



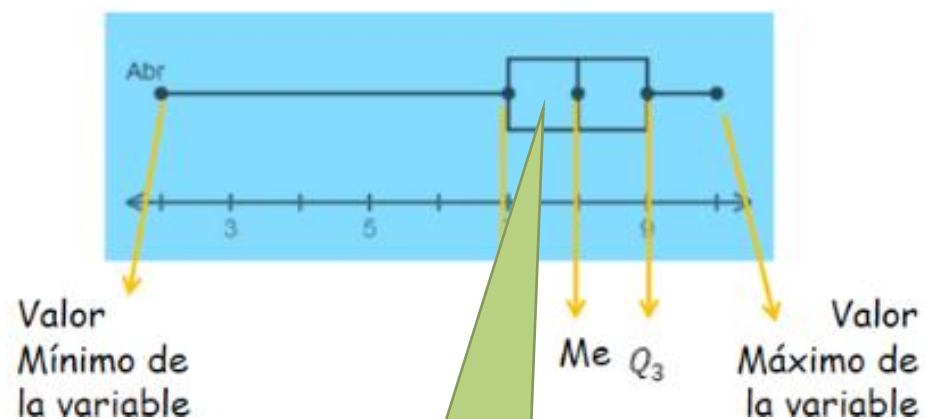
# ***BOXPLOT, Diagrama de caja***

$\pi$

***Jeanette Badilla – Ingrid Ríos***

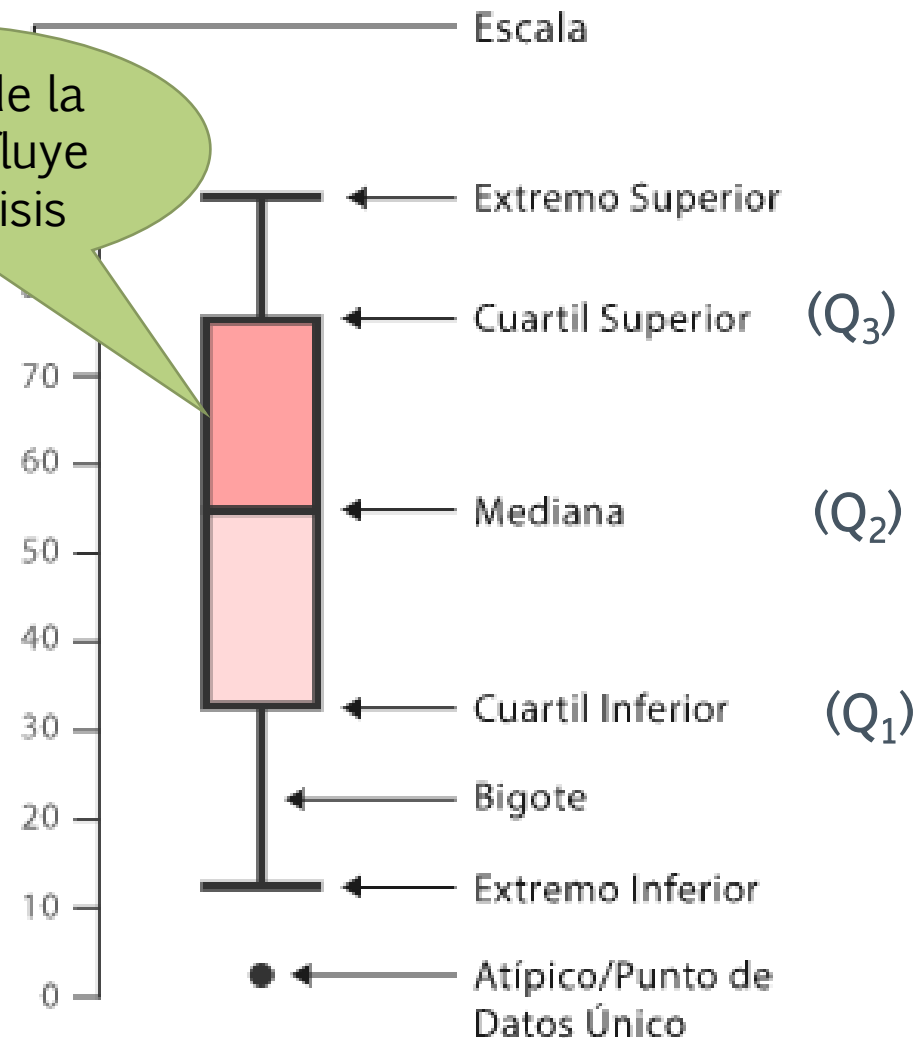
- › Un **boxplot**, **diagrama de caja o cajón con bigote** corresponde a una representación gráfica complementaria a un histograma, basada en la información entregada por la **mediana, cuartiles y valores mínimos y máximos** de un conjunto de datos.
- › Estos 5 valores que dan origen a un **bloxplot** suelen denominarse **las 5 medidas de resumen** de un conjunto de observaciones.
- › Las líneas que se extienden paralelas a las cajas se conocen como **«bigotes»**, y se usan para indicar variabilidad fuera de los cuartiles superior e inferior. Los **valores atípicos** se representan a veces como puntos individuales que están en línea con los bigotes. Los diagramas de cajas y bigotes se pueden **dibujar vertical u horizontalmente**.
- › Normalmente utilizado en estadísticas descriptivas, los gráficos de cajas y bigotes son una excelente forma de examinar rápidamente uno o más conjuntos de datos gráficamente. Aunque parezcan primitivos en comparación con un **Histograma**, tienen la ventaja de ocupar menos espacio, lo cual es útil cuando se comparan distribuciones entre muchos grupos o conjuntos de datos.

