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/
College Algebra and Precalculus

Each year, more than 1,000,000 students take college algebra and precalculus courses.

The focus in most of these courses is on preparing the students for calculus.

BUT, only a small percentage of these students ever go on to start calculus.
Based on several studies of enrollment flows from college algebra to calculus:

- Less than 5% of the students who start college algebra courses ever start Calculus I

- The typical DFW rate in college algebra is typically well above 50% (as high as 90%)

- Virtually none of the students who pass college algebra courses ever start Calculus III

- Perhaps 30-40% of the students who pass precalculus courses ever start Calculus I
The Provost of one of the largest two year colleges in the nation recently singled out college algebra as the one course that, by far, is most responsible for the school losing students each year.

Their DFW rate in college algebra has been as high as 90%!!!
The mayor’s Economic Development Council of San Antonio recently identified college algebra as one of the major impediments to the city developing the kind of technologically sophisticated workforce it needs.

The mayor appointed special task force including representatives from all 11 colleges in the city plus business, industry and government to change the focus of college algebra to make the courses more responsive to the needs of the city, the students, and local industry.
Some Questions

Why do the majority of these 1,000,000+ students a year take college algebra courses?

Are these students well-served by standard “college algebra” courses?

If not, what kind of mathematics do these students really need?

And, what does this imply about high school algebra courses?
The Status of Calculus

At many colleges, calculus is treated as the holy grail – the ultimate goal of virtually all students.

And, the same can be said about most high school programs as well – everything is designed to prepare and lead students toward calculus.
The Status of Calculus

Does it make sense to treat calculus as the holy grail – the ultimate goal of virtually all students?

Remember the knight in the 3rd Indiana Jones movie -- his sole purpose for centuries was to protect the holy grail from everyone. Do we, as mathematicians, want to play that same role?
Calculus Enrollments

- Fewer than 500,000 students take calculus in college, a number that at best has held steady over the last 20 years.

- Today, on the order of 1,000,000 students take calculus in high school each year – a number that has been growing at over 6% per year.

- College calculus is rapidly heading toward becoming a developmental offering.
High School Enrollment Flows

Historically, roughly 50% of students who successfully completed any course continued on to the succeeding course.

Over the last 15 years, the continuation rate from introductory to intermediate algebra has increased to about 85%, largely due to implementation of NCTM-inspired courses and curricula.

While high school students are taking far more math, our developmental enrollment has been increasing at alarming rates. There is a huge disconnect here.
Developmental programs and traditional algebra-oriented courses in college are a great abyss into which millions of students are pushed each year and vanishingly few of them ever manage to climb out.

We are offering the **Wrong Courses** to the **Wrong Students** for the **Wrong Reasons**!
Why Do So Many Students Fail?

They have seen virtually all of a standard skills-based algebra course in high school.

They do not see themselves ever using any of the myriad of techniques and tricks in the course (and they are right about that).

They equate familiarity with mastery, so they don’t apply themselves until far too late and they are well down the road to failure.
The Needs of Our Students

The reality is that virtually none of the students we face in these courses today or in the future will become math or STEM majors.

They take these courses to fulfill Gen Ed requirements or requirements from other disciplines.

What do those other disciplines want their students to bring from math courses?
CRAFTY held a series of workshops with leading educators from 23 quantitative disciplines to inform the mathematics community of the current mathematical needs of each discipline.

The results are summarized in two MAA Reports volumes.
The Common Threads

• Conceptual Understanding, not rote manipulation.

• Realistic applications via mathematical modeling that reflect the way mathematics is used in other disciplines and on the job.

• Statistical reasoning is the *primary* mathematical topic needed for almost all other disciplines.

• Emphasis on data and data analysis.

• The use of technology (though typically Excel, not graphing calculators).
What the Physicists Said

- Students need conceptual understanding first, and some comfort in using basic skills; then a deeper approach and more sophisticated skills become meaningful.
- Conceptual understanding is more important than computational skill.
- Computational skill without theoretical understanding is shallow.
• The sciences are increasingly seeing students who are quantitatively ill-prepared.

• The biological sciences represent the largest science client of mathematics education.

• The current mathematics curriculum for biology majors does not provide biology students with appropriate quantitative skills.
What the Biologists Said

• Mathematics provides a language for the development and expression of biological concepts and theories. It allows biologists to summarize data, to describe it in logical terms, to draw inferences, and to make predictions.

• Statistics, modeling and graphical representation should take priority over calculus.

• Students need descriptive statistics, regression analysis, multivariate analysis, probability distributions, simulations, significance and error analysis.
What the Biologists Said

The quantitative skills needed for biology:

- The meaning and use of variables, parameters, functions, and relations.
- To formulate linear, exponential, and logarithmic functions from data or from general principles.
- To understand the periodic nature of the sine and cosine functions.
- The graphical representation of data in a variety of formats – histograms, scatterplots, log-log graphs (for power functions), and semi-log graphs (for exponential and log functions).
What Business Faculty Said

• Courses should stress problem solving, with the incumbent recognition of ambiguities.
• Courses should stress conceptual understanding (motivating the math with the “why’s” – not just the “how’s”).
• Courses should stress critical thinking.
• An important student outcome is their ability to develop appropriate models to solve defined problems.
Mathematical Needs of Partners

• In discussions with faculty from the lab sciences, it becomes clear that most courses for non-majors (and even those for majors in many areas) use almost no mathematics in class.

• Mathematics arises almost exclusively in the lab when students have to analyze experimental data and then their weak math skills become dramatically evident.
Mathematical Needs of Partners

• In discussions with faculty from the social sciences, it is clear that most courses use little traditional mathematics, but heavily use statistical reasoning and methods.

