

La sélection des assurés et la détermination des primes d'assurances par l'analyse discriminante : une application à l'assurance automobile

Autor(en): **Beuthe, M.**

Objektyp: **Article**

Zeitschrift: **Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuairees Suisses = Bulletin / Association of Swiss Actuaries**

Band (Jahr): **75 (1975)**

PDF erstellt am: **27.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-967111>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

B.

Wissenschaftliche Mitteilungen

La sélection des assurés et la détermination des primes d'assurances par l'analyse discriminante

Une application à l'assurance automobile

Par M. Beuthe, Ph. Van Namen

Toute société de crédit ou d'assurance doit évaluer, d'une façon ou d'une autre, le risque qu'elle encourt en accordant un crédit ou en acceptant d'assurer un client. Selon cette évaluation, elle détermine les conditions sous lesquelles elle peut passer contrat et, éventuellement, refuser certains contrats.

Afin de résoudre ce problème, une méthode statistique appelée l'analyse discriminante est utilisée depuis quelques années¹. Celle-ci permet de classer les clients en bons ou mauvais clients tout en tenant compte des conditions particulières à chaque contrat. Nous nous proposons de présenter, en cet article, une application de cette méthode au problème de la sélection des assurés. Après avoir reformulé les principes de l'analyse discriminante en termes appropriés, nous appliquerons la méthode aux données d'assurance automobile d'une compagnie et en discuterons les résultats. Nous pourrions également calculer les primes d'assurances qui correspondent au risque particulier de certains conducteurs et proposer ainsi des éléments d'appréciation du barème actuel d'assurance automobile.

I. Le modèle théorique

Définissons:

- un assuré par le vecteur $X = (X_1, \dots, X_m)$ des m caractéristiques particulières à cet assuré: sexe, âge, puissance de la voiture, type de contrat, etc. ...;
- q_1 la probabilité a priori qu'un assuré quelconque ne cause pas de sinistre et $q_2 = 1 - q_1$;

Les auteurs ont bénéficié des commentaires fort utiles de *L. Eeckhoudt, J.M. Gagnon, S. Moreland* et *R. Roll*. Ils désirent les en remercier au début de cet article. Certains des résultats de cet article ont été publiés sous le titre « la sélection des assurés et la détermination, des primes d'assurances » dans les Cahiers Economiques de Bruxelles, n° 64, 4^e trimestre, 1974, p. 563–577.

¹ Voir, par exemple, *J. Degueldre*: «Une application de la recherche opérationnelle dans l'analyse des demandes de crédits bancaires – La méthode de crédit scoring», Cahiers Economiques de Bruxelles, n° 50, 2^e trimestre, p. 279–302.

- $p_1(X)$, la fonction de densité des caractéristiques de la population des assurés ne causant pas de sinistre, et $p_2(X)$ celle des assurés causant un sinistre;
- $R(X)$, la prime nette de taxes et de la commission payée à l'intermédiaire, fonction des caractéristiques X ;
- $C(X, u)$, le coût d'un sinistre, fonction de X et de u une variable aléatoire indépendante distribuée selon $f(u)$, avec $E[u/X] = 0$.

Au lieu de suivre l'approche habituelle de l'analyse discriminante² qui minimise le coût attendu de mal classer une observation, considérons son équivalent qui consiste à maximiser le profit attendu d'un bon classement. Dans le cas présent, on considère que la compagnie d'assurance désire déterminer la règle de sélection pour ses assurés qui lui permettra de maximiser l'espérance mathématique du profit d'un contrat. Pour ce faire, elle devra choisir la région d'acceptation S qui maximise:

$$\int_S R(X)q_1p_1(X)dX + \int_S \int_{-\infty}^{\infty} (R(X) - C(X, u))q_2p_2(X)f(u)dXdu \quad (1)$$

ou encore

$$\int_S R(X)q_1p_1(X)dX + \int_S [R(X) - \tilde{C}(X)]q_2p_2(X)dX \quad (2)$$

après avoir effectué l'intégration par rapport à u , $\tilde{C}(X) = E[C(X, u)/X]$. Tous les assurés ou contrats appartenant à S seront acceptés, tandis que les autres seront refusés. Si l'on définit

$$P(X) = \frac{q_1p_1(X)}{q_1p_1(X) + q_2p_2(X)} \quad (3)$$

la probabilité conditionnelle que l'assuré X ne cause pas de sinistre, et

$$1 - P(X) = \frac{q_2p_2(X)}{q_1p_1(X) + q_2p_2(X)}$$

la probabilité conditionnelle qu'il cause un sinistre, la région d'acceptation R sera telle que tous les assurés pour lesquels

$$R(X)P(X) + [R(X) - \tilde{C}(X)](1 - P(X)) > 0 \quad (4)$$

² C. R. Rao, «Linear Statistical Inference and its applications», John Wiley, 1965.

seront acceptés. Ceci va de soi, si l'on considère que le côté gauche de cette expression est le profit attendu du contrat avec un assuré X^3 .

