

Strumenti per l'Analisi Incrociata dei Dati

Nei primi spreadsheet evoluti, parliamo dei primi 123 della Lotus, dei primi Quattro Pro della Borland, del primo MS Excel, il 2.0, sotto Windows, esistevano già delle funzioni di Database, che permettevano di gestire delle semplici liste di dati, che venivano, un po' scorrettamente, chiamate Database. Le liste si potevano ordinare, se ne potevano estrarre dei sottoinsiemi, si potevano sottoporre ad alcune funzioni «chiocciola» di categoria statistica di database. Insomma si poteva non tanto gestire la lista, inserendo o modificandone righe, quanto manipolare i dati in essa contenuti. Con il passare del tempo, pardon delle versioni, in concomitanza con il sempre maggior interesse da parte degli utilizzatori verso quella particolare attività che si chiama analisi dati, i prodotti si sono via via specializzati, proponendo ulteriori strumenti da utilizzare in questa specifica attività

di Francesco Petroni

Tra i vari strumenti introdotti quello più sofisticato è quello che produce una analisi incrociata, ed è quello che analizzeremo, vedendolo in vari applicativi, in questo articolo.

Chiariamo, andando per punti, i concetti ai quali ci riferiamo:

- per Analisi Dati con lo spreadsheet intendiamo quella attività che consiste nel maneggiare uno o più elenchi di dati, preesistenti e disponibili in qualsiasi formato. Chiameremo per comodità tali elenchi Tabelle Dati. I dati non vengono gestiti, nel senso che l'analisi dei dati non si occupa di creare o di modificare record, ma vengono manipolati allo scopo di estrarne ulteriori informazioni prevalentemente di tipo statistico, mediante aggregazioni, ordinamenti, calcoli di somme, di medie, ecc.

- i moderni spreadsheet trattano indifferentemente Tabelle Dati, in formato interno o in formato esterno, permet-

tendo di eseguire, specie con i dati in formato esterno, delle operazioni di Join (unione di dati provenienti da più Tabelle),

- indipendentemente dall'origine delle Tabelle Dati, interne/esterne, uniche o ottenute da unioni, le manipolazioni che si possono eseguire sono comunque le stesse,

- nelle prime versioni degli spreadsheet esistevano strumenti rudimentali che potevano essere utilizzati, da persone esperte, per ottenere anche analisi di tipo complesso,

- nelle ultime versioni degli spreadsheet esistono degli strumenti molto più evoluti, che possono essere usati per lo stesso scopo anche da persone poco esperte,

- lo strumento principale per l'analisi dei dati è oggi, con le attuali versioni dei vari prodotti, il CrossTab, in italiano Analisi Incrociata oppure Campi Incrociati,

ormai presente in tutti gli ultimi spreadsheet sottoforma di funzionalità a sé stante, che vuole in input una Tabella Dati e produce in output una Tabella di Risultati.

Definiti questi aspetti iniziali vediamo cosa ci ripromettiamo di fare in questo articolo.

Vogliamo fare sostanzialmente due cose: la prima è esplorare la funzionalità di CrossTab presente in numerosi prodotti, la seconda è simulare la funzionalità di CrossTab con gli altri strumenti più rudimentali, già presenti nei primi fogli elettronici.

Riteniamo che un utilizzatore evoluto debba conoscere le une e le altre e debba saper valutare quando sia meglio usarle, le une o le altre.

Come strumenti software useremo i più recenti prodotti di tipo spreadsheet e DBMS. Per sperimentare la stessa problematica con un linguaggio di programmazione più tradizionale useremo, ovviamente, il Visual Basic.

Il nostro semplice caso studio

Abbiamo a disposizione una Tabella Dati con 2149 record, la vediamo apparire (parzialmente) in numerose figure.

Si tratta di una tabella unica con pochi campi. Diamo per scontato che le stesse operazioni di analisi che stiamo per compiere possono essere eseguite su più tabelle dati unite da operazioni di Join, e su più campi. Essendo pochi i campi li elenchiamo tutti:

COD, che corrisponde ad un numero di matricola,
COGNOME, nulla da dichiarare,
SEDE, le sedi sono 5: BARI, MILANO, NAPOLI, ROMA, TORINO,

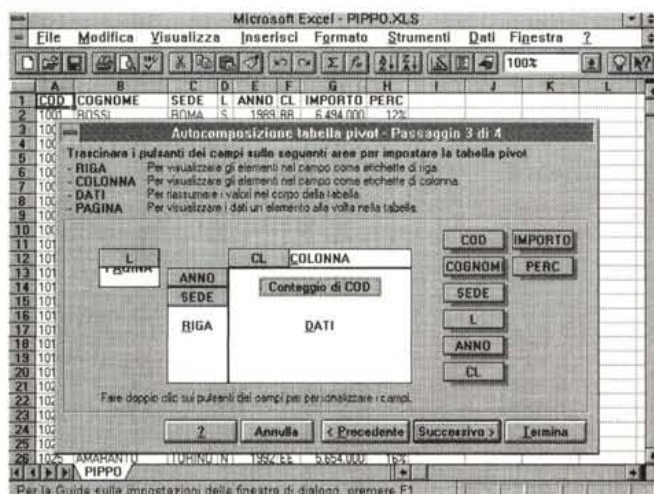


Figura 1 - MS Excel 5.0 - Tabella Pivot con l'autocomposizione.

Nell'ultima versione di Excel ha fatto la sua apparizione la Tabella Pivot, che in pratica costituisce uno strumento di Analisi Incrociata dinamica, nel senso che è possibile, anche in un secondo momento, variare i campi coinvolti, i calcoli e i tipi di calcolo desiderati e ottenere l'aggiornamento dei risultati. La prima «versione» dell'Analisi si costruisce facendosi guidare dal Wizard.

L, un campo Logico che può essere S oppure N, ANNO, che può andare dal 1989 al 1993 (5 anni), CL, un campo di classificazione, da AA a HH (8 classi), IMPORTO, un campo numerico, PERC, un campo numerico percentuale.

Esistono quindi numerosi campi che potrebbero essere oggetto di operazioni di analisi. Ad esempio:

- conteggi per Sede, per Anno, per Classe,
- somme degli Importi per Sede, per Anno, per Classe,
- distribuzione percentuale per Sede e per Classe,
- ecc.

Anche il campo Importo potrebbe essere usato per eseguire dei raggruppamenti creando delle classi di importo, ad esempio da 1 a 2 milioni, da 2 a 3, ecc.

Analisi a una, a due e a più dimensioni

Una lista ha un aspetto tabellare, righe per record, colonne per campi. Una analisi condotta sui dati lista può essere monodirezionale, ad esempio la somma degli Importi per ciascuna Città. Se l'analisi è monodirezionale si possono facilmente inserire più calcoli per riga. Ad esempio conteggio delle persone, percentuale delle persone rispetto al totale, somma degli importi, medie degli importi per città, ecc.

