

1 RILEVAZIONI STATISTICHE

DEFINIZIONI BASE

Una **rilevazione statistica** è il complesso di operazioni rivolte ad acquisire una o più informazioni su un insieme di elementi oggetti di studio. Una frase che riassume le relazioni tra le definizioni successive: *“La Popolazione si specifica nelle unità statistiche mentre il carattere (che varia nella popolazione) si specifica nella modalità (assunta nell’unità statistica).”*

La **Popolazione** è qualsiasi insieme di elementi che forma oggetto di uno studio statistico. Si distingue tra popolazione reale (effettivamente esistente) e popolazione virtuale (definibile con accuratezza ma non osservata né osservabile).

Una **unità statistica** (o soggetto) è l’elemento base della popolazione sul quale viene effettuata la rivelazione.

Carattere (o variabile) è il fenomeno oggetto dello studio, rilevato o misurato sulle unità statistiche. Un carattere quantitativo si definisce **variabile** se assume per modalità dei numeri reali (a loro volta suddivisi in variabili continue, se capaci di assumere qualsiasi valore contenuto in un intervallo reale predefinito, e variabili discrete se capaci di assumere al massimo un insieme numerabile di modalità) mentre un carattere qualitativo si definisce **mutabile** se assume attributi non numerici della più varia natura. I caratteri si distinguono in funzione della scala di misurazione:

- **I caratteri con scala nominale** costituiscono mutabili le cui modalità (o attributi) non assumono alcun ordine precostituito. L’unico confronto ammissibile consiste nello stabilire se possiedono o no lo stesso attributo.
- **I caratteri con scala ordinale** sono mutabili che, pur non riferendosi a valori numerici, assumono modalità logicamente sequenziali, in ordine crescente o decrescente (è possibile stabilire una relazione d’ordine).
- **I caratteri con scala ad intervallo** sono variabili che consentono un confronto solo per differenza tra le modalità che i soggetti assumono poiché essi fanno riferimento ad un’origine arbitraria. (esempio temperatura).
- **I caratteri con scala di rapporto** sono variabili per le quali è intrinseca ed univoca la definizione dello zero assoluto.

Modalità è l’espressione concreta del carattere nelle unità statistiche, cioè il numero (per caratteri quantitativi) o l’attributo (per caratteri qualitativi) che l’unità statistica manifesta. L’elenco di tutte le possibili modalità di un carattere si dice esaustivo se è completo e le modalità si dicono disgiunte se una unità statistica può manifestare il carattere in una ed una sola modalità tra quelle indicate.

Frequenza è il numero delle volte che una determinata modalità si verifica nel collettivo di riferimento. Se la frequenza è un numero intero non negativo si parla di frequenza assoluta mentre quando tale frequenza è rapportata al totale delle unità statistiche della popolazione si parla di frequenza relativa.

Serie è una distribuzione di frequenza organizzata rispetto ad un criterio qualitativo

Seriazione è una distribuzione di frequenza organizzata secondo un criterio quantitativo, rispetto alle modalità del carattere.

Informazione statistica è ogni risultato ottenuto da un'indagine sui collettivi esaminati (=popolazione o campione) rispetto ai loro costituenti (=unità statistiche) e in rapporto ad uno o più fenomeni (=caratteri).

MISURE STATISTICHE ELEMENTARI

La **differenza assoluta** tra due modalità di un carattere quantitativo: $x_2 - x_1$.

La **differenza relativa** (non dipende dall'unità di misura o l'ordine di grandezza): $\frac{x_2 - x_1}{x_1}$ dove la modalità x_1 è antecedente sul piano logico (o temporale) rispetto a x_2 .

Il **tasso di variazione** (percentuale se moltiplicato per 100) corrisponde alla differenza relativa tra la modalità assunta da una variabile X al tempo t_2 rispetto a t_1 . Se il carattere X assume valori strettamente positivi, allora la differenza tra i logaritmi (neperiani) delle modalità tra due tempi distinti equivale, con buona approssimazione, al tasso di variazione.

RAPPORTI STATISTICI

Il **rapporto di composizione** si ottiene dividendo il valore rilevato in una certa circostanza per l'analogo valore rilevato per l'intera popolazione. Esso esprime, quindi, la frazione relativa (o percentuale) posseduta o registrata dalla modalità in questione rispetto al totale (es. composizione percentuale della forza lavoro rispetto al sesso, al titolo di studio).

I **rapporti di derivazione** si ottengono dividendo la modalità di un fenomeno per quella corrispondente di un altro che, sul piano logico e/o temporale, ne costituisce causa o ne è antecedente necessario o presupposto logico (es. indice di mortalità come rapporto tra morti e vivi). Importante stabilire correttamente il collettivo di riferimento più idoneo, ponendo al denominatore la "effettiva" popolazione che può aver generato il dato collocato al numeratore.

I **rapporti di densità** sono definiti mediante il confronto tra la dimensione globale di un fenomeno (al numeratore) e la dimensione spaziale, temporale o caratterizzante cui esso fa riferimento (densità sostanza inquinante nell'atmosfera).

I **rapporti di coesistenza** riguardano ogni rapporto tra frequenza (o quantità) di una modalità rispetto alla frequenza (o quantità) corrispondente di un'altra modalità. Utilizzati negli studi demografici: es. indice di vecchiaia come rapporto tra residenti con più di 65 anni e residenti con meno di 15.

L'**indice di eccedenza** è una misura sintetica di un carattere che viene esaminato accorpando le sue manifestazioni in due sole modalità x_1, x_2 con rispettive frequenze f_1, f_2 . Esso è rivolto alla valutazione dello squilibrio di una modalità sull'altra rispetto al totale delle unità statistiche ed è così definito: *Indice di eccedenza* = $\frac{f_1 - f_2}{f_1 + f_2}$

TIPOLOGIA E RAPPRESENTAZIONI DELLE RILEVAZIONI STATISTICHE

I dati statistici si possono presentare in forma enumerativa, tabellare, oppure secondo una rappresentazione grafica che evidenzia aspetti particolare dei dati. Tra le più importanti ci sono:

- **Distribuzioni di frequenza** – indicano come le unità della popolazione si distribuiscono rispetto alle modalità del carattere in esame. Se i caratteri rilevati sono più di uno bisognerà studiare la distribuzione congiunta ecc.

- **Serie Storiche** – esprimono la dinamica di un certo fenomeno registrato istantaneamente (variabile di flusso) o conteggiato nel periodo definitivo (variabile di stato). Es. i nati vivi registrati ogni anno a Roma.
- **Serie territoriali** – esprimono la distribuzione di un fenomeno in rapporto al territorio e, se l'analisi è condotta anche rispetto al tempo, si parla di serie spazio-temporali.
- **Matrici dei dati** – particolare rappresentazione tabellare mediante la quale si schematizzano le informazioni raccolte su ciascuna unità statistica in rapporto ad una molteplicità di fenomeni. Ogni colonna della matrice esprimerà, quindi, una variabile o mutabile rilevata sulle diverse unità statistiche. Dualmente, ciascuna riga della matrice esprimerà ordinatamente le misurazioni ottenute sulla singola unità statistica

NUMERI INDICE

I **numeri indice** sono ottenuti **rapportando due valori differenti di uno stesso fenomeno in circostanze differenti**. Essi sono utili per confrontare l'ammontare di un fenomeno in tempi e/o luoghi differenti mediante una misura che è sempre positiva e indipendente dall'unità di misura: è prassi moltiplicare il rapporto per 100.

Così un numero indice di 100 indica che non ci sono state variazioni, un numero indice del 90% indica che c'è stata una diminuzione dei prezzi del 10%, un numero indice del 104% indica un aumento dei prezzi del 4%.

I **numeri indice più diffusi riguardano le variazioni dei prezzi di beni e servizi** che consentono, nel tempo e sul territorio, di valutare la dinamica dell'inflazione mediante confronti tra i prezzi (rispetto ad un anno base o una media nazionale). Quando vi sono più beni e servizi eterogenei, **il rapporto avviene tra la somma dei prezzi moltiplicati per le quantità commercializzate** (vendute, acquistate ecc.): in tal modo, il numero indice ponderato si risolve in un numero indice semplice ottenuto come rapporto tra il valore complessivo (=prezzi*quantità) dei beni o servizi nelle due circostanze.

Se la quantità di riferimento è fissa, l'indice sintetico corrispondente si chiama di **Laspeyrs**; se essa è variabile, l'indice sintetico di riferimento si chiama **Paasche**. L'indice di Laspeyrs risulta computazionalmente preferibile perché fa riferimento ad un paniere "fisso" di beni.

I principali indici dei prezzi al consumo costruiti mensilmente dall'ISTAT sono due: **i numeri indice dei prezzi al consumo per l'intera collettività nazionale** (transizioni relative a merci o a servizi scambiati tra gli operatori economici e l'intero universo dei consumatori finali) e **i numeri indice dei prezzi al consumo per le famiglie e gli operai** (beni e servizi acquistati dalle famiglie dei lavoratori dipendenti non agricoli, numeri indici del costo della vita). La differenza tra i due riguarda la popolazione di riferimento: il primo riguarda l'universo dei consumatori, il secondo riguarda una parte di quell'universo. La rilevazione dei prezzi al consumo per il calcolo del secondo indice è fatta su un paniere di prodotti mantenuto costante per un determinato arco di tempo. Il paniere contiene prodotti che vengono individuati tra quelli maggiormente acquistati dai consumatori. È anche utile sottolineare che tale paniere non contiene alcun prodotto obsoleto, ma invece prodotti largamente in uso. I prezzi rilevati sono quelli realmente applicati dagli esercenti al netto di sconti (se essi non ricorrono con una certa regolarità).

NUMERI INDICI E DEFLAZIONE

I numeri indice costituiscono uno strumento fondamentale per misurare le fluttuazioni dei prezzi, ma vengono anche utilizzati per confrontare nel tempo aggregati espressi in moneta corrente quindi con diverso potere di acquisto. L'operazione con cui si depurano aggregati monetari correnti riferiti a tempi diversi dagli effetti derivanti dalla variazione dei prezzi è detta deflazione. Si considerano due procedimenti:

- Deflazione diretta – utilizzabile solo per grandezze esprimibili come somma di prodotti di prezzi unitari per quantità e consiste nel ricalcolare anno per anno gli aggregati moltiplicando i prezzi dell'anno considerato come base per le quantità dei singoli anni considerati.
- Deflazione indiretta – consiste nel dividere ciascun dato della serie a prezzi correnti per la serie dei numeri indice ritenuti più idonei.

INDICE DI FISHER

L'indice di Fisher è definito come la radice quadrata del prodotto tra l'indice di Laspeyrs e l'indice di Paasche. Esso gode della proprietà di inversione dei fattori: dato un aggregato $X_t = \sum_i p_{it} q_{it}$ (ovvero una media dei prodotti prezzi per quantità) l'indice dell'aggregato $\frac{X_t}{x_0}$ è uguale al prodotto tra l'indice dei prezzi $\frac{p_t}{p_0}$ e l'indice delle quantità $\frac{q_t}{q_0}$.

2 DISTRIBUZIONI DI FREQUENZA

DISTRIBUZIONI STATISTICHE

La **distribuzione di quantità** è un'organizzazione dei dati in forma tabellare tale che **ad ogni modalità di una certa variabile fa corrispondere una quantità (assoluta o relativa)**, idealmente trasferibile tra le unità della popolazione. La **distribuzione di frequenza** è un'organizzazione dei dati in forma tabellare tale che **ad ogni modalità di una certa variabile (qualitativa o quantitativa) fa corrispondere la rispettiva frequenza** (assoluta o relativa).

Ad esempio, supponendo che si rilevi il numero dei dipendenti di un insieme di aziende, si ha una distribuzione di frequenze se per ciascun numero di dipendenti, o per ciascuna sua classe, si mostra il numero delle aziende che hanno quel numero di dipendenti; si ha invece una distribuzione di quantità se si mostra il numero dei dipendenti in quella classe.

DISTRIBUZIONI DI FREQUENZA PER VARIABILI DISCRETE

Frequenze assolute – numero di volte in cui una modalità si verifica nel collettivo

Frequenze relative – numero di volte in cui una modalità si verifica diviso il numero di elementi della popolazione

Diagramma cartesiano – è preferibile rappresentare la distribuzione di frequenza in un diagramma cartesiano ponendo in ascissa i valori delle modalità ed in ordinata le corrispondenti frequenze, ottenendo una tipica rappresentazione grafica detta “a barre verticali”.

DISTRIBUZIONI DI FREQUENZA PER VARIABILI CONTINUE

Nel caso di una variabile continua non è possibile far corrispondere ai valori che essa assume le rispettive frequenze (assolute e relative) perché tra due modalità qualsiasi ve ne possono essere infinite altre, conviene quindi suddividere il campo di variazione della variabile X in classi di modalità.

Poiché le classi possono avere ampiezza differente bisogna disegnare per ogni classe $(x_{i-1}, x_i]$ un rettangolo di area n_i (numero di unità che assumono un valore all'interno dell'intervallo). L'altezza del rettangolo con base $(x_{i-1}, x_i]$ sarà pari a: $h_i = \frac{n_i}{x_i - x_{i-1}}$. **La quantità h_i viene chiamata a volte densità di frequenza della classe i-esima.**

Istogramma – La rappresentazione grafica che alle classi della modalità di una variabile continua (in ascissa) fa corrispondere un rettangolo di area pari alla frequenza delle unità statistiche appartenenti a quella classe.

Secondo il **criterio della classi equi-ampie** si suddivide l'intervallo di definizione per la variabile X in k intervalli di uguale ampiezza $d = \frac{x_k - x_0}{k}$; secondo il **criterio delle classi equi-frequenti**, ovvero ogni classe contiene frequenze assolute o relative costanti

FUNZIONE DI RIPARTIZIONE EMPIRICA

Se consideriamo la distribuzione delle modalità ordinate di una variabile X , rilevate sulle n unità della popolazione: $x_1 \leq x_2 \leq \dots \leq x_n$. È possibile attribuire a ciascuna di esse frequenza assoluta 1 e frequenza relativa $\frac{1}{n}$.

La funzione che associa ad ogni valore reale x_0 la frazione delle unità che sono inferiori o uguali a x_0 si chiama funzione di ripartizione empirica: $F(x_0) = \{\text{frequenza relativa delle unità con modalità} \leq x_0\} = \frac{\#(X \leq x_0)}{n}$

La funzione di sopravvivenza è il complementare della funzione di ripartizione: $S(x) = 1 - F(x)$

CONFRONTO TRA DISTRIBUZIONI DI FREQUENZA: LA DISSOMIGLIANZA

Qualora si possedano due o più distribuzioni di frequenza per la stessa variabile, rilevate in tempi, luoghi o circostanze diverse, è opportuno individuare dei criteri di confronto. Per graduare la diversità tra due distribuzioni di frequenza relative $\{f_{i1}, f_{i2}, i = 1, 2, \dots, k\}$ occorre ricercare una misura: 1) nulla quando $f_{i1} = f_{i2}$; 2) crescente quando aumentano le differenze $|f_{i1} - f_{i2}|$; 3) normalizzata perché compresa tra un minimo e un massimo finiti.

$$Diss = \text{Indice di dissomiglianza} = \frac{1}{2} \sum_{i=1}^k |f_{i1} - f_{i2}|$$

$$0 (= \text{distribuzioni coincidenti}) \leq Diss \leq 1 (= \text{distribuzioni concentrate su valori diversi})$$

INDICATORI SINTETICI DELLA DISTRIBUZIONE DI FREQUENZA

Gli aspetti più importanti di una distribuzione di frequenza riguardano:

- **La posizione**, cioè misura della sua centralità complessiva in rapporto a modalità e rispettive frequenze. Valore “rappresentativo” della variabile nella sua globalità e capace di sostituire in qualche modo tutte le osservazioni.
- **La variabilità**, cioè la mutevolezza dei dati nella popolazione, ovvero l’attitudine della variabile ad assumere diverse modalità.
- La **forma** cioè l’aspetto complessivo della distribuzione di frequenza rispetto a configurazioni standard, ad esempio misurando la simmetria della distribuzione.

Tutti gli indici statistici possono essere suddivisi in tre categorie:

- Gli **indici assoluti** sono misure il cui campo di variazione dipende dalla variabile che si sta esaminando. Quindi, sono espressi in una unità di misura che dipende strettamente dall’unità di misura della variabile in oggetto.
- Gli **indici relativi** sono misure svincolate dall’unità di misura perché costituiscono rapporti tra indici assoluti; sono pertanto numeri puri, utili per confrontare fenomeni simili.
- Gli **indici normalizzati** sono particolari indici relativi che variano in un insieme finito (generalmente in 0/1 oppure in -1/1). Possono quindi essere utilizzati per effettuare sintesi e confronti tra qualsiasi tipo di fenomeno per i quali essi siano logicamente e analiticamente calcolabili. Per normalizzare $J^* = \frac{j - J_{\min}}{J_{\max} - J_{\min}}$ o $J^* = \frac{1}{1 + \exp(-j)}$

3 INDICI STATISTICI DI POSIZIONE

MEDIA SECONDO CAUCHY

La media di una variabile X è qualunque valore reale M , intermedio tra il minimo x_1 e il massimo x_n di una distribuzione di frequenza: $x_1 \leq M \leq x_n$. Per quanto ovvio e convincente, tale requisito costituisce più un controllo delle definizioni successive, essendo generalmente infiniti i numeri reali che soddisfano tale criterio di internalità.

MEDIA SECONDO CHISINI

La media di una variabile X è quel valore M che, rispetto ad una funzione sintetica delle osservazioni, ne lascia inalterato il valore: $f(x_1, \dots, x_n) = f(M, \dots, M)$. Specificando la funzione f si perviene a diversi tipi di media.

MEDIA SECONDO WALD

Valore che minimizza la funzione di perdita complessiva:

$$M: \varphi(d(x_1, M), \dots, d(x_n, M)) \text{ è minimo, } \begin{cases} d(x_i, M) = \text{perdita individuale} \\ \varphi = \text{funzione che sintetizza perdite} \end{cases}$$

MEDIA ARITMETICA

Se definiamo $f(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ allora il criterio del Chisini implica:

$$f(x_1, \dots, x_n) = f(\mu, \dots, \mu) \Leftrightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n \mu \Leftrightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

La media viene chiamata aritmetica perché in una progressione aritmetica di un numero dispari di n termini, $\{x, x+q, x+2q, \dots, x+(n-1)q\}$, la media aritmetica μ coincide proprio con il termine centrale della progressione.

PROPRIETA' DELLA MEDIA ARITMETICA

- È sempre compresa tra il massimo e il minimo delle modalità delle variabili.
- **La somma degli scarti dalla media aritmetica è nulla** (media costituisce il baricentro di una distribuzione di frequenza): $\sum_{i=1}^n (x_i - \mu) = \sum_{i=1}^n x_i - n\mu = \dots = 0$. Scarti positivi e negativi rispetto alla media si compensano.
- Se la variabile X ha media aritmetica μ , allora la variabile $\alpha + \beta X$ possiede media aritmetica pari a $\alpha + \beta\mu$.
- **È associativa**: la media di una variabile osservata in più gruppi è ottenuta come media dei singoli gruppi, tenuto ovviamente conto delle (eventuali) differenze tra le numerosità dei gruppi.
- **La media aritmetica è l'unico valore che rende minima la somma degli scarti al quadrato**. Per la dimostrazione mi basta derivare la funzione $g(\delta) = \sum (x_i - \delta)^2$ uguagliando tale derivata a zero e risolvendo l'equazione ottengo $\delta = \frac{\sum x_i}{n}$.

MEDIA GEOMETRICA

Se definiamo $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i$ allora il criterio del Chisini implica:

$$f(x_1, \dots, x_n) = f(G, \dots, G) \leftrightarrow \prod_{i=1}^n x_i = \prod_{i=1}^n G \leftrightarrow G = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} = e^{\frac{1}{n} \sum_{i=1}^n \log(x_i)}$$

Dove l'ultima espressione mostra che il logaritmo della media geometrica è pari alla media aritmetica dei logaritmi.

MEDIA ARMONICA

Se definiamo $f(x_1, \dots, x_n) = \sum_{i=1}^n \frac{1}{x_i}$ allora il criterio del Chisini implica:

$$f(x_1, \dots, x_n) = f(A, \dots, A) \leftrightarrow \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n \frac{1}{A} \leftrightarrow A = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Tale media si chiama armonica perché, per n dispari, rappresenta il termine centrale di una progressione armonica, cioè di una progressione aritmetica degli inversi (prendo i termini della progressione aritmetica e li inverte).

MEDIA ARITMETICA PONDERATA

Se assegno ad ogni modalità x_i un peso w_i proporzionale all'importanza di ciascun x_i , la media aritmetica ponderata è definita come: $\mu_w = \frac{\sum x_i w_i}{\sum w_i}$. Anche in questa formulazione la media è il baricentro delle osservazioni (cui vengono applicati dei pesi, come in fisica avviene per le masse); possiede le stesse proprietà della media aritmetica semplice.

MODA

Si definiscono medie lasche quelle che utilizzano, per l'individuazione sintetica della posizione di una variabile, alcuni valori specifici della distribuzione di frequenza, individuati sulla base della loro collocazione relativa rispetto a tutti gli altri, ma senza coinvolgere nel calcolo tutte le modalità della variabile X.

La moda MO (detta talvolta "norma" o "valore normale") di una distribuzione di frequenza è la **modalità cui corrisponde la massima frequenza, assoluta o relativa**.

Sintetizzare una variabile X tramite la moda significa assumere come valore "più rappresentativo" della distribuzione quello che si è verificato più spesso degli altri. Quindi la moda è **"quel valore in corrispondenza del quale gli scarti dalla moda sono nulli con maggiore frequenza"**. Il calcolo della moda per variabili raggruppate in classi avviene mediante la individuazione della classe di modalità cui corrisponde la massima frequenza, per cui si parla di classe modale. Tuttavia, se le classi non sono equi-ampie si confrontano le "densità di frequenza": pertanto, la classe modale è quella con densità di frequenza più elevata.

MEDIANA

La mediana è quel **valore della variabile che bipartisce la distribuzione ordinata delle modalità**, cioè tale che metà delle osservazioni sia inferiore alla mediana e metà sia ad essa superiore. In altri termini, la mediana è la modalità della unità statistica che occupa il posto centrale nella distribuzione ordinata delle osservazioni.

Ricordando la definizione della funzione di ripartizione, si vede subito che la mediana è quel valore nel quale la funzione di ripartizione vale $1/2$ cioè: $F(Me)=1/2$.

- Per variabili discrete $ME = \begin{cases} \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{se } n \text{ è pari} \\ x_{\frac{n+1}{2}}, & \text{se } n \text{ è dispari} \end{cases}$
- Per variabili X continue, il raggruppamento in classi delle modalità consente al più di determinare solo una classe mediana nella quale ricade l'unità statistica che bipartisce la distribuzione ordinata della modalità.

Dalla definizione della mediana discende subito che il numero degli scarti $(x_i - ME)$ positivi è esattamente uguale al numero degli scarti negativi. Tuttavia, la proprietà più importante della mediana (che la caratterizza) è la seguente: **la mediana ME è quel valore che minimizza la somma degli scarti assoluti** (con il modulo!).

