Review:

Four steps in hypothesis testing:

1. **Formulate Hypothesis**

Review:

Four steps in hypothesis testing:

1. Formulate Hypothesis

2. **Identify Test Statistic**
   - Quantity from the population that is used to evaluate how ‘weird’, or extreme, the results would be if the null hypothesis was true
   - Example: Mean, proportion, t -test score
   - Probability distribution of test statistics is the null distribution

Review:

Four steps in hypothesis testing:

1. Formulate Hypothesis
2. Identify Test Statistic
3. **(a)Calculate P-value or Critical Value**
   - Probability of obtaining data that are *as extreme or even more extreme* than the value of the test statistic obtained

Review:

<u>Four steps in hypothesis testing:</u>

1. Formulate Hypothesis
2. Identify Test Statistic
3. **Calculate P-value or Critical Value**

(b) Compare to fixed significance

- Compare P-value (or critical value if you can't find P-value) to desired significance level to **<u>Reject</u>** or **<u>Fail-to-Reject</u>** the null hypothesis

Review:

Four steps in hypothesis testing:

1. Formulate Hypothesis
2. Identify Test Statistic
3. Calculate P-value or Critical Value/Compare to fixed significance

4. Conclude

- Always explicitly state your conclusion about the null hypothesis and any interpretation that goes with it
- Almost always include a confidence interval

Review:

Four steps in hypothesis testing:

1. **Formulate Hypothesis**
2. **Identify test statistic**
3. **Calculate P-value or critical value/ Compare to fixed significance**
4. **Conclude**

Review:

Four steps in hypothesis testing:

1. Formulate Hypothesis
2. Identify test statistic
3. Calculate P-value or critical value/ Compare to fixed significance
4. Conclude

What does this mean?

Errors in hypothesis testing:

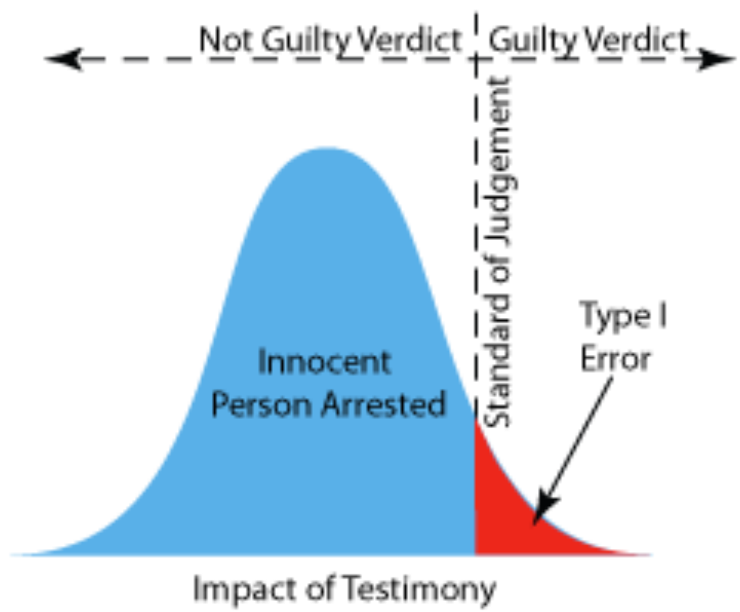
Type I (α)

Type II (β)

Type I (alpha) error:

Rejecting a true null hypothesis

$$P(\text{reject } H_0 | H_0 = \text{true}) = \alpha$$

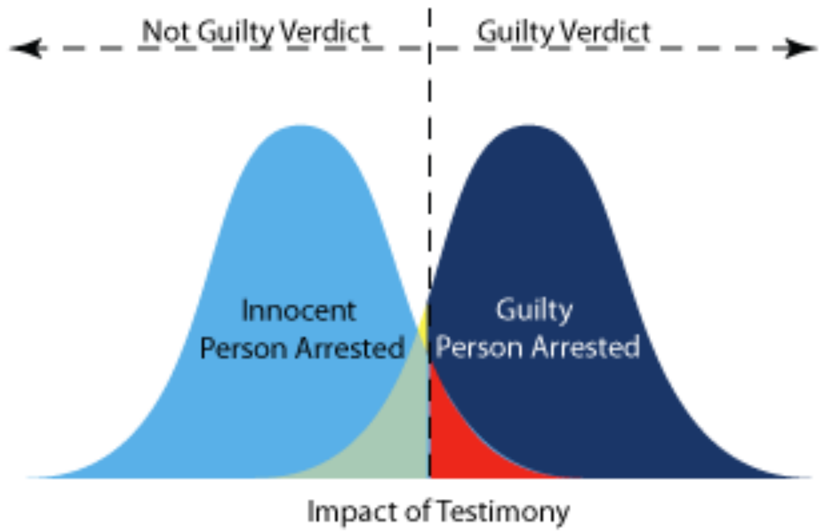


http://www.intuitor.com/statistics/T1T2Errors.html

Type II (beta) error:

Not rejecting a false null hypothesis

$$P(\text{Fail to reject } H_0 | H_A) = \beta$$



http://www.intuitor.com/statistics/T1T2Errors.html

	No Disease (H_0 true)	Disease (H_A true)
Fail To Reject H_0	No Error	Type II
Reject H_0	Type I	No Error

<u>Power</u> is the ability of a test to reject a false null hypothesis

$$\textbf{Power = 1 -}\ \beta$$

Power = 1 - P(FTR H_0| H_A)= P(Reject H_0| H_A)

* Power can be increased by increasing the sample size, *n* *

Add more data points, n, and you are able to discriminate between smaller differences in the null and alternate hypotheses!

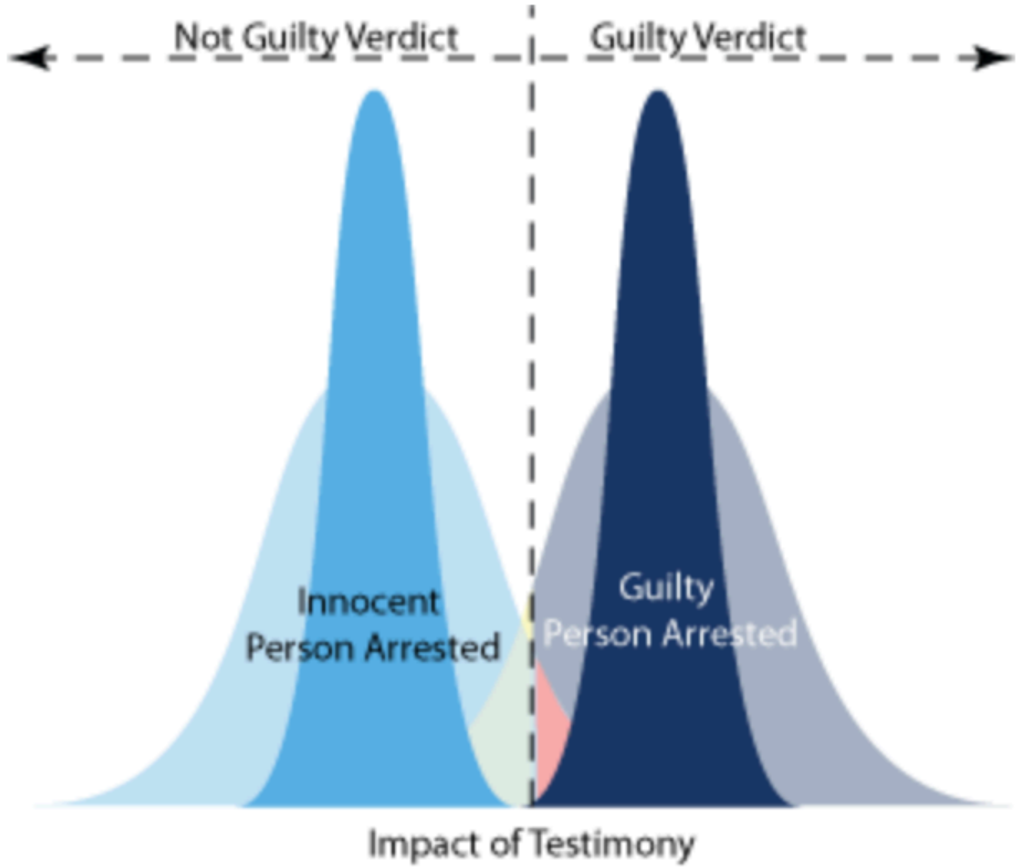


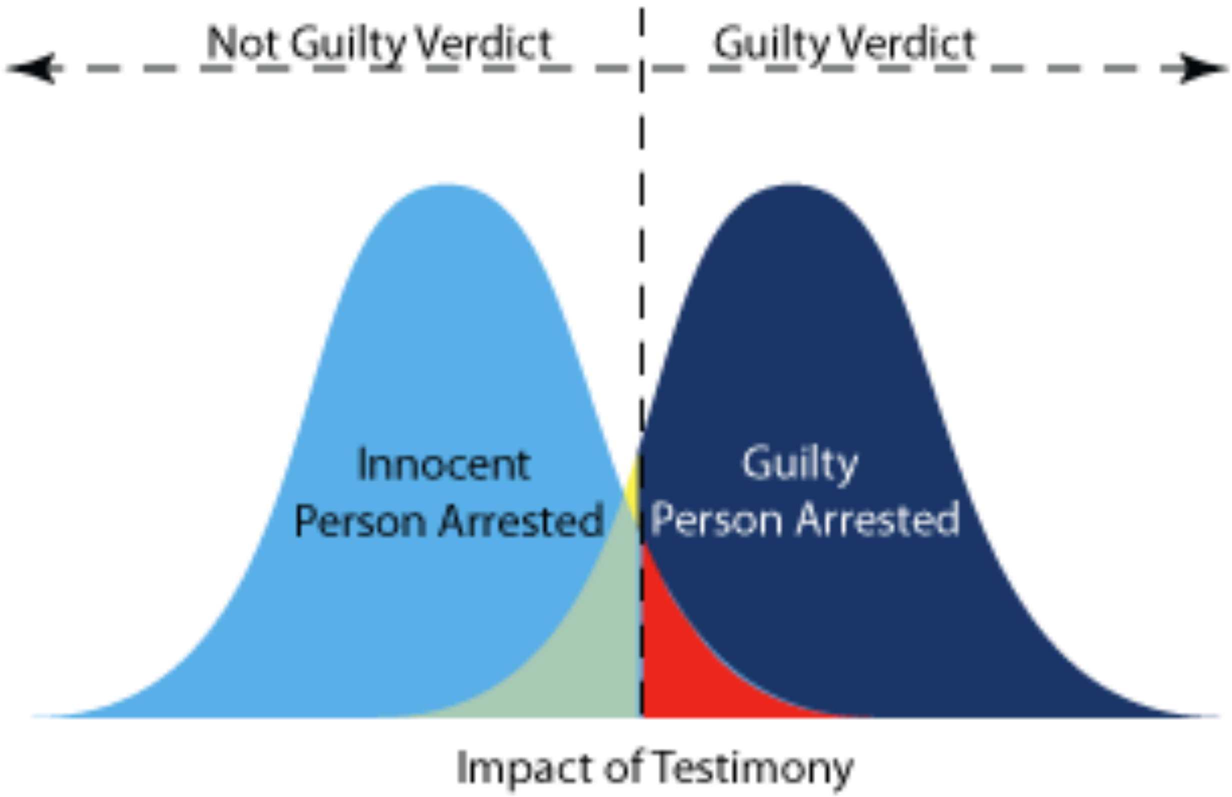
figure 5. The effects of increasing sample size or in other words, number of independent witnesses.

http://www.intuitor.com/statistics/T1T2Errors.html

- In general, type I errors are the ones that we are concerned with in biology. Although, there are circumstances when we are more concerned with type II errors (ie) Medicine

- There is a trade-off between type I error and type II error

More Hypothesis Testing



http://www.intuitor.com/statistics/T1T2Errors.html

Two clinical trials are carried out which both test the same null hypothesis under the same conditions with $\alpha = 0.05$. Trial A has 45 individuals and Trial B has 100 individuals.

Which of the following is true about the two trials described above:

a. Study A has higher probability of type I error than Study B and Study B has a higher probability of type II error than Study A

b. Study A has a lower probability of type I error than Study B and Study B has a lower probability of type II error than Study A

c. Study A has the same type I error as Study B and Study A has a higher probability of type II error than Study B

d. Study A has the same type I error as Study B and Study B also has a higher probability of type II error than Study A

Two clinical trials are carried out which both test the same null hypothesis under the same conditions with $\alpha = 0.05$. Trial A has 45 individuals and Trial B has 100 individuals.

Which study, **A** or **B**, has higher power?

Two independent studies are performed to test the same null hypothesis.

What is the probability that one or both of the studies obtains a significant result and rejects the null hypothesis ***even if the null hypothesis is true***? Assume that in each study there is a 0.05 probability of rejecting the null hypothesis.

