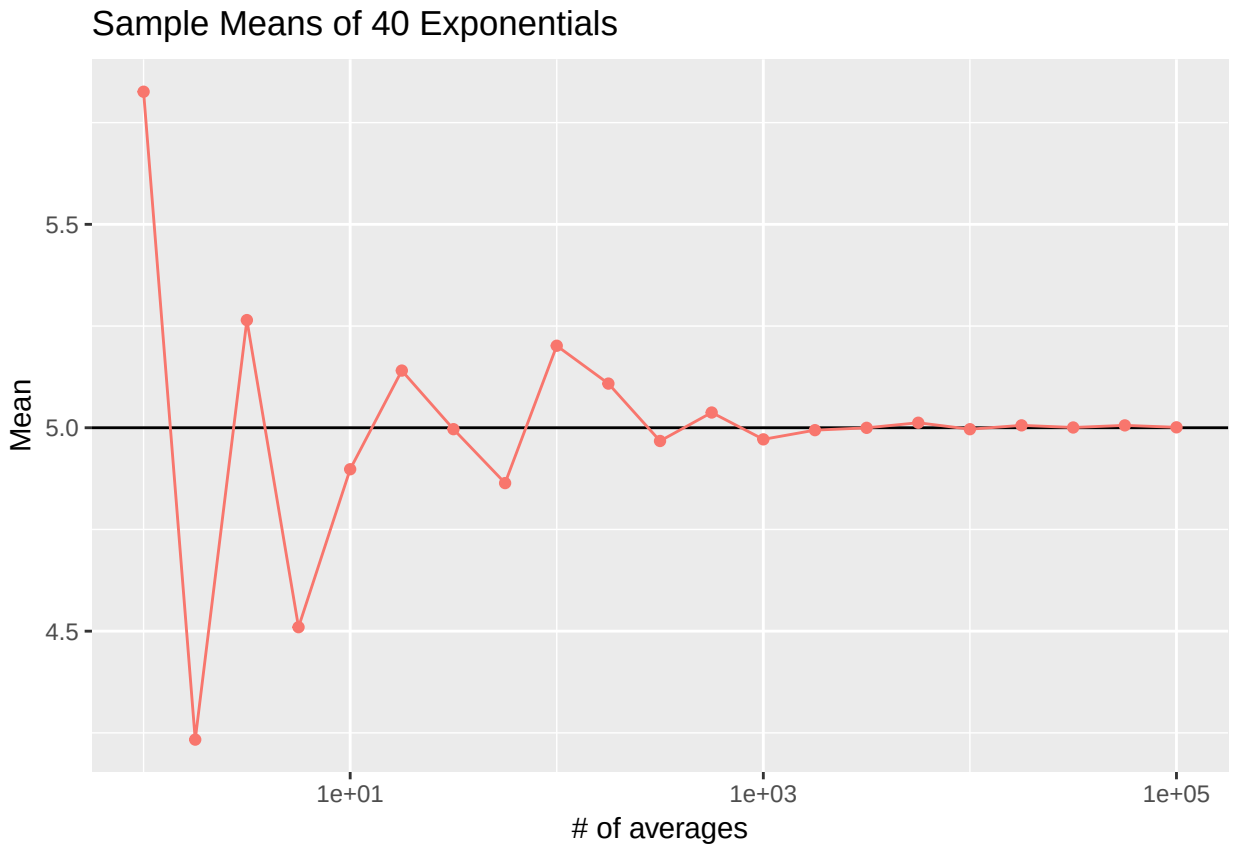
# Statistical Inference Final Project

*Steven Vasquez-Grinnell*

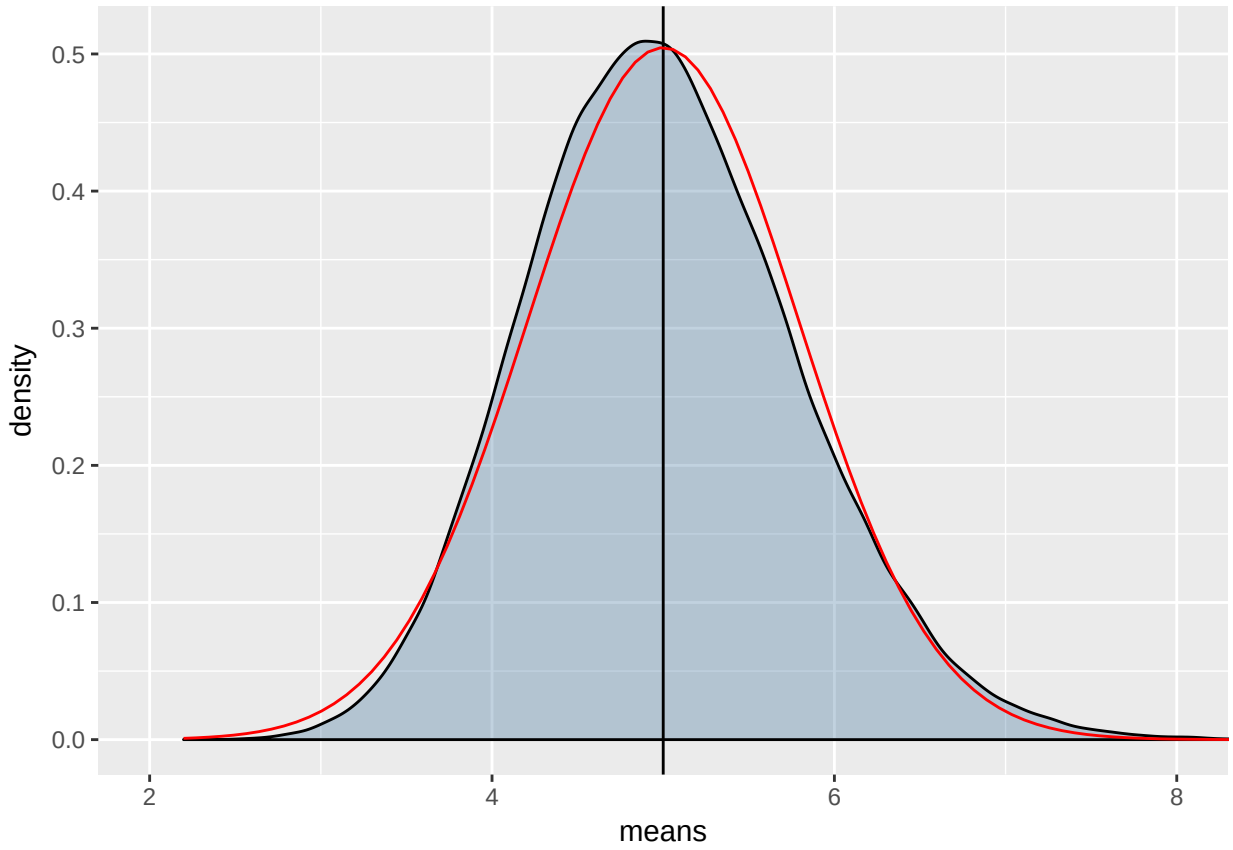
*May 3, 2017*

```
set.seed(122)
overall_means <- data.frame(index = 1:21, overall_mean = 1:21)
all_the_means <- vector("list", length = 21)

for (i in 1:21){
  step <- 10^((i-1)/4)
  tmp_means <- vector(length = step)

  for (j in 1:step) {
    tmp_means[j] <- mean(rexp(40,0.2))
  }

  all_the_means[[i]]<-tmp_means

  overall_means$index[i] <- step
  overall_means$overall_mean[i] <- mean(tmp_means)
}
```

```
ggplot(overall_means, mapping = aes(x = index, y = overall_mean, color = "red")) +
  scale_x_log10() +
  geom_hline(yintercept = 5) +
  geom_point() +
  geom_line() +
  labs(title = "Sample Means of 40 Exponentials", x = "# of averages", y = "Mean") +
  guides(color = "none")
```




```
means_count <- length(tmp_means)
means_df <- data.frame(row = means_count, means = tmp_means, cond = 1)
theoretical <- data.frame(y = rnorm(means_count, mean = 5, sd = 5))

ggplot(means_df, mapping = aes(x = means, fill = cond)) +
  coord_cartesian(xlim = c(2,8)) +
  geom_density(alpha = 0.3) +
  geom_vline(aes(xintercept = 5)) +
  stat_function(fun = dnorm, color = "red", args = list(mean = 5, sd = 5/sqrt(40))) +
  guides(fill = "none")
```




```
theoretical_vs_real<- lm(means_df$means ~ theoretical$y)

par(mfrow = c(2,2), oma = c(0,0,4,0))
plot(theoretical_vs_real)
title(main = "Figure 3: Final Model Diagnostics",  outer = TRUE)
```

**Figure 3: Final Model Diagnostics**