

Taller 2: Análisis ANOVA-MANOVA

JRojas, MRamirez, LRomero

6/24/2020

Introducción

A continuación se presentan una serie de ejercicios propuesto que permitiran interiorizar el análisis de datos multivariados utilizando las técnicas ANOVA y MANOVA. Estos ejercicios hacen parte del contenido académico desarrollado por el profesor Aquiles Enrique Darghan Contreras referente a la asignatura de Métodos Multivariados.

Previo a la solución de los ejercicios es necesario instalar y cargar las siguientes librerías.

```
library(readxl)
library(ggplot2)
library(mvShapiroTest)
library(biotools)

## ---
## biotools version 3.1
library(outliers)
```

Datos

Los ejercicios mostrados son desarrollados a partir de los siguientes datos correspondientes a una toma realizada con un espectroradiometro en los anchos de bandas correspondientes a 560 nm y 720 nm de las especies *SS*, *JL* y *LP*

| D_560nm | D_720nm | Species |
|---------|---------|---------|
| 9.33 | 19.14 | SS |
| 8.74 | 19.55 | SS |
| 9.31 | 19.24 | SS |
| 8.27 | 16.37 | SS |
| 10.22 | 25 | SS |
| 10.13 | 25.32 | SS |
| 10.42 | 27.12 | SS |
| 10.62 | 26.28 | SS |
| 15.25 | 38.89 | SS |
| 16.22 | 36.67 | SS |
| 17.24 | 40.74 | SS |
| 12.77 | 67.5 | SS |
| 12.07 | 33.03 | JL |
| 11.03 | 32.37 | JL |
| 12.48 | 31.31 | JL |
| 12.12 | 33.33 | JL |
| 15.38 | 40 | JL |

| D_560nm | D_720nm | Species |
|---------|---------|---------|
| 14.21 | 40.48 | JL |
| 9.69 | 33.9 | JL |
| 14.35 | 40.15 | JL |
| 38.71 | 77.14 | JL |
| 44.74 | 78.57 | JL |
| 36.67 | 71.43 | JL |
| 37.21 | 45 | JL |
| 8.73 | 23.27 | LP |
| 7.94 | 20.87 | LP |
| 8.37 | 22.16 | LP |
| 7.86 | 21.78 | LP |
| 8.45 | 26.32 | LP |
| 6.79 | 22.73 | LP |
| 8.34 | 26.67 | LP |
| 7.54 | 24.87 | LP |
| 14.04 | 44.44 | LP |
| 13.51 | 37.93 | LP |
| 13.33 | 37.93 | LP |
| 12.77 | 60.87 | LP |

Los datos se encuentran disponibles en un archivo de google sheets para quien desee realizar los mismos ejercicios presentados.

