Application of Discriminant Analysis on Romanian Insurance Market

■

Constantin Anghelache
Dan Armeanu
Academy of Economic Studies, Bucharest

Abstract. *Discriminant analysis is a supervised learning technique that can be used in order to determine which variables are the best predictors of the classification of objects belonging to a population into predetermined classes. At the same time, discriminant analysis provides a powerful tool that enables researchers to make predictions regarding the classification of new objects into predefined classes. The main goal of discriminant analysis is to determine which of the N descriptive variables have the most discriminatory power, that is, which of them are the most relevant for the classification of objects into classes. In order to classify objects, we need a mathematical model that provides the rules for optimal allocation. This is the classifier. In this paper we will discuss three of the most important models of classification: the Bayesian criterion, the Mahalanobis criterion and the Fisher criterion. In this paper, we will use discriminant analysis to classify the insurance companies that operated on the Romanian market in 2006. We have selected a number of eigth (8) relevant variables: gross written premium (GR_WRI_PRE), net mathematical reserves (NET_M_PES), gross claims paid (GR_CL_PAID), net premium reserves (NET_PRE_RES), net claim reserves (NET_CL_RES), net income (NE—_INCOME), share capital (SHARE_CAP) and gross written premium ceded in Reinsurance (GR_WRI_PRE_CED). Before proceeding to discriminant analysis, we performed cluster analysis on the initial data in order to identify classes (clusters) that emerge from the data.*

Key words: discriminant analysis; classifier; classification cost; prediction Fisher classifier; Bayesian classifier; Mahalanobis classifier; insurance.

■

JEL Codes: G22.
REL Codes: 11B, 11C.

In order to understand how discriminant analysis operates, first we must state the classification problem. Consider a finite population P whose objects are described by N variables. The finite population is called training set (or learning set) and the variables are called descriptive variables. These variables are called predictive variables. Let $v_1, v_2, \ldots, v_N$ be the descriptive variables and $v'_1, v'_2, \ldots, v'_n$ be the predictive variables. Therefore we have

$$\{v'_1, v'_2, \ldots, v'_n\} \subseteq \{v_1, v_2, \ldots, v_N\} \quad (1)$$

Generally speaking, the classification problem requires an algorithm to identify the criteria according to which objects are assigned to classes. Recall the population P we defined earlier. Let us now consider that P is partitioned into T classes $p_1, p_2, \ldots, p_T$ called initial classes. The initial classes satisfy the properties:

$$p_i \subset P \qquad i = 1,2,\ldots,T \quad (2)$$

$$\bigcup_{i=1}^{T} p_i = P \quad (3)$$

It is important to note that we do not require the initial classes to be disjoint subsets of P:

$$p_i \cap p_j \neq \Phi \qquad i \neq j \quad (4)$$

As mentioned previously, discriminant analysis is a prediction tool. The core of this technique consists of determining an efficient way of partitioning the training set P into T disjoint classes (subsets) called predictive classes. Let $\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_T$ be the predictive classes. We have

$$\tilde{p}_i \subseteq P \qquad i = 1,2,\ldots,T \quad (5)$$

$$\bigcup_{i=1}^{T} \tilde{p}_i = P \quad (6)$$

$$\tilde{p}_i \cap \tilde{p}_j = \Phi \qquad i \neq j \quad (7)$$

It follows immediately from (7) that, in general, the predictive class $\tilde{p}_i$ and the initial class p_i are different:

$$\tilde{p}_i \neq \tilde{p}_j \quad (8)$$

This happens because the predictive classes result from the truncation of the initial classes. As such, the predictive class $\tilde{p}_i$ represents a subset of the corresponding initial class p_i:

$$\tilde{p}_i \subseteq p_i \qquad i = 1,2,\ldots,T \quad (9)$$

Of course, perfect classification requires that

$$\tilde{p}_i = p_i \qquad i = 1,2,\ldots,T \quad (10)$$

In order to classify objects, we need a mathematical model that provides the rules for optimal allocation. This is the classifier. In this paper we will discuss three of the most important models of classification:

- the Bayesian criterion;
- the Mahalanobis criterion;
- the Fisher criterion.

The Bayesian criterion is based on the minimization of classification costs. A correct classification has cost zero and an incorrect classification has cost c:

$$C(\tilde{p}_j, p_i) = \begin{cases} 0, & \tilde{p}_j = p_i \\ c, & \tilde{p}_j \neq p_i \end{cases} \quad (11)$$

where $C(\tilde{p}_j, p_i)$ is the cost generated by the classification into class $\tilde{p}_j$ of an object actually belonging to class p_i and c is a positive constant. Obviously, the perfect classification requires that the predictive and the initial classes be identical.

