

Solutions of homework 6

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

4.6 (a) Graph is omitted. Speed is explanatory.

(b) The relationship is curved – low in the middle, higher at the extremes. Because low “mileage” is actually *good* (it means that we use less fuel to travel 100 km), this makes sense: moderate speeds yield the best performance. Note that 60 km/hr is about 37 mph.

(c) Above-average (that is, bad) values of “fuel used” are found with both low and high values of “speed”.

(d) The relationship is very strong – there is little scatter around the curve, and it is very useful for prediction.

4.8 (a) Because there is no obvious choice for the responsive variables, either could go on the vertical axis. The plot (omitted) shows a strong positive linear relationship, with no outliers. There appears to be only one species represented.

(b) $\bar{x} = 58.2$ cm and $s_x \approx 13.20$ cm (for the femur measurements); $\bar{y} = 66$ cm and $s_y \approx 15.89$ cm (for the humerus). The standard values for z_x are -1.53048 , -0.16669 , 0.06061 , 0.43944 , 1.19711 ; for z_y are -1.57329 , -0.18880 , 0.25173 , 0.37759 , 1.13277 ; for $z_x z_y$ are 2.40789 , 0.03147 , 0.01526 , 0.16593 , 1.35605 . The correlation $r = 3.97659/4 = 0.994$.

(c) Obviously, the calculator value should be the same.

4.11 $\bar{x} = 40$ and $s_x \approx 15.8114$ mph; $\bar{y} = 26.8$ cm and $s_y \approx 2.6833$ mph. The denominators are all the same $s_x s_y$. The numerator of the terms is $\sum (\bar{x} - x_i)(\bar{y} - y_i) = (-20)(-2.8) + (-10)(1.2) + (0)(3.2) + (10)(1.2) + (20)(-2.8) = 56 - 12 + 0 + 12 - 56 = 0$. The correlation is 0 because these variables do not have a straight-line relationship; the association is neither positive nor negative.

4.13 (a) A positive association between IQ and GPA would mean 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 relation is positive, roughly linear, and moderately strong (except for three outliers).

(c) The lowest point on the plot is for a student with an IQ 103 and a GPA of about 0.5.

4.17 (a) The correlation for the data in figure 4.6 is positive but not near 1; the plot clearly shows a positive association, but with quite a bit of scatter (even if we ignore the three outliers).

(b) The correlation for the data in figure 4.7 is closer to 1 because the spread is considerably less in this scatter plot (and the outlier in the scatter plot strengthens the positive association).

4.25 (a) Yes, planting rate is explanatory.

(b) Graph is omitted.

(c) The pattern is curved – high in the middle, and lower on the ends. The association is not linear, and is neither positive nor negative.

(d) The means are 131.025, 143.15, 146.225, 143.06 and 134.75 bushels/acre. The mean yields first increase with plant density, then decrease; the greatest yield occurs at or around 20,000 plants per acre.

4.28 The person who wrote the article interpreted a correlation close to 0 as if it were a correlation close to -1 (implying a negative association between teaching ability and research productivity). Professor McDaniel's findings mean there is little linear association between teaching and research – for example, knowing that a professor is a goof researcher gives little information about whether she/he is a good or bad teacher.

5.2 (a) The slope is 0.0138 minutes per meter. On the average, if the depth of the dive is increased by one meter, it adds 0.0138 minutes (about 0.83 seconds) to the time spent underwater.

(b) When $D = 200$, the regression formula estimates DD to be 5.45 minutes.

(c) To plot the line, compute $DD = 3.242$ minutes when $D = 40$ meters, and $DD = 6.83$ minutes when $D = 300$ meters. Graph is omitted.

5.7 (a) Graph is omitted.

(b) No, the pattern is curved. Therefore, a linear formula is not the appropriate choice for prediction.

(c) The sum is actually -0.01 .

(d) The first two and last four residuals are positive, and those in the middle are negative. Graph is omitted.

5.28 (a) The slope is $b = r * s_x / s_y = 0.6 * 8 / 30 = 0.16$, and the intercept $a = \bar{y} - b\bar{x} = 30.2$.

(b) Julie's predicted score is $\hat{y} = 78.2$.

(c) $r^2 = 0.36$; only 36% of the variability in y is accounted for by the regression, so the estimated $\hat{y} = 78.2$ could be quite different from the real score.

5.29 $r = \sqrt{0.16} = 0.40$ (high attendance goes with high grades, so the correlation must be positive).