Ryan Timbrook

Applied Data Science
IST687 Intro to Data Science, Spring 2019
Due Date: 05/28/2019
Homework: 8
NetID: RTIMBROO
SUID: 386792749

#R Code - unexecuted

```
## Homework Week 8: Viz Map HW: Making Predictions - Explore Antelope Populations

#--- Data Details ----------------------------------------------------------------
# 4 columns, 8 years of observations (n=8)
# 1st Column: Number of fawn in a given spring
# 2nd Column: Population of adult antelope
# 3rd column: Annual precipitation that year
# 4th column: value representing how bad the winter was during that year
#

#--- Preprocess Steps: -----------------------------------------------------------------

### Clear objects from Memory
rm(list=ls())
### Clear Console:
cat("\014")
### Set Working Directory
setwd("C:\\workspaces\\ms_datascience_su\\IST687-IntroDataScience\\R_workspace\\hw")

#---- Global Variable Assignments -------------------------------------------
antelopeDataSetURL <-
"http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/
excel/mlr01.xls"
# Path to local Perl Interpreter - Needed to gdata package for read.xls()
perlPath <- "C:\\Strawberry\\perl\\bin\\perl"

#---- Load Required Packages ------------------------------------------------
if(!require("devtools")) {install.packages("devtools")}
if(!require("RCurl")) {install.packages("RCurl")}
if(!require("gdata")) {install.packages("gdata")}
if(!require("ggplot2")) {install.packages("ggplot2")}
```

```
if(!require("MASS")) {install.packages("MASS")}
#if(!require("tidyverse")) {install.packages("tidyverse")}

# ---- Step 1: Load the data ---------------------------------------------------------------------------
## 1.1: Read the data:
readDataSetasXLSX <- function(dsURL,locPerlPath){

  df <- read.xls(dsURL, perl=locPerlPath)

  return(data.frame(df))
}

antelopeDf <- readDataSetasXLSX(antelopeDataSetURL,perlPath)

# ---- Step 2: Explore the data ---------------------------------------------------------------------------
str(antelopeDf)

# ---- Step 3: Clean the data ---------------------------------------------------------------------------
# Rename Column Headings
newColumnNames <-
c('Fawn_Cnt','Adult_Antelope_Pop','Annual_Percipitation','Bad_Winter_Scale')
colnames(antelopeDf) <- newColumnNames

# ---- Step 4: Create bivariate plots ---------------------------------------------------------------------------
#
# Create bivariate plots of number of baby fawns versus adult antelope population,
# the precipitation that year, and the severity of the winter. Your code should produce
# three separate plots. Make sure the Y-axis and X-axis are labeled. Keeping in mind
# that the number of fawns is the outcome (or dependent) variable, which axis should
# it go on in your plots?
# Attribute Type:
#   - Precipitation: Independent / Numeric:Continuous
#   - Severity of Winter: Independent / Categorical (Scale[1-5])
#   - Baby Count: Dependent / Numeric:Continuous
#   - Adult Population: Independent / Numeric:Continuous


## Plots - Focus is on the dependent variable, Baby Fawn Count

# Plot 1: Number of baby fawns versus adult antelope population
# x-axis: Adult_Antelope_Pop
# y-axis: Fawn_Cnt
g.fawn.adult <- ggplot(antelopeDf, aes(x=Adult_Antelope_Pop, y=Fawn_Cnt)) +
```

```
  geom_point() +
  stat_smooth(method='lm',col='red') +
  labs(title="Bivariate Plot: \nNumber of baby fawns born \n  versus \nAdult antelope population",
x="Annual Adult Antelope Population", y="Number of Fawns born in Spring")
g.fawn.adult
ggsave("Bivariate_Plot_Baby_Fawns_Adult_Population.jpg", width = 6, height = 6)

# Plot 2: Number of baby fawns given the precipitation that year
# x-axis: Annual_Percipitation
# y-axis: Fawn_Cnt
g.fawn.percipitation <- ggplot(antelopeDf, aes(x=Annual_Percipitation, y=Fawn_Cnt)) +
  geom_point() +
  stat_smooth(method='lm',col='red') +
  labs(title="Bivariate Plot: \nNumber of baby fawns born \n  versus \nAnnual percipitation",
x="Annual Percipitation", y="Number of Fawns born in Spring")
g.fawn.percipitation
ggsave("Bivariate_Plot_Baby_Fawns_Precipitation.jpg", width = 6, height = 6)

# Plot 3: Number of baby fawns given the severity of the winter
# x-axis: Bad_Winter_Scale
# y-axis: Fawn_Cnt
g.fawn.winter <- ggplot(antelopeDf, aes(x=Bad_Winter_Scale, y=Fawn_Cnt)) +
  geom_point() +
  stat_smooth(method='lm',col='red') +
  labs(title="Bivariate Plot: ", x="Winter Severity Rating", y="Number of Fawns born in Spring")
g.fawn.winter
ggsave("Bivariate_Plot_Baby_Fawns_Bad_Weather_Rating.jpg", width = 6, height = 6)

# ---- Step 5: Create regression models ---------------------------------------------------------------------------
#
# Create three regression models of increasing complexity using lm().

# Step 5.1: Predict the number of fawns from the severity of the winter
# Output: Multiple R-squared = 0.5459 | p-value = 0.03626
fawn.winter.lm <- lm(Fawn_Cnt ~ Bad_Winter_Scale, data=antelopeDf)
sum.f.w <- summary(fawn.winter.lm)
sum.f.w
range(antelopeDf$Bad_Winter_Scale)
newWinterData <- data.frame(Bad_Winter_Scale=3)
predict(fawn.winter.lm,newWinterData,type="response")

# Step 5.2: Predict the number of fawns from two variables (one should be the severity of the winter)
# Output: Adjusted R-squared = 0.8439 | p-value = 0.004152
```

```
fawn.AntelopePop.Winter.lm <- lm(Fawn_Cnt ~ Adult_Antelope_Pop+Bad_Winter_Scale,
data=antelopeDf)
sum.f.a.w <- summary(fawn.AntelopePop.Winter.lm)
sum.f.a.w
range(antelopeDf$Adult_Antelope_Pop)
newAntelopeWinterData <- data.frame(Adult_Antelope_Pop=9, Bad_Winter_Scale=1)
predict(fawn.AntelopePop.Winter.lm,newAntelopeWinterData,type='response')

# Step 5.3: Predict the number of fawns from the three other variables
# Output: Adjusted R-squared = 0.955 | p-value = 0.001229
fawn.AntelopePop.Percipitation.Winter.lm <- lm(Fawn_Cnt ~
Adult_Antelope_Pop+Annual_Percipitation+Bad_Winter_Scale, data=antelopeDf)
sum.f.a.p.w <- summary(fawn.AntelopePop.Percipitation.Winter.lm)
sum.f.a.p.w
range(antelopeDf$Annual_Percipitation)
newAntelopePopPercipitationWinterData <- data.frame(Adult_Antelope_Pop=5,
Annual_Percipitation=14 ,Bad_Winter_Scale=4)
predict(fawn.AntelopePop.Percipitation.Winter.lm, newAntelopePopPercipitationWinterData,
type='response')

# Step 5.4: Questions to Answer
# Question 5.4.1: Which model works best?
# The third model, fawn.AntelopePop.Percipitation.Winter.lm, it has an Adjust R-squared score of
0.955 and p-value of 0.001 (statistically significant)
paste("Adjusted R-squared:",sum.f.a.p.w$adj.r.squared)
paste("P-Value:")
sum.f.a.p.w$coefficients[,4]

# Question 5.4.2: Which of the predictors are statistically significant in each model?
# Model 1: Number of fawns from the severity of the winter
# Answer: Bad Winter Rating has a predictor confidence of 55% accuracy
paste("Multiple R-squared:",sum.f.w$r.squared)
paste("P-Value:")
sum.f.w$coefficients[,4]

# Model 2: Number of fawns from the Adult Antelope Population and Bad Winter Rating
# Answer: Adult Antelope Population has a prediction confidence accuracy of 84%, Bad Winter is not
statistically significant in this model
paste("Adjusted R-squared:",sum.f.a.w$adj.r.squared)
paste("P-Value:")
sum.f.a.w$coefficients[,4]
```

