

Sampling Distributions of the Sample Mean—Pocket Pennies

You will need

- 25 pennies collected from recent day-to-day change

Some of the distributions of data that you have studied have had a roughly normal shape, but many others were not normal at all. What kind of distribution tends to emerge when you create sampling distributions of the mean from these non-normal populations? Here you'll explore that question.

COLLECT DATA

1. Enter the dates on your random sample of 25 pennies into a Fathom case table, using the attribute *Year*. Create a second attribute, *Age*, with a formula that calculates the difference between the current year and *Year*.

	Year	Age
=		2005 - Year
1	1996	9
2	1985	20
3	1985	20
4	1993	12

- Q1** If you were to make a histogram of the ages of all the pennies from all the students in your class, what do you think the shape of the distribution would look like? Sketch your prediction.
 2. Combine everyone's data into one collection and make sure everyone has a copy of that Fathom document.
 3. Using the complete collection, make a histogram of the ages of all the pennies in the class.
- Q2** How does the actual distribution compare with your prediction in Q1?
- Q3** Estimate the mean and standard deviation of the distribution. Confirm these estimates by computing the mean and standard deviation in Fathom. Either plot the values on the histogram or use a summary table.

You can copy cases from one case table to another by choosing **Select All Cases**, **Copy**, and **Paste** from the **Edit** menu.

INVESTIGATE

Building a Sampling Distribution

Next you'll take a random sample of size 5 from the ages of your class's pennies.

4. Select the collection, and choose **Collection | Sample Cases**. By default, Fathom takes a sample of ten cases with replacement and places them in a new collection named Sample of Pocket Pennies. You'll change this to five cases without replacement.

Sampling Distributions of the Sample Mean—Pocket Pennies

continued

Notice that animation is on by default. You may want to change this later.

5. Double-click the Sample of Pocket Pennies collection to show its inspector. On the **Sample** panel, change the settings to match these. Click **Sample More Cases** to re-collect your sample.

Animation on With replacement
 Replace existing cases
 Collect new sample when source changes
 5 cases

The *SampleSize* measure will allow you to compare different sample sizes later on.

6. Go to the **Measures** panel of the inspector and define these measures.

Q4 If you were to make a histogram of the mean ages from several samples, do you think the mean of the values in this histogram would be larger

than, smaller than, or the same as the mean of the population of the ages of all pennies? Regardless of your choice, try to make an argument to support each choice. Estimate what the standard deviation of the distribution of the mean ages will be.

Measure	Value	Formula
MeanAge	9.4	mean (Age)
SampleSize	5	count ()
<new>		

7. Collect the mean ages from several samples by selecting the sample collection, then choosing **Collection | Collect Measures**. Show the inspector for the measures collection and change to these settings. Click **Collect More Measures**.

Animation on
 Replace existing cases
 Re-collect measures when source changes
 100 measures

8. Make a histogram of *MeanAge*. Compute the mean and standard deviation of *MeanAge* by plotting values on the graph or using a summary table.

Q5 Which of the three choices in Q4 appears to be correct?

Changing the Sample Size

You'll now collect measures for samples of size 10 and size 25. You'll be able to create a split histogram to compare the effect of sample size.

9. Show the inspector for the sample collection. On the **Sample** panel, change the sample size to 10.
10. Show the inspector for the measures collection. On the **Collect Measures** panel, uncheck **Replace existing cases**. This allows you to put the measures from all the samples into one collection. Then click **Collect More Measures**.

You may want to turn off animation in the sample and measures collections.

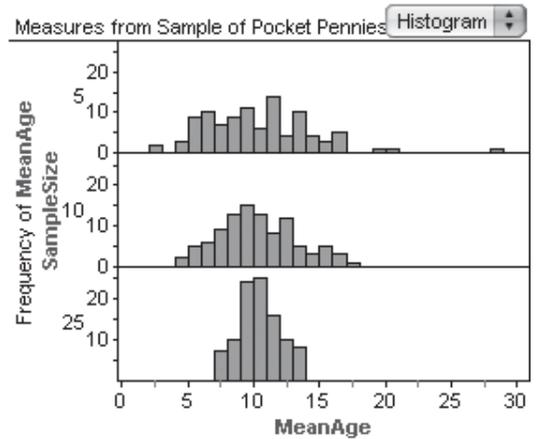
Sampling Distributions of the Sample Mean—Pocket Pennies

continued

Holding down the Shift key tells Fathom to use the numerical values of the attribute as categories.

11. Repeat steps 9 and 10 to change the sample size to 25 and collect 100 more measures.

12. Drag the attribute *SampleSize* of the measures collection and drop it on the vertical axis of the histogram for *MeanAge* while holding down the Shift key. Your histogram should split three ways, showing distributions for each sample size (5, 10, and 25).



