Amanda Cabrera

 11/20/09

 MONT 106N

Lab on Correlation, Regression, Data Analysis

**Lab Investigation A**

1. There does not seem to be a linear relation between rank (x) and population (y) according to the SMA data. The value of r is approximately -0.7 which does not show a strong correlation between variables. The residuals show a consistent decreasing/increasing and positive/negative pattern which indicates a “lack of fit” and not a strong linear relation.
2. There does not seem to be a linear relation between x and ln (y) either. Although the correlation coefficient r is -0.95 (closer to -1), the residuals still show a strong pattern which does not indicate a fitting relationship.
3. There does seem to be a linear relation between ln (x) and ln (y). The value of r is -0.98, very close to perfectly linear at -1. Also, the residuals do not increase/decrease consistently which signifies lack of pattern and a “better fit.”
4. If the first eight of the SMA data are eliminated, the recalculated correlation coefficients for each data set are closer to negative one. The new coefficients for numbers 1, 2, and 3 are approximately -0.89, -0.98, and -0.99, respectively. The residuals also show less of a pattern without the first eight statistics which would also promote a more linear relation between the variables.
5. Based on the output data, there is not a functional relation between the rank of the SMA (x) and the population (y). The correlation coefficient of -0.7 does not signify a strong correlation between the variables. Also, the residuals show a consistency of increasing/decreasing and positive/negative patterns which indicates “lack of fit” for the relation.

**Lab Investigation B**

1. The standard deviation can be used to measure the uniformity of the polysilicon thickness across all the sites on one of the wafers. (Computations in excel)
2. Site 13 can be excluded as an unrepresentative “outlier” because the numbers of all wafers at site 13 are significantly lower than all the other locations; this would be considered an anomaly of the data.
3. (Computations in excel)
4. Based on the output, I would say that the relation is a good fit because both the scatter plots are relatively “football-shaped.” This means that the endpoints of the “football” shape are the averages of the x and y plus/minus 2 standard deviations which represents 95% of the data. This represents the data well.
5. The signs of the coefficients indicate which variable plays a greater role in determining the measure of polysilicon thickness uniformity. Because the sign of the oxide thickness is negative, it plays a lesser role than the deposition time with a positive coefficient.