____

a. **0.10**

b. **0.075**

c. **0.05**

d. **0.0975**

You can consider the previous question in one of two equally valid ways:

P(at least 1 study obtains significant results)

= 1-P(neither study obtains significant results)

$= 1 - (1-0.05)^2 = 0.0975$

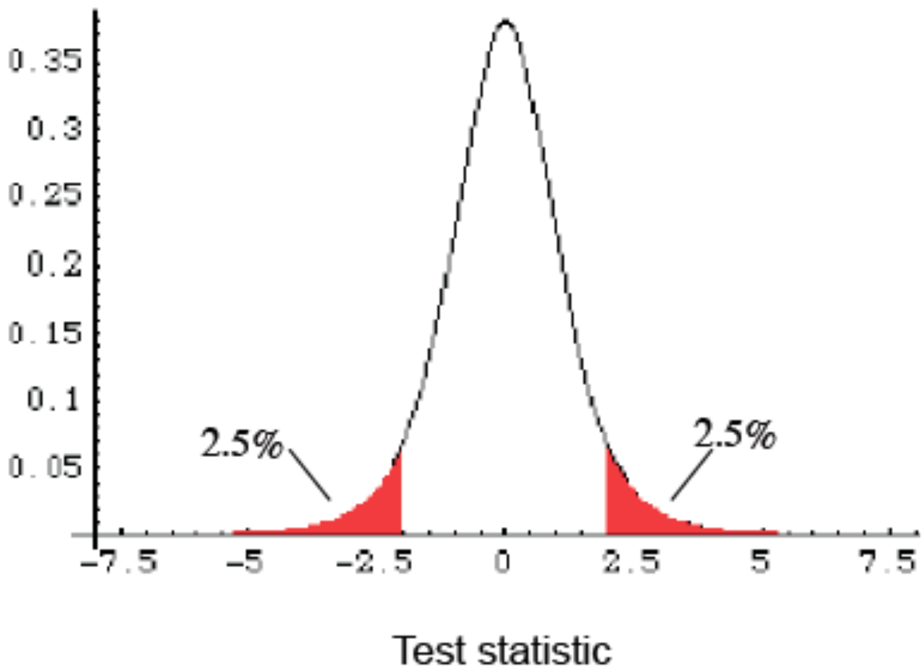
P[1st study significant OR 2nd study is significant]

$= (0.05)+(0.05) - (0.05)^2 = 0.0975$

One tailed and two tailed tests:

- Most tests are two-tailed tests
- This means that a deviation in either direction would reject the null hypothesis
  - this means that α is divided into $\frac{\alpha}{2}$ on the one side and $\frac{\alpha}{2}$ on the other

More Hypothesis Testing



One Tailed Tests:

Only used when the other tail is nonsensical

- **Example:**
  - Comparing grades on a multiple choice test to random guessing
  - Do dogs resemble their owners?

- **Example:**
  - Do daughters resemble their biological fathers?
    - Experiment involves a subject who examines photo of one girl and two adult men and guesses the father
    - If subjects pick father correctly > 0.5 then the hypothesis being tested would FTR
    - Wouldn’t make sense that daughters would, on average, resemble their biological fathers less than other men.

- Some parting words:

  – FTR does not mean ACCEPT
    - We ***never*** accept the null hypothesis

  – If FTR the null hypothesis, we can conclude that the data is compatible with the hypothesis

- <u>Why use hypothesis testing at all?</u>

  - Why don't we skip hypothesis testing since confidence intervals give us similar information **plus** gives us information about the actual magnitude of the parameter?

    - *Main purpose of hypothesis testing is to determine if sufficient evidence has been presented to support a scientific claim*

- Deploy these tools wisely
  - Just because something is statistically significant does not mean it is biologically important or interesting
  - **Almost any null hypothesis can be rejected with a large enough sample**

Some potential misconceptions about p-values

- If the result is statistically significant there is a temptation to believe that the effect is large. <u>DO NOT GIVE IN THIS TO ERRONEOUS BELIEF</u>.
  - Nor does it mean that the effect is interesting
  - If the sample size is large (and measurements have little variation) then even inconsequential differences will be significant
- P-values are calculated from the data itself. In contrast, the alpha value is set by the experimenter prior to conducting the experiment. P-values and alpha are related BUT THEY ARE NOT THE SAME!