STA 570 Spring 2011

Lecture 17 *Thursday, March 24*

- **Significance Tests for Mean and Proportion**
- **Power Calculations**

Type I and Type II Errors

- Terminology:
  - ***Alpha*** = Probability of a Type I error
  - ***Beta*** = Probability of a Type II error
  - ***Power*** = 1 – 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

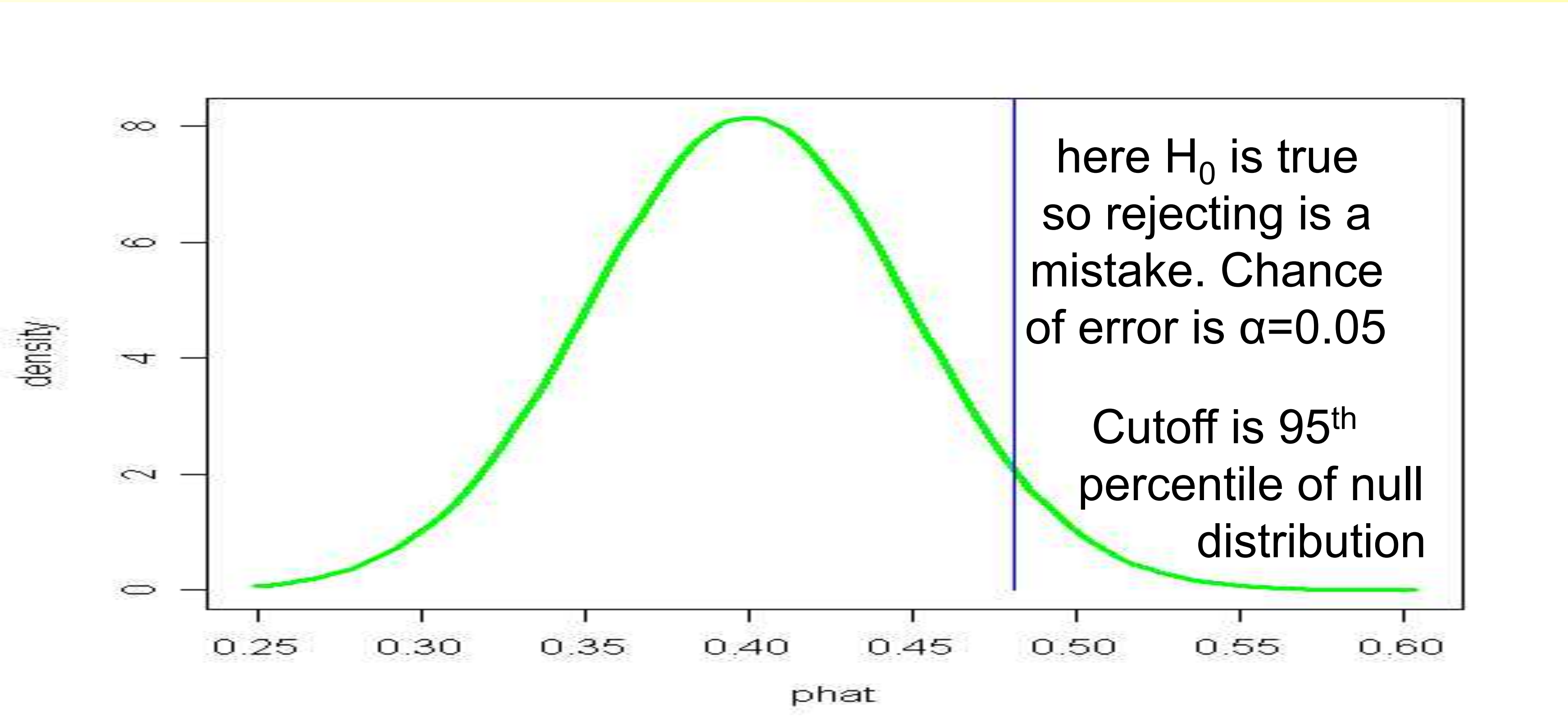
Power – Why Important?

- 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 α=0.05
- Cutoff?
- 95th 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



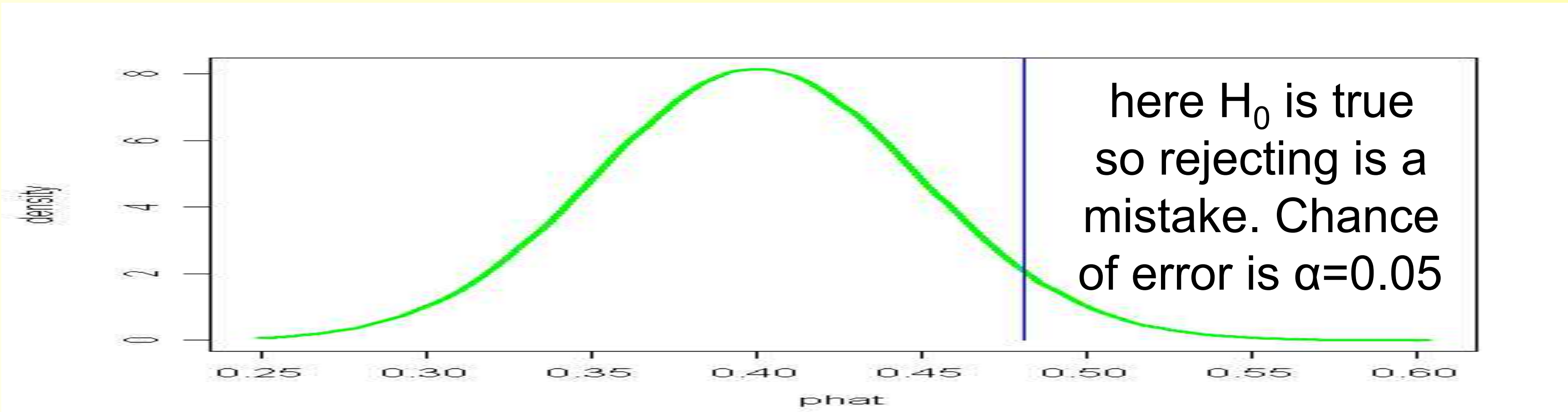
do not reject this side of blue line

reject this side of blue line

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 p-hat>0.4804.
- Thus, we need the probability that p-hat is greater than 0.4804, ***given*** p=0.45.

