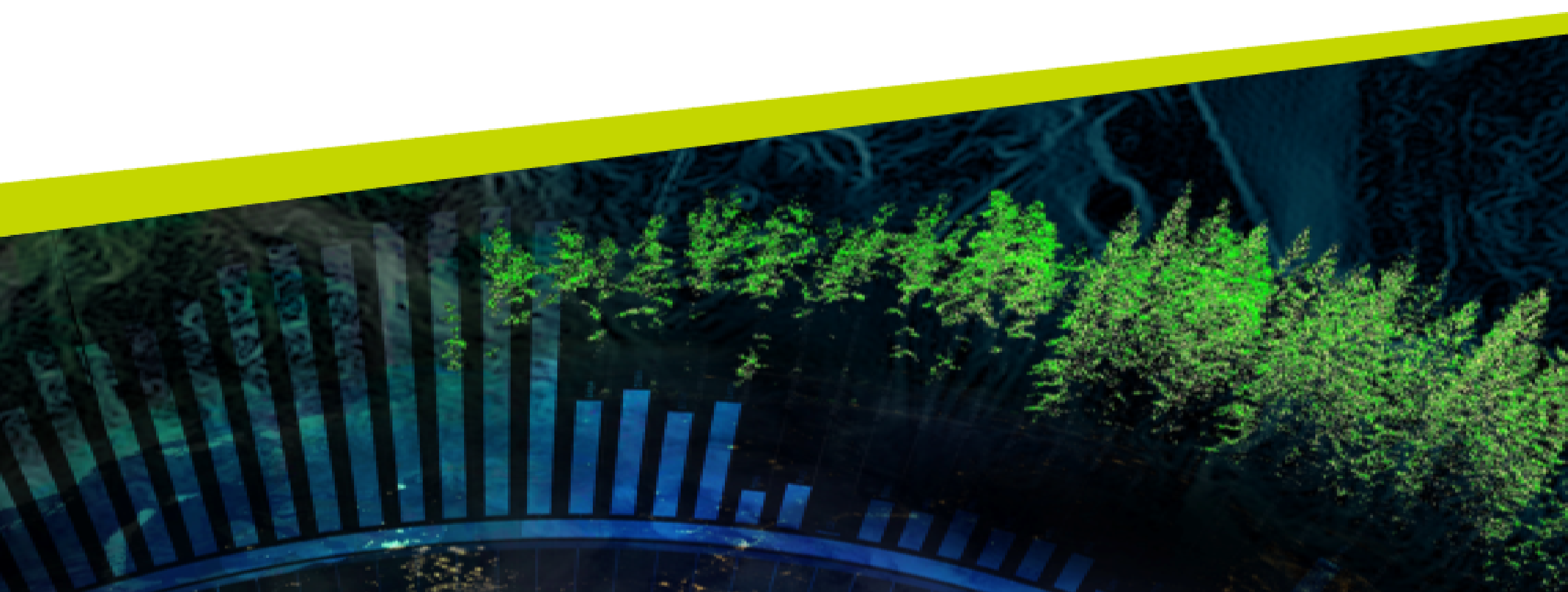
INFERENCIA Y MODELOS ESTADÍSTICOS

Jacqueline Köhler C. y José Luis Jara V.



CAPÍTULO 7. INFERENCIA CON PROPORCIONES MUESTRALES

En el capítulo5 conocimos las pruebas Z y t de Student para contrastar hipótesis con una y dos medias. Ahora estudiaremos los métodos de Wald y de Wilson para inferir acerca de una y dos proporciones, basándonos para ello en los textos de Diez y col. (2017, pp. 274-286), NIST/SEMATECH (2013, pp. 7.2.4, 7.2.4.1), Pértega y Pita (2004), Champely, Ekstrom, Dalgaard, Gill, Weibelzahl, Anandkumar, Ford, Volcic y de Rosario (2020) y Kabacoff (2017).

7.1 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.

