

informatique

un exemple d'enseignement assisté par ordinateur en statistique

*par Alain Baccini, Marie-Pierre Martin
et Robert More
Université de Toulouse-le-Mirail*

Dans cet article, nous présentons un didacticiel utilisé dans l'enseignement de la statistique descriptive en I.U.T. et en DEUG. Le programme informatique réalisé permet d'observer des nuages de points sur un écran afin de déterminer intuitivement les valeurs des caractéristiques statistiques suivantes : moyenne, écart-type, coefficient de corrélation linéaire, paramètres de la droite de régression.*

Après avoir indiqué nos objectifs, nous décrivons le fonctionnement du didacticiel et ses caractéristiques techniques.

D'autre part, en nous appuyant sur cette réalisation, nous envisageons le rôle que peut jouer l'informatique dans l'enseignement ; nous montrons les avantages réels de l'ordinateur, mais aussi les limites pédagogiques de la "machine".

* Didacticiel : programme informatique à but pédagogique.

1. Introduction : objectifs du didacticiel proposé

Pour un étudiant débutant en statistique (notamment dans un I.U.T. ou un premier cycle universitaire), arriver à maîtriser une notion nouvelle telle que l'écart-type, la corrélation linéaire ou la régression linéaire n'est pas aisé. Cela nécessite en général le passage par diverses phases qui nous paraissent être les suivantes :

- approche du concept, de sa signification, de son utilisation, de ses propriétés... ;
- acquisition des mécanismes de calcul permettant de déterminer la valeur de la caractéristique considérée sur une série statistique donnée ;
- entraînement à la détermination intuitive de l'ordre de grandeur de cette caractéristique, à la simple vue d'un graphique approprié.

La première phase, de compréhension générale du concept, est essentiellement abstraite et se fait d'habitude au moyen d'un cours de statistique ou de la lecture d'un ouvrage adapté.

La seconde phase, d'acquisition des mécanismes de calcul, est un des principaux objectifs que l'on se donne en général dans les séances de travaux dirigés ; contrairement à la précédente, elle est très concrète et consiste pour l'étudiant à mener à bien les calculs permettant d'obtenir la valeur de la caractéristique considérée sur divers exemples ; elle complète naturellement la phase précédente.

La dernière phase consiste à faire trouver aux étudiants, sur un certain nombre de cas typiques, l'ordre de grandeur de la caractéristique envisagée, uniquement en observant une représentation graphique. Souvent négligée, cette phase nous paraît pourtant indispensable à la maîtrise complète de la notion considérée. Nous avons ainsi pu observer que des étudiants d'I.U.T., connaissant la formule et les principales propriétés du coefficient de corrélation linéaire (dit de *Bravais-Pearson*) et capables d'en calculer correctement la valeur sur divers exemples, ont beaucoup de difficultés, devant un nuage de points, à indiquer l'ordre de grandeur du coefficient en question, voire son signe. Cette phase nous paraît d'autant plus importante qu'elle est souvent la première approche effectuée par le statisticien lors d'un traitement réel de données.

Bien qu'il soit possible d'effectuer cette phase d'entraînement durant les travaux dirigés, il nous a paru intéressant de l'envisager dans un autre contexte ; en effet, l'ordinateur, pouvant mémoriser de nombreux exemples typiques et "dialoguer" avec l'utilisateur, est un outil approprié pour aider le débutant en statistique à améliorer son approche intuitive de certaines notions ; ces notions sont, pour l'instant, la moyenne arithmétique, l'écart-type, le coefficient de corrélation linéaire et la droite de régression.

L'objectif du didacticiel décrit ici est de présenter aux étudiants (en première année d'I.U.T., département de Statistique, Etudes Economiques

et Techniques Quantitatives de Gestion ou en première année de DEUG Mathématiques Appliquées aux Sciences Sociales), sur un écran, des nuages de points (relatifs à deux variables quantitatives discrètes) et de leur faire trouver :

- la moyenne de chaque variable ;
- l'écart-type de chaque variable ;
- le coefficient de corrélation linéaire entre les deux variables ;
- les paramètres de la droite de régression d'une variable sur l'autre.

Au passage, un rappel de chaque notion intervenant est fourni et un certain nombre de questions concernant leurs propriétés sont posées (tout ceci en option).

