



Tema 5: Introducción a la inferencia estadística

1. Planteamiento y objetivos
2. Estadísticos y distribución muestral
3. Estimadores puntuales
4. Estimadores por intervalos
5. Contrastes de hipótesis

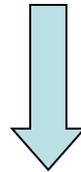
Lecturas recomendadas:

- Capítulos 20 y 21 del libro de Peña y Romo (1997)

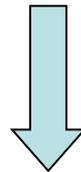


5.1 Planteamientos y objetivos

Estadística Descriptiva: la edad media de una muestra de 20 votantes del PP es de 55 con desviación típica 5.



Modelo Probabilístico: La edad de un votante del PP sigue una distribución normal $N(\mu, \sigma^2)$



Inferencia:

- Predecimos que $\mu = 55$.
- Rechazamos la posibilidad de que $\mu < 50$.



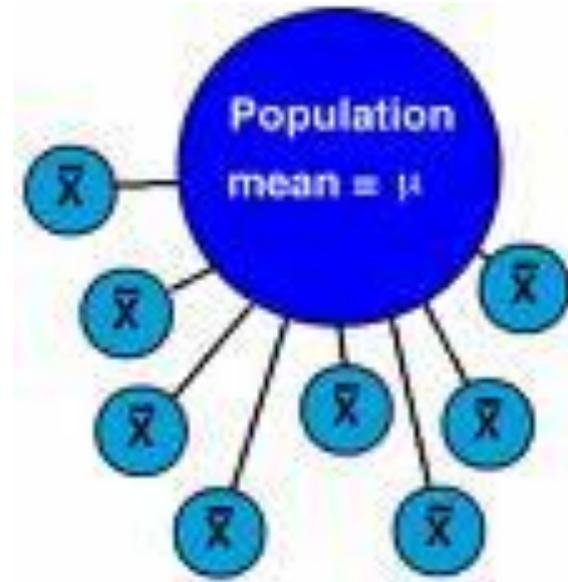
5.2 Estadísticos y distribución muestral

Distintas muestras tienen distintas medias. Antes de sacar la muestra, la media es una **variable**.

La media y varianza de la media son

$$E[\bar{X}] = \mu \quad V[\bar{X}] = \sigma^2/N$$

Si N es suficientemente grande, la distribución de la media es **normal**



Para ver cómo varia la media de distintas muestras:

<http://www.est.uc3m.es/amalonso/esp/Ejemplo%20Muestreo.xls>

<http://www.stat.tamu.edu/~west/ph/sampledist.html>



5.3 Estimadores puntuales

Usamos \bar{X} como **estimador** de la media poblacional μ .

Dada una muestra, \bar{x} es una **estimación** de μ .

Buenas propiedades estadísticas:

Insensibilidad,

Eficiencia

Máxima verosimilitud, etc.



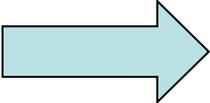
Igualmente S^2 es un **estimador** (razonable) de σ^2 .

Dada una muestra, s^2 es una **estimación** de σ^2 .



5.4 Estimadores por intervalos

Queremos calcular un intervalo donde estamos bastante seguros de que esté μ .

Intervalo ancho		muy impreciso
Intervalo pequeño		más probabilidad de cometer un error.

Método probabilístico de construcción de un intervalo de confianza:

- Elegir un **nivel de confianza**, por ejemplo 95% (ó 90% ó 99%)
- Elegir variables $L(X_1, \dots, X_N)$, $U(X_1, \dots, X_N)$ tales que $P(L < \mu < U) = 95\%$
- Dados los datos de la muestra, un intervalo de 95% de confianza es

$$(L(x_1, \dots, x_N), U(x_1, \dots, x_N))$$



Interpretación

Si construimos muchos intervalos con el mismo método y el mismo nivel de confianza de 95%, esperamos que el 95% de estos intervalos contendrán el parámetro que queremos estimar.



Si hemos construido un intervalo de 95% de confianza, **no es correcto** afirmar que la probabilidad de que μ esté dentro del intervalo es de 95%.

¿Y cuál es esa probabilidad?

