

F. BENZÉCRI

Introduction à l'analyse des correspondances d'après un exemple de données médicales

Les cahiers de l'analyse des données, tome 5, n° 3 (1980),
p. 283-310

http://www.numdam.org/item?id=CAD_1980__5_3_283_0

© Les cahiers de l'analyse des données, Dunod, 1980, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

INTRODUCTION A L'ANALYSE DES CORRESPONDANCES D'APRÈS UN EXEMPLE DE DONNÉES MÉDICALES [INT. CORR. MED.]

par F. Benzécri (1)

L'analyse des correspondances est un maillon dans une étude que nous schématiserons en nous efforçant d'éclaircir ce qu'il importe au médecin de connaître des rouages de la statistique multidimensionnelle.

Le médecin pose un problème : e.g. les effets de la Trinitrine sur des malades présentant, au départ, divers caractères (cf Pratique de l'Analyse des Données t.1, IV § 10 et Nouvelle Presse Médicale, t.4 p. 283, Janv. 79).

Il recueille les données, mettant en oeuvre tous les moyens dont il dispose. Pour rester dans le même exemple, pour chaque malade, il mesure des pressions artérielles, des débits cardiaques etc., il calcule des indices de diverses natures... et cela, avant et après administration de Trinitrine.

Pour rassembler ces données, il constitue un ou plusieurs tableaux récapitulatifs. C'est ainsi que le Dr Baligadoo a construit un tableau donnant, pour chaque malade, les différences des valeurs des divers paramètres avant et après administration de Trinitrine, et aussi des renseignements complémentaires comme : présence ou absence d'infarctus récent. Un tel tableau est appelé *tableau de données* : il constitue la base des réflexions et opérations ultérieures, mais ce n'est pas en général ce tableau qu'on soumettra tel quel à l'analyse des correspondances.

1 Du tableau de données aux profils multidimensionnels

A partir du tableau de données le statisticien construit le *tableau de correspondance* proprement dit qui, lui, sera soumis à l'analyse des correspondances. Comme nous le dirons par la suite, les nombres du tableau de correspondance ne sont pas considérés en eux-mêmes, mais en rapport les uns avec les autres. Il est donc essentiel, s'il s'agit de mesures, que ces mesures soient exprimées dans la même unité.

1.1 Codage par classes des paramètres (planche I)

Mais lorsque les paramètres sont de diverses natures comme : les pressions, les débits, etc., on ne peut espérer harmoniser tous ces nombres. En ce cas, on *code les paramètres par classes*. Par exemple, pour : Différence de la Pression Capillaire Pulmonaire (D.C.P.), mesurée en mm. de mercure (mm Hg), on a enregistré des valeurs allant de -17 mm Hg à 0. Sur l'échelle des D.C.P. que l'on peut figurer comme

(1) Dr ès Sciences ; (Exposé au séminaire du 18 Juin 1980 ; Hôpital Tenon)

soumis à l'analyse des correspondances. Mais on peut aussi lui associer un *tableau de fréquence* appelé *tableau de Burt* et croisant l'ensemble des modalités avec lui-même : à la croisée de deux modalités, on inscrit le nombre de malades présentant à la fois ces deux modalités.

	A1	A2	A3	B1	B2		A1	A2	A3	B1	B2
i1	0	1	0	1	0	A1	2	0	0	1	1
i2	0	1	0	0	1	A2	0	4	0	2	2
i3	0	0	1	1	0	A3	0	0	1	1	0
i4	1	0	0	1	0	B1	1	2	1	4	0
i5	1	0	0	0	1	B2	1	2	0	0	3
i6	0	1	0	1	0						
i7	0	1	0	0	1						

↑
↑
tableau logique et *tab. de Burt associé*

e.g. la case A1, B2 contient le nombre de malades qui, dans le tableau logique, présentent à la fois la modalité A1 et la modalité B2, soit : 1. On peut remarquer que l'on a le même nombre dans la case B2, A1. Autrement dit : le tableau de Burt est symétrique par rapport à sa 1^o diagonale. Autre particularité de ce tableau : dans le bloc $A \times A$ ou $B \times B$ il y a des zéros partout sauf sur la 1^o diagonale, car aucun malade ne peut posséder à la fois deux modalités d'un même paramètre.

A des coefficients près, l'analyse du tableau logique et celle du tableau de Burt donnent les mêmes résultats. Et c'est ce qui justifie l'analyse des tableaux en (0,1). Pourquoi ? parce que le tableau de Burt est un *tableau de fréquence* et que c'est pour les tableaux de fréquence que l'analyse des correspondances a été conçue.

1.3 Tableaux de fréquence (planche II)

D'une façon générale, on appelle *tableau de fréquence* un tableau résultant de comptages. On vient d'en voir un exemple avec le tableau de Burt ; en voici un autre exemple : dans une liasse de dossiers provenant d'une enquête auprès de médecins praticiens, on relève combien de fois chacun des six médicaments *pénicilline, tifyomycine, tétracycline, érythromycine, tiophénicol et gentalline* a été prescrit dans les 7 maladies *TYPHOÏDE, SALMONELLOSE DIGESTIVE, AFFECTION ORL, PNEUMOPATHIE, MENINGITE, AFFECTION DES VOIES URINAIRES, STAPHYLOCOCCIE*. On peut disposer ces données en un vrai tableau de correspondance, en portant en lignes les maladies, en colonnes les médicaments (ou l'inverse : mais pour des raisons matérielles, on met en colonnes le plus petit des deux ensembles) ; donc en lignes les maladies, en colonnes les médicaments ; et en inscrivant à la croisée d'une maladie et d'un médicament le nombre de fois que ce médicament a été prescrit pour cette maladie.

Répétons-le : c'est pour les tableaux de fréquence que l'analyse des correspondances a été conçue.

- Ce sont des tableaux homogènes : les nombres y représentent des faits de même nature.
- Mieux : ils sont les faits eux-mêmes, objets de l'étude, dénombrement des rencontres entre éléments de deux ensembles.
- C'est aussi sur ces tableaux que les opérations que l'on fait en A. des cor. sont le plus facile à comprendre ; par exemple : additionner les deux lignes du tableau AF. ORL et PNEUMOPATHIE revient à ne pas

distinguer l'une de l'autre ces deux maladies, mais à les considérer comme constituant une seule classe de maladies. Si *pénicilline* a été prescrite 7 fois pour PNEUMOPATHIE et 8 fois pour AF. ORL, elle a été prescrite 15 fois pour l'une ou l'autre de ces deux maladies.

	<i>pénicilline</i>	<i>tifomycine</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>tiophénicol</i>	<i>gentalline</i>
TYPHOÏDE	0	4	0	0	2	0
SALMONELLOSE DIG.	0	2	0	0	1	0
AFFECTION ORL	8	0	5	3	0	3
PNEUMOPATHIE	7	0	5	2	0	3
MENINGITE	2	2	0	0	0	1
AFF. VOIES URINAIRES	4	0	2	0	0	6
STAPHYLOCOCCIE	3	0	1	3	0	0

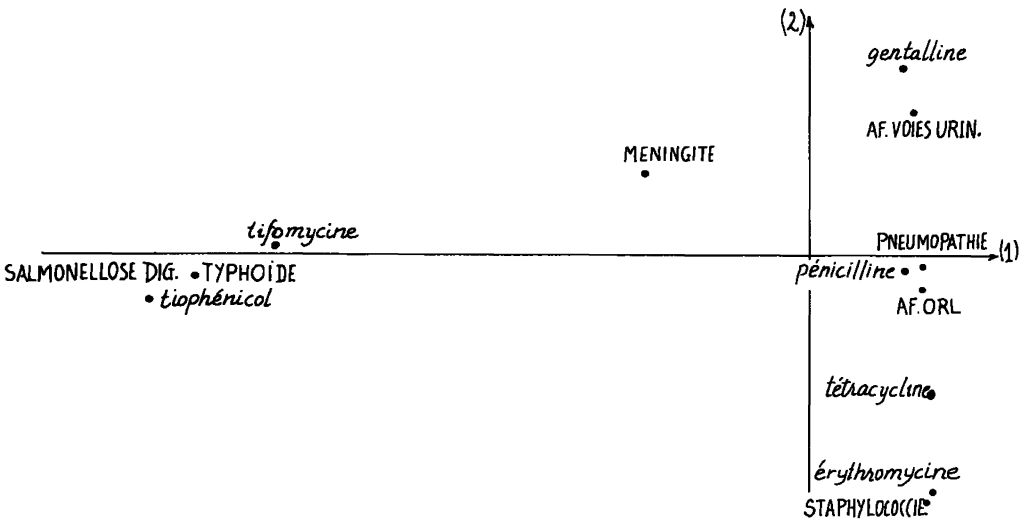
nombre de fois que la tétracycline a été prescrite pour une pneumopathie.

