Statistical Practices in Excel and Octave: Script

Siddhi Krishna

Professor Tim Hickey, CS177

April 2011

Hello! Welcome to this Octave Statistics Tutorial! Today, we will be comparing how to perform basic statistics on both Excel and Octave.

We want to find the mean, standard deviation, and mode of some data, and then perform an unpaired t-test.
Our data is already loaded into Excel. To find the mean, standard deviation, and mode, we simply pick a cell and give it the label we want. Then, next to it, we simply type "$= MEAN($" and highlight the data we would like to analyze. We continue this for both data sets.

Great! We have our vital statistics for each data set.

But we now want to perform an UNPAIRED t-test. This means that the data sets come from unique samples; that is, any participant took either of the surveys but not both.

We perform an unpaired t-test by selected a cell and labeling it "t-stat" so that we know what the statistic is. Then, we select the adjacent cell and type "$= TTEST($". Notice

that this requires 4 parameters: array #1, array #2, the number of tails, and the type of test. So we highlight array #1, press "," to indicate a change in parameters, then array #2. We are testing the null hypothesis that the two samples are not equal, rather than one being less than the other. Therefore, we type 2. Similarly, since this is an unpaired t-test, we type 2 again. Pressing enter, we get our P-value!

We also want to extract from the P-value the t-statistic. This, too, is easy! We again select a cell, and then type "$= TINV$", following by the location of the P-value, and then the degrees of freedom, which is the sum of the sample sizes -2.

Now we have all our vital statistics in Excel. We should now do this in Octave!

We load in our data by using the command "load data.csv". Notice that we must load in a .csv or a .dat file. We can't load in an excel file! We know the data is divided with column one being one set of data, and column 2 being the second set. We can find the mean, standard deviation and mode by typing "$m = mean(data(:, 1))$", and then "$stddev = std(data(:, 1))$", and "$mode = mode(data(:, 1))$". This can be continued for the second set of data.

To perform an unpaired T-Test, we must use the "T_TEST_2" command. This command returns the p-value, t-statistic, and degrees of freedom all at once in an array. Therefore, we can retrieve all this information at once by simply typing "$[pvaltstatdof] = t\_test\_2(data(:, 1), data(:, 2))$" This assigns the appropriate statistics to their respective locations in the array. Now, we just type pval, tstat, and dof to retrieve their values.

We have now successfully completed finding vital statistics in both Excel and Octave.