lab1_writeup_chochan2

September 29, 2020

0.1 Lab 1 - Number estimation (free exploration)

For this lab, please work independently. The required Python packages have been imported for you below.

[1]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

The target number (i.e. ground truth) for each experimental trial is provided in the following python array.

[2]:
```
targets = np.array([3, 8, 40, 2, 5, 30, 7, 35, 6, 15, 10, 20, 9, 25, 4]);
```

Read in the experimental data collected in-class. `df` is a dataframe of size *(Participants x Trials)*.

[3]:
```
df = pd.read_csv('data-number-estimation.csv')
```

Display the content of the data frame.

[4]:
```
df
```

[4]:

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	\
0	3	8	27	2	5	25	8	30	
1	3	8	24	2	5	25	7	25	
2	3	8	30	2	5	40	7	30	
3	3	8	30	2	5	25	7	25	
4	3	9	27	2	5	24	7	35	
5	3	9	36	2	6	38	8	32	
6	3	7	25	1	5	25	7	30	
7	3	8	30	2	6	30	8	30	
8	3	8	30	2	5	25	6	40	
9	3	8	25	2	5	30	7	27	
10	3	8	26	2	5	19	7	30	
11	3	8	22	2	5	26	7	25	
12	3	8	20	2	5	20	7	25	
13	3	9	32	2	5	27	9	38	
14	3	9	40	2	5	50	7	40	
15	3	8	39	2	5	40	7	40	

16	3	7	23	2	5	26	7	25
17	3	8	50	2	5	30	6	30
18	3	8	25	2	5	30	7	35
19	3	9	20	2	6	20	7	30
20	3	7	25	2	4	20	7	30
21	3	8	23	2	5	25	7	29
22	3	8	40	2	5	20	7	25
23	3	9	27	2	5	38	7	38
24	3	8	31	2	5	25	7	50
25	3	8	27	2	5	25	7	22
26	3	7	35	2	5	20	7	25
27	3	10	30	2	6	27	8	23
28	3	8	27	2	5	21	7	29
29	3	10	8	2	5	25	8	30
30	3	8	35	2	5	30	7	35
31	3	9	40	2	5	30	7	30
32	3	9	30	2	5	40	10	35
33	3	10	30	2	5	25	7	28
34	3	8	47	2	5	26	7	45
35	3	8	20	2	5	15	9	18
36	3	7	21	2	5	25	7	19
37	3	8	27	2	5	25	8	30
38	3	8	30	2	5	35	7	45
39	3	8	21	2	5	27	7	26
40	3	8	20	2	5	20	7	31
41	3	8	20	2	5	25	7	28
42	3	15	50	2	6	50	8	60
43	3	8	45	2	5	46	6	40
44	3	9	19	2	5	20	7	25
45	3	7	25	2	5	25	8	25
46	3	7	23	2	5	18	6	20

	Trial 9	Trial 10	Trial 11	Trial 12	Trial 13	Trial 14	Trial 15
0	6	20	17	25	8	26	4
1	6	26	12	35	12	30	4
2	6	20	9	20	9	35	4
3	6	15	9	20	9	25	4
4	6	15	10	25	9	23	4
5	6	15	20	28	15	23	4
6	6	17	11	23	10	36	4
7	7	15	10	20	9	30	4
8	6	20	11	30	9	30	4
9	6	15	10	20	9	20	4
10	6	16	10	21	9	32	4
11	6	12	9	18	9	23	4
12	6	12	10	18	12	14	4
13	6	16	11	34	17	28	4

14	8	20	11	24	10	30	4
15	6	13	10	15	9	38	4
16	6	18	10	20	9	21	4
17	6	15	10	20	9	25	4
18	6	15	11	25	9	26	4
19	6	15	12	18	9	20	4
20	6	15	10	25	9	15	4
21	6	15	11	18	9	22	4
22	6	18	8	15	12	25	4
23	6	17	10	29	10	32	4
24	6	19	10	20	9	23	4
25	6	18	10	23	9	20	4
26	6	17	10	18	9	20	4
27	6	14	8	19	7	29	3
28	6	18	10	27	9	24	4
29	6	17	12	18	10	21	4
30	6	13	9	20	10	20	5
31	6	20	13	25	10	35	4
32	6	15	10	20	10	30	4
33	6	18	15	28	10	28	4
34	6	14	10	22	9	23	4
35	6	15	10	20	9	18	4
36	6	17	9	18	7	23	4
37	7	15	11	17	9	28	4
38	6	16	9	19	9	23	4
39	6	14	11	16	9	21	4
40	6	19	12	16	16	27	4
41	6	18	10	20	9	19	4
42	7	20	10	30	10	40	4
43	6	18	9	20	9	23	4
44	6	14	11	18	9	23	4
45	6	15	12	18	11	15	4
46	5	13	9	12	8	21	4

Get a column from the data frame.

[5]:
```
df['Trial 1']
```

[5]:
```
0     3
1     3
2     3
3     3
4     3
5     3
6     3
7     3
8     3
```

```
9     3
10    3
11    3
12    3
13    3
14    3
15    3
16    3
17    3
18    3
19    3
20    3
21    3
22    3
23    3
24    3
25    3
26    3
27    3
28    3
29    3
30    3
31    3
32    3
33    3
34    3
35    3
36    3
37    3
38    3
39    3
40    3
41    3
42    3
43    3
44    3
45    3
46    3
Name: Trial 1, dtype: int64
```

Get multiple columns.

