

Actividad 3

Parte 1: Análisis Exploratorio de Datos

Tarea 1

Ejercicio 1

Filtra el conjunto de datos para conservar solamente los precios de ventas al público en general en las provincias mencionadas de los carburantes de interés. Notar que tenemos dos condiciones que cumplir: - Provincia: Madrid, Barcelona, Valencia y Sevilla. - Tipo de venta: P (venta al público en general). para las variables Precio de gasolina 95 E5 y Precio de gasóleo A.

Respuesta Vamos a leer el archivo proporcionado por UNIR. Este viene en formato `.xlsx`, por lo que utilizaremos la función `read_excel()` del paquete `readxl` para leerlo. Esta es la opción recomendada por RStudio para leer archivos de Excel. Posteriormente crearemos un nuevo dataset con las columnas de interés así como las filas que cumplan las condiciones indicadas.

```
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
dataset_0 <- read_excel("dataset.xlsx")
dataset <- dataset_0 |>
  filter(Provincia %in% c("MADRID", "BARCELONA", "VALENCIA / VALÈNCIA", "SEVILLA") & `Tipo venta` == "P")
  select(Provincia, `Tipo venta`, `Precio gasolina 95 E5`, `Precio gasóleo A`)

head(dataset)
```

```
## # A tibble: 6 x 4
##   Provincia `Tipo venta` `Precio gasolina 95 E5` `Precio gasóleo A`
##   <chr>      <chr>      <chr>      <chr>
## 1 BARCELONA P          1,549      1,489
## 2 BARCELONA P          1,559      1,499
```

## 3 BARCELONA P	<NA>	1,499
## 4 BARCELONA P	<NA>	1,370
## 5 BARCELONA P	<NA>	<NA>
## 6 BARCELONA P	1,599	1,519

Ejercicio 2

Realiza un análisis exploratorio del precio del carburante gasolina 95 E5 por provincia. Este análisis deberá contener resúmenes numéricos y gráficos adecuados al tipo de variable con la que estás trabajando. Puede ser conveniente que, en primer lugar, definas claramente la/s variable/s con la/s que vas a trabajar en este punto.

Respuesta Como primer paso vamos a comprobar que tipo de datos son las columnas seleccionadas. Para ello, utilizamos la función `str()`, la cual nos devuelve el tipo de dato de cada columna así como los primeros valores de la misma.

```
# Comprobamos que tipos de datos tenemos en nuestro dataset
str(dataset)
```

```
## tibble [2,683 x 4] (S3: tbl_df/tbl/data.frame)
## $ Provincia      : chr [1:2683] "BARCELONA" "BARCELONA" "BARCELONA" "BARCELONA" ...
## $ Tipo venta     : chr [1:2683] "P" "P" "P" "P" ...
## $ Precio gasolina 95 E5: chr [1:2683] "1,549" "1,559" NA NA ...
## $ Precio gasóleo A : chr [1:2683] "1,489" "1,499" "1,499" "1,370" ...
```

Todas las columnas son de tipo `character`, por lo que vamos a convertir las columnas de precios a tipo `numeric` para poder realizar operaciones con ellas. Igualmente tenemos valores `NA` en las columnas de precios, por lo que vamos a eliminarlos. Otra observación es que los precios de las columnas contienen comas para separar los decimales, por lo que vamos a sustituirlos por puntos.

```
# Eliminamos los NA
dataset <- dataset |>
  filter(!is.na(`Precio gasolina 95 E5`) & !is.na(`Precio gasóleo A`))

# Sustituimos las comas por puntos
dataset$`Precio gasolina 95 E5` <- gsub(",", ".", dataset$`Precio gasolina 95 E5`)
dataset$`Precio gasóleo A` <- gsub(",", ".", dataset$`Precio gasóleo A`)

# Convertimos las columnas de precios a tipo numeric
dataset$Precio_gasolina_95_E5 <- as.numeric(dataset$`Precio gasolina 95 E5`, na.rm = TRUE)
dataset$Precio_gasoleo_A <- as.numeric(dataset$`Precio gasóleo A`, na.rm = TRUE)

# Volvemos a comprobar los tipos de datos
str(dataset)
```

```
## tibble [2,565 x 6] (S3: tbl_df/tbl/data.frame)
## $ Provincia      : chr [1:2565] "BARCELONA" "BARCELONA" "BARCELONA" "BARCELONA" ...
## $ Tipo venta     : chr [1:2565] "P" "P" "P" "P" ...
## $ Precio gasolina 95 E5: chr [1:2565] "1.549" "1.559" "1.599" "1.659" ...
## $ Precio gasóleo A : chr [1:2565] "1.489" "1.499" "1.519" "1.579" ...
## $ Precio_gasolina_95_E5: num [1:2565] 1.55 1.56 1.6 1.66 1.49 ...
## $ Precio_gasoleo_A : num [1:2565] 1.49 1.5 1.52 1.58 1.38 ...
```

Vemos que se han creado dos nuevas columnas las cuales contienen los precios en el formato correcto.

Ahora podemos proceder a realizar un análisis exploratorio de las variables de interés. Para ello, vamos a utilizar la función `summary()` para obtener un resumen de las variables numéricas.

```
library(dplyr)

# Precio Gasolina por Provincia
dataset_gasolina <- dataset |>
  filter(`Tipo venta` == "P") |>
  select(Provincia, Precio_gasolina_95_E5, Precio_gasoleo_A)

# Precio Medio Por Provincia
precio_medio_gasolina <- dataset_gasolina |>
  group_by(Provincia) |>
  summarise(
    Media = mean(Precio_gasolina_95_E5, na.rm = TRUE),
    Mediana = median(Precio_gasolina_95_E5, na.rm = TRUE),
    Minimo = min(Precio_gasolina_95_E5, na.rm = TRUE),
    Maximo = max(Precio_gasolina_95_E5, na.rm = TRUE),
    SD = sd(Precio_gasolina_95_E5, na.rm = TRUE),
    Varianza = var(Precio_gasolina_95_E5, na.rm = TRUE)
  )
print(precio_medio_gasolina)
```

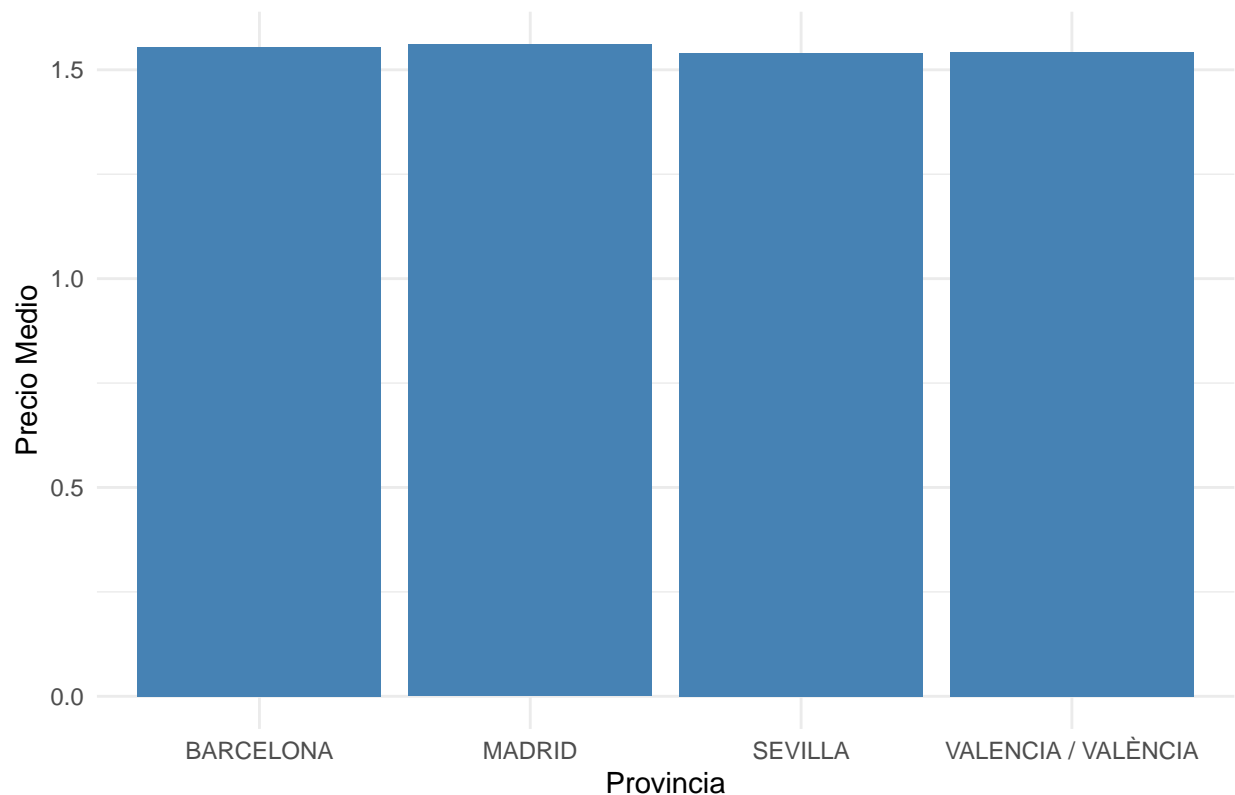
```
## # A tibble: 4 x 7
##   Provincia      Media Mediana Minimo Maximo      SD Varianza
##   <chr>          <dbl>   <dbl>  <dbl>  <dbl>   <dbl>   <dbl>
## 1 BARCELONA      1.55     1.59   1.35   1.69 0.0976  0.00953
## 2 MADRID         1.56     1.59   1.39   1.75 0.0740  0.00547
## 3 SEVILLA        1.54     1.57   1.32   1.72 0.0872  0.00761
## 4 VALENCIA / VALENCIA 1.54     1.57   1.34   1.69 0.0965  0.00931
```

Visualizamos los resultados en un gráfico de barras.

```
library(ggplot2)

# Gráfico de barras
ggplot(precio_medio_gasolina, aes(x = Provincia, y = Media)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_minimal() +
  labs(title = "Precio Medio de la Gasolina 95 E5 por Provincia",
       x = "Provincia",
       y = "Precio Medio")
```

Precio Medio de la Gasolina 95 E5 por Provincia



Resumen de percentiles de la variable Precio_gasolina_95_E5 por provincia y rango intercuartílico.

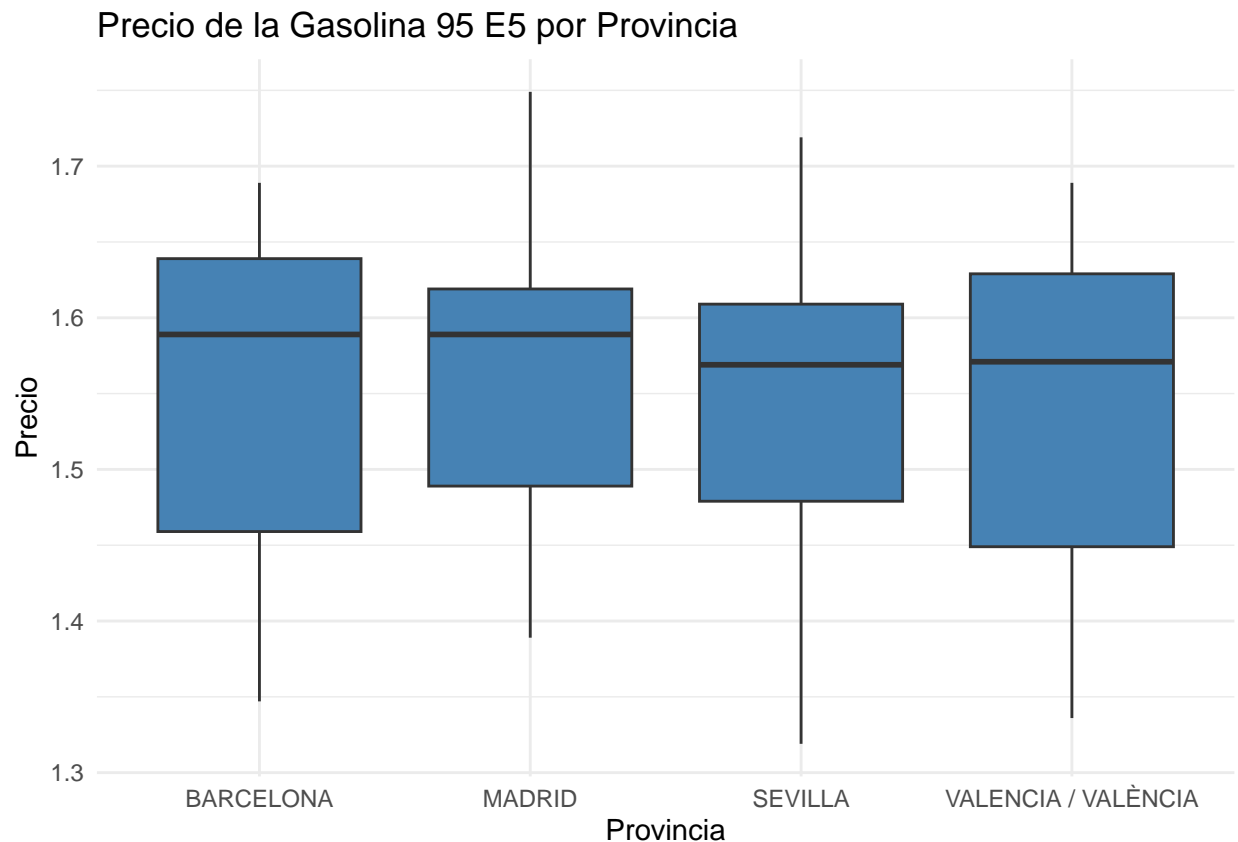
```
# Resumen de percentiles
percentiles_dataset_gasolina <- dataset_gasolina |>
  group_by(Provincia) |>
  summarise(
    Percentil_25 = quantile(Precio_gasolina_95_E5, 0.25, na.rm = TRUE),
    Percentil_75 = quantile(Precio_gasolina_95_E5, 0.75, na.rm = TRUE),
    Rango_intercuartilico = IQR(Precio_gasolina_95_E5, na.rm = TRUE)
  )
print(percentiles_dataset_gasolina)
```

```
## # A tibble: 4 x 4
##   Provincia      Percentil_25 Percentil_75 Rango_intercuartilico
##   <chr>          <dbl>         <dbl>          <dbl>
## 1 BARCELONA      1.46          1.64           0.18
## 2 MADRID         1.49          1.62           0.130
## 3 SEVILLA        1.48          1.61           0.130
## 4 VALENCIA / VALÈNCIA 1.45          1.63           0.18
```

