Ratio Math. n. 7 - 1994
25 - 38

LA DIMENSIONE CAMPIONARIA NELLA VERIFICA DI IPOTESI STATISTICHE

Enzo Ballone e Vittorio Colagrande***
**Facoltà di Medicina e Chirurgia, Università "G. d'Annunzio"- Chieti*
***Liceo Scientifico "G. Galilei" - Lanciano (CH)*

1. INTRODUZIONE

Nel presente lavoro si esporranno alcune problematiche attinenti la dimensione minima "ottimale" di un campione per la verifica di ipotesi sul valore di una proporzione o sul confronto di due proporzioni. Si presenterà anche un software per il calcolo della dimensione stessa.

La trattazione è uno sviluppo della relazione presentata dagli autori al Convegno Interregionale della MATHESIS su "Calcolo delle Probabilità, Statistica e Crittografia nell'insegnamento", tenutosi a Isernia nei giorni 2-4 aprile 1992. Essa vuole essere una proposta didattica nell'ambito dell'insegnamento della Statistica nelle scuole secondarie superiori. La stessa scelta di riportare in appendice un software sviluppato da C. DI MASCIO (Facoltà di Medicina e Chirurgia, Università di Chieti), secondo la metodologia della programmazione strutturata, ha una valenza espressamente didattica.

Si supporranno noti concetti e tecniche concernenti il campionamento e la verifica di ipotesi statistiche.

2. PRESENTAZIONE DELLA PROBLEMATICA

Il processo inferenziale della verifica di ipotesi statistiche consiste nel confrontare un'ipotesi relativa ad una popolazione con i risultati sperimentali che scaturiscono da un campione estratto dalla popolazione stessa.

Problemi importanti (perché spesso si incontrano nella realtà quotidiana) sono quelli relativi a inferenze statistiche su "popolazioni dicotomiche".

A tal proposito si consideri il seguente esempio. Si vuole studiare la prevalenza della tossicofilia (uso occasionale di droghe leggere) nell'ambito degli studenti delle scuole superiori di una città (popolazione) e si decide di condurre un'indagine campionaria.

Si stabilisce allora di scegliere casualmente, con opportune tecniche, un campione di studenti che possa ritenersi "rappresentativo" di tutta la popolazione studentesca oggetto di studio e su di esso si effettua l'indagine statistica.

Se dall'indagine risultasse una percentuale di tossicofilia del 9%, qualora valutazioni a livello nazionale ritengano verosimile che circa il 15% dei giovani delle scuole medie superiori facciano uso occasionale di droghe leggere, un problema potrebbe essere quello di verificare, attraverso l'applicazione di un appropriato test statistico, se nella popolazione scolastica della città inesame la prevalenza del fenomeno sia statisticamente inferiore.

Un secondo problema potrebbe essere quello di stabilire se la tossicofilia aumenti con l'età. Ciò si potrebbe verificare, ad esempio, confrontando statisticamente le percentuali di prevalenza tra le prime due classi e le ultime due degli istituti superiori in esame.

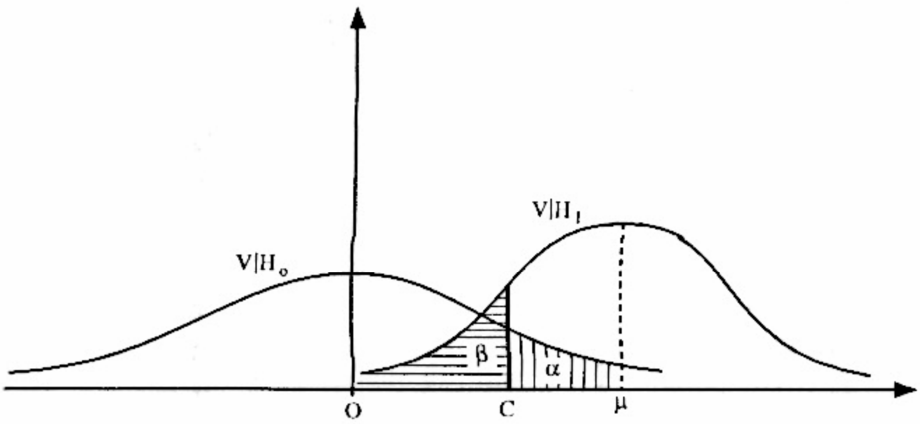
Per ottenere risultati attendibili da indagini campionarie e poter applicare quindi i processi dell'inferenza statistica, è necessario che il campione da analizzare sia "rappresentativo" della popolazione di provenienza, cioè rifletta il più possibile le sue caratteristiche.