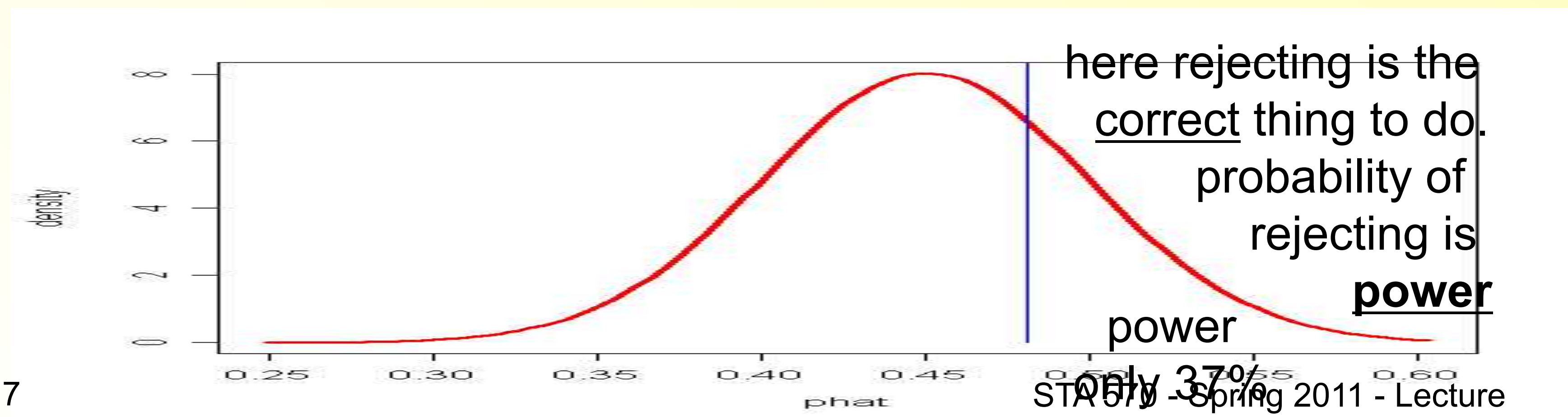
Green curve is distribution for p=0.4
Red curve is distribution for p=0.45



here H_0 is true so rejecting is a mistake. Chance of error is α=0.05

do not reject this side of blue line

reject this side of blue line



here rejecting is the <u>correct</u> thing to do. probability of rejecting is **<u>power</u>**

power only 37%

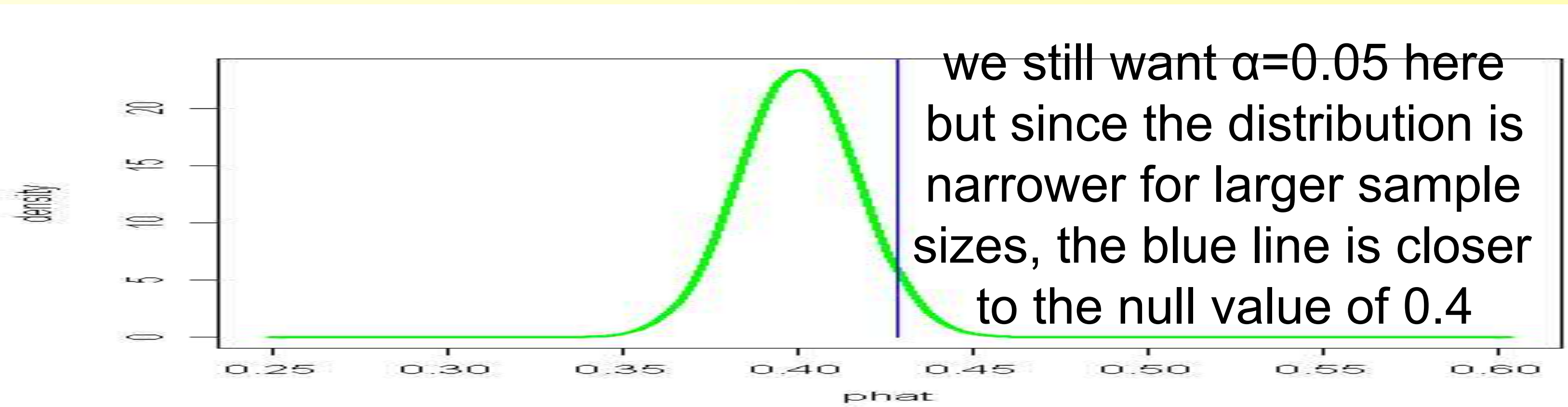
What This Means…

- 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.

How Does a Larger Sample Size Help?

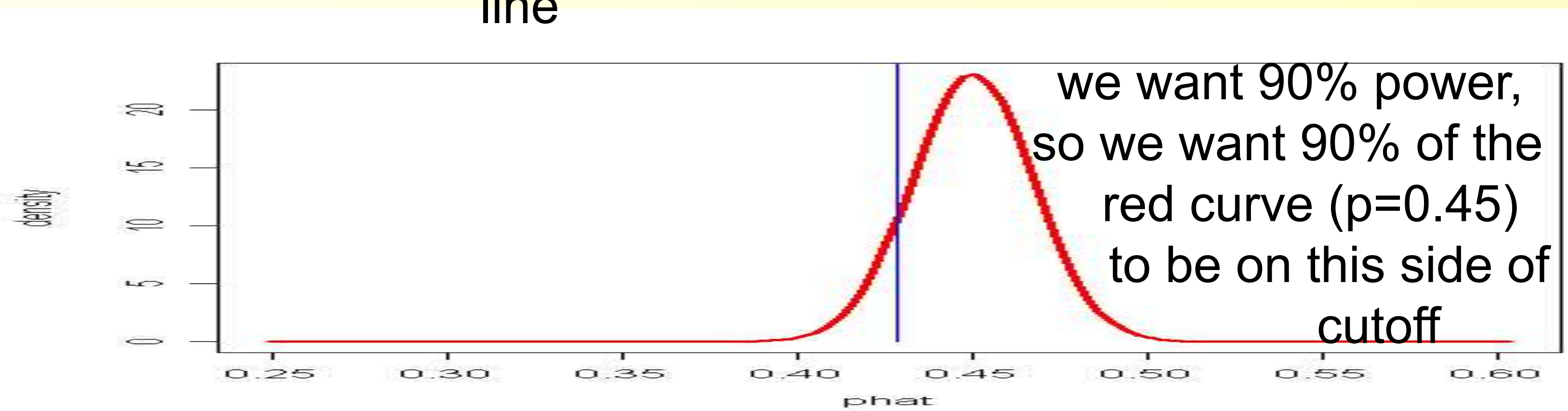
- A larger sample size reduces the standard deviations of both distributions, making them overlap less.
- Because the null distribution is narrower, the cutoff gets closer to 0.4, the null value.
- In fact, given a particular power, we can find an appropriate sample size.
- Suppose we want 90% power for p=0.45

