



# Métodos Estadísticos y Simulación

Unidad 5: Inferencia Estadística: Estimación

Mg. J. Eduardo Gamboa U.

2025-06-09

## Estimación puntual

Sea  $X_1, \dots, X_n$  una muestra de tamaño  $n$  de una población con parámetro  $\theta$ . Se denomina estimador puntual de  $\theta$  a cualquier estadístico  $\hat{\Theta} = h(X_1, \dots, X_n)$  cuyo valor  $\hat{\theta} = h(x_1, \dots, x_n)$  dará una estimación puntual de  $\theta$ . En este caso,  $\Theta$  es una variable aleatoria y  $\hat{\theta}$  es un número (aunque en el caso particular de la moda podría no serlo).

Estimador puntual de la media:

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Estimador puntual de la variancia:

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Estimador puntual de la proporción:

$$\hat{\pi} = p = \frac{\text{Número de éxitos}}{n}$$

## Estimación intervalar

Sea  $X_1, \dots, X_n$  una muestra aleatoria de tamaño  $n$  de una población con parámetro  $\theta$ , cuyos valores observados o datos respectivos son  $x_1, \dots, x_n$ .

Entonces, siendo  $a = h_1(x_1, \dots, x_n)$  y  $b = h_2(x_1, \dots, x_n)$ , valores numéricos calculados a partir de los datos de la muestra, se dice que el intervalo  $[a, b]$  tiene un nivel de confianza del  $(1 - \alpha) \times 100\%$  de contener el parámetro  $\theta$ , o que  $\theta \in [a, b]$  con un nivel de confianza del  $(1 - \alpha) \times 100\%$ .

**Interpretación:** Con un nivel de confianza del  $(1 - \alpha) \times 100\%$ , se estima que el parámetro  $\theta$  está contenido en el intervalo  $[a, b]$ .

Dicho de otro modo, si se repitiera este experimento muchas veces, cada vez obteniendo una nueva muestra aleatoria del mismo tamaño y construyendo un nuevo intervalo mediante la misma regla, entonces aproximadamente el  $(1 - \alpha)\%$  de esos intervalos incluirían el verdadero valor del parámetro  $\theta$ .

# Estimación de la media por intervalo de confianza

## Caso teórico: varianza poblacional conocida

Si  $X_1, \dots, X_n$  es una muestra aleatoria de una población Normal con media  $\mu$  y  $\sigma^2$  conocida, el intervalo con un nivel de confianza del  $(1 - \alpha) \times 100\%$  para la media  $\mu$  se obtiene mediante:

$$\left( \underbrace{\bar{X} - Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}}_a \leq \mu \leq \underbrace{\bar{X} + Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}}_b \right)$$

donde  $a$  y  $b$  son valores numéricos que representan el Límite inferior y Límite superior del intervalo de confianza. En caso  $n > 30$ , el requisito de normalidad de la variable se flexibiliza.

**Interpretación:** Con un nivel de confianza del  $(1 - \alpha) \times 100\%$ , se estima que la media poblacional  $\mu$  esté contenida en el intervalo  $[a, b]$ , es decir el procedimiento utilizado para construir este intervalo genera, en el largo plazo, intervalos que contienen la verdadera media poblacional  $\mu$  en aproximadamente el  $(1 - \alpha) \times 100\%$  de los casos.

## Caso realista: varianza poblacional desconocida

Si  $X_1, \dots, X_n$  es una muestra aleatoria de una población Normal con media  $\mu$  y  $\sigma^2$  desconocida, el intervalo con un nivel de confianza del  $(1 - \alpha) \times 100\%$  para la media  $\mu$  se obtiene mediante:

$$\left( \underbrace{\bar{X} - t_{(1-\alpha/2; n-1)} \frac{s}{\sqrt{n}}}_a \leq \mu \leq \underbrace{\bar{X} + t_{(1-\alpha/2; n-1)} \frac{s}{\sqrt{n}}}_b \right)$$

donde  $a$  y  $b$  son valores numéricos que representan el Límite inferior y Límite superior del intervalo de confianza. En caso  $n > 30$ , el requisito de normalidad de la variable se flexibiliza.

**Interpretación:** Con un nivel de confianza del  $(1 - \alpha) \times 100\%$ , se estima que la media poblacional  $\mu$  esté contenida en el intervalo  $[a, b]$ , es decir el procedimiento utilizado para construir este intervalo genera, en el largo plazo, intervalos que contienen la verdadera media poblacional  $\mu$  en aproximadamente el  $(1 - \alpha) \times 100\%$  de los casos.

