

ENTREGABLE SENATI

Soledad

2024-05-15

R Markdown

```
# <span style='color:orange'>Análisis de Datos del Conjunto "Auto  
MPG"</span>
```

```
## <span style='color:orange'>Paso 1: Importar y Visualizar los  
Datos</span>
```

```
```r  
Cargando la biblioteca readr para Leer el archivo CSV
library(readr)
library(dplyr)

Adjuntando el paquete: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

Importar el archivo auto_mpg.csv
auto_mpg <- read_csv("auto-mpg.csv")

Rows: 398 Columns: 9

— Column specification

Delimiter: ","
chr (2): horsepower, car name
dbl (7): mpg, cylinders, displacement, weight, acceleration, model
year, origin

i Use `spec()` to retrieve the full column specification for this
data.
i Specify the column types or set `show_col_types = FALSE` to quiet
this message.
```

```
Visualizar las primeras filas del conjunto de datos
```

```
head(auto_mpg)
```

```
A tibble: 6 × 9
```

```
mpg cylinders displacement horsepower weight acceleration `model`
year`
```

```
<dbl> <dbl> <dbl> <chr> <dbl> <dbl>
<dbl>
```

```
1 18 8 307 130 3504 12
70
```

```
2 15 8 350 165 3693 11.5
70
```

```
3 18 8 318 150 3436 11
70
```

```
4 16 8 304 150 3433 12
70
```

```
5 17 8 302 140 3449 10.5
70
```

```
6 15 8 429 198 4341 10
70
```

```
i 2 more variables: origin <dbl>, `car name` <chr>
```

Paso 2: Entender la Base de Datos

```
Renombrar las columnas al español
```

```
auto_mpg <- auto_mpg %>%
```

```
 rename(
 consumo_combustible = mpg,
 cilindros = cylinders,
 desplazamiento = displacement,
 potencia = horsepower,
 peso = weight,
 aceleracion = acceleration,
 ano_modelo = `model year`,
 origen = origin,
 nombre_auto = `car name`
)
```

```
Muestra la estructura del conjunto de datos
```

```
str(auto_mpg)
```

```
spc_tbl_ [398 × 9] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
```

```
$ consumo_combustible: num [1:398] 18 15 18 16 17 15 14 14 14 15 ...
```

```
$ cilindros : num [1:398] 8 8 8 8 8 8 8 8 8 8 ...
```

```
$ desplazamiento : num [1:398] 307 350 318 304 302 429 454 440
455 390 ...
```

```
$ potencia : chr [1:398] "130" "165" "150" "150" ...
```

```
$ peso : num [1:398] 3504 3693 3436 3433 3449 ...
```

```
$ aceleracion : num [1:398] 12 11.5 11 12 10.5 10 9 8.5 10 8.5
...
```

```
$ ano_modelo : num [1:398] 70 70 70 70 70 70 70 70 70 70 ...
```

```
$ origen : num [1:398] 1 1 1 1 1 1 1 1 1 1 ...
$ nombre_auto : chr [1:398] "chevrolet chevelle malibu" "buick
skylark 320" "plymouth satellite" "amc rebel sst" ...
- attr(*, "spec")=
.. cols(
.. mpg = col_double(),
.. cylinders = col_double(),
.. displacement = col_double(),
.. horsepower = col_character(),
.. weight = col_double(),
.. acceleration = col_double(),
.. `model year` = col_double(),
.. origin = col_double(),
.. `car name` = col_character()
..)
- attr(*, "problems")=<externalptr>
```

*# Resumen estadístico del conjunto de datos*

```
summary(auto_mpg)
```

```
consumo_combustible cilindros desplazamiento potencia
Min. : 9.00 Min. :3.000 Min. : 68.0 Length:398
1st Qu.:17.50 1st Qu.:4.000 1st Qu.:104.2 Class :character
Median :23.00 Median :4.000 Median :148.5 Mode :character
Mean :23.51 Mean :5.455 Mean :193.4
3rd Qu.:29.00 3rd Qu.:8.000 3rd Qu.:262.0
Max. :46.60 Max. :8.000 Max. :455.0
peso aceleracion ano_modelo origen
Min. :1613 Min. : 8.00 Min. :70.00 Min. :1.000
1st Qu.:2224 1st Qu.:13.82 1st Qu.:73.00 1st Qu.:1.000
Median :2804 Median :15.50 Median :76.00 Median :1.000
Mean :2970 Mean :15.57 Mean :76.01 Mean :1.573
3rd Qu.:3608 3rd Qu.:17.18 3rd Qu.:79.00 3rd Qu.:2.000
Max. :5140 Max. :24.80 Max. :82.00 Max. :3.000
nombre_auto
Length:398
Class :character
Mode :character
##
##
##
```

Visualización de Datos Gráfico de dispersión: Consumo de combustible vs. Desplazamiento del motor

*# Gráfico de dispersión*