Gráfico de cajas y bigotes.

```
# Gráfico de cajas y bigotes
ggplot(dataset_gasolina, aes(x = Provincia, y = Precio_gasolina_95_E5)) +
  geom_boxplot(fill = "steelblue") +
  theme_minimal() +
```

```
labs(title = "Precio de la Gasolina 95 E5 por Provincia",
     x = "Provincia",
     y = "Precio")
```



Ejercicio 3

A partir de los gráficos y resúmenes numéricos realizados en el punto anterior, ¿qué puedes concluir sobre el precio medio de los carburantes en cada una de las cuatro provincias analizadas? ¿Dirías que la variabilidad del precio es similar en las distintas provincias?

Respuesta Los precios de la gasolina entre provincias son muy similares. La variabilidad de los precios es similar en todas las provincias, aún así podemos observar que los precios en la provincia de Barcelona y Valencia, son ligeramente más variables que en el resto de provincias. Esto es debido a que en el gráfico de cajas y bigotes el rango intercuartílico es mayor en Barcelona y Valencia que en el resto de provincias.

Ejercicio 4

Repite los puntos 2. y 3. pero considerando ahora los precios del gasóleo A.

Respuesta Previamente ya convertimos la columna de precios relacionada con el gasóleo A a tipo `numeric`, por lo que podemos proceder a realizar el análisis exploratorio de la misma.

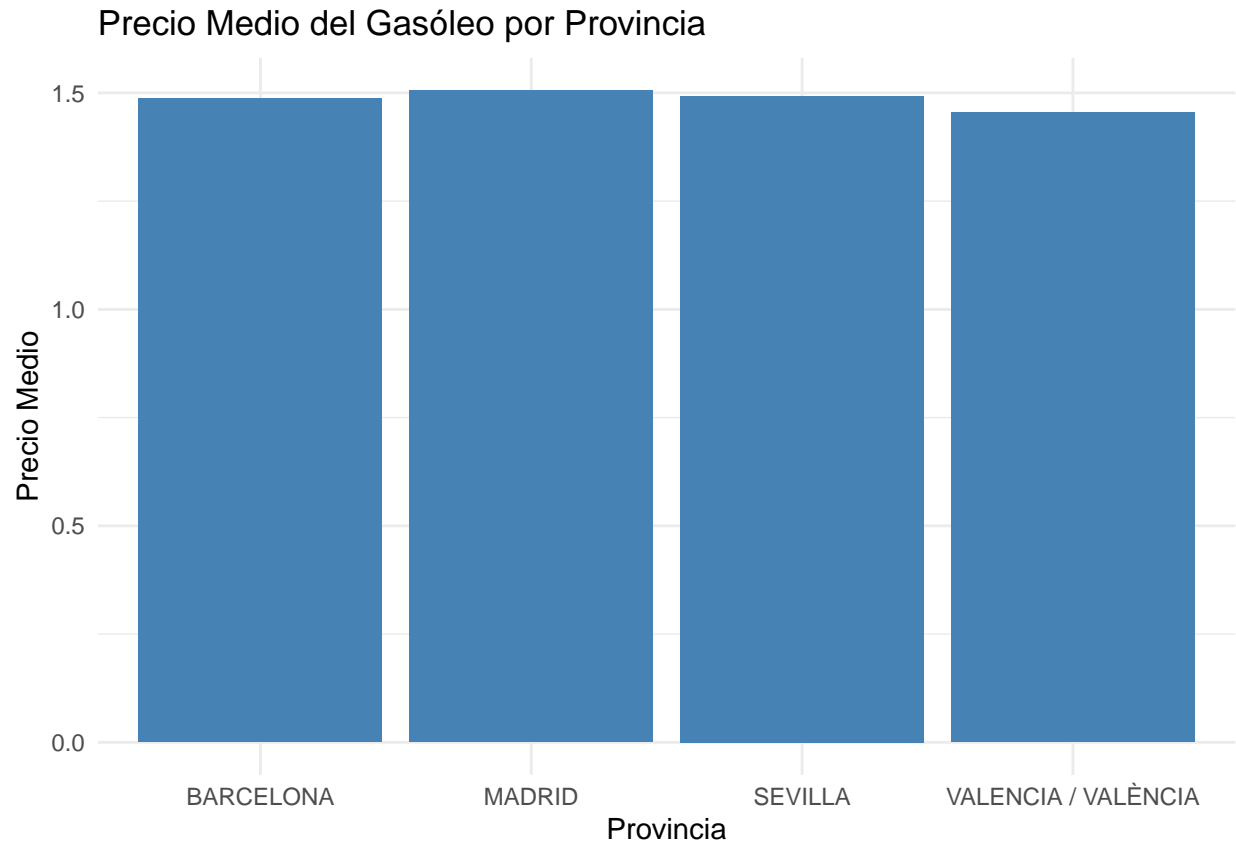
```
# Precio Medio Por Provincia
precio_medio_gasoleo <- dataset_gasolina |>
  group_by(Provincia) |>
  summarise(
    Media = mean(Precio_gasoleo_A, na.rm = TRUE),
    Mediana = median(Precio_gasoleo_A, na.rm = TRUE),
    Minimo = min(Precio_gasoleo_A, na.rm = TRUE),
    Maximo = max(Precio_gasoleo_A, na.rm = TRUE),
    SD = sd(Precio_gasoleo_A, na.rm = TRUE),
    Varianza = var(Precio_gasoleo_A, na.rm = TRUE)
  )
print(precio_medio_gasoleo)
```

```
## # A tibble: 4 x 7
##   Provincia      Media Mediana Minimo Maximo      SD Varianza
##   <chr>      <dbl>   <dbl>  <dbl>  <dbl>  <dbl>   <dbl>
## 1 BARCELONA      1.49     1.53   1.28   1.65  0.0932  0.00869
## 2 MADRID         1.51     1.56   1.25   1.71  0.101   0.0103
## 3 SEVILLA        1.49     1.55   1.22   1.73  0.120   0.0145
## 4 VALENCIA / VALENCIA 1.45     1.49   1.26   1.63  0.0951  0.00904
```

