LEAN SIX SIGMA DEFINE PHASE

"There is no substitute for knowledge." - DR.William Edwards Deming

Coverage:
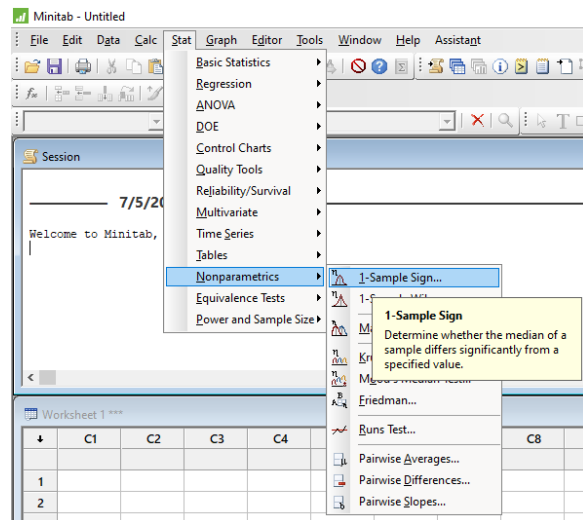
➢ Non parametric test for non normal data

➢ Other discrete hypothesis testing tools

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non - Normal
Equal Variance (More than 1 sample data set)	

1 Sample Sign test

Determine whether the median of a sample differs significantly from a specified value

Business Scenario: Admin manager of a MSME company want to check if the fuel efficiency of the new truck is meeting the number committed by the truck maker (6KM per litre)

Minitab exercise

Refer : 1 Sample Sign test

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non - Normal
Equal Variance (More than 1 sample data set)	

1 Sample Sign test

Null hypothesis Median of fuel efficiency = 6 KM/L

Alternative hypothesis Median of fuel efficiency < 6 KM/L
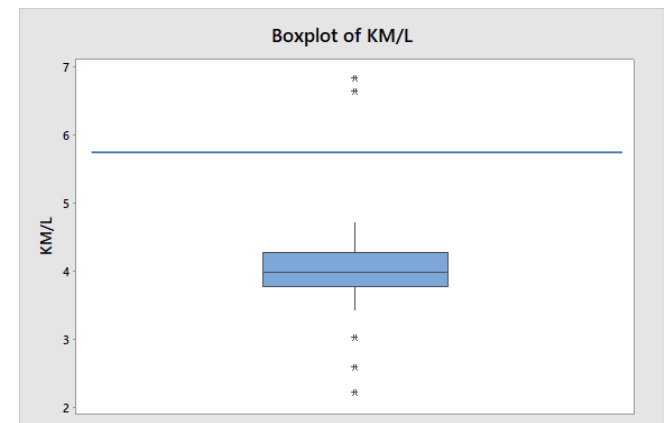
Sign test of median = 6.000 versus ≠ 6.000

	N	Below	Equal	Above
KM/L	30	28	0	2

P	Median
0.0000	3.986

Boxplot of KM/L

P-Value < 0.05 go with Alternative hypothesis
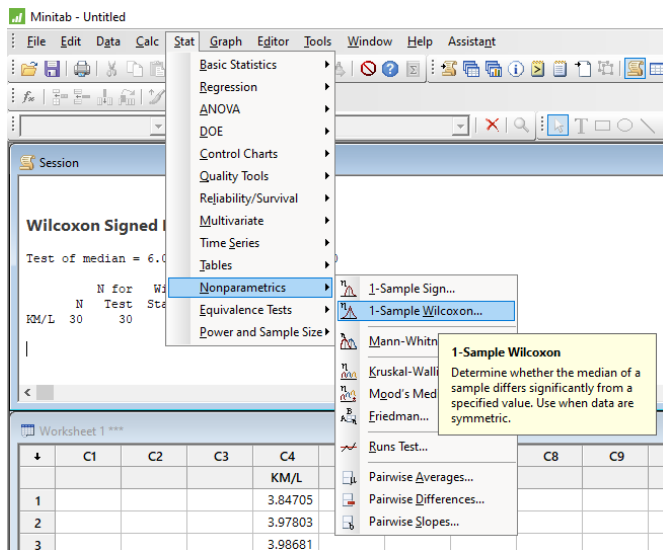P-Value > 0.05 go with Null hypothesis

