Who Sends the Most Text Messages?

Anna Bargagliotti
Loyola Marymount University
abargagl@lmu.edu

Jeanie Gibson
Hutchison School
jgibson@hutchisonschool.org

Project-SET
www.project-set.com

Published: January 2013

Overview of Lesson
This activity allows students to perform a hands-on investigation in which they explore various factors affecting the shape of the sampling distribution of the sample mean. Students will be divided into four groups, with each group being given a bag of numbered index cards. The numbers on the cards represent the number of text messages sent/received by students in a statistics class. Students will sample from these populations in order to learn about the behavior of the sample mean. Students will construct dot plots showing both the distribution of each population as well as construct dot plots for the sampling distributions of the sample means. They will discover that although the population distributions have distinctly different shapes, the distribution of the sample means have approximately the same shape. The distribution of the sample mean will follow a normal distribution when the sample size is large enough. This lesson provides an informal introduction to concepts surrounding the Central Limit Theorem.

GAISE Components
This activity follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) *Report.* The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level C activity.

Common Core State Standards for Mathematical Practice
4. Model with mathematics.
7. Look for and make use of structure.

Common Core State Standards Grade Level Content (High School)
S-ID. 1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
S-ID. 2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
S-ID. 3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
S-IC. 1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

NCTM Principles and Standards for School Mathematics
Data Analysis and Probability Standards for Grades 9-12

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:

- compute basic statistics and understand the distinction between a statistic and a parameter.

Select and use appropriate statistical methods to analyze data:

- for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.

Develop and evaluate inferences and predictions that are based on data:

- use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions;
- understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Prerequisites

Prior to completing this activity students should be able to identify the population and sample in a given situation involving random sampling. They should also understand that information from a sample is used to draw conclusions about the entire population. They should have a basic understanding of how to construct a dot plot and how to informally describe shape, center and spread. They should be able, with the aid of a calculator, to calculate the mean and standard deviation of a distribution.

Learning Targets

After completing the activity, students will have learned that the distribution of the sample mean follows a normal distribution for a sufficiently large sample size. Students will also thus have an understanding of the effect of sample size on the sampling distribution of the sample mean $\bar{X}$, and be introduced to concepts surrounding the Central Limit Theorem. They will be familiar with the process of taking repeated samples of the same size, and constructing dot plots for both the population and the sampling distributions of sample mean $\bar{X}$. Students will be able to provide quantitative descriptions of the variability of the sampling distributions they have constructed.

Time Required

The time required for this activity is roughly 90 minutes.

Materials Required

For this activity, students will need a pencil, paper, the activity sheet, and bags with cards labeled according to the following instructions (see sample cards to print and cut out in the linked .pdf files named: *Sally*, *Jamal*, *Ivy*, and *Tom*).

The following numbers will be written on 60 cards and inserted into a bag with the indicated name written on the outside. Index cards cut into fourths works well for this.

SALLY - Roughly Normal Distribution with Mean $\mu = 60$ and Standard Deviation $\sigma = 4.18$

Number	**# of Cards**
60	6
58, 59, 61, 62	5
56, 57, 63, 64	4
55, 65	3
53, 54, 66, 67	2
51, 52, 68, 69	1

JAMAL - Bimodal Distribution with Mean $\mu = 60$ and Standard Deviation $\sigma = 39.78$

Number	**# of Cards**
40, 80	1
5, 35, 85, 115	2
10, 110	3
15, 30, 90, 105	4
20, 25, 95, 100	7

IVY - Skewed Distribution with Mean $\mu = 74.25$ and Standard Deviation $\sigma = 31.36$

Number	**# of Cards**
5, 60	1
25, 65, 75, 80, 85, 105	2
10, 30, 40, 50	3
55, 70	4
90, 115	6
100	14

TOM - Scattered Distribution with Mean $\mu = 60$ and Standard Deviation $\sigma = 35.64$

Number	**# of Cards**
45	1
0, 15, 75, 80, 90	2
5, 50	3
85	4
10	6
70	8
40	11
105	14

Instructional Lesson Plan

Before beginning the activity, the teacher may wish to review the concepts of population, sample, population parameter and sample statistic, reinforcing student understanding of these foundational concepts for the lesson. If students have not had adequate practice constructing and analyzing dot plots, this process should also be reviewed.

