TP2 ModeloLinealMultivariado

Marcos Buccellato

2023-06-30

Table of Contents

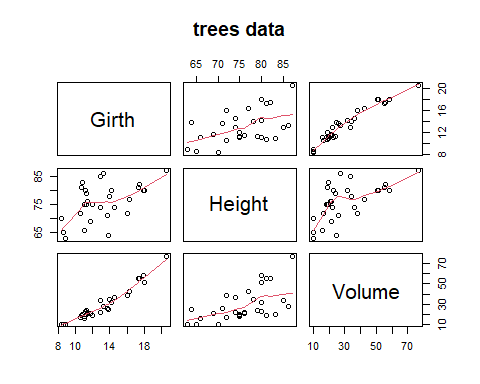
# 2.1. Modelo Aditivo

## Ejercicio 2.1 Con el set de datos trees, disponible en la biblioteca dplyr de R, pretendemos ajustar un modelo que estimo el volumen (en pies cúbicos) de los árboles de cerezo en función de la longitud de su circunferencia (en pulgadas) y de su altura (en pies).

library(dplyr)

### a) Visualizarla asociación entre las variables de a pares.

pairs(trees, panel = panel.smooth, main = "trees data")

 Sin embargo si hacemos el test de binormalidad vemos que no todas las variables son binormales. Por lo cual no sirve hacer correlación de perason para ambas

rbind(mvn(data = trees[c(1,2)], mvnTest = "hz")$multivariateNormality,  
 mvn(data = trees[c(1,3)], mvnTest = "hz")$multivariateNormality,  
 mvn(data = trees[c(2,3)], mvnTest = "hz")$multivariateNormality  
)

## Test HZ p value MVN  
## 1 Henze-Zirkler 0.6636803 0.127490598 YES  
## 2 Henze-Zirkler 1.0912171 0.007796525 NO  
## 3 Henze-Zirkler 0.9019672 0.026926833 NO

Entonces cambiamos de método:

library(ggplot2)  
library(GGally)

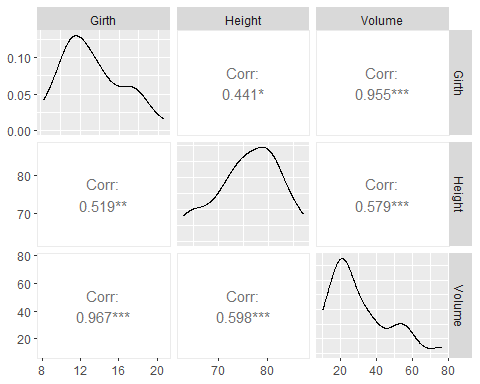
## Warning: package 'GGally' was built under R version 4.3.1

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

ggpairs(trees, upper = list(continuous = wrap("cor",method="spearman")),  
 lower =list(continuous = wrap("cor",method="pearson")))

## Warning in cor.test.default(x, y, method = method, use = use): Cannot compute  
## exact p-value with ties

## Warning in cor.test.default(x, y, method = method, use = use): Cannot compute  
## exact p-value with ties  
  
## Warning in cor.test.default(x, y, method = method, use = use): Cannot compute  
## exact p-value with ties



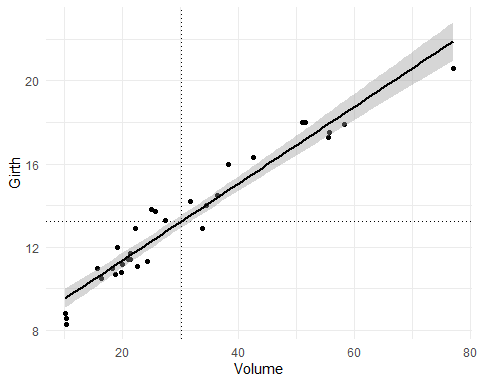
Para girth y height hay que mirar la posicion 2,1. Para girth y volume 1,3 y para height volume 2,3

### b) Ajuste un modelo lineal simple para cada una de las dos predictoras disponibles.

**Volume ~ Girth**

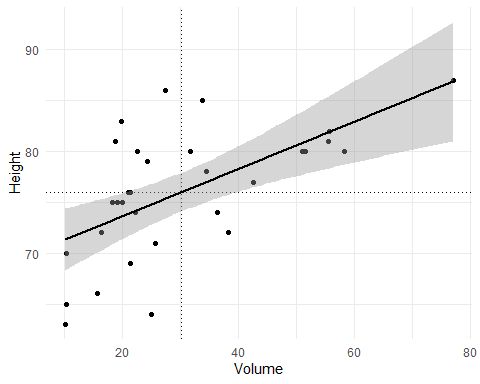
model21\_vg <- lm(Volume ~ Girth, data = trees)  
promedios21 <- colMeans(trees)  
ggplot(trees, aes(Volume, Girth)) +   
 geom\_point() +  
 geom\_vline(xintercept=promedios21[3],linetype="dotted") +   
 geom\_hline(yintercept=promedios21[1],linetype="dotted") +  
 geom\_smooth(method = "lm", se = TRUE, color = "black") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'

