How can you describe a set of data?

Measures of Center:

one number to summarize data

EXAMPLE:

3,2,1,4,5,8,2,4,5,0,2,4,7,6

	When to use:	How to use:
mean	average; when data is nearly symmetric	add up all # and divide by the # of values
median	middle number; when data is skewed	find the middle number, or average of the middle 2
mode	most frequent; when want to know popular	count what's most popular

Measures of Spread:

tells you how "spread out" your data is

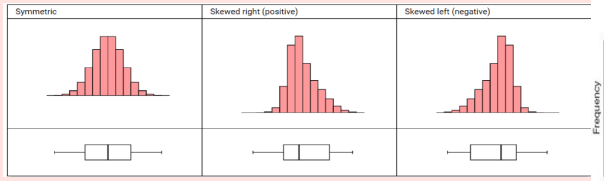
	When to use:	How to use:
Standard Deviation	with mean	fancy formula- (don't need)
Range	with median	max - min
IQR	with median/ to tell if there are outliers	$Q_3 - Q_1$

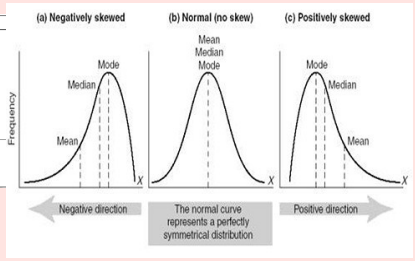
Types of Distribution:

Symmetric

Skewed

Bimodal



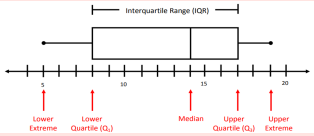


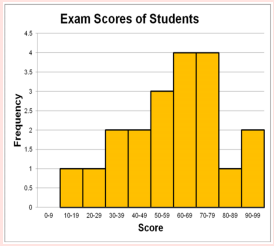
Data Displays:

Dot Plot

Box Plot (&Whisker)

Histogram





5 number summary

min
Q_1 → median of lower half
med
Q_3 → median of upper half
max

To estimate MEAN:

--find the middle # in each group & multiply by frequency

then divide by total frequency

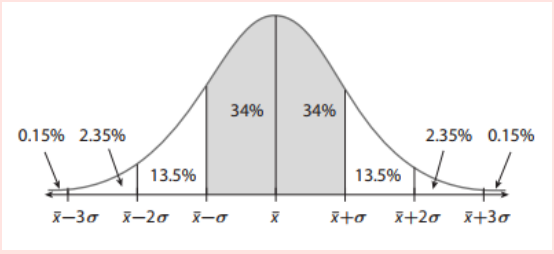
Outlier

a value that is much greater or less than most other values

$x<Q_1-1.5(IQR)$ or $x>Q_3+1.5(IQR)$

How can we use statistics to make conclusions and predictions about a set (or sets) of data?

Normal Distribution



percent of population within intervals

Empirical Rule

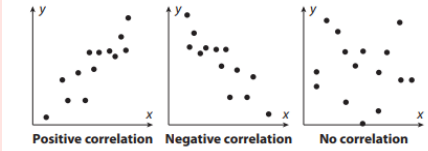
68- 95- 99

1 sd - 2 sd - 3 sd

Probability

The likelihood something will happen (percents)

Scatter Plot



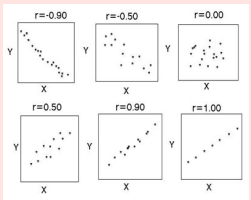
Correlation

How closely cor**related** 2 variables are

Correlation Coefficient

-1 0 1

*the closer to |1| the strong it is, closer to 0, the weaker



Linear Regression/ Line of Best Fit/ Least Squares Line

$y = mx + b$

* create a line based on points that go through the "center" of the data

Residual

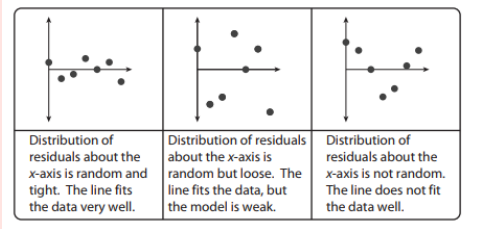
Actual Y - Predicted Y

Residual Plot

x is predicted value

y is residual (how far away)

*If there is a pattern in the residual plot, it is NOT a good fit!



Interpolation VS Extrapolation

INside the data range ← this is better!

OUTSIDE the data range

Causation

Correlation DOES NOT imply causation!