UNIVERSIDAD NACIONAL DE COLOMBIA

REGRESIÓN LINEAL MULTIPLE PARTE 1

Autor:

Daniela Pico

Jhonatan Smith

Profesor:

Isabel Cristina Ramirez

2021-01

1. Ajuste un modelo de regresión lineal múltiple, muestre la tabla de parámetros ajustados y escriba la ecuación ajustada. Calcule la Anova del modelo Es significativo el modelo? ¿Qué proporción de la variabilidad total de la respuesta es explicada por el modelo? Opine sobre esto úlltimo.
2. Calcule los coeficientes de regresión estandarizados y concluya acerca de cúal de las variables aporta más a la respuesta según la magnitud en valor absoluto de tales coeficientes (cuidado, no confunda esto con la significancia de los coeficientes de regresión).
3. Pruebe la significancia individual de cada uno de los parámetros del modelo (excepto intercepto), usando la prueba t, y para dos cualesquiera de las predictoras, usando la prueba F con sumas de cuadrados extras con test lineal general; en cada caso, especifique claramente el modelo reducido y completo, estadistico de la prueba, su distribución, cálculo de valor P, decisión y conclusión a la luz de los datos.
4. Calcule las sumas de cuadrados tipo I (secuenciales) y tipo II (parciales) ¿Cuál de las variables tienen menor valor en tales sumas? ¿Que puede significar ello?
5. Construya y analice gráficos de los residuales estudentizados vs. Valores ajustados y contra las variables de regresión utilizadas. Qué información proporcionan estas gráficas?
6. Construya una gráfica de probabilidad normal para los residuales estudentizados. ¿Existen razones para dudar de la hipótesis de normalidad sobre los errores en este modelo?
7. Diagnostique la presencia de observaciones atipicas, de balanceo y/o influenciales. Recuerde que cada unidad de observación es una institución hospitalaria. En particular, ¿las observaciones ID = 47 e ID = 112 se diferencian del resto? Ajuste el modelo de regresión sin estas dos observaciones, presente solo la tabla de parámetros ajustados resultante ¿Cambian notoriamente las estimaciones de los parámetros, sus errores estandard y/o la significancia? ¿Qué concluye al respecto? Evalúe el gráfico de normalidad para los residuales estudentizados para este ajuste ¿mejoró la normalidad? Concluya sobre los efectos de este par de observaciones.
8. Para el modelo con todas las variables y sin las observaciones con ID = 47 e ID = 112, realice diagnósticos de multicolinealidad mediante

- Matriz de correlación de las variables predictoras
- VIF's
- Proporciones de varianza

9. En el modelo ajustado sin las observaciones con ID = 47 e ID = 112, construya modelos de regresión utilizando los métodos de selección (muestre de cada método sólo la tabla de resumen de este y la tabla ANOVA y la de parámetros estimados del modelo ánalogamente resultante):

- Selección según el R2 adj
- Selección según el estadadístico Cp
- Stepwise

- Selección hacia adelante o forward * Selección hacia atrás o backward

10. Con base en los anteriores numerales, ¿Cúal modelo sugiere para la variable respuesta? ¿por qué?

Resultados

1. El siguiente informe proporciona datos recolectados de un estudio sobre la eficacia del control de infecciones nosocomiales, cuyo objetivo principal fue determinar si los programas de vigilancia y control de infecciones han reducido las tasas de infección nosocomialen hospitales de Estados Unidos. Estos datos consisten de una muestra aleatoria de n = 80; 90; 100; 70; 65; 85 hospitales, respectivamente, seleccionados de los 338 hospitales originales investigados. Los datos presentados corresponden al periodo de estudio 1975-76. Se presentaron las siguientes variables:

- ID: Número de indentificación de registro.
- DPERM: Longitud de permanencia.
- EDAD: Edad.
- RINF: Riesgo de infección.
- RRX: Razón de rutina de rayos X del pecho.
- NCAMAS: Número de camas.
- AEM: Afiliación de escuela de medicina.
- PDP: Censo promedio diario.
- NENFERM: Número de enfermeras.
- FSD: Facilidades y servicios disponibles.
- REGION: Región.
- RRC: Razón de rutina de cultivos.

Se desea estudiar la longitud de permanencia (Y) en función de las variables predictorias:

- X_1= EDAD
- X_2= RINF
- X_3= RRC
- X_4= RRX
- X_5= NCAMAS
- X_6= PDP
- X_7= NENFERM
- X_8= FSD

En primera instancia se realiza un analisis descriptivo del comportamiento de los datos a través de un gráfico de dispersión.

Se plantea un modelo de RLM para el problema:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_8 X_{i8} + \varepsilon_i, \quad i = 1, 2, \ldots, 80$$

Que tiene como supuestos lo siguiente:

$$\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, 80$$

También se puede especificar el modelo en términos matriciales, así:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{con} \quad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

Especificación del modelo de RLM, ANOVA y parámetros estimados

```
Call:
lm(formula = DPERM ~ EDAD + RINF + RRC + RRX + NCAMAS + PDP +
    NENFERM + FSD)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1671 -0.8795 -0.2033  0.7582  6.1729

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.052672   1.982646  -0.531  0.59712
EDAD         0.112843   0.036214   3.116  0.00265 **
RINF         0.478359   0.157108   3.045  0.00327 **
RRC          0.019804   0.018805   1.053  0.29586
RRX          0.018959   0.009117   2.080  0.04118 *
NCAMAS      -0.010422   0.004683  -2.226  0.02921 *
PDP          0.023132   0.005267   4.392 3.84e-05 ***
NENFERM     -0.007527   0.003061  -2.458  0.01639 *
FSD          0.004740   0.017084   0.277  0.78222
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.366 on 71 degrees of freedom
Multiple R-squared:  0.6251,    Adjusted R-squared:  0.5829
F-statistic:  14.8 on 8 and 71 DF,  p-value: 1.685e-12

      Sum_of_Squares DF Mean_Square F_Value     P_value
Model        221.000  8    27.62506 14.7988 1.68478e-12
Error        132.537 71     1.86671
```

El modelo ajustado es:

$$Y_i = -1.052672 + 0.112843X_{i1} + 0.478359X_{i2} + 0.019804X_{i3} + 0.018959X_{i4} - 0.010422X_{i5} + 0.023132X_{i6}$$

$$-0.007527X_{i7} + 0.004740X_{i8} + \varepsilon_i$$

$$i = 1, 2, \ldots, 80$$

Prueba de Significancia de la regresión

Se quiere probar:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_8 = 0, \quad \text{vs.}$$
$$H_1 : \text{Algún } \beta_j \neq 0, j = 1, \ldots, 8.$$

Para ello se usa la tabla de análisis de varianza. De ella se obtienen los valores del estadístico de prueba $F_0 = 14.7988$ y su correspondiente valor-P $\text{vp} = 1.68478e - 12$.

Como $\text{vp} < 0.05 = \alpha$ se rechaza H_0 concluyendo que el modelo de RLM propuesto es significativo. Esto quiere decir, que la logitud de permanencia es afectada significativamente por al menos una de las predictoras consideradas.

Cálculo e interpretación del coeficiente de determinación

Sabemos que $R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$, de manera que se puede calcular de la tabla ANOVA.

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{221.000}{221.000 + 132.537} = 221.000/(221.000 + 132.537)‘$$

Según el R^2 el 62.51% de la variabilidad total de la longitud de permanencia es explicado por el modelo propuesto.

Por otro lado, se puede calcular el R^2 ajustado como una medida de bondad de ajuste, así:

$$R^2_{\text{adj}} = 1 - \frac{(n-1)\,\text{MSE}}{\text{SST}} = 1 - \frac{(80-1)\,1.86671}{221.000 + 132.537} = 0.5828723$$

Según el $R^2 ajustado$ el 58.29% de la variabilidad total de la longitud de permanencia es explicado por el modelo propuesto.

Teniendo en cuenta que R^2_{adj} penaliza la varianza a medida que se agregan covariables (factor que no tiene en cuenta por si solo R^2) se prefiere usar para el caso de Regresion Lineal Multiple (RLM) el ajustado.

2. Como los X_j tienen diferente escala de medida, no se puede determinar cual de ellas tiene mayor o menor efecto parcial sobre la respuesta media, por esto tiene sentido realizar una estandarización de las variables, de tal forma que queden en la misma escala y puedan ser comparadas.

```
## Coeficientes estimados y Coeficientes estimados estandarizados
```

```
##                Estimacion    Coef.Std
## (Intercept) -1.052672111  0.00000000
## EDAD         0.112843416  0.24325881
## RINF         0.478358700  0.30053371
## RRC          0.019803538  0.10255179
## RRX          0.018959029  0.18201393
## NCAMAS      -0.010422183 -0.91448479
## PDP          0.023131707  1.61358192
## NENFERM     -0.007526505 -0.46158384
## FSD          0.004740396  0.03301155
```

Según la magnitud del valor absoluto de los coeficientes estudentizados la variable que tiene mayor efecto en la respuesta será el censo promedio diario (PDP) con un valor de 1.61358192 y la variable con menor efecto en la respuesta media es facilidades y servicios disponibles(FSD) con un valor de 0.03301155

3. **Prueba de significancia individual de los parametros usando la prueba t**

Estas pruebas establecen el siguiente juego de hipótesis:

$$\begin{array}{l} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{array} \quad \text{para} \quad j = 1, 2, \ldots, 8.$$

De la tabla de parámetros estimados, a un nivel de significancia $\alpha = 0.05$ se rechaza H_0 si $|T_0| > T_{\frac{\alpha}{2}}, n-k-1$

Para este caso con la $T_{(1-\frac{0.05}{2},71)} = 1.993943$ basta comparar con los datos suministrados en la tabla anterior en la columna de t-values:

- $EDAD = |3.116| > 1.99$
- $RINF = |3.045| > 1.99$
- $RRC = |1.053| < 1.99$
- $RRX = |2.080| > 1.99$
- $NCAMAS = |-2.226| > 1.99$

- $PDP = |4.392| > 1.99$
- $NENFERM = |-2.458| > 1.99$
- $FSD = |0.277| < 1.99$

Se concluye que los parámetros individuales $\beta_1, \beta_2, \beta_4, \beta_5, \beta_6, \beta_7$ son significativos cada uno en presencia de los demás parámetros; por otro lado, se encuentra que β_3, β_8 son individualmente no significativos en presencia de los demás parámetros.

Interpretación de los parámetros estimados

$\widehat{\beta}_1 = 0.112843$ indica que por cada unidad de aumento en la edad el promedio de la longitud de permanencia aumenta en 0.112843 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_2 = 0.478359$ indica que por cada unidad de aumento en el riesgo de infección el promedio de la longitud de permanencia aumenta en 0.478359 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_4 = 0.018959$ indica que por cada unidad de aumento en la razón de rutinas de rayos X en el pecho el promedio de la longitud de permanencia aumenta en 0.018959 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_5 = -0.010422$ indica que por cada unidad de aumento en el número de camas el promedio de la longitud de permanencia disminuye en 0.010422 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_6 = 0.023132$ indica que por cada unidad de aumento en el censo promedio diario el promedio de la longitud de permanencia aumenta en 0.023132 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_7 = -0.007527$ indica que por cada unidad de aumento en el numero de enfermeras el promedio de la longitud de permanencia disminuye en 0.007527 unidades, cuando las demás variables predictoras se mantienen fijas.

Prueba F con sumas de cuadrados extras

Para esta prueba se elegiran convenientemente Dos parametros no significativos del modelo, en este caso β_3 y β_8. Se plantean las siguientes hipotesis:

$$H_0: \ \beta_3 = \beta_8 = 0 \quad \text{vs.} \quad H_1: \ \text{Algún } \beta_j \neq 0, \ j = 3, 8$$

Esta prueba se desarrolla usando sumas de cuadrados extra y se requiere la tabla de todas las regresiones posibles como se presenta a continuación.

Modelo FULL

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_8 X_{i8} + \varepsilon_i, \quad i = 1, 2, \ldots, 80$$
$$\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, 80$$

Modelo reducido

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + E_i$$

$$\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, 80$$

Estadistico de prueba F0

$$F_0 = \frac{(SSE(MR) - SSE(MF))/2}{MSE(MF)}$$

$$= \frac{(SSE(X_1, X_2, X_4, X_5, X_6, X_7) - SSE(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8))/(n-7) - (n-9)}{MSE(X_1, X_2, X_3, X_4X_5, X_6, X_7, X_8))}$$

$$= \frac{SSR(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8|X_3, X_8)/2}{MSE(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)} \sim f_{2,71}\, bajo H_0$$

```
## Linear hypothesis test
##
## Hypothesis:
## RRC = 0
## FSD = 0
##
## Model 1: restricted model
## Model 2: DPERM ~ EDAD + RINF + RRC + RRX + NCAMAS + PDP + NENFERM + FSD
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     73 134.75
## 2     71 132.54  2    2.2096 0.5918  0.556
```

Para el criterio de decisión se requiere obtener el valor crítico de una distribución $f_{v,n-k-1} = f_{2,71}$ a un nivel de significancia $\alpha = 0.05$, esto es, $f_{0.05,2,71} = 3.1257642$.

Como $F_0 = 0.5918 < f_{0.05,2,87} = 3.1257642.$, entonces no se rechaza H_0 y se concluye que el conjunto de predictoras simultaneamente no son significativas, en presencia de los demás parámetros lo que implica que la variable RRC y FSD no son significativas para explicar la Longitud de permanencia. Notese que este resultado coincide con la prueba de significancia individual de los parametros.

4. **Suma de cuadrados Tipo I**

```
##            Sum Sq
## EDAD       33.536
## RINF      102.794
## RRC         2.147
## RRX        10.713
## NCAMAS     21.358
## PDP        38.693
## NENFERM    11.616
## FSD         0.144
## Residuals 132.537
```

La suma de cuadrados extra tipo I de 1 grado de libertad agrega secuencialmente las variables según el orden establecido en el modelo full, es decir:

1) $SSR(EDAD) = 33.536$ Y este es el resultado de ajustar la recta de regresio $Y_i = \beta_0 + \beta_1 * X_{i1}$ que es Longidtud de permanencia (DPERM) vs Edad

2) $SSR(RINF|EDAD) = 102.794$ Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable RINF al modelo de la longitud de permanencia dado que estaba la variable edad.

3) $SSR(RRC|EDAD, RINF) = 2.147$ Es el incremento del SSR (Sumas de cuadrados de la regresion) al ingresar la variable RRC (Razon de rutina de cultivos) al modelo de la longitud de permanencia dado que estaban las variables edad y riesgo de infección.

Procediendo de la misma manera para cada uno de los casos, para el ultimo caso se tendria que:

- $SSR(FSD|EDAD, RINF, ..., NENFERM) = 0.144$ Es el cambio marginal en SSR al ingresar la variable FSD (facilidad de servicios disponibles) al modelo de longitud de permanencia dado que estaban las anteriores predictoras, en este caso al ingresar la ultima covariable se presento la menor reducción marginal en la suma de cuadrados extras cuando las demás predictorias fueron agragadas al modelo.

