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

1. Significance level.
2. Type I and Type II errors.
3. Power.
4. Confidence intervals, Marine sniffing cancer example.
5. 1.96SE and theory-based CIs for a proportion, ACA example.
6. 1.96SE and theory-based CIs for a mean, used car example.

http://www.stat.ucla.edu/~frederic/13/W24 .

Read chapter 4.

The hw2 problems are on the next 3 slides.

Needles

Exercises 2.3.15 and 2.3.16 refer to the needle data.

2.3.15 Consider a manufacturing process that is producing hypodermic needles that will be used for blood donations. These needles need to have a diameter of 1.65 mm—too big and they would hurt the donor (even more than usual), too small and they would rupture the red blood cells, rendering the donated blood useless. Thus, the manufacturing process would have to be closely monitored to detect any significant departures from the desired diameter. During every shift, quality control personnel take a sample of several needles and measure their diameters. If they discover a problem, they will stop the manufacturing process until it is corrected.

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

b. State the appropriate null and alternative hypotheses using the symbol defined in (a).

c. Describe what a Type I error would be in this study. Also, describe the consequence of such an error in the context of this study.

d. Describe what a Type II error would be in this study. Also, describe the consequence of such an error in the context of this study.

3.3.18 Reconsider the investigation of the manufacturing process that is producing hypodermic needles. Using the data from the most recent sample of needles, a 90% confidence interval for the average diameter of needles is found to be (1.62 mm, 1.66 mm). For each of the following statements, say whether VALID or INVALID.

a. We are 90% confident that the average diameter of the sample of 35 needles is between 1.62 and 1.66 mm.

b. Based on the 90% confidence interval, there is evidence that the average diameter of needles produced by this manufacturing process is 1.65 mm.

c. Based on the 90% confidence interval, there is evidence that the average diameter of needles produced by this manufacturing process is different from 1.65 mm.

d. We are 90% confident that the average diameter of needles produced by this manufacturing process is between 1.62 and 1.66 mm.

e. About 90% of the needles produced by this manufacturing process have a diameter between 1.62 and 1.66 mm.

f. If we want to be more than 90% confident, we should take a larger sample of needles.

Colds and exercise

4.1.23 In November 2010, an article titled "Frequency of Colds Dramatically Cut with Regular Exercise" appeared in *Medical News Today*. The article was based on the findings of a study by researchers Nieman et al. (*British Journal of Sports Medicine*, 2010) that followed 1,002 people aged 18–85 years for 12 weeks, asking them to record their frequency of exercise (5 or more days a week? Yes or No) as well as incidences of upper respiratory tract infections (Cold during last week? Yes or No).

a. Identify the explanatory variable in this study. Also classify this variable as categorical or quantitative.

b. Identify the response variable in this study. Also classify this variable as categorical or quantitative.

c. Identify a confounding variable that provides an alternative explanation for the lower frequency of colds among those who exercised 5 or more days per week, compared to those who were largely sedentary.

1. Significance level

Section 2.3

Significance Level

- We think of a p-value as telling us something about the strength of evidence from a test of significance.
- The lower the p-value the stronger the evidence.
- Some people think of this in more black and white terms. Either we reject the null or not.

Significance Level

- The value that we use to determine how small a p-value needs to be to provide convincing evidence whether or not to reject the null hypothesis is called the **significance level**.
- We reject the null when the p-value is less than or equal to (≤) the significance level.
- The significance level is often represented by the Greek letter alpha, α.

Significance Level

- Typically we use 0.05 for our significance level. There is nothing magical about 0.05. We could set up our test to make it
  - harder to reject the null (smaller significance level say 0.01) or
  - easier (larger significance level say 0.10).

2. Type I and Type II errors

- In medical tests:
  - A type I error is a false positive. (conclude someone has a disease when they don’t.)
  - A type II error is a false negative. (conclude someone does not have a disease when they actually do.)
- These types of errors can have very different consequences.

Type I and Type II Errors

TABLE 2.9 A summary of Type I and Type II errors

		What is true (unknown to us)	
		Null hypothesis is true	Null hypothesis is false
What we decide (based on data)	Reject null hypothesis	Type I error (false alarm)	Correct decision
	Do not reject null hypothesis	Correct decision	Type II error (missed opportunity)

•

Type I and Type II errors

TABLE 2.10 Type I and Type II errors summarized in context of jury trial

		What is true (unknown to the jury)	
		Null hypothesis is true (defendant is innocent)	Null hypothesis is false (defendant is guilty)
What jury decides (based on evi-dence)	Reject null hypothesis (Jury finds defendant guilty)	Type I error (false alarm)	Correct decision
	Do not reject null hypothe-sis (Jury finds defendant not guilty)	Correct decision	Type II error (missed opportunity)

The probability of a Type I error

- The significance level is the probability of a type I error, when Ho is true.
- Suppose the significance level is 0.05. If the null is true we would reject it 5% of the time and thus make a type I error 5% of the time.
- If you make the significance level lower, you have reduced the probability of making a type I error, but have increased the probability of making a type II error.

