Unit 2 Homework

Assignment 1

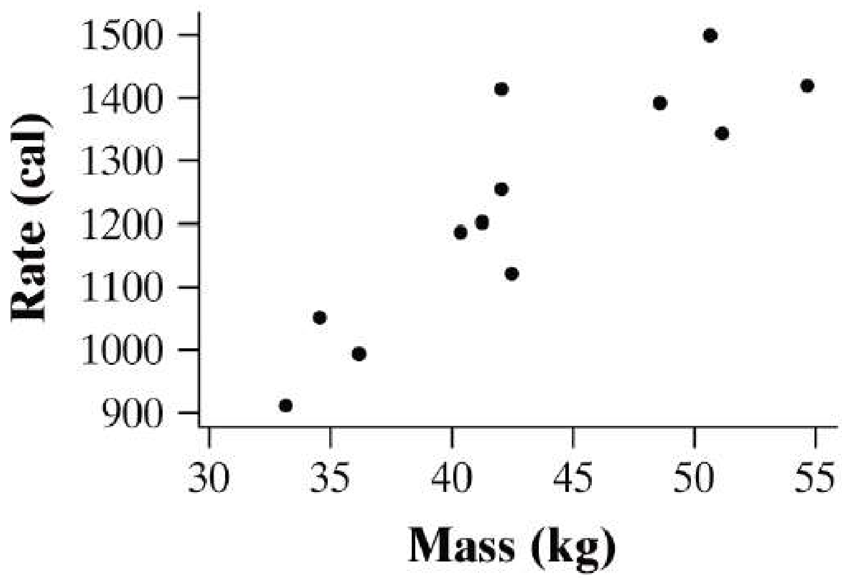
3.1 Water temperature is the explanatory variable, and weight change (growth) is the response variable. Both are quantitative.

3.2 The explanatory variable is the type of treatment—removal of the breast or removal of only the tumor and nearby lymph nodes, followed by radiation, and survival time is the response variable. Type of treatment is a categorical variable, and survival time is a quantitative variable.

3.3 (a) A positive association between IQ and GPA means that students with higher IQs tend to have higher GPAs, and those with lower IQs generally have lower GPAs. The plot does show a positive association. (b) The form of the relationship is roughly linear, because a line through the scatterplot of points would provide a good summary. The positive association is moderately strong (with a few exceptions) because most of the points would be close to the line. (c) The lowest point on the plot is for a student with an IQ of about 103 and a GPA of about 0.4.

3.4 (a) The plot shows a negative association. This is what we would expect because as the temperature gets warmer, she would not need to heat her house as much, so her use of gas would decline. (b) The scatterplot shows a strong linear relationship. It is linear because a line through the scatterplot of points would provide a good summary and it is strong because the points would all be close to the line. (c) The point in the bottom right of the plot represents a month when the average temperature was about 58 degrees and the gas usage was about 260 cubic feet.

3.10 (a) A scatterplot with mass as the explanatory variable is shown below.



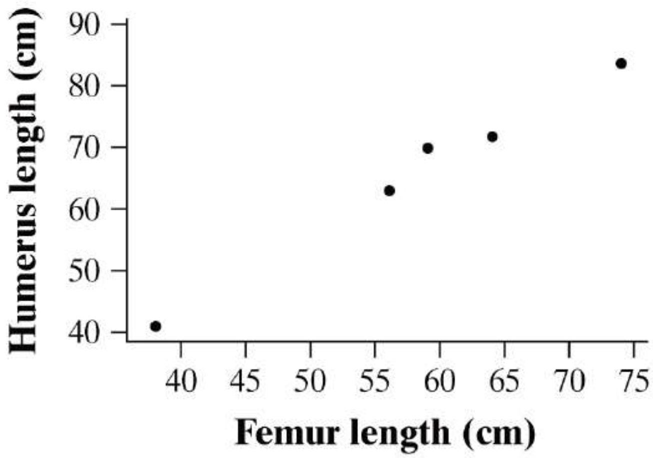
(b) The association is positive, and the relationship is linear and moderately strong.

3.12 The scatterplot shows that the pattern of the relationship does hold both for men and women. However, the relationship between mass and rate is not as strong for men as it is for women. The group of men has higher lean body masses and metabolic rates than the group of women.

3.15 (a) $r = 0.9$ (b) $r = 0$ (c) $r = 0.7$ (d) $r = -0.3$ (e) $r = -0.9$

3.17 (a) Gender is a categorical variable and the correlation coefficient r measures the strength of linear association for two quantitative variables. (b) The largest possible value of the correlation coefficient r is 1. (c) The correlation coefficient r has no units.

3.19 (a) The scatterplot below shows a strong, positive, linear relationship between the two measurements. Thus, all five specimens appear to be from the same species.



femur	Humerus	zfemur	zhumerus	product
38	41	-1.53048	-1.57329	2.40789
56	63	-0.16669	-0.18880	0.03147
59	70	0.06061	0.25173	0.01526
64	72	0.43944	0.37759	0.16593
74	84	1.19711	1.13277	1.35605

(b) The femur measurements have mean of 58.2 and a standard deviation of 13.2. The humerus measurements have a mean of 66 and a standard deviation of 15.89. The table below shows the standardized measurements (labeled zfemur and zhumerus) obtained by subtracting the mean and dividing by the standard deviation. The column labeled "product" contains the product (zfemur×zhumerus) of the standardized measurements. The sum of the products is 3.97659, so the correlation coefficient is $r = \frac{1}{4}(3.97659) = 0.9941$.

3.21 (a) There is a strong, positive linear association between sodium and calories. High-calorie hot dogs tend to be high in sodium, and low-calorie hot dogs tend to be low in sodium.
(b) The hot dog with the lowest calorie content increases the correlation. It falls in the linear pattern of the rest of the data and observations with unusually small or unusually large values of x have a big influence on the correlation.

3.23 (a) The correlation would not change, because correlation is not affected by a change of units for either variable. In terms of the formula for correlation, multiplying both the x and y values by 10 will also multiply their standard deviations by 10, so the z-scores will not change.
(b) The correlation would not change. The correlation measures the strength of the linear relationship between two quantitative variables. It does not distinguish between the explanatory and response variables. In terms of the formula for correlation, reversing the x and y values won't change the correlation because $z_x z_y$ is the same as $z_y z_x$.