Hypothesis testing - MEDIAN

1 Sample Wilcoxon test

Determine whether the median of a sample differs significantly from a specified value. Used When the data follow a symmetric distribution

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal (but symmetric)
Equal Variance (More than 1 sample data set)	

Business Scenario: Admin manager of a MSME company want to check if the fuel efficiency of the new truck is meeting number committed by the truck maker 6KM per litre

Minitab exercise
Refer : 1 Sample Sign test

Hypothesis testing - MEDIAN

1 Sample Wilcoxon test

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal (but symmetric)
Equal Variance (More than 1 sample data set)	

Null hypothesis Median of fuel efficiency = 6 KM/L
Alternative hypothesis Median of fuel efficiency < 6 KM/L
Wilcoxon Signed Rank Test: KM/L

Test of median = 6.000 versus median ≠ 6.000

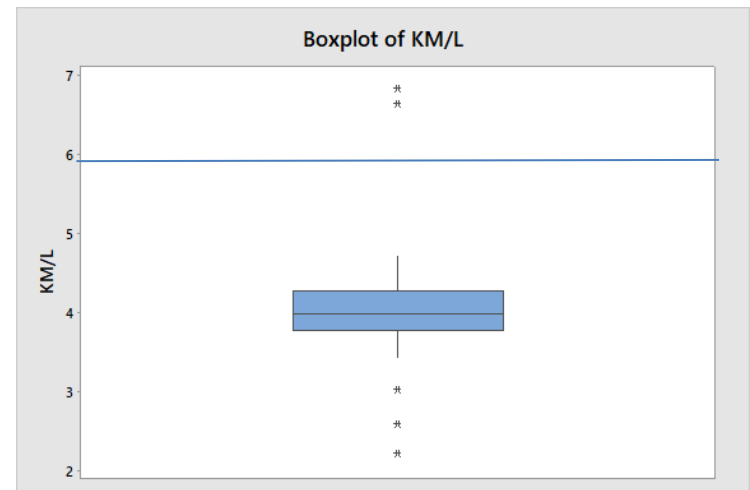|	N	N for Test	Wilcoxon Statistic	P	Estimated Median
KM/L	30	30	3.0	0.000	4.009

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

Boxplot of KM/L

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal (but Similar)
Equal Variance (More than 1 sample data set)	

Mann Whitney test

Determine whether the median two groups differ when the data for both the groups have similar distribution
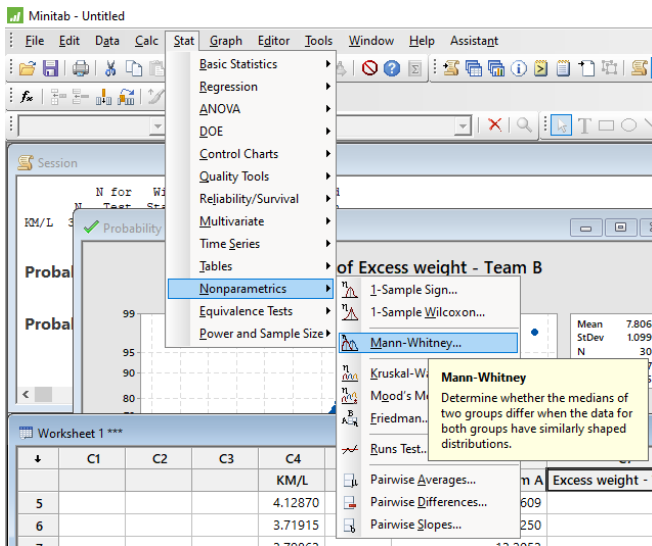
Business Scenario: Packing department supervisor in chemical company need to check if the excess weight for a 300 KG chemical bag packed by team A and team B are same

