

# Análisis Bootstrap Completo Dataset mtcars

Luna Turpo, Rosmery

16 de setiembre 2025

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Descripción de los Datos</b>	<b>2</b>
2.1. Dataset mtcars . . . . .	2
<b>3. Metodología Bootstrap</b>	<b>2</b>
<b>4. Bootstrap para la Media</b>	<b>3</b>
4.1. Resultados Numéricos . . . . .	3
4.2. Intervalos de Confianza al 95 % . . . . .	3
4.3. Visualizaciones de la Media . . . . .	4
<b>5. Bootstrap para Regresión</b>	<b>6</b>
5.1. Modelo de Regresión . . . . .	6
5.2. Visualizaciones de Regresión . . . . .	7
<b>6. Bootstrap para Correlación</b>	<b>9</b>
6.1. Análisis de Correlación MPG vs Peso . . . . .	9
<b>7. Resumen de Resultados</b>	<b>10</b>
7.1. Estadísticas Principales . . . . .	10
7.2. Interpretación de Resultados . . . . .	10
<b>8. Conclusiones</b>	<b>10</b>
<b>9. Código R Utilizado</b>	<b>11</b>

## 1. Introducción

El bootstrap es una técnica estadística de remuestreo que permite estimar la distribución de un estadístico sin hacer supuestos sobre la distribución poblacional. En este análisis utilizamos el dataset `mtcars` con 1000 remuestreos bootstrap para analizar diferentes estadísticas.

## 2. Descripción de los Datos

### 2.1. Dataset `mtcars`

- **Tamaño de muestra:** 32 automóviles
- **Variables:** 11 características por vehículo
- **Variable objetivo:** mpg (millas por galón)
- **Variables predictoras principales:** wt (peso), hp (caballos de fuerza)

Cuadro 1: Primeras 6 observaciones del dataset `mtcars`

Modelo	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3

## 3. Metodología Bootstrap

Se realizaron 1000 remuestreos bootstrap utilizando la librería `boot` de R. El proceso consistió en:

1. Muestreo con reposición de los 32 vehículos
2. Cálculo del estadístico de interés para cada muestra bootstrap
3. Construcción de intervalos de confianza usando diferentes métodos

## 4. Bootstrap para la Media

### 4.1. Resultados Numéricos

Cuadro 2: Estadísticas Bootstrap para la Media de MPG

Estadística	Valor
Media original	20.091
Sesgo estimado	0.049
Error estándar bootstrap	1.048
Error estándar teórico	1.065
Ratio EE_bootstrap/EE_teorico	0.984

### 4.2. Intervalos de Confianza al 95 %

Cuadro 3: Comparación de Intervalos de Confianza para la Media

Método	Límite Inferior	Límite Superior
Normal	17.988	22.096
Básico	17.969	22.058
Percentil	18.123	22.212
BCa	18.121	22.203

