¿Cero casos, cero prevalencia?

FELIPE ANDRÉS MEDINA MARÍN*

PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA
FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

Imaginemos que una campaña de testeo de una enfermedad convoca a un total de 200 personas, y que del total de 200 tests realizados ninguno resultó ser positivo. Está claro que estas 200 personas son sólo una muestra de una población, por lo que cualquier estadística calculada a partir de su información será una estimación de un parámetro poblacional. Entre los parámetros que podrían ser de interés estimar está la prevalencia de la enfermedad, θ. ¿Cuál sería la estimación en este caso? ¿Cero? En esta nota revisaremos qué hacer para obtener una estimación sensata cuando de un total de n observaciones (tests, ensayos) se observan 0 eventos de interés (resultados positivos, éxitos). Primero supondremos que el instrumento de medición es perfecto, es decir, un test con sensibilidad y especificidad iguales a 1, para posteriormente revisar un caso más general en el que el instrumento no es perfecto, es decir, existen falsos positivos y falsos negativos.

Antes de comenzar nuestra revisión debemos hacer un gran supuesto: la muestra a analizar es una muestra aleatoria y representativa de la población sobre la cual deseamos hacer inferencia. Esto es cuestionable en el caso presentado, ya que quienes se someten al test seguramente tienen un mayor interés en participar de tal actividad que quienes no lo hacen (sesgo de voluntario) y es muy improbable que las personas que ya saben que están enfermas se sometan al test. Esto resulta en que la población a la cual se tiene acceso no sea la misma sobre la cual se desea inferir y por tanto puedan diferir en cuanto a sus parámetros. Es decir, lo que estimaremos es la prevalencia entre quienes no han sido previamente diagnosticados con la enfermedad.

Cuando se trabaja con un test perfecto

Si el test es perfecto, tendremos que cada resultado positivo coincide con que la persona tiene la enfermedad, mientras que un resultado negativo coincidirá con que la persona está sana; i.e., un resultado positivo sólo puede ser un verdadero positivo y un resultado negativo sólo puede ser un verdadero negativo. En ese caso, el primer estimador de la prevalencia θ que se nos viene a la mente es la proporción muestral de tests positivos $P = \frac{X}{n}$. Sin embargo, rápidamente intuimos que algo no anda bien, ya que en nuestro caso de cero tests positivos este estimador resultaría en la estimación $p = \frac{0}{200} = 0$, i.e., estimaríamos que la prevalencia es igual a cero.

El problema está en que una prevalencia muy pequeña, pero mayor que cero, podría también dar origen a la nuestra observada y con una probabilidad muy alta. Por ejemplo, la probabilidad de observar cero resultados positivos, $\mathbb{P}(X = 0)$, en una muestra de 200 personas cuando la prevalencia es de $0,00001$ (o 1 en cada $100{,}000$ personas), es aproximadamente $0,998$ (en R se calcula ejecutando `pbinom(q = 0, size = 200, prob = 0.00001)`). Un mejor estimador en este caso sería entonces un conjunto de posibles valores de prevalencia que tengan altas probabilidades de haber dado origen al resultado observado $x = 0$. Lo anterior sería un estimador por intervalos o intervalar (en comparación con el primero que sería puntual), también conocido como intervalo de confianza (IC).

Existen varios métodos para estimar proporciones mediante intervalos de confianza, pero cada uno hace supuestos que podrían no hacerlos útiles en el caso planteado. Por ejemplo, varios suponen que alguna estadística pivotal sobre la cual se construye el intervalo distribuye aproximadamente normal, lo cual estaría lejos de cumplirse cuando la proporción que deseamos estimar (es decir, θ) tiene un valor cercano a cero, como planteamos en este caso. Uno de los estimadores que no hace tal supuesto distribucional es el de Clopper-Pearson, el cual consiste en una inversión del test binomial exacto (Clopper & Pearson, 1934). Con inversión de un test nos referimos a que se usa como IC aquellos valores del parámetro que, de ser usados como valor bajo la hipótesis nula (H_0), no resultarían en el rechazo de H_0.