Minitab exercise

Refer : Mann Whitney test

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal (but Similar)
Equal Variance (More than 1 sample data set)	

Mann Whitney test

Null hypothesis : Median Excess weight - Team A = Median Excess weight – Team B
Alternative hypothesis : Median Excess weight - Team A ≠ Median Excess weight – Team B

 N Median
Excess weight - Team A 30 14.070
Excess weight - Team B 30   7.623

Point estimate for $\eta_1 - \eta_2$ is 6.095
95.2 Percent CI for $\eta_1 - \eta_2$ is (5.581,6.577)
W = 1364.0
Test of $\eta_1 = \eta_2$ vs $\eta_1 \neq \eta_2$ is significant at 0.0000

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis

Boxplot of Excess weight - Team A, Excess weight - Team B

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal & No Similar
Equal Variance (More than 1 sample data set)	
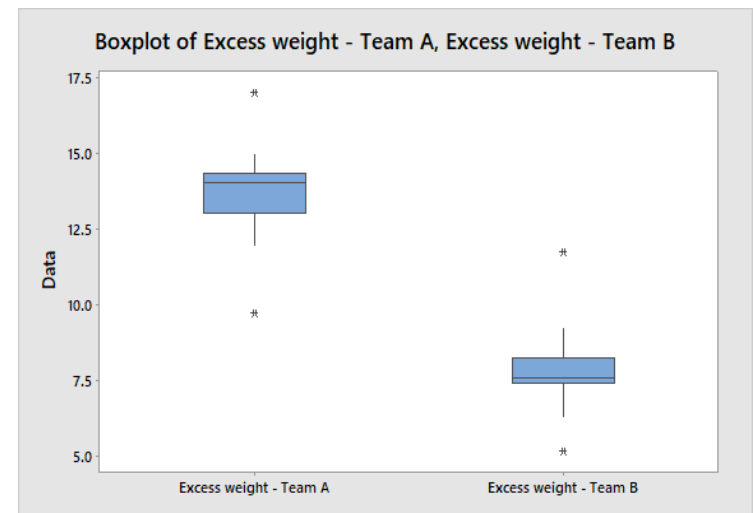
Moods Median test

Determine whether the median two or more groups differ when you are not confident if groups have similar distribution
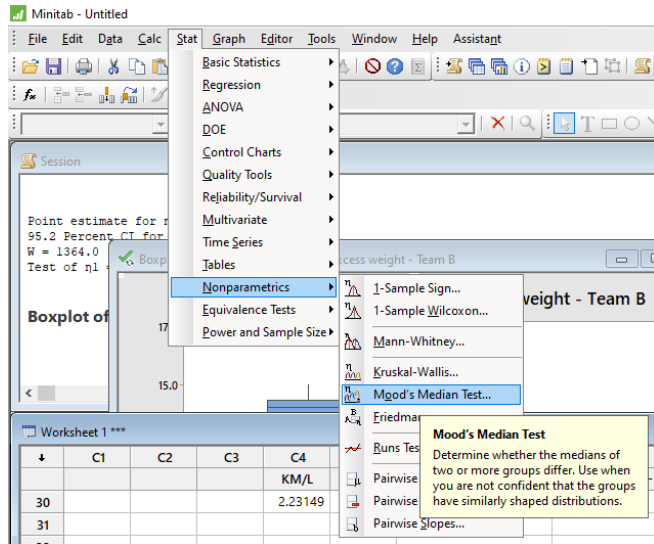
Business Scenario: Maintenance supervisor of a commercial facility want to check if the CO_2 emission of three brands of diesel generators in the facility are same.

