ACTIVITY 2: Worksheet Answers (Example)

You are now ready to collect your own simple random sample (SRS)!

Take an SRS of 5 countries, another SRS of 15 countries, and one more SRS of 30 countries and fill in the tables on the next page. Then, find the data from the following websites on life expectancy for 2022.

https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy

Note: To alphabetize the data, click on the arrow in the 1st column labeled "Countries & Territories" on the website.

https://www.cia.gov/the-world-factbook/about/archives/2022/field/life-expectancy-at-birth/country-comparison

Note: Clicking "Control-F" will allow you to search for the countries more quickly.

Record the data you collect on the next page.

Each group of students will collect their own data and hence answers will vary. However, we provide a specific example to demonstrate the process.

Sample *n*=5:

Random Number and Country	Life Expectancy	
	Wiki	CIA
108 – Marshall Islands	65.15	74.65
28 - Burundi	61.98	67.42
132 - Pakistan	66.43	69.67
82 - Ireland	83.06	81.66
48 - Djibouti	62.86	65.30

Sample *n*=15:

Random Number and Country	Life Expectancy	
	Wiki	CIA
159 - Slovenia	81.28	81.82
167 - Sudan	65.58	67.12
7 - Argentina	76.06	78.31
111 - Mexico	74.83	72.32
3 - Algeria	77.13	78.03
46 – Democratic Republic of Congo	59.74	61.83
190 - Vanuatu	70.49	75.14
139 - Philippines	72.19	70.14
118 - Mozambique	59.62	57.10
103 - Malawi	62.90	72.44
39 - Comoros	63.68	67.20
66 - Ghana	63.95	69.37
38 - Colombia	73.66	74.89
117 - Morocco	74.97	73.68
24 - Brazil	73.42	75.92

Sample *n*=30:

Random Number and Country	Life Expectancy	
	Wiki	CIA
131 - Oman	73.94	76.90
183 - Uganda	63.64	68.96
65 - Germany	80.71	81.51
133 - Palau	N/A	74.64
27 – Burkina Faso	59.77	63.44
103 - Malawi	62.90	72.44
59 - Fiji	68.31	74.27
174 - Thailand	79.68	77.66
50 – Dominican Republic	74.17	72.56
167 - Sudan	65.58	67.12
104 - Malaysia	76.26	63.13
46 – Democratic Republic of Congo	59.74	61.83
178 – Trinidad and Tobago	74.71	75.94
179 - Tunisia	74.26	76.82
109 - Mauritania	64.69	65.22
102 - Madagascar	65.23	68.17
153 - Senegal	67.91	69.96
67 - Greece	80.64	81.49
78 - India	67.74	67.22
143 - Romania	75.30	75.75
165 - Spain	83.03	82.55
142 - Qatar	81.56	79.81
3 - Algeria	77.13	78.03
88 - Kazakhstan	74.44	72.53
21 - Bolivia	64.93	72.50
157 - Singapore	82.90	86.35
33 - Canada	81.30	83.80
74 – Holy See	N/A	N/A
97 - Liberia	61.10	65.45
118 - Mozambique	59.62	57.10

When you are done finding your data, answer the following questions:

1. Did both websites give you the same value?
    No, the values are very different for the two websites.

2. Do you think the values from one website are more accurate than the values from the other?
    Yes. The CIA.gov website is most likely more accurate than Wikipedia.

3. For the website that you felt was more accurate, do you think the values given are 100% correct? If not, why?
    It is unlikely that any website would have 100% accurate data for all countries, especially for developing nations that lack the infrastructure to keep good records of births and deaths.

4. Did the website explain how they found the data and if not, how do you think they found those values?
    Wikipedia stated that finding life expectancies is difficult due to the disparities in data reporting and collection across countries. They obtained their data from a different source. The CIA Factbook did not talk about how they obtained their data.

5. Do you think the list that you received in class contained all countries in the world?
    No, the list that we were given had 195 countries whereas the two websites listed more countries than 195.

6. Did you get a country with no information available on life expectancy from either of the websites?"

    Answers will vary based on the data each group obtains.

    For the example provided here, we note that Holy See did not show any information on either website.

Let's consider three types of bias that can occur during the sampling process. Read the definition of each type of bias and discuss whether you encountered it.

Undercoverage bias:
If the sampling design systematically excludes a portion of the population, undercoverage sampling bias may be introduced into the study.

Note: The list that you were given had 195 countries whereas the two websites listed more countries than 195. According to the United Nations, there are 195 recognized countries in the world, although, if all Independent Nations-States, dependent Nations, Antarctica, and other areas were included, then there would be 253 countries.

If you consider all 253 countries as mentioned above or the number of countries listed on one of the websites as having a more appropriate count of the number of countries compared to the list of 195 given, there would be substantial undercoverage bias. The 58 countries that were not included on the population list given would have a 0% chance of being selected into the sample. If these countries differ with respect to life expectancy to the countries on the list, undercoverage bias would be present. The best way to avoid this type of bias is to make sure all units/individuals you are extending your results to are included on the list you take the samples from.

Non-response bias:
Non-response bias may occur when information cannot be found on a selected unit/individual either because the information is not available for that unit/individual, the individual cannot be contacted, or the individual refuses to cooperate.

Answers will vary as to whether the student encountered non-response bias.

For the sample collected for this example, there was no information for Holy See on either website. It is unlikely that not finding information for one country would lead to non-response bias.

Additional Teacher Notes: Non-response bias is more prevalent when individuals are being surveyed. Offering incentives to fill out surveys will help reduce non-response. Using multiple websites to find information is not recommended as sometimes data is not measured in the same way. A common example is birthrate which can be measured as a crude birth rate or total fertility rate which are very different. For example, the crude birthrate is the number of live births per 1,000 women, whereas the fertility rate may be computed as the annual number of births per woman who are at childbearing age.

Response bias: When a respondent lies or a recording error occurs for a unit/individual, response bias is present. The behavior of the interviewer or the wording of a question can cause response bias.

There was a substantial difference in life expectancies between the two websites. Although the CIA website is likely more accurate than Wikipedia, it is bound to have some errors. For a country to accurately record its life expectancy, the record-keeping would need to be very good, and the process would be extremely time-consuming. It is unlikely that data collected from countries will contain no errors.

Note to teacher: Often students confuse the undercoverage and response bias because they do not consider how the unit is defined in their sample. If individuals within the country did not participate when the country was determining their life expectancy, that would lead to an incorrect value for that country which would be a response bias and not undercoverage. The unit is a country, so if all countries are on the list that the sample was taken from, then there is no undercoverage bias.

Based on what you learned today, what additional questions do you think you should ask yourself when reading statistics online?

1. What is the intended population, and would it be easy to get a complete list of all units in the intended population?
2. What was the non-response rate?
3. How easy would it be to get accurate measures from the units? Is response bias likely?