Dec 7, 2020 201-114

Example: Alex timed 21 people in the sprint race to the nearest second. This is his observations:

59, 65, 61, 62, 53, 55, 60, 70, 64, 56, 58, 58, 62, 62, 68, 65, 56, 59, 68, 61, 67

The Actual Mean is: $\bar{x} = \frac{59+65+\cdots+61+67}{21} = 61.38095 \ldots$

To find the Median, Alex places the numbers in increasing order to get the following ordered data:

53, 55, 56, 56, 58, 58, 59, 59, 60, 61, **61**, 62, 62, 62, 64, 65, 65, 67, 68, 68, 70

We can see that the middle number is the 11th one which means that Median = 61.

To find the Mode, we have to see which number occurs the most. It is easy to see that it is the number 62. Therefore, the Mode is 62.

The next thing that Alex does is to group the data into 4 classes. We can see that 70 – 53 = 17. To make the classes of equal width, he adds 3 to 17 to make it also divisible by 4. So now 20/4 = 5. Therefore, each class should contain 5 numbers. Since the last number is 70, then we could make the first class to be 51 – 55, the second to be 56 – 60, the third to be 61 – 65, and the last class to be 66 – 70.

So our table would like the following

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4
	Total: 21

Each class has 5 numbers and that means the width is 5.

Suppose that the original data was lost and Alex is left only with the grouped data table above. And suppose that he hadn't found the Mean, the Median, and the Mode. Can he make an estimation for these Central Tendency Values? Yes, he can. Let's see how.

He would have to take the midpoint of each class. So the first class contains the number 51, 52, 53, 54, 55. The midpoint would be 53. The second class contains the numbers 56, 57, 58, 59, 60. Therefore, the midpoint of the second class is 58. Likewise, the midpoints of the third class and the fourth class will be 63 and 68 respectively. So the frequency table becomes a Midpoint Frequency Table as follows:

Midpoint (x)	Frequency (f)	(f)(x)
53	2	(2)(53) = 106
58	7	(7)(58) = 406
63	8	(8)(63) = 504
68	4	(4)(68) = 272
	Total: 21	Total: 1288

Therefore, the Estimated Mean is $M_E = \frac{1288}{21} = 61.333333 \ldots$ which is very close to the actual mean.

From the Midpoint Frequency Table, we can see that we are imagining that our observations are:

53, 53, 58, 58, 58, 58, 58, 58, 58, 63, **63**, 63, 63, 63, 63, 63, 63, 68, 68, 68, 68

So the middle number is 63. And it lies in the class 61 – 65. This class is called the Median Class or Median Group.

Now if we want to estimate the median value, we have to look at the Median Class 61 – 65. We do call it 61 – 65 but really it includes values from 60.5 up to (but not including 65.5). Because remember Alex rounded off the observations to the nearest second. So 60.5 was recorded as 61 and 65.4 was recorded as 65.

So at 60.5, we already have a cumulative frequency of 2 + 7 = 9 runners. By the next boundary number 65.5, we have 17 runners.

The formula for the Estimated Median Md_E is:$Md_E = L + \frac{(\frac{n}{2} - B)w}{G}$, where:

L is the lower boundary of the Median Class.

n is the total number of observations

B is the cumulative frequency of the groups before the Median Class

G is the frequency of the Median Class

w is the class width

So in our example: L = 60.5, n = 21, B = 9, G = 8, and w = 5

Therefore $Md_E = 60.5 + \frac{(\frac{21}{2} - 9)(5)}{8} = 60.5 + \frac{7.5}{8} = 61.4375$

Again this value of the Estimated Median is pretty close to the actual Median which was 61

Noe let's see how we can estimate the Mode:

The Modal Group or Modal Class is the group with the highest frequency. So, in our example it is the class 61 – 65.

The Estimated Mode, Mo_E, is given by the following formula:

$$Mo_E = L + \frac{(f_m - f_{m-1})w}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$, where

L is the lower boundary of the Modal Group or Modal Class

f_m is the frequency of the Modal Group

f_{m-1} is the frequency of the group before the Modal group

f_{m+1} is the frequency of the group after the Modal group.

w is the width of the class or group

So in our example, L = 60.5, $f_m = 8$, $f_{m-1} = 7$, $f_{m+1} = 4$, and w = 5

Inserting these values in our formula, we get:

$$Mo_E = 60.5 + \frac{(8-7)5}{(8-7)+(8-4)} = 61.5$$

So our results are:

Actual Mean = 61.38095… Estimated Mean = 61.333333…

Actual Median = 61 Estimated Median = 61.4375

Actual Mode = 62 Estimated Mode = 61.5

Quartiles are the values that divide a list of numbers into quarters. So what you would have to do first is to order your numbers in increasing order. Then you cut the list into 4 equal parts. The quartiles are the cuts.

<u>Example</u> Suppose we have this list of data 5, 7, 4, 4, 6, 2, 8

Listing them in increasing order, we have: 2, 4, 4, 5, 6, 7, 8 The middle number is 5 so we make a cut there. And then make a cut at the second number and second last one:

2, ~~4~~, 4, ~~5~~ , 6,~~7~~, 8

Therefore the first quartile is Q1 = 4, the second Quartile is Q2 = 5, and the third Quartile is Q3 = 7.

Notice that Q2 is the Median value also.

<u>Example</u> If the sample size is even such as the following set of data which has already been ordered in increasing form: 1, 3, ~~3~~, 4, 5 | 6, 6, ~~7~~, 8, 8, then

Q2 = Median = (5+6)/2 = 5.5

Therefore, our Q1 = 3 and Q3 = 7

The Interquartile range is always Q3 – Q1.

Therefore, in the first example about the quartiles, the interquartile range is 7- 4 = 3

And in the second example about quartiles, the interquartile range is 7 – 3 = 4