1.4 Modeling With Technology

In the previous section, we discussed the use of a linear model to solve a problem that we may see in economics: finding the equilibrium point in a system governed by traditional rules of supply versus demand. You were also given the functions that showed the supply and demand behaviors; these functions were both "determined by previous analysis." In this section, we will *do the analysis* to develop the supply and demand curves so we can find the equilibrium point of an economic system.

Suppose we are analysts for a large oil company and we want to analyze the behavior of the American oil market to make more money. The company has access to data collected through years of business. We will conduct an analysis of the price of oil (U.S. dollars per barrel), the demand of oil (millions of barrels per day in the United States), and the supply of oil (same units as demand).

First, we will develop a linear model that enables us to represent oil demand, given the price of oil. See Table 1.4 for the data (data is fictional, also linked on the course website).

Price ($ per Barrel)	Demand (Millions of Barrels per Day)
75	17.22
80	16.58
85	15.94
90	15.05
95	14.90
100	13.72
105	13.43

Table 1.4: Oil Demand vs. Price

We now have a problem that we must solve: develop a linear model for the data above, so it is possible to determine the equilibrium price of the system. What must we do to solve the problem? A process for solving the problem would certainly be in order!

Step 1. *Transform the Problem.* Given in Table 1.4 is the data we will analyze. The independent variable is the price of oil because that determines the demand of oil (the dependent variable). We must find a model that fits through the data so we can make predictions. Our first step in finding the most appropriate model is to use the skills learned in our first problem solving lab to plot the data. See Figure 1.12 for a plot of the data found in Table 1.4.

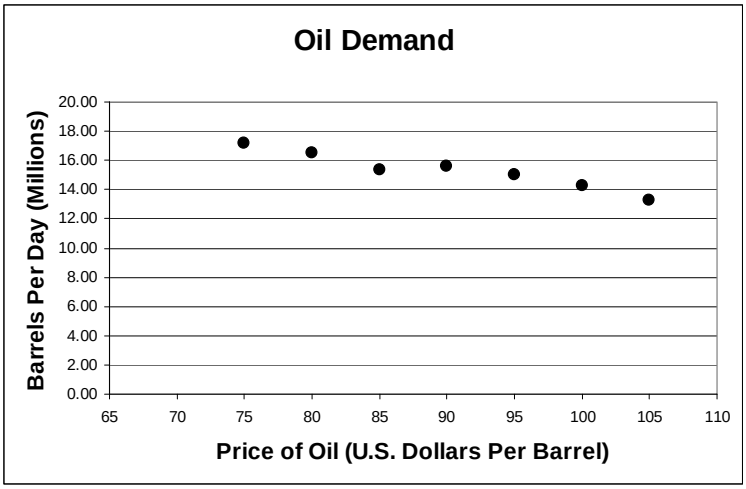

Figure 1.12: Plot of Oil Price vs. Demand

It is easy to see that a straight line will not perfectly fit through the data. That's OK! A model that you fit to actual data *very* rarely will fit perfectly through every data point – but we must do the best we can. It seems that a line will be able to give us a model that is "good enough" to solve the problem at hand: find the equilibrium point of our system. The plan we will use will be to find values for the parameters associated with a line (the slope and y-intercept).

Step 2. *Solve the Problem Using the Most Appropriate Techniques.* To finalize our solution plan, we should remind ourselves of the general form of the model we selected. The general form of a line is: $y = ax + d$. Notice that we have two parameters, 'a' and 'd.' We have two variables, 'x' and 'y.' We can solve for our parameters if we select values for the variables, forming a system of equations. Because we have two parameters we need to solve for, we will need two equations, selecting two data points. If the first and last data points are representative of the general trend of the data, it is common to select the first and last data points to estimate an initial model. Let's see what happens when we use these points: $(x, y) = (75, 17.22)$ and $(105, 13.43)$. The model we will have after solving for our two parameters should go through the two data points we used to develop the parameters – a useful fact when reflecting on our solution. Now that we have our two data points, let's form our equations:

$$17.22 = 75a + d$$
$$13.43 = 105a + d$$

We could solve these equations using substitution. Let's use one of the technology tools we have at our disposal – Mathematica, which would be much faster than substitution! In the problem solving lab, we learned how to solve two equations in two unknowns. Let's apply that knowledge here.

```
Solve[{17.22 == 75 a + d, 13.43 == 105 a + d}, {a, d}]

{{a → -0.126333, d → 26.695}}
```

Figure 1.13: Solving Two Equations in Two Unknowns (Mathematica)

The Mathematica output indicates that a line with a slope of about -0.1263 and *y*-intercept of approximately 26.695 will run through the two data points we selected. The final equation is: $y = -0.1263x + 26.695$.

