Asignatura	**:**	**INFERENCIA ESTADÍSTICA I**
Titulación	**:**	**DIPLOMADO EN ESTADÍSTICA**
Profesor	**:**	**ISABEL MOLINA PERALTA**
Capítulo 5	**:**	**INTERVALOS DE CONFIANZA**

5.1 MÉTODO DE LA CANTIDAD PIVOTAL

$X \sim F(x;\theta)$, $\theta \in \Theta \subset I\!R$, $(X_1, \ldots, X_n)$ m.a.s. de X

$P(x_1, \ldots, x_n; \theta)$ función de probabilidad conjunta

Intervalo de confianza al nivel α:

$$IC_{1-\alpha}(\theta) = [T_1(x_1, \ldots, x_n), T_2(x_1, \ldots, x_n)],$$

donde $T_1 = T_1(X_1, \ldots, X_n)$ y $T_2 = T_2(X_1, \ldots, X_n)$ son estadísticos que verifican

$$P(T_1 \leq \theta \leq T_2) \geq 1 - \alpha, \ \forall \theta \in \Theta.$$

Nota 1 Hay que fijar el *nivel de significación* α antes de construir el intervalo.

Nota 2 Interpretación del intervalo: Si calculamos el intervalo para muchas muestras, el $(1-\alpha)100\,\%$ de ellos contendrá al verdadero valor de θ.

Nota 3 T_1 y T_2 se llaman *límite inferior* y *límite superior* de confianza para θ. Es posible que solo se necesite uno de estos límites.

Pivote: Función de la muestra y del parámetro desconocido $Z = Z(\theta; X_1, \ldots, X_n)$ biyectiva en θ y con distribución de probabilidad completamente conocida.

Método de la cantidad pivotal:

(1) Encontrar un pivote $Z(\theta)$. A menudo se encuentra a partir de un estimador $\hat{\theta}$.

(2) Seleccionar dos constantes c_1 y c_2 que verifiquen

$$P_Z(c_1 \leq Z(\theta) \leq c_2) \geq 1 - \alpha.$$

(4) Despejar θ de las desigualdades $Z(\theta) \leq c_2$ y $Z(\theta) \geq c_1$, y se obtienen otras de la forma $\theta \geq T_1$ y $\theta \leq T_2$. Entonces, el intervalo de confianza al nivel α es $IC_{1-\alpha}(\theta) = [T_1, T_2]$.

Valor crítico α de una v.a. $X \sim \mathcal{F}$: Valor de X que deja a su derecha una probabilidad de α. Es decir, es el valor $\mathcal{F}_\alpha$ tal que

$$P(X \geq \mathcal{F}_\alpha) = \alpha.$$

Nota 4 Si en el paso (2), $Z(\theta) \sim \mathcal{F}$ y la probabilidad α se reparte por igual a ambos lados de la distribución $\mathcal{F}$, entonces

$$c_2 = \mathcal{F}_{\alpha/2}, \quad c_1 = \mathcal{F}_{1-\alpha/2}.$$

Ejemplo 1 Se obtiene una observación X de una distribución exponencial de media θ. Utilice X para construir un intervalo de confianza para θ al nivel de confianza 0.90. Para ello, calcule la distribución de la variable aleatoria $Z = X/\theta$.

...

5.2 INTERVALOS DE CONFIANZA PARA POBLACIONES NORMALES

$X \sim N(\mu, \sigma^2)$, $(X_1, \ldots, X_n)$ m.a.s. de X.

Intervalo de confianza para μ con σ^2 conocida:

(1) Pivote: Tomamos un estimador de μ con distribución conocida; por ejemplo, $\hat{\mu} = \bar{X}$,

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

Una transformación que hace desaparecer μ de la distribución de $\bar{X}$ es

$$Z = \bar{X} - \mu \sim N(0, \sigma^2/n).$$

No se dispone de tablas de probabilidades para esta distribución.

Por tanto, un pivote más práctico es

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

(2) Las constantes c_1 y c_2 que verifican $P(c_1 \leq Z \leq c_2) = 1 - \alpha$ son

$$c_1 = Z_{1-\alpha/2} = -Z_{\alpha/2}, \quad c_2 = Z_{\alpha/2}.$$

(3) Despejamos μ dentro de la probabilidad

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

Un intervalo de confianza para μ es

$$IC_{(1-\alpha)}(\mu) = \bar{X} \mp Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Intervalo de confianza para μ con σ^2 desconocida:

(1) En este caso,

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

no es un pivote, ya que σ^2 es desconocido. Estimando σ^2 de forma insesgada con la cuasivarianza muestral S'^2, obtenemos

$$T = \frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t_{n-1},$$

(2)...

