

LE TRACÉ D'ELLIPSES ÉQUIPROBABLES À L'AIDE D'UN ORDINATEUR (*)

par

R. HAUSPIE, E. DEFRISE-GUSSENHOVEN ET C. SUSANNE

Laboratorium voor Antropogenetica.
Vrije Universiteit Brussel.

Introduction.

Dans le domaine de la biométrie et de la biologie humaine, on a souvent besoin d'exprimer graphiquement la relation entre deux variables ou mensurations. En général, dans ces cas, on fait appel à des diagrammes représentant des nuages de points où les régions de forte densité correspondent à une haute fréquence. (Twisselmann, 1954).

Cependant, une façon élégante de condenser de tels diagrammes, sans perdre l'essentiel de l'information contenue dans les données primitives, consiste à utiliser des ellipses concentriques d'égale fréquence ou équiprobables, enfermant un nombre de plus en plus élevé de sujets du groupe (Defrise, 1955).

Quoique la machine à calculer électronique et l'ellipsographe facilitent les calculs et le dessin des ellipses, ces procédés restent cependant fastidieux. Aussi, nous avons élaboré un programme d'ordinateur qui permet non seulement d'obtenir en un temps très court les résultats des calculs, mais aussi le dessin des ellipses équiprobables avec une très grande précision. La structure souple du programme fait que son application est très facile, même pour les non-programmeurs (**).

(*) Communication présentée le 29 septembre 1975.

(**) Le présent programme a été mis au point pour l'utilisation sur l'ordinateur du Centre de Calcul U.L.B./V.U.B., qui est équipé d'une table traçante BENSON 441. L'emploi du programme sur un autre ordinateur nécessitera néanmoins une adaptation de quelques routines utilisées dans le sous-programme PLOT. La structure et le mode d'emploi des routines actuellement utilisées sont décrits par Glowacki et Meurrens (1971).

Dans les paragraphes suivants nous allons décrire la structure globale du programme, la manière dont on peut s'en servir et les différentes options possibles relatives à l'exécution du graphique.

1. — STRUCTURE DU PROGRAMME

La structure du programme est conçue d'une façon telle qu'on peut traiter simultanément plusieurs échantillons de manière à voir dans quelle mesure ces groupes se superposent (le nombre maximal étant limité à cinq).

Dans l'exemple que nous allons considérer, toutes les informations sont sur des cartes perforées, que ce soient des données proprement dites ou des instructions de programme. Il va de soi que tout autre support d'information pourrait être utilisé, tel que des bandes perforées, des bandes magnétiques ou des disques magnétiques.

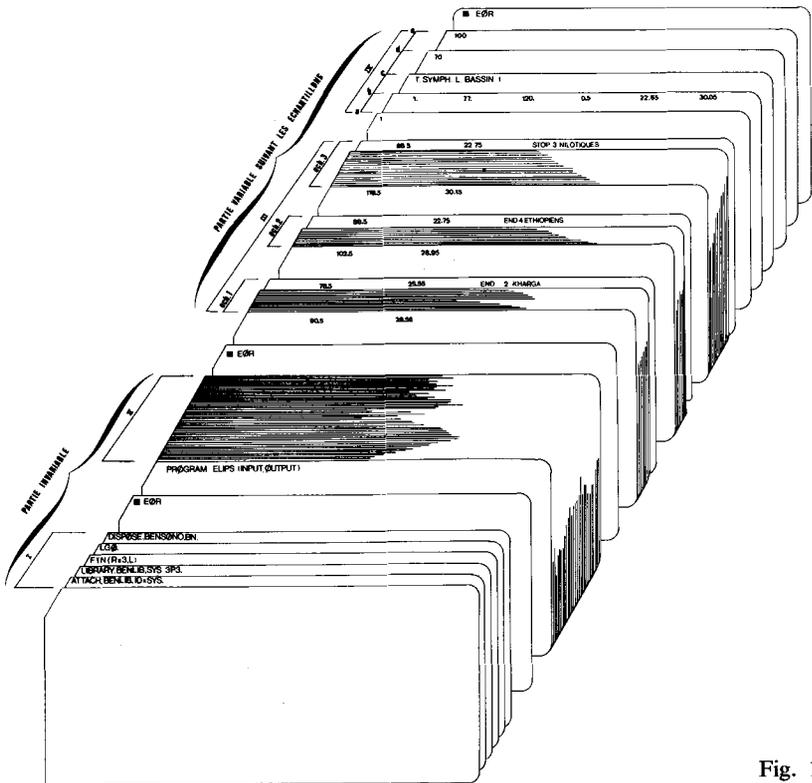


Fig. 1.

