Chapter 12
Learning to Reason About Samples and Sampling Distributions

Why is the role of sample size in sampling distributions so hard to grasp? One consideration is that . . . the rule that the variability of a sampling distribution decreases with increasing sample size seems to have only few applications in ordinary life. In general, taking repeated samples and looking at the distribution of their means is rare in the everyday and only recent in scientific practice.

(Sedlmeier & Gigerenzer, 1997, p. 46)

Snapshot of a Research-Based Activity on Sampling

The activity begins with the teacher holding up a bag of Reese's Pieces candies, which includes a mixture of three different colors (orange, yellow, and brown) and asking students to make conjectures about what they might expect the number of orange candies to be if they take repeated samples of 25 candies from the bag. Students propose different numbers such as 10 or 15. They are also asked if they would be surprised (and complain) if they had only 5 orange candies in their sample of 25 candies. Students consider the variation they would expect to see from sample to sample in the proportion of orange candies, acknowledging that they would not expect every sample of 25 to have exactly the same number of orange candies. Then they count proportions of orange Reese's Pieces in samples of 25 candies presented to them in small paper cups, reporting their counts to be numbers like 11, 12, 13, and 14.

The students plot their individual sample proportions of orange candies in a dot plot on the board and use the data to discuss the variability between samples, as well as to estimate the proportion of candies in the large bag from which the candies were sampled. Then students use a Web applet that simulates sampling Reese's Pieces from a hypothetical population (Reese's Pieces Sampling applet from the *Rossman-Chance* Website, http://rossmanchance.com, Fig. 12.3). Students then compare their individual proportions to distributions of sample proportions produced by a claim about the population proportion of orange candies. They predict and test the effect of taking larger or smaller samples on the closeness of each sample proportion to the population parameter, using simulated data from the Web applet. By the end of this activity, students have determined if a result of 5 orange candies in a random sample of 25 Reese's Pieces is surprising and would be cause for complaint.

J.B. Garfield, D. Ben-Zvi, *Developing Students' Statistical Reasoning*,

Rationale for This Activity

This *Reese's Pieces Activity* helps students focus on how they may expect sample statistics to vary from sample to sample when taking repeated samples from a population, an idea that develops slowly with repeated experience with random samples. Students' intuitions about variability between samples are often misleading: many students think samples should be more similar to each other and to the population parameter or more variable than what we can expect due to chance. In this lesson, students also develop an informal sense of statistical inference when they determine if a particular sample result is surprising or unusual, by comparing this result to a simulated distribution of sample statistics.

The approach used in this lesson is to develop conceptual thinking before students are introduced to the formal ideas of sampling distribution and statistical inference. By preceding the computer simulation with a parallel concrete activity, students are more likely to understand and believe the results of the simulation tool, and are able to develop reasonable ideas of sampling variability. Finally, by having students make conjectures that they test with data, they become more engaged in reasoning about the statistical concepts than if the concepts were presented to them.

The Importance of Understanding Samples and Sampling Distributions

Taking samples of data and using samples to make inferences about unknown populations are at the core of statistical investigations. Although much of data analysis involves analyzing a single sample and making inferences based on this sample, an understanding of how samples vary is important in order to make reasoned estimates and decisions. Most introductory statistics classes include distinctions between samples and populations, and develop notions of sampling variability by examining similarities and differences between multiple samples drawn from the same population. In high school and college classes, the study of sampling variability is extended to examining distributions of sample statistics. Looking at distributions of sample means for many samples drawn from a single population allows us to see how one sample compares to the rest of the samples, leading us to determine if a sample is surprising (unlikely) or not surprising. This is an informal precursor to the more formal notion of P-value that comes with studying statistical inference. We note that simulations based on randomization activities, such as those described in Chapter 7 on models, can also be used to examine the place of an observed value in a distribution of sample statistics to judge whether this result may be explained by chance or due to a particular treatment or condition.

Comparing means of samples drawn from the same population also helps build the idea of sampling variability, which leads to the notion of sampling error, a fundamental component of statistical inference whether constructing confidence intervals or testing hypotheses. Sampling error indicates how much a sample statistic may be expected to differ from the population parameter it is estimating. The sampling error is used in computing margins of error (for confidence intervals) and is used in computing test statistics (e.g., the t statistic when testing hypotheses)

The idea of a sample is likely to be a familiar one to students. They have all taken samples before (for example, tasting a food sample at a grocery store) and have an idea of a sample as something that is drawn from or represents something bigger. Students seem to have an intuitive sense that each sample may differ from the other samples drawn from the same larger entity. It may, therefore, seem surprising that students have such difficulty understanding the behavior of samples when they study statistics, how they relate to a population, and what happens when many samples are drawn and their statistics accumulated in a sampling distribution.

The two central ideas of sampling: sampling representativeness and sampling variability have to be understood and carefully balanced in order to understand statistical inference. Rubin, Bruce, and Tenney (1991) cautioned that over reliance on sampling representativeness leads students to think that a sample tells us everything about a population, while over reliance on sampling variability leads students to think that a sample tells us nothing useful about a population. In fact, the ideas of sample and sampling distribution build on many core concepts of statistics, and if these concepts are not well understood, students may never fully understand the important ideas of sampling. For example, the fundamental ideas of distribution and variability underlie an understanding of sampling variability (how individual samples vary) and sampling distribution (the theoretical distribution of a sample statistic computed from all possible samples of the same size drawn from a population). The idea of center is also involved (understanding the mean of the sampling distribution) as is the idea of model (the Normal Distribution as a model that fits sampling distributions under certain conditions). We also interpret empirical sampling distributions of simulated or collected data in similar ways, sometimes referring to these as, for example, a distribution of 500 sample means, rather than referring to it as a (theoretical) sampling distribution. Finally, samples and sampling variability also build on basic ideas of randomness and chance, or the study of probability.

The Place of Samples and Sampling Distributions in the Curriculum

The teaching of sampling variability typically comes later in introductory statistics courses, after the study of the foundational concepts of distribution, center, and variability. It is often introduced after the formal study of the Normal Distribution and its characteristics, and after studying probability. However, in recent years, some textbooks introduce ideas of sampling early in a course, along with informal ideas of inference that are revisited throughout the course until formalizing them later in chapters on significance tests and confidence intervals.

Sampling and sampling distributions are prerequisite topics that precede the formal methods of making statistical inferences (constructing confidence intervals and finding P-values used to compare a sample of data to a distribution based on a null hypothesis). Often the culminating lesson on sampling distributions is one that introduces and illustrates the Central Limit Theorem (CLT).

The implications of this important theorem are that for large sample sizes, distributions of sample statistics (typically, means and proportions) will be normal or approximately normal. This fact allows us to use the normal probability distribution to estimate probabilities for different intervals of values of sample means (given a particular population or hypothesis) and allows us to use the z or t distribution when making formal inferences. Although this theorem is usually included in every introductory statistics course as a prerequisite for learning statistical inference, most students never appear to understand this theorem during a course, although some are able to repeat or identify correct definitions.

Before technology tools were readily available, most text books showed some pictures of different shaped curves, and then presented what sampling distributions would look like for a few different sample sizes. The purpose was to show that curves of these sampling distributions became more normal and narrower as the sample size increased. Then students were told how to calculate the standard error (*sigma* over the square root of n, $\frac{\sigma}{\sqrt{n}}$, for the sample mean) and to use this to find probabilities (e.g., that a sample mean is larger than 3, under a particular null hypothesis). Students often confuse the standard error with the sample standard deviation (s) and the population value (σ). Some typical textbook questions ask students to calculate a probability for a single value (convert x to z using σ) and then repeat the problem for the same value but of a sample mean (convert $\bar{x}$ to z using $\frac{\sigma}{\sqrt{n}}$). Students often fail to notice or understand the difference between these two procedures.

