Soil Analysis

Project Effect Fire Dung Beetle

David Vanegas-Alarcón

November 12, 2024

Read data

Read data from "DataGLM.csv".

```
ACC <- read.csv("ACC/DataACC.csv", header = T, sep = ";")
GLMSoil <- read.csv("ACC/DataACC.csv", header = T, sep = ";")

## Convert factor
ACC$FireRegime <- factor(ACC$FireRegime,
                         levels = c("Null", "Low", "High"))

ACC$Cover <- factor(ACC$Cover, levels = c("Secondary Forest",
                                          "Biodiverse pasture",
                                          "Grass Monoculture"))

GLMSoil$FireRegime <- factor(ACC$FireRegime,
                         levels = c("Null", "Low", "High"))

GLMSoil$Cover <- factor(ACC$Cover, levels = c("Secondary Forest",
                                          "Biodiverse pasture",
                                          "Grass Monoculture"))
```

Data exploration and visualization

Empezamos a realizar un análisis exploratorio previo sobre estos.

```
# Data soil
dim(ACC)
```

```
[1] 15 13
```

```
names(ACC)
```

```
 [1] "ID"         "Cover"      "FireRegime" "Abundance"  "Richness"
 [6] "pH"         "Humidity"   "DA"         "N"          "C"
[11] "P_disp"     "P_total"    "Class"
```

Posteriormente vamos a utilizar la función *describe*() del paquete *psych* para conocer las características de nuestros datos.

```
# Describe soil
ACCDescribe <- describe(ACC)
ACCDescribe
```

```
            vars  n   mean    sd median trimmed   mad    min    max  range
ID*            1 15   8.00  4.47   8.00    8.00  5.93   1.00  15.00  14.00
Cover*         2 15   2.00  0.85   2.00    2.00  1.48   1.00   3.00   2.00
FireRegime*    3 15   2.00  0.85   2.00    2.00  1.48   1.00   3.00   2.00
Abundance      4 15  71.00 47.99  70.00   67.15 59.30  13.00 179.00 166.00
Richness       5 15   7.00  1.81   7.00    7.00  1.48   4.00  10.00   6.00
pH             6 15   5.95  0.28   5.91    5.93  0.21   5.52   6.65   1.13
Humidity       7 15  29.86  6.80  28.73   29.53  5.87  20.78  43.30  22.52
DA             8 15   2.03  0.14   2.07    2.04  0.18   1.82   2.23   0.41
N              9 15   0.18  0.13   0.16    0.16  0.10   0.03   0.51   0.48
C             10 15   2.84  1.40   2.55    2.73  1.28   1.17   5.88   4.71
P_disp        11 15   6.11  9.47   2.77    4.32  1.70   0.72  34.77  34.05
P_total       12 15 306.87 79.76 304.39  307.57 82.65 156.69 447.94 291.25
Class*        13 15   2.27  1.83   1.00    2.08  0.00   1.00   6.00   5.00
             skew kurtosis    se
ID*          0.00    -1.44  1.15
Cover*       0.00    -1.69  0.22
FireRegime*  0.00    -1.69  0.22
Abundance    0.56    -0.67 12.39
Richness     0.20    -1.16  0.47
pH           0.78     0.17  0.07
Humidity     0.54    -0.86  1.76
DA          -0.14    -1.35  0.04
N            1.16     0.60  0.03
C            0.84    -0.39  0.36
P_disp       2.12     3.18  2.44
P_total     -0.03    -0.90 20.59
Class*       0.87    -0.99  0.47
```

Como podemos observar la salida de la función nos permite realizar algunas inferencias en relación con los datos presentados. Así mismo, podemos ver que los valores del *sesgo* y *kurtosis* nos sugieren un comportamiento de una distribuciones con sesgos negativos (*DA*) y positivos (*pH*, *K_int*, *Humidity*, *Ca_int*, *CIC*, *MO*, *Abundance* y *Richness*), leptocúrtica (*pH* y *K_int*) y platicúrticas (*Humidity*, *Ca_int*, *CIC*, *DA*, *MO*, *Abundance* y *Richness*).

```
# Función para buscar NA´s en Soil
summary(is.na(ACC))
```

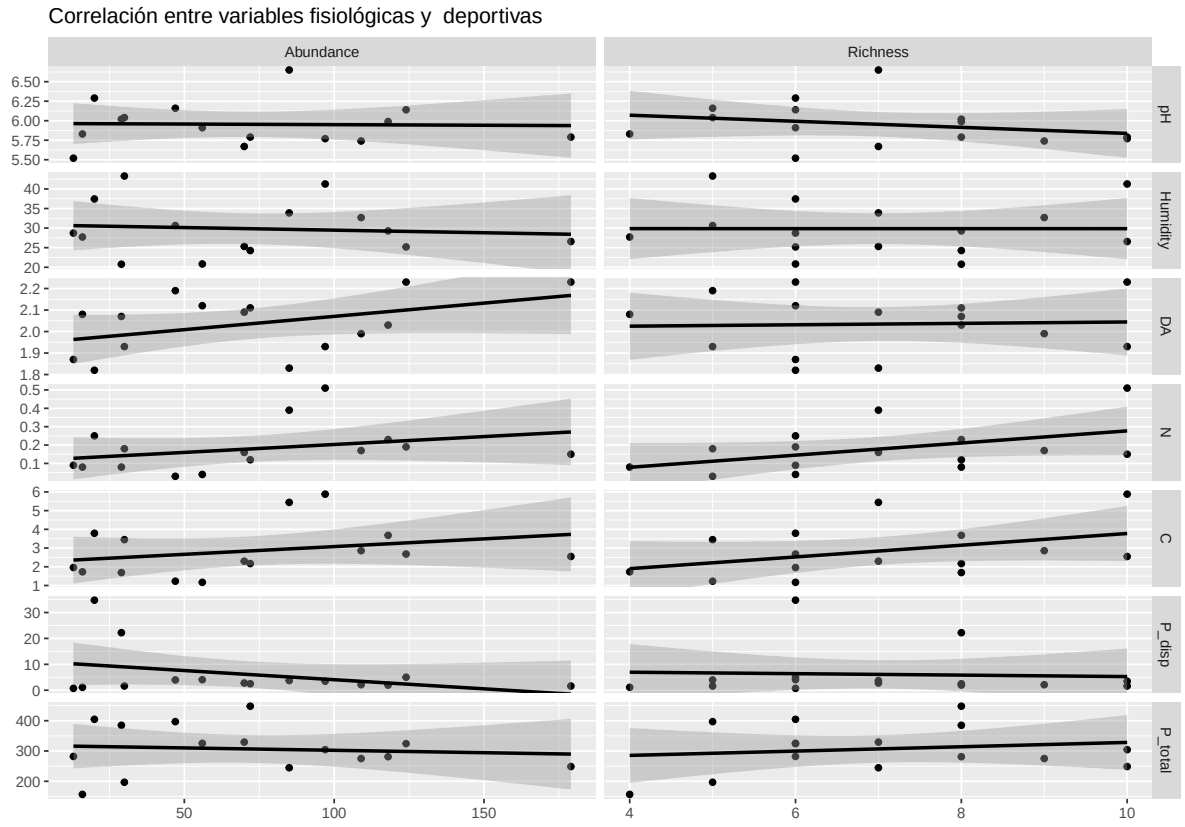
```
     ID            Cover         FireRegime      Abundance
 Mode :logical   Mode :logical   Mode :logical   Mode :logical
 FALSE:15        FALSE:15        FALSE:15        FALSE:15
  Richness           pH           Humidity           DA
 Mode :logical   Mode :logical   Mode :logical   Mode :logical
 FALSE:15        FALSE:15        FALSE:15        FALSE:15
     N               C             P_disp          P_total
```

