

# STA 321

# Spring 2014

## Lecture 18

*Thursday, April 3*

➤ **Hypothesis Tests**

# Significance Test

- A ***significance test*** is a way of statistically testing a hypothesis by comparing the data to values predicted by the hypothesis
- Data that fall far from the predicted values provide ***evidence against the hypothesis***

# Elements of a Significance Test

- Assumptions
- Hypotheses
- Test Statistic
- P-value
- Conclusion

# Assumptions

- What type of data do we have?
  - Qualitative or quantitative?
  - Different types of data require different test procedures
- What is the population distribution?
  - Is it normal? Symmetric?
  - Some tests require normal population distributions
- Which sampling method has been used?
  - We always assume simple random sampling
  - Other sampling methods are discussed in STA 675
- What is the sample size?
  - Some methods require a minimum sample size (like  $n=30$ )

# Hypotheses

- The ***null hypothesis*** ( $H_0$ ) is the hypothesis that we test (and try to find evidence against)
- The name null hypothesis refers to the fact that it often (not always) is a hypothesis of “no effect” (no effect of a medical treatment, no difference in characteristics of countries, etc.)
- The ***alternative hypothesis*** ( $H_a$ ) is a hypothesis that contradicts the null hypothesis
- When we reject the null hypothesis, the alternative hypothesis is judged acceptable
- Often, the alternative hypothesis is the actual research hypothesis that we would like to “prove” by finding evidence against the null hypothesis (proof by contradiction)

# Hypotheses

***The hypothesis is always a statement about one or more population parameters.***

# Test Statistic

- The ***test statistic*** is a statistic that is calculated from the sample data
- Often, the test statistic involves a point estimator of the parameter about which the hypothesis is stated
- For example, the test statistic may involve the sample mean or sample proportion if the hypothesis is about the population mean or population proportion

# P-Value

- How unusual is the observed test statistic when the null hypothesis is assumed true?
- The ***P-value*** is the probability, assuming that  $H_0$  is true, that the test statistic takes values at least as contradictory to  $H_0$  as the value actually observed
- The smaller the P-value, the more strongly the data contradict  $H_0$



# Conclusion

- In addition to reporting the P-value, a formal decision is made about rejecting or not rejecting the null hypothesis
- Most studies choose a cutoff of 5%.
- This corresponds to rejecting the null hypothesis for P-values smaller than 0.05.
- Smaller P-values provide more significant evidence against the null hypothesis
- “The results are significant at the 5% level”

# Elements of a Significance Test

- Assumptions
  - Type of data, population distribution, sample size
- Hypotheses
  - Null and alternative hypothesis
- Alpha-level (Type I error probability)
  - Specify alpha-level before looking at data
  - Alpha-level determines rejection region
- Test Statistic
  - Compares point estimate to parameter value under the null hypothesis
- P-value
  - Uses sampling distribution to quantify evidence against null hypothesis
  - Small P is more contradictory
- Conclusion
  - Report P-value
  - Rejection if test statistic in rejection region or  $P\text{-value} < \alpha\text{-level}$

# *P-Value*

- How unusual is the observed test statistic when the null hypothesis is assumed true?
- The ***p-value*** is the probability, assuming that  $H_0$  is true, that the test statistic takes values at least as contradictory to  $H_0$  as the value actually observed
- ***The p-value is not the probability that the hypothesis is true***
- The smaller the  $p$ -value, the more strongly the data contradict  $H_0$

# Alpha-Level

- Alpha-level (significance level) is a number such that one rejects the null hypothesis if the  $p$ -value is less than or equal to it.
- Often,  $\alpha=0.05$
- Choice of the alpha-level reflects how cautious the researcher wants to be
- Significance level  $\alpha$  needs to be chosen ***before*** analyzing the data

# Rejection Region

- The rejection region is a range of values such that if the test statistic falls into that range, we decide to reject the null hypothesis in favor of the alternative hypothesis

# Type I and Type II Errors

- Type I Error: The null hypothesis is rejected, even though it is true.
- Type II Error: The null hypothesis is not rejected, even though it is false.

