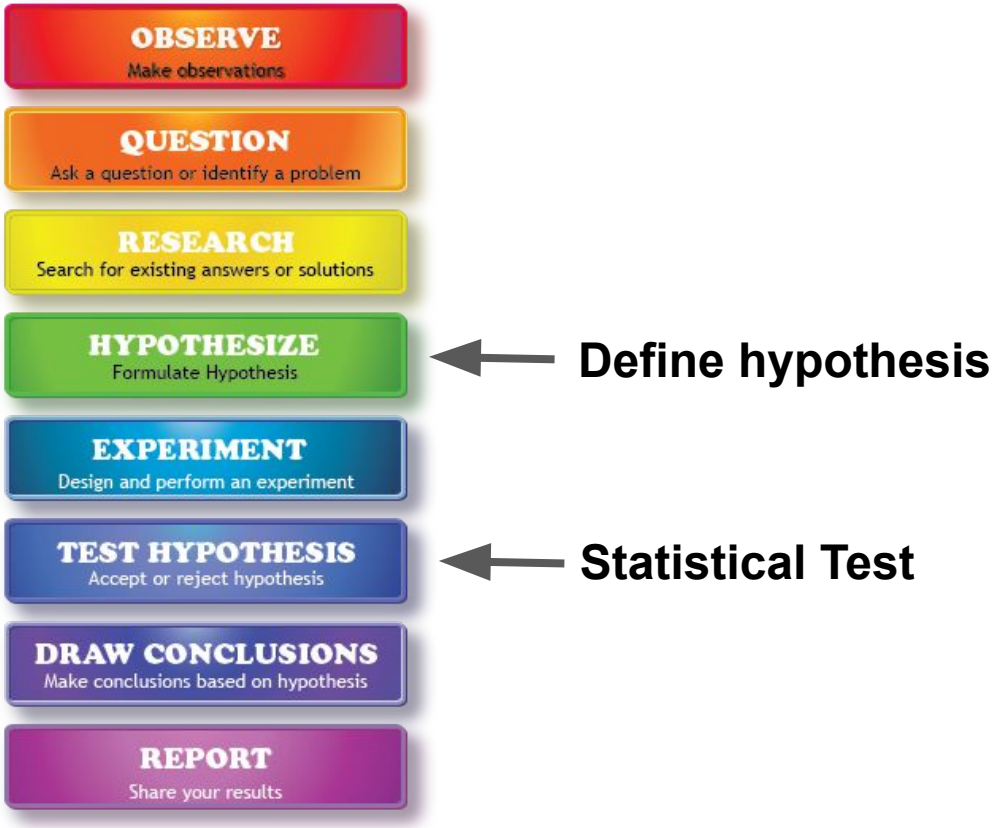
Lecture 21
Hypothesis Testing

Scientific Method

Scientific Method



How to test a hypothesis

- Proving a something is “right” is very hard to do

- Proving something is “wrong” is much easier

How to test a hypothesis

- The standard approach is to make 2 hypotheses
  - Null hypothesis - the “accepted”/”default” hypothesis
    - No signal, no new physics, etc.
  - Alternative hypothesis - Your science hypothesis
- Your goal is then to “rule out” or disprove the null hypothesis

Hypothesis Examples

- We want to determine if a source is variable
  - Null hypothesis - source has a constant flux
  - Alternative - the source flux is not constant

- Is there a source detected or is it just background?
  - Null - background fluctuation, follows bkg statistics
  - Alternative - there are signal counts + bkg counts

Hypothesis Examples

- Do we need a change to the standard model of particle physics, say a 4th neutrino flavor
  - Null hypothesis - standard model
  - Alternative - standard model + a 4th neutrino flavor

How to choose a hypothesis?

- It helps to state the hypotheses as math
  - Is there a source? In my N counts
  - Null - an expected average of Nbkg counts
  - Alt - an expected average of Nbkg + Nsig counts, where Nsig > 0

- Then a test needs to be defined that tries to “reject” the null hypothesis

Test Statistic

- A test statistic (TS) is defined
  - Generally chosen such that
    - its distribution under the null hypothesis is known
    - If the alt is true, will give a large value
- Is there a source? In my N counts
  - Let’s make our TS a signal to noise ratio
    - TS = (N - Nbkg) / σ_{bkg}
    - High with lots of signal counts
    - If there’s no signal counts (null is True) it should follow a standard normal

p-value

We then find the probability of observing that TS assuming the null is True

$$P(>=TS) = \int_{TS}^{\infty} PDF(x \mid Null)\, dx$$

$$= 1 - \int_{0}^{TS} PDF(x \mid Null)\, dx$$

We pick a predefined threshold, if it surpassed the threshold, reject the null!

Common thresholds are 0.05, 3 sigma (0.003), 5 sigma (~0.000001)

p-value

- Is there a source? In my N counts
  - For counting, we can use the normal distribution
  - PDF = Normal distribution
  - Say we know the bkg is constant with average Nbkg counts
  - Mu = Nbkg, sigma = sqrt(Nbkg)
  - P-value is the integrated tail at >= measured S/N (our TS)

What if we don’t know our PDF(TS | Null)?

- In this case, we have to approximate it
- This is often done by
  - doing many “trials” of existing data where we known the null is true
  - simulations of data where the null is true

p-value

- Is there a source? In my N counts
  - For counting, we can use the normal distribution
  - PDF = Normal distribution
  - Say we know the bkg is constant with average Nbkg counts
  - Mu = Nbkg, sigma = sqrt(Nbkg)
  - P-value is the integrated tail at >= measured S/N (our TS)