Distribuzioni di Probabilità

L'istogramma (o il box plot o lo scatter plot) viene utilizzato per descrivere i dati campionari.

Un campione è un insieme di misurazioni selezionate da una popolazione più ampia.

Ad esempio, le misurazioni del diametro dei bulloni sono ottenute da un campione di bulloni selezionati dal processo di fabbricazione.

La popolazione in questo esempio è la raccolta di tutti i diametri dei bulloni prodotti da quel processo.

L’utilizzo dei metodi statistici permette di analizzare i dati sul diametro dei bulloni del campione e trarre alcune conclusioni sull’intero processo che produce bulloni.

Distribuzioni di Probabilità

Una distribuzione di probabilità è un modello matematico che collega il valore della variabile alla probabilità che tale valore si trovi all'interno della popolazione.

Le distribuzioni di probabilità vengono utilizzate per modellizzare il comportamento di un fenomeno di interesse in relazione alla popolazione di riferimento, ovvero alla totalità dei casi di cui lo sperimentatore osserva un dato campione.

Distribuzioni di Probabilità

Ad esempio, il diametro dei bulloni può essere visto come una variabile casuale perché assume valori diversi nella popolazione secondo un meccanismo casuale, quindi la distribuzione di probabilità del diametro dei bulloni descrive la probabilità che nella popolazione si presenti qualsiasi valore del diametro dei bulloni.

Distribuzioni di Probabilità

Ripetendo più volte la misura di una grandezza nelle stesse identiche condizioni, non si ottiene in generale sempre lo stesso valore, bensì valori diversi, ciascuno con una sua propria probabilità di essere osservato.

Dunque, si è portati a postulare una *incertezza di principio* sul valore di una grandezza, che prende il posto della univocità oggettiva.

Distribuzioni di Probabilità

Quando la variabile da misurare è espressa mediante una scala continua, la sua distribuzione di probabilità viene definita una distribuzione continua. Ad esempio, la distribuzione di probabilità dello spessore dello strato di metallo è continua.

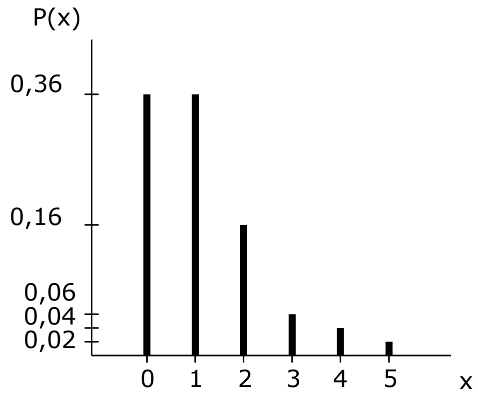
Quando la variabile da misurare può assumere solo alcuni valori (quali gli interi 0, 1, 2, 3, …) la distribuzione di probabilità viene definita una distribuzione discreta. Ad esempio, la distribuzione del numero di non conformità o difetti in un certo prodotto.

Distribuzioni di Probabilità

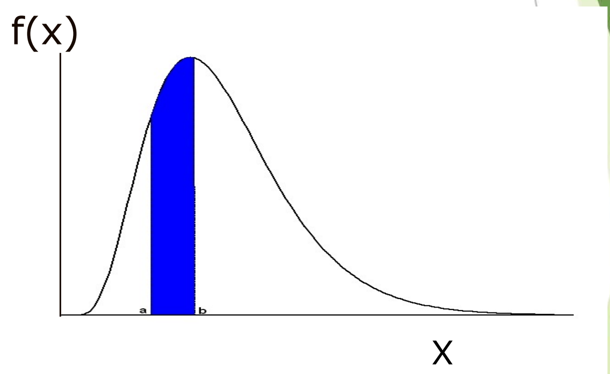
Formalmente, le distribuzioni di probabilità vengono espresse da una legge matematica detta **funzione di densità di probabilità** (indicata con $f(x)$) o **funzione di probabilità** (indicata con $p(x)$) rispettivamente per le distruzioni continue o discrete.

Distribuzioni di Probabilità

Funzione di probabilità



Funzione di densità di probabilità



In corrispondenza di ogni valore, la barra verticale ha un'altezza proporzionale alla probabilità.

La somma di tutte le barre è pari a 1.

L'area sottostante alla curva è uguale alla probabilità.

Distribuzioni Discreta

Una azienda produce migliaia di chip di semiconduttori al giorno.

In media, l'1% di questi chip non è conforme alle specifiche.

Ogni ora, un ispettore seleziona un campione casuale di 25 chip e classifica ogni chip nel campione come conforme o non conforme. Se x è la variabile casuale che rappresenta il numero di chip non conformi nel campione, allora la distribuzione di probabilità di x è

$$p(x) = \binom{25}{x}(0.01)^x(0.99)^{25-x} \qquad x=0, 1, 2, \ldots, 25$$

Dove $\binom{25}{x}$ = 25!/[x!(25-x)!

Distribuzioni Discreta

Questa è una distribuzione discreta, poiché il numero osservato di non conformità è x = 0, 1, 2,…, 25 e viene chiamata *distribuzione binomiale*.

Distribuzioni Continua

Supponiamo che x sia una variabile casuale che rappresenta il contenuto effettivo in kg di un sacco da 1 kg di chicchi di caffè.

