

# Técnicas de Análisis Cuantitativas y Cualitativas

## Resolución del ejercicio de evaluación 1

Marcos Rial Docampo

16 de enero de 2016

En el ejercicio propuesto se presentan dos tipos de resultados de forma gráfica aparecidos en prensa. Los primeros resultados son los referentes a la información extraída del resultado de las elecciones municipales de Madrid del 24 de mayo de 2015 en la que se muestra en forma de gráfico el porcentaje de votos recibidos por las diferentes formaciones políticas. Los segundos resultados son extraídos de una encuesta de estimación de voto en el municipio de Madrid hecha entre los días 4 y 10 de septiembre de 2015.

El primer error, en lo referente a la presentación de los datos, es la de no incluir en los datos de la encuesta los intervalos de confianza. En el primer caso no sería correcto incluir el intervalo de confianza puesto que se trata de datos obtenidos de unos resultados electorales donde se convocaba a participar al total de la población con derecho a voto del municipio de Madrid (censo electoral de 2.597.411 personas<sup>1</sup>). No se trata de una encuesta, como en el segundo caso, donde el tamaño muestral es de 502 personas consultadas.

Para empezar, podemos tomar el ejercicio como un conjunto de cuatro apartados individuales en los que cada uno trata un modelo discreto distinto bajo una distribución binomial. Por ejemplo, tomar para el caso del PP que un 33,7 % de la población lo votaría frente al 66,7 % que no lo haría (tomamos, por tanto, un éxito de 0,337).

Para obtener los datos de límites superior e inferior del intervalo de confianza mostrados en el cuadro 1 utilizamos el comando *prop.test()* de R como se especifica en la figura 1. En él simplemente necesitamos especificar el número de elementos de éxito en la muestra frente al tamaño de la misma así como el nivel de confianza (para este ejercicio será del 95 %). Estos datos podrían verse reflejados en una gráfica similar a la de la figura 1 del enunciado del ejercicio pero, en lugar de los porcentajes y número de concejales, se pondrían los intervalos de confianza calculados.

Se indica en la noticia que el error de muestreo es de  $\pm 4,5\%$ . Esto quiere decir que la diferencia entre los resultados de la encuesta y los reales (preguntando a toda la población) sería de un 4,5 %. Una forma de reducir este error podría ser la de ampliar el tamaño de la muestra pero es un valor que no debería generalizarse a toda la encuesta puesto que tenemos resultados porcentuales de lo que se espera obtener en la encuesta

---

<sup>1</sup>Fuente: Datos electorales del Instituto Nacional de Estadística para las elecciones del 24 de mayo de 2015.

Partido	Elecciones (mayo)	Encuesta (septiembre)		
	Valor	Valor	LI*	LS*
PP	34,6 %	33,7 %	29,6	38,0
Cs	11,4 %	13,8 %	10,9	17,1
AM	31,9 %	29,1 %	25,2	33,3
PSOE	15,3 %	17,4 %	14,2	21,0

\* LI y LS límites inferior y superior del intervalo de confianza al 95 %

Cuadro 1: Resumen de los resultados de las elecciones de mayo y la encuesta de septiembre.

para cada uno de los partidos políticos (reflejados en los resultados electorales de mayo). Es decir, que cada dato para cada partido político obtendría su propio dato de error de muestreo.

Un error muestral del  $\pm 4,5\%$  se antoja bastante alto para afirmar con rotundidad, como dice el titular del artículo, que la candidata del Partido Popular (PP) recuperaría la alcaldía de haber un pacto con Ciudadanos (C's). Si a esto añadimos el hecho de que los intervalos de confianza son bastante amplios refuerza la incorrecta inferencia hecha por el periodista. Puede deberse a que en encuestas electorales publicadas en prensa este valor se toma como si fuera el intervalo de confianza haciendo pensar que se trata de un margen de error<sup>2</sup>. El error de muestreo correcto para esta encuesta, siguiendo la ecuación 1 sería de un  $\pm 2,2\%$ .

$$E = \sqrt{\frac{p \cdot (1 - p)}{n}} \quad (1)$$

siendo  $p$  el valor tomado de 0,5 y  $n$  el tamaño de la muestra de 502 elementos.

