**Access the document: https://go.ncsu.edu/nctm2019**

# 3 Part Task*: Informal Inferential Reasoning

**Emily Elrod (ekelrod@ncsu.edu) & Asli Multu (mamutlu@ncsu.edu)**

**Presentation for NCTM Annual Meeting & Exposition 2019**

## Context for Three Part Task

Have you ever wondered how long coins stay in circulation? Are you a collector? Imagine that the president of "COINS UNLIMITED" has just hired you as his chief statistician for research on the age of pennies. You are charged with the task of estimating the average age of pennies in circulation within one year of age with high confidence based on your findings from the data. The distribution of the ages, not the year the coin was minted, will be examined.

A little background on the circulation of pennies:
The Federal Reserve and the US Treasury department take currency, including pennies, out of circulation after so many years. Most coins remain in circulation for about 30 years, whereas paper currency has a shorter lifespan, with the $1 bill having an estimated lifespan of approximately 5 years. [Data from the US Government Accountability Office's report in 2012 (USGAO, 2012)]. Why do you think the Treasury department would need to take currency out of circulation?

Suppose we want to answer the question: Typically, how old are your pennies? What information do you think you would need to know to answer this question?

*This task was adapted from an activity found in the Georgia Department of Education Mathematics IV Unit 1 document (found at http://bit.ly/2CfiJDk)

**Informal Inferential Statistics for finding the ages of Pennies***

**Task 1**

Working with your group, you will collect two samples from the "class bowl" of pennies. **You will be determining the age of pennies, not simply recording the year they were made!**

1. You will **collect one sample of size of 10 pennies**. Have one person record the year of all the pennies on the google sheet here, then return the pennies to the bag.
2. **Next, collect a sample of size of 30 pennies**. Again, have one person record the year of all the pennies on the same google sheet as above (in the column for sample size of 30), then return the pennies to the bag.
3. Calculate the sample mean for each of your samples and record the sample means below in the table. Once your group has collected your samples, one person should also enter your group's sample mean data into another google sheet found here.

| **Sample Size** | **Sample Mean** |
|---|---|
| 10 | |
| 30 | |

***While you are waiting for all of the class to input their data, move to question 4.***

4. **Highlight the correct vocabulary** in **Column 2** for each aspect of the experiment in **Column 1** of the table. In column 3, explain how you know this is the correct vocabulary.

| **Column 1** | **Column 2** | **Column 3** |
|---|---|---|
| Each group's scoop of pennies (either size 10 or 30) | Population/Sample/ Sampling Distribution | |
| Class bowl of pennies | Population/Sample/ Sampling Distribution | |
| Collection of means from each group's scoop of pennies (either size 10 or 30) | Population/Sample/ Sampling Distribution | |

5. When the whole class enters their data into the google sheet, it will be uploaded into a FREE dynamic online software program called CODAP, and sampling distribution graphs will be created. On the screen you will be seeing the boxplots with sample size 10 and sample size 30 for each group for each sample of pennies from each group. What similarities/differences do you see from these boxplots?

6. From the boxplots for a sample size of 10,
    a. What does the minimum value represent in the context of our penny distribution?
    b. What does the maximum value represent in the context of our penny distribution?
    c. What does the range represent in the context of our penny distribution?

7. From the boxplots for a sample size of 30,
    a. What does the minimum value represent in the context of our penny distribution?
    b. What does the maximum value represent in the context of our penny distribution?
    c. What does the range represent in the context of our penny distribution?

8. What parts of questions 6 and 7 were different? **Explain why.**

9. All pennies were drawn from the same bowl, why do all of the sample means not the same value?

Now we will create two sampling distributions. One of them will be the sample means for all samples of size 10. The other, the sample means for all samples of size 30.

