UNIVERSIDAD DE ATACAMA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE MATEMÁTICA
ESTADÍSTICA APLICADA

PRÁCTICA 7

Profesor: Hugo S. Salinas. Segundo Semestre de 2024

---

1. Interpretación del Intervalo de Confianza (IC): ¿por qué el término *confianza*?

Para la determinación de un IC nos apoyaremos en el hecho de que, *antes de obtener la muestra*, la media muestral $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es una variable aleatoria *con distribución* $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. A partir de aquí hemos deducido que:

$$P\left(\mu \in \left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha \tag{1}$$

Por tanto, *mientras no se haya obtenido la muestra*, los extremos del intervalo son variables aleatorias y se puede calcular la probabilidad de que dicho intervalo contenga a μ. Ahora bien, una vez que se ha obtenido una muestra, los extremos del intervalo son valores fijos. En este momento, el valor de μ estará comprendido entre ellos o no, pero ya no cabe hablar de la probabilidad de que ésto ocurra.

Podemos utilizar el símil de un lanzador de cuchillos circense que se dispone a lanzar un cuchillo contra una manzana que está arriba de la cabeza de su ayudante y lo hace con los ojos vendados. Él sabe, por su experiencia, que la probabilidad de acertar en la manzana es del 95 %. Ahora bien, una vez que ha lanzado el cuchillo habrá acertado o no, pero ya no se puede hablar de la probabilidad de que acierte. Si el lanzador continúa con los ojos vendados tras el lanzamiento, *puede confiar* en que ha acertado (incluso, tener mucha confianza en ello, ya que sabe que tiene muy buena puntería), pero no puede estar del todo seguro.

La situación de un investigador que construye un IC a partir de unos datos experimentales es análoga a la del lanzador de cuchillos que nunca se quita la venda de los ojos: *antes de tomar la muestra* sabe que la probabilidad de que el intervalo contenga al parámetro es del 95 %; por tanto, cuando tome los datos y obtenga un intervalo concreto, puede tener mucha confianza (que puede valorar en ese mismo 95 %) en que el intervalo habrá *capturado* al parámetro, pero no puede saber con seguridad si lo ha capturado o no, ya que el valor del parámetro sigue siendo desconocido.

De un modo más general, si para un parámetro θ de una distribución de probabilidad disponemos de dos estadísticos $\theta_1(\boldsymbol{X})$ y $\theta_2(\boldsymbol{X})$ tales que:

$$P(\theta \in [\theta_1(\boldsymbol{X}), \theta_2(\boldsymbol{X})]) = 1 - \alpha$$

siendo $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ una muestra aleatoria de dicha distribución, entonces cabe esperar que el $100(1-\alpha)\,\%$ de los intervalos construidos de esta manera contengan a θ y, obviamente, que el restante $100\alpha\,\%$ no lo contengan.

Una vez que obtenemos una muestra particular $(x_1, x_2, \ldots, x_n)$ y calculamos los valores $\widehat{\theta}_1 = \theta_1(x_1, x_2, \ldots, x_n)$ y $\widehat{\theta}_2 = \theta_2(x_1, x_2, \ldots, x_n)$, tenemos un intervalo concreto $[\widehat{\theta}_1, \widehat{\theta}_2]$. En realidad *no sabemos* si este intervalo contiene o no a θ, pero *confiamos* en que sea uno de entre el $100(1-\alpha)\,\%$ de intervalos que contienen al parámetro. De ahí que valoremos nuestra confianza en $1-\alpha$.

1. Copia el siguiente código R que simula la obtención de 1000 muestras de tamaño 50 de una variable aleatoria $X \sim N(\mu = 2, \sigma = 2)$. Para cada muestra se calculan la media muestral $\overline{X}$ y el IC para μ calculado a partir de la expresión dada en (1) donde $\sigma = 2$ y $1-\alpha = 0.95$.

```
simulacion <- function(n) {
muestra <- rnorm(n, 10, 2)
intervalo <- mean(muestra) + c(-1, 1) * qnorm(0.975) * 2/sqrt(n)
return(intervalo)
}
IC <- t(replicate(100, simulacion(50)))
IC
```

   Ejecutar este código en Jamovi y comenta los resultados.

2. Ahora contamos cuántos de los 100 intervalos contienen a μ. Como hemos elegido una confianza del 95 %, esperamos que aproximadamente el 95 % de los intervalos (esto es, unos 95), contengan al parámetro. Ejecuta el siguiente código en Jamovi y comenta los resultados:

```
n.IC <- 0 # Vamos a contar los intervalos
for (k in 1:100) if ((IC[k, 1] <= 10) & (10 <= IC[k,2])) n.IC <- n.IC + 1
n.IC
```

3. La Figura 1 ilustra los 100 IC generados en la simulación anterior. La línea vertical indica el valor del parámetro $\mu = 10$. De los intervalos presentados, 93 abarcan el valor de μ, mientras que 7 (destacados en rojo) no lo incluyen. Es importante recordar que, en la práctica, el investigador solo dispone de una única muestra, no de 100 ni de 1000. Con un nivel de confianza del 95 %, el investigador asume que ha capturado el parámetro. Sin embargo, si la muestra única resulta en un intervalo marcado en rojo, el parámetro se le habría escapado, y lamentablemente, el investigador no tendría forma de saberlo.

