Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Syllabus, etc.
2. Textbook and hw.
3. Example with organ donations.
4. Rough interpretation of distribution and standard deviation.
5. Sample size.
6. Statistic and parameter.
7. Categorical and quantitative variables.
8. Statistical significance and testing.
9. Null and alternative hypotheses.
10. Z statistic.

Read preliminaries, chapter 1, and p592, the first page of Appendix A.
Hw1 is due Fri Aug14, 10pm. 1.3.16 and 1.4.26. Also, on the bottom of your hw, print the names and emails of two other students in the class.
HW should be submitted BY EMAIL to STATGRADER@STAT.UCLA.EDU if you are in Section A or B, and to STATGRADER2@STAT.UCLA.EDU if you are in Section C or D

The course website is http://www.stat.ucla.edu/~frederic/13/sum26 .

1. Syllabus, etc.

Read the syllabus, especially the hw policy, the gradegrubbing policy,
and the 1 question not to ask me.
Here are things worth emphasizing.

The CCLE/Canvas/Bruinlearn website for this course is not maintained by me.
Your TA might use them, but the only course website for my lectures and homework is
http://www.stat.ucla.edu/~frederic/13/sum26 .

I do not give hw hints in office hours or by email. Conceptual questions only.

Attendance is not mandatory in lecture nor in section and lab.

You can only switch sections if someone will switch with you. If you would like to switch,
please email me at frederic@stat.UCLA.edu and maybe I can find someone to switch with you.

2. Textbook and hw.

Tintle N, Chance BL, Cobb GW, Rossman AJ, Roy S, Swanson T, and Vanderstoep J. (2016).
Introduction to Statistical Investigations, Wiley, NY.

Emphasizes concepts, not formulas.

Emphasizes randomization tests and other nonparametric methods.

Verbose, and some examples are phony or unimportant.

Optional reading, "Statistics for the Life Sciences", by Samuels and Witmer.

a. Define the parameter of interest in the context of the study and assign a symbol to it.

b. State the null hypothesis and the alternative hypothesis using the symbol defined in part (a).

c. Of the 124 kissing couples, 80 were observed to lean their heads right. What is the observed proportion of kissing couples who leaned their heads to the right? What symbol should you use to represent this value?

d. Determine the standardized statistic from the data. (*Hint:* You will need to get the standard deviation of the simulated statistics from the null distribution.)

e. Interpret the standardized statistic in the context of the study. (*Hint:* You need to talk about the value of your observed statistic in terms of standard deviations assuming ________ is true.)

f. Based on the standardized statistic, state the conclusion that you would draw about the null and alternative hypotheses.

1.3.14 Suppose that instead of H_0: $\pi = 0.50$ like it was in the previous exercise, our null hypothesis was H_0: $\pi = 0.60$.

a. In the context of this null hypothesis, determine the standardized statistic from the data where 80 of 124 kissing couples leaned their heads right. (*Hint:* You will need to get the standard deviation of the simulated statistics from the null distribution.)

b. How, if at all, does the standardized statistic calculated here differ from that when H_0: $\pi = 0.50$? Explain why this makes sense.

Love, first*

1.3.15 A previous exercise (1.2.16) introduced you to a study of 40 heterosexual couples. In 28 of the 40 couples the male said "I love you" first. The researchers were interested in learning whether these data provided evidence that in significantly more than 50% of couples the male says "I love you" first.

a. State the null hypothesis and the alternative hypothesis in the context of the study.

b. Determine the standardized statistic from the data. (*Hint:* You will need to get the standard deviation of the simulated statistics from the null distribution.)

c. Interpret the standardized statistic in the context of the study. (*Hint:* You need to talk about the value of your observed statistic in terms of standard deviations assuming ________ is true.)

d. Based on the standardized statistic, state the conclusion that you would draw about the research question of whether males are more likely to say "I love you" first.

Rhesus monkeys

Revisit Exercise 1.2.18 about the study on Rhesus monkeys. When given a choice between two boxes, 30 out of 40 monkeys approached the box that the human had gestured toward, and 10 approached the other box. The purpose is to investigate whether rhesus monkeys can interpret human gestures better than random chance.

1.3.16 For this study:

a. State the null hypothesis and the alternative hypothesis in the context of the study.

b. Determine the standardized statistic from the data. (*Hint:* You will need to get the standard deviation of the simulated statistics from the null distribution in an applet.)

c. Interpret the standardized statistic in the context of the study. (*Hint:* You need to talk about the value of your observed statistic in terms of standard deviations assuming ________ is true.)

d. Based on the standardized statistic, state the conclusion that you would draw about the research question of whether rhesus monkeys have some ability to understand gestures made by humans.

Tasting tea*

Revisit Exercise 1.1.12 about the study on a lady tasting tea. When presented with eight cups containing a mixture of milk and tea, she correctly identified whether tea or milk was poured first for all eight cups. Is she doing better than if she were just guessing?

1.3.17 For this study:

a. Define the parameter of interest in the context of the study and assign a symbol to it.

b. State the null hypothesis and the alternative hypothesis using the symbol defined in part (a).

c. What is the observed proportion of times the lady correctly identified what was poured first into the cup? What symbol should you use to represent this value?

d. Suppose that you were to generate the null distribution of the sample proportion of correct answers, that is, the distribution of possible values of sample proportion of correct identifications if the lady always guesses. Where would you anticipate this distribution would center? Also, do you anticipate the SD of the null distribution to be negative, positive, or 0? Why?

e. Use an applet to generate the null distribution of sample proportion of correct identifications and use it to determine the standardized statistic.

f. Interpret the standardized statistic in the context of the study. (*Hint:* You need to talk about the value of your observed statistic in terms of standard deviations assuming ________ is true.)

g. Based on the standardized statistic, state the conclusion that you would draw about the research question of whether the lady does better than randomly guess.

1.4.23 For the "leaning" version of the study from the previous question:

a. *Statistic:* How many times did Krieger choose the correct object? Out of how many attempts? Thus, what proportion of the time did Krieger choose the correct object?

b. *Simulate:* Using an applet, simulate 1,000 repetitions of having the dog choose between the two objects if he is doing so randomly. Report the null and standard deviation.

c. Based on the study's result, what is the p-value for this test?

d. Approximately what proportion of the 10 attempts would Krieger have needed to get correct in order to yield a p-value of approximately 0.05?

1.4.24

a. Based on the study's result, what is the standardized statistic for this test?

b. *Strength of evidence:* What are your conclusions based on the p-value you found in part (d) from the previous exercise? Are the conclusions the same if you base them off the standardized-statistic you found in (a)?

c. Revisit your conjecture in Exercise 1.4.22, part (d). Did the p-value behave the way you had conjectured?

The sign test

io far, the outcome has always been binary—Yes/No, tight/Wrong, Heads/Tails, etc. What if outcomes are uantitative, like heights or percentages? Although there re specialized methods for such data that you will learn in later chapter, you can also use the methods and logic you ave already learned for situations of a very different sort:) outcomes are quantitative, (2) you want to compare ro conditions A and B, and (3) your data come in pairs, ıe A and one B in each pair. To apply the coin toss model, ›u simply ask for each pair, "Is the A value bigger than e B value?" The resulting test is called the "sign test" beuse the difference $(A - B)$ is either plus or minus. Here's ummary table:

:oin toss	Heads	P(Heads)	Null hypothesis	Statistic
zz's guess	Right	$\pi = P(\text{Right})$	$\pi = 0.50$	$\hat{p}$
:h pair	$A > B$	$\pi = P(A > B)$	$\pi = 0.50$	$\hat{p}$

ine providence

25* Refer to Exercises 1.4.8 to 1.4.12. Dr. Arbuthnot's al analysis was different from the analysis you saw er. Instead of using each individual birth as a coin toss, ıthnot used a sign test with each of the 82 years as a toss, and a year with more male births counted as a :ess."

a. Complete the following table of comparisons:

Analysis method	Sample size n	Null value π_0	Value of $\hat{p}$
A: 1.4.8 – 1.4.12	______	______	______
B: 1.4.25	______	______	______

b. For each method of analysis, rate the strength of evidence against the null hypothesis, as one of: inconclusive, weak but suggestive, moderately strong, strong, or overwhelming.

