*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

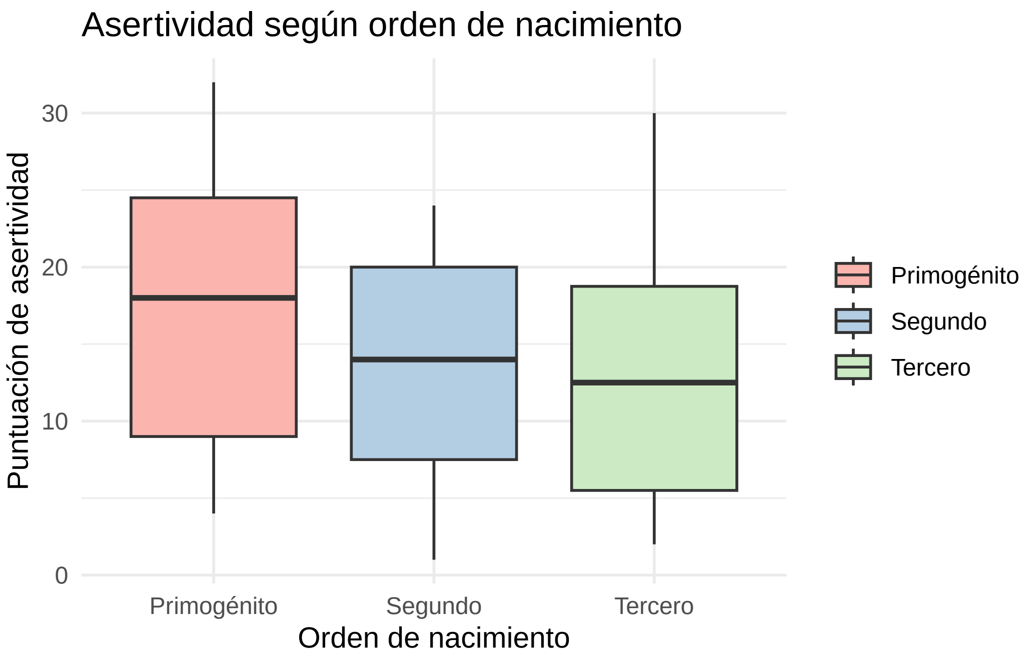
La prueba de Kruskal-Wallis evalúa si las medianas de los grupos son diferentes, basándose en los rangos de todas las observaciones. Es especialmente útil cuando las suposiciones de homogeneidad de varianzas o la normalidad de los grupos no se cumplen.

