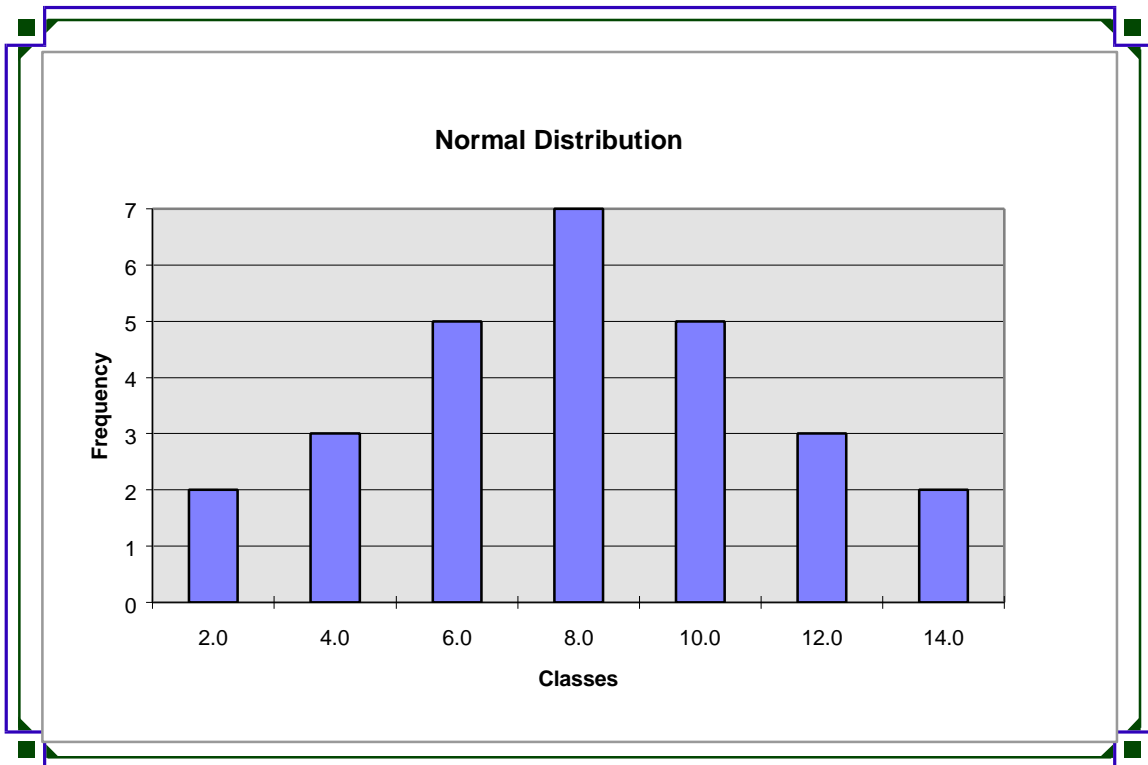


HISTOGRAM (CENTRAL TENDENCY & DISTRIBUTION)

