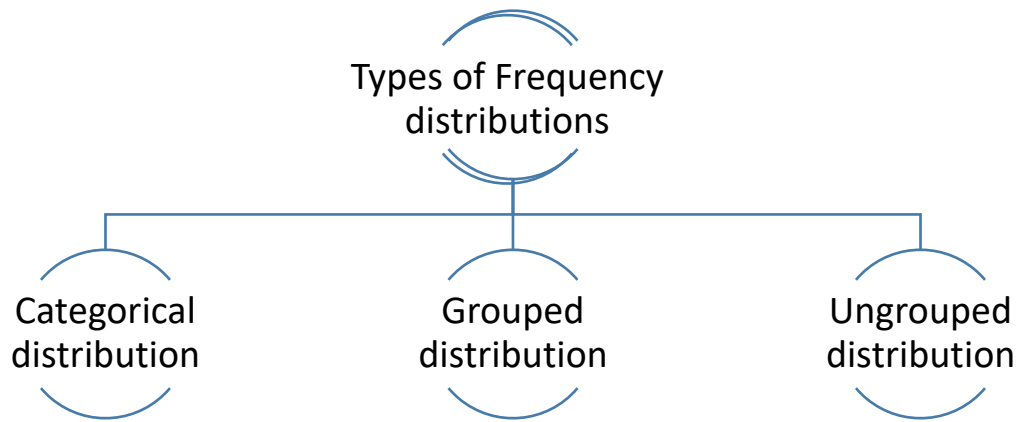


CHAPTER TWO

Data Organization

➤ Frequency distribution table



When data are collected in original form, they are called **raw data**.
For example: row data

2	5	8	7	2	2
6	8	5	2	5	7
4	5	6	2	8	6

A frequency distribution is the organization of raw data in table form, using classes and frequencies. The researches organized the raw data into

Score	<i>f</i>
8	3
7	2
6	3
5	4
4	1
2	5

Categorical Frequency Distribution

Categorical Frequency Distribution: can be used for data that can be placed in specific categories, such as nominal- or ordinal-level data.

Example: Twenty-five army indicates were given a blood test to determine their blood type.

Raw Data: A,B,B,AB,O O,O,B,AB,B B,B,O,A,O
A,O,O,O,AB AB,A,O,B,A

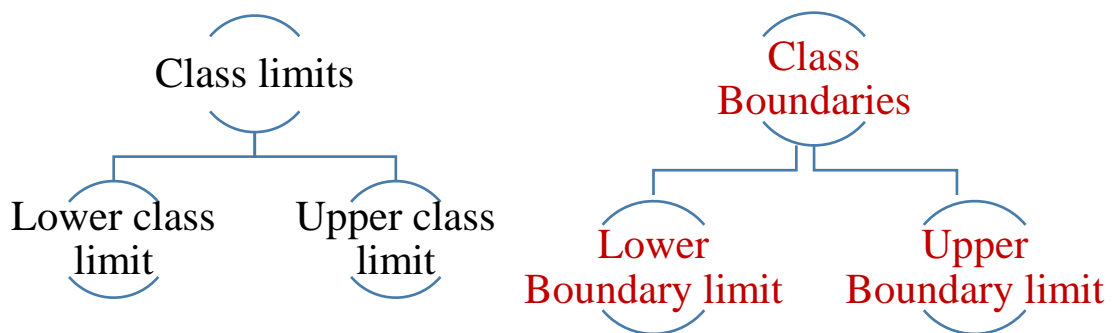
$$\text{Percent} = \frac{f}{n} * 100$$

Class	Tally	Frequency (<i>f</i>)	Percent
A	IIII	5	20
B	IIII II	7	28
O	IIII IIII	9	36
AB	IIII	4	16
		n=25	100

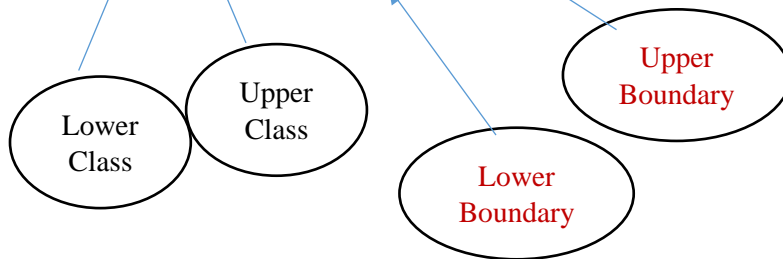
Grouped Frequency Distribution

- **Grouped frequency distributions** can be used when the range of values in the data set is very large. The data must be grouped into classes that are more than one unit in width. For example the life of boat batteries in hours.
- The smallest and largest possible data values in a class are the *lower* and *upper class limits*. *Class boundaries* separate the classes.
- To find a class boundary, average the upper class limit of one class and the lower class limit of the next class.

- The **class width** can be calculated by subtracting
 - successive lower class limits (or boundaries)
 - successive upper class limits (or boundaries)
 - upper and lower class boundaries
- The **class midpoint X_m** can be calculated by averaging
 - upper and lower class limits (or boundaries)



Class limits	Class Boundaries	Tally	Frequency (f)
24 - 30	23.5 - 30.5	III	3
31 - 37	30.5 - 37.5	I	1
38 - 44	37.5 - 44.5	III	5
45 - 51	44.5 - 51.5	III III	9
52 - 58	51.5 - 58.5	III I	6
59 - 65	58.5 - 65.5	I	1



- In the life of boat batteries example, the values 24 and 30 of the first class are the **class limits**.
- The **lower class** limit is 24 and the **upper class** limit is 30.
- **The Class boundaries are used to** separate the classes. So that there are no gaps in the frequency distribution

➤ Lower boundary = lower limit - 0.5

➤ Upper boundary = upper limit + 0.5

- Class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.

