UNDERSTANDING TRUE LIMITS AND MID-POINTS IN THE ORGANIZATION OF DATA

Organization of data involves conversion of raw Data into a more meaningful form. To do so requires requires the following steps:

1. Convert it into an array - data arranged into an ascending or descending order.

2. Then converted the array into a frequency distribution.

3. Two types of frequencies can be generated:

    i. **Ungrouped frequency distribution.**
    ii. **Grouped frequency distribution.**

4. Ungrouped frequency distribution - the number of times each item figure appears separately.

5. Grouped frequency distribution - the number of items appear in groups.

ILLUSTRATION

Given a sample of **n=25** workers at UNZA and their ages as raw data.

29	24	54	41	29	25	27	20	
22	33	24	33	25	35	27	20	
43	20	31	51	22	36	20	48	46

FROM RAW DATA INTO AN ARRAY

1. Converting the raw data into an array in an ascending order, and we have:

x_i
20
20
20
20
22
22
24
24
25
25
27
27
29
29
31

33
33
35
36
41
43
46
48
51
54

FROM ARRAY INTO AN UNGROUPED FREQUENCY DISTRIBUTION

1. Converting the array into ungrouped frequency distribution, we have:

X_i	Frequency
20	4
22	2
24	2
25	2
27	2
29	2
31	1
33	2
35	1
36	1
41	1
43	1
46	1
48	1
51	1
54	1
Total	25

FROM UNGROUPED TO GROUPED FREQUENCY DISTRIBUTION

1. To reduce the number of figures, summarize the main features of the data and present information more effectively requires a grouped frequency distribution with class intervals

2. In creating a grouped frequency distribution, ensure that the class intervals created are mutually exclusive.

3. Also ensure number of observations in the class intervals are not too many or too few.

4. Also ensure that the first class interval accommodates the smallest observation and the last class interval accommodates the largest observation.

5. An example of this is given below:

Class interval	Frequency	Relative frequency (%)
20-24	8	32
25-29	6	24
30-34	3	12
35-39	2	8
40-44	2	8
45-49	2	8
50-54	2	8
Total	**25**	**100**

DETERMINATION OF CLASS INTERVALS

1. Ideally, the number of class intervals should be between 5 and 20.

2. It is therefore important to create a number of class intervals that guarantees the least distortion of information and also gives maximum convenience.

3. Too many intervals result in cumbersomeness in analysis.

4. Too few intervals may group unrelated cases together and result in misleading interpretations.

5. Once the required number of class intervals is decided on, decide on the size of interval or width of the classes.

6. The size or width of the class interval is:

   Size of the Class Interval = Range / Required Number of Intervals

7. In this example, the range (the difference between the highest and lowest value) is:

   Range = 54-20

   = 34.

8. Therefore, if we require **7 class intervals** then: -

   Size of the Class Interval = 34 / 7

   = 4.8 or 5

9. Choose the lowest class that brackets the lowest value, 20, and the highest class interval that brackets the highest value, 54.

10. Ideally, choose starting point for the first class interval so that no measurement falls on the point of division between the intervals.

11. We now have:

Class interval
20-24
25-29
30-34
35-39
40-44
50-54
Total

12. The class intervals we have created through this process have what are known as the **stated limits**.

13. The **stated limits** have **lower** and **upper limits**, i.e., the **stated lower limit** and the **stated upper limit**.

14. In his example, for the class interval, 20 is the **stated lower limit** and 24 is the **stated upper limit**.

.

THE CONCEPT OF TRUE LIMITS AND THE MID-POINT

1. In the example of the stated limits above, it would be difficult to accommodate a value like 24.3 or 24.9.

2. To accommodate numbers continuous data, like 24.3 and 24.9 in this case, one should ensure that there are no gaps between the end of the previous interval and the beginning of the next interval.

3. This means using **true (or real) limits** instead of **stated limits**.

4. The **stated limits** are the **given limits** created by the researcher such as the ones we have, i.e., as 20-24,

   25-29, etc.

5. The **true (or real)** is theoretically the highest or lowest value that can be assigned to that interval.
6. For an upper limit like 24, this can conceivably go as far as 24.999999 before reaching 25 in the next class interval.

7. To arrive at the **true limit** in easily, a simpler approach formula is used as indicated below:

8. This involves computing a **factor** by subtracting the **stated upper limit** from the preceding class interval from the **stated lower limit** of the next class and then dividing the result by 2.

Factor = Stated lower limit of the next class interval - stated upper limit of the previous class interval divided by 2.

9. The **stated upper limit** of first class interval ,20-24, is 24; and the **stated lower limit** of the second class interval, 25-29, is 25.