## **Ejemplo 1**

Una investigadora estudia la longitud de las hojas de una planta nativa y desea estimar la media poblacional. Para ello, mide una muestra aleatoria de 8 hojas, obteniendo los siguientes valores (en centímetros):

7.8, 8.2, 7.5, 8.0, 8.3, 7.9, 7.6, 8.1

Se desea construir un intervalo de confianza del 95% para la media poblacional de la longitud de las hojas, asumiendo que la variable sigue una distribución normal y que la varianza poblacional es desconocida.

```
hojas <- c(7.8, 8.7, 7.2, 8.0, 8.3, 7.9, 7.6, 8.1)
library(magrittr)
hojas |> t.test() |> use_series(conf.int)
```

```
[1] 7.573459 8.326541
attr(,"conf.level")
[1] 0.95
```

Con un nivel de confianza del 95%, podemos afirmar que el verdadero valor de la longitud promedio de las hojas está entre 7.69 cm y 8.16 cm.

```
hojas |> t.test(conf.level = 0.90) |> use_series(conf.int) # para otro nivel
```

```
[1] 7.648309 8.251691
attr(,"conf.level")
[1] 0.9
```

```
n      <- hojas |> length()
x_bar  <- hojas |> mean()
s      <- hojas |> sd()
gl     <- n - 1
alpha  <- 0.05
tcrit  <- qt(1 - alpha/2, df = gl)
se     <- s / sqrt(n) # error estándar
li     <- x_bar - tcrit * se
ls     <- x_bar + tcrit * se
c(li,ls)
```

```
[1] 7.573459 8.326541
```



Notar que aquí no es válida la aproximación a la distribución Normal, ya que  $n = 8$

```
n      <- hojas |> length()
x_bar  <- hojas |> mean()
s      <- hojas |> sd()
alpha  <- 0.05
zcrit  <- qnorm(1 - alpha/2)
se     <- s / sqrt(n) # error estándar
li     <- x_bar - zcrit * se
ls     <- x_bar + zcrit * se
c(li,ls)
```

```
[1] 7.637897 8.262103
```

## Ejemplo 2

Una municipalidad metropolitana desea estimar el tiempo promedio que tardan los ciudadanos en llegar a sus centros de trabajo usando transporte público. Con este objetivo, se encuestó a una muestra aleatoria de 60 personas que se movilizan diariamente en bus o metro.

Los tiempos registrados (en minutos) fueron los siguientes:

78, 82, 85, 80, 77, 83, 79, 81, 86, 80, 82, 84, 78, 87, 81, 79, 83,  
82, 80, 84, 81, 79, 85, 80, 82, 78, 86, 81, 80, 83, 77, 79, 84, 82,  
80, 78, 86, 83, 79, 80, 82, 81, 77, 85, 80, 79, 83, 82, 78, 84, 81,  
79, 80, 77, 86, 82, 80, 81, 83, 78

Se desea construir un intervalo de confianza del 95% para el tiempo promedio de traslado al trabajo, asumiendo que la varianza poblacional es desconocida.

```
tiempos = c(78, 82, 85, 80, 77, 83, 79, 81, 86, 80, 82, 84, 78, 87,  
81, 79, 83, 82, 80, 84, 81, 79, 85, 80, 82, 78, 86, 81, 80, 83, 77,  
79, 84, 82, 80, 78, 86, 83, 79, 80, 82, 81, 77, 85, 80, 79, 83, 82,  
78, 84, 81, 79, 80, 77, 86, 82, 80, 81, 83, 78)  
tiempos |> t.test() |> use_series(conf.int)
```

```
[1] 80.52268 81.87732  
attr(,"conf.level")  
[1] 0.95
```

Con un nivel de confianza del 95%, podemos afirmar que el verdadero tiempo promedio de traslado al trabajo está contenido entre 80.52 y 81.88 minutos.

```
n      <- tiempos |> length()
x_bar  <- tiempos |> mean()
s      <- tiempos |> sd()
gl     <- n - 1
alpha  <- 0.05
tcrit  <- qt(1 - alpha/2, df = gl)
se     <- s / sqrt(n) # error estándar
li     <- x_bar - tcrit * se
ls     <- x_bar + tcrit * se
c(li,ls)
```

```
[1] 80.52268 81.87732
```

Notar que aquí sí es válida la aproximación a la distribución Normal, ya que  $n = 60 > 30$

```
n      <- tiempos |> length()
x_bar  <- tiempos |> mean()
s      <- tiempos |> sd()
alpha  <- 0.05
zcrit  <- qnorm(1 - alpha/2)
se     <- s / sqrt(n) # error estándar
li     <- x_bar - zcrit * se
ls     <- x_bar + zcrit * se
c(li,ls)
```

```
[1] 80.53657 81.86343
```

## Estimación de la varianza por intervalo de confianza

Si  $X_1, \dots, X_n$  es una muestra aleatoria de una población Normal con  $\sigma^2$  desconocida, el intervalo con un nivel de confianza del  $(1 - \alpha) \times 100\%$  para la variancia  $\sigma^2$  se obtiene mediante

$$\underbrace{\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2;n-1)}}}_a \leq \sigma^2 \leq \underbrace{\frac{(n-1)s^2}{\chi^2_{(\alpha/2;n-1)}}}_b$$

donde a y b son valores numéricos que representan el Límite inferior y Límite superior del intervalo de confianza. Para cualquier valor de  $n$  se debe verificar Normalidad.

