2.5.1 EXERCISES

To see all of the help resources associated with this section, click OSttS Chapter 2.
For help with these exercises, click on one or more of the resources below:

- Using a calculator to find a linear regression model (least squares line)
- Interpreting slope as a rate of change
- Using a calculator to find a quadratic regression model
- Using the vertex formula

1. According to this website[6], the census data for Lake County, Ohio is:

Year	1970	1980	1990	2000
Population	197200	212801	215499	227511

   (a) Find the least squares regression line for these data and comment on the goodness of fit.[7] Interpret the slope of the line of best fit.

   (b) Use the regression line to predict the population of Lake County in 2010. (The recorded figure from the 2010 census is 230,041)

   (c) Use the regression line to predict when the population of Lake County will reach 250,000.

2. According to this website[8], the census data for Lorain County, Ohio is:

Year	1970	1980	1990	2000
Population	256843	274909	271126	284664

   (a) Find the least squares regression line for these data and comment on the goodness of fit. Interpret the slope of the line of best fit.

   (b) Use the regression line to predict the population of Lorain County in 2010. (The recorded figure from the 2010 census is 301,356)

   (c) Use the regression line to predict when the population of Lake County will reach 325,000.

3. Using the energy production data given below

Year	1950	1960	1970	1980	1990	2000
Production (in Quads)	35.6	42.8	63.5	67.2	70.7	71.2

[6] http://www.ohiobiz.com/census/Lake.pdf
[7] We'll develop more sophisticated models for the growth of populations in Chapter 6. For the moment, we use a theorem from Calculus to approximate those functions with lines.
[8] http://www.ohiobiz.com/census/Lorain.pdf

(a) Plot the data using a graphing calculator and explain why it does not appear to be linear.

(b) Discuss with your classmates why ignoring the first two data points may be justified from a historical perspective.

(c) Find the least squares regression line for the last four data points and comment on the goodness of fit. Interpret the slope of the line of best fit.

(d) Use the regression line to predict the annual US energy production in the year 2010.

(e) Use the regression line to predict when the annual US energy production will reach 100 Quads.

4. The chart below contains a portion of the fuel consumption information for a 2002 Toyota Echo that I (Jeff) used to own. The first row is the cumulative number of gallons of gasoline that I had used and the second row is the odometer reading when I refilled the gas tank. So, for example, the fourth entry is the point (28.25, 1051) which says that I had used a total of 28.25 gallons of gasoline when the odometer read 1051 miles.

Gasoline Used (Gallons)	0	9.26	19.03	28.25	36.45	44.64	53.57	62.62	71.93	81.69	90.43
Odometer (Miles)	41	356	731	1051	1347	1631	1966	2310	2670	3030	3371

Find the least squares line for this data. Is it a good fit? What does the slope of the line represent? Do you and your classmates believe this model would have held for ten years had I not crashed the car on the Turnpike a few years ago? (I'm keeping a fuel log for my 2006 Scion xA for future College Algebra books so I hope not to crash it, too.)

5. On New Year's Day, I (Jeff, again) started weighing myself every morning in order to have an interesting data set for this section of the book. (Discuss with your classmates if that makes me a nerd or a geek. Also, the professionals in the field of weight management strongly discourage weighing yourself every day. When you focus on the number and not your overall health, you tend to lose sight of your objectives. I was making a noble sacrifice for science, but you should <u>not</u> try this at home.) The whole chart would be too big to put into the book neatly, so I've decided to give only a small portion of the data to you. This then becomes a Civics lesson in honesty, as you shall soon see. There are two charts given below. One has my weight for the first eight Thursdays of the year (January 1, 2009 was a Thursday and we'll count it as Day 1.) and the other has my weight for the first 10 Saturdays of the year.

Day # (Thursday)	1	8	15	22	29	36	43	50
My weight in pounds	238.2	237.0	235.6	234.4	233.0	233.8	232.8	232.0

Day # (Saturday)	3	10	17	24	31	38	45	52	59	66
My weight in pounds	238.4	235.8	235.0	234.2	236.2	236.2	235.2	233.2	236.8	238.2

(a) Find the least squares line for the Thursday data and comment on its goodness of fit.

(b) Find the least squares line for the Saturday data and comment on its goodness of fit.

(c) Use Quadratic Regression to find a parabola which models the Saturday data and comment on its goodness of fit.

(d) Compare and contrast the predictions the three models make for my weight on January 1, 2010 (Day #366). Can any of these models be used to make a prediction of my weight 20 years from now? Explain your answer.

(e) Why is this a Civics lesson in honesty? Well, compare the two linear models you obtained above. One was a good fit and the other was not, yet both came from careful selections of real data. In presenting the tables to you, I have not lied about my weight, nor have you used any bad math to falsify the predictions. The word we're looking for here is 'disingenuous'. Look it up and then discuss the implications this type of data manipulation could have in a larger, more complex, politically motivated setting. (Even Obi-Wan presented the truth to Luke only "from a certain point of view.")

