

Lesson 8 Example Code

I'm not going to re-hash the example code from Lesson 6 here so please review that document if you have any questions. Instead I'm going to demonstrate finding critical values using the t-distribution. I will use the time estimates data from Lesson 7.

```
library(tidyverse)

estimates = read_table2("http://www.isi-stats.com/isi/data/chap3/TimeEstimate.txt")

sample_stat = mean(estimates$Time)

sample_size = nrow(estimates)

standard_deviation = sd(estimates$Time)
```

Once again, if we are interested in finding a confidence interval (CI), we will start with the sample statistic as the center of our range of plausible values for parameter value. From here we need to establish our critical values or “the multiplier” in the book’s equation. Unfortunately, as the t-distribution is shaped a little bit differently depending on the degrees of freedom (related to the sample size here), we don’t have something similar to the *68-95-99 Rule*. Fortunately, *R* is awesome and that’s not a big deal.

Just like `pnorm()` has a sibling `qnorm()`, `pt()` has a sibling `qt()`. They’re basically a superhero family that we can use to calculate our critical values. For a 95% CI, we would use:

```
qt(0.025, df = sample_size - 1)
```

```
## [1] -2.063899
```

```
qt(0.975, df = sample_size - 1)
```

```
## [1] 2.063899
```

That’s pretty close to 2 which is why the book suggests the *2SD Method* for both the *one-sample z-interval* and the *one-sample t-interval*. Unlike the *one-sample z-interval* we can’t easily “relocate” our t-distribution to be centered at our sample statistic using parameters in the `qt()` function but the equation offered in your book basically accomplishes this.

```
lower = sample_stat + qt(0.025, df = sample_size - 1) *
  (standard_deviation / sqrt(sample_size))

upper = sample_stat + qt(0.975, df = sample_size - 1) *
  (standard_deviation / sqrt(sample_size))

paste("(", lower, ", ", upper, ")")
```

```
## [1] "( 11.5643745096081 , 16.0356254903919 )"
```

I want to stress taking a moment at this point and asking yourself: “Does this make sense as a range of plausible values for my parameter?” I, of course, never make a mistake in coding but you just might and so it helps to examine your answer as a check. Is it plausible, given what we know (our sample), that we can be 95% confident that the population average time estimate is between 11.5 and 16.0 seconds? I think it is and

therefore I can be fairly comfortable that I've accurately calculated this CI. It's not a definitive check but it just might save you from making a silly mistake.