Hypothesis Testing

Become Future Fit

You will learn

Review the basics of Hypothesis Testing &

understand the scope of this lesson

Level of Difficulty

Medium

Basics of Hypothesis Testing

Canopus Business Management Group

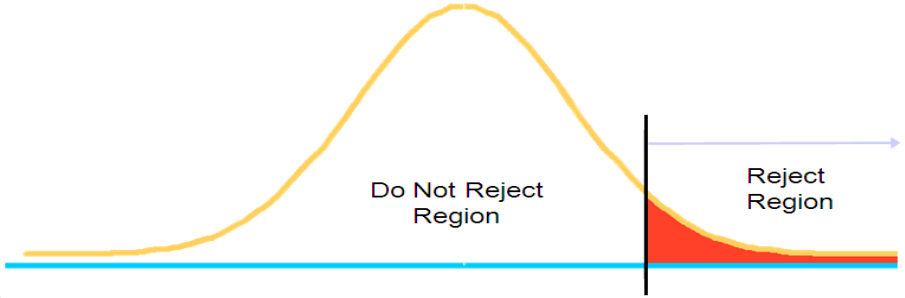
- Framing of Hypothesis Statements
- Statistical Vs Practical Significance
- Role of test statistic, critical statistic, P-value
- Errors associated with Hypothesis testing
- Selection of appropriate hypothesis tests
- Hypothesis testing procedure
- Hypothesis Tests – t Tests, ANOVA, Chi-square Tests & Proportions Tests

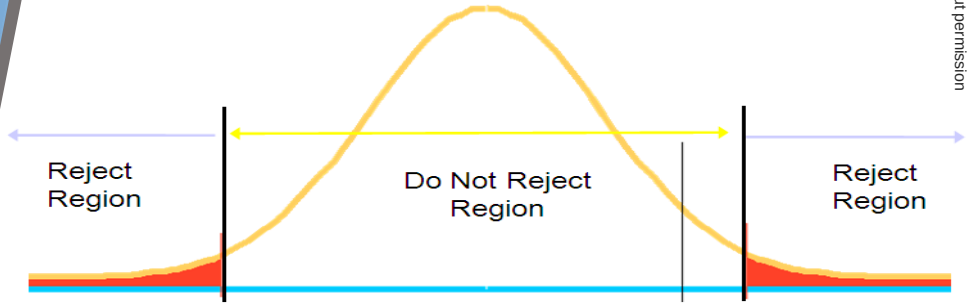
Practical scenarios & intricacies of performing hypothesis tests

- 1-t, 2-t, Paired-t
- 1-way ANOVA, 2-way ANOVA, Balanced & Unbalanced Designs, General Linear Model
- Types of Chi-squares Tests
- Proportions tests

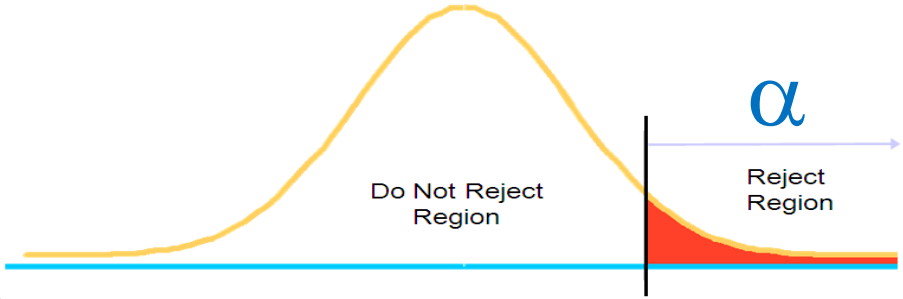
One Vs Two-Sided Tests

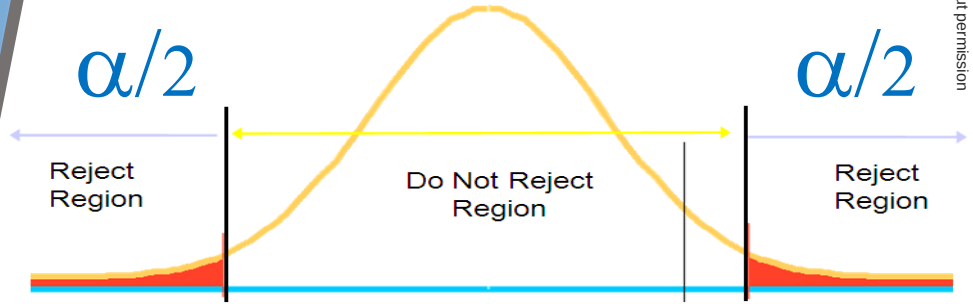
One Vs Two-Sided Tests





One Vs Two-Sided Tests





One-Sided vs Two-Sided Tests

A test that is concerned only with whether the difference is greater than orless than (but not both) is called a one-sided test. We can put all the a risk in one tail.

