Universidad Nacional de Colombia

Sede Medellín, Facultad de Ciencias

Estadística

Practica 1

Autores:
Juan Esteban Restrepo
Juan Manuel Saldarriaga
Josue Gonzales Otálvaro
David Esteban Cartagena Mejia

Docente:
Juan Carlos Salazar

2024-1S

Diseño de Experimentos

7 de abril de 2024

Tabla de contenidos

Listado de Figuras

Listado de Tablas

Punto 1

Se desea ver si la temperatura en la ciudad 1 es superior a la tempe- ratura en la ciudad 2, las temperaturas tomadas en las dos ciudades, en el verano, son las siguientes:

Tabla 1: Temperaturas registradas cuidad

Ciudad	1	2	3	4	5	6	7	8	9
Ciudad 1	83	89	89	90	91	91	92	94	96
Ciudad 2	77	78	79	80	81	81	81	82	

Use $\alpha = 0.05$

solucion

Sea

- x: Temperatura[1] de la cuidad 1
- y: Temperatura[2] de la cuidad 2

A continuacion se plantea el siguiente juego de hipotesis

$$\begin{cases} H_0 \, : \, E(x) \leq E(y) \\ H_1 \, : \, E(x) > E(y) \end{cases}$$

Equivalente a:

- H_o : la temperatura en la ciudad 1 es inferior a la temperatura en la ciudad 2
- H_1 : la temperatura en la ciudad 1 es superior a la temperatura en la ciudad 2

Supongamos que las mediciones de temperatura se realizaron de forma aleatoria, garantizando así la independencia en la selección de la muestra. Observemos que también se cumple la independencia entre las temperaturas registradas en ambas ciudades, ya que la medición en una ciudad no debería influir en la otra. Además, es importante destacar que la temperatura se mide en una escala de intervalo.Por tanto es valido aplicar el test Mann-Whitney

En el contexto de la prueba de Mann-Whitney, los puntajes se ordenan y clasifican siguiendo un método específico para su análisis comparativo entre dos grupos independientes.

[1] no se indico las unidades en que fueron reportadas las temperaturas
[2] no se indico las unidades en que fueron reportadas las temperaturas

Tabla 2: Rangos

Tr	tem	R
c2	77	1.0
c2	78	2.0
c2	79	3.0
c2	80	4.0
c2	81	6.0
c2	81	6.0
c2	81	6.0
c2	82	8.0
c1	83	9.0
c1	89	10.5
c1	89	10.5
c1	90	12.0
c1	91	13.5
c1	91	13.5
c1	92	15.0
c1	94	16.0
c1	96	17.0

Al asignar los rangos a cada una de las diferentes temperaturas en Tabla 2 se observo que existen 3 grupos de empates en medida de la temperatura.

$n = 9$ $m = 8$ $N = m + n = 17$

$T = \sum_{i=1}^{9} R_i(x) = 117$ $\sum_{i=1}^{17} R_i^2 = 1580$

Veamos que tenemos una muestra pequeña a demas existe una cantidad considerable de empates con respecto al total de datos de los que se dispone, por lo cual lo mas adecuado es el uso del factor de correccion al estadistico

$$T_1 = \frac{T - n\left(\frac{N+1}{2}\right) - 0.5}{\frac{nm}{N(N-1)} \sum_{i=1}^{17} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}}$$

Remplazando los valores obtenemos que

$$T_1 = \frac{117 - 9\left(\frac{17+1}{2}\right) - 0.5}{\sqrt{\frac{9*8}{17(17-1)} * 1580 - \frac{9*8(17+1)^2}{4(17-1)}}} = 4.6426$$

Ahora calculemos el valor p el cual correponde a

$$P(Z > 4.6426) = 1.7202601 \times 10^{-6}$$

```
    Wilcoxon rank sum test with continuity correction

data:  c1 and c2
W = 72, p-value = 0.0003033
alternative hypothesis: true location shift is greater than 0
```

Como el valor P< 0,05, Rechazamos H0 , y concluimos que los datos muestran que la temperatura de la cuidad 1 es mayor ala temperatura de la cuidad 2

Punto 3

H_0: No existe diferencia en los niveles de asertividad entre los diferentes órdenes de nacimiento.

