

L'Analyse discriminante

Introduction

Présentation

L'analyse discriminante connue dans la pratique marketing sous le nom "**scoring**" essaye de déterminer la contribution des variables qui expliquent l'appartenance des individus à des groupes. Deux ou plusieurs groupes sont comparés, pour déterminer s'il s'ils diffèrent et pour comprendre la nature de ces différences. En marketing il est important d'identifier des caractéristiques qui déterminent l'appartenance à des groupes, qui permettraient de distinguer entre:

- utilisateurs permanents et occasionnels d'un produit;
- acheteurs d'une marque et les acheteurs de marques concurrentes;
- clients loyaux et aloyaux
- vendeurs bons, médiocres et mauvais ¹

Exemple:

Par exemple suite à un concours de vente on a réussi de séparer trois groupes de vendeurs. Pour identifier les éléments (les variables) qui permettent de distinguer entre un bon et un mauvais vendeur, on a appliqué un questionnaire au participants a ce concours dans le quel on s'est intéressé d'une manière prioritaire à quatre caractéristiques : le nombre de contacts avec des nouveaux clients ; la proportion de contacts avec rendezu d'avance; les coups de téléphone aux prospects et le nombre de nouveau comptes visités².

Tableau 1

Activations des vendeurs concernant les nouveaux clients

Individu	Nombre de contacts nouveaux clients	Proportion de contacts avec rendezu d'avance	Coups de téléphone aux prospects	Nombre de nouveau comptes visités
	X1	X2	X3	X4
Gagneurs du grand prix (G)				
1 KZV	130	62	148	42
2 BOR	122	70	186	44
3 NUA	89	68	171	32
4 NOA	104	58	135	40

¹Gilbert A. Churchill, "Marketing Research, Methodological Foundations", 5e Ed., Dryden Press, 1991.

²ibidem

5 HJV	116	40	160	36
6 VRX	100	65	151	30
7 BFO	85	66	183	42
8 MIG	113	59	130	25
9 WLL	108	52	163	41
10 GJN	116	48	154	48
11OKX	99	57	188	32
12KAA	78	70	190	40
13 UHO	106	61	157	38
14 GCI	94	58	173	29
15 CSN	98	64	37	36
Gagneurs du prix de consolation (C)				
1 XDC	105	39	155	45
2 OAH	86	60	140	33
3 VKS	64	48	132	36
4 EXL	104	36	119	29
5 PQT	102	53	143	41
6 ITE	73	62	128	30
7 OXX	94	51	152	36
8 DON	59	64	130	28
9 CIG	84	31	102	32
10 GID	91	47	96	35
11 PLO	83	40	87	30
12 KOJ	95	42	114	28
13 DIN	68	52	123	26
14 RZI	101	51	98	24
15 UOE	89	39	117	33
Vendeurs sans succès (S)				
1 MJB	80	23	69	32
2 VKN	47	42	74	33
3 XOG	26	37	132	20
4 AUT	94	24	68	26
5 ZCU	57	32	94	23
6 AKY	38	41	83	28
7 HZL	29	52	96	22
8 NHN	48	24	73	26
9 KZW	57	36	82	28
10 ZUB	39	37	98	21
11 NFJ	51	38	117	24
12 TFJ	40	42	112	22
13RLM	64	21	67	29
14 XGF	35	32	78	25
15AMA	51	29	81	26

Modèle géométrique

Cas à deux variables

Pour simplifier on prend deux groupes de sujets (vendeurs) mesurés sur deux variables X, Y . Une première variable indiquant le nombre de contacts avec des nouveaux clients sera notée Y et la deuxième variable, proportion des contacts par rendez-vous sera notée par X .

Les réponses des individus à ces questions, sont représentées dans la figure 1. Les deux questions représentent les variables selon les quelles on aimerait arriver à distinguer les deux groupes.

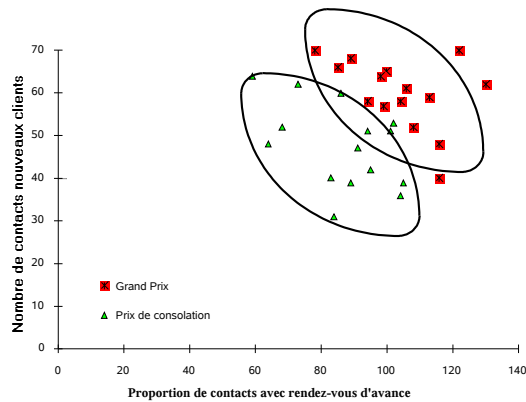


Figure 1

En analysant le graphique on se rend compte que sur chacune des axes représentées par les deux variables il y a une région importante d'incertitude, dans la quelle pour les même valeurs des variables on trouve des individus appartenant au deux groupes (bons et mauvais vendeurs).

Objectif

Le but de l'analyse discriminante serait de trouver un nouvelle axe comme combinaison linéaire des variables qui permettrait de réduire cette zone d'incertitude. Une telle axe dans l'exemple donné est illustré dans la figure 2.

