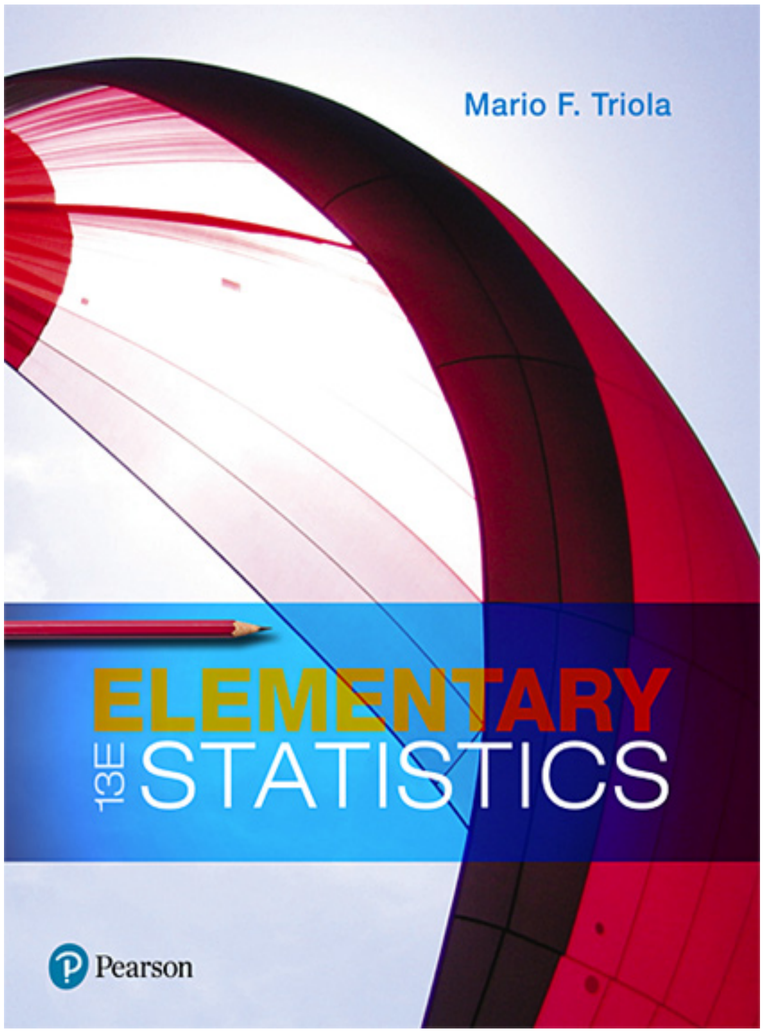
Elementary Statistics

Thirteenth Edition



Chapter 10

Correlation and Regression

ALWAYS LEARNING

Pearson

Correlation and Regression

10-1 Correlation

10-2 Regression

10-3 Prediction Intervals and Variation

10-4 Multiple Regression

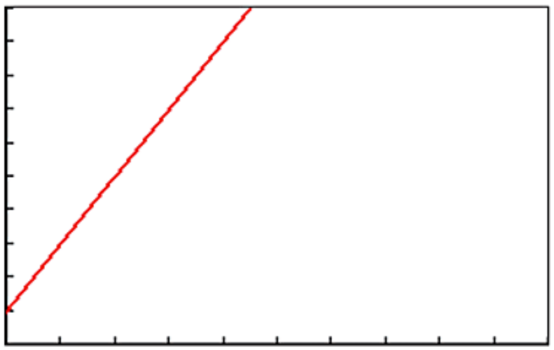
10-5 Nonlinear Regression

Key Concept

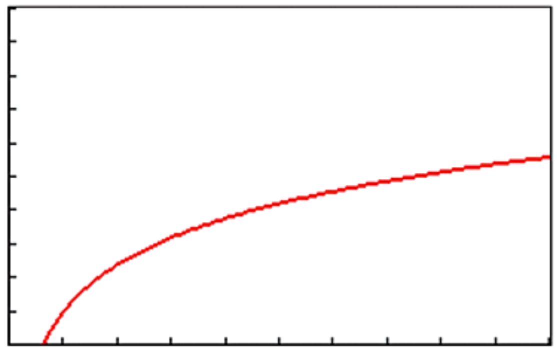
This section is a brief introduction to methods for finding some **nonlinear** functions that fit sample data. We focus on the use of technology because the required calculations are quite complex.