10. What would you predict for the mean (i.e *mean of means*) of the sampling distribution for sample size 10?

11. What would you predict for the mean (i.e *mean of means*) of the sampling distribution for sample size 30?

12. Which sampling distribution of sample means (sample size 10 or sample size 30) would be more accurate to make inferences about the population mean? **Explain.**

13. Can you make an ***approximate*** **assumption** about the average age of the pennies in the class bowl?

14. Can you make an ***approximate*** **assumption** about the standard deviation of the ages of pennies in the class bowl?

**Inferential Reasoning  for finding the ages of Pennies***

**Task 2**

Do you remember "COINS UNLIMITED"? The previous portion of this task utilized hands-on data collection to examine the age of pennies in circulation.  Think about what inferences you were able to make about samples as the sample size increases.  What about if the number of samples increases?

Now that you have that knowledge, you are now going to explore data collection through the use of an online applet to examine the age of pennies in circulation.  Instead of going to get a roll of pennies, we will use simulation techniques to form a distribution of their ages (NOT the year on the coin).

We are lucky that we have various applets in our disposal. One of them is Rossman and Chance applet. By using this applet, we will try to simulate the event of having either 1,000, 4,000 or 40,000 pennies of population size.

**Getting Familiar with the data set.**
a.) Please visit the website of Rossman and Chance

b.)  Then, click on:

- *Pennies*
- *Show sampling options*
- *Overlay Normal Distribution*

**Part A:** Our penny data file can be found here.

**Question 1:**

i. Keeping the sample size as 10 and number of samples 1, and population size 400, click on "draw samples". Explain, why do you and your neighbor have different values and visuals after the "draw samples" action? Pay particular attention to the visuals: "population data", "Most recent sample" and "statistics."

ii. When we click on "draw on samples", what action are we actually simulating?

iii. Why do you think your neighbor and you did not have the exact same values or data visuals?

**Question 2:**

i. There are three data visuals/graphs. Please find the graph which is on most left hand side, what do you think each blue dot indicates?

ii. Now look at the middle graph, there are also some blue dots on this one, what do those blue dots indicate?

iii. On the right hand side graph, we have only one blue dot, what does that blue dot indicate?

iv. By looking at these graphs, can you identify some values as "parameters"? Please describe where you see these values and give a reason why you think that each value is a parameter.

v. By looking at these graphs, can you identify some values as "statistics"? Please describe where you see these values and give a reason why you think that each value is a statistic.

**Part B:**

Click on the **"Reset"** button. **Start with the sample size of 10**, with the number of samples equal to 10. Don't forget to click on "draw samples" for each iteration. Fill in the table below, as you repeat this action.

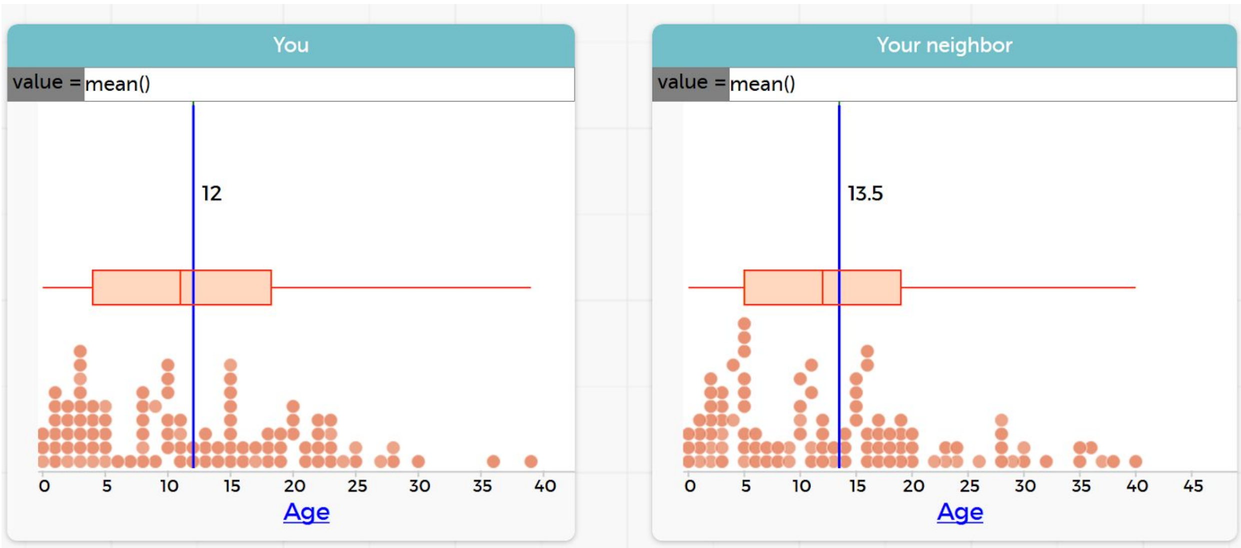
| **Sample size for n** | **Statistics Values** | **Parameter Values** |
|---|---|---|
| **10** | | |
| **30** | | |
| **50** | | |
| **100** | | |
| **200** | | |

