

# TP Inicial

Alan Erdei, Nicolás Ian Rozenberg, Mateo Suffern

2024-03-26

## Carga y preparación de los datos

```
datos_encuesta <- read.table("./ENNyS_menorA2.txt", header = TRUE)
head(datos_encuesta)
```

##	Sexo	Tipo_embarazo	Edad	Peso	Perim_encef	Talla
## 1	Varon	Simple	1.84109589	14.48	50.0	87.5
## 2	Mujer	Simple	1.49589041	11.88	47.5	76.7
## 3	Mujer	Simple	0.58082192	6.78	42.0	69.0
## 4	Mujer	Simple	0.07945205	4.18	37.8	49.9
## 5	Varon	Simple	1.64931507	11.68	48.1	83.7
## 6	Mujer	Simple	0.05479452	3.98	36.0	52.0

Cambiamos el tipo de las columnas Sexo y Tipo\_embarazo a factor (categórico)

```
datos_encuesta$Sexo <- as.factor(datos_encuesta$Sexo)
datos_encuesta$Tipo_embarazo <- as.factor(datos_encuesta$Tipo_embarazo)
attach(datos_encuesta)
```

---

## Ejercicio 1

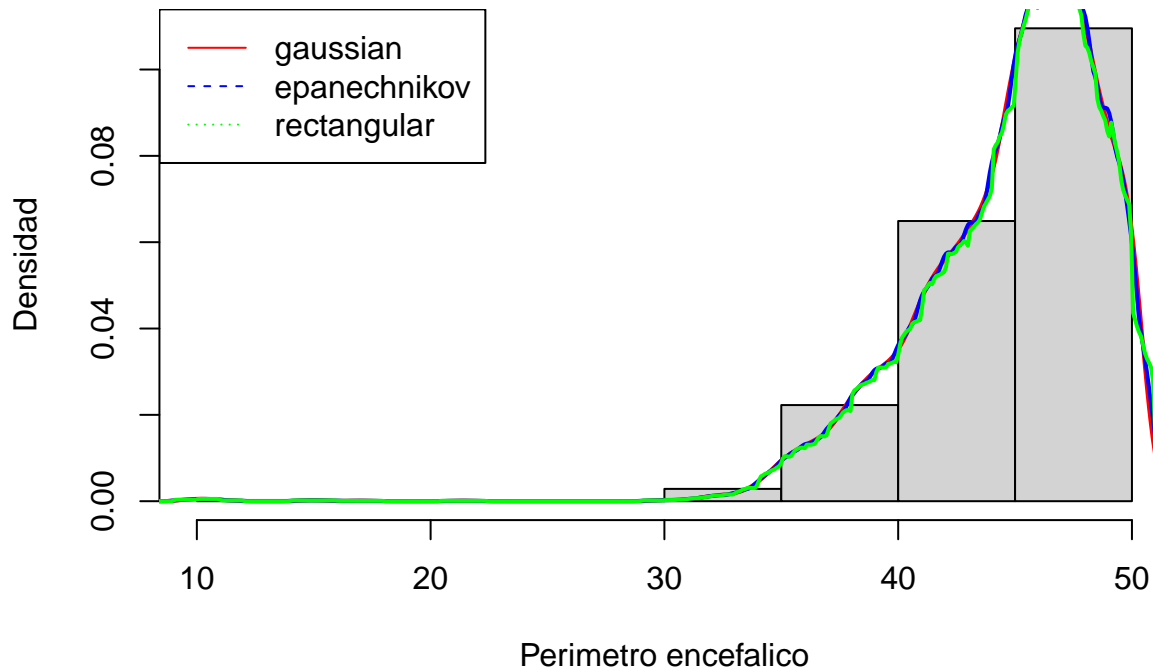
```
hist(
  Perim_encef,
  probability = TRUE,
  main="Histograma de perimetro encefalico comparado con densidades estimadas",
  xlab="Perimetro encefalico",
  ylab="Densidad"
)

kernels <- c("gaussian", "epanechnikov", "rectangular")
colors <- c("red", "blue", "green")
kde_perim_encef <- list() # A ser utilizado en siguientes ejercicios

for (i in seq_along(kernels)) {
  kde <- density(
    Perim_encef,
    kernel = kernels[i]
  )
  lines(kde, col = colors[i], lw = "2")
  kde_perim_encef[[kernels[i]]] <- kde
}
```

```
legend(x="topleft", legend = kernels, col = colors, lty = 1:3)
```