Our goal – to get 90% power for p=0.45

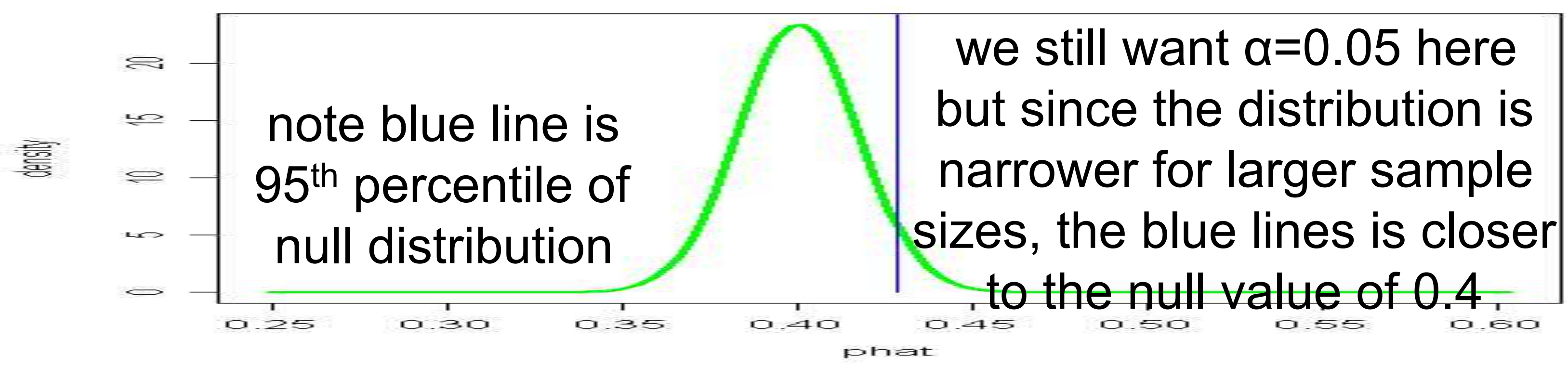


do not reject this side of blue line

reject this side of blue line

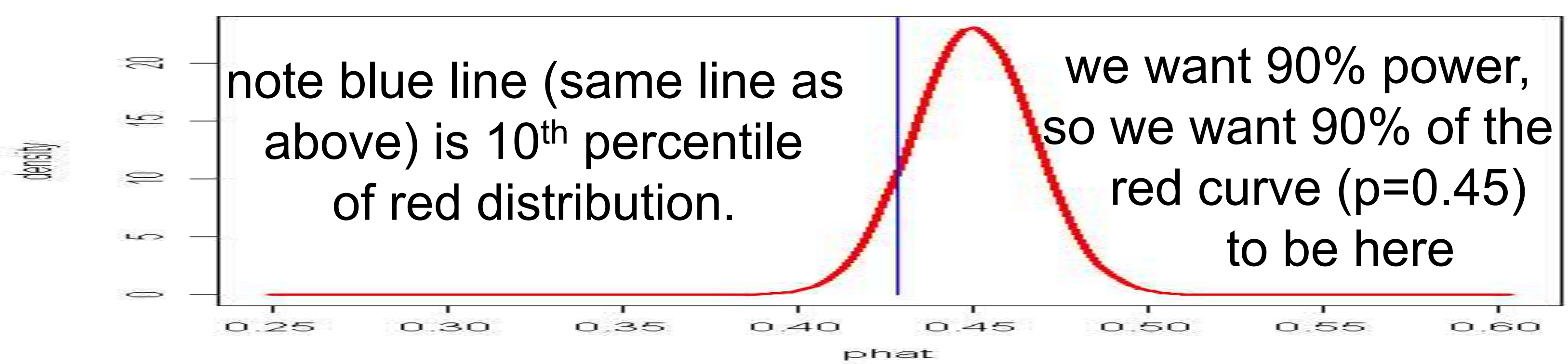


Our Goal – Get 90% Power for p=0.45



do not reject this side of blue line

reject this side of blue line



What Sample Size Should We Choose?

- The previous graph shows:
  - We need to equate the 95th percentile (**Z=1.64**) of the null distribution (using p=0.40) to the 10th percentile (**Z=-1.28**) of the distribution using p=0.45
  - With p=0.40, the standard deviation of p-hat is sqrt(0.40*0.60/n) = 0.4899/sqrt(n)

    p-hat ~ N(0.40, 0.4899/sqrt(n))
  - With p=0.45, the standard deviation of p-hat is sqrt(0.45*0.55/n) = 0.4975/sqrt(n)

    p-hat ~ N(0.45, 0.4975/sqrt(n))

The dreaded power equation

- With p=0.4, phat ~ N(0.40, 0.4899/sqrt(n))
- With p=0.45, phat ~ N(0.45, 0.4975/sqrt(n))
- We are equating the 95th percentile of a distribution (Z=1.64) to the 10th percentile of another distribution (Z=(-1.28))
- The 95th percentile when p=0.40 is
  (0.4899/sqrt(n))*(1.64) + 0.40
- The 10th percentile when p=0.45 is (0.4975/sqrt(n))*(-1.28) + 0.45

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

Cheating….

- If you want to cheat, you could think “can’t I just test 100 people, and if I don’t find anything go to 200, and if I don’t find anything go to 300, etc.”
- This is called “testing to a foregone conclusion”, and is not good science.
- You run a much bigger chance of a type I error than you claim to be.
- The way to do this right is called “sequential analysis” (not studied in this course).

Another example with a 2-sided alternative

- One of the most basic tests of a random number generator is that it produces numbers in the right frequency (more complicated tests involve verifying no discernible pattern is present in the numbers)
- One random number generator is supposed to produces successes and failures in the ratio of 20% successes and 80% failures.
- Thus, H_0 : $p=0.2$ and H_1 : $p \neq 0.2$, since being off in either direction is bad.

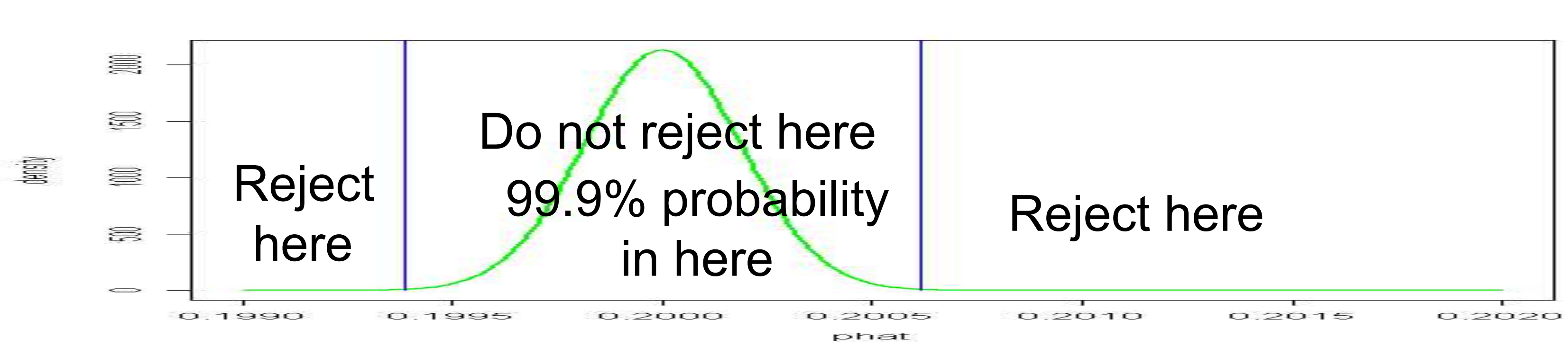
Example setup, continued

- We want to be quite sure of our random number generator (plus, generating large samples is cheap), so we want $\alpha=0.001$ and 98% power for $p=0.201$. Thus, we want to be quite likely to determine even a small difference from $p=0.2$.
- The resulting sample size will be quite large.