Certains pourraient objecter à cette approche qu'elle aboutit à refuser l'assurance à certaines catégories de personnes. Cependant, il est possible en principe de fixer la prime $R(X)$ à un niveau qui permette d'accepter tout contrat. Cette analyse procure d'ailleurs une méthode très simple pour calculer la prime d'assurance minimale correspondant au risque particulier à un assuré. Cette prime devrait être telle que

$$R^*(X)P(X) + [R^*(X) - \tilde{C}(X)](1 - P(X)) = 0$$

c'est-à-dire

$$R^*(X) = (1 - P(X))\tilde{C}(X) \quad (5)$$

C'est la prime qui permettrait de couvrir en moyenne les indemnités payées par la compagnie aux assurés définis par X par les revenus procurés par ces assurés. Elle procurerait à la compagnie un profit moyen nul.

II. Le modèle économétrique

Nous devons résoudre deux problèmes: le premier est de séparer les assurés actuels ou potentiels en deux classes de «bons» et de «mauvais» assurés, le second est d'estimer $P(X)$, la probabilité conditionnelle qu'un assuré ne causera pas de sinistre durant la période considérée. La méthode utilisée dans cet article est celle de l'analyse discriminante linéaire et quadratique⁴.

Celle-ci est basée sur l'inégalité (4). Elle peut se réécrire

$$\frac{P(X)}{1 - P(X)} = \frac{q_1 p_1(X)}{q_2 p_2(X)} > \frac{\tilde{C}(X) - R(X)}{R(X)}$$

Sous l'hypothèse que X est distribué selon une loi normale multi-variée et que les matrices de variance-covariance de $p_1(X)$ et $p_2(X)$ sont identiques, on obtient:

³ Cette règle de sélection suppose implicitement l'indifférence par rapport au risque. Voir à ce sujet *M. V. Beuthe et Ph. Van Namen*, «Discriminant Analysis with Risk Aversion», rapport de recherche, Faculté Universitaire Catholique de Mons, juin 1974.

⁴ Une autre application de cette méthode peut être trouvée dans *J. Degueldre*, op.cit.

$$\ln \frac{p_1(X)}{p_2(X)} = \alpha + X\beta > \ln \frac{q_2}{q_1} + \ln \frac{(\tilde{C} - R)}{R} \quad (6)$$

où α est l'ordonnée à l'origine et β le vecteur des autres coefficients de la fonction discriminante normale⁵. Lorsque l'hypothèse d'égalité des matrices de variances et covariances ne peut être soutenue, la fonction discriminante est une fonction quadratique.

C'est la règle de décision utilisée par l'analyse discriminante pour classer les observations sans dériver explicitement les probabilités conditionnelles $P(X)$. Cependant, celles-ci peuvent être obtenues par la formule

$$P(X) = \frac{1}{1 + \frac{q_2 p_2(X)}{q_1 p_1(X)}} = \frac{1}{1 + e^{K - (\alpha + \beta X)}} \quad (7)$$

où $K = \ln \frac{q_2}{q_1}$, pour le cas de la fonction linéaire.

Dans l'analyse empirique qui portera sur les données d'assurance automobile d'une compagnie particulière, l'hypothèse de normalité ne pourra être satisfaite, car toutes les variables explicatives seront des variables qualitatives binaires. Cependant, l'analyse discriminante permet d'introduire un grand nombre de variables explicatives, même avec un échantillon de taille modérée. Elle opère directement sur les observations originales sans nécessiter leur regroupement sous forme de fréquences observées. Ceci évite de délicats problèmes de regroupement d'observations qui résulteraient du grand nombre de variables explicatives. De plus, E.S. Gilbert a montré, par méthode de simulation, que les résultats d'une analyse discriminante sur des variables qualitatives se compareraient très favorablement avec ceux d'autres méthodes théoriquement mieux fondées⁶.

Ensuite, on peut considérer que la fonction discriminante n'est définie qu'à un coefficient de proportionnalité près. Elle peut être également trouvée en maximant le rapport du carré de l'écart entre les deux moyennes de la fonction discriminante sur la somme des carrés des écarts dans les deux populations; c'est la fonction discriminante de Fisher, proportionnelle à la première fonction, mais

⁵ Anderson T.W., «An Introduction to Multivariate Statistical Analysis», Wiley 1958, Ch. 6.

