

## Resultados Trabajo Práctico N° 1

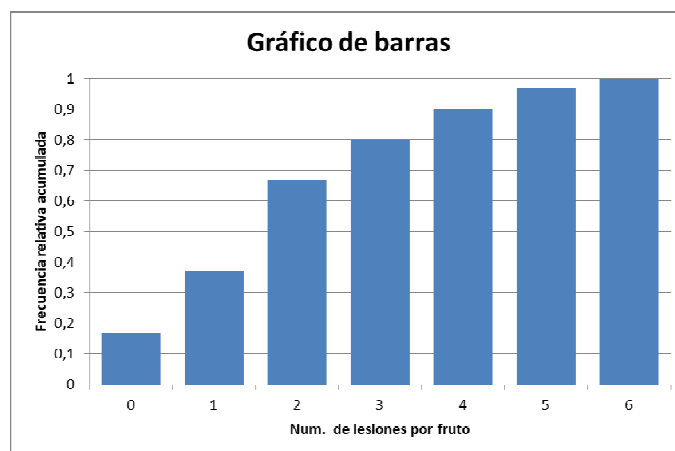
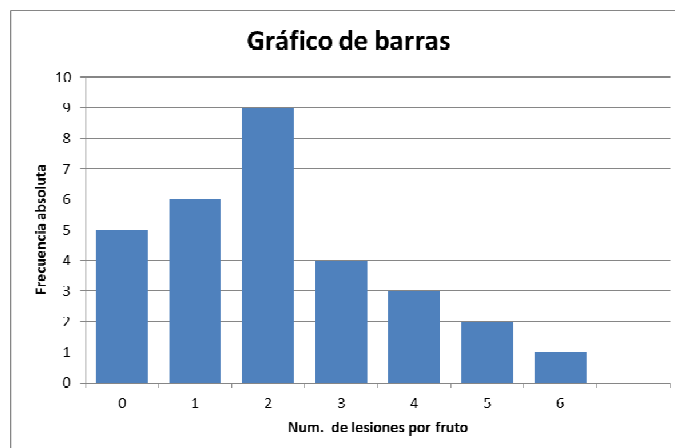
### Ejercicio 1.1

- a) **Variable:** Número de lesiones por fruto  
**Tipo:** Cuantitativa, discreta  
**Recorrido:**  $R_x \{0, 1, 2, 3, 4, 5, 6\}$

b)

$X_i$	$f_i$	$h_i$	$F_i$	$H_i$
0	5	0,17	5	0,17
1	6	0,20	11	0,37
2	9	0,30	20	0,67
3	4	0,13	24	0,80
4	3	0,10	27	0,90
5	2	0,07	29	0,97
6	1	0,03	30	1
$\Sigma$	<b>30</b>	<b>1</b>		

c)

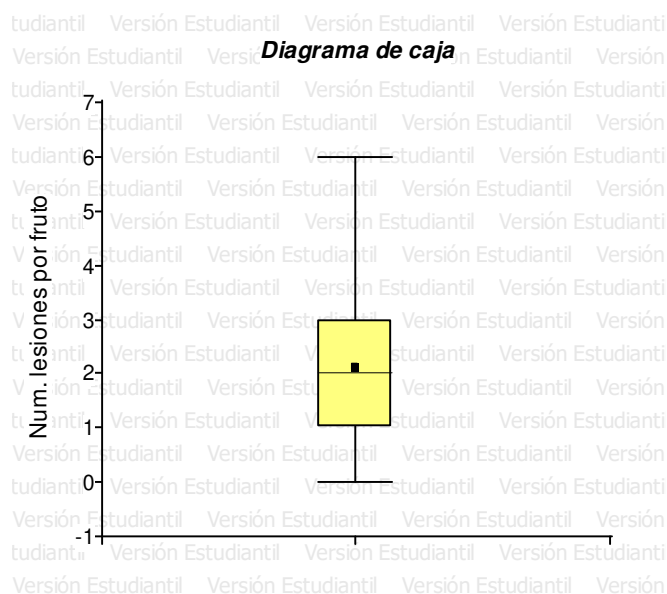


- d)  $f_{x=2} = 9$  ; indica que hay 9 frutos con 2 lesiones por fruto.  
 $h_{x=5\%} = 7\%$  ; indica que el 7% de los frutos tienen 5 lesiones.  
 $F_{x=4} = 27$  ; indica que hay 27 frutos que tienen 4 o menos lesiones

e)  $\bar{x} = 2,133$

f)  $S^2 = 2,602$   
 $S = 1,613$

e)



