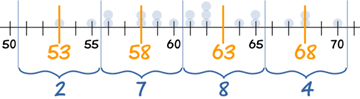
**Estimating the Mean from Grouped Data**

So all we have left is:

|  |  |
| --- | --- |
| **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



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

So, how does this work?

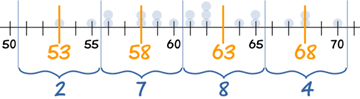
Think about the 7 runners in the group **56 - 60**: all we know is that they ran somewhere between 56 and 60 seconds:

* Maybe all seven of them did 56 seconds,
* Maybe all seven of them did 60 seconds,
* But it is more likely that there is a spread of numbers: some at 56, some at 57, etc

So we take an average and **assume** that all seven of them took 58 seconds.

Let's now make the table using midpoints:

|  |  |
| --- | --- |
| **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:

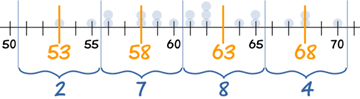
**Estimated Mean** =  *1288***21**  = **61.333...**

Very close to the exact answer we got earlier.

**Estimating the Median from Grouped Data**

Let's look at our data again:

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



The median is the middle value, which in our case is the 11th one, which is in the 61 - 65 group:

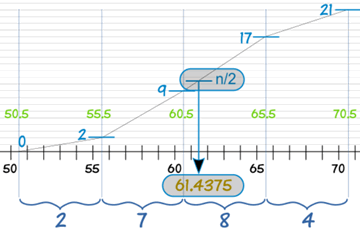
We can say "the **median group** is 61 - 65"

But if we want an estimated **Median value** we need to look more closely at the 61 - 65 group.

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.

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:

**Estimated Median** = L +  *(n/2) − B***G** × 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:

* **L** = 60.5
* **n** = 21
* **B** = 2 + 7 = 9
* **G** = 8
* **w** = 5

**Estimated Median**= 60.5 + *(21/2) − 9***8** × 5

 = 60.5 + 0.9375

 = **61.4375**

**Estimating the Mode from Grouped Data**

Again, looking at our data:

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

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:

**Estimated Mode** = L +  *fm − fm-1***(fm − fm-1) + (fm − fm+1)** × w

where:

* L is the lower class boundary of the modal group
* fm-1 is the frequency of the group before the modal group
* fm is the frequency of the modal group
* fm+1 is the frequency of the group after the modal group
* w is the group width

In this example: