title: "George\_Smith\_HW8"

author: "George Smith"

date: "6/7/2021"

output: pdf\_document

### 1. Read in the data

### 2. 2. You have the option of saving the file save this file to your computer and read it into R, or reading the data directly from the web into a data frame.

```{r}

Antelope <- readXL("C:/Users/GeorgeSmith/Desktop/Antelope.xlsx",

rownames=FALSE, header=TRUE, na="", sheet="Antelope", stringsAsFactors=TRUE)

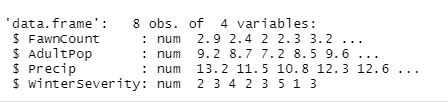
```

### 3. You should inspect the data using the str() command to make sure that all of the cases have been read in (n=8 years of observations) and that there are fourvariables.

```{r}

str(Antelope)

```

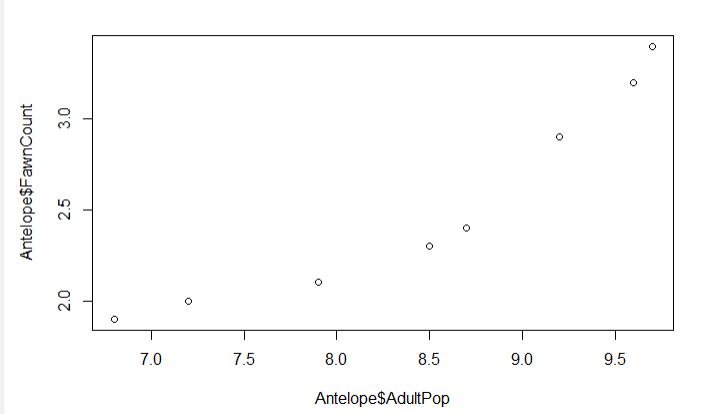


### 4. Create bivariate plots of number of baby fawns versus adult antelope population, the precipitation that year, and the severity of the winter. Your code should produce three separate plots. Make sure the Y-axis and X-axis are labeled. Keeping in mind that the number of fawns is the outcome (or dependent) variable, which axis should it go on in your plots?

```{r}

plot(Antelope$AdultPop, Antelope$FawnCount)

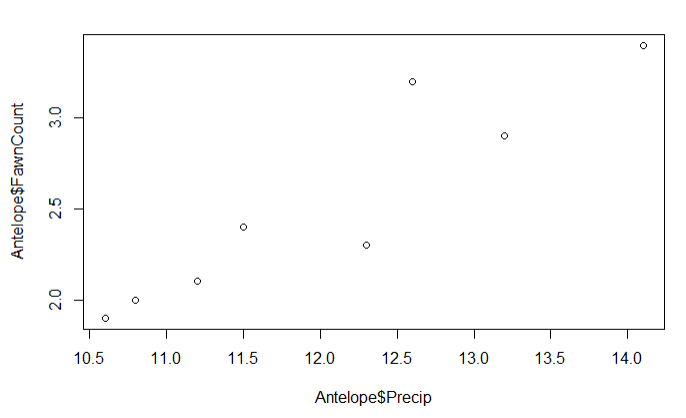
```



```{r}

plot(Antelope$Precip, Antelope$FawnCount)

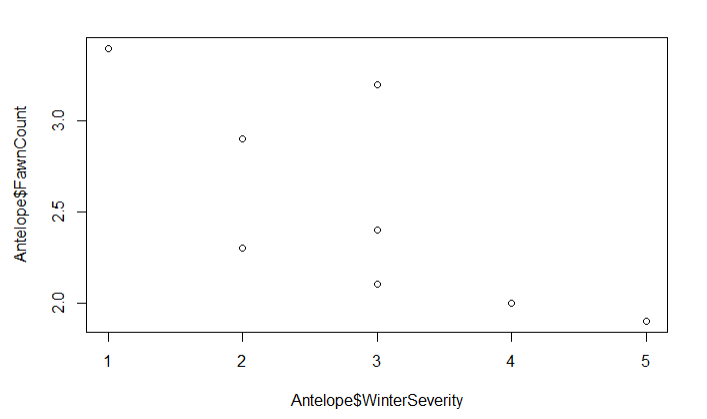
```



```{r}

plot(Antelope$WinterSeverity, Antelope$FawnCount)