**Question 3:**

i. When we click on "draw samples", what did we actually simulate?
ii. **Reset** and repeat this step **by changing the sample size** to 30, 50, 100, and then 200. (You should reset between each sample size). Record your values of "statistics" each time. Is the values of parameter changing?
iii. Explain the relationship between the values of parameter and statistics **in various sample sizes** of 10, 30, 50, 100 and 200

**Part C:**

Suppose, by using the Rossman and Chance applet, from the **same population** of pennies, you and your neighbor draw 10 samples with sample size of 100 pennies. This time you have obtained the data visuals below.




**Question 4:**

i. Compare and contrast the information you and your neighbor have from these two data visuals. Make statements about the shape, center and spread of the data distributions.

ii. What can you conclude about the population shape, center and spread? Do you think you have enough information to be able to make inferences about the population of pennies? Explain.

**Inferential Reasoning for finding the ages of Pennies**

**Task 3**

You have investigated data collected for penny ages in the previous **two** activities. Now it is time **to synthesize** all of the information.

- The first activity was a hands on activity as we were collecting the data from pennies.
- The second activity was about the simulation of collecting pennies with different sample sizes.

**Part 1:**
For question 1, please look at the two graphs below, Figure 1 and Figure 2.

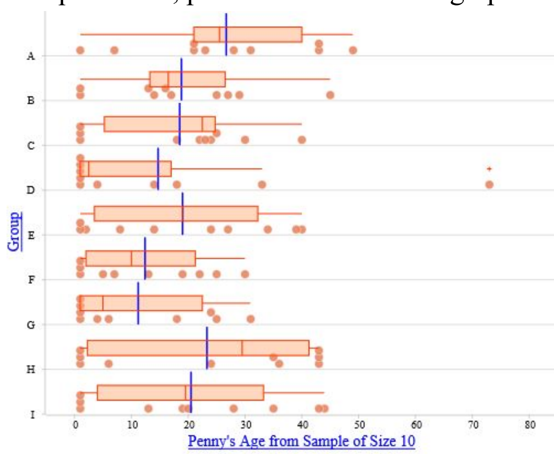



Figure 1

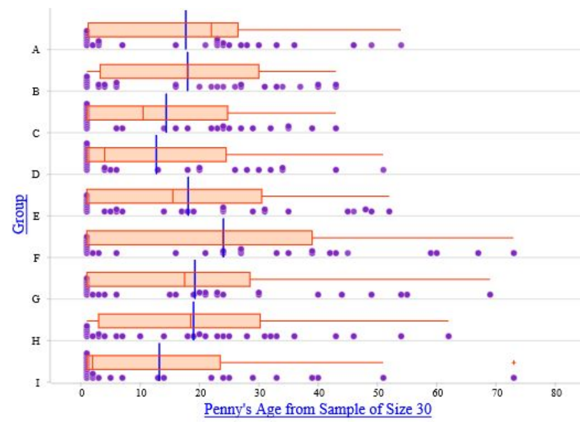



Figure 2

**Question 1:**

a.) In Figure 1, what does each orange dot represent?

b.) In Figure 2, what does each purple dot represent?

c.) What can you tell about the shape of each distribution in Figure 1 and Figure 2 ?

