

L'ANALYSE DISCRIMINANTE

OUTIL DE CONTROLE QUALITE EN FABRICATION

Thierry CEMBRZYNSKI

RENAULT SA

Direction de la Validation des Processus Industriels

Exploitation Statistique des Données Service 60225

Quai de Stalingrad

Boulogne Billancourt

Dans le cadre du contrôle en fabrication et plus particulièrement de la maîtrise statistique des processus, nous avons construit une méthode fondée sur l'analyse discriminante permettant, au fur et à mesure de l'obtention des mesures processus, la détection des dérives de performance Qualité en trois niveaux :

Bon - Moyen - Mauvais

Cet algorithme qui requiert un traitement informatique pour les calculs, n'est pas une « recette de cuisine », car il a été pensé pour la maîtrise de la variabilité des processus et construit de manière à fournir en temps quasi réel des résultats graphiques très simples à interpréter et parfaitement compatibles avec la formation des opérateurs en atelier.

Notre objectif est de mettre à la disposition du plus grand nombre « l'outil supplémentaire » de la Qualité Totale qui doit permettre d'atteindre le « zéro défaut statistique » ; à cette fin cet article n'est volontairement pas théorique et s'articule autour d'un exemple détaillé dont le lecteur trouvera en annexe tous les programmes SAS (® SAS Institute Inc, Cary USA) ainsi que les données.

Introduction

Une approche traditionnelle de la fabrication consiste à compter sur la production pour fabriquer le produit et sur le contrôle qualité pour inspecter le produit et repérer les articles ne satisfaisant pas aux spécifications.

Nous sommes alors en présence d'une *stratégie de correction*. C'est un gaspillage car on perd du temps et des matières premières pour des produits qui ne sont pas toujours utilisables ; de plus, l'inspection après l'événement est peu économique car elle est coûteuse, peu fiable, et le gaspillage a déjà été produit. Il est beaucoup plus judicieux d'éviter le gaspillage en ne produisant pas de produits inutilisables. Il faut alors mettre en place une *stratégie de prévention*. C'est l'objet de la Maîtrise Statistique des Processus (MSP ou SPC chez les Anglo-Saxons) où l'on cherche à améliorer la qualité des produits en agissant sur les facteurs de variabilité des processus de production (manuels et ou automatisés), par la mise en œuvre de méthodes avancées d'analyses statistiques comme *l'analyse discriminante*.

Une remarque s'impose alors. Dans cet article, *nous avons choisi de laisser de côté le contrôle classique portant sur une seule variable* (monodimensionnel) car les méthodes sont bien connues et relativement simples à mettre en œuvre (carte de contrôle,...) *pour nous placer dans le cas plus complexe mais fréquent du contrôle concernant de nombreuses variables* (multidimensionnel) *qui peuvent être corrélées entre elles, et, éventuellement de nature différente (qualitatives et quantitatives)*.

Renault possède déjà un acquis ; en collaboration avec la direction de l'Ingénierie Véhicule et la direction des Fabrications nous avons développé et mis en exploitation en peinture à l'usine de Sandouville; un logiciel temps réel d'alerte et de diagnostic sur la chaîne de la Safrane. Un autre logiciel développé par P. Fogel est quant à lui en exploitation dans les principales usines de montage (Sandouville, Douai, Flins) pour le suivi de l'emboutissage.

Nous suivrons le plan suivant dans cet article:

Dans une première partie, après quelques définitions de base sur le contrôle du processus, nous ferons un rappel succinct sur l'analyse discriminante, puis nous décrirons l'objectif, les enjeux, et sa mise en œuvre en MSP, pour terminer par son paramétrage et sa validation.

Dans une seconde partie, le lecteur trouvera une illustration de la méthode sur un exemple dont nous fournirons les données et les programmes.

1. Contrôle du Processus et Analyse Discriminante

1.1. Le contrôle du processus : définition de base.

Le contrôle du processus peut se présenter comme *un système d'information avec retour* dont les principaux éléments sont les suivants:

- *Le processus* : par processus, nous entendons la combinaison complète dite des 5M dont le travail en commun a pour résultat le produit : la *main d'œuvre*, les *machines*, les *matières*, le *milieu* et la *méthode* ; on y rajoute parfois la *mesure*.
- *Renseignement sur la performance* : on peut apprendre beaucoup sur la performance réelle du processus en étudiant la qualité du produit, autrement dit; ses défauts, ceux-ci sont généralement réunis sur un support qu'on appelle *la carte de contrôle*. Si elle est interprétée correctement, elle permet de montrer, s'il est nécessaire d'intervenir pour corriger le processus, quand on constate une dégradation de la qualité du produit, c'est-à-dire quand on *détecte une rupture de la performance qualité du processus*.
- *Intervention sur le processus* : elle est orientée sur le futur, car ce sont des mesures prises (selon la nécessité) pour empêcher la dégradation totale ou partielle du processus. Ces mesures peuvent consister en des changements sur les moyens, dans les matières... Bien évidemment les effets des actions sur le processus menées par les fabricants doivent aussi être contrôlés pour vérifier si elles ont permis de retrouver un régime normal.

1.2. L'Analyse Discriminante

On désigne sous le nom d'analyse discriminante une famille de techniques destinées à décrire et à *classer* (affecter à des classes pré-existantes) des individus caractérisés par un nombre important de variables numériques.

L'origine de cette méthode remonte aux travaux de Fisher (1936) et de Mahalanobis (1936).

L'analyse factorielle discriminante est une méthode à la fois descriptive et prédictive dont les exemples les plus classiques appartiennent sans doute au domaine médical.

Certains diagnostics ou certaines interventions ayant été réalisés sur un ensemble de patients caractérisés par une série d'analyses et d'examens, on se propose de faire un diagnostic ou de décider une intervention sur un nouveau patient ayant subi les mêmes analyses et examens. L'analyse discriminante tentera à partir de ces dernières informations et de leur réseau d'inter-relations de prévoir le diagnostic le plus probable ou l'intervention la plus favorable.

Remarque : La méthode utilisée ici est une *Analyse Canonique Discriminante* et non une analyse factorielle discriminante classique; cela génère de fait des différences dont il convient de préciser l'origine. Sans entrer dans le détail, il s'agit ainsi d'un cas particulier de l'analyse canonique entre un groupe de variables quantitatives (ici les variables processus) et une variable qualitative (ici la variable niveau de qualité).

La méthode proposée s'apparente fortement à une Analyse en Composantes Principales dans la mesure où il s'agit de « résumer » l'inertie inter-classes de la même manière que l'ACP résume l'inertie totale, et que les plans « factoriels » produits s'interprètent en terme de distance euclidienne usuelle.

La méthode proposée est aussi un cas particulier *de l'analyse de la variance multivariée* dans la mesure où l'on discrimine non pas directement les niveaux (bon, moyen, mauvais), mais les différences entre les niveaux (bon - mauvais, moyen - mauvais), étant entendu qu'une contrainte géométrique existe entre les classes ; en effet, si $X_{i,k}$ désigne l'indicatrice d'appartenance de l'individu i à l'une des classes k

(k =bon, moyen, mauvais), alors $\sum_{k=1}^3 X_{i,k} = 1$. Cela explique alors pourquoi il n'y a

que deux ($k-1$) composantes canoniques discriminantes contre trois (k) pour l'analyse factorielle discriminante classique de Fisher.

D'un point de vue plus « mathématique », on dispose de N individus ou observations décrites par un ensemble de p variables (x_1, x_2, \dots, x_p) réparties en q classes définies a priori par la variable Y nominale à q modalités

L'analyse discriminante se propose dans un premier temps de séparer au mieux les q classes à l'aide des p variables explicatives. Dans un deuxième temps, elle cherche à résoudre le problème de l'affectation d'individus nouveaux, caractérisés par les mêmes p variables, à certaines classes $(1, \dots, q)$ déjà identifiées sur l'échantillon des N individus (appelé échantillon d'apprentissage).

On distingue donc deux démarches successives, l'une d'ordre descriptif, l'autre d'ordre décisionnel :

- Chercher les fonctions linéaires discriminantes sur l'échantillon d'apprentissage de taille N qui sont les combinaisons linéaires des variables explicatives (x_1, x_2, \dots, x_p) , dont les valeurs séparent au mieux les q classes.
- Connaître la classe d'affectation de n nouveaux individus également décrits par les p variables explicatives explicatives (x_1, x_2, \dots, x_p) , mais dont on ignore la classe

d'appartenance. Il s'agit ici d'un problème de classement dans des classes préexistantes.

L'application de l'analyse discriminante au contrôle qualité est facile à imaginer, c'est *de prévoir la qualité d'un produit* et donc de prédire la performance du processus à partir de mesures externes. Elle est alors particulièrement appréciable pour les processus de fabrication où le retour d'information du contrôle qualité est tardif (longue manipulation, stockage intermédiaire...) ou à faible échantillonnage (contrôle destructif).

Les variables explicatives (x_1, x_2, \dots, x_p) explicatives seront des paramètres du processus pris dans les 5M et mesurés, la variable à expliquer y , un niveau de performance qualité (Bon, Moyen, Mauvais).

L'analyse discriminante est alors un outil de la MSP qui correspond à une évolution du contrôle du mode curatif au mode préventif. Dès la sortie de fabrication on estime le niveau de qualité du produit $Y_e = F(X)$ en fonction des mesures externes prises sur le processus $X (x_1, x_2, \dots, x_p)$; le contrôle qualité final Y permet alors de valider les prévisions Y_e par comparaison (comptage des erreurs d'affectation du modèle discriminant) et de le recalculer régulièrement en fonction des nouvelles données qui alimentent alors l'échantillon d'apprentissage, permettant ainsi d'affiner le modèle.

La *mise en œuvre* doit être « simple et amusante » pour être utilisable par le plus grand nombre, comme l'a demandé M. Jouslin de Noray vice-président du MFQ à la Société des Ingénieurs de l'Automobile en novembre 1995 ; en particulier il faut que l'interprétation des résultats soit élémentaire et naturellement compatible avec la formation des opérateurs.

A cette fin, pour le contrôle qualité, il convient d'associer au plan factoriel discriminant, une représentation de l'affectation aux classes Y_e sous une forme familière aux fabricants (une carte de contrôle par exemple) et naturellement la fabrication se doit d'intervenir dès qu'un saut apparaît.

2. Un exemple

Soit un processus de fabrication où sont mesurées sur chaque unité i produite $p=3$ variables processus $(x_1, x_2, x_3)_i$ (pression, température, hygrométrie) et sur lequel on cumule au contrôle qualité final, le nombre de défauts $y_i = \text{def } i$.

La variable y est alors qualitative discrète et suit une loi de Poisson [M. Chambon 82].

Très généralement en MSP, les variables, pression, température et hygrométrie sont suivies par cartes de contrôle (aux mesures), la variable y par une carte de contrôle aux attributs de type C (on supposera sans restriction que les unités i sont de même taille N).

L'atelier a collecté un tableau de données (X + Y) comportant 120 unités d'apprentissage et un tableau (X' + Y') comportant 22 unités supplémentaires pour comparer les estimations calculées Y'e à la réalité Y'.

Pour mettre en œuvre l'Analyse Discriminante nous allons :

- lire les 120 unités d'apprentissage X
- coder la variable de défaut Y en trois classes : bon, moyen, mauvais ; cette phase est cruciale, et conditionne naturellement les résultats finaux du modèle
- représenter la variable Y codée dans l'espace(x_1, x_2, x_3)
- calculer les fonctions discriminantes F et représenter dans le plan factoriel discriminant les 120 unités
- lire les données de l'unité X' et les classer Y'e = F (X')
- représenter les ruptures de performance qualité multidimensionnelles, la carte de contrôle de prédiction de la qualité du produit et comparer ses prévisions Y'e à la réalité Y' (connue pour la validation).

2.1. Lecture des données d'apprentissage (Annexe I).

Le fichier des données MODULAD est composé de 120 unités comportant le numéro de l'unité, les p=3 variables explicatives press, temps, hydr et la variable à expliquer y=def. On trouve 3 enregistrements par ligne dans le fichier de données.

Le programme LITDON.SAS lit les données MODULAD et crée un tableau SAS permanent : BASE.DONNEES.

2.2. Codage de la variable de défaut Y en trois classes (Annexes 2,3).

Le programme carte_c_def.sas effectue une analyse de la distribution de la variable y=def (PROC UNIVARIATE) qui souligne une distribution bimodale de $\bar{C} = 9.13$. On peut alors construire la carte de contrôle aux attributs de type C de la variable def dont les limites de contrôle supérieure (LSC) et inférieure (LIC) sont alors :

$$LSC = \bar{C} + 3\sqrt{\bar{C}} = 18,19$$

$$LIC = \bar{C} - 3\sqrt{\bar{C}} = 0,06.$$

Ces limites de contrôle normalisées (NF X06-031) correspondent à un intervalle de confiance Poissonien (pour une loi de Poisson $\mu=\sigma^2$ [CHAMBON 82]) à 3 écarts-types (99.7% de confiance) autour de la moyenne ($\mu=\bar{C}$).

Toute unité y_i au delà de la limite supérieure donc représentant au moins 19 défauts est « hors contrôle », en se limitant à cette règle d'interprétation de la carte de contrôle, on remarque une période anormalement défavorable dans l'intervalle (90 - 104), mais aucune « anormalement favorable » (en dessous de LIC).

Si l'on souhaite un codage en trois niveaux de performance (bon, moyen, mauvais), les limites de contrôle LIC, LSC se révèlent trop sévères (probabilité $\alpha = 0.3 \%$ de les franchir), les effectifs des classes extrêmes « bonnes » et « mauvaises » seraient trop faibles pour être représentatifs. C'est pourquoi *il est indispensable que l'utilisateur, à l'examen de la carte de contrôle aux attributs du nombre de défauts (variable def), définisse lui-même ses trois niveaux*; ce qui revient d'un point de vue opérationnel à définir une zone technologiquement inacceptable (niveau mauvais) et une zone d'amélioration cible (niveau bon).

Le programme codage_def.sas effectue le codage désiré pour cet exemple.

Les niveaux fixés par l'utilisateur à partir de l'étude de distribution de la variable déf (PROC UNIVARIATE du programme carte_c_def-sas) sont dans cet exemple :

Bon	déf. ≤ 7 défauts (Q50 = 7 = médiane)
Moyen :	$7 < \text{déf.} < 18$ défauts (P90 = 17,5 Percentile à 90 %)
Mauvais :	$18 \leq \text{déf.}$ (c'est presque la LSC dans ce cas).

Cela permet alors de rechercher l'explication des zones favorables et fortement défavorables tout en conservant des effectifs suffisants pour l'analyse discriminante.

2.3. Représentation des classes dans l'espace des mesures processus (Annexe 4).

Le programme visu_classe_3D_mesures.sas effectue une représentation des classes de défaut dans l'espace des mesures physiques (pression, température, hygrométrie); cette représentation est facultative pour la méthode.

Un symbole coloré est associé à chaque classe pour la représentation graphique :

Bon	Etoile verte
Moyen	Drapeau orange
Mauvais	Croix rouge

L'objet de l'analyse canonique discriminante est de séparer au mieux ces classes de défaut.

2.4. Calcul des fonctions discriminantes (*Annexes 5,6*).

Nous ne présenterons pas le principe du calcul, le lecteur se reportera à des ouvrages spécialisés [LEB MOR PIR 1995 p 255].

La procédure SAS CANDISC effectue une analyse discriminante canonique et recherche les q-1 composantes factorielles qui séparent au mieux les q classes.

Le programme discrim_classe.sas effectue l'AD c'est-à-dire calcule les fonctions discriminantes et les stocke dans le tableau SAS permanent : BASE.FONCDISC :

$$\text{CAN1} = 0,1599 \text{ HYGR}^* + 0,8629 \text{ PRES}^* + 0,9527 \text{ TEMP}^*$$

$$\text{CAN2} = 0,1423 \text{ HYGR}^* + 0,1155 \text{ PRES}^* - 0,675 \text{ TEMP}^*$$

où HYGR* désigne la variable HYGR centrée réduite.

Le coefficient de corrélation canonique (cosinus entre le groupe de variables explicatives et la variable à discriminer) est $\rho^2 = 0,58$, sans être réellement mauvais, il reste relativement moyen et sous entend que le modèle peut donner des réponses inexactes (erreur d'affectation), que l'on peut quantifier sur le tableau de contingence : Ye * Y (affecté * réel).

La première valeur propre 1.397 expliquant 99,6% de l'inertie souligne que la décision porte quasiment sur la seule première composante CAN1.

La représentation graphique des 120 unités de l'échantillon d'apprentissage (programme visu_plan_disc.sas) permet d'imaginer la nature des erreurs d'affectation.

Les observations « moyennes » sont situées entre les « bonnes » et les « mauvaises ».

Très visiblement une unité sera bonne si $\text{CAN1}(i) < 0$ et mauvais si $\text{CAN1}(i) > 2$.

Les autres centres de gravité des classes sont repérés sur le plan par BON, MOYEN, MAUVAIS et serviront pour l'affectation des données supplémentaires; il sont stockés par le programme dans le tableau sas BASE.NIVEAUX.

Il est assez rare que la décision d'affectation ne porte que sur la première composante, bien qu'elle extrait dans cet exemple 99,6% de l'information, c'est pourquoi nous avons tout de même bâti la procédure normale d'affectation fondée sur les deux composantes discriminantes, par souci de généralité.

2.5. Lecture et affectation des observations supplémentaires (*Annexe 7*).

Les fonctions discriminantes étant construites à partir de l'échantillon d'apprentissage, il faut maintenant utiliser la modélisation pour la prédiction.

A cette fin le programme affect_classe_sas.sas :

- lit les données supplémentaires à classer X' (modulad.test) dont, en principe on ne dispose pas de la variable y'=def que l'on cherche précisément à prévoir;
- affecte aux classes connues a priori et dont les centres de gravité calculés sur les données d'apprentissage sont stockés dans le tableau sas BASE.NIVEAUX.

Le programme utilise la PROC SCORE pour calculer les valeurs des fonctions discriminantes après centrage et réduction (BASE.FONCDISC contient les moyennes et variance des données d'apprentissage), et l'algorithme des Nuées Dynamiques [DIDAY et Coll 1979] pour effectuer le classement des observations supplémentaires (PROC FASTCLUS) à leur plus proche centre de gravité selon la distance Euclidienne usuelle. La variable CLUSTER contenue dans le tableau SAS BASE.NEWPOINT est la valeur Y'e.

