# Estimación puntual e intervalos de confianza



UNIVERSIDAD EAFIT®

# Contenido

Considere

$$\begin{array}{cccc}
y_1^1 & y_2^1 & \cdots & y_{10000}^1 \\
y_1^2 & y_2^2 & \cdots & y_{10000}^2 \\
y_1^3 & y_2^3 & \cdots & y_{10000}^3 \\
\vdots & \vdots & \ddots & \vdots \\
y_1^{100} & y_2^{100} & \cdots & y_{10000}^{100} \\
\\
\hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_{10000}
\end{array}$$

Entonces los $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_{10000}$ son 10000 estimaciones puntuales de la variable aleatoria $\hat{\Theta}$ que a su vez se distribuye con la función de distribución acumulada $F$. Es decir $\hat{\Theta} \sim F$. A $\hat{\Theta}$ se le llama **estadístico**, del parametro $\theta$, que es el parámetro real de la población.

Podríamos encontrar otro estadístico para $\theta$, llamesmoslo $\widetilde{\Theta}$ que se distribuye $\widetilde{\Theta} \sim G$

$$
\begin{array}{cccc}
y_1^1 & y_2^1 & \cdots & y_{10000}^1 \\
y_1^2 & y_2^2 & \cdots & y_{10000}^2 \\
y_1^3 & y_2^3 & \cdots & y_{10000}^3 \\
\vdots & \vdots & \ddots & \vdots \\
y_1^{100} & y_2^{100} & \cdots & y_{10000}^{100} \\
\widetilde{\theta}_1 & \widetilde{\theta}_2 & \cdots & \widetilde{\theta}_{10000}
\end{array}
$$

$\widetilde{\theta}_1, \widetilde{\theta}_2, \cdots, \widetilde{\theta}_{10000}$ son valores de la variable aleatoria o estadístico $\widetilde{\Theta} \sim G$.

En muchas investigaciones interesa determinar entre estos dos estadísticos cuál es mejor, si $\hat{\Theta}$ o $\widetilde{\Theta}$.

A continuación veremos formas generales de caracterizar un estadístidco.

Se dice que un estadístico $\hat{\Theta}$ es un **estimador insesgado** del parámetro $\theta$ si

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta.$$

Para el contexto mencionado anteriormente puede pasar que tanto $\hat{\Theta}$, como $\widetilde{\Theta}$ sean insesgados para el parámetro $\theta$. Pero, ¿Cómo escoger el mejor?

Si consideramos todos los posibles estimadores de algún parámetro $\theta$, al que tiene la menor varianza lo llamaremos estimador más eficaz de $\theta$.
Esta podría ser una buena alternativa para escoger el mejor. Es decir, decimos que $\hat{\Theta}$ es más eficaz que $\widetilde{\Theta}$ si