### 4.3. Visualizaciones de la Media

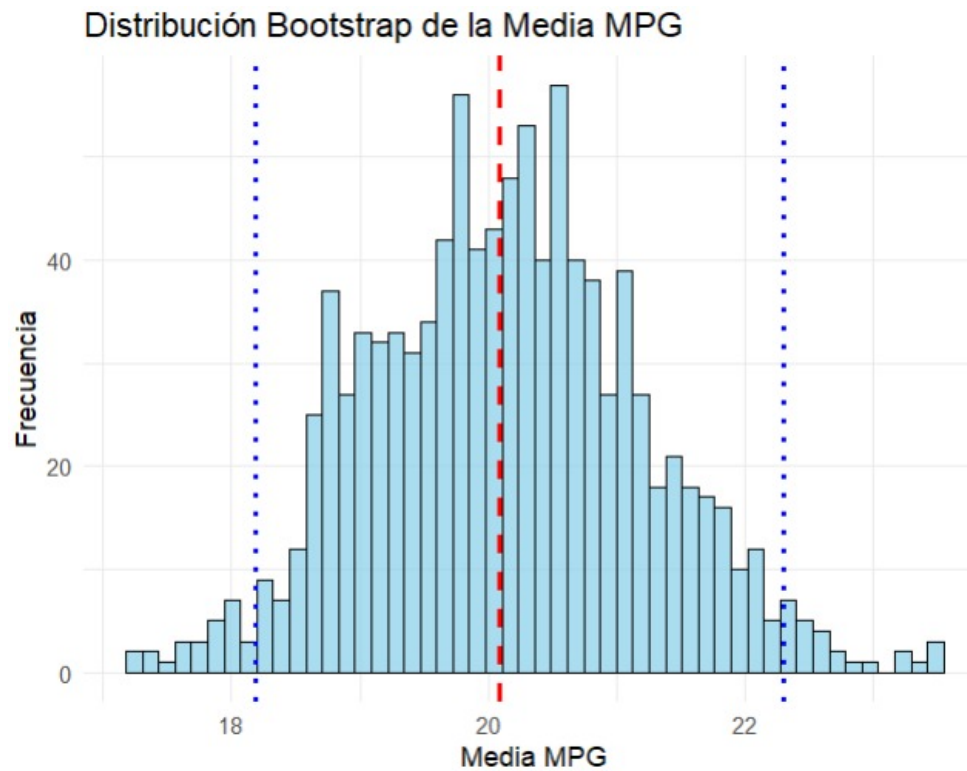


Figura 1: Distribución Bootstrap de la Media MPG. La línea roja punteada muestra la media original (20.091) y las líneas azules punteadas los percentiles 2.5 % y 97.5 %.

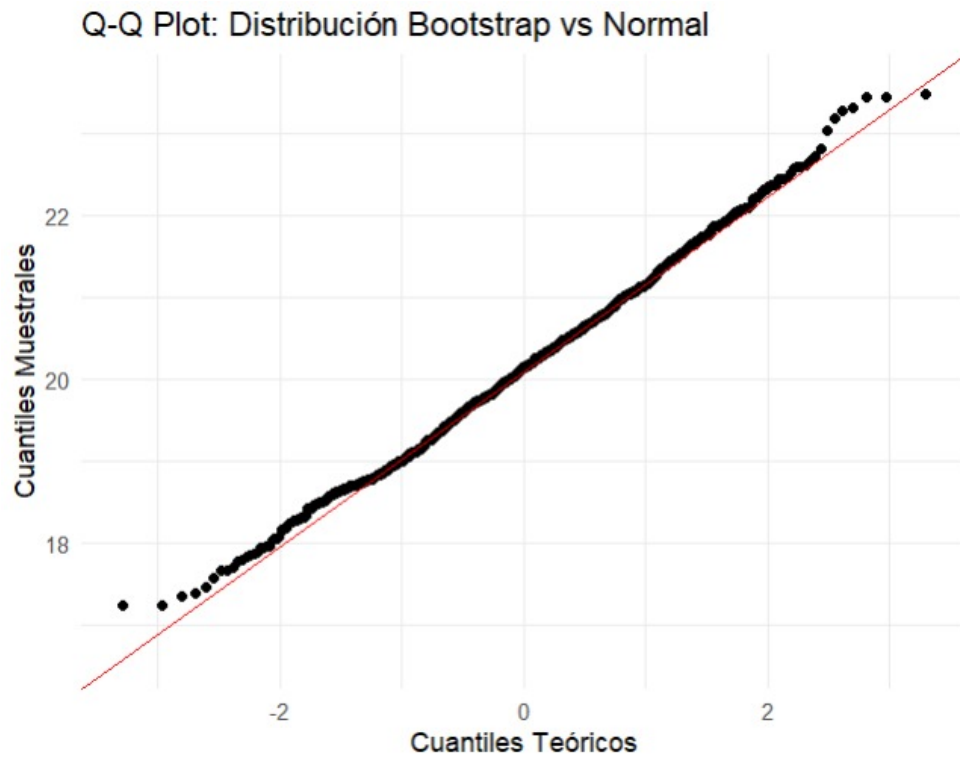


Figura 2: Q-Q Plot: Distribución Bootstrap vs Normal. Los puntos siguen aproximadamente la línea roja, indicando que la distribución bootstrap es aproximadamente normal.