In generale può affermarsi che quanto più il campione è numeroso, tanto più sarà rappresentativo; d'altra parte, però, al crescere della dimensione del campione crescono i costi e le difficoltà tecniche di organizzazione dell'indagine. La determinazione della dimensione campionaria costituisce, quindi, uno dei momenti di maggiore incertezza e riflessione da parte di un ricercatore che si appresti a realizzare un'indagine o un esperimento. Egli deve chiedersi, preventivamente, di quale grandezza debba essere il campione da esaminare; una scarsa attenzione alle dimensioni lascia il ricercatore nell'incertezza circa l'attendibilità dei risultati campionari e quindi può vanificare tutta la ricerca.

3. RELAZIONE FONDAMENTALE PER IL CALCOLO DELLA DIMENSIONE

Con riferimento alla problematica attinente la verifica di ipotesi nella ricerca statistica, si supponga che una variabile casuale test V, utilizzata per la verifica stessa, sia di tipo gaussiano con media m e varianza s^2, cioè che la sua densità di probabilità sia $N_{m,\,s}$ (v). Essa abbia densità di probabilità $N_{o,\,s1}$ (v) sotto l'ipotesi "H_o : media = 0 e varianza = s_1^2" e $N_{\mu,\,s2}$ (v) sotto l'ipotesi alternativa "H_1: media = μ e varianza = s_2^2".

Si può avere così, per $\mu > 0$ (ipotesi H_1 unidirezionale) una situazione come quella illustrata in figura seguente:



Nella figura sono individuate anche le aree relative agli errori di prima (Prob. = α) e seconda specie (Prob. = β), per un particolare valore c di V detto valore soglia, determinato in modo tale da minimizzare entrambi gli errori. È stato preso $0 < c < \mu$ e in quel che subito segue, si manterrà tale assunto.

Supponendo vera H_o, il livello di fiducia è espresso dalla probabilità

$$1 - \alpha = \int_{-\infty}^{c} N_{o,s1}\ (v)\ dv = 1/s_1 \int_{-\infty}^{c} N_{o,1}\ (v/s_1)\ dv = \int_{-\infty}^{c/s1} N_{o,1}\ (t)\ dt,$$

Dunque c/s_1 è la determinazione numerica $z_{1-\alpha}$ della variabile normale standardizzata che soddisfa la relazione

Prob $(z < z_{1-\alpha}) = 1 - \alpha$ e, quindi, $c = z_{1-\alpha}\ s_1$.

Supponendo vera H_1 si ha, invece, che l'errore di seconda specie è misurato dalla β

$$\beta = \int_{-\infty}^{c} N_{\mu,s2}\ (v)\ dv = 1/s_2 \int_{-\infty}^{c} N_{o,1}\ ((v - \mu)/s_2)\ dv =$$

$$= \int_{-\infty}^{t_c} N_{o,1}\ (t)\ dt,$$

avendo posto $t_c = (z_{1-\alpha}\ s_1 - \mu)/s_2$.

Pertanto t_c è proprio quel valore $-z_{1-\beta}$ della variabile normale standardizzata tale che Prob $(z < -z_{1-\beta}) = \beta$. Si può scrivere allora

$$-z_{1-\beta} = \frac{z_{1-\alpha}\ s_1 - \mu}{s_2}$$

da cui, osservato che $\mu = |\mu|$, si ricava la seguente

$$z_{1-\alpha}\ s_1 + z_{1-\alpha}\ s_2 = |\mu|. \tag{1}$$

La precedente formula ha validità anche nel caso $\mu < 0$.

