

# Análisis descriptivo de una base de datos

Gómez Jiménez Aaron Mauricio

2023-02-23

El archivo wine contiene trece variables numéricas derivadas de un análisis químico en vinos de Italia de tres viñedos diferentes, realizaremos un análisis estadístico de las muestras para encontrar relaciones o diferencias entre los viñedos.

Cargamos la base de datos que analizaremos

```
base = read_excel("wine.xlsx")
datos = data.frame(base)
head(datos)
```

```
##   Class Alcohol Malic_acid  Ash Alcalinity Magnesium Total_phenols Flavanoids
## 1     1   14.23     1.71 2.43     15.6      127         2.80      3.06
## 2     1   13.20     1.78 2.14     11.2      100         2.65      2.76
## 3     1   13.16     2.36 2.67     18.6      101         2.80      3.24
## 4     1   14.37     1.95 2.50     16.8      113         3.85      3.49
## 5     1   13.24     2.59 2.87     21.0      118         2.80      2.69
## 6     1   14.20     1.76 2.45     15.2      112         3.27      3.39
## Nonflavanoid_phenols Proanthocyanins Intensity Hue    OD Proline
## 1                   0.28              2.29    5.64 1.04 3.92   1065
## 2                   0.26              1.28    4.38 1.05 3.40   1050
## 3                   0.30              2.81    5.68 1.03 3.17   1185
## 4                   0.24              2.18    7.80 0.86 3.45   1480
## 5                   0.39              1.82    4.32 1.04 2.93    735
## 6                   0.34              1.97    6.75 1.05 2.85   1450
```

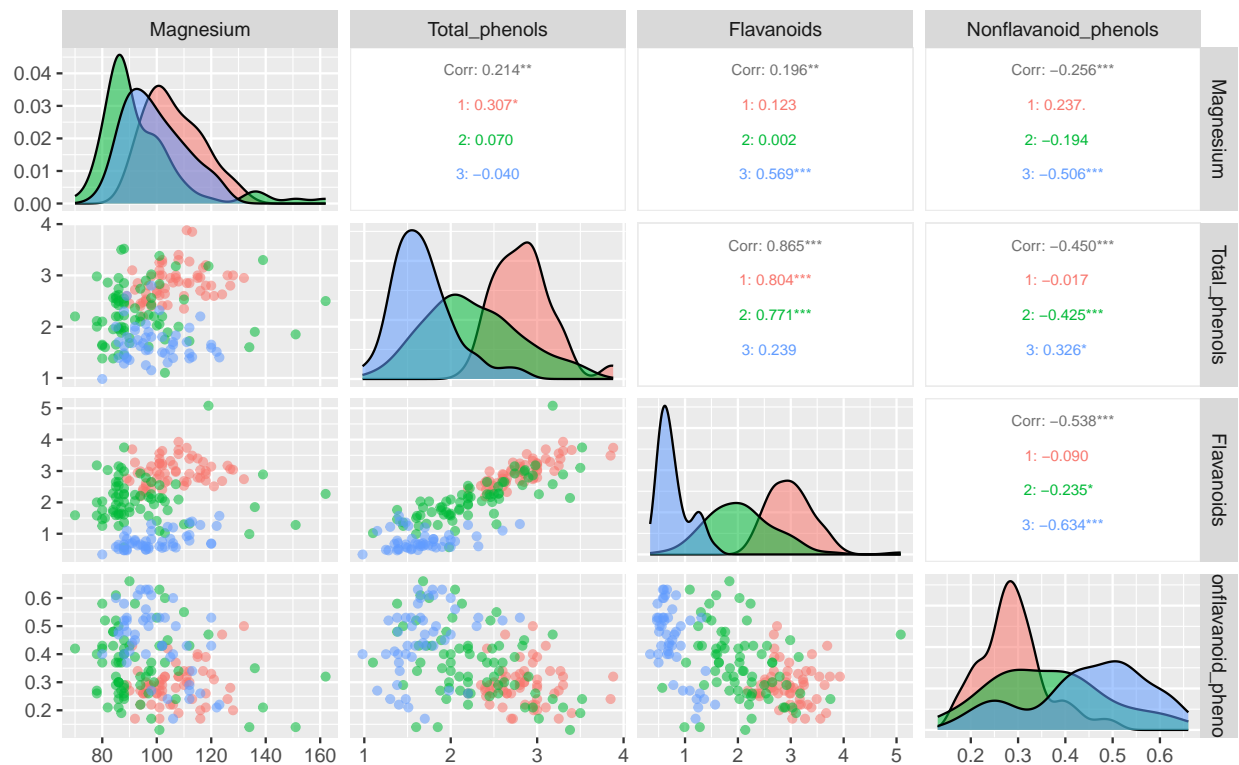
Gráficamos para identificar la distribución de los datos y analizar si existe relación entre las variables

En la siguiente gráfica podemos observar que los viñedos tienen diferencias en la variable alcohol, sin embargo, las diferencias no son significativas, en contraparte para la variable ácido malico el viñedo 1 tiene mediciones más altas en comparación a los otros viñedos, en las variables restantes se mantienen en el rango los datos obtenidos para los viñedos.

```
ggpairs(datos, columns = 2:5, aes(color = Class, alpha = 0.5),
        upper = list(continuous = wrap("cor", size = 2.5)))
```

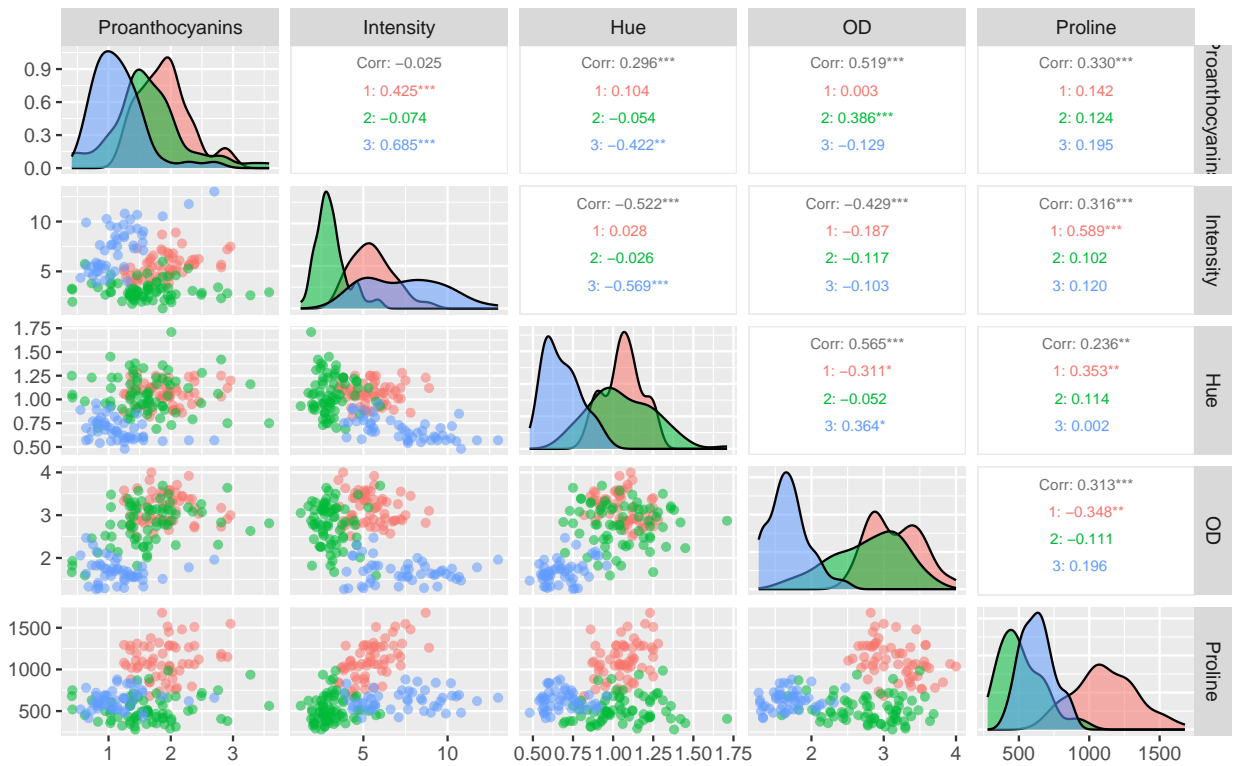


```
ggpairs(datos, columns = 6:9, aes(color = Class, alpha = 0.5),
  upper = list(continuous = wrap("cor", size = 2.5)))
```



En general en esta gráfica no hay diferencias significativas, solamente se distingue que para el total de fenoles y flavonoides las distribuciones de los datos tienen medias distintas.

```
ggpairs(datos, columns = 10:14, aes(color = Class, alpha = 0.5),
        upper = list(continuous = wrap("cor", size = 2.5)))
```



Finalmente para la tercer visualización no notamos diferencias importantes.