13. Compute the mean and standard deviation for each of the three sampling distributions, using a summary table. Again, hold down the Shift key when you drop *SampleSize* in the summary table.

Measures from Sample of Pocket Pennies		MeanAge
SampleSize	5	
	10	
	25	
Column Summary		

- Q6** Look at the four histograms you constructed. As the sample size increases, what can you say about the shape of the histogram of sample means? About the center? About the spread?
- Q7** Compare the values you got in step 13 for the mean and SD for the three sampling distributions with the values you got in Q3 for the whole population. Then figure out formulas for the mean and SD of a sampling distribution that relate them to the population mean and SD.

14. On your histogram, plot the values mean, mean *plus* 2SD, and mean *minus* 2SD.

MeanAge

| mean () = 10.2621
 | mean () + 2 stdDev () = 16.3766
 | mean () - 2 stdDev () = 4.14756

- Q8** What percentage of sample means are within 2 SD's of the population mean for each sample size?
- Q9** For which sample sizes would it be reasonable to use the rule stating that 95% of all sample means lie within approximately 2 SD's of the population mean?

EXPLORE MORE

Open the Fathom document **LifeExp.ftm**. In this file you will find data on the life expectancy for females in Asia and Africa. Discuss the shapes of the original population. Take 200 samples of size 5 from each population. Do your conclusions from Q6–Q9 still hold up?

Objectives

- Understanding the concept of a sampling distribution of the sample mean and how to generate one
- Discovering the properties of the shape, mean, and standard deviation of the sampling distribution of the sample mean
- Recognizing that the mean of the sampling distribution of the sample mean is approximately the mean of the population
- Seeing that the standard deviation of the sampling distribution of sample means decreases as the sample size increases
- Being introduced to the Central Limit Theorem: The sampling distribution of the sample mean approaches the normal distribution as the sample size increases, regardless of the shape of the original population distribution.

Activity Time: 30–50 minutes (the shorter time is when data collection is done the day before the activity)

Setting: Paired/Individual Activity (collect data using **PenniesTemplate.ftm** or use **Pennies.ftm**) or Whole-Class Presentation (use **Pennies.ftm**)

Optional Document: **LifeExp.ftm** (Explore More)

Materials

- 25 pennies collected by each student from recent day-to-day change

Statistics Prerequisites

- Familiarity with taking a sample
- Comparing distributions graphically
- Measures of center and spread

Statistics Skills

- Sampling distributions of the sample mean
- Properties of the shape, center, and spread of the sampling distribution of the sample mean
- Introduction to a geometric distribution
- Normal distribution
- Reasonably likely sample means
- Central Limit Theorem

AP Course Topic Outline: Part II B (4); Part III C, D (2, 3, 6)

Fathom Prerequisites: Students should be able to make case tables and graphs, plot values, find statistics in a summary table, and define attributes.

Fathom Skills: Students sample from a collection, define and collect measures, and combine measures for different sample sizes. *Optional:* Students use the normal density function (Extension 2) and a normal quantile plot (Extension 3).

General Notes: In this activity, students discover the properties of the shape, mean, and standard deviation of the sampling distribution of the mean for samples taken from a distribution that is decidedly not normal. It involves a lot of repeated random sampling to create distributions of means for different sample sizes. Because Fathom automates the sampling process, students will spend less time on the busywork and more time examining the results.

Procedure: During the week before the activity, have each student collect the first 25 pennies that he or she receives in change from various purchases and emphasize that these should not be some collection of pennies stored from years long past. Students should bring in the pennies and a list of the 25 dates on the pennies.

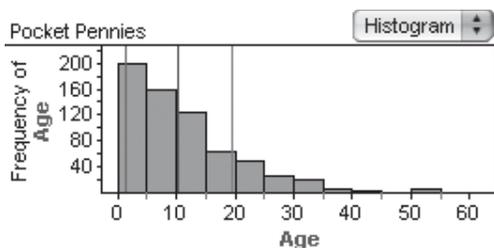
You will need to collate all the data into a single Fathom document that students can work with. The student worksheet suggests that this be done using copy and paste. Alternatively, you could use the master document, **PenniesTemplate.ftm**, into which all students enter data. A third alternative is to collect a list of the dates of the pennies from the previous class session, type them into a Fathom document, and distribute this document to the class. Another possibility is to have students type in their data and then email their document to one person who will copy and paste the data into a new class document. If you don't have time to collect data, the document **Pennies.ftm** contains sample data.

continued

COLLECT DATA

Q1–Q3 Few students realize that the shape of the distribution of the ages of all pennies will be roughly geometric. Many will believe that it should be normal (“A few pennies are new, a few pennies are old, most are lumped in the middle.”).