Suma de cuadrados tipo II

```
## Anova Table (Type II tests)
##
## Response: DPERM
##            Sum Sq Df F value    Pr(>F)
## EDAD       18.125  1  9.7096  0.002645 **
## RINF       17.306  1  9.2706  0.003265 **
## RRC         2.070  1  1.1091  0.295856
## RRX         8.073  1  4.3246  0.041178 *
## NCAMAS      9.247  1  4.9537  0.029211 *
## PDP        36.011  1 19.2913 3.841e-05 ***
## NENFERM    11.283  1  6.0442  0.016393 *
## FSD         0.144  1  0.0770  0.782217
## Residuals 132.537 71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Las sumas SS2 corresponden a las sumas de cuadrados extras de cada variable ex- plicatoria en el modelo, dadas las demas, decir:

1) $SSR(EDAD|RINF, RRX, NCAMAS, AEM, PDP, NENFERM, FSD) = 18.125$: Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable edad al modelo de la longitud de permanencia dadas las demas variables en el modelo.

2) $SSR(RINF|EDAD, RRX, NCAMAS, AEM, PDP, NENFERM, FSD) = 17.306$: Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable riesgo de infeción al modelo de la longitud de permanencia dadas las demas variables en el modelo.

3) $SSR(RRX|EDAD, RINF, NCAMAS, AEM, PDP, NENFERM, FSD) = 2.070$: Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable Razon de rutina de rayos X al modelo de la longitud de permanencia dadas las demas variables en el modelo.

Procediendo de la misma manera para cada uno de los casos, para el ultimo caso se tendria que:

- $SSR(FSD|EDAD, RINF, RRX, NCAMAS, AEM, PDP, NENFERM) = 0.144$: Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable facilidad y servicios disponibles al modelo de la longitud de permanencia dadas las demas variables en el modelo. En este caso al mirar el la suma de cuadrados de la regresión la menor fue al ingresar FSD dado que las demas predictorias estaban presentes en el modelo, note que para este caso en especifico, coincide con la suma de cuadrados Tipo I ya que el SSR calculado en teoría fue el mismo (0.144) y no hubo otro valor superior a este luego de realizar la suma de cuadrados tipo II

Las sumas de cuadrados secuenciales son especialmente utiles para realizar pruebas de hipotesis con una prueba F, esta prueba esta dada por $F_0 = \frac{SSR(X_j|X_1,X_2,...,X_k)}{MSE(X_1,X_2,...X_k)}$, Según esto un valor muy pequeño del SSR va a dar un valor pequeño en la F_0, por lo tanto, no se va a rechazar la hipotesis nula, lo que significa que el modelo no es significativo, en este caso, se puede afirmar que tiene sentido que la SSR(FSD|EDAD,RINF,RRX,NCAMAS,AEM,PDP,NENFERM)=0.144 tuviera menor suma de cuadrados extras ya que la covariable FSD no es significativa en presencia de los demás parametros.

A continuación se hara la prueba de significancia para sustentar esta idea

se quiere probar:

$$H_0 : \beta_8 = 0$$

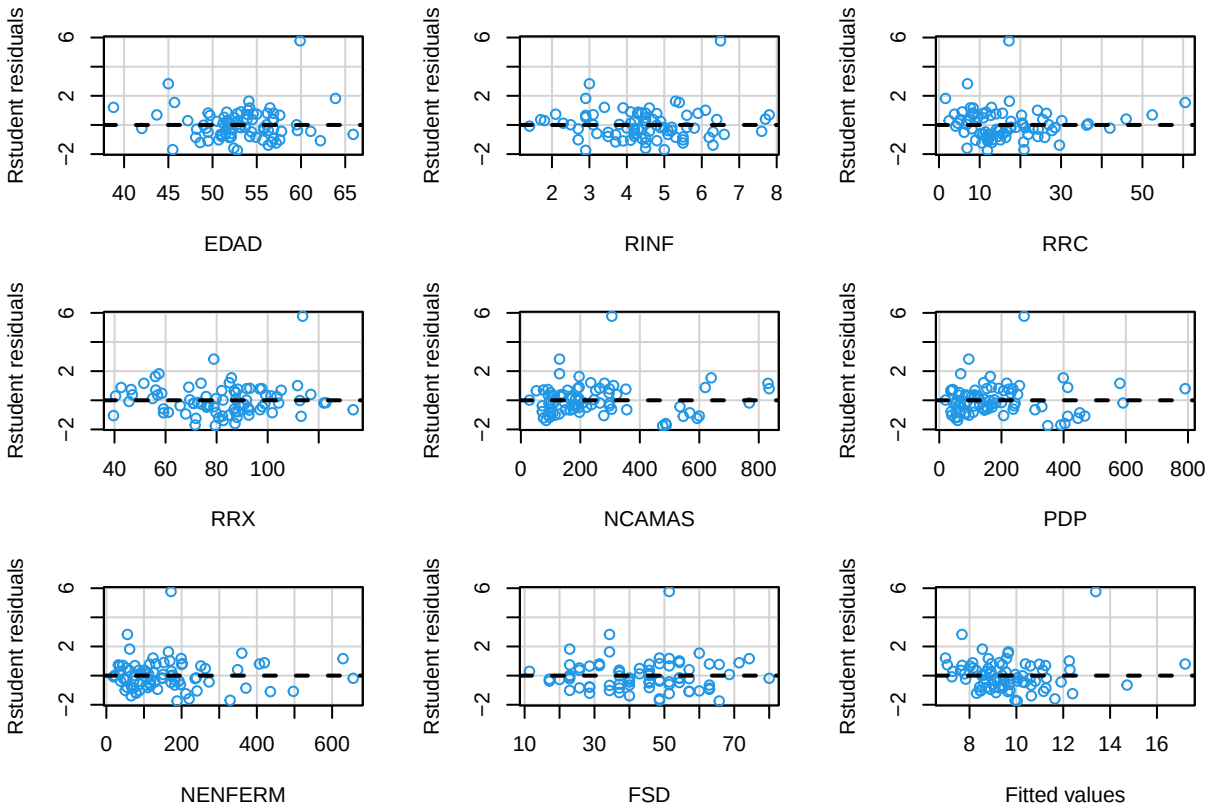
$$vs$$

$$H_a : \beta_8 \neq 0$$

La $F_0 = \frac{SSR(FSD|EDAD,RINF,RRX,NCAMAS,AEM,PDP,NENFERM)}{MSE(EDAD,RINF,RRX,NCAMAS,AEM,PDP,NENFERM,FSD)} = \frac{0.144}{1.8667} = 0.0770$

se rechaza H_0 si $F_0 > F_0.05, 1, 71$=3.98, como 0.0770<3.98, no se rechaza la hipotesis nula y se concluye que β_8 no es significativa en presencia de los demas parametros, esto quiere decir que no tiene un efecto significativo en el modelo de regresión cuando las demas variables permanencen constantes y se confirma lo mencionado anteriormente

5. **Graficos de residuales estudentizados vs valores ajustados**



Interpretación

1) Se observa que los datos son aleatorios alrededor de 0, no se identifican patrones dentro de las graficas de estudentizados vs valores ajustados lo que indica que no hay problema de varianza constante.

2) Se evidencian valores alejados de la nube de puntos, en un principio se perciben puntos atipicos(estan por encima de 3). Por ejemplo, en la grafica de EDAD vs Residuales estudentizados, se ve claramente un valor en aproximadamente x= 60 que toma valores cercanos a 6. En general todas las graficas muestran un punto atípico.

3) se observan algunos puntos alejados en direccion del eje x que da indicios de la existencia de puntos de balanceo, como en la grafica de NENFERM, valores >600.

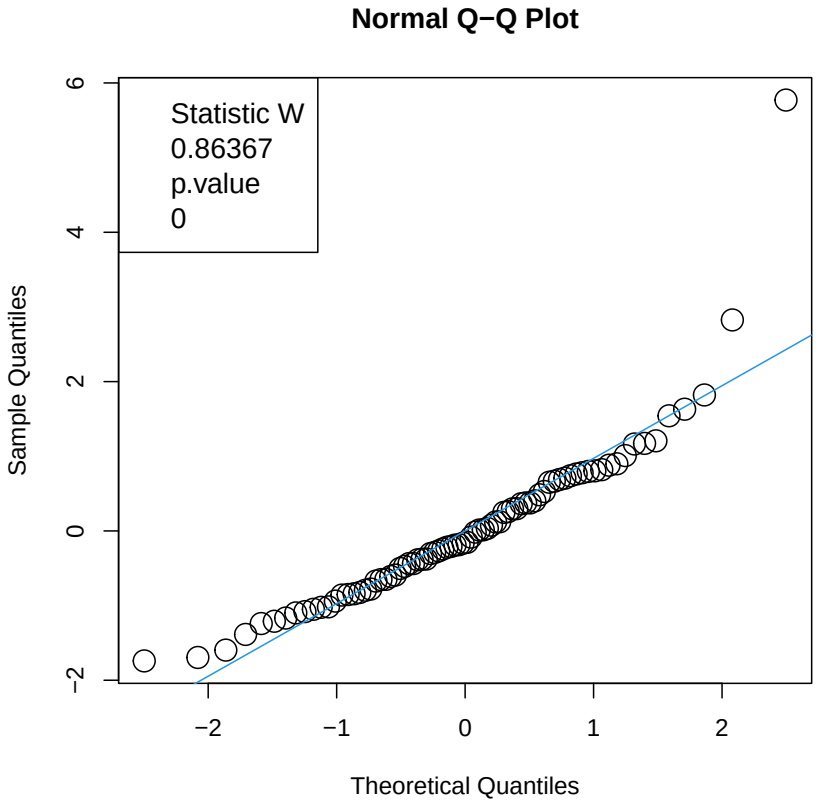
4) No se observan puntos influenciales a primera vista, sin embargo en la gráfica de residuales estudentizados vs valores ajustados se puede pensar que la observación atípica tambíen es un punto influencial.

Graficamente se da una aproximación de diagnosticos acerca del modelo, sin embargo estos se deben respaldar con pruebas estadisticas.

6. **Gráfico de normalidad**

$$H_0 : \varepsilon_i \sim N(0, \sigma^2)$$

$$H_1 : \varepsilon_i \nsim N(0, \sigma^2)$$



En un principio los residuales se ajustan a la línea azul que representa el ajuste de la distribución de los residuales a una distribución normal, se podría asegurar normalidad, sin embargo, esta normalidad se ve

afectada por el dato atipico del extremo derecho por lo que en la prueba de normalidad de Shapiro Wilk no se cumple el criterio de normalidad. se recomienda realizar un analisis de observaciones anomalas.

7. **Tabla de diagnosticos**

```
##        Y      ri      ei   se.yhat residuals hii.value
## 1   8.84 -1.3864 -1.3774 0.4034951   -1.7980    0.0872
## 2  11.07 -0.1843 -0.1856 0.7626285   -0.2104    0.3116
## 3  12.78  0.4032  0.4056 0.6127881    0.4954    0.2012
## 4  11.62  0.3750  0.3773 0.3345115    0.4998    0.0599
## 5   9.31  0.2980  0.2999 0.3647897    0.3949    0.0713
## 6   8.19 -0.5915 -0.5942 0.3056650   -0.7912    0.0501
## 7  11.77  1.6303  1.6116 0.3973442    2.1067    0.0846
## 8  13.59  1.0090  1.0089 0.3498376    1.3325    0.0656
## 9  10.33 -0.1901 -0.1914 0.3337987   -0.2535    0.0597
## 10 11.48  0.6658  0.6684 0.3596103    0.8811    0.0693
## 11  8.63  0.7063  0.7088 0.2879326    0.9466    0.0444
## 12 11.15  1.1723  1.1692 0.2472962    1.5710    0.0328
## 13  8.03 -0.7949 -0.7970 0.3217807   -1.0583    0.0555
## 14 10.05  0.0768  0.0773 0.6101839    0.0945    0.1995
## 15  8.90  0.6522  0.6549 0.3061913    0.8720    0.0502
## 16  8.54  0.0143  0.0144 0.5183814    0.0182    0.1440
## 17 12.01  0.7789  0.7811 0.3579405    1.0299    0.0686
## 18 13.95 -0.6489 -0.6516 0.6765612   -0.7735    0.2452
## 19 19.56  5.7709  4.7843 0.4494079    6.1729    0.1082
## 20 17.94  0.7969  0.7989 1.0092471    0.7358    0.5457
## 21  9.76 -0.6518 -0.6544 0.3003712   -0.8723    0.0483
## 22 11.18  1.5422  1.5274 0.9266575    1.5336    0.4600
## 23  7.58 -1.1670 -1.1641 0.3028647   -1.5509    0.0491
## 24  9.06  0.2587  0.2604 0.2705362    0.3487    0.0392
## 25 10.24 -0.0160 -0.0161 0.4135174   -0.0209    0.0916
## 26  8.28 -1.0983 -1.0967 0.5879826   -1.3526    0.1852
## 27  9.89  0.2485  0.2501 0.2888121    0.3340    0.0447
## 28  9.84 -0.1534 -0.1545 0.4613979   -0.1987    0.1140
## 29  9.74 -0.4749 -0.4775 0.4241311   -0.6201    0.0964
## 30 10.47  0.8992  0.9004 0.3631727    1.1860    0.0707
## 31  8.37 -1.0231 -1.0227 0.3195378   -1.3586    0.0547
## 32 10.90 -1.2408 -1.2361 0.6310163   -1.4980    0.2133
## 33  9.23  0.8806  0.8820 0.5662476    1.0967    0.1718
## 34 12.07  0.6888  0.6913 0.6642204    0.8254    0.2363
## 35 10.02  0.8203  0.8222 0.3385614    1.0883    0.0614
## 36  7.39 -0.7789 -0.7810 0.2674583   -1.0465    0.0383
## 37  8.45  1.2069  1.2030 0.6025264    1.4752    0.1945
## 38  8.88 -0.3029 -0.3049 0.2163343   -0.4113    0.0251
## 39 10.30 -0.3864 -0.3887 0.4181206   -0.5056    0.0937
## 40  7.94 -0.5029 -0.5056 0.3583358   -0.6666    0.0688
## 41 10.39  0.7617  0.7640 0.4419986    0.9877    0.1047
## 42  8.02  0.7385  0.7409 0.4005544    0.9678    0.0859
## 43  8.34 -0.2674 -0.2692 0.3829509   -0.3531    0.0786
## 44  9.68 -0.4375 -0.4401 0.2428889   -0.5917    0.0316
## 45  8.67 -0.2884 -0.2903 0.2523286   -0.3898    0.0341
## 46  9.00 -0.3798 -0.3821 0.3678501   -0.5028    0.0725
## 47  9.84 -1.0810 -1.0797 0.6051322   -1.3226    0.1962
## 48  8.48 -0.6717 -0.6744 0.2852136   -0.9011    0.0436
## 49 11.20  2.8258  2.6963 0.4012494    3.5214    0.0862
```

```
## 50  7.67  0.2955  0.2974 0.5259022    0.3751    0.1482
## 51  8.88 -0.6102 -0.6129 0.2531235   -0.8229    0.0343
## 52 11.41 -0.4373 -0.4398 0.7613280   -0.4989    0.3105
## 53 11.46  0.8046  0.8066 0.3052768    1.0742    0.0499
## 54  7.78 -1.6949 -1.6730 0.4346069   -2.1671    0.1012
## 55  9.61 -1.5964 -1.5793 0.4506822   -2.0370    0.1088
## 56  9.53 -0.3765 -0.3788 0.2923227   -0.5056    0.0458
## 57  8.09  0.3640  0.3662 0.4709427    0.4697    0.1188
## 58  9.05  0.1225  0.1233 0.2406542    0.1658    0.0310
## 59  7.91 -1.7412 -1.7168 0.5801034   -2.1237    0.1803
## 60  8.86  0.5282  0.5309 0.3191570    0.7052    0.0546
## 61  8.66  0.0342  0.0345 0.2287318    0.0464    0.0280
## 62  7.95  0.1202  0.1210 0.3510664    0.1598    0.0660
## 63 10.15 -0.8592 -0.8608 0.4564871   -1.1084    0.1116
## 64  9.44  0.4839  0.4865 0.3538570    0.6420    0.0671
## 65 10.80  1.8209  1.7919 0.5364336    2.2517    0.1542
## 66  7.14 -0.0828 -0.0834 0.4515430   -0.1075    0.1092
## 67  9.50 -0.2189 -0.2203 0.5573758   -0.2749    0.1664
## 68  9.41  0.0121  0.0122 0.4589593    0.0157    0.1128
## 69  7.13 -1.0498 -1.0491 0.4747409   -1.3440    0.1207
## 70  8.95 -0.1643 -0.1654 0.5355435   -0.2079    0.1536
## 71  8.28 -0.8519 -0.8535 0.3058993   -1.1365    0.0501
## 72  7.53 -0.2419 -0.2436 0.5201698   -0.3077    0.1449
## 73  9.20 -0.2081 -0.2095 0.3155275   -0.2785    0.0533
## 74 10.16  1.1642  1.1613 0.7085409    1.3566    0.2689
## 75  6.70 -1.2098 -1.2059 0.3188796   -1.6021    0.0545
## 76  7.63 -0.8242 -0.8261 0.4258703   -1.0725    0.0972
## 77  8.77  0.3751  0.3774 0.4448808    0.4875    0.1060
## 78  8.15 -0.8416 -0.8433 0.3570832   -1.1121    0.0683
## 79  7.14 -1.0171 -1.0169 0.4485025   -1.3124    0.1078
## 80  7.70 -0.9400 -0.9407 0.4946272   -1.1981    0.1311
```

Observaciones atipicas

Se asume que la observación i es atipica si un ei grande ($|ei| > 3$) y Se considera potencialmente atipica con ri grande ($|ri| > 3$) Deacuerdo a la columna ei y ri se observa que la observación 19 es atípica.

Observaciones de balanceo

Se asume que la observación i es un punto de balanceo si hii $> 2p/n$. En esta práctica tenemos como criterio que: hii $> 2(k+1)/n = 2(9/80) = 0.225$. De acuerdo a la columna hii.value la observación

- 2=0.3116
- 18=0.2452
- 20=0.5457
- 22=0.46
- 34=0.2363
- 54=0.3105
- 74=0.2689

son puntos de balanceo

Observaciones influenciales

```
##    dfb.1_ dfb.EDAD dfb.RINF dfb.RRC dfb.RRX dfb.NCAM dfb.PDP dfb.NENF dfb.FSD
```

```
## 1   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 2   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 3   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 4   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 5   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 6   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 7   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 8   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 9   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 10  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 11  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 12  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 13  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 14  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 15  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 16  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 17  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 18  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 19   TRUE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 20  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 21  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 22  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 23  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 24  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 25  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 26  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 27  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 28  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 29  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 30  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 31  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 32  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 33  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 34  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 35  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 36  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 37  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 38  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 39  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 40  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 41  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 42  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 43  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 44  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 45  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 46  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 47  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 48  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 49  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 50  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 51  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 52  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 53  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 54  FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
```

