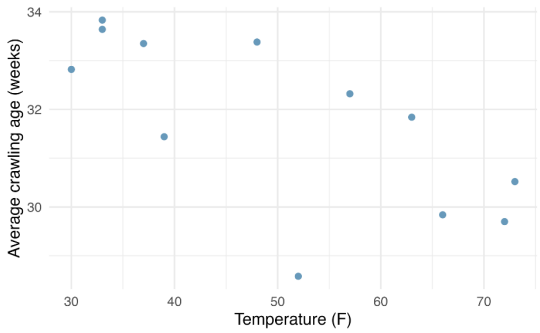
SDS 220 - Lecture 7 Handout

IMS Section 7.1

1. **[IMS 7.12] Crawling babies, correlation.** A study conducted at the University of Denver investigated whether babies take longer to learn to crawl in cold months, when they are often bundled in clothes that restrict their movement, than in warmer months. Infants born during the study year were split into twelve groups, one for each birth month. We consider the average crawling age of babies in each group against the average temperature when the babies are six months old (that's when babies often begin trying to crawl). Temperature is measured in degrees Fahrenheit (F) and age is measured in weeks (Benson 1993)



   (a) Describe the relationship between temperature and crawling age.

   (b) How would the relationship change if temperature was measured in degrees Celsius (C) and age was measured in months?

   (c) The correlation between temperature in F and age in weeks was $r = -.70$. If we converted the temperature to C and age to months, what would the correlation be?

(a) The relationship is moderate, negative, and linear. There is also an outlying month when the average temperature is about 53F and average crawling age of about 28.5 weeks. (c) Changing the units will not change the form, direction or strength of the relationship between the two variables. After all, if higher temperatures measured in F are associated with lower average crawling age measured in weeks, higher temperatures measured in C will be associated with lower average crawling age measured in months. (c) Since changing units doesn't affect correlation, $r = -.7$.

2. **[IMS 7.13] Partners' ages**. What would be the correlation between the ages of partners if people always dated others who are

   (a) 3 years younger than themselves?

   (b) 2 years older than themselves?

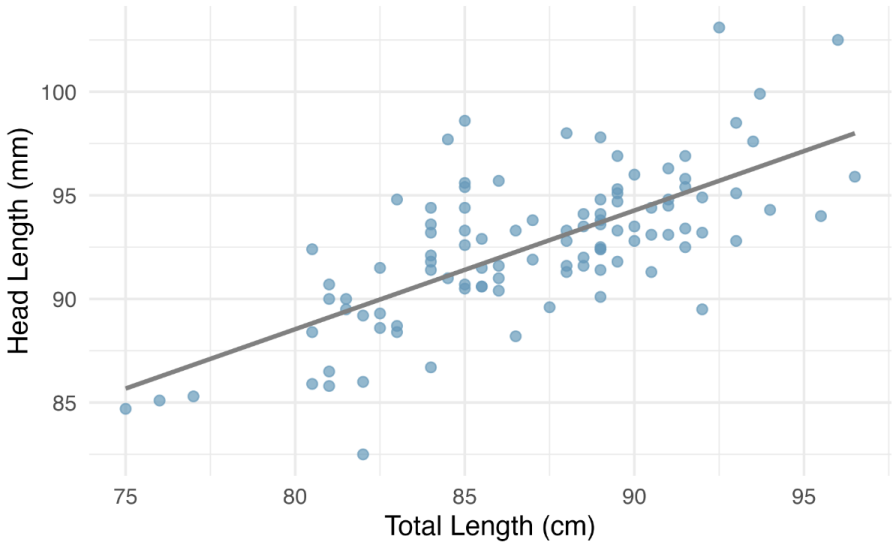
   (c) half as old as themselves?

In each part, we can write the age of one partner as a linear function of the other. Since the slopes are positive and these are perfect linear relationships, the correlation will be exactly 1 in all three parts.

3. **[IMS 7.17] Over-under, I**. Suppose we fit a regression line to predict the shelf life of an apple based on its weight. For a particular apple, we predict the shelf life to be 4.6 days. The apple's residual is -0.6 days. Did we over or under estimate the shelf-life of the apple? Explain your reasoning.

   Over-estimate. Since the residual is calculated as observed-predicted, a negative residual means that the predicted value is higher than the observed value.

4. **[Adapted from IMS 7.1]**. Brushtail possums are marsupials that live in Australia. Researchers captured 104 of these animals and took body measurements before releasing the animals back into the wild. We consider two of these measurements: the total length of each possum (`total`), from head to tail, and the length of each possum's head (`total`). The researchers then fitted a linear regression line $\widehat{\texttt{head}} = 41 + .59\texttt{total}$.



   (a) Name two other variables that may also explain head length.

   (b) If a new possum had a `total` length of 80, what would the corresponding predicted head length be?

   (c) If a new possum had a `total` length of 110, what would the corresponding predicted head length be?

   (d) Do you believe the two predicted head lengths above are equally valid? Explain.

   (e) There are three possums that have a total length of less than 80. What are their residuals?

   (f) Describe what you think the corresponding residual plot would look like. Are the points uniformly spread out? Do the points have a pattern?

   (g) What do you think the corresponding correlation would be for these variables?

   (a) Varies. (b) 88.2. (c) 105.9. (d) 110 is outside the range of the data collected, we do not know if the linear line is still appropriate outside the range of data collected. (e) The points are at approximately (75, 84.7), (77, 85.3), and (76, 85.1). The fitted values are 85.25, 86.43, 85.84. Hence the residuals are -0.55, -1.13, -0.74. (f) Somewhat random evenly scattered, there is a slight fan shape with values being more spread out as total length increases. (g) Negative, moderately correlated.