

Estudio de datos biométricos sobre el ciervo volante.

Olivia Marcos Diaz de Otazu Irene Canales Giménez
Samuel Melián Benito

2025-12-21

Índice

Objetivo del estudio	3
Metodología	4
Resultados	5
Análisis de la anchura de la cabeza de los individuos de la muestra	5
Comparación de la anchura de la cabeza por sexos	5
Comparación de la anchura de la cabeza por provincias	6
Correlación de la anchura de la cabeza y longitud de élitros	7
Correlación de la anchura de la cabeza y longitud de élitros por sexo	8
Modelo de regresión lineal	8
Conclusiones	12
Anexo:Código R	13

Objetivo del estudio

El estudio de Estadística Descriptiva que vamos a realizar tiene como foco el análisis de datos biométricos del **ciervo volante** (*Lucanus cervus*), el mayor escarabajo de Europa y una especie cuya supervivencia está amenazada por la destrucción de su hábitat, bosques con abundancia de madera muerta.

El interés biométrico de esta especie nos lleva a dos preguntas principales:

Dimorfismo Sexual: analizaremos si existen diferencias significativas en tamaño entre machos y hembras.

Influencia del Hábitat: indagaremos si el tamaño adulto de los ejemplares está correlacionado con la calidad del hábitat de procedencia (Asturias, Cantabria u otra).

Se han tomado medidas en machos y hembras de ciervo volante de tres procedencias, midiendo las variables indicadas en la figura siguiente:



El fichero CIERVO.txt recoge, para un total de 265 ejemplares de ciervo volante, parte de estas variables. Concretamente:

- Anchura de la cabeza (KB) expresada en milímetros
- Longitud de los élitros (EL) también en milímetros
- Sexo del individuo (hembra o macho)
- Provincia de procedencia del individuo (Asturias, Cantabria u otra)

Los objetivos de este estudio son:

- Realizar un análisis descriptivo de la anchura de la cabeza de una muestra de 150 ejemplares que refleje sus principales características.
- Comparar la anchura de la cabeza de las hembras con la de los machos.
- Comparar la anchura de la cabeza de los ejemplares de las tres procedencias.
- Analizar la correlación entre la anchura de la cabeza y la longitud del élitro para la muestra.
- Analizar, para cada sexo por separado, la correlación entre la anchura de la cabeza y la longitud del élitro.
- En caso de que para alguno de los sexos ambas variables estén relacionadas, ajustar una recta de regresión que explique la anchura de la cabeza en función de la longitud del élitro e interpretarla.

Metodología

Como hemos mencionado, el siguiente estudio se basa en un análisis estadístico descriptivo de datos biométricos del ciervo volante (*Lucanus cervus*). En concreto, en relación a patrones de variación morfológica por sexo y por procedencia de los individuos.

Datos y variables

La muestra global comprende a 265 ejemplares. Sobre ellos se tienen datos de la anchura de la cabeza (KB) y la longitud de los élitros (EL). La procedencia de los ejemplares comprende Asturias, Cantabria, Otras y Desconocida.

En concreto, se espera que la variable KB permita observar el dimorfismo sexual, mientras que la variable EL se toma también como medida representativa del tamaño, global en este caso, del ejemplar. Por tanto, se espera también una posible relación entre ambas variables

Selección de la muestra

Se tomará una muestra de 150 ejemplares, de forma que sea una selección representativa del total de la población. Se hará de forma arbitraria y empleando una semilla para asegurar la aleatoriedad y evitar posibles sesgos.

Análisis estadístico

Para comenzar se llevará a cabo un análisis descriptivo para la anchura de la cabeza (KB) para la muestra escogida. En concreto, se analizarán los estadísticos siguientes: media, mediana, cuartiles, desviación típica, rango intercuartílico y valores extremos.

Esto permitirá llevar a cabo una comparación entre machos y hembras, así como un estudio en función del hábitat. A través de las tablas y gráficos adecuados, se buscará observar diferencias según las variables, así como dispersión y valores atípicos.

A continuación, un análisis entre ambas variables: anchura de la cabeza (KB) y longitud de los élitros (EL). El objetivo será encontrar una posible correlación y observar su dispersión. Se llevará cabo sobre la muestra, y por sexo.

Para terminar, se empleará regresión lineal entre las variables para explicar la influencia de una sobre los valores de la otra, y se interpretarán los coeficientes obtenidos para ello.

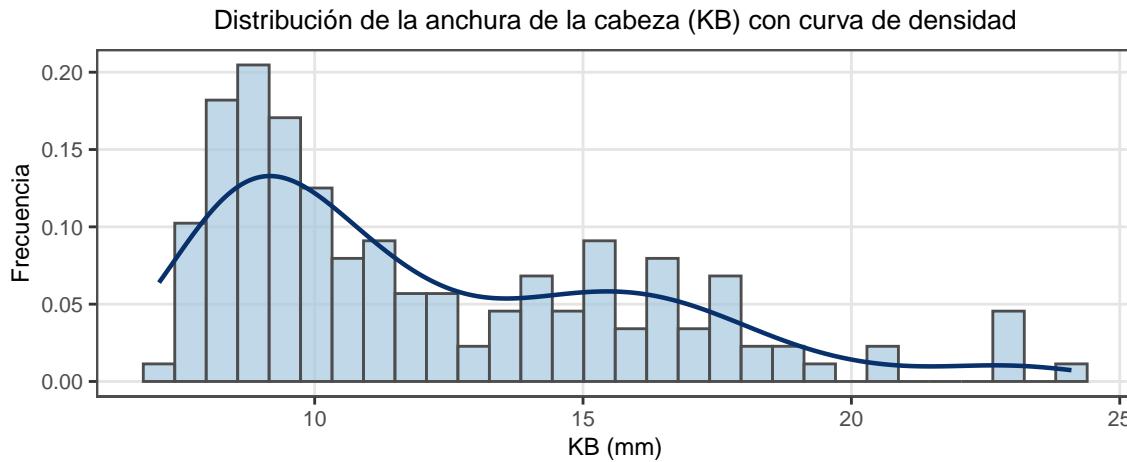
Todo este estudio se realizará a través del software R.