```
## 55  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 56  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 57  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 58  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 59  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 60  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 61  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 62  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 63  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 64  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 65  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 66  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 67  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 68  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 69  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 70  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 71  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 72  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 73  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 74  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 75  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 76  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 77  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 78  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 79  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
## 80  FALSE    FALSE    FALSE   FALSE   FALSE    FALSE   FALSE    FALSE   FALSE
##    dffit cov.r cook.d   hat
## 1  FALSE FALSE  FALSE FALSE
## 2  FALSE  TRUE  FALSE FALSE
## 3  FALSE  TRUE  FALSE FALSE
## 4  FALSE FALSE  FALSE FALSE
## 5  FALSE FALSE  FALSE FALSE
## 6  FALSE FALSE  FALSE FALSE
## 7  FALSE FALSE  FALSE FALSE
## 8  FALSE FALSE  FALSE FALSE
## 9  FALSE FALSE  FALSE FALSE
## 10 FALSE FALSE  FALSE FALSE
## 11 FALSE FALSE  FALSE FALSE
## 12 FALSE FALSE  FALSE FALSE
## 13 FALSE FALSE  FALSE FALSE
## 14 FALSE  TRUE  FALSE FALSE
## 15 FALSE FALSE  FALSE FALSE
## 16 FALSE FALSE  FALSE FALSE
## 17 FALSE FALSE  FALSE FALSE
## 18 FALSE  TRUE  FALSE FALSE
## 19  TRUE  TRUE  FALSE FALSE
## 20 FALSE  TRUE  FALSE  TRUE
## 21 FALSE FALSE  FALSE FALSE
## 22  TRUE  TRUE  FALSE  TRUE
## 23 FALSE FALSE  FALSE FALSE
## 24 FALSE FALSE  FALSE FALSE
## 25 FALSE FALSE  FALSE FALSE
## 26 FALSE FALSE  FALSE FALSE
## 27 FALSE FALSE  FALSE FALSE
```

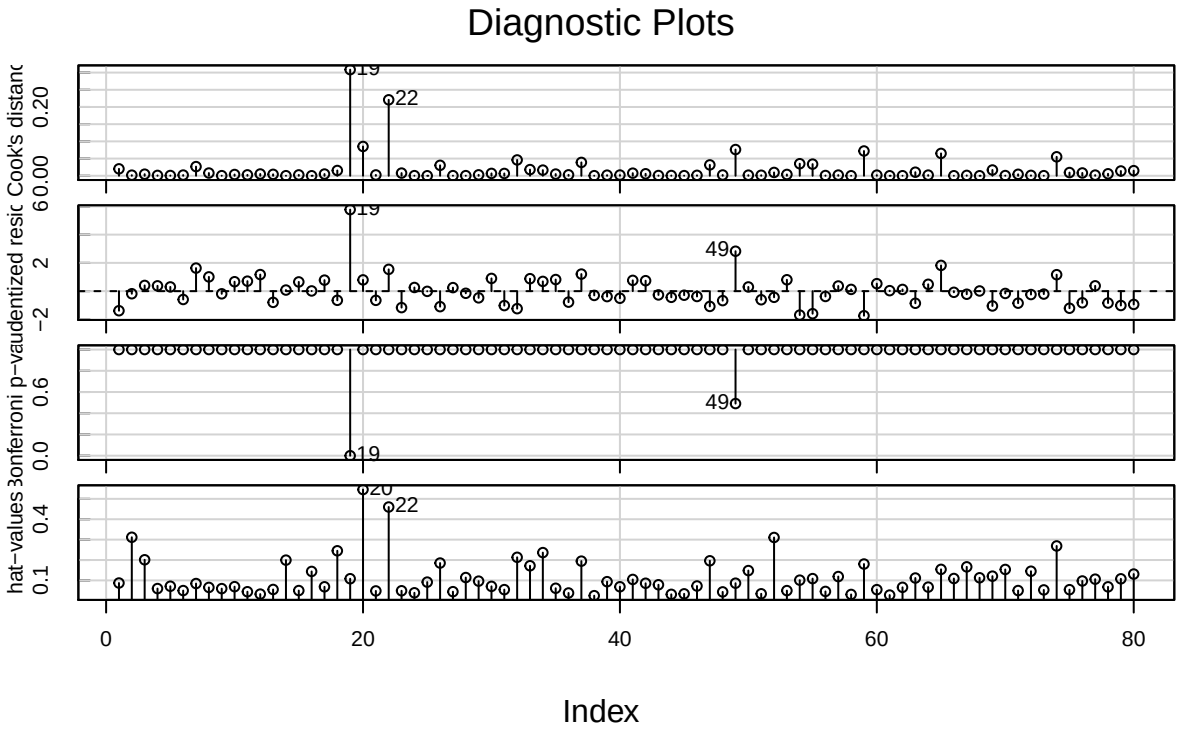
```
## 28 FALSE FALSE  FALSE FALSE
## 29 FALSE FALSE  FALSE FALSE
## 30 FALSE FALSE  FALSE FALSE
## 31 FALSE FALSE  FALSE FALSE
## 32 FALSE FALSE  FALSE FALSE
## 33 FALSE FALSE  FALSE FALSE
## 34 FALSE  TRUE  FALSE FALSE
## 35 FALSE FALSE  FALSE FALSE
## 36 FALSE FALSE  FALSE FALSE
## 37 FALSE FALSE  FALSE FALSE
## 38 FALSE FALSE  FALSE FALSE
## 39 FALSE FALSE  FALSE FALSE
## 40 FALSE FALSE  FALSE FALSE
## 41 FALSE FALSE  FALSE FALSE
## 42 FALSE FALSE  FALSE FALSE
## 43 FALSE FALSE  FALSE FALSE
## 44 FALSE FALSE  FALSE FALSE
## 45 FALSE FALSE  FALSE FALSE
## 46 FALSE FALSE  FALSE FALSE
## 47 FALSE FALSE  FALSE FALSE
## 48 FALSE FALSE  FALSE FALSE
## 49 FALSE  TRUE  FALSE FALSE
## 50 FALSE FALSE  FALSE FALSE
## 51 FALSE FALSE  FALSE FALSE
## 52 FALSE  TRUE  FALSE FALSE
## 53 FALSE FALSE  FALSE FALSE
## 54 FALSE FALSE  FALSE FALSE
## 55 FALSE FALSE  FALSE FALSE
## 56 FALSE FALSE  FALSE FALSE
## 57 FALSE FALSE  FALSE FALSE
## 58 FALSE FALSE  FALSE FALSE
## 59 FALSE FALSE  FALSE FALSE
## 60 FALSE FALSE  FALSE FALSE
## 61 FALSE FALSE  FALSE FALSE
## 62 FALSE FALSE  FALSE FALSE
## 63 FALSE FALSE  FALSE FALSE
## 64 FALSE FALSE  FALSE FALSE
## 65 FALSE FALSE  FALSE FALSE
## 66 FALSE FALSE  FALSE FALSE
## 67 FALSE FALSE  FALSE FALSE
## 68 FALSE FALSE  FALSE FALSE
## 69 FALSE FALSE  FALSE FALSE
## 70 FALSE FALSE  FALSE FALSE
## 71 FALSE FALSE  FALSE FALSE
## 72 FALSE FALSE  FALSE FALSE
## 73 FALSE FALSE  FALSE FALSE
## 74 FALSE FALSE  FALSE FALSE
## 75 FALSE FALSE  FALSE FALSE
## 76 FALSE FALSE  FALSE FALSE
## 77 FALSE FALSE  FALSE FALSE
## 78 FALSE FALSE  FALSE FALSE
## 79 FALSE FALSE  FALSE FALSE
## 80 FALSE FALSE  FALSE FALSE
```

Para identificar estos valores utilizaremos 3 criterios, que son:

- Se dice que la observacion sera influencial si Di > 1.
- una observacion sera influencial si |DFFITS| $> 2(p/n)\hat{}0.5$
- observaciones con un covratio tal que |COVRATIO-1| $> 3(p/n)$ son cadidatas a ser influenciales
- De acuerdo a la columna Cooks.D de distancias de Cook tenemos que la observacion 19 es influencial
- De acuerdo a la columna Dffits de valores DFFITS tenemos que las observaciones 19 y 22 son influenciales.
- De acuerdo a la columna Covratio de observaciones covratio tenemos que 2,3,14,18,19,20,22,34,49,52 son influenciales

Las observaciones con ID=47 y ID=112 corresponden con las observaciones 19 y 20 respectivamente, según el análisis anterior se observa que la observación 20 es un punto de balanceo esto quiere decir que es una observacion en el espacio de las predictorias alejada del resto de la muestra y que puede controlar ciertas propiedades del modelo ajustado, además se tiene que la observación 20 y 19 son también puntos influenciales osea que tienen un impacto notable sobre los coeficientes de regresión ya que jalan el modelo a su dirección, tambien se observa que la observación 19 es un dato atipico que puede afectar el ajuste del modelo de regresión, en general la observación 19 podría generar un impacto notable negativo al modelo ya que es un punto atipico, de balanceo e influencial.

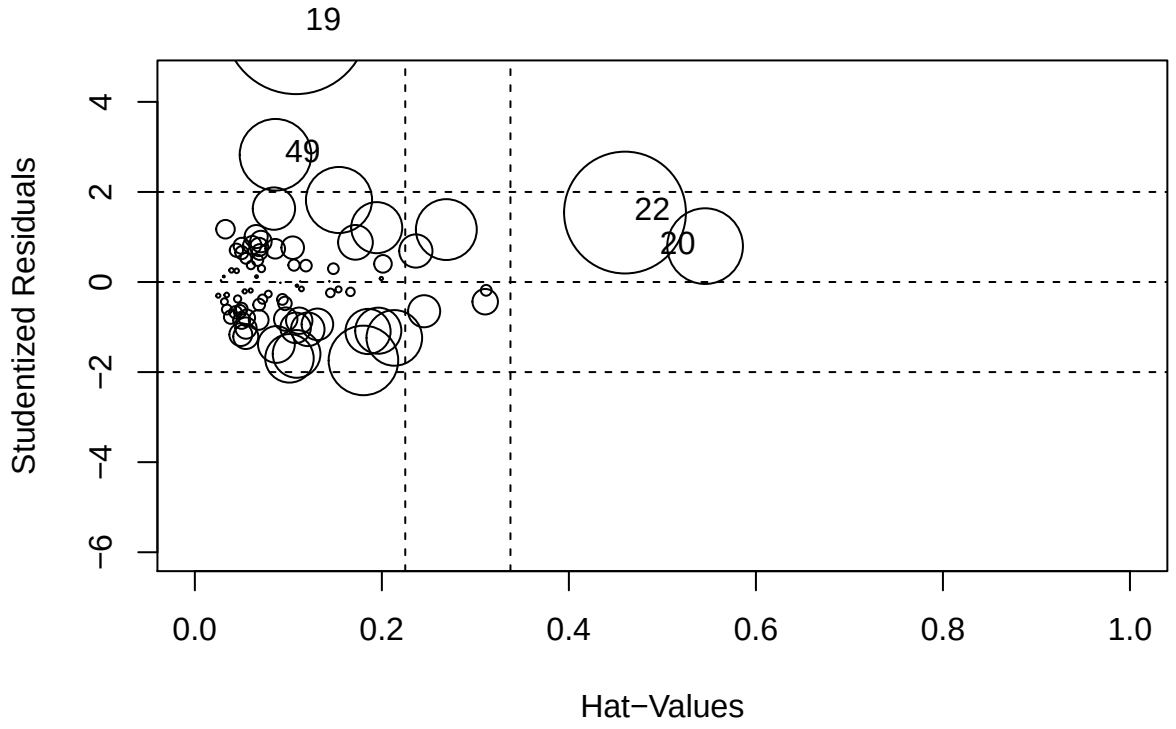
Gráficas de chequeos y diagnosticos



1) En la primera gráfica de distancias de cook, se observa que ningún valor es mayor a 1, por lo tanto no se puede concluir puntos de balanceo.

2) la segunda y tercer grafica muestra a el punto 19 como un punto atipico. La observacion numero 49 aparece con un valor alto pero no lo suficiente para considerarla como un dato atipico.

3) En la gráfica de hat.values se ve que la observación 20 y 22 que están por encima de 0.4, por lo tanto son puntos de balanceo




```
##      StudRes        Hat      CookD
## 19 5.7708858 0.10819423 0.30854688
## 20 0.7968692 0.54565457 0.08517293
## 22 1.5422011 0.46000359 0.22083064
## 49 2.8257952 0.08624851 0.07624458
```

En el grafico circular, se confirma la presencia de puntos de balanceo, en particular las observaciones 20 y 22 que estan significativamente alejadas del resto (en el eje horizontal). Ademas, el circulo correspondiente a la observacion 19 se aleja muy por encima de los demás, ratificando su condicion de dato atipico.

Las observaciones con ID=47 y ID=112 corresponden con las observaciones 19 y 20 respectivamente, por todo lo anterior mencionado en el análisis anterior se observa que la observación 20 es un punto de balanceo esto quiere decir que es una observacion en el espacio de las predictorias alejada del resto de la muestra y que puede controlar ciertas propiedades del modelo ajustado, además se tiene que la observación 20 y 19 son también puntos influenciales osea que tienen un impacto notable sobre los coeficientes de regresión ya que jalan el modelo a su dirección.

Modelo sin observaciones 19 y 20

```
##
## Call:
## lm(formula = DPERM ~ EDAD + RINF + RRC + RRX + NCAMAS + PDP +
##     NENFERM + FSD, data = datar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9405 -0.6319 -0.0813  0.6570  3.3962
##
```

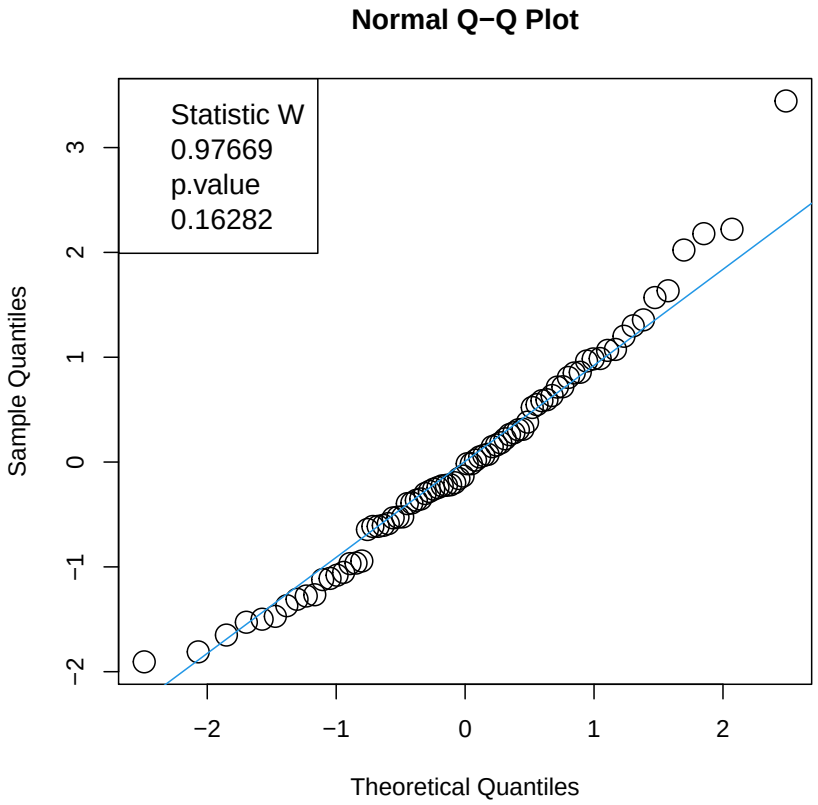
```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.828477   1.639030   0.505  0.61484
## EDAD         0.088895   0.029708   2.992  0.00384 **
## RINF         0.410158   0.129190   3.175  0.00224 **
## RRC          0.021844   0.015706   1.391  0.16874
## RRX          0.013878   0.007486   1.854  0.06804 .
## NCAMAS      -0.005095   0.004227  -1.205  0.23225
## PDP          0.013167   0.005623   2.342  0.02209 *
## NENFERM     -0.004729   0.002626  -1.801  0.07610 .
## FSD          0.005044   0.013995   0.360  0.71963
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.11 on 69 degrees of freedom
## Multiple R-squared:  0.5333, Adjusted R-squared:  0.4792
## F-statistic: 9.856 on 8 and 69 DF,  p-value: 5.055e-09
```

En efecto, existe un cambio sustancial en los parametros del modelo ajustado sin las observaciones exlcuidas, esto puede explicarse dado que la observación 19 era un punto atipico.

- El error estandar resitual pasó de 1.366 a 1.11
- El R^2 ajustado pasó de 0.5829 a 0.4792

La proporcion de variabilidad que el modelo ajustado explica en los datos disminuyó a un 47%, sin embargo; el residual estimado tambien disminuyó.

Normalidad modelo reducido



Ahora, los datos se ajustan mejor a la linea (cerca de un 97%), acorde a la prueba de Shapiro-Wilk; se concluye que sin las obsrevaciones 19 y 20, los residuos del modelo ajustado se distribuyen normal y se confirma que la observación 19 afectaba la prueba de Shapiro-Wilk.

NOTA: Profe, le deseo suerte calificando estos trabajos. Si este es largo, no me imagino como serán los otros. Exitos y fuerza en esta tarea tan larga.

8. **Multicolinealidad para modelo full**

Matriz de correlación de variables predictorias

```
##             DPERM        EDAD      RINF        RRC          RRX      NCAMAS
## DPERM   1.0000000  0.30799062 0.5674545  0.2982259  0.41410472 0.41417527
## EDAD    0.3079906  1.00000000 0.1005422 -0.2223435 -0.01000860 0.07916828
## RINF    0.5674545  0.10054220 1.0000000  0.5400434  0.45797068 0.33886636
## RRC     0.2982259 -0.22234346 0.5400434  1.0000000  0.48844754 0.10343079
## RRX     0.4141047 -0.01000860 0.4579707  0.4884475  1.00000000 0.06901589
## NCAMAS  0.4141753  0.07916828 0.3388664  0.1034308  0.06901589 1.00000000
## PDP     0.4881007  0.09919355 0.3498938  0.0952719  0.09164430 0.97933716
## NENFERM 0.3176143  0.07081883 0.3524609  0.1411993  0.09204234 0.91155110
## FSD     0.3473883  0.12628933 0.3550138  0.1389761  0.13699210 0.75454053
##                PDP    NENFERM       FSD
## DPERM   0.48810067 0.31761427 0.3473883
## EDAD    0.09919355 0.07081883 0.1262893
## RINF    0.34989385 0.35246089 0.3550138
## RRC     0.09527190 0.14119931 0.1389761
```

```
## RRX     0.09164430 0.09204234 0.1369921
## NCAMAS  0.97933716 0.91155110 0.7545405
## PDP     1.00000000 0.88798442 0.7346366
## NENFERM 0.88798442 1.00000000 0.7747349
## FSD     0.73463664 0.77473494 1.0000000
```

Se detecta una asociación lineal entre NCAMAS y PDP, NCAMAS y NENFERM, NCAMAS y FSD, PDP y NENFERM, FSD y PDP, FSD y NENFERM porque nos da un valor de correlación de 0.9793, 0.9111,0.7545, 0.887, 0.7346 y 0.7747 respectivamente.

VIF'S Para analizar problemas de multicolinealidad con los VIF's, se analiza la siguiente tabla:

```
##      EDAD      RINF       RRC       RRX    NCAMAS       PDP   NENFERM       FSD
##  1.154238  1.845165  1.795919  1.450856 31.972847 25.561085  6.676044  2.680535
```

Para VIF's con valores >10 indica que hay problemas de multicolinealidad. Según este criterio se detecta problemas de multicolinealidad ya que la variable PDP tiene un valor de 25.56.

Proporciones de varianza

Como en los datos β_0 no tiene interpretabilidad, se trabaja con los datos centrados. Para ello:

```
##   Val.propio cond.index      Pi.EDAD     Pi.RINF      Pi.RRC      Pi.RRX
## 1 3.81943910   1.000000 0.0007993112 0.011495431 0.003388687 0.003356113
## 2 1.79814710   1.457428 0.0170524930 0.058985070 0.113169255 0.118632172
## 3 1.03763495   1.918569 0.7103524829 0.031543972 0.012239881 0.026446605
## 4 0.52703742   2.692025 0.0132768364 0.238942629 0.133462845 0.813807308
## 5 0.35909760   3.261321 0.1953573907 0.611176182 0.653776646 0.013166556
## 6 0.32176467   3.445328 0.0565225181 0.042233548 0.064977719 0.004276321
## 7 0.11874468   5.671432 0.0011247237 0.001601290 0.013101782 0.002887946
## 8 0.01813448  14.512666 0.0055142439 0.004021878 0.005883184 0.017426980
##      Pi.NCAMAS       Pi.PDP   Pi.NENFERM       Pi.FSD
## 1 1.891082e-03 0.0023300974 0.0088309100 1.856022e-02
## 2 5.627655e-04 0.0006425017 0.0017908241 2.191849e-03
## 3 3.165530e-04 0.0002054415 0.0014716425 2.290033e-05
## 4 7.349521e-05 0.0002497838 0.0001662117 4.079627e-03
## 5 1.766970e-04 0.0008707253 0.0009670766 7.608678e-02
## 6 8.719051e-03 0.0146879499 0.0059253151 8.013412e-01
## 7 1.872581e-02 0.0585752051 0.9063134931 9.061810e-02
## 8 9.695345e-01 0.9224382953 0.0745345268 7.099280e-03
```

Segun el indice de condicion, existe problemas graves de multicolinealidad si dicho indice es mayor de 31. En ninguna de estas variables hay problemas graves de multicolinealidad. Ahora, hay un valor (octava fila) de casi 15. Esto implica que hay problemas leves de multicolinealidad.

Las otras columnas representan la descomposicion de varianzas para cada una de las variables y, se dice que existe problemas de multicolnialidad entre dos variables cuando los valores superen 0.5

De esta manera se tiene que existe problemas de multicolinealidad entre RINF y RRC pues los dos valores anexos en su quinta fila superan 0.5 (0.611176182 y 0.653776646)

Finalmente se concluye que existen problemas de multicolinealidad leve entre las variables RINF y RRC.

Multicolinealidad para modelo sin observaciones 19 y 20

Matriz de correlación de variables predictorias

```
##             DPERM        EDAD       RINF          RRC          RRX     NCAMAS
## DPERM   1.0000000  0.26211098 0.59787420  0.35320007  0.41698764 0.34862410
## EDAD    0.2621110  1.00000000 0.06463682 -0.23552665 -0.04552916 0.04976037
## RINF    0.5978742  0.06463682 1.00000000  0.54234607  0.43763604 0.31602477
## RRC     0.3532001 -0.23552665 0.54234607  1.00000000  0.49324268 0.07332880
## RRX     0.4169876 -0.04552916 0.43763604  0.49324268  1.00000000 0.04586272
## NCAMAS  0.3486241  0.04976037 0.31602477  0.07332880  0.04586272 1.00000000
## PDP     0.3825332  0.06011336 0.32502720  0.05614816  0.06099174 0.98399790
## NENFERM 0.3109517  0.05549688 0.34020687  0.12372867  0.08183184 0.91627723
## FSD     0.3484597  0.10769404 0.33775745  0.12626755  0.12111975 0.75876374
##                PDP    NENFERM       FSD
## DPERM   0.38253316 0.31095172 0.3484597
## EDAD    0.06011336 0.05549688 0.1076940
## RINF    0.32502720 0.34020687 0.3377575
## RRC     0.05614816 0.12372867 0.1262676
## RRX     0.06099174 0.08183184 0.1211198
## NCAMAS  0.98399790 0.91627723 0.7587637
## PDP     1.00000000 0.91469606 0.7584934
## NENFERM 0.91469606 1.00000000 0.7699048
## FSD     0.75849336 0.76990477 1.0000000
```

Se detecta una asociación lineal entre NCAMAS y PDP, NCAMAS y NENFERM, NCAMAS y FSD, PDP Y NENFERM, FSD y PDP, FSD y NENFERM ya que presentan valores de correlación de 0.98399790,0.91627723,0.75876374, 0.916277230, 0.758493360 y 0.7699048 respectivamente.

VIF'S

```
##      EDAD      RINF       RRC       RRX    NCAMAS       PDP   NENFERM       FSD
##  1.137185  1.803994  1.879683  1.430469 34.340725 34.359257  7.106006  2.653071
```

Para VIF's con valores >10 indica que hay problemas de multicolinealidad.Según este criterio se detecta problemas de multicolinealidad en almenos dos variables ya que tiene un valor de 34.359(PDP) Y 34.3407(NCAMAS) respectivamente sin embargo, no se tiene certeza de cuales son las que presentan dicho problema.

Proporciones de varianza

```
##   Val.propio cond.index      Pi.EDAD     Pi.RINF      Pi.RRC       Pi.RRX
## 1 3.79530088   1.000000 3.465940e-04 0.010480595 0.002476588 0.0024986796
## 2 1.84838167   1.432938 2.002664e-02 0.060802436 0.107078655 0.1155609942
## 3 1.02504400   1.924207 7.390268e-01 0.036912946 0.006391669 0.0238425248
## 4 0.54348631   2.642584 3.929304e-03 0.256587834 0.104017022 0.8097687327
## 5 0.36489137   3.225086 1.603262e-01 0.574578220 0.638148976 0.0280824656
## 6 0.30550268   3.524647 7.407117e-02 0.048419153 0.083611334 0.0029840755
## 7 0.10207440   6.097681 2.263565e-03 0.002709220 0.026603866 0.0003033364
## 8 0.01531869  15.740275 9.715052e-06 0.009509595 0.031671889 0.0169591911
##      Pi.NCAMAS       Pi.PDP   Pi.NENFERM       Pi.FSD
## 1 1.800421e-03 0.0018022225 0.0085936696 1.927416e-02
## 2 4.298181e-04 0.0004294900 0.0011903714 1.506240e-03
## 3 2.727590e-04 0.0001877814 0.0008377657 2.543906e-04
## 4 5.598216e-05 0.0001682869 0.0001617732 3.731138e-03
## 5 1.973613e-04 0.0007380377 0.0006578944 9.241563e-02
## 6 7.951285e-03 0.0073646587 0.0187957328 8.596159e-01
## 7 4.064871e-02 0.0404849890 0.9696863854 2.311876e-02
## 8 9.486437e-01 0.9488245338 0.0000764076 8.372702e-05
```

Segun el indice de condicion existe problemas de multicolinealidad leve en la fila 8, ya que toma un valor de 15.7502>10, hay problemas de multicolinealidad entre RINF y RRC pues los dos valores anexos en su quinta fila superan 0.5 (0.574578220 y 0.638148976)

Finalmente se concluye que existen problemas de multicolinealidad leve entre las variables RINF y RRC.

En terminos generales los problemas de multicolinealidad no mejoraron sin las observaciones 19 y 20, pues en ambos modelos hay presencia de multicolinealidad leve entre las variables RINF(Riesgo de infección) y RRC(Razón de rutina de cultivos.)

9. **Selección de variables**

Bajo el modelo ajustado sin las observaciones, se procede a realizar el analisis correspondiente al mejor modelo a ajustar acorde a los criterios solicitados.

comparando todos los posibles modelos y teniendo siempre presente el principio de parsimonia:

- Por criterio de R^2_{adj} y R^2, se selecciona el modelo con el valor mas alto
- Por ciriterio de MSE se selecciona el modelo con menor valor de este estadistico, aunque es equivalente con el anterior (A menor MSE mayor R^2_{adj} y R^2 asi que se esperan resultados similares)
- Por criterior de C_p para el valor mas pequeño de dicho estadistico; estadistico que está dado por:

$$C_p = \frac{SSE_p}{MSE(X_1, X_2, ..., X_k)} - (n - 2p)$$

Donde SSE_p es el SSE del moderlo de regresion con $p-1 \leq k$ variables predictoras y el MSE del denominador es el SSE con todas las k predictoras.

	Index	N	Predictors	R-Square	Adj. R-Square
2	1	1	RINF	0.35745356	0.34899900
4	2	1	RRX	0.17387869	0.16300867
6	3	1	PDP	0.14633161	0.13509914
3	4	1	RRC	0.12475029	0.11323384
5	5	1	NCAMAS	0.12153877	0.10998007
8	6	1	FSD	0.12142415	0.10986394
7	7	1	NENFERM	0.09669097	0.08480533
1	8	1	EDAD	0.06870216	0.05644824
9	9	2	EDAD RINF	0.40760025	0.39180292
19	10	2	RINF PDP	0.39705984	0.38098144
17	11	2	RINF RRX	0.38729912	0.37096043
18	12	2	RINF NCAMAS	0.38578067	0.36940149
21	13	2	RINF FSD	0.38168719	0.36519884
20	14	2	RINF NENFERM	0.37053476	0.35374902
16	15	2	RINF RRC	0.35864052	0.34153760
28	16	2	RRX PDP	0.30187550	0.28325885
27	17	2	RRX NCAMAS	0.28267773	0.26354914
30	18	2	RRX FSD	0.26397717	0.24434989
24	19	2	RRC PDP	0.25671880	0.23689797
11	20	2	EDAD RRX	0.25305782	0.23313936
29	21	2	RRX NENFERM	0.25102954	0.23105699
10	22	2	EDAD RRC	0.25098424	0.23101048
23	23	2	RRC NCAMAS	0.22946436	0.20891675
26	24	2	RRC FSD	0.21857833	0.19774042
13	25	2	EDAD PDP	0.20371526	0.18248100
22	26	2	RRC RRX	0.20263908	0.18137612
25	27	2	RRC NENFERM	0.19728364	0.17587787

```
## 12     28 2                    EDAD NCAMAS 0.18159655    0.15977246
## 15     29 2                       EDAD FSD 0.17245394    0.15038605
## 31     30 2                     NCAMAS PDP 0.17065295    0.14853703
## 14     31 2                   EDAD NENFERM 0.15682974    0.13434520
## 34     32 2                    PDP NENFERM 0.15562005    0.13310325
## 35     33 2                        PDP FSD 0.15433785    0.13178686
## 33     34 2                     NCAMAS FSD 0.13814420    0.11516138
## 36     35 2                    NENFERM FSD 0.12589518    0.10258571
## 32     36 2                  NCAMAS NENFERM 0.12198747    0.09857380
## 38     37 3                   EDAD RINF RRX 0.44439958    0.42187524
## 40     38 3                   EDAD RINF PDP 0.44366124    0.42110697
## 64     39 3                    RINF RRX PDP 0.43417638    0.41123758
## 39     40 3                 EDAD RINF NCAMAS 0.43366505    0.41070553
## 42     41 3                   EDAD RINF FSD 0.42603351    0.40276460
## 63     42 3                  RINF RRX NCAMAS 0.42269391    0.39928961
## 70     43 3                 RINF PDP NENFERM 0.42064020    0.39715264
## 37     44 3                      EDAD RINF RRC 0.42031466    0.39681390
## 41     45 3                EDAD RINF NENFERM 0.41893077    0.39537391
## 67     46 3                  RINF NCAMAS PDP 0.41799618    0.39440143
## 66     47 3                      RINF RRX FSD 0.41328490    0.38949915
## 65     48 3                 RINF RRX NENFERM 0.40380587    0.37963584
## 60     49 3                      RINF RRC PDP 0.40132295    0.37705226
## 71     50 3                      RINF PDP FSD 0.39730100    0.37286725
## 68     51 3               RINF NCAMAS NENFERM 0.39396922    0.36940041
## 59     52 3                  RINF RRC NCAMAS 0.38886994    0.36409439
## 69     53 3                  RINF NCAMAS FSD 0.38808246    0.36327500
## 58     54 3                      RINF RRC RRX 0.38795345    0.36314075
## 62     55 3                      RINF RRC FSD 0.38378254    0.35880075
## 72     56 3                 RINF NENFERM FSD 0.38168871    0.35662203
## 61     57 3                 RINF RRC NENFERM 0.37241660    0.34697403
## 49     58 3                      EDAD RRX PDP 0.36913213    0.34355641
## 45     59 3                      EDAD RRC PDP 0.36481398    0.33906320
## 48     60 3                   EDAD RRX NCAMAS 0.35269482    0.32645272
## 44     61 3                   EDAD RRC NCAMAS 0.34082942    0.31410629
## 74     62 3                       RRC RRX PDP 0.32713132    0.29985286
## 51     63 3                      EDAD RRX FSD 0.32585152    0.29852118
## 50     64 3                 EDAD RRX NENFERM 0.32143247    0.29392298
## 47     65 3                      EDAD RRC FSD 0.31770495    0.29004435
## 83     66 3                    RRX NCAMAS PDP 0.31749151    0.28982225
## 86     67 3                   RRX PDP NENFERM 0.31678224    0.28908423
## 43     68 3                      EDAD RRC RRX 0.31362070    0.28579451
## 46     69 3                 EDAD RRC NENFERM 0.30813443    0.28008583
## 73     70 3                    RRC RRX NCAMAS 0.30535747    0.27719629
## 87     71 3                       RRX PDP FSD 0.30384550    0.27562303
## 77     72 3                    RRC NCAMAS PDP 0.29308722    0.26442860
## 85     73 3                    RRX NCAMAS FSD 0.28849680    0.25965208
## 84     74 3                 RRX NCAMAS NENFERM 0.28645082    0.25752315
## 76     75 3                       RRC RRX FSD 0.28555672    0.25659280
## 80     76 3                   RRC PDP NENFERM 0.28185273    0.25273865
## 75     77 3                   RRC RRX NENFERM 0.27163483    0.24210651
## 88     78 3                   RRX NENFERM FSD 0.26940645    0.23978779
## 81     79 3                       RRC PDP FSD 0.25894092    0.22889799
## 79     80 3                    RRC NCAMAS FSD 0.23824397    0.20736196
## 78     81 3                 RRC NCAMAS NENFERM 0.23413863    0.20309020
```

```
## 52     82 3                     EDAD NCAMAS PDP 0.22430888    0.19286194
## 82     83 3                      RRC NENFERM FSD 0.22147901    0.18991735
## 55     84 3                    EDAD PDP NENFERM 0.21306234    0.18115946
## 56     85 3                       EDAD PDP FSD 0.20819292    0.17609263
## 54     86 3                     EDAD NCAMAS FSD 0.19222904    0.15948157
## 53     87 3                  EDAD NCAMAS NENFERM 0.18233951    0.14919111
## 90     88 3                       NCAMAS PDP FSD 0.18204926    0.14888909
## 57     89 3                    EDAD NENFERM FSD 0.17833674    0.14502607
## 89     90 3                   NCAMAS PDP NENFERM 0.17460531    0.14114337
## 92     91 3                      PDP NENFERM FSD 0.17064097    0.13701831
## 91     92 3                   NCAMAS NENFERM FSD 0.14181607    0.10702483
## 99     93 4                     EDAD RINF RRX PDP 0.48789004    0.45982922
## 98     94 4                  EDAD RINF RRX NCAMAS 0.47783656    0.44922486
## 105    95 4                EDAD RINF PDP NENFERM 0.46696924    0.43776207
## 101    96 4                    EDAD RINF RRX FSD 0.46412681    0.43476390
## 95     97 4                    EDAD RINF RRC PDP 0.46381073    0.43443050
## 102    98 4                 EDAD RINF NCAMAS PDP 0.46153483    0.43202989
## 100    99 4                EDAD RINF RRX NENFERM 0.45916470    0.42952989
## 141   100 4                 RINF RRX PDP NENFERM 0.45876143    0.42910452
## 94    101 4                 EDAD RINF RRC NCAMAS 0.45135405    0.42129126
## 138   102 4                  RINF RRX NCAMAS PDP 0.45111036    0.42103422
## 93    103 4                    EDAD RINF RRC RRX 0.44712569    0.41683120
## 106   104 4                    EDAD RINF PDP FSD 0.44367570    0.41319217
## 103   105 4               EDAD RINF NCAMAS NENFERM 0.44258737    0.41204421
## 97    106 4                    EDAD RINF RRC FSD 0.44019370    0.40951938
## 104   107 4                 EDAD RINF NCAMAS FSD 0.43434459    0.40334977
## 142   108 4                     RINF RRX PDP FSD 0.43419912    0.40319634
## 129   109 4                     RINF RRC RRX PDP 0.43418001    0.40317617
## 96    110 4                EDAD RINF RRC NENFERM 0.43342870    0.40238370
## 144   111 4                RINF NCAMAS PDP NENFERM 0.43329305    0.40224061
## 139   112 4                RINF RRX NCAMAS NENFERM 0.43254579    0.40145241
## 135   113 4                   RINF RRC PDP NENFERM 0.42837064    0.39704849
## 107   114 4                EDAD RINF NENFERM FSD 0.42611131    0.39466535
## 132   115 4                  RINF RRC NCAMAS PDP 0.42538655    0.39390088
## 147   116 4                 RINF PDP NENFERM FSD 0.42429554    0.39275009
## 140   117 4                  RINF RRX NCAMAS FSD 0.42389138    0.39232378
## 128   118 4                  RINF RRC RRX NCAMAS 0.42276201    0.39113253
## 109   119 4                     EDAD RRC RRX PDP 0.42218585    0.39052480
## 145   120 4                  RINF NCAMAS PDP FSD 0.41901086    0.38717584
## 131   121 4                     RINF RRC RRX FSD 0.41351310    0.38137684
## 143   122 4                 RINF RRX NENFERM FSD 0.41347019    0.38133158
## 130   123 4                 RINF RRC RRX NENFERM 0.40415736    0.37150844
## 108   124 4                  EDAD RRC RRX NCAMAS 0.40265612    0.36992495
## 136   125 4                     RINF RRC PDP FSD 0.40146365    0.36866714
## 146   126 4                RINF NCAMAS NENFERM FSD 0.39963988    0.36674343
## 133   127 4                RINF RRC NCAMAS NENFERM 0.39797225    0.36498443
## 112   128 4                  EDAD RRC NCAMAS PDP 0.39761201    0.36460445
## 115   129 4                 EDAD RRC PDP NENFERM 0.39507604    0.36192952
## 134   130 4                  RINF RRC NCAMAS FSD 0.39103492    0.35766697
## 121   131 4                 EDAD RRX PDP NENFERM 0.38432361    0.35058792
## 137   132 4                 RINF RRC NENFERM FSD 0.38378269    0.35001735
## 118   133 4                  EDAD RRX NCAMAS PDP 0.38131282    0.34741215
## 111   134 4                     EDAD RRC RRX FSD 0.37336311    0.33902685
## 122   135 4                     EDAD RRX PDP FSD 0.36945758    0.33490732
```

