goskills

Hypothesis Testing

GoSkills online course syllabus

Friday, August 19, 2022

Skill level
Intermediate

Lessons
25

Accredited by
CPD

Pre-requisites
None

Video duration
2h 35m

Estimated study time
14h for all materials

Instructor
Ray Sheen

Creating an Hypothesis

1 Concept of Hypothesis Testing

The scientific method of analysis is to create an hypothesis, develop experiments that generate applicable data, then analyze that data to prove or disprove the hypothesis. This approach allows us to confidently answer inquiry questions with data. This lesson explains the concepts of hypotheses in problem solving.

2 Hypothesis Test Process

Effective hypothesis testing is a disciplined process. From writing the process, to designing the study or experiments, and finally analyzing the data, there are proven best practices that should be applied. This lesson presents and explains the hypothesis testing process as used in Lean Six Sigma.

3 Writing Hypotheses

A well written hypothesis contains two elements, the Null hypothesis and the Alternate hypothesis. Writing a clear hypothesis that can be quickly analyzed with a statistical test is a skill that will be illustrated and practiced in this lesson.

4 Hypothesis Tests

There are many different statistical tests that can be used with data to analyze an hypothesis. Which test depends upon the nature of the hypothesis and the test data. This lesson provides a roadmap for selecting an appropriate hypothesis test. The key decision factors for making the hypothesis test selection are addressed and illustrated.

Statistical Analysis Principles for Hypothesis Testing

5 Inferential Statistics

Hypothesis testing relies on the principle of inferential statistics. A sample of a data set is statistically analyzed and the results area used to infer information about the entire data set. This lesson will discuss the application of this principle to Lean Six Sigma projects and provide groundrules for data sampling.

6 Confidence Intervals

It is often impossible to analyze all the items in a data set, so a sample is used. But that means that the sample statistics might not perfectly represent the full data set. Depending upon the nature of the data and the sample, a confidence interval can be established around any sample statistic to show the range in which the actual statistic occurs. This lesson discusses how to calculate this range and criteria for sampling that impact this range.

7 Alpha and Beta Risk

The statistic created in an hypothesis test is only 100% valid for the sample used to conduct the test. The application to the broader data set includes some uncertainty about the statistic in the entire data set. Therefore, it is possible to reach a wrong conclusion about the entire data set based upon the analysis of the sample. The elements of that risk are classified as Alpha risk and Beta risk. This lesson explains these risks and their significance.

8 The P Value

Inferential statistics relies on a statistical measure of goodness known as the "P" Value to determine whether to accept or reject the Null hypothesis. This P value is based upon the type of test conducted and the confidence interval and Alpha risk that are applied to the situation. This lesson explains the principle of the "P" value and its use in Lean Six Sigma projects.

9 Normal Versus Non-Normal

One of the most important criteria for selecting an hypothesis test is based upon whether the data is normal or non-normal. The normality question does not prove or disprove the hypothesis, rather it steers the nature of the analysis. This lesson reviews this concept and its application in hypothesis testing.

10 Uni-, Bi-, Multi-Variate Tests

Different tests are designed to test different quantities of test samples or test parameters. The correct test will ensure a meaningful analysis.

11 Classes of Distribution

Data sets are often displayed in distributions. Different distributions are indicative of different physical phenomena. The ability to recognize a distribution will aid in the identification of process performance issues.

Regression Analysis

12 Correlation

A common investigation in Lean Six Sigma problem solving is to determine if two factors are correlated. This insight will often point to an underlying cause of the problem. This lesson explains how to do correlation analysis using both Excel and Minitab. It also includes a discussion of the Pearson correlation coefficient.

13 Simple Linear Regression

Simple linear regression analysis creates an equation that correlates two factors. This equation both assists in understanding problems, and it can also be used to manage the problem or process going forward. This lesson shows how to calculate this line with the help of either Excel or Minitab.

14 Residual Analysis

A statistical analysis or test creates a mathematical model to fit the data in the sample. The real world data seldom precisely fits the model. The differences between the model and the actual data is known as residuals. The residuals in any analysis, whether a regression analysis or another statistical analysis, will indicate how well the statistical model fits the data. When the residuals indicate a bad fit, a different analytical approach should be selected. This lesson explains how to read residual graphs and analysis.

15 Multiple Linear Regression

Many times there are multiple factors that are influencing the response variable in a problem. Multiple regression determines the relationship between the factors and the response, including interaction effects between factors. Like with simple linear regression, a formula is created that allows both analysis and prediction of the process and problem.

16 Non-Linear Regression

Many problems encountered in Lean Six Sigma projects are not a straight-line correlation effect like those discussed in simple linear regression or multiple regression analysis. The relationship is better modeled by an exponential curve or a parabola. These are referred to as nonlinear relationships. Minitab can help to determine the best fit curve and factors. In addition, the Box Cox Transformation can be used to assist in modeling these effects within the statistical analysis. This lesson discusses these topics and illustrates how to conduct this analysis.

Applying Statistical Hypothesis Tests

17 Test of Proportions

The One-Sample and Two-Sample Test of Proportions are used with discrete data. These tests determine whether the percentage of a particular attribute being studied is similar to or different from the selected target value. These tests are illustrated using both Excel and Minitab.

18 Chi Square Test

When the data is discrete data, but there are more than two samples to be compared, the Chi Square test is used. This test is quickly accomplished using Minitab. It can be done using Excel, but requires several intermediate steps. This lesson explains this test approach.

19 One Sample Tests

One sample tests are tests of a single dataset that compares the descriptive statistics of that data set against target values.

20 Variance Tests

Many of the hypothesis test approaches will change depending upon whether the continuous data has equal variances or unequal variances between data sets. Therefore, the F Test or Bartlett's Test must be completed to determine if variances are equal. The F test can be done with either Excel or Minitab. The Bartlett's Test can only be done using Minitab. This lesson explains these tests and how to read the statistic.

21 T Tests

T Tests compare the mean of two data sample to each other.

22 ANOVA

The Analysis of Variance (ANOVA) test is a commonly used test in Lean Six Sigma projects. It allows the comparison from multiple data sets to determine whether there is a statistical difference in those data sets. The analysis can be easily done in both Excel and Minitab. This course addresses the basic of ANOVA.

23 One-Sample Sign / One-Sample Wilcoxon

The One-Sample Sign Test and the One-Sample Wilcoxon Test accomplish the same purpose, but each has strengths and weaknesses. When the data is not normal, or it is not known that it definitely is normal, these tests can be used to determine if the data set statistics meets or exceeds a target value. The application of this test using Minitab is illustrated in this lesson.

24 Levene's Test & Mann-Whitney Test

These tests are used when comparing two non-normal data samples. The Levene's Test checks for the variance and the Mann-Whitney checks the Medians. These tests are both standard tests in Minitab and the lesson will illustrate how to conduct them.

25

Mood's Median, Kruskal-Wallis, and Friedman

These three tests are for multiple samples of non-normal data. Each test has its strengths and weaknesses. The appropriate test will depend upon what is known, or not known, about the data in the samples. The Minitab interface to accomplish each of these tests is similar. This lesson will explain the differences and show how to conduct the test and read the results.

Go to GoSkills.com