Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date\_\_\_\_\_\_\_\_\_\_\_\_\_\_

AP STATISTICS – MS. KLIMCZUK

Using the Calculator: Re-Expressing Data to Achieve Linearity

Step 1: MAKING A SCATTERPLOT

Enter the following data into your calculator and make a scatterplot. This shows the change in tuition costs at Arizona State University during the 1990s.

|  |  |
| --- | --- |
| Year (Starting with 1990) | Tuition (In Dollars) |
| 0 | 6,546 |
| 1 | 6,996 |
| 2 | 6,996 |
| 3 | 7,350 |
| 4 | 7,500 |
| 5 | 7,978 |
| 6 | 8,377 |
| 7 | 8,710 |
| 8 | 9,110 |
| 9 | 9,411 |
| 10 | 9,800 |

Analyze the scatterplot. What do you notice? Does a linear model seem appropriate here? Explain.

Write down the regression model for this data.

What does the slope and intercept mean?

Step 2: MAKING THE RESIDUALS PLOT

Now check the residuals to see if a linear model is really appropriate for this data. What do you notice?

Step 3: RE-EXPRESSING THE DATA

When you made the residuals plot, you should have seen a distinct curve. Residuals are high (positive) at the left, low in the middle of the decade, and high again at the right.

This curved pattern indicates that data re-expression may be in order. If you have no clue which re-expression to try, the Ladder of Powers may help. Which one do you think we should use?

It is reasonable to suspect that tuition prices increase at a relatively consistent percentage year by year. This suggests that the logarithm of tuition may help straighten our data.

Now, tell the calculator to find the log of the tuitions and store them in a new list. Go to the home screen. Type in ***log (L2) 🡪 L3***. Now your third list will have all the logs of your tuition costs.

Check the scatterplot for your re-expressed data. Go to ***STATPLOT*** and change your ***Xlist*** to ***L1*** and your ***Ylist*** to ***L3***. Don’t forget to press ***ZoomStat***.

Analyze the scatterplot. What do you notice? Does a linear model seem appropriate here? Explain.

Write down the regression model for this data and graph it with your scatterplot. Use the following command to do this: ***LinReg(a+bx) L1, L3, Y1.*** This will create the residuals and report the details about the model.

What does the slope and intercept mean?

Step 4: MAKING THE RESIDUALS PLOT OF THE RE-EXPRESSED DATA

Now check the residuals to see if a linear model is now appropriate for the log of tuition costs. Go to ***STATPLOT*** and change your ***Xlist*** to ***L1*** and your ***Ylist*** to ***RESID***. Don’t forget to press ***ZoomStat***.

What do you notice?

While the residuals for the second and fifth years are comparatively large, the curvature we saw before is gone. The pattern of the residuals seems essentially horizontal and random. This re-expressed model seems better than the first one.

Step 5: ANALYZING THE MODEL

Remember that your model involves a log re-expression. Your calculator does not indicate that! Write down the model with the appropriate variables now.

This model doesn’t predict tuition. It predicts the logarithm of tuition.

Question: Predict the tuition for 2001.

Note: When you are working with models that involve re-expression, you will often need to “backsolve” like this to find the correct predictions.