# Diagramas de cajas y bigotes **vertical u horizontal**.



El alto de la caja no influye en el análisis

El ancho de la caja no influye en el análisis

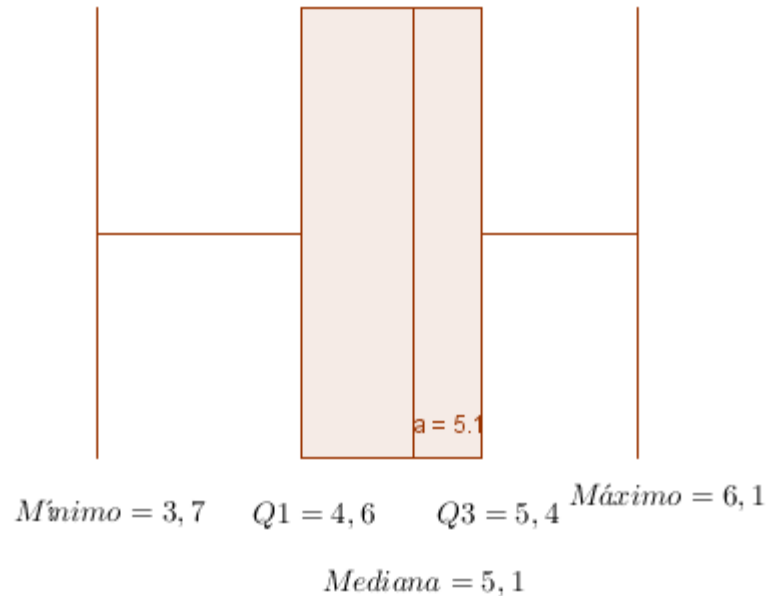


## Construcción e interpretación de un bloxplot

- › **Ejemplo:** Las siguientes son las notas de una primera evaluación de los alumnos de un curso, el segundo medio A, cuyas 5 medidas de resumen se muestran en la siguiente tabla.

