Tarea final - Estadística Aplicada a la Forestería II

Christian Richard Alberto Vásquez Velasco

2021-09-12

# 1. Medidas repetidas. DBCA con medidas repetidas: AUDPC evaluado en el tiempo

## Reconocimiento de la información:

* VARIABLE DEPENDIENTE (y): AUDPC.
* VARIABLE INDEPENDIENTE (Var): Variedad (Tratamiento). Niveles: Cruza148, Musuq, Pimpernell y Tomasa.
* VARIABLE INDEPENDIENTE (Block): Bloque (1, 2, 3 y 4)
* VARIABLE INDEPENDIENTE (tiempo): Evaluaciones realizadas (44, 51, 58, 65, 72, 79, 86 y 93 días después de la inoculación).

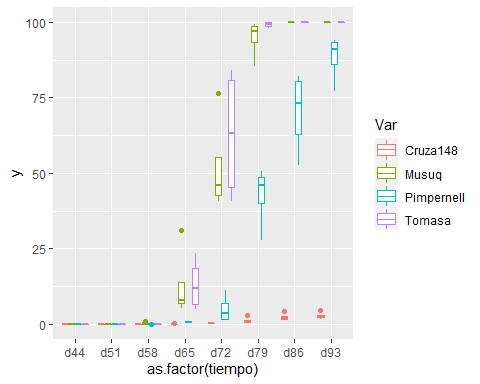
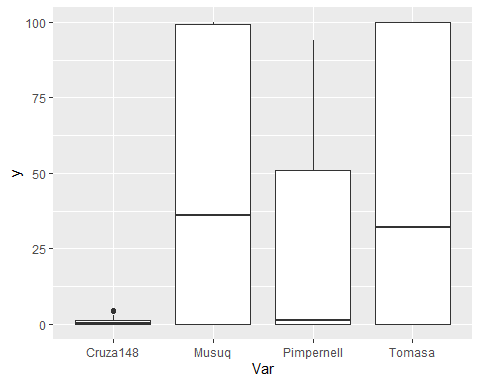
Según la estructura del objeto, el bloque y la Variedad deben ser factores.

`summarise()` has grouped output by 'Block'. You can override using the `.groups` argument.

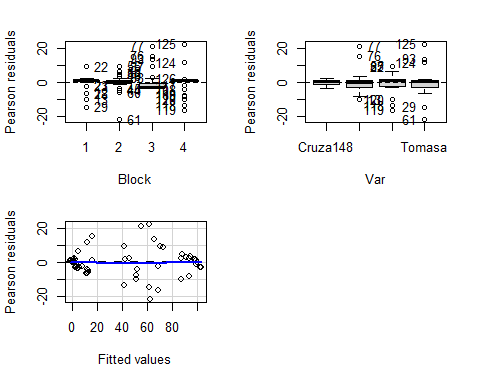
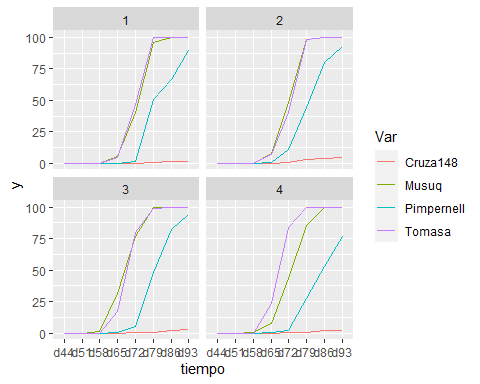
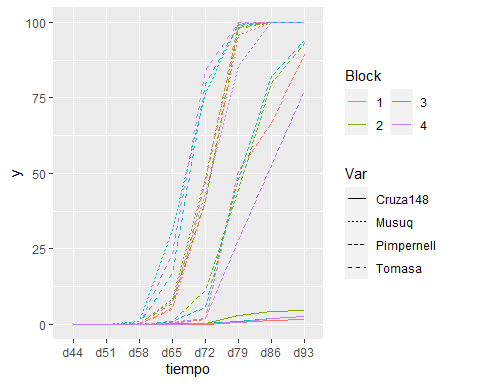
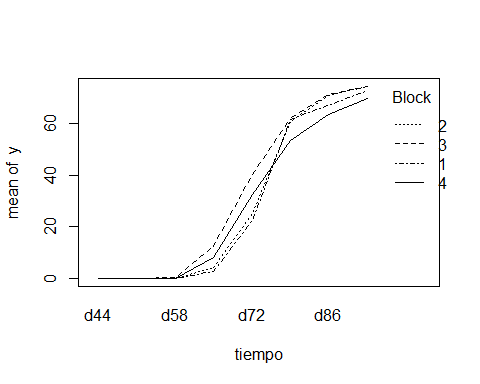
grouped\_df [128 x 4] (S3: grouped\_df/tbl\_df/tbl/data.frame)  
 $ Block : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...  
 $ Var : Factor w/ 4 levels "Cruza148","Musuq",..: 1 1 1 1 1 1 1 1 2 2 ...  
 $ tiempo: chr [1:128] "d44" "d51" "d58" "d65" ...  
 $ y : num [1:128] 0 0 0 0 0 0.65 1.15 1.6 0 0 ...  
 - attr(\*, "groups")= tibble [4 x 2] (S3: tbl\_df/tbl/data.frame)  
 ..$ Block: Factor w/ 4 levels "1","2","3","4": 1 2 3 4  
 ..$ .rows: list<int> [1:4]   
 .. ..$ : int [1:32] 1 2 3 4 5 6 7 8 9 10 ...  
 .. ..$ : int [1:32] 33 34 35 36 37 38 39 40 41 42 ...  
 .. ..$ : int [1:32] 65 66 67 68 69 70 71 72 73 74 ...  
 .. ..$ : int [1:32] 97 98 99 100 101 102 103 104 105 106 ...  
 .. ..@ ptype: int(0)   
 ..- attr(\*, ".drop")= logi TRUE

## Análisis exploratorio

### Normalidad y Homogeneidad de varianzas



### Interacción bloque a bloque



Test stat Pr(>|Test stat|)  
Block   
Var   
Tukey test 1.5485 0.1215

Test Pvalue   
1.5484755 0.1215079

## Analisis del experimento

### Mediante el analisis de variancia multivarial (MANOVA)\*\*

La matriz Y es la siguiente:

d58 d65 d72 d79 d86 d93  
 [1,] 0.00 0.00 0.00 0.65 1.15 1.60  
 [2,] 0.10 5.20 40.75 95.75 100.00 100.00  
 [3,] 0.05 0.30 1.60 50.50 66.50 89.25  
 [4,] 0.00 5.05 47.00 99.75 100.00 100.00  
 [5,] 0.00 0.30 0.45 2.90 4.30 4.50  
 [6,] 0.10 8.25 48.25 98.25 100.00 100.00  
 [7,] 0.05 1.00 11.25 44.25 79.75 93.00  
 [8,] 0.15 7.00 40.75 98.00 100.00 100.00  
 [9,] 0.00 0.05 0.20 0.90 1.95 2.60  
[10,] 0.95 31.25 76.50 99.25 100.00 100.00  
[11,] 0.00 0.90 5.50 48.00 82.25 94.00  
[12,] 0.10 17.00 79.50 99.00 100.00 100.00  
[13,] 0.00 0.05 0.30 0.75 2.05 2.45  
[14,] 0.15 7.50 43.50 85.50 100.00 100.00  
[15,] 0.05 0.65 1.80 27.75 52.50 77.00  
[16,] 0.05 23.50 84.00 100.00 100.00 100.00

#### Modelo de analisis en R

manova(Y ~ variedad, data=A)

El analisis de la varincia se resume mediante 4 metodos: Wilks, Pillai, Roy y Hotelling-Lawley.

El estadístico de Wilks es el más popular en la literatura, pero Hand y Taylor (1987) recomiendan el estadístico predeterminado de Pillai-Bartlett y el mas severo es el metodo de Roy.

##### Método de Wilks

Df Wilks approx F num Df den Df Pr(>F)   
Var 3 0.000005 50.038 18 11.799 0.00000001834 \*\*\*  
Block 3 0.119122 0.735 18 11.799 0.73   
Residuals 9   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### Método de Pillai

Df Pillai approx F num Df den Df Pr(>F)   
Var 3 2.4621 4.5775 18 18 0.001151 \*\*  
Block 3 1.3198 0.7855 18 18 0.693014   
Residuals 9   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### Método de Roy

Df Roy approx F num Df den Df Pr(>F)   
Var 3 891.58 891.58 6 6 0.00000001404 \*\*\*  
Block 3 3.01 3.01 6 6 0.1029   
Residuals 9   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Otra alternativa para el analisis de variancia

Se proporcionan varias estadísticas de prueba para modelos lineales multivariados producidos por lm o manova en el paquete car con la funcion Anova()

#### Modelo de analisis en R

Anova(m1, idata=data.frame(rfactor), idesign=~rfactor, type="III")

Univariate Type III Repeated-Measures ANOVA Assuming Sphericity  
  
 Sum Sq num Df Error SS den Df F value Pr(>F)   
(Intercept) 7 1 1077.1 9 0.0571 0.8165   
Var 58805 3 1077.1 9 163.7921 0.00000003604 \*\*\*  
Block 491 3 1077.1 9 1.3671 0.3140   
rfactor 207 5 1724.3 45 1.0800 0.3843   
Var:rfactor 32037 15 1724.3 45 55.7369 < 0.00000000000000022 \*\*\*  
Block:rfactor 915 15 1724.3 45 1.5923 0.1144   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
  
Mauchly Tests for Sphericity  
  
 Test statistic p-value  
rfactor 0.0012744 0.000033787  
Var:rfactor 0.0012744 0.000033787  
Block:rfactor 0.0012744 0.000033787  
  
