

# Interpretare i dati ...

Disegnare e confrontare  
i diagrammi Box and Whisker  
"scatole con baffi"

# Obiettivi

- Calcolare i quartili e disegnare i diagrammi
- Interpretare i diagrammi e utilizzarli per confrontare le raccolte di dati
- Usare la terminologia corretta e il campo di variazione interquartile per identificare i dati marginali da escludere

# Una lista di dati

- Il peso (KG) di 15 bambini:

37, 42, 31, 35, 48, 29, 50, 36, 44, 28,  
63, 35, 41, 52, 43

*È difficile CAPIRE dalla lista di numeri com' é questo gruppo di bambini ...*

- Il peso (KG) di 15 bambini:  
28, 29, 31, 35, 35, 36, 37, 41, 42, 43, 44,  
48, 50, 52, 63
- Minimo = 28KG
- Massimo = 63KG
- Escursione = 35KG (campo di variazione)
- Moda = 35KG
- Mediana = 41KG
- Media = 40.9KG

# Stem and leaf (gambo e foglia)

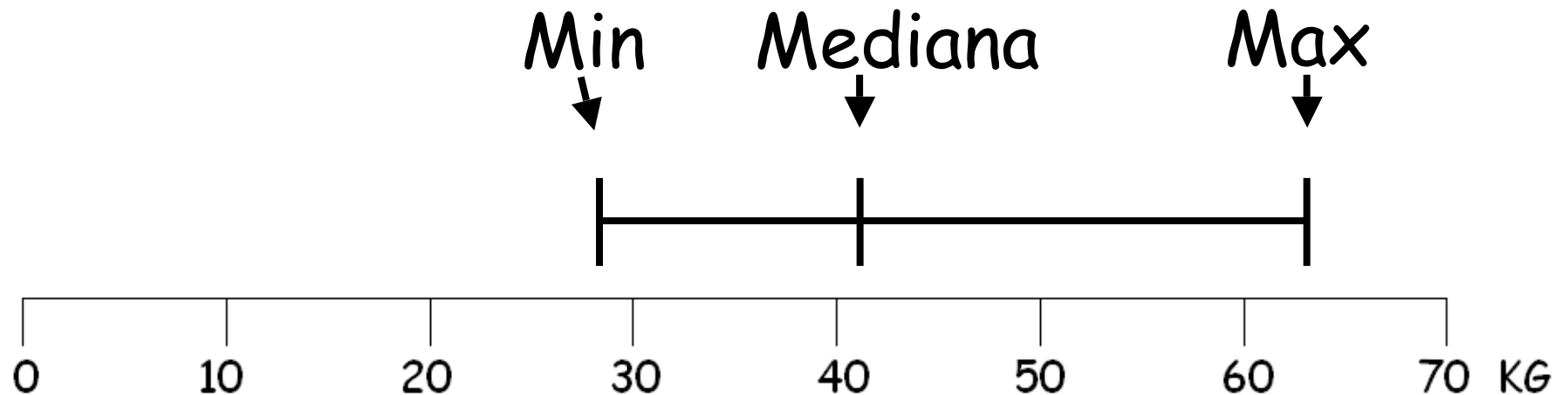
- 28, 29, 31, 35, 35, 36, 37, 41, 42, 43, 44, 48, 50, 52, 63

- |   |  |   |   |   |   |   |
|---|--|---|---|---|---|---|
| 2 |  | 8 | 9 |   |   |   |
| 3 |  | 1 | 5 | 5 | 6 | 7 |
| 4 |  | 1 | 2 | 3 | 4 | 8 |
| 5 |  | 0 | 2 |   |   |   |
| 6 |  | 3 |   |   |   |   |

Lettura: 2 | 9 significa 29

# Un altro utile diagramma

- Un diagramma che rappresenta:  
min (28KG),  
max (63KG),  
mediana (41KG) ...



# Mediana

- $1/2(n + 1)^{\text{mo}}$  posto della lista (ordinata)

28, 29, 31, 35, 35, 36, 37, **41**,  
42, 43, 44, 48, 50, 52, 63

15 dati ...  $n = 15$

$$1/2(n + 1) = 1/2(15 + 1) = 8^{\text{vo}} \text{ dato}$$

# Quartile Inferiore

- $1/4(n + 1)^{\text{mo}}$  posto della lista (ordinata)

28, 29, 31, **35**, 35, 36, 37, 41,  
42, 43, 44, 48, 50, 52, 63

15 dati ...  $n = 15$

$$1/4(n + 1) = 1/4(15 + 1) = 4^{\text{to}} \text{ dato}$$



# Quartile Superiore

- $3/4(n + 1)^{\text{mo}}$  posto della lista (ordinata)

