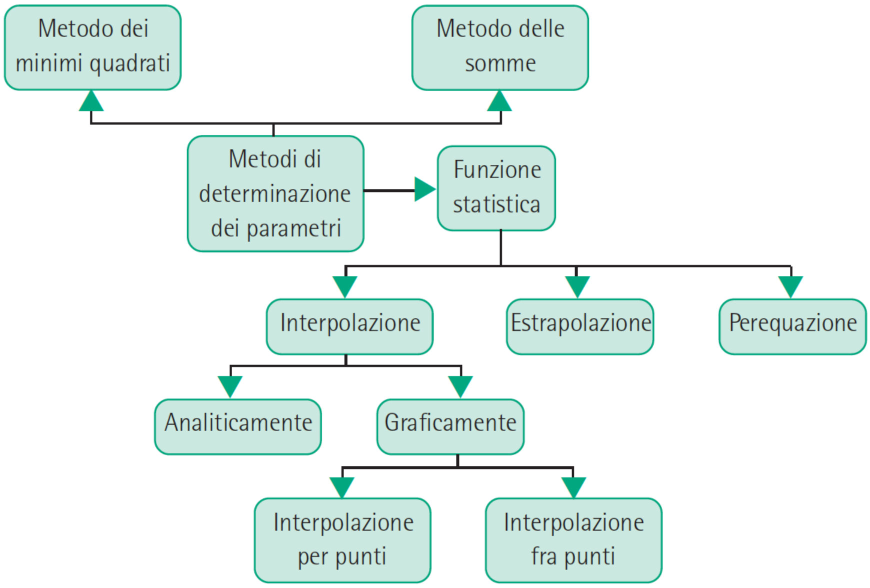
5. RAPPRESENTAZIONE ANALITICA DI VARIABILI. INTERPOLAZIONE, ESTRAPOLAZIONE E PEREQUAZIONE



1) INTERPOLAZIONE, ESTRAPOLAZIONE E PEREQUAZIONE

L'analisi statistica offre dei procedimenti per risolvere diverse tipologie di problemi connessi alla traduzione, in termini matematici, di una legge statistica.

1.1 INTERPOLAZIONE

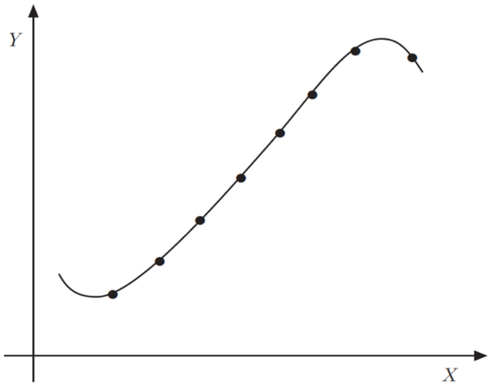
Non sempre, a seguito di una rilevazione statistica, si deducono, in corrispondenza delle modalità di un carattere, le rispettive frequenze o intensità.
L'**interpolazione** è il processo di determinazione di una successione di valori (tutti o parte teorici) di frequenze o intensità, ottenuti in corrispondenza di valori osservati di modalità di un carattere quantitativo in una distribuzione di frequenza, o modalità di tempo in una serie storica.
Il procedimento si attua sia *analiticamente* sia *graficamente*.

ANALITICAMENTE

La **rappresentazione analitica** consiste nel trovare una funzione matematica che rappresenti nel miglior modo possibile la distribuzione osservata del fenomeno.

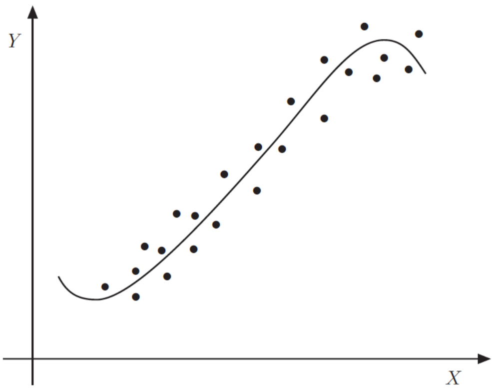
GRAFICAMENTE

La **rappresentazione grafica** consiste nel sostituire al diagramma rappresentativo dei dati osservati (**diagramma a scatter**) una curva teorica associata ad una funzione matematica.
Si parla di **interpolazione per punti** o **matematica** se le variabili X e Y non sono affette da errori e la distribuzione è costituita da n coppie di valori (x_i, y_i) cui corrispondono n punti-immagine nel piano cartesiano. La rappresentazione analitica si attua mediante una funzione matematica cui corrisponde una curva che passi **per** tutti i punti disponibili, i quali si devono trovare sulla curva teorica e le loro coordinate devono soddisfarne l'equazione.



