MAE → Primera Entrega

Tema 1 - Ejercicio 2 (Bootstrap) ⇒ $X_1 - X_n \sim F(\mu, \sigma^2)$; $X_1^* - X_n^* \sim F_n$

a) $E_{F_n}[\bar{X}_n^*] := E[\bar{X}_n^* \mid X_1, \dots, X_n]$

Como vimos en clase: $E_{F_n}[X^*] = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$;

$$\bar{X}_n^* = \frac{X_1^* + \dots + X_n^*}{n};$$

Juntando estas observaciones llegamos a lo siguiente:

$$E_{F_n}[\bar{X}_n^*] = E[\bar{X}_n^* \mid X_1, \dots, X_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^*\right] = \frac{1}{\cancel{n}}\cancel{n}\, E[X_i^*] = \boxed{\bar{X}}$$

b) $E_F[\bar{X}_n^*] = \underbrace{E_F[E[\bar{X}_n^* \mid X_1, \dots, X_n]]}_{E[Y] = E[E[Y|X]]} = E_F[\bar{X}] = E_F\left[\frac{X_1 + \dots + X_n}{n}\right] =$

$$= \frac{1}{n} E_F[X_1 + \dots + X_n] = \boxed{\frac{1}{\cancel{n}}\cancel{n}\, E_F[X_1] = \mu}$$

c) $\boxed{Var_{F_n}[\bar{X}_n^*] = Var[\bar{X}_n^* \mid X_1, \dots, X_n]}$

Propiedad: $Var[X] = E[X^2] + (E[X])^2 \Rightarrow E[X^2] = \sum_{i=1}^{n} X_i^2 P(X = x_i) \Rightarrow$

$$\Rightarrow Var[X^*] = E_{F_n}[(X^*)^2] - (E_{F_n}[X^*])^2 = \sum_{i=1}^{n} X_i^2 P(X^* = x_i) - \bar{X}^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2 = S_n^2$$

$$Var_{F_n}[\bar{X}_n^*] = Var_{F_n}[\bar{X}_n^*] = Var_{F_n}\left[\frac{X_1^* + \dots + X_n^*}{n}\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var_{F_n}[X_i^*] =$$

$$= \frac{1}{n^2}\cancel{n}\, S_n^2 = \boxed{\frac{1}{n} S_n^2}$$

d) $Var_F[\bar{X}^*_n]$

Se Sabe: $Var[Y] = E[Var[Y|X]] + Var[E[Y|X]] \Rightarrow$

$$\Rightarrow Var_F[\bar{X}^*_n] = \underbrace{E[Var[\bar{X}^*_n|X_1,\dots X_n]]}_{2} + \underbrace{Var[E[\bar{X}^*_n|X_1,\dots,X_n]]}_{1} = \circledast \Rightarrow \text{Abajo}$$

1) $\rightarrow Var[E[\bar{X}^*_n|X_1,\dots,X_n]] = Var[\bar{X}] = Var\left[\frac{X_1+\dots+X_n}{n}\right] =$

$$\frac{1}{n^2}Var[X_1+\dots+X_n] = \frac{1}{n^2}nVar[X_1] = \frac{\sigma^2}{n} \Leftrightarrow Var[X_i]=\sigma^2$$

2) $\rightarrow E[Var[\bar{X}^*_n|X_1,\dots,X_n]] = E\left[\frac{S_n^2}{n}\right] = \frac{1}{n}E[S_n^2] = \frac{1}{n}E\left[\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n}\right] =$

$$= \frac{1}{n}E\left[\frac{\sum_{i=1}^n X_i^2}{n} + \frac{\sum_{i=1}^n \bar{X}^2}{n} - \frac{2\bar{X}\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n}E\left[\frac{\sum_{i=1}^n X_i^2}{n} + \bar{X}^2 - 2\bar{X}^2\right] =$$