H_1: Existe al menos una diferencia en los niveles de asertividad entre los diferentes órdenes de nacimiento.

Dado que los datos no se distribuyen de manera normal y solo podemos asumir una escala ordinal, se necesita una prueba no paramétrica. La prueba de Kruskal-Wallis es la más adecuada aquí, ya que es un equivalente no paramétrico de la ANOVA de un solo factor y puede utilizarse para comparar más de dos grupos independientes sin la necesidad de normalidad.

Estadístico de Prueba: Kruskal-Wallis

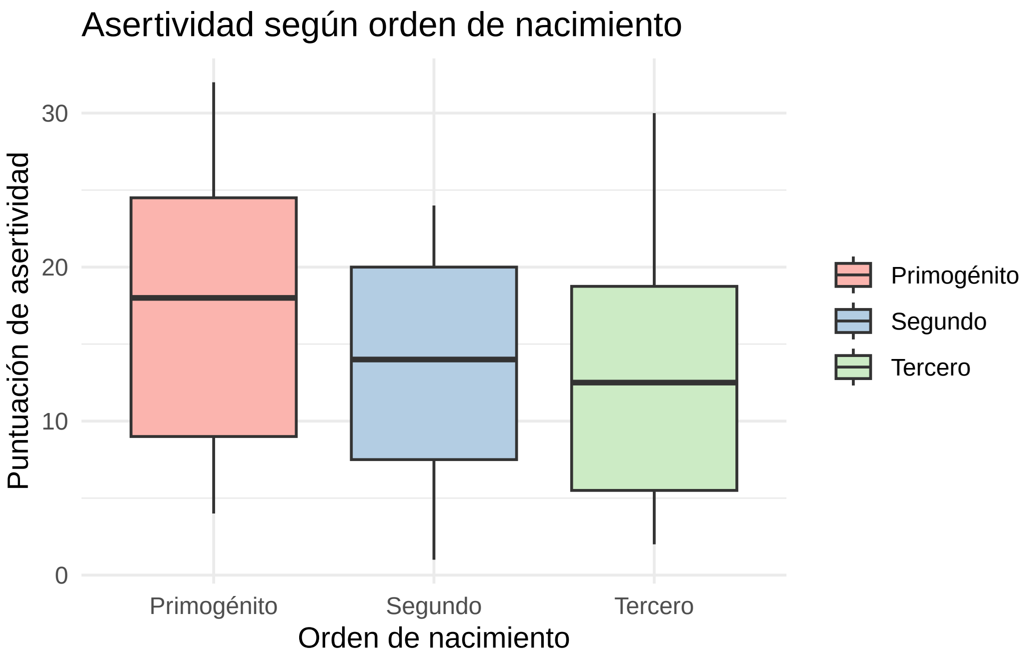
La prueba de Kruskal-Wallis evalúa si las medianas de los grupos son diferentes, basándose en los rangos de todas las observaciones. Es especialmente útil cuando las suposiciones de homogeneidad de varianzas o la normalidad de los grupos no se cumplen.

