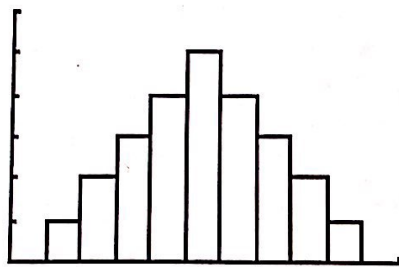


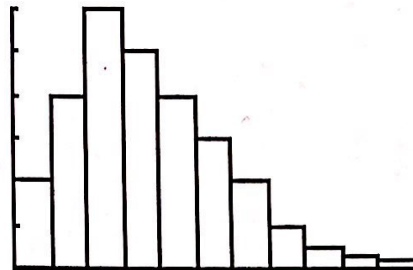
Describing Distributions

The graphs of distributions can be described by their shape, center, spread and outliers.

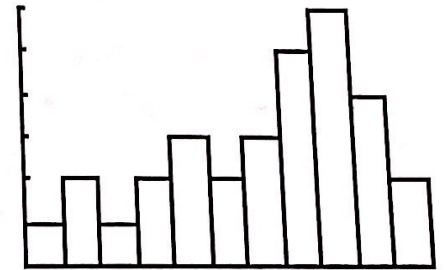
A unimodal distribution has its data in one clump (mode):



Symmetric

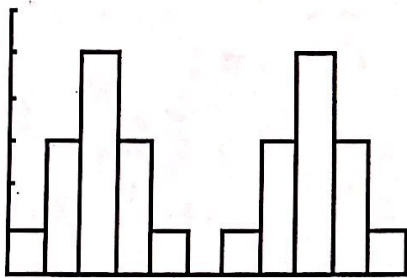


Skewed right
"positively skewed"

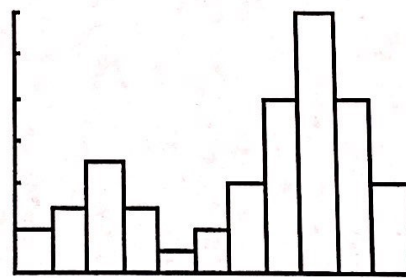


Skewed left
"negatively skewed"

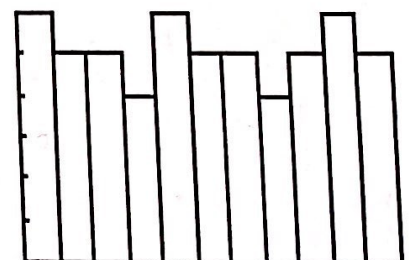
Some other important shapes are bimodal (two clumps of data) and uniform.



Bimodal



Bimodal



Uniform

The center of a symmetric data set is best described by its

mean \bar{x} : $\bar{x} = \frac{\sum x_i}{n}$. For a probability distribution, the mean

is the same as the Expected Value: $E(X) = \sum X_i P_i$

For a non-symmetric distribution, the median is a better measure of center.

↑
Skewed right or left

↑
Middle value

Example: The test scores of 16 randomly chosen Analysis students are:

94, 92, 89, 87, 85, 83, 81, 81, 80, 79, 75, 73, 70, 65, 55, 49

Go to Stat \rightarrow 1:Edit and enter the data in L_1

Window - these are good values

x-min: 40

x-max: 100

y-min: 0

y-max: 4

Use your calculator to make a frequency histogram. What is the shape of the data? Go to Stat Plot ($2^{nd} \rightarrow y=$) \rightarrow turn "on", select type (use the bar graph one), hit graph, adjust window for your list values. **Skewed Left**

Find the mean and median of the data. Which is the better measure of center? Go to Stat \rightarrow Calc \rightarrow 1:1-VarStats. Hit enter all the way down to "calculate" scroll down to find answers...

Mean: $\bar{x} = 77.375$ Median: $Med = 80.5$ \rightarrow Better measure of center since data is skewed

Notice that the median divides the data set into two parts.

The median of the lower half of the numbers is the first quartile (Q_1). The median of the upper half is the third quartile (Q_3).

Example: Find Q_1 and Q_3 for the test scores. (Scroll down the list in calc above)

$$\left. \begin{array}{l} Q_1 = 71.5 \\ Q_3 = 86 \end{array} \right\} \text{ Given on calc}$$

The simplest measure of spread, or variability, for a data set is the range, the difference between the maximum and minimum value. The range is influenced heavily by extreme minimum or maximum values (outliers).

A better measure of spread for a non-symmetric data set is the Inter-Quartile Range (IQR).

$$IQR = Q_3 - Q_1 = 86 - 71.5 = \boxed{14.5}$$

\downarrow \downarrow
 Upper Lower
 Half Half

For a symmetric data set, the sample standard deviation, s is usually used as the measure of spread.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

} Don't use

↓
 $S_x = 12.6$
in calc

We will usually use the calculator STAT CALC 1-Var Stat to obtain the standard deviation.

(S_x - Look for this)

Outliers are another important feature of a distribution. One method of finding outliers is called the 1.5 IQR rule. Any data point that is larger than the Upper Fence or smaller than the Lower Fence is considered an outlier.

Know these formulas

$$UF = Q_3 + 1.5 IQR$$

$$LF = Q_1 - 1.5 IQR$$

Example: Use the test score data to obtain the upper fence and lower fence. Are any of the scores outliers?

$$UF = 86 + 1.5(14.5)$$

$$= \boxed{107.75}$$

$$LF = 71.5 - 1.5(14.5)$$

$$= \boxed{49.5}$$

So "49" is an outlier

Another method of displaying data graphically is called a box plot. A boxplot displays Min, Q1, Median, Q3, Max, and any outliers.

Go to Stat Plot (Hit 2nd → y =) → Go to 1:Plot →

Change type to the "Box & whiskers"

Go to "Graph", hit "trace" to scroll through important values.