The GAISE Statistical Problem-Solving Procedure

I. Formulate Question(s)

This activity will be driven by a teacher-directed question; however, students will have the opportunity to formulate their own question as well. The overall question of interest for the activity is: **Using the population text message data on four students; how can we determine which of the four students does the most text messaging?**

Students will be asked to formulate the specific statistical question that needs to be explored in order to answer the question of interest. Example student questions include: what is the average mean pictured in the approximation to the sampling distribution, what do the average means tell us about the comparison of the student text messaging habits, or what does the spread of the sampling distribution tell us about the comparison of the student text messaging habits?

II. Design and Implement a Plan to Collect the Data

Students will be divided into groups of four, with each group being given a bag containing a population of 60 cards with numbers on them representing the data collected by Sally, Jamal, Ivy, and Tom in a statistics class. Each card will have a number on it representing the number of text messages sent/received on one of the 60 days the data was collected. Text message data is thus previously collected for the students. Students will perform their own data collection in this activity by random sampling from the bag population with 20 samples each of varying sample sizes $n = 4$, $n = 10$, and $n = 30$.

III. Analyze the Data

Students will construct dot plots showing the sampling distribution for the sample mean $\bar{X}$ for each sample size (see Figures 1 and 2 below). Students will then construct a dot plot for the 60 cards in the total population (see Figures 3 through 6 below). The mean and standard deviation will be calculated for each dot plot. Students will then be asked to compare the sampling distributions for the sample mean $\bar{X}$ for sample sizes $n = 4$, $n = 10$, and $n = 30$ in order to see the effect of sample size on the distribution. They will then be asked to observe the four population distributions and their corresponding sampling distributions for sample size $n = 30$. Students should observe that: (1) the spread of the sampling distribution decreases with increased sample size, and (2) even when the population distribution is not normal, the sampling distribution of the sample mean $\bar{X}$ is approximately normal.

Examples of Sampling Distribution Dot Plots

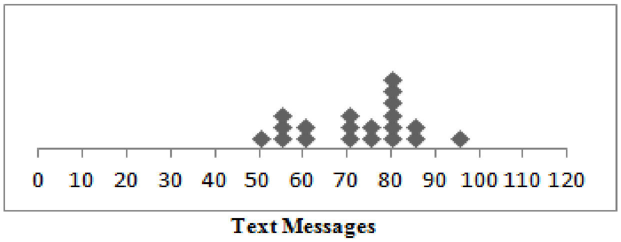

Figure 1. Sample dot plot showing sampling distribution for sample mean for skewed distribution (Ivy) – 20 samples of sample size $n = 4$.

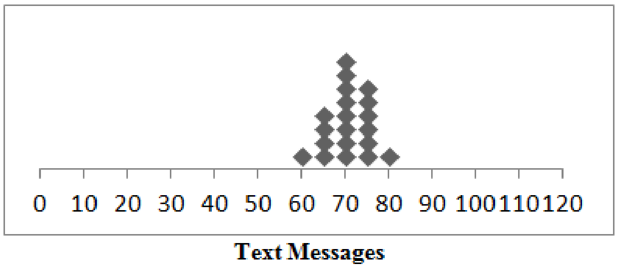

Figure 2. Sample dot plot showing sampling distribution for sample mean for skewed distribution (Ivy) – 20 samples of sample size $n = 30$.

