Measures of Central Tendency

An essential statistical concept is the "measure of central tendency". This measure is an important way to summarize the dataset with one representative value. This measure provides a rough picture of where data points are centered. The commonly used measures of central tendency are:

- Mean
- Median
- Mode
- **Mean**: "Average" value is termed as the mean of the dataset. It is very easy to calculate the mean.
  Steps to calculate Mean:
  1. Count the number of data values. Let it be n.
  2. Add all the data values. Let the sum be s.
  3. Mean = Sum of all data values (s)/Total number of data values(n)

Python Code for Mean:

1: Mean

- Mean can be defined as

$$\text{Mean } (\bar{x}) = \frac{\sum x}{n}$$

```
In [5]:  # Sample Data
         arr  = [5, 6, 11]
         arr2 = [1,2,3,4,5,6,7,8,9,10]
         # Mean
         mean = np.mean(arr)
         mean2= np.mean(arr2)
         print("Mean = ", mean)
         print("Mean = ", mean2)

         Mean =  7.333333333333333
         Mean =  5.5
```

- **Median**: The middle value of the sorted dataset is called the median. Consider a dataset comprising 'n' elements.
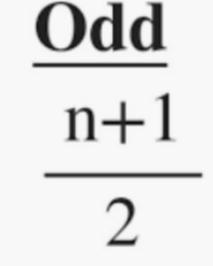 Steps to calculate median:
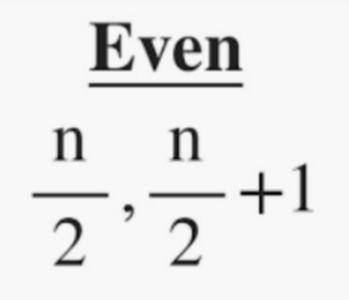 1. The dataset is arranged in either increasing or decreasing order.
  2. If the data set has an odd number of data values (n=odd), then the middlemost value of the sorted dataset is computed as the median. In other words, the data at (n + 1)/2 place is the median of the dataset.
  3. If the dataset has an even number of data values (n = even), the average of two middle values is computed as the median. i.e. the mean of (n/2) and {(n/2) + 1}th is the median of the dataset.

$$\underline{\text{Odd}} \qquad \underline{\text{Even}}$$

$$\frac{n+1}{2} \qquad \frac{n}{2}, \frac{n}{2}+1$$

```python
In [7]:
# sample Data
arr  =[1, 2, 3, 4]
arr2 = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]
# Median
median  = np.median(arr)
median2 = np.median(arr2)

print("Median = ", median)
print("Median = ", median2)
```

```
Median =  2.5
Median =  8.0
```

- **Mode**: The most frequently occurring value in the dataset is called mode.

  Steps to calculate mode:

  1. Use tally marks to identify how many times each data value occurs in the dataset.
  2. The data value with maximum tally is the mode of the dataset.

```python
In [4]:
# sample Data
arr =[1, 2, 2, 3]
arr2=[1,1,1,2,2,3,3,3,3,3,4,4]
# Mode
mode = stats.mode(arr)
mode2= stats.mode(arr2)
print("Mode = ", mode)
print("Mode = ", mode2)
```

```
Mode =  ModeResult(mode=array([2]), count=array([2]))
Mode =  ModeResult(mode=array([3]), count=array([5]))
```

Examples

Example 1. Consider the weight (in kg) of 5 children as 36, 40, 32, 42, 30. Let's compute mean, median, and mode:
Solution:
Mean = (36 + 40 + 32 + 42 + 30)/5 = 180/5 = 36kg
Median: Arrange the data in ascending order: 30, 32, 36, 40, 42 The middle value is 36. So, median = 36kg.
Mode: Thus, there are no repeating numbers in a given list, there is no mode existing for that particular list

Example 2. Consider the ages of five employees as 30, 30, 32, 38, 60 years. Calculate the measures of central tendency.
Solution:
Mean = (30 + 30 + 32 + 38 + 60)/5 = 190/5 = 38 years
Median: Arrange the data in ascending order: 30, 30, 32, 38, 60. The middlemost value is 32. So, median = 32 years
Mode: 30 years occurs most number of ties, so mode = 30 years
In this example, we saw that mean, median and mode have different values.

Mark as Read

Dash

Articles

Videos

Quiz