**IN402: Unit 7 Seminar, Laurence Burden**

* Unit covers the most important prediction techniques
  + Hypothesis testing
  + Linear regression
  + Logistic regression
  + Market basket analysis
  + Etc.
* AI sits upon the mountain of statistics
  + Probability is the soul of the mountain
* We don’t pursue exact solutions
  + We pursue the model that is most likely to happen
    - Models that outperform others and have higher probabilities
* Chapter 4: Understanding probabilities
  + Measure of likelihood that an event will occur
  + Two types
    - Discrete
      * Finite options
      * Ex: dice (1, 2, 3, 4, 5, 6)
    - Continuous
      * Infinite
      * Ex: all numbers between 0 and 1 (ie. 0.1, 0.11, etc.)
  + P = nE/nT
    - nE = number of predicted event
    - nT = number of total events
  + Binomial distribution
    - Most important
    - Ex: flipping a fair coin 10 times
      * Measure the distribution of heads
      * Each flip is a random test
  + Common mistakes
    - Always state probability as a percentage
    - Misinterpreting conditional probabilities
      * Conditional probability: a probability of an event that relies on another event happening first
      * Ex: class of 100 students in a stats class. 30 students also take an English (E) course. 40 also take a biology (B) class. 10 are taking both English and biology.
        + P(of taking both E and B) = 10%
        + P(of taking E given B) = .1/.4 = .25 = 25%
    - Confusing probability with possibility
      * Possibility is the ability for an event to happen
        + Simple yes or no
      * Probability is the likelihood of a given event to happen
    - Ignoring the denominator
      * Denominator in the formula represent the total number of events to be measured against
      * This must be taken into account to get a true probability measurement
      * Probability is a measure of relative frequency
* Chapter 5: Basic Statistical Ideas
  + Measures of central tendency
    - The central measurement
      * Mean, median, and mode
  + Measures of variability
    - Measures how variable the range is
    - Types:
      * Range (Max – Min)
      * Standard Deviation
        + Average difference of each observation in contrast to the mean
  + Hypothesis testing
    - Test between two groups
    - Null hypothesis (H0) and alternative hypothesis (Ha)
    - Ex: is the average height of men higher than women?
      * H0 = the average height is the same for both groups
      * Ha = The average height is significantly different for each group
  + Common mistakes
    - Misinterpreting central tendency measures
    - Confusing variability measures
    - Ignoring significance level
* Descriptive statistical concepts
  + Summarize and describe data
    - Central tendency
      * Mean, median, mode
    - Variability
      * SD, variance, range
  + Visualizing data is a great way to understand the data and to identify patterns