Healthy lungs

1.4.26 Researchers wanted to test the hypothesis that living in the country is better for your lungs than living in a city. To eliminate the possible variation due to genetic differences, they located seven pairs of identical twins with one member of each twin living in the country, the other in a city. For each person, they measured the percentage of inhaled tracer particles remaining in the lungs after one hour: the higher the percentage, the less healthy the lungs. They found that for six of the seven twin pairs the one living in the country had healthier lungs.

a. Is the alternative hypothesis one-sided or two-sided?

b. Based on the sample size and distance between the null value and the observed proportion, estimate the strength of evidence: inconclusive, weak but suggestive, moderately strong, strong, or overwhelming.

c. Here are probabilities for the number of heads in seven tosses of a fair coin:

# Heads	0	1	2	3	4	5	6	7
Probability	0.0078	0.0547	0.1641	0.2734	0.2734	0.1641	0.0547	0.0078

Compute the p-value and state your conclusion.

Bee stings

1.4.27* Scientists gathered data to test the research hypothesis that bees are more likely to sting a target that has already been stung by other bees. On eight separate occasions, they offered a pair of targets to a hive of angry bees; one target in each pair had been previously stung, the other was pristine. On six of the eight occasions, the target that had been previously stung accumulated more new stingers.

a. Is the alternative hypothesis one-sided or two-sided?

b. Based on the sample size and distance between the null value and the observed proportion, estimate the strength of evidence: inconclusive, weak but suggestive, moderately strong, strong, or overwhelming.

c. Here are probabilities for the number of heads in eight tosses of a fair coin:

# Heads	0	1	2	3	4	5	6	7	8
Probability	0.0039	0.0313	0.1094	0.2188	0.2734	0.2188	0.1094	0.0313	0.0039

Compute the p-value and state your conclusion.

2. Textbook, continued.
If you have a different edition of the textbook than the 2016 edition, then make sure you are doing the correct hw problems.
Hw1 is 1.3.16, 1.4.26, and the names and emails of 2 students.

1.3.16 is on p84 and is about Rhesus monkeys, exercise 1.2.18, which is on p80. In part b, it says "Hint: you will need to get the standard deviation of the simulated statistics from the null distribution in an applet." But you don't need an applet. You can use the formula

$$\text{SE for a proportion} = \sqrt{\pi(1-\pi)/n}\ ,$$

where π is the probability of the monkey getting it right under the null hypothesis, or do simulations in *R.* For instance, in *R* you could do:

```
pi2 = ## insert your answer to the null hypothesis part of question a here.
a = rep(0,10000)
for(i in 1:10000){
      b = runif(40)
      c = (b < pi2)
      a[i] = mean(c)
}
sd(a)
## compare with
sqrt(pi2 * (1-pi2) / n)
```

Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

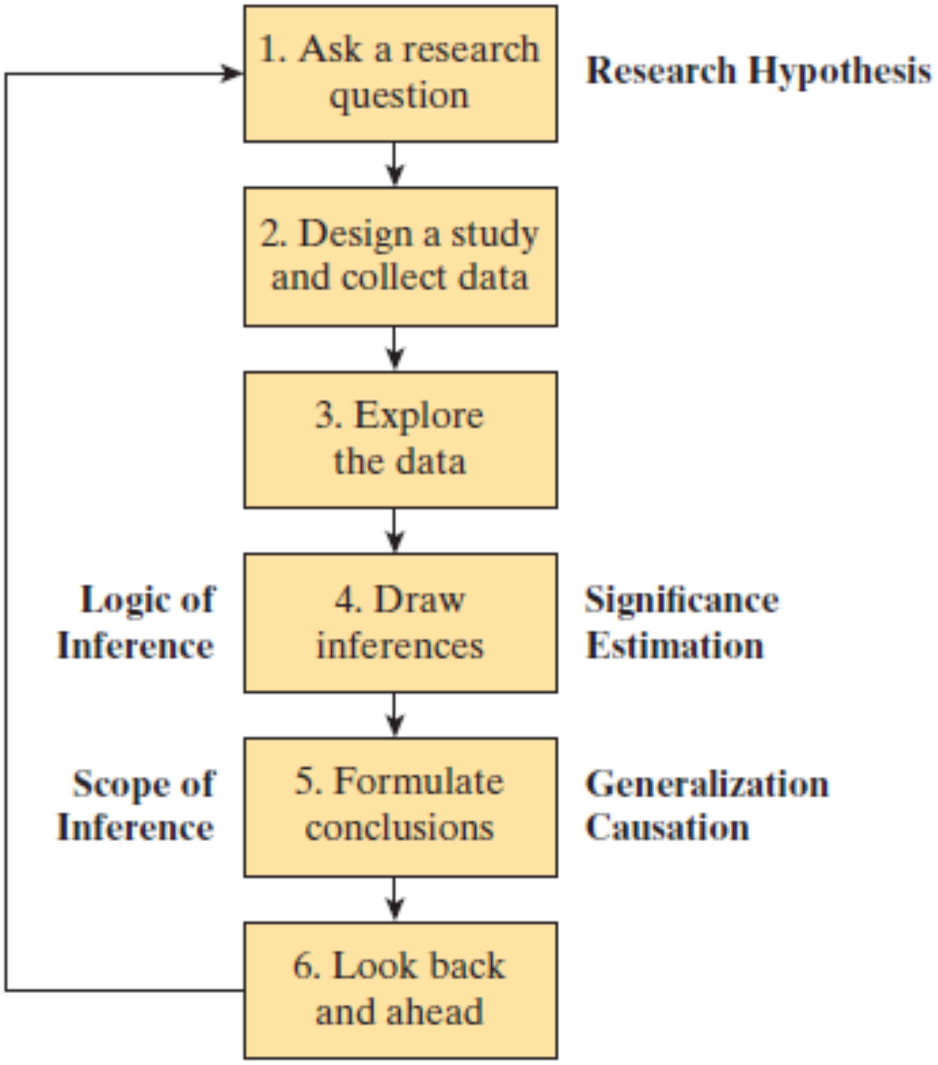
Hw1 is 1.3.16, 1.4.26, and the names and emails of 2 students.