Population Distribution Dot Plots

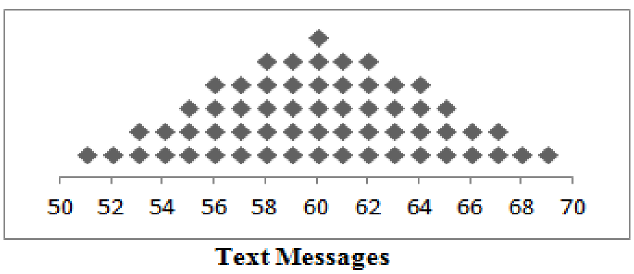

Figure 3. Dot plot showing population distribution for Sally.

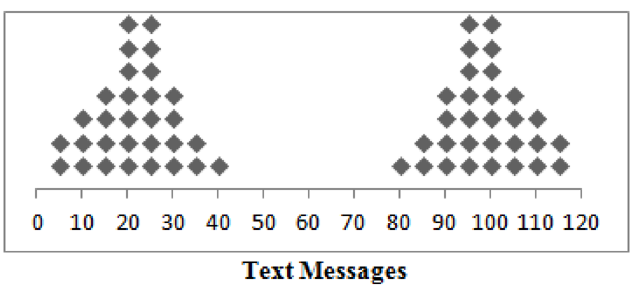

Figure 4. Dot plot showing population distribution for Jamal.

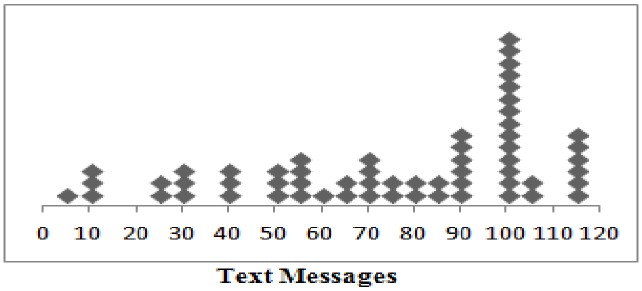

Figure 5. Dot plot showing population distribution for Ivy.

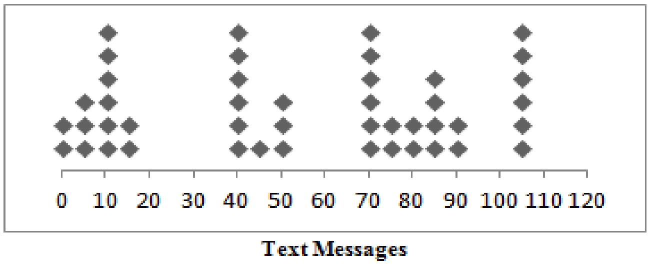

Figure 6. Dot plot showing population distribution for Tom.

IV. Interpret the Results

During the activity students were asked to determine which of four students in a statistics class sent/received the most text messages in a 60-day period. After students complete the activity in groups, teachers will complete the lesson by leading a class discussion to interpret the findings. The guiding questions for the whole-class discussion, along with possible desired student responses, are as follows:

1. What happens to the shape of the sampling distributions as the size of the sample varies?

   Students should conclude that for each of the sampling distributions, the shape becomes closer to being normal as the size of the sample increases.

2. How is the shape of the sampling distribution related to the shape of the population from which the samples were drawn?

   As the size of the sample increases, the shape of the sampling distribution resembles the shape of a normal distribution and no longer resembles the shape of the population distribution.

3. How do the population distributions compare to each other?

   Students should make the following comparisons of the population distributions:
   - SHAPE: The population distribution for Sally is approximately normal, Jamal's is bimodal, Ivy's is skewed to the left, and Tom's is scattered.
   - CENTER: Sally's, Jamal's and Tom's population distributions all have mean $\mu = 60$, while the mean of Ivy's population distribution is considerably higher at $\mu = 74.25$.
   - SPREAD: The standard deviation of the population distribution for Sally is $\sigma = 4.18$, which is much smaller than the standard deviations of the other three, which range from $\sigma = 31.36$ for Ivy to $\sigma = 39.78$ for Jamal.

