participacao

April 19, 2024

1 Exercício 1

Rafael Ragozoni Conrado 290268

[ ]:
```
import pandas as pd
import numpy as np
from plotnine import *
```

[ ]:
```
galton_df = pd.read_csv("Galton.csv")
galton_df.head()
```

[ ]:

	rownames	family	father	mother	sex	height	nkids
0	1	1	78.5	67.0	M	73.2	4
1	2	1	78.5	67.0	F	69.2	4
2	3	1	78.5	67.0	F	69.0	4
3	4	1	78.5	67.0	F	69.0	4
4	5	2	75.5	66.5	M	73.5	4

Load the height data set and create a vector x with just the male heights used in Galton's data on the heights of parents and their children from his historic research on heredity.

[ ]:
```
male_heights_df = galton_df[galton_df['sex'] == 'M']['height']
male_heights = male_heights_df.values
male_heights.shape
```

[ ]:
```
(465,)
```

1. Compute the average and median of these data.

[ ]:
```
mean_height = male_heights.mean()
mean_height
```

[ ]:
```
69.22881720430108
```

[ ]:
```
median_height = np.median(male_heights)
median_height
```

[ ]:
```
69.2
```

2. Compute the median and median absolute deviation of these data.

[ ]:
```
absolute_diff = np.abs(male_heights - median_height)
mad = np.median(absolute_diff)
mad
```

[ ]:
```
1.7999999999999972
```

3. Now suppose Galton made a mistake when entering the first value and forgot to use the decimal point.

How many inches does the average grow after this mistake?

[ ]:
```
male_heights_with_error = male_heights
male_heights_with_error[0] = male_heights_with_error[0]*10
male_heights_with_error[0]
```

[ ]:
```
732.0
```

[ ]:
```
mean_height_with_error = male_heights_with_error.mean()
print(f"{mean_height_with_error - mean_height} inches")
```

```
1.4167741935483775 inches
```

4. How many inches does the SD grow after this mistake?

[ ]:
```
sd = np.std(male_heights)
sd
```

[ ]:
```
2.628763088448231
```

[ ]:
```
sd_with_error = np.std(male_heights_with_error)
print(f"{sd_with_error - sd} inches")
```

```
28.185625167283536 inches
```

5. How many inches does the median grow after this mistake?

[ ]:
```
median_height_with_error = np.median(male_heights_with_error)
print(f"{median_height_with_error - median_height} inches")
```

```
0.0 inches
```

6. How many inches does the MAD grow after this mistake?

[ ]:
```
absolute_diff_with_error = np.abs(male_heights_with_error -␣
↪median_height_with_error)
mad_with_error = np.median(absolute_diff_with_error)
print(f"{mad_with_error - mad} inches")
```

```
0.0 inches
```

7. How could you use exploratory data analysis to detect that an error was made?

- a. Since it is only one value out of many, we will not be able to detect this.
- b. We would see an obvious shift in the distribution.
- c. A boxplot, histogram, or qq-plot would reveal a clear outlier.
- d. A scatterplot would show high levels of measurement error.

A resposta correta é a c