Tale proprietà della mediana viene richiamata per risolvere il seguente problema: “dove collocare un deposito (di merci, carburante, pezzi di ricambio, etc.) lungo un'autostrada con punti vendita ai km x_1, x_2, \dots, x_n in modo da minimizzare i costi di rifornimento dei punti vendita?

Un aspetto importante della mediana è la sua capacità di essere rappresentativa della posizione della distribuzione anche in presenza di valori notevolmente diversi da tutti gli altri: tale requisito è detto **resistenza** ed è, ovviamente, non soddisfatto dalla media aritmetica e dalle altre medie perché il valore diverso entra nel calcolo e, quindi, modifica il valore di posizione.

QUANTILE

Data la funzione di ripartizione $F(x)$ di una variabile X , si definisce quantile j-esimo di ordine p ($x_{p,j}$) quella modalità di X tale che:

$$F(x_{p,j}) = j * p \leftrightarrow x_{p,j} = F^{-1}(jp), \quad \text{per } j = 1, 2, \dots, \left[\frac{1}{p}\right]$$

Così, per esempio, quando $p = \frac{1}{10}$ i decili sono i dieci valori D_1, \dots, D_{10} che dividono in dieci parti di uguale numerosità la distribuzione della variabile X .

Mediana, quantili e moda esistono sempre all'interno della distribuzione in esame. Invece, le medie sono valori astratti che non necessariamente si ritrovano nei casi osservati. Se le variabili sono qualitative non essendo possibile alcuna operazione algebrica tra le modalità non sono applicabili le medie analitiche. Se la variabile è misurata su una scala nominale l'unico indice applicabile è la moda. Se la variabile è misurata su scala ordinale, oltre che la moda, può essere calcolata anche la mediana.

MEDIE TRONCATE E MEDIE SECONDO WINSOR

Effettuo una media aritmetica delle modalità ordinate $x_{(i)}$ di una variabile X , con pesi π_i da determinare in funzione di criteri statistici e tali che:

$$L(\pi_i) = \sum_{i=1}^n \pi_i * x_{(i)}, \quad \pi_i \geq 0, \Sigma \pi_i = 1$$

Considerando quelle combinazioni lineari che escludono dal calcolo di una media i “valori estremi” (outliers) della distribuzione, eliminando una frazione α dei dati più piccoli e una frazione β dei dati più elevati (assegno $\pi_i = 0$ ai valori che desidero eliminare), ottenendo così le medie troncate. La media aritmetica si ottiene ponendo $\pi_i = \frac{1}{n}$.

Se invece che eliminare i valori, replico quest’ultimi rispettivamente con il minimo e il massimo dei valori rimasti (cioè dei valori ordinati esclusi quelli che ho eliminato) ottengo le medie secondo Winsor.

MEDIE DI WALD

Specificando opportunamente, nella definizione di media data da Wald, sia la perdita individuale che la perdita complessiva, si possono ottenere come casi particolari sia la moda, sia la media aritmetica sia la mediana. È anche possibile accostare alle medie una “misura di accuratezza” normalizzando la media ottenuta con Wald rispetto al proprio massimo.

4 INDICI STATISTICI DI VARIABILITA'

La variabilità di un fenomeno è la sua attitudine ad assumere diverse modalità. Operativamente, occorre pervenire ad una misura di tale attitudine e questo può avvenire in via assiomatica oppure mediante una costruzione statistica degli indici verificando, poi, il soddisfacimento degli assiomi.

Una misura di variabilità $V(x_1, \dots, x_n)$ definita sulle osservazioni (x_1, \dots, x_n) deve soddisfare i seguenti assiomi:

- $V(x_1, \dots, x_n) \geq 0$
- $V(c, \dots, c) = 0$
- $V(x_1 + c, \dots, x_n + c) = V(x_1, \dots, x_n)$
- $V(x_1, \dots, x_n) \geq V(y_1, \dots, y_m)$ allora X è più variabile di Y

Bisogna distinguere le misure di variabilità in due categorie:

- Variabilità delle singole modalità x_1, \dots, x_n rispetto ad un valore di posizione mediante una sintesi degli scarti tra le singole modalità e il valore di riferimento;
- Variabilità reciproca (mutua) tra tutte le modalità considerate a due a due

VARIABILITA' RISPETTO AD UN CENTRO

Varianza

L'indice più importante per misurare la variabilità di una distribuzione è espresso dalla media degli squarti al quadrato:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

Essa è sempre non negativa ed è un indice assoluto espresso nell'unità di misura del fenomeno al quadrato. Varia tra 0 e un massimo che può essere infinito. Tuttavia, se si suppone che la media sia fissa (cioè sia fisso e finito l'ammontare complessivo di un carattere supposto trasferibile tra le n unità), **la situazione di massima variabilità si presenta quando una sola modalità racchiude l'intero fenomeno e le rimanenti n-1 posseggono 0**. Lo scarto quadratico medio rappresenta la media quadratica degli scarti dalla media ed ha stessa unità del fenomeno osservato.

Il massimo scarto possibile è limitato dalla varianza mediante la relazione seguente: $\max |x_i - \mu| \leq \sigma \sqrt{n-1}$.

Coefficiente di variazione

La variabilità e lo scarto quadratico medio sono indici assoluti per cui è opportuno introdurre indici relativi o normalizzati. Un indice relativo molto utilizzato, purché $\mu > 0$, è il coefficiente di variazione CV:

$$CV = \frac{\sigma}{\mu} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\mu} \right)^2}$$

CV è indipendente dall'unità di misura, cioè è un numero puro, e misura la variazione media del fenomeno in rapporto alla sua media aritmetica ed è utile per confrontare la variabilità relativa di un fenomeno in circostanze differenti.

Scostamento medio della mediana

Se l'indice di posizione prescelto per misurare la dispersione è la mediana, allora si definisce lo scostamento semplice medio dalla mediana mediante la formula:

$$S(Me) = \frac{1}{n} \sum_{i=1}^n |x_i - Me|$$

Che è il minimo tra tutti gli scarti assoluti.

Scostamento semplice medio

$$S(\mu) = \frac{1}{n} \sum_{i=1}^n |x_i - \mu|, \quad \text{non richiede ordinamento dei dati}$$

VARIABILITA' E FUNZIONE DI RIPARTIZIONE EMPIRICA

La funzione di ripartizione empirica può essere utilizzata per derivare informazioni e misure circa la variabilità di un fenomeno. Se essa è molto ripida le modalità del carattere sono tutte assai vicine al valore di posizione (media, mediana, valore centrale etc.); se essa tarda a raggiungere il valore 1 allora vi è una grande variabilità nelle osservazioni. Per questo il confronto tra i quantili costituisce un approccio idoneo per costruire altri indici di variabilità.

Campo di variazione

Definito come la differenza tra il valore massimo x_n e il valore minimo x_1 delle modalità X, cioè:

$$Range(X) = \max(x) - \min(X) = x_{(n)} - x_{(1)}$$

Tale misura è influenzata anche da un solo valore atipico e rende l'indice molto vulnerabile ad errori e situazioni eccezionali.

Campo di variazioni interquantile

Meno vulnerabili ai valori atipici è ed è definito come la differenza tra il terzo e il primo quartile

$$I_{QR} = Q_3 - Q_1 = F^{-1}(0,75) - F^{-1}(0,25)$$

MUTUA VARIABILITA'

La mutua variabilità è un concetto che si applica principalmente ad un carattere trasferibile poiché solo la trasferibilità di un bene, servizio, territorio, evento, etc. rende possibile la determinazione teorica della variabilità minima e massima all'interno di una prefissata distribuzione (quindi, anche se è possibile calcolare gli indici di mutua variabilità per caratteri concettualmente non trasferibili come l'età appare impossibile formulare per essi misure normalizzate non avendo significato le misure estreme di riferimento).

Per un prefissato ammontare complessivo del carattere trasferibile, e quindi per una prefissata media $\mu = \frac{\sum x_i}{n}$, la variabilità minima nella distribuzione del carattere X si verifica quando $x_1 = \dots = x_n = \mu$. Per contro la massima variabilità si verifica se una sola unità statistica racchiude in sé l'intero ammontare $\sum x_i$ del carattere X lasciando alle altre (n-1) unità il valore 0, cioè quando $x_1 = \sum x_i = n\mu, x_2 = \dots = x_n = 0$.

Indice di Champernowne

Se tutte le modalità sono maggiori di zero, $x_i > 0$, posso sintetizzare i rapporti $\left(\frac{x_i}{\mu}\right)$ mediante la loro media geometrica.

Per le proprietà di μ tali rapporti (indipendenti dall'unità di misura) saranno alcune volte maggiori o uguali ad 1, altre volte minori o uguali a 1. Quindi il carattere X sarà tanto più variabile per quanto più tali rapporti si allontaneranno ad

1. La media geometrica di tali rapporti $\left(\prod_{i=1}^n \frac{x_i}{\mu}\right)^{\frac{1}{n}} = \frac{G}{\mu}$ risulterà sempre compresa tra 0 e 1 perché $G \leq \mu$.

$$\text{Indice di Champernowne: } Ch = 1 - \frac{G}{\mu}$$

Differenza semplice media

Esamino tutte le differenze tra le modalità a due a due $|x_i - x_j|$ facendone una sintesi tramite una opportuna media. Evidentemente è necessario considerare il valore assoluto delle differenze per evitare che ogni confronto $(x_i - x_j)$ s'annulli con il suo opposto. Poiché il numero di tutti i possibili confronti tra le n unità statistiche è $n(n-1)$, avendo escluso i confronti tra una unità e se stessa, la sintesi più immediata consiste nella media aritmetica di tali differenze prese in valore assoluto. Tale indice è la differenza semplice media: $\Delta = \frac{\sum_{i \neq j=1}^n |x_i - x_j|}{n(n-1)}$.

CONCENTRAZIONE PER CARATTERI TRASFERIBILI

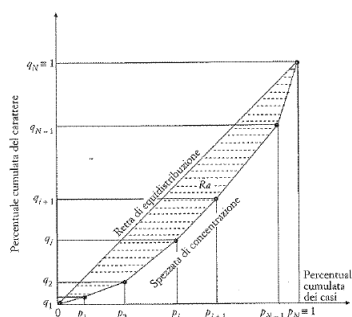
La concentrazione di un carattere X deriva dalla possibilità di "trasferire" l'ammontare del fenomeno da una unità statistica ad un'altra, avvicinandosi o allontanandosi dalla situazione di equidistribuzione dell'ammontare complessivo del carattere. Esempio classico è la distribuzione della concentrazione dei redditi tra un gruppo di unità statistiche ben

definite, sapendo che “il reddito di un paese è tanto più concentrato quanto più il reddito complessivo è posseduto da una frazione modesta delle unità statistiche, ovvero quanto più poveri vi sono in quel paese”.

Ordiniamo le modalità del carattere in senso non decrescente, per cui $x_{(1)}, \dots, x_{(n)}$ ed indichiamo con:

$$p_i = \text{frazione cumulata dei primi } i \text{ redditieri} = \frac{i}{n}, \quad q_i = \text{frazione cumulata reddito posseduto dai redditieri} = \frac{\sum_{j=1}^i x_j}{\sum_{j=1}^n x_j}$$

Dove p_i è la funzione di ripartizione empirica delle unità statistiche che assegna peso $\frac{1}{n}$ a ciascuna di esse, mentre q_i è una funzione cumulata della frazione di reddito posseduto dai primi i individui.



Una rappresentazione grafica efficace, chiamata curva di Lorenz, è ottenuta ponendo in ascissa i valori p_i , e in ordinata i corrispondenti valori q_i ed unendoli tra loro convenendo che $p_0 = q_0 = 0$.

Evidentemente $p_i \geq q_i$ perché, avendo ordinato i dati dal più povero al più ricco, il primo 10% (per esempio) delle unità può possedere al più il 10% (per esempio) del reddito complessivo. Se invece si ha $p_i = q_i$ si ha una situazione di equidistribuzione perché il 10%, il 20%, ..., il 90% possiede rispettivamente il 10%, il 20%, ... il 90%.

Quindi le differenze $(p_i - q_i)$ costituiscono misure dirette della concentrazione del carattere perché la concentrazione aumenta in modo diretto con il valore di tali differenze.

Indice di concentrazione di Gini

Si può anche dimostrare che l'indice di concentrazione di Gini è **in relazione con l'area di concentrazione, ovvero l'area che si trova tra la retta di equidistribuzione e la spezzata di concentrazione**. Più la spezzata è tesa verso la bisettrice degli assi più siamo in presenza di una concentrazione bassa; più è adagiata verso l'asse delle ascisse più ci avviciniamo ad una concentrazione alta. L'area di concentrazione varia da un minimo di 0 a un massimo di $\frac{1}{2}$.

$$R = \frac{n}{n-1} * 2 * \text{Area di concentrazione}$$

Indice concentrazione di Bonferroni

Confronta le medie progressive delle unità statistiche ordinate in senso non decrescente con la media complessiva. Nel caso di concentrazione minima si ha $\mu_{(i)} = \mu$, invece in caso di concentrazione massima si ha $\mu_{(i)} = 0$, $i = 1, \dots, n-1$ e $\mu_{(n)} = \mu$, $i = n$. Varia tra 0 (concentrazione minima) ad 1 (concentrazione massima).

Indice di Amato

La curva di Lorenz si può caratterizzare non solo mediante l'area ma anche tramite la lunghezza della spezzata di concentrazione. Nel caso di concentrazione minima la spezzata ha lunghezza pari alla diagonale del quadrato $\sqrt{2}$, nel caso di concentrazione massima ha lunghezza pari a quella di due lati del quadrato unitario, cioè 2.

5 FORMA DISTRIBUZIONE DI FREQUENZA

Due variabili possono avere la stessa variabilità e la stessa posizione ma differire per il peso dei valori più grandi o più piccoli rispetto al valore centrale, a causa del comportamento differenziato delle coda della distribuzione, cioè dalla forma della distribuzione di frequenza. La forma di una distribuzione di frequenza è stata storicamente descritta dalla mancanza di simmetria e dall'appuntimento rispetto al massimo.

ASIMMETRIA DI UNA DISTRIBUZIONE DI FREQUENZA

Se la media supera la mediana si parla di **asimmetria positiva** (in questo caso ci saranno meno valori a destra che a sinistra della media e la distribuzione avrà una coda a destra). Per contro, se la media è inferiore alla mediana, saremo in presenza di una distribuzione di frequenza che presenta una “coda” verso sinistra e, quindi, si parla di asimmetria negativa.

Indice di asimmetria di Pearson

Qualsiasi indice di forma della distribuzione dovrebbe prescindere dalla posizione e dalla variabilità per consentire confronti tra fenomeni di natura diversa. Per questo, per ogni variabile X , si definisce la variabile standardizzata $Z = Z(X)$ mediante la seguente trasformazione lineare:

$$Z(X) = \frac{X - \mu}{\sigma}$$

È agevole dimostrare che la variabile Z , oltre a non dipendere dall'unità di misura di X , possiede media aritmetica pari a 0 e varianza pari a 1: in questo la variabile Z possiede misure standard di posizione e variabilità.

L'indice di asimmetria di Pearson, in analogia con il calcolo delle probabilità, è misurata come il momento 3 standardizzato $\gamma_1(X) = Asym(X) = \bar{\mu}_3$ tuttavia al quadrato:

$$\beta_1 = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 \right]^2$$

Indice di asimmetria di Fisher

Perfettamente speculare all'analogo momento terzo standardizzato:

$$\gamma_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3 = \sqrt{\beta_1}$$

CURTOSI

Maggiore o minore appuntimento: sostanzialmente misura il rapporto relativo tra “corpo centrale” e “code” di una distribuzione di frequenza, e tale rapporto è connesso all’appiattimento o all’appuntimento della distribuzione di frequenza. In analogia alla teoria del calcolo della probabilità, indice di curtosi di Pearson è pari a:

$$\beta_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4$$

Vale 3 per una distribuzione teorica simmetrica di forma campanulare con due flessi equidistanti dal valore centrale; è maggiore di 3 per distribuzioni di frequenza relativamente “più appuntite”; è inferiore a 3 per distribuzioni piuttosto piatte. Per ottenere un indice da confrontare con lo zero si introduce l’indice di curtosi di Fisher: $\gamma_2 = \beta_2 - 3$

FORMA DI UNA DISTRIBUZIONE E ISTOGRAMMA

La costruzione di un istogramma presuppone alcuni elementi di arbitrarietà che possono condizionare ogni giudizio sulla forma di una distribuzione: come il numero delle classi (se troppo piccolo appiattisce ogni caratteristica di forma della distribuzione), l’ampiezza della classi, gli estremi del campo di definizione.

Un approccio molto buono per disegnare l’istogramma è l’**istogramma perequato**. In breve invece che fissare un numero di k classi si sostituisce alla sequenza di valori osservati (x_1, \dots, x_n) una funzione continua dell’ascissa x ottenuta come media ponderata di un opportuno intorno dei valori (x_1, \dots, x_n) (in altre parole ad ogni x si sostituisce il numero delle osservazioni che cadono vicino ad x). Si ottiene una funzione regolare che consente una visualizzazione continua dell’andamento complessivo della distribuzione di frequenza e, soprattutto, della sua forma. Bisogna scegliere la banda (tramite il parametro di smoothing), ovvero l’intorno di quanti valori includere nella perequazione di una unità x, ovvero quanti termini vicini ad x bisogna includere nella media ponderata.

METODI ESPLORATIVI PER UNA DISTRIBUZIONE DI FREQUENZA

Una prima rappresentazione grafica è quella definita **stem-and-leaf-plot** (=grafico a ramo e foglia) per la quale la distribuzione di frequenza dei dati si articola attorno a un “ramo”, ove sono i valori fondamentali, e alle “foglie”, ove sono i valori che si diramano dal ramo perché differenti nelle cifre decimali. Ad esempio se l’osservazione è costituita da numeri interi con virgola, scelgo come ramo il numero intero (e metto tutti i numeri interi in colonna) e come foglia il primo numero dopo la virgola (metto i numeri decimali in riga associandoli al numero intero a cui corrispondono).

Il **Box-plot** (o grafico a scatola) riesce a racchiudere, in una sola rappresentazione grafica, molti aspetti sintetici di una distribuzione di frequenza. Nella sua forma originaria ed essenziale, il box-plot è un grafico fondato sui quartili (inclusi il minimo e il massimo) rappresentati rispetto ad un conveniente asse che diviene una “scatola tra il primo e il terzo quartile” su cui è evidenziata la mediana della distribuzione. Il box-plot indicando la posizione della variabilità mediante l’ampiezza della scatola (che è la differenza interquartile) mostra anche l’eventuale asimmetria. Il box-plot mostra anche un cardine inferiore pari al minimo e un cardine superiore sopra il quale vengono mostrati gli eventuali outliers.

6 DISTRIBUZIONI STATISTICHE MULTIPLE

Una delle finalità più comuni nella raccolta di dati è la ricerca di relazioni del tipo causa-effetto, allo scopo di interpretare, prevedere, simulare, controllare i fenomeni reali. Le distribuzioni multiple di frequenze e le matrici dei dati costituiscono il modo più diffuso di raccogliere e presentare informazioni su una pluralità di variabili statistiche.

DISTRIBUZIONI MULTIPLE DI FREQUENZA

Quando su ogni unità appartenente ad una determinata popolazione si rilevano più caratteri (qualitativi e/o quantitativi) si parla di distribuzione multipla, o multivariata. Le variabili multiple si distinguono in discrete se tutte le variabili componenti sono discrete; in continue se tutte le variabili componenti sono continue; in miste (ma la determinazione è ambigua) se la rilevazione include variabili discrete e continue.

Per ottenere la **distribuzione doppia di frequenza di variabili entrambe discrete** (X, Y) occorre considerare per ciascuna coppia ordinata di valori (x_i, y_i) il numero n_{ij} delle unità statistiche che possiedono, ordinatamente, quelle modalità. Nel caso di variabili (X, Y) entrambe **continue** occorre individuare, sia per X che per Y , classi di modalità in modo che il numero n_{ij} rappresenti la frequenza assoluta delle unità statistiche le cui variabili possiedono valori che ricadono nella i -esima e j -esima classe di modalità, rispettivamente, per le variabili X e Y .

Un **legame statistico** tra X e Y , infatti, significa che il verificarsi di un certo valore per X implica un qualche effetto sul valore di Y , il che si traduce nel fatto che in corrispondenza di taluni valori di X si osserveranno più spesso taluni valori di Y , e non tutti con la stessa frequenza.

In una distribuzione di frequenza doppia ci troviamo di fronte ad una tabella a doppia entrata che registra quante volte (cioè la frequenza assoluta) una coppia di modalità si presenta contemporaneamente per X e per Y . Se indichiamo le k modalità di X con x_1, \dots, x_k e le h modalità di Y con y_1, \dots, y_h allora per frequenza assoluta n_{ij} intendiamo il numero di elementi tra gli n della popolazione che possiedono contemporaneamente le modalità x_i per X e y_j per Y .

Le frequenze assolute presenti nel riquadro sono connesse alla distribuzione doppia, quelle corrispondenti ai totali di riga riguardano esclusivamente la variabile X , mentre quelle corrispondenti ai totali di colonna riguardano esclusivamente la variabile Y . Per la loro collocazione, le frequenze corrispondenti alla variabile X e alla variabile Y sono definite frequenze assolute marginali e, ovviamente, costituiscono le distribuzioni di frequenza marginali univariate di X e Y ottenute dalla distribuzione doppia di frequenza (X, Y) , sommando rispettivamente le frequenze n_{ij} per riga e per colonna.

$$\sum_{i=1}^k \sum_{j=1}^h n_{ij} = \sum_{j=1}^h n_{i.} = \sum_{j=1}^h n_{.j} = n, \quad n_{i.} = \sum_{j=1}^h n_{ij}, \quad n_{.j} = \sum_{i=1}^k n_{ij}$$

La frequenza marginale $n_{i.}$ esprime il numero dei soggetti che possiedono la modalità x_i a prescindere da quello che avviene per il carattere Y .

TAB. 7.1. Distribuzione doppia di frequenza per la variabile (X, Y)

X ↓ Y →	y ₁	y ₂	...	y _j	...	y _b	Totale
x ₁	n ₁₁	n ₁₂	...	n _{1j}	...	n _{1b}	n _{1.}
x ₂	n ₂₁	n ₂₂	...	n _{2j}	...	n _{2b}	n _{2.}
...
x _i	n _{i1}	n _{i2}	...	n _{ij}	...	n _{ib}	n _{i.}
...
x _k	n _{k1}	n _{k2}	...	n _{kj}	...	n _{kb}	n _{k.}
Totale	n _{.1}	n _{.2}	...	n _{.j}	...	n _{.b}	n

DISTRIBUZIONI MARGINALI E CONDIZIONATE

Nel caso delle distribuzioni multiple le frequenze assolute si possono dividere per tutte le marginali deducibili da una distribuzione multipla, ottenendo così frequenze relative di differente significato. Nel caso più semplice di una distribuzione multipla si ottiene: i) le frequenze relative doppie della variabile (X, Y) $\rightarrow f_{ij} = \frac{n_{ij}}{n}$; ii) le frequenze relative marginali della variabile componente X $\rightarrow f_{i.} = \frac{n_{i.}}{n} = \sum_{j=1}^h f_{ij}$; iii) le frequenze relative marginali della variabile componente Y $\rightarrow f_{.j} = \frac{n_{.j}}{n} = \sum_{i=1}^k f_{ij}$.

