Hypothesis Testing

Start with Hypothesis about a Population Parameter

Collect Sample Information

Reject/Do Not Reject Hypothesis

	Ho is TRUE	H1 is TRUE
Fail to Reject Ho	Right Decision Confidence 1-alpha	Type II error beta
Reject Ho	Type I error alpha	Right Decision Power 1-beta

The factors that affect the power of a test include sample size, effect size, population variability, and α.Power and α are related as increasing α decreases β. Since power is calculated by 1 minus β, if you increase α,You also increase the power of a test. The maximum power a test can have is 1, whereas the minimum value is 0.

Slide - 48

Hypothesis testing case studies exercise:

Write Null and alternate Hypothesis for the following case studies:

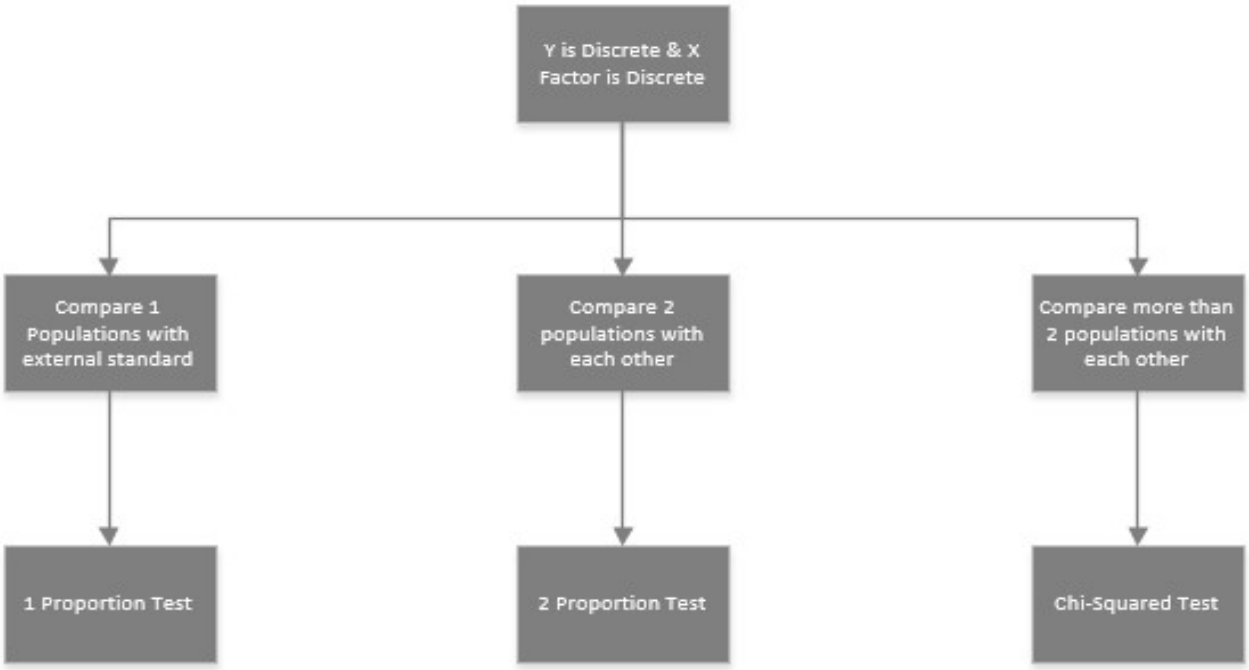
1) Our quality will not improve after the consulting project
2) The retail market will grow by 50% in the next 5 years
3) We will acquire 8,000 new customers if I open a store in this area
4) Less than 5% clients will default on their loans
5) We will need 400 more person hours to finish this project

6) Our potential customers do not spend more than 60 minutes on the web every day