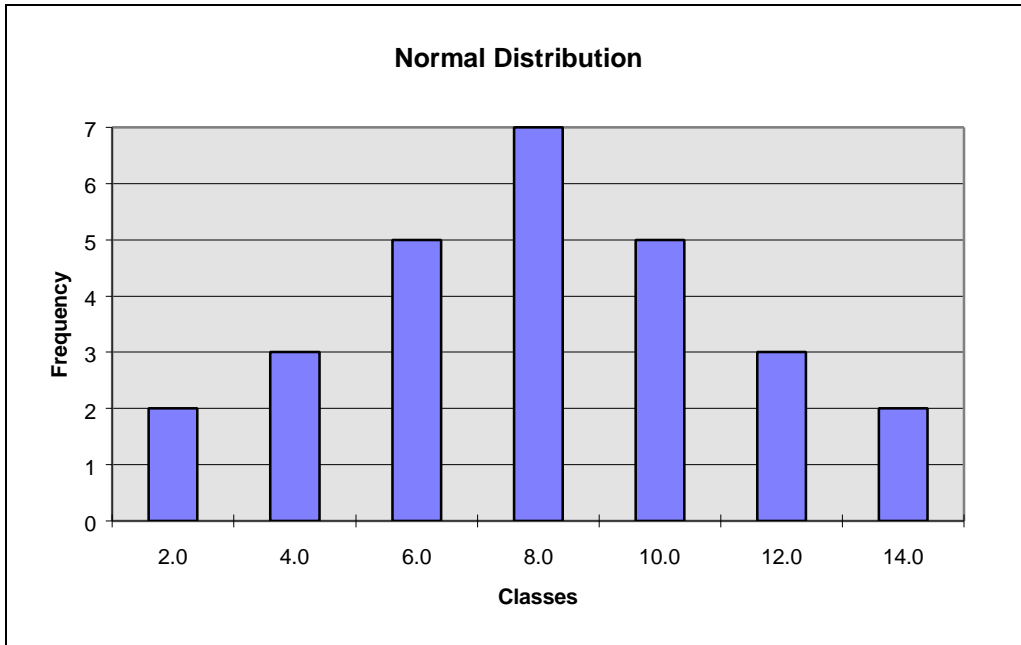


Purpose

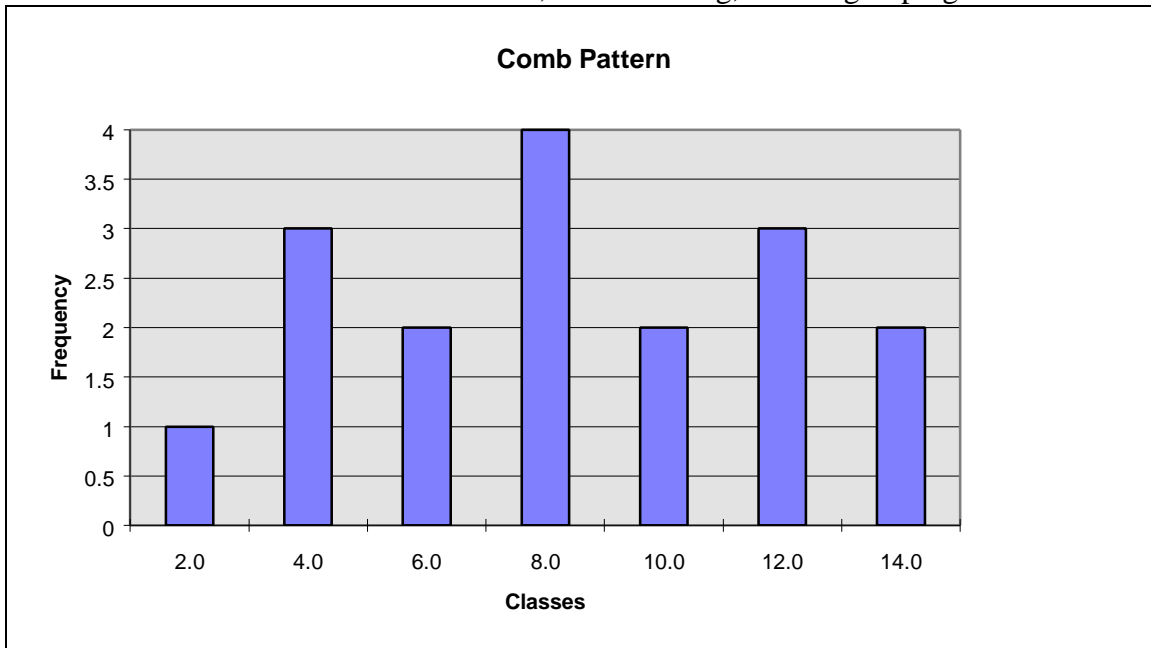
A **histogram** is a chart that shows the spread or distribution of data that can be measured but not easily categorized. Categories used in this data display are merely ranges of data. Data is arrayed on the chart showing the central tendency of data collected.

A **histogram** is used to analyze and evaluate the effectiveness of a proposed process change. The data portrayed may take on several patterns.

