Práctica 5. Relación de Problemas de Intervalos de Confianza.

Juan de Dios Luna del Castillo, Antonio Martín Andrés, Christian J. Acal González

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo **BioestadisticaR** formado por:
Antonio Martín Andrés
Juan de Dios Luna del Castillo,
Pedro Femia Marzo,
Miguel Ángel Montero Alonso,
Christian José Acal González,
Pedro María Carmona Sáez,
Juan Manuel Melchor Rodríguez,
José Luis Romero Béjar,
Manuela Expósito Ruíz,
Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de **BioestadisticaR**.

Práctica 5. Relación de Problemas de Intervalos de Confianza

Juan de Dios Luna del Castillo, Antonio Martín Andrés y Christian J. Acal González

Ejercicio 5.1

Sólo una parte de los pacientes que sufren un determinado síndrome neurológico consiguen una curación completa. Si de 64 pacientes observados se han curado 41: **a)** Dar estimación puntual y por intervalo de la proporción de los que sanan. **b)** ¿Qué número de enfermos habría que observar para estimar la proporción de curados con un error inferior a 0,05 y una confianza del 95%?

Solución Manual

a) Estimación puntual y por intervalo de confianza de una proporción:

$$n = 64; \quad x = 41; \quad n - x = 23$$

Estimador puntual de la proporción de pacientes que sanan: $\hat{p} = \frac{x}{n} = \frac{41}{64} = 0.6406 = 64.06\%$.

Como x y $n - x$ son mayores que 20, el intervalo de confianza al 95% para la proporción de pacientes que sanan puede hacerse por el método de Wald: (**R.4.4.a.i**): la segunda parte de ella da

$$p \in \frac{x \pm \left\{ z_\alpha \sqrt{\frac{x(n-x)}{n}} + 0.5 \right\}}{n} = \frac{x \pm \left\{ z_{0.05} \sqrt{\frac{x(n-x)}{n}} + 0.5 \right\}}{n} = \frac{41 \pm \left\{ 1.96 \sqrt{\frac{41 \times 23}{64}} + 0.5 \right\}}{64} =$$
$$(0.5152; 0.7660) = (51.52\%; 76.60\%)$$

lo que da igual resultado que el obtenido por la fórmula más tradicional de su primera parte:

$$p \in \hat{p} \pm \left\{ z_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}} + \frac{1}{2n} \right\} = 0.6406 \pm \left\{ 1.96 \sqrt{\frac{0.6406 \times 0.3594}{64}} + \frac{1}{2 \times 64} \right\}$$

b) Cálculo del tamaño de muestra necesario para estimar la proporción de pacientes que sanan con un error inferior a 0.05 ($d = 0.05$) y una confianza del 95% ($1 - \alpha = 0.95$): **(R.4.4.b)**.

i) *Sin información* (ANTES de tomar la muestra anterior):

$$n = \left(\frac{z_\alpha}{2d} \right)^2 = \left(\frac{1.96}{2 \times 0.05} \right)^2 = 384.16 \to n = 385$$

ii) *Con información* (DESPUÉS de tomar la muestra anterior): Si se asume que la muestra anterior es una muestra piloto, entonces $\hat{p} = 64.06\%$ y $p \in (0.5152; 0.7660)$ al 95% como se vio antes:

- Lo primero es comprobar que tal muestra no verifica la precisión solicitada (párrafo final de **R.4.4.b**). Como $|64.06\% - 51.52\%| > d = 5\% \to$ no se verifica $\to$ hace falta más muestra.
- Como el valor de p más próximo a 0.5 es 0.5152, tomando $p = 0.5152$ y $q = 1 - p = 1 - 0.5152 = 0.4848$, el tamaño muestral es:

$$n = \left(\frac{z_\alpha}{d} \right)^2 pq = \left(\frac{1.96}{0.05} \right)^2 \times 0.5152 \times 0.4848 = 383.8 \to n = 384 \ (> 64) \to \text{faltan } 320$$