$$= \frac{1}{n}E\left[\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2\right] = \frac{1}{n^2}\sum_{i=1}^n E[X_i^2] = \frac{1}{\cancel{n^2}\, n}E[X_i^2] - \frac{1}{n}E[\bar{X}^2] =$$

$$= \frac{1}{n}\left(Var[X_i] + (E[X_i])^2\right) - \frac{1}{n}\left(Var[\bar{X}] + (E[\bar{X}])^2\right) =$$

$$= \frac{1}{n}(\sigma^2+\mu^2) - \frac{1}{n}\left(\frac{\sigma^2}{n}+\mu^2\right) =$$

$$= \frac{\sigma^2}{n}\left(1-\frac{1}{n}\right)$$

Por lo tanto, Con ambos términos Resueltos

$$Var_F[\bar{X}^*_n] \Rightarrow \overset{\text{Arriba}}{\circledast} \Rightarrow \frac{\sigma^2}{n}\left(1-\frac{1}{n}\right) + \frac{\sigma^2}{n} = \boxed{\frac{\sigma^2}{n^2}(2n-1)}$$

MAE – Primera Entrega

Ejercicio 7:

Sea F una distribucion :

- media μ
- varianza σ^2
- coeficiente de asimetrıa $\gamma = EF\,[(X - \mu)^3]/\sigma^3$.

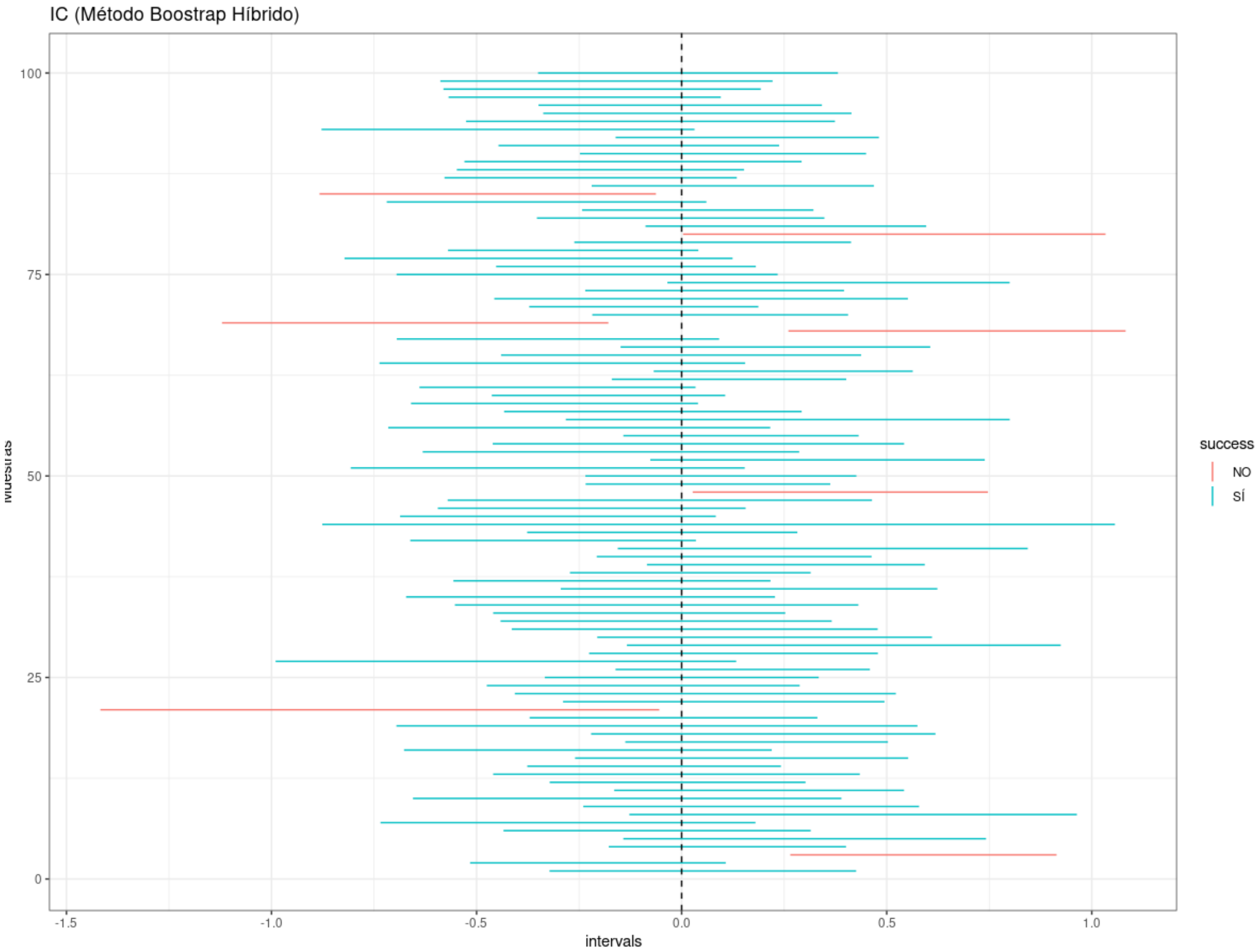
Genera $R = 1000$ muestras de observaciones iid $X1, \ldots, Xn$ con $Xi \equiv N(0, 1)$ para $n = 100$. Para cada una de ellas, calcula tres intervalos de confianza bootstrap de nivel 95 % para γ usando:

- metodo hıbrido
- metodo normal
- metodo percentil

Determina el porcentaje de intervalos que contienen al par´ametro en cada caso. Repite el ejercicio con muestras procedentes de una distribuci´on exponencial de parametro $\lambda = 1$.

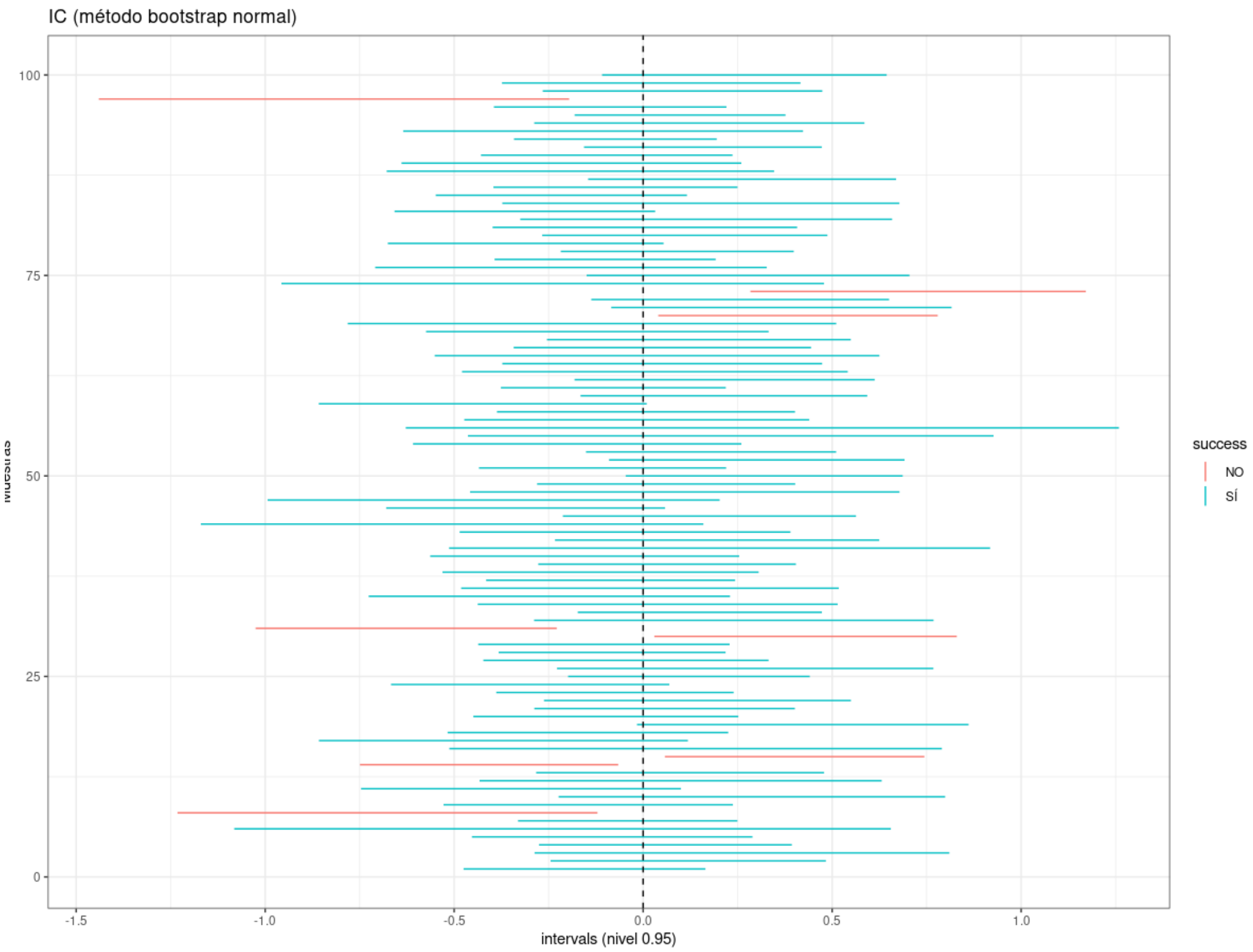
```
# Importación de Lobrerías Útiles
library(e1071)
library(ggplot2)

# Establecimiento de semilla para repetición del experimento
set.seed(123)

# Declaración de variables Globales
R <- 1000 #Remuestras
n <- 100 # 100 datos por muestras
m <- 100
a <- 0.05 # Alfa = Confianza de Boostrap = 95%

# Coeficientes de asimetría
# Coefinciente de asimetría en Normal = 0
coef_norm <- 0
# Coeficiente de asimetría en Exponencial = 2
coef_exp <- 2

# Variables Para guardar los resultados
success <- NULL
interval <- NULL

# Metodo 1, Boostrap Híbrido
for (i in 1:m){
  # Calculamos la Muestra Original Normal(0, 1)
  muestra_original <- rnorm(100,mean=0,sd=sqrt(1))
```

```
  # Calculo del coeficiente de la muestra
  coef <- skewness(muestra_original)
  # Remuestreo Bootstrap
  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  # Calculo Coeficiente Remuestreo Boostrap
  coef_boots <- apply(muestras_boots, 2, skewness)

  T_boots <- sqrt(n) * (coef_boots - coef)

  # Calculo de los quantiles de los intervals de Confianza Bootstrap
  ic_min <- coef - quantile(T_boots, 1-a/2)/sqrt(n)
  ic_max <- coef - quantile(T_boots, a/2)/sqrt(n)

  # Calculo del interval y de los acciertos
  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_norm & ic_max > coef_norm)
}

paste("% de acierto =",mean(success))

# Generamos el Gráfico
df <- data.frame(ic_min <- interval[,1],
                 ic_max <- interval[, 2],
                 ind = 1:m,
                 success = success)
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_norm), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals',
     title = ('IC (Método Boostrap Híbrido')
```