2. Description du didacticiel

2.1. Utilisation par l'étudiant

Le didacticiel proposé se compose d'une introduction et de quatre séquences principales parmi lesquelles on doit choisir à partir d'un "menu".

Généralités

La séquence d'introduction présente le didacticiel et précise notations et conventions. Ensuite, les éléments ci-dessous sont communs aux quatre autres séquences :

- un rappel de la définition de la notion considérée est proposé en option ;
- un niveau de difficulté pour le nuage à étudier est à choisir parmi trois possibles (facile, moyen, difficile) ;
- le nuage, tiré au hasard par l'ordinateur parmi ceux disponibles au niveau de difficulté choisi, s'affiche avec un système d'axes ortho-normés sur une moitié d'écran et y demeure toute la durée de la séquence ;
- chaque fois qu'une valeur numérique est demandée, en cas de réponse fausse, une indication est donnée (valeur trop petite ou trop grande) et il est demandé de recommencer ; au bout de trois réponses fausses, la valeur exacte est affichée.

Séquence sur la moyenne

- l'utilisateur peut en option être interrogé sur les deux valeurs extrêmes de chaque variable ; cela lui permet de vérifier si sa perception du nuage est correcte ;
- il est ensuite demandé laquelle des deux variables a la plus grande moyenne (trois réponses sont possibles) ;
- enfin, on demande successivement de situer chaque moyenne dans un intervalle unitaire du type $[2,3]$.

Séquence sur la variance et l'écart-type

- en option, on peut être interrogé sur les propriétés élémentaires de la variance (peut-elle être négative, nulle, dans quels cas ?) ;
- la séquence se déroule ensuite comme celle de la moyenne : valeurs extrêmes en option, comparaison des deux écarts-types, positionnement de chacun dans un intervalle unitaire.

Séquence sur la corrélation linéaire

- une série de questions sur les propriétés élémentaires de la covariance et du coefficient de corrélation linéaire est proposée en option (peuvent-ils être négatifs, plus grands que 1 ?).

On demande ensuite :

- de donner le signe du coefficient de corrélation linéaire (positif, négatif ou nul) ;
- s'il n'est pas nul, de situer sa valeur absolue par rapport à 0,5 ;
- de préciser l'intervalle du type $[0,2;0,3[$ dans lequel celle-ci se trouve ;
- de dire si, en supposant la causalité de X sur Y, il semble intéressant de réaliser une régression linéaire de Y sur X (en transformant éventuellement les données).

Séquence sur la régression linéaire

Il est demandé ici :

- de donner le signe de \hat{a} (coefficient directeur) : est-il positif, négatif ou nul ?
- de choisir entre six intervalles prédéterminés celui dans lequel on situe sa valeur absolue ;
- de préciser l'intervalle unitaire dans lequel on situe \hat{b} (ordonnée à l'origine de la droite de régression).

Compte rendu de la session

Lorsque l'utilisateur sélectionne dans le "menu" initial l'arrêt du didacticiel, la dernière image apparaissant sur l'écran est un tableau rectangulaire donnant le nombre de questions posées et le nombre de réponses exactes pour chaque type de questions et pour chaque séquence ; de plus, le pourcentage de réponses exactes au niveau global est indiqué.

2.2. Préparation par l'enseignant de la bibliothèque de nuages

Un programme particulier, uniquement accessible aux enseignants, leur permet de mettre en mémoire des nuages de points. Ils doivent pour cela indiquer le niveau de difficulté du nuage, le nombre de points, leurs coordonnées, puis pour chaque variable, les valeurs extrêmes, la moyenne et l'écart-type ; ils doivent enfin enregistrer le coefficient de corrélation linéaire entre les deux variables, les deux paramètres de la régression linéaire de la seconde sur la première et indiquer si cette régression est intéressante ou non (éventuellement après transformation des données). L'ensemble de ces nuages est conservé dans des fichiers magnétiques.

3. Caractéristiques techniques du système

3.1. Ressources matérielles et logicielles utilisées

Le matériel utilisé est une installation de taille moyenne : l'ordinateur SOLAR 16* muni d'une mémoire centrale de 192 000 mots de 16 bits, d'unités de disques et de bandes magnétiques, d'une imprimante rapide et de 12 terminaux. Ces terminaux, des écrans-claviers Televideo 925, sont des écrans semi-graphiques : on ne peut pas gérer chaque point de l'écran, mais il est possible d'arrêter le défilement de l'écran et de réaliser des schémas simples grâce à un jeu de caractères spéciaux. Ainsi, la représentation des nuages n'a pas posé de problème ; par contre, nous avons dû renoncer au dessin de la droite de régression.

