



Análisis y presentación de datos

Estudios de Campo en Ecología Vegetal

Edición 2023



CONTENIDO

- Cómo analizar los datos?
 - Medidas de tendencia central
 - Medidas de variabilidad

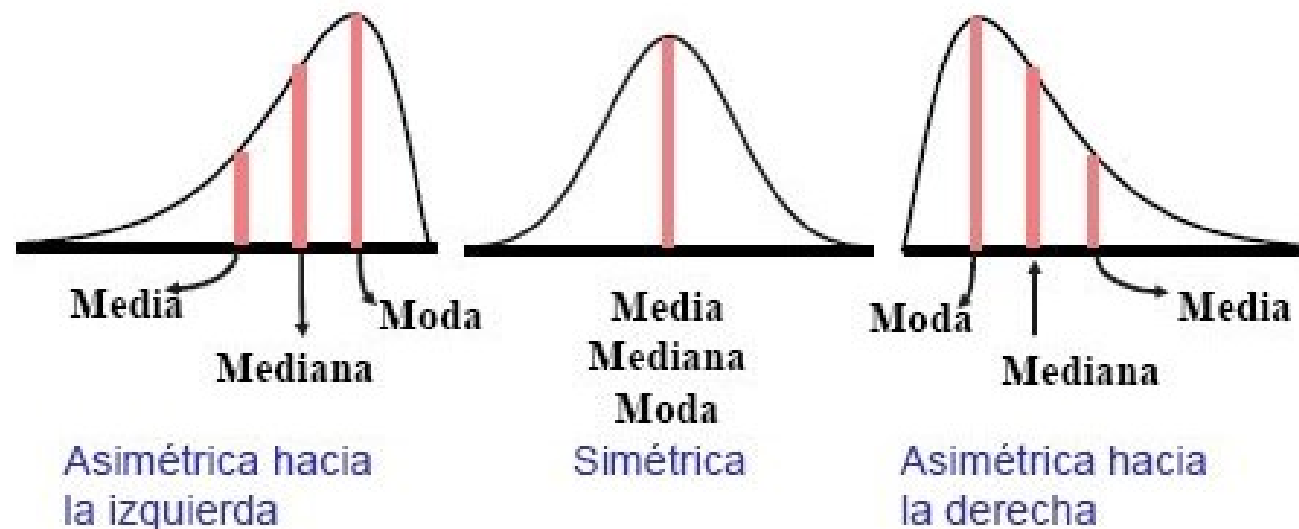
- Cómo presentamos los datos?

MEDIDAS DE TENDENCIA CENTRAL

En una muestra o población estadística, hay un valor que se encuentra cercano al centro de la distribución de observaciones y que los representa a todos

puede ser descrito por alguno de estos estadísticos (más utilizados):

- ✓ Media (promedio)
- ✓ mediana
- ✓ moda



Las tres medidas pueden coincidir o no, dependiendo de la distribución de los datos.

MEDIA (promedio)

La media (promedio) es el estadístico más utilizado para medir la tendencia central

Consiste en la suma de las observaciones individuales, cada una llamada x_i , dividida entre el número total de observaciones o tamaño de muestra, n .

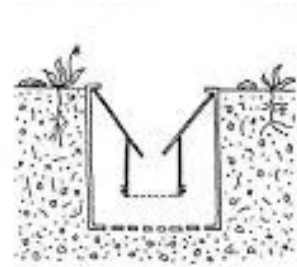
población $\mu = \frac{\sum x}{N}$

Muestra

$$\bar{x} = \frac{\sum x}{n}$$

Veamos un ejemplo!

EJEMPLO: tenemos 5 trampas para hormigas en dos ecosistemas diferentes (bosque y pastizal)



Trampa pitfall



Cuántas hormigas en promedio cayeron en cada ecosistemas?

