Problemas de la Práctica VIII. Enunciados y soluciones.

Juan de Dios Luna del Castillo

UNIVERSIDAD DE GRANADA

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo **BioestadisticaR** formado por:
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**.
Todos los materiales están protegidos por la Licencia Creative Commons **CC BY-NC-ND** que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

VIII. 1 Con objeto de detectar si la hipertensión arterial leve (HTL) da lugar a una mayor presencia de arritmias ventriculares en la vida cotidiana, se tomaron dos grupos (uno de normotensos y otro de individuos con HTL) de 51 individuos cada uno. A todos los individuos se les sometió a una monitorización de un día, observándose que en el grupo HTL 28 habían presentado alguna arritmia, mientras que en el grupo de normotensos la habían presentado 21 individuos. **a)** ¿Se puede decir que la proporción de arritmias ventriculares es distinta en ambos grupos de individuos? **b)** ¿Qué tamaño de muestra haría falta para detectar una diferencia de un 13% entre ambas proporciones con una potencia del 95%, haciendo el test a un 5% de error? (resolver el problema antes y después de conocer los datos de antes).

Solución Manual y con Tablas:

(Esta solución se presenta aquí sólo a efectos indicativos, se ruega al alumno que revise y estudie la solución con R)

		Arritmias SÍ	Arritmias NO	Total	
HTL	SÍ	$28 = x_1$	$23 = y_1$	$51 = n_1$	$\rightarrow p_1$ = % arritmias en SÍ HTL
HTL	NO	$21 = x_2$	$30 = y_2$	$51 = n_2$	$\rightarrow p_2$ = % arritmias en NO HTL
	Total	$49 = a_1$	$53 = a_2$	$102 = N$	

a)

- H_0: $p_1 = p_2$ *vs.* H_1: $p_1 \neq p_2$ en muestras independientes (**R.7.5.a**):

$$\hat{p}_1 = \frac{28}{51} = 0{,}5490 \text{ ,, } \hat{p}_2 = \frac{21}{51} = 0{,}4118 \text{ ,, } \hat{p} = \frac{49}{102} = 0{,}4804 \text{ ,, } E = \frac{49 \times 51}{102} = 24{,}5 > 7{,}7$$

$$z_{exp} = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \dfrac{2}{n_1 n_2}}{\sqrt{\hat{p}\hat{q}\dfrac{n_1+n_2}{n_1 \times n_2}}} = \frac{\left|0{,}5490 - 0{,}4118\right| - \dfrac{2}{51\times 51}}{\sqrt{0{,}4804 \times 0{,}5196 \times \dfrac{102}{51 \times 51}}} = 1{,}359 \Rightarrow 17\% < P < 18\%$$

$\Rightarrow$ hay muy leves indicios H_1 (de que la proporción de arritmias sean diferentes en ambos grupos), pero con los datos actuales hay que aceptar que ambas proporciones son iguales ($p_1 = p_2$) ($P > 17\%$).

- Para ver si es fiable esta decisión por H_0 es preciso recurrir a los valores $\beta = 5\%$ y $\delta = 13\%$ del apartado **b)**. Por (**R.7.5.c.*i***) hay que hacer un IC al error $2\beta = 10\%$ (**R.7.5.b.*ii*** con $h = 1{,}645^2/4 \approx 0{,}7$):

$$p_1 - p_2 \in \frac{28{,}7}{52{,}4} - \frac{21{,}7}{52{,}4} \pm 1{,}645\sqrt{\frac{28{,}7 \times 23{,}7}{52{,}4^3} + \frac{21{,}7 \times 30{,}7}{52{,}4^3}} \Rightarrow -2{,}55\% \leq p_1 - p_2 \leq +29{,}27\%$$

$\Rightarrow$ como $+\delta = 13\%$ está dentro $\Rightarrow H_0$ no fiable $\Rightarrow$ aumentar n_i.

b) $\alpha = \beta = 5\%$, $\delta = 13\%$:

- ANTES de conocer los datos ≡ SIN información (**R.7.5.c.*ii***):

$$n_1 = n_2 = n = \frac{1}{2}\left[\frac{z_\alpha + z_{2\beta}\sqrt{1-\delta^2}}{\delta}\right]^2 = \frac{1}{2}\left(\frac{1{,}96 + 1{,}645\sqrt{1-0{,}13^2}}{0{,}13}\right)^2 = 382$$

- DESPUÉS de conocer los datos ≡ CON información (**R.7.5.c.*i***): los datos del párrafo **a)** hacen de muestras piloto:

  1) La información sobre p_1 y p_2 se obtiene realizando un IC para cada p_i a partir de cada una de las dos muestras del enunciado: x = 28 entre n = 51 para p_1; x = 21 entre n = 51 para p_2.

