

# Estadística aplicada a la comunicación

---

## Tema 6: Análisis de datos cuantitativos II: introducción a la estadística inferencial

OpenCourseWare UPV/EHU

Unai Martín Roncero

Departamento de Sociología 2

unai.martin@ehu.eus

OCW  
OpenCourseWare



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

A partir de los datos de una encuesta realizada sobre la valoración de Telebilbao y de otras cadenas tenemos:

1. ¿Cómo es la valoración de Telebilbao? ¿Y la de la competencia de canales locales?

Valoración media de diferentes canales

Canal	
Telebilbao	4,04
Tele7	3,16
HamaikaTB	3,18

2. ¿Hay diferencias en la valoración entre hombres y mujeres?

Canal	Hombres	Mujeres
Telebilbao	4,00	4,06
Tele7	3,10	3,22
HamaikaTB	3,20	3,16

Sin embargo, la pregunta de investigación a la que se quería responder era:

¿Cómo es la valoración de Telebilbao por parte de la población residente en municipios de Bizkaia?

Pero nuestros datos nos dicen:

La valoración de Telebilbao por parte de la población encuestada

¿Qué podemos hacer?

Si hubiéramos preguntado a otras personas, ¿tendríamos los mismos resultados?

# Tema 6: Introducción a la estadística inferencial

---

La **inferencia estadística**, con base en la teoría de la probabilidad y en sus distribuciones, nos ayuda a **extrapolar** datos observados en una muestra al conjunto de la población con un determinado margen de error (que es conocido)

La **estadística descriptiva** es útil para el **resumen** y la **descripción** de los datos, sin embargo, en ocasiones no disponemos de esos datos y no podemos acceder a los de toda la población objeto de estudio

# ÍNDICE

1. Introducción: ¿Qué es la estadística inferencial y qué nos aporta? ¿Cuáles son los pasos?
2. ¿Qué es el nivel de confianza y por qué podemos calcularlo?
3. Estimación de medias y sus intervalos de confianza
4. Diferencias de medias ¿Cuándo son extrapolables a la población?
5. Asociación de variables cualitativas ¿Cuándo son extrapolables a la población?
6. ¿Qué elementos están relacionados con la amplitud del intervalo?
7. Significación estadística vs significación social

# Tema 6: Introducción a la estadística inferencial

---

**Parámetro:** medida que queremos obtener en la población a estudio y que generalmente no conocemos: porcentaje de parados, número de jóvenes, media de ingresos...

**Estadístico:** medida calculada sobre los datos de una muestra: valoración media de Telebilbao en nuestro trabajo de clase...

**Estimador:** cuando el estadístico se utiliza para aproximarnos al estimador, por ejemplo utilizamos la “valoración media del teatro por los encuestados en nuestro trabajo de clase” para aproximarnos a la valoración media del teatro de los/as alumnos/as de las carreras de comunicación.

# Tema 6: Introducción a la estadística inferencial

---

El “verdadero valor” sólo lo conocemos si tenemos los datos para toda la población, sin embargo, con una muestra nos podemos acercar a ese valor: **ESTIMACIÓN**

-Estimación puntual: por ejemplo valor de la media de nuestra muestra (si hacemos otra muestra probablemente varíe)

-Estimación por intervalos: calculamos un intervalo (dos valores) que contendrán el verdadero valor con una **probabilidad determinada**.

-¿Cómo es posible?....

# Tema 6: Introducción a la estadística inferencial

De una población de 10 individuos, podemos seleccionar una muestra de 5 individuos y obtendríamos una media y una desviación típica. Si seleccionamos otra muestra, la media y la desviación podría ser distinta, y así hasta 252 muestras...

