

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

Tarea 1

1.

```
# Filtramos el precio por provincia y tipo de venta
df_p_tp <- df |> dplyr::select(Provincia, `Tipo venta`,
                              `Precio gasolina 95 E5`) |>
  filter(Provincia %in% c("BARCELONA", "MADRID", "SEVILLA",
                          "VALENCIA / VALÈNCIA"),
         `Tipo venta`=="P") |> na.exclude(df_p_tp$`Tipo venta`)|>

# Nos quedamos con los precios por provincia
dplyr::select(Provincia, `Precio gasolina 95 E5`)
```

Figura 1. Filtro del precio de la gasolina 95 E5 al público general en las 4 provincias.

2. Haremos un **análisis exploratorio del precio del carburante gasolina 95 E5** por provincias.

Nuestra variable a analizar será `Precio gasolina 95 E5`. Analizaremos este precio en relación con la provincia en la que se registra. Para ello, separaremos el *dataset* en 4 *subdatasets* y realizaremos los respectivos **resúmenes numéricos** por provincias.

```
# Resumen numérico de los precios por provincia

# Barcelona
df_b <- df_p_tp |> filter(Provincia=="BARCELONA")
r_b <- summary(as.numeric(df_b$`Precio gasolina 95 E5`))

# Madrid
df_m <- df_p_tp |> filter(Provincia=="MADRID")
r_m <- summary(as.numeric(df_m$`Precio gasolina 95 E5`))

# Sevilla
df_s <- df_p_tp |> filter(Provincia=="SEVILLA")
r_s <- summary(as.numeric(df_s$`Precio gasolina 95 E5`))

# Valencia
df_v <- df_p_tp |> filter(Provincia=="VALENCIA / VALÈNCIA")
r_v <- summary(as.numeric(df_v$`Precio gasolina 95 E5`))

# Separación global de los resúmenes numéricos
minimos <- c(r_b[1],r_m[1],r_s[1],r_v[1])
q_1 <- c(r_b[2],r_m[2],r_s[2],r_v[2])
mediana <- c(r_b[3],r_m[3],r_s[3],r_v[3])
media <- c(r_b[4],r_m[4],r_s[4],r_v[4])
q_3 <- c(r_b[5],r_m[5],r_s[5],r_v[5])
maximos <- c(r_b[6],r_m[6],r_s[6],r_v[6])

# Creamos un dataframe para representarlos
df_sum_g95 <- data.frame(Ciudad=c("BARCELONA", "MADRID","SEVILLA",
                                   "VALENCIA"),
                        Mínimo=minimos,
                        Cuartil_1=q_1,
                        Mediana=mediana,
                        Media=media,
                        Cuartil_3=q_3,
                        Máximo=maximos)
```

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

El *dataframe* que obtenemos a partir del anterior código es el siguiente:

	Ciudad	Mínimo	Cuartil_1	Mediana	Media	Cuartil_3	Máximo
1	BARCELONA	1.347	1.459	1.589	1.554420	1.639	1.689
2	MADRID	1.389	1.489	1.589	1.560923	1.619	1.749
3	SEVILLA	1.319	1.479	1.569	1.539643	1.609	1.719
4	VALENCIA	1.336	1.449	1.569	1.542214	1.629	1.689

Figura 2. Resumen numérico del precio de la gasolina por provincia.

Una vez realizados los resúmenes numéricos de los precios de todas las provincias a estudiar, pasamos a la **elaboración de gráficos** para representar la relación entre los precios de las 4 provincias.

