

# Mostres

*Josep Alsinet i Caballeria; Carles Barceló i Vidal; Antoni Gomà i Nasarre*  
Programa d'Informàtica Educativa, 1988.

## 1. ESPECIFICACIONS GENERALS

- 1.1. Nom del programa
- 1.2. Autor
- 1.3. Temàtica
- 1.4. Assignatura
- 1.5. Nivell escolar
- 1.6. Descripció tècnica

## 2. INSTRUCCIONS DE FUNCIONAMENT

- 2.1. Dinàmica de funcionament
- 2.2. Opcions

## 3. ASPECTES PEDAGÒGICS

- 3.1. Objectiu del programa
- 3.2. Coneixements previs
- 3.3. Fonamentació teòrica
- 3.4. Metodologia d'ús

## 4. GRÀFICS

# 1. ESPECIFICACIONS GENERALS

## 1.1. Nom Programa

MOSTRES

## 1.2. Autors

Seminari Permanent d'Estadística de l'I.C.E. de la Universitat Autònoma de Barcelona:

- Josep ALSINET i CABALLERIA, professor de l'I.E.S. del Vallès, Sabadell.
- Carles BARCELÖ i VIDAL, professor de l'I.B. "Salvador Espriu", Salt.
- Antoni GOMÀ i NASARRE, professor de l'I.B. "Joanot Martorell", Esplugues de Llobregat.

## 1.3. Telemàtica

Introducció a les tècniques de mostreig.

Des d'un punt de vista gràfic i pràctic el programa tracta de reforçar els conceptes relacionats amb l'estimació de la proporció real amb què es manifesta determinat esdeveniment en una població, feta aquesta estimació a partir de l'observació de la proporció de l'esdeveniment en una mostra aleatòria.

## 1.4. Assignatura

Matemàtiques.

Tanmateix no podem deixar d'esmentar la creixent importància de les tècniques

estadístiques, sobre tot la del mostreig, en la seva aplicació a les ciències socials i de la salut.

### **1.5. Nivell escolar**

Actualment les tècniques de mostreig no apareixen explícitament, als programes de secundària. Ara bé, és fonamental que els alumnes adquireixin visió crítica d'aquelles informacions que poden obtenir a través dels mitjans de comunicació. Trobaríem una tècnica estadística més usada que les enquestes, és a dir el treball amb mostres?.

Pensem, doncs, que els conceptes exposats en aquest programa tindran el seu lloc d'introducció per als alumnes de 14-16 anys i s'hauran de treballar amb més rigor al futur batxillerat (16-18). De fet, creiem que el programa té una clara aplicació a les matemàtiques-II de COU.

### **1.6. Descripció tècnica**

L'arxiu que conté el programa es denomina MOSTRES.EXE i és escrit en llenguatge GW-BASIC i compilat posteriorment. Requereix tenir accés a l'arxiu BRUN20G.EXE i fa ús d'uns arxius auxiliars, ARX10.TXT, ARX20.TXT, ARX30.TXT, ARX50.TXT, ARX100.TXT que han d'estar ubicats al mateix disquet o subdirectori des d'on es cridi el programa, que caldrà executar amb la comanda MOSTRES.

## **2. INSTRUCCIONS DE FUNCIONAMENT**

### **2.1. Dinàmica de funcionament**

El programa, essencialment, consisteix en la simulació d'una selecció de mostres successives d'una població, els elements de la qual manifesten un de dos caràcters, [verd] o [vermell], i la visualització de les proporcions mostrals de l'esdeveniment [verd]. En aquest sentit, comporta més observació que participació per part de l'usuari. Totes les instruccions de funcionament apareixen clarament en pantalla.

La simulació no és pas interactiva i, doncs, l'usuari en serà espectador: se li mostrarà primer de tot la població amb què es treballarà (que podrà ser canviada en qualsevol moment per recomençar el treball amb una altra població) però aquesta població "desapareixerà" perquè no és pas finalitat del programa l'observació de la població total.

(Vegi's el gràfic 1).

Seguidament, amb lleugeres variacions segons l'opció de treball que s'hagi escollit, el programa anirà presentant successives mostres, seleccionades aleatòriament de la població amb què es treballi i farà visuals les proporcions mostrals de l'esdeveniment [verd] tot elaborant el diagrama de barres de la variable estadística "proporció de [verd] en les mostres seleccionades". Haurà de ser l'observació d'aquest diagrama la

que permeti millorar la comprensió dels conceptes de les tècniques de mostreig. (Vegi's el gràfic 2).