• Mathematics arises almost exclusively when students have to analyze data, usually taken from the web or published reports, and then their weak math skills similarly become dramatically evident.
Implications for Algebra

Very few students need a skills-oriented course. They need a modeling-based course that:

- emphasizes realistic applications that mirror what they will see and do in other courses;
- emphasizes conceptual understanding;
- emphasizes data and its uses, including both fitting functions to data and statistical methods and reasoning;
- better motivates them to succeed;
- better prepares them for other courses.
Further Implications

If we don’t offer courses that meet the current needs of the other disciplines who send us the bulk of our students, they will eventually stop requiring our courses.

A recent article in the New York Times weekend magazine by sociologist Andrew Hacker entitled “Is Algebra Necessary?” builds a very compelling case for either a different approach in algebra courses or no algebra at all.

The article reportedly has sparked a national debate, but mostly outside of mathematics.
A TYPICAL American school day finds some six million high school students and two million college freshmen struggling with algebra. In both high school and college, all too many students are expected to fail. Why do we subject American students to this ordeal? I've found myself moving toward the strong view that we shouldn't.
Looking Ahead

What will your department do if half of its sections are lost because of dropped requirements?

Do you seriously think that you can have requirements reinstated once they are dropped?

We must be proactive in changing the focus of our courses before very unpleasant things happen.
What Our Students Really Need

A focus on Mathematical Modeling leads to the **Right Courses** for the **Right Students** for the **Right Reasons**!
What Our Students Really Need

But, you can’t just emphasize Mathematical Modeling without a very strong emphasis on conceptual understanding!

The two **must** go together.
Further Implications

If we focus only on developing manipulative skills without developing conceptual understanding, we produce nothing more than students who are only Imperfect Organic Clones of a TI-89.
Conceptual Understanding

• What does conceptual understanding mean?
• How do you recognize its presence or absence?
• How do you encourage its development?
• How do you assess whether students have developed conceptual understanding?
What Does the Slope Mean?

Comparison of student response to a problem on the final exams in Traditional vs. Reform College Algebra/Trig

Brookville College enrolled 2546 students in 2006 and 2702 students in 2008. Assume that enrollment follows a linear growth pattern.

a. Write a linear equation giving the enrollment in terms of the year \( t \).

b. If the trend continues, what will the enrollment be in the year 2016?

c. What is the slope of the line you found in part (a)?

d. Explain, using an English sentence, the meaning of the slope.

e. If the trend continues, when will there be 3500 students?
Responses in Traditional Class

1. The meaning of the slope is the amount that is gained in years and students in a given amount of time.
2. The ratio of students to the number of years.
3. Difference of the y’s over the x’s.
4. Since it is positive it increases.
5. On a graph, for every point you move to the right on the x-axis. You move up 78 points on the y-axis.
6. The slope in this equation means the students enrolled in 2006. 
   \[ Y = MX + B \]
7. The amount of students that enroll within a period of time.
8. Every year the enrollment increases by 78 students.
9. The slope here is 78 which means for each unit of time, (1 year) there are 78 more students enrolled.
Responses in Traditional Class

10. No response
11. No response
12. No response
13. No response
14. The change in the $x$-coordinates over the change in the $y$-coordinates.
15. This is the rise in the number of students.
16. The slope is the average amount of years it takes to get 156 more students enrolled in the school.
17. Its how many times a year it increases.
18. The slope is the increase of students per year.
<table>
<thead>
<tr>
<th>Responses in Reform Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. This means that for every year the number of students increases by 78.</td>
</tr>
<tr>
<td>2. The slope means that for every additional year the number of students increase by 78.</td>
</tr>
<tr>
<td>3. For every year that passes, the student number enrolled increases 78 on the previous year.</td>
</tr>
<tr>
<td>4. As each year goes by, the # of enrolled students goes up by 78.</td>
</tr>
<tr>
<td>5. This means that every year the number of enrolled students goes up by 78 students.</td>
</tr>
<tr>
<td>6. The slope means that the number of students enrolled in Brookville college increases by 78.</td>
</tr>
<tr>
<td>7. Every year after 2006, 78 more students will enroll at Brookville college.</td>
</tr>
<tr>
<td>8. Number of students enrolled increases by 78 each year.</td>
</tr>
</tbody>
</table>
Responses in Reform Class

9. This means that for every year, the amount of enrolled students increase by 78.
10. Student enrollment increases by an average of 78 per year.
11. For every year that goes by, enrollment raises by 78 students.
12. That means every year the # of students enrolled increases by 2,780 students.
13. For every year that passes there will be 78 more students enrolled at Brookville college.
14. The slope means that every year, the enrollment of students increases by 78 people.
15. Brookville college enrolled students increasing by 0.06127.
16. Every two years that passes the number of students which is increasing the enrollment into Brookville College is 156.
17. This means that the college will enroll \(0.0128\) more students each year.
18. By every two year increase the amount of students goes up by 78 students.
19. The number of students enrolled increases by 78 every 2 years.
Both groups had comparable ability to calculate the slope of a line. (In both groups, several students used $\Delta x/\Delta y$.)

It is far more important that our students understand what the slope means in context, whether that context arises in a math course, or in courses in other disciplines, or eventually on the job.

