Lab 3: Regression and Correlation

OBJECTIVES: This lab is designed to show you how to analyze the relationship between two variables using correlation and regression. Further investigation into the regression portion of the lab will involve comparing the effects of a linear fit vs. a quadratic fit, as well as the effects of transforming the response variable.
DIRECTIONS: Follow the instructions below, answering all questions. Your answers for each of the questions, including output and any plots, should be summarized in the form of a brief report (Word), to be handed in to the instructor before the end of your assigned lab time.

Because there will be quite a few plots generated here, you may want to title each of your plots appropriately, taking into account what type of fit (linear or quadratic) you're using, whether the data is transformed or not, etc.!).

1.) *Correlation . . .*

- **In preparation for this portion of the lab, what does the correlation, r, measure?**
- **The correlation is restricted to what values? Further, what does positive correlation imply? Negative correlation?**
- **Can correlation be used to describe a curved relationship between variables, or only the strength of a linear relationship between two variables.**
- **In general, would you use correlation to completely describe two variable data?**

a.) Download the "*energy.mtw*" Minitab data worksheet from the web. *(On the Stat 280 home page, under the "Lab Assignments" section!! ;)).*

b.) Produce a scatter plot of Energy Consumption vs. Machine Settings ("Graph-> Character Graphs-> Scatter Plot") and comment on the relationship between the two variables. Can you detect any strong correlation between the two variables?

c.) Now, determine the actual correlation between Energy Consumption and Machine Settings.
(Hint: Minitab can do this very quickly if you take a look under Stat/Basic Statistics/Correlation . . .)

- **What does this correlation value tell you about these two variables, and does it go along with what you thought in part b.) above?**

2.) *Regression . . . A Linear Fit . . .*

- **In preparation for this portion of the lab, how do you define a regression line, and what is it used to do? In particular, comment on what a *least-squares* regression line is.**
- **How is regression different than correlation?**
- **What does the square of the correlation, r^2 , signify? What does an r^2 value of 1 mean?**

- **Briefly describe what residuals are and what a residual plot is used to assess.**

a.) We will now use regression to analyze how the *response* variable *"Energy Consumption"* changes as the *explanatory* variable, *"Machine Setting"* changes.

b.) Plot the simple *linear* regression line for this data. *("Stat->Regression->Fitted Line Plot").* Before doing the regression, be sure to store the resulting residuals, fits, and coefficients of this first fit in your worksheet. *(Hint: Try the "Storage" option!)*

c.) Note the results in the session window, as well as the least squares line fitted to the energy data.

- **What is the equation of the regression line that's used to predict energy consumption?**
- **Describe what the *slope* and *intercept* values really mean with regards to this data.**
- **Would you say the prediction based on this model is accurate? Why or why not? *(Hint: Consider how close the data points are to the line . . .)***
- **What does your value of r^2 tell you about the regression you've just performed.**

d.) Generate a plot of the residuals vs. the explanatory variable *(Machine Settings)* for this data. *(Hint: Look under Stat/Regression/Regression, and investigate the "Graphs" options - choose the appropriate plot accordingly . . .).*

- **What is (and *should be*) the mean of the residuals *(Note: This can be checked very quickly using the "mean" feature of the Calculator in Minitab).***
- **Comment on this plot of residuals vs. Machine Settings. Consider how this plot should appear if the regression line effectively catches the overall pattern of the data, and note if this plot achieves this form.**

3.) *<u>Regression . . . A Quadratic Fit . . .</u>*

a.) Now, apply a *quadratic* fit to this data, *(See question 2.b.) above if you don't recall how to get to this and check the option "quadratic" instead of "linear" this time)* and again perform a regression. Again, be sure to store the resulting residuals, fits, and coefficients of this second fit in your worksheet before doing the regression. Also, be sure you have your all results from the linear fit above for comparison!

b.) Note the results in the session window, as well as the quadratic fit of the energy data.

- **Analyze this quadratic fit now, comparing your conclusions from question 2.) (parts c.) and d.)). Consider which fit is better, the r^2 value now as opposed to with the linear fit, the residual plot now, etc.**

c.) Now, apply a log (base 10) transformation to the response variable *(Energy Consumption).*
(Hint: This can easily be done if you again look under Stat/Regression/Fitted Line Plot, and

look into the "Options" feature! You don't have to display the data on a log scale , although feel free to see how the plot looks on that scale if you'd like).

- **Finally, perform a regression analysis as you did above, only now with the quadratic fit and transformed Energy Consumption data, and conclude if this transformation helped the fit or not.**