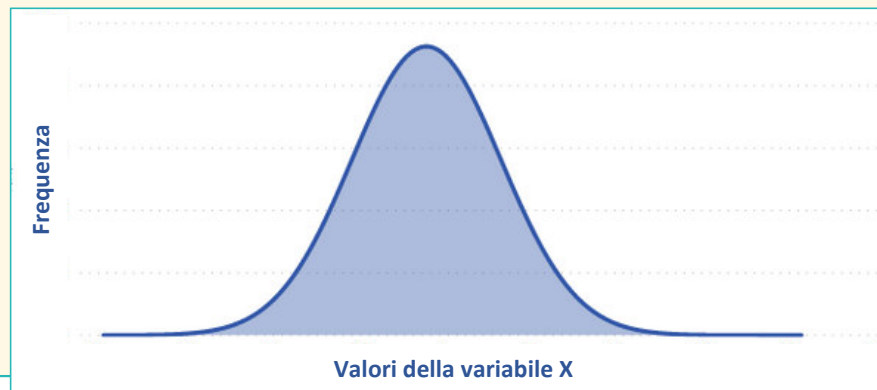


# Curva di Gauss o normale

La **curva di Gauss (curva normale)**, dalla classica forma a campana, viene utilizzata per descrivere la distribuzione di una variabile statistica continua.

L'evidenza ha dimostrato come si adatti bene a rappresentare il comportamento di una serie di fenomeni collettivi in svariati settori, es. medicina, biologia, economia etc. in cui le modalità intorno al valore medio sono più frequenti, mentre quelle alle estremità più rare.



# Esempi



Nelle discipline scientifiche la curva di Gauss approssima la distribuzione degli errori commessi nel misurare ripetutamente una stessa grandezza (curva degli errori accidentali).

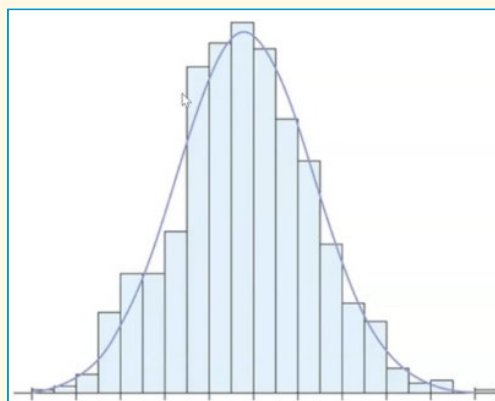
L'errore casuale varia per ogni misurazione in modo non prevedibile, una serie ripetuta di misurazioni comporta la progressiva riduzione dell'errore casuale.

Molte variabili biologiche (peso, statura, pressione arteriosa, glicemia etc,) seguono la distribuzione normale.

# Caratteristiche

Quando la popolazione è molto numerosa e le modalità sono tante, i rettangoli degli istogrammi di frequenza sono così fitti che danno luogo a una linea che sembra una vera e propria curva.

La curva continua approssima l'istogramma sottostante, ossia la distribuzione di frequenza.



# Variabile aleatoria

Una **variabile aleatoria** (detta anche variabile casuale) è l'insieme dei possibili risultati numerici di un esperimento, cioè di una prova i cui esiti non sono prevedibili con certezza perché cambiano a seconda del verificarsi di eventi aleatori.

Sono grandezze che nel corso di un esperimento possono assumere diversi valori non prevedibili a priori in modo deterministico.

Esempio; nel lancio dei dadi non si può conoscere a priori il valore della faccia che si presenterà.

# Variabile aleatoria

La relazione tra variabile aleatoria e probabilità si può ricavare assegnando a ogni esito possibile la probabilità che si verifichi. L'insieme di tutti i valori delle probabilità si chiama **distribuzione di probabilità**.

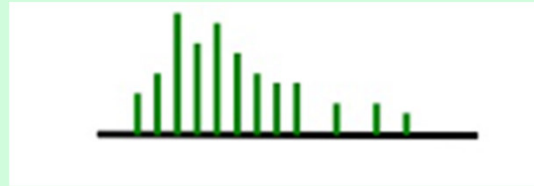
Si definisce distribuzione di probabilità (o funzione di probabilità) di una variabile aleatoria  $X$  l'insieme dei valori  $x_i$  e delle relative probabilità  $p_i$ .