Un'analisi può poi essere bidirezionale, nel senso che mostra due campi di aggregazione, uno nella prima colonna e uno nella prima riga. Il dato, all'incrocio tra riga e colonna, rappresenta il valore statistico (somma, conteggio, media che sia) per l'elemento della riga e quello per la colonna. Se a questa tabella si aggiungono anche i totali si raggiunge un livello informativo elevatissimo, perché si possono vedere nella stessa tabella, ad esempio, sia i dati per Città, che quelli per Sede, sia quelli incrociati, appunto, per quella Città e per quella Sede.

È ipotizzabile anche una terza, e perché no, ulteriori dimensioni. Il risultato dell'analisi sarà costituito da una serie di Tabelle, ad esempio una serie di Analisi Incrociate tra Sede e Classe, una per ciascun anno.

I prodotti che prevedono questo tipo di analisi permettono di scorrere le varie pagine al punto che il risultato può essere considerato una tabella 3D.

È facile prevedere le «misure» di una tabella di analisi che incroci più campi.

		MILANO								NAPOLI								ROMA								TORINO								Totale complessivo																										
		AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH																											
Totale N		16	20	19	36	28	24	27	28	16	20	19	36	28	24	27	28	16	20	19	36	28	24	27	28	16	20	19	36	28	24	27	28																											
Totale S		1989	3	0	3	0	0	0	0	0	1	7	1990	15	17	4	11	10	10	7	6	80	1991	1	10	16	7	6	8	5	13	66	1992	6	2	4	8	3	10	4	3	40	1993	2	3	7	5	6	3	5	0	31								
Totale complessivo		27	32	34	31	25	31	21	23	224	27	32	34	31	25	31	21	23	224	27	32	34	31	25	31	21	23	224	27	32	34	31	25	31	21	23	224																							

Figura 3 - MS Excel 5.0 - Variante di calcolo.

Se si analizza la zona del foglio nella quale è stata riversata l'analisi non si notano formule ma solo numeri. In altre parole la Tabella Pivot non può essere costruita scrivendo formule direttamente sul foglio, ma solo usando il generatore. Excel riconosce tuttavia che la zona è una tabella Pivot in quanto ne consente le modifiche, ad esempio consente di cambiare al volo gli elementi coinvolti.

		MILANO								NAPOLI								ROMA								TORINO								Totale complessivo							
		AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH	AA	BB	CC	DD	EE	FF	GG	HH								
Totale 1989		11,85%	24,63%	8,47%	36,62%	18,64%	100,00%	11,85%	24,63%	8,47%	36,62%	18,64%	100,00%	11,85%	24,63%	8,47%	36,62%	18,64%	100,00%	11,85%	24,63%	8,47%	36,62%	18,64%	100,00%																
Totale 1990		10,48%	22,64%	15,22%	30,57%	21,51%	100,00%	10,48%	22,64%	15,22%	30,57%	21,51%	100,00%	10,48%	22,64%	15,22%	30,57%	21,51%	100,00%	10,48%	22,64%	15,22%	30,57%	21,51%	100,00%																
Totale 1991		7,46%	18,43%	17,03%	25,45%	27,63%	100,00%	7,46%	18,43%	17,03%	25,45%	27,63%	100,00%	7,46%	18,43%	17,03%	25,45%	27,63%	100,00%	7,46%	18,43%	17,03%	25,45%	27,63%	100,00%																
Totale 1992		10,58%	10,51%	16,61%	34,82%	27,48%	100,00%	10,58%	10,51%	16,61%	34,82%	27,48%	100,00%	10,58%	10,51%	16,61%	34,82%	27,48%	100,00%	10,58%	10,51%	16,61%	34,82%	27,48%	100,00%																
Totale 1993		6,22%	20,34%	14,94%	27,97%	30,13%	100,00%	6,22%	20,34%	14,94%	27,97%	30,13%	100,00%	6,22%	20,34%	14,94%	27,97%	30,13%	100,00%	6,22%	20,34%	14,94%	27,97%	30,13%	100,00%																
Totale complessivo		7,23%	21,38%	13,22%	25,62%	30,13%	100,00%	7,23%	21,38%	13,22%	25,62%	30,13%	100,00%	7,23%	21,38%	13,22%	25,62%	30,13%	100,00%	7,23%	21,38%	13,22%	25,62%	30,13%	100,00%																

Figura 2 - MS Excel 5.0 - Variante 3D.

I campi di raggruppamento possono essere messi per riga, per colonna o per pagina. In pratica il campo (o i campi) per pagina, nel nostro caso il campo SEDE, si presenta come una Lista nella quale scegliere la Sede desiderata. La tabella presenta i dati relativi alla sola sede prescelta. Comoda la variante Tutto che produce il totale di tutte le Sedi.

COGNOME	SEDE	ANNO	CL	IMPORTO	PERC
ROSSI	ROMA	1989	BB	5.494,000	12%
ROSSI	ROMA	1990	BB	1.833,000	20%
ROSSI	ROMA	1991	BB	3.953,000	18%
ROSSI	ROMA	1992	BB	5.939,000	3%
ROSSI	ROMA	1993	BB	1.185,000	9%
ROSSI	ROMA	1989	BB	5.671,000	6%

Figura 4 - Lotus 123 versione 5.0 - XTab semplice.

L'ultima versione del '123 della Lotus, che nella beta (nella figura) si maschera ancora dietro un anonimo 4.1, dispone di due alternative, il Crosstab semplice, guidato da un Assistant molto intuitivo, e il CrossTab dinamico che in pratica «passa la pratica» al suo compagno Approach, che per l'occasione svolge compiti di OLE Server.

Ad esempio un calcolo per Sede e per Classe avrà 5 colonne (le Sedi) e 8 righe (le Classi), più un'eventuale colonna e un'eventuale riga di totali. Il calcolo può essere un conteggio delle ricorrenze (esempio: esistono 8 persone con sede Napoli e classe AA), oppure una somma, se si tratta di un campo numerico (esempio: il totale degli importi delle persone di Napoli e della Classe AA è pari a 34.500.000), oppure un calcolo percentuale (esempio: le persone di Napoli e della Classe AA, sono il 12% rispetto al totale della Sede Napoli e il 16% del totale della categoria AA).

