IST 687

Marriah Lewis

Homework 8
Due Week 8 and 11/28/2020

#read in the excel dataset

library(readxl)

>mlr01<-read_excel("data.mlr01.xls")

>View(mlr01)

	X1	X2	X3	X4
1	2.9	9.2	13.2	2
2	2.4	8.7	11.5	3
3	2.0	7.2	10.8	4
4	2.3	8.5	12.3	2
5	3.2	9.6	12.6	3
6	1.9	6.8	10.6	5
7	3.4	9.7	14.1	1
8	2.1	7.9	11.2	3

>str(mlr01)

tibble [8 x 4] (S3: tbl_df/tbl/data.frame)

$ X1: num [1:8] 2.9 2.4 2 2.3 3.2 ...

$ X2: num [1:8] 9.2 8.7 7.2 8.5 9.6 ...

$ X3: num [1:8] 13.2 11.5 10.8 12.3 12.6 ...

$ X4: num [1:8] 2 3 4 2 3 5 1 3

#rename the coloumns

>columns<-c("Number of fawn", "Adult antelope", "Annual precipitation", "Winter Condition")

>colnames(mlr01)<-columns

>head(mlr01)

```
# A tibble: 6 x 4
  `Number of fawn` `Adult antelope` `Annual precipitation` `Winter Condition`
           <dbl>          <dbl>                <dbl>              <dbl>
1          2.90           9.20                 13.2                 2
2          2.40           8.70                 11.5                 3
3          2              7.20                 10.8                 4
4          2.30           8.5                  12.3                 2
```

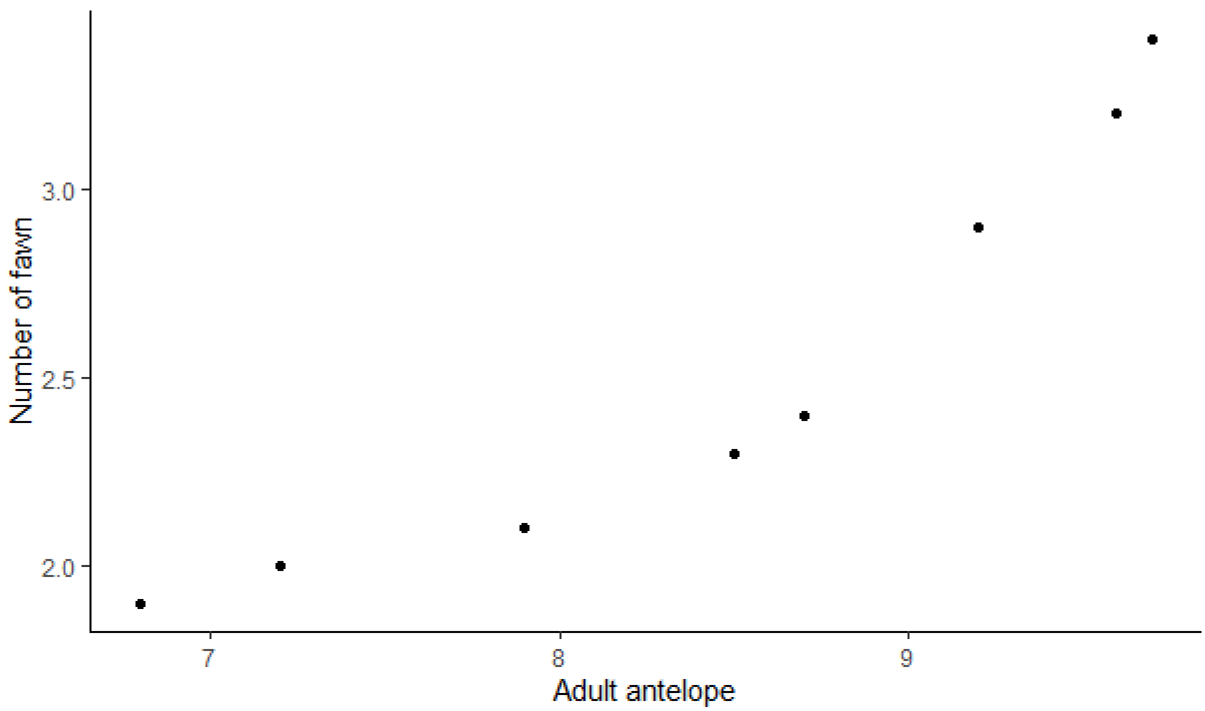
5	3.20	9.6	12.6	3
6	1.90	6.80	10.6	5

#libraries needed to create bivariate plots

```
>install.packages("cowplot")
>library(cowplot)
>library(ggplot2)
```

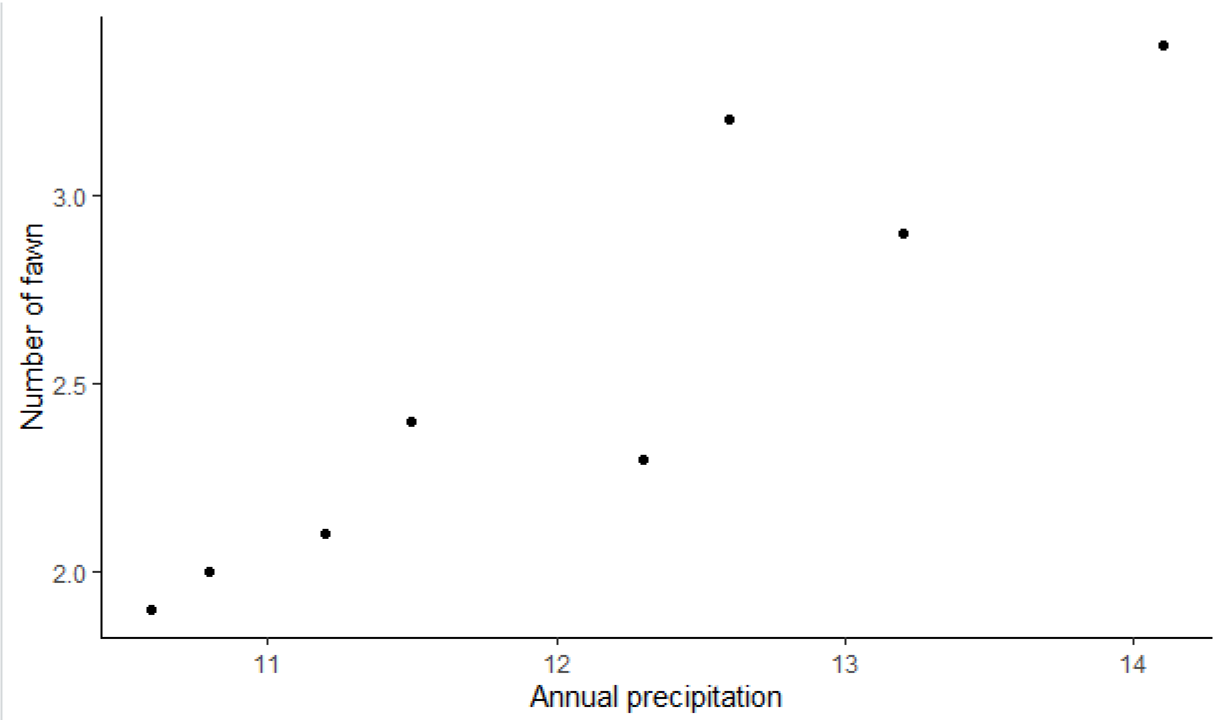
#Number of baby fawns vs adult antelope

```
>AA<-ggplot(mlr01,aes(x=`Adult antelope`, y=`Number of fawn`))+ geom_point()+theme_classic()
>AA
```



