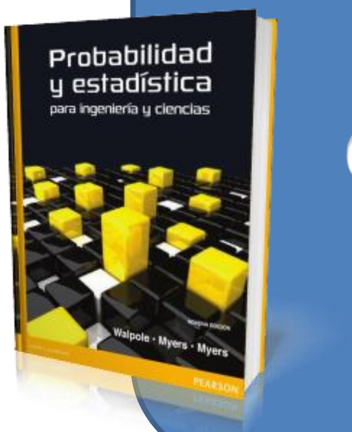


Tema 8: Problemas de estimación de una y dos muestras

w – Cap 9

Se profundizarán los conceptos dados en el tema anterior acerca de la inferencia estadística de los parámetros poblacionales a partir de muestras.

- Métodos clásicos de estimación: Puntual y por intervalos de clase.
- Error de estimación puntual y cálculo del tamaño de la muestra según el error que se fije.
- Límites de tolerancia.

- Inferencia de la media y de la diferencia de medias.
- Estimación de una proporción y de la diferencia de dos proporciones.
- Estimación de la varianza y del cociente de varianzas por intervalos de clase.

Estimación de μ usando $\overline{x}$ y la $\sim N(0{,}1)$ cuando σ es conocido:

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

Cuando x tiene $\sim N(\mu, \sigma)$ aunque n < 30
x no tiene $\sim N(\mu, \sigma)$ pero n ≥ 30

Estimación de μ usando $\overline{x}$ y la $\sim t(0, \nu)$ cuando σ es desconocido:

$$T = \frac{\overline{x} - \mu}{S / \sqrt{n}}$$

Cunado x tiene $\sim N(\mu, \sigma)$ aunque n < 30
x no tiene $\sim N(\mu, \sigma)$ pero n ≥ 30

Estimación de σ^2 usando S^2 y la $\sim \chi^2(\nu, \alpha)$:

$$V = \frac{\nu . S^2}{\sigma^2}$$

Cuando x tiene $\sim N(\mu, \sigma)$

Estimación de $\mu_1 - \mu_2$ usando $\overline{x}_1 - \overline{x}_2$ y la $\sim N(0,1)$ cuando σ_1 y σ_2 son conocidos:

$$Z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1}\sigma_1^2 + \frac{1}{n_2}\sigma_2^2}}$$

Cuando x tiene $\sim N(\mu, \sigma)$ aunque n < 30

x no tiene $\sim N(\mu, \sigma)$ pero n ≥ 30

Estimación de $\mu_1 - \mu_2$ usando $\overline{x}_1 - \overline{x}_2$ y la $\sim t(0, \nu)$ cuando σ_1 y σ_2 son desconocidos:

$$T = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1}S_1^2 + \frac{1}{n_2}S_2^2}}$$

Cuando x tiene $\sim N(\mu, \sigma)$ aunque n < 30

x no tiene $\sim N(\mu, \sigma)$ pero n ≥ 30

Estimación de ${\sigma_1^2}/{\sigma_2^2}$ usando y la $\sim f(\nu_1, \nu_2)$:

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

Cuando x tiene $\sim N(\mu, \sigma)$

Métodos Clásicos de Estimación

Se basan en los datos de una muestra seleccionada aleatoriamente.

a) Estimación Puntual

b) Estimación por Intervalos de Confianza

a) Estimación Puntual

Estima los parámetros poblacionales a partir de un estadístico de la muestra

Por ejemplo:

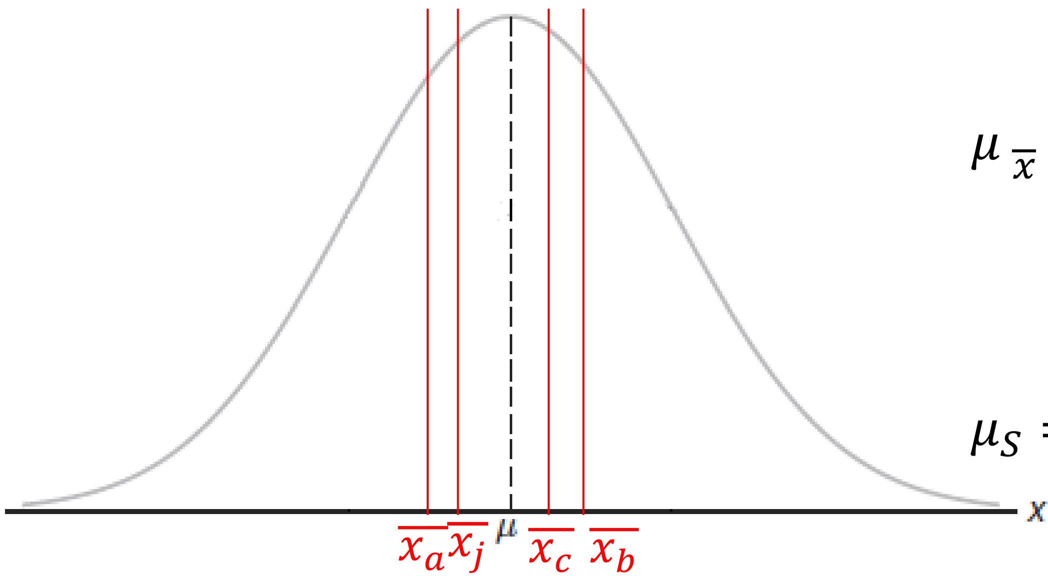
- Estimo μ a partir de $\overline{x}$.
- Estimo μ a partir de $\tilde{x}$.
- Estimo σ^2 a partir de S^2.
- Estimo σ^2 a partir de $Q_3 - Q_1$.

Llamemos $\boldsymbol{\widehat{\theta}}$ al estadístico de la muestra y $\boldsymbol{\theta}$ al parámetro poblacional a estimar.

No se espera que un estimador $\hat{\theta}$ $(\overline{x}, \tilde{x}, S^2, RIC)$ logre estimar el parámetro de la población θ sin error, lo que en realidad se espera es que no este muy alejado. Por ejemplo, no se espera que $\overline{x}$ estime μ con exactitud.

<u>Estimador insesgado</u>: Es aquel que cumple la condición:

$$\boxed{\mu_{\hat{\theta}} = E(\hat{\theta}) = \theta}$$



$$\mu_{\overline{x}} = \frac{\sum \overline{x}}{n} = E(\overline{x}) = \mu$$

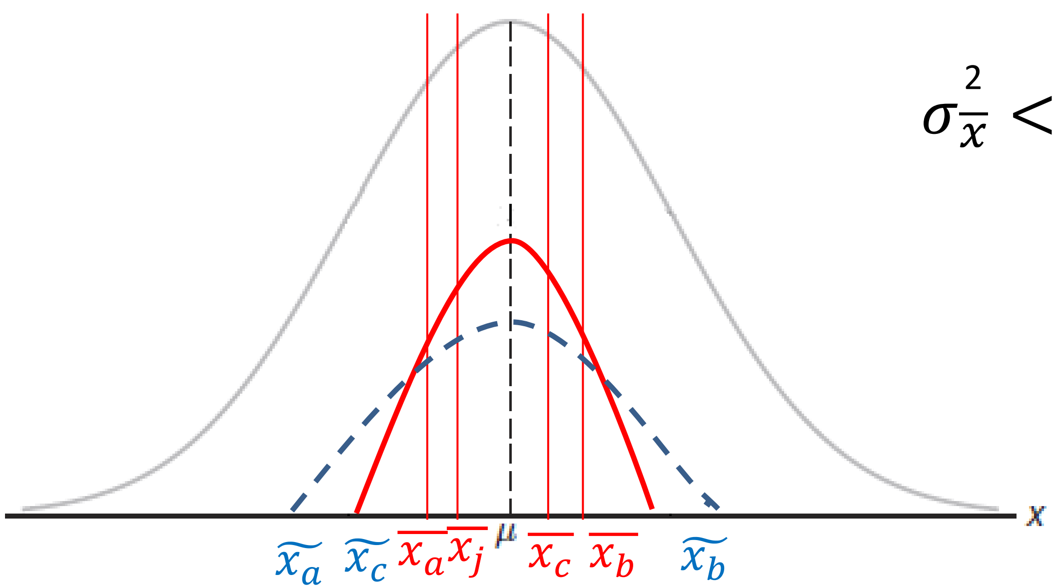
$$\mu_S = E(S) \neq \sigma$$

S^2 es un estimador insesgado de σ^2 pero S suele ser un estimador sesgado de σ, un sesgo que en el caso de muestras grandes se vuelve insignificante. Por esto es que para el cálculo de σ se debe dividir la expresión $\sum(x_i - \overline{x})^2$ por $(n-1)$ en vez de n, el último valor de x_i queda fijado por la $\overline{x}$ y no puede tomar cualquier valor.