The probability of a Type II error

- The probability of a type II error is more difficult to calculate.
- In fact, the probability of a type II error is not even a fixed number. It depends on the value of the true parameter you are estimating.
- The probability of a type II error can be very high if:
  - The effect size is small.
  - The sample size is small.

3. Power

- The probability of rejecting the null hypothesis when it is false is called the **power** of a test.
- Power = 1 – P(Type II error). It is usually expressed as a function of μ.
- We want a test with high power and this is aided by:
  - A large effect size, i.e. true μ far from the parameter in the null hypothesis.
  - A large sample size.
  - A small standard deviation.
  - A higher significance level means greater power. The downside is that you get more type I errors.

4. Estimation and confidence intervals.

Chapter 3

Chapter Overview

- So far, we can only say things like
  - "We have strong evidence that the long-run frequency of death within 30 days after a heart transplant at St. George's Hospital is greater than 15%."
  - "We do not have strong evidence kids have a preference between candy and a toy when trick-or-treating."
- We want a method that says
  - "I believe 68 to 75% of all elections can be correctly predicted by the competent face method."

Confidence Intervals

- Interval estimates of a population parameter are called **confidence intervals**.
- We will find confidence intervals three ways.
  - Through a series of tests of significance to see which proportions are plausible values for the parameter.
  - Using the standard error (the standard deviation of the simulated null distribution) to help us determine the width of the interval.
  - Through traditional theory-based methods, i.e. formulas.

Statistical Inference: Confidence Intervals

Section 3.1

Can Dogs Sniff Out Cancer?

Section 3.1

Can Dogs Sniff Out Cancer?

Sonoda et al. (2011). Marine, a dog originally trained for water rescues, was tested to see if she could detect if a patient had colorectal cancer by smelling a sample of their breath.

- She first smells a bag from a patient with colorectal cancer.
- Then she smells 5 other samples; 4 from normal patients and 1 from a person with colorectal cancer
- She is trained to sit next to the bag that matches the scent of the initial bag (the “cancer scent”) by being rewarded with a tennis ball.

Can Dogs Sniff Out Cancer?

In Sonoda et al. (2011). Marine was tested in 33 trials.

- Null hypothesis: Marine is randomly guessing which bag is the cancer specimen (π = 0.20)
- Alternative hypothesis: Marine can detect cancer better than guessing (π > 0.20)

π represents her long-run probability of identifying the cancer specimen.

Can Dogs Sniff Out Cancer?