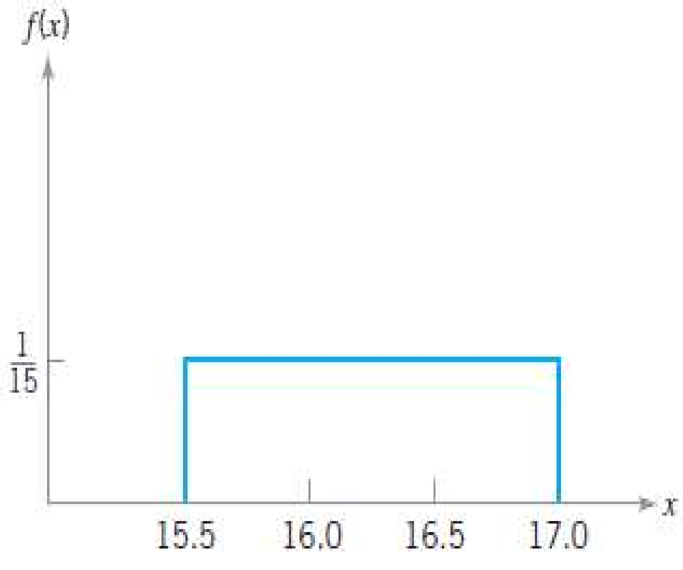
Si presume che la distribuzione di probabilità di x sia

$$f(x) = \frac{1}{1.5} \qquad 15.5 \leq x \leq 17.0$$

Distribuzioni Continua

Questa è una distribuzione continua, poiché l'intervallo di x è il intervallo [15.5, 17.0].

Questa distribuzione è chiamata *distribuzione uniforme* .

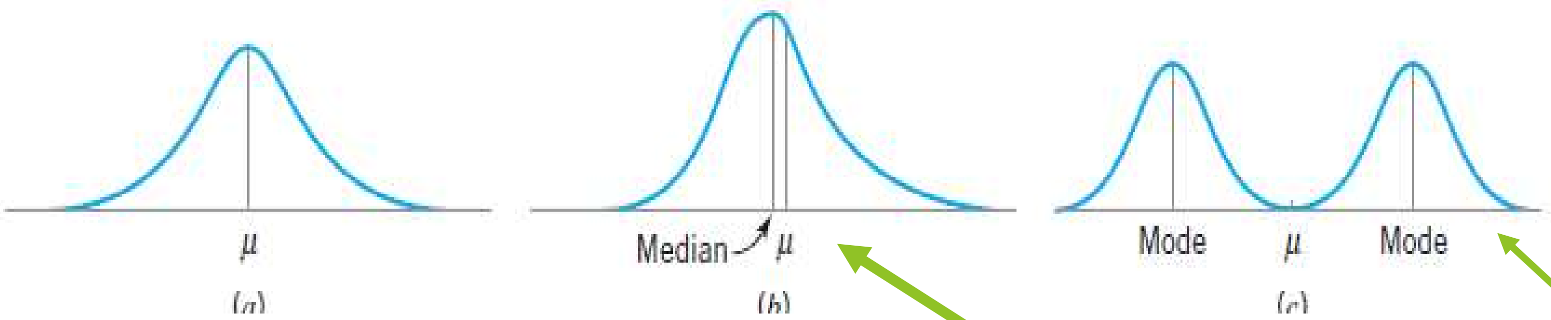


La media di una distribuzione

La media di una distribuzione di probabilità è una misura della tendenza centrale nella distribuzione o della sua posizione.

La media è il punto in cui la distribuzione "bilancia" esattamente.

Pertanto, la media è semplicemente il centro della distribuzione di probabilità.

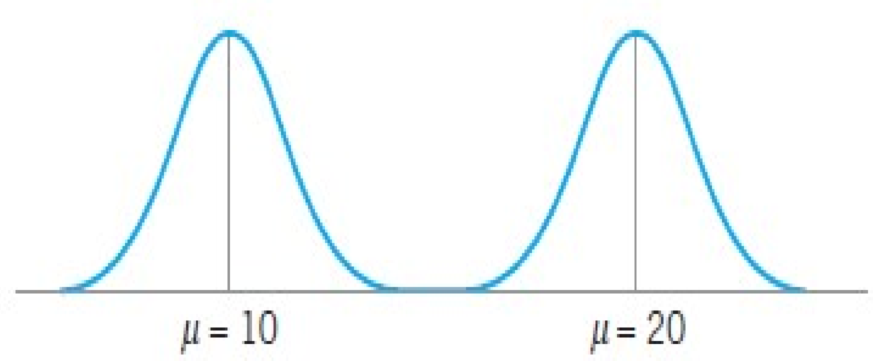


la media non è necessariamente la mediana

La media non è necessariamente il valore più probabile della variabile (la moda)

La media di una distribuzione

La media determina semplicemente la posizione della distribuzione.

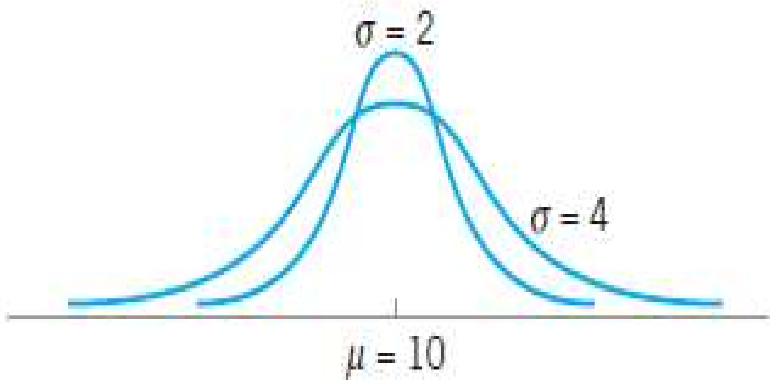


La variabilità di una distribuzione

La dispersione o variabilità in una distribuzione è espressa dalla varianza.

