**Assignment #2: Doing Some Basic Data Analysis in Excel**

Download “wapo-data.xlsx” from Canvas. This file contains information about 1,019 people fatally shot by on-duty police officers in 2020, per *The Washington Post.*

1. Open the file in Microsoft Excel.
2. Create a frequency distribution for the “race” variable (Column F). Do this by highlighting Column F, clicking “Insert” on the toolbar, selecting “Recommended Charts,” and clicking “OK” in the box that pops up. Record the **percentage** for each category in the **second** column of the table on the next page.

*Hint: To get the percentage, you need to divide each “Count” by the total (in this case, 1,019) and multiply by 100.*

1. In cell A2, type “=RAND( )” and press enter. A random decimal between 0 and 1 should appear.
2. Now, copy cell A2 and paste it to cells A3:A1020. This should insert random decimals for the rest of Column A.
3. Highlight the entire dataset (cells A1:P1020). Right click anywhere in the selection, then select “Sort” and “Sort Smallest to Largest.” This should randomly sort the dataset (note that Column B, which was previously numbered sequentially from 1 to 1019, should now be shuffled).
4. Copy the first **11 rows** of the data (A1:P11) and paste them into a new worksheet. This new worksheet (Sheet2) will represent a random sample of 1% of the data. Using this worksheet, repeat Step #2. Record the percentage for each category in the **third** column of the table on the next page.
5. Go back to the original dataset (Sheet1). Copy the first **101 rows** of the data (A1:P51) and paste them into a new worksheet. This new worksheet (Sheet3) will represent a random sample of 10% of the data. Using this worksheet, repeat Step #2. Record the percentage for each category in the **fourth** column of the table on the next page.
6. Go back to the original dataset (Sheet1). Copy the first **301 rows** of the data (A1:P101) and paste them into a new worksheet. This new worksheet (Sheet4) will represent a random sample of 30% of the data. Using this worksheet, repeat Step #2. Record the percentage for each category in the **last** column of the table on the next page.
7. Look at the four columns in the table you’ve created. How did sample size influence your results?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Population  (N=1019) | 1% Sample  (N=10) | 10% Sample  (N=100) | 30% Sample  (N=300) |
| Asian |  |  |  |  |
| Black |  |  |  |  |
| Hispanic |  |  |  |  |
| Native |  |  |  |  |
| Other |  |  |  |  |
| White |  |  |  |  |
| Undetermined |  |  |  |  |