A continuación realizaremos un análisis numérico de los datos, para esto utilizaremos un resumen de estadísticas básicas.

```
describeBy(datos, group=datos$Class, fast=TRUE)
```

```
##
## Descriptive statistics by group
## group: 1
##
```

	vars	n	mean	sd	median	min	max	range
## Class	1	59	1.00	0.00	1.00	1.00	1.00	0.00
## Alcohol	2	59	13.74	0.46	13.75	12.85	14.83	1.98
## Malic_acid	3	59	2.01	0.69	1.77	1.35	4.04	2.69
## Ash	4	59	2.46	0.23	2.44	2.04	3.22	1.18
## Alcalinity	5	59	17.04	2.55	16.80	11.20	25.00	13.80
## Magnesium	6	59	106.34	10.50	104.00	89.00	132.00	43.00
## Total_phenols	7	59	2.84	0.34	2.80	2.20	3.88	1.68
## Flavanoids	8	59	2.98	0.40	2.98	2.19	3.93	1.74
## Nonflavanoid_phenols	9	59	0.29	0.07	0.29	0.17	0.50	0.33
## Proanthocyanins	10	59	1.90	0.41	1.87	1.25	2.96	1.71
## Intensity	11	59	5.53	1.24	5.40	3.52	8.90	5.38
## Hue	12	59	1.06	0.12	1.07	0.82	1.28	0.46

```

## OD          13 59      3.16    0.36      3.17    2.51      4.00    1.49
## Proline     14 59 1115.71 221.52 1095.00 680.00 1680.00 1000.00
##            skew kurtosis    se
## Class       NaN      NaN  0.00
## Alcohol     0.07     -0.68  0.06
## Malic_acid  2.03      2.84  0.09
## Ash         0.52      0.58  0.03
## Alcalinity  0.20      0.86  0.33
## Magnesium   0.47     -0.63  1.37
## Total_phenols 0.78      0.89  0.04
## Flavanoids  0.25     -0.65  0.05
## Nonflavanoid_phenols 0.69      0.62  0.01
## Proanthocyanins 0.60      0.02  0.05
## Intensity   0.56     -0.09  0.16
## Hue        -0.08     -0.81  0.02
## OD         0.30     -0.85  0.05
## Proline     0.22     -0.40 28.84
## -----
## group: 2
##            vars  n    mean      sd median    min    max    range    skew
## Class         1 71    2.00    0.00    2.00    2.00    2.00    0.00    NaN
## Alcohol       2 71   12.28    0.54   12.29   11.03   13.86    2.83    0.56
## Malic_acid    3 71    1.93    1.02    1.61    0.74    5.80    5.06    1.56
## Ash           4 71    2.24    0.32    2.24    1.36    3.23    1.87    0.17
## Alcalinity    5 71   20.24    3.35   20.00   10.60   30.00   19.40    0.41
## Magnesium     6 71   94.55   16.75   88.00   70.00  162.00   92.00    2.03
## Total_phenols  7 71    2.26    0.55    2.20    1.10    3.52    2.42    0.35
## Flavanoids    8 71    2.08    0.71    2.03    0.57    5.08    4.51    1.12
## Nonflavanoid_phenols 9 71    0.36    0.12    0.37    0.13    0.66    0.53    0.33
## Proanthocyanins 10 71    1.63    0.60    1.61    0.41    3.58    3.17    0.65
## Intensity     11 71    3.09    0.92    2.90    1.28    6.00    4.72    1.00
## Hue           12 71    1.06    0.20    1.04    0.69    1.71    1.02    0.47
## OD            13 71    2.79    0.50    2.83    1.59    3.69    2.10   -0.43
## Proline       14 71  519.51  157.21  495.00  278.00  985.00  707.00    0.85
##            kurtosis    se
## Class       NaN    0.00
## Alcohol     0.51    0.06
## Malic_acid  2.21    0.12
## Ash         0.67    0.04
## Alcalinity  0.94    0.40
## Magnesium   4.46    1.99
## Total_phenols -0.46    0.06
## Flavanoids  3.15    0.08
## Nonflavanoid_phenols -0.48    0.01
## Proanthocyanins 1.23    0.07
## Intensity   1.03    0.11
## Hue         0.09    0.02
## OD        -0.60    0.06
## Proline     0.32  18.66
## -----
## group: 3
##            vars  n    mean      sd median    min    max    range    skew
## Class         1 48    3.00    0.00    3.00    3.00    3.00    0.00    NaN
## Alcohol       2 48   13.15    0.53   13.16   12.20   14.34    2.14    0.14

```

```
## Malic_acid      3 48   3.33   1.09   3.26   1.24   5.65   4.41   0.10
## Ash            4 48   2.44   0.18   2.38   2.10   2.86   0.76   0.34
## Alcalinity     5 48  21.42   2.26  21.00  17.50  27.00   9.50   0.44
## Magnesium      6 48  99.31  10.89  97.00  80.00 123.00  43.00   0.51
## Total_phenols   7 48   1.68   0.36   1.64   0.98   2.80   1.82   0.96
## Flavanoids      8 48   0.78   0.29   0.69   0.34   1.57   1.23   0.95
## Nonflavanoid_phenols 9 48   0.45   0.12   0.47   0.17   0.63   0.46  -0.50
## Proanthocyanins 10 48   1.15   0.41   1.10   0.55   2.70   2.15   1.48
## Intensity      11 48   7.40   2.31   7.55   3.85  13.00   9.15   0.28
## Hue            12 48   0.68   0.11   0.66   0.48   0.96   0.48   0.55
## OD             13 48   1.68   0.27   1.66   1.27   2.47   1.20   0.64
## Proline        14 48 629.90 115.10 627.50 415.00 880.00 465.00   0.30
##               kurtosis   se
## Class                NaN  0.00
## Alcohol              -0.76  0.08
## Malic_acid           -0.53  0.16
## Ash                  -0.92  0.03
## Alcalinity           -0.69  0.33
## Magnesium            -0.72  1.57
## Total_phenols         1.12  0.05
## Flavanoids            -0.13  0.04
## Nonflavanoid_phenols -0.70  0.02
## Proanthocyanins       3.16  0.06
## Intensity            -0.92  0.33
## Hue                  -0.63  0.02
## OD                   0.21  0.04
## Proline              -0.63 16.61
```

Notamos que hay 13 variables de 3 diferentes clases, cada clase pertenece a un viñedo distinto, encontramos variables como la cantidad de alcohol, acido malico, ceniza, alcalinidad, magnesio, total de fenol, flavonoides, no flavonoides, proantocianidinas, intensidad, matiz, oxígeno disuelto, prolina.

