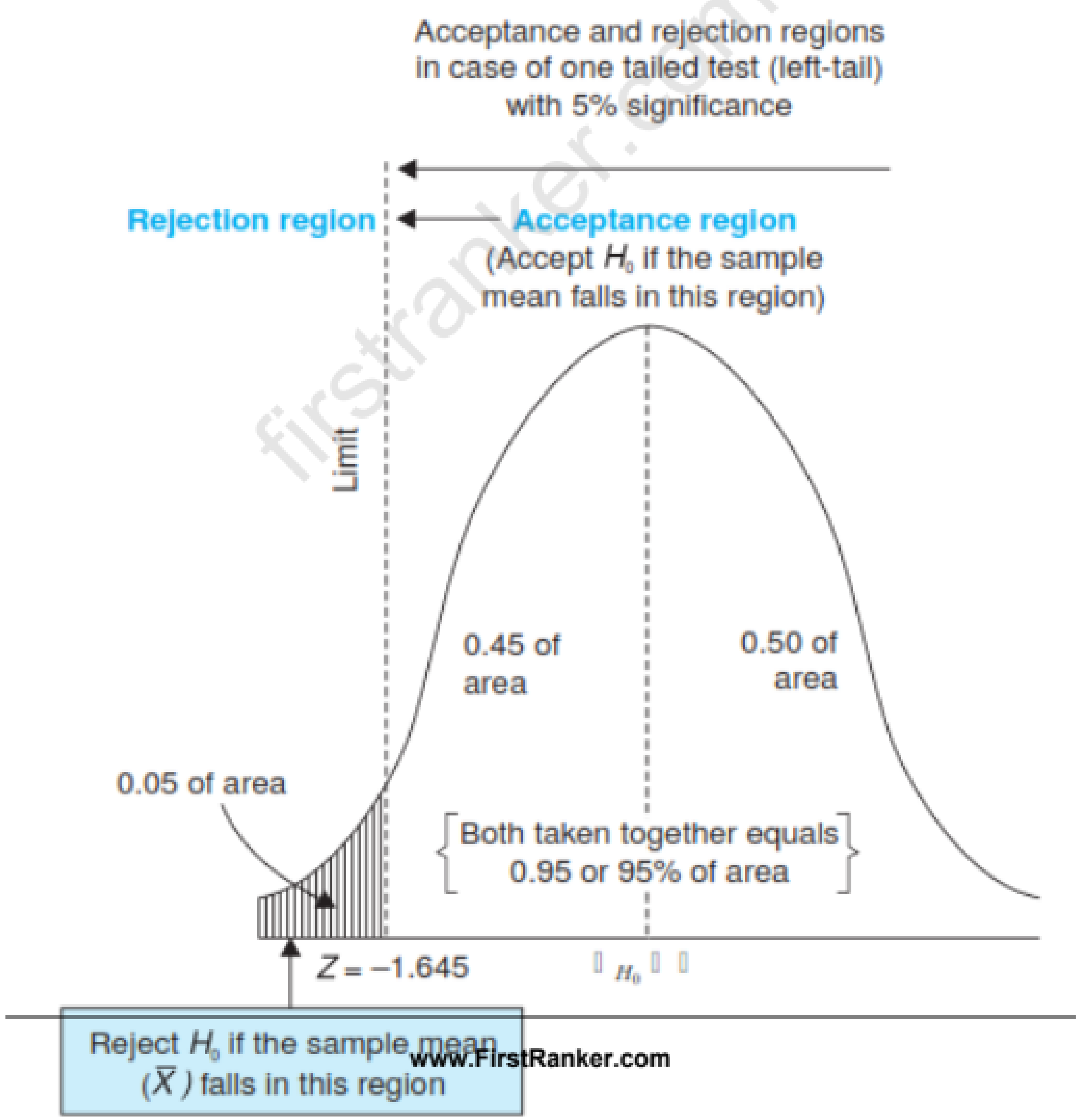
Null hypothesis $H_0 : \mu = 10$ tons

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.


