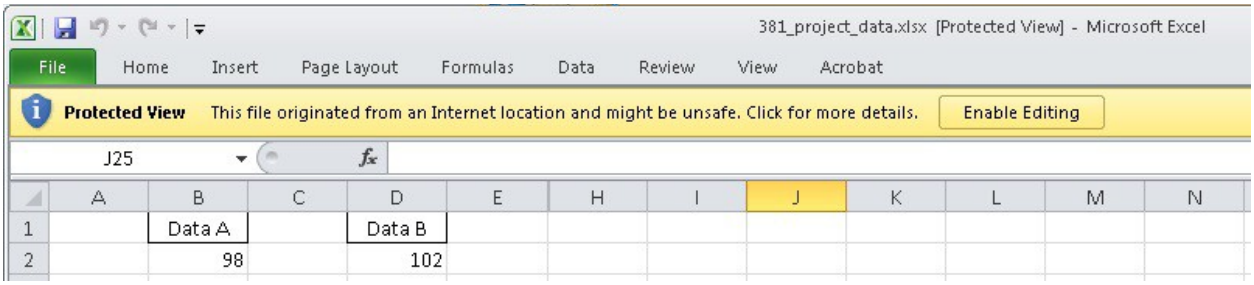
Statistics Project – Resistors

Full Project

For this project we will assume that we tested two batches of resistors (in ohms) from two different plants, **A** and **B**. The first batch was very small. The second batch was much larger. This project is going to be done entirely in MatLab. You must email your files following the instructions below by **11:59PM, Wednesday, December 9th** . All files that you need are to be found in the folder named "Project" on the main Moodle page.

First, you must retrieve your data. This can be found in the file, "STAT383_project_data.xlsx". In this file, you will see two columns of data: Data A and Data B. Press F9 to randomize the data and **copy only the data points** of the column marked "Data A" and paste them into a txt file (I would use Notepad or something) and name it "dataA.txt". Next, do the same thing for Data B in a separate txt file.

NOTE: If you see something at the top of Excel like the following,



you can simply select "Enable Editing" so that you can get the correct data sets. The data sets will be **unique** for each group.

Next, open the file that is called, "STAT383_Project_lastname_studentID.m" . Replace "lastname" and "student ID" with your correct values. This is where you are going to put your code. At the very top, there is a place for you to type in your FULL NAME and your STUDENT ID. Make sure it stays commented out.

```
% Student Names: Your FULL NAME(s)
% Student IDs: Your STUDENT ID(s)
```

Next there will be a section called `%% Init.` This section contains some code to clear up your workspace, etc., before you run your code. Also, at the bottom is something that looks like this,

```
disp(['Code by: FULL NAME  STUDENT ID SECTION #'])
disp(['Code by: FULL NAME  STUDENT ID SECTION #'])
disp(['Code by: FULL NAME  STUDENT ID SECTION #'])
disp([' ']);
```

The command *disp* will display the words between ([' … ']) in the command window. Here, you will again put your FULL NAME and STUDENT ID and also the FULL NAME of those you worked with. You will work in a group up to three people. (You can use "fprintf" if you like instead.)

Following this code, you will see there is code to import the data is into two arrays: dataA and dataB. It is important that you import the data as directed above and in the code so that I can run your data with my code to check for the correct answers.

Next, you are asked to make four plots: A box plot and a histogram for each data set. (Make sure they are labeled appropriately.)

```
%% Plot the data (dataA)

%  (Box plot) For the key parts of the plot, see figure 6.24

% (Histogram)
```

Use something like ,

```
figure;boxplot(dataA);title ('Boxplot for DataA');ylabel('ohms')
```

So that each plot has a separate figure generated. After each set of plots (A and B), you will be asked the following questions,

```
%% Questions

% Is the data skewed? Why or why not. Use your plots to support your
% analysis. HINT: Read 6.2 and 6.3 carefully.
```

You should answer the question in complete sentences and display it in the command window using the *disp* command as above. Also, to display numerical solutions you can use the *num2str* command like this (notice I also put a `disp([' ']);` to separate the sentences),

```
Y = 5
disp(['The value of Y is ' num2str(Y) ' and it is good.']);
disp([' ']);
```

Under the next section, `%% Sample Analysis` you will calculate the sample mean, mode, median and sample standard deviation for each data set. The median you can get from the box plot or the MatLab command *median(data)* and the mode may be seen from the data or use the MatLab command *mode(data)*. However, the sample mean and sample standard deviation MUST be calculated using the formulas from class, i.e. do not use software commands like *mean(X)* but you can still use *sum*.

