**Nicole Verdon**

**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.

mycars = read.csv(file="C:/Users/Verdon/Documents/DS710/Cars 2005.csv")

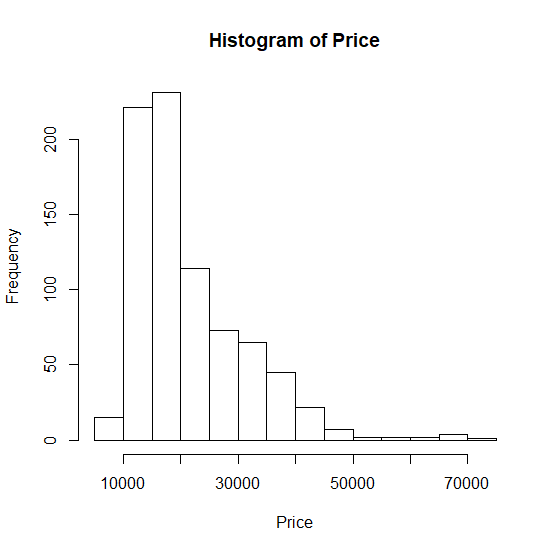
attach(mycars)

(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.)

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

with(mycars, hist(Price))

Right Skewed



1. What proportion of cars in the data set cost between $10,000 and $20,000? ~60%

with(mycars,

hist(Price, 8)

)

> with(mycars,

+ range(Price)

+ )



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

summary(mycars)

Mean :21343 Median :18025

Mean is larger, this makes sense because it is a right skewed data set, the middle price number (median) will likely be less than the mean (average price).

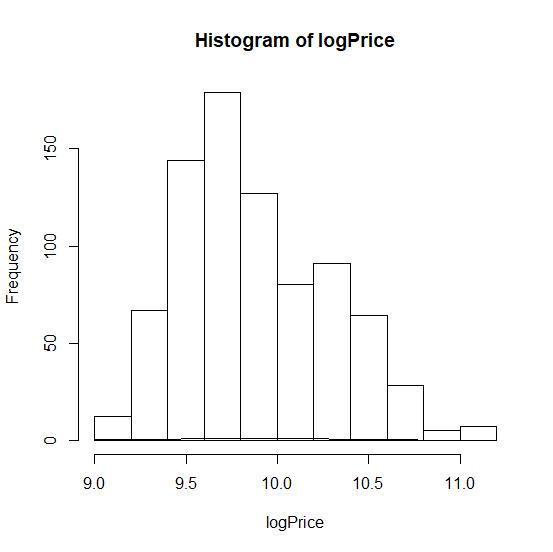
1. Add a vertical line to the histogram to denote the mean price.  Add a legend to the graph.



abline(v=21343,col=3,lty=3)

legend("topright", c("Mean"), lty=3, col=3)

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? The normal distribution is more symmetrical with the transformed data and the density curve is small, yet defined well within the data distribution.



with(mycars,

hist(logPrice)

meanlogPrice= mean(logPrice)

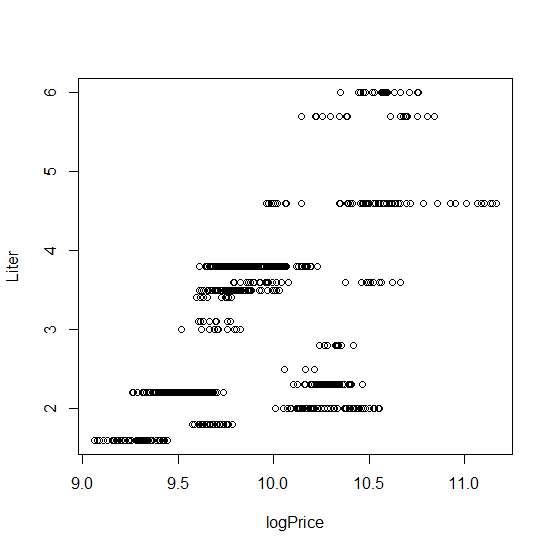
sdlogPrice= sd(logPrice)

curve( dnorm( x, meanlogPrice, sdlogPrice), add= T)