## Histograma de perímetro encefálico comparado con densidades estimadas



Se puede observar que la densidad estimada por cada uno de los núcleos proporcionados ajustan de forma muy similar y que se asimilan al histograma.

## Ejercicio 2

Primero, verificamos que no hayan registros de bebés de edad mayor a 2 años

```
sum(Edad > 2)
```

```
## [1] 0
```

No los hay. Estimamos la probabilidad de que el perímetro encefálico se encuentre en el rango de 42 cm a 48 cm utilizando tanto los datos arrojados por el histograma, como aproximando la integral de la densidad estimada con el kernel Epanechnikov

```
lower_bound <- 42
```

```
upper_bound <- 48
```

```
indices_in_range <- (kde_perim_encef$epanechnikov$x >= lower_bound &
```

```
  kde_perim_encef$epanechnikov$x <= upper_bound)
```

```
dx <- diff(kde_perim_encef$epanechnikov$x[1:2])
```

```
density_in_range <- kde_perim_encef$epanechnikov$y[indices_in_range]
```

```
prob_epa <- sum(density_in_range) * dx
```

```
cat("Probabilidad utilizando densidad estimada:", prob_epa)
```

```
## Probabilidad utilizando densidad estimada: 0.5709297
```

### Ejercicio 3

Primero, observamos las ventanas

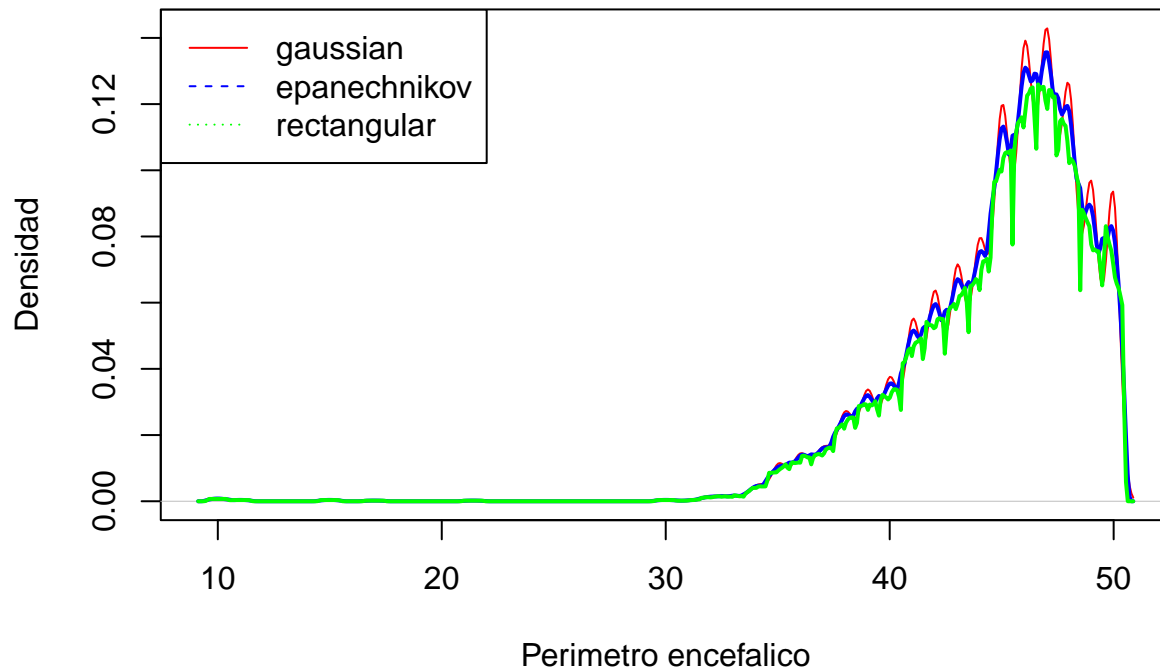
```
for (ker in kernels) {  
  dx <- diff(kde_perim_encef[[ker]]$x[1:2])  
  cat("Ventana para nucleo", ker, ":", dx, "\n")  
}
```

```
## Ventana para nucleo gaussian : 0.08513681  
## Ventana para nucleo epanechnikov : 0.08513681  
## Ventana para nucleo rectangular : 0.08513681
```