Si possono distinguere:

- **variabili aleatorie discrete** se l'insieme delle possibili realizzazioni è finito o numerabile (es. il numero dei giorni di pioggia in un anno).
- **variabili aleatorie continue**, se l'insieme delle possibili realizzazioni è un intervallo (es. la velocità di una macchina)

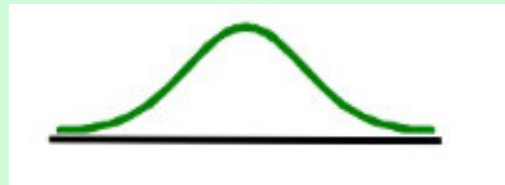
# Funzione di probabilità e di densità

Per una variabile casuale discreta, la **funzione di probabilità** fornisce la probabilità che la variabile casuale assuma un particolare valore.



Con le variabili casuali continue, la controparte della funzione di probabilità è la **funzione di densità**. La differenza è che la funzione di densità non fornisce direttamente le probabilità.

Per descrivere la distribuzione di una variabile aleatoria continua, non si può più assegnare una probabilità positiva ad ogni valore possibile.



# Funzione di probabilità

Esempio: negli ultimi 300 giorni di esercizio una concessionaria di macchine riporta i seguenti risultati di vendita

Numero di macchine vendute	Giorni	F(x)
0	54	0,18
1	117	0,39
2	72	0,24
3	42	0,14
4	12	0,04
5	3	0,01
	$\Sigma$ 300	1

# Funzione di probabilità

Dai dati storici sappiamo che  $x$  è una variabile casuale discreta che può assumere i valori 0, 1, 2, 3, 4, 5.

$f(0)$  fornisce la probabilità di vendere 0 automobili,  $f(1)$  la probabilità di vendere 1 automobile e così via.

$f(0)$  sta ad indicare che la probabilità di vendere 0 automobili nell'arco di un giorno è 0,18.

Nella costruzione di una funzione di probabilità discreta devono essere soddisfatte 2 condizioni:

$$f(x) \geq 0$$

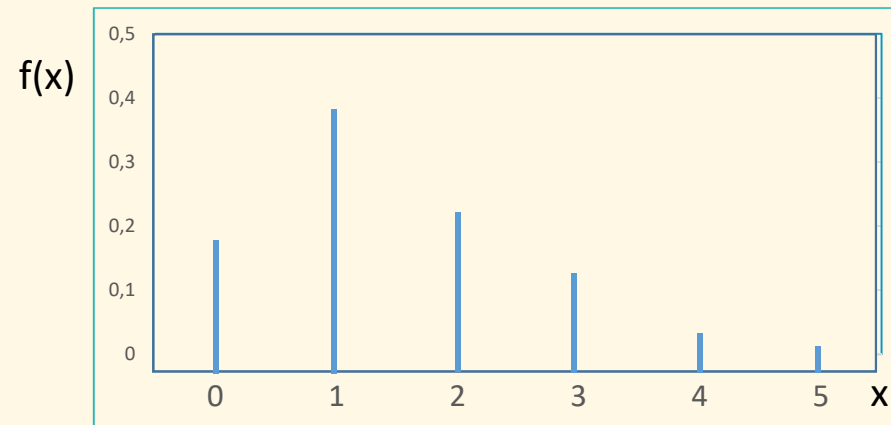
$$\sum f(x) = 1$$

Nella tabella precedente  $f(x)$  è  $\geq 0$  per tutti i valori delle  $x$  e la somma della probabilità è pari a 1



# Rappresentazione grafica

Distribuzione di probabilità del numero di automobili vendute nell'arco di un giorno



# Esercizio

Contando il numero di sale operatorie in uso presso il Gemelli, nel corso di un periodo di 20 giorni sono stati raccolti i seguenti dati:

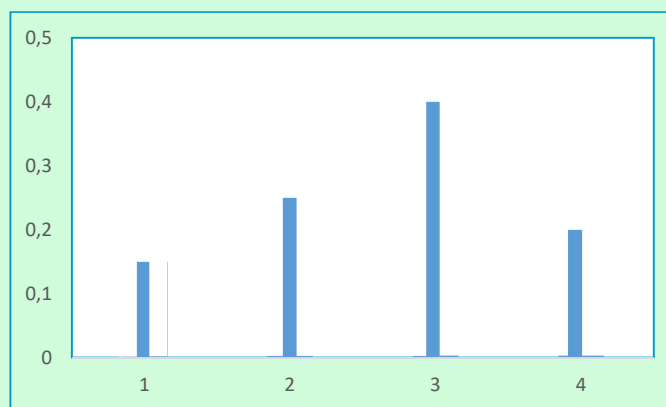
in 3 dei 20 gg è stata utilizzata solo 1 sala operatoria, in 5 dei venti giorni ne sono state utilizzate 2, in 8 dei 20 giorni ne sono state utilizzate 3 e in 4 dei 20 giorni sono state utilizzate tutte e 4 le sale.

- 1- costruire la distribuzione di probabilità del numero di sale operatorie utilizzate in un giorno
- 2- disegnare il grafico della distribuzione
- 3- mostrare che la distribuzione di probabilità ottenuta soddisfa le condizioni necessarie affinché una distribuzione di probabilità discreta sia valida.

# Soluzione

x	f(x)
1	$3/20=0,15$
2	$5/20=0,25$
3	$8/20=0,40$
4	$4/20=0,20$
	1

Soddisfa entrambe le condizioni



# Funzione di probabilità uniforme discreta

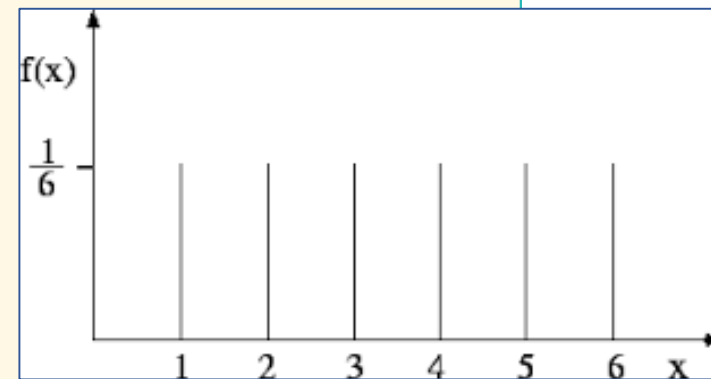
$$f(x) = 1/n$$

$n$  = numero di valori che la variabile casuale può assumere

L'esperimento consiste nel lancio di un dado, la variabile casuale  $x$  è il numero che appare.

Sono possibili  $n = 6$  valori per la variabile casuale  $x = 1, 2, 3, 4, 5, 6$ .

$x$	$f(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

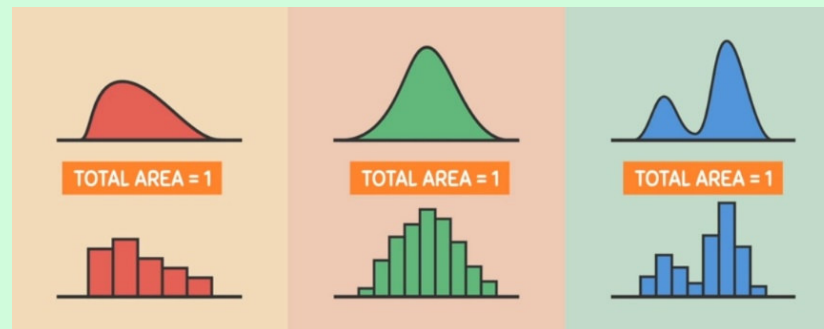


# Funzione di densità di probabilità

Quando la variabile aleatoria  $X$  è continua la distribuzione delle probabilità  $P(X)$  si chiama **funzione di densità di probabilità**.

La sua rappresentazione grafica è una curva continua identificata da un'equazione  $y = P(x)$  e l'area totale tra l'asse orizzontale e la curva stessa è uguale a 1 se parliamo di frequenze relative o al 100% se parliamo di frequenza percentuale.