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

7.1.1 Método de Wald para una proporción

El **método de Wald** permite construir intervalos de confianza y contrastar hipótesis bajo el supuesto de normalidad para una proporción. Consideremos el siguiente ejemplo: Aquiles Baeza, ingeniero en informática, desea conocer qué proporción de las ejecuciones de un algoritmo de ordenamiento para instancias con 100.000 elementos (bajo iguales condiciones de hardware y sistema) tardan menos de 25 segundos. Para ello, registró los tiempos de ejecución para 150 instancias generadas de manera aleatoria, encontrando que 64 % de dichas instancias fueron resueltas en un tiempo menor al señalado.

Si bien no conocemos la probabilidad real de éxito para la población, sabemos que $\hat{p} = 0,64$. Así, si se cumplen las condiciones para que la distribución de $\hat{p}$ sea cercana a la normal, podemos construir un intervalo de confianza para la verdadera proporción muestral.

En el enunciado del ejemplo nos indican que las instancias del problema fueron escogidas de manera aleatoria y sabemos que éstas representan menos del 10 % del total de instancias posibles, con lo que se verifica la independencia de las observaciones. Por otra parte, nos dicen que la proporción de éxito es $\hat{p} = 0,64$, por lo que esperamos encontrar $0,64 \cdot 150 = 96$ instancias que tardan menos de 25 segundos y $(1 - 0,64) \cdot 150 = 54$ fracasos (instancias que tardan 25 segundos o más), con lo que se cumple la condición de éxito-fracaso. En consecuencia, podemos asumir que la distribución muestral de $\hat{p}$ sigue aproximadamente a la normal.

Podemos estimar el error estándar usando la ecuación 7.1, reemplazando p por el estadístico $\hat{p}$:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0,64(1-0,64)}{150}} = 0,0392$$

Con ello, construimos el intervalo de confianza para un nivel de significación $\alpha = 0,05$ usando la ecuación general (4.6) con $\hat{p}$ como estimador puntual:

$$\hat{p} \pm z^* \cdot SE \rightarrow 0,64 \pm 1,96 \cdot 0,0392 \rightarrow [0,5632; 0.7168]$$

Este intervalo significa que tenemos 95 % de confianza que la proporción de instancias (de 100.000 elementos) del problema que el algoritmo ordena en menos de 25 segundos se encuentra entre 56,32 % y 71,6 %.

Desde luego, también podemos usar el modelo normal en el contexto de la prueba de hipótesis para una proporción. Para ello, se deben cumplir las condiciones de independencia y éxito-fracaso que ya verificamos para construir el intervalo de confianza, pero en este caso tenemos que verificar la segunda condición con el valor nulo, denotado p_0. Una vez verificadas ambas condiciones, el error estándar y el estadístico Z que permiten determinar el p-valor se calculan usando las ecuaciones 7.2 y 7.3, respectivamente.

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} \tag{7.2}$$

$$Z = \frac{\hat{p} - p_o}{SE} \tag{7.3}$$

Supongamos ahora, volviendo a nuestro ejemplo, que Baeza afirma que más del 70 % de las instancias de tamaño 100.000 se ejecutan en menos de 25 segundos. Sin embargo, su jefe no está seguro, por lo que decide comprobarlo mediante una prueba de hipótesis con un nivel de significación $\alpha = 0,05$ (recordemos que $n = 150$ y $\hat{p} = 0,64$):

H_0: el 70 % de las instancias se ejecutan en menos de 25 segundos.
H_A: más del 70 % de las instancias se ejecutan en menos de 25 segundos.