Vemos que ha sido seteada como la misma para los tres métodos. Graficamos las densidades estimadas utilizando tanto el doble de ventana, como la mitad.

```
dx <- diff(kde_perim_encef[["gaussian"]]  
$x[1:2])  
for (i in seq_along(kernels)) {  
  kde <- density(  
    Perim_encef,  
    kernel = kernels[i],  
    adjust = 1/2  
  )  
  
  if (i == 1){  
    plot(  
      kde,  
      col = colors[i],  
      main = sprintf("Densidad estimada con ventana %f", dx / 2),  
      xlab="Perimetro encefalico",  
      ylab="Densidad"  
    )  
  }  
  else{  
    lines(  
      kde,  
      col = colors[i],  
      lw = "2"  
    )  
  }  
}  
  
legend(x="topleft", legend = kernels, col = colors, lty = 1:3)
```

## Densidad estimada con ventana 0.042568

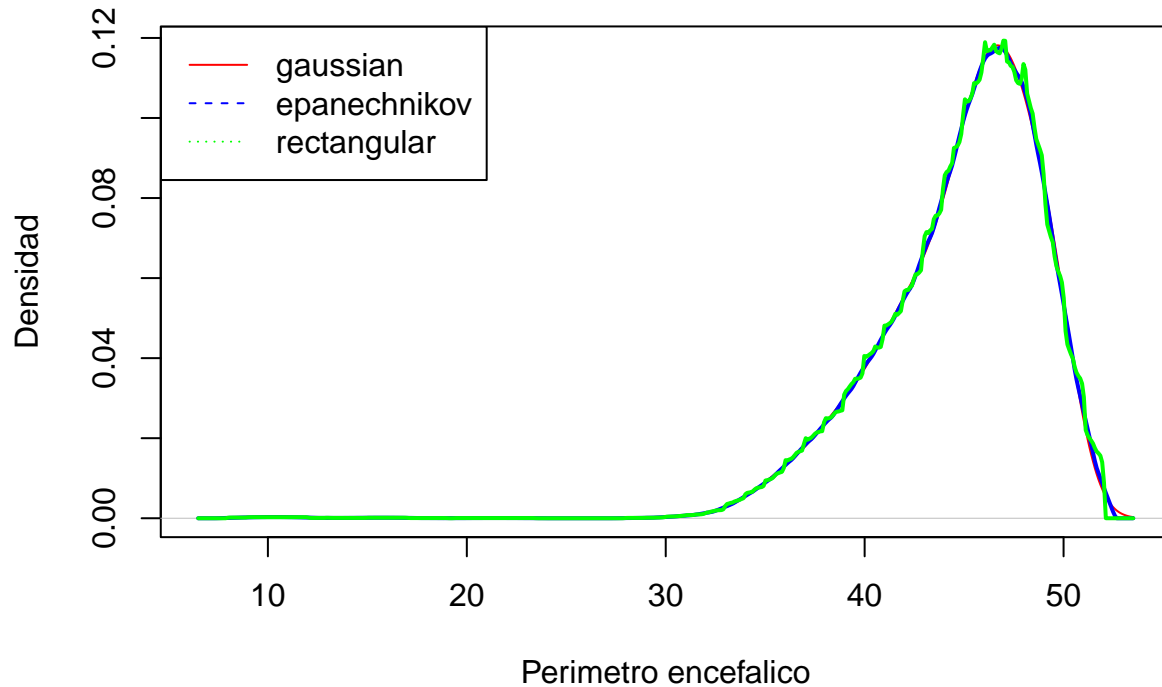


```
for (i in seq_along(kernels)) {
  kde <- density(
    Perim_encef,
    kernel = kernels[i],
    adjust=2
  )

  if (i == 1){
    plot(
      kde,
      col = colors[i],
      main = sprintf("Densidad estimada con ventana %f", dx * 2),
      xlab="Perimetro encefalico",
      ylab="Densidad"
    )
  }
  else{
    lines(
      kde,
      col = colors[i],
      lw = "2"
    )
  }
}

legend(x="topleft", legend = kernels, col = colors, lty = 1:3)
```

