

# CONTRIBUTION À L'ÉTUDE DU COÛT DES SINISTRES AUTOMOBILES

P. PICARD

Paris

## SUMMARY

### *A Contribution to the Study of Automobile claims amounts*

After having a look at the results obtained by adjustment of automobile claims amounts distribution, we research how the number and the time-configuration of past claims condition the claims law of probability.

We have statistics about a group of 471 000 cars which was followed for three years: 1970, 1971 and 1972. We use mathematical technics and among multi-dimensional analysis, we use factorial analysis of correspondance (A.F.C.). A.F.C. permits us to show the link which exists between the claim amount of the third year and the number of claims during the two years before. A quantitative analysis of the corporal claims shows that, of the frequency of corporal claims during the third year grows up in function of the number of past claims, the expected corporal claims amount of the third year decreases as the square of the material claims number during the two first years.

## I. POSITION MATHÉMATIQUE

La notion de processus de risque est désormais bien connue des actuaires. On ne rappellera donc ici que les définitions et propriétés utiles pour la suite des calculs.

Soit  $S_t$  la somme des montants des sinistres pendant la période de temps  $(0, t)$ .  $S_t$  est une variable aléatoire dépendant du temps, c'est un processus aléatoire que l'on décompose en :

- la probabilité  $P_n^m(t, s)$  pour que le nombre de sinistres passe de  $n$  à  $m$  pendant la période de temps  $(t, s)$ ;
- la fonction  $F_t(x/y)$ , probabilité pour que  $S_t$  soit inférieur à  $y$  sachant qu'à l'instant précédent  $t$ , il était égal à  $x$  et sachant que  $t$  est l'abscisse d'un saut du processus (un sinistre).

Cette fonction  $F_t(x/y)$  est l'objet de cette étude. On a :

$$F_t(x/y) = \text{Prob} [S_t < y / S_{t-\varepsilon} = x \text{ et } t = \text{abscisse d'un sinistre}]$$

soit en posant  $z = y - x$  (montant du sinistre à l'instant  $t$ )

$$F_t(z) = \text{Prob} [\Delta S_t < z / t = \text{abscisse d'un sinistre}]$$

où  $\Delta S_t$  est l'accroissement de  $S_t$  à l'instant  $t$ .

## II. AJUSTEMENTS DE LOIS

II.1. *Précautions à prendre pour analyser des coûts de sinistres automobiles*

La base statistique est un ensemble de sinistres survenus au cours d'un certain laps de temps à un groupe de véhicules bien défini. Mais pour analyser ces chiffres, des précautions doivent être prises :

- Si l'on observe des sinistres récents, beaucoup d'entre eux ne sont réglés que partiellement et la partie évaluée est peu précise. Pour avoir une meilleure connaissance des coûts, il faudra attendre le moment où la proportion des dossiers restant en évaluation est faible.
- Dans une étude de ce type, surtout si la période d'observation est longue, on est amené à comparer des sommes à des instants différents et, donc, se pose le problème du choix (ou de la construction) du type d'actualisation.

II.2. *Résultats obtenus*

Monsieur Marcel Henry a montré que la fonction  $y$ , nombre de sinistres supérieurs à une garantie  $x$ , pouvait être représentée d'une façon assez satisfaisante par la fonction de Galton-MacAlister :

$$y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \text{ avec } z = a \text{ Log } x + b$$

Il apparaît toutefois nécessaire de donner à  $a$  deux valeurs différentes, l'une pour les  $x$  inférieurs à un certain montant, l'autre pour des valeurs de  $x$  plus élevées (le nombre de gros sinistres décroît très rapidement).

On obtient des résultats comparables avec la formule de Pareto :

$$\text{Log } y = a \text{ Log } x + b \quad \text{ou} \quad y = \frac{b}{x^a}$$

qui a l'avantage de conduire à des calculs plus simples. Mais, comme dans la loi proposée par Monsieur Marcel Henry, on doit ajuster plusieurs courbes suivant l'importance des sinistres.