Si osservi, comunque, che qualora l'H_1 fosse bidirezionale ($\mu \neq 0$) si prova che la relazione (1) rimane invariata nella sua "struttura", è sufficiente sostituire $z_{1-\alpha}$ con $z_{1-\alpha/2}$.

4. DETERMINAZIONE DELLA NUMEROSITÀ PER PROPORZIONI.

Nel caso si voglia verificare un'ipotesi statistica sul valore π_o della frazione π di unità, appartenenti ad una data popolazione, che presenta una delle due modalità di un carattere dicotomico, l'ipotesi nulla H_o può essere formulata come:

$$H_o\colon \pi = \pi_o \qquad (\delta = \pi - \pi_o = 0)$$

e quella alternativa H_1 può esprimersi con una delle due ipotesi unidirezionali

$$H_1: \pi > \pi_o \qquad (\delta = \pi - \pi_o > 0),$$

oppure

$$H_1: \pi < \pi_o \qquad (\delta = \pi - \pi_o < 0)$$

Il parametro π della popolazione assume il significato di frequenza relativa delle unità che in essa possiedono l'attributo preso in esame. Nel seguito esso verrà denominato indifferentemente con termini quali "proporzione", "frequenza", "percentuale".

Per la verifica dell'ipotesi nulla occorre sottoporre a test la differenza tra il valore **p** della frequenza in un campione e quello ipotizzato π_o della frequenza nella popolazione di riferimento.

Il valore $\delta = \pi - \pi_o$ che figura nella formulazione delle ipotesi può essere preso come quella differenza minima tra proporzioni che il ricercatore ritiene "rilevante" ai fini dell'indagine e di cui vuole verificare la significatività statistica. Esso va fissato prima di intraprendere l'indagine, pertanto l'ipotesi alternativa H_1 viene a presumere una popolazione caratterizzata da una frequenza $\pi_1 = \pi + \delta$.

Facendo riferimento ad un'estrazione campionaria di tipo casuale bernoulliano (con n si denoterà la dimensione di un campione) e indicata con P la variabile casuale proporzione campionaria, la variabile test V è data, nel caso di n "grande", dalla

$$z = \frac{P - \pi_o}{[\pi_o\,(1 - \pi_o)\,/\,n]^{\frac{1}{2}}}$$

Essa ha densità di probabilità con parametri

$$\mu = \frac{\delta}{[\pi_o\,(1 - \pi_o)\,/\,n]^{\frac{1}{2}}}\,, \qquad s_1 = 1\,, \qquad s_2 = \frac{[\pi_1\,(1 - \pi_1)\,/\,n]^{\frac{1}{2}}}{[\pi_o\,(1 - \pi_o)\,/\,n]^{\frac{1}{2}}}\,.$$

Sostituendo queste formule nella (1) si ottiene:

$$z_{1-\alpha} + z_{1-\beta}\,\frac{[\pi_1\,(1 - \pi_1)\,/\,n]^{\frac{1}{2}}}{[\pi_o\,(1 - \pi_o)\,/\,n]^{\frac{1}{2}}} = \frac{|\delta|}{[\pi_o\,(1 - \pi_o)\,/\,n]^{\frac{1}{2}}}\,,$$

da cui si ricava il valore minimo della dimensione campionaria:

$$n \approx \left[\frac{z_{1-\alpha} [\pi_o (1 - \pi_o)]^{\frac{1}{2}} + z_{1-\beta} [\pi_1 (1 - \pi_1)]^{\frac{1}{2}}}{\delta} \right]^2 . \qquad (2)$$

La condizione di minimo va intesa nel senso che ogni altro valore di **n** superiore a quello determinato con la (2) permetterà di ottenere risultati statisticamente più attendibili.

La (2) esprime la dimensione ottimale del campione in dipendenza della variabilità del fenomeno, del valore di δ e dei livelli di probabilità di errore α e β.

