Agenda

1. Initial Project Proposal due on Friday
2. HW #5 Problem
3. Confidence Intervals
4. More on the Central Limit Theorem (CLT)

Confidence Interval Confidence intervals for test statistics that are normally distributed are of the form:

$$\text{point estimate} \pm z^*_{\alpha/2} \cdot SE$$

Computing the point estimate is usually easy. Once you've chosen a confidence level, finding $z^*_{\alpha/2}$ is trivial (use `qnorm()`). The difficult part is usually computing the SE, since that depends on the sampling distribution of the test statistic!

Visualizing Confidence Intervals Open the following URL in a web browser:

`http://shiny.calvin.edu/rpruim/CIs/`

- Experiment with changing the sample size. How does that change the coverage rate? How does it change the confidence intervals?
- Experiment with changing the confidence level. Does increasing the confidence level make the intervals wider or narrower?
- Experiment with changing the population distribution from normal to something non-normal. How does that change the coverage rate?

Twitter Users and News A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Conduct a hypothesis test to see if the proportion of Twitter users who get at least some of their news from twitter is different from 50%.

1. Draw a picture of the sampling distribution of the proportion of U.S. adult Twitter users who get at least some news on Twitter.

2. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter.

```r
qnorm(0.995)

## [1] 2.575829
```

3. Interpret the confidence interval in context.

Practice Problems

1. A recent study estimated the mean U.S. per capita consumption of sugar-sweetened beverages among adults 20 to 44 years of age to be 289 kcal/day with a standard error of 7 kcal/day.

   (a) The 68-95-99.7 rule says that the probability is about 0.95 that $\bar{x}$ is within y kcal/day of the population mean μ. What is y?

   (b) About 99% of all samples will capture the true mean of kcals consumed per day in the interval $\bar{x}$ plus or minus 7 kcal/day times what? Draw a labeled picture and indicate where the missing quantity is. Estimate it. What does the computer need to known in order to compute it?

2. Suppose 400 randomly selected alumni of the University of Okoboji were asked to rate the university's counseling services on a 1 to 10 scale. The sample mean was found to be 8.6. Assume that the standard error was computed to be 0.4.

   (a) Ima Bitlost computes the 99% confidence interval for the average satisfaction score as $8.6 \pm 1.96 \cdot 0.4$. What is her mistake?

   (b) After correcting her mistake in part (a), she states: "I am 95% confident that the sample mean falls between 7.82 and 9.38." What is wrong with this statement?

   (c) She quickly realizes her mistake in part (b) and instead states: "The probability the true mean is between 7.82 and 9.38 is 0.95." What misinterpretation is she making now?

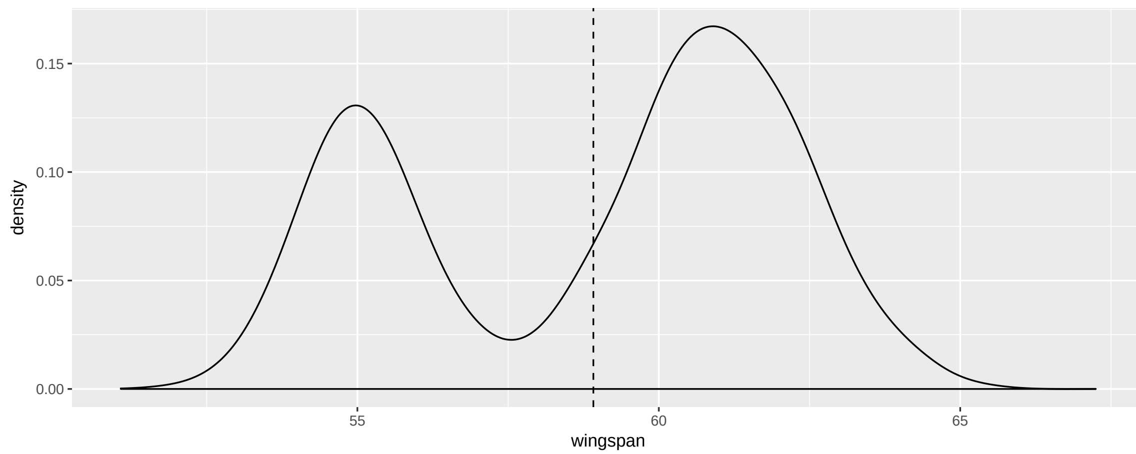
   (d) Finally in her defense for using the Normal distribution to determine the confidence interval she says "Because the sample size is quite large, the population of alumni ratings will be approximately Normal." Explain to Ima her misunderstanding and correct this statement.

Dragons' Wingspans Consider the bimodal population distribution of dragon's wingspans.

```
library(mosaic)
n <- 20000
dragon_pop <- data.frame(tiny_wings = rnorm(n, mean = 55),
                         big_wings = rnorm(n, mean = 61, sd = 1.5),
                         p = runif(n)) %>%
  mutate(wingspan = ifelse(p < 0.35, tiny_wings, big_wings))

pop_plot <- ggplot(data = dragon_pop, aes(x = wingspan)) +
  geom_density() +
  geom_vline(aes(xintercept = summarise(dragon_pop, mean(wingspan))), linetype = 2)

pop_plot
```



1. Sketch what you think the sampling distribution of the mean looks like for samples of size 1.

2. Sketch what you think the sampling distribution of the mean looks like for samples of size 2.

3. Sketch what you think the sampling distribution of the mean looks like for samples of size 10,000.

4. Sketch what you think the sampling distribution of the mean looks like for samples of size 4. Try it in R with the following code!

```
sim <- do(1000) * summarise(sample_n(dragon_pop, 4), x_bar = mean(wingspan))
pop_plot + geom_density(data = sim, aes(x = x_bar), adjust = 0.5, color = "red")
```