Parameter or statistic?

PARAMETER	POPULATION	ENTIRE SET OR CENSUS
STATISTIC	SAMPLE	SMALL AMOUNT OR SUBSET

Levels of measurement

NOMINAL LEVEL OF MEASUREMENT	CATEGORIES ONLY, THERE IS NO UNIVERSAL ORDER FOR THE CATEGORIES.	MAJORS CAR TYPES
ORDINAL LEVEL OF MEASUREMENT	CATEGORIES CAN BE PUT IN AN ORDER BUT ANY NUMERICAL DIFFERENCES ARE MEANINGLESS.	SMALL, MEDIUM, LARGE
INTERVAL LEVEL OF MEASUREMENT	DIFFERENCES BETWEEN NUMBERS ARE MEANINGFUL BUT THERE IS NO NATURAL ZERO (STARTING POINT).	YEARS, TEMPERATURE
RATIO LEVEL OF MEASUREMENT	THERE IS A NATURAL STARTING POINT (NO NEGATIVE VALUES)	HEIGHT, DISTANCE

Sampling methods:

Type of Sampling	Definition	Example
Simple Random Sampling (SRS)	**Completely random** with no chance of bias	Pulling numbers out of a hat without looking.
Cluster sampling	Divide into groups and sampling **every data point** within a group	Group 1 Group 2 Group 3 ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ Sample all people in group 2 (Sample all people highlighted)
Stratified sampling	Divide into groups and sampling a few data points out of **each group**	Group 1 Group 2 Group 3 ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ ☺☺☺☺ Sample all people highlighted
Systematic Sampling	Every n^{th} data point	☺☺☺☺☺☺☺☺☺☺☺☺ ☺☺☺☺☺☺☺☺☺☺☺☺ ☺☺☺☺☺☺☺☺☺☺☺☺ ☺☺☺☺☺☺☺☺☺☺☺☺ Sample every 8th person (all people highlighted)
Convenience Sampling	Choosing people based on **convenience**	Picking the people who look nice and will answer your question.

Pronunciation and symbol guide

SYMBOL	PRONUNCIATION	MEANING
Σ	Sigma (Uppercase)	Add the set of values attached to Σ
X	X	the variable usually used to represent the individual data values
n	N	the number of values in a sample
$\bar{x}$	X Bar	Is the mean of a set of sample values
μ	Mu	Mean of all values in a population
σ	Sigma (Lowercase)	Standard deviation of a population
σ^2	Population Variance	Standard deviation of a population squared
r	Correlation Coefficient	the linear correlation coefficient for a sample

Frequency distributions:

Frequency distribution	A table of data values and corresponding frequencies.
Frequency	The number of data points that fall within a class, or range of data points.
Class	Levels or categories used to explain the spread of a data list
Lower class limit (LCL)	Smallest whole number boundary of a class
Upper class limit (UCL)	Largest whole number boundary of a class
Class boundary	The space between the gab of each class limit
Class midpoints	The median of each class
Class width	The distance between each consecutive lower class limit

How to make a frequency distribution:
1. Decide how many classes you want to have
2. Find the class width and round the number (most of the time you round this number up)
3. Find the lowest data value in the data set, this will be the lowest class limit
4. Add the class with to the lowest class limit (from step 3). This is the second class' lower limit.
5. Repeat step 4 until you exceed the highest value in the data set.
6. Find the upper class limits, add one minus the class width to each lower limit
7. Fill out the class section of the frequency distribution (the table below).
8. Find the class frequencies (number of values that fall within each class).

Relative frequency distributions:
9. Divide each class frequency by the sum of all frequencies.

