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

1. 1.96SE and theory-based CIs for a single proportion and ACA example.
2. CIs for a single mean, used car example.
3. Factors affecting CI width, and CI meaning.
4. When to use which multiplier.
5. Statistical and practical significance, and longevity example.
6. Observational studies: association, confounding, and nightlights example.
7. Observational studies and experiments.
8. Experiments and aspirin example.
9. Random sampling and random assignment.
10. Blinding.

Reminder: lectures July 7 and 9 will be pre-recorded with links on course website.

The course website is http://www.stat.ucla.edu/~frederic/13/sum25
Read chapter 4.
HW2 is due Fri Jul11, 10pm. 2.3.15, 3.3.18, and 4.1.23.
Submit it to statgrader@stat.ucla.edu or statgrader2@stat.ucla.edu .
Midterm is Wed Jul16, 11am-12:50pm.

HW3 is due Fri Jul18, 10pm. 4.CE.10, 5.3.28, 6.1.17, and 6.3.14. In 5.3.28d, use the theory-based formula. You do not need to use an applet.

Spanking and IQ

4.CE.10 Studies have shown that children in the U.S. who have been spanked have a significantly lower IQ score on average than children who have not been spanked.

a. Is it legitimate to conclude from this study that spanking a child causes a lower IQ score? Explain why or why not.

b. Explain why conducting a randomized experiment to investigate this issue (of whether spanking causes lower IQs) would be possible in principle but ethically objectionable.

5.3.28 Recall the data from the Physicians' Health Study: Of the 11,034 physicians who took the placebo, 138 developed ulcers during the study. Of the 11,037 physicians who took aspirin, 169 developed ulcers.

a. Define the parameters of interest. Assign symbols to these parameters.

b. State the appropriate null and alternative hypotheses in symbols.

c. Explain why it would be okay to use the theory-based method (that is, normal distribution based method) to find a confidence interval for this study.

d. Use an appropriate applet to find and report the theory-based 95% confidence interval.

e. Does the 95% confidence interval contain 0? Were you expecting this? Explain your reasoning.

f. Interpret the 95% confidence interval in the context of the study.

g. Use the 95% confidence interval to state a conclusion about the strength of evidence in the context of the study.

h. Relatively speaking, is the 95% confidence interval narrow or wide? Explain why that makes sense.

5.3.29 Recall the data from the Physicians' Health Study:

every week. Be sure to compare and contrast the shape, center, and spread for study hours' distributions for males and females.

6.1.16 Reconsider the data in the previous question about number of hours spent studying.

a. Find the median number of study hours for both males and females. What do these numbers tell us about the two data sets?

b. Find the inter-quartile range for the number of study hours for both males and females. What do these numbers tell us about the two data sets?

c. Construct parallel boxplots by hand for the two data sets.

Gettysburg Address

6.1.17 The graph below displays the distribution of word lengths (number of letters) in the Gettysburg Address, which you explored in Exploration 2.1A.

a. Describe the shape of this distribution.

b. Based on this shape, do you expect the median to be less than the mean, greater than the mean, or very close to the mean? Explain.

The following table lists how often each of the word lengths appears for these 268 words.

Word length	1	2	3	4	5	6	7	8	9	10	11
Number of words	7	49	54	59	34	27	15	6	10	4	3

c. Determine the median word length of these 268 words.

d. The mean word length is 4.29 letters per word. Is the median greater than, less than, or very close to the mean? Does this confirm your answer to part (b)?

e. Calculate the five-number summary of the word lengths.

College student bedtimes*

6.1.18 In a survey, 30 college students were asked what their usual bedtime was and the results are shown in the 6.1.18 dotplot in terms of hours after midnight. Negative responses are hours before midnight.

a. Determine the five-number summary for the bed times.

b. What is the inter-quartile range?

c. The earliest bedtime is 11:30 PM (represented by −0.50 on the graph). If that person's usual bedtime is actually 9:00 PM and that change was made in the dotplot, does that change the inter-quartile range? Would it change the standard deviation?

Candy bars

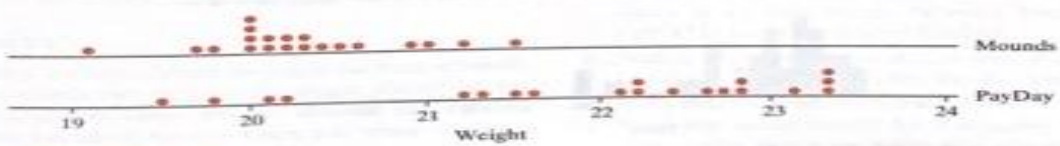
6.1.19 Weights of 20 Mounds* candy bars and 20 PayDay* candy bars, in grams, are shown in the 6.1.19 dotplots.

a. Describe how the distributions of weights of the two types of candy bars differ in both variability and center.

b. Based on your answers to part (a), which set of candy bar weights has the lowest standard deviation? Which has the lowest mean?

c. Would you say there is an association between the type of candy bar and the weight? Why or why not?

