COMPUTER GRAFICA APPLICATA

Prima parte

Cominciamo da questo numero una serie di articoli dedicati alla computer grafica applicata.

Tratteremo la vasta problematica connessa con lo studio e la realizzazione di programmi con visualizzazione grafica di dati provenienti da calcoli di qualsiasi tipo, quindi non programmi esclusivamente grafici, ma programmi di matematica, trigonometria, statistica nei quali i dati risultanti sono riprodotti in forma grafica.

Come nostra abitudine affronteremo i vari argomenti dal punto di vista pratico cercando di esaminare, capire e risolvere insieme i vari problemi che ci si presentano e quindi realizzando passo passo i programmi.

L'argomento che esamineremo in questo numero è un tipico problema di statistica: le curve di regressione. Cercheremo di trattarlo in maniera elementare per renderlo comprensibile a tutti. Nel numero seguente parleremo di trigonometria, poi si vedrà: dipenderà anche dalle vostre segnalazioni.

Tutti più o meno conoscono la statistica, sia pure solamente attraverso la storiella del pollo o del mezzo pollo.

Uno degli argomenti più interessanti della statistica, o meglio della matematica statistica, è lo studio delle curve di regressione. Introdurremo il concetto di regressione nella maniera più elementare e quindi più funzionale rispetto alla trattazione strettamente pratica che vogliamo farne.

Rimandiamo i lettori che vogliono approfondire l'argomento, anche dal punto di vista teorico, alla consultazione dei numerosissimi testi di statistica (ne esistono anche a livello elementare) nei quali i capitoli riguardanti le curve di regressione sono sicuramente presenti.

Cosa sono le Curve di Regressione

La curva di regressione è quella funzione matematica che meglio approssima i valori dei dati rilevati con osservazioni sperimentali su un dato fenomeno. Ovvero quando si eseguono delle osservazioni di tipo statistico e quindi si hanno a disposizione dei valori numerici, si cerca una legge matematica tra le varie grandezze osservate. Una volta trovata questa legge (ma non é detto che ci sia) sarà possibile valutare, tramite la interpolazione statistica, nel modo più verosimile possibile anche altri dati, senza dover eseguire direttamente (anche perché in certi casi non si può) nuove misure.

Per non complicare ulteriormente il discorso, quindi anche nel tentativo di non perdere i lettori arrivati sin qui, ci limiteremo al caso in cui il fenomeno osservato metta in relazione due variabili.

La prima si chiama variabile indipendente, la seconda, poiché dipende dalla prima, variabile dipendente.

A questo punto facciamo un esempio. Abbiamo realizzato un programma di ordinamento alfabetico con il metodo HEAP-SORT (figura 1) e vogliamo valu-

```
10  REM   METODO HEAPSORT
20  HOME  : INPUT " NUMERO DATI ";N
30  DIM A$(N): FOR H = 1 TO N
40 G =  INT ( RND (1) * 4 + 4)
50  FOR K = 1 TO G
60 A$(H) = A$(H) +  CHR$ ( RND (1) * 26 + 65)
70  NEXT K: NEXT H: PRINT : GOSUB 500
100  REM  ORDINAMENTO
101  PRINT  CHR$ (7): REM  INIZIO
110 L =  INT (N / 2) + 1:M = N
120  IF L > 1 THEN L = L - 1:B$ = A$(L): GOTO 150
130 B$ = A$(M):A$(M) = A$(1):M = M - 1
140  IF M = 1 THEN A$(1) = B$: GOTO 400
150 J = L
160 I = J:J = 2 * J
170  IF J > M THEN A$(I) = B$: GOTO 120
180  IF J < M THEN  IF A$(J) < A$(J + 1) THEN J = J + 1
190  IF B$ < A$(J) THEN A$(I) = A$(J): GOTO 160
200 A$(I) = B$: GOTO 120
400  PRINT  CHR$ (7): GOSUB 500: END
500  FOR H = 1 TO N: PRINT A$(H): NEXT : RETURN
```

Figura 1 — Listing Programma Heapsort. È uno degli algoritmi più classici e più rapidi di ordinamento.

```
100  REM  INIZIALIZZAZIONE DEL PROGRAMMA
110 E = 2.71828183:L$ = "-----------------------------------------"
120  HOME : PRINT L$: PRINT "      REGRESSIONE  ESPONENZIALE"
130 N = 4: DIM X(N),Y(N): REM  NUMERO COPPIE VALORI
140  FOR I = 1 TO N: READ X(I): NEXT I: FOR I = 1 TO N: READ Y(I): NEXT I
150  PRINT "        ";N;"  COPPIE DI NUMERI  ": PRINT L$
160  PRINT "     VAR.INDIP. X(I)  VAR.DIP. Y(I)": PRINT
170  FOR I = 1 TO N: PRINT  TAB( 10)X(I); TAB( 27)Y(I): NEXT
200  REM   CALCOLO
210  FOR I = 1 TO N:X = X(I):Y = Y(I)
220 S1 = S1 +  LOG (X) *  LOG (Y)
230 S2 = S2 +  LOG (X)
240 S3 = S3 +  LOG (Y)
250 S4 = S4 +  LOG (X) ^ 2
260 S5 = S5 +  LOG (Y) ^ 2
270  NEXT I
280 S6 = (S2 * S3) / N
290 S7 = (S2 ^ 2) / N
300 S8 = (S3 ^ 2) / N
310 C1 = (S1 - S6) / (S4 - S7)
320 C2 = E ^ ((S3 / N) - C1 * (S2 / N))
330 R = ((S1 - S6) ^ 2) / ((S4 - S7) * (S5 - S8))
400  PRINT : PRINT "CURVA DI REGRESSIONE ": PRINT "Y = ";C2;"*X^";C1
420  PRINT : PRINT "COEFF.DI DETERM. R^2 = ";R: PRINT
430  DEF  FN Y(X) = C2 * X ^ C1
500  PRINT L$: PRINT "CALCOLO NUOVI VALORI   (999) PER FINIRE"
510  PRINT "     VAR.INDIP. X(I)  VAR.DIP. Y(I)": PRINT
520  PRINT  TAB( 10)"";: INPUT "";X
530  IF X = 999 THEN  HOME : END
535  FOR K = 1 TO 20: PRINT  CHR$ (8);: NEXT
540  PRINT  TAB( 27)"";: PRINT  FN Y(X): GOTO 520
600  DATA  50,100,150,200: REM  VARIABILE INDIPENDENTE NUMERO DI PAROLE
610  DATA  8.7,20.5,32.4,46.5: REM  VARIABILE DIPENDENTE TEMPO
```