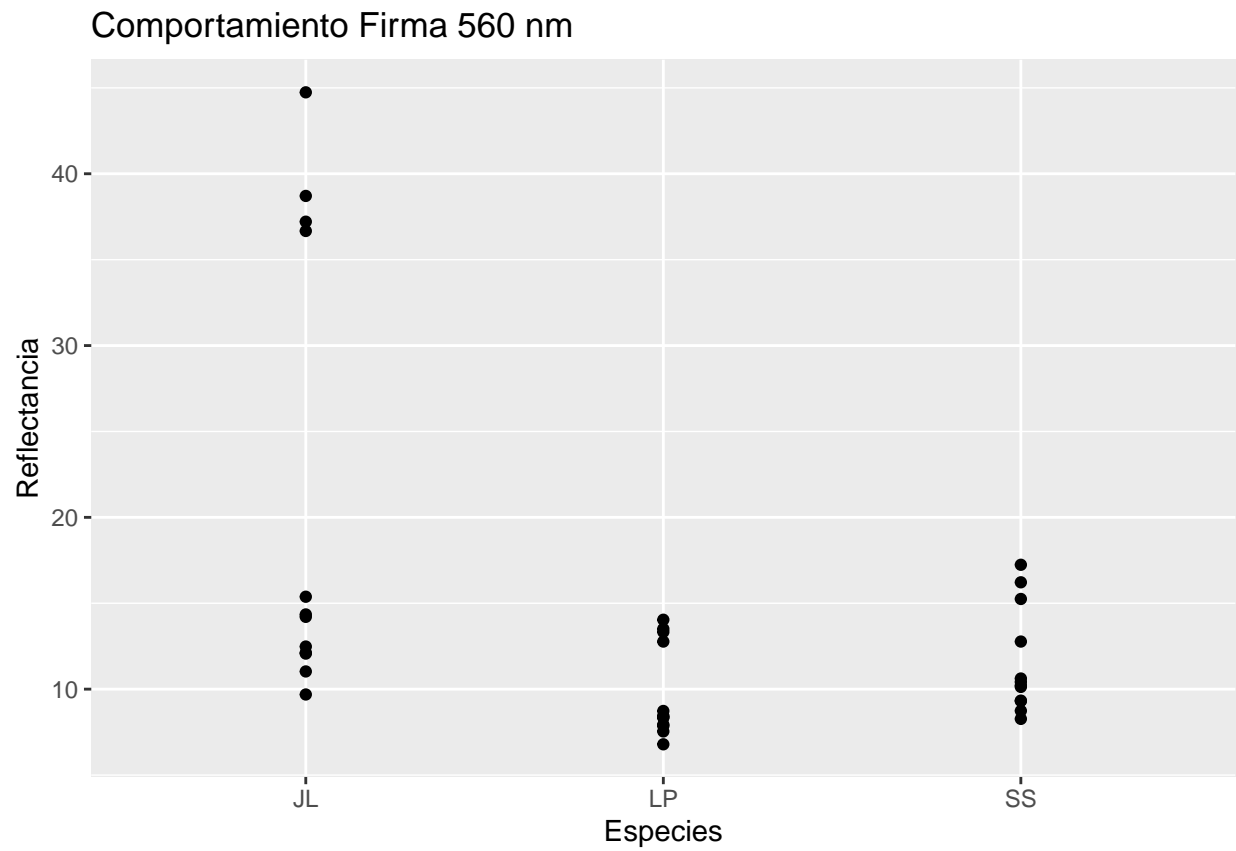
Como primer paso los datos son cargados a la variable denominada *Reflectancia*

```
Reflectancia <- read_excel("Datos_ANOVA_MANOVA.xlsx")
```