6. (Data that is neither linear nor quadratic.) We'll close this exercise set with two data sets that, for reasons presented later in the book, cannot be modeled correctly by lines or parabolas. It is a good exercise, though, to see what happens when you attempt to use a linear or quadratic model when it's not appropriate.

(a) This first data set came from a Summer 2003 publication of the Portage County Animal Protective League called "Tattle Tails". They make the following statement and then have a chart of data that supports it. "It doesn't take long for two cats to turn into 80 million. If two cats and their surviving offspring reproduced for ten years, you'd end up with 80,399,780 cats." We assume $N(0) = 2$.

Year x	1	2	3	4	5	6	7	8	9	10
Number of Cats $N(x)$	12	66	382	2201	12680	73041	420715	2423316	13968290	80399780

Use Quadratic Regression to find a parabola which models this data and comment on its goodness of fit. (Spoiler Alert: Does anyone know what type of function we need here?)

(b) This next data set comes from the U.S. Naval Observatory. That site has loads of awesome stuff on it, but for this exercise I used the sunrise/sunset times in Fairbanks, Alaska for 2009 to give you a chart of the number of hours of daylight they get on the 21^{st} of each month. We'll let $x = 1$ represent January 21, 2009, $x = 2$ represent February 21, 2009, and so on.

Month Number	1	2	3	4	5	6	7	8	9	10	11	12
Hours of Daylight	5.8	9.3	12.4	15.9	19.4	21.8	19.4	15.6	12.4	9.1	5.6	3.3

Use Quadratic Regression to find a parabola which models this data and comment on its goodness of fit. (Spoiler Alert: Does anyone know what type of function we need here?)

2.5.2 Answers

1. (a) $y = 936.31x - 1645322.6$ with $r = 0.9696$ which indicates a good fit. The slope 936.31 indicates Lake County's population is increasing at a rate of (approximately) 936 people per year.
   (b) According to the model, the population in 2010 will be 236,660.
   (c) According to the model, the population of Lake County will reach 250,000 sometime between 2024 and 2025.

2. (a) $y = 796.8x - 1309762.5$ with $r = 0.8916$ which indicates a reasonable fit. The slope 796.8 indicates Lorain County's population is increasing at a rate of (approximately) 797 people per year.
   (b) According to the model, the population in 2010 will be 291,805.
   (c) According to the model, the population of Lake County will reach 325,000 sometime between 2051 and 2052.

3. (c) $y = 0.266x - 459.86$ with $r = 0.9607$ which indicates a good fit. The slope 0.266 indicates the country's energy production is increasing at a rate of 0.266 Quad per year.
   (d) According to the model, the production in 2010 will be 74.8 Quad.
   (e) According to the model, the production will reach 100 Quad in the year 2105.

4. The line is $y = 36.8x + 16.39$. We have $r = .99987$ and $r^2 = .9997$ so this is an excellent fit to the data. The slope 36.8 represents miles per gallon.

5. (a) The line for the Thursday data is $y = -.12x + 237.69$. We have $r = -.9568$ and $r^2 = .9155$ so this is a really good fit.
   (b) The line for the Saturday data is $y = -0.000693x + 235.94$. We have $r = -0.008986$ and $r^2 = 0.0000807$ which is horrible. This data is not even close to linear.
   (c) The parabola for the Saturday data is $y = 0.003x^2 - 0.21x + 238.30$. We have $R^2 = .47497$ which isn't good. Thus the data isn't modeled well by a quadratic function, either.
   (d) The Thursday linear model had my weight on January 1, 2010 at 193.77 pounds. The Saturday models give 235.69 and 563.31 pounds, respectively. The Thursday line has my weight going below 0 pounds in about five and a half years, so that's no good. The quadratic has a positive leading coefficient which would mean unbounded weight gain for the rest of my life. The Saturday line, which mathematically does not fit the data at all, yields a plausible weight prediction in the end. I think this is why grown-ups talk about "Lies, Damned Lies and Statistics."

6. (a) The quadratic model for the cats in Portage county is $y = 1917803.54x^2 - 16036408.29x + 24094857.7$. Although $R^2 = .70888$ this is not a good model because it's so far off for small values of x. Case in point, the model gives us 24,094,858 cats when $x = 0$ but we know $N(0) = 2$.

(b) The quadratic model for the hours of daylight in Fairbanks, Alaska is $y = .51x^2 + 6.23x - .36$. Even with $R^2 = .92295$ we should be wary of making predictions beyond the data. Case in point, the model gives -4.84 hours of daylight when $x = 13$. So January 21, 2010 will be "extra dark"? Obviously a parabola pointing down isn't telling us the whole story.