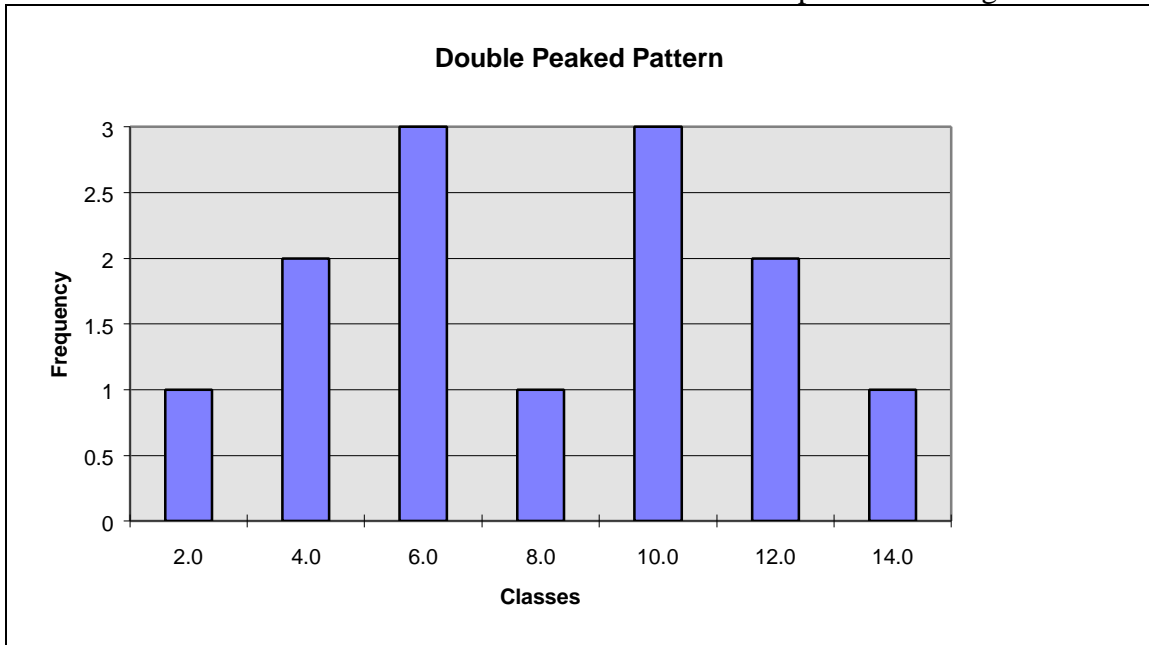
Normal (bell shaped) Pattern- Data has a central tendency with the highest point near the middle of the data.



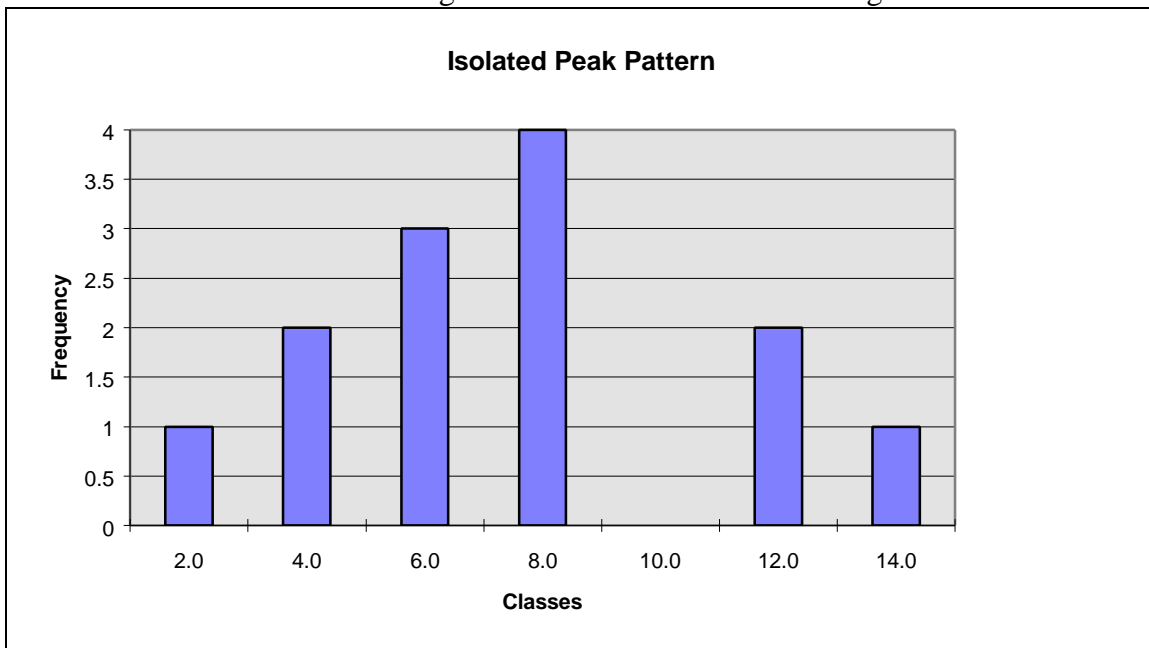
Comb Pattern- Data alternates between several high and low frequencies and may indicate either errors in data measurement, data rounding, or data grouping.



