

## CONTRIBUTION TO THE IMPROVEMENT OF THE FRAUD DETECTION IN THE CAR INSURANCE

### CONTRIBUTION A L'AMELIORATION DE LA DETECTION DE FRAUDE A L'ASSURANCE AUTOMOBILE

**Saïd Mohand MEHAH\***

*Ecole Nationale Supérieure Statistique et d'économie appliquée ENNSEA  
Laboratoire de Statistique Appliquée LASAP  
[saidoumebah@gmail.com](mailto:saidoumebah@gmail.com)*

**Nora MOSBAH**

*Ecole Nationale Supérieure Statistique et d'économie appliquée ENNSEA  
Laboratoire de Statistique Appliquée LASAP  
[noralounici@yahoo.fr](mailto:noralounici@yahoo.fr)*

**Reçu le 2018-02-12 Accepté le 2018-03-07 Publié en ligne le 2018-06-01**

**ABSTRACT:** The aim of this research is to develop a statistical model based on the linear discriminant analysis (LDA) method, which allows the classification of suspicious claims into two classes "fraudulent claims" and "non-fraudulent claims" then to provide in which class a new claim will be included. The aims of this research consist to study the claim folder transmitted by SAA to ALFA, in order to detect the clues related to the act of fraud. So, the target of this research is to study the clues related to the fraudulent act of the insured. The model developed, gives us a good ranking rate exceeding 80%, it allowed us, also, to note that 5 of the 12 indicators used were significant to predict the probability of fraud.

**Key words:** Fraud, Insurance, Asymmetric Information (moral hazard, adverse selection), linear discriminant analysis (LDA).

**RÉSUMÉ :** Le but de cette recherche est d'élaborer un modèle statistique basé sur la méthode de l'analyse discriminante linéaire (ADL), nous permettant de classer les dossiers douteux en deux classes "dossiers frauduleux" ou "dossiers non frauduleux" et de prévoir à quelle classe un nouveau dossier sera affecté. Afin d'atteindre l'objectif fixé, nous avons étudié les dossiers transmis par la SAA vers ALFA, afin de détecter les indices liés à l'acte de fraude. Ceci dit, finalité de ce travail consiste à étudier les indices liés à l'acte frauduleux de l'assuré. Le modèle élaboré, nous donne un taux de bon classement dépassant les 80%, cela nous a permis, aussi, de constater que 5 des 12 indicateurs utilisés étaient significatifs pour prédire la probabilité de fraude.

**Mots clés :** Fraude, Assurance, Asymétrie d'information (Aléa moral, Anti-sélection), Analyse discriminante linéaire (ADL).

---

\* Auteur correspondant

## 1. INTRODUCTION

La fraude à l'assurance peut être définie comme étant un acte volontaire commis par l'assuré en vue de tirer de son contrat d'assurance un profit illégitime. Cependant, pour qu'il y ait fraude, il faut que l'assuré soit de mauvaise foi.

La fraude s'est, de plus en plus, répandue dans le domaine des assurances. Son fait générateur c'est l'asymétrie d'information (aléa moral<sup>1</sup> et antisélection<sup>2</sup>).

De ce fait, la fraude à l'assurance peut se présenter sous différentes formes :

- **La fraude à la souscription** : l'assuré fournit, sciemment (de mauvaise foi), à l'assureur des circonstances inexactes du risque et ce, dans le but de fausser l'appréciation de l'assureur. Cette fraude consiste à dissimuler ou falsifier la nature du risque, pour bénéficier des primes moins élevées.

- **La fraude en cours de contrat** : le non-respect des termes de contrat ou la création volontaire des dommages, où l'assuré peut réclamer une indemnité suite à un sinistre, dont il a provoqué la réalisation.

- **La fraude à la déclaration de sinistre** : consiste à déclarer un faux sinistre, ou tout simplement à gonfler le montant des dommages.

Touchant plusieurs pays et concernant la plupart des catégories d'assurance, le phénomène de la fraude revêt divers modes opératoires coûtant, ainsi, chaque année aux assureurs des pertes colossales. Les dossiers de fraude à l'assurance automobile représentent une importante proportion, où à l'occasion de chaque remboursement, il y a probabilité (possibilité) de risque de fraude.

