INTRODUCTION

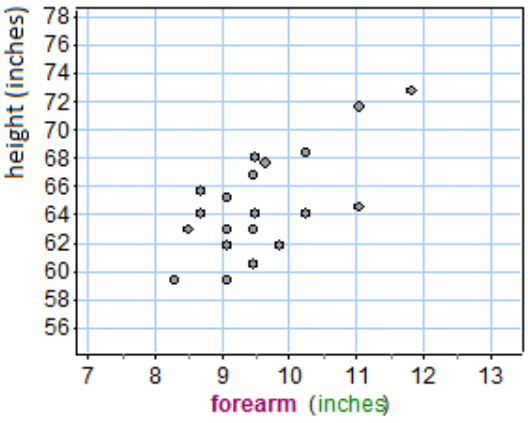
Statistical methods are used in forensics to identify human remains based on the measurements of bones. In the 1950s, Dr. Mildred Trotter and Dr. Goldine Gleser measured skeletons of people who had died in the early 1900s. From these measurements they developed statistical methods for predicting a person's height based on the lengths of various bones. These formulas were first used to identify the remains of U.S. soldiers who died in World War II and were buried in unmarked graves in the Pacific zone. Modern forensic scientists have made adjustments to the formulas developed by Trotter and Gleser to account the differences in bone length and body proportions of people living now. You will not use Trotter and Gleser's formulas in this problem, but you will use a similar process.

Activity #1

Can you identify a female student based on the length of her forearm? The mystery student has a forearm measurement of 10 inches. Height and weight measurements for three female college students are given in the table. Your task is to determine if the mystery student could be one of these 3 students.

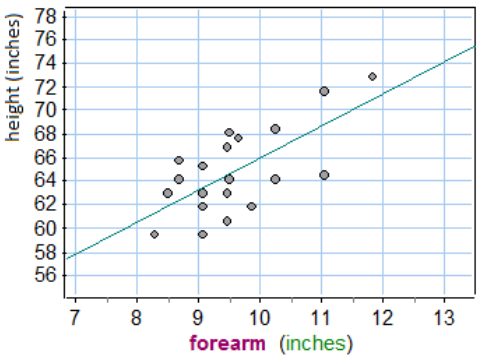
	Jane Doe 1	**Jane Doe 2**	**Jane Doe 3**
Age	18	23	33
Gender	Female	Female	Female
Height	5 feet, 5 inches	5 feet, 2 inches	6 feet
Weight	128 pounds	120 pounds	155 pounds

First, you need data that relates forearm length to either height or weight for females. The scatterplot is a graph of height versus forearm length for 21 female college students taking Introductory Statistics at CCSF in 2009.



1. Based on the scatterplot, what is a reasonable prediction for the height of the mystery student? Briefly explain or show how you made your prediction.

2. The variability in the data makes it difficult to determine if one of these students is the mystery student. Could any of the three students be eliminated as a possibility of being the mystery student? Explain your reasoning.



The scatterplot has a positive linear association. The correlation is 0.68, which is pretty strong. So, it makes sense to use a linear model to summarize the relationship between the forearm and height measurements. There is one line that is considered the best description of how height and forearm length are related. You will learn more about how to find this line later.

3. Use the graph of the best-fit line to predict the height of the mystery student.

4. The equation of this line is approximately ***predicted height*** = 2.7(***forearm length***) + 39.

$$\hat{y} = 2.7x + 39$$

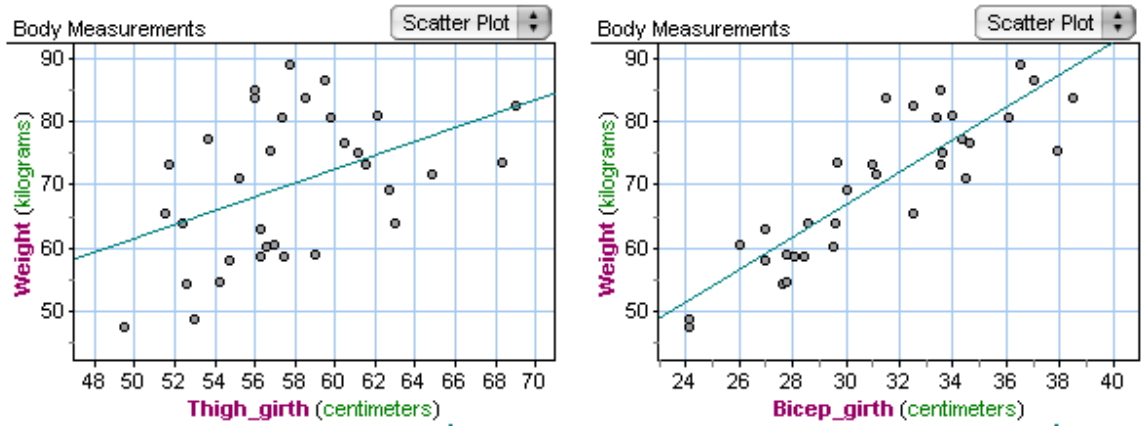
(Notice that when you use letters to represent variables in the prediction line, you put a "hat" on the *y* and write $\hat{y}$ instead of *y*. The hat is a signal that the variable is *predicted* values, not actual data values.)

Use the equation to predict the height of the mystery person.

5. Is the height of Jane Doe 1, 2, or 3 closest to the predicted height of the mystery student given by the line?

Activity #2

The scatterplots below are graphs of body measurements in centimeters for 34 adults who are physically active. These data are a random sample taken from a larger nonrandom data set gathered by researchers investigating the relationship of various body measurements and weight. Girth is the measurement around a body part.
(Retrieved from www.amstat.org/publications/jse/v11n2/datasets.heinz.html)



1. Based on these data, which do you think is a better predictor of an adult's weight: thigh girth or bicep girth? Why?

2. Adriana has a thigh girth of 57 centimeters and a bicep girth of 25 centimeters. Predict her weight using the measurement that you think will give the most accurate prediction, and then plot Adriana on the scatterplot that you used to make her weight prediction.

3. The equations of the two lines shown are

weight = 6.3 + 1.1(***thigh girth***) ***weight*** = –10.5 + 2.6(***bicep girth***)

$\hat{y} = 6.3 + 1.1x$ $\hat{y} = -10.5 + 2.6x$

Predict Adriana's weight using the equation that you think best predicts her weight.

4. Of course, you do not really know Adriana's weight. How accurate do you think the line's prediction of Adriana's weight is? Choose the option that is the most reasonable and explain your thinking.

- Very accurate (within a range of plus or minus 1 kilogram).
- Somewhat accurate (within a range of plus or minus 5 kilograms).
- Not very accurate (within a range of plus or minus 10 kilograms).

In previous lessons, you studied the concept of correlation to describe the strength and direction of the linear association between two quantitative variables. Now you are working on predicting the value of one variable based on the other. Are these two ideas related? Explain your reasoning.