1.4.26 starts "Researchers wanted to test the hypothesis that living in the country is better for your lungs than living in a city."
Be careful in part c. The table gives you P(# of heads = i), not P(# of heads ≥ i), for i = 0,1,2,3,..., 7.

3. Example P.1: Organ Donations

- While a majority of people approve of organ donation in principle, far less than that actually sign up when getting a driver’s license.
- Different states (and different countries) have different recruiting methods.
- Do these different methods result in different sign-up rates?

Six-Step Statistical Investigation Method



Recruiting Organ Donors

Step 1. Ask a Research Question

- Does the default option presented to driver's license applicants influence the likelihood of someone becoming an organ donor?

Recruiting Organ Donors

Step 2: Design a study and collect data

- The researchers decided to recruit various participants and ask them to pretend to apply for a new driver's license.
- The participants did not know in advance that different options were given for the donor question, or even that this issue was the main focus of the study.

Recruiting Organ Donors

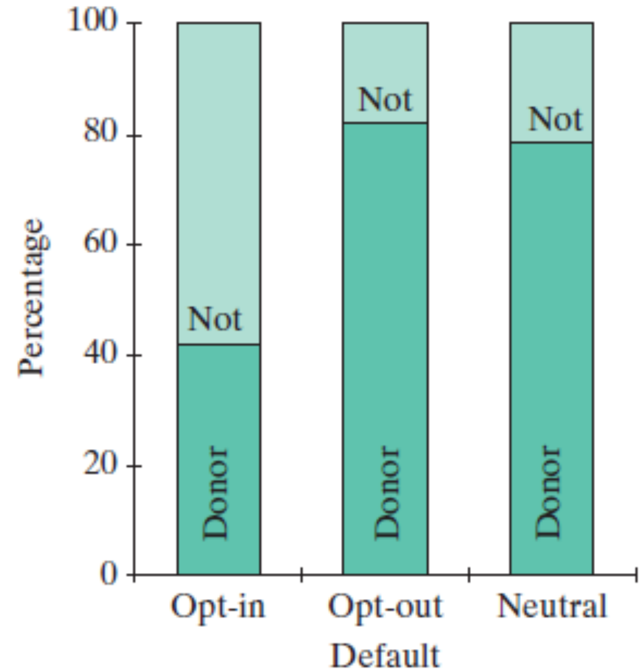
Step 2: Design a study and collect data

- Some of the participants were forced to make a choice of becoming a donor or not, without being given a default option (the “neutral” group, Michigan’s current practice).
- Other participants were told that the default option was *not* to be a donor but that they could choose *to* become a donor if they wished (the “opt-in” group, Michigan’s past practice).
- The remaining participants were told that the default option was to *be* a donor but that they could choose *not* to become a donor if they wished (the “opt-out” group, some countries use this practice).

Recruiting Organ Donors

Step 3: Explore the data.

- 23 of 55 (41.8%) participants in the opt-in group agreed to become organ donors
- 41 of 50 (82.0%) participants in the opt-out group agreed to become organ donors
- 44 of the 56 (78.6%) participants in the neutral group agreed to become organ donors



Recruiting Organ Donors

Step 4: Draw inferences beyond the data.

