James R. Kett

Teaching Sampling Distributions Using Autograph

The sampling distribution of a statistic shows the values of that statistic in all possible samples of a given size from the same population. Sampling distributions are fundamental to the study of statistics because they are the basis for confidence intervals and hypothesis testing. Students often struggle with these distributions, so teachers look for innovative ways to help students understand them. Autograph provides an excellent opportunity to illustrate sampling distributions and to demonstrate the central limit theorem. (For an introduction to Autograph, see “Technology Tips” in *MT* May 2010, vol. 103, no. 9, pp. 689–92.)

Let’s start with a simplified population. Consider a die with the number 2 on three faces and the number 6 on the other three faces. **Figure 1** shows a histogram of the six possibilities. To create this histogram, open Autograph and choose **Advanced Level** when prompted. Choose **File** and then **New 1D Statistics Page** (we could also use the corresponding icon located at the extreme left of the menu bar). Select **Data** and then **Enter Grouped Data**. In the dialogue box that appears, set the **Class Intervals** section to **Integer Data**, the **Frequencies** to **Use Raw Data**, and the **Data Type** to **Discrete** (see **fig. 2**). Choose **Edit** in the frequencies section; enter 2, 2, 2, 6, 6, 6. After click-

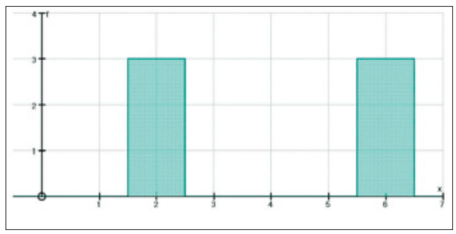


Fig. 1 The histogram of the population distribution for the simplified die shows three outcomes each of 2 and 6.

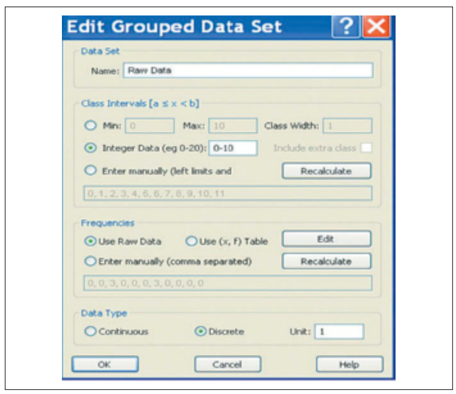


Fig. 2 The settings to enter the data are now correct.

Technology Tips, which provides a forum for innovative uses of technology in the teaching and learning of mathematics, appears seven times each year in *Mathematics Teacher*. Manuscripts for the department should be submitted via http://mt.msubmit.net. For more background information on the department and guidelines for submission, visit http://www.nctm.org/publications/content.aspx?id=10440#tech.

Edited by ***Larry Ottman***
ottmanl@hhsd.k12.nj.us
Haddon Heights High School
Haddon Heights, NJ

Russell Gordon
russellwgordon@me.com
The Bishop Strachan School
Toronto, ON, Canada

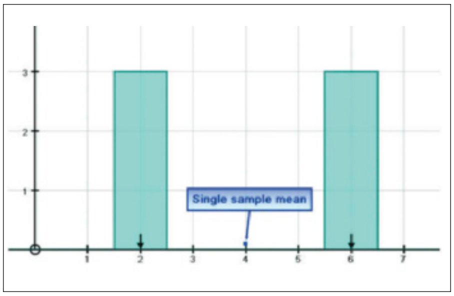


Fig. 3 Autograph selects one 2 and one 6 with a mean of 4. (The text box has been added.)

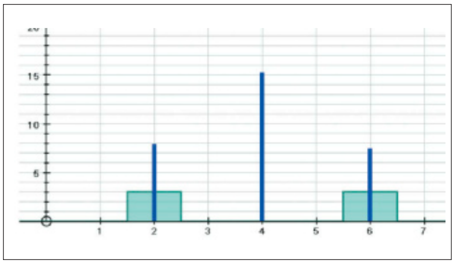


Fig. 4 After repeated sampling, Autograph shows an approximate sampling distribution of the sample mean with sample size 2.

ing on **OK** twice, select the **Histogram** icon and choose **OK**.