Monsieur B. Almer a proposé d'ajuster la distribution des sinistres par un trinôme exponentiel :

$$\Psi(x) = n [a_1 \beta_1 c^{-\beta_1 x} + a_2 \beta_2 c^{-\beta_2 x} + a_3 \beta_3 c^{-\beta_3 x}]$$

avec  $a_1 + a_2 + a_3 = 1$

Plus récemment, Monsieur Gaudibert, dans une thèse présentée devant l'Institut des Actuaire Français a ajusté, pour les *gros sinistres*, une fonction du type:

$$y = \frac{A}{x^a e^{bx}}$$

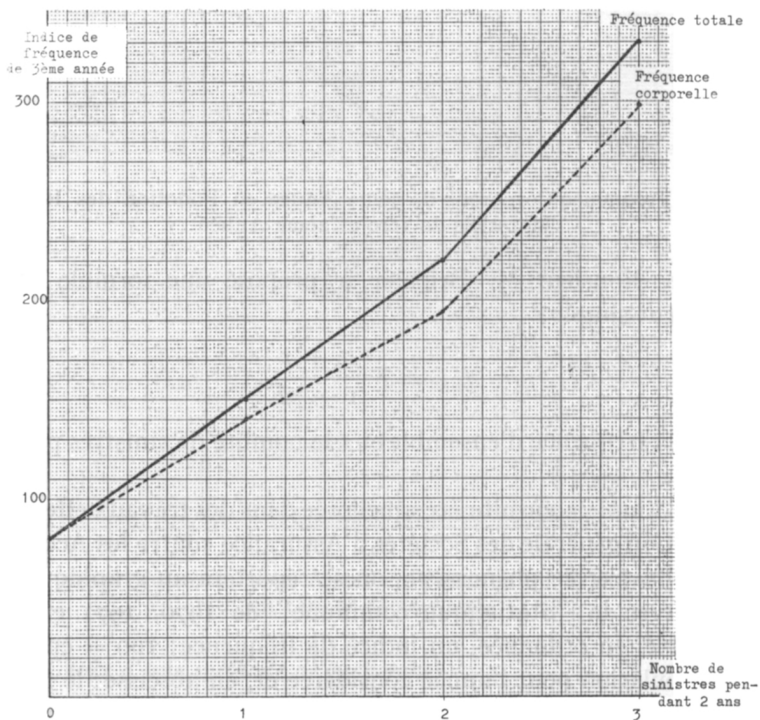
### III. LIAISONS DES RÉSULTATS DES ANNÉES SUCCESSIVES

#### III.I. Rappel des résultats obtenus pour les fréquences des sinistres

P. Depoid, dans son ouvrage "Applications de la statistique aux Assurances", fait apparaître la liaison entre les fréquences d'années successives de mêmes assurés. Monsieur Delaporte a formalisé le

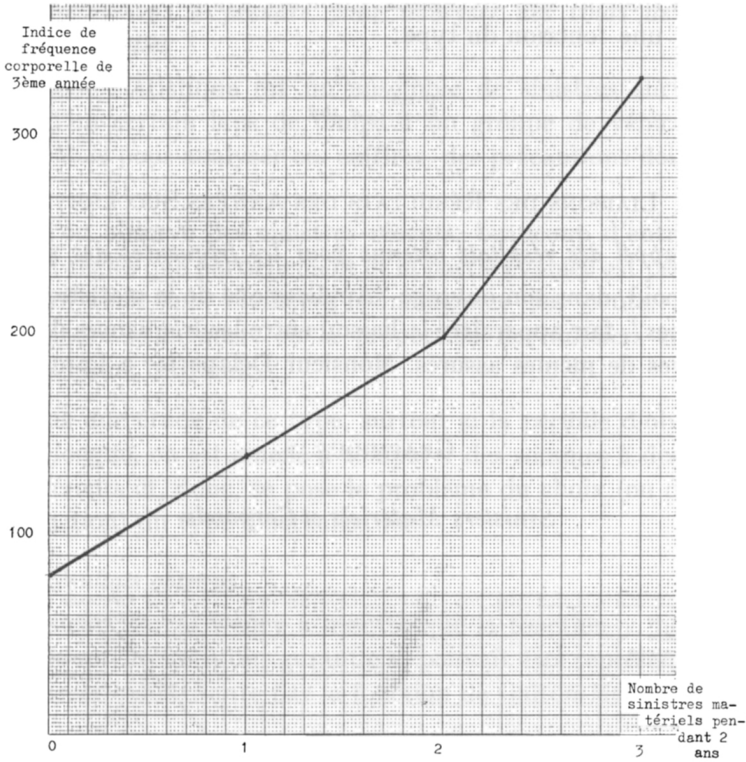
#### GRAPHIQUE I

#### *Sinistres Matériels et Corporels*



Indice de Fréquence de 3ème Année en Fonction des Résultats de 1ère et 2ème Année.

## GRAPHIQUE 2



Indice de Fréquence Corporelle de 3ème Année en Fonction des Résultats Matériels de 1ère et 2ème Année

problème. Monsieur M. Brichler a proposé une formule remarquablement simple:

$$f_{n/x} = F \frac{1 + x}{1 + nF}$$

où  $F$  est la fréquence d'ensemble et  $x$  le nombre de sinistres pendant  $n$  années.