```
plot(auto_mpg$desplazamiento, auto_mpg$consumo_combustible,
 xlab = "Desplazamiento del motor (pulgadas cúbicas)",
 ylab = "Consumo de combustible (millas por galón)",
 main = "Relación entre el consumo de combustible y
el desplazamiento del motor")
```

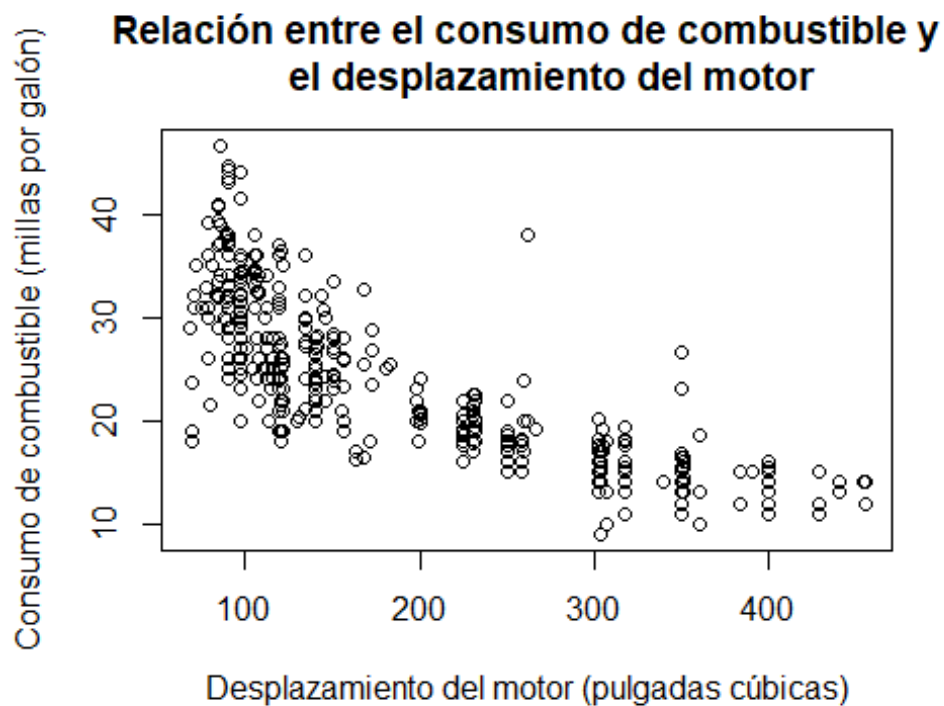


Gráfico de barras: Distribución de cilindros

```
Gráfico de barras
barplot(table(auto_mpg$cilindros),
 xlab = "Número de cilindros",
 ylab = "Frecuencia",
 main = "Distribución de cilindros")
```

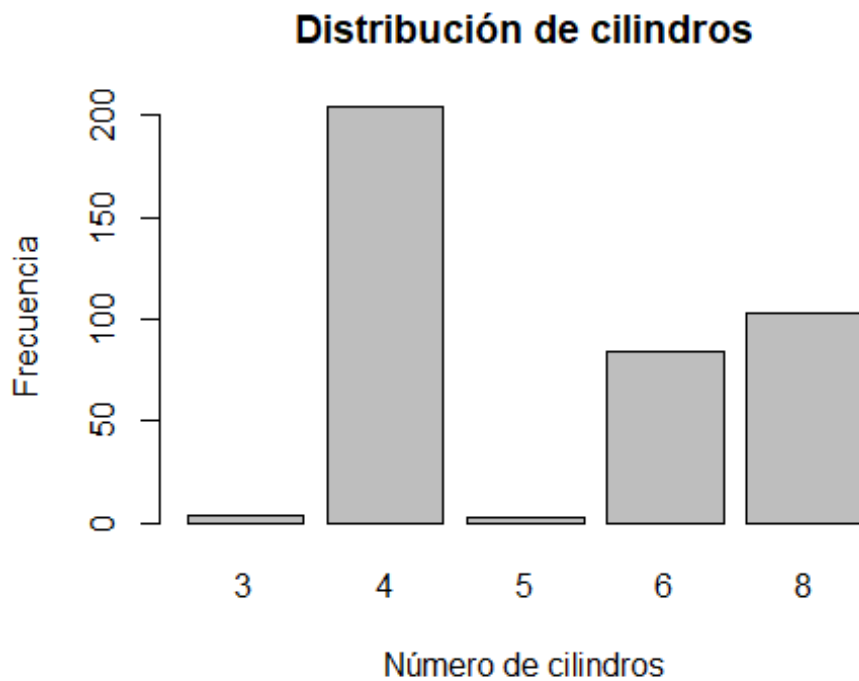


Gráfico de cajas: Peso del vehículo por número de cilindros