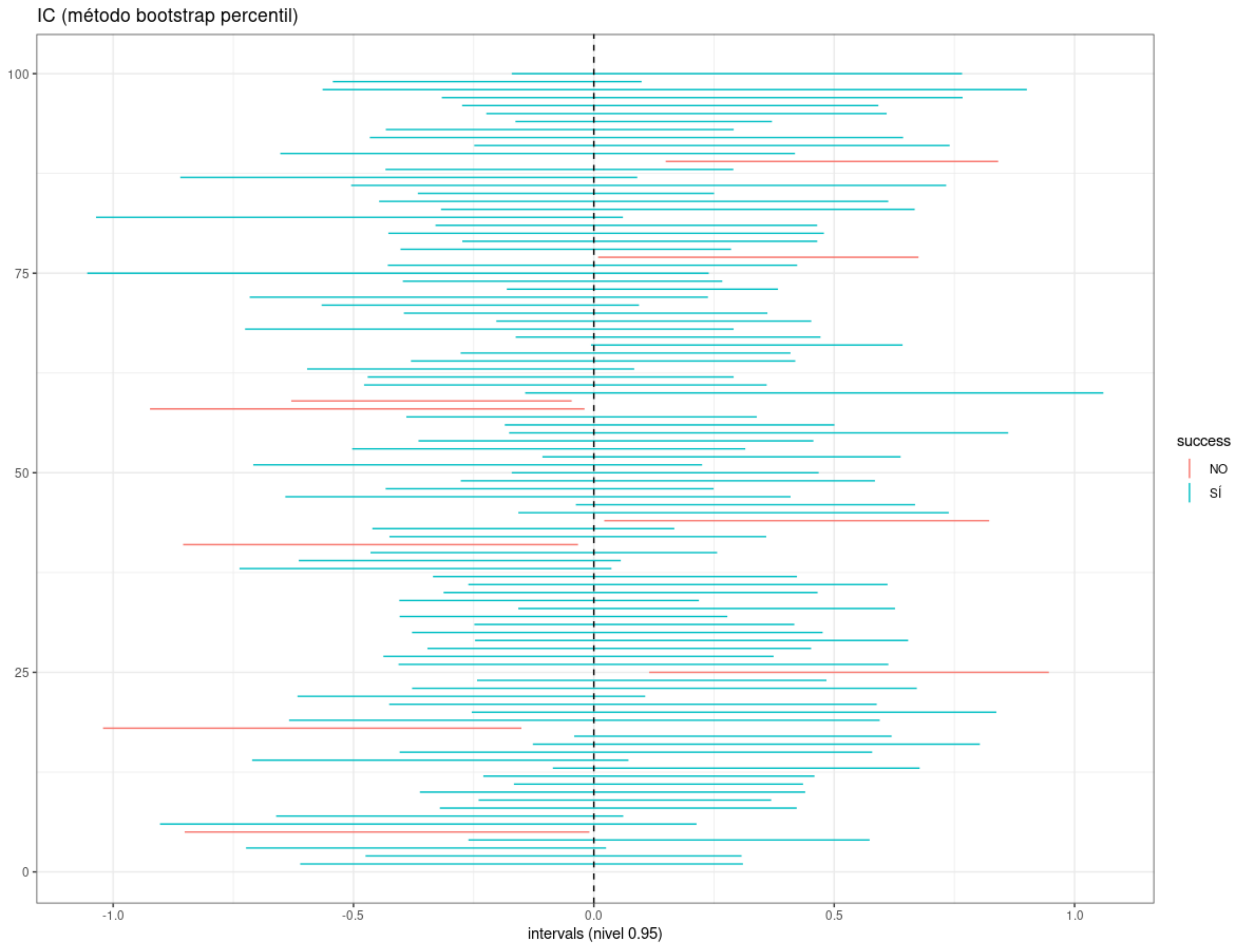

```
# Método 2, Normal
success <- NULL
interval <- NULL
for (i in 1:m){
  muestra_original <- rnorm(100,mean=0,sd=sqrt(1))
  coef <- skewness(muestra_original)

  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  coef_boots <- apply(muestras_boots, 2, skewness)
  et_boots<-sd(coef_boots)

  ic_min <- coef + qnorm(alfa/2)*et_boots
  ic_max <- coef - qnorm(alfa/2)*et_boots

  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_asim_normal & ic_max > coef_asim_normal)
}
paste("% de acierto =",mean(success))

df <- data.frame(ic_min <- interval[,1],
```

```
              ic_max <- interval[, 2],
              ind = 1:m,
              success = success)
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_asim_normal), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals (nivel 0.95)',
      title = 'IC (método bootstrap normal)')
```