Figura 2 — Listing Programma di Regressione Esponenziale. Il programma è utilizzato per la determinazione della curva di regressione num. parole/tempo di esecuzione, dell'HEAPSORT, e quindi per la valutazione delle sue prestazioni.

```
----------------------------------------------
     REGRESSIONE  ESPONENZIALE
      4  COPPIE DI NUMERI
----------------------------------------------
    VAR INDIP  X(I)  VAR DIP  Y(I)
        50                8.7
        100               20.5
        150               32.4
        200               46.5
CURVA DI REGRESSIONE
Y = .0789415328*X^1.20351505
COEFF DI DETERM  R^2 = .999680185
----------------------------------------------
CALCOLO NUOVI VALORI   (999) PER FINIRE
    VAR INDIP  X(I)  VAR DIP  Y(I)
        1000              321.996163
        2000              741.55749
```

Figura 3 — Output su printer del Programma di Regressione Esponenziale. Una volta definita la curva di regressione con la istruzione DEF FNY(X), basta immettere il valore della variabile indipendente X per avere il corrispondente valore della variabile dipendente Y.

REGRESSIONE LINEARE
Equazione del tipo Y=A*X+B
Coeff. A, B

$$A=\frac{\text{SUM(X*Y)}-\frac{\text{SUM(X)*SUM(Y)}}{N}}{\text{SUM(X\^{}2)}-\frac{\text{SUM(X)\^{}2}}{N}}$$

$$B=\frac{\text{SUM(Y)}}{N}-A\cdot\frac{\text{SUM(X)}}{N}$$

REGRESSIONE ESPONENZIALE (BASE E)
Equazione del tipo Y=A*E^(B*X)
Coeff. A, B

$$B=\frac{\text{SUM(X*Y)}-\frac{\text{SUM(X)*SUM(LN(Y))}}{N}}{\text{SUM(X\^{}2)}-\frac{\text{SUM(X)\^{}2}}{N}}$$

$$A=E\hat{}\left(\frac{\text{SUM(Y)}}{N}-B\cdot\frac{\text{SUM(X)}}{N}\right)$$

REGRESSIONE LOGARITMICA
Equazione del tipo Y=A-B*LN(X)
Coeff. A, B

$$B=\frac{\text{SUM(LN(X)*Y)}-\frac{\text{SUM(LN(X))*SUM(Y)}}{N}}{\text{SUM(LN(X)\^{}2)}-\frac{\text{SUM(LN(X))\^{}2}}{N}}$$

$$A=\frac{\text{SUM(Y)}}{N}-B\cdot\frac{\text{SUM(LN(X))}}{N}$$

REGRESSIONE ESPONENZIALE
Equazione del tipo Y=A*X^B
Coeff. A, B

$$B=\frac{\text{SUM(LN(X)*LN(Y))}-\frac{\text{SUM(LN(X))*SUM(LN(Y))}}{N}}{\text{SUM(LN(X)\^{}2)}-\frac{\text{SUM(LN(X))\^{}2}}{N}}$$

$$A=E\hat{}\left(\frac{\text{SUM(LN(Y))}}{N}-B\cdot\frac{\text{SUM(LN(X)\^{}2)}}{N}\right)$$

Figura 4 — Prospetto delle formule matematiche necessarie per il calcolo delle quattro curve di regressione. Le formule, essendo sostanzialmente simili, permettono molte semplificazioni del programma.

tarne a fondo le prestazioni.

Cronometriamo quanto impiega ad ordinare 50, 100, 150, 200 parole, poi, con i tempi cronometrati, cerchiamo una curva di regressione: questa ci permetterà di valutare teoricamente, in modo approssimato ma senza dover eseguire lunghe prove pratiche, quanto il programma impiegherebbe a ordinare 1000, 10000 o anche un milione di parole.

Il numero di parole è la variabile indipendente, in quanto la scegliamo noi, mentre il tempo di esecuzione è la variabile dipendente.

La curva di regressione che più si approssima, nel nostro esempio ai dati sperimentali è una curva esponenziale del tipo Y = A*X^B. Modificando un poco il programma presentato in seguito (figura 2) abbiamo visualizzato i risultati dell'elaborazione (figura 3). Abbiamo poi cronometrato i tempi di esecuzione nel caso di 500 e di 1000 parole ed abbiamo confrontato i tempi pratici con quelli teorici, valutati tramite la curva di regressione. Abbiamo trovato valori molto simili a quelli previsti (152 secondi contro 140 secondi e 357 secondi contro 322 secondi).

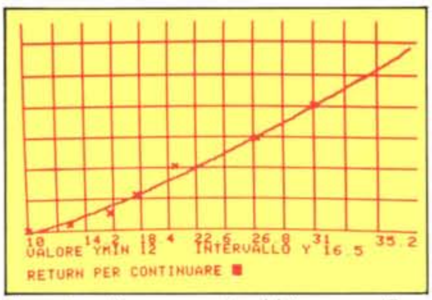

Figura 6 — Output su monitor del Programma Curve di Regressione. In questi programmi, la formattazione dei dati per renderli compatibili con le caratteristiche del video grafico, è sicuramente la parte più difficile.