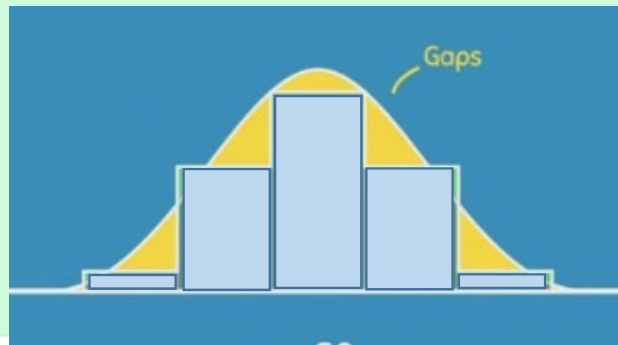
La curva di densità ci aiuta a visualizzare la forma della distribuzione e si può ricavare da ogni tipo di istogramma.



# Vantaggi rispetto all'istogramma

1- In un istogramma più intervalli si hanno a disposizione, migliore è la rappresentazione della distribuzione. Con la curva di densità non si è limitati dal numero di intervalli che si hanno.

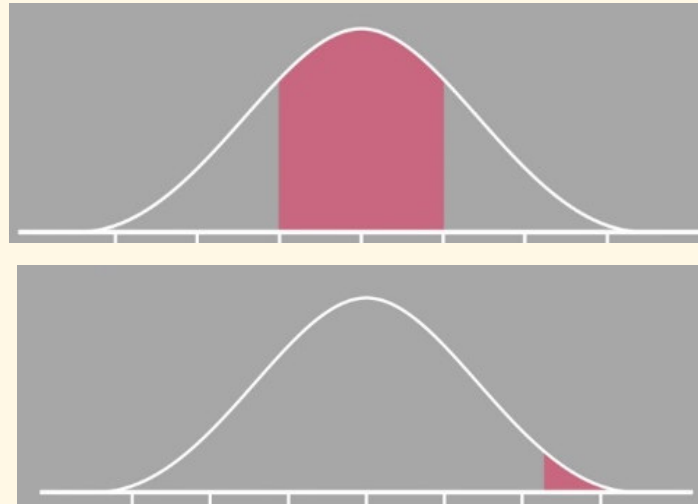
2- E' più facile lavorare con una curva soprattutto quando si ha a che fare con una popolazione di grandi dimensioni, mentre se si ha un numero esiguo di osservazioni la rappresentazione tramite una curva non è accurata, ci sono dei gap maggiori.



# Funzione di densità di probabilità

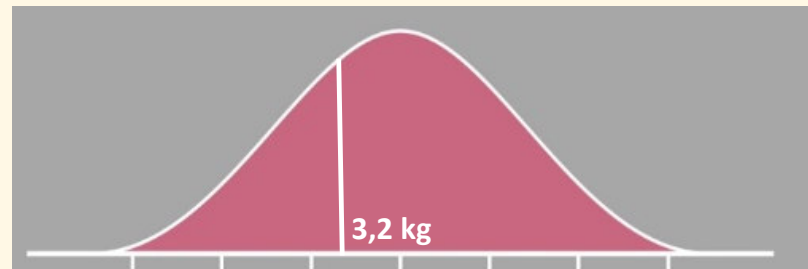
Il termine densità richiama il concetto di area sottostante la curva.

Se tutta l'area rappresenta il 100% delle probabilità, una piccola parte di essa quantifica la probabilità che la variabile casuale assuma un valore compreso nell'intervallo.



# Esempio di utilizzo

Qual è la probabilità che un neonato pesi 3,2 kg alla nascita?



La formula della probabilità si calcola come casi favorevoli sul numero dei casi possibili, in questo caso infinito. Il risultato quindi è zero.



Non ha senso calcolare la probabilità di una singola modalità, ossia la probabilità che una variabile continua assuma un determinato valore è sempre 0.



# Funzione di densità di probabilità

La funzione di densità non fornisce direttamente la probabilità, tuttavia l'area al di sotto della curva corrispondente a un dato intervallo fornisce la probabilità che la variabile casuale continua assuma un valore in quell'intervallo.