⁶ Gilbert E.S., «On Discrimination Using Qualitative Variables», American Statistical Association Journal, Vol. 63, n° 324, Dec. 1968, pp. 1399-1412.

qui ne repose pas sur l'hypothèse de normalité. Si la fonction discriminante, considérée sous cet angle, comprend un nombre suffisant de variables (indépendantes), le théorème de la limite centrale peut s'y appliquer⁷. Alors, la fonction discriminante, $z = \alpha + \beta X$ sera distribuée approximativement selon la loi normale pour chacun des deux échantillons de bons et mauvais assurés. Dans ce cas, au lieu de considérer $P(X)$, on utilisera $P(z)$ pour obtenir, grâce à une définition semblable à (3), un résultat identique à (7)⁸. Toutes ces raisons nous semblent amplement justifier l'utilisation de cette méthode comme une solution pratique à notre problème d'estimation.

III. Analyse empirique

A. Les caractéristiques de l'échantillon

Pour tous les assurés, agréés après le 31 décembre 1963, d'une petite compagnie d'assurance (1380 contrats), nous disposons des données suivantes:

1. La prime destinée à couvrir effectivement le risque propre à l'année 1971, c'est-à-dire la prime hors taxe (6%) moins la commission;
2. Le coût de la prise en charge, durant l'année 1971, de la responsabilité de l'assuré. Deux possibilités s'offrent:
 - a) la compagnie n'a rien versé à un tiers du fait de l'assurance: le coût est nul;
 - b) la compagnie a versé (ou a mis en réserve) une certaine somme d'argent, à titre de dédommagement: le coût équivaut à cette somme augmentée d'un coût fixe: 2350 francs, propre à la compagnie, correspondant aux frais de gestion du dossier;
3. l'âge atteint au 31 décembre 1970;
4. le sexe;
5. la profession;
6. la nationalité;
7. la cylindrée (en cm^3) du véhicule;
8. l'âge du véhicule;

⁷ L'analyse de *E.S. Gilbert* n'incluait qu'un maximum de six variables explicatives.

⁸ En effet, si $p_1(z)$, la densité de z pour la population des bons assurés est approximativement $n(\frac{1}{2}\theta, \theta)$, $p_2(z)$ est approximativement $n(-\frac{1}{2}\theta, \theta)$; voir *Anderson T. W.*, op. cit., p. 134-135.

- 9. l'usage auquel le véhicule est destiné;
- 10. la marque du véhicule;
- 11. le type d'assurance souscrite.

Les variables 3 à 11 seront introduites comme variables explicatives. La plupart d'entre elles sont de type qualitatif. Afin de mieux détailler le risque inhérent à certaines caractéristiques individuelles, par exemple le jeune âge, nous avons décidé de n'utiliser que des variables binaires en créant un certain nombre de catégories pour les variables continues (âge de l'assuré, âge du véhicule, puissance de la voiture). L'utilisation de variables binaires permet de définir aisément des catégories d'assurés, et facilite, dans une certaine mesure, les calculs de primes d'assurances.

Nous avons considéré comme n'ayant pas causé d'accident l'assuré pour lequel le coût supporté par la compagnie a été nul. L'échantillon a été ainsi segmenté en deux groupes: les «non-responsables d'un accident», au nombre de 930 (67,4%) et les «responsables d'un accident» au nombre de 450 (32,6%). Dans ce dernier groupe, les coûts sont ainsi distribués:

Tableau 1

Montant en fr.	%
1- 1 000	16
1 001- 2 000	14
2 001- 3 000	5
3 001- 4 000	5
4 001- 5 000	7
5 001- 6 000	6
6 001- 8 000	7
8 001- 10 000	8
10 001- 15 000	8
15 001- 20 000	7
20 001- 30 000	6
30 001- 50 000	5
50 001-500 000	5
> 500 000	1
	100

Le coût moyen par accident est obtenu en négligeant le coût d'un accident exceptionnel de deux millions de francs. Il est égal à 17 282 francs⁹. A celui-ci,

⁹ Le coût moyen national par accident était de 16452 francs en 1970. Source UPEA (Union Professionnelle des Entreprises d'Assurance).

il faut ajouter un coût fixe de 2350 francs correspondant aux frais de gestion du dossier pour cette compagnie.

