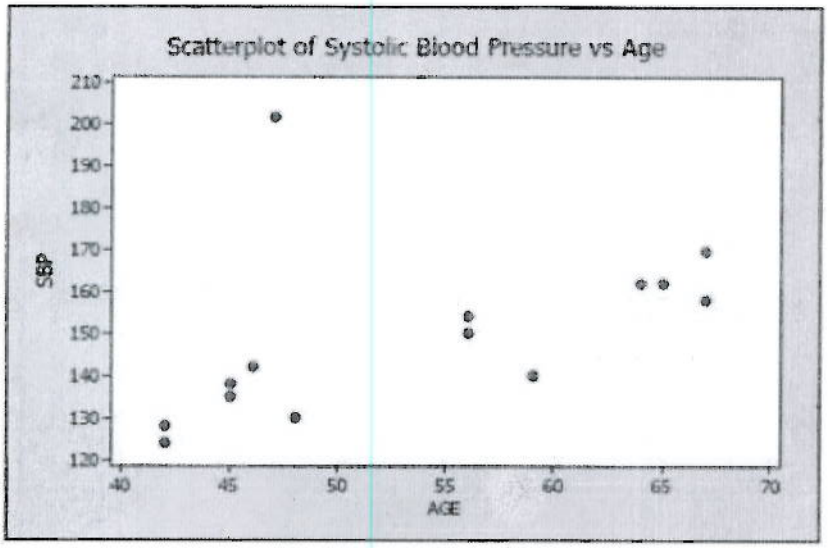
AP Statistics
Lessons 1-3 U3C3 Practice

Name: Pawlson
Period: _____ Date: _____

Linear Regression Models

1. Below is a scatterplot relating systolic blood pressure and age for 14 men from 42 to 67 years old.



a. Is there a clear explanatory variable and response variable in this setting? If so, tell which is which. If not, explain why not. Age is the explanatory variable, systolic blood pressure is the response variable. SBP does not influence age.

b. Does the scatterplot show a positive association, negative association, or neither? What does this tell you about the relationship between age and systolic blood pressure? The association is positive. This suggests that older men are more likely to suffer from high blood pressure.

c. How would you describe the shape, direction and strength of the relationship?
This data is linear, moderate and positive with an outlier at a man around age 47 with a SBP around 200.

d. Which of the following is closest to the correlation between systolic blood pressure and age for this group of 14 men? Explain.

$r = 0.9$ (r = 0.5) $r = 0.2$ $r = -0.2$

Because of the outlier, the r is decreased to around 0.5.

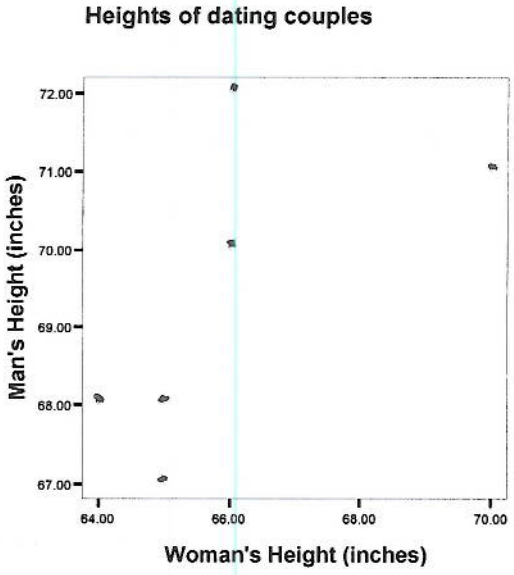
e. There is one "unusual point" on the graph. What effect would removing the "unusual point" have on the correlation? Justify your answer. Removing the outlier would make the correlation much closer to 1.

f. Suppose we rescaled the ages so that they were expressed as number of years above (+) or below (–) 50 years old. That is, suppose we subtract 50 from each value. How would the correlation change? Explain.

This would not change the correlation at all since subtracting 50 from each score would not change its distance from the mean.

2. A student wonders if tall women tend to date taller men than do short women. She measures herself, her roommate, and four friends. Then she measures the next man each woman dates. The results are shown in the table and scatterplot below.

Women (x)	66	64	66	65	70	65
Men (y)	72	68	70	68	71	67



a. Based on the scatterplot, described the pattern, if any, in the relationship between the heights of women and the heights of the men they data. while the scatterplot appears to show that the relationship between heights of woman and the heights of the men are somewhat positive, it does not appear to be a very strong relationship.

b. Use your calculator to find the correlation r between the heights of the men and women. Do the data show any evidence that taller women tend to date taller men? Explain. $r = 0.630$
Since r is positive and greater than 0.5, there is some evidence that tall woman tend to date taller men.

c. How would r change if:

- All the men were six inches shorter than the heights given in the table? Subtracting the same amount from each y-value will not change the correlation.
- Heights were measured in centimeters rather than inches? (There are 2.54 centimeters in an inch.) Multiplying each height by a constant to convert the heights into cm will not change the correlation.

d. Suppose another 70-inch-tall female who dated a 73-inch-tall male were added to the data set. How would this influence r? Adding this point to the data would increase the r-value from .630 to .77, weakening the trend.

3. Below is some data on the relationship between the price of a certain manufacturer's flat- panel LCD televisions and the area of the screen. We would like to use these data to predict the price of televisions based on size.

Screen Area (sq. inches)	Price (dollars)
154	250
207	265
289	330
437	375
584	575
683	650

$r = .979$
$\bar{x} = 392.333$
$s_x = 212.154$
$\bar{y} = 405.833$
$s_y = 163.872$

a. Find and interpret the slope of the least-squares regression line in the context of the problem. Show your work.

$b_1 = .979\left(\frac{163.872}{212.154}\right) = .756$

For every square inch increase in screen size, the price should increase by $.76.

b. Find and interpret the y-intercept of the least-squares regression line in the context of the problem. Show your work.

$b_0 = 405.833 - .756(392.333) =$

109.226

A tv with a screen size of 0 square inches should cost $109.23.

c. Find the equation of the least-squares regression line. Define any variable you use.

$\hat{y} = 109.229 + .756x$

$\hat{y}$ = price (dollars)
x = screen area (sq. inches)

d. Predict the average price for a television with a screen area of 488 square inches. Show your work.

$\hat{y} = 109.229 + .756(488) = 478.157$

$478.16

e. The manufacturer also produces a television with a screen size of 943 square inches. Would it be reasonable to use this equation to predict the price of that television? Explain.

No. 943 is well beyond the range of the screen areas used to produce the LSRL, so this would be extrapolation. We cannot be sure that the relationship described by this line holds outside the range of available data.