RIASSUNTI ARGOMENTI PRINCIPALI - STATISTICA

Con il termine popolazione, si intende l'insieme dei soggetti o individui oggetto dell'osservazione statistica.
Con il termine fenomeno, si indica il carattere che si vuole osservare in una data popolazione
Con il termine campo di definizione, si indica l'insieme di tutte i caratteri assunti dal fenomeno osservato nella popolazione di riferimento.
Variabili qualitative, si considerano come variabili qualitative quelle variabili il cui insieme di definizione è dato da un insieme di caratteri, (non numerici), riscontrati presso una popolazione di riferimento. Si parla di variabili qualitative nominale, qualora si consideri una variabile qualitativa i cui carattere non possono essere ordinati. Si parla di variabili qualitative ordinali qualora sia possibile ordinare i caratteri della variabile considerata.
Variabile quantitativa, si parla di variabili quantitative qualora si consideri della variabili il cui campo di definizione è definito da valori numerici, o più in generale qualora tale campo derivi da un processo di misurazione effettuato su una popolazione di riferimento. Si parla di variabile quantitativa intervallare o discreta, qualora i valori presenti nel campo di definizione della variabile siano approssimati ad una unita tale che non sia possibile, date due qualsiasi osservazione determinarne una intermedia.
Si parla di variabile quantitativa continua, qualora i valori presenti nel campo di definizione della variabile sia misurati in un continuo, possiamo difatti determinare, date due variabili una intermedia che potrebbe appartenere all'insieme di definizione. Queste variabili sono generalmente suddivise in classi di frequenza, data l'ampia dispersione che presenta il carattere generalmente osservato, che non permetterebbe quindi di trarre conclusioni su quanto rilevato. Può essere espressa da una variabile teorica che altri non è che una variabile continua suddivisa in classi qualora si consideri un'unica classe.
Di ognuna delle variabili sopra indicate è poi possibile calcolare la frequenza relativa che rappresenta il rapporto tra la frequenza assoluta, numero delle volte in cui si presenta un determinato valore, o classe di valori, e il numero delle osservazioni effettuate, tale valore corrisponde alla popolazione.
Considerando più nello specifico poi le variabili quantitative continue è poi possibile definire la densità di frequenza assoluta e la densità di frequenza relativa, tali valori sono rispettivamente il rapporto tra la frequenza assoluta e la frequenza relativa con l'ampiezza della classe considera, l'ampiezza della classe altri non è che la differenza tra l'estremo di destra e quello di sinistra della classe stessa.
Funzione di ripartizione, funzione necessaria alla rielaborazione di dati statistici grezzi, data quindi una funzione riscritta in maniera ordinata, dati in ordine crescente, tale implicazione esclude la possibilità di calcolare la funzione di ripartizione per le variabili qualitative nominali, la funzione di ripartizione associa ad ogni carattere dell'insieme di

definizione un valore compreso tra zero e uno, estremi inclusi, che corrisponde alla frequenza dei casi pari o inferiori al carattere stesso. La funzione di ripartizione è continua sempre per i valori compresi tra meno infinito e zero e per i valori compresi tra meno infinito e zero, e per i valori compresi tra uno e più infinito, in tali intervalli la funzione è costante e vale zero nel primo caso e uno nel secondo. Ad esclusione del caso di variabili continue suddivise in classi la funzione di ripartizione, non è continua nell'intervallo zero uno, ma si presenza come una funzione a gradini in cui si osserva un balzo per ogni valore presente nel campo di definizione della variabile statistica, tra una variabile statistica e l'altra, qualora le due variabili siano attigue, la funzione è continua per un intervallo chiuso a sinistra, che ha per estremi i due valori considerati. Nel caso di una funzione continua divisa in classi la funzione di ripartizione è continua, e presenza dei una serie di punti angolosi pari e corrispondenti agli estremi delle classi.
La funzione di ripartizione è poi crescente, mai strettamente salvo nell'intervallo zero uno qualora si consideri una variabile continua suddivisa in classi.
Data la generica operazione di sintesi dei dati secondo il minimo errore, la funzione di sintesi vale zero qualora la costante C di sostituzione corrisponda al dato considerato, vale più di zero, in proporzione al grado di errore, o uno, nel caso che la variabile non corrisponda, è possibile definire i concetto di moda, inteso come il minino grado di errore sostituendo a una qualsiasi variabile un valore C e dando alla funzione di mediana un valore pari a zero qualora sia uguale al fattore sostituito e pari a uno qualora sia differente. Quanto detto può essere semplificato assumendo di moltiplicare per uno o zero, a seconda della natura dell'errore, per la frequenza relativa della variabile osservata, questo metodo permette di limitare il numero dei confronti alle modalità assunta dalla variabile di cui si vuole calcolare la media. Quanto detto è valido per tutte le variabili non numeriche e per quelle numeriche discrete, nel caso di variabili numeriche continue è necessario, per poter utilizzare lo stesso metodo, che la variabile sia ordina secondo classi di pari ampiezza, o che si sostituisca la densità di frequenza relativa alla frequenza relativa nella funzione di errore.
Va poi detto della moda che è l'unico indice di sintesi calcolabile per tutte le tipologie di dati considerati, numerici o non, anche se presenta delle discrepanze per quanto riguarda la veridicità della sintesi ottenuta, difatti dato una variabile è possibile che esistano più di una moda, o che non ne esista nessuna, tutti i valori hanno la stessa frequenza relativa, non esiste quindi una funzione di moda, cioè tutti i valori del campo di definizione sono potenziali mode.
Riprendendo la funzione di sintesi basata sul minimo errore è necessario considerare gli altri casi, quelli in cui per un dato differente da C la funzione di errore vale più di zero, in via proporzionale all'errore stesso, osservando quindi le principali funzioni che possono definire la funzione di errore è possibile determinare altre due funzioni di media avremmo quindi:
La mediana definibile come il punto di sintesi C in cui la funzione di ripartizione vale 0,5, per poter definire tale punto in tutte le variabili numeriche si può anche considerare come il valore minimo in cui la funzione di ripartizione vale 0,5 o anche il valore per cui la

