HW8.R

xboxv

2020-04-05

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
library("gdata")
```

```
## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location of a
## gdata: valid perl intrpreter.
## gdata:
## gdata: (To avoid display of this message in the future, please ensure
## gdata: perl is installed and available on the executable search path.)
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.
```

```
##
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
```

```
##
```

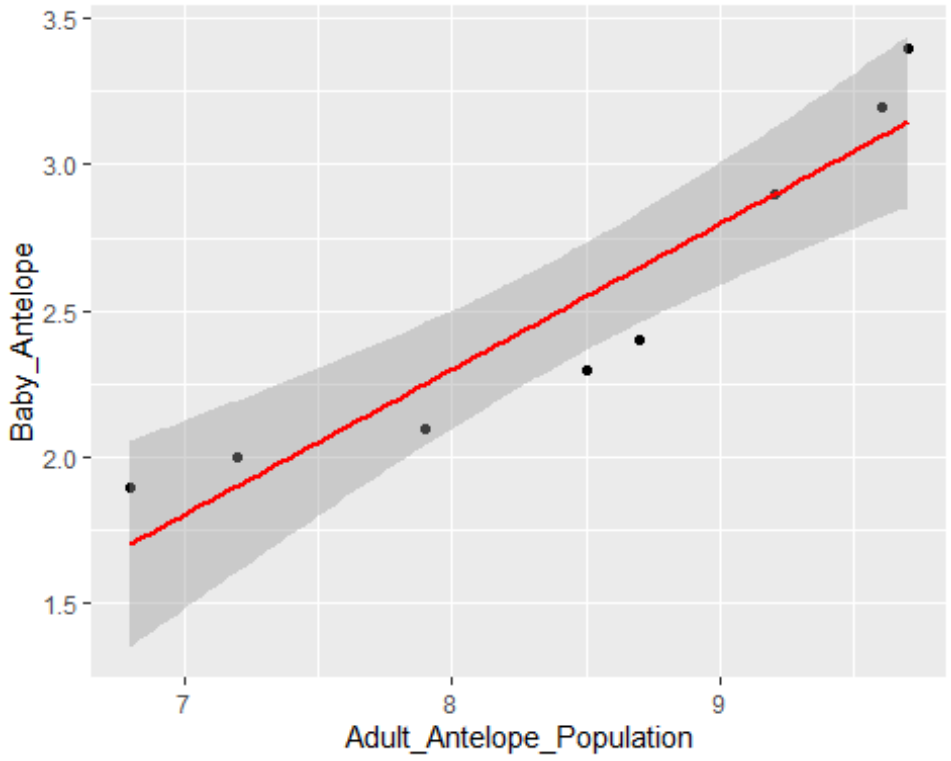
```
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.
```

```
##
## Attaching package: 'gdata'
```

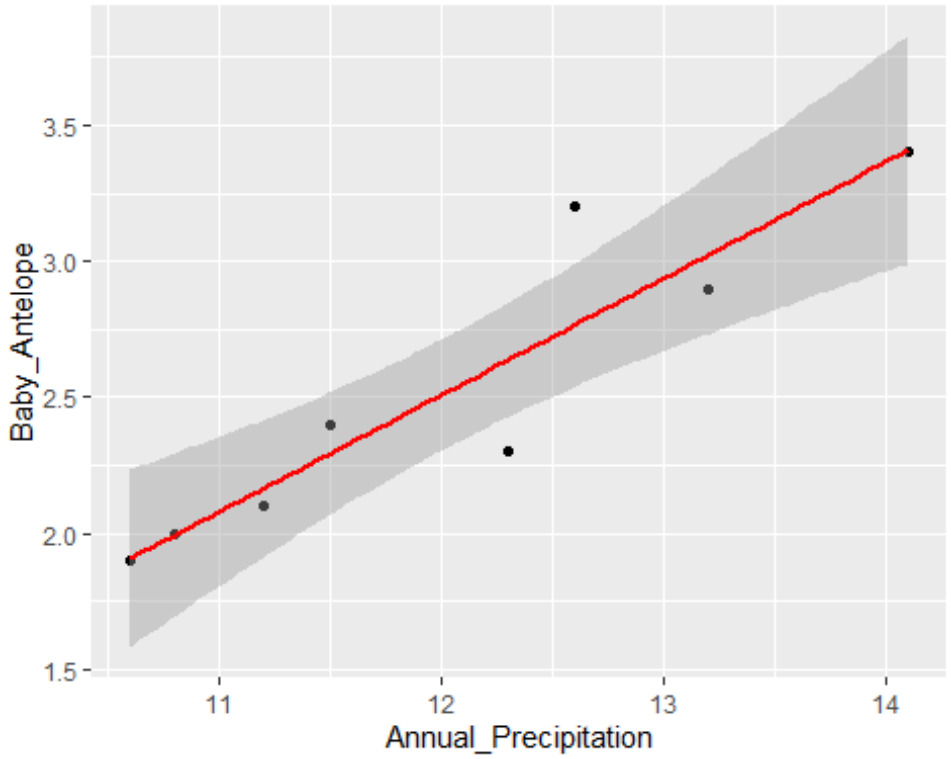
```
## The following object is masked from 'package:stats':
##
##     nobs
```

