

# Práctica 9. Construcción de intervalos de confianza alrededor de una proporción.

Jesús Martín Fernández

## Contenidos

1. Introducción . . . . .	1
2. Métodos para la estimación de los IC de una proporción . . . . .	1
2.1 Método de Wald . . . . .	1
2.2 El método de Wilson . . . . .	5
2.3 El método exacto (aproximación de Clopper-Pearson) . . . . .	6
3. Representación gráfica de los intervalos de confianza (IC) de una proporción. . . . .	8

## 1. Introducción

Los intervalos de confianza (IC) alrededor de una proporción son herramientas estadísticas fundamentales que permiten estimar el rango en el cual se espera que se encuentre la proporción verdadera de una población con un determinado nivel de confianza. Al igual que con la media, los IC para proporciones proporcionan un sentido de la incertidumbre asociada a una estimación. Existen varios métodos para calcular los IC de una proporción

## 2. Métodos para la estimación de los IC de una proporción

### 2.1 Método de Wald

El método de Wald o asintótico es el más utilizado, y se basa en la suposición de que el estadístico de la proporción sigue una distribución normal, lo que permite utilizar la aproximación normal para construir el IC. Para ello utiliza la conocida fórmula:

$$IC = \hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

donde  $\hat{p}$  es la proporción muestral,  $z$  es el valor crítico de la distribución normal correspondiente al nivel de confianza deseado, y  $n$  es el tamaño de la muestra.

Vamos a utilizarlo con la base de datos `df_iam2`

Primero estableceremos el directorio de trabajo y extraeremos `df_iam2` del directorio Práctica 9

```
#setwd ("~/Práctica 9")
getwd ()
```

```
[1] "~/Práctica 9"
```

```
df_iam2 <- read.csv("df_iam2.csv")

head (df_iam2)
```

	edad	sex	alt	peso	imc	hta	fum	colesterol	clas_soc	iam	imc_r
1	40	Mujer	157.2	54.2	21.93281	Sí	Sí	192	Alta	Sí	Normal
2	40	Varón	172.5	70.5	23.69250	No	Sí	176	Alta	No	Normal
3	84	Mujer	167.4	63.4	22.62447	Sí	No	226	Baja	No	Normal
4	87	Varón	173.3	96.3	32.06485	Sí	No	188	Baja	No	Obesidad
5	83	Varón	172.7	62.7	21.02243	No	No	178	Alta	Sí	Normal
6	60	Mujer	154.3	51.3	21.54694	Sí	No	157	Alta	No	Normal

Vamos a calcular el IC del 95% de la proporción de varones y mujeres utilizando el método de Wald, en primer lugar, creando la fórmula del IC

```
# Definir el nivel de significancia
alpha <- 0.05

# Crear tabla de frecuencias para la variable 'fum'
tabla_sex <- table(df_iam2$sex)

# Número total de observaciones
n <- sum(tabla_sex)

# Proporciones
p_varones <- tabla_sex["Varón"] / n
p_mujeres <- tabla_sex["Mujer"] / n

# Cálculo del error estándar para "Varón"
```

```
SE_varones <- sqrt(p_varones * (1 - p_varones) / n)
z <- qnorm(1 - alpha / 2) # Valor crítico

# Cálculo del intervalo de confianza para "Varón"
IC_varones <- c(p_varones - z * SE_varones, p_varones + z * SE_varones)

# Cálculo del error estándar para "Mujer"
SE_mujeres <- sqrt(p_mujeres * (1 - p_mujeres) / n)

# Cálculo del intervalo de confianza para "Mujer"
IC_mujeres <- c(p_mujeres - z * SE_mujeres, p_mujeres + z * SE_mujeres)

print(IC_varones)
```

```
      Varón      Varón
0.4647638 0.5288259
```

```
print(IC_mujeres)
```

```
      Mujer      Mujer
0.4711741 0.5352362
```

```
# Resultados (solo para que sea más fácil de leer)
print(paste("Proporción de varones:", round(p_varones, 4)))
```

```
[1] "Proporción de varones: 0.4968"
```

```
print(paste("Intervalo de confianza (95%) para varones:",
            round(IC_varones[1], 4), "a", round(IC_varones[2], 4)))
```

```
[1] "Intervalo de confianza (95%) para varones: 0.4648 a 0.5288"
```

```
print(paste("Proporción de mujeres:", round(p_mujeres, 4)))
```

```
[1] "Proporción de mujeres: 0.5032"
```

```
print(paste("Intervalo de confianza (95%) para mujeres:",
           round(IC_mujeres[1], 4), "a", round(IC_mujeres[2], 4)))
```

```
[1] "Intervalo de confianza (95%) para mujeres: 0.4712 a 0.5352"
```