Unless explicit attention is devoted to emphasizing the conceptual understanding of what the slope means, the majority of students are not able to create viable interpretations on their own. And, without that understanding, they are likely not able to apply the mathematics to realistic situations.
Further Implications

If students can’t make their own connections with a concept as simple as the slope of a line, they won’t be able to create meaningful interpretations and connections on their own for more sophisticated mathematical concepts. For instance,

• What is the significance of the base (growth or decay factor) in an exponential function?
• What is the meaning of the power in a power function?
• What do the parameters in a realistic sinusoidal model tell about the phenomenon being modeled?
• What is the significance of the factors of a polynomial?
• What is the significance of the derivative of a function?
• What is the significance of a definite integral?
Developing Conceptual Understanding

Conceptual understanding cannot be just an add-on. It must permeate every course and be a major focus of the course.

Conceptual understanding must be accompanied by realistic problems in the sense of mathematical modeling.

Conceptual problems must appear in all sets of examples, on all homework assignments, on all project assignments, and most importantly, on all tests.

Otherwise, students will not see them as important.
A Major Challenge: Statistics

The most critical need of the faculty in the lab and social sciences is for their students to know statistics. How can we integrate statistical ideas and methods into math courses at all levels?
Statistics and the Common Core

The Common Core curriculum recognizes the tremendous importance of statistics. It calls for introducing students to statistical ideas and methods starting in 6th grade.
Statistics and the Common Core

By the end of 12th grade, every student is expected to have seen the equivalent of a very solid introductory statistics course – probably considerably more material than is in most college courses today, and certainly doing it at a much more sophisticated level than is usually done in most current statistics courses.
The likelihood is that most programs based on the Common Core will introduce the statistics via independent units that are not connected to any of the algebra-oriented topics. But, it is much more effective if the statistical ideas can be integrated as natural applications of algebraic concepts and methods.
Integrating Statistics into Mathematics

• Students see the equation of a line in pre-algebra, in elementary algebra, in intermediate algebra, in college algebra, and in precalculus. Yet many still have trouble with it in calculus.

• They see all of statistics ONCE in an introductory statistics course. But statistics is far more complex, far more varied, and often highly counter-intuitive, yet they are then expected to understand and use a wide variety of the statistical ideas and methods in their lab science and other courses.
But, there are some important statistical issues that need to be addressed. For instance:

1. Most sets of data, especially in the sciences, only represent a single sample drawn from a much larger underlying population. How does the regression line based on a single sample compare to the lines based on other possible samples?

2. The correlation coefficient only applies to a linear fit. What significance does it have when you are fitting a nonlinear function to data?
Should $x$ Mark the Spot?

All other disciplines focus globally on the entire universe of $a$ through $z$, with the occasional contribution of $\alpha$ through $\omega$.

Only mathematics focuses on a single spot, called $x$.

Newton’s Second Law of Motion: $y = mx$,

Einstein’s formula relating energy and mass: $y = c^2x$,

The Ideal Gas Law: $yz = nRx$.

Students who see only $x$’s and $y$’s do not make the connections and cannot apply the techniques learned in math classes when other letters arise in other disciplines.
Should $x$ Mark the Spot?

Kepler’s third law expresses the relationship between the average distance of a planet from the sun and the length of its year.

If it is written as $y^2 = 0.1664x^3$, there is no suggestion of which variable represents which quantity.

If it is written as $t^2 = 0.1664D^3$, a huge conceptual hurdle for the students is eliminated.
Some Illustrative Examples and Problems
Topic 1: Data and Statistics

Introduction to data and statistical measures of the data, including an introduction to the notion of function in terms of two-variable data sets.
# Monthly Rainfall in Orlando

<table>
<thead>
<tr>
<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>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>

![Bar chart showing monthly rainfall in Orlando]
### Average Worldwide Temps, °C

<table>
<thead>
<tr>
<th>Year</th>
<th>1880</th>
<th>1900</th>
<th>1920</th>
<th>1940</th>
<th>1960</th>
<th>1980</th>
<th>1990</th>
<th>2000</th>
</tr>
</thead>
</table>

(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.
Topic 2: Behavior of Functions

The behavior of functions given as data in tables and as graphs, including increasing/decreasing, turning points, concave up/down, inflection points (including the logistic function and the normal distribution function patterns).
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.
Estimate the location of the inflection point(s) of the following function:

<table>
<thead>
<tr>
<th>$t$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>3</td>
<td>8</td>
<td>20</td>
<td>35</td>
<td>46</td>
<td>50</td>
</tr>
</tbody>
</table>
Topic 3: Linear Functions & Models

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 $84^\circ$.

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>
<tr>
<td>T</td>
<td>69</td>
<td>83</td>
<td>80</td>
<td>83</td>
<td>81</td>
<td>84</td>
<td>76</td>
<td></td>
</tr>
<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>
<td></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.
A set of data is normally distributed with mean \( \mu = 8.4 \) and standard deviation \( \sigma = 2.3 \).

Find the value of \( x \) that is

(a) 1.8 standard deviations above the mean;
(b) 0.6 standard deviations below the mean.

\[
z = \frac{x - \mu}{\sigma}
\]
z-Values

Car speeds on a highway are normally distributed with mean $\mu = 58$ mph and standard deviation $\sigma = 4$ mph. Find the speed $x$ of a car that is going

1. 1.4 standard deviations above the mean.
2. 0.3 standard deviations below the mean.
Student Projects

The number of sexual harassment cases filed as a function of time.
The likelihood of car crashes as a function of blood alcohol level.
The growth of the prison population as a function of time.
The time of high tide at a beach as a function of the day of the month.
The amount of solid waste generated per person as a function of time.
The time for water to come to a boil as a function of the volume of water.
The size of the human cranium over time during the last three million years.
The results of a serial dilution experiment in biology lab.
The growth in the Dow-Jones average as a function of time.
The Gini Index measuring the spread of rich versus poor in the population over time.
The number of immigrants who entered the U.S. over time.
The mean annual income as a function of the level of education.
Nonlinear families of functions:

- exponential growth and decay, especially 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.
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.
For linear functions, the successive differences are constant. That constant difference is related to the slope of the line.