# Type I and Type II Errors

- Terminology:
  - **Alpha** = Probability of a Type I error
  - **Beta** = Probability of a Type II error
  - **Power** =  $1 - \text{Probability of a Type II error}$
- The smaller the probability of Type I error, the larger the probability of Type II error and the smaller the power
- If you ask for very strong evidence to reject the null hypothesis, it is more likely that you fail to detect a real difference

# Type I and Type II Errors

- In practice, alpha is specified, and the probability of Type II error could be calculated, but the calculations are usually difficult
- **How to choose alpha?**
- If the consequences of a Type I error are very serious, then alpha should be small.
- For example, you want to find evidence that someone is guilty of a crime
- In exploratory research, often a larger probability of Type I error is acceptable
- If the sample size increases, both error probabilities decrease



# Power Calculations

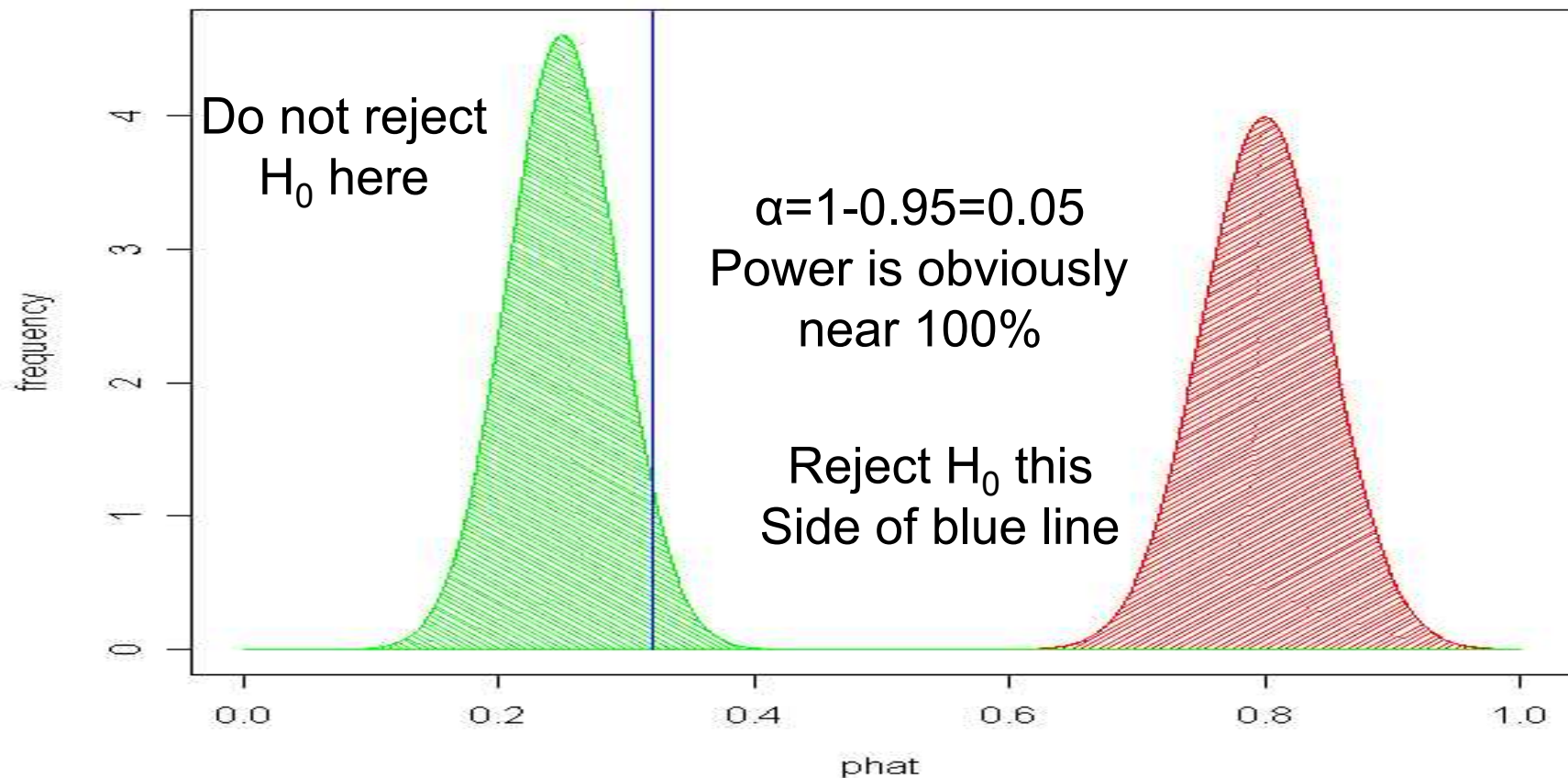
## Recall our ESP Example

- Hypothesis  $H_0 : p=0.25$  against  
 $H_1 : p>0.25.$
- The null distribution is normal with mean 0.25 and standard deviation  $\sqrt{0.25*0.75/n}$
- The cutoff is the  $1-\alpha$  percentile of this null distribution. We reject  $H_0$  if  $\hat{p}$  is above this cutoff.

