LAB ASSIGNEMENT5

KAMBALE NDULANI Sagesse

January 6, 2021

221000611

1. In the R package MASS there is a dataset called cats. Run the following commands: library(MASS) data(cats) Have a look at the dataset. The variables Bwt and Hwt give the weight of the body (kg) and the heart (g), respectively. There are both male and female cats. Make a dataset with the data from males only.

```
library(MASS)
```

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

```
data(cats)

#print the summary of data inside the dataset cats
summary(cats)
```

```
##  Sex         Bwt             Hwt
##  F:47   Min.   :2.000   Min.   : 6.30
##  M:97   1st Qu.:2.300   1st Qu.: 8.95
##         Median :2.700   Median :10.10
##         Mean   :2.724   Mean   :10.63
##         3rd Qu.:3.025   3rd Qu.:12.12
##         Max.   :3.900   Max.   :20.50
```

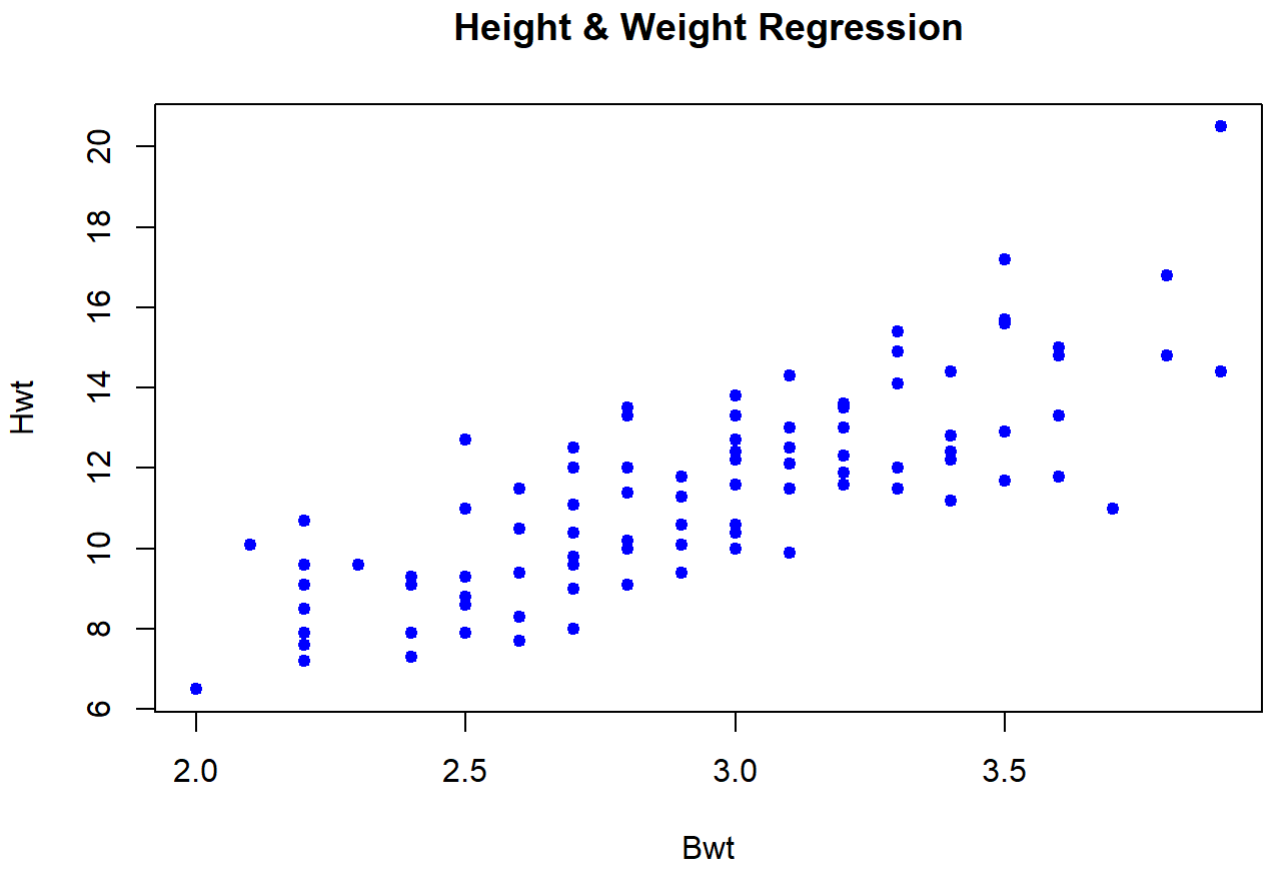
Make a dataset with the data from males only.

```
males_cats <-cats[cats$Sex=="M",] #making a subset from our dataset
print(summary(males_cats)) # printing the summary of the created dataset
```

```
##  Sex         Bwt           Hwt
##  F: 0   Min.   :2.0   Min.   : 6.50
##  M:97   1st Qu.:2.5   1st Qu.: 9.40
##         Median :2.9   Median :11.40
##         Mean   :2.9   Mean   :11.32
##         3rd Qu.:3.2   3rd Qu.:12.80
##         Max.   :3.9   Max.   :20.50
```

