PROJECT-SET
Statistics Education for Teachers

Who Sends the Most Text Messages?

Overview of Lesson
This activity allows students to perform an investigation in which they explore various factors affecting the shape of the sampling distribution of the sample mean $\bar{X}$. Students will be given data on the number of text messages sent/received by four different people over the course of the last 3 years.

Students will construct dotplots showing both the distribution of the total population, and the sampling distributions for sample means for the population. They will discover that the population distributions have distinctly different shapes for the four individuals, but the sampling distribution of the mean is approximately a normal distribution when the sample size is large enough.

One of the populations will have a higher mean than the other three, all of which have a mean equal to 60. Students will perform random sampling with sample sizes $n = 10$, $n = 30$, and $n = 50$, making dot plots for the sampling distribution of the sample mean $\bar{X}$ for each. They will also make a dot plot for the total population. The mean and standard deviation will be calculated for each dot plot. Students will be asked to analyze the effect of sample size on the sampling distribution, and to compare the shapes of the sampling distributions for sample size $n = 30$ and $n = 50$ to the shapes of the population distributions.

GAISE Components
This activity follows 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 Question, Collect Data, Analyze Data, and Interpret Results. This is a GAISE level B and C lesson.

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

Common Core State Standards Grade Level Content (Middle School & High School)
6.SP. 1. Develop understanding of sampling variability
7.SP. 1. Use random sampling to draw inferences about a population
7.SP. 3. Investigate chance processes and develop, use, and evaluate probability models
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 deviations) of two or more different data sets.
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

- 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 any 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 dotplot and do an informal analysis of the shape, center and spread. They should be able, with the aid of a calculator or software, to calculate the mean and standard deviation of a sampling distribution and of a population distribution.

Learning Targets

After completing the activity, students will have an understanding of the effect of sample size on the sampling distribution of the sample mean, $\bar{X}$, and be able to relate sampling and population distributions to the Central Limit Theorem. They will be familiar with the process of constructing dotplots for both sampling distributions of sample mean, $\bar{X}$, and for the total population. Students will 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 calculator or software, the activity sheet, and the text message data.

Instructional Lesson Plan
The GAISE Statistical Problem-Solving Procedure

1. Formulate Question(s)

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 questions of interest for the activity are: **Can the shape of the sampling distribution be predicted? How will the sampling distributions compare to samples taken from an approximately normal population? From a skewed population? From a bimodal population?**

II. Collect Data

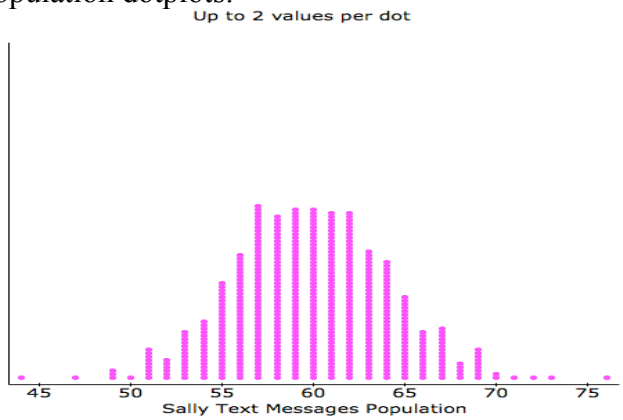
Students will be divided into four groups, with each group being assigned a person's text message data (each person's text message data is considered a population, therefore there are 4 populations of text message data in this lesson). Each entry in the data set will have a number representing the number of text

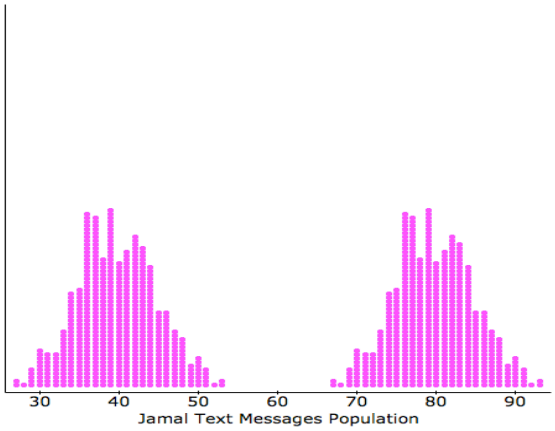
messages sent/received on one of the days collected in 3 years (total of 1085 days in three years so there are a total of 1085 text message entries in each data set).

