Untitled

Esteban Sierra

2025-09-30

D = read.csv('AlCorte.csv')

# Análisis Descriptivo

Obtener el mínimo, la mediana la media y otros valores

n = 5 #número de variables  
d = matrix(NA,ncol=8,nrow=n)  
for(i in 1:n){  
 d[i,]<-c(as.numeric(summary(D[,i])), sd(D[ ,i]), sd(D[ ,i])/mean(D[ ,i]))  
}  
m = as.data.frame(d)  
variables = names(D)  
row.names(m) = variables  
names(m) = c("Minimo","Q1","Mediana","Media","Q3","Máximo","Desv Est", "CV")  
round(m,2)

## Minimo Q1 Mediana Media Q3 Máximo Desv Est CV  
## Fuerza 25.0 30.00 35.0 35.00 40.0 45.0 4.55 0.13  
## Potencia 45.0 60.00 75.0 75.00 90.0 105.0 13.65 0.18  
## Temperatura 150.0 175.00 200.0 200.00 225.0 250.0 22.74 0.11  
## Tiempo 10.0 15.00 20.0 20.00 25.0 30.0 4.55 0.23  
## Resistencia 22.7 34.67 38.6 38.41 42.7 58.7 8.95 0.23

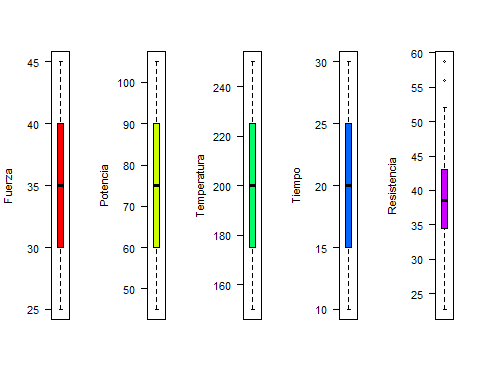
Obtener la correlación de las variables

cor(D)

## Fuerza Potencia Temperatura Tiempo Resistencia  
## Fuerza 1.0000000 0.0000000 0.0000000 0.0000000 0.1075208  
## Potencia 0.0000000 1.0000000 0.0000000 0.0000000 0.7594185  
## Temperatura 0.0000000 0.0000000 1.0000000 0.0000000 0.3293353  
## Tiempo 0.0000000 0.0000000 0.0000000 1.0000000 0.1312262  
## Resistencia 0.1075208 0.7594185 0.3293353 0.1312262 1.0000000

Obtener el gráfico de bigote

colores = rainbow(5)  
par(mfrow=c(1,5), las=1)  
boxplot(D[1], col=colores[1], ylab=variables[1])  
boxplot(D[2], col=colores[2], ylab=variables[2])  
boxplot(D[3], col=colores[3], ylab=variables[3])  
boxplot(D[4], col=colores[4], ylab=variables[4])  
boxplot(D[5], col=colores[5], ylab=variables[5])



# Obtener mejor modelo de regresión

## Criterio AIC

R = lm(Resistencia ~ . , data = D)  
step(R, direction="both", trace=1)

## Start: AIC=102.96  
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo  
##   
## Df Sum of Sq RSS AIC  
## - Fuerza 1 26.88 692.00 102.15  
## - Tiempo 1 40.04 705.16 102.72  
## <none> 665.12 102.96  
## - Temperatura 1 252.20 917.32 110.61  
## - Potencia 1 1341.01 2006.13 134.08  
##   
## Step: AIC=102.15  
## Resistencia ~ Potencia + Temperatura + Tiempo  
##   
## Df Sum of Sq RSS AIC  
## - Tiempo 1 40.04 732.04 101.84  
## <none> 692.00 102.15  
## + Fuerza 1 26.88 665.12 102.96  
## - Temperatura 1 252.20 944.20 109.47  
## - Potencia 1 1341.02 2033.02 132.48  
##   
## Step: AIC=101.84  
## Resistencia ~ Potencia + Temperatura  
##   
## Df Sum of Sq RSS AIC  
## <none> 732.04 101.84  
## + Tiempo 1 40.04 692.00 102.15  
## + Fuerza 1 26.88 705.16 102.72  
## - Temperatura 1 252.20 984.24 108.72  
## - Potencia 1 1341.01 2073.06 131.07

