After watching the video, *Regression,* complete the following problems.

1. It turns out that the time I wake up in the morning is positively correlated with the time I go to sleep at night. Here is a list of my waking and sleeping times during the week. (All times are converted to decimal numbers, so eight-thirty a.m. will be written as 8.5; military time will be used, so 10:15 p.m. will be written 22.25)

Day	Wake Up time	Bed Time
Monday	6.50	23.25
Wednesday	6.25	23.00
Friday	5.75	22.25
Saturday	9.00	26.00 (2 a.m. the next day)

One of the following is the equation of the least squares regression line for my waking and sleeping data. Figure out which one is correct, and explain your reasoning. (Hint: These lines are close enough together that it will be hard to do this by graphing; you will have to think of something else. Pay particular attention to the part of the video where Sonia talks about the squares of the residuals.)

$$y = 1.1x + 15.9$$

$$y = x + 16.1$$

$$y = 1.2x + 15$$

The mean bedtime is 23.25 (11:25 p.m.). We calculate the sums of the squares of the residuals.

	Monday	Wednesday	Friday	Saturday
1.1x + 15.9	$(23.05-23.25)^2$	$(22.775-23)^2$	$(22.225-22.25)^2$	$(25.8-26)^2$
x + 16.1	$(22.60-23.25)^2$	$(22.35-23)^2$	$(21.85-22.25)^2$	$(25.1-26)^2$
1.2 x + 15	$(22.80-23.25)^2$	$(22.50-23)^2$	$(21.9-22.25)^2$	$(25.8-26)^2$

	Monday	Wednesday	Friday	Saturday	Sum of squares
1.1x + 15.9	0.04	0.050625	0.000625	0.04	0.13125
x + 16.1	0.4225	0.4225	0.16	0.81	1.815
1.2 x + 15	0.2025	0.25	0.1225	0.04	.615

The first line is the least squares regression line because the sum of the squares is the smallest value.

2. Recreate the data for height and shoe data for ten people you know (use a single gender).

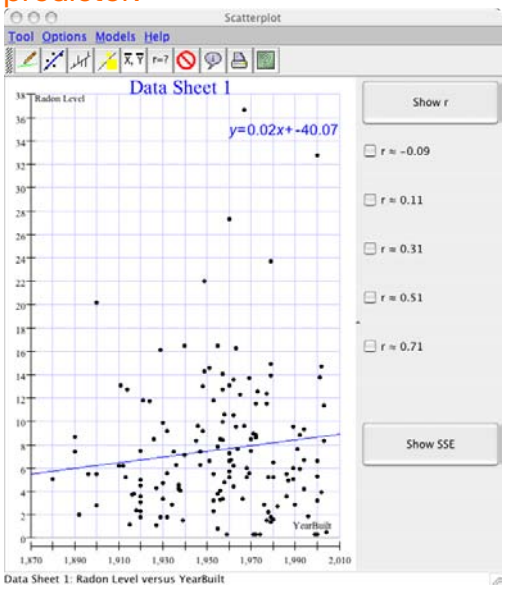
   a. Using technology, find the least squares regression line and interpret the slope and the y-intercept. Here is an online resource that may help. (http://www.shodor.org/interactivate/activities/Regression/)
   Answers will vary but the y-intercept is meaningless.

   b. Find the residual for your shoe size given your height.
   Substitute your height into the equation found in part (a). This is the predicted shoe size. To find the residual subtract your actual shoe size from the predicted shoe size.

   c. Decide whether you think the line will offer a good prediction for your own shoe size.
   The closer the points are to the line, the more likely you are to get a good prediction.

3. Sophomores at Northern University High School have been given the results of radon level for homes tested in the Cedar Valley. They are asked to analyze the data. The scatter plot shows the results of year the homes were built and the radon levels.

   a. The equation of the least-squares regression line is $y = 0.02x - 40.07$. How useful is the line for predicting radon levels in Cedar Valley homes if you know the year the home was built? The regression line is not useful because the points are scattered. It is not a good predictor.



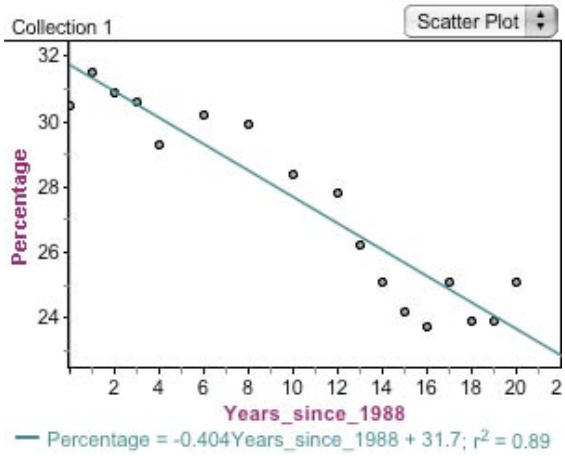
   b. What should the sophomores conclude about the association of year homes were built and radon levels? For this sample of data, there is not an association.

4. **Interpolation** is making prediction between the smallest and the largest data values in the data set. **Extrapolation** is making predictions beyond the data set. Use the following data set for parts a-c.

   Data are reported from 36 participating states from 1988–2008. The proportion of the U.S. population that reported no leisure-time physical activity decreased from about 31% in 1989 to about 28% in 2000. Then the proportion of the population reporting no leisure-time physical activity decreased to about 25% in 2008.
   (Retrieved from http://www.cdc.gov/nccdphp/dnpa/physical/stats/leisure_time.htm)

Year	%
1988	30.5
1989	31.5
1990	30.9
1991	30.6
1992	29.3
1994	30.2
1996	29.9
1998	28.4
2000	27.8
2001	26.2
2002	25.1
2003	24.2
2004	23.7
2005	25.1
2006	23.9
2007	23.9
2008	25.1

   a. Using technology, create a scatter plot and find the least squares regression line.



b. Interpret the slope and the y-intercept.
The slope is -0.404. This is a decrease of about four-tenths of a percent for each year. $-\frac{0.404}{1} \longrightarrow \frac{\text{change in percent}}{\text{per one year}}$
The y-intercept is 31.7. This means that our model predicts in 1988 (year 0), 31% of the population would report no leisure-time physical activity.

c. Is it reasonable to extrapolate the percentage for 2010? If so, make a prediction.
Yes. It is not too far from the last date given in the data. Substitute 22 (number of years since 1988) into the equation given in part (a).
Percentage=-0.404(22)+31.7 or 22.812%

d. Is it reasonable to extrapolate a percentage for 2050? If so, make a prediction.
No, it is too far away from the 1988. The prediction would be about 6.6%. Some people may argue there will be a lower limit that is reasonable for this problem.