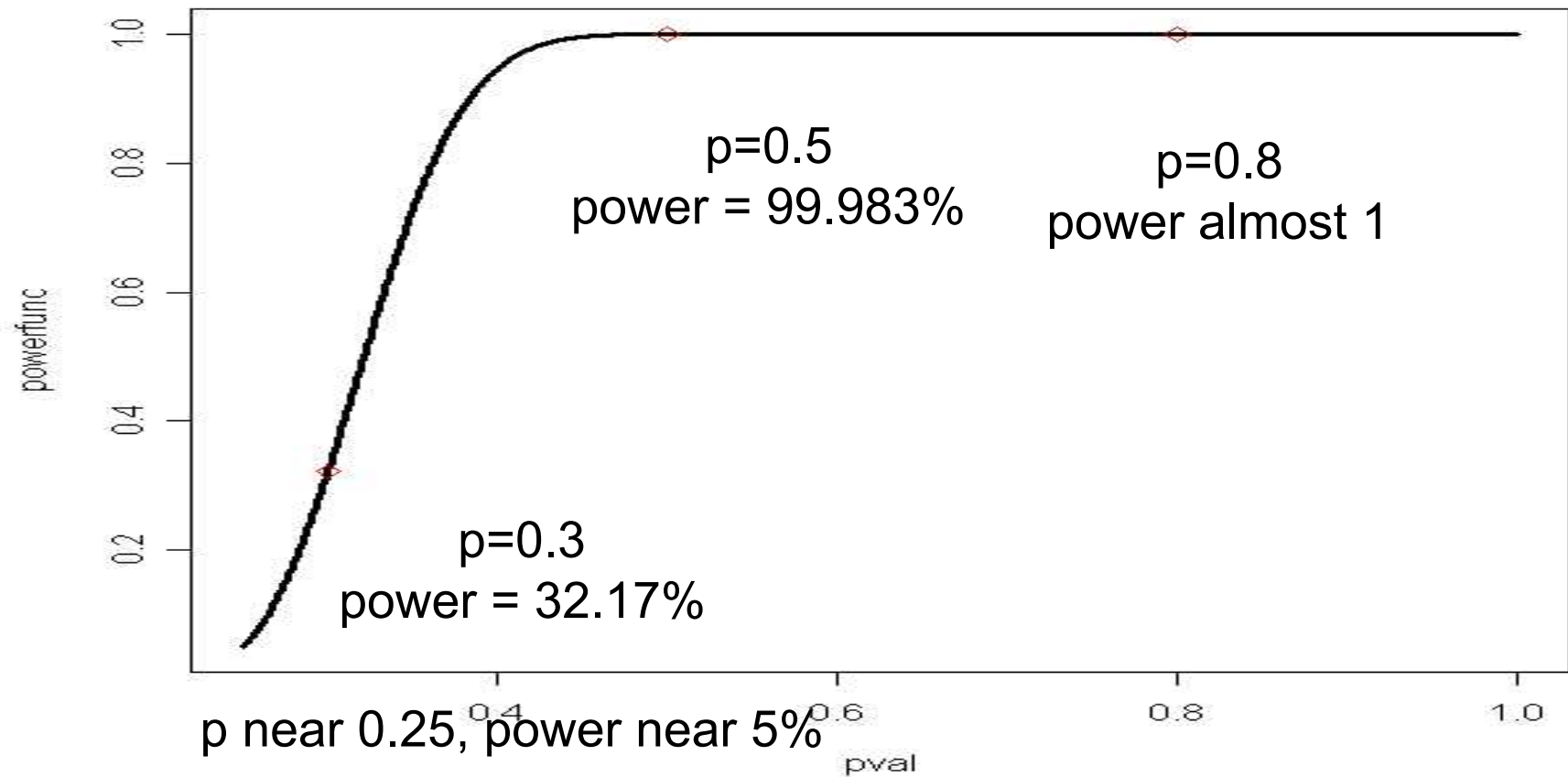
# Power Function

- What about the power of the test?
- Unfortunately under  $H_1$  we know nothing more about  $p$  than  $p > 0.25$
- BUT we can compute the power for *each*  $p > 0.25$

# What if the Alternative Is $H_1:p=0.8$ ?



# Power Function



# Questions...