El método de Wald está implementado en el paquete Rbase a través de la función `prop.test`. Vamos a realizar el ejercicio anterior con esta función

```
# Extraer el número de "Varón" y "Mujer"
n_varones <- tabla_sex["Varón"]
n_mujeres <- tabla_sex["Mujer"]

# Calcular intervalo de confianza para varones
resultado_varones <- prop.test(n_varones, n, conf.level = 0.95)

# Calcular intervalo de confianza para mujeres
resultado_mujeres <- prop.test(n_mujeres, n, conf.level = 0.95)

# Imprimir los intervalos de confianza
print(resultado_varones$conf.int[1:2])
```

```
[1] 0.4643107 0.5293057
```

```
print(resultado_mujeres$conf.int[1:2])
```

```
[1] 0.4706943 0.5356893
```

Como ves permite cambiar el nivel de confianza para el IC. Repite el último ejercicio, pero estimando intervalos de confianza del 99%. Piensa antes de ejecutar el código si serán más amplios o más estrechos.

```
resultado_varones <- prop.test(n_varones, n, conf.level = 0.99)

# Calcular intervalo de confianza para mujeres
resultado_mujeres <- prop.test(n_mujeres, n, conf.level = 0.99)

# Imprimir los intervalos de confianza
print(resultado_varones$conf.int[1:2])
```

```
[1] 0.4643107 0.5293057
```

```
print(resultado_mujeres$conf.int[1:2])
```

```
[1] 0.4607043 0.5456602
```

En situaciones donde  $p$  es muy cercana a 0 o 1 o en muestras pequeñas, el método de Wald puede subestimar o sobreestimar el IC, por lo que se recomienda considerar otros enfoques, como los métodos de Agresti-Coull o Wilson, que ofrecen resultados más precisos bajo esas condiciones.

## 2.2 El método de Wilson

El método de Wilson para calcular intervalos de confianza para proporciones presenta varias ventajas significativas, lo que lo hace preferible en diversos contextos estadísticos. Una de sus principales fortalezas es su robustez en situaciones de proporciones extremas, ya que proporciona estimaciones más precisas cuando la proporción de éxitos es muy baja o alta, algo que el método de Wald no logra manejar bien. Además, Wilson ofrece una mejor cobertura, garantizando una mayor probabilidad de que el verdadero parámetro poblacional esté dentro del intervalo calculado. Este método también es útil para tamaños de muestra pequeños, ajustando la estimación y mejorando la precisión del intervalo. Su fórmula es algo más compleja

$$IC(1 - \alpha) = \frac{p + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

donde  $p$  es la proporción muestral,  $n$  el tamaño de la muestra  $z_{\alpha/2}$  el valor crítico de la distribución normal estándar para el nivel de confianza deseado.

Para usar este método utilizaremos la función `binconf` del paquete `Hmisc`

Vamos a estimar la proporción de hipertensos con esta función

```
# Comprobar si el paquete Hmisc está instalado
if (!requireNamespace("Hmisc", quietly = TRUE)) {
  install.packages("Hmisc") # Instalar Hmisc si no está instalado
}

# Cargar el paquete Hmisc
library(Hmisc)
```

Warning: package 'Hmisc' was built under R version 4.4.1

Adjuntando el paquete: 'Hmisc'

The following objects are masked from 'package:base':

format.pval, units

```
# Contar las observaciones de "Sí" y "No"
hta_counts <- table(df_iam2$hta)

# Extraer las proporciones
n_si <- hta_counts["Sí"] # número de "Sí"
n_no <- hta_counts["No"] # número de "No"
n_total <- sum(hta_counts) # total

# Calcular los intervalos de confianza para "Sí" usando el método de Wilson
ci_si <- binconf(n_si, n_total, method = "wilson")

# Calcular los intervalos de confianza para "No" usando el método de Wilson
ci_no <- binconf(n_no, n_total, method = "wilson")

# Mostrar resultados
ci_si
```

PointEst	Lower	Upper
0.4455128	0.4139589	0.4775121

```
ci_no
```

PointEst	Lower	Upper
0.5544872	0.5224879	0.5860411

### 2.3 El método exacto (aproximación de Clopper-Pearson)

El método exacto para calcular intervalos de confianza (IC) de proporciones proporciona estimaciones más precisas en comparación con los métodos aproximados, especialmente en situaciones donde el tamaño de la muestra es pequeño o cuando la proporción es extrema (es

decir, cercana a 0 o 1). Este método es especialmente útil en estudios de muestreo donde las proporciones de éxito o fracaso pueden no seguir la distribución normal.