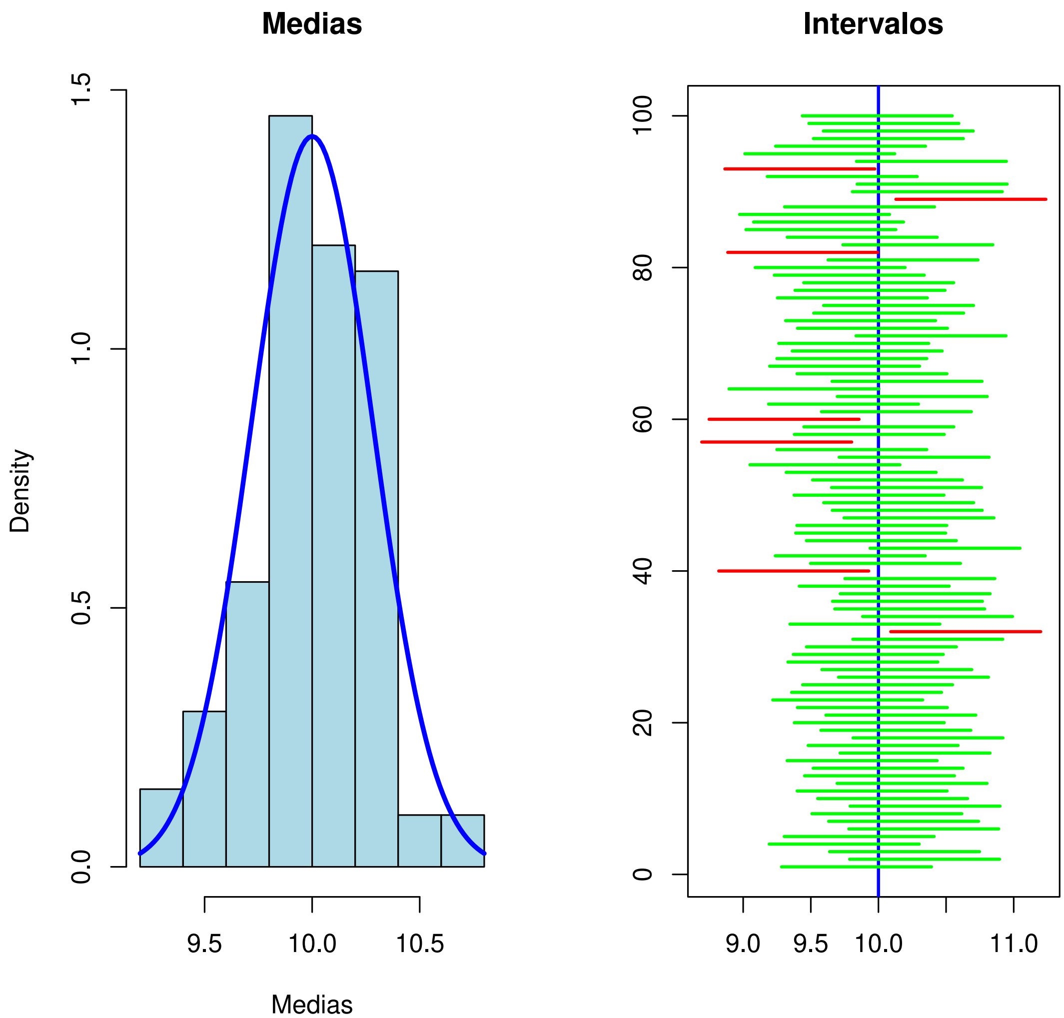

Figura 1: Distribución de la media muestral y los 100 IC para $\mu = 10$

4. Ejecutar el siguiente código en Jamovi y comentar los resultados.

```
# Configuración de los parámetros
# set.seed(123)  # Para reproducibilidad
n_muestras <- 100
n <- 50
nivel_confianza <- 0.95
alfa <- 1 - nivel_confianza

# Generación de las medias muestrales y los intervalos de confianza
medias <- numeric(n_muestras)
lim_inf <- numeric(n_muestras)
lim_sup <- numeric(n_muestras)

for (i in 1:n_muestras) {
```

```
  muestra <- rnorm(n,10,2)  # Muestra de tamaño n de una normal estándar
  medias[i] <- mean(muestra)
  error_estandar <- 2/sqrt(n)

  # Intervalo de confianza para la media
  lim_inf[i] <- medias[i] - qnorm(1 - alfa / 2) * error_estandar
  lim_sup[i] <- medias[i] + qnorm(1 - alfa / 2) * error_estandar
}

# Configurar el diseño de los gráficos en dos paneles
par(mfrow = c(1, 2))

# Gráfico del histograma de las medias
hist(medias, main = "Medias", col = "lightblue", xlab = "", ylab = "Density", freq=F)
curve(dnorm(x,10,2/sqrt(n)),add=T, col="blue", lwd=3)

# Gráfico de los intervalos de confianza
plot(c(min(lim_inf), max(lim_sup)), c(1, n_muestras), type = "n", xlab = "",
ylab = "", main = "Intervalos")
abline(v = 10, col = "blue", lwd=2)  # Línea vertical en 10

for (i in 1:n_muestras) {
  if (lim_inf[i] <= 10 && lim_sup[i] >= 10) {
    # Intervalo que contiene 0 (verde)
    segments(lim_inf[i], i, lim_sup[i], i, col = "green", lwd=2)
  } else {
    # Intervalo que no contiene 10 (rojo)
    segments(lim_inf[i], i, lim_sup[i], i, col = "red", lwd=2)
  }
}
```

5. Replicar esta actividad para encontrar un IC para la media de una población normal.