

A13_Regresión no lineal

Nallely Serna

2024-09-10

```
data(cars)
```

##Análisis de Normalidad

#Pruebas de normalidad Realizamos las pruebas de Shapiro-Wilk y Anderson-Darling para evaluar la normalidad de las variables speed y dist.

```
shapiro.test(cars$speed)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: cars$speed  
## W = 0.97765, p-value = 0.4576
```

```
shapiro.test(cars$dist)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: cars$dist  
## W = 0.95144, p-value = 0.0391
```

```
library(nortest) # Para Anderson-Darling  
ad.test(cars$speed)
```

```
##  
## Anderson-Darling normality test  
##  
## data: cars$speed  
## A = 0.26143, p-value = 0.6927
```

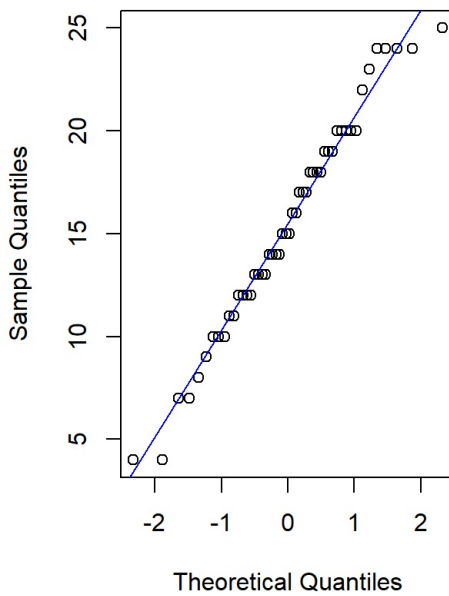
```
ad.test(cars$dist)
```

```
##  
## Anderson-Darling normality test  
##  
## data: cars$dist  
## A = 0.74067, p-value = 0.05021
```

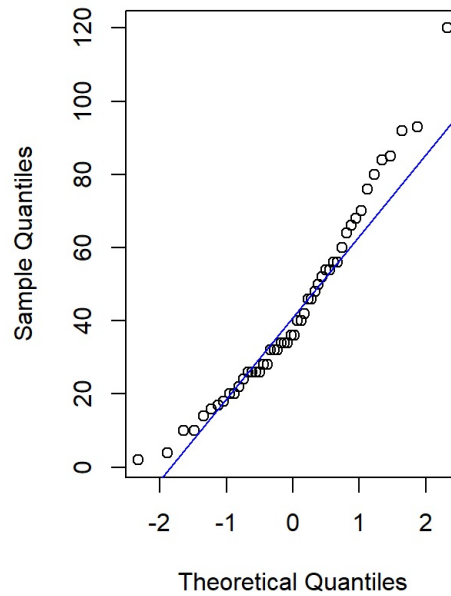
#Gráficos QQPlot y histogramas Se grafican los QQPlots y los histogramas de las variables speed y dist, con sus respectivas distribuciones teóricas superpuestas.

```
# Gráfico QQPlots  
par(mfrow=c(1,2))  
qqnorm(cars$speed, main="QQPlot de Velocidad")  
qqline(cars$speed, col="blue")  
qqnorm(cars$dist, main="QQPlot de Distancia")  
qqline(cars$dist, col="blue")
```