EXERCISE 6.1.18



EXERCISE 6.1.19

h. Summarize your conclusions about the research question of the study. Be sure to comment on statistical significance, confidence/estimation, causation, and generalization.

Perceived wealth

6.3.13 Do people tend to spend money differently based on perceived changes in wealth? In a study conducted by Epley et al. (2006), 47 Harvard undergraduates were randomly assigned to receive either a "bonus" check of $50 or a "rebate" check of $50. A week later, each student was contacted and asked whether they had spent any of that money, and if yes, how much. In this exercise we will focus on how much money they recalled spending when contacted a week later. It turned out that those in the "bonus" group spent an average of about $22, compared to $10 in the "rebate" group.

a. Identify the observational units.

b. Identify the explanatory and response variables. Identify each as either categorical or quantitative.

c. State the appropriate null and alternative hypotheses in the context of the study.

d. In the article that appeared in the *Journal of Behavioral Decision Making*, the researchers reported neither the sample size nor the sample SD of each group. In this exercise you will explore whether and how the strength of evidence is impacted by the sample size and sample SD. Complete the following table by finding the t-statistic and a p-value for a theory-based test of significance comparing two means under each of the four different scenarios.

e. Summarize what your analysis has revealed about the effects of the sample size breakdown and the sample standard deviations on the values of the t-statistic and p-value.

Nostril breathing and cognitive performance*

6.3.14 In an article titled "Unilateral Nostril Breathing Influences Lateralized Cognitive Performance" that appeared in *Brain and Cognition* (1989), researchers Block et al. published results from an experiment involving assessments of spatial and verbal cognition when breathing through only the right versus left nostril.

The subjects were 30 male and 30 female right-handed introductory psychology students who volunteered to participate in exchange for course credit. Initial testing on spatial and verbal tests revealed the following summary statistics. Note that the scores on the spatial task can range from 0 to 40, whereas those on the verbal task can go from 0 to 20. The distributions are not strongly skewed on either scale or for males or females.

Sex	Spatial		Verbal	
	Mean	SD	Mean	SD
Male	10.20	2.70	10.90	3.00
Female	7.80	2.50	15.10	3.40

a. Consider comparing males to females with regard to performance on the spatial assessment task. State the appropriate null and alternative hypotheses in the context of the study.

b. Explain why it is valid to use the theory-based method for producing a p-value to test the hypotheses stated in part (a).

c. Carry out the appropriate test to produce a p-value to test the hypotheses stated in part (a) and interpret the p-value.

d. Find a 95% confidence interval for the difference in mean scores of males and females with regard to performance on spatial assessments. Interpret the interval.

e. Based on your p-value, state a conclusion in the context of the study. Be sure to comment on statistical significance, estimation (confidence interval), causation, and generalization.

f. Repeat the investigation comparing males and females, this time on verbal performance. Be sure to address the questions asked in parts (a)–(e).

Scenario		Sample sizes	Sample means	Sample SDs	t-statistic	p-value
1	Bonus	24	22	5		
	Rebate	23	10	5		
2	Bonus	24	22	10		
	Rebate	23	10	10		
3	Bonus	30	22	5		
	Rebate	17	10	5		
4	Bonus	30	22	10		
	Rebate	17	10	10		

EXERCISE 6.3.13

1. 1.96SE and Theory-Based Confidence Intervals for a Single Proportion and ACA example.

Section 3.2

Introduction

- Section 3.1 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 (long run probability or population proportion).
- 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 find we are 95% confident that the proportion of all adult Americans that felt unaffected by the ACA is between 0.661 and 0.717.

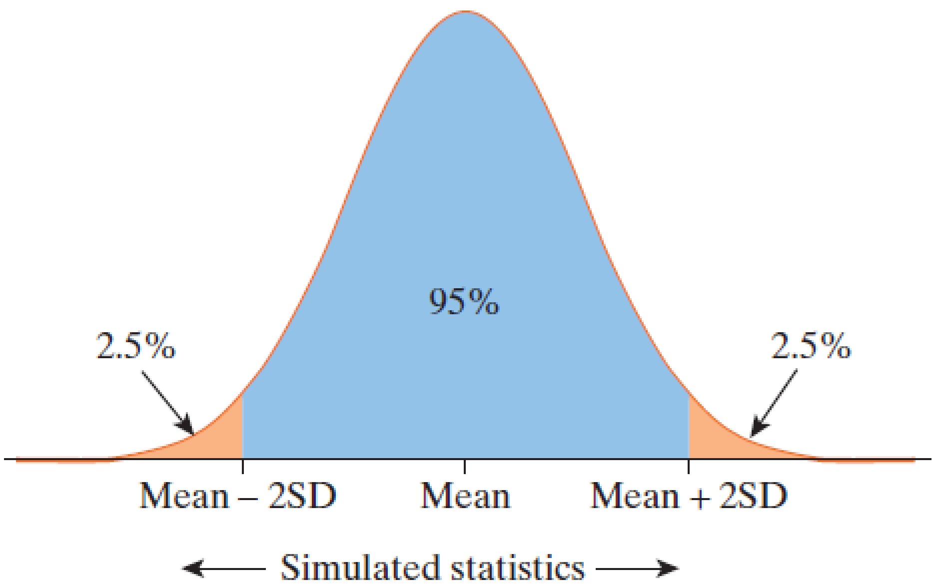
Probability under null	0.659	0.660	0.661	…………	0.717	0.718	0.719
Two-sided p-value	0.0388	0.0453	0.0514	…………	0.0517	0.0458	0.0365
Plausible value (0.05)?	No	No	Yes	…………	Yes	No	No