Cette formule a ensuite été améliorée dans des travaux effectués à l'Association générale des Sociétés d'Assurance contre les Accidents.

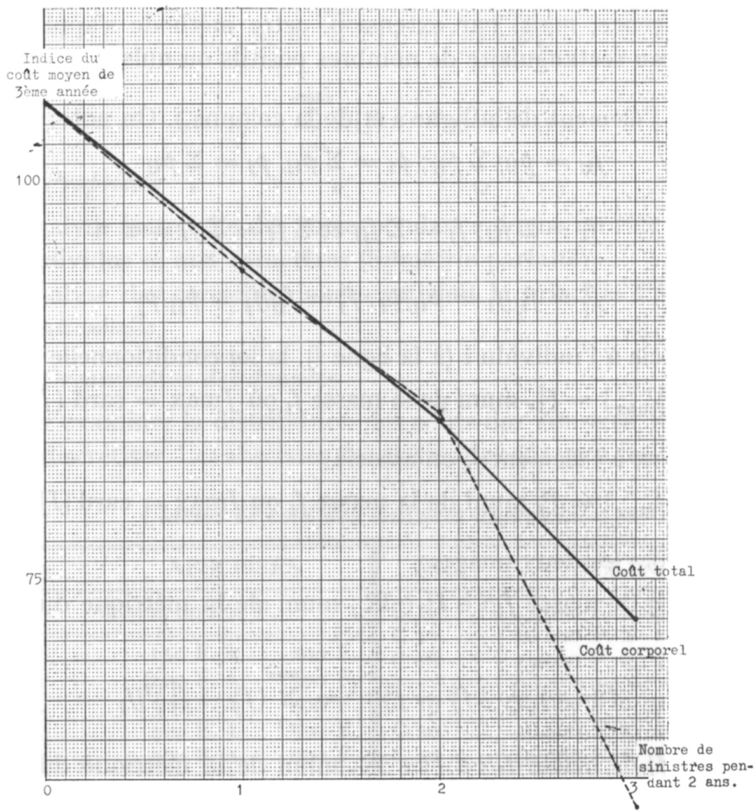
Pour illustrer ce phénomène, on se reportera au graphique n° 1 obtenu avec la "Statistique commune" de 1970, 1971 et 1972. On a porté, en ordonnée, un indice de fréquence: 100 = ensemble de la population pour l'année considérée et, en abscisse: le nombre de sinistres survenus dans les deux années précédentes. Le graphique n° 2 montre l'accroissement du risque corporel en fonction du nombre de sinistres matériels passés.

### III.2. *Liaisons entre les résultats des années précédentes et le coût moyen des sinistres automobiles*

Sur le graphique n° 3, on porte en abscisse, le nombre de sinistres des deux premières années et, en ordonnée, un indice du coût moyen des sinistres de troisième année. La décroissance du coût moyen en fonction de la gravité des antécédents est nette. En effet, pour 3 sinistres en deux ans, le coût moyen est diminué de 25%.

GRAPHIQUE 3

#### *Sinistres Matériels & Corporels*



Indice du cout Moyen de 3ème Année en Fonction du Nombre de Sinistres en 1ère et 2ème Année

### III.3. *Choix d'une méthode de recherche*

Le but de la présente étude est de rechercher comment le nombre et la configuration temporelle des sinistres passés *conditionnent la distribution des coûts* des sinistres de dernière année.

La statistique classique permet d'ajuster une loi de probabilité de forme analytique donnée aux résultats empiriques. Pour étudier la liaison temporelle dans le coût des sinistres automobiles, il est intéressant de se détacher d'une hypothèse de loi: les "analyses multidimensionnées" et, plus particulièrement, l'"Analyse factorielle des correspondances" (A.F.C.) choisie dans cette étude le permettent.

#### IV. ANALYSE FACTORIELLE DES CORRESPONDANCES (A.F.C.)

##### IV.1. *Théorie*

Cette analyse est généralement utilisée dans l'étude des tableaux de contingence, mais, par extension, cette méthode s'applique à tout tableau rectangulaire de nombres positifs ou nuls.