15, 8, 25, 199, 4

$251/5=50,2$ hormigas/trampa

35, 93, 19, 34, 68

$249/5=49,8$ hormigas/trampa

PROBLEMAS CON LA MEDIA

La media aritmética presenta el problema de que es sensible a los valores extremos de la muestra (tanto altos como bajos)

•Ej: 4; 8; 9; 12; 15; 18; 250
Media = $316/7 = 45,1$

4; 8; 9; 12; 15; 18; 25
Media = $91/7 = 13$

Si modificamos un valor “no extremo”, la media varía menos

4; 8; 9; 12; 15; **1,8**; 250
Media = $299,8/7 = 42,83$

0,4; 8; 9; 12; 15; 18; 250
Media = $312,4/7 = 44,67$



MEDIANA

es el valor que está exactamente en el medio entre el menor y el mayor en un conjunto de observaciones ordenadas

En general se halla en la posición $(n+1)/2$

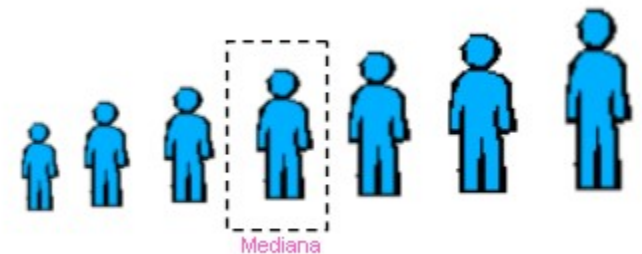
- Ej: 12; 202; 4; 8; 22; 25; 9

Para hallar la mediana lo primero es ordenar los valores de menor a mayor

- 4, 8, 9, 12, 22, 25, 202

En este caso, la mediana es el valor que está en la cuarta posición

- 4; 8; 9; **12**; 22; 25; 202



¿Cómo calculamos la mediana cuando las muestras son pares?

8; 9; 12; 22; 25; 38, 202 → 4; 8; 9; **12; 22**; 25; 38, 202 → $12+22/2$

Mediana = 17!

A diferencia de la MEDIA, la MEDIANA es un estadístico resistente a los valores extremos de una muestra

Ej.: 4, 8, 9, 12, 15, 25, 202

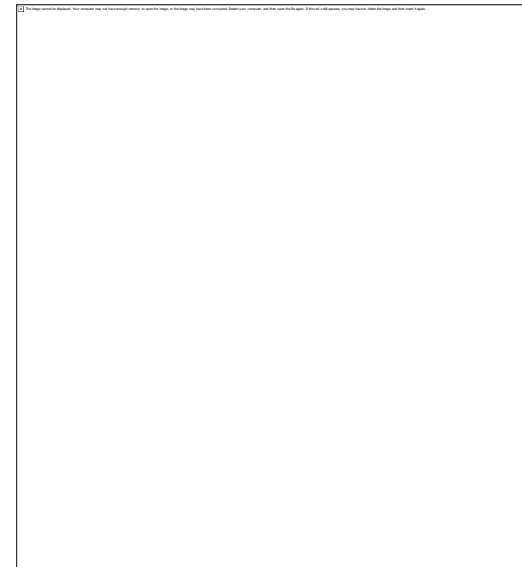
Mediana 12

4, 8, 9, 12, 15, 25, 30

Mediana 12

4, 8, 9, 12, 15, 25, 2500

Mediana 12



MODA

- Es el valor con mayor frecuencia en una distribución de datos

1 2 2 3 3 4 2 1 2 3

1 2 2 3 3 4 1 1 2 3 5 6 6 6 4 5 5 5 4 4 4 8 8 9 9 1 1 2 2 5 4 4 6 6

Unimodal



Bimodal



Multimodal



MEDIDAS DE VARIABILIDAD

Debido a la variabilidad en los datos (pesos, alturas, número de individuos, etc.), no es suficiente con brindar solamente el valor de tendencia central

Se necesita además un valor que brinde información sobre la variabilidad de las observaciones

Entre las medidas de variabilidad tenemos:

- Varianza
- Desvío estándar
- Error estándar
- Coeficiente de variación

EMPLO: queremos calcular la varianza en las mismas muestras de trampas de hormigas



ómo calcular la varianza de la siguiente muestra:

7; 35; 63; 63; 41

rimero debemos hallar la media de la muestra

uanto es? → 51,8

uego realizamos la suma de los cuadrados:

$$(57-51,8)^2+(35-51,8)^2+(63-51,8)^2+(63-51,8)^2+(41-51,8)^2$$

$$(5,2)^2+(-16,8)^2+(11,2)^2+(11,2)^2+(-10,8)^2=676,8$$

$$s^2=676,8/4$$

$$s^2=169,2$$

DESVÍO ESTÁNDAR

El desvío estándar es la medida de variabilidad de datos más utilizada

Se calcula a partir de todas las observaciones de una variable particular

Es la raíz cuadrada de la varianza

Si se calcula a partir de toda la población:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