De acuerdo a las hipótesis formuladas por el jefe de Baeza, el valor nulo es $p_0 = 0,7$, con lo que estas pueden formularse matemáticamente como:

Denotando como p a la proporción de todas las instancias de tamaño 100.000 que se ejecutan en menos de 25 segundos y considerando el valor hipotético $p_0 = 0,7$ para este parámetro:
H_0: p $= p_0$
H_A: p $> p_0$

Ya antes habíamos comprobado que se verifica la independencia de las observaciones. Además, considerando que el valor nulo fuese verdadero esperaríamos encontrar $0,7 \cdot 150 = 105$ éxitos y $(1 - 0,7) \cdot 150 = 45$ fracasos, ambos valores mayores que 10, por lo que la condición de éxito-fracaso se verifica.

Con ello, podemos calcular el estadístico de prueba:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0,7(1-0,7)}{150}} = 0,0374$$

$$Z = \frac{\hat{p} - p_o}{SE} = \frac{0,64 - 0,7}{0,0374} = -1,6043$$

El valor p asociado, calculado en R mediante la llamada a la función `pnorm(-1.6042, lower.tail = FALSE)`, es $p = 0,9456$. En consecuencia, la evidencia no es suficiente para rechazar la hipótesis nula, por lo que se concluye, con 95 % de confianza, que no es cierto que el algoritmo se ejecute en menos de 25 segundos para más del 70 % de las instancias de tamaño 100.000.

R no ofrece esta prueba, como función. Sin embargo, podemos hacerla como muestra el script 7.1 para nuestro ejemplo.

Script 7.1: método de Wald para una proporción.

```
# Fijar valores conocidos
n <- 150
p_exito <- 0.64
alfa <- 0.05
valor_nulo <- 0.7

# Construcción del intervalo de confianza.
error_est <- sqrt((p_exito * (1 - p_exito)) / n)
Z_critico <- qnorm(alfa / 2, lower.tail = FALSE)
inferior <- p_exito - Z_critico * error_est
superior <- p_exito + Z_critico * error_est
cat("Intervalo de confianza = [", inferior, ", ", superior, "]\n", sep = "")

# Prueba de hipótesis.
error_est_hip <- sqrt((valor_nulo * (1 - valor_nulo)) / n)
Z <- (p_exito - valor_nulo) / error_est_hip
p <- pnorm(Z, lower.tail = FALSE)
cat("Hipótesis alternativa unilateral\n")
cat("Z =", Z, "\n")
cat("p =", p)
```

7.1.2 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.

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

A modo de ejemplo, supongamos que la Facultad de Ingeniería de una prestigiosa universidad desea determinar si la tasa de reprobación de estudiantes que rinden la asignatura de programación por primera vez es igual para hombres y mujeres. Para ello, se examina la situación final de los estudiantes que rindieron la asignatura durante el segundo semestre de 2017. Para una muestra de 48 hombres (de un total de 632), se encontró que 26 de ellos reprobaron la asignatura. De manera similar, para una muestra de 42 mujeres (de un total de 507), se encontraron 20 reprobaciones[1], con ambas muestras tomadas de manera aleatoria.

Como ya es habitual, comencemos por verificar las condiciones de normalidad para cada una de las muestras. En ambos casos, las observaciones son independientes entre sí, pues provienen de personas diferentes que representan a menos del 10 % de la población. Además, los datos entregados evidencian que en ambos casos se cumple la condición de éxito-fracaso. Adicionalmente, ambas muestras son independientes entre sí, pues ambas categorías se excluyen mutuamente. Con esto último se verifican entonces las condiciones de normalidad para la diferencia de proporciones.

Sean $\hat{p}_1$ y $\hat{p}_2$ las proporciones de éxito muestrales (considerando en este contexto la reprobación como éxito) para hombres y mujeres, respectivamente:

$$\hat{p}_1 = 26/48 = 0,5417$$

$$\hat{p}_2 = 20/42 = 0,4762$$

$$\hat{p}_1 - \hat{p}_2 = 0,5417 - 0,4762 = 0,0655$$

El error estándar puede estimarse como:

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0,5417(1-0,5417)}{48} + \frac{0,4762(1-0,4762)}{42}} = 0,1054$$

Suponiendo un nivel de significación $\alpha = 0,05$, el intervalo de confianza corresponde a:

$$\hat{p}_1 - \hat{p}_2 \pm z^* SE_{\hat{p}_1 - \hat{p}_2} \rightarrow 0,0655 \pm 1,96 \cdot 0,1054 \rightarrow [-0,1411; 0,2721]$$

En consecuencia, podemos afirmar con 95 % de confianza que la diferencia en la tasa de reprobación de la asignatura de programación para hombres y mujeres varía entre -14,11 % y 27,21 %.