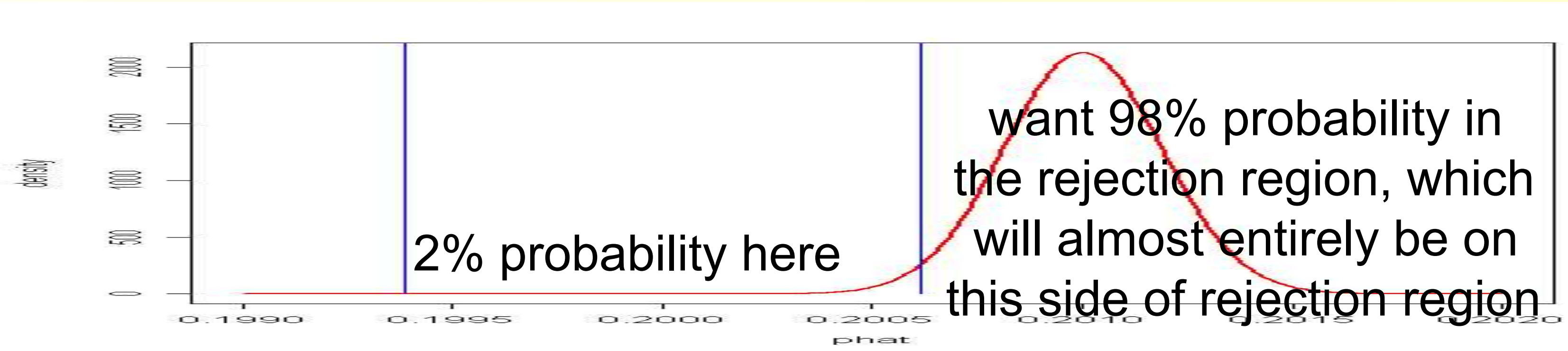
Null and Alternative distributions

- The null distribution ($p=0.2$) is normal with mean 0.2 and standard deviation sqrt(0.2*0.8/n) = 0.4/sqrt(n)
- The alternative distribution ($p=0.201$) is normal with mean 0.201 and standard deviation sqrt(0.201*0.799/n) = 0.400748/sqrt(n)
- We have a 2-sided alternative, so we reject in both directions from $p_0=0.2$

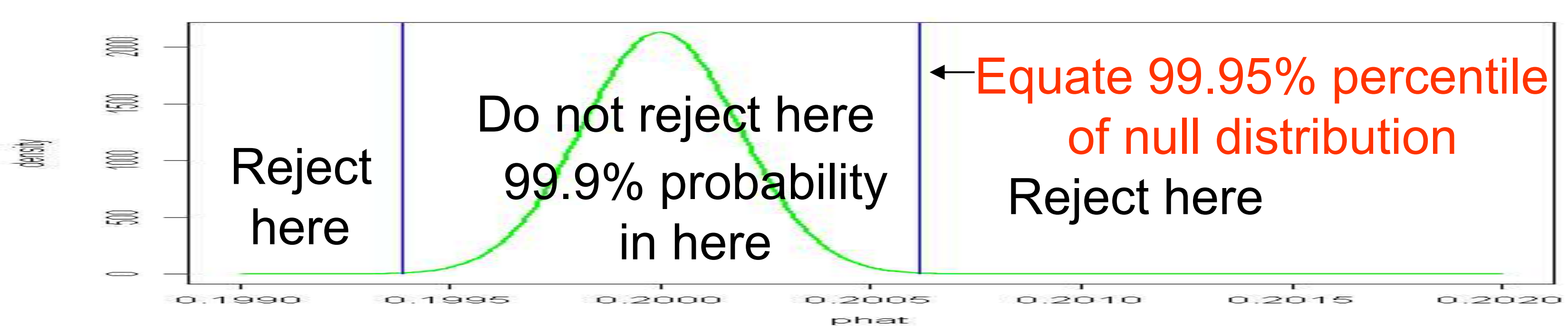
We want α=0.001 and 98% power at p=0.201



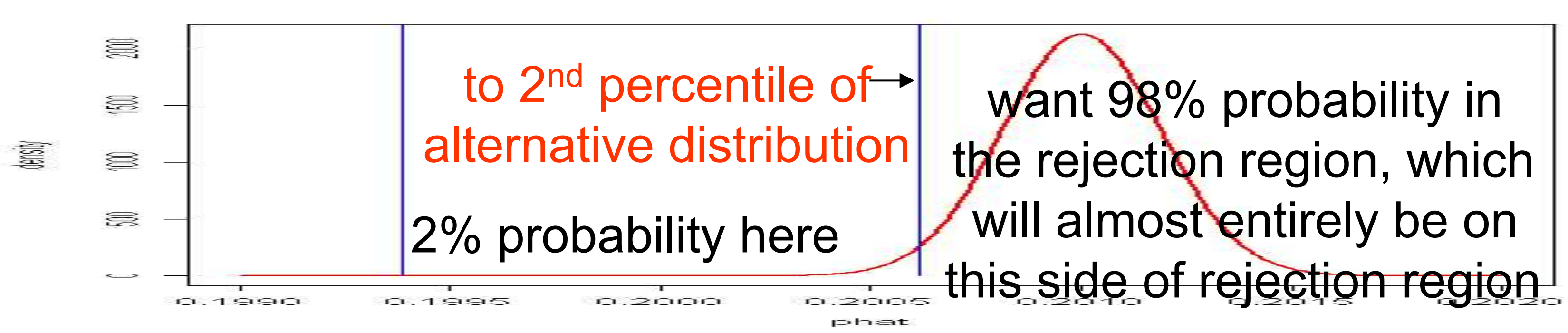
Only the right cutoff really matters for the power calculation (this is the cutoff in the direction from p=0.200 to p=0.201



We want α=0.001 and 98% power at p=0.201



Only the right cutoff really matters for the power calculation (this is the cutoff in the direction from p=0.200 to p=0.201



Power equation

- We need to equate the 99.95% percentile of the null distribution to the 2nd percentile of the alternative distribution.
- The 99.95% percentile corresponds to Z=3.29 (nearby values acceptable). For the null distribution, this Z corresponds to (3.29)(0.4/sqrt(n)) + 0.20 = **0.200 + 1.316/sqrt(n)**
- The 2nd percentile corresponds to Z=(-2.05). For the alternative distribution, this Z corresponds to (-2.05)(0.400748)/sqrt(n) + 0.201 = **0.201 - 0.82153/sqrt(n)**

Solving for n

- We equate **0.200+1.316/sqrt(n) = 0.201-0.82153/sqrt(n)**
- 2.13753/sqrt(n)=0.001
- 2137.53 = sqrt(n)
- n = 4,569,035
- This is huge, but remember we want virtually certainty of detecting a very small difference.