If we toss a pair of these dice and record the average of the two outcomes, what do we expect to get? There are only three possibilities: The mean can be 2, 4, or 6. To model this with Autograph once the histogram is drawn, simply right-click somewhere on the screen, select **Sample Means**, set the sample size to 2, and click on **Single Sample**. **Figure 3** shows the two numbers selected and the sample mean. Thus, we have the first step in forming a distribution of sample means.

Next, click on **Sample** three times. Doing so will add 300 sample means and graph the result (see **fig. 4**). A good class discussion should result when students are asked why 4 appears more often than 2 or 6. (Note: The vertical spacing for the sample means is 0.1, so the graph indicated in **fig. 4** shows about 80 2s.) To resize the graph, select **Axes** and then **Edit Axes** and change the **y-max** to 20.

The next step is to investigate what would happen if we change the sample size to 3 and ask students to describe the sampling distribution of the sample mean. Four different means could result from sampling from this distribution. The choices are 2, 2, 2 (sample mean = 2); 2, 2, 6 (sample mean ≈ 3.3); 2, 6, 6 (sample mean ≈ 4.7); or 6, 6, 6 (sample

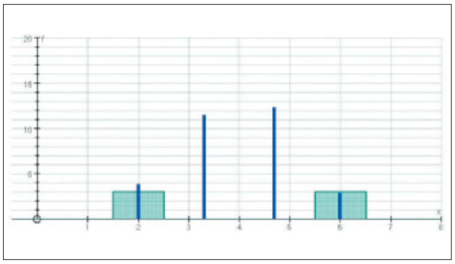


Fig. 5 An approximate sampling distribution with sample size 3 will look like this.

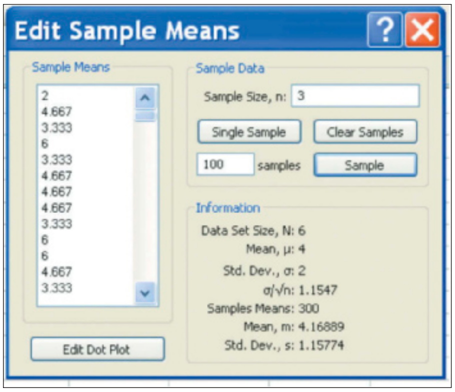


Fig. 6 The **Edit Sample Means** dialogue box shows the comparison between the theoretical values and the Autograph results.

mean = 6). To show the sampling distribution of the sample mean with Autograph, double-click on **Sample Means** in the status bar at the bottom right of the page, click on **Clear Samples**, change the sample size to 3, and select 100 samples. After 300 samples, the graph should appear similar to that shown in **figure 5**.

A fundamental result is that the mean of the sampling distribution of a sample mean is the same as the population mean. Also, the standard deviation of the sampling distribution is the population standard deviation divided by the square root of the sample size:

$$\mu_{\bar{x}} = \mu_x, \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

We can use Autograph to explore this idea by double-clicking on **Sample Means** in the status bar.

The dialogue box (see **fig. 6**) gives the mean and the standard deviation of the population (μ and σ). It also gives the theoretical standard deviation of the sampling distribution of the sample mean. At the bottom of the box is the mean (**m**) and standard deviation (**s**) of the simulated sampling distribution of

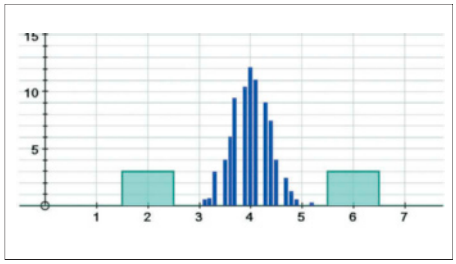


Fig. 7 An approximate sampling distribution of the sample mean from Autograph with sample size 30 shows a distribution close to normal.