  
Greenhouse-Geisser and Huynh-Feldt Corrections  
 for Departure from Sphericity  
  
 GG eps Pr(>F[GG])   
rfactor 0.33026 0.3524   
Var:rfactor 0.33026 0.000000004361 \*\*\*  
Block:rfactor 0.33026 0.2229   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
 HF eps Pr(>F[HF])  
rfactor 0.394985 0.360100947485849  
Var:rfactor 0.394985 0.000000000153393  
Block:rfactor 0.394985 0.207619265827504

### Analisis mediante un modelo mixto\*\*

#### Ajuste de modelos o análisis estadístico\*

#### Modelo utilizado:

lme(plantas ~ variedad\*tiempo, random = ~1|rep/variedad,data=B)

##### Exploración de la estructura fija

Modelo 1

Linear mixed-effects model fit by maximum likelihood  
 Data: A\_longer   
 AIC BIC logLik  
 1073.093 1101.613 -526.5463  
  
Random effects:  
 Formula: ~1 | Block  
 (Intercept) Residual  
StdDev: 0.0006549179 14.80114  
  
Fixed effects: y ~ Var \* tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -3.27117 11.853254 117 -0.275973 0.7831  
VarMusuq -129.03108 16.763033 117 -7.697359 0.0000  
VarPimpernell -99.18087 16.763033 117 -5.916642 0.0000  
VarTomasa -130.04896 16.763033 117 -7.758080 0.0000  
tiempo 0.06014 0.168483 117 0.356951 0.7218  
VarMusuq:tiempo 2.52878 0.238271 117 10.613040 0.0000  
VarPimpernell:tiempo 1.81320 0.238271 117 7.609816 0.0000  
VarTomasa:tiempo 2.57083 0.238271 117 10.789516 0.0000  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms tiempo VrMsq: VrPmp:  
VarMusuq -0.707   
VarPimpernell -0.707 0.500   
VarTomasa -0.707 0.500 0.500   
tiempo -0.974 0.688 0.688 0.688   
VarMusuq:tiempo 0.688 -0.974 -0.487 -0.487 -0.707   
VarPimpernell:tiempo 0.688 -0.487 -0.974 -0.487 -0.707 0.500   
VarTomasa:tiempo 0.688 -0.487 -0.487 -0.974 -0.707 0.500 0.500  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-2.20544819 -0.58294645 -0.01465861 0.56057647 1.88431592   
  
Number of Observations: 128  
Number of Groups: 4

Modelo 2

Linear mixed-effects model fit by maximum likelihood  
 Data: A\_longer   
 AIC BIC logLik  
 1172.334 1192.299 -579.1672  
  
Random effects:  
 Formula: ~1 | Block  
 (Intercept) Residual  
StdDev: 0.0008097511 22.32718  
  
Fixed effects: y ~ Var + tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -121.65320 9.493998 120 -12.813696 0  
VarMusuq 44.19063 5.694116 120 7.760752 0  
VarPimpernell 25.02344 5.694116 120 4.394613 0  
VarTomasa 46.05313 5.694116 120 8.087844 0  
tiempo 1.78835 0.125517 120 14.247795 0  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms  
VarMusuq -0.300   
VarPimpernell -0.300 0.500   
VarTomasa -0.300 0.500 0.500   
tiempo -0.906 0.000 0.000 0.000  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-1.92872038 -0.83475235 -0.05484837 0.82177194 1.92438150   
  
Number of Observations: 128  
Number of Groups: 4

Prueba de hipótesis para la selección de la mejor estructura fija

Model df AIC BIC logLik Test L.Ratio p-value  
data.rm1.lme 1 10 1073.093 1101.613 -526.5463   
data.rm1.lme1 2 7 1172.334 1192.299 -579.1672 1 vs 2 105.2417 <.0001

Según el AIC, se elige el modelo con la estructura fija que considera la interacción de la Variedad y el Tiempo.

##### Exploración de la estructura aleatoria.

Linear mixed-effects model fit by REML  
 Data: A\_longer   
 AIC BIC logLik  
 1064.945 1092.82 -522.4725  
  
Random effects:  
 Formula: ~1 | Block  
 (Intercept) Residual  
StdDev: 0.0008972843 15.28656  
  
Fixed effects: y ~ Var \* tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -3.27117 11.853254 117 -0.275973 0.7831  
VarMusuq -129.03108 16.763033 117 -7.697359 0.0000  
VarPimpernell -99.18087 16.763033 117 -5.916642 0.0000  
VarTomasa -130.04896 16.763033 117 -7.758080 0.0000  
tiempo 0.06014 0.168483 117 0.356951 0.7218  
VarMusuq:tiempo 2.52878 0.238271 117 10.613040 0.0000  
VarPimpernell:tiempo 1.81320 0.238271 117 7.609816 0.0000  
VarTomasa:tiempo 2.57083 0.238271 117 10.789516 0.0000  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms tiempo VrMsq: VrPmp:  
VarMusuq -0.707   
VarPimpernell -0.707 0.500   
VarTomasa -0.707 0.500 0.500   
tiempo -0.974 0.688 0.688 0.688   
VarMusuq:tiempo 0.688 -0.974 -0.487 -0.487 -0.707   
VarPimpernell:tiempo 0.688 -0.487 -0.974 -0.487 -0.707 0.500   
VarTomasa:tiempo 0.688 -0.487 -0.487 -0.974 -0.707 0.500 0.500  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-2.13541603 -0.56443548 -0.01419314 0.54277583 1.82448105   
  
Number of Observations: 128  
Number of Groups: 4

Linear mixed-effects model fit by REML  
 Data: A\_longer   
 AIC BIC logLik  
 1068.945 1102.395 -522.4725  
  
Random effects:  
 Formula: ~tiempo | Block  
 Structure: General positive-definite, Log-Cholesky parametrization  
 StdDev Corr   
(Intercept) 0.0009423901333873 (Intr)  
tiempo 0.0000000006047612 0   
Residual 15.2865550221716529   
  
Fixed effects: y ~ Var \* tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -3.27117 11.853254 117 -0.275973 0.7831  
VarMusuq -129.03108 16.763033 117 -7.697359 0.0000  
VarPimpernell -99.18087 16.763033 117 -5.916642 0.0000  
VarTomasa -130.04896 16.763033 117 -7.758080 0.0000  
tiempo 0.06014 0.168483 117 0.356951 0.7218  
VarMusuq:tiempo 2.52878 0.238271 117 10.613040 0.0000  
VarPimpernell:tiempo 1.81320 0.238271 117 7.609816 0.0000  
VarTomasa:tiempo 2.57083 0.238271 117 10.789516 0.0000  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms tiempo VrMsq: VrPmp:  
VarMusuq -0.707   
VarPimpernell -0.707 0.500   
VarTomasa -0.707 0.500 0.500   
tiempo -0.974 0.688 0.688 0.688   
VarMusuq:tiempo 0.688 -0.974 -0.487 -0.487 -0.707   
VarPimpernell:tiempo 0.688 -0.487 -0.974 -0.487 -0.707 0.500   
VarTomasa:tiempo 0.688 -0.487 -0.487 -0.974 -0.707 0.500 0.500  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-2.13541603 -0.56443548 -0.01419313 0.54277583 1.82448105   
  
Number of Observations: 128  
Number of Groups: 4

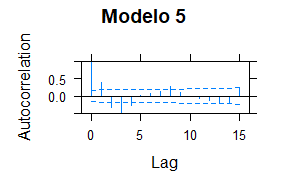
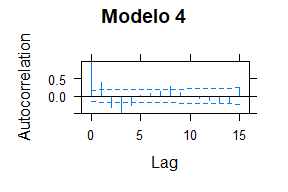
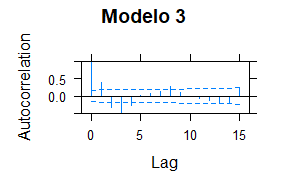
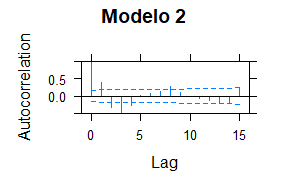
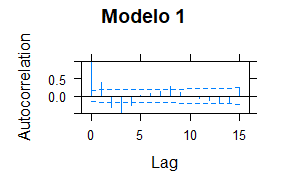
Linear mixed-effects model fit by REML  
 Data: A\_longer   
 AIC BIC logLik  
 1092.945 1159.845 -522.4725  
  
Random effects:  
 Formula: ~Var + tiempo | Block  
 Structure: General positive-definite, Log-Cholesky parametrization  
 StdDev Corr   
(Intercept) 0.001821246118048 (Intr) VarMsq VrPmpr VarTms  
VarMusuq 0.000000071957893 0   
VarPimpernell 0.000000099881448 0 0   
VarTomasa 0.000000021923554 0 0 0   
tiempo 0.000000001884305 0 0 0 0   
Residual 15.286554988799512   
  
Fixed effects: y ~ Var \* tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -3.27117 11.853254 117 -0.275973 0.7831  
VarMusuq -129.03108 16.763033 117 -7.697359 0.0000  
VarPimpernell -99.18087 16.763033 117 -5.916642 0.0000  
VarTomasa -130.04896 16.763033 117 -7.758080 0.0000  
tiempo 0.06014 0.168483 117 0.356951 0.7218  
VarMusuq:tiempo 2.52878 0.238271 117 10.613040 0.0000  
VarPimpernell:tiempo 1.81320 0.238271 117 7.609816 0.0000  
VarTomasa:tiempo 2.57083 0.238271 117 10.789516 0.0000  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms tiempo VrMsq: VrPmp:  
VarMusuq -0.707   
VarPimpernell -0.707 0.500   
VarTomasa -0.707 0.500 0.500   
tiempo -0.974 0.688 0.688 0.688   
VarMusuq:tiempo 0.688 -0.974 -0.487 -0.487 -0.707   
VarPimpernell:tiempo 0.688 -0.487 -0.974 -0.487 -0.707 0.500   
VarTomasa:tiempo 0.688 -0.487 -0.487 -0.974 -0.707 0.500 0.500  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-2.13541600 -0.56443552 -0.01419311 0.54277584 1.82448108   
  
Number of Observations: 128  
Number of Groups: 4

Model df AIC BIC logLik Test L.Ratio  
data.rm1.lme 1 10 1064.945 1092.820 -522.4725   
data.rm1.lme1 2 12 1068.945 1102.395 -522.4725 1 vs 2 0.0000000161890  
data.rm1.lme2 3 24 1092.945 1159.845 -522.4725 2 vs 3 0.0000004738711  
 p-value  
data.rm1.lme   
data.rm1.lme1 1  
data.rm1.lme2 1

Según el AIC, se elige como mejor estructura al modelo que incluye únicamente al bloque como variable aleatoria donde tiene efecto con el intercepto.

