Data presentation

Dr Abubakir M. Saleh
Assist. Prof. of Community Medicine

Outline of the lecture

- Frequency distributions and graphic presentation
  - Binary variables
  - Categorical variables
  - Ordered categorical variables
  - Discrete quantitative variables
  - Continuous variables
- Histogram
- Cumulative frequency curve
- Frequency polygon

Frequency distributions

- The frequencies with which the different possible values of a variable occur in a group of subjects is called the **frequency distribution** of the variable in the group.

Distribution of sample according to sex

Variable (Sex)	Number	(%)
Male	20	(40)
Female	30	(60)
Total	50	(100)

Frequency distribution

- The count of individuals having a particular quality is called the **frequency** of that quality. We usually use the term 'number' or 'No.'

- The proportion of individuals having the quality is called the **relative frequency** or proportional frequency. We use "%"

- The relative frequency (%) of male is 20/50 = 0.4 or 40%.

- The set of frequencies of all the possible categories is called the frequency distribution of the variable. (e.g. frequency distribution of the sex of the students)

Variable (Sex)	Frequency	Relative frequency
Male	20	40
Female	30	60
Total	50	100

Frequency distribution and graphic presentations

Binary variables

- Simple table
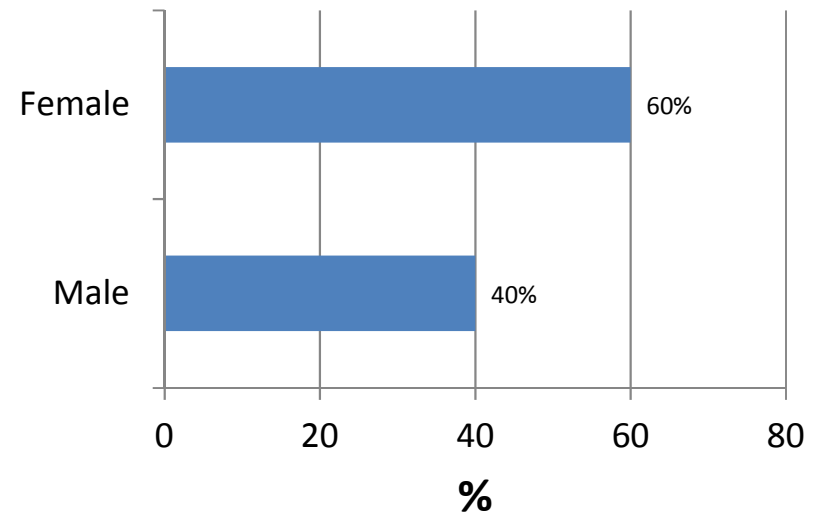- Bar chart
- Pie chart

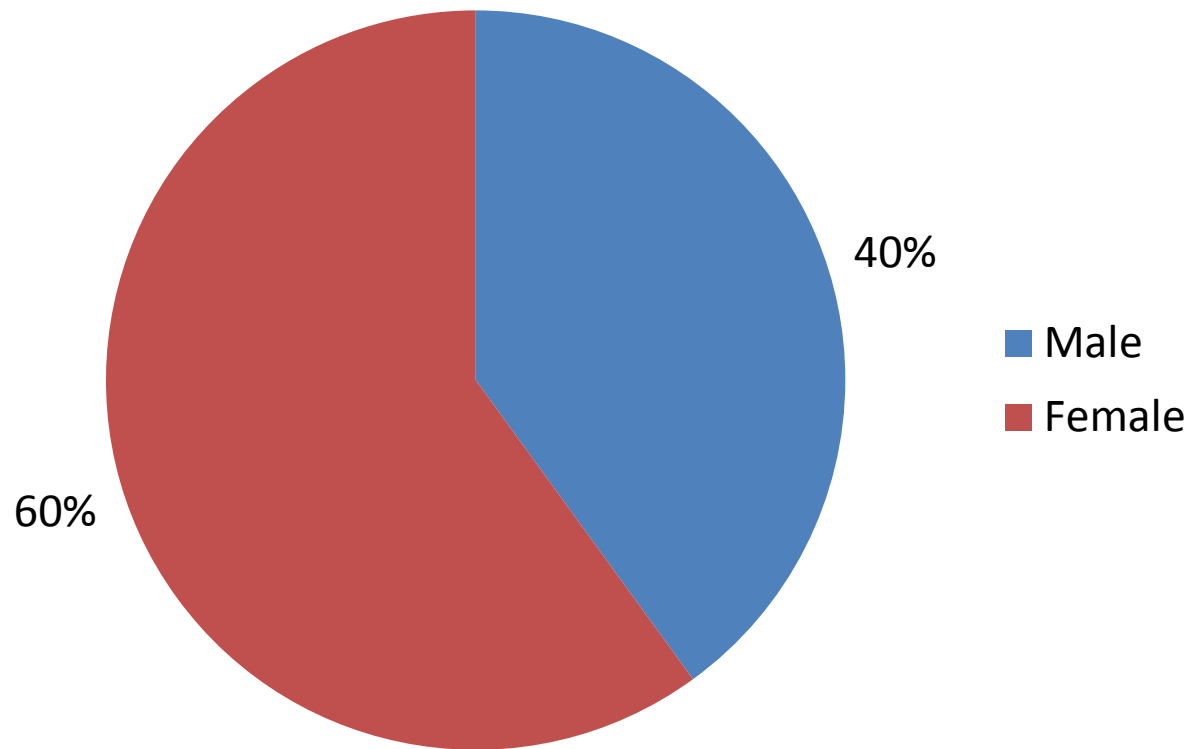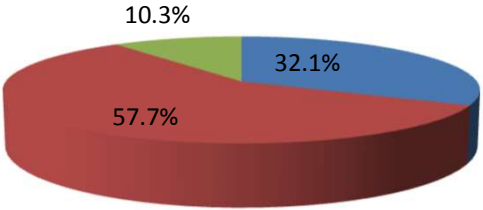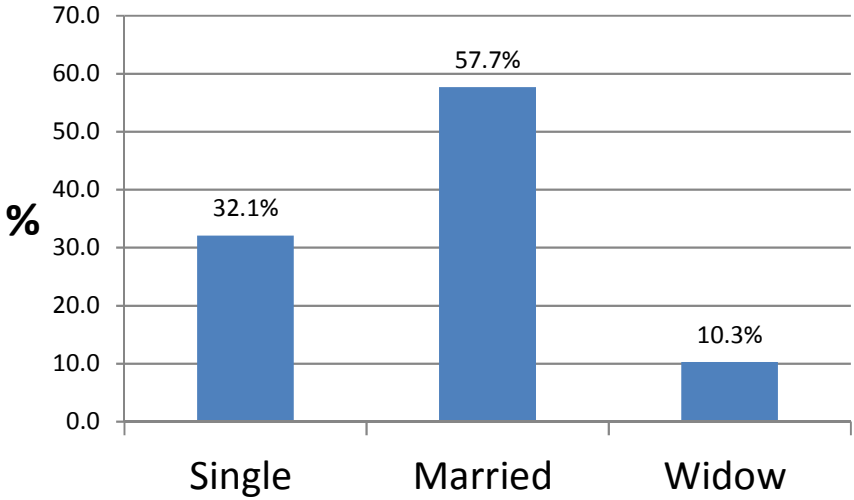| Sex	Number	(%)
Male	20	(40)
Female	30	(60)
Total	50	(100)

Bar chart

Vertical
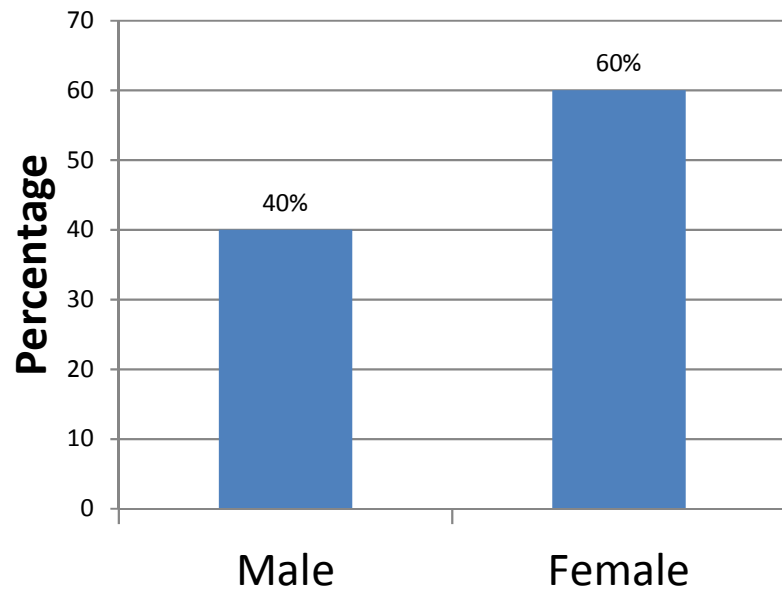
Horizontal

Pie chart

Categorical variable

- Very similar to binary variables
- Table
- Bar chart
- Pie chart

Marital status	Number	(%)
Single	25	(32.1)
Married	45	(57.7)
Widow	8	(10.3)
Total	78	(100.0)

Categorical

Ordered categorical variables

In addition to frequency and relative frequency of a value, we can show also:

- The **cumulative frequency**: the number of individuals with values less than or equal to that value.

Variable (Disease severity)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
Mild	56	29.3	56	
Moderate	87	45.5		
Severe	48	25.1		
Total	191	100.0		

Ordered categorical variables

In addition to frequency and relative frequency of a value, we can show also:

The **cumulative frequency**: the number of individuals with values less than or equal to that value.

