**Determining the Sample Size**

The first step is to estimate the sample size that will be required to meet the desired Power (80%). We can estimate this sample size by utilizing the existing sample. In this way we will consider the first sample our pilot data. If we assume the underlying data is Normally distributed, then we can calculate the minimum sample size that will produce our desired power as follows:

n = (Zα+Zβ)2 \*2\*σ2 / d2

n = the desired sample size

= the critical value for the desired confidence level

= the critical value for the desired power, where power is

= the standard deviation of the underlying distribution

d = the minimum difference in the means that we would consider significant

Since the true standard deviation are not known we will use an estimate based on our pilot data as follows:

, where is our sample standard deviation and is our sample size.

This approach produces the sample size required for each individual group (i.e. apples and oranges). We can then make a simplifying assumption that the sample proportion of apples to oranges is representative of the population to calculate the total sample size required to meet the desired power, significance level, and minimum significant difference in means.

***A note about the pilot sample data***

When reviewing the sample data provided in the oranges.txt file, several significant outliers are observed. For oranges there are 63 observations. Most of the prices listed are under $10.00, with multiple observations at or near $10.00. However, there is one observation for $1000.00, one for $129.00, and one for $54.00. If these values are included the mean price for the oranges group is: $21.79. If these three observations are removed the mean price for the orange group falls to approximately: $3.99. These outliers have similar large impacts on the calculated standard deviation. As such we removed these outliers from the data in order to calculate the desired sample size.

***Calculations***

For the purposes of the sample size calculation we used the following assumptions:

Desired Power = .8

Confidence level = 5%

Minimum difference in means considered relevant = $.10,

The minimum difference represents a 2.5% difference in the pilot sample mean price for the orange group.

The calculation for the sample size required for the orange group based on this calculation is 186. The ratio of apples vs. oranges in the pilot sample data is: 87:60 respectively. This means that in order to target 186 observations in each group the total minimum sample size should be approximately 456.

When collecting this sample care should be given to collect random samples from a group that is believed to represent the overall population in question.

**Normality Assumptions**

The above calculations assume that the underlying population is normally distributed. The assumption of

Part 2

**Section 1**