# Model 3: number of fawns from the Adult Antelope Population, Annual Percipitation and Bad Winter Rating
# Answer: Adult Antelope Population, Annual Percipitation and Bad Winter Rating are all statistically significant
#   and combined have a prediction confidence accuracy of 95%
paste("Adjusted R-squared:",sum.f.a.p.w$adj.r.squared)
paste("P-Value:")
sum.f.a.p.w$coefficients[,4]

Question 5.4.3: If you wanted to create the most parsimonious model (i.e.,
the one that did the best job with the fewest predictors), what would it contain?
Answer:
Analysis of Deviance Table
Initial Model:
Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale
Final Model:
Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale

# Apply Stepwise regression analysis, find best combination of variables for best prediction output
lm.step <- stepAIC(fawn.AntelopePop.Percipitation.Winter.lm,direction="both")
lm.step$anova
summary(lm.step)
plot(lm.step)

Other individual attribute tests

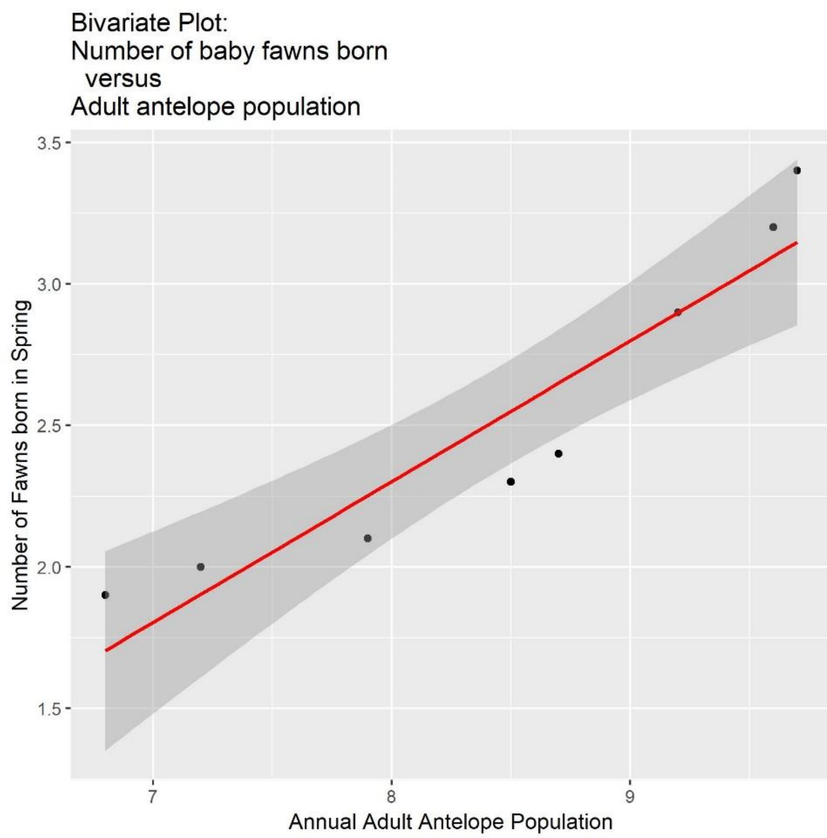
# Output: Multiple R-squared = 0.8813 | p-value = 0.0005471
fawn.apop <- lm(Fawn_Cnt ~ Adult_Antelope_Pop, data=antelopeDf)
sum.fawn.apop <- summary(fawn.apop)
sum.fawn.apop
# Output: Multiple R-squared = 0.8536 | p-value = 0.001039
fawn.aperc <- lm(Fawn_Cnt ~ Annual_Percipitation, data=antelopeDf)
sum.fawn.aperc <- summary(fawn.aperc)
sum.fawn.aperc

