STA 321 Spring 2014

Lecture 14 *Thursday, March 13*

- **Confidence Intervals**

Example: Relaxing

- In a recent Gallup poll, 455 of 1011 randomly selected adults aged 18 years and older said that they had too little time for relaxing or doing nothing.
- Construct and interpret a 95% confidence interval for the population proportion.
- Construct a 99% confidence interval and compare.

Typical Question

- *Based on a sample of n=1000 people, a 95% confidence interval for the population proportion of people voting for candidate A is calculated. It turns out to be from 67% to 73%.*
- *What does "95% confidence" mean?*
- The confidence interval (0.67, 0.73) either does or does not contain the population proportion. We don't know whether it does.
- We are 95% confident that the true population proportion is between 67% and 73%.
- That is, if we repeatedly selected random samples of the same size and each time constructed a 95% confidence interval, then in the long run about 95% of the intervals would contain the true, unknown population proportion.

Multiple Choice Question
Which of the following statements are true?

- **“95% confidence” means that**
  1. 95% of the true population parameters are in the confidence interval
  2. If we were to repeat the procedure of sampling and calculating confidence intervals from the same population, then 95% of the times our confidence interval will contain the true population parameter
  3. If we were to repeat the procedure of sampling and calculating confidence intervals from the same population, then 95% of the population parameters are going to be in every calculated interval

Multiple Choice Question
Which of the following statements are true?

- “**If we calculate a specific confidence interval based on a sample (say the interval turns out to be from 2.6 to 4.6), then**
  1. The true population parameter is in this interval with 95% probability
  2. We do not know whether the true population parameter is in this interval or not
  3. 95% of the time, the interval will be from 2.6 to 4.6.

Calculating *z*-Scores

1. **z-Score for an individual observation**
   - You need to know *Y*, *mu*, and sigma to calculate *z*

$$z = \frac{Y - \mu}{\sigma}$$

2. **z-Score for a sample mean**
   - You need to know *Y-bar*, *mu*, *sigma*, and *n* to calculate *z*

$$z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

3. **z-Score for a sample proportion**
   - You need to know *p-hat*, *p*, and *n* to calculate *z*

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Confidence Intervals for Mean and Proportion

$$\bar{Y} \pm z \cdot \frac{s}{\sqrt{n}}$$

$$\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval: Interpretation

- For a 95% confidence interval:
- If we repeatedly took random samples using the same method, then, in the long run, in 95% of the cases, the confidence interval will cover the true unknown parameter
- For one given sample, we do not know whether the confidence interval covers the true parameter
- The ***95% probability*** refers to the ***method*** that we use, but not to the individual sample

Example: Sleep

- How much do Americans sleep each night?
- A random sample of 1120 Americans (15y and older) had a mean amount of sleep per night of 7.67 hours, and the standard deviation was 1.2 hours.
- Construct and interpret a 95% confidence interval for the mean amount of sleep per night of Americans 15 years of age and older.

Choice of Sample Size

$$\bar{Y} \pm z \cdot \frac{s}{\sqrt{n}} = \bar{Y} \pm B \qquad \hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \hat{p} \pm B$$

- So far, we have calculated confidence intervals starting with z, s, n or $z, \hat{p}, n$
- These three numbers determine the precision *B* of the confidence interval
- Now we reverse the equation:
  - We specify a desired precision *B* (=bound = margin of error)
  - Given z and s, or z and $\hat{p}$, we can find the minimal sample size needed for this precision by solving the above equations for *n*

Choice of Sample Size

- The results are

$$n = s^2 \cdot \left(\frac{z}{B}\right)^2 \quad \text{and} \quad n = \hat{p} \cdot (1 - \hat{p}) \cdot \left(\frac{z}{B}\right)^2$$

- However, in practice, we usually don't know s or $\hat{p}$ before taking the sample
- **In the second formula (only there!)**, we can take the safe but conservative approach of setting p-hat=0.5, resulting in $n = 0.25 \cdot \left(\frac{z}{B}\right)^2$
- This is like a "worst case scenario" because the product can never exceed 0.25

Example: Sleep

- We would like to find out how much Americans sleep each night?
- Our last random sample had a sample standard deviation of 1.2 hours.
- How large a random sample do we need to obtain a 95% confidence interval with precision plus/minus 15 minutes?
- “Predicting the population average to within 15 minutes, with 95% confidence”

Example: Relaxing

- We would like to find out which (population) proportion of adults thinks that they have too little time for relaxing or doing nothing.
- The last poll had a proportion of 45% saying this.
- How many people do we need in our sample if we want to predict the population proportion to within five percentage points, with 90% confidence?
- What if we don’t know about the results from the last poll?