QQPlot de Velocidad



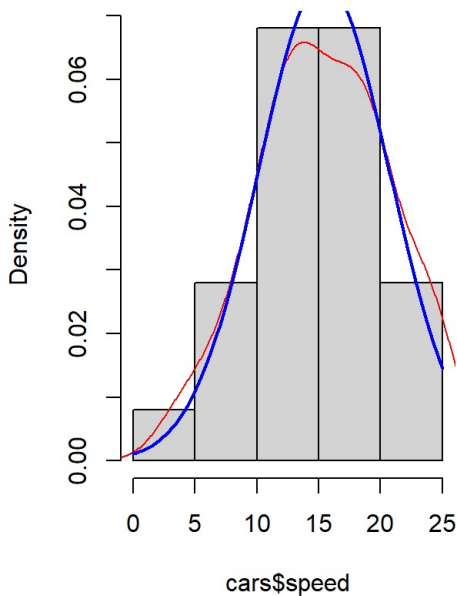
QQPlot de Distancia



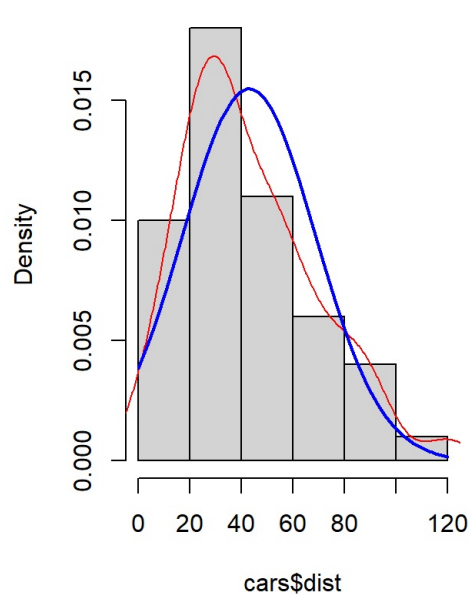
```
# Histogramas con la curva de distribución teórica
par(mfrow=c(1,2))
hist(cars$speed, freq=FALSE, main="Histograma de Velocidad")
lines(density(cars$speed), col="red")
curve(dnorm(x, mean=mean(cars$speed), sd=sd(cars$speed)), add=TRUE, col="blue", lwd=2)

hist(cars$dist, freq=FALSE, main="Histograma de Distancia")
lines(density(cars$dist), col="red")
curve(dnorm(x, mean=mean(cars$dist), sd=sd(cars$dist)), add=TRUE, col="blue", lwd=2)
```

Histograma de Velocidad



Histograma de Distancia



#Sesgo y curtosis Calculamos el

coeficiente de sesgo y el coeficiente de curtosis para las dos variables.

```
library(e1071)
skewness(cars$speed)
```

```
## [1] -0.1105533
```

```
kurtosis(cars$speed)
```

```
## [1] -0.6730924
```

```
skewness(cars$dist)
```

```
## [1] 0.7591268
```

```
kurtosis(cars$dist)
```

```
## [1] 0.1193971
```

#Conclusión del análisis de normalidad

Velocidad (speed): La prueba de Shapiro-Wilk da un p-valor de 0.4576, lo que indica que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad. El QQPlot y el histograma también muestran un buen ajuste a una distribución normal. El coeficiente de sesgo (-0.11) y curtosis (-0.67) también sugieren que no hay una desviación importante de la normalidad.

Distancia (dist): El p-valor de Shapiro-Wilk es 0.0391, lo que indica una desviación significativa de la normalidad. El histograma muestra una distribución más asimétrica y el QQPlot sugiere algunos puntos alejados de la línea de normalidad. El sesgo (0.76) y la curtosis (0.12) también refuerzan esta observación.

#Comentario final:

En general, la variable speed se ajusta mejor a la normalidad que la variable dist, lo que podría influir en la calidad de los modelos de regresión.

