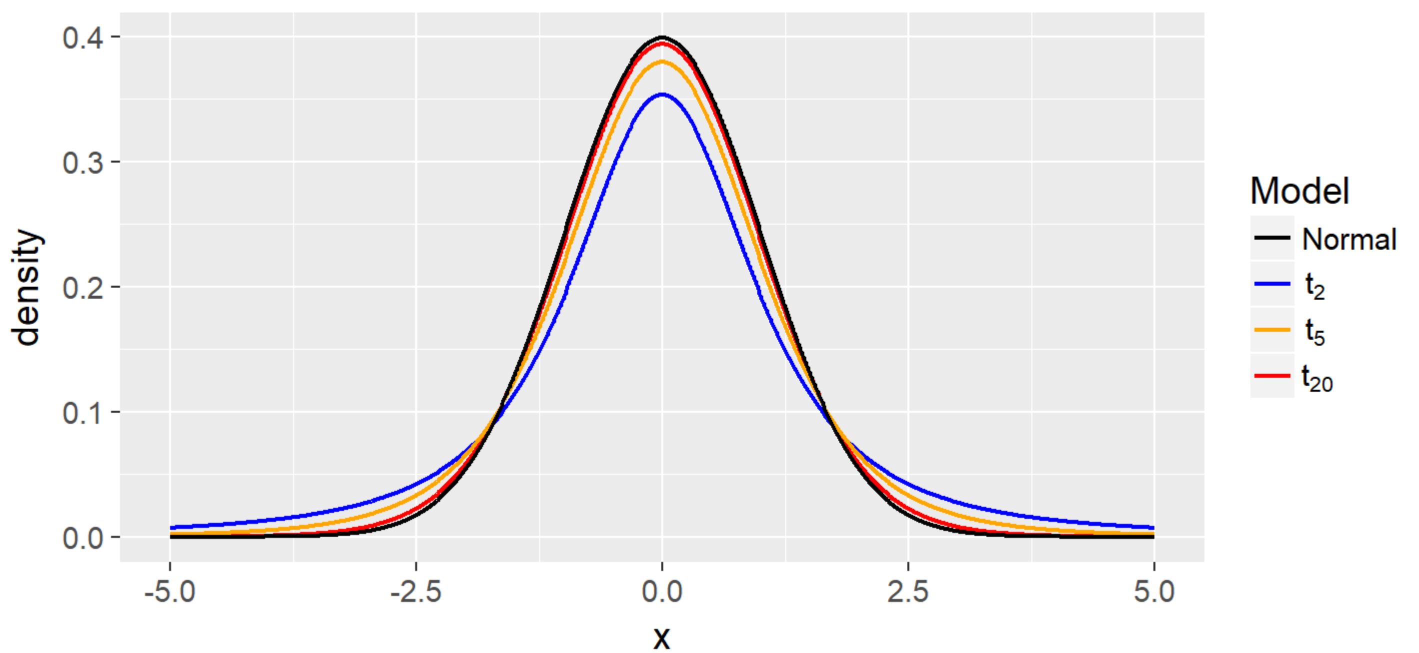
Confidence Intervals for Population Means

Evan L. Ray

November 10, 2017

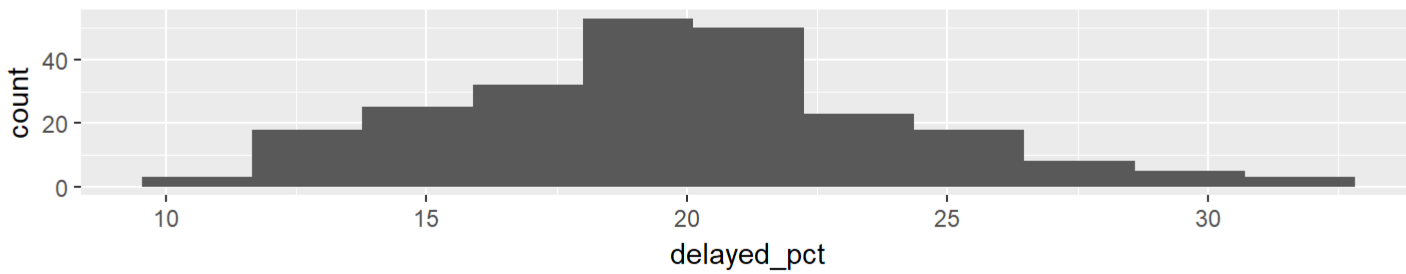
t Distribution

- The t distribution is similar to the Normal$(0, 1)$
- The t has more probability in the tails
- As the degrees of freedom increases, the t becomes more like a Normal$(0, 1)$



Example: Flight Delays

The U.S. Bureau of Transportation Statistics reported the percentage of flights that were delayed each month from 1994 through October of 2013 (238 months in total). Treat these as a representative sample of all months. Here's a histogram:



- Would it be appropriate to use these data to calculate a confidence interval for the mean percent of flights that are delayed per month?
  - **Symmetric?** Skewed a little to the right, but not too badly
  - **Unimodal?** Yes
  - **Sample Size?** The sample size is fairly large, so a normal model for the sample mean is appropriate

Example: Flight Delays

- Calculate a 95% confidence interval for the mean percent of flights that are delayed per month

```
t.test(delays$delayed_pct, conf.level = 0.95)
```

```
## 
##  One Sample t-test
## 
## data:  delays$delayed_pct
## t = 71.31, df = 237, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  19.20733 20.29872
## sample estimates:
## mean of x 
##  19.75303
```

Example: Flight Delays

- Calculate a 95% confidence interval for the mean percent of flights that are delayed per month

```
n <- nrow(delays) # 238 observations
sample_mean <- mean(delays$delayed_pct) # sample mean = 19.75
sample_sd <- sd(delays$delayed_pct) # sample standard deviation = 4.27
mean_se <- sample_sd / sqrt(n) # standard error of sample mean = 0.28
t_critical <- qt(0.975, df = n - 1) # critical value: use .975 for a 95% CI!

sample_mean - t_critical * mean_se # lower CI bound
```

```
## [1] 19.20733
```

```
sample_mean + t_critical * mean_se # upper CI bound
```

```
## [1] 20.29872
```