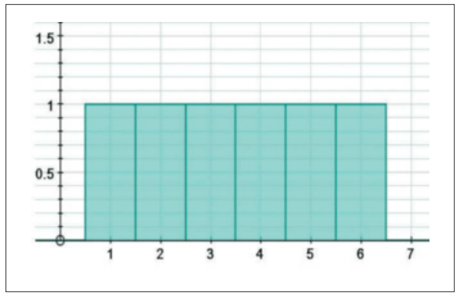


Fig. 8 When we use a fair die, the histogram shows six outcomes.

the sample mean. After discussing with students why $\sigma/\sqrt{n}$ and s do not match exactly, teachers can lead students to discover that as the number of samples taken is increased, the disparity between the two values disappears. Simply have students leave the **Edit Sample Means** box open and repeatedly press **Sample**.

Now that we have established the importance of taking many simulated samples, we can turn to the central limit theorem. This theorem describes the shape of the sampling distribution of the sample mean. As the sample size increases, the shape of the sampling distribution gets more and more normal. To demonstrate this with Autograph, double-click on **Sample Means**, change the sample size, clear the previous samples, and select **Sample**. **Figure 7** shows the results for sample size 30. Use **Edit Axes** to adjust the appearance of the graph.

We started with a simple population of only the numbers 2 and 6. The next step is to repeat the investigation with a different population. If we toss a fair die, the population histogram consists of the numbers from 1 to 6 (see **fig. 8**). Toss the die 10 times and record the mean. Repeat this process many times to approximate the sampling distribution of sample means. In the Autograph page, double-click on **Raw Data** at the bottom

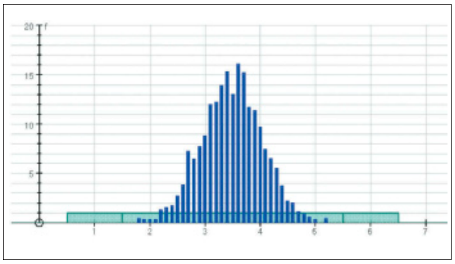


Fig. 9 Autograph constructs a histogram showing an approximate sampling distribution of the sample mean with sample size 10.

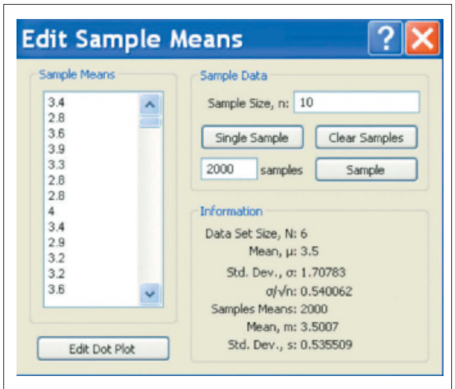


Fig. 10 Results from a sampling distribution of sample means should be very close to the theoretical values.

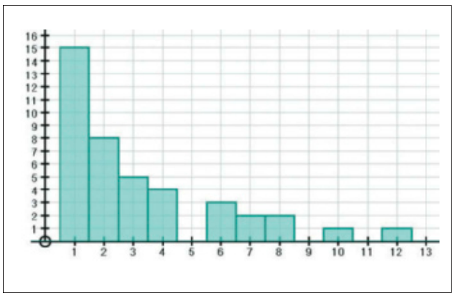


Fig. 11 What happens if we start from a strongly skewed population such as this one?

left of the Autograph page and enter the new values. Adjust the graph to match the histogram in **figure 8**. Repeat the process by changing the **Sample Means** settings, matching those mentioned above, to obtain a sampling distribution of sample means similar to the one in **figure 9**. Recall that the vertical spacing for the dot plot is 0.1.

The mean μ and standard deviation σ of the population are 3.5 and 1.7, respectively. Using the formula to calculate the theoretical standard deviation of the sampling distribution of the sample mean gives about 0.54. In the Autograph distribution, the mean (**m**) and standard deviation (**s**) should be very close to 3.5

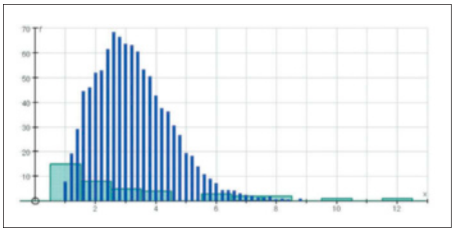


Fig. 12 An approximate sampling distribution of the sample mean of sample size 5 is still skewed.

