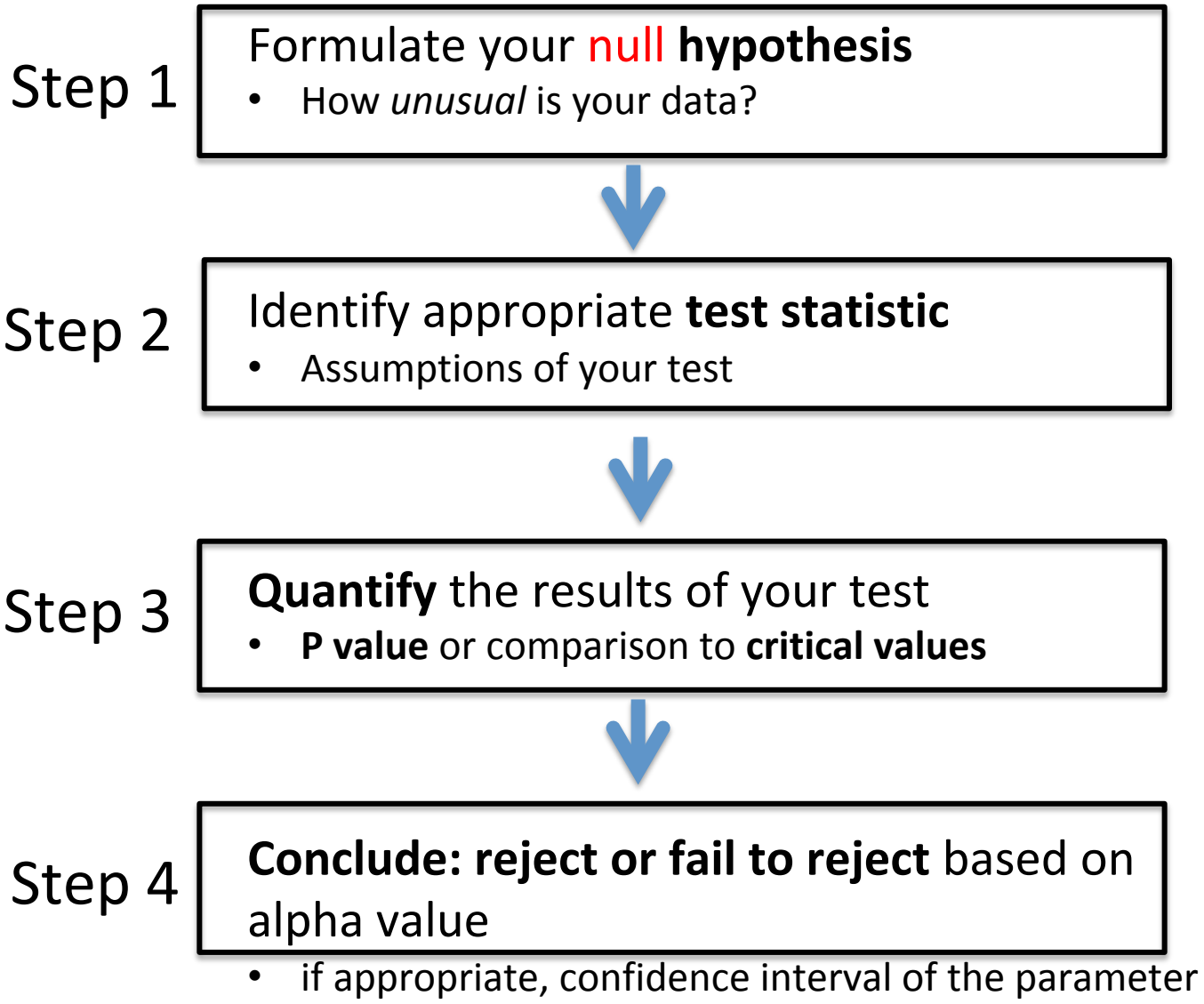
Your pipeline for hypothesis testing in statistics