##Regresion lineal

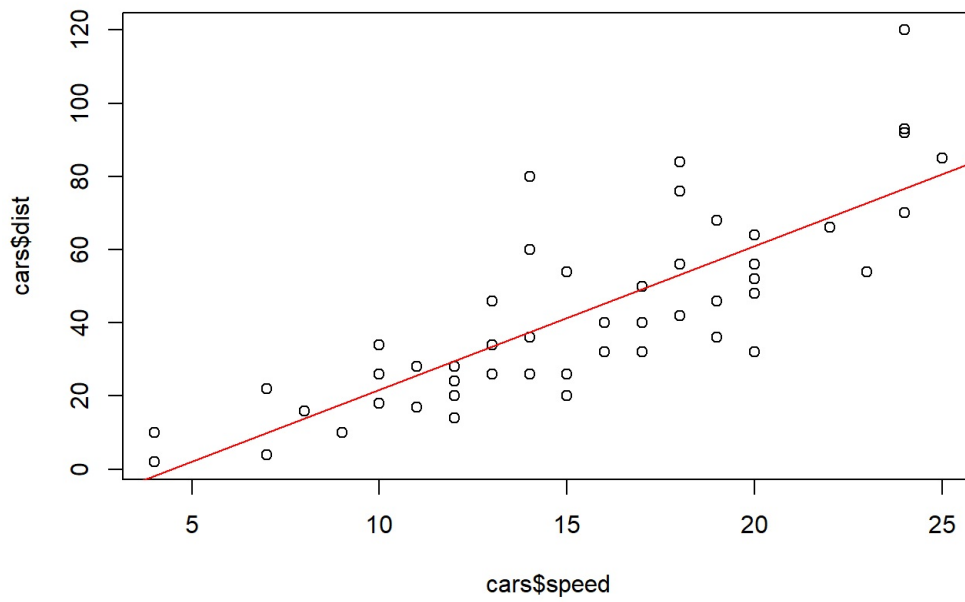
```
modelo_lineal <- lm(dist ~ speed, data = cars)
summary(modelo_lineal)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12
```

#Gráfico de regresión lineal

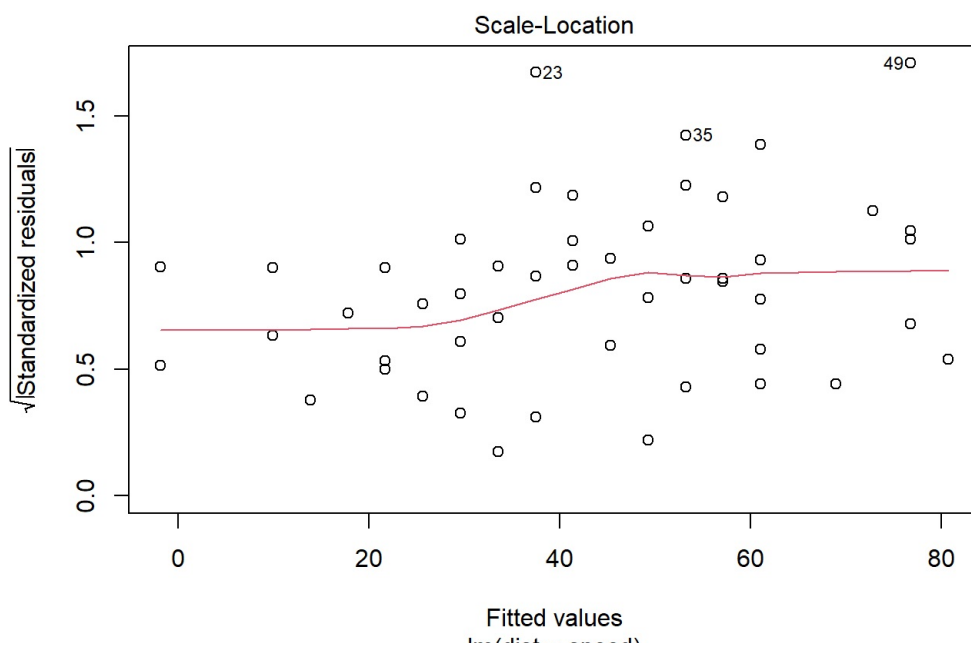
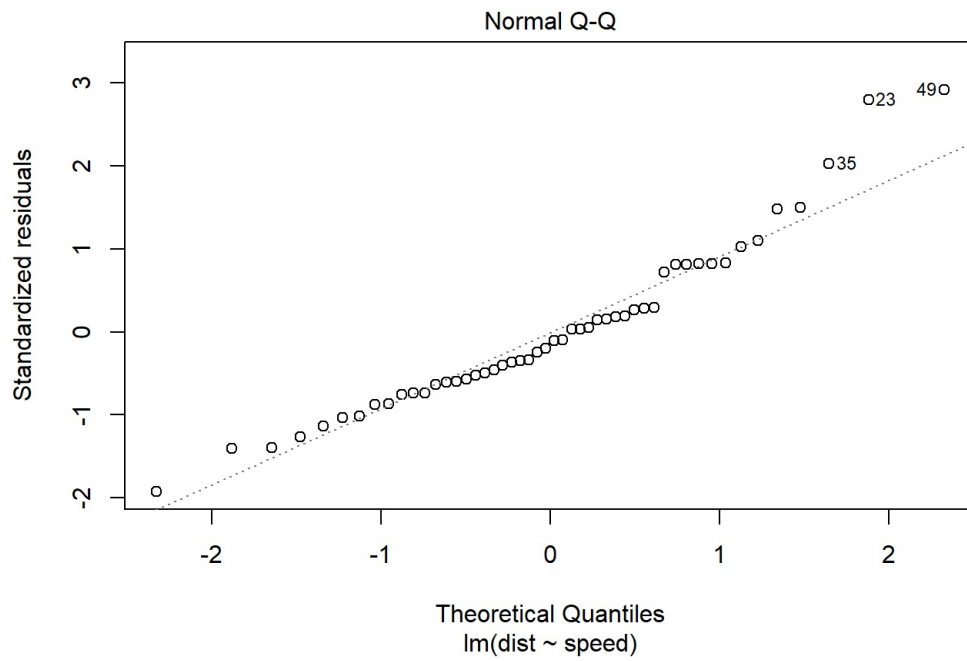
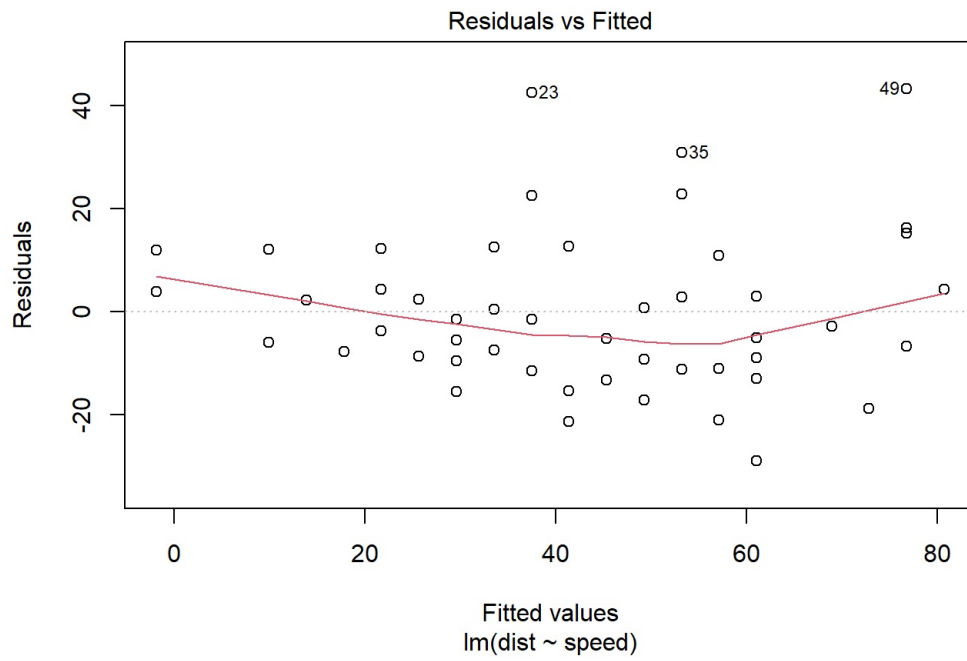
```
plot(cars$speed, cars$dist, main="Regresión lineal entre distancia y velocidad")
abline(modelo_lineal, col="red")
```