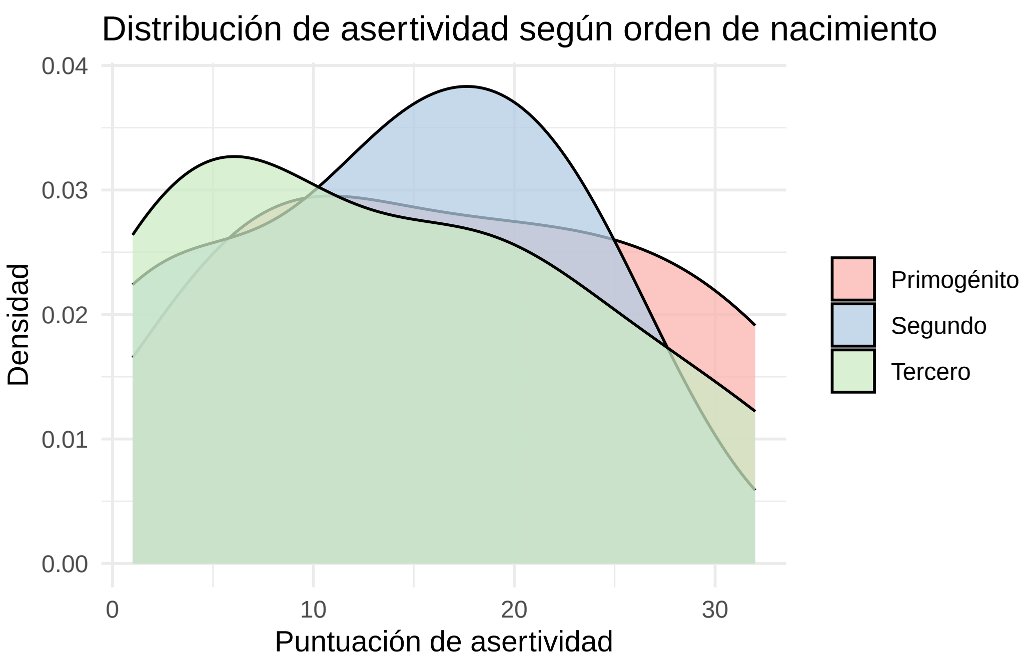
```
[1] 0.692517

    Kruskal-Wallis rank sum test

data:  asertividad by orden_nacimiento
Kruskal-Wallis chi-squared = 0.69461, df = 2, p-value = 0.7066
```

```
No hay evidencia suficiente para rechazar la hipótesis nula. No se observan diferencias significativas e
```





Punto 5

Cada vez que un carro tanqueaba, un dispositivo llevó el control de la cantidad de gasolina en galones puestos en el tanque, y la distancia en millas recorridas, los resultados fueron:

Millas	142	116	194	250	88	157	255	154	43	208
Galones	11.1	5.7	14.2	15.8	7.5	12.5	17.4	8.8	3.4	15.2

a) Suponga que la EPA estima que el millaje de este carro es 18 millas por galón. Pruebe la hipótesis de que esta cifra se aplica a este carro en particular.

Veamos que en este caso en particular, se determina por variable independiente la cantidad de gasolina, y por variable dependiente la cantidad de millas recorridas

A continuacion determinara si la tasa de consumo de gasolina de este auto es de 18 millas por galón,lo cual es equivalente a probar si la pendiente de la recta de regresión del Consumo de Gasolina respecto a las millas recorridas es dicha cantidad, para lo cual se plantea el siguiente juego de hipotesis:

$$\begin{cases} H_0 \ : \ \beta \geq 18 \\ H_1 \ : \ \beta < 18 \end{cases}$$

Para calcular dicho valor se procedera a calcular el coeficiente de Spearman entre la variable Galones y los residuos muestrales $u = Millas - 18x$

A continuacion se calculara el dichas diferencias

x_i	$R(x_i)$	u_i	$R(u_i)$
11.1	5	-57.8	5
5.7	2	13.4	10
14.2	7	-61.6	3
15.8	9	-34.4	7
7.5	3	-47.0	6
12.5	6	-68.0	1
17.4	10	-58.2	4
8.8	4	-4.4	9
3.4	1	-18.2	8
15.2	8	-65.6	2

Con dichos resultados a continuacion se mostraran resultados importantes para el el calculo del coeficiente de Spearman:

$\sum_{i=1}^{10} R(x_i)R(u_i) = 253$, $\sum_{i=10}^{10} R(u_i)^2 = 385$, $\sum_{i=10}^{10} R(x_i)^2 = 385$, $n(\frac{n+1}{2})^2 = 302.5$

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{n} R^2(x_i) - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^{n} R^2(y_i) - n\left(\frac{n+1}{2}\right)^2\right)}} = \frac{253 - 302.5}{\sqrt{(385 - 302.5)(385 - 302.5)}} = \frac{-49.5}{\sqrt{(82.5)(82.5)}} = -0.6$$

Luego con ρ ya calculado se procede a calcular, a evaluar el juego de hipotesis

Después de obtener el coeficiente de correlación ρ, se procede a realizar una evaluación del juego de hipótesis, mediante el criterio del valor p

$$valor - p = p(z < \rho\sqrt{9})$$

EL calculo del valor-p es 0.0359303, por tanto es valido afirmar que la tasa de consumo de Gasolina propuesto por la EPA no es consistente con este automovil.

b) Encuentre un intervalo de confianza del 95 % para la pendiente de la recta obtenida en el numeral anterior. Hagalo a mano.

