SECTION TWO, ESSAY: Researchers rarely have data on the full population that they are studying. In order to learn about their population, they draw a sample from it in order to infer meaning about the entire population. However, this inference is only permissible and/or credible under certain circumstances. What are these circumstances? What are the mechanisms by which these inferences are possible? What statistical principles and theories underlie these mechanisms? *In detail*, list, describe, define, and explain the process by which statistical inference is made. *(20 points)*

SUPERB! +20

In general researchers are aiming to learn or test something about a population, but due to limitations of resources, access, +/or time it is usually not possible to test everyone in that population. However, if certain principles are adhered to, researchers can make generalizations about a population by testing a subset, or sample of that population.

The ultimate goal is to produce a sample that is representative of the population of interest. This means that traits, patterns, variations w/i the population are reflected in the sample. The way researchers achieve this is by ~~char~~ selecting samples in a way so that every member of the population has an equal chance of being selected. These are known as Equal Probability of Selection Methods (EPSEM) + include simple random sample, stratified random sample, clustered random sample, + systematic random sample. These methods help reduce the chances of bias in the sample. Another concern is ensuring an equal probability of response, ensuring that people who respond aren't fundamentally different.

Once a sample has been obtained, researchers can take data on the variables they are interested in. They can describe the central tendency of those variables using mean, median, mode, etc. or dispersion, using standard deviation, range, etc. However, at this point all that data can do is describe the sample itself. In order to generalize it to the population, inferential statistics must be used.

A sample can be generalized + compared to the population it came from using the sampling distribution. The sampling distribution is a theoretical distribution of all means of all the possible samples from a given population. The Central Limit Theorem states that as long as a sample is large (100 or greater) + was selected through EPSEM standards, the sampling distribution is normally distributed. B/c the normal distribution has certain principles of distribution that are known to us (symmetrical, mean/median = Z=0, standard deviation = 1Z), we can use Z scores to ~~estimate understand~~ estimate the probability of a certain mean/proportion coming from a population, using known values from a Z table. This is important b/c even w/ EPSEM + a large population, we have the chance of sampling error. However, we can estimate ~~manage~~ that error using Z scores + the sampling distribution by setting thresholds of how far a score would need to be from the population mean (level of Z critical), where we could say the difference is real + not due to random chance/sampling error. If a sample mean exceeds that critical Z, we know it is highly likely (but not certain) that it reflects a true difference from the population mean, as there would be a very small chance of selecting a sample w/ that mean from the population using random methods. When this occurs we say a finding is statistically significant, meaning that there is a real difference + we are confident to a certain % (determined by our alpha/Z critical). It is important that while we can manage our risk of error, we can never eliminate it as there will always be a small (but existent) chance of uncertainty).