  2) Pero como las p_i hay que hacerlas lo más próximas posibles a $0{,}5 \pm 0{,}13/2 = 0{,}435$ y $0{,}565$, en tanto que las estimaciones son $\hat{p}_i \approx 0{,}55$ y $0{,}41$ (que son próximas a las anteriores) $\Rightarrow$ al calcular el IC para p_1 y p_2 van a dar valores compatibles con los primeros $\Rightarrow$ p_1 = 0,565 y p_2 = 0,435.

  3) ¡Esos son los mismos valores que ocasionan el caso de "Sin información"! $\Rightarrow$ la solución es igual.

Solución con R:

		Arritmias SÍ	Arritmias NO	Total	
HTL	SÍ	28 = x_1	23 = y_1	51 = n_1	→ p_1 = % arritmias en SÍ HTL
HTL	NO	21 = x_2	30 = y_2	51 = n_2	→ p_2 = % arritmias en NO HTL
	Total	49 = a_1	53 = a_2	102 = N	

Hemos dispuesto la tabla ahora para comodidad en el estudio y desarrollo de la solución con R.

Del enunciado del problema se deduce que tenemos definidas dos variables aleatorias:

$$x_1 \equiv \text{"}Número\ de\ individuos\ que\ sufren\ arritmias\ de\ entre\ los\ 51\ con\ HTL \rightarrow\ B(51, p_1)\text{"}$$
$$x_2 \equiv \text{"}Número\ de\ individuos\ que\ sufren\ arritmias\ de\ entre\ los\ 51\ sin\ HTL \rightarrow\ B(51, p_2)\text{"}$$

A partir de ellas que el contraste de hipótesis que vamos a realizar es el que tiene como hipótrsis nula que las dos proporciones son iguales:

$$H_0 \equiv p_1 = p_2\ vs\ H_1 \equiv p_1 \neq p_2$$

Los pacientes que participan en el estudio o son HTL o no son HTL, con lo cual cada paciente aporta un único a valor a los datos de las dos muestras. Todo ello nos afirma que las dos muestras son independientes.

Para resolver el problema tenemos que hacer un test de comparación de dos proporciones con muestras independientes.

Antes de nada, verificaremos las condiciones de validez

$$\hat{p}_1 = \frac{28}{51} = 0{,}5490\ ,,\ \hat{p}_2 = \frac{21}{51} = 0{,}4118\ ,,\ \hat{p} = \frac{49}{102} = 0{,}4804\ ,,\ E = \frac{49 \times 51}{102} = 24{,}5 > 7{,}7$$

Que pueden ser calculadas con la calculadora o directamente en R. En efecto E=24.5 que es mayor que 14.9, condición que se requeriría si acaso fuera N>500 o mayor que 7.7 si acaso N≤500, que es nuestro caso. Luego se puede emplear el test aproximado.
Con R la función que habría que emplear es la función prop.test, proporcionándole los datos de las casillas de la tabla 2x2. La orden y los resultados figuran en la siguiente tabla.

Hemos dispuesto la tabla ahora para comodidad en el estudio y desarrollo de la solución con R.