Review of the Literature Related to Reasoning About Samples and Sampling Distributions

Because sampling and sampling distributions are so confusing to students, there has been a considerable amount of research on this topic, particularly with college students. However, even at the elementary and secondary levels, studies have examined how students understand and misunderstand ideas of samples and sampling.

Studies of Students in Precollege Level Settings

A study of students' conceptions of sampling in upper elementary school by Jacobs (1999) suggested that students understood the idea that a sample is a small part of a whole and that even a small part can give you an idea of the whole. Watson and Moritz (2000a, 2000b) also studied children's intuitive ideas of samples, and identified six categories of children's thinking about this concept. They point out that while students have a fairly good "out of school" understanding of the concept of sample, they have difficulty making the transition to the formal, statistical meaning of this term and the related connotations. For example, one can make good generalizations from a small sample of food or blood to the larger entity from which it was drawn, but these intuitive ideas do not generalize to the notion of sampling variation and the need for large, representative samples in making statistical estimates. Watson and Moritz (2000a) suggest making explicit these differences (e.g., between taking a small cube of cheese which represents a homogeneous entity, with taking a sample from the population of fifth grade students to estimate a characteristic such as height, which is a population that has much variability). Watson (2004), in a summary of research on reasoning about sampling, describes how students often concentrate on fairness, and prefer biased sampling methods such as voluntary samples because they do not trust random sampling as a process producing fair samples. Saldanha and Thompson (2002) found both of these types of conceptions on sampling in high school students in a teaching experiment conducted in a statistics class. Not surprisingly, they found that only the concept of sampling as part of a repeated process with variability from sample to sample supported the notion of distribution needed to understand sampling distributions.

In a teaching experiment with eighth grade students, Bakker (2004b) was able to help students understand that larger samples are more stable (less variable) and better represent the population, using a sequence of "growing samples" activities. In a growing a sample activity (see Chapter 8), students predict and explain what happens to a graph when bigger samples are taken (Konold & Pollatsek, 2002). The goal of the growing samples activity was to use imagined and computer-simulated sets of data to build students' reasoning about sampling in the context of variability and distribution. Activities were designed to begin with students' own ideas and guide them toward more conventional notions and representations. Bakker (2004b) suggests that asking students to make conjectures about possible samples of data push them to use conceptual tools to predict the distributions, which helps them develop reasoning about samples.

Studies Involving College Students

Confusion about sampling has been found in college students and professionals. In their seminal paper, "Belief in the Law of Small Numbers," psychologists Tversky and Kahneman (1971) wrote:

> The research suggests that people have strong intuitions about random sampling; that these intuitions are wrong in fundamental aspects; that these intuitions are shared by naïve subjects and by trained scientists, and that they are applied with unfortunate consequences in the course of scientific inquiry . . . People view a sample randomly drawn from a population as highly representative, that is, similar to the population in all essential characteristics. Consequently, they expect any two samples drawn from a particular population to be more similar to one another and to the population than sampling theory predicts, at least for small samples. (p. 24).

Since the publication of this article, many researchers have examined and described the difficulties students have understanding samples, sampling variability, and inevitably, sampling distributions and the Central Limit Theorem. In a summary of articles by psychologists on this topic, Well, Pollatsek, and Boyce (1990) noted that people sometimes reason correctly about sample size (e.g., that larger samples better represent populations) and sometimes do not (e.g., thinking that both large and small samples equally represent a population). To reveal the reasons for this discrepancy, they conducted a series of experiments that gave college students questions involving reasoning about samples and sampling variability. For example, one problem described a post office that recorded the heights of local males when they turned 18. The average height of 18 year old males is known to be 5 ft 9 in. (1.75 m). Information is given on two post offices: A which registered 25 males and B which registered 100 males. Students are asked which set of heights would have a mean closer to 5 ft 9 in. Most were able to correctly pick B, which had a larger set of males. Students were also asked to estimate the percentage of days for each post office that produced an average height greater than 6 ft (1.83 m). Fewer students were able to reason correctly about this problem.

The first part of this problem looked at the accuracy of small samples compared to large samples and the second part asked students to think about which sample mean would be more likely to fall in the tail of a distribution of sample means, far away from the population mean. The researchers found that students used sample size more wisely when asked the first type of question (which sample size is more accurate) than on the question that asked them to pick which sample would produce a value in the tail of the population distribution, indicating that they do not understand the variability of sample means. They also noted that students confused distributions for large and small samples with *distributions of averages* based on large and small samples. The authors concluded that students' statistical intuitions are not always incorrect, but may be crude and can be developed into correct conceptions through carefully designed instruction.

Summarizing the research in this area as well as their own experience as statistics teachers and classroom researchers, delMas, Garfield, and Chance (2004) list the following common misconceptions about sampling distributions:

- The sampling distribution should look like the population (for $n > 1$).
- Sampling distributions for small and large sample sizes have the same variability.
- Sampling distributions for large samples have more variability.
- A sampling distribution is not a distribution of sample statistics.
- One sample (of real data) is confused with all possible samples (in distribution) or potential samples.
- The Law of Large Numbers (larger samples better represent a population) is confused with the Central Limit Theorem (distributions of means from large samples tend to form a Normal Distribution).
- The mean of a positive skewed distribution will be greater than the mean of the sampling distribution for large samples taken from this population.

In addition, students have been found to believe that a sample is only good (e.g., representative) if the sample size represents a large percentage when compared to the population (e.g., Smith, 2004). To confront the common misconceptions that develop and to build sound reasoning about samples and sampling distributions, statistics educators and researchers have turned to technological tools to illustrate the abstract processes involved in repeated sampling from theoretical populations and help students develop statistical reasoning.

In a series of studies, Sedlmeier and Gigerenzer (1997) revealed that when subjects seem to have a good understanding of the effect of sample size, they are thinking of one frequency distribution (for a sample). When subjects show confusion about sample size, they are struggling with the more difficult notion of sampling distribution (statistics from many samples). Sedlmeier (1999) continued this research, and found that if he converted items that required subjects to consider sampling distributions, to ones that instead required frequency distributions, higher percentage of correct solutions were obtained.

Building on this work, Saldanha and Thompson (2002) studied high school students' reasoning about samples and sampling distributions in a teaching experiment. They identified a multiplicative concept of samples that relates the sample to the population as well as to a sampling distribution in a visual way. This interrelated set of images is believed to build a good foundation for statistical inference, which suggests that instructors clearly help students distinguish between three levels of data: the population distribution, the sample distribution, and the sampling distribution. Lane-Getaz (2006) provides such a visual model in her Simulation Process Model, which we have adapted and called the *Simulation of Samples (SOS)* Model. This model, shown in Fig. 12.1, distinguishes between the first level of data (population), many random samples from the population (level 2) along with sample statistics for each sample, and the distribution of sample statistics (Level 3). In the last level, a sample outcome can be compared to the distribution of sample statistics to determine if it is a surprising outcome, an informal approach to statistical inference.

