Mean, Median and Mode from Grouped Frequencies

The Race.....

This starts with some raw data (not a grouped frequency yet) ...

Alex timed 21 people in the sprint race, to the nearest second:

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

After sorting of time points in dataset;

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

To find the Mean, Median & Mode Alex foloows the normal methods and return as,

Mean = 61.38095...

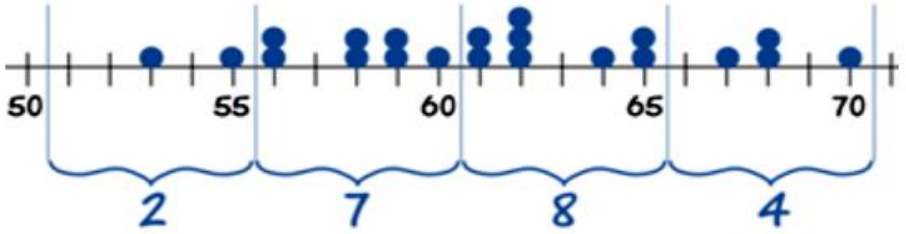
Median = 61

Mode = 62

Mean, Median and Mode from Grouped Frequencies

- Grouped Frequency Table
- Alex then makes a Grouped Frequency Table:
- *So 2 runners took between 51 and 55 seconds, 7 took between 56 and 60 seconds, etc*

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



Oh No!

Suddenly all the original data gets lost (naughty pup!)

Only the Grouped Frequency Table survived ...

... can we help Alex calculate the Mean, Median and Mode from just that table?

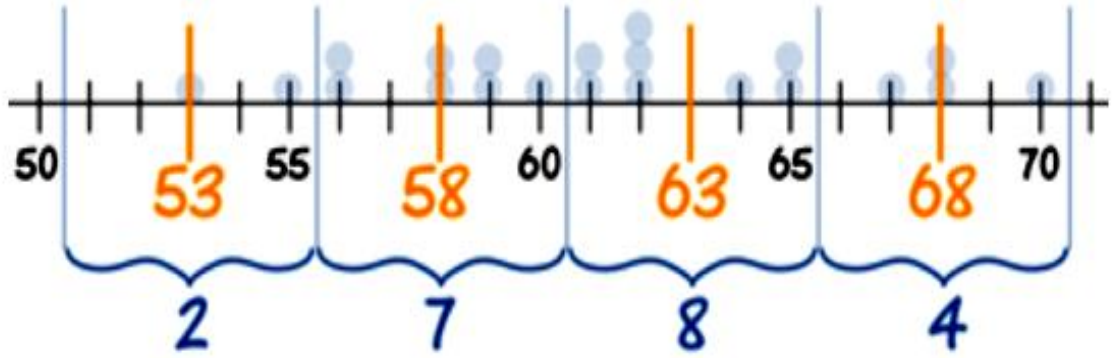
The answer is ... no we can't. Not accurately anyway. But, we can make **estimates**.

Estimating the Mean from Grouped Data

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

The groups (51-55, 56-60, etc), also called **class intervals**, are of **width** 5

The **midpoints** are in the middle of each class: 53, 58, 63 and 68



Midpoint	Frequency
53	2
58	7
63	8
68	4

We can estimate the Mean by using the ***midpoints.***

Let's now make the table using midpoints:

Estimating the Mean from Grouped Data

Midpoint	Frequency
53	2
58	7
63	8
68	4

- Our thinking is: "2 people took 53 sec, 7 people took 58 sec, 8 people took 63 sec and 4 took 68 sec". In other words we imagine the data looks like this:

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

- Then we add them all up and divide by 21. The quick way to do it is to multiply each midpoint by each frequency:

Midpoint x	Frequency f	Midpoint × Frequency fx
53	2	106
58	7	406
63	8	504
68	4	272
Totals:	**21**	**1288**

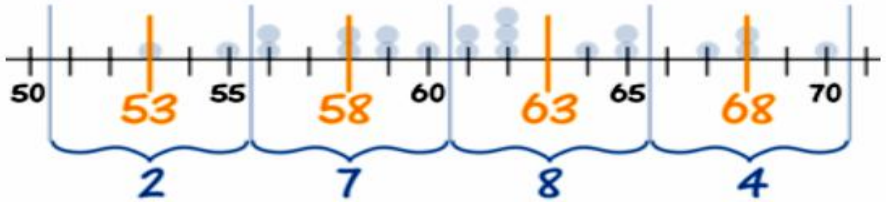
And then our estimate of the mean time to complete the race is:

$$\text{Estimated Mean} = \frac{1288}{21} = 61.333\ldots$$

Estimating the Median from Grouped Data

- Let's look at our data again:
- The median is the middle value, which in our case is the 11th one, which is in the 61 - 65 group:
- We can say "the **median group** is 61 - 65"
- But if we want an estimated **Median value** we need to look more closely at the 61 - 65 group.

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



We call it "61 - 65", but it really includes values from 60.5 up to (but not including) 65.5.

Why? Well, the values are in whole seconds, so a real time of 60.5 is measured as 61. Likewise 65.4 is measured as 65.