Step 3. *Interpret the Solution.* Perhaps the most important step in the modeling and problem solving process is interpreting the solution – communicating it in non-mathematical terms and reflecting upon whether or not the solution we attain solves the problem posed. To help interpret the solution, let's plot a graph of the linear model we developed on the same axes as the data.

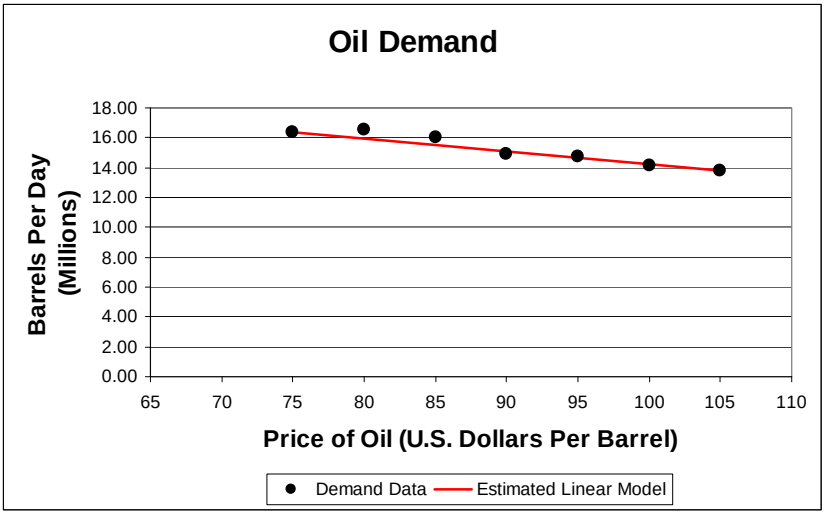

Figure 1.14: Illustration of Estimated Linear Model

The model we have developed, relating demand to price with the equation $\text{Oil_Demand} = -0.1263\text{Oil_Price} + 26.695$ seems do a nice job of describing the data provided in Table 1.4. The model does seem to fit directly through the first and last data points, as expected. The domain and range of the function follow:

$$Domain(x): \quad [75,105]$$
$$Range(y): \quad [13.43,17.22]$$

People that do modeling for a living are curious people. Our initial assumption was that the first and last data points were "good enough" from which to create an initial model. What if we changed the assumption we used about which points to choose? Let's try the third and sixth data points, (85, 15.94) and (100, 13.72).

Again, using Mathematica to solve for our parameters, our slope and *y*-intercept are -0.148 and 28.52, respectively. The slopes are different by about .02 (about a 15% difference) and the y-intercepts by less than 2.0 (7%). It appears that the model's y-intercept is less *sensitive to change* than the model's slope. We complete a **sensitivity analysis** by testing how much change in an assumption will impact the final model.

> The Government Accountability Office (GAO) defines **sensitivity analysis** as the determination of how sensitive outcomes are to changes in the assumptions. The assumptions that deserve the most attention should be those with the greatest amount of uncertainty and effect on the outcome.

In general, we prefer a model that is *robust* against change. That means, when an assumption changes, the model remains essentially the same. It is undesirable to have a model that fluctuates wildly with changes in the assumptions.

We compared the two models, algebraically. We can also compare them graphically by overlaying the second model over the first model and data. As expected, the two developed models do not look much different (see Figure 1.15). In the next section, we will discuss methods to determine which of the algebraic models is truly the best model. For now, we will choose the first model we developed, $y = -0.1263x + 26.695$.