For exponential functions $y = A \cdot b^x$, 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*. 
Behavior of Exponential Functions

- $b > 1$
- $0 < b < 1$
# Exponential Functions vs. the Base

For exponential growth functions
\[ y = A b^x, \quad b > 1, \] the larger the growth factor \( b \), the faster the function grows.

For exponential decay functions
\[ y = A b^x, \quad 0 < b < 1, \] the smaller the decay factor \( b \), the more quickly the function dies out.
Exponential Functions vs. the Base $b$

For $b = 1.5$, $b = 1.7$, $b = 1.3$, $b = 0.8$, $b = 0.6$, and $b = 0.4$. 

Graphs showing exponential growth and decay for different values of $b$. 

- For $b > 1$, the function grows exponentially. 
- For $0 < b < 1$, the function decays exponentially.
An Exponential Growth Model

In 1990, 5.01 billion metric tons of carbon dioxide were emitted into the atmosphere in the United States. In 2002, 5.80 billion metric tons were emitted.

a. Write an exponential function to model the amount of carbon dioxide emitted into the atmosphere as a function of the number of years since 1990.

b. Use the exponential function to estimate the amount emitted in 2012.
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 Prozac be after 7 days?

• How long until the Prozac level is 10 mg.?

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

Need to use logs to solve algebraically.
Modeling the Decay of a Drug

For many drugs, such as Prozac, a patient actually takes a given dose on a daily or other fixed repeated basis. The mathematics is slightly more complicated and, if time permits, we will get into that later.
Drug Level Simulation
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 $b$. 
Doubling Time/Half-Life Simulation
Doubling Time & Half-Life

- The population of Brazil was 188.1 million in 2006 and was growing at an annual rate of 1.04%.
  
a. Find an expression for Brazil's population at any time \( t \), where \( t \) is the number of years since 2006.
  
b. Predict the 2015 population if the trend continues.
  
c. Use logarithms to find the doubling time exactly.

- The famous Cro-Magnon cave paintings are found in the Lascaux Cave in France. If the level of radioactive carbon-14 in charcoal in the cave is approximately 14\% of the level of living wood, estimate how long ago the paintings were made.
Fitting Exponential Functions

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>

![Graph showing wind power generation over years]
(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: $52.497(1.1373)^t = 2 	imes 52.497$.
(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.
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.
Behavior of Power Functions $y = x^p$

$p > 1$: Increasing and concave up

$0 < p < 1$: Increasing and concave down

$p < 0$: Decreasing and concave up

All pass through $(1, 1)$
Power Functions versus the Power $p$

When $p > 1$, the larger $p$ is, the faster the function grows.

When $0 < p < 1$, the larger $p$ is, the more quickly the function grows in a concave down manner.

When $p < 0$, the more negative $p$ is, the more quickly the function dies out.
Power Functions vs. the Power \( p \)
Biologists have long observed that the larger the area of a region, the more species live there. The relationship is best modeled by a power function. Puerto Rico has 40 species of amphibians and reptiles on 3459 square miles and Hispaniola (Haiti and the Dominican Republic) has 84 species on 29,418 square miles.

(a) Determine a power function that relates the number of species of reptiles and amphibians on a Caribbean island to its area.

(b) Use the relationship to predict the number of species of reptiles and amphibians on Cuba, which measures 44218 square miles.
The accompanying table and associated scatterplot give some data on the area (in square miles) of various Caribbean islands and estimates on the number of species of amphibians and reptiles living on each.

<table>
<thead>
<tr>
<th>Island</th>
<th>Area</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redonda</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Saba</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Montserrat</td>
<td>40</td>
<td>9</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>3459</td>
<td>40</td>
</tr>
<tr>
<td>Jamaica</td>
<td>4411</td>
<td>39</td>
</tr>
<tr>
<td>Hispaniola</td>
<td>29418</td>
<td>84</td>
</tr>
<tr>
<td>Cuba</td>
<td>44218</td>
<td>76</td>
</tr>
</tbody>
</table>
How Power Regression Works

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

Method Used:

- Transform the data to \(\log y \text{ vs. } \log x\)
- Fit a linear function to the transformed data.
- Undo the transformation to create the power function.
Undoing the log-log Transformation

The linear fit to the \((\log A, \log S)\) data is:

\[
y = 0.310x + 0.485 \quad \text{or} \quad \log S = 0.310 \log A + 0.485
\]

To undo the logs, we take powers of 10:

\[
10^\log S = S = 10^{0.310 \log A + 0.485}
\]

\[
= 10^{0.310 \log A} \cdot 10^{0.485}
\]

\[
= 3.055 \cdot 10^{\log (A^{0.310})}
\]

\[
= 3.055 A^{0.310}
\]
Biologists have observed that the flying speed of animals tends to be related to their overall body length. The accompanying table gives the lengths $L$ of various organisms, in centimeters, and their top flying speed, $S$, in centimeters per second.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Length (cm)</th>
<th>Speed (cm/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit fly</td>
<td>0.2</td>
<td>190</td>
</tr>
<tr>
<td>Horse fly</td>
<td>1.3</td>
<td>660</td>
</tr>
<tr>
<td>Hummingbird</td>
<td>8.1</td>
<td>1120</td>
</tr>
<tr>
<td>Dragonfly</td>
<td>8.5</td>
<td>1000</td>
</tr>
<tr>
<td>Bat</td>
<td>11.0</td>
<td>690</td>
</tr>
<tr>
<td>Flying fish</td>
<td>34.0</td>
<td>1560</td>
</tr>
<tr>
<td>Pintail duck</td>
<td>56.0</td>
<td>2280</td>
</tr>
<tr>
<td>Swan</td>
<td>120.0</td>
<td>1880</td>
</tr>
<tr>
<td>Pelican</td>
<td>160.0</td>
<td>2280</td>
</tr>
</tbody>
</table>

a. Explain why a power function would be the appropriate function to use as a model. Which variable should be treated as the independent variable and which the dependent variable?

b. Find a power function that fits these data.

c. Transform the data to linearize it, then find the linear function that best fits the transformed data, and finally undo the transformation algebraically. Do you get the same function as in part (b)?
Identify each of the following functions (a) - (n) as linear, exponential, logarithmic, or power. In each case, explain your reasoning.