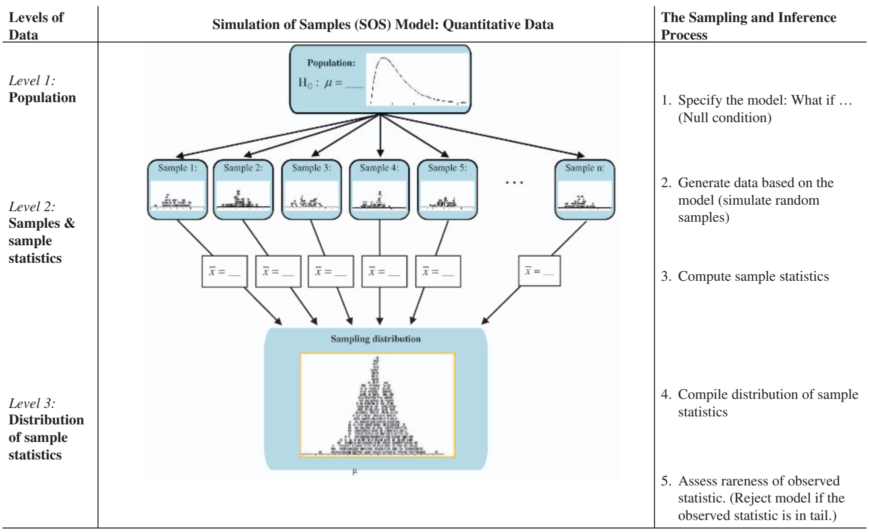

Fig. 12.1 The *Simulation of Samples* model (SOS) for quantitative data

Use of Simulations to Develop Reasoning About Sampling

There are several articles (see Mills, 2002 for a review of these articles) that discuss the potential advantage of simulations in providing examples of the process of taking repeated random samples and allowing students to experiment with variables that affect the outcomes (sample size, population parameters, etc.). In particular, technology allows students to be directly involved with the "building up" of the

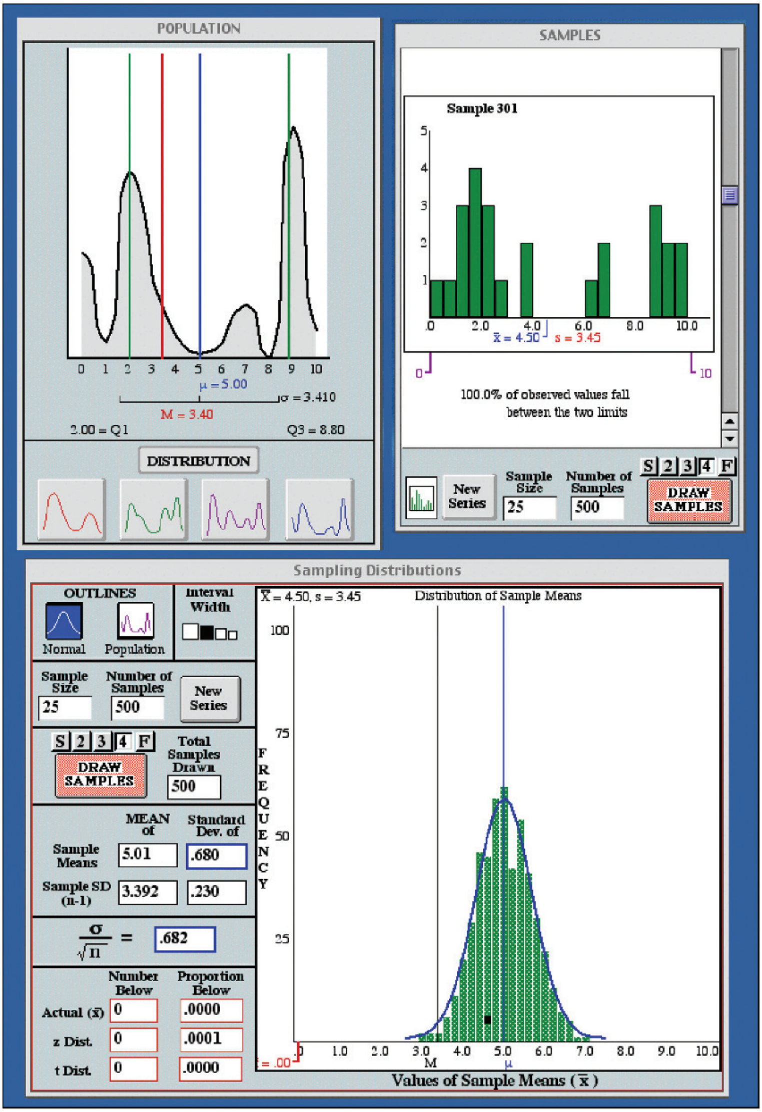

Fig. 12.2 Screen shots of the *Sampling SIM* software

sampling distribution, focusing on the process involved, instead of being presented only the end result. Recently, numerous instructional computer programs have been developed that focus on the use of simulations and dynamic visualizations to help students develop their understanding of sampling distributions and other statistical concepts (e.g., Aberson, Berger, Healy, Kyle, & Romero, 2000). However, despite the development of flexible and visual tools, research suggests that just showing students demonstrations of simulations using these tools will not necessarily lead to improved understanding or reasoning.

Chance, delMas, and Garfield (2004) report the results of a series of studies over a 10-year period that examined various ways of having students interact with the *Sampling SIM* software (delMas, 2001a). *Sampling SIM* software, described in more detail later in this chapter, is a program that allows students to specify different population parameters and generate random samples of simulated data along with many options for displaying and analyzing these samples (see Fig. 12.2). They found that it worked better to have students first make a prediction about a sampling distribution from a particular population (e.g., its shape, center, and spread), then to generate the distribution using software, and then to examine the difference between their prediction and the actual data. They then tried different ways to embed this process, having students work through a detailed activity, or be guided by an instructor. Despite students appearing to be engaged in the activity and realizing the predictable pattern of a normal looking distribution for large samples from a variety of populations, they nonetheless had difficulty applying this knowledge to questions asking them to use the Central Limit Theorem to solve problems. An approach to using the software combines a concrete activity (Sampling Reese's Pieces) with the use of some Web applets, before moving to the more abstract *Sampling SIM* Software. This sequence of activities will be described in the following section.

Implications of the Research: Teaching Students to Reason About Samples and Sampling Distributions

We believe that it is important to introduce ideas of sample and sampling to students early in a statistics course, preferably in a unit on data production and collection. By the time students are ready to study the formal ideas of sampling distributions, they should have a good understanding of the foundational concept of sample, variability, distribution, and center. They should also understand the model of the Normal Distribution and how that model may be used to determine (or estimate) percentages and probabilities (e.g., use the Empirical Rule).

As students learn methods of exploring and describing data, they should be encouraged to pay attention to ideas of samples and to consider sampling methods (e.g., where did the data come from, how was the sample obtained, how do different samples vary). By the time students begin the formal study of sampling variability, they should understand the nature of a random sample and the idea of a sample being representative of a population. They should understand how to choose a good sample and the importance of random sampling.

The study of sampling variability typically focuses on taking repeated samples from a population and comparing sample statistics, (such as the sample mean or the sample proportion). There is a lack of agreement among college statistics teachers and textbooks about whether to begin the study of sampling distributions with proportions or means. Both have their advantages and disadvantages. Chance and Rossman (2001) present both sides of the disagreement in the form of a debate. They conclude that it is more important to pay careful attention to ideas of data collection and sampling throughout the introductory course, than which statistic is introduced first. They recommend that much time should be spent on sampling distributions so that students will be able to use these ideas as a basis for understanding statistical inference.