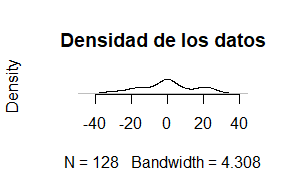
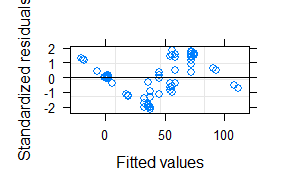
### Evaluación del modelo

#### Autocorrelación temporal

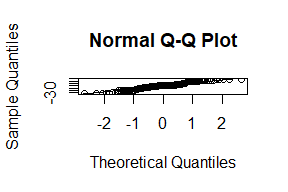


Model df AIC BIC logLik Test L.Ratio  
data.rm1.lme 1 10 1064.945 1092.820 -522.4725   
data.rm1.lme1 2 12 1068.945 1102.395 -522.4725 1 vs 2 0.0000000161890  
data.rm1.lme2 3 24 1092.945 1159.845 -522.4725 2 vs 3 0.0000004738711  
data.rm1.lme3 4 11 1066.945 1097.607 -522.4725 3 vs 4 0.0000004900603  
data.rm1.lme4 5 13 1070.945 1107.182 -522.4725 4 vs 5 0.0000001973835  
 p-value  
data.rm1.lme   
data.rm1.lme1 1  
data.rm1.lme2 1  
data.rm1.lme3 1  
data.rm1.lme4 1

#### Residuales



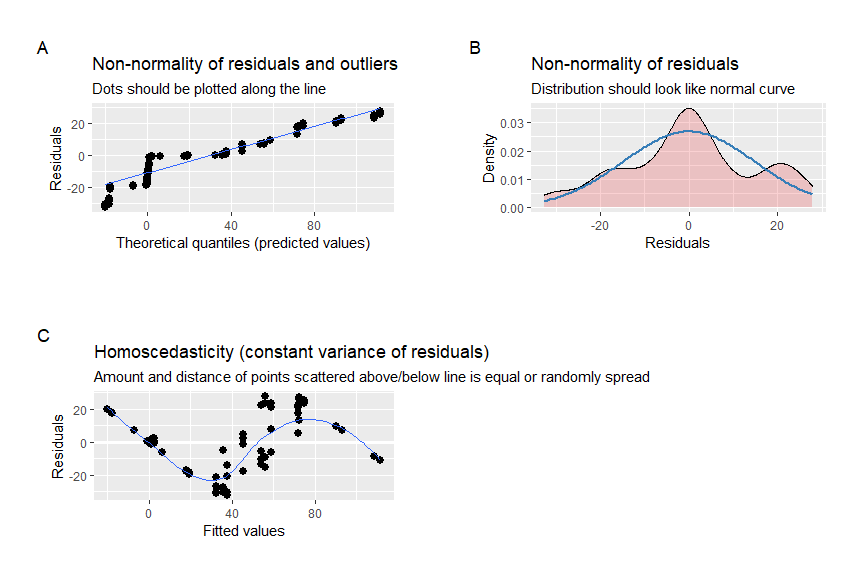
Shapiro-Wilk normality test  
  
data: residuals(data.rm1.lme)  
W = 0.9619, p-value = 0.001169



Se puede observar que los residuos poseen distribución normal en el modelo con estructura aleatoria seleccionada.

Warning in plot\_grid(plot\_model(data.rm1.lme, type = "diag")): Not enough tags  
labels in list. Using letters instead.

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

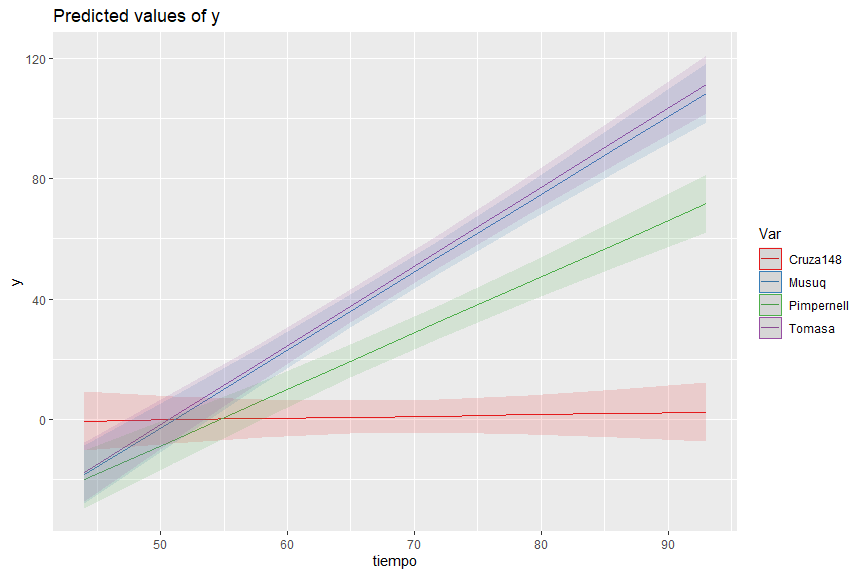
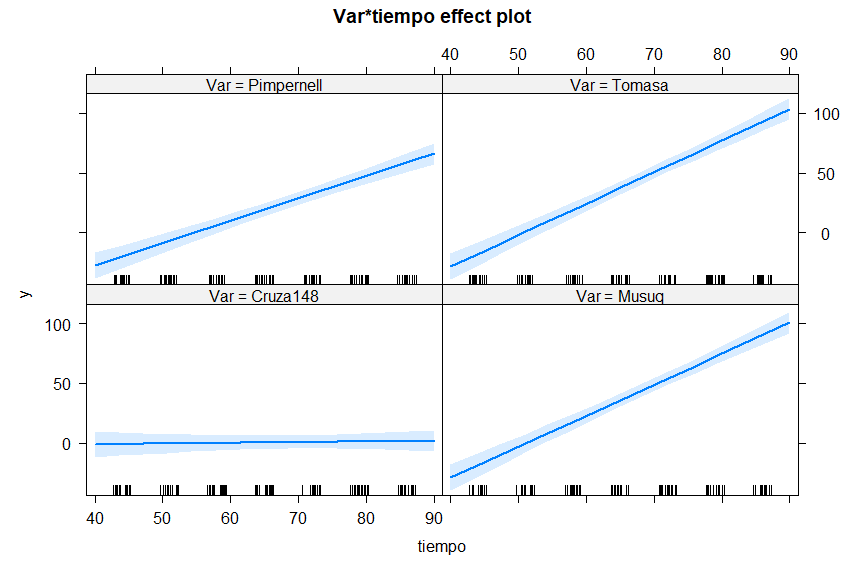


En terminos generales, se puede observar que los residuos cumplen con los supuestos de normalidad, homocedasticidad, independencia y linealidad.

### Exploración de parámetros del modelo

Linear mixed-effects model fit by REML  
 Data: A\_longer   
 AIC BIC logLik  
 1064.945 1092.82 -522.4725  
  
Random effects:  
 Formula: ~1 | Block  
 (Intercept) Residual  
StdDev: 0.0008972843 15.28656  
  
Fixed effects: y ~ Var \* tiempo   
 Value Std.Error DF t-value p-value  
(Intercept) -3.27117 11.853254 117 -0.275973 0.7831  
VarMusuq -129.03108 16.763033 117 -7.697359 0.0000  
VarPimpernell -99.18087 16.763033 117 -5.916642 0.0000  
VarTomasa -130.04896 16.763033 117 -7.758080 0.0000  
tiempo 0.06014 0.168483 117 0.356951 0.7218  
VarMusuq:tiempo 2.52878 0.238271 117 10.613040 0.0000  
VarPimpernell:tiempo 1.81320 0.238271 117 7.609816 0.0000  
VarTomasa:tiempo 2.57083 0.238271 117 10.789516 0.0000  
 Correlation:   
 (Intr) VarMsq VrPmpr VarTms tiempo VrMsq: VrPmp:  
VarMusuq -0.707   
VarPimpernell -0.707 0.500   
VarTomasa -0.707 0.500 0.500   
tiempo -0.974 0.688 0.688 0.688   
VarMusuq:tiempo 0.688 -0.974 -0.487 -0.487 -0.707   
VarPimpernell:tiempo 0.688 -0.487 -0.974 -0.487 -0.707 0.500   
VarTomasa:tiempo 0.688 -0.487 -0.487 -0.974 -0.707 0.500 0.500  
  
Standardized Within-Group Residuals:  
 Min Q1 Med Q3 Max   
-2.13541603 -0.56443548 -0.01419314 0.54277583 1.82448105   
  
Number of Observations: 128  
Number of Groups: 4

numDF denDF F-value p-value  
(Intercept) 1 117 482.0435 <.0001  
Var 3 117 62.9123 <.0001  
tiempo 1 117 450.6597 <.0001  
Var:tiempo 3 117 51.0201 <.0001



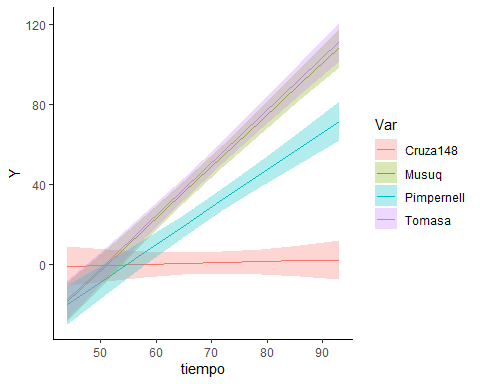
### Aproximación de R square

Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.

