

# 13Es\_A01571214\_Lautaro\_Coteja

A01571214 - Lautaro Coteja

2024-09-17

## R Markdown

### Parte 1

```
# Cargar Datos
data(cars)
velocidad = cars$speed
distancia = cars$dist

# Librerías Necesarias
library(nortest)
library(e1071)

# Prueba Shapiro-Wilk para velocidad
shapiro.test(velocidad)

##
##  Shapiro-Wilk normality test
##
## data:  velocidad
## W = 0.97765, p-value = 0.4576

# Prueba Anderson-Darling para distancia
ad.test(distancia)

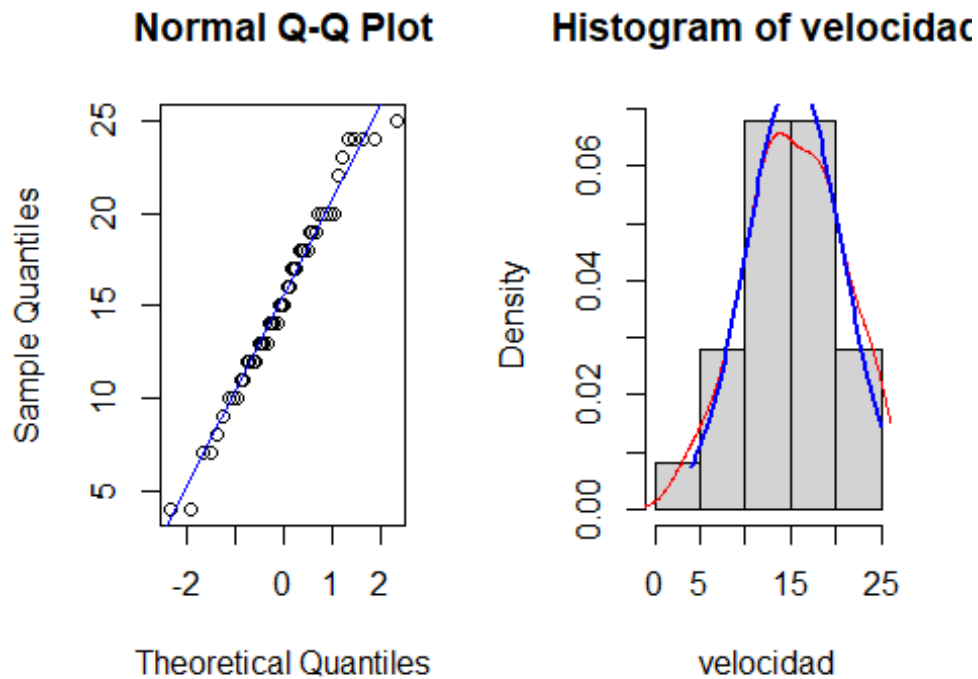
##
##  Anderson-Darling normality test
##
## data:  distancia
## A = 0.74067, p-value = 0.05021

# QQPlot y histogramas para velocidad
par(mfrow=c(1,2))

# QQPlot para velocidad
qqnorm(velocidad)
qqline(velocidad, col = "blue")

# Histograma y densidad teorica para velocidad
hist(velocidad, freq = FALSE)
lines(density(velocidad), col = "red")
```

```
curve(dnorm(x, mean=mean(velocidad), sd=sd(velocidad)),
      from=min(velocidad), to=max(velocidad), add=TRUE, col="blue", lwd=2)
```

