Discriptive Statistics

Dr. Siddiqua Mazhar

Mid Michigan College

Sep 2nd, 2021

Grouped vs. Upgrouped Data

Ungrouped data which is also known as raw data is data that has not been placed in any group or category after collection. Whereas, a grouped data is data that has been organised into groups with certain characteristic. Most often grouped data we organised into a frequency distribution.

Raw Data

Data collected in the original form is call raw data.

Example: Blood Data Raw type

A	B	O	O	O	AB	O
O	O	A	O	B	O	O
B	O	O	O	A	O	O
O	AB	O	B	O	O	O

Frequency Table

Frequency Table partitions data into classes and intervals and shows how many data values are in each class.
Frequency table of Blood Group Data:

Blood	Frequency
A	3
B	4
AB	2
O	19

Frequency Distribution of Grouped Data

A list of values with corrsponding frequencies

Class width
Difference between two lower class limits.

Lower Class Limit
Smallest Value belonging to a class

Upper Class Limit
Largest value in a class.

Procedure

- Determine number of classes
- Class width i.e. $\frac{Max\_Value - Min\_Value}{Num\ of\ Classes}$
- Start with smallest value
- Create classes using class width

Example:

A task force to encourage car pooling did a study of one-way commuting distance of workers in the downtown Lansing area. A random sample of 60 workers was taken. The commuting distance of the workers are given below.

13	47	10	3	16	20	17	40	4	2
7	25	8	21	19	15	3	17	14	6
12	45	1	8	4	16	11	18	23	12
6	2	14	13	7	15	46	12	9	18
34	13	41	28	36	17	24	27	29	9
14	26	10	24	37	31	8	16	12	16

Figure: One way Commuting Distances in miles for 60 workers in Downtown Lansing

Distance	Frequency
1 - 8	14
9 - 16	21
17 - 24	11
25 - 32	6
33 - 40	4
41 - 48	4

Steps to follow:

Class Width: $\frac{Largest\ Data\ Value\ -\ Smallest\ data\ value}{of\ classes}$

Class Width $= \frac{47-1}{6} = 7.7$ (increase to 8)

Lower Class Limit: 1, 9, 17, 25, 33, 41

Upper Class Limit: 8, 16, 24, 32, 40, 48

Class Mid Point $= \frac{Lower\ Class\ Limit\ +\ Upper\ Class\ Limit}{2}$

Class Boundaries: Used to seperate classes without gaps.

such as in this example,

Class Mid Points: 9+8/2, 17+16/2, ...

Class Boundaries: $\frac{8+9}{2}, \frac{16+17}{2}, \frac{24+25}{2}, \frac{32+33}{2}, \frac{40+41}{2}$

i.e 8.5, 16.5, 24.5, 32.5, 40.5,

Relative Frequency Distribution

Relative frequency of a particular class, divide the class frequency f by the total of all frequncies n (Sample Size).

$$Relative\ Frequency = \frac{Class\ f}{\sum f = n}$$

Cumulative Frequency Distribution

The Cumulative frequency for a class is the sum of the frequencies for that class and all previous classes.

Distance	Frequency	Rel. Freq. f/n
1 - 8	14	$14/60 \approx 0.23$
9 - 16	21	$21/60 \approx 0.35$
17 - 24	11	$11/60 \approx 0.18$
25 - 32	6	$6/60 \approx 0.10$
33 - 40	4	$4/60 \approx 0.07$
41 - 48	4	$4/60 \approx 0.07$

Frequency Distribution. You Create.
An irate customer called JC Penny 40 times during the last two weeks to see why his order had not arrived. Each time he called, her recorded the length of time he was put "on hold" before being allowed to talk to a customer service representative.
The data values are in the next page
Create a Frequency Distribution for five numer of classes.