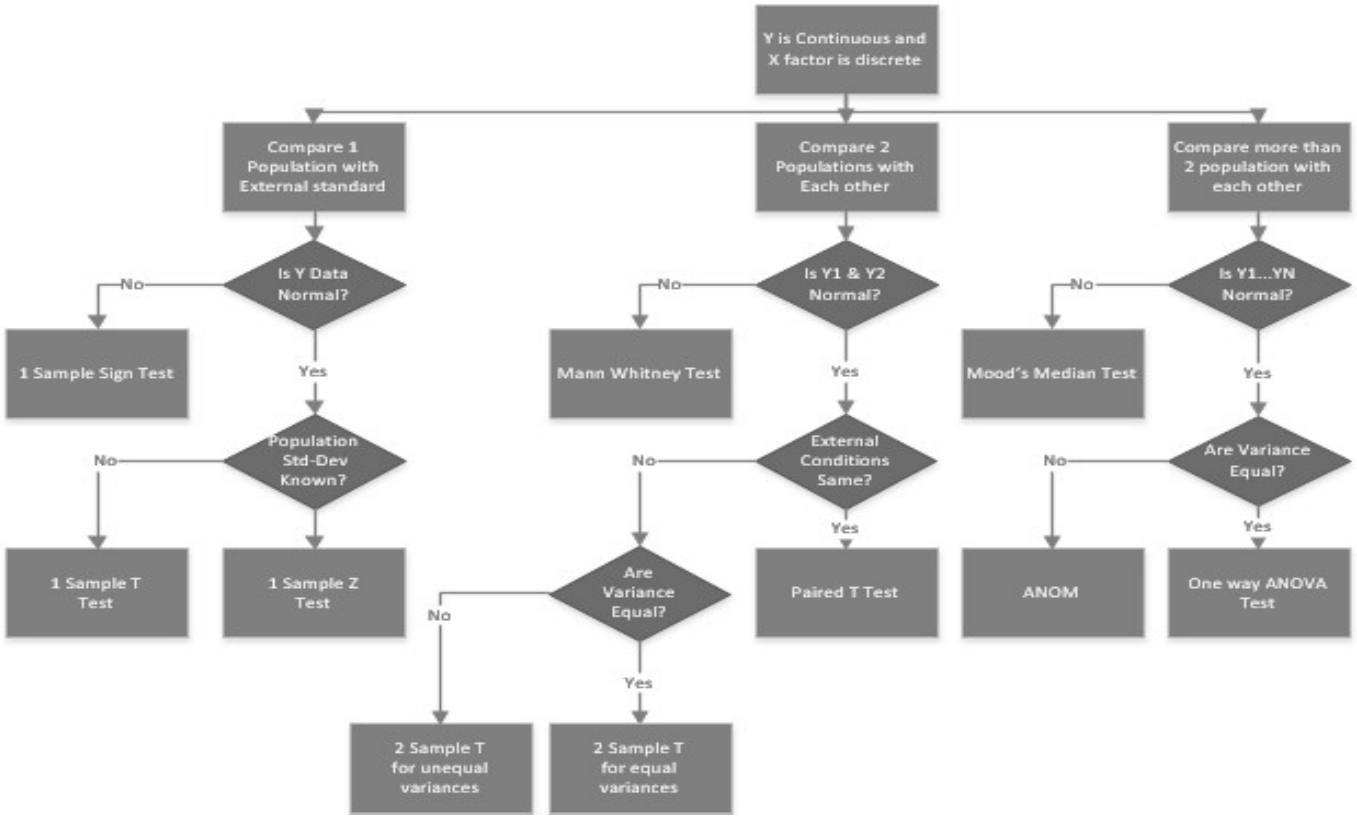
Slide-49

Different tests based on type of input and output data types and number of data types:

Y	X	Test
Continuous	Discrete in 2 categories	2 - Sample t test
Continuous	Discrete more than 2 categories	ANOVA – One Way
Discrete	Discrete in 2 categories	2 - Proportion test
Discrete	Discrete more than 2 categories	Chi-square test


Y is Discrete & X Factor is Discrete
Compare 1 Populations with external standard
Compare 2 populations with each other
Compare more than 2 populations with each other
1 Proportion Test
2 Proportion Test
Chi-Squared Test

SLIDE-50



SLIDE-51

1-Sample Z test:

Normality Test

Stat > Basic Statistics > Graphical Summary

Population Standard Deviation Known or Not

1 Sample Z Test

Stat > Basic Statistics > 1 Sample Z

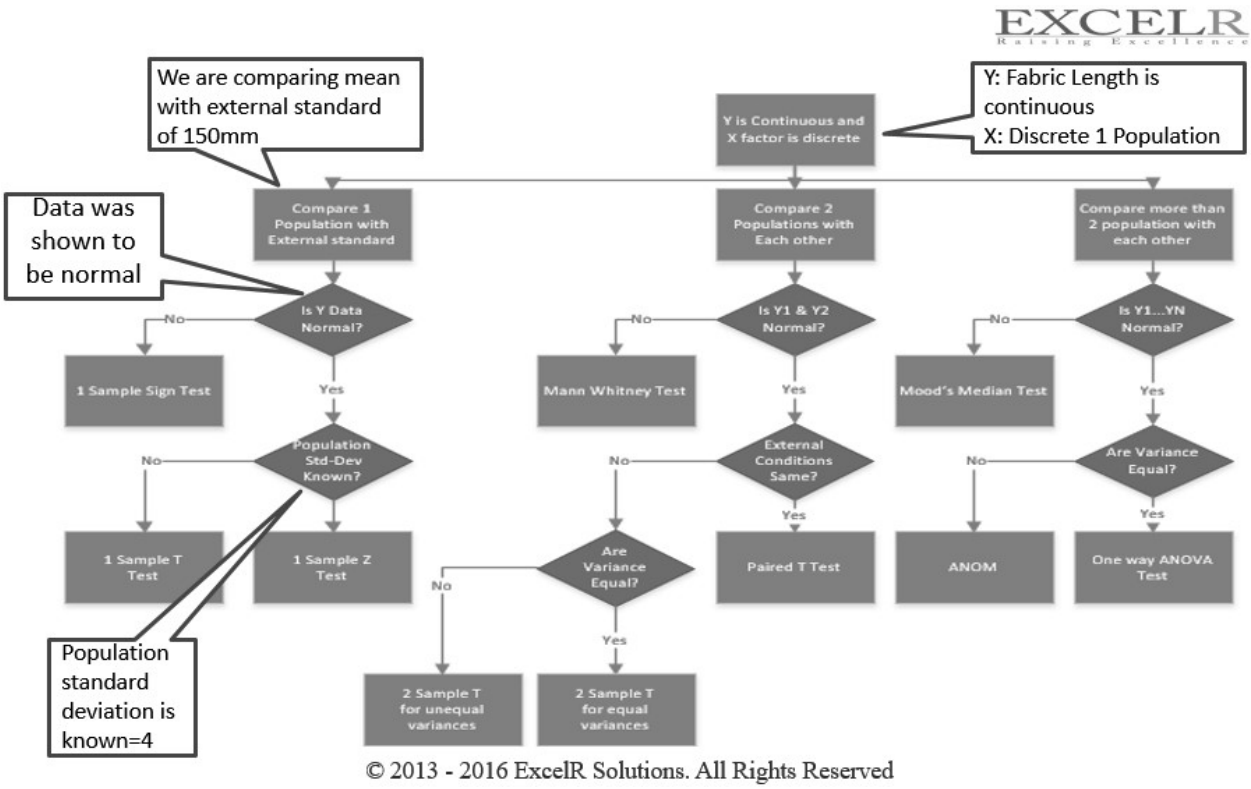
Fabric Data

The length of 25 samples of a fabric are taken at random. Mean and standard deviation from the historic 2 years study are 150 and 4 respectively. Test if the current mean is greater than the historic mean. Assume α to be 0.05