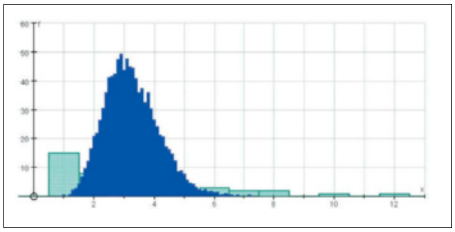


Fig. 13 An approximate sampling distribution of the sample mean of sample size 10 is also still skewed right but not nearly so much.

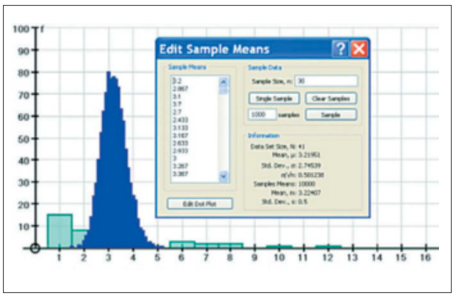


Fig. 14 An approximate sampling distribution of the sample mean for sample size 30 appears more symmetric. The **Edit Sample Means** dialogue box is also displayed.

and 0.54. Results may vary slightly, but our simulation was off by less than 0.01 (see **fig. 10**).

To complete this investigation, use Autograph to verify that the sampling distribution of the sample mean for a strongly skewed distribution becomes close to normal as the sample size increases. Select **Edit Raw Data** to create a skewed distribution. When editing the data, make sure the bin width is 1 by changing **Class Width** in the section **Class Intervals**. **Figure 11** shows the graph of one possible skewed population.

Now adjust **Sample Means** to show 10,000 trials for samples of size 5 and 10 (see **figs. 12** and **13**). Both are still skewed right, but the graph in **figure 13** is less so. The general rule taught to most beginning statistics students is that the sample size should be at least 30 to

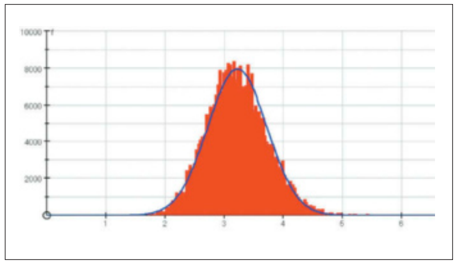


Fig. 15 A normal curve can be fit to the distribution of sample means.

use a normal model. Change the **Sample Means** settings to collect 10,000 samples of size 30. **Figure 14** shows the resulting sampling distribution of the sample mean. The results appear more bell-shaped, with noticeably less variability due to the larger sample size.

One possible way to have students investigate the normality of this distribution is to use the empirical rule. The empirical rule states that, in a normal distribution, about 68% of the values fall within 1 standard deviation of the mean; about 95% of the values fall within 2 standard deviations; and about 99.7% fall within 3 standard deviations. Copy and paste the 10,000 sample means into **New 1D Statistics Page**. To do so, **Select** (**Ctrl A**) and **Copy** (**Ctrl C**) all 10,000 numbers in the **Edit Sample Means** box. After opening **New 1D Statistics Page**, select **Enter Raw Data** in the **Data** menu and then paste (**Ctrl V**) the 10,000 sample means. Accept the data by clicking on **OK**.

Right-click on the graph window and select **Dot Plot** (or choose the icon from the toolbar). Before clicking on **OK**, change the horizontal spacing to 0.01 and the vertical spacing to 30. Right-click and select **Enter Probability Distribution** and check the box for **Normal**. After clicking on **OK**, choose **Fit to Data**, and then accept by clicking on **OK**. Autograph creates a normal curve using the mean and the standard deviation of the sample means. The graph should be similar to that in **figure 15**.

The graph's appearance alone should provide support for the approximate normality of this distribution. To help students add to the body of evidence, we can use the empirical rule. We need Autograph to produce a histogram of the sample means.