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

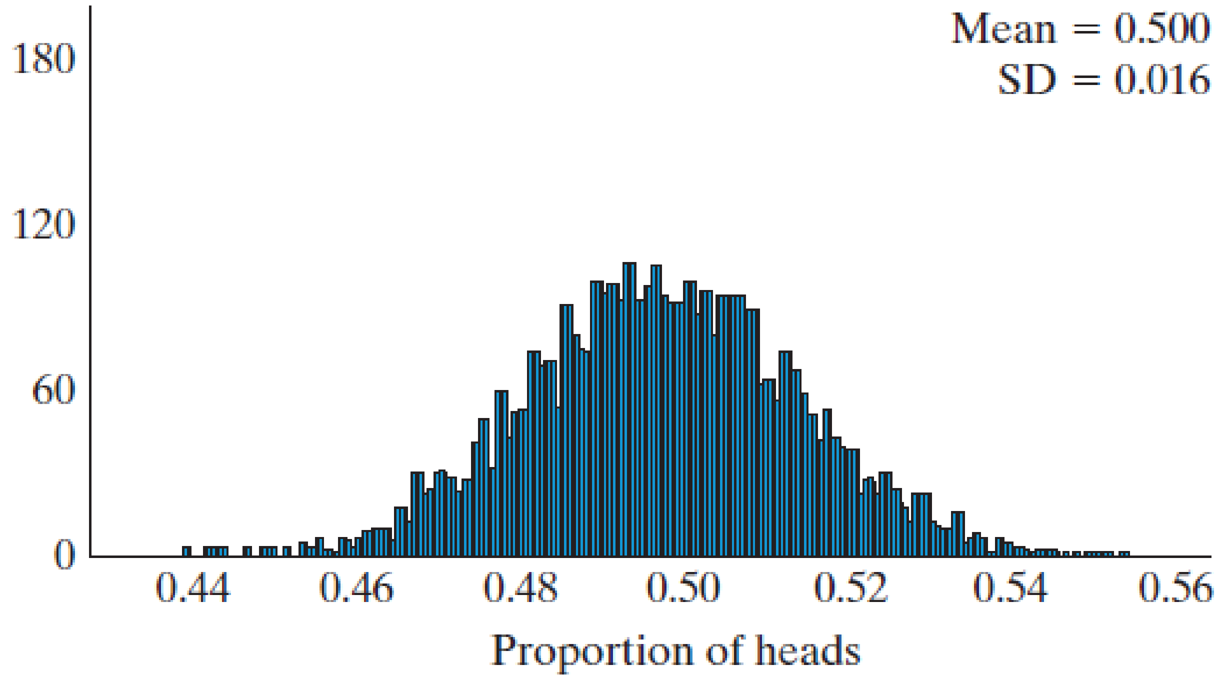
- So we could say that a parameter value is plausible if it is within 1.96 standard errors from our best estimate of the parameter, our observed sample statistic.
- This gives us the simple formula for a 95% confidence interval of

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

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?

- Null distribution for ACA with π = 0.5.



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 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 with our range of plausible values method (a bit wider).

Theory-Based Methods

- The 1.96SE method only gives us 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 provided there are at least 10 successes and 10 failures in your sample.

Theory-Based Methods

- With theory-based methods we use normal distributions to approximate our simulated null distributions.
- Therefore we can develop a formula for confidence intervals.

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

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.

- Let’s check out this example using the theory-based method.
- Remember 69% of 1034 respondents were not affected.