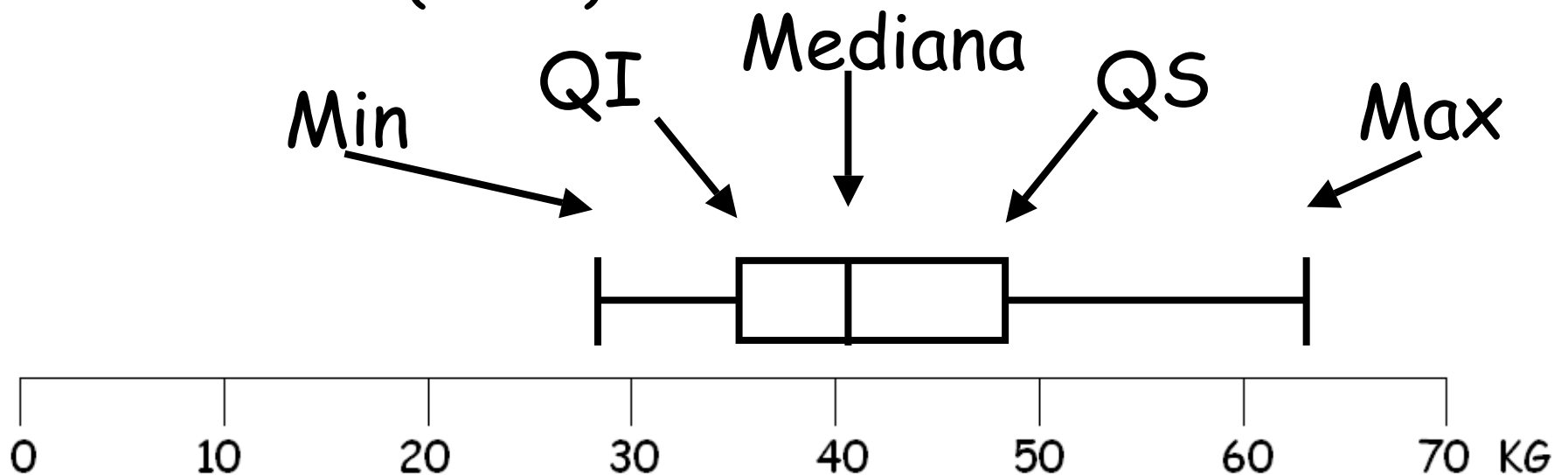
28, 29, 31, 35, 35, 36, 37, 41,  
42, 43, 44, 48, 50, 52, 63

15 dati ...  $n = 15$

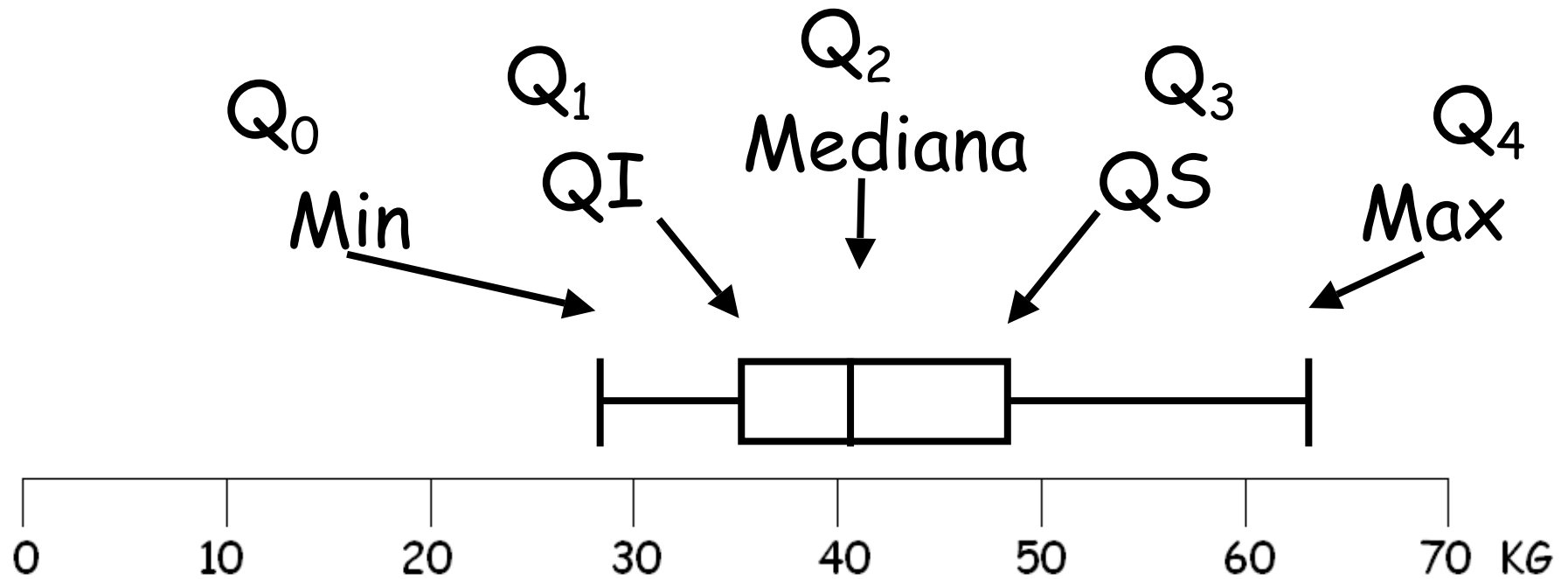
$$3/4(n + 1) = 3/4(15 + 1) = 12^{\text{mo}} \text{ dato}$$

# Rappresentazione con box

- Un diagramma che rappresenta:  
min (28KG),                      quartile inf = 35KG  
max (63KG),                      quartile sup = 48KG  
mediana (41KG) ...