(3)...

Intervalo de confianza para σ^2:

(1) Estimador insesgado de σ^2: S'^2. Distribución

(2) ...

(3) ...

Intervalo de confianza para la diferencia de medias con varianzas conocidas:

$(X_{11}, \ldots, X_{1n_1})$ m.a.s. de $N(\mu_1, \sigma_1^2)$, con σ_1^2 conocida.

$(X_{21}, \ldots, X_{2n_2})$ m.a.s. de $N(\mu_2, \sigma_2^2)$ con σ_2^2 conocida.

(1) Estimador insesgado de $\theta = \mu_1 - \mu_2$: ...

Distribución:...

Pivote:

$$Z = ...$$

(2) ...

(3) ...

Intervalo de confianza para la diferencia de medias, con varianzas desconocidas pero iguales

$(X_{11}, \ldots, X_{1n_1})$ m.a.s. de $N(\mu_1, \sigma^2)$,

$(X_{21}, \ldots, X_{2n_2})$ m.a.s. de $N(\mu_2, \sigma^2)$, σ^2 es desconocida.

(1) Estimador de $\theta = \mu_1 - \mu_2$: ...

Distribución: ...

Estimador insesgado de σ^2: ...

Distribución del estimador de σ^2: ...

Pivote:

$$T = ...$$

Distribución: ...

(2) Constantes c_1 y c_2 tales que $P(c_1 \leq T \leq c_2) \geq 1 - \alpha$,

$$c_1 = ... \quad , \quad c_2 = ...$$

(3) Despejando $\mu_1 - \mu_2$ dentro de la probabilidad encontramos un intervalo de confianza para $\mu_1 - \mu_2$ al nivel de confianza $1 - \alpha$

$$IC_{(1-\alpha)}(\mu_1 - \mu_2) = \ldots \qquad .$$

Intervalo de confianza para el cociente de varianzas

$(X_{11}, \ldots, X_{1n_1})$ m.a.s. de $N(\mu_1, \sigma_1^2)$, con μ_1 y σ_1^2 desconocidas.

$(X_{21}, \ldots, X_{2n_2})$ m.a.s. de $N(\mu_2, \sigma_2^2)$ con μ_2 y σ_2^2 desconocida.

(1) Estimador de $\theta = \sigma_1^2/\sigma_2^2$:...

Distribución: ...

(2) Repartiendo la probabilidad α igualmente entre ambas colas de la distribución, obtenemos

$$c_1 = \ldots \quad , \quad c_2 = \ldots$$

(3) Despejando el parámetro θ de las desigualdades, obtenemos el intervalo de confianza para $\theta = \sigma_1^2/\sigma_2^2$ al nivel $1-\alpha$

$$IC_{(1-\alpha)}(\sigma_1^2/\sigma_2^2) = \ldots$$

Nota 5 Se verifica

$$F \sim \mathcal{F}_{\nu_1,\nu_2} \longrightarrow F' = 1/F \sim \mathcal{F}_{\nu_2,\nu_1}.$$

Entonces

$$\alpha = P(F > \mathcal{F}_{\nu_1,\nu_2,\alpha}) = P(1/F < 1/\mathcal{F}_{\nu_1,\nu_2,\alpha}) \Leftrightarrow P(F' > 1/\mathcal{F}_{\nu_1,\nu_2,\alpha}) = 1-\alpha;$$

es decir,

$$\mathcal{F}_{\nu_2,\nu_1,1-\alpha} = 1/\mathcal{F}_{\nu_1,\nu_2,\alpha}.$$

5.3 INTERVALOS DE CONFIANZA ASINTÓTICOS

$(X_1, \ldots, X_n)$ m.a.s. de $X \sim F(x, \theta)$, $\theta \in \Theta \subset I\!R$ desconocido.

La distribución F no es normal, o es desconocida.