4. How do the sampling distributions compare to each other?
   As the sample size increases, the shapes of all four sampling distributions become approximately normal. It should also be noted that the means of the sampling distributions for Sally, Jamal and Tom are approximately $\mu = 60$, while the mean of the sampling distribution for Ivy is considerably higher at approximately $\mu = 74$. Students should observe that the spreads of the sampling distributions for Jamal, Ivy and Tom are greater than the spread of the sampling distribution for Sally.

5. What are the means of the sampling distributions?
   The means of the sampling distributions for Sally, Jamal and Tom should be approximately $\mu = 60$, while the mean of the sampling distribution for Ivy is approximately $\mu = 74.25$.

6. How would you describe the spread of the sampling distributions?
   The spreads of all four sampling distributions decrease as the sample size increases.

7. How can you use this information to make a decision about which student texted more?
   If students decide to use the mean as a way of comparison, then since the mean number of text messages sent/received by Ivy was more than the mean number for Sally, Jamal and Tom, then Ivy was the student who sent/received the most text messages. Students can observe this result by observing the sampling distributions. From the sampling distributions, students can discuss how likely it is that Sally, Jamal, Tom, and Ivy text message above a certain amount in a 60-day period.
   Students could also observe the population distributions and note that Ivy's distribution had a very large spread. This may suggest that Ivy had some peak activity of text messages on a few days but was not as consistent in her use as for example, Sally. Students may provide potential descriptions of why the population text message data looks the way it does for the four students. For example, Jamal could have gone on vacation for a part of the 60 days where he had no phone usage. Tom's distribution may suggest that he has some specific weekly patterns in his text usage.

Assessment

1. Jasmyn and Matthew are studying together for their upcoming statistics test. Matthew is attempting to explain the Central Limit Theorem to Jasmyn and says, "When you take larger and larger samples from a population, the dot plot of the sample values looks more and more normal." Did Matthew give a good explanation of the Central Limit Theorem? Why or why not?

2. A study of high school seniors' study habits found that the time (in hours) that seniors use to study each week follows a strongly skewed distribution with a mean of 5.2 hours and a standard deviation of 3.4 hours. What is the *shape* of the sampling distribution of the mean $\bar{X}$ for samples of 55 randomly selected high school seniors if 55 is considered to be a large sample? Justify your answer.

3. What does the Central Limit Theorem say about the shape of the sampling distribution of $\bar{X}$?

4. What does the Central Limit Theorem allow us to do?

Answers

1. Matthew did not give a good explanation. The dot plot of the sample values will look like the population distribution. The Central Limit Theorem says that as the sample size increases, the dot plot of the **sample means** would become more and more normal.

2. The sampling distribution of the mean $\overline{X}$ would be approximately normal in shape. The Central Limit Theorem states that when the sample size is large, the sampling distribution of the sample mean $\overline{X}$ is approximately normal.

3. The Central Limit Theorem says when we draw a simple random sample of size *n* from **ANY** shaped population with mean μ and finite standard deviation σ, that when *n* is large, the sampling distribution of the sample mean $\overline{X}$ is approximately normal.

4. The Central Limit Theorem allows us to use normal probability calculations to answer questions about sample means from many observations even when the original population distribution is not normal.

Possible Extensions

This lesson very naturally extends to the introduction of more in-depth concepts involving the mean and standard deviation of the sampling distribution of $\overline{X}$. This lesson is a natural pre-requisite to a formal discussion and introduction to the Central Limit Theorem. See Optional Step 6 in the activity below for a transition.

References

1. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) *Report*, ASA, Franklin et al., ASA, 2007 http://www.amstat.org/education/gaise/.

2. Activity background adapted from: *The Practice of Statistics*, Fourth Edition by Starnes, Daren S., Yates, Daniel S., and Moore, David S., 2012. W.H. Freeman and Company.