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

with(mycars, plot(logPrice, Liter))

Positive, linear, correlation



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

> with(mycars, cor(logPrice, Liter))

[1] 0.5904097

This means that it has a moderate correlation between the increase in price (positive) and an increase in engine size (positive), therefor it is a positive, linear, moderate 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.

hasLeather <- which(Leather == 1)

noLeather <- which(Leather == 0)

PriceWithLeather <- Price[hasLeather]

PriceNoLeather <- Price[noLeather]

with(mycars, cor(logPrice, Liter), which(1,col=2), which(2, col=5))

#I couldn’t get this one to run…. I kept getting the correlation?

legend("topright", c("With Leather"), col=2, c("Without Leather), col=5)

1. Make a barplot of the types (Sedan, Hatchback, etc.) of cars in the data.

I am very much struggling on this… I regret saving this assignment to the 11th hour…. Here is a list of what I have tried and received from the program….

> par(las=1, mar=c(3,9,1,1)

+ with(mycars, barplot(Type, names.arg=Type, horiz=TRUE))

Error: unexpected symbol in:

"par(las=1, mar=c(3,9,1,1)

with"

> with(mycars, barplot(Type, names.arg=Type, horiz=TRUE))

Error in barplot.default(Type, names.arg = Type, horiz = TRUE) :

'height' must be a vector or a matrix

> par(las=1, mar=c(3,9,1,1))

> with(mycars, barplot(Type, names.arg=Type, horiz=TRUE))

Error in barplot.default(Type, names.arg = Type, horiz = TRUE) :

'height' must be a vector or a matrix

> par(las=1, mar=c(3,9,1,1))

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> par(las=1,)

Error in par(las = 1, ) : argument is missing, with no default

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> par(las=1,)

Error in par(las = 1, ) : argument is missing, with no default

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> par(las=1))

Error: unexpected ')' in "par(las=1))"

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> with(mycars, barplot(Type, horiz=TRUE))

Error in barplot.default(Type, horiz = TRUE) :

'height' must be a vector or a matrix

> with(mycars, barplot(Type))

Error in barplot.default(Type) : 'height' must be a vector or a matrix

1. Make a barplot of the types of cars and whether they have leather interiors.  Add a legend to your plot.
2. Make a boxplot of (untransformed) price by type of car.  In words, summarize what it shows.
3. 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.
4. Create a single histogram with side-by-side bars to allow the same comparison as in part m.  Add a legend to your plot.
5. **Analyzing the running speed of mammals**
6. In R, type

install.packages("quantreg")

data(Mammals, package="quantreg")

This will load a data set called Mammals, on the maximum land speed of various species of mammal.  Attach the data and look at the first few lines.

install.packages("quantreg")

data(Mammals, package="quantreg")

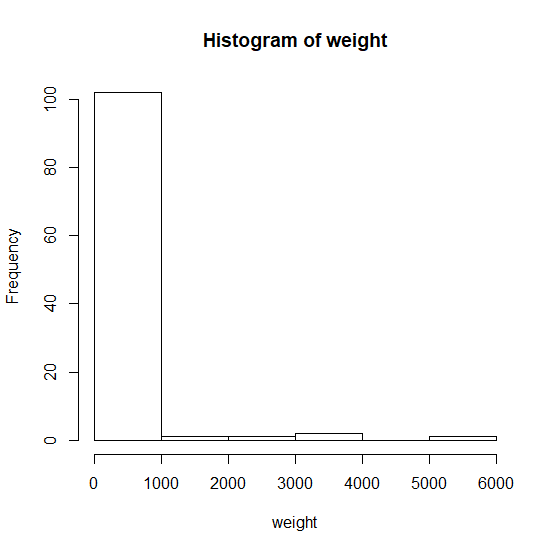
Mammals[1,]

(Source:  Garland, T. (1983) The relation between maximal running speed and body mass in terrestrial mammals, *J. Zoology*, 199, 1557-1570.