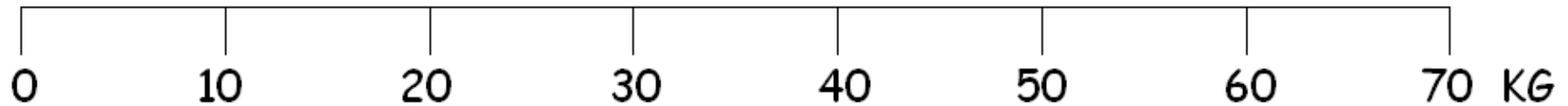
# Un po' di terminologia



Nomi alternativi dei quartili

# Un po' di terminologia

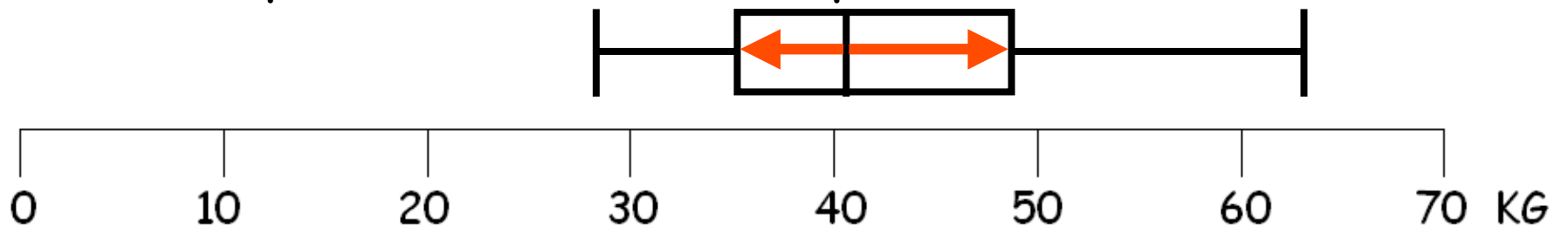
$QS - QI = \text{Variazione Interquartile IQR}$



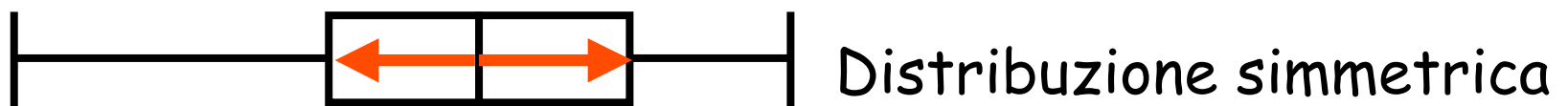
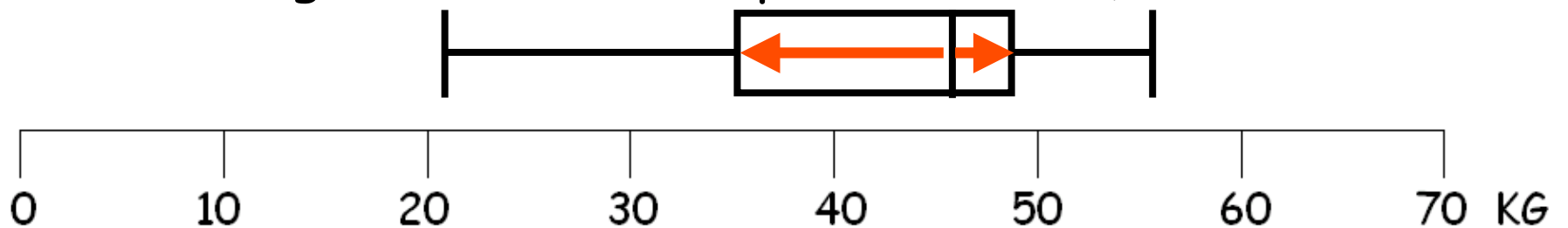
$\text{Max} - \text{Min} = \text{Escursione}$

# Un po' di terminologia

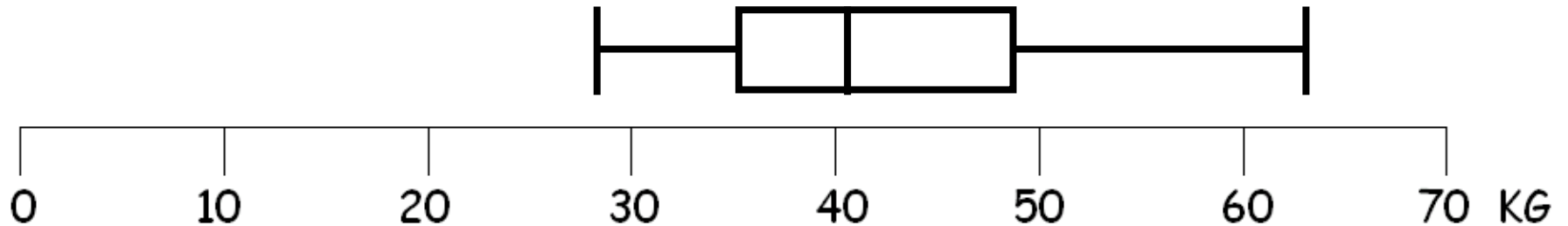
Tendenza positiva: la mediana è più vicina al QI