Resultados

Análisis de la anchura de la cabeza de los individuos de la muestra

Table 1: Estadísticos descriptivos de la anchura de la cabeza (KB)

N	Media	Q1	Mediana	Q3	Desv. típica	IQR	Mín	Máx
150	12.1493	8.9625	10.8	15.1	3.9706	6.1375	7.1	24.1



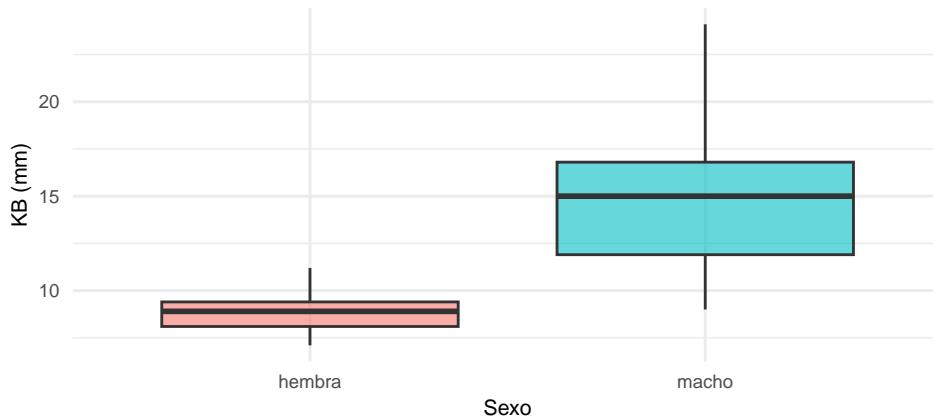
El histograma representa la distribución de la anchura de la cabeza de nuestra muestra de 150 ejemplares. Presenta un pico (moda) alrededor de los 8,5mm y asimetría positiva, con una cola alargada hacia los valores más grandes, lo que implica que la mayoría de individuos presenta una anchura de cabeza pequeña y un número reducido alcanza valores mayores. Los estadísticos obtenidos en la tabla refuerzan este argumento ya que la media (12,15mm) es mayor a la mediana (10,8mm).

Comparación de la anchura de la cabeza por sexos

Table 2: Estadísticos descriptivos de la anchura de la cabeza (KB) por sexo

Sexo	N	Media	Desv. típica	Mediana	IQR
hembra	69	8.9319	0.9107	8.9	1.3
macho	81	14.8901	3.4819	15.0	4.9

Anchura de la cabeza (KB) por sexo



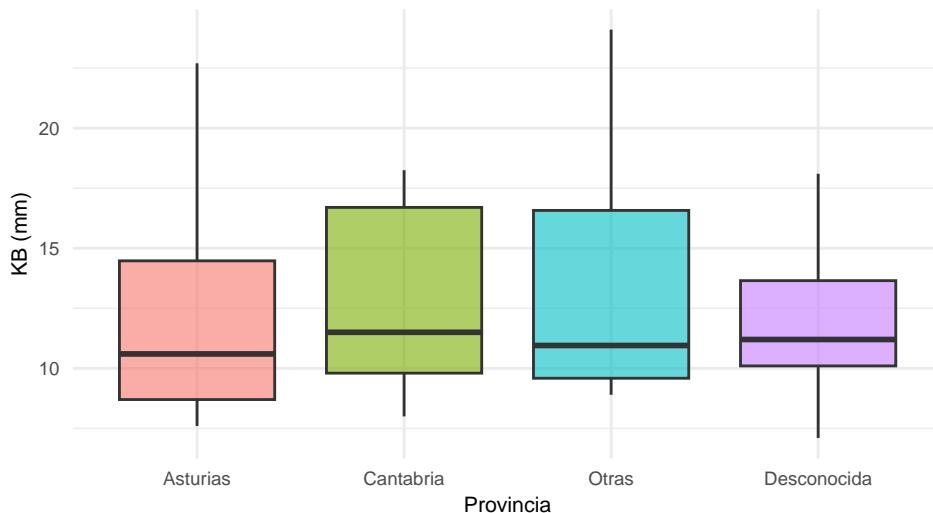
Tanto la Tabla 2 como el diagrama de caja muestran que los machos presentan una anchura de cabeza promedio mayor que las hembras, con una diferencia de 5,96mm. Además, los machos presentan mayor variabilidad, su desviación típica de 3,48 y IQR 4,9 frente a los 0,91 y 1,3 de las hembras indican mayor dispersión en los valores de los machos. Estas claras diferencias sugieren un marcado dimorfismo sexual en la anchura de la cabeza.

