7.14

(a)

There is a somewhat not strong, negative, possibly linear relationship between the temperature and the crawling age. There is one point that is about (52F, 28.5 weeks), which deviates from the possible linear regression line and should be especially considered.

(b)

Changing the units will not change the form, direction or strength of the relationship between the two variables. If higher temperatures measured in Fahrenheit are associated with smaller crawling age measured in weeks, higher temperatures measured in Celsius will be associated with smaller crawling age measured in months.

(c)

Changing units doesn't affect correlation. So r = -0.70.

7.24

(a)

There is a positive, moderate, linear association between the number of calories and amount of carbohydrates that Starbucks food menu items contain.

(b)

Explanatory: number of calories. Response: amount of carbohydrates that Starbucks food menu items contain (in grams).

(c)

We can predict the amount of carbohydrates a menu item has based on its calorie content using a regression line. This may be useful information for judging whether more calories mean more carbohydrates ingestion from the Starbucks food. This might help people of diabetes to choose lower calories food for lower carbohydrates ingestion.

(d)

Even though the relationship appears linear in the scatterplot, the residual plot actually shows a nonlinear relationship. This is not a contradiction: residual plots can show divergences from linearity that can be difficult to see in a scatterplot.

Analysis of conditions for the least squares line:

For linearity, the data show a moderate linear trend in general, so it satisfies linearity.

The residuals are not quite normal, so it does not perfectly satisfy the requirement of normal distribution.

The variability of points around the line are not constant. As the calories increase, the variability of the residuals becomes larger, which does not satisfy the condition.

A simple linear model is inadequate for modeling these data.

It is also important to consider that these data are observed sequentially, which means there may be a hidden structure not evident in the current plots but that is important to consider.

7.28

(a)

$R^2 = 0.72$, R = -0.849

The correlation between lunch and helmet is -0.849.

(b)

Slope = $b_1 = \frac{s_y}{s_x} R = \frac{0.169}{0.267} * (-0.849) = -0.537$

Intercept = $b_0 = \bar{y} - b_1 \bar{x} = 0.388 - (-0.537) * 0.308 = 0.553$

(c)

The intercept in this scenario means when receiving reduced-free lunch of children in a neighborhood is zero, there may be 55.3% of bike riders on average in the neighborhood wearing helmets.

(d)

The slope in this scenario means if the percentage of receiving reduced-free lunch of children in a neighborhood increase by 1%, the percentage of bike riders in the neighborhood wearing helmets will decrease by 0.537% on average.

(e)

$$y_i = b_0 + b_1 x_i = 0.553 - 0.537 * 0.4 = 0.3382$$
$$y_i = 0.4$$
$$e_i = y_i - y_i = 0.4 - 0.3382 = 0.0618 = 6.18\%$$

The value of the residual is 6.18%. The residual means that for a neighborhood where 40% of the children receive reduced-free lunches, the actual percentage of bike riders wearing helmets (40%) is 6.18% greater than the estimated percentage based on the model line. It is estimated that 33.82% will wear helmets of the bike riders.

7.30

(a)

$$heartweight(\text{g}) = -0.357 + 4.034 bodyweight(\text{kg})$$

(b)

Expected heart weight of cats with zero body weight is -0.357g. This is obviously not a meaningful value; it just serves to adjust the height of the regression line.

(c)

For each additional 1kg increase in body weight of cats, we expect the heart weight to be higher on average by 4.034g.

(d)

Body weight of cats explains 64.66% of the variability in heart weight of cats.

(e)

The correlation coefficient is $\sqrt{0.6466} = 0.8041$.

7.36

(a)

The relationship is positive, moderate-to-strong, and linear. There is no apparent outlier. The last point seems too high, but it does not deviate too much from the regression line. The scatter plot is a bit sparse, but the trend is still obvious.

(b)

$$BAC(\mathrm{g/\ d}) = -0.0127 + 0.0180 * cansofbeer$$

Slope: For each additional can of beer, the model predicts the average BAC to be 0.0180 (g/d) higher.