For ease of computation we will make the following notations:

- $P(\tilde{p}_j, p_i)$ = the probability of classification into class $\tilde{p}_j$ of an object actually belonging to class p_i;

- $f_{p_i}(x)$ = the probability density of objects belonging to class p_i (1). x denotes a vector of m values of variables that define the objects in the population;
- R_i = the subset of R^m in which the vectors that define objects belonging to class p_i take values;
- f (x) = the unconditional probability density of objects.

According to probability theory, can be computed using the formula:

$$P(\tilde{p}_j, p_i) = \int_{R_j} f_{p_i}(x)\,dx \qquad (12)$$

For each class p_i we define the cost of classification ($C(p_i)$) as follows:

$$C(p_i) = \sum_{j=1}^{T} C(\tilde{p}_j, p_i) \times P(\tilde{p}_j, p_i) \qquad (13)$$

The expected total cost of classification (C) is:

$$C = \sum_{i=1}^{T}\sum_{j=1}^{T} C(\tilde{p}_j, p_i) \times P(\tilde{p}_j, p_i) \times P(p_i) \qquad (14)$$

where $P(p_i)$ is the *a priori* probability of occurrence for class p_i. Unless further information is provided, we can consider

$$P(p_1) = P(p_2) = \ldots = P(p_i) = \ldots = P(p_T) = \frac{1}{T},$$

which means that all classes have equal probabilities of occurrence.

We can rewrite (14) as

$$C = \sum_{i=1}^{T} P(p_i) \times \left[\sum_{j=1}^{T} C(\tilde{p}_j, p_i) \times P(\tilde{p}_j, p_i)\right] \qquad (15)$$

Taking into account (12), (15) becomes:

$$C = \sum_{i=1}^{T} P(p_i) \times \left[\sum_{j=1}^{T} C(\tilde{p}_j, p_i) \times \int_{R_j} f_{p_i}(x)\,dx\right] \qquad (16)$$

which can be written as

$$C = \sum_{j=1}^{T} \int_{R_j} \left[\sum_{i=1}^{T} P(p_i) \times C(\tilde{p}_j, p_i) \times f_{p_i}(x)\right] dx \qquad (17)$$

where $\int_{R_j}\left[\sum_{i=1}^{T} P(p_i) \times C(\tilde{p}_j, p_i) \times f_{p_i}(x)\right] dx$ denotes the cost of classification into class $\tilde{p}_j$ of all the objects actually belonging to class p_i.

Now define

$$S_j(x) = \sum_{i=1}^{T} P(p_i) \times C(\tilde{p}_j, p_i) \times f_{p_i}(x).$$

Therefore, we have:

$$C = \sum_{j=1}^{T} \int_{R_j} S_j(x)\,dx \qquad (18)$$

As stated before, the classification rule is given by the minimum cost principle. Given the definition of $S_j(x)$, R_j can be expressed as

$$R_j = \{x \in R_P \mid S_j(x) - S_i(x) < 0, (\forall)\, j \neq i\} \qquad (19)$$

where $S_j(x) - S_i(x)$ represents the ecuation of the separation surface between classes p_i and $\tilde{p}_j$.

It can be shown that the minimization of classification costs is achieved when each object from the initial population is allocated to the class which has the greatest *a posteriori* probability of occurrence. The *a posteriori* probabilities are computed using Bayes' theorem:

$$P(p_i \mid x) = \frac{f_{p_i}(x) \times P(p_i)}{f(x)} \qquad (20)$$

where $P(p_i \mid x)$ is the *a posteriori* probability of occurrence for class p_i given x.

Let us consider again a finite population P which is partitioned into T classes. The main idea underlying the Mahalanobis classifier is the distance between the centroids of the classes and the objects subject to classification. The Mahalanobis criterion requires that each class comprise the objects that are closest to the centroid of the class in terms of Mahalanobis distance. More formally, the algorithm of the Mahalanobis classifier consists of five steps:

STEP 1. Estimate the centroids of the T classes: $\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_T$.

STEP 2. Estimate the covariance matrix $\hat{\Sigma}$.

STEP 3. Evaluate the Mahalanobis distance ($d(x, \hat{\mu}_i)$) between every object and the T centroids. This is done using the following formula:

$$d(x, \hat{\mu}_i) = (x - \hat{\mu}_i)^t \times \hat{\Sigma}^{-1} \times (x - \hat{\mu}_i), \quad i = 1,2,...,T \tag{21}$$

STEP 4. Classify the objects according to the minimum distance principle.

STEP 5. Re-compute the centroids and repeat the algorithm until all objects will have been classified. It is important to note that step 2 is not necessary anymore.

The Fisher classifier is a simple, yet robust discrimination method based on the analysis of variance. It is well known that the main purpose of pattern recognition is the classification of objects into classes so that the between-class variance is maximized and the within-class variance is minimized. Fisher (1933) addresses this issue using linear classification functions:

$$d_i = \alpha_0^{(i)} + \alpha_1^{(i)} x_1 + \alpha_2^{(i)} x_2 + ... + \alpha_m^{(i)} x_m \tag{22}$$

where d_i is discriminant function i and $\alpha_j^{(i)}, j = 1,2,...,m$ represent the coefficients of linear combination i.

Considering $\alpha^{(i)} = \left(\alpha_0^{(i)} \quad \alpha_1^{(i)} \quad ... \quad \alpha_m^{(i)}\right)^t$ and $x = (1 \quad x_1 \quad ... \quad x_m)^t$, relation (22) can be rewritten as

$$d_i = \left(\alpha^{(i)}\right)^t \times x \tag{23}$$

The coefficients of the linear combinations will be determined bearing in mind that we have to maximize the between-class variance while minimizing the within-class variance. The covariance matrix Σ can be written as sum between the between-class covariance matrix (Σ_b) and the within-class covariance matrix (Σ_w):

$$\Sigma = \Sigma_b + \Sigma_w \tag{24}$$

Now consider a discriminant function (or discriminant variable) d and a vector of coefficients . Assuming all discriminant variables are centered, it follows immediately from (23) that the variance of variable d is:

$$\begin{aligned} VAR(d) &= E\left[\left(\alpha^t \times x\right) \times \left(\alpha^t \times x\right)^t\right] = \\ &= \alpha^t \times E\left(x \times x^t\right) \times \alpha = \alpha^t \times \Sigma \times \alpha = \\ &= \alpha^t \times \Sigma_b \times \alpha + \alpha^t \times \Sigma_w \times \alpha \end{aligned} \tag{25}$$

The coefficients of the linear combination will be determined so that the following condition is satisfied:

$$\max_{\alpha} \psi = \frac{\alpha^t \times \Sigma_b \times \alpha}{\alpha^t \times \Sigma_w \times \alpha} \tag{26}$$

Differentiating with respect to α, we have:

$$\frac{\partial \psi}{\partial \alpha} = \frac{2 \times (\Sigma_b \times \alpha) \times \left(\alpha^t \times \Sigma_w \times \alpha\right) - 2 \times (\Sigma_w \times \alpha) \times \left(\alpha^t \times \Sigma_b \times \alpha\right)}{\left(\alpha^t \times \Sigma_w \times \alpha\right)^2} \tag{27}$$

The condition $\frac{\partial \psi}{\partial \alpha} = 0$ leads to $\frac{(\Sigma_b \times \alpha) \times \left(\alpha^t \times \Sigma_w \times \alpha\right) - (\Sigma_w \times \alpha) \times \left(\alpha^t \times \Sigma_b \times \alpha\right)}{\left(\alpha^t \times \Sigma_w \times \alpha\right)^2} = 0$ (28)

Multiplying by $\alpha^t \times \Sigma_w \times \alpha$, we get:

$$\Sigma_b \times \alpha - (\Sigma_w \times \alpha) \times \psi = 0 \quad (29)$$

It follows immediately that must satisfy:

$$(\Sigma_b - \Sigma_w \times \psi) \times \alpha = 0 \quad (30)$$

or

$$(\Sigma_w^{-1} \times \Sigma_b - I_m \times \psi) \times \alpha = 0 \quad (31)$$

where I_m is the identity matrix.

Equation (31) shows that α is an eigenvector of matrix $\Sigma_w^{-1} \times \Sigma_b$. In order for α to be a non-zero vector, ψ must satisfy the characteristic equation:

$$| \Sigma_w^{-1} \times \Sigma_b - I_m \times \psi | = 0 \quad (32)$$

The maximum number of eigenvalues of matrix $\Sigma_w^{-1} \times \Sigma_b$ is m (provided the matrix is non-singular(2)). Let $\lambda_1, \lambda_2, ..., \lambda_m$ be the m eigenvalues and assume further that

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \quad (33)$$

It can be easily proven that, for each eigenvalue λ, we have:

$$\psi = \lambda \quad (34)$$

Therefore, the maximum value of ψ (the between-class variance to within-class variance ratio) corresponds to the greatest eigenvalue of matrix $\Sigma_w^{-1} \times \Sigma_b$, which is λ_1. As such, eigenvector $\alpha^{(1)}$ (corresponding to eigenvalue λ_1) defines discriminant function d_1, which has the greatest discriminatory power. Eigenvector $\alpha^{(2)}$ (corresponding to eigenvalue λ_3) defines discriminant function d_2, which has less discriminatory power than d_1, and so on.