##   
## Call:  
## lm(formula = Resistencia ~ Potencia + Temperatura, data = D)  
##   
## Coefficients:  
## (Intercept) Potencia Temperatura   
## -24.9017 0.4983 0.1297

## Criterio BIC

n = length(D)  
R = lm(Resistencia ~ . , data = D)  
step(R, direction="both", k=log(n))

## Start: AIC=101.01  
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo  
##   
## Df Sum of Sq RSS AIC  
## - Fuerza 1 26.88 692.00 100.59  
## <none> 665.12 101.01  
## - Tiempo 1 40.04 705.16 101.16  
## - Temperatura 1 252.20 917.32 109.05  
## - Potencia 1 1341.01 2006.13 132.52  
##   
## Step: AIC=100.59  
## Resistencia ~ Potencia + Temperatura + Tiempo  
##   
## Df Sum of Sq RSS AIC  
## <none> 692.00 100.59  
## - Tiempo 1 40.04 732.04 100.67  
## + Fuerza 1 26.88 665.12 101.01  
## - Temperatura 1 252.20 944.20 108.30  
## - Potencia 1 1341.02 2033.02 131.31

##   
## Call:  
## lm(formula = Resistencia ~ Potencia + Temperatura + Tiempo, data = D)  
##   
## Coefficients:  
## (Intercept) Potencia Temperatura Tiempo   
## -30.0683 0.4983 0.1297 0.2583

extractAIC(R, k=log(n))

## [1] 5.0000 101.0102

## Criterio HQC

HQC = step(R, direction="both", k=2\*log(log(n)))

## Start: AIC=97.72  
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo  
##   
## Df Sum of Sq RSS AIC  
## <none> 665.12 97.722  
## - Fuerza 1 26.88 692.00 97.959  
## - Tiempo 1 40.04 705.16 98.524  
## - Temperatura 1 252.20 917.32 106.415  
## - Potencia 1 1341.01 2006.13 129.890

## Significancia

BestModel = lm(Resistencia ~ Fuerza + Potencia + Temperatura, data = D)  
summary(BestModel)

##   
## Call:  
## lm(formula = Resistencia ~ Fuerza + Potencia + Temperatura, data = D)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.3817 -2.6421 -0.5942 3.1892 8.4017   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -32.31000 12.52410 -2.580 0.01589 \*   
## Fuerza 0.21167 0.21261 0.996 0.32864   
## Potencia 0.49833 0.07087 7.032 1.82e-07 \*\*\*  
## Temperatura 0.12967 0.04252 3.049 0.00522 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.208 on 26 degrees of freedom  
## Multiple R-squared: 0.6967, Adjusted R-squared: 0.6617   
## F-statistic: 19.91 on 3 and 26 DF, p-value: 6.507e-07

# Economía de las variables  
#Significación global (Prueba para el modelo)  
#Significación individual (Prueba para cada 𝛽𝑖)  
#Variación explicada por el modelo

### Economía de las variables

### Significancia global

La significancia global en este modelo es alta ya que el p-value es menor a 0.05.

### Significancia individual

El modelo tiene una alta significancia porque - Potencia: t = 7.033 y valor p casi cero – Temperatura: t = 3.050 y valor p = 0.00499

### Variación explicada por el modelo

confint(BestModel)

## 2.5 % 97.5 %  
## (Intercept) -58.05364728 -6.5663527  
## Fuerza -0.22535738 0.6486907  
## Potencia 0.35265865 0.6440080  
## Temperatura 0.04226186 0.2170715

# Análisis de validez del modelo encontrado

## Análisis de residuos

### Homocedasticidad

### Independencia

# A1 Regresión múltiple

1. Haz un análisis descriptivo de los datos: medidas principales y gráficos
2. Encuentra el mejor modelo de regresión que explique la variable Resistencia. Analiza el modelo basándote en:
3. Significancia del modelo: 1. Economía de las variables 2. Significación global (Prueba para el modelo) 3. Significación individual (Prueba para cada 𝛽𝑖) 4. Variación explicada por el modelo
4. Analiza la validez del modelo encontrado:
5. Análisis de residuos (homocedasticidad, independencia, etc)
6. No multicolinealidad de Xi
7. Emite conclusiones sobre el modelo final encontrado e interpreta en el contexto del problema el efecto de las variables predictoras en la variable respuesta