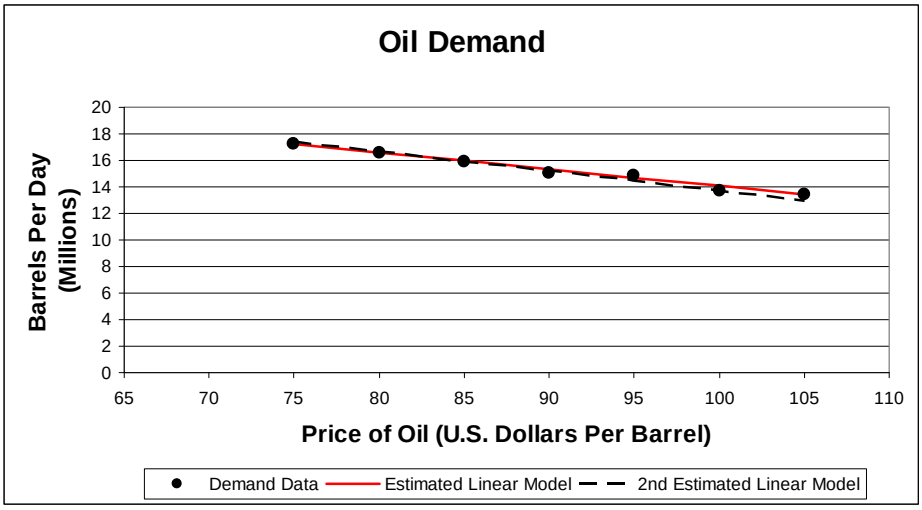

Figure 1.15: A Comparison of Linear Oil Demand Models

This linear model will enable us to find the equilibrium point of the oil supply and demand system, after the supply curve is plotted.

Question 1 *Given the data in Table 1.5 (and on course website), develop an equation for the supply of oil, in millions of barrels per day, given the price of a barrel of oil.*

Price ($ per Barrel)	Supply (Millions of Barrels per Day)
75	2.56
80	3.42
85	3.50
90	3.85
95	4.62
100	4.90
105	6.13

Table 1.5: Oil Supply vs. Price

After determining the model for calculating oil supply as a function of selling price, we are now prepared to determine the equilibrium price of the system we are analyzing. There are two possibilities for the equilibrium price: it could occur inside

the current domain values of our functions, [75, 105]. It could also occur outside those values, meaning we must extend our models.

If we predict a value *inside* the current domain of our function, it is said that we are **interpolating**. For example, if the equilibrium point occurred at a selling price of $87, it would be a price *inside* our domain; therefore, we would *interpolate*.

If we were to predict a value outside the current domain, meaning we must *extend* our model, it is said that we are **extrapolating**. For example, if the equilibrium point occurred at a selling price of $115, it would be a price outside our domain. We would have to *extend* our model (it may still be a good model); therefore, we would *extrapolate.*

> We define **interpolation** as the act of making predictions *within* the domain of known values or data. **Extrapolation** is the act of predicting values *outside* of the domain of the data.

Question 2 *Determine the equilibrium point of the oil supply and demand system described in this section. Is the answer an* interpolation *or* extrapolation *of our model?*

Example 1 *In 1986, the Space Shuttle Challenger experienced a catastrophic failure in its solid rocket booster. The explosion, 73 seconds after liftoff, claimed both the crew and shuttle. The cause of explosion was later determined to be an o-ring failure in the right solid rocket booster. The final investigative report concluded that cold weather was a contributing factor.*

The o-rings in the solid rocket boosters on the space shuttle are designed to expand when heated to seal different chambers of the rocket so that solid rocket fuel is not ignited. According to engineering specifications, the o-rings must expand by at least 5% in order to ensure a safe launch. The temperature on the day of launch was 29 degrees F. O-ring expansion data was collected on the previous nine launches and is shown in Table 1.6.

Temp (degF)	93	88	87	81	73	72	68	64	55
% Expansion	22.3	21.0	20.6	19.7	18.7	19.0	17.3	16.2	15.5

Table 1.6: O-Ring Expansion Data for Space Shuttle Challenger Launches

If you were given this data prior to launch, what would you have recommended?

Step 1. *Transform the problem.* We must answer the question posed above: should the shuttle launch. We must identify what is <u>given</u> in the problem.

- We have data from the previous nine launches.
- We know that the temperature at time of launch was 29 degrees F.
- We know that o-rings must expand by 5% to ensure a safe launch.
- Can we draw a picture (graph)? YES! (see Figure 1.16 on the next page)

What must we <u>find</u> in the problem? We need a model to determine what the percent expansion of an o-ring will be if we have a launch temperature of 29°F.

Because we need to find the expansion, that will be our *dependent* variable. The variable is the "cause" of the o-ring expansion is the temperature, the *independent* variable.

In addition to identifying what is given and what we need to find, there is an important assumption to be made: that o-rings react to temperature in a predictable manner at temperatures outside the domain of current launch temperatures (e.g., there is not a temperature at which the o-rings stop expanding).