funzione di ripartizione supera 0,5 per la prima volta. A differenza della moda tale valore di sintesi presenta uno spettro di utilizzo più limitato è difatti utilizzabile solamente per quelle variabili numeriche ordinate.
Il calcolo della mediana si differenzia a seconda della natura della variabile, discreta o continua e suddivisa in classi, nel primo caso abbiamo un problema relativo al fatto che tale valore o non esiste o esiste ma comprende uno spettro pressoché illimitato di valori, nel primo caso si assume la definizione secondo cui è la mediana il valore della variabile per cui la funzione di ripartizione supera 0,5 per la prima volta, nel secondo caso si ricorre alla definizione di mediana come il minimo dei punti in cui la funzione di ripartizione vale 0,5. Nel caso di una funzione continua il calcolo si complica, è necessario difatti definire la classe ove la mediana cade e quindi calcolare la funzione di ripartizione nel punto più piccolo a cui sarà sommato l'integrale della funzione di ripartizione inclusa tra il valore minore della classe e il punto 0,5. Tra le principali proprietà della mediana troviamo: il fatto che prescinde dalla variabile considerata continua o descritta ed è considerata come una media robusta poiché non risente di valori anomali agli estremi.
Riprendendo la generica funzione di ripartizione è possibile definire quello che è comunemente considerato come il più importante degli indici di sintesi cioè la media.
Come la mediana anche la media può essere calcolata unicamente su variabili numeriche divise in classi tal valore esprime la media di tutti i valori assunti dalla variabile considerata, la media si può esprimere come la sommatoria di tutti i valori del campo di definizione di un variabile per la rispettiva frequenza relativa. L'importanza della media come valore di sintesi dipende oltre che dalla sua capacità esplicativa anche dalla proprietà di cui gode tale valore difatti della media si può dire che: è una media dotata di baricentro si può difatti affermare che data la sommatoria di tutti gli scarti dalla media tale valore ha come risultante zero, è dotata poi della proprietà dell'internalità la media è difatti un valore maggiore del minimo e minore del massimo valore assunto dalla variabile osservata, è dotata poi della proprietà della traslabilità data difatti una media è possibile moltiplicare tutti i valori del campo di definizione per una variabile, o sommarli ad una costante, per far si che la media venga traslata della stessa dimensione, non soffre difatti di eventuali cambi dell'unità di misura o di traslazioni, la media è poi dotata della proprietà associativa, suddividendo la media in k gruppi, data la media di ogni gruppo la media complessiva dei k gruppi corrisponde alla media originale.
Data la generalizzazione della funzione di errore, in particolare considerando il caso in cui C non corrisponda al valore della variabile con cui la si sostituisce e considerando una generalizzazione della funzione di media nota come media potenziata di ordine s abbiamo una funzione generale di distribuzione da cui derivano in linea di massima tutte le funzioni di media, tale valore, s, si pone generalmente diverso da zero, date le proprietà delle esponenziali, ma analizzando il limite che tende a tale valore la funzione è definita e il limite esiste si a destra che a sinistra. A seconda del valore di

s la media prende un nome differente, per s=-1 avremo la media armonica, per s=0 avremo la media geometrica, per s=1 avremo la media aritmetica, per s=2 avremo la media quadratica, per s=3 avremo la media cubica. La media potenziata di ordine s è poi dotata di una serie di proprietà tra cui troviamo: l'internalità cioè la media potenziata è compresa tra il valore massimo e quello minimo assunti dalla variabile considerata e la monotonicità, cioè la funzione di media ha lo stesso andamento della variabile s, anche se non necessariamente della stessa quantità.
Passando dagli indici di sintesi basati sul criterio del minimo errore, agli indici di sintesi basati sul criterio delle scelte vincolate ad un obbiettivo è necessario analizzare, di quest'ultime:
L'indice di sintesi di Chisini: tale indice di sintesi prevede, oltre alla possibilità di formalizzare un insieme di numeri come espressione di un numero indice, sintesi dei numeri stessi, di legare tale indice di sintesi ad una funzione obbiettivo propria dell'insieme di numeri stesso.
Dato un indice di sintesi può rivelarsi necessario poter verificare l'effettiva capacità di tale indice di riassumere l'insieme di valori considerato, per ottemperare a questa necessità vengono utilizzati dei particolari indici di verifica che permettono di stabilire se una costante sintetizza o meno un insieme di dati. Per ottemperare a questa necessità esistono due tipi differenti di approcci che considerano due modalità con cui si può affrontare il problema: la prima modalità o variabilità da un centro punta a verificare la capacità di sintesi di un indice riferendosi a un punto focale utilizzato come centro e verificando gli scostamenti dei valori della funzione da tale valore, tale approccio è detto variabilità da un centro, un altro approccio prevede l'analisi della mutua variabilità cioè di osservare quanto tra loro le variabili sono differenti da una costante c di sintesi, all'aumentare di tale indice diminuisce la capacità di sintesi dell'indice considerato all'inizio.
Analizzando la variabilità da un centro risulta evidente come la natura dell'indice che ci troveremo a calcolare dipende essenzialmente dall'indice di riferimento che funge difatti da centro:
Considerando la mediana come indice di riferimento avremo un indice detto: scarti assoluti dalla mediana, che si prevede di misurare la sommatoria di tutte le differenze tra i valori dell'insieme definizione e la mediana, espressi in valore assoluto, e moltiplicati per la frequenza relativa della variabile considerata.
Considerando la media come indice di riferimento avremo due indici uno noto come varianza, e un altro noto come scarto quadratico medio, derivante dalla radice quadrata della varianza. Tale valore deriva dalla sommatoria delle differenze tra i valori del campo di definizione di x e la relativa media, elevati alla seconda, e moltiplicati per la frequenza relativa. Tra le proprietà della varianza ricordiamo che è considerata un indice meno relativo per quanto riguarda la verifica di una funzione di sinesi e quindi più veritiero. La varianza gode poi di tutte le proprietà riguardanti, la