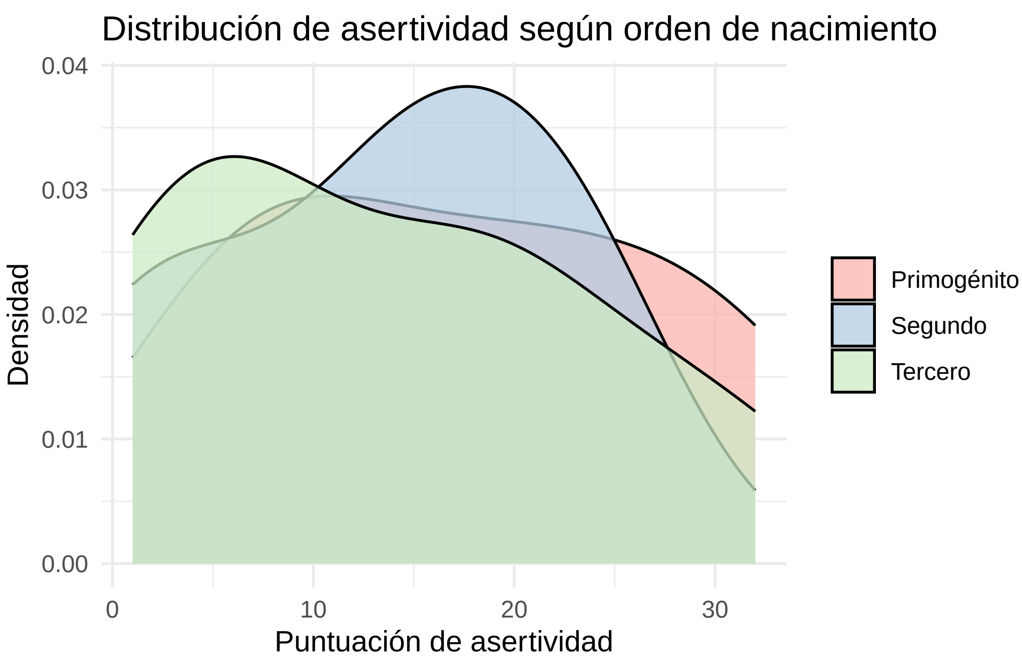
```
[1] 0.692517
```

```
    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 cuatro

Una pareja de esposos salieron a jugar bolos y guardaron sus resultados para ver si existia una relación entre dichos resultados Use $\rho$ de Pearson, el $\tau$ de kendall y el $\rho$ de spearman para realizar una prueba de independencia entre los puntajes.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Esposo | 147 | 158 | 131 | 142 | 183 | 151 | 196 | 129 | 155 | 158 |
| Esposa | 122 | 128 | 125 | 123 | 115 | 120 | 108 | 143 | 124 | 123 |

- **Coeficiente de correlación de Pearson ( $\rho$ de Pearson)**

El coeficiente de correlación de Pearson mide la relación lineal entre dos variables cuantitativas. Evalúa tanto la dirección como la magnitud de esta relación, asignando un valor entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, donde las variables se mueven juntas en la misma dirección. Un valor de -1 muestra una correlación negativa perfecta, significando que las variables se mueven en direcciones opuestas. Un valor de 0 sugiere que no existe una relación lineal entre las variables.

Se puede calcular de la siguiente forma:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]^{1/2} \left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]^{1/2}}}$$

Donde:

- $X$: Puntaje del esposo.
- $Y$: Puntaje de la esposa.
- $n$: Número total de pares.
- $\bar{X}$: Promedio de los puntajes del esposo.
- $\bar{Y}$: Promedio de los puntajes de la esposa.
- $\bar{X} \times \bar{Y}$: El producto de los promedios de los puntajes del esposo y de la esposa.

Debido a la cantidad de datos, procederemos a calcular cada componente y desglosar la ecuación por partes para una mejor comprensión.

$$n = 10$$

$$\bar{X} = \frac{147 + 158 + 131 + 142 + 183 + 151 + 196 + 129 + 155 + 158}{10} = 155$$

$$\bar{Y} = \frac{122 + 128 + 125 + 123 + 115 + 120 + 108 + 143 + 124 + 123}{10} = 123.1$$

$$\bar{X} \times \bar{Y} = 155 * 123.1 = 19080.5$$

$$\sum_{i=1}^{n} X_i Y_i - n\overline{XY} = [147 \cdot 122 - 10 \cdot 19080.5] + [158 \cdot 128 - 10 \cdot 19080.5] + [131 \cdot 125 - 10 \cdot 19080.5] + [142 \cdot 123 - 10 \cdot 19080.5] + [183 \cdot 115 - 10 \cdot 19$$

$$\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = [147^2 - 10 \cdot 155^2] + [158^2 - 10 \cdot 155^2] + [131^2 - 10 \cdot 155^2] +$$
$$[142^2 - 10 \cdot 155^2] + [183^2 - 10 \cdot 155^2] + [151^2 - 10 \cdot 155^2] +$$
$$[196^2 - 10 \cdot 155^2] + [129^2 - 10 \cdot 155^2] + [155^2 - 10 \cdot 155^2] +$$
$$[158^2 - 10 \cdot 155^2] = 63.11894$$

$$\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 = [122^2 - 10 \cdot 123.1^2] + [128^2 - 10 \cdot 123.1^2] + [125^2 - 10 \cdot 123.1^2] + [123^2 - 10 \cdot 123.1^2] + [115^2 - 10 \cdot 123.1^2] + [120^2 - 10 \cdot 123.1^2]$$

$$\frac{-1372}{[63.11894] \cdot [26.99815]} = -0.8051196$$

Implementación en R