**Interpretación:** Con un nivel de confianza del  $(1 - \alpha) \times 100\%$ , se estima que la variancia poblacional  $\sigma^2$  esté contenida en el intervalo  $[a, b]$ .

Si se desea obtener los límites de confianza para la desviación estándar se obtiene la raíz cuadrada en la expresión anterior obteniéndose:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2;n-1)}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{(\alpha/2;n-1)}}}$$

### Ejemplo 3

Un laboratorio analiza el contenido de sodio (en mg/L) en muestras de agua de una planta de tratamiento. Se toma una muestra aleatoria de  $n = 12$  unidades, obteniendo los siguientes datos:

43, 39, 45, 48, 46, 50, 44, 49, 47, 46, 42, 41

Se desea construir un intervalo de confianza del 95% para la varianza poblacional  $\sigma^2$ , asumiendo que el contenido de sodio sigue una distribución Normal.

```
sodio = c(43, 39, 45, 48, 46, 50, 44, 49, 47, 46, 42, 41)
library(EnvStats)
sodio |>
  varTest(conf.level = 0.95)|>
  use_series(conf.int)
```

```
      LCL      UCL
5.565681 31.972759
attr(,"conf.level")
[1] 0.95
```

Con un nivel de confianza del 95%, podemos afirmar que la verdadera varianza del contenido de sodio está contenida entre 5.57 y 31.97  $(mg/L)^2$



```
n      <- sodio |> length()
s2     <- sodio |> var()
alpha  <- 0.05
gl     <- n - 1
chi2_inf <- qchisq(1 - alpha/2, df = gl)
chi2_sup <- qchisq(alpha/2, df = gl)
li_var  <- (gl * s2) / chi2_inf
ls_var  <- (gl * s2) / chi2_sup
c(li_var, ls_var)
```

```
[1] 5.565681 31.972759
```

```
c(li_var, ls_var) |> sqrt() # IC para desviación estándar
```

```
[1] 2.359170 5.654446
```

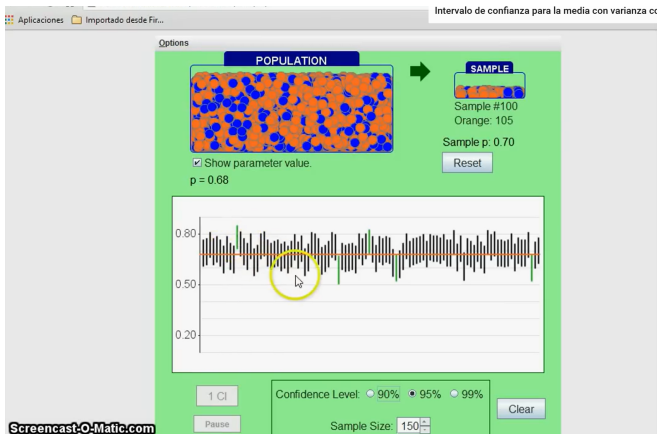
## Estimación de la proporción por intervalo de confianza

Si  $X_1, \dots, X_n$  es una muestra aleatoria donde cada  $X_i$  indica la presencia (1) o ausencia (0) de una característica,  $p$  es la proporción muestral de elementos con dicha característica,  $n > 30$ ,  $np > 0.05$  y  $n(1 - p) > 0.05$ , entonces el intervalo con un nivel de confianza del  $(1 - \alpha) \times 100\%$  para la proporción  $\pi$  se obtiene mediante

$$\underbrace{p - Z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{n}}}_a \leq \pi \leq \underbrace{p + Z_{(1-\alpha/2)} \sqrt{\frac{p(1-p)}{n}}}_b$$

donde  $a$  y  $b$  son valores numéricos que representan el Límite inferior y Límite superior del intervalo de confianza.

Ver el siguiente video sobre intervalos de confianza (click aquí)



## **Ejemplo 4**

Un equipo de investigadores está evaluando la presencia de plagas en árboles de una plantación forestal. Para ello, se selecciona aleatoriamente una muestra de 150 árboles, y se observa que 36 de ellos presentan signos visibles de infestación.

