Measures

Saturday, June 17, 2023 11:32 AM

Various Sources / Formats → Data → Exploratory Data Analysis → Plan a ML Model

Exploratory Data Analysis:
- Compute the measures
- Graphs / Visualization

Plan a ML Model ↓↑ Model Building ↓↑ Model Evaluation

Geometric Mean: a, b → $(a\,b)^{1/2}$; $a_1, a_2, \ldots, a_n$ → $(a_1 \cdot a_2 \cdot \ldots \cdot a_n)^{1/n}$

Harmonic Mean: Reciprocal of mean of reciprocals

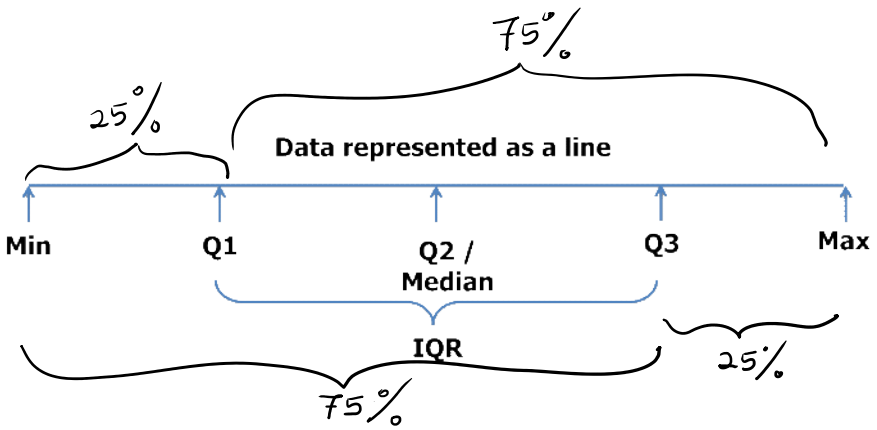
a, b → $\dfrac{1}{\frac{\frac{1}{a}+\frac{1}{b}}{2}} = \dfrac{2ab}{a+b}$

$a_1, a_2, \ldots a_n$ → $\dfrac{1}{\frac{\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}}{n}}$

1st Quartile :- 25% values in data less than 1st Q.

2nd Quartile :- 50% ——|——— 2nd Q
= median

3rd Quartile :- 75% ——|——— 3rd Q



Quartiles: Divide the data into 4 equal parts
Deciles: Divide the data into 10 equal parts
Percentiles: Divide the data into 100 equal parts

$$\text{Abs deviation about mean} = \frac{\sum |X_i - \text{mean}|}{N}$$ Mean Deviation

$$\text{Variance} = \frac{\sum (X_i - \text{mean})^2}{N}; \quad \text{Standard deviation} = \sqrt{\text{Variance}}$$