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

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

$= 69\% \pm 2.82\%.$

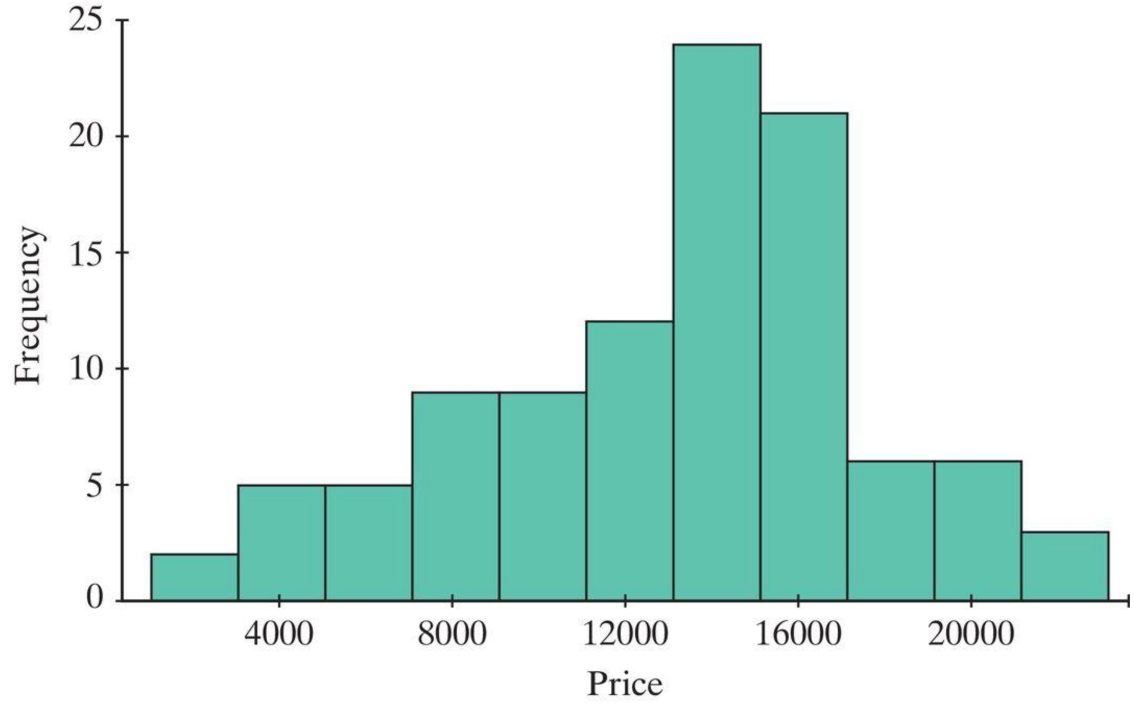
With 2 instead of 1.96 it would be 69% ± 2.88%.

2. Used Cars

Example 3.3

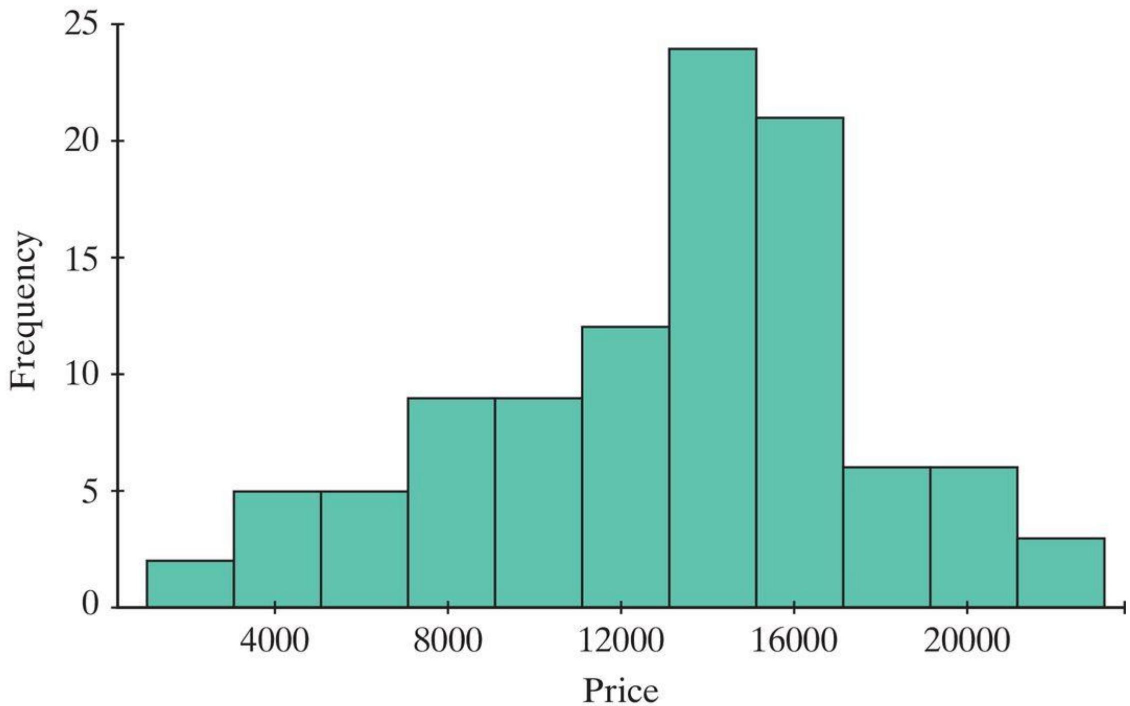
Used Cars

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 × (SD of statistic)

  SD of statistic is also called Standard Error (SE).
- In our case, the statistic is $\bar{x}$ and for a 95% CI our multiplier is 1.96, so we are 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 we can write a 1.96SE confidence interval as:

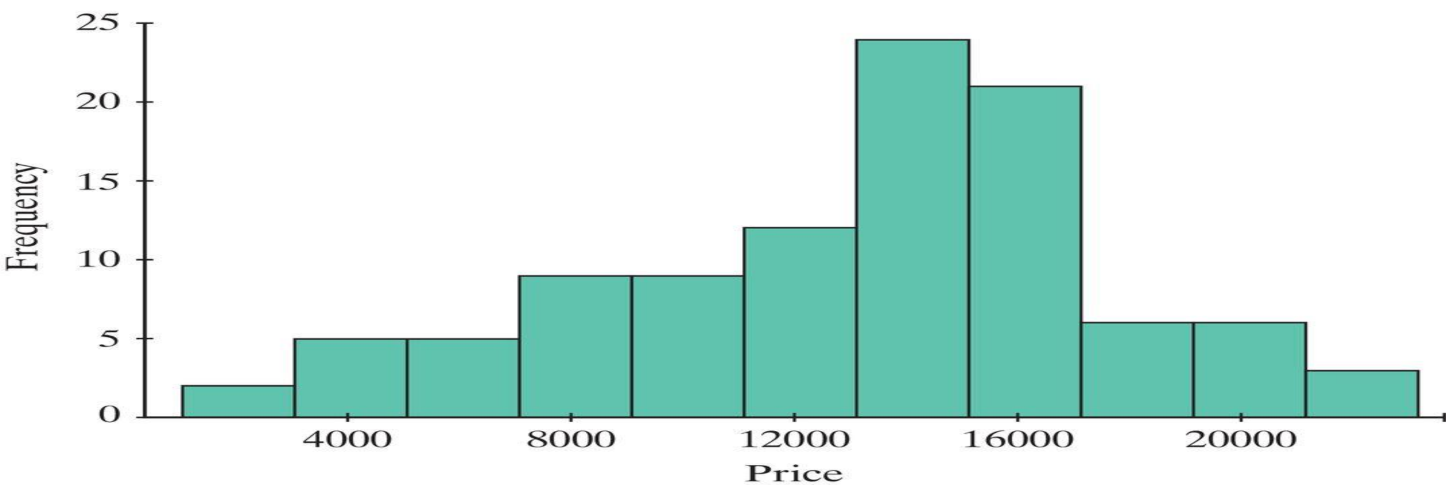
$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}}$$

Summary Statistics

- In many circumstances 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.



3. Factors that Affect the Width of a Confidence Interval

Section 3.4

Factors Affecting Confidence Interval Widths

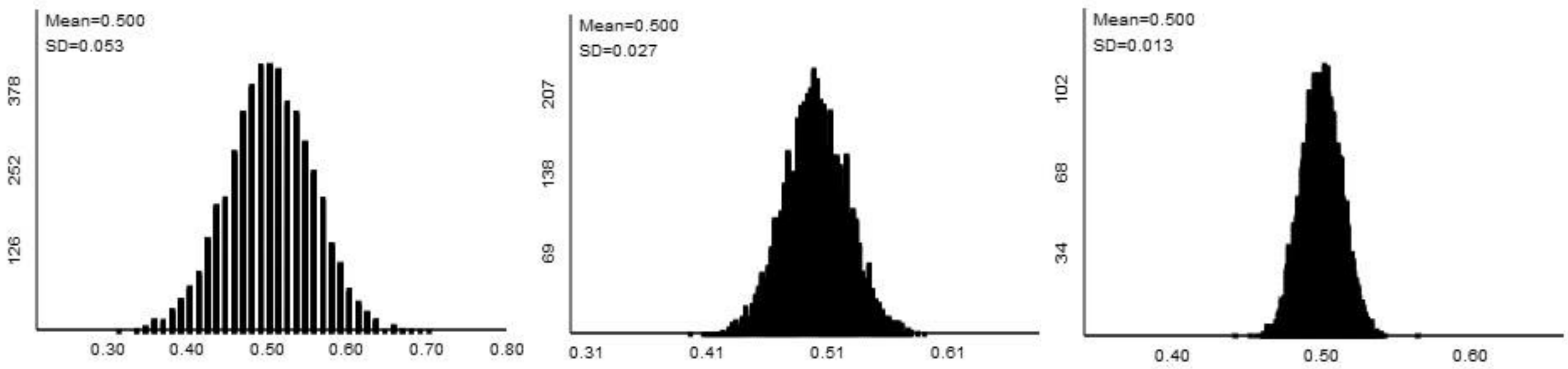
- **Level of confidence** (e.g., 90% vs. 95%)
  - As we increase the confidence level, we increase the width of the interval.
- **Sample size**
  - As sample size increases, variability decreases and hence the standard error will be smaller. This will result in a narrower interval.
- **Sample standard deviation**
  - A larger standard deviation, s, will yield a wider interval.
  - For sample proportions, wider intervals when $\hat{p}$ is closer to 0.5. $s = \sqrt{[\hat{p}(1-\hat{p})]}$.

Level of Confidence

- If we have a wider interval, we should be more confident that we have captured the population proportion or population mean.
- We could see this with repeated tests of significance.
  - A higher confidence level corresponds to a lower significance level, and one must go farther to the left and farther to the right in our tables to get our confidence interval.

