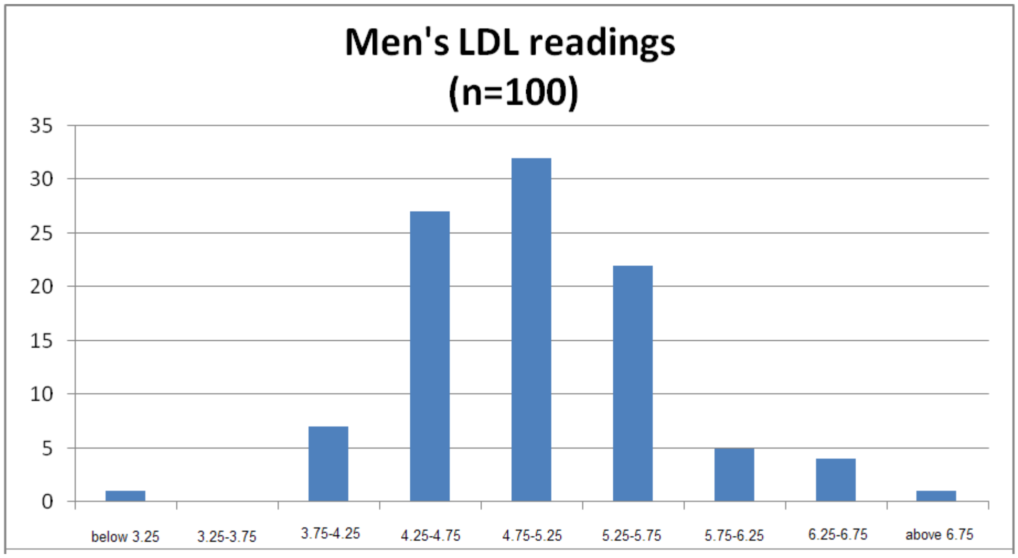
Single Mean CI Exercises

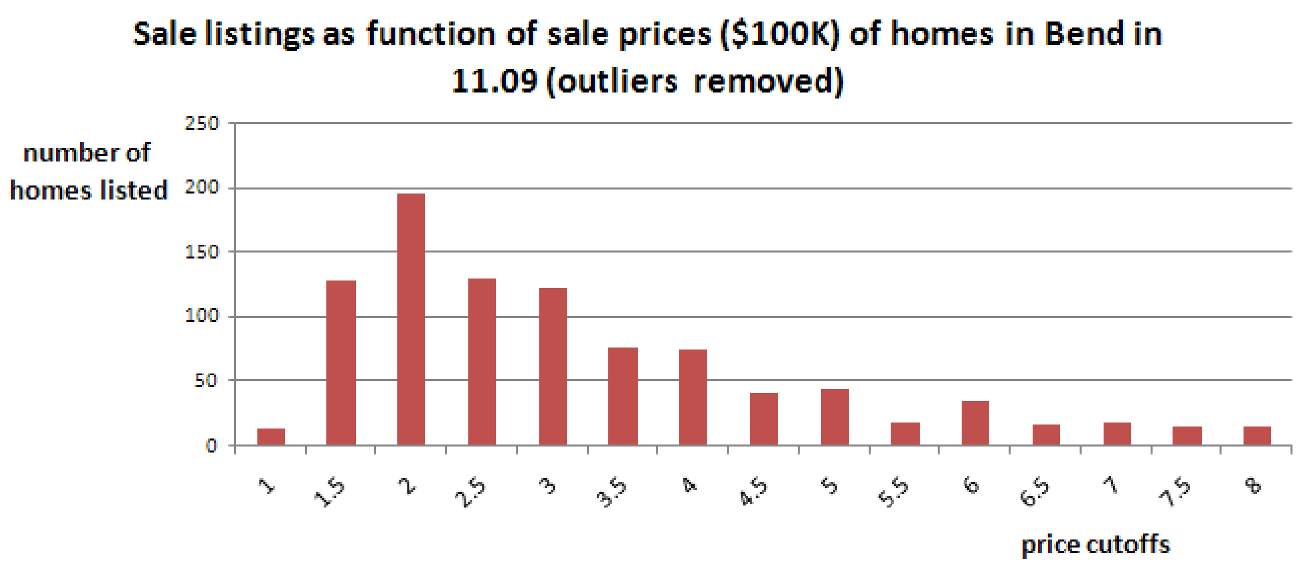
Pages	Suggested Reading
328 – 331	Section 8.3

Pages	Problems
343 – 351	(***Section 8.9***) (whenever possible, use your TI to draw a histogram of your data) 3 (b,c,d,e), 12(b,c,d), 27-29

E1. Construct the 95% CI for men's LDL ("bad" cholesterol) readings, assuming that the 100 readings at right are from a representative sample of the male population. For "below 3.25", use a data value of 3, and for "above 6.75", use 7.



