

Problemas

Estimación estadística

Vicente Manzano-Arrondo, 2012,2013

Problemas de cálculo

Ejercicio 1 resuelto

Observamos en momentos al azar en una concurrida calle de la ciudad. Nos interesa registrar cuántas personas van cogidas de la mano en cada momento de observación dentro del campo visual acotado. Estos son los resultados anotados por el observador 1.

4	4	10	4	6	4	6	8	4
6	8	4	6	6	6	4	8	6
10	10	10	10	8	6	10	8	10
6	6	6	10	8	10	8	8	4
8	6	6	8	8	6	8	6	8

Queremos estimar cuántas personas, por término medio, van cogidas de la mano en esa calle de la ciudad, sin limitaciones de momentos de observación. Utiliza los riesgos del 1%, 2%, 3%, 4% y 5% para hacer la estimación.

Solución

La población está compuesta por todos los momentos posibles de observación, mientras que la muestra consiste en una selección al azar de esos momentos. El problema consiste en estimar por intervalo el valor de una media aritmética a partir de los datos de la muestra. Para ello, necesitamos primero calcular los estadísticos implicados:

X_i	f_i	$X_i f_i$	$d^2 f_i$
4	9	36	81,0
6	17	102	17,0
8	14	112	14,0
10	10	100	90,0
Σ	50	350	202

n	50
Media	7
D.t.	2,01

El cálculo del error de precisión se lleva a cabo mediante la expresión:

$$e_p = Z \frac{\sigma}{\sqrt{n}} \simeq Z \frac{S}{\sqrt{n-1}} = Z \frac{2,01}{\sqrt{49}}$$

en la que habrá que ir sustituyendo el valor de la distancia estandarizada que representa la confianza (o el riesgo) escogido para la estimación. Dado que la muestra es grande ($n=50$, mayor que 30), podemos suponer que la distribución muestral sigue una ley

normal y acudir a la tabla de la curva normal para traducir los riesgos indicados (o sus valores de confianza o seguridad complementarios) en valores Z. Al hacerlo, encontramos:

	0,01	0,02	0,03	0,04	0,05
Z	2,58	2,33	2,17	2,05	1,96

Con esta información, incluyendo cada valor de Z en la expresión de cálculo del error de precisión y expresando los sucesivos intervalos de confianza mediante sus cotas de valores mínimo y máximo esperados para el parámetro, entonces:

Tabla de estimaciones del 1% al 5%					
	0,01	0,02	0,03	0,04	0,05
Z	2,58	2,33	2,17	2,05	1,96
ep	0,74	0,67	0,62	0,59	0,56
mín.	6,26	6,33	6,38	6,41	6,44
máx.	7,74	7,67	7,62	7,59	7,56

Observa que, coherentemente, conforme aumenta el riesgo (disminuye la confianza o seguridad), disminuye la precisión del intervalo (aumenta el error de precisión) y, por tanto, el intervalo es más amplio.

Ejercicio 2 resuelto

Nos preocupa ahora estimar el porcentaje de ocasiones en las que este observador obtiene el resultado 10. Utiliza los mismos niveles de confianza que en el problema anterior. Además, con una seguridad del 95% ¿Cuántas veces observará a 10 personas cogidas de la mano de entre 200 momentos de observación?

Solución

Una forma instructiva de solucionarlo es tomar el conjunto original de datos y transformarlo en otro donde se observan únicamente dos valores: 1 (si el original era 10) y 0 (en el resto de los casos). El resto de las operaciones son las mismas. Este ejercicio permite afianzar la idea de que las proporciones son un caso particular de media aritmética:

0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	1	0	1
0	0	0	1	0	1	0	0	0
0	0	0	0	0	0	0	0	0

X_i	f_i	$X_i f_i$	$d^2 f_i$
0	40	0	1,6
1	10	10	6,4
Σ	50	10	8

n	50
Media	0,2
D.t.	0,4

Tabla de estimaciones del 1% al 5%

	0,01	0,02	0,03	0,04	0,05
Z	2,58	2,33	2,17	2,05	1,96
ep	0,15	0,13	0,12	0,12	0,11
mín.	0,05	0,07	0,08	0,08	0,09
máx.	0,35	0,33	0,32	0,32	0,31

Estos mismos resultados se obtienen utilizando las expresiones de cálculo específicas para proporciones (compruébalo):

$$e_p = Z \sqrt{\frac{\pi(1-\pi)}{n}} \simeq Z \sqrt{\frac{p(1-p)}{n}} = Z \sqrt{\frac{0,2(1-0,2)}{50}}$$

Para resolver el segundo interrogante, aplicamos las proporciones obtenidas para $Z=1,96$ al total de 200 intentos, con lo que:

Z	1,96		
ep	0,11		22,00
mín.	0,09	→	18,00
máx.	0,31	→	62,00

Es decir, con una confianza del 95% podemos considerar que el número de momentos de observación donde este observador anotará un 10 se encuentra entre 18 y 62, entre un total de 200 observaciones.