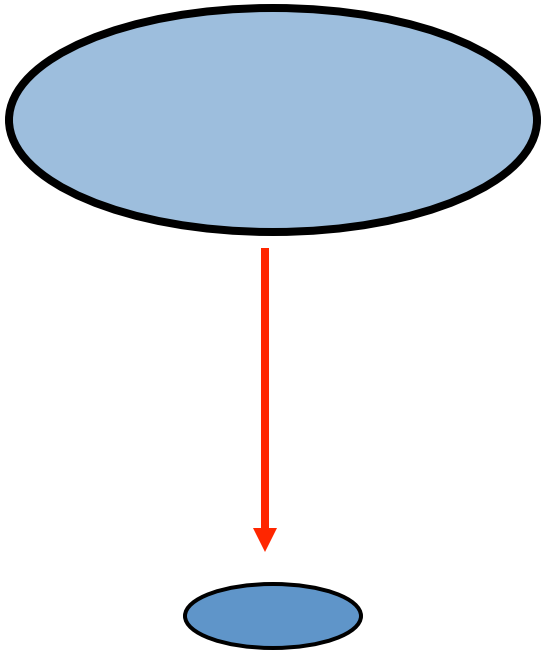
Hypothesis Testing

- Most of the work involved in statistics is clearly stating your hypothesis
  - What is your expectation? Can you quantify it?

- Hypothesis testing allows you to ask if a parameter *significantly* differs from the ***null*** expectation
  - It quantifies how unusual the data are *if you assume that the null hypothesis is true*

- Hypotheses are about populations but are tested with data from samples
  - Assumes that the sampling is random

Hypothesis Testing

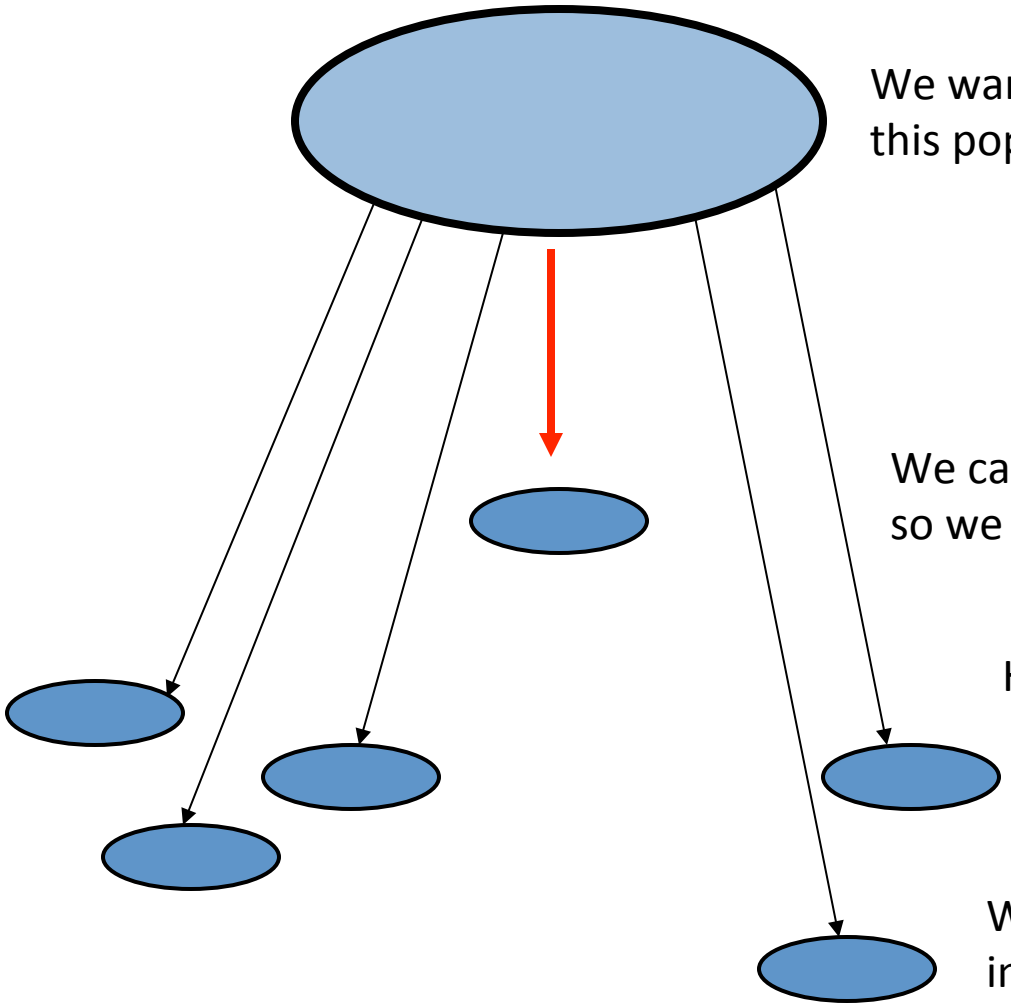


We want to know something about this population

We can't measure everyone so we take a sample

But! the sample doesn't necessarily have the same properties as the population due to chance errors.

