

# F Ratio - OLS regression

*Sri Seshadri*

*1/20/2018*

## Question

Page 76 of ISLR states that "...When there is no relationship between the response and predictors, one would expect the F-Statistic to take on a value close to 1. On the other hand, If  $H_a$  is true then  $E\{(TSS - RSS)/p\} > \sigma^2$ , so we expect F to be greater than 1"

**If there was no relationship between the predictor and the response variable shouldn't the F statistic be very less than or equal to 1? instead of close to 1?**

## Experiment

Generate random numbers and randomly assign the levels of the predictor variable to the random numbers. Fit a OLS regression and inspect the ANOVA table for F ratio. Repeat the experiment for 1000 iterations.

```
getFStats <- function() {  
  # Generate random numbers from a standard normal distribution  
  
  y <- rnorm(1000)  
  
  # generate predictor variable with levels 1 through 10  
  
  x <- rep(1:10,100)  
  
  # plot y vv x  
  
  #plot(y~x)  
  
  # Simple linear regression  
  
  model.lm <- lm(formula = y ~ x)  
  
  #summary(model.lm)  
  
  Fstats <- anova(model.lm)$`F value`[1]  
  return(Fstats)  
}  
  
result <- replicate(1000,getFStats())  
  
hist(result,breaks = 20,col = "blue",xlab = "F statistic")  
abline(v=1,col = "red",lty = 5)
```

## Histogram of result

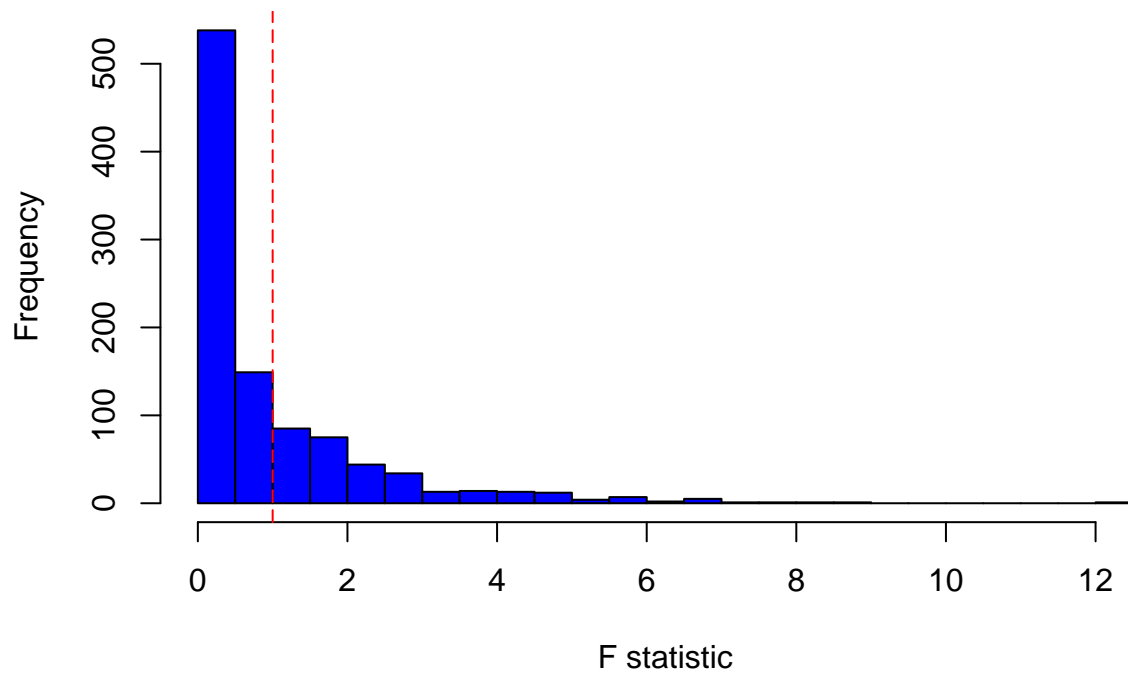


Figure 1: Histogram of F ratios

## Conclusion

The quote in ISLR is misleading and should read “F statistic should take on values less than or equal to 1”. I am open to criticism, please point me in a direction should you feel the experimentation is not set up correctly.