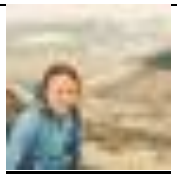
Week 3 Stats, Monday

Collect data (anonymously) as in the ss.

Return Stella Labs tomorrow in lab…. Lab KEY posted on fileshare.

 **1st stats lab due tomorrow.** Help session with Robyn Andrusyszn 3-5 in 2617.	 **here's Kara Karboski! Lab Aid during labs…**

Want your own version of JMP? Take a disk and your Evergreen ID to the CAL.

Unfortunate;y, I don't think you can get the resampling addin for excel…..

Which of these is Classic Statistics?

All ants in the Harvard Forest are in the genus *Myrmica*. I collected this ant in the Harvard Forest. → This ant is in the genus *Myrmica.*	All 25 of these ants are in the genus *Myrmica* All 25 of these ants were collected in Harvard Forest. → All ants in Harvard Forest are in the genus *Myrmica*.

Induction. From the specific to the general.

(Constrast with deduction):

Scientific Process

Initial Observation…

Scientific Hypothesis (Prediction)

Formulate a Null Hypothesis and Alternative Hypothesis

Collect data

Reject? Or Fail to Reject? The Null Hypothesis

Report Results (or continue….)

Example: Do Evergreen's MES students collect data as well as the professional field staff at HJA? I will use data from the stem maps of plots surveyed by MES gCORE students in Fall 2011. I will then pay professional field technicians to survey those same plots and compare results.

Do some background research; scientific papers on a similar experiment:

1. J. M. Laflen, M. Amemiya, and E. A. Hintz, Measuring crop residue cover, *Journal of Soil and Water Conservation November/December 1981 vol. 36 no. 6 341-343.* The photographic, meterstick, and line-transect methods of measuring the percentage of land covered with crop residue were evaluated by comparing residue cover on 10 Iowa fields farmed with conservation tillage. Although the line-transect method overestimated crop residue by 6 to 10 percent, variability in measurements proved less than with the photograph and meterstick methods. The line-transect method is preferable for field use, particularly if technicians are carefully trained in its use and in self-calibration so that overestimation of crop residue can be avoided.
2. Learning landscape ecology: a practical guide to concepts and techniques, SE Gergel… - 2002
3. The "freezing" of science: consequences of the dogmatic teaching of ecology, [PDF] from uba.ar, R GONZÁLEZ DEL SOLAR… - BioScience, 2001 – BioOne

The sample size is the number of measured trees in the plots. We will look at only DBH, and compare the DBHs for each tree. I will assume that the field technicians will make correct measurements Mf, and then subtract Mm (dbh measured by MES students) from Mf. I will use a two sided t-test, because I want to see if MES students consistently under- or over-measured.

The data table would look like something like this:

tree tag	DBH-field tech	DBH-MES student	
	Mf	Mm	Mf - Mm
7777	38	38	0
7983	27	27	0
4642	26.6	28.6	-2
4293	27.4	29.4	-2
4694	24.9	23.9	1
5915	38.5	35.5	3
4821	40.7	38.7	2
5658	34.5	35.5	-1
5401	16	16	0
5659	22	22	0
7779	27.5	27.5	0
4586	19.2	21.2	-2
2575	18.3	19.3	-1
2078	38.3	36.3	2
4798	23.1	24.1	-1
5408	39.5	36.5	3
5950	16.7	16.7	0

I will compare the differences Mf - Mm with 0. What are my null and alternative hypotheses?

My null hypothesis is that there is no significant difference, i.e., the MES students' measurements were not significantly different from those of the field technicians'.

If I can reject the null hypothesis, I will accept the alternate hypothesis that MES students' measurements are not as accurate as professional field technicians'.

H0: No difference between measurements by Evergreen MES and field technicians.
Ha: MES students' measurements are significantly different

independent variable(s)? The two data collection groups
dependent variable(s)? DBH

What are my predictions for the outcomes of your study? I think Evergreen students are still learning field techniques, so I suspect there will be a difference.

What happens next?

I perform the statistical analysis; what do I get?

H0: The difference we see is just random variation (simplest explanation!)

Ha: The observed difference is too large to be accounted for by random variation.

A test statistic and a p-value….

What is "too large"?

Comment [JBC1]: DISTRIBUTE P-VALUE HINT!

P-value: = P(data | H0)

Low P does NOT → Your data!

What determines a P-value?

1. Number of observations in the samples (n)
2. Differences between the means of the samples
   ABS((AVERAGE (Yj's) – AVERAGE(Yj'S))
3. Degree of variation among individuals (s^2)

Resampling stats provides an intuitive way of understanding p-values!

Example: Do heights of male and female students from my last class differ?

What's the p-value? (number of times out of 1000 - due to chance alone –

I find a difference as large as or larger than the observed difference.)

Talk about and demo resampling stats.

1. Open Excel
2. All Programs – Statistical Tools – Resampling Stats
3. In Excel – Enable Add-Ins, Click Add-Ins, see Resampling Toolbar
4. Resample within Columns???

How to report graphs (as in a scientific paper).

For the lab:

Outline a study, as per Section 1 in Week 3 Lab !

- **Q1**: Comparing a treatment to a control (analyze using a t-test)
- **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.

For each experiment, outline how you would test your null hypothesis. For each, include:

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.

On **Wednesday,** more on Baysian methods, and on study design….