**Statistics**

* Statistics – science of conducting studies to study, organize, summarize, analyze, and draw conclusions from data
  + **Descriptive statistics** – used to describe the situation in data
    - Consists of collection, organization, summarization, and presentation of data
    - E.g. average age of students
  + **Inferential statistics** – making inference from samples to population using probability statements
* **Variable** – a characteristic/attribute that can have different values; e.g. age, weight, gender
  + Discrete variable – has values that can be counted; e.g. # of people
  + Continuous variable – can have all values b/t any 2 specific values; e.g. height
  + Random variable – value is determined by chance; e.g. # of heads/tails
* **Population** – collection of all individuals under study
* **Sample** – a group of subjects selected from a population
* Statistical hypothesis – a claim/statement about a population
* Hypothesis testing – decision making process for evaluating a hypothesis using information obtained from samples
* **Raw data** – data collected in their original form
  + A **frequency distribution** can be constructed by organizing raw data
  + Frequency – # of values in a specific class of the distribution
* **Grouped frequency distribution**
  + There should be 5 – 20 classes
  + Class width should be an odd number – guarantees midpoints are integers
  + Classes must be mutually exclusive, continuous, exhaustive (accommodate all data), and equal in width
  + Determine the classes
    - Find the highest & lowest values
    - Find range R = highest – lowest
    - Select N = # of classes
    - Find width W = R/N
    - Select lowest class limit (≤ lowest data value)
    - Add W to get lower limit of next class; find all class boundaries
    - Note class lower limits are inclusive; upper limits are exclusive
  + Tally the data & find frequencies for each class
  + Find cumulative frequencies
  + **Histograms**
    - Bell-shaped vs. uniform
    - J-shaped vs. reverse J-shaped
    - Right vs. left skewed
      * Right skewed = mode is towards the left
    - Unimodal vs. bimodal
  + **Stem-and-leaf plots** – more useful for smallest data sets
    - Leading digit(s) = stem; trailing digit = leaf
    - Depth – count cumulative frequency from both sides until the class containing the median is reached