

# STAT 446 HW5

## Question 2

Lab 6 extension. Using the 100 samples compute a 95% confidence interval for the ratio estimator.

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.2
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(readr)
set.seed(10212019)
movies <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/movies_earnings.csv')

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

movies_sample <- movies %>% sample_n(100)
ybar <- mean(movies_sample$revenue)
xbar <- mean(movies_sample$budget)
xbar_u <- mean(movies$budget)
b <- ybar/xbar
N <- nrow(movies)
n <- nrow(movies_sample)
alpha <- 0.05
y <- movies_sample$revenue
x <- movies_sample$budget
residuals <- (y - b * x)^2
s2_e <- (1/(n-1))*sum(residuals)
var_ybar_r <- (N - n) / N * (xbar_u / xbar)^2 * s2_e / n

ybar_r <- b * xbar_u
ci_interval <- ybar_r + c(-1,1) * sqrt(var_ybar_r) * qnorm(1 - alpha/2)
```

The point estimate for revenue is 83.05 million dollars with a confidence interval of 63.05, 103.04 million dollars.