Stein's Paradox Lab:

Lab Overview

This lab will provide an opportunity to explore Stein's Paradox.

The data from the paper is available at

```r
library(readr)
stein <- read_csv('http://www.math.montana.edu/ahoegh/teaching/stat532/data/SteinData.csv')
```

```
## Parsed with column specification:
## cols(
##   Name = col_character(),
##   avg45 = col_double(),
##   avgSeason = col_double()
## )
```

1. Compute James-Stein Estimator

Using the variable `avg45` the batting average through 45 at bats, compute the James-Stein estimator. Recall the James-Stein estimator is

$$z = \bar{y} - c(y - \bar{y})$$

where $c = 1 - \frac{(k-3)\sigma^2}{\sum(y-\bar{y})^2}$ and that the authors estimate $\sigma^2 = \frac{\hat{p}(1-\hat{p})}{45}$, where $\hat{p} = \frac{1}{18}\sum y = \bar{y}$. Note you might not perfectly recreate the results in the paper

a.

Describe what role c has in the final estimate.

b.

Compute the MSE between the season ending averages (`avgSeason`) for the James-Stein estimator as well as the estimate based on 45 at bats (`avg45`).

c.

Summarize your results.

2. Create a simulation study to mimic this scenario.

Generate a set of 18 baseball players, each with some "true batting average" (typically between .150 and .350). For each batter, give them 45 at bats and record the batting average.

a.

Compute the MSE between the estimated and season ending averages for the James-Stein estimator and the observed average.

b.

Repeat this entire procedure 1000 times and record the proportion of simulations where the James-Stein estimator is better.

c.

Summarize your results.