Acceptance and rejection regions
in case of a two-tailed test
(with 5% significance level)
Rejection region
Acceptance region
(Accept H_0 if the sample
mean ($\bar{X}$) falls in this region)
Rejection r
Limit
Limit
0.475
of area
0.475
of area
0.025 of area
0.025 o
Both taken together equals
0.95 or 95% of area
Z = −1.96
μ_{H_0}
Z = 1.96
Reject H_0 if the sample mean
($\bar{X}$) falls in either of these
regions


Acceptance and rejection regions in case of one tailed test (left-tail) with 5% significance
Rejection region
Acceptance region
(Accept H_0 if the sample mean falls in this region)
Limit
0.45 of area
0.50 of area
0.05 of area
Both taken together equals 0.95 or 95% of area
$Z = -1.645$
Reject H_0 if the sample mean ($\bar{X}$) falls in this region

(iii) Deciding the distribution t

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

The choice generally remains between **norma**
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 co 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

- Comparing the probability thus calculated with t
  the significance level.
- If the calculated probability is equal to or smalle
  of one-tailed test, then reject the null hypothesi
  alternative hypothesis),
- but if the calculated probability is greater, then

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 s standard proportion, or comparing the proportio**

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.

1.Testing difference between samples (independent Sample

Alternative Hypothesis: Nourishment programme I
programme 'B' increase the children's weight signific

Solution:

	Nourishment programme A		
X	$x-\bar{x}$ = (x-46)	$(x-\bar{x})^2$	y
44	-2	4	42
37	-9	81	42
48	2	4	58
60	14	196	64
41	-5	25	64
			67
			62
230	0	310	39

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\sum x = 230 \qquad \sum y =$$

$$\sum \left(x - \bar{x}\right)^2 = 310 \qquad \sum \left(y\right.$$

$$\bar{x} = \frac{\sum x}{n_1} = \frac{230}{5} = 46$$

$$\bar{y} = \frac{\sum y}{n_2} = \frac{399}{7} = 57$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left\{\sum\left(x - \bar{x}\right)\right.$$

$$D.F = (n_1 + n_2 - 2) = (5 + 7 - 2)$$

$$s^2 = \frac{1}{10}\{310 + 674\} = 98.4$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$t =$

$=$

$=$

$=$

When is a one-tailed test appr

- If you are using a significance level of 0.05, a
  of your alpha to testing the statistical signific
  of interest.
- This means that 0.05 is in one tail of the distr
  statistic.
- When using a one-tailed test, you are testing
  relationship in **one direction** and completely
  possibility of a relationship in the other direc

- For example, imagine that you have developed a
- It is cheaper than the existing drug and, you beli
- In testing this drug, you are only interested in te than the existing drug.
- You do not care if it is significantly **more effectiv**
- You only wish to show that it is not less effective tailed test would be appropriate.

- Our null hypothesis is that the mean is equa
- A one-tailed test will test either if the mean

  than *x*

  Or

  - if the mean is significantly less than x, bu

Two-tailed test

- If you are using a significance level of 0.05, a of your alpha to testing the statistical signific half of your alpha to testing statistical s direction.
- This means that .025 is in each tail of the statistic.

- Our null hypothesis is that the mean is equal
- A two-tailed test will test both if the mean is than *x* and if the mean significantly less than
- The mean is considered significantly differen statistic is in the top 2.5% or bottom 2.5% of distribution, resulting in a p-value less than 0

2. To test Significance of the r Random sample

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$$

$\bar{X}$ = the mean of the sample
μ = the actual or hypothetical mean
n = the sample size
S = the standard deviation of the sar

$$S = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n-1}} \text{ or } S = \sqrt{\frac{\Sigma d^2 - n(}{n-1}}$$

$$= \sqrt{\frac{1}{(n-1)}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]}$$

d = deviation from the assumed me

Fiducial limits of Population M

- Assuming that the saple is a random sample
population of unknown mean the 95% fiduci
mean (μ) are

$\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.0}$

& 99% limits are.

$\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.0}$

Example

The manufacturer of a certain make of electric
bulbs have a **mean life of 25** months with a st
months. A random sample of 6 such bulbs gav

Life of months : 24 26 30 20 20 18

Can you regard the producer's claim to be vali
significance?