Il programma fornisce pure il coefficiente di determinazione che, essendo quasi uguale a uno, indica che la curva trovata approssima molto bene i dati calcolati empiricamente (cioè i tempi cronometrati).

Varie Curve di Regressione

L'interpolazione statistica, cioè la ricerca della curva di regressione e il suo impiego per la determinazione di valori teorici, pone due problemi principali.

Il primo consiste nella scelta del tipo di espressione analitica più adatta a sintetizzare l'andamento del fenomeno. Non è possibile stabilire delle regole fisse in quanto alcune curve sono adatte a rappresentare certi fenomeni ma non sono adatte per altri.

Occorre quindi fare un po' di esperienza o, magari utilizzando il programma che presentiamo, cercare più curve con gli stessi dati.

Il secondo problema consiste nella scelta del metodo di calcolo della curva. Cioè nella scelta della condizione che la funzione deve soddisfare rispetto ai valori osservati. Il metodo più usato è quello dei minimi quadrati, che si trova in tutti i testi di matematica statistica (ai quali si rimanda per approfondimenti) e che si enuncia così:
la curva di regressione deve essere tale da rendere minima la somma dei quadrati degli scarti (misurati in direzione parallela all'asse delle ordinate) tra i valori sperimentali e i valori teorici calcolati della curva.

In questo articolo presentiamo cinque curve differenti, distribuite in due programmi di regressione. La determinazione delle curve avviene in tutti e cinque i casi con il metodo dei minimi quadrati.

Il programma "curve di regressione", data una serie di coppie di valori, tira fuori quattro curve differenti (regressione lineare, esponenziale base "e", logaritmica, esponenziale).

Poiché il programma non perde i dati immessi è possibile tracciare facilmente con gli stessi dati le varie curve e quindi vedere quale li approssima meglio. Il programma fornisce inoltre il valore del coefficiente di determinazione, che indica il grado di bontà della curva trovata. Tanto più questo valore è vicino a 1 tanto più la curva trovata approssima i dati sperimentali.

Il secondo programma risolve il problema della regressione polinomiale, quando cioè la curva di regressione è una parabola di ordine N-simo.