Summary: CIs for the Mean

- Large sample confidence interval for the mean μ has the form

$$\left[\bar{Y} - z\frac{s}{\sqrt{n}}, \bar{Y} + z\frac{s}{\sqrt{n}}\right]$$

- If given a bound on the margin of error, *B*, and asked for a minimum *n* to achieve that bound:

$$n = \left\lceil s^2\left(\frac{z^2}{B^2}\right)\right\rceil$$, where ⌈ ⌉ means "round up".

Summary: CIs for the Proportion

- Large sample confidence interval for the mean *p* has the form

$$\left[\hat{p} - z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

- If given a bound on the margin of error, *B*, and asked for a minimum *n* to achieve that bound:

$$n = \left\lceil 0.25 \cdot \left(\frac{z^2}{B^2}\right)\right\rceil$$

(unless you have prior information about p-hat)

Hypothesis Testing

- Fact: It is easier to *prove* that a parameter **isn't** equal to a particular value than it is to prove it **is** equal to a particular value
- Hypothesis testing: *Proof by contradiction*:
  - we set up the belief we wish to disprove as the **null hypothesis** (**H_0**) and the belief we wish to prove as our **alternative hypothesis** (**H_1**) (or: research hypothesis)

Analogy: Court trial

- In American court trials, the jury is instructed to think of the defendant as innocent:

H_0: Defendant is innocent

- District attorney, police involved, plaintiff, etc., bring every evidence, hoping to prove

H_1: Defendant is guilty

- Which hypothesis is correct?
- Does the jury make the right decision?

Back to statistics …

Critical Concepts

- Two hypotheses: the null and the alternative
- Process begins with the assumption that the null is *true*
- We calculate a test statistic to determine if there is enough evidence to infer that the alternative is true
- Two possible decisions:
  - Conclude there is enough evidence to reject the null (and therefore accept the alternative)
  - Conclude that there is not enough evidence to reject the null
- Two possible errors?

What about those errors?

Two possible errors:

- Type I error: Rejecting the null when we shouldn't have [P(Type I error) = α]

- Type II error: Not rejecting the null when we should have [P(Type II error) = β]