CALCULATION OF $\bar{X}$ an

X	$(X - \bar{X})$ x
24	+1
26	+3
30	+7
20	−3
20	−3
18	−5
$\Sigma X = 138$	

$$\bar{X} = \frac{\Sigma X}{n} = \frac{138}{6} = 23$$

$$S = \sqrt{\frac{\Sigma x^2}{n-1}} = \sqrt{\frac{102}{5}} = \sqrt{2}$$

$$= \frac{|23 - 25|}{4.517}\sqrt{6} = \frac{2 \times 2.449}{4.517}$$

$\nu = n - 1 = 6 - 1 = 5$. For $\nu =$

3.Testing difference between me samples (Dependent Samples or observations)

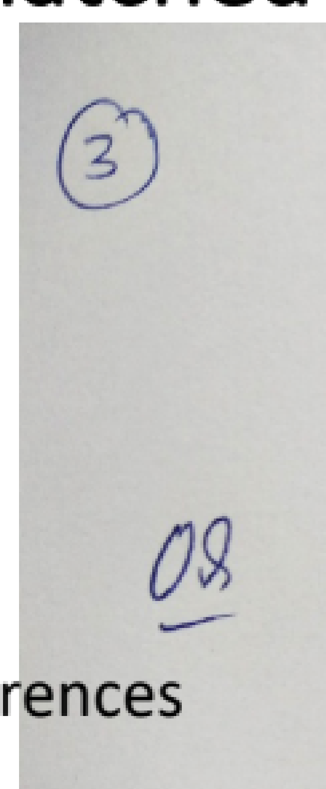
in this two samples are said to be dependent whe sample are related to those in the other in any pa on the same subject.

Two samples may consists of pairs of observation the same selected population.

Say, for effectiveness of coaching class or training

3.Testing difference between me
(Dependent Samples or matched

It is defined by



$\bar{d}$ =the mean of the differences

S = the standard deviation of the differences

If t > table value then Ho is rejected

If t < table value then Ho is accepted

- The value of the **S** is calculated by,

S =

Or

S =

- It should be noted that ***t*** is based on ***n-1*** deg

Example

- To verify whether a course in accounting imp
similar test was given to 12 participants both
course. The original marks recorded in alpha
participants were- 44,40,61,52,32,44,70,41,6

  After the course, the marks were in the same

  53,38,69,57,46,39,73,48,73,74,60 and 78. wa

Hypothesis :

there is no difference in the marks obta
course, i.e. the course has not been useful

Participants	Before (1st Test)	After (2nd Test)
A	44	53
B	40	38
C	61	69
D	52	57
E	32	46
F	44	39
G	70	73
H	41	48
I	67	73
J	72	74
K	53	60
L	72	78

$$\bar{d} = \frac{\Sigma d}{n} = \frac{60}{12} = 5$$

$$S = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}} = \sqrt{\frac{578 - 12}{12 -}}$$

$$t = \frac{5 \times \sqrt{12}}{5 \cdot 03} = \frac{5 \times 3 \cdot 464}{5 \cdot 03} = 3 \cdot 443$$

$v = n - 1 = 12 - 1 = 11$; For $v = 1$

4.Testing the Significance of a Correlation Coefficient

- Given random sample from bivariate popula
- If we are to test the hypothesis that the corr population is zero.
- i.e. the variable in the population are uncorr the test.

$$t = \frac{r}{\sqrt{1-r^2}}$$

- Here t is based on n-2 degree of freedom.
- We say that the value of ***r*** is significant at 5% le
- If $\boldsymbol{t} < \boldsymbol{t}_{0.05}$ the data are considered with the hyp uncorrelated.
- ***r***= correlation coefficient

Example 1

- A random sample of 27 pairs of observati
  normal population gives a correlation coeffic
  is it likely that the variables in the pop
  uncorrelated?

- **Solution**
  - Ho : there is no significant difference in the samp
    and correlation in the population

Example 2

- How many pairs of observations must be inc
order that an observed correlation coefficier
a calculated value of r greater than 2.72?

Solution-

Here, r value is 0.42, we have to find out n

Example 3

- The following table gives the ages in years of
  wives at marriage. Compute the correlation
  significance.
- Husband's Age : 23 27 28 29 3
  36 39
- Wife's Age : 18 22 23 24 25 2
  32

- Solution
- Ho : there is no correlation in the populatio
- Hear you need to find our

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

- The F-Test is named in honor of the great sta

- The object of the ***F-test*** is to find out wheth estimates of population variance differ sign two samples may be regarded as drawn from having the same variance. For carrying out th calculate the ratio *F*.

$$F = \frac{S_1^{\ 2}}{S_2^{\ 2}}, \text{ where } S_1^{\ 2} = \frac{\Sigma (X_1}{n_1}$$

$$\text{and} \qquad S_2^{\ 2} = \frac{\Sigma (X_2}{n_2}$$

$S_1^2 > S_2^2$ means S1 is always larger estimate o

V1 = Degree of Freedom for samples having larger va

V2 = Degree of Freedom for samples having Smaller v

$$V1 = n_1 - 1$$

If $F > table$ value then F ratio is considered as signi
H1 is accept

If $F < table$ value accept Ho, means both the sample

population having same variance.