SLIDE-52

1-Sample Z test – Write Hypothesis

SLIDE-53



SLIDE-54

1-Sample t Test

- Normality Test

  Stat > Basic Statistics > Graphical Summary

- Population Standard Deviation Known or Not
- 1 Sample t Test

  Stat > Basic Statistics > 1 Sample t

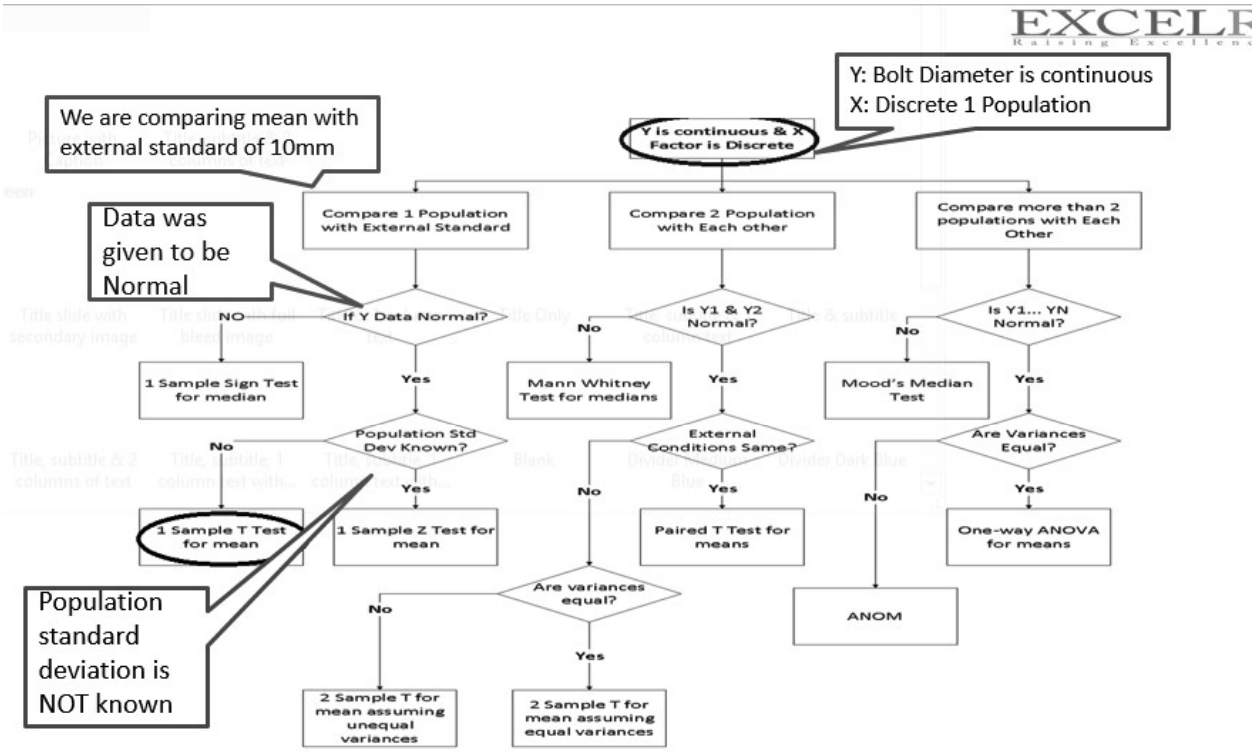
Bolt Diameter

The mean diameter of the bolt manufactured should be 10mm to be able to fit into the nut. 20 samples are taken at random from production line by a quality inspector. Conduct a test to check with 95% confidence that the mean is not different from the specification value.

SLIDE-55

1-Sample t Test – Write Hypothesis

SLIDE-56



SLIDE-57

1-Sample Sign Test

Normality Test

Stat > Basic Statistics > Graphical Summary

1 Sample Sign Test

Stat > Non Parametric >

1 Sample sign

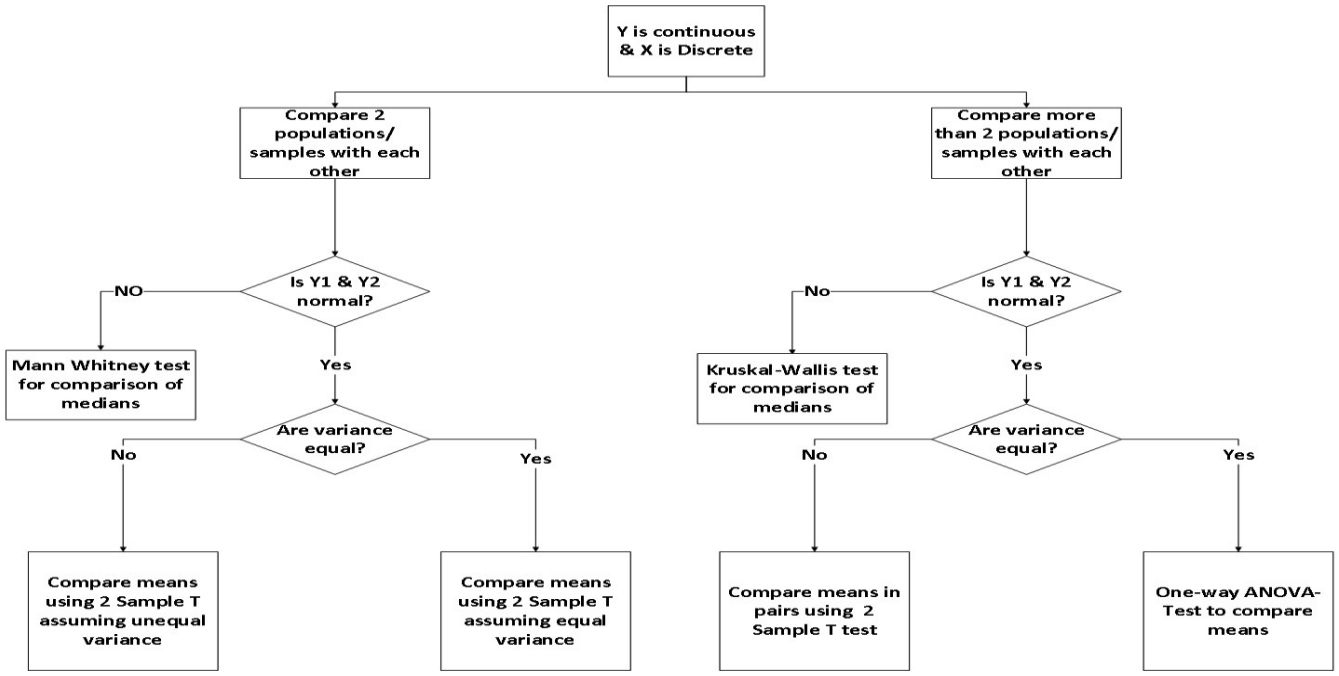
Student Scores

The scores of 20 students for the statistics exam are provided. Test if the current median is not equal to historic median of 82. Assume 'α' to be 0.05