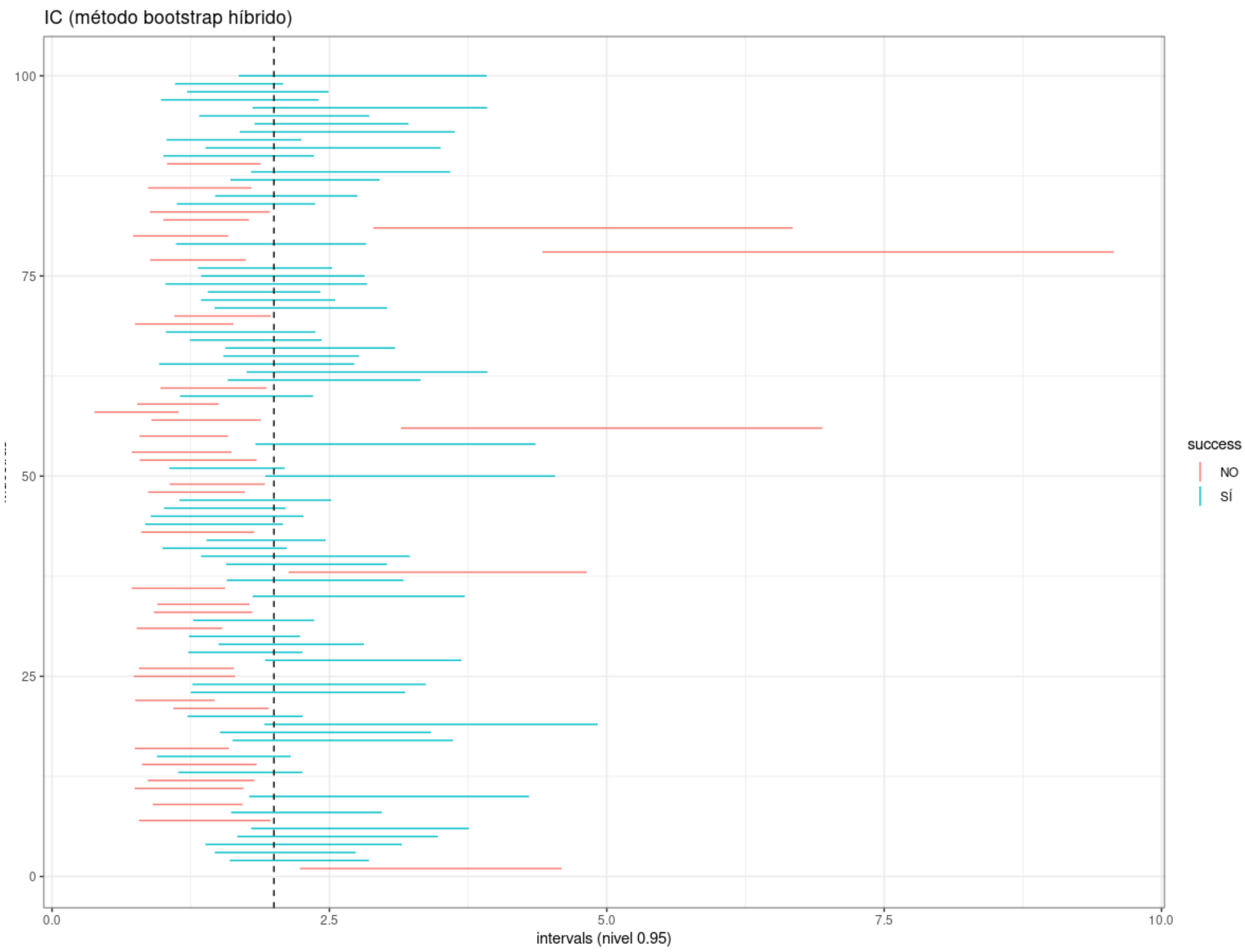



```
# Metodo 3, Percentiles
success <- NULL
interval <- NULL
for (i in 1:m){
  muestra_original <- rnorm(100,mean=0,sd=sqrt(1))
  coef <- skewness(muestra_original)
  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  coef_boots <- apply(muestras_boots, 2, skewness)
  ic_min <- quantile(coef_boots, alfa/2)
```

```
  ic_max <- quantile(coef_boots,1-alfa/2)
  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_asim_normal & ic_max > coef_asim_normal)
}
paste("% de acierto =",mean(success))

df <- data.frame(ic_min <- interval[,1],
                 ic_max <- interval[, 2],
                 ind = 1:m,
                 success = success)
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_asim_normal), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals (nivel 0.95)',
       title = 'IC (método bootstrap percentil)')
```