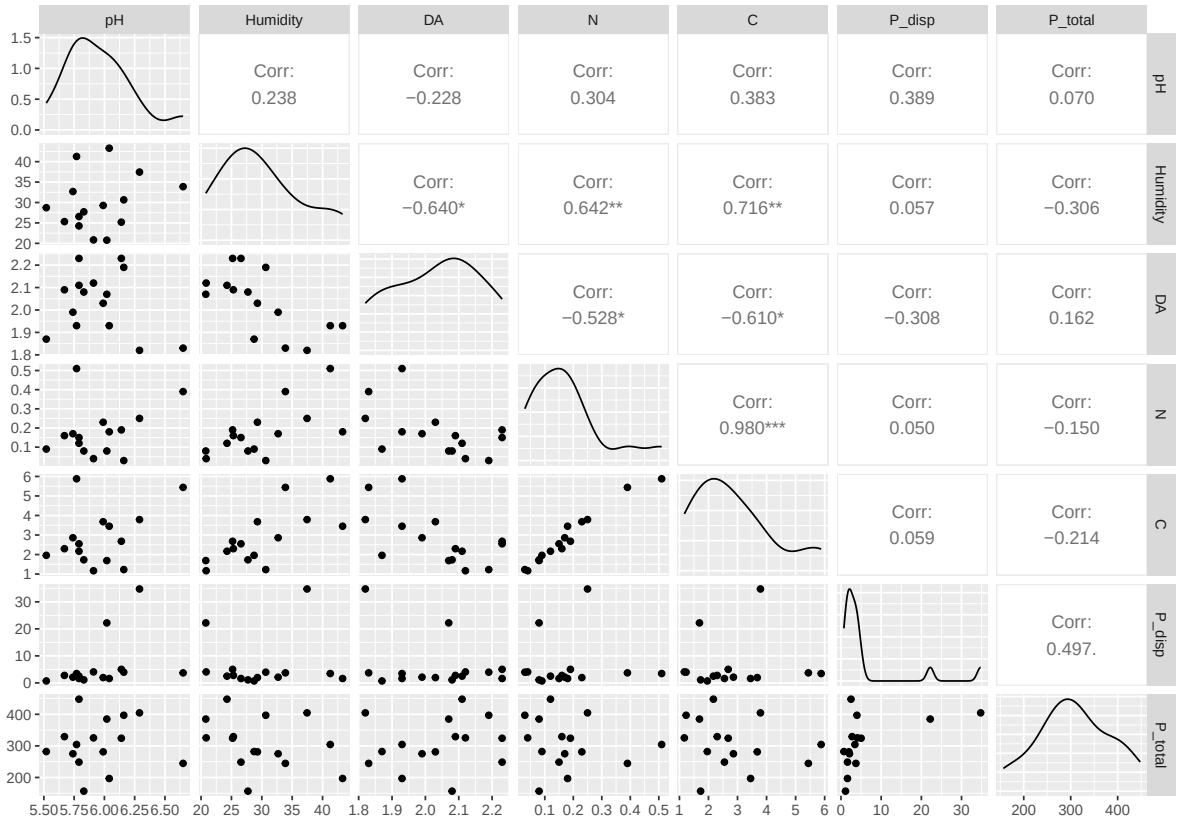
```
Mode :logical   Mode :logical   Mode :logical   Mode :logical
FALSE:15        FALSE:15        FALSE:15        FALSE:15
  Class
Mode :logical
FALSE:15
```

Observando la presencia de datos faltantes o *NA´s*, podemos ver que ninguna de las variables posee este tipo de datos.