Minitab exercise

Refer : Moods Median test

Hypothesis testing - MEDIAN

Pre Requisites	
Data Type	Continuous
Distribution	Non – Normal & No Similar
Equal Variance (More than 1 sample data set)	

Moods Median test

Null hypothesis : Median Co2 emission of all 3 gen-sets are same

Alternative hypothesis : Median Co2 emission at lest 1 gen-sets is different

Mood median test for Co2 - KG/Gal

Chi-Square = 43.47   DF = 2   P = 0.000

Gen-set	N≤	N>	Median	Q3-Q1
Gen-set A	18	12	13.28	0.68
Gen-set B	26	4	12.72	1.01
Gen-set C	1	29	15.79	0.96

```
Individual 95.0% CIs
------+---------+---------+---------+
           (--*)
 (--*--)
                          (---*-)
------+---------+---------+---------+
   13.0      14.0      15.0      16.0
```


Boxplot of Gen-set A, Gen-set B, Gen-set C
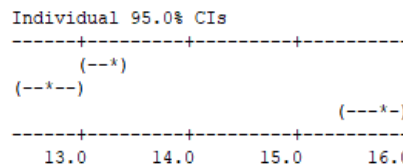
Overall median = 13.40

P-Value < 0.05 go with Alternative hypothesis
P-Value > 0.05 go with Null hypothesis
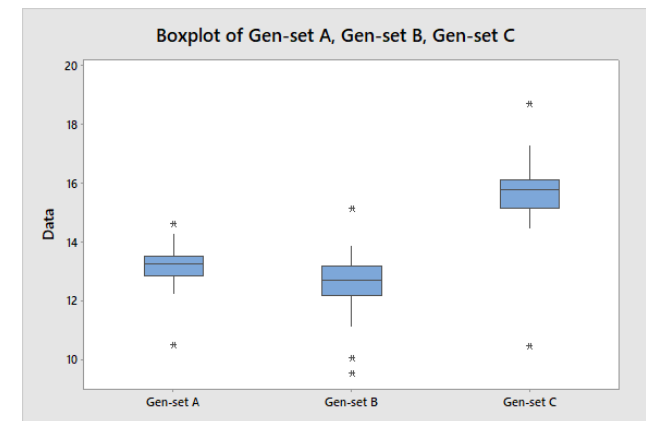
Statistics of interest : Proportion

Chi - Square Goodness of fit

Determine whether the sample Outcomes of multiple events meet target proportion

Pre Requisites	
Data Type	Discrete Ordinal or Nominal
Distribution	Chi- square

Assistant – Hypothesis Tests

Choose a Hypothesis Test

What is your objective?

Compare one sample with a target | Compare two samples with each other | Compare more than two samples

Help Me Choose | Help Me Choose | Help Me Choose

PERFORM A TEST
- μ 1-Sample t
- σ 1-Sample Standard Deviation
- p 1-Sample % Defective
- Chi-Square Goodness-of-Fit

Click to perform this test.

PERFORM A TEST
- μ 2-Sample t
- μ-μ Paired t
- σ 2-Sample Standard Deviation
- p 2-Sample % Defective
- Chi-Square Test for Association

PERFORM A TEST
- μ One-Way ANOVA
- σ Standard Deviations Test
- p Chi-Square % Defective
- Chi-Square Test for Association

Close

Business Scenario: Trainer vinoth has done 3 pilot batches for a training company and ad per the company policy participant feed back should be in below proportion to confirm vinoth as a regular trainer. As a training manager what will you do

Outcome	Target Percent
Rating 1	10
Rating 2	20
Rating 3	25
Rating 4	30
Rating 5	15

