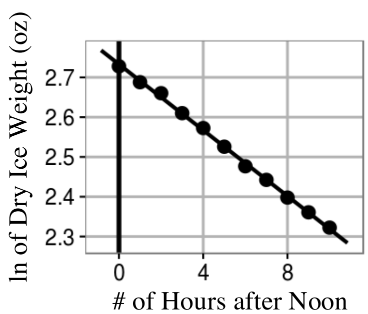
Lesson 12.2.2

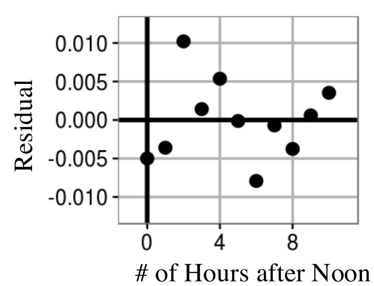
12-27. a. The residual plot shows a clear curve meaning the original data was curved. Although by observation a linear model fits the data quite well, a curved model might fit the data even better.

b. Use a logarithm! You could use any logarithm base. Using the natural log, the equation is $\ln(y) = \ln(a) + x \ln(b)$, but any logarithm (including the standard $\log_{10}$) will work fine.

c. ln or $\log(y)$ is in a linear relationship with x. You should notice that $\ln(a)$ and $\ln(b)$ are both constants, so this can be rewritten as $\ln(y) = a + bx$ (with new values for a and b).

d. Answers assuming a use of $\ln(x)$ are shown below. $\ln(y) = 2.7329 - 0.0414x$, $r^2 = 0.9986$, $r = -0.9993$; The constants in the equation will be different if using $\log(x)$, but the residual plot, R^2, etc., should be the same. The residual plot is significantly more randomly scattered, so this is a better model.





e. Using $\ln(y) = 2.7329 - 0.0414(24)$ gives $\ln(y) = 1.7393$, so $y = e^{1.7393} = 5.69$ oz. Using the linear model, $y = 15.21 - 0.52(24) = 2.73$ oz. The exponential model is likely more accurate; even though both are extrapolations, our reasonable belief that an exponential model is better than a linear model for this equation means students should trust that equation more.

f. Do a linear model of $\ln(y)$ vs. x, writing the model "$\ln(y) = a + bx$"

12-28. a. $\log(y) = \log(ax^p) \rightarrow \log(y) = \log(a) + p\log(x) \rightarrow \log(y) = k + p\log(x)$. p is the slope and also the unknown power of the model! Note that $\log(a)$ was renamed "k" since $\log(a)$ is a constant.

b. The residuals are randomly scattered about the line, and the R^2 is quite high, showing that this is a good model. The equation for the model is $\log(y) = 0.90 + 0.71 \cdot \log(x)$. Raising both sides as the exponent of 10, you get the equation $y = 10^{0.90+0.71\cdot\log(x)} = 10^{0.9}10^{0.71\cdot\log(x)} = 7.94x^{0.71}$.

c. $y = 7.94 \cdot 135^{0.71} = 258$ g

d. Take the logarithm of both x and y and graph the logs against each other. If the graph is linear, then the power model works, and the slope of the line is the exponent of the power model.

12-29. Residual plots are not provided, but it is clear from the scatterplots that the best fit for the data is the log(y) vs. x graph. Thus, an exponential model is the best choice. The equation for the model is $\log(m) = 2.2435 + 0.0278y$. Applying this model results in an estimate of 1995 MMbbl in 1938.

12-30. a. The plot of log(y) vs. x appears linear, and the model has a scattered residual plot.

b. Because log(y) vs. x was linear (the model displays a scattered residual plot), the relationship between y and x is *exponential*.

c. $\log(y) = 0.0778(40) - 2.3897 = 0.7223$, so $y = 10^{0.7223}$. The prediction is $y = 5.28$.

d. On the scatterplot of y vs. x, when $x = 40$, a predicted y of 5.28 seems very reasonable based on the trend of the actual data.

12-31. a. 81.5% of the variability in fuel efficiency can be explained by a linear relationship with weight.

b. The negative slope means there is a negative association. An increase of 1000 pounds in weight is expected to decrease the fuel efficiency by 8.4 miles per gallon.

12-32. a. Many nighttime jobs involve working in bars, casinos, or restaurants where <u>smoking</u> is prevalent. Night employment is generally considered less desirable so people who work at night may have <u>less money</u> and therefore less access to medical care.

b. This might be connected to <u>gender</u>. Men as a group eat more meat and do not live as long as women. Also, if the meat is highly <u>processed</u> (like hotdogs) it might be the additives that are harmful and not the meat itself.

12-33. a. On average student backpacks get 0.55 pounds lighter with each quarter of high school completed.

b. About 44% of the variation in student backpack weight can be explained by a linear relationship with the length of time spent in high school.

c. The “largest” residual value is about 6.2 pounds and it belongs to the student who has completed 3 quarters of high school.

d. $13.84 - 0.55(10) = 8.34$ lbs

e. A different model would be better because it looks like there is a curved pattern in the residual plot.

12-34. Mean = 159.51 lbs, s = 20.75 lbs, low = 121 lbs, Q1 = 138.9 lbs, median = 159.4 lbs, Q3 = 178.7 lbs, high = 186 lbs. Because the data is approximately symmetric and contains no strong skewing or outliers, the mean and standard deviation would be appropriate and more commonly reported than the median and IQR.

