Two Models are Better than One

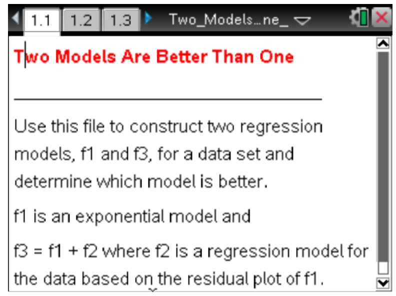
Student Activity

Name ______________

Class ______________

Open the TI-Nspire document *Two_Models_Are_Better_Than_One.tns.*

In this activity, you will construct two models for a data set involving carbon dioxide in the atmosphere– an exponential model and an improved model using the sum of the exponential model and a model for its residuals.



Move to page 1.2.

Press ctrl ▶ and ctrl ◀ to navigate through the lesson.

The concentrations of carbon dioxide were measured monthly over a one-year period. The results are listed in the spreadsheet on Page 1.2.

- The first column contains the number of the month **[mon]** starting with January as 1.
- The second column contains the concentration of carbon dioxide data collected during each month [in parts per million] **[conc]**.

Move to page 1.3.

A scatterplot of the data from the spreadsheet is shown on Page 1.3.

Using an exponential function to model the data, the regression equation is $y = 319.998 \cdot (1.0054)^x$. The residual plot is shown in the screen at the bottom of the page (the independent variable, month, is plotted along the horizontal axis, and the corresponding residual is plotted along the vertical axis.).

Recall that the residual associated with the data point (x_i, y_i) is $y_i - f(x_i)$ where y_i is the actual value of the data and $f(x_i)$ is the value predicted by the regression function.

Move to page 1.4.

The function y from Page 1.3 has been saved as $f1(x) = a \cdot b^x$ where $a = 319.998$ and $b = 1.0054$.

Move to page 1.5.

Page 1.4 shows a scatter plot of the data and a graph of $y = f1(x)$.

1. Looking at the graph of $y = f1(x)$ on the scatterplot, is the function $f1$ a good fit for this data? Give reasons for your answer.

Move back to page 1.2.

To save the residuals in the spreadsheet, "$= conc - f1(mon)$" has been entered into the cell right below the title of the third column, $resid1$. The residuals are displayed. In the residual, $y_i - f(x_i)$, y_i is the actual value of the concentration in month i, whereas $f(x_i)$ is the value of the concentration in month i predicted by the regression function, $f1$.

Move to page 1.4.

2. What is the largest residual in absolute value (or magnitude)?

Another measure of how well a function, f, fits a set of data is the sum of the squares of the residuals denoted by $SSE(f)$. Usually, the smaller the value of the $SSE(f)$, the better f fits the data.

3. Compute $SSE(f1)$, the sum of the squares of the residuals for $f1$, by typing $sum(resid1^2)$, and then record its value here:

   $SSE(f1)$ = ________ .

Move to page 1.6.

Another copy of the residual plot is included on Page 1.6, $x = mon$ and $y = resid1$.

4. Because the residual plot associated with the regression function $f1(x)$ suggests a pattern, there might be a better fit to the original data. As a first step to finding such a function, what type of function would be a good fit for the data in the residual plot? Give a reason for your answer.

Move back to page 1.4.

5. Find the regression function, $f2$, of the type you chose in Question 4 by selecting **MENU > Statistics > Stat Calculations >** *the regression function,* $f2$. Use $x = mon$ and $y = resid1$. Save **RegEqn** to $f2$, and record this function here:

   $f2(x)$ = ________________.

Name _______________
Class _______________

Move to page 1.6.

Check the fit of $f2(x)$ to the data in the residual plot. Open the entry line, move back up to $f2(x)$, and press [enter]. If $f2(x)$ does not appear to be a very good fit, make another choice for $f2$, and repeat the steps on Page 1.4 again.

Move back to page 1.4.

6. Define $f3(x) = f1(x) + f2(x)$, and record it here:

   $f3(x) =$ ____________________.

In the next few steps, you will verify that $f3(x)$ is a better fit to the original data than $f1(x)$ is.

Move to page 1.5.

To view the graph of $f3(x)$ on the scatterplot, open the entry line, move back up to $f3(x)$, and press [enter].

7. Looking at the graph of $f3(x)$ on the scatterplot, give a reason why it is a better fit to the data than $f1(x)$.

Move back to page 1.2.

Record the residuals for $f3(x)$ in the spreadsheet by selecting the cell right below the title of the fourth column, $resid2$, and entering **= conc – *f3*(mon)**. The values of the residuals will be displayed.

8. What is the largest residual in absolute value (or magnitude)?

Move to page 1.6.

Hide the graph of $f2$. Then change the function type to **scatterplot**. Display the residual plot of $f3$ by setting $x = mon$ and $y = resid2$.

9. By comparing the two residual plots, give a reason why $f3(x)$ is a better fit to the data than $f1(x)$ is.

Move back to page 1.4.

10. Finally, find $SSE(f3)$ by typing $sum(resid2^2)$, and record the value here:

    $SSE(f3)$ = ______________

11. Compare the values of and $SSE(f1)$ by completing the statement:

    $SSE(f3)$ is ____ % of $SSE(f1)$.

12. Review your answers to the previous questions, and summarize the reasons why $f3(x)$ is a better fit to the data than $f1(x)$ is.