Abalone Dataset Summary

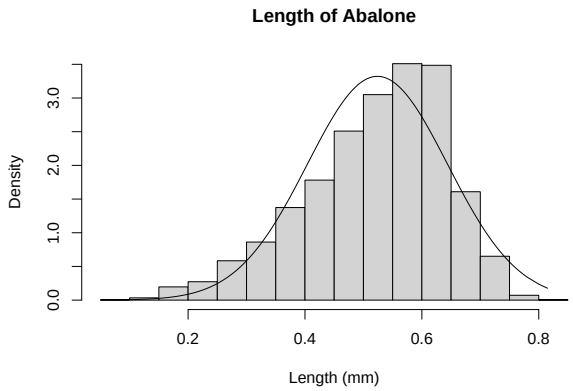
a) Abalone Data Set Summary

The Abalone data set is provided by UCI. It predicts the age of abalone from some physical measurements such as length, weight, and diameter and etc. The data originally comes from a study conducted by Warwick, Tracy, Simon, Andrew and Wes in 1994. There are a total of 4177 observations (rows) and 9 variables (columns) with the "Sex" variable being character, "Rings" integer, and the rest double. The following table outlines the detail of all attributes:

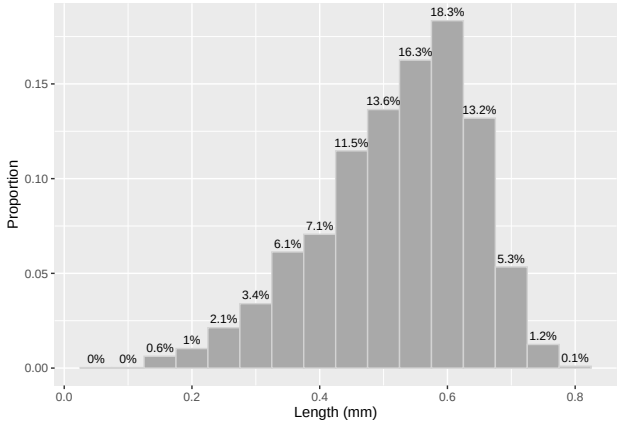
Name	Data Type	Meas.	Description
Sex	nominal	M, F, and I (infant)	
Length	Continuous	mm	Longest shell measurement
Diameter	Continuous	mm	perpendicular to length
Height	Continuous	mm	with meat in shell
Whole weight	Continuous	grams	whole abalone
Shucked weight	Continuous	grams	weight of meat
Viscera weight	Continuous	grams	gut weight (after bleeding)
Shell weight	Continuous	grams	after being dried
Rings	Integer		+1.5 gives the age in year

b) Histogram of Length

Now we will select the Length variable to be studied. In the following, we plot a density histogram with a plot of the pdf of a normal distribution superimposed:



***** Additionally, if we must display the y-axis in proportion instead of density:** The following graph, created with *ggplot*, scales the y-axis into percentage.



And if we superimpose a plot of the pdf of a normal distribution on the above graph, it scales down the y-axis and looks like this: