**WEEK 3 - 2015-11-04**

**4 - Experimental Design**

* Have a plan for data and code sharing - GitHub / figshare
* Leek group to data sharing in GitHub

**Formulating your Question in advance**

**Question:** Does changing the text on your website improve donations?

1. Randomly show visitors one version or the other
2. Measure how much they donate
3. Determine which is better - A/B - test

**Statistical inference**

Check if the small data (A/B - test) applies on the larger population

**Variability - Scenario 1**

For every 1000 visitors how much was donated depending on website structure and objectives. (Signup version or donate version)

**Confunding**

Paying attention on what variable changes the other

**Correlation is not causation**

“One reason of many reasons” - Prove to yourself what you believe.

**Randomization and blocking**

* If you can fix a variable
* If you dont fix variable, stratify it! (If you testing a website and using two colors, use both phrases equally on both of them)
* If you cant fix a variable, randomize it!

**Prediction**

* **Sensitivity** Positive l disease
* **Specificity** Negative l no disease
* **Positive Predictive Value** disease l positive
* **NegativePredictive Value** no disease l negative
* **Accuracy** correct outcome

**Beware data dreging**

* Jelly beans cause acne! - Proved NOT
* Its a specific color of Jelly Beans that cause ACNE
* After 10 tests they find what color of Jelly Beans cause acne
* **This ignores the fact that you did 10 tests!**

**Summary**

**Good experiments**

* Have replication
* Measure variability
* Generalize to the problem you care about
* Are transparent

**Prediction is not inference**

* Both can be important

**Beware of data dredging**