```
[1] -0.8051197
```

El coeficiente de Pearson de -0.805 indica una fuerte correlación negativa entre los puntajes de bolos del esposo y de la esposa. En términos prácticos, esto significa que, en general, cuando el puntaje del esposo aumenta, el puntaje de la esposa tiende a disminuir, y viceversa. Sin embargo, es importante recordar que Pearson mide relaciones lineales, por lo que este coeficiente está particularmente enfocado en cómo una variable aumenta o disminuye de manera lineal en relación con la otra.

**Coeficiente ρ de Spearman**

La correlación de Spearman evalúa cómo se relacionan dos variables basándose en el orden de sus valores, no en su magnitud real. Asigna rangos y utiliza estas posiciones para calcular la asociación, identificando tanto relaciones lineales como no lineales. La muestra de tamaño $n$, cada observación se representa como un par $(X_i, Y_i)$. Se asigna un rango $R(X_i)$ a cada $X_i$, y $R(Y_i)$ a cada $Y_i$, en función de su posición relativa dentro de los conjuntos de valores de $X$ y $Y$ respectivamente. Para datos no numéricos, los rangos se otorgan basados en categorías de calidad. En caso de empates, se asigna el promedio de rangos correspondiente, un procedimiento alineado con el test de Mann-Whitney.

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{n} R^2(x_i) - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^{n} R^2(y_i) - n\left(\frac{n+1}{2}\right)^2\right)}}$$

Como se puede ver en la tabla de datos, los datos del esposo muestran resultados iguales, debido a que el número 158 aparece en dos ocasiones. Se creó una tabla para ordenar los rangos y manejar la duplicidad del número 158 en las posiciones 7 y 8.

| x | y | R(x) | R(y) | R(x)R(y) |
|---|---|---|---|---|
| 147 | 122 | 4.0 | 4.0 | 16.00 |
| 158 | 128 | 7.5 | 9.0 | 67.50 |
| 131 | 125 | 3.0 | 8.0 | 16.00 |
| 142 | 123 | 3.0 | 5.5 | 16.50 |
| 183 | 115 | 9.0 | 2.0 | 18.00 |
| 151 | 120 | 5.0 | 3.0 | 15.00 |
| 196 | 108 | 10.0 | 1.0 | 10.00 |
| 129 | 143 | 1.0 | 10.0 | 10.00 |
| 155 | 124 | 6.0 | 7.0 | 42.00 |
| 158 | 123 | 7.5 | 5.5 | 41.25 |

En situaciones donde el valor 158 se presenta múltiples veces, procedemos a reemplazarlo por el promedio de los rangos que debería ocupar, resultando en 7.5 a partir del cálculo $\frac{7+8}{2}$. Así, sustituimos 158 por 7.5 en los lugares que corresponden.

$$\sum_{i=0}^{n} R^2(X_i) = 4.0^2 + 7.5^2 + 2.0^2 + 3.0^2 + 9.0^2 + 5.0^2 + 10.0^2 + 1.0^2 + 6.0^2 + 7.5^2 = 384.5$$

$$\sum_{i=0}^{n} R^2(y_i) = 4.0^2 + 9.0^2 + 8.0^2 + 5.5^2 + 2.0^2 + 3.0^2 + 1.0^2 + 10.0^2 + 7.0^2 + 5.5^2 = 384.5$$

$$\sum_{i=0}^{n} R(x_i)R(y_i) = 16.00 + 67.50 + 16.00 + 16.50 + 18.00 + 15.00 + 10.00 + 10.00 + 42.00 + 41.25 = 252.25$$

$$\frac{252.25 - 10\left(\frac{10+1}{2}\right)^2}{\sqrt{(384.5 - 10\left(\frac{10+1}{2}\right)^2) \cdot (384.5 - 10\left(\frac{10+1}{2}\right)^2)}} = -0.6128049$$

Implementación en R