Se desea construir un intervalo de confianza del 95% para estimar la proporción poblacional de árboles infestados en toda la plantación.

```
library(binom)
binom.confint(x = 36, n = 150, methods = "asymptotic")
```

	method	x	n	mean	lower	upper
1	asymptotic	36	150	0.24	0.1716537	0.3083463

```
n      <- 150
x      <- 36
p      <- x/n
alpha  <- 0.05
zcrit  <- qnorm(1-alpha/2)
li     <- p - zcrit*sqrt(p*(1-p)/n)
ls     <- p + zcrit*sqrt(p*(1-p)/n)
c(li,ls)
```

```
[1] 0.1716537 0.3083463
```

Con un nivel de confianza del 95%, podemos afirmar que la verdadera proporción de árboles infestados está contenida entre 0.1717 y 0.3083.

Cuando la proporción es cercana a 0 o 1, y/o la muestra es pequeña, se sugiere utilizar la aproximación de Wilson, que brinda un intervalo asimétrico e incluido siempre en el intervalo  $[0, 1]$ .

$$\text{Límite inferior} = \frac{p + \frac{z^2}{2n} - z \cdot \sqrt{\frac{p(1-p)}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$$

$$\text{Límite superior} = \frac{p + \frac{z^2}{2n} + z \cdot \sqrt{\frac{p(1-p)}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$$

## Ejemplo 5

Un equipo de biólogos está monitoreando la presencia de un insecto invasor en zonas protegidas. Se colocan trampas en 44 ubicaciones diferentes, y solo 2 trampas detectan presencia del insecto. Se desea estimar, con un intervalo de confianza del 95%, la proporción de zonas afectadas por el insecto.

```
prop.test(x = 2, n = 44, correct = FALSE)$conf.int
```

```
[1] 0.01255511 0.15134998  
attr("conf.level")  
[1] 0.95
```

```
binom.confint(x = 2, n = 44, methods = "wilson")
```

	method	x	n	mean	lower	upper
1	wilson	2	44	0.04545455	0.01255511	0.15135

Con un nivel de confianza del 95%, podemos afirmar que la verdadera proporción de zonas afectadas por el insecto está en el intervalo  $[0.0126, 0.1514]$ .

```
n      <- 44
x      <- 2
p      <- x/n
alpha  <- 0.05
zcrit  <- qnorm(1-alpha/2)
num_li <- (p+zcrit**2/(2*n)-zcrit*sqrt(p*(1-p)/n+zcrit**2/(4*n**2)))
num_ls <- (p+zcrit**2/(2*n)+zcrit*sqrt(p*(1-p)/n+zcrit**2/(4*n**2)))
li_wilson <- num_li/(1+zcrit**2/n)
ls_wilson <- num_ls/(1+zcrit**2/n)
c(li_wilson,ls_wilson)
```

```
[1] 0.01255511 0.15134998
```



Cuando la proporción es cercana (o igual) a 0 o 1, y/o el tamaño muestral es muy pequeño, se recomienda utilizar el método exacto. Este método proporciona un intervalo de confianza asimétrico, siempre contenido en el rango  $[0, 1]$  y que garantiza el nivel de confianza especificado. En el ejemplo que venimos desarrollando:

```
binom.confint(x, n, methods = "exact")
```

	method	x	n	mean	lower	upper
1	exact	2	44	0.04545455	0.005552952	0.1547316

```
binom.test(x = x, n = n) |> use_series(conf.int)
```

```
[1] 0.005552952 0.154731578  
attr("conf.level")  
[1] 0.95
```

```
li_exacto <- qbeta(alpha/2, x, n - x + 1)
ls_exacto <- qbeta(1 - alpha/2, x + 1, n - x)
c(li_exacto, ls_exacto)
```

```
[1] 0.005552952 0.154731578
```

Con un nivel de confianza del 95%, podemos afirmar que la verdadera proporción de zonas afectadas por el insecto está en el intervalo  $[0.0056, 0.1547]$ .

## **Ejemplo 6**

Durante una campaña de vacunación contra la fiebre amarilla en una zona rural, se hace un seguimiento a 18 personas vacunadas. De ellas, solo 1 reporta un efecto adverso leve (como fiebre o dolor muscular) en las 48 horas posteriores.

Se desea construir un intervalo de confianza del 95% para la proporción poblacional de personas que podrían presentar un efecto adverso leve tras recibir la vacuna.

```
x = 1  
n = 18
```

```
binom.confint(x, n, methods = "asymptotic") # Aproximación Normal
```

	method	x	n	mean	lower	upper
1	asymptotic	1	18	0.05555556	-0.05026348	0.1613746

```
binom.confint(x, n, methods = "wilson") # Aproximación de Wilson
```

	method	x	n	mean	lower	upper
1	wilson	1	18	0.05555556	0.009875191	0.257573

```
binom.confint(x, n, methods = "exact") # Método exacto
```

	method	x	n	mean	lower	upper
1	exact	1	18	0.05555556	0.001405556	0.2729436

Interpretar el intervalo adecuado para el caso.