Pour la compagnie d'assurance, la lutte contre la fraude est un enjeu de taille, il se résume à limiter efficacement les opportunités de fraude sans restreindre son marché et ses services, ni pénaliser ses bons clients. Cela dit, les stratégies des assureurs pour dissuader la fraude à l'assurance ont leur importance. Dionne et al (2009), par exemple, estiment que, les entreprises pourraient économiser jusqu'à 41% des coûts en raison de demandes frauduleuses, par la mise en œuvre de la stratégie efficace de lutte anti-fraude.

Néanmoins, il ne s'agit pas de voir dans tout assuré un fraudeur et d'instaurer un climat de suspicion généralisé. À l'inverse il convient de ne pas attendre que les circonstances et le hasard fassent éclater l'évidence d'une fraude. A ce propos, en créant des modèles à partir de l'historique des demandes d'indemnisation, on peut repérer avec une certaine exactitude les demandes frauduleuses au milieu des milliers qui parviennent chaque année.

Dans une première étape, nous proposons une procédure afin d'isoler les indicateurs les plus significatifs de manière à prédire la probabilité que la demande d'indemnisation présente un risque de fraude. Dans une seconde étape, nous élaborons un modèle statistique

---

<sup>1</sup> La question qui se pose au nom de l'aléa moral est celle de savoir si l'assuré prendra autant de précautions après l'assurance. Or, on constate globalement que trop d'assurance favorise la perte de précautions.

<sup>2</sup> Les dysfonctionnements des marchés d'assurances qui résultent de l'information cachée, dont les assurés peuvent disposer sur leurs propres risques et qui n'est pas accessible aux assureurs. Ceci constitue la source du problème d'antisélection : la coexistence de "bons" risques et de "mauvais" risques indiscernables à priori.

pour calculer les probabilités de fraude des dossiers sinistres<sup>3</sup> transmis de la SAA vers ALFA<sup>4</sup>, et décider ainsi de l'opportunité d'effectuer ou non une enquête approfondie.

Il faut, néanmoins, noter que le but de cet article n'est pas de bâtir une stratégie de lutte contre la fraude. Il s'agit d'élaborer un modèle permettant de détecter la fraude lors de règlement de sinistre. Nous dégageons les paramètres qui permettent d'extraire les indicateurs les plus significatifs permettant de prédire la probabilité qu'un dossier contienne de la fraude ou non, par la méthode ADL.

## 2. DONNEES ET METHODOLOGIE

Les données choisies pour mener cette étude concernent des dossiers d'ALFA (enquête sur dossier). L'objectif recherché par cette analyse étant de relever les indices de fraudes les plus répétitifs.

De plus, afin de préparer la procédure de la détection de fraude, nous avons opté pour la démarche suivante :

- ✓ Analyse des déclarations des assurés par un expert afin d'identifier les cas de fraude ;
- ✓ Extraction des caractéristiques à partir de ces déclarations (type d'accident, de blessures, etc...)
- ✓ Applications de la méthode ADL pour identifier les caractéristiques des déclarations qui sont fortement corrélées à la fraude.

Ainsi, lorsqu'une déclaration de sinistre est déposée auprès d'une compagnie d'assurance, l'assuré doit préciser les détails spécifiques concernant la perte. Ces détails comprennent, entre autre, *le type du dommage, la date du sinistre, les points de choc, les circonstances de l'accident...* etc.

Dans notre analyse, nous considérons les informations du rapport initial de la perte ainsi que les « anomalies » relevées par les experts. Ces informations recueillies permettent aux assureurs de savoir si la demande a le potentiel d'avoir un élément de fraude.

Dans ce qui suit, nous exposons d'abord les données de notre échantillon, ensuite, nous présentons une description statistique de l'étude. Enfin, nous abordons quelques aspects sur l'analyse discriminante et l'élaboration du modèle.

### 2.1. Présentation des données de l'échantillon

Vu l'importance de l'étude de dossier sinistre pour la détection de la fraude, la question qui se pose est la suivante : Parmi les indices de fraude, quels sont ceux qui déterminent la matérialité ou la condamnation à la fraude d'un dossier ?

Pour ce faire, nous avons analysé les dossiers de la SAA transmis vers ALFA. Cette enquête est réalisée en collaboration avec le directeur d'ALFA, que ce soit lors de la phase du choix des indices ou lors de la collecte des informations.

---

<sup>3</sup> Dossiers sinistres soupçonnés de fraude

<sup>4</sup> Agence de Lutte contre la Fraude à l'Assurance.

