Modelación lineal

Cristian Gañan, Valentina Ruiz, Maria Isabel Vasquez, Tatiana Marin, Daniel Marin, Marlon Tejada

# Punto 1

## 1.1

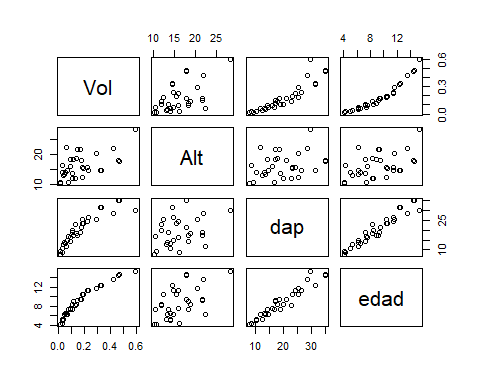
library(tidyverse)  
datos<- read.table("Bosque\_deg.txt")  
datos<- datos %>% mutate(V1= as.numeric(V1), V2= as.numeric(V2),  
 V3= as.factor(V3),V4= as.numeric(V4),  
 V5= as.numeric(V5)) %>%   
 rename(dap= V1, Alt= V2, ps= V3, edad= V4, Vol= V5)  
datos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dap | Alt | ps | edad | Vol |
| 18.72 | 21.50 | 1 | 9.33 | 0.173 |
| 10.61 | 14.00 | 2 | 5.08 | 0.030 |
| 11.82 | 22.40 | 1 | 6.25 | 0.064 |
| 13.38 | 14.04 | 2 | 6.46 | 0.056 |
| 7.88 | 10.27 | 1 | 4.05 | 0.013 |
| 16.13 | 15.85 | 1 | 7.58 | 0.092 |
| 18.98 | 13.53 | 2 | 7.30 | 0.106 |
| 26.42 | 14.68 | 2 | 11.37 | 0.235 |
| 35.07 | 17.78 | 1 | 14.54 | 0.474 |
| 21.23 | 17.84 | 1 | 9.41 | 0.175 |
| 24.42 | 15.79 | 1 | 11.37 | 0.224 |
| 28.59 | 21.86 | 1 | 13.55 | 0.425 |
| 9.47 | 10.59 | 1 | 4.19 | 0.019 |
| 25.33 | 12.44 | 2 | 10.35 | 0.183 |
| 23.34 | 15.44 | 1 | 10.52 | 0.193 |
| 16.99 | 10.57 | 1 | 7.38 | 0.075 |
| 12.64 | 13.10 | 2 | 5.24 | 0.043 |
| 8.65 | 16.24 | 2 | 4.26 | 0.028 |
| 14.47 | 18.33 | 1 | 7.28 | 0.091 |
| 14.82 | 14.76 | 2 | 6.26 | 0.069 |
| 23.09 | 11.94 | 2 | 8.22 | 0.134 |
| 25.44 | 20.34 | 1 | 11.62 | 0.291 |
| 31.51 | 14.47 | 2 | 12.41 | 0.333 |
| 18.11 | 18.77 | 1 | 8.32 | 0.137 |
| 20.11 | 11.89 | 2 | 8.38 | 0.106 |
| 13.82 | 13.18 | 2 | 6.25 | 0.051 |
| 17.17 | 18.31 | 1 | 8.90 | 0.114 |
| 29.89 | 28.24 | 1 | 15.35 | 0.599 |
| 17.41 | 21.55 | 1 | 9.23 | 0.150 |
| 31.51 | 14.50 | 2 | 12.27 | 0.326 |
| 23.34 | 15.48 | 1 | 10.41 | 0.189 |
| 10.61 | 14.02 | 2 | 5.03 | 0.029 |
| 35.07 | 17.81 | 1 | 14.39 | 0.464 |

***De este grupo de variables escoja con argumentos dasometricos la que debiera ser la variable dependiente. Use la función pairs del R y haga una descripción del comportamiento de las variables y elija las que, a su juicio, serían buenas candidatos a modelar con regresión lineal simple. Justifique sus respuestas.***

En este conjunto de datos, se puede ver en la **fig.1**, pares de comparaciones para las variables, hay que anotar dos cosas antes del análisis de este; en la columna se encontraban datos mal escritos desde la base, estos eran la forma más lógica de corregirlos fue correr el punto de la para transformarlos en respectivamente; a pesar de esto al graficar había un dato que no se ajustaba o mostraba un valor no concordante con el resto, por esto y sabiendo que al correr la coma los demás valores de esta fila serían inciertos pues no es del todo claro que estos fueran errores de entrada, se procedió a borrar la fila del valor ; se dejó en el modelo pues seguía la tendencia del resto. Teniendo estas claridades se realizó el análisis de la **fig.1**:

pairs(Vol ~ Alt\*dap\*edad, data = datos)



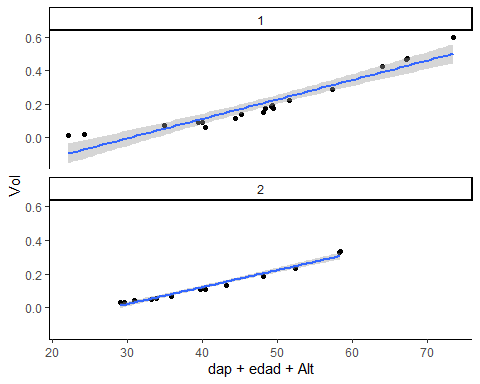
**fig.1** Matriz de relación par

\*: Aquí, hay relación sin embargo no muestra un tendencia clara para un modelo pues están un poco dispersos los datos.

\*: Si se analiza se puede observar una tendencia más ordenada a la anterior serie, pareciéndose a un modelo lineal, ya si se analiza se observa el mismo patrón presentado en el anterior par de datos , una dispersión considerable.

\*: Nuevamente los pares tienen comportamiento ordenado y reflejando la posibilidad de un modelo lineal; no muestran un patrón de comportamiento claro siendo difícil ajustar un modelo; muestran una relación evidente lineal, un patrón casi como una línea recta.

ggplot(datos, mapping = aes(x = dap+edad+Alt, y = Vol)) +  
 geom\_point() + geom\_smooth(method = "lm") + facet\_wrap(~ps, nrow= 2) +  
 theme\_classic()

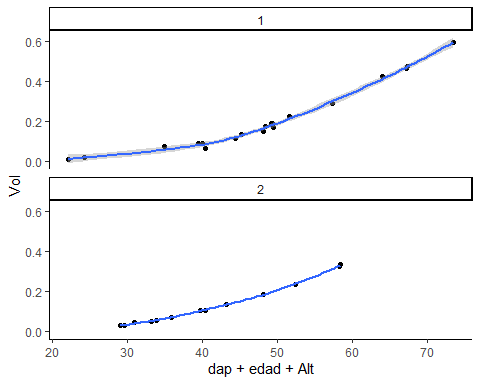


**fig.2** linea de tendencia lineal

Según el análisis anterior todas las variables tienen una relación mejor con , se puede concluir entonces que una buena opción de variable independiente sería .

Para hacer claro lo dicho anteriormente, se procedió a graficar y se hicieron dos facetas por parcela, para visualizar mejor las tendencias; en la **fig.2** se muestran los datos, con una línea de modelo lineal, los datos parecen ajustarse bien a este; sin embargo al hacer el mismo gráfico con una línea de tendencia **fig.3** se observa un mejor ajuste de la línea cuadrático.

ggplot(datos, mapping = aes(x = dap+edad+Alt, y = Vol)) +  
 geom\_point() + geom\_smooth() + facet\_wrap(~ps, nrow= 2) +  
 theme\_classic()



**fig.3** linea de tendencia

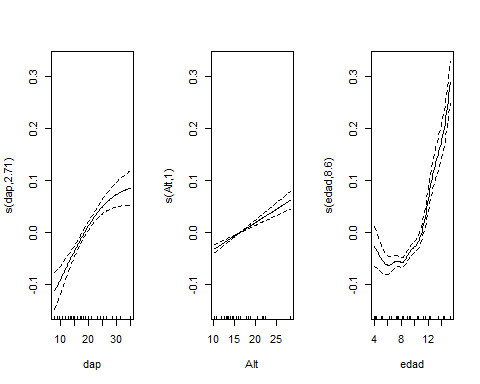
La variable es elegida como independiente por su dificultad para ser medida en campo lo que no pasa con que puede ser medida con facilidad, las otras variables también son complicadas de medir; sin embargo es claro que el volumen debería ir como independiente por su relaciòn con los demás datos, lo pertinente ahora sería mirar la posibilidad de omitir variables del modelo completo. Esto se analizará más adelante.

