

Venkatesh . K . H

linkedin.com/in/venkateshkh

My Handwritten Notes

Bafna Gold  
Date: 04/11/2019  
Page: 1

## Measures of Asymmetry

Most commonly used tool is Skewness

### Sample Skewness formula

$$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}}^3$$

Skewness indicates whether the data is concentrated on one side

Skewness can be confusing at the beginning  
So example in place.

Here we have 3 datasets

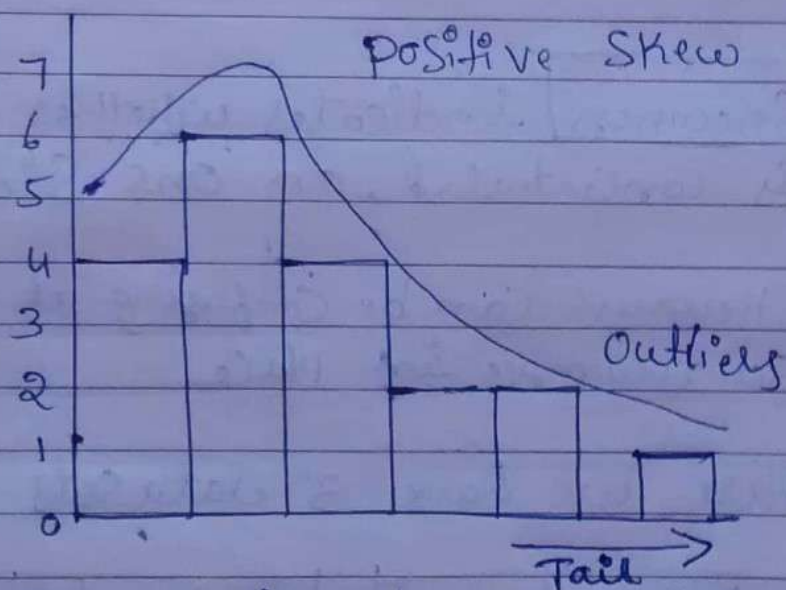
Dataset 1		Interval	Frequency
1	3	0 to 1	4
1	3	1 to 2	6
1	3	2 to 3	4
1	3	3 to 4	2
2	4	4 to 5	2
2	4	5 to 6	0
2	5	6 to 7	1
2	5		
2	7		
2			

Mean — 2.79  
 Median — 2.00  
 Mode — 2.00

$\text{Mean} > \text{Median}$

We say this is a +ve or right skew.

From the graph, we can see clearly the data points are concentrated on left side.

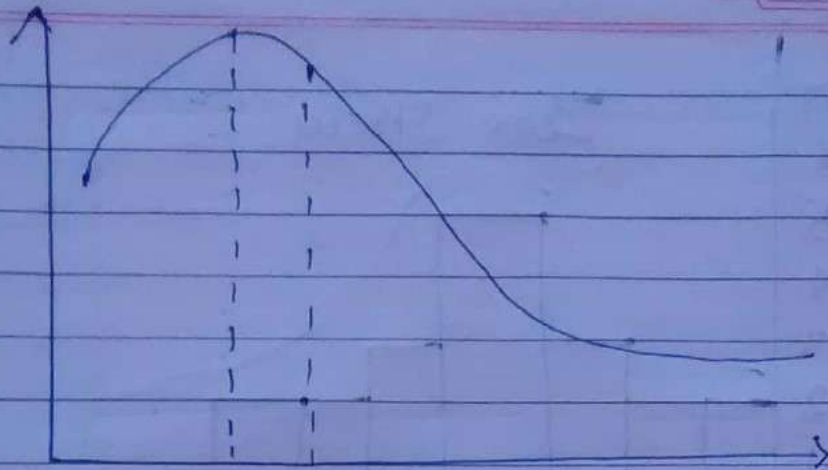


It is not depend on which side the line is leading to but rather to which side 'Tail' is leading to.

So right skewness means that the outliers are right.

When we have right skewness. The Mean is bigger than median. And the Mode is the value with the highest value representation.





Median mean.

