Graphical presentations
Lecture 3

Lecture objectives

- Represent data in frequency distributions using bar graphs, pie graphs, and pareto Charts.
- Represent data in frequency distributions graphically using histograms, frequency polygons, and ogives.
- Draw and interpret a stem and leaf plot, and time series graph

The purpose of graphs in statistics is to convey the data to the viewers in pictorial form. Statistical graphs can be used to describe the data set or to analyze it. Graphs are also useful in getting the audience's attention in a publication or a speaking presentation. They can be used to discuss an issue, reinforce a critical point, or summarize a data set. They can also be used to discover a trend or pattern in a situation over a period of time.

1 Bar Graphs

A bar Graph represents the data by using vertical or horizontal bars whose heights or length represent the frequency of the data.
bar graphs can be used to represent the data when the data is qualitative or categorical.

Example 1 *The table shows the average money spent by first-year college students in a month . Draw a bar graph for the data.*

Electronic	728
Daily expenses	344
Clothing	141
Shoes	72

Solution

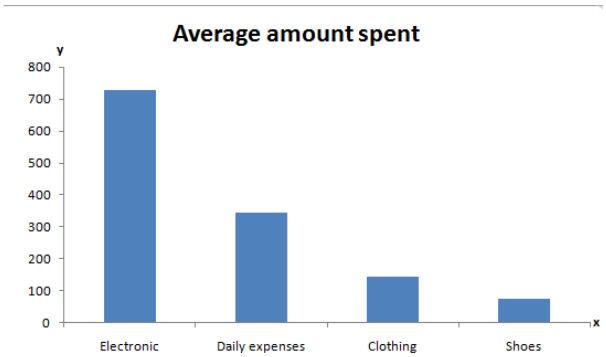
First Draw x and y axes, and Represent the frequency on the y axis

Second Using the frequencies as the heights, draw vertical bars for each class.

The graphs show that first-year college students spend the most on electronic equipment including computers, mobiles, internet etc...

2 Pie graph

A pie graph is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.



The purpose of the pie graph is to show the relationship of the parts to the whole by visually comparing the sizes of the sections, the variable is nominal or ordinal.

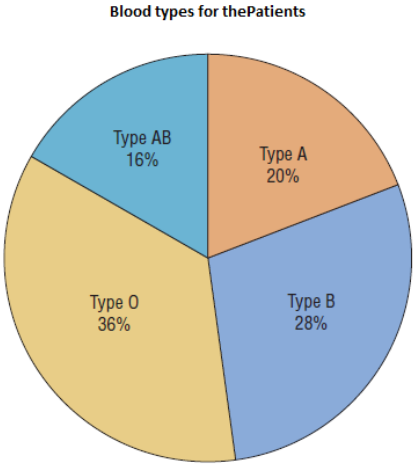
Example 2 *Construct a pie graph for the blood tests results in lecture2.*

solution

First Find the number of degrees for each class.
using the equation
$Degrees = \frac{frequency}{Total} * 360$
$A = \frac{5}{25} * 360 = 72$
$B = \frac{7}{25} * 360 = 100.8$
$O = \frac{9}{25} * 360 = 129.6$
$AB = \frac{4}{25} * 360 = 57.6$

Second Find the percentages using $p = \frac{f}{tot}$

Third Using a protractor, graph each section and write its name and corresponding percentage,



3 Pareto chart

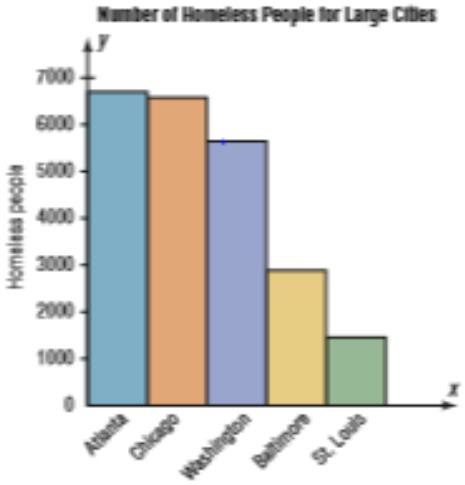
A Pareto chart is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

Example 3 *The data shown here consist of the number of homeless people for a sample of selected cities. Construct and analyze a Pareto chart for the data.*

Atlanta	6832
Baltimore	2904
Chicago	6680
St. Louis	1485
Washington	5518

Solution

First	Arrange the data from the largest to smallest according to frequency
Second	Draw x and y axes
Third	Draw the bars corresponding to the frequencies.