[6]:
```
df[['Trial 6','Trial 10']]
```

[6]:
```
    Trial 6  Trial 10
0        25        20
1        25        26
```

2	40	20
3	25	15
4	24	15
5	38	15
6	25	17
7	30	15
8	25	20
9	30	15
10	19	16
11	26	12
12	20	12
13	27	16
14	50	20
15	40	13
16	26	18
17	30	15
18	30	15
19	20	15
20	20	15
21	25	15
22	20	18
23	38	17
24	25	19
25	25	18
26	20	17
27	27	14
28	21	18
29	25	17
30	30	13
31	30	20
32	40	15
33	25	18
34	26	14
35	15	15
36	25	17
37	25	15
38	35	16
39	27	14
40	20	19
41	25	18
42	50	20
43	46	18
44	20	14
45	25	15
46	18	13

Convert a data frame column into an array.

[7]:
```
trl1 = np.asarray(df['Trial 1'])
```

View the array.

[8]:
```
trl1
```

[8]:
```
array([3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
       3, 3, 3])
```

Check the length of the array.

[9]:
```
len(trl1)
```

[9]:
```
47
```

To obtain credits for this assignment, submit through Quercus 1) a ~2-page PDF write-up addressing the points below (which can be an export of your Python notebook that includes all of your solutions), and 2) the executable Python notebook file that you have written:

- State your hypothesis in 1-2 sentences, and your rationale or scientific basis of the proposed hypothesis. **[2pts]**
- Generate one figure or table (appropriately annotated) from the data provided that helps to evaluate or illustrate the validity of your hypothesis. **[2pts]**
- Summarize the results and your observations, and suggest limitations of your approach or the experiment. **[2pts]**
- Outline a plan to extend your findings and state your conclusion in 1 sentence. **[2pts]**
- Submit your Python Jupyter notebook that reproduces all the findings reported (in a separate file). **[2pts]**

1. I hypothesize that as we continue to estimate bigger numbers (as we stare at cases where there are more dots), the result will start to become more inaccurate (the variability will increase).

2. To verify the above, I calculated the standard deviation of each trial and plotted it.

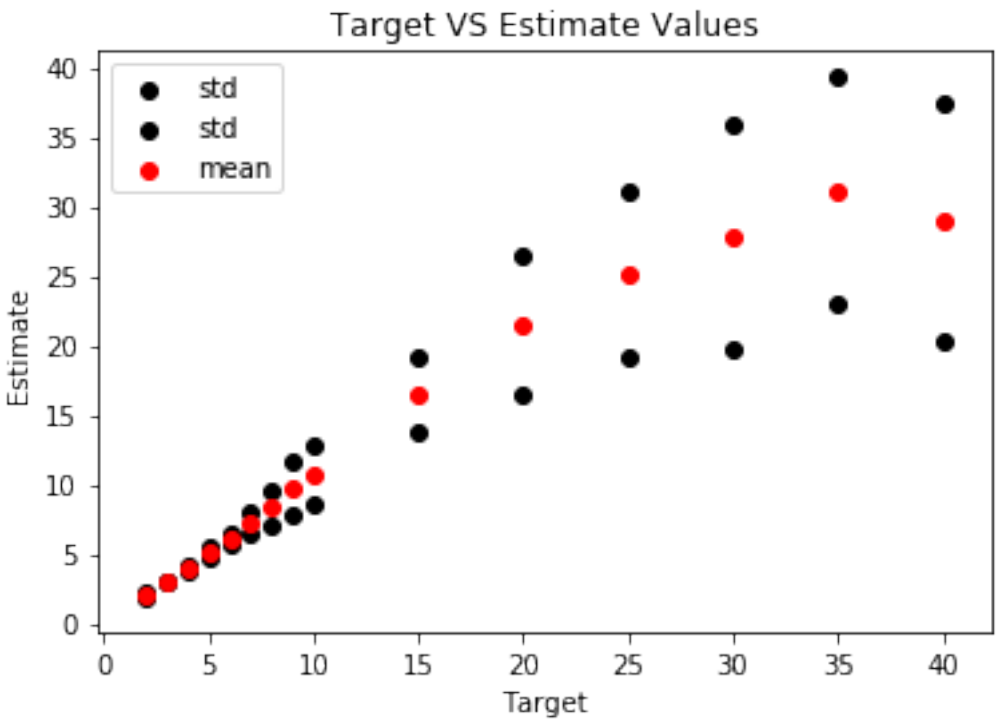
[55]:
```
mean = df.mean()
stdev = df.std()

plt.xlabel("Target")
plt.ylabel("Estimate")

plt.title("Target VS Estimate Values")
bl = plt.scatter(targets, mean+stdev, color = "black")
bl2 = plt.scatter(targets, mean-stdev, color = "black")
mean = plt.scatter(targets, mean, color = "red")

plt.legend((bl, bl2, mean), ('std', 'std', 'mean'))
```

[55]:
```
<matplotlib.legend.Legend at 0x118fc1eb8>
```



3. The above is a plot where I observe that as the number estimates become bigger, there tends to be a larger variation. We see that the standard deviation is increasing in the above plot, which tells us that we the data is getting more spread. As variability increases when standard deviation increases, this aligns with my hypothesis that more variable results will be yielded. I however address a limitation, such that the trials are relatively small so it is hard to draw a conclusion from a small subset of trials. It is also necessary to consider other factors such as humans speak different languages and therefore may think in a different linguistical numerical system. This could also have an effect when estimating variances.

4. To expand my findings, I will conduct many more multiple trials to observe a result that reflects a bigger sample. Also, if possible, I will try to find if certain groups of people who speak different languages have an easier time estimating larger numbers as compared to those who do not.