# PROBABILIDAD Y ESTADÍSTICA

Inferencia estadística – Ejemplos de estimación de la media poblacional por intervalos de confianza

Ing. Tripp, Nicolás
Ing. Miranda Virginia

2018

# Métodos de estimación clásicos

Los métodos de estimación se pueden clasificar en "clásicos" (cuando se generaliza exclusivamente a partir de los estadísticos de una muestra aleatoria tomada de la población) y "bayesianos" (que además de la información muestral incluyen conocimiento previo sobre la distribución de probabilidad de los parámetros desconocidos). Dentro de los clásicos encontramos la "estimación puntual" y la "estimación por intervalo".

Se denomina "estimador de un parámetro" a cualquier variable aleatoria que se exprese en función de una muestra aleatoria y que tenga por objetivo aproximar el valor del parámetro.
Por ejemplo, la media $\bar{X}$ y mediana $\tilde{X}$ muestrales son estimadores de la media poblacional $\mu$
Cuando se seleccionó una muestra, el valor que toma el estimador para dicha muestra se denomina "estimación puntual".

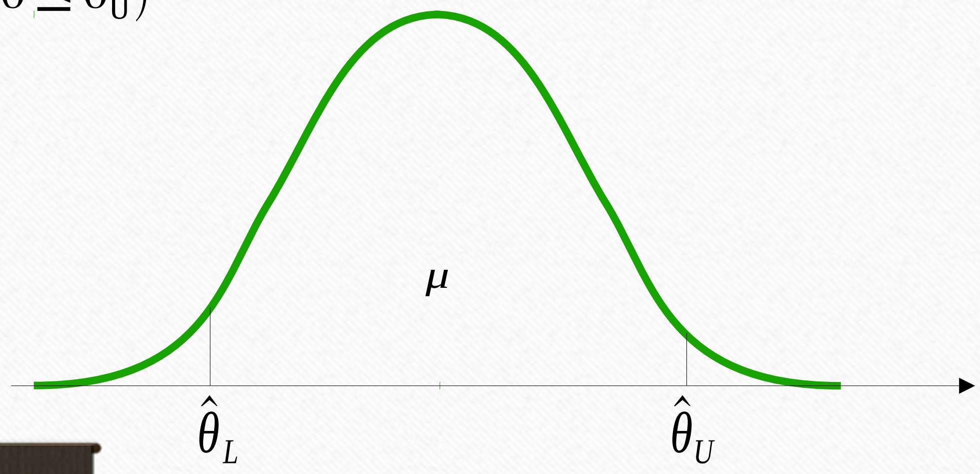
Muestra de tamaño n $(X_1, X_2, \ldots, X_n)$ $\longrightarrow$ Estimación puntual del parámetro $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

# Estimación por intervalo

Aunque se cuente con la mejor estimación puntual posible, es raro que determine exactamente el parámetro poblacional. Es preferible determinar un intervalo dentro del cual esperamos encontrar (con cierta probabilidad o confianza) el valor del parámetro.

Una "estimación por intervalo de un parámetro $\theta$" es un intervalo de la forma $\hat{\theta}_L \leq \theta \leq \hat{\theta}_U$ donde los extremos son estimaciones puntuales obtenidas de una muestra con cierta distribución muestral.

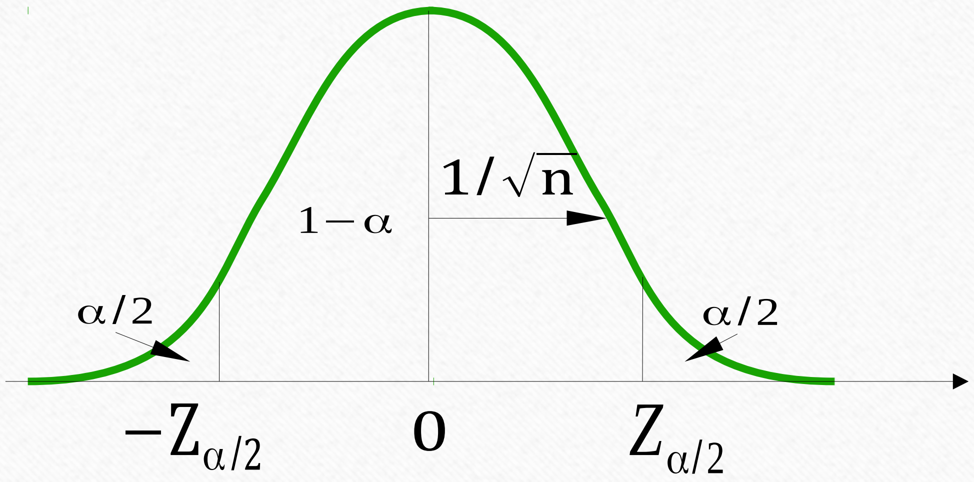
Como las estimaciones puntuales poseen una distribución muestral, se puede calcular la probabilidad de que el parámetro se encuentre dentro del intervalo $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U)$