L'ensemble des cartes perforées (fig. 1) constituant le «job» à exécuter, contient une partie invariable, notamment quelques cartes de contrôle (I), suivi du programme principal (portant le nom «ELLIPS») et une sous-routine (portant le nom «PLOT») (II). Le premier est responsable des calculs et de la sortie sur imprimante des résultats, le second est responsable de l'exécution du dessin. Le rôle des cartes de contrôle est de donner les informations nécessaires pour la traduction et le chargement des programmes ainsi que le déroulement du «job». Ce sont les seules cartes dont la forme peut légèrement différer d'un centre de calcul à l'autre. L'ensemble des cartes de contrôle, nécessaire pour l'exécution du «job» au Centre de Calcul de l'Université Libre de Bruxelles, est donné en exemple dans la figure 1.

La partie variable du «job» est constituée d'une part, d'une ou plusieurs séries de cartes contenant les données brutes (III) et d'autre part, de cinq cartes (a, b, c, d et e) contenant des codes chiffrés qui déterminent plusieurs options relatives à la représentation du dessin (IV).

Notons que les cartes de contrôle, les programmes et les données proprement dites sont séparés par une carte «END OF RECORD» contenant les perforations 7, 8 et 9 dans la première colonne. Une carte «END OF FILE» contenant les perforations 6, 7, 8 et 9 dans la première colonne termine le «job».

En fait, c'est la partie variable du «job» qui change à chaque nouveau problème et en particulier ce sont les différentes options qui vont déterminer automatiquement quelles parties du programme doivent être exécutées et quelles autres peuvent être omises.

Considérons maintenant un peu plus en détail la structure de cette partie variable (III et IV) du «job».

Pour fixer les idées nous avons pris comme exemple une étude de trois échantillons. Les données, empruntées à une étude du professeur Twiesselmann, sont issues d'une enquête anthropologique des habitants de l'Oasis de Kharga. Dans cette étude, l'auteur a comparé la relation entre la taille symphysaire et la largeur du bassin de cette population méditerranéenne, avec celles présentées par des Ethiopiens et par des Nilotes.

2. — CARTES AVEC LES MESURES INDIVIDUELLES DES SUJETS DE CHAQUE ÉCHANTILLON (III).

Le nombre maximum de sujets de chaque échantillon est fixé à cinq cents. Ces limites peuvent éventuellement être dépassées en

changeant dans le programme principal l'instruction qui réserve les positions pour les différentes valeurs, à savoir l'instruction COMMON X (500,5) Y (500,5). Pour des raisons d'économie de mémoire centrale, ces changements seront, ainsi que le nombre d'échantillons, au maximum égaux à 20 et le produit des deux indices des matrices «X» et «Y» ne dépassera pas la valeur de 10000. Par exemple : COMMON X (5000,2) Y (5000,2), ou COMMON X (500,10), Y (500,10). Le premier indice indique le nombre de sujets, le deuxième le nombre d'échantillons.

Les deux mesures d'un sujet sont perforées sur une même carte. La première valeur de chaque paire de mesures peut être perforée n'importe où entre les colonnes 1 et 20, la deuxième entre les colonnes 21 et 40. Chaque valeur doit obligatoirement être pourvue d'un point décimal, même dans le cas de nombres entiers. Le nombre de chiffres derrière le point décimal est limité à 8.

La carte du dernier sujet d'un échantillon est en outre pourvue de trois autres indications. La première indique la fin de l'échantillon. Elle correspond au mot «END» ou «STOP» selon qu'il s'agit d'un échantillon qui est suivi d'un autre ou qu'il s'agit du dernier échantillon. L'indication «STOP» est aussi utilisée dans le cas où il n'y a qu'un seul échantillon.

Les mots «END» ou «STOP» sont perforés dans les colonnes 41 à 44 de la dernière carte de l'échantillon; ils sont obligatoires. La deuxième indication de la carte terminale de chaque échantillon concerne le type de symbole qui doit être utilisé pour la représentation du nuage de points dans le dessin. La perforation d'un chiffre compris entre 1 et 5 dans la colonne 45 de la dernière carte, donne lieu à l'utilisation des symboles suivants dans la représentation du nuage de points : 1 pour +, 2 pour ○, 3 pour ·, 4 pour X, 5 pour *. Cette identification est obligatoire dans le cas où l'on prend une option pour un dessin (voir VI).