Regresión lineal entre distancia y velocidad

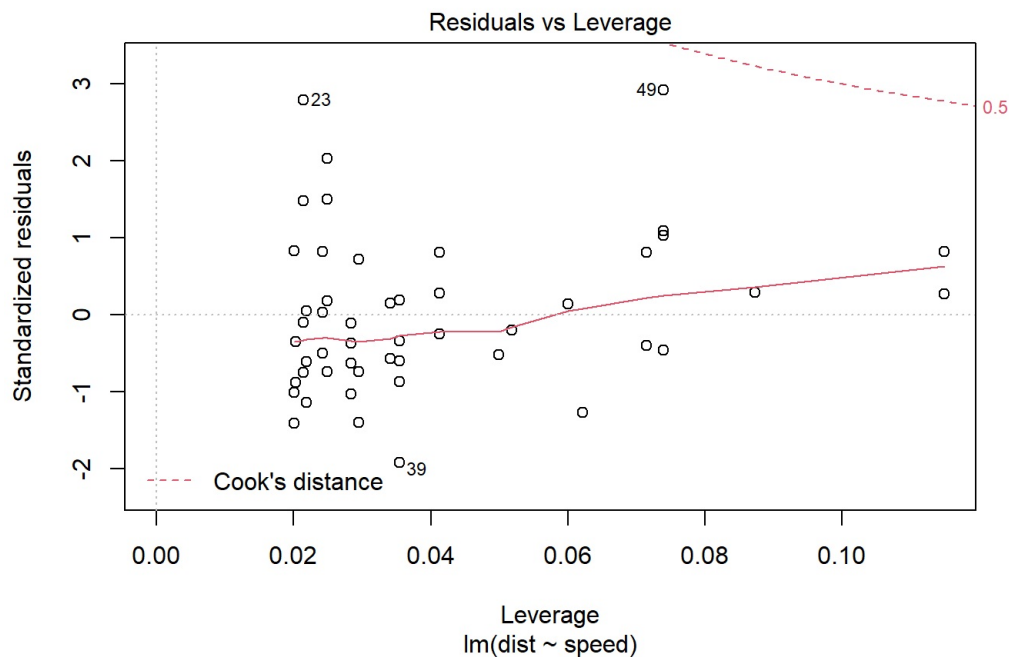


#Validación del modelo lineal

```
plot(modelo_lineal)
```



lm(dist ~ speed)



#Prueba de heterocedasticidad Usamos la prueba de Breusch-Pagan para verificar si los errores del modelo tienen varianza constante.

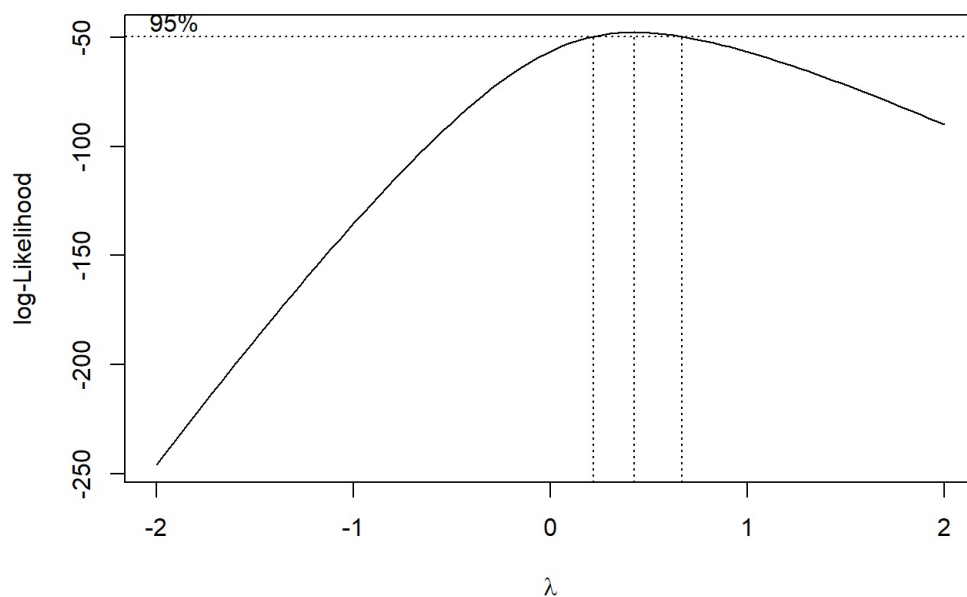
```
library(lmtest)
bptest(modelo_lineal)
```

```
##
## studentized Breusch-Pagan test
##
## data:  modelo_lineal
## BP = 3.2149, df = 1, p-value = 0.07297
```

#Conclusión sobre el modelo lineal: El modelo lineal tiene un valor de R2 ajustado de 0.6438, lo que indica que el 64% de la variabilidad en la distancia puede explicarse por la velocidad. Sin embargo, la prueba de Breusch-Pagan sugiere que podría haber heterocedasticidad (p-valor = 0.073), lo cual puede justificar explorar modelos no lineales.