A través del diagrama de cajas se puede observar que el valor mínimo de lesiones por fruto es cero y el valor máximo es de 6 lesiones por fruto. El primer cuartil equivale a 1 lesión por fruto es decir que hay un 25 % de frutos con una o menos lesiones y un 75% de frutos con más de una lesión. Por otro lado el tercer cuartil es de 3 lesiones por fruto indicando que hay un 75% de frutos con 3 o menos lesiones y un 25% con más de tres lesiones.

La mediana equivale a 2 lesiones por fruto es decir que dos lesiones por fruto es el valor que divide a la serie de datos a la mitad.

El rango tiene una amplitud de 6 unidades, siendo esta la variabilidad máxima de los datos, mientras que el rango intercuartilico (diferencia entre el tercer cuartil y el primer cuartil) es de 2 unidades, por lo tanto el 50% de los datos centrales están entre 1 y 3 lesiones.

El desplazamiento de la caja hacia los valores menores de la variable indica una tendencia de la distribución de los datos a una distribución Asimétrica positiva. Esta misma tendencia se puede observar por la diferencia entre la media y la mediana, en este caso la media es mayor a la mediana.

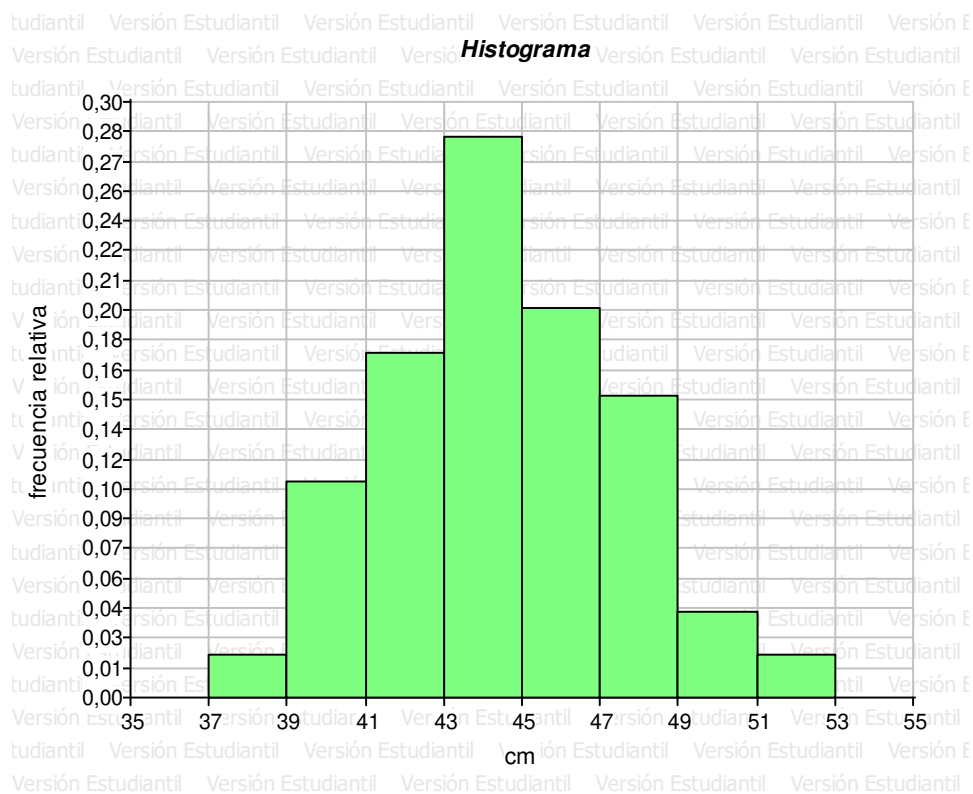
## Ejercicio 1.2

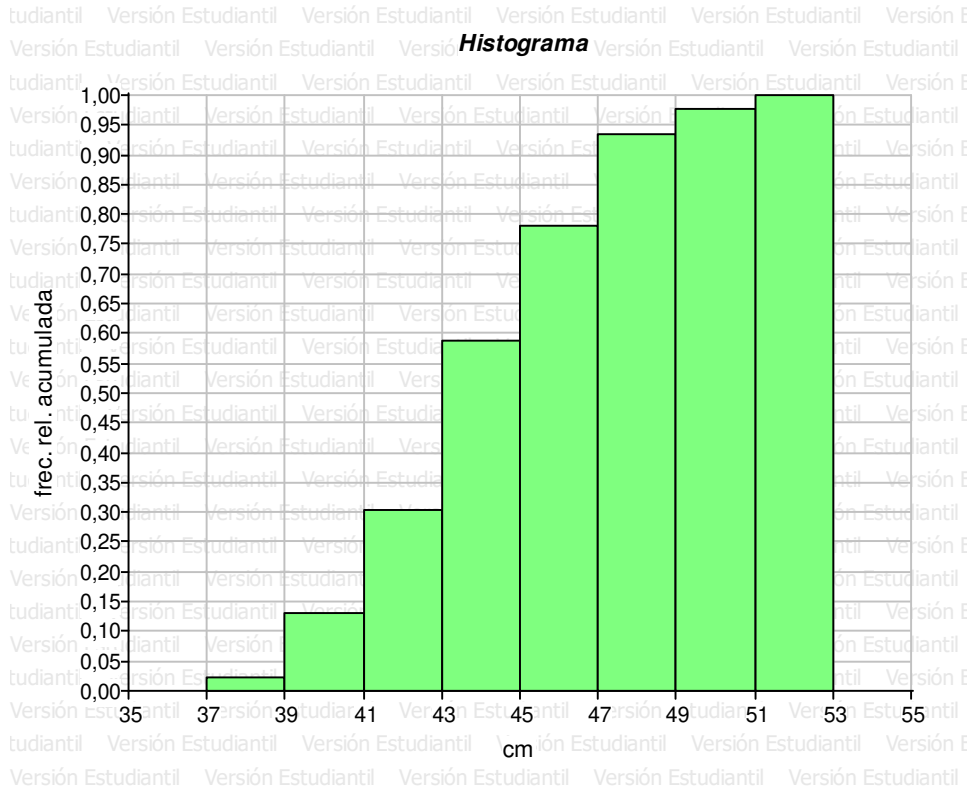
- a) **Variable:** circunferencia de árboles en cm  
**Tipo:** Variable cuantitativa continua  
**Recorrido:**  $R_x = \{37,4 ; 51,20\}$

b)