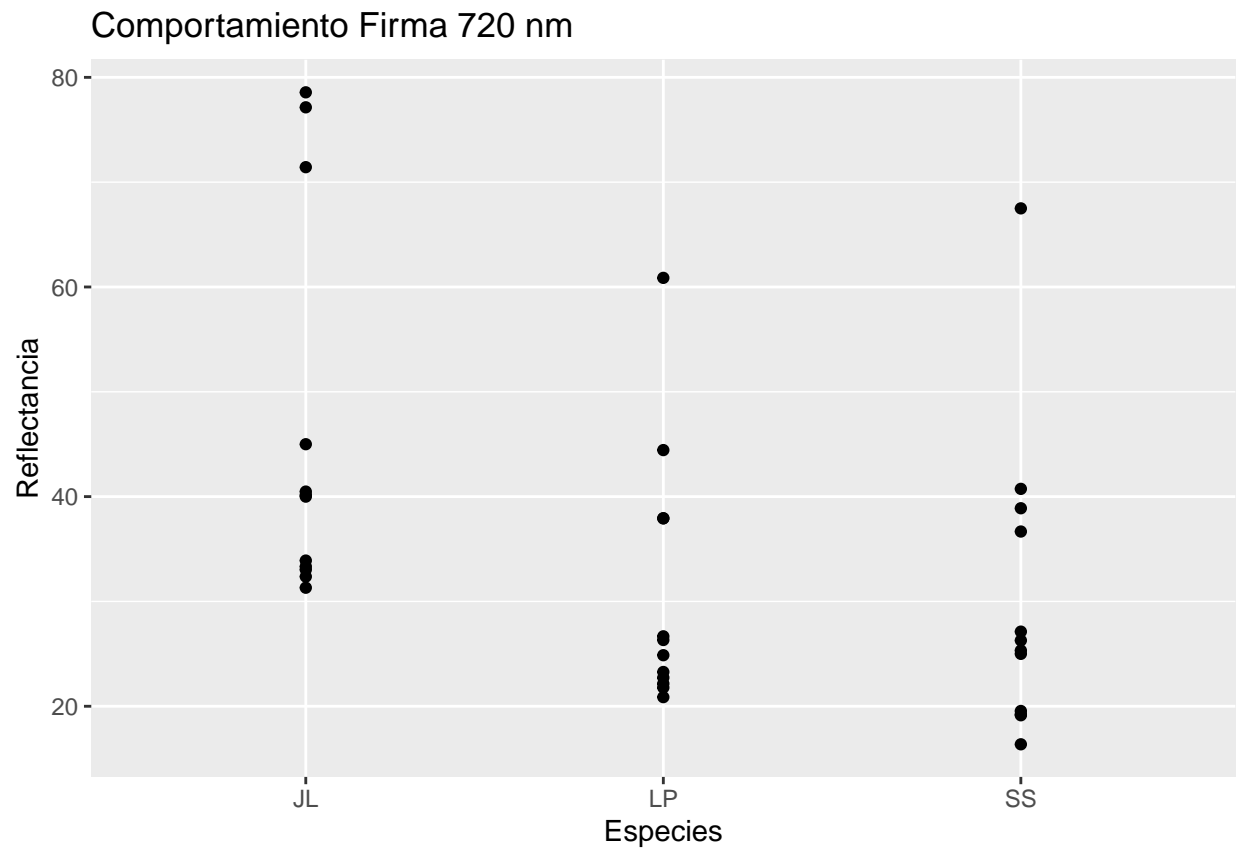
Punto 1: MANOVA

Para el análisis se realizan las gráficas para cada longitud de onda y un gráfico de dispersión que muestre la relación de las dos:

```
ggplot(data=Reflectancia, aes(Species,D_560nm)) + geom_point() + ggtitle("Comportamiento Firma 560 nm")
```

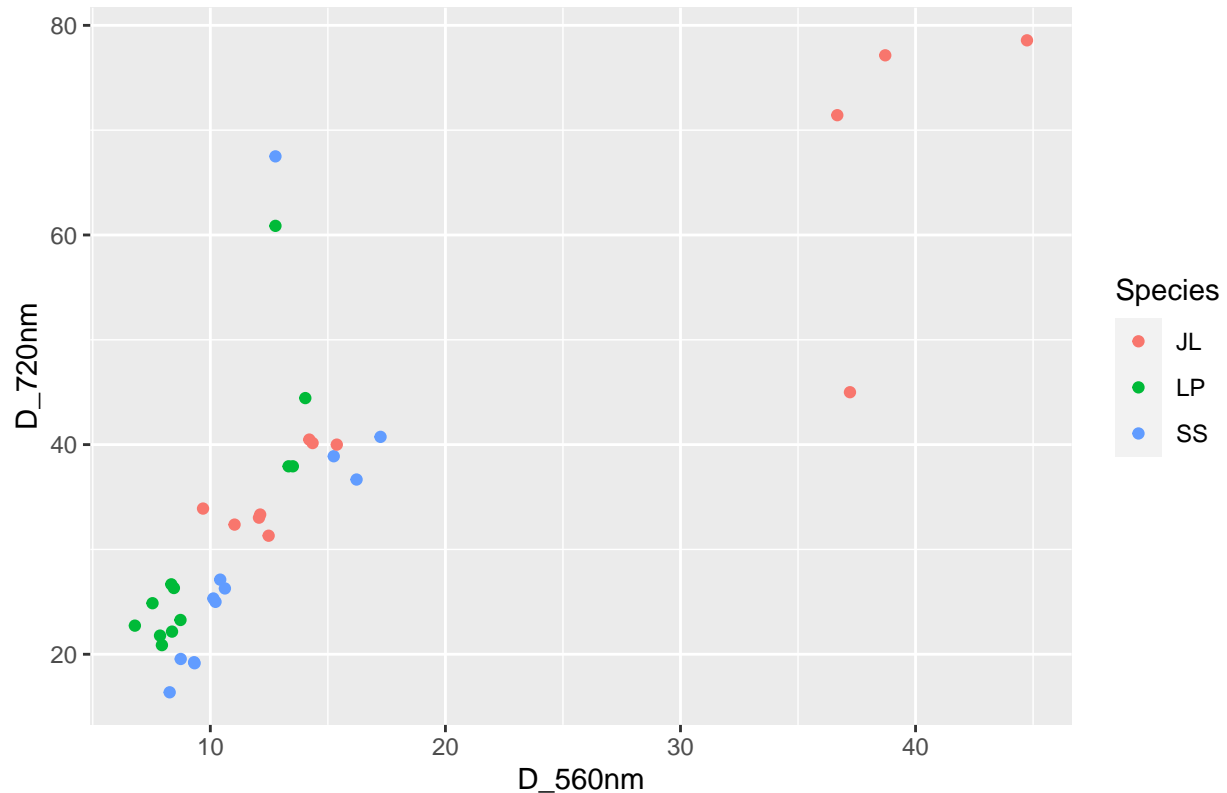


```
ggplot(data=Reflectancia, aes(Species,D_720nm)) + geom_point() + ggtitle("Comportamiento Firma 720 nm")
```



```
ggplot(data=Reflectancia, aes(D_560nm,D_720nm, color=Species)) + geom_point() + ggtitle("Comportamiento
```