Sample Size

- We know as sample size increases, the variability (and thus standard deviation) in our null distribution decreases



n = 90 (SD = 0.054) n = 361 (SD = 0.026) n = 1444 (SD = 0.013)

Sample size	90	361	1444
SD of null distr.	0.053	0.027	0.013
Margin of error	2 x SD = 0.106	2 × SD = 0.054	2 × SD = 0.026
Confidence interval	(0.091, 0.303)	(0.143, 0.251)	(0.171, 0.223)

Sample Size

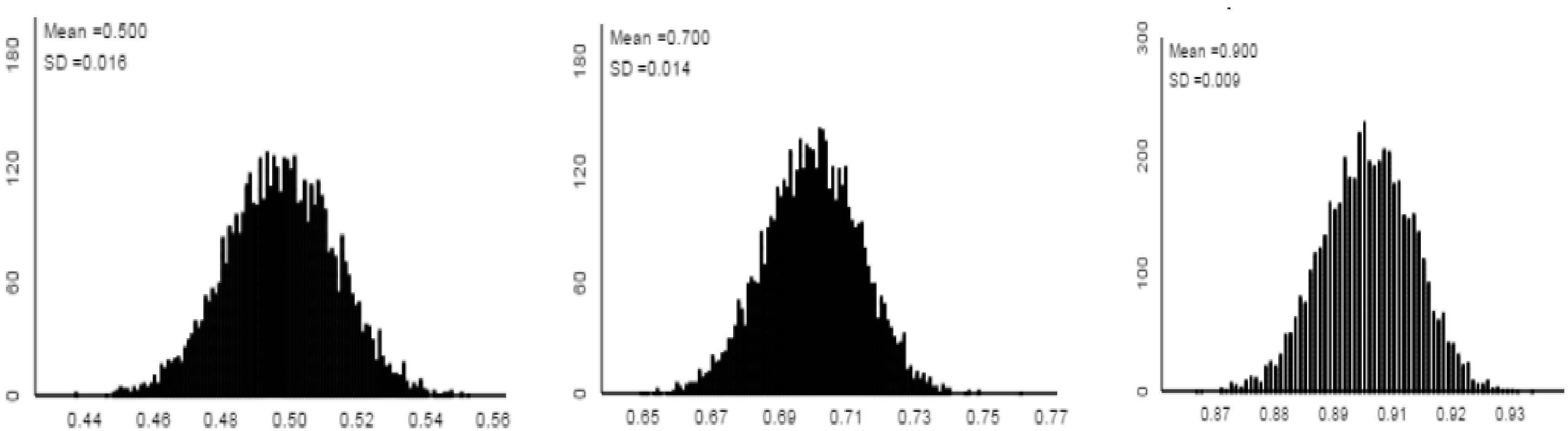
- (With everything else staying the same) increasing the sample size will make a confidence interval narrower.

Notice:

- The observed sample proportion is the midpoint. (that won't change)
- The margin of error is a multiple of the standard deviation so as the standard deviation decreases, so will the margin of error.

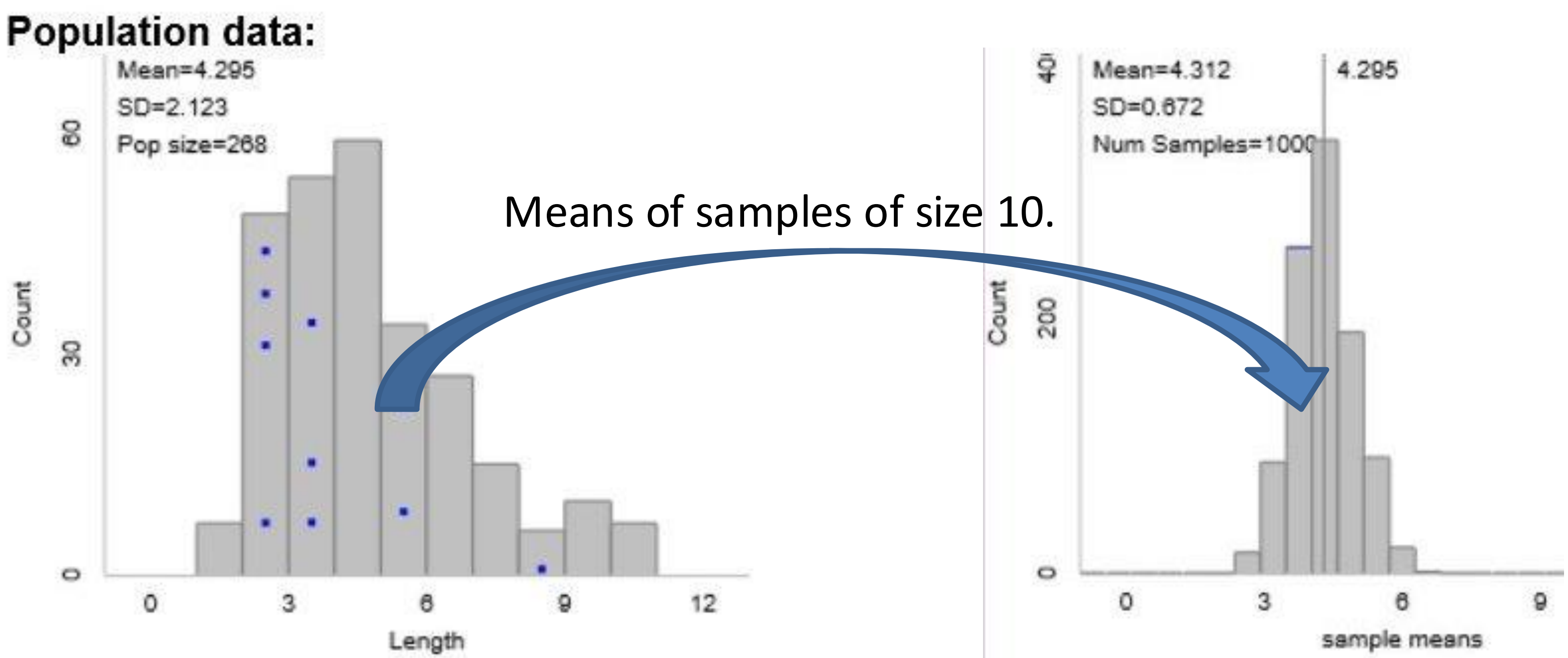
Value of $\hat{p}$
(or the value used for π under the null)