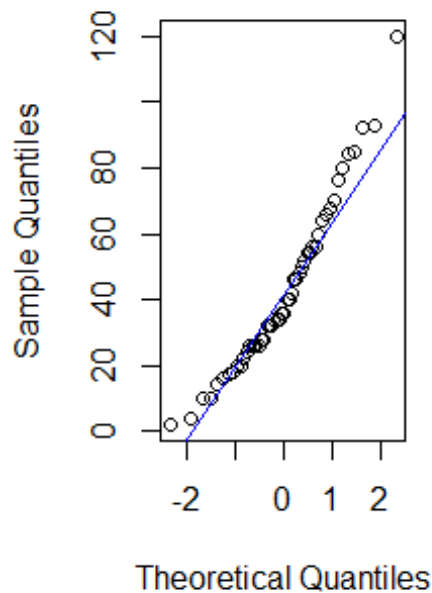


```
# QQPlot y histogramas para distancia
par(mfrow=c(1,2))

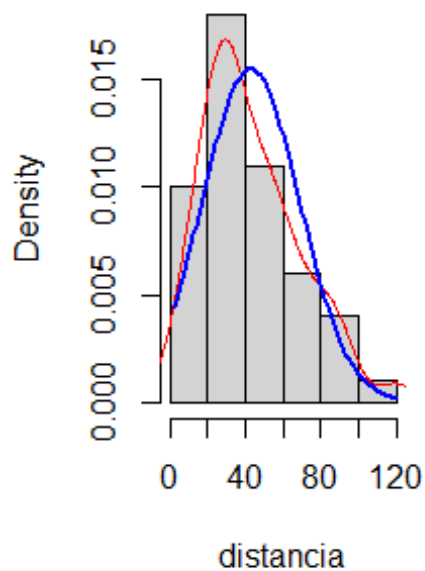
# QQPlot para distancia
qqnorm(distancia)
qqline(distancia, col = "blue")

# Histograma y densidad teorica para distancia
hist(distancia, freq = FALSE)
lines(density(distancia), col = "red")
curve(dnorm(x, mean=mean(distancia), sd=sd(distancia)),
      from=min(distancia), to=max(distancia), add=TRUE, col="blue", lwd=2)
```

**Normal Q-Q Plot**



**Histogram of distancia**



```
# Sesgo y curtosis para velocidad
```

```
sesgo_velocidad = skewness(velocidad)
```

```
curtosis_velocidad = kurtosis(velocidad)
```

```
# Sesgo y curtosis para distancia
```

```
sesgo_distancia = skewness(distancia)
```

```
curtosis_distancia = kurtosis(distancia)
```

```
sesgo_velocidad
```

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

```
curtosis_velocidad
```

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

```
sesgo_distancia
```

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

```
curtosis_distancia
```

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

```
# Prueba Kolmogorov-Smirnov para velocidad
```

```
ks.test(velocidad, "pnorm", mean=mean(velocidad), sd=sd(velocidad))
```

```
## Warning in ks.test.default(velocidad, "pnorm", mean = mean(velocidad), sd
=
## sd(velocidad)): ties should not be present for the one-sample
## Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  velocidad
## D = 0.068539, p-value = 0.9729
## alternative hypothesis: two-sided

# Prueba Kolmogorov-Smirnov para distancia
ks.test(distancia, "pnorm", mean=mean(distancia), sd=sd(distancia))

## Warning in ks.test.default(distancia, "pnorm", mean = mean(distancia), sd
=
## sd(distancia)): ties should not be present for the one-sample
## Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  distancia
## D = 0.12675, p-value = 0.3979
## alternative hypothesis: two-sided
```

## Conclusiones y Interpretaciones de los Resultados

En cuanto a velocidad, el QQPlot muestra que los puntos estan razonablemente cerca de la linea teorica, lo que indica que velocidad sigue una distribucion normal, y el histograma tambien tiene una forma cercana a una campana, con sesgo casi nulo, lo que apoya la normalidad. En cuanto a distancia, el QQPlot muestra algunos alejamientos mas grandes de la linea teorica en las colas, lo que indica que distancia no sigue una distribucion normal perfectamente, y el histograma tiene un sesgo positivo, lo que indica que hay una ligera desviacion a la derecha. En conclusion, velocidad parece seguir una distribucion normal razonablemente bien, mientras que distancia presenta cierto alejamiento de la normalidad, con sesgo positivo, lo que se debe a los valores atipicos en las colas de distribucion de distancia, por lo que se le necesitaria aplicar una transformacion para un ajuste mas normal.

