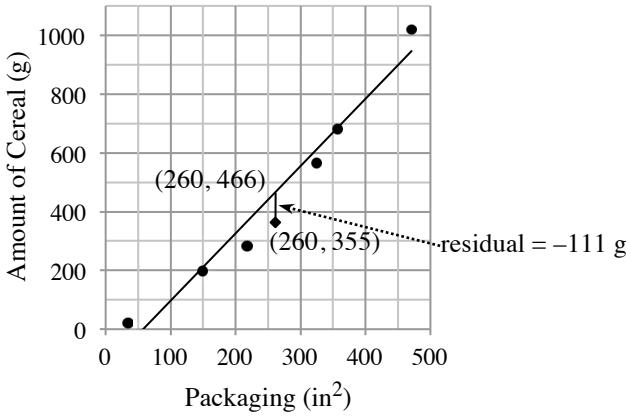
Lesson 2.1.3

2-19. Scatterplot can be found following problem 2-20. Association is a little curved, positive, moderately strong, with no outliers. One possibility for a linear equation is $\hat{m} = -132 + 2.3p$ where m is the amount of the cereal in grams and p is the amount of packaging used in square inches. Cereal in the new package is predicted to weigh ≈ 466 grams.

2-20. Teams will have different predicted values from problem 2-19. Possible answer is residual = 355 – 466 = –111 grams. A positive residual means the actual amount the box can hold is more than was predicted by the model; a negative means the box can actually hold less than predicted. See graph at right.



2-21. Teams will have different equations from problem 2-19. Possible solution is predicted amount = –132 + 2.3(471) = 951 g; residual = 1020 – 951 = 69 g

2-22. a. ≈ 1005 + (–132 + 2.3(600)) ≈ 1005 + 1248 ≈ 2253 grams

b. The large residual means the model is very wrong at that point. Statistical models often cannot be extrapolated. There is no reason to believe that a very large box will behave like smaller boxes because we have no observations near 600 in^2 to base our model on.

c. Slope = 2.3; An increase of 1 square inch in packaging cardboard will result in a predicted increase of 2.3 grams of cereal.
y-intercept = –132; Using no cardboard will result in a negative amount of cereal. That does not make sense. Statistical models often cannot be extrapolated.

2-23. a. The actual amount of sugar is less than predicted by the model; negative is better.

b. Slope = 0.23; An increase of one calorie will result in a predicted increase of 0.23 grams of sugar. *y*-intercept = –16.9; A cereal with no calories will have negative grams of sugar. Again, the statistical model cannot be extrapolated.

2-24. The predicted price for a 2800 sq ft home in Smallville is $264,800 while in Fancyville it is $804,400. The selling price is much closer to what was predicted in Smallville, so she should predict that the home is in Smallville.

2-25. a. Several fit lines are possible. Using the sample model of $y = 32 - 2.0x$, the slope of –2.0 means that for each additional second it takes a player to complete the 100 m dash, students would expect them to average 2.0 fewer points per game. The y-intercept means that if a player could run a 100 m dash in 0 seconds, the model would predict them to average 32 points per game. This is a nonsensical idea.

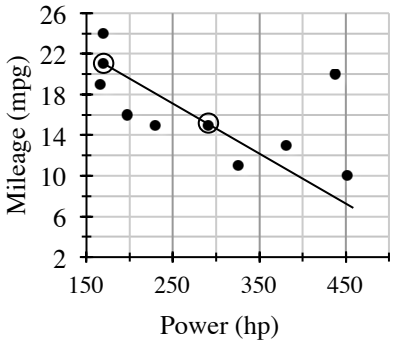
b. Using provided model, the residual is 3.5 points, meaning he averaged 3.5 points per game higher than the model estimated (good job, player!).

2-26. a. Part (*i*) is done correctly since it has both the equally spaced number line and the boxplot. Part (*ii*) has an equally spaced number line that is numbered incorrectly.

b. first quartile = 18.5, third quartile = 29.5, median = 25, IQR = 11

c. It appears to be discrete with the values limited to whole numbers.

2-27. a. See scatterplot at right.



b. Answers can vary. If you ignore the possible outlier at (438, 20), a reasonable line could pass through the points (170, 21) and (290, 15). The line of best fit has equation $\hat{y} = 29.5 - 0.05x$

c. For every increase of 1 horsepower, the gas mileage is expected to decrease by 0.05 mpg. The y-intercept means that a 0 horsepower car would get 29.5 mpg. This does not make sense because the y-axis is far from the data and thus this is extrapolation. (Prediction models of all types are unreliable when you extrapolate them.)

d. When the outlier at (438, 20) is removed, there appears to be a strong, negative, linear relationship. If the outlier is not removed, the association is more moderate (not as strong). For every increase of 1 horsepower, gas mileage is expected to decrease by about 0.05 mpg.

e. With the provided model, the residual is 2.5 mpg, meaning the car got 2.5 mpg more than expected by the model based on its power. A car owner would prefer a positive residual; that would mean the car had better fuel efficiency relative to its power than expected.