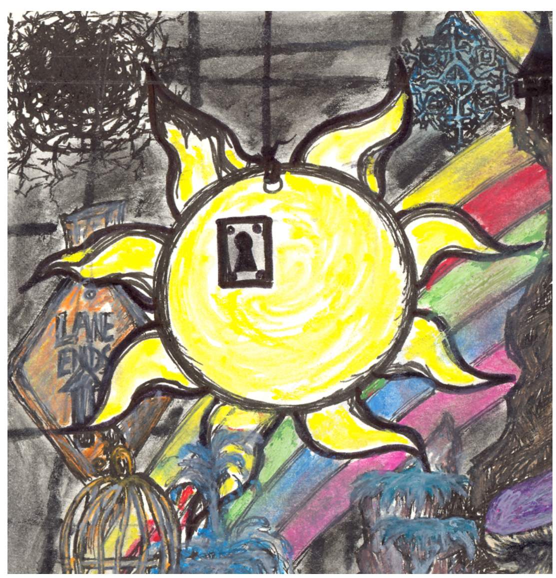
Chapter 10: Statistical Concepts

Introduction

Miranda Read, artist

As you gain experience in observing variable stars, your accuracy in estimating magnitudes will increase. Nonetheless, there will always be some scatter in your data. Doing *real* science and gathering *real* data always result in measurements that have inconsistencies. Science is a process of searching for answers that are as yet unknown. Therefore we cannot strive for "correctness." Scientists aim for precision—exactness in procedure and measurement—so that their results, whatever they may be, will be as accurate as possible. You will already have noticed in preceding activities that even when several individuals are measuring the same objects with the same measuring tool—be it string, a ruler, or the human eye—no one arrives at the exact same result. There is no way to avoid scattered data, no way to avoid the inconsistencies that come from random error. However, there are ways of eliminating the most extreme scatter so that your data are still accurate enough to be useful.

The world of science is one of continuous discovery. The excitement of discovery is in *not* knowing the correct answer before you start, and—most of the time, anyway—not knowing the *exact* answer after you finish. In most areas of science, making observations is but a tiny part of the discovery process. The bulk of the effort goes into extracting and analyzing meaningful information from observational data.



Miranda Read, artist

When dealing with quantitative data, *statistics* is the ideal mathematical tool to allow you to express the validity of your data, to view them from different perspectives, and evaluate their precision and quality. The following examples will help to explain fundamental statistical concepts, some of which you may already know.

Investigation 10.1: Finding the Average

1. Make the following two sets of measurements: the height and the arm length, in centimeters, of all the people in your classroom. For each person in the class, take three separate measurements of their height and their arm length and take the two averages. Did you all get the same arm length and height for each individual in the class? If you worked individually or in small groups, discuss with your classmates the procedures you used to take the measurements and calculate the averages. Discuss the differences in measuring techniques.

2. Obtain the average of the measurements taken for your height and arm length (in centimeters) from all of your classmates and enter them in Table 10.1. Compare the measurements. Are the measurements close together or far apart? How large is the scatter? What are some of the possible sources of random and/or systematic error which could have contributed to the differences in the results of the measurements?

3. Calculate the classroom average for your height by adding all the measurements in Table 10.1 together and dividing by the total number of measurements. Repeat the same calculation for arm length. Enter these two values, along with your name on line 1, in Table 10.2. Enter the calculated averages and names for the rest of the class. Add the measurements for heights and divide by the total number of measurements. You now have the average height for the entire class. Repeat the procedure to determine the average arm length for the class.

Table 10.1: Individual Averages

Name:		
Measurement from classmate:	Height (cm)	Arm Length (cm)
#1		
#2		
#3		
#4		
#5		
#6		
#7		
#8		
#9		
#10		
#11		
#12		
#13		
#14		
#15		
#16		
#17		
#18		
#19		
#20		
#21		
#22		
#23		
#24		
#25		
AVERAGES:		

Table 10.2: Classroom Averages

Names:	Height (cm)	Arm Length (cm)
#1		
#2		
#3		
#4		
#5		
#6		
#7		
#8		
#9		
#10		
#11		
#12		
#13		
#14		
#15		
#16		
#17		
#18		
#19		
#20		
#21		
#22		
#23		
#24		
#25		
AVERAGES:		

Core Activity 10.2: Constructing a Histogram

The *histogram* is one of the most important tools of elementary statistics. The histogram is a graph illustrating how likely it is to find any particular result in a set of data. In appearance, a histogram is similar to a bar graph. We begin by taking all possible results, and dividing them into ranges called *bins*. Then we count how many of the data points fall into each range. Finally, we divide each count by the total number of data points, to give us the *relative frequency*. Relative frequency is an estimate of the *probability* that any given data point will fall within this range. This method of graphically representing data is a powerful tool for analyzing a set of data (see Figure 10.1 below).

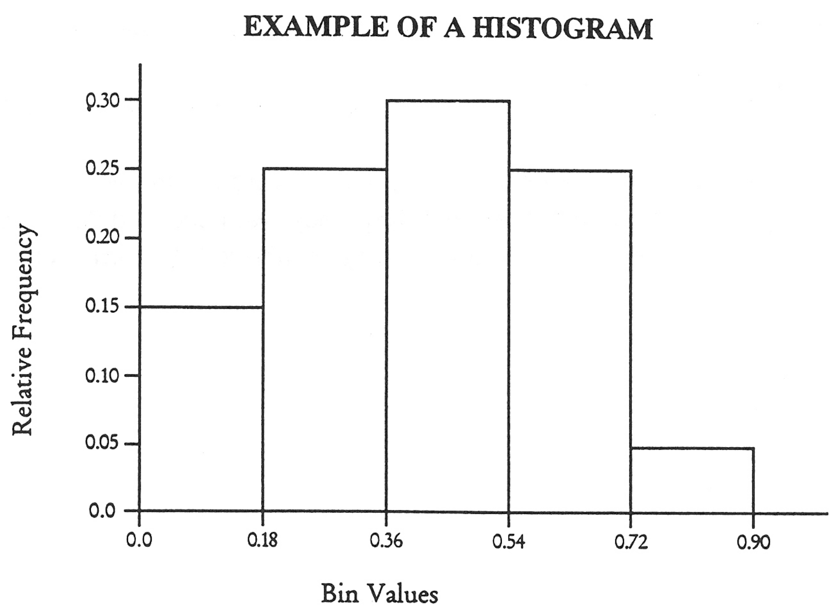


Figure 10.1

Constructing a histogram will allow you to visually examine the distribution and scatter of your data points. It is a quick way to assess the precision of a set of data and decide if the distribution of data is typical. You will be able to determine if the scatter of the data set is large or small, and how precise the measurements were. To construct a histogram, a data set has to be divided into equal groups, or bins. It will require some thought as to how a set of numbers should be divided into bins. To help make this decision, use the following rules:

a. Each bin has to be of equal value.

b. Each number falls into one and only one bin.

c. No number falls on the boundary or "in between" a bin.

d. At least 5 bins are necessary for a good representation of the data.