La varianza è espressa nel quadrato delle unità della variabile originale, pertanto, è consuetudine lavorare con la radice quadrata della varianza, chiamata deviazione standard.

La deviazione standard è una misura di dispersione nella popolazione espressa nelle unità originali.



Distribuzioni di Probabilità

Alcune distribuzioni discrete

Ipergeometrica

Binomiale

Poisson

Alcune distribuzioni continue

Normale

Esponenziale

Distribuzione Ipergeometrica

Supponiamo che ci sia una popolazione finita composta da N elementi. Alcuni numeri, per esempio, di questi elementi rientrano in una classe di interesse.

Un campione casuale di n articoli viene selezionato dalla popolazione senza sostituzione e viene osservato il numero di elementi nel campione che rientrano nella classe di interesse, ad esempio x.

Quindi x è una variabile casuale ipergeometrica con la distribuzione della probabilità definita come segue.

Distribuzione Ipergeometrica

Quindi x è una variabile casuale ipergeometrica con la distribuzione della probabilità definita come segue.

$$p(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, 2, \ldots, \min(n, D)$$

$$\mu = \frac{nD}{N}$$

$$\sigma^2 = \frac{nD}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Distribuzione Ipergeometrica

La distribuzione ipergeometrica è il modello di probabilità appropriato per la selezione di un campione casuale di n oggetti senza rimessa da un lotto di N oggetti dei quali D sono i non conformi o difettosi.

In questo caso, x rappresenta solitamente il numero di elementi non conformi trovati nel campione.

Distribuzione Ipergeometrica

Ad esempio, supponiamo che un lotto contenga 100 articoli, 5 dei quali non conformi ai requisiti.

Se vengono selezionati 10 articoli a caso senza rimessa, la probabilità di trovare uno o meno articoli non conformi nel campione è :

$$\mathrm{P}\{x \leq 1\} = P\{x = 0\} + P\{x = 1\} = \frac{\binom{5}{0}\binom{95}{10}}{\binom{100}{10}} + \frac{\binom{5}{1}\binom{95}{9}}{\binom{100}{10}} = 0.923$$

Distribuzione Binomiale

Consideriamo un processo che consiste in una sequenza di n prove indipendenti, il cui risultato è o un "successo" o un "fallimento".

Tali prove sono chiamate prove di Bernoulli.

Per prove indipendenti, si intende che il risultato di ciascuna prova non dipende in alcun modo dal risultato delle prove precedenti.

Distribuzione Binomiale

Se la probabilità di "successo" in qualsiasi prova (p) è costante, allora il numero di "successi" x in n prove di Bernoulli ha la distribuzione binomiale con i parametri n e p, definiti come segue:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

Distribuzione Binomiale

La distribuzione binomiale viene utilizzata frequentemente nel controllo di qualità.

È il modello di probabilità appropriato per il campionamento da una popolazione infinitamente grande, in cui p rappresenta la frazione di elementi difettosi o non conformi nella popolazione.

In queste applicazioni, x rappresenta in genere il numero di articoli non conformi trovati in un campione casuale di n articoli.

Distribuzione Binomiale

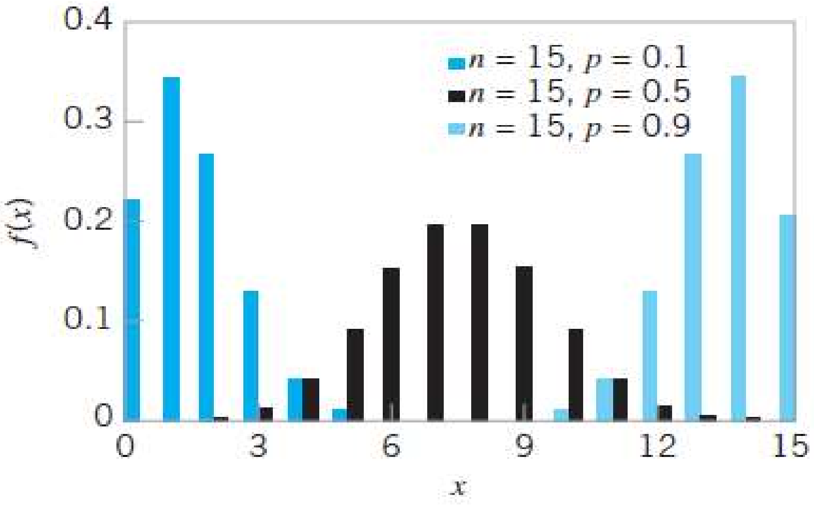
Ad esempio, se p = 0,10 e n = 15, la probabilità di ottenere x elementi non conformi è:

x	P(X=x)
0	0.2059
1	0.3432
2	0.2669
3	0.1285
4	0.0428
5	0.0105
6	0.0019
7	0.0003
8	0.0000
9	0.0000
10	0.0000
…	…
15	0.0000

