The University of Georgia

Mathematics Education
EMAT 4680/6680 Mathematics with Technology
Jim Wilson, Instructor

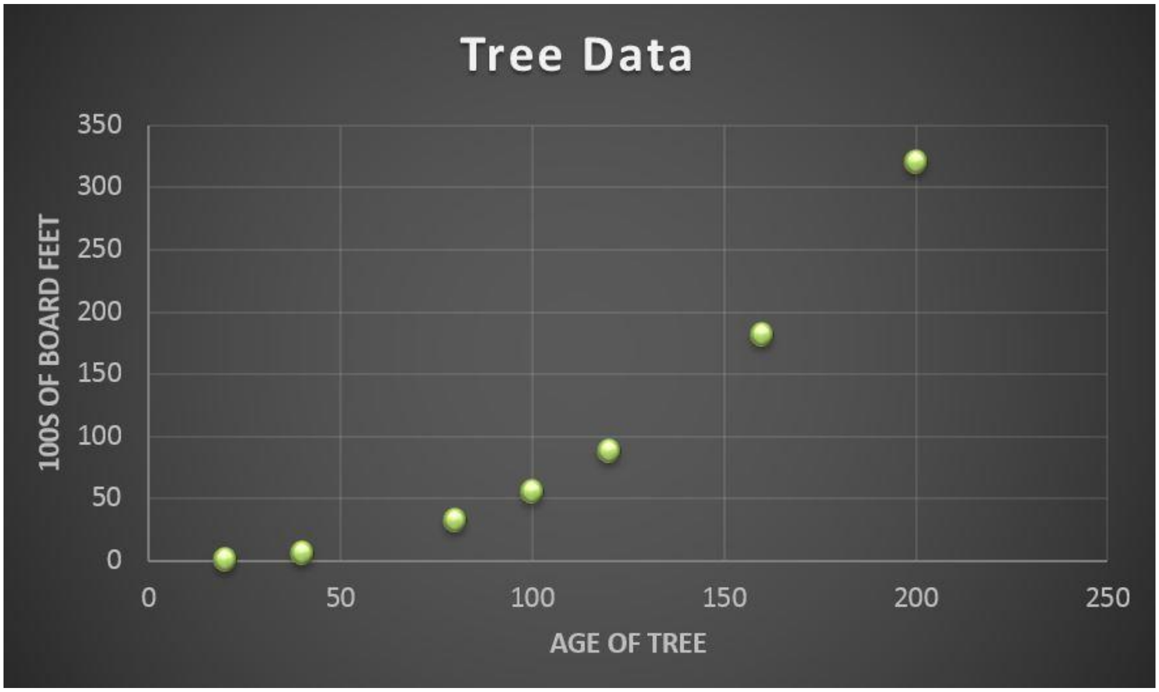
How Much Wood Could a Woodchuck Chuck?
By: Lindsey Harrison

Objective: We will first explore fitting a given data set with various functions in search of the line of best fit. To determine which fit is the best we will compute the sum of squares for error. We will then use this function to predict those data points that are incomplete in the original file.

Data and Scatterplot

Age of Tree	100s of Board Feet
20	1
40	6
60	*a*
80	33
100	56
120	88
140	*b*
160	182
180	*c*
200	320

This data is from the lumber industry, giving the approximate number of 100s of board feet of lumber per tree in a tree of a given age. Using Excel we can plot the 7 complete data points on a scatterplot (seen below).