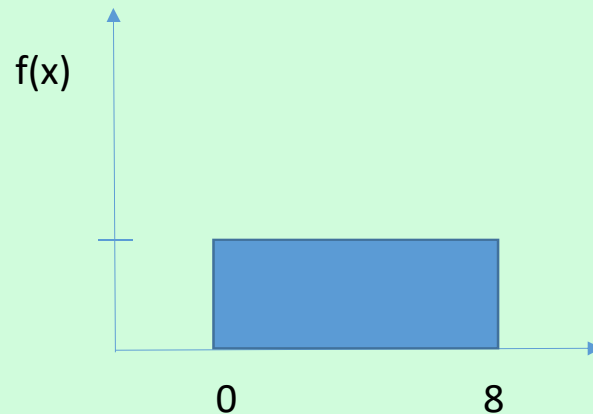
Quindi per variabili casuali continue si calcola la probabilità che la variabile assuma qualunque valore in un determinato intervallo.

Presi 2 valori  $a$  e  $b$ , l'area identificata dalle loro proiezioni sull'asse orizzontale rappresenta quante probabilità abbiamo che la variabile aleatoria assuma uno qualunque dei valori compresi tra i 2 estremi.

# Funzione di densità uniforme

Ogni volta che la probabilità è proporzionale alla lunghezza dell'intervallo, la variabile casuale si distribuisce uniformemente.

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$



Area compresa nel rettangolo = 1  
Si calcola come base per altezza

Base = 8  
Altezza =  $1:8 = 0,125$

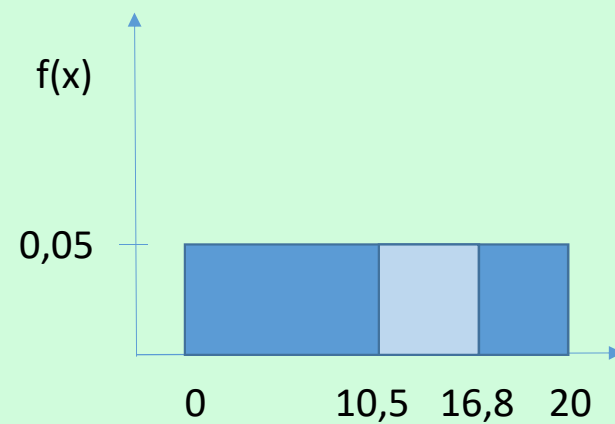
# Esempio

Qual è la proporzione dei valori compresi tra 10,5 e 16,8?

$$f(x) = 1/20 = 0,05$$

Base = 6,3

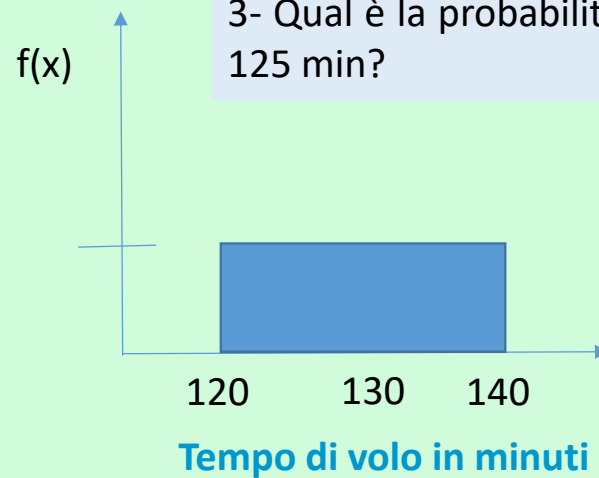
$$\text{Area} = 0,05 * 6,3 = 0,315 \text{ o } 31,15\%$$



# Esempio

Tempo di volo compreso in un intervallo di 120-140min  
con ogni intervallo di 1 min ugualmente probabile

- 1- Qual è la probabilità che il tempo di volo sia compreso tra 120 e 130 min? Quanto vale la  $P(120 \leq X \leq 130)$ ?
- 2- Qual è la probabilità che il tempo di volo sia compreso tra 128 e 136 min?
- 3- Qual è la probabilità che il tempo di volo sia 125 min?



# Caratteristiche della curva normale

1- Simmetrica rispetto alla retta  $X = \mu$ , ossia speculare rispetto al suo valore centrale. La forma della curva a sinistra è speculare rispetto a quella a destra.