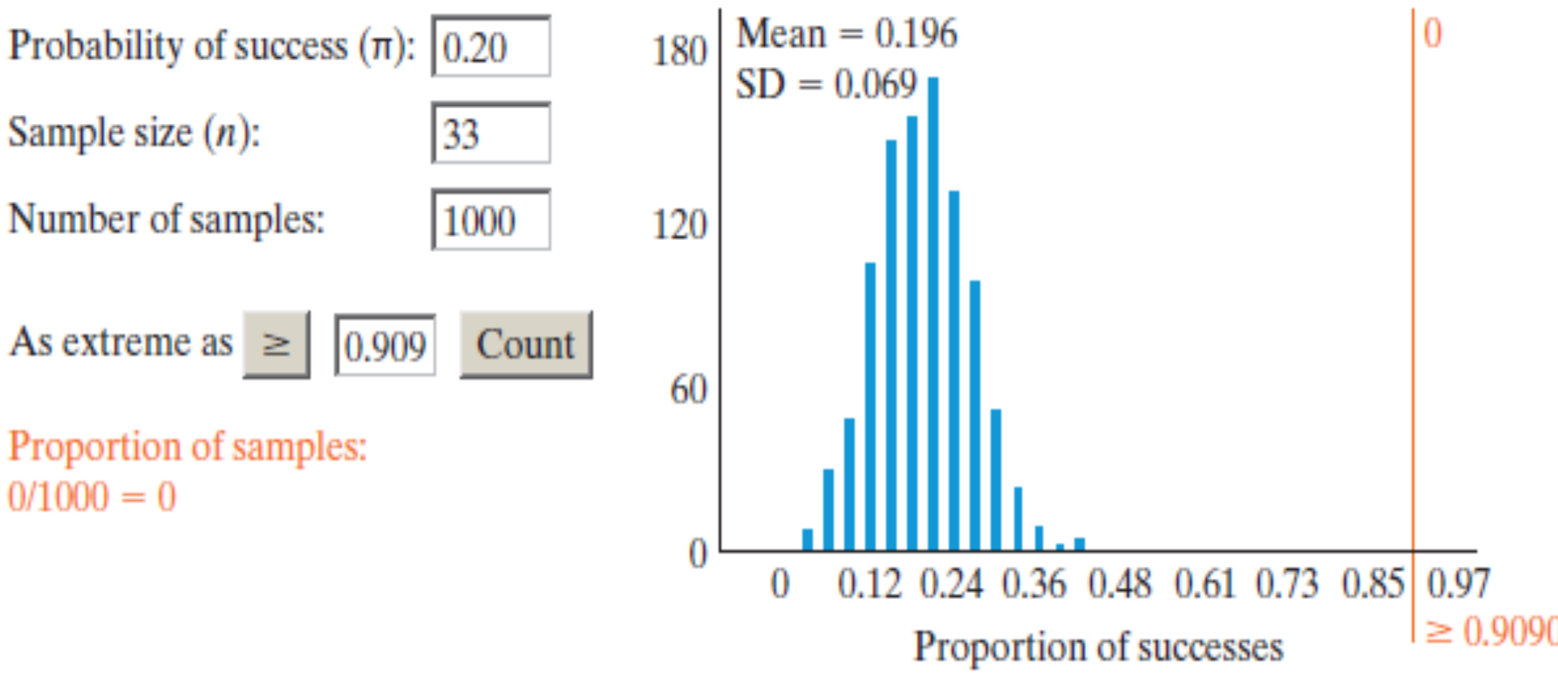
- 30 out of 33 trials resulted in Marine correctly identifying the bag from the cancer patient
- So our sample proportion is

$$\hat{p} = \frac{30}{33} \approx 0.909$$

- Do you think Marine can detect cancer?
- What sort of p-value will we get?

Can Dogs Sniff Out Cancer?

Our sample proportion lies more than 10 standard deviations above the mean and hence our p-value ~ 0.

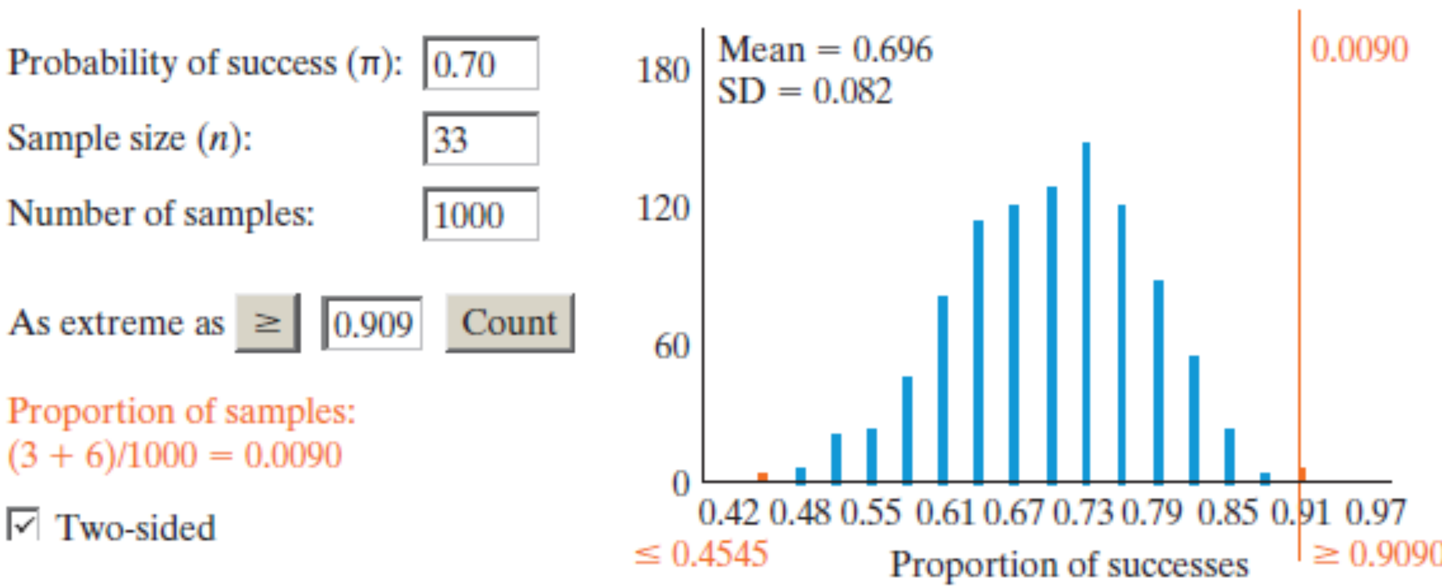


Can Dogs Sniff Out Cancer?