EXAMPLE:

Consider the following set of twenty numbers:

0.3, 0.5, 0.7, 0.6, 0.3, 0.5, 0.4, 0.1, 0.6, 0.1, 0.2, 0.8, 0.4, 0.7, 0.6, 0.3, 0.4, 0.5, 0.2, 0.5

The smallest number is 0.1 and the largest is 0.8. We need to arrange these numbers into at least five equal bins, and each of the numbers has to fall into a bin. A simple way to have equal bin values would be to center the first bin over the value of 0.1 and the last bin over the value 0.8. This ensures that all the numbers will fall into a bin and not on the boundaries between any of the bins. Therefore the value of each bin would range from .05 less than each number to .05 more than each number. If the first bin is centered on .1, the bin value would be .05 to .15. So we will have 8 bins which have values of:

[.05–.15], [.15–.25], [.25–.35], [.35–.45], [.45–.55], [.55–.65], [.65–.75], [.75–.85]

These bin values will be on the horizontal axis of our histogram. On the vertical axis is the relative frequency. To determine the relative frequency, we need to determine, by counting, how many of the data points in our set of numbers above falls into each bin.

Bin 1 [.05–.15] – 2

Bin 2 [.15–.25] – 2

Bin 3 [.25–.35] – 3

Bin 4 [.35–.45] – 3

Bin 5 [.45–.55] – 4

Bin 6 [.55–.65] – 3

Bin 7 [.65–.75] – 2

Bin 8 [.75–.85] – 1

total: 20

To determine the relative frequency with which each number occurs, the number of data points that falls into each bin is divided by the total number of data points; that is, the frequency of numbers which fall into Bin 1 is equal to 2 (0.1 occurs twice in the data set) divided by 20, or 0.1. For our 8 bin values above the relative frequencies are:

Bin 1: 2 divided by 20 = 0.10

Bin 2: 2 divided by 20 = 0.10

Bin 3: 3 divided by 20 = 0.15

Bin 4: 3 divided by 20 = 0.15

Bin 5: 4 divided by 20 = 0.20

Bin 6: 3 divided by 20 = 0.15

Bin 7: 2 divided by 20 = 0.10

Bin 8: 1 divided by 20 = 0.05

total: 1.00

Notice that the relative frequencies all add up to 1. This is because of the definition of relative frequency. It is the bin count, divided by the sum total of all the counts. So the sum of the relative frequencies is the sum of the counts, divided by the sum of the counts—which of course equals 1.

The relative frequencies and bin values can now be used to construct the histogram below.

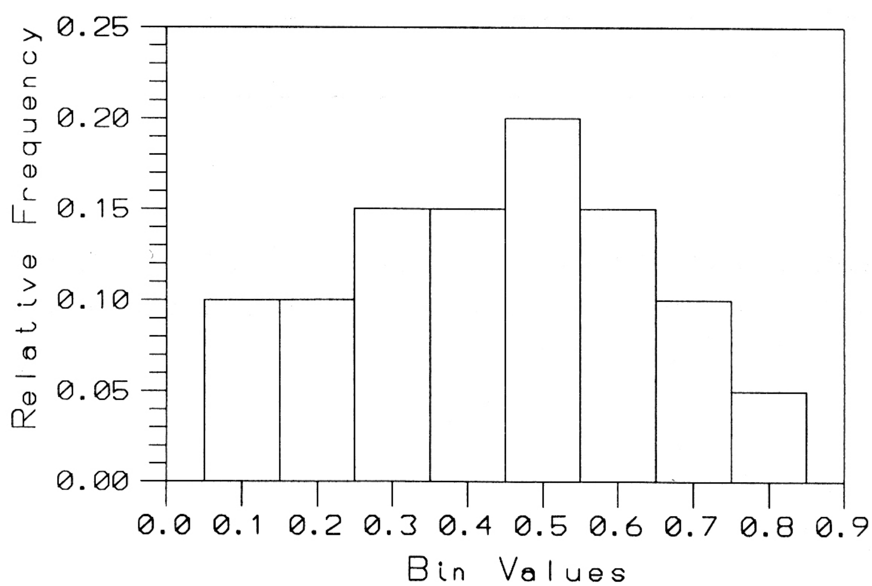


Figure 10.2

1. Using all of your height measurements from your classmates in Table 10.1, determine an appropriate number of bins, and the bin values. Enter the bin values in Table 10.3. Calculate the relative frequency of each measurement and enter it in Table 10.3. Use the information in Table 10.3 to construct a histogram.

Table 10.3

Bin Value	# of Data Points	Relative Frequency
TOTALS:		

2. On your histogram, mark the midpoint of the top of each bar (bin value) with a dot, and connect the dots with a smooth curve, as in the example histogram (Figure 10.3) on the following page.