Il peut apparaître surprenant d'utiliser un algorithme de Nuées Dynamiques pour affecter les individus aux classes, mais le plan « factoriel discriminant » produit s'interprète de la même manière que le plan factoriel des individus que produirait une Analyse en Composantes Principales classique; dès lors un individu sera affecté à la classe dont le centre de gravité lui est le plus proche; informatiquement SAS propose dans ses procédures un algorithme programmé (FASTCLUS) dont on peut imposer les noyaux (SEED=BASE.NIVEAUX), il s'agit de l'utiliser une fois (MAXITER=1) pour être le plus efficace possible mais en interdisant le recalcul des noyaux à chaque affectation (REPLACE = NONE).

NB : En appliquant la procédure affect_classe.sas aux données d'apprentissage X on obtient alors Ye que l'on peut comparer à Y (véritable classe) pour étudier les erreurs d'affectation.

Le listing affect_classe.lst issu de la PROC PRINT montre le contenu de BASE.NEWPOINT, les unités 121-127 sont affectées à la classe 1 (bon), ce qui ne semble pas incohérent à l'égard de la variable def, de même pour les unités (128-135) affectées à la classe 3 (mauvais); on a donc une rupture multidimensionnelle de performance qualité entre les unités 127 et 128. Il faut maintenant présenter ces résultats d'une manière transparente pour l'utilisateur.

2.6. Représentation des ruptures de performance Qualité (Annexes 8 et 9)

Le programme carte_multi_discr.sas représente graphiquement les observations supplémentaires en fonction des variables discriminantes CAN1 et CAN2; une annotation SAS permet de représenter les premières unités de chaque niveau de performance ainsi que les ruptures multidimensionnelles (fig: ruptures).

Cette représentation multidimensionnelle présente l'avantage de *pouvoir suivre la trajectoire des unités* dans le plan discriminant ce qui peut être précieux pour mettre en évidence des dérives lentes de la performances du processus, en revanche cette représentation n'est pas très parlante pour des opérateurs en atelier.

Il convient d'associer *la carte de contrôle prédite* de la performance qualité du processus (CLUSTER en fonction de N° d'unité) sur laquelle nous avons superposé la variable $y = \text{def}$ (qui sont normalement inconnus). On remarque que les unités (132-135) auraient dû être affectées à la classe 2 (moyenne) mais qu'aucun de ces points n'a été affecté à la classe 1 (bon). Ce résultat est largement favorable, le modèle ayant pour but de séparer au mieux les classes « bon » et « mauvais ».

Le modèle discriminant est satisfaisant pour le contrôle qualité dans la mesure où il ne commet pas d'erreur d'affectation « grave », c'est-à-dire d'affecter des unités « mauvaises » à la classe « bonne » et vice-versa.

3. Conclusion.

Nous avons proposé un outil pour le contrôle qualité multidimensionnel permettant une analyse d'impact prévisionnel de la variabilité sur la performance qualité. Cet outil est très visuel, mais requiert un traitement informatique et ne saurait donc être appliqué qu'à des processus instrumentés en temps réel, pour lesquels on souhaiterait :

- faire du pilotage industriel,
- faire de la simulation et de l'optimisation linéaire (SIMPLEXE) pour « dérégler » volontairement le processus à partir des prévisions du modèle, afin de chercher des nominaux optimaux.

Bibliographie :

- [CHAMBON 82] $\mu = \sigma^2$ Rapport Interne RENAULT 1982
- [DIDAY et coll 79] Optimisation en classification automatique, INRIA 1979
- [LEB MOR PIR 96] Statistique Exploratoire Multidimensionnelle
Lebart Morineau Piron 1995

litdon.sas Mon Sep 30 13:17:32 1996 1

```

OPTIONS PS=60 LS=72;
GOPTIONS DEVICE=XCOLOR DISPLAY GWAIT=0 CBACK=WHITE CTEXT=BLACK;
***** ROTATE=LANDSCAPE; *****
* DONNEES DU PROCESSUS *****
LIENNAME BASE ' ';
DATA BASE.DONNEES;
  LENGTH DEFAULT = 4;
  INFILE 'modulad' LRECF=132;
  IF (_N_ EQ 1) THEN DO;
    INPUT @1 N 3. (PRES TEMP HYGR) (3 * 7.3) +1 DEF 2. @@;
  END;
  ELSE IF MOD(_N_, 3) EQ 1 AND (_N_ NE 1) THEN DO;
    INPUT / @1 N 3. (PRES TEMP HYGR) (3 * 7.3) +1 DEF 2. @@;
  END;
  ELSE DO;
    INPUT +1 N 3. (PRES TEMP HYGR) (3 * 7.3) +1 DEF 2. @@;
  END;
*****
* IMPRESSION *****
PROC PRINT DATA=BASE.DONNEES(OBS=5);
RUN;

```

litdon.lst Mon Sep 30 13:17:42 1996 1

The SAS System
13:17 Monday, September 30, 1996

OBS	N	PRES	TEMP	HYGR	DEF
1	1	11.9820	22.0000	50.0340	9
2	2	11.9630	22.0450	50.4450	12
3	3	12.0280	21.8920	50.0160	4
4	4	12.0490	22.1860	50.6490	9
5	5	12.0500	21.9780	50.0640	9