#Number of baby fawns vs Annual precipitation

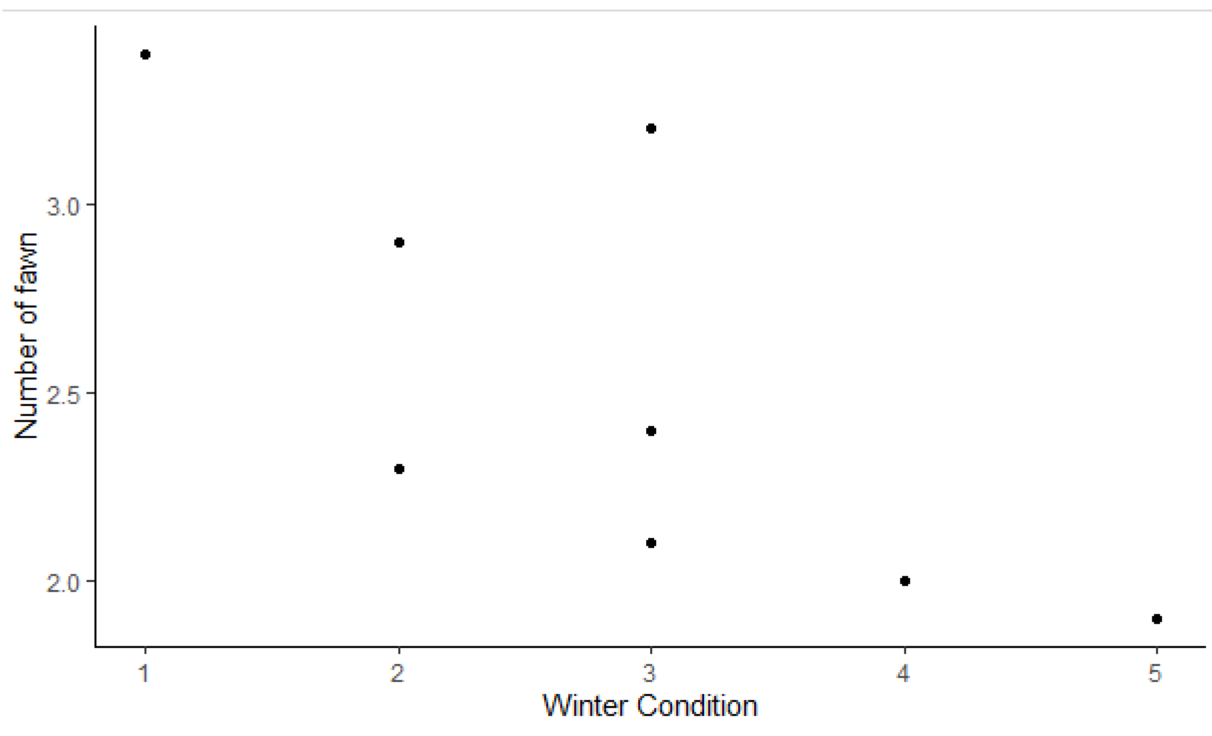
```
>P<-ggplot(mlr01, aes(x=`Annual precipitation`, y=`Number of fawn`))+geom_point()+theme_classic()
>P
```



