Name__________________________________

Recitation__________________________________

Selecting the Right Model Studio - Excel 2007 Version

We have seen linear and quadratic models for various data sets. However, once one collects data it is not always clear what model to use; that is, what type of curve works best. Using a FIXED data set we can try to fit DIFFERENT curves (a.k.a. models) and see which appears to be a better fit for the fixed data set. In this studio we are given examples of fixed data sets and we have to decide what model (exponential, or power) fits this data best.

A. Let us start with the data given by the population of the U.S. from 1790 to 1900. You already encountered this data set in Studio 2 on linear functions. It is worth emphasizing that this data set will be fixed. However for this fixed data set there will be more than one curve or model relating the points together. You are to decide which curve fits the fixed data set best. Open your spreadsheet and download the file `logstudiodataf07.xls` from the class webpage. Click on U.S.Pop tab to find the population data.

B. We will now scale down the numbers to make them easier to work with. Enter the formula =A2-1750 into cell C2 and copy it down to cell C13. Enter the formula =B2/1000000 into cell D2. The denominator is 1 million (there are 6 zeros) and the value in D2 should be 3.929. Then copy the formula from cell D2 down to cell D13.

C. Select the range C2..D13 and click on the Insert Tab to create an XY (Scatter) plot of population (in Millions) versus years since 1750. Click on one of the data points and choose Add Trendline selecting an Exponential trendline. Check the boxes to display the equation and press OK to graph an exponential model of the data.

1a. *What is the formula for this function? Does it appear to be a good fit?* Fill in your answers below:

D. We now would like to try a power function. Click on any data point and select Add Trendline selecting the Power option; again display the equation by clicking the appropriate box. To see this curve better, right click on the EQUATION for the power curve and select Format trendline. Choose ’Line Color’ from the left hand side of the dialog box. Now select ’Solid Line’ from the right hand side of the dialog box. Use the ’Color’ pull-menu to change the color to red and click ’Close’.

1b. *What is the formula for this function? Does it appear to be a good fit?* Fill in your answers below:

E. Sometimes it can be difficult to judge models just from a first glance. While Stat 325 can give you a series of tools for identifying a good model, we will look at one quick tool here that actually works very well for deciding between exponential and power models. We will look at the data using different scales on the axes. You can think of this as looking at the same picture through different “lenses.” Knowing which lens gives the sharpest image will tell you which model is appropriate. Right click on the y-axis and select Format Axis. Under ’Axis Options’, select “Logarithmic scale”. This rescales the y-axis so that the values are placed according to the logarithm of their value, rather than the values themselves. The exponential model (the black curve) now appears linear with the y-axis formated to be $\ln(y)$. The power model still appears curved. Notice that the formulas did not change. This is because the model didn’t change, just the lens through which we look changed. Evaluating the natural logarithm on both sides of $y = Ae^{kx}$ gives us $\ln(y) = \ln(A) + kx$ which is why an exponential appears linear if we change the y-axis to the $\ln(y)$-axis.

2a. *Take the natural logarithm of both sides of the formula you found in problem* $1a$*.* Note that this formula is now a line in the variables x and $\ln(y)$.

F. Now right click on the x-axis and select Format Axis. 'Axis Options', select "Logarithmic scale". It is clear the power model (the red curve) now appears linear with both the x-axis and the y-axis formatted to be $\ln(x)$ and $\ln(y)$, respectively. Meanwhile, the black exponential curve is back to a curved shape. Notice that again the formulas did not change. Again, we have not changed the model, we have changed the way we are viewing it. Evaluating the natural logarithm on both sides of $y = Ax^k$ gives us $\ln(y) = \ln(A) + k\ln(x)$ which is why a power curve appears linear if we change the x-axis to the $\ln(x)$-axis and the y-axis to the $\ln(y)$-axis.

2b. *Take the natural logarithm of both sides of the formula you found in problem* $1b$*.* Note that this formula is now a line in the variables $\ln(x)$ and $\ln(y)$.

Conclusions: Given a data set which appears to grow according to a power law or exponential model, it can sometimes be difficult to tell which model is appropriate, because the two curves look similar. We can simplify the decision by looking at the data through two different lenses. If we use a log-scale on the y-axis only (called a semi-log plot), the data points will fall roughly on a line if an exponential model is appropriate. If we use a log-scale on both the x and y axes (called a log-log plot), the data points will fall roughly on a line if a power law is appropriate. Since it is easier to look at data points and decide if they are on a line than to decide what sort of curve they follow, checking these two scales makes it easier to see which model is best. In this case, we can see an exponential model is most appropriate for describing population growth in the U.S. during the 19th century.

G Now you will get a chance to go through the process again for a new data set. Click on the Track Records tab on the spreadsheet. This will bring up a list of the Men's and Women's world record times for distances between 100 meters and 10,000 meters. This time you won't need to scale the data. Just graph the men's data in cells A3 through B12. Follow the instructions from steps C-E above to find both an exponential and power law model for the data.

3. *What are the exponential and power law models for this data?*

4 *Which model is a better fit? Briefly explain why.*

H. Of course, once we have models we want to use them to analyze a situation. Suppose we want to predict the world record for the 1 meter dash. That might sound ridiculous, but in many sports the time it takes to go from a ready position to taking the first step is vital. Track athletes have a near ideal setup for taking that first step, so measuring how long it takes them to get that first step off gives a measure of what an outstanding reaction time is. Of course, the world record for the 1 meter dash isn't just 1/100th the world record for the 100 meter dash. But with our model, we can now predict what the world record time should be for taking the first step.

5. Using the model you selected in problem 4, compute the expected world record time for the Men's 1 meter dash.

Bonus: Repeat problems 3-5 for the women's world track records. You should now be able to do this without detailed instructions. List your answers to the 3 questions below.