Visualizamos los resultados en un gráfico de barras.

```
library(ggplot2)

# Gráfico de barras
ggplot(precio_medio_gasoleo, aes(x = Provincia, y = Media)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_minimal() +
  labs(title = "Precio Medio del Gasóleo por Provincia",
       x = "Provincia",
       y = "Precio Medio")
```



Resumen de percentiles de la variable Precio_gasoleo_A por provincia y rango intercuartílico.

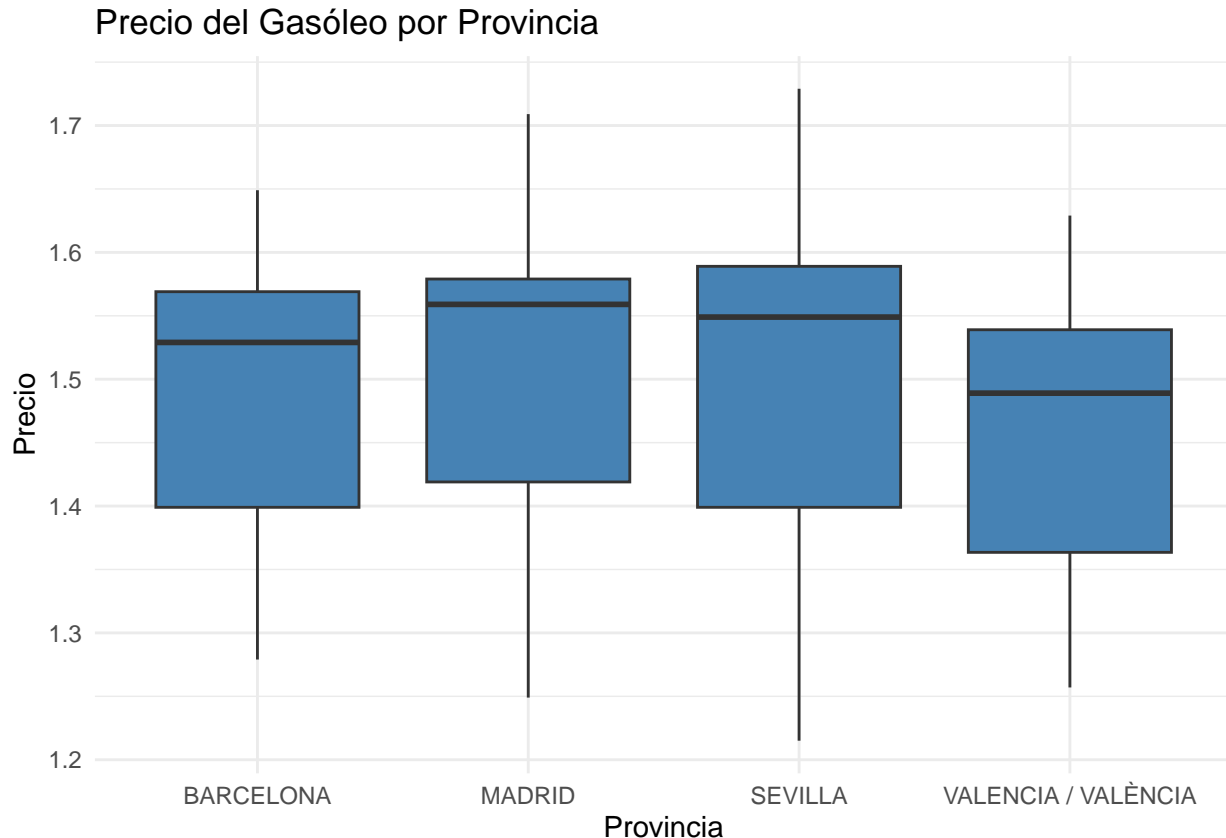
```
# Resumen de percentiles
percentiles_dataset_gasoleo <- dataset_gasolina |>
  group_by(Provincia) |>
  summarise(
    Percentil_25 = quantile(Precio_gasoleo_A, 0.25, na.rm = TRUE),
    Percentil_75 = quantile(Precio_gasoleo_A, 0.75, na.rm = TRUE),
    Rango_intercuartilico = IQR(Precio_gasoleo_A, na.rm = TRUE)
  )
print(percentiles_dataset_gasoleo)
```

```
## # A tibble: 4 x 4
##   Provincia      Percentil_25 Percentil_75 Rango_intercuartilico
##   <chr>          <dbl>         <dbl>          <dbl>
## 1 BARCELONA      1.40          1.57           0.17
## 2 MADRID         1.42          1.58           0.16
## 3 SEVILLA        1.40          1.59           0.19
## 4 VALENCIA / VALÈNCIA 1.36          1.54           0.176
```