EXAMPLE OF A HISTOGRAM

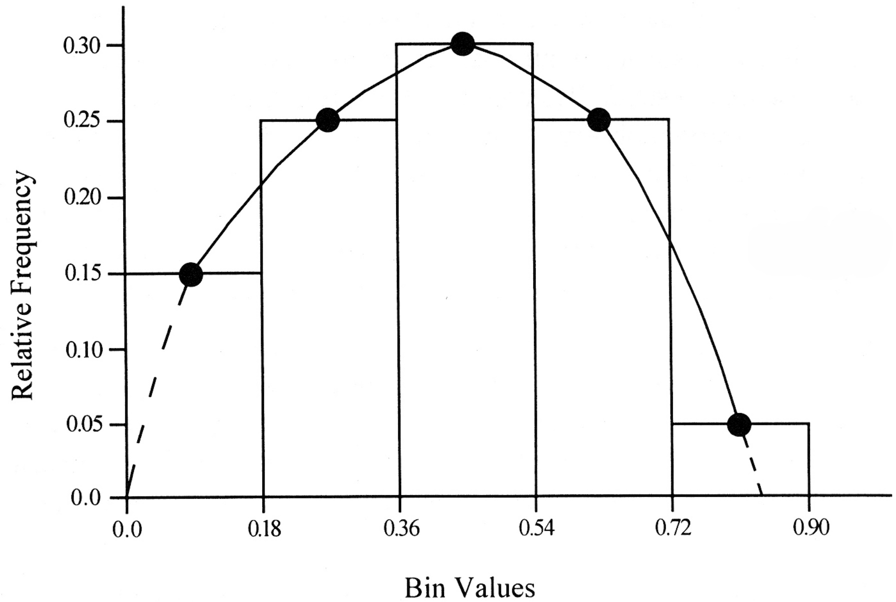


Figure 10.3

3. You now have a visual summary of your data. The resulting curve should be reasonably close to a symmetrical bell-shape. This is referred to as a normal curve, meaning that your data set follows a normal distribution and your measurements are reasonably precise. (There is an extended explanation of the normal curve on page 172 of this chapter.) Discuss possible factors which might have contributed to the amount of scatter.

Core Activity 10.3: Finding the Average Deviation

---

1. In Chapter 6 you compiled a set of magnitude estimates for Variable Star X in Table 6.6. You can now use this data set for statistical analysis. (If your class has collected a sufficient number of actual variable star observations, you may use those data instead.) Using either the information on Star X from Table 6.6 or your classroom observational data, transfer the data to columns [A] and [B] in Table 10.4. Then complete columns [C] and [D] using the data from Table 6.6. You are entering the JD, your magnitude estimation, the number of classroom estimates, and the class averages of the estimations for Star X. You will need to calculate the class average for each Julian Date.

2. The next column [E] in Table 10.4 is labeled "*Range.*" The range is simply the difference between the smallest and largest value in a set of data. Refer back to Table 6.6, which lists the estimates of the magnitude of Star X for the entire class. For each JD, look at the estimates in the row; find the smallest and largest values and take the difference between the two values. Enter the result in column [E].

3. You are now ready to calculate the class average deviation by using the Star X information in columns [B] and [D] from Table 10.4.

    a. For the first JD magnitude estimation, determine the difference of your observation from the class average and take the absolute value. This is your individual deviation. (You will use these numbers again in Core Activity 10.4.)

    b. Add together the individual deviations for the entire class for the first JD estimation, and divide the sum by the number of observations. This is the average deviation. Enter the result into column [F] in Table 10.4.

    c. Repeat for all the remaining magnitude estimates.

Determining the range gives you an idea of the amount of variability or scatter in the values within the data set. The range gives a general idea of the variation; there are other methods of analyzing the data that are more "*sensitive*" than the range—that is, they give a more detailed look at how much each data point deviates from the average. The most basic method of expressing the spread of the data in quantitative terms is called the *average deviation*.

Table 10.4

Name of Variable Star:								
Data point	[A] Julian Date	[B] Magnitude of Star (your estimate)	[C] # of class obser-vations	[D] Class Average	[E] Range	[F] Average Deviation	[G] Standard Deviation	[H] Standard Error of Average
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								

	Table 10.5 Two Hypothetical Data Sets	
	Sample 1	**Sample 2**
Measurements	1, 2, 3, 4, 5	2, 3, 3, 3, 4
Dot Diagram (showing frequency distribution)	* * * * * 1 2 3 4 5	* * * * * 1 2 3 4 5
Average	$\frac{1 + 2 + 3 + 4 + 5}{5} = \frac{15}{5} = 3$	$\frac{2 + 3 + 3 + 3 + 4}{5} = \frac{15}{5} = 3$
Distance of Measurement from Average or Deviation from Average	(1-3), (2-3), (3-3), (4-3), (5-3) or -2, -1, 0, 1, 2	(2-3), (3-3), (3-3), (3-3), (4-3) or -1, 0, 0, 0, 1

Study the two hypothetical data sets in Table 10.5 above. The first row shows the frequency distribution of each data point, the second shows the calculated average, and the third row shows the calculations for determining the distance each data point is from the average. This number is calculated by subtracting the average from each number. The distance from the average is called the *deviation* from the average.

To calculate the average deviation, drop the negative signs from the third row above. Use the absolute value of the numbers (they are all treated as though they are positive). Then the absolute values of each number are added together, and the sum is divided by the number of data points.

For Sample 1 above:

$$\frac{2 + 1 + 0 + 1 + 2}{5} = \frac{6}{5} = 1.2$$

For Sample 2 above:

$$\frac{1 + 0 + 0 + 0 + 1}{5} = \frac{2}{5} = 0.4$$

Now relate these results to the dot diagram showing frequency distribution in Table 10.5 on the previous page. Sample 1 has an average deviation of 1.2, showing that there is a larger spread in the data set than there is in Sample 2, which has a lower average deviation of 0.4. Sample 2, with a lower average deviation, must have less variation or scatter in the data: all of the data are close to the average. Conversely, Sample 1 has a larger range and the data are not as closely centered to the average.

In summary, to calculate the average deviation:

1. Calculate the average of the data set.
2. Subtract the average from each data point to get the difference.
3. Take the absolute value of each difference.
4. Add the absolute values together.
5. Divide the sum of the absolute values by the number of data points.

In mathematical terms, the process described above is represented by the following:

$$\frac{\Sigma \left| x_i - \bar{x} \right|}{n}$$

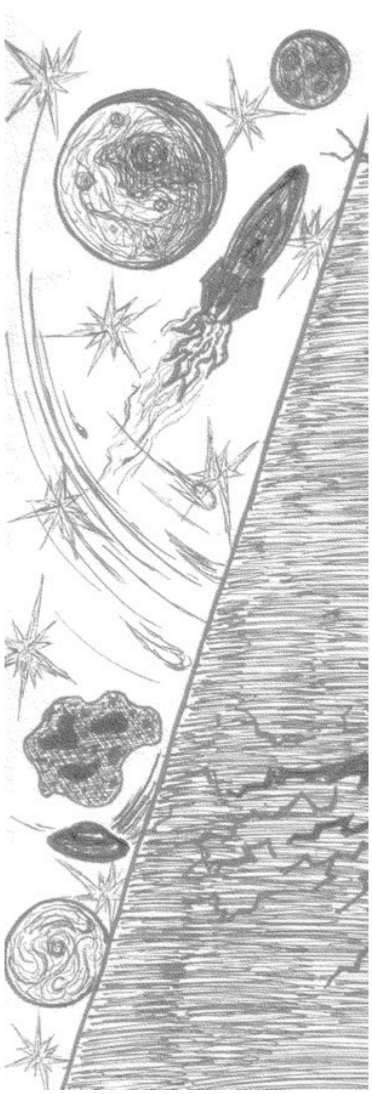
where	x_i	=	the value of each data point
	$\bar{x}$	=	the average of all the data points
	Σ	=	the Greek letter sigma, meaning "sum of"
	n	=	the total number of data points
	$\| \; \|$	=	the "absolute value of"

You will calculate the Standard Deviation [G] for Table 10.4 in Core Activity 10.4 and the Standard Error of the Average [H] in Core Activity 10.5.