E2. Suppose a cholesterol – lowering drug claims that it can lower LDL levels in men significantly. As evidence, it presents a study where those who took the drug have cholesterol levels that averaged 4.5 (whereas, the general population averaged 5, as you can see from the graph above). What do you make of their claim?



E3. Construct the 95% CI for the price of a Bend home in November of 2009 (when the data above was collected) assuming that the homes in the above chart are representative of all Bend homes.

E4. In 2008, after many, many emails to the City of Bend, I was supplied with the following speeds of cars (taken from police radar along Archie Briggs Road, where speed limit is 30 mph).

42	40	33	27	47	38	44	26
39	28	38	47	39	36	43	34
36	35	40	39	25	42	29	33
43	21	27	32	47	45	31	32
34	31	41	27	23	30	34	19

Does this data imply that the cars are obeying the speed limit of 30 miles per hour?

And, at last, a somewhat ethereal – seeming, perhaps mystical – feeling (but, still, very important) point: one of so – called ***"degrees of freedom"*** (usually shortened to "DF"). In this course (and in others in your future, if your future takes you further down the statistical path), whenever you use a distribution other than the **z** (standard normal) distribution, in the background, the DF are at play. The mathematics behind many of the definitions of degrees of freedom are many and varied and, often, confusing. Fortunately, you don't ever need to calculate DF directly (thank you, technology), but I wanted you to get an idea of just why their presence in the background is so vital[1].

The idea of "degrees of freedom" is usually described in a confusing manner in textbooks (yours is no different), so I'm going to keep it simple. Here's what you should know (and when I say "know", that does not imply "memorize"…it implies "understand"):

1) DF are always tied, somehow, to sample sizes. Larger sample sizes (assuming the data is good) are always preferential to smaller ones, so a distribution with more DF is better (all else being equal) than one with fewer.

2) DF are the number that's left once you subtract the number of parameters you're trying to estimate from your original sample size. For example, suppose you have randomly selected n data points, and have calculated their sample mean. You're obviously trying to estimate μ (that's what this whole section is about, right?). So, therefore, you have $n - 1$ degrees of freedom (which, not coincidentally, is why the denominator of s is $n - 1$).

3) Suppose, to illustrate the final point, you sampled n data points, and calculated $\bar{x}$ and **s** (thus, you've estimated the parameters μ and σ). Logical enough. Now, suppose you wanted to go and test something further…at this point, how many degrees of freedom do you have? It has to be $n - 2$, doesn't it? You have one less DF than the previous example because, once you have the data described two related ways (mean and standard deviation…you need the mean to calculate the standard deviation), the remaining data points aren't as "free to vary".

Well, that oughta do for a rough intro to DF. Find the DF in each of the following scenarios:

E5. You sample a set of n data points and calculate $\bar{x}$, **s**, and the skew value of the sample, which is dependent on the mean and standard deviation. Thus, you have estimated 3 parameters.

E6. You sample two datasets of size n_1 and n_2, and calculate their respective sample means.

E7. You sample 5 datasets of sizes n_1 through n_5, and calculate their respective sample means.

[1]. I was going to relegate this to the enrichment page, but thought it a tad too important to leave there (on the off chance you've never visited the work of art that is the enrichment page).

Answers

E1. (4.89, 5.14)

E2. Without a margin of error, it's impossible to know if their claim has merit. Why?

E3. ($304,500, $325,000)…yours might be a tad different, since you might "ish" the frequencies differently than I do.

E4. Whaddaya think?