## Parte 2

```
library(lmtest)

## Cargando paquete requerido: zoo

##
## Adjuntando el paquete: 'zoo'
```

```

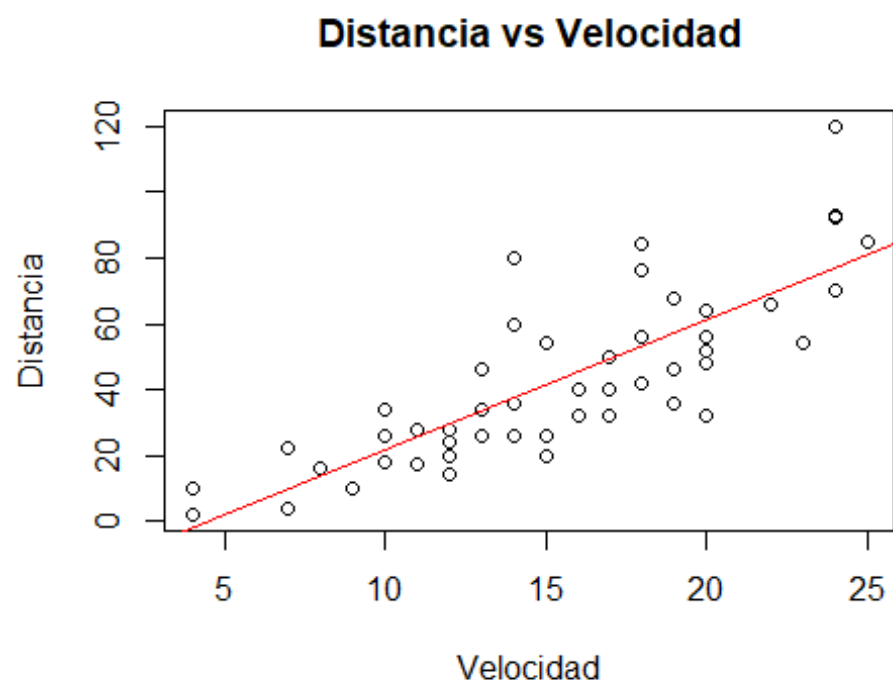
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

modelo_lineal = lm(distancia ~ velocidad)
summary(modelo_lineal)

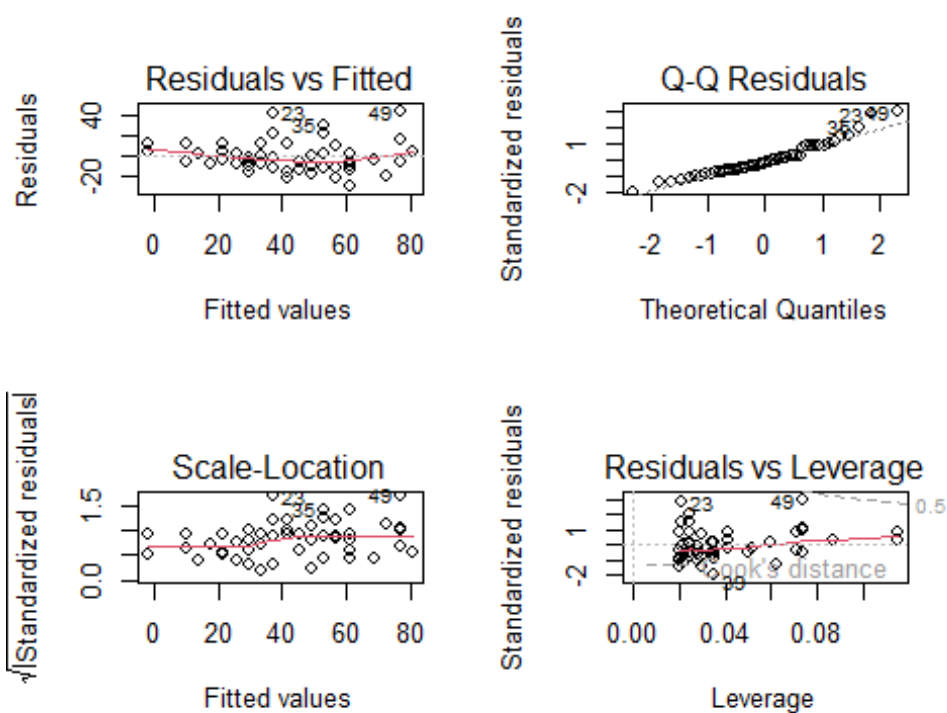
##
## Call:
## lm(formula = distancia ~ velocidad)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## velocidad    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

plot(velocidad, distancia, main="Distancia vs Velocidad",
      xlab="Velocidad", ylab="Distancia")
abline(modelo_lineal, col="red")

