

# TP Inicial

Alan Erdei, Nicolás Ian Rozenberg, Mateo Suffern

2024-03-28

```
# Librería utilizada para realizar bagplot
library(aplpack)
library(ggplot2)
```

## 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

```
perim_encef_hist <- hist(
  Perim_encef,
  probability = TRUE,
  main="Histograma de perimetro encefalico comparado con densidades estimadas",
  xlab="Perimetro encefalico (cm)",
  ylab="Densidad",
  ylim=c(0, 0.15)
)
```

```

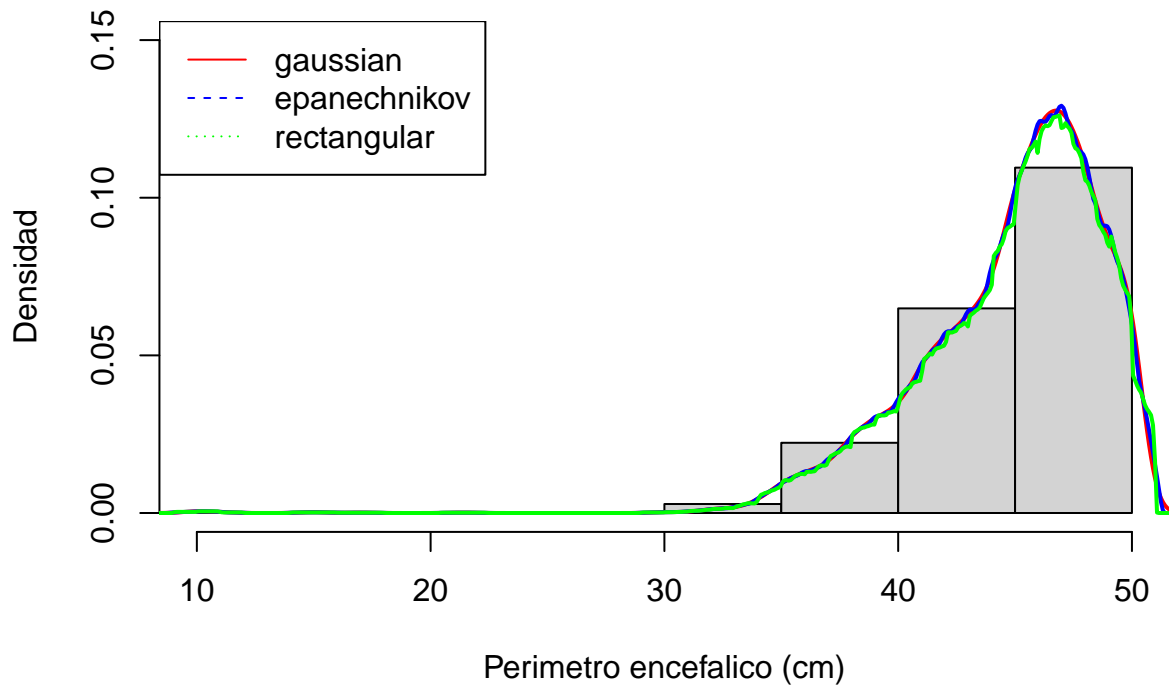
kernels <- c("gaussian", "epanechnikov", "rectangular")
colors <- c("red", "blue", "green")
kde_perim_encef <- list() # A ser utilizado en siguientes ejercicios

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 primero aproximando la integral de la densidad estimada con el kernel Epanechnikov

```
lower_bound <- 42
upper_bound <- 48

bw <- kde_perim_encef$epanechnikov$bw

prob_estim_epa <- function(datos, h, x){
  n <- length(datos)
  u <- (x - datos)/h
  return((3/(4*n)) * sum(
    (u - (u^3)/3 + (2/3)) * (abs(u) <= 1)
    + (4/3) * (u > 1)
  ))
}

prob_res <- prob_estim_epa(
  Perim_encef,
  bw,
  upper_bound
) - prob_estim_epa(
  Perim_encef,
  bw,
  lower_bound
)
cat("Probabilidad integrando densidad estimada:", prob_res)
```

```
## Probabilidad integrando densidad estimada: 0.5795061
```

Ahora utilizando tanto los datos arrojados por el histograma. Primero, observamos que los límites se encuentran en los últimos dos bins.