Minitab exercise

Refer : Chi-Square GOF

Statistics of interest : Proportion

Chi - Square Goodness of fit
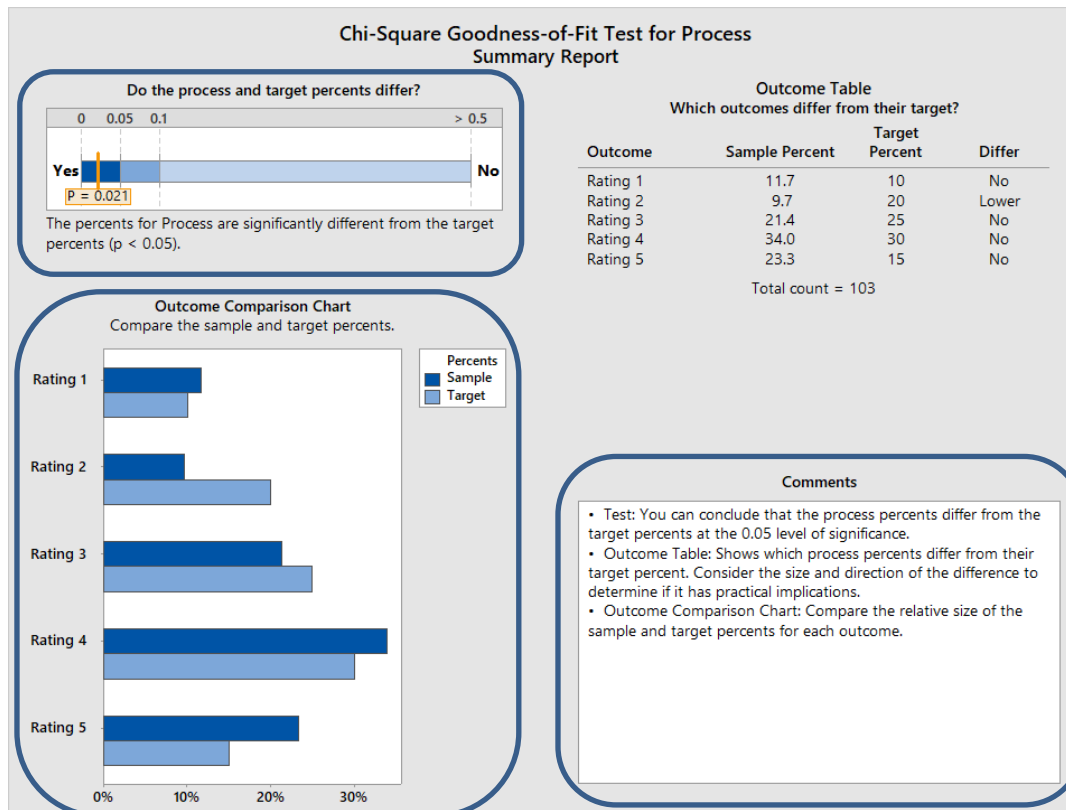

Chi-Square Goodness-of-Fit Test for Process
Summary Report

Do the process and target percents differ?

| 0 | 0.05 | 0.1 | | | > 0.5 |

Yes ... No

P = 0.021

The percents for Process are significantly different from the target percents (p < 0.05).

Outcome Table
Which outcomes differ from their target?

Outcome	Sample Percent	Target Percent	Differ
Rating 1	11.7	10	No
Rating 2	9.7	20	Lower
Rating 3	21.4	25	No
Rating 4	34.0	30	No
Rating 5	23.3	15	No

Total count = 103

Outcome Comparison Chart
Compare the sample and target percents.

Percents
■ Sample
■ Target

Comments

• Test: You can conclude that the process percents differ from the target percents at the 0.05 level of significance.
• Outcome Table: Shows which process percents differ from their target percent. Consider the size and direction of the difference to determine if it has practical implications.
• Outcome Comparison Chart: Compare the relative size of the sample and target percents for each outcome.