If the test involves inequality, then the test is called two sided. We have to split the a risk between both tails. In such cases, the Za value is substituted by Za/2 . There is no change in b or Zb. This applies to the t distribution as well.

Most software automatically halve the a if we select the greater than or less than option for the alternative hypothesis.

Performing t-Tests III

Become Future Fit

You will learn

Learn how to perform 1t, 2t & Paired-t tests

Contd. from previous lecture....

Level of Difficulty

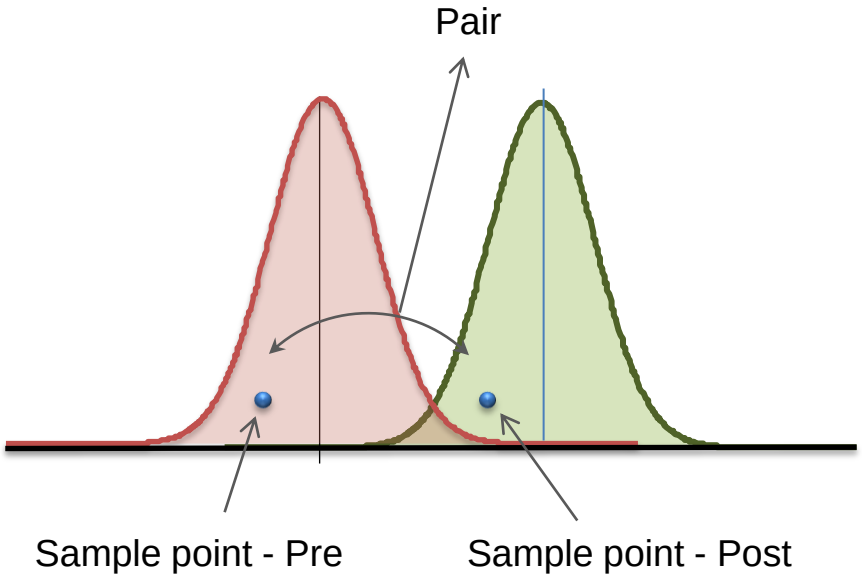
High

Weight Improvement

Students of 5 grade were given dietary complement for 3 months and their weights were compared.

Is there any positive impact of dietary complement?

Paired t test



- **Null hypothesis:** Difference between each pair = 0
- **Alternate hypothesis:** Difference between each pair < 0

ANOVA Fundamentals I

Become Future Fit

Objective

Learn about theory behind ANOVA testing

Level of Difficulty

High

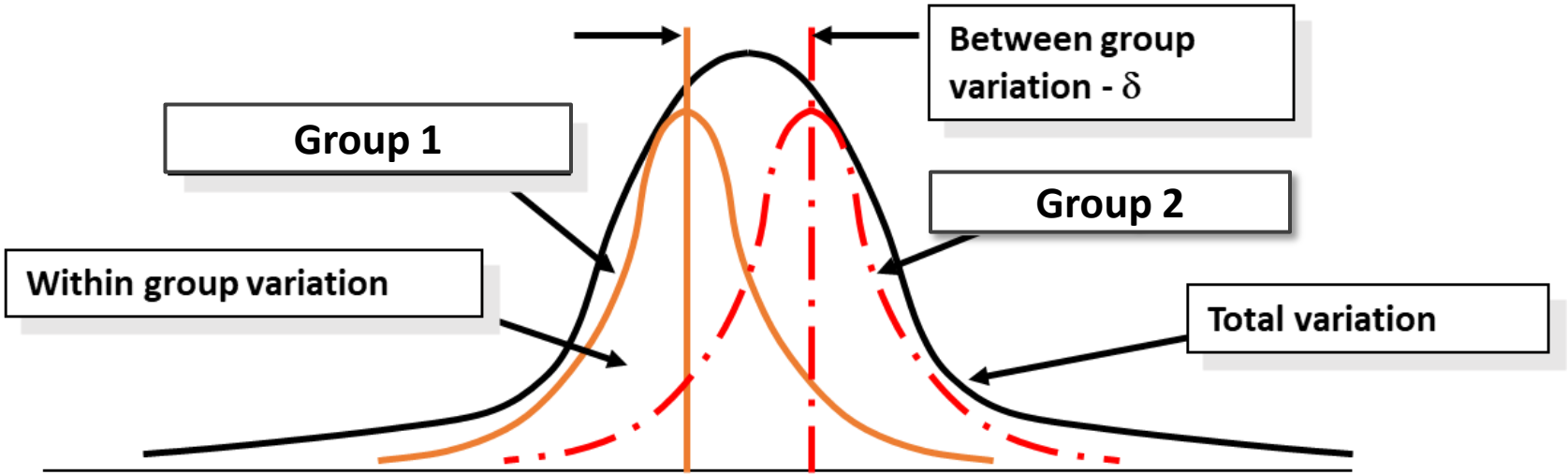
Analysis of Variance

Used to study the differences between means based on the variation between two or more groups