```



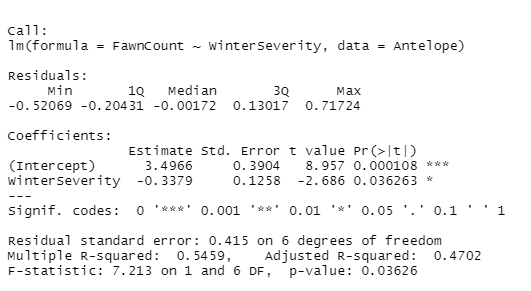
### 5. Next, create three regression models of increasing complexity using lm(). In the first model, predict the number of fawns from the severity of the winter. In the second model, predict the number of fawns from two variables (one should be the severity of the winter). In the third model predict the number of fawns from the three other variables. Which model works best? Which of the predictors are statistically significant in each model? 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?

```{r}

model1 <- lm(formula=FawnCount ~ WinterSeverity, data=Antelope)

summary(model1)

```



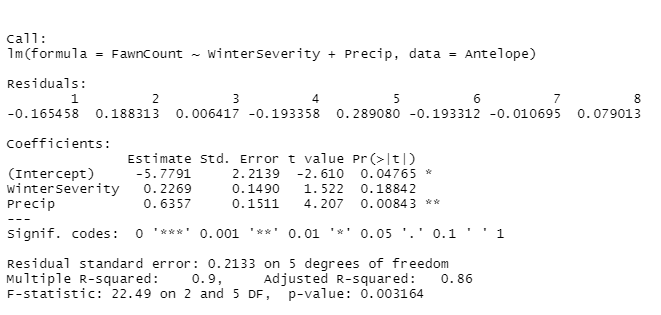
## winter severity has a p value of .036 which is less than .05 so it would be considered significant. the r squared is .47

```{r}

model2 <- lm(formula=FawnCount ~ WinterSeverity + Precip, data=Antelope)

summary(model2)

```



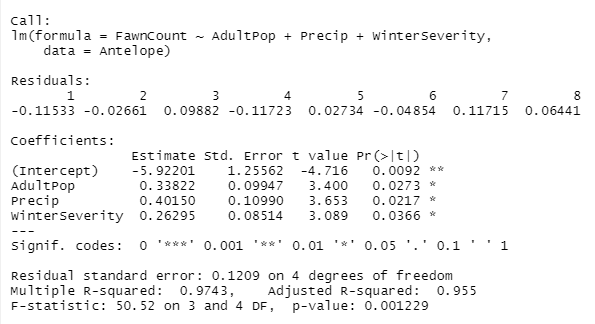
## winter severity in not conisdered significant in this model as the p value is .19. Precip would be considered significant as the p value is .008. this model has an r squared value of .86

```{r}

model3 <- lm(formula=FawnCount ~ AdultPop + Precip + WinterSeverity, data=Antelope)

summary(model3)

```



## all input variables in this model are considered significant as each p value is less than .05 the r squared value is .96

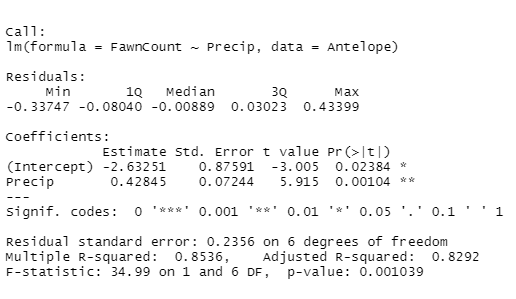
## The most parsimonious model has 1 input variable Precip. The r-squared value of .82 means this 1 value explains 82 percent of the fawn count

```{r}

model4 <- lm(formula=FawnCount ~ Precip, data=Antelope)

summary(model4)

```



## example predictions

```{r}

test=data.frame(Precip=9)

predict(model4,test,type="response")

```



```{r}

test=data.frame(Precip=16)

predict(model4,test,type="response")

```