Comportamiento Firma 560 nm vs 720 nm



Se procede a realizar el análisis MANOVA y se presenta un resumen del mismo con la hipótesis con el test de Wilks

```
MN=manova(cbind(Reflectancia$D_560nm,Reflectancia$D_720nm) ~ Reflectancia$Species)
MN
```

```
## Call:
## manova(cbind(Reflectancia$D_560nm, Reflectancia$D_720nm) ~ Reflectancia$Species)
##
## Terms:
##              Reflectancia$Species Residuals
## resp 1              965.181  2147.714
## resp 2             2026.856  7536.997
## Deg. of Freedom              2          33
##
## Residual standard errors: 8.067357 15.11271
## Estimated effects may be unbalanced
```

```
summary(MN,test="Wilks")
```

```
##              Df    Wilks approx F num Df den Df Pr(>F)
## Reflectancia$Species  2 0.67704   3.4452      4    64 0.013 *
## Residuals              33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observandose que se presentan diferencias significativas entre las dos especies

Punto 2: ANOVA

```
AV1=aov(Reflectancia$D_560nm ~ Reflectancia$Species)
summary(AV1)
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## Reflectancia$Species  2   965.2    482.6    7.415 0.00219 **
## Residuals            33  2147.7     65.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AV2=aov(Reflectancia$D_720nm ~ Reflectancia$Species)
summary(AV2)
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## Reflectancia$Species  2   2027   1013.4    4.437 0.0196 *
## Residuals            33   7537    228.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se concluye que Las especies no contribuyen significativamente a cada una de las bandas analizadas.

Punto 3: Test de Correlación

A continuación es realizado el test de correlación de *Pearson*

```
cor.test(Reflectancia$D_560nm,Reflectancia$D_720nm)
```

```
##
## Pearson's product-moment correlation
##
## data:  Reflectancia$D_560nm and Reflectancia$D_720nm
## t = 8.144, df = 34, p-value = 1.691e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6611588 0.9009499
## sample estimates:
##          cor
## 0.8130816
```

Observandose que si existe relación lineal entre las dos variables, por lo tanto no se puede asumir que sea nula la correlación.

Punto 4: Normalidad univariada

A continuación es realizado el test de *Shapiro-Wilk* para cada una de las respuestas

```
shapiro.test(Reflectancia$D_560nm)
```

```
##
## Shapiro-Wilk normality test
##
## data:  Reflectancia$D_560nm
## W = 0.64673, p-value = 4.538e-08
```

```
shapiro.test(Reflectancia$D_720nm)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: Reflectancia$D_720nm  
## W = 0.84251, p-value = 0.0001286
```

No se puede asumir normalidad en alguna de las dos variables.

Punto 5: Normalidad multivariada

A continuación es realizado el test multivariado de *Shapiro-Wilk*

```
mvShapiro.Test(as.matrix(Reflectancia[,1:2]))
```

```
##  
## Generalized Shapiro-Wilk test for Multivariate Normality by  
## Villasenor-Alva and Gonzalez-Estrada  
##  
## data: as.matrix(Reflectancia[, 1:2])  
## MVW = 0.81281, p-value = 3.948e-08
```

No se puede asumir normalidad multivariada.