Varianza de un Estimador puntual:

Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores insesgados del parámetro poblacional θ elegimos el estimador cuya distribución muestral tenga menor varianza.

$$\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2} \Longrightarrow$$ $\hat{\theta}_1$ es un estimador más eficiente de θ que $\hat{\theta}_2$



$$\sigma^2_{\overline{x}} < \sigma^2_{\tilde{x}}$$

Entonces, $\overline{x}$ es mejor estimador de μ que $\tilde{x}$

b) Estimación por intervalos de confianza

Como sabemos que con el estimador no pretendemos obtener el valor exacto del parámetro poblacional al estimar, este método propone fijar un nivel o grado de confianza $(1-\alpha)$, donde α es el grado de significancia.

De esta forma, a la estimación del parámetro poblacional θ la podemos **plantear como una probabilidad** de que se encuentre entre ciertos límites $\hat{\theta}_1$ y $\hat{\theta}_2$ con una confianza del $(1-\alpha)$.

$$P\left(\hat{\theta}_1 < \theta < \hat{\theta}_2\right) = 1-\alpha$$

Planteado como probabilidad con:

$$IC = \left[\hat{\theta}_1 < \theta < \hat{\theta}_2\right]$$

Intervalo de confianza y:

$$LC = \left[\hat{\theta}_1, \hat{\theta}_2\right]$$

Límites de confianza

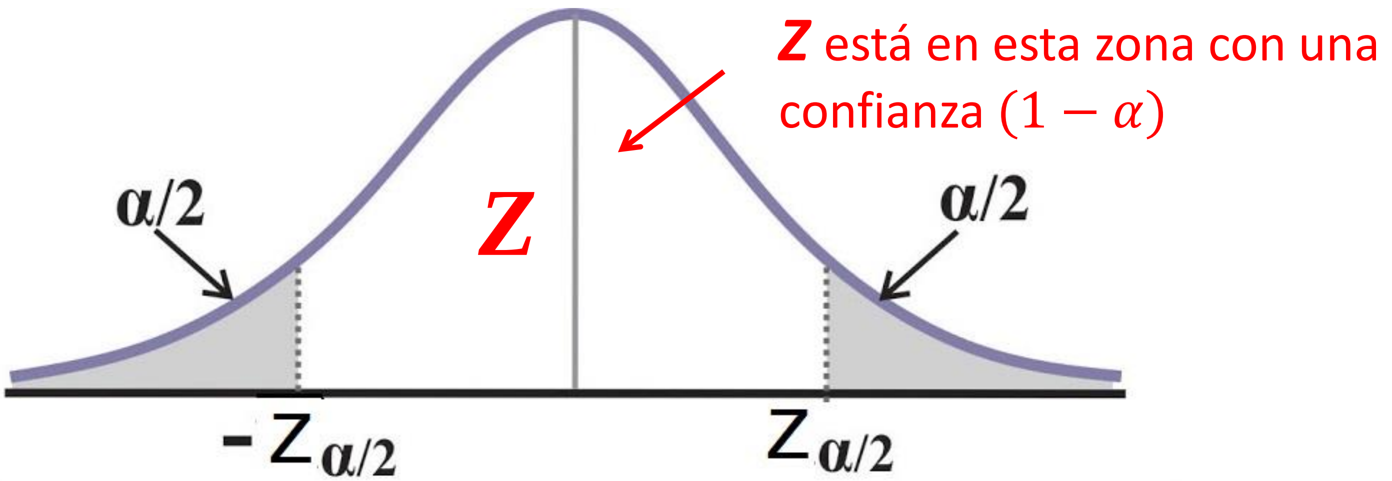
Ejemplo 1: Estimar μ a partir del estimador $\overline{x}$ por IC con $(1-\alpha)$ nivel de confianza donde σ es conocido.