```
Gráfico de cajas
auto_mpg$potencia <- as.numeric(auto_mpg$potencia)

Warning: NAs introducidos por coerción

boxplot(auto_mpg$peso ~ auto_mpg$cilindros,
 xlab = "Número de cilindros",
 ylab = "Peso del vehículo (libras)",
 main = "Peso del vehículo por número de cilindros")
```

## Peso del vehículo por número de cilindros

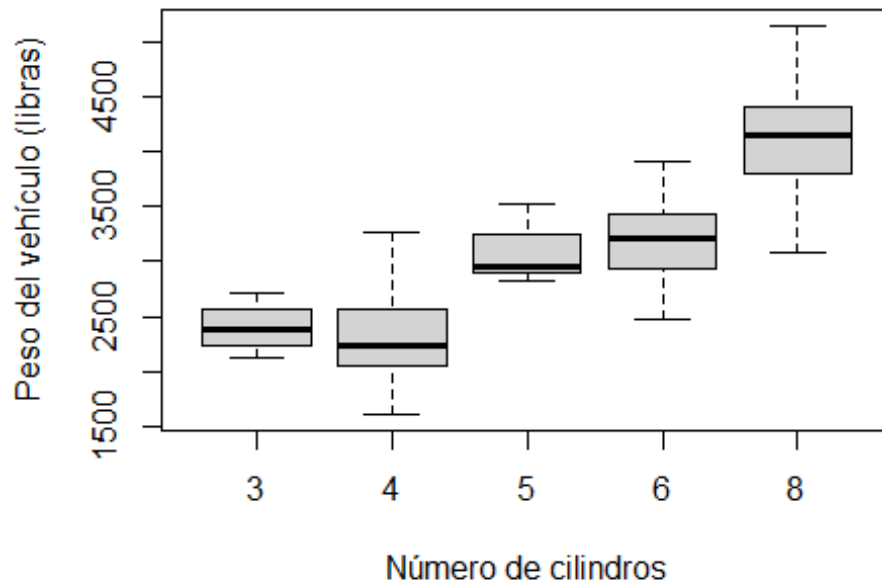
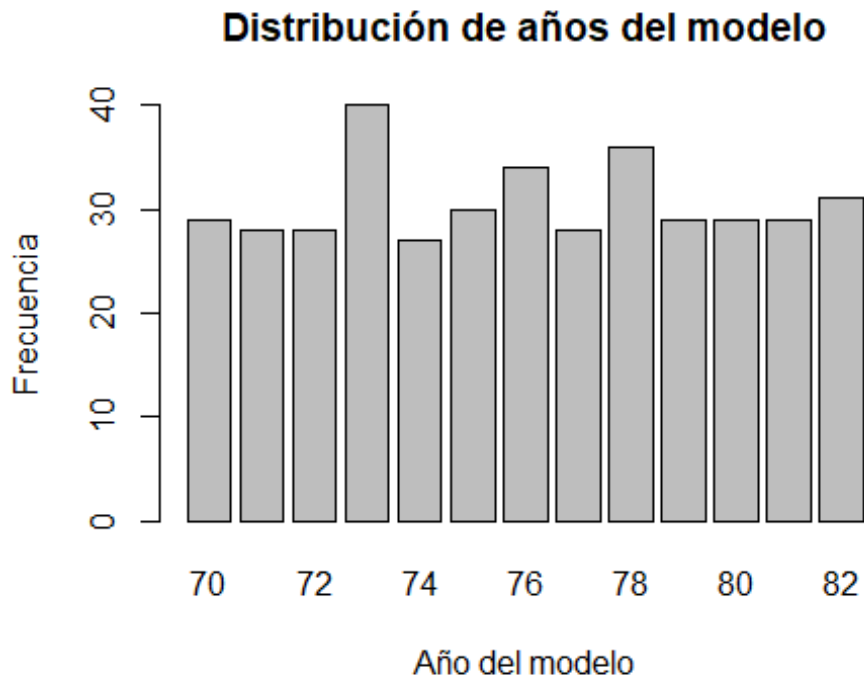


Gráfico de barras: Distribución de años del modelo