Interpolazione per punti

Si parla di **interpolazione fra punti** o **statistica** se una delle due variabili *X* e *Y* o entrambe sono affette da errori; in tal caso l'interpolazione consiste nel determinare valori teorici delle variabili non affetti da errori. La rappresentazione analitica si attua determinando una funzione matematica associata ad una curva che passi **tra** i punti-immagine del diagramma a dispersione della distribuzione osservata.



Interpolazione fra punti

1.2 ESTRAPOLAZIONE

Talvolta si rende necessario disporre di frequenze o intensità esterne al campo di osservazione del carattere investigato.
L'**estrapolazione** è il processo di determinazione di una successione di valori teorici di frequenze o intensità, ottenuti in corrispondenza di modalità di un carattere quantitativo in una distribuzione di frequenza, o modalità di tempo in una serie storica, esterne all'intervallo di osservazione.
Il procedimento si attua sia *analiticamente* sia *graficamente*. Esso presuppone che il fenomeno si sia svolto in passato con una certa regolarità.
In quanto basato solo sulla regolarità, in passato, del fenomeno che rappresenta, l'estrapolazione è poco attendibile non tenendo conto di cause perturbatrici che potrebbero verificarsi in futuro.

1.3 PEREQUAZIONE

Alcuni dati possono essere affetti da errori accidentali a danno o a vantaggio di quelli che immediatamente li seguono o li precedono nella successione. La tecnica statistica che consente di eliminare tali errori dai dati e, in particolare, irregolarità di andamento nelle serie storiche, è definita **perequazione**.
Il metodo di perequazione più semplice è la **media mobile** che consiste nel sostituire a ciascun termine il valore medio aritmetico di un gruppo di tre, cinque, sette, ... termini, di cui il termine dubbio è il centrale.
La funzione che consente di adattare ai dati osservati dati teorici è detta **funzione perequatrice**.

2) METODI DI DETERMINAZIONE DEI PARAMETRI

Esistono diversi metodi per determinare i parametri di una funzione matematica in un procedimento di interpolazione.

2.1 METODO DEI MINIMI QUADRATI

Il **metodo dei minimi quadrati** consiste nel determinare valori dei parametri della curva teorica che *rendono minima la somma dei quadrati degli scarti tra valori teorici e valori osservati*.

Date due variabili *X* e *Y* se la funzione teorica è lineare nei parametri, cioè è del tipo:

$$\hat{Y} = \beta_0 + \beta_1 X$$

i parametri da determinare sono β_0 e β_1. Siano $\hat{y}_i$ e y_i, rispettivamente il valore teorico e quello osservato sulla *i-esima* variabile, il metodo dei minimi quadrati è tale per cui:

$$S = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \min$$

ossia:

$$S = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i - y_i)^2 = \min$$

I coefficienti incogniti sono:

$$\beta_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$\beta_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

2.2 METODO DELLE SOMME

Il metodo delle somme è un metodo di interpolazione per cui, una volta scelta la funzione rappresentatrice dei dati osservati, si suddivide la distribuzione data in tante sub-distribuzioni quanti sono i parametri e si impone, per ciascuna di esse, l'uguaglianza tra la somma dei valori teorici e la somma dei valori osservati.

Esempio

Sia data la distribuzione riportata nella tabella 1.
Determiniamo l'equazione della retta interpolante ottenuta utilizzando il metodo delle somme.

Per determinare l'equazione della retta interpolante attraverso il metodo delle somme, si divide la distribuzione in due sub-distribuzioni e si usano le ascisse di comodo, ricavando i seguenti valori teorici:

Anno	Ascisse di comodo	$\hat{Y}_i$	y_i
2000	0	β_0	800
2001	1	$\beta_0+\beta_1$	980
2002	2	$\beta_0+2\beta_1$	1.040
Somma	—	$3\beta_0+3\beta_1$	2.820
2003	3	$\beta_0+3\beta_1$	1.200
2004	4	$\beta_0+4\beta_1$	1.240
2005	5	$\beta_0+5\beta_1$	1.550
Somma	—	$3\beta_0+12\beta_1$	3.990

Per ciascuna sub-distribuzione si impone la condizione:

somma dei valori teorici = somma dei valori osservati

ottenendo il sistema nei parametri incogniti β_0 e β_1:

$$\begin{cases} 3\beta_0+3\beta_1=2.820 \\ 3\beta_0+12\beta_1=3.990 \end{cases}$$

da cui:

$$\beta_0=\frac{\begin{vmatrix} 2.820 & 3 \\ 3.990 & 12 \end{vmatrix}}{\begin{vmatrix} 3 & 3 \\ 3 & 12 \end{vmatrix}}=\frac{33.840-11.970}{36-9}=\frac{21.870}{27}=810$$

$$\beta_1=\frac{\begin{vmatrix} 3 & 2.820 \\ 3 & 3.990 \end{vmatrix}}{\begin{vmatrix} 3 & 3 \\ 3 & 12 \end{vmatrix}}=\frac{11.970-8.460}{36-9}=\frac{3.510}{27}=130$$

L'equazione della retta interpolante è, dunque:

$$\hat{Y}=810+130X$$