R2m R2c  
[1,] 0.861875 0.861875

El fue de 0.86, por lo tanto, con el modelo creado se puede explicar el 86 % de la variancia de la enfermedad.

### Resumen gráfico



## Conclusión final

Segun los analisis de Roy, analisis de variancia tipo III y el analisis de variancia del modelo lineal de efectos mixtos (fijo y aleatorio), se puede afirmar que, con respecto a la enfermedad las variedades tienen un comportamiento diferentes en alguno, con un riesgo de equivocarnos del . Además, la enfermedad tiene diferencias estadísticamente significativas a través del tiempo.

Referencias: <https://www.flutterbys.com.au/stats/tut/tut9.3a.html>

# 2. Regresión lineal múltiple. Estudio de suelo: influencia en el crecimiento de la Ralstonia (Marchitez bacteriana)

Regresión múltiple: Utilizar el archivo suelo2005.txt. Se tiene datos de la enfermedad relativa de plantas (audpc) y análisis de suelo. Buscar el mejor modelo para explicar la enfermedad. Utilizar el análisis 7 como referencia. Pueden incluir otros gráficos.

La Ralstonia Solanacearum es conocida como la “marchitez bacteriana en tomate”, las altas temperaturas, por encima de 25°C, y la humedad en el terreno, en especial luego de fuertes lluvias, propician el nacimiento de ésta bacteria que habita en el agua y en el suelo.

El presente estudio es determinar que elementos del suelo influyen en la proliferacion de la bacteria (aumenta o disminuye la poblacion)

## Reconocimiento de variables

Son 35 muestras de suelo, en donde se inoculo con la bacteria en igual cantidad. Los analisis de suelo fueron realizados en la UNALM en una lectura completa de todos los elementos fisico-quimico.

Para crear un modelo de regresión lineal se tomó como variable dependiente al *audcp* que es la enfermedad relativa de plantas. Las posibles variables predictoras fueron:

pH: Potencial de hidrógeno (pH) del suelo muestreado. Su dominio está en un rango de 0 a 14. Es una variable discreta, debido a que la medición con la herramienta se da en valores hasta con una décima y no existen valores en centésimas o milésimas entre dos valores decimales.

CE: Conductividad eléctrica del suelo muestreado (en dS / cm). Es una variable continua.

: Calcareo (%) del suelo muestreado. Es una variable con distribución de Bernoulli con dominio de 0 a 100 %.

MO: Contenido de Materia orgánica (%) del suelo muestreado. Es una variable con distribución de Bernoulli con dominio de 0 a 100 %.

CIC: Capacidad de intercambio catiónico del suelo muestreado. Representa la cantidad de cationes que las superficies pueden retener. Es una variable con continua.

P: Contenido de Fósforo del suelo muestreado. Es una variable continua expresada en ppm.

K: Contenido de Potasio del suelo muestreado. Es una variable continua expresada en ppm.

Arena: Contenido de Arena (%) del suelo muestreado. Es una variable con distribución de Bernoulli con dominio de 0 a 100 %.

Limo: Contenido de Limo (%) del suelo muestreado. Es una variable con distribución de Bernoulli con dominio de 0 a 100 %.

Arcilla: Contenido de Arcilla (%) del suelo muestreado. Es una variable con distribución de Bernoulli con dominio de 0 a 100 %.

Ca: Contenido de Calcio del suelo muestreado. Es una variable continua expresada en ppm.

Mg: Contenido de Magnesio del suelo muestreado. Es una variable continua expresada en ppm.

K.Cambiable: Contenido de Potasio cambiable del suelo muestreado. Es una variable continua expresada en ppm.

Na: Contenido de Sodio del suelo muestreado. Es una variable continua expresada en ppm.

B: Contenido de Boro del suelo muestreado. Es una variable continua expresada en ppm.

Cu: Contenido de Cobre del suelo muestreado. Es una variable continua expresada en ppm.

Fe: Contenido de Fierro del suelo muestreado. Es una variable continua expresada en ppm.

Mn: Contenido de Manganeso del suelo muestreado. Es una variable continua expresada en ppm.

Zn: Contenido de Zinc del suelo muestreado. Es una variable continua expresada en ppm.

## Análisis exploratorio y descriptivo

### Análisis descriptivo

Los datos cuenta con un total de 35 observaciones (n=35).

Primeras 6 observaciones

pH CE CaCO3 MO CIC P K Arena Limo Arcilla Ca Mg K.Cambiable  
1 3.8 0.72 0.0 1.7 5.76 8.1 28 80 16 4 0.74 0.15 0.10  
2 7.2 1.00 1.4 2.6 20.00 17.5 166 26 42 32 17.21 2.06 0.47  
3 8.4 0.32 9.6 2.0 18.88 17.6 124 44 32 24 14.91 3.37 0.36  
4 4.8 0.83 0.0 1.9 17.28 15.8 86 58 30 12 12.49 1.90 0.26  
5 4.4 0.49 0.0 2.0 6.40 7.2 30 78 14 8 0.76 0.13 0.09  
6 7.6 1.09 0.9 1.3 13.60 8.9 350 48 38 14 10.56 1.79 0.88  
 Na B Cu Fe Mn Zn  
1 0.27 0.9 0.4 295.4 1.3 2.1  
2 0.26 0.2 12.1 12.9 0.5 62.2  
3 0.24 0.4 2.5 23.2 7.9 6.7  
4 0.32 0.5 3.0 352.4 5.1 7.0  
5 0.26 0.4 0.6 334.4 1.8 1.7  
6 0.37 2.5 5.1 12.9 2.1 3.1

Resumen de la base de datos

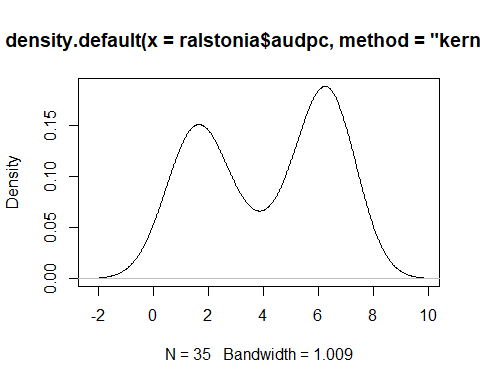
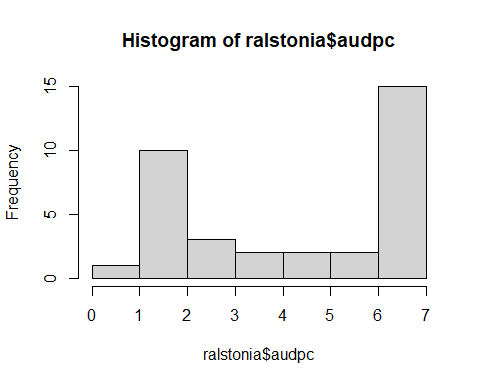
pH CE CaCO3 MO CIC   
 Min. :3.800 Min. :0.0900 Min. :0.000 Min. :0.2 Min. : 4.48   
 1st Qu.:4.950 1st Qu.:0.3250 1st Qu.:0.000 1st Qu.:1.5 1st Qu.: 9.44   
 Median :7.200 Median :0.4800 Median :0.000 Median :2.0 Median :15.04   
 Mean :6.543 Mean :0.9051 Mean :1.214 Mean :2.4 Mean :14.91   
 3rd Qu.:7.800 3rd Qu.:0.8350 3rd Qu.:1.050 3rd Qu.:2.6 3rd Qu.:19.44   
 Max. :8.400 Max. :7.1700 Max. :9.600 Max. :6.7 Max. :26.24   
 P K Arena Limo   
 Min. : 3.60 Min. : 28.0 Min. :12.00 Min. :14.00   
 1st Qu.: 8.70 1st Qu.: 89.5 1st Qu.:40.00 1st Qu.:29.00   
 Median :15.80 Median :166.0 Median :50.00 Median :32.00   
 Mean :21.75 Mean :212.5 Mean :49.03 Mean :33.51   
 3rd Qu.:34.85 3rd Qu.:329.5 3rd Qu.:58.00 3rd Qu.:38.50   
 Max. :64.10 Max. :537.0 Max. :84.00 Max. :62.00   
 Arcilla Ca Mg K.Cambiable   
 Min. : 2.00 Min. : 0.740 Min. :0.130 Min. :0.0900   
 1st Qu.:10.00 1st Qu.: 5.255 1st Qu.:0.570 1st Qu.:0.1950   
 Median :14.00 Median : 9.570 Median :1.540 Median :0.3800   
 Mean :17.17 Mean : 9.765 Mean :1.723 Mean :0.4677   
 3rd Qu.:23.00 3rd Qu.:14.625 3rd Qu.:2.205 3rd Qu.:0.6800   
 Max. :49.00 Max. :22.740 Max. :6.250 Max. :1.4200   
 Na B Cu Fe   
 Min. :0.040 Min. :0.0000 Min. : 0.400 Min. : 4.3   
 1st Qu.:0.115 1st Qu.:0.2500 1st Qu.: 1.500 1st Qu.: 9.2   
 Median :0.250 Median :0.4000 Median : 2.200 Median : 26.5   
 Mean :0.244 Mean :0.9857 Mean : 4.266 Mean : 186.5   
 3rd Qu.:0.290 3rd Qu.:1.1000 3rd Qu.: 5.200 3rd Qu.: 124.1   
 Max. :0.920 Max. :4.9000 Max. :24.900 Max. :1715.0   
 Mn Zn   
 Min. : 0.50 Min. : 0.90   
 1st Qu.: 6.65 1st Qu.: 2.45   
 Median : 29.80 Median : 4.60   
 Mean : 181.21 Mean :10.69   
 3rd Qu.: 210.30 3rd Qu.: 6.85   
 Max. :1567.80 Max. :66.40