modulad Mon Sep 30 13:14:32 1996 1

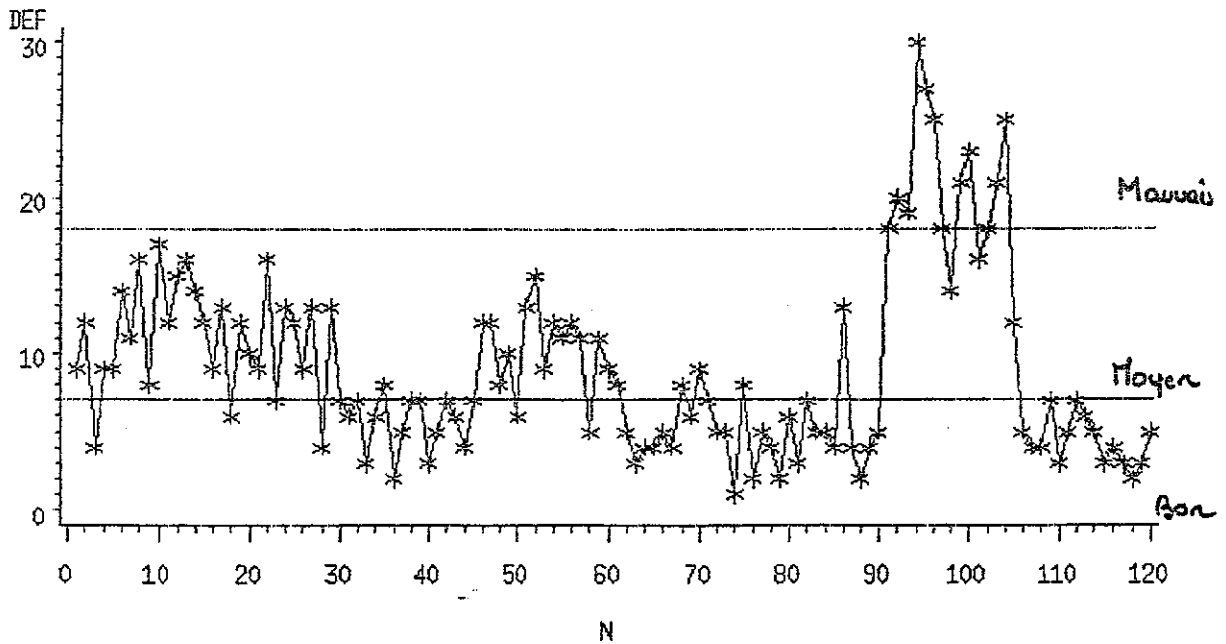
1	11.982	22.000	50.034	9	2	11.963	22.045	50.445	12	5	12.028	21.892	50.016	4
4	12.049	22.186	50.649	9	5	12.050	21.978	50.064	9	6	11.898	21.965	49.730	14
7	12.000	22.017	49.789	11	8	11.945	21.941	49.786	16	9	12.035	22.035	50.327	8
10	11.935	22.064	49.616	17	11	11.997	22.047	50.046	12	12	11.988	22.031	50.432	15
13	12.014	22.080	50.312	16	14	11.947	22.104	50.766	14	15	12.075	22.046	49.969	12
16	11.999	20.002	60.321	9	17	11.932	19.949	60.070	13	18	12.017	19.985	59.559	6
19	12.070	20.023	60.229	16	23	12.036	20.000	59.853	7	24	12.015	20.034	60.018	13
23	12.014	20.043	59.870	12	26	12.012	19.907	59.743	9	27	11.976	19.862	59.613	13
28	12.020	20.035	60.272	4	29	11.958	20.003	59.996	13	30	12.109	20.001	59.862	7
31	9.911	22.006	50.261	6	32	10.043	21.939	50.111	7	33	9.999	22.031	49.826	3
34	10.027	21.991	49.358	6	35	10.090	22.022	50.030	8	36	10.031	22.017	50.164	2
37	10.077	21.937	49.765	5	38	10.106	22.094	50.402	7	39	10.054	21.982	50.430	7
40	9.898	21.995	49.828	3	41	9.924	22.110	50.514	5	42	9.956	22.110	49.916	7
43	10.020	22.035	49.812	6	44	9.970	21.915	49.930	4	45	10.013	22.008	50.320	7
46	10.028	22.058	60.235	12	47	9.978	22.080	59.779	12	48	9.991	21.998	59.999	8
49	10.021	21.943	59.699	10	50	9.967	21.989	59.997	6	51	10.072	22.084	60.119	13
52	9.923	22.008	60.115	15	53	10.046	21.994	59.800	9	54	10.074	21.807	59.810	12
55	10.076	22.011	59.824	11	56	9.972	21.908	59.719	12	57	10.036	22.034	60.274	11
58	9.922	22.066	60.109	5	59	10.113	22.097	60.417	11	60	10.018	22.024	60.255	9
61	10.006	19.991	50.480	8	62	10.006	20.012	49.868	5	63	10.012	20.097	50.455	3
64	9.940	19.948	49.724	4	65	9.971	20.037	50.072	4	66	9.946	19.825	49.509	5
67	9.997	20.052	49.941	4	68	10.031	20.013	49.797	8	69	10.043	19.994	49.535	6
70	9.949	19.939	49.995	9	71	10.067	19.992	50.062	7	72	10.072	20.005	49.961	5
73	9.984	19.979	49.960	5	74	9.975	20.115	50.609	1	75	9.932	20.034	50.167	8
76	12.030	19.983	49.866	2	77	12.013	20.014	50.166	5	78	12.029	19.906	49.706	4
79	11.954	20.118	50.389	2	80	12.016	20.008	50.157	6	81	11.986	19.979	49.642	3
82	12.068	19.954	49.463	7	83	11.980	19.818	49.445	5	84	11.939	20.076	49.610	5
85	11.996	20.037	50.780	4	86	11.901	20.120	50.097	13	87	12.103	20.015	50.013	4
88	12.030	20.010	49.510	2	89	12.036	19.992	50.212	4	90	12.049	20.040	50.350	5
91	11.973	21.973	59.889	18	92	11.991	22.011	59.435	20	93	12.014	22.086	60.436	19
94	11.990	21.995	59.952	30	95	11.993	22.032	59.961	27	96	11.994	22.060	59.858	25
97	12.092	21.920	60.118	18	98	12.039	22.071	60.064	14	99	11.951	22.000	59.726	21
100	11.910	22.031	60.217	23	101	12.047	21.931	59.593	16	102	12.023	21.972	59.910	18
103	11.993	21.928	60.105	21	104	12.043	22.022	59.940	25	105	12.008	22.068	60.562	12
106	9.918	19.943	59.579	5	107	9.999	19.925	59.736	4	108	10.136	19.990	59.724	4
109	9.903	19.912	59.281	7	110	10.025	19.961	60.434	3	111	10.081	20.001	59.956	5
112	9.997	19.892	59.819	7	113	9.929	19.866	59.836	6	114	10.014	19.776	60.018	5
113	10.052	19.955	59.666	3	116	10.005	19.902	59.547	4	117	9.866	19.957	59.988	3
118	9.921	19.979	59.334	2	119	9.951	19.995	60.287	3	120	10.052	20.153	60.507	5