Let's use Mathematica to plot this graph, instead of Excel. The following command will
result in the plot of our data.

```
ListPlot[{{93, 22.3}, {88, 21}, {87, 20.6}, {81, 19.7}, {73, 18.7}, {72, 19}, {68, 17.3}, {64, 16.2}, {55, 15.5}} ,
 PlotRange → {14, 23}, AxesLabel → {Temperature, O - ring_Expand}]
```

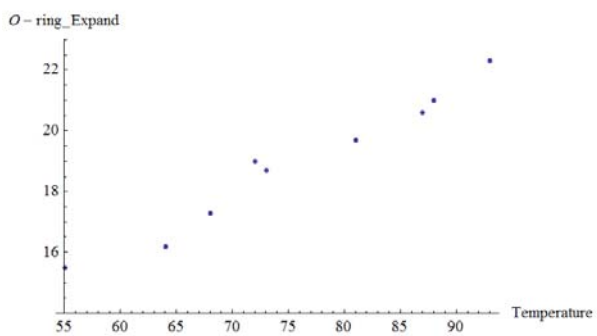

Figure 1.16: Mathematica Graph of O-Ring Data

The ListPlot command plots the data points, while the PlotRange command tells Mathematica the "y" values to plot, rather than giving the standard plot starting at zero. The AxesLabel command enable you to label the x and y axes. From this point on, it will be your choice whether to use Excel or Mathematica to plot data.

After graphing the data point, we can see that the data seem to follow a linear pattern; therefore, we should use our knowledge of linear models to develop a model that we can use to **extrapolate** whether or not the shuttle should launch at a temperature of 29°F. Because we will use a linear model, the general form of the model is: $Expansion(Temp) = a(Temp) + d$.

Our <u>plan</u> will be to estimate values for the parameters (a and d) of our linear model. To estimate two parameters, we need two data points that represent the data well. We see the data in Figure 1.16. For this model we will assume that the first and last data points will lead us to reasonable estimates of our parameters. Keep in mind that when choosing data points to estimate the parameters, the goal should be to minimize the deviation from the general trend of the data.

Step 2. *Solve Using the Most Appropriate Technique.* Our plan will be to use the first and the last point to formulate two equations with two unknowns (a and d). We will then solve these two equations simultaneously and use the solutions to estimate the two parameters of the linear function. Finally, we'll need to determine our model's

domain and range. Using our first and last data points, we create two equations with two unknowns. Our two equations are:

$$22.3 = a(93) + d$$
$$15.5 = a(55) + d$$

Figure 1.17 highlights the use of our *Mathematica* skills to solve both equations simultaneously for *a* and *d.* We find our estimated slope is 0.18 and our *y*-intercept is 5.66. Our final model is: $Expansion(Temp) = .18(Temp) + 5.66$.

```
Solve[{22.3 == a (93) + d, 15.5 == a (55) + d}, {a, d}]

{{a → 0.178947, d → 5.65789}}
```

Figure 1.17: Using Mathematica to Solve Two Simultaneous Equations

Thus, our model is: $Expansion(Temp) = 0.18 \cdot Temp + 5.66$. The data show our domain and range to be:

$$Domain: \{x \mid 55 \le x \le 93\}$$
$$Range: \{y \mid 15.5 \le y \le 22.3\}$$

Step 3. *Interpret the Solution.* The final step of the modeling and problem solving process is to communicate and reflect upon the solution that we have derived. Through interpreting the model developed, we see that on the day of launch, the temperature was 29 degrees F; therefore, our model predicts: $Expansion(29) = 0.18(29) + 5.66 = 10.88$

According to the model we created, the o-rings would have expanded nearly 11%. Our results indicate the o-ring expansion adheres to the engineering standards for a safe launch. It is upon reflection that we realize that our model may very well be flawed. After all, the Space Shuttle exploded. Was there something in the data that could have led us to that predict a dangerous launch?

Let's test an assumption to determine the sensitivity of our model. We assumed that the first and last data points were representative of the trend in our data. So, let's make a minor change to our selection and see what happens. Instead, we'll choose the second and last data points and see how much of an affect this has on our solution.

```
Solve[{21 == a (88) + d, 15.5 == a (55) + d}, {a, d}]

{{a → 0.166667, d → 6.33333}}
```

Figure 1.18: Using Technology to Test the Sensitivity of our Assumptions

Making this change slightly decreased our slope and increased our y-intercept (Figure 1.18). This also causes our new prediction on the day of launch to be: $Expansion(29) = 0.17(29) + 6.33 = 11.26$