## 1.2

***Una buena idea para modelar es el uso de una función suavizadora smooth() en modelos lineales aditivos library(mgcv), para lo cual corra un modelo suavizado como:***

***modesuav<-gam(y~s(X1)+s(X2)+s(X3),family=”gaulss)***

library(mgcv)  
modesuav<-with(datos,gam(Vol~s(dap)+s(Alt)+s(edad)))  
  
par(mfrow=c(1,3))  
plot(modesuav)



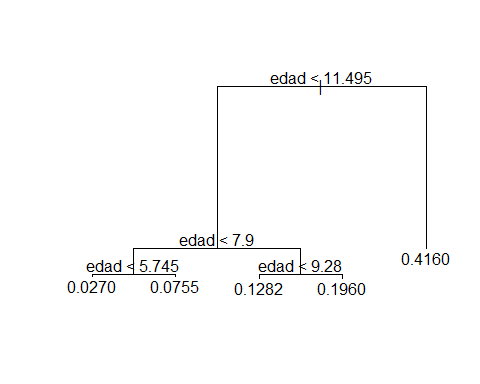
**fig.4** Modelos suavisados

En la **fig.4** se pueden ver los resultados de hacer modelos suavizados con las variables; el modelo en general resulta significativo, en el gráfico de es claro observar que a medida que el aumenta el también lo hace hasta llegar a un punto de estabilidad; los intervalos de confianza describen lo esperado para un modelo, a medida que los datos se alejan de la media la confianza es menor. Para el gráfico es posible encontrar anomalías pues es de esperarse que a más edad el volumen incremente o al menos no disminuya, sin embargo, al inicio del gráfico se encuentra esto, no muy pronunciado, los intervalos de confianza son anchos en este punto, esperándose entonces estimaciones con confianza baja como predicciones poco confiables para edades tempranas. Finalmente el gráfico de muestra una tendencia casi perfecta en línea recta e intervalos estrechos.

## 1.3

***Ejecute un dendromodelo (tree) para ver si aparecen interacciones complejas entre las variables, para lo cual debe bajar la librería tree, incorpórela al R.***

library(tree)  
modarb<-tree(Vol~edad,data=datos)  
plot(modarb)  
text(modarb)



**fig.5**Partición arborea

En la **fig.5** se muestra una partición arbórea para los datos, es evidente que la es la variable más importante en el modelo para determinar el , para los valores de tenderán a pero cuando importara si en este punto hay dos ramas para para lo cual los valores respectivos seran: .

### Modelo

modelo1<-lm(y~X1\*X2\*…\*Xk+I(X1^2)+I(X2^2)+…..I(Xk^2))

modelo1<-with(datos, lm(Vol~dap\*Alt\*edad+I(dap^2)+I(edad^2)+I(Alt^2)) )  
summary(modelo1)

##   
## Call:  
## lm(formula = Vol ~ dap \* Alt \* edad + I(dap^2) + I(edad^2) +   
## I(Alt^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.010146 -0.003810 -0.001023 0.003259 0.013295   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 7.058e-02 5.639e-02 1.252 0.224  
## dap 8.411e-03 8.822e-03 0.953 0.351  
## Alt -2.378e-03 5.823e-03 -0.408 0.687  
## edad -4.369e-02 2.789e-02 -1.566 0.132  
## I(dap^2) -3.376e-04 7.423e-04 -0.455 0.654  
## I(edad^2) -1.210e-03 5.228e-03 -0.231 0.819  
## I(Alt^2) -5.697e-05 2.334e-04 -0.244 0.809  
## dap:Alt -2.505e-04 8.493e-04 -0.295 0.771  
## dap:edad 2.172e-03 3.945e-03 0.551 0.587  
## Alt:edad 1.673e-03 2.276e-03 0.735 0.470  
## dap:Alt:edad 1.058e-05 1.986e-05 0.533 0.600  
##   
## Residual standard error: 0.006655 on 22 degrees of freedom  
## Multiple R-squared: 0.9986, Adjusted R-squared: 0.998   
## F-statistic: 1598 on 10 and 22 DF, p-value: < 2.2e-16

* **fig.1** Resumen del modelo original

Al probar el modelo inicial, se puede ver en la **fig.1** que el modelo no es significativo, partiendo de esto se hace entonces actualizaciones del modelo backward para escoger el mejor modelo

modbw<- step( modelo1, direction = "backward", trace=T)

## Start: AIC=-322.2  
## Vol ~ dap \* Alt \* edad + I(dap^2) + I(edad^2) + I(Alt^2)  
##   
## Df Sum of Sq RSS AIC  
## - I(edad^2) 1 2.3729e-06 0.00097678 -324.12  
## - I(Alt^2) 1 2.6379e-06 0.00097704 -324.11  
## - I(dap^2) 1 9.1604e-06 0.00098356 -323.89  
## - dap:Alt:edad 1 1.2563e-05 0.00098696 -323.77  
## <none> 0.00097440 -322.20  
##   
## Step: AIC=-324.12  
## Vol ~ dap + Alt + edad + I(dap^2) + I(Alt^2) + dap:Alt + dap:edad +   
## Alt:edad + dap:Alt:edad  
##   
## Df Sum of Sq RSS AIC  
## - I(Alt^2) 1 5.4700e-07 0.00097732 -326.10  
## - dap:Alt:edad 1 1.2483e-05 0.00098926 -325.70  
## - I(dap^2) 1 3.4369e-05 0.00101114 -324.97  
## <none> 0.00097678 -324.12  
##   
## Step: AIC=-326.1  
## Vol ~ dap + Alt + edad + I(dap^2) + dap:Alt + dap:edad + Alt:edad +   
## dap:Alt:edad  
##   
## Df Sum of Sq RSS AIC  
## - dap:Alt:edad 1 1.2499e-05 0.00098982 -327.68  
## - I(dap^2) 1 3.5250e-05 0.00101257 -326.93  
## <none> 0.00097732 -326.10  
##   
## Step: AIC=-327.68  
## Vol ~ dap + Alt + edad + I(dap^2) + dap:Alt + dap:edad + Alt:edad  
##   
## Df Sum of Sq RSS AIC  
## - dap:Alt 1 0.00000034 0.00099016 -329.67  
## <none> 0.00098982 -327.68  
## - I(dap^2) 1 0.00009381 0.00108363 -326.69  
## - dap:edad 1 0.00057811 0.00156793 -314.50  
## - Alt:edad 1 0.00070345 0.00169327 -311.96  
##   
## Step: AIC=-329.67  
## Vol ~ dap + Alt + edad + I(dap^2) + dap:edad + Alt:edad  
##   
## Df Sum of Sq RSS AIC  
## <none> 0.0009902 -329.67  
## - I(dap^2) 1 0.00009619 0.0010864 -328.61  
## - dap:edad 1 0.00063181 0.0016220 -315.38  
## - Alt:edad 1 0.00312890 0.0041191 -284.63

* **fig.2** Posibles modelos

Al terminar de correr el proceso backward se puede notar que el modelo quedó de la siguiente forma: Vol ~ dap + Alt + edad + I(dap^2) + dap:edad + Alt:edad. Se procedió a corroborar la pertinencia del modelo sin embargo como se puede ver en la **fig.3** no todas las variables del modelo son significativas, se considera la pertinencia de quitar variables del modelo.

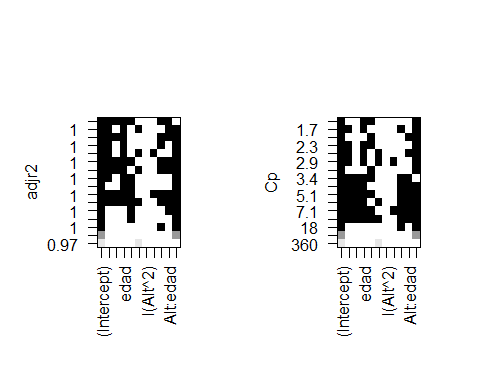
modelo2<-with(datos, lm(Vol ~ dap + Alt + edad + I(dap^2) + dap:edad + Alt:edad) )  
summary(modelo2)

##   
## Call:  
## lm(formula = Vol ~ dap + Alt + edad + I(dap^2) + dap:edad + Alt:edad)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0095830 -0.0038943 -0.0009748 0.0031644 0.0126015   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.1014926 0.0142912 7.102 1.53e-07 \*\*\*  
## dap 0.0068768 0.0028166 2.442 0.021738 \*   
## Alt -0.0048881 0.0014267 -3.426 0.002045 \*\*   
## edad -0.0467394 0.0069542 -6.721 3.94e-07 \*\*\*  
## I(dap^2) -0.0002292 0.0001442 -1.589 0.124081   
## dap:edad 0.0015994 0.0003927 4.073 0.000386 \*\*\*  
## Alt:edad 0.0014255 0.0001573 9.064 1.57e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.006171 on 26 degrees of freedom  
## Multiple R-squared: 0.9986, Adjusted R-squared: 0.9983   
## F-statistic: 3098 on 6 and 26 DF, p-value: < 2.2e-16