codage_def.sas Mon Sep 30 13:30:40 1996

```

OPTIONS PS=60 LS=72;
GOPTIONS DEVICE=XCOLOR DISPLAY GWAIT=0 CBACK=WHITE CTEXT=BLACK
ROTATE=LANDSCAPE;
LIBNAME BASE '.';
*****
* CODAGE DE LA VARIABLE DE DEFAUL DEF
*****
PROC GPLOT DATA=BASE.DONNEES;
TITLE C=BLACK 'BORNES (MOYEN/ MAUVAIS) APRES VISUALISATION';
PLOT DEF * N / VREF = 7 18 CTEXT=BLACK CAXIS=BLACK CVREF=RED;
SYMBOL I=JOIN V=STAR C=BLACK;
RUN;
DATA BASE.DONNEES(REPLACE=YES);
SET BASE.DONNEES;
IF DEF LE 7 THEN CLASSE = 1;
ELSE IF DEF LT 18 THEN CLASSE = 2;
ELSE CLASSE = 3;
KEEP N PRES TEMP HYGR CLASSE DEF;
RUN;
PROC PRINT DATA=BASE DONNEES(OBS=5);
RUN;
    
```

BORNES (MOYEN/ MAUVAIS) APRES VISUALISATION



codage_def.lst Mon Sep 30 13:31:31 1996

BORNES (MOYEN/ MAUVAIS) APRES VISUALISATION 1
13:31 Monday, September 30, 1996

OBS	N	PRES	TEMP	HYGR	DEF	CLASSE
1	1	11.9820	22.0000	50.0340	9	2
2	2	11.9630	22.0450	50.4450	12	2
3	3	12.0280	21.8920	50.0160	4	1
4	4	12.0490	22.1860	50.6490	9	2
5	5	12.0500	21.9780	50.0640	9	2

visu_classe_3d_mesures.sas

Mon Sep 30 13:41:52 1996

1

```

GOPTIONS DEVICE=XCOLOR DISPLAY GWAIT=0 CBACK=WHITE CTEXT=BLACK
      ROTATE=LANDSCAPE;
LIENAME BASE ' ';
*;
DATA BASE.DONNEES (REPLACE=YES);
  SET BASE.DONNEES (KEEP=N HYGR PRES TEMP DEF CLASSE);
  LENGTH COLORVAL $8. SHAPEVAL $8. DEFAULT = 4;
  IF CLASSE =1 THEN DO;
    SHAPEVAL='STAR';
    COLORVAL='GREEN';
  END;
  IF CLASSE =2 THEN DO;
    SHAPEVAL='FLAG';
    COLORVAL='ORANGE';
  END;
  IF CLASSE =3 THEN DO;
    SHAPEVAL='CROSS';
    COLORVAL='RED';
  END;
END;

```

RUN;

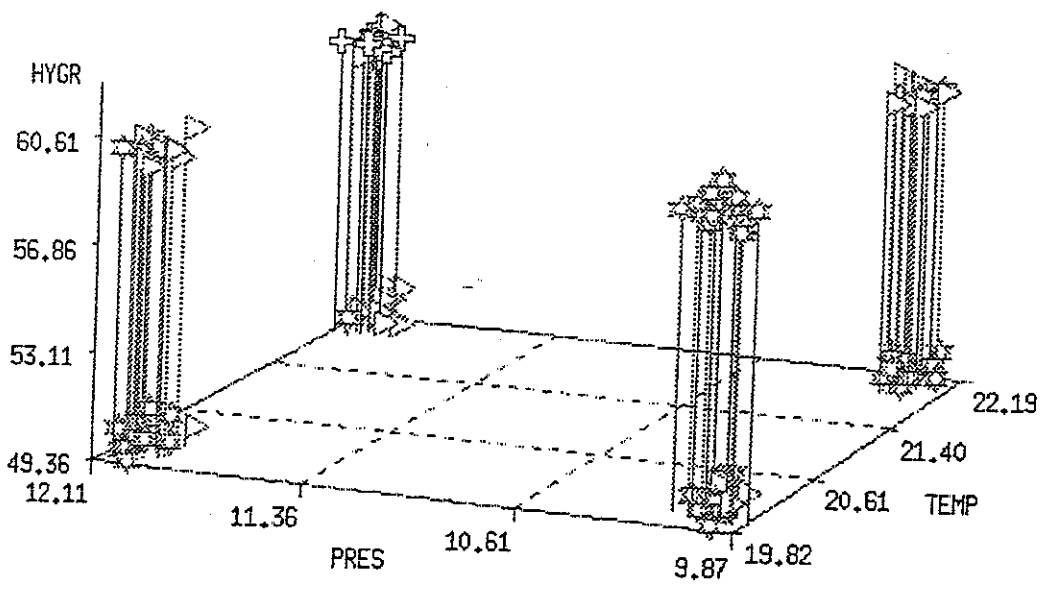
```

*****;
* CODAGE DE LA VARIABLE DE DEFAU DEF
*****;
PROC G3D DATA= BASE.DONNEES;
  TITLE1 'REPRESENTATION DES CLASSES DE DEF';
  LE2 'MESURES PHYSIQUES (HYGR PRES TEMP)';
  SCATTER PRES * TEMP= HYGR / COLOR=COLORVAL SHAPE=SHAPEVAL;
RUN;
*
```

III

REPRESENTATION DES CLASSES DE DEF

MESURES PHYSIQUES (HYGR PRES TEMP)



```

OPTIONS PS=60 LS=72;
GOPTIONS DEVICE=COLOR DISPLAY CHAIP=0 CHACK=WHITE CTEXT=BLACK
          ROTATE=LANDSCAPE;
LIBNAME BASE '/';
*****
* ANALYSE DISCRIMINANTE
*****
PROC CANDISC DATA=BASE.DONNEES(KEEP=N CLASSE DEF HYGR PRES TEMP _CHARACTER_)
  OUTSTAT=BASE.FONCDISC OUT=BASE.DONNEES(REPLACE=YES)
  PREFIX=CAN NCAN=2;
TITLE 'ANALYSE CANONIQUE DISCRIMINANTE DE CLASSE SUR : PRES TEMP HYGR';
CLASS CLASSE;
VAR HYGR PRES TEMP ;
RUN;
* /
PROC PRINT DATA=BASE.DONNEES(OBS=10);
RUN;

