

Chapter (2)

*Strategies for understanding
the meanings of Data*

Pages(19 – 27)

◆ Key words

frequency table, bar chart ,range
width of interval , mid-interval
Histogram , Polygon

Descriptive Statistics

Frequency Distribution for Discrete Random Variables

Example:

Suppose that we take a **sample** of size 16 from children in a primary school and get the following data about the number of their decayed teeth,

3,5,2,4,0,1,3,5,2,3,2,3,3,2,4,1

To construct a **frequency table**:

1- **Order** the values from the smallest to the largest.

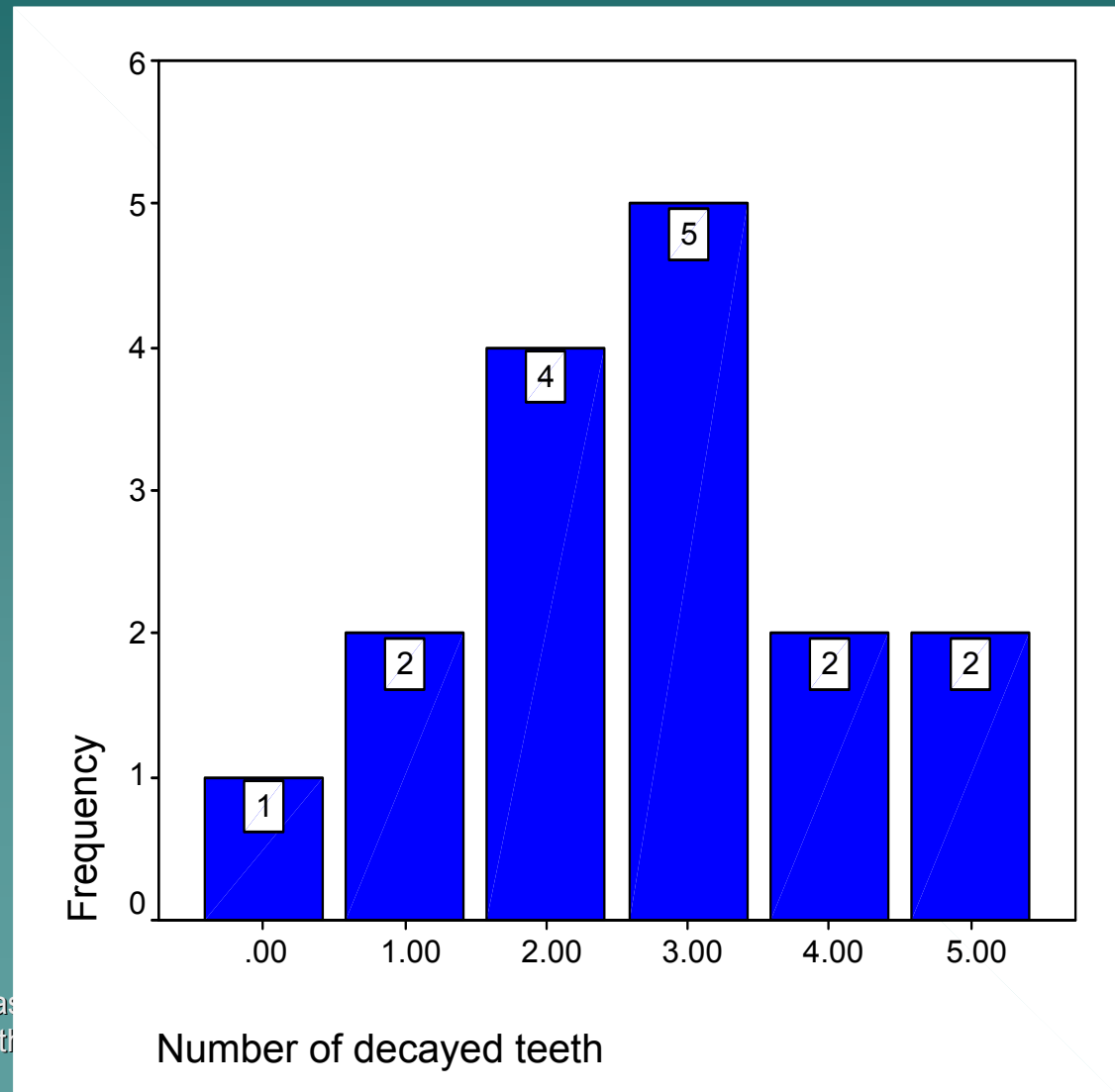
0,1,1,2,2,2,2,3,3,3,3,3,4,4,5,5

2- **Count** how many numbers are the same.

No. of decayed teeth	Frequency	Relative Frequency
0	1	0.0625
1	2	0.125
2	4	0.25
3	5	0.3125
4	2	0.125
5	2	0.125
Total	16	1

Representing the simple frequency table using the bar chart

We can represent the above simple frequency table using the bar chart.



2.3 Frequency Distribution for Continuous Random Variables

For **large samples**, we can't use the simple frequency table to represent the data.

We need to **divide** the data into **groups** or **intervals** or **classes**.

So, we need to determine:

1- The number of intervals (k).

Too few intervals are not good because information will be lost.

Too many intervals are not helpful to summarize the data.

A commonly followed rule is that $6 \leq k \leq 15$,
or the following formula may be used,

$$k = 1 + 3.322 (\log n)$$

2- The range (R).

It is the difference between the largest and the smallest observation in the data set.

3- The Width of the interval (w).

Class intervals generally should be of the **same width**. Thus, if we want k intervals, then w is chosen such that $w \geq R / k$.

Example:

Assume that the number of observations equal 100, then

$$k = 1 + 3.322(\log 100) \\ = 1 + 3.3222(2) = 7.6 \cong 8.$$

Assume that the smallest value = 5 and the largest one of the data = 61, then

$$R = 61 - 5 = 56 \text{ and}$$

$$w = 56 / 8 = 7.$$

To make the summarization more comprehensible, the class width may be 5 or 10 or the multiples of 10.

Example 2.3.1

- ◆ We wish to know how many class interval to have in the frequency distribution of the data in Table 1.4.1 Page 9-10 of ages of 189 subjects who Participated in a study on smoking cessation
- ◆ **Solution :**
- ◆ Since the number of observations equal 189, then
- ◆ $k = 1 + 3.322(\log 189)$
- ◆ $= 1 + 3.3222 (2.276) \cong 9,$
- ◆ $R = 82 - 30 = 52$ and
- ◆ $w = 52 / 9 = 5.778$
- ◆ It is better to let $w = 10$, then the intervals
- ◆ will be in the form:

Class interval	Frequency
30 – 39	11
40 – 49	46
50 – 59	70
60 – 69	45
70 – 79	16
80 – 89	1
Total	189

Sum of frequency
= sample size = n

The Cumulative Frequency:

It can be computed by adding successive frequencies.

The Cumulative Relative Frequency:

It can be computed by adding successive relative frequencies.

The Mid-interval:

It can be computed by adding the lower bound of the interval plus the upper bound of it and then divide over 2.

For the above example, the following table represents the cumulative frequency, the relative frequency, the cumulative relative frequency and the mid-interval.

$$R.f = \text{freq}/n$$

Class interval	Mid – interval	Frequency Freq (f)	Cumulative Frequency	Relative Frequency R.f	Cumulative Relative Frequency
30 – 39	34.5	11	11	0.0582	0.0582
40 – 49	44.5	46	57	0.2434	-
50 – 59	54.5	-	127	-	0.6720
60 – 69	-	45	-	0.2381	0.9101
70 – 79	74.5	16	188	0.0847	0.9948
80 – 89	84.5	1	189	0.0053	1
Total		189		1	

Example :

