Statistics

Lecture 6

Part two

Data analysis: Descriptive statistics

Lecture objectives
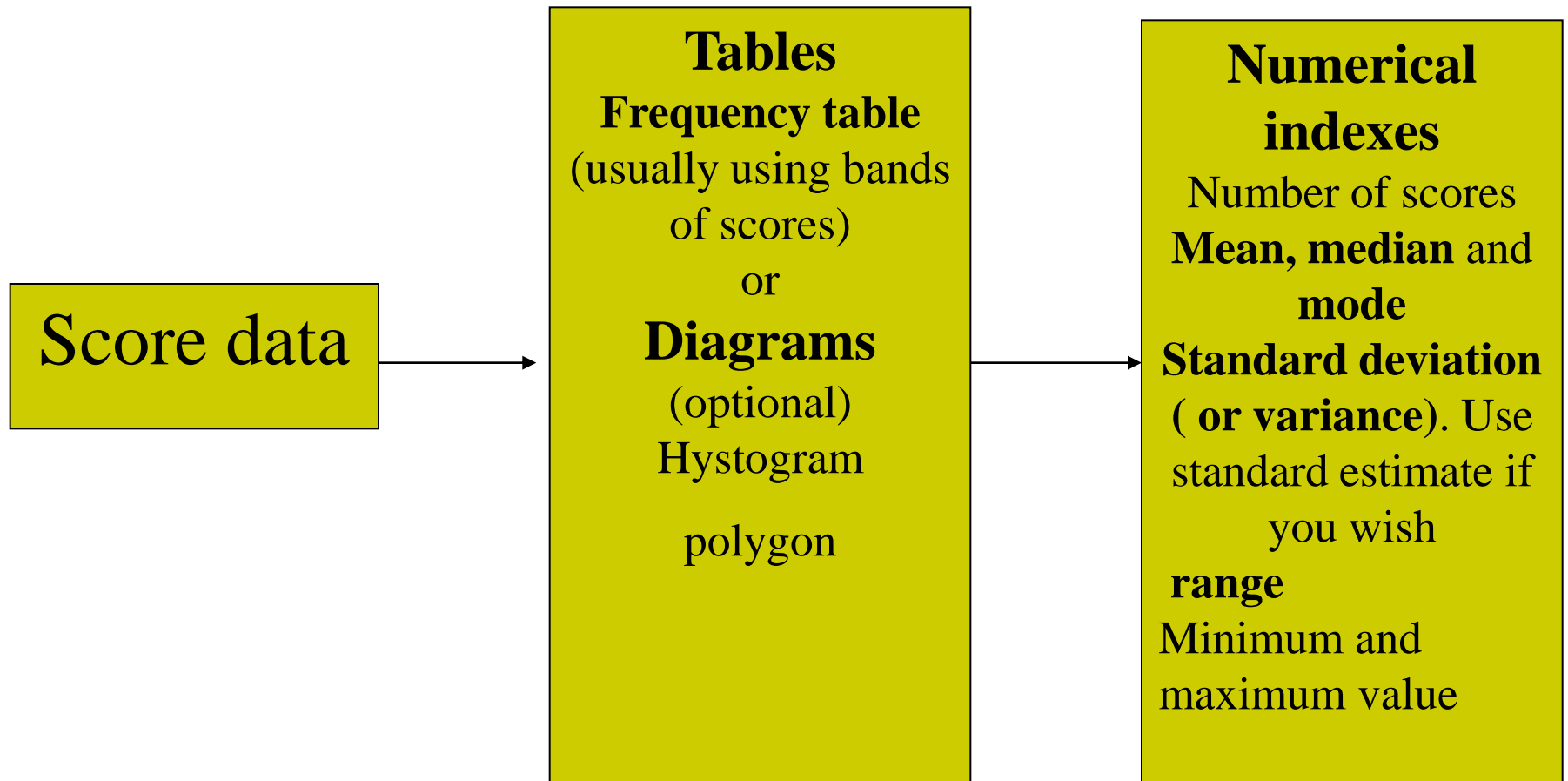
- **Describing and summarizing data**

 - Describing and summarizing numerical data/ score data (quantitative)

Introduction

After having seen how categorical data (qualitative data) can be described and summarized, this lecture introduces how numerical (score data) can also be descibed and summarized. According to (Howitt & Cramer, p. 54), 'because score data contains much more information than category data, there are many more appropriate ways of describing and summarizing score data'

Fig 7.1: Essential descriptive statistics for score variables (Howitt & Cramer, 2005, p.54)

Score data →

Tables
Frequency table
(usually using bands of scores)
or
Diagrams
(optional)
Hystogram

polygon

→

Numerical indexes
Number of scores
Mean, median and **mode**
Standard deviation (or variance). Use standard estimate if you wish
range
Minimum and maximum value