Desde luego, también podemos realizar pruebas de hipótesis en este escenario. Para el ejemplo tenemos que:

H_0: no hay diferencia en la tasa de reprobación de hombres y mujeres.
H_A: las tasas de reprobación son diferentes para hombres y mujeres.

Matemáticamente:

Denotando como p_1 y p_2 a las proporciones de hombres y mujeres, respectivamente, que reprueban la asignatura de programación la primera vez que la cursan:
H_0: $p_1 - p_2 = 0$
H_A: $p_1 - p_2 \neq 0$

Ya verificamos las condiciones para operar bajo el supuesto de normalidad cuando construimos el intervalo de confianza. Sin embargo, **cuando la hipótesis nula supone que no hay diferencia entre las proporciones**, la verificación de la condición de éxito-fracaso y la estimación del error estándar se realizan usando para ello la **proporción agrupada**, dada por la ecuación 7.5, donde $\hat{p}_1 n_1$ y $\hat{p}_2 n_2$ representan la cantidad de éxitos en la primera y segunda muestra, respectivamente.

[1] Los datos aquí presentados son ficticios, creados únicamente con fines pedagógicos.

$$\hat{p} = \frac{\text{número de éxitos}}{\text{número de casos}} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} \tag{7.5}$$

Así, en este caso tenemos:

$$\hat{p} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} = \frac{0,5417 \cdot 48 + 0,4762 \cdot 42}{48 + 42} = 0,5111$$

En consecuencia, en el caso de los hombres esperamos encontrar $\hat{p}n_1 > 24$ éxitos (reprobaciones) y $(1-\hat{p})n_1 > 23$ fracasos. Del mismo modo, para las mujeres esperamos $\hat{p}n_2 > 21$ éxitos y $(1-\hat{p})n_2 > 20$ fracasos, con lo que se verifican las condiciones para emplear el modelo normal.

El error estándar se calcula, como ya mencionamos, usando la proporción agrupada:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\frac{0,5111 \cdot (1-0,5111)}{48} + \frac{0,5111 \cdot (1-0,5111)}{42}} = 0,1056$$

El estimador puntual para la diferencia es $\hat{p}_1 - \hat{p}_2 = 0,0655$, con lo cual el estadístico de prueba está dado por:

$$Z = \frac{\text{estimador puntual} - \text{valor nulo}}{SE} = \frac{0,0655 - 0}{0,1056} = 0,6203$$

En consecuencia, el valor p correspondiente es $p = 0,5351$. Puesto que el valor p es mayor que $\alpha = 0,05$, se falla en rechazar la hipótesis nula. Así, podemos decir con 95 % de confianza que no existe evidencia suficiente para concluir que hay diferencia en la tasa de reprobación de hombres y mujeres para el primer curso de programación.

El script 7.2 muestra el desarrollo de este ejemplo en R.