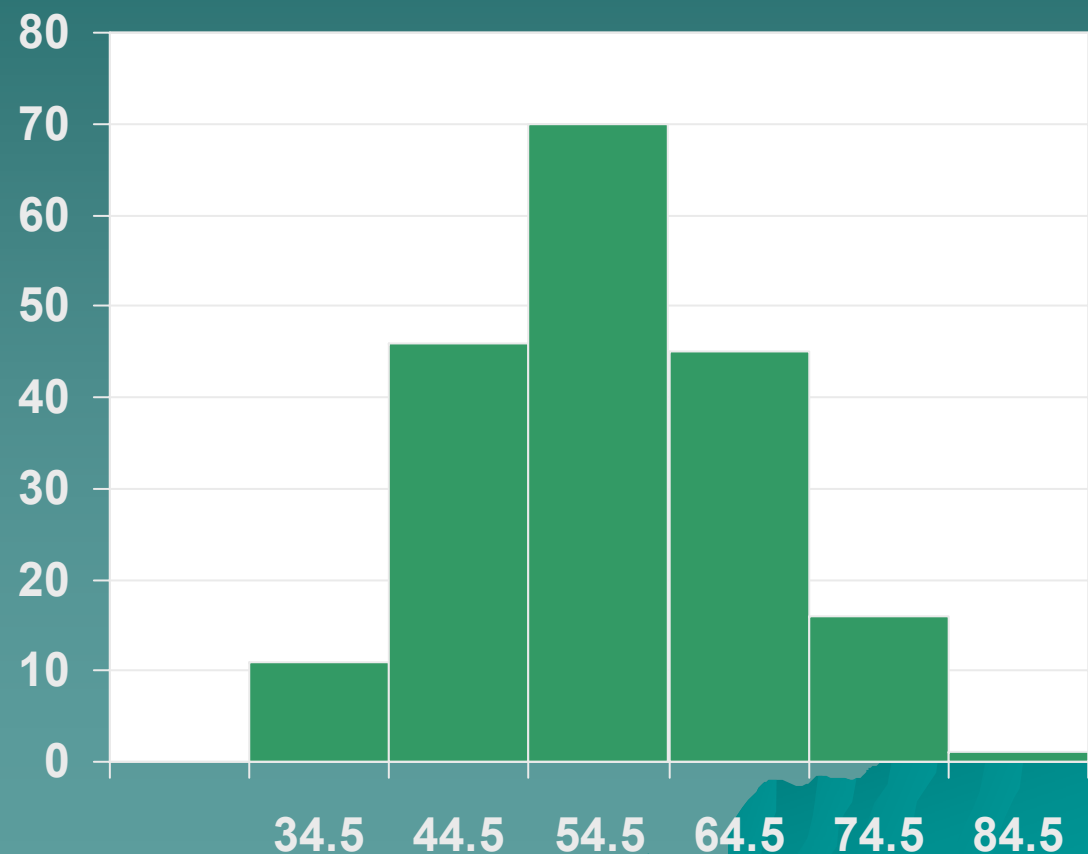
- ◆ From the above frequency table, complete the table then answer the following questions:
- ◆ 1-The number of objects with age less than 50 years ?
- ◆ 2-The number of objects with age between 40-69 years ?
- ◆ 3-Relative frequency of objects with age between 70-79 years ?
- ◆ 4-Relative frequency of objects with age more than 69 years ?
- ◆ 5-The percentage of objects with age between 40-49 years ?

- ◆ 6- The percentage of objects with age less than 60 years ?
- ◆ 7-The Range (R) ?
- ◆ 8- Number of intervals (K)?
- ◆ 9- The width of the interval (W) ?