A continuacion se procedera a encontrar cada pendiente

Pendientes del grupo 1

$$\frac{142-116}{11.1-5.7} = 4.815$$

$$\frac{142-194}{11.1-14.2} = 16.774$$

$$\frac{142-250}{11.1-15.8} = 22.979$$

$$\frac{142-88}{11.1-7.5} = 15$$

$$\frac{142-157}{11.1-12.5} = 10.714$$

$$\frac{142-255}{11.1-17.4} = 17.937$$

$$\frac{142-154}{11.1-8.8} = -5.2174$$

$$\frac{142-43}{11.1-3.4} = 12.857$$

$$\frac{142-208}{11.1-15.2} = 16.098$$

Pendientes del grupo 2

$$\frac{116-194}{5.7-14.2}=9.176$$

$$\frac{116-250}{5.7-15.8}=13.267$$

$$\frac{116-88}{5.7-7.5}=-15.556$$

$$\frac{116-157}{5.7-12.5}=6.029$$

$$\frac{116-255}{5.7-17.4}=11.88$$

$$\frac{116-154}{5.7-8.8}=12.258$$

$$\frac{116-43}{5.7-3.4}=31.7391$$

$$\frac{116-208}{5.7-15.2}=9.684$$

Pendientes del grupo 3

$$\frac{194-250}{14.2-15.8}=35$$

$$\frac{194-88}{14.2-7.5}=15.821$$

$$\frac{194-157}{14.2-12.5}=21.765$$

$$\frac{194-255}{14.2-17.4}=19.063$$

$$\frac{194-154}{14.2-8.8}=7.407$$

$$\frac{194-43}{14.2-3.4}=13.981$$

$$\frac{194-208}{14.2-15.2}=14$$

Pendientes del grupo 4

$$\frac{250-88}{15.8-7.5}=19.518$$

$$\frac{250-157}{15.8-12.5}=28.182$$

$$\frac{250-255}{15.8-17.4}=3.125$$

$$\frac{250-154}{15.8-8.8}=13.714$$

$$\frac{250-43}{15.8-3.4}=16.694$$

$$\frac{250-208}{15.8-15.2}=70$$

Pendientes del grupo 5

$$\frac{88-157}{7.5-12.5}=13.8$$

$$\frac{88-255}{7.5-17.4}=16.869$$

$$\frac{88-154}{7.5-8.8}=50.769$$

$$\frac{88-43}{7.5-3.4}=10.976$$

$$\frac{88-208}{7.5-15.2}=15.584$$

Pendientes del grupo 6

$$\frac{157-255}{12.5-17.4}=20$$

$$\frac{157-154}{12.5-8.8}=0.811$$

$$\frac{157-43}{12.5-3.4}=12.527$$

$$\frac{157-208}{12.5-15.2}=18.889$$

Pendientes del grupo 7

$$\frac{255 - 154}{17.4 - 8.8} = 11.744$$

$$\frac{255 - 43}{17.4 - 3.4} = 15.143$$

$$\frac{255 - 208}{17.4 - 15.2} = 21.364$$

Pendientes del grupo 8

$$\frac{154 - 43}{8.8 - 3.4} = 20.556$$

$$\frac{154 - 208}{8.8 - 15.2} = 8.438$$

Pendientes del grupo 9

$$\frac{43 - 208}{3.4 - 15.2} = 13.983$$

Pareja	P1	P2	P3	P4	P5	P6	P7	P8	P9
P2 = (116, 5.7)	4.8148								
P3 = (194, 14.2)	16.7742	9.1765							
P4 = (250, 15.8)	22.9787	13.2673	35.0000						
P5 = (88, 7.5)	15.0000	-15.5556	15.8209	19.5181					
P6 = (157, 12.5)	10.7143	6.0294	21.7647	28.1818	13.8000				
P7 = (255, 17.4)	17.9365	11.8803	19.0625	3.1250	16.8687	20.0000			
P8 = (154, 8.8)	-5.2174	12.2581	7.4074	13.7143	50.7692	0.8108	11.7442		
