

Correlació i regressió lineals

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

1. ESPECIFICACIONS GENERALS

- 1.1. Nom del programa, temàtica i autors
- 1.2. Assignatura i nivell escolar
- 1.3. Descripció tècnica

2. INSTRUCCIONS DE FUNCIONAMENT

- 2.1. Estructura dinàmica
- 2.2. Opcions

3. ASPECTES PEDAGÒGICS

- 3.1 Objectiu del programa
- 3.2 Coneixements previs
- 3.3 Comentaris matemàtics
- 3.4 Metodologia d'ús

1. ESPECIFICACIONS GENERALS

1.1. Nom, temàtica i autors del programa

Correlació i regressió lineals és un programa d'interpretació i exercitació gràfica dels conceptes de correlació i regressió lineals d'una distribució conjunta de dues variables numèriques, realitzat pels membres del Seminari Permanent d'Estadística de l'ICE de la UAB: Josep Alsinet i Caballeria, professor de l'IES del Vallès, Sabadell; Carles Barceló i Vidal, professor de l'IB Salvador Espriu, Salt; Antoni Gomà i Nassarre, professor de l'IB Joanot Martorell, Esplugues del Llobregat.

1.2. Assignatura i nivell escolar

En el context actual de l'ensenyament de l'Estadística aquest programa és d'aplicació en l'assignatura de Matemàtiques dels cursos 1er i 3er del BUP i en algunes especialitats de FP de 2on grau. En un futur és de suposar que podrà ésser aplicat en el 1er cicle de l'Ensenyament Secundari Obligatori.

1.3. Descripció tècnica

L'arxiu que conté el programa es denomina CORREGL.EXE i és escrit en llenguatge GW-BASIC i compilat posteriorment. Requereix la presència de l'arxiu BRUN20G.EXE. No es crea cap arxiu extern ni està pensat tampoc per fer servir altres interfícies que no la pantalla.

2. INSTRUCCIONS DE FUNCIONAMENT

2.1. Estructuració i dinàmica de funcionament del programa

El programa es posa en marxa amb el seu nom CORREGL i, després d'unes pàgines de presentació, ofereix dues opcions ben diferenciades i independents:

- 1. visualització de paràmetres.**
- 2. Exercitació i experimentació.**

2.1.1. L'Opció de visualització permet a l'alumne la interpretació gràfica dels conceptes de coeficient de correlació lineal i recta de regressió sobre diferents núvols de punts, corresponents a suposades distribucions conjuntes de dues variables numèriques X i Y .

Amb la intenció de facilitar la interpretació, s'han programat les següents tecles:

- $F1$, que permet visualitzar la posició del centre de gravetat (falta fórmula) de la distribució bidimensional i que dibuixa -centrats en aquest punt- dos eixos paral·lels als eixos de coordenades.
- $F2$, que informa del valor del coeficient de correlació.
- $F3$, que dibuixa la recta de regressió.
- $F4$, que permet a l'usuari canviar la posició d'un punt de la distribució i veure tot seguit els canvis provocats en el valor del coeficient de correlació i en la posició de la recta de regressió.

Per tal de moure un determinat punt del núvol, l'usuari ha de situar el cursor gràfic que apareix en la pantalla damunt del punt, capturar-lo tot prement la tecla *Inici*, resituar el punt amb l'ajuda de les tecles de moviment del cursor, i fixar la nova posició del punt prement la tecla Fi .

Aquesta **Opció de visualització** finalitza al prémer la tecla $F9$ tot demanant a l'usuari si vol continuar amb un nou exemple.

2.1.2. L'Opció d'exercitació ofereix les següents opcions de treball:

- 1. Exercitació sobre el coeficient de correlació.**
- 2. Exercitació sobre la recta de regressió.**
- 3. Exercitació sobre la variació del coeficient de correlació i de la recta de distribució i de la recta de regressió al moure alguns punts de la distribució.**

2.1.2.1. Exercitació sobre el coeficient de correlació.

A la vista del núvol de punts l'usuari és interrogat sobre el signe del coeficient de correlació i, seguidament, sobre el seu valor absolut. Atès que és difícil precisar-ne el valor exacte, la pregunta es formula de manera que l'usuari ha d'optar per estimar si el coeficient de correlació, en valor absolut, és entre 0 i 0.2, entre 0.2 i 0.4, entre 0.4 i 0.6, entre 0.6 i 0.8 o més gran que 0.8.