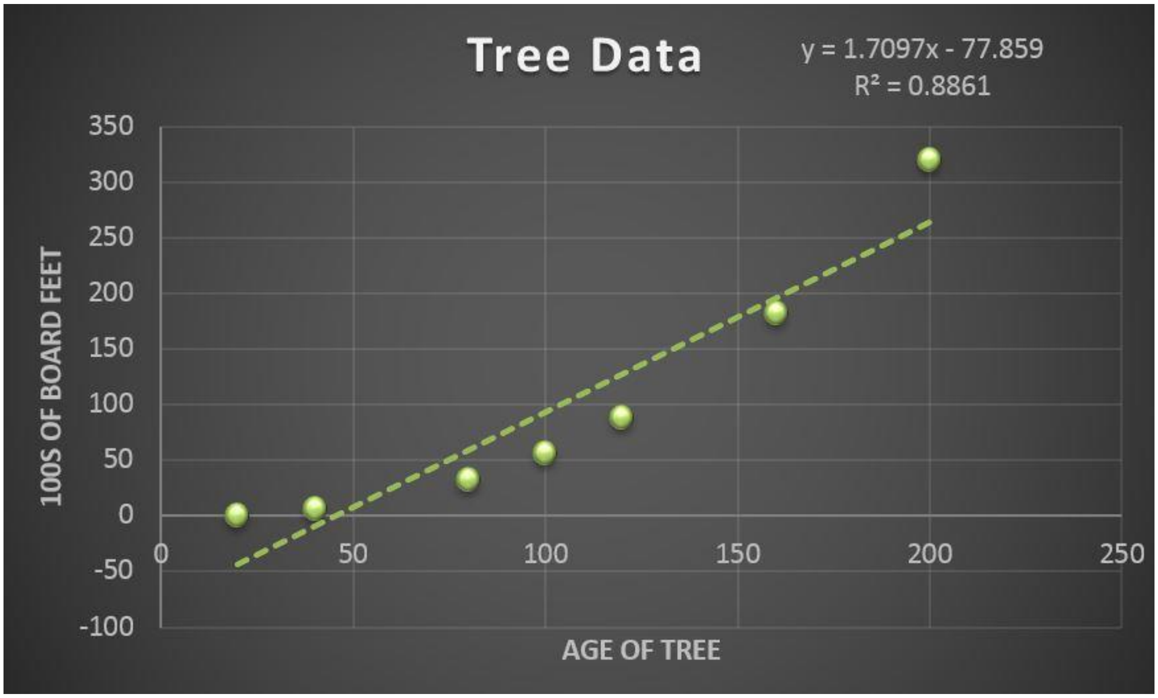


Using Excel, we are going to investigate fitting various functions to the data in order to interpolate the data to predict the missing data points.

Function Exploration

Linear

Students are most familiar with linear functions so this may be the natural first guess at fitting the data with a known function. Note that the prediction equation is posted in the top right-hand corner of the following graph.

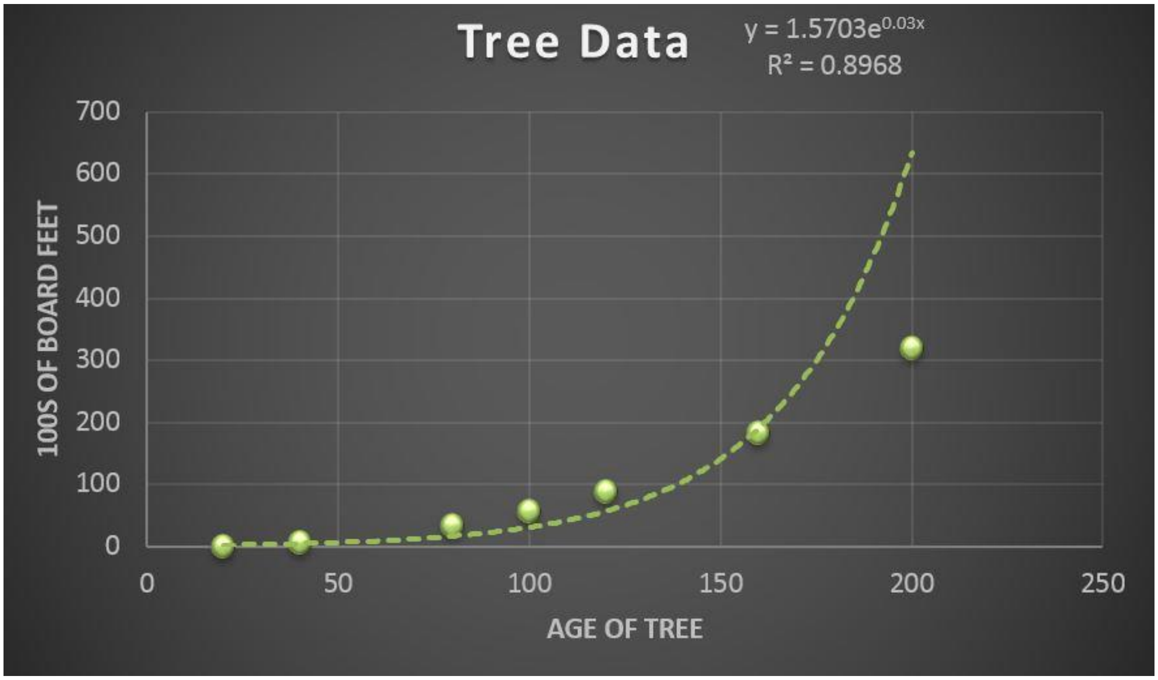


To assess the fit of the linear function, we can look at the coefficient of determination or the sum of squares of the errors. The closer the R^2 value is to one, the better the equation for prediction purposes. The linear regression model gives us an R^2 value of 0.8861. This is not a bad value, but we hope with different curves, we can get this value closer to one. The smaller the sum of squares of errors is, the better the approximating function fits the data. For the linear model the sum of squares of error is approximately 1306.