# Estimación por intervalo de la media poblacional

Si se propone un valor de error aceptable α se puede determinar un intervalo de confianza. Los extremos del intervalo se denominan límites de confianza.

Si la muestra se obtiene de una población normal o con un tamaño ≥ 30, podemos considerar que la distribución muestral de la media tipificada (Z) es la normal estándar.




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

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

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

$$\boxed{\bar{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

# Estimación por intervalo de la media poblacional

Ejemplo:

La concentración promedio de Zinc de una muestra de 36 mediciones tomadas en distintos lugares de un río dio 2,6 g/ml. Encuentre el intervalo de confianza del 95% para la concentración promedio de Zinc en el río. Considere un desvío estándar de la población de 0,3 g/ml.

Solución:

La estimación puntual es 2,6 g/ml.

Para un nivel de confianza del 95% tenemos que $1-\alpha=0{,}95 \rightarrow \alpha/2=0{,}025 \rightarrow Z_{0{,}025}=1{,}96$

el intervalo será

$$\boxed{\begin{gathered} 2{,}6-1{,}96\,\frac{0{,}3}{\sqrt{36}}<\mu<2{,}6+1{,}96\,\frac{0{,}3}{\sqrt{36}} \\ 2{,}5<\mu<2{,}7 \end{gathered}}$$

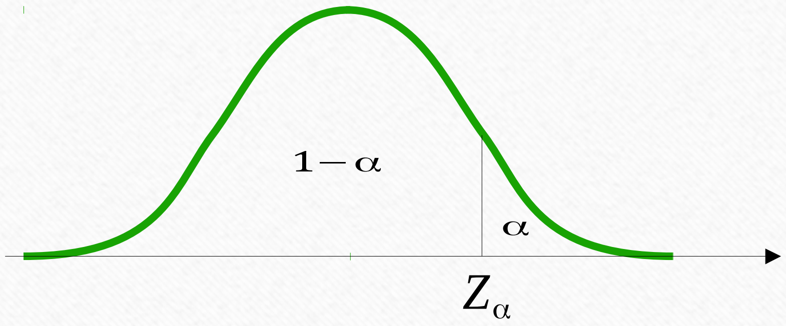
# Estimación por intervalo unilateral

En algunos estudios nos interesa conocer los valores extremos del parámetro poblacional. Por ejemplo cuando se busca determinar la resistencia mínima de un material, la concentración máxima de un tóxico en el agua, la velocidad máxima del viento, la intensidad mínima de la señal de WiFi, etc.

En estas situaciones, se utiliza la "<u>estimación por intervalos unilaterales</u>"

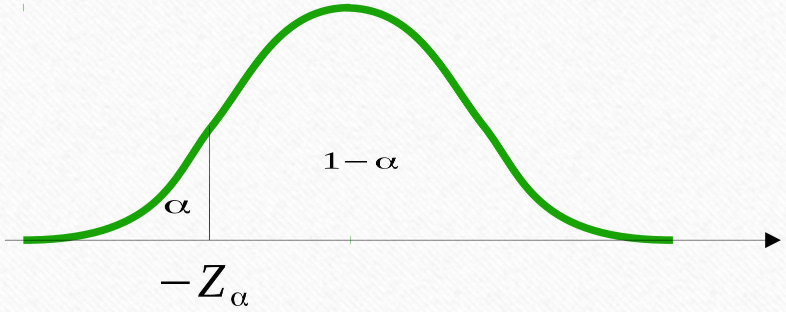
$$P(Z<Z_\alpha)=1-\alpha$$

Valor mínimo $\boxed{\bar{X}-Z_\alpha\frac{\sigma}{\sqrt{n}}<\mu}$




$$P(-Z_\alpha<Z)=1-\alpha$$

Valor máximo $\boxed{\mu<\bar{X}+Z_\alpha\frac{\sigma}{\sqrt{n}}}$



## Estimación por intervalo unilateral

Ejemplo:

En un experimento psicológico se eligen al azar 25 personas para medir su tiempo de reacción en segundos, frente a un determinado estímulo. Se conoce que la varianza de la población es 4 $seg^2$ y la distribución de los tiempos de reacción es aproximadamente normal. Para la muestra elegida el tiempo medio fue 6,2 seg. ¿Cuál es la cota superior para el tiempo medio de reacción con una confianza del 95%?

