Inferencia Estadística

# Índice

[Índice 2](#_Toc171268022)

[1. Introducción 3](#_Toc171268023)

[2. Estadística Descriptiva 4](#_Toc171268024)

[2.1. Exploración de los datos 4](#_Toc171268025)

[2.2. Emisiones de CO2 Y Tipos de Combustible 5](#_Toc171268026)

[3. Inferencia Estadística 7](#_Toc171268027)

[**Datos de la muestra** 7](#_Toc171268028)

[**Comparación de medias. Intervalo de Confianza** 7](#_Toc171268029)

[**Contraste de Hipótesis** 9](#_Toc171268030)

[**Prueba de Bondad de Ajuste** 11](#_Toc171268031)

# Introducción

En este informe se tiene como objetivo analizar el conjunto de datos sobre las calificaciones de consumo de combustible de vehículos ligeros de 2023, proporcionado por Natural Resources Canadá. Este dataset incluye información específica del modelo sobre las calificaciones de consumo de combustible y las emisiones estimadas de dióxido de carbono para vehículos nuevos disponibles en el mercado minorista de Canadá. Los datos abarcan una amplia gama de variables como el tamaño del motor, la clase del vehículo, el tipo de transmisión, el tipo de combustible y otros parámetros relevantes. Este análisis busca proporcionar una visión detallada del rendimiento de combustible de estos vehículos, identificando patrones y tendencias significativas.

# Estadística Descriptiva

## Exploración de los datos

El primer análisis estadístico que realizaremos será la extracción de los descriptivos básicos

totales de todas las variables continuas, la media, mediana, varianza, desviación típica,

curtosis, coeficiente de asimetría y cuartiles primero y tercero.

**Datos Numéricos**

A screenshot of a graph

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

## Emisiones de CO2 Y Tipos de Combustible

**Tipos de combustible por país**

X -> Gasolina Regular  
Z -> Gasolina Premium  
D -> Diesel  
E -> E85  
B -> Electricidad  
N -> Gás Natural

[Fuente](https://www.google.com/url?q=https%3A%2F%2Fnatural-resources.canada.ca%2Fenergy-efficiency%2Ftransportation-alternative-fuels%2Fpersonal-vehicles%2Fchoosing-right-vehicle%2Fbuying-electric-vehicle%2Funderstanding-the-tables%2F21383)

A screenshot of a computer

Description automatically generated

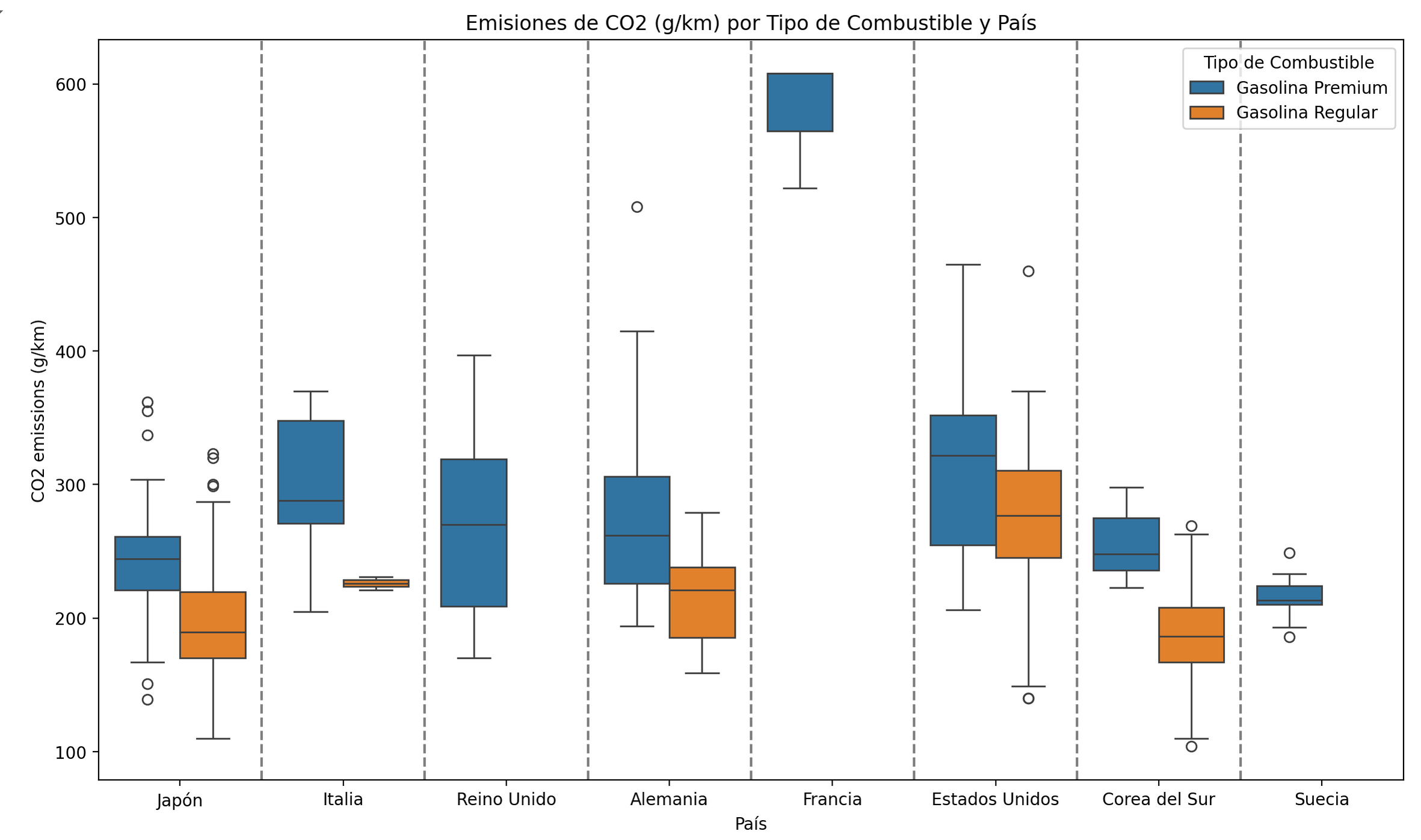
En la muestra apenas aparecen vehículos que utilicen diésel o E85. La mayoría de los vehículos utilizan gasolina regular o premium y tienen una frecuencia similar (390 y 408), por lo que solo trabajaremos con estos dos tipos de combustible.