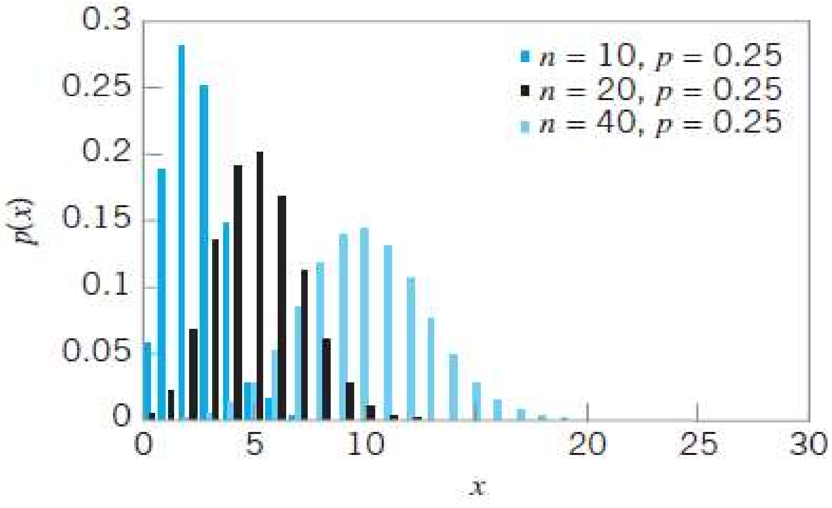
Per tutti i valori di x che si trovano tra $10 \leq x \leq 15$ la probabilità di trovare x "successi" in 15 prove è zero.

Distribuzione Binomiale

Questi esempi riportano la tipica forma delle distribuzioni binomiali.



(a) Binomial distributions for different values of p with n = 15.



(b) Binomial distributions for different values of n with p = 0.25.

Fissato un numero n, la distribuzione diventa più simmetrica quando p aumenta da 0 a 0,5 o diminuisce da 1 a 0,5.

Fissato p, la distribuzione diventa più simmetrica all'aumentare di n.

Distribuzione Binomiale

Una variabile casuale che si presenta frequentemente nel controllo statistico della qualità è:

$$\hat{p} = \frac{x}{n}$$

dove x ha una distribuzione binomiale con i parametri n e p.

Distribuzione Binomiale

Spesso $\hat{p}$ è il rapporto tra il numero osservato di articoli difettosi o non conformi in un campione (x) e la dimensione del campione (n).

Questo di solito viene chiamato frazione campionaria di elementi difettosi o non conformi.

Il simbolo "ˆ" viene utilizzato per indicare che è una stima del valore vero, sconosciuto del parametro binomiale p.

Distribuzione Ipergeometrica o Binomiale

La distribuzione ipergeometrica è una distribuzione di probabilità discreta che descrive l'estrazione senza reinserimento di alcune palline, bianca o nera, da un'urna.

L'estrazione con reinserimento (la pallina estratta viene rimessa nell'urna) viene invece descritta dalla distribuzione binomiale, la quale lascia invariata la probabilità di estrarre in seguito una pallina ad esempio bianca.

Distribuzione Ipergeometrica o Binomiale

Ad esempio, estraendo 5 palline da un'urna che ne contiene 3 bianche e 7 nere, il numero di palline bianche estratte è descritto dalla distribuzione ipergeometrica.

A differenza del caso della distribuzione binomiale, ciascuna delle n estrazioni **non** sono indipendenti l'una dalle altre.

Distribuzione Poisson

Un'utile distribuzione discreta nel controllo statistico della qualità che conta il numero di difetti per unità di misura è la distribuzione di Poisson, definita come segue:

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

X=0,1,…

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

Distribuzione Poisson

Si applica ad eventi **isolati** che accadono un certo numero di volte, in un dato intervallo di tempo (o di spazio), con una **velocità media costante.**

Esempio:

- n di guasti in un certo periodo
- n di difetti in un cavo
- n di imperfezioni in una pezza
- n di particelle contaminanti in un volume

Distribuzione Poisson

Sotto le seguenti condizioni:

- la probabilità di presentarsi di un difetto non dipende dal presentarsi o meno di nessuno degli altri difetti;
- ogni difetto ha la stessa importanza.

Distribuzione Poisson

Un'applicazione tipica della distribuzione di Poisson nel controllo di qualità è un modello del numero di difetti o non conformità che si verificano in una unità di prodotto.

Infatti, qualsiasi fenomeno casuale che si verifica in una unità (o per unità di area, per volume di unità, per unità di tempo, ecc.) è spesso ben approssimato dalla distribuzione di Poisson.

Distribuzione Poisson

È possibile costruire carte di controllo sia per il numero totale di non conformità per unità prodotta sia per il numero medio di non conformità per unità prodotta.

Per tali carte si può affermare che le non conformità contenute in un campione abbiano distribuzione di Poisson, specie se la probabilità di osservare un difetto è piccola, costante e se l'unità campionaria di riferimento è la stessa per ogni campione.

Tali ipotesi sono di difficile realizzazione anche se l'approssimazione alla Poisson è in genere ampiamente accettabile.

