

**Learning Goal:** For a linear relationship, use the least squares regression line to summarize the overall pattern and to make predictions.

**Introduction:** Statistical methods are used in forensics to identify human remains based on the measurements of bones. In the 1950's 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 WWII 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. We will not use Trotter and Gleser's formulas in this problem, but we will use a similar process.

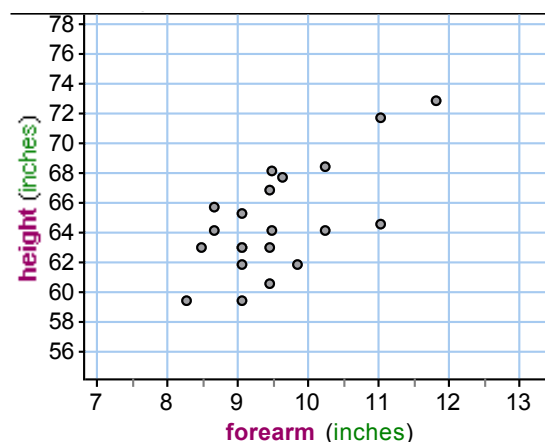
- 1) To illustrate the type of data analysis done in forensics, let's see if we can identify a female student based on the length of her forearm. The mystery student has a forearm measurement of 10 inches. (But she is alive and healthy!)

Height and weight measurements for three female college students are given in the table. Your task is to use data to determine if the mystery student could be one of these three students.

	Jane Doe #1	Jane Doe #2	Jane Doe #3
<b>Age</b>	18	23	33
<b>Gender</b>	Female	Female	Female
<b>Height</b>	5 ft. 5 in	5 ft. 2 in.	6 ft
<b>Weight</b>	128 lb	120 lb	155 lb.

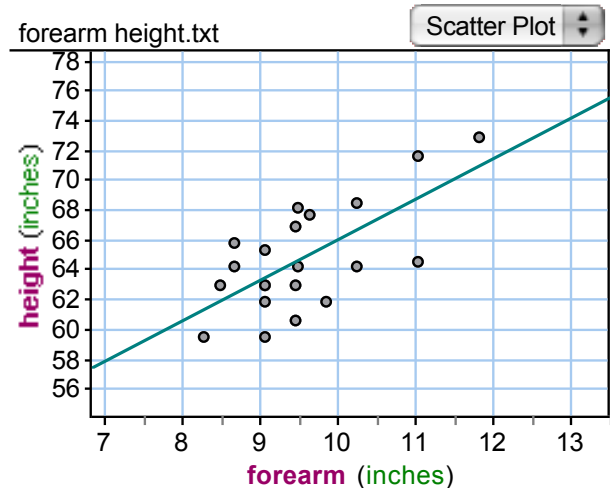
First we need data that relates forearm length to either height or weight for females. The scatterplot is a graph of height versus forearm length for 21 female college students taking Introductory Statistics at Los Medanos College in Pittsburg CA in 2009.

- a) 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.



- b) The variability in the data makes it difficult to determine if one of these students is the mystery student. Could any of the 3 students be eliminated as a possibility of being the mystery student? Why or why not?

- 2) 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. The line shown here is the line that best fits the data. (We will learn more about how to find this line in future lessons.)



- a) Use the graph of the “best fit” line to predict the height of the mystery student. Show or explain how you did this.

- b) The equation of this line is approximately  
 $\text{Predicted height} = 39 + 2.7 (\text{forearm length})$   
Use the equation to predict the height of the mystery person.

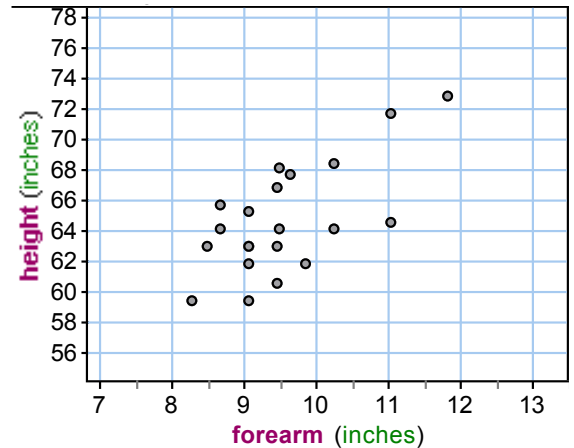
- c) Based on the predictions you have made, which student (1, 2 or 3) do you think might be the mystery student? Why?

Amelia Earhart disappeared in 1937 while flying over the Pacific Ocean. In 1941 bones comprising about a third of a human body were discovered on an uninhabited Pacific island called Nikumaroro. Could these bones be the bones of the aviation pioneer Amelia Earhart?

