

# Práctica 8. Introducción a la estadística inferencial. Construcción de intervalos de confianza alrededor de una media.

## Contenidos

1. Introducción . . . . .	1
2. Intervalos de confianza alrededor de una media . . . . .	2
2.1 Construyendo la función para la obtención del intervalo . . . . .	2
2.2 Utilizando R base . . . . .	3
2.3 Utilización de otros paquetes . . . . .	7
3. Representación gráfica de los intervalos de confianza (IC) alrededor de una media. . . . .	8

## 1. Introducción

En esta práctica, aprenderemos a construir intervalos de confianza en R, aplicados a la media muestral, utilizando diferentes métodos según las características de nuestros datos. En el caso de la media, abordaremos cómo construir intervalos de confianza, considerando la evaluación previa de la normalidad de la variable. Estos conceptos son esenciales para realizar inferencias estadísticas y comprender la precisión de nuestras estimaciones en estudios de muestras.

Recordemos que un intervalo de confianza es un rango de valores que refleja la variabilidad entre una estimación obtenida en un estudio y el valor real del parámetro poblacional que se desea estimar. Este intervalo se calcula bajo el supuesto de que los datos siguen una distribución determinada, y en el caso de un intervalo de confianza del 95 %, indica que si se repitieran muchos estudios similares, el 95 % de los intervalos construidos contendrían el verdadero valor del parámetro, siempre que se cumplan todas las suposiciones necesarias para el cálculo. Es importante destacar que un intervalo de confianza específico no tiene una probabilidad asociada de contener el verdadero valor del parámetro; este valor está o no dentro del intervalo con certeza (100 % o 0 %). El “95 %” se refiere a la frecuencia con la que los intervalos generados por múltiples estudios incluirían el valor verdadero si los supuestos fueran correctos. Para una

interpretación probabilística del intervalo, se requiere un enfoque bayesiano, que genera un intervalo creíble basado en una distribución previa del parámetro.

## 2. Intervalos de confianza alrededor de una media

El intervalo de confianza alrededor de una media es un rango de valores que, con un nivel de confianza determinado (generalmente 95% o 99%), estima el valor real de la media poblacional. En otras palabras, indica el rango dentro del cual se espera que caiga la media verdadera de una población con cierta certeza, basándose en los datos de una muestra. Para calcularlo, se utiliza la media muestral y se ajusta con un margen de error que depende de la desviación estándar y el tamaño de la muestra. Este margen de error se multiplica por un factor correspondiente al nivel de confianza (extraído de la distribución t de Student o la distribución normal, dependiendo del tamaño de la muestra y si la desviación estándar poblacional es conocida). Su fórmula es, como ya sabemos

$$IC = \bar{x} \pm Z_{\alpha/2} \left( \frac{\sigma}{\sqrt{n}} \right)$$

Vamos a calcular los intervalos de confianza alrededor de una media de diversas formas

### 2.1 Construyendo la función para la obtención del intervalo

En primer lugar debemos establecer el directorio de trabajo y obtener la base de datos que vamos a utilizar

```
#setwd("~/Práctica 8")  
getwd
```

```
function ()  
  .Internal(getwd())  
<bytecode: 0x0000025e652bc308>  
<environment: namespace:base>
```

```
df_iam2 <- read.csv("df_iam2.csv")  
head (df_iam2)
```

	edad	sex	alt	peso	imc	hta	fum	colesterol	clas_soc	iam	imc_r
1	40	Mujer	157.2	54.2	21.93281	Sí	Sí	192	Alta	Sí	Normal
2	40	Varón	172.5	70.5	23.69250	No	Sí	176	Alta	No	Normal
3	84	Mujer	167.4	63.4	22.62447	Sí	No	226	Baja	No	Normal
4	87	Varón	173.3	96.3	32.06485	Sí	No	188	Baja	No	Obesidad
5	83	Varón	172.7	62.7	21.02243	No	No	178	Alta	Sí	Normal
6	60	Mujer	154.3	51.3	21.54694	Sí	No	157	Alta	No	Normal