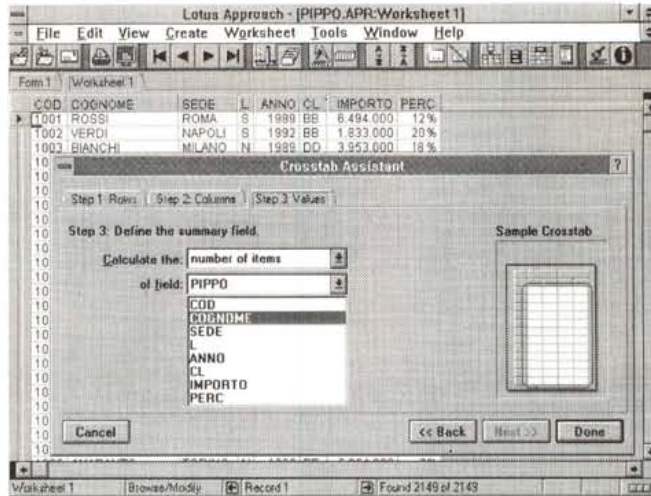


Figura 5 - Lotus Approach - L'Assistant di Approach contro il Wizard. La risposta della Lotus al Wizard della Microsoft è l'Assistant che vanta, rispetto all'avversario, una maggiore elasticità operativa, nel senso che è possibile passare, tramite le ormai tradizionali linguette, da un passo all'altro dell'operazione senza dovere rispettare obbligatoriamente una sequenza. I passi sono, manco a dirlo, tre e riguardano le colonne, le righe e il campo da calcolare.

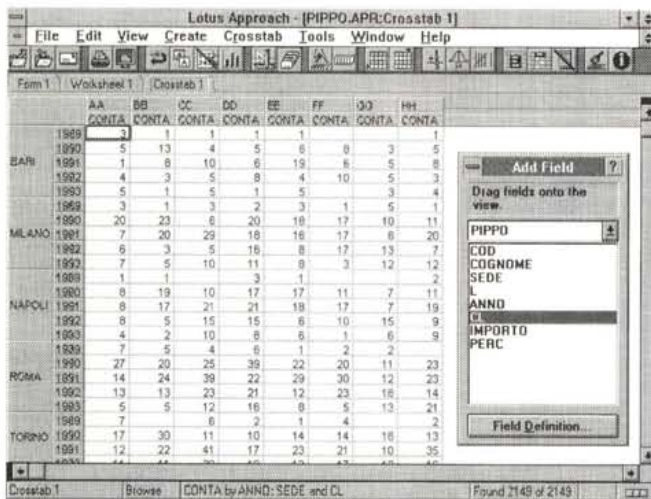
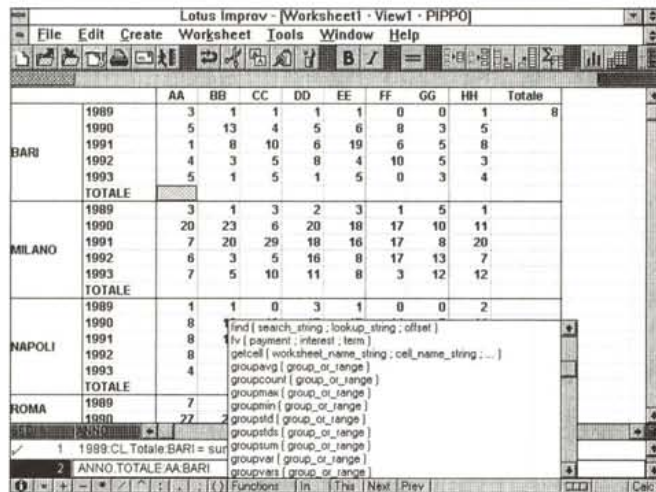


Figura 6 - Lotus Approach - CrossTab dinamico. Ecco qui l'Approach versione 3.0, che dispone di un XTab dinamico. Vediamo con l'occasione come le varie viste sui dati, anche la Crosstab è una vista, possano essere individuate dalla linguetta. Vediamo anche come sia possibile dinamicamente aggiungere e/o spostare, con un semplice Drag and Drop, i campi oggetto dell'analisi.

Figura 7 - Lotus Improv - il pioniere. L'Analisi Incrociata con tutte le sue possibili varianti costituisce uno strumento completo per qualsiasi tipo di analisi dati. Al punto che la Lotus attorno a questa funzionalità realizzò un paio di anni fa un prodotto, Improv, che ebbe un ottimo successo di critica, ma poco successo presso il pubblico che lo ha un po' confuso con uno spreadsheet. Ora che il XTab dinamico è presente in tutti i prodotti il problema, e forse anche l'Improv, è superato.



Dapprima constateremo il fatto che ormai lo strumento operativo che esegue l'analisi incrociata c'è in tutti i prodotti più recenti.

Poi faremo una capatina in un linguaggio di programmazione, abbiamo scelto quello più di moda, il Visual Basic, per verificare cosa possa costare, in termini di programmazione, la realizzazione di una tabella di analisi incrociata.

Successivamente verificheremo che le stesse cose che si fanno con lo strumento Analisi Incrociata si possono ottenere usando strumenti operativi presenti da sempre negli spreadsheet. L'utente evoluto deve essere in grado, a seconda dei casi, di usare lo strumento automatico, che però va preso per quello che può dare, oppure di raggiungere gli stessi risultati sfruttando una modalità più manuale ma più adattabile ad eventuali esigenze particolari.

Tabella Pivot in casa Microsoft e CrossTab dinamico in casa Lotus

Le ultimissime novità in materia di Analisi Incrociata sono quelle proposte da Excel 5.0 della Microsoft e da 123 per Windows 5.0 della Lotus.

Il primo dispone di una nuova funzionalità che si chiama Tabella Pivot, che non solo crea, sfruttando un Wizard che accompagna l'utente passo passo nella scelta delle varie impostazioni, una tabella di Analisi Incrociata, ma permette, sul risultato finale, una manipolazione dinamica, semplicemente spostando con degli intuitivi Drag and Drop i campi da una parte all'altra.

In figura 1 una fase dell'Autocomposizione che mostra come sia possibile ed intuitivo posizionare i campi coinvolti nell'operazione.

Nella 2 vediamo invece come si presenta un'Analisi 3D.

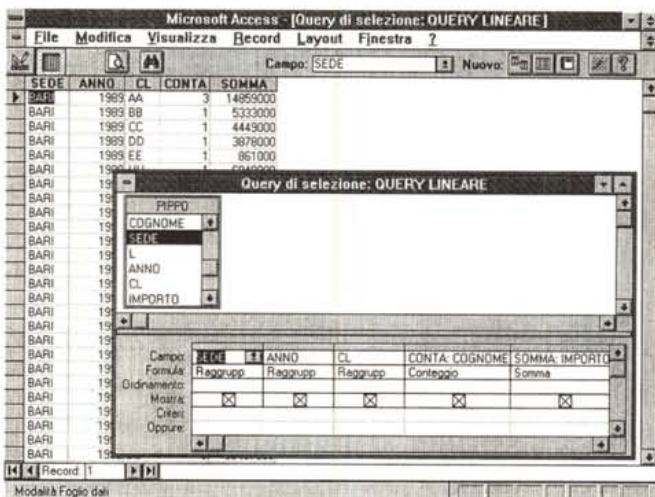
In figura 3 verifichiamo il fatto che la generazione della Tabella Pivot non pro-

In altre parole va ben distinta l'organizzazione dell'analisi, e quindi i campi su cui vengono eseguiti i vari raggruppamenti, dai campi sui cui vengono eseguiti i calcoli e dal tipo di calcolo

(conteggio, somma, media, percentuale rispetto al totale di riga e/o di colonna, ecc.) desiderato.