The rest of the instructions are in the m.file.

Read the commented text and follow the directions carefully. Ask if you any questions but ask sooner than later.

For your project to be successful (i.e. full points), you need to send me ZIP file of your data files and your MatLab file (m.file) . In addition, the following must be followed:

1) I should be able to run your file with **no errors**. This is easy to check before you send it and may result in the loss of quite a few points if I cannot resolve your errors.

2) You must send me a ZIP file that contains your data files and your m.file code. **Failure to do this will result in ZERO credit**.

3) Label your completed zip file – "STAT383_Project_lastname_studentID.zip" where you need to include your last name and student ID. **Failure to do this will result in ZERO credit**.

Below is what I expect to see displayed in the *command* window when I run your code. The values will of course be different. Also, below are the graphs properly labeled that your code should generate.

Code by: group member1 0111111
Code by: group member2 0222222
Code by: group member3 0333333

The data in the first data set, A, seem to be positively or right skewed. This is illustrated by the fact that in the boxplot, the length of the horizontal line from Q3 up to Max is much longer, 2.8333 times more than the one from Q1 down to Min. Also, the histogram seems to have a positive skew.

The data in the second data set, B, seem not to be skewed. In the boxplot, the horizontal line from Q3 up to Max and the one from Q1 down to Min seem to be roughly the same length. This is also illustrated by the fact that the histogram seems to be fairly symmetric.

For data set A:
The sample mean is 98.85.
The mode is 101.
The median is 99.
The sample standard deviation is 4.4988.

For data set B:
The sample mean is 99.62.
The mode is 99.
The median is 100.
The sample standard deviation is 4.7605.

The two-sided, 95% confidence interval for the sample mean of data set A, is (96.7445, 100.9555)

The two-sided, 95% confidence interval for the sample mean of data set B, is (98.6754, 100.5646)

The two-sided, 99% confidence interval for the sample mean of data set A, is (95.972, 101.728)

The two-sided, 99% confidence interval for the sample mean of data set B, is (98.3697, 100.8703)

The T cdf for t = 2.262 is P(T<t) = 0.97499

For H0: mu_A = 98.25, the p-value = 0.55792
For H0: mu_B = 98.25, the p-value = 0.0049053

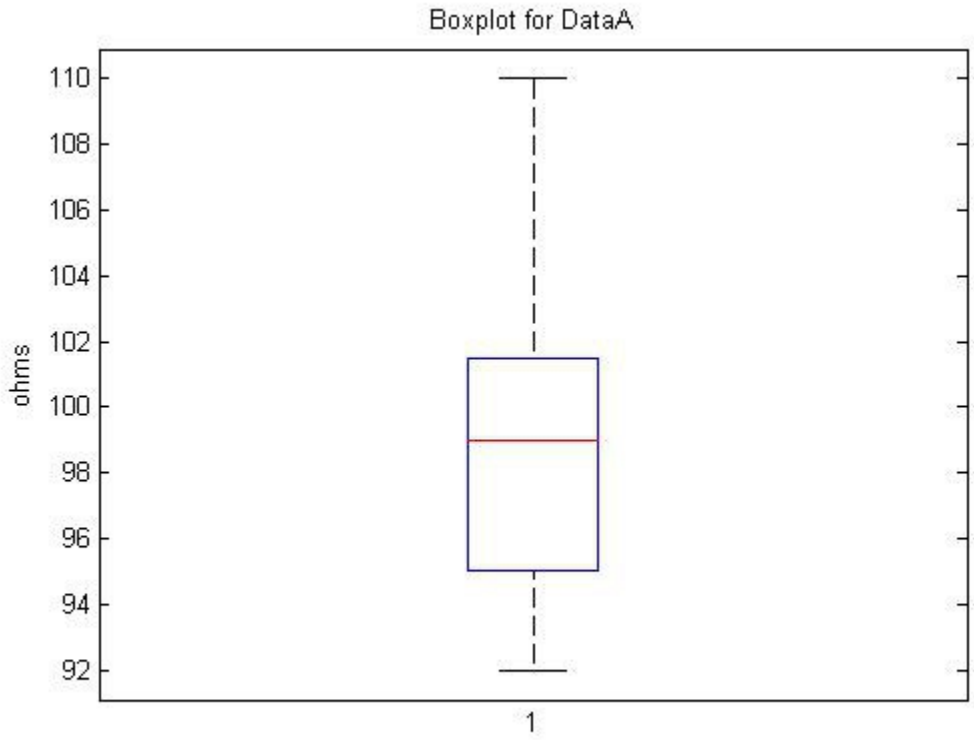
The absolute value of the t statistic for dataA (0.59644) is less than the critical value, 2.093 and therefore with a significance level of 5%, we ACCEPT the null.