Da una distribuzione doppia è anche possibile dedurre quella di una sola variabile componente dopo aver fissato la modalità dell'altra. Data una distribuzione doppia (X, Y), se fissiamo il valore x_i per la variabile X ed esaminiamo la distribuzione di Y limitatamente a quei soggetti che possiedono quel valore x_i per la variabile X, otteniamo la distribuzione condizionata di Y dato x_i cioè la variabile condizionata Y dato x_i . Tale distribuzione si indica con $Y|(X = x_i)$. Trattandosi di una variabile univariata, la variabile condizionata $Y|(X = x_i)$ sarà nota se conosciamo i valori che essa assume e le rispettive frequenze (assolute o relative). Pertanto, la distribuzione di frequenza della variabile condizionata $Y|(X = x_i)$ sarà così specificata:

$$\left\{ \begin{array}{lll} \text{Valori di } Y|(x = x_i), & y_1, y_2, \dots, y_h, & \text{Totale} \\ \text{Frequenze assolute,} & n_{i1}, n_{i2}, \dots, n_{ih}, & n_{i.} \\ \text{Frequenze relative,} & \frac{n_{i1}}{n_{i.}}, \frac{n_{i2}}{n_{i.}}, \dots, \frac{n_{ih}}{n_{i.}}, & 1 \end{array} \right.$$

Da una distribuzione doppia quindi si possono dedurre: una distribuzione per la variabile X, una per la variabile Y, k distribuzioni della Y condizionate alle corrispondenti modalità della X, h distribuzioni della X condizionate alle corrispondenti modalità delle Y.

INDIPENDENZA E MISURA DELLE RELAZIONI NELLE DISTRIBUZIONI MULTIPLE

Le eventuali relazioni deducibili da una distribuzione doppia devono essere riferite all'assenza di qualsiasi legame tra X e Y (e, per simmetria, tra Y e X), cioè alla indipendenza. Se, quindi, qualunque valore di X non modifica la distribuzione di Y (e viceversa), allora le distribuzioni condizionate di $Y|(X = x_i)$ non varieranno per $i = 1, 2, \dots, k$; il che implica che le frequenze relative condizionate $\frac{n_{ij}}{n_{i.}}$ saranno costanti al variare di $i = 1, 2, \dots, k$. Similmente, le distribuzioni condizionate di $X|(Y = y_j)$ non varieranno per $j = 1, 2, \dots, h$ e quindi le frequenze relative condizionate $\frac{n_{ij}}{n_{.j}}$ saranno costanti al variare di $j = 1, 2, \dots, h$, ciò avviene se e solo se si verifica:

$$\frac{n_{ij}}{n} = \frac{n_{i.}}{n} * \frac{n_{.j}}{n}$$

Se le variabili X e Y sono indipendenti allora le frequenze relative doppie sono esattamente il prodotto delle corrispondenti frequenze relative marginali.

$$Contingenze_{ij} = n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n} = n_{ij} - c_{ij}$$

Le contingenze esprimono la diversità tra le frequenze assolute osservate e le frequenze assolute che ci si attenderebbe per caratteri indipendenti. Si parla di attrazione tra le modalità x_i e x_j quando $Contingenza_{ij} > 0$, viceversa di parla di repulsione.

INDIPENDENZA IN MEDIA

Sia X una mutabile e Y una variabile quantitativa e sia (X, Y) la variabile doppia generata dall'osservazione congiunta di X e Y . In questo caso, nello studio della relazione doppia è possibile considerare un diverso concetto di dipendenza che coinvolge anche i valori assunti dalla variabile quantitativa. Si dice che Y è indipendente in media da X , se al variare delle modalità X , le medie delle distribuzioni condizionate di Y (medie condizionate) rimangono costanti.

$$Indipendenza\ in\ media \rightarrow M(Y|x_1) = M(Y|x_2) = \dots = M(Y|x_k) = M(Y)$$

L'indipendenza in distribuzione implica quella in media ma non è vero il contrario.

$$\eta^2_{(Y|x)} = \text{RAPPORTO CORRELAZIONE PEARSON} = \frac{\text{Devianza Between}}{\text{Devianza Totale}}$$

Vale 0 quando c'è indipendenza in media, vale 1 in assenza di variabilità interna ai gruppi, ovvero in una situazione in cui ad ogni valore prefissato di X corrisponde uno e uno solo valore di Y , cioè la massima dipendenza di Y da X .

CONNESSIONE TRA MUTABILI STATISTICHE

La connessione è un concetto introdotto da Gini per indicare la presenza di un legame molto generare tra variabili statistiche, fondato sull'analisi delle frequenze (assolute o relative) di una tabella multipla. Esso può essere specificato in termini di concordanza o discordanza se si individua anche la direzione dell'eventuale legame. Un indice di connessione sarà minimo se le frequenze doppie n_{ij} si conformeranno perfettamente a quelle c_{ij} ottenute nell'ipotesi di indipendenza. L'indice più utilizzato è l'indice del Chi quadro:

$$X^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - c_{ij})^2}{c_{ij}} = n \sum_{i=1}^k \sum_{j=1}^h \frac{(f_{ij} - f_{i.} * f_{.j})^2}{f_{i.} * f_{.j}}$$

CORRELAZIONE TRA VARIABILI STATISTICHE

La correlazione è una misura del legame statistico di tipo lineare tra due variabili. Il coefficiente di correlazione lineare di Bravaris-Pearson è il seguente:

$$\rho = \frac{\text{Covarianza}(X, Y)}{\sqrt{\text{Var}(X) * \text{Var}(Y)}} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}) * (x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$$

La correlazione soddisfa le seguenti proprietà:

- $-1 \leq \text{Corr}(X, Y) \leq +1$
- $\text{Corr}(X, Y) = \pm 1 \Leftrightarrow Y = \beta_0 \pm \beta_1 X$
- X, Y indipendenti $\rightarrow \rho = 0$
- $\text{Corr}(aX + b, cY + d) = \text{segno}(a * b) * \text{Corr}(X, Y)$

La **correlazione a blocchi** nasce quando le osservazioni riguardano due gruppi ben distinti e per i quali i valori medi delle variabili X e Y sono sufficientemente differenti, cioè siamo in presenza di due nuvole differenziate di punti. In tali casi, se si calcola la correlazione nei due gruppi separati si giunge ad un legame più basso di quello che si ottiene calcolando la correlazione sui gruppi aggregati.

La **correlazione spuria** nasce quando esiste un legame tra X e Y solo perché entrambe le variabili sono funzioni di una terza variabile Z che condiziona entrambi.

1 CALCOLO DELLE PROBABILITA'

La probabilità è un concetto primitivo, la probabilità è una misura perché associa al concetto primitivo una valutazione numerica. Gli elementi presenti negli esperimenti probabilistici sono l'incertezza del risultato, la riproducibilità dell'esperimento e l'equiprobabilità dei risultati.

DEFINIZIONE CLASSICA DI PROBABILITA'

$$\Pr(E) = \frac{m}{n}, \text{ ovvero numero di casi favorevoli su casi possibili}$$

DEFINIZIONE FREQUENTISTA

$$\Pr(E) = \lim_{n \rightarrow \infty} fr(E)/n, \quad \text{limite a cui tende il numero della frequenza relativa dei successi}$$

DEFINIZIONE SOGGETTIVISTA

La probabilità dell'evento E è la somma che un individuo coerente è disposto a scommettere in un gioco nel quale al verificarsi di E egli riceve dal banco un importo unitario.

SIGMA ALGEBRA O ALGEBRA DI BOOLE COMPLETA

Una classe \mathcal{A} di sottoinsiemi di uno spazio Ω (equivalentemente una collezione \mathcal{A} di eventi $E_1, E_2, \dots, E_n, \dots$) si dice una σ -algebra se soddisfa: 1) È non vuota; 2) Se $E \in \mathcal{A}$ allora $E^c \in \mathcal{A}$; 3) È chiusa rispetto unione numerabile, se $\{E_i\}_{i \in \mathbb{N}} \in \mathcal{A}$ allora $\bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$.

DEFINIZIONE DI PROBABILITA'

Siano $E_i, i = 1, 2, \dots$ eventi di Ω che formano una σ -algebra. La probabilità di un evento E_i è una funzione di insieme a valori reali (una corrispondenza tra elementi di Ω - cioè eventi - ed elementi di \mathbb{R} - cioè numeri reali -), che indicheremo con $\Pr(E_i)$ - e che leggeremo "probabilità di E_i " - che soddisfa i seguenti postulati:

1. $\Pr(E_i) \geq 0, \forall E_i \in \Omega$ (probabilità maggiore di zero)
2. $\Pr(\Omega) = 1$ (probabilità dell'evento certo è 1)
3. $E_i \cap E_j = \emptyset, \forall i \neq j \Rightarrow \Pr(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \Pr(E_i)$ (probabilità unione eventi disgiunti è somma di probabilità)

PROBABILITA' CONDIZIONATA

Si definisce la probabilità condizionata di B dato A, che si scrive $\Pr(B|A)$ nel modo seguente:

$$\text{Se } \Pr(A) > 0, \text{ allora } \Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

EVENTI INDIPENDENTI

Sul piano concettuale, due eventi A e B, inclusi in Ω , si dicono indipendenti se il verificarsi dell'evento A non modifica la probabilità dell'evento B cioè se, essendo $\Pr(A) > 0$, si ha: $\Pr(B|A) = \Pr(B)$.

Sul piano formale, due eventi si dicono indipendenti se e solo se:

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$

TEOREMA DELLE PROBABILITA' TOTALI

E_i , $i = 1, 2, \dots$ formano una partizione di Ω . Allora per ogni evento $A \subset \Omega$, si ha:

$$\Pr(A) = \sum_{i=1}^{+\infty} \Pr(E_i) \Pr(A|E_i)$$

TEOREMA DI BAYES

Se H_1, H_2, \dots, H_m sono eventi che costituiscono una partizione di Ω , allora per qualsiasi evento $E \subset \Omega$ la probabilità di H_i dato E è:

$$\Pr(H_i|E) = \frac{\Pr(H_i) \Pr(E|H_i)}{\sum_{j=1}^m \Pr(H_j) \Pr(E|H_j)}, \quad \forall i = 1, 2, \dots, m$$

Se E è un evento che può realizzarsi in conseguenza di m cause H_i , una delle quali certamente agisce e ognuna delle quali ha probabilità $\Pr(H_i)$ di agire, e $\Pr(E|H_i)$ è la probabilità che E si verifichi quando è noto che abbia agito la causa H_i , il teorema di Bayes esprime la probabilità a posteriori $\Pr(H_i|E)$, cioè la probabilità che avendo osservato l'evento E esso sia stato generato dalla causa H_i , in funzione della probabilità a priori $\Pr(H_i)$ e delle verosimiglianze $\Pr(E|H_i)$.

COMBINATORIA

		Gruppi	
		Distinguibili (sequenze ordinate)	Indistinguibili (sequenze non ordinate)
Estrazioni	Con ripetizioni	Disposizioni con ripetizioni N^n	Combinazioni con ripetizione $\binom{N+n-1}{n}$
	Senza ripetizioni	Disposizioni senza ripetizione $\frac{N!}{(N-n)!}$	Combinazioni senza ripetizione $\binom{N}{n}$

2 TEORIA DELLE VARIABILI CASUALI

VARIABILE ALEATORIA

Una variabile aleatoria è una funzione misurabile (l'immagine inversa di un intervallo aperto è un evento) a valori reali definita sullo spazio Ω . Per ogni $E \subset \Omega$, la v.c. X assume un valore reale x .

Se una funzione è misurabile la funzione inversa dà luogo ad un evento, cioè la funzione inversa è dotata di probabilità. Supponiamo perciò di dover calcolare la probabilità che si verifichi un evento $X \in A$, essendo A un intervallo del codominio; poiché ad A associato un evento E in \mathcal{A} tramite $E = X^{-1}(A)$, è possibile utilizzare l'assegnazione di probabilità sugli elementi di \mathcal{A} agli elementi del codominio. Infatti:

$$\Pr(X \in A) = \Pr_{\omega}(X^{-1}(A)) = \Pr_{\omega}(E)$$

V.C. DISCRETA

Una v.c. discreta è una corrispondenza tra gli eventi di $(\Omega, \mathcal{A}, \Pr)$ ed un insieme discreto (finito o numerabile) di numeri reali. Una v.c. discreta è **nota** se si conoscono i valori che può assumere e le rispettive probabilità.

VARIABILE CONTINUA

Una v.c. continua è una funzione misurabile e a valori reali che assegna ad ogni evento $E \subset \Omega$ di uno spazio di probabilità continuo un numero reale $x \in \mathbb{R}$. Una **v.c. continua è nota** se, per ogni x reale, è nota la funzione $F(x)$ oppure la funzione $f(x)$ definite dalla relazione seguente:

$$F(x) = \Pr(-\infty < X \leq x) = \Pr(X \leq x) = \int_{-\infty}^x f(w)dw; \quad f(x) = \frac{d}{dx}F(x)$$

La funzione $F(x)$ è detta **funzione di ripartizione** mentre la funzione $f(x)$ **funzione di densità**.

FUNZIONE DI RIPARTIZIONE

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^x dF(w) \begin{cases} \int_{-\infty}^x f(w)dw, & \text{se } X \text{ è una v.c. continua} \\ \sum_{x_i \leq x} p_i, & \text{se } X \text{ è v.c. discreta} \end{cases}$$

PROPRIETA' FUNZIONE DI RIPARTIZIONE

- $F(x)$ è non decrescente, cioè $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$
- $\lim_{x \rightarrow -\infty} F(x) = 0; \quad \lim_{x \rightarrow +\infty} F(x) = 1$
- $F(x)$ è continua da destra; cioè $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$

3 MOMENTI DI VARIABILI ALEATORIE

VALORE MEDIO DI g(X)

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) dF(x) = \begin{cases} \int_{-\infty}^{+\infty} g(x) f(x) dx, & \text{se } X \text{ è v.c. continua} \\ \sum_{-\infty}^{+\infty} g(x_i) p_i, & \text{se } X \text{ è v.c. discreta} \end{cases}$$

MOMENTI R-ESIMI

$$\mu_r = \mathbb{E}(X^r) = \int_{-\infty}^{+\infty} x^r dF(x) = \begin{cases} \int_{-\infty}^{+\infty} x^r f(x) dx, & \text{se } X \text{ è v.c. continua} \\ \sum_{i=1}^{+\infty} x_i^r p_i, & \text{se } X \text{ è v.c. discreta} \end{cases}$$

Si dimostra che $\mu_0 \equiv 1$.

MOMENTO PRIMO

Il **momento primo** μ_1 viene spesso indicato con la lettera $\mu = \mathbb{E}(X)$.

Si dimostra che $\mathbb{E}(X)$ è sempre compreso tra il massimo e il minimo dei valori che assume la v.c.; che il valore atteso della variabile scarto è pari a zero $\mathbb{E}(X - \mu) = 0$; che il valore medio è l'unico valore che minimizza $\mathbb{E}(X - \mu)^2$.

MOMENTI SCARTO

$$\bar{\mu}_r = \mathbb{E}(X - \mu)^r = \int_{-\infty}^{+\infty} (x - \mu)^r dF(x) = \begin{cases} \int_{-\infty}^{+\infty} (x - \mu)^r f(x) dx, & \text{se } X \text{ è v.c. continua} \\ \sum_{i=1}^{+\infty} (x_i - \mu)^r p_i, & \text{se } X \text{ è v.c. discreta} \end{cases}$$

Si dimostra che $\bar{\mu}_0 \equiv 1$ e $\bar{\mu}_1 \equiv 0$.

MOMENTO SECONDO DEI SCARTI-VARIANZA

Il **momento secondo** della v.c. scarto è chiamato varianza di una v.c.: $Var(X)$, oppure con σ^2 .

Per ogni v.c. X , si ha: $\bar{\mu}_2 = Var(x) = \mathbb{E}(X - \mu)^2 \geq 0$; $Var(cX) = c^2 Var(X)$.

Vale la relazione $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

COEFFICIENTE DI VARIAZIONE

$$Cv(x) = \frac{\sigma}{\mu}$$

VARIABILE CASUALE STANDARDIZZATA

Se la v.c. X non è degenera e possiede $\mathbb{E}(x) = \mu, Var(X) = \sigma^2 < +\infty$, si definisce **v.c. standardizzata** la v.c. Z :

$$Z = \frac{(X - \mathbb{E}(x))}{\sqrt{Var(X)}} = \frac{X - \mu}{\sigma}$$

MOMENTI R-ESIMI STANDARDIZZATI

$$\bar{\mu}_r = \mathbb{E}\left(\frac{X - \mu}{\sigma}\right)^r = \int_{-\infty}^{+\infty} \left(\frac{X - \mu}{\sigma}\right)^r dF(x) = \begin{cases} \int_{-\infty}^{+\infty} \left(\frac{X - \mu}{\sigma}\right)^r f(x) dx \\ \sum_{i=1}^{+\infty} \left(\frac{X - \mu}{\sigma}\right)^r p_i \end{cases}$$

$$\bar{\mu}_r = \mathbb{E}\left(\frac{X - \mu}{\sigma}\right)^r = \frac{\mathbb{E}(X - \mu)^r}{\sigma^r} = \frac{\bar{\mu}_r}{\sigma^r}, \quad r = 0, 1, 2, \dots$$

$$\gamma_1(X) = Asym(X) = \bar{\mu}_3, \quad \beta_2(X) = Kurt(X) = \bar{\mu}_4$$

MOMENTI ASSOLUTI

La condizione di esistenza sui momenti richiede maggiore rigore. Infatti, abbiamo già detto che $\mathbb{E}(X)$ esiste se e solo se $\mathbb{E}|X| < +\infty$, perché $|\mathbb{E}(X)| \leq \mathbb{E}|X|$, per le proprietà dei valori assoluti. Allora, poiché l'esistenza del valore medio del valore assoluto di una v.c. implica l'esistenza del valore medio della v.c., è opportuno introdurre una ulteriore classe di momenti.

$$\alpha_r = \mathbb{E}(|X|^r) = \int_{-\infty}^{+\infty} |X|^r dF(x) = \begin{cases} \int_{-\infty}^{+\infty} |x|^r f(x) dx, & \text{se } X \text{ è v.c. continua} \\ \sum_{i=1}^{+\infty} |x_i|^r p_i, & \text{se } X \text{ è v.c. discreta} \end{cases}$$

TEOREMI SUI MOMENTI

Poiché una serie (un integrale) converge se converge in modulo, si vede che l'esistenza dei momenti assoluti implica l'esistenza dei momenti (e non viceversa).

Se esiste il momento assoluto di ordine s , **esistono tutti i momenti di qualsiasi ordine $r \leq s$** .

Per ogni funzione convessa $g(\cdot)$, purché esista il valore medio di X , vale la **disuguaglianza di Jensen**:

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X))$$

TEOREMA DI MARKOV

Questa disuguaglianza permette di stabilire un limite superiore al valore di probabilità dalla sola conoscenza del valore atteso $E(X)$ a condizione che la variabile casuale sia definita non negativa:

$$\Pr(X > \alpha) \leq \frac{E(X)}{\alpha}, \quad \forall \alpha > 0$$

DISUGUAGLIANZA DI CEBYSEV

$$\Pr(|X - \mu| < k) \geq 1 - \frac{\sigma^2}{k^2}, \quad \forall k > 0$$

Regola del tre-sigma: in assenza di informazioni precise sul fenomeno in esame la probabilità di osservare valori che differiscano dal valore medio più di tre volte lo scarto quadratico medio è molto bassa (inferiore a 0,05).

FUNZIONE GENERATRICE DEI MOMENTI

La funzione generatrice dei momenti $G_X(t)$ (nel resto della trattazione per semplicità $G(t)$) della v.c. X è il valore medio della v.c. $\exp(tX)$, cioè:

$$G_X(t) = G\mathbb{E}(\exp(tX)) = \int_{-\infty}^{+\infty} \exp(tx) dF(x) = \begin{cases} \int_{-\infty}^{+\infty} \exp(tx) f(x) dx, & v.c. \text{ continua} \\ \sum_{i=1}^{+\infty} \exp(tx) p_i, & v.c. \text{ discreta} \end{cases}$$

TEOREMA SULLA FUNZIONE GENERATRICE DEI MOMENTI

Data una v.c. X , se esistono i momenti di X di qualsiasi ordine, allora $G(t)$ è esprimibile tramite di essi, se è finita per ogni t . Se $G(t)$ esiste in un intorno dell'origine, allora i momenti di X possono essere ricavati da $G(t)$ per derivazioni successive, essendo:

$$G(t) = \sum_{r=0}^{+\infty} \mu_r \frac{t^r}{r!}, \quad \left(\frac{d^r}{dt^r} G(t) \right)_{t=0} = \mu_r$$

FUNZIONE CARATTERISTICA

La funzione caratteristica $\varphi_X(t)$ (per semplicità $\varphi(t)$) della v.c. X è il valore medio della v.c. a valori complessi $\exp(itX)$, essendo $i = \sqrt{-1}$, cioè:

$$\varphi_X(t) = \varphi(t) = \mathbb{E}(\exp(itx)), \quad \text{vale } \varphi(t) = \sum_{r=0}^{+\infty} \mu_r \frac{i^r t^r}{r!}, \quad \left(\frac{d^r}{dt^r} \varphi(t) \right)_{t=0} = i^r \mu_r \text{ con } r = 1, 2, \dots$$

VARIABILI SIMMETRICHE

Una v.c. X si dice simmetrica se esiste un valore x_0 (detto centro di simmetria) tale che la probabilità che la v.c. X assuma valori alla destra di x_0 sia pari alla probabilità che la v.c. assuma valori alla sua sinistra, cioè se:

$$\Pr(X \geq x_0 + x) = \Pr(X \leq x_0 - x)$$

Se la v.c. è continua, la condizione diviene: $f(x_0 - x) = f(x_0 + x), \forall x$.

Infine se $x_0 \equiv 0$, la funzione di densità è una funzione pari.

DEFINIZIONE MODA

Se esiste un unico valore che rende massima la funzione di densità o la distribuzione di probabilità, esso è la moda della v.c. X e lo indicheremo con Mo , oppure con $Mo(X)$.

Se la v.c. X è continua, con funzione di densità che ammette le prime due derivate, allora la moda è quel valore Mo tale che: $f'(Mo) = 0; f''(Mo) < 0$.

DEFINIZIONE QUANTILE

$$F(x_p): \Pr(X \leq x_p) \geq p, \quad \Pr(X \geq x_p) \geq 1 - p, \quad \forall p \in (0,1)$$

Nel caso di una densità di probabilità la funzione di ripartizione F è continua $F(x_p) = p$.

Se $p = 0,25; 0,1; 0,01$ i quantili si chiamano rispettivamente quartili, decili, percentili.

Il quantile $x_{0,5}$ è la mediana Me della v.c. X , indicata anche con $Me(X)$.

L'intervallo interquartile è $I_{QR} = x_{0,75} - x_{0,25}$.