traslabilità, il cambio di unità d misura, e l'applicazione congiunta di queste due proprietà, come già visto per la media.
Continuando l'analisi dei possibili indici di verifica di una funzione di sintesi, è necessario prendere in considerazione il metodo della mutua variabilità, tale metodo prevede l'analisi della media delle differenze tra i valori assunti dalla funzione di riferimento, onde a verificarne le capacità di sintetizzare la funzione stessa. Tale indice deve quindi rispettare dei parametri di riferimento, deve difatti essere zero in presenza di un errore pari a zero, e deve assumere valori crescenti superiori a zero qualora la capacità di sintesi dell'intera funzione considerata. Si costruisce una tabella a doppia entrata, da un parte e dall'altra si inseriscono i valori della funzione, e nel mezzo le varie differenze, semplificando la tabella eliminando righe e colonne uguali, moltiplicando però i valori della colonna o della riga risultante per il numero di righe o colonne eliminato avremo un valore finale, frutto della somma di tutte le righe rimanenti, o tutte le colonne, se diviso per il numero delle operazioni svolte, quadrato delle dati osservati, avremo un indice che però ha valore sempre pari a zero, considerando però, invece delle differenze stessa, il loro valore assoluto, avremo un indice di sintesi che risponde alla caratteristiche necessarie definite all'inizio, esso ha difatti valori pari a zero per variabili banali e crescenti al diminuire del grado di sintesi del sistema.
Da quanto detto risulterà quindi un indice noto come indice di differenza media assoluta con ripetizione. Osservando però la tabella di riferimento si può osservare come sulla diagonale principale i valori considerati sia sempre uguali a zero e quindi escludibili dal calcolo prima considerata, avremo quindi al numeratore n^2-n, determinando siffatto l'indice di differenza media assoluta senza ripetizione. Quanto detto può essere anche scritto, oltre che come risultante della differenza dei valori assunti da una variabile della differenza tra i valori assoluti delle medie della classe considerata per il numero delle osservazioni, che considerando la funzione con ripetizione può essere semplificato considerando invece del numero delle osservazioni la loro frequenza relativa.
Del metodo appena delineato per la determinazione della verifica di un indice di sintesi è possibile, oltre alla differenzia media assoluta, deriva tutto un insieme di altri indici che genericamente derivano dall'indice s a cui può essere elevata la funzione con cui si esplicita l'errore nella sintesi, i nomi degli indici considerati sono gli stessi derivanti dalla funzione di media, è poi possibile determinare il range della variabile considerando la differenza tra il valore minimo e quello massimo assunto, il mino si determina tramite il limite della funzione a zero.
Un indice simile al range deriva dal calcolo dei quantili, valori da cui deriva anche il calcolo della mediana e che permetto di esprime il valore di una data funzione ad un prestabilito valore di frequenza relativa. In particolare considerando i quartile, cioè quei quantili che dividono la distribuzioni in quattro parti uguali, è interessante osservare la differenza tra il primo e il terzo quartile, tale indice noto come

differenza interquantile permette di analizzare la funzione di riferimento escludendo eventuali valori anomali o non presenti nella coda delle osservazioni.
Considerando gli indici di mutua variabilità un ultimo che andrebbe preso in considerazione è dato dal rapporto tra lo scarto quadratico medio e la media, tali indici difatti sono espressi nella stessa unità di misura, la risultante quindi non sarà influenzata dall'unità di misura e quindi sarà possibile confrontare tale indice anche tra due variabili che premettono unità di misura differenti, elemento di debolezza che ha influenzato tutti gli indici d variabilità da un centro e mutua variabilità sino ad ora considerati.
Continuando l'analisi della variabilità un altro degli elementi da tenere in considerazione è un ipotetico indice in grado mi misurare la variabilità stessa.
Tale indice starebbe ad indicare la tendenza di un valore a concentrarsi o distribuirsi nell'ambito della popolazione di riferimento, tale analisi non si presenta possibile per tutti gli indici ma solamente per quelli che possono essere trasferiti da un soggetto all'altro, in tal caso diventa difatti interessante poter scoprire il comportamento, concentrazione o meno di tale valore. Per svolgere tale processo è necessario definire le caratteristiche che tale indice deve avere, deve difatti corrispondere a zero per valori equamente distribuiti, e deve avere valori crescenti per l'aumentare del grado di concentrazione sino al grado massimo, valore concentrato nelle mani di un singolo individuo. per fare ciò si lavoro con la curva di laurence, uno schema grafico che racchiude la retta di equi distribuzione, bisettrice del quadrante e la spezzata che indica la concentrazione del valore considerato. Un indice che ben si presta a tale scopo è l'area compresa tra le due curve (data al curva di laurence come sistema di riferimento), spezzata di concentrazione e retta di equi distribuzione, tale valore difatti aumenta all'aumentare della concentrazione, ed è zero qualora le due rette corrispondono, per far ciò si calcola l'area compresa tra la retta e l'asse delle x, dato quindi che si considerano le distribuzione di frequenze relative, la retta di equi distribuzione ha come vertice il punto 1.
Dato il vertice si ha un lato che è corrispondente ad uno e lo stesso vale per l'altezza, determinando quindi un'area di $\frac{1}{2}$, a questa va sottratta l'area sottostante la spezzata di distribuzione, quantificabile come un triangolo sino al primo valore della retta, e come una serie di trapezi per tutte le sezioni successiva, che vanno dalla prima alla seconda, dalla seconda alla terza ecc. si deve quindi sottrarre la sommatoria delle aree sottostanti la spezzata di distribuzione per determinare l'indice di riferimento Fi. Va infine sottolineato che l'indice Fi è tale che $0<Fi<Fi_{MAX}$, è possibile quindi "schiacciare il valore Fi dividendo tutti i membri della precedente uguaglianza per uno stesso valore, Fi_{MAX}, definendo quindi la seguente situazione $0<Fi/Fi_{MAX}<1$.
Analisi statistica di due variabili di riferimento, si considera in questo caso, data una popolazione di riferimento le rilevazioni nella stessa di due variabili, di qualsiasi natura e genere, la popolazione deve essere sempre la stessa, per quanto campionata. Per costruire tale analisi è necessario riorganizzare quindi i dati in una tabella a doppia entrate, in un lato inseriremo il campo di definizione della variabile x e nell'altro il campo di definizione

