



MA34B – Estadística

Estimación por Intervalo

Prof. Rodrigo Abt B.

rabt@dim.uchile.cl

Introducción (1)

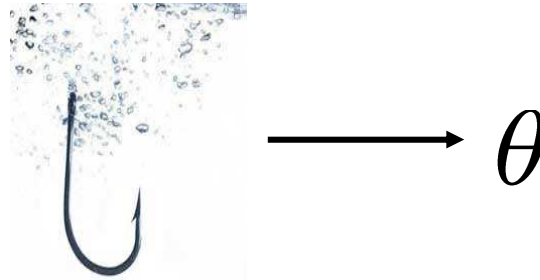
- En un problema de estimación puntual, el objetivo es determinar una fórmula o receta que permita dar un valor para el(los) parámetro(s) desconocidos de una distribución dada.
- Pero, una vez hecha una estimación, no sabemos si en realidad estamos cerca o no del parámetro estimado.
- Sólo tenemos una idea de la precisión de un estimador a través de su varianza.

Introducción (2)

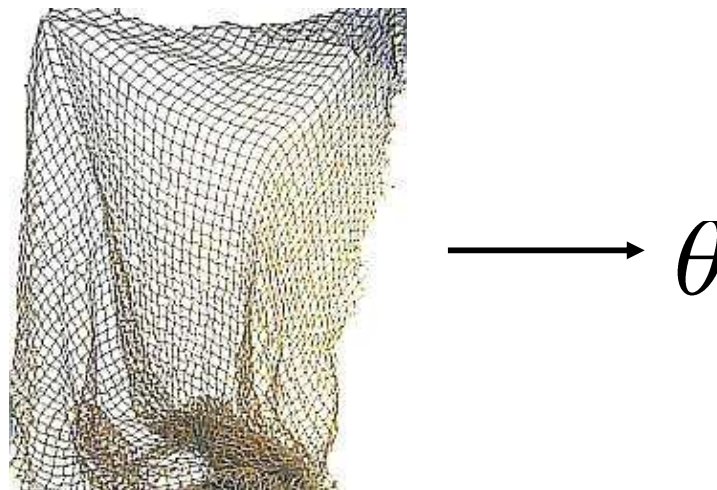
- En la práctica, la estimación puntual es rara, ya que si bien un estimador es bueno, la muestra obtenida para la estimación puede hacernos dudar de qué tan lejos o cerca está la estimación del verdadero valor de un parámetro.
- Por lo mismo, es más frecuente buscar un *intervalo* de valores entre los cuales pudiese estar el parámetro buscado con algún grado de confiabilidad.
- Por ejemplo, encontrar un intervalo de valores para la media μ de una Normal con un alto nivel de confiabilidad

Introducción (3)

- Para hacerse una idea, hacer una estimación puntual es como pescar con un anzuelo:



- Mientras que hacer una estimación por intervalo es como pescar con una red.



Introducción(4)

- Ahora bien, ¿de qué tamaño debería ser esta red para lograr “atrapar” un buen valor?, y ¿cómo nos aseguramos de que estamos “atrapando” lo correcto?
- Supongamos que estimamos la media μ de una Normal. Sabemos que μ puede tomar cualquier valor dentro de los reales, por ende, el único intervalo que nos garantiza un 100% de confiabilidad para encontrar al parámetro en él es...

$$[-\infty, \infty]$$

- Sin embargo, dicho intervalo es de muy poca utilidad. De hecho no nos dice nada informativo.
- Sin embargo, si tuviésemos un intervalo como $[2,3]$ sería BASTANTE más informativo.
- Desgraciadamente, para tener un intervalo debemos tolerar un margen de error, y “descontarlo” de la confiabilidad esperada. O sea, ¿podríamos encontrar valores “a” y “b” tales que la confiabilidad de encontrar μ en $[a,b]$ sea de un 95% por ejemplo?

Planteamiento

- Formalmente, buscamos “a” y “b” tales que:

$$P(a \leq \mu \leq b) = 1 - \alpha$$

- En este caso el intervalo $[a,b]$ se denomina “intervalo de confianza”, donde $1-\alpha$ es el “nivel de confianza” o “confianza”.
- Ahora bien, dado que no conocemos la distribución asociada a μ , debemos utilizar los valores muestrales para determinar los valores de “a” y “b”, es decir, buscamos $a=a(X_1, X_2, \dots, X_n)$ y $b=b(X_1, X_2, \dots, X_n)$.
- Por ende “a” y “b” son variables aleatorias.

Resolución

- Para resolver la ecuación anterior, debemos encontrar una variable aleatoria que dependa de μ , pero cuya distribución NO dependa de μ . Es decir, debemos transformar la ecuación original.

$$P(a \leq \mu \leq b) = 1 - \alpha$$



$$P(L \leq Z \leq U) = 1 - \alpha$$

- En que Z es una v.a. con distribución conocida.

