

Proyecto Final del Curso R Redbioma
Análisis de datos sobre densidad y biomasa del inventario de candelilla en el estado de Coahuila
Biól. Jaqueline Jeniffer Noguez Lugo

1. Objetivo del análisis:

A través de los datos del inventario de candelilla en el estado de Coahuila se busca evaluar la relación entre la densidad de plantas y su biomasa, estimar la productividad y determinar si la población es saludable y productiva.

2. Método:

Mediante el levantamiento de datos en campo, en un total de 68 predios bajo aprovechamiento de candelilla en el estado de Coahuila, se obtuvo información de densidades y biomasa total en 5,678 sitios con una densidad promedio de 2.1 ind/sitio, para los que se estimó la biomasa total por sitio. Se analizaron los datos en R siguiendo los siguientes pasos: 1) Lectura de datos, 2) Estimación de la productividad, 3) Preprocesamiento: Limpieza y acomodo de datos, 4) Análisis estadísticos para evaluar la relación entre las variables de densidad, biomasa y productividad. 5) Elaboración de gráficos.

3. Resultados: (descripción de los principales resultados). Sugiero en resultados usar una gráfica que describa los resultados.

A partir de los datos de biomasa total y densidad en los sitios se estimó la productividad para poder conocer si la población de candelilla en el estado es saludable y productiva, sin embargo, aunque los datos de densidad y biomasa presentan una alta relación, ya que, a medida que aumenta la densidad, también aumenta la biomasa, se observó que la población presenta una productividad baja esto puede deberse que en la mayoría de los sitios las densidades son bajas y que son pocos los casos con mayor biomasa reportada como se observa en la Figura 1.

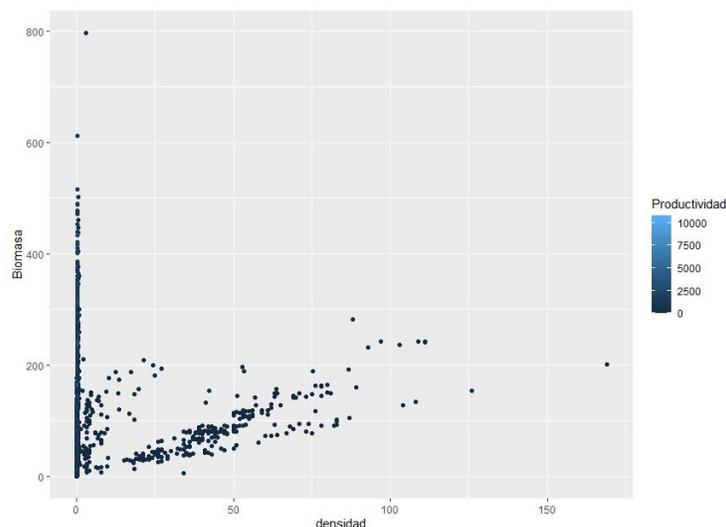


Figura 1. Diagrama de dispersión Densidad vs Biomasa de candelilla, con escala de colores para productividad.

Por otro lado, se analizó si existe correlación entre las variables de densidad, biomasa total y productividad, obteniendo los siguientes resultados:

1. Existe una correlación positiva (Correlación = 0.68, $p = 0$) entre densidad y biomasa, lo que sugiere que a medida que aumenta la densidad, también tiende a aumentar la biomasa. El valor p es cero, lo que indica que esta correlación es significativa.
2. Existe una correlación negativa (Correlación = -0.32, $p < 0.05$) entre densidad y productividad, lo que sugiere que, a mayor densidad, la productividad de la planta tiende a disminuir y dado que, el valor de p es bajo, esta relación es significativa.
3. Hay una correlación positiva pero baja (Correlación = 0.4, $p < 0.05$) entre biomasa y productividad, lo que sugiere que cuando aumenta la biomasa, también tiende a aumentar la productividad y dado que el valor de p indica es bajo, esta relación es significativa.

4. Discusión:

Se observa que la densidad en los sitios en general es baja y que los casos en los que se presenta mayor biomasa son muy pocos, por lo que la población presenta una productividad baja, lo que supone que la planta no se encuentra en un estado tan saludable y que esta ha estado sometida a condiciones no favorables para la recuperación de la especie. Se sabe que la cera de la candelilla tiene alta demanda en el mercado internacional, por lo que, se puede suponer que la sobreexplotación de las poblaciones de candelilla puede ser uno de los factores que influyen directamente en las condiciones actuales observadas de la población.

