Lesson 2.2.3

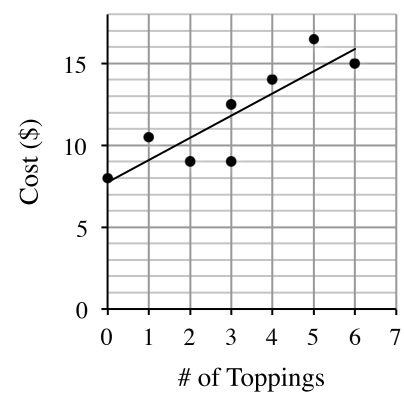
2-63. a. $y = 1.66 + 0.13x$, $r = 0.997$

b. $y = 7.49 + 0.06x$. Yes; the new point is far away from the pattern of the rest of the data. New $r = 0.530$

c. The LSRL is "pulled" toward the outlier. Amy's mistake will cause predictions close to the wall to be too large, and predictions far from the wall to be too small. Correlation decreases drastically; neither the LSRL nor correlation are resistant to outliers.

2-64. a. A scatterplot, a LSRL on the scatterplot, an equation for the LSRL, a description of the association (form, direction strength, and outliers), and possibly a description of the bounds.

b. See scatterplot at right. $y = 7.74 + 1.36x$; form: linear, direction: positive with a slope of 1.36 meaning an increase of one topping is expected to increase the cost by \$1.36, strength: strong, $r = 0.862$. About \$10.46 for a two-topping pizza.



2-65. a. I-C, II-B, III-A

b. Scatterplot II. If the data in the residual plot appears to be randomly scattered, with no pattern, the linear model fits the data well. The curved Residual Plot A for Scatterplot III indicates a curved line of best fit would be better. The fan-shaped Residual Plot C for Scatterplot I indicates that as the x-values get larger, there is more and more variability in the observed data; predictions made from smaller x-values will probably be closer to the observed value than predictions made from larger x-values.

2-66. a. Negative residual is lower cost than predicted.

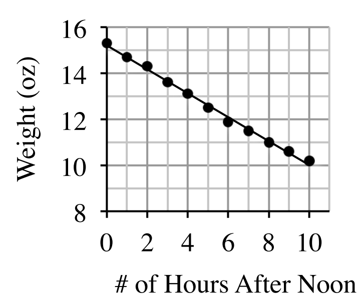
b. 0; No, the LSRL goes through the "middle" of all the data points, so students should expect the positive residuals to equal the negative residuals.

c. There does not appear to be any kind of shape or pattern to the plotted points. That means the model fits through the data points well. That is, our LSRL linear model is most appropriate.

2-67. a. See graph at right. $y = 15.21 - 0.52x$, $r = -0.997$. The direction is negative, the strength is strong, the form is … linear?

b. Residual plot shows a clear curve; original data was curved, probably exponentially!

c. NO. The correlation coefficient calculated strength of an association TREATING it as linear, but cannot tell form itself. This is less linear than the pizza problem, despite its higher r.



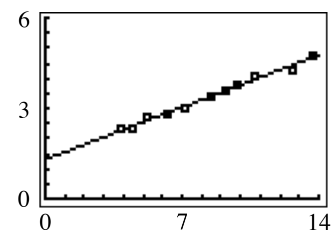
2-68. a. Yes. The scatter appears random and there is no apparent pattern in the residual plot.

b. $9.37 + 3.96(62.9) = 258$

c. The residual is about –18, so the actual observed is about 240.

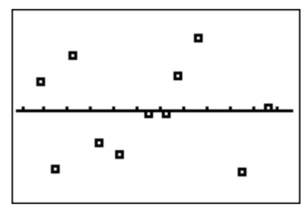
2-69. Points are not *evenly* scattered around the plot. Points can be in clusters and there can be large gaps, however, there is no discernable trend or pattern to the points.

2-70. a. See graph at right.



b. $y = 1.300 + 0.248x$

c. See graph below right.



d. Yes, the residual plot appears randomly scattered with no apparent pattern.

e. Predicted mass is $1.300 + 0.248(16.8) = 5.5$ g, residual is $6.0 - 5.5 = 0.5$ g. The measurements had one decimal place.

f. A positive residual means the pencil had more mass than was predicted by the LSRL model.

2-71. a. slope $= r \cdot \frac{s_y}{s_x} = \frac{0.0568 \cdot 39.67}{0.131} = 17.2$ Kcal/dollar. Since the LSRL must pass through the average point (\$1.071, 282.864). Substituting, you can find the y-intercept by doing $282.864 = 17.2(1.071) + a$ to solve for $a = 264.44$. Therefore the equation is $p = 264.44 + 17.2c$.

b. The scatterplot and r very close to 0 show that there is essentially no association between these variables.

c. It is! At least, it is relative to the units. If you sketch a line with slope 17.2 on the graph with the units involved, the line is very nearly horizontal.

2-72. a. Answers vary. b. Answers vary.

c. Answers vary. d. Answers vary.

e. This one is impossible; slope and r must have the same sign.

2-73. Answers may vary. This is a reasonable example. Because of a regular distance between the quartiles, the shape of the distribution of CO_2 percentages appears to be relatively uniform and symmetric. As such the mean and standard deviation are valid statistics. A typical sample CO_2 percentage is approximately 0.038%. The values range from 0.031 to 0.045% with a typical distance from the mean of 0.0044%. There are no apparent outliers.

