Hypothesis Testing Cautions and Errors

Hypothesis Testing Steps

1. Assume the null hypothesis.

2. Find the probability of observing a sample at least as extreme as the sample you have if the null is true (the **p-value**).

3. If the probability is low enough (below the **significance level**), reject the null in favor of the alternative.

Cautions about *p*-values

The use of *p*-values has become more controversial in recent years due to how often they are either misused or misunderstood.

See, for example, this Nature editorial:

https://www.nature.com/articles/d41586-019-00874-8

Cautions about *p*-values

Important:

- *p*-values do not give the likelihood that the result is due to chance

- *p*-values only summarize the data, <u>assuming the null hypothesis is true</u>! They do not say how likely the result is to be true.

- *p*-values say nothing about the size of an effect. Statistical significance is not the same as *practical* significance.

- A low *p*-value does not <u>prove</u> the alternative. Ronald Fisher, the inventor of the *p*-value, only meant for "statistical significance" to be an informal index.

Cautions about *p*-values

Another easy mistake to make with *p*-values is the **multiple comparisons/multiple testing** problem. When doing many simultaneous comparisons across a dataset, the chances increase of seeing a "statistically significant" effect which is just due to random sampling error.

See this xkcd comic: https://xkcd.com/882/ or this FiveThirtyEight interactive:

https://fivethirtyeight.com/features/science-isnt-broken/#part1

Cautions about *p*-values

When doing hypothesis testing, it is important to distinguish between exploratory analysis and hypothesis testing.

Hypothesis testing must be deliberate, which a specific hypothesis in mind prior to looking at the data.

It is not valid to first look for potential effects in a dataset and then test those effects <u>using the same data</u>.

Hypothesis Testing - Types of Errors

		Reality	
		Null is True	Null is False
Our Decision	Do not Reject Null		
	Reject Null		

Hypothesis Testing - Types of Errors

		Reality	
		Null is True	**Null is False**
Our Decision	**Do not Reject Null**	Correct Decision	
	Reject Null		Correct Decision

NASHVILLE
SOFTWARE
SCHOOL

Hypothesis Testing - Types of Errors

		Reality	
		Null is True	**Null is False**
Our Decision	**Do not Reject Null**	Correct Decision	
	Reject Null	False Positive / Type I Error	Correct Decision

Hypothesis Testing - Types of Errors

How likely are we to make a Type I/False Positive error?

Note: A Type I error is only possible if the null is true.

Assuming that the null is true, how often will we commit a Type I error?

We reject the null hypothesis if the probability of a sample at least as extreme as ours is less than the significance level.

This happens with probability equal to the significance level (usually 5% of the time).

Hypothesis Testing - Types of Errors

		Reality	
		Null is True	**Null is False**
Our Decision	**Do not Reject Null**	Correct Decision	False Negative / Type II Error
	Reject Null	False Positive / Type I Error	Correct Decision

NASHVILLE SOFTWARE SCHOOL

Hypothesis Testing - Types of Errors

How likely are we to make a Type II/False Negative error?

Note: A Type II error is only possible if the null is false.

Assuming that the null is false, how often will we commit a Type II error?

This depends on the effect size - how far off the true parameter is from the parameter value assumed by the null.

It also depends on how large a sample size we are using. At a larger sample size, we are more likely to detect a difference.

The probability of rejecting the null when it is false is the **power** of the test and can be influenced by carefully selecting a sample size.