SLIDE-58

1-Sample Sign Test – Write Hypothesis

Slide-59



Slide-60

2-Sample t-Test

- Normality test

  Stat > Basic Statistics > Graphical Summary
- 2 Variance Test

  Stat > Basic Statistics > 2 Variance
- 2 sample t- Test

  Stat > Basic Statistics > 2 sample t

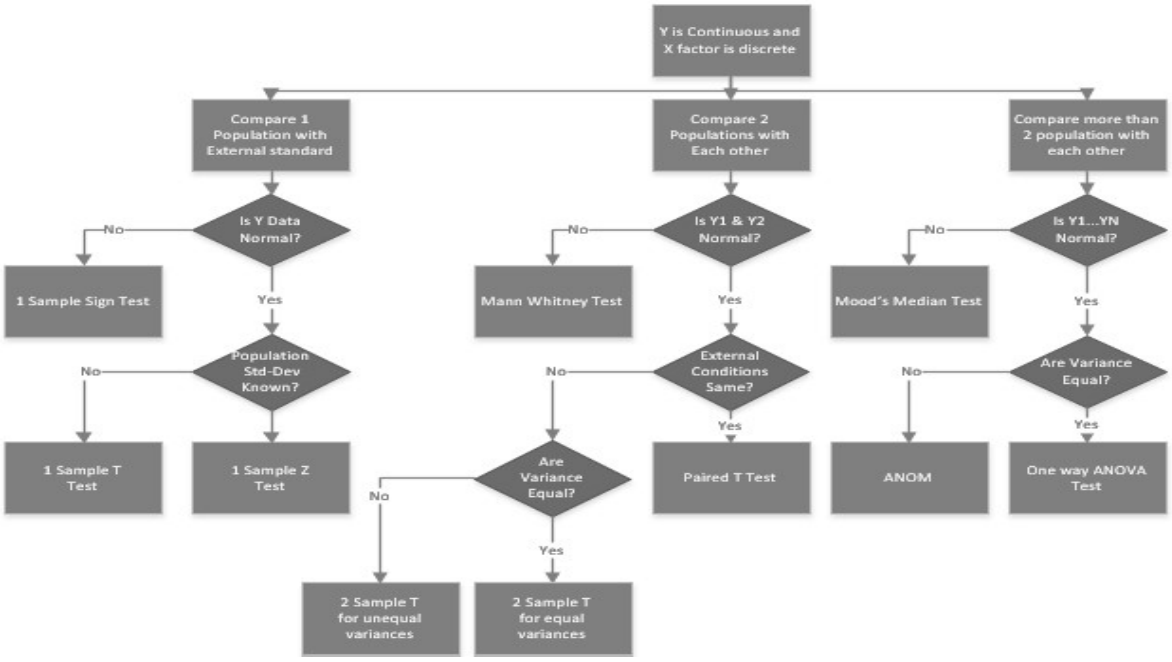
Marketing Strategy : case study

A financial analyst at a Financial institute wants to evaluate a recent credit card promotion. After this promotion, 450 cardholders were randomly selected. Half received an ad promoting a full waiver of interest rate on purchases made over the next three months, and half received a standard Christmas advertisement. Did the ad promoting full interest rate waiver, increase purchases?

SLIDE-61

2-Sample t Test – Write Hypothesis

SLIDE-62



Slide-63

Paired T Test

- This test is used to compare the means of two sets of observations when all the other external conditions are the same
- This is a more powerful test as the variability in the observations is due to differences between the people or objects sampled is factored out

Example: To find out if medication A lowers blood pressure

SLIDE-64

Trigger your thoughts!

Comparing the performance of machine A vs. machine B by feeding different raw materials to each machine	Compare the performance of machine A vs. machine B when the same raw material is fed to each machine
Compare the power output of two wind mills next to each other simultaneously when you use motor A on one wind mill and motor B on another	Compare the power output of a wind mill when you use motor A for 1 month and motor B for 1 month
Identifying resistor defects and capacitor defects in same PCB by collecting such data using 20 PCB units	Identifying resister defects on 20 PCB's and capacitor defects on 20 (different) PCB's

SLIDE-65

2-Sample t test or Paired T test

Effect of fuel additive on vehicles is being studied. Out of a total of 20 vehicles, 10 vehicles are chosen randomly and mileage is recorded. In rest of the 10 vehicles, additive to be tested is added with the fuel and their mileage is recorded. Find if the mileage increases by adding the fuel additive.

2-Sample t test

Assume the same data was recorded if only 10 vehicles were chosen and mileage was recorded before and after adding the additive. What method will you choose to find the result.

Paired T test

SLIDE-66

Mann-Whitney test

Normality Test

Stat > Basic Statistics > Graphical Summary

Mann – Whitney test for Medians

Stat > Non Parametric > Mann Whitney

Vehicle with

& without Additives

Effect of fuel additive on vehicles is being studied. Out of a total of 20 vehicles, 10 vehicles are chosen randomly and mileage is recorded. In rest of the 10 vehicles, additive to be tested is added with the fuel and their mileage is recorded. Find if the mileage increases by adding the fuel additive.

