ACTIVITY 3: Worksheet Answers

Oftentimes, researchers want an estimate for the population mean so they select a representative sample of the population and calculate the sample mean as an estimate.

Definition: The sample mean provides an estimate for the average value (mean) of the population data.

To complete this activity, please use the CIA data your group collected regarding life expectancy in Activity 2.

Part 1: Compute the sample mean and sample standard deviation as indicated below.

a) Use the given formula to calculate an estimate for the mean of life expectancy for all countries. Only use your CIA data from the simple random sample of five countries.

Note: If one of your sample countries did not have data, your sample size will be smaller.

Formula: Sample mean

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$$

n is the sample size.
$x_1, x_2, x_3, \ldots, x_n$ refers to your data values.

The formula shows that you add the data values up and then divide the sum by your sample size.

Recall: If one of your sample countries did not have data, you will need to divide by the number of data values you collected.

Answers will vary. However, for the example data given in Activity 2, the solution is as follows.

The sample mean is $\bar{x} = \frac{74.65+67.42+69.67+81.66+65.30}{5} = 71.74$ years

Researchers are also often interested in how much variation there is in the data. To measure this spread in the data, they compute and analyze the standard deviation.

Definition: The sample standard deviation provides an estimate of how much each data value differs from the sample mean. That is, it measures the spread of the data.

Note: If the data values are more spread out, they will differ more from the mean resulting in a larger standard deviation.

Formula: Sample Standard deviation

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots\ldots + (x_n - \bar{x})^2}{n-1}} = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

b) Use the formula and instructions below to calculate an estimate for the standard deviation of life expectancy for all countries. Only use your CIA data from the simple random sample of five countries.

This formula is a bit more involved. You can do this in steps by using the table below.

Answers will vary. However, for the example data given in Activity 2, the solution is as follows.

Step 1: Write down your data values	**Step 2: Find $(x_i - \bar{x})$ for each data value.** This means you should subtract the sample mean you found from each of your data values.	**Step 3: Find $(x_i - \bar{x})^2$ for each data value.** This means you should square each value you found in Step 2 (Column 2). **Note:** Would it make sense for any of these values to be non-negative?
74.65	(74.65 – 71.74) = 2.91	$(2.91)^2$ = 8.4681
67.42	(67.42 – 71.74) = -4.32	$(-4.32)^2$ = 18.6624
69.67	(69.67 – 71.74) = -2.07	$(-2.07)^2$ = 4.2849
81.66	(81.66 – 71.74) = 9.92	$(9.92)^2$ = 98.4064
65.30	(65.30 – 71.74) = -6.44	$(-6.44)^2$ = 41.4736

Step 4: Find the sum of all the values you found in Step 3 (Column 3): $\sum (x_i - \bar{x})^2$ = 171.2954

Answers will vary. However, for the example in Activity 2, the solution is provided.

<u>Step 5</u>: Divide the sum you found above by (*n*-1): $\frac{1}{n-1}\sum(x_i - \bar{x})^2$= 171.2954/(5-1) = 42.82385

Answers will vary. However, for the example given in Activity 2, the solution is provided.

<u>Step 6</u>: Take the square root of the above: $\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$ = $\sqrt{42.82385} = 6.544$

Answers will vary. However, for the example given in Activity 2, the solution is provided.

The sample standard deviation is *s* = 6.544 years

Part 2: We will now explore how to compute the sample mean and sample standard deviation using Excel. This is particularly helpful when you have data sets with a lot of values.

a) Use Excel and the instructions below to calculate an estimate for the mean life expectancy for all countries. Use your CIA data from the simple random samples of 15 and 30 countries.