Variable (Disease severity)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
Mild	56	29.3	56	
Moderate	87	45.5	143	
Severe	48	25.1		
Total	191	100.0		

Ordered categorical variables

In addition to frequency and relative frequency of a value, we can show also:
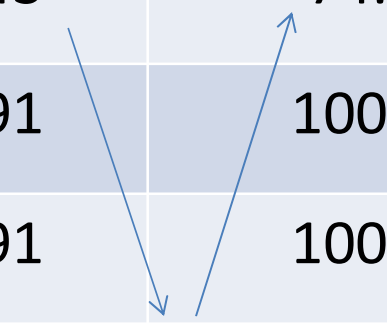
- The **cumulative frequency**: the number of individuals with values less than or equal to that value.

Variable (Disease severity)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
Mild	56	29.3	56	
Moderate	87	45.5	143	
Severe	48	25.1	191	
Total	191	100.0	191	

We can also show:

The **relative cumulative frequency**: the proportion of individuals in the sample with values less than or equal to that value.

Variable (Disease severity)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
Mild	56	29.3	56	29.3
Moderate	87	45.5	143	74.9
Severe	48	25.1	191	100.0
Total	191	100.0	191	100.0

143/191*100

Discrete quantitative variable

- We can count the number of times each possible value occurs to get the frequency distribution

Variable (Household size)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
1	53	12.6		
2	78	18.6		
3	112	26.7		
4	105	25.0		
5	72	17.1		
Total	420	100		

Discrete quantitative variable

- We can count the number of times each possible value occurs to get the frequency distribution

Variable (Household size)	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
1	53	12.6	53	12.6
2	78	18.6	131	31.2
3	112	26.7	243	57.9
4	105	25.0	348	82.9
5	72	17.1	420	100
Total	420	100	420	100

Discrete quantitative variable

- Use a bar chart
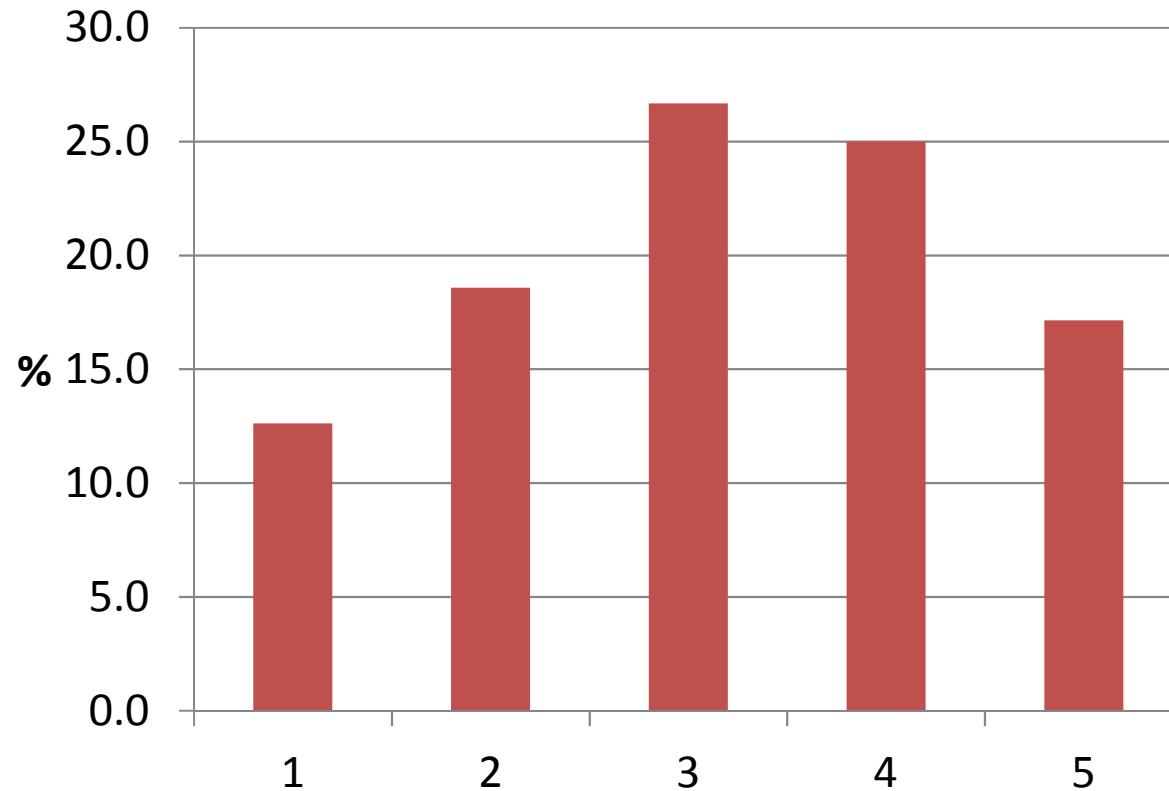
Continuous variables

Hemoglobin measurement of 40 adults (mg/dl)

7.2	14.6	10.5	13.6
13.7	11.7	10.6	10.9
14.2	12.9	11.5	13.4
13.5	11.7	15.2	12.1
8.3	12.1	11.2	10.2
12.2	12.5	11.4	14.5
13.9	9.4	12.6	8.7
11.3	10.2	11.4	9.5
12.3	14.9	12.7	12.5
11.9	14.3	13.1	13.2

Continuous variables

- As most of the values occur only once, counting the number of occurrences does not help.

