**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:

data <- read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter7/carbon\_footprint\_sandwich.csv')  
attach(data) # 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] "Ham, Cheese" "Chicken, Sweetcorn"   
## [3] "Chicken, Sweetcorn" "Cheese, Mayo"   
## [5] "Tuna, Sweetcorn" "Cheese, Onion"   
## [7] "Chicken Salad" "Egg, Rocket"   
## [9] "Egg, Rocket" "Ham, Egg"   
## [11] "Chicken, Sweetcorn" "Cheese, Tomato"   
## [13] "Ham, Mustard" "Tuna, Sweetcorn"   
## [15] "Chicken, Bacon" "Sausage, Brown Sauce"   
## [17] "Bacon, Lettuce, Tomato (BLT)" "Egg, Bacon"   
## [19] "Tuna, Cucumber" "Ham, Mayo"   
## [21] "Cheese, Pickle" "Bacon, Lettuce, Tomato (BLT)"  
## [23] "Cheese, Mayo" "Egg, Bacon"

## 

## To obtain a bootstrap sample of the rows of the dataframe, you can use data[sample(seq\_len(nrow(data)), 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 <- data[sample(seq\_len(nrow(data)), 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.   
## -0.1260 0.5071 0.6212 0.5980 0.7110 0.9705

sd(bootcorr)

## [1] 0.1541156