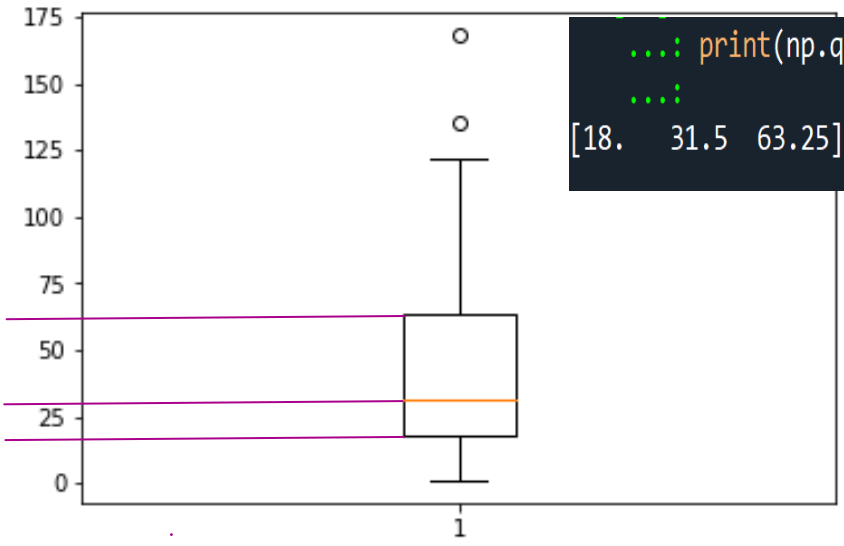
$$\text{mean} = \frac{\sum X_i}{N},$$

$$\sum_{i=1}^{N}(a_i - b_i) = \sum_{i=1}^{N} a_i - \sum_{i=1}^{N} b_i$$

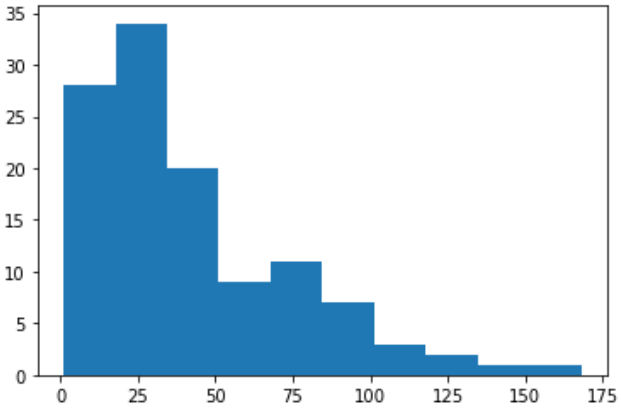
$$\sum_{i=1}^{N} (X_i - \text{mean}) = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} \text{mean}$$