My experimental data

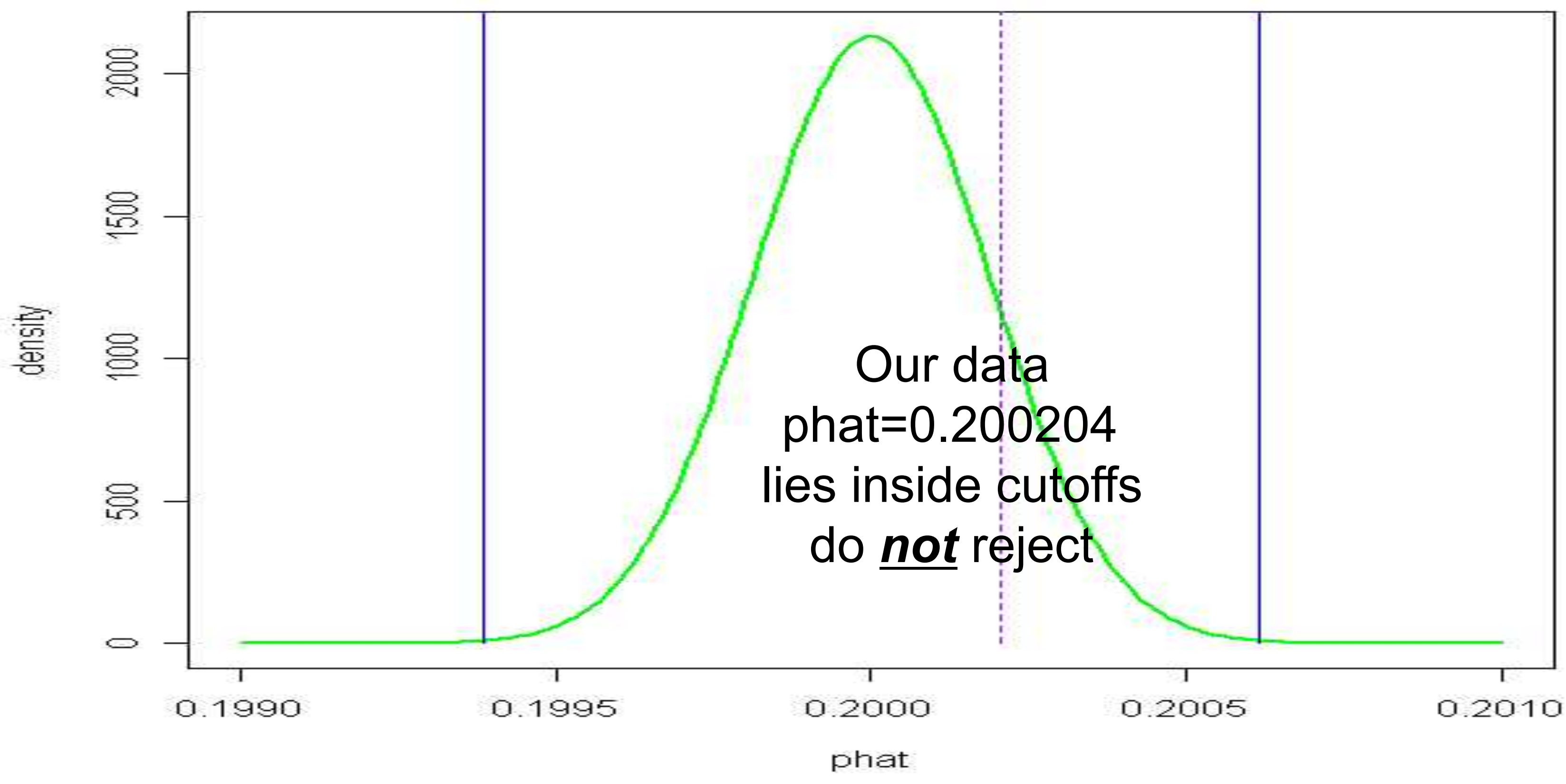
- As I said, computer random number are cheap, so we generated 5,000,000 of them, where we are supposed to get p=0.2
- I “rolled” 1,001,020 successes, which resulted in phat=1001020/5000000 = 0.200204
- Let’s conduct the hypothesis test of H_0 : p=0.2 and H_1 : p≠0.2 and determine the result

The null distribution

- The null distribution is normal with mean 0.2 and standard deviation sqrt (0.2*0.8/5000000)=0.0001789
- We are using α=0.001 and conducting a 2-sided test, so we are looking for the 0.05% and 99.95% percentiles, corresponding to Z=(±3.29).
- These percentiles are

(-3.29)(0.0001789) + 0.2 = 0.1994 and (3.29)(0.0001789) + 0.2 = 0.2006

Reject outside (0.1994,0.2006)



Conclusion

- So it appears our random number generator in R looks ok.
- As mentioned earlier, this is the EASIEST test for a random number generator to pass.
- Other tests involve trying to avoid patterns, which are not supposed to be present in random data.

Review of possible situations

- We always have a hypothesis of the form $H_0 : p=p_0$. In a sample size calculation, we are given an α, the alternative hypothesis, and a desired power POW for a specified p_1 in the alternative distribution.
- For example, "we are testing $H_0 : p=0.6$ against $H_1 : p>0.6$. We want to use $\alpha=0.05$ and achieve 90% power at $p=0.7$."