Comparación de la anchura de la cabeza por provincias

Table 3: Estadísticos descriptivos de la anchura de la cabeza (KB) por provincia

Provincia	N	Media	Desv. típica	Mediana	IQR
Asturias	102	11.7221	3.6813	10.60	5.7750
Cantabria	17	12.9353	3.7930	11.50	6.9000
Otras	24	13.4562	5.1039	10.95	6.9875
Desconocida	7	11.9857	3.6667	11.20	3.5500

Anchura de la cabeza (KB) por provincia



En este análisis por provincia podemos observar que Asturias concentra la mayor parte de la muestra. Otras presenta la mayor media seguida por Cantabria, y los individuos de procedencia desconocida ligeramente mayor que los ejemplares con menos promedio de anchura de cabeza que son los de Asturias. La variabilidad es mayor en el grupo de Otras indicado por su mayor desviación típica y mayor IQR.

En el diagrama de cajas los valores de la anchura de la cabeza de las distintas provincias se sitúan en intervalos muy parecidos por lo que las diferencias son moderadas entre provincias.

Correlación de la anchura de la cabeza y longitud de élitros

La siguiente gráfica representa el análisis comparativo de las dos características de la muestra completa.

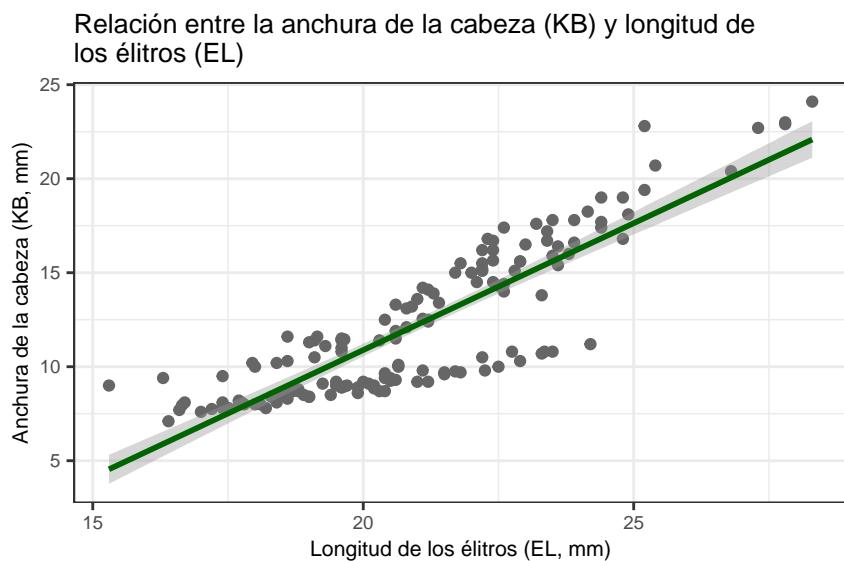


Table 4: Valores de R^2 y r para la regresión KB - EL (todos los individuos)

	Coeficiente de determinación (R^2)	Coeficiente de correlación (r)
EL	0.756	0.8695

Interpretación: La correlación es fuerte al ser la r positiva cerca de 1. Además, el valor de R^2 muestra que un 75,6% de la variabilidad de la variable KB (anchura de la cabeza) se explica en función de los valores de la otra, EL (longitud de los élitros), de nuevo reiterando la relación entre ellas.

Correlación de la anchura de la cabeza y longitud de élitros por sexo

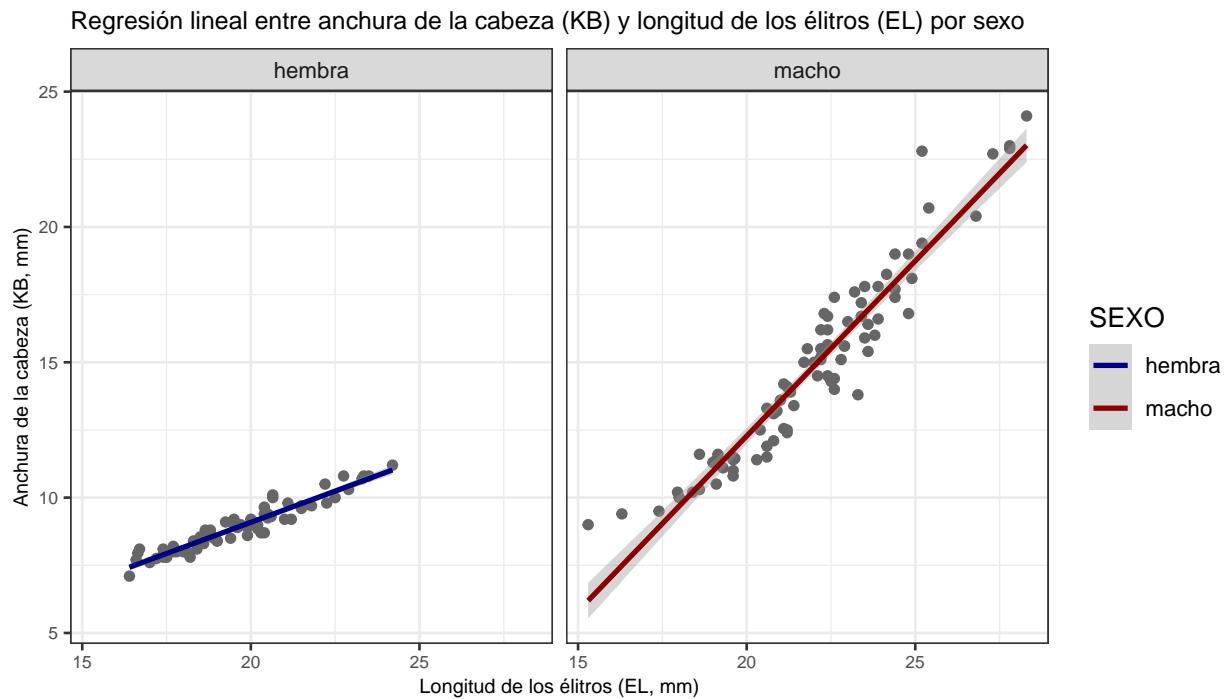


Table 5: Coeficientes de correlación y determinación de la regresión KB y EL por sexo

