Lab 11

Group Member Names - here

Lab Overview

All students attending class will earn participation points for this lab. Students not attending class will need to complete their own lab.

We will revisit the movie earnings dataset using the bootstrap estimator.

```
movies <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/movies_earnings.csv')
```

```
## Parsed with column specification:
## cols(
##   title = col_character(),
##   budget = col_double(),
##   revenue = col_double(),
##   release_date = col_date(format = "")
## )
```

```
movies <- movies %>% mutate(revenue_millions = revenue/1000000, budget_millions = budget / 1000000)
```

A SRS of 100 movies has been taken for you from the dataset. Using the ratio estimator

```
movies_sample <- movies %>% sample_n(100)
```

1. (2 points)

Using the standard SRS estimator, compute a point estimate of the mean movie revenue.

```
srs_estimate <- movies_sample %>% summarize(mean_rev = mean(revenue_millions)) %>% pull()
srs_estimate
```

```
## [1] 91.23616
```

2. (2 points)

Using a regression estimator, compute a point estimate of the mean movie earnings. Assume that you know the population mean for the movie budget,

```
xbar_U <- movies %>% summarize(mean(budget_millions)) %>% pull()
```

```
lm_movies <- lm(revenue_millions ~ budget_millions, data = movies_sample)
regression_estimate <- as.numeric(coef(lm_movies)[1]) + as.numeric(coef(lm_movies)[2]) * xbar_U
regression_estimate
```

```
## [1] 85.2037
```

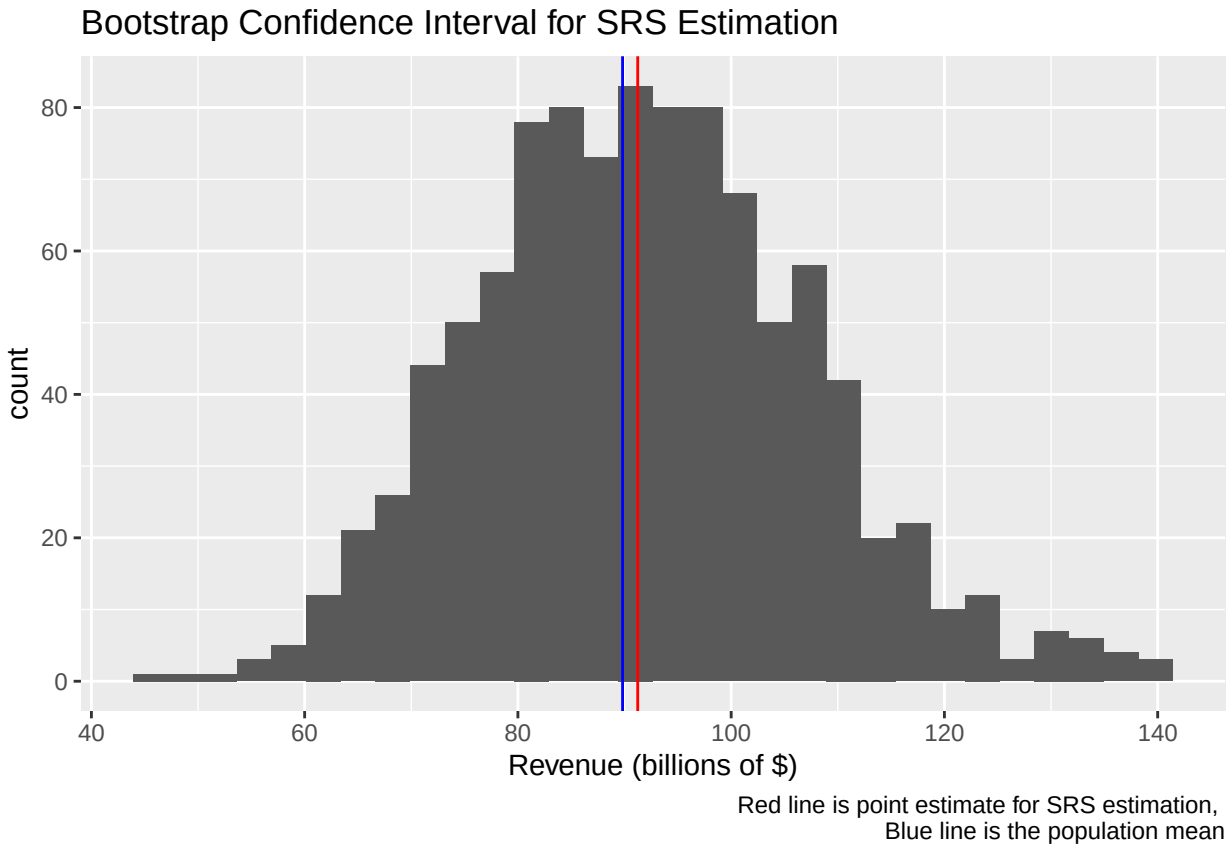
3. (3 points)