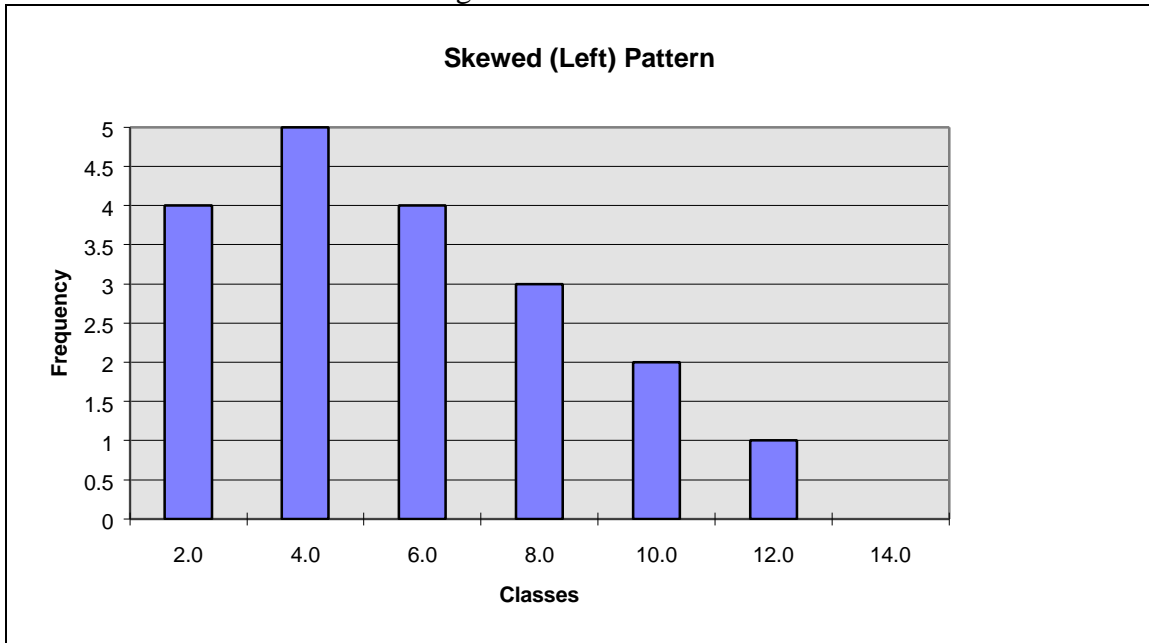
Double Peaked Pattern- Data may or may not be accurate. This could be the result of data collection differences between observers or two different processes being observed.



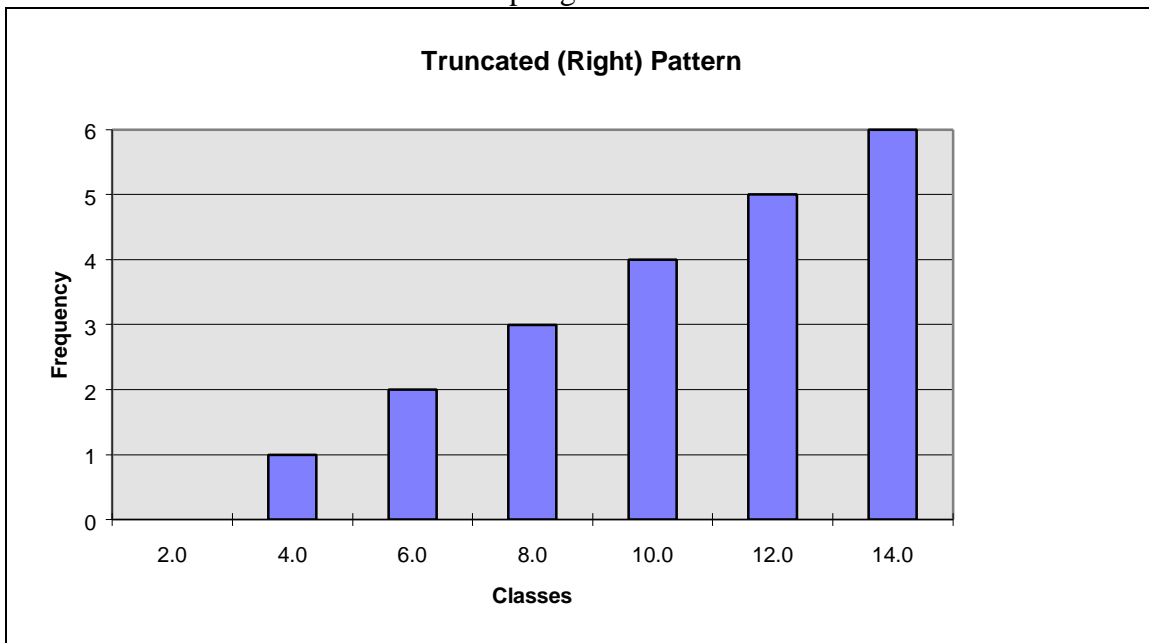
Isolated Peak Pattern- A single high or low spike in the normal data curve. May indicate either a data transfer error or a single instance out of the normal ranges.