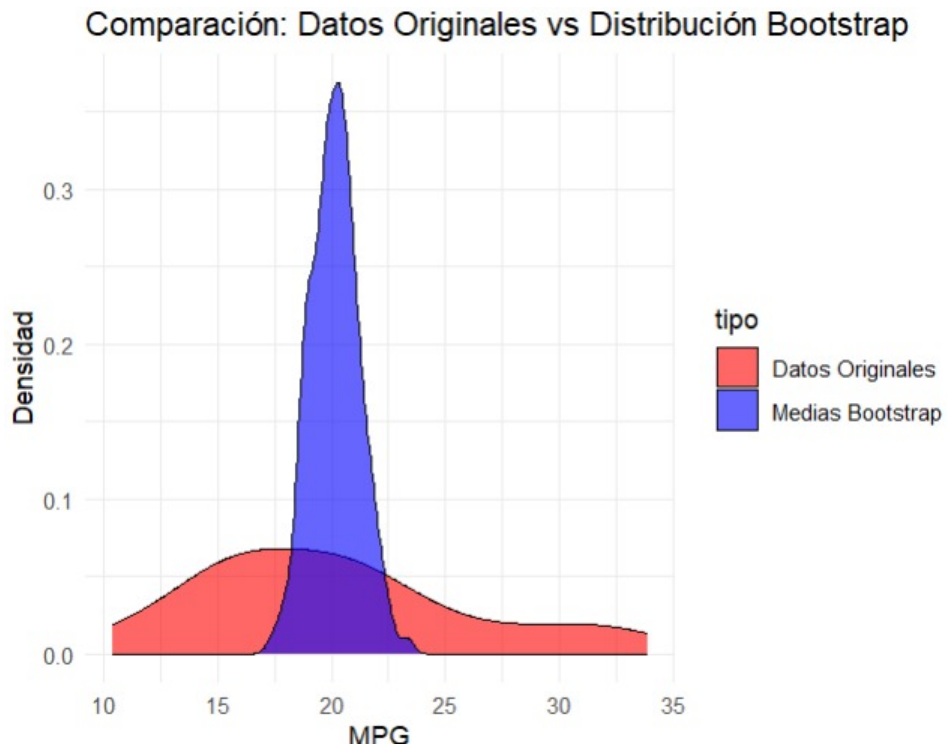


Figura 3: Comparación: Datos Originales vs Distribución Bootstrap. La distribución bootstrap (azul) está centrada alrededor de la media de los datos originales (roja).