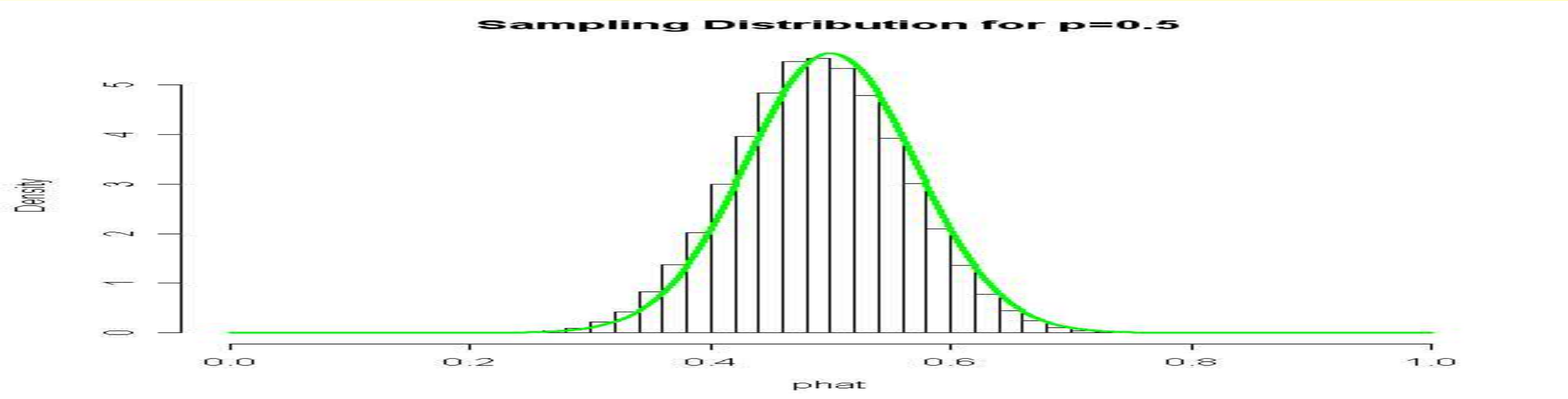
Hypothesis Testing Example

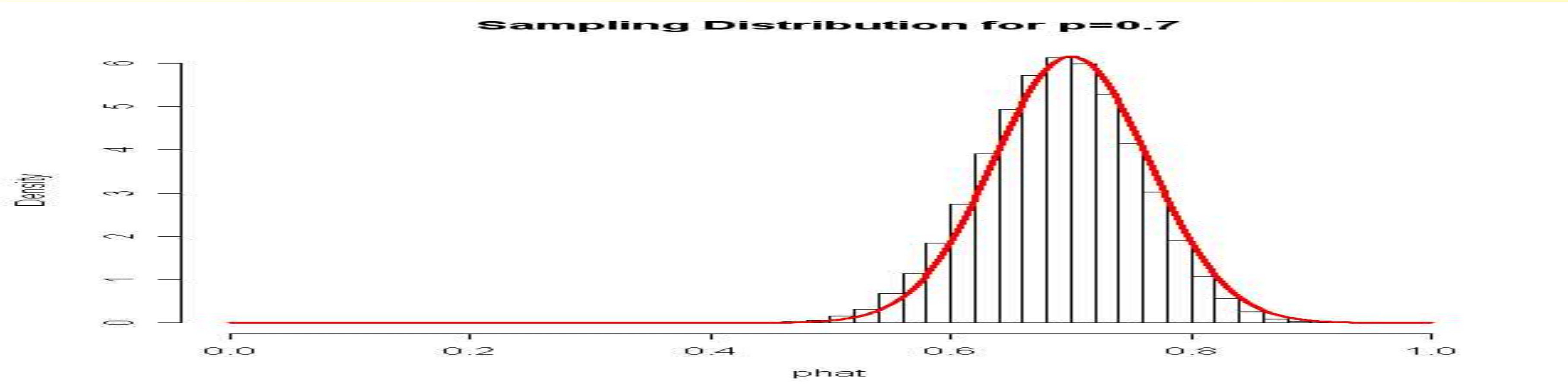
- You are given a coin. You know the coin might be fair (50% heads, 50% tails), but the coin might also be weighted (70% heads, 30% tails).
- You flip the coin 50 times and get 29 heads. Is the coin fair or weighted?

Sampling Distributions for Each Kind of Coin

- Suppose the coin is weighted, so p=0.7.
- If you flip the coin n=50 times, the sampling distribution of the proportion of heads p-hat is normal with mean p=0.7 and standard deviation sqrt(p(1-p)/n) = sqrt(0.7*0.3/50) = 0.0648.
- If the coin is fair, with p=0.5, the sampling distribution of p-hat is normal with mean p=0.5 and standard deviation sqrt (0.5*0.5/50) = 0.0707.

Need Cutoff to Separate These Groups.