DelMas et al. (2004) list desired learning outcomes for the situation where distributions of means are used. These learning outcomes include understanding that:

- A sampling distribution for means (based on quantitative data) is a distribution of sample means (statistics) of a given sample size, calculated from samples that are randomly selected from a population with mean μ and standard deviation σ. It is a probability distribution for the sample mean.
- The sampling distribution for means has the same mean as the population (parameter).
- As n gets larger, variability of the sample means gets smaller (as a statement, a visual recognition, and as a prediction of what will happen or how the next picture will differ).
- The standard error of the mean is a measure of variability of sample statistic values.
- The building block of a sampling distribution is a sample statistic. In other words, the units shown in the distribution are sample statistics, rather than individual data values (e.g., measurements).
- Some values of statistics are more or less likely than others to result from a sample drawn from a particular population.
- It is reasonable to use a normal approximation for a sampling distribution under certain conditions.
- Different sample sizes lead to different probabilities for the same value (know how sample size affects the probability of different outcomes for a statistic).
- Sampling distributions tend to look more normal than the population, even for small samples unless the population is Normal.
- As sample sizes get very large, all sampling distributions for the mean tend to have the same shape, regardless of the population from which they are drawn.
- Averages are more normal and less variable than individual observations. (Again, unless the population is Normal.)
- A distribution of observations in one sample differs from a distribution of statistics (sample means) from many samples (n greater than 1) that have been randomly selected.

- A sampling distribution would look different for different populations and sample sizes (in terms of shape, center, and spread, and where the majority of values would be found).
- Some values of the sample mean are likely, and some are less likely for different sampling distributions. For sampling distributions for a small sample size, a particular sample mean that is farther from the population mean may not be unlikely, but this same value may be unlikely for a sample mean for a larger sample.
- That the size of the standard error of the mean is determined by the standard deviation of the population and the sample size, and that this affects the likelihood of obtaining different values of the sample mean for a given sampling distribution.

Moving from Concrete Samples to Abstract Theory

The research suggests that when students view simulations of data, they may not understand or believe the results, and instead watch the simulations without reasoning about what the simulation represents. Therefore, many instructors find it effective to first provide students with concrete materials (e.g., counting Reese's Pieces candies or pennies) before moving to an abstract simulation of that activity. One nice feature of the Reese's Pieces Samples applet described below is that it provides an animation to show the sampling process that can later be turned off, to provide data more quickly when students understand where the data values are coming from.

One of the challenges in teaching about sampling distributions is that students are already familiar with analyzing samples of data, a concrete activity. When they are asked to imagine many, many samples of a given sample size, they are forced to grapple with the theory that allows them to later make inferences. Many students become confused and think the point is that they should always take many samples. It is difficult for students to understand that in order to later make inferences from one single sample from an unknown population; they must first observe the behavior of samples from a known population. This is far from a trivial task, and using technology seems to help. It is also helpful to show a model of the sampling process that distinguishes between three levels: the population, the samples, and the sample statistics calculated from those samples (see the *SOS Model* in Fig. 12.1 above).

Technological Tools to Visualize Sampling and Sampling Distributions

Perhaps the first technology tool designed to illustrate sampling and the Central Limit Theorem (CLT) was the *Sampling Laboratory*, which ran only on the Macintosh platform, described by Rubin and Bruce (1991). This program visually illustrated the process of taking samples from a population and accumulating distributions of particular characteristics (e.g., mean or median). Individual samples could also be displayed. Since then, many Web applets and simulation programs have been developed and are often used to visually illustrate the CLT. *Sampling SIM* software (delMas, 2001a) as mentioned earlier is a free-standing program that was developed to develop students' reasoning about sampling distributions. Students may select different types of populations (e.g., right-skewed, bimodal, normal) as well as different means and standard deviations, and then explore the impact of changing these parameters on the resulting simulated data. Figure 12.2 shows three windows that may be displayed simultaneously in the *Sampling SIM* program (delMas, 2001a). The Population window shows a population that may be continuous (as shown), discrete (bars), or binomial. Users may select particular populations of interest (e.g., normal, skewed, bimodal), choose one of the irregular shapes shown at the bottom of the Population window, or create their own distributions by raising and lowering the curve or bars using the mouse. The Samples window shows a histogram of each random sample that is drawn along with the sample statistics. The Sampling Distribution window shows the distributions of sample statistics as they accumulate, during the sampling process. Lines show the placement of the population mean and median, and a curve representing the Normal Distribution can be superimposed on the distribution of sample statistics, as well as the outline of the original population.

Among the best of the many Web applets that simulate sampling and sampling distributions are applets from the *RossmanChance* Website (http:// rossmanchance.com/) that help students make the bridge between sampling objects to

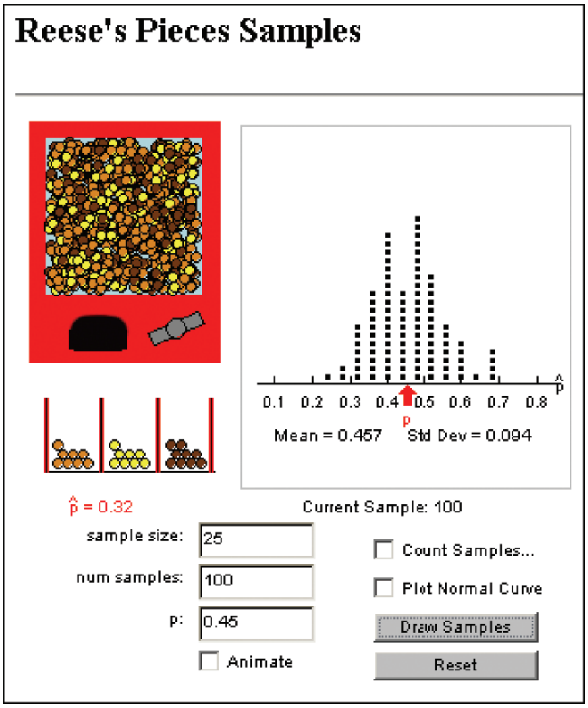

Fig. 12.3 Reese's pieces sampling applet from the *RossmanChance* Web site

sampling abstract data values. For example, the Reese's Pieces Samples applet (see Fig. 12.3) samples colored candies from a population that has a specified percent of orange candies. Different sample sizes and numbers of samples can be drawn. Animation shows the candies being sampled, but this can be turned out after a while to expedite the sampling. There are other *RossmanChance* Web applets that illustrate sampling coins to record their ages, and sampling word lengths from the Gettysburg Address (see Fig. 12.4, used in the *Gettysburg Address Activity*, Lesson 3, Unit on Data, Chapter 6). An advantage in using these applets is that they may be preceded by actual physical activities in class (taking samples of candies, coins, and words), which provide a real context for drawing samples.

