AN INTRODUCTION TO STATISTICS

LECTURE 2

BY/ ALY MAHER ABDELFATTAH

DEFINITIONS

Ungrouped data – Data that has not been organized into groups. Also called as raw data.

Data	Frequency
2	8
3	4
5	6
7	7
8	2
9	5

Grouped data - Data that has been organized into groups (into a frequency distribution).

Data	Frequency
2 – 4	5
5 – 7	6
8 – 10	10
11 – 13	8
14 – 16	4
17 – 19	3

WHY THE FREQUENCY DISTRIBUTIONS ARE CONSTRUCTED?

- Large data sets can be summarized.
- Can gain some insight into the nature of data.
- Have a basis for constructing graphs.

EXAMPLE: CONSTRUCT THE FREQUENCY TABLE FOR THE FOLLOWING UNGROUPED DATA

These data represent the record high temperatures in degrees Fahrenheit (°F) 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

Source: *The World Almanac and Book of Facts.*

- Step 1: R = highest value − lowest value =134 − 100= 34
- Step 2: Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.
- Step 3: Find the class width by dividing the range by the number of classes.

$Width = R \,/\, number\ of\ classes$ = 34 / 7 = 4.9 ≈ 5

- Round the answer up to the nearest whole number if there is a remainder: For example, 85 ÷ 6 = 14.167 ≈ 15
- Also, after dividing, if there is no remainder, you will need to add an extra class to accommodate all the data.
- Go through the data set putting a tally in the appropriate class for each data value

EXAMPLE: CONSTRUCT THE FREQUENCY TABLE FOR THE FOLLOWING UNGROUPED DATA

Number of classes k, $2^k > n$

$2^k > 50$

Classes	Tally	Frequency	Class Midpoint
100-105	//	2	$\frac{100+105}{2} = 102.5$
105-110	𝍸 ///	8	107.5
110-115	𝍸 𝍸 𝍸 ///	18	112.5
115-120	𝍸 𝍸 ///	13	117.5
120-125	𝍸 //	7	122.5
125-130	/	1	127.5
130-135	/	1	132.5
		$n = \sum f = 50$	

RELATIVE & CUMULATIVE FREQUENCY DISTRIBUTION

Relative Frequency Distribution

$$\text{relative frequency } r.f = \frac{\text{class frequency}}{\text{sum of all frequencies}}$$

Cumulative Frequency Distribution

$$\text{cumulative frequency for the upper pound of the class k } c.f = \sum_{i=1}^{k} f_i$$

SAMPLING TYPES

Classes	Tally	Frequency
100-105	//	2
105-110	𝍸 ///	8
110-115	𝍸 𝍸 𝍸 ///	18
115-120	𝍸 𝍸 ///	13
120-125	𝍸 //	7
125-130	/	1
130-135	/	1
		$n = \sum f = 50$

R. frequency	$< upper\ bound$	C. frequency
2/50=0.04	< 105	2
0.16	< 110	2+8=10
0.36	< 115	2+8+18=28
0.26	< 120	41
0.14	< 125	48
0.02	< 130	49
0.02	≤ 135	50=n
$sum = 1$		

