M5\_AI4\_PalaciosMadalengoitiaJorge

Jorge Enrique Palacios Madalengoitia

2023-10-23

# **Descripción de la tarea**

Dentro del paquete de R "MPV", se encuentra una base de datos de gasto en combustible de diferentes coches con una serie de características:

* y Miles/gallon.
* x1 Displacement (cubic in).
* x2 Horsepower (ft-lb).
* x3 Torque (ft-lb).
* x4 Compression ratio.
* x5 Rear axle ratio.
* x6 Carburetor (barrels).
* x7 No. of transmission speeds.
* x8 Overall length (in).
* x9 Width (in).
* x10 Weight (lb).
* x11 Type of transmission (1=automatic, 0=manual.
* library(MPV)  
  df<-table.b3[-c(23, 25),]  
  head(df)
* ## y x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11  
  ## 1 18.90 350 165 260 8.00 2.56 4 3 200.3 69.9 3910 1  
  ## 2 17.00 350 170 275 8.50 2.56 4 3 199.6 72.9 3860 1  
  ## 3 20.00 250 105 185 8.25 2.73 1 3 196.7 72.2 3510 1  
  ## 4 18.25 351 143 255 8.00 3.00 2 3 199.9 74.0 3890 1  
  ## 5 20.07 225 95 170 8.40 2.76 1 3 194.1 71.8 3365 0  
  ## 6 11.20 440 215 330 8.20 2.88 4 3 184.5 69.0 4215 1
* str(df)
* ## 'data.frame': 30 obs. of 12 variables:  
  ## $ y : num 18.9 17 20 18.2 20.1 ...  
  ## $ x1 : num 350 350 250 351 225 440 231 262 89.7 96.9 ...  
  ## $ x2 : num 165 170 105 143 95 215 110 110 70 75 ...  
  ## $ x3 : num 260 275 185 255 170 330 175 200 81 83 ...  
  ## $ x4 : num 8 8.5 8.25 8 8.4 8.2 8 8.5 8.2 9 ...  
  ## $ x5 : num 2.56 2.56 2.73 3 2.76 2.88 2.56 2.56 3.9 4.3 ...  
  ## $ x6 : num 4 4 1 2 1 4 2 2 2 2 ...  
  ## $ x7 : num 3 3 3 3 3 3 3 3 4 5 ...  
  ## $ x8 : num 200 200 197 200 194 ...  
  ## $ x9 : num 69.9 72.9 72.2 74 71.8 69 65.4 65.4 64 65 ...  
  ## $ x10: num 3910 3860 3510 3890 3365 ...  
  ## $ x11: num 1 1 1 1 0 1 1 1 0 0 ...

1. Proponed una especificación que a vuestra intuición sea un buen modelo para explicar la variable y en base a las x que tenemos anteriormente.

* Basado en mi conocimiento e intuición, diría que las siguiente variables explican el rendimiento de millas por galón:
  1. **Weight**: El peso del automóvil tiende a tener una influencia significativa en el consumo de combustible, siendo los más ligeros los más eficientes.
  2. **Type of transmission**: La elección entre una transmisión automática y una manual puede afectar la eficiencia de combustible. Por lo general, las transmisiones manuales permiten un mayor control sobre la velocidad y, a menudo, son más eficientes en términos de consumo de combustible.
  3. **Horsepower**: Los motores más potentes a menudo requieren más combustible para funcionar y pueden tener un impacto negativo en el consumo de combustible.
  4. **No. of transmission speeds**: Un mayor número de velocidades en la transmisión puede permitir un funcionamiento más eficiente del motor, lo que puede mejorar el consumo de combustible.
  5. **Carburetor**: El tipo y tamaño del carburador utilizado en el motor también puede tener un impacto en la mezcla de aire y combustible, lo que influye en el consumo de combustible.
* Debido a que tratamos de explicar como influyen estas variables con el consumo de millas por galón, optamos por usar un modelo de regresión lineal.
* # Factorizamos la variable x11  
  df$x11 <- relevel(as.factor(df$x11),ref=2)  
    
  # Modelo de regresión lineal  
  modelo <- lm("y ~ x2 + x6 + x7 + x10 + x11", data = df)  
    
  # Resumen del modelo  
  pander::pander(summary(modelo))

|  | * Estimate | * Std. Error | * t value | * Pr(>|t|) |
| --- | --- | --- | --- | --- |
| * **(Intercept)** | * 30.23 | * 7.661 | * 3.946 | * 0.0006037 |
| * **x2** | * -0.03196 | * 0.04838 | * -0.6606 | * 0.5152 |
| * **x6** | * 0.1806 | * 1.092 | * 0.1654 | * 0.87 |
| * **x7** | * 1.833 | * 1.893 | * 0.968 | * 0.3427 |
| * **x10** | * -0.003436 | * 0.001783 | * -1.927 | * 0.0659 |
| * **x110** | * 0.2681 | * 2.972 | * 0.09021 | * 0.9289 |

* Fitting linear model: “y ~ x2 + x6 + x7 + x10 + x11”

| * Observations | * Residual Std. Error |  | * Adjusted |
| --- | --- | --- | --- |
| * 30 | * 3.392 | * 0.7575 | * 0.707 |

* En este modelo, las variables **x2**, **x6**, **x7** y **x11** no son estadísticamente significativas en la predicción de **y** (Miles/gallon). La variable **x10** (Weight) muestra una posible influencia marginalmente significativa, y el intercepto es estadísticamente significativo.

1. Utilizar la técnica STEPWISE para elegir el modelo de tal forma que minimicemos el BIC.

* # Modelo inicial con todas las variables  
  formula\_full <- as.formula("y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11")  
  modelo\_full <- lm("y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11", data = df)  
    
  # Realizamos la selección stepwise  
  stepwise\_backward <- stepAIC(modelo\_full, trace = FALSE, direction = "backward")  
    
  # Mostramos el modelo final  
  stepwise\_backward$anova
* ## Stepwise Model Path   
  ## Analysis of Deviance Table  
  ##   
  ## Initial Model:  
  ## y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11  
  ##   
  ## Final Model:  
  ## y ~ x5 + x8 + x10  
  ##   
  ##   
  ## Step Df Deviance Resid. Df Resid. Dev AIC  
  ## 1 18 187.4007 78.96155  
  ## 2 - x11 1 0.4648362 19 187.8655 77.03587  
  ## 3 - x6 1 0.5356445 20 188.4012 75.12128  
  ## 4 - x4 1 3.4514854 21 191.8526 73.66591  
  ## 5 - x2 1 10.7796848 22 202.6323 73.30587  
  ## 6 - x7 1 10.4571693 23 213.0895 72.81545  
  ## 7 - x3 1 4.8720101 24 217.9615 71.49363  
  ## 8 - x1 1 0.7654631 25 218.7270 69.59881  
  ## 9 - x9 1 5.0970905 26 223.8241 68.28989

1. Programad vuestro propio STEPWISE (Backward o Forward) para decidir cuál sería el mejor modelo minimizando la siguiente función:  
   