Stat 13, Tue 5/15/12.

1. Hand in HW5
2. Review list.
3. Assumptions and CLT again.
4. Examples.

Hand in Hw5. Midterm 2 is Thur, 5/17. Hw6 is due Thu, 5/24.
On Thur, 5/17, I won't be able to have my usual office hour from 230 to 3:30, so it will be instead from 1:30 to 2:15pm.
The midterm is primarily on chapters 4-6, though it has a bit of probability also.
The std normal and t tables will be provided.
Again, you can have 1 page, double-sided, of notes, plus a calculator and a pen or pencil.

2. Review list, of stuff since the first midterm.

a) More probability

- (i) Expected value.
- (ii) Expected value of sums of random variables.
- (iii) Bernoulli random variables.
- (iv) Binomial random variables.
- (v) Independence.

b) Normal calculations.

- (i) Calculating the area under the normal curve between a and b.
- (ii) Normal percentiles.
- (iii) Normal probability plots.

c) CLT and CIs.

- (i) CLT.
- (ii) Construction of CIs, for the mean and for proportions.
- (iii) Interpretation of CIs.
- (iii) SE versus σ.
- (iv) Margin of error.
- (v) Sample size calculations.
- (vi) CIs using the t table.
- (vii) Assumptions behind CIs.

3. Assumptions and Central Limit Theorem (CLT), again.

If you have a SRS (or observations are iid with mean μ),
and n is large (or the population is normally distributed),
then $\bar{x}$ is approximately normally distributed with mean μ and std deviation $\frac{\sigma}{\sqrt{n}}$,
where σ is the std deviation of the population
and n is the sample size.

Equivalently, we can say $[(\bar{x} - \mu) \div \frac{\sigma}{\sqrt{n}}]$ is approximately standard normally distributed.

When to use z* and t*.

a) If it's a simple random sample (SRS) and the population is normal, σ is unknown, and n is small (< 25), then use t*.

b) If it's a SRS and the population is normal, σ is known, and n is small (< 25), then use z*.

c) If it's a SRS and n is large, then t* and z* are very close together, so it doesn't really matter which you use. Use z*, though the book recommends t*.

d) If the population might NOT be normal and n is NOT large, then neither t* nor z* is appropriate.

$\bar{x}$

4. Examples.

In order to see what tv shows Americans watch, the Nielsen corporation surveys a sample of approximately 50,000 Americans. They recently (May 20, 2009) reported that the average American watches approximately 5.1 hours of tv a day. Suppose it's a SRS and that the sample standard deviation is s = 2 hours per day.
Find an 70%-CI for the population mean.

Answer: It's a SRS and n = 50,000 is large, so the standard formulas apply, but we don't know σ so we will plug in s. For a 70%-CI, we look for a value close to 15% (or 85%) in the table, and find the closest is -1.04 (or 1.04), so z* = 1.04.
The 70% CI is $\bar{x}$ +/- (z*)s/√n = 5.1 +/- (1.04)(2) ÷ √50,000 = 5.1 +/- 0.0093.

What does this 5.1 +/- 0.0093 mean? It's a range likely to contain μ. Here, n = 50,000 is so large that we can be confident that μ is likely close to 5.1.

Q. A typical American watches ___ +/- ___ hours of tv per day?

Q. If we took another SRS of 50,000 Americans, we'd expect to get a sample mean of around _____ +/- _____ hours of tv per day?

Q. If we change this 70% CI to a 95% CI, will the margin of error increase or decrease?

4. Examples.

In order to see what tv shows Americans watch, the Nielsen corporation surveys a sample of approximately 50,000 Americans. They recently (May 20, 2009) reported that the average American watches approximately 5.1 hours of tv a day. Suppose it's a SRS and that the sample standard deviation is s = 2 hours per day.
Find an 70%-CI for the population mean.

Answer: It's a SRS and n = 50,000 is large, so the standard formulas apply, but we don't know σ so we will plug in s. For a 70%-CI, we look for a value close to 15% (or 85%) in the table, and find the closest is -1.04 (or 1.04), so z* = 1.04.
The 70% CI is $\bar{x}$ +/- (z*)s/√n = 5.1 +/- (1.04)(2) ÷ √50,000 = 5.1 +/- 0.0093.

What does this 5.1 +/- 0.0093 mean? It's a range likely to contain μ. Here, n = 50,000 is so large that we can be confident that μ is likely close to 5.1.

Q. A typical American watches _5.1_ +/- _2_ hours of tv per day?

Q. If we took another SRS of 50,000 Americans, we'd expect to get a sample mean of around _____ +/- _____ hours of tv per day?

Q. If we change this 70% CI to a 95% CI, will the margin of error increase or decrease?

4. Examples.

In order to see what tv shows Americans watch, the Nielsen corporation surveys a sample of approximately 50,000 Americans. They recently (May 20, 2009) reported that the average American watches approximately 5.1 hours of tv a day. Suppose it's a SRS and that the sample standard deviation is s = 2 hours per day.
Find an 70%-CI for the population mean.

Answer: It's a SRS and n = 50,000 is large, so the standard formulas apply, but we don't know σ so we will plug in s. For a 70%-CI, we look for a value close to 15% (or 85%) in the table, and find the closest is -1.04 (or 1.04), so z* = 1.04.
The 70% CI is $\bar{x}$ +/- (z*)s/√n = 5.1 +/- (1.04)(2) ÷ √50,000 = 5.1 +/- 0.0093.