```
## The following object is masked from 'package:utils':
##
##     object.size
```

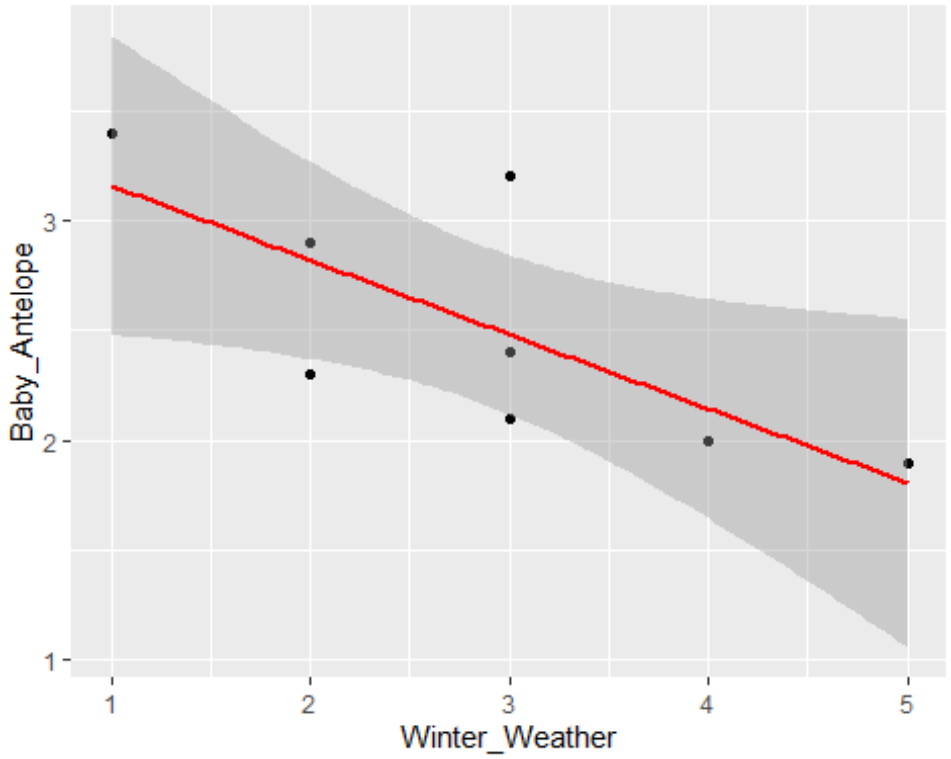
```
## The following object is masked from 'package:base':
##
##     startsWith

link <-
"http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/excel/mlr01.xls"

#download the data into an excell file
download.file(link,destfile = "./file.xls", mode = 'wb')

#view the data
datafile <- read_xls("file.xls")
str(datafile)

## Classes 'tbl_df', 'tbl' and '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: num  2 3 4 2 3 5 1 3

#columns renaming
columns <- c("Baby_Antelope",
"Adult_Antelope_Population","Annual_Precipitation","Winter_Weather")
colnames(datafile) <- columns



#bivariate  plots
Graph1 <- ggplot(datafile, aes(x=Adult_Antelope_Population, y=Baby_Antelope))
+ geom_point()  + stat_smooth(method = "lm", col ="red")
Graph1

## `geom_smooth()` using formula 'y ~ x'
```




```
Graph2 <- ggplot(datafile, aes(x=Annual_Precipitation, y=Baby_Antelope)) +
geom_point() + stat_smooth(method = "lm", col = "red")
Graph2

## `geom_smooth()` using formula 'y ~ x'
```




```
Graph3 <- ggplot(datafile, aes(x=Winter_Weather, y=Baby_Antelope)) +
geom_point() + stat_smooth(method = "lm", col = "red")
Graph3

## `geom_smooth()` using formula 'y ~ x'
```




```
#Regression model 1
#using winter condition to predict number of fawns
M1_data <- datafile[,c(1,4)]
M1 <- lm(Baby_Antelope ~ ., data = M1_data)
summary(M1)

##
## Call:
## lm(formula = Baby_Antelope ~ ., data = M1_data)
##
## 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_Weather  -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

#model 2
#using winter condition and adult population to predict number of fawns
```

```r
M2_data  <- datafile[,-3]

M2 <- lm(Baby_Antelope ~ ., data = M2_data)
summary(M2)
```

```
##
## Call:
## lm(formula = Baby_Antelope ~ ., data = M2_data)
##
## 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_Population   0.56594    0.14439   3.920   0.0112 *
## Winter_Weather              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
```

```r
#model 3
#using all three variables
M3 <- lm(Baby_Antelope ~ ., data = datafile)
summary(M3)
```

```
##
## Call:
## lm(formula = Baby_Antelope ~ ., data = datafile)
##
## 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_Population   0.33822    0.09947   3.400   0.0273 *
## Annual_Precipitation        0.40150    0.10990   3.653   0.0217 *
## Winter_Weather              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
```

```
#Which model works best?
#Ans: Model 3 because it had the strongest P value.

#Which of the predictors are statistically significant in each model?
#Ans:In model 1 Adult population is the significant predictor. model 2 and 3
all the predictors are significant with about the same P values.

#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?
#Something similar to model 2 because it was a good middle ground between P
level and use of predictors.
```