Greg Monfette

November 20, 2009

Identifying Patterns

Excel Spreadsheet Write up

In this exercise there were two data sets that were used in creating and analyzing spreadsheets. The first set of data described the populations of seventy five of the largest standard metropolitan areas used by government agencies to study economic and demographic trends. The second set of data was developed by a professor at North Carolina State. It describes an experiment about the certain techniques used for manufacturing the semiconductors used in electronic devices. The following will analyze critical cetails of these data sets.

After the excel spreadsheet was made for the first set of data, it was time to analyze some of the intracasies of it. First, when you look at the data it can be determined that the is no real linear relationship between the rank and the population because the residual graph has a definite pattern. Then when you look at the natural log of y and x it is clear that there is no relationship between them because the residual has a definite pattern. Next, I looked at the natural log of y and the natural log of x and there was a clear relationship there because there was no pattern for the plots on the residual. Next, the eight largest cities were taken out and only cities nine through seventy five were analyzed. When this was done there was a much greater linear relationship because the top populations are very extreme and skew the results. The residual plot would be more scattered because the relationship was greater. It can be said that the relationship between rank and population is still the same with an equation of y=x to the =1.13 + e to the 19.5. I then decided to go online and find some extra data on other countries in the world. There would be a similar pattern in these countries because most of them have at least a few cities with a very high population compared to others. This follows the pattern established in the excel spreadsheet of the top seventy five cities.

After the excel spreadsheet was made for the data on the top seventy five cities, it was time to make the second spreadsheet on the manufacturing of semiconductors used in electronic devices. The frist question that needed to be asked was which statistic we would use to measure the uniformity of the polysilicon thickness across all the sites of the wafers. It was determined that the standard deviation would be used. The standard deviations are recorded on the spreadsheet. Next, it was noticed that when you take out thirteen from the data set that the standard deviation went down significantly. It went down almost twenty units. This fact makes a good case for excluding 13 as an unrepresentative outlier. After the data without 13 was analyzed, I put the 13 back in and analyzed the data to see if the relation is a good fit for the data or not. I decided that it was not a good fit. Finally, I analyzed the signs of the coefficients m1 and m2 to see about the roles of the oxide thickness and the deposition time in determining the measure of polysilicon thickness uniformity. All of the coefficients are small negatives or close to 0.