Tendenza negativa: mediana è più vicina al QS

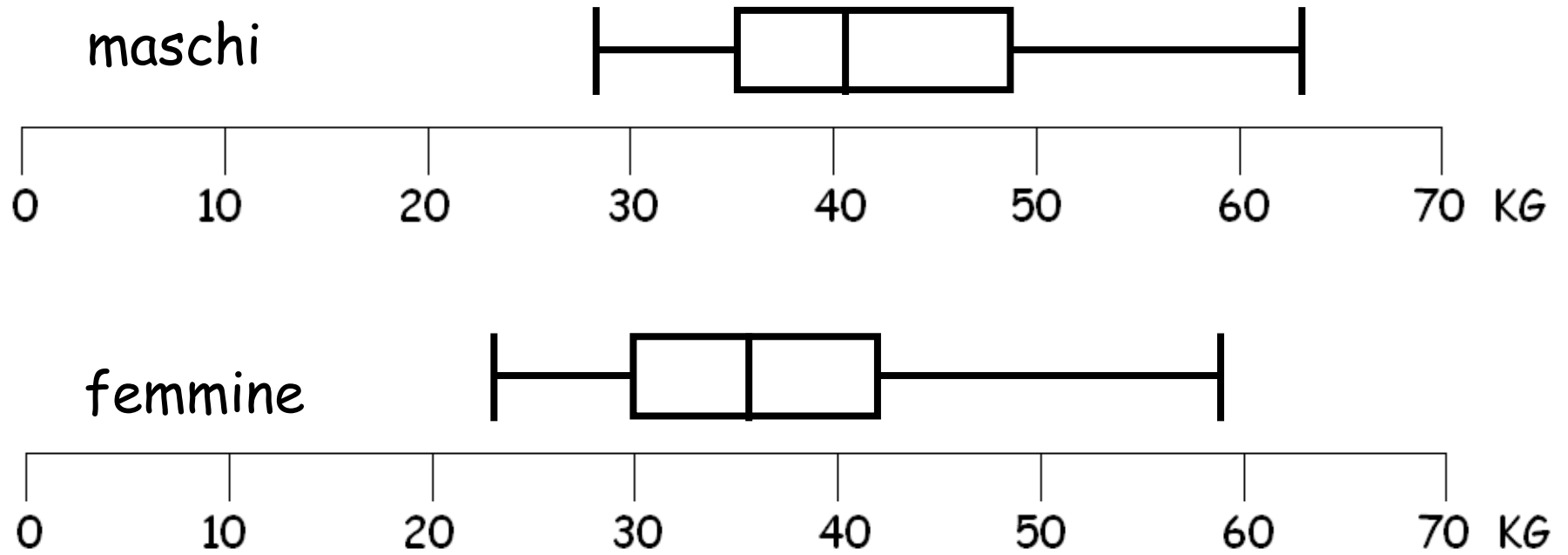


# Interpretare i diagrammi box and whisker



- Individuare facilmente il minimo/massimo e l'intervallo
- La 'scatola' contiene il 50% dei dati (la metà maggiormente rappresentativa)
- I 'baffi' mostrano gli estremi (25%) inferiore e superiore della distribuzione

# Confrontare sottoinsiemi

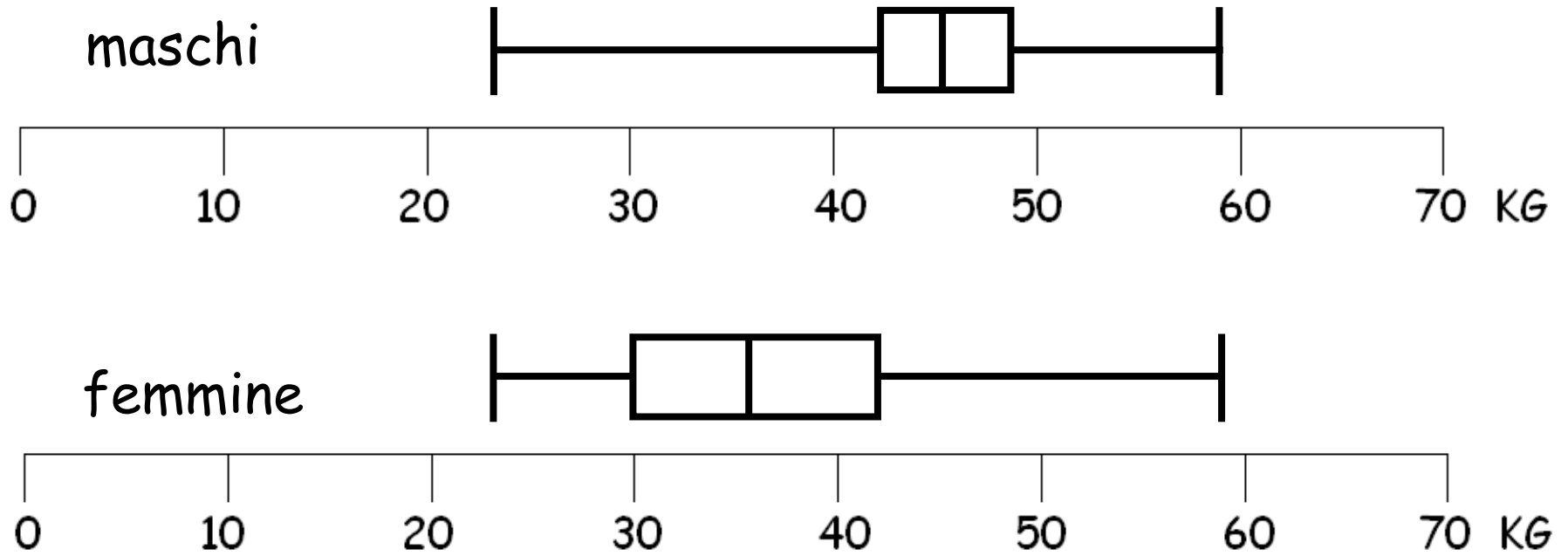


“la bimba più leggera pesa meno del bimbo più leggero”

“il bimbo più pesante pesa più della bimba più pesante”

“La metà più rappresentativa delle bambine pesa in genere meno della metà più rappresentativa dei bambini”