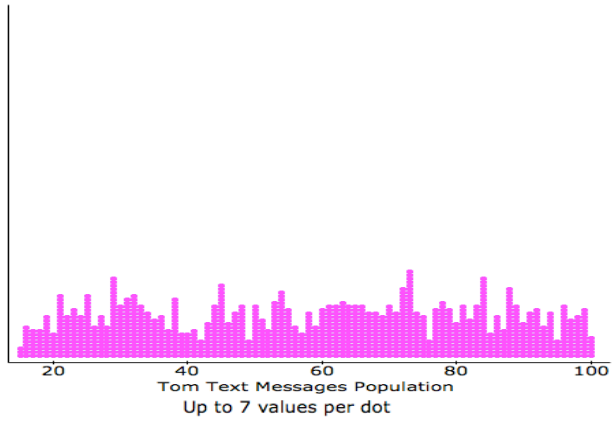
III. Analyze the Data By Measures and Graphs

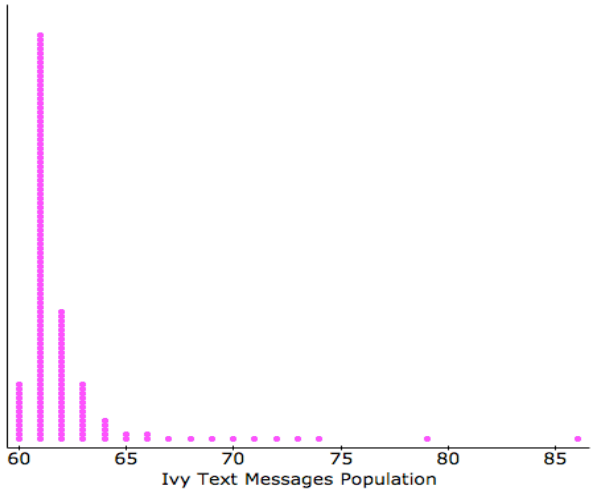
Students will be divided into four groups, with each group being given a text message population. The students will perform, using a calculator or software, random sampling from the population with 100 samples each of sample sizes $n = 10$, $n = 30$, and $n = 50$ in order to create an approximate sampling distribution. A dotplot showing the sampling distribution for sample mean $\bar{X}$ will be constructed for each sample size (see examples below). Students will also construct a dotplot for their population data. The mean and standard deviation will be calculated for each distribution (sampling and population). Students will then be asked to compare the sampling distributions for sample mean $\bar{X}$ for sample sizes $n = 10$,
$n = 30$, and $n = 50$ in order to see the effect of sample size on the distribution. Then, they will be asked to observe the four population distributions and their corresponding sampling distributions for sample size $n = 30$ and n=50.

Population dotplots:

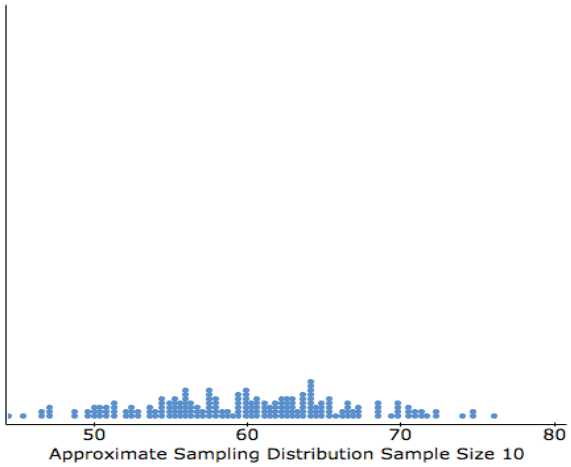


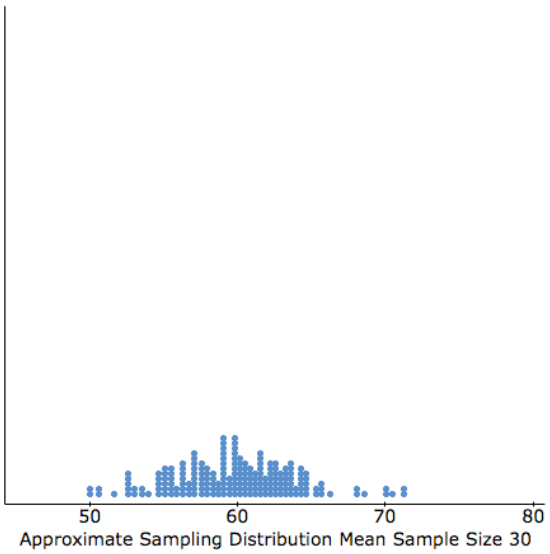


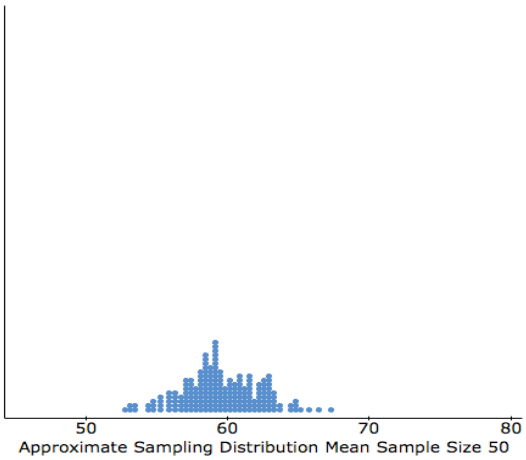




Taking repeated samples from, for example, Jamal's population, and finding an approximate sampling distribution for the sample mean for samples of size 10, 30, and 50 could yield the following results:







From this, s**tudents 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. They should also note that the mean stays the same regardless of the sample size and in particular, it equals the population mean.**

IV. Interpret the Results

During the activity, students are asked to compare the relationship of the population text messaging data with that of the sampling distribution. More specifically, they should notice that regardless of the shape of the original population distribution, their sampling distribution of the sample mean approaches a normal shape as the sample size increases, the standard deviation decreases as the sample size increases, and the mean of the sampling distribution is the same as the population distribution regardless of the sample size.

Possible Extensions

1. After constructing a dotplot for each distribution, students have been exposed to the Central Limit Theorem informally. Therefore, an extension to this lesson is to formally state and study the CLT.
2. This lesson very naturally extends to the introduction of more in-depth concepts involving the mean and standard deviation of the sampling distribution of $\bar{X}$. For example, the formula for the standard deviation of the sampling distribution of $\bar{X}$, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, can be introduced.

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.

Activity

Who Sends the Most Text Messages?

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 do more text messaging than Tom and Jamal 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 Jamal 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. From their service carriers they obtain data about their text messaging use from the last three-years. Each of them records the number of text messages they send or receive every day for 1085 days. They then use statistical analysis to compare their text messaging patterns.

Problem

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

Instructions: Your class will be divided into four groups. Each group will be given a text messaging data set for one of the students: Sally, Jamal, Tom or Ivy. Each data set will contain 1085 entries, one for each of the 1085 days that the service carriers recorded the number of text messages they sent and received. The number of text messages for each day is one entry in the data set. Your teacher will assign a section of the board on which you will record information and draw required graphs.

Name of the Student Assigned_____

Step 1: Random Sampling with Sample Size $n = 10$

Construct an approximate sampling distribution for the sample mean $\bar{X}$ with samples of size 10. Take 200 repeated samples. Cut and paste your sampling distribution here and use it to answer the questions below.

1. The dotplot is an approximation to the sampling distribution of what population parameter?

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

3. Use technology to calculate the mean and standard deviation of the approximate sampling distribution of the sample mean $\bar{X}$ for $n = 10$.