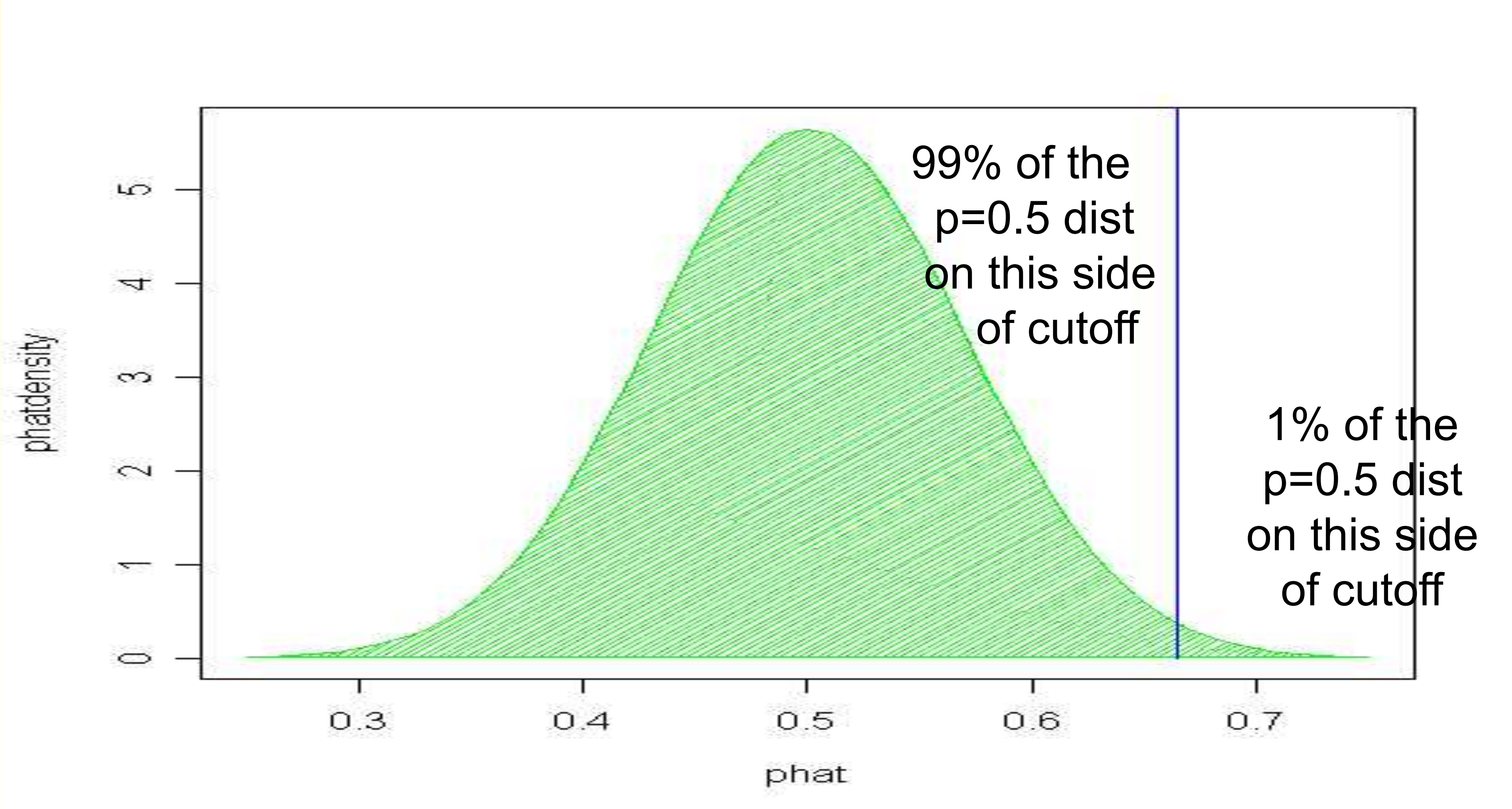
Determining a Cutoff

- We choose the cutoff to either
  - set the proportion of fair coins that are classified correctly, or
  - set the proportion of weighted coins that are classified correctly.
- We assume that the coin is fair (null hypothesis), and we try to find evidence against this.
- That is, H_0 : p=0.5 vs. H_1 : p=0.7

Example, contd.

- Return to n=50, which resulted in
  p-hat ~ N(0.5,0.0707) for H_0 : p=0.5
  p-hat ~ N(0.7,0.0648) for H_1 : p=0.7
- Let us determine a cutoff where the probability that a p=0.5 coin is classified correctly is 99%.
- We classify a coin as p=0.5 if p-hat is below the cutoff, so we need a cutoff that is the 99^{th} percentile of a N(0.5,0.0707) distribution.