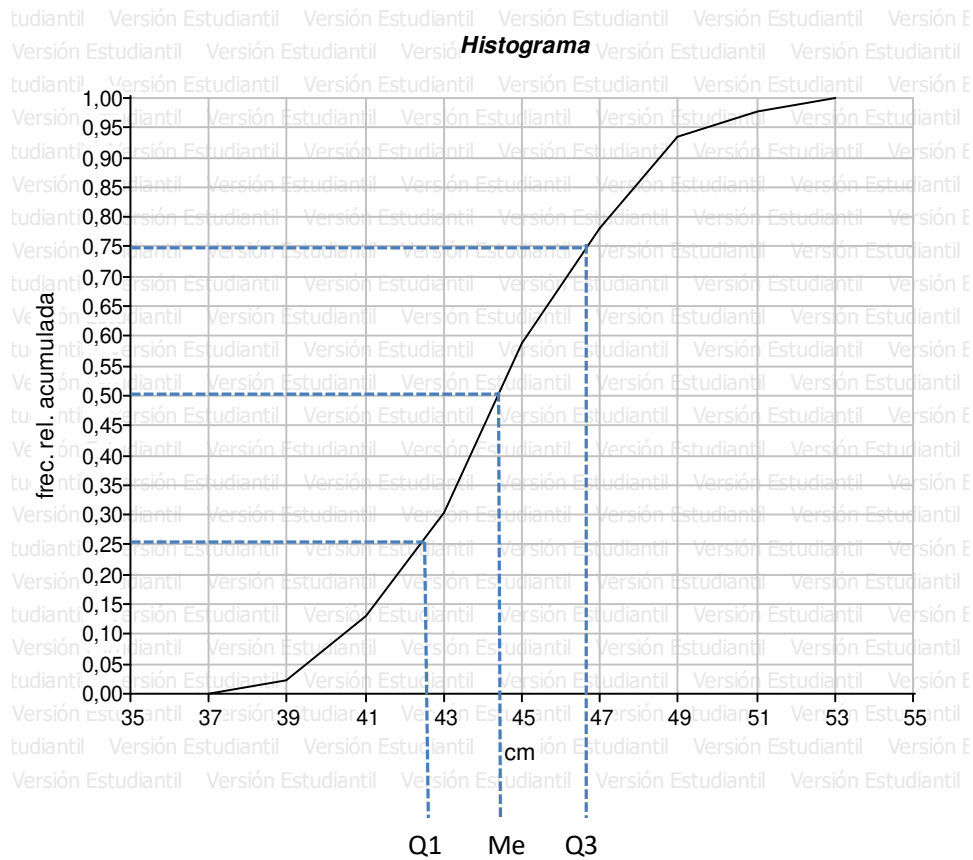
Clase	Límite Inferior	Límite Superior	$X_i$	$f_i$	$h_i$	$F_i$	$H_i$
1	[37	39]	38	1	0,02	1	0,02
2	(39	41]	40	5	0,11	6	0,13
3	(41	43]	42	8	0,17	14	0,3
4	(43	45]	44	13	0,28	27	0,59
5	(45	47]	46	9	0,2	36	0,78
6	(47	49]	48	7	0,15	43	0,93
7	(49	51]	50	2	0,04	45	0,98
8	(51	53]	52	1	0,02	46	1

c)





d)

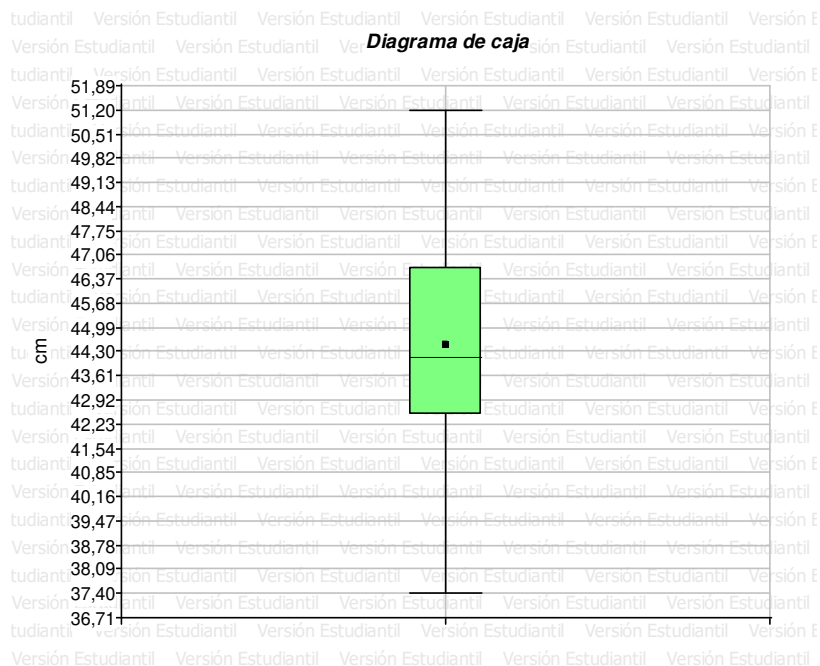


- e)  $\bar{x}$  datos originales = 44,49  
 $\bar{x}$  datos agrupados = 44,52

Se observa que los valores, si bien son muy cercanos, no coinciden y ello se debe a que el cálculo de los estimadores a partir de datos agrupados para variables continuas pierde algo de precisión, ya que estamos forzando a una variable continua a agruparla en clases. Esta pérdida de precisión no se da en variables discretas.

