Determinazione della frequenza con cui un carattere assume valori **internamente** o **esternamente** ad un intervallo

- Nota la distribuzione del carattere
  $\longrightarrow$ determino **ESATTAMENTE**
  la frequenza

- NON nota la distribuzione del carattere
  $\longrightarrow$ determino **APPROSSIMATIVAMENTE**
  la frequenza

ESEMPIO: nota la distribuzione

I seguenti dati mostrano il numero di mesi impiegato per laurearsi, per 20 studenti della Facoltà di Economia

48 52 60 48 72 56 76 54 58 50

84 48 50 52 56 60 54 68 52 66

Cosa posso dire della frequenza con cui gli studenti impiegano fra i 43 ed i 73 mesi per laurearsi?

$$Fr[X \in (43, 73)] = 18/20 = 0.9$$

Disuguaglianza di Chebyshev: LEGAME FRA MEDIA e VARIANZA

– Se non è nota la distribuzione del carattere ma sono note solo la media e la varianza
– Se gli intervalli sono centrati intorno alla media posso usare la disuguaglianza di Chebyshev per determinare
– un limite **inferiore** alla frequenza con cui il carattere assume valori **internamente** all'intervallo

$$Fr(\mu - \lambda\sigma < X < \mu + \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}$$

oppure, in modo equivalente

$$Fr(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}$$

oppure, ponendo $k = \lambda\sigma$

$$Fr(\mu - k < X < \mu + k) \geq 1 - \frac{\sigma^2}{k^2}$$

oppure, in modo equivalente

$$Fr(|X - \mu| < k) \geq 1 - \frac{\sigma^2}{k^2}$$

In modo simmetrico, poichè la frequenza totale è pari ad uno, posso usare la disuguaglianza di Chebyshev per determianre

–un limite **superiore** alla frequenza con cui il carattere assume valori **esternamente** all'intervallo

$$Fr(X < \mu - \lambda\sigma \quad \text{e} \quad X > \mu + \lambda\sigma) \leq \frac{1}{\lambda^2}$$

oppure, in modo equivalente

$$Fr(|X - \mu| > \lambda\sigma) \leq \frac{1}{\lambda^2}$$

oppure, ponendo $k = \lambda\sigma$

$$Fr(X < \mu - k \quad \text{e} \quad X > \mu + k) \leq \frac{\sigma^2}{k^2}$$

oppure, in modo equivalente

$$Fr(|X - \mu| > k) \leq \frac{\sigma^2}{k^2}$$

Dalla disuguaglianza di Chebyshev deduco che

per qualunque carattere quantitativo, indipendentemente dalla forma della distribuzione

- almeno il $1 - 1/2^2 = 75\%$ delle osservazioni deve essere contenuto in un intervallo centrato intorno alla media di ampiezza 4σ

- almeno il $1 - 1/3^2 = 88{,}89\%$ delle osservazioni deve essere contenuto in un intervallo centrato intorno alla media di ampiezza 6σ

- almeno il $1 - 1/4^2 = 93{,}75\%$ delle osservazioni deve essere contenuto in un intervallo centrato intorno alla media di ampiezza 8σ

Esempio

Una variabile casuale X ha media $E(X) = 3$ e varianza $V(X) = 4$.
Cosa si può dire della frequenza con cui X assume valori nell'intervallo $(-1, 7)$?

RISPOSTA: poiché si tratta di un intervallo simmetrico centrato intorno alla media e poiché non conosciamo la distribuzione di X possiamo usare la disuguaglianza di Chebyshev:

$$Fr(\mu - \lambda\sigma < X < \mu + \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}$$

Nel nostro caso abbiamo $\sigma = 2$ ($\sigma =$ deviazione standard = radice quadrata della varianza) e dobbiamo ricavare il valore di λ sfruttando il fatto che

$$\mu + \lambda\sigma = 7$$

$$3 + \lambda 2 = 7$$

quindi $\lambda = 2$ e risulta che

$$Fr(-1 < X < 7) \geq 1 - \frac{1}{2^2} = 0.75$$

PRIMO ESERCIZIO rivisto

Nell'esercizio dei mesi per laurearsi dei 20 studenti, se non avessi avuto i dati ma solo la media (58,2) e la deviazione standard (9.86) la disuguaglianza di Chebyshev avrebbe fornito una risposta molto meno precisa della precedente.

Infatti abbiamo che:

$$\mu + \lambda\sigma = 73$$

$$58.2 + \lambda 9.86 = 73$$

da cui ricavo che $\lambda = 1.5$.

Abbiamo inoltre che

$$58.2 - 1.5 * 9.86 = 43.41 \sim 43$$

Quindi per la disuguaglianza di Chebyshev:

$$F[X \in (43, 73)] \geq 1 - \frac{1}{1.5^2} = 0.55$$

mentre prima sapevo esattamente che
la frequenza richiesta è

$$Fr[X \in (43, 73)] = 18/20 = 0.9$$

ESERCIZIO
So che in media aspetto l'autobus 4.5 minuti e che la deviazione standard del tempo di attesa è di 2.6 minuti. Cosa posso dire sulla frequenza con cui aspetto l'autobus meno di 7.1 e più di 1.9 minuti?

RISPOSTA: poiché si tratta di un intervallo simmetrico centrato intorno alla media e poiché non conosciamo la distribuzione di X possiamo usare la disuguaglianza di Chebyshev.

Ponendo $7.1 = \mu + \lambda\sigma$ ottengo che $\lambda = 1$ e quindi la disuguaglianza di Chebyshev non è molto significativa in questo caso perchè dice che

$$Fr(1.9 \leq X \leq 7.1) \geq 0$$

ma questo già lo so perchè tutte le frequenze sono non negative

ESERCIZIO

In uno studio sul numero di sportelli delle banche si è constatato che in media le banche possiedono 10 sportelli con uno scarto quadratico medio di 2.
Cosa si può dire della frequenza con cui le banche hanno un numero di sportelli superiore a 13 o inferiore a 7?

RISPOSTA: uso la disuguaglianza nella forma

$$Fr(|X - \mu| > k) \leq \frac{\sigma^2}{k^2}$$

con $\mu = 10$ e poichè l'intervallo di interesse è (13, 7) ricavo che $k = 3$ quindi

$$Fr(|X - 10| > 3) \leq \frac{2^2}{3^2}$$

ovvero

$$Fr(X > 13 \text{ o } X < 7) \leq 0.44$$

ESERCIZIO: determinare un intervallo tale per cui un carattere con media pari a 10 e deviazione standard pari a 3 assuma valori internamente all'intervallo con frequenza almeno pari al 61%

RISPOSTA: devo risolvere rispetto a λ l'equazione

$$1 - 1/\lambda^2 = 0.61$$

la soluzione è $\lambda = 1.6$ quindi l'intervallo ha estremo inferiore

$$\mu - \lambda \times \sigma = 10 - 1.6 \times 3 = 5.2$$

ed estremo superiore

$$\mu + \lambda \times \sigma = 10 + 1.6 \times 3 = 14.8$$

quindi l'intervallo cercato è $(5.2\ ;\quad 14.8)$