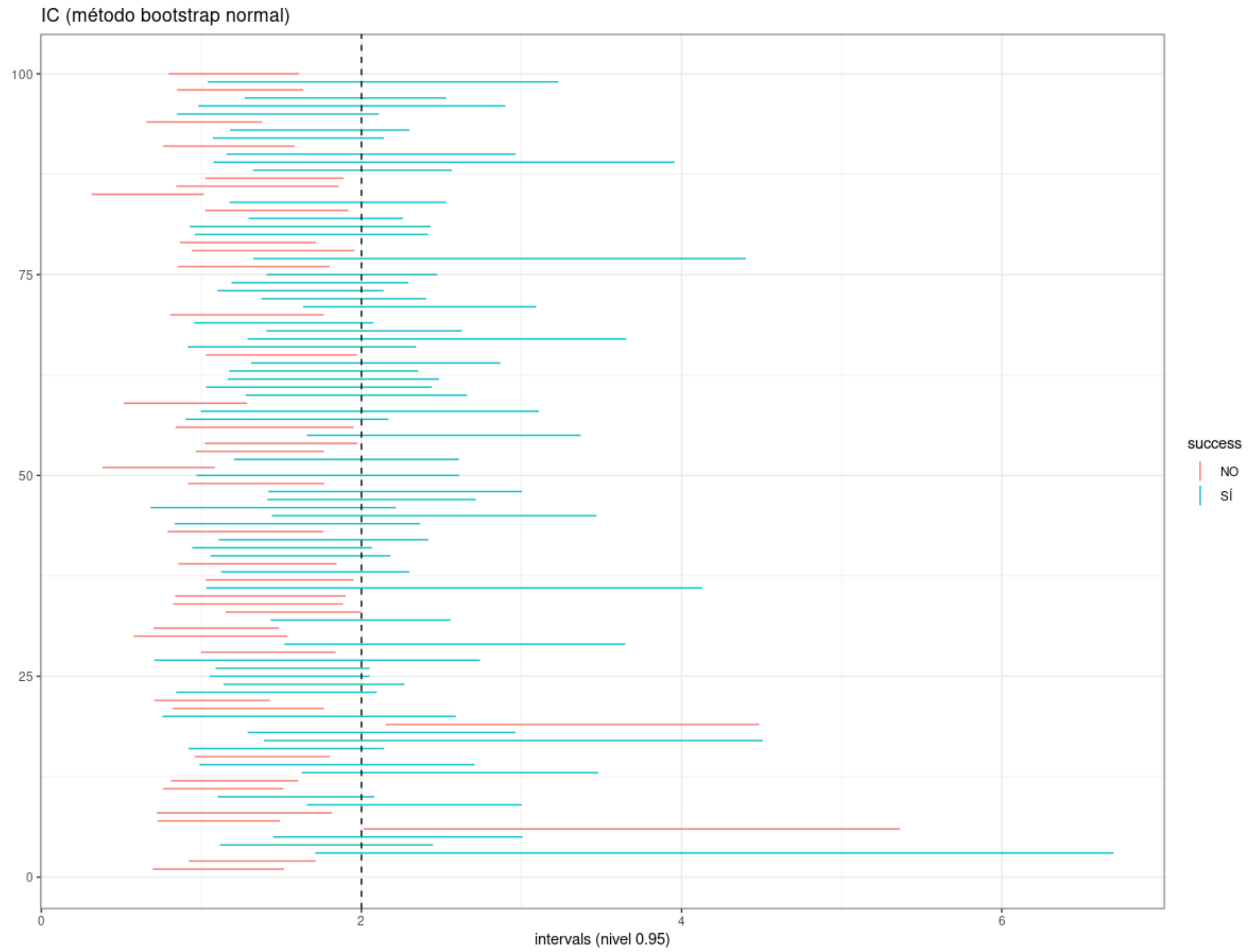


```
## Pasamos a los ejercicios de muestra exponencial los cuales se realizan exactamente igual.

# Metodo Hibrido
success <- NULL
interval <- NULL
for (i in 1:m){
  muestra_original <- rexp(100,1)
  coef <- skewness(muestra_original)
  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  coef_boots <- apply(muestras_boots, 2, skewness)
  T_boots <- sqrt(n) * (coef_boots - coef)
  ic_min <- coef - quantile(T_boots,1-alfa/2)/sqrt(n)
  ic_max <- coef - quantile(T_boots,alfa/2)/sqrt(n)
  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_exp & ic_max > coef_exp)
}
paste("% de acierto =",mean(success))

df <- data.frame(ic_min <- interval[,1],
                 ic_max <- interval[, 2],
                 ind = 1:m,
                 success = success)
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_exp), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals (nivel 0.95)',
       title = 'IC (método bootstrap híbrido)')
```




```
# Metodo Normal
success <- NULL
interval <- NULL
for (i in 1:m){
  muestra_original <- rexp(100,1)
  coef <- skewness(muestra_original)
  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  coef_boots <- apply(muestras_boots, 2, skewness)
  et_boots<-sd(coef_boots)
  ic_min <- coef + qnorm(alfa/2)*et_boots
  ic_max <- coef - qnorm(alfa/2)*et_boots
  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_exp & ic_max > coef_exp)
