In This Section

- → Example 1, Level A Revisited: Choosing a Band for the School Dance
- → Connecting Two Categorical Variables
- → Questionnaires and Their Difficulties
- → Measure of Location—The Mean as a Balance Point
- → A Measure of Spread—The Mean Absolute Deviation
- → Representing Data Distributions—The Frequency Table and Histogram
- → Comparing Distributions—The Boxplot
- → Measuring the Strength of Association between Two Quantitative Variables
- → Modeling Linear Association
- → The Importance of Random Selection
- → Comparative Experiments
- → Time Series
- → Misuses of Statistics
- → Summary of Level B

Level B

Instruction at Level B should build on the statistical base developed at Level A and set the stage for statistics at Level C. Instructional activities at Level B should continue to emphasize the four main components in the investigative process and have the spirit of genuine statistical practice. Students who complete Level B should see statistical reasoning as a process for solving problems through data and quantitative reasoning.

At Level B, students become more aware of the statistical question distinction (a question with an answer based on data that vary versus a question with a deterministic answer). They also should make decisions about what variables to measure and how to measure them in order to address the question posed.

Students should use and expand the graphical, tabular, and numerical summaries introduced at Level A to investigate more sophisticated problems. Also, when selecting a sample, students should develop a basic understanding of the role probability plays in random selection—and in random assignment when conducting an experiment.

At Level B, students investigate problems with more emphasis placed on possible associations among two or more variables and understand how a more sophisticated collection of graphical, tabular, and numerical summaries is used to address these questions. Finally, students recognize ways in which statistics is used or misused in their world.

Specifically, Level B recommendations in the *Investigative Process* include:

I. Formulate Questions

→ Students begin to pose their own questions.

→ Students address questions involving a group larger than their classroom and begin to recognize the distinction among a population, a census, and a sample.

II. Collect Data

→ Students conduct censuses of two or more classrooms.

→ Students design and conduct nonrandom sample surveys and begin to use random selection.

→ Students design and conduct comparative experiments and begin to use random assignment.

III. Analyze Data

→ Students expand their understanding of a data distribution.

→ Students quantify variability within a group.

→ Students compare two or more distributions using graphical displays and numerical summaries.

→ Students use more sophisticated tools for summarizing and comparing distributions, including:
  - Histograms
  - The IQR (Interquartile Range) and MAD (Mean Absolute Deviation)
  - Five-Number Summaries and boxplots

→ Students acknowledge sampling error.

→ Students quantify the strength of association between two variables, develop simple models for association between two numerical variables, and use expanded tools for exploring association, including:
  - Contingency tables for two categorical variables
  - Time series plots
  - The QCR (Quadrant Count Ratio) as a measure of strength of association
  - Simple lines for modeling association between two numerical variables

IV. Interpret Results

→ Students describe differences between two or more groups with respect to center, spread, and shape.

→ Students acknowledge that a sample may not be representative of a larger population.

→ Students understand basic interpretations of measures of association.

→ Students begin to distinguish between an observational study and a designed experiment.

→ Students begin to distinguish between "association" and "cause and effect."

→ Students recognize sampling variability in summary statistics, such as the sample mean and the sample proportion.

Example 1, Level A Revisited: Choosing a Band for the School Dance

Many of the graphical, tabular, and numerical summaries introduced at Level A can be enhanced and used to investigate more sophisticated problems at Level B. Let's revisit the problem of planning for the school dance introduced in Level A, in which, by conducting a census of the class, a Level A class investigated the question:

What type of music is most popular among students?

Recall that the class was considered to be the entire population, and data were collected on every member of the population. A similar investigation at Level B would include recognition that one class may not be representative of the opinions of all students at the school. Level B students might want to compare the opinions of their class with the opinions of other classes from their school. A Level B class might investigate the questions:

What type of music is most popular among students at our school?

How do the favorite types of music differ between classes?

As class sizes may be different, results should be summarized with relative frequencies or percents in order to make comparisons. Percentages are useful in that they allow us to think of having comparable results for groups of size 100. Level B students will see more emphasis on proportional reasoning throughout the mathematics curriculum, and they should be comfortable summarizing and interpreting data in terms of percents or fractions.

Table 3: Frequencies and Relative Frequencies

Favorite	Frequency	Relative Frequency Percentage
Class 1		
Country	8	33%
Rap	12	50%
Rock	4	17%
Total	**24**	**100%**
Class 2		
Favorite	Frequency	Relative Frequency Percentage
Country	5	17%
Rap	11	37%
Rock	14	47%
Total	**30**	**101%**

The results from two classes are summarized in Table 3 using both frequency and relative frequency (percents).

The bar graph below compares the percent of each favorite music category for the two classes.

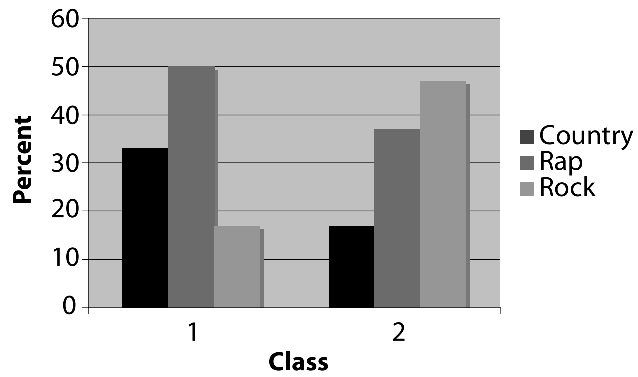

Figure 8: Comparative bar graph for music preferences

Students at Level B should begin to recognize that there is not only variability from one individual to another within a group, but also in results from one group to another. This second type of variability is illustrated by the fact that the most popular music is rap music in Class 1, while it is rock music in Class 2. That is, the mode for Class 1 is rap music, while the mode for Class 2 is rock music.

The results from the two samples might be combined in order to have a larger sample of the entire school. The combined results indicate rap music is the favorite type of music for 43% of the students,

“With the use of a scatterplot, Level A students can visually look for trends and patterns.”

rock music is preferred by 33%, while only 24% of the students selected country music as their favorite. Level B students should recognize that although this is a larger sample, it still may not be representative of the entire population (all students at their school). In statistics, randomness and probability are incorporated into the sample selection procedure in order to provide a method that is “fair” and to improve the chances of selecting a representative sample. For example, if the class decides to select what is called a simple random sample of 54 students, then each possible sample of 54 students has the same probability of being selected. This application illustrates one of the roles of probability in statistics. Although Level B students may not actually employ a random selection procedure when collecting data, issues related to obtaining representative samples should be discussed at this level.