Programs such as *Fathom* (Key Curriculum Press, 2006) can also be used to illustrate the sampling process, allowing for parameters to be varied such as sample size, number of samples, and population shape. Although not as visually effective as a computer screen, samples may also be taken and accumulated using graphing calculators. Despite the numerous software tools that currently exist to make this difficult concept more concrete, there is still little research on the most effective ways to use these tools. For example, if a teacher shows a demonstration of the software to the students, is that effective? Is it better for students to interact with the

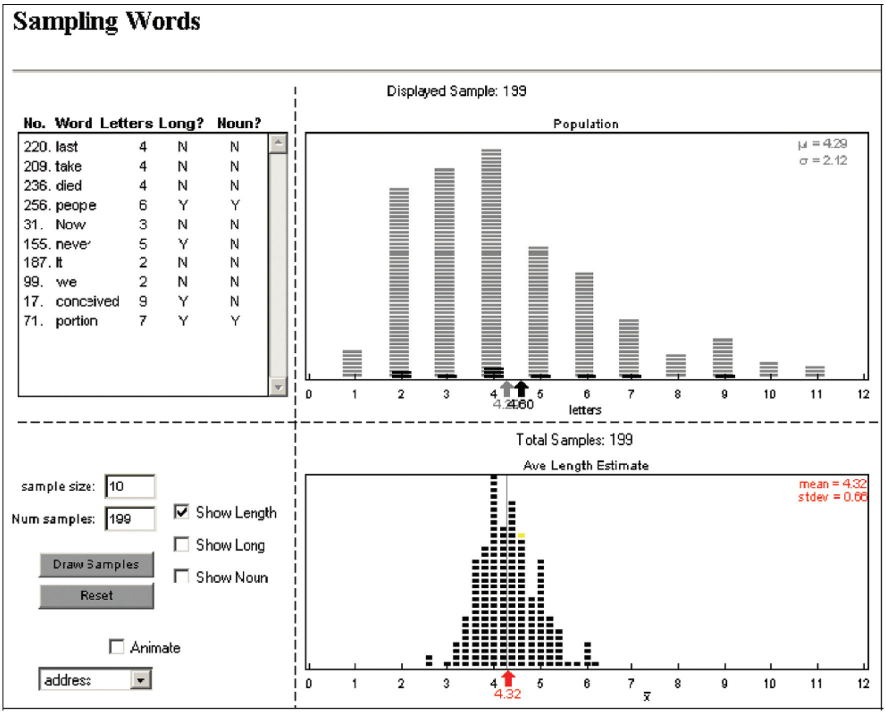

Fig. 12.4 Sampling words applet from the *RossmanChance* Website

software directly, taking their own samples? How much guidance should be given to students when using the software? Are some tools more effective than others? All of these research questions are waiting to be investigated.

Progression of Ideas: Connecting Research to Teaching

Introduction to the Sequence of Activities to Develop Reasoning About Samples and Sampling Distributions

Back in 1991, Rubin and Bruce proposed a sequence of ideas to lead high school students to understand sampling as they participated in activities using *Sampling Lab* software. Their list included:

1. A sample statistic is not necessarily the same as the corresponding population parameter, but it can provide good information about that parameter.
2. Random samples vary, especially for small samples. Therefore, the sample statistics will vary for each sample as well.
3. The variation from sample to sample is not due to error, but is a consequence of the sampling process. It occurs even with unbiased sampling methods and carefully chosen samples.
4. Although sample statistics vary from population parameter, they vary in a predictable way. Most sample statistics are close to the population parameter, and fewer are extremely larger or smaller than the population value.
5. Despite sampling variation, a large enough random sample can be used to make a reasonably good prediction for a population parameter.
6. The goodness of a particular estimate is directly dependent on the size of the sample. Samples that are larger produce statistics that vary less from the population value.

The research and related literature reviewed suggest a progression of activities that can be used to help students develop the ideas of sampling variability and sampling distributions described above. Table 12.1 contains a progression of ideas that build on those suggested by Rubin and Bruce (1991) along with types of activities that may be used to develop these ideas.

One important implication from the research is that it takes time to help students develop the ideas related to sampling distribution, longer than just one class session which is the amount of time typically allotted. In addition, prior to a formal unit on sampling distribution, students need experience in taking samples and learning how samples do and do not represent the population. This may be part of an early unit on collecting data through surveys and experiments, where they learn characteristics of good samples and reasons for bad samples (e.g., bias). Another implication is that a visual model (e.g., the *SOS model*, Fig. 12.1 above) may help students develop a deeper understanding of sampling distribution and inference, if it is repeatedly used when dealing with repeated samples and simulations.

Table 12.1 Sequence of activities to develop reasoning about samples and sampling distributions[1]

Milestones: ideas and concepts	Suggested activities
Informal ideas prior to formal study of samples and sampling distributions	
• Population parameter is fixed, but sample statistics vary from sample to sample	• The Gettysburg Address Activity (Lesson 3, Data Unit, Chapter 6)
• The idea of a random sample	• The Gettysburg Address Activity (Lesson 3, Data Unit, Chapter 6)
• As a sample grows, or as more data are collected, at some point the sample provides a stable estimate of the population parameter	• Growing a Distribution Activity (Lesson 1, Distribution Unit, Chapter 8)
• Larger random samples are more likely to be representative of the population than small ones	❖ An Activity where samples are taken from a specified population and the size of the sample is increased to determine at what point the estimates of the population are stable. (The symbol ❖ indicates that this activity is not included in these lessons.)
• The size of a representative sample is not related to a particular percentage of the population. A large well-chosen sample can be a good one even if it is a small percent of the population	❖ An activity where different sample sizes are examined in light of how well they represent the population in terms of shape, center, and spread
Formal ideas of samples and sampling distributions	
• Sample variability: Samples vary for a given sample size, for a random sample from the same population	• Reese's Pieces Activity (Lesson 1: "Sampling from a Population")
• Variability of sample statistics from sample to sample	• Reese's Pieces Activity (Lesson 1)
• There are three levels of data involved in taking random samples: the population, the individual samples, and the distribution of sample statistics	• Reese's Pieces Activity (Lesson 1)
• How and why statistics from small samples vary more than statistics from large samples	• Reese's Pieces Activity (Lesson 1)
• Sample statistics can be graphed and summarized in a distribution, just as raw data may be graphed and summarized	• Reese's Pieces Activity (Lesson 1)

[1] See page 391 for credit and reference to authors of activities on which these activities are based.

Table 12.1 (continued)

Milestones: ideas and concepts	Suggested activities
• Understanding that a simulation of a large number (e.g., 500) sample statistics is a good approximation of a sampling distribution	❖ An activity using a simulation computer tool that draws students' attention to these ideas
• Understanding that for a large number of trials (simulations) what is important to focus on is the change in sample size, not the change in number of simulations	❖ An activity using a simulation computer tool that draws students' attention to these ideas
• Although sample statistics vary from population parameter, they vary in a predictable way	• Body Temperatures, Sampling Words, and Sampling Pennies Activities (Lesson 2: "Generating Sampling Distributions")
• When and why a distribution of sample statistics (for large enough samples) looks bell shaped	• Central Limit Theorem Activity (Lesson 3: "Describing the Predictable Pattern: The Central Limit Theorem")
• Distributions of sample statistics tend to have the same predictable pattern for large random samples	• Central Limit Theorem Activity (Lesson 3)
• Understanding how the Central Limit Theorem describes the shape, center, and spread of sampling distributions of sample statistics	• Central Limit Theorem Activity (Lesson 3)
Building on formal ideas of samples and sampling distributions in subsequent topics	
• Understand the role of sample variability in making statistical inferences	• Activities (Lessons 1, 2, 3, and 4, Statistical Inference Unit, Chapter 13)

Introduction to the Lessons

Building on the basic idea of a sample, these lessons provide students with experience across different contexts with how samples vary and the factors that affect this variability. This leads to the idea of accumulating and graphing multiple samples from the same population (of a given sample size), which leads to the more abstract idea of a sampling distribution. Different empirical sampling distributions are generated and observed, to see the predictable pattern that is a consequence of the Central Limit Theorem (CLT). Finally, students use the CLT to solve problems involving the likelihood of different values of sample means. Believing that it is more intuitively accessible to students, we begin the study of sampling with proportions and then move to sample means.