Gráfico de cajas y bigotes.

```
# Gráfico de cajas y bigotes
ggplot(dataset_gasolina, aes(x = Provincia, y = Precio_gasoleo_A)) +
  geom_boxplot(fill = "steelblue") +
```

```
theme_minimal() +
labs(title = "Precio del Gasóleo por Provincia",
     x = "Provincia",
     y = "Precio")
```



Los precios del gasóleo repite un patrón similar al de la gasolina. Los precios entre provincias son muy similares. Igualmente, la variabilidad de los precios es similar en todas las provincias, pero podemos observar que los precios en la provincia de Sevilla son ligeramente más variables que en el resto de provincias. Por lo que para el gasóleo A, la provincia de Sevilla es la que presenta una mayor variabilidad de precios.

5. Exploratoriamente, ¿dirías que los precios de estos dos carburantes son diferentes? Los precios de los dos carburantes son muy similares. La variabilidad de los precios es también muy similar entre provincias. Por lo que podemos concluir que los precios de los dos carburantes no difieren mucho entre provincias

Parte 2: Inferencia

En este apartado queremos estudiar si existen diferencias en el precio medio de cada uno de los carburantes según la provincia y también si el precio medio es diferente según el tipo de carburante.

El procedimiento para contrastar diferencias entre dos medias que veremos aquí requiere la suposición de normalidad de los promedios, y además es necesario determinar si las varianzas de las dos poblaciones deben considerarse iguales o no. Esto último, porque el estadístico que se utiliza para llevar a cabo el contraste (y

también para obtener los intervalos de confianza) es distinto según sea la suposición de igualdad o no de las varianzas de ambas poblaciones.

Para esto último, podemos llevar a cabo un contraste sobre las varianzas poblacionales. En R podemos utilizar la función `var.test()` para realizar dicho contraste.

Una vez realizado el test sobre las varianzas, podemos utilizar la función `t.test()` para contrastar las medias.

Tarea 2

En esta segunda parte, queremos contrastar distintas hipótesis sobre los precios de los carburantes en relación con la provincia considerada.

Pregunta 6

Comencemos estudiando si hay diferencias de precios entre la provincia de Madrid y las demás. Para ello, considera las hipótesis:

- $H_0 : \mu_1 - \mu_2 = 0$
- $H_2 : \mu_1 - \mu_2 \neq 0$

Donde μ_1 es el precio medio de la gasolina 95 E5 en la provincia de Madrid y μ_2 es el precio medio de la gasolina 95 E5 en las demás provincias (Sevilla, Valencia y Barcelona).

Utilizando el nivel de significación $\alpha = 0.01$, ¿tenemos evidencia suficiente de que el precio es diferente en Madrid respecto de las demás provincias consideradas? Si la respuesta es afirmativa, ¿dónde el precio es mayor?

Ten en cuenta que, al realizar el análisis deberás considerar si las varianzas de las poblaciones debemos suponerlas iguales o no.