- Can we estimate Marine's long run frequency of picking the correct specimen?
- Since our sample proportion is about 0.909, it is plausible that 0.909 is a value for this frequency. What about other values?
- Is it plausible that Marine's frequency is actually 0.70 and she had a lucky day?
- Is a sample proportion of 0.909 unlikely if $\pi = 0.70$?

Can Dogs Sniff Out Cancer?

- H_0: $\pi = 0.70$ H_a: $\pi \neq 0.70$
- We get a small p-value (0.0090) so we can essentially rule out 0.70 as her long run frequency.

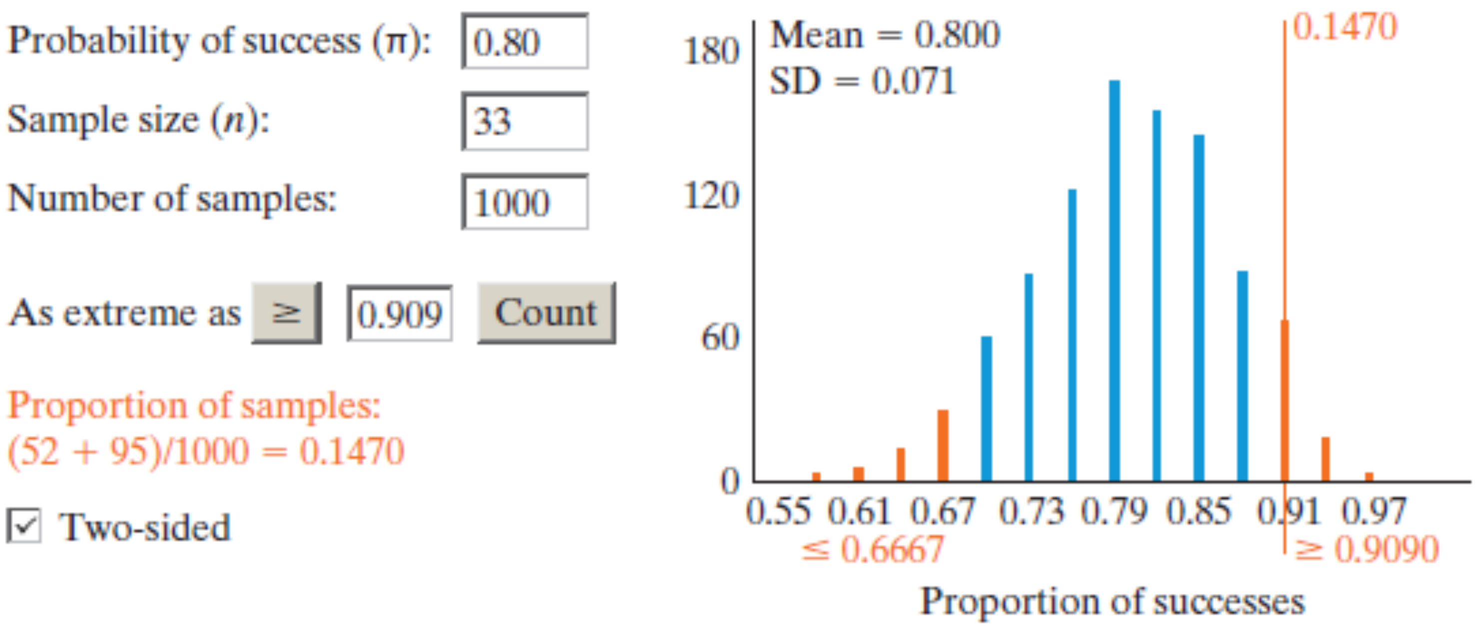


Can Dogs Sniff Out Cancer?

- What about 0.80?
- Is 0.909 unlikely if π = 0.80?

Can Dogs Sniff Out Cancer?

- H_0: $\pi = 0.80$ H_a: $\pi \neq 0.80$
- We get a large p-value (0.1470) so 0.80 is a ***plausible*** value for Marine's long-run frequency.



Developing a range of plausible values

- If we get a small p-value (like we did with 0.70) we will conclude that the value under the null is not plausible. This is when we reject the null hypothesis.
- If we get a large p-value (like we did with 0.80) we will conclude the value under the null is plausible. This is when we can't reject the null.

