

VERSION 1.5



# DATA ANALYTICS COURSE

Presented by:  
Georges Van Cauwenbergh  
Master Black Belt

### TARGET AUDIENCE

This course is recommended for all those in an organisation who want to get a good understanding of Data Analysis, the methodology, the main tools and tests, and when to use them and when not.

In addition we will spend some time looking at various common errors you must be aware of to avoid wrong results.

#### Day 1 and Day 2 – All participants foundation course

- Day 1: Data Types, Distributions, Measurement System Analysis and Graphical tools
- Day 2: Hypothesis testing – Part I

#### Day 3 and Day 4 – Advanced participants

- Day 3: Hypothesis testing – Part II, Statistical Process Control, Regression and Capability
- Day 4: Common Errors in the following areas (Sources, Data Collection, Data Quality, Estimation, Graphics, Choosing the test, Statistical Procedures, Interpreting results, Reporting)

---

### WHY DATA ANALYTICS IS IMPORTANT FOR YOUR BUSINESS

Organisations handle a lot of data every day and sometimes the quantity of data can be overwhelming. Data Analytics is a method to make sense of your data and extract valuable intelligence to guide the business in the all-important decision making process. You cannot change or manage successfully what you don't measure.

‘Without data you are just another person with an opinion’ – Dr. Deming

---

### OBJECTIVES OF THE COURSE

At the end of the course participants will understand what data analytics methodology is and will have a fundamental understanding of how to use the various tools listed below. Even though we will practice as much as possible, it is advised that the participants keep practicing to reinforce the learning.

The knowledge acquired will provide an understanding of how Data Analysis can be integrated into the company's reporting and decision making structure.

---

### METHOD OF CONTACT

Phone: 00353 1 639 0050 / Email: [info@trigraph.ie](mailto:info@trigraph.ie) / Web: <http://www.trigraph.ie/data-analytics>

Data Analytics Course

**COST**

All prices quoted are without VAT – There is no VAT on Training.

Details	Max Number of people	Costs
2 Days Participants Foundation Course	12	€1,395
2 days Advanced participants Course	12	€1,595
Full 4 day Data Analytics Course	12	€2,695

**MORE DETAILS**

**Description:** This course aims to provide delegates with sufficient basic knowledge to analyse data in the best way without making any of the common errors.

The main headings we will cover during the course are described below. Each section will first be explained theoretically, followed by practical exercises using example data (or data that is provided by the customer). We will use Minitab data analysis software to practice the various analysis. Each participant will be provided with a free 30 day trial version of Minitab. Some of the topics will also be illustrated using MS EXCEL.

**Audience:** This course is recommended for all those in an organisation who want to get a good understanding of Data Analysis, the methodology, the main tools and tests and when to use them and when not.

In addition we will spend some time looking at various common errors you must be aware to avoid wrong results.

**COURSE AGENDA DAY 1 AND 2**

**Day 1**

1. **Data Analysis Methodology – CRISP-DM** (Cross-Industry Standard Process for Data Mining)
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modelling
  - Evaluation
  - Deployment

Introduction to Minitab and Excel (History in Minitab & Change settings in Excel)

Data Analytics Course

## 2. Distributions

- Types of data: Continuous, Attribute (Ordinal, Discrete and categorical)
- Normal Distribution
- Binomial Distribution
- Poisson Distribution

## Day 2

### 3. Hypothesis testing – Part I

When you want to compare averages or medians of some samples of data to decide if they are statistically different.

When you want to compare the standard deviation of some samples of data to decide if their variation is statistically different.

When you want to compare proportions or percentages that came from different samples of data to decide if they are statistically different.

- **Principles of Hypothesis testing:** What is it and why and when do we use hypothesis testing
- **1 Sample T-Test:** for comparing the averages of one sample against a specific target or historical average
- **2 Sample T-Test:** for comparing the averages of two samples against each other.
- **One way ANOVA:** for comparing the averages of 3 or more samples against each other
- **Paired T-Test:** for comparing the averages of two samples that contain data that is linked in pairs.

### 4. Exploratory data analysis via graphical tools

- Time series
- Scatter
- Pareto
- Box and Whiskers

## Day 3

### 5. Measurement System Analysis (MSA)

This is a technique for understanding the quality of data by challenging its sources and their potential errors.

- Gage R&R (Repeatability and Reproducibility) study for continuous data
- Attribute R&R study for attribute data

### 6. Hypothesis Testing – Part II

What is it and why and when do we use hypothesis testing?

**Levenes Test:** for comparing the standard deviations of 2 or more samples that are not normally distributed

**F-Test:** for comparing the standard deviations of 2 samples that are normally distributed

**Bartlett's Test:** for comparing the standard deviations of 3 or more samples that are normally distributed.

**1 Proportion Test:** for comparing a proportion against a specific target or historical proportion

**2 Proportions Test:** for comparing 2 proportions against each other

**Chi-Square Test:** for comparing 3 or more proportions against each other.

### 7. Graphical Tools – Part II

**Histogram:** The data is summarised into bars with the most frequent values being represented by the higher bars. The overall shape of the distribution can be assessed.

**Probability Plot:** used to decide if a sample data fits a specific distribution

**Matrix plot:** produces an array of scatter plots

**Box Plot:** Shows the distribution of a sample data as a box and whiskers

**Individual value Plot:** for comparing the distribution of several samples against each other

**Fitted line Plot:** is a scatter plot in which the relationship between the input and the output is represented mathematically by a single line (a regression line) which can be linear or curved.

## 8. Statistical Process Control (SPC)

**I-MR chart:** for analysing individual data points of continuous data

**U- Chart:** for analysing the count or defects per unit

**Xbar R Chart:** for analysing the averages of small sub-groups (2 to 5)

**P chart:** for analysing the proportions or percentages

**Xbar S Chart:** for analysing the averages of large sub-groups (more than 6)

## 9. Regression

- Simple regression: model the relationship between one X variable and a response variable Y
- Multiple Regression: to model the relationship between two to five X variables and a response variable Y
- Optimise response: using multiple regression to model the relationship between two to five X variables and a response variable Y and identify X values that optimise Y.

## 10. Capability Analysis

- **Capability Analysis:** determine whether the process is capable of producing output that meets customer requirements.
- **Binomial Capability:** determine whether the % defective meets customer requirements
- **Poisson Capability:** determine whether the defect rate (DPU) meets customer requirements