- Primero, elaboraremos un diagrama de cajas y bigotes para observar si hay o no valores atípicos que afecten a nuestra media, comparar el tamaño de nuestras muestras, sus mínimos, máximos, etcétera.

```
# Creación de boxplot
ggplot(df_p_tp, aes(x = Provincia, y = `Precio gasolina 95 E5`,
  fill = Provincia)) +
  geom_boxplot() +
  xlab("Provincias") +
  ylab("Precio Gasolina 95 E5") +
  scale_fill_manual(values = c("BARCELONA" = "pink",
    "MADRID" = "lightblue",
    "SEVILLA" = "lightgreen",
    "VALENCIA / VALÈNCIA" = "purple"))
```

Al ejecutar el código, obtendremos el siguiente gráfico:

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

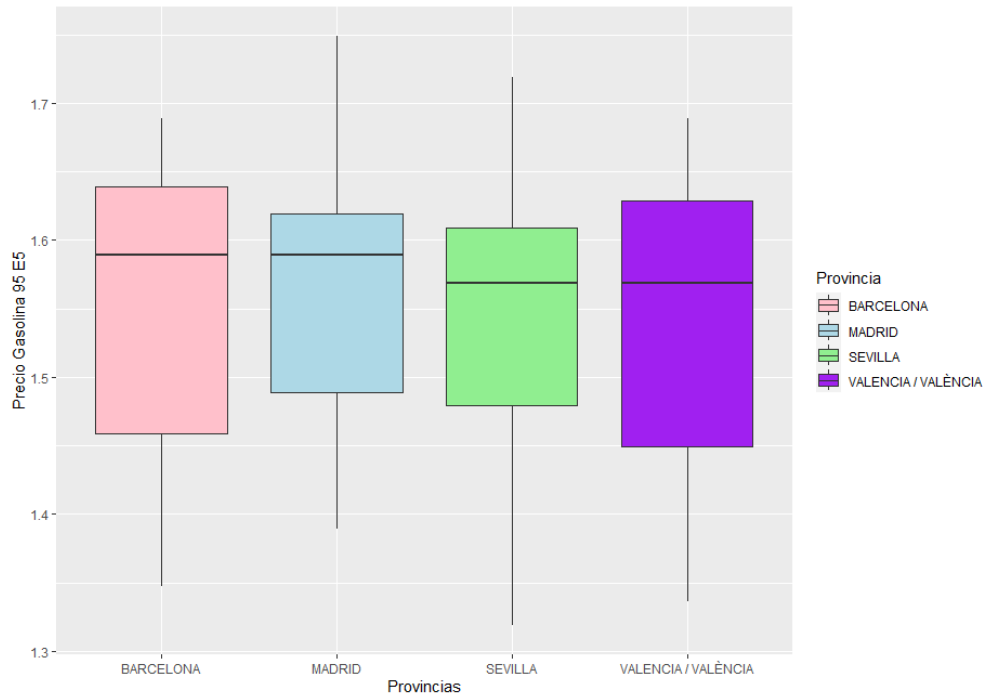


Figura 3. Diagrama de cajas y bigotes del precio de la gasolina por provincia.

También, sería conveniente comparar la media de precios por provincia en un intervalo reducido para observar la diferencia de precios de una manera más clara:

```
# Crear un dataframe con las medias de los precios y las provincias
df_means <- data.frame(Provincia = c("Barcelona", "Madrid",
                                     "Sevilla", "Valencia"),
                       Precio_Gasolina = c(r_b[4], r_m[4], r_s[4], r_v[4]))

# Graficar las barras de las medias de los precios por provincia
ggplot(df_means, aes(x = Provincia, y = Precio_Gasolina)) +
  geom_bar(stat = "identity", fill = "skyblue", width = 0.3) +
  xlab("Provincia") +
  ylab("Precio Gasolina (Media)") +
  ggtitle("Media de precios de gasolina 95 E5 por provincia")+
  coord_cartesian(ylim = c(1.535, 1.565))
```