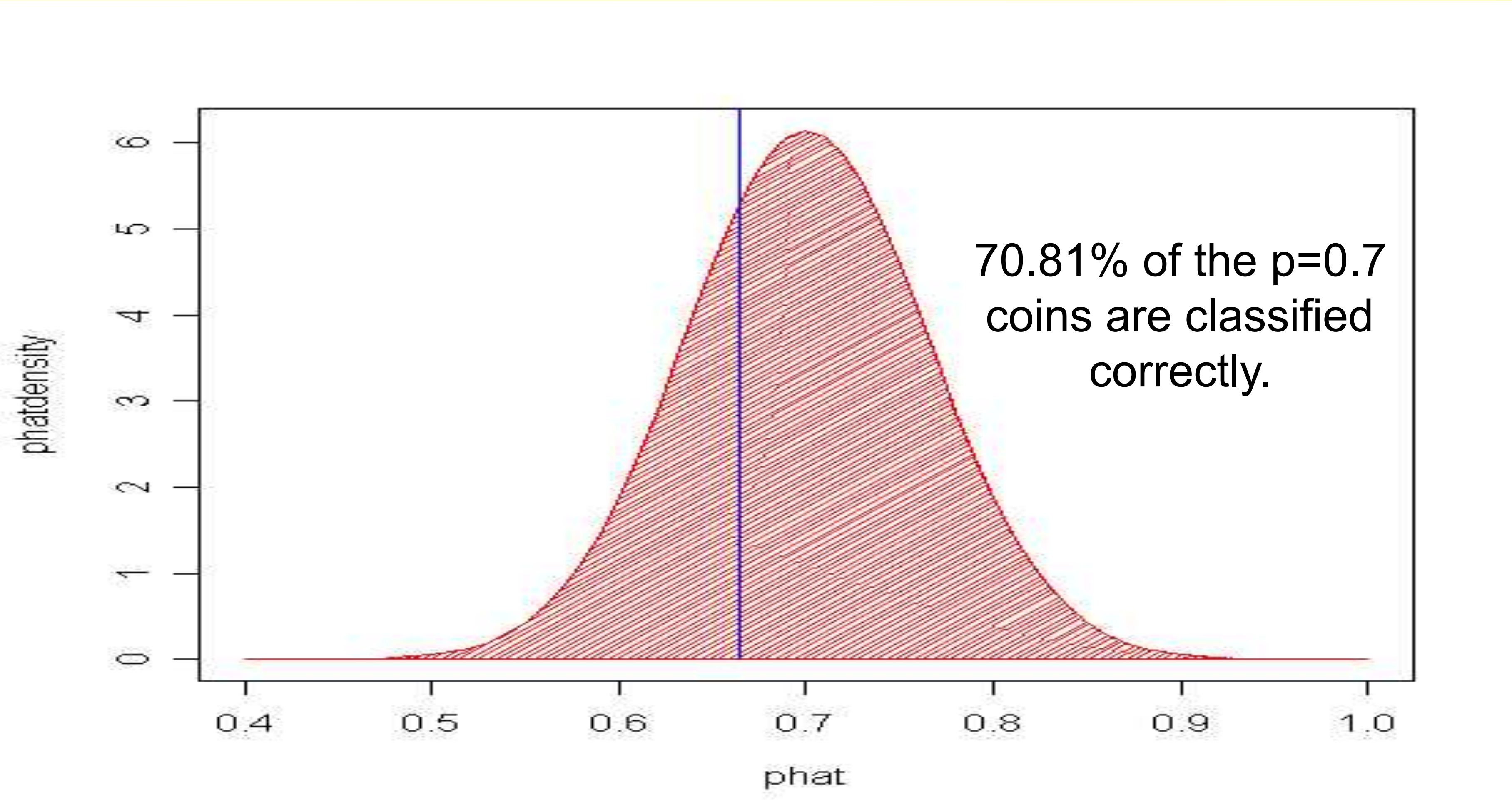
A cutoff of 0.6645 results in 99% correct classification of p=0.5 coins.



What about the weighted (p=0.7) coins?

- For the chosen cutoff of 0.6645, we can also find the probability a p=0.7 coin is correctly classified.
- We need to find the probability a N (0.7,0.0648) is greater than the cutoff of 0.6645, which is 70.81%.