Análisis de correspondencia canónica

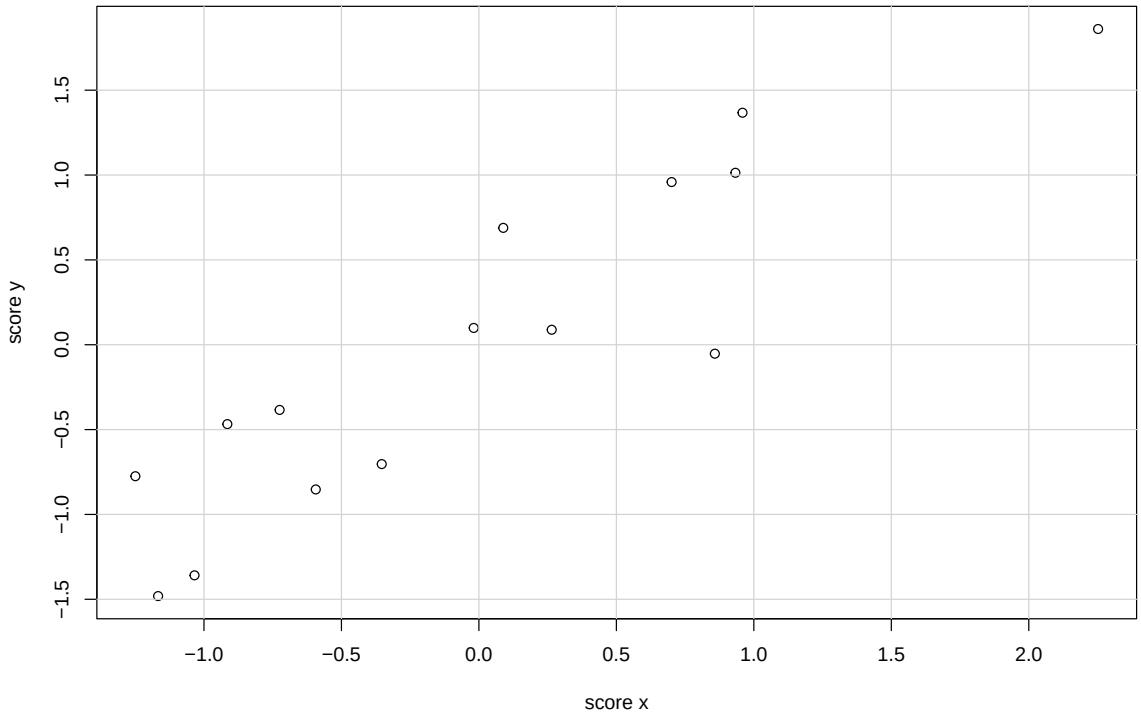
```
diversity <- ACC[, 4:5]
soil <- ACC[, 6:12]
```

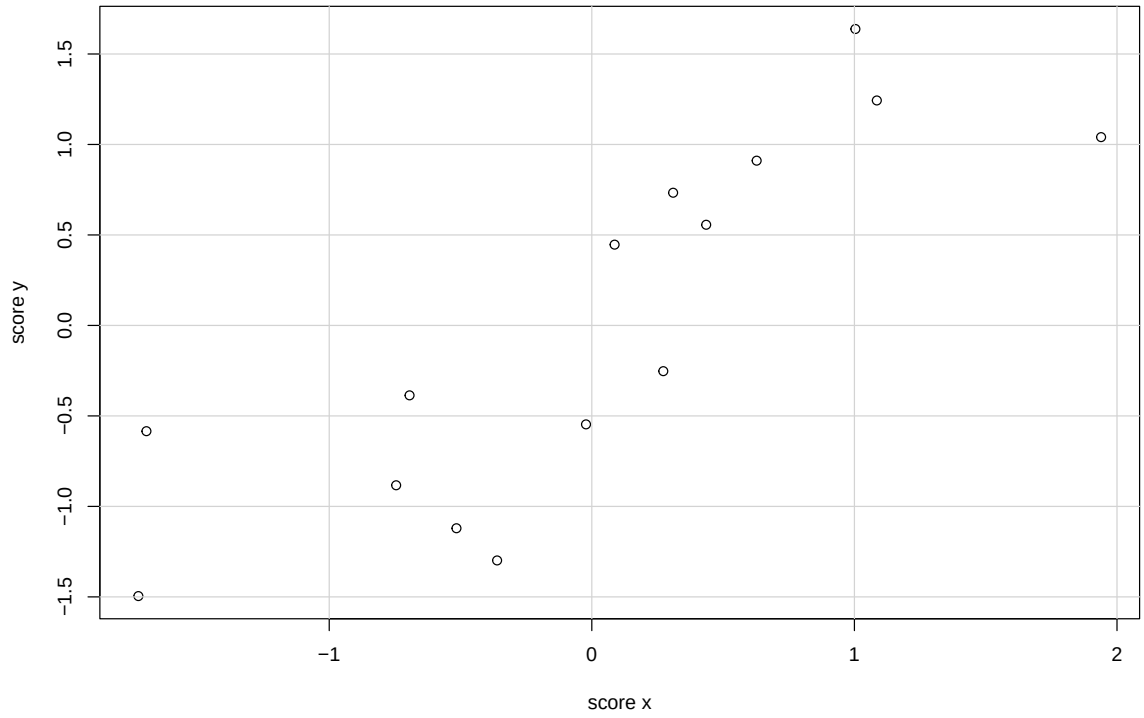





```
cc1 <- CCA::cc(diversity, soil)
cc1$cor
```

```
[1] 0.9084595 0.8271848
```