```

discrim_classe.lst Mon Sep 30 14:10:36 1996

```

ANALYSE CANONIQUE DISCRIMINANTE DE CLASSE SUR : PRES TEMP HYGR
14:03 Monday, September 30, 1996

Canonical Discriminant Analysis

120 Observations
3 Variables
3 Classes

Class Level Information
CLASS Frequency Weight Proportion
1 61 61.0000 0.508333
2 47 47.0000 0.391667
3 12 12.0000 0.100000

Multivariate Statistics and F Approximations
S=2 M=0 N=56.5
Statistic Value F Num DF Den DF Pr > F
Wilks' Lambda 0.414816462 21.1847 6 230 0.0001
Pillai's Trace 0.588399802 16.1175 6 232 0.0001
Hotelling-Lawley Trace 1.402951251 26.6561 6 228 0.0001
Roy's Greatest Root 1.39740277 54.0329 3 116 0.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

```

```

Canonical Discriminant Analysis

Adjusted Approx Squared
Canonical Canonical Standard Canonical
Correlation Correlation Error Correlation
1 0.763467 0.757900 0.038237 0.582882
2 0.074282 0.018790 0.091164 0.005518

Eigenvalues of INV(E)*H
= CanReq/(1-CanReq)
Eigenvalue Difference Proportion Cumulative
1 1.3974 1.3919 0.9960 0.9960
2 0.0055 0.0040 0.0040 1.0000

Test of H0: The canonical correlations in the
current row and all that follow are zero
Likelihood Ratio Approx F Num DF Den DF Pr > F
1 0.41481646 21.1847 6 230 0.0001
2 0.99448213 0.3218 2 116 0.7255

```

Test of H0: The canonical correlations in the current row and all that follow are zero

Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1 0.41481646	21.1847	6	230	0.0001
2 0.99448213	0.3218	2	116	0.7255

Composants canoniques (var. centrées réduites)

Raw Canonical Coefficients	CAN1	CAN2
HYGR	0.1599895763	0.1423564726
PRES	0.8629883232	0.1155298477
TEMP	0.9527501925	-0.6753745907

discrim_classe.lst Mon Sep 30 14:10:36 1996

```

ANALYSE CANONIQUE DISCRIMINANTE DE CLASSE SUR : PRES TEMP HYGR
14:03 Monday, September 30, 1996

Canonical Discriminant Analysis

Multivariate Statistics and F Approximations
S=2 M=0 N=56.5
Statistic Value F Num DF Den DF Pr > F
Wilks' Lambda 0.414816462 21.1847 6 230 0.0001
Pillai's Trace 0.588399802 16.1175 6 232 0.0001
Hotelling-Lawley Trace 1.402951251 26.6561 6 228 0.0001
Roy's Greatest Root 1.39740277 54.0329 3 116 0.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

```


affect_classe.sas

Mon Sep 30 15:41:41 1996

1

```

OPTIONS PS=60 IS=72;
GOPTIONS DEVICE=XCOLOR DISPLAY GWAIT=0 CBACK=WHITE CTEXT=BLACK
          ROTATE=LANDSCAPE;
LIBNAME BASE ' ';
*****;
* LECTURE DES POINTS A CLASSER
*****;
DATA NEWPOINT;
  LENGTH DEFAUT = 4;
  INFILE 'modulad.test' IRECL=132;
  IF (_N_ EQ 1) THEN DO;
    INPUT @1 N 3. (PRES TEMP HYGR) (3 * 7 3) +1 DEF 2. @@;
  END;
  ELSE IF MOD (_N_, 3) EQ 1 AND (_N_ NE 1) THEN DO;
    INPUT / @1 N 3. (PRES TEMP HYGR) (3 * 7.3) +1 DEF 2. @@;
  END;
  ELSE DO;
    INPUT +1 N 3. (PRES TEMP HYGR) (3 * 7.3) +1 DEF 2. @@;
  END;
RUN;
*****;
* CALCUL DES COORDONNEES FACTORIELLES DANS LE PLAN DISCRIMINANT
*****;
PROC SCORE DATA=NEWPOINT(KEEP=HYGR TEMP PRES DEF N)
  SCORE=BASE.FONCDISC(WHERE=(CLASSE EQ .)) TYPE=SCORE
  OUT=NEWPOINT;
VAR HYGR PRES TEMP;
RUN;
*****;
* CALCUL DES COORDONNEES FACTORIELLES DANS LE PLAN DISCRIMINANT
* ET CLASSIFICATION
*****;
PROC FASTCLUS DATA=NEWPOINT
  SEED=BASE.NIVEAUX(KEEP=X Y RENAME=(X=CAN1 Y=CAN2))
  MAXITER=1 REPLACE=NONE MAXCLUSTERS=3 NOPRINT
  OUT=BASE.NEWPOINT(KEEP=N HYGR PRES TEMP DEF CAN1 CAN2 CLUSTER);
VAR CAN1 CAN2;
RUN;
*****;
* FUSION DES DONNEES
*****;
PROC PRINT DATA=BASE.NEWPOINT;
  ITITLE1 "DONNEES AFFECTEES (ECH:121 -- 142)";
  TITLE2 "EN PRINCIPE LES DEFAUTS DEF NE SONT PAS CONNUS";
RUN;