L'échantillon comprend 209 dossiers de sinistres, la collecte des données de l'échantillon a duré une année, du 01/01/2013 au 31/12/2013.

Les indices ont été retenus après la consultation de la littérature dans le domaine, ainsi que les différents entretiens que nous avons eus avec les professionnels de l'assurance, entre autres, le directeur d'ALFA, les différents directeurs de la SAA et les enquêteurs d'ALFA.

Dans cet article, on s'intéresse uniquement à la fraude à l'assurance automobile lors de la déclaration de sinistre, elles sont exclues de notre périmètre d'étude la fraude à la souscription du contrat et la fraude en cours du contrat.

## **2.2. Présentation et description des variables**

Les données dont nous disposons comportent, des informations sur les dossiers douteux relatives au contrat souscrit et à la déclaration du sinistre, ainsi qu'à l'acte de fraude lui-même.

### **2.2.1. Les variables relatives au contrat d'assurance**

- *le véhicule est gagé* : il s'agit de vérifier si le véhicule est lié à un contrat ANSEJ, crédit-bail, location ou autres organismes de crédit.

- *la date d'échéance ou de souscription* : il s'agit de vérifier si le sinistre est intervenu peu de temps après la souscription du contrat ou bien peu de temps avant la date d'échéance, ainsi, la souscription, ou la modification, peut être faite dans le seul but de monter un faux sinistre avec un véhicule déjà accidenté.

### **1.2.2. Les variables relatives à la déclaration du sinistre**

- *Incompatibilité des dommages sur les deux véhicules* : les dommages sont disproportionnés ou incompatibles sur les deux véhicules. Cette discordance reflète souvent un sinistre simulé. Dans ce cas, il faudrait demander à l'expert de vérifier la compatibilité des dommages subis par chacun des véhicules.

- *Flou et contradiction sur les circonstances du sinistre* : les explications sur les circonstances du sinistre sont peu claires, contradictoires ou trop préparées.

- *Les points de chocs en contradiction* : les points de chocs sont en contradiction avec les déclarations de l'assuré.

- *Préexistence d'un sinistre non réparé* : le véhicule a été sinistré précédemment et les dommages n'ont pas été réparés.

- *Déclaration tardive ou le déclarant est différent du souscripteur* : cela peut aussi permettre de découvrir une dénaturation du risque quant au conducteur habituel. L'assuré peut avoir tardé à faire la déclaration à son assureur dans le seul but de présenter des documents falsifiés.

- *Les parties sont de la même famille ou habitent le même quartier* : cet indice permet de détecter toute mise en scène de sinistre par l'assuré et ses complices.

- *Le dossier est interne ou bien inter-compagnies* : certains assurés profitent du manque d'échange d'informations entre les compagnies d'assurance pour frauder.

- *Le montant des dommages* : cet indice est relatif au gonflement des factures de réparation ou des montants des dommages subis par le véhicule.

### 1.2.3. Les variables relatives à l'acte de fraude

-L'assuré a agi seul ou bien avec des complices : il est question de voir si certaines formes de fraude nécessitent des complices ou non.

-La forme de fraude : il y a plusieurs formes de fraude, mais on s'est contenté de mentionner 3 seulement (le nombre des autres formes de fraude est insignifiant dans notre étude). Il s'agit, bien évidemment, de la fraude à la déclaration de sinistre : simulation du sinistre ou provocation de sinistre, falsification des factures, description du sinistre de manière inexacte.

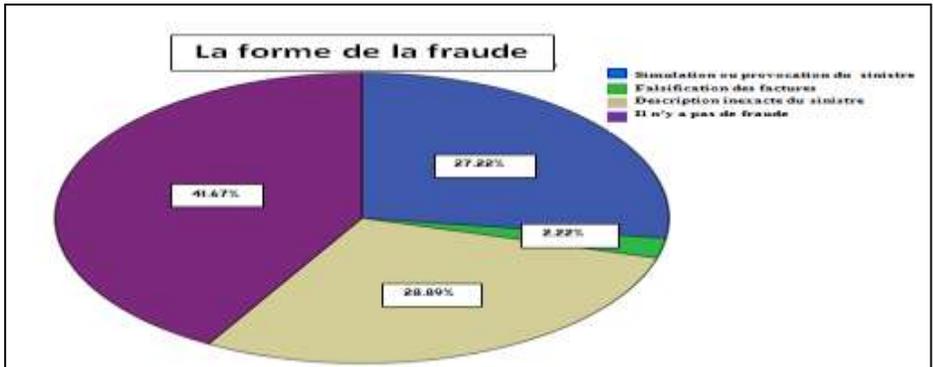
## 2. Approche statistique descriptive de l'échantillon

Après la présentation des données de l'échantillon et la description des variables, on verra comment subdiviser les fréquences de proportion des variables.