- f)  $S = 3,05$   
 $CV = 0,068$  o 6,8%

g)



A través del diagrama de cajas se puede observar que la circunferencia mínima de los árboles es 37,4 y la máxima 51,20 cm. El primer cuartil con un valor aproximado de 42,5 cm indica que hay un 25 % de árboles con una circunferencia menor o igual a ese valor y un 75% de árboles con una circunferencia mayor a 42,5 cm. Por otro lado, el tercer cuartil con un valor aproximado de 46,7 cm indica que hay un 75 % de árboles con una circunferencia menor o igual a ese valor y un 25% de árboles con una circunferencia mayor a 46,7 cm.

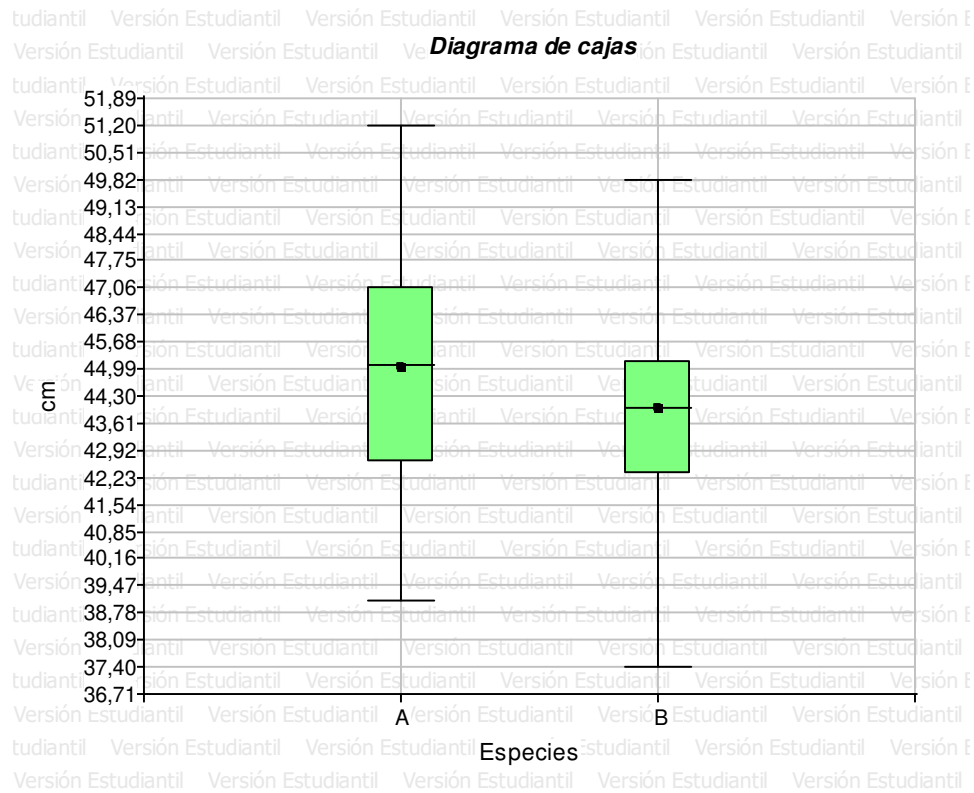
La mediana con un valor aproximado de 44,5 cm es el valor de la circunferencia que divide a la serie de árboles a la mitad.

El rango tiene una amplitud de 13,8 cm, siendo esta la variabilidad máxima de los datos, mientras el rango intercuartílico (diferencia entre el tercer cuartil y el primer cuartil) es de 4,2 cm, por lo tanto, el 50% de los datos centrales están concentrados en 4,2 cm.