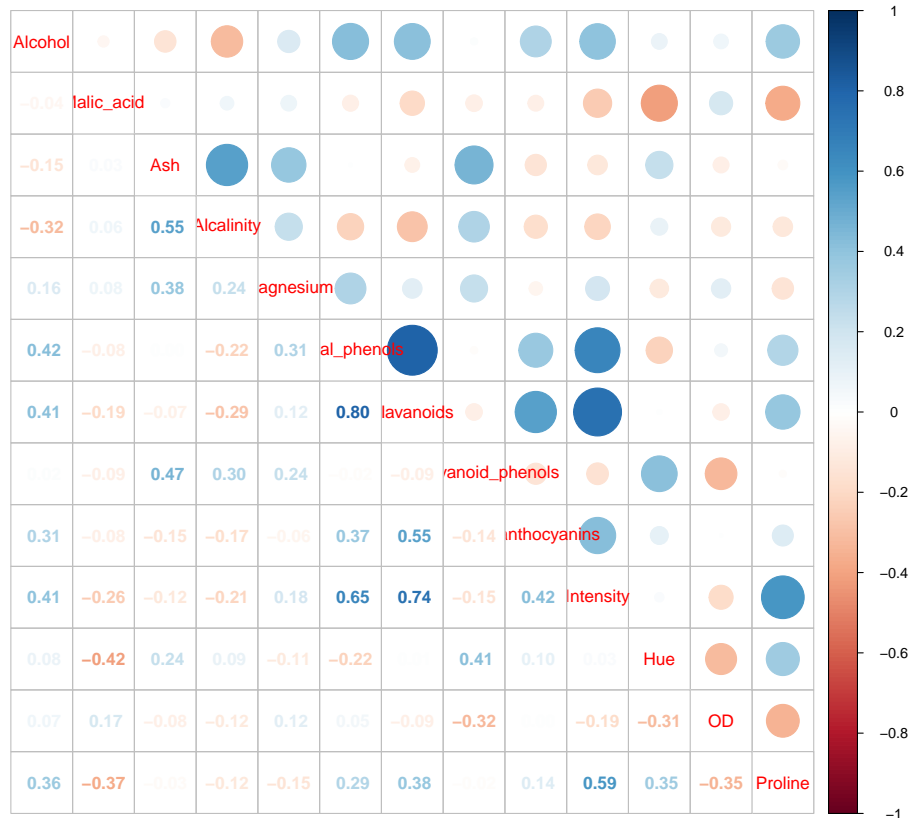
Al analizar las estadísticas podemos hacer algunas observaciones, al comparar el nivel de alcohol promedio, **notamos que en el segundo viñedo el alcohol es menor en promedio con un 12.28 comparado al 13.74 y 13.15 del viñedo 1 y 3**, también al comparar la alcalinidad el viñedo 1 es menor con un 17.04 contra un 20.24 y 21.42 del viñedo 2 y 3, el magnesio en promedio es menor en el viñedo 2 con un 94 en promedio versus un 106 y 99.31 del viñedo 1 y 3, también el nivel de intensidad en el viñedo 2 es menor con un 3.08 contra un 5.4 y 7.39, para el viñedo 3 encontramos que los flavonoides son menores con un .78 en comparación a un 2.98 y 2.08, al analizar la prolina notamos que el viñedo 1 tiene niveles 2 veces mas altos que los otros viñedos, con un 1115.7 versus 519 y 629, en las otras variables los índices se mantienen semejantes o con una varianza pequeña.

## Análisis Unitario de los Viñedos

### Viñedo 1

Hacemos un correlograma para identificar que variables están relacionadas

```
corrplot.mixed(cor(vinedo_1[2:14]))
```



Como podemos ver en los coeficientes de correlación, la alcalinidad y la ceniza es de .55, lo cual nos indica una correlación positiva débil-moderada, lo cual es logico ya que la alcalinidad se debe en parte a la ceniza contenida en el vino, tambien notamos una relación positiva fuerte entre el total de fenoles y flavonoides de .80, el total de fenoles tambien se relaciona fuertemente con la intensidad de sabor con un coeficiente de correlación de .65, las proantocianidinas y los flavonoides se relacionan moderadamente con un .55, tambien la intensidad con la prolina se relacionan moderadamente con un .59, la intensidad con lo flavonoides tiene una correlación de .74, es decir una relación moderada, tambien se observa una relación negativa débil entre el ácido malico y el matiz -.42, la prolina y el ácido malico se relacionan negativamente con un -.37.

Para el siguiente análisis tomaremos como referencia a las variables en las que la correlación es mayor a .50 ya que esto nos indica una relación moderada, ya que los flavonoides estan relacionados de forma positiva con el total de fenoles y con la intensidad, y la intensidad esta relacionada con la prolina haremos una gresión lineal multiple para ver como describen las variables antes mencionadas a la intensidad

```
intensidad_1<-lm(vinedo_1$Intensity ~ vinedo_1$Total_phenols
+ vinedo_1$Flavanoids + vinedo_1$Proline , data = datos)
summary(intensidad_1)
```

```
##
## Call:
## lm(formula = vinedo_1$Intensity ~ vinedo_1$Total_phenols + vinedo_1$Flavanoids +
##     vinedo_1$Proline, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39536 -0.41213  0.01407  0.35549  2.23542
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.8136313   0.8525036  -3.300 0.001699 **
## vinedo_1$Total_phenols  0.5980422   0.4763314   1.256 0.214599
## vinedo_1$Flavanoids    1.4732719   0.4196723   3.511 0.000901 ***
## vinedo_1$Proline       0.0020162   0.0004689   4.300 7.04e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7314 on 55 degrees of freedom
## Multiple R-squared:  0.6693, Adjusted R-squared:  0.6513
## F-statistic: 37.11 on 3 and 55 DF,  p-value: 3.004e-13
```

Notamos que aunque el total de fenoles y los flavonoides estan altamente relacionados al momento de hacer el modelo de regresión, el total de fenoles no es significativo en la explicación de los flavonoides, por esta razón quitaremos esta variable del modelo.

Realizando un nuevo modelo

```
intensidad_1.1<-lm(vinedo_1$Intensity ~ vinedo_1$Flavanoids
+ vinedo_1$Proline , data = datos)
summary(intensidad_1.1)
```

```
##
## Call:
## lm(formula = vinedo_1$Intensity ~ vinedo_1$Flavanoids + vinedo_1$Proline,
##     data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.53137 -0.45339  0.04512  0.38981  2.03096
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.3320155   0.7652313  -3.047  0.00352 **
## vinedo_1$Flavanoids  1.8856354   0.2625903   7.181 1.74e-09 ***
## vinedo_1$Proline     0.0020047   0.0004712   4.255 8.04e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7351 on 56 degrees of freedom
## Multiple R-squared:  0.6599, Adjusted R-squared:  0.6477
## F-statistic: 54.32 on 2 and 56 DF,  p-value: 7.7e-14
```

Notamos que aunque eliminamos la variable de total de fenoles, el modelo no difiere en la R cuadrada ajustada con el modelo que sí contiene a la variable total de fenoles ya que obtenemos que el nuevo modelo explica el 64% de variabilidad dados los regresores, además en este modelo todos los regresores son significativos con un  $\alpha = 0.05$  y en conjunto las variables son significativas basandonos en la prueba F.

## Intervalos de confianza para los estimadores

Ahora daremos un intervalo de confianza para los estimadores de los regresores.

```
confint(intensidad_1.1, level = 0.95)
```

```
##                2.5 %      97.5 %  
## (Intercept)    -3.864958004 -0.79907309  
## vinedo_1$Flavonoids  1.359603709  2.41166700  
## vinedo_1$Proline    0.001060783  0.00294859
```

## Verificación de Supuestos

Ahora verificaremos los supuestos de linealidad, autocorrelación, homocedasticidad y normalidad de residuales

### Linealidad

Verificaremos este supuesto calculando la media de los residuos, queremos que este valor sea cercano a 0.