```
perim_encef_hist$breaks

## [1] 10 15 20 25 30 35 40 45 50

nbins <- length(perim_encef_hist$density)
prob_estim_hist <- (
  perim_encef_hist$density[nbins-1] * (45-42)
  + perim_encef_hist$density[nbins] * (48-45)
)
cat("Probabilidad obtenida mediante histograma:", prob_estim_hist)
```

```
## Probabilidad obtenida mediante histograma: 0.5233615
```

### Ejercicio 3

Graficamos las densidades estimadas utilizando tanto el doble de ventana, como la mitad.

```

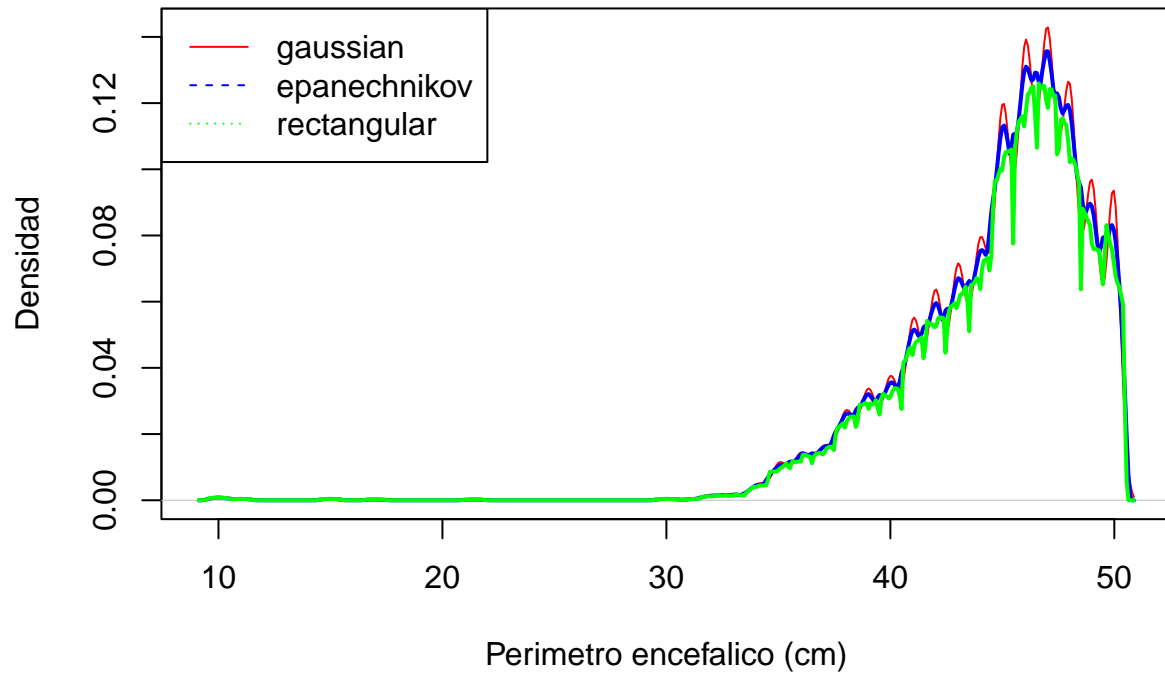
for (i in seq_along(kernels)) {
  kde <- density(
    Perim_encef,
    kernel = kernels[i],
    adjust = 1/2
  )