(g) \( y = 1.05^x \)  
(h) \( y = x^{1.05} \)  

(i) \( y = (0.7)^x \)  
(j) \( y = x^{0.7} \)  

(k) \( y = x^{-\frac{1}{2}} \)  
(l) \( 3x - 5y = 14 \)

(m) | x | y  | (n) | x | y  |
--- | --- | --- | --- | --- | --- |
| 0  | 3  | 0  | 5  |
| 1  | 5.1| 1  | 7  |
| 2  | 7.2| 2  | 9.8|
| 3  | 9.3| 3  | 13.7|
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?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>You could, but probably shouldn’t, use only the correlation coefficient.</td>
</tr>
<tr>
<td>2.</td>
<td>You could use the sum of the squares to measure and compare different fits.</td>
</tr>
<tr>
<td>3.</td>
<td>You can “eyeball” how well the function appears to fit the data.</td>
</tr>
</tbody>
</table>
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.

Topic 5: Modeling with Polynomial Functions

Emphasis on the behavior of polynomials and modeling, primarily by fitting polynomials to data
For the polynomial shown,

a. What is the minimum degree? Give two different reasons for your answer.
b. What is the sign of the leading coefficient? Explain.
c. What are the real roots?
d. What are the linear factors?
e. How many complex roots does the polynomial have?
# 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.
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.
Topic 6: Extended Families of Functions

Extending the basic families of functions using shifting, stretching, and shrinking, including:

- applying ideas on shifting and stretching to fitting extended families of functions to sets of data
- statistical ideas such as the distribution of sample means, the Central Limit Theorem, and confidence intervals.
For the zigzag function \( y = \text{zig} \ (x) \) shown, sketch the graph of:

(a) \( y = 3 + \text{zig} \ (x) \)
(b) \( y = \text{zig} \ (x - 2) \)
(c) \( y = \text{zig} \ (x + 3) \)
(d) \( y = 3 \text{ zig} \ (x) \)
(e) \( y = -4\text{zig} \ (x) \)
(f) \( y = \text{zig} \ (2x) \)
(g) \( y = \text{zig} \ (\frac{1}{2}x) \)

Be sure to mark all turning points and zeros.
A Temperature Experiment

An experiment is conducted to study the rate at which temperature changes. A temperature probe is first heated in a cup of hot water and then pulled out and placed into a cup of cold water. The temperature of the probe, in °C, is measured every second for 36 seconds and recorded in the following table.

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>42.3</td>
<td>36.03</td>
<td>30.85</td>
<td>26.77</td>
<td>23.58</td>
<td>20.93</td>
<td>18.79</td>
<td>17.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.78</td>
<td>8.78</td>
<td>8.78</td>
<td>8.78</td>
<td>8.66</td>
<td>8.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Find a function that fits this data.
The data suggest an exponential decay function, but the points don’t decay to 0.

To find a function, one first has to shift the data values down to get a transformed set of data that decay to 0.

Then one has to fit an exponential function to the transformed data. Finally, one has to undo the transformation by shifting the resulting exponential function. \[ T = 8.6 + 35.439(0.848)^t \].
The normal distribution function is

\[ N(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} \]

It makes for an excellent example involving both stretching and shifting functions and a composite function.
Match each of the four normal distributions (a)-(d) with one of the corresponding sets of values for the parameters \( \mu \) and \( \sigma \). Explain your reasoning.

(i) \( \mu = 85, \sigma = 1 \)
(ii) \( \mu = 100, \sigma = 12 \)
(iii) \( \mu = 115, \sigma = 12 \)
(iv) \( \mu = 115, \sigma = 8 \)
(v) \( \mu = 100, \sigma = 6 \)
(vi) \( \mu = 85, \sigma = 7 \)
Normal Distribution Rule of Thumb

In any normal distribution:

1. Approximately 68% of all entries lie within 1 standard deviation, $\sigma$, of the mean $\mu$.

2. Approximately 95% of all entries lie within 2 standard deviations, $2\sigma$, of the mean $\mu$.

3. Approximately 99.7% of all entries lie within 3 standard deviations, $3\sigma$, of the mean $\mu$. 
Typical Problem

Major league batting averages are roughly normally distributed with mean 268 and standard deviation 12.

1. What percentage of hitters have averages between 280 and 292?
2. What percentage of hitters have averages below 256?
3. What percentage of hitters have averages between 244 and 280?
4. What is the probability that a randomly selected hitter has an average above 280?
5. What is the lowest possible batting average for a player who is in the top 2.5% of batters?
The Two Most Important Notions About Variability in Statistics

1. Randomness and Variability
   Within Samples (measured by the standard deviation or the interquartile range)
   Between Samples

2. The Effects of Sample Size on various statistical measures
The distribution of sample means is the population consisting of all possible sample means of a given size $n$ drawn from an underlying population having mean $\mu$ and standard deviation $\sigma$. The Central Limit Theorem describes the shape of this population, its mean, and its standard deviation.
CLT Simulation
The Central Limit Theorem

1. If the underlying population is roughly normal, then the distribution of sample means is effectively normally distributed.