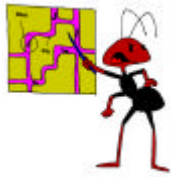


Skewed Pattern- A large peak close to either end of the data distribution indicates that some limit existed that caused a long tail to the data distribution.



Truncated Pattern- Has the data peak at one end of the data display tapering off to a tail at the other end. This may actually be a normal curve but part of the data has been screened out due to a lack of 100% sampling.





Process

1. Determine the purpose of the proposed data collection and decide on a unit of measure that will support the purpose of the data collection.

2. Decide how data will be collected. A common method of data collection is to time transactions. Others include counting occurrences such as defects, responses, arrival times, departure times, waiting times, etc.

3. Collect data using the method agreed upon.

4. Count the number of data points collected for the data set.

5. Determine the range of the entire data set. The range is the distance from the smallest value to the largest value.

6. Determine the number of classes (groups) to be used:

# Observations	# Classes
1-50	5-7
51-100	7-10
101-200	8-12
over 200	10-20



Example

1. The purpose of the proposed data collection is to determine the right amount of time required for a DMV weight enforcement officer to stop and weigh a truck using portable scales. Amount of time required will be measured in whole minutes.

2. Data will be collected using time measurements of weight enforcement officers using portable scales.

3. Data collected:

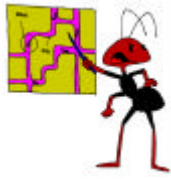
3	10	25	9	3
11	12	7	12	12
11	13	6	3	3
7	7	2	12	6
14	10	7	5	17
13	7			

4. Number of times measurements taken: 27

5. Range of data is:

Smallest value: 2 Largest value: 25
R=Max-Min R=25-2 R=23

6. Based on 27 observations, it appears that 7 classes of data would be effective.



Process

7. Determine the class width (range of data that will be included within each group). A formula that can be used to determine class width is: $W=R/C$ (where W =class width, R =range of data, and C =the number of classes agreed on). At this time it is helpful to round off the class width to one more decimal than the original data collected. Data with no decimal places (whole numbers) would be rounded off to one decimal place.

8. Determine the class boundary. The lower end point or class boundary will be the smallest value of the data set with the upper boundary being the largest value. To the lower end point add the class width, these values will be the class boundary for the first class. Continue the process until all class boundaries are calculated.

9. Construct a frequency table based on the values computed above. Count the number of data points that fall within each class boundary and enter on the frequency table.



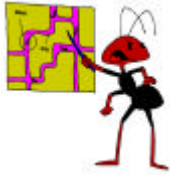
Example

7. Calculate class width:
 $W = R / C$ $W = 23 / 7$
 $W = 3.2857142$ $W = 3.3$

8. Calculate class boundaries:
 $2+3.3=5.3$ $5.3+3.3=8.6$
 $8.6+3.3=11.9$ $11.9+3.3=15.2$
 $15.2+3.3=18.5$ $18.5+3.3=21.8$
 $21.8+3.3=25.1$

9. Frequency Table:

Class	Boundaries	Frequency	Total
1	2.00-5.29	xxxxxxx6	
2	5.30-8.59	xxxxxxx	7
3	8.60-11.89	xxxxx	5
4	11.90-15.19	xxxxxxx	7
5	15.20-18.49	x	1
6	18.50-21.79	0	
7	21.80-25.09	x	1



Process

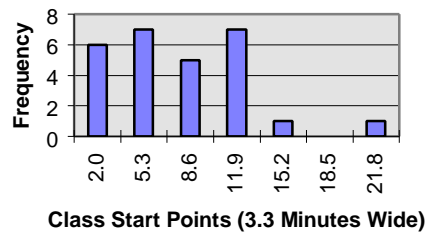
10. Construct the **histogram** based on the frequency table developed above. In most cases, class boundaries are depicted on the histogram. However, if the software or space will not allow this it is acceptable to use start points (as shown) or class mid points. Then it is good to indicate the class width to help prevent misunderstandings.



Example

10. Histogram:

Time Required to Weigh Truck



NOTE: There are several alternative methods of determining boundaries for classes as well as beginning and ending points. All are useable.



Key Points

- A **histogram** may display data with a central tendency. Deviations from the central tendency (or normal bell shaped) may require further analysis. Other factors may be impacting the data distribution. The data sample may be too small, data may have needed to be more specific and not rounded off during collection, or other factors not accounted for.
- Some processes will always deviate from the normal or central tendency distribution.
- The number of classes chosen for the graph determines how much of the pattern will be visible. The larger the number of classes chosen increases the data that will be visible. Too many classes will make spikes in data more visible.