2. Make a scatterplot of the data for the male cats (Bwt on x-axis, Hwt on y-axis). Does it look reasonable to use a linear regression model for the data?

```
Bwt<-males_cats$Bwt
Hwt<-males_cats$Hwt
plot(Bwt,Hwt,col = "blue",main = "Height & Weight Regression",cex = 1.3,pch = 20,xlab = "Bwt",ylab = "Hwt")
```

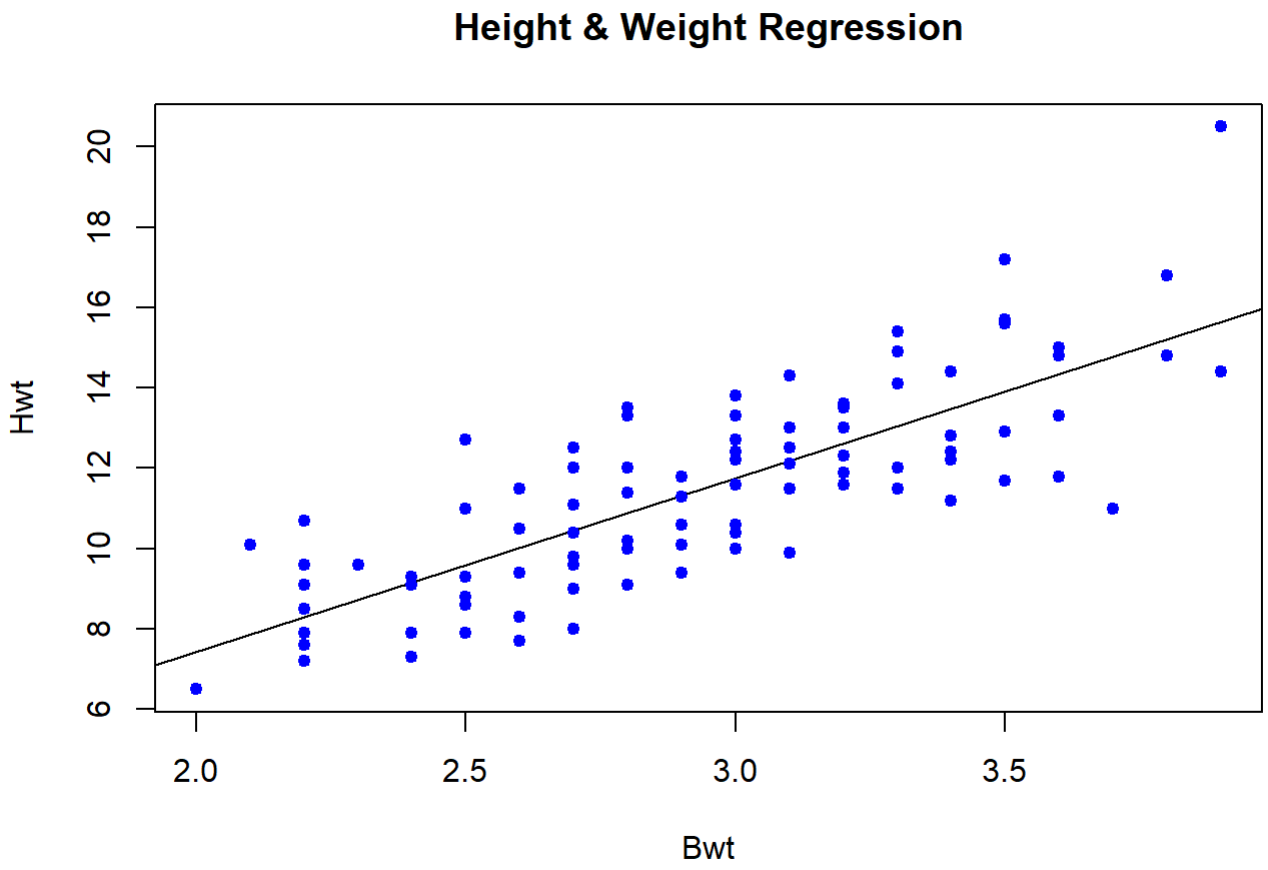


Does it look reasonable to use a linear regression model for the data?

To better understand the relationsheep between the body weight and the heart weight we have to use a linear regression model.From the above plot, we find that the scatterplot indicates an increasing trends so by using a linear regression model, we will get a probable association between the Body weight and the Heart weight

3. Fit a linear regresison model for the male cats, that allows for prediction of the heart weight given the body weight. Add the fitted regression line to the scatterplot from the previous question.

```
relation<-lm(Hwt~Bwt) #creating a linear model with lm function
plot(Bwt,Hwt,col = "blue",main = "Height & Weight Regression",
abline(relation),cex = 1.3,pch = 20,xlab = "Bwt",ylab = "Hwt")
```




```
print(summary(relation)) #printing the summary of our model for a better understanding of our data
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7728 -1.0478 -0.2976  0.9835  4.8646
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1841     0.9983  -1.186    0.239