Il programma Curva di Regressione

In figura 4 abbiamo riportate le for-

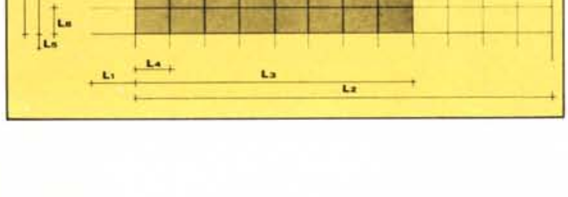

Figura 7 — Progetto Output su plotter. Programmi grafici complessi, richiedono preliminarmente uno studio a tavolino dei problemi di traduzione dei dati elaborati in dati visualizzabili.

```
100  REM  INIZIALIZZAZIONE DEL PROGRAMMA
110  CLEAR : DIM X(30),Y(30),S(34):E = 2.71828183
120 L$ = "--------------------------------------"
130  HOME : VTAB (12): PRINT " PROGRAMMI DI REGRESSIONE"
140  VTAB (20): INPUT " QUANTE COPPIE DI NUMERI ";N
150  HOME : PRINT L$;: PRINT "    IMMISSIONE DATI "
160  PRINT " ";N;"  COPPIE DI NUMERI  ": PRINT L$
170  FOR I = 1 TO N: PRINT "X(";I;"),Y(";I;")  ";
180  INPUT " ";X(I),Y(I): GOSUB 1000: NEXT
200  TEXT : HOME : PRINT L$: PRINT " SCEGLI LA REGRESSIONE DESIDER
     ATA ": PRINT L$
210  PRINT "  1- REGRESSIONE LINEARE": PRINT
220  PRINT "  2- REGRESSIONE ESPONENZIALE (BASE E)": PRINT
230  PRINT "  3- REGRESSIONE LOGARITMICA": PRINT
240  PRINT "  4- REGRESSIONE ESPONENZIALE": PRINT
250  PRINT "  5- IMMISSIONE NUOVE COPPIE DI VALORI": PRINT
260  PRINT "  9- FINE": PRINT : PRINT L$: INPUT "  ";S$
280  IF  VAL (S$) = 9 THEN : HOME : END
282  IF  VAL (S$) < 1 OR  VAL (S$) > 5 THEN 200
284  IF  VAL (S$) = 5 THEN 100
290  ON  VAL (S$) GOSUB 1500,1600,1700,1800
300  GOSUB 2500: GOSUB 2000: HGR : HCOLOR= 3: GOTO 3000
1000  REM  SUBROUTINE CALCOLO DATI
1010 S(11) = S(11) + X(I) * Y(I): REM  SUM(X*Y)
1020 S(12) = S(12) + X(I) *  LOG (Y(I)): REM  SUM (X*LN(Y))
1030 S(13) = S(13) +  LOG (X(I)) * Y(I): REM   SUM (LN(X)*Y)
1040 S(14) = S(14) +  LOG (X(I)) *  LOG (Y(I)): REM   SUM (LN(X)*L
     N(Y))
1100 S(21) = S(21) + X(I): REM  SUM(X)
1110 S(22) = S(22) +  LOG (X(I)): REM  SUM(LN(X))
1120 S(23) = S(23) + Y(I): REM  SUM(Y)
1130 S(24) = S(24) +  LOG (Y(I)): REM  SUM(LN(Y))
1200 S(31) = S(31) + X(I) ^ 2: REM   SUM(X^2)
1210 S(32) = S(32) +  LOG (X(I)) ^ 2: REM  SUM(LN(X)^2)
1220 S(33) = S(33) + Y(I) ^ 2: REM   SUM(Y^2)
1230 S(34) = S(34) +  LOG (Y(I)) ^ 2: REM  SUM(LN(Y)^2)
1290  RETURN
1500  REM  LINEARE
1510 S1 = S(11):S2 = S(21):S3 = S(23):S4 = S(31):S5 = S(33): GOSUB
     1900
1520  DEF  FN Y(X) = C1 * X + C2: RETURN
1600  REM  ESPONENZIALE (BASE E)
1610 S1 = S(12):S2 = S(21):S3 = S(24):S4 = S(31):S5 = S(34): GOSUB
     1900
1620 C2 = E ^ C2: DEF  FN Y(X) = C2 * E ^ (C1 * X): RETURN
1700  REM  LOGARITMICA
1710 S1 = S(13):S2 = S(22):S3 = S(23):S4 = S(32):S5 = S(33): GOSUB
     1900
1720  DEF  FN Y(X) = C2 + C1 *  LOG (X): RETURN
1800  REM  ESPONENZIALE
1810 S1 = S(14):S2 = S(22):S3 = S(24):S4 = S(32):S5 = S(34): GOSUB
     1900
1820 C2 = E ^ C2: DEF  FN Y(X) = C2 * X ^ C1: RETURN
1900  REM  VALORI COMUNI
1910 S6 = (S2 * S3) / N:S7 = (S2 ^ 2) / N:S8 = (S3 ^ 2) / N
1920 C1 = (S1 - S6) / (S4 - S7):C2 = (S3 / N) - C1 * (S2 / N)
1930 R = ((S1 - S6) ^ 2) / ((S4 - S7) * (S5 - S8)): RETURN
2000  REM  STAMPA DATI
2010  HOME : PRINT "I"; TAB( 5)"X(I)"; TAB( 15)"Y(I)": PRINT
2020  FOR I = 1 TO N: PRINT I; TAB( 5)X(I); TAB( 15)Y(I): NEXT
2030  PRINT : PRINT "CURVA DI REGRESSIONE": PRINT
2100  IF  VAL (S$) = 1 THEN  PRINT "Y = ";C1;"*X + ";C2
2110  IF  VAL (S$) = 2 THEN  PRINT "Y = ";C2;"*E^(";C1;"*X)"
2120  IF  VAL (S$) = 3 THEN  PRINT "Y = ";C2;" - ";C1;"*LOG(X)"
2130  IF  VAL (S$) = 4 THEN  PRINT "Y = ";C2;"*X^";C1
2140  PRINT : PRINT "COEFF.DI DETERM. R^2 = ";R: PRINT
2150  PRINT : PRINT : INPUT "PREMI RETURN PER CONTINUARE ";W$: PRINT

2160  PRINT "FORMATTAZIONE DEI DATI"
2170  PRINT
2310  PRINT "  X MAX  = ";MX: PRINT "  Y MAX  = ";MY
2320  PRINT "  X MIN  = ";NX: PRINT "  Y MIN  = ";NY: PRINT
2330  PRINT "  INT X  = ";DX: PRINT "  INT Y  = ";DY: PRINT
2340  PRINT "  SCALX  = ";SX: PRINT "  SCALY  = ";SY: PRINT : PRINT

2350  INPUT "PREMI RETURN PER CONTINUARE ";W$: HOME
2500  REM  RICERCA FATTORI DI SCALA
2510 MX =  - 9999:MY =  - 9999:NX = 9999:NY = 9999
2520  FOR I = 1 TO N
2530  IF X(I) < NX THEN NX = X(I)
2540  IF X(I) > MX THEN MX = X(I)
2550  IF Y(I) < NY THEN NY = Y(I)
2560  IF Y(I) > MY THEN MY = Y(I)
2570  NEXT
2580 DX = MX - NX:DY = MY - NY:SX = 210 / DX:SY = 96 / DY: RETURN

3000  REM   ROUTINE DISEGNO : SQUADRATURA
3010  FOR I = 2 TO 270 STEP 21: HPLOT I,0 TO I,157: NEXT
3020  FOR I = 13 TO 157 STEP 24: HPLOT 2,I TO 279,I: NEXT
3300  REM  DISEGNO PUNTI
3310  FOR I = 1 TO N:X% = (X(I) - NX) * SX + 2.5:Y% = 156.5 - (Y(I
     ) - NY) * SY
3320  HPLOT X% - 2,Y% - 2 TO X% + 2,Y% + 2
3330  HPLOT X% - 2,Y% + 2 TO X% + 2,Y% - 2: NEXT
3400  REM  TRACCIATO CURVA
3410 IX = DX / 30:LX = DX / 5:LY = DY / 4
3420 X% = 2.5:Y% = 156.5 - ( FN Y(NX) - NY) * SY
3430  HPLOT X%,Y%
3440  FOR X = NX TO MX + DX STEP IX
3450 X% = (X - NX) * SX + 2.5:Y% = 156.5 - ( FN Y(X) - NY) * SY
3460  IF Y% < 0 OR X% > 279 THEN 3500
3470  HPLOT  TO X%,Y%: NEXT X
3500  REM  SCRITTE DI RIFERIMENTO
3510  HOME : FOR K = 0 TO 6:A$ =  LEFT$ ( STR$ (NX + K * LX) + "
         ",4)
3520  VTAB (21): HTAB ((K + 1) * 6 - 5): PRINT A$: NEXT K
3530  VTAB (22): PRINT "VALORE YMIN ";NY;"    INTERVALLO Y ";LY
3540  VTAB (24): INPUT "RETURN PER CONTINUARE ";W$: GOTO 200
```