Sexo	Coeficiente de correlación (r)	Coeficiente de determinación (R^2)
Hembra	0.9549	0.9118
Macho	0.9533	0.9089

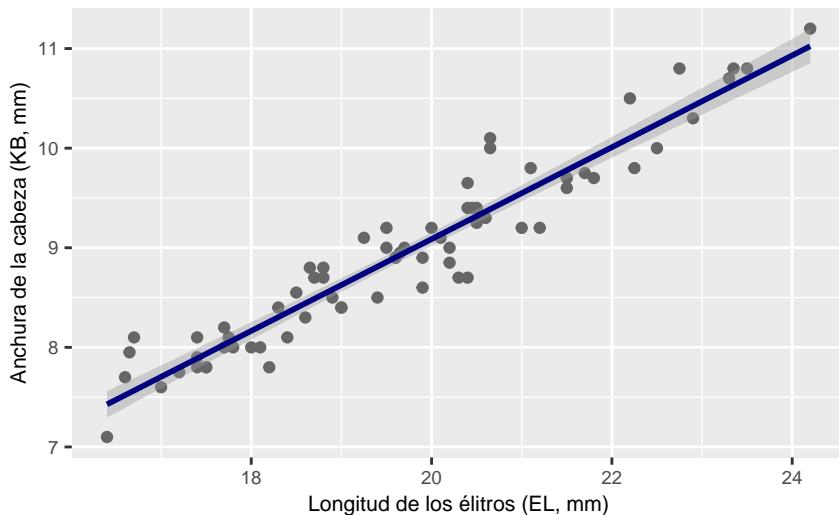
Interpretación: la correlación es muy fuerte tanto en el caso de las hembras como en el de los machos, y se observa una mejora considerable al separarlo por sexos.

Modelo de regresión lineal

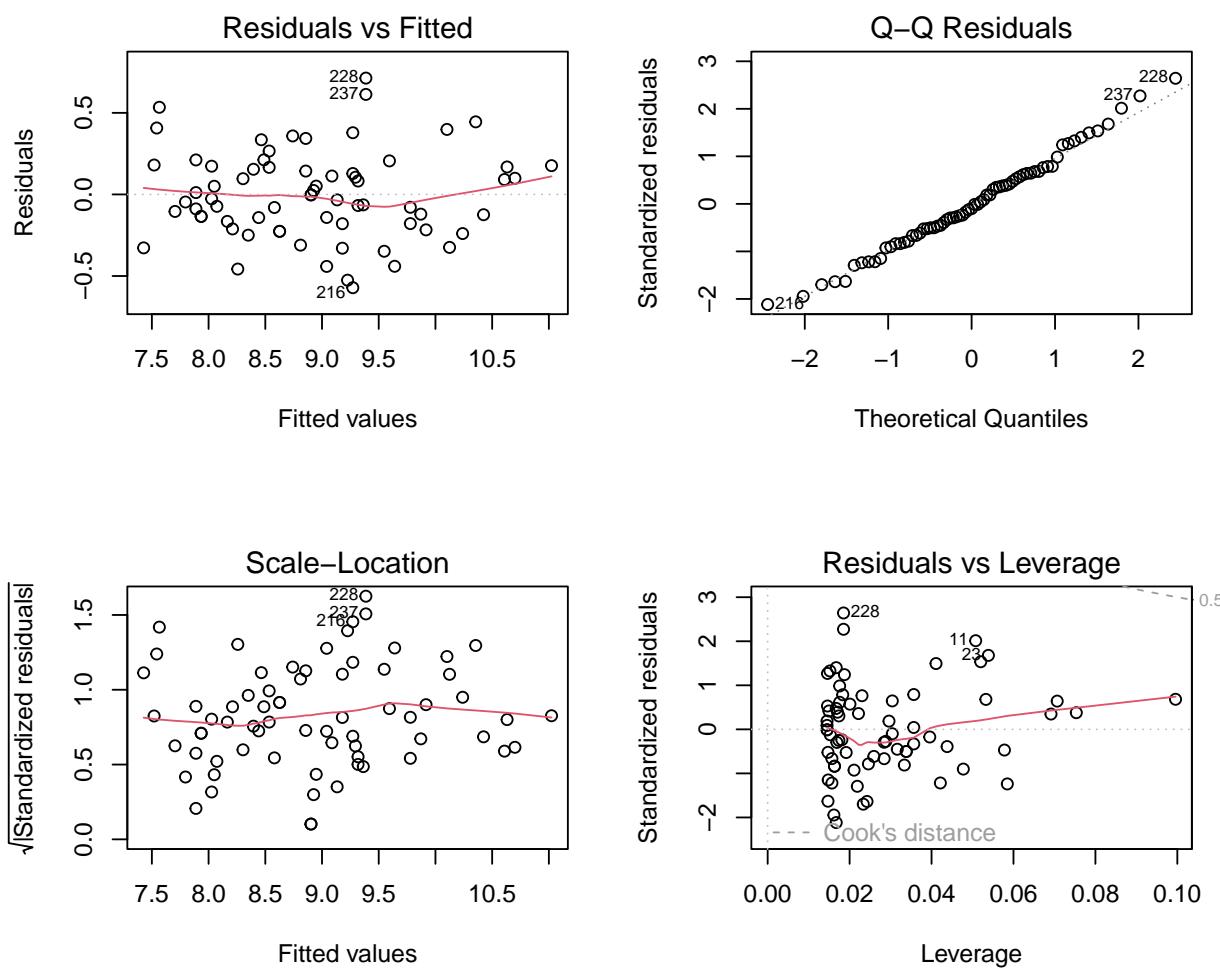
Table 6: Coeficientes del modelo de regresión lineal KB y EL para hembras

