

Basic Statistics Reference Sheet

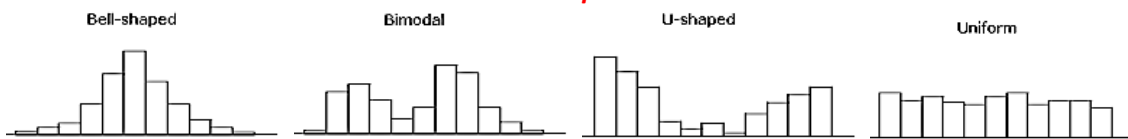
Quantitative Data: describes the individual by assigning a numeric value. Examples – weight, height, IQ. This must take the form of a number and it DOES make sense to find the center.

Distribution Includes: Shape, Center, Spread, and Outliers

Method 1: Using the Median	Method 2: Using the Mean
Center	Center
The Median splits the graph into 2 equal halves (left area = right area). The number such that half the observations are smaller and the other half are larger.	The Mean is the arithmetic average. This point is the balancing point on a graph. This is found by adding the values and dividing by sample size.
Quartiles Q ₁ is the first quartile. It is the median of the bottom half of the observations. 25% of observations are less than this number and 75% are above this number. Q ₂ is the median. 50% of observations are less than this number and 50% are above this number. Q ₃ is the third quartile. It is the median of the top half of the observations. 75% of observations are less than this number and 25% are above this number.	$\bar{x} = \frac{1}{n} \sum x_i$
Spread	Spread
The Range is the difference between the largest and smallest observation (max – min).	The standard deviation is the average distance away from the mean the data values fall.
The Interquartile Range (IQR) is the range of the middle 50% of the data: Q ₃ – Q ₁	Standard Deviation of population $\sigma = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$ Standard Deviation of a sample $s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$

Five Number Summary: Min(lowest value) Q1 Median Q3 Max (highest value)

Shape



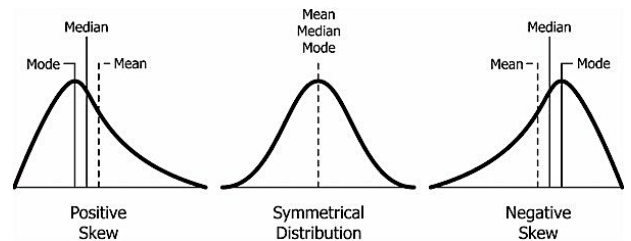
Symmetry: A distribution is symmetric if the right and left sides of the histogram are *approximately* mirror images of each other.

Gaps: holes in the data

Clusters: are the data grouped? Is there a large group that appears unexpected?

Modes: 0, 1 (unimodal), or 2 modes (bimodal)?

Skewness: if the graph is not symmetric, then the graph will be longer on one side.



Right Skew/Positive Skew: tail is on the right

Left Skew/Negative Skew: tail is on the left

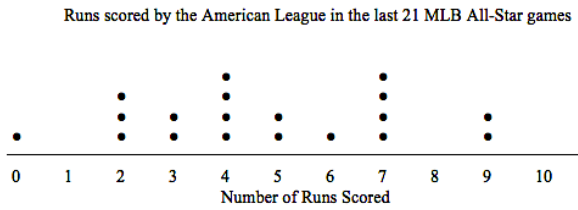
Outliers

Observations that are so far above or below the majority of the data that they do not fit. Here is the process to find the cutoff values.

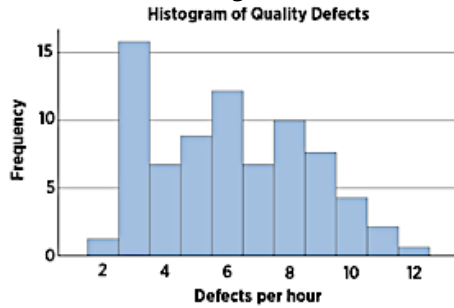
Cutoff Value Formulas: Q₁ – 1.5(IQR) Q₃ + 1.5(IQR)

Quantitative Graphs

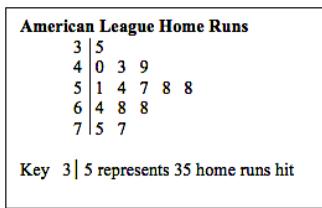
Dotplot



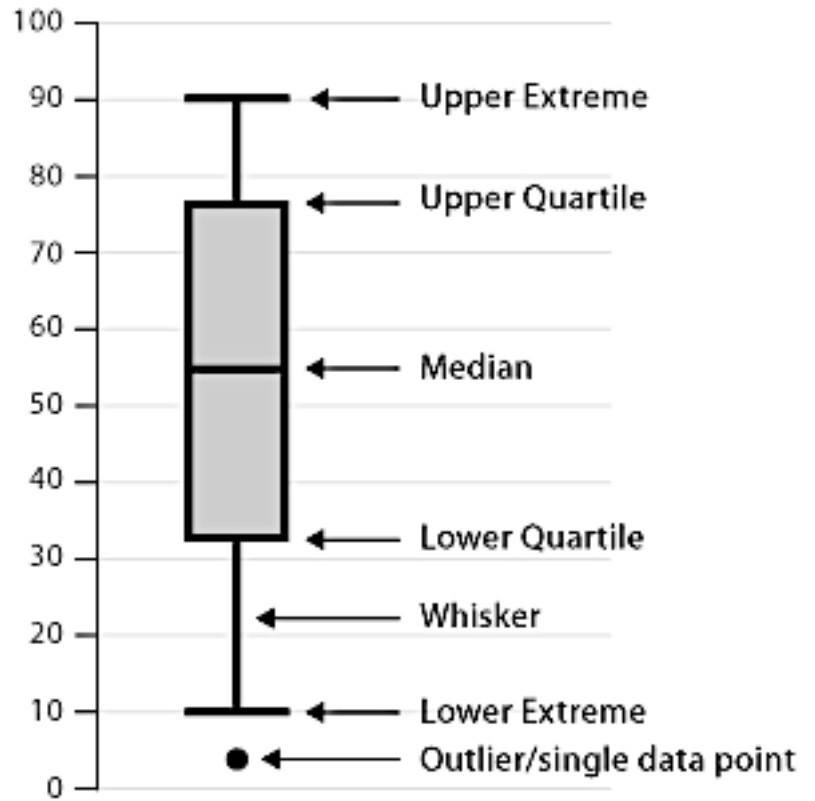
Histogram



Stem Plot



Box Plot



Categorical Data: Describes the individual by placing them in a group or category. Examples – favorite color, race, state in which you were born. This does not typically take the form of a number and it does NOT make sense to find the center. This is sometimes referred to as a qualitative variable.

Frequency: the number of times a value occurs (count them)

Relative Frequency: the percent that the value occurs relative to the others (frequency divided by total)

Categorical Graphs: Bar Chart, Pie Chart

Two Way Table - (also called a contingency table) a useful tool for examining relationships between categorical variables. The entries in the cells of a two-way table can be frequency counts or relative frequencies (just like a one-way table).

Marginal Frequency - Entries in the "Total" row and "Total" column are called marginal frequencies or the marginal distribution.

Joint frequencies – Entries in the body of the table.

Association - We say there is an association between two variables if specific values of one variable tend to occur in common with specific values of the other. A two-way table must be a conditional distribution in order for association to be noticeable. Example: In Football, there is an association between turnovers and losing.

	All Student Data		
	Pass	Fail	Total
Fraker	6 20%	24 80%	30 100%
Sroufe	10 33%	20 67%	30 100%
Total	16 27%	44 73%	60 100%