### 2.1. Les formes de fraude de l'échantillon

On constate que les sinistres décrits de manière inexacte représentent la forme de fraude la plus répandue (52 cas) soit près de 29%, suivie de la simulation ou provocation du sinistre avec 49 cas sur les 180 traités (27,2%). Ce qui prouve que, pour les fraudeurs, il est plus facile de simuler un sinistre que de se procurer des factures falsifiées. En revanche pour la description inexacte de sinistre, il peut y avoir deux possibilités, soit que l'assuré est sous le choc de l'accident, ou bien qu'il est fautif et dans l'acte de fraude il compte dissimuler sa faute.

**Figure 1** : Répartition de l'échantillon selon la forme de fraude



Elaboré par nos soins sous SPSS 22

### 2.2. La fraude selon les garanties souscrites

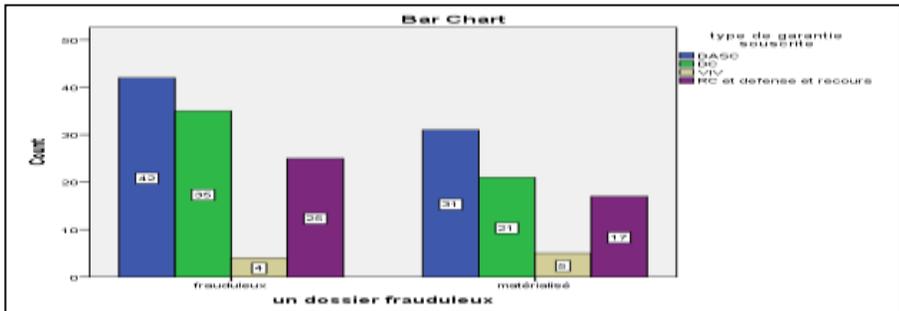
D'après la Figure 2, sur un total de 41% des dossiers douteux " de la garantie DASC<sup>5</sup>" 23% ont été jugés frauduleux par les enquêteurs d'ALFA. La raison avancée est que les assurés pensent qu'ils ont tout à gagner en simulant un sinistre puisque, dans tous les cas, ils seront remboursés. En ce qui concerne la garantie DC<sup>6</sup> 19% des cas frauduleux,

<sup>5</sup> Dommages avec ou sans collisions ; c'est la garantie tous risques

<sup>6</sup> Dommages et collisions

les assurés recourent à la fraude afin de couvrir les risques exclus de la garantie, idem pour la RC et défense et recours.

Figure 2 : Répartition de l'échantillon selon la garantie souscrite



Elaboré par nos soins sous SPSS 22

### 3. L'élaboration d'un modèle à l'aide de l'ADL

Étant donné un ensemble de variables indépendantes, l'analyse discriminante tente de trouver des combinaisons linéaires de ces variables qui permettent de mieux distinguer les différents groupes de cas. Ces combinaisons sont appelées fonctions discriminantes. D'après (Saporta G.2006, p 439), On distingue deux aspects en analyse discriminante<sup>7</sup> :

- **Le premier descriptif** : Cherche quelles sont les combinaisons linéaires de variables qui permettent de séparer le mieux possible les  $k$  catégories et donner une représentation graphique qui rende compte de cette séparation.

- **Le second décisionnel** : Un nouvel individu, dont on connaît les caractéristiques, se présente à l'assurance. Aidé du modèle, il s'agit alors de décider dans quelle catégorie il faut l'affecter. C'est un problème de classement et non de classification<sup>8</sup>.

#### 3.1. Le critère de regroupement qualitatif

La variable cible, dans notre étude est la variable «dossier douteux» qui sera considérée comme critère de regroupement qualitatif. Cette variable est constituée de 2 classes : «frauduleux» et «matérialisé».

