Lesson 2.1.4

2-28. a. Answers will vary. Possible answer $p = -50 + 0.4t$ where p is the total points scored in a season and t is the time played in minutes.

b. Scottie Sordan is an outlier. Using the model from part (a), his residual is $2491 - 1186 = 1305$ points. John Bailey is also far from the bulk of the data.

c. A positive residual is better because the player's actual points are greater than predicted by the model.

d. Answers will vary. Possible answer: $-50 + 0.4(2103) \approx 791$ points

2-29. Possible explanation: One draws the line of best fit by trying to get as close as possible to the most points. The best model is the one that is closest to the most points. One way to measure how close the line is to the points is with the residuals. Calculate the standard deviation of the residuals, S, and use that to decide which line is the best.

2-30. a. The sum of the squared residuals.

b. Because it is the line with the least possible sum of squared residuals.

2-31. a. Answers vary.

b. Answers vary.

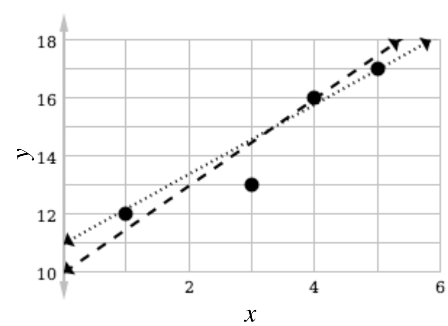
c. Answers vary.

d. You cannot beat the LSRL—that is why it is the *least* squares line.

e. A perfectly average basketball player, with both average minutes played and average points scored.

f. The LSRL goes right through the mean lines. This makes sense: the LSRL should correctly predict the "perfectly average" player.

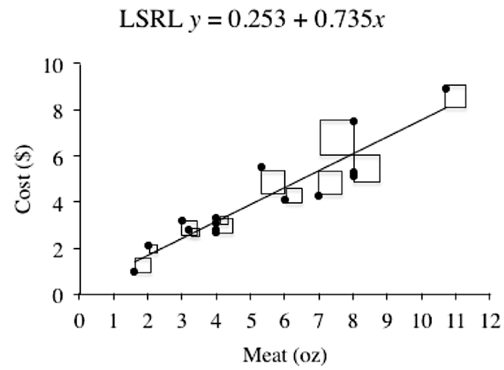
2-32. a. Answers vary. See scatterplot at right: dotted line is Kristen's, dashed is Joe's.



b. SSR for Joe = 2.75, SSR for Kristen = 2.64, so Kristen's line is slightly better. You could also choose to calculate S; Joe's S is 1.17 while Kristen's is 1.15, slightly better.

c. SSR for this line is 1.91, lower than Joe or Kristen's numbers. The S for this line is 0.98, again lower than Joe or Kristen's values.

2-33. a. See graph at right. The total area of the squares represents the sum of the squared residuals.



b. This is a fairly strong, positive, linear association.

c. The slope means that for every increase of one ounce in the patty size you can expect to see a price increase of $0.74. The *y*-intercept would be the cost of the hamburger with no meat. The *y*-intercept of $0.25 seems low for the cost of the bun, toppings, and service, but is not entirely unreasonable.

d. One would expect to pay 0.253 + 0.735(3) = $2.46 for a hamburger with a 3 oz patty while the cost of the given 3 oz patty is $3.20, so it has a residual of $3.20 – $2.46 = $0.74. The 3 oz burger costs $0.74 more that predicted by the LSRL model.

e. $S = \sqrt{\frac{\text{SSR}}{n-2}} = \sqrt{\frac{8.923}{13}} = 0.828$. The typical difference between the predicted cost of a burger and the actual cost is around 82.8 cents.

f. The LSRL model would show the expected cost of a 16 oz burger to be 0.253 + 0.735(16) = $12.01. 16 oz represents an extrapolation of the LSRL model, however $14.70 is more than $2 overpriced.

g. Average weight = 5.253 ounces. Average cost = $4.113. Is this on the line? 0.253 + 0.734(5.253) = $4.108. Close enough (within rounding error)!

h. The slope of the LSRL will increase as the line gets "pulled" toward the new point (which could be considered an outlier it is so far from the pattern). The value of *S* will increase, since there will be a new, larger residual. The strength of the association will therefore decrease.

2-34. a. School A; because of the age range (from 5 to 14). School A has more students in K-2 because half of the students are less than 8 years old according to the median.

b. 11 year-olds make up about 50% of the school.

c. It cannot be determined from the boxplot.

d. All three schools would have to be visited; the registration records at the schools give students birth dates; from these, the ages of students could be determined and recorded.

e. School B has about the same number of students in of each age, with the lower fourth spread over two years; median is 11 years, the distribution is symmetric, IQR is two years, and there are no apparent outliers.

