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

Download “wapo-data.csv” from Canvas and save it somewhere on your computer. This file contains information about 1,019 people fatally shot by on-duty police officers in 2020, per *The Washington Post’s* [Fatal Force Database](https://github.com/washingtonpost/data-police-shootings).

1. First, [download JASP](https://jasp-stats.org/download/).
2. Open wapo-data.csv. You should see something like this:

A screenshot of a computer

Description automatically generated

1. Create a frequency distribution for the “race” variable. To do this, click “Descriptives” up at the top, then highlight “race” from the box (on the left) and move it to the “variable” box (on the right). Once you’ve done that, click “Tables” and select “Frequency tables.”

Graphical user interface, application

Description automatically generated

JASP will produce a table. From this table, record the percentage for each category in the **second** column of the table on the last page of this assignment.

1. Next, you need to draw a random sample of **1% of the data**. To do this, click the black **+** symbol near the top right corner of your screen. The following box should pop up:

Graphical user interface, text, application, chat or text message

Description automatically generated

Type “Random1” in the name box, select “Nominal,” and click “Create Column.” The top of your screen should now look like the screenshot below. Over on the right-hand side, scroll down until you see “binomDist(y).” Select it.

Graphical user interface, text, application

Description automatically generated

The following formula should appear in the box:



Click “trials” and type “1.” Now click “prob” and type “.01” (for 1 percent). Once you’ve done that, click “Compute column” (see screenshot above). This will create a new variable, which you’ve titled “Random1,” containing approximately 1,009 “0s” and approximately 10 “1s.”

1. Filter the data on “Random1.” Do this by clicking the variable’s name, which is atop the right-most column in the data.



The following should pop up:



Click the checkmark next to “0.” It should change to an “x” and 99% of your dataset should now appear to be grey instead of black. Now, repeat step 3, and it should give you the same table as before, only this time the entries are based on the 1% random sample you just drew. From this table, record the percentage for each category in the **third** column of the table on the last page of this assignment.

1. Repeat steps 4 and 5 – but this time draw a random sample of **10% of the data**. Remember to record the percentage for each category in the **fourth** column of the table on the last page of this assignment.

*Hint: First, turn off your current filter by clicking the “x” next to “0” so that it changes back to a checkmark. Name your new filter variable “Random10” and be sure to type “.10” for prob instead of “.01.”*

1. Repeat steps 4 and 5 – but this time draw a random sample of **30% of the data.** Remember to record the percentage for each category in the **fifth** column of the table on the last page of this assignment.

*Hint: First, turn off your current filter by clicking the “x” next to “0” so that it changes back to a checkmark. Name your new filter variable “Random30” and be sure to type “.30” for prob instead of “.01.”*

1. Look at the four columns in the table you’ve created. How did sample size influence your results? Which sample most closely mirrored the population in terms of the distribution of race?

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

*\* Note: The “~” means “approximately.” Because they were randomly drawn, your samples may not equal precisely 10, 100, or 300 – but they should be very close.*