En ce qui concerne le logiciel de base utilisé, nos programmes ont été réalisés en BASIC et dans un contexte de temps partagé, grâce au système d'exploitation TSM (Time Sharing Monitor) disponible sur ce matériel. Nous n'avons pas utilisé de système de "langage d'auteur" (ce type de logiciel permet aux non-informaticiens d'écrire des didacticiels), alors que notre application s'y prêtait bien, car ce système n'existe pas sur la gamme d'ordinateurs SOLAR 16.

3.2. Description du système réalisé (*)

L'ensemble du système comporte un peu plus de 1100 instructions en langage BASIC. Il est fait de deux programmes :

- * Le premier programme (100 instructions) est utilisé par les enseignants ; il permet de stocker des nuages de points sous forme numérique, avec leurs caractéristiques, dans un fichier à accès direct.

- * Le deuxième programme (1000 instructions) est la partie principale du système ; il propose à l'utilisateur de choisir l'une des séquences décrites en 2.1., puis sélectionne, de manière aléatoire dans le fichier, le nuage à observer. Il gère ensuite l'ensemble des questions et des réponses, confronte les réponses données par l'utilisateur avec les réponses prévues mises dans le fichier, puis enregistre l'ensemble des résultats. Du fait de sa taille, ce programme a été découpé en unités qui, selon le besoin, sont appelées en mémoire centrale. Cependant, le temps de réponse reste bon, les réponses étant pratiquement instantanées.

La réalisation de ce travail a demandé 100 heures de programmation ; pour l'utilisateur, l'emploi de ce système représente à peu près une heure de présence au terminal.

* Ce matériel appartient conjointement à l'Université de Toulouse-le-Mirail et à l'I.U.T. B de Toulouse, rattaché à cette Université.

(*) Le listing correspondant peut être obtenu en s'adressant aux auteurs.

4. Pédagogie et enseignement assisté par ordinateur (E.A.O)

La réalisation de ce didacticiel statistique nous a paru être une bonne illustration des problèmes rencontrés en E.A.O. Nous indiquons ici ces problèmes, quelques difficultés et les raisons de nos choix.

4.1. L'utilisation de l'ordinateur était-elle nécessaire ?

Pour notre application à la statistique, il est clair que tout ce qui se passe au terminal peut être réalisé par l'enseignant avec du matériel traditionnel, donc sans ordinateur.

Cependant, plusieurs éléments nous ont poussés à réaliser ce travail.

*** Les avantages de la mise en œuvre informatique :**

Il faut d'abord mentionner la possibilité pour un utilisateur de travailler à son rythme et sans mobiliser la présence d'un enseignant ; ce double argument est fondamental et peut suffire, à lui seul, à justifier le choix de l'E.A.O.

D'autre part, l'instauration d'un dialogue entre l'étudiant et la machine permet une plus grande liberté et supprime la timidité ou l'inhibition dues à la présence de l'enseignant.

Enfin, une fois les programmes réalisés, plusieurs enseignants peuvent utiliser le système, chacun avec ses données (ici les nuages de points).

*** La motivation des étudiants :**

La plupart d'entre eux sont très motivés par l'utilisation de l'ordinateur et il nous a paru intéressant de profiter de cet engouement dans une matière autre que l'informatique.

*** L'aspect original de ce mode d'enseignement :**

Il est intéressant de présenter un enseignement classique d'une manière nouvelle ; de plus, la statistique descriptive avec ses graphiques et ses calculs, est un bon domaine d'application pour l'E.A.O.

4.2. La fabrication du didacticiel

Elle a été réalisée en équipe, celle-ci comprenant un enseignant en statistique et deux informaticiens.

Dans la pratique, nous avons suivi les étapes suivantes, habituelles dans toute réalisation informatique :

- analyse fonctionnelle : le statisticien a présenté sa demande en expliquant de manière détaillée ses objectifs ;
- analyse organique : en fonction du système disponible, les informaticiens ont proposé une configuration informatique de l'application ;
- programmation : écriture et mise au point des programmes.