1.4 Correspondance entre les deux ensembles croisés par le tableau

Donc, après codage ou non, on se trouve avoir constitué un tableau de correspondance entre deux ensembles I et J : pour tout couple d'un élément de I et d'un élément de J, le tableau de correspondance donne un nombre, et c'est par là que les deux ensembles I et J sont en correspondance. L'analyse de ce tableau fournira :

- une représentation géométrique sur des cartes planes de chacun des deux ensembles, et, ce qui est propre à l'A. des cor., des deux ensembles I et J simultanément ;
- en outre, des listes de contributions permettant de critiquer les résultats.

Par exemple, voici une carte issue de l'analyse du tableau croisant les 7 maladies avec les 6 médicaments (planche III). On y voit un groupe composé de deux maladies confondues au même point : TYPHOÏDE et SALMONELLOSE DIGESTIVE, et deux médicaments : *tifomycine* et *tiophénicol*. Le reste des maladies et des médicaments, sauf MENINGITE, s'égrène dans la direction verticale. Quant à MENINGITE, elle occupe une position intermédiaire.



Comparons ces résultats avec le tableau de données (planche II) :

tiophénicol n'apparaît qu'avec TYPHOÏDE et SALMONELLOSE DIG. ;

tifomycine apparaît principalement avec TYPHOÏDE et SALMONELLOSE, mais aussi deux fois avec MENINGITE, d'où la position intermédiaire, sur le graphique, de cette maladie.

Quant aux deux maladies TYPHOÏDE et SALMONELLOSE DIG., pourquoi sortent-elles au même point sur le graphique ? On peut constater que la ligne TYPHOÏDE du tableau de correspondance est exactement deux fois la ligne SALMONELLOSE. Autrement dit, le nombre total des occurrences de chacune de ces deux maladies se répartit dans les mêmes proportions sur les six médicaments : on dit que les deux maladies ont *même profil* ; et c'est pourquoi elles sont confondues sur les graphiques.

1.5 Profils

De fait, il faut remarquer que dans le tableau de correspondance, chaque nombre pris isolément n'a pas de sens ; dire que TYPHOÏDE a été traitée 4 fois par *tifomycine* ne prend un sens que si l'on précise : sur 6 cas de TYPHOÏDE. En sorte que ce qui importe, c'est le rapport 4/6 plutôt que le nombre 4.

C'est pourquoi la première opération effectuée dans l'analyse sur le tableau de correspondance est de calculer les totaux des lignes et des colonnes (planche IV).

	<i>pénicilline</i>	<i>tifomycine</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>tiophénicol</i>	<i>gentalline</i>	col de marge
TYPHOÏDE	0	4	0	0	2	0	6
SALMONELLOSE DIG	0	2	0	0	1	0	3
AFFECTION ORL	8	0	5	3	0	3	19
PNEUMOPATHIE	7	0	5	2	0	3	17
MENINGITE	2	2	0	0	0	1	5
AF URINAIRE	4	0	2	0	0	6	12
STAPHYLOCOCCIE	3	0	1	3	0	0	7
ligne de marge	24	8	13	8	3	13	69

← total de la ligne TYPHOÏDE, i.e. nombre des occurrences de TYPHOÏDE sur la totalité des dossiers étudiés.

↑ total de la colonne pénicilline, i.e. nombre des occurrences de pénicilline sur la totalité des dossiers étudiés.

Puis on calcule le profil de chaque ligne du tableau de correspondance en divisant tous les nombres de la ligne par le total de cette même ligne ; e.g. le profil de MENINGITE, de total 5, est la suite de rapports :

{2/5, 2/5, 0, 0, 0, 1/5},

ce qui veut dire : sur l'ensemble des cas de MENINGITE recensés dans les dossiers étudiés, les 2/5 sont traités par *pénicilline*, les 2/5 par *tifomycine*, le 1/5 par *gentalline* ; ce qui s'énonce aussi en pourcentages : 40% des cas recensés de MENINGITE sont traités par *pénicilline*, 40% par *tifomycine* et 20% par *gentalline*. Remarquons que la somme

des rapports constituant un profil vaut 1 :

$$2/5 + 2/5 + 1/5 = 1$$

On calcule de même le profil de chaque colonne du tableau de cor. en divisant tous les nombres de la colonne par le total de cette même colonne ; e.g. le profil de la colonne pénicilline, de total 24, est la suite des rapports :

$$\{0, 0, 8/24, 7/24, 2/24, 4/24, 3/24\} \text{ (de total 1).}$$

D'où deux nouveaux tableaux (planche V) : le tableau des profils des lignes (i.e. des maladies) du tableau de correspondance et le tableau des profils des colonnes (i.e. des médicaments) du tab. de cor..

	<i>pénicilline</i>	<i>tifomycine</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>tiophénicol</i>	<i>gentalline</i>
TYPH	0	4/6	0	0	2/6	0
SALM	0	2/3	0	0	1/3	0
ORL	8/19	0	5/19	3/19	0	3/19
PNEUM	7/17	0	5/17	2/17	0	3/17
MENIN	2/5	2/5	0	0	0	1/5
URIN	4/12	0	2/12	0	0	6/12
STAPH	3/7	0	1/7	3/7	0	0

	<i>pénicilline</i>	<i>tifomycine</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>tiophénicol</i>	<i>gentalline</i>	<i>col de marge</i>
TYPH	0	4/8	0	0	2/3	0	6/69
SALM	0	2/8	0	0	1/3	0	3/69
ORL	8/24	0	5/13	3/8	0	3/13	19/69
PNEUM	7/24	0	5/13	2/8	0	3/13	17/69
MENIN	2/24	2/8	0	0	0	1/13	5/69
URIN	4/24	0	2/13	0	0	6/13	12/69
STAPH	3/24	0	1/13	3/8	0	0	7/69

<i>L.demrge</i>	24/69	8/69	13/69	8/69	3/69	13/69
-----------------	-------	------	-------	------	------	-------

*tableau des profils des maladies
(à lire horizontalement).*

*tableau des profils des médicaments
(à lire verticalement).*

On remarquera que l'on calcule les profils de la ligne de marge et de la colonne de marge comme ceux des autres lignes et colonnes : nous verrons plus loin le rôle joué dans l'analyse par ces deux profils (cf § 2.1).

1.6 Masses

Mais l'effectif de chaque maladie n'est pas perdu, ni celui de chaque médicament. On le retrouve sous forme de masse affectée à chaque maladie et à chaque médicament, e.g. la masse de TYPHOÏDE est égale à 6/69 (total de la ligne TYPHOÏDE dans le tab. de cor. divisé par le total général du tab. de cor.). Ici encore, on rapporte le nombre des occurrences d'une maladie (ou d'un médicament) au total général du tableau, i.e. au nombre total des occurrences des 7 maladies (ou des 6 médicaments).

Voici les masses des 7 maladies et des 6 médicaments :

TYPHOÏDE	6/69	<i>pénicilline</i>	24/69
SALMONELLOSE DIG	3/69	<i>tifomycine</i>	8/69
AFFECTION ORL	19/69	<i>tétracycline</i>	13/69
PNEUMOPATHIE	17/69	<i>érythromycine</i>	8/69
MENINGITE	5/69	<i>tiophénicol</i>	3/69
AF. URINAIRE	12/69	<i>gentalline</i>	13/69
STAPHYLOCOCCIE	7/69		

On notera que la somme des masses des 7 maladies vaut 1 ; ainsi que la somme des masses des 6 médicaments.

1.7 Les nuages (planche V)

les points que l'on voit sur les graphiques sont définis par les profils munis de masses calculés comme on vient de le dire ; e.g. MENINGITE n'est pas autre chose, dans l'analyse, que la suite des six rapports : {2/5, 2/5, 0, 0, 0, 1/5} munie de la masse 5/69.

Or, une suite de six nombres, c'est un point dans un espace géométrique à six dimensions : de même qu'un nombre équivaut à un point sur un axe, qu'une suite de deux nombres équivaut à un point dans un plan rapporté à deux axes de coordonnées, et qu'une suite de trois nombres équivaut à un point dans l'espace à trois dimensions où nous évoluons, quand on le rapporte à trois axes de coordonnées, de même une suite ordonnée de six nombres équivaut à un point dans un espace à six dimensions, et une suite ordonnée de sept nombres équivaut à un point dans un espace à sept dimensions.

Donc, les maladies, par le tableau des profils, sont définies comme des points dans un espace à six dimensions, et ces points sont pesants. Leur ensemble constitue le *nuage des maladies*.