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

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

```

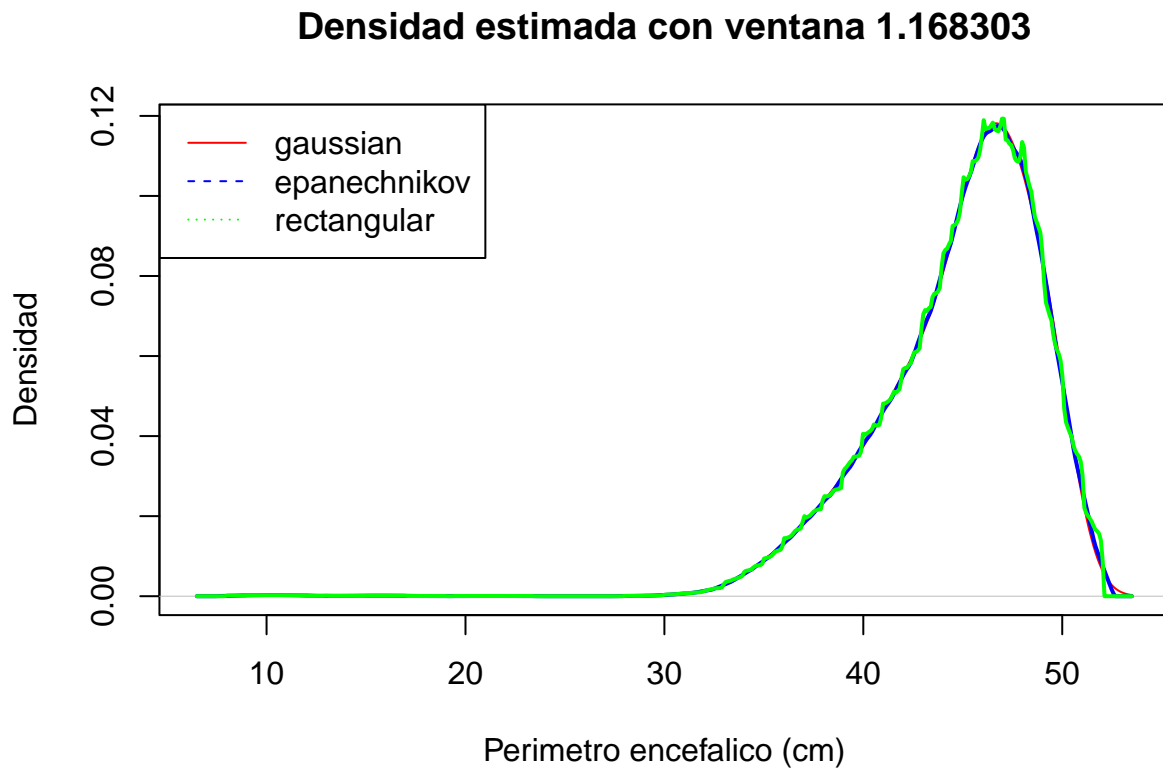
## Densidad estimada con ventana 0.292076



```
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", kde$bw),
      xlab="Perimetro encefalico (cm)",
      ylab="Densidad"
    )
  }
  else{
    lines(
      kde,
      col = colors[i],
      lw = "2"
    )
  }
}