Connecting Two Categorical Variables

As rap was the most popular music for the two combined classes, the students might argue for a rap group for the dance. However, more than half of those surveyed preferred either rock or country music. Will these students be unhappy if a rap band is chosen? Not necessarily, as many students who like rock music also may like rap music. To investigate this problem, students might explore two additional questions:

Do students who like rock music tend to like or dislike rap music?

Do students who like country music tend to like or dislike rap music?

To address these questions, the survey should ask students not only their favorite type of music, but also whether they like rap, rock, and country music.

The *two-way frequency table* (or *contingency table*) below provides a way to investigate possible connections between two categorical variables.

Table 4: Two-Way Frequency Table

		Like Rap Music?		
		Yes	No	Row Totals
Like Rock Music?	Yes	27	6	33
	No	4	17	21
Column Totals		31	23	54

According to these results, of the 33 students who liked rock music, 27 also liked rap music. That is, 82% (27/33) of the students who like rock music also like rap music. This indicates that students who like rock music tend to like rap music as well. Once again, notice the use of proportional reasoning in interpreting these results. A similar analysis could be performed to determine if students who like country tend to like or dislike rap music. A more detailed discussion of this example and a measure of association between two categorical variables is given in the Appendix for Level B.

Questionnaires and Their Difficulties

At Level B, students should begin to learn about surveys and the many pitfalls to avoid when designing and conducting them. One issue involves the wording of questions. Questions must be unambiguous and easy to understand. For example, the question:

Are you against the school implementing a no-door policy on bathroom stalls?

is worded in a confusing way. An alternative way to pose this question is:

The school is considering implementing a no-door policy on bathroom stalls. What is your opinion regarding this policy?

Strongly Oppose   Oppose   No Opinion   Support   Strongly Support

Questions should avoid leading the respondent to an answer. For example, the question:

Since our football team hasn't had a winning season in 20 years and is costing the school money, rather than generating funds, do you feel we should concentrate more on another sport, such as soccer or basketball?

is worded in a way that is biased against the football team.

The responses to questions with coded responses should include all possible answers, and the answers should not overlap. For example, for the question:

How much time do you spend studying at home on a typical night?

the responses:

none   1 hour or less   1 hour or more

would confuse a student who spends one hour a night studying.

There are many other considerations about question formulation and conducting sample surveys that can be introduced at Level B. Two such issues are how the interviewer asks the questions and how accurately the responses are recorded. It is important for students to realize that the conclusions from their study depend on the accuracy of their data.

Measure of Location—The Mean as a Balance Point

Another idea developed at Level A that can be expanded at Level B is the mean as a numerical summary of center for a collection of numerical data. At Level A, the mean is interpreted as the "fair share" value for data. That is, the mean is the value you would get if all the data from subjects are combined and then evenly redistributed so each subject's value is the same. Another interpretation of the mean is that it is the balance point of the corresponding data distribution. Here is an outline of an activity that illustrates the notion of the mean as a balance point. Nine students were asked:

How many pets do you have?

The resulting data were 1, 3, 4, 4, 4, 5, 7, 8, 9. These data are summarized in the dotplot shown in Figure 9. Note that in the actual activity, stick-on notes were used as "dots" instead of Xs.

"At Level A, the mean is interpreted as the 'fair share' value for data."

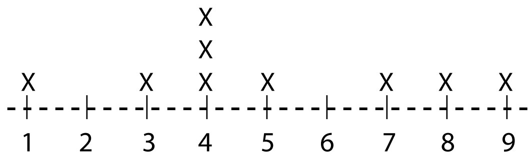

Figure 9: Dotplot for pet count

If the pets are combined into one group, there are a total of 45 pets. If the pets are evenly redistributed among the nine students, then each student would get five pets. That is, the mean number of pets is five. The dotplot representing the result that all nine students have exactly five pets is shown below:

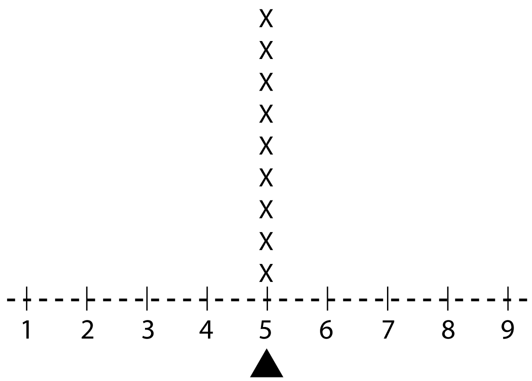

Figure 10: Dotplot showing pets evenly distributed

It is hopefully obvious that if a pivot is placed at the value 5, then the horizontal axis will "balance" at this pivot point. That is, the "balance point" for the horizontal axis for this dotplot is 5. What is the balance point for the dotplot displaying the original data?

We begin by noting what happens if one of the dots over 5 is removed and placed over the value 7, as shown below:

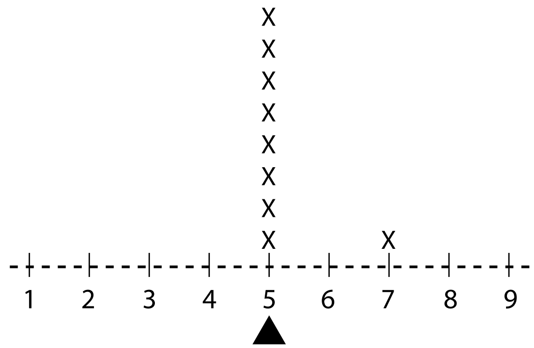

Figure 11: Dotplot with one data point moved

Clearly, if the pivot remains at 5, the horizontal axis will tilt to the right. What can be done to the remaining dots over 5 to "rebalance" the horizontal axis at the pivot point? Since 7 is two units *above* 5, one solution is to move a dot two units *below* 5 to 3, as shown below:

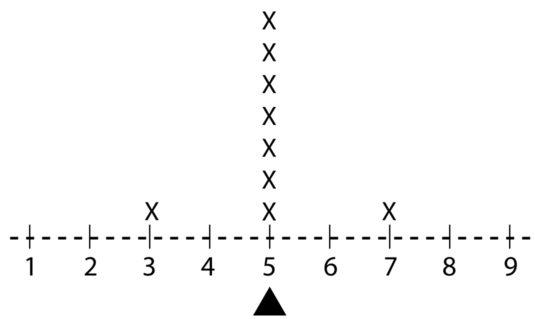

Figure 12: Dotplot with two data points moved

The horizontal axis is now rebalanced at the pivot point. Is this the only way to rebalance the axis at 5? No. Another way to rebalance the axis at the pivot point would be to move two dots from 5 to 4, as shown below:

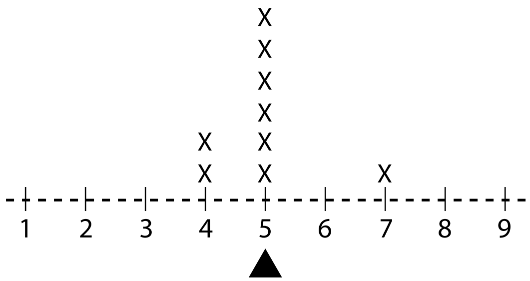

Figure 13: Dotplot with different data points moved

The horizontal axis is now rebalanced at the pivot point. That is, the "balance point" for the horizontal axis for this dotplot is 5. Replacing each "X" (dot) in this plot with the distance between the value and 5, we have:

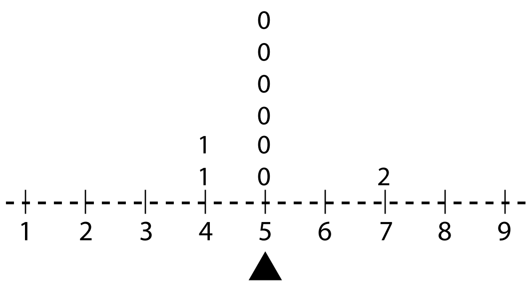

Figure 14: Dotplot showing distance from 5

Notice that the total distance for the two values below the 5 (the two 4s) is the same as the total distance for the one value above the 5 (the 7). For this reason, the balance point of the horizontal axis is 5. Replacing each value in the dotplot of the original data by its distance from 5 yields the following plot:

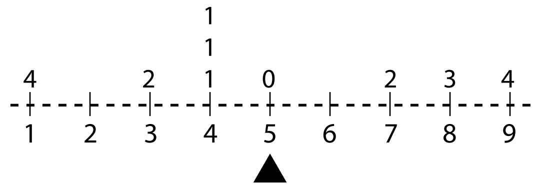

Figure 15: Dotplot showing original data and distance from 5

The total distance for the values below 5 is 9, the same as the total distance for the values above 5. For this reason, the mean (5) is the balance point of the horizontal axis.

Both the mean and median often are referred to as *measures of central location*. At Level A, the median also was introduced as the quantity that has the same number of data values on each side of it in the ordered data. This "sameness of each side" is the reason the median is a measure of central location. The previous activity demonstrates that the total distance for the values below the mean is the same as the total distance for the values above the mean, and illustrates why the mean also is considered to be a measure of central location.

A Measure of Spread—The Mean Absolute Deviation

Statistics is concerned with variability in data. One important idea is to quantify how much variability exists in a collection of numerical data. Quantities that measure the degree of variability in data are called *measures of spread*. At Level A, students are introduced to the *range* as a measure of spread in numerical data. At Level B, students should be introduced to the idea of comparing data values to a central value, such as the mean or median, and quantifying how different the data are from this central value.

In the number of pets example, how different are the original data values from the mean? One way to measure the degree of variability from the mean is to determine the total distance of all values from the mean. Using the final dotplot from the previous example, the total distance the nine data values are from the mean of 5 pets is 18 pets. The magnitude of this quantity depends on several factors, including the number of measurements. To adjust for the number of measurements, the total distance from the mean is divided by the number of measurements. The resulting quantity is called the *Mean Absolute Deviation,* or MAD. The MAD is the average distance of each data value from the mean. That is:

$$\text{MAD} = \frac{\text{Total Distance from the Mean for all Values}}{\text{Number of Data Values}}$$

The MAD for the data on number of pets from the previous activity is:

$$\text{MAD} = 18/9 = 2$$

The MAD indicates that the actual number of pets for the nine students differs from the mean of five pets by two pets, on average. Kader (1999) gives a thorough discussion of this activity and the MAD.

The MAD is an indicator of spread based on all the data and provides a measure of average variation in the data from the mean. The MAD also serves as a precursor to the standard deviation, which will be developed at Level C.

Representing Data Distributions—The Frequency Table and Histogram

At Level B, students should develop additional tabular and graphical devices for representing data distributions of numerical variables. Several of these build upon representations developed at Level A. For example, students at Level B might explore the problem of placing an order for hats. To prepare an order, one needs to know which hat sizes are most common and which occur least often. To obtain information about hat sizes, it is necessary to measure head circumferences. European hat sizes are based on the metric system. For example, a European hat size of 55 is designed to fit a person with a head circumference of between 550 mm and 559 mm. In planning an order for adults, students might collect preliminary data on the head circumferences of their parents, guardians, or other adults. Such data would be the result of a nonrandom sample. The data summarized in the following stemplot (also known as stem and leaf plot) are head circumferences measured in millimeters for a sample of 55 adults.

51 | 3

52 | 5

53 | 133455

54 | 2334699

55 | 12222345

56 | 0133355588

57 | 113477

58 | 02334458

59 | 1558

60 | 13

61 | 28

51 | 3 means 513 mm

Figure 16: Stemplot of head circumference

Based on the stemplot, some head sizes do appear to be more common than others. Head circumferences in the 560s are most common. Head circumferences fall off in a somewhat symmetric manner on both sides of the 560s, with very few smaller than 530 mm or larger than 600 mm.

In practice, a decision of how many hats to order would be based on a much larger sample, possibly hundreds or even thousands of adults. If a larger sample was available, a stemplot would not be a practical device for summarizing the data distribution. An alternative to the stemplot is to form a distribution based on dividing the data into groups or intervals. This method can be illustrated through a smaller data set, such as the 55 head circumferences, but is applicable for larger data sets as well. The *grouped frequency* and *grouped relative frequency* distributions and the *relative frequency histogram* that correspond to the above stemplot are:

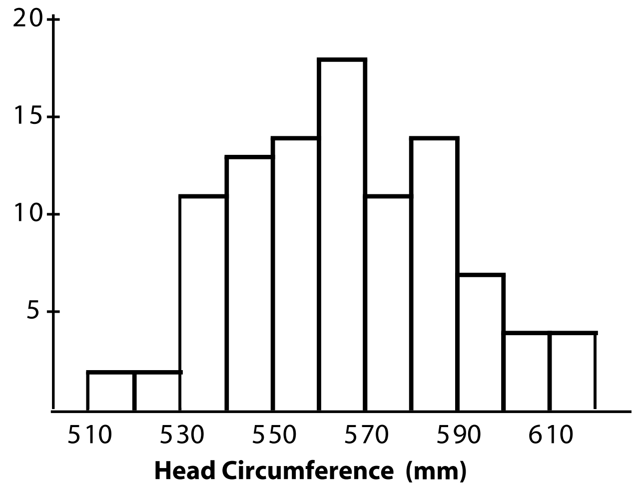

Figure 17: Relative frequency histogram

Table 5: Grouped Frequency and Grouped Relative Frequency Distributions

Stem	Limits on Recorded Measurements on Head Circumference	Interval of Actual Head Circumferences	Frequency	Relative Frequency (%)
51	510–519	510–<520	1	1.8
52	520–529	520–<530	1	1.8
53	530–539	530–<540	6	10.9
54	540–549	540–<550	7	12.7
55	550–559	550–<560	8	14.5
56	560–569	560–<570	10	18.2
57	570–579	570–<580	6	10.9
58	580–589	580–<590	8	14.5
59	590–599	590–<600	4	7.3
60	600–609	600–<610	2	3.6
61	610–619	610–<620	2	3.6
		Total	**55**	**99.8**

If the hat manufacturer requires that orders be in multiples of 250 hats, then based on the above results, how many hats of each size should be ordered? Using the relative frequency distribution, the number of hats of each size for an order of 250 hats is shown in Table 6.

Once again, notice how students at Level B would utilize proportional reasoning to determine the number of each size to order. Kader and Perry (1994) give a detailed description of "The Hat Shop" problem.

Comparing Distributions—The Boxplot

Problems that require comparing distributions for two or more groups are common in statistics. For example, at Level A students compared the amount of sodium in beef and poultry hot dogs by examining parallel dotplots. At Level B, more sophisticated representations should be developed for comparing distributions. One of the most useful graphical devices for comparing distributions of numerical data is the *boxplot*. The boxplot (also called a box-and-whiskers

Table 6: Hat Size Data

Hat Size	Number to Order
51	5
52	5
53	27
54	32
55	36
56	46
57	27
58	36
59	18
60	9
61	9

plot) is a graph based on a division of the ordered data into four groups, with the same number of data values in each group (approximately one-fourth). The four groups are determined from the *Five-Number Summary* (the minimum data value, the first quartile, the median, the third quartile, and the maximum data value). The Five-Number Summaries and comparative boxplots for the data on sodium content for beef (labeled B) and poultry (labeled P) hot dogs introduced in Level A are given in Table 7 and Figure 18.

Interpreting results based on such an analysis requires comparisons based on global characteristics of each distribution (center, spread, and shape). For example, the median sodium content for poultry hot dogs is

Table 7: Five-Number Summaries for Sodium Content

	Beef Hot Dogs (n = 20)	Poultry Hot Dogs (n = 17)
Minimum	253	357
First Quartile	320.5	379
Median	380.5	430
Third Quartile	478	535
Maximum	645	588

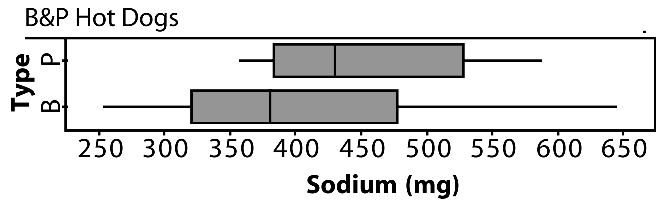

Figure 18: Boxplot for sodium content

430 mg, almost 50 mg more than the median sodium content for beef hot dogs. The medians indicate that a typical value for the sodium content of poultry hot dogs is greater than a typical value for beef hot dogs. The range for the beef hot dogs is 392 mg, versus 231 mg for the poultry hot dogs. The ranges indicate that, overall, there is more spread (variation) in the sodium content of beef hot dogs than poultry hot dogs. Another measure of spread that should be introduced at Level B is the *interquartile range,* or IQR. The IQR is the difference between the third and first quartiles, and indicates the range of the middle 50% of the data. The IQRs for sodium content are 157.5 mg for

beef hot dogs and 156 mg for poultry hot dogs. The IQRs suggest that the spread within the middle half of data for beef hot dogs is similar to the spread within the middle half of data for poultry hot dogs. The boxplots also suggest that each distribution is somewhat skewed right. That is, each distribution appears to have somewhat more variation in the upper half. Considering the degree of variation in the data and the amount of overlap in the boxplots, a difference of 50 mg between the medians is not really that large. Finally, it is interesting to note that more than 25% of beef hot dogs have less sodium than all poultry hot dogs. On the other hand, the highest sodium levels are for beef hot dogs.

Note that there are several variations of boxplots. At Level C, performing an analysis using boxplots might include a test for *outliers* (values that are extremely large or small when compared to the variation in the majority of the data). If outliers are identified, they often are detached from the "whiskers" of the plot. Outlier analysis is not recommended at Level B, so whiskers extend to the minimum and maximum data values. However, Level B students may encounter outliers when using statistical software or graphing calculators.

Measuring the Strength of Association between Two Quantitative Variables

At Level B, more sophisticated data representations should be developed for the investigation of problems that involve the examination of the relationship between two numeric variables. At Level A, the problem of packaging sweat suits (shirt and pants together or separate) was examined through a study of the relationship between height and arm span. There are several statistical questions related to this problem that can be addressed at Level B with a more in-depth analysis of the height/arm span data. For example:

How strong is the association between height and arm span?

Is height a useful predictor of arm span?

