LO 1. Identify variables as numerical and categorical.

- If the variable is numerical, further classify as continuous or discrete based on whether or not the variable can take on an infinite number of values or only non-negative whole numbers, respectively.

LO 2. Define associated variables as variables that show some relationship with one another. Further categorize this relationship as positive or negative association, when possible.

LO 3. Define variables that are not associated as independent.

  * *Reading: Section 1.1 and 1.2 of Advanced High School Statistics*

  * *Video: Types of variables, Dr. Çetinkaya-Rundel (3:53)*

  * *Test yourself: Give one example of each type of variable you have learned.*

LO 4. Identify the explanatory variable in a pair of variables as the variable suspected of affecting the other. However, note that labeling variables as explanatory and response does not guarantee that the relationship between the two is actually causal, even if there is an association identified between the two variables.

LO 5. Classify a study as observational or experimental, and determine and explain why whether the study's results can be generalized to the population and whether they suggest correlation or causation between the variables studied.

- If random sampling has been employed in data collection, the results should be generalizable to the target population.

- If random assignment has been employed in study design, the results suggest causality.

LO 6. Question confounding variables and sources of bias in a given study.

LO 7. Distinguish between simple random, stratified, and cluster, and multistage sampling, and recognize the benefits and drawbacks of choosing one sampling scheme over another.

- Simple random sampling: Each subject in the population is equally likely to be selected.

- Stratified sampling: First divide the population into homogenous strata (subjects within each stratum are similar, across strata are different), then randomly sample from within each strata.

- Cluster sampling: First divide the population into clusters (subjects within each cluster are non-homogenous, but clusters are similar to each other), then randomly sample a few clusters, and then sample all cases within those clusters.

- Multistage sampling: First divide the population into clusters, then randomly sample a few clusters, and then randomly sample from within each cluster.

LO 8. Identify the four principles of experimental design and recognize their purposes: control any possible confounders, randomize into treatment and control groups, replicate by using a sufficiently large sample or repeating the experiment, and block any variables that might influence the response.

LO 9. Identify if single or double blinding has been used in a study.

LO 10. Identify the number of factors in an experiment, the number of levels for each factor, and the total number of treatments.

* *Reading: Sections 1.3 - 1.5 of Advanced High School Statistics*

* *Videos:*

    - *Random sampling vs. random assignment, Dr. Çetinkaya-Rundel (3:18)*

    - *Quick video on correlation vs. causation, YouTube (2:19)*

    - *Slower video on correlation vs. causation, KhanAcademy (10:45)*

* *Article: How Anecdotal Evidence Can Undermine Scientific Results, Scientific American, 2008*

* *Test yourself:*

    1. *Describe when a study's results can be generalized to the population at large and when causation can be inferred.*

    2. *Explain why random sampling allows for generalizability of results.*

    3. *Explain why random assignment allows for making causal conclusions.*

    4. *Describe a situation where cluster sampling is more efficient than simple random or stratified sampling.*

    5. *Explain how blinding can help eliminate the placebo effect and other biases.*