Jonathan Ye

November 20, 2009

Identifying Patterns

Prof. Little

Lab on Correlation, Regression, and Data Analysis

**Part A**

1. There does not appear to be a strong linear relationship between the population and the rank. For the scatter plot, R= 0.7, which means the fit is pretty close to all the points; however, the value of the residual is extremely large (3.08 x 1014). This means that it most likely does not have a linear relationship.
2. Although this seems to have a better correlation, there is still a change in slope from the left to right. The residual value is still too large (in the 40s), but the R value is even closer to one.
3. This is the best linear fit of the data given in this scatter plot. The residual value is close to 1 and the r value is also close to 1. Therefore taking the natural log of both sides makes the points line up in a pretty straight line, which can be given a linear equation.
4. Using normal x and y values, the results are still quite a bit off. The residual value is much too large, but the r value is close to 1 (0.89). Compared to question one, where the first few points were included, there is not much of a difference. Taking the ln of both does not give a good residual value (too big), but a good r value (0.97). Once again, compared to the data from question three, this gives a worse linear approximation. Finally, question two has the best linear approximation of the data. ln(y) and x is a good fit of the points because both R and the residual value are close to one. Therefore, in this case, the first nine points improved the approximation.
5. The smaller the rank, the larger the population is. The best fit line for the data (from which this conclusion is drawn from) is using the natural log of both variables (x and y).
6. I used data from China to check if the same pattern holds true for all other countries, and it appears to be similar. The trend follows that the best fit for the graph appears when the natural log of both the rank and population are taken. The correlation coefficients tend to increase as more variables are changed to natural logs. However, there were still a few outlier points when the natural log was taken. The smallest cities appeared to have such a small population that they did not fit well into the graph of the best fit line. Overall, the pattern appears to be similar for these two countries; however, more would need to be tested to see if this trend holds true for almost all countries.

**Part B**

1. We can use the Standard Deviation to measure uniformity of the polysilicon thickness between all the sites for each wafer. The values are located next to the locations in the excel spread sheet.
2. The standard deviation of the points without location 13 is much better for all wafers so location 13 is an outlier. For each wafer, the SD is significantly lower, except for Wafter ID B2, in which the standard deviation is still lower, but only within a difference of one.
3. This was done and under the tab, “Multiple Regression”. The y-intercept turned out to be 115.8632. The first slope value was equal to -0.07167 while the second slope value was equal to 1.211901. All the data is still available under the “multiple regression” tab.
4. The relation between the standard deviation and oxide thickness seem to have a few outlier points, however, there is a general trend where both increase as the other increases. These two are directly proportional. A relationship between the standard deviation and the deposition time is not nearly as strong and the relationship seems more scattered. There also seems to be no relationship between oxide thickness and deposition time. The relation using the multiple relation shows this because the Multiple R value is 0.934… which is very close to 1 and therefore a relationship exists. However, the regression value does not appear to be very close to one and therefore the linear line may not be the best fit.
5. The slopes tell us that as deposition time increases, the uniformity (calculated using the standard deviation) decreases. Oppositely, as the oxide thickness increases, uniformity also increases.