Problemas resueltos
Muestreo

Vicente Manzano-Arrondo, 2013

En todos los casos vamos a suponer que las muestras se obtienen siguiendo un muestreo aleatorio simple desde poblaciones de tamaño prácticamente infinito, salvo en los problemas 1 y 7, situaciones donde la población es claramente finita.

Problema 1. Extenso

Tenemos una población formada por tres datos de valores 1, 2 y 6. a) Obtén todas las posibles muestras aleatorias simples de n=2 y calcula para cada una de ellas la media aritmética y la proporción de datos con valores inferiores a 5; b) A partir de la distribución muestral de medias, calcula el valor esperado y el error tipo; c) haz lo mismo con la distribución muestral de proporciones; d) Calcula los dos valores esperados y los dos errores tipo pero esta vez a partir de los datos de la población.

Solución

a)

	Elementos	Media	Proporción
1	1, 2	1,5	1,0
2	1, 6	3,5	0,5
3	2, 6	4,0	0,5

b) Para el valor esperado y el error tipo de la media:

$$E(\bar{X}) = \frac{1{,}5 + 3{,}5 + 4}{3} = 3$$

$$\sigma_{\bar{X}} = \sqrt{\frac{(1{,}5 - 3)^2 + (3{,}5 - 3)^2 + (4 - 3)^2}{3}} = 1{,}08$$

c) Para el valor esperado y el error tipo de la proporción:

$$E(p) = \frac{1 + 0{,}5 \cdot 2}{3} = 0{,}667$$

$$\sigma_p = \sqrt{\frac{(1 - 0{,}667)^2 + 2\,(0{,}5 - 0{,}667)^2}{3}} = 0{,}236$$

d) Para el caso de las medias:

$$\mu = \frac{1+2+6}{3} = 3$$

$$\sigma = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (6-3)^2}{3}} = 2{,}16$$

$$E(\bar{X}) = \mu = 3$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{2{,}16}{\sqrt{2}} \sqrt{\frac{3-2}{3-1}} = 1{,}08$$

Y para el caso de las proporciones:

$$\pi = \frac{2}{3} = 0{,}667$$

$$\sigma = \sqrt{\pi\,(1-\pi)} = \sqrt{0{,}667\,(1-0{,}667)} = 0{,}471$$

$$E(p) = \pi = 0{,}667$$

$$\sigma_p = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{0{,}471}{\sqrt{2}} \sqrt{\frac{3-2}{3-1}} = 0{,}236$$

Otros problemas

2. En una población de gran tamaño, el porcentaje de personas que leen un periódico al menos cinco días a la semana es del 45%. a) ¿Cuál es la desviación tipo poblacional? b) Si extraemos

muestras de 49 personas, ¿Cuál es el error tipo de la proporción?

3. Según el texto AR4, la ansiedad-rasgo se distribuye en la población con una desviación tipo de valor 16. ¿Cuál es la desviación tipo de la distribución muestral de medias obtenidas a partir de muestras con n=64?

4. ¿En qué situación la desviación tipo poblacional es el doble que el error tipo de medias o de proporciones?

5. ¿Cuál es el valor de la desviación tipo poblacional si el error tipo de la media aritmética con muestras de tamaño n=49 tiene el valor 0,4?

6. ¿Qué valor de proporción hace que la varianza de proporciones sea máxima?

7. En una clase hemos preguntado por el grado en que consideran que la población estudiantil tiene poder para cambiar el mundo. La respuesta se concreta en el intervalo (0,10). Los resultados son:
   4, 7, 5, 6, 3, 5, 3, 0, 10, 5, 0, 4, 6, 5, 4, 5, 6, 3, 10, 2,
   6, 7, 4, 10, 5, 0, 5, 7, 6, 10, 7, 6, 4, 0, 7, 3, 5, 6,

5, 4

a) Calcula la media y la desviación tipo de la distribución muestral de medias con muestras de 9 personas.

b) Obtén la media y la desviación tipo de la distribución muestral de proporciones donde las respuestas son mayores a 5 y n=25.

8. En un caso de medias aritméticas, la desviación tipo poblacional y el error tipo tienen alguno de estos valores: 0,5 y 5, aunque no se especifica cuál de estos valores se corresponde con la desviación tipo poblacional y cuál con el error tipo. Calcula el tamaño de las muestras utilizadas en la construcción de la distribución muestral, sabiendo que la población se ha considerado como infinita.

9. ¿Cuál es el valor del error tipo de proporciones si estamos utilizando muestras de tamaño 36 provenientes de una población infinita y el valor esperado es 0,4?