score y
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
score x

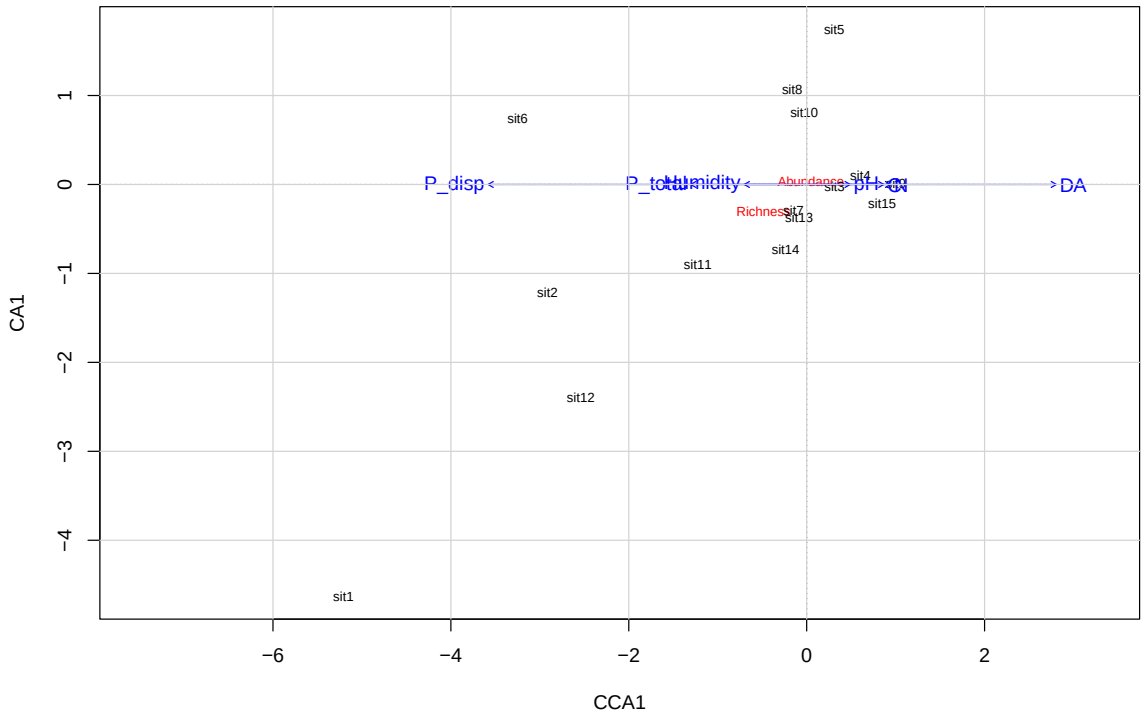


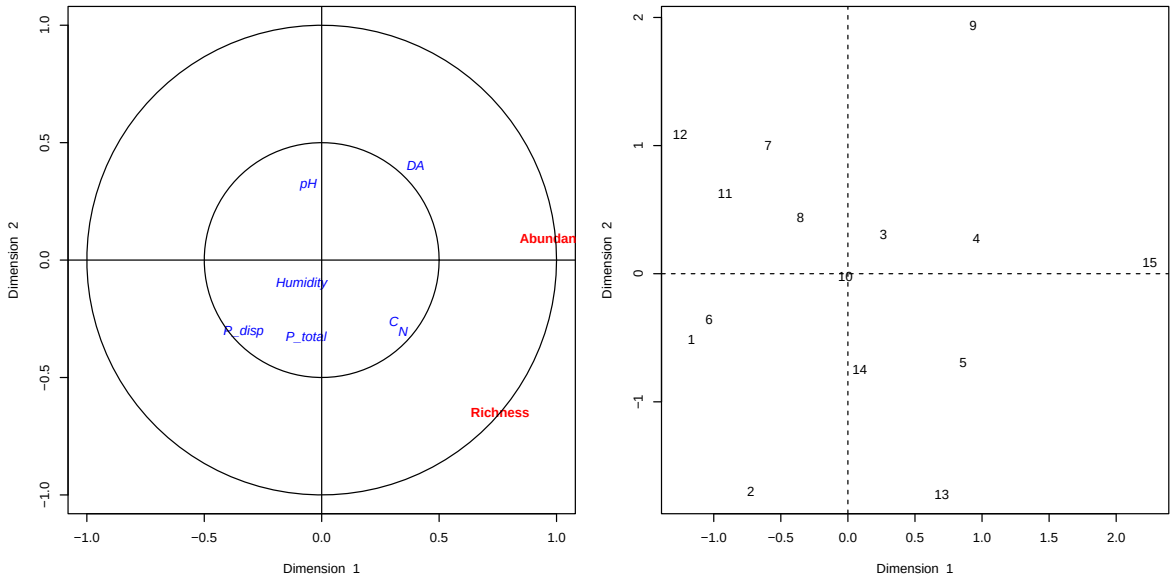

```
cc1
```

```
$cor
[1] 0.9084595 0.8271848

$names
$names$Xnames
[1] "Abundance" "Richness"

$names$Ynames
[1] "pH"       "Humidity" "DA"       "N"        "C"        "P_disp"   "P_total"

$names$ind.names
 [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14" "15"


$xcoef
               [,1]        [,2]
Abundance 0.01891686  0.02211615
Richness  0.06968700 -0.76725924

$ycoef
                [,1]         [,2]
```

```
pH        -1.840660716  3.224441069
Humidity  -0.078041285  0.060007576
DA         8.717213242  2.724497028
N        -20.347685328  5.693723257
C          3.024741230 -1.087289276
P_disp     0.006940182 -0.038436103
P_total    0.001141786 -0.005385234

$scores
$scores$xscores
             [,1]        [,2]
 [1,] -1.16686468 -0.51547763
 [2,] -0.72482097 -1.69613767
 [3,]  0.26483599  0.30962614
 [4,]  0.95877926  0.27219994
 [5,]  0.85821456 -0.69410468
 [6,] -1.03444669 -0.36066456
 [7,] -0.59337857  1.00373082
 [8,] -0.35343985  0.43551695
 [9,]  0.93290639  1.93941535
[10,] -0.01891686 -0.02211615
[11,] -0.91496513  0.62775622
[12,] -1.24948812  1.08538932
[13,]  0.70089928 -1.72675775
[14,]  0.08860386 -0.74514309
[15,]  2.25208152  0.08676679

$scores$yscores
             [,1]       [,2]
 [1,] -1.48111016 -1.1207231
 [2,] -0.38409294 -0.5843804
 [3,]  0.08826513  0.7333503
 [4,]  1.36717156 -0.2526884
 [5,] -0.05277746 -0.3863298
 [6,] -1.35884518 -1.2986388
 [7,] -0.85284869  1.6380667
 [8,] -0.70328838  0.5563195
 [9,]  1.01327152  1.0405451
[10,]  0.09897208 -0.5467652
[11,] -0.46777798  0.9105315
[12,] -0.77482754  1.2431404
[13,]  0.95869337 -1.4955714
[14,]  0.68865913 -0.8832780
[15,]  1.86053552  0.4464217

$scores$corr.X.xscores
              [,1]        [,2]
Abundance 0.9959007  0.09045357
Richness  0.7599327 -0.65000175

$scores$corr.Y.xscores
               [,1]        [,2]
pH       -0.05541963  0.32108946
Humidity -0.08502250 -0.09904535
```

```
DA         0.40074996  0.40164000
N          0.34775220 -0.30370807
C          0.30771102 -0.26264590
P_disp    -0.33164703 -0.30493682
P_total   -0.06572707 -0.32717475

$scores$corr.X.yscores
               [,1]        [,2]
Abundance 0.9047354  0.07482182
Richness  0.6903681 -0.53767160

$scores$corr.Y.yscores
                [,1]       [,2]
pH       -0.06100397  0.3881713
Humidity -0.09358976 -0.1197379
DA        0.44113136  0.4855505
N         0.38279330 -0.3671586
C         0.33871738 -0.3175178
P_disp   -0.36506530 -0.3686441
P_total  -0.07235003 -0.3955280
```