Trois critères permettent de comparer notre échantillon au parc automobile belge:

a) Le nombre de véhicules par classe de cylindrée

Tableau 2

Classe	Parc belge ¹⁰ au 1.1.71, en %	Echantillon en %
< 801 cm ³	12	14
801-1000	16	23
1001-1200	23	25
1201-1600	24	24
> 1600	25	14
	—	—
	100	100

b) La cylindrée moyenne par classe de cylindrée

Tableau 3

Classe	Parc belge ¹⁰ au 1.1.71	Echantillon
< 801 cm ³	610	610
801-1000	908	897
1001-1200	1145	1134
1201-1600	1425	1393
> 1600	2067	1951

¹⁰ Source UPEA.

*c. Le nombre de véhicules par marque**Tableau 4*

Marque	Parc belge ¹¹ au 1.8.71, en %	Echantillon en %
A	15,7	13,7
B	11,3	6,7
C	10,6	6,5
D	8,1	13,4
E	6,8	9,5
F	5,1	4,6
G	8,5	11,5
H	6,2	11,1
I	27,7	23,0
	100,0	100,0

B. La fonction discriminante linéaire

Des 1380 individus, un échantillon test (10%) a été tiré suivant une table de nombres au hasard, laissant un échantillon de 1242 individus pour l'estimation. Les tableaux suivants décrivent cet échantillon en fonction des diverses caractéristiques. Ils donnent également les coefficients de la fonction discriminante linéaire.

¹¹ *Comaubel* (Chambre Syndicale du Commerce Automobile de Belgique). Etude du marché 72, pp.13-15.

Tableau 5

Le risque automobile comparé

Catégories	% d'assurés dans la catégorie	% de responsables d'accidents	Coefficient de discrimination
<i>à l'âge de l'assuré</i> ¹²			
< 23 ans	7,7	46,9	-1,15*
23-25	13,2	42,7	-0,95*
26-30	15,9	39,6	-0,89*
31-40	23,5	28,4	-0,26
41-50	23,4	26,5	-0,10
> 50	16,3	24,8	0
	100	moyenne 32,4	
<i>au sexe de l'assuré</i>			
Hommes	86,5	31,9	0
Femmes	13,5	35,7	-0,05
	100		
<i>à la profession de l'assuré</i>			
Employés – fonctionnaires	24,9	33,0	0
Indépendants	11,8	34,7	-0,10
Sans profession (pensionnés – étudiants – ménagères)	10,2	40,9	-0,53*
Ouvriers	49,6	29,7	0,20
Représ. de l'ordre	3,5	37,2	-0,23
	100		
<i>à la nationalité de l'assuré</i>			
Belge	90,3	30,7	0
Etranger	9,7	49,2	-0,88*
	100		

¹² Dans ce tableau et les suivants, les coefficients marqués d'un astérisque sont statistiquement significatifs selon un test en $F(0,95)$ basé sur le D^2 de Mahalanobis. Voir, par exemple, C. R. Rao, «Linear Statistical Inference and its Application», Wiley 1965, p.483.

Tableau 6

Le risque automobile comparé

Catégories	% d'assurés dans la catégorie	% de responsables d'accidents	Coefficient de discrimination
<i>à l'usage du véhicule</i>			
Tourisme	92,8	32,3	0
Tourisme et affaires	7,2	34,4	-0,01
	100		
<i>au type d'assurance souscrite</i>			
Responsabilité civile avec ou sans contre-assurance spéciale	91,9	32,9	0
Avec dégâts matériels	8,1	27,0	0,22
	100		
<i>à la date de souscription de l'assurance</i>			
1970	24,3	37,7	0,06
1969	18,0	34,8	0
1968	21,4	35,8	-0,05
1967-66	20,3	31,0	0,24
1965-64	16,0	19,1	0,62*
	100	moyenne 32,4	

Les anciens assurés représentent un risque moindre. Ceci pouvait être attendu, car la compagnie résilie les contrats après certains accidents. Il n'est cependant pas possible de savoir si cette politique est la seule explication du nombre plus réduit des accidents par les anciens assurés. Il faut noter également que les détenteurs d'une assurance dégâts matériels (omnium) sont plus prudents que les autres.

Le coefficient de la catégorie d'âge «plus de 50» est égal à zéro, car cette dernière a été choisie comme catégorie de référence. Le pourcentage des responsables d'un accident y est le moins élevé, c'est la raison pour laquelle les valeurs prises par les coefficients des autres catégories sont négatifs. Il apparaît nettement que les conducteurs les plus jeunes causent le plus d'accidents, de même que les conducteurs «sans profession». Les étrangers se révèlent de

dangereux conducteurs, bien que la compagnie applique une politique de sélection assez stricte à leur égard (conditions de résidence, d'âge, etc....).