Soit  $p_{ij}$  l'élément de la ligne  $i$  et de la colonne  $j$ . On note:

$$\begin{aligned} f_{ij} &= p_{ij} / \sum_{i,j} p_{ij}; \quad p_i = \sum_i p_{ij}, \quad p_j = \sum_j p_{ij} \\ f_i &= \sum_j f_{ij}; \quad f_j = \sum_i f_{ij} \text{ et } \sum_i f_i = \sum_j f_j = 1 \\ f_j^i &= f_{ij} / f_i = p_{ij} / p_i; \quad f_j^i = f_{ij} / f_j = p_{ij} / p_j \end{aligned}$$

On associe à l'individu  $i$  (à la ligne  $i$ ), la loi conditionnelle sur  $J$ :  $\{f_1^i, \dots, f_{J_{\max}}^i\} = f_J^i$ , muni de la masse  $f_i$  où  $J_{\max} = \text{card}(J)$ .

De même, à l'élément  $j$ , on associe:

$\{f_1^j, \dots, f_{I_{\max}}^j\} = f_I^j$  muni de la masse  $f_j$  ou  $I_{\max} = \text{card}(I)$ .

On a alors les deux nuages:

$N(I) = \{f_J^i \text{ de masse } f_i / i \in I\} \subset \mathbb{R}_J$  muni de la métrique du  $\chi^2$  de centre  $f_J$ :

$$d^2(i, i') = \|f_J^i - f_J^{i'}\|^2 = \sum_{j=1}^{J_{\max}} (f_j^i - f_j^{i'})^2 / f_j$$

$N(J) = \{f_I^j \text{ de masse } f_j / j \in J\} \subset \mathbb{R}_I$  muni de la métrique du  $\chi^2$  de centre  $f_I$ .

Le meilleur espace de dimension  $k$  représentant  $N(I)$  est engendré par  $k$  vecteurs orthonormés de  $R_J \{(e_1)_I \dots (e_k)_J\}$ , auquel correspondent les opérateurs de projection (facteurs):  $\phi_1^J \dots \phi_k^J$  appartenant à  $\mathbb{R}^J = (\mathbb{R}_J)^*$  qui sont les vecteurs propres de  $m \circ \sigma$  ( $m$  est la métrique et  $\sigma$  est la forme quadratique d'inertie) correspondant aux  $k$  plus grandes valeurs propres:  $\lambda(\phi_1), \dots, \lambda(\phi_k)$ .

Le nuage  $N(I)$  est approximé par sa projection sur la variété précédente, à l'individu  $i$  on associe ses coordonnées:  $G(i, 1), \dots, G(i, k)$ , et on a avec le  $n$ ième facteur:

$$G(i, n) = \phi_n^J(f_j^i) = \sum_{j=1}^{J_{\max}} \phi_N^j f_j^i = \sqrt{\lambda(\phi_n)} \phi_n^i$$

De même pour  $N(J)$ , à  $j$  on associe:

$$F(j, 1) \dots F(j, k)$$

avec

$$F(j, n) = \sqrt{\lambda(\phi_n)} \phi_n^j$$

et

$$G(i, n) = \sum_{j=1}^{J_{\max}} F(j, n) f_j^i / \sqrt{\lambda(\phi_n)}$$

$$G(i, n) = \sum_{j=1}^{J_{\max}} F(j, n) P_{ij} / P_i \sqrt{\lambda(\phi_n)}$$

#### IV.2. Propriétés

L'analyse factorielle des correspondance donne un rôle identique aux individus et aux caractères (Symétrie parfaite).

L'analyse des correspondances satisfait le principe d'équivalence distributionnelle.

#### IV.3. Organisation des données

Pour examiner la distribution des coûts des sinistres, on est conduit à discrétiser. Les abscisses des classes de coût sont les suivantes:

##### Classes de coût des sinistres automobiles

No.	ABSCISSES	No.	ABSCISSES
C 01	Moins de 150 F.	C 16	5 000 F. à 10 000 F.
C 02	150 F. à 200 F.	C 17	10 000 F. à 15 000 F.
C 03	200 F. à 300 F.	C 18	15 000 F. à 20 000 F.
C 04	300 F. à 400 F.	C 19	20 000 F. à 30 000 F.
C 05	400 F. à 500 F.	C 20	30 000 F. à 40 000 F.
C 06	500 F. à 600 F.	C 21	40 000 F. à 50 000 F.
C 07	600 F. à 700 F.	C 22	50 000 F. à 100 000 F.
C 08	700 F. à 800 F.	C 23	100 000 F. à 150 000 F.
C 09	800 F. à 900 F.	C 24	150 000 F. à 200 000 F.
C 10	900 F. à 1 000 F.		
C 11	1 000 F. à 1 500 F.	C 25	200 000 F. à 500 000 F.
C 12	1 500 F. à 2 000 F.		
C 13	2 000 F. à 3 000 F.	C 26	Plus de 500 000 F.
C 14	3 000 F. à 4 000 F.		
C 15	4 000 F. à 5 000 F.		

