Tema 11: Estimación por Intervalo

Profesora: Carmen Elvira Ramos Domínguez

Índice

- Estimador por Intervalo. Estimación por Intervalo.
- Nivel de Confianza.
- Tipos de Intervalos. Bilaterales y Unilaterales.
- Método del Pivote.
- Intervalos de Confianza más Usuales.
- Determinación del Tamaño Muestral.

Estimador por Intervalo

En la Estimación por Intervalo se emplean dos estadísticos de la muestra. A dicha pareja se la denomina Estimador por Intervalo o Intervalo de Confianza Aleatorio $[L_1,L_2]$. En realidad, es una variable aleatoria bidimensional, cuyas componentes se denominan Límites de Confianza Inferior y Superior de Intervalo Aleatorio.

Definición: Un Intervalo de Confianza Aleatorio de un parámetro θ es un intervalo que se obtiene a partir de los datos aportados por la muestra, al que pertenece el verdadero valor del parámetro con una alta probabilidad.

Definición: Al valor que toma el estimador por intervalo para una realización particular de la muestra se llama Estimación por Intervalo o Intervalo de Confianza.

Nivel de Confianza

Definición: Se denomina Nivel de Confianza a la probabilidad de que el intervalo de confianza aleatorio o estimador por intervalo cubra al verdadero valor del parámetro. Esto es, $P\{L_1 \leq \theta \leq L_2\} = 1-\alpha$.

Dicha probabilidad suele ser alta. Valores típicos:

1. $1-\alpha = 0.99$;   2. $1-\alpha = 0.95$;   3. $1-\alpha = 0.90$;

A α se le llama Nivel de Significación y suele valer:

1. $\alpha = 0.01$;   2. $\alpha = 0.05$;   3. $\alpha = 0.10$;

- Una vez evaluado el estimador por intervalo para una realización particular de la muestra, no tiene sentido de hablar de probabilidad, ya que la estimación por intervalo es un intervalo numérico.

Tipos de Intervalos

Existen dos tipos de Intervalos de Confianza Aleatorios:

- ❖ Sean L_1 y L_2 los límites inferior y superior del Intervalo de Confianza Aleatorio, de manera que $P(L_1 \leq \theta \leq L_2) = 1-\alpha$. Entonces a dicho intervalo $[L_1, L_2]$ se le dice Intervalo de Confianza Bilateral.
- ❖ Sea L_1 el límite inferior del Intervalo de Confianza Aleatorio, de forma que $P(L_1 \leq \theta) = 1- \alpha$. Entonces a dicho intervalo $[L_1, \infty)$ se le dice Intervalo de Confianza Unilateral Superior.
- ❖ Sea L_2 el límite superior del Intervalo de Confianza Aleatorio, de forma que $P(\theta \leq L_2) = 1- \alpha$. Entonces a dicho intervalo $(-\infty, L_2,)$ se le dice Intervalo de Confianza Unilateral Inferior.

Nos centramos en los Intervalos de Confianza bilaterales $[L_1, L_2]$, centrados en un estimador insesgado del parámetro y equilibrados, en el sentido de que la probabilidad de no cubrir el parámetro sea $\alpha/2$, a ambos lados.

Método del Pivote

Los pasos a seguir para construir un intervalo de Confianza son:

1. Determinar la distribución de la variable de interés X en la población. Dicha distribución F_θ dependerá del parámetro θ, para el que buscamos el intervalo. Establecer el nivel de confianza.
2. Buscar el estimador $\hat{\theta}$ del parámetro θ para el que buscamos el intervalo. Se suele elegir insesgado y determinar su distribución.
3. A partir de dicho estimador $\hat{\theta}$, buscar una transformación $g(\theta, \hat{\theta})$ que sea una variable aleatoria, que depende de θ pero su distribución no. A dicha transformación se la llama Pivote.
4. Buscar la región tal que $P(g(\theta, \hat{\theta}) \in A) = 1- \alpha$, ya que se conoce como se distribuye el pivote, y despejar θ, $P(\theta \in h(A)) = 1- \alpha$, entonces $h(A)$ es el intervalo de confianza buscado.

Intervalos de Confianza más Usuales

Intervalo de Confianza para la media de una Población Normal.

1. **Con varianza conocida.** Sea X_1, X_2, … X_n una m.a.s. de una población N(μ, σ) con σ^2 conocida. Buscamos un intervalo de confianza para μ y se elige como estimador insesgado la media muestral $\overline{\mathrm{X}} = \frac{1}{\mathrm{n}}(\mathrm{X}_1 + \ldots + \mathrm{X}_\mathrm{n})$.