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

```
# Fijar valores conocidos
n_hombres <- 48
n_mujeres <- 42
exitos_hombres <- 26
exitos_mujeres <- 20
alfa <- 0.05
valor_nulo <- 0

# Calcular probabilidades de éxito.
p_hombres <- exitos_hombres / n_hombres
p_mujeres <- exitos_mujeres / n_mujeres

# Estimar la diferencia.
diferencia <- p_hombres - p_mujeres

# Construcción del intervalo de confianza.
error_hombres <- (p_hombres * (1 - p_hombres)) / n_hombres
error_mujeres <- (p_mujeres * (1 - p_mujeres)) / n_mujeres
error_est <- sqrt(error_hombres + error_mujeres)
Z_critico <- qnorm(alfa / 2, lower.tail = FALSE)
inferior <- diferencia - Z_critico * error_est
superior <- diferencia + Z_critico * error_est
cat("Intervalo de confianza = [", inferior, ", ", superior, "]\n", sep = "")

# Prueba de hipótesis.
```

```r
p_agrupada <- (exitos_hombres + exitos_mujeres) / (n_hombres + n_mujeres)
error_hombres <- (p_agrupada * (1 - p_agrupada)) / n_hombres
error_mujeres <- (p_agrupada * (1 - p_agrupada)) / n_mujeres
error_est_hip <- sqrt(error_hombres + error_mujeres)
Z <- (diferencia - valor_nulo) / error_est_hip
p <- 2 * pnorm(Z, lower.tail = FALSE)
cat("Hipótesis alternativa bilateral\n")
cat("Z =", Z, "\n")
cat("p =", p)
```

Cuando contrastamos hipótesis para la **diferencia entre dos proporciones con un valor nulo distinto de 0**, el procedimiento es ligeramente diferente. En este caso, la comprobación de la condición de éxito-fracaso se realiza de manera independiente para ambas muestras y el error estándar se calcula, como ya se estudió para los intervalos de confianza, mediante la ecuación 7.4.

Supongamos ahora que la Facultad de Ingeniería de la Universidad anterior ha decidido replicar el estudio realizado para el curso de programación, esta vez para una asignatura de física. No obstante, las autoridades están convencidas de que la tasa de reprobación es 10 % mayor para los hombres y que, incluso, la diferencia podría ser mayor. Desean comprobar con un nivel de confianza de 95 % y para ello, seleccionaron aleatoriamente a 89 de los 1.023 hombres y a 61 de las 620 mujeres de la cohorte correspondiente al primer semestre de 2019. En las muestras se encuentran, respectivamente, 45 y 21 reprobaciones.

Las hipótesis son, en este caso:

H_0: la tasa de reprobación de los hombres es exactamente 10 % más alta que la de las mujeres.
H_A: la tasa de reprobación de los hombres es más de 10 % más alta que la de las mujeres.

Matemáticamente:

Denotando como p_1 y p_2 a las proporciones de hombres y mujeres, respectivamente, que reprueban la asignatura de física estudiada la primera vez que la cursan:
H_0: $p_1 - p_2 = 0,1$
H_A: $p_1 - p_2 > 0,1$

Al igual que en los ejemplos previos, las observaciones de cada muestra son independientes entre sí pues corresponden a menos del 10 % de la población y fueron escogidos aleatoriamente. A su vez, los datos proporcionados indican que se cumple la condición de éxito-fracaso para cada muestra. Como ambas muestras pertenecen a grupos diferentes de estudiantes, son independientes entre sí. En consecuencia, se cumplen las condiciones para operar bajo el modelo normal.

En el caso de los hombres, la tasa de éxito se estima como:

$$\hat{p}_1 = \frac{45}{89} = 0,5056$$

Análogamente, para las mujeres tenemos:

$$\hat{p}_2 = \frac{21}{61} = 0,3443$$

Con lo que el estimador puntual para la diferencia es:

$$\hat{p}_1 - \hat{p}_2 = 0,5056 - 0,3443 = 0,1613$$

Ahora calculamos el error estándar:

$$SE = \sqrt{\frac{0,5056(1-0,0,5056)}{89} + \frac{0,3443(1-0,3443)}{61}} = 0,0807$$

Con lo cual podemos calcular el estadístico de prueba:

$$Z = \frac{\text{estimador puntual} - \text{valor nulo}}{SE} = \frac{0,1613 - 0,1}{0,0807} = 0,7596$$

Con lo que se puede obtener el valor p, correspondiente a $p = 0.2237 > \alpha = 0,05$.