HISTOGRAMS, FREQUENCY POLYGONS, AND OGIVES FOR GROUPED DATA

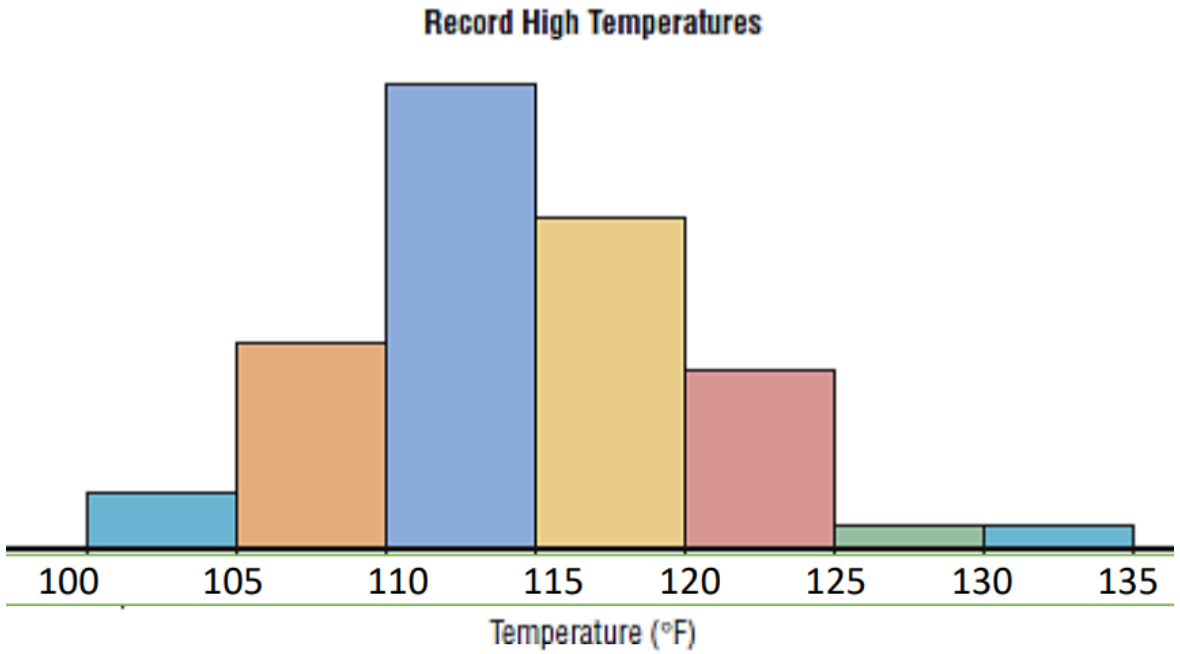
Example: Construct a histogram, polygon and Ogive to represent the data shown for the record high temperatures for each of the 50 states.

Classes	Tally	Frequency
100-105	//	2
105-110	𝍸 ///	8
110-115	𝍸 𝍸 𝍸 ///	18
115-120	𝍸 𝍸 ///	13
120-125	𝍸 //	7
125-130	/	1
130-135	/	1
		$n = \sum f = 50$

HISTOGRAMS, FREQUENCY POLYGONS, AND OGIVES FOR GROUPED DATA

Example: Construct a histogram, polygon and Ogive to represent the data shown for the record high temperatures for each of the 50 states.

Histogram:

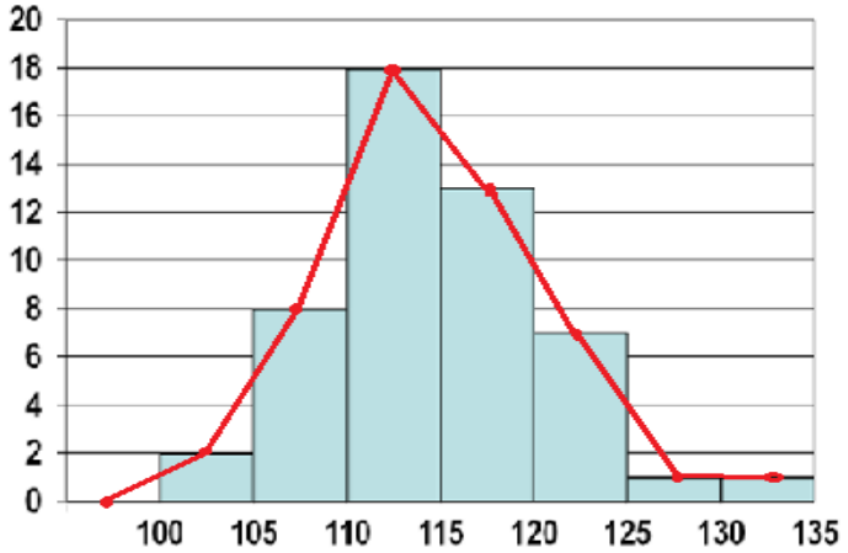


HISTOGRAMS, FREQUENCY POLYGONS, AND OGIVES FOR GROUPED DATA

Example: Construct a histogram, polygon and Ogive to represent the data shown for the record high temperatures for each of the 50 states.