Figura 5 — Listing del Programma Curve di Regressione. Il programma si divide in due parti. La prima, esclusivamente matematica, fornisce la curva di regressione e il coefficiente di determinazione. La seconda, dalla riga 2150 in poi, fornisce la visualizzazione su monitor.

```
100  REM  INIZIALIZZAZIONE
110 D$ =  CHR$ (4):L$ = "----------------------------------------"

200  REM  INPUT DATI
210  HOME : PRINT L$: PRINT " REGRESSIONE POLINOMIALE ": PRINT " I
     MMISSIONE DATI  ": PRINT " VARIABILE INDIPENDENTE"
220  PRINT : PRINT L$: INPUT "  COPPIE DI VALORI        ";M: PRINT
     : PRINT L$
230  PRINT : FOR I = 1 TO M: PRINT " X(";I;"),(Y";I;")   ";: INPUT
     "";X(I),Y(I): NEXT
240  PRINT : PRINT L$: GOSUB 500: GOSUB 2000: END
500  REM  ROUTINE CALCOLO REGRESSIONE
510  INPUT " GRADO DEL POLINOMIO ";N:N2 = 2 * N:N1 = N + 1: IF N >
     M - 1 OR N > 5 THEN 510
520  FOR I1 = 1 TO N2: FOR J = 1 TO M:S1(I1) = S1(I1) + X(J) ^ I1:
      NEXT J,I1
530  FOR I2 = 1 TO N1: FOR J = 1 TO M:S2(I2) = S2(I2) + Y(J) * (X(
     J) ^ (I2 - 1)): NEXT J,I2
540  FOR IC = 1 TO N1:ID = IC - 1: FOR IR = 1 TO N1
550  IF IC = 1 AND IR = 1 THEN C(1,1) = M:ID = ID + 1: GOTO 570
560 C(IR,IC) = S1(ID):ID = ID + 1
570  NEXT IR,IC
580 IC = N1 + 1: FOR IR = 1 TO N1:C(IR,IC) =  - S2(IR): NEXT IR: GOSUB
     700
590  PRINT L$: PRINT : FOR I = 1 TO N1: PRINT "COEFF.GRADO ";I - 1
     ; TAB( 24)XS(I): NEXT
595  FOR K = N + 2 TO 5:XS(K) = 0: NEXT
600  FOR K = 1 TO 1999: NEXT : RETURN
700  REM  SOLUZIONE SISTEMA
710 NC = N1 + 1: FOR IR = 1 TO N1
720  FOR II = IR TO N1: IF C(II,IR) = 0 GOTO 740
730  NEXT II: GOTO 800
740  IF II <  > IR GOTO 770
750  FOR J = IR TO N1: IF C(J,IR) <  > 0 GOTO 770
760  NEXT J: PRINT " SOLUZIONE IMPOSSIBILE ": END
770 NS = N1 - 1
780  FOR JR = IR TO NS: FOR JC = IR TO NC:C(JR,JC) = C(JR,JC) + C(
     JR + 1,JC): NEXT JC,JR
790  FOR JC = IR TO N1:C(NR,JC) = C(NR,JC) + C(1,JC): NEXT JC: GOTO
     720
800 IS = IR + 1
810  FOR JR = IR TO N1  FOR JC = IS TO NC:C(JR,JC) = C(JR,JC) / C(
     JR,IR): NEXT JC,JR: IF IR = NR THEN 830
820  FOR JC = IS TO NC: FOR JR = IS TO N1:C(JR,JC) = C(JR,JC) - C(
     IR,JC): NEXT JR,JC,IR
830 XS(NC) = 1: FOR J = 1 TO N1:I = N1 - J + 1:I1 = I + 1:XS(I) =
     0
840  FOR JJ = I1 TO NC:XS(I) = XS(I) + C(I,JJ) * XS(JJ): NEXT JJ:X
     S(I) =  - XS(I): NEXT J: RETURN
2000  REM  FORMATTAZIONE DISEGNO
2010  HOME : PRINT L$: PRINT "------ IMMISSIONE DATI FORMATTAZIONE
     -----"
2020  PRINT L$: PRINT : INPUT "  MARGINE SCRITTE ORIZZONTALI     "
     ;L1
2022  PRINT : INPUT "  LARGHEZZA TOTALE DISEGNO        ";L2
2024  PRINT : INPUT "  RAPPORTO LARGH.TOT./ZONA DATI   ";R1:L3 = L
     2 / R1
2026  PRINT : INPUT "  QUADRETTATURA ORIZZONTALE       ";N1:L4 = L
     2 / N1
2030  PRINT L$: INPUT "  MARGINE SCRITTE VERTICALI       ";L5
2032  PRINT : INPUT "  ALTEZZA TOTALE DISEGNO          ";L6
2034  PRINT : INPUT "  RAPPORTO ALTEZZ.TOT./ZONA DATI  ";R2:L7 = L
     6 / R2
2036  PRINT : INPUT "  QUADRETTATURA VERTICALE         ";N2:L8 = L
     6 / N2
2100  REM  RICERCA MAX E MIN
2110 MM = 9999:MX =  - MM:MY =  - MM:NX = MM:NY = MM: FOR I = 1 TO
     M
2130  IF X(I) < NX THEN NX = X(I)
2140  IF X(I) > MX THEN MX = X(I)
2150  IF Y(I) < NY THEN NY = Y(I)
2160  IF Y(I) > MY THEN MY = Y(I)
2170  NEXT I
2200  REM  RICERCA FATTORI DI SCALA
2210 DX = MX - NX:DY = MY - NY:SX = L3 / DX:SY = L7 / DY:IX = DX /
     100
2300  REM  DISEGNO SQUADRATURA
2310  FOR I = L1 TO L1 + L2 STEP L4:X = I:Y = L5 + L6: GOSUB 3000:
     Y = 1: GOSUB 3100: NEXT
2320  FOR I = L5 TO L5 + L6 STEP L8:Y = I:X = L1 + L2: GOSUB 3000:
     X = 1: GOSUB 3100: NEXT
2400  REM  DISEGNO PUNTI
2410  FOR I = 1 TO M:XR = (X(I) - NX) * SX + L1:YR = (Y(I) - NY) *
     SY + L5
2420 X = XR - 20:Y = YR - 20: GOSUB 3000:X = XR + 20:Y = YR + 20: GOSUB
     3100
2430 X = XR + 20:Y = YR - 20: GOSUB 3000:X = XR - 20:Y = YR + 20: GOSUB
     3100
2440  NEXT I
2500  REM  SCRITTE SUL DISEGNO
2510  FOR XD = NX TO MX - .01 STEP DX / N1
2520 X = ((XD - NX) * SX) + R1 + L1 + 10:Y = 1: GOSUB 3000
2530 X$ =  LEFT$ ( STR$ ((XD - NX) + R1 + NX) + "        ",5): GOSUB
     3200: NEXT
2540  FOR YD = NY TO MY - .01 STEP DY / N2
2550 Y = ((YD - NY) * SY) + R2 + L5 + 10:X = 1: GOSUB 3000
2560 X$ =  LEFT$ ( STR$ ((YD - NY) + R2 + NY) + "        ",5): GOSUB
     3200: NEXT
2600  REM  TRACCIAMENTO CURVA
2610  DEF  FN YC(XC) = XS(1) + XS(2) * XC + XS(3) * XC ^ 2 + XS(4)
     * XC ^ 3 + XS(5) * XC ^ 4 + XS(6) * XC ^ 5
2620 XC = NX:YC =  FN YC(XC)
2630 X = (XC - NX) * SX + L1:Y = (YC - NY) * SY + L5: GOSUB 3000
2640 XC = XC + IX:YC =  FN YC(XC)
2650 X = (XC - NX) * SX + L1:Y = (YC - NY) * SY + L5
2660  IF X > (L1 + L2) OR Y > (L5 + L6) THEN  RETURN
2670  GOSUB 3100: GOTO 2640
3000  REM  MOVE
3010 X =  INT (X):Y =  INT (Y)
3020  PRINT D$"PR#1": PRINT "M";X;",";Y: PRINT D$"PR#0"
3030  RETURN
3100  REM  DRAW
3110 X =  INT (X):Y =  INT (Y)
3120  PRINT D$"PR#1": PRINT "D";X;",";Y: PRINT D$"PR#0"
3130  RETURN
3200  REM  SCRITTURA STRINGA
3210  PRINT D$"PR#1": PRINT "P";X$: PRINT D$"PR#0"
3220  RETURN
```