##Regresión NO lineal

```
library(MASS)
boxcox_result <- boxcox(lm(dist ~ speed, data = cars))
```



```
lambda <- boxcox_result$x[which.max(boxcox_result$y)]
lambda
```

```
## [1] 0.4242424
```

#Transformamos los datos y ajustamos un nuevo modelo lineal.

```
if (lambda == 0) {
  y_trans <- log(cars$dist)
} else {
  y_trans <- (cars$dist^lambda - 1) / lambda
}

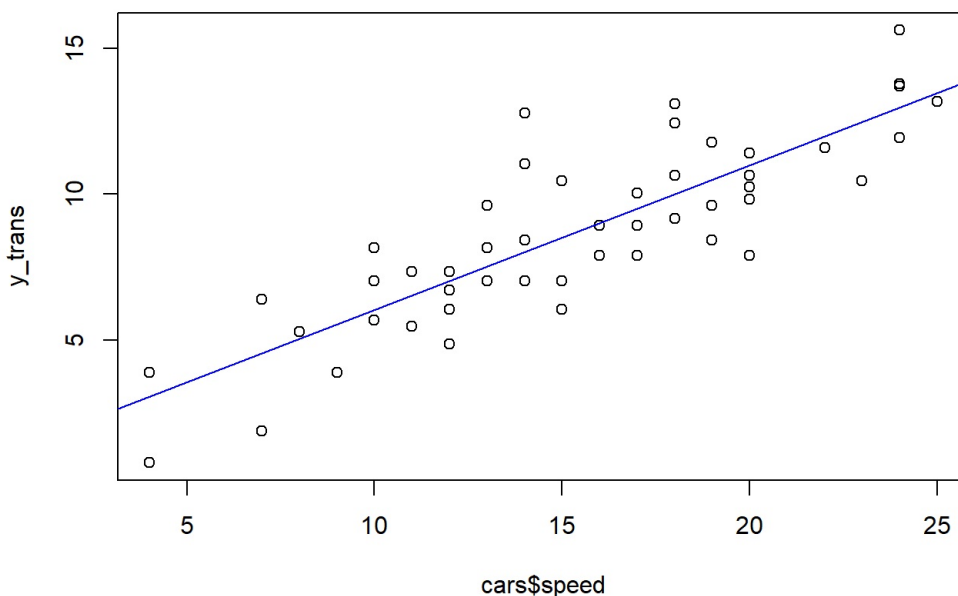
modelo_transformado <- lm(y_trans ~ speed, data = cars)
summary(modelo_transformado)
```

```
##
## Call:
## lm(formula = y_trans ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0926 -1.0444 -0.3055  0.7999  4.7520
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.08227    0.73856   1.465   0.149
## speed        0.49541    0.04541  10.910 1.35e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.681 on 48 degrees of freedom
## Multiple R-squared:  0.7126, Adjusted R-squared:  0.7066
## F-statistic: 119 on 1 and 48 DF, p-value: 1.354e-14
```

#Gráfico de regresión transformada

```
plot(cars$speed, y_trans, main="Regresión lineal con transformación de Box-Cox")
abline(modelo_transformado, col="blue")
```

Regresión lineal con transformación de Box-Cox



#Conclusión del modelo no lineal El modelo transformado tiene un R2 ajustado de 0.7066, lo que representa una mejora con respecto al modelo lineal original. Aunque la diferencia no es enorme, el nuevo modelo es más adecuado al corregir ciertos problemas de heterocedasticidad.

#Conclusión final: En resumen, la transformación de Box-Cox mejora ligeramente el ajuste del modelo al abordar algunos problemas de heterocedasticidad. Si bien el modelo lineal es más simple, la transformación ofrece un ajuste mejorado y más robusto, haciendo que sea una opción preferible en este caso.