Who Sends the Most Text Messages? Activity Sheet

Introduction

Sally, Jamal, Ivy, and Tom are students in Mr. Miller's Statistics class. One day in class they get into a discussion about text messaging. Sally and Ivy are convinced that they do more text messaging than Jamal and Tom because they think that girls like to communicate with each other more than guys do. Tom thinks he sends/receives as many text messages as Sally and Ivy because he recently received a new smartphone for his birthday and is still enjoying being able to communicate more easily with his friends. Jamal thinks he does just as much text messaging as the others in the group because texting is one of the main ways he and his girlfriend communicate. In their statistics class, they have recently learned how to collect data and calculate summary statistics. They decide to use what they have learned to investigate and compare the amount of text messaging they do. Each of them records the number of text messages they send or receive every day for 60 days. They then use statistical analysis to determine who does the most text messaging.

Problem

Use the data collected by the four students to perform the following investigation to determine which of the four students did the most text messaging.

Instructions

Your class will be divided into groups. Each group will be given a bag with the name of one of the four students (Sally, Jamal, Ivy or Tom) on the outside. Each bag will contain 60 cards, one for each of the 60 days that the person recorded the number of text messages they sent/received. The number of text messages for each day is written on one of the cards. Your teacher will assign a section of the board on which you will record information and draw required graphs.

Name on Bag ___________________________

Step 1: Random Sampling with Sample Size *n* = 4

Draw four dot plots with values ranging from 0 to 120, marked off in increments of 10, on your group's section on the board. Write the name from the outside of the bag below the dot plots.

Have one student in your group draw 4 cards blindly from the bag and write the numbers from the cards on the board. Each person in your group will now calculate the sample mean $\bar{X}$ of the 4 observations and one person will record this value on the board below the dot plot.

1. Sample mean $\bar{X}$ of the 4 observations _______________
   Show the steps for calculating the sample mean.

You have just taken one random sample of size $n = 4$ from the population of the 60 cards in your bag and calculated the sample mean $\bar{X}$.

Return the 4 cards to the box and shuffle the cards thoroughly. Have a student draw another random sample of size 4, record the numbers, and find $\bar{X}$. Repeat this process at least 20 times. Each time a new sample mean $\bar{X}$ is calculated:

(1) Record the value on the attached data sheet.

(2) Plot the $\bar{X}$-value on one of the dot plots. Label this dot plot $n = 4$.

After you have completed the random sampling process with sample size $n = 4$ and drawn the dot plot, answer the following questions:

2. The dot plot is an approximation to the sampling distribution of what sample statistic?

3. The center of the sampling distribution of the sample mean $\bar{X}$ for $n = 4$ is around ______ and its shape is ____.

4. Use your calculator to calculate the mean and standard deviation of the sampling distribution of the sample mean $\bar{X}$ for $n = 4$.

   Mean of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 4$_____________

   Standard Deviation of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 4$_____________

Step 2: Random Sampling with Sample Size *n* = 10

Repeat the process followed in Step 1, only this time draw 10 cards each time instead of 4. After you have completed the random sampling process with sample size $n = 10$, recorded the values of the sample means on the attached data sheet, draw the dot plot, label the dot plot $n = 10$ and answer the following questions.

5. The dot plot is an approximation to the sampling distribution of what sample statistic?

6. The center of the sampling distribution of the sample mean $\bar{X}$ for $n = 10$ is around ______ and its shape is ____.

7. In what ways does the sampling distribution of the sample mean $\bar{X}$ for $n = 10$ differ from the sampling distribution for $n = 4$?

8. Use your calculator to calculate the mean and standard deviation of the sampling distribution of the sample mean $\bar{X}$ for $n = 10$.

   Mean of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 10$ ___________

   Standard Deviation of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 10$ ___________

Step 3: Random Sampling with Sample Size $n = 30$
Repeat the process followed in Step 1, only this time draw 30 cards each time instead of 4. TIME-SAVING HINT: As you draw the samples of 30 cards, enter the values in a list in your calculator. Then use the one-variable statistics calculation option under the Statistics menu to find the sample mean. After you have completed the random sampling process with sample size $n = 30$, record the values of the sample means on the attached data sheet, draw the dot plot, label the dot plot $n = 30$ and answer the following questions.