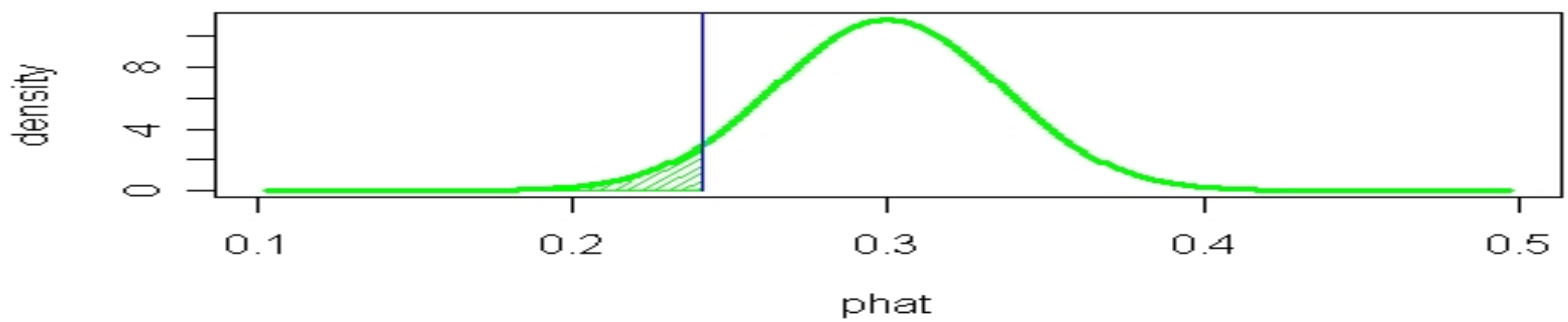
Review of possible situations

- There are three possible alternative hypotheses, $H_1 : p<p_0$, $H_1 : p>p_0$, or $H_1 : p \neq p_0$.
- For $H_1 : p \neq p_0$, the alternative value p_1 may be either less than p_0 or greater than p_0.
- Thus, we have four possible scenarios.
- In each scenario, power is computed by equating a particular percentile of the null distribution to a percentile of the alternative distribution.

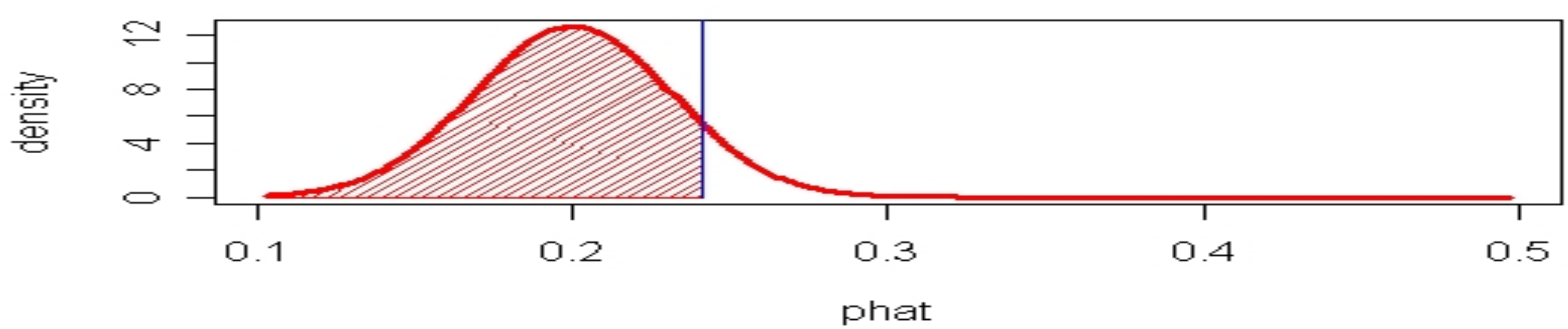
Review of possible situations

- The null distribution is always normal with mean p_0 and standard deviation $\mathrm{sqrt}(p_0(1-p_0)/n)$.
- The alternative distribution is always normal with mean p_1 and standard deviation $\mathrm{sqrt}(p_1(1-p_1)/n)$

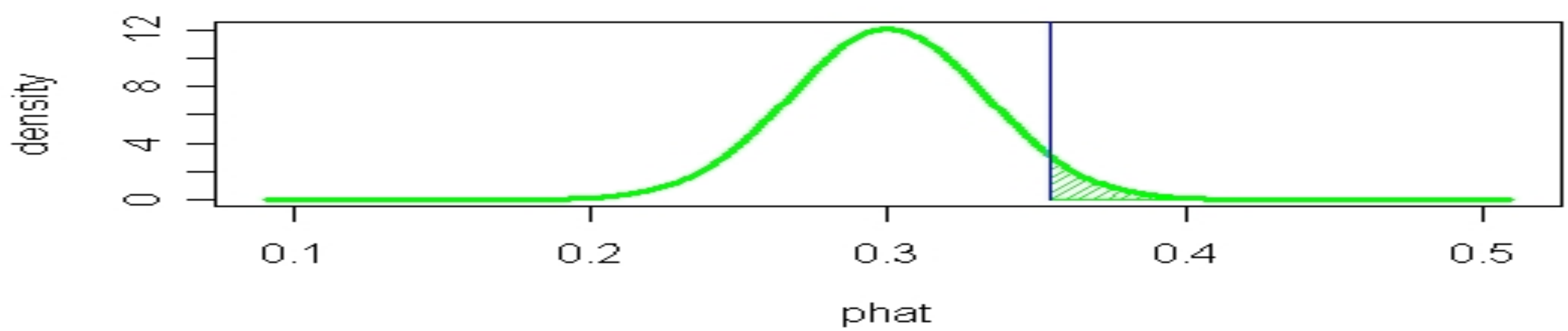
$H_0 : p=p_0, H_1 : p<p_0$



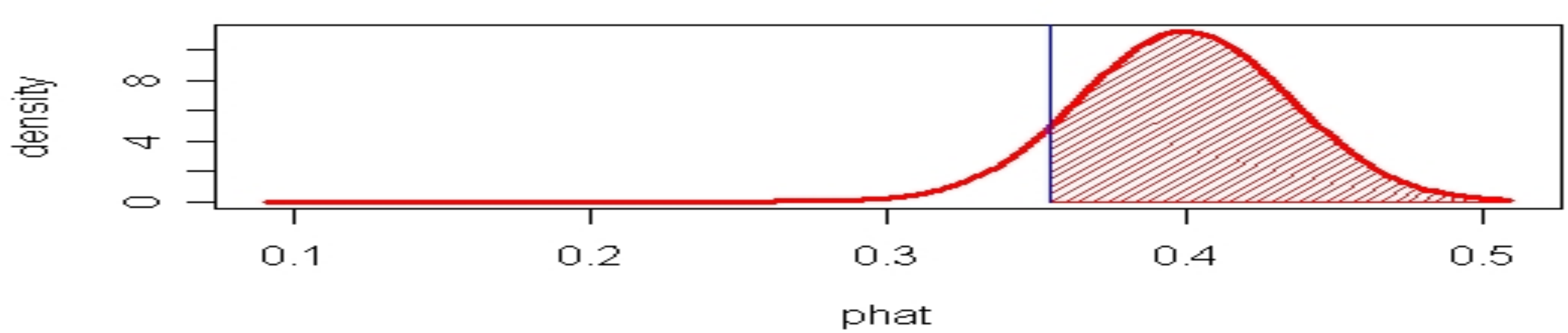
Equate the α percentile of the null to the POW percentile of the alternative