Hands-On Universe

Hands-On Universe (HOU) is a program that enables high school students to request their own observations from professional observatories. HOU students download Charge Coupled Device (CCD) images to their classroom computers and use HOU's powerful image processing software to visualize and analyze their data. The HOU program collaborates with telescopes in Hawaii, Illinois, California, Washington, Sweden, and Australia to form a network of automated telescopes for educational use. Student requests are processed by the network to decide which telescope is best suited for the particular request, considering weather, geography, scheduling, and equipment. The network provides fast turn-around for student requests and allows real-time observing in certain cases because of the location of the telescope in various time zones. A key component of the HOU project is student research and investigation. Many students have used HOU to explore astronomical phenomena and have written web-based reports of their work.

Cerro Tololo Observatory in Chile



Miranda Read, artist

Students have a chance to work with scientists on original research projects, such as the HOU Asteroid Search. The asteroid search uses images from the Berkeley Cosmology Project, which is composed of a team of scientists searching for very distant supernovae. They use world-class telescopes such as the Cerro Tololo International Observatory (CTIO) in Chile to search for type Ia supernovae near the edge of the visible universe. The scientists share their data with HOU classes so that students can search for very faint asteroids in the same regions of the sky. To date, five previously unknown asteroids have been recorded by HOU students.

The HOU website is: http://hou.lbl.gov

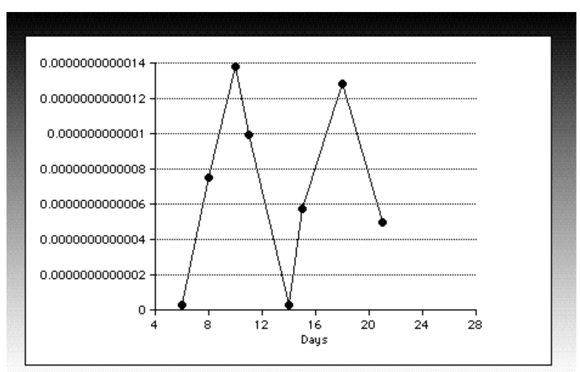
Hands-On Universe classes are involved with several research projects, including searching for supernovae and asteroids, creating H-R diagrams from images of open star clusters, and performing photometric measurements of Cepheid variable stars. One HOU project report is summarized below. *(used with permission from Hands-On Universe, Lawrence Berkeley Laboratory)*

Calculating Distance for Cepheid Variable Stars

by Adam A. Bier-high school student (http://hou.lbl.gov/studentreports/adamcv/cv.html)

	A	B	C	D	E	F
1	Image (Days)	Ref Bright	Img CV Bright	Norm Factor	Real CV Bright	
2	(Known Bright)	2.28e-12	n/a	n/a	n/a	
3	6	285847	3488	7.976295e-18	2.782131e-14	
4	8	5301	1746	4.301075e-16	7.509677e-13	
5	10	133451	80907	1.708492e-17	1.382289e-12	
6	11	111359	48670	2.047432e-17	9.964852e-13	
7	14	289890	3603	7.865052e-18	2.833778e-14	
8	15	289707	72637	7.870020e-18	5.716546e-13	
9	18	271794	152711	8.388706e-18	1.281047e-12	
10	21	285991	61895	7.972278e-18	4.934441e-13	
11						
12					Days	CV Bright
13					6	2.782131e-14
14	Average CV Bright -------->		6.915060e-13		8	7.509677e-13
15	[Calculated as SUM(F13...F20)/8]				10	1.382289e-12
16					11	9.964852e-13
17					14	2.833778e-14
18					15	5.716546e-13
19					18	1.281047e-12
20					21	4.934441e-13
21						

- Normalization was achieved by calculating Norm Factor based on ratio between known Reference Star brightness (in Joules/second/meter^2) and an image Reference Star brightness (in pixel counts).
- Period of Cepheid: 8 days (as shown in the graph below).
- Apparent brightness of Reference Star: 2.28×10^{-12} J/s/m^2 (used as B2 in the table above).
- Average apparent brightness of Cepheid: 6.1915×10^{-13} J/s/m^2 (calculated by taking the sum of the normalized brightness values and dividing it by 8, the number of values; shown as C4 in the table above). This gives E, the real or absolute Cepheid brightness.
- Luminosity of Cepheid: 1.71×10^{29} J/s (calculated by using the period-luminosity relationship to get luminosity in Solar Units, 3000, then converting that into J/s by multiplying it by 5.7×10^{25}). This gives P, the period.
- Distance to Cepheid in meters using the light/distance equation:
  $E = P / (4 \pi d^2)$
  $6.915 \times 10^{-13} = 1.71 \times 10^{29} / (4 \pi d^2)$
  $6.915 \times 10^{-13} \times (4 \pi d^2) = 1.71 \times 10^{29}$
  $8.690 \times 10^{-12} \times d^2 = 1.71 \times 10^{29}$
  $d^2 = 1.97 \times 10^{40}$
  $d = 1.40 \times 10^{20}$ meters
- Distance to Cepheid: 14,800 light-years (calculated by dividing the distance in meters by 9.46×10^{15}).



Core Activity 10.4: Variance and the Standard Deviation

Using *variance* and *standard deviation* is an even more meaningful method of measuring data variability than average deviation.

