**PARTE II. ANALISI DI REGRESSIONE LINEARE**

**Osservazione:** nel seguito una **variabile casuale** viene indicata più appropriatamente come **numero aleatorio (n.a.)**.

**Funzione di regressione**

 Considerati due numeri aleatori (n.a.) Y e X, è detta “**funzione di regressione di Y rispetto X**” **il n.a. E(Y|X)** dipendente da X secondo una funzione g(.) [cioè E(Y|X) = g(X)] determinata dalla distribuzione subordinata di Y rispetto X.

I valori del n.a. g(X) [ossia i valori del n.a. E(Y|X) ] sono le speranze matematiche condizionate E(Y|X = x). Pertanto l’evento **{g(X) = E(Y|X = x)}** coincidecon l’evento “Il n.a. E(Y|X) assume il valore E(Y|X = x) “ ed esso ha probabilità pari a Prob(X = x) quando il n.a. X è discreto; altrimenti ad esso è associata la densità di probabilità marginale del n.a. X nel punto x, cioè f(x).

La conoscenza della funzione di regressione E(Y|X) permette quindi di conoscere per ogni possibile valore x assunto dal n.a. X quale sarà il valore medio (condizionato) di Y. Inoltre al variare di x, saremo in grado di sapere come varia in media (condizionata) Y.

 Alcune proprietà della funzione di regressione sono le seguenti:

1. E( aY + b Z | X) = a E(Y|X) + b E(Z|X) , se a e b sono numeri certi (ossia delle costanti) e X, Y e Z sono numeri aleatori (**proprietà di linearità della funzione di regressione**);
2. E[E(Y|X)] = E(Y) (**legge del valore atteso iterato**);
3. E[Y\*h(X) |X] = h(X)\*E(Y|X) ; come caso particolare della 3) vale

3.1) E[E(Y|X) | X] = E[ g(X) | X] = g(X)=E(Y|X), essendo E(Y|X) una funzione g(.) di X.

1. se i numeri aleatori X e Y sono stocasticamente indipendenti, E(Y|X) = E(Y).
2. E[Y – E(Y|X)]2 ≤ E[Y - ψ(X)]2 per ogni funzione reale ψ(.) tale che E[ψ(X)]2 < ∞ .

 Una definizione assiomatica di funzione di regressione E(Y|X) è la seguente: essa è quel numero aleatorio dipendente da X che verifica la condizione  per ogni funzione  per la quale esista la speranza matematica a primo membro.

 La nozione di funzione di regressione è fondamentale nella teoria e calcolo delle probabilità; è il caso di osservare che in lingua inglese E(Y|X) è detta “conditional expectation”, ma con lo stesso nome si indicano anche i suoi possibili valori E(Y|X=x) e ciò può ingenerare fraintendimenti. In questi appunti indicheremo con il nome di funzione di regressione il numero aleatorio E(Y|X) e con quello di valor medio condizionato le possibili determinazioni, E(Y|X=x), di E(Y|X) che sono invece numeri certi.

E’ immediato generalizzare la definizione di funzione di regressione di Y rispetto X al caso in cui ci si condiziona rispetto a più di un n.a. Ad esempio, la funzione di regressione di Y rispetto X e Z è indicata con E(Y|X,Z) ed è una funzione g(X,Z). Quindi E(Y|X,Z) è un n.a. i cui valori sono le speranze matematiche condizionate E(Y|X = x, Z=z). L’evento “Il n.a. E(Y|X,Z) assume il valore E(Y|X = x, Z=z) “ ha probabilità pari alla probabilità congiunta Prob(X = x, Z=z) quando entrambi i n.a. X e Z sono discreti; se X e Z sono entrambi (assolutamente) continui ad esso è associata la densità di probabilità congiunta dei n.a. X e Z nel punto (x,z), cioè f(x,z).

Le proprietà di E(Y|X) sopra esposte sono immediatamente generalizzabili. In particolare, se X e Y sono stocasticamente indipendenti risulta E(Y|X,Z) = E(Y|Z). Un’ultima importante proprietà è la seguente generalizzazione della legge del valore atteso iterato

1. E[E(Y|X) | X,Z] = E[E(Y | X,Z) |X] = E(Y/X) ;

**Un esempio di funzione di regressione**.

 Con riferimento ad un unico lancio di un dado regolare si considerino il numero aleatorio N che individua il numero uscito nel lancio e l’indicatore |D| dell’evento D = “Esce un numero dispari”, cioè il n.a. che vale 1 se esce uno dei tre risultati 1, 3, 5 e che vale 0 se esce uno dei tre risultati 2, 4, 6. Se si è convinti che il lancio del dado è fatto senza trucchi è plausibile che si assegni probabilità 1/6 ad ognuno dei sei risultati possibili; in questo caso si ha anche P(D) = ½ .

 La distribuzione di probabilità congiunta dei n.a. |D| ed N è la seguente

 N

 1 2 3 4 5 6

  

in quanto, per esempio, l’evento [(N = 1) ∩ D] coincide con (N = 1) (si ricordi che (N = 1) implica D), per cui è P[(N = 1) ∩ D] = P(N = 1) = 1/6 ed inoltre l’evento [(N = 1) ∩ ] è impossibile, per cui è P[(N = 1) ∩ ] = 0.