En este caso invertimos el test binomial exacto de nivel de significancia α y una cola (o unidireccional) cuya H_0 es que $\theta \geq \theta_0$. Invirtiendo este test de hipótesis obtenemos un IC del $(1 - \alpha) \times 100\,\%$ de límite inferior 0 y límite superior $\widehat{\theta}_U$ tal que éste es el valor mínimo de θ para el cual se cumple que $\mathbb{P}(X \leq x) = \sum_{0 \leq u \leq x} \binom{n}{u} \theta^u (1 - \theta)^{n-u} > \alpha$ es verdadera. Cuando $x = 0$ la desigualdad anterior se simplifica a $\mathbb{P}(X = 0) > \alpha$, la cual puede ser resuelta para despejar θ, obteniendo $\theta \leq 1 - \alpha^{1/n}$ y por tanto el límite superior del IC es igual a $1 - \alpha^{1/n}$. Siguiendo con nuestro ejemplo, un intervalo del $95\,\%$ de confianza

*f.medina@uchile.cl.

estaría dado por $(0; 1-\alpha^{1/n}) = (0; 1-0,05^{1/200}) = (0; 1-0,985133) \approx (0; 0,0149)$. Nótese que este es un IC unilateral y que a diferencia de uno bilateral el complemento del nivel de confianza se ha distribuido de un solo lado del intervalo. Un intervalo bilateral, como el calculado con `binom.exact` del paquete `epitools` (Aragon, 2020), resultaría en $(0; 0,0183)$ y para obtener el intervalo que calculamos antes necesitaríamos ejecutar `binom.exact(x = 0, n = 200, conf.level = 0.90)`, i.e., usar un nivel de confianza de $1-2\alpha$ en el cálculo de un IC bilateral y quedarnos con el límite superior del IC.

Cuando se trabaja con un test imperfecto, pero de propiedades conocidas

Cuando el test no es perfecto, existe la posibilidad de que un resultado positivo observado sea en realidad un falso positivo (es decir, la realidad es que la persona no tiene la enfermedad), mientras que un resultado negativo podría ser un falso negativo (es decir, la persona en verdad sí tiene la enfermedad). Las tasas de falsos positivos y falsos negativos dependen de las cualidades de la prueba diagnóstica: su especificidad y sensibilidad. En nuestro ejemplo, al no observar casos, podemos sospechar que algunos de estos resultados podrían ser falsos negativos. Separemos los resultados positivos en los verdaderos positivos y falsos positivos. La probabilidad de un resultado positivo es

$$\begin{aligned}
\mathbb{P}(\text{test } +) &= \mathbb{P}(\text{test } + \text{ y enfermo}) + \mathbb{P}(\text{test } + \text{ y sano}) \\
&= \mathbb{P}(\text{test } +|\text{enfermo})\mathbb{P}(\text{enfermo}) \\
&\quad + \mathbb{P}(\text{test } +|\text{sano})\mathbb{P}(\text{sano}) \\
&= \text{sensibilidad} \times \text{prevalencia} \\
&\quad + (1 - \text{especificidad}) \times (1 - \text{prevalencia}) \\
&= s \times \theta + (1-e) \times (1-\theta)
\end{aligned}$$

Para obtener un IC trabajamos de la misma manera que antes, sólo que ahora en lugar de trabajar directamente con la prevalencia θ trabajamos con $\mathbb{P}(\text{test } +)$ como parámetro de la distribución binomial que suponemos dio origen a los datos. Comenzamos usando la fórmula del límite superior del IC:

$$\begin{aligned}
Pr(X \leq 0 | X \sim \text{Binomial}(n, \mathbb{P}(\text{test } +))) &> \alpha \\
[1 - s \times \theta + (1-e) \times (1-\theta)]^n &> \alpha \\
1 - s \times \theta + (1-e) \times (1-\theta) &> \alpha^{1/n}
\end{aligned}$$

Finalmente, luego de algo de álgebra despejamos θ, obteniendo

$$\theta \leq \frac{1 - \alpha^{1/n} - (1-e)}{s - (1-e)}$$

El límite resultante es numéricamente igual al estimador de Rogan-Gladen, $\widehat{\theta}_{RG} = \frac{P+(e-1)}{s+(e-1)}$, cuando se sustituye la prevalencia muestral P por el límite superior del intervalo de Clopper-Pearson (Rogan & Gladen, 1978).

Existen dos posibles problemas con el límite superior $\widehat{\theta}_U = \frac{1-\alpha^{1/n}-(1-e)}{s-(1-e)}$. Primero, esta fracción puede ser menor que cero si $1-\alpha^{1/n} < (1-e)$ o si $s < (1-e)$, y segundo, puede ser mayor que uno si $s < 1-\alpha^{1/n}$ o $(1-s)^n > \alpha$. En ambos casos el IC no es satisfactorio y no puede ser arreglado truncando el límite superior a 0 o 1 dependiendo del caso, ya que lo primero resulta en un intervalo que sólo contiene un elemento (el cero) y el otro en un intervalo que incluye a todo el espacio paramétrico, el rango de valores entre 0 y 1, incluidos. Estos problemas son análogos a los que presenta el estimador puntual $\widehat{\theta}_{RG}$.