Respuesta Para realizar el contraste de hipótesis, vamos a utilizar la función `var.test()` para contrastar las varianzas de las poblaciones. Posteriormente, utilizaremos la función `t.test()` para contrastar las medias.

```
# Contraste de varianzas
madrid <- subset(dataset_gasolina, Provincia == "MADRID")
otras_provincias <- subset(dataset_gasolina, Provincia != "MADRID")
var.test(
  madrid$Precio_gasolina_95_E5, otras_provincias$Precio_gasolina_95_E5,
  alternative = "two.sided",
  conf.level = 0.95)

##
## F test to compare two variances
##
## data: madrid$Precio_gasolina_95_E5 and otras_provincias$Precio_gasolina_95_E5
## F = 0.60566, num df = 807, denom df = 1756, p-value = 8.636e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5389799 0.6823693
## sample estimates:
## ratio of variances
##          0.6056605
```

De estos resultados podemos observar que el **p-valor** es extremadamente pequeño. Este valor tan pequeño nos indica que las varianzas de las poblaciones son diferentes. Por lo que debemos utilizar el parámetro `var.equal = FALSE` en la función `t.test()`.

```
# Contraste de medias
t.test(
  madrid$Precio_gasolina_95_E5, otras_provincias$Precio_gasolina_95_E5,
  alternative = "two.sided",
  conf.level = 0.95,
  var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: madrid$Precio_gasolina_95_E5 and otras_provincias$Precio_gasolina_95_E5
## t = 4.0416, df = 1974.6, p-value = 5.511e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.007181954 0.020722139
## sample estimates:
## mean of x mean of y
##  1.561012  1.547060
```

El **p-valor** obtenido es muy bajo y nuestro nivel de significación es $\alpha = 0.01$. Por lo que podemos rechazar la hipótesis nula y aceptar la hipótesis alternativa. Por lo que podemos concluir que el precio de la gasolina 95 E5 en Madrid es diferente al precio de la gasolina 95 E5 en el resto de provincias.

Ejercicio 7

Repita el análisis realizado en 7, pero considerando el precio del gasóleo A.

Respuesta Para realizar el contraste de hipótesis, vamos a utilizar la función `var.test()` para contrastar las varianzas de las poblaciones. Posteriormente, utilizaremos la función `t.test()` para contrastar las medias.

```
# Contraste de varianzas
var.test(
  madrid$Precio_gasoleo_A, otras_provincias$Precio_gasoleo_A,
  alternative = "two.sided",
  conf.level = 0.95)

##
## F test to compare two variances
##
## data: madrid$Precio_gasoleo_A and otras_provincias$Precio_gasoleo_A
## F = 0.98715, num df = 807, denom df = 1756, p-value = 0.8357
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.878470 1.112177
## sample estimates:
## ratio of variances
##  0.9871509
```

El **p-valor** obtenido para el gasóleo es 0.8357 por lo que no podemos concluir que las varianzas entre provincias para el gasóleo sean diferentes. Por tanto, debemos utilizar el parámetro `var.equal = TRUE` en la función `t.test()`.

```
# Contraste de medias
t.test(
  madrid$Precio_gasoleo_A, otras_provincias$Precio_gasoleo_A,
  alternative = "two.sided",
  conf.level = 0.95,
  var.equal = TRUE)

##
## Two Sample t-test
##
## data: madrid$Precio_gasoleo_A and otras_provincias$Precio_gasoleo_A
## t = 6.5885, df = 2563, p-value = 5.374e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.02004778 0.03703788
## sample estimates:
## mean of x mean of y
##  1.505934  1.477392
```

El **p-valor** obtenido es $5.374e-11$, y nuestro nivel de significación es $\alpha = 0.01$. Por lo que podemos rechazar la hipótesis nula y aceptar la hipótesis alternativa. Por lo que podemos concluir que el precio del gasóleo A en Madrid es diferente al precio del gasóleo A en el resto de provincias.

Ejercicio 8

Finalmente realiza un contraste de hipótesis que te permita decidir si el precio medio de la gasolina 95 E5 es significativamente mayor que el del gasóleo A, diferenciando el análisis según las 4 provincias bajo estudio.

Respuesta Hipótesis nula: $\mu_1 - \mu_2 = 0$ Hipótesis alternativa: $\mu_1 - \mu_2 > 0$