della variabile y, nei punti ove si incontrano x e y si inserirà la frequenza con cui le relativa x e y si presentano congiuntamente.
Sarà quindi possibile definire dei totali di riga e colonna, tali valori indicano quante volte complessivamente si presenta la variabile x o y considerata, e si chiamano frequenze marginali, il totale delle frequenze marginali definirà infine il numero totale delle osservazioni effettuate. È facile poi passare dalle distribuzione di frequenza assoluta a quella relativa, è sufficiente difatti dividere il valore iscritto in ogni distribuzione congiunta per il totale delle osservazioni. È poi possibile scomporre l'intera tabella considerando solamente una riga o colonna, e la totalità delle modalità delle colonne, se si considera solamente una riga e delle righe se si considera solamente una colonna, determinando quindi le distribuzioni di frequenza condizionate della modalità presente nella riga alle variabile espresse nelle colonne o viceversa, il rapporto tra le singole distribuzioni di frequenze, assoluta, rapportate con il totale di riga o colonna determinano le distribuzioni di frequenza marginali condizionate.
Lo scopo dell'analisi bivariata risulta quindi essere quello di osservare come, al variare di una delle due variabile si comporta l'altra, e quindi il grado di influenza reciproca presente nel sistema considerata.
La parte più semplice di questa analisi è la verifica dell'indipendenza delle due variabili.
Tale processo di verifica inizia partendo da una tabella che esprime sulle colonne le modalità che può esprimere la variabile di cui si vuole analizzare l'indipendenza e sulle righe le modalità dell'altra, nella tabella vengono poi inserite le distribuzioni di frequenza relative condizionate alla variabile di riferimento, e nell'ultima si inseriscono le distribuzioni condizionate di frequenza relativa marginali, qualora tutti i valori presenti nella tabella siano costanti, per ordine di riga, si parlerà di indipendenza della variabile espressa nelle colonne rispetto a quella espressa nelle righe, l'indipendenza di una variabile non comporta necessariamente anche il processo inverso, è necessario difatti effettuare un processo di verifica.
Semplificando il processo appena descritto è possibile ridurre l'operazione di confronto ad un paragone tra la frequenza relativa di un valore di una riga con la frequenza relativa marginale della riga considerata, esplicitando l'uguaglianza, operando le opportune rielaborazioni matematiche, è possibile definire che si ha indipendenza lineare di una variabile all'altra qualora la densità di frequenza congiunta assoluta sia uguale al prodotto del totale di riga e di colonna, ove troviamo la variabile considerata, divise per n.
È quindi possibile costituire una tabella che contenga tutti questi valori, ove al posto del generico valore n_{ij} densità di frequenza assoluta, troveremo e_{ij}, valore che esprime la risultante del processo di calcolo sopra considerato. Per esprimere quindi la dipendenza o indipendenza delle variabili condizione è quindi possibile definire un indice chi-quadro, che rappresenti la sommatoria, per tutte le modalità osservate, sia di x che di y, degli scarti al quadrato tra il valore della densità di frequenza assoluta

osservato, meno l'indice che indica calcolato tramite la condizione di indipendenza e_{ij}, elevati al quadrato e rapportati all'indice e_{ij} stesso.
Analizzando tale indice la prima cosa che risulta evidente è come esso perda capacità esplicativa qualora sia considerino dati che hanno la stessa densità di frequenza relativa, ma frequenze assolute differenti, difatti tali dati dovrebbero avere lo stesso grado di indipendenza/dipendenza, per risolvere questo problema è necessario definire un indice che nel nostro caso chiameremo phi-quadro e che sia immune da questa influenza, per far ciò il modo più semplice è definire l'indice phi-quadro come il rapporto tra chi-quadro e il numero n delle condizioni osservate. Tale indice, phi-quadro, è anche matematicamente ricavabile dall'indice chi-quadro dopo opportune rielaborazioni.
Dato quindi l'indice phi-quadro anch'esso può essere schiacciato tra zero ed uno, dati i limiti normali che sarebbero zero e phi-quadro massimo, è sufficiente dividere tutti i valori dell'uguaglianza per phi-quadro massimo, in questo modo si definisce anche un indice in grado di determinare la forza con cui le due variabili si attraggono, va infine considerato che phi-quadro massimo varia a seconda della variabile che rasenta difatti il massimo di phi-quadro, il discorso può quindi essere generalizzato considerando come divisore di phi-quadro non il numero minimo tra le colonne meno uno e le righe meno uno, considerando difatti un sistema in condizioni di massima dipendenza tale risulterebbe il valore di phi-quadro.
L'analisi sino ad ora svolta è valida a prescindere dalla natura stessa delle variabili, considerando più nello specifico delle varabili numeriche, o almeno uno delle due variabili come numerica, è possibile amplificare lo spettro del discorso, inserendo quindi la dipendenza/indipendenza in media di due variabili, è difatti possibile calcolare la media assoluta e condizionata, variabile calcolabile in funzione delle righe o delle colonne, delle due variabili, tale valore se rapportato con la "media teorica", definisce il grado di dipendenza in media delle due variabili.
Analizzando più approfonditamente la dipendenza in media è possibile, come per la dipendenza in valori assoluti, definire un indice in grado di quantificare la dipendenza delle due variabili, tale indice noto come n^2, e dato dal rapporto tra la media delle varianze e la varianza delle medie, data dallo scostamento delle media condizionata di y secondo x_i e la media di y.
L'analisi appena considera, valida si qualora si considerino anche due variabili quantitative, presenta però il limite di analizzare la variabilità di una variabile rispetto all'altra, tralasciando la possibilità di un'analisi congiunta, per sviluppare questo tipo di analisi è possibile procedere secondo due metodi, tali metodi partono dall'assunzione generale che si deve determinare una variabile che influenza e una che viene influenzata, fatto ciò si procede come segue:
Nel primo caso o spezzata di regressione è necessario costruire uno schema che riassuma le medie condizionate prima della variabile x al variare di y e poi viceversa, fatto ciò è necessario rappresentare i punti identificati sugli assi cartesiani e determinare quindi una