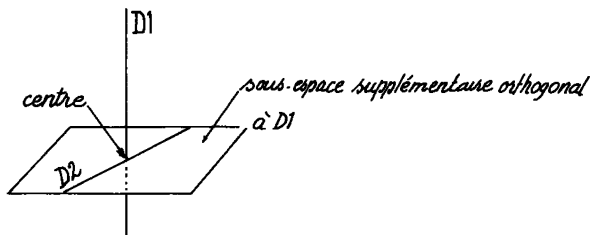
De même, les médicaments, par le tableau des profils, sont définis comme des points dans un espace à sept dimensions, et ces points sont pesants. Leur ensemble constitue le *nuage des médicaments*.

En fait, chaque nuage, dans son espace ambiant, est cantonné dans un sous-espace de dimension égale au nombre d'éléments, moins un, du plus petit des deux ensembles en correspondance. Dans notre exemple, les deux nuages ont dimension : $6 - 1 = 5$. Nous verrons qu'on peut les représenter tous les deux dans un même espace : la carte plane considérée au § 1.4 comprend à la fois des médicaments et des maladies ; c'est une projection de deux nuages représentés dans un même espace à 5 dimensions sur un plan bien choisi de cet espace.

Car il suffit de considérer (planche VI) une figure dans l'espace usuel à trois dimensions pour voir que le choix du plan de projection est essentiel pour obtenir une image fidèle de la réalité. Toutefois, même pour le meilleur plan, des points alignés sur une même projetante seront confondus en projection, bien que distincts en réalité. C'est pourquoi on ne se contentera pas en général d'observer un seul plan de projection.

1.8 Comment choisir le meilleur plan de projection d'un nuage

On détermine le centre de gravité du nuage ; par le centre, on mène la droite dont le nuage s'écarte le moins et le long de laquelle il se disperse le plus (écart et dispersion se mesurent, cf § 2, en terme d'inertie, fonction de la masse et du carré de la distance). Cette droite D1 sera la 1° axe factoriel ou axe principal d'allongement du nuage. On décompose alors l'espace ambiant en cette droite D1 et le sous-espace supplémentaire orthogonal à D1 passant par le centre (l'équivalent d'un plan perpendiculaire à une droite en dimension 3) ; on projette le nuage sur ce sous-espace et l'on recherche pour cette projection, à nouveau, l'axe principal d'allongement qui sera le 2° axe factoriel du nuage ; etc. .



Donc, chaque maladie i apporte son inertie au nuage par le terme $f_i \times r(i)^2$.

- Par f_i (masse de la maladie i) on donne à chaque maladie l'importance qui est la sienne (plus une maladie sera fréquente relativement à la totalité des cas étudiés, plus elle aura d'importance dans l'inertie totale).

- Pourquoi $(r(i))^2$ et non simplement $r(i)$? (planche IX). Parce qu'on se place dans le cadre de la géométrie euclidienne i.e. celui de la géométrie élémentaire usuelle; et que dans cette géométrie, le théorème de Pythagore bien connu permet de décomposer le carré d'une distance dans deux directions perpendiculaires; or, on veut aussi décomposer l'inertie du nuage dans deux directions perpendiculaires (cf § 1.8). C'est pourquoi on prend $(r(i))^2$ et non $r(i)$ dans la formule d'inertie.

2.2.2 Comment calcule-t-on $(r(i))^2$ dans l'espace à 6 dimensions où sont représentées les maladies ?

Dans le plan rapporté à deux axes rectangulaires, le théorème de Pythagore permet d'exprimer $(r(i))^2$ (carré de la distance du point i à l'origine) comme somme des carrés des coordonnées de i , avec éventuellement des coefficients, selon les unités de graduation que l'on a choisies sur les axes.

En dimension 6, on posera pour un point M quelconque de l'espace défini par la suite ordonnée de 6 nombres $(x_1(M), x_2(M), x_3(M), x_4(M), x_5(M), x_6(M))$,

$$(r(M))^2 = \alpha_1 (x_1(M))^2 + \alpha_2 (x_2(M))^2 + \alpha_3 (x_3(M))^2 + \alpha_4 (x_4(M))^2 + \alpha_5 (x_5(M))^2 + \alpha_6 (x_6(M))^2$$

↑
carré de la distance à l'origine du point M

- Reste à choisir les coefficients $\alpha_1, \alpha_2, \dots, \alpha_6$: ils sont déterminés à partir du tableau de correspondance de telle sorte que soit satisfait le principe d'équivalence distributionnelle qu'on peut énoncer ainsi:

On ne change pas les résultats de l'analyse en cumulant plusieurs lignes (ou plusieurs colonnes) proportionnelles entre elles.

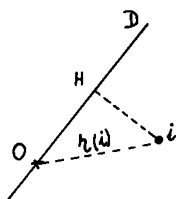
Par exemple, on aurait pu remplacer les deux lignes proportionnelles TYPHOÏDE et SALMONELLOSE DIGESTIVE par une seule ligne somme des deux premières, colonne par colonne, rien n'aurait changé dans les résultats: le point {TYPHOÏDE + SALMONELLOSE D.} serait sorti à l'endroit même où sont sortis les deux points confondus TYPHOÏDE et SALMONELLOSE (cf figure § 1.4).

- Pour que le principe d'équivalence distributionnelle soit vérifié, on doit prendre pour coefficients, dans la formule de distance donnant $(r(M))^2$, les inverses des coordonnées du profil de la ligne de marge du tableau de correspondance (cf § 1.5, tableau des profils des maladies). Dans notre exemple on aura (planche V):

$$\alpha_1 = 69/24 ; \alpha_2 = 69/8 ; \alpha_3 = 69/13$$

$$\alpha_4 = 69/8 ; \alpha_5 = 69/3 ; \alpha_6 = 69/13$$

La distance ainsi définie dans l'espace à 6 dimensions, à partir du tableau de correspondance, est appelée *distance distributionnelle*.



Si l'on fait pivoter la droite D autour du centre O, la répartition de l'inertie totale en inertie transversale à D et inertie le long de D varie : la *meilleure droite* est celle pour laquelle l'inertie transversale est minima et, par suite, l'inertie longitudinale maxima. Cette droite D1 est le support du 1^o *axe factoriel* et on appelle 1^o *valeur propre* (notée λ_1) l'inertie du nuage le long de ce 1^o axe factoriel. La valeur propre λ_1 , et surtout le *taux d'inertie* (noté τ_1) qui est le rapport de λ_1 à l'inertie totale du nuage, mesurent la qualité de ce 1^o axe factoriel et sont donnés sur les listages de l'analyse factorielle (planche X).

Pour trouver le meilleur plan, on projette la figure comme nous l'avons dit au § 1.8, sur le sous-espace supplémentaire orthogonal à D1 en O, centre du nuage ; et l'on cherche pour cette projection, la meilleure droite : ce sera D2, support du 2^o *axe factoriel*. On appelle encore 2^o *valeur propre* (notée λ_2) l'inertie du nuage le long de D2 et taux d'inertie relatif au 2^o axe le rapport (noté τ_2) de λ_2 à l'inertie totale du nuage ; λ_2 et τ_2 figurent sur les listages : ils mesurent la qualité de l'axe 2. On a nécessairement λ_2 inférieure à λ_1 , sinon on aurait trouvé la droite D2 avant de trouver la droite D1.

De proche en proche on détermine un système d'axes deux à deux perpendiculaires et, pour chacun d'eux, une valeur propre qui est l'inertie du nuage le long de l'axe (ou encore l'inertie par rapport au centre de la projection du nuage sur cet axe) et un taux d'inertie qui est le rapport de la valeur propre à l'inertie totale du nuage. Dans l'analyse des 7 maladies croisées avec les 6 médicaments, on obtient cinq axes factoriels et les valeurs propres et taux d'inertie relatifs à ces cinq axes (planche XI).

$$\lambda_1 = 0,87 \quad ; \quad \tau_1 = 75,7 \%$$

$$\lambda_2 = 0,20 \quad ; \quad \tau_2 = 17,56\%$$

$$\lambda_3 = 0,049 \quad ; \quad \tau_3 = 4,24 \%$$

$$\lambda_4 = 0,028 \quad ; \quad \tau_4 = 2,4 \%$$

$$\lambda_5 = 0,00005 \quad ; \quad \tau_5 = 0,004\%$$

Comme l'inertie se décompose sur des directions perpendiculaires, on a :

$$\text{inertie totale du nuage} = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 ;$$

la somme des valeurs propres est appelée *trace* ; la trace est donc égale à l'inertie totale du nuage.

2.3 Représentation simultanée des deux nuages ; principe barycentrique

Si l'on fait la même recherche des axes factoriels pour le nuage des médicaments, on trouve les mêmes valeurs propres, et c'est ce qui permet d'identifier les axes un à un dans les deux espaces où sont définis les deux nuages, et par là, de représenter les deux nuages dans un seul espace rapporté aux cinq axes factoriels.

