

Tarea 4

COA-501 Herramientas de cómputo para investigadores

Iván F. Quiroz Ibáñez

25 de septiembre de 2022

Entrada de datos

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.1
```

```
t4.df.datos <- read_xlsx("T4_df_auto.xlsx")
head(t4.df.datos)
```

```
## # A tibble: 6 x 9
##   mpg cylinders displacement horsepower weight acceleration year origin name
##   <dbl>      <dbl>      <dbl>      <dbl>  <dbl>      <dbl> <dbl> <dbl> <chr>
## 1    18         8        307        130   3504         12     70     1 chev~
## 2    15         8        350        165   3693        11.5    70     1 buic~
## 3    18         8        318        150   3436         11     70     1 plym~
## 4    16         8        304        150   3433         12     70     1 amc ~
## 5    17         8        302        140   3449        10.5    70     1 ford~
## 6    15         8        429        198   4341         10     70     1 ford~
```

```
names(t4.df.datos) = c("ID", "millasporgalon", "cilindros", "desplazamiento",
                       "caballosdefuerza", "peso", "aceleracion", "anio", "origen", "nombre")
```

```
## Warning: The 'value' argument of 'names<-' must have the same length as 'x' as of tibble 3.0.0.
## 'names' must have length 9, not 10.
```

```
names(t4.df.datos)
```

```
## [1] "ID"                "millasporgalon"    "cilindros"         "desplazamiento"
## [5] "caballosdefuerza"  "peso"              "aceleracion"       "anio"
## [9] "origen"
```

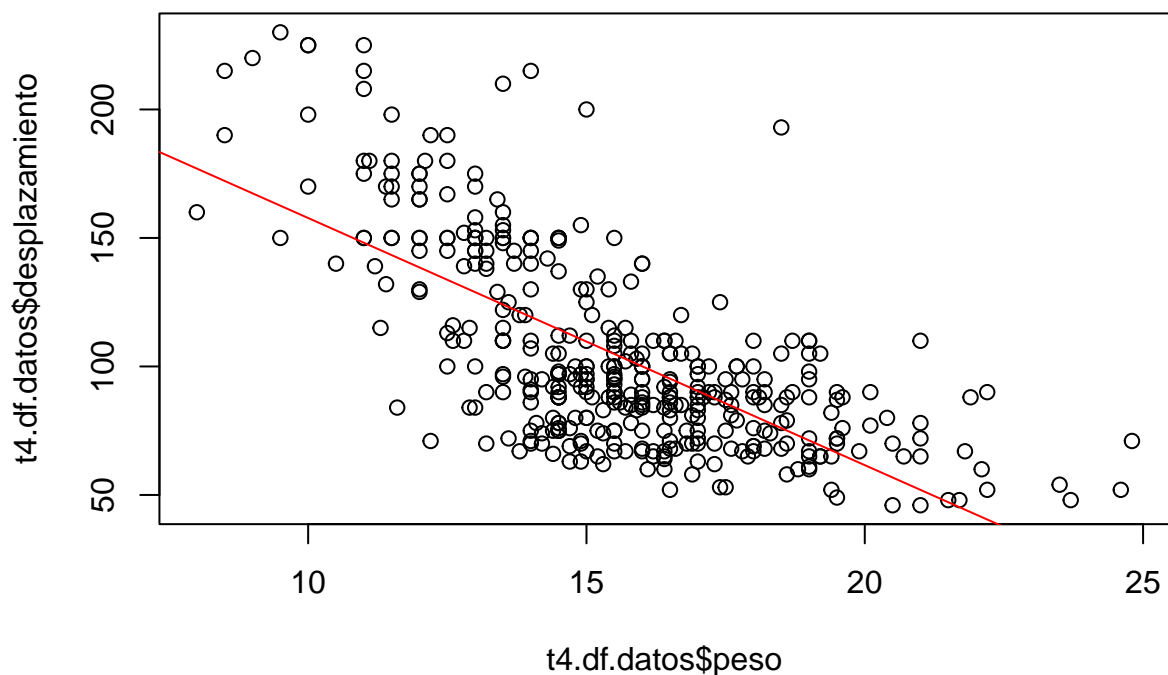
Modelo de regresión