In a number of places, students are asked to compute the mean and standard deviation of a distribution. Two ways of doing that are shown here. Plotting values for the mean and the mean plus or minus the standard deviation on top of a histogram is very satisfying, but it doesn’t give you a direct value for the standard deviation. Using a summary table gives you the numbers but no visual context.



| mean() = 10.4398
 | mean() + s() = 19.6527
 | mean() - s() = 1.22692

Pocket Pennies	
Age	10.439815 9.212891

S1 = mean()
 S2 = s()

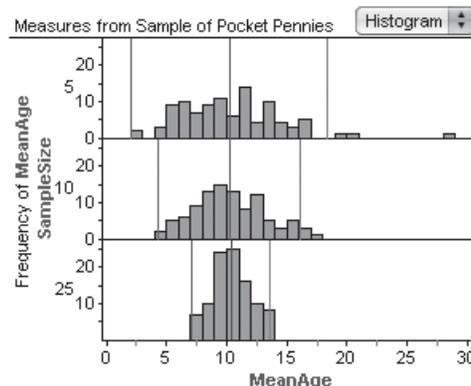
The mean age for the population tends to be between 7 and 8 years, with standard deviation about 8 years. If your results for Q3 are contrary to this standard, you may want to discuss potential causes.

INVESTIGATE

Q4–Q5 The mean will be the same. Most students will not realize this. A typical answer is to say it will be smaller. Most students will realize that the standard deviation will be smaller.

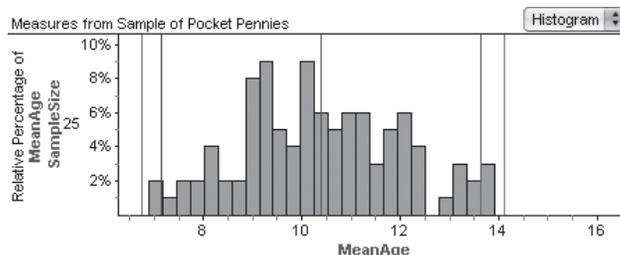
Q6–Q7 This split histogram shows the kind of results students are likely to get. The shape of the distributions becomes approximately normal as the sample size increases, the mean stays the same, and the standard deviation decreases. Specifically, the mean of the three sampling distributions should

be approximately the same as the mean of the population of all pennies: $\mu_{\bar{x}} = \mu$. The standard deviation of the sampling distributions should approximately equal the standard deviation of the population divided by the square root of the sample size: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.



| mean() = 10.2621
 | mean() + 2 stdDev() = 16.3766
 | mean() - 2 stdDev() = 4.14756

Q8 It is easiest to use the histogram. For both samples of size 5 and 10, using the sample statistics as shown in the above histogram will give nearly the same values as would plotting the theoretical values. For samples of size 5 in the sample document, 97% are within 2 standard deviations of the population mean (10.439). For samples of size 10, 97% are within 2 standard deviations of the population mean. For samples of size 25, 100% are within 2 SD’s of the mean using the sample statistics as shown below. Using the population parameters and the Central Limit Theorem, the percentage within 2 SD’s is exactly 95%.



| mean() = 10.3852
 | 10.4398 + 2 * (9.212891 / sqrt(25)) = 14.125
 | 10.4398 - 2 * (9.212891 / sqrt(25)) = 6.75464
 | mean() - 2 stdDev() = 7.14559
 | mean() + 2 stdDev() = 13.6248
 SampleSize = 25

Q9 For this sample of pennies, it seems that we could use the rule for any of them. However, it would be safer to try sampling again to see if this holds up. Typically, it would not be a good idea to use this rule with samples of size 5.

DISCUSSION QUESTIONS

- What does one case in the Pennies collection represent? One case in Sample of Pennies? One case in Measures from Sample of Pennies?
- What did you observe as you increased the sample size from 5 to 10 to 25? Explain why this makes sense.
- For samples of size 25, what characteristics does the sampling distribution of the sample mean share with the normal distribution?

EXTENSIONS

1. Have students extend the sample sizes to 50 and 100. Does this confirm or modify the conclusions they made in Q6–Q9?
2. Have students figure out how to plot a normal curve on top of the histograms of *MeanAge* for each sample size. (Try searching Fathom Help for “normal distribution.”) What do they observe about the fit of the histogram with the curve? [Students should plot the function $100 \cdot \text{normalDensity}(x, \text{population mean}, \text{stdDev}())$. The 100 scales it vertically, and the population mean translates the center. Students should find that as sample size increases, the distribution approaches the normal curve—the Central Limit Theorem. This is especially visible if students try collecting measures for larger sample sizes, such as 100.]
3. Students can investigate using a normal quantile plot to determine closeness to normality. (Use Fathom Help to learn about normal quantile plots.) Which sample size produces a distribution closest to normal? [The normal quantile plot should show each sample size to be normal. The sample of size 25 should appear slightly more normal than the others, however.]