For question 2, please look at the two graphs below, Figure 3 and Figure 4.

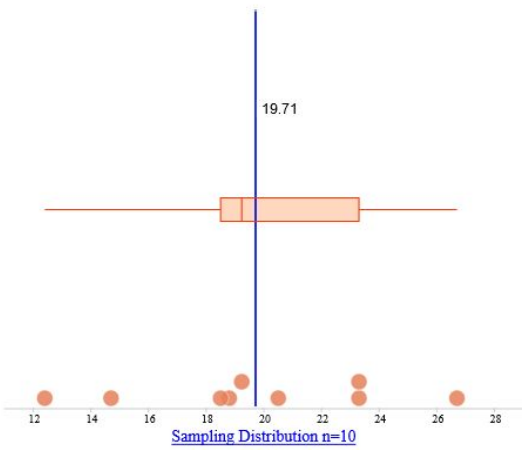



Figure 3

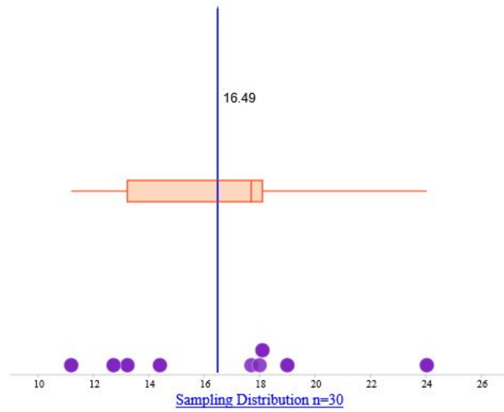



Figure 4

**Question 2:**

a.) In Figure 3 , what does each orange dot represent?

b.) In Figure 4, what does each purple dot represent?

c.) What can you tell about the shape of each distribution in Figure 3 and Figure 4 ? How do these two graphs in Figure 3 and Figure 4 differ in shape as compared to the two graphs in Figure 1 and Figure 2?

d.) The blue perpendicular lines indicate the mean of each group of data distribution of sample size n=10 and n=30. Think about the mean of the population distribution. Which graph (Figure 3 or Figure 4) will have a closer estimate value for the mean of the population? **Explain.**

**Part 2:**

The previous three activities bring us to a review of the **Central Limit Theorem for Sample Means:**

**The properties of the sampling distribution of the sample means:**

1. It is centered at the _____________ mean (symbol).
2. It has a smaller spread when the sample size of each sample is _____________ .
3. It has a shape that becomes more _____________ like as the sample size gets larger.
4. If you have a simple random sample of size *n* from any population, regardless of the original shape of the distribution, with mean $\mu$ and finite standard deviation $\sigma$. When *n* is large, the sampling distribution of the sample means $\mu_{\bar{x}}$ is approximately normal with mean _____________ and standard deviation _________________.

**Part 3:**

Question 1:

Lance's iPod has about 9,000 songs. The distribution of playtimes for these songs has a mean of 200 seconds and a standard deviation of 50 seconds. Anne has found out that one of her favorite songs also belongs to Lance's song list. Calculate the probability that the average song length of Anne's song will be less than 216 seconds?

Question 2:

Lance's iPod has about 9,000 songs. The distribution of playtimes for these songs has a mean of 200 seconds and a standard deviation of 50 seconds.

a. Laura has found that there are 36 songs that Laura and Lance have in common. Determine the average playtimes of these 36 songs which are sampled from Lance's iPod. Use the correct statistical symbol.

b. Determine the standard error of playtimes of 36 songs from Lance's iPod are sampled. Use the correct statistical symbol.

c. If these 36 songs from Lance's iPod are sampled, calculate the probability that the average song length will be less than 216 seconds.