Stat Support Activity: Using Coffee Data to Create a Sampling Distribution of Sample Means

Activity Part 1 Directions:

- Each student will go to www.matt-teachout.org and click on the "data sets" menu. Open "Sampling Distribution Data 1" in Excel. These are random samples of Columbian Mild Coffee prices in cents per pound. Also go to www.lock5stat.com and open "StatKey".
- Your instructor will assign you 2 or 3 columns of data. For each column assigned, go to "one quantitative variable" in StatKey, and copy and paste each column in the "edit data" menu. Do not click the box that says "identifier". Do click the box that says "header row". Then push ok.
- Write down the sample mean average of each of your data sets.
- Once each student has their means, they will come up to the white board and put magnets on the number line corresponding to their means. (If magnets are not available, draw a dot with a white-board pen.)
- If there are enough magnets, have each student calculate more means and repeat the process.

Activity#1 Questions:

1. How many numbers were in each of your coffee price samples?

2. What were your sample means calculated with StatKey and put on number line on the white board?

3. Look at the magnets on the white-board. Draw a picture of the number line and magnets. *(This is called a "sampling distribution", a bunch of sample statistics on the same graph. In this case, a bunch of sample means on the same graph.)*

4. Is the shape of the sampling distribution of magnets on the board nearly normal (bell-shaped)?

5. How many total magnets are on the board?

6. The population mean average price for the coffee data census was 136.427 cents per pound. Did all of the sample means you calculated come out to 136.427 cents per pound? *(This is called sampling variability.)*

7. Did all the sample proportions calculated come out to be the same? *(Even if they are not all 136.427 cents per pound) (This is also sampling variability.)*

8. For each of the sample means you calculated, calculate how far off they are from the population mean 136.427 cents per pound. *(This is called margin of error. Your answer may be positive or negative.)*

   Margin of Error = Each Sample Mean – 136.427

Stat Support Activity: Using Coffee Data to Create a Sampling Distribution of Sample Means

9. What is the center of the sampling distribution? Is the center close to the population mean 136.427 cents per pound? *(This is an important discovery. The center of a sampling distribution is close to the population parameter we are looking for!)*

10. The standard deviation of the sampling distribution is called "Standard Error". Let's see if we can estimate the standard error. Draw a bell curve over the magnets. One standard deviation is about half way down the curve where the curve stops being a hill and starts being a valley. This is sometimes called an "inflection point". Estimate the distance from the center to the inflection point. This will be approximate standard error for your sample distribution.

    Approximate Standard Error = ???

11. There is a famous formula for computing standard error for sample means. In this case our population standard deviation is σ = 55.399 and the sample size was *n = 45,* so see if you can calculate the following. Divide 55.399 by the square root of 45. Is your standard error estimate in #11 close to your answer with this formula?

    $$Standard\ Error\ for\ Mean\ Ave\ of\ Coffee\ Price\ Data = \frac{\sigma}{\sqrt{n}} = \frac{55.399}{\sqrt{45}} = ???$$

Activity Part 2 Directions:

- Let's have StatKey take random samples and calculate sample means for us. Go to www.matt-teachout.org and click on the "data sets" menu. Open the "Coffee Data". The first column is the census data of the Columbian mild coffee prices in cents per pound.
- Go to www.lock5stat.com and click on "StatKey". Under "Sampling Distributions" click on "Means".
- Go to "Edit Data", paste the Columbian Mild Coffee Census data column and press "OK".
- Click on the box that says "Choose samples of size n = " and change it to 45. *(Remember our random coffee samples all had a sample size of 45.)*
- Click on "Generate 1 sample". Notice the computer chose a random sample of 45 coffee prices from the census (population), calculated the mean average for that sample of 45 and put a dot on the sampling distribution to represent that mean.
- Click on "Generate 10 samples". Notice the computer chose a 10 random samples, each with 45 coffee prices from the census (population), calculated the mean averages for each of the 10 samples of 45 coffee prices, and put 10 dots on the sampling distribution to represent the 10 mean averages. You should now have a total of 11 dots (11 mean averages on the sampling distribution).
- Click on "Generate 1000 samples". Notice the computer chose a 1000 random samples, each with 45 coffee prices from the census (population), calculated the mean averages for each of the 1000 samples of 45 coffee prices, and put 1000 dots on the sampling distribution to represent the 1000 mean averages. You should now have a total of 1011 dots (1011 mean averages on the sampling distribution).

Stat Support Activity: Using Coffee Data to Create a Sampling Distribution of Sample Means

Activity#2 Questions:

12. Is the shape of the StatKey sampling distribution for sample means nearly normal (bell-shaped)?

13. The population mean average price for the coffee data census was 136.427 cents per pound. Did all of the sample means calculated come out to 136.427 cents per pound? *(This is called sampling variability.)*

14. Did all the sample proportions calculated come out to be the same? *(Even if they are not all 136.427 cents per pound) (This is also sampling variability.)*

15. What is the center of the sampling distribution? *(This is located under "mean" on the top right corner of the sampling distribution in StatKey. This is the mean average of all 1011 sample means.)* Is the center close to the population mean of 136.427 cents per pound? *(This is an important discovery. The center of a sampling distribution is close to the population parameter!)*

16. What is the standard error for the sampling distribution? *(This is located on the top right corner of the sampling distribution in StatKey.)* Is the standard error close to the formula answer we calculated in #11?