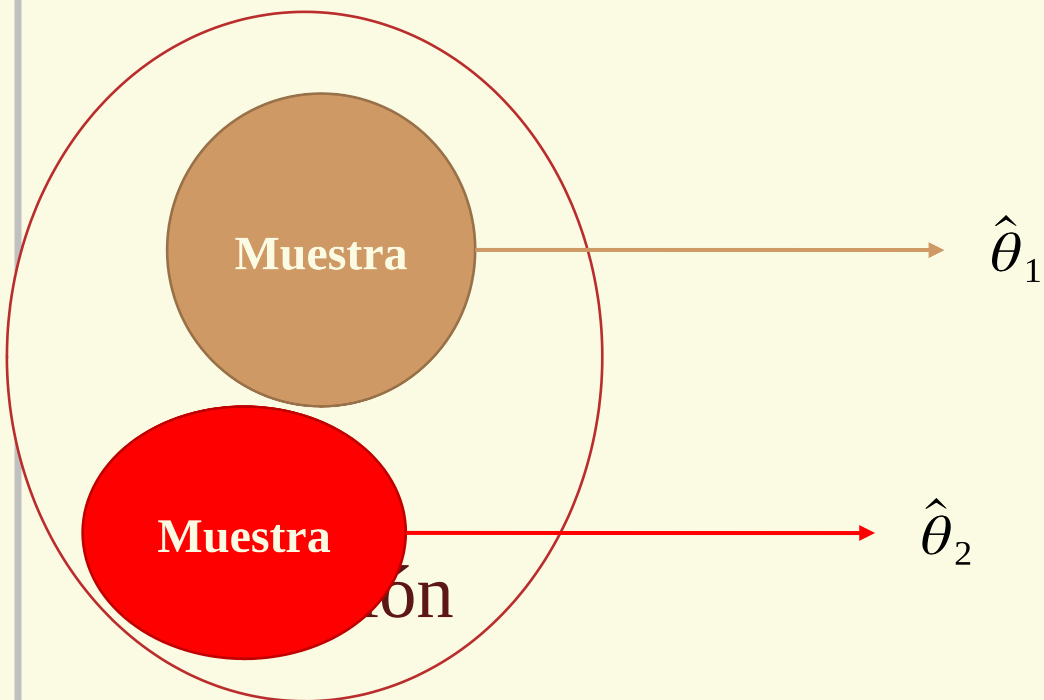
Estimación



Distintas muestras pueden producir estimaciones distintas. ¡El resultado de una estimación es una variable aleatoria!!!

Estimador: variable aleatoria correspondiente a las estimaciones de basados en una muestra al azar

Estimación de la media

$\mu = 10$ $\sigma = 2$

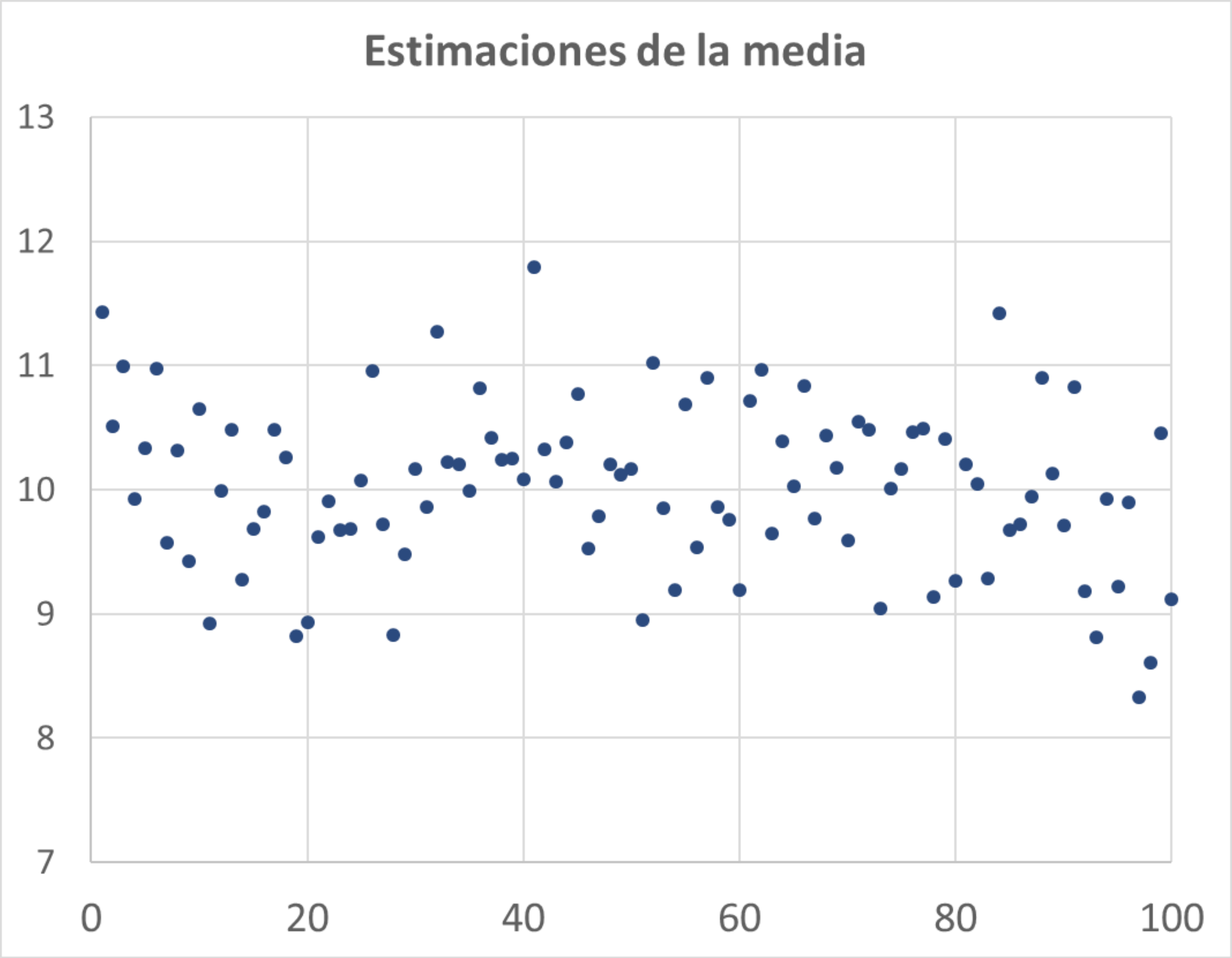
	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
	9.14	9.40	14.23	8.10	9.52
	6.29	11.87	9.20	9.52	14.72
	9.39	6.56	10.16	11.87	10.89
	8.30	8.20	10.62	11.72	11.33
	10.26	11.01	7.34	11.35	6.39
	10.07	5.84	7.49	10.01	8.75
	11.22	12.27	11.01	13.36	11.39
	8.10	9.39	8.56	14.27	11.27
	14.84	8.84	11.72	12.83	11.32
	11.55	9.31	4.62	9.05	11.44
$\hat{\mu}$	**9.92**	**9.27**	**9.49**	**11.21**	**10.70**