```
ml <- lm(desplazamiento~peso,t4.df.datos)
summary(ml)
```

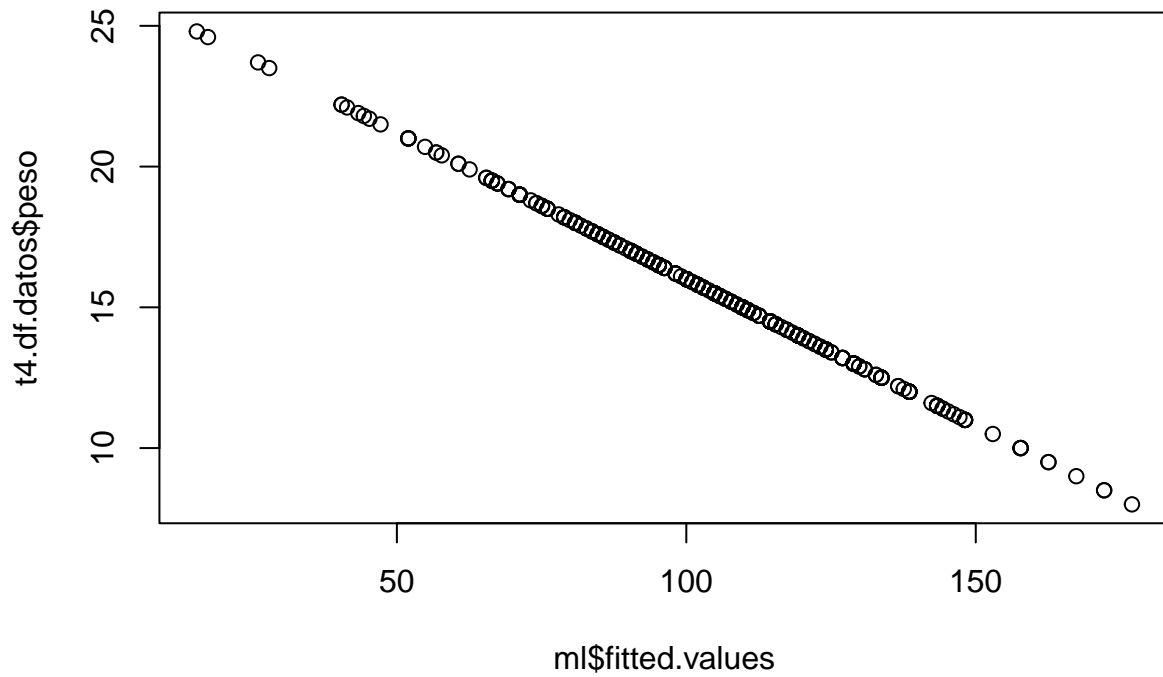
```
##
## Call:
## lm(formula = desplazamiento ~ peso, data = t4.df.datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -65.598 -17.713  -2.463   17.873  116.980
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  253.9074     8.0796   31.43  <2e-16 ***
## peso         -9.6155     0.5119  -18.78  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.93 on 390 degrees of freedom
## Multiple R-squared:  0.475, Adjusted R-squared:  0.4736
## F-statistic: 352.8 on 1 and 390 DF, p-value: < 2.2e-16
```