# Confrontare sottoinsiemi



“La bimba più leggera pesa come il bimbo più leggero”

“Il bimbo più pesante pesa come la bimba più pesante”

“La metà più rappresentativa delle bambine è più leggera della metà più rappresentativa dei bambini”

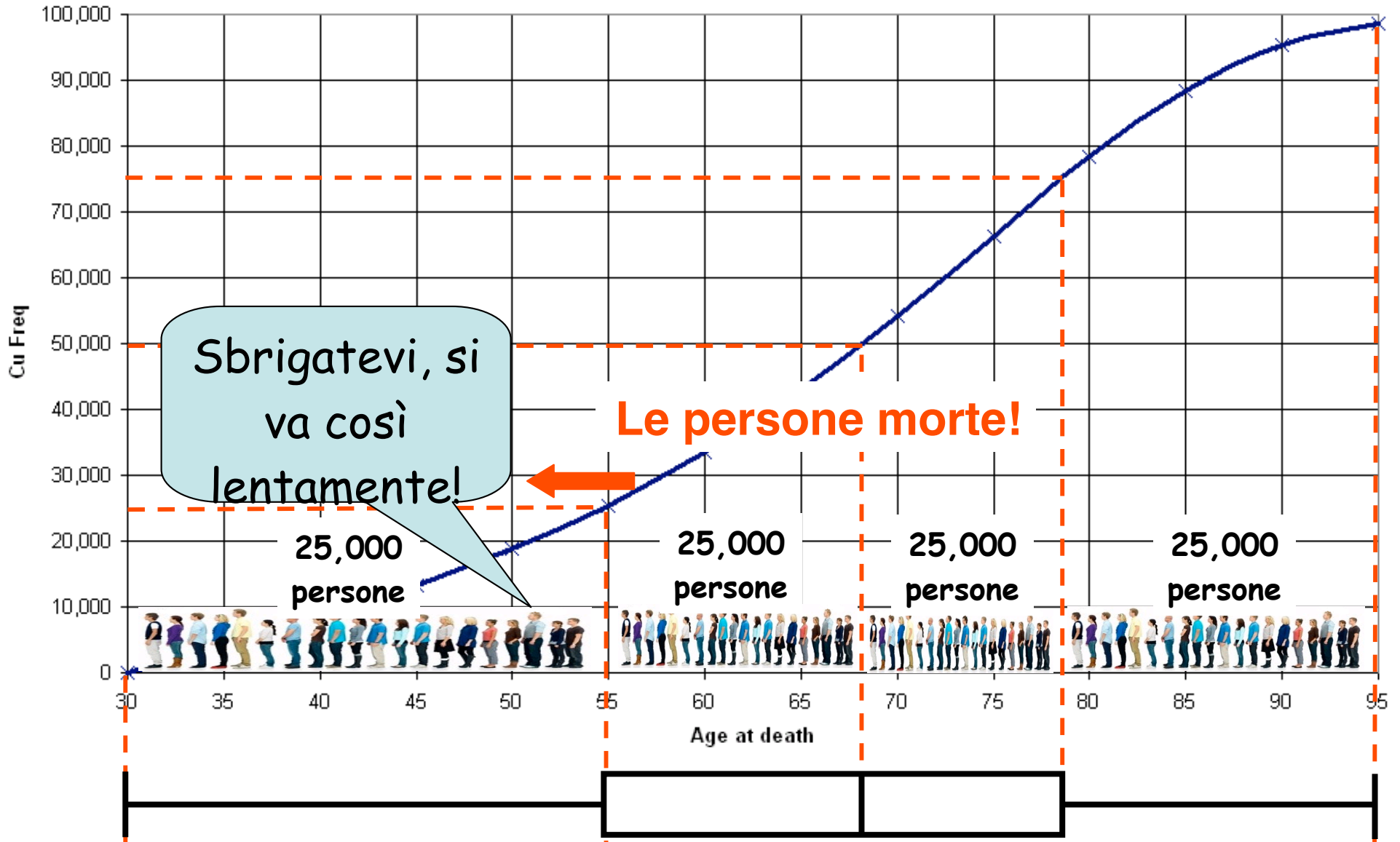
“Tre quarti delle bambine pesano meno dei tre quarti dei bambini”