A X_1	$(X_1 - \overline{X}_1)$ x_1	x_1^2	B X_2
66	−14	196	64
67	−13	169	66
75	−5	25	74
76	−4	16	78
82	+2	4	82
84	+4	16	85
88	+8	64	87
90	+10	100	92
92	+12	144	93
			95
			97
$\Sigma X_1 = 720$	$\Sigma x_1 = 0$	$\Sigma x_1^2 = 734$	$\Sigma X_2 = 913$

H_0 = Two Populations Have Same Variance

$V1 = 10$

$V2 = 8$

$F_{0.05} = 3.36$

CALCULATED VALUE OF F = 0.707

The Calculated F Value Is Less Than Table Value HENCE THE HYPO

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.

Mann–Whitney-*U* test

- This test is to determine whether two indepe
  been drawn from the same population.
- This test applies in very general conditions ar
  populations sampled are continuous.

Mann–Whitney-*U* test Steps t

- We first of all rank the data jointly, taking th
single sample in either an increasing or decre
magnitude.
- We usually adopt low to high ranking proces
rank 1 to an item with lowest value, rank 2 t
so on.
- In case there are ties, then we would assign
observation the mean of the ranks which the
  - (for 11 11 11 rank may be (6+7+8)3=7)

Mann–Whitney-*U* test Steps t

- After this we find the sum of the ranks assign first sample(R1) and also the sum of the rank of the second sample(R2)
- Then we work out the test staststic i. e.U wh the difference between the ranked observat under
  - $U = n1 \times n2 + \frac{n1\,(n1+1)}{2} - R1$

n1 and n2 are the sample sizes and R1 is the su the values of the first Sample

Mann–Whitney-*U* test Steps t

- Ho : two samples are from identical populati
- If the null hypothesis that the n1 +n2 observ identical population is true, the said U statsti distribution with

$$\text{Mean} = \mu = \frac{n1\ \ n2}{2}$$

$$\text{\& SD } \sigma = \frac{\sqrt{n1 n2\ (n1+n2+1)}}{12}$$

- If n1 and n2 are sufficiently large (i.e. both g
sampling distribution of U can be approxima
distribution and the limits of the acceptance
determined in the usual way at a given level

Example 1

- The value in one sample are 53 38 69 57 46 3
  In another sample they are 44 40 61 52 32 4

  test at the 10% level the hypothesis that the
with the same mean. Apply *U-test*

Solution

size of sample item in ascending order	Rank	Re
32	1	B
38	2	A
39	3	A
40	4	B
41	5	B
44	6.5	B
44	6.5	B

- R1 = 2+3+8+9+11.5+13+14+17+21.5+21.5+23
- R2= 1+4+5+6.5+6.5+10+11.5+15+16+18+19.5
- N1 = 12
- N2 = 12

- U = 54.5

- Since in the given proble n1 and n2 both are sampling distribution of U approximately clo
- Keeping this in view, we work out the mean hypothesis that the two samples come from under

μ = 72

σ = 17.32

- As the alternative hapothesis is that the mea
populations are not equal, a two-tailed test i
the limits of acceptance region, keeping in vi
significance as given, can be worked out as u

- As Z value for 0.45 of the are under the norm
following limits of acceptance region
- Upper limit = μ +1.64 σU = 100.40
- Lower limit = μ -1.64 σU = 43.60

The Kruskal-Wallis test (or H t

- This test is conducted in a way similar to the
- This test is used to test the null hypothesis th random samples come from identical univers alternative hypothesis that the means of the equal.
- This test is analogous to the one-way analysi the latter it does not require the assumption from approximately normal populations or t same standard deviation.

The Kruskal-Wallis test (or H t

- In this test, like the *U* test, the data are ranke high or high to low as if they constituted a si
- The test statistic is *H* for this test which is wo

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

where n = n1 + n2 + ... + nk and Ri being the sum of observations in the ith sample.

The Kruskal-Wallis test (or H te

- If the null hypothesis is true that there is no sample means and each sample has at leas sampling distribution of H can be approxim distribution with $(k - 1)$ degrees of freedom.
- As such we can reject the null hypothesis at a g if H value calculated, as stated above, excee value of chi-square.