Developing a range of plausible values

- One could use software (like the one-proportion applet the book recommends) to find a range of plausible values for Marine's long term probability of choosing the correct specimen.
- We will keep the sample proportion the same and change the possible values of π.
- We will use 0.05 as our cutoff value for if a p-value is small or large. (Recall that this is called the **significance level**.)

Can Dogs Sniff Out Cancer?

- It turns out values between 0.761 and 0.974 are plausible values for Marine's probability of picking the correct specimen.

Probability under null	**0.759**	**0.760**	**0.761**	**0.762**	**……**	**0.973**	**0.974**	**0.975**	**0.976**
p-value	0.042	0.043	**0.063**	**0.063**		**0.059**	**0.054**	0.049	0.044
Plausible?	No	No	Yes	Yes	……… Yes	Yes	Yes	No	No

Can Dogs Sniff Out Cancer?

- (0.761, 0.974) is called a ***confidence interval*.**
- Since we used 5% as our significance level, this is a 95% confidence interval. (100% – 5%)
- 95% is the ***confidence level*** associated with the interval of plausible values.

Can Dogs Sniff Out Cancer?

- We would say we are 95% confident that Marine's probability of correctly picking the bag with breath from the cancer patient from among 5 bags is between 0.761 and 0.974.
- This is a more precise statement than our initial significance test which concluded Marine's probability was more than 0.20.
- Sidenote: We do not say $P\{\pi \text{ is in } (.761, .974)\} = 95\%$, because π is not random. The *interval* is random, and would change with a different sample. If we calculate an interval this way, then $P(\text{interval contains } \pi) = 95\%$.

Confidence Level

- If we increase the confidence level from 95% to 99%, what will happen to the width of the confidence interval?

Can Dogs Sniff Out Cancer?

- Since the confidence level gives an indication of how sure we are that we captured the actual value of the parameter in our interval,
  to be more sure our interval should be wider.
- How would we obtain a wider interval of plausible values to represent a 99% confidence level?
  - Use a 1% significance level in the tests.
  - Values that correspond to 2-sided p-values larger than 0.01 should now be in our interval.

5. 1.96SE and Theory-Based Confidence Intervals for a Single Proportion and ACA example.

Section 3.2

Introduction

- Previously we found confidence intervals by doing repeated tests of significance (changing the value in the null hypothesis) to find a range of values that were plausible for the population parameter.
- This is a very tedious way to construct a confidence interval.
- We will now look at two others way to construct confidence intervals [1.96SE and Theory-Based].

The Affordable Care Act

Example 3.2

The Affordable Care Act

- A November 2013 Gallup poll based on a random sample of 1,034 adults asked whether the Affordable Care Act had affected the respondents or their family.
- 69% of the sample responded that the act had no effect. (This number went down to 59% in May 2014 and 54% in Oct 2014.)
- What can we say about the proportion of all adult Americans that would say the act had no effect?

The Affordable Care Act

- We could construct a confidence interval just like we did last time. We get (0.661, 0.717).
- We are 95% confident that the proportion of all adult Americans that felt unaffected by the ACA is between 0.661 and 0.717.

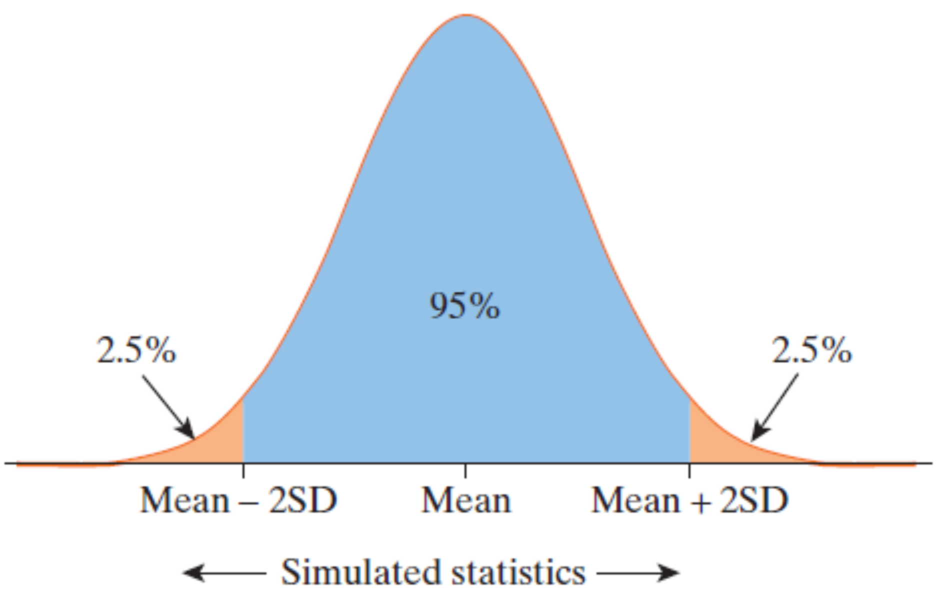
Probability under null	0.659	0.660	0.661	…………	0.717	0.718	0.719
Two-sided p-value	0.0388	0.0453	0.0514	…………	0.0517	0.0458	0.0365
Plausible value (0.05)?	No	No	Yes	…………	Yes	No	No

Short cut?