70.81% of the p=0.7 coins are classified correctly with cutoff=0.6645



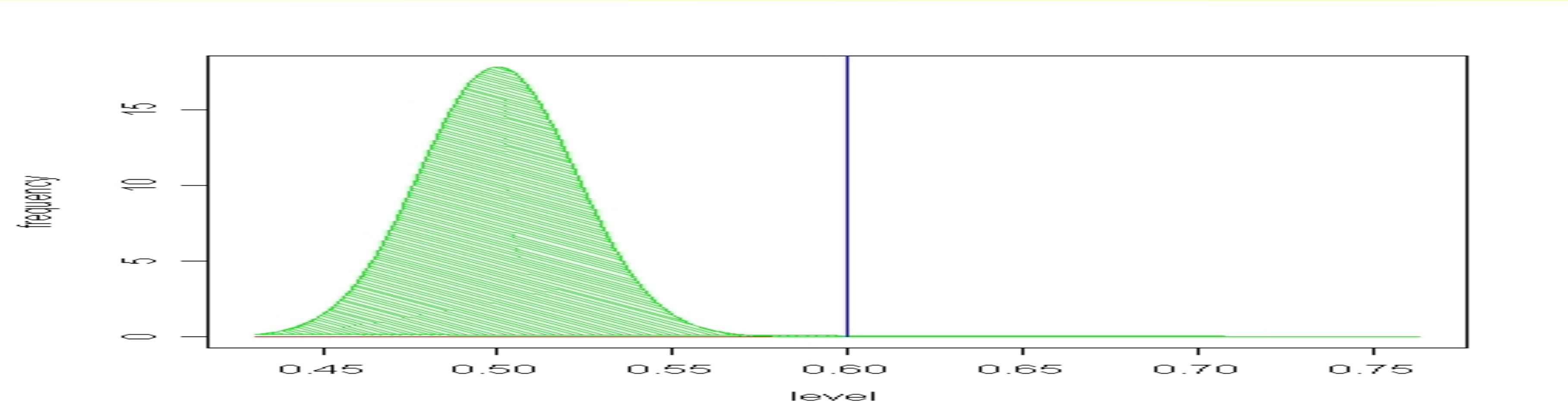
Larger Sample Sizes Increase Accuracy

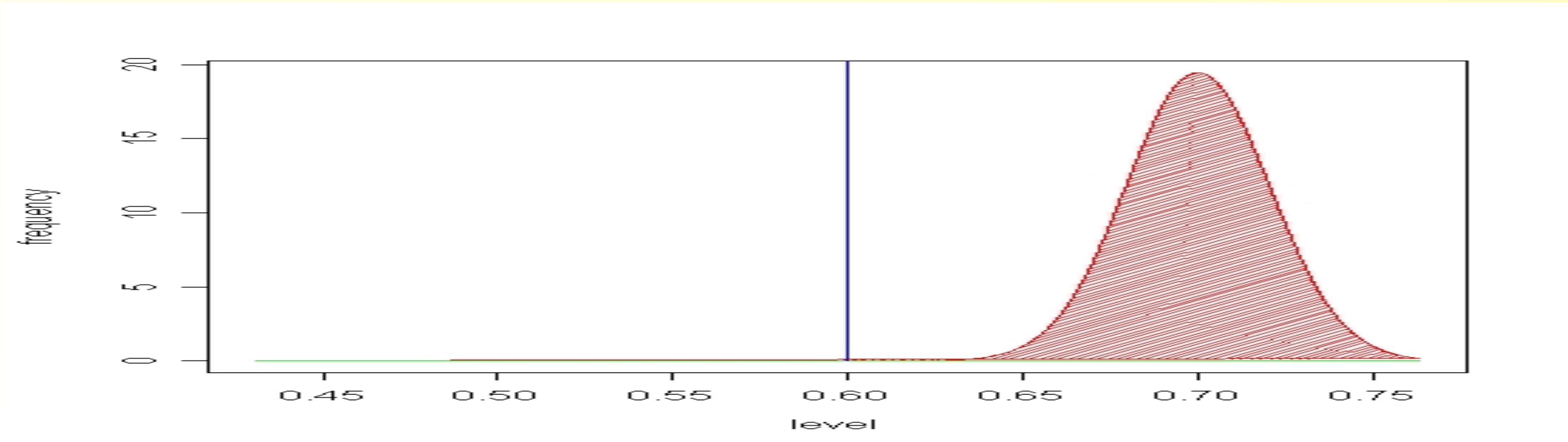
- Simple way to make a stronger test:
- Flip the coin more than 50 times.
- This changes both sampling distributions, reducing the standard deviation of both.