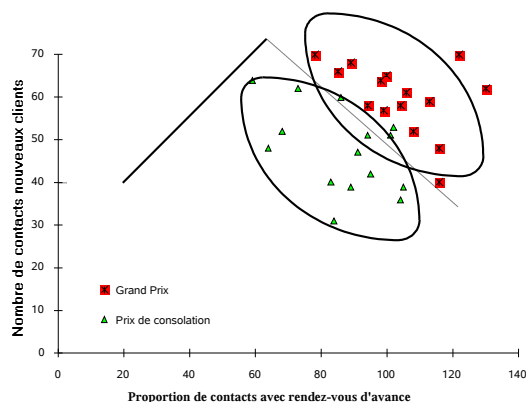


Figure 2

Cas générale

Imaginons maintenant, d'une manière plus générale, que ces variables sont centrées pour l'ensemble des deux groupes.

On illustre dans un espace à deux dimensions les coordonnées de chaque sujet sur les variables en question et on trace les lignes contours de 95 % ($1,96 \sigma$) Ou 68 % (1 σ). Les hypothèses de normalité des variables conjointes et d'homogénéité des variances et covariances sont maintenues³.

La figure 3 en est un exemple.

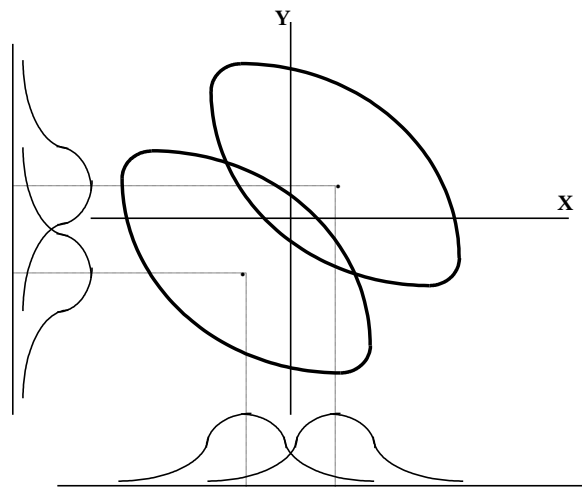


Fig.3 Lignes contours de deux échantillons de variables conjointes homogènes et normalement distribuées.

Dans la démarche habituelle de l'analyse, on éprouvera d'abord le besoin de déterminer s'il faut considérer les moyennes ou centroides de chaque population comme distinctes: L'analyse de variance multiple est un instrument bien indiqué pour cette tâche. Il apparaît ensuite intéressant, connaissant les scores d'un éventuel sujet sur les deux variables, de déterminer, à l'intérieur d'un pourcentage d'erreur, l'appartenance de ce sujet à l'une ou l'autre de ces populations.

Zones d'incertitude et attribution d'appartenance

Si l'on se base, pour une telle attribution d'appartenance, sur la connaissance du seul score X , on voit que l'incertitude de la décision va de a à b et touche les sujets situés dans la région hachurée de la figure 4.

³Hubert Laforge "Analyse multivariée pour les sciences sociales et biologiques avec applications des logiciels BMP, BMDP, SPSS, SAS", Ed. Etudes Vivantes, Montréal, 1981.

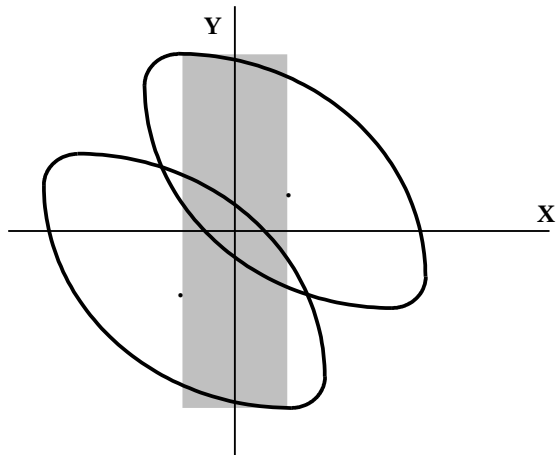


Fig.4 Région d'incertitude basée sur la variable X

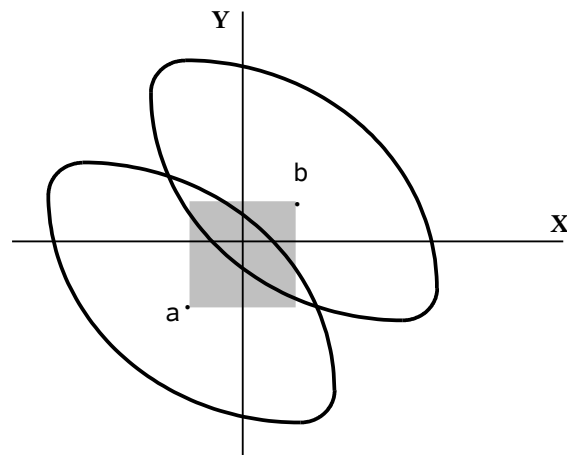


Fig. 6 Région d'incertitude basée sur deux variables

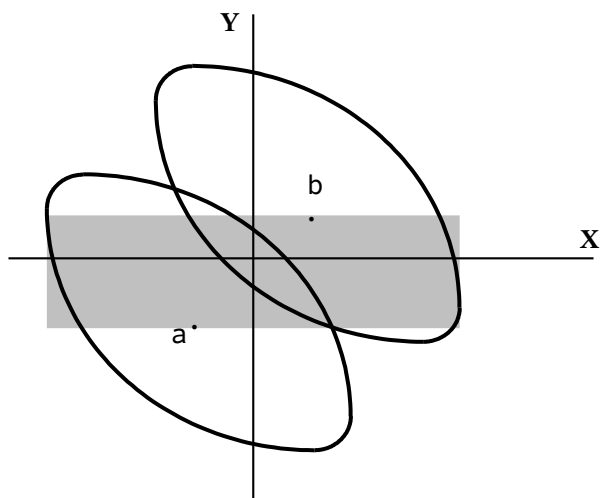


Fig. 5 Région d'incertitude basée sur la variable Y

La région sure

Si cette décision est prise en tenant compte des deux variables, on constate que la région sure correspond alors à la surface hachurée