- What happens to the power function when we use  $\alpha=0.001$  instead of  $\alpha=0.05$ ?
- What happens to the power function when the sample size  $n$  is increased?
- How can we be sure we get a particular amount of power in our experiment?

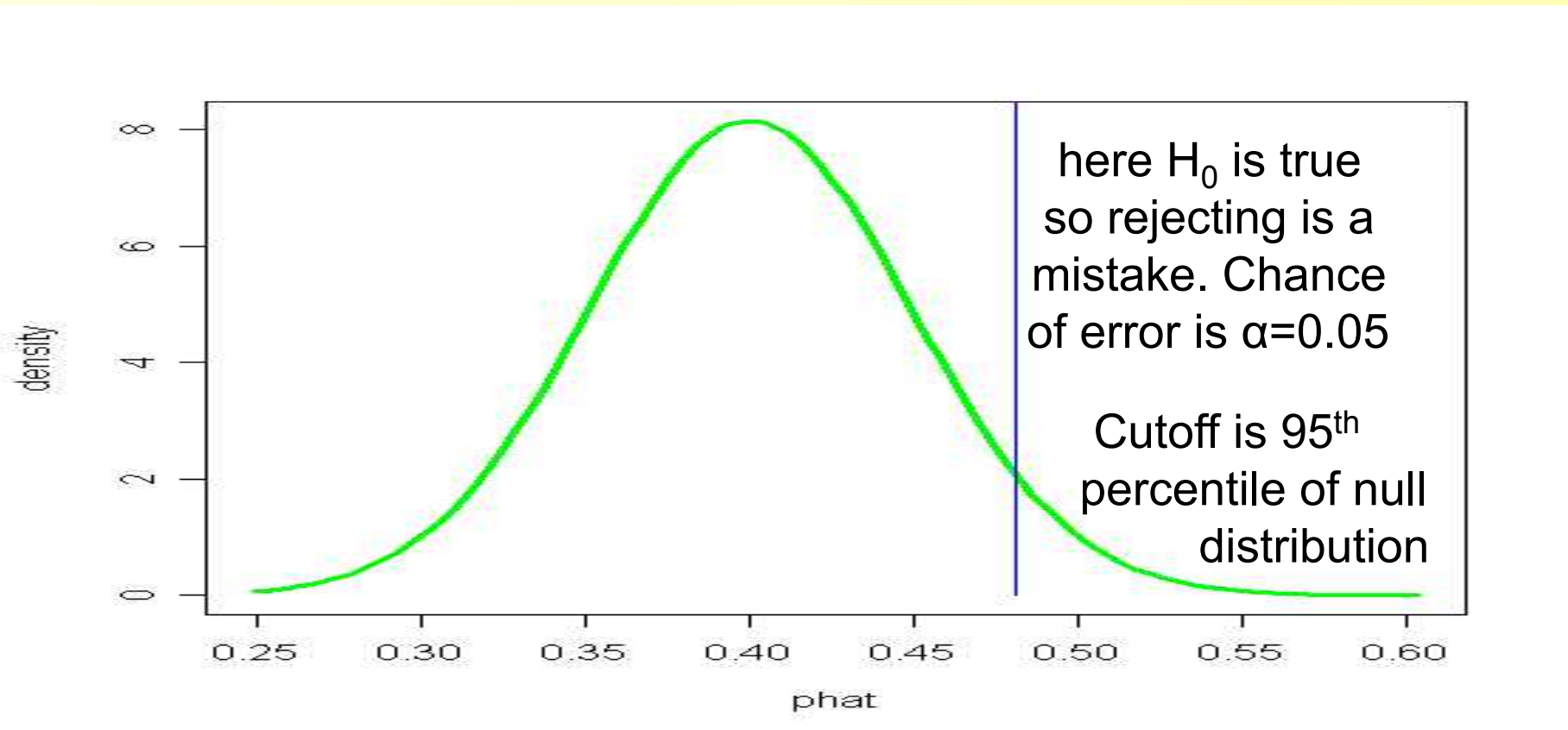
# Power – Why Important?