En primer lugar vamos a construir la función para la obtención de IC de la media, y luego la aplicaremos a la variable `edad`

```
# Función para calcular intervalo de confianza
mean.ci <- function(x, alpha) {
  MU <- mean(x)
  SD <- sd(x)
  n <- length(x)
  SE <- SD / sqrt(n)
  ci <- MU + c(0, qt(alpha / 2, n - 1) * SE, qt(1 - alpha / 2, n - 1) * SE)
  results <- data.frame(Mean = ci[1], lower = ci[2], upper = ci[3])
  print(results)
}

# Aplicación de la función a la variable "edad"
mean.ci(df_iam2$edad, 0.05) # Nivel de confianza 95%
```

	Mean	lower	upper
1	62.6485	61.55513	63.74188

## 2.2 Utilizando R base

Es posible calcular intervalos de confianza (IC) para la media utilizando las funciones disponibles en R base. Una de las funciones más comunes para este propósito es `t.test()`, que permite realizar pruebas t y, al mismo tiempo, proporciona los intervalos de confianza asociados para la media de un conjunto de datos.

```
t.test(df_iam2$edad)
```

One Sample t-test

```
data: df_iam2$edad
t = 112.45, df = 935, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 61.55513 63.74188
sample estimates:
mean of x
 62.6485
```

Observamos que, en primer lugar nos ofrece un p-value, resultado de contrastar la hipótesis nula de que el IC de la media contiene el valor nulo (  $p\text{-value} < 2.2 \times 10^{-16}$ ). Luego nos ofrece el IC del 95% y, finalmente, la media.

Se puede cambiar el nivel de confianza elegido

```
t.test(df_iam2$edad, conf.level = 0.99)
```

#### One Sample t-test

```
data: df_iam2$edad
t = 112.45, df = 935, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 61.21049 64.08652
sample estimates:
mean of x
 62.6485
```

Como se ve, al ampliar la confianza, el IC es más amplio también.

Podemos pedir que nos estime los IC para la variable **edad** dentro de cada grupo de la variable **sexo** usando la función **by**

```
by(df_iam2$edad, df_iam2$sex, t.test)
```

```
df_iam2$sex: Mujer
```

#### One Sample t-test

```
data: dd[x, ]
t = 81.185, df = 470, p-value < 2.2e-16
```

```

alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 61.27416 64.31395
sample estimates:
mean of x
 62.79406

```

```
-----
df_iam2$sex: Varón
```

```
One Sample t-test
```

```

data: dd[x, ]
t = 77.816, df = 464, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 60.92273 64.07942
sample estimates:
mean of x
 62.50108

```

O también podemos pedir que nos estime los IC para la variable `edad` dentro de cada grupo de la variable `sexo` usando la función `tapply`

```
tapply(df_iam2$edad,df_iam2$sex, t.test)
```

```
$Mujer
```

```
One Sample t-test
```

```

data: X[[i]]
t = 81.185, df = 470, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 61.27416 64.31395
sample estimates:
mean of x
 62.79406

```

```
$Varón
```

#### One Sample t-test

```
data: X[[i]]
t = 77.816, df = 464, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 60.92273 64.07942
sample estimates:
mean of x
 62.50108
```

Para cambiar la amplitud del IC usando la función `tapply` se debe proceder así

```
tapply(df_iam2$edad, df_iam2$sex, function(x) t.test(x, conf.level = 0.99))
```

\$Mujer

#### One Sample t-test

```
data: x
t = 81.185, df = 470, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 60.79360 64.79451
sample estimates:
mean of x
 62.79406
```

\$Varón

#### One Sample t-test

```
data: x
t = 77.816, df = 464, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 60.42365 64.57850
sample estimates:
mean of x
 62.50108
```

## 2.3 Utilización de otros paquetes

Los IC se pueden estimar utilizando funciones de otros paquetes. Vamos a instalar y a cargar el paquete `Rmisc` (`install.packages(Rmisc)` ; `library(Rmisc)` )