* **fig.3**  Modelo elegido 1

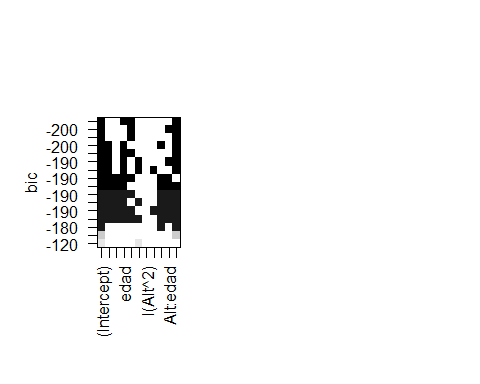
Se hizo entonces, gráficos para los criterios , y (**fig.4**) se puede ver para los gráficos de y se podrían omitir edad^2, Alt^2, dap:Alt, dap:Alt:edad y dap, Alt, dap^2, edad^2, Alt^2, dap:Alt, dap:edad, Alt:edad, dap:Alt:edad respectivamente; con el criterio de se podrían omitir dap, Alt, dap^2, edad^2, Alt^2, dap:Alt, dap:edad. Es claro entonces que para el modelo elegido (Vol ~ dap + alt + edad + I(dap^2) + dap:edad + alt:edad) se debería quitar ya sea dap:edad ò alt:edad o quizás las dos en y se deberían quitar, pero con ninguna debería quitarse del modelo. Como dos pruebas de las tres apoyan la eliminación de las variables dap:edad y alt:edad se procede a hacerlo de este modo.

library(leaps)  
  
s<- regsubsets(Vol~dap\*Alt\*edad+I(dap^2)+I(edad^2)+I(Alt^2),   
 data = datos, nbest = 2, method=c("exhaustive", "backward", "forward", "seqrep"))  
  
par(mfrow=c(1,2))  
plot(s, scale = "adjr2")  
plot(s, scale = "Cp")



**fig.4** criterios para la elecciòn de variables

plot(s)



**fig.4** criterios para la elecciòn de variables

Al quitar la variable alt:edad el modelo es significativo, lo que sería una muestra de la pertinencia de este. Lo que es preciso ahora es hacer pruebas para la viabilidad del modelo, esto se hará en el siguiente apartado.

modelo3<-with(datos, lm(Vol ~ dap + Alt + edad + I(dap^2) + dap:edad ) )  
summary(modelo3)

##   
## Call:  
## lm(formula = Vol ~ dap + Alt + edad + I(dap^2) + dap:edad)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0186431 -0.0101059 -0.0008209 0.0067953 0.0197423   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0107240 0.0204070 0.526 0.604   
## dap 0.0219356 0.0045522 4.819 4.96e-05 \*\*\*  
## Alt 0.0071463 0.0010451 6.838 2.40e-07 \*\*\*  
## edad -0.0817110 0.0115800 -7.056 1.38e-07 \*\*\*  
## I(dap^2) -0.0012892 0.0001690 -7.628 3.32e-08 \*\*\*  
## dap:edad 0.0046147 0.0004177 11.049 1.60e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01235 on 27 degrees of freedom  
## Multiple R-squared: 0.9942, Adjusted R-squared: 0.9931   
## F-statistic: 923.8 on 5 and 27 DF, p-value: < 2.2e-16

## 1.4

**¿Hay heterocedasticidad?**

shapiro.test(rstudent(modelo3))

##   
## Shapiro-Wilk normality test  
##   
## data: rstudent(modelo3)  
## W = 0.96654, p-value = 0.391

* **fig.1** Normalidad de los residuales

library(lmtest)  
bptest(modelo3)

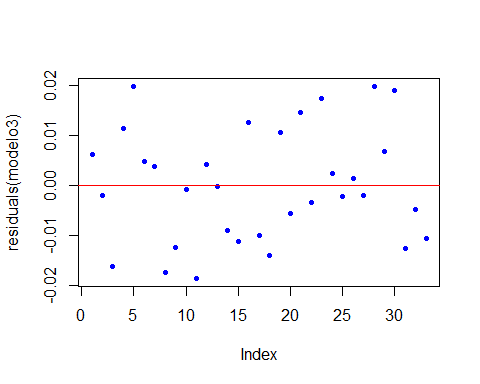
##   
## studentized Breusch-Pagan test  
##   
## data: modelo3  
## BP = 6.5594, df = 5, p-value = 0.2555

* **fig.2** Prueba de Homogeneidad de variazas

Al hacer pruebas de shapiro.test() para normalidad y bptest() para homogeneidad, se puede concluir de estas que el los residuales del modelo son normales, ademas, la variaza es constante a lo largo de la muestra.

En la **fig.3** se observa el grafico de residuales.

plot(residuals(modelo3), pch= 20, col= "blue")  
abline(0,0, col= "red")



**fig.3** Residuales

Es claro en la **fig.3** que no hay puntos palaca u observaciones influenciales, sin embargo, para corroborar esto se hace la prueba influence.measures, el resultado de este test arrojo que las observaciones son posibles puntos de influencia. Se procede a correr el modelo sin estas observaciones, sin embargo, al hacer una análisis de varianza resulta la variable èdad no significativa (**fig.5**), lo que no pasa con el modelo anterior con todas la observaciones (fig.6), por lo que se puede concluir que no habría que borrar puntos, pues el modelo pierde información que puede ser importante.

influence.measures(modelo3)

Modeloult3<- with(datos,lm(Vol ~ dap + Alt + edad + I(dap^2) + dap:edad, subset=(1:length(Vol)!= c(5,28))))  
anova(Modeloult3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| dap | 1 | 0.4468464 | 0.4468464 | 4372.476700 | 0.0000000 |
| Alt | 1 | 0.0237051 | 0.0237051 | 231.958436 | 0.0000000 |
| edad | 1 | 0.0001704 | 0.0001704 | 1.667557 | 0.2083989 |
| I(dap^2) | 1 | 0.0194750 | 0.0194750 | 190.566870 | 0.0000000 |
| dap:edad | 1 | 0.0063604 | 0.0063604 | 62.238197 | 0.0000000 |
| Residuals | 25 | 0.0025549 | 0.0001022 | NA | NA |

* **fig.5** Anova modelo sin observaciones 5 y 28

anova(modelo3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| dap | 1 | 0.5857209 | 0.5857209 | 3839.33456 | 0.0e+00 |
| Alt | 1 | 0.0640174 | 0.0640174 | 419.62679 | 0.0e+00 |
| edad | 1 | 0.0046891 | 0.0046891 | 30.73661 | 7.1e-06 |
| I(dap^2) | 1 | 0.0316099 | 0.0316099 | 207.19919 | 0.0e+00 |
| dap:edad | 1 | 0.0186238 | 0.0186238 | 122.07679 | 0.0e+00 |
| Residuals | 27 | 0.0041191 | 0.0001526 | NA | NA |

* **fig.6** Anova con todas las observaciones

# Punto 2

## 2.1

***Calcule grado de asociación entre las variables alt y dap, alt\*dap2 y vol, alt y vol, vol y la variable combinada . Y califíquelos por lo menos con dos pruebas, y además los intervalos de estimación para los coeficientes de correlación obtenidos.***

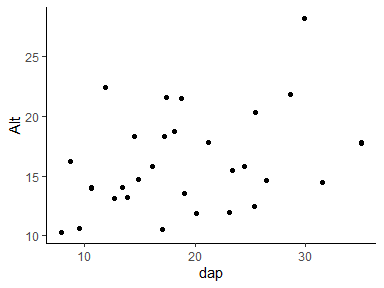
A partir los datos se procedió a calcular el grado de asociación “covarianza” definida como entre alt vs dap, dap^2\*alt vs vol,alt vs vol y vol vs dap , ver **table.1**,estas relaciones se decidieron mostrar en las **fig.1,2,3, 4**\* respectivamente. Cabe destacar que la covarianza tiene algunos problemas para su respectivo análisis dado que esta no está acotada en un número específico, pudiendo arrojar números positivos o negativos muy elevados dificultando la comprensión de cada una de estas, además las unidades en algunas veces son inconsistentes en la realidad física. En estas cuatro covarianzas lo único que podemos decir es que a medida que una variable del eje “x” está aumentando en una unidad, su homóloga en el eje “Y” está creciendo: y unidades respectivamente.

cov(datos\_okyalt) cov(datos\_okydap2\_alt) cov(datos\_okyalt) cov(datos\_okyVol

summarise(.data = datos,   
 cov(dap,Alt), cov(Vol,Alt), cov(dap,Vol),cov(Vol,(dap^2\*Alt)))

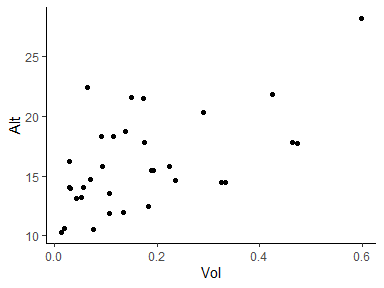
|  |  |  |  |
| --- | --- | --- | --- |
| cov(dap, Alt) | cov(Vol, Alt) | cov(dap, Vol) | cov(Vol, (dap^2 \* Alt)) |
| 9.755694 | 0.3408108 | 1.045475 | 967.2185 |

ggplot(datos, mapping = aes(x = dap, y = Alt))+   
 geom\_point()+ theme\_classic()



