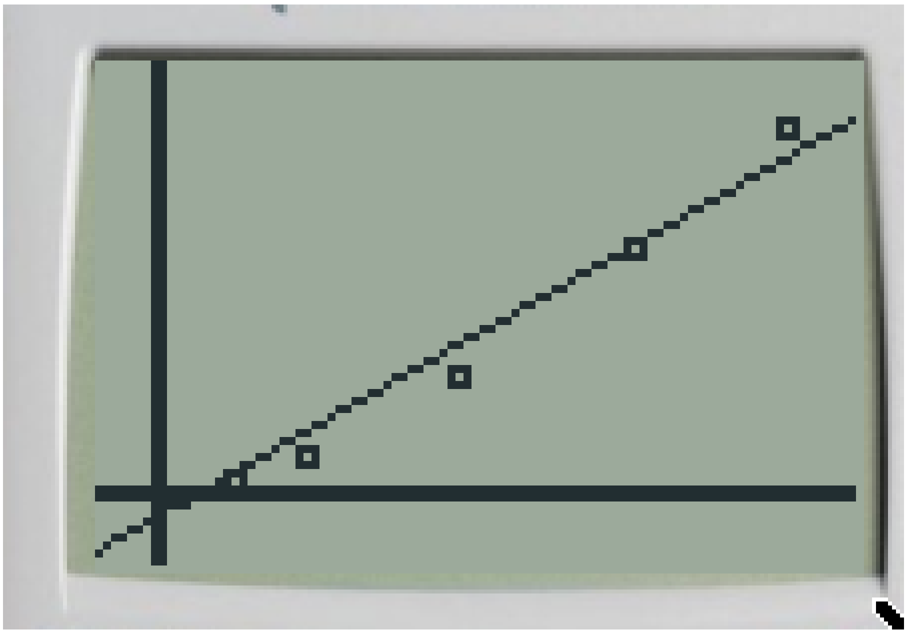
Transformations to Achieve Linearity

1. Draw a scatterplot of Distance vs. Year with the least-squares regression line. Does the line seem to model the relationship well?



Yes, the line does model the relationship quite well.

2. On your calculator, do a linear regression (`STAT` **`CALC`** `8`) for these different combinations. Which transformation yields the highest correlation coefficient? Sketch a scatter-plot of this transformation and show the least-squares line. What is the value of r and r^2 for that transformation, and what regression equation does it yield?

(see graphs at bottom.)

The translation that yields the highest r is the last translation, finding the natural logarithm of both variables. The value of r is 0.999... and r^2 is 0.999... as well. The regression equation is

$$ln(\hat{y}) = -6.80 + 1.50(\ln(x)) \tag{1}$$

3. Using the regression equation from the previous question that *best fits the data*, place the values of the residuals into L5. Create a residual plot on your calculator and interpret it; you don't need to draw the plot.

The residual plot is randomly distributed around $y - \hat{y} = 0$. Therefore, the linear regression line is very appropriate for the data, because of the random distribution and the extremely small size of the residuals themselves.

4. Using algebra, convert your regression equation into a power equation...

$$\hat{y} = e^{-6.8+1.50\cdot\ln(x)} \tag{2}$$

$$\hat{y} = e^{-6.803} \cdot x^{1.499} \tag{3}$$

$$\hat{y} = 0.00111 \cdot x^{1.499} \tag{4}$$

In questions 1-4, we have found the linear relationship between the natural logs of our two variables and derived the power equation from this, using the method of taking e to the power of our regression equation. We can then find a linear equation with good correlation and therefore a power equation with good correlation.

5. The purpose of the transformations you're studying is to find a simple model to describe the relationship in a data set...

$$ln(\hat{y}) = -6.804 + 1.499 \cdot \ln(x) \tag{5}$$

$$\ln(\hat{y}) = 1.193 \tag{6}$$

$$\hat{y} = e^{1.193} = \boxed{3.298} \tag{7}$$

6. Finally calculate the length of the year for 951 Gaspra from the power function you developed in question 4.

$$\hat{y} = 0.00111 \cdot 207.16^{1.499} = \boxed{3.30} \tag{8}$$

7. Make a scatterplot of the *untransformed* data and tell which kind of relationship the points seem to follow. Also name the best type of transformation needed to straighten the plot.

The scatterplot of the untransformed data appears to be logarithmic. Therefore, you take the natural log of the independent/explanatory variable, or the number of years since 1900. Subtracting 1900 from all the years would help as well.

8. Now try some transformations to get the data as close to linear as possible. Then find the regression line, r , and r^2.

Simply plug in the new x to find the new regression equation:

$$\hat{y} = 11.7178 + 14.117 \cdot \ln(x - 1900) \tag{9}$$

(See graphs at bottom.)

9. Using the transformed data and the regression equation for it, create a plot of residuals vs x-values. Sketch the plot and interpret it.

(see graphs below)

The residuals are randomly distributed and close to zero. Therefore, the LRE is a good fit.

10. The data below represents Medicare expenditures from 1970 to 1996, in billions of dollars.

    1. Create a scatterplot of the data

    (see graphs below)

    2. Assume that the relationship of this data is exponential. Transform the data, find the regression equation, r, and r^2. Based only on the value of r would you consider this a good model for extrapolating increases in Medicare spending?

    We can find the transformed variable regression equation to be

$$\hat{y} = -41.334 + 10.234x \tag{10}$$

    The original variable regression equation:

$$\ln(\hat{y}) = -41.334 + 10.234 \cdot ln(x - 1900) \tag{11}$$

$$\hat{y} = e^{-41.334} \cdot e^{10.234 \cdot \ln(x-1900)} \tag{12}$$

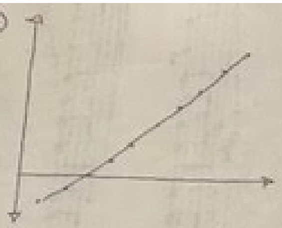
$$\boxed{\hat{y} = 1.120 \cdot 10^{-18}} \tag{13}$$

    r is 0.99 and r^2 is as well. Since the value of r is so high, this is a good model for extrapolating increases in spending in Medicare.

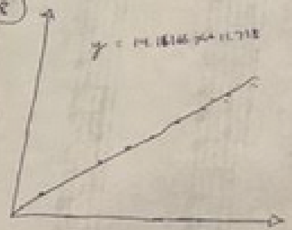
    3. Make a residual plot.

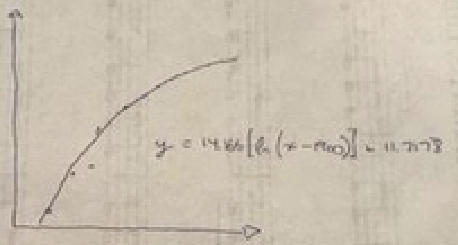
    The residual doesn't change my mind at all, since most of the residuals are really small and randomly distributed.

7

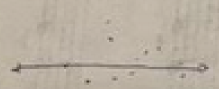


8



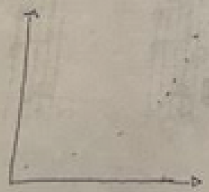


9



10

A



C