Data Set 2	Interval	Frequency
1	0 to 1	2
1	1 to 2	2
2	2 to 3	3
2	3 to 4	5
3	4 to 5	3
3	5 to 6	2
3	6 to 7	2

Mean = Median = Mode

Mean - 4.00

Median - 4.00

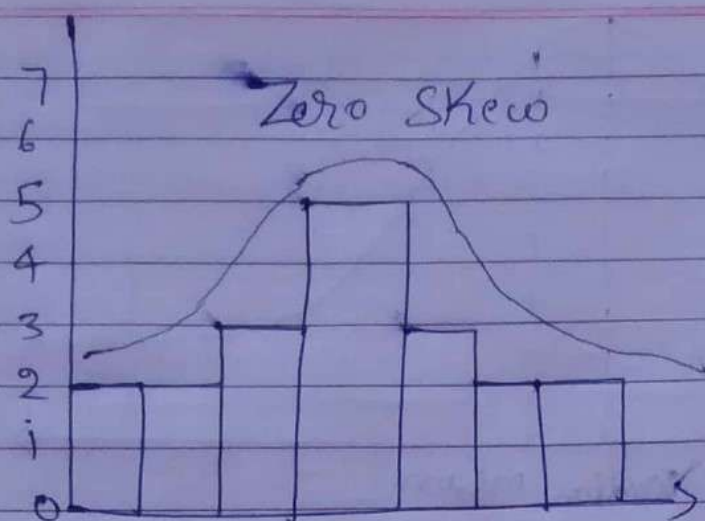
Mode - 4.00

6

6

7

7



The frequency of currency is completely symmetrical and we call this is Zero or no skew.

Most often, distribution is symmetrical.

Negative (left)

Data Set 3		Interval	Frequency
1	6	0 to 1	1
2	6	1 to 2	1
3	6	2 to 3	2
3	6	3 to 4	3
4	7	4 to 5	4
4	7	5 to 6	6
4	7	6 to 7	3
5			
5			
5			
5			
6			
6			

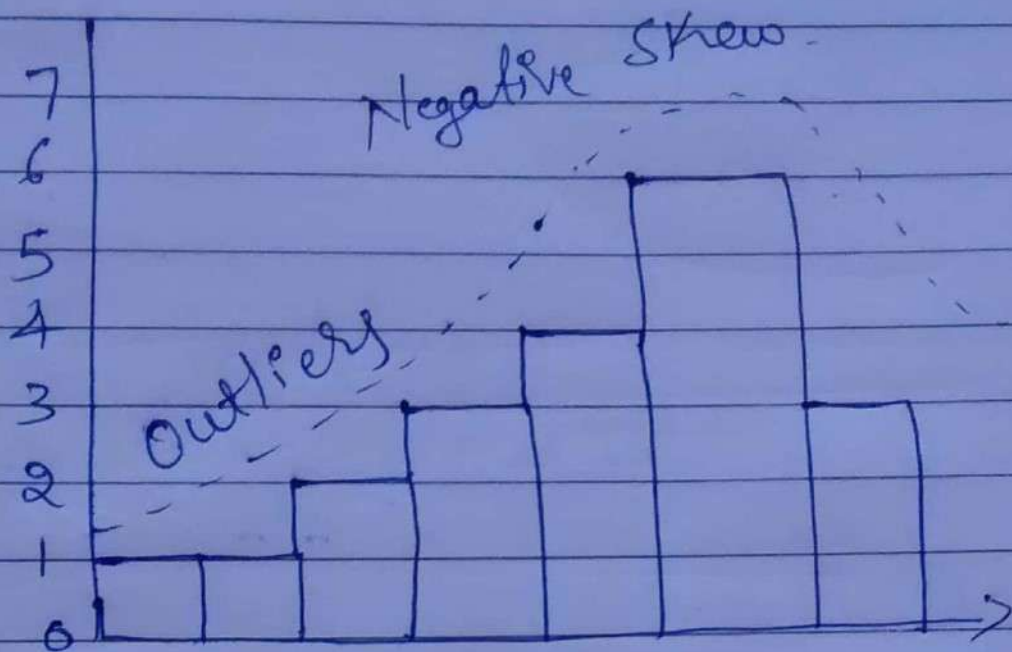
Mean - 4.90

Median - 5.00

Mode - 6.00



Mean < Median



We say there is -ve left skew.  
Once again the highest point define  
by the mode.

Why is that called left skew?  
Because outliers are at left.