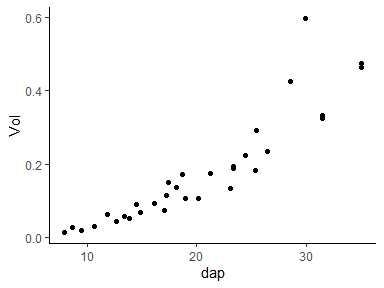
**fig.1**

ggplot(datos, mapping = aes(x = Vol, y = Alt))+   
 geom\_point()+ theme\_classic()



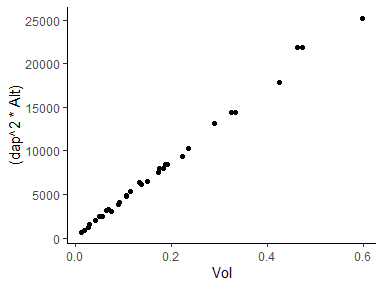
**fig.2**

ggplot(datos, mapping = aes(x = dap, y = Vol))+   
 geom\_point()+ theme\_classic()



**fig.3**

ggplot(datos, mapping = aes(x = Vol, y = (dap^2\*Alt)))+   
 geom\_point()+ theme\_classic()



**fig.4**

En **fig.1** y **fig.3** el grado de asociación es bajo pues los datos presentan un patrón difícil de describir, por lo que su tendría que ser bajo; de **fig.2** y **fig.4** se puede concluir que hay una asociación fuerte siendo más en $alt vs dap^2$ que en $dap vs vol$, esto es cierto, si se tiene en cuenta los patrones presentados mientras que en el primero es evidente una línea recta, en el segundo el patrón ya no es de línea recta si más bien una curva; entonces el sería más alto en $alt vs dap^2$ que en $dap vs vol$.

## 2.2

***Usando la prueba de cor.test(x, y, alternative = c("two.sided", "less", "greater"),method = c("pearson", "kendall", "spearman"),exact = NULL, conf.level = 0.95, continuity = FALSE, ...) que conclusiones saca de sus resultados anteriores***

with(datos, cor.test(Alt,dap, alternative = c("two.sided", "less", "greater"),method = c("kendall, pearson, spearman"),exact = NULL, conf.level = 0.95, continuity = FALSE))  
  
with(datos, cor.test(Alt,Vol, alternative = c("two.sided", "less", "greater"),method = c("kendall, pearson, spearman"),exact = NULL, conf.level = 0.95, continuity = FALSE))  
  
with(datos, cor.test(Vol,dap, alternative = c("two.sided", "less", "greater"),method = c("kendall, pearson, spearman"),exact = NULL, conf.level = 0.95, continuity = FALSE))  
  
with(datos, cor.test(Vol,(dap^2\*Alt), alternative = c("two.sided", "less", "greater"),method = c("kendall, pearson, spearman"),exact = NULL, conf.level = 0.95, continuity = FALSE))

Como la covarianza es difícil de interpretar dado sus múltiples problemas, en la práctica, lo que se hace es trabajar con una corrección de esta, llamada “coeficiente de correlación” el cual tiene varios métodos entre estos: Pearson, Kendall, Spearman. este coeficiente es adimensional, con un rango entre y , mostrando que si es mayor a cero, la relación entre las dos variables es directa, si es menor a cero la relación sería inversa, y si es igual o muy cercana a cero podría ocurrir que las variables son independientes o que no existe relación lineal entre ellas.

Luego de esto se parte de dos supuestos, *dice que las dos variables en estudio son independientes*,mostrando no correlación entre ellas, *dice que las dos variables sí están correlacionadas*;se encontró unas relaciones en P-values que se muestran en la tabla, allí podemos evidenciar que las variables que están correlacionadas linealmente y de forma positiva en los tres métodos de correlación trabajados son la alt vs vol, el vol vs dap y vol vs alt\*dap^2, el alt vs dap no se relaciona linealmente, hay que tener en cuenta que el coeficiente de correlación de pearson es paramétrico mientras los otros dos no, pues se basan en rangos, por consiguiente al ser los datos normales es más confiable la prueba de pearson, muestra de ello es que los resultados coinciden con los analizados en el punto anterior gráficamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| ——— | —————- | —————- | —————- | ——————— |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| ——— | —————- | —————- | —————- | ——————— |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## 2.3

***Ejecute el modelo de regresión lineal v=b0 + b1\*dap + b2\*alt + b3\*dap\*alt+ b4\*dap^2\*alt y entregue el modelo minimal y, todos los juicios posibles con base en gráficas y pruebas teóricas de residuales vistas.***

El análisis de regresión consiste en encontrar un modelo que relacione los valores medidos de un conjunto de variables, en este caso, el dap, alt y sus combinaciones para poder estimar el vol de los árboles, se sabe que los valores medidos en el mundo real nunca se ajustan de forma perfecta a un modelo, debido en primer lugar a errores de medida, pero también a que cualquier modelo matemático es una simplificación del mundo real, y si tuviera en cuenta todos los factores que influyen en un conjunto de variables, sería inmanejable. Por tanto, no tiene sentido aspirar a encontrar un modelo que prediga exactamente los valores medidos, y se debe admitir que el modelo cometerá un cierto error. Un modelo útil encuentra una relación funcional sencilla en conjuntos de pocas variables. Se trata de explicar una variable que tiene importancia, en función de otro conjunto de variables mejor conocidas o más fáciles de medir.

mod\_vol<-with(datos ,lm(Vol~dap+Alt+(dap\*Alt)+I(dap^2\*Alt)))  
summary(mod\_vol)

##   
## Call:  
## lm(formula = Vol ~ dap + Alt + (dap \* Alt) + I(dap^2 \* Alt))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.020904 -0.004377 -0.001170 0.005899 0.017701   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.246e-02 1.929e-02 1.683 0.10349   
## dap -2.191e-03 8.915e-04 -2.457 0.02045 \*   
## Alt -4.173e-03 1.624e-03 -2.570 0.01578 \*   
## I(dap^2 \* Alt) 1.748e-05 1.779e-06 9.824 1.42e-10 \*\*\*  
## dap:Alt 3.535e-04 1.074e-04 3.292 0.00269 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.008875 on 28 degrees of freedom  
## Multiple R-squared: 0.9969, Adjusted R-squared: 0.9964   
## F-statistic: 2243 on 4 and 28 DF, p-value: < 2.2e-16

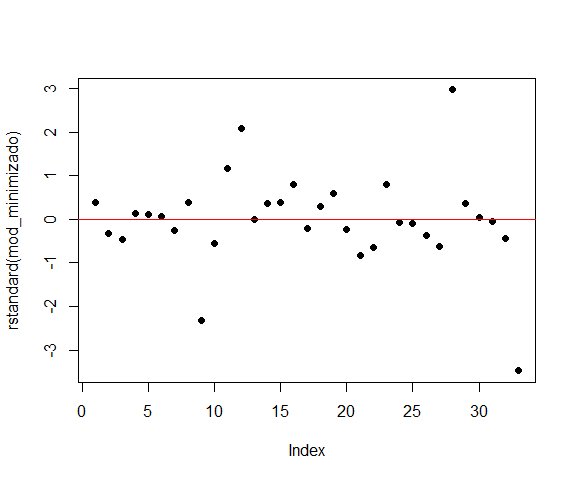
se encontró con este modelo propuesto que todas las variables independientes para calcular el volumen resultan significativas, sin embargo, la variable combinada alt\*dap^2 es la que más tiene un valor cercano a cero, indicando que esta variable combinada por sí sola mostraría valores estimados para el volumen.

mod\_minimizado <- with(datos , lm(Vol~I(dap^2\*Alt)))  
summary(mod\_minimizado)

##   
## Call:  
## lm(formula = Vol ~ I(dap^2 \* Alt))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.032654 -0.003769 -0.000041 0.003878 0.026553   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.606e-03 2.815e-03 -0.926 0.362   
## I(dap^2 \* Alt) 2.279e-05 2.813e-07 81.027 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01037 on 31 degrees of freedom  
## Multiple R-squared: 0.9953, Adjusted R-squared: 0.9951   
## F-statistic: 6565 on 1 and 31 DF, p-value: < 2.2e-16

* **Modelo minimizado**

plot(rstandard(mod\_minimizado), pch= 16)  
abline(0,0, col= "red")



**fig.1** Residuales estanadarizados del modelo minimizado

Se realiza un modelo con esta variable, arrojando que es significativo, sin embargo no es normal, ademas, es heterocedástico.

shapiro.test(rstandard(mod\_minimizado))

##   
## Shapiro-Wilk normality test  
##   
## data: rstandard(mod\_minimizado)  
## W = 0.8685, p-value = 0.0008938

bptest(mod\_minimizado)

##   
## studentized Breusch-Pagan test  
##   
## data: mod\_minimizado  
## BP = 20.555, df = 1, p-value = 5.794e-06

Para resolver este problema, se hace una transformación a todo el modelo.

mod\_tras <- with(datos , lm(log(Vol)~log(I(dap^2\*Alt))))  
summary(mod\_tras)

##   
## Call:  
## lm(formula = log(Vol) ~ log(I(dap^2 \* Alt)))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.130403 -0.029696 0.008462 0.028202 0.135070   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -11.05932 0.09279 -119.19 <2e-16 \*\*\*  
## log(I(dap^2 \* Alt)) 1.03873 0.01076 96.53 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05698 on 31 degrees of freedom  
## Multiple R-squared: 0.9967, Adjusted R-squared: 0.9966   
## F-statistic: 9318 on 1 and 31 DF, p-value: < 2.2e-16

Al hacer esto, las pruebas de breusch pagan y shapiro.wilk arrojan lo esperado, modelo homocedástico y normal.