SLIDE-67

Mann-Whitney Test – Write Hypothesis

SLIDE-68

Paired T test

Normality Test

Paired T Test

Stat > Basic Statistic > Paired T

- Since the data was not normal, the cause of non-normality was investigated and it was found that the first data point for "with additive" was wrongly entered. This value should have been 20. Now, proceed with the rest of the analysis.
- If the data were truly non-normal our analysis would stop here.
- Vehicle with & without Additives

- Effect of fuel additive on vehicles is being studied. Out of a total of 20 vehicles, 10 vehicles are chosen randomly and mileage is recorded. In rest of the 10 vehicles, additive to be tested is added with the fuel and their mileage is recorded. Find if the mileage increases by adding the fuel additive. Assume the same data was recorded if only 10 vehicles were chosen and mileage was recorded before and after adding the additive.

SLIDE-69

Paired T test – Write Hypothesis

SLIDE-70

One-Way ANOVA :

Normality Test

- *Stat > Basic Statistics > Graphical Summary*

Variance Test

- *Stat > ANOVA > Test for Equal Variances*

ANOVA

Stat > ANOVA > One-Way….

Contract Renewal: Case Study

A marketing organization outsources their back-office operations to three different suppliers. The contracts are up for renewal and the CMO wants to determine whether they should renew contracts with all suppliers or any specific supplier. CMO want to renew the contract of supplier with the least transaction time. CMO will renew all contracts if the performance of all suppliers is similar.

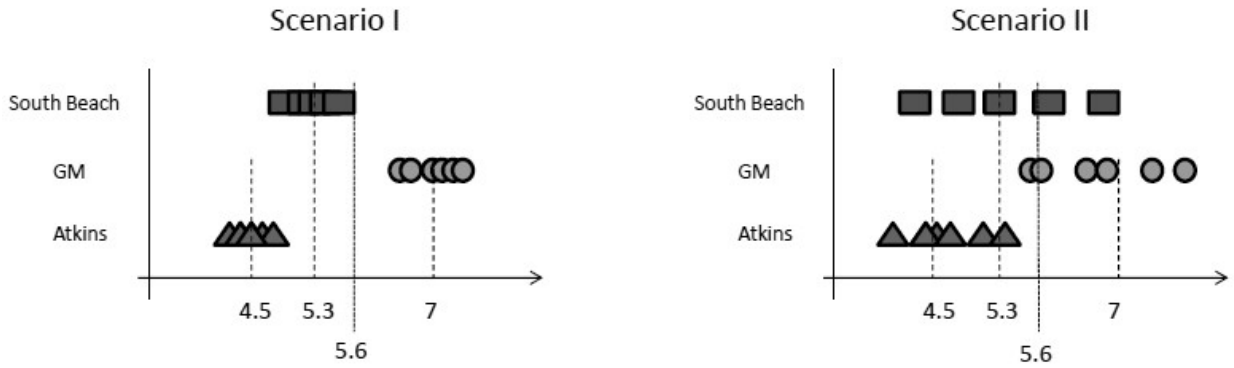
Slide-71

Example: More weight reduction programs

- Suppose the nutrition expert would like to do a comparative evaluation of three diet programs(Atkins, South Beach, GM)
- She randomly assigns equal number of participants to each of these programs from a common pool of volunteers
- Suppose the average weight losses in each of the groups(arms) of the experiments are 4.5kg, 7kg, 5.3kg
- What can she conclude?

Slide-72

Two kinds of variation matter



- Diet program having high variances for the sample weight losses
- Diet program having low variance for the sample weight losses

  Not every individual in each program will respond identically to the diet program

  Easier to identify variations across programs if variations within programs are smaller

  Hence the method is called Analysis of Variance(ANOVA)

Slide-73

Formalizing the intuition behind variations

- It should be obvious that for every observation : $Tot_{ij} = t_i + e_{ij}$
- What is more surprising and useful is:

Sum of squares total (SST) = $\sum_{i=1}^{r} \sum_{j=1}^{ni} Tot_{ij}^2$

Sum of squares Treatment (SSTR) = $\sum_{i=1}^{r} n_i t_i^2$

Sum of squares Error (SSE) = $\sum_{i=1}^{r} \sum_{j=1}^{ni} e_{ij}^2$

SST = SSTR + SSE

Slide-74

Statistically test for equality means

- n subjects equally divided into r groups
- Hypothesis
  - H0: $\mu 1 = \mu 2 = \mu 3 = \ldots = \mu r$
  - Not all μ_i are equal

Calculate

- Mean Square Treatment MSTR = SSTR / (r-1)
- Mean Square Error MSE = SSE / (n-r)
- The ratio of two squares f = MSTR/MSE
- Strength of this evidence p-value = Pr(F(r-1,n-r) ≥ f)

• Reject the null hypothesis if p-value < α

Slide-75

Analysis of variance(ANOVA)