Para asegurar lo dicho en el párrafo anterior podemos tomar los votos conjuntos a PP y C's para evaluarlos separadamente. Obtenemos así para un 47,5 % de los votos un intervalo de confianza al 95 % de 43,0 - 51,9.

Por extensión al ejercicio podemos realizar un contraste de hipótesis también con el comando `prop.test()` utilizando como hipótesis nula que el valor obtenido en la encuesta es igual al de los resultados electorales del 24 de mayo. Por lo tanto, aplicaríamos el comando como se muestra en la figura 2 obtenemos los resultados siguientes:

Vemos un valor bajo de  $p$ -value de C's, pero que no indica que sea correcto rechazar la hipótesis nula. En general ninguno lo indica, no pudiendo rechazar en ningún caso dicha hipótesis nula, así que podemos aceptar que los valores de la encuesta se adaptan a los resultados obtenidos en las elecciones.

Concluimos que una propuesta para un posible titular apto para la noticia teniendo en cuenta lo anteriormente expuesto sería "Según la encuesta de GAD3 para ABC existiría

<sup>2</sup>Entrada en Explorable.com donde se trata el tema del error de muestreo en encuestas de opinión de periódicos <https://explorable.com/es/error-de-muestreo-aleatorio>

Partido	<i>p-value</i>
PP	0,6941
Cs	0,1134
AM	0,1916
PSOE	0,2294

*Cuadro 2: Datos obtenidos del contraste de hipótesis.*

```

1 #Aplicacion de prop.test para definir los intervalos de confianza con n
  = 502 y nivel de confianza del 95%
2 #PP con 33.7% de los votos
3 prop.test(round(0.337*502, digits = 0),502,
4           conf.level = 0.95)
5
6 #Cs con 13.8% de los votos
7 prop.test(round(0.138*502, digits = 0),502,
8           conf.level = 0.95)
9
10 #AM con 29.1% de los votos
11 prop.test(round(0.291*502, digits = 0),502,
12           conf.level = 0.95)
13
14 #PSOE con 17.4% de los votos
15 prop.test(round(0.174*502, digits = 0),502,
16           conf.level = 0.95)

```

*Figura 1: Script en R con la aplicación de prop.test a los datos de la encuesta. Se aplica redondeo para ofrecer un número entero de votantes por cada partido.*

la posibilidad de que el PP recuperara Madrid tras 90 días”.

```

1 #Contraste de hipotesis
2 #H0 = porcentajes de votos de la encuesta iguales al resultado
   electoral
3 #PP con 33.7% de los votos frente al 34.6% obtenido en las elecciones
4 prop.test(round(0.337*502, digits = 0),502,
5           p = 0.346, alternative = "two.sided",
6           conf.level = 0.95)
7
8 #Cs con 13.8% de los votos frente al 11.4% obtenido en las elecciones
9 prop.test(round(0.138*502, digits = 0),502,
10          p = 0.114, alternative = "two.sided",
11          conf.level = 0.95)
12
13 #AM con 29.1% de los votos frente al 31.9% obtenido en las elecciones
14 prop.test(round(0.291*502, digits = 0),502,
15          p = 0.319, alternative = "two.sided",
16          conf.level = 0.95)
17
18 #PSOE con 17.4% de los votos frente al 15.3% obtenido en las elecciones
19 prop.test(round(0.174*502, digits = 0),502,
20          p = 0.153, alternative = "two.sided",
21          conf.level = 0.95)

```

*Figura 2: Script en R con la aplicación de `prop.test` para realizar un contraste de hipótesis. Se aplica redondeo para ofrecer un número entero de votantes por cada partido.*