Para ilustrar la proporción de intervalos que contienen al valor verdadero:

<http://www.est.uc3m.es/amalonso/esp/intervalosconfianza.xls>

http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/index.html



Un intervalo de 95% de confianza para la media de una población normal (varianza conocida)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \sim N(0, 1)$$

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} < 1.96\right) = 0.95$$

$$P\left(\bar{X} - 1.96\sigma/\sqrt{N} < \mu < \bar{X} + 1.96\sigma/\sqrt{N}\right) = 0.95$$



Dada una muestra, x_1, \dots, x_N , un intervalo de 95% de confianza para μ es

$$\left(\bar{x} - 1.96\sigma/\sqrt{N}, \bar{x} + 1.96\sigma/\sqrt{N}\right)$$

¿De dónde viene el 1,96?

¿Cómo sería un intervalo de 90% (y del 99%) de confianza?



Ejemplos

En una muestra de 20 extremeños, su sueldo medio era de € 2000 mensuales. Suponiendo que la desviación típica de los sueldos en Extremadura es de € 500, hallar un intervalo de 95% de confianza para el sueldo medio en Extremadura.

En una muestra de 10 estudiantes de ciencias políticas, la altura media era de 170cm. Suponiendo que la desviación típica de las alturas es de 5cm, hallar un intervalo de 99% de confianza para la altura media.



¿Hay una manera más rápida de hacer este cálculo?

Computación en Excel

	A	B	C	D
1	n	20		Datos
2	media	2000		
3	desviación	500		
4				
5	alpha	0,05		Computación de z
6	alpha/2	0,025		
7	1-alpha/2	0,975		
8	z	1,96		=DISTR.NORM.ESTAND.INV(0,975)
9				
10	z*sigma/raíz(n)	219,13		=B8*B3/RAIZ(B1)
11				
12	Intervalo	1780,9	2219,13	
13		=B2-B10	=B2+B10	
14				



En Excel 2010 se utiliza INTERVALO.CONFIANZA.NORM()

Argumentos de función

INTERVALO.CONFIANZA

Alfa	0,05		= 0,05
Desv_estándar	500		= 500
Tamaño	20		= 20

= 219,1306351

Devuelve el intervalo de confianza para la media de una población.

Tamaño es el tamaño de la muestra.

Resultado de la fórmula = 219,1306351

[Ayuda sobre esta función](#)

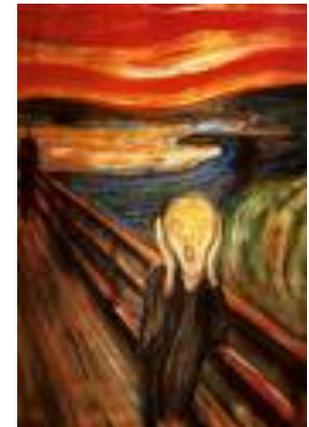
Aceptar Cancelar

Sólo tenemos que restar y sumar el valor de la media para calcular el intervalo.



Un intervalo de 95% de confianza para una proporción

$$\begin{aligned}X &\sim \text{Bi}(N, p) \Rightarrow \\X &\approx N(Np, Np(1-p)) \\ \hat{p} = \frac{X}{N} &\approx N\left(p, \frac{p(1-p)}{N}\right)\end{aligned}$$



Dada una muestra de tamaño N con proporción muestral \hat{p} , un intervalo de 95% de confianza para p es

$$\left(\hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \right)$$



Ejemplos

En una muestra aleatoria de 100 votantes, 45 de ellos votaron al PSOE en las últimas elecciones. Usar esta información para estimar la proporción de los votantes en España que votaron al PSOE. Dar una estimación puntual y un intervalo de confianza de 95%.

20 personas en una muestra de 30 norteamericanos están a favor de la legalización de la mariguana. Estimar la proporción de la población americana que esté a favor y dar un intervalo de 90%.



Computación en Excel

	A	B	C	D
1	n	100		Datos
2	x	45		
3				
4	p	0,45	=B2/B1	Proporción
5	p(1-p)/n	0,002475	=B4*(1-B4)/B1	Varianza
6				
7	alfa	0,01		Computación de z
8	alfa/2	0,005		
9	1-alfa/2	0,995		
10	z	2,58		=DISTR.NORM.ESTAND.INV(0,995)
11				
12	z*raíz(p(1-p)/n)	0,1281		
13				
14	Intervalo	0,3219	0,5781	
15		=B4-B12	=B4+B12	

¿Podemos utilizar INTERVALO.CONFIANZA.NORM?



¡Si! ... pero tiene “truco”

Argumentos de función

INTERVALO.CONFIANZA

Alfa	0,01	=	0,01
Desv_estándar	raíz(0,45*(1-0,45))	=	0,497493719
Tamaño	100	=	100

= 0,12814589

Devuelve el intervalo de confianza para la media de una población.

Tamaño es el tamaño de la muestra.

Resultado de la fórmula =

[Ayuda sobre esta función](#)

Aceptar Cancelar

Ahora la desviación estándar es $\sqrt{p^{\wedge} \times (1-p^{\wedge})}$.

Restando y sumando el resultado de 0,45, obtenemos el resultado.



Ejemplo: (Pregunta de Examen)

Los datos proviene del “último” barómetro del CIS. Las valoraciones siguen distribuciones normales con desviaciones típicas iguales a las que aparecen en la tabla.

Hallar intervalos de confianza al 95% para las verdaderas valoraciones medias de Alfredo Pérez Rubalcaba y de Mariano Rajoy.

¿Es razonable suponer que las valoraciones medias son iguales?

Comentar su respuesta.

	Media	Desviación típica	(N)
Enrique Álvarez Sostres	2.72	2.38	(133)
Joan Baldoví Roda	3.06	2.76	(104)
Uxue Barkos	4.27	2.74	(302)
Alfred Bosch	3.69	2.78	(211)
Rosa Díez	4.33	2.50	(1594)
Josep A. Durán i Lleida	2.63	2.40	(1454)
Josu Erkoreka	2.85	2.53	(491)
Mikel Errekondo	2.50	2.67	(224)
Francisco Jorquera	3.01	2.48	(216)
Cayo Lara	3.88	2.62	(1379)
Ana María Oramas	3.43	2.50	(170)
Alfredo Pérez Rubalcaba	3.40	2.57	(2314)
Mariano Rajoy	2.81	2.69	(2372)
Carlos Salvador	2.28	2.25	(96)



Ejemplo: (Pregunta de Examen)

La siguiente tabla proviene del barómetro del CIS de abril de 2011.

PREGUNTA 2

Y, ¿cree Ud. que la situación económica actual del país es mejor, igual o peor que hace un año?

	%	(N)
Mejor	5.3	(130)
Igual	35.1	(865)
Peor	57.6	(1418)
N.S.	1.7	(42)
N.C.	0.3	(8)
TOTAL	100.0	(2463)

Hallar un intervalo de 95% de confianza para la proporción de personas en España que piensan que la situación económica ahora es igual o mejor que hace un año.



Ejemplo: (Pregunta de Examen)

La siguiente noticia apareció en *La Razón*, el 9 de diciembre de 2009.

Una encuesta sostiene por primera vez que los catalanes quieren la independencia

Un 50'3% de los catalanes votarían a favor de la independencia de Cataluña y sólo un 17,8% votaría en contra en el caso de que esta comunidad celebrase un referéndum democrático de autodeterminación, según un estudio de la Universitat Oberta de Catalunya (UOC) dirigido por Miquel Strubell.

El estudio se ha realizado mediante encuesta telefónica a 2.614 personas, de entre 15 y 74 años que residen en Cataluña, y basándose en una muestra aleatoria estratificada en seis veguerías. Ésta es la primera vez que un sondeo de estas características apunta a una victoria soberanista, aunque condicionada a la participación.

Preguntadas por la decisión que tomarían en el caso de que en Cataluña se celebrase un referéndum de autodeterminación, un 50,3% de las personas entrevistadas votarían a favor de esta consulta, mientras que **un 24,6% se abstendrían**, un 17,8% votaría en contra y un 7,2% no sabe o no contesta.

A pesar de estas cifras, **un 58,1% de los encuestados cree que Cataluña no llegará a ser independiente**, frente al 31,1% que lo ven posible.

Utilizando los datos de la encuesta, calcular un intervalo de 95% de confianza para la verdadera proporción de catalanes que votarían a favor de la independencia.



Ejemplo: (Pregunta de Examen)

La siguiente tabla proviene del Barómetro del Real Instituto Elcano (BRIE) de marzo-abril 2012. Los resultados provienen de una muestra de 1000 españoles adultos.

1. CONDENA UNÁNIME DE LA NACIONALIZACIÓN DE REPSOL-YPF EN ARGENTINA

- Como consecuencia de los acontecimientos recientes, el **90%** de los españoles considera muy grave la nacionalización de Repsol-YPF por el gobierno argentino. Y en este mismo sentido valora las trabas a otras empresas inversoras españolas en el país.

¿Cómo valoraría la situación creada por la nacionalización de Repsol-YPF por parte de Argentina?

Muy grave	Bastante grave	Algo grave	Poco grave	Nada grave	NS	NC
14,2%	40,7%	27,7%	5,3%	2,9%	8,0%	1,2%

- Uno de cada dos españoles cree que las relaciones diplomáticas entre España y Argentina empeorarán como consecuencia de estas tensiones.

¿Y cómo cree que afectará a las relaciones entre España y Argentina?

Mejorarán	Seguirán igual	Empeorarán	NS	NC
11,8%	34,7%	44,1%	8,2%	1,2%

Calcular un intervalo de 95% de confianza para la verdadera proporción de los españoles que piensan que la nacionalización de Repsol-YPF por Argentina es una situación nada grave.



Material Adicional



Un intervalo de 95% de confianza para la media de una población normal (varianza desconocida)

Hasta ahora, hemos supuesto que la varianza es conocida para construir el intervalo de confianza. En la práctica, este supuesto puede ser incorrecto.

¿Qué hacemos?

Si la muestra es grande (> 30), podemos construir el intervalo como anteriormente, sustituyendo la desviación muestral, s^2 , por la poblacional que desconocemos.



Si la muestra es más pequeña utilizamos un intervalo basado en La distribución *t de Student*.

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{n-1}(0,975)$$

¿Qué es t?

Parece un más difícil construirlo pero en Excel 2010 ... se puede hacer de manera fácil.



Ejemplo

Se tienen datos sobre la duración de las sentencias de 19 criminales convictos de homicidios en España. La media y desviación típica de las sentencias son de 72,7 y 10,2 meses respectivamente.

Calcular un intervalo de 95% para la duración media de las sentencias por homicidio en España.



Usamos la función INTERVALO.CONFIANZA.T

	A	B	C	D
1	n	19		Datos
2	media	72,7		
3	s	10,2		
4				
5	alfa	0,05		
6				
7	$t*s/raíz(n-1)$	5,721		=INTERVALO.CONFIANZA.T(alfa;s;n)
8				
9	intervalo	66,979	78,421	
10		=B2-B7	=B2+B7	

Con los datos originales es aún más fácil construir un intervalo de confianza ...



Ejemplo

Se ha diseñado una encuesta para estimar el sueldo medio anual de los banqueros en España. Una muestra de 10 ejecutivos proporcionaron los siguientes resultados (en miles de euros).

1200, 1000, 1500, 800, 750, 2400, 1000, 1600, 700, 600

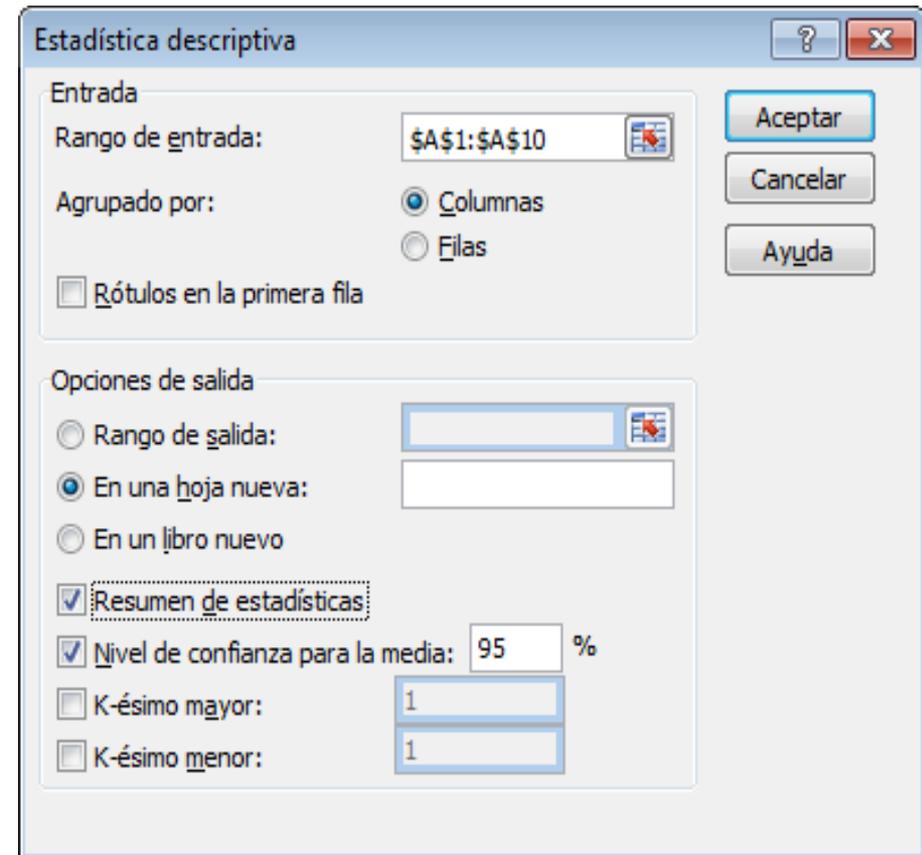
Hallar un intervalo de 95% de confianza para el verdadero sueldo medio de los banqueros en España.



Utilizamos la opción de Estadística Descriptiva en Data Analysis de Excel.

Columna1

Media	1155,00
Error típico	173,92
Mediana	1000,00
Moda	1000,00
Desviación estándar	549,97
Varianza de la muestra	302472,22
Curtosis	1,95
Coefficiente de asimetría	1,40
Rango	1800,00
Mínimo	600,00
Máximo	2400,00
Suma	11550,00
Cuenta	10,00
Nivel de confianza(95,0%)	393,43



Simplemente restamos y sumamos para calcular el intervalo.

(€761570, €1548430)

761,57 1548,43



Un intervalo de 95% de confianza para la diferencia de las medias de dos poblaciones normales (datos emparejados)

¿Qué son datos emparejados?

	Año	
Banquero	2012	2013
1	1300	1200
2	1100	1000
3	1200	1500
4	900	800
5	800	750
6	2000	2400
7	1100	1000
8	1500	1600
9	700	700
10	500	600

Ejemplo

Tenemos los datos de los sueldos de los mismos banqueros tanto en el año 2012 como en el año 2013. Queremos estimar el crecimiento en el sueldo medio de los banqueros en los dos años.



	Año		
Banquero	2012	2013	Diferencia
1	1300	1200	-100
2	1100	1000	-100
3	1200	1500	300
4	900	800	-100
5	800	750	-50
6	2000	2400	400
7	1100	1000	-100
8	1500	1600	100
9	700	700	0
10	500	600	100
	1110	1155	45

Calculamos los sueldos medios en cada año y calculamos la diferencia:

$$1155 - 1110 = 45$$

o calculamos los cambios en sueldo y sacamos la media de las diferencias:

$$(-100 - 100 + 300 + \dots + 100) / 10 = 45$$

Una estimación razonable del crecimiento del sueldo de los banqueros es €45000.

¿Cómo sacamos un intervalo de confianza?



Banquero	Diferencia
1	-100
2	-100
3	300
4	-100
5	-50
6	400
7	-100
8	100
9	0
10	100
	45

Consideramos sólo la muestra de diferencias.

Ya tenemos sólo una muestra y podemos calcular un intervalo de confianza basado en la distribución t de Student

Columna1

Media	45,00
Error típico	56,98
Mediana	-25,00
Moda	-100,00
Desviación estándar	180,20
Varianza de la muestra	32472,22
Suma	450,00
Cuenta	10,00
Nivel de confianza(95,0%)	128,91

El intervalo es de $45 \pm 128,91$ miles de euros, es decir (- € 83910, €173910).

Parece verosímil que no haya cambiado el verdadero sueldo medio.