```
## 110   136 4                  EDAD RRC RRX NENFERM 0.36878331    0.33419610
## 116   137 4                      EDAD RRC PDP FSD 0.36483036    0.33002655
## 119   138 4                EDAD RRX NCAMAS NENFERM 0.35749946    0.32229396
## 120   139 4                    EDAD RRX NCAMAS FSD 0.35482528    0.31947324
## 148   140 4                     RRC RRX NCAMAS PDP 0.35046366    0.31487263
## 151   141 4                     RRC RRX PDP NENFERM 0.35001902    0.31440363
## 113   142 4                EDAD RRC NCAMAS NENFERM 0.34873852    0.31305296
## 114   143 4                    EDAD RRC NCAMAS FSD 0.34312920    0.30713628
## 123   144 4                   EDAD RRX NENFERM FSD 0.33304323    0.29649766
## 152   145 4                       RRC RRX PDP FSD 0.32810841    0.29129243
## 158   146 4                RRX NCAMAS PDP NENFERM 0.32666103    0.28976574
## 161   147 4                   RRX PDP NENFERM FSD 0.32358206    0.28651807
## 117   148 4                   EDAD RRC NENFERM FSD 0.32186819    0.28471028
## 159   149 4                     RRX NCAMAS PDP FSD 0.32102405    0.28381989
## 149   150 4                RRC RRX NCAMAS NENFERM 0.31143678    0.27370729
## 150   151 4                    RRC RRX NCAMAS FSD 0.31008834    0.27228496
## 154   152 4                RRC NCAMAS PDP NENFERM 0.30771331    0.26977979
## 155   153 4                    RRC NCAMAS PDP FSD 0.29738607    0.25888668
## 160   154 4                RRX NCAMAS NENFERM FSD 0.29587500    0.25729281
## 157   155 4                   RRC PDP NENFERM FSD 0.29067164    0.25180434
## 153   156 4                   RRC RRX NENFERM FSD 0.28981879    0.25090475
## 156   157 4                RRC NCAMAS NENFERM FSD 0.24783513    0.20662061
## 125   158 4                   EDAD NCAMAS PDP FSD 0.23129796    0.18917730
## 124   159 4                EDAD NCAMAS PDP NENFERM 0.22866579    0.18640090
## 127   160 4                   EDAD PDP NENFERM FSD 0.22291152    0.18033133
## 126   161 4                EDAD NCAMAS NENFERM FSD 0.19572531    0.15165547
## 162   162 4                 NCAMAS PDP NENFERM FSD 0.19121004    0.14689278
## 176   163 5              EDAD RINF RRX PDP NENFERM 0.51227067    0.47840058
## 173   164 5               EDAD RINF RRX NCAMAS PDP 0.50153829    0.46692290
## 170   165 5              EDAD RINF RRC PDP NENFERM 0.49465573    0.45956238
## 164   166 5                  EDAD RINF RRC RRX PDP 0.49402898    0.45889210
## 174   167 5            EDAD RINF RRX NCAMAS NENFERM 0.48873270    0.45322802
## 177   168 5                  EDAD RINF RRX PDP FSD 0.48818474    0.45264201
## 167   169 5               EDAD RINF RRC NCAMAS PDP 0.48723455    0.45162583
## 163   170 5               EDAD RINF RRC RRX NCAMAS 0.48258701    0.44665555
## 175   171 5               EDAD RINF RRX NCAMAS FSD 0.47792793    0.44167293
## 179   172 5            EDAD RINF NCAMAS PDP NENFERM 0.47724690    0.44094460
## 182   173 5              EDAD RINF PDP NENFERM FSD 0.46859641    0.43169338
## 208   174 5             RINF RRX NCAMAS PDP NENFERM 0.46807016    0.43113059
## 166   175 5                  EDAD RINF RRC RRX FSD 0.46746447    0.43048284
## 178   176 5              EDAD RINF RRX NENFERM FSD 0.46480201    0.42763548
## 171   177 5                  EDAD RINF RRC PDP FSD 0.46408889    0.42687284
## 168   178 5            EDAD RINF RRC NCAMAS NENFERM 0.46287648    0.42557624
## 165   179 5              EDAD RINF RRC RRX NENFERM 0.46252040    0.42519542
## 180   180 5               EDAD RINF NCAMAS PDP FSD 0.46167602    0.42429242
## 211   181 5               RINF RRX PDP NENFERM FSD 0.46127265    0.42386103
## 201   182 5               RINF RRC RRX PDP NENFERM 0.45937499    0.42183158
## 198   183 5                RINF RRC RRX NCAMAS PDP 0.45172501    0.41365036
## 169   184 5               EDAD RINF RRC NCAMAS FSD 0.45167133    0.41359295
## 209   185 5                RINF RRX NCAMAS PDP FSD 0.45151206    0.41342262
## 186   186 5               EDAD RRC RRX PDP NENFERM 0.44991301    0.41171253
## 181   187 5            EDAD RINF NCAMAS NENFERM FSD 0.44546190    0.40695231
## 183   188 5                 EDAD RRC RRX NCAMAS PDP 0.44389681    0.40527853
## 204   189 5             RINF RRC NCAMAS PDP NENFERM 0.44344457    0.40479488
```

```
## 172   190 5              EDAD RINF RRC NENFERM FSD 0.44042427    0.40156485
## 212   191 5             RINF NCAMAS PDP NENFERM FSD 0.43768988    0.39864057
## 210   192 5             RINF RRX NCAMAS NENFERM FSD 0.43659173    0.39746616
## 202   193 5                    RINF RRC RRX PDP FSD 0.43420206    0.39491054
## 199   194 5             RINF RRC RRX NCAMAS NENFERM 0.43254707    0.39314061
## 207   195 5                RINF RRC PDP NENFERM FSD 0.43183725    0.39238151
## 205   196 5                 RINF RRC NCAMAS PDP FSD 0.42618654    0.38633838
## 200   197 5                 RINF RRC RRX NCAMAS FSD 0.42396145    0.38395878
## 187   198 5                    EDAD RRC RRX PDP FSD 0.42224902    0.38212742
## 189   199 5             EDAD RRC NCAMAS PDP NENFERM 0.41680423    0.37630452
## 203   200 5                RINF RRC RRX NENFERM FSD 0.41369527    0.37297967
## 184   201 5             EDAD RRC RRX NCAMAS NENFERM 0.41194472    0.37110755
## 206   202 5             RINF RRC NCAMAS NENFERM FSD 0.40360365    0.36218723
## 185   203 5                 EDAD RRC RRX NCAMAS FSD 0.40343401    0.36200582
## 190   204 5                 EDAD RRC NCAMAS PDP FSD 0.39810655    0.35630840
## 192   205 5                EDAD RRC PDP NENFERM FSD 0.39786495    0.35605002
## 193   206 5             EDAD RRX NCAMAS PDP NENFERM 0.39138679    0.34912199
## 196   207 5                EDAD RRX PDP NENFERM FSD 0.38740155    0.34485999
## 194   208 5                 EDAD RRX NCAMAS PDP FSD 0.38233371    0.33944022
## 188   209 5                EDAD RRC RRX NENFERM FSD 0.37889604    0.33576382
## 213   210 5              RRC RRX NCAMAS PDP NENFERM 0.36533582    0.32126192
## 195   211 5             EDAD RRX NCAMAS NENFERM FSD 0.36210505    0.31780679
## 216   212 5                 RRC RRX PDP NENFERM FSD 0.35575770    0.31101865
## 191   213 5             EDAD RRC NCAMAS NENFERM FSD 0.35432210    0.30948336
## 214   214 5                  RRC RRX NCAMAS PDP FSD 0.35273191    0.30778274
## 218   215 5               RRX NCAMAS PDP NENFERM FSD 0.33441740    0.28819639
## 215   216 5               RRC RRX NCAMAS NENFERM FSD 0.32020868    0.27300095
## 217   217 5               RRC NCAMAS PDP NENFERM FSD 0.31779080    0.27041517
## 197   218 5              EDAD NCAMAS PDP NENFERM FSD 0.23983581    0.18704663
## 222   219 6           EDAD RINF RRC RRX PDP NENFERM 0.52289792    0.48257944
## 229   220 6         EDAD RINF RRX NCAMAS PDP NENFERM 0.51914791    0.47851253
## 232   221 6            EDAD RINF RRX PDP NENFERM FSD 0.51301893    0.47186560
## 219   222 6             EDAD RINF RRC RRX NCAMAS PDP 0.51128905    0.46998953
## 225   223 6         EDAD RINF RRC NCAMAS PDP NENFERM 0.50860702    0.46708085
## 230   224 6             EDAD RINF RRX NCAMAS PDP FSD 0.50154652    0.45942369
## 228   225 6            EDAD RINF RRC PDP NENFERM FSD 0.49566713    0.45304745
## 220   226 6         EDAD RINF RRC RRX NCAMAS NENFERM 0.49479092    0.45209720
## 223   227 6                EDAD RINF RRC RRX PDP FSD 0.49455883    0.45184549
## 231   228 6         EDAD RINF RRX NCAMAS NENFERM FSD 0.49024979    0.44717230
## 226   229 6             EDAD RINF RRC NCAMAS PDP FSD 0.48723469    0.44390241
## 221   230 6             EDAD RINF RRC RRX NCAMAS FSD 0.48263413    0.43891307
## 233   231 6         EDAD RINF NCAMAS PDP NENFERM FSD 0.47937561    0.43537918
## 245   232 6          RINF RRX NCAMAS PDP NENFERM FSD 0.47116143    0.42647085
## 240   233 6          RINF RRC RRX NCAMAS PDP NENFERM 0.46960531    0.42478323
## 224   234 6            EDAD RINF RRC RRX NENFERM FSD 0.46822041    0.42328129
## 227   235 6         EDAD RINF RRC NCAMAS NENFERM FSD 0.46519396    0.41999908
## 234   236 6          EDAD RRC RRX NCAMAS PDP NENFERM 0.46309245    0.41771998
## 243   237 6             RINF RRC RRX PDP NENFERM FSD 0.46188501    0.41641050
## 241   238 6              RINF RRC RRX NCAMAS PDP FSD 0.45210659    0.40580573
## 237   239 6             EDAD RRC RRX PDP NENFERM FSD 0.45140699    0.40504702
## 244   240 6          RINF RRC NCAMAS PDP NENFERM FSD 0.44767511    0.40099977
## 235   241 6              EDAD RRC RRX NCAMAS PDP FSD 0.44396232    0.39697322
## 242   242 6          RINF RRC RRX NCAMAS NENFERM FSD 0.43659174    0.38897978
## 238   243 6          EDAD RRC NCAMAS PDP NENFERM FSD 0.42034716    0.37136241
```

```
## 236   244 6          EDAD RRC RRX NCAMAS NENFERM FSD 0.41505916    0.36562754
## 239   245 6          EDAD RRX NCAMAS PDP NENFERM FSD 0.39508215    0.34396233
## 246   246 6           RRC RRX NCAMAS PDP NENFERM FSD 0.37208387    0.31902053
## 247   247 7     EDAD RINF RRC RRX NCAMAS PDP NENFERM 0.53241491    0.48565640
## 250   248 7        EDAD RINF RRC RRX PDP NENFERM FSD 0.52346928    0.47581620
## 252   249 7     EDAD RINF RRX NCAMAS PDP NENFERM FSD 0.52020932    0.47223025
## 248   250 7         EDAD RINF RRC RRX NCAMAS PDP FSD 0.51135819    0.46249401
## 251   251 7     EDAD RINF RRC NCAMAS PDP NENFERM FSD 0.51004873    0.46105361
## 249   252 7     EDAD RINF RRC RRX NCAMAS NENFERM FSD 0.49620310    0.44582341
## 254   253 7      RINF RRC RRX NCAMAS PDP NENFERM FSD 0.47273054    0.42000359
## 253   254 7      EDAD RRC RRX NCAMAS PDP NENFERM FSD 0.46511598    0.41162758
## 255   255 8 EDAD RINF RRC RRX NCAMAS PDP NENFERM FSD 0.53329357    0.47918268
##     Mallow's Cp
## 2     20.996987
## 4     48.137530
## 6     52.210215
## 3     55.400897
## 5     55.875703
## 8     55.892648
## 7     59.549313
## 1     63.687306
## 9     15.583072
## 19    17.141414
## 17    18.584484
## 18    18.808978
## 21    19.414177
## 20    21.063002
## 16    22.821501
## 28    31.213900
## 27    34.052185
## 30    36.816962
## 24    37.890072
## 11    38.431328
## 29    38.731198
## 10    38.737895
## 23    41.919492
## 26    43.528932
## 13    45.726357
## 22    45.885464
## 25    46.677236
## 12    48.996486
## 15    50.348171
## 31    50.614438
## 14    52.658125
## 34    52.836970
## 35    53.026536
## 33    55.420680
## 36    57.231631
## 32    57.809364
## 38    12.142492
## 40    12.251651
## 64    13.653937
## 39    13.729533
## 42    14.857816
```

```
## 63    15.351557
## 70    15.655186
## 37    15.703316
## 41    15.907917
## 67    16.046091
## 66    16.742628
## 65    18.144050
## 60    18.511137
## 71    19.105761
## 68    19.598345
## 59    20.352247
## 69    20.468670
## 58    20.487745
## 62    21.104391
## 72    21.413953
## 61    22.784783
## 49    23.270374
## 45    23.908789
## 48    25.700541
## 44    27.454775
## 74    29.479964
## 51    29.669176
## 50    30.322508
## 47    30.873601
## 83    30.905158
## 86    31.010019
## 43    31.477436
## 46    32.288551
## 73    32.699109
## 87    32.922646
## 77    34.513199
## 85    35.191867
## 84    35.494355
## 76    35.626543
## 80    36.174158
## 75    37.684819
## 88    38.014272
## 81    39.561543
## 79    42.621475
## 78    43.228426
## 52    44.681702
## 82    45.100082
## 55    46.344441
## 56    47.064359
## 54    49.424530
## 53    50.886644
## 90    50.929556
## 57    51.478431
## 89    52.030103
## 92    52.616209
## 91    56.877813
## 99     7.712664
## 98     9.199017
## 105   10.805691
```

```
## 101   11.225928
## 95    11.272659
## 102   11.609139
## 100   11.959549
## 141   12.019171
## 94    13.114312
## 138   13.150340
## 93    13.739452
## 106   14.249514
## 103   14.410418
## 97    14.764309
## 104   15.629068
## 142   15.650574
## 129   15.653400
## 96    15.764477
## 144   15.784532
## 139   15.895010
## 135   16.512283
## 107   16.846313
## 132   16.953465
## 147   17.114765
## 140   17.174518
## 128   17.341489
## 109   17.426670
## 145   17.896076
## 131   18.708889
## 143   18.715233
## 130   20.092085
## 108   20.314035
## 136   20.490335
## 146   20.759970
## 133   21.006520
## 112   21.059780
## 115   21.434708
## 134   22.032166
## 121   23.024396
## 137   23.104369
## 118   23.469525
## 111   24.644846
## 122   25.222257
## 110   25.321945
## 116   25.906367
## 119   26.990200
## 120   27.385564
## 148   28.030406
## 151   28.096143
## 113   28.285459
## 114   29.114766
## 123   30.605921
## 152   31.335507
## 158   31.549495
## 161   32.004703
## 117   32.258090
## 159   32.382891
```