# A3-Regresión Múltiple-Detección datos atípicos

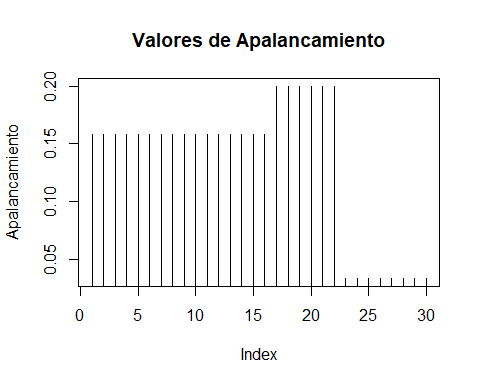
1. Haz un análisis descriptivo de los datos: medidas principales y gráficos (ya lo hiciste en la actividad A2)
2. Encuentra el mejor modelo de regresión que explique la variable Resistencia (ya lo hiciste en la actividad A2)
3. Analiza la validez del modelo encontrado (ya lo hiciste en la actividad A2)
4. Haz el análisis de datos atípicos e incluyentes del mejor modelo encontrado.

## Datos atípicos o con alto leverage.

Comenta todos los datos atípicos o con alto leverage que encuentres. Comenta por qué son influyentes o no lo son según el caso.

Matriz sombrero:

leverage = hatvalues(BestModel)  
plot(leverage, type="h", main="Valores de Apalancamiento", ylab="Apalancamiento")  
abline(h = 2\*mean(leverage), col="red") # Límite comúnmente usado

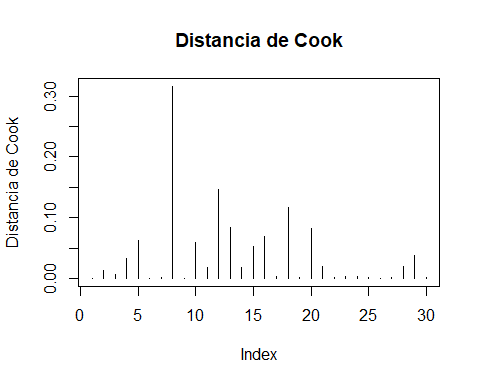


high\_leverage\_points = which(leverage > 2\*mean(leverage))  
D[high\_leverage\_points, ]

## [1] Fuerza Potencia Temperatura Tiempo Resistencia  
## <0 rows> (o 0- extensión row.names)

Detección de datos influyentes

# Distancia de cook  
#cooks.distance(BestModel)  
#I = influence.measures(BestModel)  
#summary(I)  
  
cooksdistance <- cooks.distance(BestModel)  
plot(cooksdistance, type="h", main="Distancia de Cook", ylab="Distancia de Cook")  
abline(h = 1, col="red")

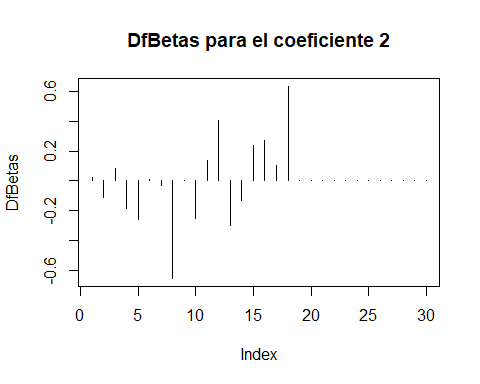


puntos\_influyentes = which(cooksdistance > 1)  
  
D[puntos\_influyentes, ]

## [1] Fuerza Potencia Temperatura Tiempo Resistencia  
## <0 rows> (o 0- extensión row.names)

Se detectan DfBetas mayores a |1|

dfbetas\_values = dfbetas(BestModel)  
#Calcula la DfBeta de los n datos para cada 𝛽𝑗  
#Gráfico auxiliar, para la variable 2:  
plot(dfbetas\_values[, 2], type="h", main="DfBetas para el coeficiente 2", ylab="DfBetas")  
abline(h = c(-1, 1), col="red") # Límites comunes



#Cuenta e identifica cuántos datos atípicos hay:  
puntos\_influyentes = which(abs(dfbetas\_values[, 2]) > 1)