

Curso de Estadística Aplicada a las Ciencias Sociales

Tema 11. Estimación de una media

(Cap. 21 del libro)

1

Tema 11. Estimación de una media

Introducción

1. Distribución de la media en el muestreo
2. La media muestral es centrada
3. El error típico de la estimación
4. Intervalos de confianza
5. Estimación para poblaciones pequeñas

Resumen

Ejercicios

Introducción

- Tema anterior: tenemos una muestra. Calculamos una proporción primero en la muestra y luego en la población.
- Ahora: lo que queremos inferir es la media de una variable en la población
- Ejemplos: gasto, edad, puntuación de un líder, número de hijos de cada familia...

Introducción (2)

- Procedimiento (como antes):
 - ◆ tomar una muestra aleatoria simple (lo vimos en tema 8);
 - ◆ Calcular la media que tiene la variable en la muestra (estimador)
 - ◆ Calcular el valor de la media en la población, para lo que tenemos que tener una idea sobre la “precisión” del estimador

Introducción (3)

- El mismo estimador puede tener niveles de precisión muy diferentes
- Ejemplo, estimación de tres variables de gasto con una muestra de 5 estudiantes

	Est 1	Est 2	Est 3	Est 4	Est 5	Media muestral
Desplaz.	2.000	2.200	2.000	1.800	2.000	2.000
Ocio	1.000	3.500	4.000	1.500	0	2.000
Libros	0	0	0	0	10.000	2.000

Introducción (4)

- La media muestral es igual en los 3 casos
- Es fácil deducir que la estimación no es igualmente precisa en los tres
- La precisión dependerá de
 - ◆ La variabilidad de los datos en la población: más variabilidad, menos precisa la estimación
 - ◆ El tamaño muestral: más grande la muestra, más precisa la estimación

1. Distribución de la media en el muestreo

- Como en lección anterior, vamos a imaginar situación ficticia con una variable con distribución conocida
- Ejemplo: variable número de hijos en las familias
- Distribución en la población es:

■ La media es 1,95

■ La desviación típica es

1,359

Número	Proporción
0	0,15
1	0,25
2	0,30
3	0,15
4	0,10
5	0,05

1. Distribución de la media en el muestreo

- Partimos de una población muy grande
 - ◆ Representamos la población como una urna con bolas. Una bola por familia, con valor de 0 a 5 (urna A), en proporción correspondiente
 - ◆ Sacamos muestra con 10 elementos. Calculamos media de la muestra. Ejemplo 1,6. Escribimos la media en una papeleta: otra urna distinta (urna B)
 - ◆ Devolvemos bolas a urna A y seguimos sacando muestras de 10 elementos

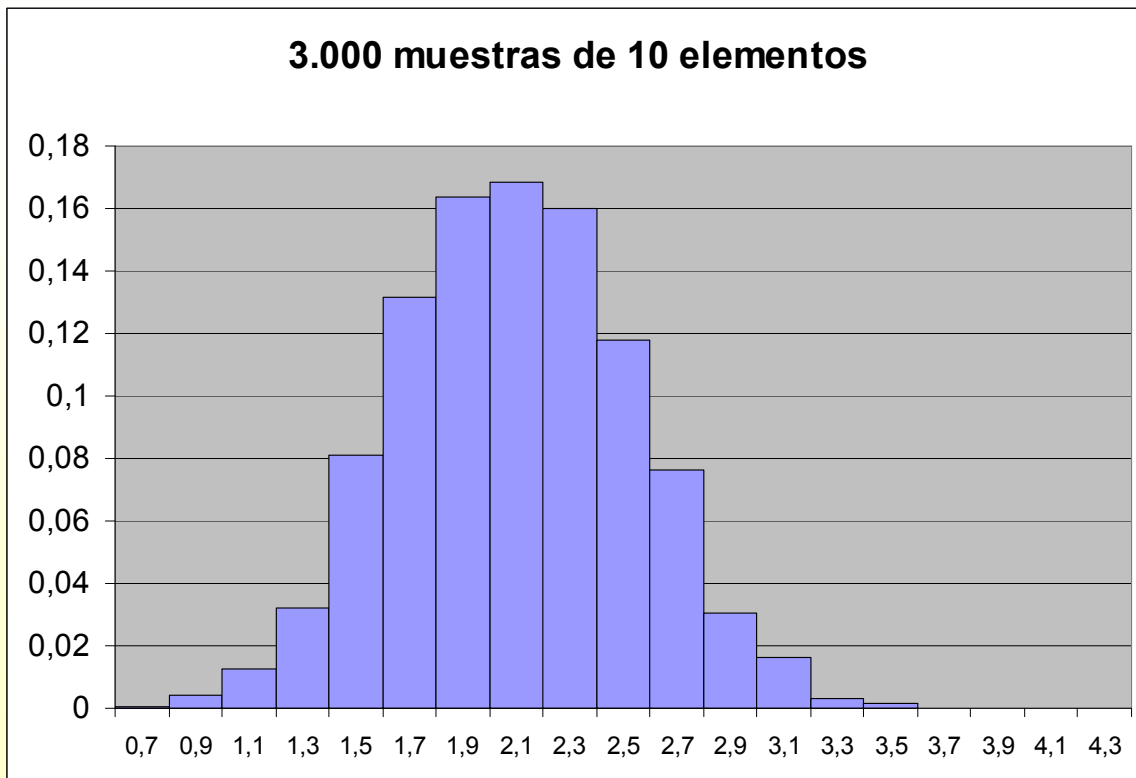
1. Distribución de la media en el muestreo

- Al final, urna B, llena de papeletas que representan las medias de cada una de las muestras
- ¿Qué valores tendríamos en urna B?
- La mayor parte de los valores: en torno a la media en la población (1,95)
- Si las muestras han sido pequeñas (10, 15, 25): habrá algunas muestras con valores alejados de 1,95: 0,5; 1; 2,5; 3
- Si muestras más grandes (más de 30 elementos): casi todas las muestras con poca variación respecto a 1,95

1. Distribución de la media en el muestreo

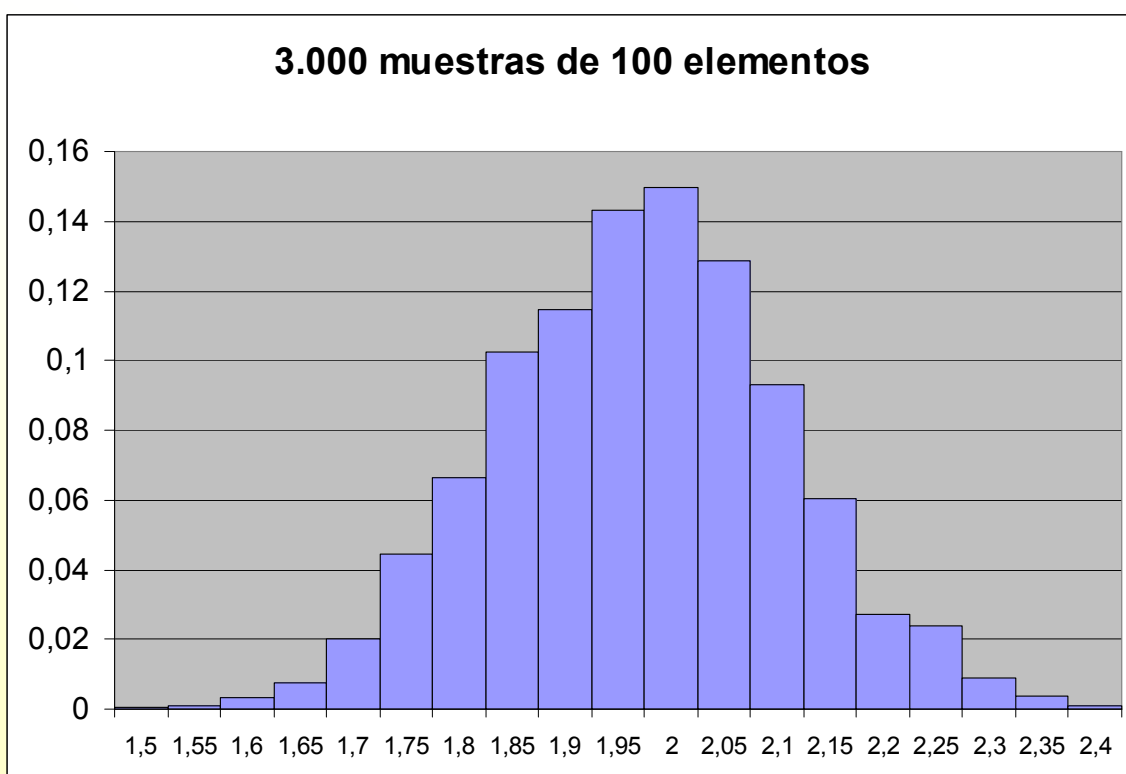
- La forma del histograma de las medias muestrales es la de una distribución normal (propiedades conocidas)
- Como en lección anterior: cuanto mayor sea la muestra, más se acerca la forma de la distribución de las medias muestrales a la distribución normal

1. Distribución de la media en el muestreo



11

1. Distribución de la media en el muestreo



12

2. La media muestral es centrada

- La variable media muestral estimada (la variable cuyos valores están en la urna B) se puede describir como cualquier variable
- Se puede demostrar que la media de la urna B = media de la urna A
- Lógico: media urna B = media valores urna A contados muchas veces

3. El error típico

- Recuperamos el concepto visto en el tema 10
- El error típico de estimación es la desviación típica de la variable aleatoria constituida por los valores de la media de todas las muestras que potencialmente podríamos sacar
- Es una medida de la dispersión de los estimadores respecto de su media

3. El error típico

- La fórmula es la misma que veíamos en el tema anterior:

$$ET = \frac{s_x}{\sqrt{n}}$$

- Es decir, como entonces:
 - ◆ el error típico es siempre menor que la desviación típica en la población
 - ◆ Y será menor cuanto mayor sea la muestra
- Pero tenemos el mismo problema que en el tema anterior: si la fórmula para calcular ET incluye la desviación típica en la población (s_x), estamos en un círculo vicioso

3. El error típico

- Solución: se puede demostrar (aquí no), que la desviación típica de la población se aproxima bastante al valor modificado de la desviación típica de la muestra, según la siguiente fórmula

$$\hat{s}_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

- Cuando el tamaño muestral es grande, la diferencia entre usar en el denominador n o $n-1$ no es demasiado importante
- En resumen, calculamos ET con la fórmula

$$ET = \frac{\hat{s}_x}{\sqrt{n}} = \frac{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}}{\sqrt{n}}$$

3. El error típico

- Ejemplo: muestra de 1.000 encuestados. Preguntamos cuántas veces han ido al cine en el último mes. La media de la muestra es 1,3 y la desviación típica (corregida con $n-1$) es 2,1.

$$ET = \frac{\hat{s}_x}{\sqrt{n}} = \frac{2,1}{\sqrt{1000}} = \frac{2,1}{31,623} = 0,0664$$

4. Intervalos de confianza

- Como en el tema anterior: podemos usar el ET para estimar un rango de valores, o intervalo dentro del que debe de estar el valor medio en la población
- Ejemplo anterior: muestra de 1.000 personas, media de 1,3 y ET de 0,0664
- Cuando la muestra es grande (más de 30 elementos), la variable aleatoria formada por las medias de las muestras tiene también una distribución normal
 - ◆ Su media es la media de la población
 - ◆ Su desviación típica es ET
 - ◆ Sabemos exactamente qué proporción de los casos está a diferentes distancias de la media, medidas en ET (valores z)