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



#### Ejercicio 4

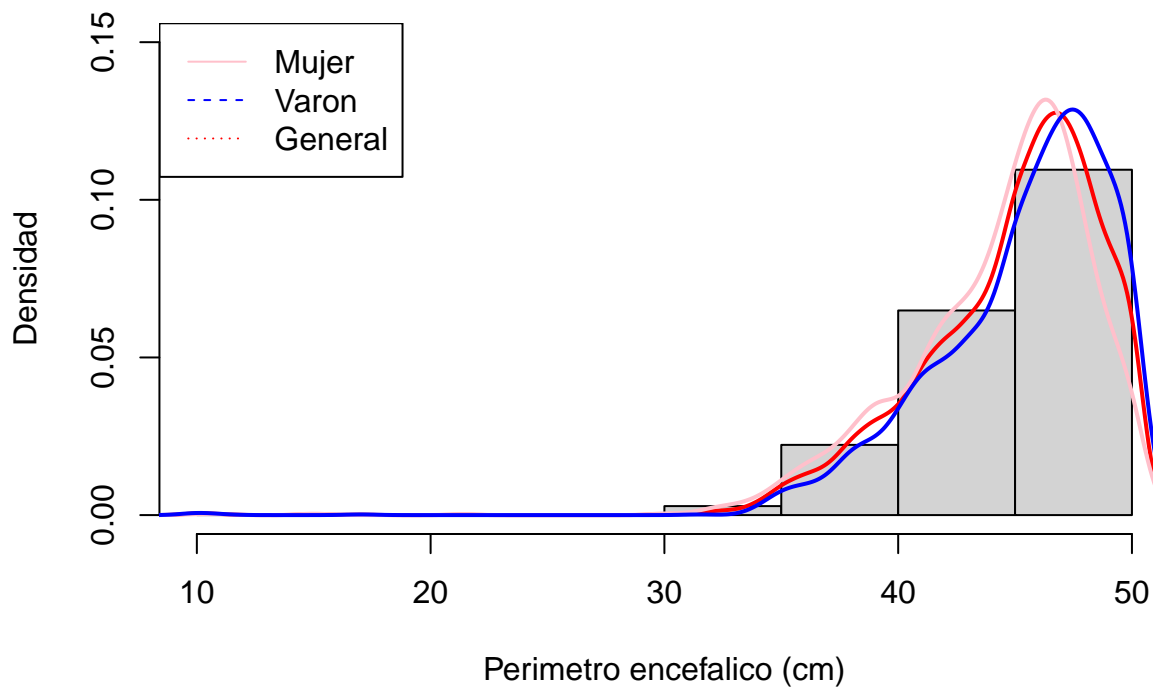
```
hist(Perim_encef,
     probability = TRUE,
     main = "Histograma comparado con densidades de acuerdo al sexo",
     xlab="Perimetro encefalico (cm)",
     ylab="Densidad",
     ylim=c(0, 0.15)
)
lines(
  kde_perim_encef[["gaussian"]],
  col = "red",
  lw = "2"
)
sexos <- c("Mujer", "Varon")
colors <- 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 = c(sexos, "General"), col = c(colors, "red"), 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 (cm)",
  ylab = "Densidad",
  ylim = c(0, 0.15)
)
lines(
  kde_perim_encef[["gaussian"]],
  col = "red",
  lw = "2"
)

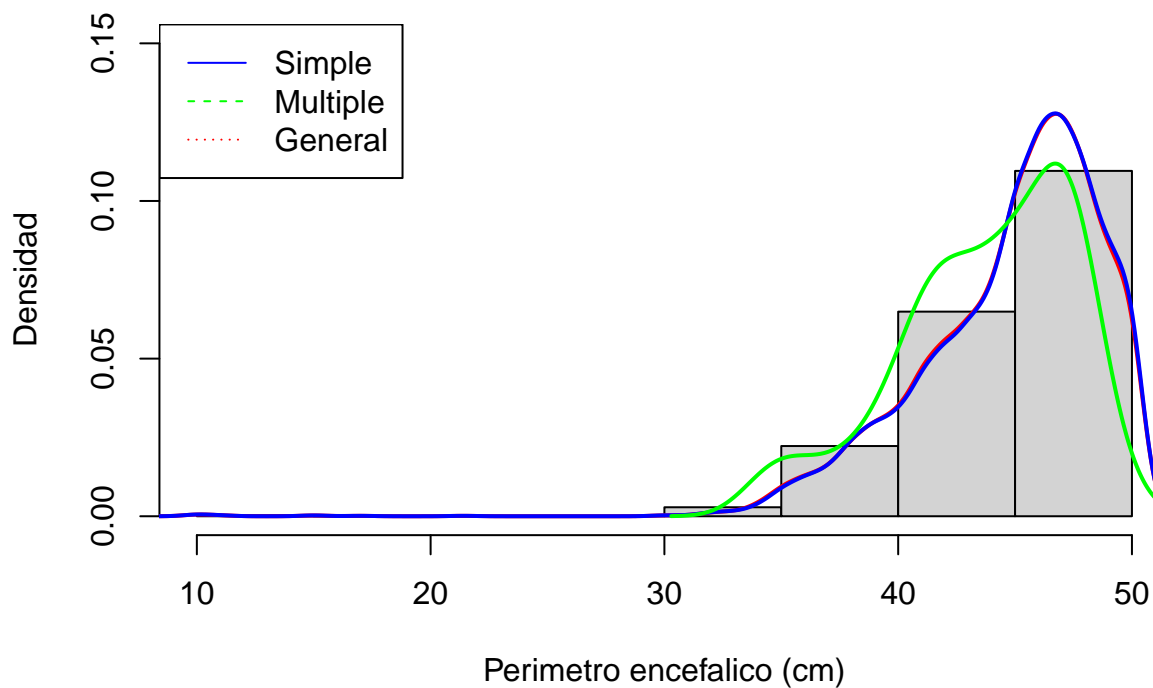
```

```

)
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", c(tipos_embarazo, "General"), col = c(colors, c("red")),lty = 1:3)