Punto 6: Igualdad de Varianza Univariante

A continuación es realizado el test multivariado de *Bartlett*

```
bartlett.test(Reflectancia$D_560nm ~ Reflectancia$Species)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: Reflectancia$D_560nm by Reflectancia$Species  
## Bartlett's K-squared = 32.714, df = 2, p-value = 7.876e-08
```

```
bartlett.test(Reflectancia$D_720nm ~ Reflectancia$Species)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: Reflectancia$D_720nm by Reflectancia$Species  
## Bartlett's K-squared = 1.7462, df = 2, p-value = 0.4177
```

Existe igualdad de varianza en los valores de reflectancia en los valores dados a 560 nm mientras que con 720 no se puede asumir esto.

Punto 7: Igualdad de matrices de varianza y covarianza

Para evaluar la igualdad de la matriz de varianza y covarianza fue realizado el test M de *Box*

```
boxM(as.matrix(Reflectancia[, -3]), as.matrix(Reflectancia$Species))
```

```
##  
## Box's M-test for Homogeneity of Covariance Matrices
```

```
##
## data:  as.matrix(Reflectancia[, -3])
## Chi-Sq (approx.) = 41.969, df = 6, p-value = 1.865e-07
```

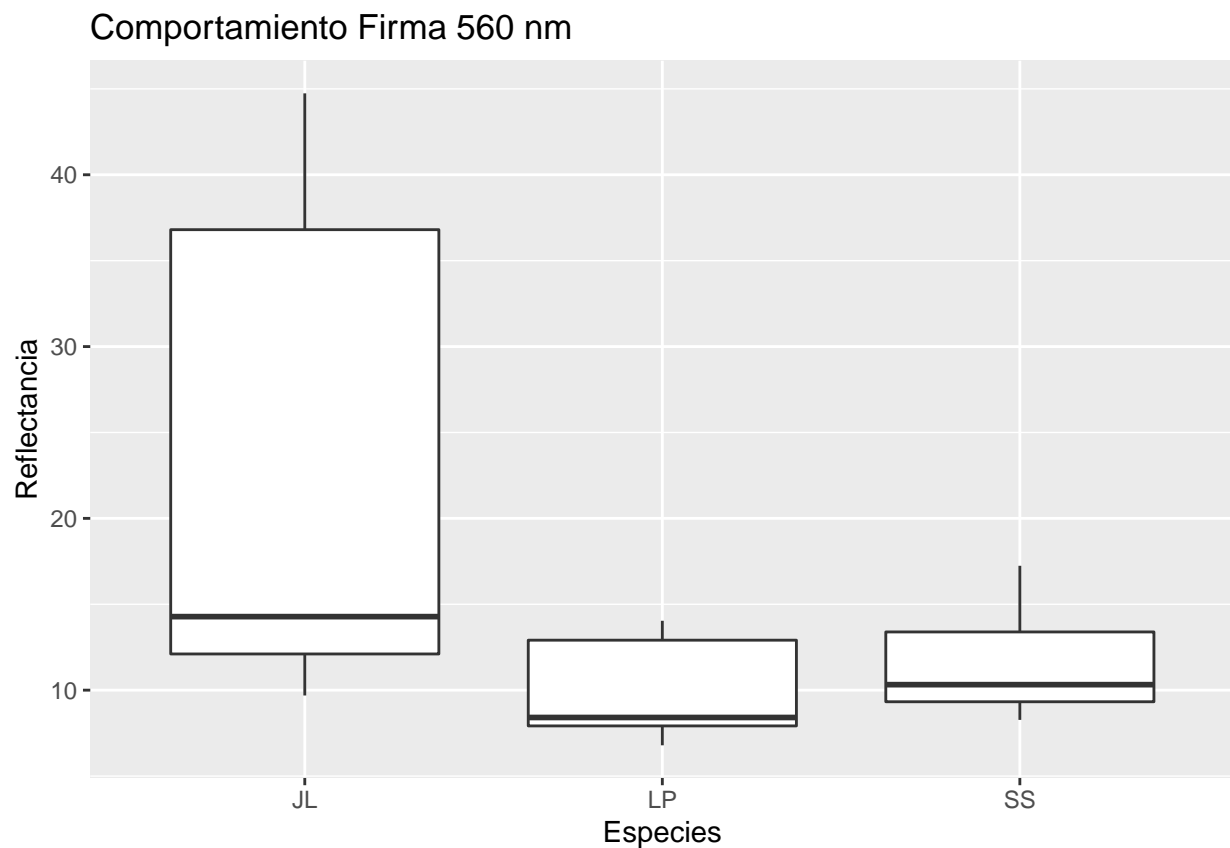
Concluyendo que no es posible asumir que las matrices de varianza y covarianza por cada una de las especies son iguales.