Solución:

La estimación puntual es 6,2 seg

Para un nivel de confianza del 95% tenemos que $Z_{0,05} = 1{,}654$

el máximo será

$$\boxed{\mu < \bar{X} + Z_{\alpha}\frac{\sigma}{\sqrt{n}}}$$

$$\boxed{\begin{array}{c} \mu < 6{,}2 + 1{,}645\dfrac{\sqrt{4}}{\sqrt{25}} \\ \mu < 6{,}858\,\text{seg} \end{array}}$$

# Estimación de media con varianza desconocida

Si tenemos una muestra aleatoria obtenida a partir de una población normal cuya varianza no se conoce, el estimador apropiado es el T que tiene una distribución t de Student con n-1 grados de libertad.

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

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$$\boxed{\bar{X} - T_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}}$$

Intervalo bilateral

$$P\left(T < T_{\alpha}\right) = 1 - \alpha$$

Valor mínimo $\boxed{\bar{X} - T_{\alpha}\frac{S}{\sqrt{n}} < \mu}$

$$P\left(-T_{\alpha} < Z\right) = 1 - \alpha$$

Valor máximo $\boxed{\mu < \bar{X} + T_{\alpha}\frac{S}{\sqrt{n}}}$

Intervalo unilateral

# Estimación de media con varianza desconocida

Ejemplo:

Se observó el contenido de ácido sulfúrico de 7 recipientes similares. Los valores registrados fueron 9,8; 10,2; 10,4; 9,8; 10,0; 10,2 y 9,6 litros. Determine un intervalo de confianza del 95% para el contenido medio de todos los contenedores, asumiendo que pertenecen a una población normal.

Solución:

La estimación puntual es 10,0 litros.

Para una muestra de 6 grados de libertad, con desvío estándar de 0,283 y un nivel de confianza del 95% tenemos que $T_{0,025} = 2,447$

el intervalo será

$$\boxed{\begin{gathered} 10,0 - 2,447\frac{0,283}{\sqrt{7}} < \mu < 10,0 + 2,447\frac{0,283}{\sqrt{7}} \\ 9,74 < \mu < 10,26 \end{gathered}}$$

# Intervalos de predicción y detección de datos aberrantes

En ciertos controles de calidad resulta de interés poder predecir el valor que se obtendrá al tomar una muestra adicional. Suponemos que se toma una muestra aleatoria a partir de una población normal con varianza conocida $\sigma^2$. Si tomamos una nueva observación $x_0$ de la misma población, su estadístico normalizado será

$$Z_0 = \frac{x_0 - \bar{X}}{\sqrt{\sigma^2 + \sigma^2/n}} = \frac{x_0 - \bar{X}}{\sigma\sqrt{1+1/n}}$$

Como la población es normal podemos calcular el intervalo de confianza

$$P\left(-Z_{\alpha/2} < Z_0 < Z_{\alpha/2}\right) = 1 - \alpha$$

$$\boxed{\bar{X} - Z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{X} + Z_{\alpha/2}\sigma\sqrt{1+1/n}}$$

Una observación se considera aberrante si cae fuera del intervalo de predicción calculado a partir de la muestra sin incluir la observación en cuestión.

# Intervalo de predicción

Ejemplo:

Debido a las corridas cambiarias, un banco recibió una gran cantidad de pedidos de extracción de dólares por ventanilla. Una muestra reciente de 50 peticiones de extracción arrojó un valor medio del monto solicitado de \$257.300. Asumiendo un desvío estándar poblacional de \$25.000, determine un intervalo de predicción de 95% para la próxima petición de extracción.

Solución:

La estimación puntual es \$257.300.

Para una muestra de 50, con desvío estándar poblacional de 25.000 y un nivel de confianza del 95% tenemos que $Z_{0,025} = 1,96$

el intervalo será

$$257.300 - 1,96\,\frac{25.000}{\sqrt{1+1/50}} < x_0 < 257.300 + 1,96\,\frac{25.000}{\sqrt{1+1/50}}$$

$$\$\,207.812,43 < x_0 < \$\,306.787,57$$