Mathematical Modeling:
The Right Courses
for the Right Students
for the Right Reasons

Sheldon P. Gordon
gordonsp@farmingdale.edu
Accessing the Presentation

This PowerPoint presentation and the DIGMath Excel files that will be used can all be downloaded from:

farmingdale.edu/faculty/sheldon-gordon/
## Monthly Rainfall in Orlando

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>2.43</td>
<td>2.35</td>
<td>3.54</td>
<td>2.42</td>
<td>3.74</td>
<td>7.35</td>
<td>7.15</td>
<td>6.25</td>
<td>5.76</td>
<td>2.73</td>
<td>2.32</td>
<td>2.31</td>
</tr>
</tbody>
</table>

*Note: The table and chart show the rainfall in inches for each month.*
(a) Decide which is the independent variable and which is the dependent variable.
(b) Decide on appropriate scales for the two variables for a scatterplot.
(c) State precisely which letters you will use for the two variables and state what each variable you use stands for.
(d) Draw the associated scatterplot.
(e) Raise some predictive questions in this context that could be answered when we have a formula relating the two variables.
Behavior of functions as data and as graphs, including increasing/decreasing, turning points, concave up/down, inflection points (including the logistic function and the normal distribution function patterns).
Behavior of Functions

The graph shows the amount of profit from an investment over the course of 40 days after it was purchased.

a. On which days is the investment a gain?
b. On which days is the investment a loss?
c. On which intervals is the profit increasing?
d. On which intervals is the profit decreasing?
e. On which intervals is the profit function concave up?
f. On which intervals is the profit function concave down?
g. Estimate the coordinates of all turning points.
h. Estimate the coordinates of all inflection points.
i. If the investment was sold on the 40th day, estimate how much of a profit or loss resulted.
Linear functions, with emphasis on the meaning of the parameters and fitting linear functions to data, including the linear correlation coefficient to measure how well the regression line fits the data.
The snowy tree cricket, which lives in the Colorado Rockies, has a chirp rate, \( R \), in chirps per minute, that is related to the temperature, \( T \), in Fahrenheit.

<table>
<thead>
<tr>
<th>Temp</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chirp Rate</td>
<td>40</td>
<td>60</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
</tr>
</tbody>
</table>

a. Find a linear function that models the chirp-rate as a function of temperature.

b. Discuss the practical meaning of the slope and the vertical intercept and give reasonable values for the domain and range.

c. Predict the chirp rate at a temperature of 84°F.

d. Predict the temperature if a cricket is chirping 138 times per minute.
Regression by hand

The following table gives measurements for the chirp rate (chirps per minute) of the striped ground cricket as a function of the temperature.

<table>
<thead>
<tr>
<th>$T$</th>
<th>89</th>
<th>72</th>
<th>93</th>
<th>84</th>
<th>81</th>
<th>75</th>
<th>70</th>
<th>82</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>20</td>
<td>16</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T$</th>
<th>69</th>
<th>83</th>
<th>80</th>
<th>83</th>
<th>81</th>
<th>84</th>
<th>76</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>15</td>
<td>16</td>
<td>15</td>
<td>17</td>
<td>16</td>
<td>17</td>
<td>14</td>
</tr>
</tbody>
</table>
Regression Simulation
Regression using Calculators and/or Excel
The Sum of the Squares

The *regression line* is the line that comes *closest* to all of the data points in the *least-squares sense*, meaning that the sum of the squares of the vertical deviations is a minimum.
Nonlinear families of functions:

- exponential growth and decay, applications such as population growth and decay of a drug in the body; doubling time and half-life;
- power functions;
- logarithmic functions;
- Fitting each family of functions to data based on the behavioral characteristics of the functions and deciding on how good the fit is.
Families of Functions

After linear functions, the most common and useful families of functions in terms of realistic applications and models are, in order:

1. exponential functions
2. power functions
3. logarithmic functions
4. sinusoidal functions
5. polynomial functions

In contrast, in traditional mathematics, the major emphasis is on polynomials because they provide so many opportunities to practice algebraic skills.
Exponential Functions

For linear functions, the successive differences are constant. That constant difference is related to the slope of the line.

For exponential functions $y = A \ b^x$, the successive ratios are constant.

When the ratio of successive values $b$ is greater than 1, it is an exponential growth function and $b$ is the growth factor.

When the ratio of successive values $b$ is less than 1, it is an exponential decay function and $b$ is the decay factor.
Modeling Drug Levels in the Blood

Every drug is washed out of the bloodstream, usually by the kidneys, though at a different rate that is characteristic of the particular drug.

