Capítulo 6

Intervalos de confianza

6. Intervalos de Confianza

Sea $X \sim f_X(x;\theta)$, donde θ es parámetro por estimar y sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria. Por alguna razón, ya no estamos interesados en encontrar un valor admisible para el parámetros sino que más bien, queremos un rango o intervalo de valores admisibles para θ. Por ejemplo, deseamos un rango de valores alternativos para la inflación θ del próximo año en vez de la inflación promedio del mismo.

6.1 Definición
Sean L_1 y L_1 dos estadísticas y sea $(1-\alpha)$ una probabilidad predeterminada. Diremos que $[L_1, L_2]$ **forma un Intervalo de Confianza (I.C.) para** θ**, al nivel de** $100(1-\alpha)\%$ **de Confianza**, **si para todo valor posible de** θ**, se cumple que** $P(L_1 \le \theta \le L_2) = 1-\alpha$

Observación:
También se suele presentar el I.C. escribiendo: $L_1 \le \theta \le L_2$ **con** $100(1-\alpha)\%$ **de Confianza.**
La probabilidad predeterminada $(1-\alpha)$ **se llama "Nivel de confianza" y el estándar es** $(1-\alpha) = 0.95$ **o sea 95% de confianza**

La construcción de un I.C. busca encontrar un intervalo que tenga alta probabilidad fija $(1-\alpha)$ ***y que a la vez sea de longitud mínima.***

Usualmente la construcción del IC. Implica resolver el problema de optimización:
$Mín\ (L_2 - L_1) \quad sa\ \ P(L_1 \le \theta \le L_2) = 1-\alpha = Valor\ \ dado$

Un método que genera intervalos óptimos o casi óptimos es el *"Método de la variable base o variable pívote"*

6.2 Metodología (Método de la Variable Base)
Sea $X \sim f_X(x;\theta)$ y sea $(X_1, X_2, ..., X_n)$ m.a. tomada de la población de X. Dado un nivel de confianza $(1-\alpha)$, para construir un Intervalo ***de*** Confianza de nivel $100(1-\alpha)\%$ para θ:

a) Determinar una variable base $W = W(X_1, X_2, ..., X_n;\theta)$ que contenga al parámetro θ pero cuya distribución $g_W(w)$ no dependa de θ.

b) Buscar en la distribución de W dos valores a y b tales que $P(W \le a) = \alpha/2$ y $P(W \le a) = \alpha/2$ de modo que $P(a \le W \le b) = P(a \le W(X_1, X_2, ..., X_n;\theta) \le b) = 1-\alpha$.

c) Despejar θ del interior del intervalo, de modo que
$P[L_1(X_1, X_2, ..., X_n;a,b) \le \theta \le L_2(X_1, X_2, ..., X_n;a,b)] = 1-\alpha$

El Intervalo de Confianza para θ al nivel $100(1-\alpha)\%$ es $[L_1(X_1, X_2, ..., X_n;a,b)\ ,\ L_2(X_1, X_2, ..., X_n;a,b)]$
y se escribe
$L_1(X_1, X_2, ..., X_n;a,b) \le \theta \le L_2(X_1, X_2, ..., X_n;a,b)$ al $100(1-\alpha)\%$ *de confianza*

Nota:
Este sistema por lo general proporciona intervalos de confianza (I.C.) de longitud promedio mínima.
Usualmente la variable base W se forma partiendo del estimador Máximo Verosímil $\hat{\theta}_{MV}$, aprovechando que si n es grande, el estimador tiene distribución normal $\hat{\theta} \sim N(\theta,\sigma^2_{\hat{\theta}}) \Leftrightarrow W = \dfrac{\hat{\theta}-\theta}{\sqrt{\sigma^2_{\hat{\theta}}}} \sim N(0,1)$

6.3 Ejemplos

Ejemplo 1
Sea $X \sim Exp(\beta = 1/\theta)$, asuma una m.a. de tamaño n=49 (grande) y construya un I.C. para θ de nivel 100(1-α)%=95% de confianza.