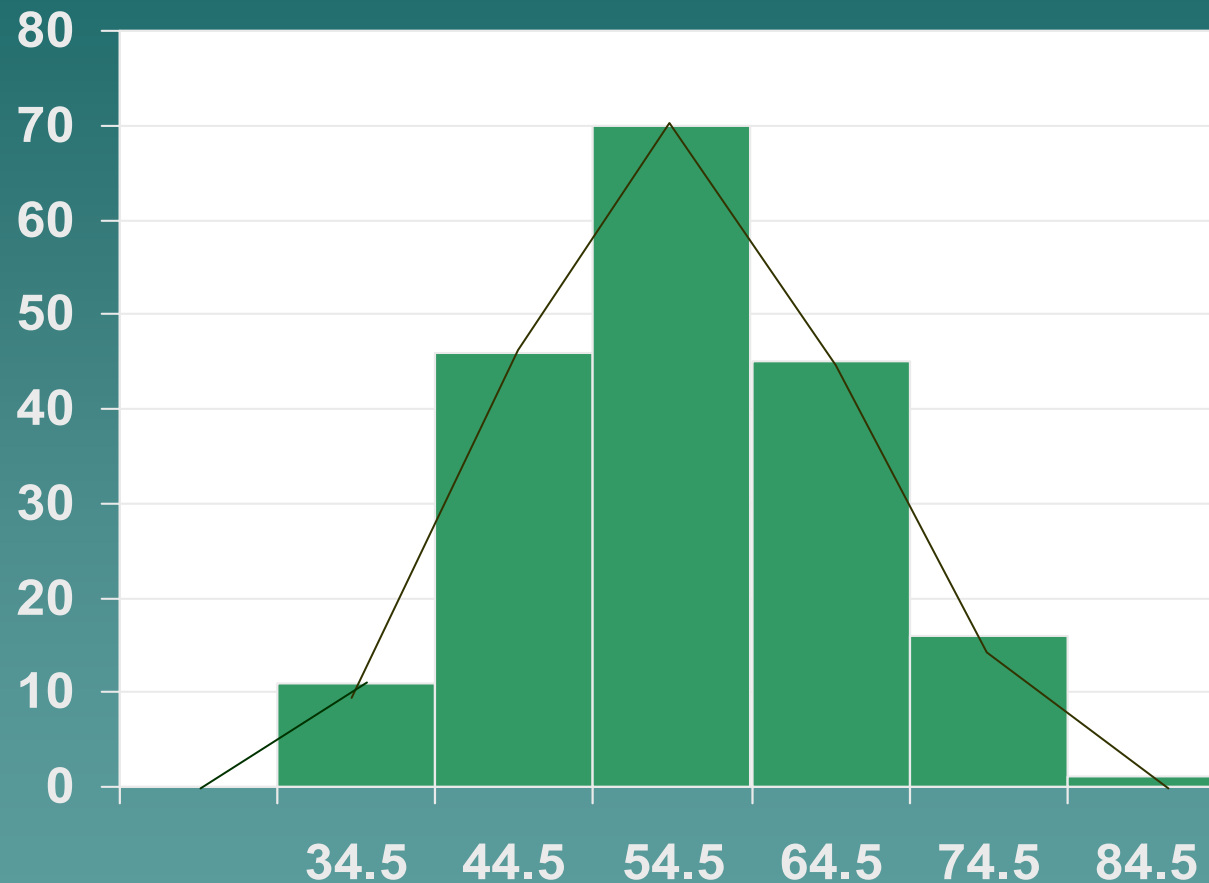
Representing the grouped frequency table using the histogram

To draw the histogram, the **true classes limits** should be used. They can be computed by **subtracting** 0.5 from the **lower** limit and **adding** 0.5 to the **upper** limit for each interval.

True class limits	Frequency
29.5 – <39.5	11
39.5 – < 49.5	46
49.5 – < 59.5	70
59.5 – < 69.5	45
69.5 – < 79.5	16
79.5 – < 89.5	1
Total	189



Representing the grouped frequency table using the Polygon



Exercises

- ◆ Pages : 31 – 34
- ◆ Questions: 2.3.2(a) , 2.3.5 (a)
- ◆ H.W. : 2.3.6 , 2.3.7(a)