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

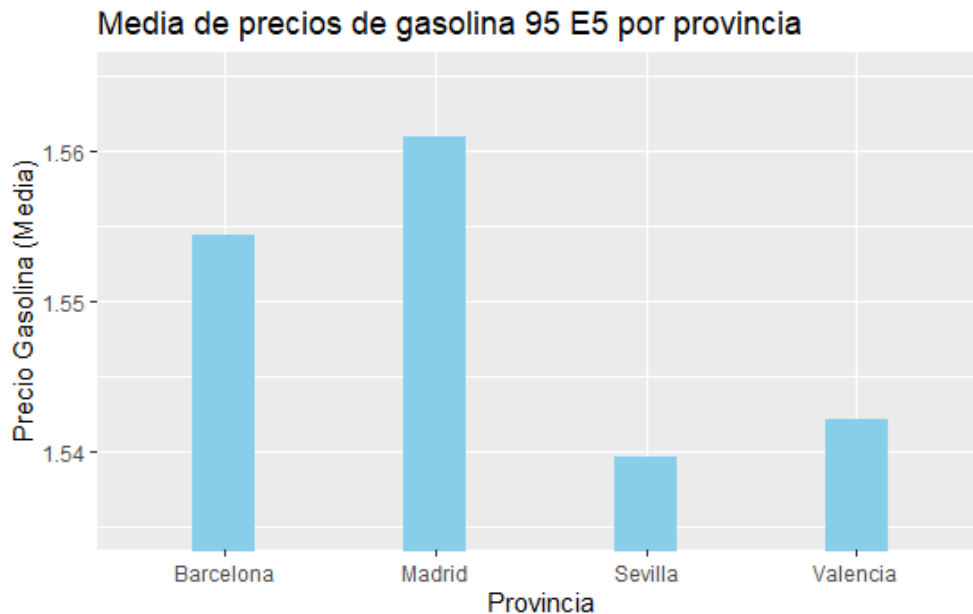


Figura 4. Diagrama de barras del precio medio de la gasolina por provincia.

Por último, representaremos gráficamente la diferencia entre los precios máximos y mínimos por provincia.

```
# Crear dataframe minimos y maximos por provincia
df_min_max <- data.frame(Provincia=c("Barcelona", "Madrid",
                                     "Sevilla", "valencia"),
                        Mínimo=c(r_b[1],r_m[1],r_s[1],r_v[1]),
                        Máximo=c(r_b[6],r_m[6],r_s[6],r_v[6]))

# Gráfica

p <- ggplot(df_min_max, aes(x = Provincia, y = Mínimo),color="blue") +
  geom_point(color="blue") +
  geom_text(aes(label = Mínimo), vjust = -0.5,color="blue") +
  coord_cartesian(ylim = c(1.315, 1.75)) +
  labs(x = "Provincia", y = "Precio Gasolina") +
  ggtitle("Mínimos y Máximos de Precio de Gasolina 95 E5 por Provincia")

# Agregar los máximos
p + geom_point(aes(y = Máximo), color = "red") +
  geom_text(aes(label = Máximo), vjust = -15, color = "red")
```

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

Obtenemos el siguiente gráfico:

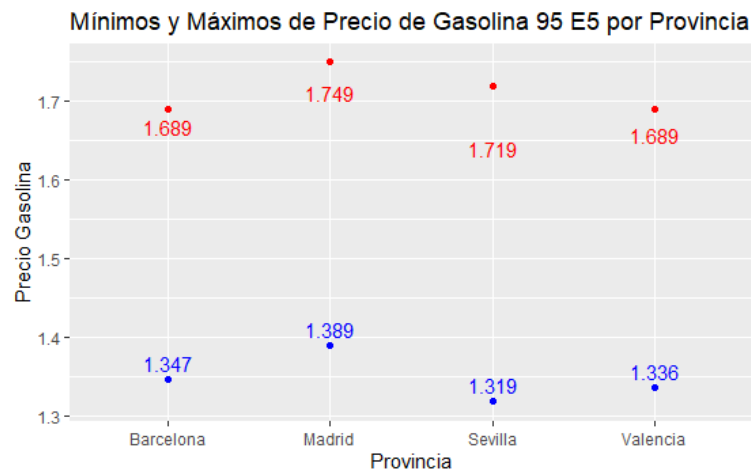


Figura 5. Gráfico de dispersión entre los precios máximos y mínimos de la gasolina por provincia.