Resumen de la base de datos con summarytools

Descriptive Statistics   
bacteria   
N: 35   
  
 Arcilla Arena B Ca CaCO3 CE CIC Cu Fe  
----------------- --------- -------- -------- -------- -------- -------- -------- -------- ---------  
 Mean 17.17 49.03 0.99 9.76 1.21 0.91 14.91 4.27 186.48  
 Std.Dev 10.69 16.87 1.33 5.90 2.42 1.32 6.03 5.62 392.63  
 Min 2.00 12.00 0.00 0.74 0.00 0.09 4.48 0.40 4.30  
 Q1 10.00 40.00 0.20 5.11 0.00 0.32 9.28 1.50 9.00  
 Median 14.00 50.00 0.40 9.57 0.00 0.48 15.04 2.20 26.50  
 Q3 24.00 58.00 1.10 14.91 1.20 0.84 19.52 5.30 154.50  
 Max 49.00 84.00 4.90 22.74 9.60 7.17 26.24 24.90 1715.00  
 MAD 8.90 14.83 0.44 7.07 0.00 0.27 7.35 1.63 31.88  
 IQR 13.00 18.00 0.85 9.37 1.05 0.51 10.00 3.70 114.90  
 CV 0.62 0.34 1.35 0.60 1.99 1.46 0.40 1.32 2.11  
 Skewness 1.02 -0.11 1.69 0.24 2.04 3.37 0.00 2.40 2.68  
 SE.Skewness 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40  
 Kurtosis 0.71 -0.33 1.60 -0.95 3.26 12.21 -1.10 5.41 6.36  
 N.Valid 35.00 35.00 35.00 35.00 35.00 35.00 35.00 35.00 35.00  
 Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00  
  
Table: Table continues below  
  
   
  
 K K.Cambiable Limo Mg Mn MO Na P  
----------------- -------- ------------- -------- -------- --------- -------- -------- --------  
 Mean 212.46 0.47 33.51 1.72 181.21 2.40 0.24 21.75  
 Std.Dev 145.84 0.34 9.73 1.45 326.90 1.49 0.17 16.97  
 Min 28.00 0.09 14.00 0.13 0.50 0.20 0.04 3.60  
 Q1 86.00 0.19 28.00 0.56 5.90 1.40 0.11 8.50  
 Median 166.00 0.38 32.00 1.54 29.80 2.00 0.25 15.80  
 Q3 340.00 0.70 39.00 2.27 214.80 2.60 0.29 39.40  
 Max 537.00 1.42 62.00 6.25 1567.80 6.70 0.92 64.10  
 MAD 130.47 0.31 8.90 1.38 41.51 0.89 0.10 11.42  
 IQR 240.00 0.48 9.50 1.64 203.65 1.10 0.17 26.15  
 CV 0.69 0.73 0.29 0.84 1.80 0.62 0.68 0.78  
 Skewness 0.66 0.92 0.18 1.16 2.60 1.13 2.01 0.87  
 SE.Skewness 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40  
 Kurtosis -0.84 -0.02 0.77 1.00 7.19 0.59 5.82 -0.61  
 N.Valid 35.00 35.00 35.00 35.00 35.00 35.00 35.00 35.00  
 Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00 100.00 100.00  
  
Table: Table continues below  
  
   
  
 pH Zn  
----------------- -------- --------  
 Mean 6.54 10.69  
 Std.Dev 1.48 18.51  
 Min 3.80 0.90  
 Q1 4.80 2.30  
 Median 7.20 4.60  
 Q3 7.80 7.00  
 Max 8.40 66.40  
 MAD 1.48 3.41  
 IQR 2.85 4.40  
 CV 0.23 1.73  
 Skewness -0.44 2.32  
 SE.Skewness 0.40 0.40  
 Kurtosis -1.42 3.74  
 N.Valid 35.00 35.00  
 Pct.Valid 100.00 100.00

### Gráficos exploratorios

Histograma de AUDPC



Se observa que la distribución del AUDPC es muy similar a una bimodal, con una asimetría negativa.

Diagrama de Correlación

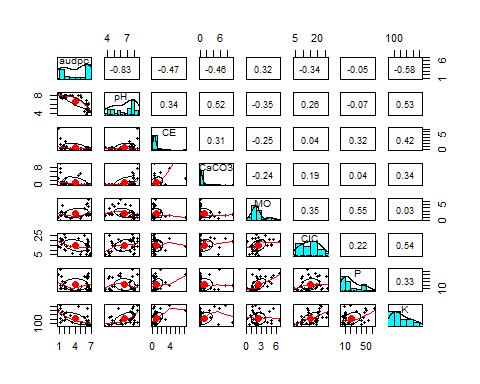


Diagrama de Correlación

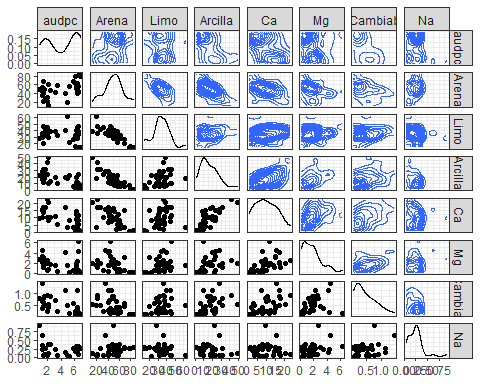


Diagrama de Correlación

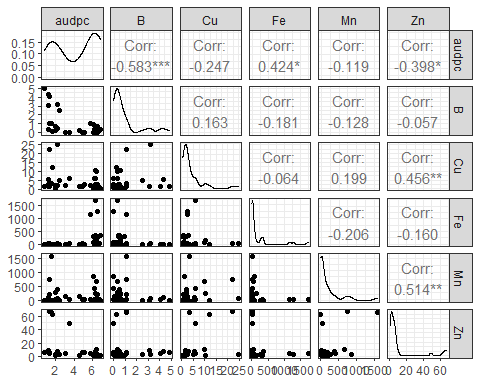
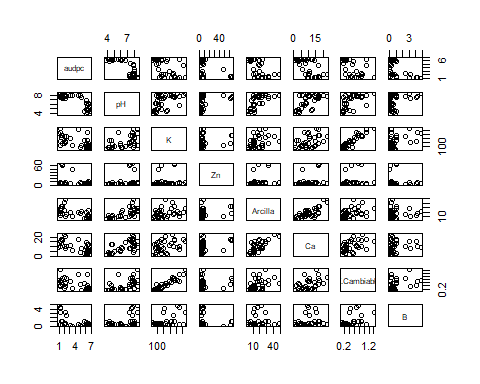


Diagrama de puntos



### Explorar las relaciones

Se tiene una variable dependiente (audpc) y 19 independientes (variables edáficas) en 35 muestras de suelo.

Coeficientes de correlación de Pearson entre AUDPC y las variables regresoras

ralstonia$audpc  
pH -0.83  
CE -0.47  
CaCO3 -0.46  
MO 0.32  
CIC -0.34  
P -0.05  
K -0.58  
Arena 0.46  
Limo -0.15  
Arcilla -0.62  
Ca -0.67  
Mg -0.26  
K.Cambiable -0.55  
Na -0.16  
B -0.58  
Cu -0.25  
Fe 0.42  
Mn -0.12  
Zn -0.40

Las variables de mayor correlacion son:

pH, K, Arcilla, Ca, K.Cambiable y B

## Modelo de regresión lineal multiple completo

### Resumen del modelo

A continuación se presenta el modelo completo:

Tener en cuenta que:

se distribuye de manera similar a la normal (N) con una media 0 y variancia constante.