4. Intervalos de confianza

- Recordamos que las las distribuciones normales tienen esta propiedad:
 - ◆ 68,3% de los valores a menos de 1dt de la media
 - ◆ 95,5% de los valores a menos de 2dt de la media
 - ◆ 99,7% de los valores a menos de 3 dt de la media

4. Intervalos de confianza

- Por esa razón, cuando tenemos una muestra, en la que hemos calculado una media y un ET (que es la desviación típica de las muestras)
- Podemos decir que la media en la población estará con un
 - ◆ 68,3% de confianza (o de probabilidad) en el intervalo: $\text{estimador} \pm \text{ET}$
 - ◆ 95,5% de confianza (o de probabilidad) en el intervalo: $\text{estimador} \pm 2 \cdot \text{ET}$
 - ◆ 99,7% de confianza (o de probabilidad) en el intervalo: $\text{estimador} \pm 3 \cdot \text{ET}$
- El intervalo de confianza es rango de valores que incluye, con una determinada probabilidad, el valor del parámetro en la población

4. Intervalos de confianza

- En el ejemplo que poníamos (media de 1,3 y ET de 0,0664)
- Podemos decir que el verdadero parámetro, con un
 - ◆ 68,3% de confianza está en el intervalo de $1,3 \pm 0,0664$ (es decir, entre 1,2336 y 1,3664)
 - ◆ 95,5% de confianza está en el intervalo de $1,3 \pm 2 \cdot 0,0664$ (es decir, entre 1,1672 y 1,4328)
 - ◆ 99,7% de confianza está en el intervalo de $1,3 \pm 3 \cdot 0,0664$ (es decir, entre 1,1008 y 1,4992)
- (Precisión exagerada)

5. Estimación en poblaciones pequeñas

- Al igual que sucedía con la estimación de proporciones, cuando el tamaño de la población es pequeño comparado con el tamaño de la muestra y se hace un muestreo sin reposición (por ejemplo, se entrevista a 60 alumnos de clase de 90): la fórmula del ET cambia
- La media de las muestras sigue siendo la media de la población

5. Estimación en poblaciones pequeñas

- Pero: el error típico (la desviación típica de las medias en las muestras respecto a la media de la población) es menor que el caso de población muy grande
- Fórmula: si la fracción de muestreo es $f = n/N$

$$ET = \frac{\hat{s}_x}{\sqrt{n}} \cdot \sqrt{1-f}$$

5. Estimación en poblaciones pequeñas

- Ejemplo. Si el estudio anterior sobre entradas de cine estuviera hecho de una población de 5.000 personas ($f = 1.000/5.000=0,2$)

$$ET = \frac{\hat{s}_x}{\sqrt{n}} \cdot \sqrt{1-f} = \frac{2,1}{\sqrt{1000}} \cdot \sqrt{1-0,2} =$$

$$\frac{2,1}{31,623} \cdot \sqrt{0,8} = 0,0664 \cdot 0,894 = 0,05936$$

- Correspondiente, los intervalos de confianza serán también más pequeños

Resumen

- Distribución en el muestreo
- Media de las medias de las muestras= media población
- Desviación típica de las medias muestrales= error típico
- Fórmula del ET $ET = \frac{\hat{s}_x}{\sqrt{n}}$
- Intervalos de confianza: para el 95,5% de confianza, media muestral ± 2 ET
- ET corregido en poblaciones pequeñas

Ejercicios recomendados

- Del manual:
 - ◆ 21.4 (El intervalo con confianza del 95% es la media $\pm 1,96$ ET)
 - ◆ 21.6
- De los exámenes:
 - ◆ Feb04: 12
 - ◆ Jul04: 11
 - ◆ Feb06: 8
 - ◆ Jul06: 9
 - ◆ Jul07: 8
 - ◆ Ene08, Jun08: 9