Dans cet espace commun, les deux nuages sont liés l'un à l'autre

par le *principe barycentrique*: à des retouches près, une maladie, e.g. :

STAPHYLOCOCCIE, de profil : 3/7 ; 0 ; 1/7 ; 3/7 ; 0 ; 0

sur : *pénicill. tifomyc. tétracycl. érythrom. tiophén. gental*

est centre de gravité des 6 médicaments affectés chacun, comme masse, de la coordonnée sur lui du profil de la maladie ; ainsi STAPHYLOCOCCIE est centre de gravité du système de points : *pénicilline* affecté de 3/7, *tétracycline* affecté de 1/7, *érythromycine* affecté de 3/7.

De même un médicament est centre de gravité des 7 maladies affectées chacune de la coordonnée sur elle du profil du médicament.

Cette propriété permet de calculer à la main la position d'un élément nouveau (maladie ou médicament) décrit par son profil (sur les médicaments si c'est une maladie ; sur les maladies si c'est un médicament).

2.4 Les facteurs

Les axes factoriels une fois déterminés, on calcule les facteurs des divers points des nuages, c'est-à-dire leurs *coordonnées* sur les axes factoriels.

Ainsi, chaque maladie qui au départ était définie par les 6 rapports (de total 1) de son profil, rapports qui correspondaient chacun à un médicament, est maintenant définie par cinq nouveaux nombres: ses *facteurs* ou coordonnées sur les axes factoriels.

2.4.1 Expression de la distance en fonction des facteurs (planche XII)

Dans ce nouveau système de coordonnées, le carré de la distance d'un point i au centre s'exprime par la formule euclidienne la plus simple, grâce à un choix convenable des unités de graduation sur les axes factoriels :

si l'on note $F_1(i)$ l'abscisse du point i sur le 1^o axe factoriel, $F_2(i)$ l'abscisse du point i sur le 2^o axe factoriel etc. le carré de la distance du point i au centre s'écrit :

$$r(i)^2 = (F_1(i))^2 + (F_2(i))^2 + (F_3(i))^2 + (F_4(i))^2 + (F_5(i))^2$$

2.4.2 La contribution $COR_\alpha(i)$ du facteur α à l'écart $(r(i))^2$ du point

i au centre : Dans la somme donnant $(r(i))^2$ en fonction des facteurs (§ 2.4.1 précédent) chaque facteur apporte sa part, plus ou moins importante ; et par là, chaque axe explique, en partie, l'écart du point i au centre. On mesure l'importance d'un axe de n^o α dans l'explication de l'écart du point i au centre par le rapport, noté $COR_\alpha(i)$, du terme $(F_\alpha(i))^2$ au tout : $(r(i))^2$; on a donc :

$$\begin{aligned} COR_\alpha(i) &= (F_\alpha(i))^2 / (r(i))^2 \\ &= (F_\alpha(i))^2 / ((F_1(i))^2 + (F_2(i))^2 + \dots + (F_5(i))^2) \end{aligned}$$

Remarque : $COR_\alpha(i)$ n'est autre que le carré du cosinus de l'angle que fait le rayon O_i avec l'axe α . Lorsque $COR_\alpha(i)$ est voisin de 1, le point i s'écarte de l'origine dans la direction de l'axe α qui explique à lui seul l'écart du point (si l'axe α mesure une qualité, cette qualité est aussi ce qui fait l'originalité du point i). Au contraire, si $COR_\alpha(i)$ est voisin de zéro, le point i s'écarte de l'origine dans une direction perpendiculaire à l'axe α qui n'a alors aucune part dans l'explication de l'écart du point i à l'origine.

2.4.3 La contribution $CTR\alpha(i)$ d'un point i à la valeur propre λ_α (planche XIII)

Les axes eux-mêmes sont faits par les points du nuage : comment reconnaître par quels individus principalement est déterminé un axe ? Ce qui distingue un axe factoriel parmi toutes les droites passant par le centre, c'est l'inertie, relativement forte, du nuage le long de cet axe. Or chaque point participe plus ou moins à cette inertie. De façon précise on a, si l'on considère les maladies :

$$\lambda_\alpha = \underbrace{(masse \text{ de TYPHOIDE}) (F\alpha(\text{TYPHOIDE}))^2}_{\substack{\text{inertie de TYPHOIDE} \\ \text{le long de l'axe } \alpha}} + \dots + \underbrace{(masse \text{ STAPHYL}) (F\alpha(\text{STAPHYL}))^2}_{\substack{\text{inertie de STAPHYLOCOCCIE} \\ \text{le long de l'axe } \alpha}}$$

la part relative, e.g. de TYPHOIDE, dans la valeur propre λ_α est le rapport, noté $CTR\alpha(\text{TYPHOIDE})$, de l'inertie le long de l'axe α de TYPHOIDE à la valeur propre λ_α :

$$CTR\alpha(\text{TYPHOIDE}) = \frac{(6/69) \times (F\alpha(\text{TYPHOIDE}))^2}{\lambda_\alpha}$$

↑
masse de
TYPHOIDE

2.4.4 Les tableaux des facteurs sur les listages (planche XI)

Le listage fournit un tableau pour les maladies et un tableau pour les médicaments ; chaque tableau comporte avec les facteurs les contributions COR et CTR définies ci-dessus. Pour chaque axe α on repère d'un seul regard les fortes valeurs de $CTR\alpha$, indiquant quels points font l'axe α , et les fortes valeurs des $COR\alpha$, indiquant quels points sont typiques de l'axe α (en ce sens que presque tout ce qui distingue ces points du point moyen, origine des axes factoriels, est expliqué par l'axe α). Les nombres figurant dans ces tableaux sont des millièmes.

Ainsi dans notre exemple (pl. XI), ressortent sur l'axe 1 :

■ dans le tableau des maladies,

TYPHOIDE $F_1 = -2,314$; $COR_1 = 994/1000$; $CTR = 530/1000$

SALMONELLOSE $F_1 = -2,314$; $COR_1 = 994/1000$; $CTR = 265/1000$

géométriquement ces deux points sont confondus ; c'est pourquoi ils ont le même COR_1 , d'ailleurs presque égal à 1, ce qui signifie qu'ils sont presque sur l'axe 1 ; mais TYPHOIDE étant deux fois plus lourde que SALMONELLOSE l'inertie de TYPHOIDE le long de l'axe 1 est double de celle de SALMONELLOSE, ce qui explique que la contribution relative de l'une soit $530/1000$ et celle de l'autre $265/1000$; à elles deux elles totalisent $795/1000$ de λ_α ;

■ dans le tableau des médicaments,

tifomycine : $G_1 = -2,019$; $COR_1 = 976/1000$; $CTR_1 = 538/1000$;

tiophénicol : $G_1 = -2,470$; $COR_1 = 915/1000$; $CTR_1 = 302/1000$;

ici encore les deux médicaments contribuent fortement à la détermination de l'axe par leur forte part relative à λ_1 (CTR) et sont situés, géométriquement, au voisinage de l'axe 1 (COR_1 voisin de 1).

Nous verrons, au § 3 suivant, les conclusions à tirer de cette situation.

Sur l'axe 2 ressortent (en millièmes)

AFFECTION DES V. URINAIRES	: F2 = 758 ; COR2 = 724 ; CTR2 = 491
STAPHYLOCOCCIE	: F2 = - 941 ; COR2 = 725 ; CTR2 = 441
<i>érythromycine</i>	: G2 = - 921 ; COR2 = 737 ; CTR2 = 482
<i>gentalline</i>	: G2 = 737 ; COR2 = 759 ; CTR2 = 502

Les signes des facteurs montrent que s'opposent sur l'axe 2 les deux groupes : AFF. des V. URINAIRES et *gentalline* du côté positif ; STAPHYLOCOCCIE et *érythromycine* du côté négatif. L'écart à l'origine de chacun de ces quatre éléments est en grande partie (724/1000 à 759/1000) expliqué par l'axe 2 mais toutefois non exclusivement. Leur contribution à λ_2 est sans commune mesure avec celle des autres éléments : les deux maladies totalisent 932/1000 de λ_2 et les deux médicaments 984/1000 ; ce sont donc ces éléments qui font l'axe 2.

Nous n'examinerons pas les axes suivants car nous n'avons rien à attendre de cet exemple fictif : nous voulons seulement montrer l'usage qu'on peut faire des indications fournies sur les listages à la sortie de l'ordinateur.