Per cada població i cada opció de treball està previst, com a norma general, que l'usuari pugui triar tres mides diferents per a les mostres que s'aniran seleccionant i que indiqui quantes mostres successives vol que el programa li presenti. Tanmateix, en alguna ocasió i segons com s'hagin esdevingut els diagrames, a fi i efecte que la presentació gràfica sigui correcta, el mateix programa deturarà l'exemple en curs.

## 2.2. Opcions

El programa ofereix quatre opcions de treball perfectament diferenciades, lligades les tres primeres amb cadascun dels conceptes que hom intenta aprofundir. L'altre opció és un exemple d'aplicació. A saber:

- a) Observacions de proporcions en mostres.
- b) Treball amb la proporció real coneguda: interval de tolerància.
- c) Treball amb la proporció real no coneguda: interval de confiança.
- d) És el [verd] un factor de risc per [tenir X] ?

La primera opció simplement mostra el diagrama de les proporcions mostrals de [verd] i permet que l'usuari tregui les seves pròpies conclusions a la vista del diagrama que es va construint a la pantalla. En acabar la selecció de mostres d'una mida determinada l'usuari ha d'indicar en quin interval creu que han quedat "la majoria" de les proporcions mostrals observades.

En la segona opció, d'antuvi es mostra la proporció real de [verd] en la població i, amb cada mida de les mostres, s'assenyala sobre el gràfic l'interval de tolerància corresponent, tal com s'explica als aspectes pedagògics (secció 3), a fi i efecte de poder observar com, realment, la majoria de proporcions de les mostres seleccionades queden dins de l'interval de tolerància. El programa en dóna una confirmació numèrica. (Vegi's el gràfic 3).

La tercera opció de treball és la que ens col·loca en una situació més semblant a la real. Hom extreu mostres sense conèixer la proporció real en la població. El programa fa visual l'interval de confiança de l'estimació de la proporció real a través de les proporcions en les mostres i això amb tres mides de mostres diferents. Per acabar l'opció s'indica el valor de la proporció real de [verd] en la població que, excepte situacions anòmales provocades per l'atzar, quedarà sempre dins dels intervals de confiança de les estimacions fetes mitjançant mostres. (Vegi's el gràfic 4).

La quarta opció de treball és un exemple pràctic. Ara els elements de la població poden o no manifestar un cert caràcter [tenir X] i es tracta de treballar amb la proporció real de [verd] i el corresponent interval de tolerància coneguts i llavors, tot prenent mostres relatives, únicament, als elements que manifesten [tenir X], donar ajuts per contestar a la pregunta formulada en l'opció, com s'explica amb més deteniment a l'apartat següent.

### 3. ASPECTES PEDAGÒGICS

#### 3.1. Objectiu del programa

Ajudar el professor, amb la possibilitat de prendre moltes mostres aleatòries d'una població, a fer entendre als alumnes els conceptes d'interval de tolerància i d'interval de confiança en l'estimació, mitjançant tècniques de mostreig, de la proporció amb la que un esdeveniment apareix en una població.

Així doncs, hem de dir que aquesta unitat didàctica s'encamina molt més a donar facilitats per adquirir la idea intuïtiva, això sí, clara i correcta, que no pas a introduir conceptes rigorosos.

#### 3.2. Coneixements previs

El plantejament que es fa del tema requereix únicament que l'usuari estigui familiaritzat amb el llenguatge de la probabilitat i entengui què representa seleccionar una mostra aleatòria d'una població. Convé saber, també, interpretar els diagrames de barres / histogrames.

Tanmateix, com ja es pot deduir d'allò que s'ha dit anteriorment, serà convenient que el programa vagi acompanyat de les explicacions per part del professor dels conceptes que es treballen.

#### 3.3. Fonamentació teòrica

Quan prenem moltes mostres d'una determinada mida partint d'una mateixa població els resultats de la proporció mostral de l'esdeveniment estudiat tendeixen "a concentrar-se". És important poder veure que és ben possible seleccionar un interval on hi hagi "la majoria" d'observacions, tot eliminant aquelles que intuïtivament hom pugui entendre que no responen "al funcionament normal de l'atzar". Aquesta idea és la que prepara la introducció del primer concepte que treballa el programa: el de l'interval de tolerància.