Metadata:  <http://vincentarelbundock.github.io/Rdatasets/doc/quantreg/Mammals.html>, accessed 7 June 2015.)

1. Decide whether either of the quantitative variables should be transformed.  Justify your decision using appropriate plots and/or descriptive statistics.

The median weight of the mammals is 34, the mean is 279… this shows that the data is skewed right, as also shown in the histogram. Utilizing the log function would assist in transforming the data for more appropriate analysis.



> with(Mammals, hist(logweight))

Error in hist(logweight) : object 'logweight' not found

> with(mycars,

+ hist(logPrice)

+ logPrice= log(Price)

Error: unexpected symbol in:

"hist(logPrice)

logPrice"

> meanlogPrice= mean(logPrice)

> sdlogPrice= sd(logPrice)

> meanlogWeight= mean(logWeight)

Error in mean(logWeight) : object 'logWeight' not found

> sdlogWeight= sd(logWeight)

Error in is.data.frame(x) : object 'logWeight' not found

> logWeight = log10

> meanlogWeight= mean(logWeight)

Warning message:

In mean.default(logWeight) :

argument is not numeric or logical: returning NA

> sdlogWeight= sd(logWeight)

Error in as.double(x) :

cannot coerce type 'builtin' to vector of type 'double'

> summary(Mammals)

weight speed hoppers specials

Min. : 0.016 Min. : 1.60 Mode :logical Mode :logical

1st Qu.: 1.700 1st Qu.: 22.50 FALSE:96 FALSE:97

Median : 34.000 Median : 48.00 TRUE :11 TRUE :10

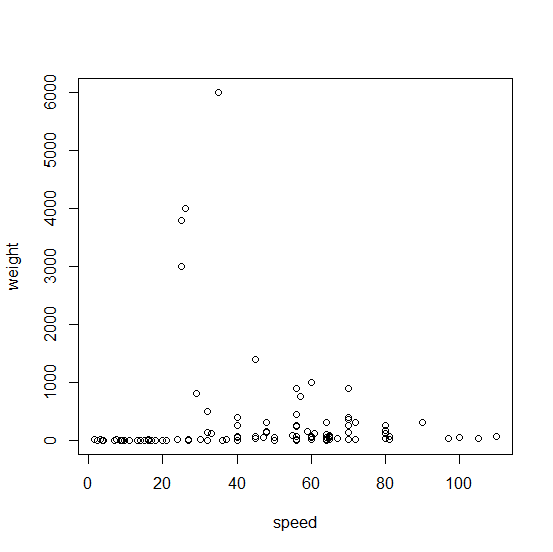
Mean : 278.688 Mean : 46.21

3rd Qu.: 142.500 3rd Qu.: 65.00

Max. :6000.000 Max. :110.00

>

c.  Use appropriate graphs and/or descriptive statistics to describe the relationship between maximum land speed and body weight.  Does it matter whether the animal is a “hopper” (such as a kangaroo)?  Explain why you chose the graphs and/or statistics that you chose.

  
> with(Mammals, plot(speed, weight))

> with(Mammals, cor(weight,speed))

[1] -0.06653467

The correlation shows a very minimal negative correlation between weight and speed.

If we were to look at the correlation of weight and speed of hoppers, it would be a negative correlation, much stronger than weight and speed correlation. As the weight decreases, the speed increases for hoppers.

Submit a **single** .docx or .pdf document containing your R code, output and graphs, and interpretations (where requested).  To facilitate a quick turnaround on grading, please keep all parts of a problem together, rather than putting code at the end of the file.

* Please include your full name at the top of your document.
* It is not necessary to include R code where you were testing code or where you made a mistake--just submit the final version.