Table 8: Height and Arm Span Data

Height	Arm Span	Height	Arm Span
155	151	173	170
162	162	175	166
162	161	176	171
163	172	176	173
164	167	178	173
164	155	178	166
165	163	181	183
165	165	183	181
166	167	183	178
166	164	183	174
168	165	183	180
171	164	185	177
171	168	188	185

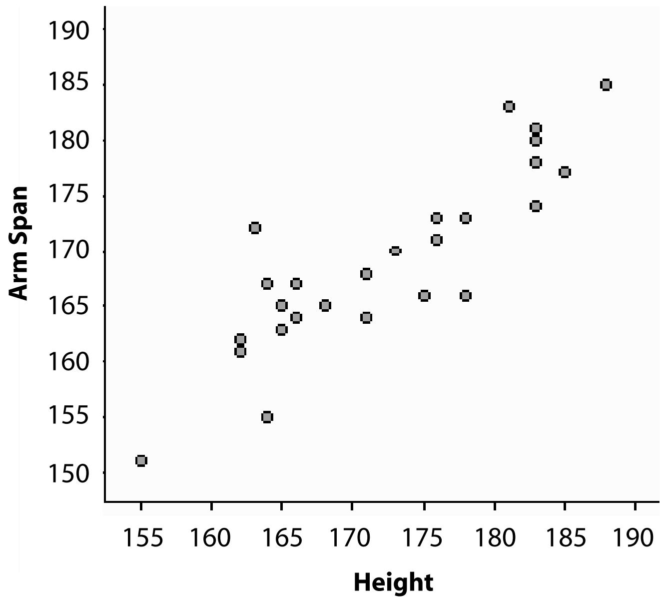

Figure 19: Scatterplot of arm span vs. height

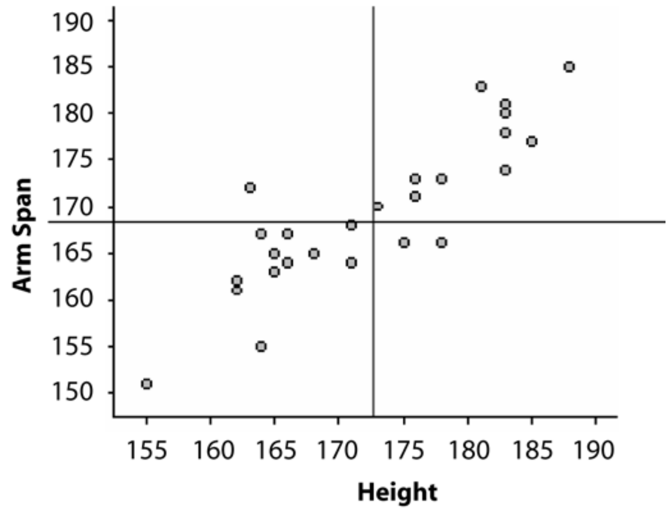

Figure 20: Scatterplot showing means

Table 8 provides data on height and arm span (measured in centimeters) for 26 students. For convenience, the data on height have been ordered.

The height and arm span data are displayed in Figure 19. The scatterplot suggests a fairly strong increasing relationship between height and arm span. In addition, the relationship appears to be quite linear.

Measuring the strength of association between two variables is an important statistical concept that should be introduced at Level B. The scatterplot in Figure 20 for the height/arm span data includes a vertical line drawn through the mean height (x = 172.5) and a horizontal line drawn through the mean arm span (y = 169.3).

The two lines divide the scatterplot into four regions (or quadrants). The upper right region (Quadrant 1) contains points that correspond to individuals with above average height and above average arm span. The upper left region (Quadrant 2) contains points that correspond to individuals with below average height and above average arm span. The lower left region (Quadrant 3) contains points that correspond to individuals with below average height and below average arm span. The lower right region (Quadrant

“A correlation coefficient is a quantity that measures the direction and strength of an association between two variables.”

4) contains points that correspond to individuals with above average height and below average arm span.

Notice that most points in the scatterplot are in either Quadrant 1 or Quadrant 3. That is, most people with above average height also have above average arm span (Quadrant 1) and most people with below average height also have below average arm span (Quadrant 3). One person has below average height with above average arm span (Quadrant 2) and two people have above average height with below average arm span (Quadrant 4). These results indicate that there is a *positive association* between the variables *height* and *arm span*. Generally stated, two numeric variables are *positively associated* when above average values of one variable tend to occur with above average values of the other and when below average values of one variable tend to occur with below average values of the other. *Negative association* between two numeric variables occurs when below average values of one variable tend to occur with above average values of the other and when above average values of one variable tend to occur with below average values of the other.

A *correlation coefficient* is a quantity that measures the direction and strength of an association between two variables. Note that in the previous example, points in Quadrants 1 and 3 contribute to the positive association between height and arm span, and there is a total of 23 points in these two quadrants. Points in Quadrants 2 and 4 do not contribute to the positive association between height and arm span, and there is a total of three points in these two quadrants. One correlation coefficient between height and arm span is given by the QCR (*Quadrant Count Ratio*):

$$\text{QCR} = \frac{23 - 3}{26} = .77$$

A QCR of .77 indicates that there is a fairly strong positive association between the two variables height and arm span. This indicates that a person’s height is a useful predictor of his/her arm span.

In general, the QCR is defined as:

The QCR has the following properties:

$$\frac{(\text{Number of Points in Quadrants 1 and 3}) - (\text{Number of Points in Quadrants 2 and 4})}{\text{Number of Points in all Four Quadrants}}$$

→ The QCR is unitless.

→ The QCR is always between –1 and +1 inclusive.

Holmes (2001) gives a detailed discussion of the QCR. A similar correlation coefficient for 2x2 contingency tables is described in Conover (1999) and discussed in the Appendix for Level B. The QCR is a measure of the strength of association based on only the number of points in each quadrant and, like most summary measures, has its shortcomings. At Level C, the shortcomings of the QCR can be addressed and used as foundation for developing Pearson's correlation coefficient.