   Mean of the approximate sampling distribution of sample mean $\bar{X}$ for $n = 10$ is _____

   Standard Deviation of the approximate sampling distribution of sample mean $\bar{X}$ for $n = 10$ is _____

Step 2: Random Sampling with Sample Size *n* = 30

Construct an approximate sampling distribution for the sample mean $\bar{X}$ with samples of size 30. Take 200 repeated samples. Cut and paste your approximate sampling distribution here and use it to answer the questions below.

1. The dot plot is an approximation to the sampling distribution of what population parameter?

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

3. Use technology to calculate the mean and standard deviation of the approximate sampling distribution of the sample mean $\bar{X}$ for $n = 30$.

   Mean of the approximate sampling distribution of sample mean $\bar{X}$ for $n = 30$ is _____

   Standard Deviation of sampling distribution of sample mean $\bar{X}$ for $n = 30$ is

   _____

Step 3: Random Sampling with Sample Size *n* = 50

Construct an approximate sampling distribution for the sample mean $\bar{X}$ with samples of size 50. Take 200 repeated samples. Cut and paste your approximate sampling distribution here and use it to answer the questions below.

1. The dot plot is an approximation to the sampling distribution of what population parameter?

2. The center of the approximate sampling distribution of the sample mean $\bar{X}$ for $n = 50$ is around _____ and its shape is _____.

3. Use technology to calculate the mean and standard deviation of the approximate sampling distribution of the sample mean $\bar{X}$ for $n = 50$.

   Mean of sampling distribution of sample mean $\bar{X}$ for $n = 50$ is _____

   Standard standard deviation of sampling distribution of sample mean $\bar{X}$ for $n = 50$ is _____

Step 4: Dot Plot Showing Population Distribution

On the fourth dotplot, show the text message population distribution for the student you were assigned to. Cut and paste your sampling distribution here and use it to answer the questions below.

1. How does the shape of the dotplot of the total population compare with the shapes of the approximate sampling distributions?

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

   Mean of total population _____

   Standard deviation of total population _____

Step 5:

At this point in the investigation, you should have the following four dotplots in front of you for your population:

- Approximation of the sampling distribution of sample mean $\bar{X}$ for $n = 10$
- Approximation of the sampling distribution of sample mean $\bar{X}$ for $n = 30$
- Approximation of the sampling distribution of sample mean $\bar{X}$ for $n = 50$
- Entire Population of 1085 days of text messages

Display the dotplots so that your classmates can see them. After observing and discussing the dot plots for each of the students, collaborate within your group and answer the following questions:

1. Can the shape of the sampling distribution be predicted?

2. How would the sampling distributions compare if they were taken from an approximately normal population? From a skewed population? From a bimodal population?

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. Population 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:

1. Population mean.
2. Answers will vary.
3. 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.
4. Answers will vary, but they should be the mean and standard deviation of the sample means listed on the data sheet in the $n = 10$ column.

Step 3:

1. Population mean.
2. Answers will vary.
3. At this point, students should see a definite decrease in the spread. The shape should be very close to approximating a Normal distribution.
4. Answers will vary, but they should be the mean and standard deviation of the sample means listed on the data sheet in the $n = 30$ column.

Step 4:

1. 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
2. The means and standard deviations of the populations should be as follows:
   Sally: Mean = 60, $\sigma = 4.18$
   Jamal: Mean = 60, $\sigma = 41.13$
   Ivy: Mean = 74.25, $\sigma = 31.36$
   Tom: Mean = 60, $\sigma = 35.64$

Step 5:

1. 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?"
2. Ivy was the student who sent/received the largest mean number of text messages over the 60-day period, so he did the most text messaging. The other three students sent/received the same mean number (60) of text messages.
3. Sally: Normal Jamal: Bimodal Ivy: Skewed Tom: Scattered
4. 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.
5. Yes, at sample size $n = 30$, the sampling distribution of the sample mean $\bar{X}$ for these 3 students became approximately Normal.