Ejercicio 3 resuelto

Queremos saber cuál es la opinión de los cirtenos (habitantes de la famosa ciudad de Cirte) respecto a las últimas declaraciones de su concejal de urbanismo sobre cómo utilizar los terrenos colindantes al parque central. Para ello, hemos pedido a una muestra aleatoria que puntúen las declaraciones de 0 a 10 como si las calificaran con una nota. Los resultados son

3, 1, 7, 2, 3, 2, 7, 3, 4, 2, 3, 2, 1, 8, 3, 1, 4, 1, 3, 1,
4, 2, 3, 2, 2, 3, 1, 4, 8, 1, 3, 2, 1, 5, 3, 1, 2, 7, 1, 4

Con estos datos, responde a la inquietud del problema mediante una confianza del 95%.

Solución

Los datos se han recogido en una muestra, mientras que el interés es conocer qué opina la población. Por tanto, hay que realizar una estimación por intervalo. Es decir, respondemos a la necesidad de dar forma concreta a la expresión:

$$\mu \in \{\bar{X} \pm e_p\}_{conf}$$

Para responder a ello, necesitamos calcular la media de la muestra y el error de precisión. Vamos a generar una tabla de frecuencias básica que nos permita realizar los cálculos de forma más sencilla (aprovechando las frecuencias):

X_i	f_i	$X_i f_i$	$d^2 f_i$
1	10	10	40
2	9	18	9
3	10	30	0
4	5	20	5
5	1	5	4
7	3	21	48
8	2	16	50
40		120	156

$$\bar{X} = \frac{\sum X_i f_i}{n} = \frac{120}{40} = 3$$

Ya tenemos el valor de la media aritmética. Para calcular el error de precisión:

$$e_p = Z \cdot \frac{\sigma}{\sqrt{n}}$$

Vamos a realizar una estimación habitual con una confianza del 95%. Dado que la muestra es grande ($n = 39 \geq 30$), podemos suponer que la distribución muestral de medias es normal. En la curva normal, un área centrada del 95% se corresponde con una distancia estandarizada de 1,96. Luego, $Z=1,96$.

Por otro lado, no contamos con el valor de la desviación tipo poblacional, así que acudimos a la cuasidesviación tipo de la muestra:

$$\hat{S} = \sqrt{\frac{\sum (X_i - \bar{X})^2 f_i}{n-1}} = \sqrt{\frac{156}{39}} = 2$$

Con esta información podemos completar el problema:

$$e_p = Z \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{2}{\sqrt{40}} = 0,62$$

Por tanto:

$$\mu \in \{\bar{X} \pm e_p\}_{conf} \rightarrow \mu \in \{3 \pm 0,62\}_{0,95} \rightarrow \mu \in \{2,$$

En otras palabras: con una confianza del 95% afirmamos que la nota que los cirtenos otorgan a las últimas declaraciones del concejal de urbanismo se encuentra entre 2,38 y 3,62 lo que implica un suspenso en cualquier caso.

Ejercicios propuestos

1. Hemos preguntado a una muestra aleatoria de personas de la tercera edad por su valoración en torno a cómo han cambiado las relaciones entre la generación de los mayores y de los jóvenes. Tras analizar las respuestas, una de las variables generadas es si cada persona entrevistada considera que las relaciones intergeneracionales van hoy mejor, igual o peor que antes. Los datos correspondientes a esa variable son

mejor, igual, peor, peor, igual, igual, mejor, peor, mejor,
peor,

igual, peor, igual, peor, peor, peor, igual, mejor, igual,
peor,
igual, peor, mejor, mejor, mejor, peor, peor, igual, peor,
mejor

¿Cuántas personas en la población de la tercera edad consideran que las relaciones intergeneracionales son hoy diferentes a las que había antes? Utiliza un nivel de confianza del 97%.

2. ¿Cuál es el tamaño de la muestra utilizado en una investigación donde se han registrado 50 personas casadas y se ha estimado que el porcentaje de personas casadas en la población se encuentra entre 20% y 30%?

3. En la estimación de una media aritmética se ha generado el intervalo de confianza $\{12; 16\}$. Si la media de la muestra hubiera sido el doble, mientras que el error de precisión fuera la mitad, ¿cuál sería el nuevo intervalo de confianza?

4. Con los datos de la siguiente distribución, estima la media aritmética de la población (de la que sabemos que es normal), el porcentaje de datos con un valor superior a 4 y en cuántas ocasiones se obtendría un 2 de un total de 1500. Para responder a estas preguntas, utiliza un riesgo del 3%.

X_i	f_i
2	4
3	9
4	1
6	3
7	2
9	1
Σ	20