3. A partir de los resúmenes y gráficos obtenidos en el punto 2, podemos extraer ciertas **conclusiones acerca del precio medio del carburante**.

Como vemos en Figura 4, el precio medio del carburante en Sevilla es el más bajo de todos (1,539€/Litro), muy lejano del precio medio más elevado: el de Madrid (1,56€/Litro), con una diferencia de 0,02128 €/Litro, calculado a partir de los datos en Figura 2. El precio medio en Sevilla va casi a la par con el de Valencia, mientras que en Barcelona es algo más elevado que en estas dos provincias, acercándose al precio medio en Madrid.

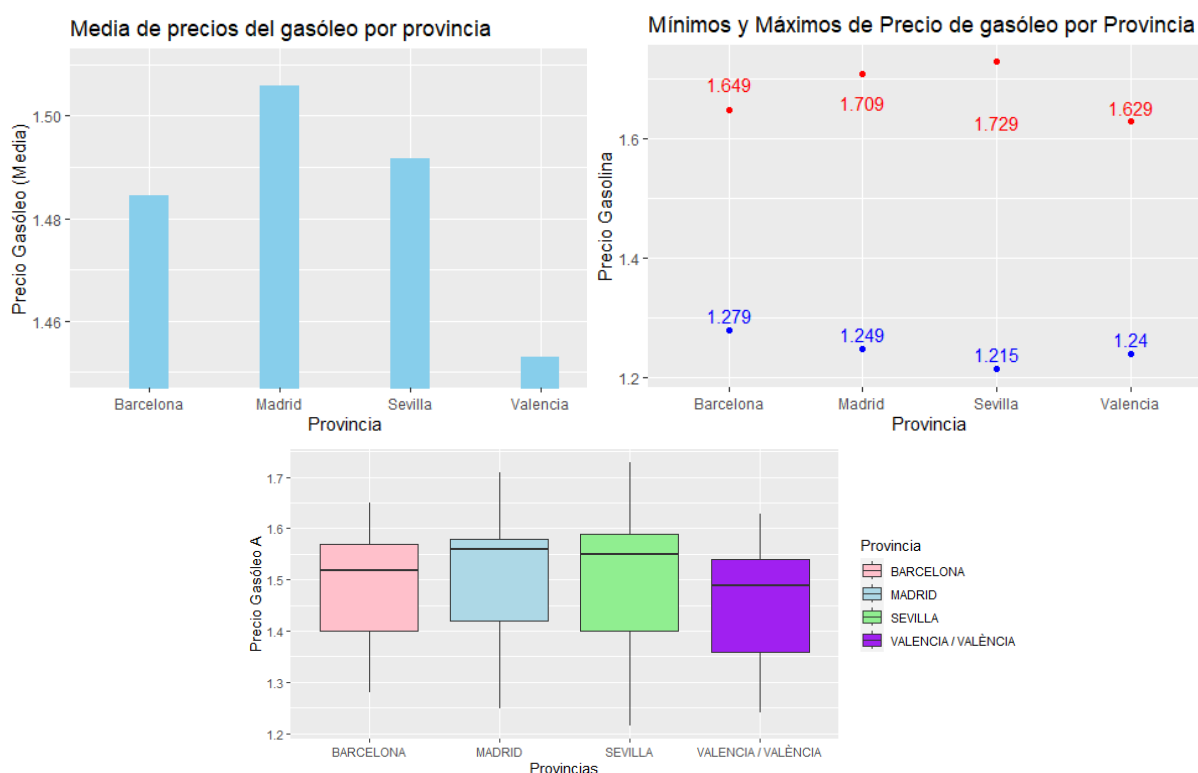
La variabilidad entre el precio mínimo y máximo por provincia es de unos 0,35€, excepto en Sevilla, que es de 0,4€. Esto se aprecia en Figura 5.

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

4.