Now use a bootstrap procedure to estimate the confidence interval for the SRS estimator. Plot the confidence interval along with your point estimate and the true value.

```
num_boot <- 1000
boot_estimates_srs <- rep(0, num_boot)
for (iter in 1:num_boot){
  boot_sample <- movies_sample %>% sample_n(100, replace = T)
  boot_estimates_srs[iter] <- boot_sample %>% summarize(mean_rev = mean(revenue_millions)) %>% pull()
}

tibble(x = boot_estimates_srs) %>% ggplot(aes(x)) + geom_histogram() + geom_vline(xintercept = srs_esti
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



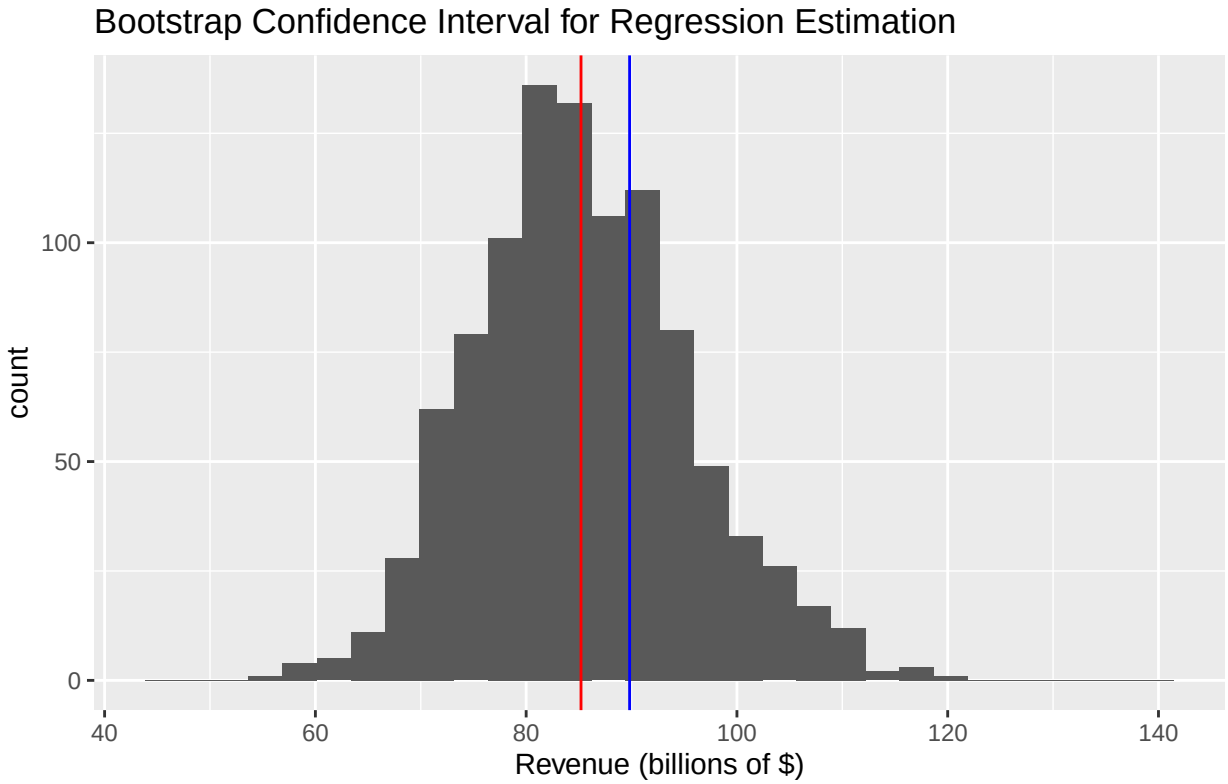
4. (3 points)

Now use a bootstrap procedure to estimate the confidence interval for the regression estimator. Plot the confidence interval along with your point estimate and the true value.

```
num_boot <- 1000
boot_estimates_reg <- rep(0, num_boot)
for (iter in 1:num_boot){
  boot_sample <- movies_sample %>% sample_n(100, replace = T)
  lm_movies <- lm(revenue_millions ~ budget_millions, data = boot_sample)
   boot_estimates_reg[iter] <- as.numeric(coef(lm_movies)[1]) + as.numeric(coef(lm_movies)[2]) * xbar_U
}

tibble(x = boot_estimates_reg) %>% ggplot(aes(x)) + geom_histogram() + geom_vline(xintercept = regressi
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Bootstrap Confidence Interval for Regression Estimation
count
100
50
0
40
60
80
100
120
140
Revenue (billions of $)
Red line is point estimate for regression estimation,
Blue line is the population mean