Vi proponiamo ora una serie di esperimenti di vario tipo.

Figura 8 - MS Access - Query Lineare.
Gli stessi dati coinvolti in un'analisi incrociata possono essere visti in un'analisi non incrociata, che potremo definire lineare. Vediamo come in MS Access si possa creare una Query che totalizza per SEDE, per ANNO, per CLASSE, e che conta le ricorrenze e somma gli Importi. In questo caso la lettura dei dati può avvenire da una sola direzione.



duce «formule», cosa che potrebbe consentire all'utente evoluto di comporre l'Analisi direttamente sul foglio, magari per impostare dei calcoli non previsti dall'Autocomposizione (si pensi per esempio ad una Somma degli Importi divisa per 1.000.000).

La zona con i dati è comunque individuata da Excel che solo nelle celle di questa zona rende disponibili i comandi di modifica delle caratteristiche della Tabella Pivot.

La Lotus, dal suo canto, nel suo 123 ha da tempo inserito un semplice XTab, statico, e che accetta un solo campo per riga e uno solo per colonna. Lo vediamo in figura 4.

Pareggia il conto con la Microsoft con il CrossTab dinamico inserito nella nuova versione di Approach, la 3.0, e lanciabile direttamente dall'123 versione 5.0, come se fosse una propria funzionalità. Il livello qualitativo è analogo a quello dell'Excel anche tenendo conto del fatto che viene utilizzato un Assistente, più servile, per il fatto che non obbliga ad una rigida sequenza di operazioni, del Mago di Excel (fig. 5). Anche in questo caso abbiamo una totale dinamicità. Si possono inserire al volo nuovi campi da coinvolgere nell'analisi (fig. 6).

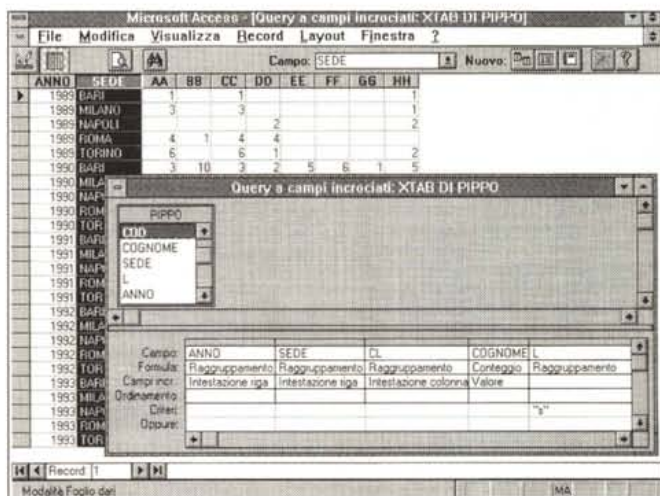
Lotus Improv e MS Access

Uno strumento che consenta di eseguire un'Analisi Incrociata, con tutte le sue possibili varianti (più campi per riga, per colonna, per pagina, più tipi di calcolo, inserimento dei totali di gruppo e sottogruppo, ecc.) costituisce uno strumento sufficiente per eseguire qualsiasi tipo di analisi dati, di qualsiasi dimensione, lineare, bidimensionale e quindi tabellare, tridimensionale, quadrimensionale, ed oltre.

Questo è tanto vero che la Lotus, attorno a questa funzionalità, realizzò un paio di anni fa l'Improv (fig. 7), che per

Figura 9 - MS Access - Query Campi Incrociati.
MS Access affida la realizzazione di un'Analisi Incrociata ad un particolare tipo di Query, che poi viene memorizzata come istruzione SQL.

Non si tratta di un XTab dinamico, ma essendo facile modificare, nell'ambiente GObE, i vari campi coinvolti, è come se lo fosse. Possono essere indicati più campi per riga, un solo campo per colonna e un solo campo di calcolo. Possono essere inseriti altri criteri di selezione, ma non possono essere inseriti calcoli di totalizzazione e di percentualizzazione rispetto ai totali.



questa sua intuizione geniale ebbe un ottimo successo da parte della critica, che fu colpita soprattutto dal fatto che l'Analisi era dinamica, e questa possibilità è a dir poco «spettacolare». Ebbe invece poco successo presso il pubblico che lo ha un po' confuso con uno spreadsheet. L'idea però piacque al punto che successivamente la stessa funzionalità è stata inserita un po' ovunque.

Va detto comunque, ad onore dell'Improv, che pur essendo basato sul concetto dell'Analisi Incrociata disponeva di altre funzionalità alcune delle quali molto innovative, come quella che consentiva di scrivere le formule in un linguaggio «in chiaro».

È però indubbio che ormai, essendo il XTab dinamico presente in tutti i prodotti, l'Improv è superato, cosa che succede a molti prodotti pionieri, che aprono la strada, rischiando in proprio, e sono copiati e superati dagli altri.

A dimostrazione del fatto che l'Anali-

si XTab sia ormai una funzionalità pressoché obbligatoria non solo nei prodotti a base spreadsheet, vediamo come sia stata introdotta in un prodotto DBMS, l'Access della Microsoft.

MS Access dispone dello strumento operativo Query che permette di costruire, in un ambiente Query by Example grafico, delle «interrogazioni», che possono servire a selezionare, ordinare, raggruppare, aggiornare, cancellare, ecc. i dati delle varie tabelle presenti nell'applicazione.

È ad esempio molto facile realizzare delle analisi monodimensionali come

quella mostrata, e commentata, in figura 8.

Esistono vari tipi di Query, tra queste una che si chiama Campi Incrociati, che non è dinamica, ma la cosa non ha molta importanza data la facilità con cui, nell'ambiente QbE grafico, è facile modificare i campi in gioco (fig. 9).

È possibile scegliere più campi di raggruppamento per riga, è possibile però scegliere un solo campo per colonna e un solo campo su cui eseguire i calcoli. Non sono possibili operazioni di totalizzazione e di percentualizzazione.

Però tenendo conto del fatto che è possibile inserire ulteriori campi sui quali impostare «regole» o «calcoli» e soprattutto del fatto che è possibile organizzare una catena di Query, molti dei limiti di questo tipo di Query possono essere superati dall'utente più smaliziato.

Una pausa di riflessione

Tutte le funzionalità viste fino ad ora

sono in grado di risolvere da sole problematiche applicative complesse. Per ottenere dei risultati analoghi, usando strumenti di programmazione tradizionale, occorrerebbero parecchie ore di lavoro.