	Estimación	Error estándar	Estadístico t	Valor p
(Intercept)	-0.1333	0.3459	-0.3853	0.7012
EL	0.4610	0.0175	26.3257	0.0000

Modelo de regresión lineal KB vs EL (Hembra)



Diagnóstico del modelo de regresión lineal (Hembras)



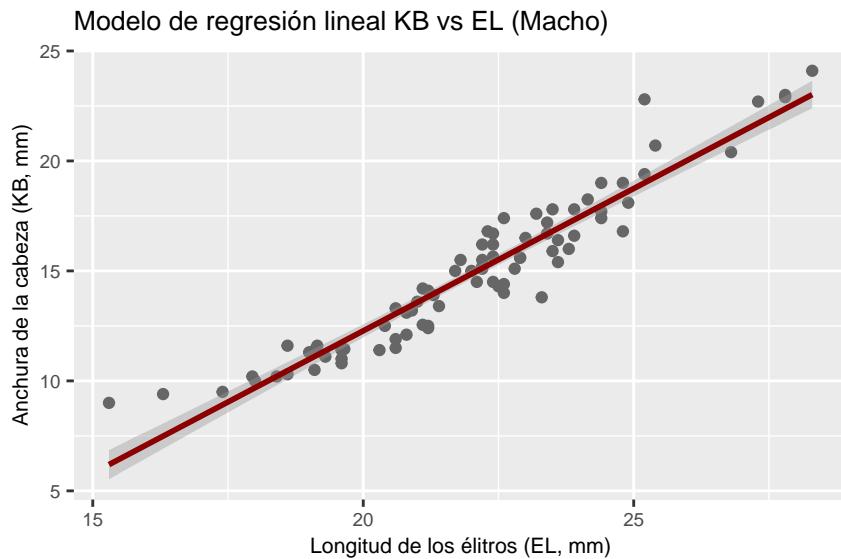
El modelo de regresión lineal ajustado para las hembras viene dado por

$$\widehat{KB} = -0.1333 + 0.4610 EL.$$

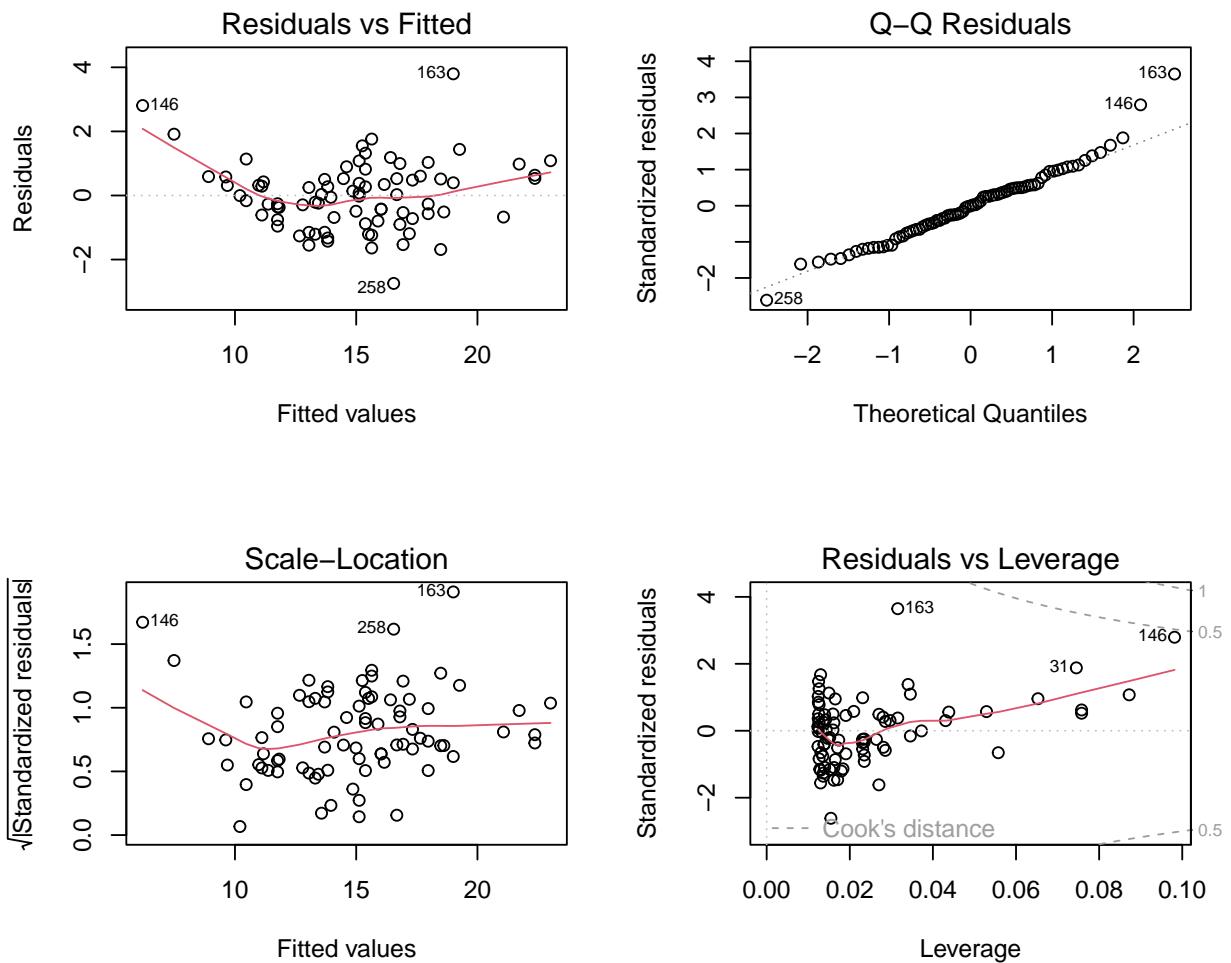
La pendiente positiva indica que, en promedio, la anchura de la cabeza aumenta aproximadamente 0.46 mm por cada incremento de 1 mm en la longitud de los élitros. El coeficiente de determinación es $R^2 \approx 0.9118$, lo que implica que alrededor del 91.2% de la variabilidad de la anchura de la cabeza queda explicada por la longitud de los élitros. Además, las gráficas de diagnóstico no muestran desviaciones importantes de los supuestos del modelo lineal.