És important esmentar ara (encara que no es comenti després amb els alumnes) que el model teòric o distribució de probabilitat que s'associa amb aquesta presa de mostres, si pensem en el nombre d'elements de la mostra que manifesten el caràcter estudiat, és una distribució binominal de mitjana  $np$  essent  $p$  la proporció real de l'esdeveniment en la població i  $n$  la mida de la mostra. La mitjana d'una distribució de probabilitat rep també el nom de "valor esperat" perquè és un bon predictor del resultat que hom estima que succeirà en les diverses realitzacions de l'experiència. En el cas que ens ocupa això vol dir que podem estimar que la proporció en la mostra serà  $np / n = p$ , és a dir la mateixa que la proporció real en la població. Ara, les mateixes lleis de la probabilitat ja ens expliquen quin tipus de "garanties" podem tenir en aquesta predicció.

Aquesta idea és la que porta a la definició de l'interval de tolerància amb un risc del 5%:

És aquell interval  $[p-t, p+t]$  tal que la probabilitat que la proporció mostral quedi en aquest interval sigui igual o superior al 95%.

Així, sempre que el programa treu un missatge que esmenta "l'interval de tolerància de [verd]" haurem d'entendre que ens referim a l'interval de tolerància, amb risc del 5%, que la mostra seleccionada tingui una proporció de [verd] pertanyent a aquest interval, centrat en la proporció real de [verd] en la població amb què es treballa.

Ara bé, el camí de la inferència estadística és invers a aquest. No treballarem amb la proporció real coneguda sinó que tendirem a inferir el valor real de la proporció de l'esdeveniment estudiat prenent com a base el valor de la proporció observada en la mostra. En aquesta situació es defineix "l'interval de confiança del 95%" com l'interval format per totes aquelles possibles proporcions reals tals que la proporció mostral pertanyeria al corresponent interval de tolerància amb risc del 5%.

Així quan, en l'opció 3 de treball, el programa dibuixa "l'interval de confiança de la mitjana de proporcions" observada en les mostres que s'han seleccionat, allò que en realitat cal entendre és que, amb risc d'error del 5%, hom pot inferir que la proporció de l'esdeveniment [verd] en la població estudiada pertany a aquell interval.

### **3.4. Metodologia d'ús**

El nostre suggeriment metodològic d'ús del programa va ben lligat a un determinat plantejament didàctic que podria començar amb la simulació a la classe, amb un ampolla de mostreig -o una bossa amb fitxes de colors-, de la presa de mostres d'una població per tal d'observar què succeeix amb les proporcions mostrals de l'esdeveniment estudiat. De fet, el programa farà això mateix amb molta més rapidesa.

Per al treball informàtic pot ser bo un ús inicialment col·lectiu, encara que, més endavant, serà interessant que els alumnes puguin experimentar-hi individualment o en petits grups.

#### **3.4.1. Introducció dels conceptes d'interval de tolerància i d'interval de confiança**

Comencem per fer veure als alumnes què passa si s'extreuen moltes mostres d'una població. Primer amb mostres de la mateixa mida i després comparant els resultats entre mostres d'una mida i mostres d'una mida més gran.

En aquesta opció no hi ha (apart de les derivades de la presentació gràfica) cap limitació en el nombre de mostres que hom pot fer visuals. És per això que també podrà servir el programa per donar la idea intuïtiva que, quan es va repetint molts cops l'experiència, el diagrama que representa les proporcions mostrals tendeix a acostar-se a un model ben concret que, en el seu moment, ja serà presentat com la distribució binominal que, per altra banda, en la majoria de situacions que presenta el programa pot ser aproximada per la distribució normal.

Quan s'hagi acabat l'extracció de mostres d'una mida, l'usuari haurà d'indicar quin és l'interval on ha quedat "la majoria" de proporcions mostrals, eliminant, de forma ben intuïtiva, les que es puguin considerar "anòmales".

Convé que l'alumne s'adoni que, quan augmentem la mida de la mostra, aquest interval

es fa cada vegada més petit. Per això el programa ens permetrà treballar amb diverses mides. Finalment, quan haguem acabat la fase d'observació es podrà comprovar com el valor real de la proporció de l'esdeveniment estudiat en la població total queda dins dels intervals seleccionats.