2.1.2.2 Exercitació sobre la recta de regressió.

Apareix sobre el núvol de punts una recta mòbil que l'usuari pot anar movent tot prement les tecles *suma* (+) i *resta* (-) segons vulgui augmentar o disminuir el pendent de la recta per tal d'aproximar-la a la posició on es suposa està situada la recta de regressió. La recta es mou pivotant sobre el centre de gravetat (falta fórmula) del núvol. L'opció incorpora també la possibilitat -tot prement la tecla $F1$ - de visualitzar els eixos de coordenades centrats en (falta fórmula).

Aquesta opció finalitza quan l'usuari aconsegueix estimar correctament la posició de la recta de regressió o bé quan prem la tecla $F3$ per tal de visualitzar la recta de regressió

de la distribució bidimensional representada en pantalla.

2.1.2.3 Exercitació sobre la variació del coeficient de correlació i de la recta de regressió al moure alguns punts de la distribució.

En aquesta opció l'usuari ha de moure un punt de la distribució i a continuació contestar les preguntes formulades pel programa sobre les variacions -augment o disminució- experimentades pel valor absolut del coeficient de correlació i pel pendent de la recta de regressió.

El procés es repeteix fins que l'usuari manifesta que no vol moure cap més punt de la distribució.

2.2 Opcions del programa.

En l'apartat anterior s'han comentat les dues parts de què consta el programa i s'ha fet referència contínuament al núvol de punts sobre el qual s'aplica tant l'**Opció de visualització** com la **d'exercitació**.

El programa ofereix a l'usuari la possibilitat d'escollir el núvol de punts a manipular:

1. Núvol de punts totalment aleatori.
2. Núvol de punts escollit a l'atzar entre els exemples preparats.
3. Núvol de punts escollit d'entre els exemples preparats.
4. Núvol de punts construït per l'usuari.

2.2.1. L'Opció 1 fa que el programa esculli totalment a l'atzar un núvol de 20 punts dins la finestra gràfica.

2.2.2. L'Opció 2 ofereix a l'usuari la possibilitat de que el núvol de punts tingui:

1. Correlació molt alta; positiva ($r \geq 0.8$).
2. Correlació molt alta, negativa ($r \leq -0.8$).
3. Correlació elevada, positiva ($0.4 \leq r \leq 0.8$).
4. Correlació elevada, negativa ($-0.8 \leq r \leq -0.4$).
5. Correlació nul·la o molt baixa ($-0.2 \leq r \leq 0.2$).

En tots els casos el núvol consta de 20 punts i és escollit d'un repertori d'exemples preparats que incorpora el programa.

2.2.3. L'Opció 3 es limita a escollir a l'atzar un dels núvols exemple de què disposa el programa.

2.2.4. L'Opció 4 ofereix a l'usuari la possibilitat de dissenyar la seva pròpia distribució tot indicant sobre la pantalla la posició dels punts que la configuren. El núvol ha de constar d'un mínim de 2 punts fins un màxim de 20, amb l'única precaució de que els punts no estiguin situats tots ells sobre la mateixa vertical.

En el centre de la finestra gràfica apareix un cursor gràfic i la col·locació dels punts es realitza amb l'ajuda de les tecles de moviment del cursor -que permeten situar el cursor gràfic en el lloc desitjat- i amb la tecla *Fi* que dibuixa el punt en el lloc indicat.

3. ASPECTES PEDAGÒGICS

3.1. Objectiu del programa

El programa ha estat dissenyat amb la finalitat de facilitar la comprensió intuitiva i gràfica dels conceptes de correlació lineal i recta de regressió d'una distribució bidimensional.

3.2. Coneixements previs

L'execució d'aquest programa pressuposa que l'alumne ha d'estar mínimament familiaritzat amb els següents conceptes estadístics:

- Distribució conjunta de dues variables numèriques.
- Representació gràfica d'una distribució bidimensional.
- Centre de gravetat d'una distribució bidimensional.
- Coeficient de correlació lineal. Si bé no és necessari conèixer la seva fórmula sí que resulta indispensable tenir una primera idea de la seva significació.
- Recta de regressió. Tampoc cal saber com es calcula tot i que resulta imprescindible que l'alumne tingui una primera idea del seu significat i conegui el fet que passa sempre pel centre de gravetat de la distribució.