Cumulative frequency distributions:
9. Using the upper class limits, at each limit add all frequencies below the upper class limit

Class midpoints: add the lower class limit and the upper class limit and divide

Box plot:
1. Find the 5-number summary
2. Construct a scale with values that include the minimum and maximum values
3. Construct a box extending from the first quartile to the third quartile and draw
    a line in the box at the median value
4. Draw lines extending outward from the box to the minimum and the maximum
    data values.
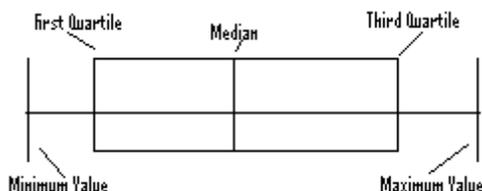
Properties of the correlation coefficient:
1. The value of r is always between –1 and +1
2. The value of r does not change if all values of either variable are converted to a different scale
3. The value of r is not affected by the choice of x or y Interchange all values for x
and y and the value of r will not change
4. r measures the strength of only linear relationships

Test for correlation
First, find the correlation coefficient and refer to table A-6 if the correlation coefficient is more than the .05 or the .01 there is a significant linear correlation.
If the correlation coefficient is significant use the regression equation: y= a + b x   and solve for y
If the correlation coefficient is not significant use the population mean.

Probability:
Unusual events are < 0.05
Compliments $P(A) + P(B) = 1$
RULE $= 0 < P(A) < 1$
$P(A \text{ or } B) = P(A) + P(B) – P(A \text{ and } B)$
$P(A \text{ and } B) = P(A) \cdot P(B \text{ given } A)$
Conditional Probability: $P(B | A) = P(A \text{ and } B) / P(A)$
Independence: $P(A \text{ and } B) = P(A) \cdot P(B)$

Binomial Distribution:
1. Only 2 results for each trial procedure success (p) or failure (q)
2. Fixed # of trials
3. Trials must be independent
4. P(s) must be same for each trial
Written P(s)= p   P(f) = q

BINOMIAL PROBABILITY FORMULA

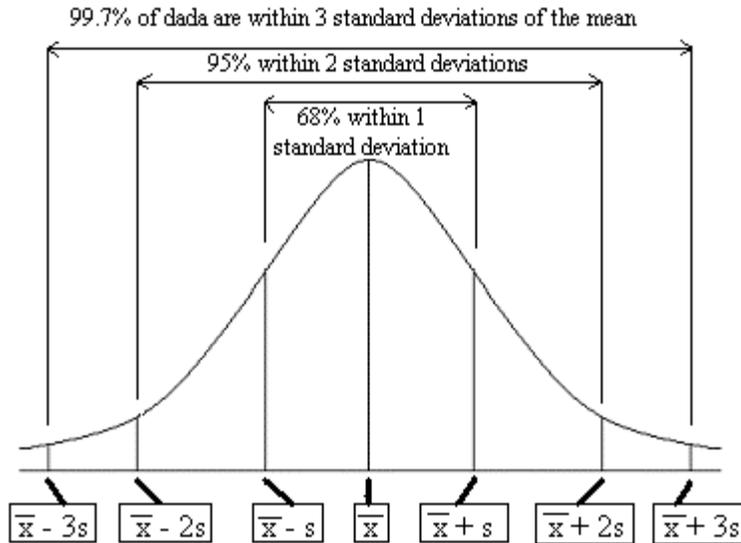$$P(x) = \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot q^{(n-x)}$$

Finding Probabilities when given z-scores:
1. Use z scores table when the mean is 0 and the standard deviation is 1.
2. When finding Z-Scores use the left column and the top row
3. When finding Areas use the values in the body of the table

To convert values to standard scores use the formula below.
$$z = \frac{x - \mu}{\sigma}$$

The Empirical Rule Standard Normal Distribution

99.7% of dada are within 3 standard deviations of the mean

95% within 2 standard deviations

68% within 1 standard deviation

| $\overline{x} - 3s$ | $\overline{x} - 2s$ | $\overline{x} - s$ | $\overline{x}$ | $\overline{x} + s$ | $\overline{x} + 2s$ | $\overline{x} + 3s$ |

Common mistakes when finding values from Known Areas:
1. Don't confuse z scores and areas. Remember, z scores are distances along the horizontal scale but areas are regions under the normal curve.
2. Choose the correct right/left side of the graph.
3. A z score must be negative whenever it is located in the left half of the normal distribution.
4. Areas (or probabilities) are positive or zero values, but they are never negative.