#Number of baby fawns vs Winter conditions

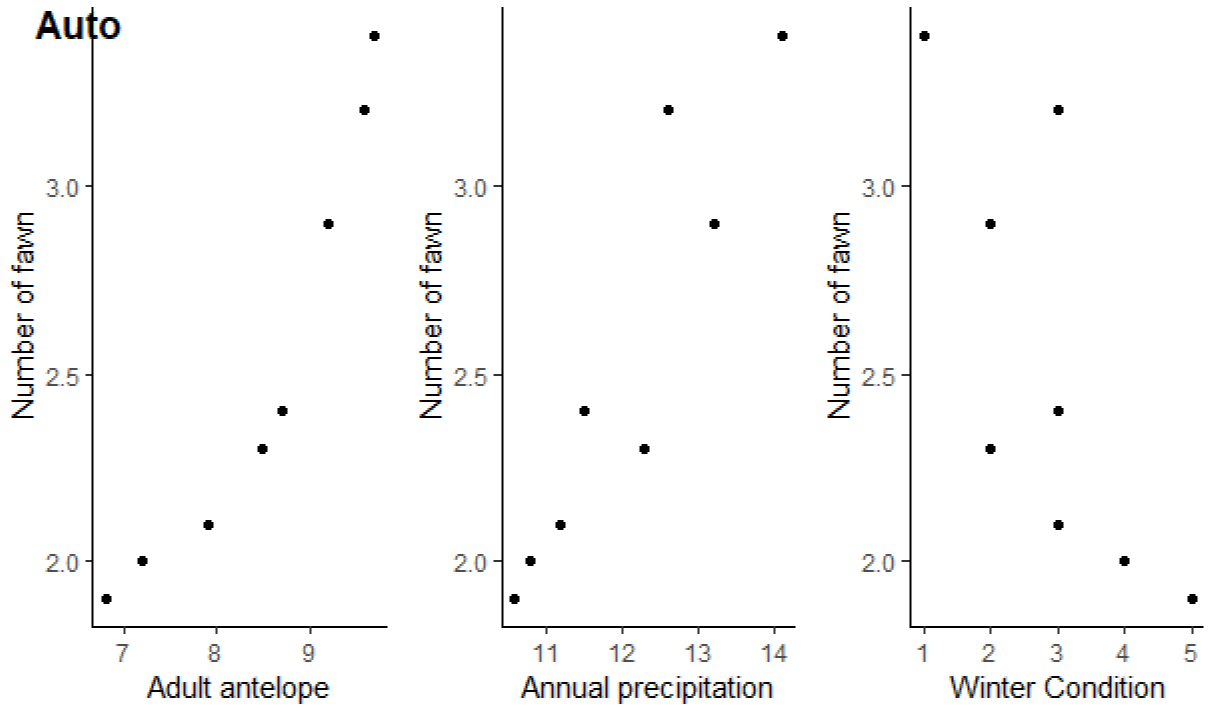
>W<-ggplot(mlr01, aes(x=`Winter Condition`, y=`Number of fawn`))+geom_point()+theme_classic()

>W



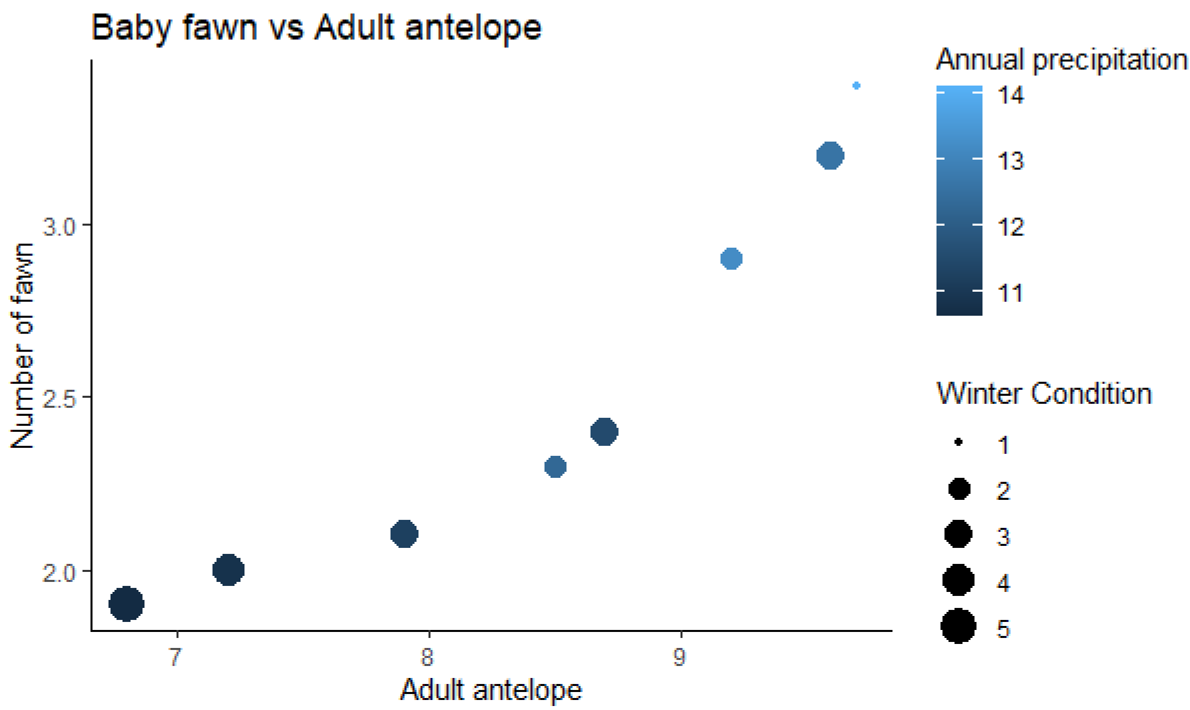
#All three plots

>plot_grid(AA, P, W, nrow= 1, ncol= 3, labels= "Auto")