Tanto las instrucciones como los resultados están en la tabla siguiente. Los comentarios de dicha información se harán al final de la casilla en la que se presentan.

```
> #Problema 8.1.
> #Apartado a)
> #Cálculo de E para la condición de validez.
> E=49*51/102
> E
[1] 24.5
> x=c(28,21)
> prop.test(x=c(28,21), n=c(51,51))

        2-sample test for equality of proportions with continuity correction

data:  c(28, 21) out of c(51, 51)
X-squared = 1.4139, df = 1, p-value = 0.2344
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.0744309  0.3489407
sample estimates:
   prop 1    prop 2
0.5490196 0.4117647

> #Apartado b)
> # Solo para hacer el intervalo de Agresti y Caffo
> prop.test(x=c(29,22), n=c(53,53), correct = FALSE)

        2-sample test for equality of proportions without continuity correction

data:  c(29, 22) out of c(53, 53)
X-squared = 1.8517, df = 1, p-value = 0.1736
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.05648858  0.32063952
sample estimates:
   prop 1    prop 2
0.5471698 0.4150943

> #Apartado c)
> #Dar intervalos de confianza para p1 y p2
> library(BioestadisticaR2)
> icp(x=28, n=51)

Intervalo de confianza para una proporción binomial
---------------------------------------------------

Información muestral:
   Tamaño de muestra: n =  51
   Estimación puntual clásica: p=x/n =  0.549 , q=(1-p)= 0.451
   Casos observados : x =  28

Método exacto (Clooper-Pearson):
   Pseudo-estimación puntual: p' =  0.5512 , q'=(1-p')= 0.4488
   95 %-IC(π): ( 0.4137 ,  0.6887 )
   Semiamplitud: 0.1375

Método de Wilson (con cpc):
   Pseudo-estimación puntual: p' =  0.5453 , q'=(1-p')= 0.4547
   95 %-IC(π): ( 0.4045 ,  0.6862 )
   Semiamplitud: 0.1408

Método de Wald (con cpc):
   Estimación puntual (clásica): p=x/n =  0.549 , q=(1-p)= 0.451
   95 %-IC(π): ( 0.4027 ,  0.6954 )
```

```
   Precisión: 0.1464

Método de Wald ajustado (Agresti-Coull):
   Estimación puntual: p=(x+2)/(n+4) =  0.5455 , q=(1-p)= 0.4545
   95 %-IC(π): ( 0.4139 ,  0.677 )
   Precisión: 0.1316

> icp(x=21,n=51)

Intervalo de confianza para una proporción binomial
---------------------------------------------------

Información muestral:
   Tamaño de muestra: n =  51
   Estimación puntual clásica: p=x/n =  0.4118 , q=(1-p)= 0.5882
   Casos observados : x =  21

Método exacto (Clooper-Pearson):
   Pseudo-estimación puntual: p' =  0.4192 , q'=(1-p')= 0.5808
   95 %-IC(π): ( 0.28 ,  0.5583 )
   Semiamplitud: 0.1391

Método de Wilson (con cpc):
   Pseudo-estimación puntual: p' =  0.4184 , q'=(1-p')= 0.5816
   95 %-IC(π): ( 0.2789 ,  0.5579 )
   Semiamplitud: 0.1395

Método de Wald (con cpc):
   Estimación puntual (clásica): p=x/n =  0.4118 , q=(1-p)= 0.5882
   95 %-IC(π): ( 0.2669 ,  0.5566 )
   Precisión: 0.1449

Método de Wald ajustado (Agresti-Coull):
   Estimación puntual: p=(x+2)/(n+4) =  0.4182 , q=(1-p)= 0.5818
   95 %-IC(π): ( 0.2878 ,  0.5485 )
   Precisión: 0.1304

> power.prop.test(p1=0.565,p2=0.435, power=0.95, sig.level=0.05)

     Two-sample comparison of proportions power calculation

              n = 381.4873
             p1 = 0.565
             p2 = 0.435
      sig.level = 0.05
          power = 0.95
    alternative = two.sided

NOTE: n is number in *each* group
```

En primer lugar, hemos calculado el valor de E, para ver si podíamos aplicar el test aproximado explicado en las clases de Teoría. El valor que sale ha resultado ser de 24.5 superior a 7.7 que es el punto de corte cuando N≤500 que es el caso. Por tanto, se puede aplicar el test aproximado que aparece en los resultados a continuación.

Hecho el test se obtiene

```
        2-sample test for equality of proportions with continuity correction

data:  c(28, 21) out of c(51, 51)
X-squared = 1.4139, df = 1, p-value = 0.2344
```

```
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.0744309  0.3489407
sample estimates:
   prop 1    prop 2
0.5490196 0.4117647
```

Dónde los resultados del test son, χ^2_{exp}=1.41, 1g.l., P=0.2344.