Solución: Como σ es conocido se puede plantear el Teorema del Límite Central

$$Z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

Que planteado como probabilidad entre dos valores de la variable estandarizada Z con α nivel de significancia resulta:

$$P\left(-Z\alpha_{/2} < Z < Z\alpha_{/2}\right) = 1 - \alpha$$



$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) = 1 - \alpha$$

Multiplicando m.a.m. por $\sigma/\sqrt{n}$ y sumando $\bar{x}$

$$P\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \underbrace{\bar{x} + Z\frac{\sigma}{\sqrt{n}}}_{\mu} < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

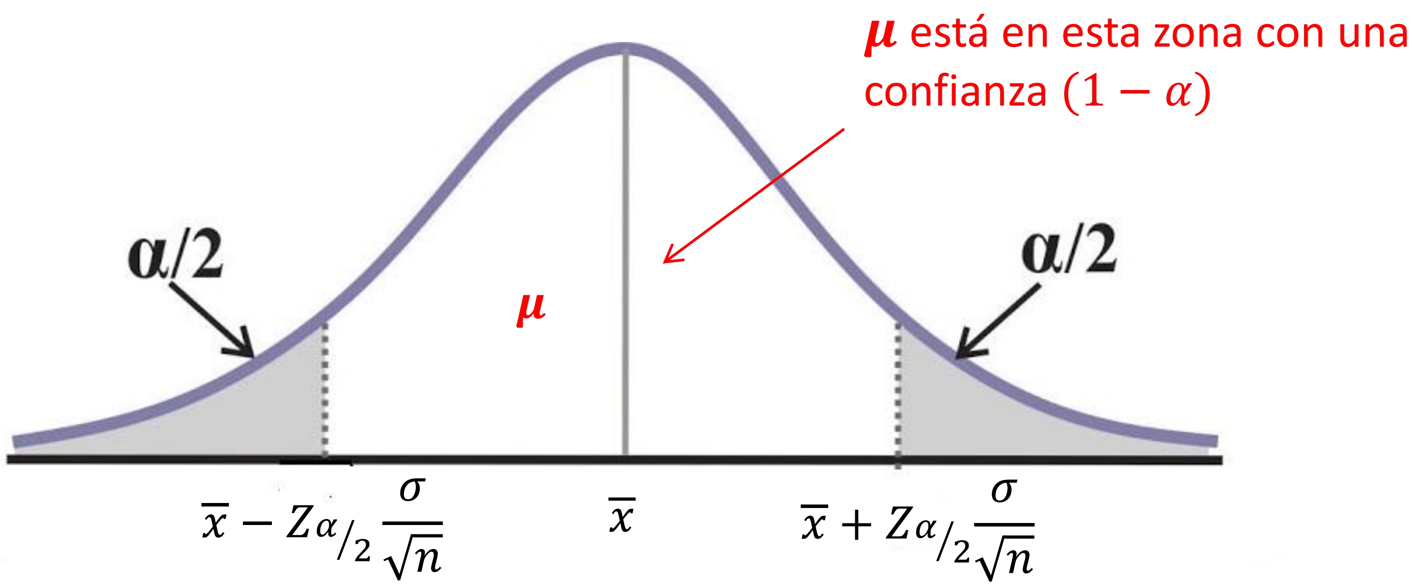
$$\boxed{P\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha}$$

$$\boxed{IC = \left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Intervalo de confianza