- Example
  - A pharmaceutical company knows that the old treatment for a disease cures 40% of the people.
  - They hope their new treatment is better.
  - They hope they can get the cure rate to 45%.

# Power Example Continued

- Our hypothesis test will test the null hypothesis  $H_0 : p=0.4$  (the old treatment proportion) against  $H_1 : p>0.4$  (we want our treatment to do better, hence this alternative).
- We intend to give the treatment to 100 people, and using  $\alpha=0.05$
- Cutoff?
- 95<sup>th</sup> percentile of the null distribution.
- $Z=1.64$ , thus  $Y = (0.049)(1.64) + 0.4 = 0.4804$

# Null Distribution Centered at $p_0=0.4$

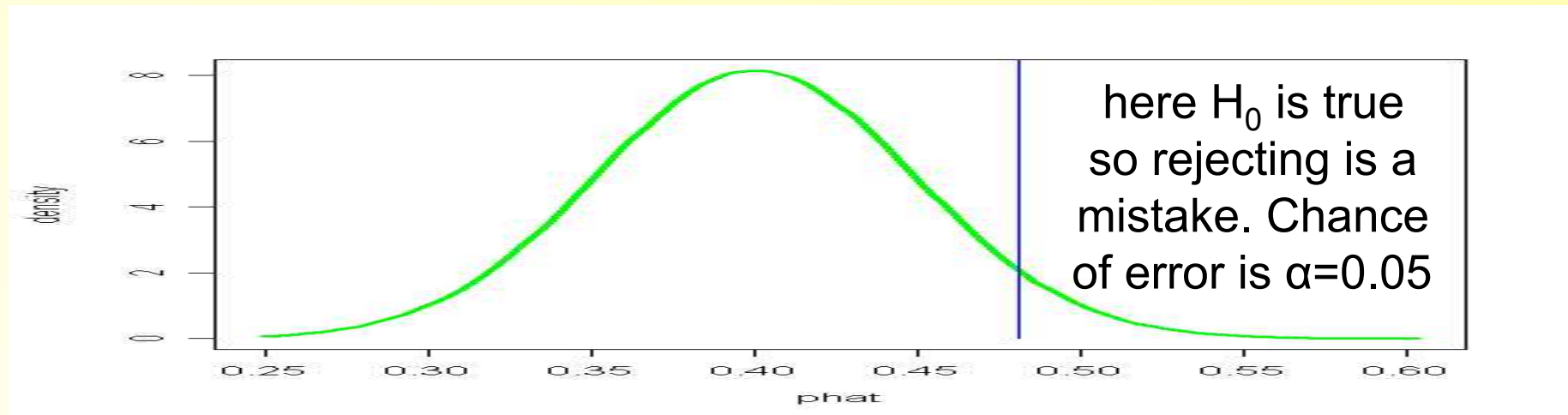




## Can this Experiment Find Anything?

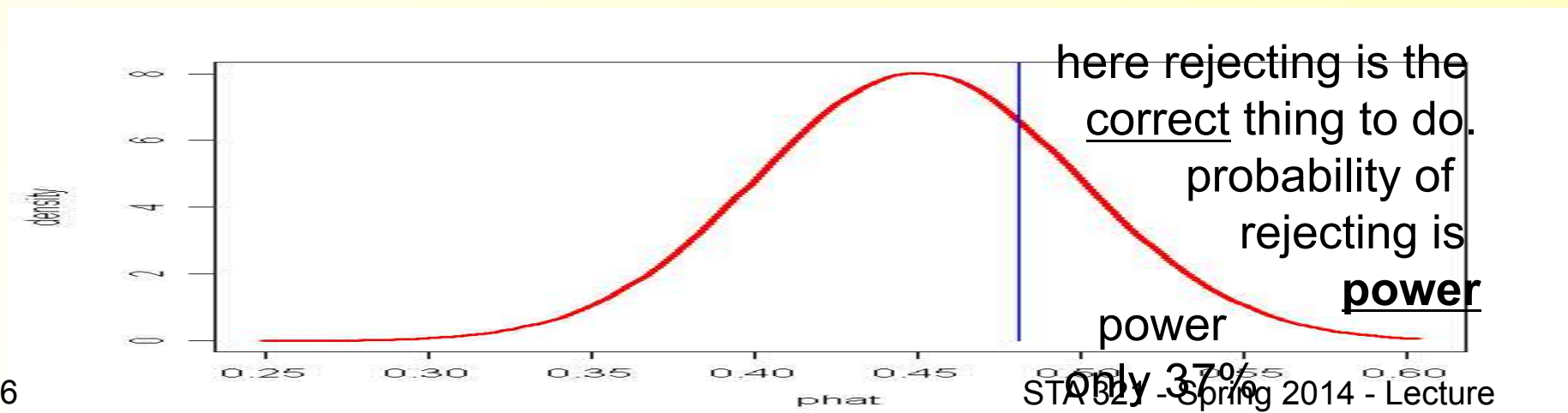
- We only are **guessing** our treatment can get the cure rate to 45%.
- What is the power for 45%?
- Remember, the power is the chance that, **when  $p=0.45$** , we reject  $H_0$  (the right decision in that case).
- We reject when  $\hat{p} > 0.4804$ .
- Thus, we need the probability that  $\hat{p}$  is greater than 0.4804, **given**  $p=0.45$ .

Green curve is distribution for  $p=0.4$   
 Red curve is distribution for  $p=0.45$



do not reject this side of blue line

reject this side of blue line



# What This Means...

- Suppose our treatment works (this is the assumption under which the power is calculated).
- Then we only have a 37.09% chance of getting a “reject  $H_0$ ” conclusion.
- That is not great. We could have a beneficial treatment and miss it.
- Solution – choose a higher sample size.

# Power Equation

- We need to solve for the  $n$  that satisfies

$$\frac{0.4899(1.645)}{\sqrt{n}} + 0.4 = \frac{0.4975(-1.28)}{\sqrt{n}} + 0.45$$

$$\frac{1.4427}{\sqrt{n}} = 0.05$$

- $n=832.5$ , so  $n$  must be at least 833