2. If the sample size \( n \) is large enough (\( n > 30 \)), the distribution of sample means is effectively normally distributed, regardless of the shape of the underlying distribution.

3. The mean of the distribution of sample means is

\[
\mu_{\bar{x}} = \mu
\]

4. The standard deviation is

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]
The Central Limit Theorem can be interpreted in terms of stretching and shifting functions -- the mean of the distribution of sample means corresponds to a horizontal shift of a standard normal distribution function and its standard deviation can be thought of in terms of a stretch or a squeeze, depending on the sample size $n$. 

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \]
Key Methods of Inferential Statistics

1. Constructing Confidence Intervals to estimate a population parameter based on a sample.

2. Hypothesis Testing to verify whether a claim about a population parameter may be valid.

Both involve the Central Limit Theorem, which algebraically entails ideas on shifting and stretching functions.
Confidence Intervals for the Mean

Given a sample of size $n$ with mean $\bar{x}$ and standard deviation $s$, we seek to estimate the population mean $\mu$ using an interval

$$\bar{x} \pm z \frac{\sigma}{\bar{x}}$$

centered at the sample mean in such a way that we have a given level of confidence (say 95%) that the interval actually does contain $\mu$. Equivalently, 95% of the intervals so constructed should contain $\mu$. 
Confidence Interval Simulation
Other Confidence Intervals

The comparable ideas and virtually the identical methods can be applied to produce other types of confidence intervals; for instance:

1. Confidence intervals to estimate the population proportion $\pi$.

2. Confidence intervals for the difference of means $\mu_1 - \mu_2$ (based on two samples).

3. Confidence intervals for the difference of proportions $\pi_1 - \pi_2$ (based on two samples).
Hypothesis Testing for the Mean

Suppose that there is some belief or claim made about the mean $\mu$ of some population. We wish to test whether this claim is valid by considering a single random sample of size $n$ with mean $\bar{x}$ and standard deviation $s$.

The claim about the mean is called the null hypothesis $H_0$ and the challenge to the claim is the alternate hypothesis $H_a$. 
Hypothesis Testing for the Mean

The null hypothesis asserts that the claimed value $\mu_0$ for the mean is true:

$$H_0: \mu = \mu_0$$

The alternate hypothesis may state that

1. the claim is wrong – that is, $H_a: \mu \neq \mu_0$
2. the true mean is larger – that is, $H_a: \mu > \mu_0$
3. the true mean is smaller – that is, $H_a: \mu < \mu_0$
Hypothesis Testing for the Mean

The decision of whether we *reject* the null hypothesis in favor of the alternate hypothesis or *fail to reject* the null hypothesis depends on how far from or close to the claimed mean the sample mean $\bar{x}$ falls.

Note that we do not *accept* the alternate hypothesis; we only *fail to reject* it and these are not the same thing.
Hypothesis Testing for the Mean

There are always two possible errors in any such conclusion: either failing to reject a false null hypothesis or rejecting a valid null hypothesis. In the process, we must accept a predetermined element of risk of making an erroneous decision. The risk levels are typically $\alpha = 10\%, 5\%, 2\%, 1\%$, or $0.5\%$. The key is whether or not the sample mean $\bar{x}$ falls among the least likely of all possible sample means for a given risk level.
Hypothesis Testing

There are very strong parallels between the ideas on which hypothesis testing are based and the principles of our criminal justice system.

It is helpful to draw these parallels to help students understand the statistical ideas.
Hypothesis Testing Simulation
Other Hypothesis Tests

There are many other types of hypothesis tests. However, comparable ideas and virtually the identical methods apply to many of them, including:

1. Hypothesis tests on the population proportion $\pi$.
2. Hypothesis tests for the difference of means $\mu_1 - \mu_2$ (based on two samples).
3. Hypothesis tests for the difference of proportions $\pi_1 - \pi_2$ (based on two samples).
Topic 7: Functions of Two Variables

Functions of several variables using tables, contour plots, and formulas with multiple variables.

These are the same multiple representations of functions used normally with functions of a single variable.
One of the major themes that comes from many different disciplines is a request that students need to know about functions of several variables.

When we hear this, most mathematicians interpret it as a request that we teach about partial derivatives or multiple integrals.

Instead, what the other disciplines need is for their students to be able to read/interpret tables and contour diagrams and to work with simple expressions involving three or more interrelated variables, such as the Ideal Gas Law: $PV = nRT$. 
## The Temperature Humidity Index

<table>
<thead>
<tr>
<th>Temperature</th>
<th>80</th>
<th>82</th>
<th>84°</th>
<th>86°</th>
<th>88°</th>
<th>90°</th>
<th>92°</th>
<th>94°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rel Hum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>78</td>
<td>80</td>
<td>81</td>
<td>83</td>
<td>84</td>
<td>86</td>
<td>88</td>
<td>89</td>
</tr>
<tr>
<td>20%</td>
<td>79</td>
<td>80</td>
<td>82</td>
<td>83</td>
<td>85</td>
<td>86</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>30%</td>
<td>79</td>
<td>80</td>
<td>82</td>
<td>84</td>
<td>86</td>
<td>88</td>
<td>90</td>
<td>93</td>
</tr>
<tr>
<td>40%</td>
<td>80</td>
<td>81</td>
<td>84</td>
<td>87</td>
<td>89</td>
<td>93</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>45%</td>
<td>80</td>
<td>82</td>
<td>84</td>
<td>87</td>
<td>89</td>
<td>93</td>
<td>96</td>
<td>100</td>
</tr>
</tbody>
</table>

What are reasonable values for the domain and range of the THI function? Predict the THI value for a temperature of 100° and 10% relative humidity. Predict the value of the THI for \( T = 85 \) and \( H = 40\% \) relative humidity. If the temperature is 90° and the THI is 90, what is the relative humidity? Does the THI value increase more rapidly at low temperature values or high temperature values? Does the THI value increase more rapidly at low relative humidity levels or high relative humidity values?
Contour diagrams are used very extensively in most other fields to represent functions of two variables graphically. In comparison, mathematics usually use surface plots.