P9 = (43, 3.4)	12.8571	31.7391	13.9815	16.6935	10.9756	12.5275	15.1429	20.5556	
P10 = (208, 15.2)	16.0976	9.6842	14.0000	70.0000	15.5844	18.8889	21.3636	8.4375	13.9831

Veamos que segun tabla A.11 del libro de, cuando n=10, el valor de dicho de $W_{0.975} = 21$

N= 10C2 =45, luego

$r = \frac{1}{2}(N - W_{0.975}) = 0.5(45 - 21) = 12$

$s = N + 1 - r = 45 + 1 - 12 = 34$

Raro esto no esta dando – Kendall {SuppDists}

Un IC al 95% esta dado por: $[S^{(12)}, S^{(34)}] = [10.975, 19.062]$

```
 [1] -15.5555556  -5.2173913   0.8108108   3.1250000   4.8148148   6.0294118
 [7]   7.4074074   8.4375000   9.1764706   9.6842105  10.7142857  10.9756098
[13]  11.7441860  11.8803419  12.2580645  12.5274725  12.8571429  13.2673267
[19]  13.7142857  13.8000000  13.9814815  13.9830508  14.0000000  15.0000000
[25]  15.1428571  15.5844156  15.8208955  16.0975610  16.6935484  16.7741935
[31]  16.8686869  17.9365079  18.8888889  19.0625000  19.5180723  20.0000000
[37]  20.5555556  21.3636364  21.7647059  22.9787234  28.1818182  31.7391304
[43]  35.0000000  50.7692308  70.0000000
```

c) Ajuste una recta de regresión lineal usando el método de mínimos cuadrados.

Para aplicar la refresion lineal por minimos cuadrados, se hara uso del la funcion lm del paquete stats, del software estadistico R, para encontrar el ajuste de estos datos

```
Call:
lm(formula = df.5$Millas ~ df.5$Galones)

Coefficients:
 (Intercept)  df.5$Galones
       5.875        13.873
```

d) Obtenga una recta de regresión monótona como la vista en clase. hagalo a mano y luego usando software.

A continuacion se procede a calcular cada la regresion monotona

Tabla 6: Rangos de las variable

x	y	$R(X_i)$	$R(Y_i)$
11.1	142	5	4
5.7	116	2	3
14.2	194	7	7
15.8	250	9	9
7.5	88	3	2
12.5	157	6	6
17.4	255	10	10
8.8	154	4	5

x	y	$R(X_i)$	$R(Y_i)$
3.4	43	1	1
15.2	208	8	8

Veamos que en Tabla 6 se encuentran los rangos de la variable Galones de gasolina(x) y de Millas recorridas(y) con las cuales se procede a estimar el intercepto y la pendiente de la recta de regresion monotona

$b_2 = \frac{\sum_{i=1}^{10} R(X_i)R(Y_i) - n((n+1)^2/4)}{\sum_{i=1}^{10} R(X_i)^2 - n((n+1)^2/4)}$ $a_2 = \frac{(1-b_2)(n+1)}{2}$

Con lo cual se obtiene y=21.0015 -3.667*x

Se procede a encontrar la estimacion el rango estimado de y $(\widehat{R(y_i)}) = a_2 + b_2 X$