In this paper, we will use discriminant analysis to classify the insurance companies that operated on the Romanian market in 2006. We have selected a number of eigth (8) relevant variables: gross written premium (GR_WRI_PRE), net mathematical reserves (NET_M_PES), gross claims paid (GR_CL_PAID), net premium reserves (NET_PRE_RES), net claim reserves (NET_CL_RES), net income (NE—_INCOME), share capital (SHARE_CAP) and gross written premium ceded in Reinsurance (GR_WRI_PRE_CED).

Before proceeding to discriminant analysis, we performed cluster analysis on the initial data in order to identify classes (clusters) that emerge from the data. The results are displayed in the table below:

Class	Class members
PREMIUM	ASIROM, ALLIANZ-TIRIAC, ING ASIGURARI DE VIATA
A	BCR ASIGURARI, OMNIASIG, ASIBAN
B	GENERALI, UNITA, BT ASIGURARI TRANSILVANIA, ASTRA, ARDAF
C	GARANTA, CARPATICA ASIG, ASITRANS, AIG ROMANIA
D	INTERAMERICAN, OMNIASIG VIATA, GRAWE, BCR ASIGURARI DE VIATA, AVIVA, AIG LIFE
E	THE REST OF THE COMPANIES

As we can see, we have six classes of insurance companies. PREMIUM class includes the largest and most profitable companies on the market (ALLIANZ-TIRIAC, ASIROM, ING ASIGURARI DE VIATA). Class A comprises companies with important market share and good levels of profitability (BCR ASIGURARI, OMNIASIG, ASIBAN). Class B includes companies with significant market share but weak profitability (ASTRA) or companies that incur substantial losses (GENERALI, UNITA and especially ARDAF). Class C groups companies with average market share and variable return on capital (CARPATICA ASIG, ASITRANS, GARANTA, AIG ROMANIA). Class D consists of small companies which are in the life insurance business (OMNIASIG VIATA, GRAWE, BCR ASIGURARI DE VIATA, AVIVA, AIG LIFE). Finally, class E includes small companies with weak financial performances (the rest of the companies).

The table below summarizes the role of the descriptive variables in the discriminant analysis performed:

As we can see in the header, Wilks' Lambda[3] is only 0.00015, which means that the model has significant discriminatory power. The test-statistic F is 19.605 and the p-value is below 10^{-4}, thus assuring the goodness of the model.

The second column of the table contains the Wilks' Lambda statistic computed for each descriptive variable and it shows that all variables included in the model have important discriminatory power.

The Partial Lambda statistic (computed in the third column of the table) illustrates the contribution of the variables to the classification and, in this respect, it is clear that the descriptive variables aren't very different one from another.

Discriminant analysis summary		
Wilks' Lambda: 0.00015	F-test: 19.605	p-value < 0.0000
	Wilks' Lambda	**Partial Lambda**
GR_WRI_PRE	0.000505	0.302171
NET_M_RES	0.000413	0.369350
GR_CL_PAID	0.000769	0.198484
NET_PRE_RES	0.000587	0.260160
NET_CL_RES	0.000715	0.213281
NET_INCOME	0.000510	0.299359
SHARE_CAP	0.000735	0.207650
GR_WRI_PR_CED	0.000793	0.192521

The Fisher discriminant functions are:

	d1	d2	d3	d4	d5
GR_WRI_PRE	-0.000000073293551	0.000000037616167	0.000000035455884	-0.000000021791953	-0.000000027802918
NET_M_RES	-0.000000026315923	-0.000000007664877	-0.000000008638985	0.000000007265270	0.000000010319762
GR_CL_PAID	-0.000000102115120	0.000000019705988	-0.000000005064990	0.000000009658398	0.000000042580728
NET_PRE_RES	0.000000087054118	-0.000000076412717	-0.000000032406245	0.000000084062284	0.000000005330565
NET_CL_RES	0.000000303233204	-0.000000129275626	-0.000000164373520	-0.000000159754400	-0.000000006432826
NET_INCOME	0.000000101887694	-0.000000016161207	-0.000000100967733	-0.000000059705624	-0.000000009966636
SHARE_CAP	0.000000067035005	-0.000000052824959	0.000000011426392	0.000000016347683	0.000000010384906
GR_WRI_PR_CED	0.000000083852117	-0.000000039164855	-0.000000019043765	0.000000019130481	0.000000003542834
Constant	5.098413743419430	1.885731801860240	-0.454078109241892	-0.294166354414800	0.127152340049423
Eigenval	164.986328649779000	5.695449444939720	2.102466164517110	0.789207162386037	0.062290267356944
Cum.Prop	0.950186448048853	0.982987583284832	0.995096070535591	0.999641258955379	1.000000000000000

As the previous table demonstrates, matrix $\Sigma_w^{-1} \times \Sigma_b$ has only five positive eigenvalues. The first discriminant function (corresponding to the first and greatest eigenvalue) is, by far, the most important, as it accounts for over 95% of total discrimination.

