

# R Forward MTCARS

Yosef Guevara Salamanca

26/11/2020

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(MASS)
```

```
library(nortest)
```

```
library(crayon)
```

```
##
```

```
## Attaching package: 'crayon'
```

```
## The following object is masked from 'package:lmtest':
```

```
##
```

```
##      reset
```

```
data(mtcars)
```

```
attach(mtcars)
```

**PASO 0:** Para el primer paso ajustamos el modelo nulo con las 3 covariables

```
mt.fit3 <- lm(mpg ~ wt + disp + hp)
```

```
summary(mt.fit3)
```

```
##
```

```
## Call:
```

```
## lm(formula = mpg ~ wt + disp + hp)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -3.891 -1.640 -0.172  1.061  5.861
```

```
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.105505   2.110815  17.579 < 2e-16 ***
## wt          -3.800891   1.066191  -3.565  0.00133 **
## disp        -0.000937   0.010350  -0.091  0.92851
## hp          -0.031157   0.011436  -2.724  0.01097 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.639 on 28 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8083
## F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

Como analisis previo a los calculos de los modelos propuesto de regresion mediante el summary podemos ver que el modelo Full no es el mas apto para explicar a mpg pues el p.value de disp es 0.92

```
anova(mt.fit3)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## wt          1 847.73   847.73 121.7305 1.052e-11 ***
## disp         1  31.64    31.64   4.5433  0.04196 *
## hp           1  51.69    51.69   7.4228  0.01097 *
## Residuals  28 194.99     6.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
SCReg.full<-sum(anova(mt.fit3)[1:3,2])
SCReg.full
```

```
## [1] 931.0565
```

```
CME.full<-anova(mt.fit3)[4,3]
CME.full
```

```
## [1] 6.963953
```

**PASO 1: Eliminar una covariable de ser posible.** Lo primero que tenemos que hacer es eliminar una covariable, para ellos generaremos un f.criticoFull con lso siguientes parametros

Alpha input <- 0.1 G.L SSR = No. Covariables = 3 G.L SSE = (n - 1) - G.L SSR = 32 - 1 - 3 = 28

```
# G.L SST
# G.L SSR
# G.L SSE

#qf(1-alpha, Gl. SSR, Gl. SSE)

alpha <- 0.1

f.critico.1.back<-qf(1 - alpha ,3,28)
f.critico.1.back
```

```
## [1] 2.290595
```

Se eliminara la covariable con Fparcial menor que 4.24 para ello calcularemos los 3 valores de fparciales de cada modelo con 2 covariables

a) Ajustamos el modelo con parejas de 2 covariables X1: wt; X2: hp; X3: disp;

```
cat(red("mpg ~ wt + hp\n"))
```

```
## mpg ~ wt + hp
```

```
mt.fit12<-lm(mpg~ wt + hp)
anova(mt.fit12)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1  847.73   847.73  126.041 4.488e-12 ***
## hp         1   83.27    83.27   12.381 0.001451 **
## Residuals 29  195.05     6.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat(red("mpg~ wt + disp\n"))
```

```
## mpg~ wt + disp
```

```
mt.fit13<-lm(mpg~ wt + disp)
anova(mt.fit13)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1  847.73   847.73  99.6586 6.861e-11 ***
## disp        1   31.64    31.64   3.7195 0.06362 .
## Residuals 29  246.68     8.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat(red("mpg~ hp + disp\n"))
```

```
## mpg~ hp + disp
```

```
mt.fit23<-lm(mpg~ hp + disp)
anova(mt.fit23)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## hp          1  678.37   678.37   69.394 3.498e-09 ***
## disp        1  164.18   164.18   16.795 0.0003063 ***
## Residuals   29  283.49     9.78
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b) **Calculamos SCReg de los 3 modelos** Se extrae las sumas cuadradas del error de cada subconjunto de covariables

```
SCReg.12<-sum(anova(mt.fit12)[1:2,2])
SCReg.13<-sum(anova(mt.fit13)[1:2,2])
SCReg.23<-sum(anova(mt.fit23)[1:2,2])
```

c) **CALCULAMOS tres fparciales**  $f_{\text{parcial.sin.3}} <- (\text{SCReg.full} - \text{SCReg.12}) / (1) / (\text{SCE.full} / g.l.)$

```
fparcial.sin.3<-(SCReg.full-SCReg.12)/CME.full
fparcial.sin.2<-(SCReg.full-SCReg.13)/CME.full
fparcial.sin.1<-(SCReg.full-SCReg.23)/CME.full

decidir.paso1.B<-cbind(f.critico.1.back, fparcial.sin.3,fparcial.sin.2,fparcial.sin.1 )
decidir.paso1.B
```

```
##           f.critico.1.back fparcial.sin.3 fparcial.sin.2 fparcial.sin.1
## [1,]           2.290595      0.008196498      7.422771      12.7087
```

Podemos ver que apartir de los Fparciales es posible descartar la variable disp pues el F parical del modelo  $\text{lm}(\text{mpg} \sim \text{wt} + \text{hp})$ , se propone este modelo para su posterior evaluacion de modelos.

Por ende podemos decir que el modelo propuesto que mejor explica a mpg mediante la metodologia Backward es:

$$\text{mpg} = B_0 + B_1 * \text{wt} + B_2 * \text{hp}$$

## Validacion de supuestos

Para la validacion de los de supuestos, se utiliza el modelo propuesto modelo y sus residuales, para lo cual se ha creado la siguiente funcion.

```
ValidarSupuestos <- function (respuesta ,modelo,confianza){

  print("En conclusión: ")

  # Test de normalidad de los residuales
```

```

shapiro <- shapiro.test(modelo$residuals)
shapiro <- shapiro$p.value

lillie <- lillie.test(modelo$residuals)
lillie <- lillie$p.value

ifelse((shapiro > confianza) & (lillie > confianza), print("Normalidad de los residuales, no se rechaza H0"), print("No se rechaza H0"))

# Test de homocedasticidad (varianza constante de los residuales)

Breusch <- bptest(modelo)
Breusch <- Breusch$p.value

Goldfeld <- gqtest(modelo)
Goldfeld <- Goldfeld$p.value

ifelse((Breusch > confianza) & (Goldfeld > confianza), print("Existe homocedasticidad, no se rechaza H0"), print("No se rechaza H0"))

# Test de independencia de errores

independencia <- dwtest(mpg ~ modelo$residuals)
independencia <- independencia$p.value

ifelse((independencia > confianza), print("Hay independencia de errores, No se rechaza H0"), print("No se rechaza H0"))

# Construccion de la Tabla de respuestas

tabla <- rbind(shapiro, lillie, Breusch, Goldfeld, independencia)
rownames(tabla) <- c("Shapiro", "Lillie", "Breusch", "Goldfeld", "independencia")
colnames(tabla) <- c("p.value")

print(tabla)
}

```

```

modelo_propuesto <- lm(mpg ~ wt + hp)

```

```

ValidarSupuestos(mpg, modelo_propuesto, alpha)

```

```

## [1] "En conclusión: "
## [1] "No existe normalidad de los residuales, se rechaza H0"
## [1] "No existe homocedasticidad, se rechaza H0"
## [1] "No hay independencia de errores, se rechaza H0"
##           p.value
## Shapiro    0.0342747606
## Lillie      0.3674906935
## Breusch     0.6438038145
## Goldfeld    0.0004864276
## independencia 0.0022188827

```