Tableau 7

Le risque comparé à

Catégories	% d'assurés dans la catégorie	% de responsables d'accidents	Coefficient de discrimination
<i>la cylindrée du véhicule</i>			
< 801 cm ³	14,2	26,1	0,74
801-1000	23,3	33,8	0
1001-1200	24,6	31,0	0,11
1201-1600	23,8	30,8	0,08
> 1600	14,1	41,7	-0,27*
	100		
<i>l'âge du véhicule</i>			
1 an	23,5	35,3	0,34
2 et 3 ans	29,9	26,7	0,67*
4 à 6 ans	28,8	32,4	0,28
plus de 6 ans	17,8	38,5	0
	100		
<i>la marque du véhicule</i>			
A	13,1	27,0	0,54
B	6,4	31,2	0,16
C	11,9	28,4	0,44
D	9,6	31,1	-0,25
E	6,4	40,0	-0,06
F	13,6	34,3	0,03
G	11,4	29,1	0,44
H	4,6	36,8	0,16
I	30,0	36,1	0
	100	moyenne 32,4	

Les conducteurs de grosses cylindrées sont responsables d'un plus grand nombre d'accidents. Les conducteurs de voitures de deux à trois ans d'âge sont plus prudents, ou connaissent mieux les possibilités de leur voiture que les propriétaires de nouvelles voitures. Les propriétaires de voitures plus anciennes

semblent moins soucieux d'éviter les accidents. Peut-être comprennent-ils un bon nombre de débutants. Les défauts des voitures anciennes peuvent également expliquer cette fréquence plus élevée d'accidents. On enregistre des variations sensibles du risque d'accident d'une marque à l'autre, mais aucun coefficient n'est statistiquement significatif. Notons enfin que le terme constant est égal à $-0,23$. Il donne la valeur de la fonction discriminante lorsque toutes les variables linéaires ont valeur zéro, c'est-à-dire lorsque l'observation correspond aux catégories de référence.

La fonction discriminante incluant toutes ces variables se révèle significative. Le test de l'égalité des moyennes dans les deux groupes, Λ de Wilks, donne une valeur de F égale à $3,65$ ($F[0,95] = 1,46$). Par contre, l'hypothèse d'égalité des matrices de variances et covariances est rejetée par le test de Box qui produit un $F = 1,48$, pour un seuil critique $F(0,95) = 1^{13}$. Il faut noter cependant que ce test, aussi bien que les précédents, est basé sur l'hypothèse d'une distribution normale de X . La portée du test sur les variances et covariances est donc assez douteuse. En ce qui concerne les tests sur les coefficients, ils ne peuvent être interprétés de façon classique, mais comme une mesure de l'accroissement du D^2 que la variable considérée contribue. Pour cette interprétation, il est utile de se rappeler que D^2 est une mesure descriptive de la distance entre les moyennes des éléments de X des deux populations.

Nous avons calculé la fonction discriminante qui ne reprenait que les huit variables significatives. Ensuite, nous l'avons recalculée en excluant la variable

Tableau 8

*Fonction discriminante avec 7 variables explicatives*¹⁴

Etranger	-0,84
Moins de 23 ans	-1,09
Entre 23 et 25 ans	-0,88
Entre 26 et 30 ans	-0,70
Plus de 1599 cm ³	-0,64
Sans profession	-0,60
Véhicule de 2 à 3 ans d'âge	0,37
Terme constant	0,49

¹³ Box G. E. P., «A General Distribution Theory for a Class of Likelihood Criteria», *Bimetrika* 36, 1949, p. 317-346.

¹⁴ Les variables sont placées dans ce tableau selon leur ordre d'introduction dans l'analyse discriminante par étapes (stepwise).

«assurance souscrite en 1964–65», car nous voulions centrer notre intérêt sur la sélection des nouveaux assurés. Les résultats de cette fonction sont donnés dans le tableau 8.

La fonction et tous ses coefficients sont significatifs. L'hypothèse d'égalité des variances et covariances est encore rejetée avec une valeur de F égale à 6,01 ($F[0,95] \simeq 1,46$).

Lorsque cette fonction réduite est utilisée pour classifier les observations de l'échantillon de calcul et celles de l'échantillon-test, on obtient les résultats donnés dans le tableau 9.