Caso 1: Media μ para Normal con varianza conocida

- Retomemos el caso de la media
 - Buscamos “a” y “b” tales que $P(a \leq \mu \leq b) = 1 - \alpha$
 - Sabemos por otro lado que:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

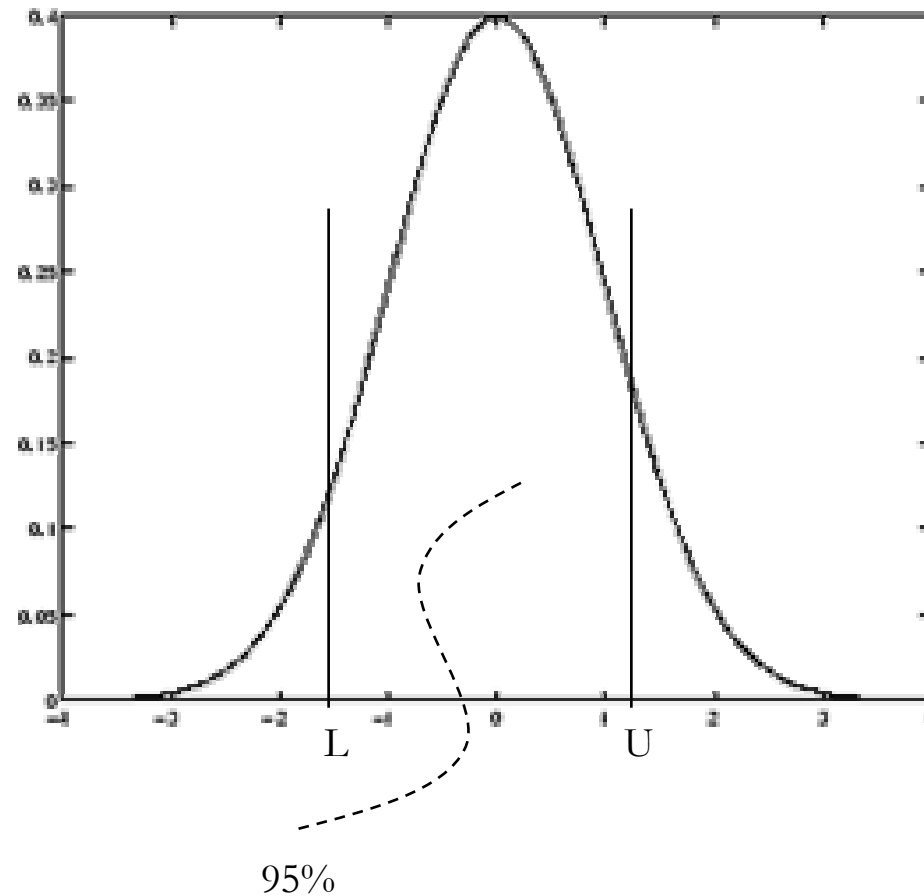
- Si hacemos las transformaciones algebraicas correspondientes, tenemos:

$$P\left(\frac{\bar{X} - b}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - a}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

The diagram shows three labels at the bottom: L, Z, and U. Arrows point from each label to a specific part of the probability expression above. An arrow from 'L' points to the first fraction $\frac{\bar{X} - b}{\sigma/\sqrt{n}}$. An arrow from 'Z' points to the middle fraction $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$. An arrow from 'U' points to the third fraction $\frac{\bar{X} - a}{\sigma/\sqrt{n}}$.

Análisis Gráfico

- En términos gráficos, si fijamos un nivel de confianza del 95% buscamos L y U tales que:



Condiciones de Borde (1)

- Si nos detenemos un poco en el análisis gráfico, veremos que existen infinitos valores de L y U que cumplen la condición. En efecto, fijando uno podemos obtener el otro.
- ¿Cómo elegimos entre uno u otro intervalo?
- Criterio: Largo del intervalo
- Claramente elegiremos aquellos intervalos que presenten un largo menor, ya que significa que son más precisos.

Condiciones de Borde (2)

- Idealmente preferiremos aquellos de largo mínimo.
- En el caso de las distribuciones simétricas, el problema está resuelto: Basta tomar como centro del intervalo el punto de simetría y tendremos un intervalo de largo mínimo.
- Luego, debemos encontrar L tal que $P(-L \leq Z \leq L) = 1 - \alpha$, que es equivalente a buscar $P(Z > L) = 0,025 \Rightarrow L = 1,96$

Solución

- Luego, despejando tenemos que $L = 1,96$, pero:

$$L = \frac{\bar{X} - a}{\sigma/\sqrt{n}} \Rightarrow a = \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}$$

- Y por otro lado:

$$b = \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}$$

- Finalmente el intervalo buscado es:

$$[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}]$$

Observaciones del Resultado

- Notemos que el intervalo está centrado en la media muestral.
- Además crece con la varianza, y disminuye con el tamaño muestral, lo que es intuitivamente deseable.
- Si queremos aumentar la confianza, debemos estar dispuestos a que crezca el tamaño del intervalo.

Notas

- Si no conocemos la varianza, debemos utilizar la distribución t-Student y el estadístico asociado.
- Los intervalos de confianza para el caso Bayesiano se pueden derivar directamente de la distribución a posteriori.
- La interpretación correcta es que el intervalo de confianza “cubre” al parámetro el $(1-\alpha)100\%$ de las veces.
- Las distribuciones muestrales como la Normal, t-Student o Chi cuadrado se utilizan como distribuciones límite cuando el tamaño muestral es grande.
- Lo más complicado suele ser encontrar el estadístico que dependa del parámetro a estimar, pero cuya distribución NO dependa del mismo.
- A este tipo de estadísticos se les denomina “pivotes”.