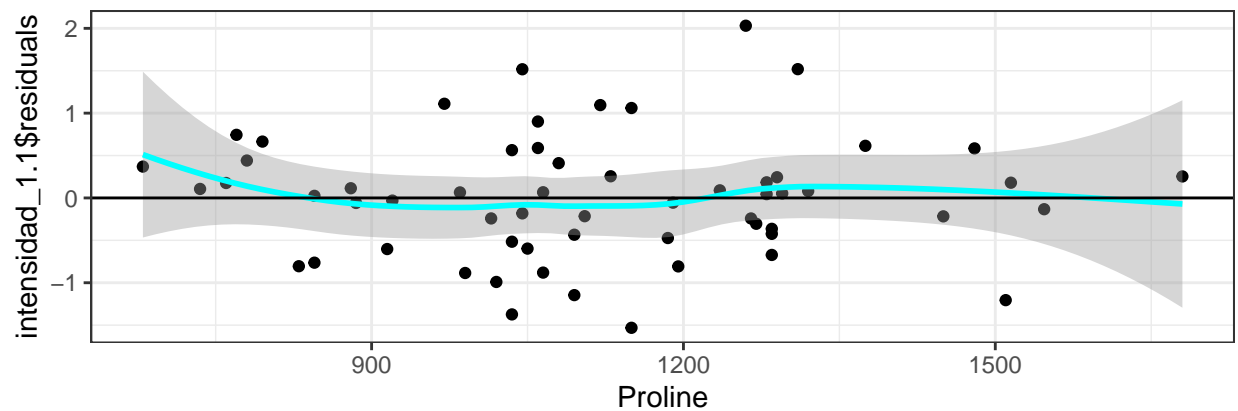
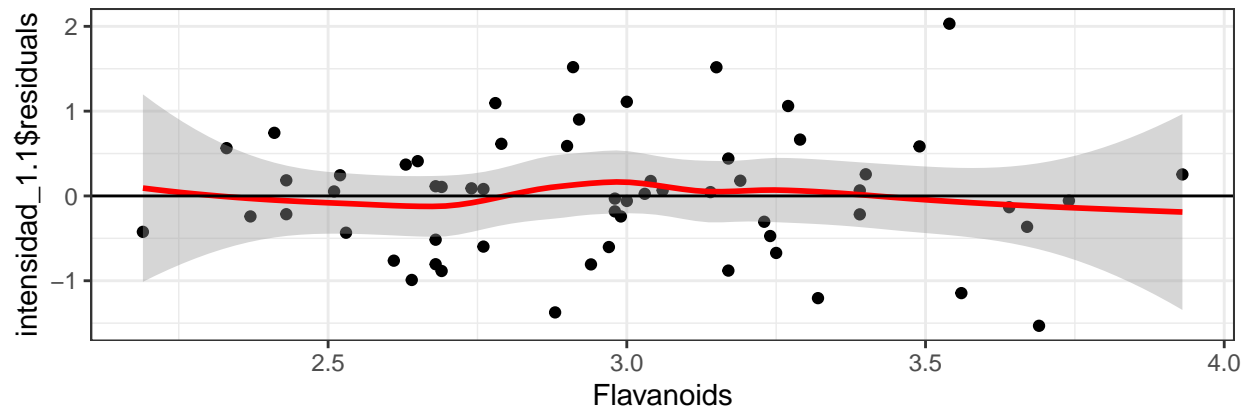
```
mean(intensidad_1.1$residuals)
```

```
## [1] 2.747534e-17
```

De forma gráfica

```
plot_1<-ggplot(data = vinedo_1, aes(Flavonoids, intensidad_1.1$residuals))+  
  geom_point() + geom_smooth(color="red") +  
  geom_hline(yintercept = 0)+  
  theme_bw()  
  
plot_2<-ggplot(data = vinedo_1, aes(Proline, intensidad_1.1$residuals))+  
  geom_point() + geom_smooth(color="cyan")+  
  geom_hline(yintercept = 0)+  
  theme_bw()  
  
grid.arrange(plot_1, plot_2)
```





### Independencia o No Autocorrelación

Realizaremos un test Durbin-Watson donde la hipótesis nula sostiene que la autocorrelación es igual a cero

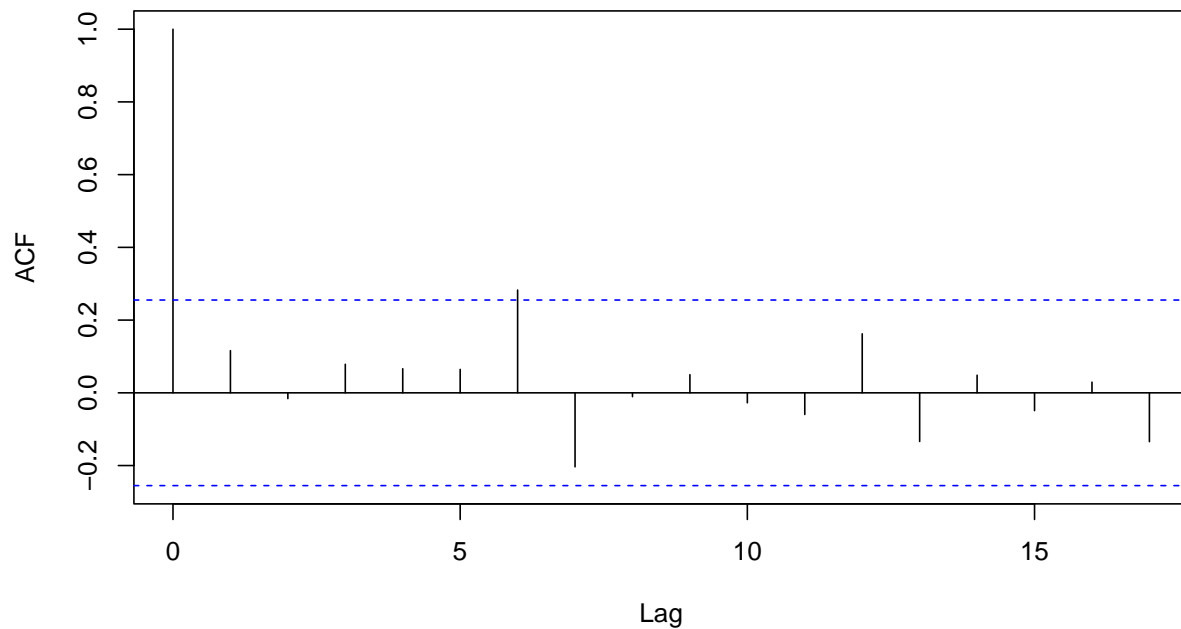
```
dwtest(intensidad_1.1)
```

```
##
## Durbin-Watson test
##
## data: intensidad_1.1
## DW = 1.7634, p-value = 0.1673
## alternative hypothesis: true autocorrelation is greater than 0
```

Con una significancia de  $\alpha = 0.05$  no se rechaza la hipótesis nula, por lo tanto podemos concluir que la autocorrelación es cero con un 95% de confianza, además que el valor del estadístico DW es cercano a 2 lo que indica que no hay autocorrelación.

```
acf(intensidad_1.1$residuals, main="Supuesto de covarianza de residuales igual a cero")
```

### Supuesto de covarianza de residuales igual a cero



Para verificar este supuesto haremos otra prueba Breusch-Godfrey, con hipótesis nula no existe autocorrelación

```
bgtest(intensidad_1.1)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: intensidad_1.1
## LM test = 0.8195, df = 1, p-value = 0.3653
```

Por lo tanto no se rechaza la hipótesis nula, es decir no existe autocorrelación.

### Homocedasticidad

Para este supuesto haremos una prueba Breusch-Pagan, con hipótesis nula se cumple la homocedasticidad.