```

## Histograma comparado con densidades



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)
```

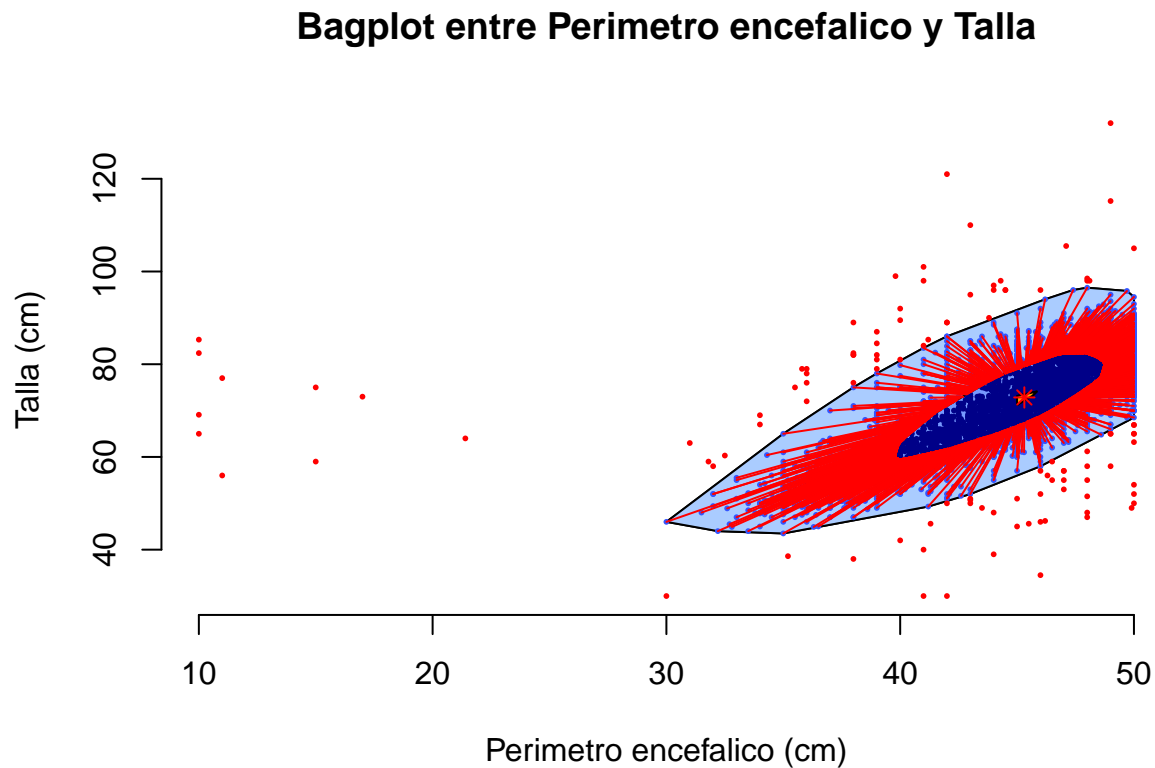
```
## Tipo_embarazo
## Multiple      Simple
## 0.03347428 0.96652572
```



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.

## Ejercicio 6