Figura 8 — Listing Programma di Regressione Polinomiale. Trattandosi di un programma molto generalizzato, va curata particolarmente la fase di immissione dati. Immissioni non corrette, potrebbero produrre grafici non comprensibili, oppure bloccare il programma.

mule matematiche risolutive con le quali si trovano i coefficienti A, B necessari per determinare le curve di regressione nei quattro casi esaminati.

Per la cronaca abbiamo tratto tali formule dal libretto di Programmi Applicativi dell'HP 25, che le presenta in maniera molto adatta per una rapida traduzione in BASIC.

Poiché i quattro casi in esame (regressione lineare, regressione esponenziale in base e, regressione logaritmica, regressione esponenziale) sono sostanzialmente simili, abbiamo realizzato un programma unico (listing in figura 5).

Durante l'input vengono direttamente calcolati e caricati negli accumulatori S (I) i valori delle grandezze necessarie al calcolo. Terminato l'input delle N coppie di valori, il programma chiede quale curva si intende calcolare e, indicata l'opzione, esegue una delle piccole subroutine (righe 1500, 1600, 1700, 1800) per la determinazione dei valori S1... S8 che permettono il calcolo dei coefficienti A,B e del coefficiente di determinazione R^2 della curva.

Dal punto di vista analitico, il programma finisce qui.

Premuto il tasto RETURN inizia la parte relativa alla visualizzazione dei dati sul monitor APPLE II.

Questa seconda parte si divide in varie fasi, distinguibili sul listato dai REMARKS.

1 - Formattazione dei dati output
tramite la solita ricerca di massimo e minimo vengono trovati i dati necessari per la formattazione e che vengono visualizzati. Tramite questi dati è possibile tradurre i valori teorici della curva in valori visualizzabili sul monitor.

2 - Squadratura del disegno
viene tracciata una quadrettatura per facilitare la lettura dei valori di interpolazione.

3 - Scrittura valori di riferimento
con gli stessi coefficienti di traduzione trovati per la formattazione del disegno si calcolano i valori di riferimento della quadrettatura e quindi del grafico. Questi valori vengono scritti in basso rispetto al disegno.

Questo sistema è molto rudimentale e quindi suggerisco, a chi volesse trascrivere il programma, di modificare la routine in funzione sia del software che possiede per tracciare scritte alfanumeriche anche sulla pagina HGR2, sia in funzione del tipo di dati da rappresentare.

4 - Disegno dei punti
sulla squadratura vengono tracciati i punti (individuati da crocette) rappresentanti i valori empirici della regressione.

5 - Disegno curva
la curva dal punto di vista analitico è definita dall'istruzione DEF FNY (X). I valori di Y trovati nel loop della X vengono, anche questi, tradotti in coordinate schermo (tramite i coefficienti di formattazione di cui abbiamo tanto parlato).

Ora, esaminando il grafico, si può sia valutare quanto la curva approssima i dati sperimentali, sia determinare i valori teorici della curva.

Anche esaminando il listato del programma si può valutare quanto sia più difficile risolvere il problema della formattazione del disegno, della sua squadratura e del tracciamento delle scritte di riferimento rispetto al disegno puro e semplice della curva.

Anzi quest'ultimo, una volta definita con l'istruzione DEF FNY (X) la funzione da visualizzare, è facile la realizzazione tramite un loop sulla X.

La difficoltà maggiore è invece quella di combinare opportunamente la formattazione del disegno con il tracciamento delle scritte di riferimento. Questo perchè, essendo il programma generalizzato, non si sa a priori né il tipo dei valori che saranno immessi, né la loro unità di misura, né la loro entità, né il loro intervallo, (cioè posso utilizzare il programma per esaminare l'andamento di un fenomeno nel corso degli anni, o al limite, per valutare l'errore di lettura del braccio di un giradischi).