Riprendendo l'esempio delle tossicofilie introdotto nel par. 2, se a livello nazionale si stima la frequenza del fenomeno, nei giovani delle scuole superiori, intorno al 15% (π_o) e si ritiene che nelle scuole superiori della città esaminata il fenomeno sia meno frequente qualora si riscontrasse un valore della differenza tra frequenze minore o uguale al 6% (δ = - 0.06), si può calcolare il numero di studenti da includere nel campione. Dalla (2), per valori di $\alpha = 0.05$ e $1- \beta = 0.90$ cui corrispondono $z_{1-\alpha} = 1.645$ eo $z_{1-\beta}$ =1.282 (come si ricava consultando opportune tavole della distribuzione normale o utilizzando un software specifico come quello proposto nel successivo par. 6), si ottiene

$$n = \text{ROUND} \left[\left[\frac{1.645 \cdot (0.15 \cdot 0.85)^{\frac{1}{2}} + 1.282 \cdot (0.09 \cdot 0.91)^{\frac{1}{2}}}{0.06} \right]^2 \right] = 253 ,$$

dove con ROUND[...] si è indicato il minimo intero maggiore o uguale della quantità entro parentesi.

Per realizzare le condizioni suddette occorre, quindi, un campione di studenti di almeno 253 unità.

Qualora si volesse una differenza tra frequenze minore o uguale al 9% (δ = - 0.09), si può ricavare che, agli stessi livelli di errori di prima e seconda specie, la dimensione campionaria minima sarebbe di 98 unità. Sempre a titolo di esempio, qualora risultasse a livello nazionale un valore di $\pi_o = 0.25$ e si volesse una differenza minore o uguale al 10% (δ = - 0.10) agli stessi livelli di errore, la numerosità campionaria richiesta sarebbe di n = 137.

Nel caso si desideri verificare se esiste una diversa prevalenza della tossicofilia tra le prime due classi e le ultime due, è necessario fissare la differenza δ di percentuale di casi che si ritiene rilevante tra i due gruppi e determinare quindi le dimensioni minime dei due campioni.

Le ipotesi da saggiare sono quelle relative al confronto tra due proporzioni π_1 e π_2 di una stessa modalità di un carattere in due popolazioni dicotomiche distinte.

L'ipotesi nulla può essere formulata come:

$$H_o: \delta = \pi_1 - \pi_2 = 0$$

e quella alternativa prendendo in esame una delle ipotesi unidirezionali:

$$H_o: \delta = \pi_1 - \pi_2 > 0 \text{ oppure } H_1: \delta = \pi_1 - \pi_2 < 0.$$

Considerando anche in questo caso campioni casuali bernoulliani (n_1 e n_2 indicheranno le dimensioni di due campioni) e indicata con Pi - P_2 la variabile casuale delle proporzioni campionarie, la funzione test da utilizzare nel caso di "grandi campioni" è la seguente:

$$Z = \frac{P_1 - P_2}{[p^* (1 - p^*) (1/n_1 + 1/n_2)]^{\frac{1}{2}}} .$$

Z è distribuita con parametri

$$\mu = \frac{\delta}{[p^* (1 - p^*) (1/n_1 + 1/n_2)]^{\frac{1}{2}}}, \qquad s_1 = 1,$$

$$s_2 = \frac{[p_1 (1 - p_1)/n_1 + p_2 (1 - p_2)/n_2]^{\frac{1}{2}}}{[p^* (1 - p^*) (1/n_1 + 1/n_2)]^{\frac{1}{2}}},$$

dove p_1, p_2 sono opportune frequenze campionarie note assunte come stime delle frequenze π_1 e π_2 delle due popolazioni (studenti delle prime due classi e delle ultime due nel problema in esame) e la proporzione p* può essere data dalla media ponderata delle due frequenze campionarie pi ($p^* = (n_1p_1 + n_2p_2) / (n_1 + n_2)$), dalla prevalenza registrata in altre indagini similari o essere desunta dalla bibliografia.