Ciudad	Mínimo	Cuartil_1	Mediana	Media	Cuartil_3	Máximo
BARCELONA	1.279	1.399	1.519	1.484379	1.569	1.649
MADRID	1.249	1.419	1.559	1.505885	1.579	1.709
SEVILLA	1.215	1.399	1.549	1.491701	1.589	1.729
VALENCIA	1.240	1.359	1.489	1.453068	1.539	1.629

Figura 6. Resumen numérico del precio del gasóleo A por provincia.



Podemos ver a partir de estos resúmenes numéricos y gráficos que el precio medio del gasóleo más elevado es el de Madrid (1,505€/Litro), muy distante al de Valencia, con una diferencia de 5 céntimos el litro. Excepto Valencia, las demás provincias presentan una media de unos 1,5€/Litro, aunque sus precios mínimos rondan entre los 1,2-1,3€/Litro.

A pesar de no tener la media más alta, Sevilla presenta el precio máximo del gasóleo, además del mínimo. Es decir, su variabilidad es bastante más considerable que la de las otras provincias.

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

5. Exploratoriamente, podemos decir que el precio medio de la gasolina 95 en las cuatro provincias es más elevado que el precio del gasóleo A, con una diferencia de 7 céntimos el litro.

Esto también se refleja en los máximos y mínimos, pues hay una diferencia de 10 céntimos entre el precio mínimo de los carburantes, siendo el gasóleo el más barato. En el máximo esta diferencia no es tan notoria, ya que es de apenas 2 céntimos. Pero, en conclusión, el gasóleo es más barato.

Parte 2: Inferencia

Tarea 2

6. Estudiamos si hay diferencia de precios entre la provincia de Madrid y las demás.

Para ello, consideramos las hipótesis:

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

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

μ_1 : precio medio de gasolina 95 E5 en la provincia de Madrid.

μ_2 : precio medio de gasolina 95 E5 en las demás provincias.

Utilizando un nivel de significancia $\alpha = 0,01$, probamos si tenemos evidencia suficiente para afirmar que el precio es diferente en Madrid respecto de las demás provincias consideradas. Primero, comprobamos si tomamos las varianzas de población como iguales:

```
# Hacemos el var.test
var.test(df_m$`Precio gasolina 95 E5`,df_p_tp3$`Precio gasolina 95 E5`,
        conf.level = 0.99,
        alternative = 'two.sided')
```

Al tener p-valor < 0.01 , no tenemos suficiente evidencia como para suponer las varianzas iguales.

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

Teniendo en cuenta esto, ya podemos realizar el `t.test()`.

```
# a = 0.01 == conf.level 0.99
t.test(df_m$`Precio gasolina 95 E5`, df_p_tp3$`Precio gasolina 95 E5`,
       conf.level = 0.99,
       var.equal = FALSE,
       alternative = "two.sided")
```

Después de realizar el test bilateral, obtenemos un p-valor menor que 0.01, por lo tanto, con un nivel de significancia $\alpha = 0,01$, tenemos suficiente evidencia para rechazar la hipótesis nula de que la media de Madrid es igual a la de Barcelona, Sevilla y Valencia. El `t.test()` nos devuelve la media del precio de la gasolina 95 en Madrid (1,56), la cual es mayor que la media en las demás provincias (1,547).

7. A continuación, realizaremos el mismo contraste de hipótesis con los precios del gasóleo A. Procederemos, de igual manera, haciendo el test de varianza.

```
# u_1 - Precio medio Gasóleo A MADRID
u_1_1 = mean(df_m2$`Precio gasóleo A`)

# u_2 - Precio medio Gasóleo A BCN,SVQ,VAL
u_2_1 = mean(df_p_tp4$`Precio gasóleo A`)

# Hacemos el var.test
var.test(df_m2$`Precio gasóleo A`, df_p_tp4$`Precio gasóleo A`,
        conf.level = 0.99,
        alternative = 'two.sided')
```

De acuerdo con el p-valor obtenido (0.8401), tenemos evidencia suficiente para asumir las varianzas de las poblaciones como distintas. Ya podemos realizar el “t.test”.

```
# a = 0.01 == conf.level 0.99
t.test(df_m$`Precio gasolina 95 E5`, df_p_tp3$`Precio gasolina 95 E5`,
       conf.level = 0.99,
       var.equal = FALSE,
       alternative = "two.sided")
```

Esto nos devuelve un p-valor minúsculo, así que teniendo en cuenta las hipótesis:

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

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

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

μ_1 : precio medio de gasóleo en la provincia de Madrid.