   Se sabe que $\overline{\mathrm{X}} \sim \mathrm{N}(\mu, \frac{\sigma}{\sqrt{\mathrm{n}}})$ entonces tipificando: $\frac{\overline{\mathrm{X}} - \mu}{\sigma/\sqrt{\mathrm{n}}} \sim \mathrm{N}(0,1)$ cuya distribución no depende de μ. Por consiguiente:

$$P\left(-z_{\alpha/2} \leq \frac{\overline{\mathrm{X}} - \mu}{\sigma/\sqrt{\mathrm{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha \rightarrow$$

$$P\left(-z_{\alpha/2}\,\sigma/\sqrt{\mathrm{n}} \leq \overline{\mathrm{X}} - \mu \leq z_{\alpha/2}\,\sigma/\sqrt{\mathrm{n}}\right) = 1 - \alpha \rightarrow$$

$$P\left(\overline{\mathrm{X}} - z_{\alpha/2}\,\sigma/\sqrt{\mathrm{n}} \leq \mu \leq \overline{\mathrm{X}} + z_{\alpha/2}\,\sigma/\sqrt{\mathrm{n}}\right) = 1 - \alpha$$

   Por tanto, el intervalo de confianza es: $\left[\overline{\mathrm{X}} \pm z_{\alpha/2}\,\sigma/\sqrt{\mathrm{n}}\right]$

Intervalos de Confianza más Usuales

2. **Con varianza desconocida.** Sea $X_1, X_2, \dots X_n$ una m.a.s. de una población $N(\mu, \sigma)$ con σ^2 desconocida. Buscamos un intervalo de confianza para μ y se elige como estimador insesgado la media muestral $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$.

   Se sabe que $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ entonces tipificando: $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$,

   cuya distribución no depende de μ, pero él si depende de σ.

   Se distinguen dos casos:

   1. Si la muestra es grande, $n > 30$, entonces: $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$,

      entonces: $P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha \rightarrow$

      $$P\left(\overline{X} - z_{\alpha/2}\, S/\sqrt{n} \leq \mu \leq \overline{X} + z_{\alpha/2}\, S/\sqrt{n}\right) = 1 - \alpha$$

      Por tanto, el intervalo de confianza es: $\left[\overline{X} \pm z_{\alpha/2}\, S/\sqrt{n}\right]$

Intervalos de Confianza más Usuales

2. Si la muestra es pequeña, n < 30, entonces: $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}{}^2$,

entonces: $\frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2\,(n-1)}}} \sim t_{n-1} \rightarrow \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

Entonces podemos considerar:

$$P\left(-t_{\alpha/2,(n-1)} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,(n-1)}\right) = 1-\alpha \;\rightarrow$$

$$P\left(\bar{X} - t_{\alpha/2,(n-1)}\, S/\sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2,(n-1)}\, S/\sqrt{n}\right) = 1-\alpha$$

Por tanto, el intervalo de confianza es: $\left[\bar{X} \pm t_{\alpha/2,(n-1)}\, S/\sqrt{n}\right]$

Determinación del Tamaño Muestral

Hasta ahora suponíamos conocido el tamaño muestral. Sin embargo, la determinación del tamaño muestral no es fácil, ya que si es muy grande puede elevar el coste del estudio y si es pequeño puede que las conclusiones no sean adecuadas.

¿Cuál debería ser el tamaño de la muestra que nos permita hacer estimaciones con una determinada precisión y al mínimo coste?

Sea ε el error de estimación que estamos dispuestos a tolerar y $1-\alpha$ el nivel de confianza, buscamos el mínimo n tal que

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

Notar que los intervalos de confianza tienen una longitud que es inversamente proporcional al tamaño muestral (cuanto más grande es el tamaño muestral menor longitud y mayor precisión), por supuesto fijado el nivel de confianza.

Determinación del Tamaño Muestral

Ejemplo: Sea X = N(μ, σ) con σ^2 conocida, y queremos determinar el tamaño n para estimar μ de forma que el error cometido en las estimaciones sea ε, y el grado de confianza 1-α. Esto es,

$$\mu \in [\bar{X} - \varepsilon, \bar{X} + \varepsilon]$$

con probabilidad 1-α.

Por otra parte,

$$\mu \in \left[\bar{X} - z\alpha/_2\, {}^{\sigma}/_{\sqrt{n}}, \bar{X} + z\alpha/_2\, {}^{\sigma}/_{\sqrt{n}}\right]$$

con probabilidad 1-α.

Entonces:

$$\varepsilon = z\alpha/_2\, {}^{\sigma}/_{\sqrt{n}}$$

Y despejando n nos queda:

$$n \geq \left(z\alpha/_2\, {}^{\sigma}/_{\varepsilon}\right)^2$$