UNIVERSIDAD NACIONAL DE COLOMBIA

*Universidad Nacional de Colombia*

*Sede Medellín, Facultad de Ciencias*

*Estadística*

# Practica 1

***Autores:***
Jessy Mark Mena Hernandez
David Esteban Cartagena Mejia

***Docente:***
Victor Ignacio Lopez Rios

**2024-1S**

*Diseño de Experimentos*

31 de marzo 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

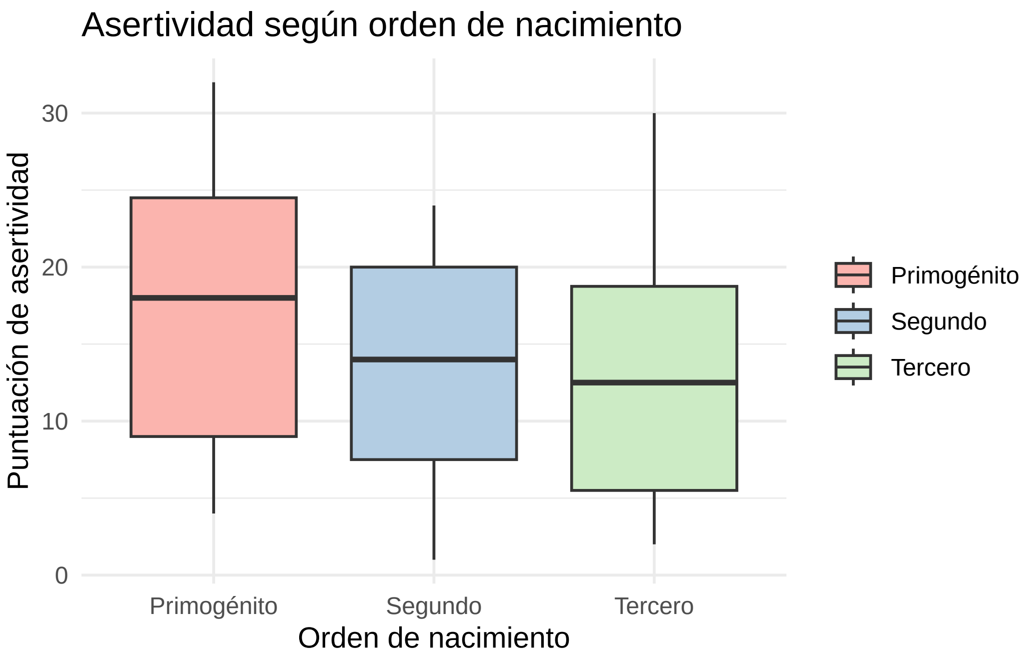
La prueba de Kruskal-Wallis evalúa si las medianas de los grupos son diferentes, basándose en los rangos de todas las observaciones. Es especialmente útil cuando las suposiciones de homogeneidad de varianzas o la normalidad de los grupos no se cumplen.

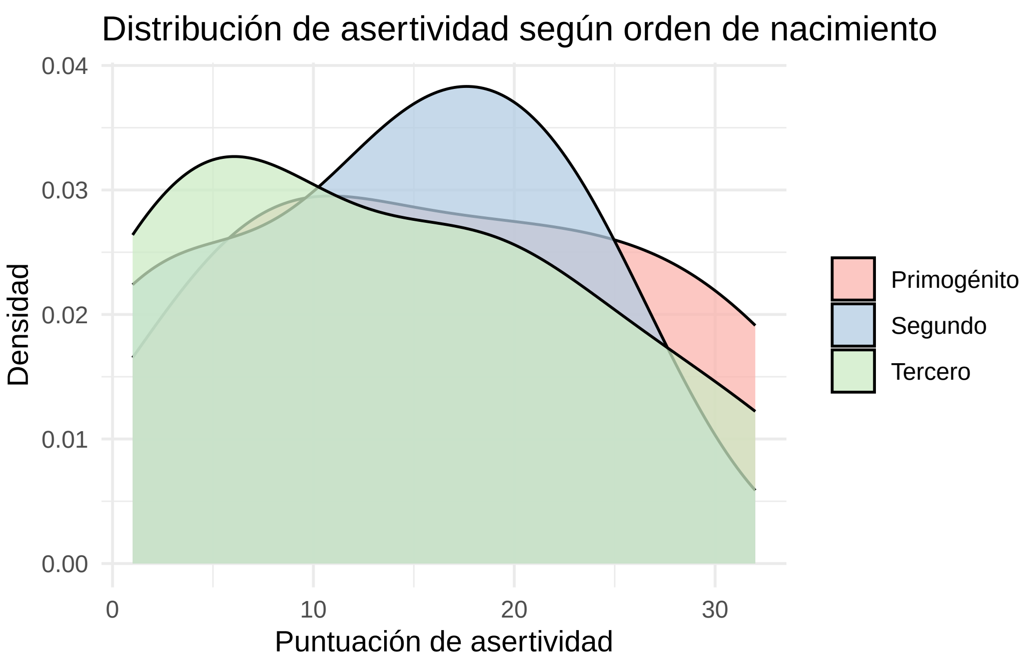
```
[1] 0.692517

    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 4

- Coeficiente de correlación de Pearson ( $\rho$ de Pearson): Mide la correlación lineal entre dos variables cuantitativas. Se calculará a partir de la covarianza de las variables dividida por el producto de sus desviaciones estándar.
- Coeficiente de correlación de Kendall ($\tau$ de Kendall): Mide la similitud de los órdenes de los datos cuando se clasifican por cada una de las cantidades. Se basa en la diferencia entre la probabilidad

de que dos observaciones disponibles se ordenen de la misma manera en ambas listas contra la probabilidad de que no lo sean.

- Coeficiente de correlación de Spearman ( $\rho$ de Spearman): Es un medida de correlación no paramétrica que evalúa la relación monótona entre dos variables cuantitativas. Se calcula en base al rango de los datos.

Procedemos a calcular cada uno de estos coeficientes para los puntajes dados.

Los coeficientes de correlación calculados entre los puntajes de bolos de la pareja de esposos son los siguientes:

```
[1] -0.8051197
```

Coeficiente de Correlación de Pearson: -0.805

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.

```
[1] -0.5227273
```

Coeficiente τ de Kendall: -0.523

El coeficiente τ de Kendall 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.

```
[1] -0.6128049
```

Coeficiente ρ de Spearman: -0.613

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.

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.

## Apendice de codigo

```
library(tidyverse)
library(knitr)
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
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)
```

```r
# 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)

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)
```

```r
esposa <- c(122, 128, 125, 123, 115, 120, 108, 143, 124, 123)

#  Pearson
cor_pearson <- cor(esposo, esposa, method = "pearson")
cor_pearson

# Kendall
cor_kendall <- cor(esposo, esposa, method = "kendall")
cor_kendall

#  Spearman
cor_spearman <- cor(esposo, esposa, method = "spearman")
cor_spearman
```

## Bibliografía