La troisième indication, qui est toujours facultative, nous fournit la possibilité de munir l'échantillon d'un nom ou d'une notice explicative. Cette notice éventuelle figure dans les «listings» et dans une légende qui accompagne le dessin. Elle peut être composée au maximum de 15 caractères alphanumériques, qui seront perforés dans les colonnes 50 à 65 de la dernière carte de l'échantillon (voir exemple fig. 1).

3. — LES CINQ CARTES a, b, c, d et e CONTENANT LES OPTIONS OBLIGATOIRES ET FACULTATIVES (IV).

3.1. — Option dessin (a).

Quelles que soient les options, le «job» donnera toujours un «listing» de paramètres statistiques (fig. 2), mais la représentation graphique ne se fait que sur option.

L'exécution du dessin peut être demandée en perforant dans la première colonne de la carte «a» le chiffre 1. La présence de tout autre chiffre ou l'absence d'indication dans cette colonne résultera en une non-exécution du dessin. Cependant, cela ne signifie pas que les cartes «b», «c», «d» et «e» peuvent être omises; seulement, les options qu'elles contiennent éventuellement ne seront pas prises en considération et par conséquent ces cartes peuvent aussi bien être vierges.

Si au contraire, on a choisi l'exécution d'un dessin, l'indication des options sur les autres cartes devient importante.

3.2. — Echelle et graduation sur les deux axes (b).

La carte «b» contient six indications chacune dans une zone de 10 positions. Les trois premières indications concernent l'abscisse, les trois dernières, l'ordonnée. La première indication donne le nombre d'unités de la variable X par centimètre de papier réel sur le dessin. Sur le dessin, les graduations des axes sont toujours distantes de 1 cm. Il faut donc indiquer à quelles valeurs de mensurations elles correspondent.

Si par exemple on a porté en abscisse la stature d'un groupe d'enfants et si on sait que les valeurs extrêmes de la taille se situent entre 60 cm et 190 cm, on pourrait choisir un nombre d'unités par centimètres de papier égal à 10. Cela signifierait que la distribution des valeurs de la stature couvrirait au maximum 15 cm sur la graphique.

Le nombre 10 doit alors être perforé en nombre décimal (avec point décimal) sur la carte «b», quelque part dans la première zone de dix positions (de la colonne 1 à 10).

La deuxième indication de la carte «b» spécifie la valeur de l'origine de l'axe X. Cette valeur doit également être exprimée en nombre décimal et figure dans la deuxième zone de la carte «b», c'est-à-dire de la colonne 11 à 20. Dans l'exemple cité ci-dessus ce paramètre sera par exemple la valeur 50.

La troisième indication de la carte «b» donne la valeur maximale de la graduation le long de l'abscisse et doit être perforée en nombre

décimal dans la zone comprise entre la 21^e et la 30^e colonne. Dans l'exemple cité, il pourrait prendre une valeur égale à 200. Les longueurs d'une unité de mesure de X et Y ne doivent pas être égales sur le dessin.

Les trois autres indications à porter sur la carte «b» se rapportent à l'axe Y, mais elles ont une signification analogue aux trois premières. Elles se trouvent dans le même ordre et doivent être perforées respectivement de la même façon que les indications précédentes dans les zones 31 à 40, 41 à 50 et 51 à 60 de la carte «b».

3.3. — *Commentaires et chiffres sur les axes (c).*

La carte «c» contient trois paramètres dont les deux premiers nous permettent de pourvoir les deux axes d'un commentaire qui définit les deux variables en question. Le troisième paramètre nous donne la possibilité de munir les deux axes gradués d'une échelle chiffrée. Ces options sont néanmoins facultatives.

Le commentaire qui doit accompagner l'axe X est perforé dans les colonnes 1 à 10 de la carte «c» et peut être composé d'une suite éventuellement non continue de caractères alphanumériques ou de blancs. Ce commentaire figurera d'une part dans les tableaux statistiques (listing) et d'autre part sur le graphique au niveau de l'extrémité droite de l'axe X.

Le commentaire relatif à l'axe X s'inscrit dans la deuxième zone de la carte «c», c'est-à-dire de la colonne 11 à 20; il figurera également dans les tableaux statistiques ainsi que sur le graphique au niveau de l'extrémité supérieure de l'axe Y. Ce commentaire pourrait être par exemple : «stature», «long. bras», «diam. bicr.», etc. L'absence de tout caractère alphanumérique dans ces deux zones de la carte «c» résultera évidemment en une non-exécution de cette option.