retta; tale retta nota come spezzata di regressione rappresenta l'andamento e quindi l'influenza che una variabile esercita sull'altra.
Data la spezzata di regressione è possibile verificarne la veridicità tramite l'indice n^2, che incluso tra zero e uno, indica in percentuale il grado di veridicità della spezzata di regressione, tale metodo presenta molti limiti per quanto riguarda la determinazione dell'influenza delle due variabili, non è difatti possibile correggere n^2 e determinare una spezzata con una capacità "migliore" di rappresentare tale relazione, oltretutto tale metodo è considerato estremamente "grezzo" per determinare la relazione tra due variabili.
Un metodo decisamente più valido, in quanto malleabile e riproducibile qualora i risultati ottenuti non fossero consoni all'analisi, o al grado di precisione, che si voglia raggiungere è il metodo dei minimi quadrati; tale metodo si struttura in più punti in primis si determina una funzione z che sintetizza la relazione tra x e y, inserendo opportunamente anche valori esterni al sistema x e y, costanti ed espressi con altre lettere dell'alfabeto; la seconda parte del processo consiste nel determinare queste variabili ricorrendo appunto al metodo dei minimi quadrati, tale metodo da il nome all'intero sistema che si sta considerando, la terza parte consiste nella determinazione dell'indice R^2, tale indice valuta il grado di veridicità delle operazioni svolte.
È infine possibile ripetere l'intero processo per un qualsiasi numero di volte, onde ad ottenere un maggiore grado di veridicità delle variabili ottenute.
Continuando l'analisi quindi secondo la metodologia dei minimi quadrati è necessario in primis definire una ipotetica relazione che leghi le due variabili, ipotizzando che tale relazione sia: $y=a+bx$, è quindi possibile determinare un valore di y reale, come da dati osservati e uno teorico, traslato in relazione alla funzione appena espletata, dalla differenza tra questi due valore è possibile determinare un valore noto come e_{ij}, volendo quindi determinare tale valore, in maniera assoluta e non relativa ad un valore di y, è sufficiente determinare la sommatoria tra tale valore alla seconda per p_{ij}, densità di frequenza relativa, tale funzione complessivamente altri non è che un'operazione di media. Operando opportune semplificazioni matematiche è possibile infine giungere alla conclusione che data una relazione lineare i valori di a e b della nostra funzione altri non sono che la covarianza diviso la varianza alla seconda meno la varianza di x alla seconda, per b, e la media di y meno la media di x per b, considerando la variabile a.
Considerando quanto detto sull'analisi del metodo dei minimi quadrati considerando una funzione lineare e le due relative determinanti, nel nostro caso a e b è possibile sostituire l'indice R^2 con un ipotetico indice p che risulta di più facile determinazione, tale processo prevede modifiche dell'indice R^2 sino a giungere alla conclusione che tale indice è determinabile come il rapporto tra la varianza e il prodotto della varianza x per la varianza di y.

Il calcolo delle probabilità si occupa della determinazione del generico grado con cui può verificarsi un dato evento, detta in termini più specifici la probabilità quantifica il grado con cui un evento può essere vero o falso. Data una probabilità pari a zero, avremo quindi un numeratore h, eventi favorevole alla nostra affermazione, pari a zero, data una probabilità pari ad uno, numerato h e denominatore n uguali, avremo un evento certo. Va poi detto che non è sempre possibile calcolare la probabilità di un evento, qualora ad esempio gli eventi considerati non siano a due a due escludibili, quindi il verificarsi dell'uno non preclude al verificarsi dell'altro, o qualora vi sia un numero complessivamente infinito di eventi, qualora difatti gli eventi non siano quantificabili.
Dato il generico calcolo probabilistico è possibile ricondurre le varie metodologie di calcolo a differenti scuole di "pensiero", relative al calcolo probabilistico: in primis abbiamo la scuola classica che definisce il calcolo probabilistico come il rapporto tra i casi favorevoli ad un data affermazione, h, e il numero di casi osservabili, questo approccio presenta tra gli altri alcuni inconvenienti legati all'impossibilità di determinare una ipotetica base di riferimento n, espressione dei casi totali possibili; per sopperire a questa carenza si può far riferimento all'approccio o metodo frequentista, tale metodo prevede la determinazione, tramite la ripetizione di un esperimento di una base di riferimento n, numero totale degli esperimenti effettuati, da utilizzare come base per un rapporto con k, numero totale degli eventi favorevoli alla nostra operazione registrati, tale approccio vuole che qualora il numero degli esperimenti effettuati sia sufficientemente grande il rapporto tra k ed n corrisponde alla probabilità di tale evento; un terzo approccio, detto anche approccio delle preferenze rivelate prevede che, per la determinazione della possibilità si ricorra alla fiducia che un dato soggetto ha in tale evento, determinabile tramite l'escamotage della scommessa.
Si prevede quindi una funzione di guadagno, data da –p, prezzo pagato, più uno qualora il soggetto vinca, e –p, qualora il soggetto perda, si valuta quindi la p che il soggetto sarebbe disposto a pagare per partecipare a tale scommessa, la p risulta quindi essere la probabilità dell'evento, per essere valida a tal scopo gli eventi della scommessa devono essere a due a due escludibili, non devono dare vita a guadagni o perdite certe ed infine il guadagno in caso di affermazione vera deve essere di segno opposto alla perdita registrata in caso di affermazione falsa.
Analizzando il metodo soggettivista, in particolare i casi estremi escludibili di vittoria certa, e perdita certa, abbiamo che nel primo caso il soggetto sia disposta a pagare la quota massima possibile per l'evento stesso, data quindi da –p con segno opposto, che nel nostro caso corrisponde ad uno, e nel secondo caso non sia disposto a pagare praticamente niente, abbiamo quindi un valore di p che può definirsi come compreso tra due estremi zero ed uno, tale p in generale non si considera come dipendente dalla ricchezza del soggetto.

Dato quanto detto sul calcolo della probabilità è possibile definirne una tipologia di calcolo che prescinda le condizioni effettive della variabile rispettando dei criteri generali.
Tale metodologia di calcolo si dice calcolo assiomatico della probabilità, poiché per determinare la veridicità del calcolo effettuato si deve far riferimento a tre assiomi di formalizzazione: dato un evento qualsiasi la sua probabilità deve necessariamente essere un numero compreso tra zero ed uno, il secondo assioma vuole che scommettendo sull'evento certo la probabilità di tale evento corrisponde con quella di omega, totalità degli eventi, quindi con uno, dato un evento impossibile la sua probabilità corrisponde con quella dell'insieme vuoto, quindi con zero; data una successione di eventi incompatibili, che costituiscono una partizione di omega, si che la sommatoria per i che va da uno ad n, di E_i probabilità di tale evento, corrisponde alla probabilità di omega e quindi con uno.
Il calcolo delle probabilità appena definito può essere formalizzato per determinare la probabilità di alcune particolari categorie di eventi, in primis abbiamo l'estrazione da un'urna senza reimmissione, dato il verificarsi in un qualsiasi evento, con una data probabilità, si modificheranno anche il verificarsi di tutti gli altri, compreso l'evento stesso che diventerà quindi impossibile, il verificarsi quindi di un evento Y tale che sia una risultante degli eventi considerati, avremo che la probabilità del verificarsi di Y sia data dalla probabilità del verificarsi di uno degli eventi favorevoli al verificarsi di Y moltiplicato per la probabilità che dato questo evento se ne verifichi quello complementare che determina il nostro valore Y, sommato alla probabilità che si verifichino gli altri eventi favorevoli ad Y secondo le espressioni appena dati; va osservato in particolare che qualora il verificarsi di uno degli eventi considerati, primo degli eventi favorevoli ad Y, modifichi la possibilità che l'evento successivo si verifichi si parla di eventi che si respingono qualora la probabilità sia minore di quella dell'evento stesso, eventi che si attraggono nel caso contrario e eventi indifferenti qualora la probabilità rimanga costante. Quanto detto può anche essere formalizzato per il caso dell'estrazione con reimmissione, in tal caso avremo che data una funzione di Y combinatoria degli elementi estratti, avremo che la probabilità che si verifichi un valore di Y, è data dal prodotto di uno qualsiasi degli eventi considerati, favorevoli alla Y presa come esempio, per l'evento complementare favorevole al valore di Y che stiamo considerando, sommato a tutti gli altri eventi che possono portare a tale valore, senza escludere la possibilità che un evento si verifichi due volte di seguito come nel caso precedente.
Analizzando le variabili casuali teoriche è necessario attuare una prima distinzione in funzione della partizione dell'insieme, omega, noto come spazio degli eventi, dato un insieme omega finito o infinito numerabile avremo una variabile casuale discreta, dato un insieme omega infinito, quindi non numerabile, avremo una variabile casuale infinita, le cardinalità degli insiemi corrispondono rispettivamente a $|\mathbb{N}$ e $|\mathbb{R}$.
Variabile casuale di Bernoulli o Bernoulliana, si ha un insieme omega, a prescindere dalla sua natura, suddiviso in due sotto insieme S e S*, tali eventi sono incompatibili,