Tableau 9

Classification par la fonction linéaire à sept variables suivant le critère $\ln q_2/q_1$

Echantillon de calcul

Situation réelle	Classification					
	Non-responsables d'un accident		Responsables d'un accident			
Non-responsables	789		50		839	
		94,0%		6%		67,6%
Responsables	339		64		403	
		84,1%		15,9%		32,4%
Pourcentage de classements corrects: 68,7%						

Echantillon de test

Situation réelle	Classification					
	Non-responsables d'un accident		Responsables d'un accident			
Non-responsables	86		5		91	
		95,5%		4,5%		65,9%
Responsables	37		10		47	
		78,7%		21,3%		34,1%
Pourcentage de classements corrects: 69,6%						

Cette classification est faite en ne retenant comme critère que le premier terme sur la droite de la règle de décision (6), c'est-à-dire $\ln q_2/q_1$. Elle ne tient donc pas compte de l'importance relative des profits rapportés par les bons conducteurs et par ceux qui causent un accident. On voit que les bons conducteurs sont bien repérés par la fonction (94,0% pour l'échantillon de calcul), tandis que les responsables d'accident le sont assez mal. Cependant, le pourcentage moyen de classement correct est assez élevé (68,7% pour l'échantillon de calcul). Les résultats sont encore meilleurs lorsque la fonction est appliquée à l'échantillon de test, ce qui dénote une sélection efficace.

Un test en χ^2 , qui n'est pas basé sur l'hypothèse d'une distribution normale, nous permet d'évaluer d'une autre façon la qualité de la classification¹⁵. Pour l'échantillon de calcul, on obtient un χ^2 de 173,3 qui rejette l'hypothèse que la classification aurait pu résulter d'une sélection au hasard ($\chi^2_1 [0,999] = 10,83$). Pour l'échantillon-test de taille plus réduite, la valeur du χ^2 calculé doit normalement être plus petite. Il est égal à 21,1 et rejette également l'hypothèse de sélection au hasard.

Les résultats de la classification opérée par la fonction incluant les trente-cinq variables sont à peine différents. Notons seulement que le pourcentage de classements corrects s'élève à 69,2% pour l'échantillon de calcul et à 67,4% pour l'échantillon-test. Il ne semble donc pas utile de compliquer l'analyse et les calculs en retenant toutes les variables. De même, il y a peu de différence entre les classifications opérées par la fonction à huit variables et celle à sept variables qui exclut la variable «assurance souscrite en 1964–65». En fait, le rôle de cette dernière est repris par les variables de catégories d'âge qui lui sont fortement corrélées. Nous ne donnerons pas ici les résultats de ces autres fonctions et limiterons l'application à la fonction à sept variables.

Il nous est possible maintenant de calculer le profit qu'aurait fait la société d'assurance si elle avait utilisé la fonction discriminante réduite pour opérer une sélection parmi ses assurés actuels. Nous prendrons, comme profit rapporté par un assuré qui ne cause pas d'accident, 4340 francs, la prime moyenne de l'échantillon, et comme perte pour un assuré qui cause un accident 15 292 francs, le coût moyen moins la prime moyenne. En calculant sur ces bases le profit de la société, celle-ci a fait une perte de 2 845 200 francs durant l'année considérée. Si elle avait sélectionné ses assurés en appliquant la fonction réduite, elle aurait ramené sa perte à 1 937 000 francs. Ce résultat peut être amélioré si l'on tient compte de l'importance relative des profits rapportés par les bons

¹⁵ Voir *S. J. Press*, «Applied Multivariate Analysis», Holt, Rinehart et Winston, 1972, p. 382.

conducteurs et par ceux qui provoquent un accident. C'est d'ailleurs ce que propose la règle de décision (6).

C. L'influence des coûts sur la sélection des contrats

Deux remarques préliminaires doivent être faites au sujet de l'application de la règle de décision (6). En premier lieu, une régression des coûts d'accidents, comme variable dépendante, sur les caractéristiques des conducteurs n'a indiqué aucune relation statistique entre ces variables ($R^2 = 0,03$). En conséquence, nous avons renoncé à considérer \tilde{C} comme une fonction de X , ainsi que nous l'avons fait dans le modèle théorique. \tilde{C} sera donc une constante, le coût moyen de l'accident dans l'échantillon 19 632 francs (voir ci-dessus). En second lieu, c'est la prime moyenne, 4340 francs, qui sera prise en compte et non pas la prime effectivement payée par chaque assuré. Les résultats de cette procédure sont donnés dans le tableau 10.

Cette règle de sélection est beaucoup plus restrictive et peu d'assurés sont retenus dans le portefeuille, ce qui traduit l'influence des coûts d'accidents relativement élevés par rapport aux primes. Si cette sélection avait été appliquée par la compagnie, le profit total, calculé sur les mêmes bases qu'auparavant, aurait été de 174 800 francs.