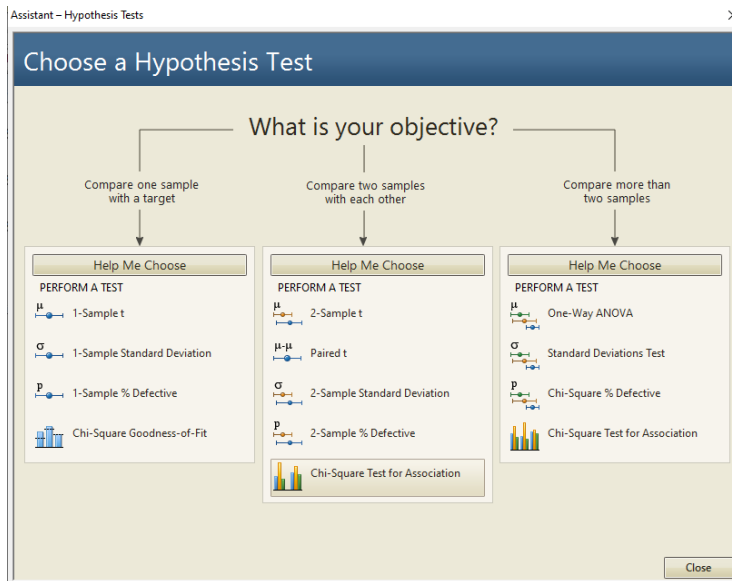| **Pre Requisites**	
Data Type	Discrete Ordinal or Nominal
Distribu tion	Chi- square

Statistics of interest : Proportion

Chi - Square test of association

The chi-square test for association determines whether the percentage of items in each outcome category is significantly different for two samples.

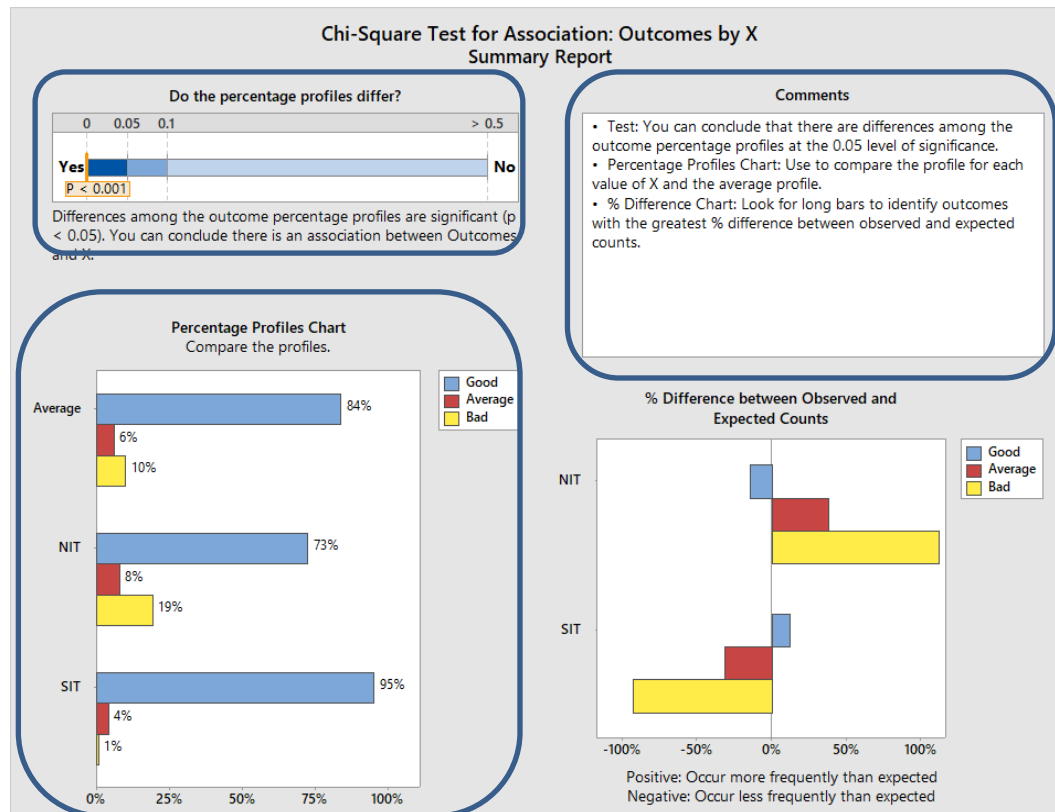| Pre Requisites	
Data Type	Discrete Ordinal or Nominal
Distribution	Chi- square

Business Scenario: Compare the Good, Average and Bad ratings of North Indian Thali and south Indian Thali in a restaurant