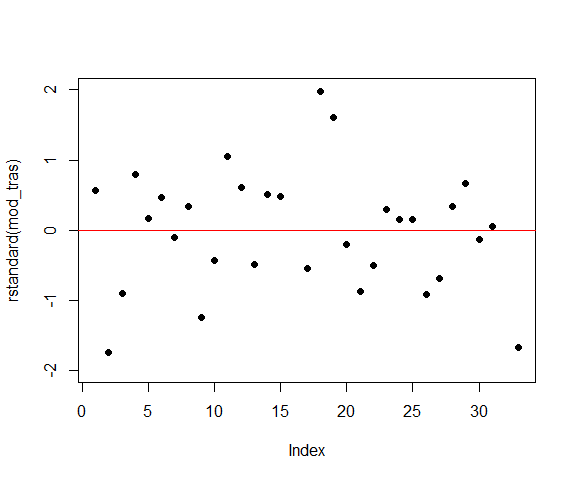
shapiro.test(rstandard(mod\_tras))

##   
## Shapiro-Wilk normality test  
##   
## data: rstandard(mod\_tras)  
## W = 0.9805, p-value = 0.8014

bptest(mod\_tras)

##   
## studentized Breusch-Pagan test  
##   
## data: mod\_tras  
## BP = 2.9738, df = 1, p-value = 0.08462

plot(rstandard(mod\_tras), pch= 16, ylim = c(-2,2))  
abline(0,0, col= "red")



**fig.2** Residuales estanadarizados del modelo minimizado transformado

## 2.4

***Proponga y ejecute algunas modelaciones para el factor mórfico:*** \* ***salidas de un volumen con variable combinada*** \* ***salidas de los datos***

datos<- datos %>% mutate(A.basal= (pi/40000)\*dap^2) %>%   
 mutate(F.forma= Vol/(A.basal\*Alt)) %>%   
 mutate(volumen.ff= round(0.2711151\*(A.basal\*Alt), 3))  
  
datos

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| dap | Alt | ps | edad | Vol | A.basal | F.forma | volumen.ff |
| 18.72 | 21.50 | 1 | 9.33 | 0.173 | 0.0275234 | 0.2923520 | 0.160 |
| 10.61 | 14.00 | 2 | 5.08 | 0.030 | 0.0088414 | 0.2423665 | 0.034 |
| 11.82 | 22.40 | 1 | 6.25 | 0.064 | 0.0109730 | 0.2603797 | 0.067 |
| 13.38 | 14.04 | 2 | 6.46 | 0.056 | 0.0140605 | 0.2836735 | 0.054 |
| 7.88 | 10.27 | 1 | 4.05 | 0.013 | 0.0048769 | 0.2595557 | 0.014 |
| 16.13 | 15.85 | 1 | 7.58 | 0.092 | 0.0204342 | 0.2840534 | 0.088 |
| 18.98 | 13.53 | 2 | 7.30 | 0.106 | 0.0282932 | 0.2769018 | 0.104 |
| 26.42 | 14.68 | 2 | 11.37 | 0.235 | 0.0548221 | 0.2920023 | 0.218 |
| 35.07 | 17.78 | 1 | 14.54 | 0.474 | 0.0965965 | 0.2759848 | 0.466 |
| 21.23 | 17.84 | 1 | 9.41 | 0.175 | 0.0353989 | 0.2771107 | 0.171 |
| 24.42 | 15.79 | 1 | 11.37 | 0.224 | 0.0468362 | 0.3028898 | 0.201 |
| 28.59 | 21.86 | 1 | 13.55 | 0.425 | 0.0641975 | 0.3028451 | 0.380 |
| 9.47 | 10.59 | 1 | 4.19 | 0.019 | 0.0070435 | 0.2547228 | 0.020 |
| 25.33 | 12.44 | 2 | 10.35 | 0.183 | 0.0503918 | 0.2919244 | 0.170 |
| 23.34 | 15.44 | 1 | 10.52 | 0.193 | 0.0427850 | 0.2921584 | 0.179 |
| 16.99 | 10.57 | 1 | 7.38 | 0.075 | 0.0226713 | 0.3129750 | 0.065 |
| 12.64 | 13.10 | 2 | 5.24 | 0.043 | 0.0125483 | 0.2615852 | 0.045 |
| 8.65 | 16.24 | 2 | 4.26 | 0.028 | 0.0058765 | 0.2933931 | 0.026 |
| 14.47 | 18.33 | 1 | 7.28 | 0.091 | 0.0164447 | 0.3018923 | 0.082 |
| 14.82 | 14.76 | 2 | 6.26 | 0.069 | 0.0172499 | 0.2710045 | 0.069 |
| 23.09 | 11.94 | 2 | 8.22 | 0.134 | 0.0418734 | 0.2680172 | 0.136 |
| 25.44 | 20.34 | 1 | 11.62 | 0.291 | 0.0508305 | 0.2814608 | 0.280 |
| 31.51 | 14.47 | 2 | 12.41 | 0.333 | 0.0779806 | 0.2951135 | 0.306 |
| 18.11 | 18.77 | 1 | 8.32 | 0.137 | 0.0257589 | 0.2833541 | 0.131 |
| 20.11 | 11.89 | 2 | 8.38 | 0.106 | 0.0317625 | 0.2806790 | 0.102 |
| 13.82 | 13.18 | 2 | 6.25 | 0.051 | 0.0150005 | 0.2579579 | 0.054 |
| 17.17 | 18.31 | 1 | 8.90 | 0.114 | 0.0231542 | 0.2688970 | 0.115 |
| 29.89 | 28.24 | 1 | 15.35 | 0.599 | 0.0701684 | 0.3022877 | 0.537 |
| 17.41 | 21.55 | 1 | 9.23 | 0.150 | 0.0238061 | 0.2923860 | 0.139 |
| 31.51 | 14.50 | 2 | 12.27 | 0.326 | 0.0779806 | 0.2883121 | 0.307 |
| 23.34 | 15.48 | 1 | 10.41 | 0.189 | 0.0427850 | 0.2853641 | 0.180 |
| 10.61 | 14.02 | 2 | 5.03 | 0.029 | 0.0088414 | 0.2339534 | 0.034 |
| 35.07 | 17.81 | 1 | 14.39 | 0.464 | 0.0965965 | 0.2697073 | 0.466 |

summarise(datos, promedio= mean(F.forma))

|  |
| --- |
| promedio |
| 0.279917 |

Al realizar el promedio del factor de forma de los 33 árboles nos da un valor de lo cual equivale al factor mórfico del bosque degradado. Al multiplicar el valor de factor de forma por el volumen de un cilindro de referencia se obtiene una estimación del volumen, al comparar estos datos con los del volumen dados se puede ver que es una buena estimación.

En ingeniería forestal tener el dato del factor mórfico es importante ya que con este se puede obtener aproximaciones del volumen de la madera que hay en un bosque, y es mucho más fácil ya que solo se necesita las mediciones del DAP y alturas, esto nos ahorra el trabajo de tener que que tumbar el árbol y cortarlo en trozas para hallar el volumen.