The next table contains the means of the previously defined discriminant variables:

Means of discriminant variables

Class	d1	d2	d3	d4	d5
E	5.4680	1.38377	-0.25422	-0.173957	0.148278
D	3.4618	-0.71992	0.42856	1.831199	-0.140463
C	4.3425	0.99157	-0.35342	-0.925392	-0.622381
PREMIUM	-39.7480	2.16861	0.39354	0.078882	0.002275
B	-0.2577	-3.47241	2.56004	-0.715147	0.064743
A	-7.1665	-5.02740	-3.43611	-0.213782	0.061490

The first discriminant function distinguishes PREMIUM class from the other classes of insurance companies, the second discriminant functions distinguishes companies pertaining to class A and the third discriminant function distinguishes class B companies from the others. The fourth and fifth functions have very little discriminatory power. This is normal considering the eigenvalues determined earlier.

The discriminant scores are:

Insurance company	Class	d1	d2	d3	d4	d5
ALLIANZ-TIRIAC	PREMIUM	-39.9280	0.56399	-0.01384	-1.34926	-2.12665
ASIROM	PREMIUM	-39.6342	3.25933	0.32720	0.52237	3.16470
ING ASIGURARI DE VIATA	PREMIUM	-39.6819	2.68250	0.86725	1.06353	-1.03122
ASIBAN	A	-6.8393	-6.60250	-3.92334	0.96281	0.52981
BCR ASIGURARI	A	-7.7692	-1.36025	-2.56639	-0.97737	0.29722
OMNIASIG	A	-6.8910	-7.11944	-3.81862	-0.62678	-0.64256
ARDAF	B	-0.2911	-4.47046	4.01559	-3.04192	1.42167
ASTRA	B	-0.8659	-3.72637	4.56805	-1.84197	-0.87956
BT ASIGURARI TRANSILVANIA	B	1.5743	-2.55761	0.02579	-0.36178	-0.33126
GENERALI	B	-0.3632	-2.33102	3.38964	1.98220	-2.13192
UNITA	B	-1.3426	-4.27661	0.80113	-0.31227	2.24478
AIG ROMANIA	C	6.3841	1.23344	0.50994	-0.65662	-1.52766
ASITRANS	C	4.8983	0.85383	-0.25421	-0.60909	-0.47031
CARPATICA ASIG	C	2.3389	1.26386	-1.03491	-2.85404	-0.67614
GARANTA	C	3.7485	0.61518	-0.63451	0.41819	0.18458
AIG LIFE	D	0.5759	-1.14606	-1.17700	2.40960	-1.41603
AVIVA	D	3.5399	-2.01577	3.12010	2.57199	0.62727
BCR ASIGURARI DE VIATA	D	4.8550	0.38203	-0.39716	0.23042	-0.46659
GRAWE	D	3.9777	0.06726	0.12905	0.99363	-0.60267
INTERAMERICAN	D	4.5128	-0.79149	1.10791	1.49238	0.49555
OMNIASIG VIATA	D	3.3092	-0.81549	-0.21153	3.28918	0.51968
ABC ASIGURARI	E	5.4765	1.33486	-0.14938	-0.10738	0.15478
AGRAS	E	4.8886	1.39471	-0.41105	-0.40577	0.02500
ASIMED	E	5.1375	1.80852	-0.43160	-0.26843	0.14603