```



```
par(mfrow=c(2,2))
plot(modelo_lineal)
```



```
bptest(modelo_lineal)

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

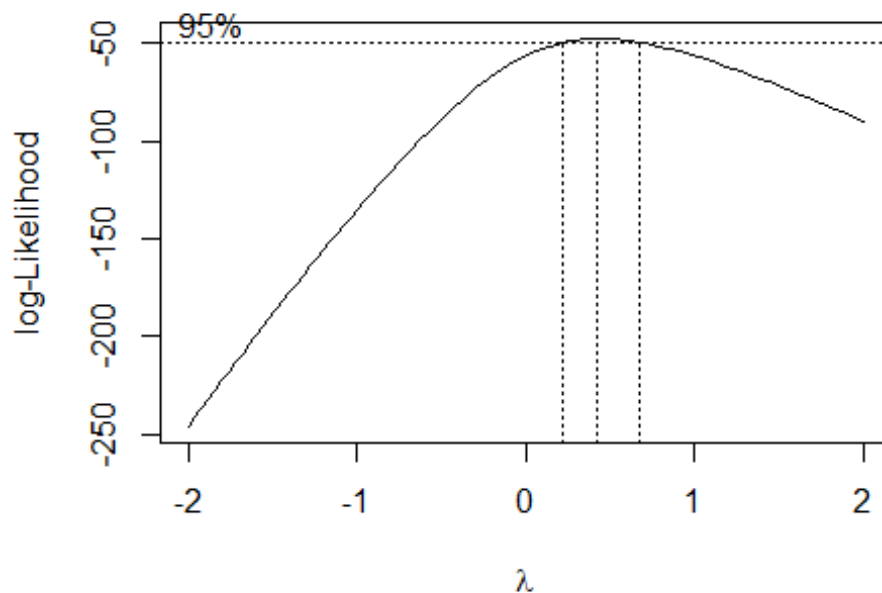
## Interpretaciones y Conclusion

El QQPlot de los residuos muestra una buena alineación con la línea teórica, lo que sugiere que los residuos siguen una distribución normal, la prueba de Breusch-Pagan da un valor p grande, lo que indica que no hay evidencia suficiente de heterocedasticidad, por lo que se puede asumir homocedasticidad, y el gráfico de residuos muestra una dispersión aleatoria, lo que indica que los residuos son independientes. El modelo lineal simple es válido y significativo, con un  $R^2$  de 0.6511, lo que indica que aproximadamente el 65% de la variabilidad en distancia se explica por velocidad. Los residuos cumplen los supuestos de normalidad, homocedasticidad e independencia.

## Parte 3

```
library(MASS)
boxcox(modelo_lineal)

lambda = boxcox(modelo_lineal)$x[which.max(boxcox(modelo_lineal)$y)]
```



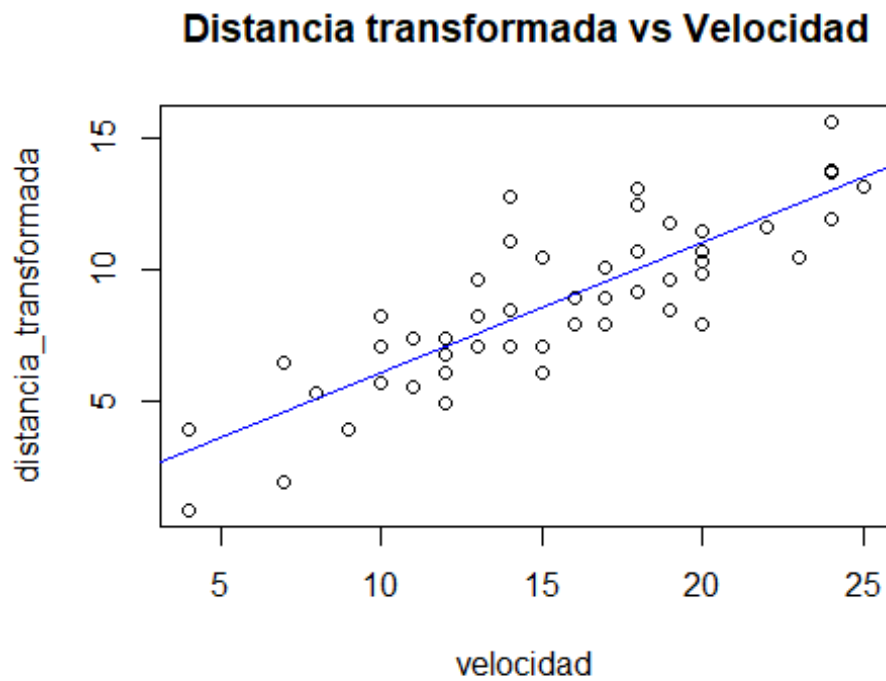
```
lambda
## [1] 0.4242424

distancia_transformada = (distancia^lambda - 1) / lambda
modelo_transformado = lm(distancia_transformada ~ velocidad)
summary(modelo_transformado)