Estadísticos	Valor
Mínimo	3,7
Primer cuartil	4,6
Mediana	5,1
Tercer cuartil	5,4
Máximo	6,1

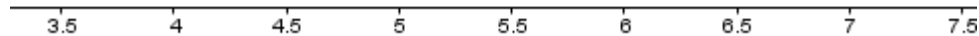
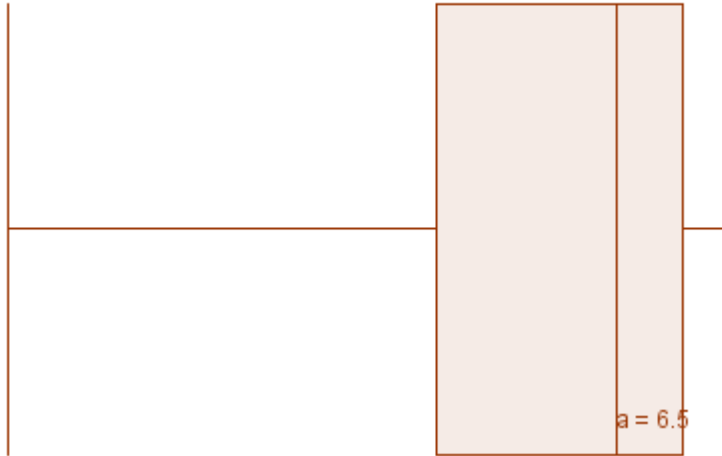
## Notas segundo medio A



Al mirar el eje de coordenadas podemos extraer la siguiente información:

- La mediana es ligeramente más cercana al tercer cuartil que al primero. Esto dice que los valores de las observaciones está ligeramente más concentrada en la zona superior del cajón.
- Las amplitudes de los «bigotes» superior e inferior son similares. Esto nos habla de cierta similitud entre los extremos de la distribución.
- La altura del cajón es ligeramente menor a la suma de las longitudes de los «bigotes». Esto dice que los datos están levemente más concentrados en el centro.

## Notas en la segunda evaluación del segundo medio A



Al mirar el eje de coordenadas podemos extraer la siguiente información:

- La mediana no está en el centro del cajón. Esto indica que la distribución de las observaciones no es simétrica.
- La mediana está bastante más cercana al tercer cuartil que al primero- Esto dice que los valores de las observaciones se encuentran más concentradas en la zona superior del cajón, es decir, en los valores altos de las notas.
- Las amplitudes de los «bigotes» superior e inferior son bastante diferentes. Esto indica que los extremos de la distribución son distintos; hay notas bajas que se alejan del grueso de los datos.
- La distribución tiene, marcadamente, una mayor concentración en las notas altas que en las bajas, lo que se observa en las distancias entre la mediana y la nota máxima es bastante menor que la distancia entre la misma y la nota mínima.