Sostituendo i parametri, la (1) diventa

$$z_{1-\alpha} + z_{1-\beta} \frac{[p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2]^{\frac{1}{2}}}{[p^*(1-p^*)(1/n_1+1/n_2)]^{\frac{1}{2}}} =$$

$$= \frac{|\delta|}{[p^*(1-p^*)(1/n_1+1/n_2)]^{\frac{1}{2}}} .$$

Un primo dimensionamento campionario lo si può ottenere assumendo che i due campioni abbiano la stessa numerosità, ossia $n_1 = n_2 = n$. In tal caso, ricavando *n dalla* formula precedente, si ha

$$n \approx \left[\frac{z_{1-\alpha}[2p^*(1-p^*)]^{\frac{1}{2}} + z_{1-\beta}[p_1(1-p_1) + p_2(1-p_2)]^{\frac{1}{2}}}{\delta}\right]^2 .$$

Nell'esempio della tossicofilia, allora, iotizzando che le percentuali di studenti interessati dal fenomeno siano, ad esempio, $p_1 = 0.07$ e $p_2 = 0.13$, con un livello di significatività $\alpha = 0.05$ e una potenza $1 - \beta = 0.90$, dalla (3) si ottiene

$$n = \mathrm{ROUND}\left[\left[\frac{1.645 \cdot (2 \cdot 0.10 \cdot 0.90)^{\frac{1}{2}} + 1.282 \cdot (0.07 \cdot 0.93 + 0.13 \cdot 0.87)^{\frac{1}{2}}}{0.06}\right]^2\right] = 426 ,$$

per cui vanno indagati almeno 426 ragazzi delle prime e seconde e 426 delle ultime due classi.

Se si ritiene che questa ricerca, tendente a scoprire una eventuale differenza tra due sottogruppi di studenti, sia più particolare, potrebbe imporsi un livello di significatività "più rigoroso", ad esempio $\alpha = 0.01$ e, a parità di potenza, la numerosità di ciascuno dei due campioni dovrebbe essere di almeno 649, come risulta dal calcolo seguente

$$n = \mathrm{ROUND}\left[\left[\frac{2.328 \cdot (2 \cdot 0.10 \cdot 0.90)^{\frac{1}{2}} + 1.282 \cdot (0.07 \cdot 0.93 + 0.13 \cdot 0.87)^{\frac{1}{2}}}{0.06}\right]^2\right] = 649 .$$

Si fa osservare che la numerosità campionaria è inversamente proporzionale al quadrato della differenza δ prefissata, a parità di errori di prima e seconda specie. Così se si dimezza la differenza nella percentuale di casi da scoprire nel campione (rispetto alla popolazione) o tra due campioni, è necessario quadruplicare le dimensioni dell'indagine o ricerca. D'altra parte n aumenta al diminuire dei due tipi di errori, condizione questa abbastanza ovvia in quanto, qualora si voglia minimizzare la probabilità di errore, necessita aumentare la numerosità del campionamento.

5. CONSIDERAZIONI CONCLUSIVE

Con questa breve nota non si ritiene di aver esaurito l'argomento, ma di avere, se non altro, presentato il quesito fondamentale che tutti i ricercatori dovrebbero porsi nel pianificare la loro indagine, e cioè quanti soggetti o oggetti sono necessari per soddisfare gli obiettivi della ricerca. Ciò affinchè i risultati della ricerca stessa siano attendibili e, principalmente, permettano di fare dell'inferenza statistica, ossia di effettuare la verifica delle ipotesi sulla base dei risultati campionari.

Si fa rilevare che il criterio adottato nel lavoro è quello di assegnare una differenza minima tra frequenze relative che si ritiene rilevante, differenza definita in termini di $\delta = \pi_o - \pi_1$ oppure $\delta = \pi_1 - \pi_2$. Si potrebbe specificare un valore δ_o diverso da δ, al fine di cautelarsi maggiormente rispetto agli errori di campionamento. Infatti non si può mai "garantire" una differenza significativa dal momento che le fluttuazioni campionarie possono portare ad un valore di $p - \pi_o$ oppure $p_1 - p_2$ molto più piccolo di δ e, quindi, ad un non-rifiuto dell'ipotesi nulla.

È opportuno far presente, infine, che la determinazione della numerosità campionaria è stata sviluppata nella presente nota secondo un criterio di "ottimalità" ma che tuttavia esistono diverse situazioni pratiche che possono limitare la dimensione ottimale. Ad esempio, per ragioni di carattere economico, etico, ecc., può essere desiderabile che il numero di unità da sottoporre ad esperimento sia il minimo possibile. Si pensi a quando l'esperimento comporta la distruzione dell'unità da controllare oppure quando esso presuppone la somministrazione di un farmaco del quale non è ancora del tutto nota la tossicità sull'uomo.