```
#grafica predictor vs predicho
```

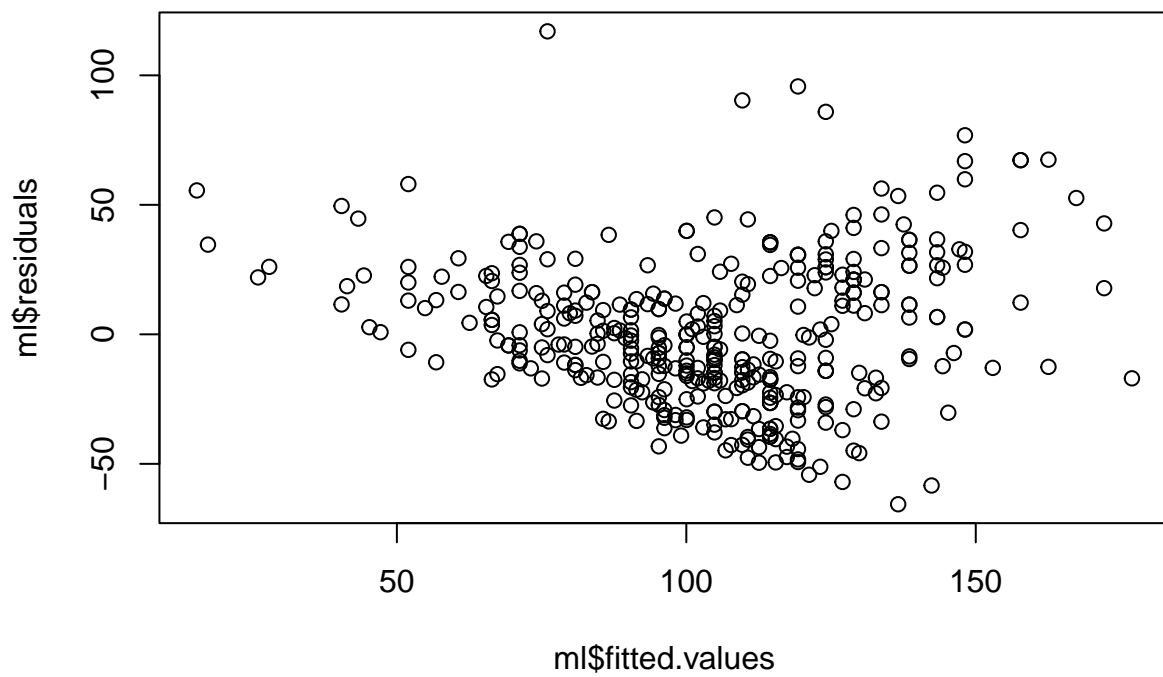
```
plot(t4.df.datos$peso,t4.df.datos$desplazamiento)
abline(ml,col="red")
```



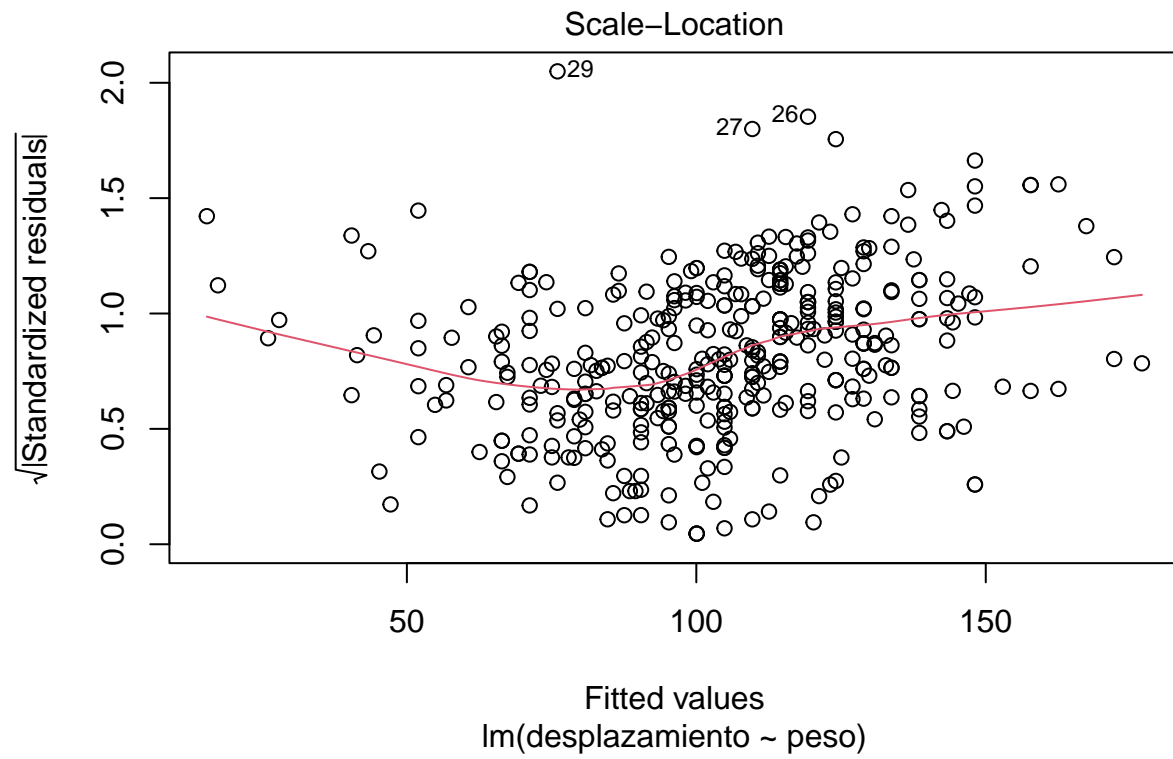
```
#grafica predictor vs ajustados  
plot(ml$fitted.values,t4.df.datos$peso)
```



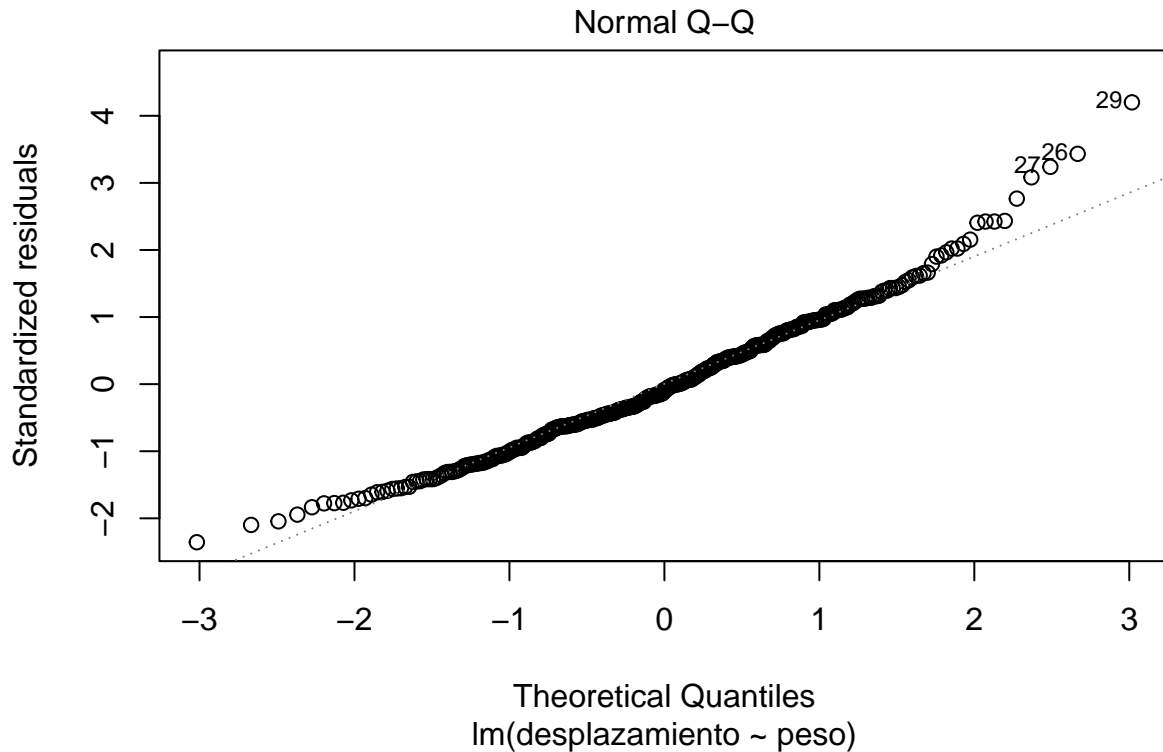
```
#grafica residuales vs predichos  
plot(ml$fitted.values,ml$residuals)
```



```
plot(ml,3)
```



```
#qqplot  
plot(m1,2)
```



```
#Kolmogorov-Smirnov test
ks.test(ml$residuals,"pnorm")
```

```
## Warning in ks.test.default(ml$residuals, "pnorm"): ties should not be present
## for the Kolmogorov-Smirnov test
```

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: ml$residuals
## D = 0.49733, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

Interpretación de gráficas y resultados

De acuerdo con los resultados, se observa que la relación de desplazamiento y peso es fuertemente lineal; realizando el ajuste del modelo lineal, se visualiza que existe heterocedasticidad en el modelo y con el qqplot y la prueba de Kolmogorov-Smirnov, se concluye que los residuales no tienen una distribución normal. Se propone usar GLM o transformar a las variables.