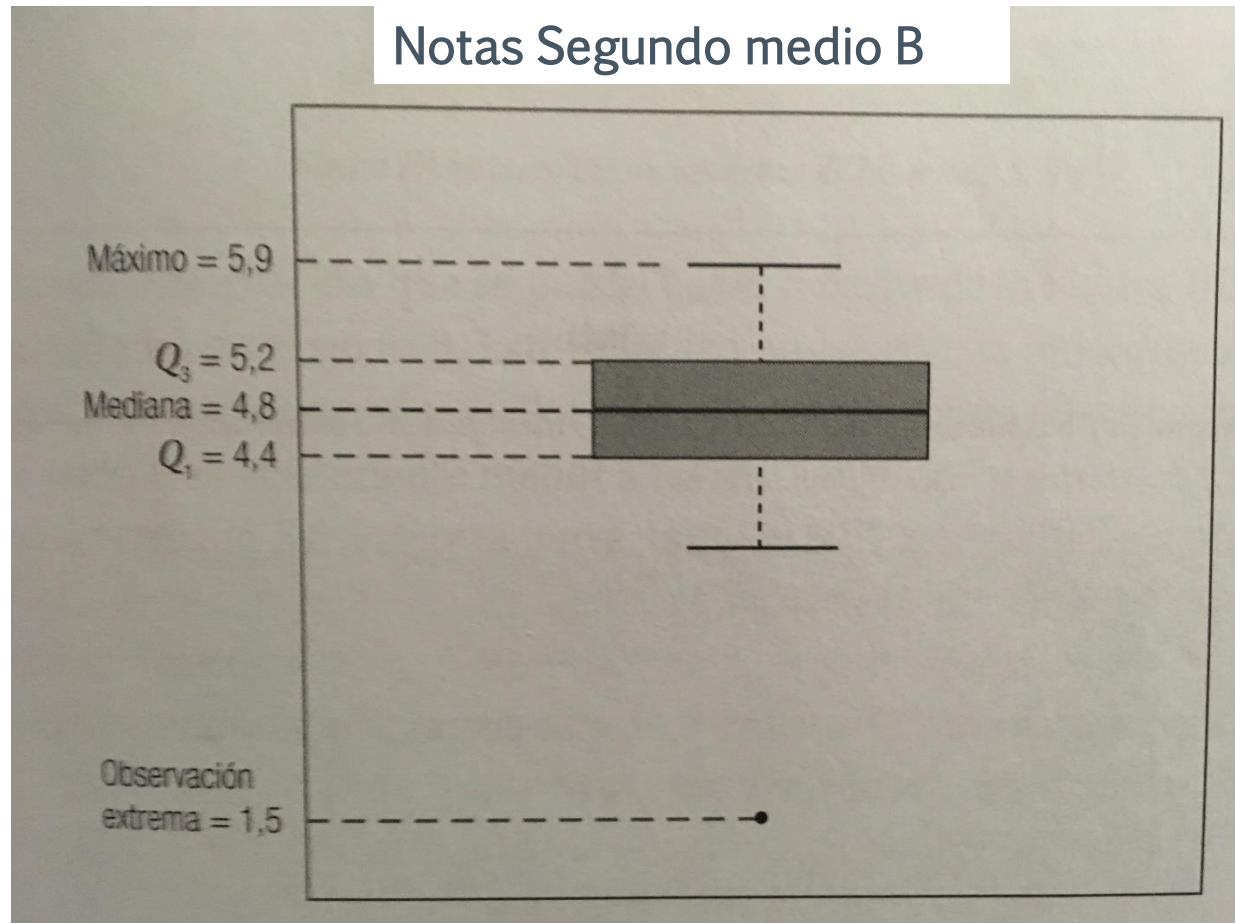
## Tratamiento de observaciones extremas en un bloxplot

( observaciones que se alejan del grueso de los datos)

- › **Ejemplo:** Las siguientes son las notas obtenidas por el curso paralelo, el segundo medio B, en la primera evaluación. Las 5 medidas de resumen se muestran en la siguiente tabla.

Estadísticos	Valor
Mínimo	1,5
Primer cuartil	4,4
Mediana	4,8
Tercer cuartil	5,2
Máximo	5,9