Resumen del modelo de regresión lineal completo

Call:  
lm(formula = audpc ~ ., data = ralstonia)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.27214 -0.33213 0.05206 0.40037 1.32604   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 30.03032467 16.25886168 1.847 0.08457 .   
pH -0.72346888 0.29990136 -2.412 0.02911 \*   
CE 0.30227977 0.44008899 0.687 0.50265   
CaCO3 -0.05783932 0.13083683 -0.442 0.66474   
MO -0.10055396 0.31174567 -0.323 0.75149   
CIC -0.10009972 0.07647224 -1.309 0.21024   
P 0.02065278 0.01956208 1.056 0.30779   
K -0.00147510 0.00528680 -0.279 0.78404   
Arena -0.19080212 0.15073390 -1.266 0.22489   
Limo -0.18194248 0.14784602 -1.231 0.23741   
Arcilla -0.24913813 0.16535317 -1.507 0.15266   
Ca 0.06698372 0.12278429 0.546 0.59340   
Mg 0.19748922 0.28885327 0.684 0.50459   
K.Cambiable 0.73006164 2.45312002 0.298 0.77009   
Na -0.91112647 3.30980408 -0.275 0.78686   
B -0.93526752 0.29124980 -3.211 0.00583 \*\*  
Cu -0.02329033 0.04302825 -0.541 0.59626   
Fe 0.00085749 0.00099312 0.863 0.40150   
Mn 0.00006228 0.00103473 0.060 0.95280   
Zn -0.02713599 0.01259531 -2.154 0.04787 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.9541 on 15 degrees of freedom  
Multiple R-squared: 0.923, Adjusted R-squared: 0.8254   
F-statistic: 9.462 on 19 and 15 DF, p-value: 0.00003071

### Análisis de variancia

ANOVA del modelo de regresión lineal completo

Analysis of Variance Table  
  
Response: audpc  
 Df Sum Sq Mean Sq F value Pr(>F)   
pH 1 122.392 122.392 134.4445 0.000000006901 \*\*\*  
CE 1 7.329 7.329 8.0509 0.012480 \*   
CaCO3 1 0.000 0.000 0.0003 0.987121   
MO 1 0.009 0.009 0.0096 0.923182   
CIC 1 4.709 4.709 5.1724 0.038067 \*   
P 1 0.642 0.642 0.7049 0.414311   
K 1 0.306 0.306 0.3361 0.570710   
Arena 1 0.711 0.711 0.7812 0.390726   
Limo 1 3.600 3.600 3.9551 0.065292 .   
Arcilla 1 2.396 2.396 2.6323 0.125533   
Ca 1 0.784 0.784 0.8611 0.368122   
Mg 1 4.075 4.075 4.4763 0.051506 .   
K.Cambiable 1 0.537 0.537 0.5895 0.454545   
Na 1 0.211 0.211 0.2313 0.637495   
B 1 8.237 8.237 9.0479 0.008828 \*\*   
Cu 1 1.255 1.255 1.3787 0.258635   
Fe 1 1.820 1.820 1.9991 0.177809   
Mn 1 0.420 0.420 0.4618 0.507148   
Zn 1 4.226 4.226 4.6417 0.047869 \*   
Residuals 15 13.655 0.910   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Selección de variables paso a paso

### Forward

#### Resumen del modelo

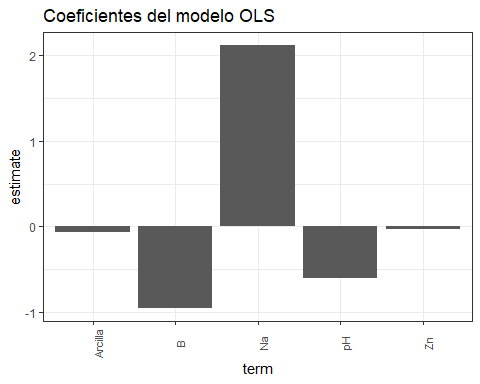
Resumen del modelo de regresión lineal por selección paso a paso con el método Forward

Call:  
lm(formula = audpc ~ pH + B + Zn + Arcilla + Na, data = ralstonia)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-2.02514 -0.55853 0.07554 0.58921 1.40056   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 9.852979 0.854159 11.535 0.00000000000235 \*\*\*  
pH -0.606634 0.152614 -3.975 0.000428 \*\*\*  
B -0.952721 0.173494 -5.491 0.00000647996503 \*\*\*  
Zn -0.025638 0.008361 -3.066 0.004658 \*\*   
Arcilla -0.060741 0.018724 -3.244 0.002964 \*\*   
Na 2.118857 1.271503 1.666 0.106397   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8392 on 29 degrees of freedom  
Multiple R-squared: 0.8848, Adjusted R-squared: 0.865   
F-statistic: 44.56 on 5 and 29 DF, p-value: 0.0000000000009723

#### AIC del modelo

[1] 94.47131

#### Coeficientes del modelo



#### VIF de las variables seleccionadas

pH B Zn Arcilla Na   
2.448247 2.579156 1.156266 1.935981 2.161278

#### MSE del entrenamiento

[1] 0.5835083

#### Indicadores predictivos con la base de ralstonia de entrenamiento

RMSE Rsquared MAE   
0.7638771 0.8848211 0.6272570

### Backward

#### Resumen del modelo

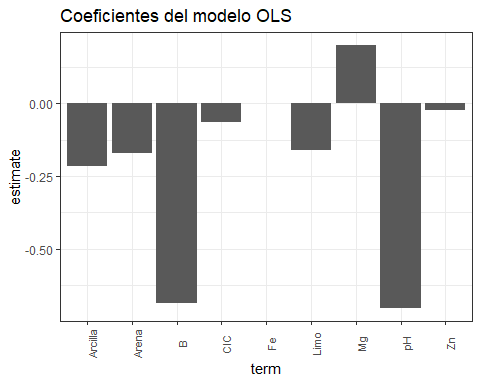
Resumen del modelo de regresión lineal por selección paso a paso con el método Backward

Call:  
lm(formula = audpc ~ pH + CIC + Arena + Limo + Arcilla + Mg +   
 B + Fe + Zn, data = ralstonia)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.69065 -0.33683 -0.09582 0.36596 1.21090   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 27.5376217 8.8990579 3.094 0.004807 \*\*   
pH -0.7038632 0.1610804 -4.370 0.000191 \*\*\*  
CIC -0.0637006 0.0330439 -1.928 0.065319 .   
Arena -0.1703402 0.0879948 -1.936 0.064278 .   
Limo -0.1592396 0.0920441 -1.730 0.095953 .   
Arcilla -0.2155386 0.0872641 -2.470 0.020689 \*   
Mg 0.2002498 0.1299191 1.541 0.135798   
B -0.6869073 0.1341575 -5.120 0.0000273 \*\*\*  
Fe 0.0006994 0.0004217 1.658 0.109726   
Zn -0.0230746 0.0082634 -2.792 0.009885 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8085 on 25 degrees of freedom  
Multiple R-squared: 0.9078, Adjusted R-squared: 0.8746   
F-statistic: 27.36 on 9 and 25 DF, p-value: 0.00000000008788

#### AIC del modelo

[1] 94.67262

#### Coeficientes del modelo



#### VIF de las variables seleccionadas

pH CIC Arena Limo Arcilla Mg B   
 2.938073 2.062963 114.592904 41.739067 45.300200 1.836162 1.661309   
 Fe Zn   
 1.426054 1.216602

#### MSE del entrenamiento

[1] 0.4669578

#### Indicadores predictivos con la base de ralstonia de entrenamiento

RMSE Rsquared MAE   
0.6833431 0.9078271 0.5259072

### Stepwise

#### Resumen del modelo

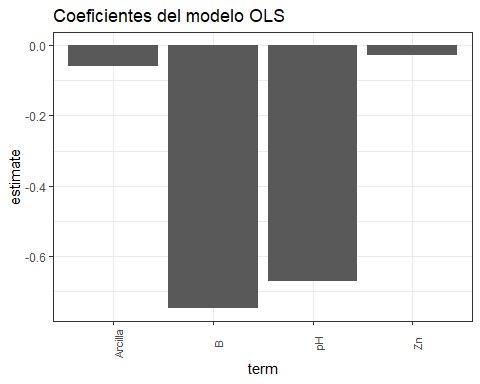
Resumen del modelo de regresión lineal por selección paso a paso con el método Stepwise

Call:  
lm(formula = audpc ~ pH + B + Zn + Arcilla, data = ralstonia)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.8198 -0.5362 0.0769 0.6261 1.5572   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 10.569647 0.759533 13.916 0.0000000000000127 \*\*\*  
pH -0.672064 0.151782 -4.428 0.000117 \*\*\*  
B -0.747847 0.125988 -5.936 0.0000016682630285 \*\*\*  
Zn -0.027721 0.008509 -3.258 0.002787 \*\*   
Arcilla -0.057900 0.019190 -3.017 0.005162 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8637 on 30 degrees of freedom  
Multiple R-squared: 0.8738, Adjusted R-squared: 0.857   
F-statistic: 51.93 on 4 and 30 DF, p-value: 0.0000000000004632

#### AIC del modelo

[1] 95.6719

#### Coeficientes del modelo



#### VIF de las variables seleccionadas

pH B Zn Arcilla   
2.286193 1.284032 1.130407 1.919938

#### MSE del entrenamiento

[1] 0.6393832

#### Indicadores predictivos con la base de ralstonia de entrenamiento

RMSE Rsquared MAE   
0.7996144 0.8737919 0.6670123

### Comparación de modelos

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Método* |  | ajustado |  |  |  |
| *Forward* | 94.47131 | 0.865 | 0.7638771 | 0.8848211 | 0.6272570 |
| *Backward* | 94.67262 | 0.8746 | 0.6833431 | 0.9078271 | 0.5259072 |
| *Stepwise* | 95.6719 | 0.857 | 0.7996144 | 0.8737919 | 0.6670123 |

Se selecciona al modelo Barkward, por lo que el modelo final considerará a las variables predictoras: pH, CIC, Arena, Limo, Arcilla, Mg, B, Fe y Zn. Se sustenta la selección de este modelo por obtener el menor AIC, RMSE y MAE.

## Modelo de regresión lineal final

### Resumen del modelo

A continuación se presenta el modelo final:

Tener en cuenta que:

se distribuye de manera similar a la normal (N) con una media 0 y variancia constante.

Resumen del modelo de regresión lineal final

La prueba de hipótesis para los coeficientes es la siguiente:

: = 0. : .