Freq. Polygon:

Frequency	Class Midpoint
2	$\frac{100+105}{2} = 102.5$
8	107.5
18	112.5
13	117.5
7	122.5
1	127.5
1	132.5
$n = \sum f = 50$	

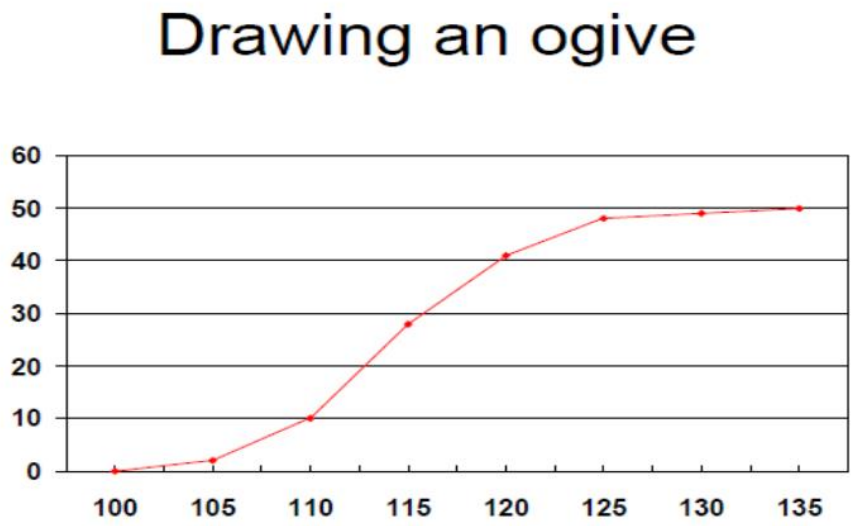


HISTOGRAMS, FREQUENCY POLYGONS, AND OGIVES FOR GROUPED DATA

Example: Construct a histogram, polygon and Ogive to represent the data shown for the record high temperatures for each of the 50 states.

Ogive:

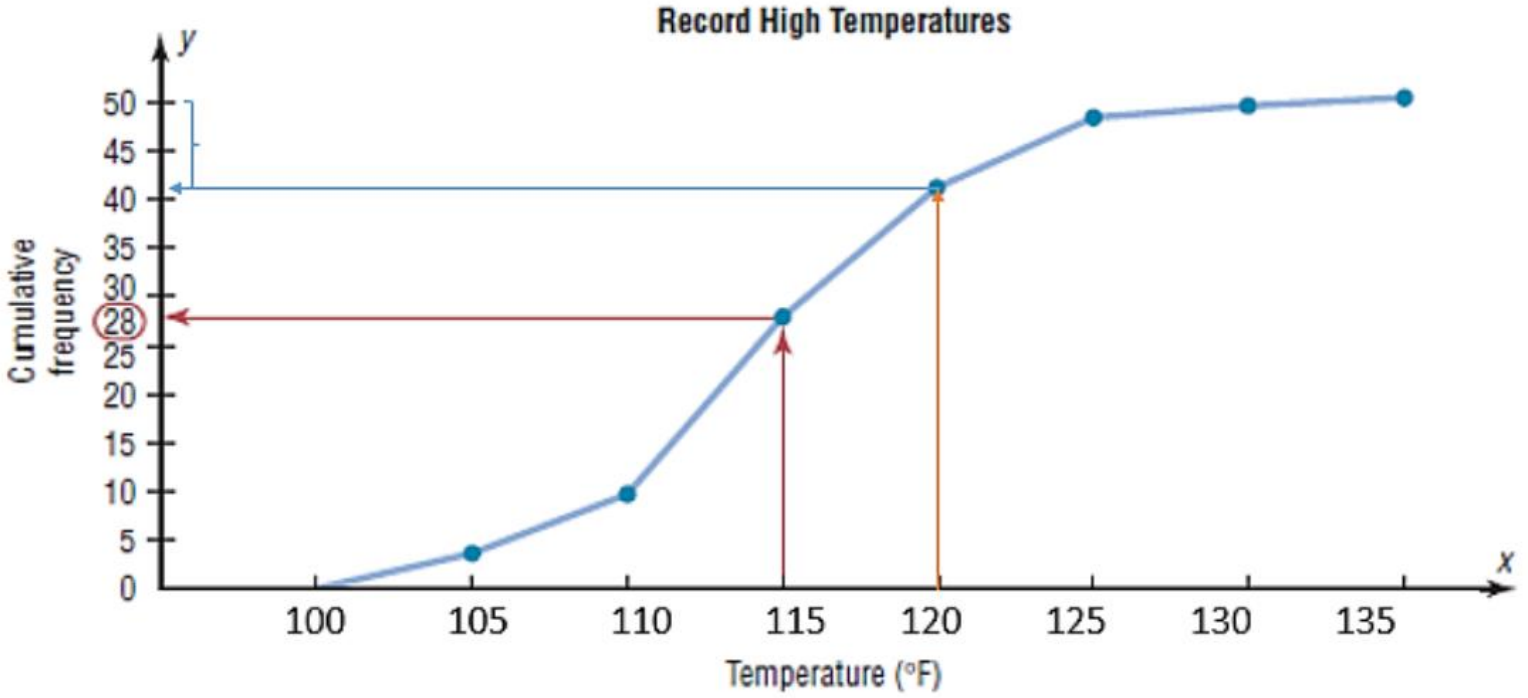
< upper bound	C. frequency
< 105	2
< 110	2+8=10
< 115	2+8+18=28
< 120	41
< 125	48
< 130	49
≤ 135	50=n