Hb	Frequency	Relative frequency
7.2	1	2.5
8.3	1	2.5
8.7	1	2.5
9.4	1	2.5
9.5	1	2.5
10.2	2	5
10.5	1	2.5
10.6	1	2.5
10.9	1	2.5
11.2	1	2.5
11.3	1	2.5
11.4	2	5
11.5	1	2.5
11.7	2	5
11.9	1	2.5
12.1	2	5
12.2	1	2.5
12.3	1	2.5
12.5	2	5
12.6	1	2.5
12.7	1	2.5
12.9	1	2.5
13.1	1	2.5
13.2	1	2.5
13.4	1	2.5
13.5	1	2.5
13.6	1	2.5
13.7	1	2.5
13.9	1	2.5
14.2	1	2.5
14.3	1	2.5
14.5	1	2.5
14.6	1	2.5
14.7	1	2.5
15.2	1	2.5

Continuous variables

- To get a useful frequency distribution we need to divide the hemoglobin measure into class intervals,

  e.g. from 7.0 to 8, from 8.0 to 9, etc, and count the number of individuals with hemoglobin measure in each class interval.

- The class intervals should not overlap, so we must decide which interval contains the boundary point to avoid it being counted twice.

- It is usual to put the lower boundary of an interval into that interval and the higher boundary into the next interval.

- Thus the interval starting at 7.0 and ending at 8.0 contains 7.0 but not 8.0.

- So it is better to write it in this way, , from 7.0 to 7.99, from 8.0 to 8.99, etc.

Continuous variable

Hb measurement class	Frequency	Relative frequency
7 to 7.9	/	
8 to 8.9	//	
9 to 9.9	//	
10 to 10.9	////	
11 to 11.9	//// ///	
12 to 12.9	//// ////	
13 to 13.9	//// //	
14 to 14.9	////	
15 to 15.9	/	
Total		

Hb	Frequency	Relative frequency
7.2	1	2.5
8.3	1	2.5
8.7	1	2.5
9.4	1	2.5
9.5	1	2.5
10.2	2	5
10.5	1	2.5
10.6	1	2.5
10.9	1	2.5
11.2	1	2.5
11.3	1	2.5
11.4	2	5
11.5	1	2.5
11.7	2	5
11.9	1	2.5
12.1	2	5
12.2	1	2.5
12.3	1	2.5
12.5	2	5
12.6	1	2.5
12.7	1	2.5
12.9	1	2.5
13.1	1	2.5
13.2	1	2.5
13.4	1	2.5
13.5	1	2.5
13.6	1	2.5
13.7	1	2.5
13.9	1	2.5
14.2	1	2.5
14.3	1	2.5
14.5	1	2.5
14.6	1	2.5
14.7	1	2.5
15.2	1	2.5

• Here we changed continuous variables to ordered categorical variables

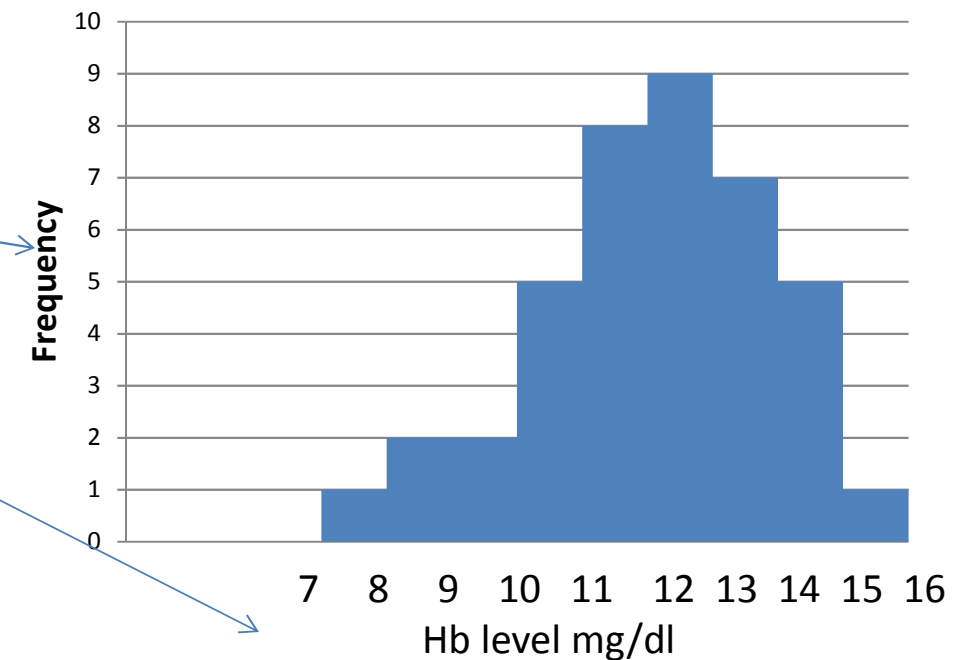| Hb measurement class	Frequency	Relative frequency
7 to 7.9	1	2.5
8 to 8.9	2	5
9 to 9.9	2	5
10 to 10.9	5	12.5
11 to 11.9	8	20
12 to 12.9	9	22.5
13 to 13.9	7	17.5
14 to 14.9	5	12.5
15 to 15.9	1	2.5
Total	40	100