#Baby fawn vs Adult antelope

>ggplot(mlr01, aes(x=`Adult antelope`, y=`Number of fawn`))+geom_point(aes(color=`Annual precipitation`, size=`Winter Condition`))+ggtitle("Baby fawn vs Adult antelope")+theme_classic()



#Model 1: Predict the number of fawns from winter condition

```
>model1<-lm(`Number of fawn`~ `Winter Condition`, data = mlr01)

>summary(model1)

Call:

lm(formula = `Number of fawn` ~ `Winter Condition`, data = mlr01)


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 ***

`Winter Condition`  -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
>anova(model1)
Analysis of Variance Table

Response: Number of fawn
                   Df Sum Sq Mean Sq F value  Pr(>F)
`Winter Condition`  1 1.2419 1.24190  7.2126 0.03626 *
Residuals           6 1.0331 0.17218
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#Model 2: Number of fawns with two variables(Winter Condition and Adult antelope population)
>model2<-lm(formula = `Number of fawn` ~ `Winter Condition`+ `Adult antelope`, data = mlr01)
>summary(model2)
Call:
lm(formula = `Number of fawn` ~ `Winter Condition` + `Adult antelope`,
    data = mlr01)

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
`Winter Condition`  0.07058    0.12461   0.566   0.5956
`Adult antelope`    0.56594    0.14439   3.920   0.0112 *
```

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

>anova(model2)

Analysis of Variance Table

Response: Number of fawn

```
                    Df  Sum Sq Mean Sq F value   Pr(>F)
`Winter Condition`  1 1.24190 1.24190  24.478 0.004294 **
`Adult antelope`    1 0.77943 0.77943  15.363 0.011187 *
Residuals           5 0.25367 0.05073
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#Model 3: All 3 of the variables

>model3<-lm(formula = `Number of fawn` ~ `Winter Condition`+ `Adult antelope`+ `Annual precipitation`, data = mlr01)

>summary(model3)

Call:

```
lm(formula = `Number of fawn` ~ `Winter Condition` + `Adult antelope` +
    `Annual precipitation`, data = mlr01)
```

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 **
`Winter Condition`     0.26295    0.08514   3.089   0.0366 *
`Adult antelope`       0.33822    0.09947   3.400   0.0273 *
`Annual precipitation`  0.40150    0.10990   3.653   0.0217 *
---
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
>anova(model3)
Analysis of Variance Table

Response: Number of fawn
                    Df  Sum Sq Mean Sq F value    Pr(>F)
`Winter Condition`      1 1.24190 1.24190  84.925 0.0007704 ***
`Adult antelope`        1 0.77943 0.77943  53.300 0.0018717 **
`Annual precipitation`  1 0.19518 0.19518  13.347 0.0217072 *
Residuals             4 0.05849 0.01462
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Questions:

Which model works best? Which of the predictors are statistically significant in each model? If you wanted to create the most parsimonious model, what would it contain?

- Model 3 works best to predict which variable is statistically significant. The two variables that are significant are winter conditions and adult antelopes. Regression model provides evidence that as the number of adult antelopes increase, the number of fawns increase as well. As winter conditions increase the number of fawns decrease. The y-intercept and the slope coefficient for adult antelopes and winter conditions are statistically significant because the p-values for each are much smaller than 0.05. The variable that is most significant is winter conditions due to the pr(|>t|) = 0.0007704, which is considerably less than the p-value cutoff point of 0.05. I would create a model that contains x= Annual Winter Conditions and the total number of Adult Antelopes in cold states and y= Total number of fawns.