Insurance company	Class	d1	d2	d3	d4	d5
ASIROM CONCORDIA	E	5.7092	1.39201	-0.39602	-0.08419	0.21027
ASITO KAPITAL	E	4.8505	1.59226	-0.23307	-0.60552	-0.01944
ATE INSURANCE	E	6.3424	0.83535	-0.11121	0.10185	0.34666
CERTASIG	E	5.2772	1.63414	-0.24463	-0.14182	0.14778
CITY INSURANCE	E	5.3289	1.57219	-0.19398	-0.19527	0.13260
CLAL ROMANIA	E	5.1891	1.37690	0.36534	0.30644	0.31293
DELTA	E	7.6640	0.77592	-1.74619	-1.10241	0.00394
DELTA ADDENDUM	E	5.5204	1.40654	-0.30207	-0.10285	0.16955
EUROASIG	E	5.2987	1.52207	-0.16043	-0.08560	0.15016
FATA ASIGURARI AGRICOLE	E	5.5884	1.39372	-0.21785	0.01291	0.09965
GERROMA	E	5.3770	1.67182	-0.40018	-0.35734	0.09080
IRASIG	E	5.2940	1.62531	-0.24090	-0.11468	0.17971
KD LIFE ASIGURARI	E	5.7951	0.99354	0.21822	0.34225	0.30791
NBG INSURANCE	E	5.0743	1.37045	-0.04783	0.02069	0.18842
OTP GARANCIA ASIGURARI	E	5.6130	0.79942	0.29161	0.23553	0.31975
R.A.I.	E	4.4668	1.79196	-0.41897	-0.75360	-0.14920

Even though all five discriminant functions could be used in our analysis, we are going to use only the first two of them, because of the following reasons:

- together they account for over 98% of the total discrimination
- they enable us to plot the discriminant scores and thus emphasize the grouping of the objects in the reduced discriminant space. The chart is shown below:

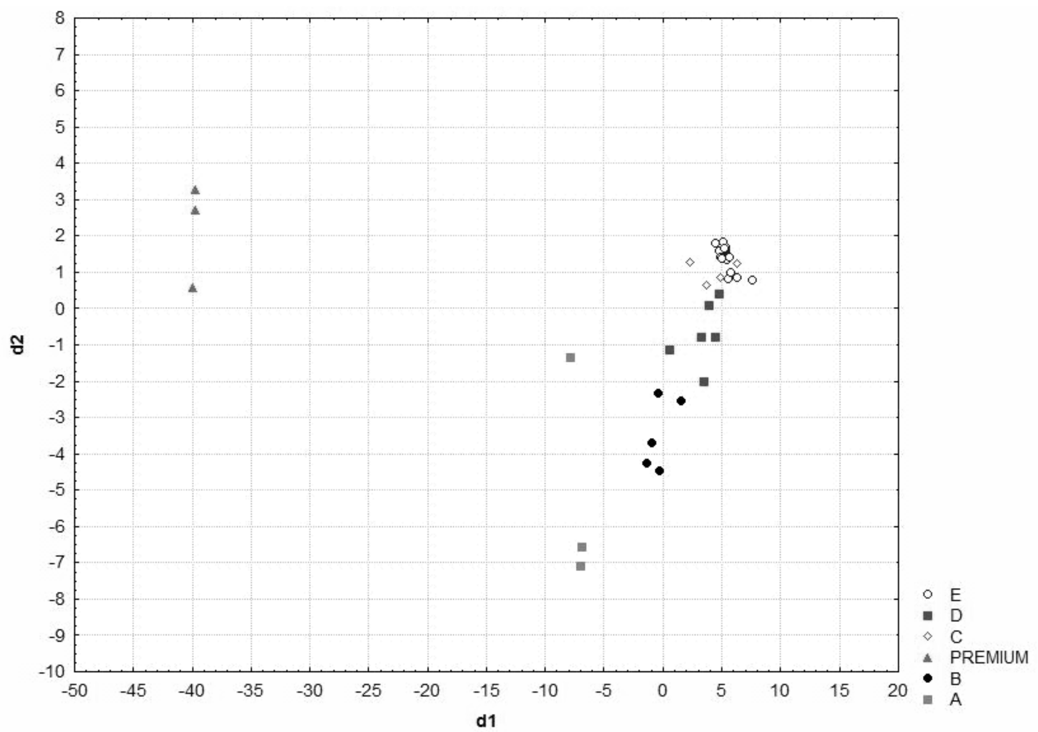

Figure 1. *Plot of discriminant scores*

The plot shows that a clear distinction can be made only between PREMIUM and A classes and the others. At the same time, it is difficult to distinguish between classes C and E; once again, we have clear proof that the first two discriminant functions are the most significant. Discriminant variable d_1 is negatively correlated with predictive variables *gross written premium*, *net mathematical reserves* and *gross claims paid*, indicating that companies with low d_1 values have a good chance of being included in the PREMIUM category. Discriminant variable d_2 is positively correlated with *gross written premium* and *gross claims paid* and negatively correlated with the other six predictive variables. Companies with negative d_2 are very likely to be allocated to class A, but they can also be classified in class B. As for classes C, D and E, they are all characterised by similar d_1 and d_2 scores, making it very difficult to distinguish between them.