#R Code – executed

```
> ### Set Working Directory
> setwd("C:\\workspaces\\ms_datascience_su\\IST687-IntroDataScience\\R_worksp
ace\\hw")
>
> #---- Global Variable Assignments ------------------------------------------
---
```

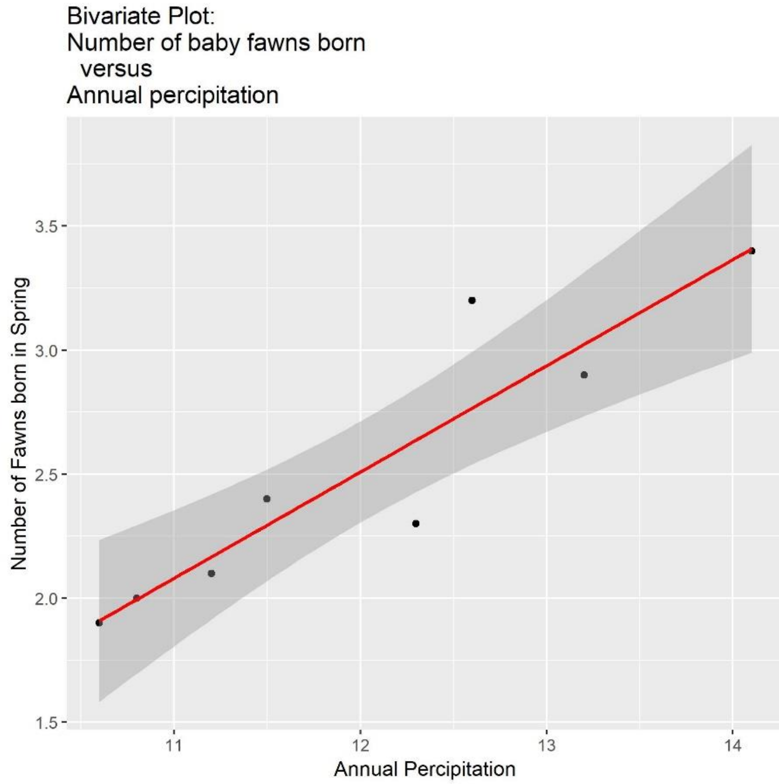
```
> antelopeDataSetURL <- "http://college.cengage.com/mathematics/brase/underst
andable_statistics/7e/students/datasets/mlr/excel/mlr01.xls"
> # Path to local Perl Interpreter - Needed to gdata package for read.xls()
> perlPath <- "C:\\Strawberry\\perl\\bin\\perl"
>
> #---- Load Required Packages ---------------------------------------------
---
> if(!require("devtools")) {install.packages("devtools")}
> if(!require("RCurl")) {install.packages("RCurl")}
> if(!require("gdata")) {install.packages("gdata")}
> if(!require("ggplot2")) {install.packages("ggplot2")}
> if(!require("MASS")) {install.packages("MASS")}
> #if(!require("tidyverse")) {install.packages("tidyverse")}
>
> # ---- Step 1: Load the data ----------------------------------------------
-------------------------------
> ## 1.1: Read the data:
> readDataSetasXLSX <- function(dsURL,locPerlPath){
+
+   df <- read.xls(dsURL, perl=locPerlPath)
+
+   return(data.frame(df))
+ }
>
> antelopeDf <- readDataSetasXLSX(antelopeDataSetURL,perlPath)
trying URL 'http://college.cengage.com/mathematics/brase/understandable_stati
stics/7e/students/datasets/mlr/excel/mlr01.xls'
Content type 'application/vnd.ms-excel' length 5632 bytes
downloaded 5632 bytes

>
> # ---- Step 2: Explore the data -------------------------------------------
----------------------------------
> str(antelopeDf)
'data.frame':   8 obs. of  4 variables:
 $ X1: num  2.9 2.4 2 2.3 3.2 ...
 $ X2: num  9.2 8.7 7.2 8.5 9.6 ...
 $ X3: num  13.2 11.5 10.8 12.3 12.6 ...
 $ X4: int  2 3 4 2 3 5 1 3
>
> # ---- Step 3: Clean the data ---------------------------------------------
--------------------------------
> # Rename Column Headings
> newColumnNames <- c('Fawn_Cnt','Adult_Antelope_Pop','Annual_Percipitation',
'Bad_Winter_Scale')
> colnames(antelopeDf) <- newColumnNames
>
> # ---- Step 4: Create bivariate plots -------------------------------------
-----------------------------------------
> #
> # Create bivariate plots of number of baby fawns versus adult antelope popu
lation,
> # the precipitation that year, and the severity of the winter. Your code sh
ould produce
> # three separate plots. Make sure the Y-axis and X-axis are labeled. Keepin
g in mind
```

```
> # that the number of fawns is the outcome (or dependent) variable, which ax
is should
> # it go on in your plots?
> # Attribute Type:
> #    - Precipitation: Independent / Numeric:Continuous
> #    - Severity of Winter: Independent / Categorical (Scale[1-5])
> #    - Baby Count: Dependent / Numeric:Continuous
> #    - Adult Population: Independent / Numeric:Continuous
>
>
> ## Plots - Focus is on the dependent variable, Baby Fawn Count
>
> # Plot 1: Number of baby fawns versus adult antelope population
> # x-axis: Adult_Antelope_Pop
> # y-axis: Fawn_Cnt
> g.fawn.adult <- ggplot(antelopeDf, aes(x=Adult_Antelope_Pop, y=Fawn_Cnt)) +
+   geom_point() +
+   stat_smooth(method='lm',col='red') +
+   labs(title="Bivariate Plot: \nNumber of baby fawns born \n  versus \nAdul
t antelope population", x="Annual Adult Antelope Population", y="Number of Fa
wns born in Spring")
> g.fawn.adult
> ggsave("Bivariate_Plot_Baby_Fawns_Adult_Population.jpg", width = 6, height
= 6)
```