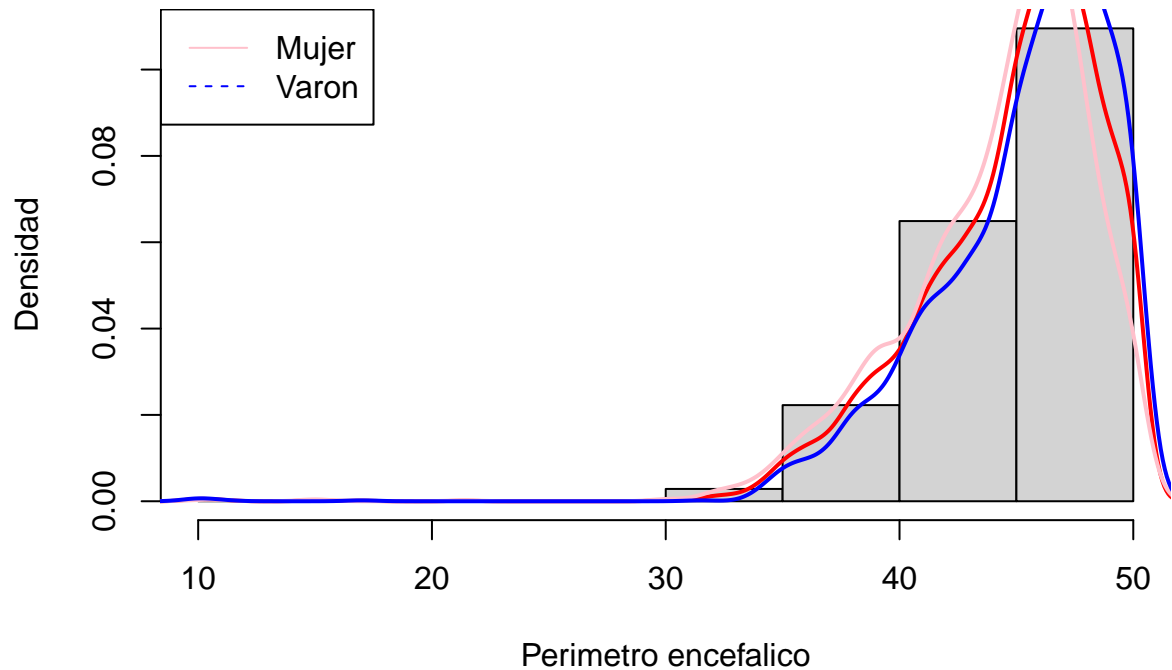
### Densidad estimada con ventana 0.170274



#### Ejercicio 4

```
hist(Perim_encef,
     probability = TRUE,
     main = "Histograma comparado con densidades de acuerdo al sexo",
     xlab="Perimetro encefalico",
     ylab="Densidad"
)
lines(
  kde_perim_encef[["gaussian"]],
  col = "red",
  lw = "2"
)
sexos <- c("Mujer", "Varon")
colors <- c("pink", "blue")
for (i in seq_along(sexos)){
  kde <- density(
    Perim_encef[Sexo == sexos[i]],
    kernel = "gaussian",
  )
  lines(
    kde,
    col = colors[i],
    lw = "2"
  )
}
legend(x="topleft", legend = sexos, col = colors, lty = 1:3)
```

