

## Intervalos de Confianza

- *Interval that contains an unknown quantity with a given frequency-*  
*All of Statistics, Wasserman*

1. Implemente una función **intervalo.mu.asin** que tenga por input un conjunto de datos  $x_1, \dots, x_n$ , provenientes de una muestra de  $X$ , el nivel  $1 - \alpha$  y devuelva el intervalo de confianza asintótico  $1 - \alpha$  para  $\mu = \mathbb{E}(X)$ .

**Nivel de Cubrimiento empírico:** El nivel de cubrimiento empírico (de un procedimiento) se define como la proporción de veces que los intervalos (construidos con el procedimiento) utilizando datos simulados contiene a  $\mu$  (o  $\theta$ ), en cierta cantidad de  $Nrep$  replicaciones.

2. **Simulación 1: bajo normalidad** Genere variables con distribución normal de media  $\mu = 0$  y  $\sigma = 0,1, 1, 10$ . Calcule el cubrimiento empírico del intervalo de confianza asintótico de nivel  $1 - \alpha$ , definido en **intervalo.mu.asin**, para  $\alpha = 0,05$ ,  $n = 5$ ,  $n = 10$ ,  $n = 30$ ,  $n = 50$ ,  $n = 100$ ,  $n = 1000$ , utilizando  $Nrep = 1000$  replicaciones, y complete la siguiente tabla. En cada caso, calcule el promedio de las longitudes en las  $Nrep = 1000$  replicaciones e incluya el valor en la tabla (long). Comente los resultados observados.

Modelo	Normales con media $\mu = 0$ . Nivel nominal=0.9					
	n=5	n=10	n=30	n=50	n=100	n=1000
$\sigma = 0,1$						
$\sigma = 1$			cub (long)			
$\sigma = 10$						

Cuadro 1: Nivel de Cubrimiento Empírico

Indique a que valor debe aproximarse el nivel de cubrimiento empírico. Comente los resultados obtenidos.

3. **Simulación 3: Uniformes** Genere variables con distribución uniforme en el intervalo  $[0, \theta]$ , para  $\theta = 3, 10, 100$ . Calcule el cubrimiento empírico del intervalo de confianza asintótico de nivel  $1 - \alpha$ , definido en **intervalo.mu.asin**, para  $\alpha = 0,05$ ,  $n = 5$ ,  $n = 10$ ,  $n = 30$ ,  $n = 50$ ,  $n = 100$ ,  $n = 1000$ , utilizando  $Nrep = 1000$  replicaciones. En cada caso, calcule el promedio de las longitudes en las  $Nrep = 1000$  replicaciones e informe los valores obtenidos. (long). Comente los resultados observados.

4. Implemente una función **intervalo.mu.exacto.normal** que tenga por input un conjunto de datos  $x_1, \dots, x_n$ , provenientes de una muestra  $\mathcal{N}(\mu, \sigma^2)$ , el nivel  $1 - \alpha$  y devuelva el intervalo de confianza exacto de nivel  $1 - \alpha$  para  $\mu$ .
5. Repita el ítem 2. utilizando ahora la función **intervalo.mu.exacto.normal** y compare los resultados obtenidos (los de ahora con los del ítem 2.) Comente los resultados observados.
6. **Diferencias de medias:** Supongamos que tenemos dos conjuntos de observaciones provenientes de dos distribuciones  $N(\mu_X, \sigma_X^2)$  y  $N(\mu_Y, \sigma_Y^2)$ , asumamos que  $\sigma_X = \sigma_Y$ . El objetivo es decidir a partir de las observaciones si las medias son o no iguales.
  - a) Implemente una función **dif.mu.interseccion** que tenga por input el nivel  $1 - \alpha$ , un conjunto de datos  $x_1, \dots, x_n$  y un conjunto de datos  $y_1, \dots, y_m$  y devuelva un 1 si los intervalos exactos de nivel  $1 - \alpha$  para  $\mu_X$  y para  $\mu_Y$  se intersecan y un 0 si no se intersecan. Es decir, si los intervalos obtenidos en el ítem 4 para cada conjunto de datos se intersecan o no.
  - b) Implemente una función **dif.mu.intervalo** que tenga por input el nivel  $1 - \alpha$ , un conjunto de datos  $x_1, \dots, x_n$  y un conjunto de datos  $y_1, \dots, y_m$  y devuelva un intervalo exacto de nivel  $1 - \alpha$  para  $\mu_X - \mu_Y$  y un 1 si el 0 está en el intervalo y un 0 si no.
  - c) Genere variables  $X$  con distribución normal de media  $\mu_X = 1$  y  $\sigma_X = 2$  y variables  $Y$   $\mu_Y = 2$  y  $\sigma_Y = 2$ . Calcule la proporción de veces que el método de intersección decide que ambas medias son iguales y calcule el cubrimiento empírico del intervalo de confianza exacto de nivel  $1 - \alpha$  para la diferencia de medias (es decir la proporción de veces que el intervalo incluye al 0). Considere  $\alpha = 0,05$ ,  $n = 5$ ,  $n = 10$ ,  $n = 30$ ,  $n = 50$ ,  $n = 100$ ,  $n = 1000$ , utilizando  $Nrep = 1000$  replicaciones. Comente los resultados observados.
7. El objetivo de este ejercicio es comparar empíricamente la performance de diferentes intervalos. Por un lado, utilizaremos intervalos construidos estimando la varianza asintótica mediante (i) bootstrap y (ii) propagación de errores; luego consideraremos (iii) intervalos bootstrap por percentil. Finalmente, contruiremos (iv) intervalos para una proporción, siendo que el parámetro de interés es una probabilidad. Estudiaremos el nivel de cubrimiento empírico y la longitud media de cada uno procedimientos mencionados.

Generaremos datos asumiendo que provienen de una familia exponencia  $\mathcal{E}(\lambda): X \sim \mathcal{E}(\lambda)$ . Procuramos estimar

$$\theta = \mathbb{P}(X > 1)$$

- a) Obtenga una fórmula para el parámetro de interés en función de  $\lambda$ .
- b) Genere datos utilizando  $\lambda = 3$ . Considere  $n = 100, 200, 500$  y estudie el nivel de cubrimiento empírico en  $Nrep = 1000$  replicaciones de los intervalos propuestos, tomando  $Nboot = 1000$  muestras bootstrap.