El método exacto también se puede utilizar con la función `binconf`. Vamos a repetir el ejercicio anterior con el método exacto

```
# Calcular los intervalos de confianza para "Sí" usando el método exacto
ci_e_si <- binconf(n_si, n_total, method = "exact")

# Calcular los intervalos de confianza para "No" usando el método exacto
ci_e_no <- binconf(n_no, n_total, method = "exact")

# Mostrar resultados

ci_e_si
```

PointEst	Lower	Upper
0.4455128	0.4133521	0.4780174

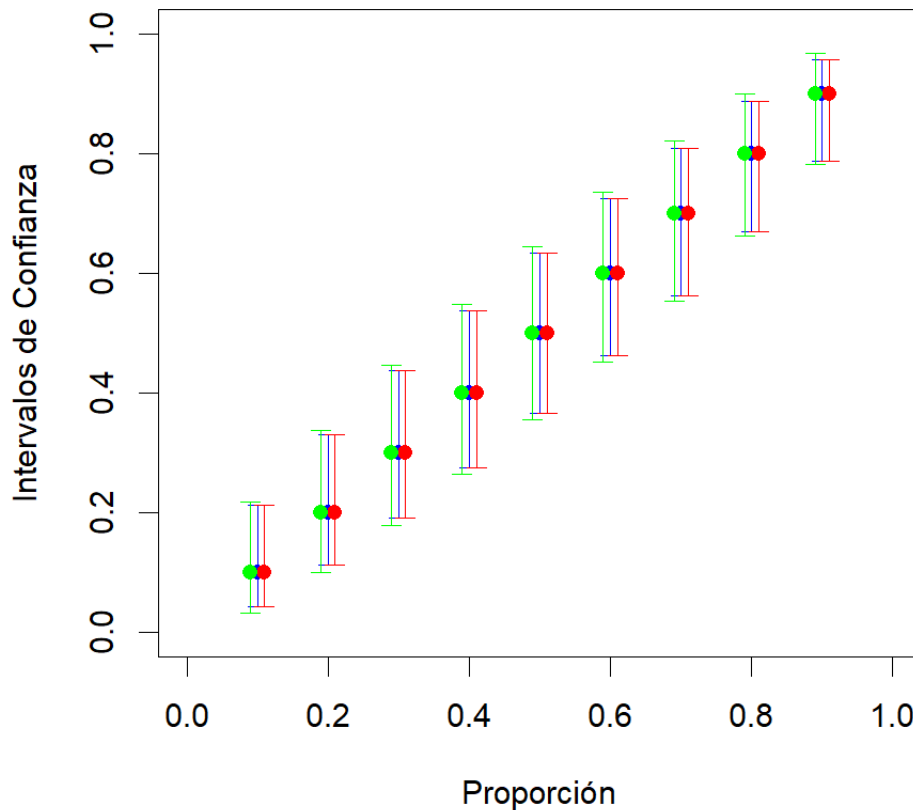
```
ci_e_no
```

PointEst	Lower	Upper
0.5544872	0.5219826	0.5866479

La función `binconf` también permite usar el método de Wald con la subintrucción “asymptotic”.

Como curiosidad, vamos a mostrar los IC para una muestra de 50 sujetos de diferentes proporciones mediante los tres métodos

## Comparación de IC para 50 Sujetos



Clopper-Person (verde) Wald (azul) Wilson (rojo)

### 3. Representación gráfica de los intervalos de confianza (IC) de una proporción.

Para representar IC de una proporción vamos a utilizar la función `plotCI` del paquete `Hmisc`, aunque su uso puede ser un poco más complejo debido a la forma en que se deben estructurar los datos. Esta función permite visualizar no solo proporciones, sino también cualquier tipo de estimación que incluya un margen de error, lo que la hace útil en una variedad de contextos estadísticos.

Vas a representar las proporciones de las categorías de la variable `fumy` sus IC del 95%. Analiza y utiliza el código siguiente,



```

tabla_fum <- table(df_iam2$fum)

# Número total de observaciones
total_fum <- sum(tabla_fum)

# Calcular el intervalo de confianza para cada categoría
resultado_si <- prop.test(tabla_fum["Sí"], total_fum, correct = FALSE)
resultado_no <- prop.test(tabla_fum["No"], total_fum, correct = FALSE)

# Graficar las proporciones con IC
# Ajustar los márgenes del gráfico
par(mar = c(5, 5, 2, 5))
# Aumentar el margen inferior (5) y lateral izquierdo (5)
plot(c(resultado_si$estimate, resultado_no$estimate),
      xaxt = "n",          # Evitar etiquetas automáticas en x
      ylim = c(0.2, 0.8),  # Ajustar el límite de y
      pch = 19,            # Tipo de punto
      col = "darkblue",    # Color de los puntos
      xlab = "Categorías de Fumador", # Etiqueta del eje x
      ylab = "Proporción",   # Etiqueta del eje y
      main = "Proporciones de Fumadores con IC", # Título
      cex = 1,             # Tamaño de los puntos
      xlim = c(0.5, 2.5)   # Ajustar límite de eje x
)

# Añadir los nombres de las categorías en el eje x
axis(1, at = 1:2, labels = c("Sí", "No"))

# Añadir los intervalos de confianza para "Sí"
arrows(
  x0 = 1,
  y0 = resultado_si$conf.int[1],
  y1 = resultado_si$conf.int[2],
  angle = 90,
  code = 3,
  length = 0.1,
  col = "blue"
)

# Añadir los intervalos de confianza para "No"
arrows(
  x0 = 2,

```

```
y0 = resultado_no$conf.int[1],  
y1 = resultado_no$conf.int[2],  
angle = 90,  
code = 3,  
length = 0.1,  
col = "blue"  
)
```