- As the value that is used under the null gets farther away from 0.5, the standard error decreases.
- When this standard error is used in the 1.96 SE method, the interval gets gradually narrower.



Standard Deviation

- Suppose we are taking repeated samples of a population.
- How do we estimate what the standard error (standard deviation of the null distribution) will be? $s/\sqrt{n}$.



Standard Deviation

- The SE (SD of the null distribution) is approximated by $s/\sqrt{n}$.
- Remember that $1.96\, s/\sqrt{n}$ is the **margin** of error for a 95% confidence interval, so as the standard deviation of the sample data, s, increases so does the width of the confidence interval.
- Intuitively this should make sense, more variability in the data should be reflected by a wider confidence interval.

Formulas for Theory-Based Confidence Intervals

$$\hat{p} \pm multiplier \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \bar{x} \pm multiplier \times \frac{s}{\sqrt{n}}$$

- The width of the confidence interval increases as level of confidence increases (multiplier)
- The width of the confidence interval decreases as the sample size increases
- The value $\hat{p}$ also has a more subtle effect. The farther it is from 0.5 the smaller the width.
- The width of the confidence interval increases as the sample standard deviation increases.

What does 95% confidence mean?

- If we repeatedly sampled from a population and constructed 95% confidence intervals, 95% of our intervals will contain the population parameter.
- Notice the interval is the random event here.

What does 95% confidence mean?

- Suppose a 95% confidence interval for a mean is 2.5 to 4.3. We would say we are 95% confident that the population mean is between 2.5 and 4.3.
  - Does that mean that 95% of the data fall between 2.5 and 4.3?
    - No
  - Does that mean that in repeated sampling, 95% of the sample means will fall between 2.5 and 4.3?
    - No
  - Does that mean that there is a 95% chance the population mean is between 2.5 and 4.3?
    - Not quite but close.

What does 95% confidence mean?

- What does it mean when we say we are 95% confident that the population mean is between 2.5 and 4.3?
  - It means that if we repeated this process (taking random samples of the same size from the same population and computing 95% confidence intervals for the population mean) repeatedly, 95% of the confidence intervals we find would contain the population mean.
  - P(confidence interval contains μ) = 95%.

4. For CIs, when to use 1.96 from the normal, & when to use a multiplier based on the t distribution.

iid = independent and identically distributed.

if the observations are iid. and n is large, then

$P(\mu$ is in the range $\bar{x}$ +/- $1.96\ \sigma/\sqrt{n}) \sim 95\%$.

If the observations are iid and normal, and σ is known, then

$P(\mu$ is in the range $\bar{x}$ +/- $1.96\ \sigma/\sqrt{n}) \sim 95\%$.

If the obs. are iid and normal and σ is unknown, then

$P(\mu$ is in the range $\bar{x}$ +/- $t_{mult}\ s/\sqrt{n}) \sim 95\%$.

where t_{mult} is the multiplier from the t distribution.

This multiplier depends on n.

5. Statistical and Practical significance.

- *Statistically significant* means that the results are unlikely to happen by chance alone.
- *Practically important* means that the difference is large enough to matter in the real world.

Cautions

- Practical importance is context dependent and somewhat subjective.
- Well designed studies try to equate statistical significance with practical importance, but not always.
- Look at the sample size.
  – If very large, expect significant results.
  – If very small, don’t expect significant results. (A lot of missed opportunities---type II errors.)

Longevity example.

According to data from the WHO (2014) and World Cancer Report (2014), the average number of cigarettes smoked per adult per day in the U.S. is 2.967, and in Latvia it is 2.853.

The sample sizes are huge, so even this little difference is stat. sig. (In the U.S., the National Health Interview Survey has $n > 87000$).

If you do not like cigarette smoke around you, should you move to Latvia?