}
paste("% de acierto =",mean(success))

df <- data.frame(ic_min <- interval[,1],
                 ic_max <- interval[, 2],
                 ind = 1:m,
                 success = success)
```

```
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_exp), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals (nivel 0.95)',
       title = 'IC (método bootstrap normal)')
```




```
# Metodo percentil
success <- NULL
interval <- NULL
for (i in 1:m){
  muestra_original <- rexp(100,1)
  coef <- skewness(muestra_original)
  muestras_boots <- sample(muestra_original, n*R, rep = TRUE)
  muestras_boots <- matrix(muestras_boots, nrow = n)
  coef_boots <- apply(muestras_boots, 2, skewness)
  ic_min <- quantile(coef_boots, alfa/2)
  ic_max <- quantile(coef_boots,1-alfa/2)
  interval <- rbind(interval, c(ic_min, ic_max))
  success <- c(success, ic_min < coef_exp & ic_max > coef_exp)
```

```
}
paste("% de acierto =",mean(success))

df <- data.frame(ic_min <- interval[,1],
                 ic_max <- interval[, 2],
                 ind = 1:m,
                 success = success)
ggplot(df) +
  geom_linerange(aes(xmin = ic_min, xmax = ic_max, y = ind, col = success)) +
  scale_color_hue(labels = c("NO", "SÍ")) +
  geom_vline(aes(xintercept = coef_exp), linetype = 2) +
  theme_bw() +
  labs(y = 'Muestras', x = 'intervals (nivel 0.95)',
       title = 'IC (método bootstrap percentil)')
```