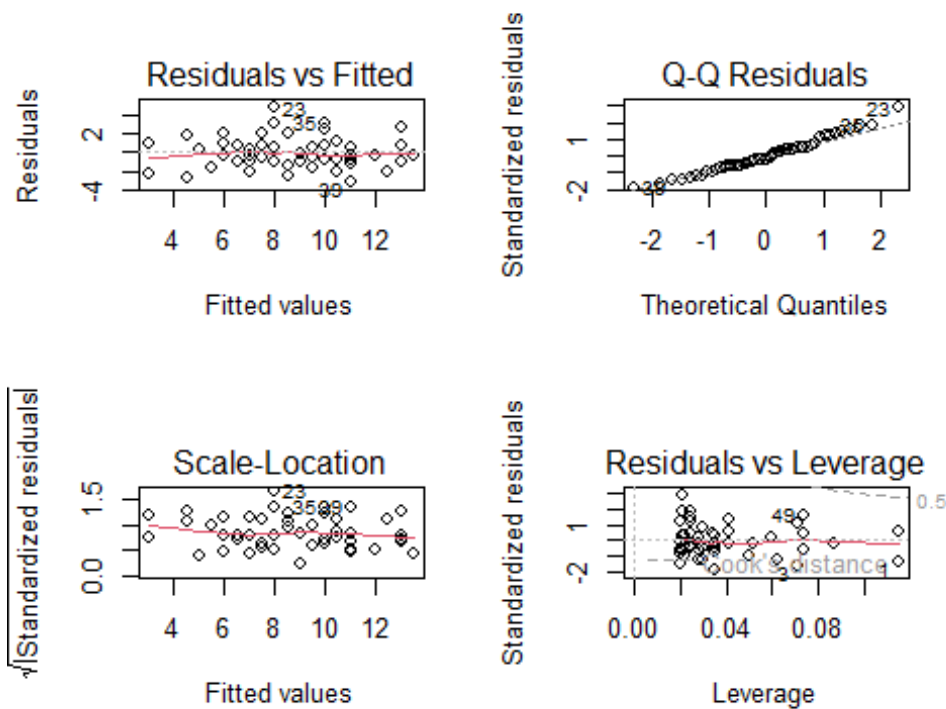
##
## Call:
## lm(formula = distancia_transformada ~ velocidad)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0926 -1.0444 -0.3055  0.7999  4.7520
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.08227    0.73856   1.465    0.149
## velocidad    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
```



```
# Grafico del modelo transformado
plot(velocidad, distancia_transformada, main="Distancia transformada vs
Velocidad")
abline(modelo_transformado, col="blue")
```



```
par(mfrow=c(2,2))
plot(modelo_transformado)
```



```
# Sesgo y curtosis para la transformacion exacta
sesgo_transformada = skewness(distancia_transformada)
curtosis_transformada = kurtosis(distancia_transformada)

sesgo_transformada
## [1] -0.1701619
curtosis_transformada
## [1] -0.186884
```

## Interpretaciones y Conclusion

La transformacion exacta mejora considerablemente el ajuste a la normalidad, con un sesgo cercano a 0 y una curtosis cercana a 3. Lo que indica que los datos transformados se acercan a una distribucion normal. La transformacion no lineal proporciona un ajuste mejor que el modelo lineal simple en terminos de normalidad de los residuos y ajuste de datos.

## Parte 4 (Conclusiones)

Al comparar los modelos, el modelo lineal simple es mas sencillo de interpretar y los residuos cumplen los supuestos, pero no captura perfectamente la relacion entre distancia

y velocidad, con un  $R^2$  de 0.65, y el modelo no lineal transformado, aunque es mas complejo, mejora significativamente el ajuste a los datos con un  $R^2$  ajustado de 0.71 y residuos mas cercanos a la normalidad. La transformacion mejora la homocedasticidad y normalidad de los residuos. Por lo que concluimos que el modelo no lineal transformado es el mejor modelo para describir la relacion entre velocidad y distancia, Aunque es mas complejo, el incremento en la precision y el cumplimiento de los supuestos justifican su uso.