Nonlinear Regression: When Linear Regression Doesn't Cut It

Amber, Brandon, and Cody have a new data set and are trying to create the most accurate regression possible, and they need your help determining who is right – again.

The data below depict the number of women athletes at each of the 28 Summer Olympics since 1896 (minus 1940 and 1944, when the Olympics were canceled due to WWII).

Olympic Number	# of Women Athletes	Olympic Number	# of Women Athletes	Olympic Number	# of Women Athletes	Olympic Number	# of Women Athletes
1	0	8	156	15	613	22	2202
2	23	9	312	16	680	23	2721
3	6	10	202	17	783	24	3520
4	6	11	361	18	1060	25	4068
5	44	12	446	19	1260	26	4303
6	53	13	521	20	1123	27	4611
7	78	14	371	21	1567	28	4655

Enter this information into a table in Desmos (www.desmos.com/calculator). See your notes on how to do so if you have forgotten.

Does this data look linear? ________________

Create a linear regression for this data (see your previous notes).

What is the equation for the line that models this data? ______________________________________

$r =$ _________________________ $r^2 =$ ________________________________

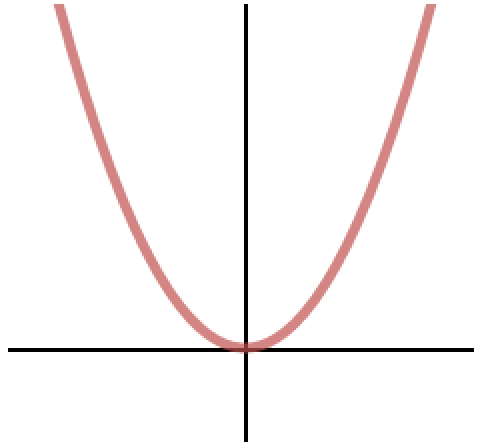
Turn on the residuals for this line. Sketch the plot below. Is this a good residuals plot? Why or why not?

Do you think that this data can be accurately modeled with a linear regression?

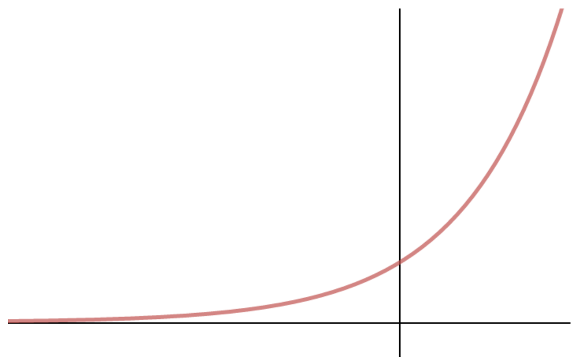
This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Amber, Brandon, and Cody decided that a line of best fit wasn't appropriate for this data set. They looked over the different parent functions they know, and each picked a different one that they think will produce the most accurate curve of best fit.

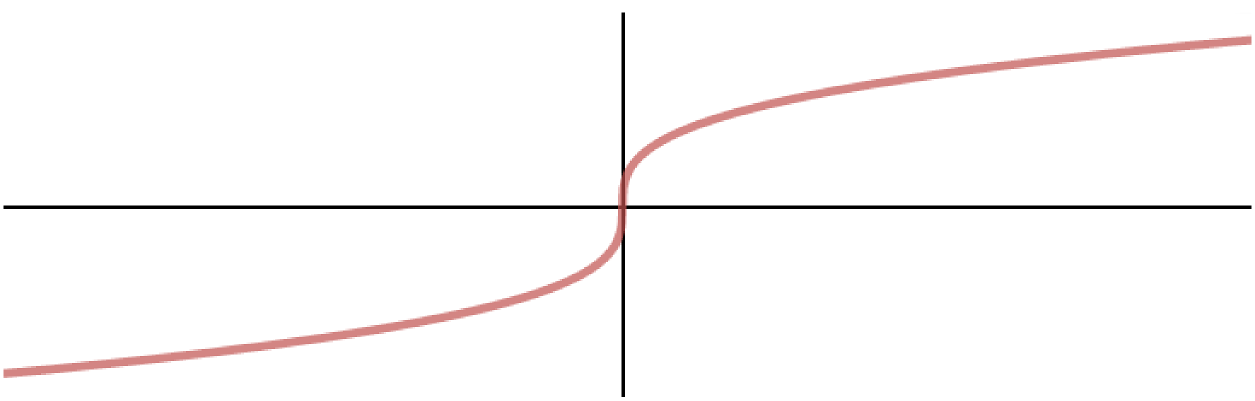
Amber thinks that the data looks like it models the right side of quadratic function ($y = x^2$)



Brandon thinks that the data looks like it models an exponential function ($y = b^x$)



Cody thinks that the data looks like it models a cube root function ($y = \sqrt[3]{x + b} + c$)



Who do you think is right? Why?

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Test Amber's theory using Desmos. Leaving the same data set in the table, graph a new regression line by typing "y1~ax1^2" to produce a quadratic regression line.

What is the equation for the quadratic regression line? ____

$r^2 =$ _________________________

Turn on the residuals for this line. Sketch the plot below. Is this a good residuals plot? Why or why not?

Do you think that this data can be accurately modeled with a quadratic regression?

Test Brandon's theory using Desmos. Leaving the same data set in the table, graph a new regression line by typing "y1~b^(x1+a)" to produce an exponential regression line.

What is the equation for the exponential regression line? _____________________________________

$r^2 =$ _________________________

Turn on the residuals for this line. Sketch the plot below. Is this a good residuals plot? Why or why not?

Do you think that this data can be accurately modeled with an exponential regression?

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Test Cody's theory using Desmos. Leaving the same data set in the table, graph a new regression line by typing "y1 = a nthroot 3 (x1+b) + c to produce a cube root regression line. (This one can be difficult to type in. Ask for help if you need it.)

What is the equation for the cube root regression line? ______________________________

$r^2 =$ ____________________

Turn on the residuals for this line. Sketch the plot below. Is this a good residuals plot? Why or why not?

Do you think that this data can be accurately modeled with a cube root regression?

Which regression most accurately models the data on these female Olympic athletes? How did you make your decision?

Use the model to predict how many women athletes will be at the 2016 Rio Olympics (Olympics #29).

This activity was created by Jessica Thomas (jmthomas891@gmail.com).

Summarizing the Investigation

A line of best fit, when data is strongly correlated, should have a correlation coefficient *r* as close as possible to _________ or ___________and a coefficient of determination r^2 as close as possible to __________

A curve of best fit cannot have a correlation coefficient, but its coefficient of determination r^2 should still be as close as possible to __________.

A line of best fit should produce a residuals plot with _________________________ correlation, and the average of residuals should be as close as possible to __________.

A curve of best fit should also produce a residuals plot with _________________________ correlation, and the average of residuals should also be as close as possible to ______________

Why is it important to construct a curve of best fit if a line of best fit does not provide a good coefficient of determination or residuals plot?

Enrichment

What other types of functions might model data? How would you write an equation to test these in Desmos?

This activity was created by Jessica Thomas (jmthomas891@gmail.com).