This is an example of how to use the linear regression calculations shown in class to perform a regression analysis on a set of data points. We will use the same example for each set of calculations.

Problem Statement: We are trying to calibrate a force sensor by measuring the resistance of the sensor when different levels of force are applied. We have recorded the following data points:

Force (N)	**Resistance (M-Ω)**
1	0.9
2	0.6
4	0.3
7	0.2

We need to know: which form of the regression equation (linear, exponential, or power) will best fit the data points?

First, note that *n*, the number of data points, equals 4 for this example.

Linear Form of the Regression Equation

For an equation following the linear form ($y = mx + b$), the slope (m) is calculated as follows:

$$m = \frac{n\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}$$

We need to calculate all the sums that are required by the equation. We'll let "x" represent the independent variable (Force) and "y" represent the dependent variable (Resistance). In the table below, the values in bold type at the bottom of each column are the sums of the values in each column.

Force (N)	**Resistance (M-Ω)**			
x	**y**	**xy**	**x^2**	**y^2**
1	0.9	0.9	1	0.81
2	0.6	1.2	4	0.36
4	0.3	1.2	16	0.09
7	0.2	1.4	49	0.04
14	**2.0**	**4.7**	**70**	**1.3**

Now we can plug the numbers into the equation for slope:

$$m = \frac{4 \cdot 4.7 - 14 \cdot 2.0}{4 \cdot 70 - 14^2}$$

$$m=\frac{18.8-28}{280-196}$$

$$m=-0.110$$

Now we can calculate the intercept (b)

$$b=\frac{\sum y_i - m\left(\sum x_i\right)}{n}$$

$$b=\frac{2.0-(-0.110\cdot 14)}{4}$$

$$b=0.885$$

The best equation in linear form for these data points is therefore:

$$y=-0.110x+0.885$$

Now we need to determine how good a fit the regression equation is for the data that we have. We do this by calculating the coefficient of determination, or r^2, value.

$$r^2=\left[\frac{n\left(\sum x_i y_i\right)-\left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\left(\sum x_i^2\right)-\left(\sum x_i\right)^2}\sqrt{n\left(\sum y_i^2\right)-\left(\sum y_i\right)^2}}\right]^2$$

$$r^2=\left[\frac{4(4.7)-(14)(2.0)}{\sqrt{4(70)-(14)^2}\sqrt{4(1.3)-(2.0)^2}}\right]^2$$

$$r^2=\left[\frac{18.8-28}{\sqrt{280-196}\sqrt{5.2-4}}\right]^2$$

$$r^2=\left[\frac{-9.2}{\sqrt{84}\sqrt{1.2}}\right]^2$$

$$r^2=\left[-0.915\right]^2$$

$$r^2=0.840$$

From the r^2 value, we can see that 84.0% of the variability in resistance ("y") can be explained by the changes in force ("x"). Though this isn't bad, we should check the other forms to see if we can find a better equation that fits our data.

Next, let's try the exponential form ($y = be^{mx}$).

Exponential Form of the Regression Equation

For all the regression equations for slope, intercept, and r^2, we must replace all the "y" terms with "ln(y)". The equation for the slope (*m*) is therefore

$$m = \frac{n\left(\sum x_i \ln(y_i)\right) - \left(\sum x_i\right)\left(\sum \ln(y_i)\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}$$

We now need to recalculate some of the values in our table, as follows:

Force (N)	**Resistance (M-Ω)**				
x	**y**	**ln(y)**	**x*ln(y)**	**x^2**	**$(\ln(y))^2$**
1	0.9	-0.105	-0.105	1	0.011
2	0.6	-0.511	-1.022	4	0.261
4	0.3	-1.204	-4.816	16	1.450
7	0.2	-1.609	-11.266	49	2.590
14		**-3.430**	**-17.209**	**70**	**4.312**

Note that we no longer need the sum of the y terms, because we have replaced "y" with "ln(y)".

Now, we can calculate *m* as follows:

$$m = \frac{4(-17.209) - (14)(-3.430)}{4(70) - (14)^2}$$

$$m = \frac{-68.836 + 48.020}{280 - 196}$$

$$m = -0.248$$

We need to calculate the intercept (*b*), but we must remember that *for the exponential form, we first calculate the natural log of b* (ln(*b*), as follows:

$$\ln(b) = \frac{\sum \ln(y_i) - m\left(\sum x_i\right)}{n}$$

$$\ln(b) = \frac{-3.430 - (-0.248)(14)}{4}$$

$$\ln(b) = 0.0105$$

To calculate b, we take the anti-log of both sides as follows:

$$b = e^{\ln(b)}$$

Therefore,

$$b = e^{0.0105}$$

$$b = 1.011$$

The best equation to fit the data points that follows the exponential form is therefore:

$$y = 1.011e^{-0.248}$$

Now let's determine how well the exponential form of the regression equation fits our data. Again, remember that we have to replace every occurrence of "y" with "ln(y)".

