Nonlinear Regression Act 2
Correcting the Standard Deviation of the Residual Errors for Exponential Functions (Statcrunch)

Directions: Statcrunch’s calculation of standard error is usually pretty accurate except in the case of exponential functions. Because Statcrunch uses a “Log y” transformation to compute its exponential functions, all the y values have been compressed, which makes the standard deviation of the residual errors \( S_e \) on the screen totally wrong. Standard Deviation of the errors is an important statistic in the analysis of nonlinear functions. It not only tells us how far on average ordered pairs are from the curve, but it also tells us how much error we can expect to have when we make a prediction in the scope of the data. Let’s see if we can use Excel to calculate the real \( S_e \). Here is the formula for standard deviation of the errors that we will be using. SSE refers to the “sum of the squares of the residual errors” and \( n \) is the number of ordered pairs in your data set.

\[
S_e = \sqrt{\frac{SSE}{n - 2}}
\]

1. We looked at the exponential growth relationship between the years since 1995 (x) and worldwide wind power capacity in MW (megawatts) (y). Statcrunch gave the following exponential equation and scatterplot. Notice the Standard Error is very wrong. Use the equation and Excel to calculate the real \( S_e \). Then write two sentences explaining the two meanings of the standard error in this problem.

\[
\text{LN} \ ( y ) = 8.5061554 + 0.25017053 \ ( x )
\]

\[
\text{Standard Error} = 0.078793843 \ (\text{Wrong!})
\]

a) Open the Mod 4 data sets. Copy and paste the years since 1995 and wind power data into a fresh excel spread sheet.

b) We will be using Excel to first calculate the y values on the curve (predicted y values). First plug in the x values into the right side of the log equation below. In cell C2 type in “=A2*0.25017053 +8.5061554”. Push enter. Now drag cell C2 down to the bottom to calculate the rest of the column.

c) Now to solve for y, we will need to raise the right side of the equation to the power of “e” (2.71828). In cell D2 type in “=exp(C2)”. Push enter. Now drag cell D2 down to complete column D. Title column D “predicted Y value”

d) Now we will calculate the residuals by subtracting the actual y values in column B minus the predicted y values in column D. So in cell E2, type in “=B2-D2” and push enter. Drag down cell E2 to complete the residuals.

e) We need to square all the residuals and add them up to find SSE (sum of squares of errors). In cell F2, type in “=E2^2” and press enter. Drag down cell F2 to complete the
squares column. Now we need to add the squares. In any open cell, type in “=sum(F2:F9)” and push enter. This is the SSE.

f) Now we are finally ready to calculate standard error. Since there were only 8 ordered pair data points, our n=8. Use the standard error formula $s_e = \sqrt{\frac{SSE}{n-2}}$ and SSE to calculate the standard error.

g) Now write two sentences interpreting the two meanings of standard error in this context.

2. We looked at the exponential decay relationship between months since January 1st 2010 (x) and a retirement account balance in thousands of dollars (y). Statcrunch gave the following exponential equation and scatterplot. Notice the Standard Error is very wrong. Use the equation and Excel to calculate the real standard error. Then write two sentences explaining the two meanings of the standard error in this problem.

\[ \ln(y) = 4.1565898 - 0.020360086(x) \]
Standard Error = 0.068367959 (Wrong!)

a) Open the Mod 4 data sets. Copy and paste the months since January 1st and the retirement account data into a fresh excel spread sheet.

b) We will be using Excel to first calculate the y values on the curve (predicted y values). First plug in the x values into the right side of the log equation below. In cell C2 type in “=A2*(-0.020360086) + 4.1565898”. Push enter. Now drag cell C2 down to the bottom to calculate the rest of the column.

c) Now to solve for y, we will need to raise the right side of the equation to the power of “e” (2.71828). In cell D2 type in “=exp(C2)”. Push enter. Now drag cell D2 down to complete column D. Title column D “predicted Y value”
d) Now we will calculate the residuals by subtracting the actual y values in column B minus the predicted y values in column D. So in cell E2, type in “=B2-D2” and push enter. Drag down cell E2 to complete the residuals.

e) We need to square all the residuals and add them up to find SSE (sum of squares of errors). In cell F2, type in “=E2^2” and press enter. Drag down cell F2 to complete the squares column. Now we need to add the squares. In any open cell, type in “=sum(F2:F26)” and push enter. This is the SSE.

f) Now we are finally ready to calculate standard error. Since there were 25 ordered pair data points, our \( n = 25 \). Use the standard error formula \( s = \sqrt{\frac{SSE}{n-2}} \) and SSE to calculate the standard error.

g) Now write two sentences interpreting the two meanings of standard error in this context.