Our new model predicts an even greater expansion than the last (former model is solid line, latter, dashed). Figure 1.19 shows the data versus the two possible models described in the text, and how the models are plotted.

```
Plot1 = ListPlot[{{93, 22.3}, {88, 21}, {87, 20.6}, {81, 19.7}, {73, 18.7}, {72, 19}, {68, 17.3}, {64, 16.2}, {55, 15.5}, {29, 11.26}} ,
  PlotRange → {5, 23}, AxesLabel → {Temperature, O - ring_Expand}]

f[x_] = .178947 x + 5.65789

5.65789 + 0.178947 x

Plot2 = Plot[f[x], {x, 29, 93}, PlotRange → {5, 23}]

g[x_] = .166667 x + 6.33333

6.33333 + 0.166667 x

Plot3 = Plot[g[x], {x, 29, 93}, PlotRange → {5, 23}, PlotStyle → {Dashed}]

Show[Plot1, Plot2, Plot3]
```

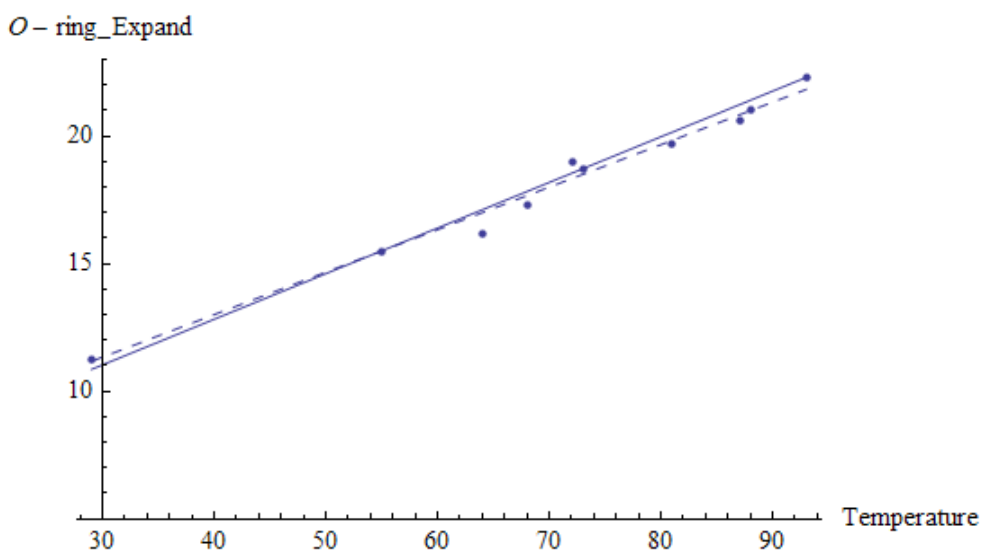

Figure 1.19: O-Ring Sensitivity Analysis

Our analysis concludes that the launch should have been safe; however, since we know this wasn't the case, perhaps our assumptions were not valid. Specifically, our assumption to select a linear model and our assumption that the data accurately represent what happens to o-rings at other temperatures may be invalid. We will revisit this idea later in the course.

Question 3 The data represented in Table 1.6 are posted on the course website. Using this data and a linear model, would it have been *possible to predict the Challenger's explosion?*

Question 4 *Given the following height and weight data from R-Day of 10 male new cadets,*

Weight (pounds)	Height (inches)
150	68
155	70
140	67
138	66
170	71
185	73
195	74
200	75
175	72
165	70

answer the following questions.

a. *What type of model do you think would be appropriate to predict a male new cadet's height given his weight?*
b. *Predict the height of a male cadet that weighs 100, 160 and 300 pounds.*
c. *Does your model have any limitations? What impact do these limitations have on the domain and range of your model?*

***Question 5** *The National Collegiate Athletic Association (NCAA) exists to promote a commitment to excellence in both the classroom and the "fields of friendly strife." Data is provided below that describes NCAA's membership since 1950*[1].

Year	Active Members
1950	362
1955	449
1960	524
1965	579
1970	645
1975	704
1980	738
1985	793
1990	828
1995	903
2000	977
2001	977
2002	1005
2003	1024
2004	1028
2005	1027

a. *Determine a model for NCAA membership as a function of time. Be sure to give the domain and range of your model.*

In its 2004 Annual Report, the NCAA noted that it anticipated that its membership growth rate will slow in coming years.

b. *Based on the prediction in the Annual Report, what characteristics would best describe a model of NCAA membership in the future?*

[1] 2004 NCAA Membership Report, Retrieved 26 January, 2006, from http://www.ncaa.org/library/membership/membership_report/2004/2004_ncaa_membership_report.pdf