Week 3 Stats, Wednesday

Announcements:

- Returned Stella Labs in lab…. Lab KEY posted on fileshare. Didn't get it back?
- Excel Tips – posted on web site, and fileshare

Today:

1. Recap from last week and Monday….
2. Experimental Design
3. Questions about this week's lab!

I.Recap from last week and Monday….

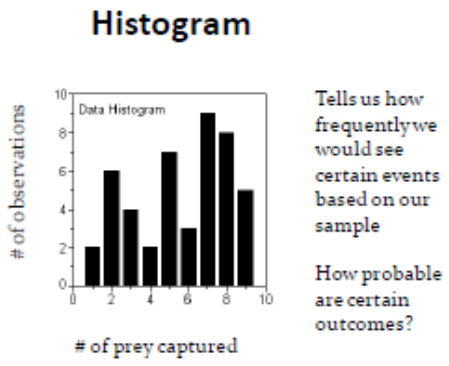
1. Terms you should know….
    a. Event, outcome, trial, replicate, experiment
    b. Populations (parameters, e.g., μ), samples (statistics, e.g., Y), census
    c. Independent, dependent variable
    d. variable types - categorical, continuous, discrete/ordinal
    e. Probability
        i. Complex event – P(A v B v C) = P(A) + P(B) + P(C)
        ii. Shared events – P(A ∩ B) = P(A) x P(B)
        iii. Combined Events – P(A ∪ B) = P(A) + P(B) - P(A ∩ B)
        Some events are not independent of each other, e.g., whirligig beetle fitness, if fitness in Year 2 depends on # offspring in Year 1
        iv. Conditional Probabilities: event B depends on event A ~ Bayes' Theorem
        P (A | B) = P(A ∩ B) / P(B)
    f. Summary statistics (Law of Large Numbers)
    g. p-value
2. Graphs we've done:
    a. Histogram how frequently would see we certain events based on our sample

Comment [JBC1]: Outcomes mutually exclusive & exhaustive

Comment [JBC2]: Advanced! milkweed example

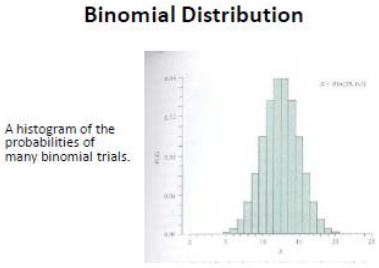
Comment [JBC3]: Requires "prior probability"

Comment [JBC4]: INSERT EXAMPLE where are independent and dependent vars….???



b. Box and whisker plot

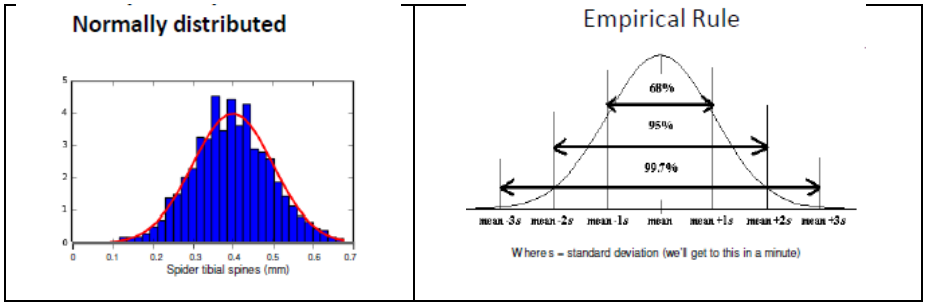
3. Distributions
   a. Binomial



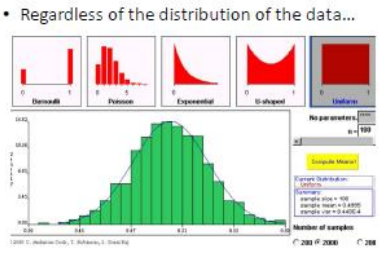
b. Poisson – rare events – $P(X) = \lambda^x / X! * e^{-x}$

Comment [JBC5]: base of natural log 2.718

c. Gaussian – continuous variables – probability density function



4. Central Limit Theorem



http://www.intuitor.com/statistics/CentralLim.html

II.Experimental Design

Monday…

Initial Observation…& Background Research

Scientific Hypothesis (Prediction)

What are the variables?