- ANOVA can be used to test equality of means when there are more then 2 populations
- ANOVA can be used with one or two factors
- If only one factor is varying, then we would use a one-way ANOVA
  - Example: We are interested in comparing the mean performance of several departments within a company. Here the only factor is the name of department
  - If there are two factors, we would use a two way ANOVA. Example: One factor is department and the second factor is the shift.(day vs. Night)

SLIDE-76

Analysis of variance (ANOVA)

ONE-WAY ANOVA

Source of Variation	Sum of Squares (SS)	Degrees of Freedom	Mean Square (MS)	F Test Statistic
Between Treatments	SSFactor	K-1	MSFactor = SSFactor / DFFactor	F = MSFactor / MSError
Within Treatment	SSError	N-k	MSError = SSError / DFError	
Total	SSTotal	N-1		

Source of Variation	Sum of Squares (SS)	Degrees of Freedom	Mean Square (MS)	F Test Statistic
Factor A	SS_A	$n_A - 1$	$MS_A = SS_A / (n_A - 1)$	$F_A = MS_A / MS_E$
Factor B	SS_B	$n_B - 1$	$MS_B = SS_B / (n_B - 1)$	$F_B = MS_B / MS_E$
Interaction A * B	SS_{AB}	$(n_A - 1)(n_B - 1)$	$MS_{AB} = SS_{AB} / (n_{AB} - 1)$	$F_{AB} = MS_{AB} / MS_E$
Error	SS_E	$n - n_A * n_B$	$MS_E = SS_E / (n - n_A * n_B)$	
Total	SS_T	n - 1		

Slide-77

Dichotomies

2-Sample t test

Is the Transaction time dependent on whether person A or B processes the transaction?

Is medicine 1 effective or medicine 2 at reducing heart stroke?

Is the new branding program more effective in increasing profits?

ANOVA – One-Way

Does the productivity of employees vary depending on the three levels? (Beginner, Intermediate and Advanced)

Three different sale-closing methods were used. Which one is most effective?

Four types of machines are used. Is weight of the Rugby ball dependent on the type of machine used?

Slide-78

Non-Parametric equivalent to ANOVA

- When the data are not normal or if the data points are very few to figure out if the data are normal and we have more than 2 populations, we can use the Mood's Median or Kruskal Wallis test to compare the populations

H_o: All the medians are the same

H_a: One of the medians is different

- Mood's median assigns the data from each population that is higher than the overall median to one group, and all points that are equal or lower to another group. It then uses a Chi-Square test to check if the observed frequencies are close to expected frequencies
- Kruskal Wallis is another test that is non-parametric equivalent of ANOVA. Kruskal Wallis is the extension of Mann-Whitney test

Slide-79

Mood's Median & Kruskal Wallis

Mood's Median – handles outliers well

- *Stat > Nonparametric > Mood's Median*

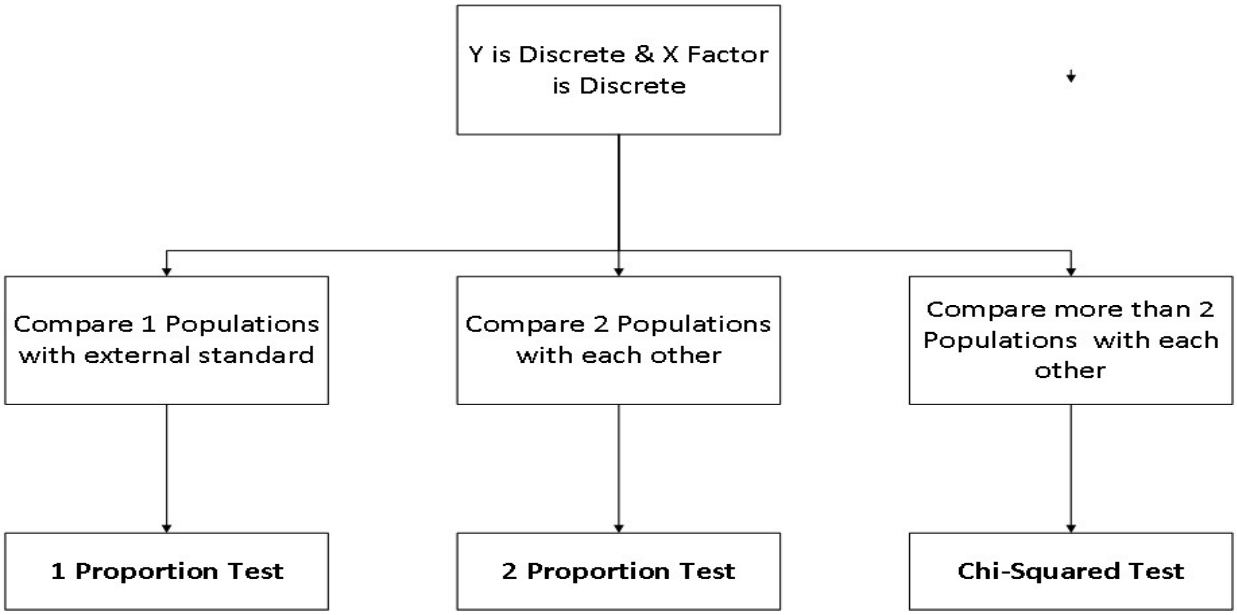
Kruskal Wallis – more powerful than Mood's Median

- *Stat > Nonparametric > Kruskal Wallis*

Height Growth:

Growth is measured for three treatments as shown in the case study. Compare the effect of the three treatments on growth.

