**Week 2 – Lesson Learning Objectives**

LO1 Identify variables as numerical or categorical

* Numerical
  + Continuous – can take an infinite number of values within a given range
    - Eg. Height
  + Discrete – can take only non-negative whole numbers
    - Eg. Number of cars a household owns
    - Eg. Count data
  + Take on numerical values
  + Sensible to add, subtract, take averages, perform numerical operations
* Categorical
  + Take on a limited number of distinct categories
  + Not sensible to perform arithmetic operations
  + Eg. Male = 0 Female = 1 is often used
  + Further check if ordinal based on whether or not the levels have a natural ordering
    - Levels have an inherent ordering
    - Eg. Survey that has options from very satisfied to very unsatisfied
      * Inherent ordering i.e. satisfied is better than unsatisfied

LO2 Define associated variables as variables that show some relationship with one another. Further categorize this relationship as positive or negative association

* Correlation does not imply causation

LO3 Define variables that are not associated as independent

LO4 Identify the explanatory variable in a pair of variables as the variable suspected of affecting the other, however note that labelling variables as explanatory and response does not guarantee that the relationship between the two is actually causal, even if there is association identified between the two variables

LO5 Classify a study as observational or experimental, and determine and explain whether the study’s results can be generalized to the population and whether the results suggest correlation or causation between the quantities studied

* If random sampling has been employed in data collection, the results should be generalized to the target population
* If random assignment has been employed in study design, the results suggest causality
* Observational studies generally only allow us to make correlation based conclusions
* Experimental studies generally allow us to make causation based conclusions
* The difference is that an experimental study uses random assignment whereas an observational study does not

LO6 Question confounding variables and sources of bias in a given study

* Some sources of sampling bias:
  + Convenience bias: individuals who are easily accessible are more likely to be included in the sample
  + Non-response: if only a random fraction of the randomly sampled people responds to a survey such that the sample is no longer representative of the population
    - People who choose to respond are not representative of the population
    - An initial random sample, but not everyone in the random sample is reached
  + Voluntary response: occurs when the sample consists of people who volunteer to respond because they have strong opinions on the issue
    - No initial random sample

LO7 Distinguish between simple random, stratified and cluster sampling, and recognize the drawbacks and benefits of choosing one sampling scheme over another

* Simple random: each subject in the population is equally likely to be selected
* Stratified: First divide the population into homogenous strata then randomly sample a few clusters, and then randomly sample from within each stratum
* Cluster: 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 randomly sample from within each cluster

LO8 Identify the four principles of experimental design and recognize their purposes

* Control any possible confounders
* Randomize into treatment and control groups
* Replicate by using sufficiently large sample or repeating the experiment
* Black any variables that might influence the response

LO9 Identify if single or double blinding has been used in the study

* Blinding: when subjects don’t know if they are in the experimental or control group
* Double blinding: when both the subjects and researchers don’t know