#### 3.2. Les variables indépendantes

Les variables indépendantes expliquent la fonction score, elles sont au nombre de douze : le véhicule est gagé, la date d'échéance ou de souscription, incompatibilité des dommages sur les deux véhicules, flou et contradiction sur les circonstances du sinistre, les points de chocs en contradiction, préexistence d'un sinistre non réparé, déclaration tardive ou le déclarant est différent du souscripteur, les parties sont de la même famille ou habitent le même quartier, le dossier est interne ou bien inter-compagnies, le montant des dommages.

<sup>7</sup> SADI Khadîdja,(2009). Élaboration d'un modèle d'évaluation du risque de crédit d'exploitation à l'aide des réseaux de neurones artificiel (RNA) et de l'analyse discriminante linéaire (ADL) thèse de doctorat en science de gestion, option : management.

<sup>8</sup>Dans le cas de classement, les classes sont connues et le problème devient un problème d'affectation des individus. Alors que le but de la classification est de construire une ou plusieurs partitions (classes). Les classes formées doivent être le plus homogènes possible.

### 3.3. Répartition de l'échantillon

L'échantillon global doit être divisé en deux : échantillon de construction (180 dossiers) et l'échantillon de validation ou test (29).

### 3.4. La fonction score

D'après de Table ci-dessous, l'examen du F montre que cinq variables sont discriminantes : Incompatibles des dommages sur deux véhicules, flou et contradiction sur les circonstances du sinistre, points de chocs en contradiction et Préexistence d'un sinistre non réparé. De plus, selon le test du Lambda de Wilks<sup>9</sup>, seules les variables "flou et contradiction sur les circonstances du sinistre" et " incompatibles des dommages sur les deux véhicules" semblent avoir une influence.

**Table 1:**Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
incompatibilité des dommages sur les deux véhicules	,812	41,110	1	178	,000
flou et contradiction sur les circonstances du sinistre	,722	68,444	1	178	,000
Les points de chocs en contradiction	,859	29,130	1	178	,000
Date de souscription ou d'échéance est proche	,987	2,263	1	178	,134
Préexistence d'un sinistre non réparé	,933	12,850	1	178	,000
une déclaration tardive ou le déclarant est différent du souscripteur	,997	,450	1	178	,503
Les parties sont de la même famille ou habitent le même quartier	,967	6,083	1	178	,015
le véhicule est gagé	,999	,144	1	178	,705
type de garantie souscrite	1,000	,001	1	178	,977
le dossier	,890	21,895	1	178	,000
le montant des dommages	,989	2,047	1	178	,154
la délégation régionale	,998	,313	1	178	,577

### 3.5. Vérification de la validité de l'étude

On estime la validité d'une analyse discriminante à partir des indicateurs :

- ✓ Le test de Box.
- ✓ La Lambda de Wilks

**Table 2 :** Test Results

Box's M	126,157
F Approx.	8,146
df1	15
df2	98714,907
Sig.	,000

Tests null hypothesis of equal population covariance matrices.

Pour que le modèle soit valide, il faut que le M soit le plus élevé possible et que la significativité du test de F tende vers 0, si elle est supérieure à 0,05, l'analyse n'est pas valide. Ainsi, on observe sur le Table 2 que la sig de F= 0, et le M =126.157 ce qui nous amène à conclure que cette analyse est valide.

<sup>9</sup> Le lambda de Wilks univarié correspond au rapport de la somme des carrés des écarts intraclasse par la somme totale des carrés des écarts.

**Table 3 : Wilks' Lambda**

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	,572	98,110	5	,000

Sur le Table 3 la significativité de Wilks' Lambda plus elle est tend vers 0 et plus le modèle est bon.

### 3.6. Estimation des coefficients de la fonction discriminante

**Table 4 : Canonical Discriminant Function Coefficients**

	Function
	1
X1 : incompatibles des dommages sur les deux véhicules	,923
X2 : flou et contradiction des sur les circonstances du sinistre	1,669
X3 : Préexistence d'un sinistre non réparé	1,125
X4 : dossier	-,738
X5 : montant des dommages	-,415
(Constant)	-4,029

Unstandardized coefficients

Les variables les plus significatives sont indiquées dans le Table ci-dessus.