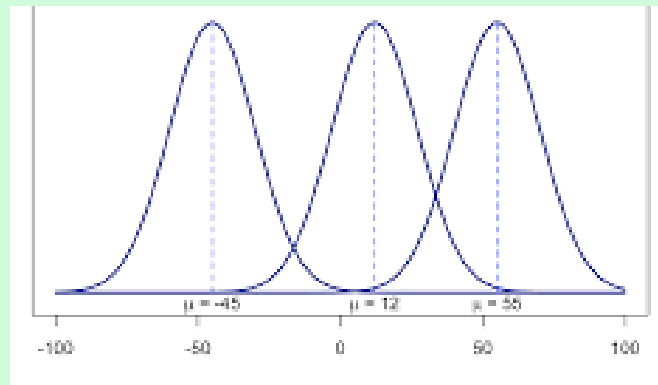
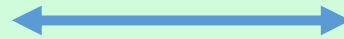
2- La moda si trova nel punto di **massima frequenza** e quindi il punto di massimo della curva.  
Moda, mediana e media coincidono.

3- Bastano 2 valori per disegnare la curva normale, la  $\mu$  e la  $\sigma$ . Tali valori determinano la posizione e la forma della distribuzione normale e la differenziano da un'altra curva normale.

# Caratteristiche

Il valore della media che indica il centro della distribuzione (della campana), caratterizza la posizione della curva rispetto all'asse delle ordinate.

Al variare della media la curva si sposta lungo l'asse x, mentre la sua forma non si modifica.  
La media può assumere anche valore negativo.



Uguale  $\sigma$  ma  
 $\mu$  diversa

Fonte; Hartmann, K., Krois, J., Waske, B. (2018)

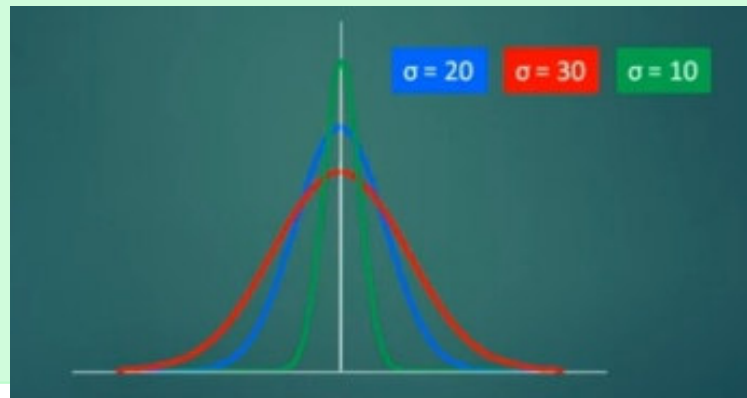
# Caratteristiche

La deviazione standard caratterizza la forma della curva, in quanto è una misura della dispersione dei valori attorno al valore medio: al variare di  $\sigma$  la curva cambia forma.

Al crescere di  $\sigma$  la curva si appiattisce e si allarga, mentre al diminuire di  $\sigma$  la curva si restringe e si alza.

In corrispondenza dei punti di flesso si manifesta un cambiamento di curvatura.

I punti  $\mu - \sigma$  e  $\mu + \sigma$  sono punti di flesso, ossia si trovano una  $\sigma$  sopra la media e una  $\sigma$  sotto la media.



# Caratteristiche

4- Trattandosi di variabili continue l'intera area sotto la curva a campana è sempre uguale a 1. L'area alla sinistra della media è pari a 0,5 così come quella a destra.

5- Asintotica rispetto all'asse delle  $x$  (al tendere di  $x$  verso  $-\infty$  o  $+\infty$ ). Allontanandosi dal valore centrale si avvicina sempre più all'asse orizzontale senza toccarlo.

Non potendo sapere con certezza quale sarà l'ultimo valore con almeno una osservazione e tenendo presente che le frequenze verso gli estremi diminuiscono sempre di più, si disegna una curva che non si interrompe in un punto specifico.