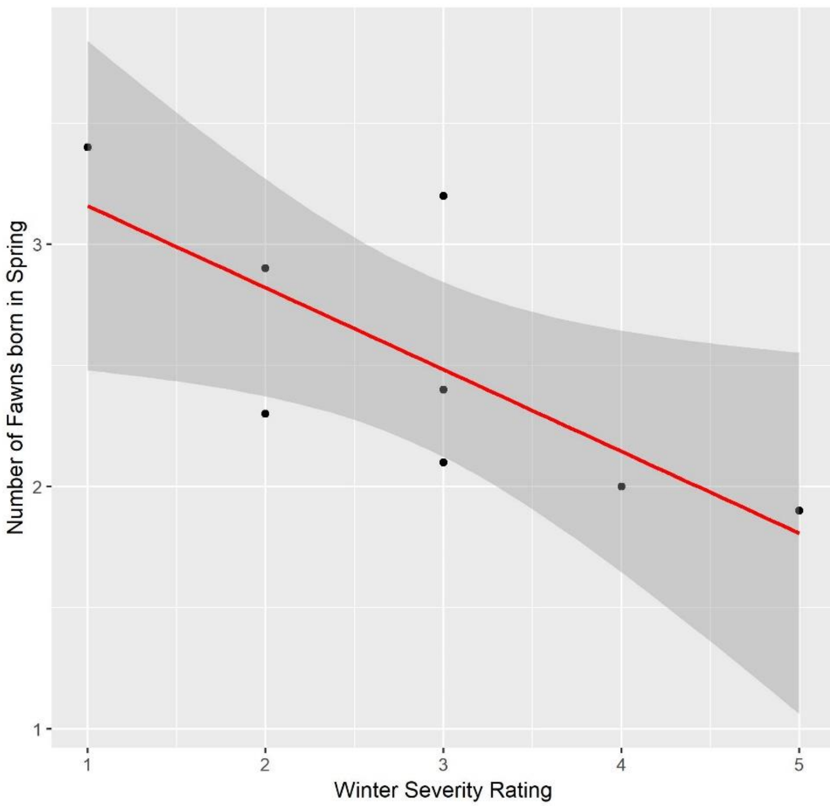

```
> # Plot 2: Number of baby fawns given the precipitation that year
> # x-axis: Annual_Percipitation
> # y-axis: Fawn_Cnt
> g.fawn.percipitation <- ggplot(antelopeDf, aes(x=Annual_Percipitation, y=Fa
wn_Cnt)) +
```

```
+   geom_point() +
+   stat_smooth(method='lm',col='red') +
+   labs(title="Bivariate Plot: \nNumber of baby fawns born \n  versus \nAnnu
al percipitation", x="Annual Percipitation", y="Number of Fawns born in Sprin
g")
> g.fawn.percipitation
> ggsave("Bivariate_Plot_Baby_Fawns_Precipitation.jpg", width = 6, height = 6
)
```




```
> # Plot 3: Number of baby fawns given the severity of the winter
> # x-axis: Bad_Winter_Scale
> # y-axis: Fawn_Cnt
> g.fawn.winter <- ggplot(antelopeDf, aes(x=Bad_Winter_Scale, y=Fawn_Cnt)) +
+   geom_point() +
+   stat_smooth(method='lm',col='red') +
+   labs(title="Bivariate Plot: ", x="Winter Severity Rating", y="Number of F
awns born in Spring")
> g.fawn.winter
> ggsave("Bivariate_Plot_Baby_Fawns_Bad_Weather_Rating.jpg", width = 6, heigh
t = 6)
```