Una volta capito a cosa servono e una volta imparato ad utilizzarle correttamente, vediamo di analizzarne i vantaggi e gli svantaggi.

I vantaggi sono l'immediatezza nell'uso e il fatto che sono strumenti alla portata di un utente anche poco evoluto. Gli svantaggi consistono nel fatto che a volte capita di dover eseguire dei calcoli non permessi dallo strumento.

Per esempio le somme di importi in lire, eseguite per categoria, spesso è comodo vederle divise per 1.000.000, non fosse altro che per limitare la di-

Calcolo e Caricamento in Griglia										
	AA	BB	CC	DD	EE	FF	GG	HH	TTOT	
BARI	1989	3	1	1	1	1	1	1	1	8
	1990	5	13	4	5	6	8	3	5	48
	1991	1	8	10	6	19	6	5	8	63
	1992	4	3	5	8	4	10	5	3	42
	1993	5	1	5	1	5	3	4	24	
		18	26	25	21	35	24	16	21	186
MILANO	1989	3	1	3	2	3	1	5	1	19
	1990	20	23	5	20	18	17	10	11	125
	1991	7	20	29	18	16	17	9	20	135
	1992	6	3	5	16	8	17	13	7	75
	1993	7	5	10	11	8	3	12	12	68
		43	52	53	67	53	55	48	51	422
NAPOLI	1989	1	1	3	1	1	1	1	2	8
	1990	8	19	10	17	17	11	7	11	100
	1991	9	17	21	21	18	17	7	19	128
	1992	8	5	15	15	6	10	15	9	83
	1993	4	2	10	8	6	1	5	9	46
		28	44	56	64	48	39	35	50	365
ROMA	1989	7	5	4	6	1	2	2	2	27
	1990	27	20	25	39	22	20	11	23	187
	1991	14	24	39	22	29	30	12	23	193
	1992	13	13	23	21	12	23	18	14	137
	1993	5	5	12	16	8	5	13	21	85
		66	67	103	104	72	80	56	81	628
TORINO	1989	7	5	6	2	1	4	2	2	22
	1990	17	30	11	10	14	14	18	13	127
	1991	12	22	41	17	23	21	10	35	181
	1992	11	11	29	10	12	17	12	16	118
	1993	7	8	12	13	8	12	15	24	99
		54	71	99	52	58	68	55	90	547
		210	260	336	308	266	258	210	293	2149

Figura 10 - MS Visual Basic 3.0 - Tutto a mano.

Lo scopo di questo esercizio è quello di valutare l'impegno necessario per realizzare un'analisi incrociata con uno strumento di programmazione tradizionale. Usiamo il Visual Basic con il quale abbiamo creato una semplice Form con una semplice griglia. I vari programmi e sottoprogrammi servono per riempire la griglia di dati.

mensione della tabella in uscita.

Oppure in caso di categorizzazione su campi numerici spesso è necessario

creare degli intervalli di valori. Ad esempio contare quanti sono i record il cui importo è compreso tra 1 e 2 milioni, quanti tra 2 e 3 milioni, e così via.

In questo caso occorre preparare i dati, ad esempio si può facilmente creare un campo calcolato ottenuto come valore intero dell'importo diviso per 1.000.000 e in tal modo il valore ottenuto corrisponde ad una classe di appartenenza. 1 vuole dire importo che va da 1 milione compreso a 2 milioni escluso, ecc. Oppure si può manipolare ulterio-

mente la tabella ottenuta, ad esempio dividendo tutti i valori risultanti, per 1.000.000.

Tutte queste possono essere soluzioni accettabili.

Noi riteniamo che un utilizzatore esperto debba trovare strade alternative. La prima è quella di raggiungere gli stessi risultati per altre strade, meno automatizzate ma più controllabili, la seconda può essere quella di ricorrere alla programmazione, non fosse altro che per porsi gli stessi problemi che si sono posti gli sviluppatori che hanno creato gli strumenti di XTab.

Usiamo il Visual Basic

Cominciamo proprio da qui, ipotizzando una serie di semplificazioni che in programma di XTab completo non ci sono:

- l'analisi che vogliamo condurre è fissa, già sono impostati rigidamente i campi di categorizzazione e il campo di calcolo,
- le varie categorie, SEDI, ANNI e CLASSI, sono dati del problema, non debbono essere calcolati dinamicamente. Il problema della creazione dinamica delle categorie (per intenderci il numero e il nome delle Sedi viene letto dall'archivio) richiede due «passaggi» sui dati, il primo per creare le categorie, il secondo per eseguire i conteggi per categoria.

La nostra applicazione semplificata consiste in una Form che contiene una Griglia già dimensionata per contenere SEDI, ANNI, CLASSI e Totali, per Anni, per Classe e Totali Generali. La dimensione della griglia, che vediamo in figura 10, è quindi di 33 righe per 11 colonne.

Il programma si svolge in un'unica sequenza e comporta una serie di momenti eseguiti al verificarsi dell'evento Sub Form Load, ovvero al lancio dell'applicazione (tutti i programmi e i sottoprogrammi li vediamo in figura 11):

```
Dim Db As Database, Qd As Dynaset, Sn As Snapshot
Dim S$(5), A(5), G$(9), I, J, K, L, SS, AA, GG

Sub CARICA ()
G$(1) = "AA": G$(2) = "BB": G$(3) = "CC": G$(4) = "DD"
G$(5) = "EE": G$(6) = "FF": G$(7) = "GG": G$(8) = "HH": G$(9) = "TOT."
S$(1) = "BARI": S$(2) = "MILANO": S$(3) = "NAPOLI"
S$(4) = "ROMA": S$(5) = "TORINO"
A(1) = 1989: A(2) = 1990: A(3) = 1991: A(4) = 1992: A(5) = 1993
End Sub

Sub Form_Load ()
CARICA ' caricamento dei tre vettori
Set Db = OpenDatabase("d:\db3", False, False, "dBASE III;")
Q$ = "Select pippo.cognome,pippo.sede,pippo.cl,pippo.anno from pippo"
Set Qd = Db.CreateDynaset(Q$)
TITOLI ' preparazione della griglia
While Not Qd.EOF
T1.Text = Qd("Cognome"): T2.Text = Qd("Sede") ' Output
T3.Text = Qd("Anno"): T4.Text = Qd("Cl")
RICERCA ' ricerca della posizione nella griglia
T5.Text = SS: T6.Text = AA: T7.Text = GG ' Output
G1.Col = GG + 1: G1.Row = SS * 6 + AA - 6
G1.Text = Val(G1.Text) + 1
Qd.MoveNext
Wend
TOTALI ' calcolo e posizionamento dei totali
End Sub

Sub RICERCA ()
For I = 1 To 5: If S$(I) = T2.Text Then SS = I ' ricerca la Sede
Next I
For I = 1 To 5: If A(I) = Val(T3.Text) Then AA = I ' ricerca l'Anno
Next I
For I = 1 To 8: If G$(I) = T4.Text Then GG = I ' ricerca la Classe
Next I
End Sub

Sub TITOLI ()
G1.Row = 0: For I = 2 To 10: G1.Col = I: G1.Text = G$(I - 1)
G1.ColAlignment(I) = 1: Next I
G1.Col = 1: For L = 0 To 4: For I = 1 To 5
G1.Row = I + L * 6: G1.Text = A(I): Next I: Next L
G1.Col = 0: For L = 1 To 5: G1.Row = L * 6 - 5: G1.Text = S$(L): Next L
For I = 1 To 30: G1.RowHeight(I) = 200: Next I
End Sub

Sub TOTALI ()
For I = 1 To 5: For J = 2 To 9: TT = 0: For K = 1 To 5
RR = I * 6 + K - 6: G1.Row = RR: G1.Col = J: TT = TT + Val(G1.Text)
Next K: G1.Row = RR + 1: G1.Text = TT: Next J: Next I
For J = 2 To 9: TT = 0: For I = 6 To 31 Step 6
G1.Row = I: G1.Col = J: TT = TT + Val(G1.Text): Next I
G1.Row = 31: G1.Text = TT: Next J
For I = 1 To 31: TT = 0: For J = 2 To 9
G1.Row = I: G1.Col = J: TT = TT + Val(G1.Text): Next J
G1.Col = 10: G1.Text = TT: Next I
End Sub
```