Région minimum d'incertitude
 Cette seconde façon de procéder, quoique meilleure que la première peut comporter encore de grosses imprécisions (sauf cas particuliers) car la véritable région d'incertitude est celle de l'intersection des deux ellipses

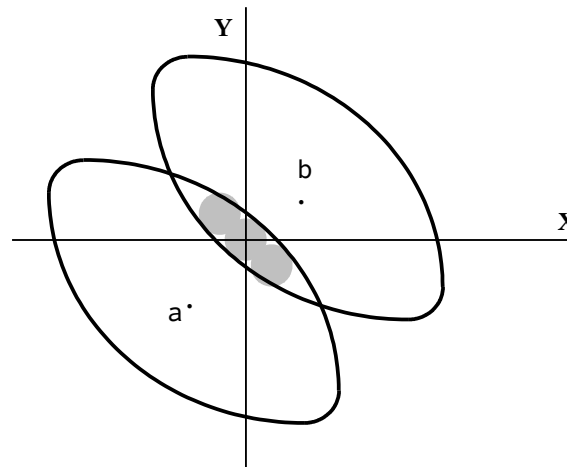


Fig. 7 Région minimum d'incertitude

Recherche des axes discriminants

On propose de considérer une nouvelle variable qui soit une combinaison linéaire des précédentes; géométriquement, cette nouvelle variable est représentée par un axe sur lequel on projette les divers points des groupes de sujets. Aux fins d'illustration, on limite le nombre de variables de départ ainsi que le nombre de groupes à deux; dans la pratique, le modèle mathématique dont il sera question à l'article suivant ne comporte pas de telles limites.

L'axe est appelé axe de la fonction discriminante. Les points projetés sur cet axe se distribuent normalement pour chacun des groupes; les valeurs de cette fonction comprises entre les lignes pointillées correspondent à la

région d'incertitude. On désire que l'axe occupe une position telle que la projection des points donne lieu au minimum de superposition des divers groupes de sujets. La figure 8 illustre la situation particulière de la recherche de la fonction de deux variables qui discrimine au maximum deux groupes de cas.

Maximiser le pouvoir discriminant

La qualité de la discrimination est liée à la superposition des deux distributions de projections sur l'axe. On peut mesurer la qualité de la dispersion à la grandeur du rapport de la variance entre les moyennes à la variance à l'intérieur d'un groupe

; ce rapport est analogue au F de l'analyse de variance. On suppose que la variance des scores à l'intérieur de chaque groupe répond au critère d'homogénéité de telle sorte que cette variance intra est la moyenne des variances intra des groupes considérés.

Un rapport maximum est lié non seulement à la grandeur de son numérateur mais aussi à l'étroitesse du dénominateur: la variance inter atteindra son maximum pour l'axe parallèle au segment joignant les centroides tandis que la variance intra sera minimum pour un axe perpendiculaire à l'axe principal des ellipses; c'est en une position intermédiaire que se situe le rapport maximum des variances inter et intra.

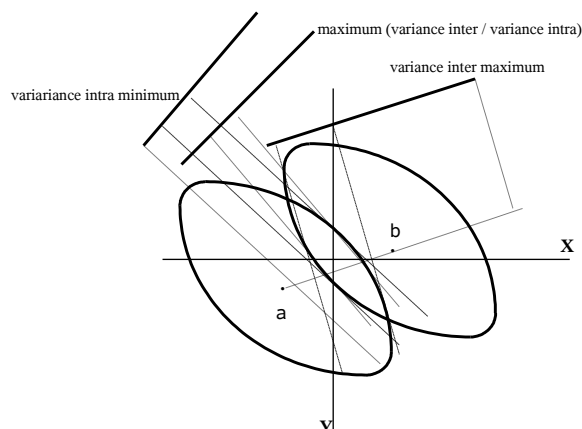


Figure 8. Modèle géométrique de la fonction discriminante.

Resultats

L'analyse discriminante a pour but de déterminer cet axe optimum de la fonction discriminante, c'est-à-dire de calculer les éléments d'un vecteur k qui définissent une combinaison linéaire des variables X et Y . Ces exigences géométriques peuvent être avantageusement traitées à la manière de l'extraction des composantes principales, car des fonctions discriminantes orthogonales successives s'apparentent à ces composantes. On comprend alors que le nombre possible de ces fonctions soit limité par le nombre de variables et aussi par le nombre de groupes: ce qui est dans le sens d'une interprétation plus aisée de la discrimination. Ceci est d'ailleurs facilité par le fait que

des fonctions discriminantes additionnelles sont responsables d'une proportion de moins en moins élevée du pouvoir discriminatoire total; il est souvent souhaitable de limiter à deux ou à trois ces fonctions discriminantes afin d'en rendre possible l'illustration géométrique et donc l'interprétation (voir Cooley et Lohnes, 1971, p. 244-246).