Lesson 1: Sampling from a Population

In this lesson, students make and test conjectures about sample proportions of orange-colored candies. They take physical samples from a population of colored candies (Reese's Pieces) and construct distributions of sample proportions. Students then use a Web applet to generate a larger number of samples of candies, allowing them to examine the distribution of sample proportions for different sample sizes. Students map the simulation of sample proportions to the *Simulation of Samples (SOS)* Model (Fig. 12.1), a visual scheme that distinguishes between the population, the samples, and the distribution of sample statistics (see also Chapter 6 where this model is first introduced). Student learning goals for this lesson include:

1. Understand variability between samples (how samples vary).
2. Build and describe distributions of sample statistics (in this case, proportions).
3. Understand the effect of sample size on: how well a sample resembles a population, and the variability of the distribution of sample statistics.
4. Understand what changes (samples and sample statistics) and what stays the same (population and parameters).
5. Understand and distinguish between the population, the samples, and the distribution of sample statistics.

Description of the Lesson

The lesson begins with a discussion of questions relating to what information a small sample can provide about a population. Students discuss their opinions about whether a small sample of Reese's Pieces can provide a good estimate of the proportion of orange Reese's Pieces candies produced by Hershey Company, and whether or not they would be surprised if they found only 5 Orange Reese's Pieces in a cup of 25 candies. These questions lead to a general discussion of sampling that reviews previous material covered in the introductory course, such as: What is a sample? Why sample? What do we do with samples? How should we sample? What is a good sample?

Students are then guided through the *Reese's Pieces* activity. They are first asked to guess the proportion of each color candy in a bag of Reese's Pieces and predict the number of orange candies that they would expect in 10 samples of 25 candies. Next, each student is given a small paper cup of Reese's Pieces candies, and is instructed to count out 25 without paying attention to the color. Then they count the number of orange candies and find the proportion of orange candies for their sample (of 25).

These proportions are collected and graphed and the class is asked to describe the graph in terms of shape, center, and spread. Note: this is not an actual sampling distribution because it does not consist of all possible samples from the population, but it is a *distribution of sample means* (and can serve as an approximation to the sampling distribution). This is an important distinction to make when generating or simulating sample data.

Students are asked to consider what they know and do not know: they do know the sample statistics but do not know the population parameter. They are asked to consider what is fixed and what changes (i.e., the sample statistics change from sample to sample; the population proportion stays the same regardless of the sample). Attention is drawn to the variability of the sample statistics, and students refer back to their predicted sample statistics for 10 samples earlier in the activity. Finally, students are asked to produce an estimate of the proportion of orange candies in the population of all Reese's Pieces candies.

Next, students use a Web applet from *RossmanChance* Website that has a picture of a large container of Reese's Pieces and allows them to draw multiple samples of any size. They use the proportion of orange candies given at the Website (.45), and then draw samples size 25 and watch as distributions of samples proportions are visually created (see Fig. 12.3 above). They are asked to consider what kinds of sample statistics they might expect if there were 10 candies in each sample instead of 25, or 100 candies in each sample. It is easy to change the sample size in the applet and simulate data to see how it affects the variability of the sample proportions for different sample sizes. Students are asked to complete a blank copy of the *SOS Model* for the simulation of Reese's Pieces they completed.

After completing the activity using the applet, a wrap-up discussion is used to draw students' attention to the important aspects of the concept they have just seen demonstrated. Students are asked to distinguish between how samples vary from each other, and variability of data *within* one sample of data. The Law of Large Numbers is revisited as it relates to the fact that larger samples better represent the population from which they were sampled. There can be a discussion of how big a sample needs to be to represent a population. (Students mistakenly think a sample has to be a certain percentage of the population, so a random sample of 1000 is not big enough to represent a population of one million.) Finally, students discuss the use of a model to simulate data, and the value of simulation in allowing us to determine if a sample value is surprising (e.g., 5 orange candies in a cup of 25 candies).

Lesson 2: Generating Sampling Distributions

In this lesson, students first speculate about the distribution of normal body temperatures and then contrast typical and potentially unusual temperatures for an individual person with a typical and potentially unusual values of means for samples of people. Students contrast the variability of individual values with the variability of sample means, and discover the impact on variability for different sample sizes. Web applets are used to simulate samples and distributions of sample means from two additional populations, revealing a predictable pattern as they generate sample means for increasingly large sample sizes, despite the differing population shapes. At the end of the lesson, students discuss how to determine if one value of a sample statistic is surprising or unusual, a precursor to formal statistical inference. Student learning goals for this lesson include:

1. Be able to generate sampling distributions for the sample mean, for different populations shapes.
2. Observe a predictable pattern (more normal, narrower, centered on the population mean) as the sample size increases.
3. Be able to distinguish between the population distribution, sample distribution, and the distribution of sample means.
4. Use the *Simulation of Samples (SOS)* Model to explain the process of creating sampling distributions.

Description of the Lesson

Class discussion begins with discussion about what is meant by a "normal" body temperature (e.g., healthy), what might be an unusual "normal" body temperature for someone who is not sick, and at what point they would consider someone sick as their temperature is too far beyond the expected range of natural variability. Students are also asked to consider and discuss at what point they would consider the mean temperature for a group of students to be unusual or suspicious. They are asked to consider any one person and how likely a person would be to have a body temperature of 98.6°F (37°C). This leads to the first activity (*Body Temperatures*), where students make conjectures about what they would expect to see in a distribution of normal body temperatures for the population of healthy adults in terms of shape, center, and spread. They draw graphs of what they would expect this distribution to look like.

Then, students are asked to consider where some particular values would be located on their graphs: 99.0, 98.0, and 97.2°F (37.2, 36.7, and 36.2°C). They are asked to think about whether or not any of these values would be considered unusual or surprising. They are asked to consider a temperature of 96°F (35.6°C). Most students will most likely say that this is a surprising body temperature for a normal adult. Then they are asked to think about how to determine what a surprising or unusual value is.

This question leads to a discussion about how much statistical work has to do with looking at samples from populations and determining if a particular result is surprising or not, given particular hypotheses. If a result is surprising, students are told, we often call that a *statistically significant* result. Students are told that they will be looking at what leads to a sample or research results that are "statistically significant."

Students are guided to begin by looking at z scores, which are computed using population values for μ and σ. They are given these values and are asked to find z scores for the body temperatures considered earlier: 99.0, 98.0, and 97.2°F. Students are next asked to think about z scores that would be more or less surprising, if they represent an individual's body temperature. The next part of the activity is designed to help move students from considering individual values to considering averages.

As the discussion continues, students are asked to consider a random sample of 10 students who have their temperatures taken using the same method. They are asked whether they would predict if the *average temperature* for this sample would be exactly 98.6°F or close to this value, and to give their reasons. They are then asked what about another sample of 10 different students and if that sample would have the same mean as the first sample of 10 or whether it would produce a different mean. Students are asked to think about what would cause differences in sample means.

Working in pairs, students are asked to write down plausible values for five students' body temperatures and then to write down plausible values of sample mean body temperatures for five random samples of ten college students ($n = 10$). Students are asked to think about and describe how they would expect these two sets of temperatures to compare and which set would have more variability and why. They are asked to think about what would happen if they took five random samples of 100 people, what the means of those samples of 100 people might be. *Fathom* is used to simulate a population of body temperatures with the given values of μ and σ, and samples sizes are varied to allow students to compare their estimates with simulated data (see Fig. 12.5 for a graph of the population).