```
bagplot_pc_t <- bagplot(  
  Perim_encef,  
  Talla,  
  main = "Bagplot entre Perimetro encefalico y Talla",  
  xlab="Perimetro encefalico (cm)",  
  ylab="Talla (cm)"  
)
```



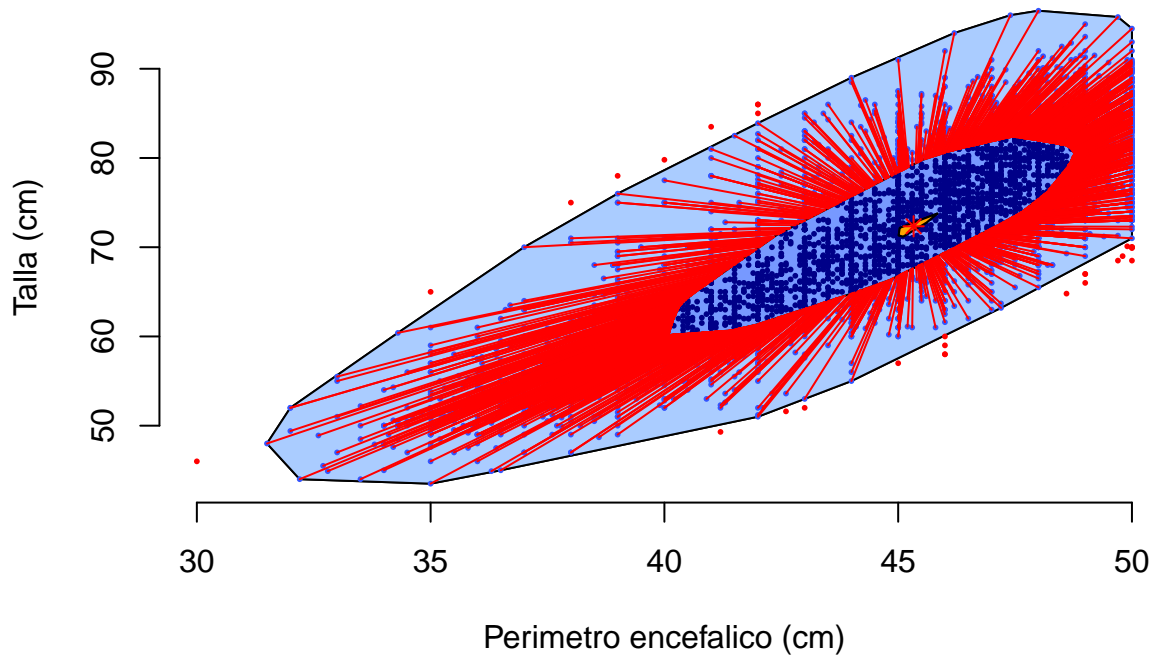
Se pueden identificar fácilmente varios puntos atípicos, especialmente con perímetro encefálico chico, y tallas alrededor de la media de la talla. No necesariamente tienen talla alta en estos casos. Sí se ven tallas muy altas (Hay puntos con tallas mayores a 90 cm). También casos donde la talla es muy baja (menor a 40 cm) y un perímetro encefálico muy alto (cercano a 50 cm).

## Ejercicio 7

```
bagplot_pc_t_2 <- bagplot(  
  c(bagplot_pc_t$pxy.bag[, 'x'], bagplot_pc_t$pxy.outer[, 'x']),
```