Age of Tree	100s of Board Feet	Linear Prediction Equation	Residual	Error^2
20	1	-43.665	-44.665	1994.96223
40	6	-9.471	-15.471	239.351841
80	33	58.917	25.917	671.690889
100	56	93.111	37.111	1377.22632
120	88	127.305	39.305	1544.88303
160	182	195.693	13.693	187.498249
200	320	264.081	-55.919	3126.93456
				9142.54711
				1306.07816

Exponential

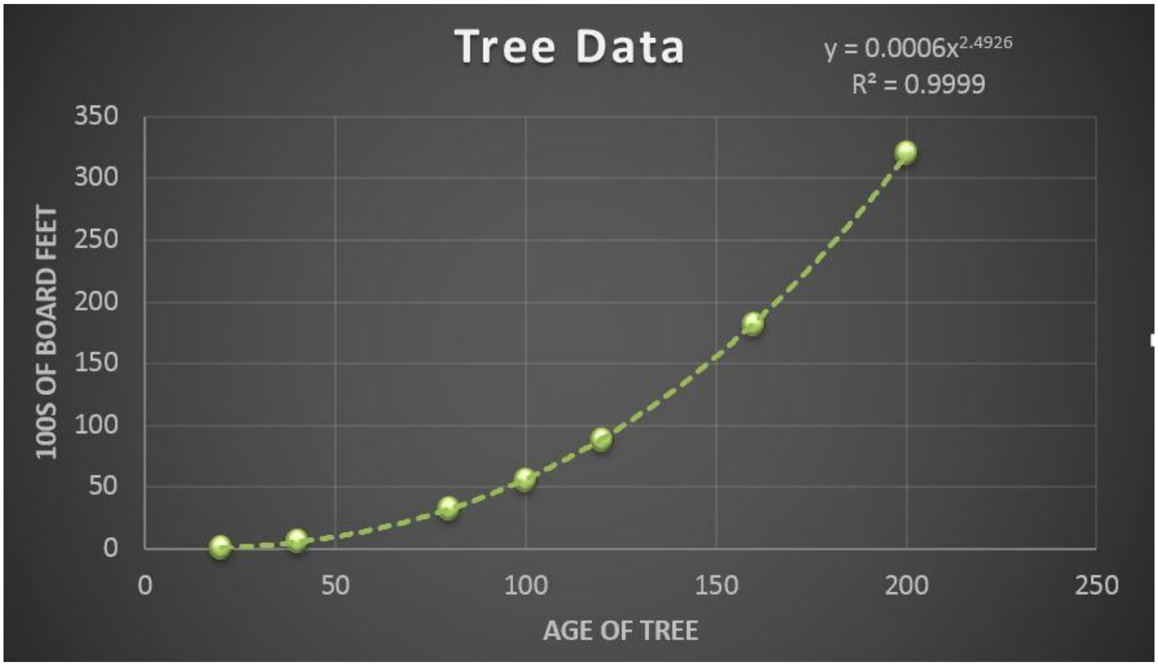
The next curve we will use to approximate the trend in tree data is an exponential regression equation. In the long run, this family of functions would not accurately model the relationship between age of a tree and 100s of board feet since the tree would eventually die; however, given our range of data, it may be useful to use an exponential curve.



The exponential regression model gives us an R^2 value of 0.8968. This is a coefficient of determination that is slightly higher than the linear regression coefficient of determination. For the exponential regression model the sum of squares of error is approximately 14,306. This show us that the exponential curve is definitely not the one we will use to describe the given data.

Age of Tree	100s of Board Feet	Exponential Prediction Equation	Residual	Error^2
20	1	2.861273152	1.861273152	3.46433775
40	6	5.213579604	-0.786420396	0.61845704
80	33	17.30969387	-15.69030613	246.185706
100	56	31.54031863	-24.45968137	598.276013
120	88	57.47020755	-30.52979245	932.068227
160	182	190.8078086	8.80780863	77.5774929
200	320	633.5042344	313.5042344	98284.905
				100143.095
				14306.1565

Power
The next curve we will use to approximate the trend in tree data is a power regression equation. Again the equation is displayed in the top right-hand corner of the following graph.