 La distribuzione condizionata di N all’evento , cioè al verificarsi di un numero pari, assegna evidentemente una probabilità pari a 1/3 ai valori 2, 4 e 6 e probabilità nulla ai rimanenti; similmente la distribuzione condizionata di N all’evento D, cioè al verificarsi di un numero dispari, assegna la probabilità 1/3 ai valori 1, 3, 5 e la probabilità nulla ai rimanenti. Di conseguenza, per le speranze matematiche condizionate di N a ciascuno dei due eventi  e *D* si ha:

 E(N|) = (2 + 4 + 6) = 4 e E(N|) = (1 + 3 + 5) = 3 .

 In questo esempio, la funzione di regressione di N rispetto all’indicatore |D| , indicata con

 E( N | |D| ) , è il numero aleatorio dipendente da |D| che assume il valore 4 , cioè E(N|) , se |D| = 0 (evento di probabilità ½) ed il valore 3 , cioè E(N|) , se |D| = 1 (evento di probabilità ½) . Avendo perciò specificato i valori possibili del n.a. E( N | |D| ) e le corrispondenti probabilità, esso è compiutamente noto. Si osservi infine che risulta

 E[ E( N | |D| ) ] = 4 . ½ + 3 . ½ = 3.5 = E( N ) .

 Lasciando al lettore i facili calcoli, ci limitiamo ad affermare che la funzione di regressione dell’indicatore |D| rispetto ad N, E( |D| | N ) , è il n.a. che assume valore 1 quando N assume valori dispari e 0 quando N è pari ; ovviamente, risulta E [E( |D| | N ) ] = ½ = P (D) .

**Un altro esempio di funzione di regressione**

 Supponiamo che due n.a. X e Y abbiano una distribuzione congiunta di tipo normale bivariato con valori medi E(X) ed E(Y), varianze V(X) e V(Y) e covarianza Cov(X,Y) = c ; si prova che le due distribuzioni condizionate sono entrambe normali univariate con parametri espressi dalle:

  ; 

  ; 

Poiché i primi momenti di queste densità subordinate sono i possibili valori delle corrispondenti funzioni di regressione, si ha:

  ;  .

**Modelli di regressione lineare**

 Operativamente, l’analisi di regressione viene usata per studiare **l’influenza del n.a. X sul valor medio di Y;** se si volesse analizzare **l’influenza di X su Y** bisognerebbe far ricorso alla distribuzione subordinata di Y rispetto X, la qual cosa risulta decisamente più impegnativa (si consideri che i valori della funzione di regressione sono i momenti primi delle distribuzioni subordinate di Y rispetto ai possibili valori di X: nell’analisi di regressione gli altri infiniti momenti di quelle distribuzioni dunque non intervengono !).

 Nelle applicazioni concrete, il più delle volte non si ha nessuna idea sulla **forma funzionale** g(.) della funzione di regressione E(Y|X) = g(X) per cui è necessario introdurre un’ipotesi di lavoro per g(.): **ogni ipotesi su tale forma funzionale** (per esempio g(X) = , oppure g(X) = exp{-aX}, oppure g(X) = ,………..) **costituisce** **un modello di regressione.** Il modello di gran lunga più usato per la sua semplicità è quello di regressione lineare (affine)

 E(Y|X) = .

Una forma equivalente dello stesso modello lineare[[1]](#footnote-1) è , ove per definizione U = Y – E(Y|X).

Utilizzando le suddette proprietà della funzione di regressione si prova che risulta E(U|X) = 0 e quindi E(U) = 0, Cov (U , X) = 0. Si osservi che mentre Y e X sono variabili osservabili, non lo è la variabile U.

Dimostriamo che E(U|X) = 0 per il modello lineare affine indicato sopra con U = Y – E(Y|X). Dalla definizione di U abbiamo che E(U|X) = E[ (Y – E(Y|X))|X]= E[Y|X] – E[E(Y|X))|X]=

= E[Y|X] – E[Y|X]=0, essendo E[E(Y|X))|X]=E[h(X)|X] dove h(X)= E(Y|X).

Usando la legge del valore atteso iterato si dimostra che E(U)= E[E(U|X)] =E[0]=0. Infine, poiché E[U]=0, risulta cov(U,X)=E(X\*U). Possiamo applicare nuovamente la legge del valore atteso iterato: E(X\*U)= E[ E(X\*U |X) ]. Risulta, E(X\*U |X)= X\*E(U|X)= X\* 0= 0. Pertanto E(X\*U) = E[0] = 0.

Si noti infine che usando lo stesso procedimento sopra indicato è facile dimostrare che quando E(U|X) = 0, vale il risultato più generale: cov( *m*(X) , U)=0 per ogni funzione *m*(X) avente secondo momento finito.