```
c(bagplot_pc_t$pxy.bag[, 'y'], bagplot_pc_t$pxy.outer[, 'y']),
main = "Bagplot entre Perimetro encefalico y Talla",
xlab="Perimetro encefalico (cm)",
ylab="Talla (cm)"
)
```

## Bagplot entre Perimetro encefalico y Talla



Se visualizan datos atípicos, mas muchos menos que con los datos anteriores y más cercanos a la cápsula convexa.

## Ejercicio 8

```
#plot(NULL, xlim = range(Perim_encef), ylim = range(Talla), xlab = "Perimetro encefalico", ylab = "Talla")

xlim = c(min(Perim_encef), max(Perim_encef))
ylim = c(min(Talla), max(Talla))

par(mfrow = c(1, 2))

for (sexo in unique(Sexo)) {
  subset_data <- datos_encuesta[Sexo == sexo, ]

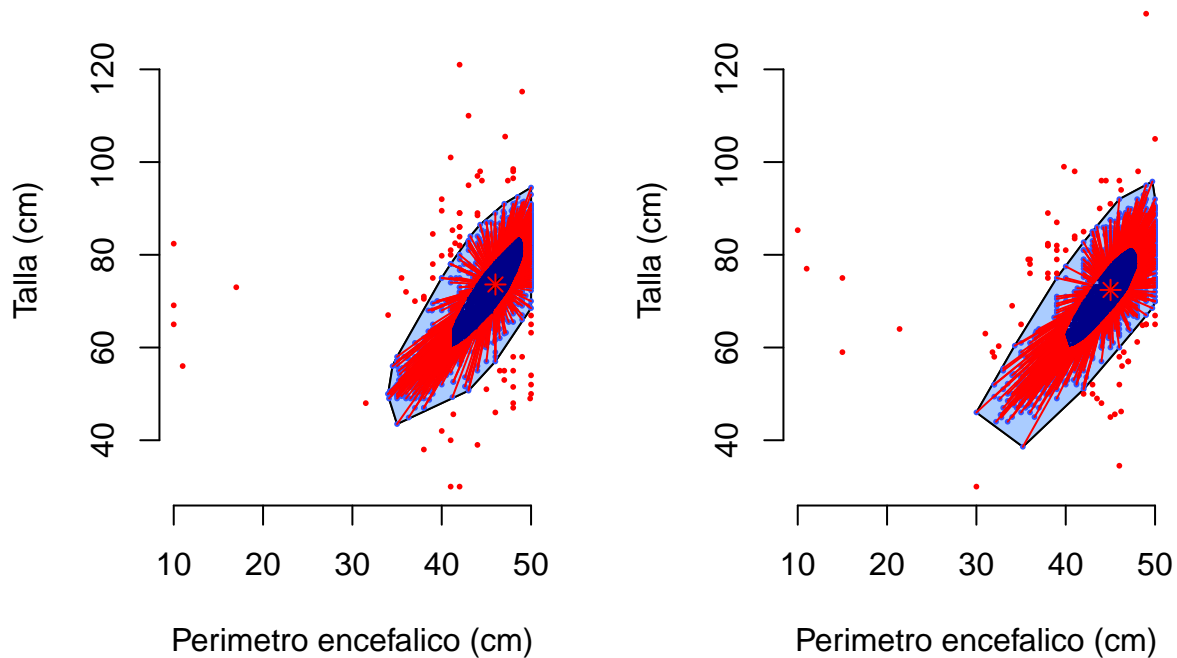
  bagplot_pc_t <- bagplot(
    subset_data$Perim_encef,
    subset_data$Talla,
```

```

main = sprintf("Bagplot entre Perimetro encefalico y Talla (%s)", sexo),
xlab="Perimetro encefalico (cm)",
ylab="Talla (cm)",
xlim=xlim,
ylim=ylim
)
}

```

## lot entre Perimetro encefalico y Tallaot entre Perimetro encefalico y Tal



Se puede observar que existe una distribución conjunta muy similar para ambos sexos, y la ubicación de outliers también muy similar. Parecería ser que la relación entre perímetro encefálico y talla no cambia mucho entre ambos, y que la aparición de outliers no parece depender del sexo.