Solución con R

La solución sigue el mismo itinerario que la solución a mano salvo que han de ponerse los comandos de R que se necesitan para obtener los resultados y los resultados.

```r
#APARTADO A).
#Nótese que habría que consultar la salida "Método de Wald (con cpc)"
#Nótese que pueden producirse ligeras diferencias en los decimales por el redondeo que
#se hace a mano y por el redondeo que se hace con R.
library(BioestadisticaR)
icp(x=41, n=64)
```

```
##
## Intervalo de confianza para una proporción binomial
## ---------------------------------------------------
##
## Información muestral:
##    Tamaño de muestra: n =  64
##    Casos observados : x =  41
##
## Método de Wilson (con cpc):
##    Estimación puntual (clásica): p=x/n =  0.6406 , q=(1-p)= 0.3594
##    95 %-IC(p): ( 0.5103 ,  0.754 )
##    Semiamplitud: 0.1218
##
## Método de Wald (con cpc):
##    Estimación puntual (clásica): p=x/n =  0.6406 , q=(1-p)= 0.3594
##    95 %-IC(p): ( 0.5153 ,  0.766 )
##    Precisión: 0.1254
##
## Método de Wald ajustado:
##    Estimación puntual: p=(x+2)/(n+4) =  0.6324 , q=(1-p)= 0.3676
##    95 %-IC(p): ( 0.5178 ,  0.747 )
##    Precisión: 0.1146
```

```r
#APARTADO B).
#Nótese que habría que mirar el apartado "Tamaño muestral requerido para ..."
np(x=41, n=64, d=.05)
```

```
##
## Tamaño de muestra para estimar una proporción binomial
## -------------------------------------------------------
##
## Información muestral
##    tamaño de la muestra: n =  64
##    N: de casos: x =  41
##    Inferencia para la proporción basada en el método de Wald ajustado:
##    95 %-IC(p): ( 0.5178 ,  0.747 )
##    precisión observada: d =  0.1146  ( 11.46 %)
##
## Tamaño muestral requerido para d =  0.05 ( 5 %), conf.= 95 %
## - Basado en la muestra piloto (po = 0.5178 ):   n =  384
## - Sin considerar la información previa: n =  385
```

Ejercicio 5.2

Se desea estimar el peso medio de los niños varones de 12 semanas de vida. Si de una muestra de 25 de tales bebés se ha obtenido un promedio de 5.900 g con una desviación típica de 94 g, **a)** obtener un intervalo de confianza para el peso medio. **b)** ¿Cuántos bebés harían falta para estimar esa media con un error no superior a 15 g y una confianza del 95%?

Solución Manual

a) Intervalo de confianza para la media de una variable aleatoria: $n = 25$, $\overline{x} = 5900$ y $s = 94$.

Suponiendo que la variable aleatoria "peso de los niños varones" sigue una distribución Normal (en otro caso no podría responderse a las preguntas pues $n = 25$ es pequeño), el intervalo de confianza al 95% para el peso medio (μ) es: **(R.4.3.a.i)**

$$\mu \in \overline{x} \pm t_\alpha \frac{s}{\sqrt{n}} = 5900 \pm 2.064 \frac{94}{\sqrt{25}} = 5900 \pm 38.8032 \approx (5.861; 5.939)$$

pues $t_{5\%}(24\ gl) = 2.064$. Tenga en cuenta que se redondea "ensanchando" el intervalo a fin de que la confianza sea $\geq 95\%$.

b) Tamaño de muestra para determinar una media con muestra piloto **(R.4.3.b.ii)**: $d = 15$ y $1 - \alpha = 0.95$:

- Párrafo final de **R.4.3.b**: Como $(5.939 - 5.861)/2 = 39 > 15 \rightarrow$ la muestra piloto no tiene la precisión deseada $\rightarrow$ hace falta más muestra.
- Usando $t_{5\%}(24\ gl)$:

$$n = \left(\frac{t_\alpha s}{d}\right)^2 = \left(\frac{2.064 \times 94}{15}\right)^2 = 167.2987 \rightarrow n = 168\ (> 25) \rightarrow \text{faltan } 143$$

Solución con R

La solución sigue el mismo itinerario que la solución a mano salvo que han de ponerse los comandos de R que se necesitan para obtener los resultados y los resultados mismos y los comentarios decir que son los anteriores.

```r
#APARTADO A).
#Nótese que pueden producirse ligeras diferencias en los decimales por el redondeo que
#se hace a mano y por el redondeo que se hace con R.
icm(n=25, m=5900, s=94)
```

```
##
## Intervalo de confianza bilateral para la media de una VA normal
## -----------------------------------------------------------------
## Información muestral:
##   Tamaño muestral: n =  25
##   Media: m =  5900
##   Desviación típica: s =  94
##   Error estándar de la media: sem =  18.8
##
## Estimación:
##   95 %-IC(μ):  ( 5861.199 ,  5938.801 )
##   Precisión obtenida: 38.8013
```

```
#APARTADO B).
nm(n=25, m=5900, s=94, d=15)
```

```
##
## Tamaño de muestra para la estimación de la media de una VA normal o su aproximación
## ---------------------------------------------------------------------------------
## Muestra piloto:
##   Tamaño muestral actual:  n =  25
##   Media: m =  5900
##   Desviación típica: s = 94
##   Error estandar de la media: sem =  18.8
##   Precisión observada: d =  38.8013
##
## Estimación del tamaño muestral:
##   Precisión deseada: d =  15
##   Tamaño muestral necesario: n =  168
```

Ejercicio 5.3

En una determinada región se tomó una muestra aleatoria de 125 individuos, de los cuales 12 padecían afecciones pulmonares. **a)** Estímese la proporción de afecciones pulmonares en dicha región. **b)** Si queremos estimar dicha proporción con un error máximo del 4%, para una confianza del 95%, ¿qué tamaño de muestra debemos tomar? (considerar la muestra anterior como muestra piloto).