$$Var(\hat{\Theta}) < Var(\widetilde{\Theta})$$

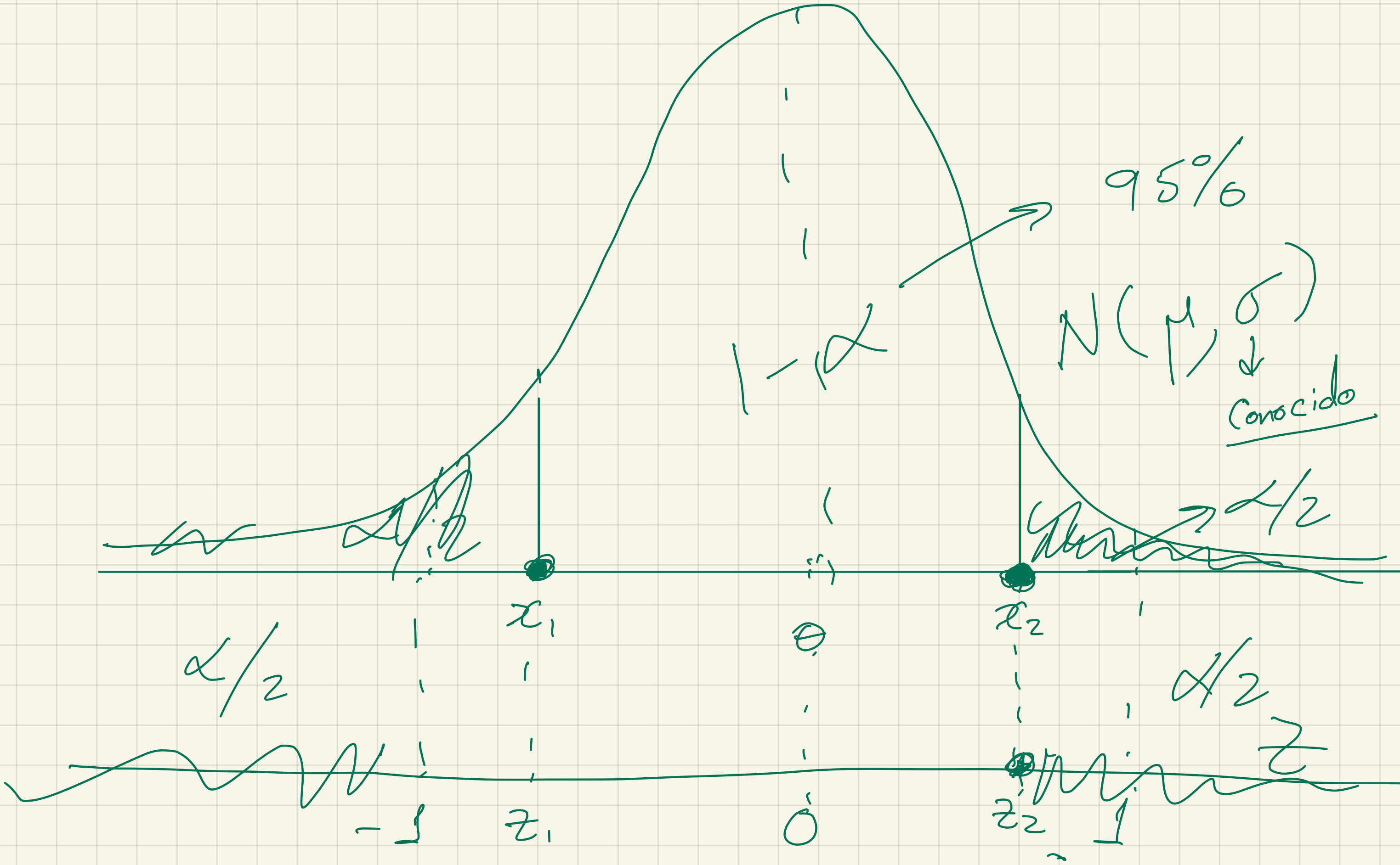
## Estimación de intervalos de confianza

Si $\bar{x}$ es la media de una muestra aleatoria de tamaño $n$ de una población de la que se conoce su varianza $\sigma^2$, lo que da un intervalo de confianza de $100(1-\alpha)\,\%$ para $\mu$ es

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

donde $z_{\alpha/2}$ es el valor $z$ que deja un área de $\alpha/2$ a la derecha. Denotaremos:

$$\hat{\theta}_L = \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ y } \hat{\theta}_U = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