3 Interprétation des résultats et enchaînement des analyses

3.0 Interprétation des résultats

On a dit (cf ANA. MED. § 5.2) qu'une étude de pénétration des médicaments devait conduire à repérer les zones vides de médicaments dans le nuage des maladies. Nous n'obtenons rien de tel ici car au lieu de prendre l'ensemble de tous les médicaments commercialisés et l'ensemble de la pathologie, nous avons choisi 6 médicaments assez bien répartis sur 7 maladies. Mais nous pouvons, sur ce modeste exemple, présenter un phénomène courant en analyse des correspondances, quand on dépouille les résultats d'une 1^o analyse : des faits saillants se remarquent immédiatement ; ici, c'est l'opposition de TYPHOÏDE, SALMONELLOSE, *tifomycine* et *tiophénicol* avec le reste des nuages ; et ces quatre éléments déterminent à eux seuls le 1^o axe factoriel (ou axe principal d'allongement du nuage). Cela suggère de refaire l'analyse en mettant ces quatre éléments en supplémentaires (§ 3.1). Il arrive aussi qu'on ait de nombreuses maladies qui sortent groupées parce que leurs profils sont voisins les uns des autres : représenter ce groupe par son centre de gravité peut rendre le graphique plus clair ; or, comme on peut s'en rendre compte en considérant la manière dont on a défini le centre de gravité du nuage tout entier (en prenant le profil de la ligne totalisant toutes les lignes du tableau), le centre de gravité de plusieurs lignes du tableau n'est autre que le profil de la ligne somme des premières. Nous faisons une telle expérience au § 3.2 où l'on cumule les lignes, voisines l'une de l'autre, AF. ORL et PNEUM.

Nous n'aurons ainsi donné qu'un faible aperçu du travail de va-et-vient des tableaux aux facteurs pour finalement dégager les axes intéressants pour le médecin qui doit résoudre un problème.

3.1 Mise d'éléments en supplémentaires

On a vu au § 2.4.6, et l'on vient de le rappeler au § 3.0, que l'axe 1 était fait par un groupe de 2 maladies confondues au même point : TYPHOÏDE et SALMONELLOSE, et deux médicaments : *tifomycine* et *tiophénicol*. En pareil cas, on doit refaire l'analyse en mettant ces quatre éléments en supplémentaires de la façon suivante (planche XIV) :

1^o) On forme un nouveau tableau, appelé tableau principal en ôtant du 1^o tableau de correspondances les deux lignes : TYPHOÏDE et SALMONELLOSE, ainsi que les deux colonnes *tifomycine* et *tiophénicol*.

	<i>pénicilline</i>	<i>tifomycine</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>tiophénicol</i>	<i>gentalline</i>
TYPHOÏDE	0	0	0	0	0	0
SALMONELLOSE DIG.	0	0	0	0	0	0
AFFECTION O.R.L.	8	5	5	3	3	3
PNEUMOPATHIE	7	5	5	2	3	3
MENINGITE	2	0	0	0	1	1
AFF. VOÏES URINAIRES	4	2	2	0	6	6
STAPHYLOCOCCIE	3	1	3	3	0	0

A ce tableau principal, on adjoint, en éléments supplémentaires, les deux maladies TYPHOÏDE et SALMONELLOSE définies par ce qui reste de leur ligne après suppression des deux colonnes *tifomycine* et *tiophénicol* ; et de même on adjoint, en éléments supplémentaires les deux médicaments *tifomycine* et *tiophénicol* définis par ce qui reste de leur colonne après suppression des deux lignes TYPHOÏDE et SALMONELLOSE :

	<i>pénicilline</i>	<i>tétracycline</i>	<i>érythromycine</i>	<i>gentalline</i>	<i>tifomycine</i>	<i>tiophénicol</i>
AFFECTION O.R.L.	8	5	3	3	0	0
PNEUMOPATHIE	7	5	2	3	0	0
MENINGITE	2	0	0	1	2	0
AFF. V. URINAIRES	4	2	0	6	0	0
STAPHYLOCOCC.	3	1	3	0	0	0
TYPHOÏDE	0	0	0	0	} <i>colonnes supplémentaires</i>	
SALMONELLOSE	0	0	0	0	} <i>lignes supplémentaires</i>	

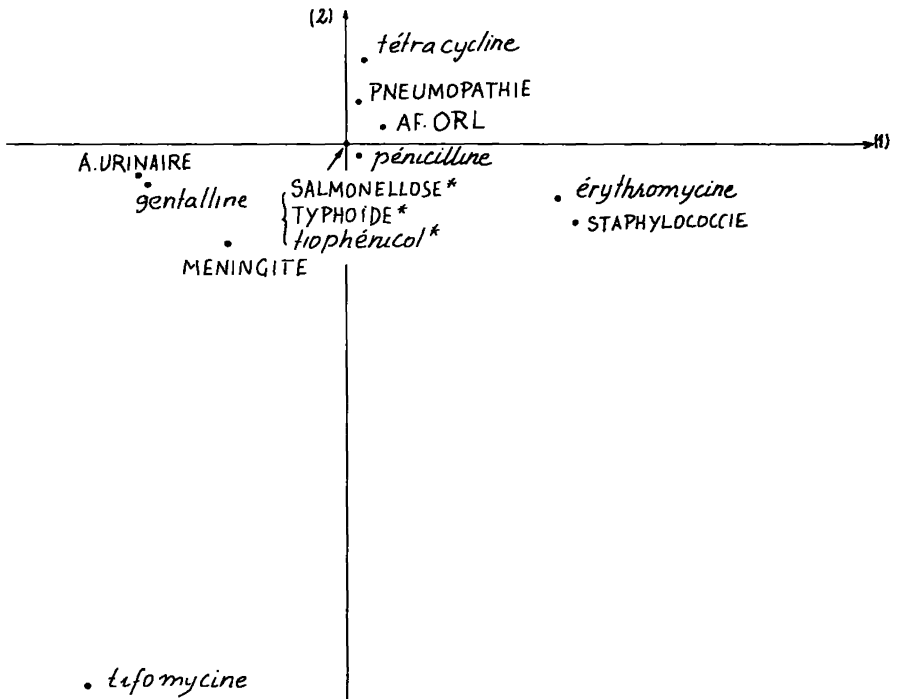
On constate que ces éléments n'ont que des zéros sauf la colonne *tifomycine* qui comporte un 2 correspondant à MENINGITE.

Le programme comporte d'une part l'analyse du tableau principal (à 5 lignes et 4 colonnes) qui fournit trois axes factoriels (le nombre 4 d'éléments du plus petit des deux ensembles en correspondance, moins 1) ; d'autre part la projection sur ces axes des éléments principaux et des éléments supplémentaires. Dans notre exemple particulier, les trois éléments supplémentaires qui n'ont que des zéros sont, par convention, placés à l'origine ; leurs facteurs sont nuls ainsi que les contributions COR et CTR ; *tifomycine* qui comporte un 2 à sa croisée avec la ligne MENINGITE se projette sur les axes du côté de MENINGITE, mais isolé à une très grande distance du reste du nuage.

Comparons les axes issus des deux analyses (planche XV).

Dans la 1° analyse, l'axe 1 sépare le groupe TYPHOÏDE, SALMONELLOSE, *tifomycine* et *tiophénicol* du reste du nuage.

Dans la 2° analyse, ce groupe n'existant plus dans le tableau principal, l'axe a disparu. Et l'on peut remarquer que le 1° axe de la 2° analyse ressemble au 2° axe de la 1° analyse ; on le voit de façon plus précise sur les tableaux des facteurs (planche XVI).



1° analyse : sur le 2° axe ressortent, avec de fortes contributions à λ_2 , et en s'opposant :

AF. V. URIN. :	F2 = 0,758	STAPHYLOCOCCIE :	F2 = - 0,941
gentalline :	G2 = 0,737	érythromycine :	G2 = - 0,921

2° analyse : sur le 1° axe ressortent de même en s'opposant :

AF. V. URIN. :	F1 = - 0,688	STAPHYLOCOCCIE :	F1 = 0,872
gentalline :	G1 = - 0,677	érythromycine :	G1 = 0,861

On remarquera que les orientations des deux axes considérés ne coïncident pas : on a bien les mêmes oppositions mais le groupe qui dans la 1° analyse sort du côté positif du 2° axe, sort dans la 2° analyse du côté négatif du 1° axe. On se souvient que ce qu'on a déterminé, ce sont des droites rendant minimum l'écart pondéré du nuage ; pour en faire des axes de coordonnées, il a fallu choisir un sens positif de parcours : ce choix est arbitraire. Il ne faut donc pas s'étonner qu'une même droite sortant comme axe factoriel dans deux analyses différentes soit orientée différemment dans l'une et l'autre analyses. On retiendra que les signes des facteurs n'ont pas de signification en eux-mêmes : ce qui est important ce sont les conjonctions et les oppositions que ces facteurs révèlent.