9. The dot plot is an approximation to the sampling distribution of what sample statistic?

10. The center of the sampling distribution of sample mean $\bar{X}$ for $n = 30$ is around ______ and its shape is ____.

11. In what ways does the sampling distribution for $n = 30$ differ from the sampling distributions for $n = 4$ and $n = 10$?

12. Use your calculator to calculate the mean and standard deviation of the sampling distribution of sample mean $\bar{X}$ for $n = 30$.

Mean of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 30$__________

Standard Deviation of Sampling Distribution of Sample Mean $\bar{X}$ for $n = 30$ ___________

Step 4: Dot Plot Showing Population Distribution
On the fourth dot plot you will plot each of the numbers on the 60 cards in the bag (the total population). After drawing the dot plot, answer the following questions.

13. How does the shape of the dot plot of the population compare with the shapes of the sampling distributions?

14. Use your calculator to calculate the mean and standard deviation of the population.

Mean of the population ____________________

Standard deviation of the population _____________________

Write the value of the population mean below the dot plot of the population.

Step 5:
At this point in the investigation, you should have the following four dot plots on the board for each of the four populations:

- Approximation of the Sampling Distribution of Sample Mean $\bar{X}$ for sample size $n = 4$
- Approximation of the Sampling Distribution of Sample Mean $\bar{X}$ for sample size $n = 10$
- Approximation of the Sampling Distribution of Sample Mean $\bar{X}$ for sample size $n = 30$
- Distribution of the entire Population of 60 days of text messages

After observing and discussing the dot plots for each of the students, collaborate within your group and answer the following questions:

15. At the beginning of the activity, the stated problem was:
    Use the data collected by the four students to investigate which of the four students did the most text messaging.
    What is the SPECIFIC question that needs to be addressed in your investigation? Be sure to use the vocabulary of statistics in your answer.

16. After exploring the data collected for the four students, how would you answer the specific question you just posed? Give a detailed explanation of how you arrived at your answer. Be sure to use the vocabulary of statistics in your explanation.

Step 6: Extension: Introduction to the Central Limit Theorem
Observe the four POPULATION dot plots on the board (one each for Sally, Jamal, Ivy, and Tom) and answer the following questions:

17. Describe the shape of the dot plot for each of the populations.

    SALLY____

    JAMAL____

    IVY____

    TOM____

18. For each of the four students, observe the 3 sampling distributions of the sample means $\bar{X}$ for sample sizes n = 4, n = 10 and n = 30. What do you notice about the dot plots as n becomes larger?

The **Central Limit Theorem** is one of the most important theorems in statistics. It states that when we draw a simple random sample of size n from **ANY** shaped population with mean μ and finite standard deviation σ, that when n is large, the sampling distribution of the sample mean $\bar{X}$ is approximately normal. This is true even if the original population is not normal!

19. You have observed that the populations for Jamal, Ivy and Tom are not normally distributed. At sample size n = 30, were the sampling distributions of the sample mean $\bar{X}$ for these 3 students approximately normal?

Why is the Central Limit Theorem so important?
The Central Limit Theorem allows us to use normal probability calculations to answer questions about sample means from many observations even when the original population distribution is not normal.