* If any of the given samples has less than fiv distribution approximation can not be used and based on table

The Kruskal-Wallis test (or H t

Illustration

- Use the Kruskal-Wallis test at 5% level of sig
  hypothesis that a professional bowler perfo
  four bowling balls, given the following result

Bowling Results in Five G		
With Ball No. *A*	271	282
With Ball No. *B*	252	275
With Ball No. *C*	260	255
With Ball No. *D*	279	242

Bowling results	*Rank*
302	1
297	2
282	3
279	4
276	5
275	6
271	7
270	8
268	9
266	10
262	11
260	12
258	13
257	14
255	15
252	16
248	17
246	18
242	19
239	20

Table 12.7 (a): Bowling Results with Different Balls a

Ball A	*Rank*	*Ball B*	*Rank*	*Ball C*	*R*
271	7	252	16	260	
282	3	275	6	255	
257	14	302	1	239	
248	17	268	9	246	
262	11	276	5	266	
$n_1 = 5$	$R_1 = 52$	$n_2 = 5$	$R_2 = 37$	$n_3 = 5$	R_3

Now we calculate *H* statistic as under:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{20(20+1)} \left\{ \frac{52^2}{5} + \frac{37^2}{5} + \frac{75^2}{5} \right.$$

$$= (0.02857)\,(2362.8) - 63 = 67.51$$

- As the four samples have five items* each, t
of *H* approximates closely with chi-square di
- Now taking the null hypothesis that the bow
with the four balls, we have the value of chi-
or 4 – 1 = 3 degrees of freedom at 5% level o
- Since the calculated value of *H* is only 4.51
c2 value of 7.815, so we accept the null hypo
bowler performs equally well with the four b

Illustration:

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 ar
  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 DISTR

- 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.

- $\chi 2$ (calculated) > $\chi 2$ (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) a
(calculated) < $\chi 2$ (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 ad patients in government hospital in all day uniform or not can be tested with the hel test.

- $\chi 2$ (calculated) < $\chi 2$ (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.

- Behaviors, emotions, cognitions, and at 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
  (behavior, emotion, cognition, etc.) in the po
- 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 assumptio

Introductio

- ANOVA is an abbreviation for the f
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

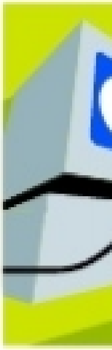
Sec: A

Sec: B

GK

CGPA

Indecipline



Continued..

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

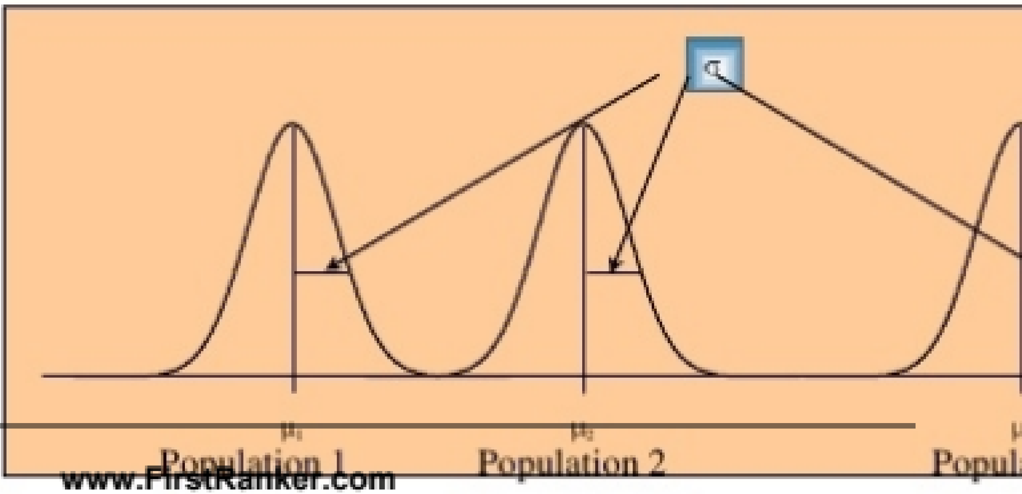
- The ANOVA is a powerful and com
procedure in the social sciences. It
variety of situations.

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 optio... compare means between more than tw...
- The problem with the multiple t-te... that as the number of groups increa... 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

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... =$$

- 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 p
  allows you to examine the effe
  independent variables concurr