A. The determination of the variance differs from average deviation in the following manner. Instead of taking the absolute value to eliminate the negative signs, the deviations are squared. Then, instead of being divided by the number of data points, the sum of the squared values is divided by the number of data points minus one. Referring back to Table 10.5, the variance for Sample 1 is as follows:

$$\frac{(-2)^2,\ (-1)^2,\ (0)^2,\ (1)^2,\ (2)^2}{(5-1)} = \frac{4 + 1 + 0 + 1 + 4}{4} = 2.5$$

The variance for Sample 2:

$$\frac{(-1)^2,\ (0)^2,\ (0)^2,\ (0)^2,\ (1)^2}{(5-1)} = \frac{1 + 0 + 0 + 0 + 1}{4} = 0.5$$

B. The standard deviation is the positive square root of the variance. Therefore for sample 1 the standard deviation (**SD**) is:

SD $= \sqrt{2.5} = 1.6$

For sample 2:

SD $= \sqrt{0.5} = 0.71$

Comparison:

Sample 1 – average deviation = 1.2, **SD** = 1.6

Sample 2 – average deviation = 0.5, **SD** = 0.71

The standard deviation shows a larger deviation than the average deviation leads us to believe. While this may seem a strange and confusing way of measuring variability, try to understand this method in the following way. If you ignore the square root for a moment and just consider the variance, this expresses the average of the *squared* differences between the values and the average, rather than the average of the absolute differences.

Squaring the differences causes the expression to be dominated by the largest differences, while the comparatively small ones become insignificant. This has the net effect of emphasizing large deviations from the average, while de-emphasizing small ones. For example, 2 squared = 4, while 5 squared = 25, a much larger number. Squaring the larger number makes a larger impact. As a result, the expression inside the square root is considerably more "sensitive" than the average deviation.

In summary, to calculate the standard deviation:

1. Calculate the average of the data set.
2. Subtract the average from each data point to find the difference.
3. Eliminate the negative signs and square the differences.
4. Add the squared differences together.
5. Divide the sum by the number of data points minus one.
6. Take the square root of the result.

The numerical expression for this method is as follows:

$$\text{variance} = \frac{\Sigma\ (x_i - \bar{x})^2}{n - 1}$$

and standard deviation (SD) = the square root of the variance, therefore:

$$\text{SD} = \sqrt{\frac{\Sigma\ (x_i - \bar{x})^2}{n - 1}}$$

where
- $\mathbf{x_i}$ = the value of each data point
- $\bar{\mathbf{x}}$ = the average of all the data points
- Σ = the Greek letter sigma, meaning “sum of”
- $\mathbf{n}$ = the total number of data points

1. To calculate the standard deviation for the Star X data:
   a. Using the individual deviation numbers you calculated for the first JD magnitude estimation from (3a) in Core Activity 10.3, square the number.
   b. Add together the squared numbers for the first JD estimation for the entire class, and divide by the number of data points minus one. This is the variance.
   c. Take the square root of the variance. This is the standard deviation. Enter into column [G] in Table 10.4.
   d. Repeat for all the remaining magnitude numbers.

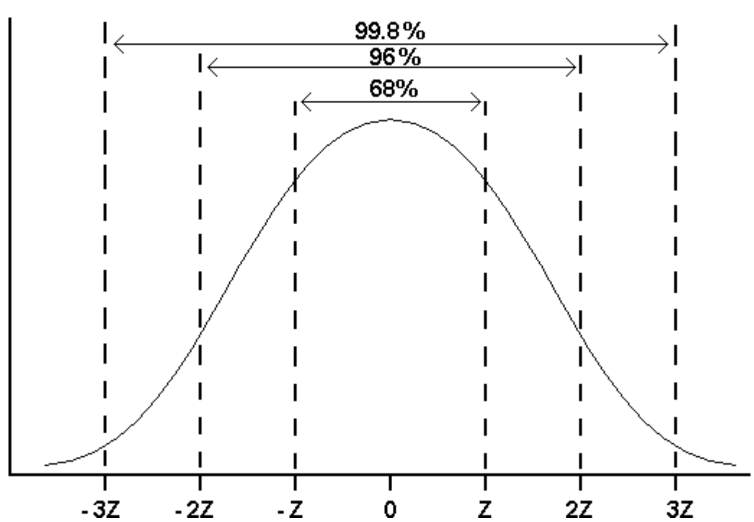
Remember, when observations are relatively precise (low variability), the standard deviation is very small. When the observations have poor precision (large scatter), then the standard deviation becomes much larger.

2. Compare your values in Table 10.4 for average and standard deviation. Compare your results with those of your classmates. Did the standard deviation change between the first and last point in Table 10.4? If so, what reasons can you suggest?

3. Find the range and the standard deviation again with only half of the class observations. How does the size of the sample affect the range and the standard deviation? Can you determine that standard deviation is a better indicator of variability than the range? How?

Standard Deviation and the Normal Curve

Standard deviation also has many other useful applications. Statisticians have created a model for random events called the *normal distribution.* This mathematically describes the likelihood of obtaining a certain value in an experiment, depending on how many standard deviations from the accepted average that value lies. If you connect the midpoint of the tops of each bar in a histogram, you will get a curve. A bell-shaped curve that closely matches the distribution of many large sets of numbers is called the *normal curve* or *bell curve*. For example, the odds of a coin-toss resulting in "heads" is 50–50, or half of 100 tosses. But if you toss a coin 100 times and keep track of the number of times you get "heads," you probably will not get "heads" exactly 50 times. But if you repeat the experiment 1000 times (100,000 coin tosses in sets of 100 each), and then draw a relative frequency histogram for the number of times you get "heads," a normal curve will result. The likelihood of a measurement being within a certain number (Z) of standard deviations from the average is assessed by finding the area under the bell curve between the points (–Z) and (Z). Statistical tables exist which give the area between (–Z) and (Z) for a range of possible Z's. The area is given in percent (%) and should be interpreted as the probability that a value will fall within Z standard deviations of the average.



In a bell-shaped histogram, we would expect about 68% of the data to lie within one standard deviation (the interval $\overline{x} \pm 1$ SD), and almost 100% within three standard deviations (the interval $\overline{x} \pm 3$ SD).

To understand what this means, consider the following set of data:

4.0, 3.9, 4.1, 4.0, 4.2, 3.9, 3.9, 4.1, 3.8, 4.0,

with an average = 4.0 and a standard deviation = 0.12.

If the measurements follow the normal distribution, then approximately:

a) 68% of the measurements fall between 4.0 ± 0.12, or between 3.88 and 4.12;

b) 96% of the measurements fall between 4.0 ± (2 x 0.12), or between 3.76 and 4.24;

c) 99.8% of the measurements fall between 4.0 ± (3 x 0.12), or between 3.64 and 4.36.

For any set of data to appear to be normal, the number of data points should be large—at least 30—and the larger the better. Then and only then an analysis of Z should be made to determine if the distribution is normal or not. The example we just considered is not a good representation of a normal distribution, even though it may give a normal curve, because the data points are fewer than 30. So let us assume that we had a large number of observations and came up with a normal curve. Now the question is, why is it so important to have a normal curve?