En consecuencia, se falla en rechaza H_0 en favor de H_A, por lo que concluimos, con 95 % de confianza, que la tasa de reprobación de los hombres es 10 % superior a la de las mujeres para el curso de física.

En R, esta prueba puede realizarse como muestra el script 7.3.

Script 7.3: método de Wald para la diferencia entre dos proporciones (ejemplo 2).

```
# Fijar valores conocidos
n_hombres <- 89
n_mujeres <- 61
exitos_hombres <- 45
exitos_mujeres <- 21
alfa <- 0.05
valor_nulo <- 0.1

# Calcular probabilidades de éxito.
p_hombres <- exitos_hombres / n_hombres
p_mujeres <- exitos_mujeres / n_mujeres

# Estimar la diferencia.
diferencia <- p_hombres - p_mujeres

# Prueba de hipótesis.
p_agrupada <- (exitos_hombres + exitos_mujeres) / (n_hombres + n_mujeres)
error_hombres <- (p_hombres * (1 - p_hombres)) / n_hombres
error_mujeres <- (p_mujeres * (1 - p_mujeres)) / n_mujeres
error_est <- sqrt(error_hombres + error_mujeres)
Z <- (diferencia - valor_nulo) / error_est
p <- pnorm(Z, lower.tail = FALSE)
cat("Hipótesis alternativa bilateral\n")
cat("Z =", Z, "\n")
cat("p =", p)
```

7.2 MÉTODO DE WILSON

El método de Wald, tratado en la sección anterior, es el método que tradicionalmente se ha usado y el que aparece en la mayoría de los libros clásicos de inferencia estadística. Sin embargo, el método está siendo muy criticado hoy en día debido a que hace importantes simplificaciones matemáticas en su procedimiento y ya hay evidencia empírica que ha demostrado sus limitaciones (Agresti & Coull, 1998).

Gracias al aumento del poder de cómputo y la disponibilidad de software estadístico, han surgido diversas alternativas, entre las cuales destaca el **método de Wilson** (junto con algunas variaciones), considerado el más robusto por diversos autores (Agresti & Coull, 1998; Brown y col., 2001; Devore, 2008; Wallis, 2013). Este método opera del mismo modo que el de Wald, aunque las fórmulas empleadas para estimar la proporción en la muestra y el error estándar son diferentes.

En R, podemos hacer esta prueba usando la función `prop.test(x, n, p, alternative, conf.level, ...)`, cuyos principales parámetros son:

- `x`: cantidad de éxitos en la muestra.
- `n`: tamaño de la muestra.
- `p`: valor nulo (por defecto, `p=0`).
- `alternative`: tipo de hipótesis alternativa, por defecto bilateral (`alternative=‘‘two.sided’’`), y valores `‘‘less’’` y `‘‘greater’’` para hipótesis unilaterales.
- `conf.level`: nivel de confianza (`conf.level=0.95` por defecto).

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.

```
# Fijar valores conocidos
n <- 150
p_exito <- 0.64
alfa <- 0.05
valor_nulo <- 0.7

# Calcular cantidad de éxitos.
exitos <- p_exito * n

# Prueba de Wilson en R.
prueba <- prop.test(exitos, n = n, p = valor_nulo,
                    alternative = "greater", conf.level = 1 - alfa)

print(prueba)
```

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

```
# Fijar valores conocidos (hombres, mujeres)
n <- c(c(48, 42))
exitos <- c(26, 20)
alfa <- 0.05
valor_nulo <- 0.0

# Prueba de Wilson en R.
prueba <- prop.test(exitos, n = n, alternative = "two.sided",
                    conf.level = 1 - alfa)

print(prueba)
```

7.3 PODER Y PRUEBAS DE PROPORCIONES

En el capítulo anterior conocimos el poder estadístico y vimos que está relacionado con el nivel de significación, el tamaño de la muestra y el tamaño del efecto que queremos detectar.