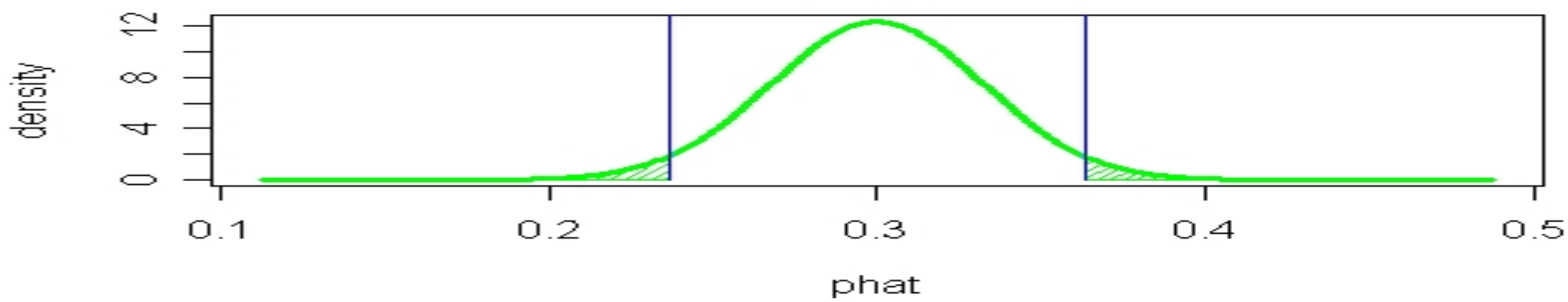
$H_0 : p=p_0, H_1 : p>p_0$



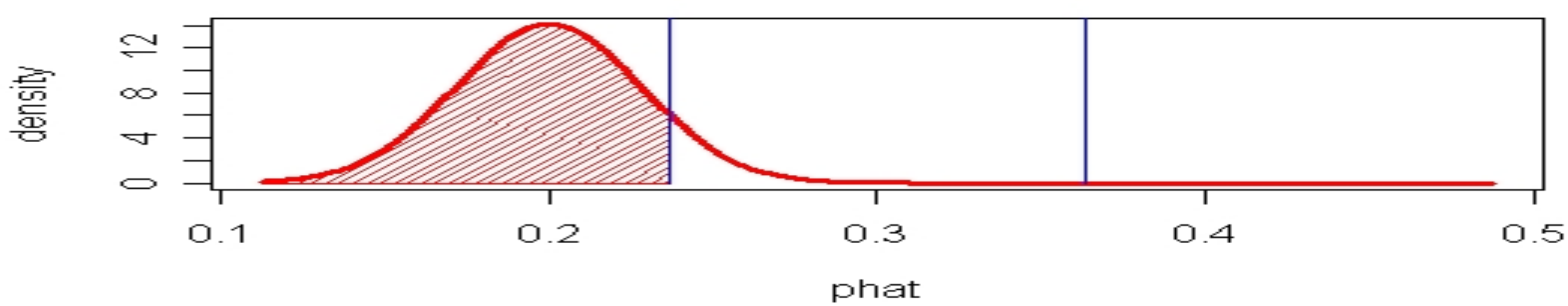
Equate the 1-α percentile of the null to the 1-POW percentile of the alternative



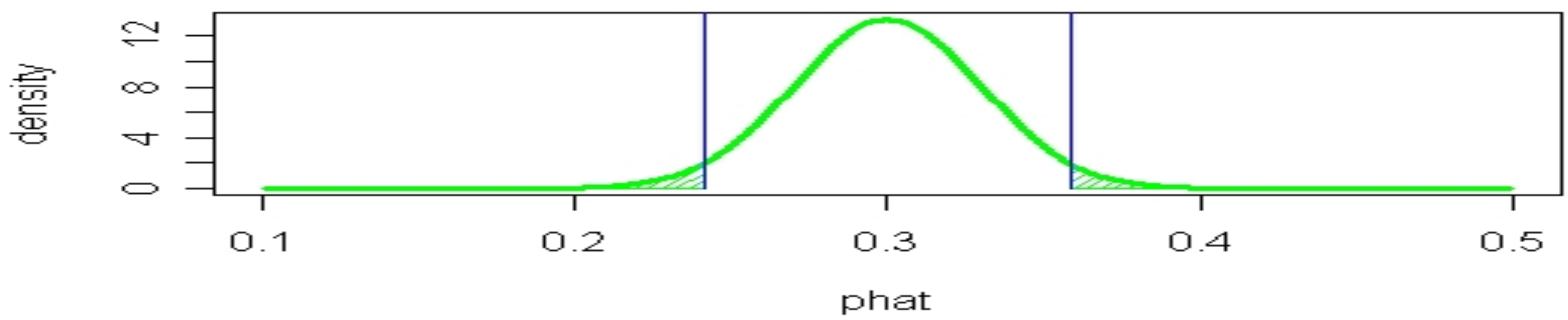
$H_0 : p=p_0$, $H_1 : p\neq p_0$, with $p_1<p_0$



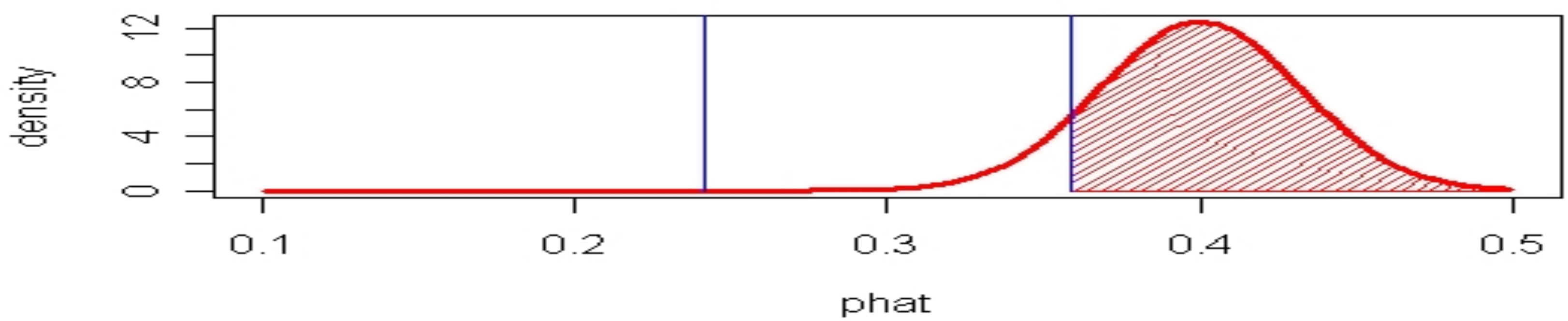
Equate the α/2 percentile of the null to the POW percentile of the alternative



