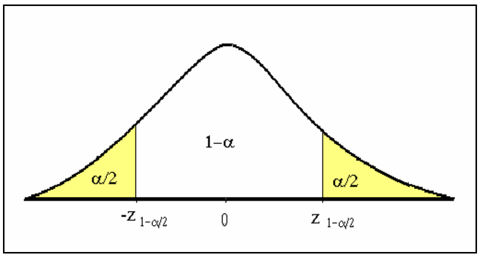
11 INFERENCIA ESTADÍSTICA I (INTERVALOS DE CONFIANZA)

Sea Ω una población y sobre ella una variable aleatoria X que sigue una ley normal N(μ; σ), con media **μ desconocida** y desviación típica **σ conocida.**

Con el fin de estimar μ se toma una muestra aleatoria simple de tamaño n M = (x_1,x_2,...,x_n), que proporciona una media $\overline{x}_n$. Por el **Teorema Central del Límite** sabemos que si la población es grande y n ≥ 30, entonces $\overline{x}_n$ sigue una ley normal $\mathcal{N}(\mu;\frac{\sigma}{\sqrt{n}})$, luego la variable tipificada $Z=\frac{\overline{x}_n-\mu}{\frac{\sigma}{\sqrt{n}}}$ sigue una ley normal $\mathcal{N}(0;1)$.

Sea $z_{1-\frac{\alpha}{2}}$ tal que $P\{Z\le z_{1-\frac{\alpha}{2}}\}=1-\frac{\alpha}{2}$; es decir $\Phi(z_{1-\frac{\alpha}{2}})=1-\frac{\alpha}{2}$, donde Φ(z) es la función de distribución acumulativa de la $\mathcal{N}(0; 1)$. Entonces

$$P\left[-z_{1-\frac{\alpha}{2}}\le Z\le z_{1-\frac{\alpha}{2}}\right]=1-\alpha$$



de donde $P\left[\overline{x}_n-z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{x}_n+z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]=1-\alpha$.

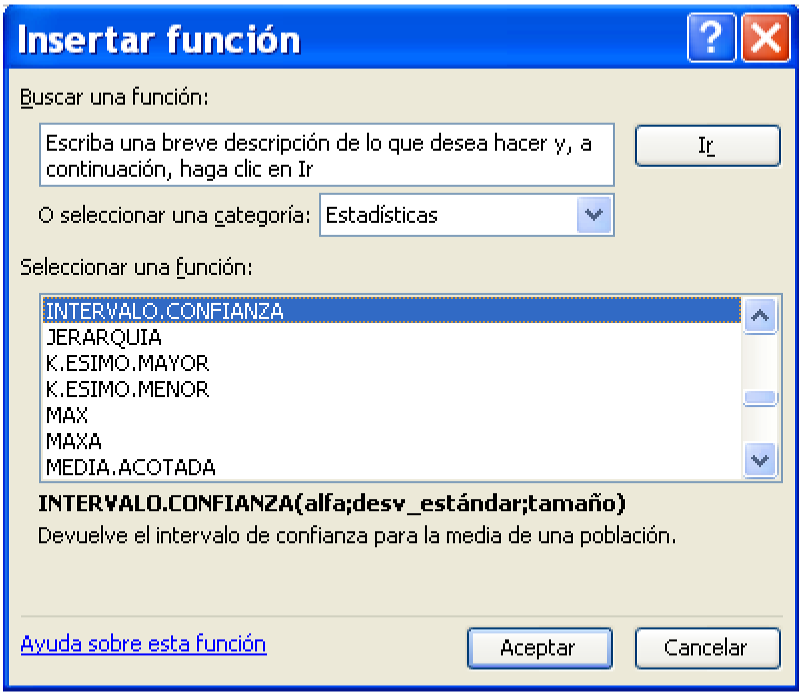
Así, $I=\left[\overline{x}_n-z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}};\quad\overline{x}_n+z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right]$ es el **intervalo de confianza para μ con el nivel de confianza 1 – α (nivel de significación α).** A $\mathbf{z_{1-\alpha/2}}$ se le llama **valor crítico** correspondiente a **α.**

Se trata de un intervalo centrado en la media de la muestra, $\overline{x}_n$, y de radio $E=z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$. Este radio es el **error máximo** que se comete al estimar μ mediante $\overline{x}_n$. Mide la **precisión** del intervalo: a menor radio más precisa es la estimación.

La librería de funciones estadísticas de EXCEL incluye la función

=INTERVALO.CONFIANZA(α;σ;n)

que devuelve el radio $E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ del intervalo con un nivel de confianza $1-\alpha$, esto es, con un nivel de significación α. El intervalo es, por tanto, $I = [a,b] = [\bar{x}_n - E\,;\ \ \bar{x}_n + E]$.



Como en el texto hay suficientes ejercicios resueltos manualmente, usando las tablas de la normal tipificada $\mathcal{N}(0;1)$, aquí obtendremos los resultados mediante EXCEL siempre que sea posible.

1 La variable estadística X sigue una ley normal de media desconocida, μ, y desviación típica $\sigma = 3$. Extraída una muestra de tamaño n = 36 se ha obtenido una media muestral $\overline{x}_{36}$. Hallar un intervalo para estimar μ con un nivel de confianza del 95%. Aplicar al caso en que $\overline{x}_{36} = 4$.

Microsoft Excel - EjPr_11_1.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

Arial

E2 f_x =INTERVALO.CONFIANZA(B2;C2;D2)

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	E=radio=error máx.
2	0,95	0,05	3	36	0,979982
3					
4	centro=media muestral=	4			
5	EXTREMOS:	a	b		
6	INTERVALO: [	3,020018	4,979982	]	

2 El consumo de gasolina de los vehículos de un empresa, durante un período de 36 días elegidos al azar durante el año 2002, proporciona una media de 4000 litros/día. Se sabe que el consumo de la empresa sigue una ley normal cuya varianza es de 1600 litros/día. Determinar un intervalo para estimar el consumo medio diario durante 2002 con un nivel de confianza del 99%

Microsoft Excel - EjPr_11_2.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

Arial

E2 f_x =INTERVALO.CONFIANZA(B2;C2;D2)

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	E=radio=error máx.
2	0,99	0,01	40	36	17,172195
3					
4	centro=media muestral=	4000			
5	EXTREMOS:	a	b		
6	INTERVALO: [	3982,8278	4017,1722	]	

Recordemos el **"TEOREMA DEL LÍMITE PARA PROPORCIONES"**:

Si la distribución de una población grande tiene una proporción p de una cierta característica A, entonces la variable aleatoria, P, de las proporciones muestrales extraídas de esta población (que se distribuye binomialmente), a medida que aumenta al tamaño de la muestra, se <u>aproxima</u> a una distribución normal:

$$\mathcal{N}\left(\mu_P = p;\ \sigma_P = \sqrt{\frac{p\cdot(1-p)}{n}}\right)$$

(La afirmación se da por buena si n ≥ 30 y N ≥ 2n)

Con el fin de estimar ***p*** se toma una muestra aleatoria simple de tamaño n M = $(x_1,x_2,...,x_n)$, que da una proporción muestral $\hat{p} = p_n = \frac{n_A}{n}$**.** La variable tipificada $Z = \frac{\hat{p}-P}{\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}}$ sigue una ley normal $\mathcal{N}(0;1)$. Sea $z_{1-\frac{\alpha}{2}}$ tal que $P\{Z \le z_{1-\frac{\alpha}{2}}\} = 1-\frac{\alpha}{2}$; es decir $\Phi(z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$, donde *Φ(z)* es la función de distribución de la N(0; 1). Entonces,

$$P\left[-z_{1-\frac{\alpha}{2}} \le Z \le z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

de donde

$$P\left[\hat{p} - z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} \le P \le \hat{p} + z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right] = 1-\alpha$$

Así,

$$I = \left[\hat{p} - z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}};\ \hat{p} + z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right]$$

es el **intervalo de confianza para la proporción, *p*, con el nivel de confianza 1 − α (nivel de significación α).**

Se trata de un intervalo centrado en la proporción de la muestra, $\hat{p} = p_n$, y de radio $E = z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = z_{1-\frac{\alpha}{2}}\cdot\frac{\sqrt{\hat{p}\cdot(1-\hat{p})}}{\sqrt{n}}$. Este es el <u>**error máximo**</u> que se comete al estimar ***p*** mediante $\hat{p} = p_n$.

Para estimar la proporción, EXCEL no proporciona directamente el radio del intervalo es $E = z_{1-\frac{\alpha}{2}}\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}$, pero lo podemos calcular: $\Phi(z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$ implica $z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ que se calcula con **=DISTR.NORM.ESTAND.INV(1−α/2)**

El centro del intervalo es la proporción muestral: $centro = \hat{p}$, y el intervalo es, $I = [a,b] = [\hat{p}-E;\quad \hat{p}+E]$

3 Una determinada marca de electrodomésticos selecciona una muestra de tamaño n = 300 de un cierto modelo a fin de estimar la proporción de toda la producción que son defectuosos. Resulta que la muestra contiene 30 defectuosos. Dar un intervalo de confianza, al 95%, para estimar la proporción de artículos defectuosos de dicho modelo.

Microsoft Excel - EjPr_11_3.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

D2 =DISTR.NORM.ESTAND.INV(1-B2/2)*RAIZ(B5*(1-B5)/C2)

	A	B	C	D
1	nivel de confianza=1-alfa	alfa	tamaño=n	E=radio=error máx.
2	0,95	0,05	300	0,033948
3				
4	núm. defectuosos = nA =	30		
5	centro=p^= nA/n =	0,1		
6	EXTREMOS:	a	b	
7	INTERVALO: [	0,0661	0,1339	]

4 A fin de determinar el tiempo medio, en minutos, que emplea un autobús de circunvalación en realizar su recorrido, se midió dicho tiempo durante 25 días, dando una media de 35 minutos. Suponiendo que el tiempo empleado en el recorrido se distribuye normalmente con media desconocida, μ, y desviación típica σ = 4 minutos, construir un intervalo de confianza para μ con un nivel de significación α = 0,01

Microsoft Excel - EjPr_11_4.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

E2 =INTERVALO.CONFIANZA(B2;C2;D2)

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	radio=error máx.
2	0,99	0,01	4	25	2,060663
3					
4	centro=media muestral=	35			
5	EXTREMOS:	a	b		
6	INTERVALO: [	32,9393	37,0607	]	

5 En una encuesta de opinión, durante una campaña electoral en una ciudad, se preguntó a una muestra aleatoria de 400 personas a cual de los dos partidos, A o B, pensaban votar: 160 declararon que votarían al partido A.

Obtener un estimador puntual y un intervalo de confianza, al 95 por 100, para la proporción de la población que votará al partido A en las elecciones.

Microsoft Excel - EjPr_11_5.xls

D2 f_x =DISTR.NORM.ESTAND.INV(1-B2/2)*RAIZ(B5*(1-B5)/C2)

	A	B	C	D
1	nivel de confianza=1-alfa	alfa	tamaño=n	E=radio=error máx.
2	0,95	0,05	400	0,048009
3				
4	núm. defectuosos=nA=	160		
5	centro=p^=nA/n=	0,4		
6	EXTREMOS:	a	b	
7	INTERVALO: [	0,3520	0,4480	]

6 Para estudiar la velocidad de lectura, en palabras por minuto, de los alumnos de Enseñanza Secundaria Obligatoria de un Instituto, se ha tomado una muestra de tamaño n = 100 alumnos, que proporciona una media muestral de $\bar{x}_{100} = 110$ palabras por minuto. Sabiendo que el intervalo de confianza para la media poblacional, al 95%, es $[107{,}65;\ 112{,}35]$, hallar la desviación típica poblacional, σ.

La media muestral es el centro del intervalo, $\bar{x}_n = \dfrac{a+b}{2}$, y el error máximo el radio del intervalo: $E = \dfrac{b-a}{2}$. Una vez calculado E, de $E = z_{1-\frac{\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}$ se despeja $\sigma = \dfrac{E \cdot \sqrt{n}}{z_{1-\frac{\alpha}{2}}}$.

Microsoft Excel - EjPr_11_6.xls

C2 f_x =E2*RAIZ(D2)/DISTR.NORM.ESTAND.INV(1-B2/2)

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	E=radio=error máx.
2	0,95	0,05	11,9900162	100	2,350000
3					
4	centro=media muestral=	110			
5	EXTREMOS:	a	b		
6	INTERVALO: [	107,6500	112,3500	]	

7 La viscosidad del aceite de las latas que produce una fábrica sigue una ley normal de media desconocida y desviación típica 0,0120. Seleccionada una muestra de tamaño n de dichas latas se ha obtenido una viscosidad media $\bar{x}_n = 4{,}6000$. El extremo superior del intervalo de confianza, al 95%, resulta ser b = 4,6029

a) Determinar el tamaño n de la muestra.
b) Hallar el extremo inferior del intervalo.

El radio del intervalo es $E = b - \bar{x}_n$; el extremo inferior es $a = \bar{x}_n - b$. De $E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ despejamos $\sqrt{n} = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \Rightarrow n = \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \right)^2$

Microsoft Excel - EjPr_11_7.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

D2 =((DISTR.NORM.ESTAND.INV(1-B2/2))*C2/E2)^2

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	E=radio=error máx.
2	0,95	0,05	0,0120	65,775276	0,002900
3					
4	centro=media muestral=	4,6000			
5	EXTREMOS:	a	b		
6	INTERVALO: [	4,5971	4,6029	]	
7					
8	Tamaño de la muestra:	66			

8 La dureza de un cierta fibra se distribuye normalmente con desviación típica 0,1. Para estimar la dureza media de dicha fibra se ha tomado una muestra de tamaño 49 y se ha encontrado un intervalo de confianza [7,372; 7,428].

a) ¿Cuál ha sido la dureza media de las 49 fibras?

b) ¿Cuál es el nivel de confianza de este intervalo?

El intervalo nos proporciona la media muestral $\bar{x}_n = \dfrac{a+b}{2}$ y el error máximo o radio del intervalo es $E = \dfrac{b-a}{2}$. De $E = z_{1-\frac{\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}$ despejamos $z_{1-\frac{\alpha}{2}} = \dfrac{E \cdot \sqrt{n}}{\sigma}$. Entonces

$$\Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \;\Rightarrow\; \Phi\left(\frac{E \cdot \sqrt{n}}{\sigma}\right) = 1 - \frac{\alpha}{2} \;\Rightarrow\; \alpha = 2 \cdot \left[1 - \Phi\left(\frac{E \cdot \sqrt{n}}{\sigma}\right)\right]$$

Microsoft Excel - EjPr_11_8.xls

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

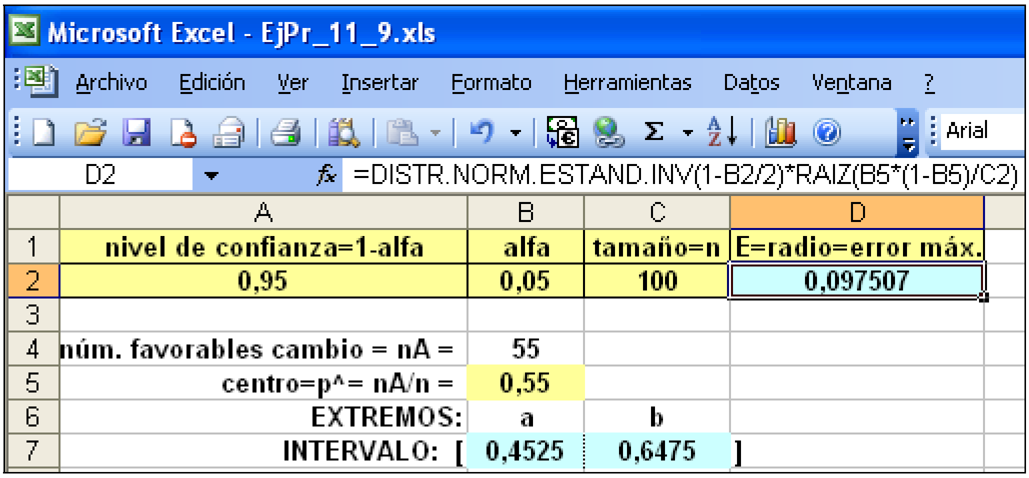
B2 =2*(1-DISTR.NORM.ESTAND(E2*RAIZ(D2)/C2))

	A	B	C	D	E
1	nivel de confianza=1-alfa	alfa	sigma	tamaño=n	E=radio=error máx.
2	0,95	0,05	0,1	49	0,028000
3					
4	centro=media muestral=	7,4000			
5	EXTREMOS:	a	b		
6	INTERVALO: [	7,3720	7,4280	]	

9 Un Instituto con 1000 alumnos tiene en su horario un único recreo de 30 minutos. Se estudia la posibilidad de cambiar a dos recreos de 15 minutos, para lo cual se consulta a una muestra de 100 alumnos y resulta que el 55% están a favor del cambio y el resto en contra. Hallar un intervalo de confianza, al 95%, para la proporción poblacional de alumnos favorables al cambio.

El intervalo es:

$$I=\left[\hat{p}-E,\,\hat{p}+E\right]=\left[\hat{p}-z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\,;\,\hat{p}+z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right]$$



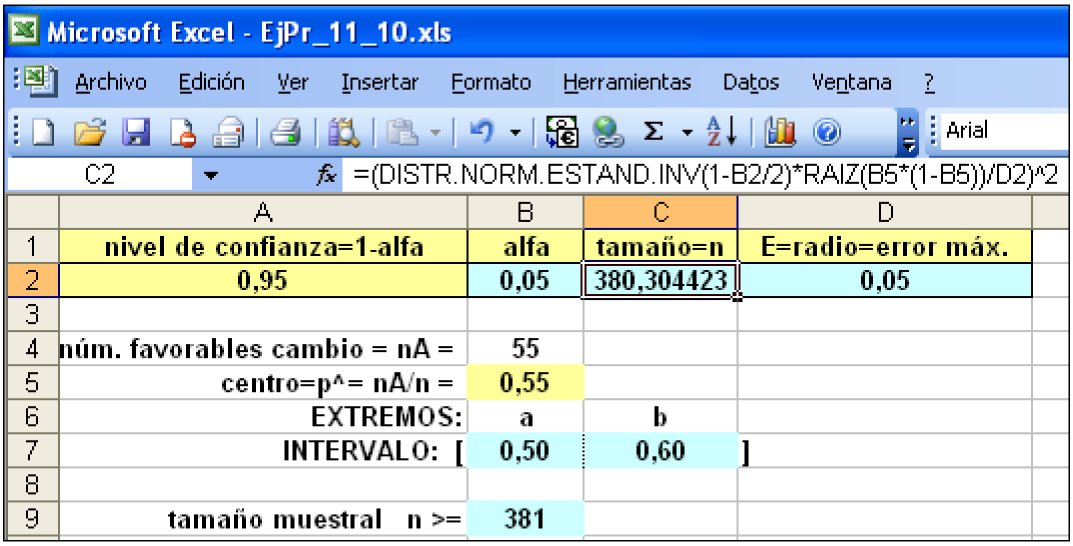
	A	B	C	D
1	nivel de confianza=1-alfa	alfa	tamaño=n	E=radio=error máx.
2	0,95	0,05	100	0,097507
3				
4	núm. favorables cambio = nA =	55		
5	centro=p^= nA/n =	0,55		
6	EXTREMOS:	a	b	
7	INTERVALO: [	0,4525	0,6475	]

10 En relación con el ejercicio anterior, ¿qué tamaño debería darse a la muestra para que el nivel de confianza de que fuese elegido el cambio sea del 95%?

El cambio será elegido si la proporción poblacional es mayor al 50%; es decir, si el extremo inferior, a, es mayor que 0,50. Entonces, $a = \hat{p} - E = 0{,}55 - E > 0{,}50$,

Por tanto, ha de ser $E < 0{,}05$. Así,

$$E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p}\cdot(1-\hat{p})}}{\sqrt{n}} < 0{,}05 \Rightarrow \sqrt{n} > z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p}\cdot(1-\hat{p})}}{0{,}05} \Rightarrow n > \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\hat{p}\cdot(1-\hat{p})}}{0{,}05} \right)^2$$



Microsoft Excel - EjPr_11_10.xls

C2 =(DISTR.NORM.ESTAND.INV(1-B2/2)*RAIZ(B5*(1-B5))/D2)^2

	A	B	C	D
1	nivel de confianza=1-alfa	alfa	tamaño=n	E=radio=error máx.
2	0,95	0,05	380,304423	0,05
3				
4	núm. favorables cambio = nA =	55		
5	centro=p^= nA/n =	0,55		
6	EXTREMOS:	a	b	
7	INTERVALO: [	0,50	0,60	]
8				
9	tamaño muestral n >=	381		