Punto 8: Outliers univariados

Se realiza un gráfico box-plot para cada uno de los valores de longitud de onda y también fue realizada la prueba de *Dixon* que también hace la identificación de los valores atípicos:

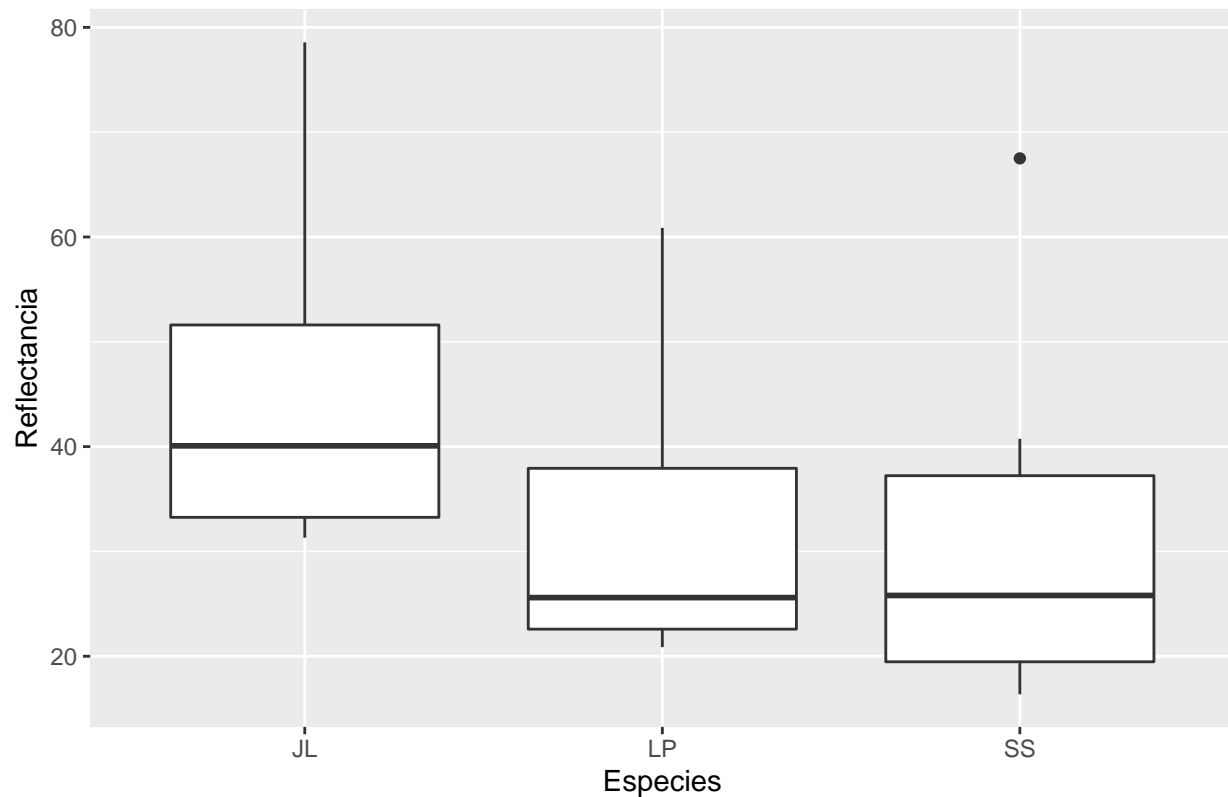
#Gráficas Box-plot

```
ggplot(data=Reflectancia, aes(Species,D_560nm)) + geom_boxplot() + ggtitle("Comportamiento Firma 560 nm")
```



```
ggplot(data=Reflectancia, aes(Species,D_720nm)) + geom_boxplot() + ggtitle("Comportamiento Firma 720 nm")
```