```
## 149   33.800317
## 150   33.999676
## 154   34.350812
## 155   35.877637
## 160   36.101041
## 157   36.870329
## 153   36.996420
## 156   43.203475
## 125   45.648405
## 124   46.037557
## 127   46.888293
## 126   50.907626
## 162   51.575185
## 176    6.108121
## 173    7.694845
## 170    8.712394
## 164    8.805057
## 174    9.588082
## 177    9.669096
## 167    9.809576
## 163   10.496689
## 175   11.185508
## 179   11.286195
## 182   12.565123
## 208   12.642925
## 166   12.732473
## 178   13.126104
## 171   13.231535
## 168   13.410783
## 165   13.463428
## 180   13.588264
## 211   13.647901
## 201   13.928460
## 198   15.059467
## 169   15.067403
## 209   15.090951
## 186   15.327361
## 181   15.985434
## 183   16.216824
## 204   16.283686
## 172   16.730219
## 212   17.134484
## 210   17.296839
## 202   17.650140
## 199   17.894822
## 207   17.999764
## 205   18.835191
## 200   19.164158
## 187   19.417331
## 189   20.222314
## 203   20.681956
## 184   20.940766
## 206   22.173948
## 185   22.199028
```

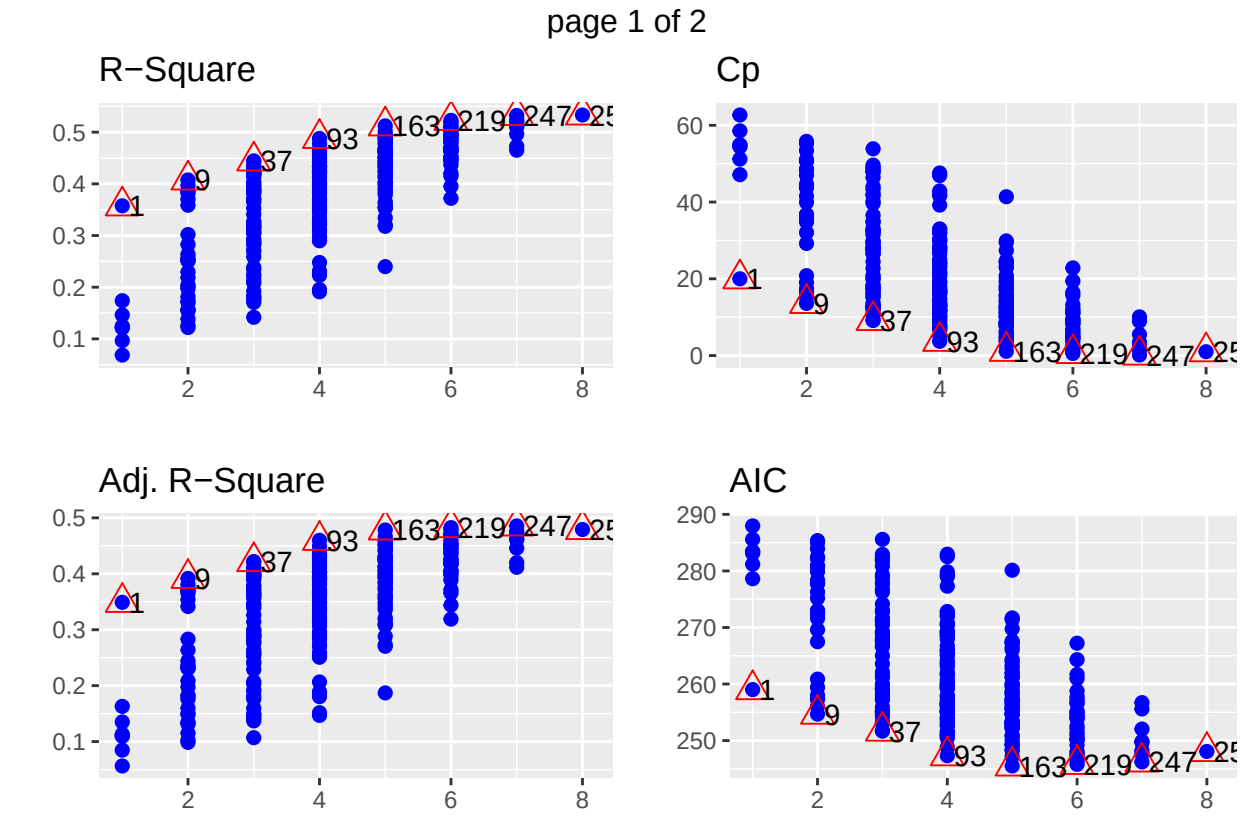
```
## 190   22.986664
## 192   23.022383
## 193   23.980143
## 196   24.569341
## 194   25.318593
## 188   25.826833
## 213   27.831638
## 195   28.309290
## 216   29.247710
## 191   29.459956
## 214   29.695057
## 218   32.402757
## 215   34.503439
## 217   34.860909
## 197   46.386130
## 222    6.536940
## 229    7.091358
## 232    7.997496
## 219    8.253248
## 225    8.649773
## 230    9.693628
## 228   10.562865
## 220   10.692407
## 223   10.726720
## 231   11.363789
## 226   11.809555
## 221   12.489722
## 233   12.971477
## 245   14.185898
## 240   14.415962
## 224   14.620712
## 227   15.068157
## 234   15.378853
## 243   15.557367
## 241   17.003053
## 237   17.106484
## 244   17.658222
## 235   18.207138
## 242   19.296838
## 238   21.698512
## 236   22.480313
## 239   25.433806
## 246   28.833976
## 247    7.129905
## 250    8.452469
## 252    8.934436
## 248   10.243027
## 251   10.436623
## 249   12.483624
## 254   15.953915
## 253   17.079685
## 255    9.000000
```

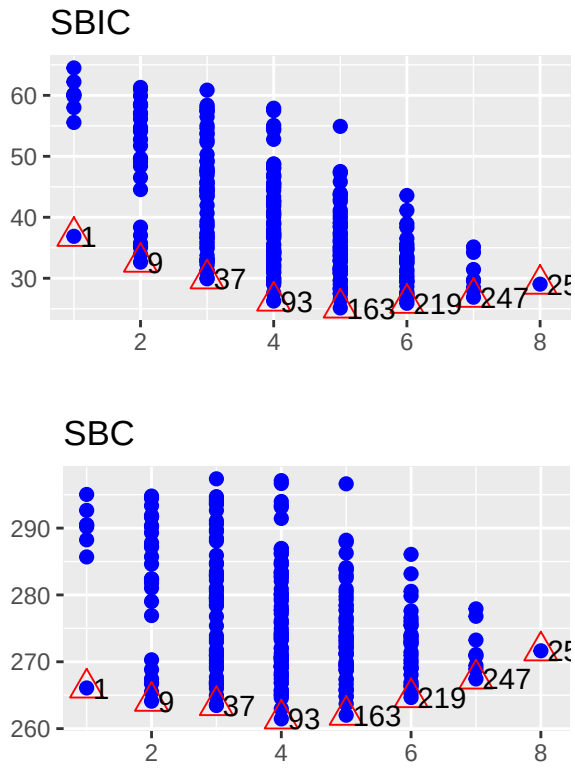
Con esto en mente, se seleccionaria acorde a los criterios mencionados; sin embargo, revisar uno a uno los

criterios de la tabla puede ser un trabajo poco efectivo y engorroso. Para ello, se apoya en la siguiente grafica:

```
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.

## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
```

```
## use `guide = "none"` instead.
```





Donde en el eje x se encuentra el numero de covariables tomadas y en el eje Y se representa el valor de cada prueba.

R^2 y R^2_{adj}:

Se seleccionan como candidatos los valores resaltados en las graficas anteriores siguiendo las indicaciones de los criterios, de esta manera se tiene que:

Bajo el criterio del R cuadrado, con 6 variables el modelo 219 es: 0.52289792

Bajo el criterio del R cuadrado ajustado, con 6 variables el modelo 219 es: 0.48257944

Bajo el criterio del R cuadrado, con 7 variables el modelo 247 es: 0.53241491

Bajo el criterio del R cuadrado ajustado, con 7 variables el modelo 247 es: 0.48565640

Bajo el criterio del R cuadrado, con 5 variables el modelo 163 es: 0.51227067

Bajo el criterio del R cuadrado ajustado, con 5 variables el modelo 163 es: 0.47840058

Al analizar todos los posibles valores del R cuadrado y ajustado, el mejor modelo a tener en cuenta es el modelo 163 que incluye a las variables EDAD, RINF, RRX, PDP y NENFERM debido a que los otros modelos tienen valores ligeramente mas altos pero no es una diferencia representativa.

NOTA: Se seleccionan los modelos 219,247 y 163 debido a que son los valores mas altos de esta medida, teniendo en cuenta el principio de parsimonia, intentando tener los valores mas altos de R con el menor numero de variables.

Estadistico Cp

Por este criterio, en la grafica buscamos los valores mas pequeños y nuveamente se resaltan los modelos 163,219, 247 y 255. Analizando dichos modelos en la tabla anterior se tiene que:

Para el modelo 163 es |6.108121-6|= 0.108121

Para el modelo 219 es |6.536940-7| = 0.46306

Para el modelo 247 es |7.129905-8| = 0.870095

Para el modelo 255 es |9.0-9| = 0

Teniendo en cuenta que se busca el minimo valor de Cp y minimo de la resta; Bajo este criterio, se selecciona nuevamente el modelo 163; correspondiente a las variables EDAD, RINF, RRX, PDP y NENFERM.

Como en ambos metodos el modelo ajjustado 163 es este:

```
##
## Call:
## lm(formula = DPERM ~ EDAD + RINF + RRX + PDP + NENFERM, data = datar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9360 -0.8765  0.0334  0.6191  3.2273
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.129183   1.595847   0.708  0.48149
## EDAD         0.078833   0.028049   2.811  0.00637 **
## RINF         0.496985   0.114354   4.346 4.47e-05 ***
## RRX          0.018157   0.007021   2.586  0.01173 *
## PDP          0.006667   0.002381   2.800  0.00656 **
## NENFERM     -0.004658   0.002455  -1.897  0.06182 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 72 degrees of freedom
## Multiple R-squared:  0.5123, Adjusted R-squared:  0.4784
## F-statistic: 15.12 on 5 and 72 DF,  p-value: 3.876e-10
```

De este modelo, se muestran los parametros estimados y demas valores.

```
## Anova Table (Type II tests)
##
## Response: DPERM
##           Sum Sq Df F value    Pr(>F)
## EDAD       9.748  1  7.8992  0.006366 **
## RINF      23.308  1 18.8879 4.473e-05 ***
## RRX        8.252  1  6.6875  0.011731 *
## PDP        9.674  1  7.8397  0.006557 **
## NENFERM    4.441  1  3.5991  0.061820 .
## Residuals 88.848 72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nota: Teniendo en cuenta que dentro de un modelo, se busca la eficiencia basado en el principio de parsimonia; el modelo resultante Incluye una variable que no es significativa a un nivel de 0.05. sin embargo, dicha variable no es significativa por muy poco (pues es de casi 0.06 y la significancia es de 0.05).

Se recomienda interrogar al experto para buscar claridad en esta cuestión específica pues, el podría reconocer o no la importancia de la variable NENFERM; dependiendo de esto se podría eliminar esta variable.

Seleccion Forward

Este metofo ajusta primero al modelo aquella X_j que sea estadisticamente significativa al modelo con menor error MSE (Con el p-valor) y asi, se agregan variables una a una hasta ir reduciendo significativamente el SSE en presencia de las que ya estan en el modelo.

Seleccion backward

Este metodo parte del modelo FULL a trabajar y de este, se seleccionan una a una variables X_j que no resulten significativa al modelo (la que menos), de tal manera que al eliminarla se reduzca el SSE, en presencia de las demas.

Seleccion StepWise

Combinacion de los dos metodos anteriores. Comienza agregando variables una a la vez segun el metodo forward (segun las mas significativas).

Una vez agregada una nueva variable, utilizar backward para verificar que todas las variables presentes son significativas.

Y se detiene el proceso cuando ya todas las variables son significativas.

Para este caso, se fija un $\alpha = 0.05$

Teniendo en cuenta que cada una de las selecciones para cada metodo se basa en las pruebas F de significancia de las variables, de esta manera; se tiene que:

Step Backward

```
##
##
##                          Elimination Summary
## ----------------------------------------------------------------------
##         Variable                  Adj.
## Step    Removed     R-Square    R-Square     C(p)       AIC        RMSE
## ----------------------------------------------------------------------
##    1    FSD           0.5324      0.4857    7.1299    246.2215    1.1031
##    2    NCAMAS        0.5229      0.4826    6.5369    245.7931    1.1064
##    3    RRC           0.5123      0.4784    6.1081    245.5115    1.1109
##    4    NENFERM       0.4879      0.4598    7.7127    247.3162    1.1305
## ----------------------------------------------------------------------
```

Despues de realizas la seleccion de variables a traves del metodo backward, el modelo de regresion resultante tiene a las variables EDAD, RINF, RRX y PDP; pues en esta tabla se muestran las variables removidas del modelo FULL (FSD,NCAMAS,RRC,NENFERM).

Se tienen las medidas resumenes calculadas anterior mente para un modelo con estas covariables.

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0641 -0.7924  0.0631  0.5754  3.3557
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.199063   1.623576   0.739 0.462559
## EDAD        0.078983   0.028544   2.767 0.007162 **
## RINF        0.476710   0.115863   4.114 0.000101 ***
## RRX         0.017938   0.007144   2.511 0.014256 *
## PDP         0.002585   0.001038   2.490 0.015056 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.13 on 73 degrees of freedom
## Multiple R-squared:  0.4879, Adjusted R-squared:  0.4598
## F-statistic: 17.39 on 4 and 73 DF,  p-value: 4.636e-10
```

El modelo ajustato tendrá los anteriores coeficientes y su ANOVA será:

```
## Anova Table (Type II tests)
##
## Response: DPERM
##           Sum Sq Df F value    Pr(>F)
## EDAD       9.785  1  7.6567 0.0071617 **
## RINF      21.634  1 16.9286 0.0001008 ***
## RRX        8.057  1  6.3047 0.0142558 *
## PDP        7.923  1  6.1995 0.0150561 *
## Residuals 93.289 73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

De lo anterior, como resultado se tiene un modelo con 4 variables predictoras, un R^2 de 0.4879 y el ajustado de 0.4598. Esto implica que el modelo explica aproximadamente entre un 49 y un 46% de la variabilidad total existente (dependiendo de que medida se tome en cuenta) ademas, Los residuales son de 1.13

Step Forward

```
##
##                             Selection Summary
## ------------------------------------------------------------------------
##         Variable                  Adj.
## Step    Entered     R-Square    R-Square     C(p)        AIC        RMSE
## ------------------------------------------------------------------------
##    1    RINF          0.3575      0.3490    20.9970    259.0144    1.2410
##    2    EDAD          0.4076      0.3918    15.5831    254.6763    1.1995
##    3    RRX           0.4444      0.4219    12.1425    251.6740    1.1695
##    4    PDP           0.4879      0.4598     7.7127    247.3162    1.1305
## ------------------------------------------------------------------------
```

Finalmente, el modelo resultante es el dado en la tabla resumen. A traves del metodo de seleccion forward y backward se llega al mismo modelo. Si se compara los coeficientes del modelo ajustado por el metodo Backward:

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Coefficients:
## (Intercept)         RINF         EDAD          RRX          PDP
##    1.199063     0.476710     0.078983     0.017938     0.002585
```

Se comprueba que el modelo es exactamente el mismo, con sus medidas de interes iguales.

Step Stepwise

```
##
##                               Stepwise Selection Summary
## ----------------------------------------------------------------------------------------
##                      Added/                   Adj.
## Step    Variable     Removed     R-Square    R-Square     C(p)        AIC         RMSE
## ----------------------------------------------------------------------------------------
##    1      RINF       addition       0.357       0.349    20.9970    259.0144    1.2410
##    2      EDAD       addition       0.408       0.392    15.5830    254.6763    1.1995
##    3      RRX        addition       0.444       0.422    12.1420    251.6740    1.1695
##    4      PDP        addition       0.488       0.460     7.7130    247.3162    1.1305
## ----------------------------------------------------------------------------------------
```

Finalmente con stepwise se llega al mismo modelo ajustado, concluyendo que todas las pruebas resultantes coinciden en que, el mejor modelo a ajustar es el ya calculado.

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Coefficients:
## (Intercept)         RINF         EDAD          RRX          PDP
##    1.199063     0.476710     0.078983     0.017938     0.002585
```

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

Entonces, reemplazando los coeficientes estimados, el modelo finalmente es:

$$Y = 1.199063 + 0.476710 X_{i1} + 0.078983 X_{i2} + 0.017938 X_{i3} + 0.002585 X_{i4+} \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

- X_1 es la variable RINF
- X_2 es la variable EDAD
- X_3 es la variable RRX
- X_4 es la variable PDP

10. **Seleccion del modelo**

Para seleccionar el modelo mas optimo, se debe de realizar una comparativa de los posibles modelos resultantes a trabajar, en este caso, el modelo 163 y el modelo resultante de los metodos de seleccion recien hechos (forward, backward, stepwise).

Pruebas sobre el nuevo modelo

Se deben de realizar pruebas sobre este modelo para verificar que cumpla todos los supuestos y sea un modelo optimo a trabajar.

El modelo a ajustar (acorde a lo ya realizado) es el siguiente:

```
modelo_final =lm(DPERM~EDAD+RINF+RRX+PDP,data = datar)
summary(modelo_final)
```

```
##
## Call:
## lm(formula = DPERM ~ EDAD + RINF + RRX + PDP, data = datar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0641 -0.7924  0.0631  0.5754  3.3557
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.199063   1.623576   0.739 0.462559
## EDAD        0.078983   0.028544   2.767 0.007162 **
## RINF        0.476710   0.115863   4.114 0.000101 ***
## RRX         0.017938   0.007144   2.511 0.014256 *
## PDP         0.002585   0.001038   2.490 0.015056 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.13 on 73 degrees of freedom
## Multiple R-squared:  0.4879, Adjusted R-squared:  0.4598
## F-statistic: 17.39 on 4 and 73 DF,  p-value: 4.636e-10
```

Se muestran, nuevamente los parametros estimados del modelo en la primera columna, rapidamente; con una significancia del 0.05; se puede ver los p-values en la ultima columna, que ayudan a probar significancia de los parametros(Lo cual es consistente, pues gracias a las pruebas de eliminacion de variables; el modelo ha sido construido de tal manera que las variables sean significativas tanto global como parcialmente)

De esta manera, se desea probar la significancia global del modelo (pues la individual de los parametros ya ha sido probada)

Significancia global:

El modelo planteado es de la forma:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i, \quad i = 1, 2, \ldots, 80$$

Que tiene como supuestos lo siguiente:

$$\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, 80$$

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_4 = 0, \quad \text{vs.}$$
$$H_1 : \text{Algún } \beta_j \neq 0, j = 1, \ldots, 4.$$

Para ello se usa la tabla de análisis de varianza. De ella se obtienen los valores del estadístico de prueba $F_0 = 14.7988$ y su correspondiente valor-P $\text{vp} = 1.68478e-12$.