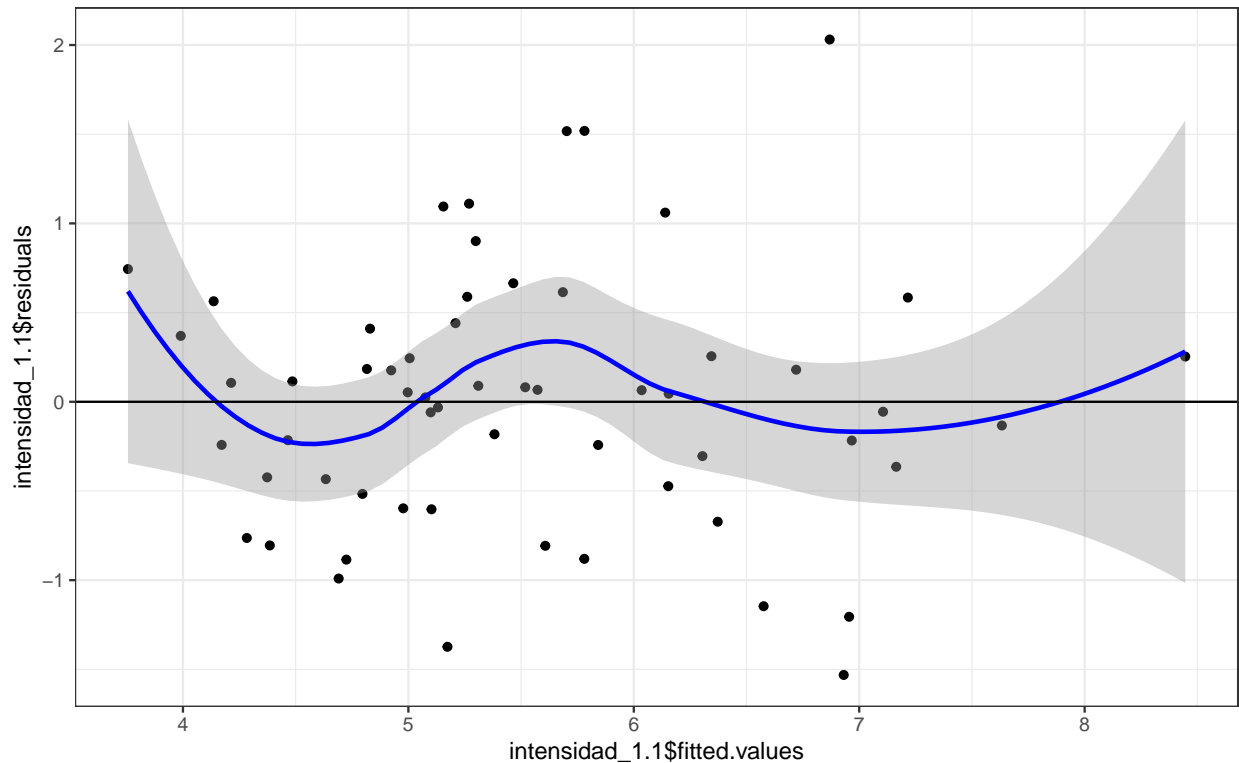
```
bptest(intensidad_1.1)
```

```
##
## studentized Breusch-Pagan test
##
## data: intensidad_1.1
## BP = 2.9706, df = 2, p-value = 0.2264
```

No se rechaza la hipótesis nula, por lo cual existe homocedasticidad.

Graficamente podemos notar que no existe algún patron alrededor del cero, lo cual es buen indicativo que existe homocedasticidad.

```
ggplot(data = vinedo_1, aes(intensidad_1.1$fitted.values, intensidad_1.1$residuals)) +  
geom_point() + geom_smooth(color = "blue", se = TRUE) + geom_hline(yintercept = 0) +  
theme_bw()
```



### Normalidad de los residuales

Realizando una prueba Shapiro-Wilks, con hipotesis nula la muestra se distribuye normal.

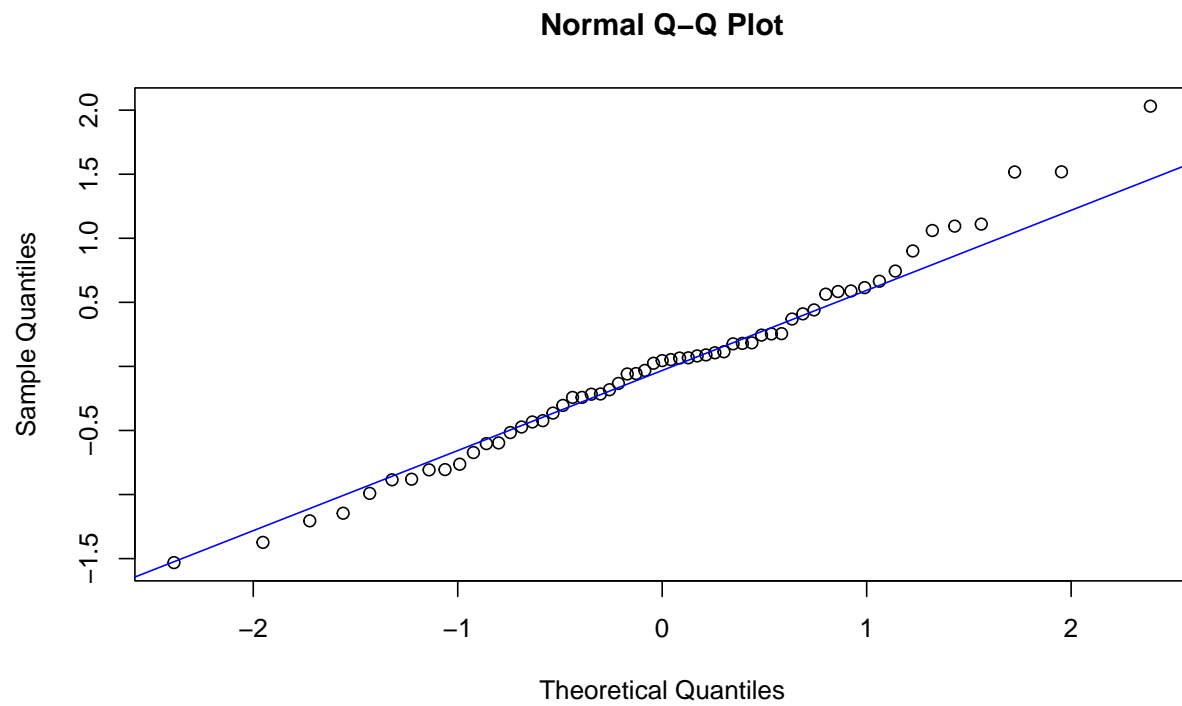
```
shapiro.test(intensidad_1.1$residuals)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  intensidad_1.1$residuals  
## W = 0.98541, p-value = 0.7012
```

No se rechaza la hipotesis nula, existe normalidad.