```
[1] -0.6128049
```

El coeficiente ρ de Spearman de -0.613 indica una correlación negativa moderada a fuerte, similar a la de Pearson pero basada en rangos en lugar de valores exactos. Al igual que Kendall, Spearman es más robusto ante datos no normales o la presencia de valores atípicos. Este coeficiente sugiere que, en términos de rangos, existe una tendencia negativa moderada a fuerte entre los puntajes de bolos de la pareja.

**Coeficiente τ de Kendall.**

El coeficiente $\tau$ de Kendall cuantifica la correlación entre dos series de datos examinando la concordancia y la discordancia en el ordenamiento de sus valores. Se obtiene mediante la diferencia entre los números de pares concordantes y discordantes, lo que refleja tanto la intensidad como la dirección de la relación entre las variables.

Cuando se presentan empates, el criterio de comparación es el siguiente: si $X_i = X_j$, la comparación entre estos pares se omite. En cambio, si $Y_i = Y_j$ pero al mismo tiempo $X_i \neq X_j$, entonces dicho par se considera como parcialmente concordante y parcia

$$\tau = \frac{N_c - N_d}{N_c + N_d'}$$

Donde $N_c$ representa el número de pares concordantes y $N_d$ el número de pares discordantes.

Se considera que hay concordancia si $\frac{Y_j - Y_i}{X_j - X_i} > 0$, y discordancia si $\frac{Y_j - Y_i}{X_j - X_i} < 0$. En el caso de que $\frac{Y_j - Y_i}{X_j - X_i} = 0$, el par se clasifica como mitad concordante y mitad discordante. Si $X_i = X_j$, entonces no se realiza ninguna comparación entre los pares. Finalmente, para facilitar el cálculo de $N_c$ (número de coincidencias) y $N_d$ (número de discrepancias), es beneficioso ordenar primero las observaciones $(X_i, Y_i)$ por los valores ascendentes de $X$ y, después, hacer lo mismo con los valores ascendentes de $Y$. Este procedimiento simplifica la comparación descendente de cada valor de $Y$. De acuerdo con esta metodología, definimos las reglas siguientes: se asigna un signo + indicando concordancia si $\frac{Y_j - Y_i}{X_j - X_i} > 0$. Si $\frac{Y_j - Y_i}{X_j - X_i} < 0$, entonces se asigna un signo − para señalar discordancia. Cuando $\frac{Y_j - Y_i}{X_j - X_i} = 0$, lo cual indica igualdad perfecta, se asigna un signo 0 que refleja una concordancia y discordancia a partes iguales, añadiendo 0.5 tanto a la cuenta de concordancia como a la de discordancia.

**Par (129,143)**

$$\frac{129 - 131}{143 - 125} = \frac{-}{+} = -$$
$$\frac{129 - 142}{143 - 123} = \frac{-}{+} = -$$
$$\frac{129 - 147}{143 - 122} = \frac{-}{+} = -$$
$$\frac{129 - 151}{143 - 120} = \frac{-}{+} = -$$
$$\frac{129 - 155}{143 - 124} = \frac{-}{+} = -$$

$$\frac{129-158}{143-123}=\frac{-}{+}=-$$
$$\frac{129-158}{143-128}=\frac{-}{+}=-$$
$$\frac{129-183}{143-115}=\frac{-}{+}=-$$
$$\frac{129-196}{143-108}=\frac{-}{+}=-$$

**Concordancias=0 Discordancias=9**

**Par (131,125)**

$$\frac{131-142}{125-125}=\frac{-}{+}=-$$
$$\frac{131-1123}{125-123}=\frac{-}{+}=-$$
$$\frac{131-147}{125-122}=\frac{-}{+}=-$$
$$\frac{131-151}{125-120}=\frac{-}{+}=-$$
$$\frac{131-155}{125-124}=\frac{-}{+}=-$$
$$\frac{131-158}{125-123}=\frac{-}{+}=-$$
$$\frac{131-158}{125-128}=\frac{-}{-}=+$$
$$\frac{131-183}{125-115}=\frac{-}{+}=-$$
$$\frac{131-196}{125-108}=\frac{-}{+}=-$$

**Concordancias=1 Discordancias=7**

**Par (142,123)**

$$\frac{142-147}{123-122}=\frac{-}{+}=-$$
$$\frac{142-151}{123-120}=\frac{-}{+}=-$$
$$\frac{142-155}{123-124}=\frac{-}{-}=+$$
$$\frac{142-158}{123-123}=\frac{-}{0}=0$$
$$\frac{142-158}{123-128}=\frac{-}{-}=+$$
$$\frac{142-183}{123-115}=\frac{-}{+}=-$$

$$\frac{142-196}{123-108}=\frac{-}{+}=-$$

**Concordancias=2.5 Discordancias=4.5**

**Par(147,122)**

$$\frac{147-151}{122-120}=\frac{-}{+}=-$$
$$\frac{147-155}{122-124}=\frac{-}{-}=+$$
$$\frac{147-158}{122-123}=\frac{-}{-}=+$$
$$\frac{147-158}{122-128}=\frac{-}{-}=+$$
$$\frac{147-183}{122-115}=\frac{-}{+}=-$$
$$\frac{147-196}{122-108}=\frac{-}{+}=-$$

**Concordancias=3 Discordancias=*3***

**Par(151,120)**

$$\frac{151-155}{120-124}=\frac{-}{-}=+$$
$$\frac{151-158}{120-123}=\frac{-}{-}=+$$
$$\frac{151-158}{120-128}=\frac{-}{-}=+$$
$$\frac{151-183}{120-115}=\frac{-}{+}=-$$
$$\frac{151-196}{120-108}=\frac{-}{+}=-$$

**Concordancias=3 Discordancias=2**

**Par(155,124)**

$$\frac{155-158}{124-123}=\frac{-}{+}=-$$
$$\frac{155-158}{124-128}=\frac{-}{-}=+$$
$$\frac{155-183}{124-115}=\frac{-}{+}=-$$
$$\frac{155-196}{124-108}=\frac{-}{+}=-$$

**Concordancias=1 Discordancias=3**

**Par(158,123)**

$$\frac{158-158}{123-128}=\frac{0}{-}=0$$
$$\frac{158-183}{123-115}=\frac{-}{+}=-$$
$$\frac{158-196}{123-108}=\frac{-}{+}=-$$

**Concordancias=0.5 Discordancias=2.5**

**Par(158,128)**

$$\frac{158-183}{128-115}=\frac{-}{+}=-$$
$$\frac{158-196}{128-108}=\frac{-}{+}=-$$

**Concordancias=0 Discordancias=2**

**Par(183,115)**

$$\frac{183-196}{115-108}=\frac{-}{+}=-$$

**Concordancias=0 Discordancias=1**

**Par(196,108)**

Como es el último por ende ambas son cero.

**Concordancias=0 Discordancias=0**

Tenemos que:

| $X_iY_i$ | Pares Concordantes | Pares Discordantes |
|---|---|---|
| (129, 143) | 0 | 9 |
| (131, 125) | 1 | 7 |
| (142, 123) | 2.5 | 4.5 |
| (147, 122) | 3 | 3 |
| (151, 120) | 3 | 2 |
| (155, 124) | 1 | 3 |
| (158, 123) | 0.5 | 2.5 |
| (158, 128) | 0 | 2 |
| (183, 115) | 0 | 1 |
| (196, 108) | 0 | 0 |

$$\tau=\frac{N_c-N_d}{N_c+N_d}=\frac{11-34}{11+34}=-0,5111111$$

Implementación en R