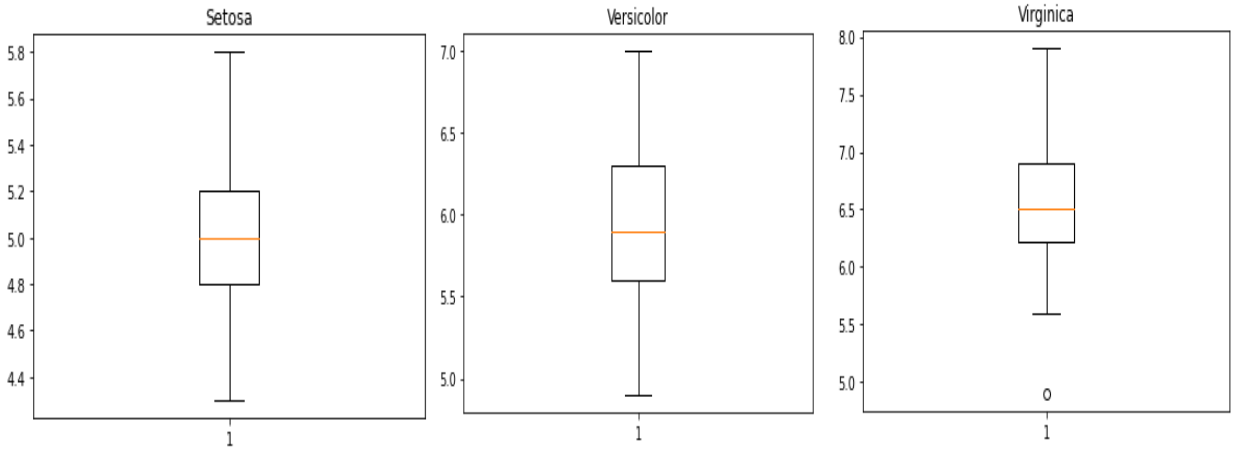
$$= N * \text{mean} - N * \text{mean}$$

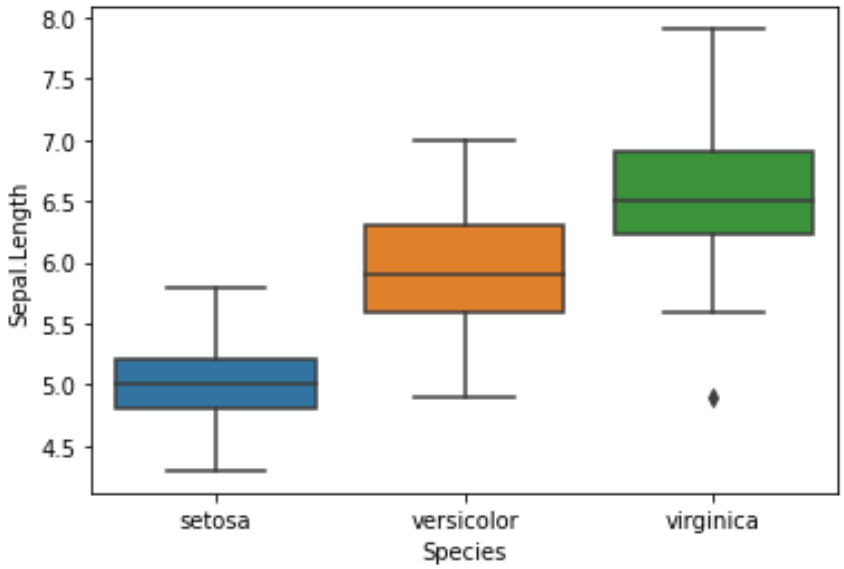
$$= 0$$

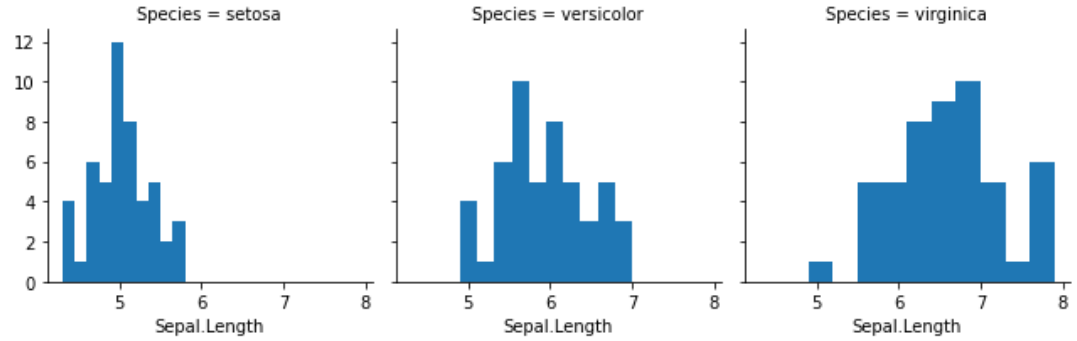


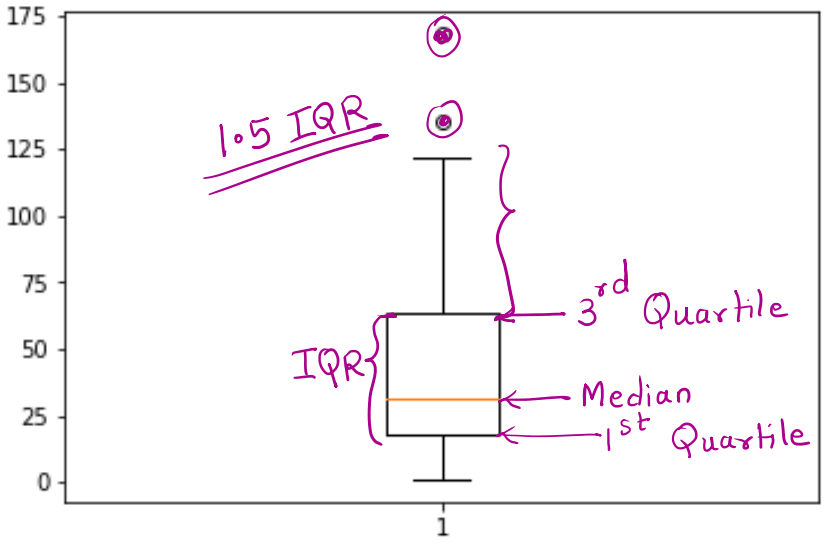

```
...: print(np.quantile(air['Ozone'].dropna(),
...:                   [0.25,0.5,0.75]))
[18.   31.5  63.25]
```


35
30
25
20
15
10
5
0
0
25
50
75
100
125
150
175

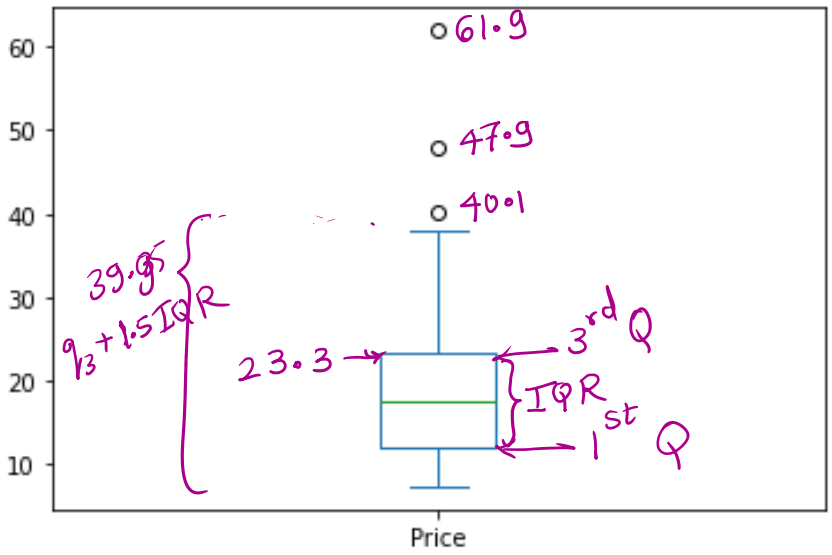

Setosa
Versicolor
Virginica
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
7.0
6.5
6.0
5.5
5.0
8.0
7.5
7.0
6.5
6.0
5.5
5.0
o
1
1
1


8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
Sepal.Length
setosa
versicolor
virginica
Species





$Q_3 - Q_1 = IQR$



Outliers :-
Any point which is 1.5 IQR away from the edge of the box

$1.5IQR = 16.65$

$q_3 + 16.65 = 39.95$

40.1
47.9
61.9

Outliers