$$\boxed{LC = \left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

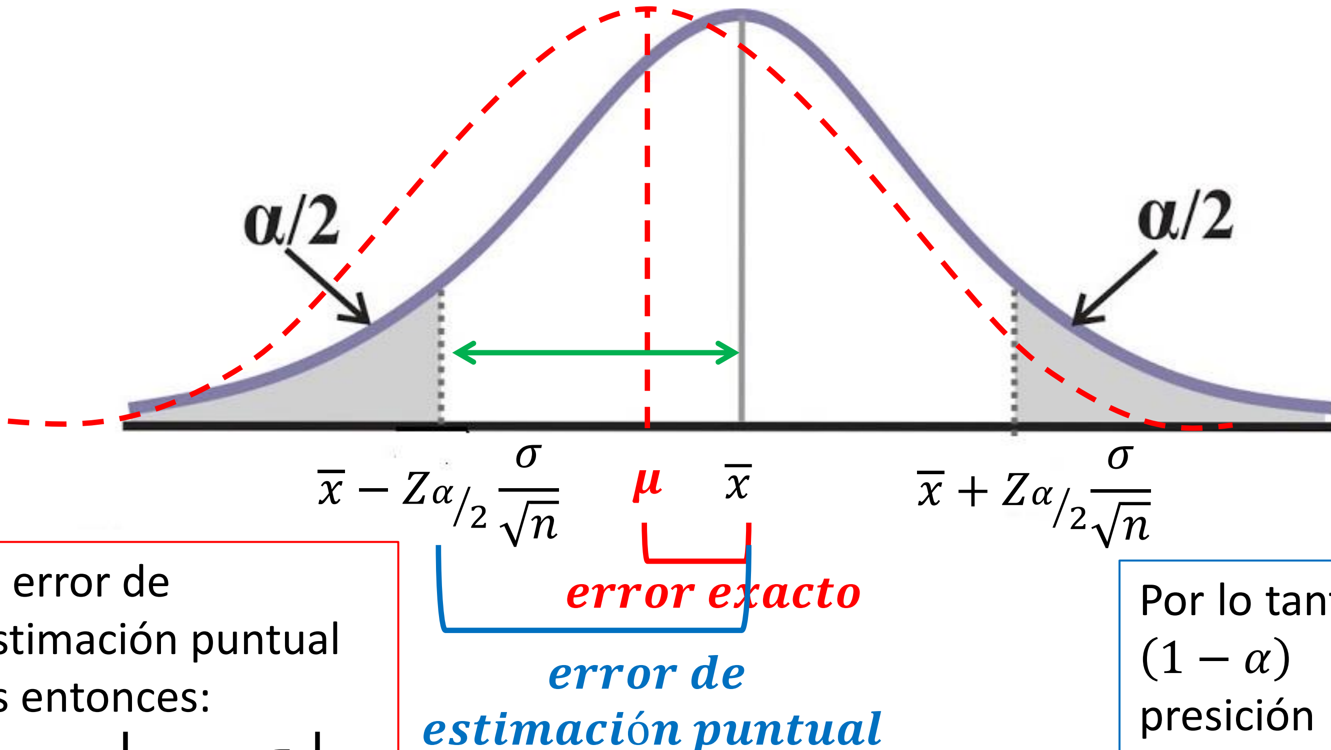
Límites de confianza



En el caso de que σ sea desconocido y $n < 30$ reemplazo Z por T y utilizo la $\sim t$

$$\boxed{IC = \left[\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]}$$

Error de estimación puntual



El error de estimación puntual es entonces:

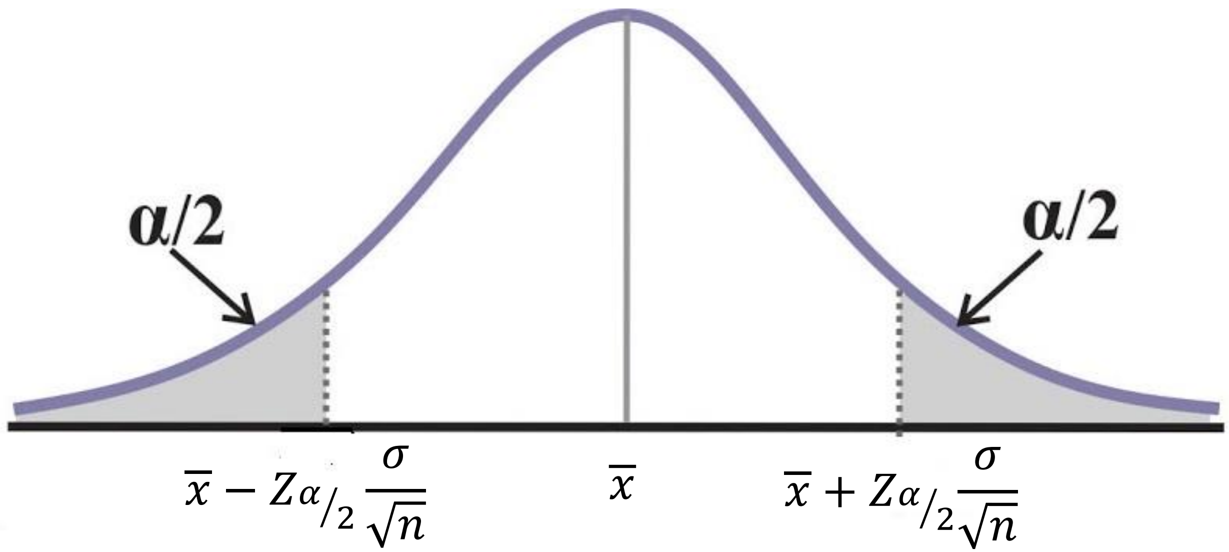
$$e = \left| Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right|$$

Por lo tanto, el nivel de confianza $(1-\alpha)$ proporciona una presición de nuestra estimación puntual.