4 VARIABILI DOPPIE

VARIABILI ALEATORIE DOPPIE

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) du dv, \quad \Pr((X,Y) \in A) = \int_A f_{X,Y}(u,v) du dv$$
$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy, \quad F_X(x) = \int_{-\infty}^x f_X(u) du$$

VARIABILI CONDIZIONATE

$$\Pr(Y = y|X = x) = \frac{\Pr(X = x \cap Y = y)}{\Pr(X = x)}, \quad f_Y(y|X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_X(x) > 0$$

VALORE ATTESI DI VARIABILI ALEATORIE DOPPIE

Per v.c. componenti indipendenti, la fattorizzazione della distribuzione multipla della probabilità implica anche la fattorizzazione dei momenti (posso scrivere il valore atteso congiunto come il prodotto di valori attesi).

COVARIANZA

$$\text{Cov}(X,Y) = \mathbb{E}[(X - \mu_x)(Y - \mu_y)] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

La covarianza è invariante per traslazione ma non per cambiamento di scala.

$$\text{Cov}(\beta_0 + \beta_1 X, Y) = \text{Cov}(X, \beta_0 + \beta_1 Y) = \beta_1 \text{Cov}(X, Y)$$

$$\text{Cov}(\beta_0 + \beta_1 X, \alpha_0 + \alpha_1 Y) = \alpha_1 \beta_1 \text{Cov}(X, Y)$$

COEFFICIENTE DI CORRELAZIONE

$$\text{Corr}(X,Y) = \rho_{xy} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \mathbb{E}\left[\left(\frac{X - \mu_x}{\sigma_x}\right)\left(\frac{Y - \mu_y}{\sigma_y}\right)\right]$$

Se la correlazione è pari a 1 allora la v.c. doppia è degenera, in tal caso lo spazio di probabilità non è a due dimensioni, ma ad una soltanto: $\text{Corr}(X,Y) = \pm 1 \Leftrightarrow \Pr(Y = \beta_0 \pm \beta_1 X) = 1$.

CORRELAZIONE E INDIPENDENZA

Se X e Y sono v.c. indipendenti, allora $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ per cui $\text{Corr}(X,Y) = 0$; ma non è necessariamente vero il viceversa. Un coefficiente di correlazione pari a 0 (le v.c. si dicono incorrelate) non implica l'indipendenza. Ciò è ragionevole perché la correlazione misura un legame probabilistico di tipo lineare e quindi l'incorrelazione esclude solo un legame lineare tra v.c. mentre l'indipendenza esclude ogni tipo di legame probabilistico.

5 TRASFORMAZIONI DI VARIABILI CASUALI E CONVERGENZA IN PROBABILITA'

TEOREMA FUNZIONE GENERATRICI

Se (X_1, \dots, X_m) è una v.c. multivariata a componenti indipendenti, ciascuna delle quali possiede funzione caratteristica $\varphi_i(t)$ allora la v.c. $S_m = X_1 + \dots + X_m$ possiede funzione caratteristica $\varphi(t)$ definita dalla relazione:

$$\varphi(t) = \varphi_1(t) * \dots * \varphi_m(t)$$

Il teorema, se esistono ben definite nell'intorno dell'origine, vale anche per le funzioni generatrici.

SUCCESSIONI DI VARIABILI ALEATORIE

Definiamo successione di v.c. X_n una regola che associa ad ogni $n = 1, 2, \dots$ una v.c. la cui funzione di ripartizione è $F_n(x)$. Scriviamo: $X_n \sim F_n(x)$, $n = 1, 2, \dots$

CONVERGENZA IN DISTRIBUZIONE

$$X_n \xrightarrow{d} X \text{ cioè converge in distribuzione} \Leftrightarrow \lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

CONVERGENZA IN PROBABILITA'

$$X_n \xrightarrow{p} X \text{ (plim } X_n = X) \text{ cioè converge in probabilità} \Leftrightarrow \lim_{n \rightarrow +\infty} \Pr(|X_n - X| < \epsilon) = 1$$

CONVERGENZA IN MEDIA QUADRATICA

$$X_n \xrightarrow{m} X \text{ cioè converge in media quadratica} \Leftrightarrow \lim_{n \rightarrow +\infty} E(X_n - X)^2 = 0$$

CONVERGE QUASI CERTAMENTE

$$X_n \xrightarrow{qc} X \text{ cioè converge quasi certamente} \Leftrightarrow \lim_{n \rightarrow +\infty} \Pr(|X_n - X| < \epsilon, \forall m \geq n) = 1, \forall \epsilon$$

IMPLICAZIONI DI CONVERGENZA

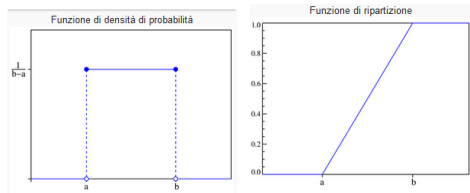
$$X_n \xrightarrow{qc} X \Rightarrow X_n \xrightarrow{p} X, \quad X_n \xrightarrow{m} X \Rightarrow X_n \xrightarrow{p} X, \quad X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

TEOREMA DI MANN E WALD

$$\begin{aligned} X_n \xrightarrow{d} X &\Rightarrow g(X_n) \xrightarrow{d} g(X), & g(\cdot) \text{ funzione continua con probabilità 1} \\ X_n \xrightarrow{p} X &\Rightarrow g(X_n) \xrightarrow{p} g(X), & g(\cdot) \text{ funzione continua con probabilità 1} \\ X_n \xrightarrow{qc} X &\Rightarrow g(X_n) \xrightarrow{qc} g(X), & g(\cdot) \text{ funzione continua con probabilità 1} \end{aligned}$$

6 MODELLI PER VARIABILI ALEATORIE CONTINUE

VARIABILE UNIFORME CONTINUA



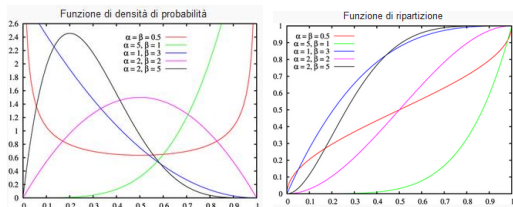
Distribuzione di probabilità continua che è uniforme su un insieme, ovvero che attribuisce la stessa probabilità a tutti i punti appartenenti ad un dato intervallo $[a, b]$ contenuto nell'insieme.

$$X \sim U(\theta_1, \theta_2)$$

$$f(x) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)}, & \text{per } \theta_1 < x < \theta_2 \\ 0, & \text{altrove} \end{cases} \rightarrow F(X) = \int_{-\infty}^x f(x) dz = \int_{\theta_1}^x f(x) dz = \begin{cases} 0, & \text{per } x \leq \theta_1 \\ \frac{x - \theta_1}{(\theta_2 - \theta_1)}, & \text{per } \theta_1 < x < \theta_2 \\ 1, & \text{per } x \geq \theta_2 \end{cases}$$

$$E(X) = \frac{\theta_1 + \theta_2}{2}, \quad V(X) = \frac{(\theta_2 - \theta_1)^2}{12}$$

DISTRIBUZIONE BETA



Questa distribuzione trova particolare utilizzo nella statistica bayesiana perché governa la probabilità p di un processo di Bernoulli a posteriori dell'osservazione di $\alpha - 1$ "successi" e $\beta - 1$ "fallimenti", quando p è a priori distribuita uniformemente tra 0 e 1.

$$X \sim Be(\theta_1, \theta_2) \text{ su intervallo unitario}$$

$$f(x) = \frac{1}{B(\theta_1, \theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1}, \quad \text{con } 0 < x < 1$$

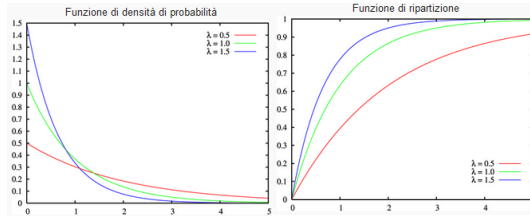
$$B(\theta_1, \theta_2) = \int_0^1 x^{\theta_1-1} (1-x)^{\theta_2-1} dx = \frac{\Gamma(\theta_1) \Gamma(\theta_2)}{\Gamma(\theta_1 + \theta_2)}, \quad \Gamma(n+1) = n! \text{ con } n \text{ intero}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$E(X) = \frac{\theta_1}{\theta_1 + \theta_2}$$

$$X \sim Be(1, 1) \rightarrow X \sim U(0, 1)$$

DISTRIBUZIONE ESPONENZIALE O ESPONENZIALE NEGATIVA



Descrive la “durata di vita” di un fenomeno che non invecchia, ovvero che è privo di memoria.

$$X \sim \text{En}(\theta)$$

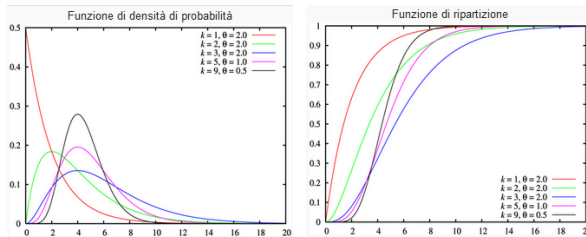
$$f(x) = \begin{cases} \theta e^{-\theta x}, & x > 0 \\ 0, & \text{altrove} \end{cases}, \quad F(x) = 1 - e^{-\theta x}$$

$$E(X) = \frac{1}{\theta}, \quad V(X) = \frac{1}{\theta^2}$$

$$\text{ASSENZA DI MEMORIA} = P(X > a + b | a) = \frac{P(X > a + b)}{P(X > a)} = \frac{e^{-(a+b)x}}{e^{-ax}} = e^{-bx} = P(X > b)$$

$$\text{Somma di Esponenziali} = \text{Gamma} \rightarrow X_1 + \dots + X_n \sim \text{En}(\theta) \Rightarrow \sum_{i=1}^n X_i = \Gamma(n, \theta)$$

DISTRIBUZIONE GAMMA



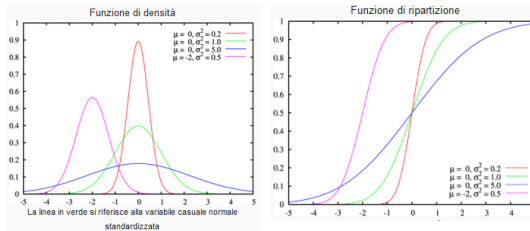
Viene utilizzata come modello generale dei tempi di attesa nella teoria delle code, soprattutto qualora siano importanti effetti che rimuovano "l'assenza di memoria" della distribuzione esponenziale.

$$X \sim \Gamma(\beta, \theta)$$

$$f(x) = \frac{\theta^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\theta x}$$

$$E(X) = \frac{\beta}{\theta}, \quad V(X) = \frac{\beta}{\theta^2}$$

$$\text{Somma di Gamma} = \text{Gamma} \rightarrow X_1 + \dots + X_n \sim \Gamma(\beta_i, \theta) \Rightarrow \sum_{i=1}^n X_i = \Gamma(\beta_1 + \dots + \beta_n, \theta)$$



DISTRIBUZIONE NORMALE

Distribuzione di probabilità continua che è spesso usata come prima approssimazione per descrivere variabili casuali a valori reali che tendono a concentrarsi attorno a un singolo valor medio.

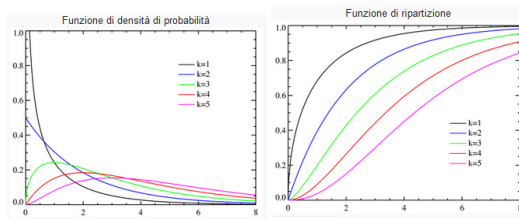
La distribuzione normale è considerata il caso base delle distribuzioni di probabilità continue a causa del suo ruolo nel teorema del limite centrale.

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

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

$$E(X) = \mu, \quad V(X) = \sigma^2, \quad \text{Asym}(X) = 0, \quad \text{Kurt} = 3$$

$$\text{Somma di Normali} = \text{Normale} \rightarrow X_1 + \dots + X_n \sim \Gamma(\mu_i, \sigma_i^2) \Rightarrow \sum_{i=1}^n a_i X_i = N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$



DISTRIBUZIONE CHI-QUADRO CENTRATA

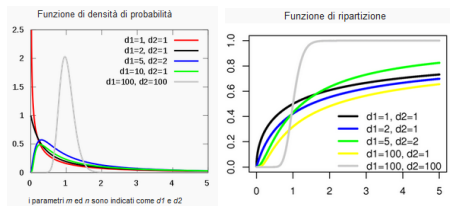
Distribuzione di probabilità della somma dei quadrati di variabili aleatorie normali indipendenti.

$$X \sim \chi^2(g)$$

$$X_1 + \dots + X_n \sim N(0,1) \Rightarrow \sum_{i=1}^n X_i = \chi^2(n)$$

$$X_1 + \dots + X_n \sim \chi^2(g_i) \Rightarrow \sum_{i=1}^n X_i = \chi^2(g_1 + \dots + g_n)$$

$$E(X) = k, \quad V(X) = 2k$$



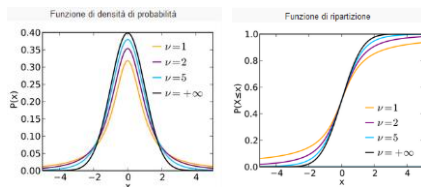
DISTRIBUZIONE FISHER

Distribuzione di probabilità continua che regola il rapporto "riscalato" tra due variabili aleatorie che seguono due distribuzioni Chi-Quadrato.

$$X \sim \mathcal{F}(g_1, g_2)$$

$$X_1 \sim \chi^2(g_1), X_2 \sim \chi^2(g_2) \Rightarrow \frac{\frac{X_1}{g_1}}{\frac{X_2}{g_2}}$$

$$E(X) = g_2 / (g_2 - 2), g_2 > 2$$



T-STUDENT

Distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, la prima con distribuzione normale e la seconda, al quadrato, segue una distribuzione chi quadrato.

$$T_g \sim \frac{N(0,1)}{\sqrt{\chi^2(g)/g}}$$

VARIABILI ORDINATE

$$\min(X_1, \dots, X_n) = Y_1 \leq \dots \leq Y_n = \max(X_1, \dots, X_n)$$

VARIABILE MASSIMO Y_n

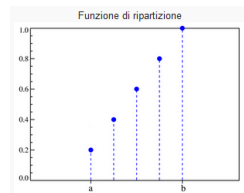
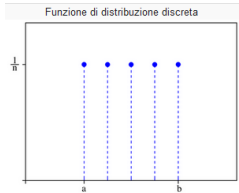
$$\rightarrow F_n(y) = \Pr(Y_n \leq y) = \Pr((X_1 \leq y) \cap \dots \cap (X_n \leq y)) = \Pr((X_1 \leq y)) * \dots * \Pr((X_n \leq y)) = F(y) * \dots * F(y) = [F(y)]^n$$

VARIABILE MINIMO

$$Y_1 \rightarrow F_1(y) = \Pr(Y_1 \leq y) = 1 - \Pr(Y_1 > y) = 1 - \Pr((X_1 > y) \cap \dots \cap (X_n > y)) = 1 - [1 - F(y)]^n$$

7 MODELLI PER VARIABILI DISCRETE

UNIFORME DISCRETA



Prova che genera la v.c. Uniforme discreta si può assimilare all'estrazione da un'urna che contiene n palline numerate da 1 a n. Evento è il seguente: "Si è estratta la pallina che reca il numero x".

$$X \sim Ud(n)$$

$$\Pr(X = x) = \frac{1}{n}, \quad F(X) = \begin{cases} 0, & x < 1 \\ \frac{k}{n}, & k \leq x < k+1, \quad k = 1, \dots, n-1 \\ 1, & x \geq n \end{cases}$$

$$E(X) = \frac{n+1}{2}, \quad V(X) = \frac{n^2-1}{12}$$

BERNOULLI

Prova nella quale si è interessati a verificare se E si è verificato (x=1) oppure no (x=0).

$$X \sim Ber(\theta)$$

$$\Pr(X = x) = \theta^x (1 - \theta)^{1-x}, \quad \text{con } x = \{0, 1\}, \quad F(X) = \begin{cases} 0, & x < 0 \\ 1 - \theta, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

$$E(X) = \theta, \quad V(X) = \theta(1 - \theta), \quad Asym(X) = \frac{1 - 2\theta}{\sqrt{\theta(1 - \theta)}}$$

IPERGEOMETRICA

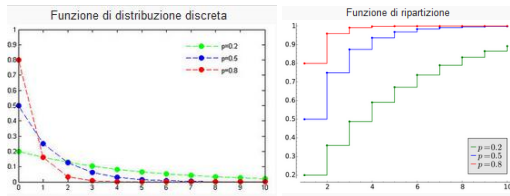
Prova: estrazione n palline senza ripetizione da un'urna che contiene H palline, b bianche e H-b nere. Evento=#successi.

$$X \sim Ip(n, b, H)$$

$$\Pr(X = x) = \frac{\binom{b}{x} \binom{H-b}{n-x}}{\binom{H}{n}}$$

$$E(X) = \frac{nb}{H}, \quad V(X) = n\theta(1 - \theta) \frac{H-n}{H-1}$$

GEOMETRICA



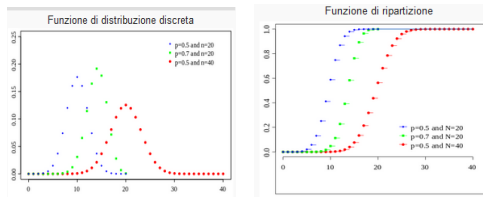
Numero prove occorrenti affinché si verifichi il primo successo.

$$X \sim \text{Geom}(\theta)$$

$$\Pr(X = x) = \theta(1 - \theta)^{x-1}$$

$$E(X) = \frac{1}{\theta}, \quad V(X) = \frac{1 - \theta}{\theta^2},$$

BINOMIALE



Descrive l'evento "numero dei successi" in un processo di Bernoulli (una serie di prove ripetute). Assimilabile ad un'estrazione con ripetizione di n palline da un'urna che ne contiene H di cui b bianche e $H - b$ nere. La probabilità di successo è $\theta = \frac{b}{H}$.

$$X \sim \text{Bin}(n, \theta)$$

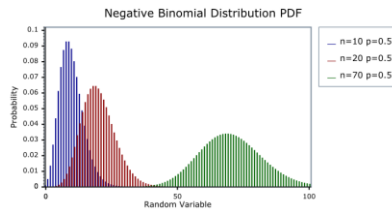
$$\Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad F(X) = \sum_{r=0}^x \binom{n}{r} \theta^r (1 - \theta)^{n-r}$$

$$E(X) = n\theta, \quad V(X) = n\theta(1 - \theta),$$

Proprietà riproduttiva $\rightarrow X_i \sim \text{Bin}(n_i, \theta) \rightarrow \text{Somma } X_i \sim \text{Bin}(n_1 + \dots + n_m, \theta)$

Per $n \rightarrow +\infty$ si approssima con normale

BINOMIALE NEGATIVA O PASCAL



Descrive il numero di fallimenti precedenti il successo k-esimo in un processo di Bernoulli (serie di prove bernoulliane)
 Numero X di sottoprobe occorrenti affinché si verifichino k successi.

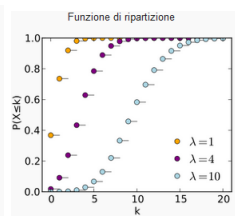
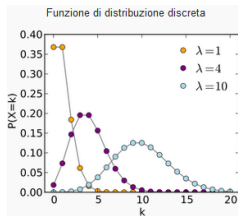
$$X \sim Bn(k, \theta)$$

$$\Pr(X = x) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}$$

$$E(X) = \frac{k}{\theta}, \quad V(X) = \frac{k(1-\theta)}{\theta^2},$$

Proprietà riproduttiva $\rightarrow X_i \sim Bn(k_i, \theta) \rightarrow \text{Somma } X_i \sim Bn(k_1 + \dots + k_m, \theta)$

POISSON



Esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero θ . Ad esempio, il numero di chiamate ricevute in un call-center in un determinato arco temporale.

$$X \sim Po(\theta)$$

$$\Pr(X = x) = \frac{\theta^x}{x!} e^{-\theta}, \quad \lim_{\substack{n \rightarrow +\infty \\ \theta \rightarrow 0}} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{e^{-\theta} \theta^x}{x!}$$

$$E(X) = VAR(X) = \theta$$

Proprietà riproduttiva $\rightarrow X_i \sim Po(\theta_i) \rightarrow \text{Somma } X_i \sim Po(\theta_1 + \dots + \theta_m)$

8 TEOREMI DEI LIMITI CENTRALI

LEGGE DEBOLE DEI GRANDI NUMERI

Data una successione di variabili aleatorie X_1, X_2, \dots, X_n IID con $E(X_i) = \mu, \text{Var}(X_i) < +\infty$
 $\Rightarrow \bar{X}_n \xrightarrow{p} \mu$ (ovvero la media campionaria converge in probabilità al valore medio della v. c.)
$$\lim_{n \rightarrow +\infty} \Pr(|\bar{X}_n - \mu| < \epsilon) = 1$$

LEGGE FORTE DEI GRANDI NUMERI

Data una successione di variabili aleatorie X_1, X_2, \dots, X_n IID con $E(X_i) = \mu, E(X_i^4) < +\infty$
 $\Rightarrow \bar{X}_n \xrightarrow{qc} \mu$ (ovvero la media campionaria converge quasi certamente al valore medio della v. c.)
$$\lim_{n \rightarrow +\infty} \Pr(|\bar{X}_n - \mu| < \epsilon, \forall m \geq n) = 1, \forall \epsilon$$

Le legge debole afferma che al crescere di n tende a 1 la probabilità che la media empirica sia entro l'intervallo $\mu \pm \epsilon$, per ogni ϵ comunque piccolo. Invece le leggi forti affermano che al crescere di n tende a 1 la probabilità che, per quanto grande si scelga una sequenza di prove a partire da n , cioè $n, n+1, n+2, \dots, n+m$, media empirica sia entro l'intervallo $\mu \pm \epsilon$, per ogni ϵ comunque piccolo

TEOREMA DI DE MOIVRE-LAPLACE

X_n IID come variabili di bernoulli con $E(X_i) = \theta$,

$$\Rightarrow Z_n = \frac{\sum X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{d} Z \sim N(0,1)$$

Applicazione teorema di lindeberg a una bernoulliana

TEOREMA CENTRALE DEL LIMITE- LINDEBERG/LEVY

X_n IID come variabili di bernoulli con $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2 < +\infty$

$$\Rightarrow Z_n = \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

9 FORMULARIO

COMBINATORIA

$$(a+b)^N = \sum_{i=0}^N \binom{N}{i} a^i b^{N-i}, \quad \binom{N}{n} = \binom{N}{N-n}, \quad \binom{N}{n} = \binom{N}{n-1} \binom{N-1}{n-1}, \quad \sum_{i=0}^N \binom{N}{i} = 2^N, \quad \sum_{i=0}^N (-1)^i \binom{N}{i} = 0$$

ESPONENZIALI

$$1^x = 1, \quad 0^x = 0, \quad a^0 = 1, \quad \text{con } a \neq 0$$
$$a^{-n} = \left(\frac{1}{a}\right)^n = \frac{1}{a^n}, \quad a^{\frac{n}{m}} = \sqrt[m]{a^n}, \quad a^{-\frac{n}{m}} = \frac{1}{a^{\frac{n}{m}}} = \frac{1}{\sqrt[m]{a^n}}$$
$$a^x * a^y = a^{x+y}, \quad a^x : a^y = a^{x-y}, \quad a^x * b^x = (a * b)^x, \quad a^x : b^x = (a : b)^x, \quad (a^x)^y = a^{x*y}$$