The *contours* (or level curves) for a function of two independent variables represent all points that are at the same height or have the same value.
Contour Diagram of Island
Rules for Contour Diagrams

1. There must be a uniform difference between contour values – say 10’s or 100’s or ...

2. Contours with different values cannot intersect (that would violate the definition of a function).

3. The closer the successive contours, the faster the function changes.

4. The further apart the successive contours, the slower the function changes.
Crater Lake
What is the greatest depth, in feet, in the lake?
Using the scale shown at the bottom left, estimate the average slope of the bottom of the lake in the northeast corner just above Skell Head.

In the north central part of the lake (just south of Pumice Point), there is a collection of nearly circular contours. Explain the shape of the terrain under the water surface there.

Just to the east and south of that collection of circular contours, there is a relatively large area enclosed by a single contour. Describe the shape of the bottom of the lake within that contour. What are the approximate dimensions of that region?

There is one spot within Crater Lake where visitors can go for a swim in the rather chilly waters. Where would you expect that beach to be located? Explain your reasoning.
Human Surface Area Function

The contour diagram for the human surface area function
\[ S(H, W) = 15.64W^{0.425}H^{0.725} \]
is to be plotted.

a. The level curves might be increasing and concave up, increasing and concave down, decreasing and concave up, or decreasing and concave down. Decide which of the four possibilities is correct.

b. Having decided on the shape of each of the contours, the next issue is whether the successive contours get larger or smaller in a certain direction. That is, as you move upward or toward the right, do the contours correspond to larger or smaller values for \( S \)?

c. The third issue with a contour diagram is to determine the spacing between the successive contours – do they get closer together or further apart. Use several different contours to decide whether the distances between the successive contours get smaller or larger. What is the practical significance of your answer?
Multivariate Regression

Just as the regression line $y = mx + b$ is the linear function that is the best fit (in the least squares sense) to a set of $x$ - $y$ data in the plane, the regression plane $y = m_1x_1 + m_2x_2 + b$ is the linear function of two independent variables that is the best fit to a set of $x_1$ - $x_2$ – $y$ data in space in the sense that the sum of the squares of the vertical distances from the points to the plane is a minimum.

This is typically found using software, not by hand.

It can be extended to data based on more than two independent variables.
### Multivariate Regression Example

A study is conducted to see how serum cholesterol level ($Y$) is related to a person’s weight ($x_1$) and systolic blood pressure ($x_2$).

<table>
<thead>
<tr>
<th>Serum Cholesterol ($y$)</th>
<th>Weight ($x_1$)</th>
<th>Systolic Blood Pressure ($x_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>152.2</td>
<td>59.0</td>
<td>108</td>
</tr>
<tr>
<td>158.0</td>
<td>52.3</td>
<td>111</td>
</tr>
<tr>
<td>157.0</td>
<td>56.0</td>
<td>115</td>
</tr>
<tr>
<td>155.0</td>
<td>53.5</td>
<td>116</td>
</tr>
<tr>
<td>156.0</td>
<td>58.7</td>
<td>117</td>
</tr>
<tr>
<td>159.4</td>
<td>60.1</td>
<td>120</td>
</tr>
<tr>
<td>169.1</td>
<td>59.0</td>
<td>124</td>
</tr>
<tr>
<td>181.0</td>
<td>62.4</td>
<td>127</td>
</tr>
<tr>
<td>174.9</td>
<td>65.7</td>
<td>122</td>
</tr>
<tr>
<td>180.2</td>
<td>63.2</td>
<td>131</td>
</tr>
<tr>
<td>174.0</td>
<td>64.2</td>
<td>125</td>
</tr>
</tbody>
</table>

Regression Equation: $Y = 0.6x_1 + 1.1315x_2 - 7.64$

Coefficient of Determination $R^2 = 0.851$

so 85.1% of variation in data is explained by linear function.

Sum of the squares = 174.9
How to Fit Polynomials to Data

The 10 points (-2,11), (-1, 6), (0, 3), (1, 5), (2, 11), (3, 23), (4, 30), (5, 25), (6, 14), and (7, 2) fall into a cubic pattern.

To fit a cubic

\[ y = ax^3 + bx^2 + cx + d \]

to these points, think of \( Y = y \) as a linear function of the three independent variables

\[ X_1 = x, \quad X_2 = x^2, \quad \text{and} \quad X_3 = x^3, \]

construct the associated 4 column table with values for each \( X \) and for \( Y \), and then “hit” the table with multivariate linear regression.
Contour Plots for Linear Functions

The contour plot for any linear function of two variables, \( z = ax + by + c \) is always a series of parallel and equally spaced lines.