This concept is critical to assessing the validity of measurements, since it helps to detect errors. Almost 100% of the data will fall within three standard deviations of the average, so if we get a measurement of 4.4 in our sample data, we can assume that the measurement is probably false. However, we have to be very careful to determine whether there is any valid reason to discard this measurement. Not all unlikely measurements are incorrect. To determine the validity of the results, the *standard error of the average* is calculated.

Core Activity 10.5: The Standard Error of the Average—The Error Bar

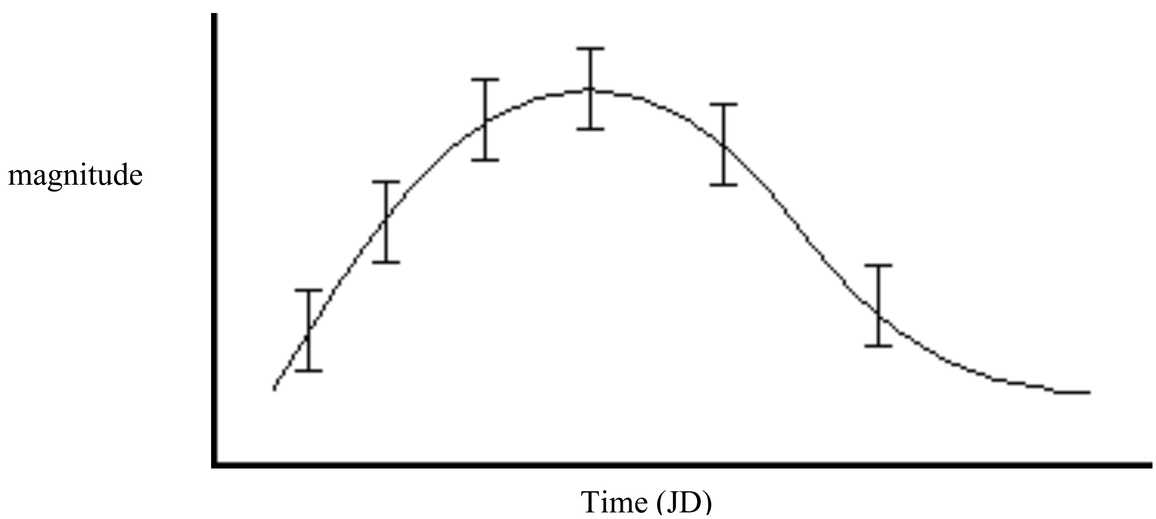
There is a mathematical calculation of the uncertainty of the average for a set of data. Since the average is calculated using a set of data that has error, the error of the average also needs to be calculated. The *standard error of the average* is the measure of how close to the exact value the average is likely to be. It is determined by dividing the standard deviation by the square root of the number of measurements. In mathematical terms:

$$\text{Standard deviation of the average} = \frac{\mathbf{SD}}{\sqrt{n}}$$

For the sample data set above (4.0, 3.9, 4.1, 4.0, 4.2, 3.9, 3.9, 4.1, 3.8, 4.0) **SD** = 0.12 and the number of observations **(n)** = 10.

Therefore $0.12/\sqrt{10} = \pm 0.038$

This value is used as the value of the error bars commonly seen on scientific graphs. To draw the *error bar* for this data point, you would draw a vertical line through the point on the graph with a 0.038 magnitude length above the point and a 0.038 magnitude length below the point, to produce the required .076 magnitude length for the entire bar (the error ranges from 0.038 to +0.038).



Exercise: Calculate the standard deviation of the average for the class data points for Star X and enter the result in column [H] of Table 10.4. You will use this information in the following chapter.

Mythological Evidence for Ancient Observations of Variable Stars

(*Adapted from a paper by Stephen R. Wilk, published in the journal of the AAVSO, Volume 24, 1996, pp. 129–133.)*

The known history of variable stars begins with David Fabricius' 1596 observations of omicron Ceti (Mira, "the Wonderful"). The eclipsing variable star Algol was first noted by Gemiani Montanari in 1667, but its period of 2.867 days was not measured until the 1783 work of Nathaniel Pigott, John Goodricke, and Johann Georg Palitzch.

However, there has long been suspicion that knowledge of variable stars extends much farther back in time. The variability of Algol or Mira is suggested in ancient Babylonian and Chinese texts. The names applied to Algol—Demon's Head," "Head of the Gorgon," "Lilith," "Satan," or "The Piled-up Corpses"—have a vaguely evil ring to them, which suggests ancient knowledge of peculiar properties. Another indication of ancient knowledge of Algol's variability is its rarely-cited Hindu name, Mayavati, meaning "The Changeful."

Perseus, Philippe La Hire, 1705

A case has been made for ancient Greek knowledge of variable stars on the basis of Greek mythology. Perseus, son of Zeus and Danae, was sent by the tyrant Polydectes to obtain the head of a Gorgon. Perseus first visited the Graeae ["Grē'-ay"], sisters of the Gorgons. The Graeae had the form of old women, and had only one eye that they shared in common. Perseus intercepted the eye as they passed it from hand to hand, and promised to return it if they gave him directions to the home of the Gorgons. They did so, but according to some accounts, Perseus threw the eye into Lake Tritonis in Africa, so he could safely escape.

Perseus went to the island of the Gorgons and found them asleep. Two of them, Stheno and Euryale, were immortal, but the third, Medusa, was not. Perseus struck off Medusa's head and stuck it in his bag. From the severed neck sprang the winged horse Pegasus and the warrior Chrysaor, Medusa's children by Poseidon. The noise roused the other Gorgons, but Perseus was able to escape. As he returned home, Perseus passed over Ethiopia and saw Princess Andromeda chained to a rock as a sacrifice to Cetus, the sea-monster. Perseus went to Andromeda's parents, King Cepheus and Queen Cassiopeia of Ethiopia, and offered to save Andromeda, provided she was given to him in marriage. They agreed. Perseus rescued Andromeda and slew Cetus, but Cepheus and Cassiopeia later plotted against Perseus, and he turned them to stone with the Gorgon's head.

The constellation of Perseus has been associated with the mythological character of that name since at least the fifth century BC. Later illustrations generally show Algol forming one of the Gorgon's eyes, but Roman and Arab authors call the star the head or face of the Gorgon. A more reasonable interpretation of the periodic fading of Algol is that it represents Perseus cutting off Medusa's head and placing it in his bag.