```
[1] -0.5227273
```

El cálculo manual del coeficiente de Tau de Kendall difiere ligeramente del obtenido con R, sin embargo, ambos indican una correlación negativa de intensidad moderada de -0.523 también sugiere una correlación negativa, aunque no tan fuerte como la indicada por Pearson. Kendall mide la concordancia en los ordenamientos de los datos entre dos variables, lo que significa que puntajes más altos de un miembro de la pareja tienden a asociarse con puntajes más bajos del otro miembro, pero con menos énfasis en la linealidad de esa relación. Este coeficiente es útil para entender las tendencias generales en los datos que pueden no ser estrictamente lineales.

**Conclusión**

Los tres coeficientes muestran una tendencia negativa en la relación entre los puntajes de bolos de la pareja, sugiriendo que no hay una dependencia directa en el sentido de que puntajes altos de uno impliquen puntajes altos del otro; de hecho, la tendencia es lo contrario. Esto podría interpretarse como que cuando uno de los dos tiene un buen día en el juego, el otro tiende a no tenerlo, aunque la interpretación exacta podría variar dependiendo de otros factores no considerados en este análisis.

**Prueba de Independencia**

Se llevaran a cabo pruebas de hipótesis utilizando las correlaciones de Pearson, Spearman y Kendall para determinar si existe independencia entre los puntajes de bolos de una pareja de esposos.

**Prueba de Independencia de Pearson**