To see why, think about what the contours mean. For each of a set of values of \( z \), say \( z = 100, 200, 300, \ldots \), you have the linear expressions \( ax + by + c = 100, ax + by + c = 200, \ldots \), all of which are normal forms for a set of lines having the same slope and so are parallel. Also, because the values for \( z \) have fixed separations, the lines will have the same fixed separation in their intercepts and so are equally spaced.
Contour Plots for Linear Functions

Suppose you are making a peanut butter and jelly sandwich on white bread. Each slice of bread contains 75 calories, each gram of peanut butter has 6 calories, and each gram of jelly has 2.5 calories. The number of calories in a sandwich with \( P \) grams of peanut butter and \( J \) grams of jelly is \( C = 6P + 2.5J + 150 \).
Topic 8: Modeling Periodic Behavior

Using the sine and cosine as models for periodic phenomena.
Jacksonville Daylight Function

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

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.
Write a possible formula for each of the following trigonometric functions:
Dallas Temperatures

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°?
# Data Everywhere: Ocean Temperatures

<table>
<thead>
<tr>
<th>Location</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>Dauphin Island AL</td>
<td>51</td>
<td>53</td>
<td>60</td>
<td>70</td>
<td>75</td>
<td>82</td>
<td>84</td>
<td>84</td>
<td>80</td>
<td>72</td>
<td>62</td>
<td>56</td>
</tr>
<tr>
<td>Daytona Beach FL</td>
<td>61</td>
<td>59</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>79</td>
<td>80</td>
<td>80</td>
<td>83</td>
<td>79</td>
<td>71</td>
<td>65</td>
</tr>
<tr>
<td>Fort Lauderdale</td>
<td>72</td>
<td>72</td>
<td>74</td>
<td>76</td>
<td>78</td>
<td>80</td>
<td>82</td>
<td>82</td>
<td>82</td>
<td>80</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>Jacksonville FL</td>
<td>57</td>
<td>56</td>
<td>61</td>
<td>68</td>
<td>74</td>
<td>80</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>78</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td>Key West FL</td>
<td>69</td>
<td>70</td>
<td>75</td>
<td>79</td>
<td>82</td>
<td>86</td>
<td>87</td>
<td>87</td>
<td>86</td>
<td>83</td>
<td>76</td>
<td>72</td>
</tr>
<tr>
<td>Miami Beach FL</td>
<td>71</td>
<td>72</td>
<td>75</td>
<td>79</td>
<td>80</td>
<td>83</td>
<td>86</td>
<td>86</td>
<td>83</td>
<td>80</td>
<td>76</td>
<td>72</td>
</tr>
<tr>
<td>Naples FL</td>
<td>66</td>
<td>66</td>
<td>71</td>
<td>77</td>
<td>82</td>
<td>86</td>
<td>87</td>
<td>87</td>
<td>86</td>
<td>81</td>
<td>73</td>
<td>68</td>
</tr>
<tr>
<td>Pensacola FL</td>
<td>58</td>
<td>58</td>
<td>63</td>
<td>71</td>
<td>78</td>
<td>84</td>
<td>86</td>
<td>86</td>
<td>82</td>
<td>74</td>
<td>68</td>
<td>58</td>
</tr>
<tr>
<td>St. Petersburg FL</td>
<td>62</td>
<td>64</td>
<td>68</td>
<td>74</td>
<td>80</td>
<td>84</td>
<td>86</td>
<td>86</td>
<td>84</td>
<td>78</td>
<td>70</td>
<td>64</td>
</tr>
<tr>
<td>Savannah GA</td>
<td>52</td>
<td>52</td>
<td>59</td>
<td>65</td>
<td>72</td>
<td>79</td>
<td>84</td>
<td>85</td>
<td>82</td>
<td>79</td>
<td>64</td>
<td>54</td>
</tr>
<tr>
<td>Cape Hatteras NC</td>
<td>49</td>
<td>46</td>
<td>52</td>
<td>58</td>
<td>68</td>
<td>72</td>
<td>76</td>
<td>80</td>
<td>78</td>
<td>72</td>
<td>58</td>
<td>55</td>
</tr>
<tr>
<td>Myrtle Beach SC</td>
<td>48</td>
<td>50</td>
<td>56</td>
<td>62</td>
<td>70</td>
<td>77</td>
<td>81</td>
<td>82</td>
<td>71</td>
<td>72</td>
<td>61</td>
<td>53</td>
</tr>
<tr>
<td>Charleston SC</td>
<td>50</td>
<td>50</td>
<td>57</td>
<td>64</td>
<td>71</td>
<td>79</td>
<td>82</td>
<td>84</td>
<td>82</td>
<td>73</td>
<td>63</td>
<td>54</td>
</tr>
<tr>
<td>Cape Charles VA</td>
<td>36</td>
<td>39</td>
<td>46</td>
<td>51</td>
<td>68</td>
<td>70</td>
<td>78</td>
<td>78</td>
<td>74</td>
<td>68</td>
<td>54</td>
<td>44</td>
</tr>
<tr>
<td>Washington DC</td>
<td>37</td>
<td>37</td>
<td>46</td>
<td>54</td>
<td>64</td>
<td>74</td>
<td>80</td>
<td>83</td>
<td>78</td>
<td>64</td>
<td>52</td>
<td>41</td>
</tr>
<tr>
<td>San Juan P.R.</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>31</td>
<td>30</td>
</tr>
</tbody>
</table>
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
What Can Be Removed?

How many of you remember that there used to be something called the Law of Tangents?

What happened to this universal law?

Did triangles stop obeying it?

Does anyone miss it?
What Can Be Removed?

- Descartes’ rule of signs
- The rational root theorem
- Synthetic division
- The Cotangent, Secant, and Cosecant

were needed for computational purposes;

Just learn and teach a new identity:

\[ 1 + \tan^2 x = \frac{1}{\cos^2 x} \]