Figura 11 - MS Visual Basic 3.0 - I nostri programmi.

La relativa esiguità dei programmi dipende dal fatto che le dimensioni della griglia, o meglio il numero di Sedi, di Anni e di Classi rispetto ai quali calcolare i nostri dati sono fissi. Ben altro impegno avrebbe comportato un programma in cui il riempimento dei titoli e il dimensionamento della griglia fosse «calcolato» leggendo e contando i dati della Tabella Dati.

Figura 12 - Quattro Pro SE - Cosa Se a 2 variabili.

La funzionalità fondamentale su cui si poggiano anche i più moderni strumenti di CrossTab è la Dati Tabella, altrimenti detta, in certi prodotti, What If, ovvero Cosa Se. Occorre scrivere due serie di valori, una nella prima colonna della tabella e una nella prima riga. Poi occorre scrivere una formula che contiene due variabili (in pratica il riferimento a due celle) che devono assumere via via i valori delle due serie di variabili. Il risultato è una tabella di valori in cui il singolo elemento è il risultato dell'applicazione di quella formula ai due valori, quello della riga e quello della colonna.

COD	COGNOME	SEDE	L	ANNO	CL	IMPORTO	PERC
1001	ROSSI	ROMA	S	1989	BB	6.494.000	12%
1002	VERDI	NAPOLI	S	1992	BB	1.833.000	20%
1003	BIANCHI	MILANO	N	1989	DD	3.953.000	16%
1004	FUCSIA	BARI	N	1993	HH	5.939.000	
1005	GIALLI	ROMA	S	1989	CC	1.185.000	
1006	MARRONI	NAPOLI	S	1992	GG	5.671.000	
1007	ROSA	ROMA	N	1992	BB	2.820.000	
1008	NERI	TORINO	S	1993	HH	5.939.000	
1009	ROSSI	ROMA	S	1989	BB	6.494.000	12%
1010	TURCHESI	TORINO	S	1993	HH	5.939.000	

COD	COGNOME	SEDE	L	ANNO	CL	IMPORTO	PERC
1001	ROSSI	ROMA	S	1989	BB	6.494.000	12%
1002	VERDI	NAPOLI	S	1992	BB	1.833.000	20%
1003	BIANCHI	MILANO	N	1989	DD	3.953.000	16%
1004	FUCSIA	BARI	N	1993	HH	5.939.000	
1005	GIALLI	ROMA	S	1989	CC	1.185.000	
1006	MARRONI	NAPOLI	S	1992	GG	5.671.000	
1007	ROSA	ROMA	N	1992	BB	2.820.000	
1008	NERI	TORINO	S	1993	HH	5.939.000	
1009	ROSSI	ROMA	S	1989	BB	6.494.000	12%
1010	TURCHESI	TORINO	S	1993	HH	5.939.000	

Figura 13 - MS Excel - Un po' di analisi.

A parte l'Analisi Incrociata sono tante altre le possibili necessità di chi deve tener sotto controllo dei dati continuamente modificati. Vediamo delle semplici Analisi lineari costruite per ciascuno dei tre campi oggetto di osservazione, Sedi, Anni e Classe, e una bidimensionale, tra Classi e Anni. L'ultima riga di tale tabella, che rappresenta i totali per classe, deve corrispondere all'analisi semplice per classe. Idem per l'ultima colonna che rappresenta il totale per anno.

- dimensionamento dei Vettori e delle Variabili,
- routine Carica, per definire le tre serie di categorie,
- apertura del Database, in formato DBF, e della sola tabella PIPPO,
- esecuzione di un comando SQL che definisce uno «snapshot»,
- routine Titoli, che inserisce nei punti giusti della griglia, i titoli,
- istruzione While Not Qd.Eof, che serve per gestire lo scorrimento di tutti i record dello snapshot Qd,
- routine Ricerca, fondamentale, che serve per localizzare le tre categorie. Vengono scorse una per una fino a quando si trova che il campo del record attivo corrisponde a quella data categoria,
- istruzioni Qd.MoveNext e Wend, per gestire l'esecuzione del Loop, ovvero il passaggio al record seguente e il ritorno all'istruzione While,
- routine Totali, che calcola e posiziona i totali di vario tipo al punto giusto. Due considerazioni:

- nelle varie routine sono presenti numerosi cicli For Next, per scorrere le varie categorie. Questi cicli comprendono piccoli calcoli che servono ad individuare, dato l'indice del ciclo, l'esatta riga o colonna,
- esistono delle istruzioni che servono per ripetere alcuni valori in apposite Caselle di Testo. Non servono a nulla, ma possono servire in fase iniziale di test, per vedere se le varie routine stanno funzionando.

Per chi sa programmare sono istruzioni molto in linea con tutti gli altri linguaggi. Chi non sa programmare... dovrà abituarci. Le istruzioni usate sono abbastanza intuitive.

Le basi, ad esempio con Borland Quattro Pro SE

Per studiare le funzionalità di base del foglio elettronico, che vengono usate dai vari «servocomandi» che costruiscono le analisi di XTab, usiamo il nostro Quattro Pro Se, che continua ad

essere un ottimo strumento per chi voglia imparare «bene» ad usare un foglio elettronico standard. Il materiale didattico usato nel corso in dotazione al Quattro Pro SE è stato sviluppato da MC.