LOGARITMI

$$\log_a a = 1, \quad \log_a 1 = 0$$
$$a^{\log_a x} = \log_a a^x = x$$
$$\log_a(xy) = \log_a(x) + \log_a(y), \quad \log_a\left(\frac{x}{y}\right) = \log_a(x) - \log_a(y)$$
$$\log_a(x^c) = c \log_a x, \quad \log_a(x) = \frac{\log_y(x)}{\log_y(a)}, \quad \log_a(x) = \frac{1}{\log_x(a)}$$

SOMMATORIE

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i, \quad \sum_{i=1}^n \sum_{j=1}^m (x_i y_j) = \sum_{i=1}^n x_i * \sum_{j=1}^m y_j, \quad \sum_{i=1}^n (x_i * y_i) \neq \sum_{i=1}^n x_i * \sum_{i=1}^n y_i, \quad \sum_{i=1}^n \left(\frac{x_i}{y_i}\right) \neq \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i}$$

DERIVATE

$$(f(g(x)))' = f'(g(x)) * g'(x), \quad (x^\alpha)' = \alpha x^{\alpha-1}, \quad (\log_b x)' = \frac{\log_b e}{x}, \quad (\ln x)' = \frac{1}{x}, \quad (e^x)' = e^x, \quad (a^x)' = a^x \ln a$$

INTEGRALI

$$\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1}, \quad \int \frac{1}{x} dx = \ln|x|, \quad \int \frac{f'(x)}{f(x)} dx = \ln|f(x)|, \quad \int \log_b x dx = x \log_b x - x \log_b e, \quad \int e^{ax} dx = \frac{e^{ax}}{a}$$

1 CAMPIONI CASUALI E DISTRIBUZIONI CAMPIONARIE

INFERENZA

Rispetto alla probabilità l'inferenza capovolge punto di vista: si suppone di conoscere soltanto i risultati della prova, e non la popolazione da cui vengono. Ogni inferenza si basa sulla specificazione accurata dei seguenti elementi:

1. **Popolazione di riferimento:** insieme informazioni statistiche che esauriscono il problema oggetto dello studio. Nel seguito "popolazione x " sarà sinonimo di "v.c. X ", quindi la popolazione sarà $X \sim f(x; \theta)$.
2. **Procedura di raccolta e selezione delle informazioni:** dalla popolazione viene estratto un campione casuale
3. **Tecnica inferenziale per giungere dal risultato parziale alla popolazione**
 - Nella **teoria della stima** si cerca di determinare un valore numerico per il parametro (o vettore di parametri) θ che caratterizza la popolazione $X \sim f(x; \theta)$ sulla base delle informazioni campionarie desumibili da (x_1, x_2, \dots, x_n) .
 - Nella **teoria del test** (o verifica) delle ipotesi statistiche si controlla quale tra due affermazioni complementari (dette "ipotesi") possa essere ritenuta maggiormente verosimile, sulla base delle informazioni campionarie.
 - Nella **teoria degli intervalli di confidenza** si cerca di determinare, sulla base dei dati campionari, un intervallo di valori reali (o una regione nel caso di un vettore di parametri) in cui riporre una prefissata "fiducia" per il parametro θ .
4. **Validità statistica della procedura utilizzata**

Tutte le procedure inferenziali si articoleranno in due momenti successivi:

 - stabilire cosa si intende per procedura ottimale;
 - individuare metodi statistici che producano procedure ottimali

CAMPIONE CASUALE

Campione casuale: collezione di v.c. $\underline{X} = (X_1, X_2, \dots, X_n)$ mutuamente indipendenti e identicamente distribuite, ottenuta con procedimento di estrazione dalla v.c. $X \sim f(x; \theta)$.

Il **campione osservato** è la determinazione numerica del campione casuale

STATISTICA

Una **statistica** $T_n = T(X_1, X_2, \dots, X_n)$ è qualunque funzione a valori reali del campione casuale (X_1, X_2, \dots, X_n) che non dipende da quantità incognite;

La **statistica calcolata** $t_n = T(x_1, x_2, \dots, x_n)$ è il valore della statistica T_n calcolata sul campione osservato (x_1, x_2, \dots, x_n) .

Nella teoria della stima la statistica T_n è chiamata stimatore, mentre nella teoria del test delle ipotesi T_n è chiamata statistica-test.

La **distribuzione campionaria** di T_n è la distribuzione di probabilità della statistica T_n , calcolata sullo spazio di probabilità del campione casuale.

STATISTICHE COMUNI

$$MEDIA CAMPIONARIA \rightarrow \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$VARIANZA CAMPIONARIA \rightarrow \tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad CORRETTA \rightarrow S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$$COEFFICIENTE DI VARIAZIONE CAMPIONARIA \rightarrow CV_n = \frac{S_n}{\bar{X}_n}$$

$$MOMENTO R - ESIMO \rightarrow M_r = M_{r,n} = \frac{1}{n} \sum_{i=1}^n (X_i)^r$$

$$COEFFICIENTE CORRELAZIONE CAMPIONARIO \rightarrow R_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$$

MOMENTI CARATTERISTICI

$$\mathbb{E}(X) = \mu; \text{Var}(x) = \sigma^2; \text{Asym}(X) = \gamma_1; \text{Kurt}(X) = \beta_2$$

$$\mathbb{E}(\bar{X}_n) = \mu; \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

$$\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1} \sum_{i=1}^n (X_i)^2 - \frac{n}{n-1} \sum_{i=1}^n (\bar{X}_n)^2\right) = \sigma^2$$

$$E(\tilde{S}_n^2) = \mathbb{E}\left(\frac{n-1}{n} S_n^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \frac{\sigma^2(n-1)}{n}$$

METODO DELTA

Approssimo momenti caratteristici di una funzione non lineare di una v.c. i cui momenti sono noti:

$$X: \mathbb{E}(X) = \mu, \text{Var}(X) = \sigma^2 \rightarrow \mathbb{E}[g(X)] \cong g(\mu) + \frac{1}{2} \sigma^2 g''(\mu) \rightarrow \text{Var}[g(X)] \cong [g'(\mu)]^2 \text{Var}(X)$$

DISTRIBUZIONI CAMPIONARIE NOTEVOLI

Alcune distribuzioni campionarie vengono definite notevoli perché forniscono risultati esatti sotto condizioni comuni nell'inferenza.

$$\begin{cases} (X_1, \dots, X_n) \text{ con } X_i \sim N(\mu, \sigma^2) \\ (Y_1, \dots, Y_m) \text{ con } Y_i \sim N(v, t^2) \end{cases} \rightarrow \begin{cases} \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \\ \bar{Y}_m \sim N\left(v, \frac{t^2}{m}\right) \end{cases} \rightarrow T_{n,m} = a\bar{X}_n + b\bar{Y}_m \sim N\left(a\mu + bv, a^2 \frac{\sigma^2}{n} + b^2 \frac{t^2}{m}\right)$$

TEOREMA

Se il campione (X_1, \dots, X_n) è generato da una v.c. $X \sim N(\mu, \sigma^2)$ allora \bar{X}_n e S_n^2 sono variabili aleatorie indipendenti, e viceversa, cioè se \bar{X}_n e S_n^2 sono indipendenti allora il campione è generato da v.c. Normali.

DISTRIBUZIONE CHI QUADRO VARIANZE

$$\begin{cases} (X_1, \dots, X_n) \text{ con } X_i \sim N(\mu, \sigma^2) \\ (Y_1, \dots, Y_m) \text{ con } Y_i \sim N(v, t^2) \end{cases} \rightarrow \begin{cases} S_n^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2 \\ \tilde{S}_n^2 \sim \frac{\sigma^2}{n} \chi_{n-1}^2 \end{cases} \rightarrow S_{x,n}^2 + S_{y,m}^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2 + \frac{t^2}{m-1} \chi_{m-1}^2$$

DISTRIBUZIONE T-STUDENT

Il risultato è importante perché mentre le distribuzioni di \bar{X}_n e S_n^2 dipendono da σ^2 la seguente distribuzione no:

$$X \sim N(\mu, \sigma^2) \rightarrow T_n = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} = \frac{\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}}{\sqrt{S_n^2 / n}} \sim t - student_{n-1}$$

RAPPORTO DI NORMALI CON VARIANZA DIVERSA

$$\begin{cases} X \sim N(\mu, \sigma^2) \\ Y \sim N(v, t^2) \end{cases} \rightarrow T_{n,m} = \frac{S_{x,n}^2}{S_{y,m}^2} \sim \frac{\sigma^2}{t^2} F_{n-1, m-1}$$

2 TEORIA DEGLI STIMATORI

$X \sim f(x; \theta)$, nota eccetto vettore dei parametri $\theta \in \Omega(\theta)$

Estraggo un campione casuale $\underline{X} = (X_1, \dots, X_n)$ la cui determinazione numerica è $\underline{x} = (x_1, \dots, x_n)$

STIMA $\rightarrow t_n = T(x_1, \dots, x_n)$ calcolo della funzione T sui numeri reali estratti

STIMATORE $\rightarrow T_n = T(X_1, \dots, X_n)$ funzione di variabili aleatorie ovvero una variabile aleatoria

SUFFICIENZA

Sul piano intuitivo uno stimatore sufficiente T_n racchiude ed esaurisce tutte le informazioni riguardanti θ e contenute nel campione casuale.

$\underline{X} = (X_1, \dots, X_n)$ campione casuale generato da $X \sim f(x; \theta), \theta \in \Omega(\theta)$ incogniti \Rightarrow

T_n è sufficiente per $\theta \Leftrightarrow \varphi_{X_1, \dots, X_n}(x_1, \dots, x_n | T_n = t_0) = \frac{h(x_1, \dots, x_n, T_n = t_0; \theta)}{g(T_n = t_0; \theta)}$ non dipende da θ

TEOREMA DI FATTORIZZAZIONE

$\underline{X} = (X_1, \dots, X_n)$ campione casuale generato da $X \sim f(x; \theta), \theta \in \Omega(\theta)$ incogniti \Rightarrow

T_n è sufficiente per θ

$\Leftrightarrow \left\{ \begin{array}{l} \text{Esistono } g(*), f(*) > 0: \text{ la funzione di verosimiglianza possa fattorizzarsi come} \\ \mathcal{L}(\theta; \underline{x}) = g(T(x_1, \dots, x_n); \theta) * h(x_1, \dots, x_n) \\ \text{con } g(*) \text{ che dipende dalle osservazioni campionarie solo attraverso la sintesi } T(x_1, \dots, x_n) \\ \text{con } h(*) \text{ funzione del campione che non dipende da } \theta \end{array} \right.$

STATISTICHE SEMPRE SUFFICIENTI

La funzione di verosimiglianza, il campione casuale, v.c. ordinata, la funzione score V'_n .

TEOREMA STATISTICA SUFFICIENTE E INFORMAZIONE DI FISHER

L'informazione di Fisher fornita da uno stimatore sufficiente T_n coincide con quella fornita dall'intero campione.

Se T_n è sufficiente per θ lo sarà anche qualsiasi funzione biunivoca di T_n .

SUFFICIENZA MINIMALE

Uno stimatore T_n è sufficiente minimale per θ , se per qualsiasi altro stimatore sufficiente T_n^* , la statistica T_n è una funzione di T_n^* .

TEOREMA DI LEHAMAN E SCHEFFE

Siano (X_1, \dots, X_n) e (Y_1, \dots, Y_n) 2 campioni casuali generati da $X \sim f(X; \theta)$. Si individui una partizione dello spazio campionario C_n tale che i 2 campioni appartengano ad essa se e solo se $\mathcal{L}(\theta; \underline{x})/\mathcal{L}(\theta; \underline{y})$ non dipenda da θ . Allora, ogni stimatore corrispondente a questa partizione è sufficiente minimale.

Uno stimatore induce una partizione sullo spazio dei campioni: due campioni appartengono allo stesso sottoinsieme solo se il risultato delle statistiche è lo stesso. Ad esempio se lo stimatore è la media allora due campioni apparterranno allo stesso sottoinsieme della partizione se la media è la stessa.

STATISTICA ANCILLARE

T_n è ancillare se la distribuzione di probabilità di T_n non dipende da θ (statistica non contiene informazioni su θ).

STIMATORE COMPLETO

$X \sim f(X; \theta)$, T_n sufficiente per θ . T_n è completo se $\forall q(T_n): E(q(T_n)) = 0, \forall \theta \Rightarrow q(T_n) = 0$

Uno stimatore T_n è completo se l'unica funzione di T_n il cui valor medio è 0 è la funzione identicamente nulla.

TEOREMA COMPLETEZZA \rightarrow Una statistica sufficiente completa è sempre minimale

STIMATORE CORRETTO

Uno stimatore è corretto (o non distorto) se $E(T_n) = \theta$, la distorsione è $b(T_n) = E(T_n) - \theta$

T_n non distorto per $\theta \Rightarrow \psi(T_n)$ non distorto per $\psi(\theta)$ solo quando ψ è funzione lineare.

Si può essere non distorti in mediana \rightarrow $\text{Mediana}(T_n) = \theta$

ERRORE QUADRATICO MEDIO

$$MSE(T_n) = E[(T_n - \theta)^2] = E[(T_n - ET_n + (ET_n - \theta))^2] = E[(T_n - ET_n)^2] + (ET_n - \theta)^2 = \text{Var}(T_n) + (b(T_n))^2$$

EFFICIENZA RELATIVA

T_{1n} è più efficiente di T_{2n} per θ se $MSE(T_{1n}) < MSE(T_{2n})$

STIMATORE UMVUE

UMVUE = Uniformly Minimum Variance Unbiased Estimator = Stimatore non distorto a varianza minima

DISUGUAGLIANZA DI CRAMER RAO

(X_1, \dots, X_n) campione casuale generato da $X \sim f(X; \theta)$, allora per ogni stimatore T_n non distorto per θ si ha:

$$\text{Var}(T_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}, \quad \text{con } I_n(\theta) = \text{informazione di Fisher}$$

TEOREMA \rightarrow Se esiste stimatore non distorto per θ che raggiunge il limite Cramer Rao \rightarrow esso è unico

STIMATORE EFFICIENTE (EFFICIENTE ASSOLUTO)

Uno stimatore T_n non distorto si dice efficiente per un parametro θ di una v.c. $X \sim f(x; \theta)$, che soddisfa le usuali condizioni di regolarità, se e solo se $\text{Var}(T_n) = [I_n(\theta)]^{-1}$. Quindi, se esiste ed è non distorto, uno stimatore efficiente è quello in cui la varianza raggiunge il limite inferiore della disuguaglianza di Cramer Rao.

$$\text{EFF}(T_n) = \frac{1/\text{Var}(T_n)}{1/(I_n(\theta))} = [\text{Var}(T_n) * I_n(\theta)]^{-1}, \quad \text{stimatore tanto più efficiente quanto più } \text{EFF}(T_n) \rightarrow 1$$

TEOREMA STIMATORE EFFICIENTE E NON DISTORTO

Condizione necessaria e sufficiente affinché esista uno stimatore T_n efficiente e non distorto per θ è che sia:

$$V'_n = \frac{\partial}{\partial \theta} \log \mathcal{L}(\theta; \underline{x}) = I_n(\theta)(T_n - \theta)$$

TEOREMA DI RAO E BLACKWELL

(X_1, \dots, X_n) campione casuale estratto da $X \sim f(x; \theta)$ e sia T_{1n} uno stimatore sufficiente per θ mentre T_{2n} è un qualsiasi stimatore non distorto di θ . Allora posto $T_n = E(T_{2n}|T_{1n})$ si ha:

- 1) T_n è funzione esclusiva di T_{1n}
- 2) $E(T_n) = \theta$
- 3) $\text{Var}(T_n) \leq \text{Var}(T_{2n})$

Posso costruire uno stimatore più efficiente di uno stimatore non distorto utilizzando uno stimatore sufficiente. Uno stimatore non distorto di θ con varianza minima deve essere funzione di una statistica sufficiente.

TEOREMA DI LEHMANN-SCHEFFE (2)

Se \exists stimatore UMVUE per θ e $E(\psi(T_n)) = \theta$, con T_n è stimatore completo sufficiente minimale, allora $\psi(T_n)$ è UMVUE. Una funzione di uno stimatore sufficiente completo minimale che non è distorto per θ è UMVUE per θ .

STIMATORE ASINTOTICAMENTE EFFICIENTE

$$\lim_{n \rightarrow +\infty} E(T_n) = \theta \Leftrightarrow \lim_{n \rightarrow +\infty} b(T_n) = 0, \quad \text{ad esempio la varianza campionaria}$$

CONSISTENTE IN MEDIA QUADRATICA

$$\lim_{n \rightarrow +\infty} MSE(T_n) = \lim_{n \rightarrow +\infty} E(T_n - \theta)^2 = 0 \Leftrightarrow \lim_{n \rightarrow +\infty} Var(T_n) = 0, \quad \lim_{n \rightarrow +\infty} [b(T_n)]^2 = 0$$

Se uno stimatore è non distorto e la sua varianza tende a zero \rightarrow è consistente in media quadratica

Se uno stimatore è consistente in media quadratica \rightarrow è asintoticamente non distorto

CONSISTENTE IN PROBABILITA'

$$T_n \xrightarrow{p} \theta \text{ (plim } T_n = \theta) \Leftrightarrow \lim_{n \rightarrow +\infty} \Pr(|T_n - \theta| < \epsilon) = 1 \quad \forall \epsilon$$

La consistente in media quadratica implica la consistenza in probabilità

CONSISTENTE QUASI CERTAMENTE

$$T_n \xrightarrow{qc} \theta, \text{ consistente quasi certamente} \Leftrightarrow \Pr\left(\lim_{n \rightarrow +\infty} T_n = \theta\right) = 1 \Leftrightarrow \lim_{n \rightarrow +\infty} \Pr(|T_n - \theta| < \epsilon, \forall m \geq n) = 1, \forall \epsilon$$

EFFICIENZA ASINTOTICA

$$\lim_{n \rightarrow +\infty} Var(T_n) = 1/I_n(\theta) \Leftrightarrow \lim_{n \rightarrow +\infty} EFF(T_n) = 1$$

ASINTOTICAMENTE NORMALE

$$\lim_{n \rightarrow +\infty} \Pr\left(\frac{(T_n - E(T_n))}{\sqrt{Var(T_n)}} \leq t\right) = \Phi(t),$$

al crescere numerosità campionaria la funzione di ripartizione stimatore standardizzato tende ad una normale

BAN-CANE

T_n per θ è BAN (Best Asymptotical Normal) oppure CANE (Consistent Asymptotically Normal Efficient) se è asintoticamente normale, consistente in probabilità e – nella classe di tutti gli stimatori asintoticamente normali e consistenti – possiede la varianza più piccola.

3 STIMATORI DI MASSIMA VEROSIMIGLIANZA

$$\begin{aligned} X \sim f(x; \theta) &\rightarrow f(\underline{x}; \theta) = f(x_1; \theta) * \dots * f(x_n; \theta), & \text{prima di estrarre il campione è probabilità congiunta} \\ X \sim f(x; \theta) &\rightarrow \mathcal{L}(\theta; \underline{x}) = f(x_1; \theta) * \dots * f(x_n; \theta), & \text{estratto il campione è funzione di verosimiglianza} \end{aligned}$$

PRINCIPIO DI VEROSIMIGLIANZA FORTE

(X_1, \dots, X_n) campione estratto da $X \sim f(x; \theta)$ con $L_f(\theta; \underline{x})$; (Y_1, \dots, Y_n) campione estratto da $Y \sim g(y; \theta)$ con $L_g(\theta; \underline{y})$
Se $L_f \propto L_g$ le conclusioni inferenziali su θ devono essere le stesse

CONDIZIONI DI REGOLARITA'

Servono per consentire di scambiare derivata ed integrale ed affinché il campo di variazione della v.c. non dipenda dal parametro θ (esistenza derivate fino al terzo ordine, derivate tutte limitate in modulo, valore atteso verosimiglianza al quadrato compresa tra 0 e infinito).

LOG-VEROSIMIGLIANZA

$$V_n = \log \mathcal{L}(\theta; \underline{x}) = \log(f(x_1; \theta) * \dots * f(x_n; \theta)) = \sum_{i=1}^n \log f(x_i; \theta)$$

FUNZIONE SCORE

$$V'_n = \frac{\partial \log \mathcal{L}(\theta; \underline{x})}{\partial \theta} = \frac{\mathcal{L}'(\theta; \underline{x})}{\mathcal{L}(\theta; \underline{x})}, \quad E(V'_n) = 0$$

INFORMAZIONE DI FISHER

$$\begin{aligned} \text{Var}(V'_n) &= E[(V'_n)^2] = -E(V''_n) = I_n(\theta) \\ \text{se campione iid} &\rightarrow I_n(\theta) = nI(\theta) = nE\left[\left(\frac{\partial \log f(x; \theta)}{\partial \theta}\right)^2\right] = -nE\left[\frac{\partial^2 \log f(x; \theta)}{\partial \theta^2}\right] \end{aligned}$$

METODO MASSIMA VEROSIMIGLIANZA

$$\text{Stimatore di massima verosimiglianza } (\theta_{ML}) \theta: V'_n = \frac{\partial \log \mathcal{L}(\theta; \underline{x})}{\partial \theta} = 0, \quad V''_n = \frac{\partial^2 \log \mathcal{L}(\theta; \underline{x})}{\partial \theta^2} \leq 0$$

PROPRIETA' MASSIMA VEROSIMIGLIANZA

Se stimatore ML T_n per il parametro θ esiste unico, allora T_n è funzione della statistica sufficiente minimale per θ .

PROPRIETA' INVARIANZA

Se T_n è lo stimatore ML per il parametro θ e se $\psi(\theta)$ è una funzione biunivoca di θ allora $\psi(T_n)$ è lo stimatore ML per $\psi(\theta)$

TEOREMA DI CRAMER

Sotto le usuali condizioni di regolarità, lo stimatore ML $T_n \xrightarrow{qc} \theta$, è asintoticamente normale ed asint. efficiente:

$$T_n \xrightarrow{d} N\left(\theta, (I_n(\theta))^{-1}\right), \quad \text{con } I_n(\theta) \text{ informazione Fisher, Stimatore ML è BAN}$$

PROPRIETA' INVARIANZA DISTRIBUZIONE

$\psi(\theta) \rightarrow$ per invarianza, se T_n stimatore ML per $\theta \rightarrow \psi(T_n)$ stimatore ML di $\psi(\theta)$

per teorema di Cramer $\rightarrow \psi(T_n) \xrightarrow{d} N(\psi(\theta), (H' I_n(\theta) H)^{-1})$, con H matrice derivate seconde di $\psi(\theta)$

TEOREMA

Se esiste uno stimatore T_n non distorto ed efficiente per θ e se T_n^ è la soluzione ML di $V_n'(\theta) = 0 \rightarrow T_n = T_n^*$*

Lo stimatore ML possiede, quindi, una duplice validità. Se esiste uno stimatore efficiente per θ , lo stimatore ML coincide con esso, ed è quindi efficiente per θ per ogni n finito. D'altra parte, anche se non esiste uno stimatore efficiente per θ , lo stimatore ML è comunque asintoticamente efficiente. Nel caso di v.c. appartenenti alla famiglia esponenziale, il metodo ML assicura l'esistenza di stimatore efficienti.