De la comparaison des analyses 1 et 2 il ressort que la mise en supplémentaires des éléments qui faisaient l'axe 1 de la 1° analyse entraîne la disparition de cet axe et le décalage des axes suivants. C'est un phénomène courant. Toutefois, il se peut que des éléments qui font le 1° axe dans une analyse contribuent aussi de façon non négligeable aux axes suivants. En ce cas, leur mise en supplémentaires peut changer beaucoup l'analyse.

3.2 Retour au principe d'équivalence distributionnelle : cumul des lignes AF. ORL et PNEUMOPATHIE

Nous avons énoncé ainsi le principe d'équivalence distributionnelle : on ne change pas les résultats de l'analyse en cumulant des lignes proportionnelles entre elles. Ici, nous ajoutons : si les lignes ne sont pas exactement proportionnelles, mais ont des profils voisins les uns des autres, on change peu les résultats de l'analyse en cumulant ces lignes. Outre TYPHOÏDE et SALMONELLOSE qui ont des lignes exactement proportionnelles, donc des profils exactement égaux, le tableau de correspondance que nous avons pris pour exemple comporte deux lignes qui diffèrent peu l'une de l'autre : AFFECTION ORL et PNEUMOPATHIE (cf tableau 2). On peut vérifier sur le graphique (§ 3.1) que leurs profils sont voisins sur le plan (1, 2) ; quant au 3^o facteur, il ne les sépare pas, comme on le voit sur le listage :

$$F3 (ORL) = 0,023 \quad \text{et} \quad F3 (PNEUMOPATHIE) = 0,030.$$

On peut vérifier que les résultats de l'analyse changent peu si l'on cumule ces deux lignes : tableau 3 (planche XVII).

	pénicil.	tétracy.	érythr.	gental.
AF. ORL	8	5	3	3
PNEUMOP	7	5	2	3
MENINGITE	2	0	0	1
AF. URIN.	4	2	0	6
STAPHYLOC	3	1	3	0

Tableau 2

	pénicil.	tétracy.	érythr.	gental.
AF. ORL	} ORP 15	10	5	6
PNEUMOP.				
MENINGITE	2	0	0	1
AF. V URIN	4	2	0	6
STAPHYLOC	3	1	3	0

Tableau 3

	F1	F2	F3
AF ORL (2)	143	91	23
PNEUMO (2)	44	179	30
ORP (3)	97	132	27
MENING. (2)	-437	-388	478
" (3)	-440	-388	475
AF. VURI (2)	-688	-121	-140
" (3)	-689	-115	-141
STAPHYL. (2)	872	-309	-101
(3)	870	-316	-104

	G1	G2	G3
pénicilline (2)	40	-46	154
(3)	40	-46	154
tétracycline (2)	73	332	-66
(3)	79	326	-64
érythromycine (2)	861	-199	-163
(3)	854	-197	-164
gentalline (2)	-677	-125	-118
(3)	-677	-120	-118

Les lignes marquées (2) sont extraites du listage de l'analyse (2) (5 maladies principales) ; les lignes marquées (3) sont extraites du listage de l'analyse (3) (où deux maladies sont cumulées en une seule classe : ORP = AF ORL + PNEUMOP.)

On peut constater en comparant les abscisses sur les axes de même rang des divers points que ceux-ci ne se sont guère déplacés d'une analyse à l'autre ; hormis bien sûr AF. ORL et PNEUMOPATHIE qui ont été remplacés par un point intermédiaire ORP : le facteur 1 de ORP prend une valeur comprise entre celles prises par ce même facteur en les points AF. ORL et PNEUMOP. : $44 < 97 < 143$; de même pour les facteurs 2 et 3.

Remarque : Dans l'analyse 3, on aurait pu mettre AF. ORL et PNEUMOPATHIE en éléments supplémentaires : ORP, qui correspond à la somme de ces deux maladies serait apparue comme leur centre de gravité (leurs masses étant, dans cette analyse, 19/58 et 17/58 : cf planche XVII pour les totaux).

4 Listage de sortie de l'ordinateur : Nous n'avons considéré au § 2 que les principaux résultats imprimés sur le listage de sortie : suite des valeurs propres, facteurs et contributions ; on trouvera ici des indications complémentaires qui auraient alourdi l'exposé.

4.1 Tableau des valeurs propres

- colonne VAL PROPRE : suite des valeurs propres, ou inerties du nuage le long des axes factoriels. Elles sont toutes inférieures à 1 ; elles vont en décroissant de la 1-ère à la dernière ligne : $\lambda_1 > \lambda_2 > \dots$; leur somme, ou trace, donne l'inertie totale du nuage par rapport à son centre : $\lambda_1 + \lambda_2 + \dots = \text{trace}$.

- colonne POURCENT : à chaque ligne α , est imprimé le rapport $\tau_\alpha = \lambda_\alpha / \text{trace}$ exprimé en centièmes, ou % : ce rapport donne l'importance relative de l'axe α quant à l'inertie du nuage.

- colonne CUMUL : à la 1^o ligne on lit τ_1
à la 2^o ligne on lit $\tau_1 + \tau_2$
etc.
à la ligne α on lit $\tau_1 + \tau_2 + \tau_3 + \dots + \tau_{\alpha-1} + \tau_\alpha$,

somme qui mesure l'importance relative de l'espace des α premiers axes quant à l'inertie du nuage.

- l'HISTOGRAMME DES VAL. PROPRES représente chaque valeur propre par un segment dont la longueur lui est proportionnelle : on peut ainsi évaluer d'un regard la décroissance relative de la suite des valeurs propres. Une valeur propre nettement séparée de celle qui précède et de celle qui suit est relative à un axe bien individualisé qu'on cherchera à interpréter en lui même. Deux valeurs propres voisines l'une de l'autre mais bien séparées des autres correspondent à un plan bien individualisé qu'on cherchera à interpréter plutôt que les axes eux-mêmes qui le définissent.

PREMIERE VALEUR PROPRE (TRIVIALE) VAL(1)=1.00000763

NUMI	ITER	VAL PROPRE	POURCENT	CUMUL
2	2	.87832409	75.726	75.726
3	1	.20369929	17.562	93.288
4	2	.04922582	4.244	97.532
5	4	.02857690	2.464	99.996
6	2	.00005007	.004	100.000

NUMI HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE

2	xxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxx
3	xxxxxxxxxxxxxxxx
4	xxx
5	xx
6	

N.B. Les numéros des val. pr donnés dans la colonne NUM sont décalés d'une unité, la val. pr triviale (égale à 1 et ne correspondant pas à un axe factoriel) étant numérotée 1.

La colonne ITER se réfère au procédé de calcul des val pr qui comporte un nombre variable d'itérations, donné dans cette colonne.

IV

TABLEAU DE CORRESPONDANCE AVEC SES MARGES

total de la ligne TYPHOÏDE, + le nombre des occurrences de cette maladie sur la totalité des docteurs étudiés

total de la colonne pénicilline, + le nombre des occurrences de ce médicament sur la totalité des docteurs étudiés

	pénicilline		tétracycline		érythromycine		staphylocoque		total de marge	
TYPHOÏDE	0	4	0	0	2	0	0	0	24	69
SALMONELLOSE DIG	0	2	0	0	1	0	0	0	8	13
AFFECTION ORL	8	0	5	3	0	3	19		13	69
PNEUMOPATHIE	7	0	5	2	0	3	17		13	69
MENINGITE	2	2	0	0	0	1	5		8	13
AFFECT VONES URIN	4	0	2	0	0	6	12		3	13
STAPHYLOCOCCIE	3	0	1	3	0	0	7		8	13
ligne de marge	24	8	13	8	3	13	69			

total de la colonne pénicilline, + le nombre des occurrences de ce médicament sur la totalité des docteurs étudiés

EXEMPLE DE CALCUL DE PROFIL

MENINGITE: 2/5, 2/5, 0, 0, 0, 1/5

ce qui se peut exprimer en pourcentages:

% des cas de MENING 40% 40% 0% 0% 0% 20%

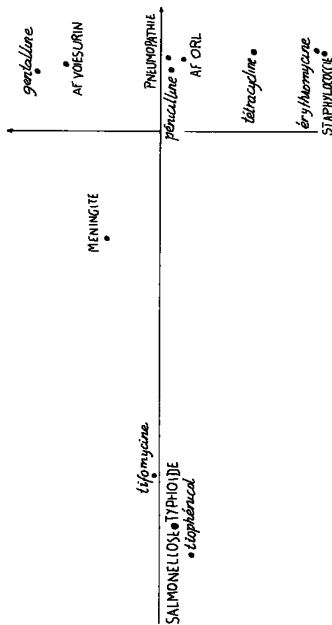
se rencontrant avec: pénicilline tétracycline érythromycine gentamicine

pénicilline 0, 0, 8/24, 7/24, 2/24, 4/24, 3/24

ce qui est donc %: 0%, 0%, 30%, 29%, 8%, 16%, 12%

avec TYPH SAL ORL PNEUM MENING AF URI STAPH

III



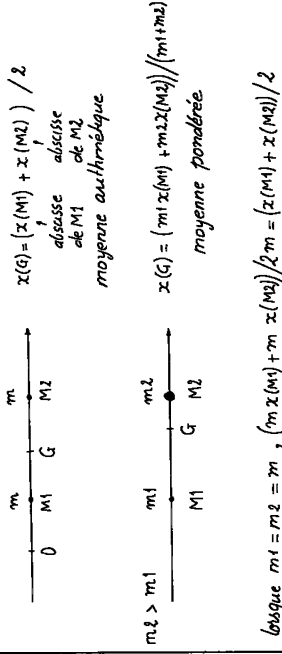
Le graphique ci-dessus est à comparer au tableau de correspondance ci-dessous

pénicilline tétracycline érythromycine gentamicine

TYPHOÏDE	0	4	0	0	2	0
SALMONELLOSE	0	2	0	0	1	0
AFFECT ORL	8	0	5	3	0	3
PNEUMOPATHIE	7	0	5	2	0	3
MENINGITE	2	0	0	0	0	1
AF V URIN	4	0	2	0	0	6
STAPHYLOC	3	0	1	3	0	0

VII

CENTRE DE GRAVITE



Centre de gravité du nuage des maladies:
 1° coordonnées (sur pénicilline). moyenne pondérée des 1° coord des 7 profils de maladies (la somme des masses des maladies valant 1, on n'a eut que le numérateur)

$0 \cdot 0 + [(8/19) \times (19/69)] + [(7/17) \times (17/69)] + [(2/5) \times (5/69)]$
 $+ [(4/12) \times (12/69)] + [(3/7) \times (7/69)]$
 ↑ absc. ORL ↑ masse PNEUM ↑ absc. PNEUM ↑ masse MENING
 ↑ absc. V URIN ↑ masse V URIN ↑ absc. STAPHYL ↑ masse STAPHYL

après simplification, il reste $(8 + 7 + 2 + 4 + 3) / 69$
 total de la col pénicill dans le tab de Cor
 total général du tab de Cor
 i.e la 1° coordonnée (sur pénicilline) du profil de la ligne de marge du tableau de correspondance

→ on trouve ainsi, coordonnées par coordonnées, que le Centre de grav. du nuage des maladies est le profil de la ligne de marge du tableau de correspondance.

V

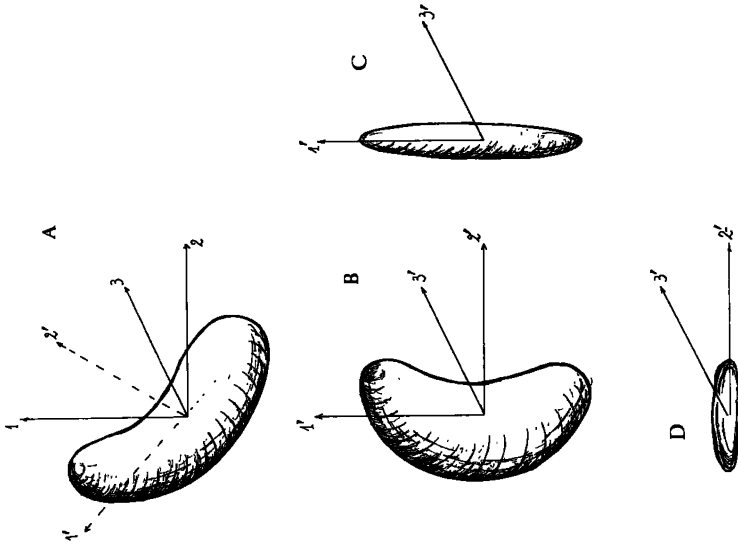
TABLEAU DES PROFILS DES MALADIES AVEC LES MASSES à lire horizontalement

total de la Tab de cor.	pénicilline	tétracycline	érythromyc.	thiophène	gentamicine	masse
6	TYPHOIDE 0	4/6	0	0	2/6	0
3	SALMONEL 0	2/3	0	0	1/3	0
19	AF ORL 8/19	0	5/19	3/19	0	3/19
17	PNEUMOP 7/17	0	5/17	2/17	0	3/17
5	MENINGITE 2/5	2/5	0	0	0	1/5
12	AF V URIN 4/12	0	2/12	0	0	6/12
7	STAPHYLOC 3/7	0	1/7	3/7	0	0
69	de marge 24/69	8/69	13/69	8/69	3/69	43/69

69/69=1

TABLEAUX DES PROFILS DES MEDICAMENTS AVEC LES MASSES à lire verticalement

total de la Tab de cor.	24	8	13	8	3	13	69
col de marge	pénicilline	tétracycline	érythromyc.	thiophène	gentamicine	masse	col de marge
6/69	TYPHOIDE 0	4/8	0	0	2/3	0	6/69
3/69	SALMONEL 0	2/8	0	0	1/3	0	3/69
19/69	AF ORL 8/24	0	5/13	3/8	0	3/13	19/69
17/69	PNEUMOP 7/24	0	5/13	2/8	0	3/13	17/69
5/69	MENINGITE 2/24	2/8	0	0	0	1/13	5/69
12/69	AF V URIN 4/24	0	2/13	0	0	6/13	12/69
7/69	STAPHYLOC 3/24	0	1/13	3/8	0	0	7/69
69/69=1	masse 24/69	8/69	13/69	8/69	3/69	43/69	69/69=1



VI

On considère à-combe un objet de forme néphroïde aplati.
 En haut, à gauche, fig A, il est désigné de façon quelconque
 par rapport au système des axes 1, 2, 3. Il s'empare de chaque
 un nouveau 1' axe 1' suivant la direction principale d'allonge-
 ment, et un deuxième axe 2', perpendiculaire à 1' et défini-
 ment avec 1' le plan moyen de l'objet. Le 3' axe 3' sera pris
 perpendiculaire aux deux premiers, 1' et 2'.

Dans les figures B, C, D, les nouveaux axes sont pris comme
 directions principales de la représentation :

sur la figure B, le plan moyen de l'objet, (1', 2'), soit de plan
 frontal et l'objet apparaît avec l'épaisseur maximum;

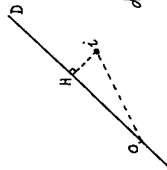
sur la figure C, on considère la projection sur le plan (1', 3') :
 la largeur 2', apparaît seulement l'épaisseur 3' qui est la mon-
 due des deux dimensions de l'objet;

enfin, sur la figure D, on a la projection sur le plan (2', 3'),
 perpendiculaire à l'axe principal de l'objet : la dimension lon-
 gitudinale n'est plus visible, seuls subsistent les deux dimen-
 sions mineures de largeur et d'épaisseur.

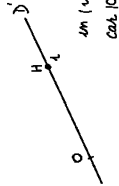
VIII

INERTIE (1)

i de masse f_i point masse
 inertie de i ; f_i par rapport au centre O
 $f_i \times d^2(O, i)$
 $(f_i \cdot i)^2$
 masse distance au carré
 du point du point i au centre O



$d^2(O, i) = d^2(O, H) + d^2(H, i)$
 ainsi $f_i \cdot d^2(O, i) = f_i \cdot d^2(O, H) + f_i \cdot d^2(H, i)$
 inertie de i inertie de i
 par rapport à O le long de D transversalement
 $m_i \cdot i$ m_i / D $m_i \cdot D$



$m_i \cdot i = m_i / D \cdot i$; $m_i \cdot l \cdot D \cdot i = 0$
 car $l \cdot H = 0$



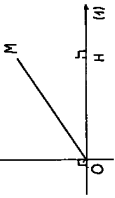
$m_i \cdot i = m_i / D' \cdot i$; $m_i / D' \cdot i = 0$
 car $l \cdot i \cdot H \cdot i$; $car l \cdot O \cdot H = 0$

IX

DISTANCE

(2)

Dans le triangle OMH, rectangle en H,
 $d^2(O, M) = d^2(O, H) + d^2(M, H)$
 distance O, M distance O, H distance M, H
 au carré au carré au carré



à condition de choisir comme unité de graduation sur les axes l'unité de distance dans le plan (e.g. le cm), on a

$d(O, H) = |x_1(M)|$, val absolue de la coord de M sur l'axe (1),
 $d(M, H) = |x_2(M)|$, " " " " " "
 d'où: $d^2(O, M) = (x_1(M))^2 + (x_2(M))^2$

Un changement d'unité de graduation sur les axes introduit des coefficients dans la formule de distance.

si l'axe (1) est gradué tous les 3 cm, l'axe (2) tous les 1/4 cm,

$d(O, H) = 3 \cdot |x_1(M)|$
 $d(M, H) = (1/4) |x_2(M)|$
 où: $d^2(O, M) = 9 (x_1(M))^2 + 1/4 (x_2(M))^2$

Le usage des mètres est enclavé dans un espace à 6 dimensions; Dans l'espace à 6 dimensions, la formule générale de distance euclidienne s'écrit

$d^2(O, M) = \alpha_1 (x_1(M))^2 + \alpha_2 (x_2(M))^2 + \alpha_3 (x_3(M))^2$
 $+ \alpha_4 (x_4(M))^2 + \alpha_5 (x_5(M))^2 + \alpha_6 (x_6(M))^2$

Principe d'équivalence dimensionnelle: on ne change pas l'analyse en cumulant en une seule ligne plusieurs lignes proportionnelles entre elles (et idem pour les colonnes).