Distribuzione Poisson

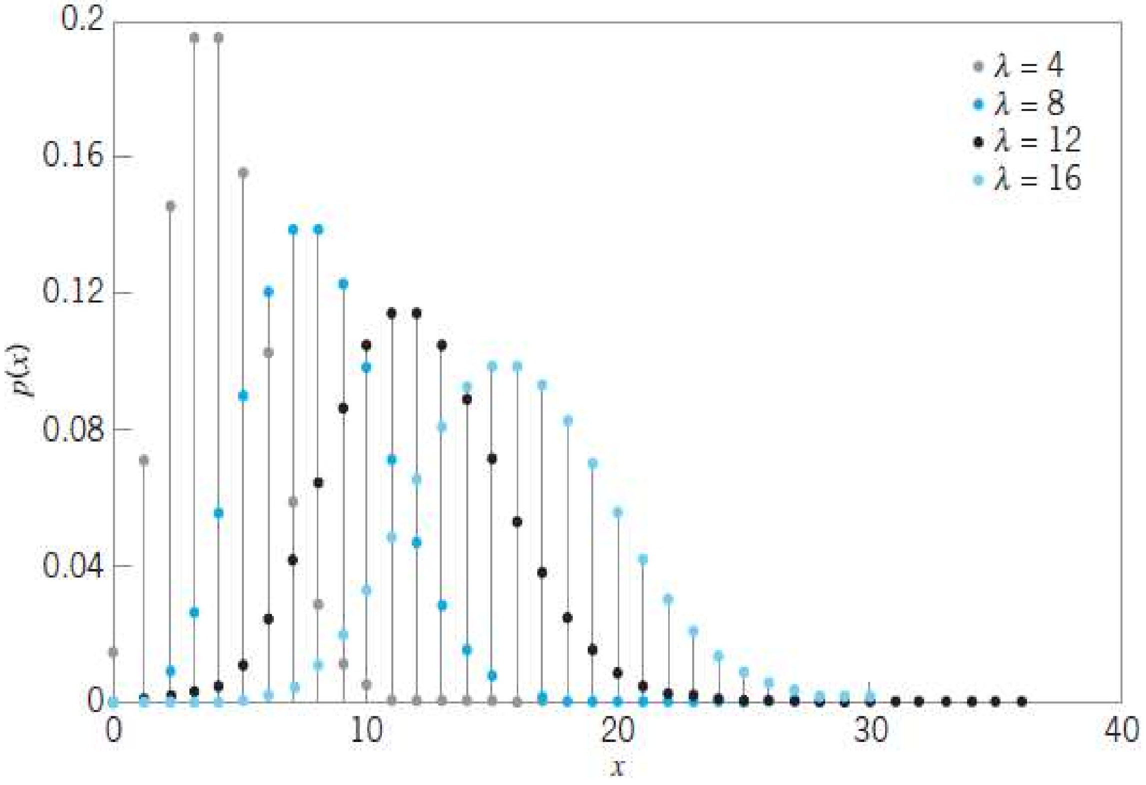
Ad esempio, supponiamo che il numero di difetti di collegamento dei fili per unità che si verificano in un semiconduttore sia distribuito secondo una Poisson con il parametro λ =4.

Quindi la probabilità che un semiconduttore selezionato casualmente conterrà due o meno difetti di collegamento dei fili è:

$$P\{x \leq 2\} = \sum_{x=0}^{0} \frac{e^{-4} 4^{x}}{x!} = 0.018 + 0.073 + 0.147 = 0.238$$

Distribuzione Poisson

Distribuzioni di probabilità di Poisson per valori selezionati di λ.



Si noti che la distribuzione è inclinata; cioè, ha una lunga coda a destra.

Man mano che il parametro diventa più grande, la distribuzione di Poisson diventa più simmetrica.

Distribuzione Normale

La distribuzione normale è probabilmente la distribuzione più importante sia nella teoria che nell'applicazione della statistica.

Molti fenomeni nella realtà si distribuiscono secondo una tipica forma a campana, con la maggioranza della popolazione concentrata nel mezzo e con dei valori che scendono gradualmente man mano che si va verso gli estremi a sinistra e a destra.

Distribuzione Normale

Se x è una normale variabile casuale, la distribuzione di probabilità di x è definita come segue:

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

$-\infty < x < +\infty$

La media della distribuzione normale è μ $(-\infty < \mu < +\infty)$ e la varianza è $\sigma^2 > 0$.

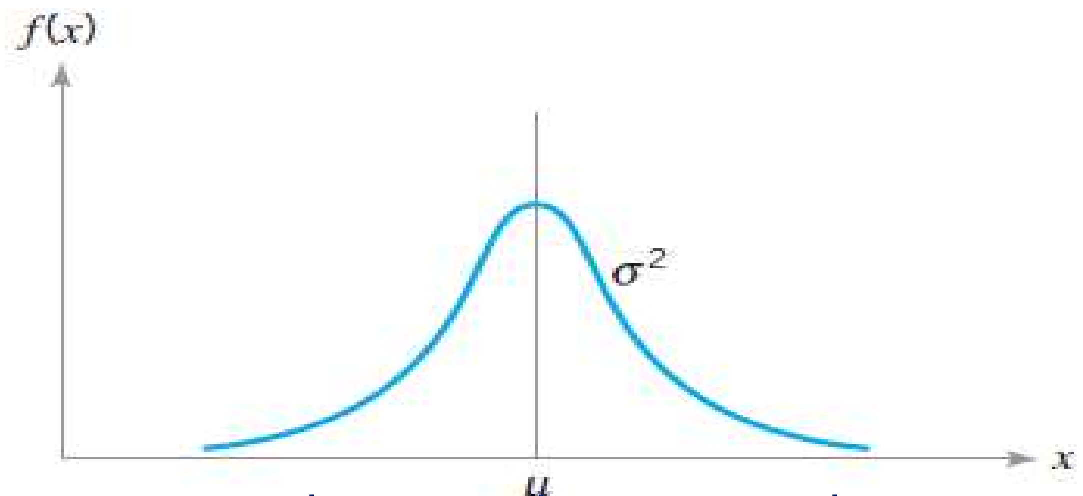
La distribuzione normale è utilizzata tanto e la sua notazione è:

$$x \sim N(\mu, \sigma^2)$$

Cioè, x si distribuisce normalmente con media μ e varianza σ^2.