Hypothesis Testing



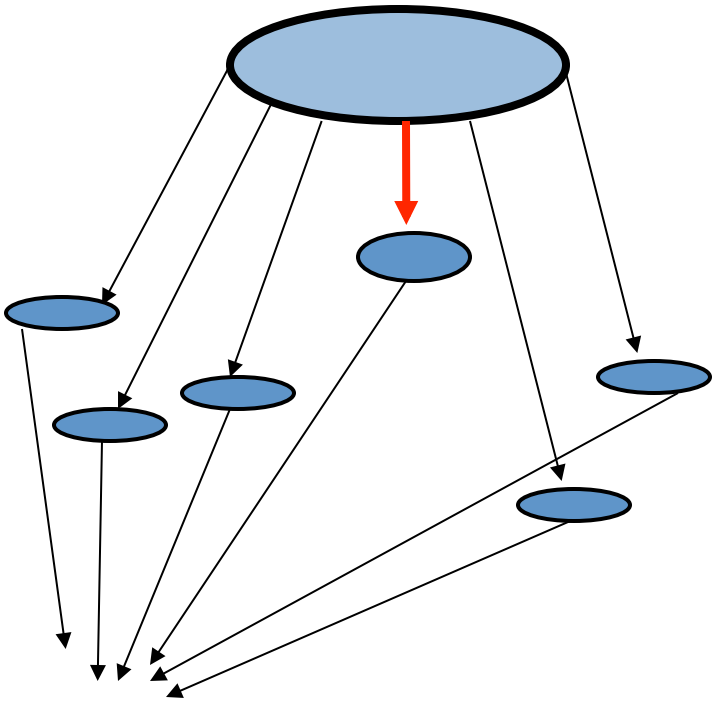
We want to know something about this population

We can't measure everyone so we take a sample

How '*good*' is the sample?

We imagine taking an infinite number of samples from the ***null distribution***

Hypothesis Testing

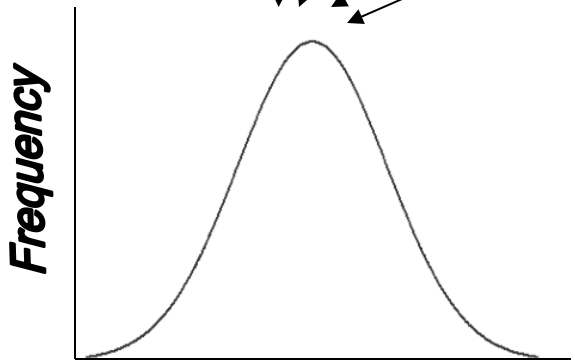


- We want to know something about this population
- We can't measure everyone

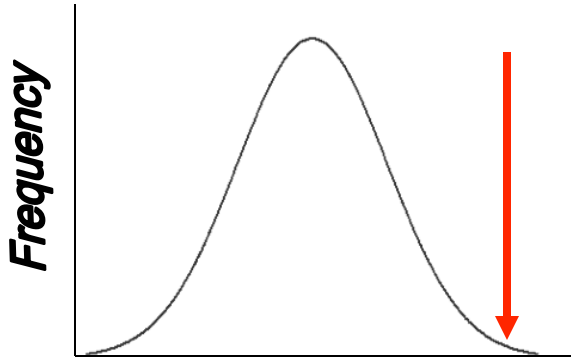
so we take a sample

How 'good' is the sample?