$\widehat{R(Y_1)}$ = 21.0015 -3.667* $R(X_1)$ = 2.666

$\widehat{R(Y_2)}$ = 21.0015 -3.667* $R(X_2)$ = 13.668

$\widehat{R(Y_3)}$ = 21.0015 -3.667*$R(X_3)$ = -4.667

$\widehat{R(Y_4)}$ = 21.0015 -3.667*$R(X_4)$ = -12.002

$\widehat{R(Y_5)}$ = 21.0015 -3.667*$R(X_5)$ = 10.001

$\widehat{R(Y_6)}$ = 21.0015 -3.667*$R(X_6)$ = -1

$\widehat{R(Y_7)}$ = 21.0015 -3.667*$R(X_7)$ = -15.669

$\widehat{R(Y_8)}$ = 21.0015 -3.667*$R(X_8)$ = 6.334

$\widehat{R(Y_9)}$ = 21.0015 -3.667*$R(X_9)$ = 17.334

$\widehat{R(Y_{10})}$ = 21.0015 -3.667*$R(X_{10})$ = -8.334

En el siguiente paso es calacular el rango estimado de y $(\widehat{R(x_i)})$:

$$\widehat{R(x_i)} = \frac{R(y_i) - a_2}{b_2}$$

$$\widehat{R(x_1)} = \frac{4 - 21.0015}{-3.667} = 4.636$$

$$\widehat{R(x_2)} = \frac{3 - 21.0015}{-3.667} = 4.909$$

$$\widehat{R(x_3)} = \frac{7 - 21.0015}{-3.667} = 3.818$$

$$\widehat{R(x_4)} = \frac{9 - 21.0015}{-3.667} = 3.273$$

$$\widehat{R(x_5)} = \frac{2 - 21.0015}{-3.667} = 5.182$$

$$\widehat{R(x_6)} = \frac{6 - 21.0015}{-3.667} = 4.091$$
$$\widehat{R(x_7)} = \frac{10 - 21.0015}{-3.667} = 3$$

$$\widehat{R(x_8)} = \frac{5 - 21.0015}{-3.667} = 4.364$$
$$\widehat{R(x_9)} = \frac{1 - 21.0015}{-3.667} = 5.454$$

$$\widehat{R(x_{10})} = \frac{8 - 21.0015}{-3.667} = 3.546$$

e) gráfique la funciones de los numerales anteriores en un mismo plano ¿Que puede concluir?

f) Estime el millaje, para Galones=16.

