Finding CIs

This project looks at how to find and think about confidence intervals (CIs). We see there are different ways of finding them. We start with the commands to find them by hand. The formula for a C CI for μ based on $\bar{x}$ is

$$\bar{x} \pm z^* \sigma / \sqrt{n}$$

To compute this in R, we start with weight data from cars in the `mtcars` data set. Let's assume for this data $\sigma = 1$. So

```
> x <- mtcars$wt
> xbar <- mean(x)
> sigma <- 1
> zstar <- qnorm(0.975)
> xbar + c(-1, 1) * zstar * sigma

[1] 1.257286 5.177214
```

Question 0.1. We can do question 6.26 from the text, once we download the data and store it in the variable `x`, as follows:

```
> f <- "http://www.math.csi.cuny.edu/verzani/classes/MTH214/R/Data/ex06_026.txt"
> d <- read.table(f, header = TRUE)
> str(d)

'data.frame':        20 obs. of  1 variable:
 $ MPG: num  41.5 50.7 36.6 37.3 34.2 45 48 43.2 47.7 42.2 ...

> x <- d$MPG
```

Now, as the problem says suppose $\sigma = 3.5$.

1. What is $\sigma_{\bar{x}}$? $\bar{x}$?
2. Give a 95% CI for μ.
3. Verify graphically that the data seems to come from a normal or nearly normal distribution. How did you do so?

Question 0.2. If you have data, you can also type it in of course. Suppose our data from some experiment is

```
7, 12, 13, 6
```

If $\sigma = 2$, find a 90% CI. (You can type in the data as

```
> x <- c(7, 12, 13, 6)
```

The t-statistic

We will see in class that assuming we know σ when we don't know μ is not really satisfactory in practice. Instead we will use the sample standard deviation instead of the population standard deviation in our calculations.

We will need to make two changes to our basic formula. Before dicussing those, we note that t he `t.test` function will find these confidence intervals for us. It is straightforward to use, but one must learn to read its output. For example, we have

```
> x <- mtcars$mpg
> t.test(x, conf.level = 0.95)

        One Sample t-test

data:  x
t = 18.8569, df = 31, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 17.91768 22.26357
sample estimates:
mean of x
 20.09062
```

The key lines are

```
95 percent confidence interval:
 17.91768 22.26357
```

indicating a confidence interval (17.9,22.3) miles per gallon.

Question 0.3. Compare a 95% CI where we use the `t.test` function to that we already did for the data from problem 6.26. What is similar, what is different?

Question 0.4. The assumption on the population is that the data is normal or nearly normal. Investigate this for the `mpg` data. Comment.

Simulation studies

We can do a simulation study of some properties of confidence intervals. To open a window to assist in the study, do the following:

```
> f <- "http://www.math.csi.cuny.edu/verzani/classes/MTH214/Computer/ci.R"
> source(f)
```

The resulting window will run a simulation every time you make a change or click the resample "again" button. The graphic shows confidence intervals computed over several simulations from a normal population with the specified parameters. A CI that "misses" is marked differently from the others.

Question 0.5. With the default values are the lengths of the CIs equal or different?

Question 0.6. What is the mark that makes the missed CIs different?

Question 0.7. For a sample of size 10, how long are the confidence intervals (assume $\mu = 1$, $\sigma = 1$)? How about a sample of size 20? 40?

Question 0.8. Make a general statement: keeping all other things equal, as sample size goes up, the margin or error ....

Question 0.9. Run a simulation of 25 intervals. How often is the mean less than μ?

Question 0.10. Run a simulation of 25 intervals with a confidence level of 0.95. How many intervals "Miss?" Repeat with a confidence level of 0.8. Now how many? Repeat both with 100 intervals.

Question 0.11. What is the distribution for the random number computed in the last question?

Question 0.12. Change to "Use s, not sigma" uses the t-statistic. What can you say about the CIs now?

Question 0.13. Toggle back and forth between the statistics, keeping everything else the same. Can you make a general statement comparing the margin of errors?

Question 0.14. Does it appear that the CIs based on assuming σ contain μ more often than those estimating σ?