Significance Testing and Errors

Clicker Question

After taking a SAT review course, those who scored below 250 improved their scores by an average of 30 points. This improvement

A. Must be attributed to the quality of the review course
B. Probably represents the hot hand phenomenon
C. Is likely an instance of regression to the mean
D. Must be due to the students trying harder

Clicker Question

A study based on a sample of 100 UCSD students reported a difference in interest in partying between men and women (p<.01)

- This result is not reliable because of the small sample size
- This result is not reliable because of the small p-value
- There is less than 1 in 100 likelihood that there is a difference in the actual population
- There is less than 1 in 100 likelihood that the difference in the sample is due to chance

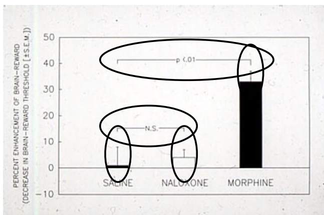
Clicker Question

A study showing a positive correlation between age and interest in the Beatles re[prts $p < .01$. This means

A. There is little correlation between age and interest in the Beatles
B. The correlation is highly important
C. The correlation has low significance
D. There is less than 1 in 100 chance that the correlation in the sample is due to chance

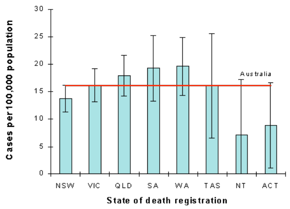
Showing Statistically Significant Differences with Error Bars

- Error bars can be used to identify 1 or more standard deviations above and below the mean
- If the error bars overlap, the difference is not statistically significant
- If they do not, the difference may be statistically significant



Showing Statistical (non)-Significance with Error Bars

- The bar graph to the right shows suicide rates of people between 15 and 24 in the different States and territories of Australia
- Error bars show 95% confidence interval
- No differences are statistically significant



Non-significant Difference versus No Difference

- If the difference in your sample is not significant, you conclude that you cannot tell whether there is actually a difference in the real population
  - There may be one, but the power of your test was too weak to find it
- **No significant difference does not mean there is no difference**
  - There may well be a difference, but one that has not been detected given the tests employed
  - All we can say is that we have not detected any difference

Clicker Question

If the attempt to find a difference in means based on a sample is reported to be non-significant, that means

- There is no difference between means in the actual population
- The probability that the null hypothesis was true was less than 5%
- No difference between the means in the actual population has been detected
- The result is not important

Caught Between Two Errors

Type I error (over confidence): Concluding there is a difference between means when there is none

- Use higher significance levels: instead of requiring only p<.05, require p<.01 or even p<.001

Type II error (false humility): Concluding there is no difference between means when there is one

- Use a larger sample, which has a greater chance of finding a significant difference if one is to be found

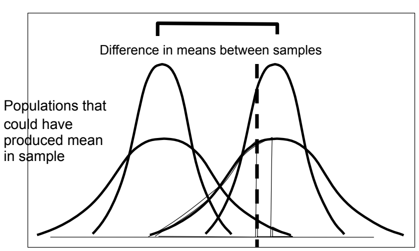
Two dangers - 2

	H_o is true	H_o is false
Did not reject H_o	**Correct failure to reject**	**Type II error (β)**
Did reject H_o	**Type I error (α)**	**Correct rejection**

α and β levels

- α-level is the probability of rejecting the null hypothesis when it is true
  - Statistical significance and p-value
- β-level is the probability of failing to reject the null hypothesis when it is false
  - (1- β) is probability that the researcher will correctly reject the null when the null is indeed false
  - The statistical *power* of the test

Two types of error - 2



Reduce type I error by increasing p-value

Increase Type II error

Increase sample size to reduce Type II error

Type I error Type II error

Clicker Question

Under what conditions should one focus more on reducing type II errors than type I errors?

A. When it is critical not to claim a difference when there isn't one
B. One should always be more concerned with type I errors
C. When it is critical not to miss a difference when there is one
D. When there is little worry about being wrong

Clicker Question

In which type of situation would you most likely expect that a Type II error has been committed?

- When the difference between means in a small or moderate-sized sample is not found to be statistically significant
- When an extremely large sample has been used
- When the difference between means in a sample has been found to be significant (p<.01)
- When the difference between means in an extremely large sample is not found to be statistically significant

Clicker Question

To reduce the likelihood of a Type II error, one should

- Always insist on using p-values <.01
- Not worry about the p-value and just look at the differences produced in the sample
- Use a large enough sample so that if there is a difference, it will produce a significant difference in the sample
- Use a small sample since then if there is a significant difference, there is likely to be a large difference in the real population

Science without Error?

One can reduce the risk of type I and type II errors to whatever level one desires if one

- insists on a small enough p-value
- uses a large enough sample

But one cannot eliminate the risk of error

- It is always possible that there is no difference in means despite obtaining a significant result in one's sample
- It is always possible that there is a real difference in means, but the difference in the sample is not significant

This is one more example of how scientific knowledge remains fallible!

Clicker Question

Is the following a good argument for confirming a correlational claim based on a sample:

If there is a difference between means in the population, the result in the sample will be statistically significant (p<.X)

The result in the sample is statistically significant (p<.X)

∴There is a difference between means in the population

- Yes, the argument is valid
- Yes, the argument is sound
- No, the argument affirms the consequent
- No, the argument denies the antecedent

The Logic of Correlational Research

To confirm or falsify a correlational claim based on a sample, we use *modus tollens*. The first premise in each case, though, is different

Confirming a correlational claim:

- If there is *no* difference between means in the population, then there will *not* be a statistically significant (p < ?) difference in my sample
- There is a statistically significant difference (p < ?) in means in my sample
- ∴There is a difference between means in the population

We pick the level of significance in the first premise according to how great a risk of error in our conclusion we can accept

The Logic of Correlational Research - 2

Falsifying a correlational claim

- If there is a *detectable* difference between means in the population, then there will be a statistically significant difference ($p < ?$) in my sample
- There is no statistically significant difference ($p < ?$) in means in my sample
- ∴There is no *detectable* difference between means in the population

The truth of the first premise depends upon using a large enough sample

NOTE: The conclusion refers to *DETECTABLE* differences