Si $(1-\alpha) = 0{,}95$ entonces tengo una confianza de 95% que el error de estimación de μ no excede de $Z_{0{,}025}\frac{\sigma}{\sqrt{n}}$.

Si $(1-\alpha) = 0{,}99$ entonces tengo una confianza de 99% que el error de estimación de μ no excede de $\pm Z_{0{,}005}\frac{\sigma}{\sqrt{n}}$.

¿Cuál de los dos casos es más exigente, el 95% o el 99%?



El caso de α más pequeño es el menos exigente porque las zonas de rechazo son más pequeñas.

Entonces, el 95% de confianza es más exigente que el 99% de confianza.

Cálculo del tamaño de la muestra con un error fijado

Ahora, con una idea de los errores de estimación que se comenten según el nivel de confianza podemos preguntarnos ¿qué tan grande debe ser la muestra para cometer un error menor que un valor fijado (por ejemplo por alguna reglamentación)?

Podemos calcular el tamaño de la muestra *n* de antemano para asegurar que nuestro error de estimación puntual no excederá de un valor *e* establecido.

$$e = \left| Z\alpha_{/2} \frac{\sigma}{\sqrt{n}} \right| \xrightarrow{\text{Despejamos } n} \boxed{n = \left(\frac{Z\alpha_{/2} . \sigma}{e} \right)^2}$$

Error estándar de estimación puntual

$s.e.$ *(standard error)* es el desvío estándar del error de estimación puntual. En el caso de estimar μ con $\bar{x}$ y σ es conocido, el $Z_{\alpha/2}$ está fijado por α, entonces el error varía con $\frac{\sigma}{\sqrt{n}}$, por lo tanto:

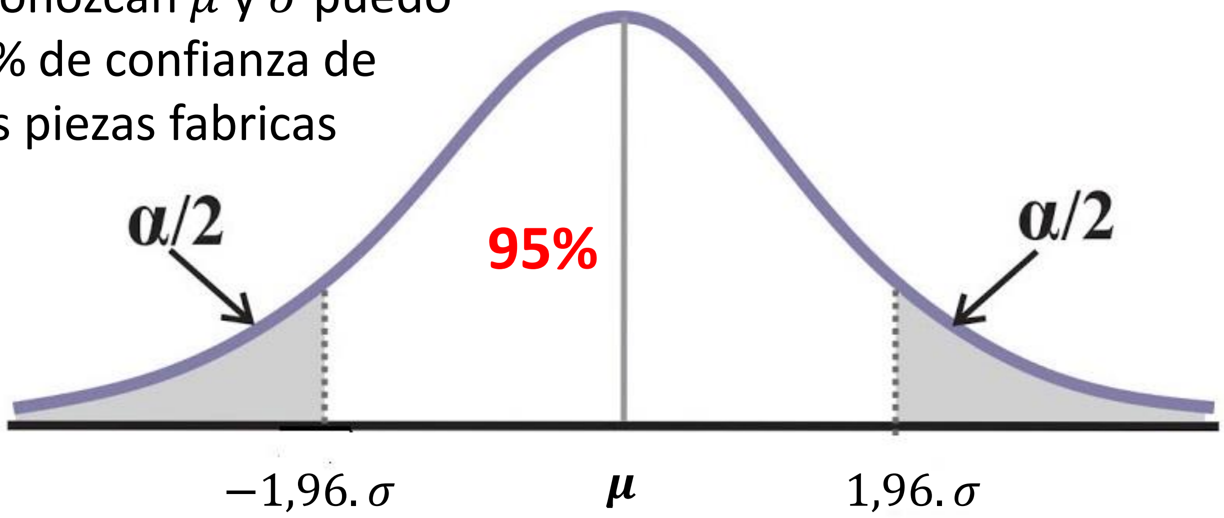
$$\boxed{s.e. = \frac{\sigma}{\sqrt{n}}}$$

Límites de tolerancia

Ahora en lugar de preocuparnos en estimar μ, queremos determinar límites de tolerancia para aceptar o rechazar elementos, por ejemplo, en un proceso de control de calidad.