Distribuzione Normale

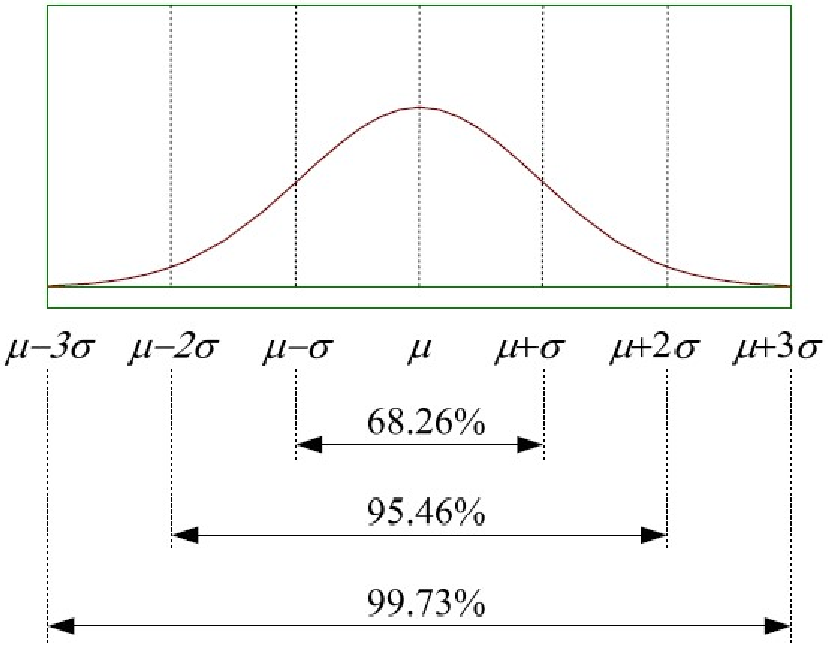
L'aspetto grafico della distribuzione normale è una curva simmetrica, unimodale o a campana.



Media = Mediana = Moda

Distribuzione Normale

Aree sottese dalla distribuzione normale



Il 68.26% dei valori della popolazione è contenuto nei limiti definiti dalla media meno o più **1** deviazione standard.

Il 95.46% dei valori della popolazione è contenuto nei limiti definiti dalla media meno o più **2** deviazioni standard.

Il 99.73% dei valori della popolazione è contenuto nei limiti definiti dalla media meno o più **3** deviazione standard.

Distribuzione Normale

Il tempo per risolvere i reclami dei clienti è una qualità fondamentale caratteristica per molte organizzazioni.

Supponiamo che in un'organizzazione finanziaria, il tempo per risolvere i reclami (x) sia normalmente distribuito con $\mu = 40$ ore e σ= 2 ore.

Qual è la probabilità che un reclamo del cliente venga risolto in meno di 35 ore?

Distribuzione Normale

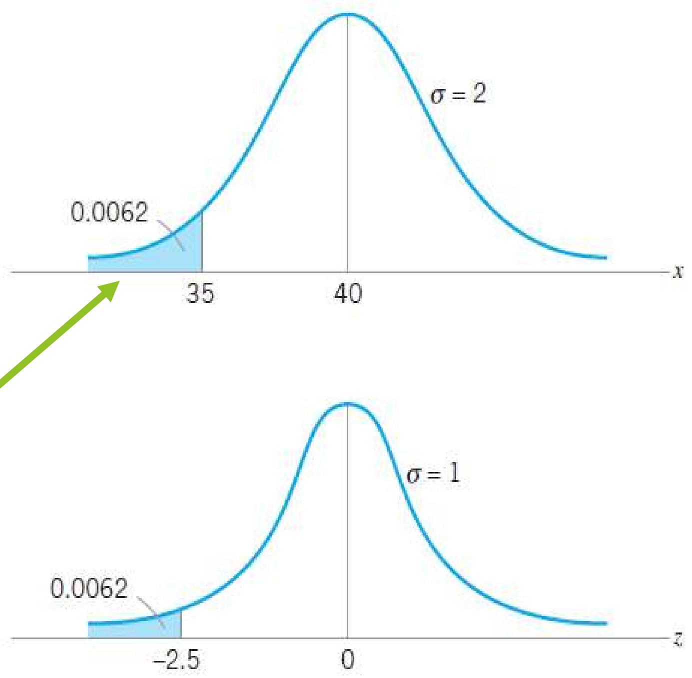
La probabilità desiderata è $P\{x \leq 35\}$.

Per valutare questa probabilità dalla tavola della normale standardizzata, standardizziamo il punto 35 e troviamo che la probabilità desiderata è:

$$P\{x \leq 35\} = P\left\{z \leq \frac{35-40}{2}\right\} =$$

$$P\{z \leq -2.5\} = \Phi(-2.5) = 0.0062$$

L'area ombreggiata a sinistra di 35 ore rappresenta la frazione dei reclami dei clienti risolti in meno o uguale a 35 ore.



Distribuzione Normale

Il diametro di un albero metallico utilizzato in un'unità a dischi è normalmente distribuito con una media di 0,2508 cm e una deviazione standard di 0,0005 cm.

Le specifiche sull'albero sono state stabilite come 0,2500 ± 0,0015 cm (ossia 0.248 e 0.2515).

Quale frazione degli alberi prodotti è conforme alle specifiche?