What does this 5.1 +/- 0.0093 mean? It's a range likely to contain μ. Here, n = 50,000 is so large that we can be confident that μ is likely close to 5.1.

Q. A typical American watches _5.1_ +/- _2_ hours of tv per day?

Q. If we took another SRS of 50,000 Americans, we'd expect to get a sample mean of around _5.1_ +/- _0.009_ hours of tv per day?

Q. If we change this 70% CI to a 95% CI, will the margin of error increase or decrease?

4. Examples.

In order to see what tv shows Americans watch, the Nielsen corporation surveys a sample of approximately 50,000 Americans. They recently (May 20, 2009) reported that the average American watches approximately 5.1 hours of tv a day. Suppose it's a SRS and that the sample standard deviation is s = 2 hours per day.
Find an 70%-CI for the population mean.

Answer: It's a SRS and n = 50,000 is large, so the standard formulas apply, but we don't know σ so we will plug in s. For a 70%-CI, we look for a value close to 15% (or 85%) in the table, and find the closest is -1.04 (or 1.04), so z* = 1.04.
The 70% CI is $\bar{x}$ +/- (z*)s/√n = 5.1 +/- (1.04)(2) ÷ √50,000 = 5.1 +/- 0.0093.

What does this 5.1 +/- 0.0093 mean? It's a range likely to contain μ. Here, n = 50,000 is so large that we can be confident that μ is likely close to 5.1.

Q. A typical American watches _5.1_ +/- _2_ hours of tv per day?

Q. If we took another SRS of 50,000 Americans, we'd expect to get a sample mean of around _5.1_ +/- _0.009_ hours of tv per day?

Q. If we change this 70% CI to a 95% CI, will the margin of error increase or decrease? Increase. z* will go from 1.04 to 1.96. Margin of error will almost double.

According to the CDC, the largest number of reported cases in the U.S. for any condition is chlamydia, a sexually transmitted disease reported each year in about 0.4% of people overall. In 2009-2010, the National Health and Nutrition Examination Survey (NHANES) took a SRS of 10,253 Americans, and among the (roughly) 510 females age 15-24, they found the prevalence of chlamydia to be 8.0%.

Find a 95%-CI for the population percentage of females age 15-24 with chlamydia.

Answer: This is a 0-1 question. It's a SRS and n is large because the number of females with chlamydia in the sample = 8% x 510 = 41 ≥ 10, and the number without = 92% x 520 = 469 ≥ 10.

For a 95%-CI, z* = 1.96 from the bottom row of Table 4.
The formula for the 95%-CI is $\bar{x}$ +/- z* σ/√n.

We don't know σ so use s = $\sqrt{\hat{p}\hat{q}}$ = √ (8.0% x 92.0%) ~ 0.271.

Our 95%-CI is 8.0% +/- (1.96) (0.271) / √510
which is 8.0% +/- 2.35%.

Suppose you take a SRS of 17 UCLA students and find their IQs. You find that $\bar{x}$ is 120 and s = 20. Find a 95% CI for μ, the mean IQ for UCLA, assuming these IQs are normally distributed.

Here we have a SRS, the pop. is normal, and σ is unknown, so use the t table.
df = n-1 = 17-1 = 16. From Table 4, for a 95% CI, with df = 16, t* = 2.12.
So, the 95% CI is $\bar{x}$ +/- t* s/√n = 120 +/- 2.12 (20)/√17 = 120 +/- 10.28.

What is the standard error?
s/√n = 20/√17 = 4.85.

A typical UCLA student has an IQ of about ____ +/- ________ .

If you take other samples each of 17 UCLA students, you'd expect your samples typically to have a mean of about _______ +/- ________ .

Suppose you take a SRS of 17 UCLA students and find their IQs. You find that $\bar{x}$ is 120 and s = 20. Find a 95% CI for μ, the mean IQ for UCLA, assuming these IQs are normally distributed.

Here we have a SRS, the pop. is normal, and σ is unknown, so use the t table.
df = n-1 = 17-1 = 16. From Table 4, for a 95% CI, with df = 16, t* = 2.12.
So, the 95% CI is $\bar{x}$ +/- t* s/√n = 120 +/- 2.12 (20)/√17 = 120 +/- 10.28.

What is the standard error?
s/√n = 20/√17 = 4.85.

A typical UCLA student has an IQ of about _120_ +/- __20___ .

If you take other samples each of 17 UCLA students, you'd expect your samples typically to have a mean of about ______ +/- ______ .

Suppose you take a SRS of 17 UCLA students and find their IQs. You find that $\bar{x}$ is 120 and s = 20. Find a 95% CI for μ, the mean IQ for UCLA, assuming these IQs are normally distributed.

Here we have a SRS, the pop. is normal, and σ is unknown, so use the t table.
df = n-1 = 17-1 = 16. From Table 4, for a 95% CI, with df = 16, t* = 2.12.
So, the 95% CI is $\bar{x}$ +/- t* s/√n = 120 +/- 2.12 (20)/√17 = 120 +/- 10.28.

What is the standard error?
s/√n = 20/√17 = 4.85.

A typical UCLA student has an IQ of about _120_ +/- __20___ .

If you take other samples each of 17 UCLA students, you'd expect your samples typically to have a mean of about _120__ +/- _4.85__ .