E5. $n - 3$

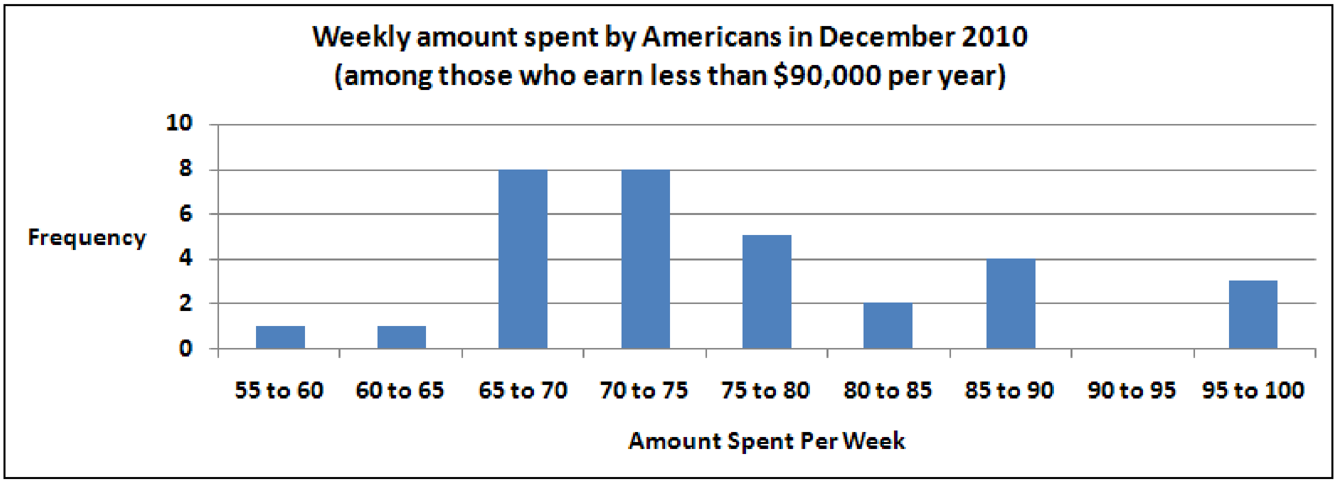
E6. $n_1 + n_2 - 1 - 1$, or $n_1 + n_2 - 2$. This is what is used when running controlled experiments (we'll see them later in the course).

E7. $n_1 + n_2 + n_3 + n_4 + n_5 - 5$. This is one of two DF used in ANOVA testing, one of the last things we'll learn in MTH 244.

Single Mean CI Quizzes

Quiz 1.

Gallup recently (1.13.11) released poll results concerning weekly spending of Americans. Although they hid their raw data (surprising, I know), I was able to generate some that fit their image of the spending of Americans who earned less than $90,000 in December 2010:



(**4 points**) Find the 95% CI for the weekly amount spent by Americans in December 2010, among those who make less than $90,000 per year. Round to the nearest dollar.

(**2 point**) According to this Gallup Poll, the average weekly spending among this same group in January 2011 was $58. Treat this as a parameter (I can't find the source's MOE, so why not?). Based on your CI above, is the headline of the article "*U.S. Consumer Spending Down in Early January*" accurate? "Yes" or "no" for this one, please.

(**4 points**) Why or why not?

Quiz 2.

Answer question **E4**.

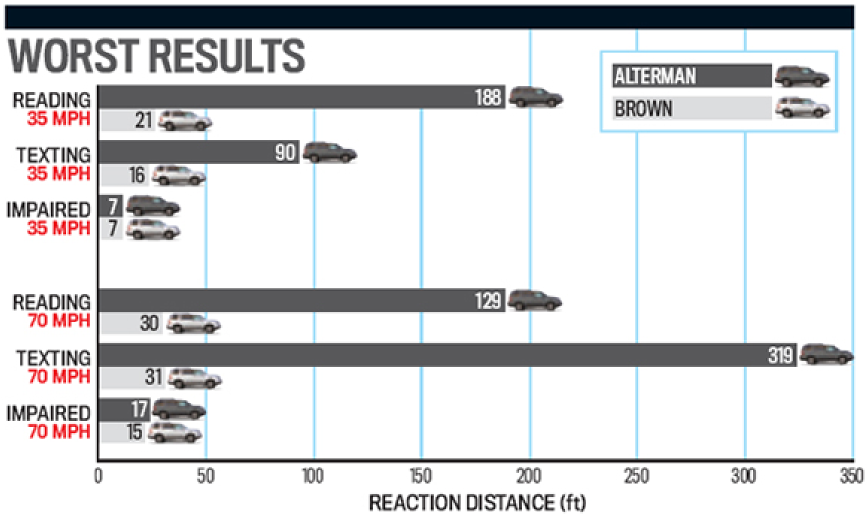
(**10 points**…**4 points for interval, 6 points for explanation**). Feel free to round the interval's endpoints to the nearest whole mile per hour. Be sure to use the number"30" in your answer!