 **Volume ~ Height** **Volume ~ Girth**

model21\_vh <- lm(Volume ~ Height, data = trees)  
promedios21 <- colMeans(trees)  
ggplot(trees, aes(Volume, Height)) +   
 geom\_point() +  
 geom\_vline(xintercept=promedios21[3],linetype="dotted") +   
 geom\_hline(yintercept=promedios21[2],linetype="dotted") +  
 geom\_smooth(method = "lm", se = TRUE, color = "black") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'

 ### c) Realice un análisis diagnóstico en cada caso y señale en caso de haber los puntos influyentes y outliers.

#### Volume ~ Girth\*\*

**Modelo lineal**

summary(model21\_vg)

##   
## Call:  
## lm(formula = Volume ~ Girth, data = trees)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.065 -3.107 0.152 3.495 9.587   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -36.9435 3.3651 -10.98 7.62e-12 \*\*\*  
## Girth 5.0659 0.2474 20.48 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.252 on 29 degrees of freedom  
## Multiple R-squared: 0.9353, Adjusted R-squared: 0.9331   
## F-statistic: 419.4 on 1 and 29 DF, p-value: < 2.2e-16

Explica el 93% de los datos y los coeficientes tinene p-valores del test de wald por debajo de 0.05

**Normalidad**

shapiro.test(model21\_vg$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: model21\_vg$residuals  
## W = 0.97889, p-value = 0.7811

No rechaza normalidad

**Independencia**

dwtest(model21\_vg,alternative ="two.sided",iterations=1000)

##   
## Durbin-Watson test  
##   
## data: model21\_vg  
## DW = 1.4372, p-value = 0.06941  
## alternative hypothesis: true autocorrelation is not 0

La hipótesis nula es que NO hay autocorrelación, en este caso no rechaza pero por muy poco…..

**Homocedasticidad**

bptest(model21\_vg)

##   
## studentized Breusch-Pagan test  
##   
## data: model21\_vg  
## BP = 5.6197, df = 1, p-value = 0.01776

Los residuos son heterocedasticos porque se rechaza la hipotesis nula de homocedasticidad. Es decir que la varianza NO es constante para los diversos valores estimados.

**Resumen:**

* Normalidad: Sí
* Independencia: Sí
* Homcedasticidad: NO

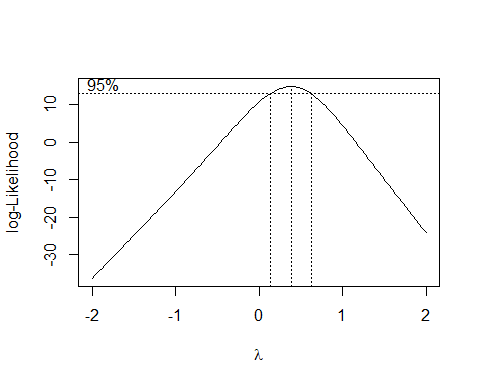
No se cumplen los supuestos entonce primero aplico una transformación de Box y Cox para ver si se resuelve el problema

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

box\_cox\_result <-boxcox(Volume ~ Girth, data = trees)



best\_box\_cox <- box\_cox\_result$x[which.max(box\_cox\_result$y)]  
model21\_vg\_bc <- lm((Volume)^(best\_box\_cox) ~ Girth, data = trees)

summary(model21\_vg\_bc)

##   
## Call:  
## lm(formula = (Volume)^(best\_box\_cox) ~ Girth, data = trees)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2733 -0.1277 -0.0056 0.1088 0.3595   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.559614 0.122535 4.567 8.44e-05 \*\*\*  
## Girth 0.228160 0.009008 25.329 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1548 on 29 degrees of freedom  
## Multiple R-squared: 0.9568, Adjusted R-squared: 0.9553   
## F-statistic: 641.6 on 1 and 29 DF, p-value: < 2.2e-16

Nuevamente los coeficientes tienen p-valores bajos para los test de wald de cada coeficiente dan ok y ahora explica el 95% de los datos.