Per donar la màxima flexibilitat en el treball d'aquest i dels altres apartats, per cada població i per cada tipus de treball que es vol fer amb ella, l'usuari escollirà mostres de 3 mides diferents i, per cada mida, quantes mostres vol considerar.

Després del treball comentat és probable que l'alumne hagi adquirit ja el concepte d'interval de tolerància i podrem passar a la part central del programa (opcions 2 i 3).

Amb l'opció 2 treballarem coneixen la proporció real i l'interval de tolerància corresponent a la mida de mostres seleccionada. Després d'observar un nombre relativament gran de mostres de cada mida és probable que alguna de les proporcions mostrals hagi quedat fora de l'interval de tolerància. En acabar la fase d'observació d'aquesta opció serà mostrat en pantalla el percentatge de proporcions mostrals que han quedat dins de l'interval de tolerància i, com que els intervals de tolerància ho son amb risc del 5%, convé prendre un gran nombre de mostres a fi i efecte que quedi clar el concepte.

El pas, conceptualment ben important d'interval de tolerància a interval de confiança haurà de ser explicat pel professor. Aquest és el que convé aplicar en casos d'enquestes -presa de mostres- reals quan volem tenir una certa mesura del grau de certesa que podem donar a les nostres conclusions. Per fer-ho més semblant a un cas real, quan treballem aquesta part és aconsellable prendre un nombre reduït de mostres. Acabada la presa de mostres de cadascuna de les mides es dibuixen els corresponents intervals de confiança (prenent com a estimació el valor de la mitjana de les proporcions si s'ha pres més d'una mostra). El programa mostra finalment el valor real de la proporció estudiada en la població total i així es pot comprovar si aquest valor queda dintre dels intervals de confiança, cosa que succeirà quasi sempre (pensem que el risc és, només, del 5%).

Vegi's el gràfic 5 que transcriu un bolcat de pantalla corresponent a aquesta opció de treball.

### **3.4.2. Exemple en relació al "factor de risc" (opció 4 del programa)**

S'ha pensat que era convenient poder fer un exemple d'aplicació, en el mateix programa, dels conceptes treballats.

Sovint sentim a dir frases del tipus "l'homosexualitat és un factor de risc per tenir la SIDA" o bé "l'hàbit de fumar entre la classe mèdica és significativament més alt que en el conjunt de la població". Quin plantejament teòric tenen afirmacions com aquestes?.

Per tal de poder-hi treballar, des del punt de vista intuïtiu com en la resta del programa, hom mostrarà, en la mateixa població amb què es treballa, com a caràcter suplementari, [tenir X] o [no tenir X].

Així, tindrem quatre subpoblacions:

[verds amb X], [verds sense X], [vermells amb X], [vermells sense X], i voldrem estudiar si el caràcter [verd] és un "factor de risc" per "tenir X". (Vegi's els gràfic 6).

Caldrà suposar que és coneguda la proporció de [verd] en la població total (i, doncs, es mostrarà de seguida) i els intervals de tolerància en l'extracció de successives mostres. Llavors es durà a terme el treball d'extracció de mostres, restringides a la població que manifesta el caràcter [tenir X]. Com s'ha dit per a l'opció 3, a fi i efecte de fer una actuació semblant a la de situacions reals, serà recomanable triar un nombre ben reduït de mostres de cada mida. L'acostumada visualització de resultats permetrà decidir la resposta al nostre estudi. (Vegi's el gràfic 7).

Perquè es diu que "l'homosexualitat és un factor de risc per tenir la SIDA"? Perquè si prenem una mostra de malalts de SIDA, la proporció en ella d'homosexuals és "molt més alta" que la proporció d'homosexuals en la població total. Rigoritzar aquesta idea de "molt més alta" porta a dir que la proporció en la mostra queda fora de l'interval de tolerància de la proporció del caràcter estudiat en la població total.

Semblantment podrem treure conclusions -no massa alentadores- si veiem que en una mostra aleatòria presa d'entre la classe mèdica la proporció de fumadors no solament és més gran que la proporció de fumadors en la població total sinó que, fins i tot, queda fora de l'interval de tolerància.

