KEY_Practice15_Intro_Stats_II

August 28, 2019

1 Practice with Statistics (Part 2)!

Remember: * Count statistics are a useful way of summarizing the items in a set of measurements. * `Counter` provides a useful class for counting lists of items. * Percentages tell you what fraction of a list consists of a given category.

First, import numpy and pandas and Counter:

```
[0]: # load numpy and pandas and Counter

     import numpy as np
     import pandas as pd
     from collections import Counter
```

```
[2]: # mount Google Drive
     from google.colab import drive
     drive.mount('/content/gdrive')
     path = '/content/gdrive/My Drive/SummerExperience-master/'
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive", force_remount=True).

Load in the sample data from the Lesson:

```
[0]: # load the csv file: 'SampleData/detroit_weather.csv'

     data_table = pd.read_csv(path + 'SampleData/detroit_weather.csv')
```

```
[4]: # Print the beginning of the table using the head function to remind you of the␣
     ↪format:

     data_table.head()
```

```
[4]:    Unnamed: 0  YEAR  MONTH  DAY   Rain   Snow
    0           0  1950      1    1   True  False
    1           1  1950      1    2   True  False
    2           2  1950      1    3   True  False
    3           3  1950      1    4   True   True
    4           4  1950      1    5  False  False
```

During the lesson, we looked at the rates of snow occurance, now we will repeat the same analysis for the occurance of rain.

```
[5]: # Count the number of days that have been raining since 1950
     # and the number of days that haven't been

     raining = Counter(data_table["Rain"])
     raining
```

```
[5]: Counter({False: 17263, True: 8051})
```

```
[6]: # What percentage of days since 1950 have been spent raining?

     raining[True] / (raining[True] + raining[False])
```

```
[6]: 0.3180453503989887
```

```
[7]: # How man days have been spent raining AND snowing?
     # HINT: use a `and` statement in pandas

     len(data_table.query('Rain and Snow'))
```

```
[7]: 1095
```

```
[8]: # What percentage of days have been spent raining AND snowing?

     len(data_table.query('Rain and Snow')) / len(data_table)
```

```
[8]: 0.043256695899502255
```

```
[9]: # Calculate the percentage of days during the month you were born that were␣
     ↪spent raining:

     june_rain = Counter(data_table.query('MONTH == 6')["Rain"])
     june_rain[True] / (june_rain[True] + june_rain[False])
```

```
[9]: 0.3685990338164251
```

CHALLENGE In the next lesson, we will look at climate change between the early 20th century and today, can you calculate a difference in days spent snowing between the 1950's and 2000's?

```
[10]: # Calculate a change in the percentage of days spent snowing
      # during the 1950's and 2000's


      snow_1950 = Counter(data_table.query('YEAR < 1960')["Snow"])
      snow_2000 = Counter(data_table.query('YEAR < 2010 and YEAR >= 2000')["Snow"])
```

```
print("Snow days in 1950's", snow_1950[True] / (snow_1950[True] +␣
 ↪snow_1950[False]))
print("Snow days in 2000's", snow_2000[True] / (snow_2000[True] +␣
 ↪snow_2000[False]))
```

```
Snow days in 1950's 0.20208105147864183
Snow days in 2000's 0.16374589266155531
```

By how much did the percentage change from the 1950's to the 2000's? Did it increase or decrease?

Answer: $0.202 - 0.164 = 0.038 =$ decreased 3.8%

Nice job! You just practiced:

- Turning categorical variables into counts using `Counter`
- Calculating percentages from count variables
- Interpreting the results from basic statistical analysis