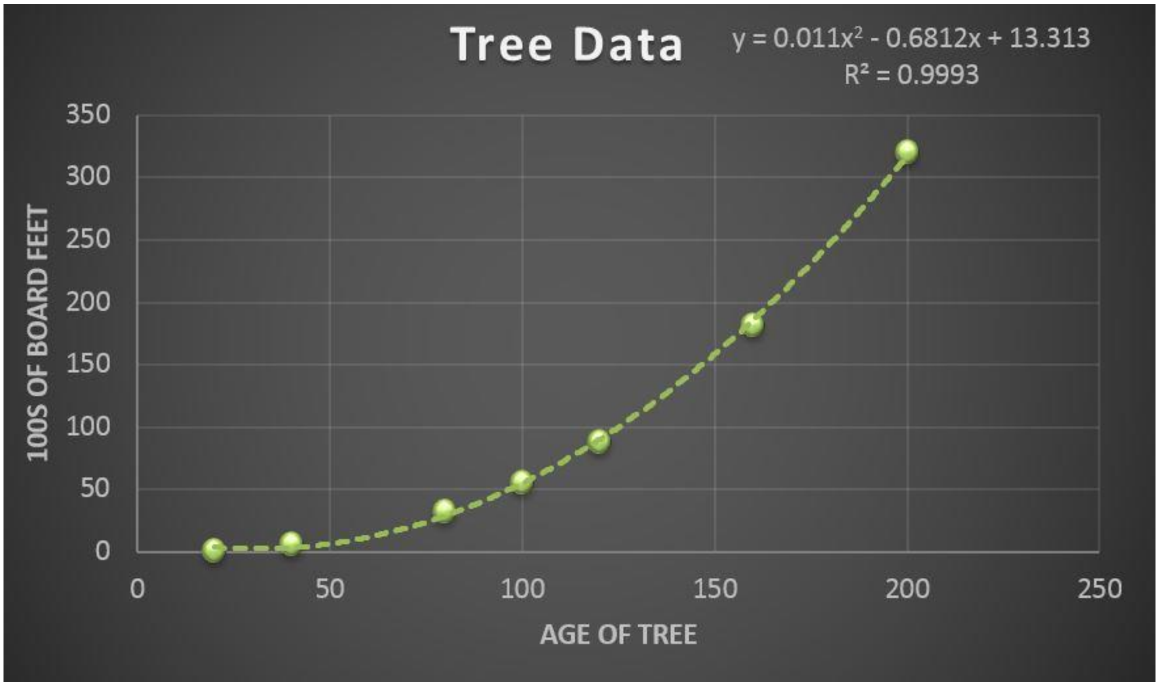


The power regression model gives us an R^2 value of 0.9999. This high of a coefficient of determination is an indicator that this model is a strong candidate to be used for the prediction of our unknown values. For the power regression model the sum of squares of error is approximately 12. This statistical summary shows that the power function does a good job of explaining the relationship between the tree data points.

Age of Tree	100s of Board Feet	Power Prediction Equation	Residual	Error^2
20	1	1.049780781	0.049780781	0.00247813
40	6	5.908074795	-0.091925205	0.00845024
80	33	33.25013033	0.250130332	0.06256518
100	56	57.98975186	1.989751859	3.95911246
120	88	91.35207664	3.352076636	11.2364178
160	182	187.1288373	5.128837293	26.304972
200	320	326.3612723	6.361272333	40.4657857
				82.0397815
				11.7199688

Polynomial

The polynomial curve of degree two is the last curve we will investigate in an attempt to find a prediction equation for the tree data.



The polynomial regression model gives us an R^2 value of 0.9993. This coefficient of determination for this polynomial regression equation is much higher than our first two curves (linear and exponential); however, it is not as high as the coefficient of determination from the power function. For the polynomial of degree two regression model the sum of squares of error is approximately 8. This statistical summary shows that the polynomial function also does a good job of explaining the relationship between the tree data points.

Age of Tree	100s of Board Feet	Polynomial Prediction Equation	Residual (prediction-data)	Error^2
20	1	4.089	3.089	9.541921
40	6	3.665	-2.335	5.452225
80	33	29.217	-3.783	14.311089
100	56	55.193	-0.807	0.651249
120	88	89.969	1.969	3.876961
160	182	185.921	3.921	15.374241
200	320	317.073	-2.927	8.567329
				57.775015
				8.25357357

Making Predictions

The two statistical quantities we have been using to assess "goodness of fit" for our functions are sum of squares due to error (SSE) and the coefficient of determination (R^2). Note the sum of squares of error measures the total deviation of the response values (100s of board feet) from the fit to the response values (predicted 100s of board feet). The coefficient of determination measures how successful the fit is in explaining variation in the data. If we use the sum of squares due to error to select the best fit, it would be the polynomial function, but if we use the coefficient of determination to select the best fit, it would be the power function. It is for this reason that we have both the polynomial model and power model below to predict the 100s of board feet for trees of age 60, 140, and 180. There is not a significant amount of variation between the two model's predicted quantities which tells us the both could be useful in predicting 100s of board feet from the age of a given tree.

Age of Tree	100s of Board Feet	Polynomial Prediction Equation
20	1	4.089
40	6	3.665
60	*a*	12.041
80	33	29.217
100	56	55.193
120	88	89.969
140	*b*	133.545
160	182	185.921
180	*c*	247.097
200	320	317.073

Age of Tree	100s of Board Feet	Power Prediction Equation
20	1	1.049780781
40	6	5.908074795
60	*a*	16.23196349
80	33	33.25013033
100	56	57.98975186
120	88	91.35207664
140	*b*	134.1497884
160	182	187.1288373
180	*c*	250.9825322
200	320	326.3612723