```
Gráfico de barras
barplot(table(auto_mpg$ano_modelo),
 xlab = "Año del modelo",
 ylab = "Frecuencia",
 main = "Distribución de años del modelo")
```



### Paso 3: Generar la Regresión Lineal

```
Ajustar un modelo de regresión lineal
regresion <- lm(consumo_combustible ~ cilindros + desplazamiento +
potencia + peso + aceleracion + ano_modelo + origen, data = auto_mpg)
summary(regresion)
```

```
##
Call:
lm(formula = consumo_combustible ~ cilindros + desplazamiento +
potencia + peso + aceleracion + ano_modelo + origen, data =
auto_mpg)
##
Residuals:
Min 1Q Median 3Q Max
-9.5903 -2.1565 -0.1169 1.8690 13.0604
##
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435 4.644294 -3.707 0.00024 ***
cilindros -0.493376 0.323282 -1.526 0.12780
desplazamiento 0.019896 0.007515 2.647 0.00844 **
potencia -0.016951 0.013787 -1.230 0.21963
peso -0.006474 0.000652 -9.929 < 2e-16 ***
aceleracion 0.080576 0.098845 0.815 0.41548
ano_modelo 0.750773 0.050973 14.729 < 2e-16 ***
origen 1.426141 0.278136 5.127 4.67e-07 ***
```

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
Residual standard error: 3.328 on 384 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8182
F-statistic: 252.4 on 7 and 384 DF, p-value: < 2.2e-16
```

Paso 4: Evaluar y Ajustar la Regresión Lineal si es Necesario En este paso, se evalúa el modelo y se pueden ajustar las variables predictoras según la significancia estadística y el rendimiento del modelo.

Paso 5: Construir un Modelo de Predicción

*# Crear Los nuevos datos para La predicción*

```
nuevos_datos <- data.frame(
 cilindros = 8,
 desplazamiento = 350,
 potencia = 165,
 peso = 3693,
 aceleracion = 11.5,
 ano_modelo = 70,
 origen = 1
)
```

*# Hacer La predicción*

```
predict(regresion, nuevos_datos)
```

```
1
13.9993
```

Conclusiones Analiza los resultados de la regresión y las predicciones para sacar conclusiones sobre el modelo predictivo.

Ajuste Adicional del Modelo

*# Ajustar el modelo de regresión Lineal nuevamente*

```
auto_mpg$potencia <- as.numeric(auto_mpg$potencia)
regresion_ajustada <- lm(consumo_combustible ~ peso + ano_modelo +
 origen, data = auto_mpg)
summary(regresion_ajustada)
```

```
##
Call:
lm(formula = consumo_combustible ~ peso + ano_modelo + origen,
data = auto_mpg)
##
Residuals:
Min 1Q Median 3Q Max
-10.0019 -2.0996 -0.0485 1.7371 13.2227
##
```



```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.788e+01 3.958e+00 -4.518 8.27e-06 ***
peso -6.023e-03 2.523e-04 -23.873 < 2e-16 ***
ano_modelo 7.559e-01 4.781e-02 15.808 < 2e-16 ***
origen 1.166e+00 2.578e-01 4.524 8.04e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
Residual standard error: 3.353 on 394 degrees of freedom
Multiple R-squared: 0.8174, Adjusted R-squared: 0.816
F-statistic: 587.7 on 3 and 394 DF, p-value: < 2.2e-16
```

### Predicción con el Modelo Ajustado

```
Crear un nuevo conjunto de datos con las características del automóvil para predecir
nuevos_datos <- data.frame(
 peso = 3000,
 ano_modelo = 80,
 origen = 1
)

Hacer la predicción
prediccion <- predict(regresion_ajustada, nuevos_datos)
prediccion

1
25.68415

'''
```