Suppose we flip the coin n=500 times

- When p=0.5, the sampling distribution is normal with mean 0.5 and standard deviation sqrt(0.5*0.5/500) = 0.0224
- When p=0.7, the sampling distribution is normal with mean 0.7 and standard deviation sqrt(0.7*0.3/500) = 0.0205
- The two groups are now well separated.

Now the groups are well separated





Terminology

- Choosing the group corresponding to H_1 is called “rejecting the null hypothesis”
- Choosing the group corresponding to H_0 is called “not rejecting the null hypothesis”

More Terminology

- Note: We set the probability that the correct decision is made when H_0 is true. This probability is typically termed 1-α.
- Thus, α is the probability of making a mistake when H_0 is true (rejecting when you should NOT reject).
- Once α is set, the cutoff is determined. The most common value for α is 0.05.
- Aside from making α small (it is the chance of a mistake) there is no absolute justification for this choice compared to others.

More Terminology

- The choice of α determines the cutoff, and thus the probability of making the correct decision when H_1 is true (when H_1 is true the correct decision is to reject the null hypothesis).
- This probability is called the ***power*** of the test.
- Thus, we want 1-α and the power to be high (close to 1).

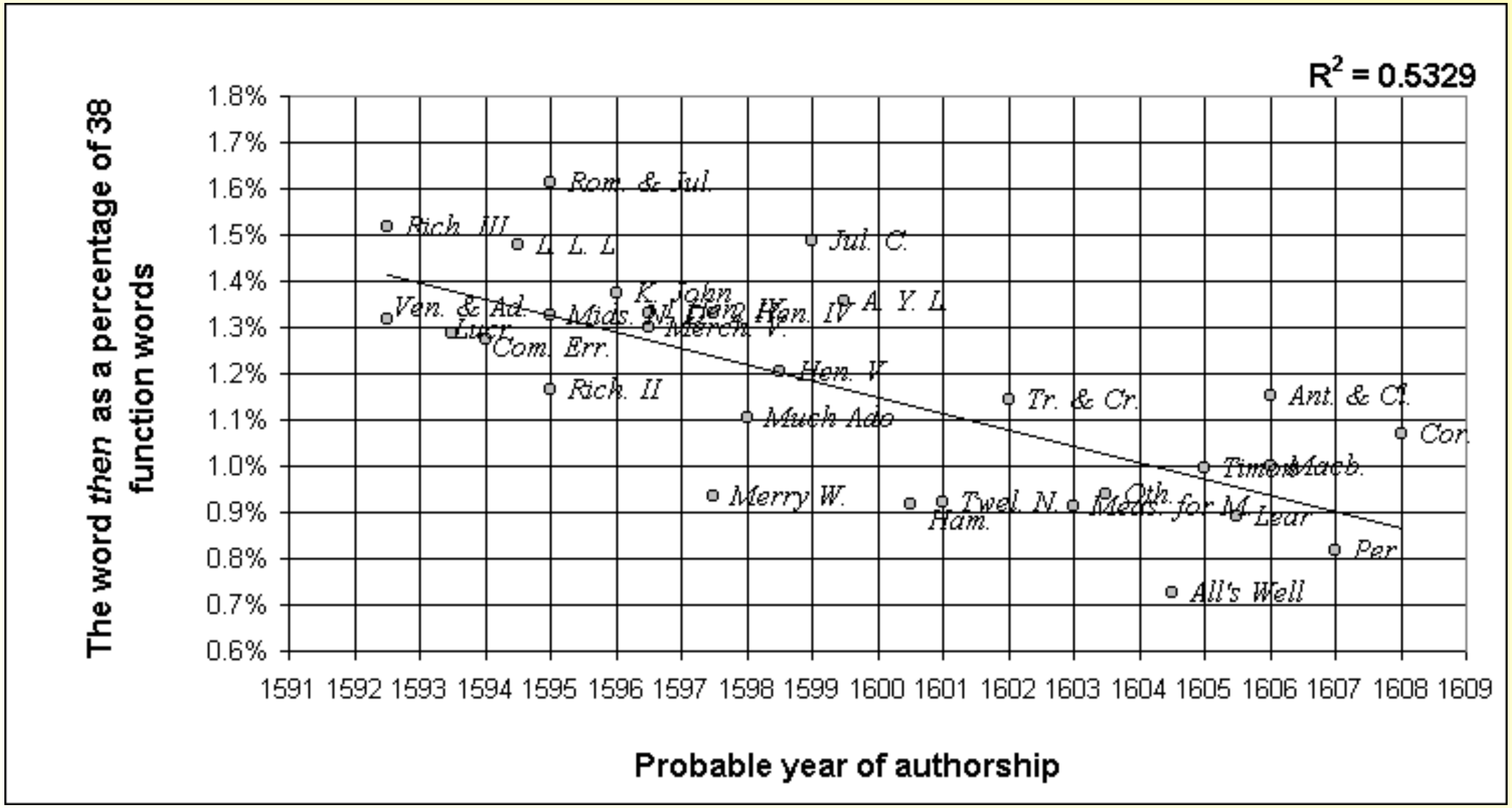
Four Possibilities

- There are two possible states of the world – either the null hypothesis is true or the alternative hypothesis is true.
- You can make two decisions – either reject the null hypothesis or do not reject the null hypothesis.
- Thus, there are four possibilities. Two correspond to correct decisions and two are errors.

By the way..regression is also useful for this

- Which of the Shakespeare Sonnets were really within the Shakespeare canon?

http://www.sonnets.org/canon.htm



Example

- Suppose you have a document that may or may not have been written by author Bill
- You observe that Bill uses a particular form X of the word 80% of the time
- The document in question has 84 instances of the word choice, and word X is used 58 times.

Example continued

- Null hypothesis H_0: $p=0.8$ vs. H_1 : $p≠0.8$