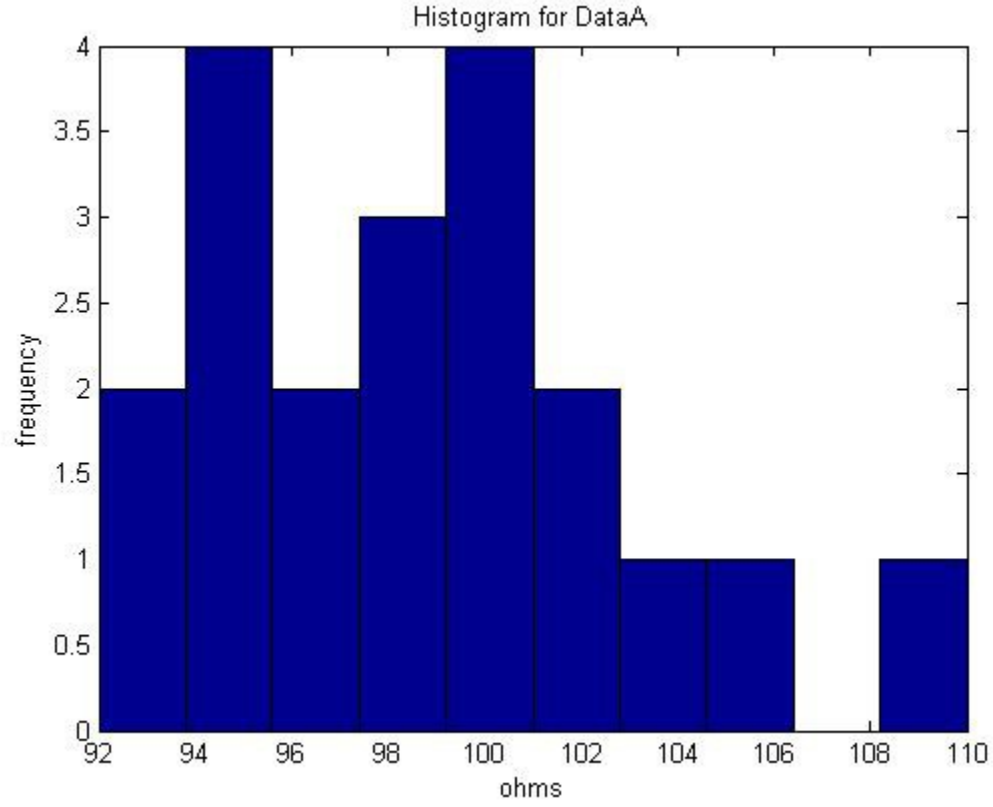
The absolute value of the t statistic for dataB (2.8779) is greater than the critical value, 1.9842 and therefore with a significance level of 5%, we REJECT the null.

