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

Download “opd\_incidents.csv” from Canvas and save it somewhere on your computer. This file contains basic, de-identified information about 52,245 criminal incidents that were reported to the Omaha Police Department in 2023.[[1]](#footnote-1)

1. First, [download JASP](https://jasp-stats.org/download/) and start it up on your computer.
2. Click “So open a data file and take JASP for a spin!” and open **opd\_incidents.csv**. Once it’s open, you should see something like this:



1. Create a frequency distribution for the **incident\_type** variable. To do this, click “Descriptives” up at the top. Your screen should now look like the screenshot on the next page. Highlight **incident\_type** 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.”



JASP will produce a table. From this table, record the **percent** 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 **0.1% of the data**. To do this, navigate back to the data (using the on the left side of the window) and click the black **+** near the top right corner of the window. The following box should pop up:



Type “Random\_001” in the name box, select “Nominal,” and click “Create Column.” **Random\_001** will now appear at the top of the rightmost column in the data. Double-click it. 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.



The following formula should appear in the box:



Click “trials” and type “1.” Now click “prob” and type “.001” (for 0.1 percent). Once you’ve done that, click “Compute column” (see screenshot above). This will create a new variable, which you’ve titled **Random\_001**, containing approximately 52,193 “0s” and approximately 52 “1s.”

1. Filter the data on “Random\_001.” Do this by clicking the  near the upper-left corner of the window. The following should appear at the top of the JASP window:



Scroll to the bottom of the variable list (on the left side) until you see **Random\_001**. Click it and it will appear in the “drag and drop filter” box in the center. Then, click the “**=**”above the box. Three dots should appear in the box, to the right of **=** . Click the three dots and type “1”, then click **Apply pass through filter** at the bottom. You should see a message that says “Filter applied” and most of the data should now appear in grey, rather than black, text. Now, repeat step 3, and it should give you the same table as before, only this time the entries are based on the 0.1% random sample you just drew. From this table, record the **percent** 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 **1% of the data**. Record the **percent** for each category in the **fourth** column of the table on the last page of this assignment.

*Hint: First, remove your current filter by dragging “Random\_001 = 1” to the trash can in the bottom right corner of the drag and drop filter box, then clicking “Apply pass-through filter.” You should see a message that says “Filter cleared.” Name your new filter variable “Random\_01” and be sure to type “.01” for prob instead of “.001”.*

1. Repeat steps 4 and 5 – but this time draw a random sample of **10% of the data.** Record the **percent** for each category in the **fifth** column of the table on the last page of this assignment.

*Hint: First, remove your current filter by dragging “Random\_01 = 1” to the trash can in the bottom right corner of the drag and drop filter box, then clicking “Apply pass-through filter.” You should see a message that says “Filter cleared.” Name your new filter variable “Random\_05” and be sure to type “.05” 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 incident types? Answer these questions in the space below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Incident Type | Population(N=52,245) | 0.1% Sample(N= ~52) | 1% Sample(N= ~522) | 5% Sample(N= ~2,662) |
| Homicide |  |  |  |  |
| Sexual assault |  |  |  |  |
| Robbery |  |  |  |  |
| Aggravated assault |  |  |  |  |
| Simple assault |  |  |  |  |
| Burglary/theft |  |  |  |  |
| Motor vehicle theft |  |  |  |  |
| Arson |  |  |  |  |
| Other |  |  |  |  |

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

1. The raw data were downloaded from [this webpage](https://police.cityofomaha.org/crime-information/incident-data-download) on 1/15/2014. For this assignment, I’ve already cleaned the data. [↑](#footnote-ref-1)