Modelo pH

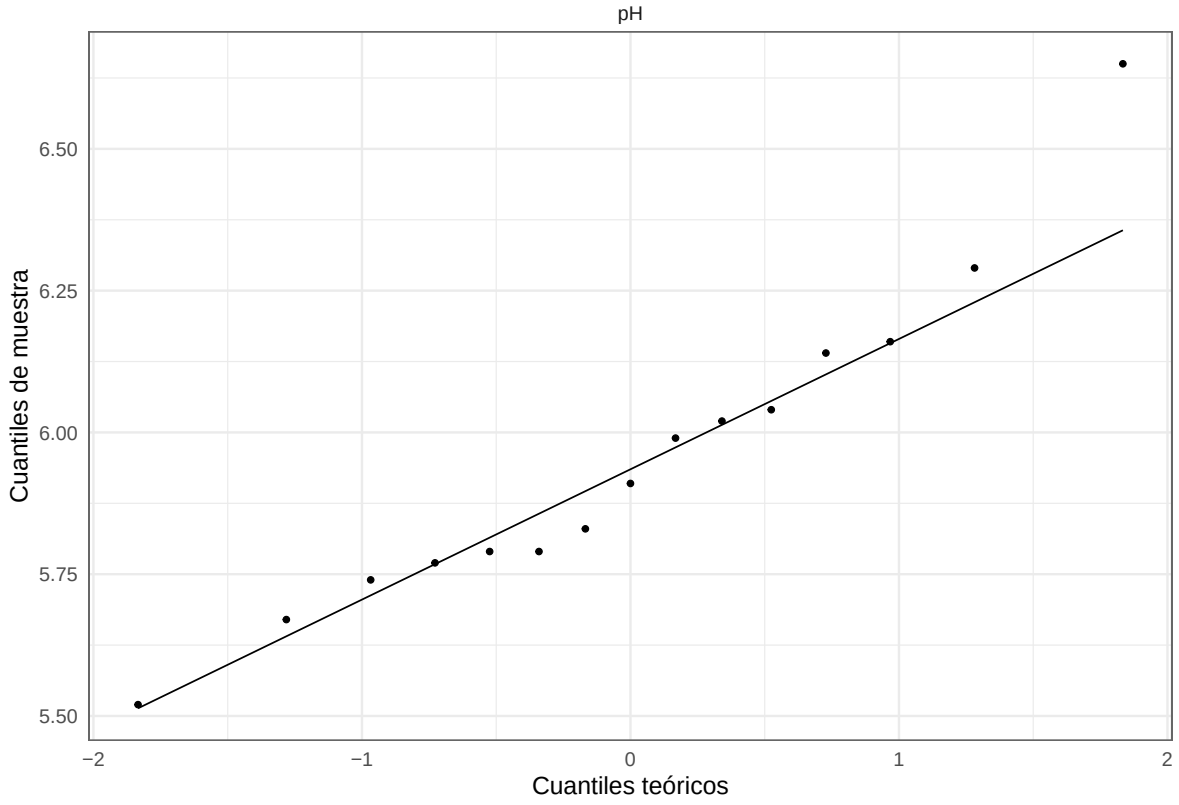
Analizaremos la cantidad de *outliers* para la variable *pH*

```
boxplot.stats(GLMSoil$pH)$out
```

```
[1] 6.65
```

Para apoyar el análisis de datos, realizaremos un gráfico para poder observar la disposición de la distribución de los datos.

Los gráficos cuantil cuantil *(quantile-quantile (qq))* son una ayuda para explorar si un conjunto de datos o muestra proviene de una población con cierta distribución, en nuestro caso, exploraremos algún comportamiento en relación con su sesgo y su distribución.



Relación entre el pH y el Régimen de quemas

Modelo nulo

```
nullpHDB <- glm(pH ~ 1, family = gaussian(),
                data = GLMSoil)
summary(nullpHDB)
```

```
Call:
glm(formula = pH ~ 1, family = gaussian(), data = GLMSoil)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.95400    0.07261      82   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.07908286)

    Null deviance: 1.1072  on 14  degrees of freedom
```

```
Residual deviance: 1.1072  on 14  degrees of freedom
AIC: 7.4744

Number of Fisher Scoring iterations: 2
```

