**MATH 3210**

**Assignment #3 - Messy Data Exercise**

**Scenario:** The dataset Diabetes.xls contains the values of eight attributes for individuals classified as either healthy or sick. Several of the attributes have missing or incorrect data. We are going to attempt to resolve some of these messy data issues.

**Part 1 (Triceps Skin Fold Thickness):** The Triceps Skin Fold Thickness data contains many missing items. One could make an argument that there is a correlation between the Triceps Skin Fold Thickness and the Body Mass Index.

* Is there a correlation between these two attributes?
* If so, can we use this relationship to fill in some of the missing values?

**Part 2 (Diastolic Blood Pressure):** The Diastolic Blood Pressure data contains many missing items. Recommend a methodology to resolve these missing values. Can your recommended solution help simplify the data issue in Part 1 of this assignment?

**This will be an in-class exercise. Be prepared to participate in this exercise.**