Intercept: If no beer is drunk, the BAC will be -0.0127. This is obviously not possible. BAC cannot be negative. Here, the y-intercept serves only to adjust the height of the line and is meaningless by itself.

(c)

H_0: The true slope coefficient of BAC is zero (β_1 = 0).

H_A: The true slope coefficient of BAC is greater than zero (β_1 > 0). A two-sided test would also be acceptable for this application. The p-value for the two-sided alternative hypothesis ($\beta_1 \neq 0$) is incredibly small, so the p-value for the one-sided hypothesis will be even smaller. That is, we reject H_0. The data provide convincing evidence that BAC and cans of beer are positively correlated. The true slope parameter is indeed greater than 0.

(d)

$R^2 = 0.89^2 = 0.7921$. Approximately 79.21% of the variability in BAC can be explained by the number of cans of beer.

(e)

The relationship may not be as strong as that found in the Ohio State University. Because people in the bar probably do not drink randomly numbers of cans of beer. The numbers of women and men may not be equal either. They may also not be quite independent with each other. Their weights and drinking habits may not vary randomly. So there may be some clusters in the scatter plot.

7.42

(a)

Predicted head circumference $= 3.91 + 0.78 * 28 = 25.75$ cm.

(b)

If we are not sure whether the association exists or not, we can use two tails t value to compute the confidence interval. $t_{23}^{*} = 2.07$ for 95% confidence interval. CI: (0.78-2.07*0.35, 0.78+2.07*0.35) , i.e. (0.0555, 1.5045) Since the 95% confidence interval does not include 0, the model provides strong evidence that gestational age is significantly associated with head circumference.

7.44

(a)

H_0: $\beta_1 = 0$; H_A: $\beta_1 > 0$

(whether positively associated)

(b)

The p-value for the two-sided test is approximately 0, so for one-sided test the p-value is much smaller. Therefore, we reject H_0. The data provide convincing evidence that body weight is a significant predictor of heart weight for cats.

(c)

n =144; df = 142; t^*_{142} = 1.66 (one tail for 95%);

95% confidence interval: 4.034 $\pm$ 1.66*0.25 = (3.619, 4.449)

For each additional increase of body weight (1kg), the heart weight is expected to be higher on average by 3.619 to 4.449 grams.

(d)

Yes, we rejected H_0 and the confidence interval is greater than 0.

8.2

(a)

$$birthweight = 120.07 - 1.93 * parity$$

(b)

The slope means that the estimated body weight of babies of first born is 1.93 ounces greater than babies of other parities. First born: 120.07 ounces. Others: 120.07-1.93 = 118.14 ounces.

(c)

H_0: β_1 = 0. H_A: $\beta_1 \neq 0$. The p-value is greater than 0.1. So we cannot reject H_0. The data does not provide strong evidence that the true slope parameter is different than 0. There is no significant relationship between the average birth weight and parity.

8.4

(a)

$$days = 18.93 - 9.11 * eth + 3.10 * sex + 2.15 * lrn$$

(b)

β_{eth}: The model predicts 9.11 days' decrease in the absent days for not aboriginal students, all else held constant.

β_{sex}: The model predicts 3.1 days' increase in the absent days for male students, all else held constant.

β_{lrn}: The model predicts 2.15 days' increase in the absent days for slow learners, all else held constant.

(c)

$$y_i = 18.93 + 3.1 + 2.15 = 24.18$$
$$y_i = 2$$
$$e_i = y_i - y_i = 2 - 24.18 = -22.18$$

(d)

$$Var(e_i) = 240.57$$
$$Var(\mathrm{y}_i) = 264.17$$
$$R^2 = 1 - \frac{Var(e_i)}{Var(\mathrm{y}_i)} = 1 - \frac{240.57}{264.17} = 0.089$$
$$R^2_{adj} = 1 - \frac{Var(e_i)}{Var(\mathrm{y}_i)} \times \frac{n-1}{n-k-1} = 1 - \frac{240.57}{264.17} \times \frac{146-1}{146-3-1} = 0.07$$