Queda dit, doncs, que perquè puguem respondre afirmativament sobre el fet que el caràcter [verd] és un factor de risc per [tenir X], caldrà que les mostres considerades (restringides a la subpoblació que té X) ens donin proporcions de [verd] no incloses a l'interval de tolerància. Només si això passa, i no solament si hi ha una discrepància entre unes proporcions i l'altra, voldrà dir que podem parlar d'un "factor de risc" o lligam especial entre els caràcters estudiats.

Per tal de comprovar si les conclusions tretes de la part gràfica de presentació de proporcions mostrals concorden amb la realitat, el programa facilita una sortida numèrica de proporcions reals, tal com es pot veure a gràfic 8. Se'ns dona:

\* La població real de [verd] en la població total i l'interval de tolerància corresponent a la mida de mostres considerada en darrer lloc (que és recomanable que sigui la més gran).

\* La proporció de [verd] restringida a la subpoblació que manifesta el caràcter [tenir X].

\* La proporció de [verd] observada en les mostres considerades en darrer lloc, que ho són restringides a aquells que [tenen X].

La denominació de "factor de risc" es refereix, és clar, a aquells casos en els que les proporcions observades (relatives a [tenir X]) són més grans que les de l'interval de tolerància i, a més ens fan pensar que [tenir X] és un esdeveniment "dolent". Però el mateix concepte s'aplica a tots els casos de "comportament molt anòmal" de les proporcions relatives a la població que [té X] respecte les proporcions de [verd] en la població total, cosa que fa pensar en una especial associació entre aquests dos caràcters.

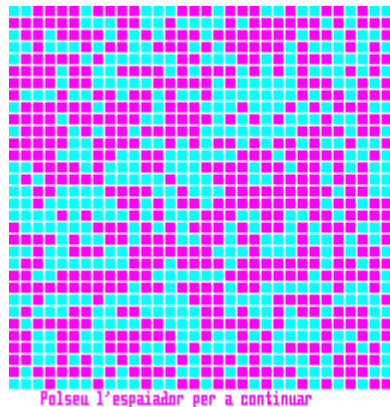
### 3.4.3. La tria de població

Per acabar la descripció metodològica del programa ens queda per dir com és triada en cada cas la població i, si s'escau, el tenir X o no.

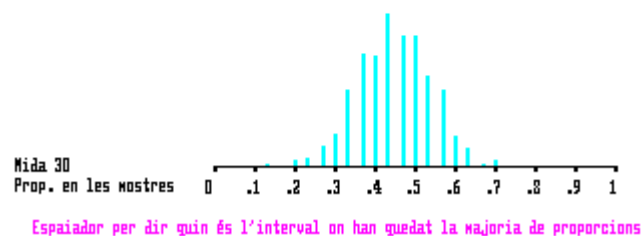
Per tal de decidir el color de cadascun dels 1024 individus el programa tria aleatòriament un PROBABILITAT i va escollint valor de RND. Si el valor surt menor que la PROBABILITAT, serà de color verd; si surt major, vermell.

En cas d'haver de posar X a alguns individus s'actua, una vegada ja triada la població d'una forma semblant. Es trien PROB1 i PROB2. Quan considerem un individu amb color verd, si el RND que es tria és menor que PROB1, aquell individu tindrà X; si és major, no. Si l'individu és de color roig s'actua semblantment amb PROB2.

És fàcil que algú plantegi si podria ser interessant que l'usuari pogués triar aquests valors PROBABILITAT, PROB1, PROB2. No s'ha fet així perquè creiem que no donen cap valor afegit al programa ni permeten cap variació teòrica. Creiem, en canvi, que és interessant l'aleatorietat.

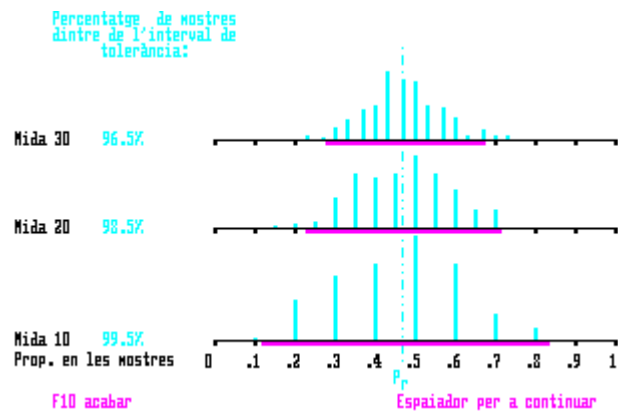


Gràfic núm 1. Presentació de la població amb què es treballa.