Minitab exercise
Refer : Chi-Square GOF

Statistics of interest : Proportion

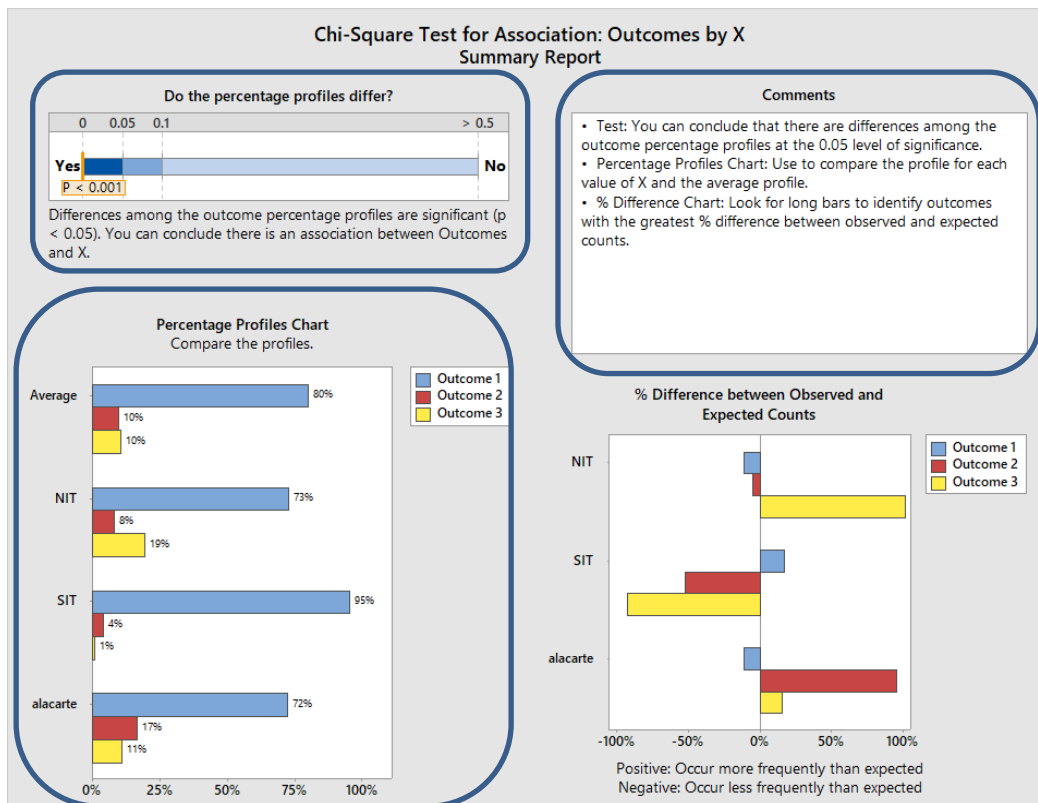## Chi - Square Goodness of fit

Chi-Square Test for Association: Outcomes by X
Summary Report

Do the percentage profiles differ?

Differences among the outcome percentage profiles are significant (p < 0.05). You can conclude there is an association between Outcomes and X.

Comments
- Test: You can conclude that there are differences among the outcome percentage profiles at the 0.05 level of significance.
- Percentage Profiles Chart: Use to compare the profile for each value of X and the average profile.
- % Difference Chart: Look for long bars to identify outcomes with the greatest % difference between observed and expected counts.

Percentage Profiles Chart
Compare the profiles.

% Difference between Observed and Expected Counts
Positive: Occur more frequently than expected
Negative: Occur less frequently than expected

Pre Requisites	
Data Type	Discrete Ordinal or Nominal
Distribution	Chi- square

Statistics of interest : Proportion

Chi - Square Goodness of fit

Chi-Square Test for Association: Outcomes by X
Summary Report

Pre Requisites	
Data Type	Discrete Ordinal or Nominal
Distribution	Chi- square

Six Sigma is a quality program that, when all is said and done, improves your customers' experience, lowers your costs, and builds better leaders.

— Jack Welch —