Una vez cargado e instalado el paquete, usaremos la función `CI` para que nos estime los IC para la variable `edad`

```
CI(df_iam2$edad)
```

	upper	mean	lower
	63.74188	62.64850	61.55513

Se puede variar la amplitud del intervalo de confianza introduciendo un argumento adicional

```
CI(df_iam2$edad, ci = 0.99)
```

	upper	mean	lower
	64.08652	62.64850	61.21049

Observa cómo varía la amplitud del intervalo

Y también podríamos estimar los IC para la variable `edad` dentro de cada grupo de la variable `sexo` , con la función `group.CI`

```
group.CI(edad ~ sex, data = df_iam2, ci = 0.95)
```

	sex	edad.upper	edad.mean	edad.lower
1	Mujer	64.31395	62.79406	61.27416
2	Varón	64.07942	62.50108	60.92273

En este último caso, la amplitud del IC se puede variar modificando el último argumento

```
group.CI(edad ~ sex, data = df_iam2, ci = 0.99)
```

	sex	edad.upper	edad.mean	edad.lower
1	Mujer	64.79451	62.79406	60.79360
2	Varón	64.57850	62.50108	60.42365

### 3. Representación gráfica de los intervalos de confianza (IC) alrededor de una media.

Vamos a utilizar la función `plotmeans` del paquete `gplots` para visualizar las medias de diferentes grupos junto con sus intervalos de confianza. La propia función estima las medias y los IC antes de representarlos.

Usaremos esta función para estimar las medias de `edad` en cada grupo de la variable `sex`. Recuerda instalar y cargar el paquete `gplots`(`install.packages(gplots)` ; `library(gplots)`)

Ahora vamos a volver a estimar las medias de `edad` en cada grupo de la variable `imc_r`

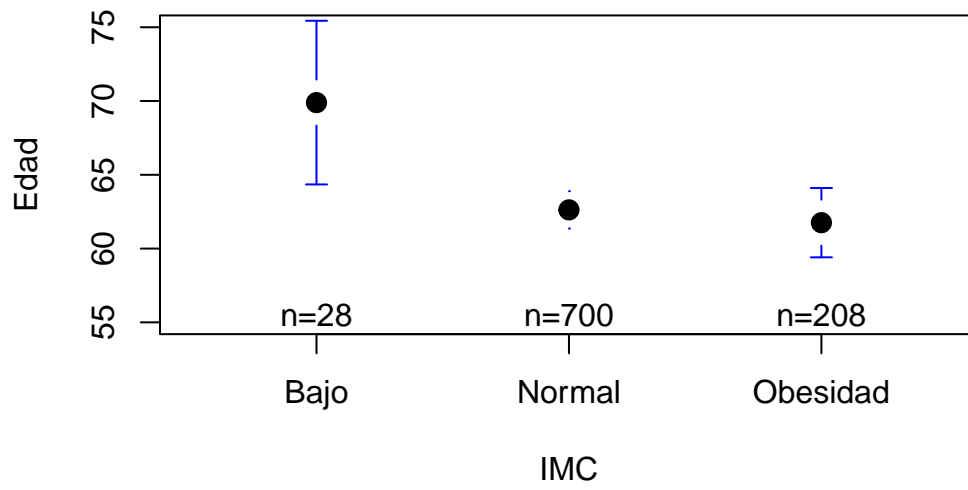
```
plotmeans(
  edad ~ imc_r,
  data = df_iam2,           # Datos
  ci.label = FALSE,        # No mostrar los IC
  n.label = TRUE,          # Mostrar n de la muestra
  connect = FALSE,         # No conectar las medias
  mean.label = FALSE,      # No mostrar valores de media
  barcol = "blue",         # Color de las barras
  xlab = "IMC",             # Etiqueta del eje x
  ylab = "Edad",           # Etiqueta del eje y
  main = "Media de Edad por grupo IMC", # Título
  ylim = c(55, 75),        # Ajustar los límites de y
  lwd = 5,                 # Grosor de las barras de IC
  pch = 20,                # (20= circulo pequeño)
  cex = 2                  # Tamaño del punto
)
```

```
Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :
zero-length arrow is of indeterminate angle and so skipped
```

```
Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :
zero-length arrow is of indeterminate angle and so skipped
```



### Media de Edad por grupo IMC



```
remove.packages(c("Rmisc", "gplots", "binom"))
```