- The method we used last time to find our interval of plausible values for the parameter is tedious and time consuming.
- Might there be a short cut?
- Our sample proportion should be the middle of our confidence interval.
- We just need a way to find out how wide it should be.

1.96SE method

- When a statistic is normally distributed, about 95% of the values fall within 1.96 standard errors of its mean with the other 5% outside this region



1.96SE method

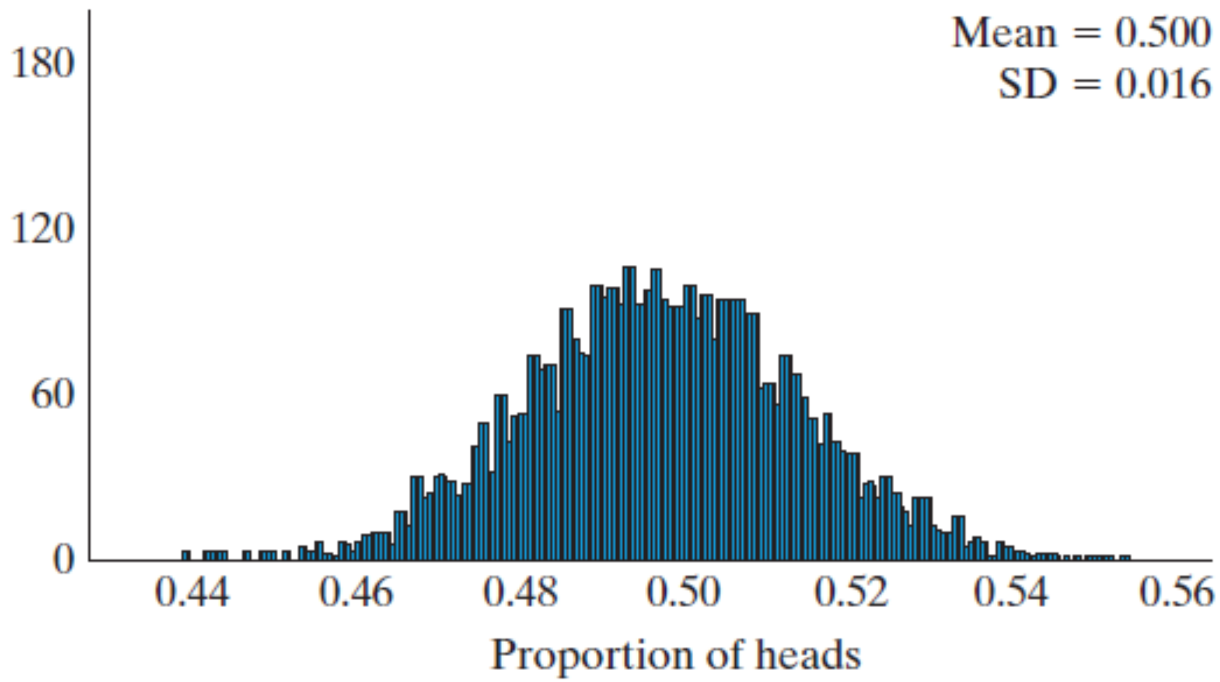
- So we could say that a parameter value is plausible if it is within 1.96 standard errors from our best estimate of the parameter, which is our observed sample statistic.
- This gives us the simple formula for a 95% confidence interval of

$$\boldsymbol{\hat{p} \pm 1.96SE}$$

Note that your book calls this the 2SD method but it really should be called the 1.96SE method.

Where do we get the SE?

- One way is via simulation.
- When the null hypothesis is π = 0.5, the SE = 0.016.



1.96SE method

- Using the 1.96SE method on our ACA data we get a 95% confidence interval

$$0.69 \pm 1.96(0.016)$$
$$0.69 \pm 0.031$$

- The $\pm$ part, like the 0.031 in the above, is called the **margin of error.**
- The interval can also be written as we did before using just the endpoints; (0.659, 0.721)
- This is approximately what we got using simulations, with our range of plausible values method. We had (0.661, 0.717).