Call:  
lm(formula = audpc ~ pH + CIC + Arena + Limo + Arcilla + Mg +   
 B + Fe + Zn, data = ralstonia)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.69065 -0.33683 -0.09582 0.36596 1.21090   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 27.5376217 8.8990579 3.094 0.004807 \*\*   
pH -0.7038632 0.1610804 -4.370 0.000191 \*\*\*  
CIC -0.0637006 0.0330439 -1.928 0.065319 .   
Arena -0.1703402 0.0879948 -1.936 0.064278 .   
Limo -0.1592396 0.0920441 -1.730 0.095953 .   
Arcilla -0.2155386 0.0872641 -2.470 0.020689 \*   
Mg 0.2002498 0.1299191 1.541 0.135798   
B -0.6869073 0.1341575 -5.120 0.0000273 \*\*\*  
Fe 0.0006994 0.0004217 1.658 0.109726   
Zn -0.0230746 0.0082634 -2.792 0.009885 \*\*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.8085 on 25 degrees of freedom  
Multiple R-squared: 0.9078, Adjusted R-squared: 0.8746   
F-statistic: 27.36 on 9 and 25 DF, p-value: 0.00000000008788

A un nivel de significancia de 0.1, los coeficientes que estadísticamente diferentes de 0 se observaron para las variables pH, CIC, Arena, Limo, Arcilla, B y Zn.

### Análisis de variancia

La prueba de hipótesis para cada variable regresora es:

: La variable regresora es buena predictora. : La variable regresora no es buena predictora.

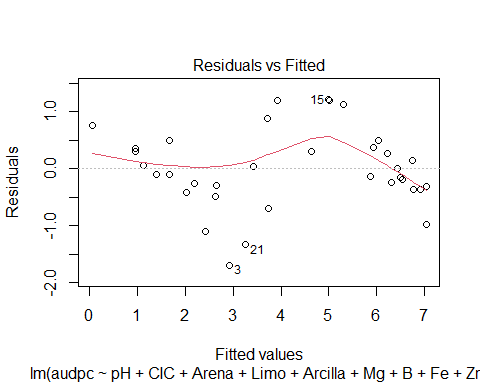
ANOVA del modelo de regresión lineal final

Analysis of Variance Table  
  
Response: audpc  
 Df Sum Sq Mean Sq F value Pr(>F)   
pH 1 122.392 122.392 187.2179 0.0000000000004104 \*\*\*  
CIC 1 2.912 2.912 4.4538 0.044999 \*   
Arena 1 3.245 3.245 4.9644 0.035105 \*   
Limo 1 4.680 4.680 7.1585 0.012971 \*   
Arcilla 1 2.456 2.456 3.7566 0.063970 .   
Mg 1 1.111 1.111 1.6994 0.204246   
B 1 16.664 16.664 25.4901 0.0000328538507304 \*\*\*  
Fe 1 2.413 2.413 3.6911 0.066175 .   
Zn 1 5.097 5.097 7.7974 0.009885 \*\*   
Residuals 25 16.344 0.654   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

A un nivel de significancia de 0.1, se observó que las variables pH, CIC, Arena, Limo, Arcilla, B, Fe y Zn fueron buenas predictoras. La variable Mg no es una buena variable que explique al AUDPC pero según la selección Backward ayuda a crear un mejor modelo predictivo. La eliminación de esta variable para un modelo de inferencia puede ser opcional según el criterio del investigador, sin embargo, en un modelo predictivo debería dejarse porque su uso si mejora la capacidad predictiva.

### Análisis de supuestos

#### Valores predichos vs Residuos



Valores predichos vs Residuos del AUDPC.

Se puede observar que las observaciones 3, 21 y 15 poseen valores de los residuos del AUDPC más distantes a la media de los residuos observados del AUDPC, calculada según el efecto de los valores predichos generados por las variables predictoras sobre el AUDPC. La línea de tendencia generada para los residuos del AUDPC según los valores predichos tiene una forma no lineal, lo que evidencia que los residuos del AUDPC dependen de los valores predichos y que el AUDPC estaría explicado también por otros factores no evaluados en el modelo. Los residuos no son aleatorios, por lo que es recomendable emplear una prueba de hipótesis para la evaluación del cumplimiento de linealidad y aditividad.

#### Gráfico de probabilidad normal de residuos

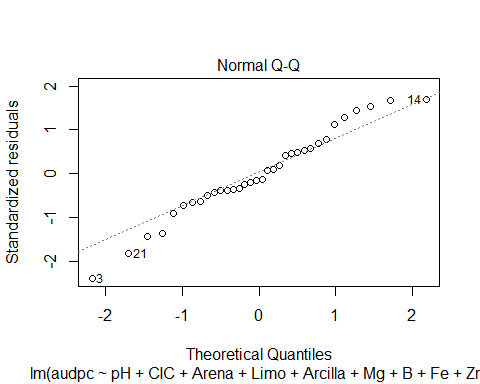
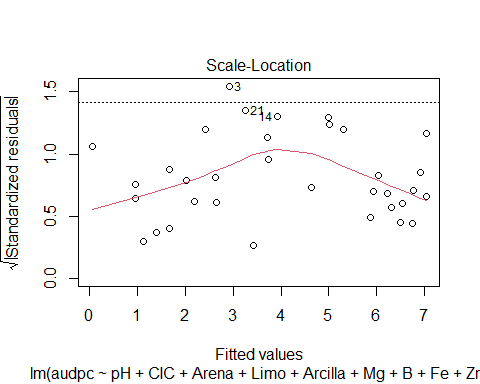


Gráfico Quantil – Quantil de los residuos del AUDPC.

El gráfico de quantil teórico (distribución normal teórica) vs los quantiles observados (residuos estandarizados del AUDPC), evidencia que las observaciones 3, 21 y 14, están más distantes del cumplimiento de la distribución teórica normal formando una asimetría negativa. Es necesario comprobar la normalidad de los residuos con el uso de una prueba estadística.

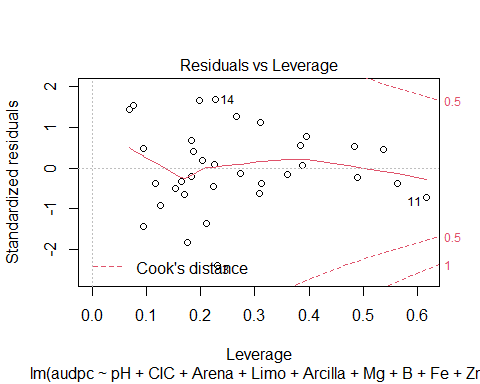
#### Valores predichos vs Raíz de los residuos estandarizados absolutos



Valores predichos vs Raíz cuadrada de los residuos estandarizados absolutos del AUDPC.

La dispersión de la raíz cuadrada de los residuos estandarizados absolutos del AUDPC, no es constante y a medida que los valores predichos del AUDPC son mayores, los residuos estandarizados absolutos también aumentan hasta que en un cierto punto disminuyen, lo que evidencia ciertos problemas de no contancia de variancia. Por estos motivos, se debe recurrir a pruebas de hipótesis para la homocedasticidad.

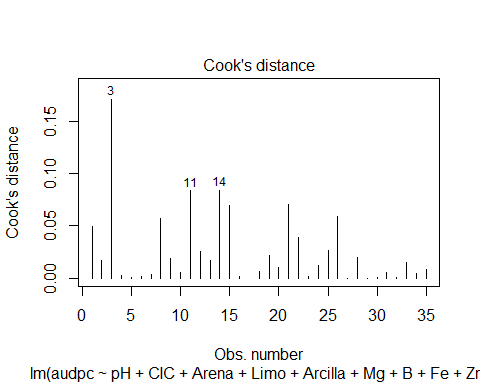
#### Leverage vs Residuos estandarizados



Puntos de apalancamiento vs residuos estandarizados absolutos del AUDPC.

Se observa que los puntos 3, 11 y 14 son los puntos más influyentes en el modelo, por lo que podría ser necesario evaluar que tanta influencia poseen sobre el modelo en caso de incumplimiento de algún supuesto.

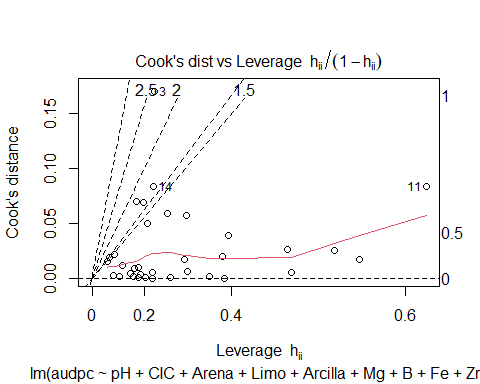
#### Distancias de Cook



Distancias de Cook de los residuos del AUDPC.

Las observaciones 3, 11 y 14, podrían ser posibles valores influyentes, pues, poseen las mayores distancias de Cook.

#### Distancias de Cook vs Leverage



Distancias de Cook vs Puntos de Apalancamiento de los residuos del AUDPC.

Las observaciones 3 y 14, serían valores influyentes, y la observacion 11 sería un punto de apalancamiento.

#### Prueba de hipótesis para la linealidad del modelo

Para evaluar el cumplimiento de este supuesto, se elaboró la siguiente hipótesis:

: El modelo creado cumple con el supuesto de linealidad y aditividad.

: El modelo creado no cumple con el supuesto de linealidad y aditividad.