For example: Class limit 7.8 – 8.8

Class boundary 7.75 – 8.85

➤ Lower boundary = lower limit - 0.05
= 7.8 - 0.05 = 7.75

➤ Upper boundary = upper limit + 0.05
= 8.8 + 0.05 = 8.85

Class width = lower of second class limit - lower of first class limit

Or

Class width = Upper of second class limit - Upper of first class limit

Class width: 31 – 24 = 7

The class midpoint X_m is found by adding the lower and upper class limit (or boundary) and dividing by 2.

$$X_m = \frac{\text{lower limit} + \text{upper limit}}{2}$$

Or

$$X_m = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

For Example: $\frac{24+30}{2} = 27$, $\frac{23.5 + 30.5}{2} = 27$

- Find the boundaries for the following class limits:
 - 44 - 37
 - 10.3 - 11.5
 - 22.2 – 23.0
 - 547.04 - 553.20
- Find the class width for the following class limits:
 - 37 – 44
 - 45 – 52
 - 625 – 654
 - 655 - 684
- Find the class width for the following class boundaries:
 - 10.5 – 11.5
 - 22.15 – 27.15

Rules for Classes in Grouped Frequency Distributions

1. There should be 5-20 classes.
2. The class width should be an odd number.
3. The classes must be mutually exclusive.

Age
10 – 20
20 – 30
30 – 40
40 – 50

Better way to construct
a frequency distribution



Age
10 – 20
21 – 31
32 – 42
43 – 53

4. The classes must be continuous.
5. The classes must be exhaustive.
6. The classes must be equal in width (except in open-ended distributions).

Procedure for Constructing a Grouped Frequency Distribution

- **STEP 1** Determine the classes.
 - ✓ Find the highest and lowest value
 - ✓ Find the range
 - ✓ Select the number of classes desired.
 - ✓ Find the width by divided the range by the number of classes and rounding up.
 - ✓ Select a starting point (usually the lowest value), add the width to get the lower limits.
 - ✓ Find the upper class limits.
 - ✓ Find the boundaries.
- **STEP 2** Tally the data.
- **STEP 3** Find the frequencies.
- **STEP 4** Find the cumulative frequencies by keeping a running total of the frequencies.

Constructing a Grouped Frequency Distribution

Example

The following data represent the record high temperatures for each of the 50 states. Construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

STEP 1 Determine the classes. Find the class width by dividing the range by the number of classes 7.

Range = High – Low

$$= 134 - 100 = 34$$

$$\text{Width} = \frac{\text{Range}}{7} = \frac{34}{7} = 5$$

Note: Rounding Rule: Always round up if a remainder.

STEP 2 Tally the data.

STEP 3 Find the frequencies.

Class Limits	Class Boundaries	Frequency	Cumulative Frequency
100 - 104	99.5 - 104.5	2	
105 - 109	104.5 - 109.5	8	
110 - 114	109.5 - 114.5	18	
115 - 119	114.5 - 119.5	13	
120 - 124	119.5 - 124.5	7	
125 - 129	124.5 - 129.5	1	
130 - 134	129.5 - 134.5	1	

STEP 4 Find the cumulative frequencies by keeping a running total of the frequencies.

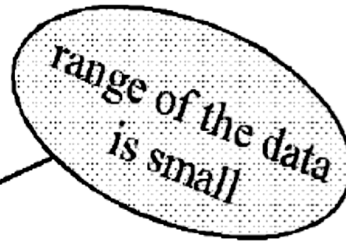
Class Limits	Class Boundaries	Frequency	Cumulative Frequency
100 - 104	99.5 - 104.5	2	2
105 - 109	104.5 - 109.5	8	10
110 - 114	109.5 - 114.5	18	28
115 - 119	114.5 - 119.5	13	41
120 - 124	119.5 - 124.5	7	48
125 - 129	124.5 - 129.5	1	49
130 - 134	129.5 - 134.5	1	50

Ungrouped Frequency distribution:

Example : The data shown here represent the number of miles per gallon that 30 selected four-wheel- drive sports utility vehicles obtained in city driving.

12 17 12 14 16 18
 16 18 12 16 17 15
 15 16 12 15 16 16
 12 14 15 12 15 15
 19 13 16 18 16 14

STEP 1 Determine the classes.
 The rang of the data set is small .



$$\text{Range} = \text{High} - \text{Low}$$

$$= 19 - 12 = 7$$

So the class consisting of the single data value can be used.
 They are 12,13,14,15,16,17,18,19.

□ This type of distribution is called ungrouped frequency distribution

STEP 2 Tally the data.

STEP 3 Find the frequencies.

Class Limits	Class Boundaries	Frequency	Cumulative Frequency
12	11.5-12.5	6	0
13	12.5-13.5	1	6
14	13.5-14.5	3	7
15	14.5-15.5	6	10
16	15.5-16.5	8	16
17	16.5-17.5	2	24
18	17.5-18.5	3	26
19	18.5-19.5	1	29
			30