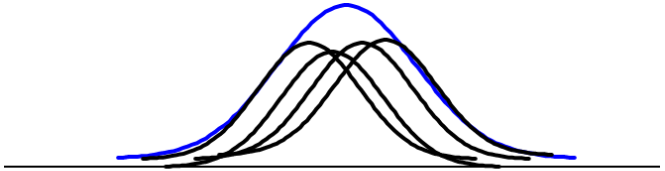
$$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Between & Within Variation

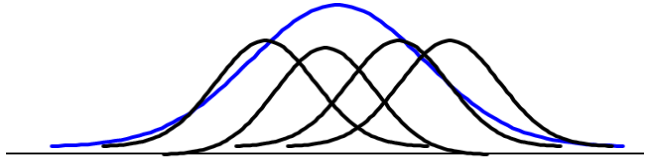
Rational Subgroups



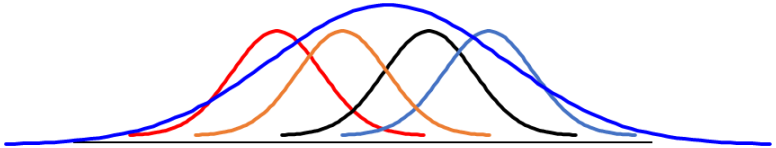
Between Vs Within Sample



Scenario 1



Scenario 2



Scenario 3

Signal-to-Noise Ratio

Nah nah nah I can't hear you!

Analysis of Variance

Total Variation can be split into two parts:

Total = Between + Within

SST = SSB + SSW

SST = Total Sum of Squares

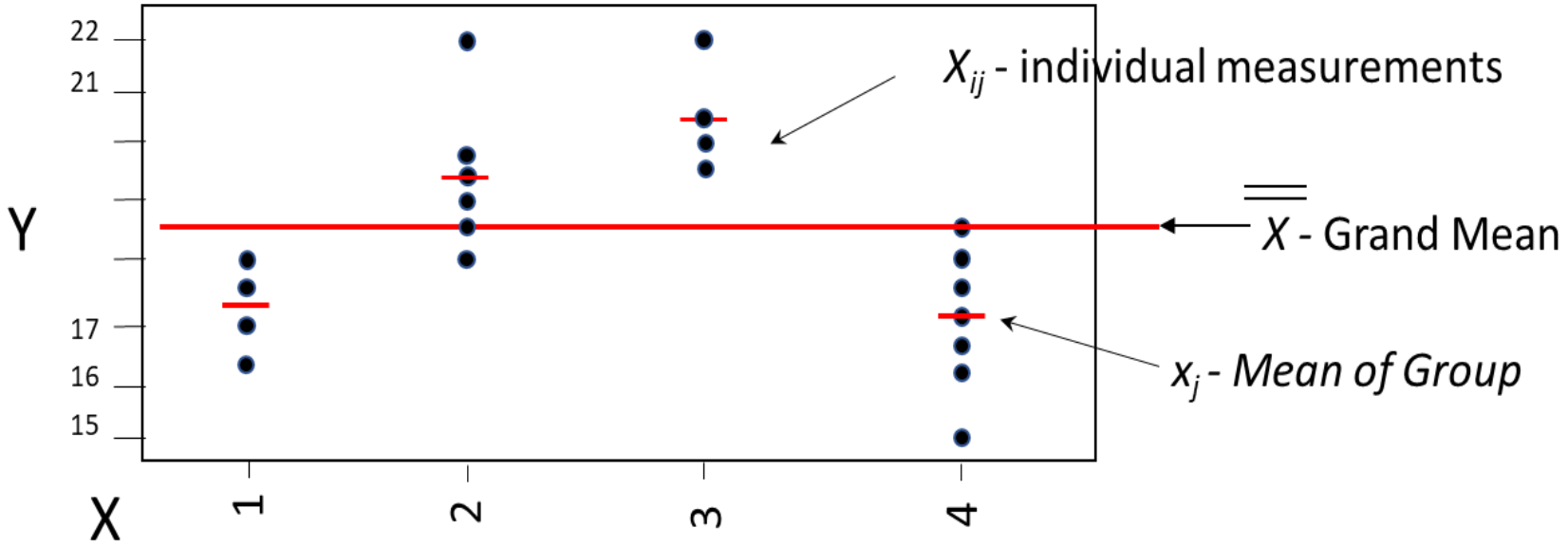
(Total Variation)