Comportamiento Firma 720 nm



Prueba de Dixon

```
dixon.test(sample(Reflectancia$D_560nm, size=30))
```

```
##
## Dixon test for outliers
##
## data: sample(Reflectancia$D_560nm, size = 30)
## Q = 0.21929, p-value = 0.7239
## alternative hypothesis: highest value 44.74 is an outlier
```

```
dixon.test(sample(Reflectancia$D_720nm, size=30))
```

```
##
## Dixon test for outliers
##
## data: sample(Reflectancia$D_720nm, size = 30)
## Q = 0.12034, p-value = 0.5587
## alternative hypothesis: highest value 78.57 is an outlier
```

Viendo que para la longitud de onda de 560 no existen valores atípicos mientras que para el de 720 nm si para la especie SS.

Punto 9: Outliers multivariados

Para observar los valores atípicos multivariados es realizado a partir del estadístico T^2 y comparado con el percentil que este posee con la distribución chi cuadrado.

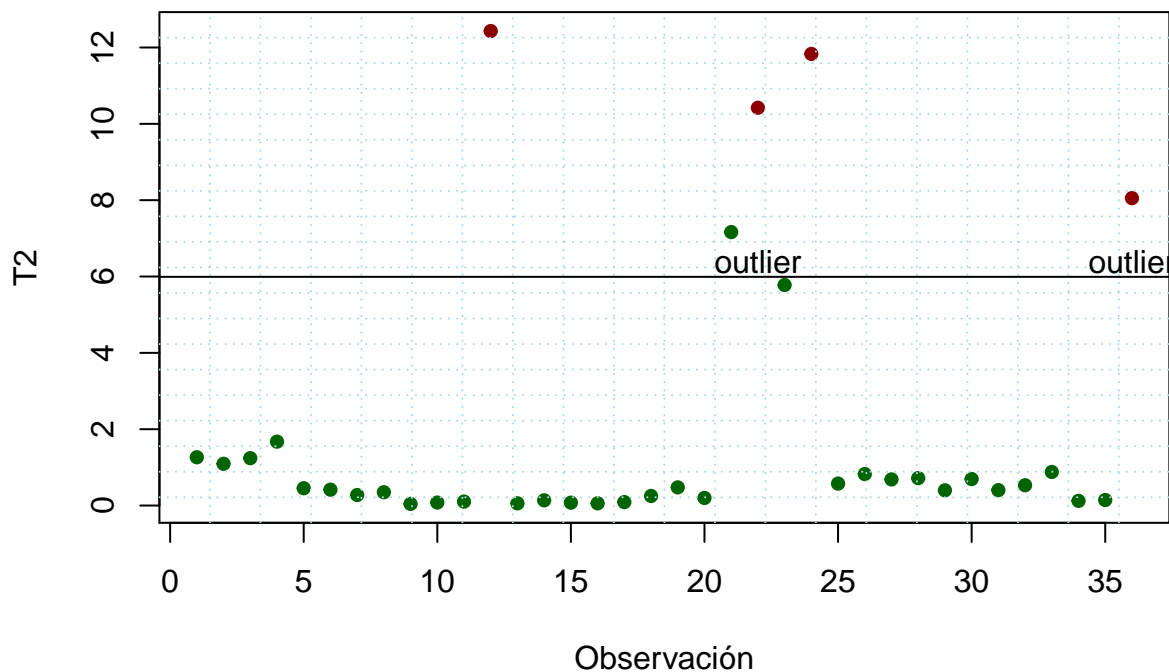
```
vec.medias=apply(Reflectancia[,1:2],2,mean);vec.medias
```

```
## D_560nm D_720nm
## 14.30139 35.78806
```