Table 7: Coeficientes del modelo de regresión lineal KB y EL para machos

	Estimación	Error estándar	Estadístico t	Valor p
(Intercept)	-13.6044	1.0219	-13.3125	0
EL	1.2940	0.0461	28.0695	0



Diagnóstico del modelo de regresión lineal (Machos)



Para los machos, el modelo de regresión lineal ajustado es

$$\widehat{KB} = -13.6044 + 1.2940 EL.$$

En este caso, la pendiente es algo mayor que en el modelo para las hembras, lo que indica que un aumento de 1 mm en la longitud de los élitros se asocia con un incremento medio de aproximadamente 1.29 mm en la anchura de la cabeza. El coeficiente de determinación es $R^2 \approx 0.9089$, lo que nos hace ver que el modelo se ajusta muy bien. El análisis de las gráficas de diagnóstico también confirma la adecuación del ajuste del modelo lineal en este caso.

Conclusiones

En este análisis estadístico se ha realizado un análisis descriptivo de datos biométricos del ciervo volante (*Lucanus cervus*) a partir de una muestra aleatoria de 150 individuos. El estudio en su totalidad se ha centrado en el análisis de la anchura de la cabeza, su variabilidad en función del sexo y procedencia geográfica y la relación con la longitud de los élitros.

En resumen, la variable KB (anchura de la cabeza) presenta una distribución asimétrica (a la derecha), con una media de aproximadamente 12,14 mm, algo superior a la mediana (11,2 mm) y una dispersión relativamente moderada. La anchura de la cabeza toma valores comprendidos entre 7 y 20 mm aproximadamente, lo que nos ha hecho ver que existe variabilidad morfológica en nuestra muestra tomada.

A continuación, el análisis por sexos nos ha permitido comprobar la existencia de un claro dimorfismo sexual. Los machos presentan una anchura de la cabeza media de 14,91 mm, lo que resulta casi 6 mm más que el promedio de las hembras de la muestra. Además, la variabilidad en los machos es aproximadamente el triple que en las hembras, mostrando estas últimas valores bastante concentrados en torno a la media como se puede comprobar en la respectiva caja del diagrama. Estas conclusiones han sido posibles gracias a estadísticos descriptivos y diagramas de caja.

Pasando ahora a analizar la variación geográfica, podemos ver que se detectan pequeñas diferencias en la anchura de la cabeza, pero no muy significativas. Los valores no nos han permitido ver contrastes tan marcados como los del sexo. No obstante, los marcados con origen desconocido son los que presentan mayor promedio de anchura de la cabeza (13,2 mm), seguidos de los de Cantabria (12,9 mm), Asturias (12,1 mm) y otras procedencias (11,2 mm).

Ahora pasamos a analizar otro punto muy importante del estudio, que es la relación entre la anchura de la cabeza (KB) y la longitud de los élitros (EL). Podemos afirmar que la correlación entre ambas variables es fuertemente lineal y positiva (el coeficiente de correlación es aproximadamente 0,9), lo que nos indica que existe una relación fuerte entre ambas variables. No obstante, comprobamos que esta relación se refuerza todavía más al realizar el análisis por sexos, siendo los valores de coeficiente de correlación de 0,95 para las hembras y 0,93 para los machos. Esto nos permite sacar la conclusión de que una gran parte de la variabilidad de la anchura de la cabeza puede explicarse mediante la longitud de los élitros.

Cuando hacemos el ajuste de regresión lineal por sexos podemos confirmar esto último que hemos comentado. En ambos casos (machos y hembras), la pendiente estimada es positiva y estadísticamente significativa, lo que indica que incrementos en la longitud del élito se asocian con aumentos en la anchura de la cabeza. Esta pendiente es mayor en los machos que en las hembras, lo que significa que en este grupo la anchura de la cabeza crece más en función de la longitud de los élitros. Los valores de R^2 son muy elevados en ambos casos, lo que indica que existe un buen ajuste de los modelos y por tanto resultan realmente útiles para predecir.

A modo de resumen, estos resultados nos permiten concluir que el sexo es el factor más influyente (de los analizados) en la anchura de la cabeza del ciervo volante, mientras que la procedencia geográfica tiene un efecto mucho menor. Además, la fuerte relación entre la anchura de la cabeza y la longitud de los élitros nos permite afirmar que la longitud de los élitros es una buena variable para explicar el tamaño de la cabeza, y especialmente cuando el análisis se hace de manera separada por sexos.

