Stats Lecture – Week 2 - Monday

**Announcements:**

* Quiz – 9:30-9:45am Tuesdays – will cover major concepts from Ch 1-3.
* Reading this week: by Tuesday Ch. 1, by Wednesday Ch 2-3.
* Seminar assignments posted on the doors to seminar this afternoon.
* Lab 1 Modeling: If you have NOT handed in your lab, last chance is tomorrow 9:30!

**Today: Overview of this week’s reading, and then Lab Preview.**

1. Why statistics?
   1. Are patterns we see due to chance alone?
   2. How likely are they?
   3. Example – the scientific process….
      1. Go into nature, observe and wonder…”why are there more ants in the forest?
      2. Formulate a testable hypothesis,
      3. devise an experiment to test that hypothesis
      4. ???
      5. ???
      6. ???
      7. ???

1. Probability & uncertainty – it’s everywhere. …
   1. 40% chance of rain
   2. 100 year flood, e.g., usgs.gov
   3. Chance of winning, e.g., 1 in 4.5
2. Precision vs. accuracy (in stats)
   1. Precision: Agreement among many measurements, e.g., on same instrument.
      1. Contrast to “precision” for computer scientists….
   2. Accuracy: how close is the value to truth?
3. Terms
   1. Event – insect visits to a plant
   2. Outcome – 2 possibilities (prey capture, prey escape)
   3. Trial
   4. Probability -- #outcomes / # trials – 0.0, 1.0, 0.5
   5. Sample – can’t measure all plants in a population. vs census (all)
4. 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

* 1. Complex event P(A v B v C v D) = P(A) + P(B) + P(C) + P(D)

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

* 1. Shared (combined) events and conditional probabilities – WEDNESDAY

Ch. 2 – probability distributions.

1. Binomial – flip coin = n! / X!(n-X)! \* p\*\*x (1-p)\*\*n-x
   1. Histogram showing probabilities of many binomial trials.
   2. Problems –
      1. very cumbersome when n large, or p small
      2. must know number of attempts (number of seeds that didn’t germinate)
2. Other distributions – WEDNESDAY….
3. The bell curve! Continuous variables….
   1. Mean
   2. Variance (standard deviation).
      1. 68% all values – 1 SD from mean,
      2. 95% - 2 SD,
      3. 99.7% - 3 SD.

**Central Limit Theorem** – regardless of the underlying distribution of a variable, the distribution of many, many means of that variable will be normally distribution.

**Law of large numbers** – Observations must be random, independent, and drawn from large population

**Types of variables**

* Discrete
* Continuous
* Categorical

**Populations, samples, censuses –**

**Summary statistics**

1. Arithmetic mean – Ybar - write this in summation notation
   1. Geometric Mean – for rates - GMy – always a little smaller than arithmetic mean
2. median, mode
3. Spread –
   1. sample variance, standard deviation, standard error – in graphs use SD or SE – convention by field.
   2. Range, quantiles, quartiles, box and whisker plot
   3. Coefficient of variation – unitless –standardized by the mean, comparable to other measures. JMP multiplies this by 100

**The Lab:**

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)

****

1. Trial, sample, population
2. Measures of dispersion (spread):
   1. range
   2. sample variance: sum of each of the differences or deviations

between each individual value and the mean value.

****

* 1. **Standard deviation:** The square root of the variance.
  2. **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.

****

* 1. **Coefficient of variation (CV).** Measures the variability of values in a sample relative to the magnitude of the sample mean () – gives an index of population variability that is comparable across measurements and populations.

****

* 1. Quantiles, quartiles

**Tools you will need to use to do the lab:**

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

**About graphs – and reporting them!**

1. Dependent variable on Y-axis, independent on x-axis.
2. ALWAYS use a figure legend!
3. Know what kind of variables you have: continuous, categorical, discrete….
   1. Bar chart – categorical variables – example – male/female heights.
   2. Scatterplot - linear analysis.

**On Wednesday – what you (should have) learned in this lab!**