Agenda

1. Applying the Normal Model
2. Confidence Intervals

Applying the Normal Model Recall the baseball example from last time.

```
## Warning:  package 'dplyr' was built under R version 3.3.2
## Warning:  package 'bindrcpp' was built under R version 3.3.2

##    playerID yearID      BAvg
## 1 willite01   1941 0.4057018
## 2 brettge01   1980 0.3897550
## # A tibble: 2 x 4
##   yearID     N mean_BAvg    sd_BAvg
##    <int> <int>     <dbl>      <dbl>
## 1   1941    98 0.2806367 0.03279026
## 2   1980   148 0.2788247 0.02757441
```

George Brett, who hit .390 in 1980, won the AL MVP. The player who finished second in the balloting, Reggie Jackson, hit .300 (with 41 home runs). Let's examine Jackson's batting average in the context of his peers. What we need is a way to understand the *distribution* of batting average in the AL in 1980. We have three different ways to do this:

1. Use the actual batting averages from the 148 players with at least 400 at-bats:

```
pdata(~BAvg, q = .300, data = filter(mlb, yearID == 1980))
```

2. *Assume that batting average is distributed normally* and use the observed mean and standard deviation to specify the distribuion:

```
xpnorm(.300, mean = .279, sd = 0.0276)
pnorm(.300, mean = .279, sd = 0.0276)
```

3. Simulate the distribution using R's random number generating capabilities:

```
sim <- data.frame(BAvg = rnorm(10000, mean = .279, sd = 0.0276))
pdata(~BAvg, q = .300, data = sim)
```

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.

1. Draw a picture of the sampling distribution of the proportion of U.S. adult Twitters 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.

```
qnorm(0.995)

## [1] 2.575829
```

3. Interpret the confidence interval in context.

4. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

   (a) The data provide statistically significant evidence that more than half of U.S. adult Twitter users get some news through Twitter. Use a significance level of $\alpha = 0.01$.

   (b) Since the standard error is 2.4%, we can conclude that 97.6% of all U.S. adult Twitter users were included in the study.

   (c) If we want to reduce the standard error of the estimate, we should collect less data.

   (d) If we construct a 90% confidence interval for the percentage of U.S. adults Twitter users who get some news through Twitter, this confidence interval will be wider than a corresponding 99% confidence interval.

   (e) If we repeated this study 1,000 times and constructed a 99% confidence interval for each study, then approximately 990 of those confidence intervals whould contain the true fraction of U.S. adult Twitter users who get at least some news on Twitter.

   (f) The margin of error in this poll is less than 3 percentage points.