We imagine taking an infinite number of samples from the ***null distribution***



Difference in mean height



Difference in mean height

Making and using hypotheses:

The Null Hypothesis (H_0):

A specific statement about a population parameter made for the purpose of the argument

The Alternate Hypothesis (H_A):

Represents all other possible parameter values except that stated in H_0

Hypothesis Testing

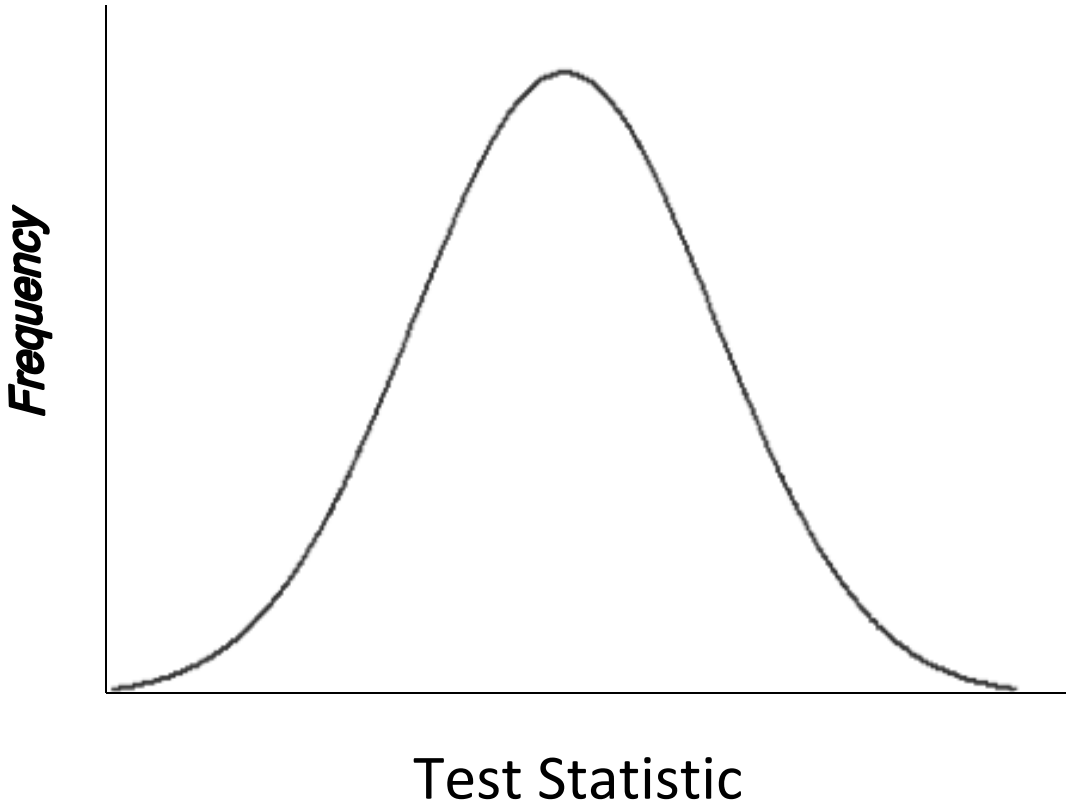
H_0:

- The *only hypothesis actually tested by the data*
- *Usually the skeptical POV*
  - *Claims **NO difference/effect***
  - *Observations are just due to chance*
- *Reject or Fail-To-Reject BUT **NEVER EVER** accept*
- *Rejecting H_0 reveals nothing about the magnitude of a parameter*

H_A:

- Usually the statement that the researchers *hope* is true

Hypothesis Testing



Test Statistic:

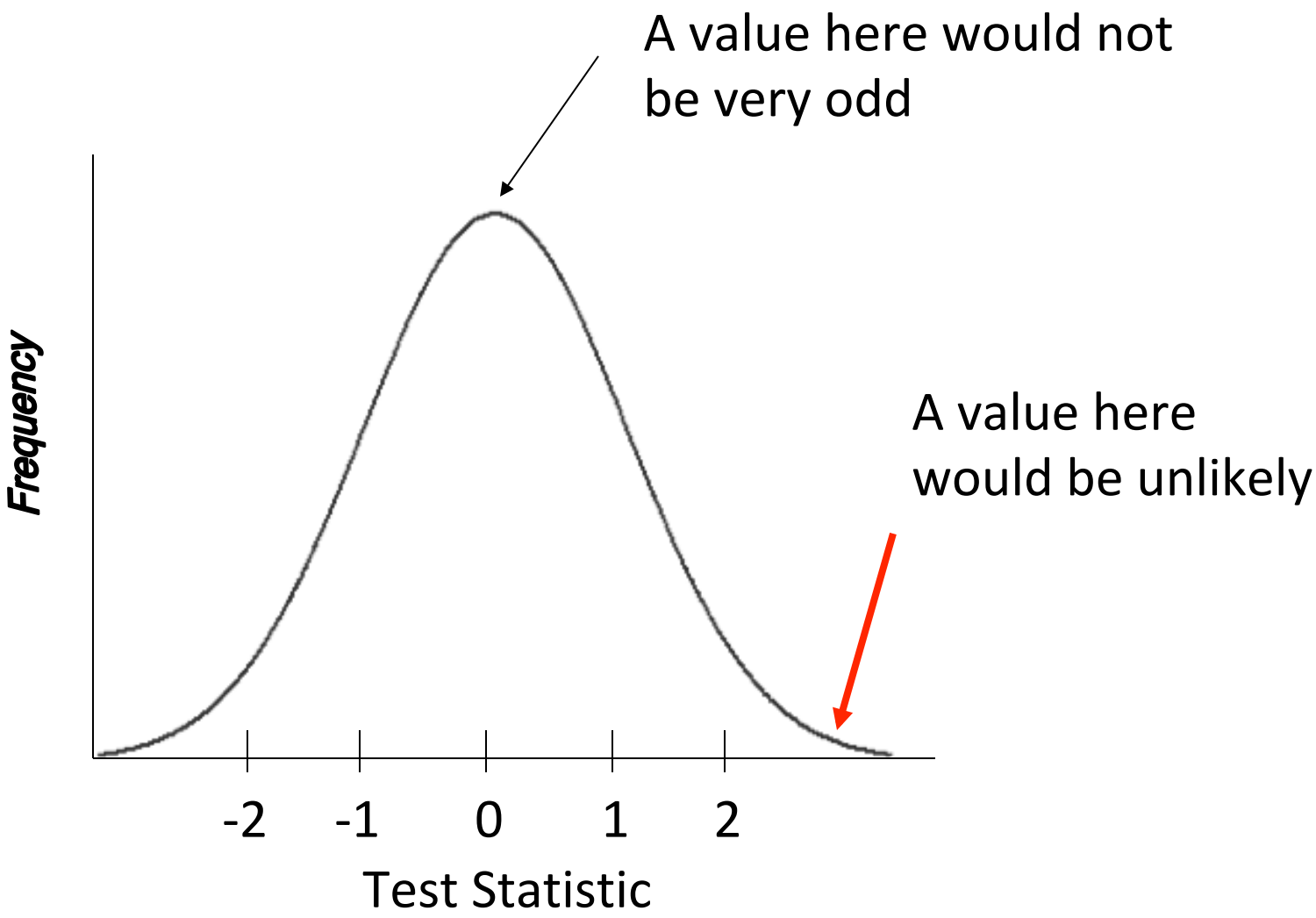
Quantity calculated from the data that is used to evaluate How compatible the results are with those expected the null hypothesis

Null Distribution:

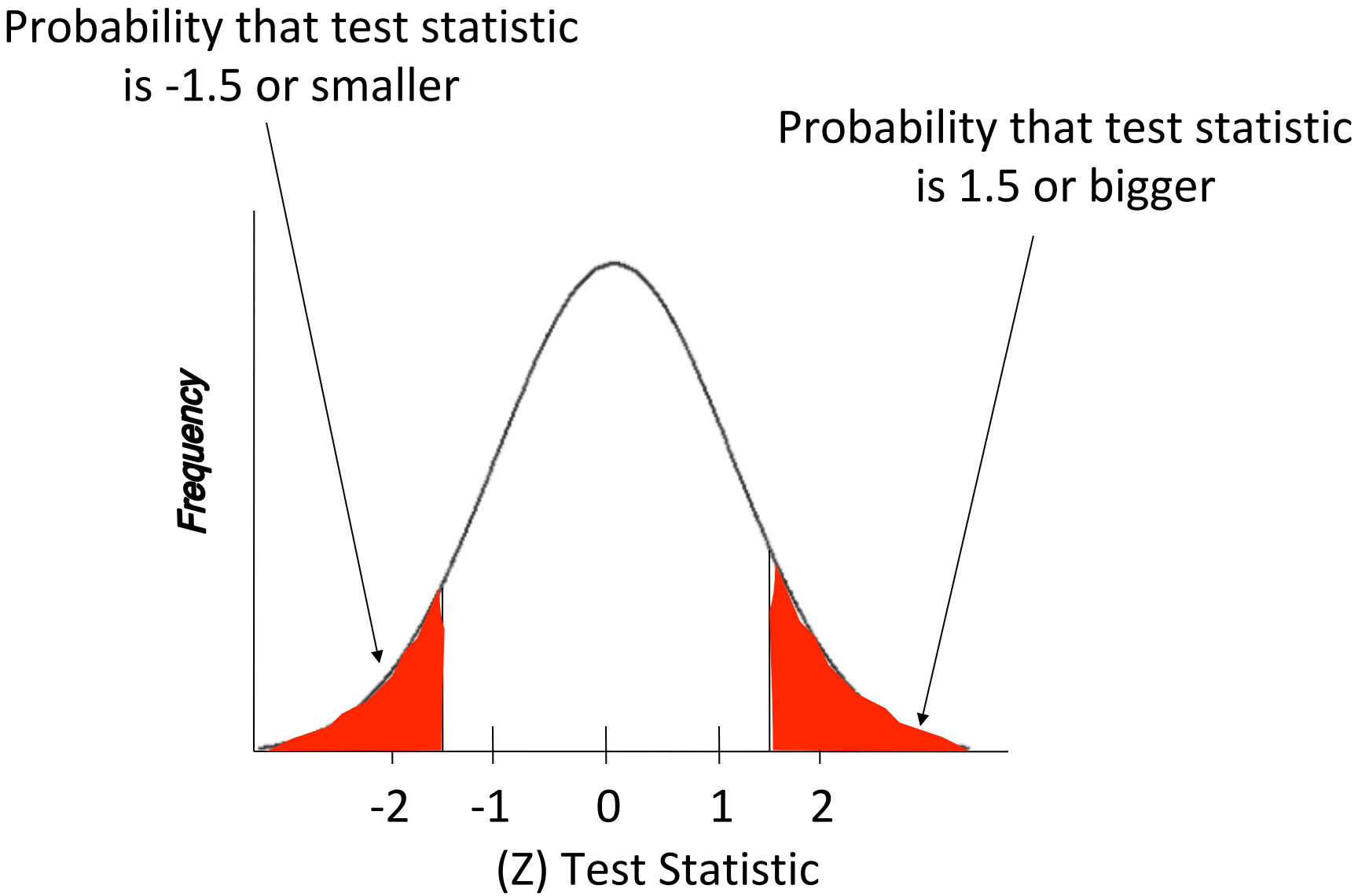
Probability of the test statistic assuming the null hypothesis

- often the null distribution is acquired via computer simulations/modeling

Hypothesis Testing

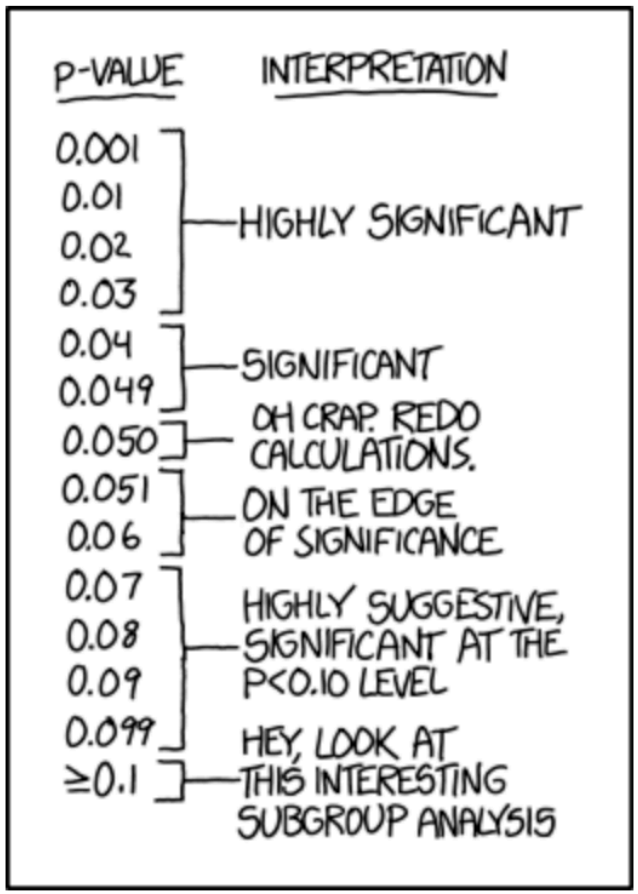


P-value


Probability that test statistic is -1.5 or smaller
Probability that test statistic is 1.5 or bigger
Frequency
-2
-1
0
1
2
(Z) Test Statistic

P-Value:

Probability of obtaining data that are equal to or even more extreme than the value assuming the null hypothesis is true



http://xkcd.com/1478/

How are P-values found?