10. The required factor is therefore:

= (25-24)/2
= 0 .5

11. The factor of **0.5** is then subtracted from all **stated lower limits** and added to all **stated upper limits** to get the **true limits** for the entire distribution as indicated below:
.

Class interval	Frequency
19.5 - 24.5	8
24.5 - 29.5	6
29.5 - 34.5	3
34.5 - 39.5	2
39.5 - 44.5	2
44.5 - 49.5	2
49.5 - 54.5	2
Total	**25**

THE SIZE OF THE CLASS INTERVAL

1. This is the difference between the **true upper limit** and the **true lower limit**.

   Size = True Upper Limit – True Lower Limit

   Size = 24.5 - 19.5 = 5

2. If you use the stated limits, you will get a misleading result like the one below:

   Size = 24 - 29 = 4.

INTERPRETATION

1. **True limits** extend boundaries of intervals and bracket the **stated limits.**

2. This eliminates ambiguity in assigning cases to intervals especially for boundary values such as 24.3

3. In interpreting the true limits in the example given, the first interval includes those older than 19.5 but younger than 24.5 years.

REASONS FOR COMPUTING TRUE LIMITS

True limits are recommended for the following reasons:

1. To avoid gaps between intervals for continuous data like weight e.g., 24.3.
2. To avoid ambiguity by assigning cases to mutually exclusive intervals.
3. To ensure accuracy in computing medians, percentiles, deciles, quartiles, etc.
4. To aid the construction of the graphs that use continuous data such as histograms, polygons, line charts, etc.

THE CONCEPT OF A MID-POINT

The mid-point represents the middle point in a class interval.

REASONS FOR COMPUTING THE MID-POINT

The mid – point has a number of important uses. It is used in the following way:

1. To represent $\mathbf{X_i}$ values in grouped data.
2. To aid in the construction of frequency polygons - points on graph are connected through the mid-points of the intervals.

COMPUTATION OF THE MID-POINT

1. The formula for the mid-point is:

$$\textbf{Mid - point} = \frac{\textbf{True Lower Limit + True Upper Limit}}{\textbf{2}}$$

$$= \frac{\textbf{19.5 + 24.5}}{\textbf{2}}$$

$$= \textbf{22}$$

You can also use the stated limits and get the same result.

$$\textbf{Mid - point} = \frac{\textbf{Stated Lower Limit + Stated Upper Limit}}{\textbf{2}}$$

$$= \frac{\textbf{20+ 24}}{\textbf{2}}$$

$$= \textbf{22}$$

THE RELATIVE FREQUENCY DISTRIBUTION

1. This shows the percentage of the total number observed at each score value (for ungrouped data) and four each class interval (for grouped data).

2. It is useful in making comparisons because using absolute frequency distributions can be misleading.

3. For example, one say that there are twice as many workers at ZCCM compared to UNZA, but in relative terms or percent terms, the percentages in each group are equal.

	UNZA		ZCCM	
Class interval	**Frequency**	**Percent**	**Frequency**	**Percent**
19.5 - 24.5	8	32	16	32
24.5 - 29.5	6	24	12	24
29.5 - 34.5	3	12	6	12
34.5 - 39.5	2	8	4	8
39.5 - 44.5	2	8	4	8
44.5 - 49.5	2	8	4	8
49.5 - 54.5	2	8	4	8
Total	**25**	**100**	**50**	**100**

THE CUMULATIVE DECUMULATIVE FREQUENCY DISTRIBUTION

1. This shows the percentage on number of observations located below or above a certain limit.

2. To know the percentage or number of observations below a certain limit - use a **cumulative frequency distribution (less than).**

- **Cumulate down (add the frequencies downwards).**

3. To know the percentile or number of observations above a certain limit - use a **decumulative frequency distribution.**

- Cumulate up (add the frequencies upwards).

			UNZA			
			Less than		**Greater than**	
Class interval	**Frequency**	**Frequency (%)**	**CF**	**CF (%)**	**DCF**	**DCF (%)**
19.5 - 24.5	8	32	8	32	25	100

24.5 - 29.5	6	24	14	56	17	68
29.5 - 34.5	3	12	17	68	11	44
34.5 - 39.5	2	8	19	76	8	32
39.5 - 44.5	2	8	21	84	6	24
44.5 - 49.5	2	8	23	92	4	16
49.5 - 54.5	2	8	25	100	2	8
Total	**25**	**100**				

INTERPRETATION

1. **Cumulative Distribution:**

   Use the true upper limit as the point of comparison thus:

   Seventeen (17) UNZA employees or 68% of them were aged below 34.5 years.

2. **Decumulative distribution**:

   Use the true lower limit or the point of comparison thus:

   Eleven (11) UNZA employees or 44% of them were aged above 29.5 years.