As mentioned previously, discriminant analysis is also an important prediction tool: based on the data in the training set, it produces a model that can be used to classify new objects. The quality of classification can be assessed using the classification matrix:

Rows: observed classifications							
Columns: predicted classifications							
	% correct	E	D	C	PREMIUM	B	A
E	94.7368	18	0	1	0	0	0
D	83.3333	1	5	0	0	0	0
C	75.0000	1	0	3	0	0	0
PREMIUM	100.0000	0	0	0	3	0	0
B	100.0000	0	0	0	0	5	0
A	100.0000	0	0	0	0	0	3
Total	92.5000	20	5	4	3	5	3

As we can see, the total percentage of correct classification is 92.5%, which can be considered excellent. The application of the model resulted in perfect classification for objects in classes PREMIUM, A and B. The percentage of correct classification for objects in classes C, D and E is not 100% because, as explained in the previous section, it is difficult to distinguish between companies belonging to these classes. Still, over 75% of objects have been classified correctly, which proves that we developed a robust prediction tool.

The Bayesian criterion produces exactly the same classification as the Fisher linear discriminant functions. The *a posteriori* probabilities of classification are presented in the table below:

A posteriori probabilities							
Incorrect classifications are marked with *							
	Observed	**E**	**D**	**C**	**PREMIUM**	**B**	**A**
1	E	0.792934	0.001575	0.205490	0.000000	0.000000	0.000000
2	E	0.588580	0.001601	0.409820	0.000000	0.000000	0.000000
3	D	0.000001	0.999808	0.000115	0.000000	0.000076	0.000000
* 4	C	0.665418	0.000225	0.334357	0.000000	0.000000	0.000000
5	PREMIUM	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
6	B	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
7	A	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
8	E	0.730372	0.000633	0.268995	0.000000	0.000000	0.000000

A posteriori probabilities							
Incorrect classifications are marked with *							
	Observed	E	D	C	PREMIUM	B	A
9	PREMIUM	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
10	E	0.840944	0.000809	0.158247	0.000000	0.000000	0.000000
11	E	0.556741	0.000827	0.442432	0.000000	0.000000	0.000000
12	C	0.410196	0.002916	0.586888	0.000000	0.000000	0.000000
13	B	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
14	E	0.918427	0.001357	0.080216	0.000000	0.000000	0.000000
15	D	0.000005	0.999863	0.000003	0.000000	0.000129	0.000000
16	A	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
* 17	D	0.481964	0.049176	0.468860	0.000000	0.000000	0.000000
18	B	0.000026	0.369984	0.004412	0.000000	0.625578	0.000000
19	C	0.006686	0.000024	0.993291	0.000000	0.000000	0.000000
20	E	0.768181	0.001063	0.230757	0.000000	0.000000	0.000000
21	E	0.765919	0.001016	0.233064	0.000000	0.000000	0.000000
22	E	0.814868	0.008210	0.176922	0.000000	0.000000	0.000000
23	E	0.928964	0.000004	0.071033	0.000000	0.000000	0.000000
24	E	0.805603	0.001141	0.193256	0.000000	0.000000	0.000000
25	E	0.773573	0.001537	0.224890	0.000000	0.000000	0.000000
26	E	0.822372	0.001423	0.176205	0.000000	0.000000	0.000000
27	C	0.307588	0.181682	0.510730	0.000000	0.000000	0.000000
28	B	0.000000	0.000356	0.000000	0.000000	0.999644	0.000000
29	E	0.751512	0.000466	0.248022	0.000000	0.000000	0.000000
30	D	0.093725	0.741787	0.164488	0.000000	0.000000	0.000000
31	PREMIUM	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
32	D	0.015387	0.979071	0.005542	0.000000	0.000000	0.000000
33	E	0.778769	0.001113	0.220118	0.000000	0.000000	0.000000
34	E	0.882592	0.005748	0.111660	0.000000	0.000000	0.000000
35	E	0.736508	0.004175	0.259317	0.000000	0.000000	0.000000
36	A	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
37	D	0.000090	0.999873	0.000037	0.000000	0.000000	0.000000
38	E	0.839954	0.010030	0.150016	0.000000	0.000000	0.000000
* 39	E	0.412300	0.000590	0.587110	0.000000	0.000000	0.000000
40	B	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000

It is important to mention that the a priori probabilities of occurrence for each class have been chosen equal to $\frac{1}{6}$.