D. La fonction discriminante quadratique

Puisque, de fait, les tests d'égalité des matrices de variances et covariances n'étaient pas concluants, nous avons calculé la fonction discriminante quadratique pour les sept variables retenues précédemment. Il serait trop long d'en détailler les coefficients, et nous nous bornerons ici à mentionner que les résultats de classification, si l'on tient compte des profits relatifs, sont exactement les mêmes que pour la fonction linéaire, c'est-à-dire les résultats donnés au tableau 10.

E. Les primes d'assurance minimales

Il nous reste à calculer les primes qui devraient être appliquées si l'on désirait tenir compte du risque particulier aux diverses catégories d'assurés. Etant donné les sept variables retenues, il y a 64 catégories différentes. Pour chacune de ces

Tableau 10

Classification par la fonction linéaire à sept variables suivant le critère (5)

Echantillon de calcul

Situation réelle	Classification					
	Non-responsables d'un accident		Responsables d'un accident			
Non-responsables	125		714		839	
		14,9%		85,1%		67,6%
Responsables	28		375		403	
		6,9%		93,1%		32,4%
Pourcentage de classement corrects: 40,0%						

Echantillon de test

Situation réelle	Classification					
	Non-responsables d'un accident		Responsables d'un accident			
Non-responsables	21		70		91	
		23,1%		76,9%		65,9%
Responsables	2		45		47	
		4,3%		95,7%		34,1%
Pourcentage de classements corrects: 48,0%						

catégories, les tableaux 11 et 12 donnent la prime d'assurance minimale telle qu'elle est définie par l'équation (4). Ce sont donc les primes qui procureraient, en moyenne, un profit nul à la compagnie d'assurance. Elles n'incluent pas la commission de l'intermédiaire ni la taxe, et sont basées sur le coût moyen par accident de la compagnie (17 282 francs). Il faut noter cependant que leur calcul sur base du coût moyen national (16 452 francs) n'altérerait pas substantiellement les conclusions qu'on en peut tirer. Pour procurer à la compagnie un profit moyen positif quelconque $\bar{\Pi}$, il suffirait de majorer ces primes de $\bar{\Pi}$.

Tableau 11

Primes (en francs) calculées selon la formule $R = [1-P(X)]\tilde{C}$.

		Belge			
		2 à 3 ans		≠ 2 à 3 ans	
		< 1600	≥ 1600	< 1600	≥ 1600
< 23	Profession	7 440	10 542	9 207	12 290
	Sans profession	10 366	13 330	12 132	14 802
23-25	Profession	6 518	9 521*	8 206	11 308
	Sans profession	9 345	12 427	11 151	14 017
26-30	Profession	5 732*	8 618*	7 342	10 424
	Sans profession	8 461	11 563	10 248	13 232
> 30	Profession	3 337**	5 497**	4 496*	7 067*
	Sans profession	5 359*	8 167*	6 910	9 953*

Les primes marquées** sont inférieures aux primes du barème; celles marquées* ne sont inférieures que pour les grosses cylindrées de la catégorie. A noter que 5359 francs est la prime de barème pour une voiture de 1185 cm³, 7067 francs celle pour 2335 cm³ et 8167 francs celle pour 3222 cm³.

Tableau 12

Primes (en francs) calculées selon la formule $R = [1-P(X)]\tilde{C}$.

		Etranger			
		2 à 3 ans		≠ 2 à 3 ans	
		< 1600	≥ 1600	< 1600	≥ 1600
< 23	Profession	11 504	14 312	13 193	15 607
	Sans profession	14 174	16 314	15 490	17 217
23-25	Profession	10 503	13 467	12 270	14 901
	Sans profession	13 310	15 705	14 783	16 726
26-30	Profession	9 620	12 663	11 406	14 213
	Sans profession	12 505	15 097	14 076	16 255
> 30	Profession	6 341	9 305*	8 010	11 112
	Sans profession	9 148	12 231	10 955	13 840

Nous ne décrivons pas ici les mécanismes de ce barème qui prévoit un système de personnalisation a posteriori, dit bonus-malus. Il nous suffit de préciser ici que la comparaison fut faite sur base du degré 10 de l'échelle bonus-malus pour un contrat «responsabilité civile» de dix ans sans appliquer de réduction pour usage limité de la voiture (15%) ni pour une franchise éventuelle, mais en soustrayant la commission de l'intermédiaire (19%).