Les coefficients de la fonction discriminante

Utilisation des coefficients

Une manière de déterminer quelles sont les variables qui discriminent entre les deux types de gagnants aux concours de ventes est de construire un index, basé sur les valeurs des caractéristiques mesurées, qui sépare les deux groupes, formant une combinaison linéaire de ces dernières du genre:

$$Y = v_1 X_1 + v_2 X_2 + v_3 X_3 + v_4 X_4,$$

ou v_1 ; v_2 ; v_3 ; v_4 sont des coefficients arbitraires. En analyse discriminante les coefficients sont dérivés de telle manière que la variation des scores de Y entre les (deux) groupes soit si large que possible et la variation des scores de Y à l'intérieur des groupes (within group - intra groupe) soit si petite que possible. Autrement dit on calcule les coefficients qui maximisent le rapport

Ce ci fait les groupes aussi distincts que possible du point de vue des nouveaux scores de l'index.

Pour l'exemple analysé les coefficients discriminants sont $v_1 = 0,059$, $v_2 = 0,063$, $v_3 = 0,034$ et $v_4 = -0,032$, et la combinaison linéaire qui différencie de manière maximales entre les groupes est⁴

$$Y = 0,059 X_1 + 0,063 X_2 + 0,034 X_3 - 0,032 X_4,$$

Calcul des scores sur les axes discriminants

Ayant les coefficients discriminants on peut calculer le score de chaque vendeur, si ce score est plus proche de la moyenne des scores du groupe des gagnants du grand prix, l'individu sera affecté à ce groupe si non il sera affecté à l'autre groupe.

On peut observer que l'approche de l'analyse discriminante est proche de celle de la **regression**. Dans les deux cas on essaye d'expliquer (prévoir) une variable dépendante par une combinaison linéaire de variables indépendantes. En regression la variable expliquée est continue. En analyse discriminante la variable dépendante est l'appartenance à un groupe. On peut transformer le problème d'analyse discriminante pour de groupes en problème de regression en utilisant une variable muette (dummy) comme variable

⁴ les coefficients (poids) sont donnés par le vecteur propre qui résulte de la résolution de l'équation:

$$(W^{-1}B - \lambda I)v = 0,$$

où W est la matrice des variances-covariances intra-groupes (**within group**) et B est la matrice des sommes des carrées et produits croisés inter-groupes (**between group**)

dépendante. Les coefficients de regression résultants seront proportionnels à ceux obtenus par l'analyse discriminante.

Tableau 2
Scores calculés des gagnants du grand prix et du prix de consolation utilisant la fonction discriminante

$$Y = 0,059 X_1 + 0,063 X_2 + 0,034 X_3 - 0,032 X_4$$

	X1	X2	X3	X4	Y
Gagnants du grand prix (G)					
1 KZV	130	62	148	42	15,1
2 BOR	122	70	186	44	16,4
3 NUA	89	68	171	32	14,2
4 NOA	104	58	135	40	13,0
5 HJV	116	40	160	36	13,5
6 VRX	100	65	151	30	14,1
7 BFO	85	66	183	42	14,0
8 MIG	113	59	130	25	13,9
9 WLL	108	52	163	41	13,8
10 GJN	116	48	154	48	13,5
11 OKX	99	57	188	32	14,7
12 KAA	78	70	190	40	14,1
13 UHO	106	61	157	38	14,1
14 GCI	94	58	173	29	14,1
15 CSN	98	64	37	36	9,8
Moyenn e	103,9	59,9	155,1	37,0	13,9
Gagnants du prix de consolation (C)					
1 XDC	105	39	155	45	12,4
2 OAH	86	60	140	33	12,5
3 VKS	64	48	132	36	10,1
4 EXL	104	36	119	29	11,4
5 PQT	102	53	143	41	12,8
6 ITE	73	62	128	30	11,5
7 OXX	94	51	152	36	12,7
8 DON	59	64	130	28	11,0
9 CIG	84	31	102	32	9,3
10 GID	91	47	96	35	10,4
11 PLO	83	40	87	30	9,3
12 KOJ	95	42	114	28	11,1
13 DIN	68	52	123	26	10,6
14 RZI	101	51	98	24	11,6
15 UOE	89	39	117	33	10,5
Moyenn e	86,5	47,7	122,4	32,4	11,1

Interpretation de la fonction discriminante

Tester la différenciation obtenue

Dans une approche rigoureuse avant d'interpréter la fonction discriminante, on doit tester si au niveau des scores discriminants on obtient une différenciation significative entre les groupes. Cela se fait en appliquant un test en F aux valeurs de la statistique D^2 de Mahalanobis (qui mesure la distance de chaque case à la moyenne du groupe, tout en permettant des axes corrélés et des unités de mesure différentes).