La posición de la caja en la parte central del gráfico nos indica una tendencia de la distribución de los datos a una distribución Simétrica. Si bien en el gráfico se ve una diferencia entre la media y la mediana, la misma está magnificada por la escala del eje Y, ya que en valor absoluto la diferencia entre ambas es mínima.

### Ejercicio 1.3

a)



Como puede apreciarse en el gráfico la Especie A con respecto a la especie B presenta valores de circunferencia mayores, lo cual se puede apreciar en que tanto la circunferencia mínima como la máxima son más grandes. Los valores tanto de la media como de la mediana son mayores en la especie A indicando que son árboles de mayor circunferencia media.

Con respecto a la dispersión de los datos el rango no manifiesta grandes diferencias de variabilidad, en cambio el rango intercuartílico (largo de la caja) es mucho más grande en la especie A que en la B, lo cual nos indica que el 50% central de los árboles de la especie A presentan una mayor variabilidad en las circunferencias.

Respecto a la Asimetría no hay grandes diferencias y podríamos estimar que ambas especies tienen una distribución bastante simétrica.

b) Para resolver este punto se requiere construir las tablas de frecuencias para ambas especies por separado

Especie A	Clase	LI	LS	$X_i$	$f_i$	$h_i$	$F_i$	$H_i$
A	1	[ 37,00	39,00 ]	38,00	0	0,00	0	0,00
A	2	( 39,00	41,00 ]	40,00	3	0,15	3	0,15
A	3	( 41,00	43,00 ]	42,00	3	0,15	6	0,30
A	4	( 43,00	45,00 ]	44,00	4	0,20	10	0,50
A	5	( 45,00	47,00 ]	46,00	4	0,20	14	0,70
A	6	( 47,00	49,00 ]	48,00	4	0,20	18	0,90
A	7	( 49,00	51,00 ]	50,00	1	0,05	19	0,95
A	8	( 51,00	53,00 ]	52,00	1	0,05	20	1,00

Especie B	Clase	LI	LS	$X_i$	$f_i$	$h_i$	$F_i$	$H_i$
B	1	[ 37,00	39,00 ]	38,00	1	0,04	1	0,04
B	2	( 39,00	41,00 ]	40,00	2	0,08	3	0,12
B	3	( 41,00	43,00 ]	42,00	5	0,20	8	0,32
B	4	( 43,00	45,00 ]	44,00	9	0,36	17	0,68
B	5	( 45,00	47,00 ]	46,00	4	0,16	21	0,84
B	6	( 47,00	49,00 ]	48,00	3	0,12	24	0,96
B	7	( 49,00	51,00 ]	50,00	1	0,04	25	1,00
B	8	( 51,00	53,00 ]	52,00	0	0,00	25	1,00

En la especie A el porcentaje de árboles entre 43 y 47 cm es del 0,4 o sea del 40%, mientras en la especie B es del 0,52 o sea del 52%.

#### Ejercicio 1.4

a) Resuelto en página 5 de la “Guía Uso InfoStat”

b) Los estadísticos que más varían entre las localidades 2 y 3 son los de variabilidad y forma y en menor medida los de posición.

Esas diferencias están indicando que las plántulas provenientes de la localidad 2 son más homogéneas es decir presentan menor variabilidad que las provenientes de la localidad 3 y tienen una distribución más Simétrica mientras los crecimientos en la localidad 3 tienden a una distribución Asimétrica positiva.

c) Gráficos en páginas 7 a 9 de la “Guía Uso InfoStat”

En los gráficos de caja la mayor variabilidad de la localidad 3 se ve en el mayor rango y rango intercuartilico mientras la tendencia de los crecimientos a una distribución Asimétrica positiva se puede apreciar por el desplazamiento de la caja hacia los valores menores de crecimiento y al mayor valor de la media respecto a la mediana.

En los histogramas la mayor variabilidad de la localidad 3 se ve en la mayor amplitud del gráfico ya que tiene crecimientos desde la clase que tiene como límite inferior 11 mm/día hasta la clase que tiene como límite superior 23 mm/día, mientras estos valores para la localidad 2 son 15 y 21 respectivamente.