## Histograma comparado con densidades de acuerdo al sexo



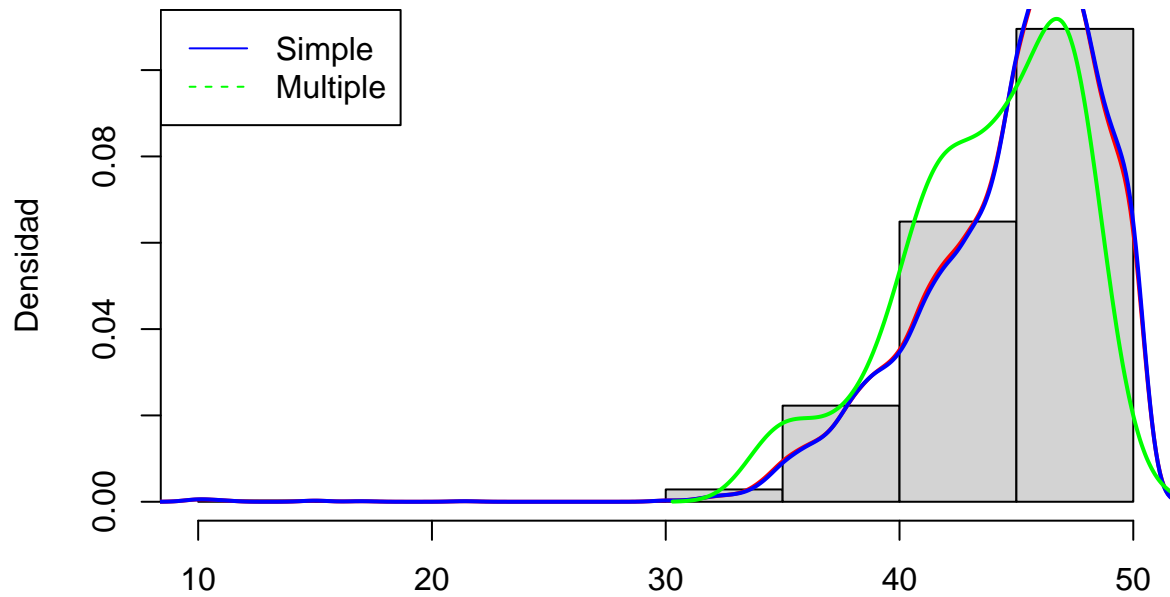
Se puede observar que el perímetro encefálico tiende a ser de menor longitud que el de los varones, puesto que la curva de densidad estimada para las mujeres es más alta para valores más pequeños.

### Ejercicio 5

```
hist(  
  Perim_encef,  
  probability = TRUE,  
  main = "Histograma comparado con densidades",  
  xlab="Perimetro encefalico",  
  ylab="Densidad"  
)  
lines(  
  kde_perim_encef[["gaussian"]],  
  col = "red",  
  lw = "2"  
)  
tipos_embarazo <- c("Simple", "Multiple")  
colors <- c("blue", "green")  
for (i in seq_along(tipos_embarazo)){  
  kde <- density(  
    Perim_encef[Tipo_embarazo == tipos_embarazo[i]],  
    kernel = "gaussian",  
  )  
  lines(  
    kde,  
    col = colors[i],  
    lw = "2",  
  )  
}
```

```
legend(x="topleft", tipos_embarazo, col = colors, lty = 1:3)
```

## Histograma comparado con densidades



### Perimetro encefalico

La densidad estimada del perímetro encefálico para los casos de nacimiento múltiple difiere considerablemente de la densidad general estimada, y la de los casos de nacimiento simple. Calculamos la frecuencia relativa de cada tipo de embarazo.

```
table(Tipo_embarazo) / length(Tipo_embarazo)
```

```
## Tipo_embarazo
##   Multiple      Simple
## 0.03347428 0.96652572
```

Vemos que aproximadamente el 3% de los registros provienen de embarazos múltiples, por lo que consideramos que no podemos sacar conclusiones acerca de esta diferencia.