- Simulation
- Parametric tests
- Re-sampling

How do you use a P-value?

In hypothesis testing you can do one of two things:

Reject or Fail-to-Reject H_0

Statistical Significance:

α is used as the basis for rejecting the null hypothesis (α is set by the experimenter, p-values are calculated from the sample)

If P-value <= α , H_0 Rejected

If P-Value > α, FTR H_0

*α is often 0.05

Hacking p-values: getting the p-value you need to publish your results

- Nate Silver has a widget that demonstrate ‘p hacking’ here:

https://projects.fivethirtyeight.com/p-hacking/

- should we get rid of p-values? http://fivethirtyeight.com/features/statisticians-found-one-thing-they-can-agree-on-its-time-to-stop-misusing-p-values/

- Even well intentioned, honest researchers can accidentally “p-hack”

  - Stopping the study when p-value is significant (**n** individuals) but continuing other studies with more **n** when p-value isn’t yet significant (so you end up with a bias towards studies that have greater **n** and so are more likely to pick up smaller differences)

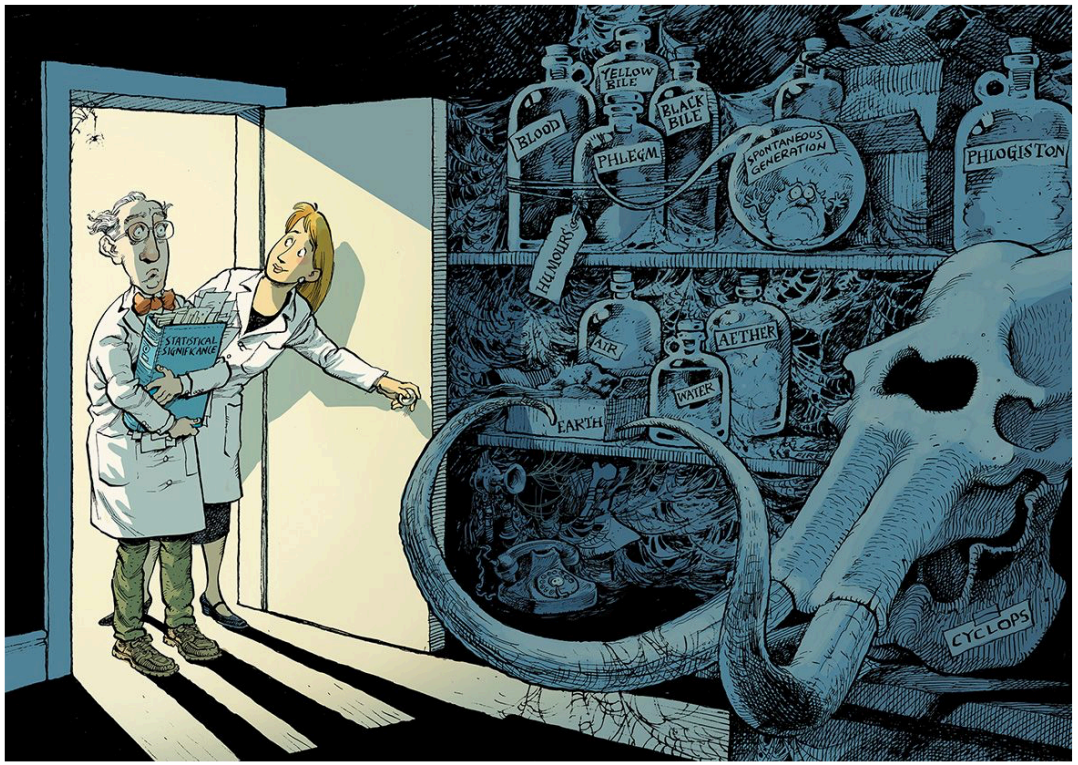
  - Play with outliers (include or exclude) until a significant p-value is achieved.

Recent NATURE article about rethinking **p values**

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein, Sander Greenland & Blake McShane



Which statement(s) is true about p-values?

---

a. p-value is the probability that the null hypothesis is true or false

b. p-value reflects the weight of evidence against the null hypothesis

c. p-value measures the size of the effect

d. if p value is less than or equal to the significance level, then the null hypothesis is not rejected.