R base nos ofrece la función `power.prop.test(n, p1, p2, sig.level, power, alternative)`, donde:

- `n`: número de observaciones por cada grupo.
- `p1`: probabilidad de éxito en un grupo.

- `p2`: probabilidad de éxito en otro grupo.
- `sig.level`: nivel de significación.
- `power`: poder de la prueba.
- `alternative`: tipo de hipótesis alternativa (“`one.sided`” si es unilateral, “`two.sided`” si es bilateral).

Al igual que vimos en el capítulo anterior para la función `power.t.test()`, recibe cuatro de los primeros argumentos y al restante debe asigrnársele el valor `NULL`. Como resultado, retorna un objeto que incluye el valor calculado para el argumento faltante.

Una vez más, el paquete `pwr` de R nos ofrece varias funciones que podemos usar como alternativa:

- `pwr.p.test(h, n, sig.level, power, alternative)`: para pruebas con una única proporción.
- `pwr.2p.test(h, n, sig.level, power, alternative)`: para pruebas con dos proporciones donde ambas muestras son de igual tamaño.
- `pwr.2p2n.test(h, n1, n2, sig.level, power, alternative)`: para pruebas con dos proporciones y muestras de diferente tamaño.

Donde:

- `h`: tamaño de efecto.
- `n`, `n1`, `n2`: tamaño(s) de la(s) muestra(s).
- `sig.level`: nivel de significación.
- `power`: poder.
- `alternative`: tipo de hipótesis alternativa (“`two.sided`”, “`less`” o “`greater`”).

El funcionamiento de esta familia de funciones es igual al que ya conocimos en el capítulo anterior para la función `pwr.t.test()`. Se entrega el parámetro `alternative` y todos los demás excepto uno (al cual debe asignarse explícitamente el valor `NULL`). Como resultado, la función calcula dicho valor.

El tamaño del efecto puede calcularse como muestra la ecuación 7.6, implementada en R en la función `ES.h(p1, p2)` del paquete `pwr`.

$$h = 2\arcsin(\sqrt{p_1}) - 2\arcsin(\sqrt{p_2}) \tag{7.6}$$

En el caso de una única proporción, los autores del paquete `pwr` sugieren usar $p_2 = 0,5$ (Champely y col., 2020).

Otra función que nos puede ser de ayuda es `bsamsize(p1, p2, fraction, alpha, power)`, del paquete `Hmisc`. En el caso de una prueba de Wilson con dos muestras, calcula los tamaños de cada grupo dados los siguientes argumentos:

- `p1`: probabilidad de la población para el grupo 1.
- `p2`: probabilidad del grupo 2.
- `fraction`: fracción de las observaciones en el grupo 1 ($n1/(n1+n2)$).
- `alpha`: nivel de significación.
- `power`: poder deseado.

7.4 EJERCICIOS PROPUESTOS

1. ¿En qué condiciones la distribución muestral de una proporción tiene comportamiento aproximadamente normal?
2. ¿Cómo se calcula la desviación estándar de la distribución muestral de las proporciones bajo estas condiciones (según el método de Wald)?
3. ¿Cómo se calcula un intervalo de confianza para la verdadera proporción (según el método de Wald)?