Algol B eclipses Algol A approximately every third day. This could explain why there are three Gorgon sisters, and why only Medusa is mortal. The two days during which Algol is not eclipsed represent the two immortal sisters, Stheno and Euryale. Medusa is the third day, during which the star is eclipsed, and the Gorgon "loses her head."

The eclipsing of Algol can be interpreted another way within the same myth. The three Graeae are virtual doubles of the Gorgons-they are both sets of three sisters, and they share the same parents. Maybe they, rather than the Gorgons, are the actual monsters from a parallel version of the myth, in which the task set to Perseus was to steal the eye of the Graeae. The fading of Algol in this case represents Perseus intercepting the eye (Algol) as it is passed from one sister to another.

There is an interesting corollary to this interpretation. The spectacular Perseid meteor shower every mid-August appears to originate from the arm of the constellation of Perseus. It is very easy to see in the display Perseus hurling the eye of the Graeae into Lake Tritonis.

It is also notable that the constellations representing characters in the myth of Perseus and Andromeda are grouped so close together in the sky (Figure 1). Most of these constellations harbor naked-eye variable stars. Three of them are so noticeable that they have given the names to their types. Algol, the preeminent example of eclipsing variables, has already been mentioned, as has Mira (omicron Ceti), the first historical variable star to be officially discovered.

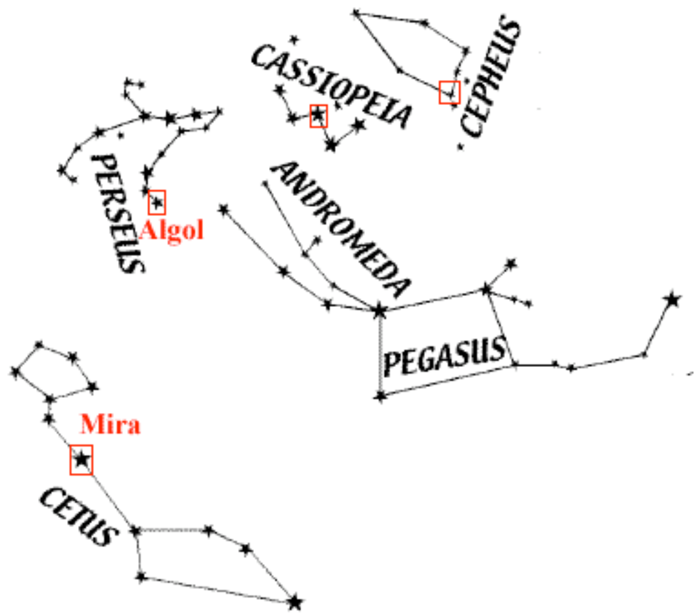


Figure 1.

Goodricke, co-discoverer of Algol, also discovered delta Cephei, the prototype for Cepheid variables. One must also note that gamma Cassiopeiae, the center star of the "W" of Cassiopeia, is an irregular variable star which varies between 1.6 and 3.1. Besides Algol in Perseus, that are naked-eye variable stars in the constellations of Cetus, Cepheus, and Cassiopeia. These are all constellations representing Perseus' enemies in the myths. In addition, Cetus is the mother of both the Gorgons and the Graeae.

Evidence of this sort can never be certain, but the set of coincidences strongly suggests that the ancient myth-makers and proto-astronomers knew of the variability of Algol, Mira, delta Cephei, and gamma Cassiopeiae, and on that basis associated their constellations together in a common myth.

There are other myths and interpretations associated with Perseus. One of the oldest and most peculiar images associated with Perseus, the birth of Chrysaor and Pegasus from the neck of Medusa, is first referred to in one of the most ancient Greek poems extant-Hesiod's *Theogeny*. The real meaning of this old myth is apparent from the constellations of Perseus (with Medusa's head) and Pegasus. If Hesiod's words mean that Pegasus and Chrysaor sprang from the stump of the neck that is attached to the head, rather than from the stump attached to the body, then the scene is pictured in that grouping of stars. The constellation of Perseus stands in for the person of Chrysaor, springing to the East. Pegasus, the winged horse, faces and springs to the West (Figure 2).

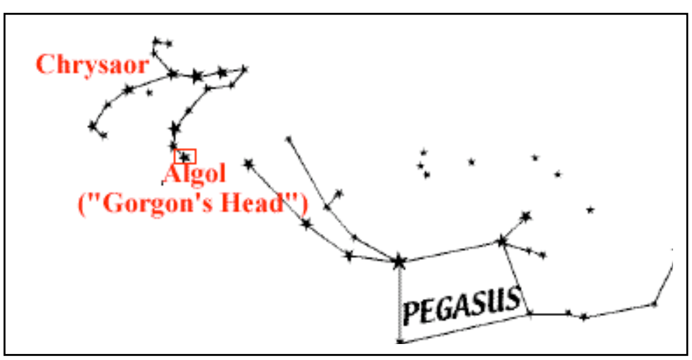


Figure 2.

The correlations between the variations of Algol and elements of the myth of Perseus and the Gorgon suggest ancient knowledge of that variability. The further association of surrounding constellations, which contain most of the naked-eye variable stars visible from Greece, with characters in the same myth, suggests that these variable stars were also known in preclassical Greece, whence the myths arose.

Activity 10.6: Statistical Analysis of Delta Cephei

If you have classroom observational data for delta Cep, use them to repeat the above processes of calculating the average and range, constructing a histogram, and calculating average deviation, variance and standard deviation, and the standard error of the average.

MATH TALK

Have you ever heard the expression, "Four out of five doctors recommend...?" Or "... 42% more relief from heartburn"? Or "...better highway mileage than any other sub-compact hatchback sedan costing under $10,000 made in America"?

Perhaps you suspected that these claims were not completely true. It is wise to be suspicious, because **statistics** (and numbers in general) can be manufactured to make any idea sound convincing. When used properly, statistics is a powerful tool for uncovering truth; when used improperly, it can be manipulated to prove almost anything.

Try, try again

There are lots of ways to misuse statistics. One way is perseverance: if at first you don't succeed (i.e., get the result you wanted), try, try again. Suppose you want to claim in a TV commercial that 4 out of 5 dentists recommend your toothpaste. You ask 5 dentists, but only 1 of them recommends your brand. So, forget you ever asked them! Ask another 5 dentists! This time, 2 of them recommend your brand. Forget them! Ask another 5! Keep trying until, by random fluctuation, you get lucky and 4 out of 5 recommend your brand. Then, show your TV commercial. Whatever you do, *do not* talk about the 13,925 dentists you had to survey before you got lucky, and don't mention that only 8% of *them* recommended your brand.