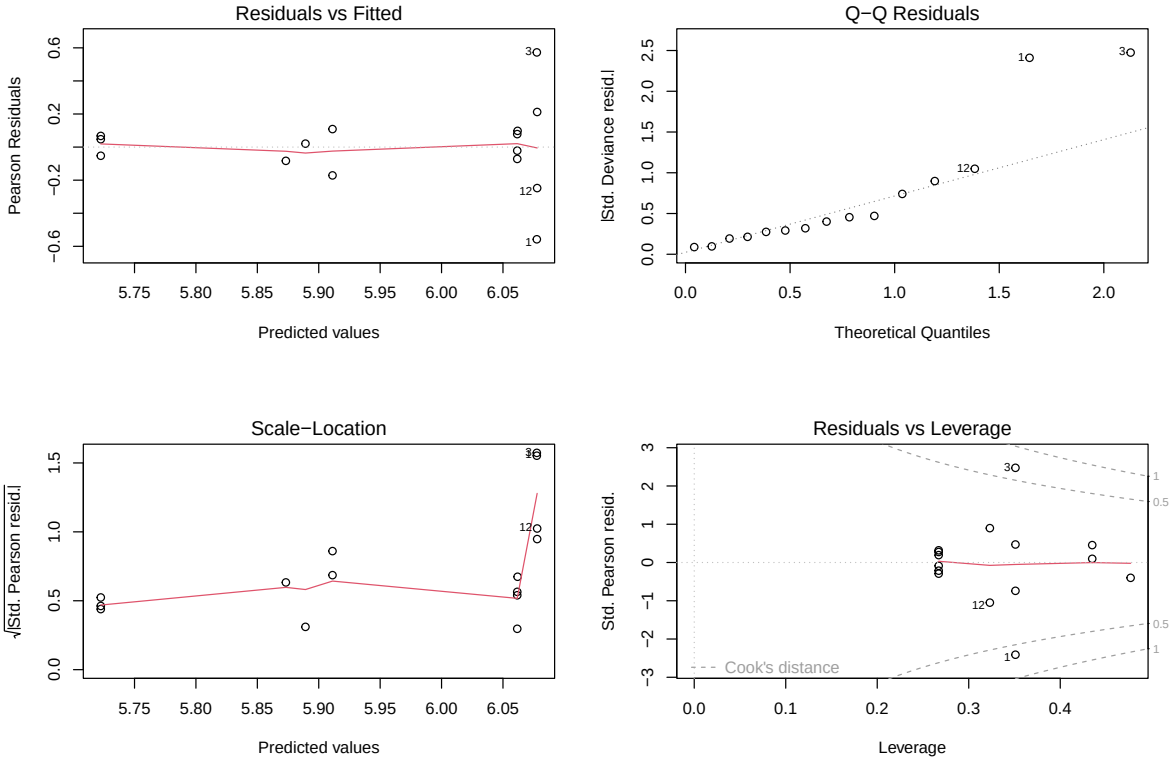
Modelo *pH ~ FireRegimen + Cover*

Luego procederemos a realizar un modelo entre las variables *pH* y *FireRegimen + Cover*. Esto con el fin de poder conocer una posible asociación entre las variables.

```r
pHDB <- glm(pH ~ FireRegime + Cover, family = gaussian(),
            data = GLMSoil)
summary(pHDB)
```

```
Call:
glm(formula = pH ~ FireRegime + Cover, family = gaussian(), data = GLMSoil)

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               5.7225     0.1484  38.561 3.28e-12 ***
FireRegimeLow             0.1887     0.1860   1.014    0.334
FireRegimeHigh            0.1884     0.2148   0.877    0.401
CoverBiodiverse pasture   0.1507     0.2148   0.702    0.499
CoverGrass Monoculture    0.1666     0.2009   0.830    0.426
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.08244341)

    Null deviance: 1.10716  on 14  degrees of freedom
Residual deviance: 0.82443  on 10  degrees of freedom
AIC: 11.052

Number of Fisher Scoring iterations: 2
```

Sin embargo, vamos a validar el modelo de regresión a partir de los gráficos de los residuales.




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

```
    Shapiro-Wilk normality test

data:  pHDB$residuals
W = 0.92193, p-value = 0.2061
```

```
result <-  fligner.test(pH ~ interaction(FireRegime, Cover), data = GLMSoil)

print(result)
```

```
    Fligner-Killeen test of homogeneity of variances

data:  pH by interaction(FireRegime, Cover)
Fligner-Killeen:med chi-squared = 13.02, df = 7, p-value = 0.07162
```

Interpretación de los Resultados

Intercepto (β_0):

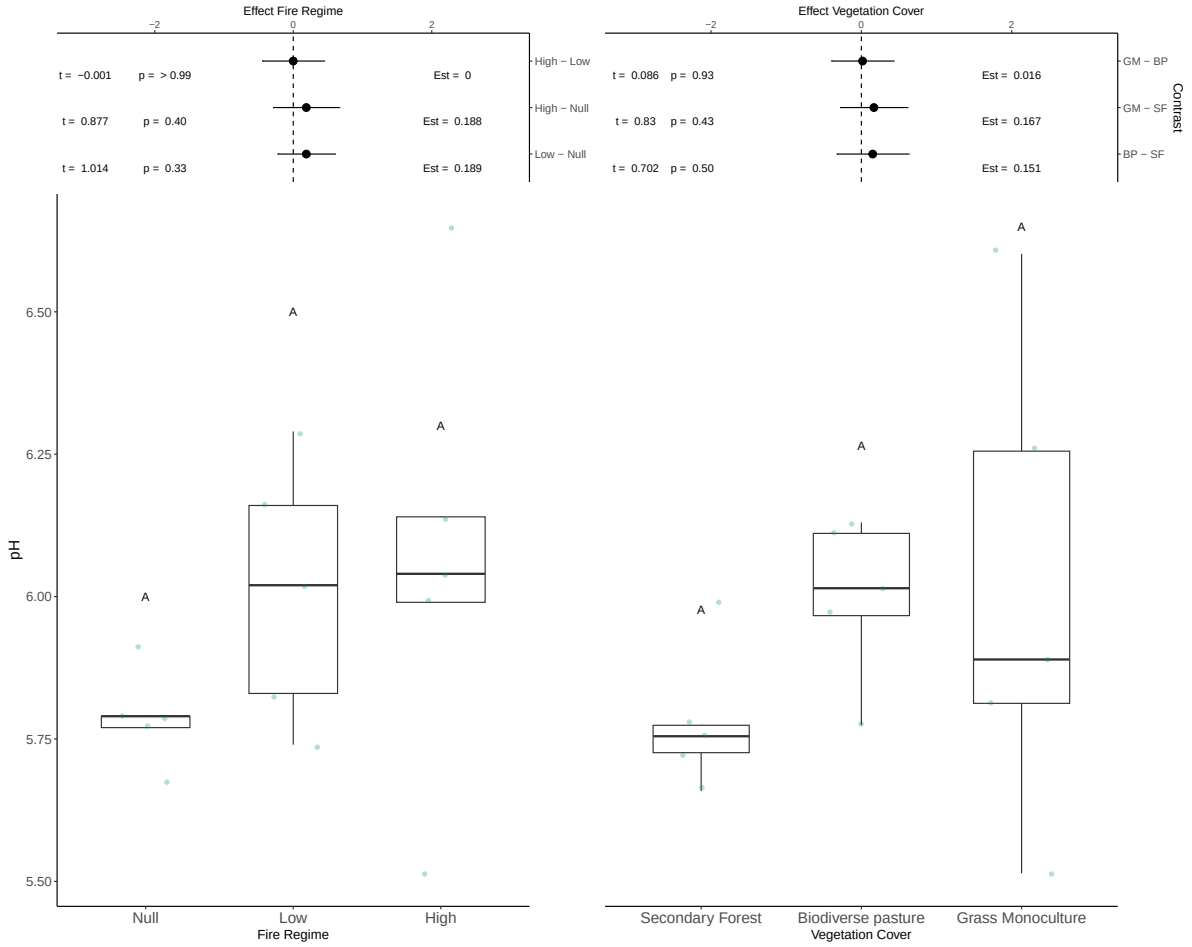
FireRegimenLow (β_1):

