Stats Lecture – Week 2 - Wednesday

Announcements:
- Quiz – 9:30-9:45am Tuesdays –covered some major concepts from Ch 1-3 – to reopen Thursday
- Reading this week: by Tuesday Ch. 1, by Wednesday Ch 2-3.

Today:

- Lab Recap….
- More on probabilities…. (exercises)
- CS Subdiscipline overlap - Database w/ Data Mining!

1. Axiom: Sum i = 1,n P(Ai) = 1.0 Sum of all probabilities of all outcomes Ai = 1.0.
                    Outcomes must be mutually exclusive & exhaustive

    a. Complex events. Composites of simple events in the sample space
       sum individual probabilities
            P(A v B v C v D) = P(A) + P(B) + P(C) + P(D)

            Whirligig beetle offspring – example!

    b. Shared events .
        multiple simultaneous occurrences of simple events in the sample space
        If 2 events are independent of one another,
                the probability that both occur (shared event) is
                        product of individual probabilities

            EXAMPLE 1 – P(ace) = **1/52** + 1/52 + 1/52 + 1/52 = **4/52** = 1/13

            EXAMPLE 2 - whirligig beetle offspring - exercise!

            EXAMPLE 2 (cont) – exercise!

    c. What if you have both complex and shared events? Pp 15-21
            P (A | B) = P (A Ω B) / P (B)

Concepts you will need to understand to do the lab (the lab will help!):

1. Binomial Probability
2. Combining probabilities
3. Measures of Central Tendency: Mode, Median, Mean (summation notation)

   Mean (X-bar): A sample estimate of the population mean (µ)

   $$\overline{Y} = \frac{\sum\limits_{i=1}^{n} Y_i}{n} = (Y_1 + Y_2 + Y_3 + Y_4 + ... + Y_n)/n$$

4. Trial, sample, population
5. Measures of dispersion (spread):
   a. range
   b. sample variance: sum of each of the differences or deviations between each individual value and the mean value.

   $$s^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (Y_i - \overline{Y})^2$$

   c. **Standard deviation:** The square root of the variance.
   d. **Standard error:** the standard deviation divided by the square root of the sample size (n): most commonly used estimate of variance around means for figures in scientific papers in biology.

   $$SE = \frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{1}{n-1} \sum\limits_{i=1}^{n} (Y_i - \overline{Y})^2}}{\sqrt{n}}$$

   e. **Coefficient of variation (CV).** Measures the variability of values in a sample relative to the magnitude of the sample mean ($\overline{x}$) – gives an index of population variability that is comparable across measurements and populations.

   $$CV = \frac{\sqrt{\frac{1}{n-1} \sum\limits_{i=1}^{n} (Y_i - \overline{Y})^2}}{|\overline{x}|}$$

   f. Quantiles, quartiles

Tools you will need to use to do the lab (the lab will tell you how to do these):

1. Excel
   a. Enter equations using Excel functions to compute summary statistics
   b. Sort data
   c. Create a histogram
   d. Create a Bar Graph (either in JMP or Excel)
2. JMP
   a. Copy/paste data from Excel into a JMP table
   b. Run summary statistics in JMP
   c. Copy/paste JMP tables back into Excel (to make a nice table)
   d. Create a Bar Graph (either in JMP or Excel)
   e.