Anexo:Código R

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
library(tidyverse)
library(summarytools)
library(GGally)
library(gt)
library(flextable)
library(knitr)
library(corrplot)
library(ggplot2)
library(dplyr)
library(kableExtra)
knitr::include_graphics("imagenes/imagen.png")
datos <- read.table("./datos/CIERVO.txt", fill=TRUE, dec = ",", na.strings =
  c("", "NA"), header=TRUE );

datos$PROVINCIA[is.na(datos$PROVINCIA)] <- "Desconocida"

datos$SEXO <- factor(datos$SEXO, levels = c("hembra", "macho"))
datos$PROVINCIA <- factor(datos$PROVINCIA, levels = c("Asturias", "Cantabria",
  "Otras", "Desconocida"))
set.seed(78) #Ponemos semilla random para asegurar la aleatoriedad a la hora de
  elegir la muestra de individuos y así evitar posibles sesgos

indices <- sample(1:nrow(datos), size = 150, replace = FALSE) #Tomamos la muestra
  aleatoria de 150 individuos

muestra <- datos[indices, ] #Creamos el dataframe de esta nueva muestra de 150
  individuos
tabla_kb <- muestra |>
  summarise(
    N = n(),
    Media = mean(KB, na.rm = TRUE),
    Q1 = quantile(KB, 0.25, na.rm = TRUE),
    Mediana = median(KB, na.rm = TRUE),
    Q3 = quantile(KB, 0.75, na.rm = TRUE),
    DE = sd(KB, na.rm = TRUE),
    IQR = IQR(KB, na.rm = TRUE),
    Min = min(KB, na.rm = TRUE),
    Max = max(KB, na.rm = TRUE)
  )

tabla_kb |>
  kable(
    col.names = c(
      "N", "Media", "Q1", "Mediana", "Q3",
```

```

    "Desv. típica", "IQR", "Mín", "Máx"
),
caption = "Estadísticos descriptivos de la anchura de la cabeza (KB)",
digits = 4,
booktabs = TRUE,
align = "c"
) |>
kable_styling(
  latex_options = c("hold_position","striped"),
  font_size = 10
)
muestra |> ggplot(aes(KB)) +
  geom_histogram(
    aes(y = after_stat(density)),
    fill = "#A6C8E0",
    color = "grey30",
    bins = 30,
    alpha = 0.7
  ) +
  geom_density(
    color = "#08306B",
    linewidth = 0.8,
    alpha = 0.5
  ) +
  labs(
    title = "Distribución de la anchura de la cabeza (KB) con curva de densidad",
    x = "KB (mm)",
    y = "Frecuencia"
  ) +
  theme_bw() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 10),
    axis.title = element_text(size = 9),
    axis.text = element_text(size = 8),
    panel.grid.major = element_line(color = "grey90"),
    panel.grid.minor = element_blank(),
    panel.border = element_rect(color = "grey30", fill = NA, size = 0.8)
  )
tabla_kb_sexo <- muestra |>
  group_by(SEXO) |>
  summarise(
    N = n(),
    Media = mean(KB, na.rm = TRUE),
    DE = sd(KB, na.rm = TRUE),
    Mediana = median(KB, na.rm = TRUE),

```

```

IQR = IQR(KB, na.rm = TRUE),
)

tabla_kb_sexo |>
  kable(
    col.names = c("Sexo", "N", "Media", "Desv. típica", "Mediana", "IQR"),
    caption = "\ Estadísticos descriptivos de la anchura de la cabeza (KB) por
      ↵ sexo",
    digits = 4,
    booktabs = TRUE,
    align = "c",
    escape = FALSE
  ) |>
  kable_styling(
    latex_options = c("hold_position","striped"),
    font_size = 10
  )

ggplot(muestra, aes(x = SEXO, y = KB, fill = SEXO)) +
  geom_boxplot(alpha = 0.6) +
  labs(
    title = "Anchura de la cabeza (KB) por sexo",
    x = "Sexo",
    y = "KB (mm)"
  ) +
  theme_minimal() +
  theme(legend.position = "none") +
  theme(
    plot.title = element_text(size = 10),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 7)
  )
tabla_kb_provincia <- muestra |>
  group_by(PROVINCIA) |>
  summarise(
    N = n(),
    Media = mean(KB, na.rm = TRUE),
    DE = sd(KB, na.rm = TRUE),
    Mediana = median(KB, na.rm = TRUE),
    IQR = IQR(KB, na.rm = TRUE),
  )
)

tabla_kb_provincia |>
  kable(
    col.names = c(
      "Provincia", "N", "Media",
      "Desv. típica", "Mediana", "IQR"
    ),

```

```

caption = "Estadísticos descriptivos de la anchura de la cabeza (KB) por
  ↵ provincia",
digits = 4,
booktabs = TRUE,
align = "c"
) |>
kable_styling(
  latex_options = c("hold_position", "striped"),
  full_width = FALSE
)
ggplot(muestra, aes(x = PROVINCIA, y = KB, fill = PROVINCIA)) +
  geom_boxplot(alpha = 0.6) +
  labs(
    title = "Anchura de la cabeza (KB) por provincia",
    x = "Provincia",
    y = "KB (mm)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")+
  theme(
    plot.title = element_text(size = 10),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 7)
  )
muestra |>
  select(KB, EL) |>
  ggplot(aes(x=EL, y = KB)) +
  geom_point(
    color = "grey40",
  ) +
  theme_bw() +
  labs(
    x = "Longitud de los élitros (EL, mm)",
    y = "Anchura de la cabeza (KB, mm)",
    title = "Relación entre la anchura de la cabeza (KB) y longitud de\nlos
      ↵ élitros (EL)"
  ) +
  geom_smooth(
    method = lm,
    se = TRUE,
    color = "darkgreen"
  ) +
  theme(
    plot.title = element_text(size = 10),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 7)
  )
modelo <- muestra |>

```