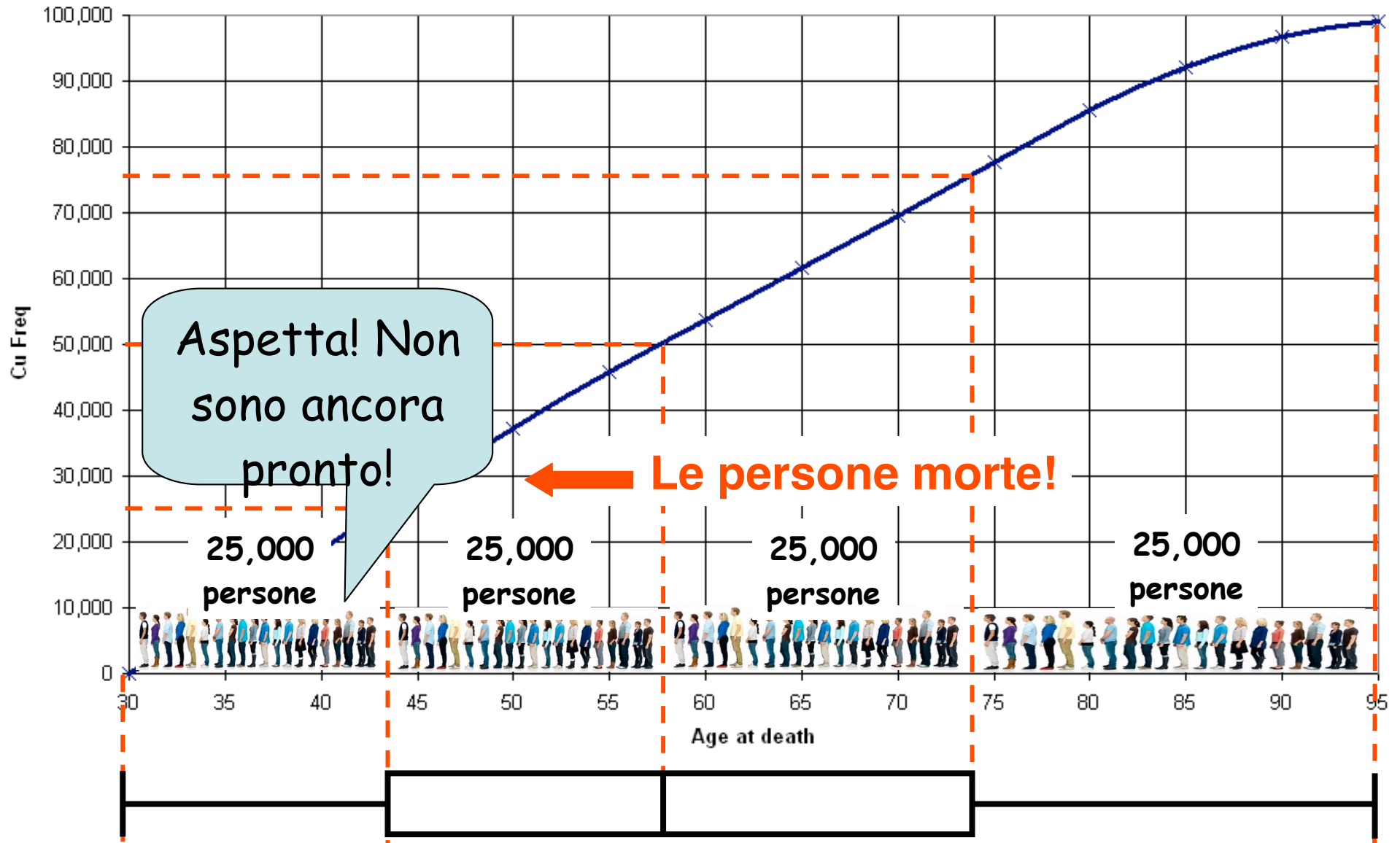
Ordine di altezza crescente ...

Cumulative Frequency (non-smoker deaths)