Solución Manual

a) Estimación puntual y por intervalo de confianza de una proporción: $n = 125$, $x = 12$ y $n - x = 113$.

- $\hat{p} = \frac{x}{n} = \frac{12}{125} = 0.096 = 9.6\%$
- Como x y $n - x$ son mayores que 5 (y una de ellas no es mayor que 20), el intervalo de confianza al 95% para la proporción de afecciones pulmonares es el de Wilson **(R.4.4.a.i)**

$$p \in \frac{(x \pm 0.5) + \frac{z_\alpha^2}{2} \pm z_\alpha \sqrt{\frac{z_\alpha^2}{4} + (x \pm 0.5)\left(1 - \frac{x \pm 0.5}{n}\right)}}{n + z^2\alpha} =$$

$$= \frac{(12 \pm 0.5) + \frac{1.96^2}{2} \pm 1.96\sqrt{\frac{1.96^2}{4} + (12 \pm 0.5)\left(1 - \frac{12 \pm 0.5}{125}\right)}}{125 + 1.96^2} = (0.0528; 0.1651) = (5.28\%; 16.51\%)$$

b) Tamaño de muestra con muestra piloto (**R.4.4.b.i**): $d = 4\%$ y $1 - \alpha = 0.95$:

- Como $|16.51\% - 9.6\%| = 6.91\% > 4\% \to$ la muestra piloto no tiene la precisión deseada $\to$ hace falta más muestra.
- Como según el IC $p \in (0.0528; 0.1651)$ al 95% de confianza $\to$ el valor más próximo a 0.5 es 0.1651 $\to$ tomando $p = 0.1651$ y $q = 1 - p = 1 - 0.1651 = 0.8349$, el tamaño muestral será:

$$n = \left(\frac{z_\alpha}{d}\right)^2 pq = \left(\frac{1.96}{0.04}\right)^2 \times 0.1651 \times 0.8349 = 330.96 \to n = 331\ (> 125) \to \text{faltan } 206$$

Solución con R

La solución sigue el mismo itinerario que la solución a mano salvo que han de ponerse los comandos de R que se necesitan para obtener los resultados y los resultados mismos y los comentarios decir que son los anteriores.

```
#APARTADO A).
#Nótese que habría que consultar la salida "Método de Wilson (con cpc)"
icp(x=12, n=125)
```

```
##
## Intervalo de confianza para una proporción binomial
## ---------------------------------------------------
##
## Información muestral:
##    Tamaño de muestra: n =  125
##    Casos observados : x =  12
##
## Método de Wilson (con cpc):
##    Estimación puntual (clásica): p=x/n =  0.096 , q=(1-p)= 0.904
##    95 %-IC(p): ( 0.0528 ,  0.1651 )
##    Semiamplitud: 0.0561
##
## Método de Wald (con cpc):
##   No aplicable: x= 12 <20 , n-x= 113
##
## Método de Wald ajustado:
##    Estimación puntual: p=(x+2)/(n+4) =  0.1085 , q=(1-p)= 0.8915
##    95 %-IC(p): ( 0.0549 ,  0.1622 )
##    Precisión: 0.0537
```

```
#APARTADO B).
#Nótese que habría que mirar el apartado "Tamaño muestral requerido para ..."
np(x=12, n=125, d=.04)
```

```
##
## Tamaño de muestra para estimar una proporción binomial
## -------------------------------------------------------
##
## Información muestral
##    tamaño de la muestra: n =  125
##    N: de casos: x =  12
##    Inferencia para la proporción basada en el método de Wald ajustado:
##    95 %-IC(p): ( 0.0549 ,  0.1622 )
##    precisión observada: d =  0.0537  ( 5.37 %)
##
## Tamaño muestral requerido para d =  0.04 ( 4 %), conf.= 95 %
## - Basado en la muestra piloto (po = 0.1622 ):   n =  327
## - Sin considerar la información previa: n =  601
```

Se observa que con la función ***np*()** no se consigue el mismo resultado que el obtenido a mano (con la salida de R n sería igual a 327 y a mano se ha calculado que n es 331). Esto se debe a que en dicha función el intervalo de confianza se elabora por el método de Wald ajustado (el cual es siempre válido) y no por el de Wilson. Por tanto, para alcanzar el mismo resultado habría que definir todos los parámetros y escribir la fórmula "a mano".

```
alpha=0.05
z=round(qnorm(alpha/2,lower.tail = F),2)
d=0.04
p=0.1651
q=1-p
n=round(p*q*(z/d)^2)
n
```

```
## [1] 331
```