Tutti i fogli elettronici dispongono di una funzionalità per il calcolo della distribuzione statistica.

Occorrono due serie di valori, quelli, tanti, oggetto di analisi, i nostri duemila e rotti importi, e quelli, pochi, rispetto ai quali raggruppare. Fino a 500.000, fino a 1.000.000, ecc. Il comando produce una colonna di valori che indicano le ricorrenze, ovvero quanti record ricadono in quell'intervallo. La successiva colonna, quella con le ripartizioni percentuali, l'abbiamo messa successivamente a completamento dell'analisi.

L'esempio di figura 12 mostra un'analisi XTab fatta con strumenti operativi più tradizionali. L'analisi incrocia le Sedi con le Classi, limitatamente ai dati dell'anno 1991.

Una tabella di questo genere si ottiene sfruttando la sinergia tra la funzionalità Cosa Se, che esegue il calcolo di una stessa formula funzione di due variabili per tutta una tabella, e una funzione @ che esegue un calcolo statistico di database.

In pratica si sfrutta il dinamismo della funzione Cosa Se (Dati Tabella in altri prodotti) che sposta dinamicamente i valori della prima riga e i valori della prima colonna in due celle di appoggio, richieste dal comando. Se le celle di appoggio (che vengono alimentate via via in modo trasparente, in quanto appare direttamente il risultato completo sull'intera tabella) sono le stesse celle del Blocco dei Criteri, usato dalla Cosa Se, si ottiene un calcolo di Analisi Incrociata.

Nel nostro caso il Blocco dei Criteri contiene i due campi di nostro interesse, SEDE e CLASSE, ed un altro campo, l'ANNO, che usiamo per impostare un ulteriore Criterio. In pratica Cosa Se calcola la distribuzione tra SEDI e CLASSI per quell'anno.

A questo punto si potrebbe cambiare la formula del Cosa Se, posta in corrispondenza dell'incrocio tra prima riga e prima colonna. Ad esempio la formula:

```
@SOMMADB((PIPPO)A1..H2150;6;A1..C2)/1000000
```

in cui:
[PIPPO], è il foglio con i Dati
A1..H2150, è il blocco con i dati,
6, è la colonna degli Importi, la settima del blocco dei dati,
A1..C2, è il blocco dei criteri
/1000000, per vedere l'analisi espressa in milioni e non in ingombranti lire.

Non è facile, ma è un punto impor-

tantissimo per rendersi autonomi e per riuscire a trovare soluzioni «geniali» per problemi complessi.

Provate e riprovate fino a quando non saprete costruire un'analisi di questo genere senza nessuna incertezza.

Per chiudere e concludere

Il comando Dati Tabella, o Cosa Se, in realtà presenta due possibili varianti. Dati Tabella ad 1 variabile, e a 2 variabili. Con quello a due variabili si eseguono le Analisi Incrociate ora viste, occorre riempire una riga di valori X, una colonna di valori Y e occorre nella cella di incrocio, in alto a sinistra, scrivere la formula. Nella versione a 1 variabile si usa una sola colonna di valori (si potrebbe anche sviluppare in orizzontale ed allora sarebbe una riga) e poi si scrivono le formule, anche più di una, nella riga in alto a destra rispetto ai valori in colonna.

In figura 13 mostriamo un'applicazione Excel che analizza, in tre Tabelle a 1 variabile e in una a 2 variabili, i nostri dati. È chiaro che, una volta impostate le formule e impostati i quattro Dati Tabella, il foglio con l'analisi, che si chiama Pluto, viene aggiornato ogni volta che viene aggiornata la Tabella Dati, che è il foglio Pippo. Anzi in genere si lavora liberamente sulla Tabella Dati, e poi al semplice caricamento del foglio di analisi, le varie Dati Tabella vengono immediatamente ricalcolate.

Nell'esercizio di figura 14, stavolta realizzato con Lotus 123, vi presentiamo un'interessante variante. Supponiamo che uno dei campi sui quali vogliamo classificare sia l'Importo. Vogliamo delle classi di importo, ad esempio da 1 a 2 milioni, da 2 a 3 milioni, e così via.

Il processo è simile a quelli di prima solo che, come evidente nella figura, in cui abbiamo a fianco alle celle con le formule scritto a mo' di documentazione le formule stesse, il criterio non conterrà valori secchi ma conterrà delle espressioni logiche, tra la prima cella dei dati, che è quella che serve per eseguire dei test del tipo «maggiore di», e che nel nostro caso è la G2, e una fantomatica cella J2 che sembrerebbe non entrarci per nulla. Invece la cella J2 è quella d'appoggio per i valori di colonna del comando Dati Tabella. Insomma la funzionalità Dati Tabella legge via via i valori di colonna, li posiziona nella cella di appoggio J2, nel criterio viene subito eseguito il confronto tra G2 (convenzionale) e il valore in transito in J2, per tutti gli importi della Tabella Dati. La formula del Dati Tabella (@COUNT ..) esegue il conteggio statistico delle righe che soddisfano il criterio e quindi riporta il risultato

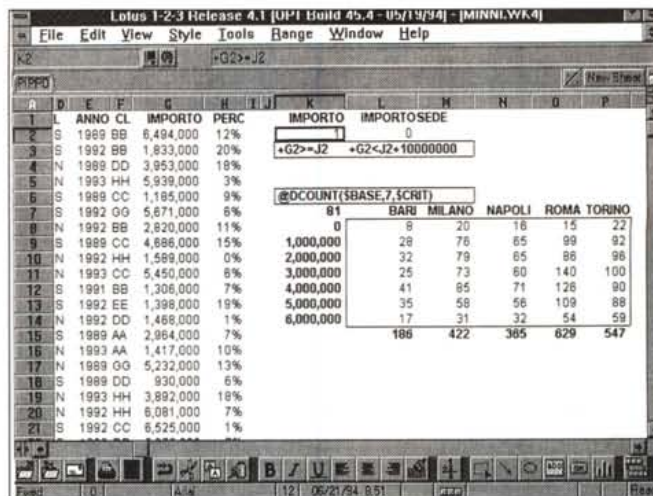
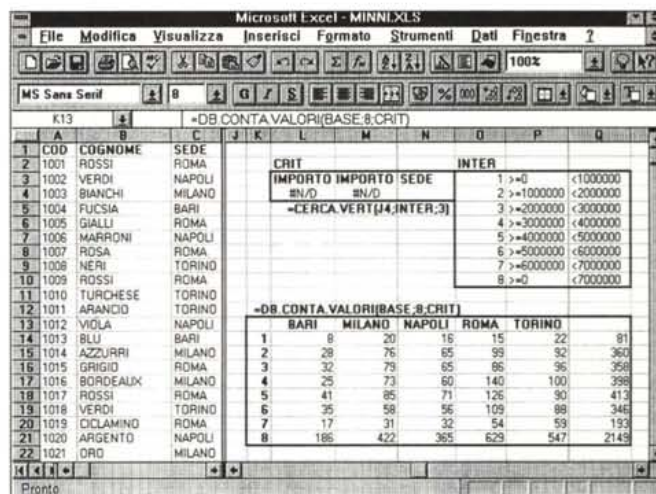


Figura 15 - MS Excel 5.0 - A tutta velocità.