Modeling Linear Association

The height/arm span data were collected at Level A in order to study the problem of packaging sweat suits. Should a shirt and pants be packaged separately or together? A QCR of .77 suggests a fairly strong positive association between height and arm span, which indicates that height is a useful predictor of arm span and that a shirt and pants could be packaged together. If packaged together, how can a person decide which size sweat suit to buy? Certainly, the pant-size of a sweat suit depends on a person's height and the shirt-size depends on a person's arm span. As many people know their height, but may not know their arm span, can height be used to help people decide which size sweat suit they wear? Specifically:

Can the relationship between height and arm span be described using a linear function?

Students at Level B will study linear relationships in other areas of their mathematics curriculum. The degree to which these ideas have been developed will determine how we might proceed at this point. For example, if students have not yet been introduced to the equation of a line, then they simply might draw a line through the "center of the data" as shown in Figure 21.

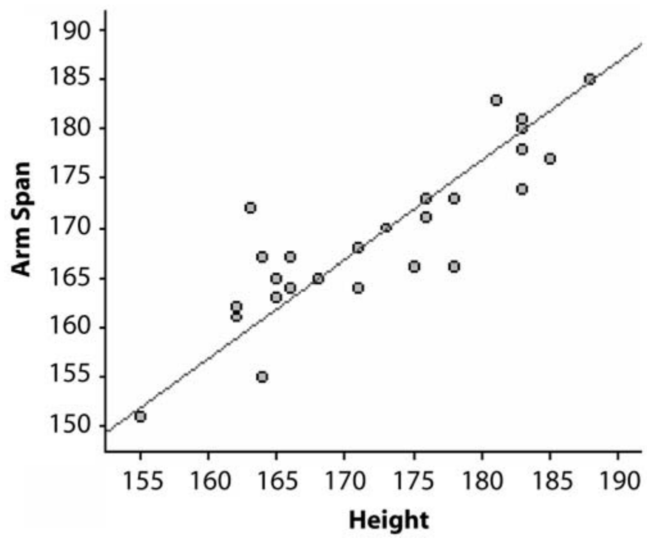

Figure 21: Eyeball line

This line can be used to predict a person's arm span if his or her height is known. For example, to predict the arm span for a person who is 170 cm tall, a vertical segment is drawn up from the X-axis at Height = 170. At the point this vertical segment intersects the segment, a horizontal line is drawn to the Y-axis. The value where this horizontal segment intersects the Y-axis is the predicted arm span. Based on the graph above, it appears that we would predict an arm span of approximately 167 cm for a person who is 170 cm tall.

If students are familiar with the equation for a line and know how to find the equation from two points, then they might use the Mean – Mean line, which is determined as follows. Order the data according to the X-coordinates and divide the data into two "halves" based on this ordering. If there is an odd number of measurements, remove the middle point from the analysis. Determine the means for the X-coordinates and Y-coordinates in each half and find the equation of the line that passes through these two points. Using the previous data:

Lower Half (13 Points)	Upper Half (13 Points)
Mean Height = 164.8	Mean Height = 180.2
Mean Arm Span = 163.4	Mean Arm Span = 175.2

The equation of the line that goes through the points (164.8, 163.4) and (180.2, 175.2) is Predicted Arm Span $\approx$ 37.1 + .766(Height). This equation can be used to predict a person's height more accurately than an eyeball line. For example, if a person is 170 cm tall, then we would predict his/her height to be approximately 37.1 + .766(170) = 167.3 cm. A more sophisticated approach (least squares) to determine a "best-fitting" line through the data will be introduced in Level C.

The Importance of Random Selection

In statistics, we often want to extend results beyond a particular group studied to a larger group, the *population*. We are trying to gain information about the population by examining a portion of the population, called a *sample*. Such generalizations are valid only if the data are representative of that larger group. A representative sample is one in which the relevant characteristics of the sample members are generally the same as those of the population. Improper or biased sample selection tends to systematically favor certain outcomes, and can produce misleading results and erroneous conclusions.

Random sampling is a way to remove bias in sample selection, and tends to produce representative samples. At Level B, students should experience the consequences of nonrandom selection and develop a basic understanding of the principles involved in random selection procedures. Following is a description of an activity that allows students to compare sample results based on personal (nonrandom) selection versus sample results based on random selection.

Consider the 80 circles on the next page. What is the average diameter for these 80 circles? Each student should take about 15 seconds and select five circles that he/she thinks best represent the sizes of the 80 circles. After selecting the sample, each student should find the average diameter for the circles in her/his personal sample. Note that the diameter is 1 cm for the small circles, 2 cm for the medium-sized circles, and 3 cm for the large circles.

Next, each student should number the circles from one to 80 and use a random digit generator to select a random sample of size five. Each student should find