DATA SHEET

Sample Means for *n* = 4	**Sample Means for *n* = 10**	**Sample Means for *n* = 30**
1. $\bar{X}$ = __________	1. $\bar{X}$ = __________	1. $\bar{X}$ = __________
2. $\bar{X}$ = __________	2. $\bar{X}$ = __________	2. $\bar{X}$ = __________
3. $\bar{X}$ = __________	3. $\bar{X}$ = __________	3. $\bar{X}$ = __________
4. $\bar{X}$ = __________	4. $\bar{X}$ = __________	4. $\bar{X}$ = __________
5. $\bar{X}$ = __________	5. $\bar{X}$ = __________	5. $\bar{X}$ = __________
6. $\bar{X}$ = __________	6. $\bar{X}$ = __________	6. $\bar{X}$ = __________
7. $\bar{X}$ = __________	7. $\bar{X}$ = __________	7. $\bar{X}$ = __________
8. $\bar{X}$ = __________	8. $\bar{X}$ = __________	8. $\bar{X}$ = __________
9. $\bar{X}$ = __________	9. $\bar{X}$ = __________	9. $\bar{X}$ = __________
10. $\bar{X}$ = __________	10. $\bar{X}$ = __________	10. $\bar{X}$ = __________
11. $\bar{X}$ = __________	11. $\bar{X}$ = __________	11. $\bar{X}$ = __________
12. $\bar{X}$ = __________	12. $\bar{X}$ = __________	12. $\bar{X}$ = __________
13. $\bar{X}$ = __________	13. $\bar{X}$ = __________	13. $\bar{X}$ = __________
14. $\bar{X}$ = __________	14. $\bar{X}$ = __________	14. $\bar{X}$ = __________
15. $\bar{X}$ = __________	15. $\bar{X}$ = __________	15. $\bar{X}$ = __________
16. $\bar{X}$ = __________	16. $\bar{X}$ = __________	16. $\bar{X}$ = __________
17. $\bar{X}$ = __________	17. $\bar{X}$ = __________	17. $\bar{X}$ = __________
18. $\bar{X}$ = __________	18. $\bar{X}$ = __________	18. $\bar{X}$ = __________
19. $\bar{X}$ = __________	19. $\bar{X}$ = __________	19. $\bar{X}$ = __________
20. $\bar{X}$ = __________	20. $\bar{X}$ = __________	20. $\bar{X}$ = __________

Answers to Activity Sheet

Step 1:

1. Answers will vary, but students should show correct steps for calculating the mean of the numbers on the four cards drawn from their bag.
2. Sample mean. 3. Answers will vary.
4. Answers will vary, but they should be the mean and standard deviation of the sample means listed on the data sheet in the n = 4 column.

Step 2:

5. Sample mean. 6. Answers will vary.
7. Answers will vary, but students may begin to notice the following two changes:
   - a decrease in the spread
   - shape is becoming closer to approximating a normal distribution.
8. Answers will vary, but they should be the mean and standard deviation of the sample means listed on the data sheet in the n = 10 column.

Step 3:

9. Sample mean. 10. Answers will vary.
11. At this point students should see a definite decrease in the spread and the shape should be very close to approximating a normal distribution.
12. Answers will vary, but they should be the mean and standard deviation of the sample means listed on the data sheet in the n = 30 column.

Step 4:

13. Answers will vary depending on the bag of cards the student received. The answers should correspond to the following shapes for each bag of cards.
    Sally – Normal Jamal – Bimodal Ivy – Skewed Tom – Scattered
14. The means and standard deviations of the populations should be as follows:
    Sally: Mean = 60, $\sigma = 4.18$ Jamal: Mean = 60, $\sigma = 39.78$
    Ivy: Mean = 74.25, $\sigma = 31.36$ Tom: Mean = 60, $\sigma = 35.64$

Step 5:

15. The specific question that needs to be addressed is, "Which of the four students in the statistics class sent/received the largest mean number of text messages over the 60-day period for which they collected data?"
16. Ivy was the student who sent/received the largest mean number of text messages over the 60-day period, so she did the most text messaging. The other three students sent/received the same mean number (60) of text messages.
17. Sally: Normal Jamal: Bimodal Ivy: Skewed Tom: Scattered
18. For each of the sampling distributions, the shape becomes closer to being normal as n increases. When n = 30, all four of the sampling distributions have a shape that approximates a normal distribution.
19. Yes, at sample size n = 30, the sampling distribution of the sample mean $\overline{X}$ for these 3 students became approximately normal.