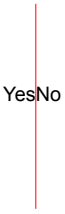
Lesson 36: Is hypothesis testing a yes/no question? Explain.



Is there a fine line with P-value?

No, but...
'The lower the P-value, the more comfortable we feel about our decision to reject the null hypothesis, but the null hypothesis doesn't get any more false.'

The question of whether the diabetes drug Avandia increased the risk of heart attack was raised by a study in the *New England Journal of Medicine*. This study estimated the seven-year risk of heart attack to be 28.9% and reported a P-value of 0.03 for a test of whether this risk was higher than the baseline seven-year risk of 20.2%. An earlier study (the ADOPT study) had estimated the seven-year risk to be 26.9% and reported a P-value of 0.27.

Why did the researchers in the ADOPT study not express alarm about the increased risk they had seen?

A P-value of 0.27 means that a heart attack rate at least as high as the one they observed could be expected in 27% of similar experiments even if, in fact, there were no increased risk from taking Avandia. That's not remarkable enough to reject the null hypothesis. In other words, the ADOPT study wasn't convincing.

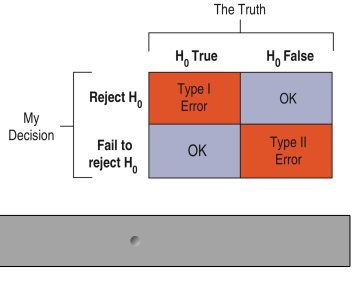
Alpha level (α): threshold for p-value. If our p-value falls below that point, we'll reject the null hypothesis. Common α levels are .10, .05, .01. You must choose an appropriate one for a given situation.

- The alpha level is also called significance level. When we reject the null hypothesis, we say the test is significant at that level.
- When we say that a test is statistically significant, we mean that the test statistic had a p-value lower than our α level.

TYPE I and TYPE II Errors

I. The null hypothesis is true, but we mistakingly reject it.

II. The null hypothesis is false, but we fail to reject it.



Example: Statistics Final Exam

H_0: the student has learned only 60% of the material

Type I	A student who learned 60% passes class
Type II	A student fails who knew enough to pass

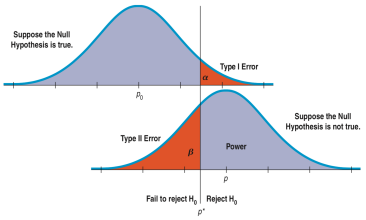
Probability(Type I error) = α level

Probability(Type II error) = β

we can reduce type II errors by increasing α but this would mean more Type I errors.

Power of the test is the probability that it correctly rejects a false null hypothesis = 1- β

A larger sample decreases the probability of type II error
-- increasing the power of the test.



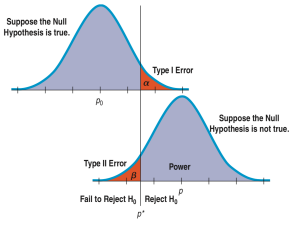
H_0: $p=p_0$	Then, we reject H_0 if $\hat{p} > p^*$ and make Type I error with P=α
H_A: $p>p_0$	$\hat{p} < p^*$, we fail to reject H_0

Power = 1- β

* Reducing α to lower the chance of Type I error will move p* to the right --increase β(prob. type II error) --reduce power

* Larger effect size (difference between p_0 and true p)
  - ⟶ smaller the chance of type II error
  - ⟶ greater the power of test

Can we reduce both Type I and Type II errors?



Yes. Make both curves narrower.
Reduce Standard deviations by increasing sample size. To half SD, quadruple sample size.

A company manufacturing computer chips finds that 8% of all chips manufactured are defective. Management is concerned that employee inattention is partially responsible for the high defect rate. In an effort to decrease the percentage of defective chips, management decides to offer incentives to employees who have lower defect rates on their shifts. The incentive program is instituted for one month. If successful, the company will continue with the incentive program.

1. Write the company's null and alternative hypotheses.
2. In this context describe a Type I error and the impact such an error would have on the company.
3. In this context describe a Type II error and the impact such an error would have on the company.
4. Based on the data they collected during the trial program, management found that a 95% confidence interval for the percentage of defective chips was (5.0%, 7.0%). What conclusion should management reach about the new incentive program? Explain.
5. What level of significance did management use?
6. Describe to management an advantage and disadvantage of using a 1% alpha level of significance instead.
7. Management decided to extend the incentive program so that the decision can be made on three months of data instead. Will the power increase, decrease, or remain the same?
8. Over the trial month, 6% of the computer chips manufactured were defective. Management decided that this decrease was significant. Why might management choose not to permanently institute the employee incentive program?