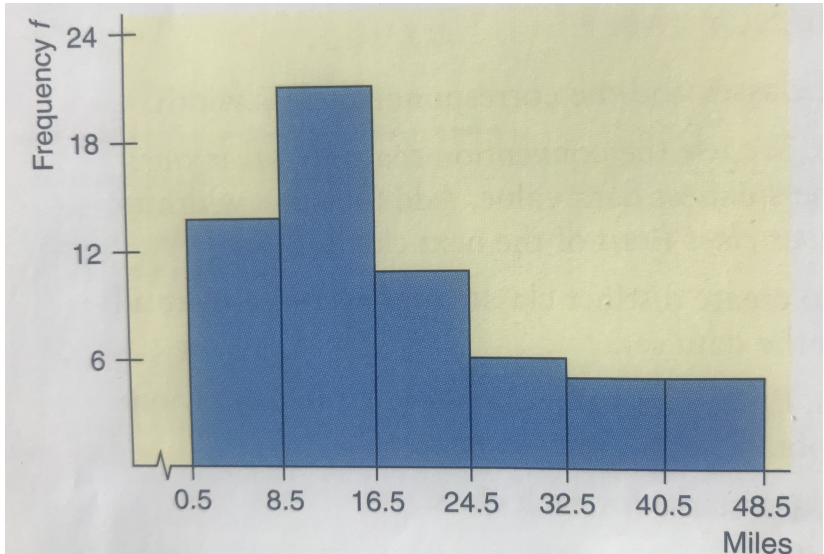| 1 | 5 | 5 | 6 | 7 | 4 | 8 | 5 | 6 | 5 |
5	6	7	6	6	5	8	9	9	10
7	8	11	2	4	6	5	12	13	6
3	7	8	8	9	9	10	9	8	9

Graphs of Data

- Create Visual Representation of Data
- Know and recognise all graphs

Histograms

Histogram is a graph that displays the data by using contigous vertical bars of various heights to represent the frequencies of classes.

Histograms

Histograms

Create Histogram with the time on hold data frequncy table.

Line Graph

A **Line Graph** will plot frequency against a changing value, such as time. It is used for Ungrouped Data.
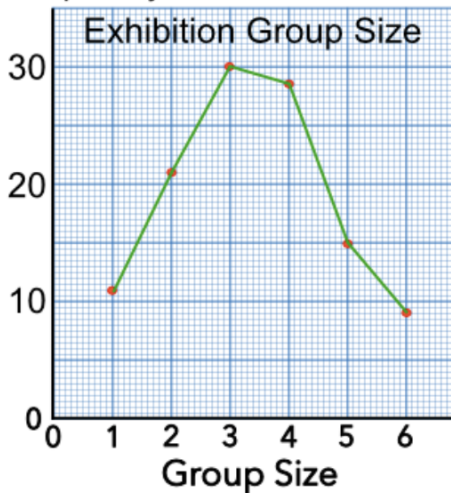
- Points should be plotted using the values provided.
- Only points that are plotted are joined.

Example

In this survey, the number of people in each group arriving at an exhibition:

Group size	Frequency
1	12
2	22
3	30
4	27
5	15
6	8

Exercise: Draw a line graph with the blood group data.

Blood Group	Frequency
A	3
B	4
AB	2
O	19

Frequency Polygon

A **Frequency Polygon** is a line graph that plots the frequency of *grouped data*.
Note: The points plotted are the *middle value* of the class interval and the frequency.

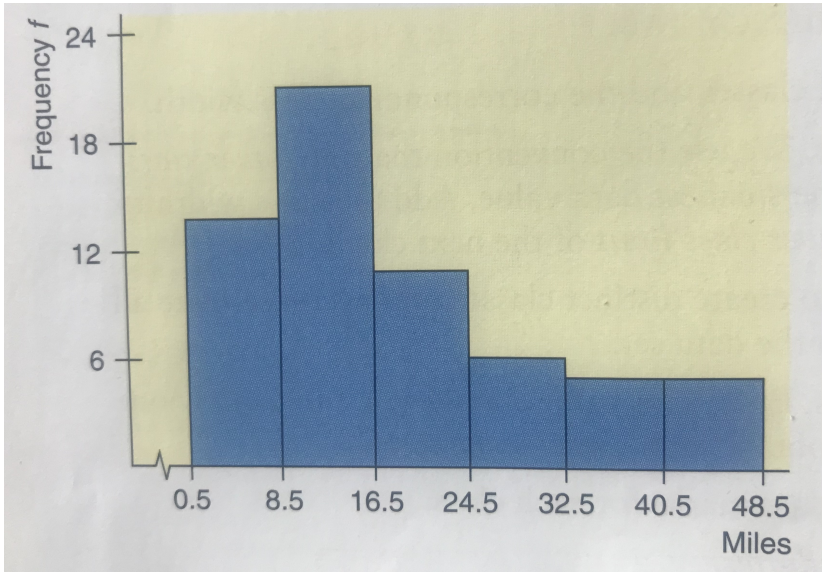. **Frequency Polygon of Distance values covered in miles**
Step1 : Find the mid points of each interval.
Mid Point $(a - b) = \frac{a+b}{2}$