Muestra 1=4,2,3,3,6

Media=3,6

Desviación=1,36

Muestra 2= 6,4,4,5,2

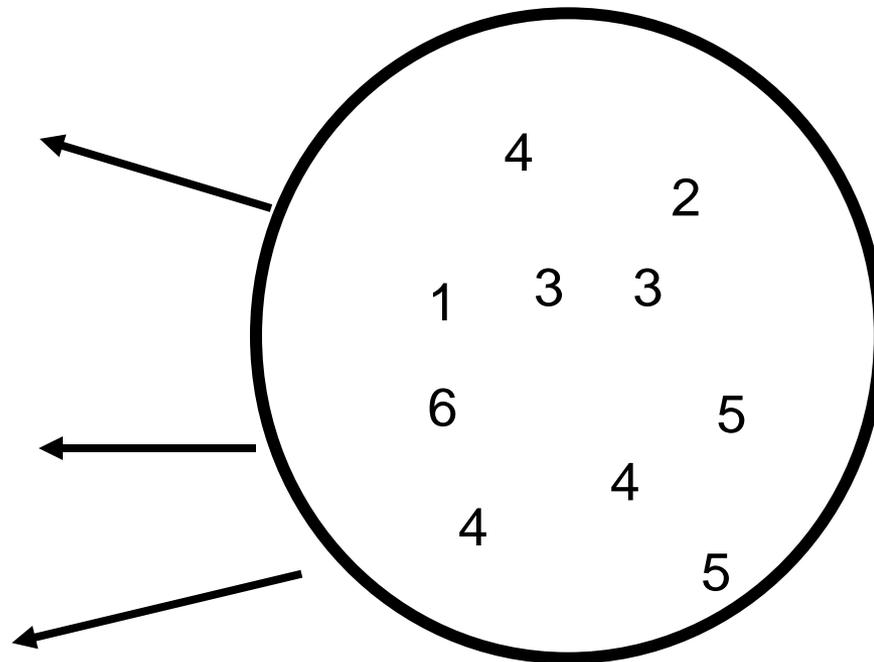
Media=4,2

Desviación=1,33

Muestra 3= 4,5,3,3,2

Media=3,4

Desviación=1,01



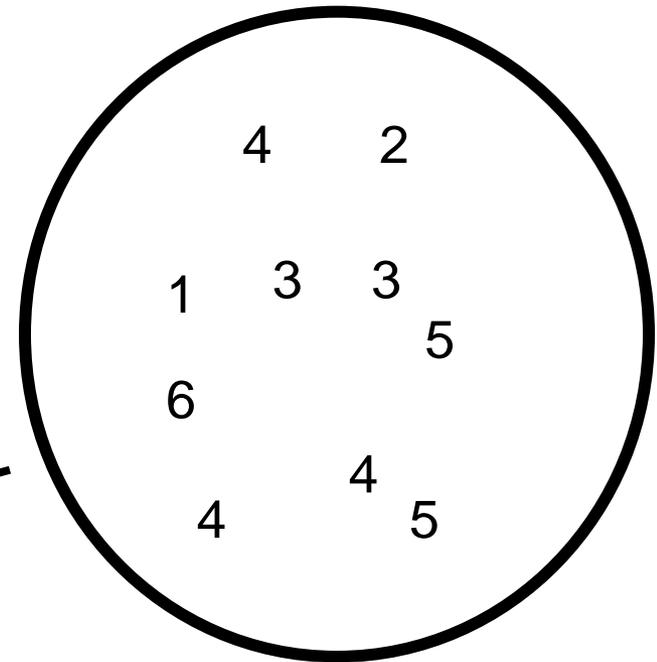
Media=3,7

Desviación típica=1,42

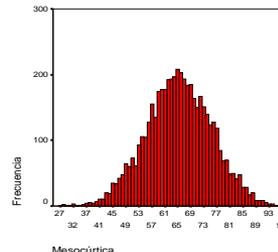
N=10

# Tema 6: Introducción a la estadística inferencial

Si seleccionamos las 252 muestras posibles, obtendríamos 252 medias. Esas medias siguen una distribución normal y su media (la media de las medias) coincide con la media de la población.



Media 1	3,6
Media 2	4,2
Media 3	3,2
...	
Media 252	3,8
<b>Media</b>	<b>3,7</b>



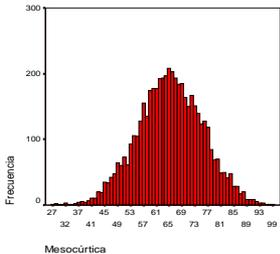
Media=3,7

Desviación típica=1,42

N=10

# Tema 6: Introducción a la estadística inferencial

Esta propiedad nos permite calcular a partir de una muestra un intervalo que con una probabilidad determinada contenga el verdadero valor de la población, ya que también podemos calcular la desviación.



$$EE = \frac{DE}{\sqrt{n}}$$

$$IC95\% = \bar{x} \pm z * EE$$

$$IC95\% = 3,36 \pm 1,96 * 0,43(2,52;4,21)$$

La media de la población estará entre 2,52 y 4,21 con una probabilidad del 95%

# Tema 6: Introducción a la estadística inferencial

¿Cómo era la valoración de Telebilbao, y de la competencia?