5. Anexo. Código en R utilizado para: la importación de datos, el análisis, las figuras (pdf y script adjuntos)

```

# Obtener directorio raíz
getwd()

install.packages("readxl")
library(readxl)

# Ruta al archivo Excel
archivo <- "C:/Users/jnoguez/Documents/Proyecto final R/Proyecto final Curso
R/Candelilla_2021.xlsx"

# Leer datos excel
datos <- read_excel("Candelilla_2021.xlsx",
                    sheet = "Candelilla")
class(datos)
head(datos)
str(datos)

# Skimr
# Explorar datos
install.packages("skimr")
library("skimr")
skim(datos)

View(datos)

# biomasa
# Primero vemos si hay NA
range(datos$densidad)

# De todas maneras ver min y max
range(datos$densidad,
      na.rm = TRUE)

# Primero vemos si hay NA
range(datos$`Biomasa total`)

range(datos$`Biomasa total`,
      na.rm = TRUE)

indices <- which(is.na(datos$`Biomasa total`))
datos$`Biomasa total`[indices]

datos <- read_excel("Candelilla_2021.xlsx",
                    sheet = "Candelilla")

install.packages("dplyr")
dplyr::glimpse(datos$`Biomasa total`)

# Gráficos
plot(datos$`densidad`,
      datos$`Biomasa total`)

plot(datos$`Biomasa total`)

plot(datos$`densidad (Ind/400m2)`, datos$`Biomasa total`,
      xlim = c(0,150),
      ylim = c(0,800))

# Histograma
hist(datos$`densidad`)
hist(datos$`Biomasa total`)

library(dplyr)

```

```

# Calcular la productividad
datos <- datos |>
  mutate(Productividad = `Biomasa total`/ `densidad`)

# Ver los resultados
print(datos)

library(tidyverse)
library(ggplot2)

ggplot(datos, aes(x = `Biomasa total`, y = Productividad, fill = `Biomasa total`)) +
  geom_point(stat = "identity") +
  labs(title = "Indicador de Productividad por Biomasa total",
       x = "Biomasa total",
       y = "Productividad (kg/individuo)") +
  theme_minimal()

library(rstatix)
library(tidyverse)
library(readxl)

rstatix::filter()
stats::filter()

# Preprocesamiento limpieza de datos
datos <- datos |>
  mutate(across(c(densidad, `Biomasa total`,`Productividad`), ~ifelse(.x >= 99999, NA,
.x))) |>
  # Eliminar NA
  drop_na(densidad) |>
  drop_na(`Biomasa total`) |>
  drop_na(`Productividad`) |>
  # Seleccionar columnas de interés
  select(`Predio`,`densidad`,`Biomasa total`,`Productividad`) |>
  # Renombrar columnas
  rename("Parcela" = "Predio",
        "densidad" = "densidad",
        "Biomasa" = "Biomasa total",
        "Productividad" = "Productividad")

# Estadística descriptiva
datos |>
  get_summary_stats(type = "mean_sd")

datos |>
  freq_table(densidad)

datos |>
  freq_table(Biomasa)

datos |>
  identify_outliers(densidad)

datos |>
  identify_outliers(Biomasa)

# Pruebas para ver si usamos paramétricas o no paramétricas

# Homogeneidad de varianzas
datos |>
  levene_test(densidad ~ biomasa)

# Correlación
datos |>
  select(densidad, Biomasa, Productividad) |>

```

```
cor_test(vars = c(densidad, Biomasa, Productividad),
         vars2 = c(densidad, Biomasa, Productividad),
         method = "spearman")

ggplot(datos, aes(x = `densidad`, y = Productividad, fill = `densidad`)) +
  geom_point(stat = "identity") +
  labs(title = "Indicador de Productividad por densidad",
       x = "densidad",
       y = "Productividad (kg/individuo)") +
  theme_minimal()

datos |>
  ggplot(aes(x = densidad,
            y = Biomasa,
            color = Productividad)) +
  geom_point()
```