Di solito nei libri di testo il modello di regressione lineare affine viene presentato scrivendo l’equazione:



e ponendo come prima assunzione del modello l’ipotesi:

1. E[U|X]=0.

E’ facile dimostrare che partendo da tali due asserzioni la funzione di regressione di Y dato X coincide con

E(Y|X) = .

Ossia quest’ultima formulazione e quella da noi inizialmente presentata risultano tra loro equivalenti. Dimostriamo quanto affermato. Poiché  risulta che E(Y|X) = . Usando la proprietà di linearità possiamo riscrivere=, dove . Infine=.

 La stima dei parametri del modello di regressione si effettua con uno dei procedimenti statistici di stima puntuale: il metodo di massima verosimiglianza, il metodo dei minimi quadrati, etc. Nel seguito useremo il secondo dei due, e cioè il **metodo dei minimi quadrati**, che ora brevemente richiamiamo.

 Supponendo di poter conoscere i valori di T coppie di variabili osservabili  e avendo introdotto per esse il modello di regressione lineare   , t = 1,……….,T , si stimano i due coefficienti, , con quei valori numerici che rendono minima la funzione

 ;

si prova che i valori che annullano le due derivate parziali di  sono le coordinate del punto di minimo.

**Proprietà di consistenza di uno stimatore**

Ricordiamo a questo punto la proprietà di consistenza di uno stimatore  (costituito da una qualche funzione di T variabili osservabili) di un qualche parametro incognito θ:  è consistente se p-lim  = θ al divergere di T, ossia se **lo stimatore converge in probabilità al parametro da stimare**. Il significato della condizione p-lim  = θ è il seguente: fissato arbitrariamente un ε > 0, accade che lim Pr per  . In termini discorsivi, uno stimatore è consistente se al crescere del numero T delle variabili osservabili diminuisce la probabilità di commettere errori di stima maggiori di ε in modulo. Si osservi che ciò che diminuisce al crescere dell’informazione campionaria (rappresentata dal numero T delle osservazioni) non è l’errore di stima, ma la probabilità che l’errore  sia maggiore di ε ! La condizione deve valere per ogni ε > 0 **piccolo a piacere**, anche pari a 0,00000001. E’ possibile dimostrare che se:

1. $\lim\_{T\to \infty }E\left[S\_{T}\right]=θ$;
2. $\lim\_{T\to \infty }Var\left[S\_{T}\right]=0$;

lo stimatore  è consistente (condizione sufficiente ma non necessaria).

**Alcune proprietà concernenti il modello di regressione lineare semplice e gli stimatori OLS dei coefficienti**.

 Enunceremo alcuni risultati che si riveleranno utili nel seguito.

Si verifica facilmente che la retta di regressione stimata  contiene il punto  le cui coordinate sono le medie aritmetiche dei valori osservati.

Per quanto concerne i residui  sussistono le relazioni seguenti, facilmente dimostrabili:

 1);

 2) ;

 3) .

 In forza della proprietà 3) risulta  ed il rapporto /, indicato con il simbolo  e denominato **indice di determinazione**, viene usato come indice di affidabilità del modello di regressione stimato nel senso che quanto più prossimo ad 1 risulta essere, tanto più affidabile è ritenuto il modello stimato . Si noti che la definizione data di  è valida anche per i modelli di regressione multipla.

**Alcune proprietà del valore atteso della varianza e della covarianza**

Siano *a,* *b, c* e *d* delle costanti(n. certi) e X, Z e W n.a. Vale:

1. E[a]=a
2. E[a\*X]=a\*E[X]
3. E[a\*X+*b*\*Z]=a\*E[X]+*b*\*E[Z]
4. E[a\*X-*b*\*Z+c]=a\*E[X]-*b*\*E[Z]+*c*

Pertanto E[ X-E[X] ]=0

1. Var(X)= E[(X-E[X])2]=E[X2]-{E[X]}2
2. Var(*a*)=0
3. Var(*a*\*X)= *a2* \* Var(X)

Pertanto Var( [1/Var(X)1/2 ]\*X )=1

Quindi il n.a. [ (X-E[X])/ Var(X)1/2] ha media nulla e varianza unitaria.

1. Var(*a*X+(-)*b*Z)= *a2* \* Var(X)+(-)2\**a*\**b*\*Cov(X,Z)+*b2* \* Var(Z)
2. Cov(X,Z)=E[(X-E[X])\*(Z-E[Z])]=E[X\*Z]-E[X]\*E[Z]
3. Cov(*a*,X)=0
4. Cov(*a*\*X,Z)= *a*\*Cov(X,Z)
5. Cov(X,*b*\*Z)= *b*\*Cov(X,Z)
6. Cov(*a*X+(-)b\*W+(-)*c*, *d*\*Z)= *a\*d*\*Cov(X,Z)+(-)*b*\**d*\* Cov(W,Z)
7. Corr(X,Z)=Cov(X,Z) / [Var(X)\* Var(Z)]1/2
8. -1 ≤ Corr(X,Z) ≤ +1
1. Ottenuta partendo dall’ipotesi che la funzione di regressione di Y rispetto X è lineare affine. [↑](#footnote-ref-1)