Bivariate Plot:



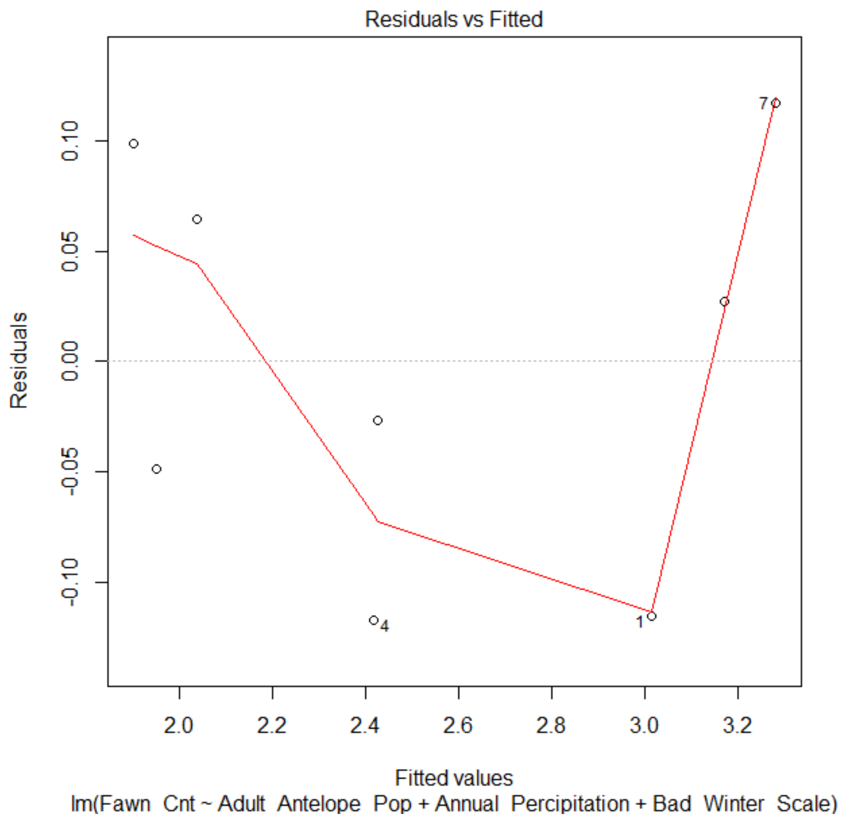

```
> # ---- Step 5: Create regression models ----------------------------------
-------------------------------------------
> #
> # Create three regression models of increasing complexity using lm().
>
> # Step 5.1: Predict the number of fawns from the severity of the winter
> # Output: Multiple R-squared = 0.5459 | p-value = 0.03626
> fawn.winter.lm <- lm(Fawn_Cnt ~ Bad_Winter_Scale, data=antelopeDf)
> sum.f.w <- summary(fawn.winter.lm)
> sum.f.w

Call:
lm(formula = Fawn_Cnt ~ Bad_Winter_Scale, data = antelopeDf)

Residuals:
     Min       1Q   Median       3Q      Max
-0.52069 -0.20431 -0.00172  0.13017  0.71724

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.4966     0.3904   8.957 0.000108 ***
Bad_Winter_Scale  -0.3379     0.1258  -2.686 0.036263 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.415 on 6 degrees of freedom
Multiple R-squared:  0.5459,	Adjusted R-squared:  0.4702
F-statistic: 7.213 on 1 and 6 DF,  p-value: 0.03626
```