Continuando con el ejemplo, si la sensibilidad y especificidad del test fuesen 1 y $0,99$, respectivamente, calculamos el límite superior del intervalo del $95\,\%$ de confianza como $\frac{1-\alpha^{1/n}-(1-e)}{s-(1-e)} = \frac{1-0,05^{1/200}-(1-0,99)}{1-(1-0,99)} \approx 0,0049$, aproximadamente cuatro veces menor que la prevalencia estimada cuando supusimos que la prueba era perfecta. Nótese que este es un IC unilateral. Para obtener este intervalo podemos usar la función `epi.prev` del paquete `epiR` (Stevenson & Sergeant, 2025), usando un nivel de confianza de $1-2\alpha$, ejecutando `epi.prev(pos = 0, tested = 200, se = 1.00, sp = 0.99, method = "c-p", units = 1, conf.level = 0.90)`.

Cabe señalar que el ejemplo presentado fue articulado para no resultar en un intervalo aberrante. Si la especificidad del test hubiese sido $0,95$ en lugar de $0,99$ habríamos obtenido un límite superior negativo, el cual al ser truncado a cero habría resultado en un intervalo degenerado en cero.

Cuando se trabaja con un test imperfecto de propiedades desconocidas

Finalmente nos queda considerar la posibilidad de que las propiedades del test de diagnóstico no sean del todo conocidas. Una forma de lidiar con este problema es usando una aproximación bayesiana (Lesaffre & Lawson, 2012). Esto consiste en asignar una distribución inicial (previa o *a priori*) al conjunto de los tres parámetros desconocidos del problema: la prevalencia poblacional θ, la sensibilidad del test diagnóstico σ y su especificidad ϵ. Luego, el algoritmo o método de Monte Carlo basado en cadenas de Markov (MCMC) realiaza un muestreo iterativo desde la distribución conjunta de parámetros y el límite superior del intervalo de credibilidad (no de confianza, ya que hemos pasado de una aproximación frecuentista a una bayesiana) se calcularía como el cuantil $0,95$ de los valores de θ que se muestrearon.

Para simplificar el algoritmo MCMC podemos de-

finir una distribución inicial independiente para cada parámetro del modelo. Luego tendríamos que programar el modelo en un lenguaje como JAGS o Stan para realizar el análisis. El Código 1 en el Anexo corresponde a una posible implementación del algoritmo en JAGS para ser ejecutado en R usando el paquete `rjags` (Plummer, 2025). En esta implementación, a los parámetros desconocidos del modelo se les asignaron distribuciones iniciales beta independientes: $\theta \sim \text{Beta}(1,1)$, $\sigma \sim \text{Beta}(20,5)$ y $\epsilon \sim \text{Beta}(20,5)$. Después de muestrear la distribución final (ver Figura 1), se obtiene que el límite superior del intervalo del $95\,\%$ de credibilidad calculado es aproximadamente igual a $0,0192$.

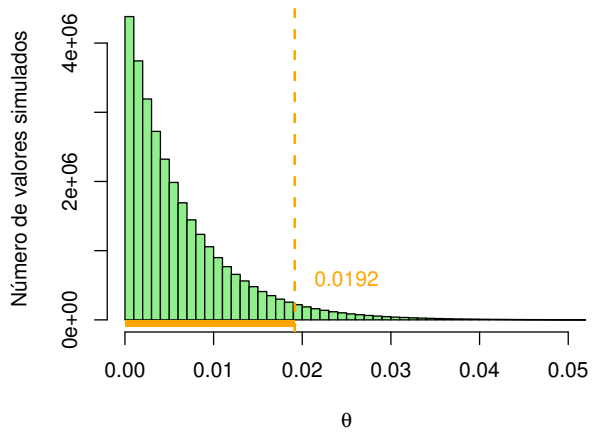

Figura 1: Histograma de la distribución final de θ estimada mediante MCMC usando `rjags`. La línea horizontal naranja debajo del histograma indica el intervalo unilateral del 95 % de credibilidad estimado.