Solución:
El tamaño n=49>30 (grande) implica que $\overline{X} \sim N(\mu=\theta, \dfrac{\sigma^2}{n} = \dfrac{\theta^2}{49})$, pues en una distribución exponencial $\mu = \dfrac{1}{\beta} = \theta$ y $\sigma^2 = \dfrac{1}{\beta^2} = \theta^2$. Luego:

a) $Z = \dfrac{\overline{X}-\theta}{\sqrt{\dfrac{\theta^2}{n}}} = \dfrac{\overline{X}-\theta}{\dfrac{\theta}{\sqrt{n}}} = \dfrac{\overline{X}-\theta}{\dfrac{\theta}{7}} \sim N(0,1)$ es una **variable base**

b) Como el nivel de confianza es 100(1- α)%=95%, entonces α/2=0.025 y si planteamos valores **a** y **b** tales que $P(a \le Z \le b) = 0.95$ y $P(Z \le a) = 0.025$, $P(b \le Z) = 0.025$, dada la simetría de la distribución N(0,1) se tiene que **b=1.96 y a=-b=-1.96 .**
Por tanto tenemos $P(-1.96 \le \dfrac{\overline{X}-\theta}{\dfrac{\theta}{7}} \le 1.96) = 0.95$

c) "Despejando" θ en el interior del paréntesis:

$$0.95 = P(-1.96 \le \frac{\overline{X}-\theta}{\frac{\theta}{7}} \le 1.96) = P(-1.96 \le \frac{\frac{\overline{X}}{\theta}-1}{\frac{1}{7}} \le 1.96) = P(\frac{-1.96}{7} \le \frac{\overline{X}}{\theta} - 1 \le \frac{1.96}{7})$$

$$= P(1-1.96/7 \le \frac{\overline{X}}{\theta} \le 1+1.96/7) = P(\frac{1-1.96/7}{\overline{X}} \le \frac{1}{\theta} \le \frac{1+1.96/7}{\overline{X}}) = P(\frac{\overline{X}}{1+1.96/7} \le \theta \le \frac{\overline{X}}{1-1.96/7}) = 0.95$$

Y todo lo anterior ocurre sea cual fuere el valor de θ, por tanto si definimos:
$L_1 := \dfrac{\overline{X}}{1+1.96/7}$ y $L_2 := \dfrac{\overline{X}}{1-1.96/7}$. **Entonces podemos decir que** $\left[L_1 := \dfrac{\overline{X}}{1+1.96/7}, L_2 := \dfrac{\overline{X}}{1-1.96/7}\right]$, **es un Intervalo de Confianza para** θ **al 95% de confianza.**

Ejemplo 2 (Intervalo de Confianza de 95% para la Media μ de una $N(\mu,\sigma^2)$ en el caso de σ^2 conocida)
Tenemos una distribución normal $N(\mu,\sigma^2)$ donde $\sigma^2=\sigma_0^2$ es conocida y una muestra de tamaño n, ($X_1, X_2, ...X_n$). Dado un nivel de confianza de 95%, se desea construir un I.C. para la media μ.

a) Sea $(X_1,X_2,...X_n)$ una m.a. tomada de una distribución $N(\mu,\sigma^2)$ donde $\sigma^2=\sigma_0^2$ es conocida. Sabemos que $\overline{X} \sim N(\mu, \dfrac{\sigma_0^2}{n})$ y por tanto $Z = \dfrac{\overline{X}-\mu}{\dfrac{\sigma_0}{\sqrt{n}}}$ ~ N(0,1) sirve como **variable base**

b) Como el nivel de confianza es 100(1- α)%=95%, entonces α/2=0.025 y si planteamos valores **a** y **b** tales que $P(a \le Z \le b) = 0.95$ y $P(Z \le a) = 0.025$, $P(b \le Z) = 0.025$, dada la simetría de la distribución N(0,1) se tiene que **b=1.96 y a=-b=-1.96;**