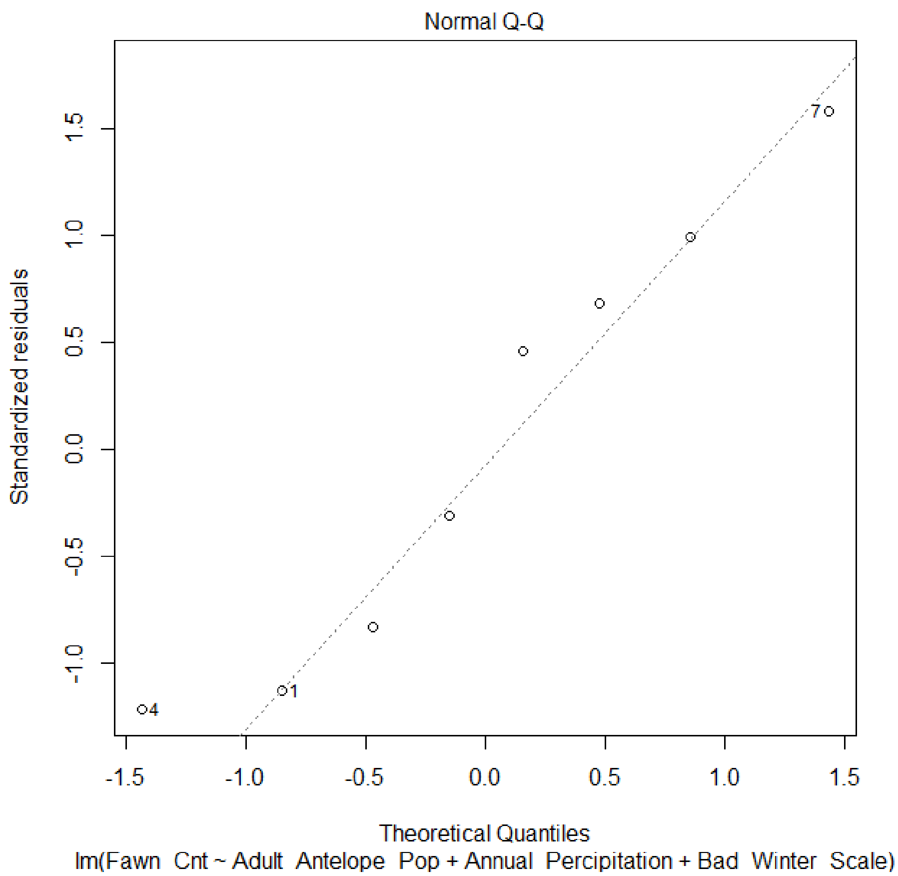
```
> range(antelopeDf$Bad_Winter_Scale)
[1] 1 5
> newWinterData <- data.frame(Bad_Winter_Scale=3)
> predict(fawn.winter.lm,newWinterData,type="response")
       1 
2.482759 
> 
> # Step 5.2: Predict the number of fawns from two variables (one should be t
he severity of the winter)
> # Output: Adjusted R-squared = 0.8439 | p-value = 0.004152
> fawn.AntelopePop.Winter.lm <- lm(Fawn_Cnt ~ Adult_Antelope_Pop+Bad_Winter_S
cale, data=antelopeDf)
> sum.f.a.w <- summary(fawn.AntelopePop.Winter.lm)
> sum.f.a.w

Call:
lm(formula = Fawn_Cnt ~ Adult_Antelope_Pop + Bad_Winter_Scale, 
    data = antelopeDf)

Residuals:
       1        2        3        4        5        6        7        8 
 0.01231 -0.27531  0.10301 -0.19154  0.01535  0.15880  0.29992 -0.12256 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)  
(Intercept)        -2.46009    1.53443  -1.603   0.1698  
Adult_Antelope_Pop  0.56594    0.14439   3.920   0.0112 *
Bad_Winter_Scale    0.07058    0.12461   0.566   0.5956  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2252 on 5 degrees of freedom
Multiple R-squared:  0.8885,	Adjusted R-squared:  0.8439 
F-statistic: 19.92 on 2 and 5 DF,  p-value: 0.004152

> range(antelopeDf$Adult_Antelope_Pop)
[1] 6.8 9.7
> newAntelopeWinterData <- data.frame(Adult_Antelope_Pop=9, Bad_Winter_Scale=
1)
> predict(fawn.AntelopePop.Winter.lm,newAntelopeWinterData,type='response')
      1 
2.70392 
> 
> # Step 5.3: Predict the number of fawns from the three other variables
> # Output: Adjusted R-squared = 0.955 | p-value = 0.001229
> fawn.AntelopePop.Percipitation.Winter.lm <- lm(Fawn_Cnt ~ Adult_Antelope_Po
p+Annual_Percipitation+Bad_Winter_Scale, data=antelopeDf)
> sum.f.a.p.w <- summary(fawn.AntelopePop.Percipitation.Winter.lm)
> sum.f.a.p.w

Call:
lm(formula = Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + 
    Bad_Winter_Scale, data = antelopeDf)

Residuals:
       1        2        3        4        5        6        7        8 
-0.11533 -0.02661  0.09882 -0.11723  0.02734 -0.04854  0.11715  0.06441 
```

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -5.92201    1.25562  -4.716   0.0092 **
Adult_Antelope_Pop     0.33822    0.09947   3.400   0.0273 *
Annual_Percipitation   0.40150    0.10990   3.653   0.0217 *
Bad_Winter_Scale       0.26295    0.08514   3.089   0.0366 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1209 on 4 degrees of freedom
Multiple R-squared:  0.9743,   Adjusted R-squared:  0.955
F-statistic: 50.52 on 3 and 4 DF,  p-value: 0.001229