Si se calcula a partir de una muestra de la población:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

ERROR ESTÁNDAR

Es un estadístico relacionado al desvío estándar

Es el desvío estándar dividido entre la raíz cuadrada del número de observaciones de la muestra (menos 1)

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

PRESENTACIÓN DE LOS DATOS

- En un trabajo científico, los resultados pueden ser presentados en forma escrita, en forma de tabla o gráficos
- La forma en que se presentan los datos dependerá de la **estrategia de difusión** que tenga el investigador
- Hay que tener en cuenta que no todos los gráficos son iguales, por lo que el gráfico que se elija dependerá del tipo de dato que se tenga. Ej.: proporciones (gráfico de torta); comparaciones (gráfico de barras); correlaciones (gráficos de puntos)
- Lo que se debe evitar es presentar la misma información en forma de tabla y gráfico (repetir información) en un mismo trabajo

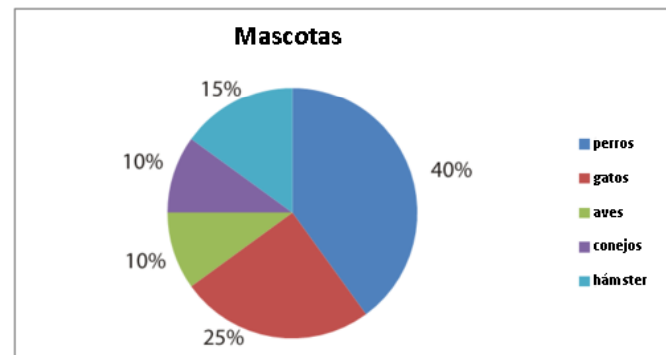
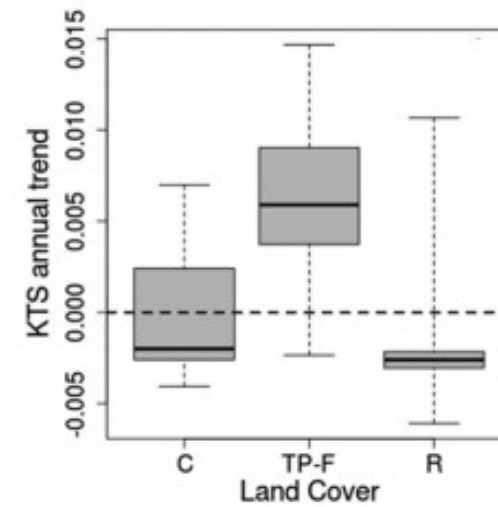
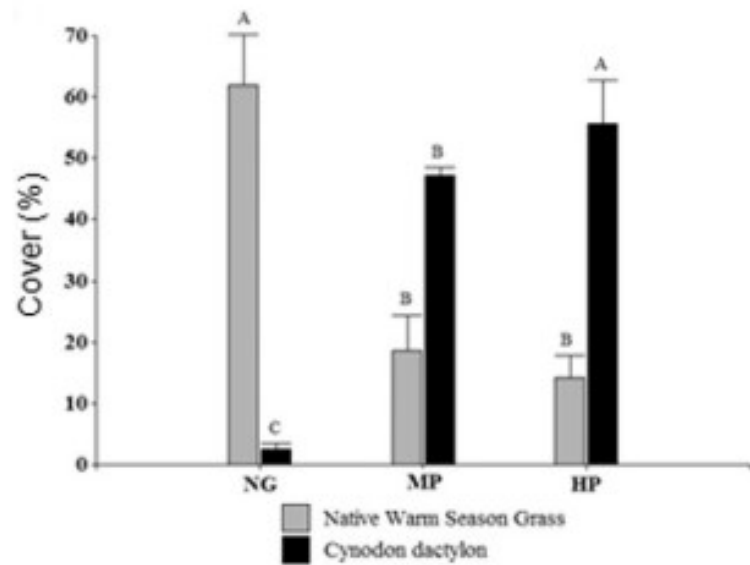


Table 1. Sampling periods covered between the date of placement and removal of ingrowth cores, the duration in days and the mean temperature of each period.

| Período de muestreo | Fecha de muestreos | Duración (días) | Temperatura (°C) |
|---------------------|---------------------|-----------------|------------------|
| Otoño-invierno | 22/04/09 - 30/07/09 | 99 | 16.7 |
| Invierno-primavera | 30/07/09 - 27/10/09 | 89 | 13.0 |
| Primavera-verano | 27/10/09 - 30/01/10 | 95 | 21.3 |
| Verano-otoño | 30/01/10 - 12/05/10 | 102 | 19.3 |