Como $\text{vp} < 0.05 = \alpha$ se rechaza H_0 concluyendo que el modelo de RLM propuesto es significativo. Esto quiere decir, que la logitud de permanencia es afectada significativamente por al menos una de las predictoras consideradas.

```
      Sum_of_Squares DF Mean_Square F_Value     P_value
Model        88.8773  4    22.21934 17.3869 4.63553e-10
Error        93.2894 73     1.27794
```

Para ello se usa la tabla de análisis de varianza. De ella se obtienen los valores del estadístico de prueba $F_0 = 17.3869$ y su correspondiente valor-P $\text{vp} = 4.63553e-10$.

Como p-value $< 0.05 = \alpha$ se rechaza H_0 concluyendo que el modelo de RLM propuesto es significativo. Esto quiere decir, que la logitud de permanencia es afectada significativamente por al menos una de las predictoras consideradas.

Coeficiente de determinacion:

De la tabla resumen del modelo se obtiene rapidamente que los valores de $R^2 y R^2_{adj}$ son 0.4879 y 0.4598 respectivamente.

El modelo ajustado con las 4 variables selecionadas explica segun estas cifras, un 48.79 y 45.98 % de la varianza total. Nuevamente, se concluye que hace falta una "variable clave" para dar una respuesta optima al problema, pues aun, despues de todos los analisis y pruebas realizadas, el modelo final "abarca" un porcentaje bajo de la variabilidad total como para considerarse como un buen modelo.

Se recomienda tomar mas datos, muestreando nuevas variables a tener en cuenta.

Variables mas influyentes

Se desea saber cual variable/s tienen mas influencia en el modelo final. Para ello:

```
miscoeficientes(modelo_final,datar)
```

```
## Coeficientes estimados y Coeficientes estimados estandarizados
```

```
##              Estimacion  Coef.Std
## (Intercept) 1.199063119 0.0000000
## EDAD        0.078983075 0.2331713
## RINF        0.476710385 0.4076068
## RRX         0.017937925 0.2357010
## PDP         0.002585365 0.2216573
```

Al observar los coeficientes estandarizados se sabe que, la magnitud del valor absoluto de la variable con mayor valor será la más influyente. De esta forma:

- La variable RINF es la mas influyente para la variable respuesta DPERM
- La variable PDP es la menos influyente para la variable respuesta DPERM

Prueba de significancia individual de los parametros:

Se procede a probar la significancia individual de los parametros. Estas pruebas establecen el siguiente juego de hipótesis:

$$\begin{array}{l} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{array} \quad \text{para} \quad j = 1, 2, \ldots, 4.$$

De la tabla de parámetros estimados, a un nivel de significancia $\alpha = 0.05$ se rechaza H_0 si $|T_0| > T_{\frac{\alpha}{2}}, n-k-1$

Para este caso con la $T_{(1-\frac{0.05}{2},75)} = 1.992102$ basta comparar con los datos suministrados en la tabla anterior en la columna de t-values (note, nuevamente que este modelo fue creado de tal manera que cada parametro sea significativo, esta prueba lo ratifica):

- $EDAD = |2.767| > 1.992102$
- $RINF = |4.114| > 1.992102$
- $RRX = |2.5110| > 1.992102$
- $PDP = |2.490| > 1.992102$

concluye que los parámetros individuales $\beta_1, \beta_2, \beta_3, \beta_4$ son significativos cada uno en presencia de los demás parámetros. $\widehat{\beta}_1 = 0.078983$ indica que por cada unidad de aumento en la edad el promedio de la longitud de permanencia aumenta en 0.078983 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_2 = 0.476710$ indica que por cada unidad de aumento en el riesgo de infección el promedio de la longitud de permanencia aumenta en 0.476710 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_3 = 0.017938$ indica que por cada unidad de aumento en la razón de rutinas de rayos X en el pecho el promedio de la longitud de permanencia aumenta en 0.017938 unidades, cuando las demás variables predictoras se mantienen fijas.

$\widehat{\beta}_4 = 0.002585$ indica que por cada unidad de aumento en el censo promedio diario el promedio de la longitud de permanencia aumenta en 0.002585 unidades, cuando las demás variables predictoras se mantienen fijas.

Suma de cuadrados Tipo I

```
##           Sum Sq
## EDAD      12.515
## RINF      61.736
## RRX        6.704
## PDP        7.923
## Residuals 93.289
```

La suma de cuadrados extra tipo I agregada secuencialmente las variables según el orden establecido en el modelo full, es decir:

1) $SSR(EDAD) = 12.515$ Y este es el resultado de ajustar la recta de regresio (Suma de cuadrados de la regresion) $Y_i = \beta_0 + \beta_1 * X_{i1}$ que es Longidtud de permanencia (DPERM) vs Edad
2) $SSR(RINF|EDAD) = 61.736$ Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable RINF al modelo de la longitud de permanencia (DPERM) vs EDAD.
3) $SSR(RRX|EDAD, RINF) = 6.704$ Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable RINF al modelo de la longitud de permanencia (DPERM) vs EDAD.
4) $SSR(PDP|EDAD, RINF, RRX) = 7.923$ Este es el incremento de SSR (Suma de cuadrados de la regresion) al ingresar la variable RINF al modelo de la longitud de permanencia (DPERM) vs EDAD.

Teniendo en cuenta que se busca que el SSR sea mayor Y el SSE sea menor aqui se obvserva el efecto parcial de ir agregando cada nueva variable de manera secuencial, comenzando en X1 y terminando con X4.

En este caso al ingresar la ultima covariable se presento la menor reducción marginal en la suma de cuadrados extras cuando las demás predictorias fueron agragadas al modelo.

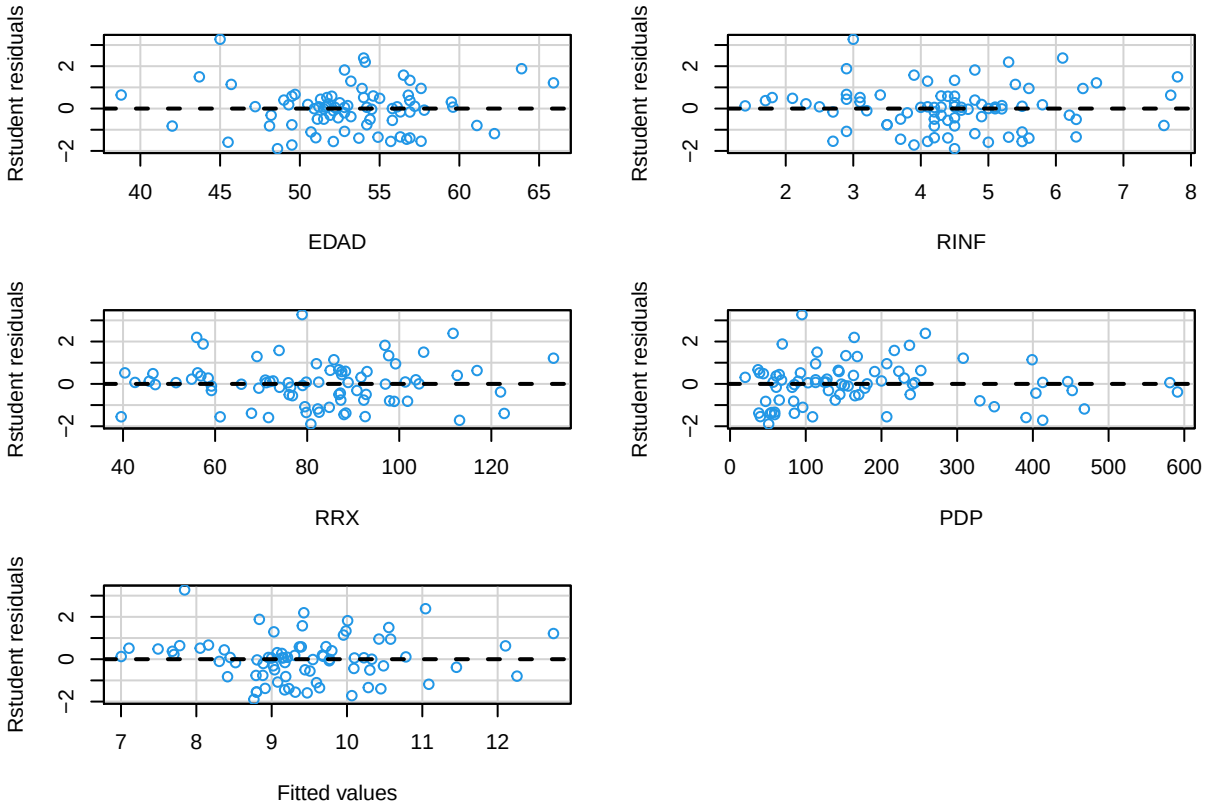
Suma de cuadrados tipo II

```
##           Sum Sq
## EDAD       9.785
## RINF      21.634
## RRX        8.057
## PDP        7.923
## Residuals 93.289
```

En este caso, se parte del modelo con todas las variables y se agrega una covariable, la idea es analizar el cambio que tiene en el modelo cada una de esta dado que en el modelo ya se encuentran las otras covariables (todas).

Se nota que en efecto, la variable con menor suma de cuadrado es PDP, indicando ser una que minimiza.

Graficos de residuales estudentizados vs valores ajustados



Interpretación

1) Se observa que los datos son aleatorios alrededor de 0, no se identifican patrones dentro de las graficas de estudentizados vs valores ajustados lo que indica que no hay problema de varianza constante.

2) Se evidencian valores alejados de la nube de puntos, en un principio se perciben puntos atipicos(estan por encima de 3). Por ejemplo, en la grafica de EDAD vs Residuales estudentizados, se ve claramente un valor en aproximadamente x= 45 que toma valores cercanos a 3. En general todas las graficas muestran un punto por encima de 3.

3) se observan algunos puntos alejados en direccion del eje x que da indicios de la existencia de puntos de balanceo

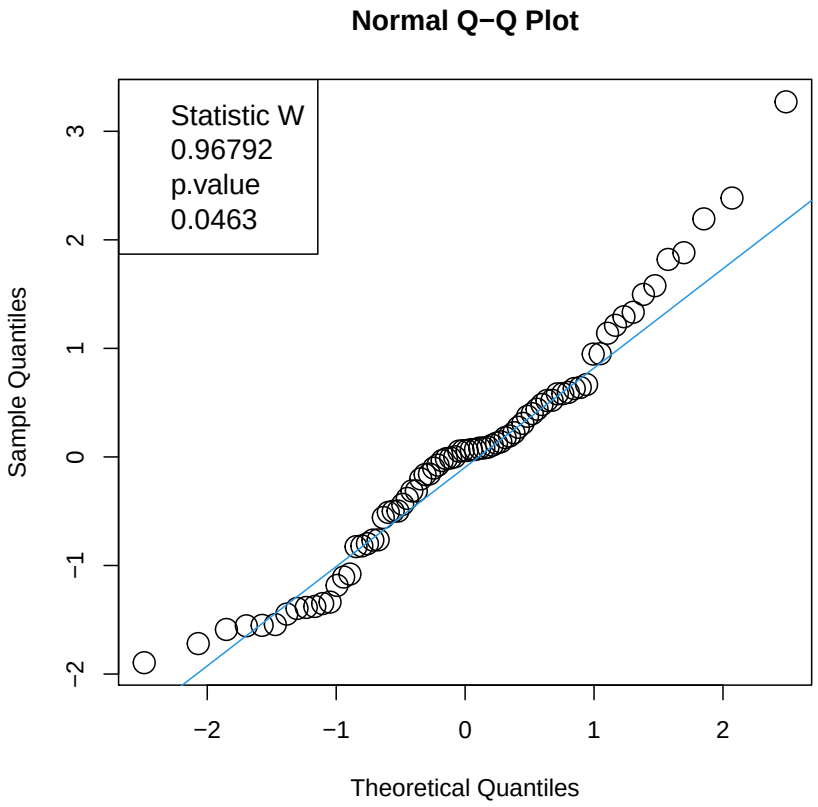
4) No se observan puntos influenciales a primera vista.

Graficamente se da una aproximación de diagnosticos acerca del modelo, sin embargo estos se deben respaldar con pruebas estadisticas.

Gráfico de normalidad Se desea probar que:

$$H_0 : Errores \varepsilon_i \sim N(0, \sigma^2)$$

$$H_1 : \varepsilon_i \nsim N(0, \sigma^2)$$



Los puntos sobre la recta no se ajustan lo suficiente para asegurar normaliad. De hecho, en los extremos, claramente los puntos se alejan de la recta lo que incita a sospechar sobre problemas de normalidad en los errores.

El valor p dado en la grafica es cero (0) por tanto, segun test shapiro wilk,se rechaza hipotesis nula y se concluye que en este caso, los errores no se distribuyen normal.

Este problema en la normalidad de los errores podria estar dado a observaciones atipicas.