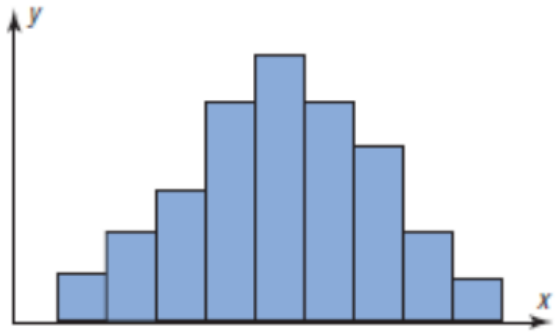
HISTOGRAMS, FREQUENCY POLYGONS, AND OGIVES FOR GROUPED DATA

Use the Ogive curve to find out how many record high temperatures are less than 114.5F?

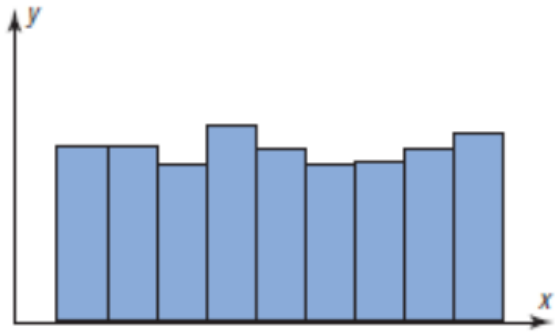
or Find the upper temperature level that obtained by 28 states?