$$r^2 = \left[\frac{n\left(\sum x_i \ln(y_i)\right) - \left(\sum x_i\right)\left(\sum \ln(y_i)\right)}{\sqrt{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}\sqrt{n\left(\sum \ln(y_i)^2\right) - \left(\sum \ln(y_i)\right)^2}} \right]^2$$

$$r^2 = \left[\frac{4(-17.209) - (14)(-3.430)}{\sqrt{4(70) - (14)^2}\sqrt{4(4.312) - (-3.430)^2}} \right]^2$$

$$r^2 = \left[\frac{-68.836 + 48.02}{\sqrt{84}\sqrt{5.483}} \right]^2$$

$$r^2 = \left[\frac{-20.816}{21.461} \right]^2$$

$$r^2 = [-0.970]^2$$

$$r^2 = 0.941$$

Therefore, 94.1% of the variability in resistance can be explained by the changes in force. This value of r^2 is definitely better than the one we calculated for the linear form of the regression equation (0.840), but let's check the power form to see if it will fit our data even better than the exponential form.

Power Form of the Regression Equation

The power form of the regression equation is: $y = bx^m$. To calculate the slope and intercept for this form of the equation, we must take the logarithms of both sets of data ("x" and "y"). Note that are free to use EITHER common OR natural logs; however, we must be consistent in all 3 equations. Since we have already calculated the natural log of the y terms, we'll stick with the natural log.

$$m = \frac{n\left(\sum \ln(x_i)\ln(y_i)\right) - \left(\sum \ln(x_i)\right)\left(\sum \ln(y_i)\right)}{n\left(\sum \ln(x_i)^2\right) - \left(\sum \ln(x_i)\right)^2}$$

The table of values we need is:

Force (N) x	Resistance (M-Ω) y	ln(x)	ln(y)	ln(x)*ln(y)	$(\ln(x))^2$	$(\ln(y))^2$
1	0.9	0.000	-0.105	0.000	0.000	0.011
2	0.6	0.693	-0.511	-0.354	0.480	0.261
4	0.3	1.386	-1.204	-1.669	1.922	1.450
7	0.2	1.946	-1.609	-3.132	3.787	2.590
		4.025	**-3.430**	**-5.155**	**6.189**	**4.312**

$$m = \frac{4(-5.155) - (4.025)(-3.430)}{4(6.189) - (4.025)^2}$$

$$m = \frac{-20.62 + 13.806}{24.756 - 16.201}$$

$$m = -0.796$$

Again, to calculate the intercept, we first calculate the log of the intercept. We will continue to use natural log for this calculation.

$$\ln(b) = \frac{\sum \ln(y_i) - m\left(\sum \ln(x_i)\right)}{n}$$

$$\ln(b) = \frac{-3.430 - (-0.796)(4.025)}{4}$$

$$\ln(b) = -.0565$$

$$b = e^{\ln(b)}$$

Therefore,

$$b = e^{-0.0565}$$

$$b = 0.945$$

The best equation to fit the data points that follows the power form is therefore:

$$y = 0.945x^{-0.796}$$

Finally, to calculate the r^2 value,

$$r^2 = \left[\frac{n\left(\sum \ln(x_i)\ln(y_i)\right) - \left(\sum \ln(x_i)\right)\left(\sum \ln(y_i)\right)}{\sqrt{n\left(\sum \ln(x_i)^2\right) - \left(\sum \ln(x_i)\right)^2}\sqrt{n\left(\sum \ln(y_i)^2\right) - \left(\sum \ln(y_i)\right)^2}} \right]^2$$

$$r^2 = \left[\frac{4(-5.155) - (4.025)(-3.430)}{\sqrt{4(6.189) - (4.025)^2}\sqrt{4(4.312) - (-3.430)^2}} \right]^2$$

$$r^2 = \left[\frac{-20.62 + 13.806}{\sqrt{24.756 - 16.201}\sqrt{17.248 - 11.765}} \right]^2$$

$$r^2 = \left[\frac{-6.814}{\sqrt{24.756 - 16.201}\sqrt{17.248 - 11.765}} \right]^2$$

