

## Lectura 7

# Inferencia con proporciones muestrales

## Método de Wald

En el capítulo 3 vimos que, cuando queremos responder preguntas del tipo “¿qué proporción de la ciudadanía apoya al gobierno actual?”, estamos hablando de una variable aleatoria que sigue una distribución binomial. En general, no conocemos la **probabilidad de éxito**  $p$  de la población, por lo que tenemos que usar el estimador puntual (correspondiente a la proporción de éxito de la muestra), denotado por  $\hat{p}$ . Este estimador se distribuye de manera cercana a la normal cuando se cumplen las siguientes condiciones:

1. Las observaciones de la muestra son independientes.
2. Se cumple la **condición de éxito-fracaso**, que establece que se espera observar al menos 10 observaciones correspondientes a éxito y al menos 10, correspondientes a fracasos. Matemáticamente,  $np \geq 10$  y  $n(1 - p) \geq 10$ .

Así, si la distribución muestral de  $\hat{p}$  cumple con las condiciones anteriores, se dice que es cercana a la normalidad con media  $\mu = p$ , desviación estándar  $\sigma = \sqrt{p(1 - p)}$  y error estándar dado por la ecuación 7.1.

$$SE_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} \quad (7.1)$$

## Método de Wald para una proporción

-> Este método permite construir intervalos de confianza y contrastar hipótesis bajo el supuesto de normalidad para una proporción.

Script 7.1: método de Wald para una proporción.

```
1 # Fijar valores conocidos
2 n <- 150
3 p_exito <- 0.64
4 alfa <- 0.05
5 valor_nulo <- 0.7
6
7 # Construcción del intervalo de confianza.
8 error_est <- sqrt((p_exito * (1 - p_exito)) / n)
9 Z_critico <- qnorm(alfa / 2, lower.tail = FALSE)
10 inferior <- p_exito - Z_critico * error_est
11 superior <- p_exito + Z_critico * error_est
12 cat("Intervalo de confianza = [", inferior, ", ", superior, "]\n", sep = "")
13
14 # Prueba de hipótesis.
15 error_est_hip <- sqrt((valor_nulo * (1 - valor_nulo)) / n)
16 Z <- (p_exito - valor_nulo) / error_est_hip
17 p <- pnorm(Z, lower.tail = FALSE)
18 cat("Hipótesis alternativa unilateral\n")
19 cat("Z =", Z, "\n")
20 cat("p =", p)
```

## Método de Wald para dos proporciones

También podemos usar el método de Wald para estudiar la **diferencia entre las proporciones** de dos poblaciones, considerando para ello como estimador puntual la diferencia  $\hat{p}_1 - \hat{p}_2$ .

De manera similar a lo que ya vimos para una única proporción, también en este caso debemos verificar ciertas condiciones antes de poder aplicar el modelo normal:

1. Cada proporción, por separado, sigue el modelo normal.
2. Las dos muestras son independientes una de la otra.

El error estándar para la diferencia entre dos proporciones muestrales está dado por la ecuación 7.4, donde  $p_1$  y  $p_2$  corresponden a las proporciones de las poblaciones, y  $n_1$  y  $n_2$ , a los tamaños de las muestras. La construcción del intervalo de confianza se realiza, una vez más, con la ecuación general 4.6.

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (7.4)$$

Script 7.2: método de Wald para la diferencia entre dos proporciones (ejemplo 1).

```
1 # Fijar valores conocidos
2 n_hombres <- 48
3 n_mujeres <- 42
4 exitos_hombres <- 26
5 exitos_mujeres <- 20
6 alfa <- 0.05
7 valor_nulo <- 0
8
9 # Calcular probabilidades de éxito.
10 p_hombres <- exitos_hombres / n_hombres
11 p_mujeres <- exitos_mujeres / n_mujeres
12
13 # Estimar la diferencia.
14 diferencia <- p_hombres - p_mujeres
15
16 # Construcción del intervalo de confianza.
17 error_hombres <- (p_hombres * (1 - p_hombres)) / n_hombres
18 error_mujeres <- (p_mujeres * (1 - p_mujeres)) / n_mujeres
19 error_est <- sqrt(error_hombres + error_mujeres)
20 Z_critico <- qnorm(alfa / 2, lower.tail = FALSE)
21 inferior <- diferencia - Z_critico * error_est
22 superior <- diferencia + Z_critico * error_est
23 cat("Intervalo de confianza = [", inferior, ", ", superior, "]\n", sep = "")
24
25 # Prueba de hipótesis.
```

## Método de Wilson

Este método opera del mismo modo que el de Wald, aunque con las fórmulas empleadas para estimar la proporción en la muestra y el error estándar son diferentes

El script 7.4 muestra el uso de esta función con el mismo ejemplo que usamos para presentar la prueba de Wald para una proporción. Del mismo modo, el script 7.5 usa la función `prop.test()` para el primer ejemplo del método de Wald para la diferencia entre dos proporciones. Sin embargo, esta función tiene la limitante de que, al trabajar con dos proporciones, no permite establecer un valor nulo distinto de cero para la diferencia.

Script 7.4: método de Wilson para una proporción.

```
1 # Fijar valores conocidos
2 n <- 150
3 p_exito <- 0.64
4 alfa <- 0.05
5 valor_nulo <- 0.7
6
7 # Calcular cantidad de éxitos.
8 exitos <- p_exito * n
9
10 # Prueba de Wilson en R.
11 prueba <- prop.test(exitos, n = n, p = valor_nulo,
12                     alternative = "greater", conf.level = 1 - alfa)
13
14 print(prueba)
```

Script 7.5: método de Wilson para la diferencia entre dos proporciones.

```
1 # Fijar valores conocidos (hombres, mujeres)
2 n <- c(c(48, 42))
3 exitos <- c(26, 20)
4 alfa <- 0.05
5 valor_nulo <- 0.0
6
7 # Prueba de Wilson en R.
8 prueba <- prop.test(exitos, n = n, alternative = "two.sided",
9                     conf.level = 1 - alfa)
10
11 print(prueba)
```