In the next activity (*Sampling Words*), students return to the Web applet at *RossmanChance* Website that visually illustrates sampling words from the Gettysburg Address, as shown in Fig. 12.4 (Note, Lesson 3 in Chapter 6 uses this applet to take random samples). Students use the applet to sample five words and list the word lengths, then discuss how they vary. They find the mean and standard deviation for their set of five word lengths and consider what would be an unusual value (word length).
Students are asked to compare the mean and standard deviation for their sample of five word lengths to the given values of μ and σ for the population. They next begin taking samples of five words at a time, using the software, and examine the

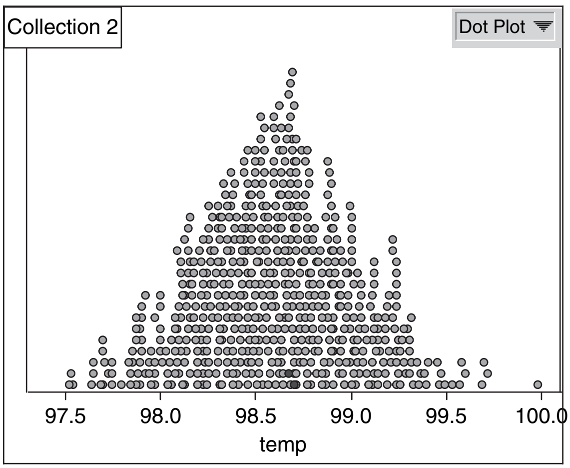

Fig. 12.5 *Fathom* simulation of human body temperatures

sample means. They are asked to consider what would be unusual values of means for samples of this size and to consider criteria to use in determining what a surprising value would be. Next, they generate a distribution of 500 sample means for samples of size 5 and use it to determine where their hypothetical "unusual" value would be. They are instructed to change the sample size to 10 and determine what differences will result for a new distribution of 500 sample means. They then repeat this for sample size of 20 and 50.

A whole class discussion follows where students are asked to compare these distributions in terms of shape, center, and spread, to discuss what is different about them, and how their unusual value fits on each. They are then guided to compute z scores for this value and compare them.

In the third activity (*Sampling Pennies*), students take simulated samples from a third population (ages of pennies) to see if they notice the same predictable pattern. They use the Sampling Pennies applet at RossmanChance.com, which shows a skewed distribution of dates on pennies from which they sample and use to generate distributions of sample means. Students vary parameters on the applet starting with small sample sizes and increasing to large sample sizes. Each time, they take 500 samples and look at the distribution of sample statistics (means or proportions) to describe a predictable pattern.

In a wrap-up discussion, students refer back to the beginning activity that looked at the variability of individual body temperatures from a normal (distribution) population and note that:

- Some temperatures are more likely than others.
- To see if a value is more or less likely (or surprising) we needed to look at their relative position in the distribution.
- Number of standard deviations above and below the mean (z scores) can tell us if something is unlikely or unusual. This can also be done for sample means after we learn the appropriate way to find z scores for a sample mean.
- Sample means vary too, but they tend to vary less than individual values.
- Means from smaller samples vary more than from large ones.
- There was a predictable pattern when we took larger sample sizes and plotted their means. The predictable pattern was: Symmetric, bell shaped (even when the populations were skewed), centered on μ, and smaller spread (less variability).
- To determine if a sample mean is unusual or surprising, we need to compare it to many other sample means, of the same size, from the same population. This is a distribution of sample means.
- If a distribution of sample means is normal, we can use z scores to help us see if values are unlikely or surprising.

Students also discuss how each of the simulations is modeled by the *SOS model*. The next lesson helps us determine when it is appropriate to assume that a distribution of sample means is normal so that we may use z scores to see if values are unlikely or surprising.

Lesson 3: Describing the Predictable Pattern – The Central Limit Theorem

This lesson moves students from noticing a predictable pattern when they generate distributions of sample statistics to describing that pattern using mathematical theory (i.e., the Central Limit Theorem, CLT). Students investigate the impact of sample size and population shape on the shape of the sampling distribution, and distinguish between sample size and number of samples. Students then apply (when appropriate) the Empirical Rule (68, 95, 99.7% within 1, 2, and 3 standard deviations from the mean) to estimate the probability of sample means occurring in a specific interval. Student learning goals for this lesson include:

1. Discover the Central Limit Theorem by examining the characteristics of sampling distributions.
2. See that the Central Limit Theorem describes the predictable pattern that students have seen when generating empirical distributions of sample means.
3. Describe this pattern in terms of shape, center, and spread; contrasting these characteristics of the population to the distribution of sample means.
4. See how this pattern allows us to estimate percentages or probabilities for a particular sample statistic, using the Normal Distribution as a model.
5. Understand how the *SOS Model* represents the Central Limit Theorem.
6. Understand how we determine if a result is surprising.

Description of the Lesson

The lesson begins with a review discussion of the previous lessons on samples and distributions of sample statistics. This includes revisiting ideas of variability in data values and in sample means, how to determine if a particular value or sample statistics is unlikely (surprising) using z scores, that statistics from small samples have more variability than those from large samples, and there is a predictable pattern when plotting the means of many large random samples of a given sample size from a population.

Students are asked to compare z scores for individual words to z-scores for means of word lengths (from the Gettysburg Address) and use z-scores as a yardstick for distance from a mean. They are asked to consider which is important: the number of samples or sample size, and what these terms mean. This is because students often confuse number of samples (an arbitrary number when performing simulations) with sample size. The point is made that we often used 500 samples because the resulting distribution of sample means is very close to what we expect the sampling distribution to look like in terms of shape, center, and spread.

The focus of the main activity of this lesson (*Central Limit Theorem* activity) is to examine and describe in detail the predictable pattern revealed with different populations, parameters, and sample sizes. The students are asked to make predictions, generate simulations to test them, and then evaluate their predictions. Their predictions are about what distributions of sample means will look like as they change the population shape and sample size. Each time, they take 500 samples, which provides a picture that is very close to what they would get if they took all possible samples.

This activity can be done using different simulation tools, but we prefer to use *Sampling SIM* software (delMas, 2001a) and a set of stickers that can be printed using a template available at the Website. These stickers show three columns of distributions, each headed by a specified population (see Fig. 12.6). Students are