Steps: Sample mean

1. Open Excel and type your collected data for the sample size of 15 countries into Column 1, and 30 countries into Column 2.
2. Save your data to use in Activity 5.
3. Click the box below your first set of data.
4. Select "Formulas" at the top of the page.
5. Select "More Functions" followed by "Statistical" followed by "AVERAGE."
6. For Number 1 you should see the following set up:
   A1:A15 (This might differ a little if you had some non-response.*)
7. Click "OK" and the mean should show in the box below the data.
8. Repeat the process for Column 2 (Remember you have more data, so you should see B1:B30 - Again, this might differ a little if you had some non-response.)

Answers will vary. However, for the example given in Activity 2, the solution is provided.

Sample Mean for *n*=15: $\bar{x}$ = 71.69 years
Sample Mean for *n*=30: $\bar{x}$ = 72.87 years

b) Use Excel and the instructions below to calculate an estimate for the standard deviation of life expectancy for all countries. Use your CIA data from the simple random samples of 15 and 30 countries.

Steps: Sample Standard Deviation

1. Click the box below the mean that was calculated for Column 1.
2. Select "Formulas"
3. Select "More Functions" followed by "Statistics" followed by "STDEV.S"
4. For Number 1, you need to type in the following:
   A1:A15 (This might differ a little if you had some non-response.)
5. Click "OK" and the standard deviation should show in the box below the mean.
6. Repeat the process for Column 2 (Remember you have more data, so you should type in B1:B30 - Again, this might differ a little if you had some non-responses.)

Answers will vary. However, for the example given in Activity 2, the solution is provided.

Sample Standard Deviation for *n*=15: *s* = 6.495 years

Sample Standard Deviation for *n*=30: *s* = 7.259 years

As each group shares their calculated sample mean, record the values in the table below. Answers will vary and the number of sample means will be dependent on the class size as each group will share their calculated sample mean. The data below was collected from 17 simple random samples of *n* countries.

Sample Means

***n* = 5**		***n* = 15**		***n* = 30**
71.74		71.69		72.87
72.81		75.71		73.33
76.55		75.91		72.78
73.89		72.98		74.06
73.44		71.74		74.20
70.41		73.68		73.47
77.10		74.21		73.71
79.83		72.87		75.30
77.87		73.83		75.50
76.58		77.60		75.64
73.70		73.08		74.39
71.04		70.49		72.99
71.31		71.76		72.71
76.33		75.70		74.68
71.66		74.38		74.35
77.83		72.16		73.29
75.71		74.63		73.68

The population mean and standard deviation were calculated using the complete list of countries provided in class and the CIA website (see below).

Population Mean Life Expectancy: $\mu = 73.76$ years
Population Standard Deviation Life Expectancy: $\sigma = 6.95$ years

Refer to the information on the previous page to answer the following questions.

1. How did your sample mean and sample standard deviation compare to the population mean and population standard deviation for n=5? n=15? n=30?
   Most likely none of the sample means will equal the population mean given.

2. Did every group get the same sample mean for n=5? n=15? n=30?
   Most likely all the sample means will be different.

3. How much did the sample means in the table vary from one another for n=5? n=15? n=30?
   The means in the n=5 sample size group varied the most, followed by the sample means in the n=15 sample size group with the means in the n=30 sample size group varying the least.

4. How does the sample size impact the amount of variability in the sample means?
   The larger the sample size, the lower the variability from one calculation of the sample mean to the next.

5. Why would choosing a larger sample size be better than a smaller one?
   The calculated sample mean would be closer to the population mean regardless of the sample selected.

6. Do you think that with a simple random sample, it is possible to get the n countries that have the lowest life expectancy?
   Yes, it is possible since with the SRS, every sample of size n has the same chance of being selected.

7. How likely do you think it is that you would get the n countries with the lowest life expectancies in your sample?
   With a larger sample size, it is a lot less likely for this to happen because of the number of possible samples of size n.

Based on the concepts of sampling variability and sampling error, what question do you think you should ask yourself when reading statistics online?

You should ask how large the sample size is. Smaller sample sizes will cause more variability, so the estimate will lack precision. The lack of precision means that your estimate could have substantial sampling error.