4 Historgram

is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

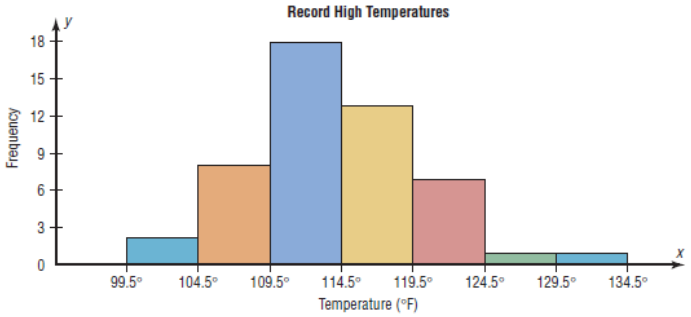
Example 4 *Construct a histogram to represent the data shown for the record high temperatures for each of the 50 cities.*

Class boundaries	Frequency
99.5−−104.5	2
104.5−−109.5	8
109.5−−114.5	18
114.5−−119.5	13
119.5−−124.5	7
124.5−−129.5	1
129.5−−134.5	1
	Total=50

Solution

First Draw x and y axes, and Represent the frequency on the y axis and the class boundaries on the x axis.

Second Using the frequencies as the heights, draw vertical bars for each class.



As the histogram shows, the class with the greatest number of data values (18) is 109.5 − −114.5, followed by 13 for 114.5 − −119.5. The graph also has one peak with the data clustering around it.

5 Frequency Polygon

is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.

Example 5 *using the previous data construct a frequency polygon.*

solution

First Find the midpoints of each class

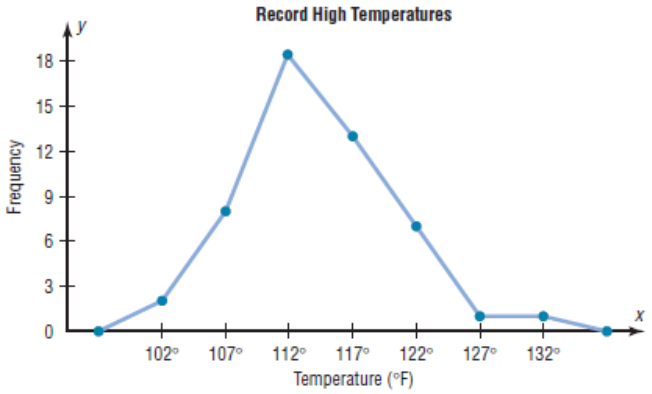
$$x_m = \frac{lowerboundary + uperboundary}{2}$$

Class boundaries	Mid points	Frequency
99.5−−104.5	102	2
104.5−−109.5	107	8
109.5−−114.5	112	18
114.5−−119.5	117	13
119.5−−124.5	122	7
124.5−−129.5	127	1
129.5−−134.5	132	1
		Total=50

Second Draw the x and y axes. Label the x axis with the midpoint of each class, and then use a suitable scale on the y axis for the frequencies

Third Using the midpoints for the x values and the frequencies as the y values, plot the points.

Fourth Connect adjacent points with line segments. Draw a line back to the x axis at the beginning and end of the graph at the same distance that the previous and next midpoints would be located.