> range(antelopeDf$Annual_Percipitation)
[1] 10.6 14.1
> newAntelopePopPercipitationWinterData <- data.frame(Adult_Antelope_Pop=5, A
nnual_Percipitation=14 ,Bad_Winter_Scale=4)
> predict(fawn.AntelopePop.Percipitation.Winter.lm, newAntelopePopPercipitati
onWinterData, type='response')
       1
2.441915
>
> # Step 5.4: Questions to Answer
> # Question 5.4.1: Which model works best?
> # The third model, fawn.AntelopePop.Percipitation.Winter.lm, it has an Adju
st R-squared score of 0.955 and p-value of 0.001 (statistically significant)
> paste("Adjusted R-squared:",sum.f.a.p.w$adj.r.squared)
[1] "Adjusted R-squared: 0.955004704934087"
> paste("P-Value:")
[1] "P-Value:"
> sum.f.a.p.w$coefficients[,4]
         (Intercept)   Adult_Antelope_Pop Annual_Percipitation     Bad_Winter
_Scale
         0.009196072          0.027272444          0.021707219          0.036
626174
>
> # Question 5.4.2: Which of the predictors are statistically significant in
each model?
> # Model 1: Number of fawns from the severity of the winter
> # Answer: Bad Winter Rating has a predictor confidence of 55% accuracy
> paste("Multiple R-squared:",sum.f.w$r.squared)
[1] "Multiple R-squared: 0.545888574072554"
> paste("P-Value:")
[1] "P-Value:"
> sum.f.w$coefficients[,4]
     (Intercept) Bad_Winter_Scale
     0.000108158      0.036263036
>
> # Model 2: Number of fawns from the Adult Antelope Population and Bad Winte
r Rating
> # Answer: Adult Antelope Population has a prediction confidence accuracy of
84%, Bad Winter is not statistically significant in this model
> paste("Adjusted R-squared:",sum.f.a.w$adj.r.squared)
[1] "Adjusted R-squared: 0.84389367311556"
> paste("P-Value:")
[1] "P-Value:"
```

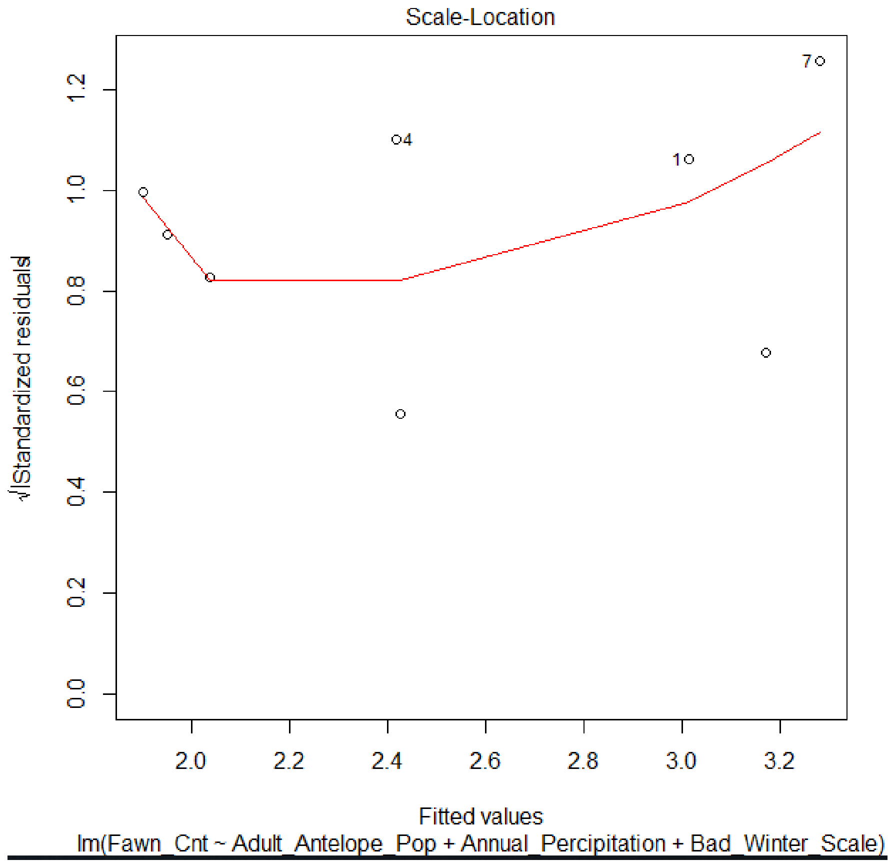
```
> sum.f.a.w$coefficients[,4]
      (Intercept) Adult_Antelope_Pop   Bad_Winter_Scale
       0.16977988         0.01118699         0.59557987
>
> # Model 3: number of fawns from the Adult Antelope Population, Annual Perci
pitation and Bad Winter Rating
> # Answer: Adult Antelope Population, Annual Precipitation and Bad Winter Ra
ting are all statistically significant
> #         and combined have a prediction confidence accuracy of 95%
> paste("Adjusted R-squared:",sum.f.a.p.w$adj.r.squared)
[1] "Adjusted R-squared: 0.955004704934087"
> paste("P-Value:")
[1] "P-Value:"
> sum.f.a.p.w$coefficients[,4]
         (Intercept)    Adult_Antelope_Pop Annual_Percipitation     Bad_Winter
_Scale
         0.009196072          0.027272444          0.021707219          0.036
626174

 # Question 5.4.3: If you wanted to create the most parsimonious model (i.e.,
 > #                the one that did the best job with the fewest predictors), what would
 > # Answer:
 > #   Analysis of Deviance Table
 > #   Initial Model:
 > #     Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale
 > #   Final Model:
 > #     Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale
 >
 > # Apply Stepwise regression analysis, find best combination of variables for best predi
 > lm.step <- stepAIC(fawn.AntelopePop.Percipitation.Winter.lm,direction="both")
 Start:  AIC=-31.35
 Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale

                        Df Sum of Sq      RSS     AIC
 <none>                              0.058494 -31.346
 - Bad_Winter_Scale      1   0.13950 0.197989 -23.592
 - Adult_Antelope_Pop    1   0.16907 0.227561 -22.478
 - Annual_Percipitation  1   0.19518 0.253673 -21.609
 > lm.step$anova
 Stepwise Model Path
 Analysis of Deviance Table

 Initial Model:
 Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale

 Final Model:
 Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale


   Step Df Deviance Resid. Df Resid. Dev       AIC
 1                          4 0.05849389 -31.3462
 > summary(lm.step)