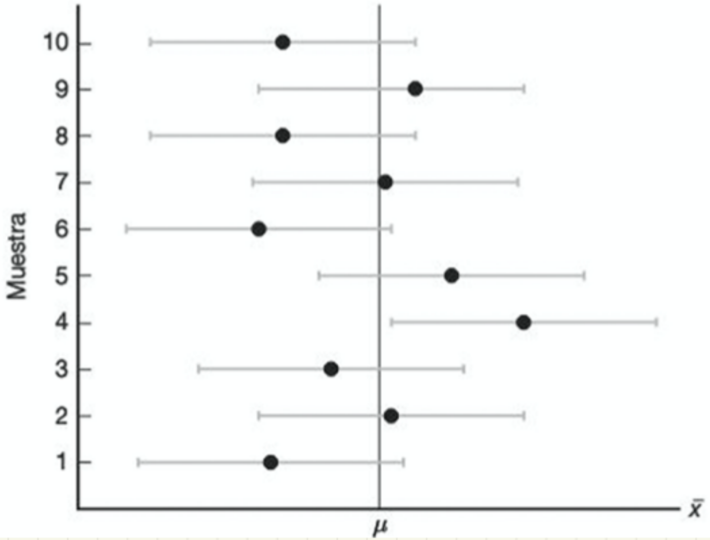



$$x_2 = \bar{x} + z_2 \frac{\sigma}{\sqrt{n}}$$

donde

$$z_2 = Z_{1-\alpha/2}$$

$$x_1 = \bar{x} - z_2 \frac{\sigma}{\sqrt{n}}$$


Muestra
10
9
8
7
6
5
4
3
2
1
$\mu$
$\bar{x}$

**Teorema:**
**Si utilizamos $\bar{x}$ como una estimación de $\mu$, podemos tener $100(1-\alpha)\,\%$ de confianza en que el error no excederá a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.**

**Teorema: Si usamos $\bar{x}$ como una estimación de $\mu$, podemos tener $100(1-\alpha)\,\%$ de confianza en que el error no excederá a una cantidad específica $e$ cuando el tamaño de la muestra sea**

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

$$\sqrt{n} = \frac{z_{\alpha/2}\,\sigma}{e} \Longrightarrow e = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$$

Shara $\begin{cases} 95\% & (2.5\ ,\ 2.7) \\ 99\% & (2.46\ ,\ 2.7362) \end{cases}$ 1 décima

## Ejemplo

Manuela $\begin{cases} 95\% & (2.498, 2.701) \\ 99\% & (2.46\ ,\ 2.736) \end{cases}$ 1 décima

Se encuentra que la concentración promedio de zinc que se obtiene en una muestra de mediciones en 36 sitios diferentes de un río es de 2.6 gramos por mililitro. Calcule los intervalos de confianza del 95 % y 99 % para la concentración media de zinc en el río. Suponga que la desviación esntándar de la población es de 0.3 gramos por mililitro.

Mario José $\begin{cases} 95\% & (2.50\ ,\ 2.698) \\ 99\% & (2.47\ ,\ 2.728) \end{cases}$ 3 décimas

$$(1-\alpha)100\% = 95\% \Longrightarrow 1-\alpha = 0.95$$

$$\boxed{\alpha = 0.05}$$

$$\bar{X} = 2.6\, g/ml \qquad n = 36$$

$$\frac{\alpha}{2} = 0.025$$

$$\sigma = 0.3\, g/ml$$

$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

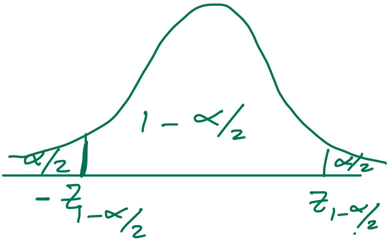
$$a_L = \bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$= 2.6 - \text{qnorm}(1-\alpha/2)\,\frac{0.3}{\sqrt{36}}$$

$$= 2.6 - \text{qnorm}(0.975)\cdot\frac{0.3}{6}$$

$$= 2.6 - 1.96\left(\frac{0.3}{6}\right)$$

$$= 2.502$$



$$a_u = \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$= 2.6 + 1.96\left(\frac{0.3}{6}\right)$$

$$= 2.698$$

$$99\% = (1-\alpha)\,100\%$$

$$1-\alpha = 0.99$$

$$\alpha = 0.01$$

$$\frac{\alpha}{2} = 0.005$$

$$z_{1-\alpha/2} = z_{0.995} = \text{qnorm}(0.995) = 2.57$$

$$a_L = \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = qnorm\left(\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}}\right)$$

$$= qnorm\left(0.005, \bar{x}, \frac{\sigma}{\sqrt{n}}\right)$$

$$= 2.476$$

$$a_u = \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = qnorm\left(1-\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}}\right)$$

$$= qnorm\left(0.995, \bar{x}, \frac{\sigma}{\sqrt{n}}\right)$$

$$= 2.728$$

| Confianza | $d_L$ | $d_U$ | ME |
|---|---|---|---|
| 90% | 2.518 | 2.68 | 0.08224 |
| 95% | 2.502 | 2.698 | 0.09799 |
| 99% | 2.476 | 2.728 | 0.128791 |