Pour chaque configuration de sinistres durant les trois années consécutives observées, on note la distribution des coûts des sinistres de troisième année. C'est-à-dire que:

- pour tout triplet n°  $i(a, b, c)$  avec  $0 \leq a \leq 3, 0 \leq b \leq 3$  et  $1 \leq c \leq 3$ , on a les nombres  $P_{ij}$  des sinistres de la classe de coût n°  $j$ .

CONFIGURATIONS				TRANCHES DE COÛTS					
No.	1 <sup>ère</sup> année	2 <sup>ème</sup> année	3 <sup>ème</sup> année	$C_{01}$	$C_{02}$	....	$C_j$	....	$C_N$
1	0	0	1	$P_{11}$	$P_{12}$	....	$P_{1j}$	....	$P_{1N}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$i$	$a$	$b$	$c$	$P_{i1}$	$P_{i2}$	....	$P_{ij}$	....	$P_{iN}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
M	3	3	3	$P_{M1}$	$P_{M2}$	....	$P_{Mj}$	....	$P_{MN}$

M = 48 configurations      N = 26 classes

*REMARQUES: On a retenu comme nombre maximum de sinistres par an, le nombre trois, afin de posséder dans chaque cas un nombre d'observations suffisant pour l'analyse.*

V. ANALYSES DES RÉSULTATS

Les taux d'inertie des axes factoriels sont faibles. Cependant, les plans des axes 1 et 2 et des axes 1 et 3 possèdent des parts d'inertie suffisantes pour permettre une interprétation.

Pour améliorer la commodité de lecture des graphiques en cas de points superposés, on imprime un identificateur d'autant plus noir que la multiplicité est grande.

Les configurations 331 et 133 se distinguent particulièrement et rendent l'interprétation du reste des nuages difficile. Les figures n° 4 et n° 5 représentent les projections des nuages sur les plans (1, 2) et (1, 3) après suppression de points 331 et 133 (sans modification des actes factoriels).

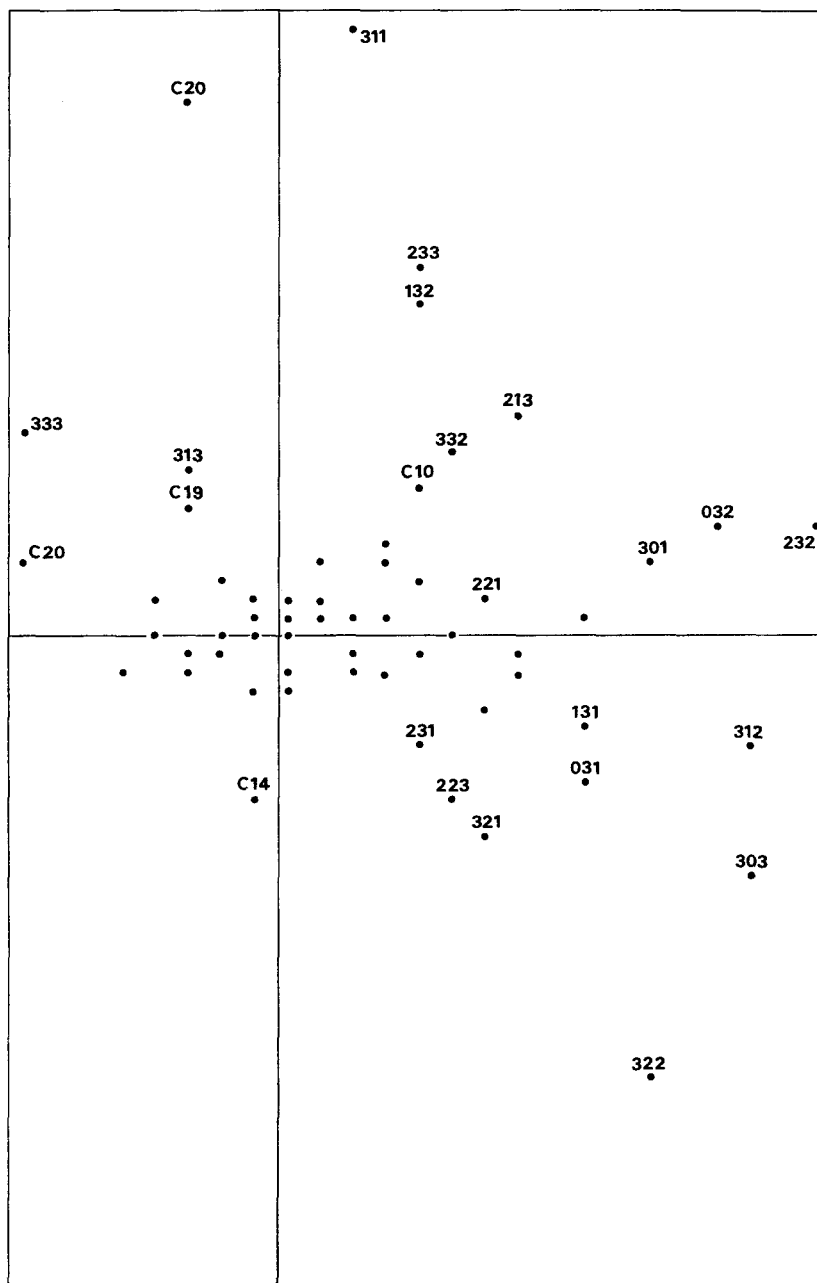
L'étude du plan (1, 3), figure n° 5, permet de mettre en évidence un effet Guttman, c'est-à-dire que l'on peut disposer le tableau de



