

La regressione: riepilogo (1)

A cosa serve la regressione?

La regressione lineare è un metodo statistico finalizzato a tentare di individuare una relazione tra una variabile Y e una o più variabili X che la determinano tramite una relazione di tipo analitico (una funzione, lineare o non lineare).

In ogni modello c'è una sola variabile *dipendente* e una (regressione semplice) o più variabili *indipendenti* (regressione multipla).

La regressione: riepilogo (2)

Il tipo di variabile usato come «dipendente» influisce sulla scelta del modello da usare?

Decisamente sì.

Quando la variabile dipendente è cardinale si usa la regressione lineare (semplice o multipla).

Quando la variabile dipendente è categoriale si usa la regressione logistica (in questo caso la dipendente è espressa tramite una variabile dicotomica).

La regressione: riepilogo (3)

Che tipo di variabili possono essere usate come «indipendenti»?

Variabili cardinali (dopo aver controllato la forma della distribuzione di ciascuna variabile).

Variabili categoriali, che vanno sottoposte ad un processo di semplificazione e dicotomizzazione, al fine di produrre delle variabili «dummy» (una dicotomica, ad esempio, produrrà una variabile «dummy» dove una delle due modalità sarà la «categoria di riferimento»; vedi lucidi pubblicati in precedenza).

La regressione: riepilogo (4)

A partire da una variabile categoriale usata come dipendente, quante «dummy» vanno prodotte?

Dopo aver controllato la distribuzione di frequenza, con l'accortezza di avere una distribuzione sufficientemente bilanciata (ed eventualmente ricodificando la variabile accorpendone le modalità), le «dummy» che si ottengono da una variabile categoriale sono « $n-1$ ».

Se la categoriale, per es., ha tre modalità, essa produrrà due variabili dicotomiche (con una terza, non in matrice, che funge da «categoria» di riferimento).

L'uso di variabili "dummy"

Il caso della variabile di "genere" come indipendente

<i>Femmina</i>	1	0		
<i>Maschio</i>	0	0	—————→	Categoria di riferimento

Il caso della variabile "zona di residenza" come indipendente

<i>Nord</i>	1	0	0	
<i>Centro</i>	0	1	0	
<i>Sud</i>	0	0	0	—————→ Categoria di riferimento

La scelta della categoria di riferimento **non è arbitraria**.

Essa **comporta importanti implicazioni** sul piano dell'interpretazione dei risultati.

Essa deve **rispondere alle ipotesi teoriche** del ricercatore .

La regressione: riepilogo (5)

Quali sono i principali coefficienti da leggere per interpretare i risultati del modello?

R^2 (r-square), cioè il coefficiente di determinazione, che indica quanta della varianza della dipendente è riprodotta dal modello.

I coefficienti «beta standard» che ci informano sull'entità dell'effetto di ciascuna variabile indipendente sulla variazione della variabile dipendente (nel caso della regressione logistica i coefficienti «beta standard» esprimono una probabilità). Vi sono poi i valori di significatività per ogni coefficiente.