**Emisiones de CO2 por cada tipo de combustible**

**A graph of a graph with blue and green lines

Description automatically generated with medium confidence**

**Emisiones de CO2 por tipo de combustible y país**

****

En este gráfico se puede apreciar como los vehículos con gasolina premium emiten ligeramente más CO2 que los vehículos con gasolina regular. Se puede apreciar un valor atípico en los vehículos con Gasolina Premium en Francia que emiten considerablemente más CO2 que el resto de los vehículos.

**A screenshot of a cell phone

Description automatically generated**

En Francia solo hay 3 vehículos que corresponden a vehículos de muy alta gama, y que presentan un consumo y unas emisiones de CO2 muy superiores al resto de vehículos.  
Esto explica el valor atípico en el gráfico anterior.

# Inferencia Estadística

### **Datos de la muestra**

Vamos a evaluar la influencia del tipo de combustible sobre otras variables como las emisiones de CO2 o el consumo de combustible en ciudad y en autopista.

A white rectangular sign with black text

Description automatically generated

Como podemos observar en la imagen, apenas disponemos datos de Diesel y E85 por lo que no usaremos estos en el estudio.

### **Comparación de medias. Intervalo de Confianza**

Se ha realizado una comparación de medias de distintos los tipos de métricas para los combustibles Gasolina Regular y Gasolina Premium mediante intervalo de confianza

¿Se puede afirmar que hay diferencia en las emisiones de CO2 entre los 2 tipos de combustible?

0.05

**390**  
**-39.8077** = **83.6157**

**1.9661**

±**8.3245**

**-48.1322-31.4832**

Se Rechaza la hipótesis nula H0 porque 0 no está incluido en el intervalo de confianza.

Por tanto se puede afirmar que uno de los 2 combustibles contamina más.

A graph on a white background

Description automatically generated

En el código también se han hecho pruebas donde se puede comprobar que entre los 2 tipos de combustibles también hay diferencias para el consumo de combustible en ciudad y en autopista.

### **Contraste de Hipótesis**

Vamos a realizar un nuevo contraste de hipótesis para determinar si las emisiones de CO2 son mayores en vehículos que utilizan Gasolina Premium en comparación con los que utilizan Gasolina Regular.

Planteamiento del Contraste

Utilizaremos un contraste de hipótesis para la diferencia de medias de dos muestras independientes con varianzas desconocidas y diferentes, utilizando la siguiente fórmula:

= **8.421**

**794.541**

**1.6468**

**8.421** > **1.6468**  Se puede suponer que las emisiones de CO2 son mayores en vehículos que utilizan Gasolina Premium en comparación con los que utilizan Gasolina Regular

Conclusión

Los resultados indican que las emisiones de CO2 de los vehículos que utilizan Gasolina Premium son significativamente mayores que las de los vehículos que utilizan Gasolina Regular. Este análisis respalda el estudio anterior donde mediante intervalos de confianza, se pudo comprobar que existía una diferencia apreciable en cuanto a las emisiones de CO2 entre los dos tipos de combustibles.

A screenshot of a graph

Description automatically generated

### **Prueba de Bondad de Ajuste**

En este apartado se quiere comprobar si las emisiones de CO2 de la muestra sigue algún tipo de distribución (Normal, Gamma, Exponencial…).

En este caso, visualmente la muestra se asimila mucho a una distribución normal por lo que se procede a hacer un contraste de bondad de ajuste por el método de prueba χ² de Pearson.

0.05

390

**64.2243**

**257.47**

**29.1178**

= **36.415**

**0.2157**

A graph of a graph

Description automatically generated with medium confidence

En el código se presenta una herramienta para el cálculo de la prueba de Bondad de Ajuste donde, de forma interactiva, se pueden modificar las variables y evaluar la prueba para distintas distribuciones y métricas.

A screenshot of a computer

Description automatically generated

Las diferencias que aparecen en verde representan por cada bin y se puede activar su visualización activando la casilla “*Mostrar Diferencias*”.

La muestra de Emisiones de CO2 en concreto también se ajusta a otras distribuciones como la Gamma.

A graph of a graph

Description automatically generated A graph showing the number of different colored lines

Description automatically generated with medium confidence

A graph with blue and red dots

Description automatically generated 