Valoración media de las cadenas

	95% Intervalo de confianza		
	Media	Inferior	Superior
Telebilbao	4,04	3,87	4,21
Tele7	3,17	2,99	3,34
HamaikaTB	3,18	2,99	3,38

# Tema 6: Introducción a la estadística inferencial

## Valoración de Sálvame

Media	95% Intervalo de confianza	
	Inferior	Superior
3,15	2,97	3,34

¿Son diferentes las valoraciones medias de Sálvame entre hombres y mujeres?

	Media
Hombres	2,07
Mujeres	4,13
Total	3,15

	Media	95% Intervalo de confianza	
		Inferior	Superior
Hombres	2,07	1,84	2,29
Mujeres	4,13	3,87	4,39

# Tema 6: Introducción a la estadística inferencial

¿Son diferentes las valoraciones medias de Sálvame entre hombres y mujeres?

Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
		Inferior	Superior
-2,066	0,178	-2,416	-1,717

La diferencia de valoración media de Sálvame entre hombres y mujeres es de 2 puntos en nuestra muestra. Las mujeres valoran el programa de media con dos puntos más que los hombres

*¿Es posible extrapolar esta diferencia al conjunto de la población?*

Calculamos el intervalo de confianza y vemos que el “valor real” de la diferencia entre hombres y mujeres va a ser en todo caso distinto de cero (IC95%: -2,416; -1,717), es decir, siempre se van a encontrar diferencias en la valoración.

Por tanto, la valoración de Sálvame por parte de las mujeres es estadísticamente significativa mayor que la de los hombres (extrapolable a la población)

# Tema 6: Introducción a la estadística inferencial

## Valoración de Telebilbao

Media	95% Intervalo de confianza	
	Inferior	Superior
4,04	3,87	4,21

¿Son diferentes las valoraciones medias de Telebilbao entre hombres y mujeres?

	Media
Hombres	4,00
Mujeres	4,06
Total	4,04

	Media	95% Intervalo de confianza	
		Inferior	Superior
Hombres	4,00	3,76	4,24
Mujeres	4,06	3,82	4,31

# Tema 6: Introducción a la estadística inferencial

¿Son diferentes las valoraciones medias de Telebilbao entre hombres y mujeres?

Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
		Inferior	Superior
-0,060	0,174	-0,401	0,280

La diferencia de valoración media de Sálvame entre hombres y mujeres es de 0,06 puntos en nuestra muestra. Las mujeres valoran el programa de media con 0,06 puntos más que los hombres

*¿Es posible extrapolar esta diferencia al conjunto de la población?*

Calculamos el intervalo de confianza y vemos que el “valor real” de la diferencia entre hombres y mujeres puede ser igual a cero (IC95%: -0,401; 0,280), es decir, puede que no existan diferencias en la valoración.

Por tanto, no existen diferencias significativas en la valoración de TeleBilbao entre hombres y mujeres (no extrapolable a la población)

# Tema 6: Introducción a la estadística inferencial

---

CUIDADO:

La amplitud del intervalo depende de:

- Tamaño de la muestra
- Variabilidad de la variable en la población que estudiemos
- Nivel de confianza elegido (95%, 97%, 99%...)

Por eso, el tamaño de la muestra influye en que unas diferencias sean o no significativas

- ✓ Muestras pequeñas
- ✓ Muestras grandes: más fácil encontrar significación estadística. Además debemos preguntarnos si los resultados son significativos socialmente

# Tema 6: Introducción a la estadística inferencial

## Asociación entre variables cualitativas: ¿es significativa?

Hipótesis de partida:

En la **población** no existe asociación entre variables. Encontraríamos la tabla de valores esperados

Tabla de contingencia Eje izquierdo-derecho - ¿Has hecho la confirmación?

		¿Has hecho la confirmación?	
		No	Si
Eje izquierdo-derecho	Extremo izquierdo	Recuento 42	7
	Frecuencia esperada % dentro de ¿Has hecho la confirmación?	39.2 18.4%	9.9 12.3%
	Izquierdo	Recuento 107	13
		1.0 3%	24 31.4%
		1.0 3%	6 7.9%
		2	1.2
		4.8 1.8%	3.5%
No se significó		Recuento 17	5
Frecuencia esperada % dentro de ¿Has hecho la confirmación?		17.6 7.3%	4.4 8.8%
Total		Recuento 228	57
Frecuencia esperada % dentro de ¿Has hecho la confirmación?		228.0 100.0%	57.0 100.0%

**Tabla esperada**

Obtenemos una muestra de nuestra población (real) ¿Qué puede pasar?:

a) Obtener una tabla que es poco probable que hubiéramos obtenido según nuestra hipótesis

b) Obtener una tabla que es probable obtener según nuestra hipótesis

**Rechazo la hipótesis: existe asociación**

**No rechazo la hipótesis**

**¿Cuándo es poco probable?**

# Tema 6: Introducción a la estadística inferencial

**Vamos a verlo con un ejemplo:**

**¿Qué elementos se asocian con la ubicación ideológica en el eje izquierda-derecha?**

**- ¿El sexo? ¿La creencia religiosa?**

**Eje izquierda-derecha**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Extrema izquierda	49	16,1	17,2	17,2
	Izquierda	120	39,3	42,1	59,3
	Centro	67	22,0	23,5	82,8
	Derecha	21	6,9	7,4	90,2
	Extrema derecha	6	2,0	2,1	92,3
	No sé significado	22	7,2	7,7	100,0
	Total	285	93,4	100,0	
Perdidos	Sistema	20	6,6		
Total		305	100,0		

# Tema 6: Introducción a la estadística inferencial

Calculamos la tabla de contingencia para ver si las dos variables se encuentran relacionadas

Tabla de contingencia Eje Izquierda-derecha \* ¿Has hecho la confirmación?

			¿Has hecho la confirmación?	
			No	Sí
Eje izquierda-derecha	Extrema izquierda	Recuento	42	7
		Frecuencia esperada	39,2	9,8
		% dentro de ¿Has hecho la confirmación?	18,4%	12,3%
	Izquierda	Recuento	107	13
		Frecuencia esperada	96,0	24,0
		% dentro de ¿Has hecho la confirmación?	46,9%	22,8%
	Centro	Recuento	43	24
		Frecuencia esperada	53,6	13,4
		% dentro de ¿Has hecho la confirmación?	18,9%	42,1%
	Derecha	Recuento	15	6
		Frecuencia esperada	16,8	4,2
		% dentro de ¿Has hecho la confirmación?	6,6%	10,5%
	Extrema derecha	Recuento	4	2
		Frecuencia esperada	4,8	1,2
		% dentro de ¿Has hecho la confirmación?	1,8%	3,5%
	No sé significado	Recuento	17	5
		Frecuencia esperada	17,6	4,4
		% dentro de ¿Has hecho la confirmación?	7,5%	8,8%
	Total	Recuento	228	57
		Frecuencia esperada	228,0	57,0
		% dentro de ¿Has hecho la confirmación?	100,0%	100,0%

# Tema 6: Introducción a la estadística inferencial

A través de la prueba de Chi-cuadrado comprobamos la significación de la asociación entre las dos variables

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	19,517 <sup>a</sup>	5	,002
Razón de verosimilitudes	18,947	5	,002
N de casos válidos	285		

Si es menor de 0,05 rechazamos la hipótesis.

a. 4 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,20.

En este caso la relación encontrada en nuestra muestra entre la posición ideológica y las creencias religiosas puede ser extrapolable al conjunto de la población

# Tema 6: Introducción a la estadística inferencial

Tabla de contingencia Eje izquierda-derecha \* Sexo

			Sexo		Total
			Hombre	Mujer	
Eje izquierda-derecha	Extrema izquierda	Recuento	22	27	49
		Frecuencia esperada	18,1	30,9	49,0
		% dentro de Sexo	21,0%	15,0%	17,2%
	Izquierda	Recuento	40	80	120
		Frecuencia esperada	44,2	75,8	120,0
		% dentro de Sexo	38,1%	44,4%	42,1%
	Centro	Recuento	20	47	67
		Frecuencia esperada	24,7	42,3	67,0
		% dentro de Sexo	19,0%	26,1%	23,5%
	Derecha	Recuento	10	11	21
		Frecuencia esperada	7,7	13,3	21,0
		% dentro de Sexo	9,5%	6,1%	7,4%
	Extrema derecha	Recuento	4	2	6
		Frecuencia esperada	2,2	3,8	6,0
		% dentro de Sexo	3,8%	1,1%	2,1%
	No sé significado	Recuento	9	13	22
		Frecuencia esperada	8,1	13,9	22,0
		% dentro de Sexo	8,6%	7,2%	7,7%
Total	Recuento	105	180	285	
	Frecuencia esperada	105,0	180,0	285,0	
	% dentro de Sexo	100,0%	100,0%	100,0%	

# Tema 6: Introducción a la estadística inferencial

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	6,907 <sup>a</sup>	5	,228
Razón de verosimilitudes	6,786	5	,237
N de casos válidos	285		

Si es mayor de 0,05  
no podemos  
rechazar la hipótesis

a. 2 casillas (16,7%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,21.

En este caso la relación encontrada en nuestra muestra entre la posición ideológica y el sexo no puede ser extrapolable al conjunto de la población