In addition to the frequency, we can use also use **relative frequency** and **cumultative frequency.**

* **Relative frequency refers to how ofeten something happens divided by the number of observations or f/ n (** **as seen in the previous slides)**

* **Cumulative frequency is obtained by summing the frequencies (relative frequencies) of all classes up to the specific class**

Fig 1: Frequency distribution table (example of employees weekly earnings frequencies)

Weekly earnings (dollar)	Frequency
801 to 1000	9
1001 to 1200	22
1201 to 1400	39
1401 to 1600	15
1601 to 1800	9
1801 to 200	6

Fig 2: Frequencies/relative frequencies od students marks (n=50)

Class marks	Frequency	Relative frequency
< 10	13	0.26
10	8	0.16
12	12	0.24
13	10	0.2
14	4	0.08
15	2	0.04
17	1	0.02
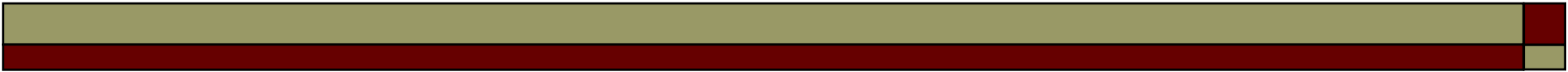
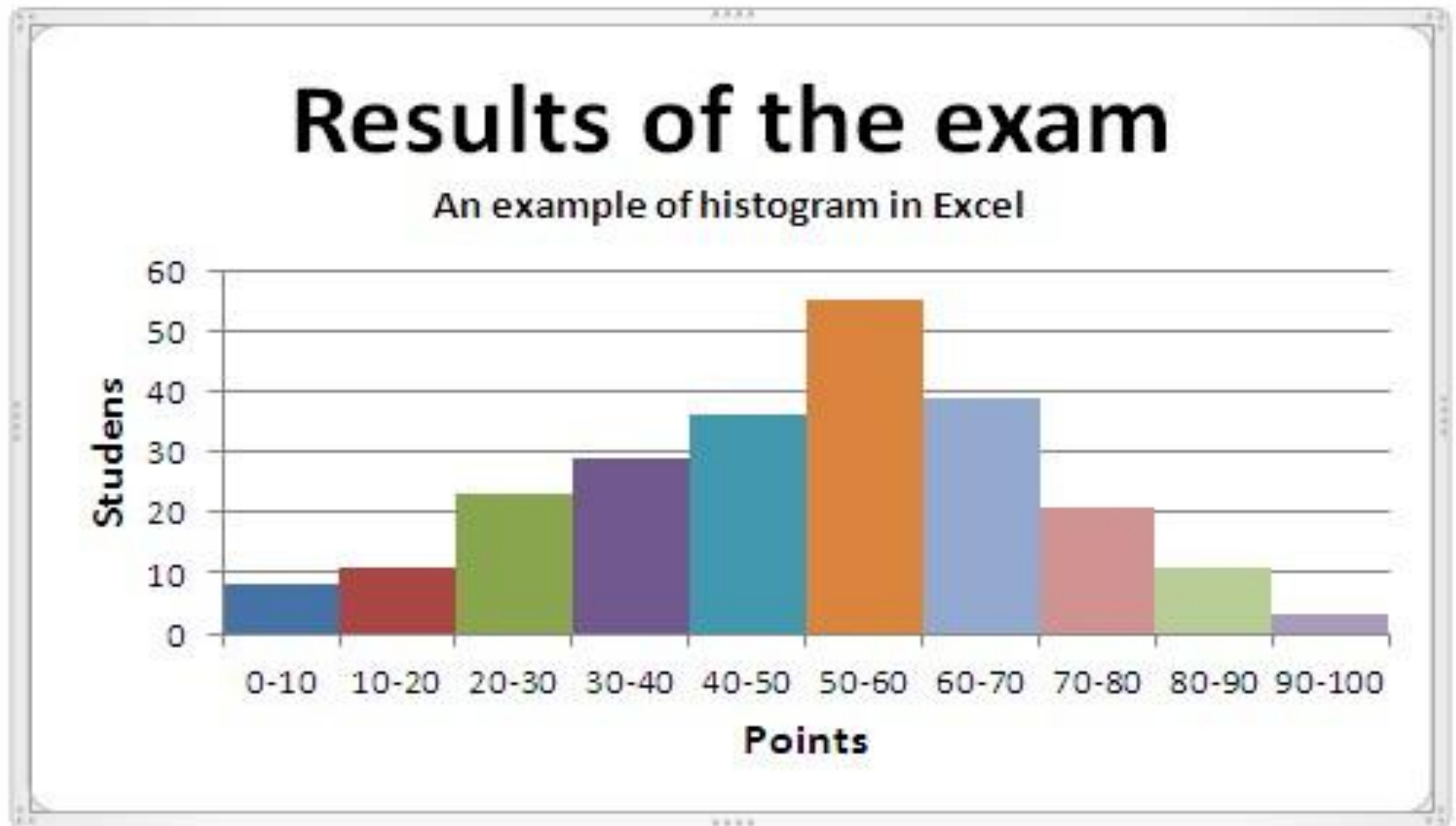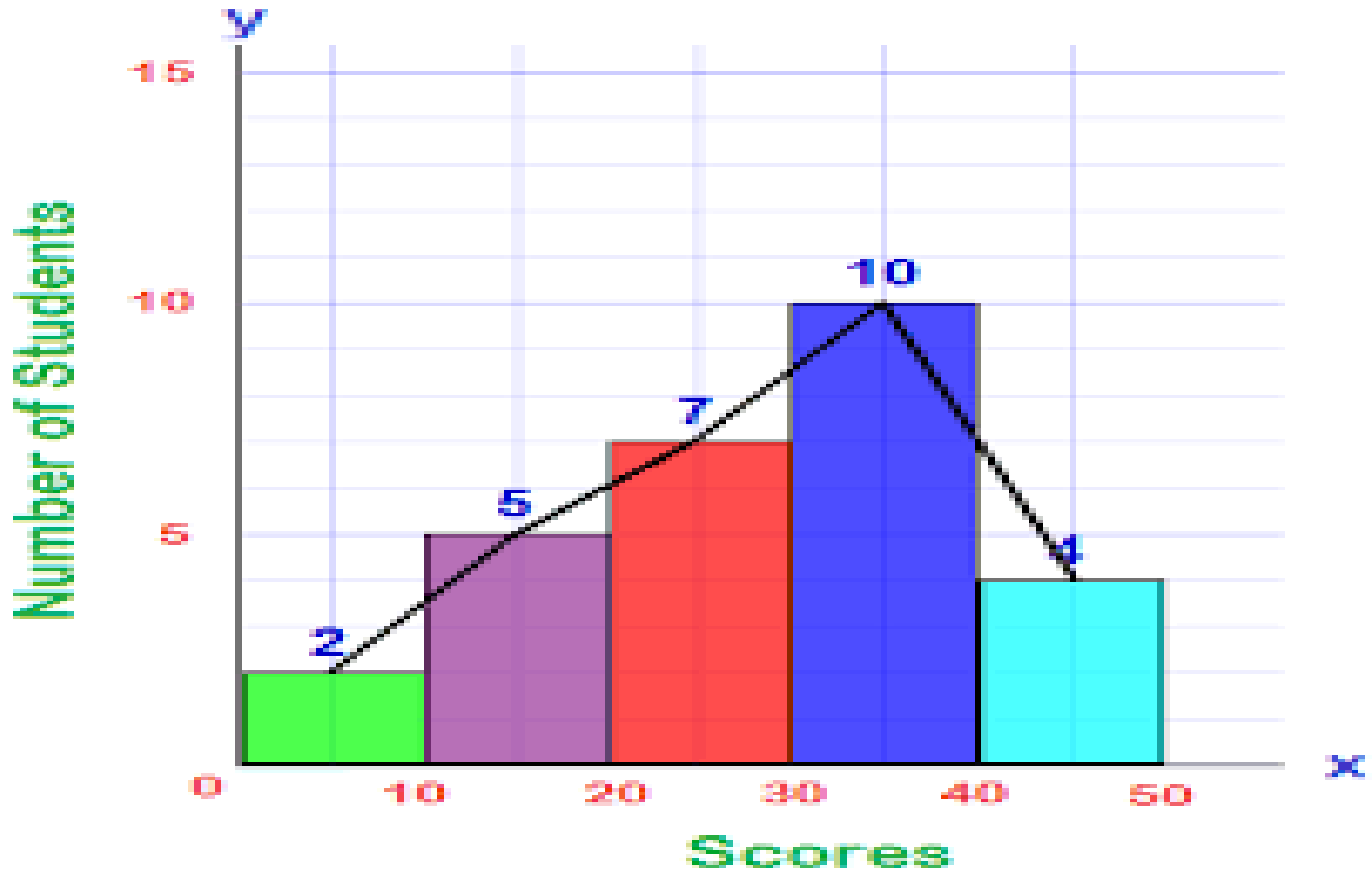| Class marks	Frequency	Cumulative frequency
< 10	13	13
10	8	21
12	12	33
13	10	43
14	4	47
15	2	49
17	1	40

Graphing score data:
1.Histograms

A hystogram can be grawn from a frequency distribution, a relative frequency distribution or a percentage distribution. The bars in a hystogram are drawn adjacent to each other with no gap between them.