La tendencia a la Asimetría positiva se puede apreciar por la mayor concentración de datos en las clases de crecimiento menor respecto las de mayor crecimiento, mientras en la localidad 2 la distribución es más simétrica (en forma de campana).

d) Cálculos y gráficos páginas 5 a 9 de la “Guía Uso InfoStat”

e) Si hubiera que elegir una localidad como fuente de semillas para una siembra siendo mi objetivo maximizar el rendimiento, seleccionaría las semillas de la localidad 2 porque son las de mayor crecimiento y menor variabilidad.

f) Las dos muestras con crecimiento medio más alto son las semillas provenientes de la localidad 1 y 2.

**g)** Al haber un valor atípico (alejado) en la muestra 1 (valor 49), deberíamos utilizar como estadístico de posición la mediana (en lugar de la media) y como estadístico de dispersión el rango intercuartilico (en lugar del rango o del desvío estándar) ya que ambos estadísticos son resistentes/robustos a la presencia de estos valores.

La presencia de valores atípicos suele distorsionar los gráficos y por ello no son buenas herramientas para su análisis.

**h)** Las dos localidades con mayor probabilidad de originar plantas establecidas cuando el crecimiento mínimo requerido es mayor a 13 mm/día, serán las localidades 2 y 3 porque son las que tienen mayor cantidad de ejemplares con crecimiento mayor a 13 mmm/día. En cambio, para el caso de 17 mm/día serán las localidades 2 y 4.

**i)** Las mayores diferencias entre las localidades 3 y 4 están dadas por que la localidad 3 es más variable y tiene una tendencia en la distribución de los datos Asimétrica positiva, mientras la localidad 4 es menos variable y tiene una tendencia en la forma de la distribución de los datos Asimétrica negativa.

En los gráficos de caja la mayor variabilidad de la localidad 3 se ve en el mayor rango y rango intercuartilico mientras la tendencia de los crecimientos a una distribución Asimétrica positiva se puede apreciar por el desplazamiento de la caja hacia los valores menores de crecimiento y al mayor valor de la media respecto a la mediana contrario a lo que ocurre en la localidad 4 donde la tendencia de los datos a una distribución Asimétrica negativa se puede apreciar por el desplazamiento de la caja hacia los valores más alto de crecimiento.

En los histogramas la mayor variabilidad de la localidad 3 se ve en la mayor amplitud del gráfico ya que tiene crecimientos desde la clase que tiene como límite inferior 11 mm/día hasta la clase que tiene como límite superior 23 mm/día, mientras estos valores para la localidad 4 son 11 y 18 respectivamente.

La tendencia a la Asimetría positiva en la localidad 3 se puede apreciar por la mayor concentración de datos en las clases de crecimiento menor respecto las de mayor crecimiento, mientras en la localidad 4 la mayor concentración de datos se da en las clases con mayor crecimiento.

**j)** Como se requiere que el crecimiento de las plántulas utilizadas sea homogéneo utilizaría las plántulas provenientes de la localidad 2, ya que los estadísticos de dispersión me indican que es la muestra de menor variabilidad.

**k)**

Datos	Localidad 2	Localidad 2 (nueva muestra)
n	20	30
Media	17,57	15
Desvío estándar	0,94	0,946
Error estándar	<b>0,21</b>	<b>¿?</b>

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0,946}{\sqrt{30}} = 0,17 \text{ mm / día}$$



El error estándar de la nueva muestra de la localidad 2 es menor y su diferencia se debe a la mayor cantidad de semillas que se utilizaron para su estimación.

**I)** Se espera la mayor diferencia respecto de la media de la muestra anterior para la localidad 3 y la menor diferencia para la localidad 2. Dicha conclusión se basa en la comparación de los errores estándar entre las localidades, siendo la localidad 3 la de mayor error estándar y la localidad 2 la de menor error estándar. (No se contempla en la comparación la localidad 1 por tener un valor atípico).