**Category:**

Lean Thinking

**Course Prerequisite:**

None

**Course Length:**

16 hours

**Materials:**

Minitab or JMP Statistical Software (can be downloaded on a trial basis if necessary).

**Cancellation Policy:**

Less than 14 days, 50% non-refundable , less than 7 days 100% non-refundable

**Minimum Number of Students:**

6

**Maximum Number of Students:**

16

**Delivery:**vILT or cILT  
(virtual or classroom)

**Pre-work:**None

**Basic Data Mining**

**Course Description**

In today’s data rich environment, vast amounts of data are routinely collected. These are termed ‘happenstance’, ‘non-experimental’, or ‘observational’ data. The role of statistics with such observational data is to extract all available information – often called Data Mining – and in particular to identify the Key Process Input Variables (KPIVs) for use in process improvement and process control. With a suitable sampling plan and a knowledge of how to prepare data for analysis, the engineer or researcher can then use statistical methods, much like a detective looking for clues, to release otherwise hidden information from data, providing the basis for correct decisions.

Observational data require special techniques and care to extract meaningful information and reach valid conclusions. Observational data are common in most process industries and can yield valuable information from normal process data without resorting to designed experimental data, which may be more costly to obtain. This course gives basic methods to compare a single input to a single output. It covers discrete or continuous inputs with continuous outputs and discrete inputs with discrete outputs. The methods introduced here are building blocks for more advanced data mining techniques as well as the basis for single factor experiments.

**Who Should Attend**This course is designed for engineers, quality professionals, researchers, and managers who need to understand and extract information from observational data such as key process input variables or process drivers.

**Learning Objectives**Through training, participants will:

* Understand statistical reasoning
* Be able to plan a multi-vari study and clean datasets
* Learn the different types of statistical tests based on data type (t-Tests, ANOVA, non-parametric tests, simple linear regression and chi-square test)
* How to avoid the pitfalls and perils of analyzing observational data
* Improve utilization of available data to extract relevant information