Soluciones a los problemas 2 a 9

2:

a) $\sigma = \sqrt{\pi\left(1-\pi\right)} = \sqrt{0{,}45\left(1-0{,}45\right)} = 0{,}497$ b)

$$\sigma_p = \frac{\sigma}{\sqrt{n}} = \frac{0{,}497}{\sqrt{49}} = 0{,}071$$

3:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{64}} = 2$$

4:

$$\sigma_{\bar{X}\,p} = \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \sigma = \sqrt{n}\,\sigma_{\bar{X}\,p} = 2\,\sigma_{\bar{X}\,p} \quad \rightarrow \quad \sqrt{n} = 2 \quad \rightarrow \quad n$$

5:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \sigma = \sigma_{\bar{X}}\sqrt{n} = 0{,}4\sqrt{49} = 2{,}8$$

6:

Como sabemos, una forma cómoda de calcular la varianza en el caso de trabajar con proporciones, donde hemos codificado con unos y ceros los datos (1: se cumple la condición y 0: no se cumple), es multiplicar la proporción por su distancia a 1, es decir:

$$S^2 = p\left(1-p\right) \qquad \sigma^2 = \pi\left(1-\pi\right)$$

Una forma de plantear el problema desde nuestras habilidades es probar sistemáticamente varios valores. Lo podemos hacer en la siguiente tabla:

0,1	0,2	0,3	0,4	0,5	0,6	0,7	0
0,9	0,8	0,7	0,6	0,5	0,4	0,3	0
0,09	0,16	0,21	0,24	0,25	0,24	0,21	0,1

Hemos trabajado demasiado. Observa que, como es obvio porque p+(1-p)=1, al llegar a p=0,5, ya no es necesario seguir porque 1-p coincide con los valores previos de p y las parejas p & (1-p) se repiten a partir de ese momento. Pero es mejor verlo que explicarlo. Como puedes ver, el valor máximo para la varianza ocurre con p = 1-p = 0,5. Por muchos ensayos que lleves a cabo, incluyendo más decimales, no encontrarás un valor de varianza superior a 0,25 para el trabajo con proporciones. Es, además, intuitivo, puesto que los valores de proporción muy pequeños (cercanos a 0), disminuyen sensiblemente el valor del producto (cercano a 0). Y lo más lejano a 0, es la situación ya indicada de p = 1-p = 0,5.

7:

Xi	fi	Xifi	d2fi
0	4	0	100
2	1	2	9
3	4	12	16
4	6	24	6
5	9	45	0
6	7	42	7
7	5	35	20
10	4	40	100
	40	200	258

Para responder a las preguntas del problema, va a resultar útil construir previamente una tabla de frecuencias con los datos de la población, donde incluyamos las columnas intermedias que facilitan los cálculos:

a) $$\mu = \frac{\sum X_i}{n} = \frac{200}{40} = 5$$

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{258}{40}} = 2{,}54$$

$$E(\bar{X}) = \mu = 5$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{2{,}54}{\sqrt{9}}\sqrt{\frac{40-9}{40-1}} = 0{,}75$$

Luego, la media y la desviación tipo de la distribución muestral de medias son, respectivamente, 5 y 0,75.

b) Tanto mediante la tabla como observando el listado de datos, podemos contabilizar 16 datos con valores superiores a 5, lo que implica una proporción poblacional de valor 0,4. Luego:

$$E(p) = \pi = \frac{16}{40} = 0{,}4$$

$$\sigma_p = \sqrt{\frac{\pi\,(1-\pi)}{n}} = \sqrt{\frac{0{,}4\,(1-0{,}4)}{25}} = 0{,}098$$

Así pues, la media y la desviación tipo de la distribución muestral de proporciones son, respectivamente 0,4 y 0,098.

8:

El error tipo es siempre inferior a la desviación tipo de la población (pues surge de dividir esta entre la raíz cuadrada del tamaño de la muestra, es decir, una cantidad superior a 1). Por ello, la desviación tipo vale 5 y el error tipo 0,5. Introduciendo estos valores en la fórmula, despejamos el valor del tamaño de las muestras utilizadas:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \sqrt{n} = \frac{\sigma}{\sigma_{\bar{X}}} \quad \rightarrow \quad n = \left(\frac{\sigma}{\sigma_{\bar{X}}}\right)^2 = \left(\frac{5}{0,5}\right)^2 = 100$$

9:

$$\pi = \grave{E}(p) = 0,4$$

$$\sigma_p = \sqrt{\frac{\pi\,(1 - \pi)}{n}} = \sqrt{\frac{0,4\,(1 - 0,4)}{36}} = 0,082$$