

# KEY\_Practice15\_Intro\_Stats\_II

July 12, 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:

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

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

```
[5]: # mount Google Drive
```

```
from google.colab import drive
drive.mount('/content/gdrive')
path = '/content/gdrive/My Drive/SummerExperience-master/'
```

```

      □
↳ -----
ModuleNotFoundError                                Traceback (most recent call↳
↳last)

<ipython-input-5-b958c7a1dd08> in <module>
      1 # mount Google Drive
----> 2 from google.colab import drive
      3 drive.mount('/content/gdrive')
      4 path = '/content/gdrive/My Drive/SummerExperience-master/'

ModuleNotFoundError: No module named 'google'
```

Load in the sample data from the Lesson:

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

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

```

↳ -----

NameError                                Traceback (most recent call↳
↳ last)

<ipython-input-6-02a7bde06426> in <module>
      1 # load the csv file: 'Lessons/SampleData/detroit_weather.csv'
      2
----> 3 data_table = pd.read_csv(path + 'Lessons/SampleData/detroit_weather.
↳ csv')

NameError: name 'path' is not defined
```

[1]: *# Print the beginning of the table using the head function to remind you of the*  
*↳ format:*

```
data_table.head()
```

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

[2]: *# 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
```

[3]: *# What percentage of days since 1950 have been spent raining?*

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

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

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

```

↳ -----

NameError                                Traceback (most recent call↳
↳ last)
```

```

<ipython-input-7-ce803845dfba> in <module>
    2 # HINT: use a `and` statement in pandas
    3
----> 4 len(data_table.query('Rain and Snow'))

```

NameError: name 'data\_table' is not defined

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

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

```
[ ]: 0.043256695899502255
```

```
[ ]: # 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])
```

```
[ ]: 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?

```
[ ]: # 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