

FOVIEA

STAT'

TRAITEMENT
STATISTIQUE
DES ÉTUDES
CLINIQUES

ANALYSE DES DONNÉES / II
MÉTHODES EXPLICATIVES

L'analyse statistique classique ne permet d'étudier simultanément qu'un nombre restreint de caractères. Or, dans la pratique, les individus observés sont fréquemment décrits par un grand nombre de caractères, ce qui limite l'utilisation de ce type d'analyse.

L'analyse des données (ou analyse multivariée) permet, au contraire, d'étudier simultanément un grand nombre de caractères en prenant en compte les liaisons pouvant exister entre eux.

Elle s'articule autour de 2 grands types de méthodes : les méthodes descriptives et les méthodes explicatives.

Spécialisée dans le traitement statistique des études cliniques, FOVEA STAT' vous propose au travers de ce fascicule quelques notions simples pour vous familiariser avec les MÉTHODES EXPLICATIVES de l'analyse des données.

La méthode explicative à utiliser diffère selon le **type des caractères étudiés** :

CARACTÈRE À EXPLIQUER \ CARACTÈRES EXPLICATIFS	QUALITATIFS	QUANTITATIFS
QUALITATIFS	ANALYSE DES MESURES CONJOINTES	ANALYSE DE VARIANCE
QUANTITATIFS	ANALYSE DISCRIMINANTE	RÉGRESSION MULTIPLE

1 Analyse des mesures conjointes

L'analyse des mesures conjointes a pour objectif de mesurer l'**effet simultané** (ou conjoint) de **plusieurs caractères qualitatifs explicatifs** x_i sur les différentes modalités que peut prendre le **caractère qualitatif à expliquer** y .

Lorsque les caractères explicatifs sont indépendants les uns des autres, cette méthode permet d'établir un **modèle additif** (sans interaction) ; dans le cas contraire, il s'agit d'un **modèle configural** (avec interaction).

Cette méthode est principalement utilisée dans le **domaine du marketing**.

2 Analyse de variance

L'analyse de variance multidimensionnelle ne diffère pas dans son principe de l'analyse de variance bidimensionnelle utilisée, par exemple, pour comparer les valeurs moyennes de PAS obtenues avec 3 traitements différents.

Analyse discriminante

L'analyse discriminante comporte **2 applications** :
 une application exclusivement **explicative** : **analyse factorielle discriminante (A.F.D.)**,
 une application également **prédictive** : **analyse discriminante décisionnelle**.

ANALYSE FACTORIELLE DISCRIMINANTE

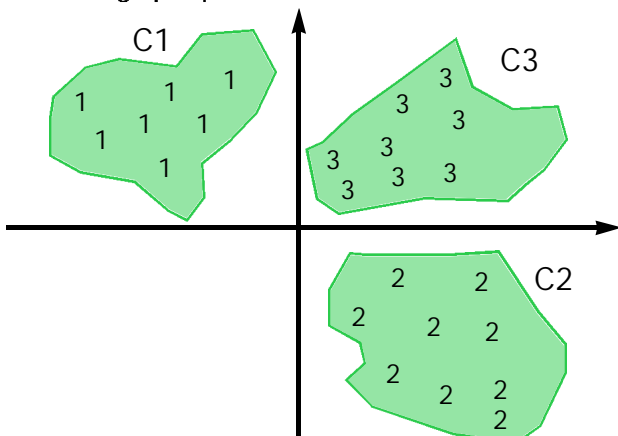
Cette méthode permet, à l'aide d'une représentation graphique appropriée, de mettre en évidence les liens existant entre le caractère à expliquer et les caractères explicatifs. Elle est **une variante de l'analyse en composantes principales (A.C.P.)**.

Elle a pour objectif d'identifier **les combinaisons linéaires (ou facteurs discriminants)** des caractères explicatifs x_i qui expliquent le mieux le classement des individus selon les **k modalités** du caractère à expliquer y .

Chaque combinaison linéaire $C = a_k x_k + \dots + a_n x_n + \dots + a_p x_p$ est un facteur d'autant plus discriminant qu'elle prend des valeurs très proches pour des individus ayant la même modalité de y , et des valeurs très éloignées pour des individus ayant des modalités différentes.

Une représentation graphique appropriée permet de **séparer le plus nettement possible les k classes du caractère y**.

Soit une variable y comportant 3 modalités (1, 2 et 3), l'A.F.D. permet d'identifier les facteurs discriminants C_1 , C_2 et C_3 , combinaisons linéaires des caractères explicatifs initiaux. On obtient la représentation graphique suivante :



ANALYSE DISCRIMINANTE DÉCISIONNELLE

Elle permet de **prévoir les modalités du caractère à expliquer à partir des valeurs prises par les caractères explicatifs**.

Ainsi, pour un individu donné, lorsque l'on connaît les valeurs des caractères explicatifs, on peut affecter cet individu, avec un risque d'erreur minimale, dans l'un des k groupes (k classes) déterminés par le caractère à expliquer.

Cette méthode a de **nombreuses applications en recherche clinique**. Elle peut notamment permettre de prévoir le succès ou l'échec d'une thérapeutique en fonction de caractères explicatifs quantitatifs (*âge, poids, NFS, ...*).

Régression multiple

La régression multiple est la **généralisation de l'ajustement linéaire simple** (*ajustement linéaire multiple*).

Elle s'exprime par **une équation du type** :

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n + b$$

y est le **caractère dépendant** ou **à expliquer**.

x_1, x_2, \dots, x_n sont les **caractères explicatifs**.

a_1, a_2, \dots, a_n sont les **paramètres** de l'équation. Ils sont estimés à partir des données au moyen de la **méthode des moindres carrés**.

b est **une constante**.

Cette méthode n'est applicable que dans le cas où les variables explicatives prises 2 à 2 sont indépendantes entre elles (*coefficient de corrélation proche de 0*) ; dans le cas contraire, il faut introduire dans l'équation une seule des 2 variables corrélées.

		TYPE DE CARACTERES { ÉTUDIÉS (ANALYSE DESCRIPTIVE) À EXPLIQUER (ANALYSE EXPLICATIVE)		
		NOMINAUX	ORDINAUX	QUANTITATIFS
ANALYSE DESCRIPTIVE	MÉTHODES GRAPHIQUES	ANALYSE FACTORIELLE DES CORRESPONDANCES (A.F.C.)	ANALYSE DES PROXIMITÉS ET DES PRÉFÉRENCES	ANALYSE EN COMPOSANTES PRINCIPALES (A.C.P.)
	MÉTHODES DE CLASSIFICATION	<ul style="list-style-type: none"> • MÉTHODES HIÉRARCHIQUES (<i>effectif faible</i>) <ul style="list-style-type: none"> → MÉTHODES ASCENDANTES → MÉTHODES DESCENDANTES • MÉTHODES NON HIÉRARCHIQUES (<i>effectif important</i>) 		
ANALYSE EXPLICATIVE	VARIABLES EXPLICATIVES QUALITATIVES	ANALYSE DES MESURES CONJOINTES		ANALYSE DE VARIANCE
	VARIABLES EXPLICATIVES QUANTITATIVES	ANALYSE DISCRIMINANTE		RÉGRESSION MULTIPLE

Récapitulatif des méthodes d'analyse des données

FOVIA STAT