# Punto 3

***Los siguientes datos corresponden a un inventario de Pinus patula, realizado en 4 rodales diferentes en el departamento del Cauca en que: lote: código de lote, parc: número de la parcela en el mapa, numarb: número de árboles medidos/parc, narha: numero promedio de árboles/ha, dappr:DAP promedio/parc, dcpr: diámetro cuadrmedio/parc, altpro: altura total promedia/arbol/parc, Vtccha: volumen total con corteza/ha, Vtscha: volumen total sin corteza/ha, vasccha:volumen aserrío con corteza/ha, Vpccha: volumen de pulpa con corteza/ha, vasescha: volumen aserrío sin corteza/ha, vpuscha: volumen de pulpa sin corteza/ha, VCSCha: volumen comercial sin corteza/ha, porvcvt: porcentaje de volumen comercial sin corteza con respecto al volumen total.***

## 3.1

Ajuste un modelo para VCSCha, con base en dappr, dcpr, altpro, narha, para cada uno de los rodales.

rodal<- read.csv2("Inventario.csv", dec = ".")  
rodal

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| lote | parc | numarb | narha | dappr | dcpr | altpro | Vtccha | Vtscha | vasccha | Vpccha | vasescha | vpuscha | VCSCha | porvcvt |
| R1 | 169 | 42 | 1680 | 17.40 | 18.52 | 20.03 | 443.69 | 364.59 | 160.68 | 255.34 | 94.72 | 238.76 | 333.48 | 0.75 |
| R1 | 170 | 35 | 1400 | 17.30 | 18.14 | 18.78 | 337.95 | 277.92 | 98.42 | 219.16 | 49.05 | 205.97 | 255.03 | 0.75 |
| R1 | 157 | 39 | 1560 | 13.97 | 14.85 | 17.15 | 236.82 | 195.75 | 49.86 | 159.78 | 24.14 | 139.54 | 163.68 | 0.69 |
| R1 | 160 | 30 | 1200 | 13.79 | 14.16 | 18.34 | 165.92 | 137.66 | 5.11 | 137.53 | 0.00 | 110.74 | 110.74 | 0.67 |
| R1 | 158 | 22 | 880 | 16.86 | 17.51 | 18.56 | 199.82 | 164.45 | 47.72 | 139.14 | 27.69 | 122.32 | 150.01 | 0.75 |
| R1 | 165 | 32 | 1280 | 15.87 | 16.29 | 16.22 | 205.18 | 169.46 | 10.67 | 175.71 | 0.00 | 148.89 | 148.89 | 0.73 |
| R1 | 166 | 30 | 1200 | 13.76 | 14.92 | 16.30 | 177.73 | 146.82 | 23.94 | 132.56 | 11.55 | 113.74 | 125.29 | 0.70 |
| R1 | 163 | 30 | 1200 | 17.11 | 17.93 | 22.15 | 320.60 | 263.87 | 89.04 | 210.35 | 48.00 | 191.48 | 239.47 | 0.75 |
| R1 | 164 | 35 | 1400 | 17.08 | 17.83 | 18.37 | 310.16 | 255.31 | 66.74 | 223.26 | 33.61 | 199.23 | 232.84 | 0.75 |
| R1 | 168 | 36 | 1440 | 16.50 | 17.58 | 16.84 | 298.81 | 245.79 | 98.38 | 179.86 | 59.88 | 163.74 | 223.62 | 0.75 |
| R1 | 167 | 38 | 1520 | 16.31 | 17.04 | 11.76 | 186.76 | 154.04 | 18.09 | 151.27 | 0.00 | 135.88 | 135.88 | 0.73 |
| R3 | 48 | 35 | 1400 | 15.88 | 16.88 | 14.90 | 241.99 | 199.21 | 66.47 | 156.75 | 32.40 | 145.93 | 178.33 | 0.74 |
| R3 | 49 | 27 | 1080 | 15.65 | 16.02 | 15.18 | 157.72 | 130.33 | 6.70 | 136.20 | 5.44 | 107.73 | 113.17 | 0.72 |
| R3 | 45 | 17 | 680 | 20.96 | 21.45 | 21.48 | 240.55 | 197.26 | 88.41 | 143.57 | 50.92 | 135.90 | 186.82 | 0.78 |
| R3 | 43 | 20 | 800 | 19.17 | 19.47 | 19.46 | 219.78 | 180.62 | 44.31 | 164.70 | 24.32 | 144.58 | 168.89 | 0.77 |
| R3 | 41 | 33 | 1320 | 13.05 | 14.77 | 14.15 | 183.68 | 151.67 | 24.15 | 140.98 | 0.00 | 132.40 | 132.40 | 0.72 |
| R3 | 47 | 31 | 1240 | 15.09 | 15.86 | 14.35 | 170.20 | 140.55 | 24.48 | 128.29 | 10.82 | 109.68 | 120.49 | 0.71 |
| R3 | 46 | 20 | 800 | 14.30 | 14.63 | 12.78 | 83.07 | 68.84 | 4.68 | 68.20 | 0.00 | 57.36 | 57.36 | 0.69 |
| R3 | 44 | 21 | 840 | 18.35 | 18.69 | 17.90 | 194.86 | 160.31 | 37.91 | 145.74 | 14.96 | 132.94 | 147.89 | 0.76 |
| R3 | 42 | 19 | 760 | 18.52 | 19.00 | 18.21 | 181.52 | 149.29 | 36.78 | 135.15 | 21.76 | 116.17 | 137.93 | 0.76 |
| R3 | 40 | 30 | 1200 | 10.03 | 11.47 | 14.91 | 89.17 | 74.36 | 12.43 | 54.92 | 10.11 | 41.07 | 51.18 | 0.57 |
| R3 | 39 | 27 | 1080 | 15.78 | 16.36 | 16.93 | 200.22 | 165.07 | 34.40 | 149.98 | 20.40 | 127.41 | 147.82 | 0.74 |
| R3 | 38 | 18 | 720 | 21.95 | 22.60 | 18.03 | 240.51 | 196.93 | 106.92 | 127.16 | 67.04 | 121.05 | 188.09 | 0.78 |
| R3 | 37 | 32 | 1280 | 16.40 | 17.25 | 16.12 | 232.25 | 191.33 | 46.44 | 167.71 | 25.48 | 144.55 | 170.03 | 0.73 |
| R3 | 36 | 19 | 760 | 19.16 | 19.38 | 18.27 | 189.45 | 155.77 | 37.93 | 142.36 | 10.77 | 134.38 | 145.15 | 0.77 |
| R3 | 35 | 17 | 680 | 18.68 | 19.40 | 16.76 | 179.79 | 147.46 | 72.55 | 99.36 | 42.77 | 95.87 | 138.64 | 0.77 |
| R3 | 105 | 12 | 480 | 25.22 | 25.63 | 24.38 | 279.77 | 228.49 | 172.66 | 102.43 | 107.92 | 115.10 | 223.02 | 0.80 |
| R3 | 100 | 28 | 1120 | 19.80 | 21.09 | 20.47 | 378.86 | 310.41 | 170.71 | 192.87 | 110.57 | 183.08 | 293.65 | 0.78 |
| R3 | 99 | 24 | 960 | 17.92 | 19.01 | 17.45 | 244.47 | 200.71 | 86.98 | 146.02 | 54.87 | 132.16 | 187.03 | 0.77 |
| R3 | 101 | 24 | 960 | 19.87 | 20.35 | 20.73 | 308.11 | 252.86 | 102.04 | 194.00 | 51.27 | 187.15 | 238.42 | 0.77 |
| R3 | 106 | 16 | 640 | 22.26 | 22.97 | 24.34 | 327.94 | 268.10 | 188.96 | 131.41 | 121.43 | 136.81 | 258.25 | 0.79 |
| R3 | 103 | 17 | 680 | 21.87 | 22.49 | 20.47 | 256.43 | 210.00 | 119.54 | 129.53 | 70.01 | 131.55 | 201.56 | 0.79 |
| R3 | 102 | 17 | 680 | 18.04 | 19.30 | 14.56 | 159.05 | 130.38 | 66.50 | 85.74 | 41.94 | 80.04 | 121.98 | 0.77 |
| R3 | 104 | 18 | 720 | 21.59 | 21.91 | 20.84 | 256.66 | 210.41 | 101.64 | 147.03 | 44.41 | 155.97 | 200.38 | 0.78 |
| R3 | 98 | 28 | 1120 | 20.78 | 21.93 | 19.84 | 395.70 | 323.99 | 196.20 | 186.47 | 130.02 | 176.95 | 306.97 | 0.78 |
| R3 | 107 | 22 | 880 | 21.88 | 22.41 | 19.70 | 316.78 | 259.44 | 139.56 | 167.61 | 71.60 | 177.69 | 249.28 | 0.79 |
| R4 | 56 | 29 | 1160 | 17.70 | 18.58 | 17.57 | 253.11 | 208.21 | 75.45 | 161.36 | 28.62 | 159.46 | 188.08 | 0.74 |
| R4 | 142 | 37 | 1480 | 18.15 | 18.68 | 20.77 | 389.87 | 320.79 | 82.99 | 284.65 | 36.29 | 259.07 | 295.36 | 0.76 |
| R4 | 143 | 29 | 1160 | 16.19 | 17.07 | 19.16 | 248.03 | 204.35 | 67.11 | 161.64 | 30.33 | 151.10 | 181.42 | 0.73 |
| R4 | 144 | 35 | 1400 | 18.62 | 19.27 | 20.19 | 390.70 | 321.04 | 117.24 | 253.52 | 56.04 | 241.47 | 297.51 | 0.76 |
| R4 | 136 | 35 | 1400 | 13.67 | 14.78 | 16.42 | 209.47 | 173.27 | 0.00 | 188.86 | 0.00 | 149.24 | 149.24 | 0.71 |
| R4 | 131 | 18 | 720 | 10.76 | 12.63 | 13.62 | 78.70 | 65.21 | 5.94 | 62.02 | 4.83 | 49.14 | 53.97 | 0.69 |
| R4 | 132 | 50 | 2000 | 14.47 | 15.58 | 15.86 | 322.75 | 266.40 | 43.93 | 249.82 | 20.15 | 214.98 | 235.13 | 0.73 |
| R4 | 127 | 45 | 1800 | 12.94 | 15.36 | 13.78 | 283.38 | 233.48 | 51.79 | 212.17 | 25.30 | 184.86 | 210.16 | 0.74 |
| R4 | 137 | 33 | 1320 | 15.34 | 16.02 | 18.95 | 250.68 | 206.93 | 40.21 | 187.81 | 14.47 | 167.38 | 181.84 | 0.73 |
| R4 | 138 | 44 | 1760 | 12.17 | 12.73 | 15.62 | 177.57 | 147.70 | 0.00 | 146.11 | 0.00 | 112.26 | 112.26 | 0.63 |
| R4 | 129 | 32 | 1280 | 14.62 | 15.33 | 15.30 | 180.04 | 148.78 | 28.50 | 132.72 | 11.05 | 115.20 | 126.24 | 0.70 |
| R4 | 51 | 14 | 560 | 20.70 | 21.38 | 19.96 | 188.01 | 154.10 | 75.58 | 106.77 | 44.77 | 100.33 | 145.10 | 0.77 |
| R4 | 55 | 27 | 1080 | 20.03 | 20.98 | 17.52 | 300.87 | 246.71 | 129.30 | 159.98 | 77.46 | 154.31 | 231.78 | 0.77 |
| R4 | 54 | 20 | 800 | 22.69 | 23.31 | 21.59 | 355.50 | 290.75 | 195.58 | 150.85 | 114.66 | 165.41 | 280.07 | 0.79 |
| R4 | 53 | 17 | 680 | 18.99 | 19.61 | 19.75 | 198.02 | 162.59 | 61.37 | 127.62 | 33.57 | 118.18 | 151.75 | 0.77 |
| R4 | 52 | 18 | 720 | 23.28 | 23.70 | 19.79 | 276.97 | 226.70 | 118.62 | 151.90 | 68.03 | 150.89 | 218.92 | 0.79 |
| R4 | 124 | 32 | 1280 | 14.65 | 15.92 | 14.75 | 195.01 | 160.91 | 29.75 | 149.04 | 12.10 | 130.50 | 142.60 | 0.73 |
| R4 | 125 | 28 | 1120 | 11.52 | 14.47 | 12.88 | 161.94 | 133.45 | 28.51 | 121.91 | 15.14 | 104.78 | 119.92 | 0.74 |
| R4 | 126 | 35 | 1400 | 16.85 | 18.91 | 16.77 | 378.38 | 310.36 | 137.34 | 226.88 | 76.68 | 217.38 | 294.05 | 0.78 |
| R4 | 130 | 34 | 1360 | 14.86 | 15.38 | 14.98 | 197.61 | 163.23 | 20.69 | 157.23 | 9.20 | 133.36 | 142.55 | 0.72 |
| R4 | 139 | 38 | 1520 | 15.58 | 16.25 | 17.09 | 270.37 | 223.04 | 40.06 | 206.86 | 16.59 | 179.72 | 196.31 | 0.73 |