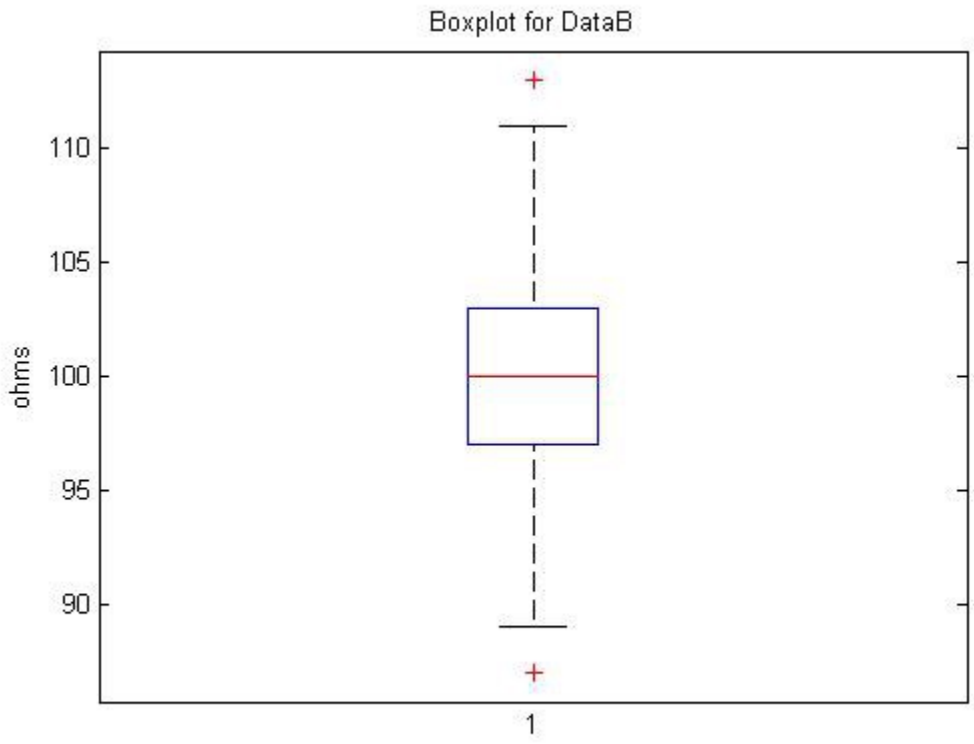
For H0: muA - muC = 0, the paired p-value = 0.33459
The absolute value of the paired t statistic (0.99005) is less than the critical value, 2.093 and therefore with a significance level of 5%, we ACCEPT the null.

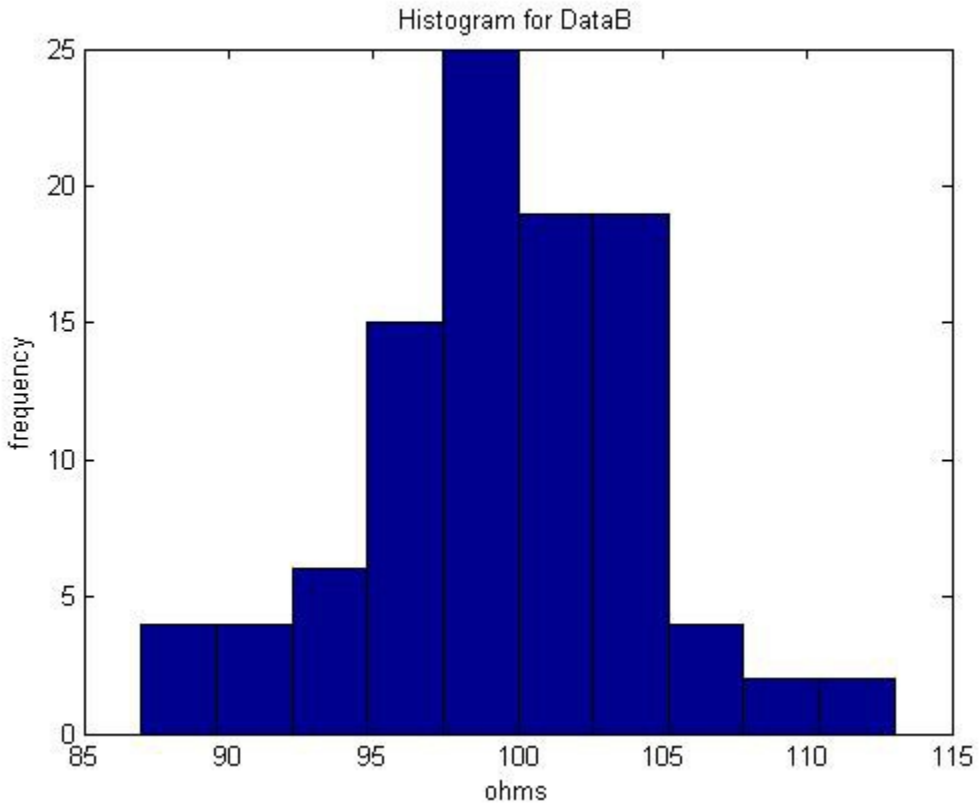
For H0: muA - muD = 0, the un-paired p-value = 0.47363
The absolute value of the un-paired t statistic (0.7244) is less than the critical value, 2.0301 and therefore with a significance level of 5%, we ACCEPT the null.