Finally, we get the same results using the Mahalanobis classifier:

Squared Mahalanobis distances from group centroids							
Incorrect classifications are marked with *							
Case	Observed	E	D	C	PREMIUM	B	A
1	E	0.019	12.461	2.719	2046.305	63.710	211.146
2	E	0.504	12.318	1.228	1993.989	59.173	195.833
3	D	41.249	13.993	32.130	1647.860	32.972	90.121
* 4	C	4.565	20.553	5.941	2132.020	73.078	241.186
5	PREMIUM	2080.620	1911.114	1975.152	21.885	1614.391	1134.908
6	B	109.730	81.418	93.262	1640.010	24.583	127.170
7	A	231.027	161.736	201.215	1180.493	99.058	5.307
8	E	0.368	14.470	2.366	2015.705	66.191	207.194
9	PREMIUM	2057.613	1895.710	1965.956	21.368	1621.917	1157.123
10	E	0.094	13.987	3.435	2067.657	68.428	216.276
11	E	0.685	13.710	1.145	1990.265	59.610	198.692
12	C	1.244	11.137	0.527	1996.207	53.585	190.778
13	B	108.339	73.956	89.528	1583.482	21.643	124.009
14	E	1.353	14.388	6.229	2126.631	70.153	228.250
15	D	36.137	11.775	37.254	1907.099	29.686	176.467
16	A	193.624	148.451	162.826	1049.921	92.144	20.064
* 17	D	2.068	6.633	2.123	1993.619	51.045	183.617
18	B	31.254	12.163	21.021	1730.525	11.112	94.877
19	C	19.404	30.703	9.402	1784.362	48.349	144.341
20	E	0.130	13.296	2.535	2028.061	64.943	209.448
21	E	0.070	13.320	2.450	2032.735	64.529	210.217
22	E	0.789	9.984	3.844	2020.194	59.175	208.531
23	E	9.519	34.471	14.661	2257.034	100.718	258.489
24	E	0.019	13.139	2.874	2050.368	65.777	212.208
25	E	0.079	12.521	2.550	2029.999	63.638	209.046
26	E	0.070	12.789	3.151	2056.400	66.122	214.346
27	C	5.661	6.714	4.647	1897.181	45.879	160.855
28	B	75.392	34.446	62.389	1593.027	18.566	114.242
29	E	0.170	14.939	2.387	2037.368	67.126	211.478
30	D	6.314	2.176	5.189	1917.917	40.029	165.044
31	PREMIUM	2064.088	1894.236	1966.324	22.241	1619.088	1157.748
32	D	10.731	2.424	12.773	1970.870	37.447	178.430
33	E	0.119	13.220	2.646	2029.580	65.054	209.782
34	E	0.805	10.873	4.940	2075.779	63.273	218.008
35	E	0.254	10.600	2.342	2009.928	59.258	202.337
36	A	240.991	174.872	206.483	1186.913	100.903	7.671
37	D	22.532	3.906	24.322	1874.675	44.608	151.262
38	E	0.913	9.768	4.358	2059.692	58.884	211.488
*39	E	1.881	14.980	1.175	1956.726	59.217	191.536
40	B	92.415	54.617	78.466	1530.325	18.295	65.674

To sum up, discriminant analysis provides an important classification tool. Regardless of the method used, the application of the model produced 92.5% correct classification, which underlines the robustness of the analytical approach. It is also worth mentioning that discriminant analysis offers a powerful predictive tool which can be used to describe future developments in the Romanian insurance market.

Notes

(1) We can also write $f_{p_i}(x) = f(x \mid x \in p_i)$. $f_{p_i}(x)$ is a conditional probability density.

(2) Recall that is a positively defined and symmetrical matrix.

(3) Computed as sum

$$\frac{\textit{sum of squares of errors between classes}}{\textit{total sum of squares of errors}}.$$

The closer its value is to zero, the more discriminatory power the model has; the closer its value is to one, the less discriminatory power the model has.

References

Armeanu, D. (2006). *Managementul riscului în asigurări*, seria business, Editura Universităţii Romano-Britanice, Bucureşti

Discriminant Function Analysis, http://www2.chass.ncsu.edu/garson/pa765/discrim.htm

Poulsen, J., French, A., *Discriminant Function Analysis*", http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf

Ruxanda, Gh., *Analiza multidimensională a datelor*, Master Baze de Date – Suport pentru Afaceri

Ruxanda, Gh. (2001). *Analiza Datelor*, Editura ASE, Bucureşti

Simar, L. (2004). *Applied Multivariate Statistical Analysis*, Springer

Spircu, Liliana (2006). *Analiza Datelor: Aplicaţii Economice*, Editura ASE, Bucureşti

Welling, Max, *Fisher Linear Discriminant Analysis*", http://www.cs.huji.ac.il/~csip/Fisher-LDA.pdf