ANÁLISIS ESTADÍSTICO

| Análisis no-paramétrico | Análisis paramétrico |
|--|---|
| Pueden ser usados con observaciones crudas o con observaciones convertidas en rangos | Son usados sólo con observaciones crudas |
| Pueden ser usados con observaciones en escala nominal, ordinal o de intervalo | Generalmente restringidos a observaciones en escala de intervalo y cociente |
| Compara medianas | Compara medias y varianzas |
| Distribución libre, no requieren que los datos tengan distribución normal ni que tengan varianza homogénea | Requieren distribución normal y varianza homogénea de los datos |
| Son aptos para datos de conteo | Conteos deben ser transformados |
| Son aptos para datos “derivados”, ej.: proporciones, índices | Datos derivados deben ser primero transformados |

Análisis paramétrico

Prueba t

Compara la **media** de dos muestras

Anova

Compara la **media** de tres o más muestras

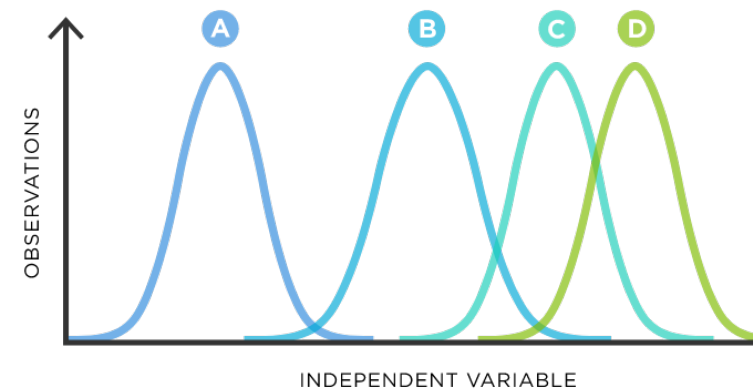
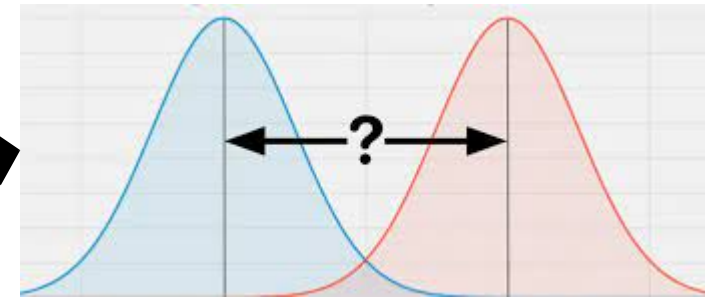
Análisis no-paramétrico