Estimación de la media

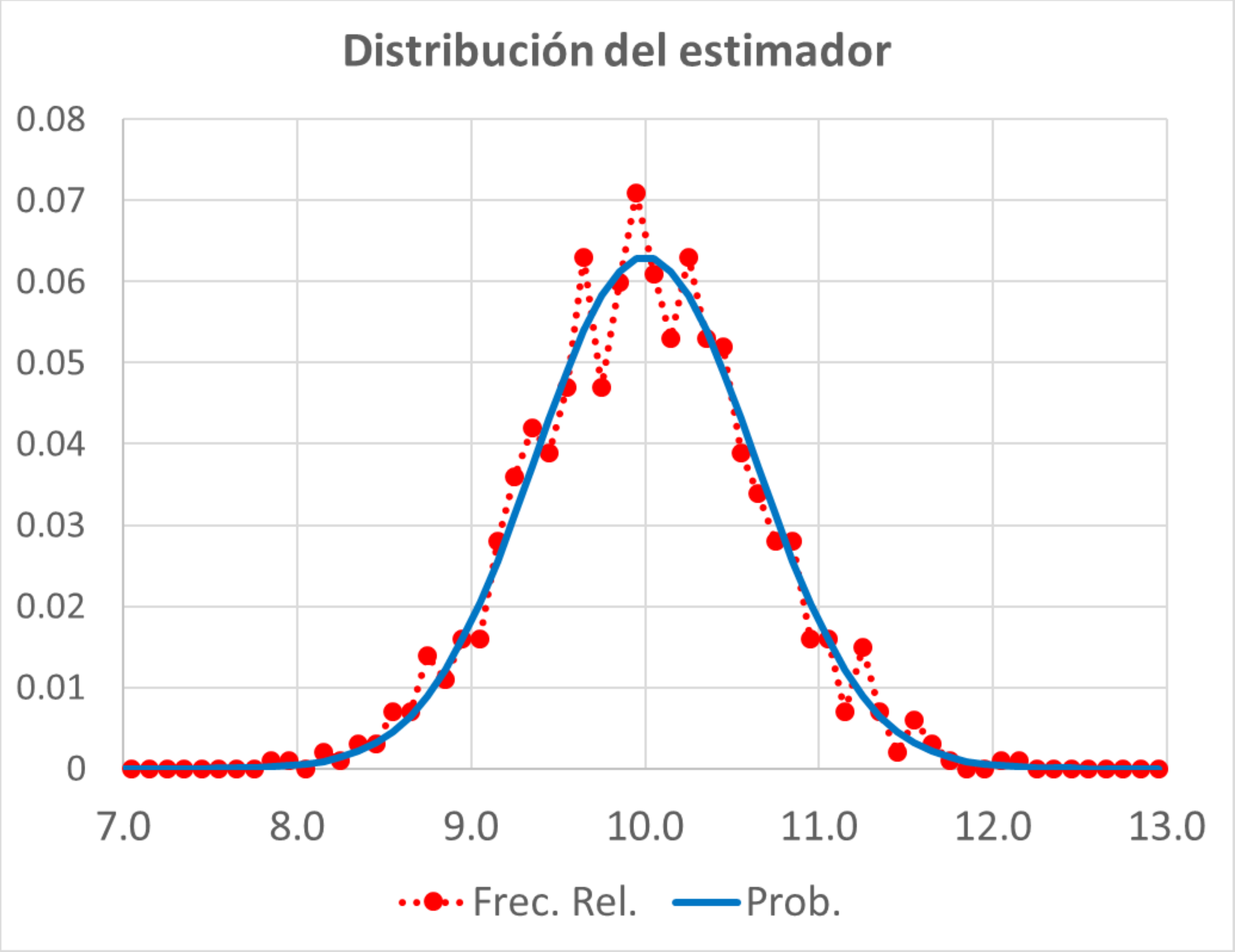
	B	C	D	E	F
1	$\mu =$	10	$\sigma =$	2	
2	**Muestra 1**	**Muestra 2**	**Muestra 3**	**Muestra 4**	**Muestra 5**
3	=NORM.INV(RAND(),C1,E1)			8.10	9.52

	$\mu =$ 10		$\sigma =$ 2	
	Muestra 1	**Muestra 2**	**Muestra 3**	**M**
	9.14	9.40	14.23	
	6.29	11.87	9.20	
	9.39	6.56	10.16	
	8.30	8.20	10.62	
	10.26	11.01	7.34	
	10.07	5.84	7.49	
	11.22	12.27	11.01	
	8.10	9.39	8.56	
	14.84	8.84	11.72	
	11.55	9.31	4.62	
$\hat{\mu}$	=AVERAGE(B3:B12)		**9.49**	

Estimación de la media



Estimación de la media



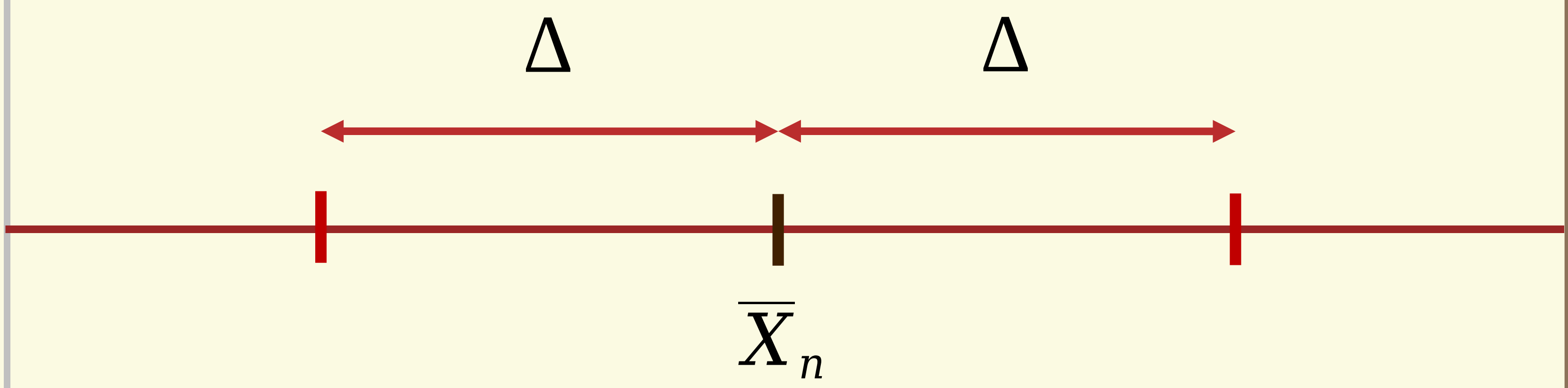
Estimación de la media

i.i.d. y se desea estimar :

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

¿Qué tan probable es que una estimación haya caído cerca de la media real?

Intervalo de confianza para la media



Se fija un **nivel de confianza** alto (cercano a 1):

¿Cómo tiene que ser para que la probabilidad de que la media esté en sea al menos ?

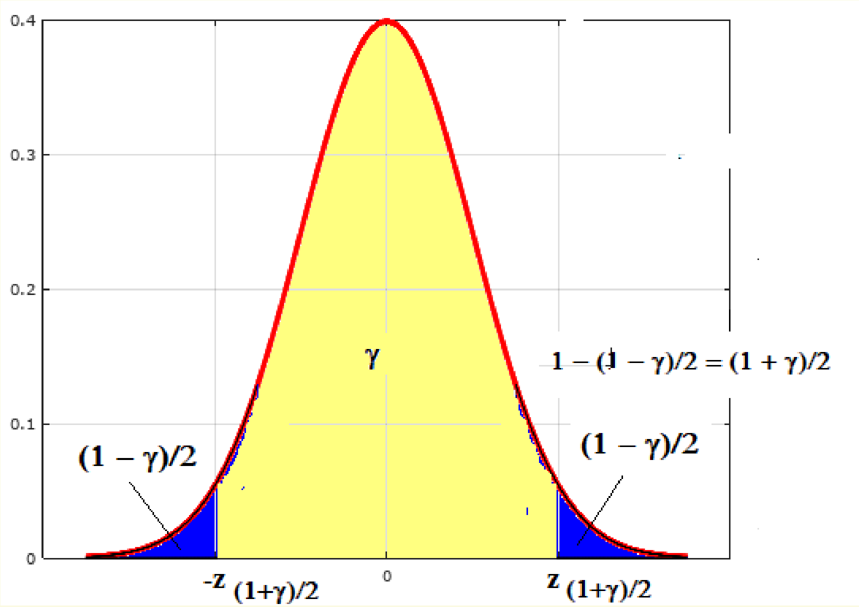
Intervalo de confianza para la media

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(\left|\overline{X}_n - \mu\right| \leq \Delta\right) = \gamma$$

$$P\left(\left|Z\right| \leq \frac{\Delta/\sigma}{\sqrt{n}}\right) = \gamma$$

$$\Longrightarrow \Delta = \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}$$



Intervalo de confianza para la media

¿Cómo tiene que ser para que la probabilidad de que la media esté en sea al menos ?

$$\Delta = \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}$$

	$Z_{(1+\ )/2}$
0.90	1.645
0.95	1.960
0.955	2.005
0.99	2.576

Intervalo de confianza para la media

$$\overline{X}_n \pm \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}$$

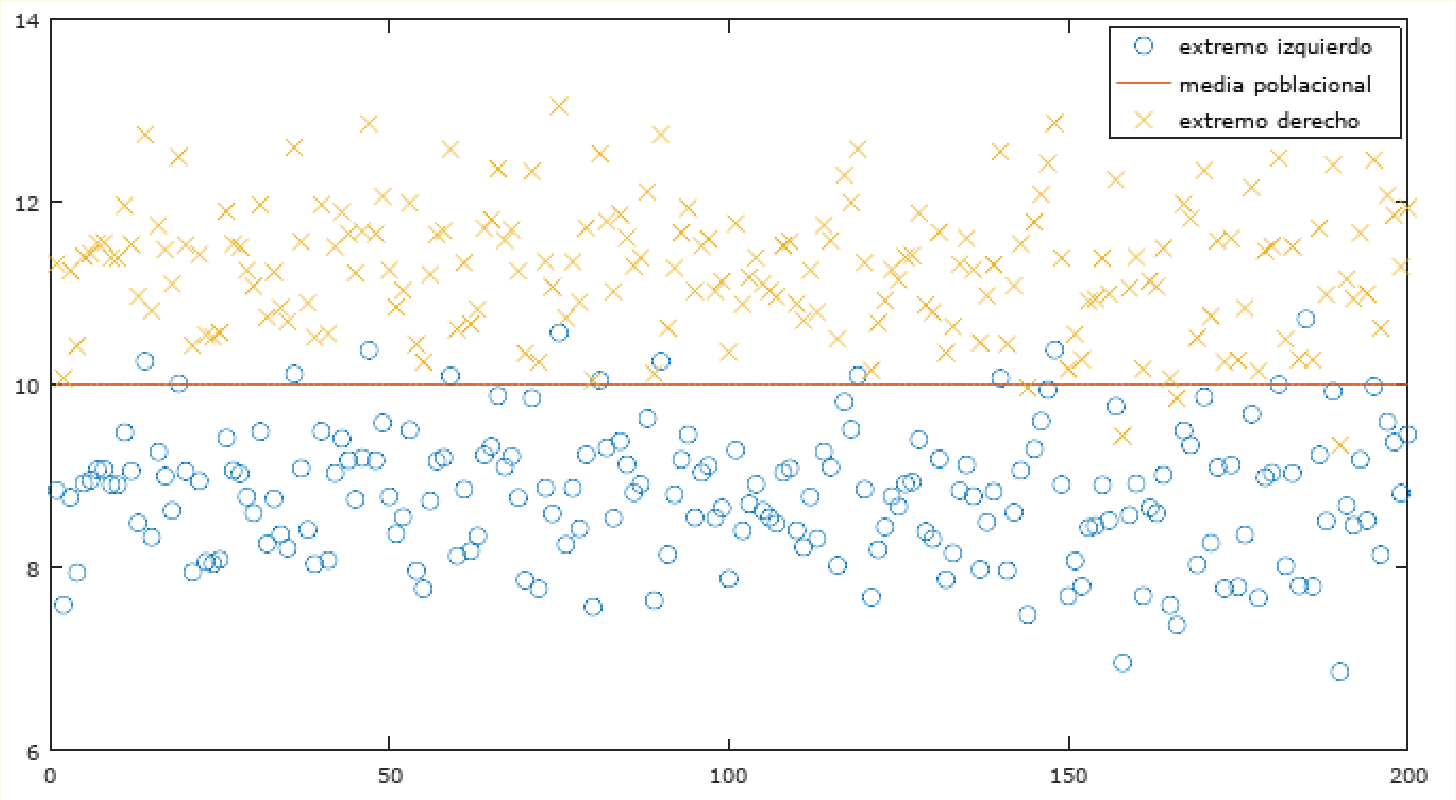
¿Qué significa esto?

NO significa que el error de una estimación es

$$\frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}$$

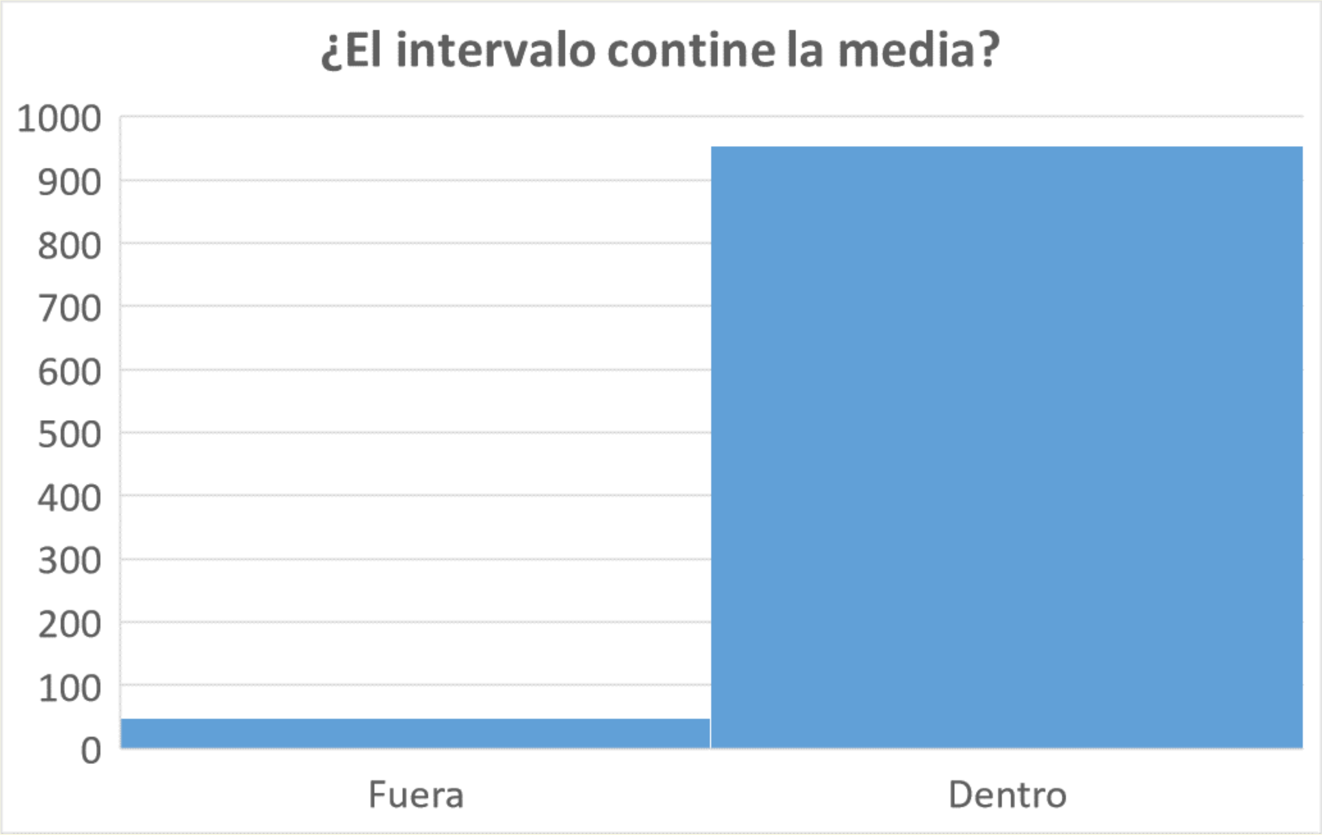
Lo que significa es que, si se hacen estimaciones, en aproximadamente de ellas la media estará dentro del intervalo

Intervalo de confianza para la media

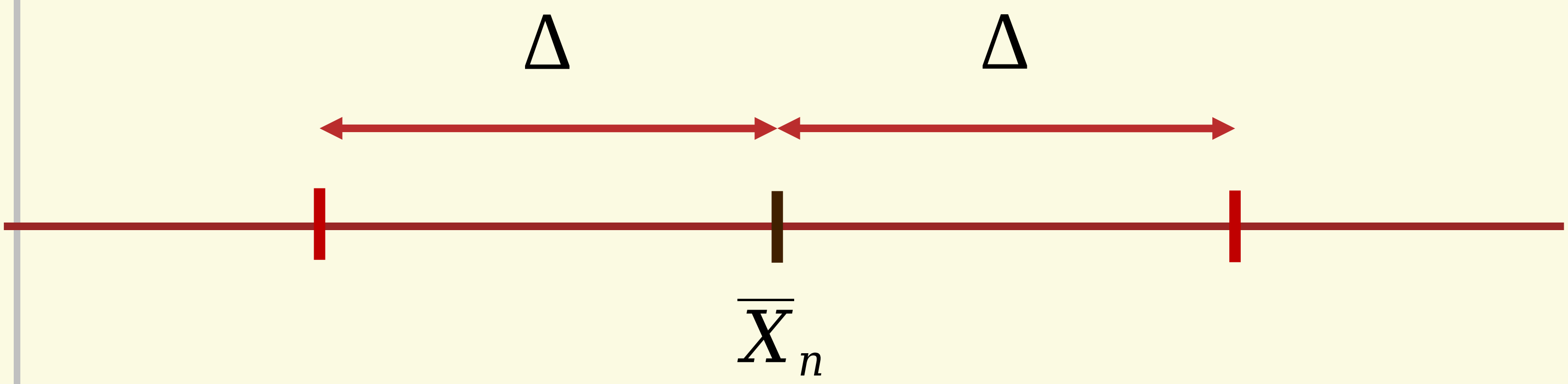


200 intervalos de confianza de nivel de confianza 0.95; para esta colección de determinaciones el 93.5 % de ellos tuvieron la propiedad de que la media perteneciese a esos intervalos.

Intervalo de confianza para la media



Intervalo de confianza para la media



Este intervalo de confianza es bilateral

Pero nos puede preocupar un límite inferior o un límite superior

Intervalo de confianza para la media

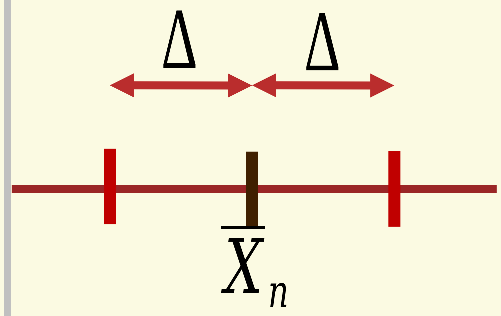
Ejemplo 1: Nos preocupa que la tensión media de rotura de un elemento no sea menor que cierto valor.

Ejemplo 2: Una máquina llena paquetes de azúcar en la empresa Medasma. El peso llenado es una variable aleatoria (es imposible llenar exactamente la misma cantidad – cual adolescente, algún grano se le escapa).

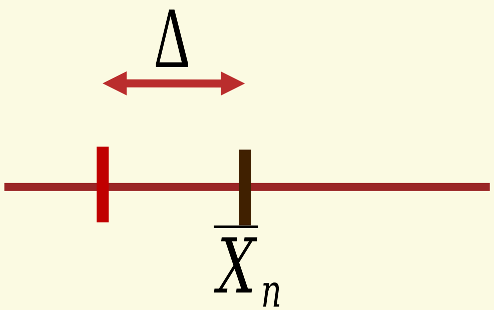
A la empresa Medasma le preocupa que el valor medio no sea superior a cierta cantidad (de lo contrario, estaría perdiendo dinero).

A la Asociación de Defensores del Consumidor de Azúcar (ADECA) le preocupa que el valor medio no sea inferior a cierto número (de lo contrario, el consumidor puede estar pagando de más).

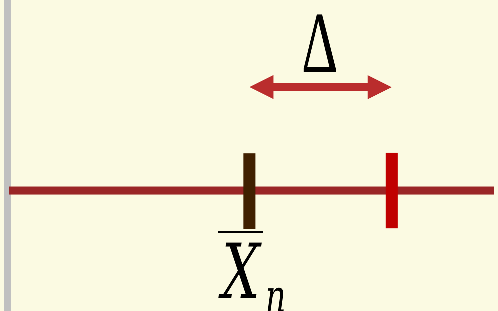
Intervalo de confianza para la media



$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)$$



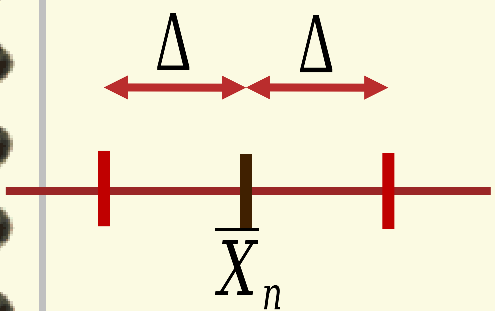
$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\gamma}, +\infty\right)$$



$$\left(-\infty, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\gamma}\right)$$

Intervalo de confianza para la media

i.i.d. y se desea estimar



$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}}\, z_{\frac{1+\gamma}{2}}\,,\ \overline{X}_n + \frac{\sigma}{\sqrt{n}}\, z_{\frac{1+\gamma}{2}}\right)$$

¿Y si las variables **no son normales**?

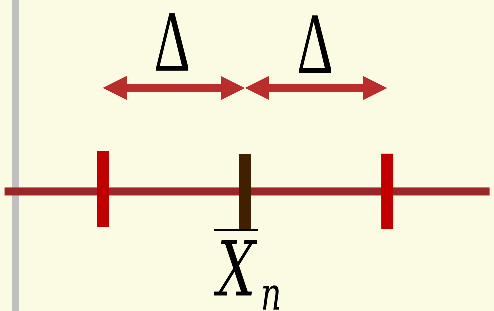
Si es grande, no hay problemas porque podemos recurrir al Teorema Central del Límite

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \ {}_{\text{apróx.}}\ N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Usamos las mismas fórmulas.

Intervalo de confianza para la media

i.i.d. y se desea estimar



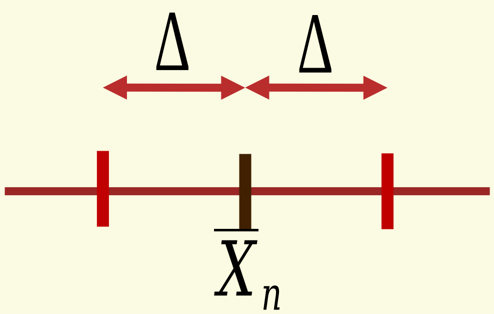
$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)$$

¿Y si las variables **no son normales y no es grande**?

El problema es complicado… hay que preguntarle a alguien que sepa…

Intervalo de confianza para la media

i.i.d. y se desea estimar



$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}} , \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}} \right)$$

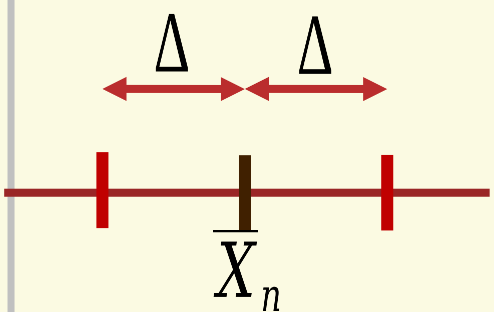
Está claro que no conozco , pero las fórmulas usan : ¿de dónde sale?

Si es muy grande, en muchos casos se puede aproximar por el desvío muestral .

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2}$$

Intervalo de confianza para la media

i.i.d. y se desea estimar



$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}}\, z_{\frac{1+\gamma}{2}} , \overline{X}_n + \frac{\sigma}{\sqrt{n}}\, z_{\frac{1+\gamma}{2}}\right)$$

Está claro que no conozco , pero las fórmulas usan : ¿de dónde sale?

Si no es muy grande, pero las variables son normales, lo vemos la próxima...

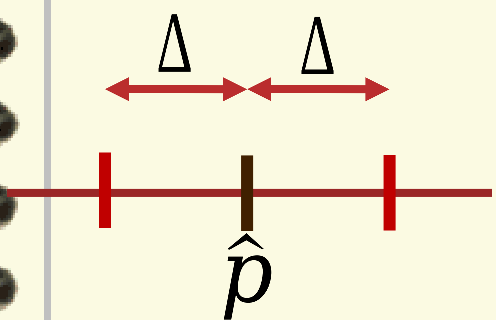
Intervalo de confianza para la proporción

Exactamente la misma idea es válida para la proporción de una población. Si es grande,

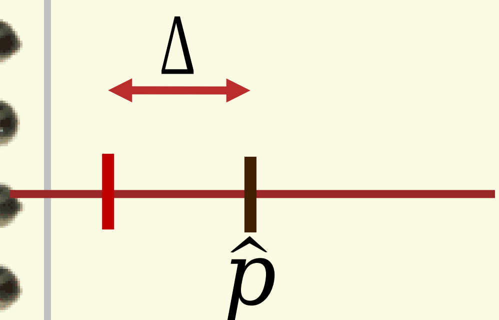
$$\frac{\hat{p}\left(1-\hat{p}\right)}{n} \approx \frac{p\left(1-p\right)}{n}$$

Por eso usamos las fórmulas con y reemplazados por y .

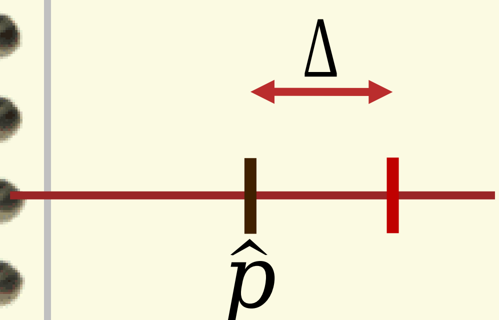
Intervalo de confianza para la proporción



$$\left(\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{\frac{1+\gamma}{2}}, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{\frac{1+\gamma}{2}}\right)$$



$$\left(\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{\gamma}, 1\right)$$



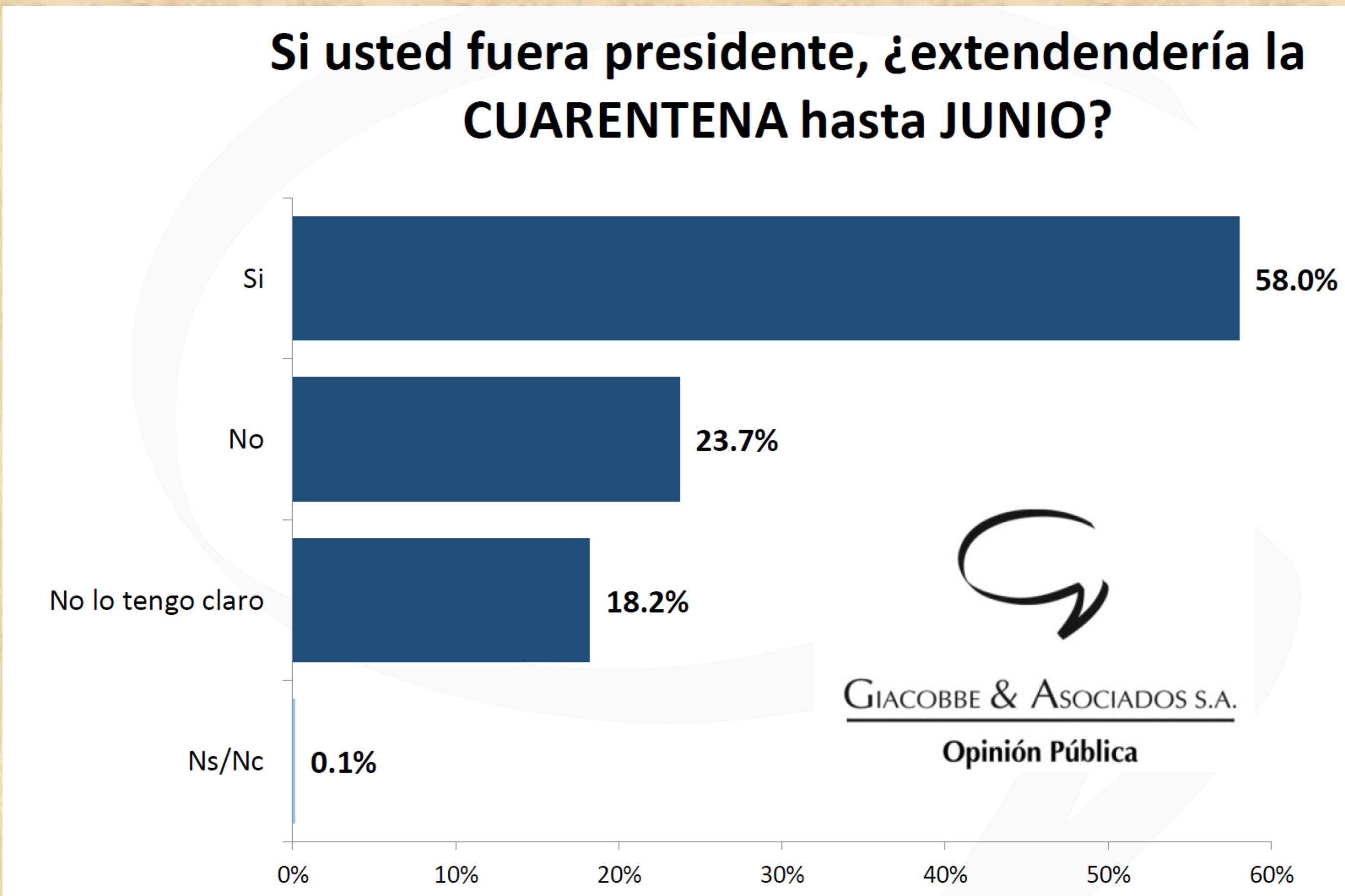
$$\left(0, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{\gamma}\right)$$

POLITICA / PANDEMIA

Jueves 7 Mayo, 2020

Encuesta: bajó la imagen de Alberto Fernández, pero la mayoría mantendría la cuarentena hasta junio

La discusión por la liberación de presos habría sido uno de los principales factores que perjudicaron al Presidente.



https://www.perfil.com/noticias/politica/encuestas-bajo-imagen-alberto-fernandez-pero-mayoria-mantendria-cuarentena-hasta-junio.phtml

FICHA TÉCNICA

Área
ARGENTINA.

Fecha de realización
27 AL 29 DE ABRIL DE 2020.

Tipo de muestreo
Ajustado por cuotas de género, edad, regiones, secciones electorales de la provincia de Buenos Aires, comunas de la Ciudad de Buenos Aires.

Tamaño de la muestra
2500 CASOS.

Margen de error
+/- 2%.

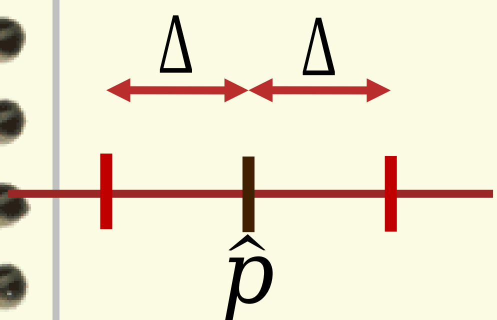
Modalidad
Cuestionario estructurado con preguntas abiertas y cerradas.

Sistema de consulta
Encuestas a dispositivos móviles.

Intervalo de confianza para la proporción

Falta el nivel de confianza

¿Qué valor ponemos?



$$\left(\hat{p}-\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,z_{\frac{1+\gamma}{2}},\ \hat{p}+\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,z_{\frac{1+\gamma}{2}}\right)$$

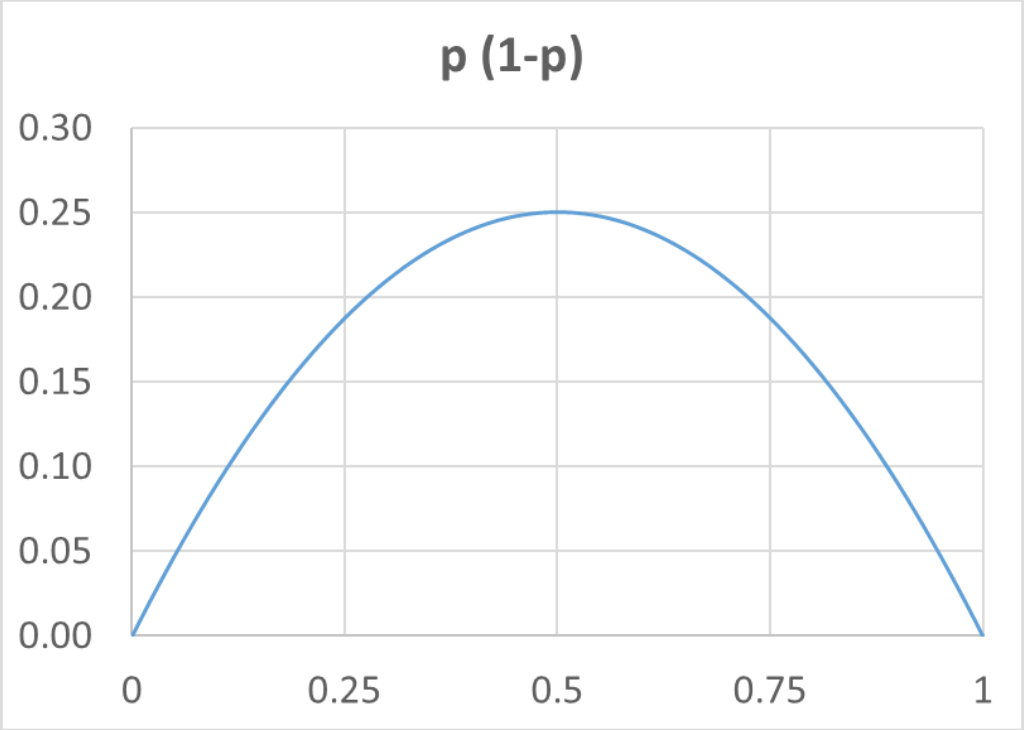
Tamaño de la muestra
2500 CASOS.

Margen de error
+/- 2%.

Intervalo de confianza para la proporción

Buscamos el peor caso

$$p(1-p) \leq \frac{1}{4}$$



Intervalo de confianza para la proporción

$$\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}\,z_{\frac{1+\gamma}{2}} \le \sqrt{\frac{1}{4n}}\,z_{\frac{1+\gamma}{2}} = \sqrt{\frac{1}{4\cdot 2500}}\,z_{\frac{1+\gamma}{2}} = 0.02$$

$$z_{\frac{1+\gamma}{2}} = 2$$

$$\frac{1+\gamma}{2} = 0.97725$$

$$\gamma = 0.9545 \approx 95\%$$