(**extra 3 points**) (**w**) Construct a one – sided confidence interval for this data, and explain to me why someone might be more interested in that than a 2 – sided one.

Quiz 3.

Texting and driving really, really upsets me. I'm a bike commuter, and therefore more vulnerable out there on the road. I know people "think" they can text and drive and remain safe, but it's my opinion that they're absolutely, horribly wrong.

Anyway – a few years back, Car and Driver ran a test to compare the effects of reading texts, writing texts, and impairment due to alcohol on braking distance of drivers going both 35 miles per hour and 70 miles per hour. Here are some of their (worst) results:



Telling? Chilling? Depressing? (Yes)^3. Also utterly unscientific (no control to speak of, ***n*** = 2, no real mention of methodology). Since then, however, lots of other studies have popped up, including this one from HFES. Although they don't compare texting and driving to drunk driving (unfortunately), they ***do*** do a nice, scientific job of establishing a baseline with a nice sized sample (n = 40….remember that, because you'll need it in a sec). Here are their results (for "brake onset time"…i.e., "reaction time"):

TABLE 1: Means and Standard Deviations of Driving Performance for Each Experimental Condition

	Condition	
Variable	Single Task	Dual Task
Brake onset time (ms)	881 (349)	1,077 (380)

1. (**2 points**) Construct the 95% CI for "single task" (i.e., driving only) brake onset time (in ms). Note: when you see, for example, 881(349), that means "the mean is 881 ms and the standard deviation is 349 ms".

2. (**2 points**) Construct the 95% CI for "dual task" (i.e., driving while texting) brake onset time (in ms).

3. (**2 points**) Put both intervals on the same number line. Please use the Snipping Tool to copy and paste, if you like!

4. (**2 points**) is there a statistically significant difference? Why or why not (hint: mention overlap or lack thereof)

5. (**2 points**) (**w**) Now – since you constructed two confidence intervals (both at 95% confidence) what confidence do you ***actually*** achieve? If you're stuck, look back at E4 from the Single Proportion CI exercises.

So, when you're comparing two parameters (like in this case – single and dual task time), you need to collapse the CI down to one single calculation to maintain confidence. We'll start that soon in class!

<u>Quiz 4.</u>

Ripped from the headlines! Here's a Rasmussen report from 10.30.13 (http://www.rasmussenreports.com/public_content/politics/current_events/healthcare/health_care_law):

Health Care Law

52% Expect Obamacare to Make Health Care System Worse

<u>The Main Text:</u>

"Voters remain overwhelmingly positive about the health care they receive but are less enthusiastic about the overall health care system. But just over half also continue to believe the health care system will get worse under the new national health care law.

The latest Rasmussen Reports national telephone survey finds that 82% of Likely U.S. Voters rate the overall quality of the health care they now receive as good or excellent. Just four percent (4%) describe that health care as poor."

<u>The Fine Print:</u>

The survey of 1,000 Likely Voters was conducted on October 26-27, 2013 by Rasmussen Reports. The margin of sampling error is +/- 3 percentage points with a 95% level of confidence.

So, there are two claims in here I'd like to address: 1) "*Voters remain overwhelmingly positive about the health care they receive*" and 2) "*[J]ust over half also continue to believe the health care system will get worse under the new national health care law*."

1) *I believe this* – with a 3% MOE, and $\hat{p}$ = 82%, at least 79% (at 95% confidence) rate their overall health care quality as good or excellent.

2) *I don't necessarily believe this* – with a 3% MOE, and $\hat{p}$ = 52%, we're 95% confident that between 49% and 55% believe health care will get worse. That's not statistically significantly "more than half".

****

(**10 points**) **Find a study (or more than one study) where the MOE makes a claim statistically significantly believable, and find another study (or the same study with a different claim, like I did above) where it makes it *not* statistically significantly believable. 5 points for each study. You might use proportions (like these) or means – your choice. Make sure to include the analyses like I did in 1 and 2 above! Also include the course of your article(s).**