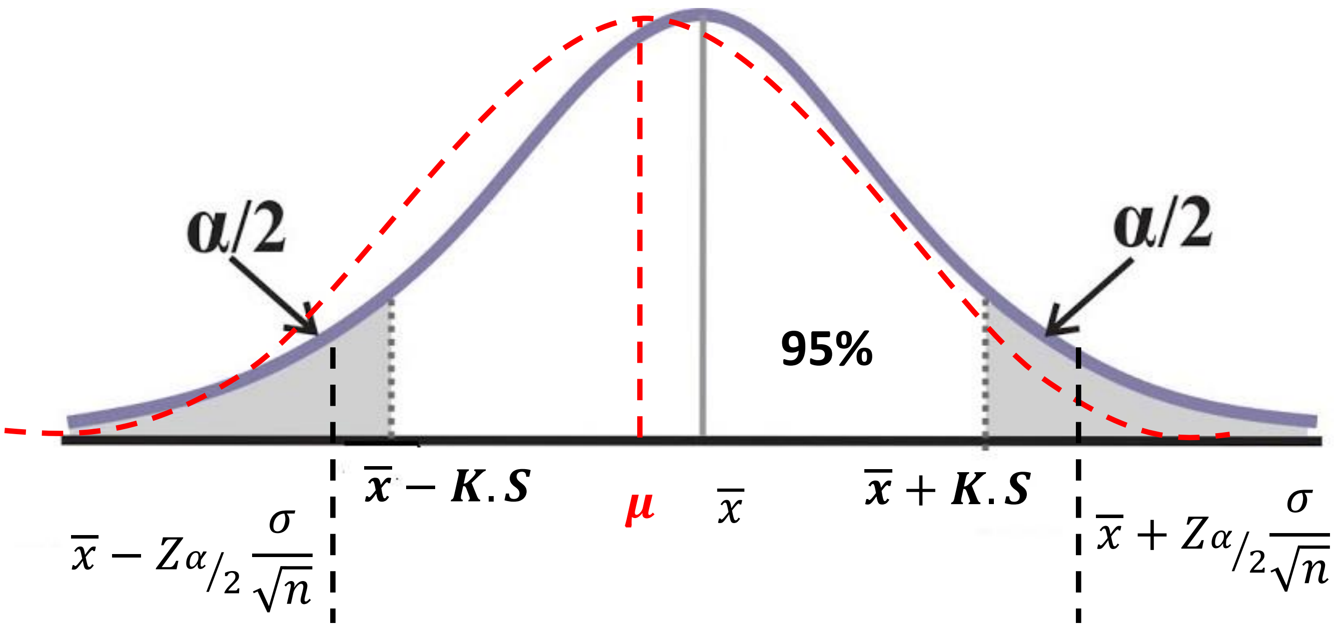
Se plantea que con una confianza de $(1-\gamma)\%$ se acepten $(1-\alpha)\%$ de los elementos realizando mediciones de los mismos en un sistema de control.

Para el caso que se conozcan μ y σ puedo proponerme un 100% de confianza de aceptar el 95% de las piezas fabricas



Esto es porque sé exactamente cuál es el valor de μ y de σ de ls piezas que se están fabricando

En el caso de que no sé exactamente o no estoy seguro de los parámetros μ y σ voy a tener que calcular los límites de tolerancia con un nivel de confianza $(1-\gamma)$.



En este ejemplo tengo un $(1-\gamma)\% = 95\%$ de confianza que μ se encuentre en el intervalo $\overline{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Necesito tomar una muestra de *n* elementos para calcular S y estimar K.

Si, por ejemplo, deseo aceptar el 99% de las piezas, el valor de K me asegura de aceptar el $(1-\alpha)\% = 99\%$ de las piezas con un $(1-\gamma)\% = 95\%$ de confianza.

