How Close Can You Get to the Finish Line?

The Goal: Get your team's car to reach the finish line without going over.
The Catch: The distance will not be revealed until later.

Test drive your car for pull-backs of 2, 4, 6, and 8 inches. Measure the distance the car travels. Repeat this process 3 times. Fill in the table below.

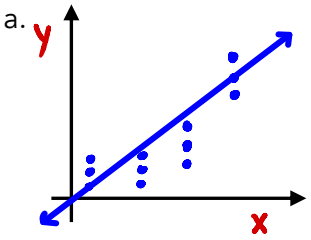
x = Pull-back Distance (in.)	2	2	2	4	4	4	6	6	6	8	8	8
y = Distance Traveled (in.)	4	7	5	12	8	10	20	22	17	45	35	38

Use the applet at statsmedic.com/applets to create three different regression models for the data For each of the three models below:

a. Sketch a scatterplot.
b. Find the equation for the line of best fit, the correlation (r), and r^2.
c. Sketch the residual plot.

Model 1: Plot (*x*, *y*): → LINEAR

a.



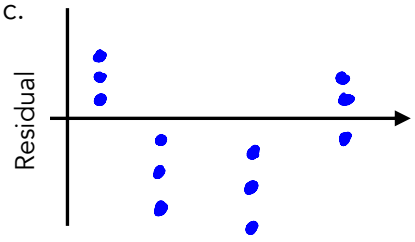
b. Equation:

$\hat{y} = -9.33 + 5.58x$

r: $r = 0.939$

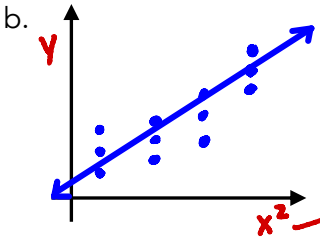
r^2: $r^2 = 0.882$

c.



Model 2: Plot (x^2, *y*): → QUADRATIC

b.



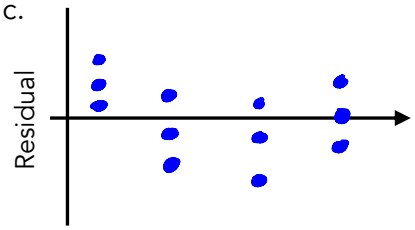
b. Equation:

$\hat{y} = 1.48 + 0.57x^2$

r: $r = 0.974$

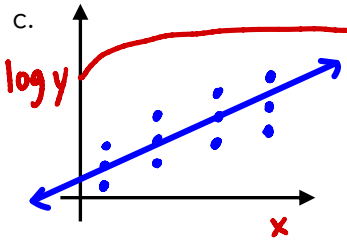
r^2: $r^2 = 0.949$

c.



Model 3: Plot (*x*, log *y*) → EXPONENTIAL

c.



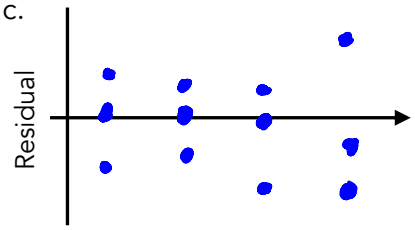
b. Equation:

$\widehat{\log y} = 0.416 + 0.146x$

r: $r = 0.978$

r^2: $r^2 = 0.956$

c.



1. Which model is the best choice for your data? Why?

Model 3. Highest r^2. Residual plot shows no pattern.

2. Calculate how many inches you will have to pull back for a finish line of 80 inches.

$\log(80) = 0.416 + 0.146x$

$1.903 = 0.416 + 0.146x$

$x = 10.2$ inches

STATS MEDIC

Lesson 3.7 – Choosing the Best Regression

QuickNotes

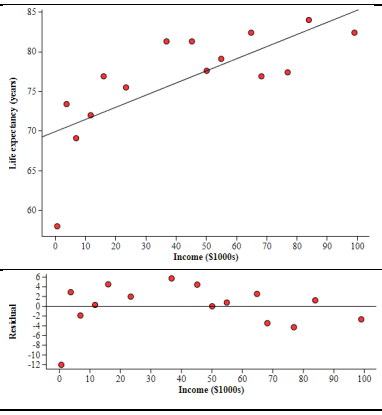
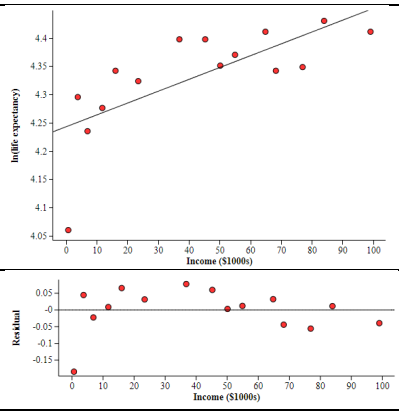
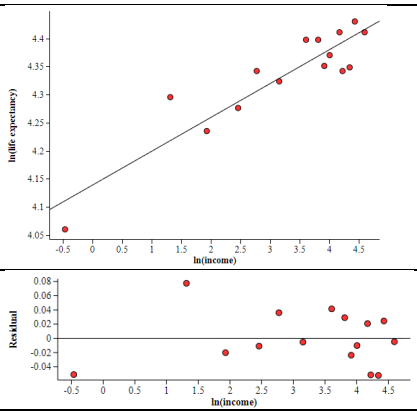
LT#1 Choosing the best regression model

① Check the scatterplot for a linear pattern.
② Check the residual plot for no leftover pattern.
③ Check for the r^2 that is closest to 1.

Check Your Understanding

Gapminder.org is an organization dedicated to studying global health. Here is a scatterplot showing average annual income (in thousands of U.S. dollars) and life expectancy (years) for a random sample of 15 countries selected from the Gapminder data base.

The output shows three possible models for predicting life expectancy from income. Model (a) is based on the original data, while Model (b) and (c) involve transformations of the original data. Each set of output includes a scatterplot with a least-squares regression line added and a residual plot. The regression equations are given below.

Model (a)	Model (b)	Model (c)
		
$\widehat{life\ expectancy} = 69.9352 + 0.1530(income)$	$\widehat{\ln(life\ expectancy)} = 4.2436 + 0.0021(income)$	$\widehat{\ln(life\ expectancy)} = 4.1393 + 0.0603 ln(income)$

1. Use each model to predict the life expectancy of residents of a country for which the average annual income is $80,000.

(a) $\widehat{\text{life expectancy}} = 69.9352 + 0.1530(80) = 82.2$ yrs

(b) $\widehat{\ln(\text{life expectancy})} = 4.2436 + 0.0021(80) = 4.4116$ $e^{4.4116} = 82.4$ yrs

(c) $\widehat{\ln(\text{life expectancy})} = 4.1393 + 0.0603 \ln(80) = 4.4035$ $e^{4.4035} = 81.7$ yrs

2. Which model will make the best predictions for life expectancy? Explain your answer.

Model (c). The scatterplot is the most linear and the residual plot has no leftover pattern.

STATS MEDIC