Les coefficients discriminants

La fonction discriminante originale contient des poids (coefficients) à appliquer aux valeurs brutes des variables. Ces poids sont influencés par l'échelle de mesure des variables. Pour éviter des effets d'échelle de mesure arbitraire *quand on compare les contributions de chaque variable on utilise des coefficients standardisés*, obtenus par multiplication des coefficients bruts de chaque variable par l'écart type pour l'ensemble des groupes (pooled standard deviation) de.

Classification des individus utilisant la fonction discriminante

Calcul de cutting score

Pour classer les individus dans un des groupes on doit fixer un score (cutting score) qui joue le rôle de frontière entre les groupes. Normalement c'est la moyenne des scores des deux groupes. Si les groupes sont de dimensions égales le **cutting score** (Y_{CS}) est égale à la moyenne des moyennes des scores des groupes.

$$Y_{CS} = (G + C) / 2 = (14,2 + 11,2) / 2 = 12,7.$$

Si les groupes ne sont pas égaux on utilisera une moyenne pondérée du genre:

$$CS =$$

où \bar{Y}_1 et \bar{Y}_2 sont les scores discriminants moyens et n_1 et n_2 sont les dimensions des groupes (on observe que la moyenne de chaque groupe est pondérée avec la dimension de l'autre).

Attribution aux groupes

Dans le cas du concours des vendeurs, une règle de décision simple serait de classer un vendeur dans le groupe des gagnants du grand prix si son score est plus proche de la moyenne des scores du groupe des gagnants du grand prix, que de la moyenne des scores du groupe des gagnants du prix de consolation, si non il sera affecté au groupe des gagnants du prix de consolation.

Tableau 3

L'appartenance de groupe prédite utilisant la règle de classification simple

Differences par rapport à la moyenne du groupe

C

0 13

Score discriminant Premier groupe Deuxieme groupe groupe predite

Yi Yi - Yg Yi - Yc

Gagneurs du grand prix (G)

1	15,1	1,0	4,0	G
2	16,4	2,3	5,3	G
3	14,2	0,1	3,1	G
4	13,0	-1,1	1,8	G
5	13,5	-0,6	2,4	G
6	14,1	0,0	2,9	G
7	14,0	-0,1	2,8	G
8	13,9	-0,2	2,7	G
9	13,8	-0,3	2,6	G
10	13,5	-0,7	2,3	G
11	14,7	0,6	3,6	G
12	14,1	0,0	3,0	G
13	14,1	0,0	3,0	G
14	14,1	-0,1	2,9	G
15	13,2	-0,9	2,1	G

Gagneurs du prix de consolation (C)

1	12,4	-1,7	1,2	C
2	12,5	-1,6	1,3	C
3	10,1	-4,0	-1,1	C
4	11,4	-2,7	0,3	C
5	12,8	-1,3	1,7	G
6	11,5	-2,6	0,4	C
7	12,7	-1,4	1,5	G
8	11,0	-3,1	-0,2	C
9	9,3	-4,8	-1,9	C
10	10,4	-3,7	-0,8	C
11	9,3	-4,8	-1,8	C
12	11,1	-3,0	0,0	C
13	10,6	-3,5	-0,6	C
14	11,6	-2,5	0,5	C
15	10,5	-3,6	-0,6	C

La qualité de la classification

La conformité de cette classification prédictive avec la réalité est illustré par le tableau suivant, appelé matrice des confusions

Appartenance prédite Appartenance réelle

	G	C
G	15	2

En générale cette matrice est un tableau de contingence $g \times g$ (où g est le nombre de groupes), en ligne figurent les appartenances réelles et en colonnes les affectations par le modèle. On peut y repérer le nombre d'affectations correctes et erronées⁵. Le **hit rate** est le pourcentage d'affectations correctes par rapport au nombre total d'individus. Pour que le modèle présente un intérêt, il faut que le hit score soit suffisamment élevé. Dans le cas de deux groupes à effectifs égaux ⁶, une procédure de répartition purement aléatoire entraînerait 50 % d'affectations correctes. La différence entre le hit score et 50 % mesure ainsi la qualité du modèle. Le caractère significatif de cette différence est repéré à l'aide de l'expression:

$$Z = (p - 0,5) / [(0,5)(0,5)/n]^{1/2}$$

où n est le nombre d'individus. Si Z est supérieur à 1,64, le modèle a significativement mieux réussi à classer les individus qu'un processus aléatoire, à un seuil de 95 % ⁷.

Critères du maximum chance et proportional chance

Quand les groupes sont de dimensions différentes le hit rate ne peut plus être comparé au critère du 50% dans ce cas on peut utiliser deux critères: le critère de la probabilité maximum (maximum chance) et le critère de la probabilité proportionnelle (proportional chance). Le critère **maximum chance** considère que tout individu choisit aléatoirement doit être classé comme appartenant au plus grand groupe. Le critère **proportional chance** est donné par la somme des carrés des proportions de chaque groupe par rapport au nombre totale d'individus (dans le cas de deux groupes $C_{pro} = p^2 + (1 - p)^2$).

Si le groupe G (grand prix) avait 20 membres et le groupe C (prix de consolations) avait 80 membres⁸ le critère maximum chance serait 80 et le critère proportional chance $(0,20)^2 + (0,80)^2 = 0,68$. Un hit rate de 85% montrera une bonne amélioration par rapport à la pure chance mais ils sera moins fort par rapport au critère maximum chance.