- Using methods that you will learn in this course, the researchers analyzed whether the observed differences between the groups was large enough to indicate that the default option had a genuine effect.
- In particular, they reported strong evidence that the neutral and opt-out versions do lead to a higher chance of agreeing to become a donor, as compared to the opt-in version currently used in many states.
- In fact, they could be quite confident that the neutral version increases the chances that a person agrees to become a donor by between 20 and 54 percentage points, a difference large enough to save thousands of lives per year in the United States.

Recruiting Organ Donors

Step 5: Formulate conclusions.

- Based on the analysis of the data and the design of the study, the researchers concluded that the neutral version *causes* an increase in the proportion who agree to become donors over the opt-in.
- But because the participants in the study were volunteers recruited from various general interest Internet bulletin boards, generalizing conclusions beyond these participants is only legitimate if they are representative of a larger group of people. (The authors believed their sample included a "broad range of demographics.")

Recruiting Organ Donors

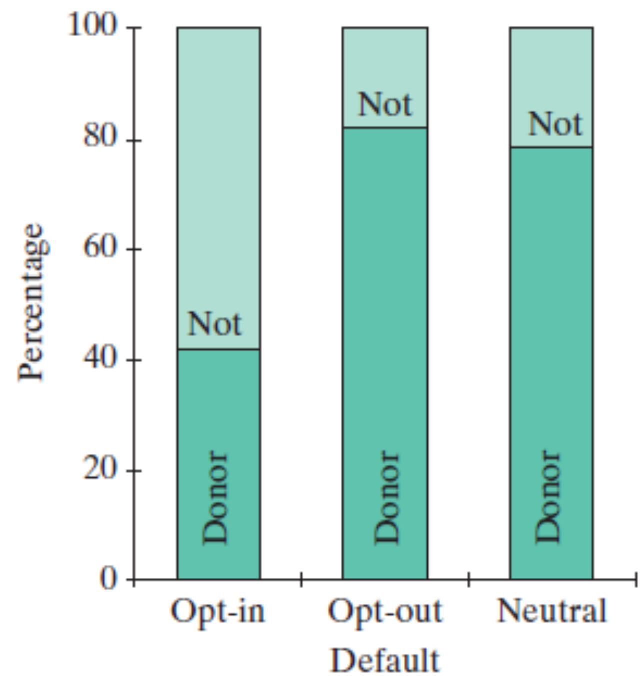
Step 6: Look back and ahead.

- One limitation of the study is that participants were asked to imagine how they would respond, which might not mirror how people would actually respond in such a situation.
- A new study might look at people's actual responses to questions about organ donation or could monitor donor rates for states that adopt a new policy.

- The individual entities on which data are recorded are called ***observational units***.
- The recorded characteristics of the observational units are the ***variables*** of interest.
- What are the observational units and variables in the Organ Donation Study?

4. Distribution and SD (rough definitions)

- The ***distribution*** of variable describes the pattern of value/category outcomes.
- For the organ donation study the bar chart shown displays the distribution of responses.



- One way to measure the center of a distribution is with the average, also called the mean.

  The sample mean $\bar{x} = \sum x_i/n$.

- One way to measure variability is with the ***standard deviation***, which is roughly the average distance between a data value in the distribution and the mean of the distribution.

- The sample std deviation s = $\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}$.
  - What is the standard deviation of the data set {7,7,7,7,7}?
  - Which data set has the largest standard deviation?
    - A {1, 3, 3, 3, 3, 3, 7}
    - B {1, 2, 3, 4, 5, 6, 7}
    - C {1, 1, 1, 4, 7, 7, 7}

5. Sample size.

Each record typically corresponds to an *observational unit*, and the number of observed units in the analysis is called the sample size, *n*.

In some situations, the population size might be known and you might have a Simple Random Sample (SRS) from the population. The sample size then is the number of people in your sample.
For instance, there are 4 million births every year in the United States.
Suppose we sample 1,000 of them at random from this population, and record for each pregnancy, the number of weeks of pregnancy, and the height, weight and gender of the baby at birth.
Here n = 1,000. Each baby is an observational unit.

6. Statistic and parameter.

A statistic is a numerical description of your sample. Another word for statistic is *random variable*. The sample is typically considered random, and if a different sample were obtained, then the statistic might be different.