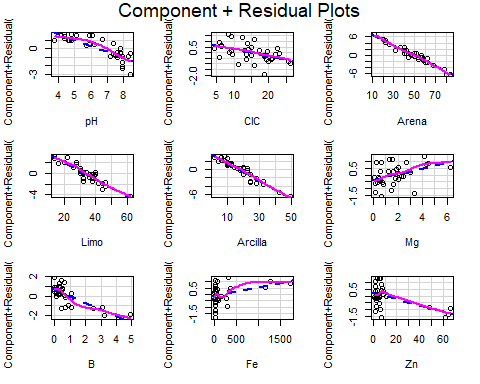
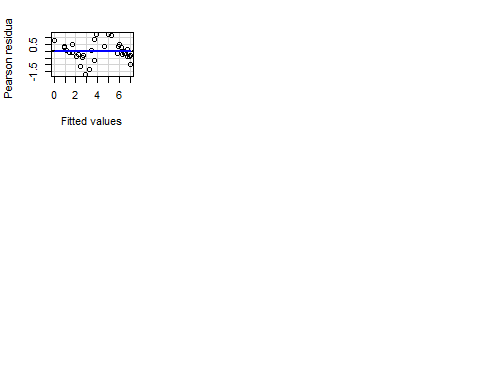
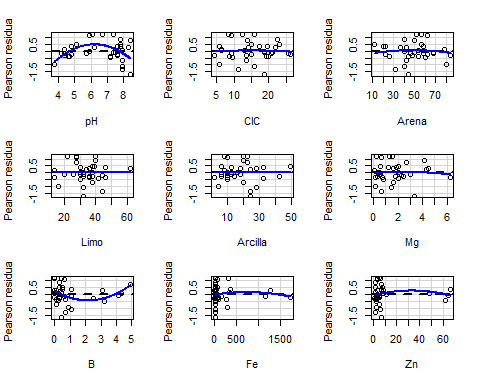


Gráfico de residuo para modelos sin interacción.



Test stat Pr(>|Test stat|)   
pH -2.8829 0.008183 \*\*  
CIC -0.1897 0.851113   
Arena -0.4977 0.623243   
Limo -0.0068 0.994632   
Arcilla 0.0575 0.954584   
Mg -0.3758 0.710374   
B 1.7981 0.084756 .   
Fe -0.5030 0.619528   
Zn -0.4846 0.632374   
Tukey test -0.0925 0.926274   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Gráfico de residuo para modelos con interacción.

Prueba de hipótesis de Tukey para evaluación de la linealidad y aditividad.

Test Pvalue   
-0.09253367 0.92627404

Prueba de hipótesis Global Stats para evaluación de la linealidad y aditividad.

Call:  
lm(formula = audpc ~ pH + CIC + Arena + Limo + Arcilla + Mg +   
 B + Fe + Zn, data = ralstonia)  
  
Coefficients:  
(Intercept) pH CIC Arena Limo Arcilla   
 27.5376217 -0.7038632 -0.0637006 -0.1703402 -0.1592396 -0.2155386   
 Mg B Fe Zn   
 0.2002498 -0.6869073 0.0006994 -0.0230746   
  
  
ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS  
USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:  
Level of Significance = 0.05   
  
Call:  
 gvlma(x = modelo\_final)   
  
 Value p-value Decision  
Global Stat 1.8532083 0.7627 Assumptions acceptable.  
Skewness 0.1885012 0.6642 Assumptions acceptable.  
Kurtosis 0.0007436 0.9782 Assumptions acceptable.  
Link Function 0.0124825 0.9110 Assumptions acceptable.  
Heteroscedasticity 1.6514810 0.1988 Assumptions acceptable.

* CONCLUSIÓN:

A un nivel de significancia de 0.05, según la prueba de Tukey y Global Stats, no se rechaza la hipótesis nula (H0), por lo tanto, se cumple con el supuesto y el modelo es lineal y aditivo.

#### Prueba de hipótesis para la normalidad de residuos

##### Simetría de los residuos

Medida de simetría para los residuos

[1] -0.1797623

##### Curtosis de los residuos

Medida de kurtosis para los residuos

[1] 3.022581

Se realizó las pruebas de hipótesis para normalidad de los residuos, con la siguiente hipótesis:

Los residuos estandarizados del modelo planteado se distribuyen de forma similar a la función normal.

Los residuos estandarizados del modelo planteado no se distribuyen de forma similar a la función normal.

Prueba Shapiro-Wilk para los residuos de la variable con hasta 5000 muestras

Shapiro-Wilk normality test  
  
data: .  
W = 0.97068, p-value = 0.4805

Prueba Lilliefors para los residuos de la variable con más 50 muestras (recomendable)

Lilliefors (Kolmogorov-Smirnov) normality test  
  
data: .  
D = 0.095365, p-value = 0.6029

Prueba de Anderson Darlin es similar a Shapiro

Anderson-Darling normality test  
  
data: .  
A = 0.33172, p-value = 0.5018

Prueba de Kolmogorov Smirnov

One-sample Kolmogorov-Smirnov test  
  
data: .  
D = 0.090375, p-value = 0.9206  
alternative hypothesis: two-sided

CONCLUSIÓN:

A un nivel de significancia de 0.05, según las pruebas de Shapiro – Wilk, Lilliefors, Anderson - Darlin y Kolmogorov - Smirnov no se rechaza la hipótesis nula (H0), por lo tanto, se cumple con el supuesto y los residuos del AUDPC poseen una distribución similar a la normal.

#### Prueba de hipótesis para la homocedasticidad

Por otro lado, al evaluar el supuesto de homocedasticidad y constrastar con la siguiente hipótesis:

La variancia de los errores del modelo planteado es homocedástica.

La variancia de los errores del modelo planteado es heterocedástica.

Prueba de Score para varianzas no constantes

Non-constant Variance Score Test   
Variance formula: ~ fitted.values   
Chisquare = 0.1892029, Df = 1, p = 0.66358

Prueba de Ols Test de Breusch Pagan

Breusch Pagan Test for Heteroskedasticity  
 -----------------------------------------  
 Ho: the variance is constant   
 Ha: the variance is not constant   
  
 Data   
 ---------------------------------  
 Response : audpc   
 Variables: fitted values of audpc   
  
 Test Summary   
 ----------------------------  
 DF = 1   
 Chi2 = 0.1892029   
 Prob > Chi2 = 0.6635808

Prueba de Breusch Pagan con residuos estandarizados

studentized Breusch-Pagan test  
  
data: .  
BP = 13.819, df = 9, p-value = 0.1289

Prueba de Breusch Pagan sin residuos estandarizados

Breusch-Pagan test  
  
data: .  
BP = 13.975, df = 9, p-value = 0.1232

CONCLUSIÓN:

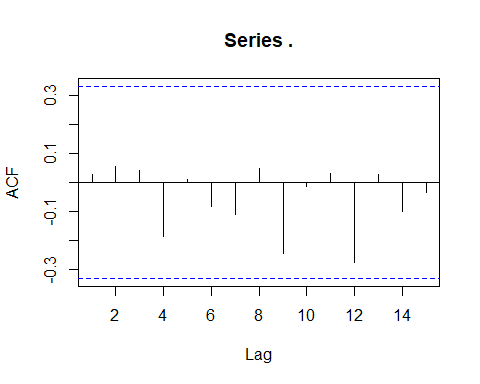
A un nivel de significancia de 0.05, según la prueba de score para variancias no constantes y las pruebas de Breush Pagan, no se rechaza la hipótesis nula (H0), por lo tanto, se cumple con el supuesto y la variancia del AUDPC es constante.

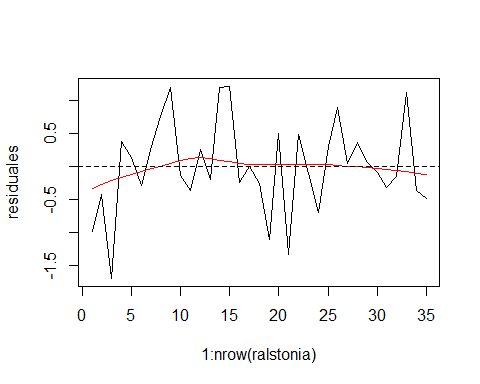
#### Prueba de hipótesis para la autocorrelación de residuos

Al evaluar el supuesto de autocorrelación y constrastar la siguiente hipótesis:

Los errores del modelo planteado no están autocorrelacionados.

Los errores del modelo planteado están autocorrelacionados.





Número de observación vs residuos del AUDPC.

La observación 2 posee residuos del AUDPC más distante del valor 0 de los residuos y de la tendencia lineal de los residuos del AUDPC. Además, se puede observar que la línea de tendencia generada con la función lowess se mantiene muy cercana al residuo 0, por lo tanto, se puede asumir que la aleatorización de los residuos según el número de observación ayuda al cumplimiento del supuesto de autocorrelación de residuos.

Prueba Durbin - Watson para autocorrelación de residuos

Durbin-Watson test  
  
data: .  
DW = 1.8706, p-value = 0.5682  
alternative hypothesis: true autocorrelation is not 0

Prueba Durbin - Watson para autocorrelación de residuos con vectorización de los residuos

lag Autocorrelation D-W Statistic p-value  
 1 0.02831062 1.870646 0.558  
 2 0.05391403 1.800269 0.762  
 3 0.04034771 1.575132 0.398  
 4 -0.18773484 2.021423 0.636  
 5 0.01002138 1.618736 0.844  
 6 -0.08309658 1.799418 0.632  
 7 -0.11209812 1.851688 0.348  
 8 0.04626779 1.492268 0.982  
 9 -0.24395975 1.984876 0.088  
 10 -0.01411727 1.475938 0.782  
 Alternative hypothesis: rho[lag] != 0

CONCLUSIÓN:

A un nivel de significancia de 0.05, según la prueba de Durbin – Watson, no se rechaza la hipótesis nula (H0), por lo tanto, se cumple con el supuesto y los residuos del AUDPC no están autocorrelacionados. El estadístico de Durbin Watson está dentro del rango de 1.5 a 2.5, y el p valor es mayor de 0.05, por lo tanto, existe independencia de los residuos.

## Conclusión final

Desde un punto de vista inferencial y predictivo, según la selección de variables paso a paso por el método Backward (hacia atrás), el mejor modelo para explicar el AUDPC es el que incluye a las variables pH, CIC, Arena, Limo, Arcilla, Mg, B, Fe y Zn.