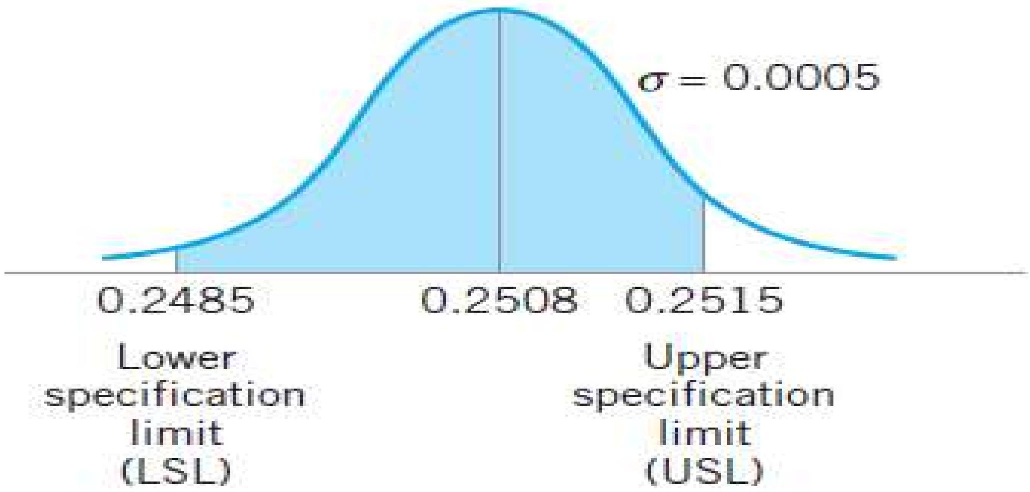
Distribuzione Normale

$$P\{0.2485 \le x \le 0.2515\} = P\{x \le 0.2515\} - P\{x \le 0.2485\}$$

$$= \Phi\left(\frac{0.2515 - 0.2508}{0.0005}\right) - \Phi\left(\frac{0.2485 - 0.2508}{0.0005}\right) = \Phi(1.40) - \Phi(-4.60)$$

$$= 0.9192 - 0.000 = 0.9192$$

Circa il 91,92% degli alberi prodotti è conforme alle specifiche.



Si noti che quasi tutti gli alberi non conformi sono troppo grandi, poiché la media del processo si trova molto vicino al limite superiore delle specifiche.

Distribuzione Normale

Se si aggiornasse il processo di produzione, magari regolando la macchina, in modo che la media del processo sia esattamente uguale al valore nominale di 0,2500, si avrebbe che:

$$P\{0.2485 \leq x \leq 0.2515\} = P\{x \leq 0.2515\} - P\{x \leq 0.2485\} = \Phi\left(\frac{0.2515-0.2500}{0.0005}\right) -$$

$$\Phi\left(\frac{0.2485- \ .2500}{0.0005}\right) = \Phi(3.00) - \Phi(-3.00) = 0.99865 - 0.00135 = 0.9973$$

Con il riadattamento della macchina, la resa del processo aumenta a circa il 99,73%.

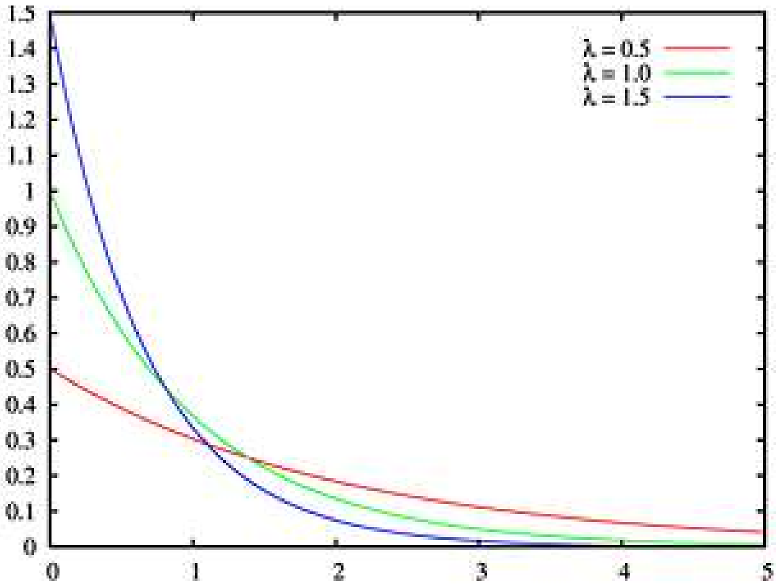
Distribuzione esponenziale

La distribuzione di probabilità della variabile casuale esponenziale è definita:

$f(x) = \lambda e^{-\lambda}$ x≥ 0

$\mu = \frac{1}{\lambda}$

$\sigma^2 = \frac{1}{\lambda^2}$



Distribuzione esponenziale

La distribuzione esponenziale è ampiamente utilizzata nel campo dell'ingegneria dell'affidabilità per modellizzare il **tempo di fallimento di un componente o sistema, oppure i tempi di vita**, di funzionamento, di attesa per sistemi di vario tipo.

In queste applicazioni, il parametro λ è chiamato **tasso di fallimento** del sistema e la **media** della distribuzione **1/**λ è chiamata **tempo medio al guasto**.

Distribuzione esponenziale

Ad esempio, supponiamo che un componente elettronico montato su un aereo abbia una durata descritta da una distribuzione esponenziale con tasso di guasto $10^{-4}/h$ ovvero $\lambda = 10^{-4}$.