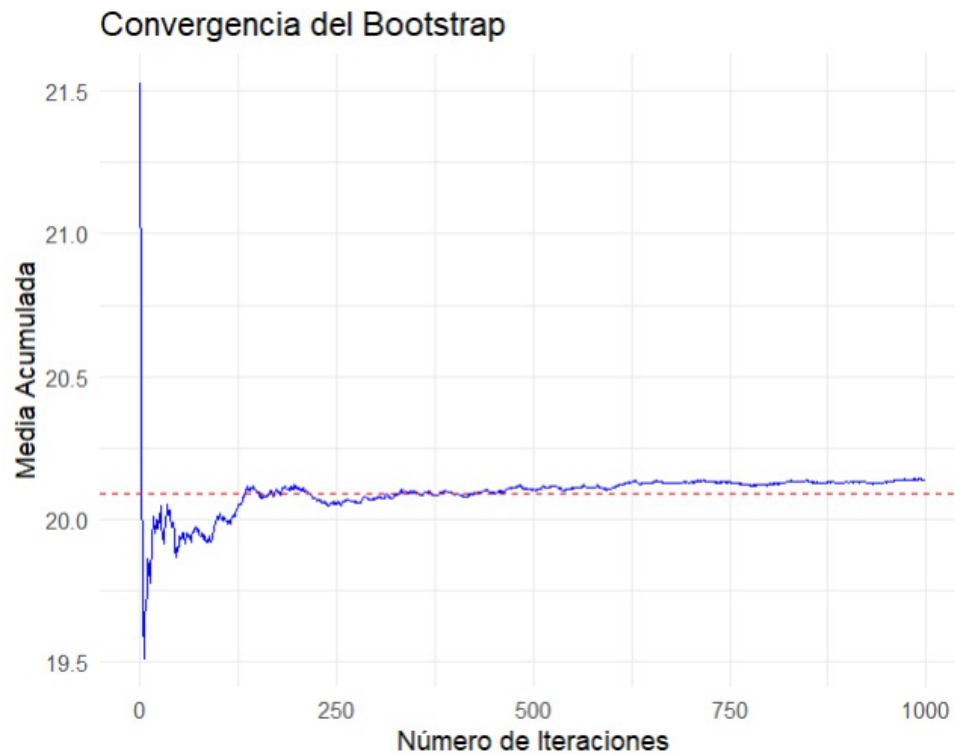


Figura 4: Convergencia del Bootstrap. La línea azul muestra cómo la media acumulada converge hacia la media poblacional (línea roja punteada) conforme aumenta el número de iteraciones.

## 5. Bootstrap para Regresión

### 5.1. Modelo de Regresión

El modelo ajustado es:  $mpg = \beta_0 + \beta_1 \cdot wt + \beta_2 \cdot hp + \varepsilon$

Cuadro 4: Coeficientes de Regresión y sus Intervalos de Confianza Bootstrap

Coeficiente	Estimación	Sesgo	EE Bootstrap	IC 95 %
Intercepto	37.227	0.023	2.170	(32.881, 41.391)
Peso (wt)	-3.878	-0.003	0.724	(-5.461, -2.399)
Potencia (hp)	-0.032	-0.001	0.008	(-0.049, -0.019)