la loro intersezione è vuota, e rappresentano una partizione di omega, tale per cui la loro unione ci da l'insieme omega stesso, tali sottoinsiemi sono uno un evento, S, e uno l'evento contrario, S*, è possibile quindi definire P(S)=|P, 0<|P<1, e P(S*)=1-|P, data quindi una relazione f tale che associ f(S)=1, e f(S*)=0, tale associazione comporta anche l'associazione tra le relative probabilità, noto quindi il valore |P è possibile conoscere tutti gli elementi del sistema.

Variabile uniforme discreta, dato un insieme di eventi, omega, necessariamente di natura finita e numerabile, si presuppone che la variabile sia equi distribuzione all'interno dell'insieme.

Che quindi ogni evento abbia la stessa probabilità di verificarsi degli altri, quindi P(E(y))=1/n, ove E(y) indica il generico evento di Omega. Data quindi la generica f(Y) che lega ogni evento di omega ad un rispettivo numero appartenete all'insieme dei reali, avremo che ad ogni evento Y corrisponde un evento x secondo la seguente funzione f(Y)=x, data quindi la variabile N che esprime il valore massimo raggiunto dalla funzione f, è possibile determinare tutti i valori significativi di tale variabile, P(x)=1/n x=1;2;…;n.

Variabile Casuale geometrica, a differenza delle precedenti questa variabile di definisce in un campo degli eventi, omega, infinito ma numerabile, la cui cardinalità è data dalla cardinalità di |N, data quindi la suddivisione di omega in S e S*, S* evento contrario di S, dato poi che i due eventi siano incompatibili, la loro intersezione è un insieme vuoto, dato che essi siano una partizione di omega, la loro unione definisce l'insieme omega, dato poiché P(S)=|P e P(S*)=1-P, come le probabilità dei due insiemi, è possibile definire una relazione che permette di definire il numero di prove necessarie per far si che si osservi il verificarsi dell'eve3nto successo per la prima volta, data la verifica del terzo assioma è possibile osservare come tale processo sia una progressione geometrica.

Variabile casuale Binomiale, data un famiglia di eventi, omega, e diviso omega in due sotto insiemi S e S*, ove S* è l'evento contrario di S, la loro intersezione è un insieme vuoto, e la loro intersezione è l'insieme omega, e data P(S)=|P e P(S*)=1-|P, con 0<|P<1, dati B=S e N=S*, avremo che P(B)=|P% e P(N)=(1-|P)%, data una qualsiasi n, e che una estrazione effettuata tra le due variabili non influenzi la successiva è possibile tramite una relazione definire le probabilità di B ed N, noto il valore |P che ne definisce la probabilità.

Variabile casuale di Poisson, data le condizioni di partenza della variabile casuale binomiale, con l'unica eccezione della cardinalità dell'insieme che invece di essere definita ad un ipotetico n tende all'infinito, è possibile osservare come in queste condizioni la probabilità dell'evento n che tende all'infinito tende a zero, è però possibile, tramite una serie di processi e trasformazioni matematiche dimostrare come in tali condizioni la probabilità della variabile considerata tendi alla variabile di Poisson, $P(x)=(TY^x/x!)e^{TY}$ dato che TY=n*|P.

Dato il l'analisi, generica, di variabili casuali continue, il principale problema riscontrato riguarda il fatto che avendo una cardinalità definito nell'insieme dei reali non risulta sempre possibile assegnare una probabilità positiva, all'evento in questione, data la probabilità schiacciata tra zero e uno, che per quanto si consideri tale insieme come definito in |R non possiede sufficienti valori per definire una probabilità ad ogni evento appartenente ad un insieme omega con cardinalità in |R. Per sopperire a ciò è necessario definire una funzione, f, tale che possa definire una nuova x, tale funzione dovrà rispondere al criterio generale di avere un integrale definito tra gli estremi dell'intervallo pari ad uno.

Variabile casuale uniforme continua, date le stesse assunzioni valide per la variabile casuale uniforme discreta è possibile definire una f(x), tale che definisca l'intervallo come delineato tra due estremi a e b, avremo una f(x)=1/(a+b), qualora a<x<b, e un valore pari a zero altrove, tali condizioni permettono di effettuare una costruzione grafica di tale funzione, definita ta a e b, e con un valore continuo che corrisponde ad 1/(a+b).

Variabile casuale esponenziale negativa, tale funzione altri non è che una trasposizione nell'insieme |R della variabile casuale geometrica, generalmente utilizzata per definire il tempo, variabile continua, necessario per osservare il caso che esprime il "fallimento" del fenomeno stesso, per la funzione di normalizzazione si calcola il valore di tale funzione sino ad un fenomeno qualsiasi salvo poi sottrarre tale valore ad uno in maniera da trovare la possibilità degli eventi superiori a tale valore, si definisce una f(x) definita come $\theta e^{-\theta x}$ per x maggiore di zero, e che valga zero altrove, conoscendo θ è possibile quindi definire tutti i valori significativi di tale variabile casuale.