En primer lugar, lo que llama la atención es el estadístico de contraste que es una Chi-cuadrado con 1 g.l., cuando en clase se ha usado una z_{exp} que sigue una N(0,1) bajo la hipótesis nula. Pues bien, no hay ninguna contradicción puesto que se puede demostrar que $z^2_{exp}=\chi^2_{exp}$ y que $z^2_{\alpha}=\chi^2_{\alpha}$; por tanto, estamos en presencia del mismo test expresado de manera diferente, lo que nos dice que el valor de P es el mismo en ambos casos. El valor de P=0.2344, que, si no hemos fijado un α de antemano, nos llevará a aceptar H_0 sin más que aplicar la Regla Automática de Decisión. Es decir, no podremos rechazar la hipótesis nula y habremos de aceptar que ambas proporcionales son iguales. No obstante, esa afirmación no debería dejarlos tranquilos puesto que el intervalo de confianza al 95% para la diferencia de proporciones es $p_1 - p_2 \in (-0.0744; 0.3489)\ 95\%\ de\ confianza$, lo que quiere decir que el porcentaje de pacientes que sufren arritmias entre los pacientes con HTL es entre un 38.48% mayor y un 7.44 % inferior que tal porcentaje en el caso de pacientes que no sufran HTL. Aunque este intervalo no esté hecho al error 2β, muestra una diferencia entre ambas proporciones de hasta un 35% que es lo suficientemente grande como para que el test la detecte.

Por tanto, no nos quedaremos tranquilos con la aceptación de la hipótesis nula, por eso atenderemos al siguiente apartado del problema. El intervalo para la diferencia de proporciones que se ha calculado es el clásico de Wald, pero sabemos que el mejor sería el de Agresti y Caffo que se consigue haciendo el de Wald pero para la tabla incrementada en 1 unidad para cada uno de los valores x_i e incrementando en 2 unidades los n_i. Eso hemos hecho con R y hemos obtenido el intervalo de Agresti y Caffo, el razonamiento correcto sería con ese intervalo pero como no hay, en este caso, diferencias importantes entre uno y otro, pues no lo repetiremos y seguiremos opinando lo mismo.

b) Se trata en este caso de calcular un tamaño de muestra para detectar una diferencia δ=13%, con una probabilidad de detectarla cuando realmente sea cierta(potencia), de un 95% si es que se hace el test a un nivel de error α=5%. Para hacer esto recuérdese que hay que proponer valores para p_1 y p_2. Empezaremos por calcular intervalos de confianza para amabas proporciones que nos dan los siguientes resultados: $p_1 \in (0.4139; 0.6770)\ 95\%\ de\ confianza$ y $p_2 \in (0.2878; 0.5485)\ 95\%\ de\ confianza$. Cojamos p_1(sería análogo con p_2) y elijamos el valor del intervalo de p_1 que esté más cerca $0.5 \pm \delta/2$, es decir 0.5-0.065=0.435 y 0.5+0.065=0.565. Cualquiera de ellos dos está dentro del intervalo para p_1, por tanto, tomaremos uno de ellos, por ejemplo 0.5655. Cogeremos para p_2 un valor que diste 0.13 de 0.565y que esté en el intervalo de p_2 y ese es sólo el valor 0.435. Así hemos pedido el tamaño de muestra con la función power.prop.test y hemos obtenido un tamaño de muestra de n_1=n_2=382. Este resultado nos lleva a concluir que la aceptación hecha H_0 no es fiable y que hay que aumentar el tamaño de muestra hasta 382 casos por muestra, resultado muy lejano de los 51 que tenemos.