La fonction discriminante est la suivante :

$$Z = 0.923 * X1 + 1.669 * X2 + 1.125 * X3 - 0.738 * X4 -$$

### 3.7. Evaluation du modèle

**Table 5 : classification de l'échantillon de construction**

Classification Results <sup>a,c</sup>					
		dossier douteux	Predicted Group Membership		Total
			Frauduleux	Matérialisé	
Original	Count	Frauduleux	93	13	106
		Matérialisé	22	52	74
	%	Frauduleux	87,7	12,3	100,0
		Matérialisé	29,7	70,3	100,0
Cross-validated <sup>b</sup>	Count	Frauduleux	93	13	106
		Matérialisé	22	52	74
	%	Frauduleux	87,7	12,3	100,0
		Matérialisé	29,7	70,3	100,0

a. 80,6% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 80,6% of cross-validated grouped cases correctly classified.

Les résultats obtenus de l'analyse discriminante sont satisfaisants, du fait que les taux de bon classement pour l'échantillon est de **80,6%**, il dépasse les **50%**. Il est de **87,8%** dans la classe de «dossiers frauduleux».

Néanmoins, il y a lieu de signifier que l'analyse n'a retenu que cinq variables citées précédemment. Par conséquent, les autres variables apportent peu d'informations au modèle.

- Sur les 106 dossiers classés à priori dans la classe "frauduleux" :
  - ✓ 93 ont été correctement classés par le modèle, soit un taux de bon classement de 87,7%.
  - ✓ 13 sont mal classés, soit un taux d'erreur de 12,3%.
- Sur les 74 dossiers classés a priori dans la classe "matérialisé" :
  - ✓ 52 sont bien classés, soit un taux de bon classement de 70,3%
  - ✓ 22 sont mal classés, soit un taux d'erreur de 29,7%

On constate que la classe des dossiers douteux jugés frauduleux est plus homogène que la classe des dossiers non frauduleux.

### 3.7.1. Test et validation

Cette fonction discriminante semble performante, les deux classes ont de bons taux de classement. Comme tout travail statistique, il reste à élaborer la dernière étape, qui est celle de la validation du modèle.

**Table 6** :Functions at Group Centroids

	Function
un dossier douteux	1
Frauduleux	-,719
Matérialisé	1,030

Unstandardized canonical discriminant functions evaluated at group means

D'après les valeurs de la fonction discriminante estimée au barycentre de deux groupes, la projection du « dossier douteux » sur la fonction linéaire discriminante le classe parmi les dossiers matérialisés « non frauduleux ».

Pour aboutir au diagnostic du classement l'échantillon de classement nous allons utiliser la méthode de la distance géométrique<sup>10</sup> afin d'affecter l'individu à la classe correspondante. La méthode se résume dans l'algorithme suivant :

<sup>10</sup> SADI Khadidja, (2009). Élaboration d'une modèle d'évaluation du risque de crédit d'exploitation à l'aide des réseaux de neurones artificiel (RNA) et de l'analyse discriminante linéaire (ADL). Thèse de doctorat en science de gestion, Option : management.

1. Calculer la distance de l'individu  $i$  à chaque classe à l'aide de la fonction au Barycentre :

$$D_{i1} = |Z_i - f_1| \quad \text{ou} \quad \left. \begin{array}{l} Z_i \text{ est le score de l'individu } i. \\ f_j \text{ est la fonction au barycentre de la classe } j. \\ D_{ij} \text{ est la distance de l'individu } i \text{ et la classe } j. \end{array} \right\}$$

$$D_{i2} = |Z_i - f_2|$$

2. Calculer le minimum des distances :  $Min = \text{minimum}(D_{i1}, D_{i2})$ .

3. Si  $Min = D_{i1}$  Alors l'individu  $i$  est affecté à la classe 1

Sinon l'individu  $i$  est affecté à la classe 2

Fin si

Après exécution de ce programme, nous obtenons les résultats de classement pour chaque individu de l'échantillon de validation. Nous vous présentons un aperçu des résultats obtenus dans le Table 7

**Table 7:** Diagnostic des observations pour l'échantillon de validation.

la classe effective	X1	X2	X3	X4	X5	Z	D1	D2	min (D1D2)	classe prévue	confrontation class effective class prédite
1	1	1	2	1	2	-0,755	0,036	1,785	0,036	1	VRAI
1	2	2	2	2	2	1,099	1,818	0,069	0,069	2	FAUX
2	1	2	2	1	2	0,914	1,633	0,116	0,116	2	VRAI
1	1	1	2	2	2	-1,493	0,774	2,523	0,774	1	VRAI
1	2	1	1	1	2	-0,957	0,238	1,987	0,238	1	VRAI
2	2	2	1	1	3	0,297	1,016	0,733	0,733	2	VRAI
2	2	1	2	2	2	-0,57	0,149	1,6	0,149	1	FAUX
1	1	1	1	1	2	-1,88	1,161	2,91	1,161	1	VRAI

FAUX= observation mal classée

VRAI= observation bien classée

Le Table reprend la fonction score (Z) pour l'échantillon de validation, constitué de 29 dossiers douteux. En fait, on compare le classement réservé pour chaque dossier, dans le modèle avec ses 5 variables, avec son classement effectif. Ceci dit, on conclue que c'est un bon classement, si seulement si, la classe effective est égale à la classe prévue par le modèle (c.-à-d. Résultat = VRAI, on reprenant la terminologie du Table).