Por tanto tenemos $P(-1.96 \le \dfrac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}} \le 1.96) = 0.95$

c) "Despejando" μ en el interior del paréntesis:

$$0.95 = P(-1.96 \le \frac{\overline{X}-\mu}{\frac{\sigma_0}{\sqrt{n}}} \le 1.96) = P(-1.96\frac{\sigma_0}{\sqrt{n}} \le \overline{X}-\mu \le 1.96\frac{\sigma_0}{\sqrt{n}}) = P(-\overline{X}-1.96\frac{\sigma_0}{\sqrt{n}} \le -\mu \le -\overline{X}+1.96\frac{\sigma_0}{\sqrt{n}})$$

$$= P(\overline{X}+1.96\frac{\sigma_0}{\sqrt{n}} \ge \mu \ge \overline{X}-1.96\frac{\sigma_0}{\sqrt{n}}) = P(\overline{X}-1.96\frac{\sigma_0}{\sqrt{n}} \le \mu \le \overline{X}+1.96\frac{\sigma_0}{\sqrt{n}}) = 0.95;$$

Y todo lo anterior ocurre sea cual fuere el valor de μ. Por tanto si definimos:

$L_1 := \overline{X} - 1.96\dfrac{\sigma_0}{\sqrt{n}}$ y $L_2 := \overline{X} + 1.96\dfrac{\sigma_0}{\sqrt{n}}$. **Entonces podemos decir que**

$\left[L_1 := \overline{X} - 1.96\dfrac{\sigma_0}{\sqrt{n}}, L_2 := \overline{X} + 1.96\dfrac{\sigma_0}{\sqrt{n}}\right]$, **es un Intervalo de Confianza para μ al 95% de confianza.**

Nota: Intervalo para un nivel general de confianza 100(1- α)%

En este caso, el procedimiento es análogo al anterior, sólo se cambia el valor 1.96 por el valor tabular $Z_{1-\alpha/2}$ que es el percentil (1-α/2) de la distribución N(0,1). El I.C. Resultante es: $\left[L_1 := \overline{X} - Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}, L_2 := \overline{X} + Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}\right]$

Observaciones:

También se escribe el I.C. mediante:

$\mu = \overline{X} \pm Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}$ con un 100(1-α)% de Confianza o equivalentemente

$\overline{X} - Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}} \le \mu \le \overline{X} + Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}$ al 100(1-α)% de Confianza.

Ejemplo 3

Si el gasto *X* (en dólares) en educación, por parte de las familias de estrato medio, tiene distribución normal de media μ y desviación estándar σ=16 y se trata de estimar μ mediante un I.C. de Nivel 95%. Si una muestra de n = 60 familias dio una media $\overline{X}$ de US$110 halle un intervalo de confianza para μ al nivel de 95%. ¿Sería aceptable pensar que la media μ es superior a los 100 dólares?

Solución:

La desviación estándar poblacional σ es conocida e igual a $\sigma_0 = 16$ y podemos usar $\mu = \overline{X} \pm Z_{1-\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}$. En este problema $100(1-\alpha)\% = 95\%$ o sea $\alpha = 0.05$ y $1-\alpha/2 = 0.975$ de modo que $Z_{1-\alpha/2} = Z_{0.975} = 1.96$

Además $n = 60$ y reemplazando en la formulación del I.C. tenemos $\mu = 110 \pm 1.96 \times \dfrac{16}{\sqrt{60}} = 110 \pm 4.05$ con 95% de confianza. O equivalentemente $105.95 \le \mu \le 114.05$ con 95% de confianza.

Observando el I.C. hallado, se ve que podemos confiar en que la media μ será, en efecto, mayor a los US$ 100

Ejemplo 4 (Intervalo de Confianza para la Media μ de una $N(\mu,\sigma^2)$ en el caso de σ^2 desconocida)
Tenemos una distribución normal $N(\mu,\sigma^2)$ donde σ^2 es **desconocida** y una muestra de tamaño n, ($X_1, X_2, \ldots X_n$). Dado un nivel de confianza de 95%, se desea construir un I.C. para la media μ .