3. We looked at the exponential growth relationship between years since 1990 (x) and the saving’s account balance in dollars (y). Statcrunch gave the following exponential equation and scatterplot. Notice the Standard Error is very wrong. Use the equation and Excel to calculate the real standard error. Then write two sentences explaining the two meanings of the standard error in this problem.

\[
\ln (y) = 6.214606 + 0.081580066 (x)
\]

Standard Error = 0.0000029724237 (Wrong!)

a) Open the Mod 4 data sets. Copy and paste the years since 1990 and savings account data into a fresh excel spread sheet.

b) We will be using Excel to first calculate the y values on the curve (predicted y values). First plug in the x values into the right side of the log equation below. In cell C2 type in “=A2*0.081580066+6.214606". Push enter. Now drag cell C2 down to the bottom to calculate the rest of the column.
c) Now to solve for $y$, we will need to raise the right side of the equation to the power of 10. In cell D2 type in “=exp(C2)”. Push enter. Now drag cell D2 down to complete column D. Title column D “predicted $Y$ value”.

d) Now we will calculate the residuals by subtracting the actual $y$ values in column B minus the predicted $y$ values in column D. So in cell E2, type in “=B2-D2” and push enter. Drag down cell E2 to complete the residuals.

e) We need to square all the residuals and add them up to find SSE (sum of squares of errors). In cell F2, type in “=E2^2” and press enter. Drag down cell F2 to complete the squares column. Now we need to add the squares. In any open cell, type in “=sum(F2:F9)” and push enter. This is the SSE.

f) Now we are finally ready to calculate standard error. Since there were only 8 ordered pair data points, our $n=8$. Use the standard error formula $s_e = \sqrt{\frac{SSE}{n-2}}$ and SSE to calculate the standard error.

g) Now write two sentences interpreting the two meanings of standard error in this context.

4. We looked at the relationship between the metal distance in millimeters (x) and the ultrasound response values (y). Statcrunch gave the following exponential equation and scatterplot. Notice the Standard Error is very wrong. Use the equation and Excel to calculate the real standard error. Then write two sentences explaining the two meanings of the standard error in this problem.

\[
\text{Ln}(y) = 4.3162985 - 0.48313114(x)
\]

\[
\text{Standard Error} = 0.24656116 \text{ Wrong!}
\]
a) Open the Mod 4 data sets. Copy and paste the metal distance and ultrasound response data into a fresh excel spreadsheet.

b) We will be using Excel to first calculate the y values on the curve (predicted y values). First plug in the x values into the right side of the log equation below. In cell C2 type in “=A2*(-0.48313114)+ 4.3162985”. Push enter. Now drag cell C2 down to the bottom to calculate the rest of the column.

c) Now to solve for y, we will need to raise the right side of the equation to the power of 10. In cell D2 type in “=exp(C2)”. Push enter. Now drag cell D2 down to complete column D. Title column D “predicted Y value”.

d) Now we will calculate the residuals by subtracting the actual y values in column B minus the predicted y values in column D. So in cell E2, type in “=B2-D2” and push enter. Drag down cell E2 to complete the residuals.

e) We need to square all the residuals and add them up to find SSE (sum of squares of errors). In cell F2, type in “=E2^2” and press enter. Drag down cell F2 to complete the squares column. Now we need to add the squares. In any open cell, type in “=sum(F2:F215)” and push enter. This is the SSE.

f) Now we are finally ready to calculate standard error. Since there were 214 ordered pair data points, our \( n = 214 \). Use the standard error formula \( s_e = \sqrt{\frac{SSE}{n-2}} \) and SSE to calculate the standard error.

g) Now write two sentences interpreting the two meanings of standard error in this context.

5. Explain why using a “Log Y” transformation would so drastically effect the standard error calculation on Statcrunch.