4 METODI DI COSTRUZIONE DEGLI STIMATORI

METODO DEI MOMENTI

Affinché sia valido: 1) Esistenza dei momenti della v.c. in numero uguale a quello dei parametri da stimare; 2) Conoscenza relazione tra momenti e parametri che caratterizzano la popolazione (non richiede esplicitamente la conoscenza della distribuzione di probabilità, ma nella pratica per costruire la relazione serve conoscerla).

$X \sim f(X; \theta)$, vettore di parametri θ di dimensione m , momento m – esimo assoluto $E(|X|^m) < +\infty$

$$\text{Uguaglianza momenti teorici e campionari} \rightarrow \mu_r(\theta) = M_{r,n} \Leftrightarrow E(X^r) = \frac{1}{n} \sum_{i=1}^n (X_i)^r, \quad \text{con } r = 1, 2, \dots, m$$

PROPRIETA' STIMATORI MOMENTI

- Consistenti in probabilità (stimatore tende al vero valore del parametro);
- Asintoticamente non distorti
- Asintoticamente normali

MOMENTI GENERALIZZATI

Conosco un numero (pari al numero di parametri) di funzioni delle osservazioni per cui il valore atteso è 0

$$m(\theta) = E[g(X; \theta)] = 0 \rightarrow \text{Sostituisco momenti campionari} \rightarrow \hat{m}(\theta) = \frac{1}{N} \sum_{i=1}^N g(x_i; \theta) = 0$$

METODO MINIMI QUADRATI

Scrivo ogni componente campione $X_i = g(\theta) + \varepsilon_i$, con ε_i errore IID con $E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2, \varepsilon_i$ incorrelati

$$G(\theta) = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (x_i - g(\theta))^2 = \min!, \quad \text{e lineari sono BLUE, stimatori consistenti e asintoticamente normali}$$

METODI MINIMA DISTANZA

$$\text{Chi}(\theta) = \sum_{i=1}^{k(=\text{classi})} \frac{(\hat{f}_i - f_i(\theta))^2}{f_i(\theta)} = \min! \rightarrow \theta \text{ che minimizza espressione}$$

Cerco quei parametri che riducono al minimo possibile la discrepanza al quadrato tra distribuzione osservata nel campione e quella teoricamente attesa dalla distribuzione prefissata da X.

5 FAMIGLIA ESPONENZIALE

DEFINIZIONE

X appartiene alla famiglia esponenziale di $\dim(1)$ se è possibile scrivere funzione di densità:
 $f(x; \theta) = \exp\{Q(\theta)A(x) + C(x) - K(\theta)\} = M(x) \exp\{Q(\theta)A(x) - K(\theta)\}$, con $M(x) = \exp\{C(x)\}$

PROPRIETA'

Se $X_i, i = 1, \dots, n$ è una collezione di variabili indipendenti tutte appartenenti alla famiglia esponenziale
→ la distribuzione congiunta (pari alla produttoria della $f(x_i; \theta)$) appartiene alla famiglia esponenziale

RISCRITTURA

Se $A(x) = x$ si definisce famiglia esponenziale regolare → $f(x; \eta) = \exp\{\eta x + C(x) - K(\eta)\}$

$$\text{Si ha che } E(X) = \frac{\partial}{\partial \eta} K(\eta), \quad \text{Var}(X) = \frac{\partial^2}{\partial^2 \eta} K(\eta)$$

ESEMPIO

$$\text{Po}(\theta) \rightarrow Q(\theta) = \log(\theta), A(x) = x, C(x) = -\log(x!), K(\theta) = \theta \rightarrow \log(\theta) = \eta \quad [Q(\theta) = \eta] \rightarrow \theta = e^\eta \rightarrow \\ K(\eta) = e^\eta \rightarrow E(X) = K'(\eta) = e^\eta = \theta \rightarrow \text{Var}(X) = K''(\eta) = e^\eta = \theta$$

PROPRIETA' GENERALI

La famiglia esponenziale è l'unica ad ammettere stimatori sufficienti di \dim pari a quello dello spazio parametrico.

Ammettendo stimatori sufficienti e completi ammette stimatori sufficienti minimali.

Esistenza stimatore corretto ed efficiente per θ è cond. necessaria e sufficiente perchè $X \in \text{fam. esponenziale}$

MASSIMA VEROSIMIGLIANZA

$$\text{Lo stimatore di massima verosimiglianza si trova risolvendo } \rightarrow \sum_{i=1}^n A(x_i) = \sum_{i=1}^n E[A(x_i)], j = 1, \dots, n$$

ESEMPIO MASSIMA VEROSIMIGLIANZA ESPONENZIALE

$$X \sim \text{En}(\theta) \rightarrow f(x) = \theta e^{-\theta x}, x > 0 \rightarrow Q(\theta) = -\theta, A(x) = x, C(x) = 0, K(\theta) = -\log(\theta)$$

$$ML \rightarrow \sum_{i=1}^n A(x_i) = \sum_{i=1}^n E[A(x_i)] \Leftrightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n E[x_i] \Leftrightarrow n\bar{X}_n = \frac{n}{\theta} \Leftrightarrow \theta_{ML} = \frac{1}{\bar{X}_n}$$

6 TEST DELLE IPOTESI STATISTICHE

$\Omega(\theta)$ = spazio parametrico della v. c $X \sim f(x; \theta)$ = insieme dei valori compatibile per il vettore di parametri θ

$\omega_0 \subset \Omega(\theta)$ insieme dei valori specificati dall'ipotesi nulla

Il test verifica quale tra $H_0: \theta \in \omega_0$ e $H_1: \theta \notin \omega_0$ non è contraddetta dai risultati campionari (x_1, \dots, x_n) . Esisteranno valori del campione $(x_1, \dots, x_n) \in R_0 \subset \mathbb{R}^n$ per i quali la regola impone di rifiutare H_0 , e altri valori $(x_1, \dots, x_n) \notin R_0$ per i quali la regola impone di non rifiutare H_0 . La regione R_0 è detta regione di rifiuto o regione critica (RC) per H_0 , la regione complementare a R_0 , è detta regione di accettazione per H_0 .

Riduco $(X_1, \dots, X_n) \in \mathbb{R}^n$ tramite la statistica $T(X_1, \dots, X_n) \in \mathbb{R}^n$, ove $m < n$:

$T_n \in C_0 \rightarrow$ rifiuto H_0 , $T_n \notin C_0 \rightarrow$ non rifiuto H_0

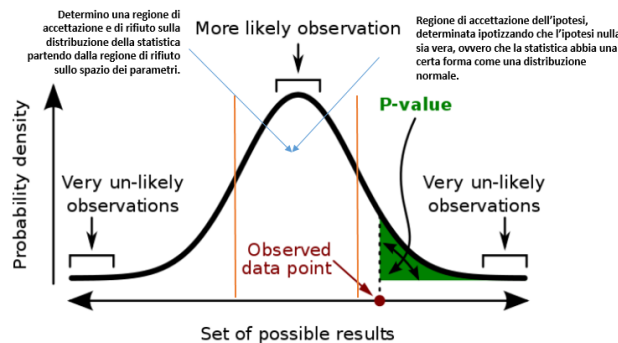
Il test delle ipotesi è una corrispondenza tra lo spazio campionario \mathbb{R}^n e lo spazio parametrico $\Omega(\theta)$, a cui si affianca una valutazione probabilistica.

ERRORI DI PRIMA E SECONDA SPECIE

$$\left\{ \begin{array}{l} \alpha = \text{PROB. ERRORE 1 SPECIE} = \text{LIV. SIGNIFICATIVITA'} = \Pr(E_1) = \Pr(\text{rifiutare } H_0 | H_0 \text{ è vera}) = \Pr(\underline{X} \in C_0 | H_0: \theta \in \omega_0) \\ \beta = \text{PROB. ERRORE 2 SPECIE} = \Pr(E_2) = \Pr(\text{non rifiutare } H_0 | H_0 \text{ è falsa}) = \Pr(\underline{X} \notin C_0 | H_1: \theta \notin \omega_0) \\ 1 - \alpha = \Pr(G_1) = \Pr(\text{non rifiutare } H_0 | H_0 \text{ è vera}) = \Pr(\underline{X} \notin C_0 | H_0: \theta \in \omega_0) \\ \gamma = 1 - \beta = \text{POTENZA DEL TEST} = \Pr(G_2) = \Pr(\text{rifiutare } H_0 | H_0 \text{ è falsa}) = \Pr(\underline{X} \in C_0 | H_1: \theta \notin \omega_0) \end{array} \right.$$

REGIONE CRITICA OTTIMALE

Si definisce "ottimale" per un prefissato α , una RC che, fra quelle di ampiezza α , possiede la potenza più elevata.



P-VALUE

Se il livello di significatività sotto l'ipotesi nulla è la probabilità di cadere nella zona di rifiuto quando l'ipotesi nulla è vera e può essere fissato a priori dallo statistico, il p-value è il livello di significatività osservato del test per il quale si rifiuterebbe l'ipotesi nulla. Ovvero il p value mi dice, con i dati campionari osservati, quale è la probabilità, supposta vera l'ipotesi H_0 , di ottenere un risultato (dai dati osservati) uguale o "più estremo" di quello effettivamente osservato. Se il p-value è più piccolo del livello di significatività l'ipotesi nulla viene rifiutata.

LEMMA DI NEYMAN E PEARSON

Sia $\underline{X} = (X_1, \dots, X_n)$ un campione casuale generato da $X \sim f(x; \theta)$ e si voglia verificare $H_0: \theta = \theta_0$ contro $H_1: \theta = \theta_1$. Se $\mathcal{L}(\theta; \underline{X})$ è la funzione di verosimiglianza di \underline{X} , allora la Rco(α) per H_0 contro H_1 è quella regione C_0 dello spazio

campionario di \mathbb{R}^n tale che soddisfa:
$$\begin{cases} \text{i) } \frac{\mathcal{L}(\theta_1; \underline{X})}{\mathcal{L}(\theta_0; \underline{X})} \geq c \\ \text{ii) } \Pr(\underline{X} \in C_0 | H_0) = \alpha \end{cases}$$

Il valore numerico della costante c viene determinato dall'ampiezza α della RC. Quindi, il vincolo i) determina la forma della RC e il vincolo ii) specifica la sua ampiezza. Conseguenza del Lemma di Neyman e Pearson: la Rco(α) è funzione dello stimatore sufficiente minimale T_n per θ (quando esso esiste). Anzi la Rco individuata è funzione del campione solo attraverso lo stimatore sufficiente minimale T_n per θ .

FUNZIONE POTENZA

$$\gamma(\theta) = \text{FUNZIONE POTENZA} = \Pr(\underline{X} \in C_0 | \theta) \text{ per ogni } \theta \in \Omega(\theta)$$

TEST UNIFORMEMENTE PIU' POTENTE

C_0 RC di ampiezza α . Test uniformemente più potente di ampiezza α , UMP_α (=Uniformly Most Powerful), se $\gamma(\theta) \geq \gamma^*(\theta)$ per tutti i $\theta \notin \omega_0$, essendo $\gamma^*(\theta)$ la funzione potenza di qualsiasi altro test di uguale ampiezza α .

Supponiamo che per verificare $H_0: \theta = \theta_0$ contro l'ipotesi semplice $H_1: \theta = \theta_1$ esista una Rco(α). Se essa non si modifica al variare di H_0 contro $H_1: \theta = \theta_1$ per ciascuno dei possibili valori di $\theta_1 \notin \omega_0$, allora tale RC è ottimale in ciascuno di questi test: quindi, essa è una regione UMP_α per verificare una ipotesi semplice $H_0: \theta = \theta_0$ contro una ipotesi composta $H_1: \theta_1 \notin \omega_0$.

Per test unidirezionali (maggiore o minore e non diverso che è bidirezionale) per le v.c. appartenenti alla famiglia esponenziale con $Q(\theta)$ monotona la UMP_α esiste sempre.

TEST CONSISTENTE

Un test si dice consistente se $\lim_{n \rightarrow \infty} \gamma_n(\theta) = 1$, per tutti i $\theta \notin \omega_0$. Test Lemma di Neyman e Pearson consistenti.

All'aumentare della numerosità campionaria è più probabile rifiutare correttamente l'ipotesi nulla.

TEST NON DISTORTO

Un test di ampiezza α si definisce non distorto se $\gamma(\theta) \geq \alpha$ per ogni $\theta \notin \omega_0$. La non distorsione garantisce che la probabilità di rifiutare correttamente H_0 sia almeno non inferiore alla probabilità di rifiutarla a torto. Un test UMP è necessariamente non distorto, tuttavia poiché può non esistere un test UMP ha senso ricercare un test UMP nella sottoclasse dei test non distorti, cioè un test UMPU (=Uniformly Most Powerful Unbiased).

TEST DEL RAPPORTO DI VEROSIMIGLIANZA LRT

Sia $\underline{X} = (X_1, \dots, X_n)$ un campione casuale generato da $X \sim f(x; \theta)$ ove $\theta \in \Omega(\theta)$ e si voglia verificare $H_0: \theta \in \omega_0$ contro $H_1: \theta \notin \omega_0$. Definiamo il rapporto di verosimiglianza $\lambda = \lambda(\underline{X}) = \frac{\max_{\theta \in \omega_0} \mathcal{L}(\theta; \underline{X})}{\max_{\theta \in \Omega(\theta)} \mathcal{L}(\theta; \underline{X})}$ e costruiamo una RC di ampiezza α in modo tale che sia $\Pr(\lambda(\underline{X}) \leq c_\alpha | H_0) = \alpha$. La RC $C_\alpha = \{ \underline{X}: \lambda(\underline{X}) \leq c_\alpha \}$ è una RC di ampiezza α costruita con il metodo LRT. Con una formula: $\Pr\left(\frac{\max_{\theta \in \omega_0} \mathcal{L}(\theta; \underline{X})}{\max_{\theta \in \Omega(\theta)} \mathcal{L}(\theta; \underline{X})} \leq c_\alpha \middle| H_0\right) = \alpha$

In pratica per ottenere la RC mediante il LRT, si calcola il rapporto tra la verosimiglianza massimizzata sotto l'ipotesi H_0 , (cioè quando $\theta \in \omega_0$) e la verosimiglianza massimizzata senza alcun vincolo (cioè quando $\theta \in \Omega(\theta)$). Questo rapporto (delle funzioni di verosimiglianza e non dei parametri) è funzione di \underline{X} e la disuguaglianza $\lambda(\underline{X}) \leq c$ rappresenta un evento la cui probabilità può essere determinata conoscendo la distribuzione della v.c. $\lambda(\underline{X})$, oppure quella di una sua trasformazione biunivoca.

PROPRIETA' LRT

$$\left\{ \begin{array}{l} 0 \leq \lambda(\underline{X}) \leq 1 \\ \text{La RC derivata dal LRT è funzione dello stimatore ML e dello stimatore sufficiente minimale per } T_n \\ \text{Se è applicabile il Lemma di Neyman e Pearson il LRT produce } R_{co}(\alpha) \text{ coincidenti con quelle del lemma} \\ \text{LRT asintoticamente UMPU} \end{array} \right.$$

TEOREMA DI WILKS

Sotto opportune condizioni di regolarità, se è vera H_0 , allora: $-2 \log[\lambda(\underline{X})] \xrightarrow{d} \chi_g^2$ ove g è pari alla differenza tra la dimensione dello spazio parametrico sotto $H_0 \cup H_1$ (esaurisce lo spazio parametrico) e la dimensione dello spazio parametrico sotto H_0 .

In generale, passando da $H_0 \cup H_1$ ad H_0 la dimensione dello spazio parametrico viene ridotta ponendo dei vincoli sui parametri. Ne consegue che g è pari al numero dei vincoli "indipendenti" specificati da H_0 .

Le condizioni di regolarità per il teorema coincidono, in buona sostanza, con quelle imposte per la validità degli stimatori ML.

Utile teorema perché se non si riesce ad esplicitare la distribuzione esatta di $\lambda(\underline{X})$ posso usare l'approssimazione.

SCORE TEST-TEST MULTIPLICATORI DI LAGRANGE

Date le usuali condizioni di regolarità e che la matrice di Fisher sia definita positiva, sia $\underline{X} = (X_1, \dots, X_n)$ un campione casuale generato da $X \sim f(x; \theta)$ ove $\theta \in \Omega(\theta)$ e si voglia verificare $H_0: \theta = \theta_0$ contro $H_1: \theta \neq \theta_1$. Se indichiamo con $V'_n(\theta)$ il vettore degli score rispetto agli m parametri, cioè:

$$V'_n(\theta) = \left(\frac{\partial \log \mathcal{L}(\theta; \underline{X})}{\partial \theta_1}, \dots, \frac{\partial \log \mathcal{L}(\theta; \underline{X})}{\partial \theta_m} \right)^T$$

allora sotto H_0 è possibile dimostrare che, asintoticamente, la forma quadratica:

$$S_n(\underline{X}) = (V'_n(\theta_0))^T [I_n(\theta_0)]^{-1} V'_n(\theta_0) \xrightarrow{d} \chi_m^2$$

e la RC(α) è della forma $S_n(\underline{X}) \geq c_\alpha$, dove $\Pr(S_n(\underline{X}) \geq c_\alpha | H_0) = \alpha$.

TEST DI WALD

Date le usuali condizioni di regolarità e che la matrice di Fisher sia definita positiva:

$$W_n(\underline{X}) = (\mathbf{T}_n - \theta_0)^T I_n(\mathbf{T}_n) (\mathbf{T}_n - \theta_0) \xrightarrow{d} \chi_m^2$$

E la RC(α) è della forma $W_n(\underline{X}) \geq c_\alpha$, dove $\Pr(W_n(\underline{X}) \geq c_\alpha | H_0) = \alpha$

TEST DI WALD MODIFICATO

$$W_n^0(\underline{X}) = (\mathbf{T}_n - \theta_0)^T I_n(\theta_0) (\mathbf{T}_n - \theta_0) \xrightarrow{d} \chi_m^2$$

CONFRONTO TEST LRT, SCORE TEST, TEST DI WALD

Qualora l'ipotesi $H_0: \theta = \theta_0$ e riguardi un solo parametro ($m = 1$) indicando con $V_n(\theta_0)$ e $V_n(\mathbf{T}_n)$ la log-verosimiglianza calcolata per $\theta = \theta_0$ e per $\theta = \mathbf{T}_n$ (=stimatore ML), rispettivamente, allora la RC $_\alpha$ dei tre test di natura asintotica introdotti si particolarizzano:

$$\left\{ \begin{array}{ll} \text{LRT,} & 2[V_n(\mathbf{T}_n) - V_n(\theta_0)] \geq \chi_{(\alpha,1)}^2 \\ \text{Score test,} & \frac{[V'_n(\theta_0)]^2}{I_n(\theta_0)} \geq \chi_{(\alpha,1)}^2 \\ \text{Test di Wald,} & (\mathbf{T}_n - \theta_0)^2 I_n(\mathbf{T}_n) \geq \chi_{(\alpha,1)}^2 \end{array} \right. \\ \text{con } \chi_{(\alpha,1)}^2 = \text{quantili } (1 - \alpha) \text{ della v. c Chi - quadrato con 1 grado di libertà}$$

7 TEST PARAMETRICI E NON PARAMETRICI

TEST NORMALITA'

\$(X_1, \dots, X_n)\$ proveniente da \$X \sim N(\mu, \sigma^2)\$, \$Rco(\alpha) = \text{Regione di rifiuto ottimale}\$

$$\left\{ \begin{array}{l} \textbf{VALORE MEDIO NOTA LA VARIANZA } \sigma^2 \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu > \mu_0 \rightarrow Rco(\alpha): \bar{X}_n \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu < \mu_0 \rightarrow Rco(\alpha): \bar{X}_n \leq \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu \neq \mu_0 \rightarrow Rc(\alpha): \begin{cases} \bar{X}_n \geq \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ \bar{X}_n \leq \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \end{cases} \\ I \text{ quantili } z_\alpha \text{ e } z_{\alpha/2} \text{ sono relativi a } Z \sim N(0,1) \text{ e tali che:} \\ \Pr(Z \leq z_\alpha) = \Phi(z_\alpha) = 1 - \alpha, \Phi(z_{\alpha/2}) = 1 - \alpha/2 \end{array} \right\} \left\{ \begin{array}{l} \textbf{VALORE MEDIO NON NOTA LA VARIANZA } \sigma^2 \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu > \mu_0 \rightarrow Rco(\alpha): \bar{X}_n \geq \mu_0 + t_{(\alpha,g)} \frac{\sigma}{\sqrt{n}} \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu < \mu_0 \rightarrow Rco(\alpha): \bar{X}_n \leq \mu_0 - t_{(\alpha,g)} \frac{\sigma}{\sqrt{n}} \\ H_0: \mu = \mu_0 \text{ contro } H_1: \mu \neq \mu_0 \rightarrow Rc(\alpha): \begin{cases} \bar{X}_n \geq \mu_0 + t_{(\frac{\alpha}{2},g)} \frac{\sigma}{\sqrt{n}} \\ \bar{X}_n \leq \mu_0 - t_{(\frac{\alpha}{2},g)} \frac{\sigma}{\sqrt{n}} \end{cases} \\ I \text{ quantili } t_{(\alpha,g)} \text{ e } t_{(\frac{\alpha}{2},g)} \text{ sono relativi a } \mathcal{T}_g \text{ di student} \\ \text{con } g = n - 1 \end{array} \right.$$

TEST SULLA VARIANZA SE E' NON NOTO IL VALORE MEDIO \$\mu\$

$$\left\{ \begin{array}{l} H_0: \sigma^2 = \sigma_0^2 \text{ contro } H_1: \sigma^2 > \sigma_0^2 \rightarrow Rco(\alpha): S_n^2 \geq \frac{\sigma_0^2}{n-1} \chi_{(\alpha,g)}^2 \\ H_0: \sigma^2 = \sigma_0^2 \text{ contro } H_1: \sigma^2 < \sigma_0^2 \rightarrow Rco(\alpha): S_n^2 \leq \frac{\sigma_0^2}{n-1} \chi_{(1-\alpha,g)}^2 \\ H_0: \sigma^2 = \sigma_0^2 \text{ contro } H_1: \sigma^2 \neq \sigma_0^2 \rightarrow Rc(\alpha): \begin{cases} S_n^2 \geq \frac{\sigma_0^2}{n-1} \chi_{(\alpha/2,g)}^2 \\ S_n^2 \leq \frac{\sigma_0^2}{n-1} \chi_{(1-\alpha/2,g)}^2 \end{cases} \end{array} \right.$$

\$I\$ quantili \$\chi_{(\alpha,g)}^2\$ sono relativi alla v. c \$\chi_{(g)}^2\$ con \$g = n - 1\$ gradi di libertà.