⁵Vedrine J.-P. "Le traitement des données en marketing", Ed. Organisation, Paris, 1991.

⁶Pour les cas d'effectifs inégaux cf. J. F. Hair, R. E. Anderson et R. L. Tatham *Multivariate data analysis*, Macmillan, 1987, p. 89 et ss.

⁷Il est conseillé de réaliser les calculs permettant d'aboutir au hit score sur un échantillon différent de celui qui a servi à construire la fonction discriminante (hold out method).

⁸ Churchill, 91

Développement mathématique du modèle^a

Matrice de covariances intra-groupes

Soit une matrice \mathbf{X} de scores centrés sur v variables. Celle-ci est partitionnée en g sous-matrices \mathbf{D}_i de n_i cas. À chaque partition correspond un centroïde de $\mathbf{m}_i = (m_{i1}, \dots, m_{iv})$, une matrice de dispersion (sommées des carrés des écarts et des produits des écarts aux moyennes) \mathbf{W}_i ainsi qu'une matrice de covariances $\mathbf{V}_i = \mathbf{W}_i / (n_i - 1)$.

Les matrices de covariances de chaque groupe étant homogènes, il y a intérêt à considérer la matrice de dispersion au sein des groupes $\mathbf{W} = \mathbf{W}_1 + \dots + \mathbf{W}_g$ à laquelle correspond la matrice moyenne de covariances intra groupes $\mathbf{V} = \mathbf{W} / (\sum n_i - g)$.

L'ensemble des centroïdes m_j constitue la matrice $\mathbf{M} =$

Soit $= 0$.

Matrice de covariances inter-groupes

La matrice correspondante des covariances est dite d'intergroupes et s'obtient ainsi:

$$\mathbf{B} = \mathbf{M}'\mathbf{M} / (g - 1)$$

La fonction discriminante y recherchée est celle qu'on obtient au moyen d'une combinaison linéaire \mathbf{k} de la matrice \mathbf{X} de telle sorte qu'à une variance intra groupes donnée corresponde un maximum de la variance inter groupes. On peut représenter ainsi la dispersion des cas et moyennes des trois groupes sur l'axe y de la fonction discriminante.

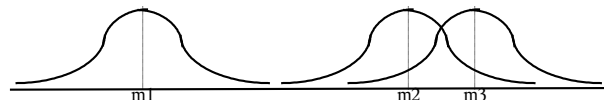


Figure 9 - Dispersion de trois groupes sur l'axe de la fonction discriminante.

La fonction discriminante expression générale

La fonction discriminante ayant pour expression générale $y = \mathbf{X}\mathbf{k}$, les projections sur y des centroïdes, c'est-à-dire des moyennes des groupes, seront $\mathbf{M}\mathbf{k}$ et la variance de ces moyennes ou variance inter sera $\mathbf{k}'\mathbf{M}'\mathbf{M}\mathbf{k} / (g - 1) = \mathbf{k}'\mathbf{B}\mathbf{k}$. De la même façon, la variance intra groupes sera $\mathbf{k}'\mathbf{V}\mathbf{k}$.

On désire considérer simultanément une dispersion maximum des moyennes pour une dispersion donnée des cas au sein des groupes. On choisit d'arrêter la condition d'une variance intra unité, c'est-à-dire $\mathbf{k}'\mathbf{V}\mathbf{k} =$

^aHubert Laforge "Analyse multivariée pour les sciences sociales et biologiques avec applications des logiciels BMP, BMDP, SPSS, SAS", Ed. Etudes Vivantes, Montréal, 1981.

1, qui s'ajoute au multiplicateur indéterminé de Lagrange pour former l'équation

$$F = \mathbf{k}'\mathbf{B}\mathbf{k} - \lambda(\mathbf{k}'\mathbf{V}\mathbf{k} - 1)$$

Critères de maximisation

On calcule alors les valeurs du vecteur \mathbf{k} qui maximisent F :

$$\begin{aligned} &= 2\mathbf{B}\mathbf{k} - 2\lambda\mathbf{V}\mathbf{k} = 0 \\ &= \mathbf{B}\mathbf{k} - \lambda\mathbf{V}\mathbf{k} = 0 \\ &= (\mathbf{B} - \lambda\mathbf{V})\mathbf{k} = 0 \end{aligned}$$

On sait que celle équation comporte une solution différente de zéro pour $|\mathbf{B} - \lambda\mathbf{V}| = 0$.

Ces équations, solutions du problème de la fonction discriminante présentent donc une grande ressemblance avec celles du problème de l'analyse factorielle. On leur donne la forme habituelle en pré multipliant par l'inverse de \mathbf{V} :

$$\begin{aligned} &(\mathbf{V}^{-1}\mathbf{B} - \lambda\mathbf{I})\mathbf{k} = 0 \\ &\text{et} \\ &|\mathbf{V}^{-1}\mathbf{B} - \lambda\mathbf{I}| = 0 \end{aligned}$$

C'est le problème de la recherche des vecteurs latents \mathbf{k} d'une matrice carrée non symétrique $\mathbf{V}^{-1}\mathbf{B}$. Il y a intérêt à faire en sorte de revenir à une matrice symétrique. On a vu (théorème 7 ...) qu'une matrice non symétrique \mathbf{M} peut être écrite comme la somme d'une matrice symétrique \mathbf{S} et d'une matrice non symétrique \mathbf{A} : $\mathbf{M} = \mathbf{S} + \mathbf{A}$. Par multiplications et additions appropriées de lignes, on rend diagonale la matrice symétrique et on remplace $\mathbf{V}^{-1}\mathbf{B}$ par $\mathbf{D}_V^{-1/2}\mathbf{B}\mathbf{D}_V^{-1/2}$ qui a la propriété d'être symétrique et d'avoir les mêmes racines et vecteurs latents que $\mathbf{V}^{-1}\mathbf{B}$.