## Ejemplo

Camila $n = 138.29 \approx 138$

¿Qué tan grande debe ser la muestra del ejemplo anterior si queremos tener 95 % de confianza en que nuestra estimación de $\mu$ diferirá por menos de 0.05?

$$n = \left(\frac{z_{1-\alpha/2}\,\sigma}{e}\right)^2 = \left(\frac{1.96 \cdot (0.3)}{0.05}\right)^2$$

$$= 138.29$$

$$P\left(|\bar{x}-\mu|<0.05\right)=0.95$$

$$P\left(-0.05<\bar{x}-\mu<0.05\right)=0.95$$

$$P\left(\frac{-0.05}{\sigma_{\bar{x}}}<z<\frac{0.05}{\sigma_{\bar{x}}}\right)=0.95$$

$$P\left(\frac{-0.05}{\frac{0.3}{\sqrt{n}}}<z<\frac{0.05}{\frac{0.3}{\sqrt{n}}}\right)=0.95$$

$$P\left(-\frac{0.05\sqrt{n}}{0.3}<z<\frac{0.05\sqrt{n}}{0.3}\right)=0.95$$

$$1 - 2F\left(-\frac{0.05\sqrt{n}}{0.3}\right) = 0.95$$

$$\frac{1-0.95}{2} = F\left(-\frac{0.05\sqrt{n}}{0.8}\right)$$

$$0.025 = F\left(-\frac{0.05\sqrt{n}}{0.3}\right)$$

$$\text{qnorm}(0.025) = -\frac{0.05\sqrt{n}}{0.3}$$

$$0.3\,\text{qnorm}(0.025) = -0.05\sqrt{n}$$

$$\frac{0.3\,\text{qnorm}(0.025)}{-0.05} = \sqrt{n}$$

$$\left(\frac{0.3\,\text{qnorm}(0.025)}{-0.05}\right)^2 = n = 138.29$$

## Límites de confianza unilaterales de $\mu$ cuando se conoce el valor $\sigma^2$

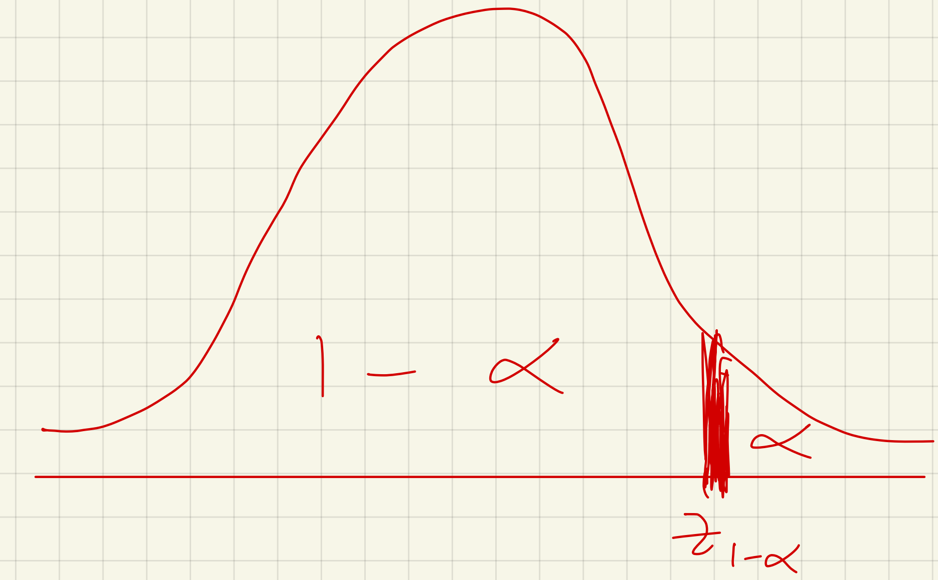
Si $\bar{X}$ es la media de una muestra aleatoria de tamaño $n$ a partir de una población con varianza $\sigma^2$, los límites de confianza unilaterales del $100(1-\alpha)\,\%$ para $\mu$ son dados por