6. SOFTWARE PER IL CALCOLO DELLA DIMENSIONE CAMPIONARIA. *(a cura di CLAUDIO DI MASCIO)*

Si riporta di seguito il listato di un programma in Turbo Pascal per il dimensionamento di campioni al fine di effettuare verifiche d'ipotesi su frequenze relative.

La metodologia utilizzata per la costruzione del programma è quella della programmazione strutturata, in quanto molto efficace da un punto di vista didattico.

Le varie procedure sono facilmente interpretabili. Va osservato comunque che il programma utilizza la Unit esterna "STAT" per la determinazione dei valori $z_{1-\alpha}$ e $z_{1-\beta}$ della variabile normale standardizzata. La Unit è stata costruita con la tecnica della Programmazione Orientata agli Oggetti (OOP). In essa viene definito l'oggetto "normale", risultato dell'"icapsulamento" di dati (i parametri di una distribuzione normale) e metodi (uno per il calcolo dell'area al di sotto della curva normale e un secondo per la determinazione del valore x_o della variabile normale tale che la Prob $(a < x < x_o)$ sia uguale ad un valore prefissato). Tale Unit può essere vista come una specie di "cassetta di strumenti" statistici incrementabile volta volta e utilizzabile per la soluzione di problemi.

Il programma richiede in input i seguenti dati:

- livello di significatività α e potenza del test $1 - \beta$;
- differenza δ che si ritiene rilevante;
- frequenza relativa π_o nota o stimata nella popolazione, se trattasi di inferenza su un campione; frequenze relative (π_1 e π_2) delle due popolazioni analizzate, se trattasi di inferenza su due campioni (in questo caso $\delta = \pi_1 - \pi_2$).

PROGRAM DIMENSIONE_CAMPIONARIA;

```
uses crt, STAT;
var     N                          :normale;
        alpha,potenza,beta         :real;
        z1,z2,po,p,d,p1,p2         :real;
        num                        :integer;
        ch                         :char;

Function SCELTA (scritta:string; ch1, ch2: char): char;
var     x,y:word;
        ch :char;
begin
write(scritta); x: = wherex; y: = wherey;
repeat
        gotoxy (x, y); clreol; ch: = readkey; ch: = upcase (ch); write (ch)
until (ch=ch1) or (ch=ch2);
SCELTA: = ch;
end;
Procedure RIGHE (R:WORD);
var i:word;
begin
        for i: = 1 to R do writeln
end;
Procedure ERRORI;
var x,y: word;
begin
        RIGHE (2); write ('Livello di significatività α = ');
        x: = wherex;  y: = wherey;
        repeat
            gotoxy (x, y); clreol; read (alpha)
        until (alpha < 0.5) and (alpha > 0);
        write(' Potenza del test 1 - β = ');
        x: = wherex;  y: = wherey;
        repeat
            gotoxy (x, y); clreol; read (potenza);
        until (potenza > 0.5) and (potenza < 1);
        RIGHE (1); x: = wherex; y: = wherey;
        gotoxy (1,24); write ('Attendere Prego...!');
        z1: = N.z (-4*N.s + N.m, 4*N.s + N.m, 1 - alpha);
        z2: = N.z (-4*N.s + N.m, 4*N.s + N.m, potenza);
        gotoxy (1,24); delline; gotoxy (x,y);
        writeln ('z1- α = ',z1 : 6 : 3); writeln ('z1 - β = ',z2 : 6 : 3);
end;
Procedure CAMPIONE;
var     x, y   :word;
        ch     :char;
begin
        RIGHE (3); writeln ('CASO DI UN CAMPIONE' :45);
writeln; writeln;
        write ('Frequenza relativa πo = '); readln (po);
        write ('Differenza 'rilevante' δ = '); readln (d);
```

```
        RIGHE (1);
        p: = po + d;
end;
Procedure DIM_CAMPIONE;
var calc: real;
begin
        RIGHE (1); d: = ABS (d);
        calc: = (z1*sqrt (po*(1 - po)) + z2*sqrt (p*(1-p))) / d;
        num: = ROUND           (calc*calc);
        gotoxy (23,22); write ('DIMENSIONE CAMPIONARIA n = ', num);
end;
Procedure CAMPIONI;
var     x, y  : word;
        ch    :char;
begin
        RIGHE (3); writeln ('CASO DI DUE CAMPIONI': 45);
writeln; writeln;
        write ('Frequenza nella Prima Popolazione π1 = '); readln (p1);
        write ('Frequenza nella Seconda Popolazione π2 = ');
readln (p2);
        d: = p1 - p2;
        write ('Differenza che si ritiene 'rilevante' δ = ', d : 4 : 2);
end;
Procedure DIM_CAMPIONI;
var calc, media: real;
begin
        RIGHE (1); media: = (p1 + p2) / 2; d: = ABS (d);
        calc: = (z1*sqrt (2*media*(1 - media)) + z2*sqrt (p1*(1- p1)
                + p2*(1 - p2))) / d;
        num: = ROUND (calc*calc);
        gotoxy (23, 22); write ('DIMENSIONE CAMPIONARIA n = ', num);
end;

{MAIN BODY}
begin
        repeat
                clrscr; N.m: = 0; N.s: = 1;
                writeln ('CALCOLO DELLA DIMENSIONE MINIMA DI UN CAMPIONE': 60);
                ERRORI; RIGHE (1);
                ch: = SCELTA ('Inferenza su (U)no o (D)ue campioni : ', 'U', 'D');
                if ch = 'U' then
                                begin CAMPIONE; DIM_CAMPIONE; end
                        else
                                begin CAMPIONI; DIM_CAMPIONI; end;
                gotoxy (1, 25);
        until SCELTA ('Vuole Continuare ? (S/N) ', 'S',' N') = N'
end.
```