3.3 Comentaris d'ordre matemàtic.

3.3.1. El coeficient de correlació que es calcula és el de Pearson, que respon a les conegudes fórmules:

(fórmula 1)

3.3.2. Pel que fa a la recta de regressió es considera la recta de regressió de Y (eix d'ordenades / variable dependent) sobre X (eix d'abscisses / variable independent) que respon a la fórmula:

(fórmula 2)

essent

(fòrmula 3)

3.3.3. Les representacions dels núvols de punts que realitza el programa fan a l'interior d'una finestra gràfica, sense que es dibuixin graduacions damunt els eixos que emmarquen la finestra ja que el que interessa és la posició relativa dels punts de la distribució respecte del centre de gravetat (*fòrmula 4*) d'aquesta.

3.3.4. No és possible que tots els punts d'una distribució estiguin situats sobre una

mateixa recta vertical, ja que en aquest cas la variable independent X prendria només un valor i no tindria sentit parlar de correlació entre ambdues variables.

3.3.5. No està prevista cap mena d'indicació en el cas que dos o més punts d'una distribució ocupin la mateixa posició dins la finestra gràfica.

3.3.6. Els resultats numèrics es presenten amb una aproximació fins als mil·lèsims. Per aquest motiu és possible que, en alguns casos, després de canviar lleugerament la posició d'un punt d'una distribució, els nous valors dels paràmetres \underline{r} i \underline{m} presentats en pantalla coincideixin amb els valors primitius, tot i que en la realitat difereixin en quantitats de l'ordre de 10^{-n} , amb $n > 3$.

3.4. Metodologia d'ús.

El programa està concebut per a la seva utilització individual o en petit grup a l'aula d'informàtica. En la primera fase de visualització dels conceptes de correlació i recta de regressió serà necessària la direcció del professor per passar posteriorment a una activitat individual per part de l'alumne en compliment d'un pla de treball prèviament establert.

3.4.1. 1a fase: visualització dels conceptes.

El professor ha de presentar ordenadament els exemples de què disposa el programa fent veure als alumnes la relació que hi ha entre la disposició dels punts al voltant del centre de gravetat (*fórmula 4*) de la distribució i el valor del coeficient de correlació. A remarcar també la relació entre el signe de \underline{r} i l'orientació del núvol de punts.

Posteriorment caldrà passar a la facilitat que ofereix el programa de dissenyar distribucions bidimensionals de 4 o 5 punts corresponents a situacions extremes de correlació 0 i de correlació 1. És important aprofitar aquests exemples per fer veure als alumnes com, moltes vegades, el desplaçament adequat d'un o dos punts d'una distribució pot fer canviar substancialment el valor de \underline{r} .

Paral·lelament es mostrarà el concepte de recta de regressió com aquella recta que intenta ajustar-se al màxim al núvol de punts de la distribució. Fer veure també com en els casos de correlació molt baixa, el pendent de la recta de regressió té tendència a acostar-se a 0. Important també que l'alumne descobreixi la relació entre el signe del coeficient de correlació \underline{r} i el signe del pendent \underline{m} de la recta de regressió.

3.4.2. 2a fase: exercitació dels conceptes.

En aquesta fase, després que l'alumne pugui comprovar per ell mateix si ha comprès o no els conceptes estadístics que s'estan tractant, és important proposar-li el disseny del núvol de punts que compleixin certs requisits tant pel que fa referència al nombre N de punts com als valors i signes dels paràmetres \underline{r} i \underline{m} . Després del disseny sobre la pantalla de l'ordinador l'alumne pot fer una còpia del núvol de punts dibuixat, acompanyat de la transcripció numèrica del valor dels paràmetres \underline{r} i \underline{m} .

Igualment el professor podrà proposar exercicis consistents en moure el menor nombre de punts d'una determinada distribució -determinada a priori- per tal que el coeficient r

i/o el pendent m de la recta de regressió variïn en un determinat sentit i dins d'un interval numèric fixat amb antelació.