Antes de cerrar, cabe remarcar que este último intervalo no se interpreta igual que los intervalos calculados en las secciones anteriores. Los intervalos calculados en las primeras secciones son frecuentistas: su nivel de confianza $1-\alpha$ describe el valor nominal de la probabilidad de cobertura o, en otras palabras, de que una muestra aleatoria resulte en un intervalo que capture al valor real del parámetro. En la estadística frecuentista la probabilidad de un evento tiene relación con el límite de la frecuencia relativa con la que se observaría tal evento si el número de repeticiones del experimento tendiera a infinito. El intervalo de esta última sección es diferente debido a que se construye bajo un paradigma en el que la probabilidad puede ser interpretada como un grado de creencia y por tanto podemos asignar distribuciones de probabilidad a los valores de los parámetros desconocidos. El intervalo calculado tiene una probabilidad del $95\,\%$ de contener el valor del parámetro de acuerdo con las creencias actualizadas (i.e., la integración de las creencias iniciales con los resultados observados).

Conclusiones

- Cero eventos no implica cero prevalencia (θ).
- Con cero eventos se puede calcular un límite superior de un intervalo de confianza (IC) unilateral de θ.
- Si el test de diagnóstico no es perfecto, pero es de sensibilidad y especificidad conocidas, entonces, bajo ciertas condiciones se puede calcular un límite superior de un IC unilateral de θ.
- Si el test de diagnóstico no es perfecto y se desconocen su sensibilidad y especificidad, entonces se puede recurrir al ajuste de un modelo bayesiano para poder calcular un límite superior de un intervalo de credibilidad unilateral de θ, aunque su interpretación difiere de la de un IC.

Referencias

Aragon, T. J. (2020). *epitools: Epidemiology Tools* [R package version 0.5-10.1]. https://doi.org/10.32614/CRAN.package.epitools

Clopper, C. J., & Pearson, E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, *26*(4), 404-413. https://doi.org/10.1093/biomet/26.4.404

Lesaffre, E., & Lawson, A. B. (2012). *Bayesian Biostatistics*. Wiley. https://doi.org/10.1002/9781119942412

Plummer, M. (2025). *rjags: Bayesian Graphical Models using MCMC* [R package version 4-17]. https://doi.org/10.32614/CRAN.package.rjags

Rogan, W. J., & Gladen, B. (1978). Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, *107*(1), 71-76.

Stevenson, M., & Sergeant, E. (2025). *epiR: Tools for the Analysis of Epidemiological Data* [R package version 2.0.83]. https://doi.org/10.32614/CRAN.package.epiR

Anexo: algoritmo de Monte Carlo basado en cadenas de Markov

Script en R para estimación de prevalencia

```
# Semilla de las simulaciones
set.seed(1984)

# Número de simulaciones
NSIM <- 10000000

# Datos observados
x <- 0
n <- 200

# Cargar paquete requerido
library(rjags)

# Definir modelo bayesiano
modelo_jags <- "model {
  # Verosimilitud
  x ~ dbin(prob_pos, n)

  # Probabilidad de un resultado positivo
  prob_pos <- sigma * theta + (1-epsil) * (1-theta)

  # Distribuciones iniciales
  theta ~ dbeta(1, 1)
  sigma ~ dbeta(20, 5)
  epsil ~ dbeta(20, 5)
}"

# Datos para JAGS
datos_jags <- list(x = x, n = n)

# Parámetros a monitorear
params <- c("theta", "sigma", "epsil")

# Inicializar el modelo
modelo <- jags.model(
  textConnection(modelo_jags),
  data = datos_jags,
  n.chains = 3)

# Burn-in
update(modelo, 1000)

# Muestreo de la distribución final
muestras <- coda.samples(
  modelo,
  variable.names = params,
  n.iter = NSIM)

# Muestra de valores de theta de la distribución final
theta_final <- as.matrix(muestras)[,"theta"]

# Límite superior del intervalo del 95% de credibilidad
(límite_sup <- quantile(theta_final, probs = 0.95))
```

```
# Figura
postscript(
  file = "figura.eps",
  width = 5,
  height = 4,
  horizontal = FALSE)
par(mar=c(5,4,2,2))
hist(
  theta_final,
  xlim = c(0, 0.05),
  breaks = 100,
  col = "lightgreen",
  main = "",
  xlab = expression(theta),
  ylab = "Número de valores simulados")
rect(
  xleft = 0,
  ybottom = -1e5,
  xright = límite_sup,
  ytop = 0,
  col = "orange",
  border = NA)
abline(
  v = límite_sup,
  col = "orange",
  lty = 2,
  lwd = 2)
text(
  x = 0.025,
  y = 600000,
  col = "orange",
  labels = round(x = límite_sup, digits = 4))
dev.off()
```