Ce principe impose comme coefficients α_i , les inverses des rapports cumulés le profil de la ligne de marge du tabl de con

XIII

CONTRIBUTIONS AUX INERTIES

$$\lambda_\alpha = \frac{\text{masse de TYPHOÏDE} \cdot (F_\alpha(\text{TYPH}))^2 + \text{masse de SALMONELLE} \cdot (F_\alpha(\text{SALMON}))^2 + \text{masse de AF-ORL} \cdot (F_\alpha(\text{ORLU}))^2 + \text{masse de MENINGITE} \cdot (F_\alpha(\text{MENINGITE}))^2 + \text{masse de STAPHYLOCOCCIE} \cdot (F_\alpha(\text{STAPHYLOCOCCIE}))^2}{\text{distance à l'origine de la projection sur l'axe } \alpha \text{ de TYPHOÏDE} + \text{distance à l'origine de la projection sur l'axe } \alpha \text{ de SALMONELLE} + \text{distance à l'origine de la projection sur l'axe } \alpha \text{ de AF-ORL} + \text{distance à l'origine de la projection sur l'axe } \alpha \text{ de MENINGITE} + \text{distance à l'origine de la projection sur l'axe } \alpha \text{ de STAPHYLOCOCCIE}}$$

CTR α (i) : part relative de la maladie i à λ_α , inertie du nuage le long de l'axe α .

e.g. CTR α (MENINGITE) = $(5/69) \cdot (F_\alpha(\text{MENINGITE}))^2 / \lambda_\alpha$
 pour $\alpha = 1$, $F_1(\text{MEN}) = -0.625$; $(F_1(\text{MEN}))^2 = 0.390$; $(6/69) \cdot 0.390 = 0.0283$
 $\lambda_1 = 0.87$; $0.0283 / 0.87 = 0.032$, nombre qui figure dans le tableau des facteurs (p.XIII) ligne MENINGITE, colonne CTR 1.

XII

LES FACTEURS

F1, F2, F3, F4, F5 : coordonnées sur les axes factoriels des maladies

G1, G2, G3, G4, G5 : coordonnées sur les axes factoriels des médicaments

Dans l'espace on est défini le nuage des maladies, distance au carré de i à l'origine:

$$d^2(0, i) = (F1(i))^2 + (F2(i))^2 + (F3(i))^2 + (F4(i))^2 + (F5(i))^2$$

centres du nuage de i sur l'axe, coordonnées au carré

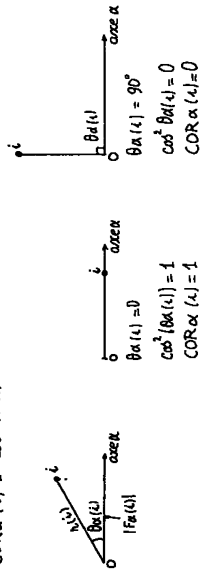
$$\text{COR } \alpha (i) : \text{part relative du facteur } F_\alpha \text{ dans } d^2(0, i) = (F_\alpha(i))^2$$

e.g. $\text{COR } 3 (i) = \frac{(F3(i))^2}{(F1(i))^2 + (F2(i))^2 + (F3(i))^2 + (F4(i))^2 + (F5(i))^2}$

$$d^2(0, i) \text{ ou encore } (d(i))^2$$

distance au carré de i à l'origine

$$\text{COR } \alpha (i) = \cos^2 \theta_{\alpha}(i)$$



XIV

MISE D'ÉLÉMENTS
EN SUPPLÉMENTAIRES

	penicilline	tétracycline	érythromycine	gentamicine
TYPHOÏDE	10	10	10	10
SALMONEL	10	10	10	10
AF ORL	8	5	3	3
PNEUMOPATH	7	5	2	3
MENINGITE	2	0	0	1
AF URINAIRES	4	2	0	6
STAPHYLOCOC	3	1	3	0

	penicilline	tétracycline	érythromycine	gentamicine	total	létomycine	létomycine
AFFECT. ORL	8	5	3	3	19	0	0
PNEUMOPATH	7	5	2	3	17	0	0
MENINGITE	2	0	0	1	3	2	0
AF V URIN	4	2	0	6	12	0	0
STAPHYLOC.	3	1	3	0	7	0	0
Total	24	13	8	13	58	2	0
TYPHOÏDE	0	0	0	0	0	autres supplém	
SALMONEL	0	0	0	0	0	lignes supplémentaires	

XXV

ANALYSE 1

	(2)	(3)	(4)
SALMONEL	• tétromycine	• tétracycline	• pénicilline
TYPHOÏDE	• Typhoïde		
• Urophérid			
MENINGITE			
PNEUMOPATHIE			
AF ORL			
érythromycine			
STAPHYLOCOC			

	(2)	(3)	(4)
AFFECT. URIN.	• gentamicine	• tétracycline	• érythromycine
SALMONELLOSIS*			
TYPHOÏDE*			
• Urophérid*			
MENINGITE			
PNEUMOPATHIE			
• AF ORL			
• pénicilline			
• STAPHYLOCOCCIE			

ANALYSE 2

4 éléments, marqués d'un *, sont en supplémentaires

• létomycine*

XVII

CUMUL DE DEUX LIGNES

ANALYSE 2:

	pénicilline tétracycline érythromyc gentalline				total
AF ORL	8	5	3	3	19
ORP { PNEUMOP	7	5	2	3	17
MENINGITE	2	0	0	1	3
AF V URIN	4	2	0	6	12
STAPHYLOC	3	1	3	0	7
total	24	13	8	13	58

ANALYSE 3:

	pénicilline tétracycline érythromyc gentalline				total
ORL { DRP	15	10	5	6	36 = 19+17
PNE	2	0	0	1	3
MENINGITE	2	0	0	1	3
AF V URIN	4	2	0	6	12
STAPHYLOC	3	1	3	0	7
total	24	13	8	13	58

	F 1	F 2	F 3	G 1	G 2	G 3
AF ORL (2)	143	91	23	pénicilline (2)	40	-46
PNEUMOP (2)	44	179	30	(3)	40	-46
ORP (3)	91	132	27	tétracycline (2)	73	332
MENINGITE (2)	437	-388	478	(3)	79	325
(3)	440	-388	475	érythromycine (2)	861	-199
AF URIN (2)	688	-121	-140	(3)	854	-191
(3)	689	-115	-141	gentalline (2)	-677	-125
STAPHYLOC (2)	872	-309	-101	(3)	-677	-120
(3)	870	-316	-104			-118

Les lignes marquées (2) se réfèrent à l'analyse 2; celles marquées 3 à l'analyse 3.

XVIII

ANALYSE 2

VAL PROPRES	λ	τ	POURCENT	CUMUL
$-.20 = -\lambda_1$	$79.9 = \tau_1$	$79.9 = \tau_1$		
$-.034 = \lambda_2$	$13.3 = \tau_2$	$93.2 = \tau_1 + \tau_2$		
$-.017 = \lambda_3$	$6.8 = \tau_3$	$100 = \tau_1 + \tau_2 + \tau_3$		

	POID	INR	1#F	COR	CTR	2#F	COR	CTR
AF ORL	328	37	143	699	33	91	283	79
PNEUMOPATH	293	40	44	54	3	179	919	273
MENINGITE	52	114	-437	335	48	-388	264	226
AF V URIN	207	406	-688	932	473	-121	29	88
STAPHYL.	121	404	872	878	444	-309	110	335
pénicilline	414	44	40	60	3	-46	76	25
tétracycline	224	104	73	44	3	332	920	715
érythromycine	138	430	861	918	494	-199	49	158
gentalline	224	423	-677	939	497	-125	32	102
AF V URIN	758	724	491					
STAPHYL	-941	725	441					
érythromycine	821	737	482					
gentalline	737	759	502					

EXTRAIT DU LISTAGE DE L'ANALYSE I