Nelle applicazioni pratiche, dove il programma viene realizzato in un campo specifico, il problema della formattazione può essere risolto una volta per tutte, magari utilizzando carta con prestampata la quadrettatura e le scritte di riferimento.

In tal caso il programma avrà una costante di Scaling predeterminata con la quale dovrà tracciare solo la curva.

Il programma (vedi output in figura 6), riesce a rendere chiaramente sul monitor, l'andamento della curva rispetto ai punti corrispondenti ai valori sperimentali.

Certo, però, per una utilizzazione del grafico anche per valutare i valori teorici della curva è indispensabile una uscita su carta.

Regressione Polinomiale con il metodo dei minimi quadrati

Il metodo più comunemente usato per le regressioni polinomiali è quello dei minimi quadrati con il procedimento detto "delle equazioni normali", che si adotta soprattutto quando la forma assunta dai punti osservati consente di scegliere a priori il tipo della funzione interpolatrice, o, almeno, di scartare a priori alcuni tipi più semplici. Questo metodo tende a minimizzare la somma dei quadrati degli scostamenti dei singoli punti osservati dalla curva teorica interpolatrice, ossia a rendere nulle tutte le derivate parziali prime di tale somma rispetto ad ognuno dei coefficienti del polinomio di interpolazione.

Scelta, per una serie di K termini, la solita funzione del tipo $Y = a_0 + a_1 * X + a_2 * X^2 + \ldots + a_K * X^K$ dobbiamo cercare, mediante un sistema di equazioni i coefficienti $a_0, a_1, a_2, \ldots a_K$ del polinomio.

Si è già detto che il metodo tende a minimizzare i quadrati degli scarti ε_1, quindi possiamo dire che:

$$f(a_1, a_2, a_3, \ldots a_K) = \sum_{i=1}^{N} \varepsilon^2_i = \text{minimo}$$

dove

$$\varepsilon_i = \varepsilon(X_i) = Y(X_i) - Y_i \quad [i=1,N]$$

sapendo che:

$$Y(X_i) = a_0 + a_1 * X_i + a_2 * X_i^2 + \ldots + a_K * X^K_i$$

Questo significa rendere nulle tutte le derivate parziali rispetto ai coefficienti del polinomio e perciò

$$\frac{\delta f}{\delta a_0} = 0 \quad \frac{\delta f}{\delta a_1} = 0 \quad \frac{\delta f}{\delta a_K} = 0$$

e poiché

$$\frac{\delta f}{\delta a_0} = 0 \rightarrow \Sigma Y_i = a_0 * N + a_1 * \Sigma X_i + \ldots + a_N * \Sigma X^N_i$$

$$\frac{\delta f}{\delta a_1} = 0 \rightarrow \Sigma Y_i * X_i = a_0 + \Sigma X_i + a_1 * \Sigma X^2_i + \ldots + a_N * \Sigma X_i^{N+1}$$

$$\frac{\delta f}{\delta a_2} = 0 \rightarrow \Sigma Y_i * X^2_i = a_0 * \Sigma X^2_i + a_1 * \Sigma X^3_i + \ldots + a_N * \Sigma X_i^{N+2}$$

$$\frac{\delta f}{\delta a_K} = 0 \rightarrow \Sigma Y_i * X^K_i = a_0 * \Sigma X^{K+1} + a_1 * \Sigma X^{K+2} + \ldots + a_N * \Sigma X_i^{2K}$$

la soluzione di questo sistema di K+1 equazioni ci permetterà di trovare i K+1 coefficienti $(a_0, a_1, a_2, \ldots, a_K)$ del polinomio cercato di grado K.

La Regressione Polinomiale

Il secondo programma calcola la curva di regressione polinomiale, di grado N-simo. Ovvero date M coppie di valori empirici, vogliamo trovare i coefficienti a,b,c,.... della equazione Y = a + b*x + c*x^2 + d*x^3... che permette la migliore approssimazione dei risultati sperimentali.

Questo secondo programma ha una uscita su plotter che ci permette di completare il "discorsetto" sui problemi che si incontrano nel preparare un'uscita grafica (su plotter o su monitor) di una data funzione, eventualmente dotata di scritte, di scale di lettura, ecc. Anzi è molto più pesante la programmazione delle parti "accessorie" del disegno, che non il semplice tracciamento della funzione.

In generale è consigliabile preparare a tavolino il progetto del disegno.

Ad esempio dovendo prevedere una uscita su plotter del programma di regressione polinomiale abbiamo fatto uno schizzo di come va organizzata l'uscita (figura 7).

Le grandezze che vanno definite per specificare il formato del disegno sono ben 8:

L(1) - è la grandezza da riservare per la zona scritte nella scala verticale.

L(2) - è la larghezza totale della zona disegno. Ovviamente L(1)+ L(2) deve essere minore della larghezza della carta su cui disegnare.

L(3) - è la larghezza entro la quale far ricadere i dati empirici della regressione. In particolare se si tratta di una interpolazione, ovvero dobbiamo valutare valori teorici interni ai valori empirici, porremo L(3) = L(2). Altrimenti se dobbiamo fare una estrapolazione, cioè con valori teorici esterni ai valori empirici, L(3) sarà una frazione di L(2).

L(4) - è l'intervallo della scala, rappresenta il passo della quadrettatura.

L(5), L(6), L(7), L(8) sono i corrispondenti valori nella direzione verticale.

Va qui notato, e vale anche per il programma precedente, che non c'è nessun legame dimensionale tra variabile indipendente, quella che va riportata sull'asse X, e quella dipendente, che va sull'asse della Y.

Questo significa che non è necessario, per formattare il disegno, utilizzare una unica scala di trasformazione tra X e Y, ma il programma le può scegliere indipendentemente l'una dall'altra.

Anche questo programma si può dividere in due parti indipendenti.

La prima parte consiste nella ricerca analitica della curva che avviene attraverso vari passi:

1 immissione delle coppie di valori sperimentali,

2 scelta del grado del polinomio,

3 ricerca del polinomio con il già ricordato metodo dei minimi quadrati.

