Clive Sombe

**DS 710**

**R Programming Assignment**

**Assignment 3**

1. **Analyzing Used Car Prices**
2. Download Cars 2005.csv, load the data into R, and attach it.

(Dataset: “Car Data," submitted by Shonda Kuiper, Grinnell College. Dataset obtained from the Journal of Statistics Education (http://www.amstat.org/publications/jse). Accessed 3 June 2015. Used by permission of author.)

**Ans:**

**> ridez = read.csv("C:/Users/sembc2622/Documents/DS710/Cars 2005.csv")**

**> attach(ridez)**

1. Make a histogram of the prices of cars in the data.  Describe the shape of the distribution.

**Ans:**

**> hist(Price)**



**Ans:The histogram has a tail on the right side. It is right-skewed.**

1. What proportion of cars in the data set cost between $10,000 and $20,000?

**Ans:**

**> > TenToTwentyK = which((Price >= 10000) & (Price <= 20000))**

**> length(TenToTwentyK)/length(Price)**

**[1] 0.5621891**

**Proportion of cars that cost between $10,000 and $20,000 is 0.5621891**

1. Find the mean and median price.  Which is larger?  Why does this make sense?

**Ans:**

**> mean(Price)**

**[1] 21343.14**

**> median(Price)**

**[1] 18025**

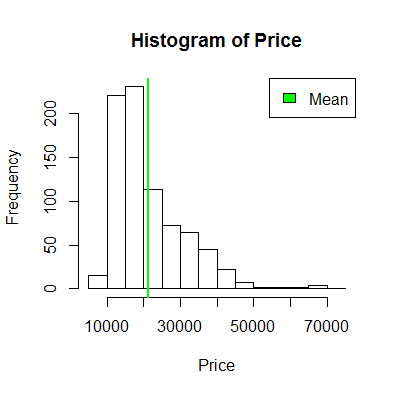
**The mean is larger. This makes sense because the data, as the histogram shows, is skewed to the right and we know that the mean tends to be pulled towards the extreme values.**

1. Add a vertical line to the histogram to denote the mean price.  Add a legend to the graph.

**Ans:**

**> abline(v = 21343.14, col = "green", lwd = 2)**

**> legend("topright",legend = c("Mean"),fill=c("green"))**



1. Transform the price to reduce its skew, and make a histogram of the transformed price.  Fit a normal distribution to the transformed price, and graph the normal density curve on the same plot as the histogram.  How well does a normal distribution fit the transformed data?

**Ans:**

**> LogP=log(Price)**

**> hist(LogP)**

**> hist(LogP, prob=T)**

****

**> m<-mean(LogP)**

**> std<-sqrt(var(LogP))**

**> hist(LogP,prob=T,**

**+ xlab="LogP", ylim=c(0,1),**

**+ main="normal curve over histogram")**

**> curve(dnorm(x,9.87905,0.4101115),**

**+ col="red",lwd=2,add=T,yaxt="n")**

****

**The normal distribution fits the transformed data very well.**

1. Make a scatterplot of transformed price versus engine size, measured in liters.  Describe the relationship between these two variables.

**Ans:**

**> plot(LogP, Liter)**



**ploThe scatterplot shows a moderately positive correlation between the transformed price of cars and the engine size in liters.**

1. Find the correlation between transformed price and engine size in liters.  Explain what it tells us.

**> cor(LogP,Liter)**

**[1] 0.5904097**

**A correlation value of 0.5904097 confirms what the scatterplot is showing – a moderately positive correlation.**

1. Modify your scatterplot in part g to use one color of plotting symbol for cars with leather interiors, and a different color for cars without leather interiors.  Add a legend to your plot.
2. Make a barplot of the types (Sedan, Hatchback, etc.) of cars in the data.
3. Make a barplot of the types of cars and whether they have leather interiors.  Add a legend to your plot.

**Ans:**

**> barplot(Leather)**

**> counts = table(Leather)**

**> barplot(counts)**

**> barplot(counts, xlab = "Leather Interior", ylab = "Number of Cars", main = "Barplot of Car Interors")**

**> counts = table(Type,Leather)**

**> barplot(counts, xlab = "Leather Interior", ylab = "Number of Cars", main = "Barplot of Car Interors")**

**> legend("topleft",legend=c("Sedan","Hatchback","Convertible","Wagon","Coupe"))**



1. Make a boxplot of (untransformed) price by type of car.  In words, summarize what it shows.



**Ans:**

**The boxplot shows outliers on its high end indicating that the car prices are right skewed. The data has a definite lower limit but no upper limit**

1. Create two different histograms in a vertical stack that allow comparison of (untransformed) price according to whether the car has a leather interior.  Use the same horizontal axis for each to enable comparison, and use informative labels for each graph and the x-axis.

**Ans:**



**boxplot(Price ~ Leather)**

1. Create a side-by-side histogram to allow the same comparison as in part l.  Add a legend to your plot.

Submit a .doc, .docx, or .pdf document containing your R code, output and graphs, and interpretations (where requested) to parts a-n.  It is not necessary to include R code where you were testing code or where you made a mistake--just submit the final version.