1 límite unilateral superior: $\bar{x} + z_\alpha \sigma/\sqrt{n}$;

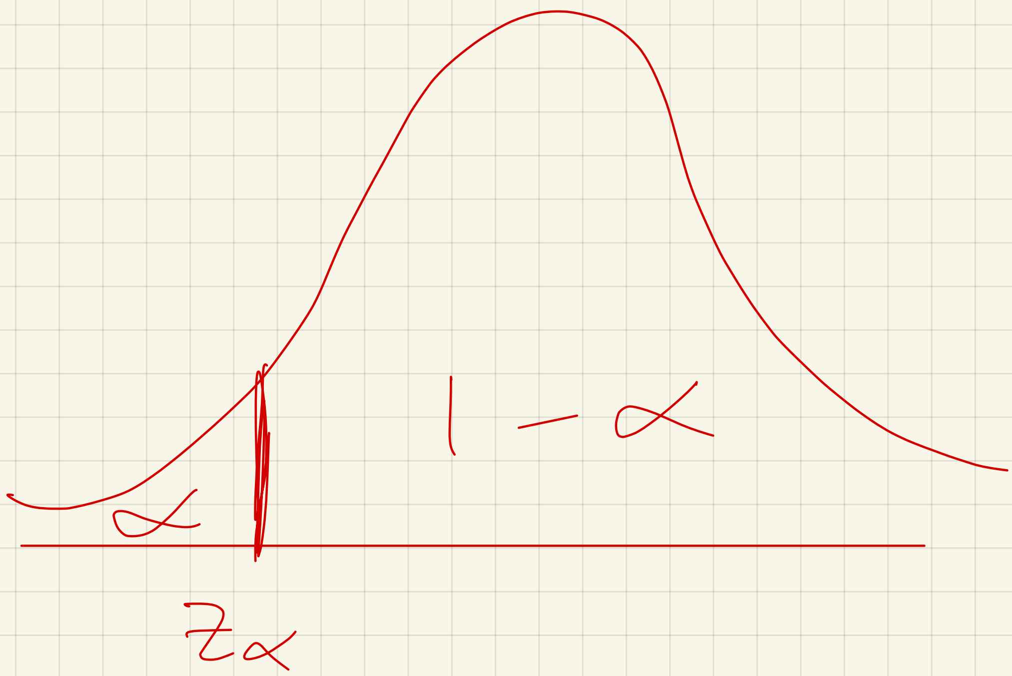
2 límite unilateral inferior: $\bar{x} - z_\alpha \sigma/\sqrt{n}$.

$$1 = \bar{X} + \text{qnorm}(1-\alpha)\frac{\sigma}{\sqrt{n}}$$

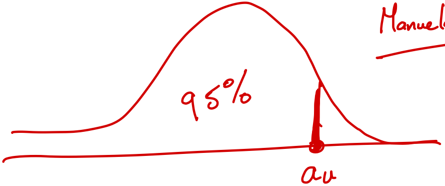
$$2 = \bar{X} - \text{qnorm}(\alpha)\frac{\sigma}{\sqrt{n}}$$



1 − α
α
z₁₋α

α
1 − α
zα




Manuela 6857

2 décimas.

## Ejemplo

En un experimento de pruebas psicológicas se seleccionan al azar 25 sujetos y se miden sus tiempos de reacción, en segundos, ante un estímulo particular. La experiencia sugiere que la varianza en los tiempos de reacción es ~~aproximadamente normal.~~ $4\ seg^2$ El tiempo promedio para los sujetos fue de 6.2 segundos. Calcule el límite superior del 95 % para el tiempo medio de reacción.

$$a_u = \bar{x} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

$$= 6.2 + z_{0.95}\frac{2}{5} = 6.2 + (1.64)\cdot\frac{2}{5}$$

$$= 6.857 \approx 6.86$$

## Intervalo de confianza para $\mu$ cuando se desconoce $\sigma^2$

Si $\bar{x}$ y $s$ son la media y la desviación estándar de una muestra aleatoria de una población normal de la que se desconoce la varianza $\sigma^2$, un intervalo de confianza del $100(1-\alpha)\,\%$ para $\mu$ es

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

donde $t_{\alpha/2}$ es el valor $t$ con $v = n - 1$ grados de libertad que deja una área de $\alpha/2$ a la derecha.

