3.1.5 Thinkful Drill - Describing Data ¶

This is a stats drill designed to describe a data via the Brady Bunch population. Calculations were done by hand first, then calculations were performed using Python.

Below are the questions asked

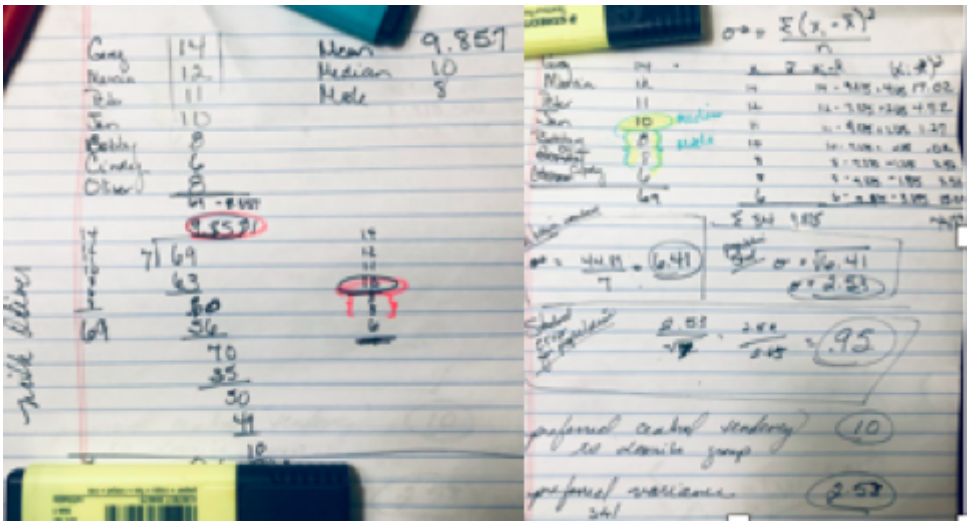
1. Greg was 14, Marcia was 12, Peter was 11, Jan was 10, Bobby was 8, and Cindy was 6 when they started playing the Brady kids on The Brady Bunch. Cousin Oliver was 8 years old when he joined the show. What are the mean, median, and mode of the kids' ages when they first appeared on the show? What are the variance, standard deviation, and standard error?
2. Using these estimates, if you had to choose only one estimate of central tendency and one estimate of variance to describe the data, which would you pick and why?
3. Next, Cindy has a birthday. Update your estimates- what changed, and what didn't?
4. Nobody likes Cousin Oliver. Maybe the network should have used an even younger actor. Replace Cousin Oliver with 1-year-old Jessica, then recalculate again. Does this change your choice of central tendency or variance estimation methods?
5. On the 50th anniversary of The Brady Bunch, four different magazines asked their readers whether they were fans of the show. The answers were: TV Guide 20% fans Entertainment Weekly 23% fans Pop Culture Today 17% fans SciPhi Phanatic 5% fans

> Based on these numbers, what percentage of adult Americans would you estimate were Brady Bunch fans on the 50th anniversary of the show?

In [53]:
```
# our boilerplate

import pandas as pd
import numpy as np
```

Initial standard calculations of the Brady's ages



In [63]:
```
#First I created a data frame

bb_df = pd.DataFrame()

bb_df['name'] = ["Greg", "Marcia", "Peter", "Jan", "Bobby", "Cindy", "Oliver
bb_df['age'] = [14, 12, 11, 10, 8, 6, 8]
bb_df.set_index("name")
```

Out[63]:

name	age
Greg	14
Marcia	12
Peter	11
Jan	10
Bobby	8
Cindy	6
Oliver	8

In [64]:
```
bb_df.describe()
```

Out[64]:

	age
count	7.000000
mean	9.857143
std	2.734262
min	6.000000
25%	8.000000
50%	10.000000
75%	11.500000
max	14.000000

In [65]:
```
print (bb_df.median())
print (bb_df.mode())
```

```
age    10.0
dtype: float64
     name  age
0   Bobby  8.0
1   Cindy  NaN
2    Greg  NaN
3     Jan  NaN
4  Marcia  NaN
5  Oliver  NaN
6   Peter  NaN
```