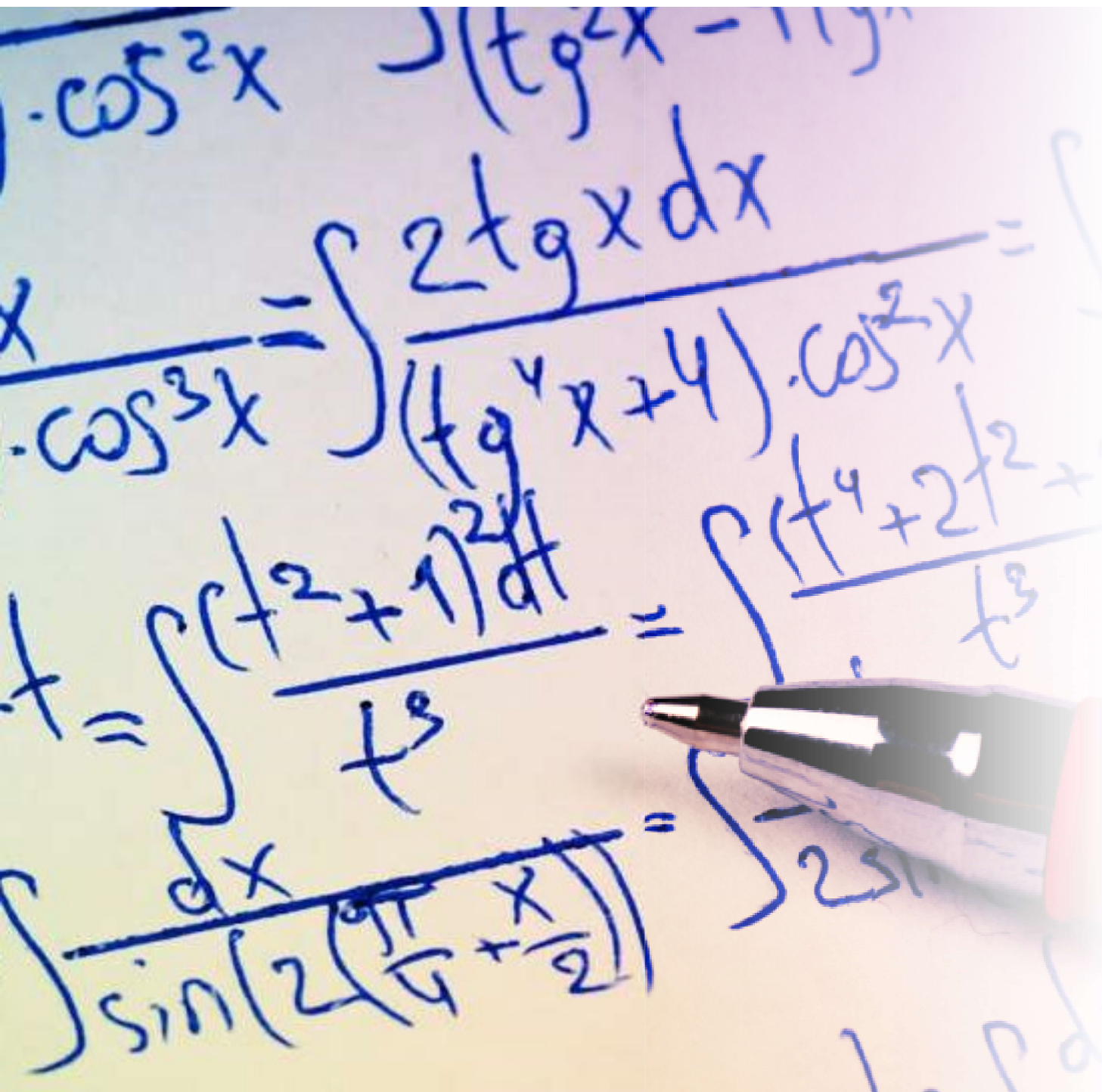
Análisis U de Mann-Whitney

Compara la **mediana** de dos muestras independientes

Prueba de Kruskal-Wallis

Compara las **medianas** de tres o más muestras





The image shows a piece of paper with handwritten mathematical work in blue ink. At the top, there are several lines of algebraic expressions involving trigonometric functions and powers of $\tan x$. Below these, there are three integral equations:

$$\int \frac{2 \tan x dx}{(\tan^4 x + 4) \cos^2 x} = \int \frac{(t^2 + 1) dt}{t^3} = \int \frac{(t^4 + 2t^2 + 1) dt}{t^3}$$
$$\int \frac{dx}{\sin\left(2\left(\frac{\pi}{4} + \frac{x}{2}\right)\right)} = \int \frac{dx}{2 \cos x}$$

A silver and black pen is lying horizontally across the bottom right of the paper, partially overlapping the handwritten work.

Con lo que vimos hasta ahora, hagamos un pequeño ejercicio!

| Trampas | Hormigas en Pastizal | Hormigas en Monte |
|---------|----------------------|-------------------|
| 1 | 125 | 25 |
| 2 | 180 | 30 |
| 3 | 155 | 26 |
| 4 | 139 | 32 |
| 5 | 208 | 32 |
| 6 | 130 | 32 |
| 7 | 127 | 32 |
| 8 | 198 | 33 |
| 9 | 204 | 39 |
| 10 | 225 | 36 |
| 11 | 136 | 25 |
| 12 | 175 | 28 |
| 13 | 152 | 27 |
| 14 | 122 | 26 |
| 15 | 138 | 22 |

1. Calcular los medidas de tendencia central y variabilidad que considere
2. Cómo presentaría los resultados?

| | | |
|-----|-------------|------------|
| POM | 160.9333333 | 29.6666666 |
| MED | 152 | 30 |
| VAR | 1117.928889 | 19.9555555 |
| DE | 34.60897214 | 4.6239541 |
| EE | 9.249636875 | 1.2358037 |