Il listato della Unit STAT è il seguente.

```
UNIT STAT;

Interface
```

```
type
        normale = object
                m,s: real; {parametri della distribuzione normale}
                function f (x: real) : real;
                function area (a, b: real) : real;
                function z (a, b, alpha: real): real;
        end;

Implementation

function pot (base, esponente: word): real;
var    i:  word;
       p:  real;
begin
       p: = 1; for i: = 1 to esponente do p: = p*base;
       pot: = p
end;

function normale.f(x:real):real;
begin
       f: = exp (- (x - m)*(x - m) / (2*s*s)) / (s*sqrt (2*Pi));
end;

function normale.area;
{Calcolo dell'area al di sotto della curva normale utilizzando
 il metodo di integrazione di Simpson}
var
       h, r, l, p, int1        : real;
       n                       : word;
begin
       n: = 100; h: = (b - a) / n; r: = a; int1: = 0;
       repeat
              l: = r + 2*h; p: = r + h;
              int1: = int1 + ((l - r) / 6)*(f (r) + f (l) + 4*f (p));
              r: = l
       until r>=(b-h);
       area:=int1
end;

function normale.z;
{Risoluzione dell'equazione Prob (a < x < xo) = alpha nell'incognita xo col metodo di
Newton}
var
       y, y1, e: real;
       i : word;
              function g (x: real): real;
              begin
                      g: = area (a, x) - alpha
              end;
```

```
begin
        i: = 1; y1: = (a + b) / 2;
        repeat
                y: = y1 - g (y1) / f (y1); e: = abs (y - y1);
                y1: = y; i: = i + 1
        until (i > 100) or (e < 1 / pot (10, 3));
        z: = y
end;
end.
```

BIBLIOGRAFIA

[1] M.L. BERENSON - D.M. LEVINE (1989), "Statistica per le Scienze Economiche", Zanichelli, Bologna.

[2] J. BISHOP (1990), "Pascal - Corso di programmazione", Addison-Wesley Masson, Milano.

[3] B.M. CESANA - E. MARUBINI, "Dimensione del campione nella ricerca bio-medica", Collana "Centro Zambon" n. 8, Milano.

[4] B.J. COX (1990), "Programmazione Orientata agli Oggetti", Addison-Wesley Masson, Milano.

[5] M. R. SPIEGEL (1992), "Statistica", Etas Libri, Milano.