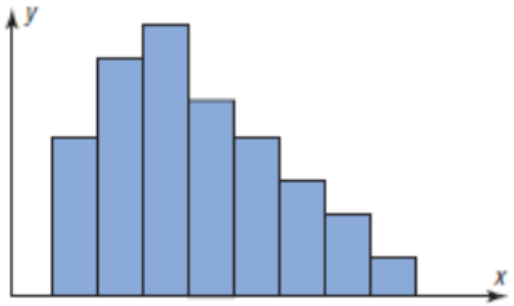
SHAPES OF HISTOGRAM



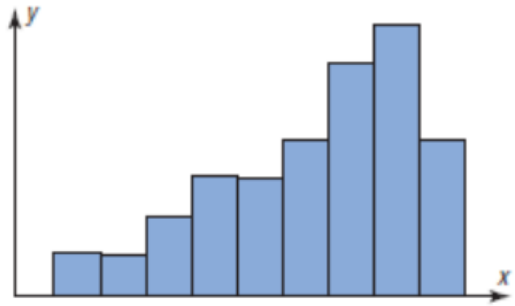
(a) Bell-shaped



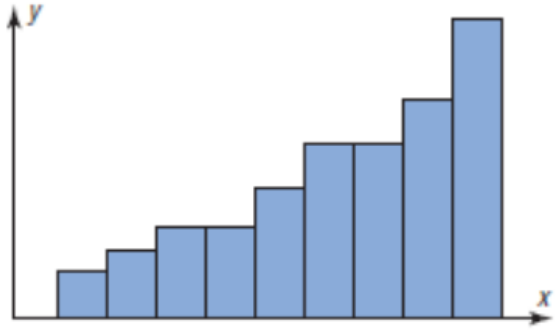
(b) Uniform



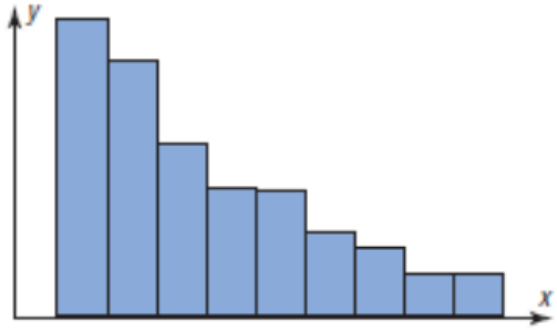
(e) Right-skewed



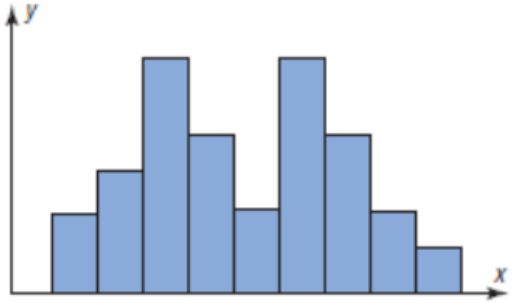
(f) Left-skewed



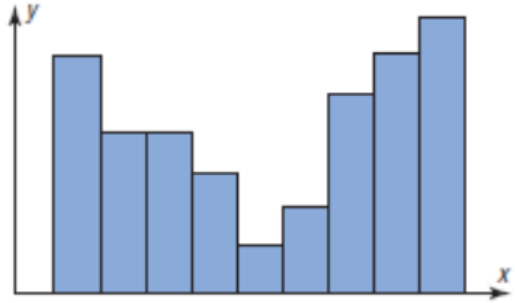
(c) J-shaped



(d) Reverse J-shaped



(g) Bimodal

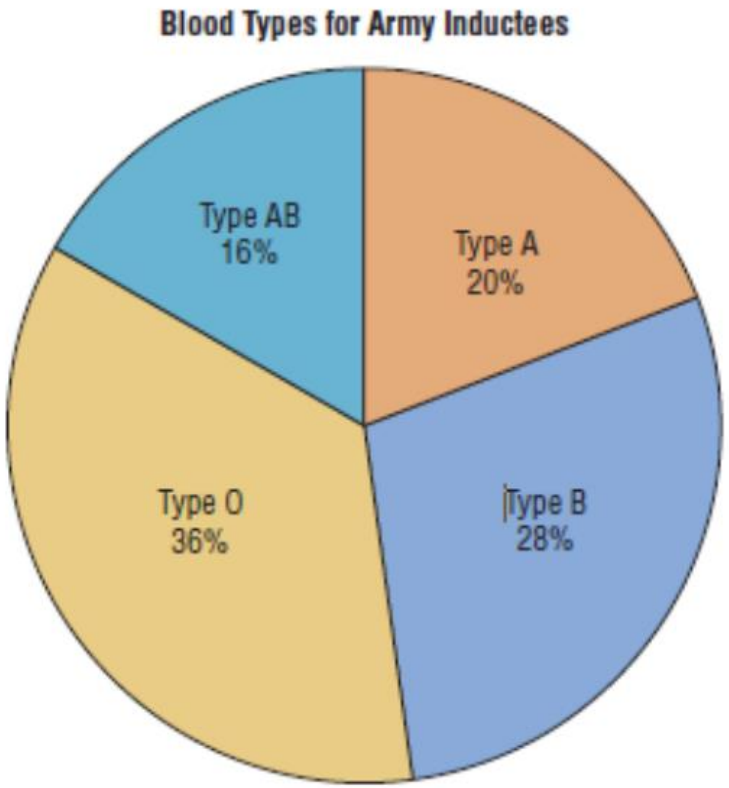


(h) U-shaped

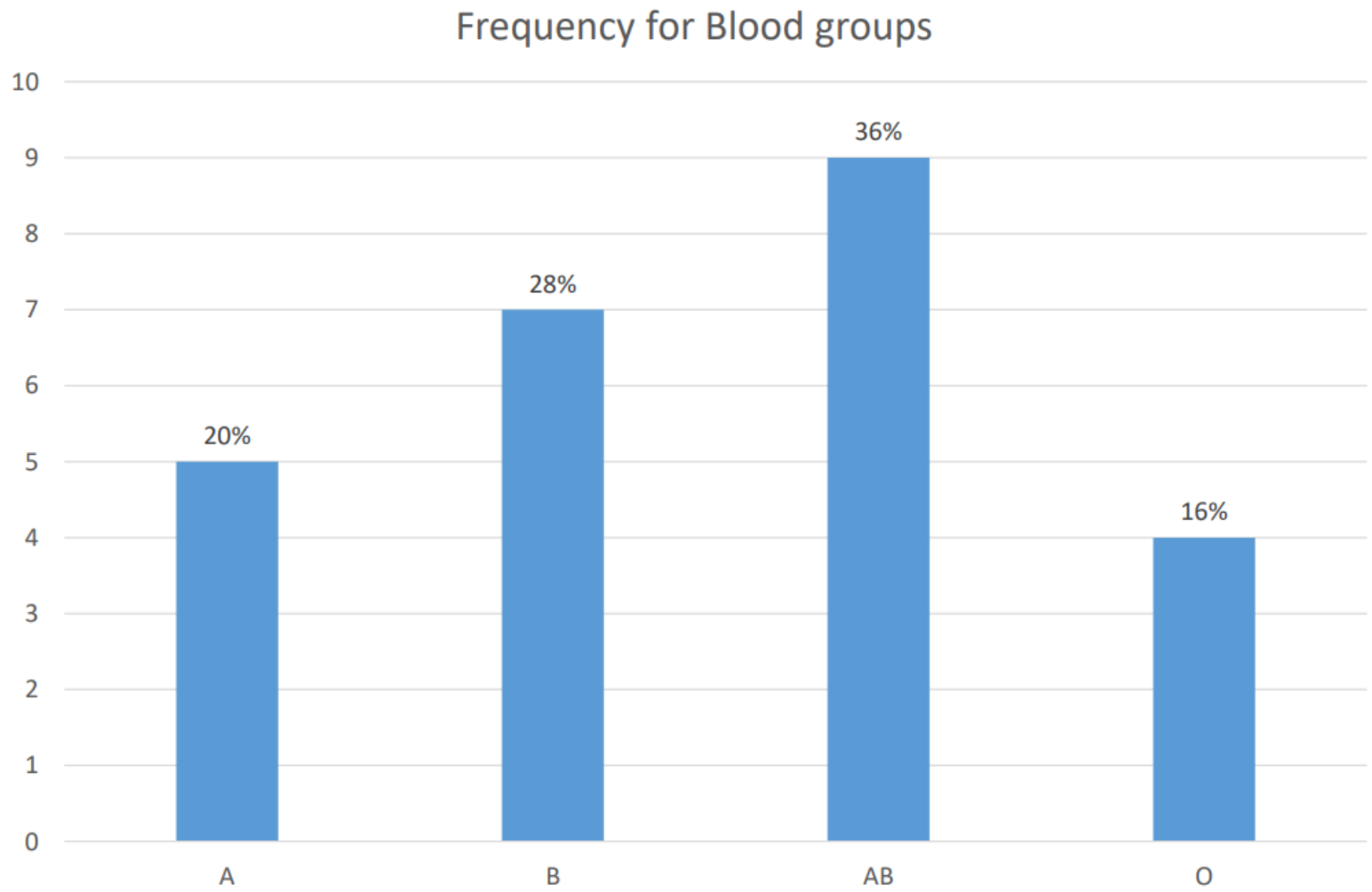
PIE AND BAR CHART FOR PRESENTING THE QUALITATIVE DATA

Construct a pie graph showing the blood types of the army inductees described in Example 2–1. The frequency distribution is repeated here.

Class	Frequency	Percent	Degree=$\frac{f}{n} * 360$
A	5	20%	$\frac{5}{25} * 360 = 72$
B	7	28%	$\frac{7}{25} * 360 = 100.8$
O	9	36%	$\frac{9}{25} * 360 = 129.6$
AB	4	16%	$\frac{4}{25} * 360 = 57.6$
	+ 25	100%	



PIE AND BAR CHART FOR PRESENTING THE QUALITATIVE DATA



STEM AND LEAF PLOT

At an outpatient testing center, the number of cardiograms performed each day for 20 days is shown. Construct a stem and leaf plot for the data.

25	31	20	32	13
14	43	02	57	23
36	32	33	32	44
32	52	44	51	45

Leading digit (stem)	Trailing digit (leaf)
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7