```

	Pearson's product-moment correlation

data:  esposo and esposa
t = -3.8394, df = 8, p-value = 0.004951
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9521021 -0.3559160
sample estimates:
       cor 
-0.8051197 
```

$$
\begin{aligned}
H_0 &: \rho = 0 \quad \text{Los puntajes son independientes,} \\
H_1 &: \rho \neq 0 \quad \text{Los puntajes no son independientes,} \\
p\text{-valor} &= 0.004951 < 0.05 \qquad \text{Rechazamos } H_0,
\end{aligned}
$$

**Prueba de Independencia de Spearman**

```
    Spearman's rank correlation rho

data:  esposo and esposa
S = 266.11, p-value = 0.05961
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6128049
```

$$H_0 : \rho_s = 0 \quad \text{Los puntajes son independientes,}$$

$$H_1 : \rho_s \neq 0 \quad \text{Los puntajes no son independientes,}$$
$$p\text{-valor} = 0.05961 > 0.05 \qquad \text{No rechazamos } H_0,$$

**Prueba de Independencia de Kendall**

```
    Kendall's rank correlation tau

data:  esposo and esposa
z = -2.0737, p-value = 0.03811
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau
-0.5227273
```

$$H_0 : \tau = 0 \quad \text{Los puntajes son independientes,}$$
$$H_1 : \tau \neq 0 \quad \text{Los puntajes no son independientes,}$$
$$p\text{-valor} = 0.03811 < 0.05 \qquad \text{Rechazamos } H_0,$$

Al llevar a cabo los análisis de correlación mediante las metodologías de Pearson, Spearman y Kendall, se han obtenido resultados que nos permiten establecer distintas inferencias sobre la relación entre las puntuaciones de los individuos en estudio. El test de Pearson resultó en un p-valor de 0.004951, indicando con claridad la existencia de una correlación lineal negativa significativa y justificando el rechazo de la hipótesis nula de independencia. Por el contrario, el método de Spearman generó un p-valor de 0.05961, que supera el nivel de significación establecido de 0.05, lo cual significa que no hay evidencia suficiente para descartar la hipótesis nula y, por ende, no se puede afirmar la presencia de una relación monótona significativa. En el caso del test de Kendall, el p-valor obtenido fue de 0.03811, que está por debajo del umbral de significancia y sugiere una correlación negativa significativa en términos de rango. Estos hallazgos indican que, según los tests de Pearson y Kendall, las puntuaciones no son independientes y comparten una relación negativa. No obstante, la ausencia de evidencia significativa en la prueba de Spearman para un nivel de confianza del 95% nos insta a considerar la posibilidad de que la naturaleza de los datos o la sensibilidad inherente a cada test puedan influir en la interpretación de la relación entre las variables.

## 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 $\rho$ ya calculado se procede a calcular, a evaluar el juego de hipotesis

Después de obtener el coeficiente de correlación $\rho$, 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 9: 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 9 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

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))
```

```r
#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
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
```

```r
# 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)

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())

esposo <- c(147, 158, 131, 142, 183, 151, 196, 129, 155, 158)
esposa <- c(122, 128, 125, 123, 115, 120, 108, 143, 124, 123)
cor_pearson <- cor(esposo, esposa, method = "pearson")
cor_pearson
cor_spearman <- cor(esposo, esposa, method = "spearman")
cor_spearman
cor_kendall <- cor(esposo, esposa, method = "kendall")
cor_kendall
pearson <- cor.test(esposo, esposa, method="pearson")
print(pearson)
```

```r
spearman<- cor.test(esposo, esposa, method="spearman")
print(spearman)
kendall <- cor.test(esposo, esposa, method="kendall")
print(kendall)
# 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]
```

```r
                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)
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)
```

```r
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