VIII. 2 Para valorar comparativamente dos métodos de diagnóstico A y B se sometió a 180 pacientes a ambos métodos. Pruebas posteriores al desarrollo de la enfermedad permitieron comprobar que el método A había dado un diagnóstico correcto en 124 casos, 116 de los cuales habían sido diagnosticados correctamente por el B, que a su vez diagnosticó correctamente a 27 de los 56 fallados por el A. **a)** ¿Es distinta la capacidad de acierto de ambos métodos?; **b)** ¿Cuánto de distintas son? (confianza del 95%)

Solución Manual y con Tablas:

(Esta solución se presenta aquí sólo a efectos indicativos, se ruega al alumno que revise y estudie la solución con R)

		Acierta B			
		SÍ	NO	Total	
Acierta A	SÍ	$116 = n_{11}$	$8 = n_{12}$	124	$\rightarrow p_A$ = % aciertos A (= p_1)
	NO	$27 = n_{21}$	$29 = n_{22}$	56	
	Total	143	37	$180 = n$	

$\downarrow$

p_B = % aciertos B (= p_2)

a) H_0: $p_A = p_B$ *vs.* H_1: $p_A \neq p_B$ con muestras apareadas (**R.7.6.a**): Como 8+27 = 35 > 10 $\Rightarrow$ test válido:

$$z_{exp} = \frac{|n_{12} - n_{21}| - 0{,}5}{\sqrt{n_{12} + n_{21}}} = \frac{|8 - 27| - 0{,}5}{\sqrt{8 + 27}} = 3{,}127 \Rightarrow H_1 \left(P < 0{,}2\% \right) \Rightarrow$$

Hay evidencias (P < 0,2%) de que $p_A < p_B$ (pues 124 < 143 ≡ 8 < 27) $\Rightarrow$ B acierta más que A.

b) R.7.6.b.*ii* (permutando el orden para que los extremos salgan positivos):

$$p_B - p_A = p_2 - p_1 \in \frac{(n_{21} - n_{12}) \pm z_\alpha \sqrt{(n_{12} + n_{21} + 1) - \frac{(n_{12} - n_{21})^2}{n+2}}}{n+2} = \frac{19 \pm \left\{ 1{,}96 \sqrt{36 - \frac{19^2}{182}} \right\}}{182}$$

$\Rightarrow$ 4,16% $\leq p_2 - p_1 \leq$ 16,72% al 95% de confianza.

Solución con R:

		Acierta B			
		SÍ	NO	Total	
Acierta A	SÍ	$116 = n_{11}$	$8 = n_{12}$	124	$\rightarrow p_A$ = % aciertos A (= p_1)
	NO	$27 = n_{21}$	$29 = n_{22}$	56	
	Total	143	37	$180 = n$	

$\downarrow$

p_B = % aciertos B (= p_2)

Hemos dispuesto la tabla ahora para comodidad en el estudio y desarrollo de la solución con R.

Tanto las instrucciones como los resultados están en la tabla siguiente. Los comentarios de dicha información se harán al final de la casilla en la que se presentan.

a) H_0: $p_A = p_B$ *vs.* H_1: $p_A \neq p_B$ con muestras apareadas

```
> #Problema 8.2.
> #Apartado a)
> #Condciones de validez
> #n12+n21>10, 8+27=35>10 La aproximación es válida
> #Test de Mcnemar
> tablamcn<-matrix(c(116,8,27,29),2,2)
> p1=8/180
> p1
[1] 0.04444444
> p2=27/180
> p2
[1] 0.15
> mcnemar.test(tablamcn)
```

```
        McNemar's Chi-squared test with continuity correction

data:  tablamcn
McNemar's chi-squared = 9.2571, df = 1, p-value = 0.002346
```

Los resultados son claros $\hat{p}_A = 0.0444\ y\ \hat{p}_B = 0.1500$ y hecho el test de McNemar (que se puede hacer porque $n_{12} + n_{21} = 7 + 28 = 35 > 10$), se obtuvo que P=0.0023 lo que nos dice que la probabilidad de acertar con A es diferente de la probabilidad de acertar con B, pero dicho esto podemos ir un poco más allá y dado que los estimadores puntuales guardan un determinado orden, podemos afirmar que la probabilidad de acertar con el método A es menor que la probabilidad de acertar con B. Atendiendo a ese parámetro, la probabilidad de acertar, el B sería el preferido.

Como en el guion de la práctica no se explica como calcular el tamaño de muestra para la diferencia de proporciones apareadas ni el cálculo del tamaño de muestra en ese caso, no seguiremos con la solución del problema empleando R.