## 5.2. Visualizaciones de Regresión

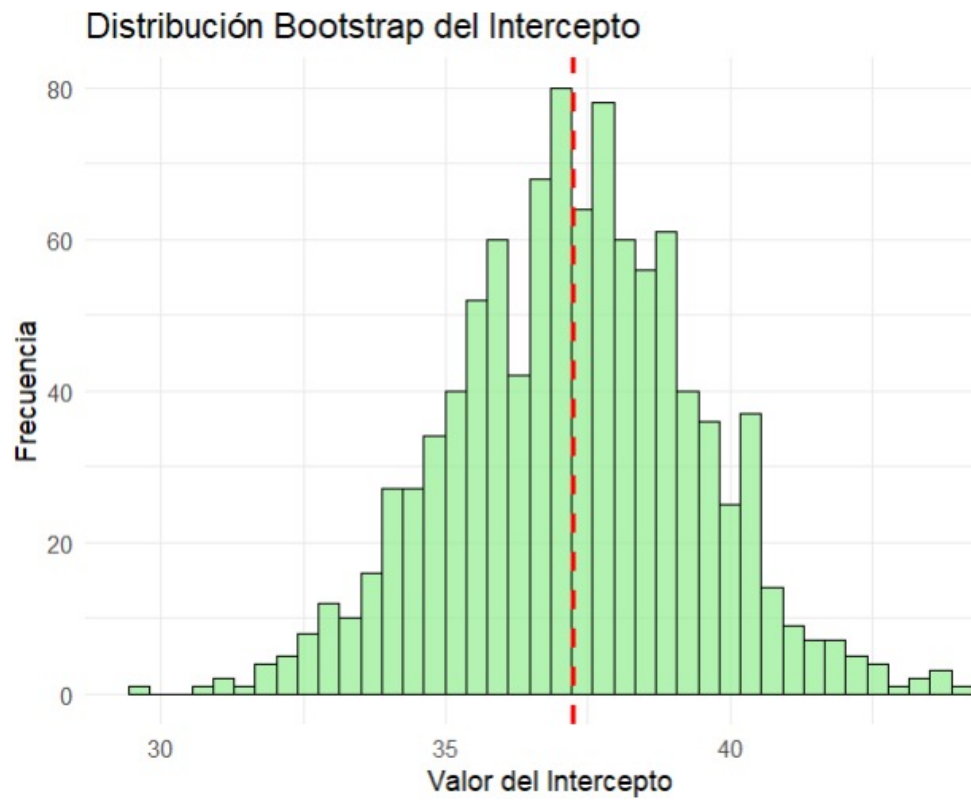


Figura 5: Distribución Bootstrap del Intercepto. La línea roja punteada marca el valor original del intercepto (37.227).

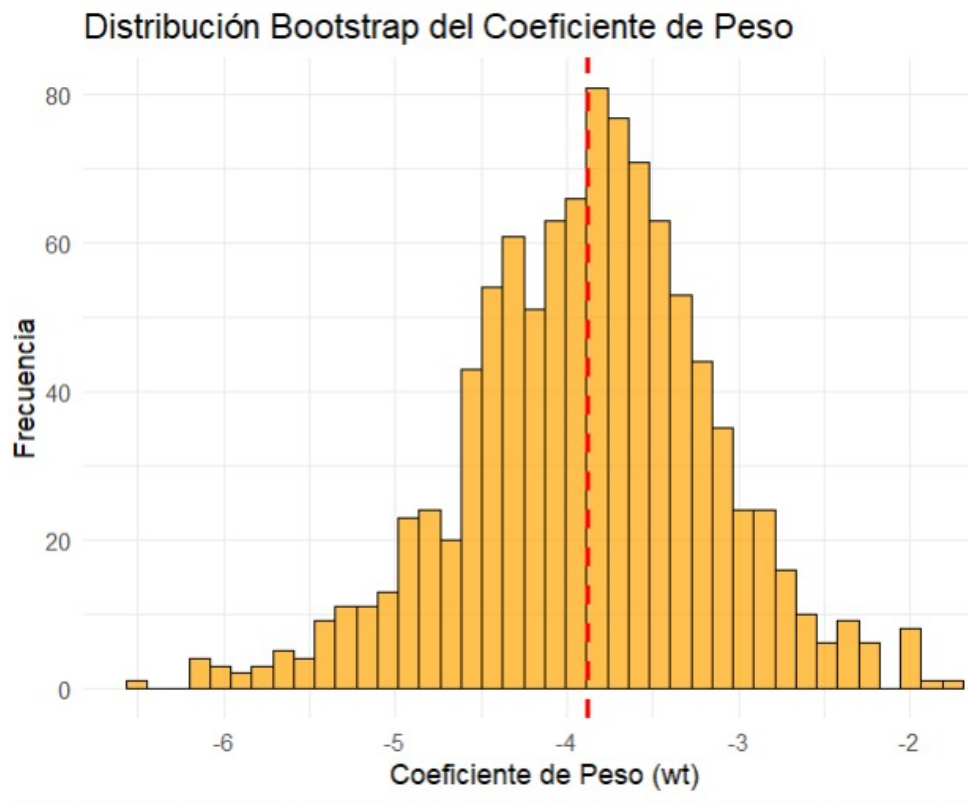


Figura 6: Distribución Bootstrap del Coeficiente de Peso. La línea roja punteada marca el valor original (-3.878).