Pivote asintótico:

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\hat{V}(\hat{\theta})}} \xrightarrow{d} N(0, 1).$$

Intervalo de confianza asintótico para θ al nivel $1 - \alpha$:

$$IC_{(1-\alpha)}(\theta) = ... \qquad\qquad\qquad\qquad\qquad\qquad .$$

Es un intervalo aproximado para θ, acercándose al intervalo correcto al aumentar el tamaño muestral n.

Intervalo de confianza asintótico para una media

X v.a. con $E(X) = \mu$ y $V(X) = \sigma^2$, siendo ambos parámetros desconocidos, $(X_1, \ldots, X_n)$ m.a.s. de X.

Estimador de μ: ...

Distribución: ...

Pivote:

$$Z = \ldots \qquad\qquad\qquad\qquad .$$

Intervalo de confianza para μ al nivel de confianza $1 - \alpha$:

$$IC_{(1-\alpha)}(\mu) = \ldots$$

Ejemplo 2 Se registraron los tiempos utilizados en la compra para $n = 64$ clientes seleccionados al azar en el supermercado local. La media y la cuasivarianza de los 64 tiempos de compra fueron 33 minutos y 256 respectivamente. Estime el promedio real, μ, del tiempo utilizado por clientes en la compra, con un nivel de confianza de 0.90.

...

Ejemplo 3 Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. con distribución $Pois(\lambda)$. Calculamos un intervalo de confianza asintótico al nivel α para λ.

...

Intervalo de confianza asintótico para una proporción

$(X_1, \ldots, X_n)$ m.a.s. $Bern(p)$

Estimador de p: ...

Distribución: ...

Pivote:

$$Z = ...$$

Intervalo de confianza asintótico para p:

$$IC_{(1-\alpha)}(p) = ...$$

Ejemplo 4 Dos marcas de refrigeradores, A y B, tienen una garantía de un año. En una muestra aleatoria de 50 refrigeradores de la marca A, 12 se estropearon antes de terminar el periodo de garantía, y de una muestra aleatoria de 60 refrigeradores de la marca B, se estropearon 12 antes del periodo de garantía. Estime la diferencia real entre las proporciones de fallo durante el periodo de garantía, al nivel de confianza del 98 %.

...

5.4 INTERVALOS DE CONFIANZA BOOTSTRAP

A menudo no existe una expresión explícita para la varianza $V(\hat{\theta})$ de un estimador $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ de θ. Consecuentemente, tampoco se puede calcular analíticamente un intervalo de confianza para θ. El método de remuestreo llamado *bootstrap* permite estimar la varianza de un estimador y también obtener intervalos de confianza.

Idea: Considerar la muestra observada $(x_1, \ldots, x_n)$ como si fuera la población y extraer muchas muestras aleatorias con reemplazamiento de ella. Calculando el estimador para cada muestra podemos estimar la distribución en el muestreo de $\hat{\theta}$.

(1) Extraer una m.a.s. con reemplazamiento de $(x_1, \ldots, x_n)$: $(x_1^*, \ldots, x_n^*)$

(2) Calcular el valor del estimador para la muestra bootstrap, $\hat{\theta}^* = \hat{\theta}(x_1^*, \ldots, x_n^*)$.

(3) Repetir (1) y (2) un número grande de veces B. Los valores del estimador en las B repeticiones son: $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

(4) Calcular la media y la varianza de los valores del estimador

$$\hat{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*, \quad \hat{V}(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b^* - \hat{\theta}^*)^2$$

(5) Ordenar los valores del estimador de menor a mayor $\hat{\theta}^*_{(1)} \leq \ldots \leq \hat{\theta}^*_{(B)}$. Seleccionar los cuantiles $\alpha/2$ y $1{-}\alpha/2$, dados por $\hat{\theta}^*_{\lfloor B\,\alpha/2\rfloor}$ y $\hat{\theta}^*_{\lfloor B\,(1-\alpha/2)\rfloor}$. El intervalo de confianza al nivel α para θ es entonces

$$IC_{(1-\alpha)}(\theta) = \left[\hat{\theta}^*_{\lfloor B\,\alpha/2\rfloor}, \hat{\theta}^*_{\lfloor B\,(1-\alpha/2)\rfloor}\right].$$

Actividad 9: Rellenar los espacios adecuados en este tema razonando las respuestas y entregarlo a la profesora el lunes 4 de Junio.