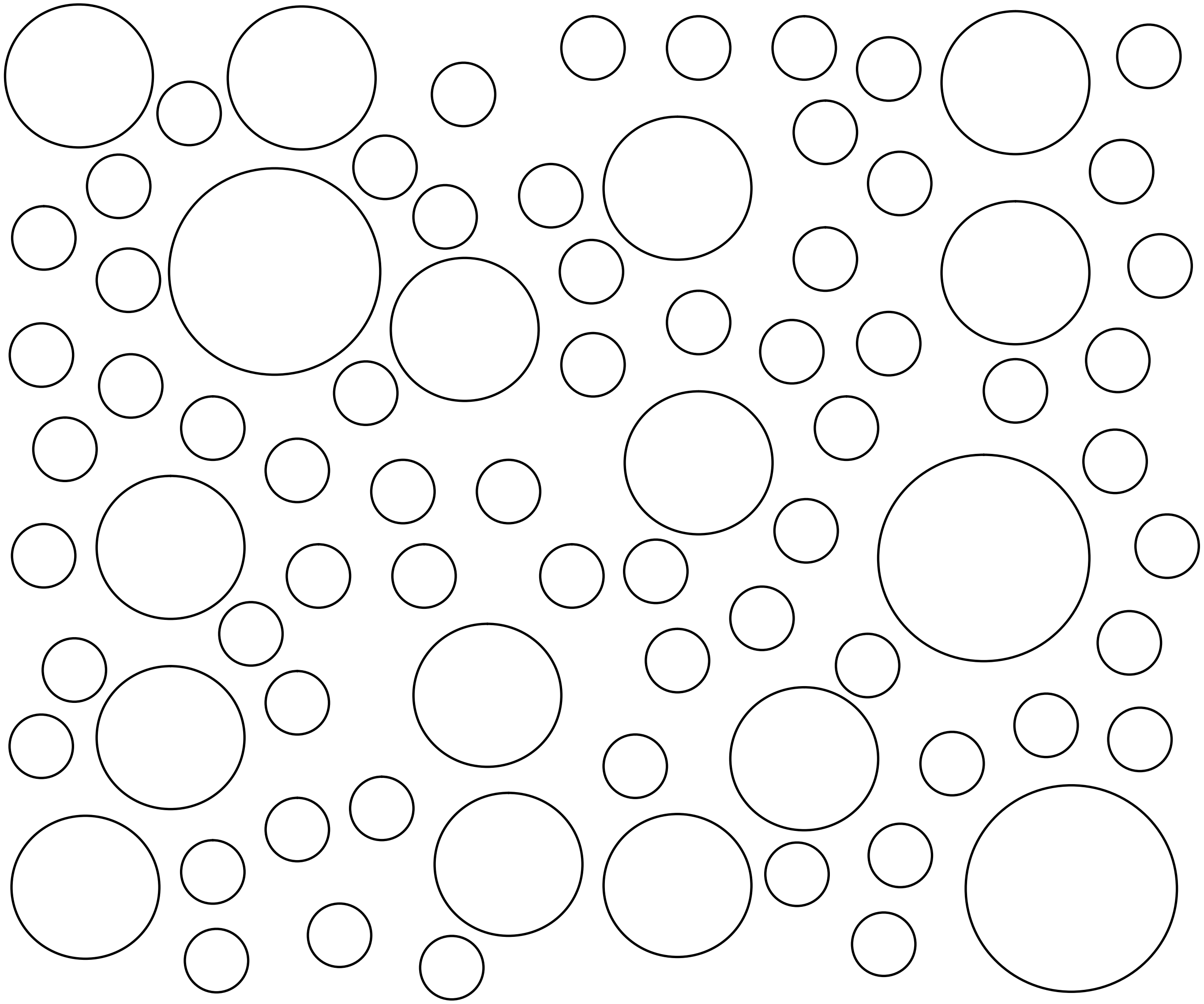

Figure 22: Eighty circles

the average diameter for the circles in his/her random sample. The sample mean diameters for the entire class can be summarized for the two selection procedures with back-to-back stemplots.

How do the means for the two sample selection procedures compare with the true mean diameter of 1.25 cm? Personal selection usually will tend to yield sample means that are larger than 1.25. That is, personal selection tends to be biased with a systematic favoring toward the larger circles and an overestimation of the population mean. Random selection tends to produce some sample means that underestimate the population mean and some that overestimate the population mean, such that the sample means cluster somewhat evenly around the population mean value (i.e., random selection tends to be *unbiased*).

In the previous example, the fact that the sample means vary from one sample to another illustrates an idea that was introduced earlier in the favorite music type survey. This is the notion of sampling variability. Imposing randomness into the sampling procedure allows us to *use probability* to describe the long-run behavior in the variability of the sample means resulting from random sampling. The variation in results from repeated sampling is described through what is called the *sampling distribution*. Sampling distributions will be explored in more depth at Level C.

Comparative Experiments

Another important statistical method that should be introduced at Level B is *comparative experimental studies*. Comparative experimental studies involve comparisons of the effects of two or more *treatments* (experimental conditions) on some response variable. At Level B, studies comparing two treatments are adequate. For example, students might want to study the effects of listening to rock music on one's ability to memorize. Before undertaking a study such as this, it is important for students to have the opportunity to identify and, as much as possible, control for any potential extraneous sources that may interfere with our ability to interpret the results. To address these issues, the class needs to develop a design strategy for collecting appropriate experimental data.

One simple experiment would be to *randomly* divide the class into two equal-sized (or near equal-sized) groups. Random assignment provides a fair way to assign students to the two groups because it tends to average out differences in student ability and other characteristics that might affect the response. For example, suppose a class has 28 students. The 28 students are randomly assigned into two groups of 14. One way to accomplish this is to place 28 pieces of paper in a box—14 labeled "M" and 14 labeled "S." Mix the contents in the box well and have each student randomly choose a piece of paper. The 14 Ms will listen to music and the 14 Ss will have silence.

Table 9: Five-Number Summaries

	Music	Silence
Minimum	3	6
First Quartile	6	8
Median	7	10
Third Quartile	9	12
Maximum	15	14

Each student will be shown a list of words. Rules for how long students have to study the words and how long they have to reproduce the words must be determined. For example, students may have two minutes to study the words, a one-minute pause, and then two minutes to reproduce (write down) as many words as possible. The number of words remembered under each condition (listening to music or silence) is the response variable of interest.

The Five-Number Summaries and comparative boxplots for a hypothetical set of data are shown in Table 9 and Figure 23. These results suggest that students generally memorize fewer words when listening to music than when there is silence. With the exception of the maximum value in the music group (which is classified as an outlier), all summary measures for the music group (labeled M in Figure 23) are lower than the corresponding summary measures for the silence group (labeled S in Figure 23). Without the outlier, the degree of variation in the scores appears to be similar for both groups. Distribution S appears to be reasonably

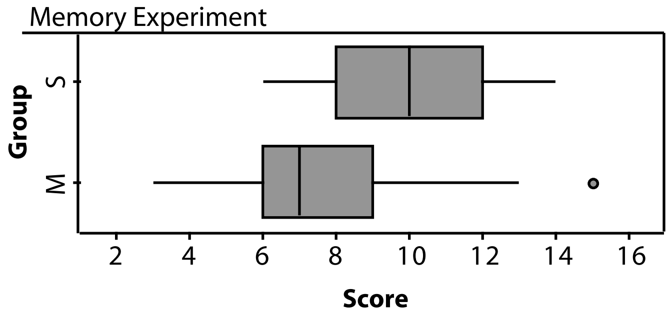

Figure 23: Boxplot for memory data

symmetric, while distribution M is slightly right-skewed. Considering the degree of variation in the scores and the separation in the boxplots, a difference of three between the medians is quite large.

Time Series

Another important statistical tool that should be introduced at Level B is a time series plot. Problems that explore trends in data over time are quite common. For example, the populations of the United States and the world continue to grow, and there are several factors that affect the size of a population, such as the number of births and the number of deaths per year. One question we ask is:

How has the number of live births changed over the past 30 years?