 Call:
 lm(formula = Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation +
     Bad_Winter_Scale, data = antelopeDf)

 Residuals:
        1        2        3        4        5        6        7        8
 -0.11533 -0.02661  0.09882 -0.11723  0.02734 -0.04854  0.11715  0.06441
```

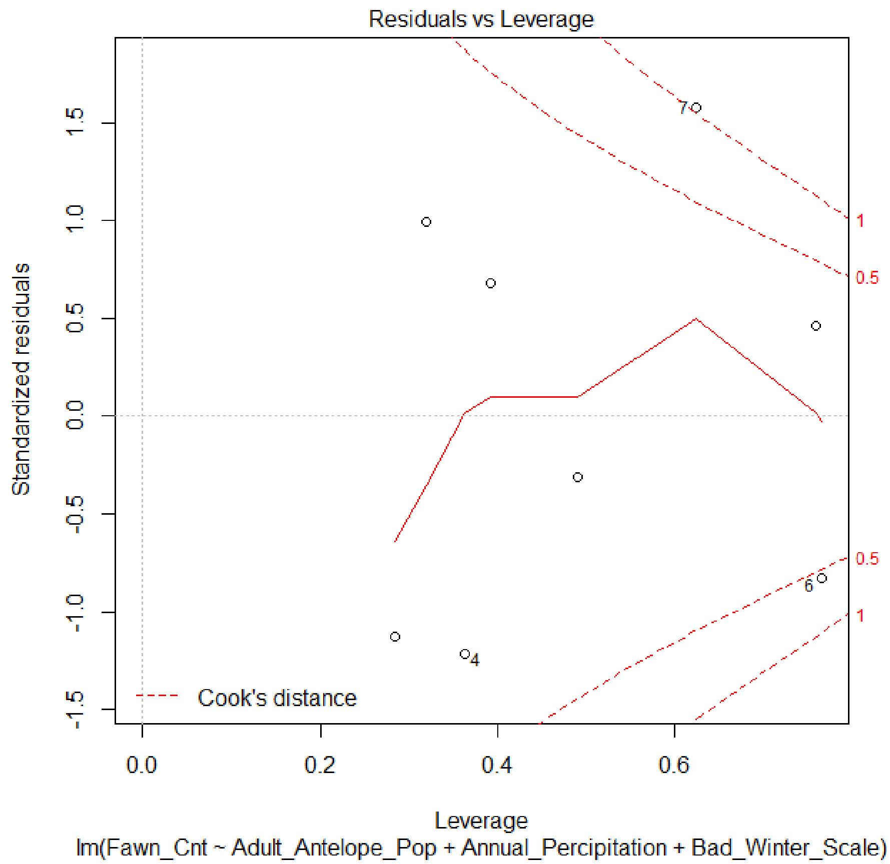
```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            -5.92201    1.25562  -4.716   0.0092 **
Adult_Antelope_Pop      0.33822    0.09947   3.400   0.0273 *
Annual_Percipitation    0.40150    0.10990   3.653   0.0217 *
Bad_Winter_Scale        0.26295    0.08514   3.089   0.0366 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1209 on 4 degrees of freedom
Multiple R-squared:  0.9743,   Adjusted R-squared:  0.955
F-statistic: 50.52 on 3 and 4 DF,  p-value: 0.001229

> plot(lm.step)
```





Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Fawn Cnt ~ Adult Antelope Pop + Annual Percipitation + Bad Winter Scale)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
7
4
1


Scale-Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
4
1
7
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Fitted values
lm(Fawn_Cnt ~ Adult_Antelope_Pop + Annual_Percipitation + Bad_Winter_Scale)




```
## Other individual attribute tests
> # Output: Multiple R-squared = 0.8813 | p-value = 0.0005471
> fawn.apop <- lm(Fawn_Cnt ~ Adult_Antelope_Pop, data=antelopeDf)
> sum.fawn.apop <- summary(fawn.apop)
> sum.fawn.apop

Call:
lm(formula = Fawn_Cnt ~ Adult_Antelope_Pop, data = antelopeDf)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.24988 -0.17586  0.04938  0.12611  0.25309 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        -1.67914    0.63422  -2.648 0.038152 *  
Adult_Antelope_Pop  0.49753    0.07453   6.676 0.000547 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2121 on 6 degrees of freedom
Multiple R-squared:  0.8813,	Adjusted R-squared:  0.8616 
F-statistic: 44.56 on 1 and 6 DF,  p-value: 0.0005471

> # Output: Multiple R-squared = 0.8536 | p-value = 0.001039
> fawn.aperc <- lm(Fawn_Cnt ~ Annual_Percipitation, data=antelopeDf)
> sum.fawn.aperc <- summary(fawn.aperc)
> sum.fawn.aperc
```

```
Call:
lm(formula = Fawn_Cnt ~ Annual_Percipitation, data = antelopeDf)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.33747 -0.08040 -0.00889  0.03023  0.43399 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)   
(Intercept)          -2.63251    0.87591  -3.005  0.02384 * 
Annual_Percipitation  0.42845    0.07244   5.915  0.00104 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2356 on 6 degrees of freedom
Multiple R-squared:  0.8536,	Adjusted R-squared:  0.8292 
F-statistic: 34.99 on 1 and 6 DF,  p-value: 0.001039


>
```