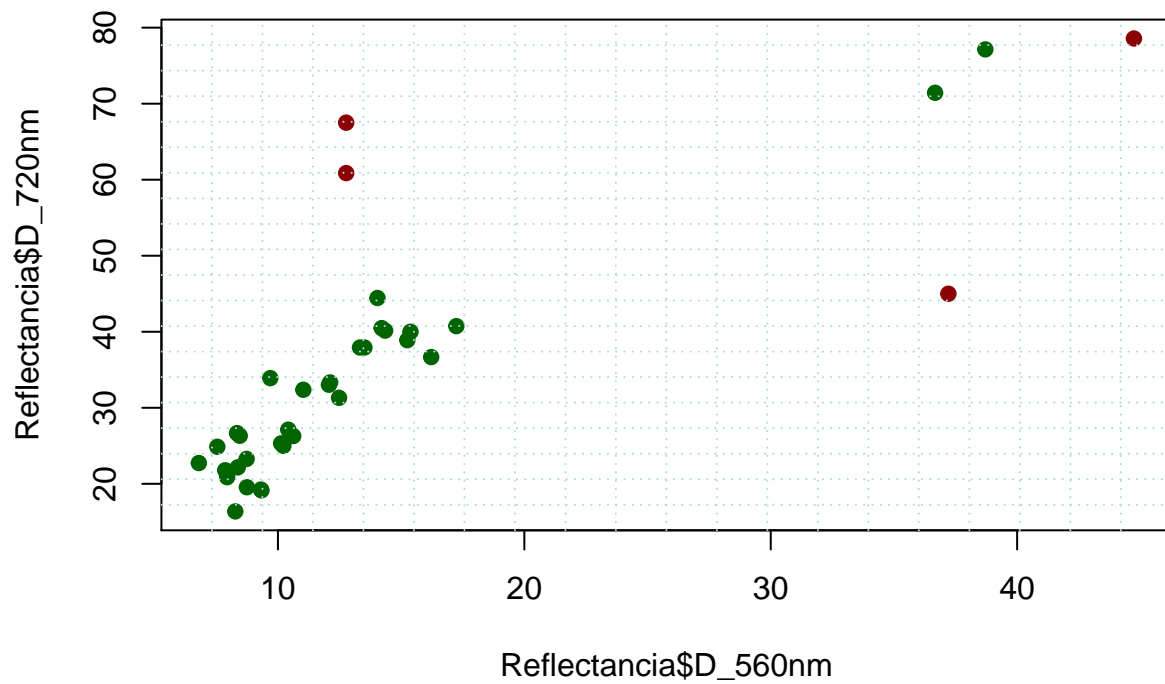
```
T2=c()
for(j in 1:dim(Reflectancia)[1]){
  T2[j]=c((t(t(Reflectancia[j,1:2])-(vec.medias)))%*%solve(var(Reflectancia[,1:2]))%*%as.matrix(t(Refle
})
T2
```

```
## [1] 1.26541020 1.09383567 1.24128875 1.67375718 0.45417223 0.41656430
## [7] 0.27561128 0.34847917 0.04318851 0.07844282 0.10339451 12.43210949
## [13] 0.05790637 0.13704612 0.07555811 0.05807567 0.09033585 0.25119830
## [19] 0.47601009 0.19901130 7.16415107 10.42111846 5.77865326 11.83215879
## [25] 0.57530618 0.82479717 0.68473177 0.71821599 0.39872455 0.69415980
## [31] 0.40375328 0.53175909 0.88021407 0.12249645 0.14489054 8.05347359
```

```
LS=qchisq(0.95,df=dim(Reflectancia)-1)
colores=ifelse(T2>LS,"darkred","darkgreen")
plot(T2,col=colores,pch=19,cex=0.85,xlab="Observación")
grid(20,20,col="lightblue")
abline(h=LS)
etiquetas=which(T2>LS)
text(etiquetas,c(LS)+0.4,"outlier")
```



```
plot(Reflectancia$D_560nm,Reflectancia$D_720nm,pch=19,col=colores)
grid(20,20,col="lightblue")
```



Lo anterior arrojo como resultado que los datos poseen 4 datos multivariados.

Punto 10: Comparación de medias por cada respuesta

Para observar los valores atípicos multivariados es realizado a partir del estadístico T^2 y comparado con el percentil que este posee con la distribución chi cuadrado.