Il s'est ajouté à ce travail classique en informatique le problème suivant : une séquence d'E.A.O. n'est pas un simple calcul où les résultats sont justes ou faux ; c'est une succession de situations pédagogiques extrêmement complexes et impossibles à prévoir intégralement. Il est indispensable d'expérimenter ces situations en se mettant au terminal et en n'hésitant pas à remettre en cause leur contenu et parfois même leurs structures. Cet aspect du travail en E.A.O. est très important ; si on le néglige, on obtient un produit finalement inutilisable en enseignement. Nous avons donc plusieurs fois remis en cause notre travail, les modifications intervenant à différents niveaux : présentation de l'écran, contenu des messages, enchaînement logique des questions etc.

Nous pensons procéder à une nouvelle série de retouches, une fois que des utilisateurs suffisamment nombreux auront fait fonctionner le didacticiel (ce n'est pas actuellement le cas). Leurs remarques seront très utiles et surtout, les résultats obtenus par les utilisateurs et affichés sur l'écran en fin de session feront probablement apparaître des imperfections. Cette observation du didacticiel en fonctionnement est très importante ; elle est un moyen efficace de l'améliorer.

4.3. Considérations didactiques

Les notions statistiques qui sont en cause sont déjà familières aux étudiants : ceux-ci ont suivi le cours correspondant et ont manipulé ces notions en travaux dirigés.

Pendant la session au terminal, l'utilisateur est interrogé sur quelques résultats fondamentaux qui sont simplement rappelés, si nécessaire. La séquence au terminal n'apporte aucun concept nouveau, tout le contexte correspondant a été construit préalablement et hors ordinateur.

L'objectif de notre système est donc modeste : il est *d'entraîner* nos étudiants à apprécier visuellement la valeur de certains coefficients statistiques, et par suite, à mieux les maîtriser.

Nous pensons que l'ordinateur peut être un bon outil d'entraînement dans certaines situations pédagogiques ; par contre, en faire un outil de transmission des connaissances se heurterait, nous semble-t-il, à des difficultés bien plus importantes.

4.4. Terminal informatique et communication

Les écrans utilisés ont 23 lignes et 80 colonnes. Nous avons voulu que la communication entre l'ordinateur et l'utilisateur soit très claire, très intelligible. Elle est donc simple et, en général, les écrans ne sont pas trop chargés. Nous avons délimité trois zones distinctes pour les questions, les rappels et le tracé des nuages.

La simplicité de la communication, la clarté des rappels et des questions, l'évidence des situations nous paraissent des conditions obligées en E.A.O. : le respect de ces impératifs nous a guidés, tant lors de la conception du système que lors de ses mises au point successives.

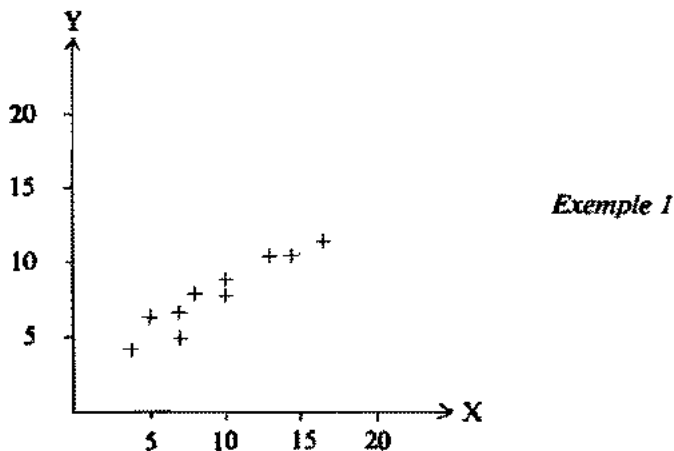
5. Conclusion

En Enseignement Assisté par Ordinateur, la principale difficulté n'est pas d'ordre informatique ; ce qui est difficile et important, c'est de réaliser un programme qui soit une véritable aide à l'enseignement.

En France et dans le monde, des milliers de didacticiens ont été réalisés pour l'ensemble des matières enseignées ; la plupart de ces programmes ont été abandonnés à cause de leur manque d'efficacité ou d'intérêt pédagogique.

Nous avons fait des hypothèses pédagogiques très faibles, l'idée de base étant celle d'entraînement ; il nous reste à vérifier, avec nos populations d'étudiants, que notre objectif, bien que modeste, peut être atteint.