```

select(KB, EL) |>
with(lm(KB ~ EL))

#Valor de r^2
r2 <- summary(modelo)$r.squared
#r2

#Valor de r
r <- sign(coef(modelo)[2]) * sqrt(r2)
#r
tabla_cor <- data.frame(
  r2 = r2,
  r = r
)

tabla_cor |>
kable(
  col.names = c("Coeficiente de determinación ( $R^2$ )", "Coeficiente de
  ↵ correlación (r)"),
  digits = 4,
  booktabs = TRUE,
  caption = "Valores de  $R^2$  y r para la regresión KB - EL (todos los
  ↵ individuos)",
  align = "c"
) |>
kable_styling(
  latex_options = c("hold_position", "striped"),
  full_width = FALSE
)

muestra |>
ggplot(aes(x = EL, y = KB)) +
facet_wrap(~ SEXO) +
geom_point(
  color = "grey40",
) +
labs(
  x = "Longitud de los élitros (EL, mm)",
  y = "Anchura de la cabeza (KB, mm)",
  title = "Regresión lineal entre anchura de la cabeza (KB) y longitud de los
  ↵ élitros (EL) por sexo"
) +
geom_smooth(
  method = "lm",
  aes(color = SEXO)
) +
scale_color_manual(values = c("hembra" = "navy", "macho" = "darkred")) +
theme_bw()+

```

```

theme(
  plot.title = element_text(size = 10),
  axis.title = element_text(size = 8),
  axis.text = element_text(size = 7)
);

modelo_hembra <- lm(
  KB ~ EL,
  data = muestra,
  subset = SEXO == "hembra"
)
modelo_macho <- lm(
  KB ~ EL,
  data = muestra,
  subset = SEXO == "macho"
)
r2_hembra <- summary(modelo_hembra)$r.squared
r2_macho <- summary(modelo_macho)$r.squared

r_hembra <- sign(coef(modelo_hembra)[2]) * sqrt(r2_hembra)
r_macho <- sign(coef(modelo_macho)[2]) * sqrt(r2_macho)
tabla_correlacion <- data.frame(
  Sexo = c("Hembra", "Macho"),
  r = c(r_hembra, r_macho),
  r2 = c(r2_hembra, r2_macho)
)

tabla_correlacion |>
  kable(
    col.names = c("Sexo", "Coeficiente de correlación (r)", "Coeficiente de
      ↵ determinación ( $R^2$ )"),
    caption = "Coeficientes de correlación y determinación de la regresión KB y
      ↵ EL por sexo",
    digits = 4,
    booktabs = TRUE,
    align = "c"
  ) |>
  kable_styling(
    latex_options = c("hold_position", "striped"),
    full_width = FALSE,
    font_size = 10
  )

summary(modelo_hembra)$coefficients |>
  as.data.frame() |>
  kable(
    col.names = c(
      "Estimación",

```

```

    "Error estándar",
    "Estadístico t",
    "Valor p"
),
caption = "Coeficientes del modelo de regresión lineal KB y EL para hembras",
digits = 4,
booktabs = TRUE,
align = "c"
) |>
kable_styling(
  latex_options = c("hold_position", "striped"),
  full_width = FALSE
)

muestra |>
  filter(SEXO == "hembra") |>
  ggplot(
    aes(x = EL, y = KB),
  ) +
  geom_point(
    color = "grey40"
  ) +
  geom_smooth(
    method = "lm",
    color = "navy"
  ) +
  labs(
    x = "Longitud de los élitros (EL, mm)",
    y = "Anchura de la cabeza (KB, mm)",
    title = "Modelo de regresión lineal KB vs EL (Hembra)",
  )+
  theme(
    plot.title = element_text(size = 10),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 7)
)
par(mfrow = c(2, 2))
plot(modelo_hembra)

mtext("Diagnóstico del modelo de regresión lineal (Hembras)",
      outer = TRUE, line = -1, cex = 1)
summary(modelo_macho)$coefficients |>
  as.data.frame() |>
  kable(
    col.names = c(
      "Estimación",
      "Error estándar",
      "Estadístico t",

```

```

    "Valor p"
),
caption = "Coeficientes del modelo de regresión lineal KB y EL para machos",
digits = 4,
booktabs = TRUE,
align = "c"
) |>
kable_styling(
  latex_options = c("hold_position", "striped"),
  full_width = FALSE
)
muestra |>
  filter(SEXO == "macho") |>
  ggplot(
    aes(x = EL, y = KB),
  ) +
  geom_point(
    color = "grey40"
  ) +
  geom_smooth(
    method = "lm",
    color = "darkred"
  ) +
  labs(
    x = "Longitud de los élitros (EL, mm)",
    y = "Anchura de la cabeza (KB, mm)",
    title = "Modelo de regresión lineal KB vs EL (Macho)", ) +
  theme(
    plot.title = element_text(size = 10),
    axis.title = element_text(size = 8),
    axis.text = element_text(size = 7)
  )
par(mfrow = c(2, 2))
plot(modelo_macho)

mtext("Diagnóstico del modelo de regresión lineal (Machos)",
      outer = TRUE, line = -1, cex = 1)

```