Tabla de diagnosticos

```
##    datar.DPERM      ri      ei   se.yhat residuals Cooks.D hii.value  Dffits
## 1         8.84 -1.3372 -1.3301 0.3169667   -1.4433  0.0302    0.0786 -0.3906
## 2        11.07 -0.3818 -0.3841 0.5314837   -0.3832  0.0084    0.2210 -0.2034
## 3        12.78  0.6302  0.6328 0.3717972    0.6756  0.0097    0.1082  0.2195
## 4        11.62  0.9475  0.9482 0.2685445    1.0412  0.0108    0.0564  0.2317
## 5         9.31  0.0901  0.0907 0.2476265    0.1000  0.0001    0.0480  0.0202
## 6         8.19 -0.1025 -0.1032 0.2033579   -0.1148  0.0001    0.0324 -0.0188
## 7        11.77  2.1931  2.1380 0.2776463    2.3429  0.0587    0.0603  0.5556
## 8        13.59  2.3851  2.3120 0.2531305    2.5472  0.0564    0.0501  0.5480
## 9        10.33  0.0011  0.0011 0.2125204    0.0013  0.0000    0.0353  0.0002
## 10       11.48  0.9509  0.9515 0.2157310    1.0559  0.0068    0.0364  0.1849
## 11        8.63  0.5219  0.5245 0.2430335    0.5791  0.0027    0.0462  0.1149
## 12       11.15  1.5773  1.5615 0.1802147    1.7426  0.0127    0.0254  0.2547
## 13        8.03 -0.7640 -0.7662 0.1998463   -0.8525  0.0038    0.0313 -0.1372
## 14       10.05  0.5829  0.5855 0.1405837    0.6568  0.0011    0.0155  0.0731
## 15        8.90  0.6676  0.6701 0.2516009    0.7386  0.0047    0.0495  0.1524
## 16        8.54  0.0824  0.0830 0.2754721    0.0910  0.0001    0.0594  0.0207
```

```
## 17       12.01  1.8197  1.7915 0.1725666    2.0015  0.0153    0.0233  0.2811
## 18       13.95  1.2128  1.2090 0.5276609    1.2087  0.0814    0.2179  0.6401
## 21        9.76 -0.0094 -0.0095 0.1801300   -0.0106  0.0000    0.0254 -0.0015
## 22       11.18  1.1397  1.1373 0.3389956    1.2265  0.0256    0.0899  0.3582
## 23        7.58 -1.4476 -1.4369 0.2233678   -1.5923  0.0168    0.0390 -0.2918
## 24        9.06  0.0538  0.0541 0.1507623    0.0607  0.0000    0.0178  0.0072
## 25       10.24  0.3991  0.4015 0.2635703    0.4413  0.0019    0.0544  0.0957
## 26        8.28 -1.7191 -1.6966 0.4143470   -1.7844  0.0893    0.1343 -0.6773
## 27        9.89  0.1932  0.1945 0.2131712    0.2159  0.0003    0.0356  0.0371
## 28        9.84  0.1394  0.1403 0.1800717    0.1566  0.0001    0.0254  0.0225
## 29        9.74 -0.5109 -0.5136 0.2711117   -0.5636  0.0032    0.0575 -0.1262
## 30       10.47  1.2917  1.2858 0.1500451    1.4407  0.0059    0.0176  0.1730
## 31        8.37 -1.1063 -1.1046 0.2202614   -1.2247  0.0096    0.0380 -0.2198
## 32       10.90  0.1085  0.1093 0.3229788    0.1184  0.0002    0.0816  0.0324
## 33        9.23  0.0687  0.0692 0.3796848    0.0736  0.0001    0.1128  0.0245
## 34       12.07  1.4974  1.4848 0.4868475    1.5149  0.1004    0.1855  0.7145
## 35       10.02  0.5815  0.5841 0.1818716    0.6517  0.0018    0.0259  0.0948
## 36        7.39 -1.3774 -1.3690 0.1993003   -1.5233  0.0120    0.0311 -0.2467
## 37        8.45  0.6402  0.6428 0.4333730    0.6711  0.0142    0.1470  0.2657
## 38        8.88 -0.5581 -0.5607 0.1527929   -0.6281  0.0012    0.0183 -0.0761
## 39       10.30  0.0680  0.0685 0.2522967    0.0755  0.0000    0.0498  0.0156
## 40        7.94 -0.7659 -0.7680 0.2160053   -0.8522  0.0045    0.0365 -0.1491
## 41       10.39  0.5943  0.5969 0.1585250    0.6682  0.0014    0.0197  0.0842
## 42        8.02  0.4829  0.4855 0.3090394    0.5279  0.0038    0.0747  0.1372
## 43        8.34 -0.1619 -0.1630 0.2497529   -0.1797  0.0003    0.0488 -0.0367
## 44        9.68 -0.0754 -0.0760 0.1903589   -0.0846  0.0000    0.0284 -0.0129
## 45        8.67 -0.3133 -0.3153 0.2006935   -0.3508  0.0006    0.0315 -0.0565
## 46        9.00 -0.1572 -0.1583 0.1991329   -0.1762  0.0002    0.0310 -0.0281
## 47        9.84 -1.1834 -1.1801 0.4034047   -1.2463  0.0406    0.1273 -0.4520
## 48        8.48 -0.4986 -0.5012 0.1909369   -0.5584  0.0015    0.0285 -0.0854
## 49       11.20  3.2713  3.0734 0.2930137    3.3557  0.1361    0.0672  0.8779
## 50        7.67  0.5198  0.5224 0.3332039    0.5643  0.0052    0.0869  0.1603
## 51        8.88 -0.5008 -0.5034 0.1640752   -0.5630  0.0011    0.0211 -0.0735
## 52       11.41 -0.7989 -0.8009 0.3986108   -0.8472  0.0182    0.1243 -0.3010
## 53       11.46  1.3332  1.3261 0.2082026    1.4735  0.0123    0.0339  0.2498
## 54        7.78 -1.5894 -1.5730 0.3486391   -1.6916  0.0520    0.0951 -0.5153
## 55        9.61 -0.4367 -0.4391 0.2735784   -0.4816  0.0024    0.0586 -0.1089
## 56        9.53 -0.0154 -0.0155 0.2210328   -0.0172  0.0000    0.0382 -0.0031
## 57        8.09  0.3746  0.3768 0.3250813    0.4080  0.0026    0.0827  0.1125
## 58        9.05  0.0821  0.0827 0.1463072    0.0927  0.0000    0.0168  0.0107
## 59        7.91 -1.0785 -1.0773 0.3130789   -1.1702  0.0193    0.0767 -0.3108
## 60        8.86  0.4399  0.4423 0.2358566    0.4890  0.0018    0.0435  0.0938
## 61        8.66 -0.2020 -0.2033 0.1530533   -0.2277  0.0002    0.0183 -0.0276
## 62        7.95  0.2235  0.2250 0.2608737    0.2475  0.0006    0.0533  0.0530
## 63       10.15 -0.3131 -0.3151 0.3894138   -0.3344  0.0027    0.1187 -0.1149
## 64        9.44  0.2732  0.2749 0.2152916    0.3051  0.0006    0.0363  0.0530
## 65       10.80  1.8814  1.8495 0.3884494    1.9634  0.0916    0.1181  0.6884
## 66        7.14  0.1269  0.1277 0.3413601    0.1377  0.0003    0.0912  0.0402
## 67        9.50  0.1780  0.1792 0.3143372    0.1945  0.0005    0.0773  0.0515
## 68        9.41  0.3093  0.3112 0.3244982    0.3370  0.0017    0.0824  0.0927
## 69        7.13 -1.5516 -1.5368 0.3151399   -1.6684  0.0398    0.0777 -0.4504
## 70        8.95 -1.3958 -1.3869 0.3292196   -1.4998  0.0356    0.0848 -0.4249
## 71        8.28 -0.8196 -0.8214 0.2446266   -0.9066  0.0066    0.0468 -0.1817
## 72        7.53 -0.8261 -0.8279 0.3709474   -0.8841  0.0165    0.1077 -0.2870
```

```
## 73        9.20  0.0577  0.0581 0.1674474    0.0650  0.0000    0.0219  0.0086
## 74       10.16  0.0592  0.0596 0.4686829    0.0613  0.0001    0.1719  0.0270
## 75        6.70 -1.8952 -1.8624 0.2228386   -2.0641  0.0280    0.0389 -0.3811
## 76        7.63 -1.5555 -1.5406 0.2887694   -1.6838  0.0331    0.0653 -0.4110
## 77        8.77 -0.0339 -0.0341 0.3165248   -0.0370  0.0000    0.0784 -0.0099
## 78        8.15 -1.3506 -1.3431 0.2338240   -1.4854  0.0161    0.0428 -0.2855
## 79        7.14 -1.5438 -1.5294 0.3159019   -1.6601  0.0396    0.0781 -0.4493
## 80        7.70 -1.3863 -1.3776 0.2167303   -1.5285  0.0145    0.0368 -0.2708
##    Covratio
## 1    1.0286
## 2    1.3616
## 3    1.1688
## 4    1.0673
## 5    1.1248
## 6    1.1064
## 7    0.8252
## 8    0.7712
## 9    1.1107
## 10   1.0446
## 11   1.1023
## 12   0.9276
## 13   1.0622
## 14   1.0629
## 15   1.0930
## 16   1.1385
## 17   0.8761
## 18   1.2381
## 21   1.0993
## 22   1.0766
## 23   0.9659
## 24   1.0906
## 25   1.1205
## 26   1.0122
## 27   1.1080
## 28   1.0978
## 29   1.1164
## 30   0.9726
## 31   1.0237
## 32   1.1657
## 33   1.2072
## 34   1.1284
## 35   1.0744
## 36   0.9709
## 37   1.2208
## 38   1.0680
## 39   1.1272
## 40   1.0678
## 41   1.0665
## 42   1.1394
## 43   1.1243
## 44   1.1022
## 45   1.0988
## 46   1.1038
## 47   1.1150
```

```
## 48   1.0840
## 49   0.5745
## 50   1.1516
## 51   1.0756
## 52   1.1707
## 53   0.9817
## 54   0.9965
## 55   1.1231
## 56   1.1140
## 57   1.1567
## 58   1.0891
## 59   1.0711
## 60   1.1052
## 61   1.0883
## 62   1.1277
## 63   1.2074
## 64   1.1060
## 65   0.9557
## 66   1.1776
## 67   1.1586
## 68   1.1599
## 69   0.9855
## 70   1.0244
## 71   1.0730
## 72   1.1454
## 73   1.0952
## 74   1.2935
## 75   0.8739
## 76   0.9716
## 77   1.1624
## 78   0.9877
## 79   0.9875
## 80   0.9750
```

Observaciones atipicas

Se asume que la observación i es atipica si un ei grande ($|ei| > 3$) y Se considera potencialmente atipica con ri grande ($|ri| > 3$) Deacuerdo a la columna ei y ri se observa que la observación 19 es atípica.

Se tiene que la observacion 49 es atipica.

Observaciones de balanceo

Se asume que la observación i es un punto de balanceo si hii $> 2p/n = 0.1282051$

De acuerdo a la columna hii.value la observación 2 (0.2210),18 (valor muy alto, de 0.2179),26,34 (valor de 0.1855),37 (0.1470) y 74 (0.1719) son puntos de balanceo.

Se sospecha, particularmente de la observacion 2 y 18.

Observaciones influenciales

Para identificar estos valores utilizaremos 3 criterios, que son:

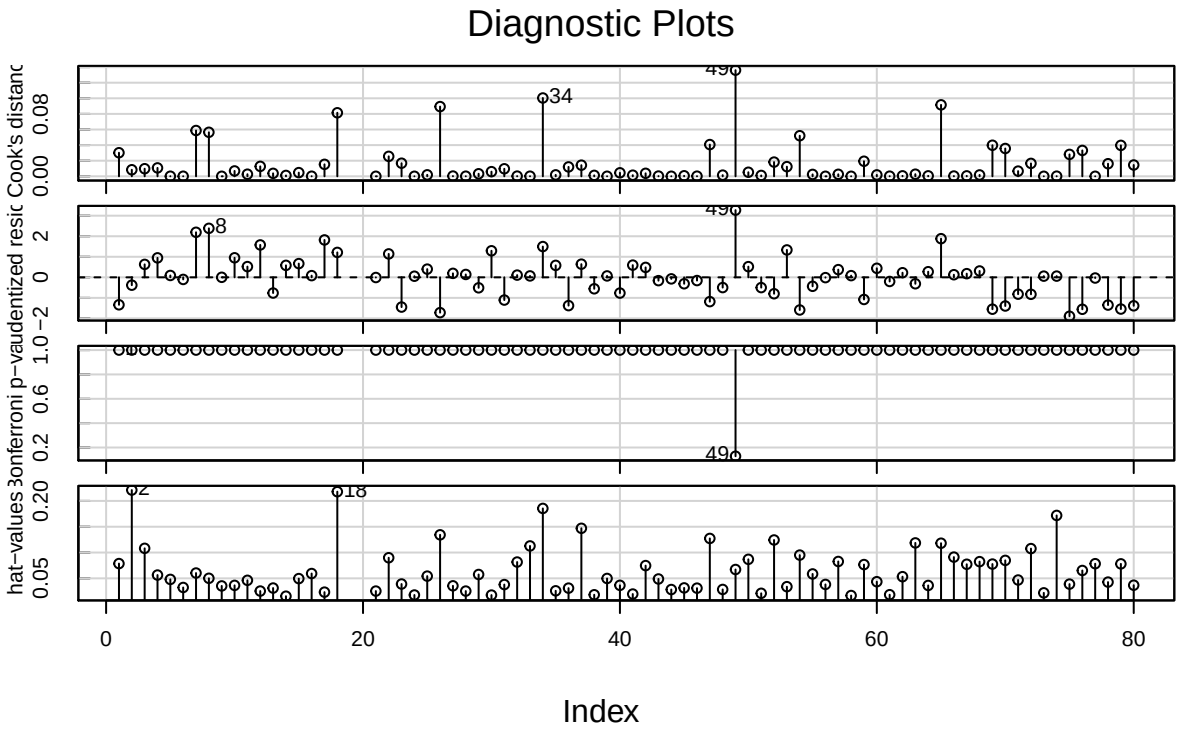
- Se dice que la observacion sera influencial si Di > 1. Segun esto, no hay datos influenciables.
- una observacion sera influencial si $|DFFITS| > 2(p/n)\hat{}0.5 = 0.5063697$

Segun esto, la observacion 7(0.5556),8(0.5480),18(0.6401),26(-0.6773),34(0.7145),49(0.8779),54(-0.5153) y 65(0.6884) son influenciables.

- observaciones con un covratio tal que |COVRATIO-1| > 3(p/n) =0.1923077

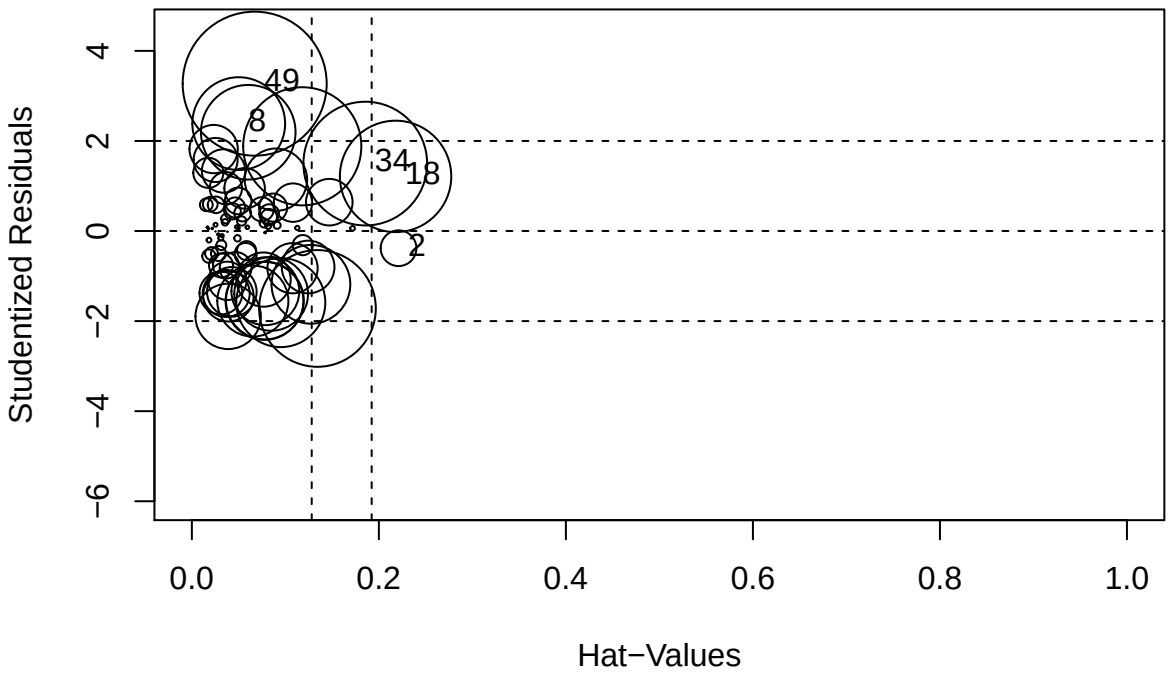
Segun esto, la observacion 2(0.3616),8(0.2288),18(0.2381),33(0.2072),37(0.2208),49(0.4255),63(0.2074) y 74(0.2935)

Gráficas de chequeos y diagnosticos



1) En la primera gráfica de distancias de cook, se observa que ningún valor es mayor a 1, por lo tanto no se puede concluir puntos de balanceo.
2) En la segunda y tercer grafica se muestra a el dato 49 como atipico, habiendo valores altos, como el 8.
3) En la gráfica de hat.values se ve que la observación 18 y 22 que están por encima de 0.4, por lo tanto son puntos de balanceo

```
influencePlot(modelo_final,xlim=c(0,1),ylim=c(-6.0,4.5))
```




```
##       StudRes        Hat       CookD
## 2  -0.3818396 0.22103978 0.008372574
## 8   2.3850640 0.05013943 0.056430713
## 18  1.2128479 0.21787151 0.081427650
## 34  1.4973619 0.18547116 0.100398250
## 49  3.2712938 0.06718407 0.136066014
```

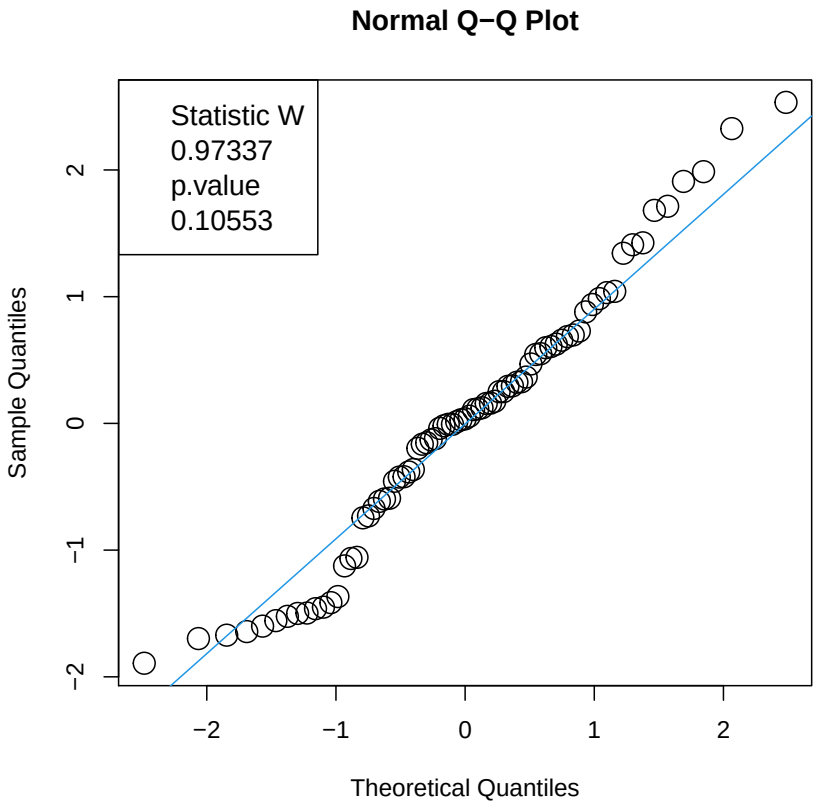
Nuevamente, se observan algunos datos ligeramente por encima de los Hat-values como el 34 y 18 pero no se alejan demasiado de los datos. Ademas a la observacion 49 como atipica.

al verificar que no son errores de digitacion, se procede a eliminar la observacion 49.

```
##
## Call:
## lm(formula = DPERM ~ EDAD + RINF + RRX + PDP, data = data_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.93498 -0.63067  0.03488  0.63228  2.52876
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.1086393  1.5613745   0.070 0.944721
## EDAD        0.0963358  0.0273370   3.524 0.000743 ***
## RINF        0.5129382  0.1094169   4.688 1.27e-05 ***
## RRX         0.0173364  0.0067144   2.582 0.011858 *
## PDP         0.0026553  0.0009758   2.721 0.008151 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.062 on 72 degrees of freedom
## Multiple R-squared:  0.5461, Adjusted R-squared:  0.5209
## F-statistic: 21.65 on 4 and 72 DF,  p-value: 9.254e-12
```

Estos son los nuevos parametros con los nuevos datos, sin la observacion 49. Verificando normalidad tenemos que:



Al eliminar la observacion 49, los residuales ahora se distribuyen normal estandar. Este dato atipico afectava negativamente la normalidad de los residuos.

En efecto, existe un cambio sustancial en los parametros del modelo ajustado si la observacion exlcuida, esto puede explicarse dado que la observación 49 era un punto atipico.

- El error estandar resitual pasó de 1.13 a 1.062
- El R^2 ajustado pasó de 0.4598 a 0.5209

La proporcion de variabilidad que el modelo ajustado explica en los datos aumentó a el 52%

Analisis multicolinealidad:

```
##     EDAD     RINF      RRX      PDP
## 1.008145 1.391833 1.256781 1.124648
```

Segun los VIFF's, no existe ningun tipo de multicolinealidad en las variables seleccionadas.

```
##   Val.propio cond.index      Pi.EDAD      Pi.RINF    Pi.RRX    Pi.PDP
## 1  1.5736486   1.000000 0.0008154207 2.162951e-01 0.1647611 0.1125050
## 2  1.0442726   1.227572 0.6676149895 3.170816e-06 0.1037641 0.1370549
```

```
## 3  0.9009342   1.321622 0.3190467351 2.446910e-03 0.1812922 0.4953226
## 4  0.4811445   1.808490 0.0125228546 7.812548e-01 0.5501826 0.2551176
```

Verificando el indice de condicion y la descomposicion de varianzas se comprueba que no hay problemas de multicolinealidad.

Finalmente, una vez probado todos los supuestos, se concluye que este es el modelo optimo que se puede ajustar a los datos obtenidos.