The difference is statistically significant, but not practically significant for most purposes.

Causation.

Chapter 4

Big Idea of Chapter 4

- Previously research questions focused on **one** proportion
  - What proportion of the time did Marine choose the right bag?
- We will now start to focus on research questions comparing **two** groups.
  - Are smokers more likely than nonsmokers to have lung cancer?
  - Are children who used night lights as infants more likely to need glasses than those who didn’t use night lights?

Big Idea of Chapter 4

- Typically we observe two groups and we also have two variables (like smoking and lung cancer).
- So with these comparisons, we will:
  - determine when there is an association between our two variables.
  - discuss when we can conclude the outcome of one variable causes a change in the other.

6. Observational studies and confounding.
Types of Variables

- When two variables are involved in a study, they are often classified as explanatory and response
- **Explanatory variable** (Independent, Predictor)
  - The variable we think may be causing or explaining or used to predict a change in the response variable. (Often this is the variable the researchers are manipulating.)
- **Response variable** (Dependent)
  - The variable we think may be being impacted or changed by the explanatory variable.
  - The one we are interested in predicting.

Roles of Variables

- Choose the explanatory and response variable:
  - Smoking and lung cancer
  - Heart disease and diet
  - Hair color and eye color

- Sometimes there is a clear distinction between explanatory and response variables and sometimes there isn't.

Observational Studies

- In observational studies, researchers *observe* and measure the explanatory variable but do not set its value for each subject.
- Examples:
  - A significantly higher proportion of individuals with lung cancer smoked compared to same-age individuals who don’t have lung cancer.
  - College students who spend more time on Facebook tend to have lower GPAs.

Do these studies prove that smoking *causes* lung cancer or Facebook *causes* lower GPAs?

Night Lights and Nearsightedness

Example 4.1

Nightlights and Near-Sightedness

- Near-sightedness often develops in childhood
- Recent studies looked to see if there is an association between near-sightedness and night light use with infants
- Researchers interviewed parents of 479 children who were outpatients in a pediatric ophthalmology clinic
- Asked whether the child slept with the room light on, with a night light on, or in darkness before age 2
- Children were also separated into two groups: near-sighted or not near-sighted based on the child's recent eye examination

Night-lights and near-sightedness

	Darkness	Night Light	Room Light	Total
Near-sighted	18	78	41	137
Not near-sighted	154	154	34	342
Total	172	232	75	479

The largest group of near-sighted kids slept in rooms with night lights. It might be better to look at the data in terms of proportions.

Conditional proportions

$18/172 \approx 0.105$ $78/232 \approx 0.336$ $41/75 \approx 0.547$

Night lights and near-sightedness

	Darkness	Night Light	Room Light	Total
Near-sighted	**10.5%** 18/172	**33.6%** 78/232	**54.7%** 41/75	137
Not near-sighted	154	154	34	342
Total	172	232	75	479

- Notice that as the light level increases, the percentage of near-sighted children also increases.
- We say there is an **association** between near-sightedness and night lights.
- Two variables are **associated** if the values of one variable provide information (help you predict) the values of the other variable.

Night lights and near-sightedness

- While there is an association between the lighting condition and nearsightedness, can we claim that night lights and room lights *caused* the increase in near-sightedness?

- Might there be other reasons for this association?

Night lights and near-sightedness

- Could parents' eyesight be another explanation?
  - Maybe parents with poor eyesight tend to use more light to make it easier to navigate the room at night and parents with poor eyesight also tend to have children with poor eyesight.
  - Now we have a third variable of *parents' eyesight*
  - *Parents' eyesight* is considered a ***confounding variable.***
  - Other possible confounders? Wealth? Books? Computers?

Confounding Variables

- A **confounding variable** is associated with both the explanatory variable and the response variable.
- We say it is confounding because its effects on the response cannot be separated from those of the explanatory variable.
- Because of this, we can’t draw cause and effect conclusions when confounding variables are present.

Confounding Variables

- Since confounding variables can be present in observational studies, we can’t conclude causation from these kinds of studies.
- This doesn’t mean the explanatory variable isn’t influencing the response variable. **Association may not imply causation, but can be a pretty big hint.**

7. Observational studies versus Experiments

Section 4.2

Observational Studies vs. Experiments

- In an **observational study**, the researchers do not set the level of the explanatory variable for each subject. Typically each subject herself decides her level of the explanatory variable. Sometimes nature decides.
- For example, the researchers didn't control which children slept with a night light on or not.
- Observational studies always have potential confounding variables present and these may prevent us from determining cause and effect.

Observational Studies vs. Experiments

- In an **experiment**, the researchers set the level of the explanatory variable for each subject.
- These levels may correspond to a treatment and control.
- Well designed experiments can control for confounding variables by making the treatment and control groups very similar except for what the experimenter manipulates.

8. Experiments and aspirin example.

Physicians' Health Study I (study aspirin's affect on reducing heart attacks.

- Started in 1982 with 22,071 male physicians.
- The physicians were **randomly assigned** into one of two groups.
    - Half took a 325mg aspirin every other day and half took a placebo.