Le troisième paramètre détermine si oui ou non les deux axes gradués du graphique doivent être munis d'une échelle chiffrée. A cet égard on donnera à ce paramètre respectivement la valeur 1 ou 0, valeur qui sera inscrite dans la colonne 21 de la carte «c». Notez que pour des raisons de limite d'espace les chiffres le long de l'axe X sont imprimés verticalement (voir fig. 2).

3.4. — *Ellipses équiprobables (d).*

Les paramètres de la carte «d» permettent de choisir le nombre d'ellipses équiprobables qui vont entourer le(s) nuage(s) de points ainsi que le pourcentage d'individus qu'elles doivent contenir. Le nombre maximal d'ellipses équiprobables qui peuvent être obtenues simulta-

L. BASSIN

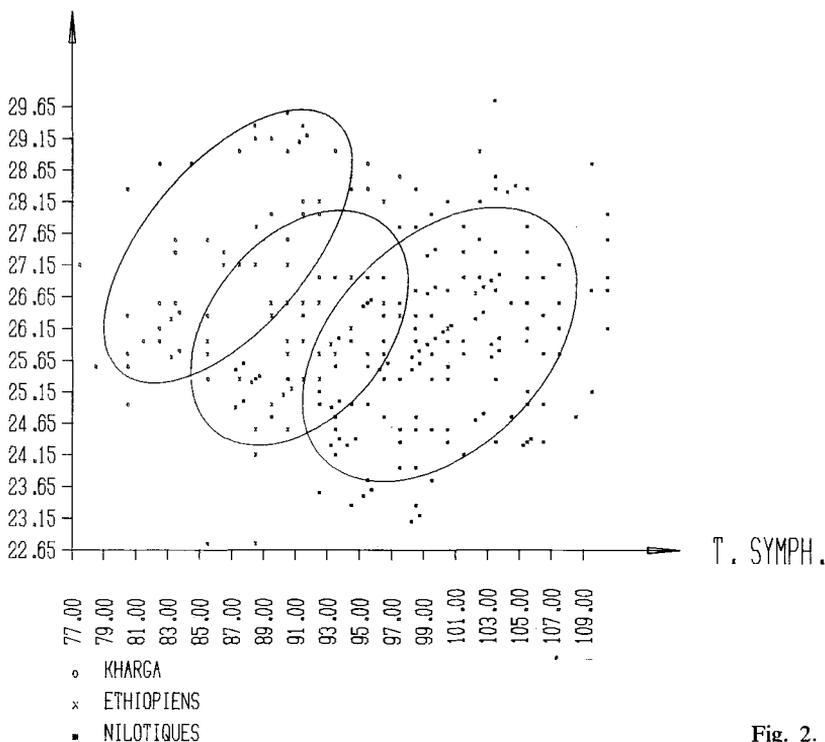


Fig. 2.

nément est limité à trois; il s'agit notamment des ellipses contenant 70 %, 95 % et 99 % des sujets de l'échantillon.

Sur la carte «d» sont réservées 3 zones de 2 positions qui correspondent aux trois ellipses, dans l'ordre mentionné ci-dessus. L'ellipse contenant 70 % des points est obtenue en perforant le chiffre «70» dans les colonnes 1 et 2 de la carte «d»; l'ellipse contenant 95 % des points est obtenue en perforant le chiffre «95» dans les colonnes 3 et 4 et celle contenant 99 % des points en perforant le chiffre «99» dans les colonnes 5 et 6 de la carte «d». Ces ellipses peuvent être obtenues indépendamment l'une de l'autre en laissant vierges certaines zones de la carte «d». Si la carte «d» reste entièrement vierge, aucune ellipse ne figurera sur le graphique.