Il tempo medio al guasto per questo componente è

$$\frac{1}{\lambda} = 10^{4} = 10.000 \text{ ore}$$

Distribuzione esponenziale

Se volessimo determinare la **probabilità** che questo componente **si guasti prima della sua durata attesa**, essa sarà pari a

$$P\left\{x \leq \frac{1}{\lambda}\right\} = \int_0^{1/\lambda} \lambda e^{-\lambda t} dt = 1 - e^{-1} = 0.632$$

Questo risultato **non dipende dal valore di λ**, cioè la probabilità che un valore di una variabile casuale esponenziale sia inferiore alla sua media è 0,632.

Questo accade, ovviamente, perché la distribuzione non è simmetrica.

Distribuzione esponenziale

Una proprietà importante della legge esponenziale è rappresentata dalla cosiddetta **assenza di memoria**.

Facciamo ad esempio riferimento al tempo di vita di un sistema.

La proprietà di assenza di memoria si manifesta quando **qualunque sia il tempo trascorso, il tempo residuo di vita non è affetto dal passato e ha la stessa distribuzione del tempo di vita originario**.

In altre parole, **l'oggetto non subisce logoramento**, per cui la sua propensione statistica a rompersi resta invariata.

Ovviamente da questo si vede che l'ipotesi di esponenzialità è piuttosto ideale nella pratica.

Distribuzione esponenziale e di Poisson

Esiste una **relazione** importante tra la distribuzione esponenziale e quella di Poisson.

Se si considera la distribuzione di Poisson come un modello del **numero di volte in cui un evento si determina in un intervallo di tempo (0,t]**, allora:

$$p(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

Se x=0, significa che non ci sono accadimenti dell'evento nell'intervallo (0,t] e che $P\{x = 0\} = p(0) = e^{-\lambda t}$.

Distribuzione esponenziale e di Poisson

Possiamo ritenere che **p(0)** rappresenti la probabilità che l'intervallo di tempo per un primo verificarsi dell'evento sia maggiore di t, ovvero:

$$P\{y > t\} = p(0) = e^{-\lambda t}$$

Dove y è la variabile casuale che indica l'intervallo di tempo al primo determinarsi dell'evento.

Quindi si ottiene che

$$f(y) = \lambda e^{-\lambda}$$

Che rappresenta la distribuzione dell'intervallo di tempo fino al determinarsi del primo evento.

Distribuzione esponenziale e di Poisson

$$f(y) = \lambda e^{-\lambda t}$$

Ha la forma di una **distribuzione esponenziale con parametro** $\boldsymbol{\lambda}$.

Pertanto, se il numero di volte in cui un evento si verifica si distribuisce secondo la distribuzione di Poisson con parametro λ, allora la distribuzione dell’intervallo tra il determinarsi di due eventi è esponenziale con parametro λ.

Grafici di probabilità

I grafici di probabilità sono molto utili e spesso sono la prima tecnica esplorativa usata per la scelta della distribuzione di probabilità più adatta a descrivere i dati.

Nell'usare i grafici di probabilità, la decisione è solitamente presa attraverso una valutazione soggettiva del grafico.

Grafici di probabilità

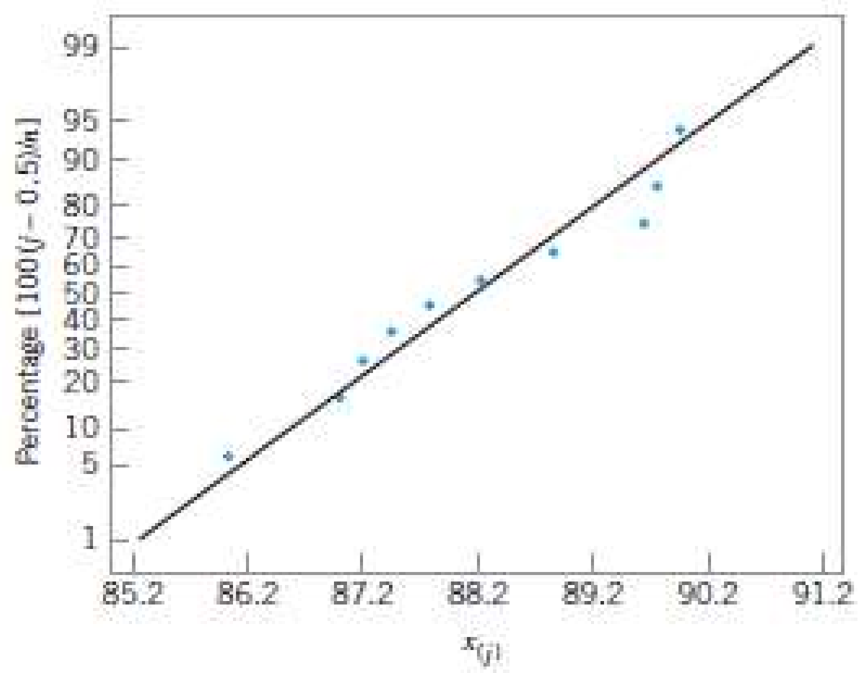
Per costruire un grafico di probabilità normale occorre dapprima **ordinare le osservazioni** dalla più piccola alla più grande.

Il grafico di probabilità viene realizzato rappresentando le **coppie** costituite da **ciascuna osservazione ordinata e dalla corrispondente frequenza cumulata**.

Se i punti del grafico si trovano approssimativamente su una linea retta immaginaria inclinata positivamente, allora si può affermare che i dati osservati si distribuiscono approssimativamente secondo la legge **normale**.

Se i punti disegnati deviano significativamente e sistematicamente da una linea retta, il modello ipotizzato non è appropriato.

Grafici di probabilità



Asse delle ordinate: % di probabilità cumulate

Asse delle ascisse: valori ordinati assunti dalla variabile casuale