assignment 6

Oyedayo Oyelowo

2 November 2017

#puudata.xls is used in every exercise  
#clear memory  
rm(list = ls())  
setwd("C:/Users/oyeda/Desktop/R\_COURSE/assignment6")  
#load the data  
data<-read.table("puudata.txt", header = T, sep = "\t")

### Exercise 1

* Construct a linear height model for spruces (PUULAJI = 2) in the - first canopy cover layer. Use
* diameter (LPM) as an explanatory variable. Report the model's - - parameters, coefficient of
* determination, and residual error. Report also the p-value of the - explanatory variable.

spruceCanopy1 <- subset(data, data$PUULAJI==2 & data$LATVKERROS==1)  
lm\_fit = lm(formula= PITUUS~LPM, data=spruceCanopy1)  
summary(lm\_fit)

##   
## Call:  
## lm(formula = PITUUS ~ LPM, data = spruceCanopy1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -298.656 -16.420 -1.221 17.060 114.284   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.726999 1.414875 22.42 <2e-16 \*\*\*  
## LPM 0.676181 0.006779 99.75 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 27.77 on 1709 degrees of freedom  
## Multiple R-squared: 0.8534, Adjusted R-squared: 0.8533   
## F-statistic: 9950 on 1 and 1709 DF, p-value: < 2.2e-16

names(lm\_fit)

## [1] "coefficients" "residuals" "effects" "rank"   
## [5] "fitted.values" "assign" "qr" "df.residual"   
## [9] "xlevels" "call" "terms" "model"

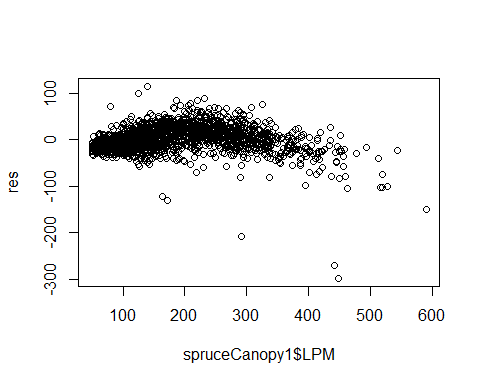
* Parameters: "LPM" and "intercept"
* coefficient of determination: 0.8534
* p-value: < 2.2e-16

## Exercise 2

### Exercise 1 continues.

* **Plot the model's residual error as a function of diameter.**
* **Does the variance change as diameter grows?**

res =residuals(lm\_fit)  
plot(res~ spruceCanopy1$LPM)



* **The variance changes slightly as the diameter grows. There are a few larger residuals as the diameter grows**

I also tried my hands on the qq-plots

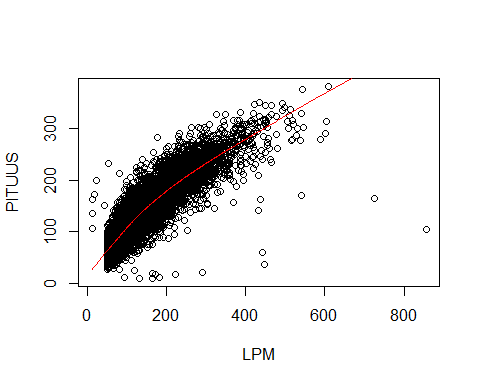
#{oldpar <- par(mfrow = c(2, 2))  
#plot(lm\_fit)  
#par(oldpar)}

## Exercise 3

### Exercise 1 continues.

* **Plot the height of spruces in the first canopy layer as a function of diameter. Add a red curve to the picture that illustrates the values predicted with the model.**

spr1<- data[data["LATVKERROS"]==1,]  
plot(PITUUS~LPM, spr1)  
lines(lowess(spr1$LPM, spr1$PITUUS), col="red")



#abline(lm\_fit, col="purple")  
#or  
#abline(coef = coef(lm\_fit), col="red")

## Exercise 4

### Exercise 1 continues.

* Develop the model further. Add new explanatory variables from the puudata and examine how model's coefficient of determination and residual error #change. Can you make the model better and if so, which explanatory variables belong into the model?

lm\_fit2= lm(formula= PITUUS~LPM+ELAVALARAJA, data=spruceCanopy1)  
lm\_fit3= lm(formula= PITUUS~LPM+ELAVALARAJA+LATVUSLEV, data=spruceCanopy1)  
summary(lm\_fit2)

##   
## Call:  
## lm(formula = PITUUS ~ LPM + ELAVALARAJA, data = spruceCanopy1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -291.78 -12.94 -0.94 14.08 116.64   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.018163 1.278822 23.47 <2e-16 \*\*\*  
## LPM 0.595353 0.007347 81.04 <2e-16 \*\*\*  
## ELAVALARAJA 0.365241 0.018412 19.84 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 25.04 on 1708 degrees of freedom  
## Multiple R-squared: 0.8809, Adjusted R-squared: 0.8807   
## F-statistic: 6314 on 2 and 1708 DF, p-value: < 2.2e-16

summary(lm\_fit3)

##   
## Call:  
## lm(formula = PITUUS ~ LPM + ELAVALARAJA + LATVUSLEV, data = spruceCanopy1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -295.252 -12.822 -0.758 13.863 116.796   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.257848 1.303859 23.973 < 2e-16 \*\*\*  
## LPM 0.620567 0.009336 66.471 < 2e-16 \*\*\*  
## ELAVALARAJA 0.366603 0.018319 20.012 < 2e-16 \*\*\*  
## LATVUSLEV -0.353131 0.081355 -4.341 1.5e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 24.91 on 1707 degrees of freedom  
## Multiple R-squared: 0.8822, Adjusted R-squared: 0.882   
## F-statistic: 4260 on 3 and 1707 DF, p-value: < 2.2e-16

summary(lm\_fit)

##   
## Call:  
## lm(formula = PITUUS ~ LPM, data = spruceCanopy1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -298.656 -16.420 -1.221 17.060 114.284   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.726999 1.414875 22.42 <2e-16 \*\*\*  
## LPM 0.676181 0.006779 99.75 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 27.77 on 1709 degrees of freedom  
## Multiple R-squared: 0.8534, Adjusted R-squared: 0.8533   
## F-statistic: 9950 on 1 and 1709 DF, p-value: < 2.2e-16

**ELAVALARAJA(lower limit of living canopy ) was added the coefficient of determination increased from 0.8534 to 0.8809 and the residual standard error reduced from 27.77 to 25. 04**

* **However adding LATVUSLEV(width of the canopy) didn't improve the prediction considerably. The coefficient of determination only increased from 0.8809 to 0.8822 while the residual standard error only reduced neglibigly from 25.04 to 24.91. The P value of LATVUSLEV was also higher and close to 0.05 compared to others there are much more lower. Also the standard error is much higher.**