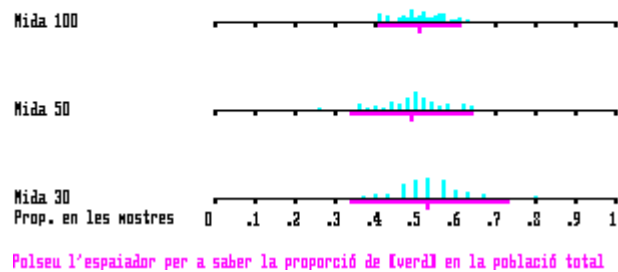


Gràfic núm 2. Opció 1 de treball. Possibilitat de selecció de moltes mostres de la mateixa mida. Es pot veure en pantalla una de les mostres seleccionades.

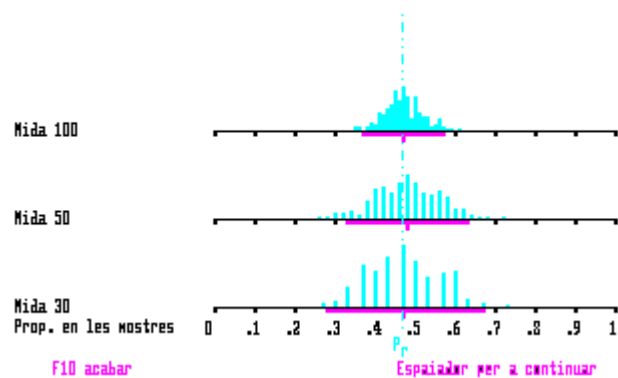




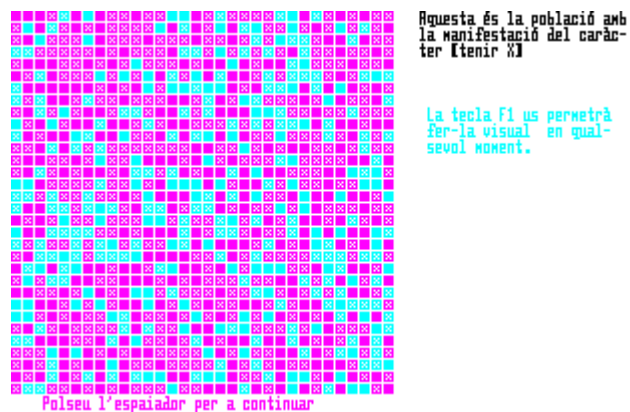
Gràfic núm 3. Opció 2 de treball. La proporció real de [verd] és coneguda i es mostren els percentatges de proporcions que han quedat dins de l'interval de tolerància.



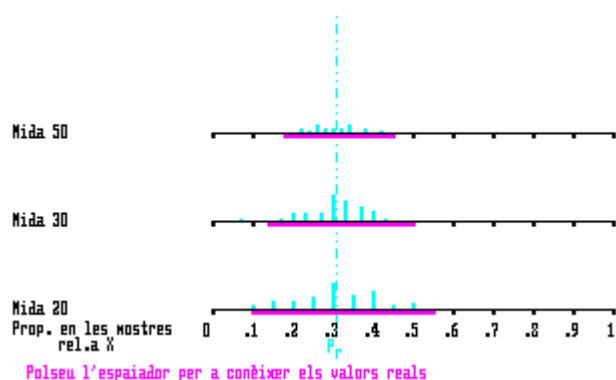
Gràfic núm 4. Opció 3 de treball: interval de confiança.



Gràfic núm 5. Opció 3 de treball: finalment es mostra el valor real de la proporció de verd en la població total.



Gràfic núm 6 Presentació de la població amb el caràcter suplementari [tenir X].



Gràfic núm 7. Opció 4 de treball en un cas en què [verd] sí es factor de risc per [tenir X].

	Prop.[verd]
Població total	31%
relativa a [%]	31%
mida 50	30%

Int. tol. [ 18% , 45% ]  
amb mida 50

Comproveu si les vostres deduccions havien estat correctes.

F10 acabar

Espaiador per a continuar

Gràfic núm 8. Sortida numèrica de dades en el mateix cas del gràfic anterior.