The U.S. Census Bureau publishes vital statistics in its annual *Statistical Abstract of the United States*. The data below are from *The Statistical Abstract of the United States* (2004–2005) and represent the number of live births

Table 10: Live Birth Data

Year	Births (x 1,000)	Year	Births (x 1,000)
1970	3,731	1985	3,761
1971	3,556	1986	3,757
1972	3,258	1987	3,809
1973	3,137	1988	3,910
1974	3,160	1989	4,041
1975	3,144	1990	4,158
1976	3,168	1991	4,111
1977	3,327	1992	4,065
1978	3,333	1993	4,000
1979	3,494	1994	3,979
1980	3,612	1995	3,900
1981	3,629	1996	3,891
1982	3,681	1997	3,881
1983	3,639	1998	3,942
1984	3,669	1999	3,959

per year (in thousands) for residents of the United States since 1970. Note that, in 1970, the value 3,731 represents 3,731,000 live births.

The time series plot in Figure 24 shows the number of live births over time. This graph indicates that:

→ from 1970 to 1975, the number of live births generally declined

→ from 1976 to 1990, the number of live births generally increased

→ from 1991 to 1997, the number of live births generally declined

And it appears that the number of live births may have started to increase since 1997.

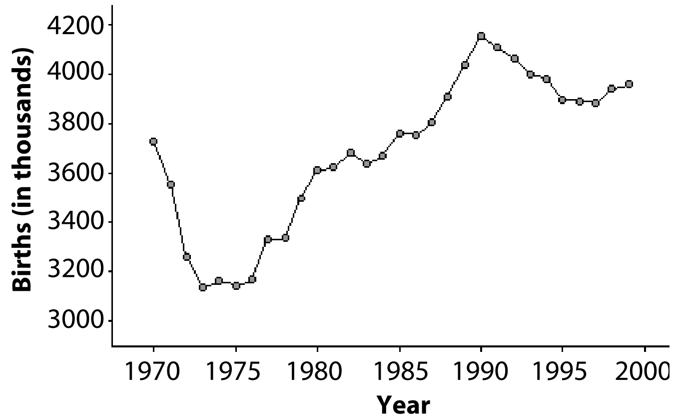

Figure 24: Time series plot of live births

Misuses of Statistics

The introduction of this document points out that data govern our lives. Because of this, every high-school graduate deserves to have a solid foundation in statistical reasoning. Along with identifying proper uses of statistics in questionnaires and graphs, the Level B student should become aware of common misuses of statistics.

Proportional reasoning allows the Level B student to interpret data summarized in a variety of ways. One type of graph that often is misused for representing data is the pictograph. For example, suppose the buying power of a dollar today is 50% of what it was 20 years ago. How would one represent that in a pictograph? Let the buying power of a dollar 20 years ago be represented by the following dollar bill:

If the buying power today is half what it was 20 years ago, one might think of reducing both the width and height of this dollar by one-half, as illustrated in the pictograph below:

Today's dollar at "half" size, representing that it buys only half of what it did 20 years ago.

Today's dollar should look half the size of the dollar of 20 years ago. Does it? Since both the length and the width were cut in half, the area of today's dollar shown above is one-fourth the original area, not one-half.

The two pictographs below show the correct reduction in area. The one on top changes only one dimension, while the other changes both dimensions, but in correct proportion so that the area is one-half the area of the original representation. This example provides the Level B student with an excellent exercise in proportional reasoning.

Today's dollar at half size, with 50% taken from the length.

Today's dollar at half size, with sides in correct proportion to the original.

Poorly designed statistical graphs are commonly found in newspapers and other popular media. Several examples of bad graphs, including the use of an unwarranted third dimension in bar graphs and circle graphs can be found at *www.amstat.org/education/gaise/2*, a web site managed by Carl Schwarz at Simon Fraser University. Students at Level B should be given opportunities to identify graphs that incorrectly represent data and then draw, with the aid of statistical computer

“Poorly designed statistical graphs are commonly found in newspapers and other popular media.”

software, the correct versions. This gives them excellent practice in calculating areas and volumes.

There are many famous misuses of data analysis in the literature, and three are mentioned here. The magazine *Literary Digest* erred in 1936 when it projected that Alf Landon would defeat Franklin Delano Roosevelt by a 57 to 43 percent margin based on responses to its survey. Each survey included a subscription form to the magazine, and more than 2.3 million were returned. Unfortunately, even large voluntary response surveys are generally not representative of the entire population, and Roosevelt won with 62% of the vote. George Gallup correctly projected the winner, and thereby began a very successful career in using random sampling techniques for conducting surveys. Learning what Gallup did right and the *Literary Digest* did wrong gives the Level B student valuable insight into survey design and analysis. A more detailed discussion of this problem can be found in Hollander and Proschan (1984).

The 1970 Draft Lottery provides an example of incorrectly applying randomness. In the procedure that was used, capsules containing birth dates were placed in a large box. Although there was an effort to mix the capsules, it was insufficient to overcome the fact that the capsules were placed in the box in order from January to December. This resulted in young men with birth dates in the latter months being more likely to have their dates selected sooner than birth dates elsewhere in the year. Hollander and Proschan (1984) give an excellent discussion of this problem.

The 25th flight of NASA's space shuttle program took off on January 20, 1986. Just after liftoff, a puff of gray smoke could be seen coming from the right solid rocket booster. Seventy-three seconds into the flight, the *Challenger* exploded, killing all seven astronauts aboard. The cause of the explosion was determined to be an O-ring failure, due to cold weather. The disaster possibly could have been avoided had available data been displayed in a simple scatterplot and correctly interpreted. The *Challenger* disaster has become a case study in the possible catastrophic consequences of poor data analysis.

Summary of Level B

Understanding the statistical concepts of Level B enables a student to begin to appreciate that data analysis is an investigative process consisting of formulating their own questions, collecting appropriate data through various sources (censuses, nonrandom and random sample surveys, and comparative experiments with random assignment), analyzing data through graphs and simple summary measures, and interpreting results with an eye toward inference to a population based on a sample. As they begin to formulate their own questions, students become aware that the world around them is filled with data that affect their own lives, and they begin to appreciate that statistics can help them make decisions based on data. This will help them begin to appreciate that statistics can help them make decisions based on data, investigation, and sound reasoning.