Sometimes this sort of thing happens even to honest people. If the results do not match our theory, it is too easy think of a "good" reason to believe that the data we do not like are not valid, so we have to do the experiment again. This happens far too often in scientific research, even today. Despite people's best intentions to be fair, there is just too much temptation to rationalize away the "bad" data. However, you rarely see any scientists rationalize away the "good" data, the data which support their theories!

Here we have the first lesson of honest statistics: you cannot ignore the data that do not fit your theory. Sometimes you have good reason to believe some piece of data should be excluded because it is just a mistake. But in your scientific report, you have to say so, and state exactly why it has been omitted. You can exclude data if you have good reason, but you cannot ignore them, or fail to report them.

How many?

A nursing home recently tried new procedures designed to reduce the number of accidental injuries to patients. They were pleased to announce that in the first four months of the year, patient accidents were down a whopping 60% compared to last year. Can't argue with that!

Or can you? How many are we talking about here? If last year there were 50 accidents, and this year only 20, then they are down 60%, and there is no doubt that this result is statistically significant. The chance of that happening by random fluctuation ("by accident") is less than 1 in 10,000.

But suppose there were 5 accidents last year, and only 2 this year. Yes, they are down 60%. But no, this result is *not* significant. The chances are better than 1 in 4 that this could happen by random fluctuation.

We have already seen that as we acquire more data, our results become more precise. They also become more *reliable*. Sometimes, an early result is based on so little data that it has no real significance. Do not put too much faith in statistical results (not even a whopping 60%) until you know how much data went into them.

Survey says!

Suppose two politicians are debating a school funding bill. They both try to show that the public is on their side by conducting a survey. Politician **A** wants to show that people favor the bill, so his survey asks, "Should we invest more in our children's future by passing the school funding bill?" Lo and behold, people *do* want to invest in their children's future, so most people say yes, and politician **A** announces that the vast majority favor his bill.

Politician **B** wants the bill to fail, so his survey asks, "Should we raise taxes to fund more and bigger government bureaucracy by passing the school funding bill?" Not surprisingly, people do not want higher taxes and more bureaucracy, so they mostly say no, and politician **B** claims that the vast majority oppose the bill.

This may seem like an exaggerated example, but it is not. This actually happens! Almost every political survey is deliberately designed to get a specific response. The questions are usually phrased to make the desired response sound good, while making the undesired response sound very bad. By doing so, the questions bias the subject's opinion about the topic of the survey. Not surprisingly, whoever paid for the survey usually gets the response they want. Politicians are not the only ones who do this. Advertising surveys are carefully designed to make the company product look good while making the competition look bad.

Even if you are trying very hard to be fair, it is actually quite difficult to phrase the question in a way that does not influence anyone's response. There are other ways surveys can go wrong, too; designing an accurate survey is a very difficult task, requiring much expertise. There are some organizations that do it well; for example, the Gallup organization specializes in conducting fair, scientifically reliable surveys. Still, it is an unfortunate fact that *most surveys just cannot be trusted* (especially political and advertising surveys).

What are you trying to prove?

It happens regularly that a government agency or private commission launches a major study of an important social issue. Too often they begin by announcing that they are going to prove some theory, which has important consequences for social policy. You can bet big money that they *will* find proof. After all, they have already made up their minds!

Any study which begins by assuming the correct answer, then looks for proof, will fail to give serious consideration to the possibility that the assumed "correct answer" is *not* correct. Any scientist who has already decided before the experiment that one result is "right" and another is "wrong" is no scientist at all.

It is very hard to avoid all bias when taking data. That is why we work very hard to make our experiments ***double blind***: we arrange that neither the scientists taking data, nor their subjects, know how the data will affect the outcome. For example, suppose we want to study the effectiveness of a new headache pill. We give half our subjects the new medication, while the other half get an inert sugar pill. We have to be sure that the subjects *do not know* which one they are getting. We also have to be sure that the scientists taking the data also do not know (at least until all the data are in). Otherwise, there is far too much temptation to "nudge" the data the way we want them to go.

Accidents happen

We have said that the standard of "unlikeliness" in statistics is 0.05, or 5%, or a 5% ***false-alarm probability***. This means that if we do a scientific experiment, and get a result that's only 5% likely to happen by accident, we have evidence that it is not an accident. We can write our results in a scientific paper, and every statistician will agree that our evidence is significant.

So we have evidence, but we do not yet have *proof*. After all, there *is* a 5% chance that it *did* happen by accident. Accidents do happen! In fact, an accident that is only 5% likely will happen about 5% of the time. After all, with a 5% false-alarm probability, we will get some false alarms.

Suppose a university employs 100 scientists, and each one does a different scientific experiment. From probability theory, we *expect* 5% of them to get a result that's only 5% likely, *by accident*! So *just by accident*, about 5 of the 100 scientists will get evidence that they can call "statistically significant" and publish in a scientific paper.

And they *do* have evidence, strong enough that their claim deserves further study. But they do not have proof. That is one of the reasons scientific experiments have to be *repeated*. If you get a "significant result" once, you have evidence. If two people get the same result, there is very strong evidence. If a dozen people do the same experiment, and they all get a significant result, then we can start to believe it.

Every year, scientists do hundreds of thousands of experiments. If they use a 5% false-alarm probability (and most of them do), we can *expect* 5% of the results to be false alarms. Five percent of 100,000 experiments is 5,000 false alarms! That means 5,000 results that seem to be significant, but really happened only by accident. Some of them will be published in important scientific journals. And they should be published: they are all *possibilities*, and deserve further study. But for most of them, we should not be convinced until the results are repeated.

Conclusion

We have seen that if you want to deceive people, statistics makes it easy. In fact, even if you want to be honest, there are so many things that can go wrong in an experiment or a survey, that we must carefully guard against bias. Even if we succeed, and get an unbiased result which is "statistically significant," it still might have happened just by accident. So the experiment has to be repeated, many times, and each time requires the same care in guarding against any bias which could affect the results.

That is a lot of work! Still, the payoff makes it well worth it. Not doing so gives us half-baked theories which sound good but really are not, supported by biased data and invalid statistics. This is worse than ignorance! But if we invest the effort to do science well, we reap the reward of knowledge that we can trust, and often can put to very good use.