- The ***null distribution*** (when H_0 is true) is normal with mean 0.8 and standard deviation sqrt(0.8*0.2/84)=0.0436.
- The null distribution is the sampling distribution of the sample statistic when the null hypothesis is assumed true.

Example continued

- We have the null distribution N(0.8,0.0436)
- Let's choose $\alpha=0.05$
- We will reject H_0 for p-hats far away from 0.8.
- The cutoff to be the 2.5^{th} and the 97.5^{th} percentile of the null distribution. This corresponds to Z=(-1.96) and Z=1.96, and a cutoff of
- (-1.96)(0.0436)+0.8=0.7145 and
- (1.96)(0.0436)+0.8=0.8855

Example continued

- (-1.96)(0.0436)+0.8=0.7145 and
- (1.96)(0.0436)+0.8=0.8855
- Our ***rejection region*** consists of those values of p-hat that are not between 0.7145 and 0.8855
- Our actual observed phat was 58/84=0.6904.
- Thus our conclusion is to reject H_0. We would conclude that H_1 is true and that Bill did not write the document.

Example continued

- How likely is 58/84=0.6904 under the null hypothesis?
- The z-score is (0.6904-0.8)/0.0436= –2.51.
- Beyond (Here="below") a z-score of –2.51 is probability 0.00597.
- All values ***at least as extreme as*** a z-score of -2.51 have together probability 2 x 0.00597 = 0.0119.
- This is the ***P-value: The probability, assuming that the null hypothesis is true, of observing anything at least as extreme as what we actually observed.***
- ***"As extreme" = "providing as much, or more evidence against the null hypothesis"***

Example continued

- By construction, the probability of type I error is $\alpha=0.05=5\%$. What is the probability of type II error?
- Type II error is not rejecting H_0 when in fact H_1 is true.
- Assume that another author, Maria, could have written the manuscript, and she chooses form X 60% of the time.
- The sampling distribution for her would be the ***alternative distribution*** (when H_1 is true), normal with mean 0.6 and standard deviation sqrt(0.6*0.4/84) =0.0535.
- We need to find the probability that the alternative distribution places above the cutoff 0.7145.
- The Z-score for a N(0.6,0.0535) is Z=(0.7145-0.6)/0.0535=2.14 and the probability of type II error is 0.01609=1.6%