Estimating the Median from Grouped Data

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

- At 60.5 we already have 9 runners, and by the next boundary at 65.5 we have 17 runners.
- By drawing a straight line in between we can pick out where the median frequency of n/2 runners is:

And this handy formula does the calculation:

$$\textbf{Estimated Median} = L + \frac{(n/2) - B}{G} \times w$$

where:

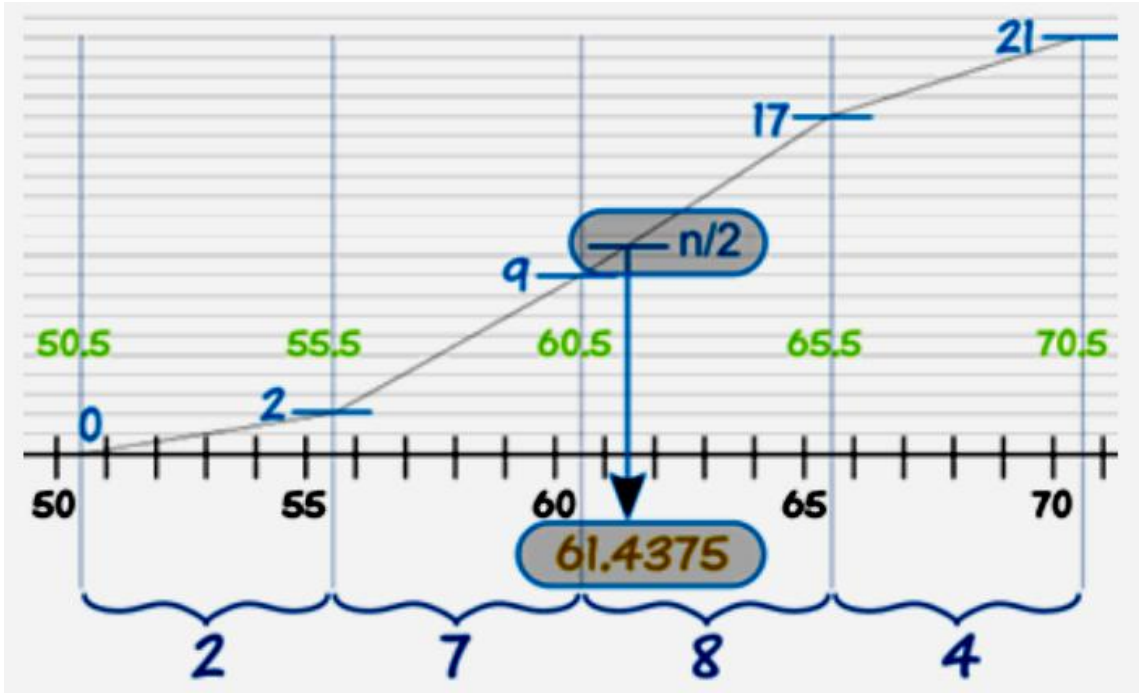
L is the lower class boundary of the group containing the median

n is the total number of values

B is the cumulative frequency of the groups before the median group

G is the frequency of the median group

w is the group width



For our example:

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

L = 60.5

n = 21

B = 2 + 7 = 9

G = 8

w = 5

$$median = L_1 + (\frac{n/2 - (\sum freq)l}{freq_{median}})width$$

age	*frequency*
1–5	200
6–15	450
16–20	300
21–50	1500
51–80	700
81–110	44

Estimated Median $= 60.5 + \frac{(21/2) - 9}{8} \times 5$

$= 60.5 + 0.9375$

$= \mathbf{61.4375}$

Estimating the Mode from Grouped Data

- Again, looking at our data:
- We can easily find the **modal group** (***the group with the highest frequency***), which is 61 - 65
- We can say **"the modal group is 61 - 65"**
- But the actual Mode may not even be in that group! Or there may be more than one mode. Without the raw data we don't really know. But, we can estimate the Mode using the following formula:

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

$$\text{Estimated Mode} = L + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times w$$

where:

L is the lower class boundary of the modal group

$\mathbf{f_{m-1}}$ is the frequency of the group before the modal group

$\mathbf{f_{m+1}}$ is the frequency of the group after the modal group

$\mathbf{f_m}$ is the frequency of the modal group

w is the group width

For our example:

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

$L = 60.5$

$f_{m-1} = 7$

$f_m = 8$

$f_{m+1} = 4$

$w = 5$

Estimated Mode $= 60.5 + \frac{8 - 7}{(8 - 7) + (8 - 4)} \times 5$

$= 60.5 + (1/5) \times 5$

$= \mathbf{61.5}$

Our final result is:

- Estimated Mean: **61.333...**
- Estimated Median: **61.4375**
- Estimated Mode: **61.5**

(Compare that with the true Mean, Median and Mode of **61.38..., 61 and 62** that we got at the very start.)

Baby Carrots Example

Example: You grew fifty baby carrots using special soil. You dig them up and measure their lengths (to the nearest mm) and group the results:

Length (mm)	Frequency
150 - 154	5
155 - 159	2
160 - 164	6
165 - 169	8
170 - 174	9
175 - 179	11
180 - 184	6
185 - 189	3

Age Example

The ages of the 112 people who live on a tropical island are grouped as follows:

Age	Number
0 - 9	20
10 - 19	21
20 - 29	23
30 - 39	16
40 - 49	11
50 - 59	10
60 - 69	7
70 - 79	3
80 - 89	1