La deducción del I.C. se hace a partir del hecho de reconocer que como σ es desconocida, debemos

a) Cambiar de variable base para la construcción del I.C., reemplazando σ por su estimación S . En este caso la variable base resultante es $t=\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ que tiene distribución t de Student con (n-1) grados de libertad.

b) Usamos la Tabla t y la simetría de la distribución t-Student, para encerrar a la variable $t=\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ entre los percentiles $-t_{1-\alpha/2}$ y $t_{1-\alpha/2}$, de modo que $P\left[-t_{1-\alpha/2}\le\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\le t_{1-\alpha/2}\right]=1-\alpha$

c) "Despejado" μ en el interior de la probabilidad:

$$1-\alpha=P\left[-t_{1-\alpha/2}\le\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\le t_{1-\alpha/2}\right]=P\left[-t_{1-\alpha/2}\frac{S}{\sqrt{n}}\le\overline{X}-\mu\le t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]=$$

$$P\left[\overline{X}-t_{1-\alpha/2}\frac{S}{\sqrt{n}}\le\mu\le\overline{X}+t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$; de modo que el I.C. final es:

$\left[\overline{X}-t_{1-\alpha/2}\dfrac{S}{\sqrt{n}},\overline{X}+t_{1-\alpha/2}\dfrac{S}{\sqrt{n}}\right]$, donde, ya sabemos, S es la desviación estándar de la muestra y $t_{1-\alpha/2}$ es el percentil $(1-\alpha/2)$ de la distribución t-Student de parámetro k=n-1 grados de libertad.

Observación:

Como en el ejemplo 1, podemos escribir el IC mediante $\mu=\overline{X}\pm t_{1-\alpha/2}\dfrac{S}{\sqrt{n}}$ con un 100(1-α)% de Confianza.

O alternativamente $\overline{X}-t_{1-\alpha/2}\dfrac{S}{\sqrt{n}}\le\mu\le\overline{X}+t_{1-\alpha/2}\dfrac{S}{\sqrt{n}}$ al 100(1-α)% de Confianza.

Ejemplo 5
En un estudio, se desea estimar el Promedio del Número diario de horas que trabaja un microempresario y para ello, se tomó una muestra piloto 7 microempresarios registrándose la cantidad de horas de trabajo en un día de semana. Los datos fueron: (12,11,14,10,9,9,8). Asuma normalidad de datos y calcule un I.C. para la Promedio de horas de trabajo/dia, al nivel 100(1-α)% =95%. ¿Se puede inferir que los microempresarios tienen una jornada promedio más larga que jornada legal del sector formal?

Solución:

- Aquí estamos en el contexto del problema 3: I.C. para μ cuando σ^2 es desconocida y por tanto aplicamos

$\mu=\overline{X}\pm t_{1-\alpha/2}\dfrac{S}{\sqrt{n}}$ con un 100(1-α)% de Confianza.

- Como vemos de los datos, n =7 y 100(1-α)%=95% ⇒ α=0.05 ⇒ 1-α/2=0.975 y $t_{1-\alpha/2}=t_{0.975}(6)=2.4469$
  (Los grados de libertad son n-1=7-1=6)

- Además, calculando los estadísticos $\overline{X}$ y S : $\overline{X}$ = (12+11+14+10+9+9+8)/7=73/7=10.43 horas y $S^2 = (\sum X_j^2 - n\overline{X}^2)/(n-1) = (787 - 7(10.43^2)/6 = 4.29 \Rightarrow S = \sqrt{4.29} = 2.07$ horas
- Luego: $\mu = 10.43 \pm 2.4469\frac{2.07}{7}$ μ = 10.43 ± 1.914 con 95% de confianza. O en formato de intervalo: 8.52 ≤ μ ≤ 12.34 con 95% de Confianza.
- Examinando el I.C., se puede inferir que, como el I.C. está totalmente a la izquierda de 8, los microempresarios tienen una jornada diaria de trabajo más larga que la jornada legal del sector formal.