## Bwt           4.3127     0.3399  12.688   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.557 on 95 degrees of freedom
## Multiple R-squared:  0.6289, Adjusted R-squared:  0.625
## F-statistic:   161 on 1 and 95 DF,  p-value: < 2.2e-16
```

4. Find the coefficients of the fitted line. How large is the expected difference in heart weight for two cats with a difference of 1 kg in bodyweight? Find a confidence interval for this difference? How large is the expected difference in heart weight for two cats with a difference of 100 g in bodyweight?

Find the coefficients

```
#find the coefficients
# applying lm function to find the relationship model and the coefficients
print(relation)
```

```
##
## Call:
## lm(formula = Hwt ~ Bwt)
##
## Coefficients:
## (Intercept)          Bwt
##      -1.184        4.313
```

Coefficients are –1.184 and 4.313

How large is the expected difference in heart weight for two cats with a difference of 1 kg in bodyweight? Find a confidence interval for this difference?

```
newdata <- data.frame(Bwt=1)
predict(relation, newdata)
```

```
##        1
## 3.128591
```

```
predict(relation, newdata, interval="confidence")
```

```
##        fit     lwr      upr
## 1 3.128591 1.80867 4.448512
```

<

The expected difference in heart weight for two cats with a difference of 1Kg is large by *3.128591* with a confidence interval *[1.80867, 4.448512]*

How large is the expected difference in heart weight for two cats with a difference of 100 g in bodyweight?