Thus we can present them in frequency distribution table and show the cumulative frequency and relative cumulative frequency

Hb measurement class	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
7 to 7.9	1	2.5	1	2.5
8 to 8.9	2	5.0	3	7.5
9 to 9.9	2	5.0	5	12.5
10 to 10.9	5	12.5	10	25.0
11 to 11.9	8	20.0	18	45.0
12 to 12.9	9	22.5	27	67.5
13 to 13.9	7	17.5	34	85.0
14 to 14.9	5	12.5	39	97.5
15 to 15.9	1	2.5.0	40	100
Total	40	100	40	100

Histogram

- A histogram is a form of bar chart that is used for quantitative variables

- The values for the variable should be grouped (like the Hb example)

- The bars touch one another to indicate the continuous nature of the variable
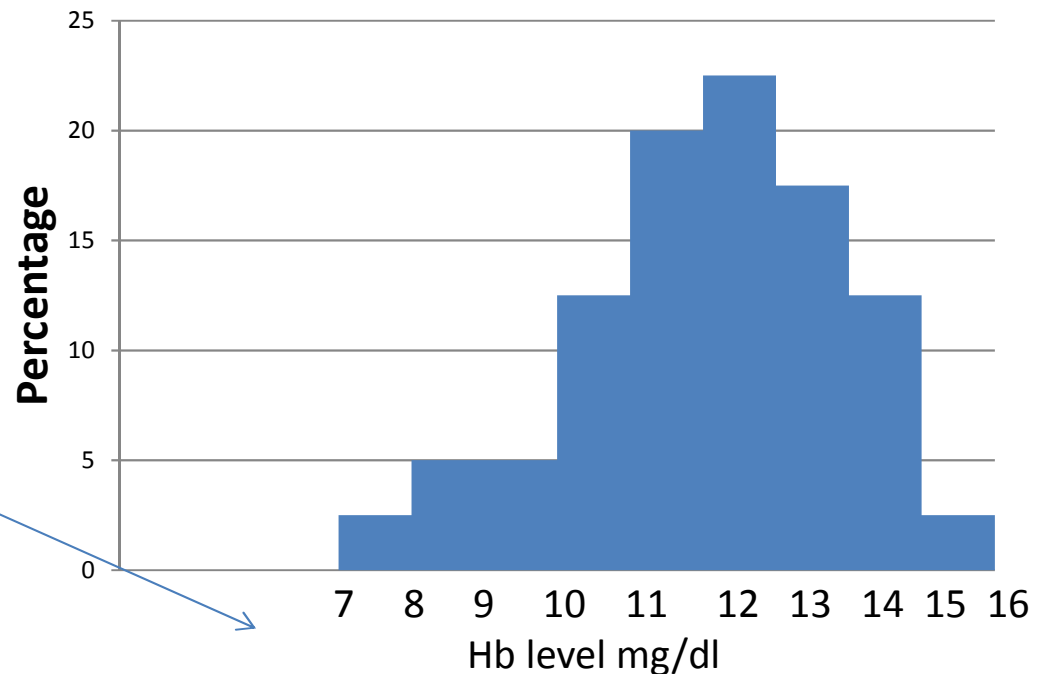
Hb level	Frequency	Relative freq.	Cumul freq	Relative cumulative freq
7 to 7.9	1	2.5	1	2.5
8 to 8.9	2	5.0	3	7.5
9 to 9.9	2	5.0	5	12.5
10 to 10.9	5	12.5	10	25.0
11 to 11.9	8	20.0	18	45.0
12 to 12.9	9	22.5	27	67.5
13 to 13.9	7	17.5	34	85.0
14 to 14.9	5	12.5	39	97.5
15 to 15.9	1	2.5.0	40	100
Total	40	100	40	100

Histogram

- In a histogram, the area of the rectangle represents the frequency (or percentage):
  - The vertical scale is measured in frequency per unit of value
  - The horizontal scale is measured in units of value.

- Note: the rectangles are drawn from 8 up to 9, 9 up to 10, etc., not from 8 up to 8.9, 9 up to 9.9, etc.