Formula, or Theory-Based Method

- The 1.96SE method is for a 95% confidence interval.
- If we want a different level of confidence, we can use the range of plausible values (hard) or theory-based methods (easy).
- The theory-based method is valid for CIs for a proportion, provided it's a Simple Random Sample (SRS) and there are at least 10 successes and 10 failures in your sample.

FORMULA FOR CIs FOR A PROPORTION.

- On the previous slides, we relied on simulations to tell us that the SE was 0.016. But we don't need this. In general for testing a proportion, under the null hypothesis, SE = $\sqrt{\pi(1-\pi)/n}$.
- For confidence intervals, we do not assume the null hypothesis, and since π is unknown, use $\hat{p}$ in its place:

$$\hat{p} \pm \textit{multiplier} \times \sqrt{\hat{p}\,(1-\hat{p}\,)/n}.$$

For a 95% CI, the book suggests a multiplier of 2. Actually people use 1.96, not 2. This comes from a property of the normal distribution.

qnorm(.975) = 1.96.

qnorm(.995) = 2.58, the multiplier for a 99% CI.

- Going back to the ACA example, recall

69% of 1034 respondents were not affected.
With no default value of π, to get a 95% CI for $\widehat{p}$, use

$\widehat{p} \pm \textit{multiplier} \times \sqrt{\hat{p}(1-\hat{p})/n}$

= 69% ± 1.96 x $\sqrt{.69(1-.69)/1034}$

= 69% ± 2.82%.

With 2 instead of 1.96 it would be 69% ± 2.88%.

This is the formula we actually use for CIs for a proportion.

$$\widehat{p} \pm \textit{multiplier} \times \sqrt{\hat{p}(1-\hat{p})/n} \; .$$

To review, the book first explains how to get a CI by repeated testing, then using the "2 SE" method where the SE is found via simulation, then gives you this formula. But the formula is actually the correct answer. The others are approximations and require simulation.

6. 1.96SE and Theory-Based Confidence Intervals for a Single Mean and used car example.

