**R Code for Examples in the book**



***“Statistics: The Art and Science of Learning from Data”***

**by Agresti, Franklin and Klingenberg, 5th edition**

**Chapter 7**

**Example 6: Sampling Distribution for the Sample Correlation Coefficient**

## Reading in the data:

sandwiches <- read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter7/carbon\_footprint\_sandwich.csv')  
attach(sandwiches) # so we can refer to variable names

## To compute the correlation coefficient between carbon footprint and energy content

cor(EnergyContent..kCal., Carbon.footprint..g.CO2.eq..)

## [1] 0.6208991

## To obtain a bootstrap sample of the sandwiches

sample(Sandwich, replace = TRUE)

## [1] "Egg, Mayo, Cress" "Ham, Salad" "Chicken, Sweetcorn"  
## [4] "Chicken Salad" "Chicken, Stuffing" "Egg, Bacon"   
## [7] "Egg, Mayo, Cress" "Tuna, Cucumber" "Chicken, Sweetcorn"  
## [10] "Cheese, Tomato" "Ham, Egg" "Cheese, Mayo"   
## [13] "Egg, Mayo, Cress" "Chicken, Sweetcorn" "Chicken, Mayo"   
## [16] "Prawn, Mayo" "Chicken, Sweetcorn" "Cheese Ploughman "   
## [19] "Chicken Salad" "Breakfast" "Chicken Salad"   
## [22] "Chicken, Bacon" "Cheese, Onion" "Ham, Egg"

## 

## To obtain a bootstrap sample of the rows of the dataframe, you can use sandwiches[sample(seq\_len(nrow(sandwiches)), replace = TRUE), ]. Then to generate 10,011 bootstrap samples and find each sample’s correlation coefficient

bootcorr <- c() # initializing  
for (i in 1:10011) {  
 bootsample <- sandwiches[sample(seq\_len(nrow(sandwiches)),   
 replace = TRUE), ]  
 bootcorr[i] <- cor(bootsample$EnergyContent..kCal.,   
 bootsample$Carbon.footprint..g.CO2.eq..)  
 }

## To obtain summary of the correlation coefficients from the bootstrap samples

summary(bootcorr)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -0.08516 0.50750 0.62058 0.59889 0.71105 0.96436

sd(bootcorr)

## [1] 0.1519422