## Ejemplo

El contenido de ácido sulfúrico de 7 contenedores similares es de 9.8, 10.2, 10.4, 9.8,10.0, 10.2, y 9.6 litros. Calcule un intervalo de confianza del 95 % para el contenido promedio de todos los contenedores suponiendo una distribución aproximadamente normal.

$$\bar{X} = 10 \quad n = 7 \quad s = 0.2828427 \quad \alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

$$a_L = \bar{X} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

$$= 10 - qt(1-\alpha/2, 6) \cdot \frac{0.2828427}{\sqrt{7}}$$

$$= 9.74$$

$$a_u = \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

$$= 10 + qt(1-\alpha/2, 6) \cdot \frac{0.2828427}{\sqrt{7}}$$

$$= 10.26159$$

## Intervalo de confianza para $p$ de una muestra grande

Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño $n$, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado del $100(1-\alpha)\,\%$ para el parámetro binomial $p$ se obtiene por medio de (método 1)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

o mediante (método 2)

$$\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} - \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}} < p < \frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} + \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}}$$

donde $z_{\alpha/2}$ es el valor $z$ que deja una área de $\alpha/2$ a la derecha.

## Ejemplo

En una muestra aleatoria de $n = 500$ familias que tiene televisores en la ciudad de Hamilton, Canadá, se encuentra que $x = 340$ están suscritas a HBO. Calcule un intervalo de confianza del 95 % para la proporción real de familias que tienen televisores en esta ciudad y están suscritas a HBO.

**Teorema:** Si $\hat{p}$ se utiliza como un estimado de $p$, podemos tener un $100(1-\alpha)\,\%$ de confianza en que el error no excederá a $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

**Teorema:** Si $\hat{p}$ se utiliza como un estimado de $p$, podemos tener un $100(1-\alpha)\,\%$ de confianza en que el error será menor que una cantidad específica $e$ cuando el tamaño de la muestra sea aproximadamente

$$n = \frac{z_{\alpha/2}^{2}\hat{p}\hat{q}}{e^{2}}.$$

**Ejemplo: ¿Qué tan grande debe de ser una muestra en el ejemplo anrerior si queremos tener un $95\,\%$ de confianza en que la estimación de $p$ esté dentro de 0.02 del valor verdadero?**

# Intervalo de confianza para $\sigma^2$

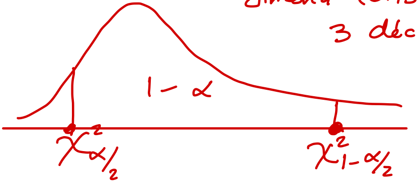
Si $s^2$ es la varianza de una muestra aleatoria de tamaño $n$ de una población normal, un intervalo de confianza del $100(1-\alpha)\,\%$ para $\sigma^2$ es

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores $\chi^2$ con $v = n-1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1-\alpha/2$, respectivamente, a la derecha.

Jimena (0.1354, 0.954)
3 décimas

## Ejemplo




Los siguientes son los pesos, en decagramos, de 10 paquetes de semillas de pasto distribuidas por cierta empresa: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2 y 46.0. Calcule un intervalo de confianza del 95 % para la varianza de todos los pesos de este tipo de paquetes de semillas de pasto distribuidos por la empresa. Suponga una población normal.

$S^2 = 0.286 \qquad n = 10 \qquad 1-\alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$\chi^2_{1-\alpha/2} = \chi^2_{0.975} = qchisq(0.975, 9) = 2.7$

$\chi^2_{\alpha/2} = \chi^2_{0.025} = qchisq(0.025, 9) = 19.02$

$a_L = \frac{9(0.286)}{19.02} = 0.135$

$a_u = \frac{9(0.286)}{2.7} = 0.953$

$0.135 < \sigma^2 < 0.953$