Formulate a Null & Alternative Hypotheses

Collect and analyze data

Reject H0?

Report Results (& continue….)

This week's Lab asks you to devise your own 2 experiments.

- **Q1**: Comparing 2 groups - a treatment to a control (are means different? t-test, or ABS(AVERAGE (x1s) – AVERAGE (X2s), as in lab)….
- **Q2**: Linear relationships: Correlation or regression (you choose which)

For each experiment, explain your experimental design (including sample size, sampling protocols, etc.) in paragraph form with complete sentences. You should have several paragraphs for each experimental design. The more detail you include the better, such as where and when your study would take place, which variables you would measure, etc.

a) What is your null hypothesis?
b) What is your research hypothesis?
c) What is (are) your independent variable(s)?
d) What is (are) your dependent variable(s)?
e) What are your predictions for the outcomes of your study?
f) Find one scientific paper that reports on a similar experiment – give a bibliographic citation for that paper, and in 2-3 sentences say what the paper is about.

What do you need to know, to do this?

How many samples? Rule of 10

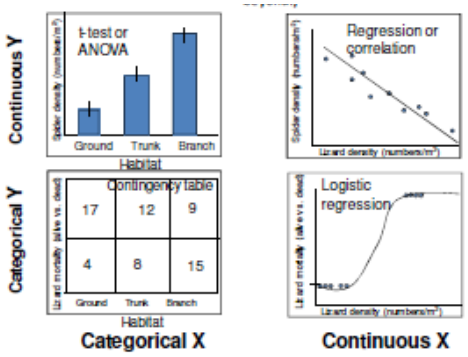
Independence and confounding factors….

- What if fertilized plots end up on sunny hillsides & control plots are shaded

Types of Experiments

- Manipulative
  - Press – treatments maintained over time (reapply fertilizer to maintain N)
  - Pulse – treatments applied once, and system allowed to recover
- Natural (observations)
  - Snapshot – replicated in space
  - Trajectory – replicated in time

Deciding what statistical test(s) to run: The 4 Basic Analyses (below). This week we worked with a categoricalX and continuous Y, and compared two means by taking their absolute value, and running resampling stats to see if the difference between the two groups was due to chance alone. Next week, we'll do t-tests (and the week after ANOVA (both still in upper left corner!)



III.This week's Lab!

Running resampling stats…. RESET is your friend. So is the red box!

Explore resampling stats (Monte Carlo) using male and female student height data from my last class. Ask -- is the apparent difference in male and female heights "real"?

a. an intuitive way of understanding p-values
number of times out of 1000 random shuffles
you have a difference as large as or larger than the observed difference –
is the difference you see between M, F due to chance alone?

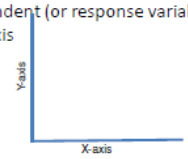
b. Use it when assumptions of a statistical test are not met

Then, do a new problem using the CPAT data!

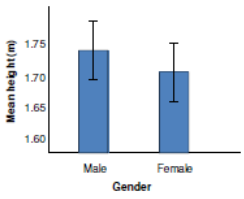
Data Validation & Cleaning!!!

- Computing the p-value
- Generating the Histogram in Excel, with error bars – see fileshare Handouts\StatsModelingVisualization\errorBarsInExcel.docx
- How to report graphs (as in a scientific paper) – see below

- Know what types of variables you have
  - Continuous, categorical, ordinal
- Your independent variable goes on the x-axis
- Your dependent (or response variable) goes on the y-axis



(Continuous Y, Categorical X; No difference between groups)



Bar height = mean
Error bars = standard error
n = sample size

Note the axis labels! Not a default setting in excel...

Categorical variables as bar charts