TEST CONFRONTO NORMALITA'

$\underline{X}_n = (X_1, \dots, X_n), \underline{Y}_m = (Y_1, \dots, Y_m)$ proveniente da $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$

Se invece che $\mu_x - \mu_y = 0$ vale $\mu_x - \mu_y = c$ vale tutto ma sostituisco a statistica $-c \rightarrow es. Z_{n,m} - c$

TEST SULLA DIFFERENZA TRA I VALORI MEDI CON σ_x^2 e σ_y^2 NOTE

$$\begin{cases} H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y > 0 \rightarrow Rco(\alpha): Z_{n,m} \geq z_\alpha \\ H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y < 0 \rightarrow Rco(\alpha): Z_{n,m} \leq -z_\alpha \\ H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y \neq 0 \rightarrow Rc(\alpha): |Z_{n,m}| \geq z_{\alpha/2} \end{cases}$$

La statistica test $Z_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \rightarrow \text{se è vera } H_0 \rightarrow Z_{n,m} \sim N(0,1)$

TEST DI STUDENT – TEST SULLA DIFFERENZA TRA I VALORI MEDI CON $\sigma_x^2 = \sigma_y^2 = \sigma^2$ IGNORE

$$\begin{cases} H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y > 0 \rightarrow Rco(\alpha): T_{n,m} \geq t_{(\alpha,g)} \\ H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y < 0 \rightarrow Rco(\alpha): T_{n,m} \leq -t_{(\alpha,g)} \\ H_0: \mu_x - \mu_y = 0 \text{ contro } H_1: \mu_x - \mu_y \neq 0 \rightarrow Rc(\alpha): |T_{n,m}| \geq t_{(\alpha/2,g)} \end{cases}$$

$T_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_x^2(n-1) + S_y^2(m-1)}} \sqrt{\frac{nm(n+m-2)}{n+m}} \rightarrow \text{se è vera } H_0 \rightarrow T_{n,m} \sim T_g, \text{ con } g = n+m-2$

Se le numerosità campionarie sono alte, $T_{n,m}$ è sostituita da $Z_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_x^2/n + S_y^2/m}} \sim N(0,1)$

TEST DI OMOSCHEDASTICITA' – TEST SULL RAPPORTO TRA 2 VARIAZE, VALORI MEDI IGNOTI

$$\begin{cases} H_0: \sigma_x^2 = \sigma_y^2 \text{ contro } H_1: \sigma_x^2 > \sigma_y^2 \rightarrow Rco(\alpha): F_{n,m} \geq F_{(\alpha, g_1, g_2)} \\ H_0: \sigma_x^2 = \sigma_y^2 \text{ contro } H_1: \sigma_x^2 < \sigma_y^2 \rightarrow Rco(\alpha): F_{n,m} \leq F_{(1-\alpha, g_1, g_2)} \\ H_0: \sigma_x^2 = \sigma_y^2 \text{ contro } H_1: \sigma_x^2 \neq \sigma_y^2 \rightarrow Rc(\alpha): \begin{cases} F_{n,m} \geq F_{(\alpha/2, g_1, g_2)} \\ F_{n,m} \leq F_{(1-\alpha/2, g_1, g_2)} \end{cases} \end{cases}$$

La statistica test $F_{n,m} = \frac{S_x^2}{S_y^2} \rightarrow \text{se è vera } H_0 \rightarrow F_{(g_1, g_2)} \text{ v.c. Fischer con } g_1 = n-1, g_2 = m-1$

TEST DEI SEGNI

Viene usato per verificare una ipotesi sul valore della mediana ($\Pr(X \geq Me) = \Pr(X \leq Me) = 1/2$) di una popolazione oppure per controllare se due campioni provengono dalla stessa popolazione accertando che la mediana delle differenze $X - Y$ sia nulla.

$$H_0: Me = Me_0 \quad vs \quad H_1: Me \neq Me_0$$

Se è vera H_0 circa metà delle osservazioni dovrebbe essere superiore (inferiore) a Me_0 , perciò si rifiuterà H_0 in funzione del fatto che questo accada.

Per un campione (X_1, \dots, X_n) , il numero di osservazioni T_n superiori a Me_0 è una binomiale con parametri n e $\theta = \Pr(X < Me_0)$. Quindi l'ipotesi nulla $H_0: Me = Me_0$ può essere riformulata sulla variabile $T_n \sim \text{Bin}(n, \theta)$ come $H_0: \theta = 1/2$ contro l'ipotesi alternativa $H_1: \theta \neq 1/2$.

Se l'ipotesi nulla è vera $T_n \sim \text{Bin}(n, 1/2)$ per cui il campione in media conterrà $n/2$ osservazioni al di sopra di Me_0 . Pertanto si può definire la seguente $Rc(\alpha)$:

$$\left| T_n - \frac{n}{2} \right| \geq c_{\frac{\alpha}{2}}, \quad c_{\frac{\alpha}{2}}: \alpha = \Pr \left(\left| T_n - \frac{n}{2} \right| \geq c_{\frac{\alpha}{2}} \right) = 1 - \Pr \left(\frac{n}{2} - c_{\frac{\alpha}{2}} < T_n < \frac{n}{2} + c_{\frac{\alpha}{2}} \right) \approx \text{appr Normale} \approx c_{\alpha/2} \approx (z_{\alpha/2} \sqrt{n} - 1)/2$$

Allora se T_n è la statistica test definita come il numero di unità superiori alla mediana Me_0 , la $Rc(\alpha)$ è:

$$\begin{cases} T_n < \frac{n+1}{2} - \frac{z_{\alpha/2} \sqrt{n}}{2} \\ T_n > \frac{n-1}{2} + \frac{z_{\alpha/2} \sqrt{n}}{2} \end{cases}$$

Il test dei segni possiede un'efficienza relativa asintotica (=ARE) comparativamente accettabile ed elevata rispetto agli omologhi test derivanti dalle procedure classiche che presuppongono la validità dell'ipotesi della distribuzione normale.

TEST SULLE FREQUENZE-BERNOULLI

Campione casuale (X_1, \dots, X_n) è generato da una popolazione $X \sim \text{Ber}(\theta)$ ove θ è la probabilità dell'evento successo. Per verificare asintoticamente se è vera un'ipotesi del tipo $H_0: \theta = \theta_0$ si utilizza la seguente procedura:

$$\mathcal{F}_n = \text{frequenza relativa} = \frac{\# \text{successi}}{\# \text{prove}} \rightarrow \text{Se è vera } H_0: \theta = \theta_0 \rightarrow Z_n = \frac{\mathcal{F}_n - \theta_0}{\sqrt{\theta_0(1 - \theta_0)/n}} \xrightarrow{d} N(0,1)$$

Applicazione teorema centrale a $\mathcal{F}_n = \frac{1}{n} \sum_{i=1}^n I_{\text{successo}}$ con I_{successo} variabile indicatrice che vale 1 in caso di successo.

$$\begin{cases} H_0: \theta = \theta_0 \text{ contro } H_1: \theta > \theta_0 \rightarrow Rc(\alpha): \mathcal{F}_n \geq \theta_0 + z_{\alpha} \sqrt{\theta_0(1 - \theta_0)/n} \\ H_0: \theta = \theta_0 \text{ contro } H_1: \theta < \theta_0 \rightarrow Rc(\alpha): \mathcal{F}_n \leq \theta_0 - z_{\alpha} \sqrt{\theta_0(1 - \theta_0)/n} \\ H_0: \theta = \theta_0 \text{ contro } H_1: \theta \neq \theta_0 \rightarrow Rc(\alpha): |\mathcal{F}_n - \theta_0| \geq z_{\alpha/2} \sqrt{\theta_0(1 - \theta_0)/n} \end{cases}$$

TEST SULLE FREQUENZE- UGUAGLIANZA TRA PROPORZIONI

Verifico l'uguaglianza di due proporzioni (per esempio, in due circostanze differenti) mediante il confronto tra le stime ottenute da due campioni indipendenti. Se $\mathcal{F}_{n_1}, \mathcal{F}_{n_2}$ sono le frequenze relative dell'attributo su due campioni indipendenti, di numerosità n_1, n_2 generati da due v.c. $X \sim \text{Ber}(\theta_1), Y \sim \text{Ber}(\theta_2)$ risulta che:

$$Z_{n_1, n_1} = \frac{\mathcal{F}_{n_1} + \mathcal{F}_{n_2}}{\sqrt{\hat{\mathcal{F}}(1 - \hat{\mathcal{F}})[1/n_1 + 1/n_2]}} \xrightarrow{d} N(0,1), \quad \text{con } \hat{\mathcal{F}} = \frac{n_1 \mathcal{F}_{n_1} + n_2 \mathcal{F}_{n_2}}{n_1 + n_2}, \quad \mathcal{F}_{n_1} = \text{freq. relativa} = \frac{\# \text{successi}}{\# \text{prove}}$$

$$\begin{cases} H_0: \theta_1 - \theta_2 = 0 \text{ contro } H_1: \theta_1 - \theta_2 > 0 \rightarrow Rco(\alpha): Z_{n_1, n_1} \geq z_\alpha \\ H_0: \theta_1 - \theta_2 = 0 \text{ contro } H_1: \theta_1 - \theta_2 < 0 \rightarrow Rco(\alpha): Z_{n_1, n_1} \leq -z_\alpha \\ H_0: \theta_1 - \theta_2 = 0 \text{ contro } H_1: \theta_1 - \theta_2 \neq 0 \rightarrow Rc(\alpha): |Z_{n_1, n_1}| \geq z_{\alpha/2} \end{cases}$$

TEST DI CHI-QUADRATO

Fa parte dei test di bontà di adattamento (test introdotti per verificare l'ipotesi che i dati campionari provengano da una v.c. la cui distribuzione è nota). Il test misura la discrepanza tra l'ipotesi formulata in H_0 e quella che emerge dal campione mediante una opportuna metrica delle differenze standardizzate tra l'istogramma osservato e l'istogramma "teorico".

Verifico che i dati siano stati generati da una funzione di ripartizione $F(x; \theta) = F_0(x)$ completamente specificato nella forma e nei parametri, contro qualsiasi altra ipotesi alternativa H_1 .

(X_1, \dots, X_n) generato da ignota $X \sim F(x; \theta)$

$H_0: F(x; \theta) = F_0(x)$, per tutti gli x vs $H_1: F(x; \theta) \neq F_0(x)$, per qualche x

Organizzo i dati campionari in k categorie C_1, C_2, \dots, C_k a cui sono associate le frequenze assolute n_1, n_2, \dots, n_k

Calcolo $p_1, \dots, p_k: p_i = \Pr(X \in C_i | H_0) = \int_{C_i} dF_0(x), \quad i = 1, \dots, k$

$$X^2 = \text{Statistica Test Chi - Quadrato} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^k \frac{(\text{freq. oss.} - \text{freq. teo.})^2}{\text{freq. teo.}} = \sum_{i=1}^k \frac{(n_i)^2}{np_i} - n$$

$$X^2 \rightarrow \text{Se è vera } H_0 \rightarrow \text{asintoticamente} \rightarrow X^2 = \chi_{(k-1)}^2$$

$$Rc(\alpha): X^2 > \chi_{(\alpha, k-1)}^2$$

Se i parametri sono stimati dal campione (ipotizzo distribuzione normale e stimo media e varianza dal campione) devo togliere al numero di gradi di libertà $g = k - 1$ il numero di parametri che stimo (Es. normale $g = k - 1 - 2$).

ANALISI DELLA VARIANZA

ANOVA (=ANalysis Of VARIance) verifica l'uguaglianza tra più valori medi. Tale confronto viene svolto mediante stime diverse della variabilità il cui rapporto, se è vera l'ipotesi H_0 di nessuna differenza tra i valori medi, si distribuisce come una v.c. di Fisher, con opportuni gradi di libertà. ANOVA sostanzialmente è un'analisi delle medie in più di due gruppi mediante il confronto tra le devianze.

Somministro k trattamenti alle n_1, n_2, \dots, n_k unità statistiche con $n_1 + n_2 + \dots + n_k = n$. y_{ij} è la variabile risposta connessa alla unità j -esima che ha ricevuto l' i -esimo trattamento.

$$\begin{cases} \text{Trattamento } A_1 \rightarrow y_{11}, \dots, y_{1n_1} \\ \text{Trattamento } A_2 \rightarrow y_{21}, \dots, y_{2n_2} \\ \dots \\ \text{Trattamento } A_k \rightarrow y_{k1}, \dots, y_{kn_k} \end{cases}$$

Il modello probabilistico imposto ai risultati sperimentali è il seguente:

$$Y_{ij} = \mu_i + \epsilon_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \text{per } i = 1, \dots, k, \quad \text{per } j = 1, \dots, n_i, \quad \epsilon_{ij} \sim IDN(0, \sigma^2)$$

Il problema statistico è quello di verificare se i valori medi sono uguali tra i differenti trattamenti perché, allora, il trattamento non modificherebbe la risposta se non per effetto del caso. Ciò equivale a sottoporre a test:

$$\begin{cases} H_0: \mu_1 = \dots = \mu_k = \mu \Leftrightarrow H_0: \alpha_1 = \dots = \alpha_k = 0 \\ H_1: \text{non è vera } H_0 \end{cases}$$

$$\bar{Y}_i = \text{Media Campionaria Trattamento } i\text{-esimo} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad \bar{Y} = \text{Media Campionaria Generale} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$$

$$\begin{cases} D_T = \text{Devianza Totale} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{Y})^2 = D_B + D_W \\ D_B = \text{Devianza tra i trattamenti} = \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2, \quad D_W = \text{Devianza entro i trattamenti} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{Y}_i)^2 \end{cases}$$

Si dimostra che:

- D_B/σ^2 si distribuisce come una v. c. $\chi^2_{(k-1)}$;
- D_W/σ^2 si distribuisce come una v. c. $\chi^2_{(n-k)}$;
- D_B, D_W sono indipendenti

$$\text{Sotto } H_0 \rightarrow T_n = \frac{\frac{D_B}{\sigma^2(k-1)}}{\frac{D_W}{\sigma^2(n-k)}} = \frac{\frac{D_B}{(k-1)}}{\frac{D_W}{(n-k)}} = \frac{S_B^2}{S_W^2} = \frac{\frac{1}{(k-1)} \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2}{\frac{1}{(n-k)} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{Y}_i)^2} \sim F_{(k-1, n-k)}, \quad Rc(\alpha): T_n > F_{(\alpha, k-1, n-k)}$$

TEST DI KOLMOGOROV

Verifica l'adattamento di un campione casuale proveniente da una v.c. continua ed utilizza il massimo della differenza assoluta tra funzione di ripartizione empirica e funzione di ripartizione teorica specificata in H_0 .

$$H_0: F(x; \theta) = F_0(x), \text{ per tutti gli } x \quad \text{vs} \quad H_1: F(x; \theta) \neq F_0(x), \text{ per qualche } x$$

$$\left\{ \begin{array}{l} F_n(x) = \text{Funzione di Ripartizione Empirica} = \frac{\#\{X_i \leq x\}}{n} = \frac{i}{n} \\ \text{per } X_{(i)} \leq x \leq X_{(i+1)}, i = 0, \dots, n, \quad X_{(i)} = \text{Statistica ordinata}, X_{(0)} = -\infty, X_{(n+1)} = +\infty \\ D_n = \text{Statistica di Kolmogorov} = \sup |F_n(x) - F_0(x)|, \quad \text{per ogni } -\infty < x < +\infty \end{array} \right.$$

Se è vera H_0 , $F_n(x) \xrightarrow{d} F_0(x)$, e la differenza tra la funzione di ripartizione empirica e la funzione di ripartizione specificata sarà minima, se H_0 è falsa tale differenza sarà notevole. Perciò si rifiuta H_0 al crescere di $D_n > 0$ rispetto ad un livello critico determinato dal livello di significatività prescelto. Calcolare la distribuzione esatta della statistica test è piuttosto arduo. Per questo si ricorre ad una approssimazione per la Regione Critica:

$$Rc(\alpha): D_n \geq \frac{d_\alpha}{\sqrt{n}} \cong \sqrt{-\frac{1}{2n} \log\left(\frac{\alpha}{2}\right)}$$

TEST DI SMIRNOV - O KOLMOGOROV-SMIRNOV

Supponiamo di possedere due campioni casuali indipendenti $(X_1, \dots, X_n), (Y_1, \dots, Y_n)$. Desideriamo verificare se è vera l'ipotesi H_0 che entrambi provengano dalla stessa popolazione, cioè:

$$H_0: F_X(w) = G_Y(w), \text{ per tutti gli } w \quad \text{vs} \quad H_1: F_X(w) \neq G_Y(w), \text{ per qualche } w$$

$$D_n = \text{Statistica Test} = \sup |F_n(w) - G_m(w)|, \quad F_n(w) \text{ e } G_m(w) \text{ funzioni empiriche}$$

$$Rc(\alpha): D_n \geq d_{\alpha,n,m} \text{ su apposite tavole si approssima con } d_{\alpha,n,m} \cong \sqrt{-\frac{1}{2} \left(\frac{1}{n} + \frac{1}{m} \right) \log\left(\frac{\alpha}{2}\right)}$$

TEST NORMALITA'-SHAPIRO E WILKS

Uno dei test ritenuto più potente per la verifica della normalità, soprattutto in piccoli campioni. Verifica la Normalità confrontando due stimatori alternati della varianza: uno stimatore BLUE non parametrico basato sulla combinazione lineare ottimale delle statistiche ordinate di una v.c. Normale standardizzata, e l'usuale stimatore parametrico, cioè la varianza campionaria. La statistica-test è:

$$W_n = \frac{[\sum_{i=1}^n a_i x_{(i)}]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Le costanti a_i sono disponibili su apposite tavole e la RC per rifiutare H_0 è del tipo $W_n < w_\alpha$, per opportuni valori critici. Infatti, se è vera H_0 , W_n dovrebbe essere vicino ad 1, mentre l'inadeguatezza di tale ipotesi implica valori calcolati di W_n molto piccoli.

TEST CHI QUADRATO SULL' INDIPENDENZA

H_0 : X e Y sono indipendenti vs H_1 : X e Y non sono indipendenti

Ho una tabella a doppia entrata ottenuta dall'osservazione della variabile (X, Y) sulle n unità statistiche classificando le osservazioni secondo le k modalità x_i di X e le h modalità y_j di Y, pervenendo all'usuale tabella a doppia entrata delle frequenze assolute n_{ij} . Calcolo le frequenze assolute di X con $n_i = \sum_{j=1}^h n_{ij}$ e quelle di Y con $n_j = \sum_{i=1}^k n_{ij}$

Poi calcolo le frequenze assolute teoriche assumendo l'indipendenza, cioè: $\hat{n}_{ij} = \frac{n_i n_j}{n}$, $\forall i, j$.

$$X, Y \text{ indipendenti} \rightarrow p_{ik} = p_i * p_j \rightarrow \text{stimo con frequenze relative} \rightarrow \frac{n_{ij}}{n} = \frac{n_i}{n} * \frac{n_j}{n}$$

$$X^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \rightarrow \text{Se } H_0 \text{ è vera} \rightarrow \chi^2_{(g)} \text{ con } g = (k-1)(h-1)$$

$$Rc(\alpha): X^2 > \chi^2_{(\alpha, g)} \text{ con } g = (k-1)(h-1)$$

TEST DI CORRELAZIONE DEI RANGHI DI SPEARMAN

Verifica l'indipendenza di 2 popolazioni X e Y mediante un campione casuale $(X_1, Y_1), \dots, (X_n, Y_n)$. Ai valori X_i e Y_i si sostituiscono i rispettivi ranghi - separatamente calcolati per ciascuno dei 2 gruppi - indicati rispettivamente con R_i, S_i .

$$T_n = R_s \sqrt{n-1} \xrightarrow{d} Z \sim N(0,1), \quad \text{con } R_s = \text{Corr}(R_i, S_i) = \frac{\sum_{i=1}^n \left[R_i - \frac{n+1}{2} \right] \left[S_i - \frac{n+1}{2} \right]}{\sqrt{\sum_{i=1}^n \left[R_i - \frac{n+1}{2} \right]^2 \sum_{i=1}^n \left[S_i - \frac{n+1}{2} \right]^2}}$$

8 INTERVALLI DI CONFIDENZA

Dato un campione casuale $\underline{X} = (X_1, X_2, \dots, X_n)$ generato da $X \sim f(x; \theta)$ si definisce pivot:

$v(\underline{X}; \theta)$ la cui distribuzione non dipende da θ .

Se $v(\underline{X}; \theta)$ è invertibile \rightarrow Intervallo di confidenza $\rightarrow \begin{cases} 1 - \alpha = \Pr(v_L \leq v(\underline{X}; \theta) \leq v_U) \\ 1 - \alpha = \Pr(v^{-1}(v_L; \underline{X}) \leq \theta \leq v^{-1}(v_U; \underline{X})) \end{cases}$

ESEMPIO

$X \sim N(\theta, \sigma^2)$ con σ^2 ignota $\rightarrow v(\underline{X}; \theta) = \frac{\bar{X}_n - \theta}{\frac{S_n}{\sqrt{n}}} \sim t_{(n-1)} \rightarrow$ intervallo confidenza \rightarrow

$$-\bar{X}_n - t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \leq -\theta \leq -\bar{X}_n + t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \rightarrow \bar{X}_n - t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \leq \theta \leq \bar{X}_n + t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}$$

TEOREMA

Se esiste uno stimatore ML T_n per un parametro di posizione $\theta \rightarrow v.c. (T_n - \theta)$ è una pivot

Se esiste uno stimatore ML T_n per un parametro di scala $\rightarrow v.c. \frac{T_n}{\theta}$ è una pivot

INTERVALLI ASINTOTICI

Per teorema di Cramer \rightarrow Stimatore ML di T_n di θ è asintoticamente normale $\rightarrow \sqrt{I_n(\theta)}(T_n - \theta) \xrightarrow{d} N(0,1)$

$\left\{ \begin{array}{l} \text{Se } I_n(\theta) = I_n \rightarrow \text{non dipende dal parametro} \rightarrow \left(T_n - z_{\frac{\alpha}{2}} * (I_n)^{-\frac{1}{2}}, T_n + z_{\frac{\alpha}{2}} * (I_n)^{-\frac{1}{2}} \right) \\ \text{Se } I_n(\theta) \text{ dipende dal parametro} \rightarrow \text{sostituisco} \rightarrow I_n(\theta) = I_n(T_n) \rightarrow \left(T_n - z_{\frac{\alpha}{2}} * (I_n(T_n))^{-\frac{1}{2}}, T_n + z_{\frac{\alpha}{2}} * (I_n(T_n))^{-\frac{1}{2}} \right) \end{array} \right.$

1 REGRESSIONE LINEARE

MODELLO E STIMA

La regressione lineare rappresenta un metodo di stima del valore atteso condizionato di una variabile dipendente, o endogena, Y , dati i valori di altre variabili indipendenti, o esogene, X_1, \dots, X_k . La regressione è un modello statistico dove si assume un legame tra variabile endogena e variabili esogene di tipo lineare.

