In previous class, we discussed about another data summarization method by which we can summarize and present our raw data set with a single value. And, number that can be describe data sets are called descriptive measures. We discussed about mean, more specifically arithmetic mean, geometric mean, and harmonic mean.

Now, let, you have two arithmetic means $\bar{X}_1$ and $\bar{X}_2$with respective numbers of observations n_1 and n_2. Then you can calculate another mean named "Combined Mean" or "Pooled mean". The formula can be written as,

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

For example, A consulting firm runs in two shifts. A random sample A of 13 employees has mean weekly salary 495$, and another random sample B of 10 employees has mean weekly salary 492$. Compute the arithmetic mean of the combined sample.

Except arithmetic mean, geometric mean, and harmonic mean, there is another special type of arithmetic mean is called "Weighted Arithmetic Mean". In arithmetic mean the frequency part is called weight for weighted arithmetic mean. The best example of "Weighted Mean" is calculating our CGPA. Formula can be written as,

$$\bar{X}_{WM} = \frac{\sum(w_i \times x_i)}{\sum w_i}$$

For example,

GPA	4.00	3.75	3.75	4.00
Credit	3	3	2	2

Question is "What is the average GPA?" We need to calculate GPA, right? That means, here GPA = Xi and Credit = Wi.

Another example, suppose a burger company offers three different kinds of burger packages small, medium and large for Tk. 100, Tk. 125 and Tk. 150. Of the last 10 burgers sold 3 were small, 4 were medium and 3 were large. To find the mean price of the last 10 burger packages sold. $\bar{X}_{WM} = \frac{(3\times100)+(4\times150)+(3\times150)}{10} = 125$

Median: Two points always keep in mind to understand the definition of "Median".

1. Middle value of the observation.
2. After they have been arranged/ordered from smallest to largest.

There two formulas of identifying median from ungrouped or raw data set.

First one, when "n" is odd:

$$Me = \left(\frac{n+1}{2}\right)^{th} Observation$$

Second one, when "n" is even:

$$Me = \frac{\left(\frac{n}{2}\right)^{th} Obs. + \left(\frac{n}{2}+1\right)^{th} Obs.}{2}$$

Let's see some mathematical examples.

To calculate median from "Odd" sample number, we have to follow some steps.

Step 1: Organize in ascending order

Step 2: $Me = \left(\frac{n+1}{2}\right)^{th} Observation$

For example: 5, 3, 9, 2, 7, 5, 8 are exam score of section "A".

To calculate median from "Odd" sample number, we have to follow some steps.

Step 1: Organize in ascending order

Step 2: $Me = \frac{\left(\frac{n}{2}\right)^{th} Obs. + \left(\frac{n}{2}+1\right)^{th} Obs.}{2}$

For example: 5, 3, 9, 2, 7, 5 are exam score of section "A".

But, for grouped data, the formula is lil bit different.

$$Me = L_m + \frac{\frac{N}{2} - F_c}{f_m} \times c$$

$L_m = Lower\ limit\ of\ median\ class$

$N = Total\ number\ of\ observations/frequency$

$F_c = Cumultative\ frequency\ of\ pre-median\ class$

$f_m = Frequency\ of\ median\ class$

$c = Class\ interval$

Steps:

1. Prepare a less than type cumulative frequency distribution.
2. Determine $\frac{N}{2}$, where N is the total frequency.
3. Locate the median class whose cumulative frequency includes the value of $\frac{N}{2}$.
4. Determine the value of $L_m, F_c, f_m, and\ c$.

Example:

Class	Frequency	CF_i
5-9	4	4
9-13	3	7
13-17	3	10
17-21	3	13

Here, $N = 13$.