Variabile casuale normale, definita con una formula nota che determina un grafico noto come Gaussiana, tale grafico ha un andamento a campana caratterizzato essenzialmente da due valori la media e la varianza, la prima definisce il centro simmetrico del grafico stesso, a destra e sinistra di tale valore, dati spostamenti di uguale misura, la funzione presenta gli stessi valori, tale valore rappresenta, oltre che la media, anche la moda e la mediana della funzione, la varianza rappresenta invece il valore che va aggiunto o sottratto alla media per determinare i punti di flesso, in cui la concavità della funzione cambia, del grafico stesso, quanto detto è definibile dall'analisi della deriva prima e seconda della funzione, che, salvo fatto l'analisi di variabili banali, riduce la nostra analisi all'analisi di una funzione normalizzata, data da (x-E(x))/V(x), che data la derivata prima e seconda della funzione Gaussiana risulta l'unico elemento non determinato da costanti.

Variabile casuale normale standardizzata: data la funzione ((x-E(x)/V(x))=Z, è possibile determinare una variabile z tale per cui la funzione risulta centrata sull'asse delle y con media pari a zero e varianza pari ad uno, l'importanza di tale procedimento risiede nel fatto che di tale funzione è nota la funzione di ripartizione partendo da zero per tutti i numeri positivi, va poi sottolineato come lo stesso processo di calcolo

è valido dati segni invertiti anche per valori negativi di z, quanto detto è riscontrabile in opportune tabelle.

Inferenza statistica (1° parte): particolare ramo della statistica che si occupa di tutti qui ragionamenti relativi ad una serie di variabili, tali per cui, non è sempre possibile o auspicabile una rilevazione completa, che consideri quindi tutta la popolazione disponibile di un dato fenomeno, quanto detto può accadere per i motivi più disparati, che vanno dal costo dell'operazione, sino alla non volontà dei soggetti di partecipare all'indagine o alla necessaria distruzione che l'indagine stessa comporta del campione di riferimento. È quindi necessario sostituire una popolazione di riferimento con un opportuno campione e sfruttando dei collegamenti prestabiliti il campione e la popolazione totale, analizzare la stessa in funzione e dato il campione.

Inferenza statistica (2° parte): è necessario in primis definire le caratteristiche del campione che deve essere quindi scelto in maniera casuale ed essere rappresentativo della popolazione stessa, va osservato come poi i campionamenti risalenti ad una unica popolazione possono essere differenti ma, date estrazioni con reimmissione è possibile osservare come le probabilità legate alla comparsa degli elementi del campione siano sempre le stesse. Analizzando quindi la media campionaria, è possibile osservare come il suo valore non è determinabile e dipende dai valori della popolazioni e dal numero di elementi che compongono l'insieme di riferimento, è quindi possibile per ogni campione, dato un x numero elementi che ne compongono il relativo insieme campionario, osservare come la media di ogni singola estrazione non sia necessariamente corrispondente con la media, se non in via del tutto casuale, è poi però possibile osservare come la media delle medie di tutte le estrazioni osservate tende a coincidere con la media del campione, e come la probabilità di ottenere dati campionari che hanno per media un valore vicino alla media della popolazione sia più alto di ottenere valori che ne siano estranei, e quindi possibile concludere che la media campionaria coincide con la media della popolazione.

Dato quanto detto è possibile definire, in maniera proporzionalmente corretta all'aumentare della dimensione del campione, che diminuisce il grado di dispersione, che al limite, dato un campione coincidente con la media della popolazione sarebbe inesistente, ad un media campionaria data dalla sommatoria di tutte le medie si ottenuto diviso n cioè il numero di campioni osservati. Avremo quindi X*(media campionaria)=(1/N)*E(x_i).

Varianza campionaria: quanto detto per la media campionaria può essere osservato con la varianza campionaria con l'unica eccezione che al momento dell'estrazione di 1/n dalla formulazione delle varianza si devo osservare come tale valore va elevato alla seconda e quindi non si semplifica come nel caso della media campionaria con la n estratta dal processo di campionamento, se non in parte, facendo si che la varianza campionaria, S^2, non corrisponda propriamente alla varianza, ma alla varianza fratto n, S^2=V(x)/n, è possibile osservare poi come tale valore può essere corretto sfruttando

quello che è l'errore medio di stima che è (n-1)/n, elemento distorcente, invertito fa si che è possibile determinare, dopo opportune rielaborazione di S^2 e V(x) il valore S^2_c=V(x)/n-1, che altri non è lo stimatore corretto della varianza di x.

Teorema del limite centrale: qualora non si conoscano esattamente le dimensioni del campionamento effettuato è possibile far riferimento ad una insita distribuzione casuale della popolazione, tale per cui data una popolazione distribuita in maniera casuale tale distribuzione rimarrà valida anche per il campione analizzato. Va poi detto che qualora non sia nota la distribuzione, casuale, del campione di partenza è comunque possibile ipotizzarne una anche per la variabile casuale, difatti si può osservare come all'aumentare del numero delle estrazione effettuate qualsiasi variabile casuale si può ipotizzare come distribuita in maniera normale, va detto che generalmente il numero delle estrazioni per cui tale fenomeno inizia a verificarsi è trenta anche se è certo solamente intorno al cento, quanto detto è poi osservabile anche analiticamente, poiché svolgendo il limite per n che tende all'infinito di una qualsiasi distribuzione di frequenza si può osservare come la media tendi a zero e la varianza ad uno, caratteri fondamentali della funzione normale.

Stimatori: dato quindi un generico parametro che caratterizza una funzione di distribuzione casuale è possibile dover essere nella condizione di dover stimare tale parametro, dati per incognita alcuni valori fondamentali della funzione è necessario quindi utilizzare uno stimatore.

Prima proprietà degli stimatori: correttezza. È possibile osservare come la prima proprietà che uno stimatore deve possedere è quella di essere una stima corretta del parametro da stimare. Dato quindi un indice di distorsione B che indica la differenza tra lo stimatore, stimato, e il parametro da stimare, è possibile definire il parametro come il valore atteso dello stimatore meno B, dato quindi un processo di stima applicato alla precedente equazione, e la sostituzione del relativo parametro da stimare come prima visto possiamo definire che l'errore nella stima dipende essenzialmente: dalla differenza tra lo stimatore meno lo stimatore stima più l'indice di distorsione B. E' facile poi osservare come in presenza di stimatori corretti la distorsione, differenza tra il valore stimato e quello della relativa popolazione, è pari a zero e quindi risulta minimo l'errore di stima. È poi possibile ricorrere ad una eventuale correzione per determinare, da uno stimatore distorto, il relativo stimatore corretto. Dato invece uno stimatore distorto è possibile verificare che esso sia asintoticamente corretto, cioè che possa diventare uno stimatore corretto qualora aumentasse la dimensione del campione, è necessario quindi analizzare il limite per n che tende all'infinito dello stimatore stesso, se tale valore tende a zero avremo uno stimatore asintoticamente corretto.

Seconda proprietà degli stimatori: consistenza o convergenza in probabilità: tale principio impone come all'aumentare delle estrazioni campionarie la differenza tra la media del parametro stimato e il relativo valore riferito alla popolazione coincidano facendo si che tale errore di stima tenda a zero. Quanto detto può essere semplicemente dimostrato osservando che la probabilità di un errore di stima,

stimatore meno parametro da stimare, qualora inferiore ad alfa, sia inferiore ad uno meno la varianza dello stimatore meno parametro stimato fratto diviso per a al quadrato. Dato uno stimatore corretto tale processo si semplifica poiché è noto che la varianza, stimatore parametro da stimare corrisponda a zero ed è sufficiente osservare come la probabilità di ottenere un valore stimatore meno parametro da stimare, in valore assoluto, minori di a, debba essere minore o al massimo uguale ad uno, ciò si semplifica osservando come uno sia il valore massimo che la P_i può assumere e quindi tale probabilità debba essere uguale ad uno. Va osservato come quanto iscritto a sinistra della nostra uguaglianza altri non è che l'errore quadratico medio, escluso l'uno, qualora quindi ciò sia vero si verifica quanto osservato sopra, probabilità complessiva pari ad uno condizione sempre vera.
È necessario osservare poi come dopo opportune rielaborazioni si può determinare che l'errore quadratico medio altri non è che la varianza, della stima meno lo stimatore, sommato all'indice B di distorsione al quadrato.
Terza proprietà degli stimatori o efficienza relativa, (NB: da non confondere con quella assoluta, non trattata). Dati quindi due stimatori di uno stesso parametro, a parità di altre condizione, è necessario scegliere quello che presenta un errore quadratico medio minore.
Date le caratteristiche che deve possedere uno stimatore è necessario osservare come si possono avere più metodi di stima in grado di produrre uno stimatore corretto di un parametro. In particolare è necessario considerare due:
Metodo di stima uno: metodo dei momenti o metodo di stima per analogia o metodo di stima per corrispondenza. Come osservabile dalla denominazione tale metodo sfrutta un'analogia presente tra il momento campionario, centrato o non, e il correlativo momento della popolazione, quanto detto viene utilizzato in funzione del fatto che, come osservato in precedenza, in particolare i momenti non centrati, possiedono delle caratteristiche, correttezza e convergenza in probabilità che sono tutti caratteri che deve possedere un buono stimatore.
Detto quanto si associa alla stima del momento r esimo della popolazione il corrispondente momento campionario. L'unica problematica di tale metodo consiste nella presenza di eventuali momenti che centrati che come già dimostrato sono stimatori distorti, anche se asintoticamente corretti del momento della popolazione.
Metodo di stima due: massima verosimiglianza: data quindi la necessità di determinare la provenienza di un campione e un'estrazione di più campioni, tra una gamma di ipotesi possibili, generalmente tutte riferibili a variabili casuali, è possibile fare ciò determinando la probabilità dello stesso ponendo l'ipotesi di indipendenza delle estrazioni effettuate, e osservando la maggiore, dato quindi la necessità di stimare il valore di uno stimatore in maniera continua, seguendo lo stesso metodo, si ottiene una moltiplicatoria che va da uno a n, valori assunti dalla variabile, della funzione di riferimento, per inciso sarebbe la funzione del modello teorico che permette di calcolare tale parametro, per risolvere tale funzione si deve poi ricorre a una

logaritmica; determinata la logaritmica è possibile derivare la funzione è determinare il punto di massimo che altri non è che la stima di massima verosimiglianza di tale parametro.

Prova di ipotesi: tale prova è un insieme di tecniche che permette, di verificare con quali livelli di probabilità le ipotesi fatte su un parametro di una popolazione sono correttamente o erratamente vere o false. Tale costruzione si base su funzioni normali, N.B. non ancora normalizzate, una prima nota come H_0, o ipotesi nulla, esprime l'ipotesi di partenza, che generalmente si intende contraddire, H_1 invece indica un'ipotesi modificata che in genere si tende a voler confermare. Abbiamo quindi degli specifici valori validi per entrambe le ipotesi, in generale la varianza di tale evento, e dei valori validi specificatamente per ogni singola ipotesi, la media dell'evento stesso, tale problematica si può affrontare anche nel senso opposto partendo dalla probabilità di errore o successo, è quindi necessario determinare un punto critico c che funge da spartiacque, determinato dall'uguaglianza delle due funzioni normali, per un valore della media campionaria minore di c siamo nella condizione in cui è vero l'evento con media minore, nel caso contrario sarà vero l'evento con media maggiore, va però osservato come anche qualora si dia per vero un dato evento in funzione del rapporto con c è possibile definire un margine di errore, per quanto minimo, di tale evento, che p rappresentabile come la coda della funzione normale che cade al di sotto dell'altra funzione normale, tale spazio è noto come a qualora riferito alla funzione H_0 e b riferito alla funzione H_1.

I valori a e b si determinano, data la media derivante dal caso pratico, traslando le due normali in z, di cui si conoscono i valori della funzione di ripartizione, tramite la differenza tra uno e la funzione di ripartizione in z calcolata per il valore traslato della media considerata, dato come media cardine quella dello specifico caso che si considera come vero. Date quindi le funzioni normalizzate è possibile con un semplice calcolo, che permette di evitale l'uguaglianza di determinare il valore di c, noto il valore di zeta e la media della variabile da considerarsi come vera.

Teorema di Neyman-Pearson: tale teorema permette di individuare l'area della funzione H_0 ove si collocano le aree di rifiuto dell'ipotesi, altresì vera, in funzione di un valore predeterminato della media di H_1, qualora quindi la media di H_1 sia maggiore di quella di H_0 l'area si colloca nella coda destra, qualora invece ci trovassimo nel caso contrario l'area di rifiuto si colloca nella coda sinistra, con entrambi i valori pari ad a, nel caso invece di una media di H_1 diversa da un dato valore, qualsiasi esso sia, l'area di rifiuto si colloca in ambedue le code della funzione H_0, in questo caso però la probabilità di rifiutare un'ipotesi vera non è pari ad a ma ad a/2.