FireRegimenHigh (β_2):

CoverBiodiverse pasture (β_3):

FireRegimenLow (β_4):

Visualización



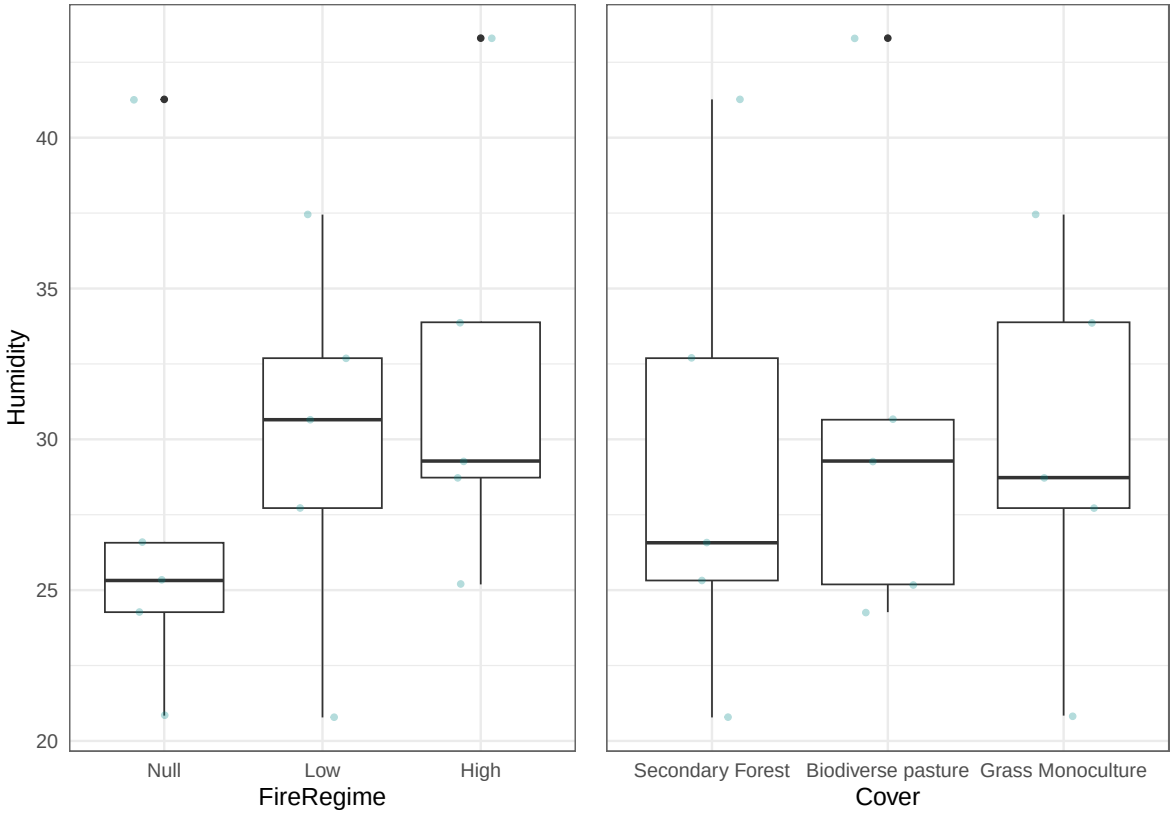
Modelo Humedad

Analizaremos la cantidad de *outliers* para la variable humedad

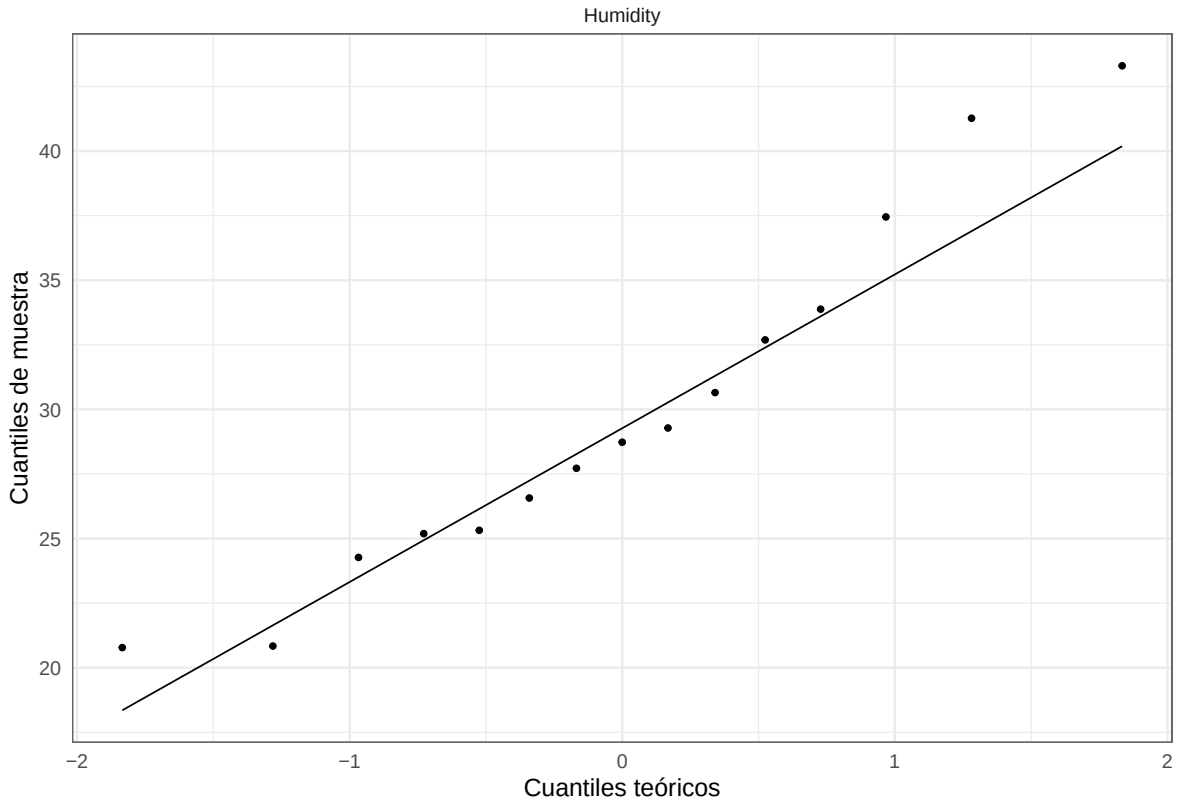
```
boxplot.stats(GLMSoil$Humidity)$out
```

```
numeric(0)
```

Para apoyar el análisis de datos, realizaremos un gráfico para poder observar la disposición de la distribución de los datos.