Step 2: The input table for the creation of the frequency polygon
is summarized below:

Distance	Frequency	Mid Point
1 - 8	14	4.5
9 - 16	21	12.5
17 - 24	11	20.5
25 - 32	6	28.5
33 - 40	4	36.5
41 - 48	4	44.5

. Note: We will be creating it on top of Histogram here.

.

Exercise: Construct a frequency polygon of a test undertaken by
some math students. shown in the table below.

| Score	Frequency
0-10	1
11 -20	3
21 - 30	7
31 - 40	15
41 - 50	28
51 - 60	32
61-70	12
71-80	2

Bar Chart

A Bar Graph (also called Bar Chart) is a graphical display most often for categorical data using bars of different heights.
Note:

- Bars can be vertical or horizontal.
- The lengths of the bars represents values of the variable being displayed, the frequency of occurance, or the percentage of occurance.
- The graph is well annotated with title, labels for each bar, and vertical scale or actual value for the length of each bar.

Bar chart to display how many different sports balls are in the gym closet. The data is diplayed in the following chart: :

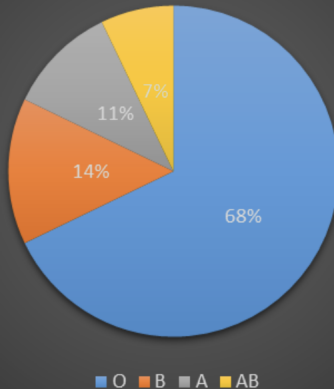| Ball Type	Amount
Soccer	8
Foorball	5
Tennis	3
Valleyball	10
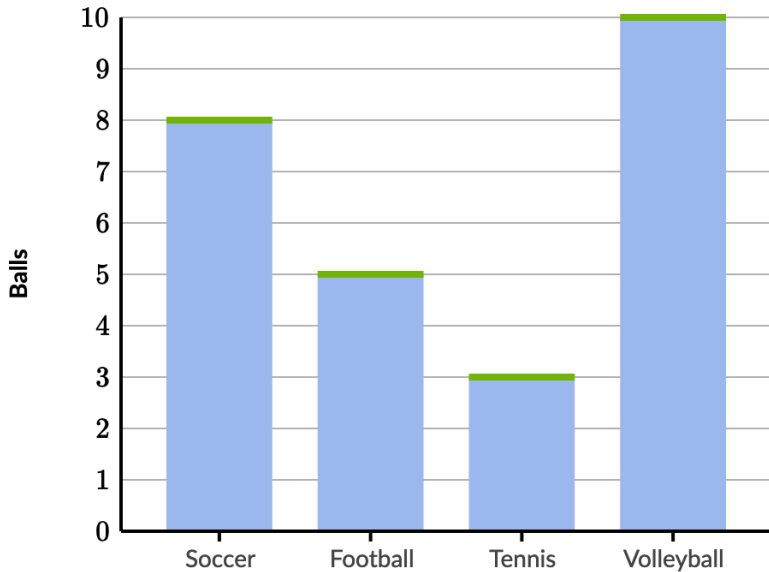
Pie Chart

Pie Graph or Pie chart is circular representation of data.

- Only makes sense when speaking about parts of a whole.
- Categories cannot overlap.
- Pieces of pie add up to 100 percent

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.

Amount of oil changes per day at local oil station.

20	27	18	14	27	31	34
22	23	12	15	29	36	30

Stem and Leaf Plot

Stem	Leaf
1	2, 4, 5, 8
2	0, 2, 3, 7, 7, 9
3	0, 1, 4, 6