$$r^2 = \left[\frac{-6.814}{\sqrt{8.555}\sqrt{5.483}} \right]^2$$

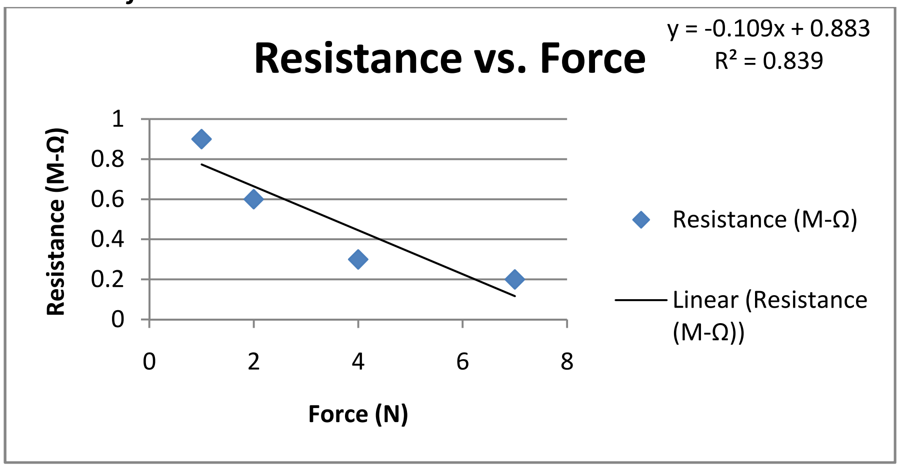
$$r^2 = [-0.995]^2$$

$$r^2 = 0.990$$

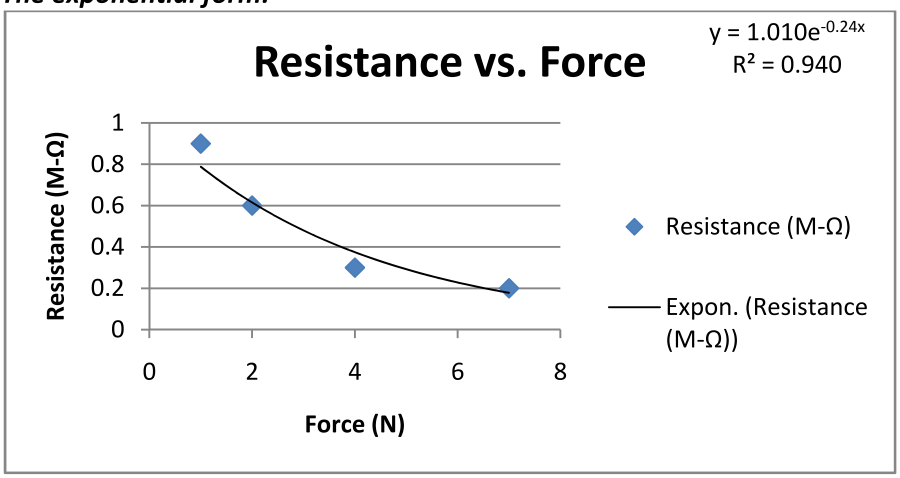
The r^2 value tells us that 99.0% of the variability in resistance can be explained by the changes in force. This is the highest of the three r^2 values that we have calculated. We therefore conclude that the power form of the equation is the best fit for the data points that we recorded.

We can check our work by plotting the data points in Microsoft Excel and adding a trendline to the graph. Note that the values for m, b, and r^2 differ slightly from the values we calculated here because we used fewer significant digits in our calculations than Excel does.

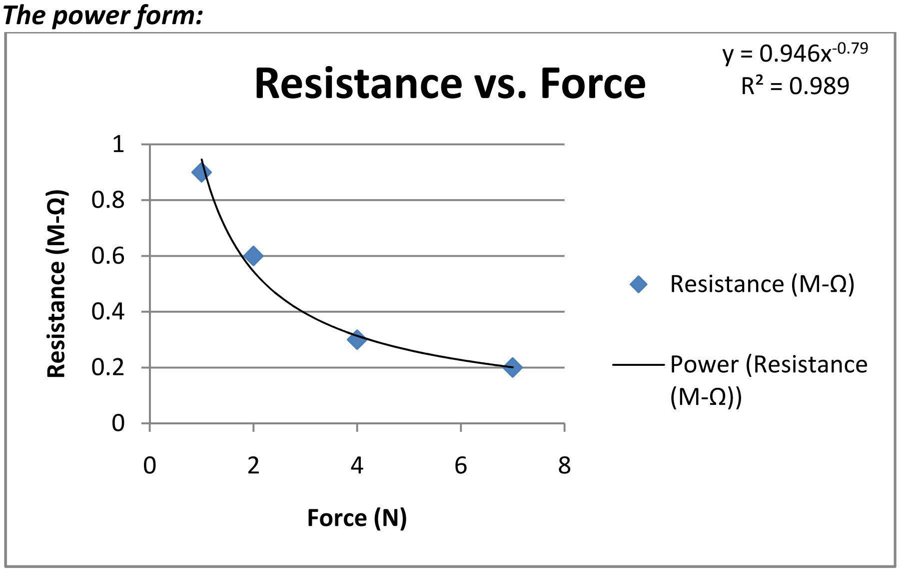
The linear form:



The exponential form:



The power form:



Note that when Microsoft Excel calculates a regression line, the format of the axis (linear or logarithmic) doesn't affect the calculations for the trendline. In the following graph, both axes have been re-formatted to a logarithmic scale. The points appear to line up almost in a straight line, as we would expect if the data points really follow a power form; however, the regression line and r^2 values haven't changed.