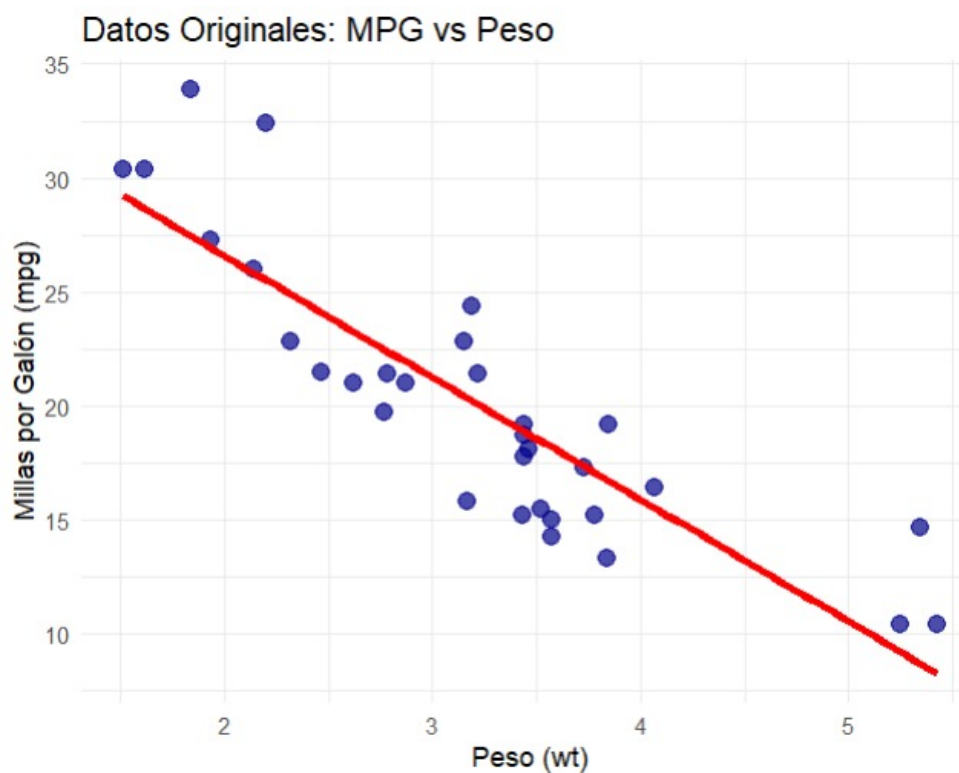


Figura 7: Datos Originales: MPG vs Peso. Se observa una clara relación negativa entre el peso del vehículo y su eficiencia de combustible.



## 6. Bootstrap para Correlación

### 6.1. Análisis de Correlación MPG vs Peso

Cuadro 5: Análisis Bootstrap de la Correlación MPG-WT

Estadística	Valor
Correlación original	-0.868
Sesgo estimado	-0.002
Error estándar bootstrap	0.034
IC 95 %	(-0.926, -0.791)