The frequency polygon and the histogram are two different ways to represent the same data set. The choice of which one to use is left to the discretion of the researcher

6 Ogive

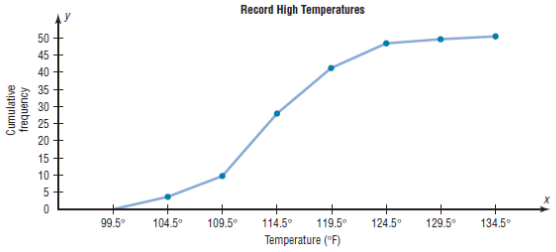
The ogive is a graph that represents the cumulative frequencies for the classes in a frequency distribution.also called *cumulative frequency graph.*and it's used to visually represent how many values are below a certain upper class boundary.

Example 6 *Construct an ogive for the frequency distribution described in the previous example.*

solution

First Find the cumulative frequency for each class.
Second Draw the x and y axes. Label the x axis with the class boundaries.
Use an appropriate scale for the y axis to represent the cumulative frequencies
Third Plot the cumulative frequency at each upper class boundary
Fourth Starting with the first upper class boundary, 104.5,
connect adjacent points with line segments, Then extend the graph to the first lower class boundary, 99.5, on the x axis.

Class boundaries	Cumulative frequency
less than 99.5	0
less than 104.5	2
less than 109.5	10
less than 114.5	28
less than 119.5	41
less than 124.5	48
less than 129.5	49
less than 134.5	50



7 Relative Frequency Graphs

The past three graphs constructed by using frequencies in terms of the raw data. These distributions can be converted to distributions using proportions instead of raw data as frequencies. These types of graphs are called *relative frequency graphs.*and are used when the proportion of data values that fall into a given class is more important than the actual number of data values that fall into that class.

Example 7 *Construct a histogram, frequency polygon and ogive using relative frequencies for the distribution.*

First Convert each frequency to a proportion or relative frequency by dividing the frequency for each class by the total number of observations.
$relative frequency = \frac{frequency}{Total}$
Second Find the cumulative relative frequencies,by adding the relative frequency in each class to the total relative frequency of the preceding class.
Third Plot the cumulative frequency at each upper class boundary
Fourth Draw each graph as stated before . The scale on the y axis uses proportions.

Class boundaries	Midpoints	frequency	Relative frequency	cumulative relative frequency
99.5–−104.5	102	2	0.05	0.05
104.5–−109.5	107	8	0.10	0.15
109.5–−114.5	112	18	0.15	0.30
114.5–−119.5	117	13	0.25	0.55
119.5–−124.5	122	7	0.20	0.75
124.5–−129.5	127	1	0.15	0.90
129.5–−134.5	132	1	0.1	1

8 Stem and leaf plot

A stem and leaf plot is a data plot that uses part of the data value as the stem and part of the data value as the leaf to form groups or classes

Example 8 *At a clinic the number of blood tests 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*

Solution

First	Arrange the data in order
Second	Separate the data according to the first digit
Third	Plot the cumulative frequency at each upper class boundary
Fourth	A display can be made by using the leading digit as the stem and the trailing digit as the leaf

Leading digits (Stem)	Trailing digits(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

the distribution peaks in the center and that there are no gaps in the data. For 7 of the 20 days, the number of blood tests was between 31 and 36. The plot also shows that the testing center treated from a minimum of 2 patients to a maximum of 57 patients in any one day

9 Time series graph

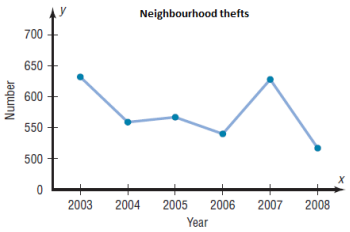
Time series graph represents data that occur over a specific period of time.

Example 9 *The number of thefts that occurred in a certain neighbourhood from 2003 to 2008 as shown below, draw and analyze time series graph for the data.*

Year	*Number*
03	*632*
04	*599*
05	*567*
06	*540*
07	*628*
08	*517*

Solution

First	Draw and label the x and y axes. Label the x axis for the years and the y axis for the numbers of thefts
Second	Plot each point according to the table.
Third	Draw line segments connecting adjacent points



There was a slight decrease in the years 04, 05 and 06, compared to 03 and again an increase in 07. The largest decrease occurred in 08. Two or more data sets can be compared on the same graph called a compound time series graph.