Gráficamente lo podemos observar en el siguiente qqplot

```
qqnorm(intensidad_1.1$residuals)  
qqline(intensidad_1.1$residuals, col="blue")
```

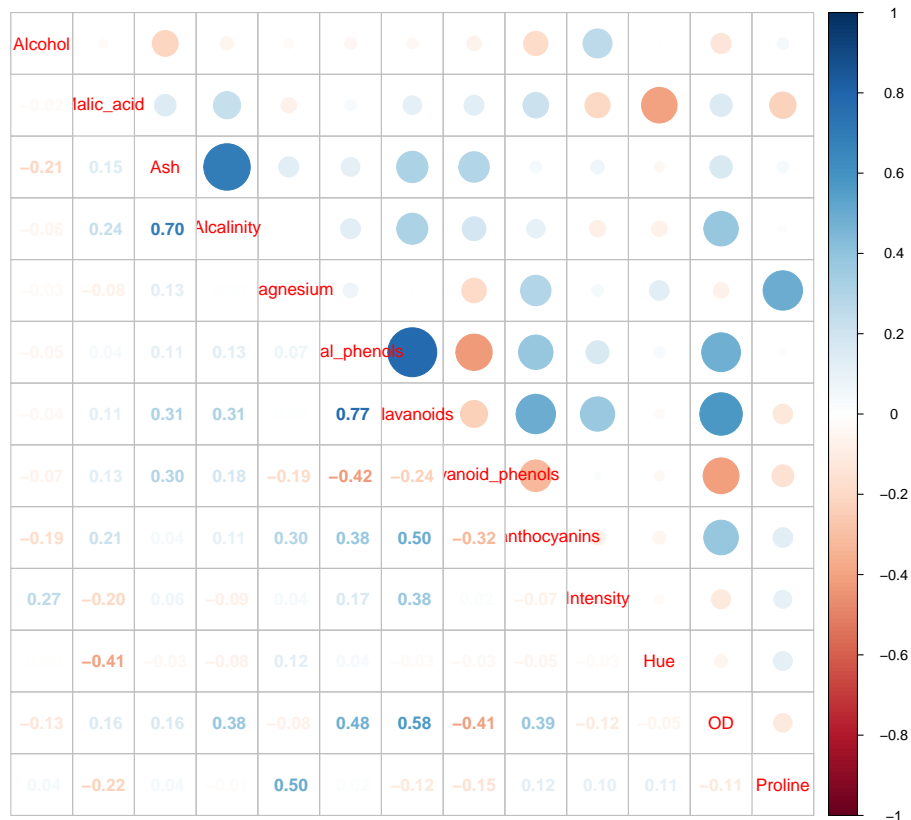


Notamos que los datos se ajustan de buena forma a la linea.

Por lo tanto podemos concluir que se cumplen los supuestos para nuestro modelo.

## Vinredo 2

```
corrplot.mixed(cor(vinedo_2[2:14]))
```



Al interpretar las graficas podemos notar que **existe una relación positiva entre la alcalinidad y la ceniza con .70** tambien los **flavonoides con el total de fenoles .77** existe una relación debil positiva entre **oxigeno diluido y el total de flavonoides con .48** y **oxigeno diluido y flavonoides con .58** tambien podemos observar una relación negativa moderada entre **los no flavonoides y total de fenoles con -.425** tambien notamos una relación entre **prolina y el magnesio con .50** y de la **proantocianidina y flavonoides con .50**

Para este viñedo nos centraremos en el estudio de la ceniza y la alcalinidad, el total de fenoles y flavonoides, flavonoides y el oxígeno diluido, flavonoides y proantocianidina.

La variable flavonoides que es la que mayor relación moderada tiene con las otras variables, por lo cual haremos una regresión lineal

```
flavonoides_2<-lm(vinedo_2$Flavanoids ~ vinedo_2$Total_phenols + vinedo_2$Proanthocyanins + vinedo_2$OD
summary(flavonoides_2)
```

```
##
## Call:
## lm(formula = vinedo_2$Flavanoids ~ vinedo_2$Total_phenols + vinedo_2$Proanthocyanins +
##      vinedo_2$OD, data = vinedo_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91257 -0.22434 -0.01541  0.18288  1.95730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.88070    0.28995  -3.037  0.00340 **
```

```
## vinedo_2$Total_phenols    0.76677    0.10569    7.255 5.29e-10 ***
## vinedo_2$Proanthocyanins  0.22007    0.09077    2.425 0.01803 *
## vinedo_2$OD              0.31261    0.11624    2.689 0.00903 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4093 on 67 degrees of freedom
## Multiple R-squared:  0.6781, Adjusted R-squared:  0.6637
## F-statistic: 47.05 on 3 and 67 DF,  p-value: < 2.2e-16
```

Podemos notar que todas las variables que incluimos en nuestro modelo son significativas, aunado a que la prueba F nos indica que los coeficientes de las variables no son 0, al realizar este modelo de regresión obtenemos que se explica el 67 % de la varianza de los flavonoides con los regresores que se incluyen en el modelo.

## Intervalos de confianza para los estimadores

Ahora encontraremos los intervalos de confianza para los estimadores de los coeficientes que obtuvimos con el modelo anterior con un nivel de confianza del 95%

```
confint(flavonoides_2, level = 0.95)
```

```
##                2.5 %    97.5 %
## (Intercept)    -1.45944947 -0.3019587
## vinedo_2$Total_phenols  0.55580647  0.9777364
## vinedo_2$Proanthocyanins 0.03889553  0.4012444
## vinedo_2$OD        0.08058618  0.5446372
```

Es importante notar que en nuestros intervalos no abarcan al 0, lo cual es congruente con las pruebas t individuales para cada coeficiente beta.

## Verificación de Supuestos

Verificando los supuestos que se deben cumplir para el modelo de regresión

### Linealidad

```
mean(flavonoides_2$residuals)
```

```
## [1] -3.338694e-18
```

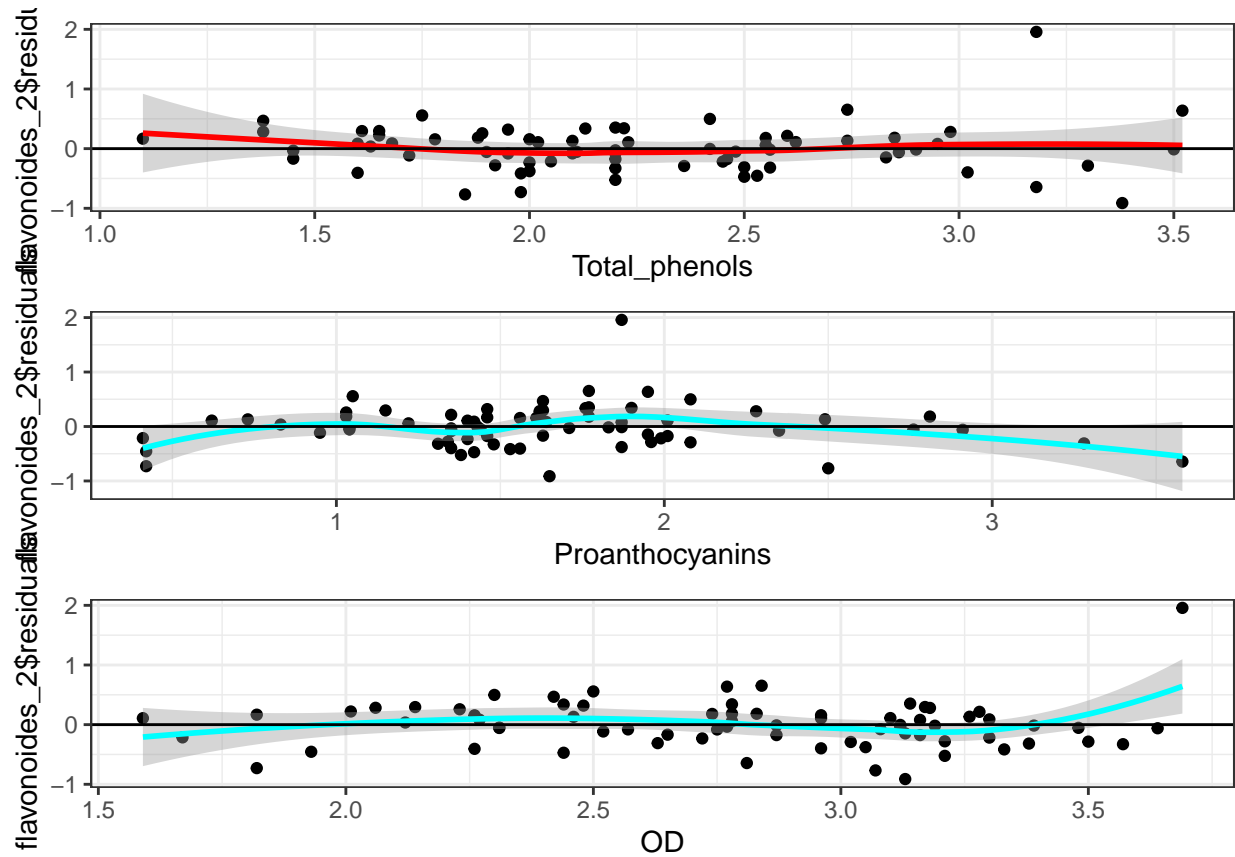
De forma gráfica

```
plot_1<-ggplot(data = vinedo_2, aes(Total_phenols, flavonoides_2$residuals))+
  geom_point() + geom_smooth(color="red")+ geom_hline(yintercept = 0)+ theme_bw()

plot_2<-ggplot(data = vinedo_2, aes(Proanthocyanins, flavonoides_2$residuals)) +
  geom_point() + geom_smooth(color="cyan") + geom_hline(yintercept = 0) + theme_bw()
```

```
plot_3<-ggplot(data = vinedo_2, aes(OD, flavonoides_2$residuals)) +
  geom_point() + geom_smooth(color="cyan")+ geom_hline(yintercept = 0) + theme_bw()

grid.arrange(plot_1, plot_2, plot_3)
```