Los gráficos cuantil cuantil *(quantile-quantile (qq))* son una ayuda para explorar si un conjunto de datos o muestra proviene de una población con cierta distribución, en nuestro caso, exploraremos algún comportamiento en relación con su sesgo y su distribución.



Relación entre el Humidity y el Régimen de quemas

Modelo nulo

```
nullHumidityDB <- glm(Humidity ~ 1, family = gaussian(),
             data = GLMSoil)
summary(nullHumidityDB)
```

```
Call:
glm(formula = Humidity ~ 1, family = gaussian(), data = GLMSoil)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   29.863      1.756      17 9.58e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 46.26539)

    Null deviance: 647.72  on 14  degrees of freedom
```

```
Residual deviance: 647.72  on 14  degrees of freedom
AIC: 103.05

Number of Fisher Scoring iterations: 2
```

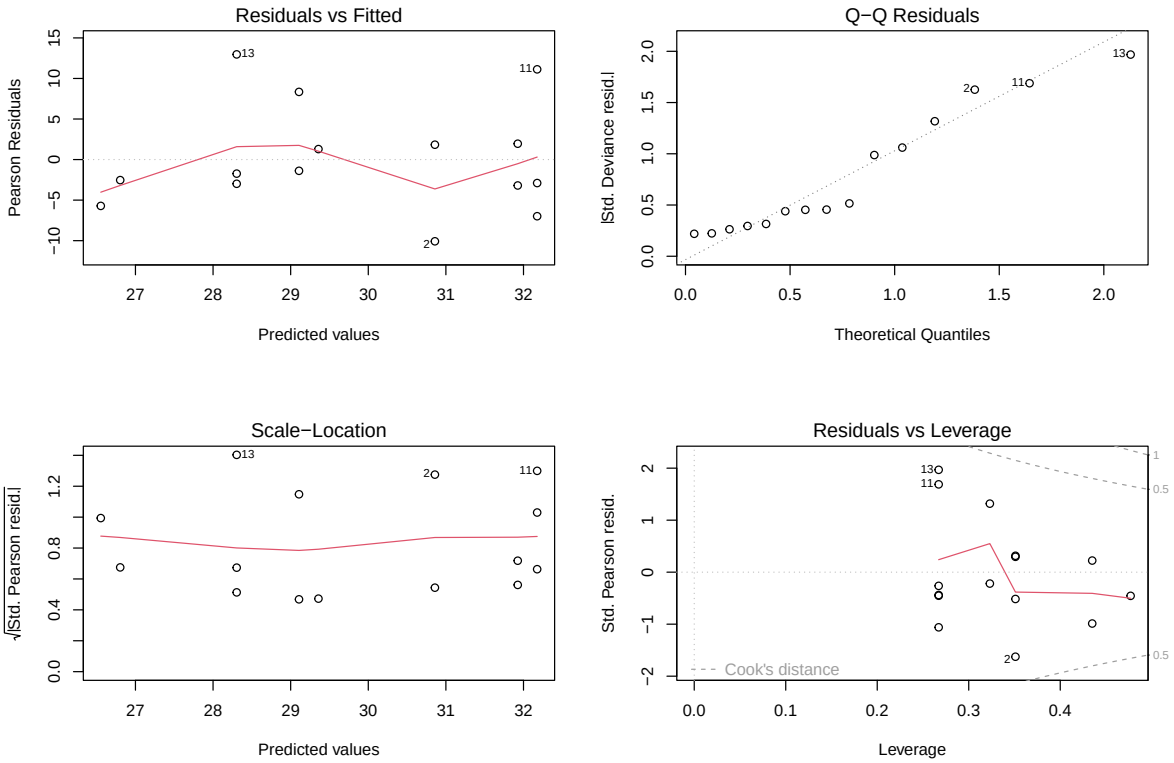
Modelo *Humidity ~ FireRegimen + Cover*

Luego procederemos a realizar un modelo entre las variables *Humidity* y *FireRegimen + Cover*. Esto con el fin de poder conocer una posible asociación entre las variables.

```r
HumidityDB <- glm(Humidity ~ FireRegime + Cover, family = gaussian(),
             data = GLMSoil)
summary(HumidityDB)
```

```
Call:
glm(formula = Humidity ~ FireRegime + Cover, family = gaussian(),
    data = GLMSoil)

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              28.304      3.977   7.116 3.23e-05 ***
FireRegimeLow             2.554      4.985   0.512    0.619
FireRegimeHigh            5.372      5.756   0.933    0.373
CoverBiodiverse pasture  -1.501      5.756  -0.261    0.800
CoverGrass Monoculture   -1.751      5.384  -0.325    0.752
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 59.21731)

    Null deviance: 647.72  on 14  degrees of freedom
Residual deviance: 592.17  on 10  degrees of freedom
AIC: 109.7

Number of Fisher Scoring iterations: 2
```

Sin embargo, vamos a validar el modelo de regresión a partir de los gráficos de los residuales.




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

```
    Shapiro-Wilk normality test

data:  HumidityDB$residuals
W = 0.92454, p-value = 0.2258
```

```
result <-  fligner.test(Humidity ~ interaction(FireRegime, Cover), data = GLMSoil)

print(result)
```

```
    Fligner-Killeen test of homogeneity of variances

data:  Humidity by interaction(FireRegime, Cover)
Fligner-Killeen:med chi-squared = 5.2303, df = 7, p-value = 0.6319
```

Interpretación de los Resultados

Intercepto (β_0):

FireRegimenLow (β_1):

FireRegimenHigh (β_2):

CoverBiodiverse pasture (β_3):

FireRegimenLow (β_4):

Visualización