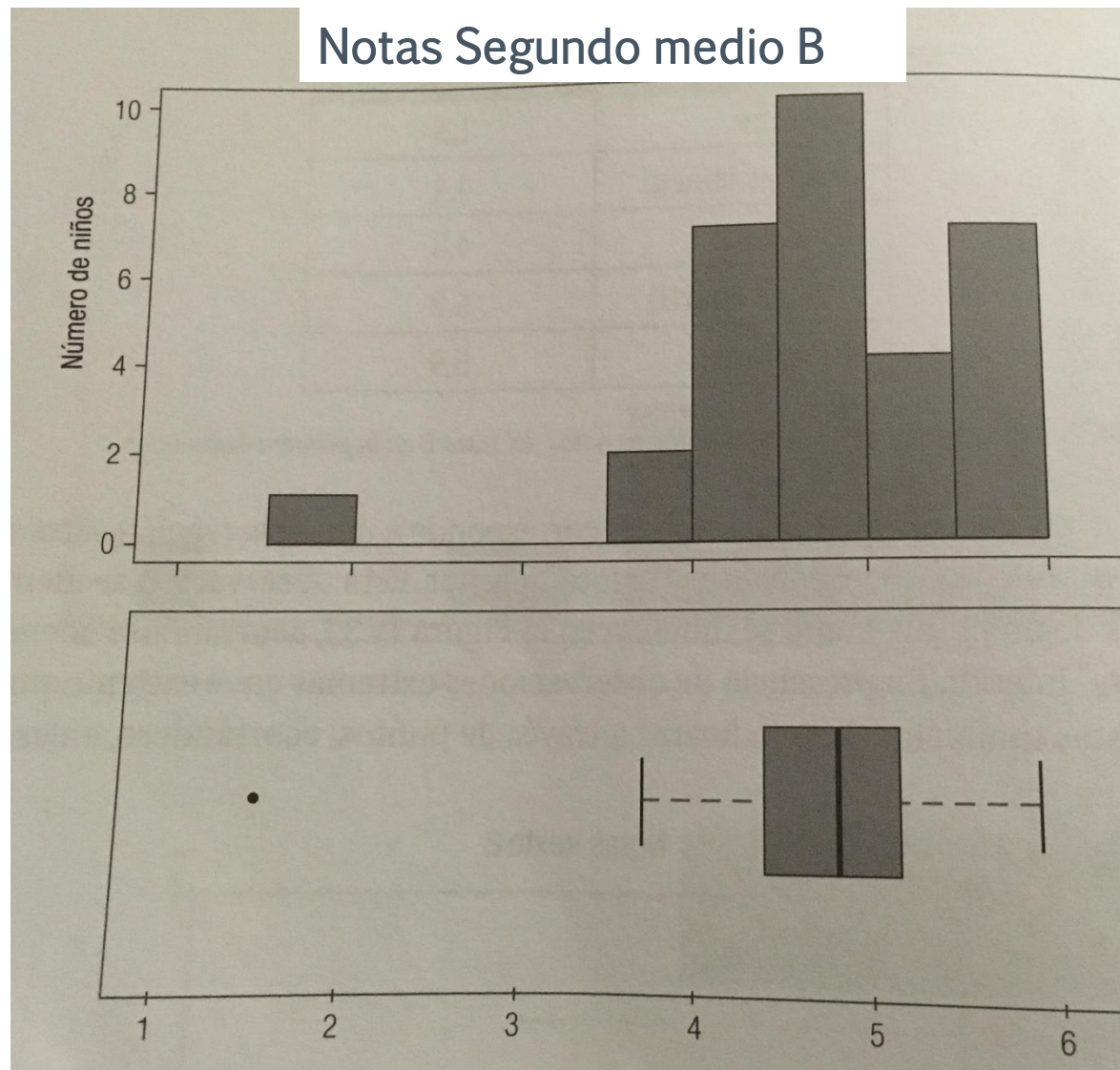
## Notas Segundo medio B



Al mirar el eje de coordenadas podemos extraer la siguiente información:

- El «bigote» inferior no llega hasta la observación mínima, debido a que esta es considerada una observación extrema. En su lugar, este «bigote» se extiende hasta la menor observación, cuyo valor es mayor o igual a una cota establecida.
- La observación identificada como observación extrema se muestra a través de un punto en la parte baja de la figura, a la altura de su valor, 1,5.

### Notas Segundo medio B



# En resumen

- › Un bloxplot, diagrama de caja o cajón con bigotes corresponde a una representación gráfica de la distribución de las observaciones.
- › En un bloxplot representado verticalmente, se observa una caja cuyo límite superior corresponde al *tercer cuartil* y el inferior al *primer cuartil*. La mediana es identificada dentro del cajón a través de una línea horizontal.
- › En general, los «bigotes» se extienden hasta los valores *mínimo* y *máximo*.
- › Una mayor amplitud de la caja implica mayor variabilidad.
- › Una mayor longitud de los «bigotes» implica mayor variabilidad.
- › En caso de existir observaciones indicadas a través de puntos más allá de los extremos de los «bigotes», corresponden a observaciones que pueden ser consideradas ***observaciones extremas***.