NOTAZIONE MATRICIALE $\rightarrow \underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, NOTAZIONE VETTORIALE $\rightarrow y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i$

TEOREMA DI GAUSS-MARKOV

In un modello lineare in cui i termini di errore (supposti indipendenti dalle variabili esogene) hanno valore atteso nullo, sono incorrelati e omoschedastici, gli stimatori lineari corretti più efficienti sono gli stimatori ottenuti con il metodo dei minimi quadrati.

$$\text{IPOTESI CLASSICHE GAUSS MARKOV} \left\{ \begin{array}{l} \text{A1) } E(\varepsilon_i) = 0, \quad i = 1, \dots, N \\ \text{A2) } \{\varepsilon_1, \dots, \varepsilon_N\} \text{ e } \{\underline{x}_1, \dots, \underline{x}_N\} \text{ indipendenti (o x deterministica)} \\ \text{A3) } V(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, N, \quad \text{omoschedasticità} \\ \text{A4) } \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i, j = 1, \dots, N (i \neq j), \quad \text{no autocorrelazione} \end{array} \right.$$

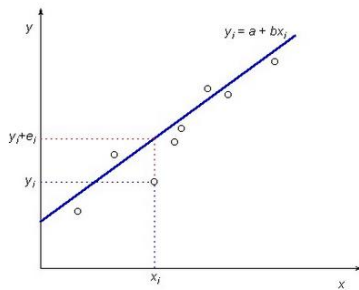
CON IPOTESI GAUSS MARKOV \underline{b} E' BLUE (Best Linear Unbiased Estimatori)

STIMATORE MINIMI QUADRATI (OLS=ORDINARY LEAST SQUARES)

$$\left\{ \begin{array}{l} \underline{b} = \text{STIMATORE MINIMI QUADRATI MATRICE} = (X'X)^{-1}X'y \\ \underline{b} = \text{VETTORE} = \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N \underline{x}_i y_i \\ \text{REGRESSIONE SEMPLICE} \left\{ \begin{array}{l} b_1 = \bar{y} - b_2 \bar{x} \\ b_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\sigma(x, y)}{\sigma^2(x)} \end{array} \right. \end{array} \right.$$

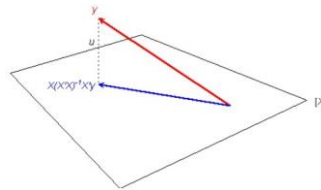
RESIDUI

$$\text{MODELLO STIMATO} \rightarrow \left\{ \begin{array}{l} \hat{y}_i = \underline{x}_i' \underline{b} \\ \hat{y} = \underline{X}\underline{b} \end{array} \right., \quad \text{RESIDUI} \rightarrow \left\{ \begin{array}{l} e_i = y_i - \hat{y}_i = y_i - \underline{x}_i' \underline{b} \\ \underline{e} = \underline{y} - \hat{\underline{y}} = \underline{y} - \underline{X}\underline{b} \end{array} \right.$$



Il metodo dei minimi quadrati, nel caso bivariato, deriva una retta che interpola uno scatter di punti minimizzando la somma dei quadrati delle distanze ε_i dei punti stessi dalla retta; il grafico fornisce un'intuizione del procedimento. La scelta di minimizzare i quadrati degli ε_i non è arbitraria. Se si minimizzasse la somma degli ε_i , distanze positive e negative si compenserebbero, rendendo peggiore la qualità dell'interpolazione; se si adottasse una funzione criterio uguale alla somma dei valori assoluti degli ε_i , non essendo la funzione valore assoluto derivabile su tutto l'asse reale, non si potrebbe ricorrere all'elegante metodo di minimizzazione sopra illustrato.

INTERPRETAZIONE GEOMETRICA DELLE STIME OLS



Il vettore di stime OLS $\underline{\hat{y}}$ consente di ottenere i valori previsti ("teorici") per la variabile dipendente $\underline{\hat{y}} = X\underline{\hat{b}} = X(X'X)^{-1}X'y$. Formalmente l'espressione corrisponde a una proiezione ortogonale del vettore delle osservazioni \underline{y} sullo spazio generato dalle colonne della matrice X .

MULTICOLLINEARITA'

Se due o più colonne della matrice dei regressori X sono linearmente dipendenti, non esiste l'inversa $(X'X)^{-1}$ per cui il vettore di stime OLS non può essere determinato. Se da un lato è assai improbabile che questa eventualità si verifichi nelle applicazioni pratiche, è comunque possibile che alcune colonne della matrice X siano prossime alla dipendenza lineare; in tal caso sarà ancora possibile ottenere un vettore di stime OLS, ma sorgerà il problema della multicollinearità.

Si parla di multicollinearità allorché una o più colonne della matrice dei regressori X sono prossime a essere linearmente dipendenti (correlate). L'effetto della multicollinearità è che la matrice $(X'X)^{-1}$ è prossima all'essere singolare. Questo ha due conseguenze di particolare rilievo nelle applicazioni:

- la significatività statistica dei singoli coefficienti risulta modesta (non si riesce a distinguere il contributo informativo delle variabili e le procedure diventano scarsamente significative);
- il fitting della regressione risulta elevato (si osservano elevati valori dell'indice R^2).

$$VIF_i = \text{Variance Inflation factor} = \frac{1}{1 - R_i^2} \rightarrow \begin{cases} R_i^2 \text{ di una regressione sulla colonna } i \text{ e gli altri regressori} \\ \text{Più } R_i^2 \text{ elevato più } VIF_i \text{ elevato} \rightarrow \text{maggiore multicollinearità} \end{cases}$$

Uso di dummy causa della Multicollinearità esatta

ADATTAMENTO AI DATI

L' R^2 è dato dal rapporto tra la devianza della regressione e la devianza totale. L' R^2 esprime con immediatezza quanta parte della variabilità complessiva del fenomeno Y , che si intende spiegare tramite X , si può attribuire al legame lineare stimato mediante la retta di regressione. L' R^2 è un numero compreso tra 0 e 1.

Se esso vale 0 allora la devianza della regressione è zero e quindi la retta stimata è parallela all'asse x (il modello non contribuisce in alcun modo a spiegare la variabile dipendente y_i in aggiunta alla sua media campionaria). Quindi la retta di regressione non ha alcuna capacità interpretativa per Y .

Viceversa se l' R^2 assume valore paria a 1 la devianza della regressione coincide con la devianza totale e, quindi, la devianza residua è nulla, cioè tutti i residui sono nulli e tutti i valori stimati per la Y sono collocati esattamente sulla retta di regressione che passa per tutti i dati osservati (i quali evidentemente sono perfettamente allineati).

$$\left\{ \begin{array}{l} R^2 = \frac{V(\hat{y}_i)}{V(y_i)} = \frac{\frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\text{DevianzaRegressione}}{\text{DevianzaTotale}} = [\text{se } \exists \text{ intercetta}] = 1 - \frac{V(e_i)}{V(y_i)} \\ \\ R_{\text{senza intercetta}}^2 = R_{\text{non centrato}}^2 = \frac{\sum_{i=1}^N \hat{y}_i^2}{\sum_{i=1}^N y_i^2} = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N y_i^2}, \quad y_i^2 \text{ senza media} \\ \\ R_{\text{modello non lineare}}^2 = \text{Corr}^2(y_i, \hat{y}_i) = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\left[\sum_{i=1}^N (y_i - \bar{y})^2 \right] \left[\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \right]}, \quad \left\{ \begin{array}{l} \text{per OLS coincide con altre def} \\ \text{riflette cmq qualità dell'aprox lineare} \end{array} \right. \\ \\ R_{\text{corretto}}^2 = 1 - \frac{\frac{1}{N-K} \sum_{i=1}^N e_i^2}{\frac{1}{N-1} \sum_{i=1}^N y_i^2} \rightarrow \text{penalizzazione per inclusione variabili esplicative aggiuntive} \end{array} \right.$$

DUMMY

Esistono Dummy additive $y = \alpha + \beta_1 D + \beta_2 x_2$ e Dummy moltiplicative $y = \alpha + \beta_2 x_2 + \beta_3 x_2 D$. Le Dummy additive modificano l'intercetta.

PROPIETA' DEI STIMATORI

$$\left\{ \begin{array}{l} E[\underline{b}] = \underline{\beta}, \quad V[\underline{b}] = \begin{cases} \sigma^2 (X'X)^{-1} \\ \sigma^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \end{cases} \\ \hat{V}[\underline{b}] = \left[\text{stimo } \sigma^2 \text{ con } \hat{\sigma}^2 = s^2 = \frac{1}{N-K} \sum_{i=1}^N e_i^2 \right] = s^2 (X'X)^{-1} = s^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \\ s^2 \text{ è la stima della varianza dei residui considerando che essi hanno valore atteso nullo} \\ \text{La var stimata di un elemento } b_k \text{ è dato da } s^2 c_{kk}, \text{ dove } c_{kk} \text{ è l'elemento } (k, k) \text{ della matrice } \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \end{array} \right.$$

SCelta TRA PIU' MODELLI - AIC, BIC

AIC e BIC sono criteri alternativi all'R quadro corretto che forniscono un trade-off tra qualità dell'adattamento dei dati, e la semplicità del modello, misurata dal numero dei parametri K .

$$AIC = \text{Akaike Information Criteria} = \log \left(\frac{1}{N} \sum_{i=1}^N e_i^2 \right) + \frac{2K}{N}, \quad BIC = \text{Bayesian ...} = \log \left(\frac{1}{N} \sum_{i=1}^N e_i^2 \right) + \frac{K}{N} \log N$$

Modello con AIC, BIC minore è il migliore. Entrambi prevedono penalizzazione per aggiunta di regressori.

IPOTESI DISTRIBUZIONE

$$A5) \underline{\varepsilon} \sim N(0, \sigma^2 I_n) = e_i \sim NID(0, \sigma^2) \rightarrow \text{sostituisce } A1), A3) \text{ e } A4)$$

$$\text{Se vale } A5) \rightarrow \underline{b} \sim N \left(\underline{\beta}, \sigma^2 (X'X)^{-1} \right) \rightarrow b_k \sim N(\beta_k, \sigma^2 c_{kk})$$

Gli stimatori dei minimi quadrati (e le loro proprietà) non derivano da alcuna ipotesi circa la distribuzione delle v.c. ε_i e di conseguenza per n finito nulla si può dire sulla distribuzione esatta degli stimatori OLS; in particolare non si possono derivare test ed intervalli di confidenza esatti per i parametri. Inoltre, non si può discutere di proprietà come la sufficienza e l'efficienza assoluta. Tuttavia con l'ipotesi che i termini di errore sono indipendenti e normali con valori medi nulli e varianze costante il campione osservato (y_1, \dots, y_n) è la realizzazione di un campione casuale $\underline{Y} = (Y_1, \dots, Y_n)$ le cui componenti $Y_i \sim N(\underline{x}_i' \underline{\beta}, \sigma^2)$ costituiscono una n -pla di v.c. indipendenti e somiglianti. Calcolando gli stimatori e le stime ML (Max Likelihood) si vede che essi coincidono con quelle OLS. Gli stimatori ML di $\underline{\beta}$ godono di tutte le proprietà di Gauss-Markov ma sono anche sufficienti ed efficienti in senso assoluto tra tutti gli stimatori non distorti (non solo lineari e non distorti) perché si può dimostrare che essi raggiungono il limite inferiore della disuguaglianza di Cramer-Rao. Inoltre, per qualsiasi n finito, essendo $\underline{\beta}$ delle combinazioni lineari di v.c. Normali gli stimatori ML/OLS \underline{b} sono distribuiti Normalmente con media varianze già esplicitate.

VERIFICA D'IPOTESI

TEST T $\rightarrow \begin{cases} H_0: \beta_k = \beta_k^0 \\ H_1: \beta_k \neq \beta_k^0 \end{cases} \rightarrow t_k = \frac{(b_k - \beta_k^0)}{s\sqrt{c_{kk}}} \sim \mathcal{T}_{N-K} \rightarrow \text{Approx normale } RC(\alpha) \rightarrow \text{rifiuto ipotesi nulla se } |t_k| > 1,96$

TEST RATIO \rightarrow Statistica test da calcolare per verificare ipotesi nulla $\beta_k = 0 \rightarrow \frac{b_k}{se(b_k)}$ con $s\sqrt{c_{kk}} = se(b_k)$

TEST F – CONGIUNTO DI SIGNIFICATIVITA' $\rightarrow J$ dei K ($J < K$) regressori sono $= 0 \rightarrow H_0: \beta_{K-J+1} = \dots = \beta_K = 0$

$S_0 (= \sum e_i^2 \text{ modello vincolato, cioè omettendo } J \text{ regressori}), \quad S_1 (= \text{somma quadrati residui modello completo})$

$f = \frac{(S_0 - S_1)/J}{S_1/(N - K)} = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(N - K)} \sim \mathcal{F}_{N-K}^J \rightarrow RC(\alpha): P(f > \mathcal{F}_{N-K,\alpha}^J) = \alpha, \quad \text{con } \begin{cases} J = \text{Numero Regressori Nulli} \\ K = \text{Numero Regressori Totali} \end{cases}$

TEST F per livello significatività del 5% può essere accettato guardando il p-value (se minore di 5% allora i coefficienti sono diversi da zero).

Table 2.2 OLS results wage equation

Dependent variable: wage			
Variable	Estimate	Standard error	t-ratio
constant	-3.3800	0.4650	-7.2692
male	1.3444	0.1077	12.4853
school	0.6388	0.0328	19.4780
exper	0.1248	0.0238	5.2530
$s = 3.0462 \quad R^2 = 0.1326 \quad \bar{R}^2 = 0.1318 \quad F = 167.63$			

La tabella di seguito può essere interpretata nel modo seguente: sotto “Estimate” viene riportato il valore dei coefficienti del modello di regressione (essi possono essere interpretati solo sotto una condizione ceteris paribus, che stabilisce che tutte le altre variabili incluse nel modello devono essere mantenute costanti); lo “Standard Error” definisce la varianza dei coefficienti di regressione; il t-ratio è la statistica test da calcolare per verificare l'ipotesi nulla $\beta_k = 0$ affinché l'ipotesi che il coefficiente sia uguale a zero sia accettata il t-

ratio deve essere compreso tra -1,96 e +1,96; la statistica test F, infine, verifica congiuntamente che tutti e tre i coefficienti di pendenza siano congiuntamente nulli.

TEST DI WALD

$$WALD \rightarrow R\underline{b} = \underline{q} \rightarrow R\underline{b} \sim N(R\underline{\beta}, \sigma^2 R(X'X)^{-1}R') \rightarrow \xi = \frac{(\underline{R\underline{b}} - \underline{q})[R(X'X)^{-1}R']^{-1}(\underline{R\underline{b}} - \underline{q})'}{\sigma^2} \sim \chi_J^2, \begin{cases} Es. R \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \\ Es. \underline{q} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, J = \dim \underline{q} \end{cases}$$

PROPRIETA' PER CORRETTEZZA

Se non vale A5) $\underline{\varepsilon} \sim N(0, \sigma^2 I_n) \rightarrow$ distribuzione \underline{b} *NON NORMALE*, se $E(\underline{\varepsilon}|X) \neq 0 \rightarrow \underline{b}$ *NON CORRETTO*

CONSISTENZA

La consistenza è una cosiddetta proprietà per grandi campioni e, intuitivamente, afferma che se utilizziamo un numero sempre più grande di osservazioni, la probabilità che lo stimatore si discosti dal vero valore diventa sempre più piccola. Valori di \underline{b} non prossimi a $\underline{\beta}$ diventano sempre più improbabili.

$$\left\{ \begin{array}{l} A6) \frac{1}{N} \sum_{i=1}^N \underline{x}_i \underline{x}_i' \text{ converge matrice non singolare } \Sigma_{xx} + A7) E(\underline{x}_i \varepsilon_i) = 0 \rightarrow \\ \rightarrow \underline{b} \text{ CONSISTENTE} \leftrightarrow \text{plim } \underline{b} = \underline{\beta} \leftrightarrow \lim_{N \rightarrow \infty} P(|b_k - \beta_k| < \epsilon) = 1, \forall \delta \end{array} \right.$$

NORMALITA' ASINTOTICA

$$\text{Sotto ipotesi A1} - \text{A4} + \text{A6} \rightarrow \text{Normalità ASINTOTICA} \rightarrow \underline{b} \sim N\left(\underline{\beta}, \sigma^2 \frac{\Sigma_{xx}^{-1}}{N}\right) \rightarrow \text{stimo } \Sigma_{xx}^{-1} \rightarrow \underline{b} \sim N\left(\underline{\beta}, s^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i'\right)^{-1}\right)$$

PREVISIONE

$$\text{Dato un } \underline{x}_0 \rightarrow \text{Previsione OLS } \hat{y}_0 = \underline{x}_0' \underline{b} \rightarrow V(\hat{y}_0) = \text{VAR PREDITTORE} = V(\underline{x}_0' \underline{b}) = \underline{x}_0' V(\underline{b}) \underline{x}_0 = \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0$$

$$\text{VARIANZA ERRORE DI PREVISIONE} \rightarrow V(y_0 - \hat{y}_0) = \sigma^2 + \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0$$

2 INTERPRETAZIONE E CONFRONTO MODELLI DI REGRESSIONE

INTERPRETAZIONE MODELLO LINEARE

Il coefficiente β_k misura la variazione attesa in y_i quando x_{ik} cambia di una unità, ma tutte le altre esplicative in \underline{x}_i' restano costanti.

$$y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \beta_k = \frac{\partial E(y_i | \underline{x}_i)}{\partial x_{ik}} \rightarrow \begin{cases} \text{Misura variazione assoluta (Ceteris paribus: altre var non cambiano)} \\ \text{della variabile } y_i \text{ rispetto ad una variazione di } x_{ik} \end{cases}$$
$$\frac{\partial E(y_i | \underline{x}_i)}{E(y_i | \underline{x}_i)} \bigg/ \frac{\partial x_{ik}}{x_{ik}} = \text{ELASTICITA' MODELLO LINEARE} = \frac{x_{ik}}{\underline{x}_i' \underline{\beta}} \beta_k \rightarrow \begin{cases} \text{Nei modelli lineari} \\ \text{le elasticità non sono costanti} \\ \text{come in quelli logaritmici} \end{cases}$$

INTERPRETAZIONE MODELLO LOGARITMICO

$$\log y_i = (\log \underline{x}_i)' \underline{\gamma} + i \rightarrow \gamma_k = \frac{\partial E(\log y_i | \log \underline{x}_i)}{\partial \log x_{ik}} \approx \left(\log(g(x))' = \frac{g'(x)}{g(x)} \right) \approx \frac{\partial E(y_i | \underline{x}_i)}{E(y_i | \underline{x}_i)} \bigg/ \frac{\partial x_{ik}}{x_{ik}}$$

Nei modelli logaritmici le variazioni attese sono approssimabili come elasticità e sono costanti -> esse, approssimativamente uguali ai coefficienti di regressione, misurano la variazione relativa (percentuale) Ceteris paribus della variabile y_i rispetto ad una variazione di x_{ik} .

DISTORSIONE DA VARIABILI OMESSE

La distorsione da variabile omessa nasce quando viene omessa una variabile dalla regressione, che è una determinante di Y ed è correlata con uno o più dei regressori. L'omissione di variabili rilevanti può rendere le stime OLS inconsistenti. Infatti, la presenza di una variabile omessa potrebbe determinare correlazione tra i residui e la variabile della regressione correlata con la variabile omessa. Una soluzione si ha utilizzando una variabile strumentale.

CONFRONTO MODELLI LINEARI NON NESTED (NESSUNO CASO PARTICOLARE DELL'ALTRO)

$$y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \text{MODELLO A}, \quad y_i = \underline{z}_i' \underline{\gamma} + i \rightarrow \text{MODELLO B}$$

Stimo una regressione uguale al modello A o B aggiungendo le variabili non presenti in un modello, ma presenti nell'altro modello. Guardo alla significatività dei coefficienti aggiunti.

DATA MINING

Stesso insieme di osservazione viene usato sia per la specificazione del modello che per verificare ipotesi.

CONFRONTO MODELLI LINEARI-LOGARITMICI NON NESTED

$$\begin{aligned}
 & y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \text{MODELLO A}, \quad \log y_i = (\log \underline{x}_i)' \underline{\gamma} + i \rightarrow \text{MODELLO B} \\
 & \hat{y}_i = \text{previsioni OLS usando A}, \quad \tilde{y}_i = \text{previsioni OLS usando B} \\
 & \left\{ \begin{array}{l} \text{Verifico linearità} \rightarrow y_i = \underline{x}_i' \underline{\beta} + \delta_{\text{lin}}(\log \hat{y}_i - \log \tilde{y}_i) + \varepsilon_i \rightarrow \text{se } \delta_{\text{lin}} = 0 \rightarrow \text{OK LINEARITA'} \\ \text{Verifico log - linearità} \rightarrow \log y_i = (\log \underline{x}_i)' \underline{\gamma} + \delta_{\log}(\hat{y}_i - \tilde{y}_i) + i \rightarrow \text{se } \delta_{\log} = 0 \rightarrow \text{OK LOG - LINEARITA'} \end{array} \right.
 \end{aligned}$$

TEST FORMA FUNZIONALE

Una strategia per verificare la correttezza della forma funzionale di $E(y_i | x_i) = \underline{x}_i' \underline{\beta}$ consiste nel verificare la significatività di termini aggiuntivi non lineari in \underline{x}_i . Si può usare test t standard, test F, o più in generale test di Wald, ma si deve essere in grado di specificare in dettaglio l'ipotesi alternativa. Altrimenti faccio un test RESET dove tramite una regressione ausiliaria $y_i = \underline{x}_i' \underline{\beta} + a_2 \hat{y}_i^2 + \dots + a_q \hat{y}_i^q + \varepsilon_i$ verifico se le $a_i = 0$. Ovvero verifico che i coefficienti di una regressione ausiliaria costruita sulle previsioni del modello lineare abbia coefficienti non nulli.

TEST DI CHOW O DI BREAK STRUTTURALE

I coefficienti del modello sono diversi per due sottogruppi del campione (ad esempio prima o dopo un importante intervento di politica macroeconomica). Definiamo la variabile indicatrice g_i : $g_1 = 0$ gruppo 1, $g_1 = 1$ gruppo 2.

$$y_i = \underline{x}_i' \underline{\beta} + g_i \underline{x}_i' \underline{\gamma} + \varepsilon_i, \quad \text{dove il vettore } K - \text{dimensionale } g_i \underline{x}_i' \text{ contiene interazioni variabile esplicative con } g_i$$

Il vettore dei coefficienti per il gruppo 1 è $\underline{\beta}$ mentre quello del gruppo 2 è $\underline{\beta} + \underline{\gamma}$

$$H_0: \underline{\gamma} = 0 \rightarrow f = \frac{(S_V - S_{NV})/K}{S_{NV}/(N - 2K)} \sim \mathcal{F}_{N-2K}^K, \quad \begin{cases} K = \text{numero regressori compresa intercetta} = 0 \\ 2K = \text{numero di regressori totali} \rightarrow \text{si noti che } \dim \underline{x}_i' = \dim g_i \underline{x}_i' = K \end{cases}$$

Un test alternativo che non fa uso delle variabili dummy confronta i residui della regressione sui due gruppi con la somma dei residui di due regressioni, una costruita sul gruppo 1 e una sul gruppo 2.

Supponendo che il modello dei dati sia $y_t = a + bx_{1t} + cx_{2t} + \varepsilon_t$. Dividiamo i dati in due gruppi:

$$\begin{cases} y_t = a_1 + b_1 x_{1t} + c_1 x_{2t} + \varepsilon_t \\ y_t = a_2 + b_2 x_{1t} + c_2 x_{2t} + \varepsilon_t \end{cases}$$

L'ipotesi nulla del test di Chow asserisce che $a_1 = a_2, b_1 = b_2, c_1 = c_2$. Sia S_c la somma dei residui del modello completo, S_1 somma dei residui del modello 1, S_2 la somma dei residui del modello 2. Allora la statistica test di Chow è la seguente:

$$f = \frac{(S_c - (S_1 + S_2))/K}{(S_1 + S_2)/(N_1 + N_2 - 2K)} \sim \mathcal{F}_{N_1+N_2-2K}^K$$