L'expression de la solution de la fonction discriminante devient donc:

$$\begin{aligned} &(\mathbf{D}_V^{-1/2}\mathbf{B}\mathbf{D}_V^{-1/2} - \lambda\mathbf{I})\mathbf{k} = 0 \\ &|\mathbf{D}_V^{-1/2}\mathbf{B}\mathbf{D}_V^{-1/2} - \lambda\mathbf{I}| = 0 \end{aligned}$$

La résolution de l'équation

$$|\mathbf{D}_V^{-1/2}\mathbf{B}\mathbf{D}_V^{-1/2} - \lambda\mathbf{I}| = 0$$

comporte un nombre de solutions égal au minimum suivant: soit le nombre de groupes moins un, soit le nombre de variables. Chacune de ces valeurs est introduite dans l'équation

$$(\mathbf{D}_V^{-1/2}\mathbf{B}\mathbf{D}_V^{-1/2} - \lambda_i\mathbf{I})\mathbf{k}_i = 0$$

entraînant un nombre correspondant de solutions \mathbf{k}_i . On a vu antérieurement que les vecteurs \mathbf{k}_i sont orthogonaux, c'est-à-dire indépendants: la matrice \mathbf{K} des vecteurs \mathbf{k}_i répond donc à la relation $\mathbf{K}'\mathbf{K} = \mathbf{I}$ lorsque normée aux racines latentes.

Le calcul se fait à partir de la valeur des matrices \mathbf{B} et \mathbf{V} des covariances inter et intra. Si on procéderait à partir des matrices de dispersion, la solution serait la même sauf pour les racines latentes qui s'en trouveraient multipliées par le rapport

$$(g - 1) / (\sum n_i - g).$$

Plusieurs auteurs présentent la fonction discriminante comme étant définie par le vecteur k tel que le rapport de la variance inter à la variance intra soit maximum:

C'est une approche équivalente à celle décrite plus haut où $k'Vk$ étant fixé à 1, on recherche les valeurs de k qui maximisent le numérateur.

Rotation des axes des fonctions discriminantes

Interpretation

Une fonction discriminante est définie par un vecteur colonnes de coefficients appliqués, par une combinaison linéaire, aux variables étudiées; ces coefficients sont dits bruts ou standard (beta weights) selon qu'ils s'appliquent à des variables brutes ou standard. Le but de telles combinaisons linéaires est de séparer au maximum les groupes les uns des autres; en plus d'être indépendantes les unes des autres, les fonctions discriminantes ont un pouvoir de discrimination qui décroît d'une fonction à l'autre. Le nombre de fonctions discriminantes est le plus petit des deux possibilités suivantes: nombre de variables ou nombre de groupes moins un.

Méthodes de rotation

On a vu qu'à beaucoup de points de vue, les fonctions discriminantes présentent une grande analogie avec les facteurs mis en évidence par l'une ou l'autre technique de l'analyse factorielle. En particulier, on peut souhaiter identifier par un nom chacune de celles-ci, en se basant sur les contributions des variables, telles qu'exprimées de façon comparable par les coefficients standard. Comme en analyse factorielle, une telle identification n'est pas toujours facile, à moins de procéder à une rotation des axes des fonctions tout en maintenant constantes les positions relatives des cas et des moyennes ou centroïdes. Une rotation VARIMAX est proposée en option dans le programme SPSS; on obtient ainsi des coefficients qui sont le plus possible voisins de 1 pour les uns, et de zéro pour les autres. L'avantage qu'on en tire est celui d'une facilité plus grande d'interprétation, mais cependant on y perd la connaissance de l'ordre des fonctions quant à leur pouvoir de discrimination. C'est pourquoi il est suggéré de n'utiliser qu'avec prudence la rotation des axes des fonctions discriminantes (voir Klecka, dans SPSS, p. 444-445).

Méthodes de sélection des variables

Plusieurs méthodes peuvent être utilisées dans le choix des variables à inclure dans l'édification des fonctions discriminantes. Celle dont il a été question jusqu'à maintenant, et qui consiste à considérer toutes les variables à la fois, est dite méthode directe.

On peut aussi faire appel à une approche hiérarchique (stepwise) où les variables sont introduites une à une selon leur capacité décroissante à mettre en évidence la différence entre les groupes. Au cours des sélections successives, il est possible que des variables déjà entrées perdent leur pouvoir de discrimination: la raison en est une redondance d'information, c'est-à-dire que le pouvoir de discrimination de cette variable est désormais inclus dans quelque combinaison de nouvelles variables retenues.

Donc à chaque étape de l'analyse, on procède à l'élimination des variables devenues inutiles. Dans le programme SPSS ce test de variable apparaît sous le titre F-TO-REMOVE.