library(tidyverse)  
library(modelr)  
  
rodales<- rodal %>% group\_by(lote) %>% nest()  
  
  
modelos<- function(data){  
 lm(VCSCha~dappr + dcpr + altpro + narha, data = data)  
}  
  
rodales<- rodales %>%   
 mutate(model= map(data, modelos))  
  
  
model\_df<- rodales %>%   
 mutate(model\_df= map(model, broom::tidy)) %>%   
 unnest(model\_df, .drop = TRUE)  
  
model\_df

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| lote | term | estimate | std.error | statistic | p.value |
| R1 | (Intercept) | -654.1512808 | 47.4371925 | -13.789840 | 0.0000090 |
| R1 | dappr | -29.2829884 | 15.9199748 | -1.839387 | 0.1154763 |
| R1 | dcpr | 52.8372206 | 15.6312020 | 3.380240 | 0.0148534 |
| R1 | altpro | 13.0297712 | 1.5322338 | 8.503775 | 0.0001448 |
| R1 | narha | 0.1470130 | 0.0190021 | 7.736677 | 0.0002449 |
| R3 | (Intercept) | -477.1584435 | 44.5882024 | -10.701451 | 0.0000000 |
| R3 | dappr | -13.9767505 | 11.3654201 | -1.229761 | 0.2330574 |
| R3 | dcpr | 32.5510719 | 11.3399873 | 2.870468 | 0.0094565 |
| R3 | altpro | 7.2814303 | 2.3801960 | 3.059173 | 0.0061916 |
| R3 | narha | 0.1672237 | 0.0228538 | 7.317095 | 0.0000004 |
| R4 | (Intercept) | -556.0815331 | 66.7340130 | -8.332805 | 0.0000003 |
| R4 | dappr | -38.7900985 | 10.4805916 | -3.701136 | 0.0019377 |
| R4 | dcpr | 54.5778792 | 10.0923511 | 5.407846 | 0.0000581 |
| R4 | altpro | 14.9838697 | 4.2282485 | 3.543753 | 0.0027019 |
| R4 | narha | 0.1371557 | 0.0156198 | 8.780896 | 0.0000002 |

model\_df<- rodales %>%   
 mutate(model\_df= map(model, broom::glance)) %>%   
 unnest(model\_df, .drop = TRUE)  
  
model\_df

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| lote | r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual |
| R1 | 0.9834447 | 0.9724078 | 11.45220 | 89.10537 | 1.79e-05 | 5 | -39.09458 | 90.18916 | 92.57653 | 786.9179 | 6 |
| R3 | 0.9307693 | 0.9169231 | 18.39903 | 67.22226 | 0.00e+00 | 5 | -105.49162 | 222.98323 | 230.29649 | 6770.4853 | 20 |
| R4 | 0.9092318 | 0.8865398 | 22.58440 | 40.06830 | 0.00e+00 | 5 | -92.40485 | 196.80970 | 203.07683 | 8160.8787 | 16 |

En el R1 hay una alta significancia en el número promedio de narha junto con la altpro, y una baja significancia en el diámetro cuadrado dcpr.

En el R3 se puede observar una alta significancia en el número promedio de , y una significancia media en la altpro y el dcpr.

En el R4 hay una alta significancia en el número promedio de narha, junto con el dcpr, y una significancia media en dappr junto con la altpro.

Al aplicar la prueba de Shapiro-Wilk en los tres rodales (R1, R3, R4) siguen una distribución aproximadamente normal, ya que el valor de dicha prueba en cada uno de los rodales ( y ) tienen valores muy cercanos a uno , también se puede ver que en los tres rodales hay una alta significancia en el número promedio de narha, y si observamos bien el ejercicio se puede afirmar que así debe de ser ya que los datos son tomados de un inventarios realizado de Pinus patula, en el Departamento del Cauca.

## 3.2

Compare los resultados de cada rodal con el modelo general con todos los datos para todo ese bosque.

general<- rodal %>% nest()  
  
  
modelos<- function(data){  
 lm(VCSCha~dappr + dcpr + altpro + narha, data = data)  
}  
  
general<- general %>%   
 mutate(model= map(data, modelos))  
  
model\_g<- general %>%   
 mutate(model\_g= map(model, broom::tidy)) %>%   
 unnest(model\_g, .drop = TRUE)  
  
model\_g

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value |
| (Intercept) | -494.757853 | 29.3691621 | -16.846169 | 0.0e+00 |
| dappr | -28.242439 | 5.7546499 | -4.907760 | 9.5e-06 |
| dcpr | 45.439162 | 5.8595326 | 7.754742 | 0.0e+00 |
| altpro | 10.299718 | 1.5219282 | 6.767545 | 0.0e+00 |
| narha | 0.143453 | 0.0096393 | 14.882083 | 0.0e+00 |

model\_g<- general %>%   
 mutate(model\_g= map(model, broom::glance)) %>%   
 unnest(model\_g, .drop = TRUE)  
  
model\_g

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual |
| 0.9139684 | 0.9073506 | 19.86901 | 138.1073 | 0 | 5 | -248.6452 | 509.2903 | 521.5486 | 20528.42 | 52 |

En el modelo general se puede observar que al aplicar el test de Shapiro-Wilk los datos siguen una distribución aproximadamente normal, dado el valor de p-evalue se puede decir que hay un no rechazo de la .

También se observa que hay una alta significancia en todos sus componentes (dappr, decpr, altpro, narha) esto se puede dar por la relación de los datos, ya que los datos utilizados para el modelo general no son tomados del mismo espacio, por lo tanto pueden existir valores muy diferentes entre rodales.

