Six Sigma
Green Belt
Introduction to Minitab
Minitab is statistical analysis software. It can be used for learning about statistics as well as statistical research. Statistical analysis computer applications have the advantage of being accurate, reliable, and generally faster than computing statistics and drawing graphs by hand. Minitab is relatively easy to use once you know a few fundamentals.

Open Minitab following the steps below.

- Click the **Start button in the bottom left hand corner of the screen.**
- Select **Programs >Minitab for Windows>Minitab (or Minitab 12 Student).**

Or

- Double click the **Minitab icon on the desktop.**
- **Minitab will open.**
Other windows include

• Graph Window: When you generate graphs, each graph is opened in its own window.

• Report Window: Version 13 has a report manager that helps you organize your results in a report.

• Other Windows: *History and Project Manager are other windows. See Minitab help for more information on these if needed.*

DATA Types

• Numerical: Numerical data is the only type Minitab will use for statistical calculations. Numerical data is aligned on the right side of the column. Minitab will not recognize numbers with commas as numbers but will consider them text.

• Text: Text cannot be used for computations. Though “text” generally means words or characters, numbers can be classified as text. If column 1 has text in it, the column label will change from C1 to C1-T. Data types can be changed. See the details in the Manipulating Data section.

• Date/Time: Minitab recognizes 3/5/00 as a date and 4:30 as a time but will store these internally as a number so you can manipulate them. The column label will indicate a date or time value by putting a D after the column name (for example C1-D).
Minitab will conduct a variety of statistical calculations. These are found under the main menu option of STAT. Each category also has subcategories. In this introductory lesson, we are interested in the Basic Statistics. A menu of the statistics categories and the subcategories for Basic Statistics from Student Version 12 are shown below. (Note: Professional Version 13 has additional options.)
Definition of Probability

Experiment: Toss a coin twice

Sample space: possible outcomes of an experiment

I. \( S = \{HH, HT, TH, TT\} \)

Event: a subset of possible outcomes

I. \( A = \{HH\}, B = \{HT, TH\} \)

Probability of an event : an number assigned to an event \( Pr(A) \)

I. Axiom 1: \( Pr(A) \geq 0 \)

II. Axiom 2: \( Pr(S) = 1 \)

III. Axiom 3: For every sequence of disjoint events

\[
Pr(\bigcup_i A_i) = \sum_i Pr(A_i)
\]

IV. Example: \( Pr(A) = \frac{n(A)}{N} \): frequentist statistics
Hypothesis Testing

- Is also called significance testing
- Tests a claim about a parameter using evidence (data in a sample)
- The technique is introduced by considering a one-sample z test
- The procedure is broken into four steps
- Each element of the procedure must be understood

Hypothesis Testing Steps

A. Null and alternative hypotheses
B. Test statistic
C. P-value and interpretation
D. Significance level (optional)
Null and Alternative Hypotheses

This is an example of a one-sample test of a mean when $\sigma$ is known. Use this statistic to test the problem:

$$z_{stat} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

where $\mu_0 \equiv$ population mean assuming $H_0$ is true

and $SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

**z statistic**
For the illustrative example, $\mu_0 = 170$
We know $\sigma = 40$
Take an SRS of $n = 64$. Therefore

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$
Null and Alternative Hypotheses

Sampling distribution of $\bar{x}$ under $H_0: \mu = 170$ for $n = 64 \Rightarrow$

$\bar{x} \sim N(170, 5)$
T-Test

**t Distributions**

- $t$ dist. are used when we know the mean of the population but not the SD of the population from which our sample is drawn.
- $t$ dist. are useful when we have small samples.
- $t$ dist is flatter and has fatter tails.
- As sample size approaches 30, $t$ looks like $z$ (normal) dist.

**Same Three Assumptions**
- Dependent Variable is scale
- Random selection
- Normal Distribution