100g is equal to 0.1Kg

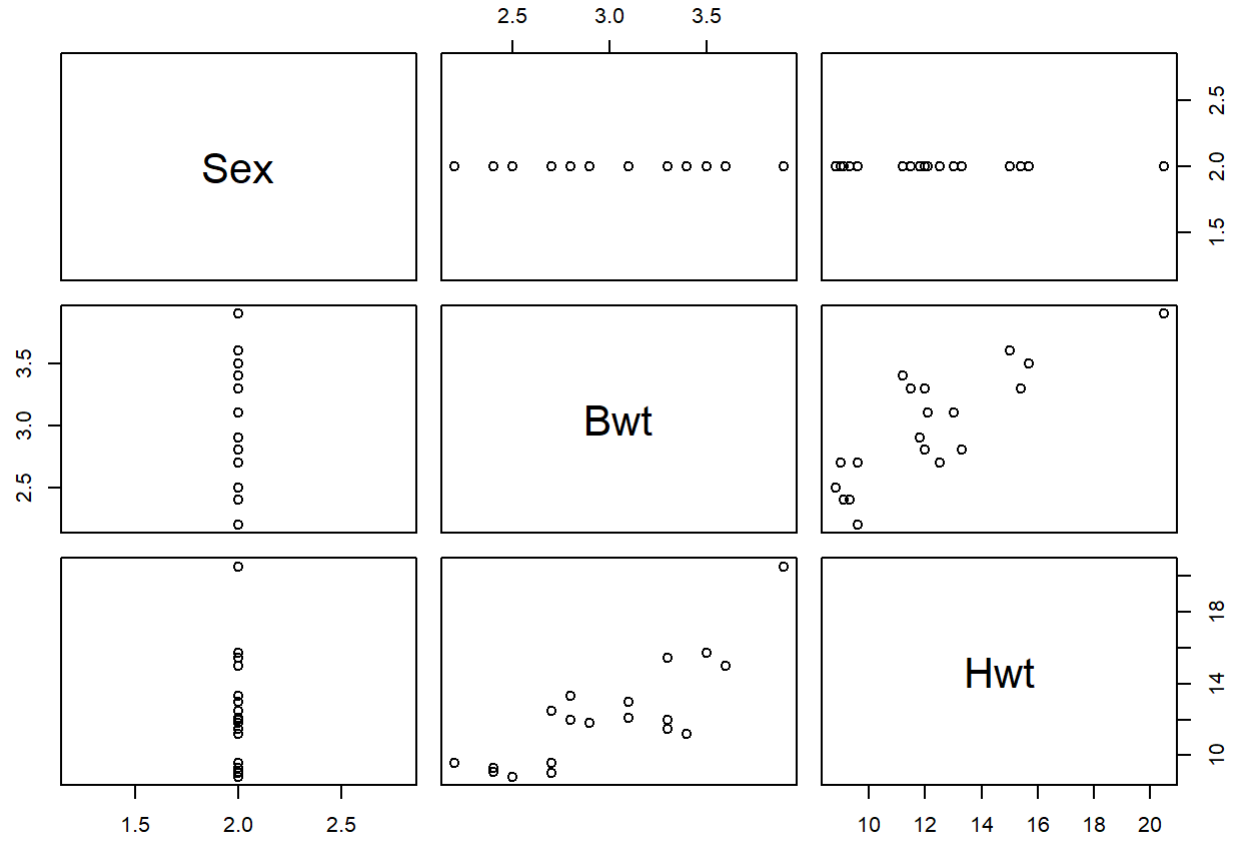
```
predict(relation, data.frame(Bwt=0.1) ,interval="confidence")
```

```
##          fit       lwr      upr
## 1 -0.7528201 -2.668075 1.162435
```

The expected difference in heart weight for two cats with a difference in heart weight of 100g is large by *-0.7528201* with a confidence interval *[-2.668075 ,1.162435]*

5. Use model validation plot to examine if the model is appropriate for the data.

```
Samples<-sample(seq(1,3),size=nrow(males_cats),replace=TRUE,prob=c(0.8,0.2,0.2))
Validate<-males_cats[Samples==3,] # validating our
plot(Validate)
```



6. Use the estimates to find the expected heart weight for a male cat that weighs 3 kg. Then try the commands (where you replace the name regModel with whatever name you gave the the model fit in question 2). newObs <- data.frame(Bwt=3) newObs predict(regModel, newObs) predict(regModel, newObs, interval="predict")

```
newObs <- data.frame(Bwt=3)
newObs
```

```
##   Bwt
## 1   3
```

```
predict(relation, newObs)
```

```
##        1
## 11.75395
```

```
predict(relation, newObs, interval="predict")
```

```
##        fit      lwr      upr
## 1 11.75395 8.646589 14.86131
```