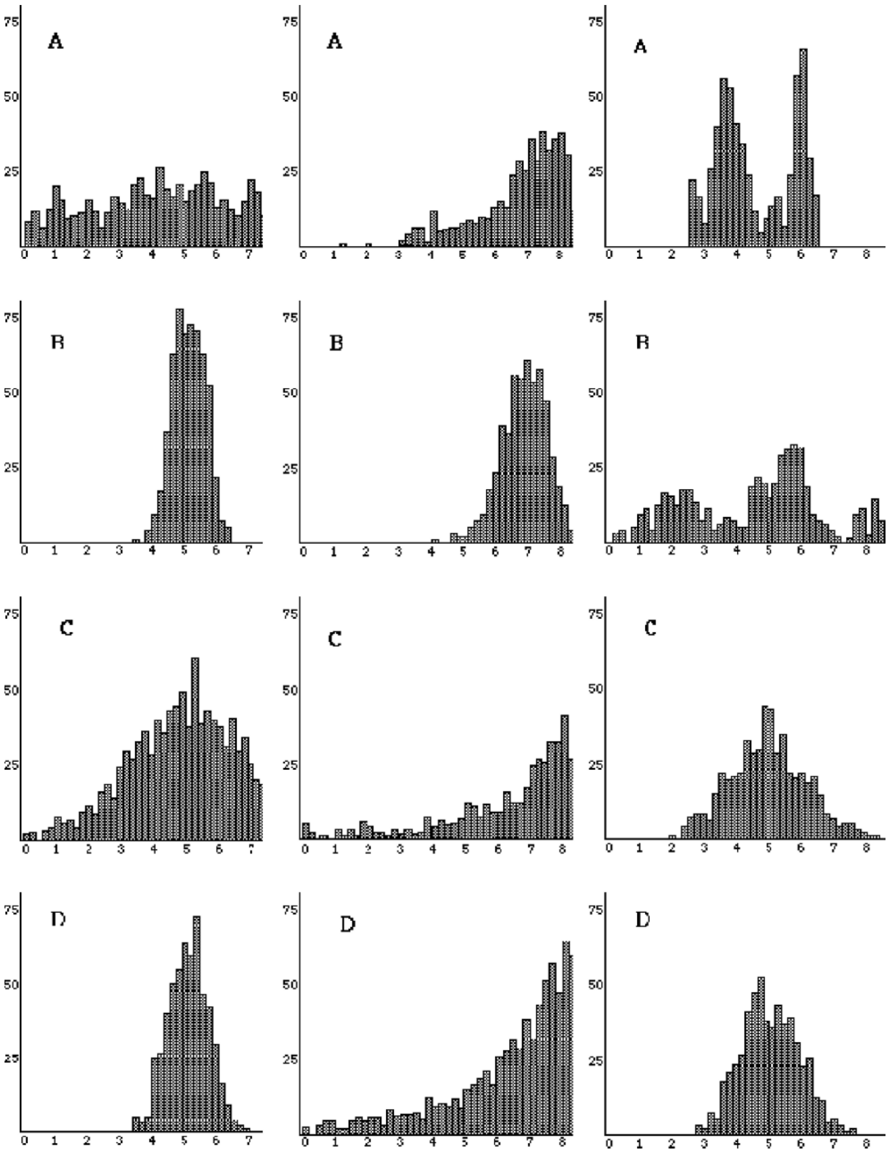

Fig. 12.6 Stickers for the *Central Limit Theorem Activity*

asked to consider three populations, one at a time, and to predict which of five distributions shown on the stickers will correspond to a distribution of 500 sample means for a particular sample size. They test their conjecture by running the simulation, then affix the sticker that matches that simulation result in a "Scrapbook." When the activity is finished, students have striking visual record that allows them to see and describe what happens when different sized samples are taken from a

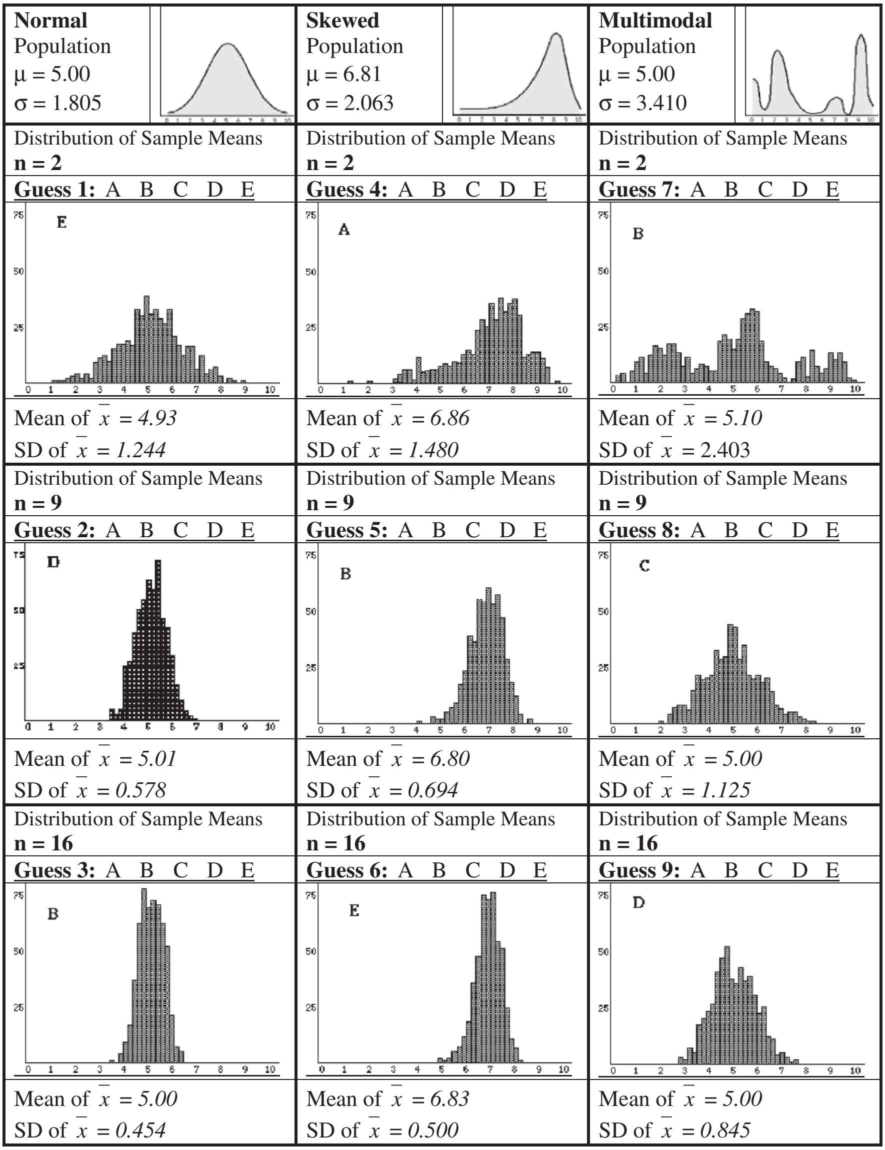

Fig. 12.7 Sample of a sampling scrapbook at the end of the *Central Limit Theorem Activity*

Normal, a skewed, and a trimodal distribution (see Fig. 12.7). If it is not possible to use stickers, copies of the pages shown in Figs 12.6 and 12.7 can be used instead.

As students work through this activity, they see that the predictable pattern is true regardless of population shape, and that (surprisingly) as the sample size increases, the distributions look more normal and less like the population (for non Normal populations). They are encouraged to develop a theory of how to evaluate whether a sample mean is surprising or not. Students are then guided to use the Central Limit Theorem to apply the Empirical Rule (68, 95, and 99.7% of data within 1, 2, and 3 standard deviations in a Normal Distribution) to sampling distributions that appear to be normal.

In a wrap-up discussion, students are asked to contrast the Law of Large Numbers to the Central Limit Theorem and also to discuss how they are connected. For example, when a large sample size is generated, individual samples better represent the population (Law of Large Numbers) and their sample statistics are closer to the population parameters. When graphing 500 of these sample means, they will cluster closer to μ resulting in less variability (smaller standard error) and a symmetric bell shape. Students are asked to distinguish between: populations, samples, and sampling distributions. They are asked to discuss what is similar and what is different and why. Finally, they work in small groups to respond to questions such as: How can we describe the sampling distribution of sample means without running simulations? For random samples size 100 from a population of human body temperatures, what would be the shape, center, and spread of the distribution of sample means? And which would be more likely to be closest to 98.6°F: A mean temperature based on 9 people or a mean temperature based on 25 people? Finally, students are asked to use the *SOS Model* to represent the Central Limit Theorem.

Summary

We believe that ideas of samples and sampling distributions should be introduced early in a statistics course and that by the time students reach the formal study of sampling distributions, they have already generated and examined different distributions of sample statistics while making informal inferences. Even given that background, we believe that the full set of activities described in this chapter are needed in order for students to understand and appreciate the Central Limit Theorem. Then, when the formal study of inference is introduced we encourage the revisiting of ideas of sampling distribution and the reference to the *Simulation of Samples (SOS)* Model (Fig. 12.1) so that students can see the role of a sampling distribution in making statistical inferences.