Five Basic Generic Models

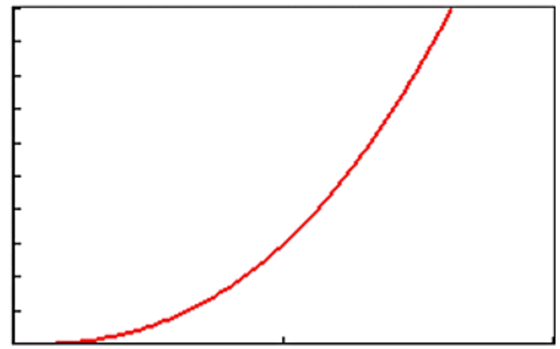
Linear: $y = a + bx$
Example: $y = 1 + 2x$



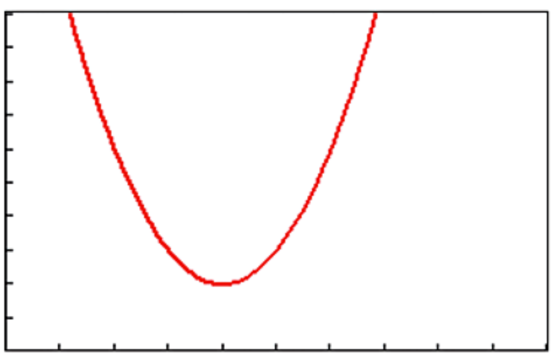
Logarithmic: $y = a + b \ln x$
Example: $y = 1 + 2 \ln x$



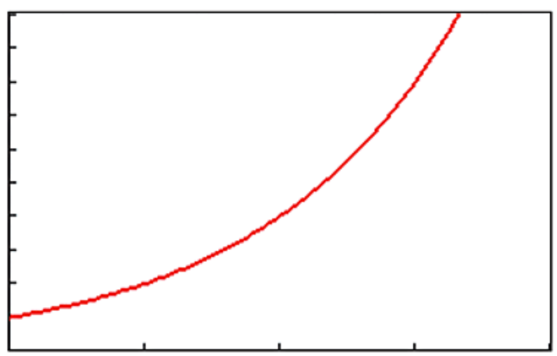
Power: $y = ax^b$
Example: $y = 3x^{2.5}$



Quadratic: $y = ax^2 + bx + c$
Example: $y = x^2 - 8x + 18$



Exponential: $y = ab^x$
Example: $y = 2^x$



Three Basic Rules for Identifying a Good Mathematical Model (1 of 3)

1. Look for a pattern in the graph. Construct a graph, compare it to those shown here, and identify the model that appears to be most similar.
2. Compare values of R^2. For each model being considered, use technology to find the value of the coefficient of determination R^2. Choose functions that result in larger values of R^2, because such larger values correspond to functions that better fit the observed sample data.

Three Basic Rules for Identifying a Good Mathematical Model (2 of 3)

2. Compare values of R^2.
   - Don't place much importance on small differences, such as the difference between $R^2 = 0.984$ and $R^2 = 0.989$.
   - Unlike in Section 10-4, we don't need to use values of adjusted R^2. Because the examples of this section all involve a single predictor variable, it makes sense to compare values of R^2.
   - In addition to R^2, another measure used to assess the quality of a model is the sum of squares of the residuals. See Exercise 18 "Sum of Squares Criterion".