For example, in any 24-hour period, about 25% of any Prozac in the blood is washed out, leaving 75% of the amount.
Modeling Drug Levels in the Blood

$D_0 = 80 \text{ mg.}$

After 24 hours (1 day), 25% is removed, leaving
$D_1 = 0.75D_0 = 60 \text{ mg.}$

After another day, 25% of $D_1$ is removed, leaving
$D_2 = 0.75D_1 = 0.75(0.75D_0) = (0.75)^2D_0 = 45 \text{ mg.}$

After another day, 25% of $D_2$ is removed, leaving
$D_3 = 0.75D_2 = 0.75(0.75)^2D_0 = (0.75)^3D_0 = 33.75 \text{ mg.}$

In general, after $n$ days, the level of Prozac is:
$D_n = (0.75)^nD_0 = 80(0.75)^n$
Predictive Questions

- What will the level of the drug be after a given number of time periods?

- How long will it take until the drug level is down to a given level?

How long until Prozac level is 10 mg.?

Solve: $80(0.75)^n = 10$

Need to use logs to solve algebraically.
An Exponential Decay Model

When a person smokes a cigarette, about 0.4 mg of nicotine is absorbed into the blood. Each hour, about 35% of any nicotine present is washed out of the blood.

a. Write the equation of a function that models the level of nicotine in the blood after a single cigarette.

b. Use your model to estimate how long it takes for the amount of nicotine to drop to 0.005 mg.
Doubling Time & Half-Life

Every exponential growth function \( y = A b^t, \ b > 1, \) has a characteristic *doubling time* – the length of time needed for it to double in size. The doubling time depends *only* on the growth factor \( b. \)

Every exponential decay function \( y = A b^t, \ 0 < b < 1, \) has a characteristic *half-life* – the length of time needed for it to decrease by half. The half-life depends *only* on the decay factor \( b. \)
The following table shows world-wide wind power generating capacity, in megawatts, in various years.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind power</td>
<td>10</td>
<td>1020</td>
<td>1930</td>
<td>4820</td>
<td>7640</td>
<td>13840</td>
<td>32040</td>
<td>47910</td>
</tr>
</tbody>
</table>
(a) Which variable is the independent variable and which is the dependent variable?
(b) Explain why an exponential function is the best model to use for this data.
(c) Find an exponential function that models the relationship between the amount of power $P$ generated by wind and the year $t$.
(d) What are some reasonable values that you can use for the domain and range of this function?
(e) What is the practical significance of the base (1.1373) in the exponential function you created in part (c)?
(f) What is the doubling time for this function? Explain what it means. Solve: $52.497(1.1373)^t = 2 \times 52.497$.
(g) According to your model, what do you predict for the total wind power generating capacity in 2012?
How Exponential Regression Works

If a set of \((x, y)\) data follows an exponential pattern, then the transformed \((x, \log y)\) data follow a linear pattern.

Method Used:

- Transform the data to \(\log y\) vs. \(x\)
- Fit a linear function to the transformed data.
- Undo the transformation to create the exponential function.
Undoing the Semi-log Transformation

The linear fit to the (log $W$, $t$) data for wind power is:

\[ y = 0.1298x + 1.6897 \quad \text{or} \quad \log W = 0.1298t + 1.6897 \]

To undo the logs, we take powers of 10:

\[ 10^{\log W} = W = 10^{0.1298t + 1.6897} \]

\[ = 10^{0.1298t} \cdot 10^{1.6897} \]

\[ = 48.944 \cdot (10^{0.1298})^t \]

\[ = 48.944 \cdot (1.3483)^t, \]

Wind power has been growing at a rate of 34.83% a year.
How Non-Linear Fits Are Done

If a set of \((x, y)\) data follows an exponential pattern, then the transformed \((x, \log y)\) data follow a linear pattern.

If a set of \((x, y)\) data follows a power function pattern, then the \((\log x, \log y)\) data follow a linear pattern.

If a set of \((x, y)\) data follows a logarithmic pattern, then the \((\log x, y)\) data follow a linear pattern.
Correlation with Non-Linear Fits

In the process, a value for the linear correlation coefficient $r$ is calculated. It only measures how well the linear function fits the transformed data, not how well the nonlinear function fits the original data. It is this value that is reported by Excel and by graphing calculators. As such, it must be interpreted with care – students need to realize that it does not measure how well the nonlinear function fits the data.
How Good is the Fit?

How does one measure how well a particular function actually fits a set of data?

How does one decide on which of several possible candidates is the “best” fit to a set of data?

1. You could, but probably shouldn’t, use only the correlation coefficient.
2. You could use the sum of the squares to measure and compare different fits.
3. You can “eyeball” how well the function appears to fit the data.
Where Do You Find Data?

Statistical Abstract of the U.S. (on-line)
www.census.gov/statab/www/

Vital signs: The Trends that are Affecting Our Lives. (annual book published by the WorldWatch Foundation)

On Size and Life, Scientific American Lib.