4. El patrón de un gran fundo de nogales está preocupado porque se ha detectado la presencia de una plaga en varios árboles. Si bien existe un pesticida para el parásito, este es bastante caro y su aplicación solo se justifica económicamente si más del 20 % de los árboles está infectado. En consecuencia, el patrón ha decidido estimar la extensión de la infestación revisando una muestra aleatoria de 200 nogales (una porción bastante pequeña de los más de 20.000 árboles en el fundo). En base a lo anterior, determina:
   a) ¿Cuál es la variable dicotómica (experimento Bernulli) en este caso?
   b) ¿Cuál es el parámetro de interés?
   c) ¿Qué estimador existe para este parámetro?
   d) ¿Qué hipótesis respondería las dudas del patrón del fundo?
5. En el experimento del ejercicio anterior se encontró que 45 árboles de la muestra estaban infectados:
   a) ¿Se puede asumir que esta proporción muestral sigue el modelo normal?
   b) Independientemente de la respuesta anterior, obtén un intervalo con 95 % confianza para la verdadera proporción de árboles infectados en el fundo.
   c) ¿Qué recomendarías al patrón del fundo?
6. Como el patrón sigue con dudas, ahora pregunta: ¿cuántos árboles debería revisar en una muestra para estar 99 % confiado que más del 20 % de los árboles están infectados, con solo 10 % de probabilidades de equivocarse si la verdadera proporción fuera 18 %? ¿Cómo se puede calcular esto? ¿Cuál debiera ser la respuesta a la pregunta del patrón?
7. ¿En qué condiciones la distribución muestral de la diferencia de dos proporciones tiene comportamiento aproximadamente normal?
8. ¿Cómo se calcula el error estándar de la diferencia entre dos proporciones (según el método de Wald)?
9. ¿Cómo se calcula un intervalo de confianza para la verdadera diferencia entre dos proporciones (según el método de Wald)?
10. Un laboratorio homeopático acaba de lanzar un tónico que asegura que ayuda a prevenir el resfrío durante el periodo invernal, con igual eficacia tanto en mujeres como en hombres. Para comprobar esta promesa, el laboratorio está realizando un estudio de la eficacia del producto en una muestra aleatoria de 100 mujeres y 200 hombres:
    a) ¿Cuál es el parámetro de interés y qué estimador se podría usar?
    b) ¿Qué hipótesis se deberían docimar para comprobar o refutar la homogeneidad de la eficacia del tónico para el resfrío?
11. El estudio anterior encontró que, durante las semanas de prueba, 38 mujeres y 102 hombres presentaron síntomas de resfrío. ¿Es homogénea la eficacia del producto con un nivel de significación de 0,05?
12. ¿Qué poder tuvo la prueba anterior?
13. ¿Qué tamaño deberían tener las muestras aleatorias de mujeres y hombres (manteniendo la proporción del ejemplo) para conseguir un poder de 0,85 con 99 % de confianza?
14. Las fórmulas presentadas en la sección 7.1, se conocen colectivamente como “Método de Wald”, el que ya no es recomendado por académicos del área. Usando la bibliografía citada, ¿cuáles son las fórmulas del método de Wilson para estimar el error estándar de la proporción y su extensión a la prueba de hipótesis e intervalos de confianza?
15. Investiga para qué sirve y cómo funciona el parámetro `correct` (verdadero por defecto) de la función `prop.test()` de R.

REFERENCIAS

Agresti, A. & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119-126.

Brown, L. D., Cai, T. T. & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical science*, *16*(2), 101-117.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R. & de Rosario, H. (2020). *pwr: Basic Functions for Power Analysis*. Consultado el 1 de octubre de 2021, desde https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html

Devore, J. L. (2008). *Probabilidad y Estadística para Ingeniería y Ciencias* (7.ª ed.). CENAGE Learning.

Diez, D., Barr, C. D. & Çetinkaya-Rundel, M. (2017). *OpenIntro Statistics* (3.ª ed.). https://www.openintro.org/book/os/.

Kabacoff, R. I. (2017). *Power Analysis*. Consultado el 1 de octubre de 2021, desde https://www.statmethods.net/stats/power.html

NIST/SEMATECH. (2013). *e-Handbook of Statistical Methods*. Consultado el 29 de abril de 2021, desde http://www.itl.nist.gov/div898/handbook/

Pértega, S. & Pita, S. (2004). *Asociación de variables cualitativas: El test exacto de Fisher y el test de Mcnemar*. Consultado el 29 de abril de 2021, desde https://www.fisterra.com/mbe/investiga/fisher/fisher.asp#Mcnemar

Wallis, S. (2013). Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, *20*(3), 178-208.