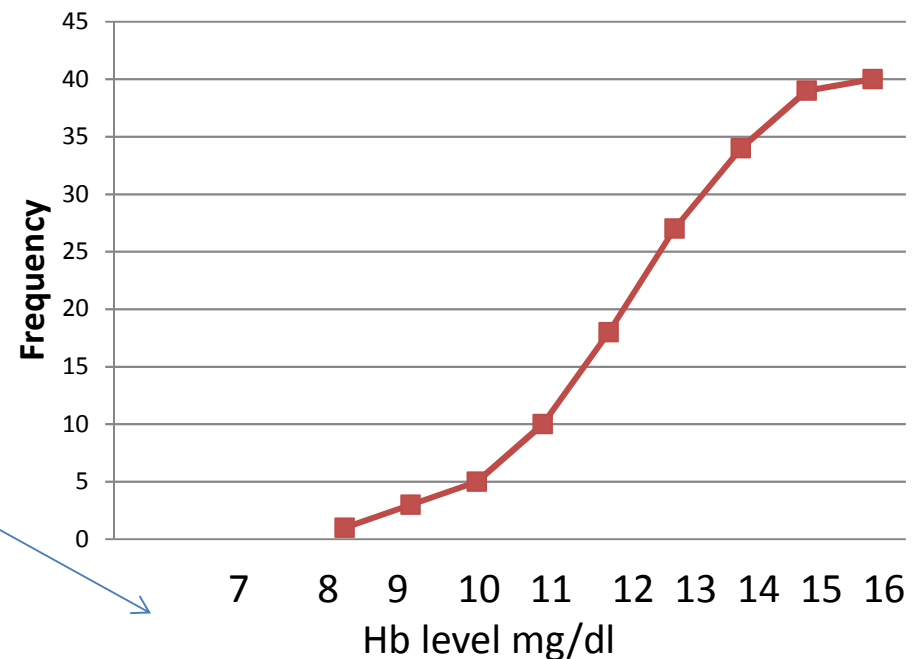| Hb level	Frequency	Relative freq.	Cumulfreq	Relative cumulat freq
7 to 7.9	1	2.5	1	2.5
8 to 8.9	2	5.0	3	7.5
9 to 9.9	2	5.0	5	12.5
10 to 10.9	5	12.5	10	25.0
11 to 11.9	8	20.0	18	45.0
12 to 12.9	9	22.5	27	67.5
13 to 13.9	7	17.5	34	85.0
14 to 14.9	5	12.5	39	97.5
15 to 15.9	1	2.5.0	40	100
Total	40	100	40	100

Cumulative frequency curves

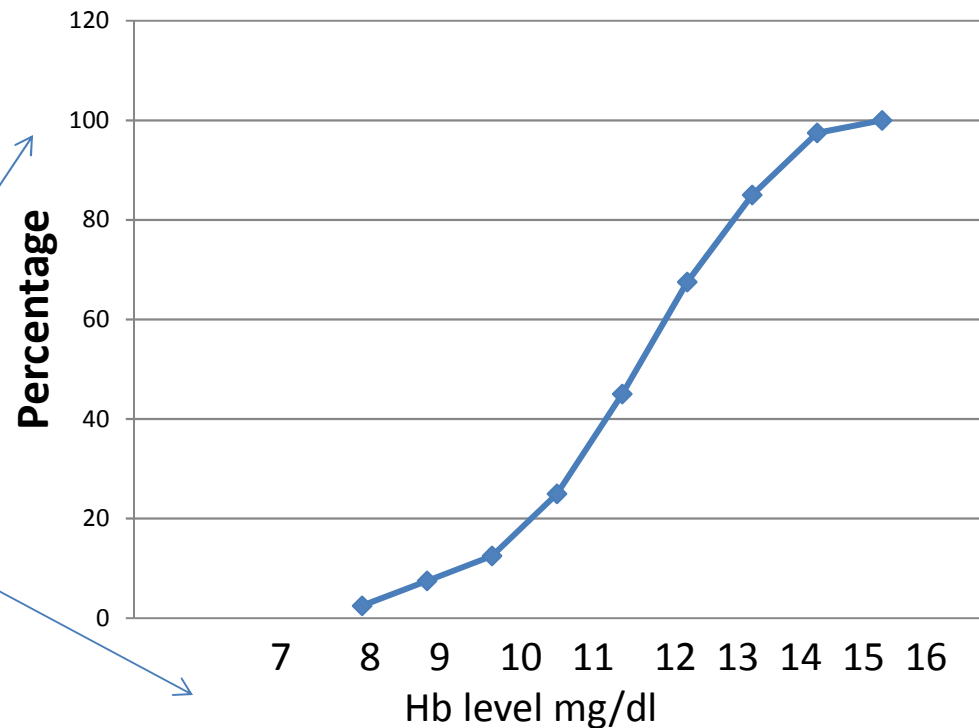- An alternative to the histogram for quantitative variables, is to display the cumulative frequencies.

| Hb measurement class	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
7 to 7.9	1	2.5	1	2.5
8 to 8.9	2	5.0	3	7.5
9 to 9.9	2	5.0	5	12.5
10 to 10.9	5	12.5	10	25.0
11 to 11.9	8	20.0	18	45.0
12 to 12.9	9	22.5	27	67.5
13 to 13.9	7	17.5	34	85.0
14 to 14.9	5	12.5	39	97.5
15 to 15.9	1	2.5.0	40	100
Total	40	100	40	100

Cumulative frequency curves

- The cumulative percentage of people whose haemoglobin level is below 8 is 2.5%, the cumulative percentage below 9 is 7.5%, and so on.