## Independencia

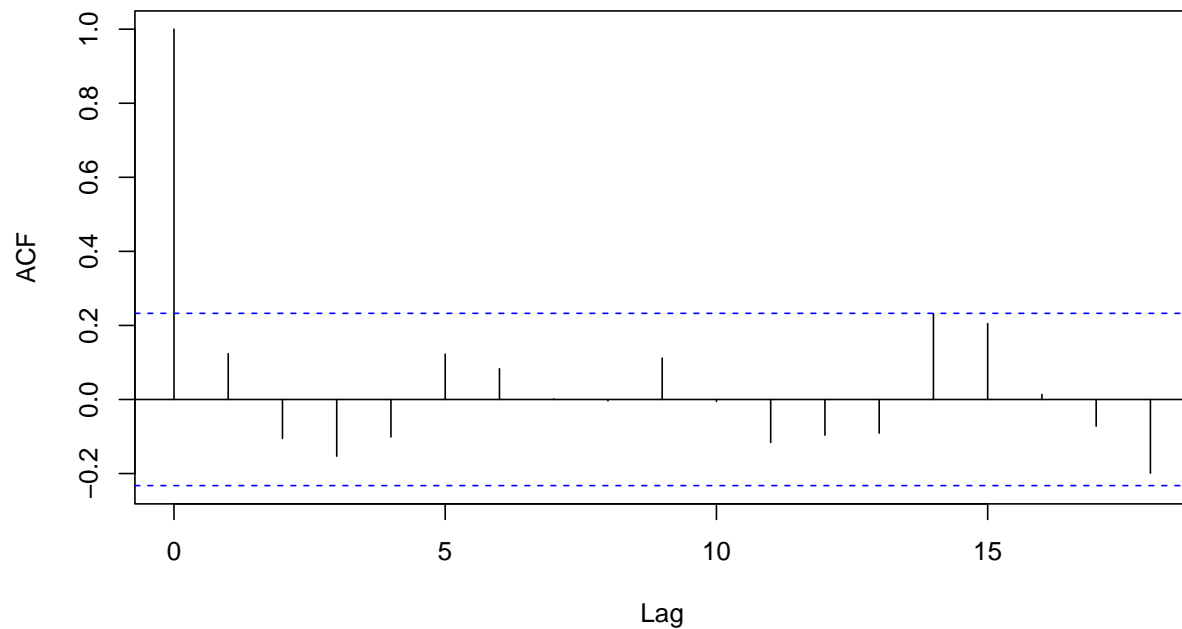
```
dwtest(flavonoides_2)
```

```
##
## Durbin-Watson test
##
## data: flavonoides_2
## DW = 1.7049, p-value = 0.09253
## alternative hypothesis: true autocorrelation is greater than 0
```

No existe autocorrelación.

```
acf(flavonoides_2$residuals, main="Supuesto de covarianza de residuales igual a cero")
```

### Supuesto de covarianza de residuales igual a cero



```
bgtest(flavonoides_2)
```

```
##  
## Breusch-Godfrey test for serial correlation of order up to 1  
##  
## data: flavonoides_2  
## LM test = 1.1368, df = 1, p-value = 0.2863
```

Por lo tanto no se rechaza la hipótesis nula, es decir no existe autocorrelación.

### Homocedasticidad

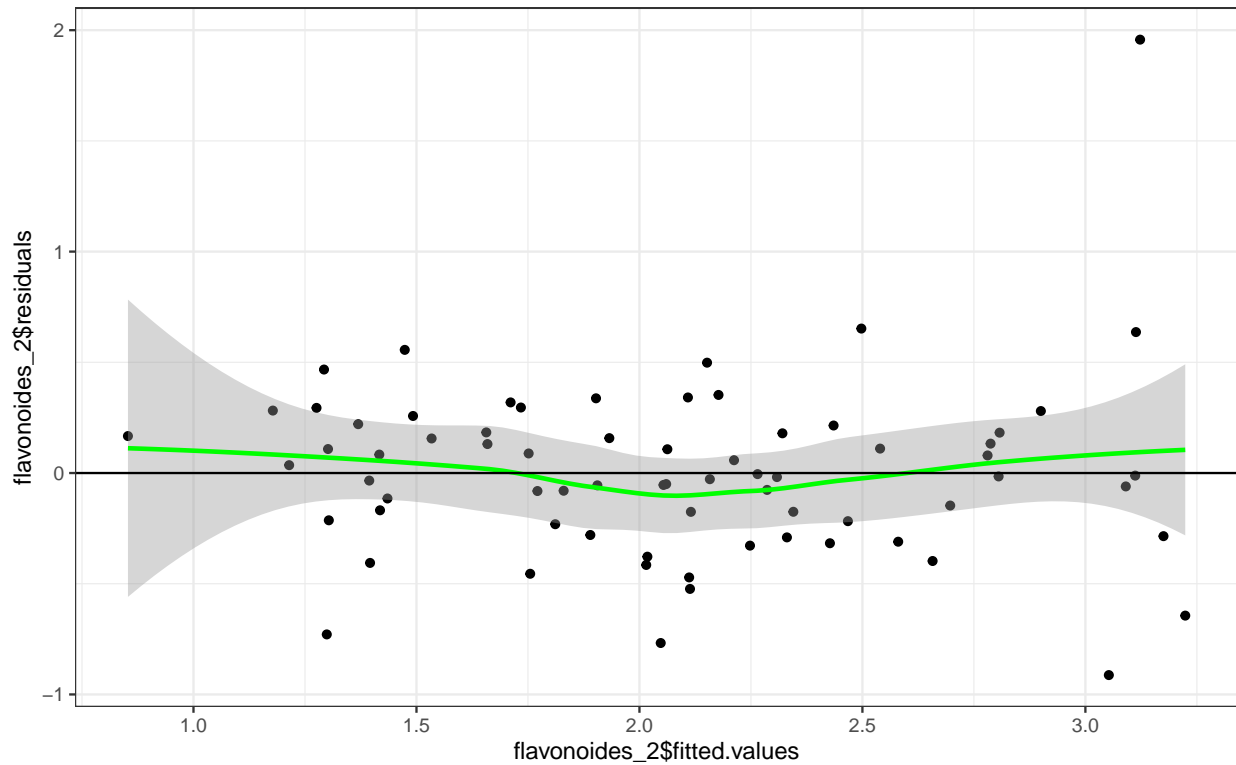
```
bptest(flavonoides_2)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: flavonoides_2  
## BP = 5.5009, df = 3, p-value = 0.1386
```