La Tabla A.7 permite obtener los valores de K para el cálculo de los límites de tolerancia: $\boxed{L.T. = \overline{x} \pm K.S}$

Tabla A.7 Factores de tolerancia para distribuciones normales

	Intervalos bilaterales						Intervalos unilaterales					
	$\gamma = 0.05$			$\gamma = 0.01$			$\gamma = 0.05$			$\gamma = 0.01$		
	$1-\alpha$			$1-\alpha$			$1-\alpha$			$1-\alpha$		
n	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
2	32.019	37.674	48.430	160.193	188.491	242.300	20.581	26.260	37.094	103.029	131.426	185.617
3	8.380	9.916	12.861	18.930	22.401	29.055	6.156	7.656	10.553	13.995	17.170	23.896
4	5.369	6.370	8.299	9.398	11.150	14.527	4.162	5.144	7.042	7.380	9.083	12.387
5	4.275	5.079	6.634	6.612	7.855	10.260	3.407	4.203	5.741	5.362	6.578	8.939
6	3.712	4.414	5.775	5.337	6.345	8.301	3.006	3.708	5.062	4.411	5.406	7.335
7	3.369	4.007	5.248	4.613	5.488	7.187	2.756	3.400	4.642	3.859	4.728	6.412
8	3.136	3.732	4.891	4.147	4.936	6.468	2.582	3.187	4.354	3.497	4.285	5.812
9	2.967	3.532	4.631	3.822	4.550	5.966	2.454	3.031	4.143	3.241	3.972	5.389
10	2.839	3.379	4.433	3.582	4.265	5.594	2.355	2.911	3.981	3.048	3.738	5.074
11	2.737	3.259	4.277	3.397	4.045	5.308	2.275	2.815	3.852	2.898	3.556	4.829
12	2.655	3.162	4.150	3.250	3.870	5.079	2.210	2.736	3.747	2.777	3.410	4.633
13	2.587	3.081	4.044	3.130	3.727	4.893	2.155	2.671	3.659	2.677	3.290	4.472
14	2.529	3.012	3.955	3.029	3.608	4.737	2.109	2.615	3.585	2.593	1.189	4.337
15	2.480	2.954	3.878	2.945	3.507	4.605	2.068	2.566	3.520	2.522	3.102	4.222
16	2.437	2.903	3.812	2.872	3.421	4.492	2.033	2.524	3.464	2.460	3.028	4.123
17	2.400	2.858	3.754	2.808	3.345	4.393	2.002	2.486	3.414	2.405	2.963	4.037
18	2.366	2.819	3.702	2.753	3.279	4.307	1.974	2.453	3.370	2.357	2.905	3.960
19	2.337	2.784	3.656	2.703	3.221	4.230	1.949	2.423	3.331	2.314	2.854	3.892
20	2.310	2.752	3.615	2.659	3.168	4.161	1.926	2.396	3.295	2.276	2.808	1.832
25	2.208	2.631	3.457	2.494	2.972	3.904	1.838	2.292	3.158	2.129	2.633	3.001
30	2.140	2.549	3.350	2.385	2.841	3.733	1.777	2.220	3.064	2.030	2.516	3.447
35	2.090	2.490	3.272	2.306	2.748	3.611	1.732	2.167	2.995	1.957	2.430	3.334
40	2.052	2.445	3.213	2.247	2.677	3.518	1.697	2.126	2.941	1.902	2.364	3.249
45	2.021	2.408	3.165	2.200	2.621	3.444	1.669	2.092	2.898	1.857	2.312	3.180
50	1.996	2.379	3.126	2.162	2.576	3.385	1.646	2.065	2.863	1.821	2.269	3.125
60	1.958	2.333	3.066	2.103	2.506	3.293	1.609	2.022	2.807	1.764	2.202	3.038
70	1.929	2.299	3.021	2.060	2.454	3.225	1.581	1.990	2.765	1.722	2.153	2.974
80	1.907	2.272	2.986	2.026	2.414	3.173	1.559	1.965	2.733	1.688	2.114	2.924
90	1.889	2.251	2.958	1.999	2.382	3.130	1.542	1.944	2.706	1.661	2.082	2.883
100	1.874	2.233	2.934	1.977	2.355	3.096	1.527	1.927	2.684	1.639	2.056	2.850
150	1.825	2.175	2.859	1.905	2.270	2.983	1.478	1.870	2.611	1.566	1.971	2.741
200	1.798	2.143	2.816	1.865	2.222	2.921	1.450	1.837	2.570	1.524	1.923	2.679
250	1.780	2.121	2.788	1.839	2.191	2.880	1.431	1.815	2.542	1.496	1.891	2.638
300	1.767	2.106	2.767	1.820	2.169	2.850	1.417	1.800	2.522	1.476	1.868	2.608
∞	1.645	1.960	2.576	1.645	1.960	2.576	1.282	1.645	2.326	1.282	1.645	2.326