Para realizar el contraste de hipótesis, vamos a utilizar la función `var.test()` para contrastar las varianzas de las poblaciones. Posteriormente, utilizaremos la función `t.test()` para contrastar las medias.

```
# Obtención de subconjuntos por provincias
madrid <- subset(dataset_gasolina, Provincia == "MADRID")
barcelona <- subset(dataset_gasolina, Provincia == "BARCELONA")
valencia <- subset(dataset_gasolina, Provincia == "VALENCIA / VALÈNCIA")
sevilla <- subset(dataset_gasolina, Provincia == "SEVILLA")
```

```
# Contraste de varianzas
var.test(madrid$Precio_gasolina_95_E5, madrid$Precio_gasoleo_A, alternative = "greater", conf.level = 0.95)

##
## F test to compare two variances
##
## data: madrid$Precio_gasolina_95_E5 and madrid$Precio_gasoleo_A
## F = 0.53158, num df = 807, denom df = 807, p-value = 1
```

```

## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
## 0.4734188      Inf
## sample estimates:
## ratio of variances
## 0.531579

var.test(barcelona$Precio_gasolina_95_E5, barcelona$Precio_gasoleo_A, alternative = "greater", conf.level = 0.95)

##
## F test to compare two variances
##
## data:  barcelona$Precio_gasolina_95_E5 and barcelona$Precio_gasoleo_A
## F = 1.0962, num df = 754, denom df = 754, p-value = 0.1037
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
## 0.9723998      Inf
## sample estimates:
## ratio of variances
## 1.096246

var.test(valencia$Precio_gasolina_95_E5, valencia$Precio_gasoleo_A, alternative = "greater", conf.level = 0.95)

##
## F test to compare two variances
##
## data:  valencia$Precio_gasolina_95_E5 and valencia$Precio_gasoleo_A
## F = 1.03, num df = 587, denom df = 587, p-value = 0.3603
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
## 0.8991144      Inf
## sample estimates:
## ratio of variances
## 1.029987

var.test(sevilla$Precio_gasolina_95_E5, sevilla$Precio_gasoleo_A, alternative = "greater", conf.level = 0.95)

##
## F test to compare two variances
##
## data:  sevilla$Precio_gasolina_95_E5 and sevilla$Precio_gasoleo_A
## F = 0.52487, num df = 413, denom df = 413, p-value = 0.4711
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
## 0.4463435      Inf
## sample estimates:
## ratio of variances
## 0.5248703

```

Provincia	p-valor	Varianzas
Madrid	1	Igual
Barcelona	0.1037	Igual
Valencia	0.3603	Igual
Sevilla	1	Igual

Procedemos con el contraste de medias para cada provincia.

```
# Contraste de medias
t.test(madrid$Precio_gasolina_95_E5, madrid$Precio_gasoleo_A, var.equal = TRUE, alternative = "greater")

##
## Two Sample t-test
##
## data: madrid$Precio_gasolina_95_E5 and madrid$Precio_gasoleo_A
## t = 12.468, df = 1614, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 90 percent confidence interval:
##  0.04941414      Inf
## sample estimates:
## mean of x mean of y
##  1.561012  1.505934

t.test(barcelona$Precio_gasolina_95_E5, barcelona$Precio_gasoleo_A, var.equal = TRUE, alternative = "greater")

##
## Two Sample t-test
##
## data: barcelona$Precio_gasolina_95_E5 and barcelona$Precio_gasoleo_A
## t = 13.843, df = 1508, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 90 percent confidence interval:
##  0.06169819      Inf
## sample estimates:
## mean of x mean of y
##  1.554634  1.486638

t.test(valencia$Precio_gasolina_95_E5, valencia$Precio_gasoleo_A, var.equal = TRUE, alternative = "greater")

##
## Two Sample t-test
##
## data: valencia$Precio_gasolina_95_E5 and valencia$Precio_gasoleo_A
## t = 15.745, df = 1174, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 90 percent confidence interval:
##  0.08078006      Inf
## sample estimates:
## mean of x mean of y
##  1.542558  1.454616
```

```
t.test(sevilla$Precio_gasolina_95_E5, sevilla$Precio_gasoleo_A, var.equal = TRUE, alternative = "greater")

##
## Two Sample t-test
##
## data: sevilla$Precio_gasolina_95_E5 and sevilla$Precio_gasoleo_A
## t = 6.3989, df = 826, p-value = 1.31e-10
## alternative hypothesis: true difference in means is greater than 0
## 90 percent confidence interval:
##  0.03739208      Inf
## sample estimates:
## mean of x mean of y
##  1.539643  1.492877
```

Resultados:

Provincia	p-valor	Hipótesis
Madrid	2.2e-16	Rechazamos H_0
Barcelona	2.2e-16	Rechazamos H_0
Valencia	2.2e-16	Rechazamos H_0
Sevilla	1.31e-10	Rechazamos H_0

Debido a los resultados obtenidos al obtener los p-valores, podemos concluir que el precio medio de la gasolina 95 E5 es significativamente mayor que el del gasóleo A en todas las provincias.

Parte 3: Conclusiones

A partir de la información obtenida del análisis realizado en los puntos anteriores, extrae conclusiones sobre el precio de los carburantes en las distintas provincias españolas.

Respuesta

Los precios de los carburantes en las distintas provincias españolas son muy similares. La variabilidad de los precios es también muy similar entre provincias. Por lo que podemos concluir que los precios de los carburantes no difieren mucho entre provincias.

La variabilidad de los precios de la gasolina 95 E5 es similar en todas las provincias, aún así podemos observar que los precios en la provincia de Barcelona y Valencia, son ligeramente más variables que en el resto de provincias. Esto es debido a que en el gráfico de cajas y bigotes el rango intercuartílico es mayor en Barcelona y Valencia que en el resto de provincias.

Igualmente existen diferencias entre los tipos de carburantes en las distintas provincias. Los precios de la gasolina 95 E5 son significativamente mayores que los del gasóleo A en todas las provincias.