Ejemplo 6

Se quiere introducir un producto nuevo en el mercado y para ello se solicitó a una muestra de consumidores potenciales del bien, que lo usaran durante 15 días, al término de los cuales, señalaron el precio que estarían dispuestos a pagar por el producto. Sin más información, un economista plantea que el precio X que se pagará, es una v.a. con distribución uniforme $X \sim U(x;0,\beta)$ y quiere usar el estimador máximo verosímil de β, $\hat{\beta} = Máx\{X_j\}$ para formar un intervalo de confianza de 95% para β. ¿Cuál sería ese intervalo?

Solución:

- $\hat{\beta} = Máx\{X_j\}$ no sirve como variable base o pivot, pues no contiene a β, pero sabemos que la función de distribución acumulativa de $\hat{\beta}$ es (ver los apuntes del capítulo 4):

  $F_{\hat{\beta}}(\hat{\beta}) = [F_X(\hat{\beta})]^n = \left(\frac{\hat{\beta}}{\beta}\right)^n \quad 0 \le \hat{\beta} \le \beta$ y por tanto la función de densidad es $f_{\hat{\beta}}(\hat{\beta}) = n\frac{\hat{\beta}^{n-1}}{\beta^n} \quad 0 \le \hat{\beta} \le \beta$

  Se intuye que definiendo $W = \frac{\hat{\beta}}{\beta}$ tendríamos una variable base pues el rango de $W = \frac{\hat{\beta}}{\beta}$ sería $0 \le W \le 1$ y

  $F_W(w) = P[W \le w] = P\left[\frac{\hat{\beta}}{\beta} \le w\right] = P[\hat{\beta} \le w\beta] = \left(\frac{w\beta}{\beta}\right)^n = (w)^n = w^n \quad 0 \le w \le 1$, que no depende de β.

  Ya tenemos la variable base $W = \frac{\hat{\beta}}{\beta}$.
- Sean valores a y b que cumplan $P(a \le W \le b) = 0.95$, trabajando dentro de la probabilidad: $0.95 = P(a \le W \le b) = P(a \le \frac{\hat{\beta}}{\beta} \le b) = F_W(b) - F_W(a) = b^n - a^n$. Por comodidad tomemos a y b tales que $P(W \le a) = P(W > b) = 0.025$ para tener el intervalo central [a,b] con probabilidad 95%. Así resulta:

  $P(W \le a) = F_W(a) = a^n = 0.025 \Rightarrow a = \sqrt[n]{0.025}$ y

  $P(W > b) = 1 - F_W(b) = 1 - b^n = 0.025 \Rightarrow b = \sqrt[n]{0.975}$

  Llegamos a $P(a \le W \le b) = 0.95 \Leftrightarrow P(\sqrt[n]{0.025} \le \frac{\hat{\beta}}{\beta} \le \sqrt[n]{0.975}) = 0.95$
- Despejando β:

  $0.95 = P(\sqrt[n]{0.025} \le \frac{\hat{\beta}}{\beta} \le \sqrt[n]{0.975}) = P(\frac{\sqrt[n]{0.025}}{\hat{\beta}} \le \frac{1}{\beta} \le \frac{\sqrt[n]{0.975}}{\hat{\beta}}) = P(\frac{\sqrt[n]{0.975}}{\hat{\beta}} \le \beta \le \frac{\sqrt[n]{0.025}}{\hat{\beta}})$ y como

  $\hat{\beta} = Máx\{X_j\}$, el I.C. de 95% para β es $\frac{\sqrt[n]{0.975}}{Máx\{X_j\}} \le \beta \le \frac{\sqrt[n]{0.025}}{Máx\{X_j\}}$

Ejemplo 7

La rentabilidad de una inversión en un sector de la economía es una v.a. $X \sim LogN(\mu,1)$ y se desea estimar μ_X mediante un Intervalo de Confianza (I.C.) de 95%, a partir de la siguiente muestra: 3, 5, 10, 3, 5, 8.

a) Construya un IC de 95% para μ, basándose en que $X \sim LogN(\mu,1) \Leftrightarrow \ln(X) \sim N(\mu,1)$.

b) Use el IC construido en a) para hallar un I.C. para μ_X. ¿Sería cierto que la rentabilidad promedio en este sector supera el 2.5%? Justifique.