Three Basic Rules for Identifying a Good Mathematical Model (3 of 3)

3. Think.
   - Use common sense. Don’t use a model that leads to predicted values that are unrealistic. Use the model to calculate future values, past values, and values for missing data, and then determine whether the results are realistic and make sense. Don’t go too far beyond the scope of the available sample data.

Example: Finding the Best Population Model (1 of 5)

The table below lists the population of the United States for different 20-year intervals. Find a mathematical model for the population size, then predict the size of the U.S. population in the year 2040.

Year	1800	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000
Coded Year	1	2	3	4	5	6	7	8	9	10	11
Population	5	10	17	31	50	76	106	132	179	227	281

Example: Finding the Best Population Model (2 of 5)

Solution

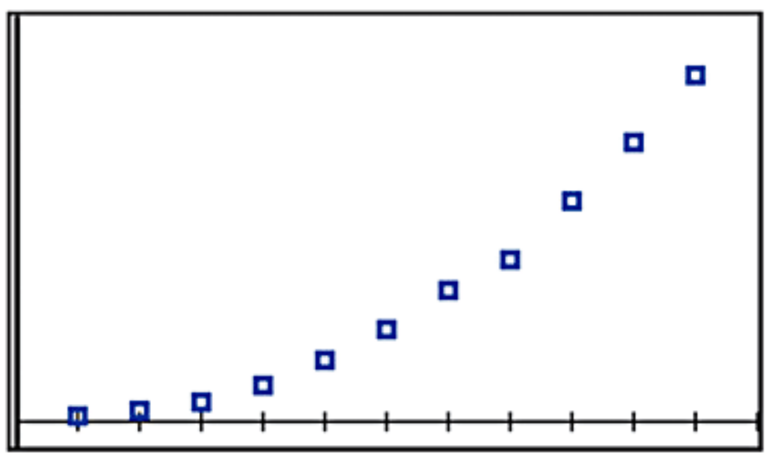
First, we “code” the year values by using 1, 2, 3,…, instead of 1800, 1820, 1840,…. The reason for this coding is to use values of *x* that are much smaller and much less likely to cause computational difficulties.

Example: Finding the Best Population Model (3 of 5)

Solution

1. **Look for a pattern in the graph.** Examine the pattern of the data values in theTI-83/84 Plus display. Good candidates for the model appear to be the quadratic, exponential, and power functions.

TI-83/84 Plus



Example: Finding the Best Population Model (4 of 5)

Solution

2. **Find and compare values of R^2.** For the quadratic model, $R^2 = 0.9992$. The table includes this result with results from two other potential models. If we select the quadratic function, we conclude that the equation $y = 2.77x^2 - 6.00x + 10.01$ is the best model.

Model	R^2	Equation
Quadratic	**0.9992**	$y = 2.77x^2 - 6.00x + 10.01$
Exponential	0.9631	$y = 5.24(1.48^x)$
Power	0.9764	$y = 3.35x^{1.77}$

Example: Finding the Best Population Model (5 of 5)

Solution

3. **Think.** The forecast result of 400 million in 2040 seems reasonable. (As of this writing, the latest figures from the U.S. Bureau of the Census use much more sophisticated methods to project that the U.S. population in 2040 will be 380 million.) However, there is considerable danger in making estimates for times that are beyond the scope of the available data.

Example: Interpretation R^2 (1 of 2)

In the previous example, we obtained the value of $R^2 = 0.9992$ for the quadratic model. Interpret that value as it relates to the predictor variable of year and the response variable of population size.

Example: Interpretation R^2 (2 of 2)

Solution

In the context of the year/population data from the first example, the value of $R^2 = 0.9992$ can be interpreted as follows: 99.92% of the variation in the population size can be explained by the quadratic regression equation that relates year and population size.

Statistical Methods

In "Modeling the U.S. Population" (**AMATYC Review,** Vol. 20, No. 2), Sheldon Gordon makes this important point that applies to all uses of statistical methods:

"The best choice (of a model) depends on the set of data being analyzed and requires an exercise in judgment, not just computation."