Exemple : On demande l'ellipse contenant 95 % des points en perforant simplement le chiffre «95» dans les colonnes 3 et 4; on demande

L. BASSIN

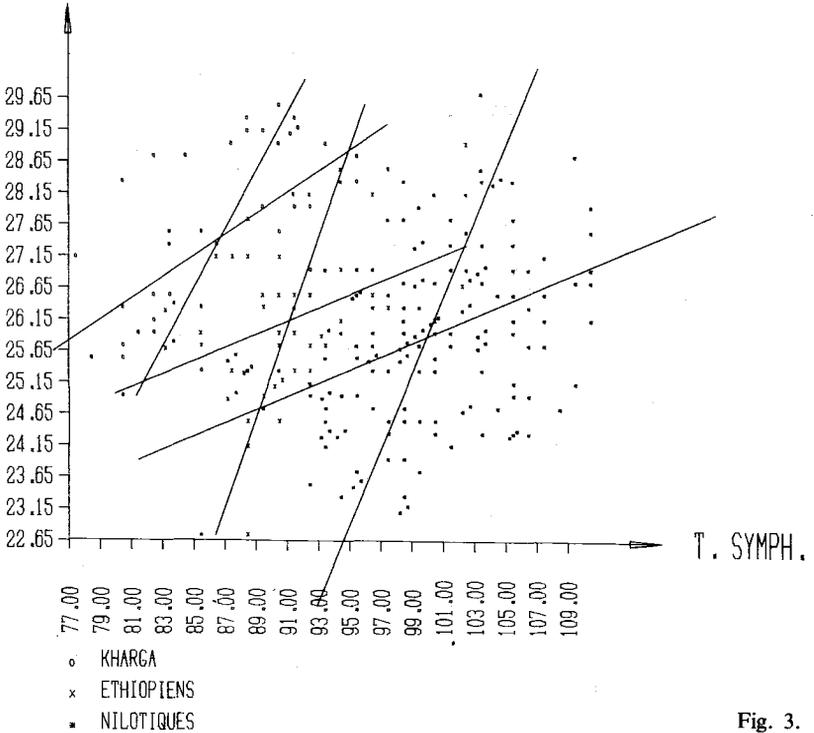


Fig. 3.

simultanément les ellipses contenant 70 % et 99 % des points en perforant le chiffre «70» dans les colonnes 1 et 2 et le chiffre «99» dans les colonnes 5 et 6 de la carte «d»; etc.

3.5. — Nuage de points, droites de régression et droites perpendiculaires aux deux axes du système et passant par le point moyen (e).

Le tracé des nuages de points se fait sur option. Elle est obtenue en donnant au premier paramètre de la carte «e» la valeur 1. Elle est omise si ce paramètre reçoit la valeur 0. Cette valeur doit être perforée dans la première colonne de la carte «e». Notons que dans le cas où ce paramètre est égal à 1, le type de symbole qui sera utilisé pour la représentation du nuage de points est celui qui aura été défini dans la dernière carte de l'échantillon (voir plus haut).

Les droites de régression sont aussi obtenues facultativement. Le paramètre qui détermine si oui ou non les droites de régression doivent

figurer dans le graphique, sera perforé dans la deuxième colonne de la carte «e» et prendra respectivement la valeur 1 ou 0.

Finalement, on peut demander de dessiner les deux droites qui passent par le point moyen de chaque échantillon et qui sont perpendiculaires aux deux axes. Cette option est prise en perforant le chiffre 1 dans la colonne 3 de la carte «e». La perforation de la valeur 0 dans cette zone résultera en une non-exécution de cette option.

3.6. — Exemples.

Les figures 2, 3 et 4 nous montrent différents modes de représentation des mêmes échantillons de l'étude du professeur Twiesselmann.

La figure 2 représente les ellipses de 70 % comme l'a fait Defrise-Gussenhoven (1955).

La figure 3 donne, pour les mêmes échantillons, les nuages de points et les droites de régression.

