Sample Sizes - Part 1

When working with data there are some fundamental concepts that we must learn. This week, we are going to explore sample sizes and begin to develop our thinking around statistics, with concrete examples during our code alongs.

Using the Adult Data set we began working with, you will create some datasets as we explore sampling. Adult Data Set can be found here: **https://archive.ics.uci.edu/ml/datasets/adult**

This will be an iterative process.

Ideally, you would generate the code in python (to learn the syntax and structure of the language) and collect your generated data from the Adult Dataset into a csv. You can use google sheets and save it as a csv.

You should explore the connection between google drive, google colaboratory, and develop an efficient workflow for storing and accessing your datasets.

Directions:

1. Collect a sample size of 100, five times
2. Collect a sample size of 1000, five times
3. Collect a sample size of 10% of the data, five times
4. Collect a sample size of 20% of the data, five times
5. Collect a sample size of 33.3% of the data, five times

Collect (write down in a spreadsheet) data for the following categories. These should be float values, a minimum of two decimal places.

Lab (submit by 12/6/2021, end of class)

- HS-grade
- Some-college
- Bachelors
- Assoc-voc
- Prof-school

A google-colab showing your derived code for each sample is required.

```
1 samp_2 = pd.DataFrame(df[['education']].sample(100))
2 samp_2.value_counts(normalize=True)
```

```
education
 HS-grad          0.32
 Some-college     0.21
 Bachelors        0.17
 7th-8th          0.05
 Masters          0.04
 5th-6th          0.04
 11th             0.04
 10th             0.04
 Prof-school      0.02
 Assoc-voc        0.02
 Assoc-acdm       0.02
 12th             0.02
 1st-4th          0.01
dtype: float64
```

```
[ ]  1 samp_3 = pd.DataFrame(df[['education']].sample(100))
     2 samp_3.value_counts(normalize=True)
```

```
education
 HS-grad          0.34
 Some-college     0.16
 Bachelors        0.16
 Masters          0.07
 10th             0.06
 11th             0.05
 Assoc-voc        0.04
 12th             0.04
 Doctorate        0.03
 5th-6th          0.02
```

You will enter your data on this form for the Lab

Homework (submit by 12/8/2021, 5:30 pm)

- Masters
- Assoc-acdm
- Doctorate

You will enter your data on this form for the Homework