Comme on peut le constater, le Table indique que le modèle nous donne un taux de bon classement global de **79,31%** pour l'échantillon de validation constitué de 29 dossiers, ainsi :

- Sur les 11 dossiers classés à priori "frauduleux" : 10 sont bien classés, seul 1 dossier est mal classé
- Sur les 18 dossiers classés à priori "matérialisés" : 13 sont bien classés, soit un taux de bon classement de 72,22 %, et 5 sont mal classés.

#### 4. CONCLUSION

Le but de cette étude a été d'essayer d'effectuer une classification des dossiers douteux transmis par la SAA vers l'agence ALFA en groupes (les dossiers frauduleux et les dossiers matérialisés) et de prévoir la classe d'un nouveau dossier entrant. Les résultats obtenus par l'analyse discriminante sont satisfaisants, du fait que les taux de bon classement pour l'échantillon est de **80,6%**, il dépasse les **50%**. Il est de **87,8%** dans la classe de «dossiers frauduleux». Néanmoins, il y a lieu de signifier que l'analyse n'a retenu que cinq variables. Cela est dû au fait que les autres variables apportent peu d'informations au modèle. La qualité des résultats a probablement été affectée par le nombre de dossiers qui ne sont pas représentatifs de la population générale, résultant notamment de la démarche purement empirique de sondage adoptée et éventuellement de l'année de référence 2013. Pour atteindre notre but, nous avons mené une étude statistique sur la fraude visant à classer les dossiers douteux en deux classes "frauduleux" ou "non frauduleux". L'originalité de ce travail a consisté à étudier les indices liés à l'acte frauduleux de l'assuré. À partir de cette analyse, nous avons pu dégager les conclusions suivantes :

- Notre modèle statistique, basé sur la méthode ADL, nous fournit des résultats très satisfaisant.
- Parmi les indices laissant présumer qu'un dossier est frauduleux, nous pouvons citer le "*fou et contradiction sur les circonstances du sinistre*", "*les points de chocs en contradiction*" et "*Préexistence d'un sinistre non réparé*".
- Il existe des problèmes implicites à la fraude à l'assurance, auxquels, les assureurs doivent apporter des réponses, entre autres, l'insuffisance de l'information et le problème d'incitation à la fraude, où dans la majorité des cas, les assureurs se contentent de refuser de prendre en charge le sinistre.

#### BIBLIOGRAPHIE :

- [1] **BELHADJI El Bachir et DIONNE Georges** ; Développement d'un système expert de détection automatique de la fraude à l'assurance automobile ; Cahier de recherche 97-04 ; Juin 1997 ;
- [2] **BERNARD Frederic**, Contrôle interne : concepts, réglementation ;
- [3] **BORGI Sameh**, Une Analyse Économique et Expérimentale de la Fraude à l'Assurance et de l'Audit, THESE en vue d'obtenir le grade de Docteur de l'Université de Paris I.
- [4] **BUJOLD Louis, DIONNE Georges et GAGNE Robert** ; Assurance valeur à neuf et vols d'automobiles : une étude statistique ; Cahier de recherche 97-01 ; Février 1997
- [5] **CARROT Georges**, Lutte contre la fraude à l'assurance, 1993 ;
- [6] **CHARBONNIER Jacques**, L'assurance du risque automobile, contrôle et assurance, 2012 ;
- [7] **DIONNE Georges** ; Fraude ou risque moral, Analyse économique de la fraude ; 1993 ;
- [8] **LANDEL James, NAMIN Lionel**, Manuel de l'assurance automobile 3<sup>ème</sup> édition, L'argus de l'assurance, 2003 ;
- [9] **REVOL Sylvie**, Déclarations frauduleuses de risque et de sinistre en assurance IARD, PU d'Aix-Marseille, 1984