A Modeling-Based Course: Topic 5

Modeling with Polynomial Functions:
Emphasis on the behavior of polynomials and modeling, primarily by fitting polynomials to data
Modeling the Spread of AIDS

The total number of reported cases of AIDS in the United States since it was first diagnosed in 1983:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS Cases</td>
<td>4,589</td>
<td>10,750</td>
<td>22,399</td>
<td>41,256</td>
<td>69,592</td>
<td>104,644</td>
<td>146,574</td>
<td>193,878</td>
</tr>
<tr>
<td>AIDS Cases</td>
<td>251,638</td>
<td>326,648</td>
<td>399,613</td>
<td>457,280</td>
<td>528,144</td>
<td>594,641</td>
<td>653,084</td>
<td>701,353</td>
</tr>
<tr>
<td>Year</td>
<td>1999</td>
<td>2000</td>
<td>2001</td>
<td>2002</td>
<td>2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIDS Cases</td>
<td>742,709</td>
<td>783,976</td>
<td>824,809</td>
<td>886,098</td>
<td>909,269</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Modeling the Spread of AIDS

An Exponential Function:

\[ A = 10.742(1.2583)^t, \]

where \( t \) is measured in years since 1980.

The correlation coefficient \( r = 0.9176. \)

The Sum of the Squares = 2,655,781.
Modeling the Spread of AIDS

A cubic polynomial:

\[ A = -0.20t^3 + 8.39t^2 - 53.23t + 97.97, \]

where \( t \) is the number of years since 1980.

The Sum of the Squares = 3,489.
Modeling the Spread of AIDS

A larger window:

\[ A = -0.20t^3 + 8.39t^2 - 53.23t + 97.97, \]
where \( t \) is the number of years since 1980.

Notice that the leading coefficient is negative.
Sinusoidal Functions and Periodic Behavior: Using the sine and cosine as models for periodic phenomena such as the number of hours of daylight, heights of tides, average temperatures over the year, etc.
Jacksonville Daylight Function

\[ H = 12 + 1.48 \sin \left( \frac{2\pi}{365} (t - 80) \right) \]

What do the four numbers represent?

What does the 365 represent in practice?

What does the 12 represent in practice?

What does the 1.48 represent in practice?

What does the 80 represent in practice?
Sinusoidal Functions
Sinusoidal Functions
Sinusoidal Functions

Using such displays, one can ask:

1. Estimate the period.

2. Estimate the maximum and minimum; estimate the midline.

3. Estimate the amplitude.

4. What is the frequency?

5. Estimate the phase shift for a cosine or sine function.

6. Write a formula to model the phenomenon.
The table gives the average daytime high temperature in Dallas roughly every 2 weeks.

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>15</th>
<th>32</th>
<th>46</th>
<th>60</th>
<th>74</th>
<th>91</th>
<th>105</th>
<th>121</th>
<th>135</th>
<th>152</th>
<th>196</th>
<th>213</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>55</td>
<td>53</td>
<td>56</td>
<td>59</td>
<td>63</td>
<td>67</td>
<td>72</td>
<td>77</td>
<td>81</td>
<td>84</td>
<td>89</td>
<td>98</td>
<td>99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day</th>
<th>227</th>
<th>244</th>
<th>258</th>
<th>274</th>
<th>288</th>
<th>305</th>
<th>319</th>
<th>335</th>
<th>349</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>98</td>
<td>94</td>
<td>90</td>
<td>85</td>
<td>80</td>
<td>72</td>
<td>66</td>
<td>61</td>
<td>58</td>
</tr>
</tbody>
</table>

Create a sinusoidal function to model the average daytime high temperature in Dallas.

The values in the table are temperatures roughly every 2 weeks, but two entries are missing. Predict the average daytime high temperature in Dallas on those dates.
The Sum of the Squares

Can you get a better fit based on the sum of the squares?

This is a lovely classroom investigation in which students suggest minor changes to the values of the parameters and immediately see whether they can improve on the fit. In the process, they get a much deeper understanding of the meaning of each of the parameters.
The average daytime high temperature in New York as a function of the day of the year varies between 32°F and 94°F. Assume the coldest day occurs on the 30th day and the hottest day on the 214th.

a. Sketch the graph of the temperature as a function of time over a three year time span.
b. Write a formula for a sinusoidal function that models the temperature over the course of a year.
c. What are the domain and range for this function?
d. What are the amplitude, vertical shift, period, frequency, and phase shift of this function?
e. Predict the most likely high temperature on March 15.
f. What are all the dates on which the high temperature is most likely 80°F?
The air conditioner in a home is set to come on when the temperature reaches 76 ° and to turn off when the temperature drops to 72 °. This cycle repeats every 20 minutes, starting at noon.

a. Sketch the graph of the temperature as a function of time over a one hour time span. (Assume the temperature is 76 ° at the start, which is at noon.)

b. Write a formula for a sinusoidal function that models the temperature over the course of time.

c. What are the domain and range for this function?

d. What are the amplitude, vertical shift, period, frequency, and phase shift of this function?

e. Predict the temperature at 12:08; at 12:26.

f. What are all the times between noon and 1 pm when temperature is 73°?
For More Information

This PowerPoint presentation and the DIGMath Excel files demonstrated can all be downloaded:

farmingdale.edu/faculty/sheldon-gordon/

Sheldon P. Gordon
gordonsp@farmingdale.edu