$H_0 : p=p_0$, $H_1 : p \neq p_0$, with $p_1 > p_0$



Equate the 1-(α/2) percentile of the null to the 1-POW percentile of the alternative



A common thread

- Every situation requires equating a percentile of the null distribution to a percentile of the alternative distribution.
- Let z_0 be the z-score for the percentile required for the null distribution, and let z_1 be the z-score for the percentile required for the alternative distribution.
- Note the standard deviation for the null distribution is $\text{sqrt}(p_0(1-p_0)/n) = \text{sqrt}(p_0(1-p_0))/\text{sqrt}(n)$. Let $s_0=\text{sqrt}(p_0(1-p_0))$, thus the standard deviation of the null is $s_0/\text{sqrt}(n)$.
- Similarly, define $s_1=\text{sqrt}(p_1(1-p_1))$, so the standard deviation of the alternative is $s_1/\text{sqrt}(n)$.

General Formula

- We need to equate a percentile of the null distribution to a percentile of the alternative distribution.
- For the null, we need $p_0 + z_0 (s_0/\text{sqrt}(n))$
- For the alternative, we need $p_1 + z_1 (s_1/\text{sqrt}(n))$
- Thus we need to solve the equation below for n

$$p_0 + z_0\left(\frac{s_0}{\sqrt{n}}\right) = p_1 + z_1\left(\frac{s_1}{\sqrt{n}}\right)$$

A general formula

$$p_0 + z_0\left(\frac{s_0}{\sqrt{n}}\right) = p_1 + z_1\left(\frac{s_1}{\sqrt{n}}\right)$$

$$p_0 - p_1 = \frac{z_1 s_1 - z_0 s_0}{\sqrt{n}}$$

$$\boxed{n = \left(\frac{z_1 s_1 - z_0 s_0}{p_0 - p_1}\right)^2}$$

Choose the next larger integer.

Sample Size: Test for a Proportion

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : p = p_0$		
Research Hypothesis	$H_1 : p < p_0$	$H_1 : p > p_0$	$H_1 : p \neq p_0$
Sample Size	$\left(\dfrac{z_{1-\beta} s_1 + z_{1-\alpha} s_0}{p_0 - p_1} \right)^2$		$\left(\dfrac{z_{1-\beta} s_1 + z_{1-\alpha/2} s_0}{p_0 - p_1} \right)^2$

z_p = pth quantile of standard normal

Sample Size: Test for a Mean

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0 : \mu = \mu_0$		
Research Hypothesis	$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
Sample Size	$\left(\sigma \dfrac{z_{1-\beta} + z_{1-\alpha}}{\mu_0 - \mu_1} \right)^2$		$\left(\sigma \dfrac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1} \right)^2$

z_p = pth quantile of standard normal

Example

- We want to test H_0 : p=0.7 against H_1 : p>0.7.
- We want to use α=0.01 and achieve 95% power when p=0.80.
- What is the minimum required sample size *n*?
- We want to equate the 1-α=0.99 percentile of the null distribution to the 1-POWER=0.05 percentile of the alternative distribution.
- For the formula: $z_{1-\alpha}$ =2.33 and $z_{1-\beta}$=1.64.
- Also s_0=sqrt(p_0(1-p_0))=sqrt(0.7*0.3)=0.4583 and s_1=sqrt(p_1(1-p_1))=sqrt(0.8*0.2)=0.4000.

Example of using the formula

$$n = \left(\frac{z_{1-\beta} s_1 + z_{1-\alpha} s_0}{p_0 - p_1} \right)^2$$

$$n = \left(\frac{(1.64)(0.4000) + (2.33)(0.4583)}{0.7 - 0.8} \right)^2 = 297.2$$

Thus n must be at least 298.

Significance Test for a Mean

Example

- The mean score for all high school seniors taking a college entrance exam equals 500.
- A study is conducted to see whether a different mean applies to those students born in a foreign country.
- For a random sample of 100 of such students, the mean and standard deviation on this exam equal 508 and 100.
- Q: Is mean significantly different from 500 for international students?

Significance Test for a Mean

Assumptions

- What type of data?
  - *Quantitative*
- What is the population distribution?
  - *No special assumptions.*
  - *The test refers to the population mean of the quantitative variable.*
- Which sampling method has been used?
  - *Random sampling*
- What is the sample size?
  - *Minimum sample size of n=30 to use Central Limit Theorem with estimated standard deviation*

Significance Test for a Mean

Hypotheses

- The null hypothesis has the form $H_0 : \mu = \mu_0$ where μ_0 is an a priori (before taking the sample) specified number like 0 or 5.3 or 500
- The most common alternative hypothesis is

$$H_1 : \mu \neq \mu_0$$

- This is called a two-sided hypothesis, since it includes values falling above and below the null hypothesis

Significance Test for a Mean

Test Statistic

- The hypothesis is about the population mean
- So, a natural test statistic would be the sample mean
- The sample mean has, for sample size of at least *n=25*, an approximately normal sampling distribution
- The parameters of the sampling distribution are, under the null hypothesis,
  - Mean = μ_0 (that is, the sampling distribution is centered around the hypothesized mean)
  - Standard error = $\frac{\sigma}{\sqrt{n}}$, estimated by $\frac{s}{\sqrt{n}}$

Significance Test for a Mean

Test Statistic

- Then, the *z*-score has a standard normal distribution

$$z = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}}$$

- The *z*-score measures how many estimated standard errors the sample mean falls from the hypothesized population mean
- The farther the sample mean falls from μ_0 the larger the absolute value of the *z* test statistic, and the stronger the evidence against the null hypothesis

Significance Test for a Mean

p-Value

- The *p*-value has the advantage that different test results from different tests can be compared: The *p*-value is always a number between 0 and 1
- It is the probability that a standard normal distribution takes values more extreme than the observed *z* score
- The smaller the *p*-value is, the stronger is the evidence against the null hypothesis and in favor of the alternative hypothesis
- Round *p*-value to two or three significant digits

Example

- The mean score for all high school seniors taking a college entrance exam equals 500.
- A study is conducted to see whether a different mean applies to those students born in a foreign country.
- For a random sample of 100 of such students, the mean and standard deviation on this exam equal 508 and 100.

1. Set up hypotheses for a significance test.
2. Compute the test statistic.
3. Report the *P*-value, and interpret.
4. Can you conclude that the population mean for students born in a foreign country equals 500? Why or why not?
5. Make a decision about H_0, using alpha=0.05
6. Construct a 95% confidence interval for mu.