A parameter, however, is a property of the whole population. If a different sample were obtained, the parameter would not change.

Parameters are properties of the population. Typically unknown. Represented by Greek letters (like μ or σ).
Statistics are properties of the sample.
Represented by Roman letters (like $\bar{x}$ or s).
Typically, you're interested in a value of a parameter. But you can't know it. So you *estimate* it with a statistic, based on the sample.

There are two means and two standard deviations.
The sample mean $\bar{x}$ and sample std deviation *s* are statistics.
Define the population average μ as the sum of all values in the population ÷ the number of subjects in the population. (parameter).
It turns out $\bar{x}$ is an unbiased estimate of μ.
That is, $\bar{x}$ is neither higher nor lower, on average, than μ, if we sampled repeatedly.

7. Categorical and quantitative variables.
For a quantitative variable, the responses are all numbers and the difference between two observations has a natural interpretable meaning. For categorical variables there is no such meaning to the difference between two observations.
The line between the two terms can sometimes be a bit blurry.
e.g. gender of baby would be categorical.
height, weight, and number of weeks would be quantitative.
eye color, birth type, or pain medication used might be examples of categorical variables here with multiple possibilities.

8. Statistical significance and Testing.

According to the CDC, 4 million babies were born in the U.S. in 2014 and 10% were born preterm (< 37 weeks). Suppose you take a simple random sample (SRS) of women with Hyperemesis Gravidarum (HG) and you want to test whether the proportion preterm among women with HG might really be different from 10%.

Suppose in the sample of n=254 mothers with HG, $\hat{p}$ = 39/254 (15.35%) are preterm.

You want to test whether something like this could reasonably have happened just by chance alone, if the populations were actually identical with respect to delivery time.

Otherwise we conclude that the two population proportions are probably not equal, i.e. the difference observed is *statistically significant*.

There are different tests, but we'll just talk about the Z-test (or normal test) for now.

Assumptions:

SRS (or obs are known to be independent)

AND n is large (or pop is known to be normally distributed).

For testing proportions, there should be ≥ 10 of each type of response in the sample.

Here we have 39 preterm and 215 not preterm.

We will talk later in the course about these assumptions and also about the t test. If n is small, pop. is normal, and σ is unknown, then use t instead of Z.

After checking assumptions, the remaining steps in testing are

* stating the hypotheses,
* computing the test statistic (Z in this case),
* computing the p-value, and
* concluding.

9. Null and alternative hypotheses.

Let π be the proportion preterm in the population from which the sample was drawn.

Null hypothesis (Ho): $\pi = 10\%$.

This means that any observed difference between the sample proportion, $\hat{p}$, and 10%, was due to chance alone. Usually we specify these hypotheses numerically.

Alternative hypothesis (Ha): $\pi \neq 10\%$. Difference is not due to chance alone. (2-sided test.)

Or Ha: $\pi > 10\%$. Or Ha: $\pi < 10\%$. (1-sided tests). We will talk about this next lecture.

When in doubt, do a two-sided test, unless there is a specific reason to do a 1-sided test.

1-sided versus 2-sided tests.

- On my tests, I will tell you explicitly whether to do a 1 or 2 sided test.
- On hw problems, you might have to decide whether to do a 1-sided or 2-sided test.
- With the hw, if in the problem you are given that you are only looking for evidence in one direction, then you do a 1-sided test. If you are looking for *any* difference in proportions, then do a 2-sided test.

10. Z-statistic.

A test statistic is a summary of the strength of the evidence in your data.

Z-statistic here = ($\hat{p}$ – 10%) ÷ SE.

SE means Standard Error. We will talk about ways to get the SE either analytically or via simulations in a bit. For proportion problems like this, SE = $\sqrt{\pi(1-\pi)/n}$.

Here, using the formula, the SE would be $\sqrt{0.1\ (1-0.1)/254}$ ~ 1.88%.

Z = (15.34% - 10%)/1.88% = 2.84.

The book calls Z a *standardized* test statistic.

It indicates how many SDs the observed statistic is above its hypothesized value under Ho.

The book also calls the SE the "standard deviation of the null distribution" but it is usually called the standard error or SE.

A value of Z far from 0 (more than 2 or less than -2) indicates strong evidence against the null hypothesis. A value of Z between -2 and 2 indicates weak evidence against the null.

|Z| > 3 indicates very strong evidence against the null.