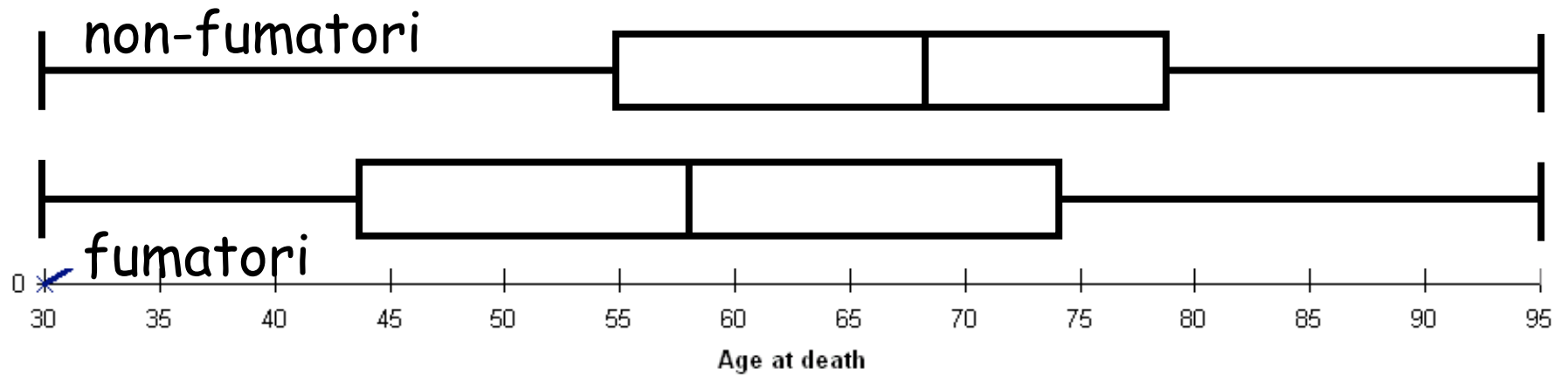


Dati: studio del Dr Pearl's 1938 su 100,000 non fumatori

### Cumulative Frequency (smoker deaths)



Dati: studio del Dr Pearl's 1938 su 100,000 fumatori

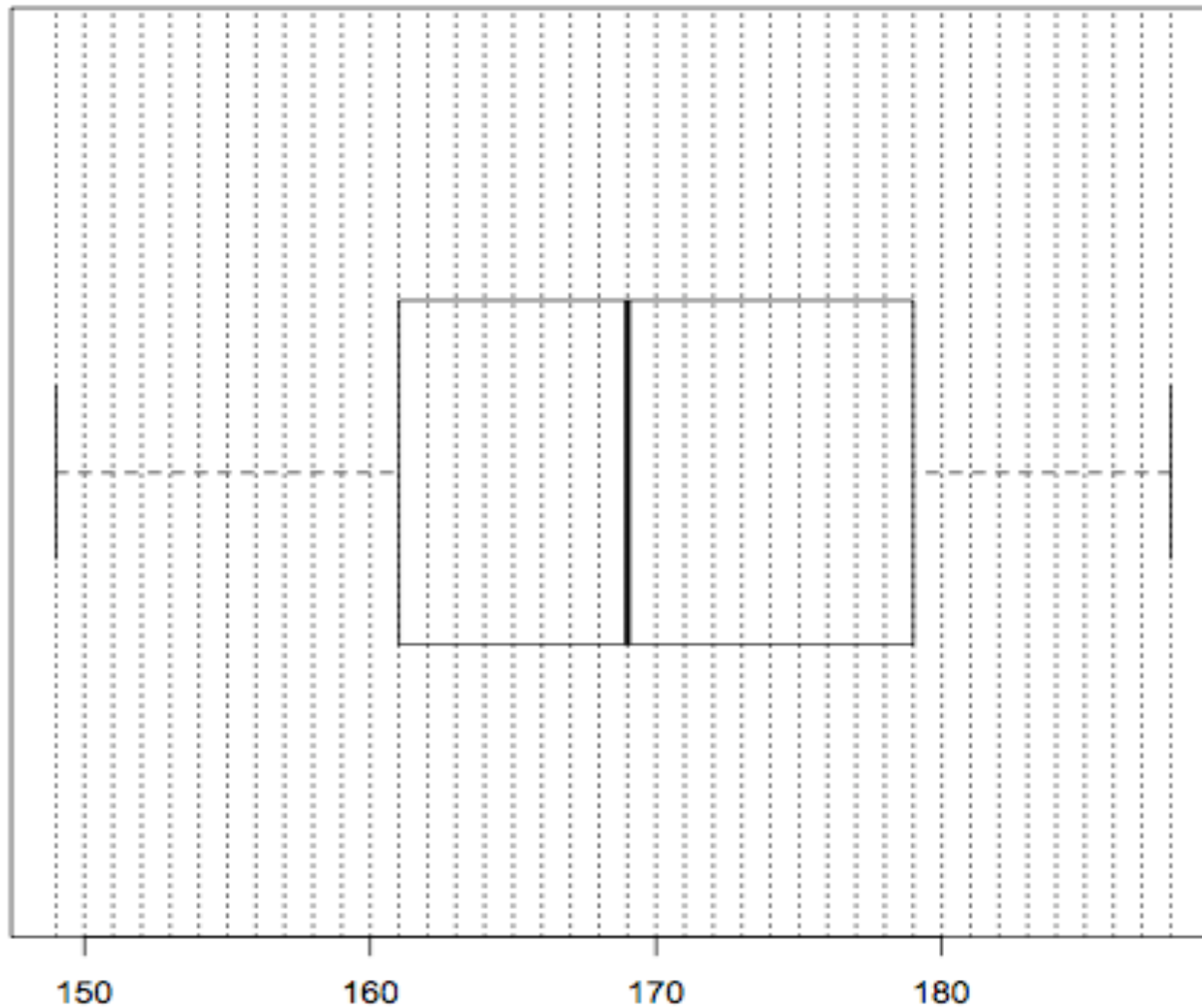


Confronto diretto con i box

# Misura dell'altezza di alcuni studenti

Escursione	39 cm
Massimo	188 cm
Mediana	169 cm
Quartile inferiore	161 cm
Variazione IQR	18 cm

Usa i dati per creare un box



23 ragazzi e 11 ragazze hanno svolto un test.

I loro punteggi sono:

ragazzi: 7, 13, 15, 19, 35, 35, 37, 43, 44, 44,  
45, 46, 47, 47, 49, 51, 52, 55, 55, 56,  
78, 82, 91

ragazze: 7, 18, 23, 47, 58, 63, 68, 72, 72, 75, 87

Usa i box per studiare le differenze dei loro punteggi.

Quali punteggi si devono escludere? Perché?

23 ragazzi e 11 ragazze hanno svolto un test di matematica.

I loro punteggi sono elencati sotto:

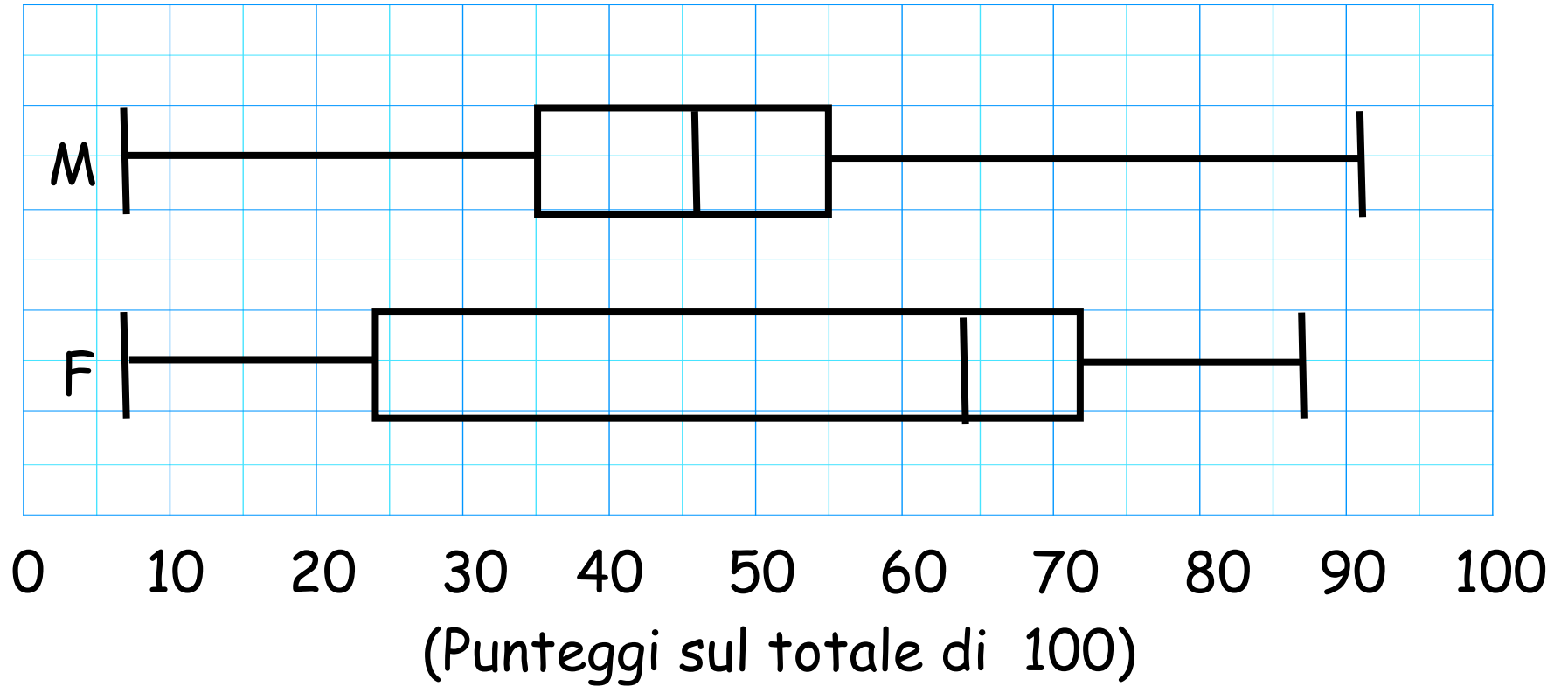
ragazzi: 7, 13, 15, 19, 35, 35, 37, 43, 44, 44, 45, 46, 47, 47, 49, 51,  
52, 55, 55, 56, 78, 82, 91

ragazze: 7, 18, 23, 47, 58, 63, 68, 72, 72, 75, 87

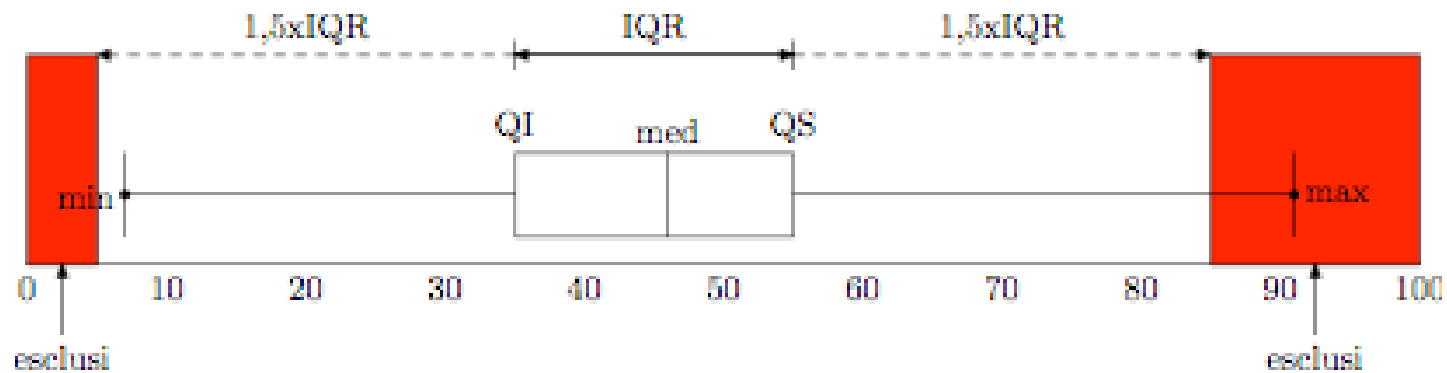
	ragazzi		ragazze
Min	7		7
QI	35	} 11 distribuzione	23 } 40 tendenza
Mediana	46	} 9 simmetrica	63 } 9 negativa
QS	55		72
Max	91		87
Variazione IQR	20		49
Escursione	84		80



## Box dei punteggi di ragazzi e ragazze



# Ricerca degli 'esclusi'



Quando considerare i dati 'estremi'  
troppo estremi?

# Esclusi

- Esclusi superiori  $> QS + 1.5 \times IQR$
- Esclusi inferiori  $< QI - 1.5 \times IQR$

Es: per i ragazzi  $1.5 \times IQR = 1.5 \times 20 = 30$

I punteggi inferiori a  $QI - 30$  ( $35 - 30 = 5$ ) sono esclusi

I punteggi superiori a  $QS + 30$  ( $55 + 30 = 85$ ) sono esclusi

L'unico escuso è 91 ... ma non è un punteggio così irragionevole! Si tratta solo di un esempio guida!

Per le ragazze gli esclusi sono  $< -50.5$  o  $> 145.5$   
... non ci sono punteggi da escludere!

# Domande

- Cosa servono i box?
- Spiega i vantaggi di usare i box per confrontare i dati.