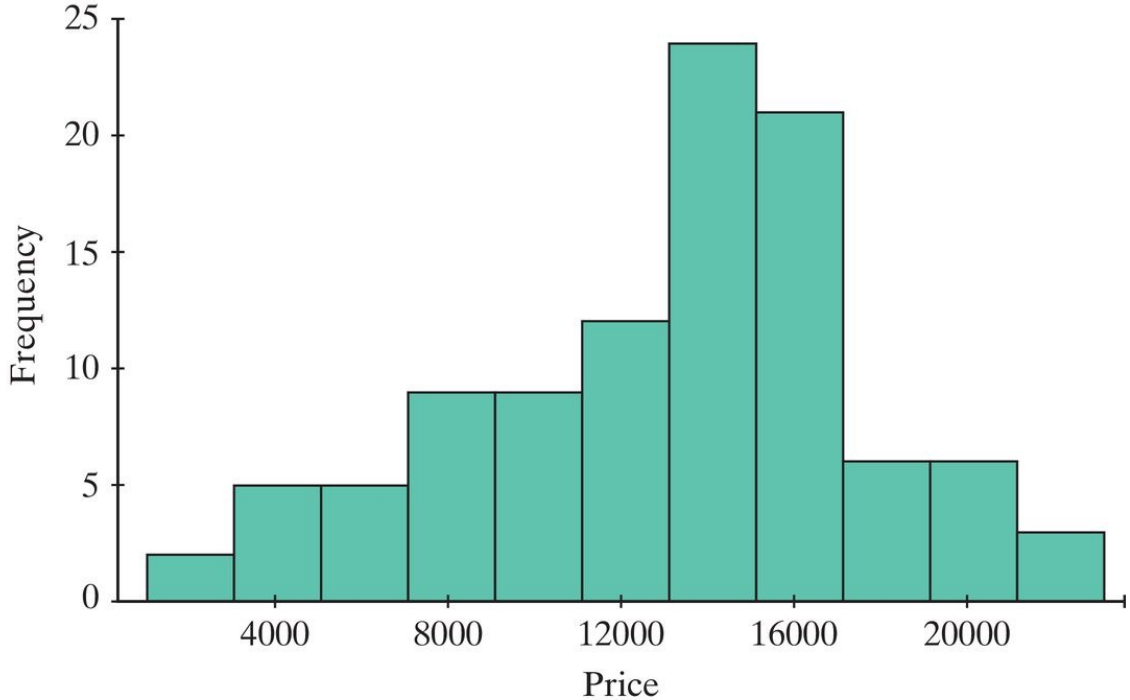
Section 3.3

Used Cars

Example 3.3

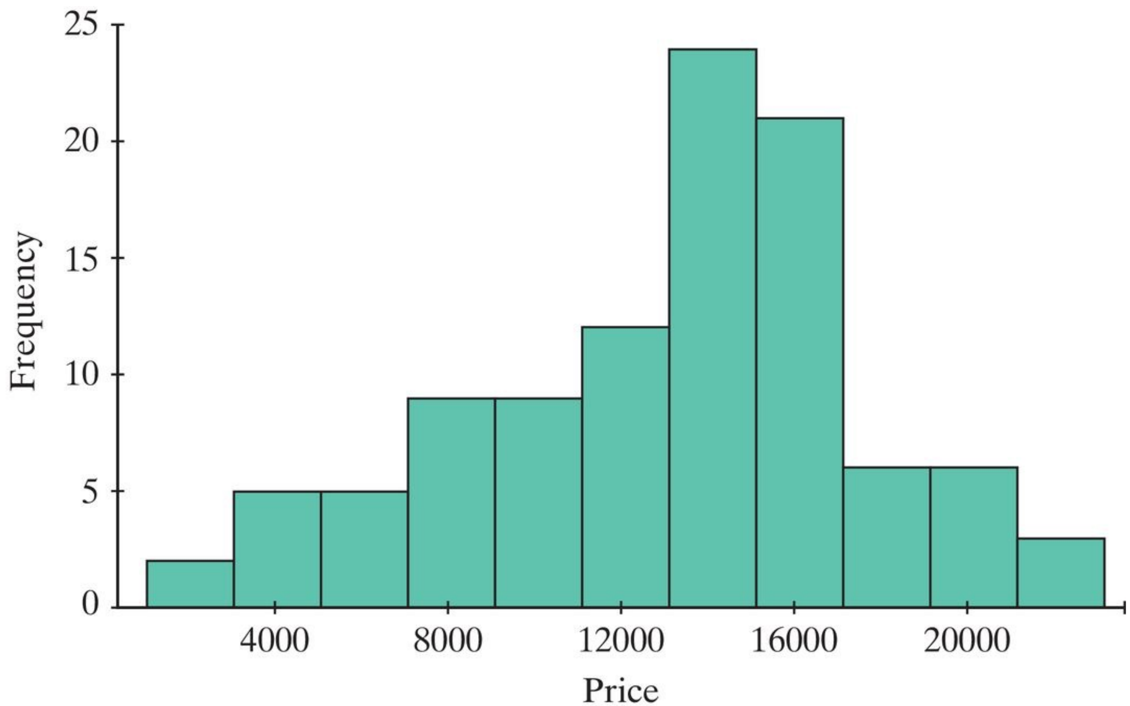
Used Cars

The following histogram displays data for the selling price of 102 Honda Civics that were listed for sale on the Internet in July 2006.



Used Cars

- The average of this sample is $\bar{x}$ = \$13,292 with a standard deviation of s = \$4,535.
- What can we say about μ, the average price of all used Honda Civics?



Used Cars

- While we should be cautious about our sample being representative of the population, let's treat it as such.
- μ might not equal \$13,292 (the sample mean), but it should be close.
- To determine how close, we can construct a confidence interval.

Confidence Intervals

- Remember the basic form of a confidence interval is:

statistic ± multiplier × SE

SE is called by the book "SD of statistic".

- In our case, the statistic is $\bar{x}$ and for large n, for a 95% CI our multiplier is 1.96, so we can write our 1.96SE confidence interval as:

$$\bar{x} \pm 1.96(\mathrm{SE})$$

Confidence Intervals

- It is important to note that the SE, which is the SD of $\bar{x}$, is not the same as the SD of our sample, s = $4,535.
- There is more variability in the data (the car-to-car variability) than in sample means.
- The SE is $s/\sqrt{n}$. Which means in general we can write a 1.96SE confidence interval for the mean as

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}}.$$

This 1.96 multiplier may be valid when n is large.

Summary Statistics

- When n is small and the population is approximately normal, we will use a multiplier that is based on a *t*-distribution, instead of 1.96. The t multiplier is dependent on the sample size and confidence level.
- For a theory-based confidence interval for a population mean (called a one-sample t-interval) to be valid, the observations should be approximately iid (independent and identically distributed), and either the population should be normal or n should be large. Check the sample distribution for skew and asymmetry.

Confidence Intervals

- We find our 95% CI for the mean price of all used Honda Civics is from $12,401.20 to $14,182.80.
- Notice that this is a much narrower range than the prices of all used Civics.
- For a 99% confidence interval, it would be wider. The multiplier would be 2.58 instead of 1.96.