Existe homocedasticidad.

```
ggplot(data = vinedo_2, aes(flavonoides_2$fitted.values, flavonoides_2$residuals)) +  
  geom_point() + geom_smooth(color = "green", se = TRUE) + geom_hline(yintercept = 0) +  
  theme_bw()
```





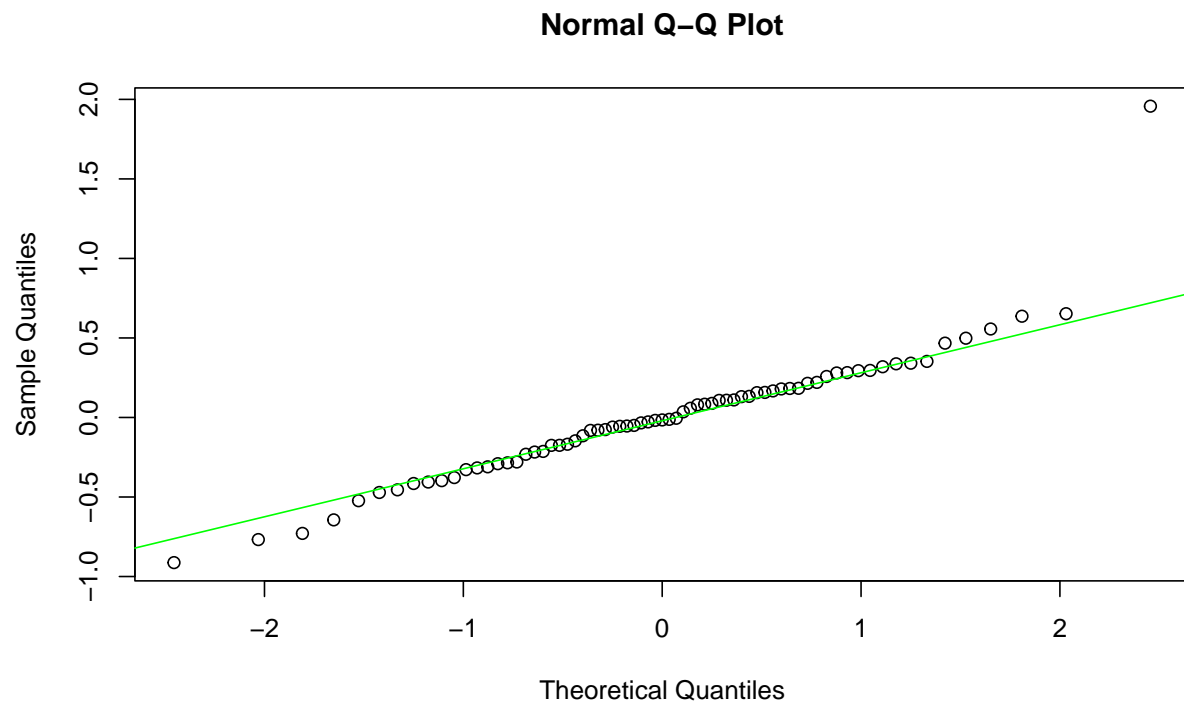
Notamos que los residuos se distribuyen aleatoriamente alrededor del cero.

### Normalidad de los residuales

```
shapiro.test(flavonoides_2$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  flavonoides_2$residuals
## W = 0.89936, p-value = 3.246e-05
```

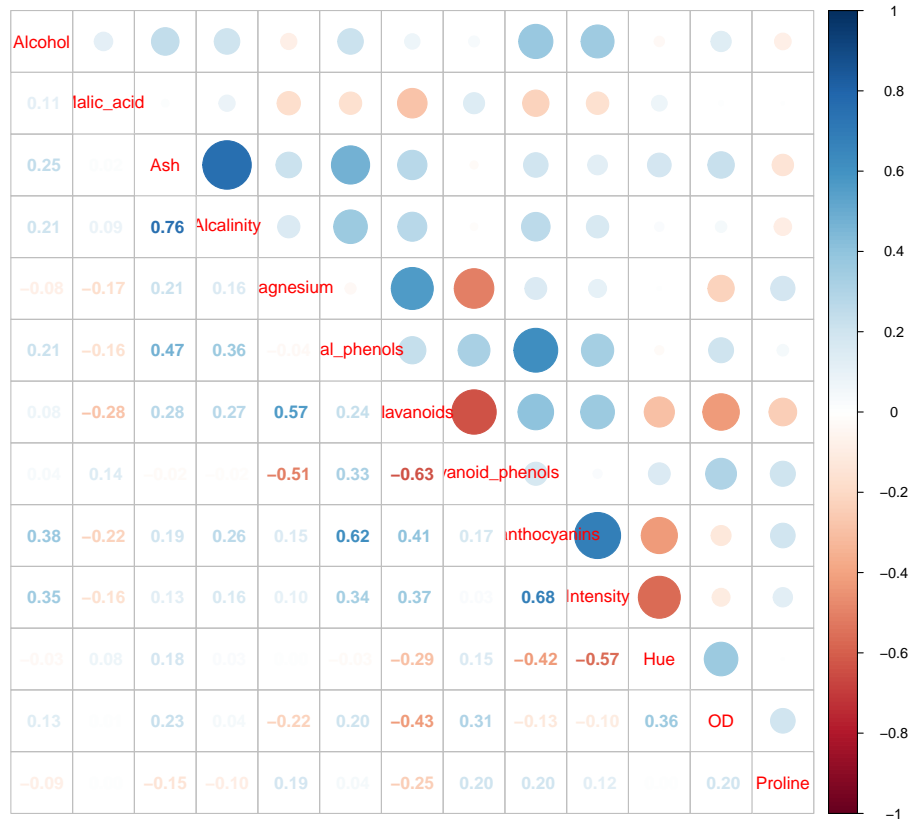
```
qqnorm(flavonoides_2$residuals)
qqline(flavonoides_2$residuals, col="green")
```



Existe Normalidad entre los residuales.

### Viñedo 3

```
corrplot.mixed(cor(vinedo_3[2:14]))
```



Encontramos relaciones positivas moderadas como alcalinidad y ceniza con .76 tambien entre intensidad con proantocianidina .68 proantocianidina con fenoles totales .62 los flavonoides con el magnesio .57 pero tambien relaciones negativas como no flavonoides fenoles con el magnesio -.51 y logicamente la de flavonoides con no flavonoides con -.63 tambien la de matiz con la intensidad con -.56 .

Sin embargo las relaciones no son lo suficiente fuertes para intentar hacer una regresión lineal múltiple, no obstante es importante notar que los estudios realizados en este viñedo muestran que tiene un mayor grado de magnesio en comparación a los otros viñedos.

## Conclusiones

### Viñedo 1

Para el primer viñedo podemos concluir que la intensidad del vino esta moderadamente influenciada por el total de fenoles, los flavonoides y la prolina, ya que basandonos en nuestro modelo de regresión, las variables antes mencionadas afectan la intensidad de forma positiva.

Además las muestras indican que este viñedo tiene una mayor cantidad de alcohol, una menor alcalinidad y casi dos veces más prolina y una mayor cantidad de magnesio en comparación con los otros viñedos.

### Viñedo 2

Para el segundo viñedo podemos inferir que el nivel de los flavonoides esta moderadamente influidos por el total de fenoles, la proantocianidina y el nivel de oxigeno diluido, en promedio tiene menor cantidad de magnesio que los otros viñedos

### **Viñedo 3**

Para el tercer viñedo el matiz es menor pero tiene mayor intensidad, duplicando la intensidad del viñedo 2, además la intensidad esta relacionada con la proantocianidina moderadamente, en contraparte tiene menor cantidad de flavonoides.