Divers critères, mettant l'accent sur l'un ou l'autre aspect de la dispersion des groupes, sont utilisés pour la sélection de variables: a) le test de Wilks vise à minimiser un rapport où entrent en considération la dispersion des centroïdes et la cohésion des cas au sein des groupes: il est semblable à un test multivarié F sur les différences entre les centroïdes;

b) plusieurs tests, reliés à la notation de distance de Mahalanobis, visent à maximiser l'écart entre les deux groupes les plus rapprochés (les méthodes MAHAL, MAXMINF, MINRESID du programme SPSS sont des variantes de cette approche); c) la méthode de Rao consiste à choisir la variable qui contribue le plus à une distance généralisée, évaluée sur les variables précédentes.

Pour tous ces critères, une variable est sélectionnée lorsque son rapport F partiel dépasse une valeur critique, c'est-à-dire lorsque sa contribution à la dispersion additionnelle des centroïdes est statistiquement significative: dans le programme SPSS ce F est dit F-TO-ENTER

Test de signification

On peut poursuivre, jusqu'à exhaustion, l'extraction des fonctions discriminantes. Mais comme dans le cas des composantes principales, l'intérêt des fonctions additionnelles va décroissant. Dans beaucoup d'applications on ne dépasse pas deux ou trois fonctions afin de tirer parti de la facilité et de l'intérêt d'une illustration de la position des groupes de sujets dans un espace à trois dimensions et moins.

L'effet de discrimination de la fonction i par rapport à toutes les fonctions est exprimé par la proportion (Hope, p. 117-120; Cooley et Lohnes, p.248-250)

Ce rapport exprime la proportion de la variance expliquée par chaque fonction discriminante. Cependant cette proportion ne conduit pas à une décision statistique au sens habituel du terme. On recourt souvent à un autre indicateur.

On montre que:

$\Lambda =$, où p = nombre de fonctions discriminantes,

peut être utilisé pour exprimer la capacité de discrimination d'un ensemble de variables (ce paramètre est similaire à Λ = de l'article 11.3.0. de Laforge). De même pour les fonctions au-delà de la k-ième fonction: $\Lambda' =$

Ce lambda (λ') est donc une mesure de l'inverse de la puissance discriminative expliquée par les fonctions discriminantes à venir. La signification de la discrimination des fonctions restantes k à p, à la suite de l'acceptation des k premières, peut se calculer au moyen de l'approximation de Bartlett:

$$\chi^2 = - [N - (v + g) - 1] \ln \lambda' \quad \text{avec} \quad \text{d.l.} = (v - k)(g - k - 1)$$

où v : nombre de variables
 g : nombre de groupes

et $\Lambda' =$

Si pour ces fonctions discriminantes (k + 1) à p, on obtient une valeur de χ^2 qui ne dépasse pas le seuil critique, on considère que les k premières fonctions calculées suffisent seules à expliquer de façon significative les écarts entre les groupes.

Une autre méthode permettant de juger de la valeur de discrimination de chaque fonction est celle de la corrélation canonique qui mesure l'étroitesse de la relation entre une fonction discriminante et l'appartenance à l'un ou l'autre groupe. Le carré de cette corrélation canonique exprime la proportion de la variance de la fonction discriminante expliquée par une fonction d'appartenance aux divers groupes. Les programmes universels d'analyse discriminante font appel, en plus d'autres méthodes, au calcul d'une telle corrélation canonique.

Remarques et résumé

Discussion

L'analyse discriminante peut être vue comme un cas spécial d'analyse factorielle. Mais le but diffère: il s'agit de faire ressortir au maximum les différences entre des groupes mesurés dans un espace multidimensionnel, en projetant chaque cas dans l'espace unidimensionnel d'un petit nombre de fonctions linéaires orthogonales.

Cette opération fait suite habituellement à celle de l'analyse de variance multivariée où, en présence d'une situation où plusieurs groupes sont mesurés sur plusieurs variables, on s'intéresse d'abord à déterminer s'il y a une différence significative entre les groupes. Dans le cas de résultats positifs, il devient intéressant de déterminer, parmi les variables, celles qui sont responsables dans un ordre décroissant d'importance des différences entre les groupes: c'est le but de l'analyse discriminante. Une exploitation plus poussée des

résultats conduit à leur utilisation dans le but de classifier (en se donnant comme objectif une probabilité minimum d'erreurs) des nouveaux sujets dans les divers groupes: l'étude de cette technique fait l'objet du chapitre suivant.

Le rôle de l'analyse discriminante peut être envisagé de deux façons quant à l'attribution des qualificatifs d'indépendance et de dépendance, aux variables mesurées sur les populations visées et aux fonctions discriminantes. En sciences d'exploration, en général, les populations sont considérées comme variables indépendantes (predictors) et les fonctions discriminantes comme variables dépendantes (critères). En sciences expérimentales, ces rôles se trouvent renversés.

L'analyse discriminante consiste donc à projeter dans un sous-espace appropriée des échantillons de mesures multidimensionnelles. L'interprétation de cette opération peut être faite en termes (voir Cooley et Lohnes, p. 243) soit du nombre et de l'importance relative des fonctions discriminantes retenues, soit de la localisation dans l'espace discriminant des populations étudiées.