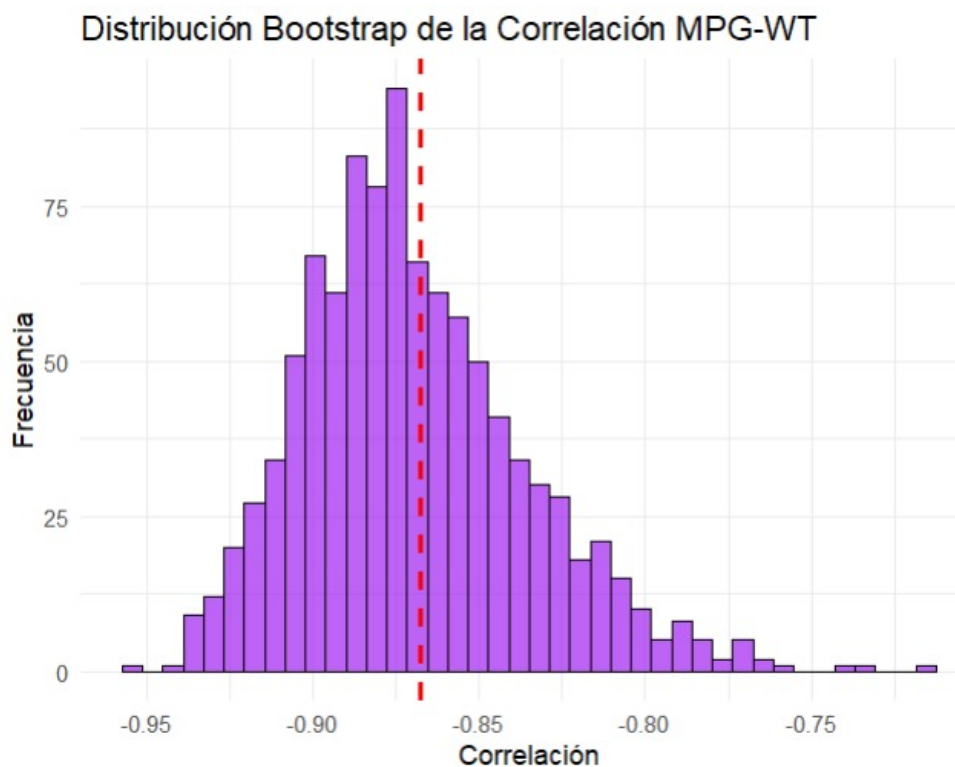


Figura 8: Distribución Bootstrap de la Correlación MPG-WT. La línea roja punteada marca la correlación original (-0.868). La distribución es aproximadamente normal.

## 7. Resumen de Resultados

### 7.1. Estadísticas Principales

Cuadro 6: Resumen Completo de Resultados Bootstrap

Estadística	Valor Original	IC 95 % Bootstrap
Media MPG	20.091	(18.121, 22.203)
Desviación estándar MPG	6.027	—
Correlación MPG-WT	-0.868	(-0.926, -0.791)
Intercepto regresión	37.227	(32.881, 41.391)
Coefficiente peso (wt)	-3.878	(-5.461, -2.399)
Coefficiente potencia (hp)	-0.032	(-0.049, -0.019)

### 7.2. Interpretación de Resultados

- **Media MPG:** La media poblacional de millas por galón está entre 18.12 y 22.20 con 95 % de confianza.
- **Correlación:** Existe una fuerte correlación negativa entre peso y eficiencia de combustible (-0.93 a -0.79).
- **Regresión:** Por cada unidad adicional de peso, el MPG disminuye entre 2.40 y 5.46 unidades.
- **Validez del Bootstrap:** El ratio entre error estándar bootstrap y teórico (0.984) indica que el bootstrap proporciona estimaciones precisas.

## 8. Conclusiones

1. El método bootstrap proporciona estimaciones robustas de intervalos de confianza sin asumir distribuciones específicas.
2. Los diferentes métodos de construcción de intervalos de confianza (Normal, Básico, Percentil, BCa) dan resultados similares, indicando estabilidad.
3. La distribución bootstrap de la media es aproximadamente normal, validando el Teorema Central del Límite.
4. El peso del vehículo es un predictor significativo y fuerte de la eficiencia de combustible.
5. Con 1000 remuestreos, el bootstrap converge adecuadamente, proporcionando estimaciones confiables.

## 9. Código R Utilizado

```
1 # Cargar librerías
2 library(ggplot2)
3 library(dplyr)
4 library(boot)
5 library(gridExtra)
6
7 # Bootstrap para la media
8 media_mpg <- function(data, indices) {
9   return(mean(data[indices, "mpg"]))
10 }
11
12 bootstrap_media <- boot(data = mtcars,
13                        statistic = media_mpg,
14                        R = 1000)
15
16 # Intervalos de confianza
17 ic_media <- boot.ci(bootstrap_media,
18                    type = c("norm", "basic", "perc", "bca"))
```

Listing 1: Script Bootstrap Principal