Apendice de codigo

```
library(tidyverse)
library(knitr)
library(magrittr)
c1 <- c(83,89,89,90,91,91,92,94,96)
c2 <- c(77,78,79,80,81,81,81,82)
df.T <- data.frame(Tr=c(rep("c1", times=length(c1)),
                          rep("c2", times=length(c2))),
                   tem=c(c1, c2))
#asignación de rangos de menor a mayor, con promedio de iguales
R <- rank(df.T$tem, ties.method = "average")
#tabla con resultados
ranked <- cbind(df.T,R)
k <- arrange(ranked,R)

wilcox.test(c1,c2,correct = T,alternative = 'g')
library(dplyr)

# Datos de asertividad
primogenitos <- c(18, 8, 4, 21, 28, 32, 10)
segundos <- c(18, 12, 3, 24, 22, 1, 14)
terceros <- c(7, 19, 2, 30, 18, 5)

# Crear un dataframe combinado con todos los datos
```

```r
datos <- data.frame(
  asertividad = c(primogenitos, segundos, terceros),
  orden_nacimiento = factor(c(rep("Primogénito", length(primogenitos)),
                              rep("Segundo", length(segundos)),
                              rep("Tercero", length(terceros))))
)
# Combinar todos los datos en un vector único
todos_los_datos <- c(primogenitos, segundos, terceros)

# Crear un vector de grupo correspondiente
grupo <- factor(c(rep("Primogénito", length(primogenitos)),
                  rep("Segundo", length(segundos)),
                  rep("Tercero", length(terceros))))

# Calcular los rangos de todas las observaciones
rangos <- rank(todos_los_datos)

# Calcular la suma de rangos para cada grupo
suma_rangos <- tapply(rangos, grupo, sum)

# Número total de observaciones
N <- length(todos_los_datos)

# Calcular el estadístico de Kruskal-Wallis
EP <- (12 / (N * (N + 1))) * sum(suma_rangos^2 / table(grupo)) - 3 * (N + 1)

# Mostrar el estadístico de prueba
EP

# Realizar la prueba de Kruskal-Wallis

kruskal_test <- kruskal.test(asertividad ~ orden_nacimiento, data = datos)


print(kruskal_test)

if(kruskal_test$p.value < 0.05) {
  cat("Hay evidencia suficiente para rechazar la hipótesis nula. Existen diferencias significativ
} else {
  cat("No hay evidencia suficiente para rechazar la hipótesis nula. No se observan diferencias si
}

library(ggplot2)
```

```r
ggplot(datos, aes(x = orden_nacimiento, y = asertividad, fill = orden_nacimiento)) +
  geom_boxplot() +
  labs(title = "Asertividad según orden de nacimiento",
       x = "Orden de nacimiento",
       y = "Puntuación de asertividad") +
  theme_minimal() +
  scale_fill_brewer(palette = "Pastel1") +
  theme(legend.title = element_blank())


ggplot(datos, aes(x = asertividad, fill = orden_nacimiento)) +
  geom_density(alpha = 0.75) + # Ajusta la transparencia con 'alpha'
  labs(title = "Distribución de asertividad según orden de nacimiento",
       x = "Puntuación de asertividad",
       y = "Densidad") +
  theme_minimal() +
  scale_fill_brewer(palette = "Pastel1") +
  theme(legend.title = element_blank())

# dataframe
df.5 <- data.frame(
  Millas = c(142, 116, 194, 250, 88, 157, 255, 154, 43, 208),
  Galones = c(11.1, 5.7, 14.2, 15.8, 7.5, 12.5, 17.4, 8.8, 3.4, 15.2)
)
n <- length(df.5$Galones)
u <- df.5$Millas-18*df.5$Galones
Ru <- rank(u)
Rx <- rank(df.5$Galones)
k <- cbind(df.5$Galones,Rx,u,Ru) %>% kable(format = 'markdown')
j <- 1 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 2 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
```

```
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 3 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 4 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 5 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 6 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
```

```
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 7 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 8 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 9 ## pareja de comparacion
h <- numeric(9)
g <- numeric(9)
y <- numeric(9)
x <- numeric(9)
for (i in j+1:10){h[i] <- paste(df.5$Galones[j],'-',df.5$Galones[i])
                g[i] <- paste(df.5$Millas[j],'-',df.5$Millas[i])
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}

j <- 1 ## pareja de comparacion
comparar <- function(j){
y <- numeric(9)
x <- numeric(9)
```

```r
for (i in seq(j+1,10,1)){
                y[i] <- df.5$Millas[j]-df.5$Millas[i]
                x[i] <- df.5$Galones[j]-df.5$Galones[i]
}
                return(y/x)
}
c1 <- comparar(1)
c2 <- comparar(2)
c3 <- comparar(3)
c4 <- comparar(4)
c5 <- comparar(5)
c6 <- comparar(6)
c7 <- comparar(7)
c8 <- comparar(8)
c9 <- comparar(9)
k <- c(c1,c2,c3,c4,c5,c6,c7,c8,c9)
k <- k[complete.cases(k)]
#cbind(c1,c2,c3,c4,c5,c6,c7,c8,c9) %>% kable(format = 'markdown')
sort(k)
lm(df.5$Millas~df.5$Galones)
nd <- data.frame(x=df.5$Galones,y=df.5$Millas,
                 Rx=rank(df.5$Galones),Ry=rank(df.5$Millas))
n <- length(df.5$Millas)
b2 <- function(Rx,Ry,n){
numerator <- sum(Rx*Ry) - n * ((n + 1)^2 / 4)
denominator <- sum(Rx^2) - n * ((n + 1)^2 / 4)

# Calculate b2
b2 <- numerator / denominator
return(b2)
}
a2 <- function(b2,n){
  numerator <- (1-b2)*(n-1)

a2 <- numerator /2
return(a2)
}
lmnp <- function(x,a2,b2){
  return(a2+b2*x)
}
lmnp <- Vectorize(lmnp)
b2 <- round(b2(nd$Rx,nd$Ryn,n),3)
a2 <- a2(b2,n)
```

Bibliografía