GRAPHIQUE 4

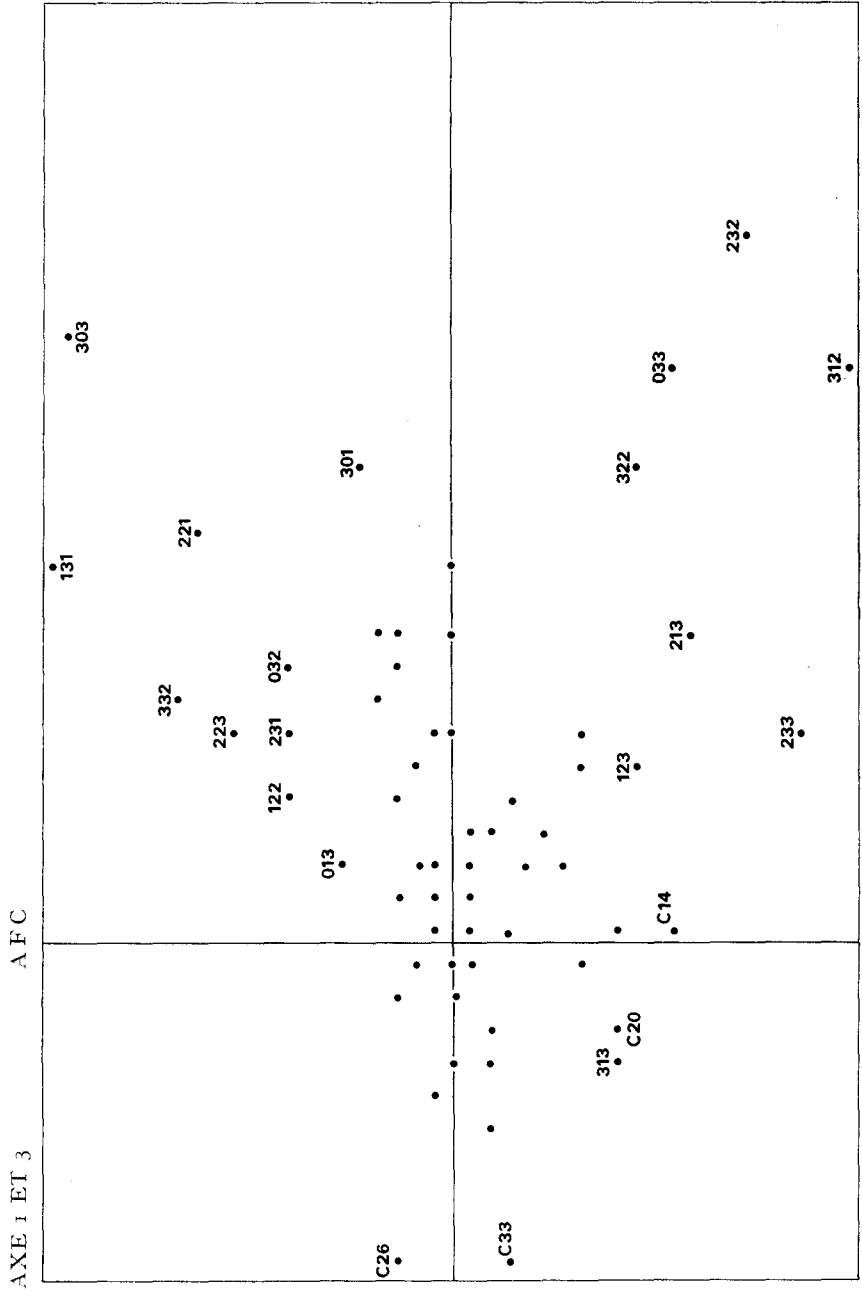
AXE 1 ET 2

AFC



Coûts des Sinistres Automobiles

GRAPHIQUE 5



Couts des Sinistres Automobiles

données, par permutation des lignes et des colonnes, sous une forme *bloc-diagonale* et donc il existe un *classement des types de configuration* qui induit un classement des coûts des sinistres de 3<sup>ème</sup> année. Cette remarque prouve l'existence d'un conditionnement de la distribution des coûts par les antécédents.