```

affect_classe.lst

Mon Sep 30 15:41:44 1996

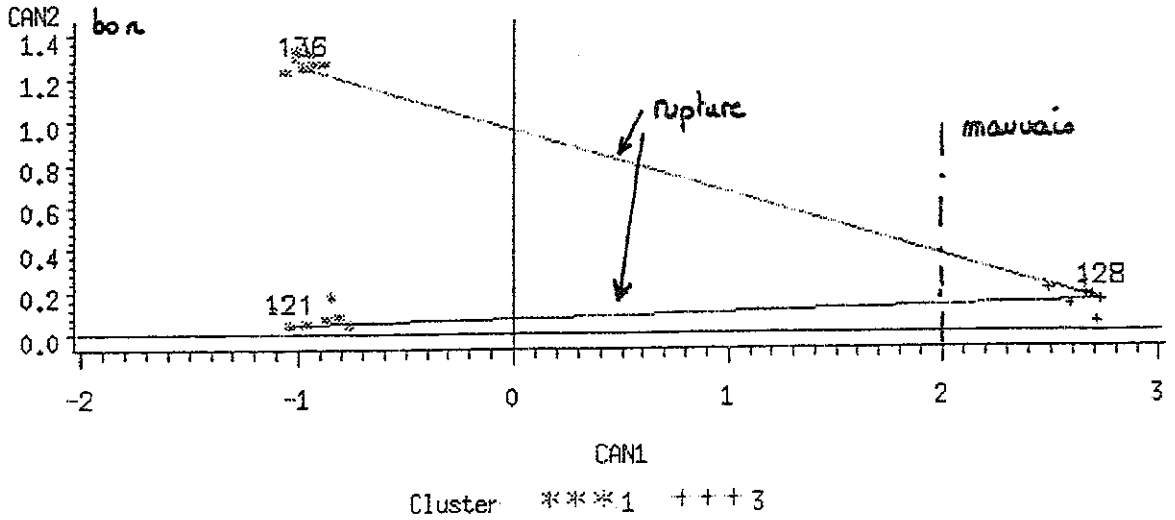
DONNEES AFFECTEES (ECH:121 -- 142) 1
 EN PRINCIPE LES DEFAUTS DEF NE SONT PAS CONNUS
 15:41 Monday, September 30, 1996

OBS	N	PRES	TEMP	HYGR	DEF	CAN1	CAN2	CLUSTER
1	121	11.9520	19.9620	49.5960	3	-1.03785	0.04469	1
2	122	12.0720	20.0680	50.0430	6	-0.76177	0.05059	1
3	123	11.9740	19.9750	50.5830	4	-0.84856	0.17895	1
4	124	11.9550	20.0680	50.3940	4	-0.80659	0.08704	1
5	125	11.9500	20.0030	49.8470	2	-0.96035	0.05250	1
6	126	12.0300	20.0070	49.9560	4	-0.87006	0.07456	1
7	127	11.8850	19.8980	49.8610	6	-1.11424	0.11789	1
8	128	12.0120	22.0670	60.2720	24	2.72751	0.14975	3
9	129	11.9090	22.0560	60.4820	23	2.66176	0.17517	3
10	130	11.9800	21.9860	60.4120	25	2.64513	0.22069	3
11	131	11.9850	21.9060	59.8730	20	2.48699	0.19857	3
12	132	12.0340	21.9870	59.7320	15	2.58390	0.12945	3
13	133	12.0850	21.9400	60.1780	15	2.65448	0.23058	3
14	134	11.9350	22.1570	60.0550	17	2.71211	0.04918	3
15	135	11.9360	22.0610	60.4570	14	2.68581	0.17136	3
16	136	9.9640	20.0590	60.1230	4	-0.97684	1.24810	1
17	137	9.9800	20.0940	60.4250	6	-0.88137	1.26930	1
18	138	10.0950	19.9550	59.8540	5	-1.00592	1.29518	1
19	139	9.9750	20.0150	59.8000	7	-1.06093	1.23310	1
20	140	10.1480	19.9500	59.9420	4	-0.95085	1.31720	1
21	141	10.0570	19.9460	60.0930	4	-1.00903	1.33088	1
22	142	10.0100	20.0560	60.1930	6	-0.92879	1.26540	1

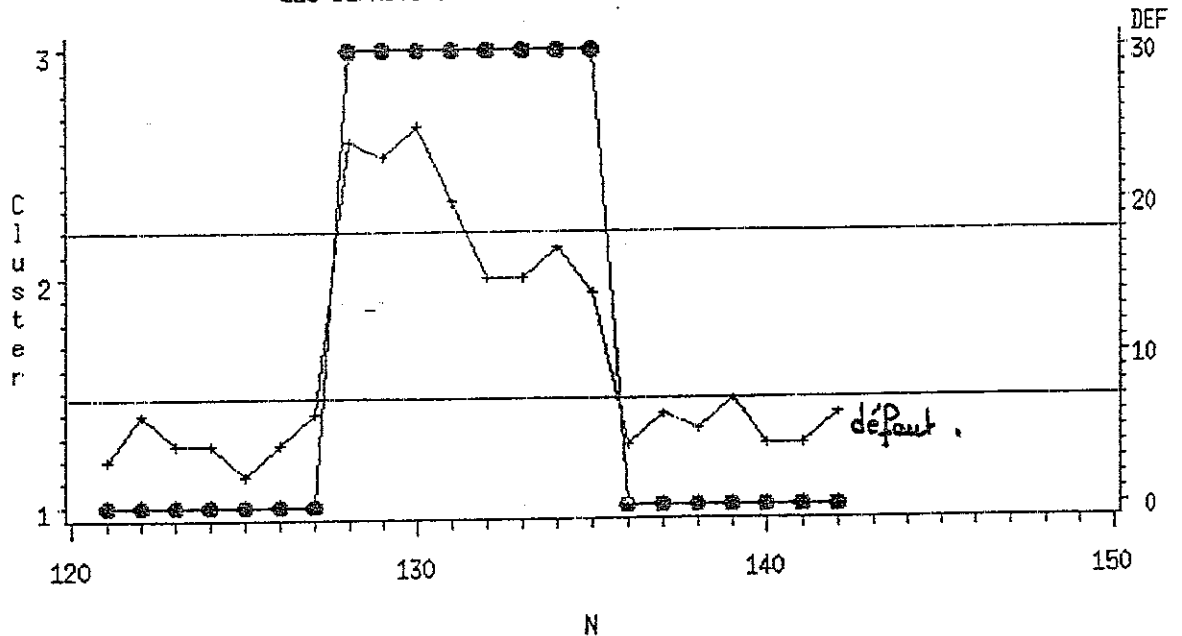
RUPTURES MULTIDIMENSIONNELLES

$$CAN1 = 0.1599*(HYGR-\mu)/s + 0.8629*(PRES-\mu)/s + 0.9527*(TEMP-\mu)/s$$

$$CAN2 = 0.1423*(HYGR-\mu)/s + 0.1155*(PRES-\mu)/s - 0.6753*(TEMP-\mu)/s$$



AFFECTATION DANS LES CLASSES A PARTIR DE L ECHANTILLON 121 LES DEFAUTS OBTENUS SONT ASSOCIES POUR VERIFIER



LES DEFAUTS NE SONT PAS CONNUS AU MOMENT DE L AFFECTATION