# Funzione di densità normale

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left[ \frac{(x-\mu)}{\sigma} \right]^2}$$

$\mu$  = media

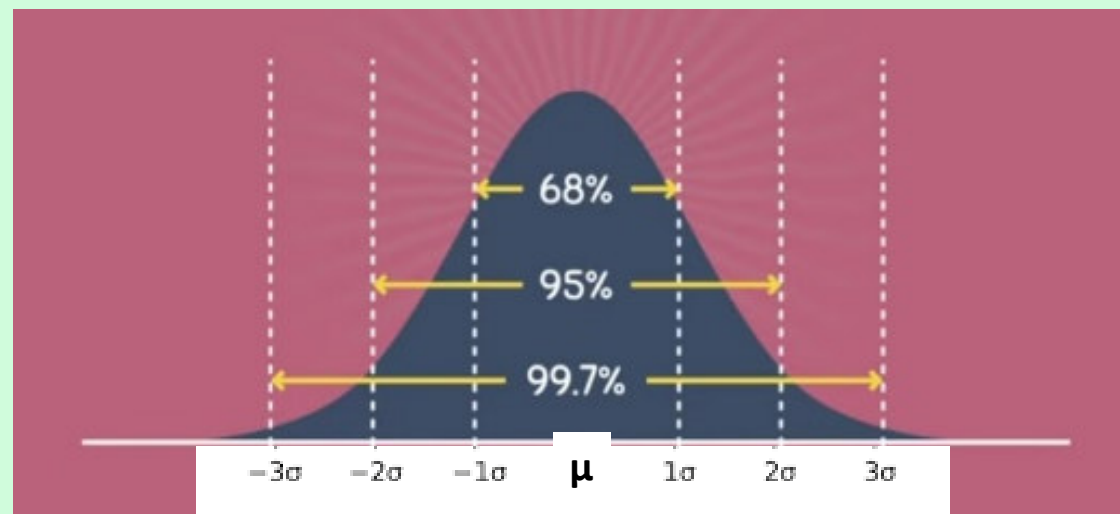
$\sigma$  = dev std

$\pi$  = 3.14159

$e$  = 2.71828 base dei  
logaritmi naturali

# Regola empirica

**Regola empirica o del 68-95-99,7** è una regola statistica che afferma che per una distribuzione normale, quasi tutti i dati osservati cadranno entro tre deviazioni standard



# Regola empirica

Nell'intervallo compreso fra  $\mu - \sigma$  e  $\mu + \sigma$  si trovano circa il 68% dei valori osservati della distribuzione (rientrano in una deviazione standard).

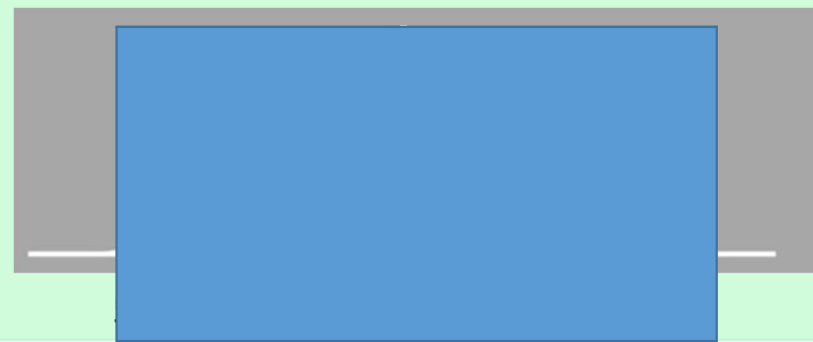
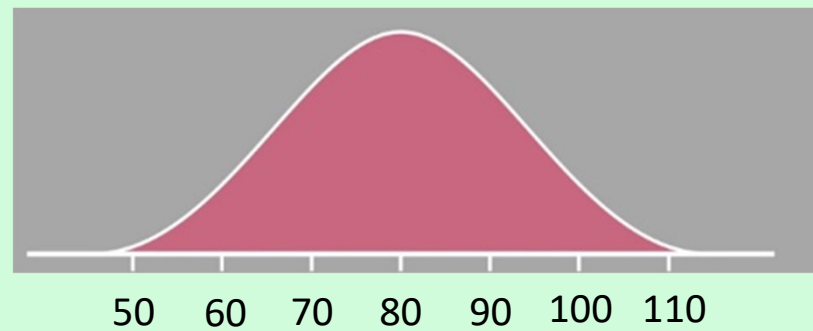
Per cui il 32% dei valori si trovano equamente presenti all'esterno dell'intervallo.

Nell'intervallo compreso fra  $\mu - 2\sigma$  e  $\mu + 2\sigma$  si trovano circa il 95% dei valori osservati. Per cui il 5% dei valori si trovano equamente presenti all'esterno dell'intervallo.