L. BASSIN

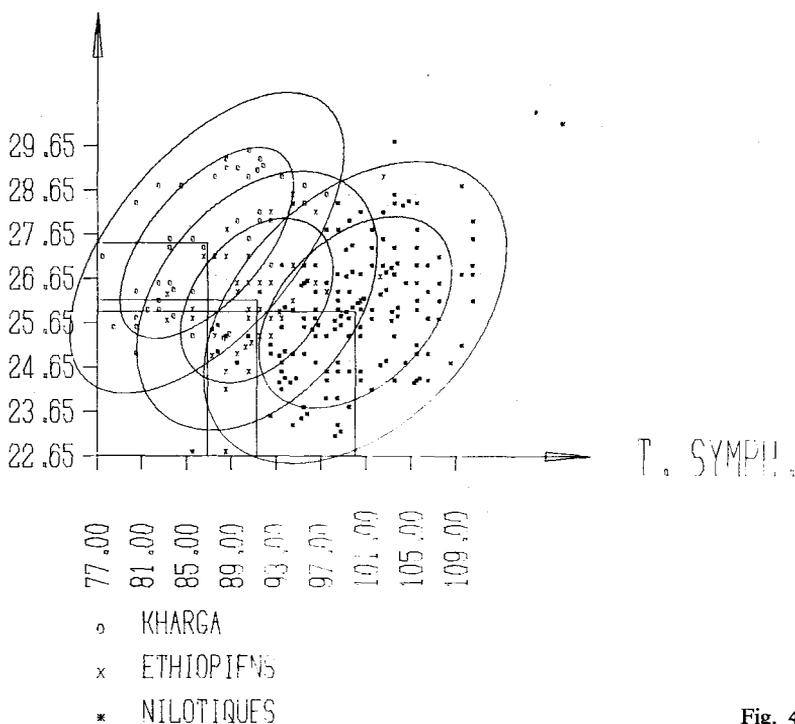


Fig. 4.

La figure 4 est réalisée en adoptant une échelle différente de celle des dessins précédents. Les ellipses autour des nuages de points contiennent 70 % et 95 % des sujets des échantillons respectifs. En outre sont dessinées les droites perpendiculaires aux deux axes du système et passant par les points moyens de chaque échantillon.

```

*****
*
* NUMERO DE L ECHANTILLON 1      KHARGA      ( 0 ) *
*
*****
*
* VARIABLE X ; T. SYMPH.      VARIABLE Y ; L. BASSIN *
* -----
*
* MOYENNE      86.79070      MOYENNE      27.45465 *
* ECART-TYPE   4.99242      ECART-TYPE   1.39037 *
* EFFECTIF     43           EFFECTIF     43 *
*****
*
* COEFFICIENT DE CORRELATION      .6031 *
* COEFFICIENTS DE REGRESSION B1   .1680 *
*                                B2   .4618 *
*****

```

```

*****
*
* NUMERO DE L ECHANTILLON 2      ETHIOPiens      ( X ) *
*
*****
*
* VARIABLE X ; T. SYMPH.      VARIABLE Y ; L. BASSIN *
* -----
*
* MOYENNE      91.25122      MOYENNE      26.15854 *
* ECART-TYPE   4.37872      ECART-TYPE   1.19587 *
* EFFECTIF     41           EFFECTIF     41 *
*****
*
* COEFFICIENT DE CORRELATION      .3841 *
* COEFFICIENTS DE REGRESSION B1   .1049 *
*                                B2   .7111 *
*****

```

```

*****
*
* NUMERO DE L ECHANTILLON 3      NILOTIQUES      ( * ) *
*
*****
* VARIABLE X : T. SYMPH.      VARIABLE Y : L. BASSIN *
* -----
*
* MOYENNE      100.02500      MOYENNE      25.89633 *
* ECART-TYPE   5.51964      ECART-TYPE   1.39413 *
* EFFECTIF     150          EFFECTIF     150 *
*****
*
* COEFFICIENT DE CORRELATION      .4203 *
* COEFFICIENTS DE REGRESSION B1   .1062 *
*                                B2   .6009 *
*****

```

Le tableau 1 résume les données statistiques relatives aux trois échantillons, telles qu'elles se trouvent sur le «listing», à savoir, un numéro de succession pour chaque échantillon (apparaissant automatiquement), le «nom» de l'échantillon, le type de symbole qui est utilisé pour la représentation du nuage de points (entre parenthèses), le nom des variables X et Y, la moyenne et l'écart-type des deux variables, l'effectif de l'échantillon, le coefficient de corrélation et les coefficients de régression.

Le «listing» du programme «ELLIPS» et du sous-programme «PLOT» peuvent être obtenus en s'adressant aux auteurs.

BIBLIOGRAPHIE

DEFRISE-GUSSENHOVEN, E.

1955 Ellipses équiprobables et taux d'éloignement en biométrie.
Bull. Inst. roy. Sc. nat. Belg., 31 (26), 1-31.

GLOWACKI, C. et M. MEURENS.

1971 *Programmes pour l'emploi de la table traçante BENSON 441*. Centre de Calcul U.L.B./V.U.B., Université Libre de Bruxelles, avenue Ad. Buyl, 91, 1050 Bruxelles.

TWIESELMANN, F.

1954 Propos sur l'Anthropologie,
in : *Volume jubilaire Victor Van Straelen*, II : 1065-1098, Inst. roy. Sc. nat.
Belg., Bruxelles.

Adresse des auteurs : Laboratorium voor Antropogenetica.
Vrije Universiteit Brussel,
Pleinlaan, 2.
B 1050 Brussel (België).