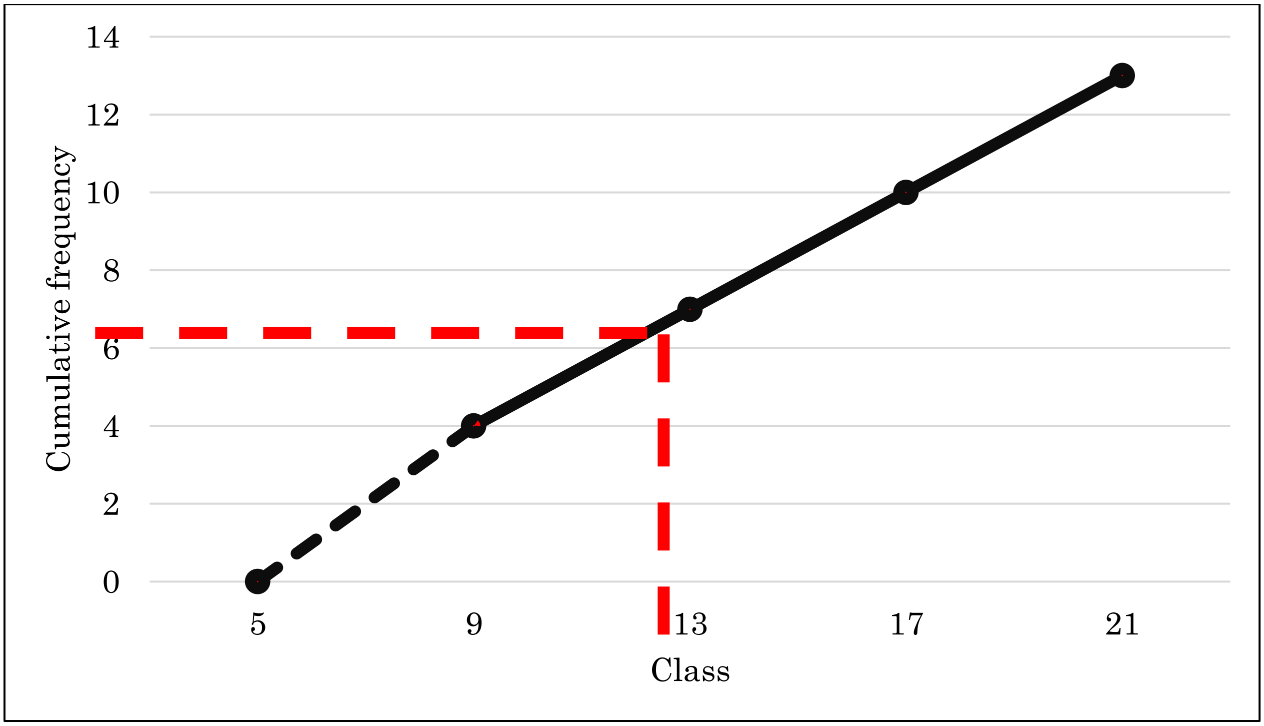
Now, $\frac{N}{2} = \frac{13}{2} = 6.5$

So, the median class is (9-13)

We know that, $Me = L_m + \frac{\frac{N}{2} - F_c}{f_m} \times c$

$$\therefore Me = 9 + \frac{6.5 - 4}{3} \times 4 = 12.33$$

One interesting thing, we can calculate median from ogive curve. This is as like your assignment's last question.



Mode: Most frequent value and occurs more than one times.

For example: 1, 2, 2, 5, 5, 3, 6, 6, 6, 4

Here the most frequent value is 6. Thus, the mode value is 6, and this is a unimodal data.

One mode = Unimodal; Two mode = Bi-modal; More than two = Multimodal

Another example: Example: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Which one is "Modal value" here?

There is no value which occur more than one time. Thus, there is no "Mode" in this data set.

For grouped data, the formula of mode can be written as,

$$Mo = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$$

Here,

$L_o = Lower\ limit\ of\ modal\ class$

$\Delta_1 = Difference\ between\ frequency\ of\ the\ modal\ class\ and\ pre - modal\ class.$

$\Delta_2 = Difference\ between\ frequency\ of\ the\ modal\ class\ and\ post - modal\ class.$

$c = Class\ interval$

Class	$\boldsymbol{f_i}$
10-20	5
20-30	8
30-40	12
40-50	7
50-60	9

Here, the class with highest frequency is $(30 - 40)$. This is our modal class.

Now, $L_o = 30, \Delta_1 = 12 - 8 = 4, \Delta_2 = 12 - 7 = 5, c = 10$

$$\therefore Mode, Mo = L_o + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c = 30 + \frac{4}{4+5} \times 10 = 34.44$$

Now, does all data have a Mean, median, and mode?

Do you remember the “Scales of measurement”? Nominal, Ordinal etc????

1. Every set of continuous data possesses a median, mode, and mean.
2. When considering ordinal data, it encompasses solely a median and mode.
3. Nominal data solely involves a mode

Measures of Location: We have already learned that median divided a set of data into two equal parts.

In the same way, we can divide a set of data into,

1. Four equal parts: If we divide four equal parts, it is called "Quartiles"
2. Ten equal parts: If we divide ten equal parts, it is called "Deciles"
3. Hundred equal parts: If we divide 100 equal parts, it is called "Percentiles"

For quartiles, we get 3 quartile points: Q1, Q2, and Q3

Similarly, for deciles, we get 9 decile points: D1, D2, D3, … , D9.

Finally, for percentiles, we get 99 percentile points: P1, P2, P3, … , P99

Now the question is, how to calculate this value. That is, how to calculate Q1, Q2, or Q3, or D1, D2, and so on. There are some simple formulas for calculating position Quartile, Deciles and Percentile.

$i = 1,\ 2,\ 3$

$N = Total\ number\ of\ observations$

For quartiles,

$$position\ of\ Q_i = \frac{i \times N}{4}$$

Similarly,

$$position\ of D_i = \frac{i \times N}{10}$$

$$position\ of P_i = \frac{i \times N}{100}$$

Steps:

- Arrange data set from smallest to largest
- Identify the position of $Q_i,\ D_i,\ or\ P_i$ by utilizing the formula,

$$J = \frac{(i \times n)}{4}$$

- If J is the integer value, then

$$\frac{\left(J^{th} Observation + (J+1)^{th} Observation\right)}{2}$$

- If J is not integer value, then take the next integer value as position.

Data: 20, 22, 27, 33, 23

Organize the data into ascending order,
20, 22, 23, 27, 33

Now, position of $Q_1 = \frac{i \times N}{4} = 1.25$

Since, the position value is not integer, thus we should go for next integer value.

$$\therefore Q_1 = 2nd\ obs. = 22$$

Data: 20, 22, 27, 23

Organize the data into ascending order,
20, 22, 23, 27

Now, position of $Q_1 = \frac{i \times N}{4} = 1$

Since, the position value is integer

$$\therefore Q_1 = \frac{1st\ Obs. \ + 2nd\ Obs}{2} = 21$$

For grouped data: Procedure are similar as “Median” calculation.

$$Q_i = L_Q + \frac{\frac{i \times N}{2} - F_c}{f_Q} \times c$$

$$D_i = L_D + \frac{\frac{i \times N}{2} - F_c}{f_D} \times c$$

$$P_i = L_P + \frac{\frac{i \times N}{2} - F_c}{f_P} \times c$$