Figura 9 — Output su plotter del Programma di Regressione Polinomiale. Il programma mostra sia i punti corrispondenti ai valori empirici, sia la curva di regressione. Se l'ordine del polinomio è pari al numero dei punti immessi meno uno, la curva passa per tutti i punti.

Da un punto di vista matematico, l'applicazione del metodo dei minimi quadrati comporta la soluzione di un sistema di equazioni. Il numero di equazioni del sistema è pari al grado del polinomio più uno che si vuole ottenere (vedi riquadro).

Nel programma tra le righe 500 e 600 vengono calcolati i coefficienti da inserire nel sistema di equazioni, mentre la soluzione è calcolata nella subroutine della riga 700.

La seconda parte del programma consiste nella visualizzazione della curva in un opportuno sistema di riferimento.

Anche qui sono facilmente individuabili i passi successivi, seguiti per la determinazione prima dei coefficienti di formattazione, poi per il disegno della squadratura, per il tracciamento dei valori di riferimento, per il disegno dei punti corrispondenti ai valori sperimentali ed infine per il disegno della curva vera e propria.

In figura 9 vediamo un esempio di output su plotter del programma. Sono stati immessi 4 punti ed è stata scelta una curva del 3° ordine. In tale caso (quando cioè il grado del polinomio è pari al numero dei punti meno uno) la curva determinata passa per tutti i punti. In figura 10 vediamo due output ottenuti con gli stessi valori empirici. I due disegni sono differenti sia perché sono stati immessi valori L(1)....L(8) di formattazione differenti, sia perché nel primo caso abbiamo un polinomio di 5° grado e nel secondo di 3° grado.

Va comunque tenuto presente che il programma richiede una immissione di dati coerenti. Ovvero in caso di dati immessi casualmente si possono presentare curve fuori scala con conseguente blocco dell'esecuzione.

Questo succede anche perché i valori di formattazione del disegno sono determinati elaborando i dati immessi inizialmente e non la curva calcolata.

Si può comunque migliorare l'affidabilità del programma, prevedendo un calcolo preventivo delle coppie di valori X, Y desunti dalla curva teorica, il loro caricamento su una matrice e l'esecuzione della routine di formattazione sui valori così immagazzinati.

Nessuna difficoltà, invece, come abbiamo più volte visto, nel prevedere l'uscita su plotter.

Tutti i comandi plotter utilizzati (MOVE, DRAW, PRINT) sono collocati nella subroutine delle righe 3000, 3100, 3200.

Chi non ha il plotter può sostituire a queste routine le corrispondenti routine per la visualizzazione su monitor.

A tale scopo è opportuno vedere lo specchietto di "traduzione" pubblicato nell'articolo sul numero scorso.

Il plotter da noi utilizzato è, come al solito, il Digiplot della Watanabe, che ha l'indiscutibile pregio di offrire prestazioni interessanti per un prezzo decisamente contenuto, aspetto fondamentale per un hobbysta o un professionista che non voglia o non possa fare un grosso investimento.

Francesco Petroni

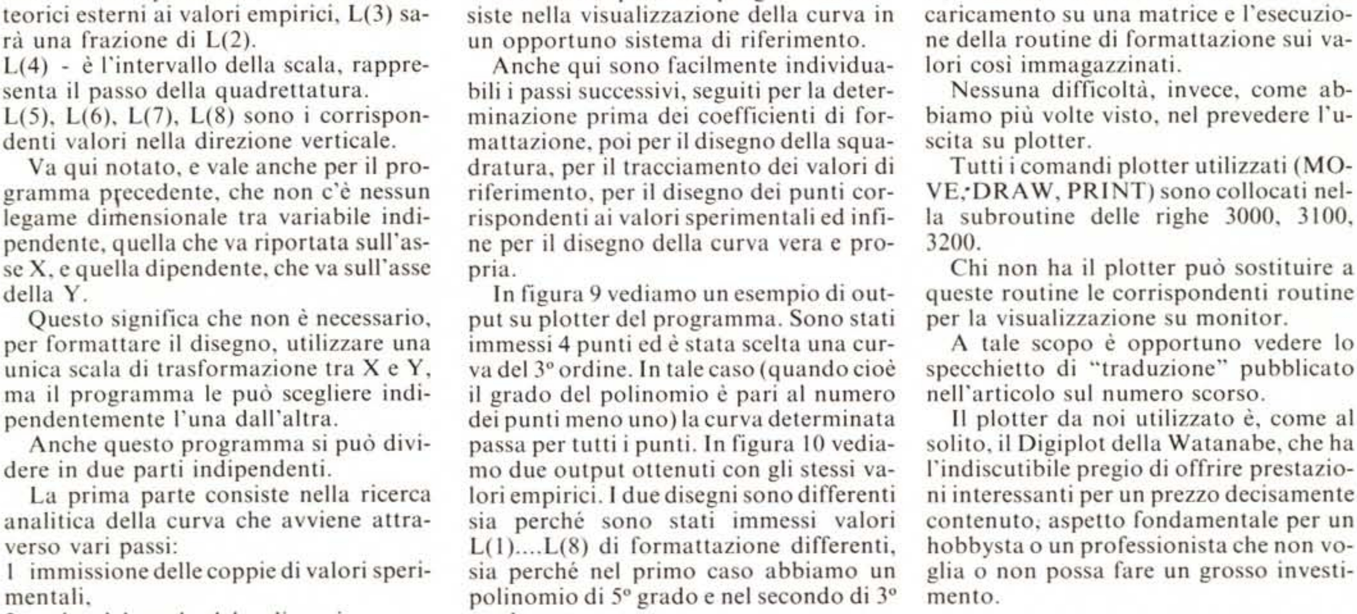

Figura 10 — Due output su carta del Programma di Regressione Polinomiale. Dati 6 punti vengono calcolati e visualizzati due polinomi di regressione del 5° e del 3° grado.