Slide-80



Slide-81

1-Proportion Test

Stat > Basic Statistics > 1-Proportion

Football Coach

- The people carry out a poll to find the acceptability of new football coach. It was decided that if the support rate for the

coach for the entire population was truly less than 25%, the coach would be fired

- 2000 people participated and 482 people supported the new coach
- Conduct a test to check if the new coach should be fired with 95% level of confidence.

Slide-82

2-Proportion Test

HO	Proportion A = Proportion B	Check p-value
Ha	Proportion A NOT = Proportion B	If p-value < alpha, we reject Ho

Johnnie Talkers : case study

Johnnie Talkers soft drinks division sales manager has been planning to launch a new sales incentive program for their sales executives. The sales executives felt that adults (>40 yrs) won't buy, children will & hence requested sales manager not to launch the program. Analyze the data & determine whether there is evidence at 5% significance level to support the hypothesis

Slide-83

Chi-Square Test

- *How can you determine whether the distribution of defects in your product or service has changed from the historic distribution over time, or exceeds an industry standard*

- Do you think mean is more significant or variance?
  - *Comparing population's variance to a standard value involves calculating the*

Chi-square test statistic

We can also:

- *Determine whether one variable is dependent over another*

$$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$$

- *Comparing observed & expected frequencies where variance is unknown.*

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

This is called as goodness-of-fit test

- *Compare multiple proportions*

Slide-84

Chi-Square Goodness-of-fit test

Goodness-of-fit test is to test assumptions about the distributions that fit the process data

Are observed frequencies (O) same or different from historical, expected or theoretical frequencies (E)?

If there's a difference between them, this suggests that the distribution model expressed by the expected frequencies does not fit the data

Slide-85

- A city has a newly opened nuclear plant, and there are families staying dangerously close to the plant. A health safety officer wants to take this case up to provide relocation for the families that live in the surrounding area. To make a strong case, he wants to prove with numbers that an exposure to radiation levels is leading to an increase in diseased population. He formulates a contingency table of exposure and disease.
- Does the data suggest an association between the disease and exposure?

	Disease		Total
Exposure	Yes	No	
Yes	37	13	50
No	17	53	70
Total	54	66	120

Slide-86

Calculate the number of individuals of exposed and unexposed groups expected in each disease category (yes and no) if the probabilities were the same

If there were no effect of exposure, the probabilities should be same and the chi-squared statistic would have a very low value.

Proportion of population exposed = (50/120) = 0.42

Proportion of population not exposed = (70/120) = 0.58

Thus, expected values:

Population with disease = 54

Exposure Yes : 54 * 0.42 = 22.5

Exposure No : 54 * 0.58 = 31.5

Population without disease = 66

Exposure Yes : 66 * 0.42 = 27.5

Exposure No : 66 * 0.58 = 38.5

Slide-87

- Calculate the Chi-squared statistic

χ^2

$$= \frac{(\sum (\text{observed frequency} - \text{expected frequency})^2)}{\text{Expected frequency}}$$

$$= \frac{(37-22.5)^2}{22.5} + \frac{(13-27.5)^2}{27.5} + \frac{(17-31.5)^2}{31.5} + \frac{(53-38.5)^2}{38.5}$$

- Calculate the degrees of freedom :

(Number of rows – 1) X (Number of columns – 1)

df = (2 – 1) X (2 – 1) = 1

- Calculate the p-value from the Chi-squared table

For chi-squared value 29.1 and degrees of freedom = 1, from the table, p-value is < 0.001

- Interpretation: There is 0.001 chance of obtaining such discrepancies between expected and observed values if there is no association

- Conclusion : There is an association between the exposure and disease

Slide-88

Chi-Square Test

HO	All proportions are equal	Check p-value
Ha	Not all proportions are equal	If p-value < alpha we reject Ho

Bahaman Research: case study

Baha ManTech Research Company uses 4 regional centers in South Asia (India, China, Srilanka and Bangladesh) to input data of questionnaire responses. They audit a certain % of the questionnaire responses versus data entry. Any error in data entry renders it defective. The chief data scientist wants to check whether the defective % varies by country. Analyze the data at 5% significance level and help the manager draw appropriate inferences. [‘1’ means not defectives & ‘0’ means defective]