| Hb measurement class	Frequency	Relative frequency	Cumulative frequency	Relative cumulative frequency
7 to 7.9	1	2.5	1	2.5
8 to 8.9	2	5.0	3	7.5
9 to 9.9	2	5.0	5	12.5
10 to 10.9	5	12.5	10	25.0
11 to 11.9	8	20.0	18	45.0
12 to 12.9	9	22.5	27	67.5
13 to 13.9	7	17.5	34	85.0
14 to 14.9	5	12.5	39	97.5
15 to 15.9	1	2.5.0	40	100
Total	40	100	40	100

Frequency polygon
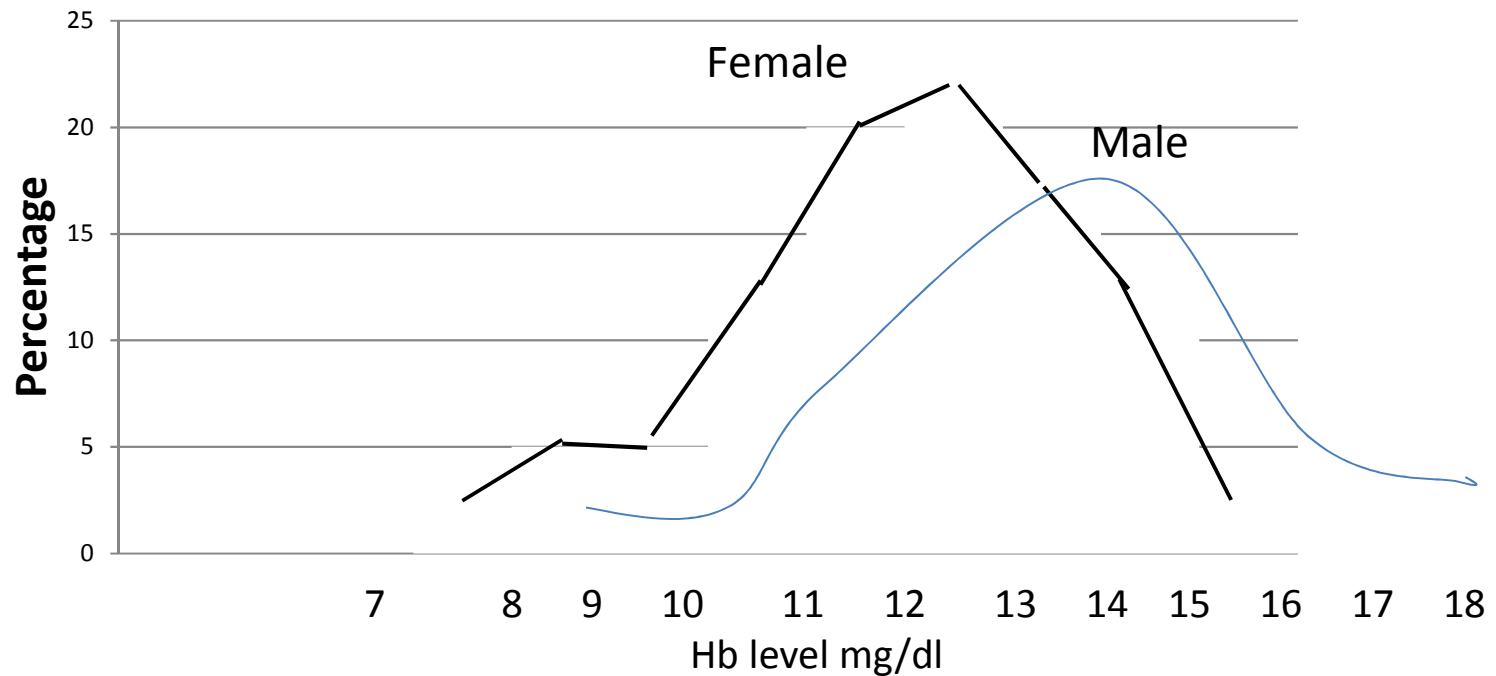
- Join the tops of the bars in the histogram
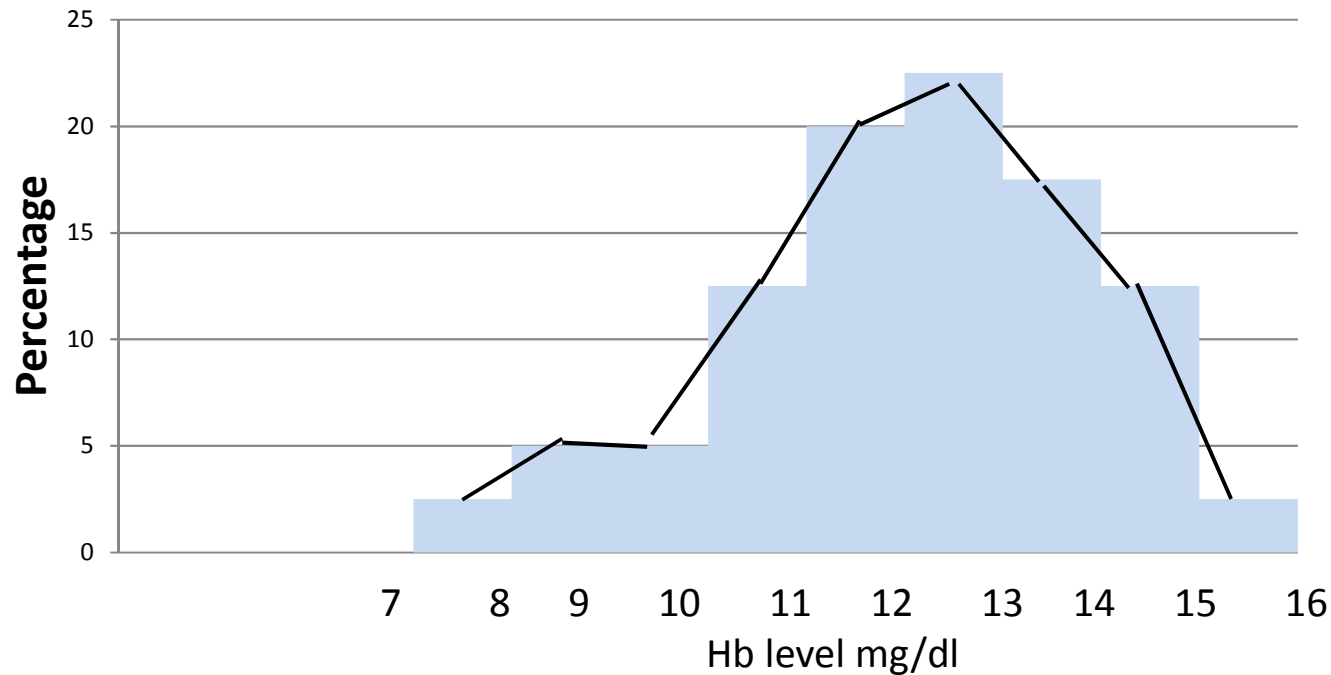
Frequency polygon
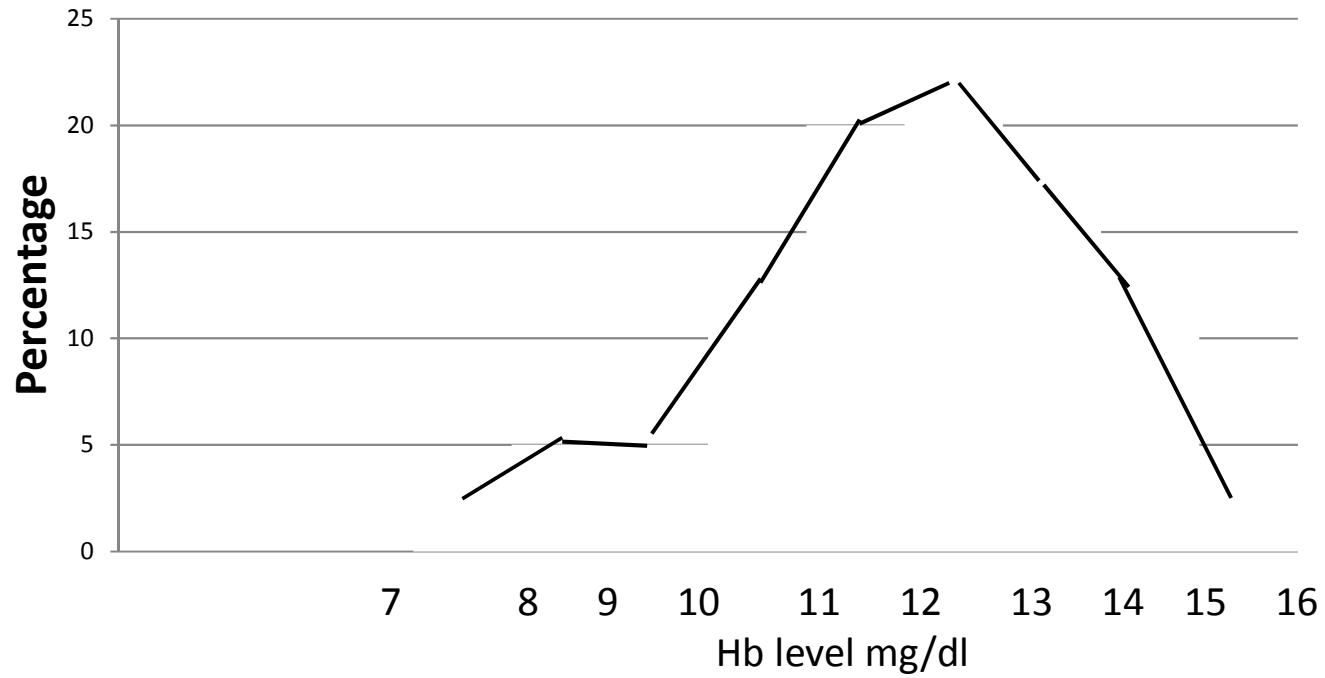
Frequency polygon

- Good for showing more than one distribution on the same axes.

Thank You