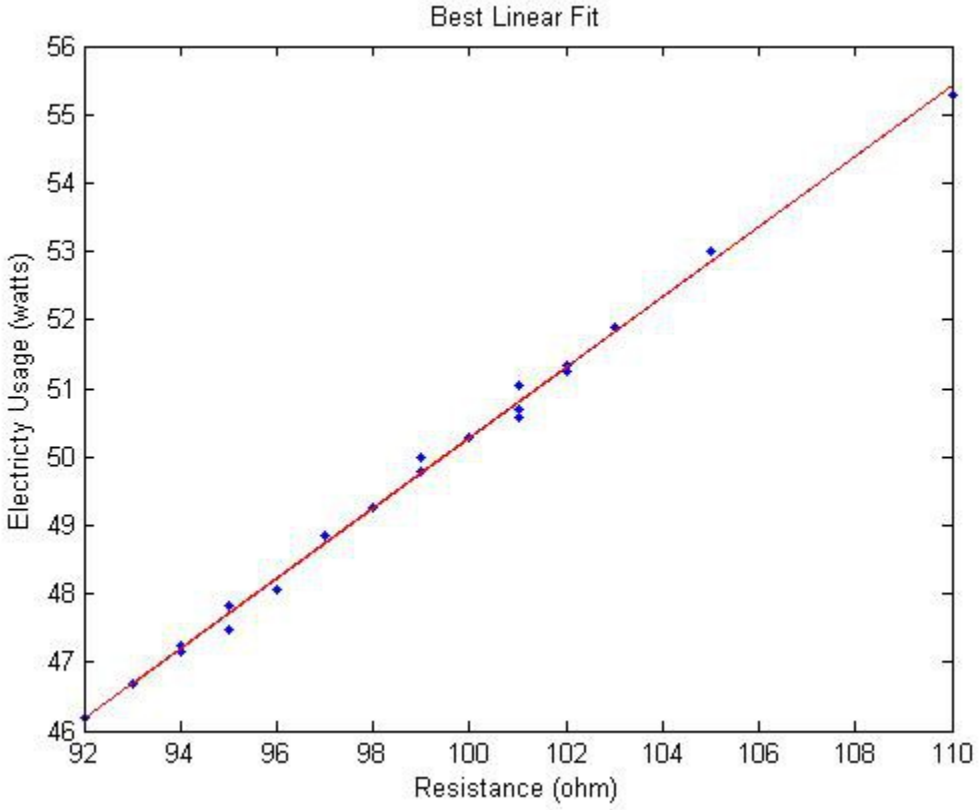
Calculated slope and intercept
~B1 = 0.5136
~B0 = -1.0781


Boxplot for DataA
ohms
110
108
106
104
102
100
98
96
94
92
1


Histogram for DataA
frequency
4
3.5
3
2.5
2
1.5
1
0.5
0
92
94
96
98
100
102
104
106
108
110
ohms


Boxplot for DataB
ohms
110
105
100
95
90
1


Histogram for DataB
frequency
ohms
0
5
10
15
20
25
85
90
95
100
105
110
115


Best Linear Fit
Electricty Usage (watts)
Resistance (ohm)
46
47
48
49
50
51
52
53
54
55
56
92
94
96
98
100
102
104
106
108
110