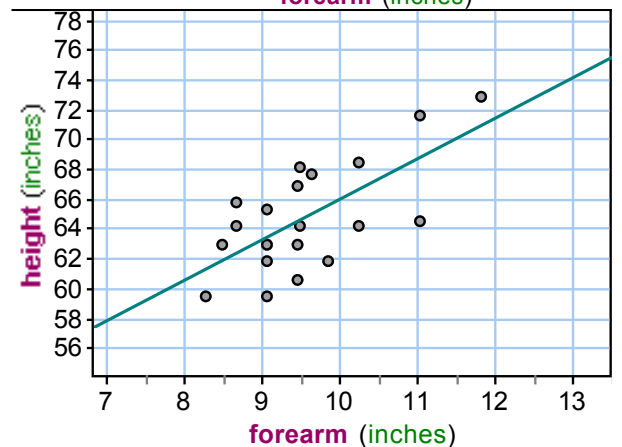
In 1941 an M.D. named Dr. Hoodless measured the bones. The length of the radius, which is the forearm bone, was 9.6 inches. Amelia Earhart said she was 5'8" tall, but other records suggest she might have been closer to 5'7".

3) Here again is a scatterplot of the data collected from 21 female college students.

- a) Plot Amelia on the scatterplot. Based on the scatterplot, do you think that the bones found on Nikumaroro could be Amelia Earhart? Why or why not?



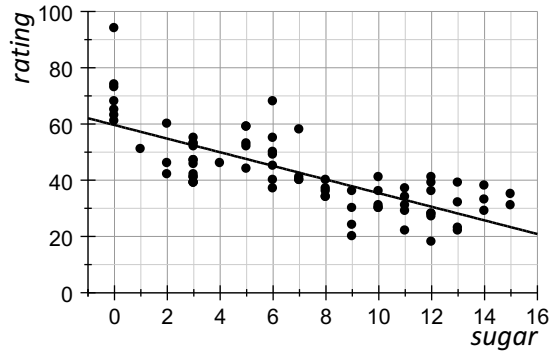
- b) Use the line to predict the height of the person whose bones Dr. Hoodless measured. Plot this person in the scatterplot.  
 $height = 39 + 2.7 (forearm)$



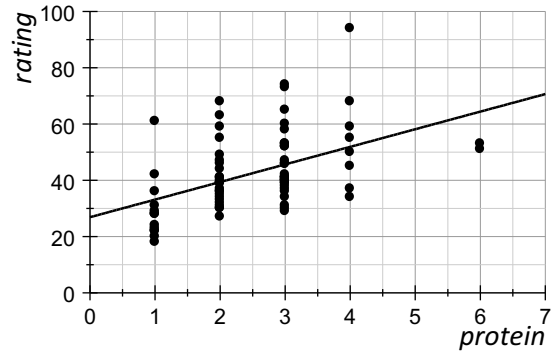
- c) Based on this investigation, do you think the bones could belong to Amelia Earhart? Why or why not?

(Dr. Hoodless concluded that the bones belonged to a short, stocky European male. The bones are now missing, but modern analysis of the notes taken by Dr. Hoodless has revealed many discrepancies and his conclusions are hotly debated.)

- 4) Consumer Reports is a non-profit organization that rates products to help consumers make smart choices. Here are the Consumer Reports ratings for 77 breakfast cereals graphed in relation to sugar and protein amount in a serving. Also shown are the equations and the graphs for the line of best fit.



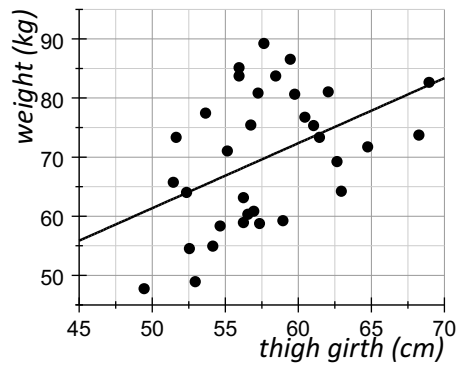
$$\text{Predicted rating} = 60 - 2.43 \cdot (\text{sugar})$$



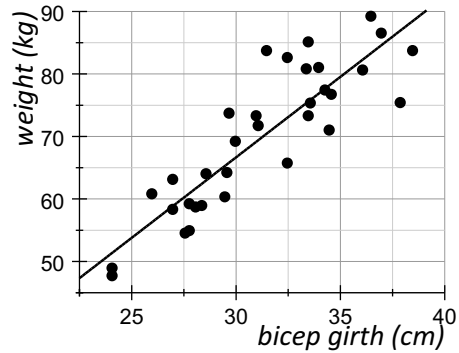
$$\text{Predicted rating} = 28 + 5.96 \cdot (\text{protein})$$

- a) Cereal A has 10.5 grams of sugar in a serving and Cereal B has 2.5 grams of protein in a serving. Using the lines of best fit, predict the Consumer Reports rating for the two cereals.
- b) For which cereal do you think your prediction is probably more accurate? That is, for which cereal do you think your prediction is likely be closer to the actual *Consumer Reports* rating? Why?

- 5) The scatterplots below show body measurements in centimeters for 34 adults who are physically active. The line graphed in each scatterplot is the line of best fit. The equation for each line is given below the scatterplot.



$$Weight = 7.85 + 1.07(\text{thigh girth})$$



$$Weight = -13.45 + 2.67(\text{bicep girth})$$

- a) Adriana has a thigh girth of 57 centimeters and a bicep girth of 25 centimeters. Predict Adriana's weight using both measurements. Show or explain your process.
- b) Which prediction do you think is probably more accurate? Why?