**Normalidad**

shapiro.test(model21\_vg\_bc$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: model21\_vg\_bc$residuals  
## W = 0.98089, p-value = 0.837

No rechaza normalidad

**Independencia**

dwtest(model21\_vg\_bc,alternative ="two.sided",iterations=1000)

##   
## Durbin-Watson test  
##   
## data: model21\_vg\_bc  
## DW = 1.9353, p-value = 0.7071  
## alternative hypothesis: true autocorrelation is not 0

La hipótesis nula es que NO hay autocorrelación, no la rechaza y por un valor más lejano

**Homocedasticidad**

bptest(model21\_vg\_bc)

##   
## studentized Breusch-Pagan test  
##   
## data: model21\_vg\_bc  
## BP = 0.00074587, df = 1, p-value = 0.9782

No se rechaza la homocedasticidad

**Resumen:**

* Normalidad: Sí
* Independencia: Sí
* Homcedasticidad: Sí

#### Volume ~ Height

**Modelo lineal**

summary(model21\_vh)

##   
## Call:  
## lm(formula = Volume ~ Height, data = trees)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.274 -9.894 -2.894 12.068 29.852   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -87.1236 29.2731 -2.976 0.005835 \*\*   
## Height 1.5433 0.3839 4.021 0.000378 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.4 on 29 degrees of freedom  
## Multiple R-squared: 0.3579, Adjusted R-squared: 0.3358   
## F-statistic: 16.16 on 1 and 29 DF, p-value: 0.0003784

Explica el 33% de los datos y los coeficientes tienen p-valores del test de wald por debajo de 0.05

**Normalidad**

shapiro.test(model21\_vh$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: model21\_vh$residuals  
## W = 0.95078, p-value = 0.164

No rechaza normalidad

**Independencia**

dwtest(model21\_vh,alternative ="two.sided",iterations=1000)

##   
## Durbin-Watson test  
##   
## data: model21\_vh  
## DW = 0.5009, p-value = 1.791e-07  
## alternative hypothesis: true autocorrelation is not 0

La hipótesis nula es que NO hay autocorrelación, en este caso si rechaza independecia o más bien NO autocorrelación de los residuos

**Homocedasticidad**

bptest(model21\_vh)

##   
## studentized Breusch-Pagan test  
##   
## data: model21\_vh  
## BP = 12.207, df = 1, p-value = 0.0004762

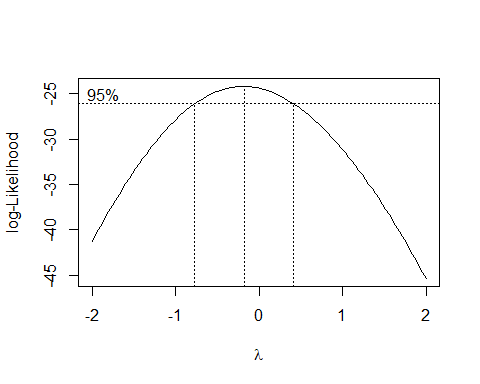
Los residuos son heterocedasticos porque se rechaza la hipotesis nula de homocedasticidad. Es decir que la varianza NO es constante para los diversos valores estimados.

**Resumen:**

* Normalidad: Sí
* Independencia: NO
* Homcedasticidad: NO

No se cumplen los supuestos entonces primero aplico una transformación de Box y Cox para ver si se resuelve el problema

library(MASS)  
box\_cox\_result <-boxcox(Volume ~ Height, data = trees)



best\_box\_cox <- box\_cox\_result$x[which.max(box\_cox\_result$y)]  
model21\_vh\_bc <- lm((Volume)^(best\_box\_cox) ~ Height, data = trees)

summary(model21\_vh\_bc)

##   
## Call:  
## lm(formula = (Volume)^(best\_box\_cox) ~ Height, data = trees)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.06174 -0.04167 0.00429 0.02815 0.06786   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.966501 0.088731 10.892 9.17e-12 \*\*\*  
## Height -0.005428 0.001164 -4.665 6.43e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04061 on 29 degrees of freedom  
## Multiple R-squared: 0.4287, Adjusted R-squared: 0.409   
## F-statistic: 21.76 on 1 and 29 DF, p-value: 6.433e-05

Nuevamente los coeficientes tienen p-valores bajos para los test de wald de cada coeficiente dan ok y ahora explica el 95% de los datos.

**Normalidad**

shapiro.test(model21\_vh\_bc$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: model21\_vh\_bc$residuals  
## W = 0.92768, p-value = 0.03792

No rechaza normalidad

**Independencia**

dwtest(model21\_vh\_bc,alternative ="two.sided",iterations=1000)

##   
## Durbin-Watson test  
##   
## data: model21\_vh\_bc  
## DW = 0.51685, p-value = 2.833e-07  
## alternative hypothesis: true autocorrelation is not 0

La hipótesis nula es que NO hay autocorrelación, la rechaza, por lo que hay autocorrelación

**Homocedasticidad**

bptest(model21\_vh\_bc)

##   
## studentized Breusch-Pagan test  
##   
## data: model21\_vh\_bc  
## BP = 0.11809, df = 1, p-value = 0.7311

No se rechaza la homocedasticidad

**Resumen:**

* Normalidad: NO
* Independencia: NO
* Homcedasticidad: Sí

Tampoco sirve.