Il 99,7% dei dati osservati si trova entro 3 deviazioni standard dalla media.

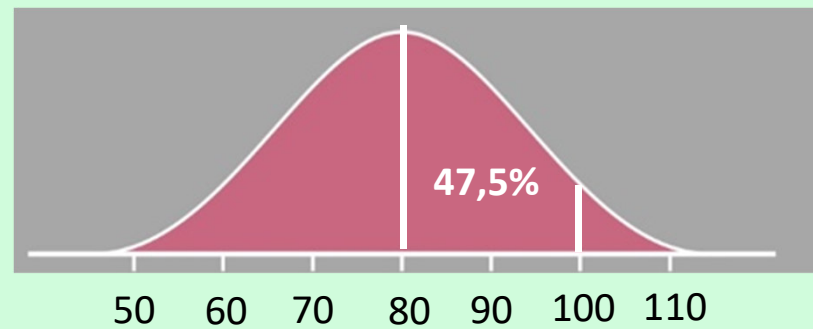
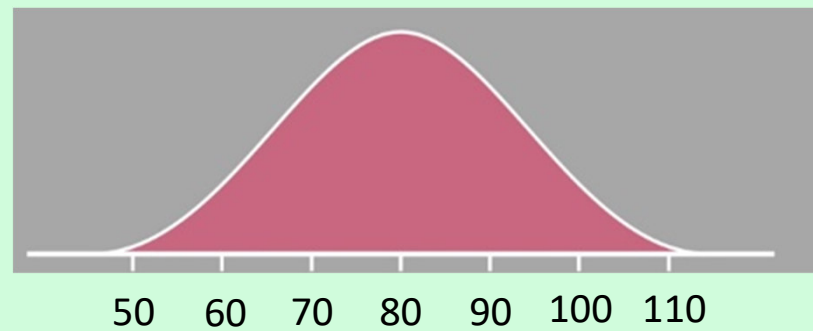
# Esempio

La distribuzione normale nel grafico ha una deviazione standard di 10. Qual è l'area contenuta tra 80 e 100?



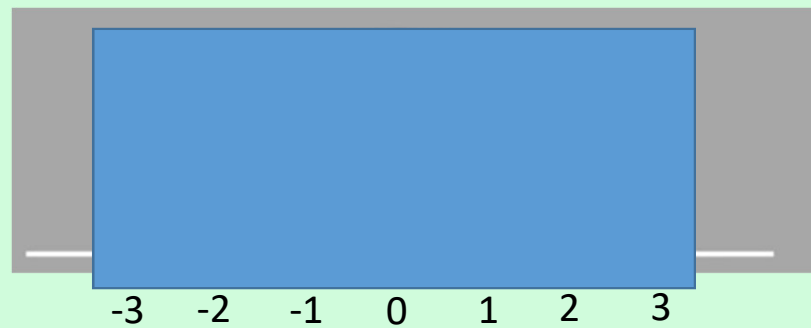
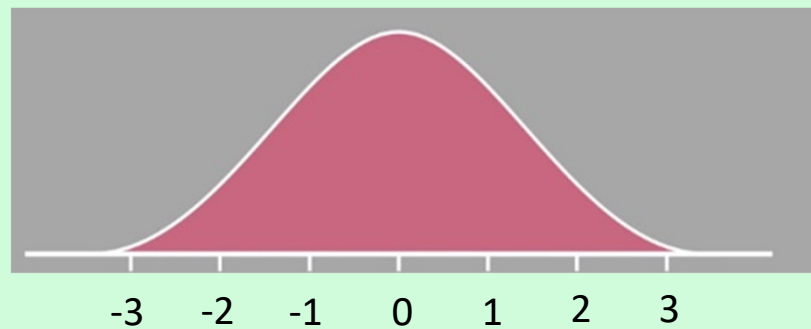
# Esempio

La distribuzione normale nel grafico ha una deviazione standard di 10. Qual è l'area contenuta tra 80 e 100?



# Esempio

Per la distribuzione normale nel grafico qual è l'area contenuta tra -2 e 1? Media = 0 e Deviazione standard 1



# Esempio

Per la distribuzione normale nel grafico qual è l'area contenuta tra -2 e 1? Media = 0 e Deviazione standard 1