* The classes are marked on the horizontal axis and the frequencies, relative frequencies or percentages are represented by the heights of the bars
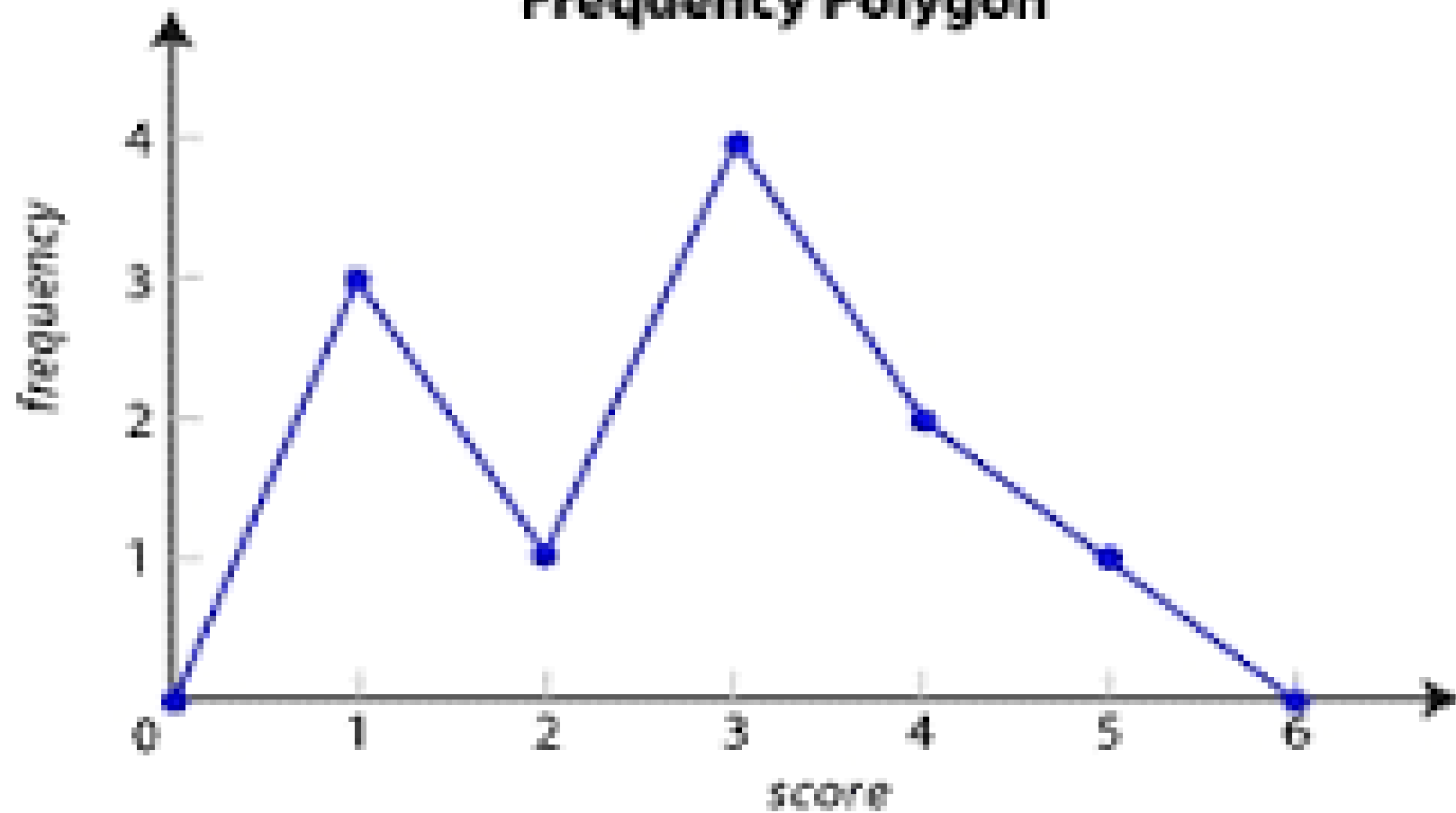
Hystogram representing students marks frequencies

2. Polygon representing students scores

Frequency Polygon

Task 1

Calculate the relative, cumulative and percent frequencies of the following and place them in a frequency distribution table.

Students marks are as follows:

- 3 students got from 1 to 4
- 15 students got from 5 to 9
- 21 student got from 10 to 13
- 5 students got from 14 to 16
- 3 students got from 17 to 19

Task 2

1. Is a frequency polygon similar to a line graph or different?

2. What is a scatter plot/ graph?

3. What is the difference between frequency percentage and relative frequency?

4. What is the importance of cumulative frequencies?