In [66]:
```
print (bb_df.var())
```

```
age    7.47619
dtype: float64
```

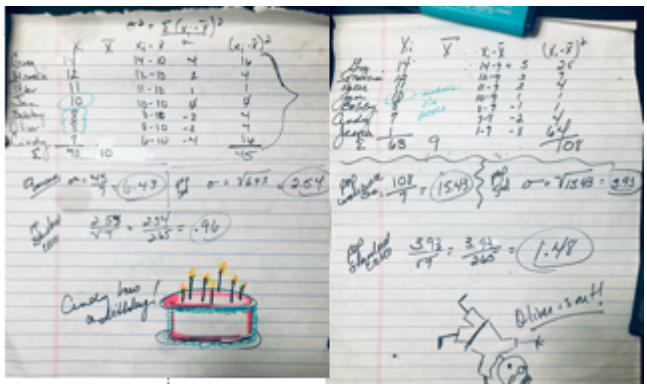
My inital calculations by hand were not far off.

- Mean at 9.857
- Median at 10
- Mode at 8

However, on the second set of calulations (σ2, σ, and standard error), I had transposed the numbers and calculated using a mean of 9.875 (typical).

Calculations with Cindy's Birthday & Oliver replaced by Jessica

The mode did not change when Cindy turned 7 nor did the median. However, when Oliver left all the calculations changed including the mode and median, the averger was changed significantly as was the std.



In [67]:
```
bb_bday_df = bb_df.replace(6,7)
```

In [68]:
```
print(bb_bday_df)
```

```
     name  age
0    Greg   14
1  Marcia   12
2   Peter   11
3     Jan   10
4   Bobby    8
5   Cindy    7
6  Oliver    8
```

In [69]:
```
bb_bday_df.describe()
```

Out[69]:

	age
count	7.000000
mean	10.000000
std	2.516611
min	7.000000
25%	8.000000
50%	10.000000
75%	11.500000
max	14.000000

In [70]:
```
bb_bday_df.var()
```

Out[70]:
```
age    6.333333
dtype: float64
```

In [71]:
```
bb_jessica = bb_bday_df
bb_jessica.drop([6], axis=0)
bb_jessica.append({"name":"Jessica", "age":1}, ignore_index=True)
```

Out[71]:

	name	age
0	Greg	14
1	Marcia	12
2	Peter	11
3	Jan	10
4	Bobby	8
5	Cindy	7
6	Oliver	8
7	Jessica	1

In [72]:
```
bb_jessica.describe()
```

Out[72]:

	age
count	7.000000
mean	10.000000
std	2.516611
min	7.000000
25%	8.000000
50%	10.000000
75%	11.500000
max	14.000000

In [73]:
```
#calculating the standard error after Cindy's birthday
np.std(bb_bday_df['age'])/np.sqrt(len(['age']))
```

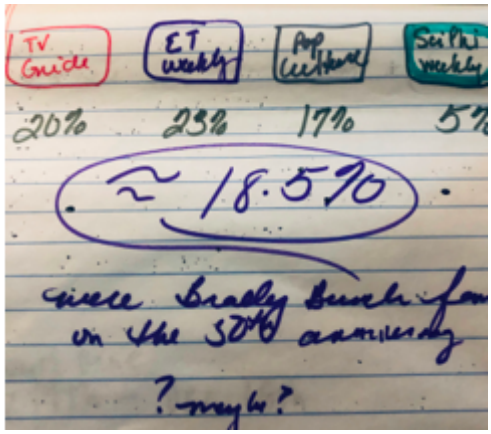
Out[73]: 2.32992949004287

In [74]:
```
#calculating the standard error with the replacement of Oliver w/ Jessica
np.std(bb_jessica['age'])/np.sqrt(len(["age"]))
```

Out[74]: 2.32992949004287

50th Anniversary Show and stats

Totally spitballing this, the overall Brady Bunch fans were around 18.5% when the 50th anniversary show aired. Loosely finding a median (with percentages) between 17% and 20%. I think the SciPhi Weekly readership would be considered an outlier. They most likely have a small subscription base, so the percentage of fans may not as accurately be representated as they would be in more pop-culture periodicals.



In [ ]:
```

```