L'explication des liaisons entre les configurations et les coûts des sinistres se déduit de l'interprétation des premiers axes factoriels. Il est pratique d'examiner successivement les projections des nuages *I* et des nuages *J* (graphiques n° 6 et n° 7).

Sur le graphique n° 6, on repère les configurations par les figures suivantes :

Nombre de sinistres en deux ans	Figures
0 .....	□
1 .....	○
2 .....	△
3 .....	▽

L'analyse des groupements ainsi obtenus permet de montrer que l'axe horizontal classe les configurations par leurs nombres de sinistres dans les deux premières années. On remarque que le nombre de sinistres de 3<sup>ème</sup> année ne semble pas avoir d'influence sur la distribution des coûts.

La forme triangulaire du graphique n° 7 est caractéristique d'éléments classés naturellement (tranches de coût). Pour l'interprétation, on doit prendre soin de se baser principalement sur les éléments ayant un poids important (population importante). A cette condition, on remarque le phénomène de classement sur l'axe horizontal suivant les coûts (coût élevé à gauche, moyen et faible à droite). La distinction entre tranches de coût moyen et tranches de coût faible peut être observée sur l'axe vertical.

Cette étude permet de mettre en évidence la liaison entre les survenances des sinistres passés et le coût des sinistres présents.

La distribution des coûts des sinistres de 3<sup>ème</sup> année est d'autant plus biaisée vers les classes inférieures que le nombre de sinistres pendant les deux premières années est important.

## VI. ÉTUDES QUANTITATIVES

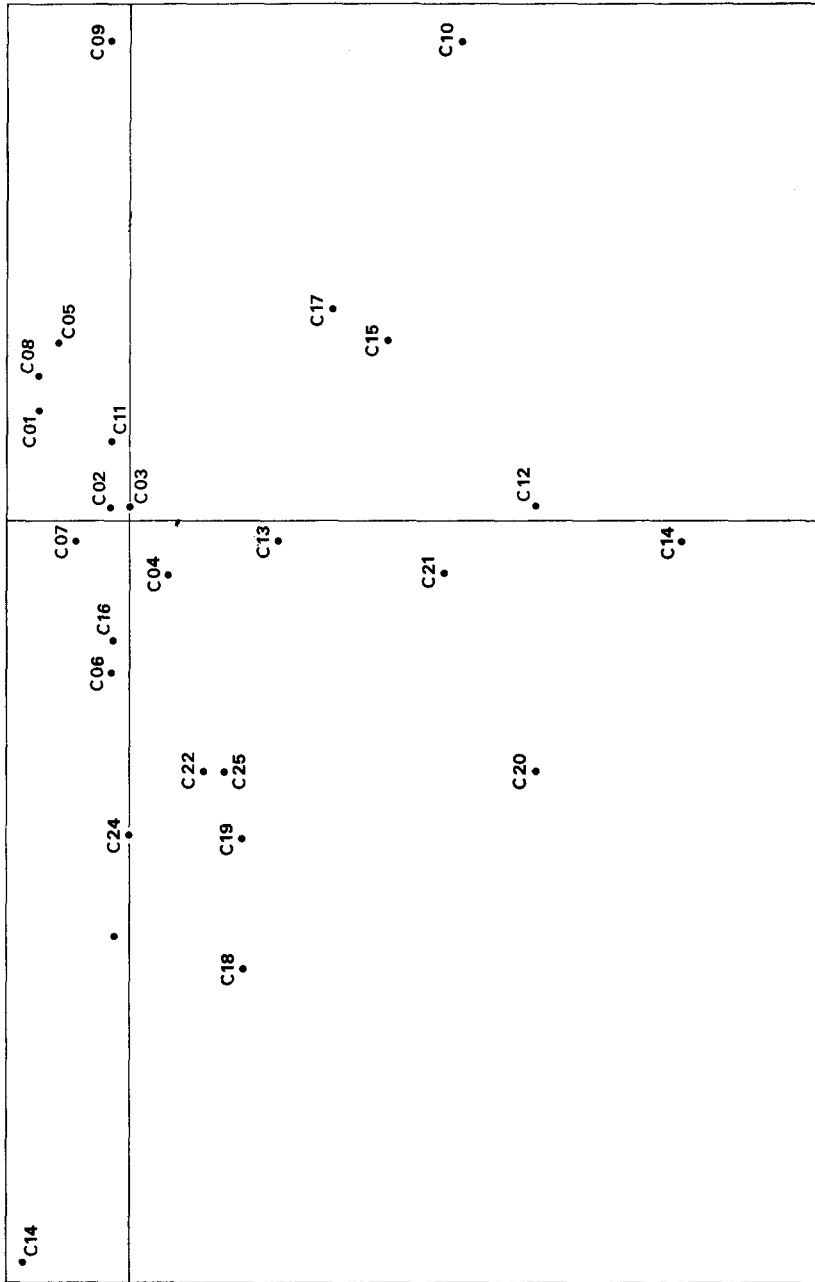
La mesure du phénomène observé réclamerait une période d'observation plus longue, car on constate que la sélection des assurés par leurs survenances est moins rapide vis-à-vis des coûts