Annexe : Quatre exemples de nuages de points



Nuage de difficulté 1 (facile)

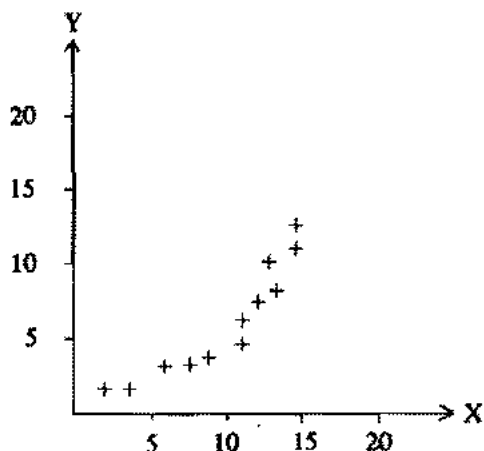
Moyenne de X : 9,5 ; de Y : 7,4 .

Ecart-type de X : 4,01 ; de Y : 2,20 .

Coefficient de corrélation linéaire : + 0,96 .

Est-il intéressant de faire une régression linéaire de Y sur X (en supposant la causalité de X sur Y) : oui, sans transformation des données.

Paramètres de la régression : $\hat{a} \approx 0,53$; $\hat{b} \approx 2,37$.

*Exemple 2*

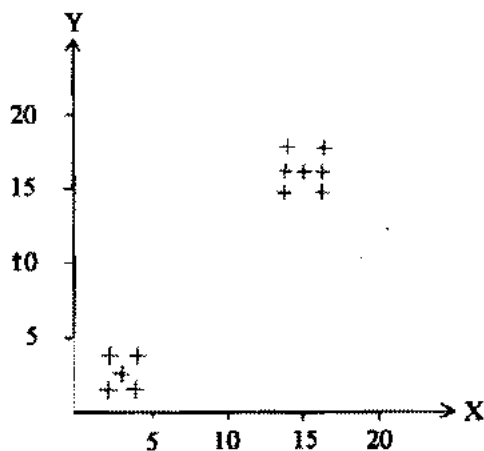
Nuage de difficulté 2 (moyen)

Moyenne de X : 9,67 ; de Y : 6,17 .

Ecart-type de X : 3,88 ; de Y : 3,53 .

Coefficient de corrélation linéaire : + 0,89 .

Est-il intéressant de faire une régression linéaire de Y sur X (en supposant la causalité de X sur Y) : oui, en appliquant à Y une transformation logarithmique.

*Exemple 3*

Nuage de difficulté 3 (difficile).

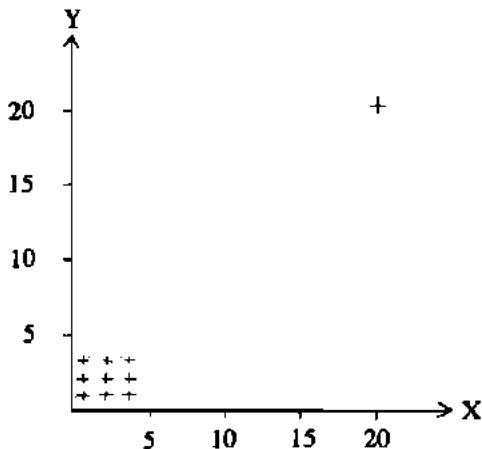
Moyenne de X : 10 ; de Y : 10,58 .

Ecart-type de X : 5,99 ; de Y : 6,46 .

Coefficient de corrélation linéaire : + 0,98 .

Est-il intéressant de faire une régression linéaire de Y sur X (en supposant la causalité de X sur Y) : non, même en transformant les données.

Paramètres de la régression ; $\hat{a} \approx 1,06$; $\hat{b} \approx 0$.



Exemple 4

Nuage de difficulté 3 (difficile).

Moyenne de X : 3,8 ; de Y : 3,8 .

Ecart-type de X : 5,46 ; de Y : 5,46 .

Coefficient de corrélation linéaire : + 0,98 .

Est-il intéressant de faire une régression linéaire de Y sur X (en supposant la causalité de X sur Y) : non, même en transformant les données.

Paramètres de la régression : $\hat{a} \approx 0,98$; $\hat{b} \approx 0,08$.