First, delete the normal curve and the dot plot by right-clicking and then

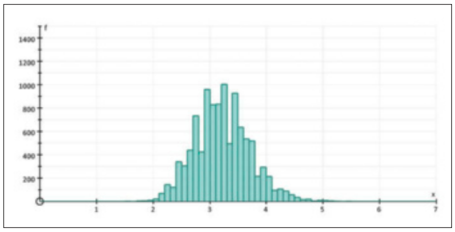


Fig 16. Autograph displays a histogram of the sample means.

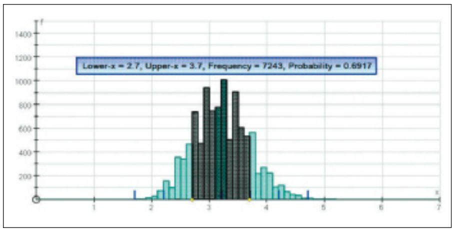


Fig. 17 The highlighted section shows the percentage of area within 1 standard deviation of the mean. (The text box has been added.)

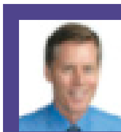
selecting **Delete Object**. Before drawing the histogram, we must group the data, so right-click and select **Group Data Set**, changing **Class Width** to 0.1. Click on the **Histogram** button to get a graph similar to that shown in **figure 16**.

Figure 14 shows that the mean is about 3.2 with a standard deviation of 0.5. On the Autograph screen, if we right-click and select **Mean ±3 Std Devs**, we will see marks on the x-axis that correspond to the points that we will use to check how well the distribution fits the empirical rule.

Select **Histogram** in the status bar on the bottom right of the page and choose **Probability by Area** in the **Object** menu. Change the **lower-x** value to 2.7, the **upper-x** value to 3.7, and observe the area within one standard deviation of the mean as a decimal. **Figure 17** shows the area to be about 69% of the total area under the curve. To check the second part of the empirical rule, double-click on **Probability by Area** in the status bar and change the limits to 2.2 and 4.2. In our example, we got the area to be 96% of the area under the curve. Finally, change the limits to 1.7 and 4.7; these produce an area of 99.7% of the area under the curve.

This demonstration provides some additional evidence that the distribution of sample means is close to normal when the sample size is 30.

Students should continue this investigation with other populations until they have a good understanding of the central limit theorem. With this foundation, students should find hypothesis testing much easier to understand.



JAMES R. KETT, dr.jimkett@gmail.com, has extensive experience teaching mathematics both in the United States and in Singapore. In addition to speaking at conferences and leading workshops, he has written two books on teaching and learning mathematics using Autograph.

SURFING NOTE

With a graphing calculator, the user enters an equation, and the calculator produces a graph. With the Inverse Graphing Calculator (www.xamuel.com/inverse-graphing-calculator.php), the user can enter a text string, and the calculator will output the equations that produce that text string. In addition to obtaining equations for their names or favorite expressions, such as "I love math," students can study the patterns in the results for similar starting points. For example, compare the equation for A (see **fig. a**) with the equation for AA (see **fig. b**). Many letters will be needed to explain how the first equation can be transformed to produce the second and then to predict the equation that would spell AAA. This site gets an A+. The "Technology Tips" editors are grateful to Ron Lancaster for providing this surfing note.

$$(y-2x+2)^2(y+2x-10)^2\left((y-3)^2+\left|x-\frac{5}{2}\right|+\left|x-\frac{7}{2}\right|-1\right)^2$$
$$+\left(y^2-6y+8+\sqrt{y^4-12y^3+52y^2-96y+64}\right)^2=0$$

(a)

$$(y-2x+2)^2(y+2x-10)^2\left((y-3)^2+\left|x-\frac{5}{2}\right|+\left|x-\frac{7}{2}\right|-1\right)^2$$
$$\cdot(y-2x+8)^2(y+2x-16)^2\left((y-3)^2+\left|x-\frac{11}{2}\right|+\left|x-\frac{13}{2}\right|-1\right)^2$$
$$+\left(y^2-6y+8+\sqrt{y^4-12y^3+52y^2-96y+64}\right)^2=0$$

(b)