Cette comparaison des primes calculées sur base du risque estimé de la catégorie et des primes de barème, est une façon d'opérer une sélection parmi les assurés potentiels selon la règle de décision (6). Elle est plus précise que celle employée ci-dessus pour les assurés actuels qui ne se réfèrent qu'à la prime moyenne. Le tableau 11 indique que seules deux catégories seraient acceptées sans restrictions. La première est précisément celle qui avait été retenue par l'analyse discriminante proprement dite et pour laquelle la prime serait de 3337 francs, tandis que la seconde est la catégorie pour laquelle la prime serait de 5497 francs. Tous les conducteurs de ces deux catégories devraient être acceptés par la compagnie si le barème 1971, degré 10, leur était appliqué. D'autres conducteurs pourraient encore être acceptés si leur voiture était d'une puissance suffisamment élevée. En effet, dans ce cas, leur prime pourrait être telle qu'il serait avantageux pour la compagnie de les assurer. Ce dernier résultat est cependant sujet à caution, car il pourrait bien être la conséquence de l'utilisation de catégories de cylindrées en lieu et place d'une variable continue.

Ces résultats sont évidemment fort utiles pour orienter la politique de prospection d'une compagnie d'assurance. Pour ne citer qu'un exemple, les preneurs d'assurance tous risques (omnium) se révèlent à tout le moins aussi bons conducteurs que les autres assurés alors que les indemnités qui leur sont payées n'apparaissent pas être plus élevées en moyenne que celles payées aux autres assurés. Cependant, ils doivent payer une prime plus élevée, dans le présent système. Ce segment de marché semble donc particulièrement intéressant pour les compagnies d'assurances.

L'intérêt du calcul des primes réside également en ce qu'il nous permet d'évaluer la structure actuelle du tarif de l'assurance automobile. Le système de bonus-malus qui ajuste a posteriori la prime d'assurance au risque permet une diversification progressive des primes d'assurances. Cependant, la différence maximale entre primes lorsque le système bonus-malus a opéré pleinement n'est que celle de 1 à 3,33 pour une cylindrée donnée. Pour les primes estimées par notre méthode, cette différence est de 1 à 4,6 dans le cas des voitures de puissance inférieure à 1600 cm³ tandis qu'elle est de 1 à 3,13 pour les puissances plus

élevées. On peut donc se demander si la diversification prévue est suffisante, d'autant plus que, pour chaque année sans accident, la prime ne diminue que de 5%. Il faut ainsi cinq ans à un bon conducteur pour arriver à la prime la plus basse (s'il débute au degré 6 pour un contrat d'usage réduit de voiture) et il faut six ans pour un mauvais conducteur, qui aurait un accident chaque année, pour en arriver à payer la prime maximale. Il serait sans doute souhaitable que l'adaptation a posteriori des primes d'assurances soit plus rapide qu'elle n'est prévue, afin d'éviter que les conducteurs prudents ne paient pour les mauvais conducteurs. Par contre, il convient d'observer que le système actuel différencie davantage les primes selon la cylindrée. D'une voiture de 395 cm³ à une de 5000 cm³, la prime (sur base 10) varie de 1 à 2,75 tandis que d'une 425 cm³ à une 3000 cm³, ce qui est un intervalle correspondant mieux à la nature de notre échantillon, la prime varie de 1 à 2. Notre analyse statistique n'accorde pas une telle importance à la puissance puisque la prime calculée varie approximativement de 1 à 1,6.

On peut se demander encore, sur base des données analysées, si le niveau moyen des primes d'assurance automobile est suffisamment élevé. En effet, à supposer que les compagnies d'assurance puissent sélectionner systématiquement leurs assurés, nous ne pourrions que leur recommander de refuser, au degré 10 du barème, la majorité de leurs clients potentiels. Ceci suggère que le secteur de l'assurance automobile pourrait bien être subsidié par les autres secteurs d'assurance. S'il en était ainsi, un problème existerait du point de vue de la rationalité économique et de la répartition des revenus: serait-il socialement souhaitable que les contrats d'assurance-vie ou d'assurance-incendie, par exemple, subsidient l'assurance automobile?

Zusammenfassung

Vorschlag, mit Hilfe der statistischen Methoden der Diskriminanz-Analyse Risikoklassen für Autohaftpflicht-Policen zu definieren und für diese Tarifprämien zu berechnen.

Résumé

Proposition de définir les classes de risques, pour les polices R.C. Automobiles, à l'aide des méthodes statistiques de l'analyse discriminante et de calculer les primes tarifaires afférentes.

Riassunto

Proposte per definire le classi di rischio per polizze R.C. Automobili con l'aiuto di metodi statistici di analisi discriminanti e per calcolare i premi di tariffa corrispondenti.

Summary

Proposition to define risk classes for Motor Liability policies with the help of the statistical methods of the discriminant analysis and to calculate the appertaining tariff premiums.