* **R1 vs modelo general(modg)**

En R1 se puede observar que los residuos del error estándar es de 11.45 mientras que en el modelo general es de . En el R1 solo las variables narha y decpr son de alta significancia, mientras que en el modelo general todas son de alta significancia. El P-evalue se encuentra en el modelo general es mucho menor al p-evalue encontrado en el R1 .

* **R3 vs modg**

En R3 los residuos del error estándar tienen un valor de , aquí se puede observar que los residuos del error estándar de tiene un valor mas cercano a los residuos del modg. En R3 solo se encuentra una alta significancia en narha, y en el modg todos los datos son de ata significancia. El p-evalue del modg sigue siendo menor al p-evalue del R3 .

* **R4 vs modg**

En R4 los residuos del error estándar es de siendo un valor mayor que los residuos del modg. Pero en R4 se tiene grados de libertad, y en el modg grados de libertad. En R4 también solo dos variables son de alta significancia narha y dcpr el p.evalue de R4 es de este valor es mayor que el p-evalue del modg.

## 3.3

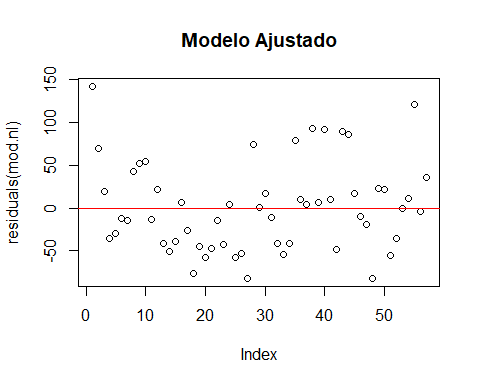
mod.nl <- nls(VCSCha~a\*dappr\*exp(-b\*altpro),   
 data = rodal, start = list(a=1, b=0))  
  
modg <- lm(VCSCha~dappr + dcpr + altpro + narha, data = rodal)  
summary(mod.nl)

##   
## Formula: VCSCha ~ a \* dappr \* exp(-b \* altpro)  
##   
## Parameters:  
## Estimate Std. Error t value Pr(>|t|)   
## a 7.09731 1.81474 3.911 0.000255 \*\*\*  
## b -0.02184 0.01342 -1.627 0.109387   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 52.87 on 55 degrees of freedom  
##   
## Number of iterations to convergence: 7   
## Achieved convergence tolerance: 4.352e-06

shapiro.test(residuals(mod.nl))

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(mod.nl)  
## W = 0.95005, p-value = 0.01978

windows()  
plot(residuals(mod.nl), main = "Modelo Ajustado")  
abline(0,0, col="red")



#Prueba de bondad de ajuste  
pred <-predict(mod.nl)  
chi.mod.nl <- sum((rodal$Vtscha-pred)^2/pred)  
chi.mod.nl

## [1] 1123.852

qchisq(0.975, 55)

## [1] 77.38047

pred1 <- predict(modg)  
chi.modg <- sum((rodal$Vtscha-pred1)^2/pred1)  
chi.modg

## [1] 532.2127

qchisq(0.975, 55)

## [1] 77.38047

#homocedasticidad  
AIC(modg, mod.nl)

|  |  |  |
| --- | --- | --- |
|  | df | AIC |
| modg | 6 | 509.2903 |
| mod.nl | 3 | 618.0581 |

Para realizar el modelo ajustado se hizo un modelo no lineal usando la función nls, donde tras varias iteraciones los valores iniciales convergieron a los valores de los parámetros y .

Para ver que tal se ajustan los modelos, entre los dos modelos el que mejor se ajusta es el modg, pero al comparar el valor del chi cuadrado de cada modelo con el valor tabular se puede ver que estos no son los modelos que mejor se ajustan a los datos. El valor de AIC para cada modelo también muestra que el mejor modelo es el modg.

Se puede observar que en ambos modelos no se presenta una simetría, en cambio se ve una asimetría en la dispersión, por lo tanto se podría decir que ningunos de los dos modelos es el más indicado para este proyecto.

En los graficos tambien se pueden observar la homocedasticidad de los datos, ya que estos no varían mucho respecto a la media.

Otra diferencias que se puede notar entre ambos modelos es que los residuales de nuestro modelo ajustado varían entre un poco menos de y , mientras que el modelo general varía entre aproximadamente y .

# Punto 4

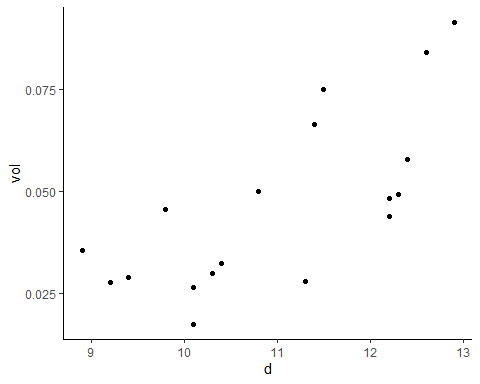
## 4.1

***Compare dasométricamente los modelos v=f(dap) y v=f(dap,alt)***

Inicialmente se debe hacer un análisis exploratorio con el fin de saber si la relación del contra y contra volumen era lineal o no lineal. En las **fig.1** y **fig.2** se puede notar la tendencia lineal que siguen los datos, además que a medida que aumenta el y la , aumenta el .

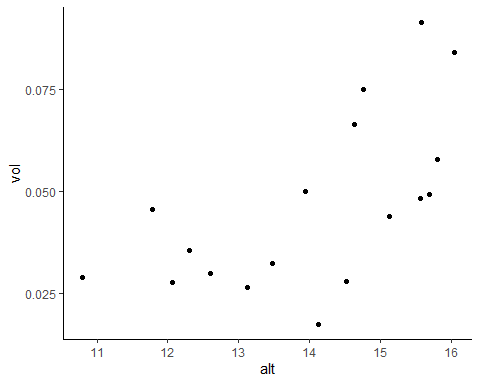
library(tidyverse)  
datos2<- read.table("volu.txt")  
  
datos2<- datos2 %>% rename(d= V1, vol= V2, alt= V3)

ggplot(datos2, mapping = aes(x = d, y = vol)) +  
 geom\_point() + theme\_classic()



**fig.1** volumen en funcion del diametro

ggplot(datos2, mapping = aes(x = alt, y = vol)) +  
 geom\_point() + theme\_classic()



**fig.2** volumen en funcion de la alt

Mediante la regresión lineal se crearon dos modelos: m1=lm(vol~d) y m2=lm(vol~d^2\*alt)con el anova se pudo observar que ambos modelos eran significativos.

Para escoger un mejor modelo entre ellos dos se compara el AIC de cada una y el RSE, así como el sesgo e incertidumbre, que se halla con la fórmula:

El resumen de los modelos se presenta en la siguiente tabla:

mod1<-lm(vol~d,data=datos2)  
summary(mod1)  
anova(mod1)  
summary(mod1)  
AIC(mod1)  
shapiro.test(residuals(mod1))  
  
predm1=predict(mod1)  
error1=100\*(predm1-datos3$vol)/datos3$vol  
mean(error1)  
sd(error1)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

De los modelos planteados no se puede decir a ciencia cierta cuál es el mejor modelo, puesto que tanto el error estándar de los residuales con el AIC son muy similares y un criterio para escoger es la diferencia entre AIC mayores a 4 por lo que se podría decir que estos modelos son estadísticamente indiferenciables. Teniendo en cuanta eso, se evalúa el criterio del error, mostrando que el modelo 1 tiene un menor sesgo e incertidumbre que el modelo 2, por lo que se podría decir que el primer modelo es más adecuado.

## 4.2

***Haga una prueba teórica de bondad de ajuste al primer modelo***

La prueba de bondad de ajuste permite establecer si el modelo se ajusta bien a los datos, para este caso, se utilizó el criterio de la suma de cuadrados de los residuales es bien sabido que si esta da como resultado la regresión se ajusta completamente a los datos, haciendo una análisis se encuentra este valor, para este caso fue de no es cero pero si es un valor pequeño, por lo cual el modelo ajustado es bueno. Para comprobar esto se miró el ajustado, el cual fue de: $0.4908 $ es un número positivo por lo cual es una asociación directa pero no es tan cercana a , sin embargo, tampoco es cercana a lo que indicaría un ajuste moderado del modelo, apoyado en esto se puede decir que es un modelo aceptable.

## 4.3

***Proponga un buen modelo con base en la última propuesta de 4.1***

El modelo propuesto de vol en función de dap y alt fue el modelo exponencial . En la siguiente tabla se presentan los resultados de algunas pruebas de analisis para el modelo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

Como este es un modelo exponencial, que inicialmente se planteó como logarítmico, no se puede comparar con los dos modelos anteriores bajo el criterio del AIC, puesto que la variable dependiente está en distinta escala; sin embargo, se puede comparar el RSE y el error.