μ_2 : precio medio de gasóleo en las demás provincias.

A partir del “t.test” realizado, podemos concluir que hay suficiente evidencia con $\alpha=0.01$ para afirmar que el precio medio del gasóleo de la provincia de Madrid es distinto al de las demás provincias, siendo mayor el de Madrid.

8. Queremos estudiar si el precio medio de la gasolina 95 E5 es significativamente mayor que el del gasóleo A. Para ello, formulamos las siguientes hipótesis:

$$H_0: \mu_1 = \mu_2$$

$$H_2: \mu_1 > \mu_2$$

μ_1 : precio medio de gasolina 95 E5 (en las 4 provincias).

μ_2 : precio medio de gasóleo A (en las 4 provincias).

```
# Hacemos el var.test
var.test(df_p_tp$`Precio gasolina 95 E5`, df_p_tp2$`Precio gasóleo A`,
         conf.level = 0.99,
         alternative = 'greater')
# p-valor = 1 --- suponemos varianzas iguales

# a = 0.01 == conf.level 0.99
t.test(df_p_tp$`Precio gasolina 95 E5`, df_p_tp2$`Precio gasóleo A`,
       conf.level = 0.99,
       var.equal = TRUE,
       alternative = "greater")
```

Tras realizar el test unilateral, obtenemos un p-valor $> 2.2e-16$, por lo que, con un nivel de significancia $\alpha=0.01$, tenemos suficiente evidencia para afirmar que el precio medio de la gasolina es significativamente mayor al del gasóleo en Barcelona, Madrid, Sevilla y Valencia.

Parte 3: Conclusiones

Asignatura	Datos del alumno	Fecha
Estadística	Apellidos: Witsman García	24/01/24
	Nombre: Carmen	

Tras realizar este análisis exploratorio de los precios de carburantes en las provincias de Barcelona, Madrid, Sevilla y Valencia, podemos resumir las siguientes conclusiones:

Diferencias en los precios: En términos generales, Madrid presenta los precios más elevados, mientras que Valencia cuenta con los precios más bajos. Esta discrepancia puede deberse a una combinación de factores, como la demanda local o los costos de distribución y transporte. A pesar de que Valencia presente precios más bajos en general, Sevilla alcanza los precios mínimos más bajos de entre todas las provincias.

Diferencia entre la gasolina y el gasóleo: Se ha observado que, en general, el precio promedio de la gasolina 95 E5 es superior al del gasóleo A en todas las provincias analizadas. Por lo tanto, a pesar de que el precio del gasóleo fluctúe más, es la opción más económica, sobre todo para los valencianos.

Variabilidad de precios: Se aprecia una variabilidad significativa en los precios de los combustibles, especialmente en el caso del gasóleo. Los precios oscilan entre 40 y 50 céntimos para el gasóleo, y entre 30 y 40 céntimos para la gasolina. Esta variabilidad puede influir en las decisiones de consumo tanto del público en general como de las empresas que dependen de estos combustibles.

En resumen, podemos concluir que **los precios de los combustibles varían según la provincia**, siendo Madrid la que presenta los precios más altos y Valencia los precios más bajos. **La gasolina tiende a ser más cara** que el gasóleo en todas las provincias, y la variabilidad de los precios puede ser relevante tanto para los consumidores como para las empresas.

Gracias a este análisis, podemos conocer un poco mejor cómo tienden a cambiar los precios de los carburantes en las diferentes provincias. Los gráficos, resúmenes numéricos y los contrastes de hipótesis nos han servido de apoyo para extraer dichas conclusiones y para reafirmar lo que podíamos suponer previamente al análisis.