

## Simple Frequency Distribution

For class of 25 student , the score of math quiz would typically look like those in table 2.1

8	6	10	9	6	6	8	7		Table 2.1
4	9	6	2	8	6	10	4		
5	6	8	4	7	8	4	7	6	

You can see that there are a lot of 6s and 8s, max value-10, min value-2

But this is not the best way to get a sense of how the class performed. A very simple and logical step makes it easier to understand the scores: **Put them in order** (customarily starting with the highest value) is often called **an array**

10	10	9	9	8	8	8	8		Table 2.2
8	7	7	7	6	6	6	6		
6	6	6	5	4	4	4	4	2	

But the table could be made more compact. If the class contained 100 students, an array would be quite difficult to look at. A more informative way to display these data is in a **simple frequency distribution**

**Frequency** —that is, how many times that score is repeated in the array.

Table consisting of two columns, The first column lists all of the possible scores, beginning with the highest

Score in the array and going down to the lowest score. The second column lists the **frequency** of each score.

**Table 2.3**

$X$	$f$
10	2
9	2
8	5
7	3
6	7
5	1
4	4
3	0
2	1
$\Sigma f =$	25

$X$  stands for any score, and  $f$  stands for the frequency of that score, and  $N$  stands for total no of subject..

score of 3 is included in the table even though it has a frequency of zero (i.e., there are no 3s in the data array). The rule is to list all the possible scores from the highest to the lowest, whether a particular score actually appears in the data or not.

The score that occurs most frequently in a distribution is called **the mode** of the distribution. For the preceding distribution, the mode is 6 because

To check whether you have included all your scores in the frequency distribution, add up all of the frequencies (i.e.,  $\Sigma f$ ), and make sure that the total is equal to the number of scores in the array (i.e., check that  $\Sigma f = N$ )

---

To evaluate his or her own performance in a class, a student will frequently ask, “How many students in the class had lower scores than mine?”

However, you can perform a procedure that will answer that question for any student in the class: You can construct **a cumulative frequency distribution**.

Pro-----Each cf entry equals the frequency of that score plus the cf for the score below.

**Table 2.4**

$X$	$f$	$cf$
10	2	25
9	2	23
8	5	21
7	3	16
6	7	13
5	1	6
4	4	5
3	0	1
2	1	1

The entry at the bottom will be same as  $f$

The entry at the top of the  $cf$  column should equal the total number of scores,  $N$ , which also equals  $\Sigma f$ .

You have scored better than other does not mean that you have got good grades.

Outscoring 15 students in a class of 25 is pretty good because you beat 3/5 of the class. Having 15 students below you in a class of 100 is not very good because in that case you have outperformed only 3/20, or .15, of the class.

The kind of table that can tell you what fraction of the scores is lower than yours is called a *cumulative relative frequency distribution*

You can create a relative frequency distribution by dividing each entry in the  $f$  column of a simple frequency distribution by  $N$ .

$X$	$f$	$cf$	$rf$	$crf$
10	2	25	.08	1.00
9	2	23	.08	.92
8	5	21	.20	.84
7	3	16	.12	.64
6	7	13	.28	.52
5	1	6	.04	.24
4	4	5	.16	.20
3	0	1	0	.04
2	1	1	.04	.04

The resulting fraction is the relative frequency ( $rf$ ), and it tells you what proportion of the group attained each score.

The cumulative relative frequencies ( $crf$ ) are then found by accumulating the  $rf$ 's starting from the bottom

Alternatively (**second method**), you can convert each entry in the *cf* column into a proportion by dividing it by *N*. (For example, the *cf* of .64 for a score of 7 can be found either by dividing 16 by 25.

Note that the *cf* for the top score in the table must be 1.0

---

Sometimes people find it more convenient to work with percentages.

If you want a **cumulative percentage frequency (*cpf*)** column, you need only multiply each *cf* entry by 100. A score of 7 is better than the scores of 52% of the class; a 6 beats only 24% of the scores. Be

<i>X</i>	<i>f</i>	<i>cpf</i>
10	2	100%
9	2	92
8	5	84
7	3	64
6	7	52
5	1	24
4	4	20
3	0	4
2	1	4

*cpf* column have a special name: They are called **percentile ranks (PR)**.

By convention, a percentile rank is defined as the percentage of the group that is **at or below a given score**.

To find the PR of a particular score we look straight across at the *cpf* entry, rather than looking at the score below. Thus, the PR of a score of 7 is 64; 64% of the group scored 7 or below. Similarly, the PR for 6 is 52.

---

Further if we want some treatment (extra class) for the bottom 20% student In this case, we want to **find the score** in the distribution that has a PR of 20. You can see from Table 2.6 **that a score of 4 has a PR of 20**.

This score is called the **20th percentile**. Anyone with this score or a lower score will have to get algebra help(extra class)

**A percentile can be defined as a score that has a given PR**—the 25<sup>th</sup> percentile is a score whose PR is 25. In other words, a percentile is the score at or below which a given percentage of the group falls.

The most interesting percentiles are either quartiles (i.e., 25%, 50%, or 75%) or deciles (i.e., 10%, 20%, etc.). Unfortunately, these convenient percentiles rarely appear as entries in a *cpf* column.(there are method to find it)

Later in this section, I will show how you can use a graph to **more precisely estimate** percentiles (and PRs) that do not appear as entries in a table.

---

Sometimes we want to know how many are above than one's score. To find the same we reverse the process by adding the frequency from top.

<b><i>X</i></b>	<b><i>f</i></b>	<b><i>cf (less than)</i></b>	<b><i>cpf</i></b>	<b><i>cf(more than)</i></b>	<b><i>cpf</i></b>
10	2	25	100%	2	8%
9	2	23	92	4	16
8	5	21	84	9	36
7	3	16	64	12	48
6	7	13	52	19	76
5	1	6	24	20	80
4	4	5	20	24	96
3	0	1	4	24	96
2	1	1	4	25	100

If you want to felicitate top 16% than what score will you choose