In questo caso la variabile in colonna è un indice che viene usato per pescare dinamicamente dalla zona INTER via via i vari Criteri, che vengono posti nella zona CRIT. In pratica vengono abilmente e spregiudicatamente messe in sinergia tre funzionalità potentissime del foglio elettronico: il comando Dati Tabella, la funzione di ricerca verticale, la funzione statistica di database.



tato nella cella corrispondente a quel valore di Importo e a quella Sede.

Visto che ci siamo facciamo qualcosa di ancora più complicato, documentato in figura 15, l'ultima, e realizzato con Excel, in cui i vari Dati Tabella sono matrici e quindi sono continuamente ricalcolati, al contrario del '123 e del Quattro Pro in cui gli stessi comandi vanno lanciati.

Nella prima colonna abbiamo inserito una serie di indici, una semplice sequenza di numeri. Dati Tabella legge l'indice, lo pone nella solita cella che sembra non entrarci nulla, in questo caso la J4.

Sulla base di questo valore esegue, nelle celle dei Criteri, una serie di formule di ricerca verticale che servono per pescare nella zona Inter il criterio corrispondente all'indice (ad esempio il criterio corrispondente al valore 3 è >=2000000 e <3000000). E poi tutto funziona come prima.

In questo modo si sfruttano:

- le formule statistiche di database, che eseguono conteggi sulla base di criteri,
- la funzionalità Dati Tabella, che calcola

Figura 14 - Lotus 123 5.0 - Variazioni sul tema.

In questa variante costruiamo una Tabella a due variabili che lavora sulle Sedi, in alto, e sugli Importi a sinistra. La cella di appoggio utilizzata è la J2. Questa è utilizzata nelle due celle dei Criteri, la prima che esegue il test sull'inizio dell'intervallo (+G2>=J2), la seconda che lo esegue sulla fine dell'intervallo (+G2<J2+1000000). In questo caso si fruttifica il dinamismo della funzionalità Dati Tabella, per impostare dinamicamente gli intervalli.

la stessa formula in un'intera tabella nella quale sia stata impostata una riga di valori e una colonna di valori,

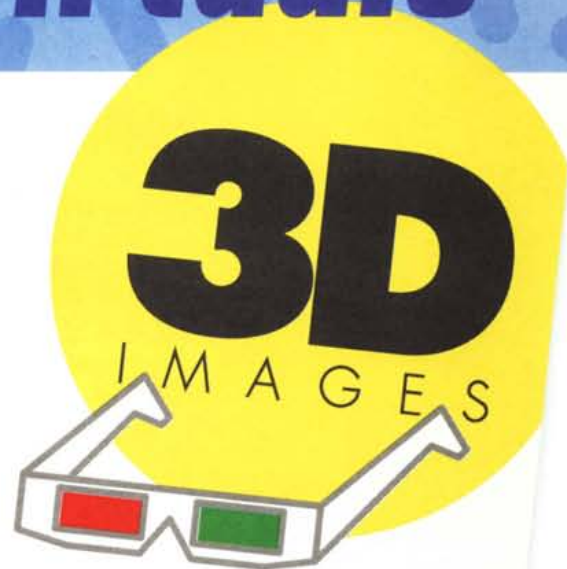
- la formula, o meglio la funzione, di ricerca verticale con la quale vengono pescati dinamicamente i criteri da una tabella, esterna e facilmente modificabile, di criteri.

La cosa in più, che complica talmente l'applicazione al punto che temo di essermi perso la gran parte dei lettori, consiste nel fatto che con questo sistema si possono gestire elenchi di criteri, più facilmente modificabili rispetto a criteri inseriti direttamente nel meccanismo di calcolo.

Ci fermiamo qui.

Con questo ultimo esercizio abbiamo dimostrato che un utente, mi sia consentito, particolarmente esperto, riesce con comandi tradizionali, gli ultimi usati ci sono da sempre, a realizzare le stesse analisi ottenibili con le varie funzionalità di XTab. Anzi può realizzare analisi anche più complesse e soprattutto può controllarne le formule direttamente sul foglio.

4 Passi Nel Virtuale



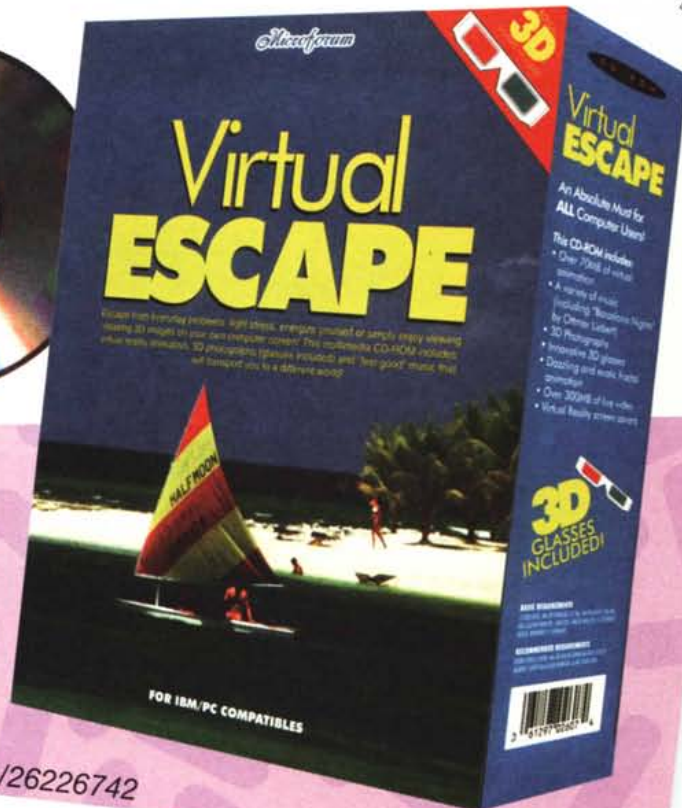
Virtual ESCAPE

ADVANCE - ROMA

UNA BELLA PASSEGGIATA...

...tra virtual animation, foto 3D, fractal animation, video ed un programma giornaliero personalizzabile di esercizi fisici.

È il nuovo "rilassante" CD-ROM dalla Microforum, in vendita nei migliori computer shop



Microforum
ORIGINAL SOFTWARE CD

Ufficio di rappresentanza per l'Italia: Tel. 02/22473137 - Fax 02/26226742