SSB = Sum of Squares Between Groups

(Between - Group Variation)

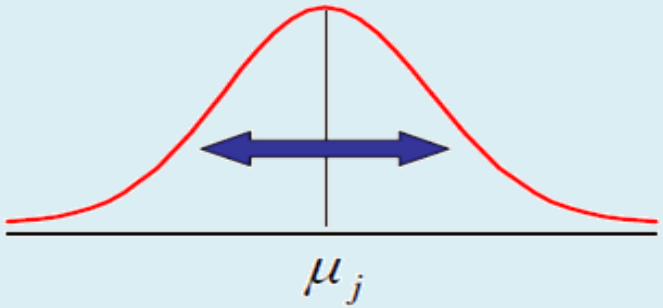
SSW = Sum of Squares Within Groups

(Within - Group Variation)



Sum of Squares (Within)

Summing the variation within each group and then adding over all groups



$$SSW = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2$$

SSW = Sum of squares within groups

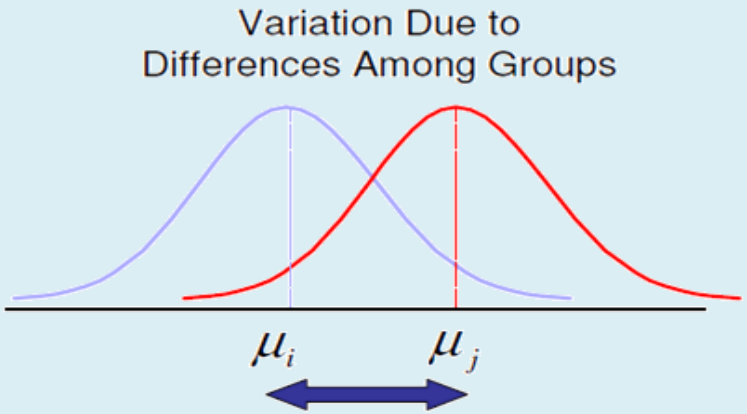
c = number of groups

n_j = sample size from group j

$\overline{X}_j$ = sample mean from group j

X_{ij} = i^{th} observation in group j

Sum of Squares (Among or Between)



$$SSB = \sum_{j=1}^{c} n_j(\bar{X}_j - \bar{\bar{X}})^2$$

SSB = Sum of squares between groups

c = number of groups

n_j = sample size from group j

$\bar{X}_j$ = sample mean from group j

$\bar{\bar{X}}$ = grand mean (mean of all data values)

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 = \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{x})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

SS(Total) = SS(Between) + SS(Within)

The Sums of Squares equation provides the method for calculating the SS values.

Source of Variation	SS	Df	MS (Variance)	F ratio
Between Groups	SSB	c-1	$MSB = \frac{SSB}{c-1}$	$F = \frac{MSB}{MSW}$
Within Groups	SSW	n - c	$MSW = \frac{SSW}{n-c}$	
Total	SST= SSB +SSW	n - 1		

c = number of groups

n = sum of the sample sizes from all groups

df = degrees of freedom

Performing ANOVA & GLM III

Become Future Fit

Objective

Learn to apply 1 way ANOVA, 2 way-balanced ANOVA & 2 way –unbalanced ANOVA (GLM) and interpret the results

Contd. from previous lecture

Level of Difficulty

High

Factors Impacting Quality Scores

The quality scores are sampled for different types of transactions across months. Is there an impact of these factors?

General Linear Model

General Linear Model (GLM) is the called so as it can be used in all the possibilities:

- Two or more than two factors
- Variables and Factors
- Balanced or unbalanced
- Nested or un-nested

Group Activity 2 – ANOVA (ME)

The procurement team wants to use data to finalize their strategy to achieve the budget for the year.

Procurement Managers have worked out 4 different approaches for few parts and estimated the potential saving per part.

They haven't taken all parts as it is not practical for initial stage.

Is there a distinct strategy evolving from this data?

File: Procurement strategy.mtw

Performing Chi-square Tests

Become Future Fit

You will learn

How to perform different types of chi-square tests and draw suitable inferences?

Level of Difficulty

High

Impact of Gender

We want to know if customer satisfaction scores obtained after delivering the product is impacted by the gender of the customers.

Are customers of either gender are biased towards product satisfaction?

Factors impacting Employee Attrition

CEO of a company wants to know if performance rating and commute distance to office are key driving factors for employees to leave the organization.

Performing Proportions Tests II

Become Future Fit

You will learn

Learn how to perform 1P & 2P tests

Contd. from previous lecture

Level of Difficulty

High

- Recently hired manager is poor in management resulting in higher errors
- Last June, out of 425 processed documents, 26 had errors
- This June, out of 266 documents, 31 had errors