Solución:

a)

$X \sim LogN(\mu,1) \Leftrightarrow Y := \ln X \sim N(\mu,1)$. Construyendo un I.C. para μ:

(a) Sea $Y := \ln X \sim N(\mu,1) \Rightarrow \overline{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{6}) \Rightarrow Z = \frac{\overline{Y}-\mu}{\frac{1}{\sqrt{6}}} \sim N(0,1)$. Podemos tomar como variable <u>base</u> del I.C. a Z

(b) Como el nivel de confianza pedido es 95%, planteamos valores *a* y *b* tales que $P(a \le Z = \frac{\overline{Y}-\mu}{\frac{1}{\sqrt{6}}} \le b) = 0.95$

donde *a* y *b* satisfacen $P(Z \le a) = P(Z > b) = 0.025$; de la tabla Z se obtiene *a*= -1.96 , *b*=1.96

(c) De b) despejamos μ, dejándola en el centro del intervalo:

$$0.95 = P(-1.96 \le \frac{\overline{Y}-\mu}{\frac{1}{\sqrt{6}}} \le 1.96) = P(-\frac{1.96}{\sqrt{6}} \le \overline{Y}-\mu \le \frac{1.96}{\sqrt{6}}) = P(-\overline{Y}-\frac{1.96}{\sqrt{6}} \le -\mu \le -\overline{Y}+\frac{1.96}{\sqrt{6}})$$

$= P(\overline{Y}-\frac{1.96}{\sqrt{6}} \le \mu \le \overline{Y}+\frac{1.96}{\sqrt{6}})$. Llamando $L_1 = \overline{Y}-\frac{1.96}{\sqrt{6}}$; $L_2 = \overline{Y}+\frac{1.96}{\sqrt{6}}$, tenemos el I.C. para μ

<u>Para tener el I.C. final, necesitamos $\overline{Y}$, como *Y=lnX*, tenemos</u>:

X	3	5	10	3	5	8
Y=lnX	1.10	1.61	2.30	1.10	1.61	2.08

así $\overline{Y} = 1.63$ y reemplazando en la fórmula del I.C. $L_1 = 1.63 - \frac{1.96}{\sqrt{6}} = 0.83$; $L_2 = 1.63 + \frac{1.96}{\sqrt{6}} = 2.43$

y con 95% de confianza podemos decir que $0.83 \le \mu \le 2.43$

b) En cuanto al I.C. para μ_X:

$X \sim LogN(\mu,1) \Rightarrow \mu_X = e^{\mu+\frac{1}{2}\sigma^2} = e^{\mu+\frac{1}{2}}$ y trabajando dentro del I.C. para μ:

$$0.95 = P(L_1 \le \mu \le L_2) = P(L_1 + \frac{1}{2} \le \mu + \frac{1}{2} \le L_2 + \frac{1}{2}) = P(e^{L_1+\frac{1}{2}} \le e^{\mu+\frac{1}{2}} \le e^{L_2+\frac{1}{2}}) = P(e^{L_1+\frac{1}{2}} \le \mu_X \le e^{L_2+\frac{1}{2}}),$$

o sea $e^{L_1+\frac{1}{2}} \le \mu_X \le e^{L_2+\frac{1}{2}}$ con 95% de confianza. Reemplazando valores

$e^{0.83+\frac{1}{2}} \le \mu_X \le e^{2.43+\frac{1}{2}} \Leftrightarrow 3.78 \le \mu_X \le 18.73$ con 95% de confianza. Finalmente, como $2.5 < 3.78 \le \mu_X$, podemos considerar cierta la afirmación.