GRAPHIQUE 7

AFC

AXE 1 ET 3



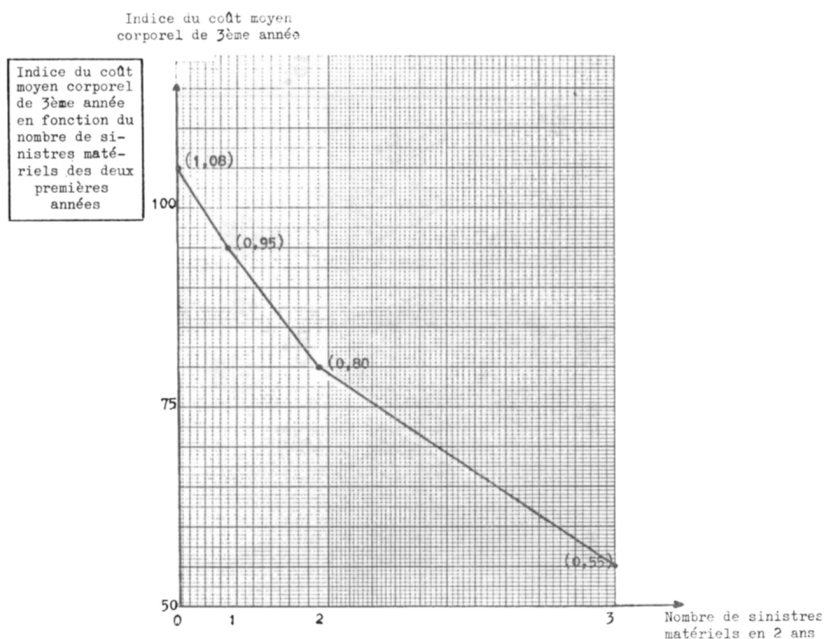
Coûts des Sinistres Automobiles

que vis-à-vis des fréquences. Cependant, sur trois ans, l'influence est particulièrement nette sur les tranches élevées de coûts, c'est-à-dire sur les sinistres corporels.

On a vu (graphique n° 2) que la fréquence corporelle croît rapidement en fonction des antécédents. On va montrer que le *coût moyen corporel* décroît en fonction du nombre de sinistres passés.

Le graphique n° 8 montre que le coût moyen corporel de 3<sup>ème</sup> année décroît en fonction du nombre de sinistres matériels des deux années précédentes. On remarque que si l'assuré n'a aucun sinistre matériel pendant deux ans, son coût moyen corporel est supérieur d'environ 5% au coût moyen corporel de l'ensemble.

GRAPHIQUE 8



La baisse du coût moyen corporel de 3<sup>ème</sup> année est proportionnelle au carré du nombre de sinistres matériels des deux premières années.

Nombre de sinistres matériels en deux ans	N <sup>2</sup>	Baisse du coût moyen corporel
1	1	5 %
2	4	20 %
3	9	45 %

## VII. CONCLUSION

L'analyse factorielle des correspondances appliquée dans cette étude permet de mettre en évidence le conditionnement de la distribution des coûts des sinistres de dernière année par le nombre de sinistres passés. La probabilité d'un sinistre de coût élevé diminue très rapidement lorsque le nombre d'antécédents augmentent.

La probabilité de survenance d'un sinistre corporel est d'autant plus forte que le nombre de sinistres passés est important. Cependant, l'étude quantitative a permis de mesurer la décroissance de gravité des sinistres corporels. Le coût moyen corporel de troisième année semble être une fonction quadratique décroissante du nombre de sinistres matériels des deux années précédentes.